TRAFFIC ENGINEERING

Roger P. Roess
Elena S. Prassas
William R. McShane

Pearson

FIFTH EDITION

Traffic Engineering

Fifth Edition

Roger P. Roess, Ph.D.

Professor Emeritus

New York University Tandon School of Engineering

Elena S. Prassas, Ph.D.

Associate Professor of Transportation Engineering

New York University Tandon School of Engineering

William R. McShane, Ph.D., PE, PTOE

Professor Emeritus

New York University Tandon School of Engineering

Chairman Emeritus, KLD Engineering, P.C.

President/CEO, KLD Associates, Inc.

Pearson

330 Hudson Street, NY NY 10013

Senior Vice President Courseware Portfolio Management: Marcia J. Horton

Director, Portfolio Management: Engineering, Computer Science & Global Editions: Julian Partridge

Executive Portfolio Manager: Holly Stark

Portfolio Management Assistant: Amanda Perfit

Product Marketing Manager: Yvonne Vannatta

Field Marketing Manager: Demetrius Hall

Marketing Assistant: Jon Bryant

Managing Content Producer: Scott Disanno

Content Producer: Amanda Brands

Project Manager: Louise Capulli

Inventory Manager: Bruce Boundy

Rights and Permissions Manager: Ben Ferrini

Manufacturing Buyer, Higher Ed, Lake Side Communications, Inc. (LSC): Maura Zaldivar-Garcia

Cover Designer: Black Horse Designs

Cover Illustration/Photo(s): JG Photography/Alamy Stock Photo

Full-Service Project Management: Integra Software Services Pvt. Ltd./Ashwina Ragunath

Printer/Binder: LSC Communications, Inc.

Credits and acknowledgments borrowed from other sources and reproduced, with permission, in this textbook appear on appropriate page within text.

Library of Congress Cataloging-in-Publication Data

Names: Roess, Roger P., author. | Prassas, Elena S., author. | McShane, William R., author.

Title: Traffic engineering/Roger P. Roess, Elena S. Prassas, William R. McShane.

Description: Fifth edition. | Hoboken, NJ : Pearson Education, 2019. | Includes bibliographical references.

Identifiers: LCCN 2017054076| ISBN 9780134599717 | ISBN 0134599713

Subjects: LCSH: Traffic engineering—United States.

Classification: LCC HE355 .M43 2018 | DDC 388.3/120973—dc23 LC record available at https://lccn.loc.gov/2017054076

1   18

ISBN-10:   0-13-459971-3

ISBN-13: 978-0-13-459971-7

Contents

Preface

The transportation system is the nation's lifeblood circulation system. Our complex system of roads and highways, railroads, airports and airlines, waterways, and urban transit systems provides for the movement of people and goods to and from the most remote outposts of the nation. It is the transportation network which allows for the concentrated production of food, goods, energy, and other material in an economically optimal manner, knowing that the systems needed to collect raw materials, and distribute final products throughout the nation are in place.

Traffic engineering deals with several critical elements of the transportation system: our streets and highways, and the transportation services they support. Because the transportation system is such a critical part of our infrastructure, the traffic engineer is involved in a wide range of issues, often in a very public setting, and must bring a broad range of skills to the table. Traffic engineers must have an appreciation for and understanding of planning, design, management, construction, operation, control, and system optimization. All of these functions involve traffic engineers at some level.

This text focuses on the key engineering skills required to practice traffic engineering in a broad setting. This is the fifth edition of the textbook, and it includes the latest standards and criteria of the *Manual on Uniform Traffic Control Devices* (2009, as updated through May 2012), the *Policy on Geometric Design of Highways and Streets* (2011), the *Highway Capacity Manual* (2016), the *Highway Safety Manual* (2010, with 2014 Supplement), and other critical documents. While this edition uses the latest versions of basic references, students must be aware that all of these are periodically updated, and (at some point), versions not available at this writing will become available, and should be used.

The text is organized into four major functional parts:

- Part I – Basic Concepts and Characteristics
- Part II – Traffic Studies and Programs

- Part III – Interrupted Flow Facilities: Design, Control, and Level of Service
- Part IV – Uninterrupted Flow Facilities: Design, Control, and Level of Service

The text is appropriate for an undergraduate survey course in traffic engineering, or for more detailed graduate (or undergraduate) courses focusing on specific aspects of the profession. A survey course might include all of Part I, a selection of chapters from Part II, and a few chapters focusing on signal design and/or capacity and level of service analysis. Over the years, the authors have used the text for graduate courses on Traffic Studies and Characteristics, Traffic Control and Operations, and Highway Capacity and Level of Service Analysis. Special courses on highway traffic safety and geometric design have also used this text.

Some chapters, particularly Traffic Impact and Mitigation Studies, are organized around case studies. These should only be used in a more advanced course with an instructor who is familiar with the many tools referenced.

What's New in This Edition

This edition of the textbook adds a significant amount of material, including, but not limited to:

1. More than 50% of the homework problems (and an available solutions manual) are new for most chapters.
2. New material on unsignalized intersections, roundabouts, alternative intersections, interchanges, operation and analysis of facilities, and more.
3. Material on signalized intersections, signal design and timing, and signal hardware has been updated and extended.
4. Material from the latest editions of key traffic engineering references is included, as noted previously.
5. Links to a number of new Web sites which students and instructors

will find valuable.

There are some additional revisions. There is no overview chapter on statistics; undergraduate engineering degrees now require coursework in statistics. We have included supporting material on statistical analyses within the applications in which they are used. An overview chapter can't cover everything, and it should be expected that modern engineering students have been exposed to this material. The text still provides details on a number of capacity and level of service applications. The *2016 HCM*, however, has over 3,000 pages of printed and electronic material, and many complicated analyses can only be presented in outline or overview form. There is material from the *Highway Safety Manual*, but complete analysis material is included for only one type of application. Again, there is simply too much material to include more than an example of its procedures and applications.

We hope that students and instructors will continue to find this text useful in learning about the profession of traffic engineering, and about many of its key components. As in the past, comments are always welcome.

Roger P. Roess

Elena S. Prassas

William R. McShane

Part I Basic Concepts and Characteristics

Chapter 1 Introduction

1.1 Traffic Engineering as a Profession

Traffic engineering has been defined in many ways over the years. It is currently described by the Institute of Transportation Engineers (ITE) in the following words [1]:

> A branch of civil engineering, traffic engineering concerns the safe and efficient movement of people and goods along roadways. Traffic flow, road geometry, sidewalks, bicycle facilities, shared lane markings, traffic signs, traffic lights, and more—all of these elements must be considered when designing public and private sector transportation solutions.

This description represents an ever-broadening profession that includes multimodal transportation systems and options, many of which use streets and highways. It also highlights that the simple objectives of safety and efficiency have become ever-more complex.

Historically, traffic engineering begins with early road-builders, which have existed since ancient times. The ancient Romans were prolific road-builders. The focus was on the physical and structural design of roadways. Civil engineering, with its focus on physical infrastructure, became the traditional home for traffic engineering.

With the advent of the automobile and its growing influence on modern transportation, the traffic engineer's purview was extended to the areas of traffic control and operations. Modern traffic engineering involves complex technologies employed to control and operate roadway facilities and networks, and touches upon virtually all of the fundamental engineering disciplines. While not technically "traffic engineering," the associated profession of transportation planning is integral, focusing on various aspects of human behavior and their impacts on travel, the forecasting of transportation demand, and the development and assessment of plans to accommodate society’s travel and mobility needs.

1.1.1 Safety: The Primary Objective

The principal goal of the traffic engineer remains the provision of a safe system for highway traffic. This is no small task. Traffic fatalities peaked at 55,600 in 1972. Improvements in vehicles, driver training, roadway design, and traffic control have helped bring that number significantly down beginning in the 1980s. The number of traffic fatalities has been less than 40,000 per year since 2008, with a low of 32,744 posted in 2014 [2].

Unfortunately, 2015 and 2016 fatalities show that the number is rising again. Traffic fatalities rose by 8.4% to 35,485 in 2015. Fatalities for 2016 show a further increase of 5.6%, resulting in 37,461 fatalities [3]. The National Safety Council (NSC) had predicted that fatalities would actually be more than 40,000 for 2016 [4]. The NSC uses a different basis to define traffic fatalities than the National Highway Transportation Administration (NHTSA), which may account for some of the discrepancy.

While total highway fatalities per year have fluctuated, accident rates based on vehicle-miles traveled have consistently declined. That is because U.S. motorists generally drive more miles each year, with the exception of 2008 and 2009, which saw a small reduction due to poor economic conditions. The increasing number of annual vehicle-miles traveled produces a declining fatality rate. The fatality rate reached its lowest point in memory in 2014, at 1.08 fatalities per 100 million vehicle-miles traveled (100 MVM). In 2015, the rate increased to 1.15, and in 2016 to 1.18.

Improvements in fatality rates reflect a number of trends, many of which traffic engineers have been instrumental in implementing. Stronger efforts to remove dangerous drivers from the road have yielded significant dividends in safety. Driving under the influence (DUI) and driving while intoxicated (DWI) offenses are more strictly enforced, and licenses are suspended or revoked more easily as a result of DUI/DWI convictions, poor accident record, and/or poor violations record. Vehicle design has greatly improved (encouraged by several acts of Congress requiring certain improvements). Today’s vehicles feature padded dashboards, collapsible steering columns, seat belts with shoulder harnesses, air bags (some vehicles now have as many as eight), and antilock braking systems. Collision avoidance systems and other driver aids now exist in a growing number of vehicles. Highway design has improved through the development and use of advanced barrier systems for medians and roadside areas. Traffic control systems communicate better and faster, and surveillance systems can alert authorities to accidents and breakdowns in the system.

The increase in fatalities over the last 2 years has generally been attributed to higher incidence of "distracted driving." The modern vehicle has many more distractions for the driver, despite all of the technological advances made to assist drivers. Electronic devices, including Bluetooth phones and other devices, a vast variety of listening options, and an increasingly busy external environment tend to lure the driver's attention from his or her primary task. Nearly 40,000 people per year still die in traffic accidents. The objective of safe travel is always number one and is never finished for the traffic engineer.

1.1.2 Other Objectives

Traffic engineers have other objectives to consider.

- Travel time
- Comfort
- Convenience
- Economy
- Environmental compatibility

Most of these are self-evident desires of the traveler. Most of us want our trips to be fast, comfortable, convenient, cheap, and in harmony with the environment. All of these objectives are also relative and must be balanced against each other and against the primary objective of safety.

While speed of travel is much to be desired, it is limited by transportation technology, human characteristics, and the need to provide safety. Comfort and convenience are generic terms that mean different things to different people. Comfort involves the physical characteristics of vehicles and roadways, and is influenced by our perception of safety. Convenience relates more to the ease with which trips are made and the ability of transport systems to accommodate all of our travel needs at appropriate times. Economy is also relative. There is little in modern transportation systems that can be termed "cheap." Highway and other transportation systems involve massive construction, maintenance, and operating expenditures, most of which are provided through general and user taxes and fees. Nevertheless, every engineer, regardless of discipline, is called upon to provide the best possible systems for the money.

Harmony with the environment is a complex issue that has become more important over time. All transportation systems have some negative impacts on the environment. All produce air and noise pollution in some forms, and all utilize valuable land resources. In many modern cities, transportation systems utilize as much as 25% of the total land area. "Harmony" is achieved when transportation systems are designed to minimize negative environmental impacts, and where system architecture provides for aesthetically pleasing facilities that "fit in" with their surroundings.

The traffic engineer is tasked with all of these goals and objectives and with making the appropriate trade-offs to optimize both the transportation systems and the use of public funds to build, maintain, and operate them.

1.1.3 Responsibility, Ethics, and Liability in Traffic Engineering

The traffic engineer has a very special relationship with the public at large. Perhaps more than any other type of engineer, the traffic engineer deals with the daily safety of a large segment of the public. Although it can be argued that any engineer who designs a product has this responsibility, few engineers have so many people using their product so routinely and frequently and depending upon it so totally. Therefore, the traffic engineer also has a special obligation to employ the available knowledge and state of the art within existing resources to enhance public safety.

The traffic engineer also functions in a world in which a number of key participants do not understand the traffic and transportation issues or how they truly affect a particular project. These include elected and appointed officials with decision-making power, the general public, and other professionals with whom traffic engineers work on an overall project team effort. Because all of us interface regularly with the transportation system, many overestimate their understanding of transportation and traffic issues. The traffic engineer must deal productively with problems associated with naïve assumptions, plans, and designs that are oblivious to transportation and traffic needs, oversimplified analyses, and understated impacts.

Like all engineers, traffic engineers must understand and comply with professional ethics codes. Primary codes of ethics for traffic engineers are those of the National Society of Professional Engineers and the American Society of Civil Engineers. The most up-to-date versions of each are available online. In general, good professional ethics requires that traffic engineers work only in their areas of expertise; do all work completely and thoroughly; be completely honest with the general public, employers, and clients; comply with all applicable codes and standards; and work to the best of their ability. In traffic engineering, the pressure to understate negative impacts of projects, sometimes brought to bear by clients who wish a project to proceed and employers who wish to keep clients happy, is a particular concern. As in all engineering professions, the pressure to minimize costs must give way to basic needs for safety and reliability.

Experience has shown that the greatest risk to a project is an incomplete analysis. Major projects have been upset because an impact was overlooked or analysis oversimplified. Sophisticated developers and experienced professionals know that the environmental impact process calls for a fair and complete statement of impacts and a *policy decision by the reviewers* on accepting the impacts, given an overall good analysis report. The process does not require zero impacts; it does, however, call for clear and complete disclosure of impacts so that policy makers can make informed decisions. Successful challenges to major projects are almost always based on flawed analysis, not on disagreements with policy makers. Indeed, such disagreements are not a valid basis for a legal challenge to a project. In the case of the Westway Project proposed in the 1970s for the west side of Manhattan, one of the bases for legal challenge was that the impact of project construction on striped bass in the Hudson River had not been properly identified or disclosed. In particular, the project died due to overlooking the impact on the reproductive cycle of striped bass in the Hudson River. While this topic was not the primary concern of the litigants, it was the legal “hook” that caused the project to be abandoned.

The traffic engineer also has a responsibility to protect the community from liability by good practice. There are many areas in which agencies charged with traffic and transportation responsibilities can be held liable. These include (but are not limited to) the following:

- Placing control devices that do not conform to applicable standards for their physical design and placement.

- Failure to maintain devices in a manner that ensures their effectiveness; the worst case of this is a “dark” traffic signal in which no indication is given due to bulb or other device failure.

- Failure to apply the most current standards and guidelines in making decisions on traffic control, developing a facility plan or design, or conducting an investigation.

- Implementing traffic regulations (and placing appropriate devices) without the proper legal authority to do so.

A historic standard has been that “due care” be exercised in the preparation of plans, and that determinations made in the process be reasonable and “not arbitrary.” It is generally recognized that professionals must make value judgments, and the terms “due care” and “not arbitrary” are continually under legal test.

The fundamental ethical issue for traffic engineers is to provide for the public safety through positive programs, good practice, knowledge, and proper procedure. The negative (albeit important) side of this is the avoidance of liability problems.

1.2 Transportation Systems and Their Function

Transportation systems are a major component of the U.S. economy and have an enormous impact on the shape of the society and the efficiency of the economy in general. Table 1.1 illustrates some key statistics for the U.S. highway system for 2015 [1].

Table 1.1: Important Statistics on U.S. Highways

Statistic	2015 Value
Miles of public roadway	4.19 million
Vehicle-miles traveled	3.11 trillion
Total population of the United States	321 million
Licensed drivers	218 million
Registered vehicles	256 million
Fatalities	35,485

Table 1.1: Full Alternative Text

America moves on its highways. While public transportation systems are of major importance in large urban areas such as New York, Boston, Chicago, and San Francisco, it is clear that the vast majority of person-travel as well as a large proportion of freight traffic is entirely dependent on the highway system.

The system is a major economic force in its own right: Over $150 billion per year is spent by state and local governments on highways. The vast majority of disbursements applied to highways and streets is made by state and local governments. The federal government provides massive funding through aid to the states. The federal government spends directly on federally owned lands, such as military bases, national parks, national forests, and Indian (Native American) reservations.

The revenue to support these expenditures comes from a variety of sources. Federal aid is disbursed from the *Highway Trust Fund*, which is funded by the federal excise tax on fuels and other highway-related items, as well as from the federal general fund. State and local funds come from state and local taxes on fuels, and from state and local general funds. Table 1.2 summarizes the sources of national highway expenditures for the year 2011 [5].

Table 1.2: Revenue Sources for 2011 Highway Disbursements

Revenue ($ billion)	Source	Percent of Total
41.2	State & local motor fuel taxes	26.9
28.0	Federal motor fuel & other excise taxes	18.3
23.2	State license fees	15.2
12.7	Tolls and other local user fees	8.3
105.1	***Subtotal road-user taxes***	**68.7**
30.0	State & local general fund allocations	19.6
18.0	Federal general fund allocations & deficit financing	11.7
48.0	***Subtotal general funds***	**31.3**
153.1	**TOTAL**	**100.0**

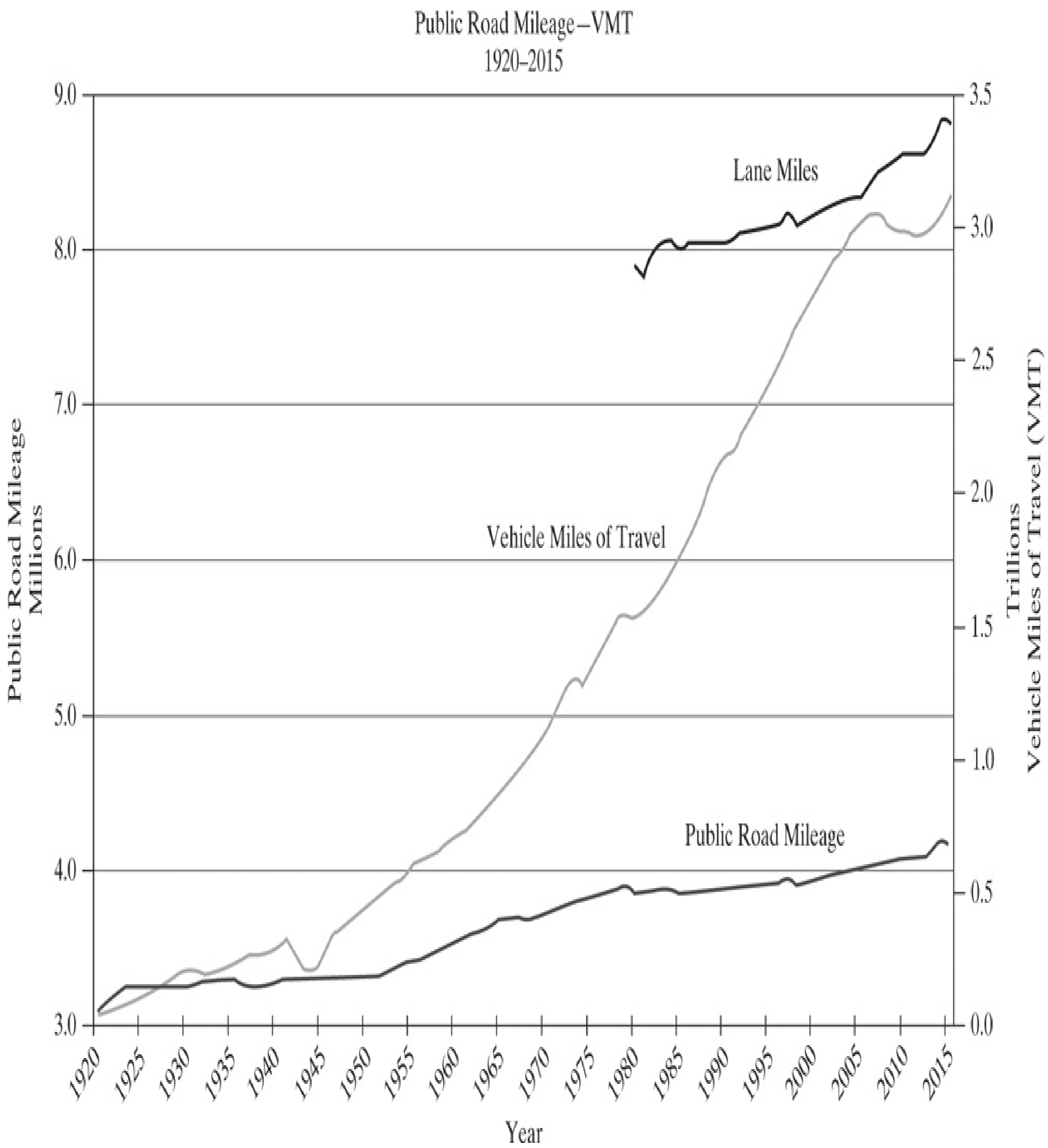
Table 1.2: Full Alternative Text

When the United States embarked on the *National System of Interstate and Defense Highways* in 1956, it created the *Highway Trust Fund*, with a host of federal road-user excise taxes to fund it. The theory was that the users of these new facilities would be the primary beneficiaries, and should therefore pay the lion's share of their cost.

Over the years, the general view of road-user taxes has changed. Many federal excise taxes were dropped in the mid-1970s—such as excise taxes on vehicle purchases, tires, oil, and parts. The federal fuel tax has not been raised since 1993. While the need for investment in highway and transportation infrastructure has greatly increased, more fuel-efficient cars have actually reduced federal fuel tax revenues. A political debate over raising the tax has been ongoing for almost a decade. On the one hand, more money for investment in this key infrastructure is badly needed. On the other hand, it is recognized that a user tax system is fairly regressive, one that hits those with lower incomes the hardest.

The American love affair with the automobile has grown consistently since the 1920s, when Henry Ford's Model T made the car accessible to the average wage earner. This growth has survived wars, gasoline embargoes, depressions, recessions, and almost everything else that has happened in society. As seen in Figure 1.1, annual vehicle-miles traveled reached the 1 trillion mark in 1968 and the 2 trillion mark in 1987, and is now over 3 trillion vehicle miles per year.

Figure 1.1: Public Highway Mileage and Annual Vehicle-Miles Traveled in the United States, 1920–2015



(Source: *Highway Statistics 2015*, Federal Highway Administration, U.S. Department of Transportation, Washington, D.C., 2015, Table VMT 421C.)

Figure 1.1: Full Alternative Text

This growth pattern is one of the fundamental problems to be faced by traffic engineers. Given the relative maturity of our highway systems and the difficulty faced in trying to add system capacity, particularly in urban areas, the continued growth in vehicle-miles traveled leads directly to increased congestion on our highways. The inability to simply build additional capacity to meet the growing demand creates the need to address alternative modes, fundamental alterations in demand patterns, and management of the system to produce optimal results.

1.2.1 The Nature of Transportation Demand

Transportation demand is directly related to land-use patterns and to available transportation systems and facilities. Figure 1.2 illustrates the fundamental relationship, which is circular and ongoing. Transportation demand is generated by the types, amounts, and intensity of land use, as well as its location. The daily journey to work, for example, is dictated by the locations of the worker's residence and employer and the times that the worker is on duty.

Figure 1.2: The Nature of Transportation Demand

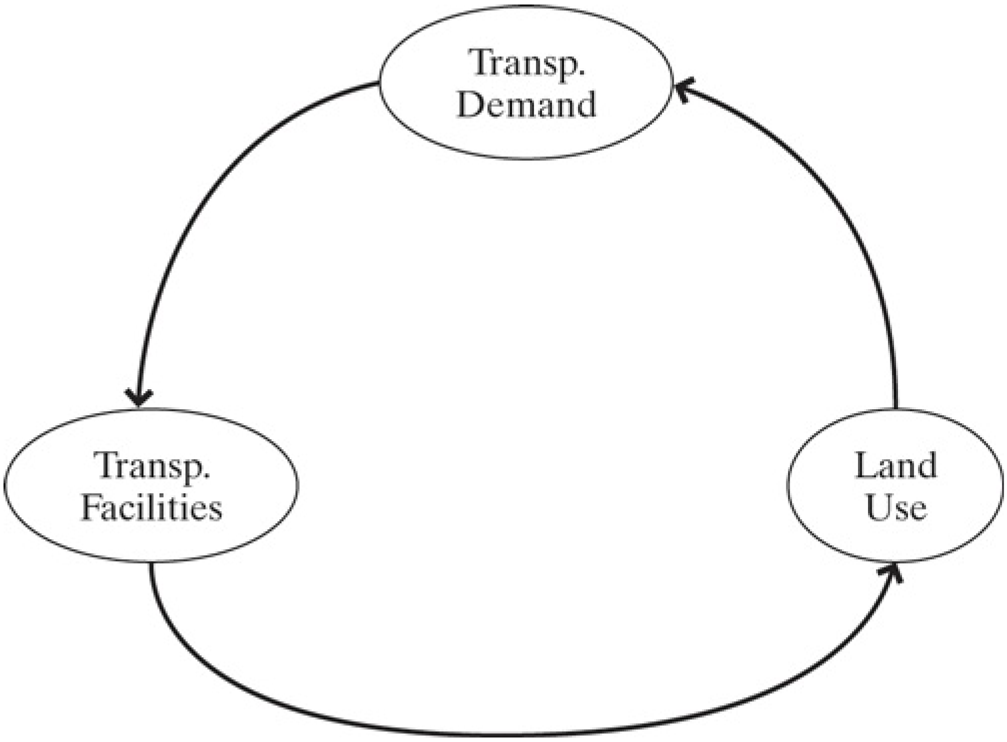


Figure 1.2: Full Alternative Text

Transportation planners and traffic engineers attempt to provide capacity for observed or predicted travel demand by building transportation systems. The improvement of transportation systems, however, makes the adjacent and nearby lands more accessible and, therefore, more attractive for development. Thus, building new transportation facilities leads to further increases in land-use development, which (in turn) results in even higher transportation demands. This circular, self-reinforcing characteristic of traffic demand creates a central dilemma: Building additional transportation capacity invariably leads to incrementally increased travel demands.

In many major cities, this has led to the search for more efficient transportation systems, such as public transit and car-pooling programs. In some of the largest cities, providing additional system capacity on highways is no longer an objective, as such systems are already substantially choking in congestion. In these places, the emphasis shifts to improvements within existing highway rights-of-way and to the elimination of bottleneck locations (without adding to overall capacity). Other approaches include staggered work hours and work days to reduce peak-hour demands, and even more radical approaches involve development of satellite centers outside of the central business district (CBD) to spatially disperse highly directional demands into and out of city centers.

Demand, however, is not constrained by capacity in all cities, and the normal process of attempting to accommodate demand as it increases is feasible in these areas. At the same time, the circular nature of the travel/demand relationship will lead to congestion if care is not taken to manage both capacity and demand to keep them within tolerable limits.

It is important that the traffic engineer understands this process. It is complex and cannot be stopped at any moment in time. Demand-prediction techniques (not covered in this text) must start and stop at arbitrary points in time. The real process is ongoing, and as new or improved facilities are provided, travel demand is constantly changing. Plans and proposals must recognize both this reality and the professional’s inability to precisely predict its impacts. *A 10-year traffic demand forecast that comes within approximately* ±20% *of the actual value is considered a significant success*. The essential truth, however, is that traffic engineers cannot simply build their way out of congestion.

If anything, we still tend to underestimate the impact of transportation facilities on land-use development. Often, the increase in demand is hastened by development occurring simply as a result of the planning of a new facility.

One of the classic cases occurred on Long Island, in New York State. As the Long Island Expressway was built, the development of suburban residential communities lurched forward in anticipation. While the expressway's link to Exit 7 was being constructed, new homes were being built at the anticipated Exit 10, even though the facility would not be open to that point for several years. The result was that as the expressway was completed section by section, the 20-year anticipated demand was being achieved within a few years, or even months. This process has been repeated in many cases throughout the nation.

1.2.2 Concepts of Mobility and Accessibility

Transportation systems provide the nation's population with both mobility and accessibility. The two concepts are strongly interrelated but have distinctly different elements. *Mobility* refers to the ability to travel to many different destinations with relative ease, while *accessibility* refers to the ability to gain entry to a particular site or area.

Mobility gives travelers a wide range of choices as to where to go to satisfy particular needs, and provides for efficient trips to get to them. Mobility allows shoppers to choose from among many competing shopping centers and stores. Similarly, mobility provides the traveler with many choices for all kinds of trip purposes, including recreational trips, medical trips, educational trips, and even the commute to work. The range of available choices is enabled by having an effective transportation network that connects to many alternative trip destinations within a reasonable time, with relative ease, and at reasonable cost. Thus, mobility provides not only access to many travel opportunities but also relative speed and convenience for the required trips.

Accessibility is a major factor in the value of land. When land can be accessed by many travelers from many potential origins, it is more desirable for development and, therefore, more valuable. Thus, proximity of land to major highways and public transportation facilities is a major factor determining its value.

Mobility and accessibility may also refer to different portions of a typical trip. Mobility focuses on the through portion of trips and is most affected by the effectiveness of through facilities that take a traveler from one general area to another. Accessibility requires the ability to make a transfer from the transportation system to the particular land parcel on which the desired activity is taking place. Accessibility, therefore, relies heavily on transfer facilities, which include parking for vehicles, public transit stops, and loading zones.

Most transportation systems are structured to separate mobility and access functions, as the two functions often compete and are not necessarily compatible. In highway systems, mobility is provided by high-type facilities, such as freeways, expressways, and primary and secondary arterials. Accessibility is generally provided by local street networks. Except for limited-access facilities, which serve only through vehicles (mobility), most other classes of highway serve both functions to some degree. Access maneuvers, however (e.g., parking and unparking a vehicle, vehicles entering and leaving off-street parking via driveways, buses stopping to pick up or discharge passengers, trucks stopped to load and/or unload goods), retard the progress of through traffic. High-speed through traffic, on the other hand, tends to make such access functions more dangerous.

A good transportation system must provide for both mobility and accessibility, and should be designed to separate the functions to the extent possible to ensure both safety and efficiency.

1.2.3 People, Goods, and Vehicles

The most common unit used by the traffic engineer is “vehicles.” Highway systems are planned, designed, and operated to move vehicles safely and efficiently from place to place. Yet the movement of vehicles is not the objective; the goal is the movement of the people and goods that occupy vehicles.

Modern traffic engineering now focuses more on people and goods. While lanes must be added to a freeway to increase its capacity to carry vehicles, its person-capacity can be increased by increasing the average vehicle occupancy. Consider a freeway lane with a capacity of 2,000 vehicles per hour (veh/h). If each vehicle carries one person, the lane has a capacity of 2,000 persons per hour as well. If the average car occupancy is increased to 2.0 persons per vehicle, the capacity in terms of people is doubled to 4,000 persons per hour. If the lane were established as an exclusive bus lane, the vehicle-capacity might be reduced to 1,000 veh/h due to the larger size and poorer operating characteristics of buses as compared with automobiles. However, if each bus carries 50 passengers, the people-capacity of the lane is increased to 50,000 persons per hour.

The efficient movement of goods is also vital to the general economy of the nation. The benefits of centralized and specialized production of various products are possible only if raw materials can be efficiently shipped to manufacturing sites and finished products can be efficiently distributed throughout the nation and the world for consumption. While long-distance shipment of goods and raw materials is often accomplished by water, rail, or air transportation, the final leg of the trip to deliver a good to the local store or the home of an individual consumer generally takes place on a truck using the highway system. Part of the accessibility function is the provision of facilities that allow trucks to be loaded and unloaded with minimal disruption to through traffic and the accessibility of people to a given site.

The medium of all highway transportation is the vehicle. The design, operation, and control of highway systems rely heavily on the characteristics of the vehicle and of the driver. In the final analysis, however, the objective is to move people and goods, not vehicles.

1.2.4 Transportation Modes

While traffic engineers focus their attention on the movement of people and goods in over-the-road vehicles, they must be keenly aware of the role of public transportation and other modes, particularly as they interface with the street and highway system. Chapter 2 presents an in-depth overview of the various transportation modes and their functions.

1.3 History of U.S. Highway Legislation

The development of highway systems in the United States is strongly tied to federal legislation that supports and regulates much of this activity. Key historical and legislative actions are discussed in the sections that follow.

1.3.1 The National Pike and the States' Rights Issue

Before the 1800s, roads were little more than trails cleared through the wilderness by adventurous travelers and explorers. Private roadways began to appear in the latter part of the 1700s. These roadways ranged in quality and length from cleared trails to plank roadways. They were built by private owners, and fees were charged for their use. At points where fees were to be collected, a barrier usually consisting of a single crossbar was mounted on a swiveling stake, referred to as a "pike." When the fee was collected, the pike would be swiveled or turned, allowing the traveler to proceed. This early process gave birth to the term "turnpike," often used to describe toll roadways in modern time.

In 1811, the construction of the first national roadway was begun under the direct supervision of the federal government. Known as the "national pike" or the "Cumberland Road," this facility stretched for 800 miles from Cumberland, MD, in the east, to Vandalia, IL, in the west. A combination of unpaved and plank sections, it was finally completed in 1852 at a total cost of $6.8 million. A good deal of the original route is now a portion of U.S. Route 40.

The course of highway development in the United States, however, was forever changed as a result of an 1832 Supreme Court case brought by the administration of President Andrew Jackson. A major proponent of states' rights, the Jackson Administration petitioned the court claiming that the U.S. constitution did not specifically define transportation and roadways as federal functions; they were, therefore, the responsibility of the individual states. The Supreme Court upheld this position, and the principal administrative responsibility for transportation and highways was forevermore assigned to state governments.

If the planning, design, construction, maintenance, and operation of highway systems is a state responsibility, what is the role of federal agencies—for example, the U.S. Department of Transportation and its components, such as the Federal Highway Administration, the National Highway Safety Administration, and others in these processes?

The federal government asserts its overall control of highway systems through the power of the purse string. The federal government provides massive funding for the construction, maintenance, and operation of highway and other transportation systems. States are not *required* to follow federal mandates and standards but must do so to qualify for federal funding of projects. Thus, the federal government does not force a state to participate in federal-aid transportation programs. If it chooses to participate, however, it must follow federal guidelines and standards. As no state can afford to give up this massive funding source, the federal government imposes strong control of policy issues and standards.

The federal role in highway systems has four major components:

1. Direct responsibility for highway systems on federally owned lands, such as national parks and Native American reservations.

2. Provision of funding assistance in accord with current federal-aid transportation legislation.

3. Development of planning, design, and other relevant standards and guidelines that must be followed to qualify for receipt of federal-aid transportation funds.

4. Monitoring and enforcing compliance with federal standards and criteria, and the use of federal-aid funds.

State governments have the primary responsibility for the planning, design, construction, maintenance, and operation of highway systems. These functions are generally carried out through a state department of transportation or similar agency. States are entrusted with:

1. Full responsibility for administration of highway systems.

2. Full responsibility for the planning, design, construction, maintenance, and operation of highway systems in conformance with applicable federal standards and guidelines.

3. The right to delegate responsibilities for local roadway systems to local jurisdictions or governmental agencies.

Local governments have general responsibility for local roadway systems as delegated in state law. In general, local governments are responsible for the planning, design, construction, maintenance, and control of local roadway systems. Often, assistance from state programs and agencies is available to local governments in fulfilling these functions. At intersections of state highways with local roadways, it is generally the state that has the responsibility to control the intersection.

Local organizations for highway functions range from a full highway or transportation department to local police to a single professional traffic or city engineer.

There are also a number of special situations across the United States. In New York State, for example, the state constitution grants “home rule” powers to any municipality with a population in excess of 1,000,000 people. Under this provision, New York City has full jurisdiction over all highways within its borders, including those on the state highway system.

1.3.2 Key Legislative Milestones

Federal-Aid Highway Act of 1916

The Federal-Aid Highway Act of 1916 was the first allocation of federal-aid highway funds for highway construction by the states. It established the “A-B-C System” of primary, secondary, and tertiary federal-aid highways, and provided 50% of the funding for construction of highways in this system. Revenues for federal aid were taken from the federal general fund, and the act was renewed every 2 to 5 years (with increasing amounts dedicated). No major changes in funding formulas were forthcoming for a period of 40 years.

Federal-Aid Highway Act of 1934

In addition to renewing funding for the A-B-C System, this act authorized states to use up to 1.5% of federal-aid funds for planning studies and other investigations. It represented the entry of the federal government into highway planning.

Federal-Aid Highway Act of 1944

This act contained the initial authorization of what became the National System of Interstate and Defense Highways. No appropriation of funds occurred, however, and the system was not initiated for another 12 years.

Federal-Aid Highway Act of 1956

The authorization and appropriation of funds for the implementation of the National System of Interstate and Defense Highways occurred in 1956. The act also set the federal share of the cost of the Interstate System at 90%, the first major change in funding formulas since 1916. Because of the major impact on the amounts of federal funds to be spent, the act also created the *Highway Trust Fund* and enacted a series of road-user taxes to provide it with revenues. These taxes included excise taxes on motor fuels, vehicle purchases, motor oil, and replacement parts. Most of these taxes, except for the federal fuel tax, were dropped during the Nixon Administration. The monies housed in the Highway Trust Fund may be disbursed only for purposes authorized by the current federal-aid highway act.

Federal-Aid Highway Act of 1970

Also known as the Highway Safety Act of 1970, this legislation increased the federal subsidy of non-Interstate highway projects to 70% and required all states to implement highway safety agencies and programs.

Federal-Aid Highway Act of 1983

This act contained the “Interstate trade-in” provision that allows states to “trade in” federal-aid funds designated for urban Interstate projects for alternative transit systems. This historic provision was the first to allow road-user taxes to be used to pay for public transit improvements.

ISTEA and TEA-21

The single largest overhaul of federal-aid highway programs occurred with the passage of the Intermodal Surface Transportation Efficiency Act (ISTEA) in 1991 and its successor, the Transportation Equity Act for the 21st Century (TEA-21), in 1998.

Most importantly, these acts combined federal-aid programs for all modes of transportation and greatly liberalized the ability of state and local governments to make decisions on modal allocations. Key provisions of ISTEA included the following:

1. Greatly increased local options in the use of federal-aid transportation funds.

2. Increased the importance and funding to Metropolitan Planning Organizations (MPOs) and requiring that each state maintain a state transportation improvement plan (STIP).

3. Tied federal-aid transportation funding to compliance with the Clean Air Act and its amendments.

4. Authorized $38 billion for a 155,000-mile National Highway System.

5. Authorized an additional $7.2 million to complete the Interstate System and $17 billion to maintain it as part of the National Highway System.

6. Extended 90% federal funding of Interstate-eligible projects.

7. Combined all other federal-aid systems into a single surface transportation system with 80% federal funding.

8. Allowed (for the first time) the use of federal-aid funds in the construction of toll roads.

TEA-21 followed in kind, increasing funding levels, further liberalizing local options for allocation of funds, further encouraging intermodality and integration of transportation systems, and continuing the link between compliance with clean-air standards and federal transportation funding.

The creation of the National Highway System (NHS) answered a key question that had been debated for years: What comes after the Interstate System? The new, expanded NHS is not limited to freeway facilities and is over three times the size of the Interstate System, which becomes part of the NHS.

SAFETY-LU

President Bush signed the most expensive transportation funding act into law on August 10, 2005. The act was a mile wide, and more than four years late, with intervening highway funding being accomplished through annual continuation legislation that kept TEA-21 in effect.

The Safe, Accountable, Flexible and Efficient Transportation Equity Act—A Legacy for Users (SAFETY-LU) has been both praised and criticized. While it retains most of the programs of ISTEA and TEA-21, and expands the funding for most of them, the act also adds many new programs and provisions, leading some lawmakers and politicians to label it “the most pork-filled legislation in U.S. history.” Table 1.3 provides a simple listing of the programs covered under this legislation. The program, which authorizes over $248 billion in expenditures, includes many programs that represent items of special interest inserted by members of Congress.

Table 1.3: Programs Covered by SAFETY-LU*

Interstate Maintenance Program	$25.1
National Highway System	$30.5
Surface Transportation System	$32.4
Congestion Mitigation/Air Quality Improvement Program	$8.5
Highway Safety Improvement Program	$5.1
Appalachian Development/Highway System Program	$2.4
Recreational Trails Program	$0.4
Federal Lands Highway Program	$4.5
National Corridor Infrastructure Improvement Program	$1.9
Coordinated Border Infrastructure Program	$0.8
National Scenic Byways Program	$0.2
Construction of Ferry Boats/Terminals	$0.3
Puerto Rico Highway Program	$0.7
Project of National and Regional Significance Program	$1.8
High-Priority Projects Program	$14.8
Safe Routes to School Program	$ 0.61
Deployment of MagLev Trans Projects	$ 0.45
Nat'l Corridor Planning/Dev of Coordinated Infrastructure Programs	$ 0.14
Highways for Life Program	$ 0.45
Highway Use Tax Evasion Projects	$ 0.12

*All amounts are stated in billions of dollars.

Table 1.3: Full Alternative Text

The legislation does recognize the need for massive funding of Interstate highway maintenance, as the system continues to age, with many structural components well past their anticipated service life. It also provides massive funding for the new NHS, which is the successor to the Interstate System in terms of new highways. It also retains the flexibility for local governments to push more funding into public transportation modes.

MAP-21

The current (as of June 2017) transportation act is the "Moving Ahead for Progress in the 21st Century" (MAP) act, signed into law by President Obama on July 12, 2012. Unlike its immediate predecessors, MAP-21 was a limited 2-year stopgap that froze spending at the 2012 level for the 2-year period covered by the legislation. It consolidated 87 programs under SAFETY-LU into 30, and gave states greater flexibility in the allocation of funds. It authorized $105 billion for 27 months.

Like its immediate predecessors, MAP-21 has yet to be replaced. It has been extended on an annual basis by Congress to provide for ongoing federal transportation funding. A replacement piece of legislation has been under discussion for some time, and is now (June 2017) being considered as part of the Trump Administration's overall infrastructure plan.

1.3.3 The National System of Interstate and Defense Highways

The "Interstate System" has been described as the largest public works project in the history of mankind. In 1919, a young army officer, Dwight Eisenhower, took part in an effort to move a complete battalion of troops and military equipment from coast to coast on the nation's highways to determine their utility for such movements in a time of potential war. The trip took months and left the young officer with a keen appreciation for the need to develop a national roadway system. It was no accident that the Interstate System was implemented in the administration of President Dwight Eisenhower, nor that the system now bears his name.

After the end of World War II, the nation entered a period of sustained prosperity. One of the principal signs of that prosperity was the great increase in auto ownership along with the expanding desire of owners to use their cars for daily commuting and for recreational travel. Motorists groups, such as the American Automobile Association (AAA), were formed and began substantial lobbying efforts to expand the nation's highway systems. At the same time, the over-the-road trucking industry was making major inroads against the previous rail monopoly on intercity freight haulage. Truckers also lobbied strongly for improved highway systems. These substantial pressures led to the inauguration of the Interstate System in 1956.

The System Concept

Authorized in 1944 and implemented in 1956, the National System of Interstate and Defense Highways is a 42,500-mile national system of multilane, limited-access facilities. The system was designed to connect all standard metropolitan statistical areas (SMSAs) with 50,000 or greater population (at the time) with a continuous system of limited-access facilities. The allocation of 90% of the cost of the system to the federal government was justified on the basis of the potential military use of the system in wartime.

System Characteristics

Key characteristics of the Interstate System include the following:

1. All highways have at least two lanes for the exclusive use of traffic in each direction.
2. All highways have full control of access.
3. The system must form a closed loop: All Interstate highways must begin and end at a junction with another Interstate highway.
4. North–South routes have odd one- or two-digit numbers (e.g., I-95).
5. East–West routes have even one- or two-digit numbers (e.g., I-80).
6. Interstate routes serving as bypass loops or acting as a connector to a primary Interstate facility have three-digit route numbers, with the last two digits indicating the primary route.

A map of the Interstate System is shown in Figure 1.3.

Figure 1.3: A Map of the Interstate System

Figure 1.3: Full Alternative Text

Status and Costs

By 1994, the system was 99.4% complete. Most of the unfinished sections were not expected to ever be completed for a variety of reasons. The total cost of the system was approximately $128.9 billion. This final estimate of cost was released in 1991. It is estimated that the cost would be over $500 billion in today’s dollars.

The impact of the Interstate System on the nation cannot be understated. The system facilitated and enabled the rapid suburbanization of the United States by providing a means for workers to commute from suburban homes to urban jobs. The economy of urban centers suffered as shoppers moved in droves from traditional CBDs to suburban malls.

The system also had serious negative impacts on some of the environs through which it was built. Following the traditional theory of benefit-cost, urban sections were often built through the low-income parts of communities where land was the cheapest. The massive Interstate highway facilities created physical barriers, partitioning many communities, displacing residents, and separating others from their schools, churches, and local shops. Social unrest resulted in several parts of the country, which eventually leading to in important modifications to the public hearing process and in the ability of local opponents to legally stop many urban highway projects.

Between 1944 and 1956, a national debate was waged over whether the Interstate System should be built into and out of urban areas, or whether all Interstate facilities should terminate in ring roads built around urban areas. Proponents of the ring-road option (including, ironically, Robert Moses, who built many highways into and out of urban cities) argued that building these roadways into and out of cities would lead to massive urban congestion. The other side of the argument was that most of the road users who were paying for the system through their road-user taxes lived in urban areas and should be served. The latter view prevailed, but the predicted rapid growth of urban congestion also became a reality.

1.4 Elements of Traffic Engineering

There are a number of key elements of traffic engineering:

1. Traffic studies and characteristics
2. Performance evaluation
3. Facility design
4. Traffic control
5. Traffic operations
6. Transportation systems management
7. Integration of intelligent transportation system technologies

Traffic studies and characteristics involve measuring and quantifying various aspect of highway traffic. Studies focus on data collection and analysis that is used to characterize traffic, including (but not limited to) traffic volumes and demands, speed and travel time, delay, accidents, origins and destinations, modal use, and other variables.

Performance evaluation is a means by which traffic engineers can rate the operating characteristics of individual sections of facilities and facilities as a whole in relative terms. Such evaluation relies on measures of performance quality and is often stated in terms of “levels of service.” Levels of service are letter grades, from A to F, describing how well a facility is operating using specified performance criteria. Like grades in a course, A is very good, while F connotes failure (on some level). As part of performance evaluation, the *capacity* of highway facilities must be determined.

Facility design involves traffic engineers in the functional and geometric design of highways and other traffic facilities. Traffic engineers, per se, are not involved in the structural design of highway facilities but should have some appreciation for structural characteristics of their facilities.

Traffic control is a central function of traffic engineers and involves the establishment of traffic regulations and their communication to the driver through the use of traffic control devices, such as signs, markings, and signals.

Traffic operations involves measures that influence overall operation of traffic facilities, such as one-way street systems, transit operations, curb management, and surveillance and network control systems.

Transportation systems management (TSM) involves virtually all aspects of traffic engineering in a focus on optimizing system capacity and operations. Specific aspects of TSM include high-occupancy vehicle priority systems, car-pooling programs, pricing strategies to manage demand, and similar functions.

Intelligent transportation systems (ITS) refers to the application of modern telecommunications technology to the operation and control of transportation systems. Such systems include automated highways, automated toll-collection systems, vehicle-tracking systems, in-vehicle GPS and mapping systems, automated enforcement of traffic lights and speed laws, smart control devices, and others. This is a rapidly emerging family of technologies with the potential to radically alter the way we travel as well as the way in which transportation professionals gather information and control facilities. While the technology continues to expand, society will grapple with the substantial “big brother” issues that such systems invariably create.

This text contains material related to all of these components of the broad and complex profession of traffic engineering.

1.5 Modern Problems for the Traffic Engineer

We live in a complex and rapidly developing world. Consequently, the problems that traffic engineers are involved in evolve rapidly.

Urban congestion has been a major issue for many years. Given the transportation demand cycle, it is not always possible to solve congestion problems through expansion of capacity. Traffic engineers therefore are involved in the development of programs and strategies to manage demand in both time and space and to discourage growth where necessary. A real question is not “how much capacity is needed to handle demand?” but rather “how many vehicles and/or people can be allowed to enter congested areas within designated time periods?”

Growth management is a major current issue. A number of states have legislation that ties development permits to level-of-service impacts on the highway and transportation system. Where development will cause substantial deterioration in the quality of traffic service, either such development will be disallowed or the developer will be responsible for general highway and traffic improvements that mitigate these negative impacts. Such policies are more easily dealt with in good economic times. When the economy is sluggish, the issue will often be a clash between the desire to reduce congestion and the desire to encourage development as a means of increasing the tax base.

Reconstruction of existing highway facilities also causes unique problems. The entire Interstate System has been aging, and many of its facilities have required major reconstruction efforts. Part of the problem is that reconstruction of Interstate facilities receives the 90% federal subsidy, while routine maintenance on the same facility is primarily the responsibility of state and local governments. Deferring routine maintenance on these facilities in favor of major reconstruction efforts has resulted from federal funding policies over the years. Major reconstruction efforts have a substantial major burden not involved in the initial construction of these facilities: maintaining traffic. It is easier to build a new facility in a dedicated undeveloped right-of-way than to rebuild it while continuing to serve 100,000 or more vehicles per day. Thus, issues of long-term and short-term construction detours as well as the diversion of traffic to alternate routes require major planning by traffic engineers.

Since 2001, the issue of security of transportation facilities has come to the fore. The creation of facilities and processes for random and systematic inspection of trucks and other vehicles at critical locations is a major challenge, as is securing major public transportation systems such as railroads, airports, and rapid transit systems.

As the fifth edition of this text is written, we are now in a new era with many unknowns. With the sharp rise in fuel prices through 2008, vehicle usage actually began to decline for the first time in decades. The upward trend, however, returned as economic conditions improved.

The economic crisis of 2008 and 2009 caused many shifts in the economy, even as the price of fuel came back to more normal levels. Major carmakers in the United States (Chrysler, GM) headed into bankruptcy, with major industry reductions and changes. Government loans to both banks and industries brought with it more governmental control of private industries. A shift of U.S. automakers to smaller, more fuel-efficient and “green” vehicles has begun, with no clear appreciation of whether the buying public will sustain the shift.

As the economy rebounded, however, some of these shifts were modified. While the emphasis on “green” vehicles continues, renewed interest and sales of sport-utility vehicles (SUVs), pickup trucks, and “muscle cars” occurred. While they still have some problems, the major U.S. automakers are more stable. Banks and other industries began to pay off their debt to the government, returning to more normal private control and management, albeit in a more stringent regulatory environment.

For perhaps the first time in many decades, transportation and traffic demand may be very much dependent upon the state of the general economy, not the usual motivators of improved mobility and accessibility. Will people learn new behaviors resulting in fewer and more efficient trips? Will people flock to hybrid or fully electric vehicles to reduce fuel costs? Will public transportation pick up substantial new customers as big-city drivers abandon their cars for the daily commute? It is an unsettling time that will continue to evolve into new challenges for traffic and transportation engineers. With new challenges, however, comes the ability for new and innovative approaches that might not have been feasible only a few years ago.

The point is that traffic engineers cannot expect to practice their profession only in traditional ways on traditional projects. Like any professional, the traffic engineer must be ready to face current problems and to play an important role in any situation that involves transportation and/or traffic systems.

1.6 Standard References for the Traffic Engineer

In order to remain up to date and aware, the traffic engineer must keep up with modern developments through membership and participation in professional organizations, regular review of key periodicals, and an awareness of the latest standards and criteria for professional practice.

Key professional organizations for the traffic engineer include the ITE, the Transportation Research Board (TRB), the Transportation Group of the American Society of Civil Engineers (ASCE), ITS America, and others. All of these provide literature and maintain journals, and have local, regional, and national meetings. TRB is a branch of the National Academy of Engineering and is a major source of research papers and reports.

Like many engineering fields, the traffic engineering profession has many manuals and standard references, most of which will be referred to in the chapters of this text. Major references include

- *Traffic Engineering Handbook, 7th Edition* [1]
- *Uniform Vehicle Code and Model Traffic Ordinance* [6]
- *Manual on Uniform Traffic Control Devices, 2009* (as updated through May 2012) [7]
- *Highway Capacity Manual, 6th Edition: A Guide for Multimodal Mobility Analysis* [8]
- *A Policy on Geometric Design of Highways and Streets* (The AASHTO Green Book), 6th *Edition* [9]
- *Traffic Signal Timing Manual, 2nd Edition* [10]
- *Transportation Planning Handbook, 4th Edition* [11]
- *Trip Generation, 8th Edition* [12]

- *Parking Generation, 4th Edition* [13]

All of these documents are updated periodically, and the traffic engineering professional should be aware of when updates are published and where they can be accessed.

Other manuals abound and often relate to specific aspects of traffic engineering. These references document the current state of the art in traffic engineering, and those most frequently used should be part of the professional's personal library.

There are also a wide variety of internet sites that are of great value to the traffic engineer. Specific sites are not listed here, as they change rapidly. All of the professional organizations, as well as equipment manufacturers, maintain web sites. The federal Department of Transportation (DOT), Federal Highway Administration (FHWA), National Highway Traffic Safety Administration (NHTSA), and private highway-related organizations maintain web sites. The entire *Manual on Uniform Traffic Control Devices* is available online through the FHWA web site, as is the *Manual of Traffic Signal Timing*.

Because traffic engineering is a rapidly changing field, the reader cannot assume that every standard and analysis process included in this text is current, particularly as the time since publication increases. While the authors will continue to produce periodic updates, the traffic engineer must keep abreast of latest developments as a professional responsibility.

1.7 Metric versus U.S. Units

This text is published in English (or Standard U.S.) units. Despite several attempts to switch to metric units in the United States, most states now use English units in design and control.

Metric and U.S. standards are not the same. A standard 12-ft lane converts to a standard 3.6-m lane, which is narrower than 12 ft. Standards for a 70-mi/h design speed convert to standards for a 120-km/h design speed, which are not numerically equivalent. This is because even units are used in both systems rather than the awkward fractional values that result from numerically equivalent conversions. That is why a metric set of wrenches for use on a foreign car is different from a standard U.S. wrench set.

Because more states are on the U.S. system than on the metric system (with more moving back to U.S. units) and because the size of the text would be unwieldy if dual units were included, this text continues to be written using standard U.S. units.

1.8 Closing Comments

The profession of traffic engineering is a broad and complex one. Nevertheless, it relies on key concepts and analyses and basic principles that do not change greatly over time. This text emphasizes both the basic principles and current (in 2017) standards and practices. The reader must keep abreast of changes that influence the latter.

References

- 1. Wolshen, B. and Pande, A. (Ed), *Traffic Engineering Handbook*, 7th Edition, John Wiley & Sons, Institute of Transportation Engineers, Washington, D.C., January 2016.

- 2. *Quick Facts 2016,* Federal Highway Administration, Washington, D.C., 2016.

- 3. "2016 Fatal Motor Crashes: Overview," *Traffic Safety Facts*, National Highway Traffic Safety Administration, Department of Transportation, Washington, D.C., October 2017.

- 4. *Motor Vehicle Deaths in 2016 Estimated to Be Highest in nine Years,* National Safety Council, Atasca, IL, 2017.

- 5. Henchman, J., "Gasoline Taxes and User Fees Pay for Only Half of State and Local Road Spending," *Tax Foundation Fiscal Facts*, Tax Foundation, Washington, D.C., January 6, 2016.

- 6. *Uniform Vehicle Code and Model Traffic Ordinance*, National Committee on Uniform Traffic Laws and Ordinances, Washington, D.C., 2002.

- 7. *Manual on Uniform Traffic Control Devices*, Federal Highway Administration, Washington, D.C., 2009, as updated through May 2012.

- 8. *Highway Capacity Manual, 6th Edition: A Guide for Multimodal Mobility Analysis*, Transportation Research Board, Washington, D.C., 2016.

- 9. *A Policy on Geometric Design of Highways and Streets*, 6th Edition, American Association of State Highway and Transportation Officials, Washington, D.C., 2011.

- 10. *Traffic Signal Timing Manual*, 2nd Edition, National Cooperative Highway Research Program Report 812, Transportation Research Board, Washington, D.C., 2016.

- 11. *Transportation Planning Handbook*, 4th Edition, John Wiley & Sons, Institute of Transportation Engineers, Washington, D.C., 2016.

- 12. *Trip Generation*, 8th Edition, ITE Informational Report, Institute of Transportation Engineers, Washington, D.C., 2008.

- 13. *Parking Generation*, 4th Edition, Institute of Transportation Engineers, Washington, D.C., 2010.

Chapter 2 Transportation Modes and Characteristics

The traffic engineer is involved in the planning, design, operation, and management of the street and highway system. While the street and highway system primarily serves vehicular traffic, it is actually multimodal in many ways.

Consider, for example, a typical major urban arterial. Within the right-of-way of the arterial and its intersections, service is provided to drivers and passengers in privately owned vehicles, passengers in bus transit operating on the arterial, goods moved in trucks along the arterial, pedestrians using the sidewalks and crosswalks, and bicyclists riding in vehicular lanes or in designated bike lanes. In some places, light rail transit may be sharing vehicular lanes. Even rapid transit lines, always on segregated rights-of-way (tunnels, elevated structures, separated facilities), interact by depositing large numbers of pedestrians onto the arterial at station locations.

Curb space along the arterial is shared by moving vehicles, bus-stops, truck loading zones, parking, and perhaps bicyclists. One of the principal functions of urban traffic engineers is the management of curb space, and its allocation to competing user groups.

On a more regional level, highway systems provide access to airports, railroad stations, ports, and other transportation facilities.

It is imperative, therefore, that traffic engineers clearly understand the many modes of transportation that impact their profession, and how these modes fit into the national and regional infrastructure that serves our total transportation needs.

2.1 Classifying Transportation Modes

There are many ways of classifying transportation modes. One significant factor is whether the transportation demand being serviced is intercity (between centralized areas) or intra-city (within a centralized area). Intercity trips typically involve longer travel distances and travel times, and occur less frequently than trips entirely within an area. Some modes of transportation serve one type of trip almost exclusively: Virtually all trips by air are intercity; virtually all pedestrian trips are local, or intra-city.

A second categorization involves whether the primary function is the movement of goods or people. While most transportation is dominated by the movement of people, goods movement is a vital function in the economy.

Both people and goods may travel intercity or intra-city. People use a wide variety of modes, but most person-travel is by private automobile. In major cities, public transportation (rapid transit, light rail, bus) can serve large components of the person-travel demand. Intercity person-trips occur in private automobiles, airplanes, on passenger railroads, and on intercity buses. Within urban centers, person-trips are accommodated by cars, bus or rail transit, and walking.

Goods move between cities in airplanes, over-the-road trucks, railroads, and ships. Where liquids are concerned, pipelines also play a major role. Within cities, most goods move by truck, but some may use a variety of rail services. One normally would not think of pipelines in an urban setting, yet they form a vital part of the urban transportation infrastructure in the delivery of natural gas and water to individual consumers, and in the removal of liquid waste.

A final way to categorize transportation modes is by whether the mode is privately or publicly operated. In intercity transportation, passenger cars are virtually always privately owned and operated. Airlines, railroads, ships, and pipelines are owned by mostly private or public operators who provide, maintain, manage, and operate the physical infrastructure. All are subject to government regulation. One might, however, consider such modes to be “public” in nature, as the individual traveler (or good) has no direct role in the operation of the service.

In urban areas, pedestrians, bicyclists, and drivers/passengers of privately owned vehicles form the core of “private” transportation, while “public” transportation includes transit and for-hire vehicles (taxis).

Table 2.1 summarizes the various transportation modes in terms of the categories discussed.

Table 2.1: Transportation Modes by Category

Function	Person-Transport	Goods Movement
Intercity	Private auto (private) Passenger railroad (public) Air (public) Intercity bus (public) Passenger boat—ferries (public)	Over-the-road trucks (public) Air freight (public) Railroad (public) Ships (public) Pipelines (public)
Intra-City	Pedestrian (private) Bicycle (private) Private auto (private) For-hire vehicle (public) Transit (public) Passenger boat—ferries (public)	Trucks (public) Railroad (public) Pipelines (public)

Note that “public” modes are so categorized because they are publicly accessible to users, whether or not the operator of the service is a private entity or a public one.

Table 2.1: Full Alternative Text

2.2 The Transportation Infrastructure and Its Use

To provide for the diverse transportation needs of the nation, a vast infrastructure must be in place. Much (but not all) of the basic infrastructure is publicly provided. Table 2.2 shows the miles of transportation infrastructure in place within the United States in 2014 [1,2].

Table 2.2: Transportation Infrastructure in the United States—2014

Category	System Mileage or Number
Public roads and highways (route mileage)	4,177,074
Interstate highway system	*47,622*
Other national highway system	*178,643*
Local	*3,950,809*
Railroad (route mileage)	127,012
Class I railroad	*94,362**
AMTRAK	*21,356*
Commuter rail	*7,795*
Heavy rail	*1,622*
Light rail	*1,877*
Navigable channels (route mileage)	25,000
Pipelines (route mileage)	2,368,436
Oil	*199,653*
Natural gas	*2,168,783*
Airports (number of airports)	19,294
Public use	*5,145*
Private use	*13,863*
Military	*286*

*Much of AMTRAK's mileage is shared with Class I railroads.

Table 2.2: Full Alternative Text

The U.S. highway system is massive and serves intercity and intra-city transport of people and goods. The Interstate System (formally the Eisenhower National System of Interstate and Defense Highways) is of particular importance. Though it makes up a bit over 1.1% of the paved route-miles in the United States, it serves approximately 20% of all vehicle-miles traveled in the United States. The planning, design, and importance of this system are discussed in later chapters.

"Navigable channels" include commercially navigable rivers and inland passages and the Great Lakes–St. Lawrence Seaway. They do *not* include ocean-going routes, which are virtually limitless.

While we may tend to think of pipelines as conveying mostly oil, the vast majority of pipelines are devoted to the delivery of natural gas. Most of these are located within urbanized areas, and they deliver natural gas right to the individual homes of users.

Table 2.3 shows the annual tonnage of goods moved by the various modes in 2015 [2]. Note that Table 2.3 shows only domestic goods movement, that is, goods moved entirely within the United States, and does *not* include goods imported to the United States from abroad or goods exported from the United States to abroad.

Table 2.3: U.S. Domestic Goods Movement by Tonnage —2015

Mode	Tonnage
Truck	10,568,000,000
Rail	1,602,000,000
Water	884,000,000
Air (includes air + truck)	10,000,000
Multiple modes and mail	1,346,000,000
Pipelines	3,326,000,000
Other or unknown	33,000,000
TOTAL	**17,997,000,000**

Table 2.3: Full Alternative Text

Table 2.4 shows similar data for passenger transportation [3], which is quantified in terms of passenger-miles of travel. Some of the modes are exclusively intercity or intra-city, but a number span both categories.

Table 2.4: Passenger-Miles of Travel in the United States—2014

Mode	Passenger-Miles Traveled
Air (intercity domestic)	**607,772,000,000**
Highway	**4,092,575**
Passenger car (intercity & intra-city)	3,731,888,000,000
Motorcycle (intercity & intra-city)	21,510,000,000
Intercity bus (intercity)	339,177,000,000
Public Transit (intra-city)	**55,321,000,000**
Motor bus*	21,429,000,000*
Light rail	2,675,000,000
Heavy rail (rail rapid transit)	18,339,000,000
Commuter rail	11,600,000,000
Demand-responsive*	864,000,000*
Ferry boat	414,000,000
Other	1,692,000,000
Railroad—AMTRAK (intercity)	**6,675,000,000**
TOTAL	**4,762,343,000,000**

*Most motor bus and demand-responsive transit occurs on streets and highways, and therefore could also be included in the "Highway" category. It was not to avoid double-counting these passenger miles.

NOTE: Air figures do not include 244,373,000,000 passenger miles flown on international flights.

Table 2.4: Full Alternative Text

From Table 2.4, it can be observed that the U.S. population accounted for almost 5 *trillion* passenger-miles of travel in 2014. The vast majority of these (approximately 86%) occur on the nation's street and highway system. This emphasizes the significant dependence on the automobile as the principal means of mobility and access for the nation's population.

Much of the service and infrastructure for heavy rail is centered in a few major cities, like New York, Chicago, and Washington, D.C. Ferry service is not widespread, and again is largely focused on a few areas such as New York (the Staten Island Ferry) and the Puget Sound region around Seattle. For many, urban travel options are limited to the automobile, bus transit, taxi, and other on-call car services. Intercity travelers have a broader range of choices available, including air, rail, intercity bus, or highways.

Note that there are no statistics shown in Table 2.4 for pedestrians, as it is almost impossible to collect meaningful data on how many pedestrian trips are made, and how far people walk for various purposes.

2.3 Modal Attributes

Travel modes for people can be divided into two general categories:

- Personal modes of transportation
- Public modes of transportation

The main characteristic of personal modes of transportation is that the traveler most often owns the “vehicle” in or on which the travel takes place. In some cases, the vehicle may be leased on a long-term basis or rented for a shorter period of time. In public modes, vehicles are generally owned and operated by an external agency that may be either publicly or privately owned. Personal modes of transportation include walking (no vehicle required), bicycling, and driving or riding in a privately owned and operated automobile. Public modes of transportation include taxi or other for-hire small vehicles, buses, light rail systems, and rail rapid transit (or “heavy rail” systems).

The primary features of personal modes of transportation are that they provide direct origin-to-destination service and are available at any time as needed. In public modes, taxis, for-hire vehicles, and other types of demand-responsive services closely mimic the characteristics of private modes. They, in general, do provide direct origin-to-destination service. They are available on call, but there may be waiting times and/or other time restrictions imposed, and come with a visible out-of-pocket cost.

Public modes of transportation, other than taxis and similar forms, primarily run on fixed routes according to a fixed schedule. The traveler must adjust his/her travel needs to accommodate these. Pickup and drop-off points may or may not be near the desired origin and destination, with the traveler responsible for making the connections between the origin and pickup location and the drop-off and destination location. Depending upon the specific circumstances, either or both of these could include significant travel time and/or travel cost.

Public modes of transportation can have numerous subcategories of characteristics that alter the type of service provided. Buses, for example, can be operated on local bus routes along local streets and arterials, or can make part or all of their trips on exclusive bus lanes or busways. Express bus services make pickups and drop-offs in defined areas, but travel nonstop between these areas to increase speed (and decrease travel time). Light rail services can operate on streets, mixed with other traffic, or in segregated lanes. They can also operate on separated rights-of-way with or without at-grade crossings.

Table 2.5 summarizes some of the fundamental service characteristics of personal and public transportation modes.

Table 2.5: Fundamental Service Characteristics of Personal and Public Transportation Modes

Type of Mode	Mode	Type of Service Provided	Typical Trip Lengths Served	Typical Average Speed of Service	Special Problems or Restrictions
Personal	Walking	– Door-to-door service provided. – Available on demand. – Limited control of personal environment.	0–1 miles	1–3 mi/h	– Depends on health/age. – Limited distance range. – Security. – Weather.
	Bicycling	– Door-to-door service provided. – Available on demand (assuming bike ownership). – Limited control of personal environment.	0–5 miles	5–15 mi/h	– Depends on health/age. – Secure storage for bike needed at both trip ends. – Limited dedicated bicycle facilities available in most areas. – Weather. – Safety of operation in mixed traffic.
	Automobile	– Door-to-door service provided. – Available on demand (assuming car ownership). – Full control of personal environment.	Unlimited	Highly variable 5–70 mi/h	– Parking needed at both ends of trip. – Safety (accident risk). – Costs of ownership, fuel, maintenance, insurance, etc.

Public	Local bus	– Bus stop must be accessed; often within ½ mile of origin/destination. – Available on schedule (usually headway-based). – No control of personal environment.	0.5–10.0 mi	7–12 mi/h	– Affected by local traffic conditions. – Waiting a problem in bad weather. - Late-night security.
	Express bus	– Bus stop must be accessed; may be a considerable distance away. – Available on schedule (usually headway-based). – No control of personal environment.	3–20 mi	10–30 mi/h	– Affected by traffic while in mixed lanes. – Access to service may be a problem. – Weather (general). – Schedules may be limited to peak periods.
	Light rail	– Stop/station must be accessed; may be a considerable distance away. – Available on schedule (may be headway-based). – No control of personal environment.	3–20 mi	10–35 mi/h	– Power outages disrupt service. – May have to cross traffic lanes to get to stations. – Weather can cause disruptions. – There may not be night service.
	Heavy rail	– Station must be accessed; often a considerable distance from origin/destination. – Available on schedule (may be headway-based). – No control of personal environment.	5–30 mi	15–45 mi/h	– Power outages disrupt service. – Station security may be a problem. – Weather can cause disruptions if above surface. – There may not be night service.
	Commuter rail	– Station must be accessed, may be far away from origin. – Available on published schedule (may be limited). – No control of personal environment.	5–50 mi	30–80 mi/h	– If electrically powered, outages may disrupt service. – Station security and comfort may be a problem. – Outdoor lines subject to weather disruptions. – Schedules may be very limited. – Fares generally quite high.

Table 2.5: Full Alternative Text

2.4 The Capacity of Transportation Modes

How big is the bucket? This is a pretty important characteristic if you are carrying water. It is no less critical for transportation systems. The bucket has a capacity of some number of gallons of fluid. Transportation systems carry people and goods, so their "capacity" involves how many people or how many tons of freight they can accommodate.

While capacity is a generically understood phrase, it was formally defined for highways in the first edition of the *Highway Capacity Manual* (*HCM*) [4]. For a highway, *capacity* is currently defined as the maximum rate of flow at which vehicles or persons can be reasonably expected to pass a point or uniform segment of a highway or lane under prevailing conditions [5]. One can easily extend the concept to other modes of transportation as well—at least in terms of the ability to carry people.

There are four key concepts embedded in this definition:

1. Rate of flow. Capacity is defined not in terms of a full-hourly volume but as a maximum rate of flow. The standard unit of time used in most cases is 15 minutes. Fifteen minutes is believed to be the minimum unit of time in which statistically stable (or predictable) traffic flow exists, although some researchers have used time periods of 5 minutes or even 1 minute in their studies.

2. Reasonable expectancy. Capacity is not a static measure. A 5-gallon bucket always has a capacity of 5 gallons. Capacity of a transportation system element is, however, a random variable depending upon traveler behavior, which is not static over time or space. Capacity is defined in terms of values that can be "reasonably expected" to be replicated at different times and at different places with similar characteristics. Thus, it is quite possible to observe actual flow rates in excess of stated capacity values on some transportation facilities.

3. Point or uniform segment. Capacity depends upon the physical

characteristics of the specific segment of the facility for which it is defined, as well as some characteristics of the travelers (or their vehicles) and control systems in place. Thus, along any given facility, capacity can only be stated for a point or a segment of limited length over which these characteristics are the same.

4. Prevailing conditions. Capacity is stated for whatever conditions prevail at the location. Prevailing conditions *for highways* fall into three broad categories:

    - – Physical conditions. This includes the geometric characteristics of the horizontal and vertical alignments, and cross-sectional elements such as lane widths and lateral clearances at the roadsides.

    - – Traffic conditions. This means the mix of vehicle types (cars, trucks, buses, etc.) making up the traffic stream.

    - – Control conditions. This means all traffic controls and operational regulations, including signalization, speed limits, lane-use controls, and other control measures.

The key idea here is that when any one of the underlying prevailing conditions is changed, so is the capacity.

While the concept of capacity transfers relatively easily to other passenger transportation modes, the issue of "prevailing conditions" is more difficult. Capacity of a highway segment refers to the maximum flow rate that the highway can accommodate. This is also true for other modes, but the list of "prevailing conditions" becomes much longer.

For example, consider the capacity of a single track of rapid transit line. Its capacity (in persons/h) depends upon several categories of issues:

- Design of the rail car. How many people can fit into a single rail car? This depends primarily on the size of the car (floor dimensions) and the number and arrangement of seats. Rapid transit lines typically service more standees than seated passengers, so the interior layout becomes critical.

- How many rail cars are in a train? The number of cars that make up a train is limited primarily by the length of station platforms. Obviously, more cars per train = more people per train.

- How many trains per hour can use a single track? There are two limits on this: the control system and station dwell times. Control systems, whether old (using fixed block signaling) or new (using moving block technology), essentially limit how close trains can get to each other during operation. If a control system allows trains to operate 2 minutes apart, then a track can handle 60/2 = 30 trains/h.

  The control system, however, is sometimes not the limiting factor. If it takes a train 4 minutes to decelerate to a station stop, let passengers on and off, and accelerate back to normal speed, then a second train cannot enter the station for a minimum of 4 minutes—regardless of the control system.

- Schedule. Unlike highways, where users essentially bring their own vehicles, public transportation systems provide vehicles on a schedule. Thus, though the track and dwell time might accommodate 30 trains/h, if the schedule only provides 20 trains/h, then the capacity is limited to the number of passengers that can be transported by 20 trains/h.

These issues together control the capacity of a segment of rail line. The issues become far more complicated when a rapid transit system involves several branch lines merging to form a trunk line. The capacity of the trunk line limits the capacity of all of the branch lines, as the total number of trains scheduled must be less than the capacity of the trunk. There may be “excess” capacity available on the branch lines, but it cannot be used. The single trunk line is, essentially, the bottleneck of the entire system.

Transit buses are similarly limited by the size and interior design of the bus, the length and number of bus stops, dwell times, and schedules. Further, bus operations are limited by the general traffic conditions on the streets they use.

Capacity values are established based upon observed vehicle and passenger volumes, and on analytic models that describe key limiting values of various system elements. For highway facilities, the *Highway Capacity Manual, 6th Edition* [5], is the standard document specifying procedures to estimate capacities of various types of facilities and facility segments. For transit facilities, the third edition of the *Transit Capacity Manual* [6] defines current standards.

Table 2.6 shows the current criteria for capacities of various types of highway facilities as specified by the *HCM*. The values shown represent fundamentally “ideal” conditions, that is, the best possible values that apply when there are only passenger cars in the traffic stream, and where all geometric elements are the most desirable—that is, 12-ft lanes, adequate lateral clearances, and so on. Highway capacity values are stated in terms of maximum flow rates in passenger cars/hour/lane (pc/h/ln). Vehicle occupancies vary over both time and space, and the *HCM* does not specify a national standard. In most places, car occupancy is between 1.3 and 1.5 persons per vehicle.

Table 2.6: Ideal Capacities of Highway Facilities

Uninterrupted Flow Facilities				
Type of Facility	Free-Flow Speed (mi/h)	Vehicle Capacity (pc/h/ln)	Person Capacity for Auto Occupancy of: (pers/h/ln)	
			1.3 pers/car	1.5 pers/car
Freeways	≥ 70	2,400	3,120	3,600
	65	2,350	3,055	3,525
	60	2,300	2,990	3,450
	55	2,250	2,925	3,375
Multilane highways	≥ 60	2,200	2,860	3,300
	55	2,100	2,730	3,150
	50	2,000	2,600	3,000
	45	1,900	2,470	2,950
Two-lane highways	All, one lane	1,700	2,210	2,550
	All, total, both dir*	3,200	4,160*	4,800*

Interrupted Flow Facilities				
Type of Facility	Green-to-Cycle Length Ratio (g/C)	Vehicle Capacity (pc/h/ln) *Based on:* (1,900 pc/hg/ln)	Person Capacity for Auto Occupancy of: (pers/h/ln)	
			1.3 pers/car	1.5 pers/car
Arterials/streets	0.30	570	741	855
	0.40	760	988	1,140
	0.50	950	1,235	1,425
	0.60	1,140	1,482	1,710
	0.70	1,330	1,729	1,995

* Total for both lanes; on two-lane highways, the directional movements interact, restricting passing maneuvers and total capacity.

Table 2.6: Full Alternative Text

For uninterrupted segments of highway facilities (freeways, multilane highways, two-lane highways), capacities are defined based upon the *free-flow speed* of the facility. An “uninterrupted” segment is any segment on a limited-access facility (no signals or other points of fixed interruption to the traffic stream) or a segment on a surface facility that is two miles or more from the nearest traffic signal. The “free-flow speed” of such a facility is the average speed that can be achieved when traffic is very light, that is, when there are few vehicles on the road. Recent studies show that free-flow speeds can exist over a wide range of flow rates, and that speeds do not begin to decline until flow rates exceed 1,000 pc/h/ln or more.

On interrupted flow facilities (arterials and streets), ideal capacities are stated in terms of *passenger cars per hour of green time per lane* (pc/hg/ln), as flow is restricted not only by prevailing geometric and traffic characteristics but also by signal timing. Thus, the capacity of an arterial, for example, is controlled by the traffic signal in the subject segment that has the minimum amount (or portion) of green time assigned. In Table 2.6, rough estimates of arterial and street capacity for green-to-cycle length (g/C) ratios of 0.30, 0.40, 0.50, 0.60, and 0.70 are shown. Other values are, of course, possible for different signal timings.

Generalized capacities for public transportation modes are shown in Table 2.7. Public transit capacities are based upon observations of highest-volume operations across the United States, documented by the American Public Transportation Association [7].

Table 2.7: Highest Observed Transit Flows in North America

Type of Transit	Route/Service	No. of Tracks/ Lanes	Trains/Hour or Buses/Hour	Passengers per Hour
Rail rapid transit	Queens E, F Express (NYC)	1	29	51,084
	Lexington Ave 4,5 Express (NYC)	1	28	34,059
	Queens Express & Local (NYC)	2	47	67,234
	Lexington Ave Express & Local (NYC)	2	50	63,234
Commuter rail	Metro-North RR, New Haven Branch	1	20	15,282
	Long Island RR, Babylon Branch	1	14	12,980
Light rail	Green-Line Subway, Boston	1*	45	9,600
	South Line, Calgary, Alberta	1	11	4,950
Bus	Lincoln Tunnel (NYC, Excl Lane)	1**	735	32,600
	West Transitway (Ottowa, Busway)	1***	225	11,100
	Madison Avenue (NYC, Bus Lanes)	2	180	10,000
	Hillside Avenue (NYC, Mixed Traffic)	–	180	10,000

* Double-track stations.

** No stops.

*** Stops; passing of stopped buses by others is possible.

Table 2.7: Full Alternative Text

It is no accident that the highest transit flows are found, for most types of transit, in New York City (NYC) and its surrounding tri-state region (which includes parts of New Jersey and Connecticut). New York has one of the largest rail rapid transit systems in the world (by revenue track-miles), as well as *the* largest local bus system in the world.

The single highest rail transit passenger flows per track are found on the Queens Line in NYC. The express track of this subway carries two routes—the E and F trains—and regularly services a peak-hour passenger flow of 51,000 passengers on one track through the critical station at Queens Plaza. When the local track is added, this four-track (two in each direction) subway carries over 67,000 passengers per hour in one direction every weekday during peak hours.

The highest single-track passenger flow on a commuter railroad is found on the Metro-North Railroad on its New Haven ranch. During peak hours, 20 trains per hour carrying over 15,000 persons per hour run every weekday. Capacities on commuter rail lines are limited primarily by schedules, but are affected by longer station dwell times than rapid transit (due to station configurations) and by railroad signal systems, which are generally less efficient than on modern rail rapid transit lines.

The highest hourly passenger flow observed on a light rail system is 9,600 passengers per hour, on Boston's Green-Line Subway. The Green-Line Subway accommodates several traditional trolley routes in downtown Boston. It has one track in each direction, but has double-track stations, which limits the impact of station dwell times. For a light rail system with single-track stations, the highest observed flows are on the South Line in Calgary, Alberta, Canada, where 4,950 persons per hour are carried during a typical weekday peak hour.

Bus system capacities are highly variable. The exclusive bus lane in the Lincoln Tunnel (New York–New Jersey) carries 735 buses per hour and

32,600 passengers per hour, but has no stops within it. Numerous bus routes converge on the bus lane, which connects directly to the Port Authority Bus Terminal in Manhattan, New York. In Ottawa, the West Transitway carries 225 buses per hour, and 11,100 passengers per hour. It is an exclusive roadway for buses, with stops. Buses may pass others while they are stopped. The highest on-street bus volumes are observed on Madison Avenue in Manhattan and Hillside Avenue in Queens, both in NYC. Both carry 180 buses per hour and approximately 10,000 passengers per hour. Madison Avenue has two exclusive bus lanes adjacent to the curb. On Hillside Avenue, buses operate in mixed traffic. These passenger volumes are extremely high, and represent multiple bus routes converging onto a common route. Bus schedules are usually the limit on capacity. A single bus per hour can carry as little as 50–60 passengers per hour, and typical single-route service can carry anything from several hundred passengers per hour to several thousand passengers per hour.

2.5 Multimodal Focus

The modern traffic engineer must keep the full range of transportation modes in mind in addressing transportation issues. Not every mode is appropriate for every demand, but in many urban cases, there may be different approaches that are feasible.

In the final analysis, most of our facilities will serve several different modes. Streets will serve cars, trucks, transit buses, pedestrians, taxis, and bicycles. Further, the integration of modes is a critical issue for the traffic engineer. After parking their car, a motorist becomes a pedestrian. After leaving a rapid transit station, a user is pedestrian, but they may use a bus or a taxi to continue their journey. The interface between and among modes is as important as the modes themselves.

“Multimodal” is a critical concept in modern transportation planning and design. Users of all modes need to be provided with a safe and efficient set of facilities to handle their unique needs. Often, the optimal approach will involve several modes of transportation. The best plans and designs will be those that provide an appropriate mix of transportation modes in a means that efficiently links and integrates them into a seamless transportation system.

References

- 1. *National Transportation Statistics*, Bureau of Transportation Statistics, U.S. Department of Transportation, Washington, D.C., April 2016.

- 2. *Freight Facts and Figures*, Bureau of Transportation Statistics, U.S. Department of Transportation, Washington, D.C., 2016.

- 3. *Passenger Travel Facts and Figures*, Bureau of Transportation Statistics, U.S. Department of Transportation, Washington, D.C., 2016.

- 4. *Highway Capacity Manual*, U.S. Government Printing Office and Highway Research Board, Washington, D.C., 1950.

- 5. *Highway Capacity Manual, 6th Edition: A Guide for Multimodal Mobility Analysis*, Transportation Research Board, Washington, D.C., 2016.

- 6. *Transit Capacity and Quality of Service Manual*, 3rd Edition, TCRP Report 165, Transit Cooperative Research Program, Transportation Research Board, Washington, D.C., 2013.

- 7. *Transit Ridership Report* (Quarterly Publication), American Public Transportation Association, Washington, D.C., 2015.

Problems

1. 2-1. What characteristics affect the capacity of a street or highway?

2. 2-2. What characteristics affect the capacity of a rapid transit line?

3. 2-3. A rapid transit line with one track in each direction uses rail cars that can accommodate 50 seated and 80 standing passengers. Stations are long enough to accommodate 10 car trains. The control system allows trains to travel 1.5 minutes apart. The critical station has a dwell time of 1.8 minutes. Estimate the capacity of one track.

4. 2-4. A six-lane urban freeway (three lanes in each direction) has a free-flow speed of 55 mi/h. Traffic includes 10% trucks and 2% express buses. Each truck and express bus displaces 2.0 passenger cars from the traffic stream. If the occupancy of passenger cars is 1.5 people per vehicle, and buses carry an average of 50 people per bus, what is the person-capacity of the freeway (in one direction)? It may be assumed that trucks carry one person (the driver).

5. 2-5. A travel demand of 30,000 people/h has been identified for a growing commercial corridor. What modal options might be considered to handle this demand, and what would (in general terms) be the advantages and disadvantages of each?

Chapter 3 Road-User, Vehicle, and Roadway Characteristics

The behavior of traffic is very much affected by the characteristics of the elements that comprise the traffic system, which are as follows:

- Road users—drivers, pedestrians, bicyclists, and passengers
- Vehicles—private and commercial
- Streets and highways
- Traffic control devices
- General environment

This chapter provides an overview of critical road-user, vehicle, and roadway characteristics. Chapter 4 provides an overview of traffic control devices and their role in the traffic system. Chapter 27 provides a more detailed look at the specific geometric characteristics of roadways.

The general environment also has an impact on traffic operations, but this is difficult to assess in any given situation. Such things as weather, lighting, density of development, and local enforcement policies all play a role in affecting traffic operations. These factors are most often considered qualitatively, with occasional supplemental quantitative information available to assist in making judgments.

3.1 Dealing with Diversity

Traffic engineering would be a great deal simpler if the various components of the traffic system had uniform characteristics. Traffic controls could be easily designed if all drivers reacted to them in exactly the same way. Safety could be more easily achieved if all vehicles had uniform dimensions, weights, and operating characteristics.

Drivers and other road users, however, have widely varying characteristics. The traffic engineer must deal with elderly drivers as well as 18-year-olds, aggressive drivers and timid drivers, and drivers subject to myriad distractions both inside and outside their vehicles. Simple subjects like reaction time, vision characteristics, and walking speed become complex because no two road users are the same.

Most human characteristics follow the normal distribution, which is discussed in Chapter 11. The normal distribution is characterized by a strong central tendency (i.e., most people have characteristics falling into a definable range). For example, most pedestrians crossing a street walk at speeds between 3.0 and 5.0 ft/s. However, there are a few pedestrians that walk either much slower or much faster. A normal distribution defines the proportions of the population expected to fall into these ranges. Because of variation, it is not practical to design a system for “average” characteristics. If a signal is timed, for example, to accommodate the average speed of crossing pedestrians, about half of all pedestrians would walk at a slower rate and be exposed to unacceptable risks.

Thus, most standards are geared to the “85th percentile” (or “15th percentile”) characteristic. In general terms, a percentile is a value in a distribution for which the stated percentage of the population has a characteristic that is less than or equal to the specified value. In terms of walking speed, for example, safety demands that we accommodate slower walkers. The 15th percentile walking speed is used, as only 15% of the population walks slower than this. Where driver reaction time is concerned, the 85th percentile value is used, as 85% of the population has a reaction time that is numerically equal to or less than this value. This approach leads to design practices and procedures that safely accommodate 85% of the population. What about the remaining 15%? One of the characteristics of normal distributions is that the extreme ends of the distribution (the highest and lowest 15%) extend to plus or minus infinity. In practical terms, the highest and lowest 15% of the distribution represent very extreme values that could not be effectively accommodated into design practices. Qualitatively, the existence of road users who may possess characteristics not within the 85th (or 15th) percentile is considered, but most standard practices and criteria do not directly accommodate them. Where feasible, higher percentile characteristics can be employed.

Just as road-user characteristics vary, the characteristics of vehicles vary widely as well. Highways must be designed to accommodate motorcycles, the full range of automobiles, and a wide range of commercial vehicles, including double- and triple-back tractor-trailer combinations. Thus, lane widths, for example, must accommodate the largest vehicles expected to use the facility.

Over the past decade, much progress has been made in the design of vehicles to make them safer and more efficient. With this emphasis, cars are getting smaller and lighter. Their relative safety within a mixed traffic stream still containing large trucks and buses becomes an important issue requiring new planning and design approaches. The traffic professional must be prepared to deal with this and other emerging issues as they arise.

Some control over the range of road-user and vehicle characteristics is maintained through licensing criteria and federal and state standards on vehicle design and operating characteristics. While these are important measures, the traffic engineer must still deal with a wide range of road-user and vehicle characteristics.

While traffic engineers have little control over driver and vehicle characteristics, design of roadway systems and traffic controls is in the core of their professional practice. In both cases, a strong degree of uniformity of approach is desirable. Roadways of a similar type and function should have a familiar “look” to drivers; traffic control devices should be as uniform as possible. Traffic engineers strive to provide information to drivers in uniform ways. While this does not assure uniform reactions from drivers, it at least narrows the range of behavior, as drivers become accustomed to and familiar with the cues traffic engineers design into the system.

3.2 Road Users and Their Characteristics

Human beings are complex and have a wide range of characteristics that can and do influence the driving task. In a system where the driver is in complete control of vehicle operations, good traffic engineering requires a keen understanding of driver characteristics. Much of the task of traffic engineers is to find ways to provide drivers with information in a clear, effective manner that induces safe and proper responses.

The two driver characteristics of utmost importance are visual acuity factors and the perception–reaction process. The two overlap, in that reaction requires the use of vision for most driving cues. Understanding how information is received and processed is a key element in the design of roadways and controls.

There are other important characteristics as well. Hearing is an important element in the driving task (i.e., horns, emergency vehicle sirens, brakes squealing, etc.). While noting this is important, however, no traffic element can be designed around audio cues, as hearing-impaired and even deaf drivers are licensed. Physical strength may have been important in the past, but the evolution of power-steering and power-braking systems has eliminated this as a major issue, with the possible exception of professional drivers of trucks, buses, and other heavy vehicles.

Of course, one of the most important human factors that influences driving is the personality and psychology of the driver. This, however, is not easily quantified and is difficult to consider in design. It is dealt with primarily through enforcement and licensing procedures that attempt to remove or restrict drivers who periodically display inappropriate tendencies, as indicated by accident and violation experience.

3.2.1 Visual Characteristics of Drivers

When drivers initially apply for, or renew, their licenses, they are asked to take an eye test, administered either by the state motor vehicle agency or by an optometrist or ophthalmologist who fills out an appropriate form for the motor vehicle agency. The test administered is a standard chart-reading exercise that measures *static visual acuity*—that is, the ability to see small stationary details clearly.

While certainly an important characteristic, static visual acuity is hardly the only visual factor involved in the driving task. The *Traffic Engineering Handbook* [1] provides an excellent summary of visual factors involved in driving, as shown in Table 3.1.

Table 3.1: Visual Factors in the Driving Task

Visual Factor	Definition	Sample Related Driving Task(s)
Accommodation	Change in the shape of the lens to bring images into focus.	Changing focus from dashboard displays to roadway.
Static visual acuity	Ability to see small details clearly.	Reading distant traffic signs.
Adaptation	Change in sensitivity to different levels of light.	Adjusting to changes in light upon entering a tunnel.
Angular movement	Seeing objects moving across the field of view.	Judging the speed of cars crossing our paths.
Movement in depth	Detecting changes in visual image size.	Judging the speed of an approaching vehicle.
Color	Discrimination between different colors.	Identifying the color of signals.
Contrast sensitivity	Seeing objects that are similar in brightness to their background.	Detecting dark-clothed pedestrians at night.
Depth perception	Judgment of the distance of objects.	Passing on two-lane roads with oncoming traffic.
Dynamic visual acuity	Ability to see objects that are in motion relative to the eye.	Reading traffic signs while moving.
Eye movement	Changing the direction of gaze.	Scanning the road environment for hazards.
Glare sensitivity	Ability to resist and recover from the effects of glare.	Reduction in visual performance due to headlight glare.
Peripheral vision	Detection of objects at the side of the visual field.	Seeing a bicycle approaching from the left.
Vergence	Angle between the eyes' line of sight.	Change from looking at the dashboard to the road.

(Source: Used with permission of the Institute of Transportation Engineers, Dewar, R, "Road Users," *Traffic Engineering Handbook*, 5th Edition, Chapter 2, Table 2-2, pg 8, 1999.)

Table 3.1: Full Alternative Text

Many of the other factors listed in Table 3.1 reflect the dynamic nature of the driving task and the fact that most objects to be viewed by drivers are in relative motion with respect to the driver's eyes.

As static visual acuity is the only one of these many visual factors that is examined as a prerequisite to issuing a driver's license, traffic engineers must expect and deal with significant variation in many of the other visual characteristics of drivers. Good static visual acuity is a key factor, as this is a prerequisite for other “good” vision characteristics. A driver with good static visual acuity could, for example, have poor dynamic visual acuity, poor depth perception, partial or complete color blindness, or other negative factors.

Fields of Vision

Figure 3.1 illustrates three distinct fields of vision, each of which is important to the driving task [2]:

- Acute or clear vision cone—3° to 10° around the line of sight; legend can be read only within this narrow field of vision.
- Fairly clear vision cone—10° to 12° around the line of sight; color and shape can be recognized in this field.
- Peripheral vision—This field may extend up to 90° to the right and left of the centerline of the pupil, and up to 60° above and 70° below the line of sight. Stationary objects are generally not seen in the peripheral vision field, but the movement of objects through this field is detected.

Figure 3.1: Fields of Vision Illustrated

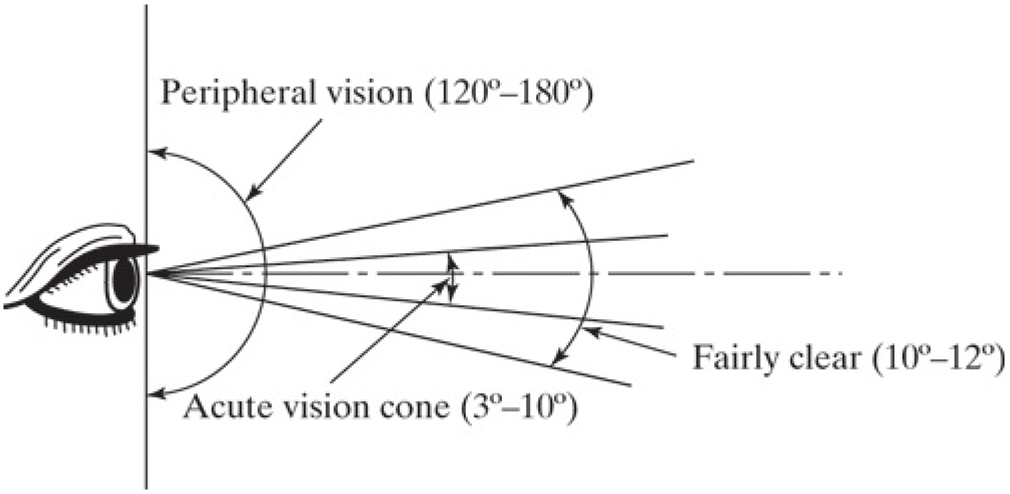


Figure 3.1: Full Alternative Text

These fields of vision, however, are defined for a stationary person. In particular, the peripheral vision field narrows, as speed increases, to as little as 100° at 20 mi/h and to 40° at 60 mi/h.

The driver's visual landscape is both complex and rapidly changing. Approaching objects appear to expand in size, while other vehicles and stationary objects are in relative motion both to the driver and to each other. The typical driver essentially samples the available visual information available and selects appropriate cues to make driving decisions.

The fields of vision affect a number of traffic engineering practices and functions. Traffic signs, for example, are placed so that they can be read within the acute vision field without requiring drivers to change their line of sight. Thus, they are generally placed within a 10° range of the driver's expected line of sight, which is assumed to be in line with the highway alignment. This leads to signs that are intended to be read when they are at a significant distance from the driver; in turn, this implies how large the sign and its lettering must be in order to be comprehended at that distance. Objects or other vehicles located in the fairly clear and peripheral vision fields may draw the driver's attention to an important event occurring in that field, such as the approach of a vehicle on an intersection street or driveway or a child running into the street after a ball. Once noticed, the driver may turn his/her head to examine the details of the situation.

Peripheral vision is the single most important factor when drivers estimate their speed. The movement of objects through the peripheral vision field is the driver's primary indicator of speed. Old studies have demonstrated time and again that drivers deprived of peripheral vision (using blinders in experimental cases) and deprived of a working speedometer have little idea of how fast they are traveling.

Important Visual Deficits

There are a number of visual problems that can affect driver performance and behavior. Unless the condition causes a severe visual disability, drivers affected by various visual deficits often continue to drive.

Reference [3] contains an excellent overview and discussion of these.

Some of the more common problems involve cataracts, glaucoma, peripheral vision deficits, ocular muscle imbalance, depth perception deficits, and color blindness. Drivers who undergo eye surgery to correct a problem may experience temporary or permanent impairments. Other diseases, such as diabetes, can have a significant negative impact on vision if not controlled. Some conditions, like cataracts and glaucoma, if untreated, can lead to blindness.

While color blindness is not the worst of these conditions, it generally causes some difficulties for the affected driver, since color is one of the principal means to impart information. Unfortunately, one of the most common forms of color blindness involves the inability to discern the difference between red and green. In the case of traffic signals, this could have a devastating impact on the safety of such drivers. To ameliorate this difficulty to some degree, some blue pigment has been added to green lights and some yellow pigment has been added to red lights, making them easier to discern by color-blind drivers. Also, the location of colors on signal heads has long been standardized, with red on the top and green on the bottom of vertical signal heads. On horizontal heads, red is on the left and green on the right. Arrow indications are either located on a separate signal head or placed below or to the right of ball indications on a mixed signal head.

3.2.2 Perception–Reaction Time

The second critical driver characteristic is perception–reaction time (PRT). During perception and reaction, there are four distinct processes that the driver must perform [4]:

- Detection or perception. In this phase, an object or condition of concern enters the driver’s field of vision, and the driver becomes consciously aware that something requiring a response is present.

- Identification. In this phase, the driver acquires sufficient information concerning the object or condition to allow the consideration of an appropriate response.

- Decision or emotion. Once identification of the object or condition is sufficiently completed, the driver must analyze the information and make a decision about how to respond.

- Response or volition. After a decision has been reached, the response is now physically implemented by the driver.

The total amount of time that this process takes is called the perception–reaction time. Some of the literature refers to this as "PIEV" time, named for the four individual actions making up the process.

Design Values

Like all human characteristics, PRTs vary widely among drivers, and are influenced by a variety of other factors, including the type and complexity of the event perceived and the environmental conditions at the time of the response.

Nevertheless, design values for various applications must be selected. The American Association of State Highway and Transportation Officials (AASHTO) mandates the use of 2.5 s for most computations involving braking reactions [5], based upon a number of research studies [6–9]. This value is believed to be approximately a 90th percentile criterion (i.e., 90% of all drivers will have a PRT as fast or faster than 2.5 s).

For signal timing purposes, the Institute of Transportation Engineers [10] recommends a PRT time of 1.0 s. Because of the simplicity of the response and the preconditioning of drivers to respond to signals, the PRT time is significantly less than that for a braking response on an open highway. While this is a lower value, it still represents an approximately 85th percentile for the particular situation of responding to a traffic signal.

AASHTO criteria, however, recognize that in certain more complex situations, drivers may need considerably more time to react than 1.0 s or 2.5 s. These are often referred to as *decision reaction times*. Table 3.2 summarizes PRT times in common use in traffic engineering.

Table 3.2: Recommended

PRT Times (AAHSTO, ITE)

Situation	Recommended PRT
Normal stop at a traffic signal	1.0 s
Normal stop on a highway	2.5 s
Avoidance maneuver: stop on a highway	3.0 s
Avoidance maneuver: stop on an urban road	9.1 s
Avoidance maneuver: speed/ path/direction change on a rural road	10.2 s–11.2 s
Avoidance maneuver: speed/ path/direction change on a suburban road	12.1 s–12.9 s
Avoidance maneuver: speed/ path/direction change on an urban road	14.0 s–14.5 s

Table 3.2: Full Alternative Text

Most of the "avoidance maneuver" categories involve complex situations requiring multiple actions from the driver. A driver might come up on a truck traveling at a very low speed, while weaving in and out of a lane. This information will take some time for the driver to process and make an appropriate decision on evasive actions.

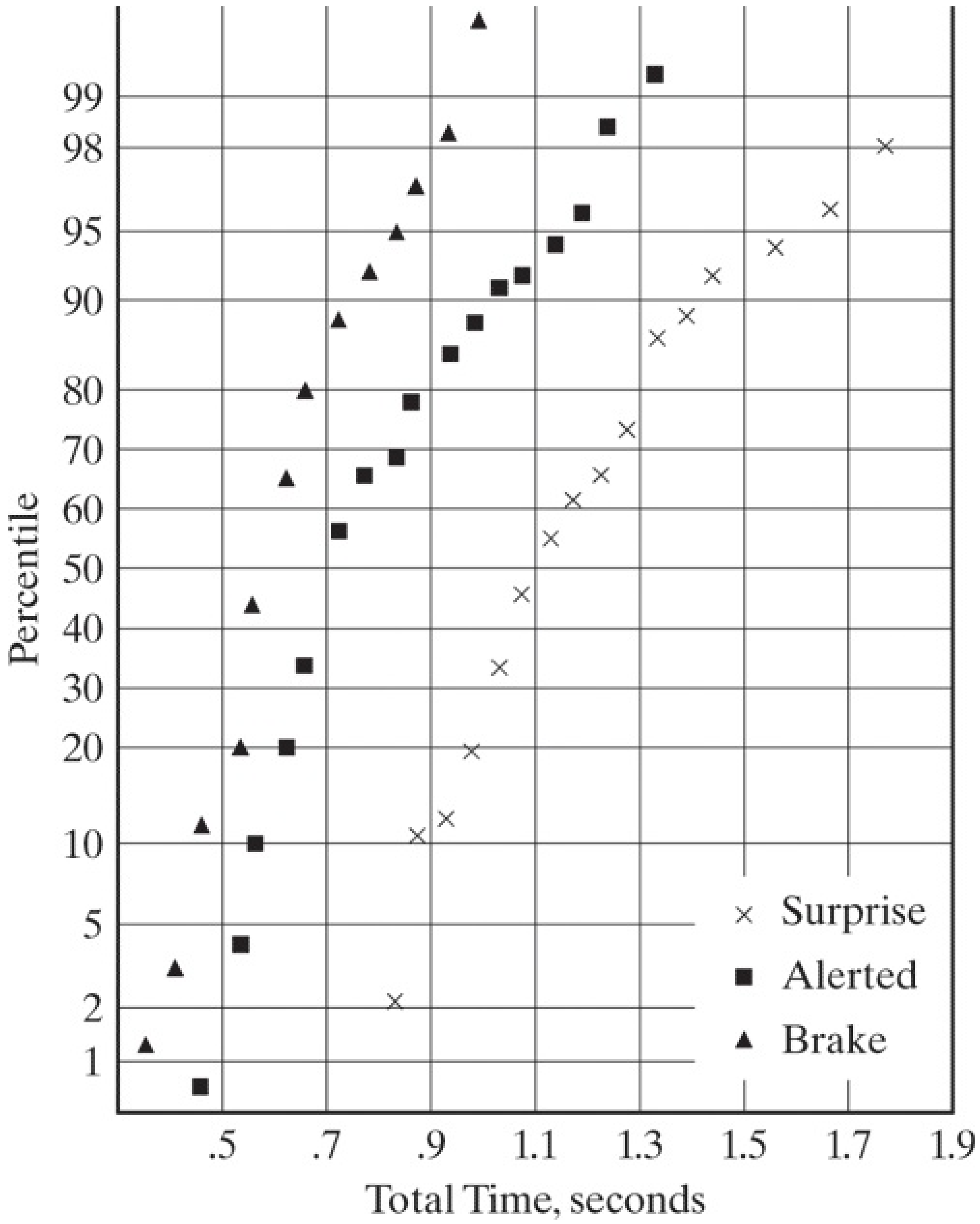
Expectancy

The concept of expectancy is important to the driving task and has a significant impact on the perception–reaction process and PRT. Simply put, drivers will react more quickly to situations they *expect* to encounter as opposed to those that they *do not expect* to encounter. There are three different types of expectancies:

- Continuity. Experiences of the immediate past are generally expected to continue. Drivers do not, for example, expect the vehicle they are following to suddenly slow down, without an obvious reason.
- Event. Things that have not happened previously will not happen. If no vehicles have been observed entering the roadway from a small driveway over a reasonable period of time, then the driver will assume that none will enter now.
- Temporal. When events are cyclic, such as a traffic signal, the longer a given state is observed, drivers will assume that it is more likely that a change will occur.

The impact of expectancy on PRT is illustrated in Figure 3.2. This study by Olsen, et al. [11] in 1984 was a controlled observation of student drivers reacting to a similar hazard when they were unaware that it would appear, and again where they were told to look for it. In a third experiment, a red light was added to the dash to initiate the braking reaction. The PRT under the “expected” situation was consistently about 0.5 s faster than under the “unexpected” situation.

Figure 3.2: Comparison of Perception–Reaction Times Between Expected and Unexpected Events



(Source: Used with permission of the Transportation Research Board, National Research Council, Olson, P., et al., "Parameters Affecting Stopping Sight Distance," *NCHRP Report 270*, Washington, D.C., 1984.)

Figure 3.2: Full Alternative Text

Given the obvious importance of expectancy on PRT, traffic engineers must strive to avoid designing "unexpected" events into roadway systems and traffic controls. If there are all right-hand ramps on a given freeway, for example, left-hand ramps should be avoided if at all possible. If absolutely required, guide signs must be very carefully designed to alert drivers to the existence and location of the left-hand ramp, so that when they reach it, it is no longer “unexpected.”

Other Factors Affecting PRT

In general, PRTs increase with a number of factors, including (1) age, (2) fatigue, (3) complexity of reaction, and (4) presence of alcohol and/or drugs in the driver’s system. While these trends are well documented, they are generally accounted for in recommended design values, with the exception of the impact of alcohol and drugs. The latter are addressed primarily through enforcement of ever-stricter DWI/DUI laws in the various states, with the intent of removing such drivers from the system, especially where repeated violations make them a significant safety risk. Some of the more general effects of alcohol and drugs, as well as aging, on driver characteristics are discussed in a later section.

Reaction Distance

The most critical impact of PRT is the distance the vehicle travels while the driver goes through the process. In the example of a simple braking reaction, the PRT begins when the driver first becomes aware of an event or object in his or her field of vision and ends when his or her foot is applied to the brake. During this time, the vehicle continues along its original course at its initial speed. Only after the foot is applied to the brake pedal does the vehicle begin to slow down in response to the stimulus.

The reaction distance is simply the PRT multiplied by the initial speed of the vehicle. As speed is generally in units of mi/h and PRT is in units of seconds, it is convenient to convert speeds to ft/s for use:

$$\frac{1\ \text{mi} \times \left(\frac{5{,}280\ \text{ft}}{\text{mi}}\right)}{1\ \text{h} \times \left(\frac{3{,}600\ \text{s}}{\text{h}}\right)} = 1.466666\cdots\ \frac{\text{ft}}{\text{s}} = 1.47\ \frac{\text{ft}}{\text{s}}$$

Thus, the reaction distance may be computed as

$$d_r = 1.47\ S\ t \quad [3\text{-}1]$$

where

d_r=reaction distance, ft,S=speed of vehicle, mi/h, and$t$=perception −reaction time, s.

The importance of this factor is illustrated in the following example: A driver rounds a curve at a speed of 60 mi/h and sees a truck overturned on the roadway ahead. How far will the driver's vehicle travel before the driver's foot reaches the brake? Applying the AASHTO standard of 2.5 s for braking reactions:

$$d_r = 1.47 \times 60 \times 2.5 = 220.5 \text{ft}$$

The vehicle will travel 220.5 ft (approximately 11–12 car lengths) before the driver even engages the brake. The implication of this is frightening. If the overturned truck is closer to the vehicle than 220.5 ft when noticed by the driver, not only will the driver hit the truck, he or she will do so at full speed—60 mi/h. Deceleration begins only when the brake is engaged —*after* the perception–reaction process has been completed.

3.2.3 Pedestrian Characteristics

One of the most critical safety problems in any highway and street system involves the interactions of vehicles and pedestrians. A substantial number of traffic accidents and fatalities involve pedestrians. This is not surprising, as in any contact between a pedestrian and a vehicle, the pedestrian is at a significant disadvantage.

Virtually all of the interactions between pedestrians and vehicles occur as pedestrians cross the street at intersections and at midblock locations. At signalized intersections, safe accommodation of pedestrian crossings is as critical as vehicle requirements in establishing an appropriate timing pattern. Pedestrian walking speed in crosswalks is the most important factor in the consideration of pedestrians in signal timing.

At unsignalized crossing locations, gap-acceptance behavior of pedestrians is another important consideration. "Gap acceptance" refers to the clear time intervals between vehicles encroaching on the crossing path and the behavior of pedestrians in "accepting" them to cross through.

Walking Speeds

Table 3.3 shows 50th percentile walking speeds for pedestrians of various ages. It should be noted that these speeds were measured as part of a controlled experiment [12] and not specifically at intersection or midblock crosswalks. Nevertheless, the results are interesting.

Table 3.3: 50th Percentile Walking Speeds for Pedestrians of Various Ages

Age (years)	50th Percentile Walking Speed (ft/s)	
	Males	Females
2	2.8	3.4
3	3.5	3.4
4	4.1	4.1
5	4.6	4.5
6	4.8	5.0
7	5.0	5.0
8	5.0	5.3
9	5.1	5.4
10	5.5	5.4
11	5.2	5.2
12	5.8	5.7
13	5.3	5.6
14	5.1	5.3
15	5.6	5.3
16	5.2	5.4
17	5.2	5.4
18	4.9	N/A
20–29	5.7	5.4
30–39	5.4	5.4
40–49	5.1	5.3
50–59	4.9	5.0
60+	4.1	4.1

(Source: Compiled from Eubanks, J., and Hill, P., *Pedestrian Accident Reconstruction and Litigation*, 2nd Edition, Lawyers & Judges Publishing Co., Tucson, AZ, 1999.)

Table 3.3: Full Alternative Text

One problem with standard walking speeds involves physically impaired pedestrians. A study of pedestrians with various impairments and assistive devices concluded that average walking speeds for virtually all categories were lower than the standard used in signal timing until recently (4.0 ft/s) [13]. Table 3.4 presents some of the results of this study. These and similar results of other studies suggest that more consideration needs to be given to the needs of handicapped pedestrians.

Table 3.4: Walking Speeds for Physically Impaired Pedestrians

Impairment/Assistive Device	Average Walking Speed (ft/s)
Cane/crutch	2.62
Walker	2.07
Wheelchair	3.55
Immobilized knee	3.50
Below-knee amputee	2.46
Above-knee amputee	1.97
Hip arthritis	2.44–3.66
Rheumatoid arthritis (knee)	2.46

(Source: Compiled from Perry, J., *Gait Analysis*, McGraw-Hill, New York, NY, 1992.)

Table 3.4: Full Alternative Text

Because of studies such as these, the approach to walking speeds has become more conservative where street crossings are involved. For pedestrian needs at signalized intersections, the *Manual on Uniform Traffic Control Devices*—referred to as the *MUTCD* [14]—now recommends the use of 3.5 ft/s for timing of pedestrian clearance intervals (flashing Upraised Hand), and 3.0 ft/s for total crossing time, which included the pedestrian WALK and the pedestrian clearance intervals.

Even lower speeds can be used where elderly or impaired pedestrians are thought to be present in significant numbers, such as near hospitals, senior residences, and similar types of facilities.

Gap Acceptance

When a pedestrian crosses at an uncontrolled (either by signals, STOP, or YIELD signs) location, either at an intersection or at a midblock location, the pedestrian must select an appropriate "gap" in the traffic stream through which to cross. The "gap" in traffic is measured as the time lag between two vehicles in any lane encroaching on the pedestrian's crossing path. As the pedestrian waits to cross, he or she views gaps and decides whether to "accept" or "reject" the gap for a safe crossing. Some studies have used a gap defined as the distance between the pedestrian and the approaching vehicle at the time the pedestrian begins his or her crossing. An early study [15] using the latter approach resulted in an 85th percentile gap of approximately 125 ft.

Gap-acceptance behavior, however, is quite complex and varies with a number of other factors, including the speed of approaching vehicles, the width of the street, the frequency distribution of gaps in the traffic stream, waiting time, and others. Nevertheless, this is an important characteristic that must be considered due to its obvious safety implications. Chapter 15, for example, presents warrants for (conditions justifying) the imposition of traffic signals. One of these is devoted entirely to the safety of pedestrian crossings.

Pedestrian Comprehension of Controls

One of the problems in designing controls for pedestrians is generally poor understanding of and poor adherence to such devices. One questionnaire survey of 4,700 pedestrians [16] detailed many problems of misunderstanding. The proper response to a flashing "DON'T WALK" (or flashing Upraised Hand) signal, for example, was not understood by 50% of road users, who thought it meant they should return to the curb from which they started. The meaning of this signal is to not start crossing while it is flashing; it is safe to complete a crossing if the pedestrian has already started to do so. Another study [17] found that violation rates for the solid “DON’T WALK” signal were higher than 50% in most cities, the use of the flashing “DON’T WALK” for pedestrian clearance was not well understood, and most pedestrians tend not to use pedestrian-actuated signals.

Most pedestrians do not understand the operation of a pedestrian push-button actuator at a signalized intersection. It does not provide an immediate WALK interval for the pedestrian. Rather, on the *next* signal cycle, the phase will be lengthened to accommodate a WALK interval. This may be anywhere between 30 s and 120 s after the time the pedestrian pushed the button. Most pedestrians don’t wait that long, and try to make an unsafe crossing. When the WALK interval finally arrives, the pedestrian is often gone.

The task of providing for a safe environment for pedestrians is not an easy one. The management and control of conflicts between vehicles and pedestrians remains a difficult one. These issues will be discussed in some detail when the use and implementation of various forms of traffic control, including signals, is discussed in subsequent chapters.

3.2.4 Impacts of Drugs and Alcohol on Road Users

The effect of drugs and alcohol on drivers has received well-deserved national attention for many years, leading to substantial strengthening of DWI/DUI laws and enforcement. These factors remain, however, a significant contributor to traffic fatalities and accidents.

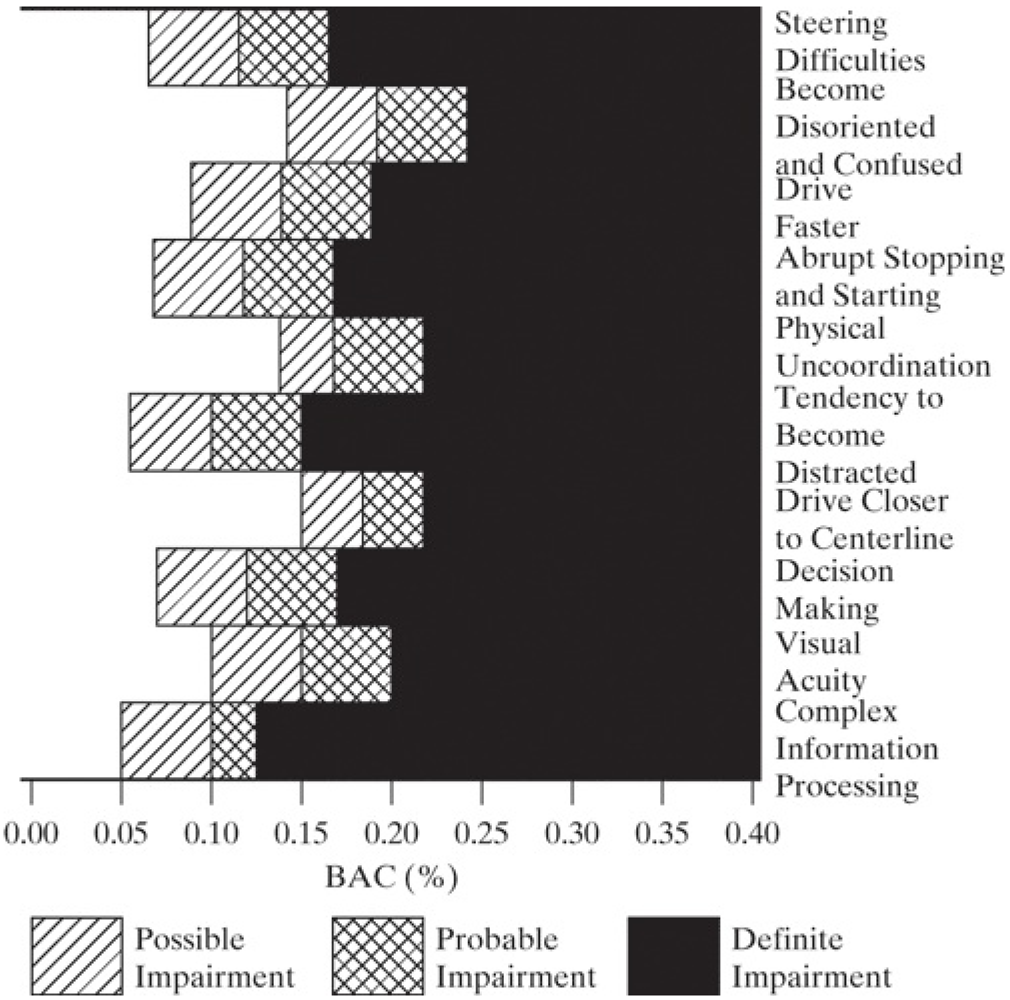
In 2015, there were 10,265 fatalities in crashes involving at least one driver with a blood-alcohol content (BAC) of 0.08 g/dL, the legal limit for impairment. This represented 29.3% of all traffic accident fatalities for the year. It is estimated that the economic cost of these fatalities was approximately $44 billion. The 2015 alcohol-related fatalities represented 3.2% increase over 2014. Total highway fatalities were 7.2% higher in 2015 than in 2014.

Of the 48,613 drivers involved in fatal crashes in 2015, 20% were legally impaired. This percentage is the same as it was in 2005 [18]. However, another 4% of these drivers had blood-alcohol levels between 0.01% and 0.08%. Of the 20% who were legally impaired, 13% had blood-alcohol levels over 0.15%.

Legal limits for DWI/DUI do not define the point at which alcohol and/or drugs influence the road user. Recognizing this is important for individuals to ensure safe driving, and is now causing many states to consider further reducing their legal limits on alcohol. Some states have instituted “zero tolerance” criteria (0.01%) for new drivers for the first year or two they are licensed.

Figure 3.3 is a summary of various studies on the effects of drugs and alcohol on various driving factors. Note that for many factors, impairment of driver function begins at levels well below the legal limits—for some factors at blood-alcohol levels as low as 0.05%.

Figure 3.3: Effects of Blood-Alcohol Level on Driving Tasks



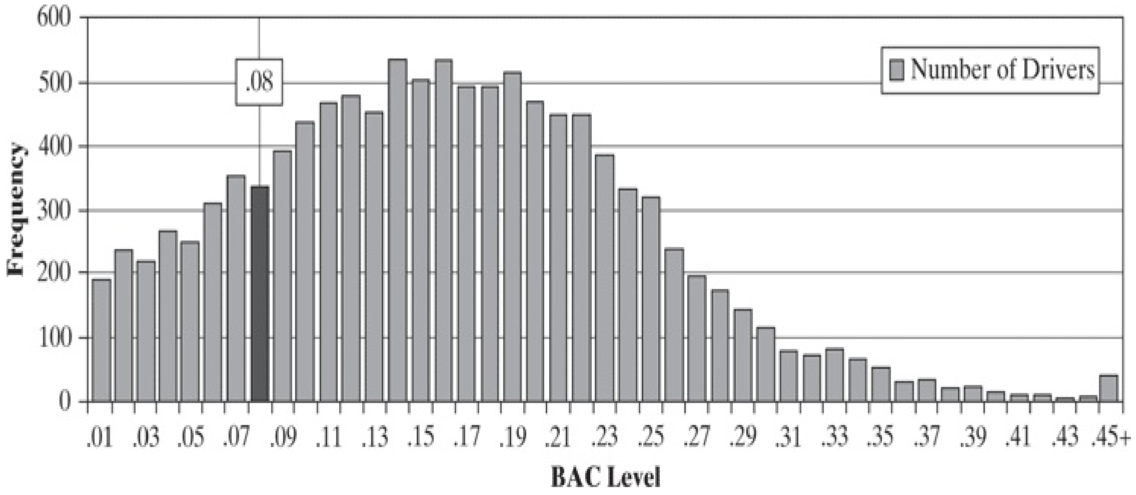
(Source: Used with permission of Institute of Transportation Engineers, Blaschke, J., Dennis, M., and Creasy, F., "Physical and Psychological Effects of Alcohol and Other Drugs on Drivers," *ITE Journal*, 59, Washington, D.C., 1987.)

Figure 3.3: Full Alternative Text

Figure 3.4 shows the distribution of blood-alcohol levels for drivers (including all drivers with BACs over 0.01 g/dL) for 2014. Clearly, there are significant numbers of drivers involved in fatal accidents at alcohol levels below the legal limit for DWI/DUI.

Figure 3.4: Distribution of Alcohol-Impaired Drivers in

Fatal Crashes, 2014 (Drivers with BAC ≥ 0.01 g/dL)



(Source: *Traffic Safety Facts: 2014 Data*, National Highway Traffic Safety Administration, U.S. Department of Transportation, Washington, D.C., 2014, Fig 3.)

Figure 3.4: Full Alternative Text

Severe impairment is also a problem, with 65% of the impaired drivers in fatal crashes (in 2015) having BACs in excess of 0.15%.

The bottom line is that an impaired driver is a dangerous driver, even when their BACs are below the legal definition of impairment. Impairment leads to longer PRT times, poor judgments, and actions that can and do cause accidents. Since few of these factors can be ameliorated by design or control (although good designs and well-designed controls help both impaired and unimpaired drivers), enforcement and education are critical elements in reducing the incidence of DWI/DUI and the accidents and deaths that result.

If impaired drivers are a menace, then impaired pedestrians are even more so—although the danger is mostly to themselves. In 2015, 5,376 pedestrians were killed in traffic accidents, an increase of 9.59% over 2014. In accidents involving a pedestrian fatality, 48% included either a driver or pedestrian with a BAC in excess of 0.08 g/dL. Of the pedestrians involved in these accidents, 34% were legally impaired, while only 14% of the drivers involved were legally impaired. Drunk walking is obviously extremely dangerous [19].

While the nation has made substantial progress in reducing the number of traffic fatalities overall, the success in reducing pedestrian fatalities has been quite limited. Between 2003 and 2012, total traffic fatalities were reduced from 42,884 to 33,461 (22%), pedestrian fatalities only decreased from 4,774 to 4,743 (0.64%). At least some of this can be attributed to impaired pedestrians.

While there has been a great deal of research on alcohol and driving impairment, there is far less information available about the influence of other drugs on traffic fatalities and crashes. This, however, is becoming more of an issue, as several states have legalized the use of recreational marijuana. Efforts to develop a “fast test” (like a breathalyzer for alcohol) to detect marijuana-impaired drivers are now underway. In 2009, 12,055 drivers killed in traffic crashes were tested for drug involvement. Of these, 33% were found to be drug-impaired. This was an increase from the results in 2005, when only 28% were found to be drug-impaired [20]. Obviously, this is becoming a critical issue in traffic safety.

Both motorists and pedestrians should also be aware of the impact of common prescription and over-the-counter medications on their performance capabilities. Many legitimate medications have effects that are similar to those of alcohol and/or marijuana. Users of medications should always be aware of the side effects of what they use (a most frequent effect of many drugs is drowsiness), and exercise care and good judgment when considering whether or not to drive. Some legitimate drugs can have a direct impact on blood-alcohol levels and can render a motorist legally intoxicated without “drinking.”

3.2.5 Impacts of Aging on Road Users

As life expectancy continues to rise, the number of older drivers has risen dramatically over the past several decades. Thus, it becomes increasingly important to understand how aging affects driver needs and limitations and how these should impact design and control decisions. Reference [21] is an excellent compilation sponsored by the National Academy of Sciences on a wide range of topics involving aging drivers.

Many visual acuity factors deteriorate with age, including both static and dynamic visual acuity, glare sensitivity and recovery, night vision, and speed of eye movements. Such ailments as cataracts, glaucoma, macular degeneration, and diabetes are also more common as people age, and these conditions have negative impacts on vision.

The increasing prevalence of older drivers presents a number of problems for both traffic engineers and public officials. At some point, deterioration of various capabilities must lead to revocation of the right to drive. On the other hand, driving is the principal means of mobility and accessibility in most parts of the nation, and the alternatives for those who can no longer drive are either limited or expensive. The response to the issue of an aging driver population must have many components, including appropriate licensing standards, consideration of some license restrictions on older drivers (e.g., a daytime only license), provision of efficient and affordable transportation alternatives, and increased consideration of their needs, particularly in the design and implementation of control devices and traffic regulations. Older drivers may be helped, for example, by such measures as larger lettering on signs, better highway lighting, larger and brighter signals, and other measures. Better education can serve to make older drivers more aware of the types of deficits they face and how to best deal with them. More frequent testing of key characteristics such as eyesight may also be helpful.

3.2.6 Psychological, Personality, and Related Factors

In the past few years, traffic engineers and the public in general have become acquainted with the term “road rage.” Commonly applied to drivers who lose control of themselves and react to a wide variety of situations violently, improperly, and almost always dangerously, the problem (which has always existed) is now getting well-deserved attention. “Road rage,” however, is a colloquial term, and is applied to everything from a direct physical assault by one road user on another to a variety of aggressive driving behaviors.

According to the testimony of Dr. John Larsen to the House Surface Transportation Subcommittee on July 17, 1997 (as summarized in Chapter 2 of Reference [1]), the following attitudes characterize aggressive drivers:

- The desire to get to one's destination as quickly as possible, leading to the expression of anger at other drivers/pedestrians who impede this desire.
- The need to compete with other fast cars.
- The need to respond competitively to other aggressive drivers.
- Contempt for other drivers who do not drive, look, and act as they do on the road.
- The belief that it is their right to "hit back" at other drivers whose driving behavior threatens them.

"Road rage" is the extreme expression of a driver's psychological and personal displeasure over the traffic situation he or she has encountered. It does, however, remind traffic engineers that drivers display a wide range of behaviors in accordance with their own personalities and psychological characteristics.

Once again, most of these factors cannot be addressed directly through design or control decisions and are best treated through vigorous enforcement and educational programs.

3.3 Vehicle Characteristics

In 2015, there were 263,610,219 registered vehicles in the United States. With a 2015 population of 320,000,000, this means that there was one registered vehicle for every 0.82 people in the United States, including children [22]. The characteristics of these vehicles vary as widely as those of the motorists who drive them. Table 3.5 summarizes these vehicles by type in four broad categories.

Table 3.5: U.S.-Registered Vehicles in 2015

Vehicle Type	Registered Vehicles
Passenger cars	112,864,228
Buses (school, transit, and intercity)	888,907
Trucks	141,256,148
Motorcycles	8,600,936
Total	**263,610,219**

Table 3.5: Full Alternative Text

Trucks and buses are not generally owned by individuals, but are more likely to be part of commercial fleets owned by various businesses, including trucking firms, transit systems, and the like. Nevertheless, there are more registered vehicles in the United States than there are licensed drivers (218,084,219 in 2015).

In general, motor vehicles are classified by AASHTO [5] into four main categories:

- Passenger cars—all passenger cars, SUVs, minivans, vans, and pickup trucks

- Buses—intercity motor coaches, transit buses, school buses, and articulated buses
- Trucks—single-unit trucks, tractor-trailer, and tractor-semi-trailer combination vehicles
- Recreational vehicles—motor homes, cars with various types of trailers (boat, campers, motorcycles, etc.)

This categorization is somewhat different from the national statistical summary of Table 3.5. Recreational vehicles are generally treated as trucks in national statistics. Motorcycles are not isolated as a separate category in AASHTO, as their characteristics do not usually limit or define design or control needs.

There are a number of critical vehicle properties that must be accounted for in the design of roadways and traffic controls. These include the following:

- Braking and deceleration
- Acceleration
- Low-speed turning characteristics
- High-speed turning characteristics

In more general terms, the issues associated with vehicles of vastly differing size, weight, and operating characteristics sharing roadways must also be addressed by traffic engineers.

3.3.1 Concept of the Design Vehicle

Given the immense range of vehicle types using street and highway facilities, it is necessary to adopt standard vehicle characteristics for design and control purposes. For geometric design, AASHTO has defined 20 “design vehicles,” each with specified characteristics. The 20 design vehicles are defined as follows:

P=passenger carSU-30=single-unit truck with two axlesSU-40=single-

unit truck with three axlesBUS-40=intercity bus with a 40-
ft wheelbaseBUS-45=intercity bus with a 45-ft wheelbaseCITY-
BUS=transit busS-BUS 36=conventional school bus for 65S-
BUS 40=large school bus for 84 passengersA-BUS=articulated busWB-
40=intermediate semi-trailer with 40-ft wheelbaseWB-62=interstate semi-
trailer with a 62-ft wheelbaseWB-67=interstate semi-trailer with a 67-
ft wheelbaseWB-67D="double-bottom" semi-trailer/trailer with a 67-
ft wheelbaseWB-92D=Rocky Mountain double semi-
trailer/trailer with a 92-ft wheelbaseWB-100T=triple semi-
trailer/trailers with a 100-ft wheelbaseWB-109D=turnpike double semi-
trailer/trailer with a 109-
ft wheelbaseMH=motor homeP/T=passenger car and camper trailerP/B=pas

Wheelbase dimensions are measured from the frontmost axle to the rearmost axle, including both the tractor and trailer in a combination vehicle.

Design vehicles are primarily employed in the design of turning roadways and intersection curbs, and are used to help determine appropriate lane widths, and such specific design features as lane-widening on curves. Key to such usage, however, is the selection of an appropriate design vehicle for various types of facilities and situations. In general, the design should consider the largest vehicle likely to use the facility with reasonable frequency.

In considering the selection of a design vehicle, it must be remembered that all parts of the street and highway network must be accessible to emergency vehicles, including fire engines, ambulances, emergency evacuation vehicles, and emergency repair vehicles, among others. Therefore, the single-unit truck is usually the minimum design vehicle selected for most local street applications. The mobility of hook-and-ladder fire vehicles is enhanced by having rear-axle steering that allows these vehicles to negotiate sharper turns than would normally be possible for combination vehicles; so the use of a single-unit truck as a design vehicle for local streets is not considered to hinder emergency vehicles.

The passenger car is used as a design vehicle only in parking lots, and even there, access to emergency vehicles must be considered. For most other classes or types of highways and intersections, the selection of a design vehicle must consider the expected vehicle mix. In general, the design vehicle selected should easily accommodate 95% or more of the expected vehicle mix.

The physical dimensions of design vehicles are also important considerations. Design vehicle heights range from 4.3 ft for a passenger car to 13.5 ft for the largest trucks. Overhead clearances of overpass and sign structures, electrical wires, and other overhead appurtenances should be sufficient to allow the largest anticipated vehicles to proceed. As all facilities must accommodate a wide variety of potential emergency vehicles, use of 14.0 ft for minimum clearances is advisable for most facilities.

The width of design vehicles ranges from 7.0 ft for passenger cars to 8.5 ft for the largest trucks (excluding special “wide load” vehicles such as a tractor pulling a prefabricated or motor home). This should influence the design of such features as lane width and shoulders. For most facilities, it is desirable to use the standard 12-ft lane width. Narrower lanes may be considered for some types of facilities when necessary, but given the width of modern vehicles, 10 ft is an absolute minimum for virtually all applications, and 11 ft is a commonly used reasonable minimum.

3.3.2 Turning Characteristics of Vehicles

There are two conditions under which vehicles must make turns:

- Low-speed turns (≤ 10 mi/h)
- High-speed turns (> 10 mi/h)

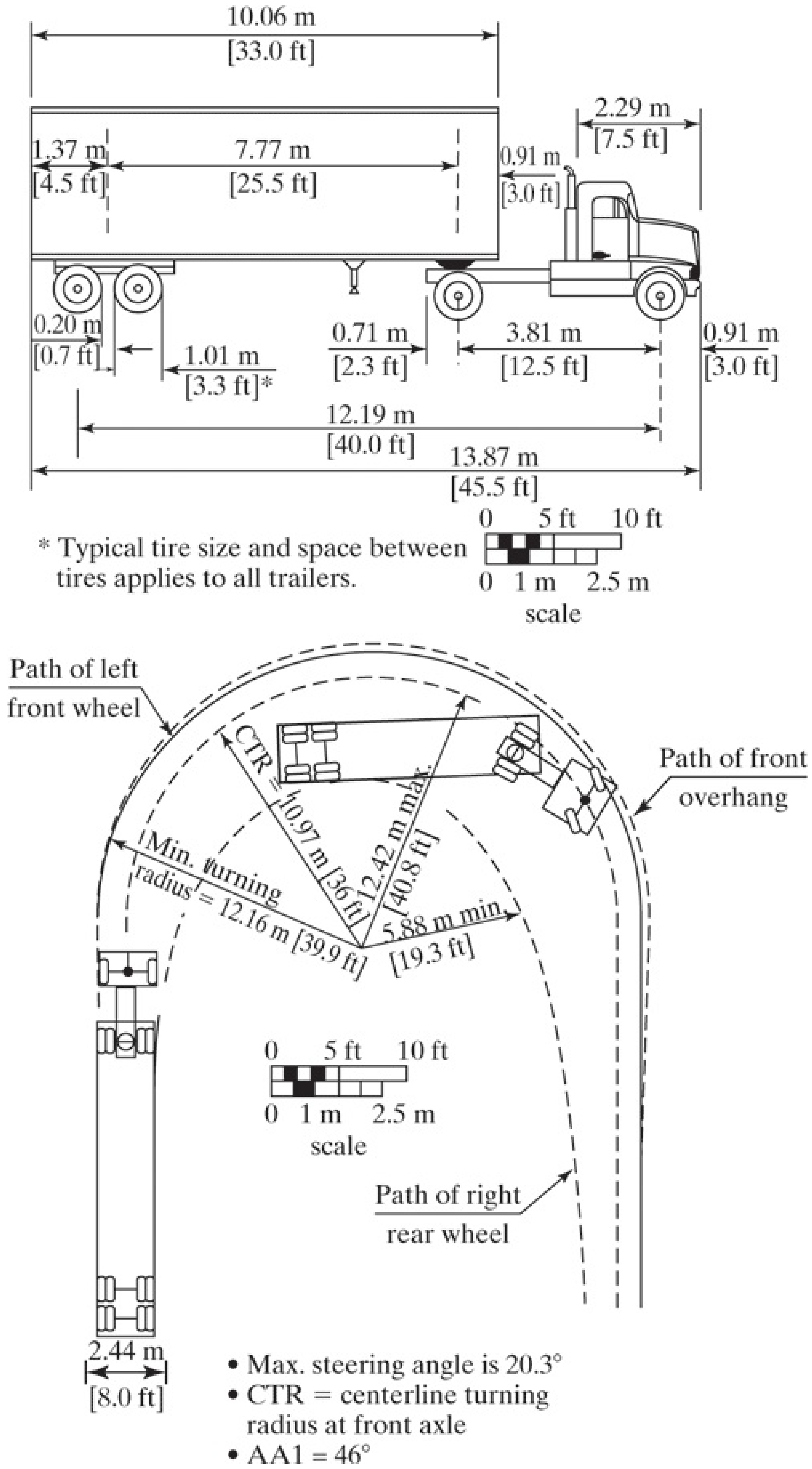
Low-speed turns are limited by the characteristics of the vehicle, as the minimum radius allowed by the vehicle’s steering mechanism can be supported at such speeds. High-speed turns are limited by the dynamics of side friction between the roadway and the tires, and by the superelevation (cross-slope) of the roadway.

Low-Speed Turns

AASHTO specifies minimum design radii for each of the design vehicles, based on the centerline turning radius and minimum inside turning radius of each vehicle. While the actual turning radius of a vehicle is controlled by the front wheels, rear wheels do not follow the same path. They "off-track" as they are dragged through the turning movement.

Reference [5] contains detailed low-speed turning templates for all AASHTO design vehicles. An example (for a WB-40 combination vehicle) is shown in Figure 3.5. Note that the minimum turning radius is defined by the track of the front outside wheel. The combination vehicle, however, demonstrates considerable "off-tracking" of the rear inside wheel, effectively widening the width of the "lane" occupied by the vehicle as it turns. The path of the rear inside wheel is not circular, and has a variable radius.

Figure 3.5: Low-Speed Turning Template for WB-40 Combination Vehicles



10.06 m
[33.0 ft]
2.29 m
[7.5 ft]
1.37 m
[4.5 ft]
7.77 m
[25.5 ft]
0.91 m
[3.0 ft]
0.20 m
[0.7 ft]
1.01 m
[3.3 ft]*
0.71 m
[2.3 ft]
3.81 m
[12.5 ft]
0.91 m
[3.0 ft]
12.19 m
[40.0 ft]
13.87 m
[45.5 ft]
* Typical tire size and space between tires applies to all trailers.
0 5 ft 10 ft
0 1 m 2.5 m
scale
Path of left front wheel
Path of front overhang
CTR = 10.97 m [36 ft]
12.42 m max. [40.8 ft]
Min. turning radius = 12.16 m [39.9 ft]
5.88 m min. [19.3 ft]
0 5 ft 10 ft
0 1 m 2.5 m
scale
Path of right rear wheel
2.44 m
[8.0 ft]
• Max. steering angle is 20.3°
• CTR = centerline turning radius at front axle
• AA1 = 46°

(Source: Adapted from *A Policy on Geometric Design of Highways and Streets*, 2011, by the American Association of State Highway and Transportation Officials, Washington, D.C. U.S.A)

Figure 3.5: Full Alternative Text

Turning templates provide illustrations of the many different dimensions involved in a low-speed turn. In designing for low-speed turns, the minimum design turning radius is the minimum centerline radius plus one-half of the width of the front of the vehicle.

Minimum design turning radii range from 23.8 ft for a passenger car to a high of 82.0 ft for the WB-92D double tractor-trailer combination vehicle. Depending upon the specific design vehicle, the minimum inside curb radius is generally considerably smaller than the minimum design turning radius, reflecting the variable radius of the rear-inside wheel's track.

Table 3.6 summarizes the minimum turning radii and minimum inside curb radii for the various defined design vehicles.

Table 3.6: Minimum Low-Speed Turning Radii for AASHTO Design Vehicles

Design Vehicle	Minimum Turning Radius (ft)	Minimum Inside Curb Radius (ft)
P	23.8	14.4
SU-30	41.8	28.4
SU-40	51.2	36.4
BUS-40	41.7	24.3
BUS-45	44.0	24.7
CITY-BUS	41.6	24.5
S-BUS 36	38.6	23.8
S-BUS 40	39.1	25.3
A-BUS	39.4	21.3
WB-40	39.9	19.3
WB-62	44.8	7.4
WB-67	44.8	7.9
WB-67D	44.8	19.1
WB-92D	82.0	55.6
WB-100T	44.8	9.7
WB-109D	59.9	13.8
MH	39.7	26.0
P/T	32.9	18.8
P/B	23.8	8.0
MH/B	49.8	35.0

Table 3.6: Full Alternative Text

In designing intersections, off-tracking characteristics of the design vehicle should be considered when determining how far from travel lanes to locate (or cut back) the curb. In a good design, the outside wheel of the turning design vehicle should be able to negotiate its path without “spilling over” into adjacent lanes as the turn is negotiated. This requires that the curb setback must accommodate the maximum off-tracking of the design vehicle.

High-Speed Turns

When involved in a high-speed turn on a highway curve, centripetal forces of momentum are exerted on the vehicle to continue in a straight path. To hold the curve, these forces are opposed by side friction and superelevation.

Superelevation is the cross-slope of the roadway, always with the lower edge in the direction of the curve. The sloped roadway provides an element of horizontal support for the vehicle. Side-friction forces represent the resistance to sliding provided across the plane of the surface between the vehicle's tires and the roadway. From the basic laws of physics, the relationship governing vehicle operation on a curved roadway is

e+f1−ef=S2gR [3-2]

where

e=superelevation rate of the roadway, ft/ft (dimensionless),f=coefficient of

The superelevation rate is the total rise in elevation across the travel lanes of the cross-section (ft) divided by the width of the travel lanes (ft), expressed as a decimal. Some publications, including AASHTO, express superelevation rate as a percentage.

Equation 3-2 is simplified by noting that the term "*ef*" is extremely small, and may be ignored for the normal range of superelevation rates and *side-friction factors*. It is also convenient to express vehicle speed in mi/h. Thus:

e+f1=(1.47 S)232.2 R

This yields the more traditional relationship used to depict vehicle operation on a curve:

R=S215 (e+f) [3-3]

where all terms are as previously defined, except that "*S*" is the speed in mi/h rather than ft/s, as in Equation 3-2.

The normal range of superelevation rates is from a minimum of approximately 0.005 to support side drainage to a maximum of 0.12. As speed increases, higher superelevation rates are used. Where icing conditions are expected, the maximum superelevation rate is generally limited to 0.08 to prevent a stalled vehicle from sliding toward the inside of the curve.

Coefficients of side friction for design are based upon wet roadway conditions. They vary with speed and are shown in Table 3.7.

Table 3.7: Side-Friction Factors (*f*) for Wet Pavements at Various Speeds

Speed (mi/h)	30	40	50	60	70
f	0.16	0.15	0.14	0.12	0.10

Table 3.7: Full Alternative Text

Theoretically, a road can be banked to fully oppose centripetal force without using side friction at all. This is, of course, generally not done, as vehicles travel at a range of speeds and the superelevation rate required in many cases would be excessive. High-speed turns on a flat pavement may be fully supported by side friction as well, but this generally limits the radius of curvature or speed at which the curve may be safely traversed.

Chapter 27 treats the design of horizontal curves and the relationships among superelevation, side friction, curve radii, and design speed in greater detail.

Equation 3-3 can be used in a number of ways as illustrated in the following sample problems. In design, a minimum radius of curvature is computed based on maximum values of e and f.

Sample Problem 3-1: Estimating

Minimum Radius of Curvature

A roadway has a design speed of 65 mi/h, with maximum values of *e* = 0.08 and *f* = 0.11, determine the minimum radius of curvature that can be used.

The minimum radius is found as:

$$R=65215\ (0.08+0.11)=1{,}482.5\ \text{ft}$$

Sample Problem 3-2: Estimating Maximum Safe Speed on a Horizontal Curve

If a highway curve with radius of 800 ft has a superelevation rate of 0.06, estimate the maximum safe speed. This computation requires that the relationship between the *coefficient of side friction*, *f*, and speed, as indicated in Table 3.7, be taken into account. Solving Equation 3-3 for *S* yields

$$S=15\ R\ (e+f)\ [3\text{-}4]$$

For the example given, the equation is solved for the given values of *e* (0.06) and *R* (800 ft) using various values of *f* from Table 3.7. Computations continue until there is closure between the computed speed and the speed associated with the coefficient of side friction selected. Thus:

$$S=15 \times 800 \times (0.06 + f)S=15 \times 800 \times (0.06 + 0.10)=43.8\ \text{mi/h (70 mi/h assu}$$

The correct result is obviously between 49.0 and 50.2 mi/h. If straight-line interpolation is used:

$$S=49.0+(50.2-49.0)\times[(50.0-49.0)(50.2-49.0)+(50.2-40.0)]=49.1\ \text{mi/h}$$

Thus, for the curve as described, 49.1 mi/h is the maximum safe speed at which it should be negotiated.

It must be noted that this is based on the design condition of a wet pavement and that higher speeds would be possible under dry conditions.

3.3.3 Braking Characteristics

Another critical characteristic of vehicles is their ability to stop (or decelerate) once the brakes have been engaged. Again, basic physics relationships are used. The distance traveled during a stop is the average speed during the stop multiplied by the time taken to stop, or

$$d_b = \left(\frac{S}{2}\right) \times \left(\frac{S}{a}\right) = \frac{S^2}{2a} \quad [3\text{-}5]$$

where

d_b = braking distance, ft, S = initial speed of the vehicle, ft/s, and a = deceleratic

It is convenient, however, to express speed in mi/h, yielding:

$$d_b = \frac{(1.47S)^2}{2a} = 1.075\frac{S^2}{a}$$

where *S* is the speed in mi/h. Note that the 1.075 factor is derived from the more exact conversion factor between mi/h and ft/s, 1.46666..... It is often also useful to express this equation in terms of the coefficient of forward rolling or skidding friction, *F*, where $F = a/g$ (or $a = Fg$), where *g* is the acceleration due to gravity, 32.2 ft/s^2. Then:

$$d_b = 1.075\,\frac{S^2}{Fg} = 1.075\,\frac{S^2}{32.2\,F} = \frac{S^2}{30\,F}$$

When the effects of grade are considered, and where a braking cycle leading to a reduced speed other than “0” are considered, the equation becomes

$$d_b = \frac{S_i^2 - S_f^2}{30\,(F \pm G)} \quad [3\text{-}6]$$

where

S_i = initial speed of the vehicle, mi/h, S_f = final speed of the vehicle (after the

When there is an upgrade, a “+” is used; a “−” is used for downgrades. This results in shorter braking distances on upgrades, where gravity helps deceleration, and longer braking distances on downgrades, where gravity causes acceleration.

In previous editions of AASHTO, braking distances were based on coefficients of forward skidding friction on wet pavements that varied with speed. In the latest standards, however, a standard deceleration rate of 11.2 ft/s^2 is adopted as a design rate. This is viewed as a rate that can be developed on wet pavements by most vehicles. It is also expected that 90% of drivers will decelerate at higher rates. This, then, suggests a standard friction factor for braking distance computations of F = 11.2/32.2 = 0.348 and Equation 3-6 becomes

$$d_b = \frac{S_i^2 - S_f^2}{30\,(0.348 \pm G)} \quad [3\text{-}7]$$

Sample Problem 3-3: Estimating Braking Distance

Consider the following case: Once the brakes are engaged, what distance is covered bringing a vehicle traveling at 60 mi/h on a 0.03 downgrade to a complete stop ($S_f = 0$). Applying Equation 3-7:

$$d_b = \frac{60^2 - 0^2}{30\,(0.348 - 0.03)} = 377.4 \text{ ft}$$

The braking distance formula is also a favorite tool of accident investigators. It can be used to estimate the initial speed of a vehicle using measured skid marks and an estimated final speed based on damage assessments. In such cases, actual estimated values of F are used, rather than the standard design value recommended by AASHTO. Thus, Equation 3-6 is used.

Sample Problem 3-4: Use of Braking Formula in Crash

Investigations

An accident investigator estimates that a vehicle hit a bridge abutment at a speed of 20 mi/h, based on his or her assessment of damage. Leading up to the accident location, he or she observes skid marks of 100 ft on the pavement (*F* = 0.35) and 75 ft on the grass shoulder (*F* = 0.25). There is no grade. An estimation of the speed of the vehicle at the beginning of the skid marks is desired.

In this case, Equation 3-6 is used to find the initial speed of the vehicle, S_i, based upon a known (or estimated) final speed, S_f. Each skid must be analyzed separately, starting with the grass skid (for which a final speed has been estimated from the observed vehicle damage). Then:

db=75=S i2 − 20230(0.25)Si=(75 × 30 × 0.25) + 202=962.5=31.0 mi/h

This is the estimated speed of the vehicle at the *start* of the grass skid; it is also the speed of the vehicle at the *end* of the pavement skid. Then:

db=00=S i2 − 962.530 × 0.35Si=
(100 × 30 × 0.35) + 962.5=2012.5=44.9 mi/h

It is, therefore, estimated that the speed of the vehicle immediately before the pavement skid was 44.9 mi/h. This, of course, can be compared with the speed limit to determine whether excessive speed was a factor in the accident.

3.3.4 Acceleration Characteristics

The flip side of deceleration is acceleration. Passenger cars are able to accelerate at significantly higher rates than commercial vehicles. Table 3.8 shows typical maximum acceleration rates for a passenger car with a weight-to-horsepower ratio of 30 lbs/hp and a tractor-trailer with a ratio of 200 lbs/hp.

Table 3.8: Acceleration

Characteristics of a Typical Car versus a Typical Truck on Level Terrain

Speed Range (mi/h)	Acceleration Rate (ft/s^2) for:	
	Typical Car (30 lbs/hp)	Typical Truck (200 lbs/hp)
0–20	7.5	1.6
20–30	6.5	1.3
30–40	5.9	0.7
40–50	5.2	0.7
50–60	4.6	0.3

(Source: Used with permission from *Traffic Engineering Handbook*, 5th Edition, Institute of Transportation Engineers, Washington, D.C., 2000, Chapter 3, Tables 3-9 and 3-10.)

Table 3.8: Full Alternative Text

Acceleration is highest at low speeds and decreases with increasing speed. The disparity between passenger cars and trucks is significant.

Consider the distance required for a vehicle to accelerate to target speed. The distance is the time taken to accelerate to the target speed x the average speed during acceleration, or

da=(Sa)×(S2) [3-8]

where

da=acceleration distance, ft,S=target speed, ft/s, anda=acceleration rate, ft/s

Again, it is useful to convert the equation for the use of speed in mi/h:

da=(1.47 Sa)×(1.47 S2)=1.075 (S2a) [3-9] where *S* is now in units of mi/h.

Once again, note that the 1.075 factor is derived using the more precise factor for converting mi/h to ft/s (1.466666.....).

Sample Problem 3-5: Acceleration Impacts

Consider the difference in acceleration distance for a passenger car and truck to accelerate from a standing stop to 20 mi/h. From Table 3.8, the acceleration rate for a passenger car is 7.5 ft/s^2, while the acceleration rate for a typical truck is 1.6 ft/s^2.

Then:

For the passenger car:

$$da=1.075\,(2027.5)=57.3\ \text{ft}$$

For the truck:

$$da=1.075\,(2021.6)=268.8\ \text{ft}$$

The disparity is striking. If a car is at a “red” signal behind a truck, the truck will significantly delay the car. If a truck is following a car in a standing queue, a large gap between the two will occur as they accelerate.

Unfortunately, there is not much that can be done about the disparity indicated in Sample Problem 3-5 in terms of design and control. In the analysis of highway capacity, however, the disparity between trucks and cars in terms of acceleration and in terms of their ability to sustain speeds on upgrades leads to the concept of “passenger car equivalency.” Depending on the type of facility, severity and length of grade, and other factors, one truck may consume as much roadway capacity as six to seven or more passenger cars. Thus, the disparity in key operating characteristics of trucks and passenger cars is taken into account in design by providing additional capacity as needed.

3.3.5 Total Stopping Distance and Applications

The total distance to bring a vehicle to a full stop, from the time the need to do so is first noted, is the sum of the reaction distance, dr, and the braking distance, db. If Equations 3-1 (for dr) and 3-8 (for db) are combined, the total stopping distance becomes

$$d_S=d_r+d_b \quad d_S=1.47\ S_i\ t+\frac{S_i^2-S_f^2}{30\ (0.348\pm G)} \quad [3\text{-}10]$$

where all variables are as previously defined.

The concept of total stopping distance is critical to many applications in traffic engineering. Three of the more important applications are discussed in the sections that follow.

Safe Stopping Sight Distance

One of the most fundamental principles of highway design is that the driver must be able to see far enough to avoid a potential hazard or collision. Thus, on all roadway sections, the driver must have a sight distance that is at least equivalent to the total stopping distance required at the design speed.

Essentially, this requirement addresses this critical concern: A driver rounding a horizontal curve and/or negotiating a vertical curve is confronted with a downed tree, an overturned truck, or some other situation that completely blocks the roadway. The only alternative for avoiding a collision is to stop. The design must provide visibility for one safe stopping distance at every point along the roadway. By ensuring this, the driver can never be confronted with the need to stop without having sufficient distance to do so.

Sample Problem 3-6: Safe

Stopping Distance

Consider a section of rural freeway with a design speed of 70 mi/h. On a section of level terrain, what safe stopping distance must be provided? Equation 3-10 is used with a final speed (Sf) of "0" and the AASHTO standard reaction time of 2.5 s. Then:

dS=1.47 (70) (2.5)+702−0230 (0.348±0.0)=257.3+469.3=726.6 ft

The results of Sample Problem 3-6 mean that for the entire length of this roadway section drivers must be able to see at least 726.6 ft ahead. Providing this safe stopping sight distance will limit various elements of horizontal and vertical alignment, as discussed in Chapter 27. Not doing so exposes drivers to the risk of seeing an object blocking the road without adequate time to stop. This is explored in Sample Problem 3-7.

Sample Problem 3-7: The Cost of Not Providing Safe Stopping Sight Distance

What could happen, if a section of the roadway described in Sample Problem 3-6 provided a sight distance of only 500 ft? It would now be possible that a driver would initially notice an obstruction when it is only 500 ft away. If the driver were approaching at the design speed of 70 mi/h, a collision would occur. Again, assuming design values of reaction time and forward skidding friction, Equation 3-10 could be solved for the collision speed (i.e., the final speed of the deceleration cycle), using a known deceleration distance of 500 ft:

500=1.47 × 70 × 2.5 + 702 − Sf230(0.348)500 − 257.3=242.7=702 − Sf210.442,533.8=4,900 − Sf2Sf=4,900 − 2,533.8=48.6 mi/h

If the assumed conditions hold, a collision at 48.6 mi/h would occur. Of course, if the weather were dry and the driver had faster reactions than the design value (remember, 90% of drivers do), the collision might occur at a lower speed, and might be avoided altogether. The point is that such a collision *could* occur if the sight distance were restricted to 500 ft.

Decision Sight Distance

While every point and section of a highway must be designed to provide at least safe stopping sight distance, there are some sections that should provide greater sight distance to allow drivers to react to potentially more complex situations than a simple stop. Previously, reaction times for collision avoidance situations were cited [5].

Sight distances based upon these collision-avoidance decision reaction times are referred to as “decision sight distances.” AASHTO recommends that decision sight distance be provided at interchanges or intersection locations where unusual or unexpected maneuvers are required: changes in cross-section such as lane drops and additions, toll plazas, and intense-demand areas where there is substantial “visual noise” from competing information (e.g., control devices, advertising, roadway elements).

The decision sight distance is found by using Equation 3-10, replacing the standard 2.5 s reaction time for stopping maneuvers with the appropriate collision avoidance reaction time for the situation from Table 3.2.

Sample Problem 3-8: Decision Sight Distance, Assuming a Stop is Required

Consider the decision sight distance required for a freeway section with a 60 mi/h design speed approaching a busy urban interchange with many competing information sources. The approach is on a 0.03 downgrade. For this case, Table 3.2 suggests a reaction time up to 14.5 s to allow for complex path and speed changes in response to conditions. The decision sight distance is still based on the assumption that a worst case would require a complete stop. Thus, the decision sight distance would be

$$d=(1.47\times60\times14.5)+[\ 602-0230(0.348-0.03\ ]=1{,}278.9+377.4=1{,}656.3\ \text{ft}$$

AASHTO criteria for decision sight distances do not assume a stop maneuver for the speed/path/direction changes required in the most complex situations. The criteria, which are shown in Table 3.9, replace the braking distance in these cases with maneuver distances consistent with maneuver times between 3.5 and 4.5 s. During the maneuver time, the initial speed is assumed to be in effect. Thus, for maneuvers involving speed, path, or direction change on rural, suburban, or urban roads, Equation 3-11 is used to find the decision sight distance.

$$d=1.47\,(t_r+t_m)\,S_i \quad [3\text{-}11]$$

Table 3.9: Decision Sight Distances Resulting from Equations 3-10 and 3-11

Design Speed (mi/h)	Assumed Maneuver Time (s)	Decision Sight Distance for Avoidance Maneuver (ft)				
		A (Eqn 3-10)	B (Eqn 3-10)	C (Eqn 3-11)	D (Eqn 3-11)	E (Eqn 3-11)
Reaction Time (s)		3	9.1	11.2	12.9	14.5
30	4.5	219	488	692	767	838
40	4.5	330	688	923	1023	1117
50	4.0	460	908	1117	1242	1360
60	4.0	609	1147	1341	1491	1632
70	3.5	778	1406	1513	1688	1852
80	3.5	966	1683	1729	1929	2117

A: Stop on a rural road.
B: Stop on an urban road.
C: Speed/path/direction change on a rural road.
D: Speed/path/direction change on a suburban road.
E: Speed/path/direction change on an urban road.

Table 3.9: Full Alternative Text

where

t_r=reaction time for appropriate avoidance maneuver, s, and t_m=maneuver t

The criteria for decision sight distance shown in Table 3.9 are developed from Equations 3-10 and 3-11 for the decision reaction times indicated for the five defined avoidance maneuvers.

Sample Problem 3-9: Decision Sight Distance Based on AASHTO Criteria

Consider the result of Sample Problem 3-8. What decision distance would be required for the roadway described in Sample Problem 3-9 using AASHTO criteria? AASHTO would not assume that a stop is required. At 60 mi/h, a maneuver time of 4.0 s is used with the 14.5 s reaction time, and

$$d=1.47\,(14.5+4.0)\,60=1{,}631.7\ \text{ft}$$

Note that the result is not very different from Sample Problem 3-8 in this case.

Other Sight Distance Applications

In addition to safe stopping sight distance and decision sight distance, AASHTO also sets criteria for (1) passing sight distance on two-lane rural highways and (2) intersection sight distances for various control options. Intersection sight distances are treated in Chapter 15.

Passing sight distance on two-lane rural highways is a critical issue in the safe design of these types of facilities. On multilane highways, the objective is to always provide the driver with at least the safe stopping distance. This is also true of two-lane highways.

An additional issue arises on two-lane highways, however: passing maneuvers take place in the opposing lane of traffic when the opportunity exists. In this situation, the passing vehicle must assess the availability of a safe gap in the opposing traffic, move into the opposing lane, overtake and pass the slower vehicle(s), and safely return to the proper lane. All of this must be done as a potential vehicle in the opposing lane is approaching at considerable speed.

Passing on two-lane highways is not permitted at all locations. It is only permitted when the *passing sight distance* is available. Passing sight distance is sufficient for a driver to assess the desired maneuver and safely complete all aspects of it before an opposing vehicle imposes a hazard. Many models have been used over the years to analyze required passing sight distances for safety. For some time, criteria presented in the *MUTCD* conflicted with criteria presented in AASHTO. The current AASHTO criteria [5] now align with those of the *MUTCD* [14]. These criteria are summarized in Table 3.10.

Table 3.10: Passing Sight Distances for Two-Lane Rural Highways

Design Speed (mi/h)	Assumed Maneuver Speeds (mi/h)		Passing Sight Distance (ft)
	Passed Vehicle	Passing Vehicle	
20	8	20	400
25	13	25	450
30	18	30	500
35	23	35	550
40	28	40	600
45	33	45	700
50	38	50	800
55	43	55	900
60	48	60	1,000
65	53	65	1,100
70	58	70	1,200
75	63	75	1,300
80	68	80	1,400

(Source: Adapted from (*A Policy on Geometric Design of Streets and Highways*), (2011), by the American Association of State Highway and Transportation Officials, Washington, D.C. U.S.A.)

Table 3.10: Full Alternative Text

Note that assumed maneuver speeds are somewhat conservative. The passed vehicle is assumed to be traveling at a speed 12 mi/h below the design speed of the facility, while the passing vehicle is assumed to travel at the design speed. In reality, passing vehicles often travel at higher speeds, particularly while in the opposing lane.

Wherever these passing sight distances are *not* available, passing must be prohibited. Signs and markings are used to mark “No Passing” zones on such highways.

Change (Yellow) and Clearance (All-Red) Intervals for a Traffic Signal

The yellow interval for a traffic signal is designed to allow a vehicle that cannot comfortably stop when the green is withdrawn to enter the intersection legally. Consider the situation shown in Figure 3.6.

Figure 3.6: Timing Yellow and All-Red Intervals at a Signal

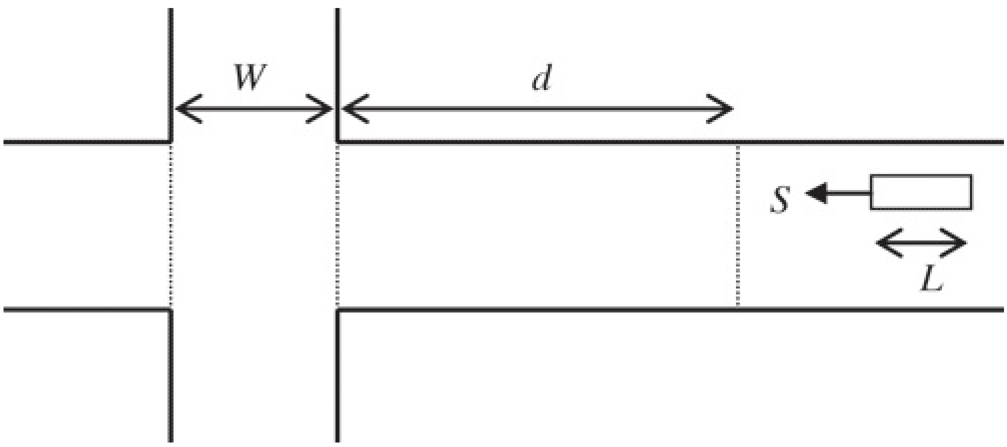


Figure 3.6: Full Alternative Text

In Figure 3.6, *d* is the safe stopping distance. At the time the green is withdrawn, a vehicle at *d* or less feet from the intersection line will not be able to stop, assuming normal design values hold. A vehicle further away than *d* would be able to stop without encroaching into the intersection area. The yellow signal is timed to allow a vehicle that cannot stop to traverse distance *d* at the approach speed (*S*). A vehicle may legally enter the intersection on yellow (in most states).

Having entered the intersection legally, the all-red period must allow the vehicle to cross the intersection width (*W*) and clear the back end of the vehicle (*L*) past the far intersection line (at a minimum).

Thus, the yellow interval must be timed to allow a vehicle to traverse the safe stopping distance. Sample Problem 3-10 illustrates how the length of the yellow and all-red intervals can be determined.

Sample Problem 3.10: Timing “Yellow” and “All-Red” Signal Intervals

Consider a case in which the approach speed to a signalized intersection is 40 mi/h. The grade is level, and the standard reaction time for a “Red” signal is 1.0 s. How long should the yellow and all-red intervals be?

The safe stopping distance is computed using a standard reaction time of 1.0 s for signal timing and level grade:

$$d=1.47 \times 40 \times 1.0 + 402 - 0230(0.348)=58.8 + 153.3=212.1\text{ft}$$

The length of the yellow signal is the time it takes an approaching vehicle to traverse 212.1 ft at 40 mi/h, or

$$y=212.11.47 \times 40=3.6\text{s}$$

A vehicle can, therefore, legally enter the intersection during the last instant of the yellow signal. Such a vehicle must be allowed to safely cross the intersection width (*W*) and the length of the vehicle (*L*) before conflicting vehicles are allowed to enter the intersection. This is the purpose of the all-red signal. If the street width in this case was 50 ft, and the length of the vehicle 20 ft, it would have to be

$$ar=50+201.47\times40=1.2\ \text{s}$$

In signal timing applications, the yellow signal is computed using a time-based equation and a standard deceleration rate. Also, for greater safety, the yellow signal uses an 85th percentile speed (speed below which 85%

of all vehicles travel), rather than an average speed. For the same reason, the all-red signal uses a 15% speed. In the equation for the yellow, speed is in the numerator, and faster vehicles would be at greater risk. In the equation for the all-red, speed is in the denominator, and slower vehicles would be at greater risk.

Sample Problem 3-10 shows how the concept of safe stopping distance is incorporated into signal timing methodologies, which are discussed in detail in Chapters 19 and 20.

3.4 Roadway Characteristics

Roadways are complex physical elements that have important impacts on traffic behavior. Roadways are, in fact, structures that bear the weight load of highway traffic. Further, roadways involve ancillary structures such as bridges, underpasses, embankments, drainage systems, and other features. This text does not treat the physical structural qualities of roadways.

Vehicle operations, however, are greatly influenced by the geometric characteristics of roadways, including horizontal and vertical curvature, and cross-sectional design elements (such as lanes, lane widths, shoulders). An overview of the specific geometric design elements of highways is presented in Chapter 27.

In this chapter, an overview of how roadway systems are organized, developed, and used by motorists is presented.

3.4.1 Highway Functions and Classification

Roadways are a major component of the traffic system, and the specifics of their design have a significant impact on traffic operations. There are two primary categories of service provided by roadways and roadway systems:

- Accessibility
- Mobility

"Accessibility" refers to the direct connection to abutting lands and land uses provided by roadways. This accessibility comes in the form of curb parking, driveway access to off-street parking, bus stops, taxi stands, loading zones, driveway access to loading areas, and similar features. The access function allows a driver or passenger (or goods) to depart the transport vehicle to enter the particular land use in question. "Mobility" refers to the through movement of people, goods, and vehicles from Point

A to Point B in the system.

The essential problem for traffic engineers is that the specific design aspects that provide for good access—parking, driveways, loading zones, and so on—tend to retard through movement, or mobility. Thus, the two major services provided by a roadway system are often in conflict. This leads to the need to develop roadway systems in a hierarchal manner, with various classes of roadways specifically designed to perform specific functions.

Trip Functions

The AASHTO defines up to six distinct travel movements that may be present in a typical trip:

- Main movement
- Transition
- Distribution
- Collection
- Access
- Termination

The *main movement* is the through portion of a trip, making the primary connection between the area of origin and the area of destination. *Transition* occurs when a vehicle transfers from the through portion of the trip to the remaining functions that lead to access and termination. A vehicle might, for example, use a ramp to transition from a freeway to a surface arterial. The *distribution* function involves providing drivers and vehicles with the ability to leave a major through facility and get to the general area of their destinations. *Collection* brings the driver and vehicle closer to the final destination, while *access* and *termination* result in providing the driver with a place to leave his or her vehicle and enter the land parcel sought. Not all trips will involve all of these components.

The hierarchy of trip functions should be matched by the design of the roadways provided to accomplish them. A typical trip has two terminals, one at the origin, and one at the destination. At the origin end, the access function provides an opportunity for a trip-maker to enter a vehicle and for the vehicle to enter the roadway system. The driver may go through a series of facilities, usually progressively accommodating higher speeds and through movements, until a facility—or set of facilities—is found that will provide the primary through connection. At the destination end of the trip, the reverse occurs, with the driver progressively moving toward facilities accommodating access until the specific land parcel desired is reached.

Highway Classification

All highway systems involve a hierarchal classification by the mix of access and mobility functions provided. There are four major classes of highways that may be identified:

- Limited-access facilities
- Arterials
- Collectors
- Local streets

The *limited-access facility* provides for 100% through movement, or mobility. No direct access to abutting land uses is permitted. *Arterials* are surface facilities that are designed primarily for through movement but permit some access to abutting lands. *Local streets* are designed to provide access to abutting land uses with through movement only a minor function, if provided at all. The *collector* is an intermediate category between arterials and local streets. Some measure of both mobility and access is provided. The term “collector” comes from a common use of such facilities to collect vehicles from a number of local streets and deliver them to the nearest arterial or limited-access facility. Figure 3.7 illustrates the traditional hierarchy of these categories.

Figure 3.7: Hierarchy of

Roadway Classifications

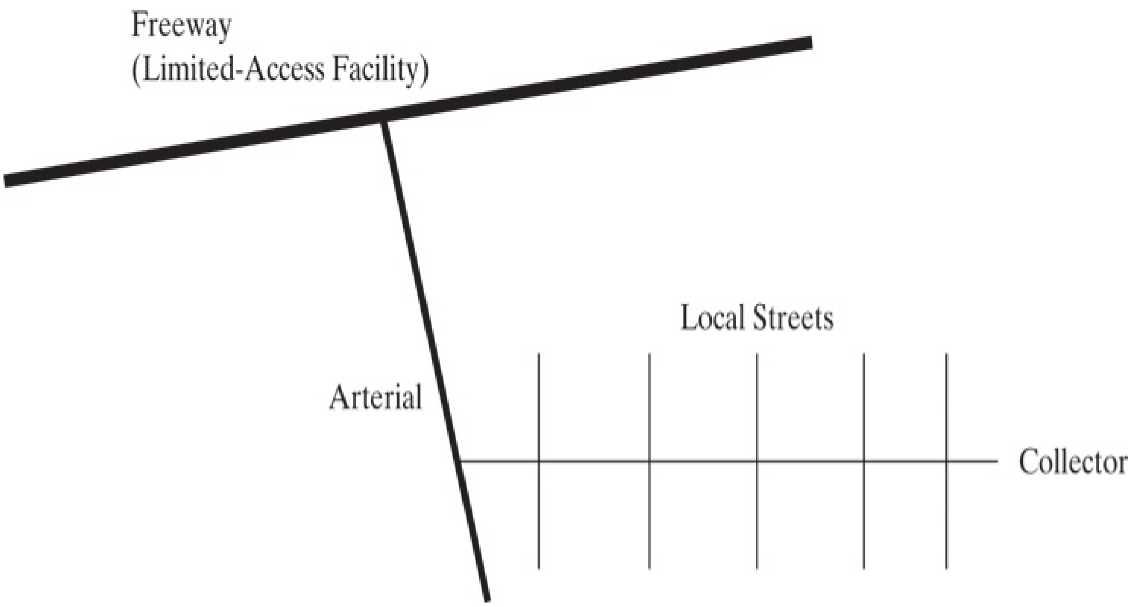


Figure 3.7: Full Alternative Text

The typical trip starts on a local street. The driver seeks the closest collector available, using it to access the nearest arterial. If the trip is long enough, a freeway or limited-access facility is sought. At the destination end of the trip, the process is repeated in reverse order. Depending upon the length of the trip and specific characteristics of the area, not all component types of facilities need be included in every trip.

Table 3.11 shows the range of through (or mobility) service provided by the major categories of roadway facility. Many states have their own classification systems that often involve subcategories. Table 3.12 provides a general description of frequently used subcategories in highway classification.

Table 3.11: Through Service Provided by Various Roadway Categories

Roadway Class	Percent through Service
Freeways (Limited-Access Facilities)	100
Arterials	60–80
Collectors	40–60
Local Streets	0–40

Table 3.11: Full Alternative Text

Table 3.12: Typical Rural and Urban Roadway Classification Systems

Sub-category	Rural	Urban
	Freeways	
Interstate freeways	All freeways bearing interstate designation.	All freeways bearing interstate designation.
Other freeways	All other facilities with full control of access.	All other facilities with full control of access.
Expressways	Facilities with substantial control of access, but having some at-grade crossings or entrances.	Facilities with substantial control of access, but having some at-grade crossings or entrances.
	Arterials	
Major or principal arterials	Serving significant corridor movements, often between areas with populations over 25,000 to 50,000. High-type design and alignment prevail.	Principal service for through movements, with very limited land-access functions that are incidental to the mobility function. High-type design prevails.
Minor arterials	Provide linkage to significant traffic generators, including towns and cities with populations below the range for principal arterials; serve shorter trip lengths than principal arterials.	Principal service for through movements, with moderate levels of access service also present.
	Collectors	
Major collectors	Serve generators of intra-county importance not served by arterials; provide connections to arterials and/or freeways.	No subcategories usually used for urban collectors:
Minor collectors	Link locally important generators with their rural hinterlands; provide connections to major collectors or arterials.	Provide land access and circulation service within residential neighborhoods and/or commercial/industrial areas; collect trips from local generators and channel them to nearby arterials; distribute trips from arterials to their ultimate destination.
	Local roads	
Residential	No subcategories generally used in rural classification schemes:	Provide land access and circulation within residential neighborhoods.
Commercial	Provide access to adjacent lands of all types; serve travel over relatively short distances.	Provide land access and circulation in areas of commercial development.
Industrial		Provide land access and circulation in areas of industrial development.

Table 3.12: Full Alternative Text

It is emphasized that the descriptions in Table 3.12 are presented as typical. Each highway agency will have its own highway classification system, and many have features that are unique to the agency. The traffic engineer should be familiar with highway classification systems, and be able to properly interpret any well-designed system.

3.4.2 Preserving the Function of a Facility

Highway classification systems enable traffic engineers to stratify the highway system by functional purpose. It is important that the intended function of a facility be reinforced through design and traffic controls.

Figure 3.8, for example, illustrates how the design and layout of streets within a suburban residential subdivision can reinforce the intended purpose of each facility.

Figure 3.8: Suburban Residential Subdivision Illustrated

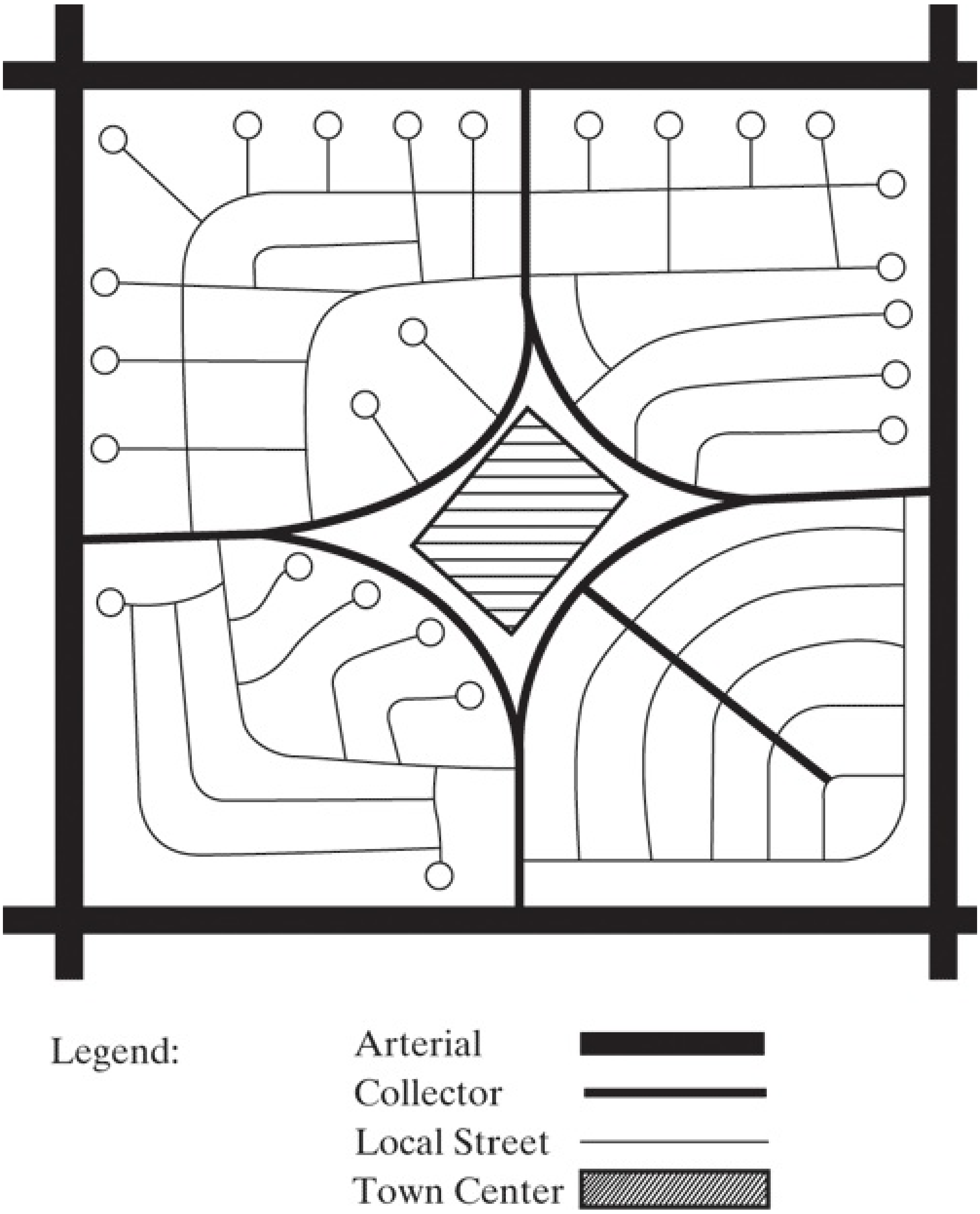


Figure 3.8: Full Alternative Text

The character of local streets is assured by incorporating sharp curvature into their design, and through the use of cul-de-sacs. No local street has direct access to an arterial; collectors within the subdivision provide the only access to arterials. The nature of collectors can be strengthened by not having any residence front on the collector.

The arterials have their function strengthened by limiting the number of points at which vehicles can enter or leave the arterial. Other aspects of an arterial, not obvious here, that could also help reinforce their function include the following:

- Parking prohibitions.
- Coordinated signals providing for continuous progressive movement at appropriate speeds.
- Median dividers to limit midblock left turns, and
- Speed limits appropriate to the facility and its environment.

In many older cities, it is difficult to separate the functions served by various facilities due to basic design and control problems. The historic development of many older urban areas has led to open-grid street systems. In such systems, local streets, collectors, and surface arterials all form part of the grid. Every street is permitted to intersect every other street, and all facilities provide some land access. Figure 3.9 illustrates this case. The only thing that distinguishes an arterial in such a system is its width and provision of progressive signal timing to encourage through movement.

Figure 3.9: An Open Grid System Illustrated

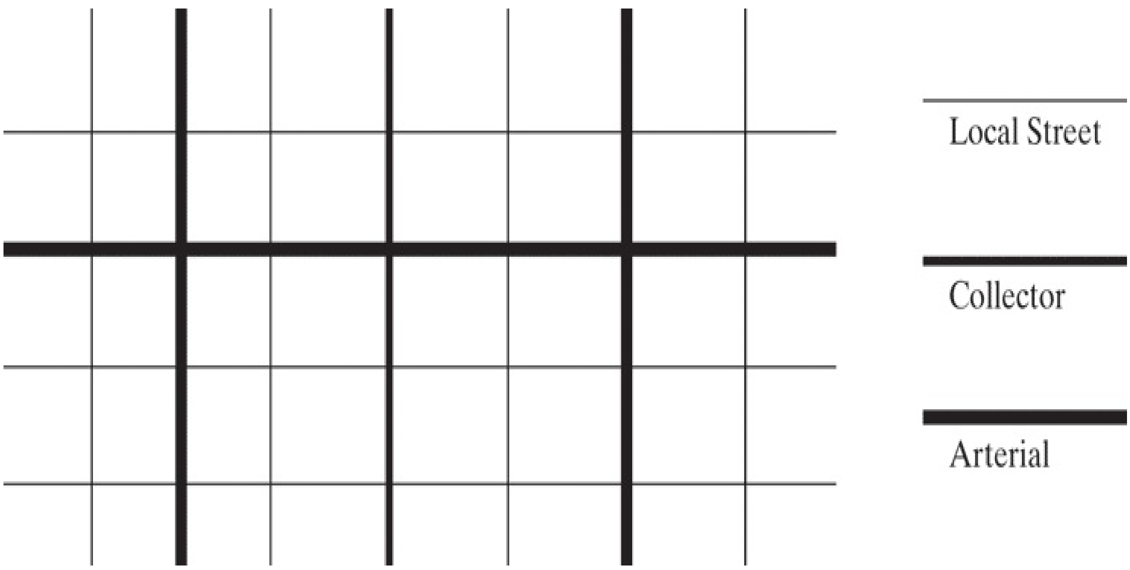


Figure 3.9: Full Alternative Text

Such systems often experience difficulties when development intensifies, and all classes of facility, including arterials, are subjected to heavy pedestrian movements, loading and unloading of commercial vehicles, parking, and similar functions. Because local streets run parallel to collectors and arterials, drivers experiencing congestion on arterials often reroute themselves to nearby local streets, subjecting residents to unwanted and often dangerous heavy through flows.

The importance of providing designs and controls that are appropriate to the intended function of a facility cannot be understated. Given that access and mobility operations do not mix well, their combination often creates safety problems and breeds congestion. It is, of course, impossible to completely separate the major functions, particularly in older cities where street systems were developed long before the era of the automobile. As far as is possible, however, traffic engineers need to be keenly aware of the conflicts that exist, and systems need to be developed to optimize the safety and operations of all elements using roadways: from huge trucks carrying goods, to cars, to pedestrians, to bicyclists, and others.

3.5 Traffic Control Systems and Characteristics

The fourth major component of the traffic system is the myriad control devices that are used to direct operations in a hopefully safe and efficient manner. Control devices are the traffic engineer's means of communicating with the driver (and other road users). There are no failsafe mechanisms on our highways: placing a STOP sign does not guarantee that all drivers will observe and obey it. A speed limit does not physically constrain vehicles from operating at lower speeds. While modern technology is rapidly developing, incorporating collision avoidance systems in many vehicles, the driver is still fundamentally in control. The driver is controlled only through provision of information and regulations, and the enforcement of those regulations.

Chapter 4 presents a significant overview of traffic control devices and their use, and how they affect the flow of traffic on highway systems.

3.6 Closing Comments

This chapter has summarized some of the key elements of driver, pedestrian, vehicle, control, and roadway characteristics that influence highway design and traffic control. These elements combine to create traffic streams. The characteristics of traffic streams are the result of interactions among and between these elements. The limitations on highway systems are directly related to the physical and other limitations of human road users, their vehicles, the roadways they travel on, and the controls they encounter.

References

- 1. Dewar, R., “Road Users,” *Traffic Engineering Handbook*, 5th Edition, Institute of Transportation Engineers, Washington, D.C., 1999.
- 2. Ogden, K.W., *Safer Roads: A Guide to Road Safety Engineering*, University Press, Cambridge, England, 1996.
- 3. Allen, M., et al., *Forensic Aspects of Vision and Highway Safety*, Lawyers and Judges Publishing Co., Inc.,Tucson, AZ, 1996.
- 4. Olson, P., *Forensic Aspects of Driver Perception and Response*, Lawyers and Judges Publishing Co., Inc., Tucson, AZ, 1996.
- 5. *A Policy on Geometric Design of Highways and Streets*, 6th Edition, American Association of State Highway and Transportation Officials, Washington, D.C., 2011.
- 6. Johansson, G. and Rumar, K., “Driver’s Brake Reaction Times,” *Human Factors*, Vol. 13, No. 1, Human Factors and Ergonomics Society, February 1971.
- 7. *Report of the Massachusetts Highway Accident Survey*, Massachusetts Institute of Technology, Cambridge, MA, 1935.
- 8. Normann, O.K., “Braking Distances of Vehicles from High Speeds,” *Proceedings of the Highway Research Board*, Vol. 22, Highway Research Board, Washington, D.C., 1953.
- 9. Fambro, D.B., et al., “Determination of Safe Stopping Distances,” *NCHRP Report 400*, Transportation Research Board, Washington, D.C., 1997.
- 10. *Determination of Vehicle Signal Change and Clearance Intervals*, Publication IR-073, Institute of Transportation Engineers, Washington, D.C., 1994.
- 11. *Human Factors*, Vol. 28, No. 1, Human Factors and Ergonomics

Society, 1986.

- 12. Eubanks, J.J. and Hill, P.L., *Pedestrian Accident Reconstruction and Litigation*, 2nd Edition, Lawyers and Judges Publishing Co., Inc., Tucson, AZ, 1998.
- 13. Perry, J., *Gait Analysis*, McGraw-Hill, New York, NY, 1992.
- 14. *Manual on Uniform Traffic Control Devices*, Federal Highway Administration, U.S. Department of Transportation, Washington, D.C., 2009, revised through 2012.
- 15. Sleight, R.B., “The Pedestrian,” *Human Factors in Traffic Safety Research*, John Wiley and Sons, Inc., New York, NY, 1972.
- 16. Tidwell, J.E. and Doyle, D., *Driver and Pedestrian Comprehension of Pedestrian Laws and Traffic Control Devices,* AAA Foundation for Traffic Safety, Washington, D.C., 1993.
- 17. Herms, B.F., “Pedestrian Crosswalk Study: Accidents in Painted and Unpainted Crosswalks,” *Pedestrian Protection*, Highway Research Record 406, Transportation Research Board, Washington, D.C., 1972.
- 18. *Traffic Safety Facts: Alcohol-Impaired Driving*, National Highway Traffic Safety Administration, U.S. Department of Transportation, Washington, D.C., December 2015.
- 19. *Traffic Safety Facts: Pedestrians*, National Highway Traffic Safety Administration, U.S. Department of Transportation, Washington, D.C., April 2014.
- 20. *Traffic Safety Facts: Drug Involvement of Fatally Injured Drivers*, National Highway Traffic Safety Administration, U.S. Department of Transportation, Washington, D.C., November 2010.
- 21. “Transportation in an Aging Society,” *Special Report 218*, Transportation Research Board, Washington, D.C., 1988.
- 22. *Highway Statistics*, Federal Highway Administration, U.S. Department of Transportation, Washington, D.C., 2015.

Problems

1. 3-1. A driver takes 3.5 s to react to a complex situation. How far does the vehicle travel before the driver initiates a physical response to the situation (i.e., putting his or her foot on the brake)? Plot the results for speeds ranging from 30 to 70 mi/h (in 5-mi/h increments).

2. 3-2. A driver traveling at 65 mi/h rounds a curve on a level grade to see a truck overturned across the roadway at a distance of 350 ft. If the driver is able to decelerate at a rate of 10 ft/s^2, at what speed will the vehicle hit the truck? Plot the result for reaction times ranging from 0.50 to 5.00 s in increments of 0.5 s. Comment on the results.

3. 3-3. A car hits a tree at an estimated speed of 25 mi/h on a 3% upgrade. If skid marks of 150 ft are observed on dry pavement (F = 0.348) followed by 200 ft (F = 0.250) on a grass-stabilized shoulder, estimate the initial speed of the vehicle just before the pavement skid was begun.

4. 3-4. Drivers must slow down from 60 to 40 mi/h to negotiate a severe curve on a rural highway. A warning sign for the curve is clearly visible for a distance of 200 ft. How far in advance of the curve must the sign be located in order to ensure that vehicles have sufficient distance to safely decelerate? Use the standard reaction time and deceleration rate recommended by AASHTO for basic braking maneuvers.

5. 3-5. How long should the “yellow” signal be for vehicles approaching a traffic signal on a 2% downgrade at a speed of 40 mi/h? Use a standard reaction time of 1.0 s and the standard AASHTO deceleration rate.

6. 3-6. What is the safe stopping distance for a section of rural freeway with a design speed of 80 mi/h on a 2% downgrade?

7. 3-7. What minimum radius of curvature may be designed for safe operation of vehicles at 70 mi/h if the maximum rate of superelevation (*e*) is 0.06 and the maximum coefficient of side

friction (*f*) is 0.10?

Chapter 4 Communicating with Drivers: Traffic Control Devices

Traffic control devices are the media through which traffic engineers communicate with drivers and other road users. Virtually every traffic law, regulation, or operating instruction must be communicated through the use of devices that fall into three broad categories:

- Traffic markings
- Traffic signs
- Traffic signals

The effective communication between traffic engineer and driver is a critical link if safe and efficient traffic operations are to prevail. Traffic engineers have no direct control over any individual driver or group of drivers. If a motorman violates a RED signal while conducting a subway train, an automated braking system would force the train to stop anyway. If a driver violates a RED signal, only the hazards of conflicting vehicular and/or pedestrian flows would impede the maneuver. Thus, it is imperative that traffic engineers design traffic control devices that communicate uncomplicated messages clearly, in a way that encourages proper observance.

This chapter introduces some of the basic principles involved in the design and placement of traffic control devices. Subsequent chapters cover the details of specific applications to freeways, multilane and two-lane highways, intersections, and arterials and streets.

4.1 The Manual on Uniform Traffic Control Devices

The principal standard governing the application, design, and placement of traffic control devices is the current edition of the *Manual on Uniform Traffic Control Devices* (MUTCD) [1]. The Federal Highway Administration publishes a national MUTCD, which serves as a minimum standard and a model for individual state MUTCDs. Many states simply adopt the federal manual by statute. Others develop their own manuals. In the latter case, the state MUTCD must meet all of the minimum standards of the federal manual. Current law requires that all states update their MUTCDs to be compliant with the current federal MUTCD within two years of the issuance of a federal update. Noncompliant devices (that were compliant with previous manuals) may be replaced as part of a regular device replacement program. They must, however, be replaced with compliant devices. As is the case with most federal mandates in transportation, compliance is enforced through partial withholding of federal-aid highway funds from states deemed in violation of federal MUTCD standards.

4.1.1 History and Background

One of the principal objectives of the MUTCD is to establish *uniformity* in the use, placement, and design of traffic control devices. Communication is greatly enhanced when the same messages are delivered in the same way and in similar circumstances at all times. Consider the potential confusion if each state designed its own STOP sign, with different shapes, colors, and legends.

Varying device design is not a purely theoretical issue. As late as the early 1950s, two-color (red, green) traffic signals had the indications in different positions in different states. Some placed the “red” ball on top; others placed the “green” ball on top. This is a particular problem for drivers with color blindness, the most common form of which is the inability to distinguish “red” from “green.” Standardizing the order of signal lenses was a critical safety measure, guaranteeing that even color-blind drivers could interpret the signal by position of the light in the display. More recently, small amounts of blue and yellow pigments have been added to “green” and “red” lenses to enhance their visibility to color-blind drivers.

Early traffic control devices were developed in various locales with little or no coordination on their design, much less their use. The first centerline appeared on a Michigan roadway in 1911. The first electric signal installation is thought to have occurred in Cleveland, Ohio, in 1914. The first STOP sign was installed in Detroit in 1915, where the first three-color traffic signal was installed in 1920.

The first attempts to create national standards for traffic control devices occurred during the 1920s. Two separate organizations developed two manuals in this period. In 1927, the American Association of State Highway Officials (the forerunner of the American Association of State Highway and Transportation Officials, AASHTO) published the *Manual and Specification for the Manufacture, Display, and Erection of U.S. Standard Road Markings and Signs.* It was revised in 1929 and 1931. This manual addressed only rural signing and marking applications. In 1930, the National Conference on Street and Highway Safety published the *Manual on Street Traffic Signs, Signals, and Markings*, which addressed urban applications.

In 1932, the two groups formed a merged Joint Committee on Uniform Traffic Control Devices and published the first complete MUTCD in 1935, revising it in 1939. This group continued to have responsibility for subsequent editions until 1972, after which the Federal Highway Administration formally assumed responsibility for the manual.

The latest official edition of the MUTCD (as of January 2018) was published in 2009, and has been formally updated through May 2012.

For an excellent history of the MUTCD and its development, consult a series of articles by Hawkins [2–5].

4.1.2 General Principles of the MUTCD

The MUTCD states that the purpose of traffic control devices is *to promote highway safety and efficiency by providing for orderly movement of all road users on streets, highways, bikeways, and private roads open to public travel throughout the Nation* [Ref *1*, pg 1].

It also defines five requirements for a traffic control device to be effective in fulfilling that mission. A traffic control device must:

1. fulfill a need;
2. command attention;
3. convey a clear, simple meaning;
4. command respect from road users; and
5. give adequate time for a proper response.

In addition to the obvious meanings of these requirements, some subtleties should be carefully noted. The first strongly implies that superfluous devices *should not* be used. Each device must have a specific purpose and must be needed for the safe and efficient flow of traffic. The fourth requirement reinforces this. Respect of drivers is commanded only when drivers are conditioned to expect that all devices carry meaningful and important messages. Overuse or misuse of devices encourages drivers to ignore them—it is like "crying wolf" too often. In such an atmosphere, drivers may not pay attention to those devices that are really needed.

Requirements 2 and 3 affect the design of a device. Commanding attention requires proper visibility and a distinctive design that attracts the driver's attention in what is often an environment filled with visual distractions. Standard use of color and shape coding plays a major role in attracting this attention. Clarity and simplicity of message are critical; the driver is viewing the device for only a few short seconds while traveling at what may be a high speed. Again, color and shape coding are used to deliver as much information as possible. Legend, the hardest element of a device to understand, must be kept short and as simple as possible.

Requirement 5 affects the placement of devices. A STOP sign, for example, is always placed at the stop line, but must be visible for at least one safe stopping distance. Guide signs requiring drivers to make lane changes must be placed well in advance of the diverge area to give drivers sufficient distance to execute the required maneuvers.

4.1.3 Contents of the MUTCD

The MUTCD addresses three critical aspects of traffic control devices. It contains the following:

1. Detailed standards for the physical design of the device, specifying shape, size, colors, legend types and sizes, and specific legend.

2. Detailed standards and guidelines on where devices should be located with respect to the traveled way.

3. Warrants, or conditions, that justify the use of a particular device.

The most detailed and definitive standards are for the physical design of the device. Little is left to judgment, and virtually every detail of the design is fully specified. Colors are specified by specific pigments and legend by specific fonts. Some variance is permitted with respect to size, with minimum sizes specified, providing optional larger sizes for use when needed for additional visibility.

Placement guidelines are also relatively definitive but often allow for some variation within prescribed limits. Placement guidelines sometimes lead to obvious problems. One frequent problem involves STOP signs. When placed in the prescribed position, they may wind up behind trees or other obstructions where their effectiveness is severely compromised. Figure 4.1 shows such a case, in which a STOP sign placed at the prescribed height and lateral offset at the STOP line winds up virtually hidden by a tree. Common sense must be exercised in such cases if the device is to be effective.

Figure 4.1: STOP-Sign Partially Hidden by Tree

(Photo courtesy of J. Ulerio and R. Roess)

Warrants are given with various levels of specificity and clarity. Signal warrants, for example, are detailed and relatively precise. This is necessary because signal installations represent a significant investment, both in initial investment and in continuing operating and maintenance costs. The warrants for STOP and YIELD signs, however, are far more general and leave substantial latitude for the exercise of professional judgment.

Chapter 15 deals with the selection of an appropriate form of intersection control and covers the warrants for signalization, two-way and multiway STOP signs, and YIELD signs in some detail. Because of the cost of signals, much study has been devoted to the defining of conditions warranting their use. Proper implementation of signal and other warrants in the MUTCD requires appropriate engineering studies to be made to determine the need for a particular device or devices.

4.1.4 Legal Aspects of the MUTCD

The MUTCD provides guidance and information in four different categories:

1. Standard. A standard is a statement of a required, mandatory, or specifically prohibited practice regarding a traffic control device. Typically, standards are indicated by the use of the term “shall” or “shall not” in the statement.

2. Guidance. Guidance is a statement of recommended, but not mandatory, practice in typical situations. Deviations are allowed if engineering judgment or a study indicates that a deviation is appropriate. Guidance is generally indicated by use of the term “should” or “should not.”

3. Option. An option is a statement of practice that is a permissive condition. It carries no implication of requirement or recommendation. Options often contain allowable modifications to a standard or a guidance. An option is usually stated using the term “may” or “may not.”

4. Support. This is a purely information statement provided to supply additional information to the traffic engineer. The words “shall,” “should,” or “may” do not appear in these statements (nor do their negative counterparts).

The four types of statements given in the MUTCD have legal implications for traffic agencies. Violating a standard leaves the jurisdictional agency exposed to liability for any accident that occurs because of the violation. Thus, placing a nonstandard STOP sign would leave the jurisdictional agency exposed to liability for any accident occurring at the location. Guidelines, when violated, also leave some exposure to liability. Guidelines should be modified only after an engineering study has been conducted and documented, justifying the modification(s). Without such documentation, liability for accidents may also exist. Options and support carry no implications with respect to liability.

It should also be understood that jurisdiction over traffic facilities is established as part of each state’s vehicle and traffic law. That law generally indicates what facilities fall under the direct jurisdiction of the state (usually designated state highways and all intersections involving such highways) and specifies the state agency exercising that jurisdiction. It also defines what roadways would fall under the control of county, town, and other local governments. Each of those political entities, in turn, would appoint or otherwise specify the local agency exercising jurisdiction.

Many traffic control devices must be supported by a specific law or ordinance enacted by the appropriate level of government. Procedures for implementing such laws and ordinances must also be specified. Many times (such as in the case of speed limits and parking regulations), public hearings and/or public notice must be given before imposition. For example, it would not be legal for an agency to post parking prohibitions during the night and then ticket or tow all parked vehicles without having provided adequate advance public notice, which is most often accomplished using local or regional newspapers.

This chapter presents some of the principles of the MUTCD, and generally describes the types of devices and their typical applications. Chapter 31 goes into greater detail concerning the use of traffic control devices on freeways, multilane, and two-lane highways. Chapter 17 contains additional detail concerning use of traffic control devices at intersections.

4.1.5 Communicating with the Driver

The driver is accustomed to receiving a certain message in a clear and standard fashion, often with redundancy. A number of mechanisms are used to convey messages. These mechanisms make use of recognized human limitations, particularly with respect to eyesight. Messages are conveyed through the use of the following:

- Color. Color is the most easily visible characteristic of a device. Color is recognizable long before a general shape may be perceived and considerably before a specific legend can be read and understood. The principal colors used in traffic control devices are red, yellow, green, orange, black, blue, and brown. These are used to code certain types of devices and to reinforce specific messages whenever possible. Recently, purple has been added as a special color for toll plaza markings designating an electronic toll collection lane.

- Shape. After color, the shape of the device is the next element to be discerned by the driver. Particularly in signing, shape is an important element of the message, either identifying a particular type of information that the sign is conveying or conveying a unique message of its own.

- Pattern. Pattern is used in the application of traffic markings. In general, double-solid, solid, dashed, and broken lines are used. Each conveys a type of meaning with which drivers become familiar. The frequent and consistent use of similar patterns in similar applications contributes greatly to their effectiveness and to the instant recognition of their meaning.

- Legend. The last element of a device that the driver comprehends is its specific legend. Signals and markings, for example, convey their entire message through use of color, shape, and pattern. Signs, however, often use specific legend to transmit the details of the message being transmitted. Legend must be kept simple and short, so that drivers do not divert their attention from the driving task, yet are able to see and understand the specific message being given.

Redundancy of message can be achieved in a number of ways. The STOP sign, for example, has a unique shape (octagon), a unique color (red), and a unique one-word legend (STOP). Any of the three elements alone is sufficient to convey the message. Each provides redundancy for the others.

Redundancy can also be provided through use of different devices, each reinforcing the same message. A left-turn lane may be identified by arrow markings on the pavement, a “This Lane Must Turn Left” sign, and a protected left-turn signal phase indicated by a green arrow. Used together, the message is unmistakable.

The MUTCD provides a set of standards, guidelines, and general advice on how to best communicate various traffic rules and regulations to drivers. The MUTCD, however, is a document that is always developing. The traffic engineer must always consult the latest version of the manual (with all applicable revisions) when considering traffic control options. The most current version (at any given time) of the MUTCD is available online through the Federal Highway Administration web site.

4.2 Traffic Markings

Traffic markings are the most plentiful traffic devices in use. They serve a variety of purposes and functions and fall into three broad categories:

- Longitudinal markings
- Transverse markings
- Object markers and delineators

Longitudinal and transverse markings are applied to the roadway surface using a variety of materials, the most common of which are paint and thermoplastic. Reflectorization for better night vision is achieved by mixing tiny glass beads in the paint or by applying a thin layer of glass beads over the wet pavement marking as it is placed. The latter provides high initial reflectorization, but the top layer of glass beads is more quickly worn. When glass beads are mixed into the paint before application, some level of reflectorization is preserved as the marking wears. Thermoplastic is a naturally reflective material, and nothing need be added to enhance drivers' ability to see them at night.

In areas where snow and snowplowing is not a problem, paint or thermoplastic markings can be augmented by pavement inserts with reflectors. Such inserts greatly improve the visibility of the markings at night. They are visible in wet weather (often a problem with markings) and resistant to wear. They are generally not used where plowing is common, as they can be dislodged or damaged during the process.

Object markers and delineators are small object-mounted reflectors. Delineators are small reflectors mounted on lightweight posts and are used as roadside markers to help drivers in proper positioning during inclement weather, when standard markings are not visible.

4.2.1 Colors and Patterns

Five marking colors are in current use: yellow, white, red, blue, and purple. In general, they are used as follows:

- *Yellow* markings separate traffic traveling in opposite directions.
- *White* markings separate traffic traveling in the same direction, and are used for all transverse markings.
- *Red* markings delineate roadways that shall not be entered or used by the viewer of the marking.
- *Blue* markings are used to delineate parking spaces reserved for persons with disabilities.
- *Purple* markings are used to delineate electronic toll lanes in toll plazas.

Black is used in conjunction with other markings on light pavements. Black markings may be used to “fill in” gaps in markings of other colors to enhance their visibility. To emphasize the pattern of the line, gaps between yellow or white markings are filled in with black to provide contrast and easier visibility.

A solid line prohibits or discourages crossing. A double-solid line indicates maximum or special restrictions. A broken line indicates that crossing is permissible. A dotted line uses shorter line segments than a broken line. It provides trajectory guidance and often is used as a continuation of another type of line in a conflict area. Normally, line markings are 4 to 6 inches wide. Wide lines, which provide greater emphasis, should be at least twice the width of a normal line. Broken lines normally consist of 10 ft line segments and 30 ft gaps. Similar dimensions with a similar ratio of line segments to gaps may be used as appropriate for prevailing traffic speeds and the need for delineation. Dotted lines usually consist of 2 ft line segments and 4 ft (or longer) gaps. MUTCD suggests a maximum segment-to-gap ratio of 1:3 for dotted lines.

4.2.2 Longitudinal Markings

Longitudinal markings are those markings placed parallel to the direction of travel. A vast majority of longitudinal markings involve centerlines, lane lines, and pavement edge lines.

Longitudinal markings provide guidance for the placement of vehicles on the traveled way cross section and basic trajectory guidance for vehicles traveling along the facility. The best example of the importance of longitudinal markings is the difficulty in traversing a newly paved highway segment on which lane markings have not yet been repainted. Drivers do not automatically form neat lanes without the guidance of longitudinal markings; rather, they tend to place themselves somewhat randomly on the cross section, encountering many difficulties. Longitudinal markings provide for organized flow and optimal use of the pavement width.

Centerlines

The yellow centerline marking is critically important and is used to separate traffic traveling in opposite directions. Use of centerlines on all types of facilities is not mandated by the MUTCD. The applicable standard is as follows:

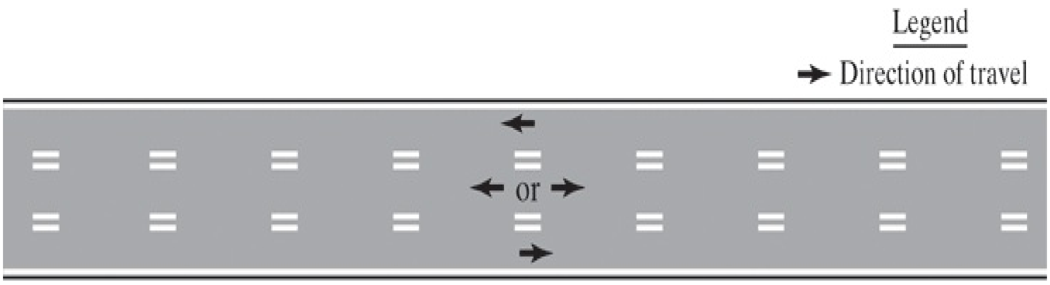
> Centerline markings shall be placed on all paved urban arterials and collectors that have a travelled way of 20 ft or more and an ADT of 6,000 veh/day or greater. Centerline markings shall also be placed on all paved, two-way streets or highways that have 3 or more traffic lanes. [Ref *1*, pg 349]

Further guidance indicates that placing centerlines is recommended for urban arterials and streets with an average daily traffic (ADT) of 4,000 or more and a roadway width of 20 ft or more, and on rural highways with a width in excess of 18 ft and ADT >3,000 veh/day. Caution should be used in placing centerlines on pavements of 16 ft or less, which may increase the incidence of traffic encroaching on roadside elements outside the traveled way.

On two-lane, two-way rural highways, centerline markings supplemented by signs are used to regulate passing maneuvers. A double-solid yellow centerline marking indicates that passing is not permitted in either direction. A solid yellow line with a dashed yellow line indicates that passing is permitted from the dashed side only. Where passing is permissible in both directions, a single dashed yellow centerline is used.

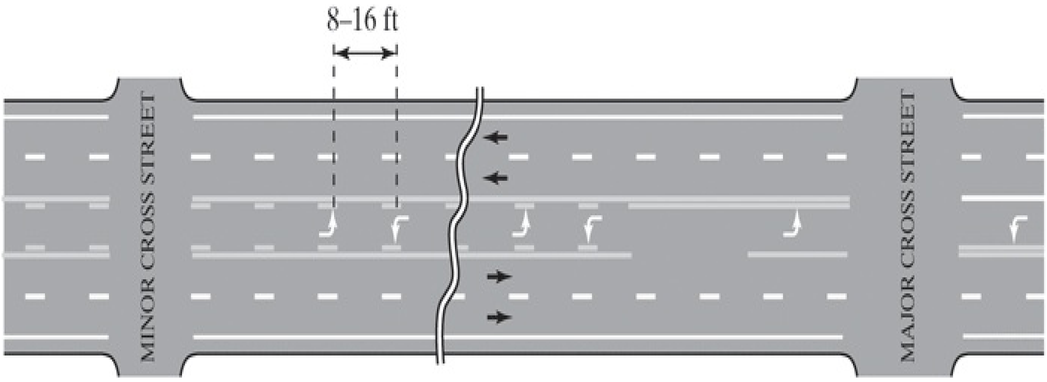
There are other specialized uses of yellow markings. Figure 4.2(a) illustrates the use of double-dashed yellow markings to delineate a reversible lane on an arterial. Signing and/or lane-use signals would have to supplement these to denote the directional use of the lane. Figure 4.2(b) shows the markings used for two-way left-turn lanes on an arterial.

Figure 4.2: Specialized Uses of Yellow Markings



(a) Reversible Lane Marking

4.2-1 Full Alternative Text



(b) Two-Way Left-Turn Lane Marking

4.2-1 Full Alternative Text

(Source: *Manual of Uniform Traffic Control Devices*, Federal Highway Administration, Washington, D.C., 2009, as revised through May 2012, Figs 3B-6 and 3B-7, pgs 356 and 357.)

Lane Markings

The typical lane marking is a single white dashed line separating lanes of traffic in the same direction. MUTCD standards require the use of lane markings on all freeways and Interstate highways and recommend their use on all highways with two or more adjacent traffic lanes in a single direction. The dashed lane line indicates that lane changing is permitted. A single solid white lane line is used to indicate that lane changing is discouraged but not illegal. Where lane changing is to be prohibited, a double-white solid lane line is used.

Edge Markings

Edge markings are a required standard on freeways, expressways, and rural highways with a traveled way of 20 ft or more in width and an ADT of 6,000 veh/day or greater. They are recommended for rural highways with ADTs >3,000 veh/day and a 20 ft or wider traveled way.

When used, right-edge markings are a single normal solid white line; left-edge markings are a single normal solid yellow line.

Other Longitudinal Markings

The MUTCD provides for many options in the use of longitudinal markings. Consult the manual directly for further detail. The manual also provides standards and guidance for other types of applications, including freeway and nonfreeway merge and diverge areas, lane drops, extended markings through intersections, and other situations.

Chapter 31 contains additional detail on the application of longitudinal markings on freeways, expressways, and rural highways. Chapter 17 includes additional discussion of intersection markings.

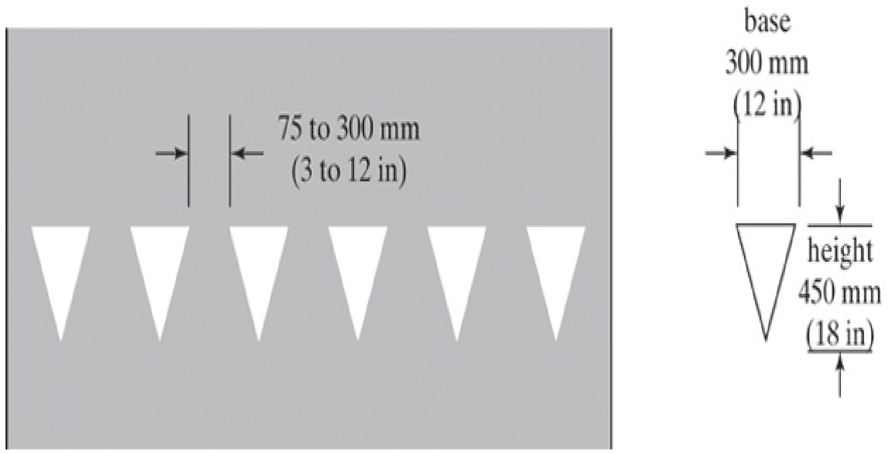
4.2.3 Transverse Markings

Transverse markings, as their name implies, include any and all markings with a component that cuts across a portion or all of the traveled way. When used, all transverse markings are white.

STOP and YIELD Lines

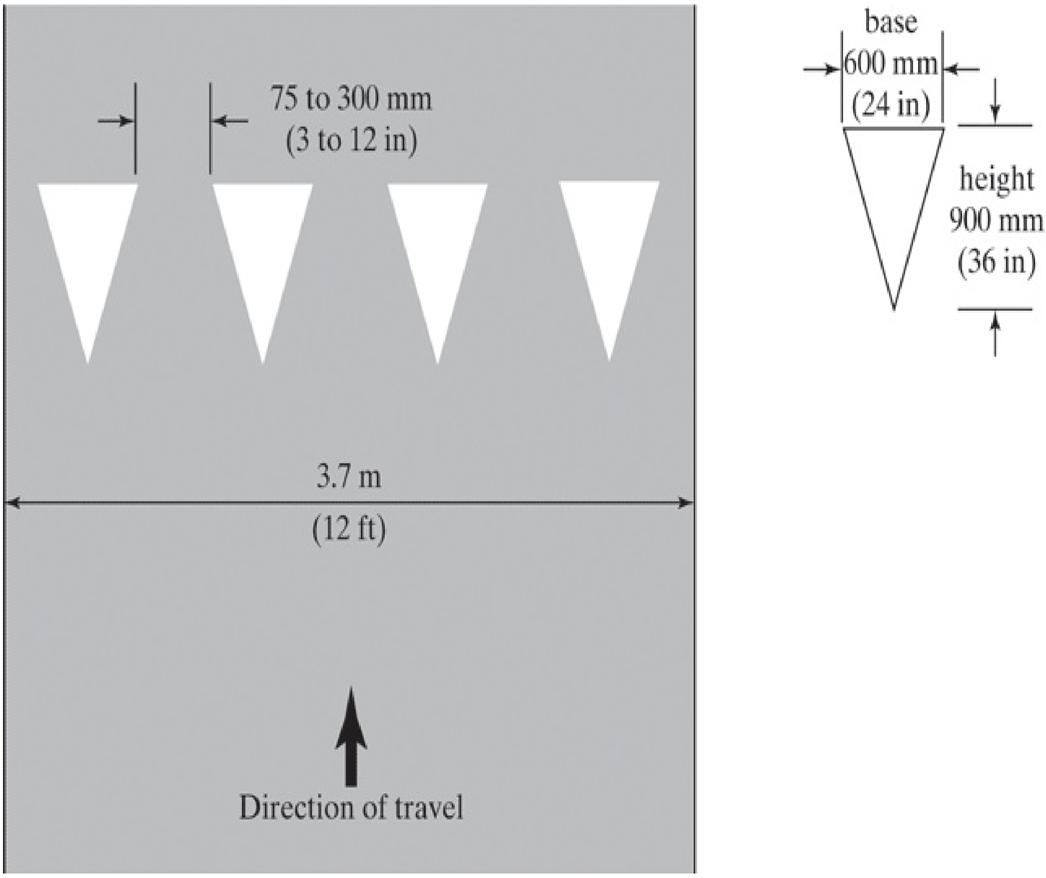
STOP lines are generally not mandated by the MUTCD. In practice, STOP lines are almost always used where marked crosswalks exist and in situations where the appropriate location to stop for a STOP sign or traffic signal is not clear. When used, it is recommended that the width of the line be 12 to 24 inches. When used, STOP lines must extend across all approach lanes. STOP lines, however, *shall not* be used in conjunction with YIELD signs; in such cases, the newly introduced YIELD line would be used. The YIELD line is illustrated in Figure 4.3.

Figure 4.3: The YIELD Line Illustrated



(a) Minimum Dimensions

4.2-1 Full Alternative Text



(b) Maximum Dimensions

4.2-1 Full Alternative Text

Notes: Triangle height is equal to 1.5 times the base dimension.

Yield lines may be smaller than suggested when installed on much narrower, slow-speed facilities such as shared-use paths.

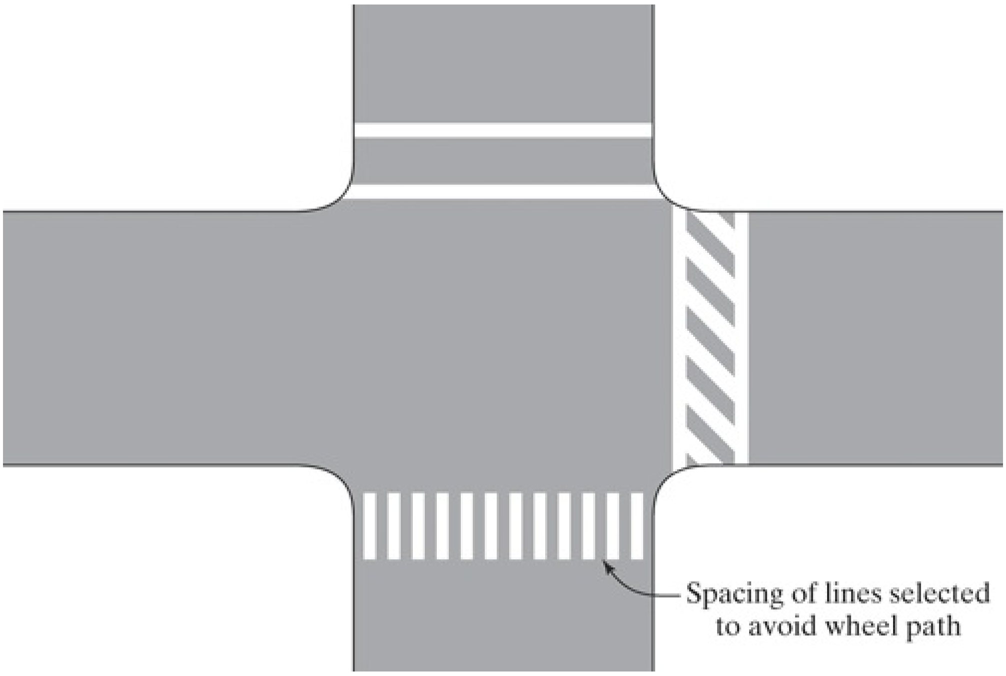
(Source: *Manual of Uniform Traffic Control Devices*, Federal Highway Administration, Washington, D.C., 2009, as revised through May 2012, Fig 3B-16, pg 382.)

Crosswalk Markings

Although not mandated by the MUTCD, it is recommended that crosswalks be marked at all intersections at which “substantial” conflict between vehicles and pedestrians exists. They should also be used at points of pedestrian concentration and at locations where pedestrians might not otherwise recognize the proper place and/or path to cross.

A marked crosswalk should be 6 ft or more in width. Figure 4.4 shows the three types of crosswalk markings in general use. The most frequently used is composed of two parallel white lines. These lines must be between 6 and 24 inches in width. Cross-hatching may be added to provide greater focus in areas with heavy pedestrian flows. The use of parallel transverse markings to identify the crosswalk is another option used at locations with heavy pedestrian flows.

Figure 4.4: Crosswalk Markings Illustrated



(Source: *Manual of Uniform Traffic Control Devices*, Federal Highway Administration, Washington, D.C., 2009, as revised through May 2012, Fig 3B-19, pg 384.)

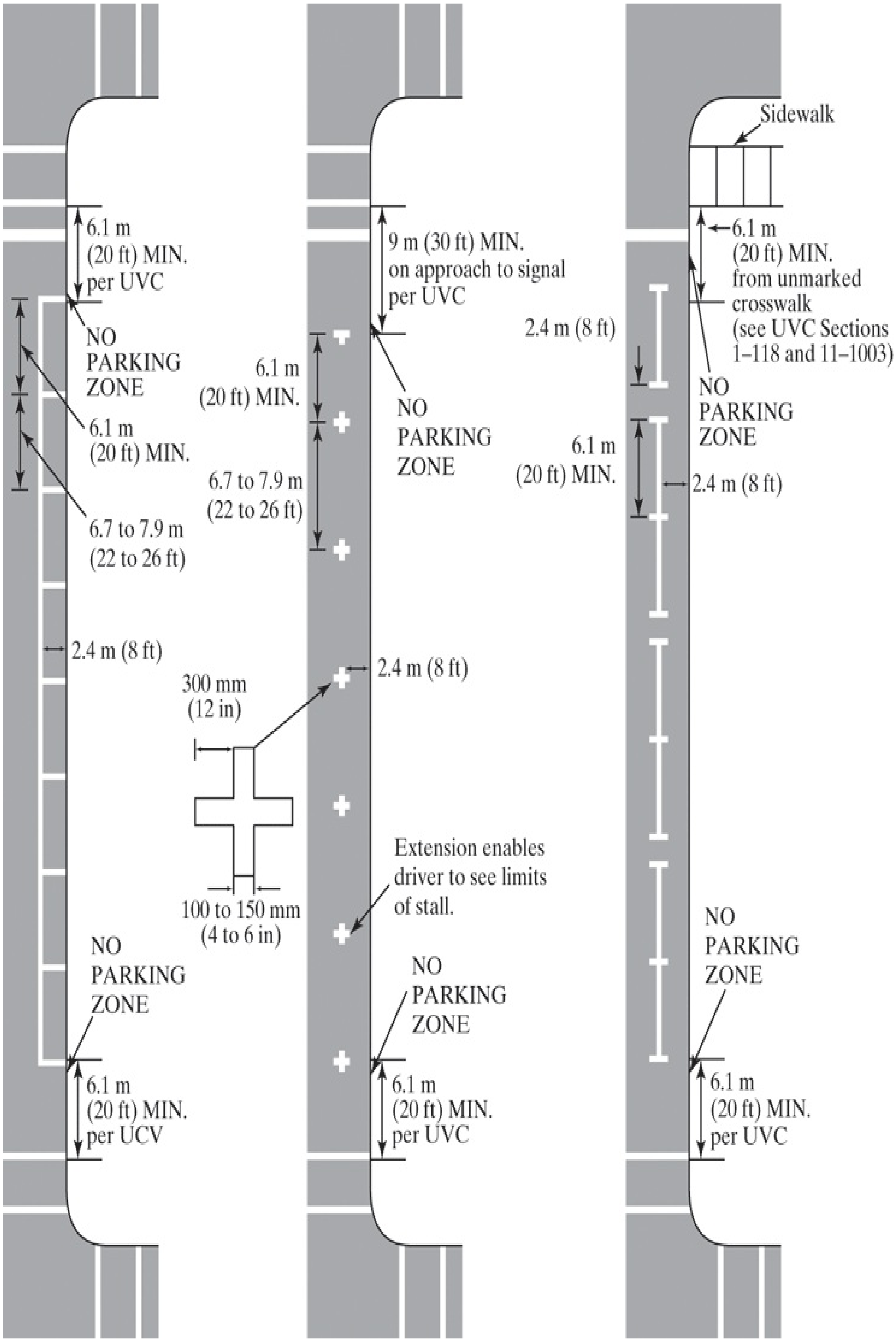
Figure 4.4: Full Alternative Text

The manual also contains a special pedestrian crosswalk marking for signalized intersections where a full pedestrian phase is included. Consult the manual directly for details of this particular marking.

Parking Space Markings

Parking space markings are not purely transverse, as they contain both longitudinal and transverse elements. They are officially categorized as transverse markings, however, in the MUTCD. They are always optional and are used to encourage efficient use of parking spaces. Such markings can also help prevent encroachment of parked vehicles into fire-hydrant zones, loading zones, taxi stands and bus stops, and other specific locations at which parking is prohibited. They are also useful on arterials with curb parking, as they also clearly demark the parking lane, separating it from travel lanes. Figure 4.5 illustrates typical parking lane markings.

Figure 4.5: Parking Space Markings



(Source: *Manual on Uniform Traffic Control Devices*, Federal Highway Administration, Washington, D.C., 2009, as revised through May 2012, Figure 3B-21, pg 386.)

Figure 4.5: Full Alternative Text

Note that the far end of the last marked parking space should be at least 20 ft away from the nearest crosswalk marking (30 ft on a signalized intersection approach).

Word and Symbol Markings

The MUTCD prescribes a number of word and symbol markings that may be used, often in conjunction with signs and/or signals. These include arrow markings indicating lane-use restrictions. Such arrows (with accompanying signs) are mandatory where a through lane becomes a left- or right-turn-only lane approaching an intersection.

Word markings include “ONLY,” used in conjunction with lane-use arrows, and “STOP,” which can be used only in conjunction with a STOP line and a STOP sign. “SCHOOL” markings are often used in conjunction with signs to demark school and school-crossing zones. The MUTCD contains a listing of all authorized word markings and allows for discretionary use of unique messages where needed. Figure 4.6 provides two examples.

Figure 4.6: Word and Symbol Marking in the Field

(a) School Marking

4.2-1 Full Alternative Text

(b) Lane-Use Arrows

4.2-1 Full Alternative Text

(Photos courtesy of J. Ulerio and R. Roess)

Other Transverse Markings

Consult the MUTCD directly for examples of other types of markings, including preferential lane markings, curb markings, roundabout and traffic circle markings, and speed-hump markings. Chapter 17 contains a detailed discussion of the use of transverse and other markings at intersections. Chapter 25 contains details on the marking of roundabouts, which include both transverse and longitudinal markings.

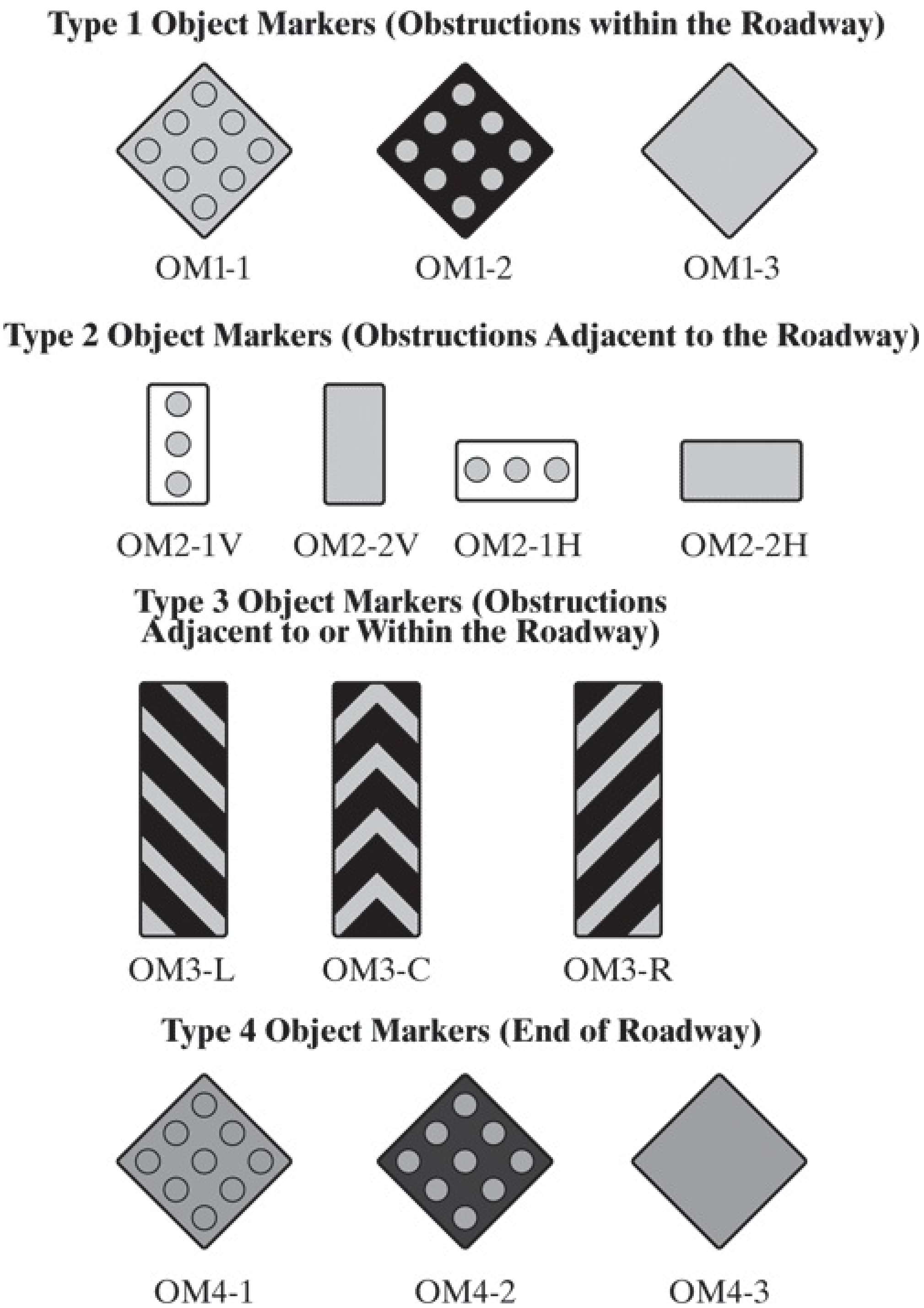
4.2.4 Object Markers

It should be noted that the MUTCD now treats object markers as signs rather than markings. Because their function is more in line with markings, however, they are discussed in this section.

Object markers are used to denote obstructions either in or adjacent to the traveled way. Object markers are mounted on the obstruction in accordance with MUTCD standards and guidelines. In general, the lower edge of the marker is mounted a minimum of 4 ft above the surface of the nearest traffic lane (for obstructions 8 ft or less from the pavement edge) or 4 ft above the ground (for obstructions located further away from the pavement edge).

There are three types of object markers used, as illustrated in Figure 4.7. Obstructions within the roadway *must* be marked using a Type 1 or Type 3 marker. The Type 3 marker, when used, must have the alternating yellow and black stripes sloped downward at a 45° angle toward the side on which traffic is to pass the obstruction. When used to mark a roadside obstruction, the inside edge of the marker must be in line with the inner edge of the obstruction.

Figure 4.7: Object Markers



(Source: *Manual of Uniform Traffic Control Devices*, Federal Highway Administration, Washington, D.C., 2009, as revised through May 2012, Fig 3C-13, pg 135.)

Figure 4.7: Full Alternative Text

4.2.5 Delineators

Delineators are reflective devices mounted at a 4 ft height on the side(s) of a roadway to help denote its alignment. They are particularly useful during inclement weather, where pavement edge markings may not be visible. When used on the right side of the roadway, delineators are white; when used on the left side of the roadway, delineators are yellow. The back of delineators may have red reflectors to indicate wrong-way travel on a one-direction roadway.

Delineators are mandated on the right side of freeways and expressways and on at least one side of interchange ramps, with the exception of tangent sections where raised pavement markers are used continuously on all lane lines, where whole routes (or substantial portions thereof) have large tangent sections, or where delineators are used to lead into all curves. They may also be omitted where there is continuous roadway lighting between interchanges. Delineators may be used on an optional basis on other classes of roads.

4.3 Traffic Signs

The MUTCD provides specifications and guidelines for the use of literally hundreds of different signs for myriad purposes. In general, traffic signs fall into one of three major categories:

- Regulatory signs. Regulatory signs convey information concerning specific traffic regulations. Regulations may relate to right-of-way, speed limits, lane usage, parking, or a variety of other functions.

- Warning signs. Warning signs are used to inform drivers about upcoming hazards that they might not see or otherwise discern in time to safely react.

- Guide signs. Guide signs provide information on routes, destinations, and services that drivers may be seeking.

It would be impossible to cover the full range of traffic signs and applications in a single chapter. The sections that follow provide a general overview of the various types of traffic signs and their use.

4.3.1 Regulatory Signs

> Regulatory signs shall be used to inform road users of selected traffic laws or regulations and indicate the applicability of the legal requirements.
>
> Regulatory signs shall be installed at or near where the regulations apply. The signs shall clearly indicate the requirements imposed by the regulations and shall be designed and installed to provide adequate visibility and legibility in order to obtain compliance.
> [Ref *1*, pg 45]

Drivers are expected to be aware of many general traffic regulations, such as the basic right-of-way rule at intersections and the state speed limit. Signs, however, should be used in all cases where the driver cannot be expected to know the applicable regulation.

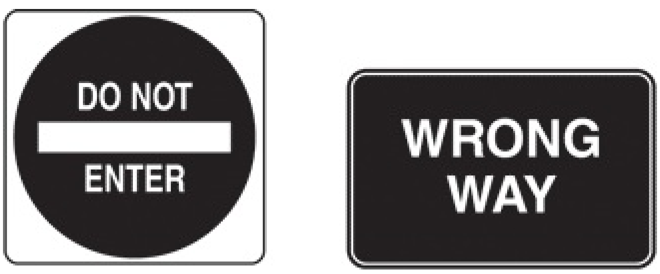
Except for some special signs, such as the STOP and YIELD signs, most regulatory signs are rectangular, with the long dimension vertical. Some regulatory signs are square. These are primarily signs using symbols instead of legend to impart information. The use of symbol signs generally conforms to international practices originally established at a 1971 United Nations conference on traffic safety. The background color of regulatory signs, with a few exceptions, is white, while legend or symbols are black. In symbol signs, a red circle with a bar through it signifies a prohibition of the movement indicated by the symbol.

The MUTCD contains many pages of standards for the appropriate size of regulatory signs and should be consulted directly on this issue.

Regulatory Signs Affecting Right-of-Way

The regulatory signs in this category have special designs reflecting the extreme danger that exists when one is ignored. These signs include the STOP and YIELD signs, which assign right-of-way at intersections and unsignalized crosswalks, and WRONG WAY and ONE WAY signs, indicating directional flow. The STOP and YIELD signs have unique shapes, and they use a red background color to denote danger. The WRONG WAY sign also uses a red background for this purpose. Figure 4.8 illustrates these signs.

Figure 4.8: Regulatory Signs Affecting Right-of-Way



(a) Do Not Enter and Wrong Way Signs

4.3-1 Full Alternative Text

(b) STOP and YIELD Signs

4.3-1 Full Alternative Text

(c) One-Way Signs

4.3-1 Full Alternative Text

(d) STOP and YIELD to Pedestrian Signs for Unsignalized Crosswalks

4.3-1 Full Alternative Text

(Source :*Manual of Uniform Traffic Control Devices*, Federal Highway Administration, Washington, D.C., 2009, as revised through May 2012, Figs 2B-1, 2B-2, 2B-11, and 2B-13, pgs 51, 55, 75, and 78.)

The “All Way” panel is mounted below a STOP sign where multiway STOP control is in use.

Consult Chapter 15 for a detailed presentation and discussion of warrants for use of STOP and YIELD signs at intersections.

Speed Limit Signs

One of the most important issues in providing for safety and efficiency of traffic movement is the setting of appropriate speed limits. To be effective, a speed limit must be communicated to the driver and should be sufficiently enforced to engender general observance. There are a number of different types of speed limits that may be imposed:

- Linear speed limits
- Areawide (statutory) speed limits
- Night speed limits
- Truck speed limits
- Minimum speed limits

Speed limits are generally stated in terms of standard U.S. units (mi/h), although a few states continue to use metric units (km/h). Speed limits must be posted in 5 mi/h increments, that is, a speed limit of 27 mi/h would be in violation of the MUTCD.

Linear speed limits apply to a designated section of roadway. Signs should be posted such that no driver can enter the roadway without seeing a speed limit sign within approximately 1,000 ft. This is not an MUTCD standard, but reflects common practice in the United States.

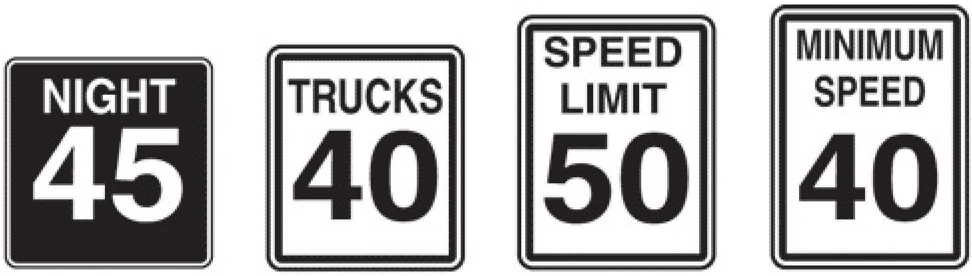
Area speed limits apply to all roads within a designated area (unless otherwise posted). A state statutory speed limit is one example of such a regulation. Cities, towns, and other local governments may also enact ordinances establishing a speed limit throughout their jurisdiction. Areawide speed limits should be posted on every facility at the boundary entering the jurisdiction for which the limit is established. Such postings *are* mandated by the MUTCD. Where an area limit is in place, a panel is posted below signs, indicating the type of area over which the limit applies.

The “reduced speed” or “speed zone ahead” signs (not shown here) should be used wherever engineering judgment indicates a need to warn drivers of a reduced speed limit for compliance. When used, however, the sign must be followed by a speed limit sign posted at the beginning of the section in which the reduced speed limit applies.

Consult Chapter 31 for a discussion of criteria for establishing an appropriate speed limit on a highway or roadway section.

Figure 4.9 shows a variety of speed signs in common use. While most signs consist of black lettering on a white background, night speed limits are posted using the reverse of this: white lettering on a black background.

Figure 4.9: Speed Limit Signs



(a) Basic Speed Limit Signs

4.3-1 Full Alternative Text

(b) Auxiliary Panels for Area Speed Limits

4.3-1 Full Alternative Text

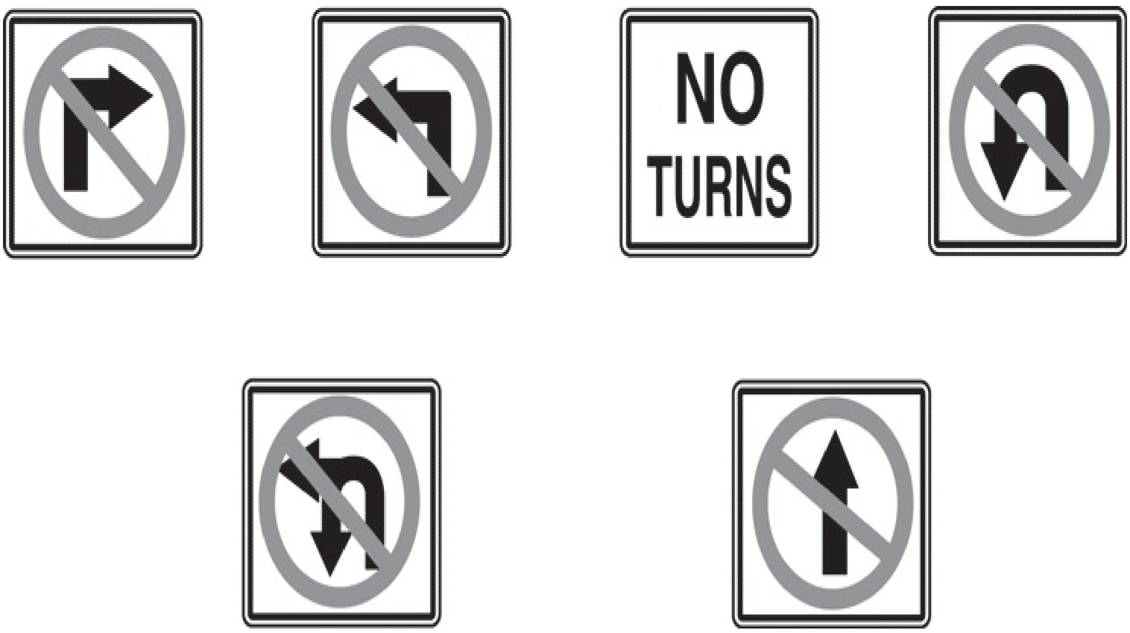
(Source: *Manual of Uniform Traffic Control Devices*, Federal Highway Administration, Washington, D.C., Dec. 2009, as revised through May 2012, Fig 2B-3, pg 57.)

Turn and Movement Prohibition

Signs

Where right, left, and/or U-turns, or even through movements, are to be prohibited, one or more of the movement prohibition signs shown in Figure 4.10 are used. In this category, international symbol signs are preferred. The traditional red circle with a bar is placed over an arrow indicating the movement to be banned.

Figure 4.10: Movement Prohibition Signs



(Source: *Manual of Uniform Traffic Control Devices*, Federal Highway Administration, Washington, D.C., 2009, as revised through May 2012, Fig 2B-4, pg 60.)

Figure 4.10: Full Alternative Text

Lane-Use Signs

Lane-use control signs are used wherever a given movement or movements are restricted and/or prohibited from designated lanes. Such situations include left-turn- and right-turn-only lanes, two-way left-turn lanes on arterials, and reversible lanes. Lane-use signs, however, may also be used to clarify lane usage even where no regulatory restriction is involved. Where lane usage is complicated, advance lane-use control signs may be used as well. Figure 4.11 illustrates these signs.

Figure 4.11: Lane-Use Control Signs

(a) Selection of Standard Lane-Use Control Signs

4.3-1 Full Alternative Text

(b) Examples of Optional Lane-Use Control Signs

4.3-1 Full Alternative Text

(c) Sample Reversible Lane Control Signs

4.3-1 Full Alternative Text

(d) Two-Way Left-Turn Lane Signs

4.3-1 Full Alternative Text

(Source: *Manual of Uniform Traffic Control Devices*, Federal Highway Administration, Washington, D.C., 2009, as revised through May 2012, Figs 2B-4 and 2B-6, pgs 60 and 65.)

Two-way left-turn lane signing must be supplemented by the appropriate markings for such a lane, as illustrated previously. Reversible lane signs must be posted as overhead signs, placed over the lane or lanes that are reversible. Roadside signs may supplement overhead signs. In situations where signing may not be sufficient to ensure safe operation of reversible lanes, overhead signals should be used.

Parking Control Signs

Curb parking control is one of the more critical aspects of urban network management. The economic viability of business areas often depends upon an adequate and convenient supply of on-street and off-street parking. At the same time, curb parking often interferes with through traffic and occupies space on the traveled way that might otherwise be used to service moving traffic. Chapter 12 provides a detailed coverage of parking issues and programs. It is imperative that curb parking regulations be clearly signed, and strict enforcement is often necessary to achieve high levels of compliance.

When dealing with parking regulations and their appropriate signing, three terms must be understood:

- Parking. A “parked” vehicle is a stationary vehicle located at the curb with the engine not running; whether or not the driver is in the vehicle is not relevant to this definition.

- Standing. A “standing” vehicle is a stationary vehicle located at the curb with the engine running and the driver in the car.

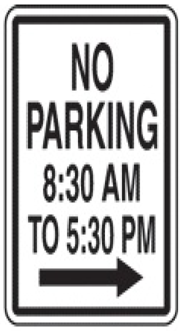
- Stopping. A “stopping” vehicle is one that makes a momentary stop at the curb to pick up or discharge a passenger; the vehicle moves on immediately upon completion of the pickup or discharge, and the driver does not leave the vehicle.

In legal terms, most jurisdictions maintain a common hierarchal structure of prohibitions. “No Stopping” prohibits stopping, standing, and parking. “No Standing” prohibits standing and parking, but permits stopping. “No Parking” prohibits parking, but permits standing and stopping.

Parking regulations may also be stated in terms of a prohibition or in terms of what is permitted. Where a sign is indicating a prohibition, red legend on a white background is used. Where a sign is indicating a permissive situation, green legend on a white background is used. Figure 4.12 illustrates a variety of parking-control signs in common use.

Figure 4.12: A Sample of Parking Control Signs

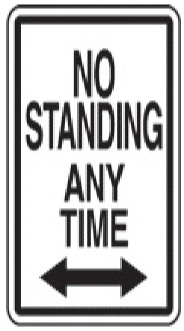
R7-1



R7-2

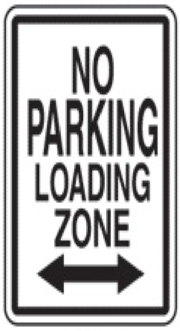
R7-2a

R7-3



R7-4

R7-5



R7-6

R7-7

R7-8

R7-8aP

R7-8bP

R7-20

R7-21

R7-21a

R7-22

R7-22a

(Source: *Manual of Uniform Traffic Control Devices*, Federal Highway Administration, Washington, D.C., 2009, as revised through May 2012, Fig 2B-24 and 2B-25, pgs 88 and 90.)

Figure 4.12: Full Alternative Text

Parking signs must be carefully designed and placed to ensure that the often complex regulations are understood by the majority of drivers. The MUTCD recommends that the following information be provided on parking-control signs, in order from top to bottom of the sign:

- The restriction or prohibition (or condition permitted in the case of a permissive sign).
- The times of the day that it is applicable (if not all day).
- The days of the week that it is applicable (if not every day).

Parking-control signs should always be placed at the boundaries of the restricted area and at intermediate locations as needed. At locations where the parking restriction changes, two signs should be placed on a single support, each with an arrow pointing in the direction of application. Where areawide restrictions are in effect, the restriction should be signed at all street locations crossing into the restricted area.

In most local jurisdictions, changes in parking regulations must be disclosed in advance using local newspapers and/or other media and/or by placing posters throughout the affected area warning of the change. It is not appropriate, for example, to place new parking restrictions overnight and then ticket or remove vehicles now illegally parked without adequate advance warning.

Other Regulatory Signs

The MUTCD provides standards and guidelines for over 100 different regulatory signs. Some of the most frequently used signs have been discussed in this section, but they are merely a sample of the many such signs that exist. New signs are constantly under development as new types of regulations are introduced. Consult the MUTCD directly for additional regulatory signs and their applications.

4.3.2 Warning Signs

> Warning signs call attention to unexpected conditions on or adjacent to a highway, street or private road open to public travel, and to situations that might not be readily apparent to road users. Warning signs alert road users to conditions that might call for a reduction of speed or an action in the interest of safety and efficient traffic operations. [Ref *1*, pg 103]

Most warning signs are diamond-shaped, with black lettering or symbols on a yellow background. A new lime-green background is being introduced for warning signs dealing with pedestrian and bicycle crossings, and school crossings. A pennant shape is used for the “No Passing Zone” sign, used in conjunction with passing restrictions on two-lane, two-way rural highways. A rectangular shape is used for some arrow indications. A circular shape is used for railroad crossing warnings.

The MUTCD specifies minimum sizes for various warning signs on different types of facilities. For the standard diamond-shaped sign, minimum sizes range from 30 in. by 30 in. to 36 in. by 36 in. On freeways, many signs must measure 48 in. by 48 in., and oversized signs up to 60 in. by 60 in. may be used when justified.

The MUTCD indicates that warning signs shall be used only in conjunction with an engineering study or based on an engineering judgment. Although this is a fairly loose requirement, it emphasizes the need to avoid overuse of such signs. A warning sign should be used only to alert drivers of conditions that they could not be normally expected to discern on their own. Overuse of warning signs encourages drivers to ignore them, which could lead to dangerous situations.

When used, warning signs must be placed far enough in advance of the hazard to allow drivers adequate time to perform the required adjustments. Table 4.1 gives the recommended advance placement distances for two conditions, defined as follows:

Table 4.1: Advance Placement Distances for Warning Signs

Posted or 85th Percentile Speed	Advance Placement Distance[1]								
	Condition A: Speed reduction and lane changing in heavy traffic[2]	Condition B: Deceleration to the listed advisory speed (mph) for the condition							
		0[3]	10[4]	20[4]	30[4]	40[4]	50[4]	60[4]	70[4]
20 mph	225 ft	100 ft[6]	N/A[5]	—	—	—	—	—	—
25 mph	325 ft	100 ft[6]	N/A[5]	N/A[5]	—	—	—	—	—
30 mph	460 ft	100 ft[6]	N/A[5]	N/A[5]	—	—	—	—	—
35 mph	565 ft	100 ft[6]	N/A[5]	N/A[5]	N/A[5]	—	—	—	—
40 mph	670 ft	125 ft	100 ft[6]	100 ft[6]	N/A[5]	—	—	—	—
45 mph	775 ft	175 ft	125 ft	100 ft[6]	100 ft[6]	N/A[5]	—	—	—
50 mph	885 ft	250 ft	200 ft	175 ft	125 ft	100 ft[6]	—	—	—
55 mph	990 ft	325 ft	275 ft	225 ft	200 ft	125 ft	N/A[5]	—	—
60 mph	1,100 ft	400 ft	350 ft	325 ft	275 ft	200 ft	100 ft[6]	—	—
65 mph	1,200 ft	475 ft	450 ft	400 ft	350 ft	275 ft	200 ft	100 ft[6]	—
70 mph	1,250 ft	550 ft	525 ft	500 ft	450 ft	375 ft	275 ft	150 ft	—
75 mph	1,350 ft	650 ft	625 ft	600 ft	550 ft	475 ft	375 ft	250 ft	100 ft[6]

[1] The distances are adjusted for a sign legibility distance of 180 ft for Condition A. The distances for Condition B have been adjusted for a sign legibility distance of 250 ft, which is appropriate for an alignment warning symbol sign. For Conditions A and B, warning signs with less than 6 in. legend or more than four words, a minimum of 100 ft should be added to the advance placement distance to provide adequate legibility of the warning sign.

[2] Typical conditions are locations where the road user must use extra time to adjust speed and change lanes in heavy traffic because of a complex driving situation. Typical signs are Merge and Right Lane Ends. The distances are determined by providing the driver a PRT of 14.0 to 14.5 s for vehicle maneuvers (2005 AASHTO Policy, Exhibit 3-3, Decision Sight Distance, Avoidance Maneuver E) minus the legibility distance of 180 ft

for the appropriate sign.

[3] Typical condition is the warning of a potential stop situation. Typical signs are Stop Ahead, Yield Ahead, Signal Ahead, and Intersection Warning signs. The distances are based on the 2005 AASHTO Policy, Exhibit 3-1, Stopping Sight Distance, providing a PRT of 2.5 s, a deceleration rate of 11.2 ft/s^2, minus the sign legibility distance of 180 ft.

[4] Typical conditions are locations where the road user must decrease speed to maneuver through the warned condition. Typical signs are Turn, Curve, Reverse Turn, or Reverse Curve. The distance is determined by providing a 2.5 s PRT, a vehicle deceleration rate of 10 ft/s^2, minus the sign legibility distance of 250 ft.

[5] No suggested distances are provided for these speeds, as the placement location is dependent on site conditions and other signing. An alignment warning sign may be placed anywhere from the point of curvature up to 100 ft in advance of the curve. However, the alignment warning sign should be installed in advance of the curve and at least 100 ft from any other signs.

[6] The minimum advance placement distance is listed as 100 ft to provide adequate spacing between signs.

(Source: *Manual on Uniform Traffic Control Devices*, Federal Highway Administration, Washington, D.C., 2009, as revised through May 2012, Table 2C-4, pg 108.)

Table 4.1: Full Alternative Text

- Condition A: High judgment required. Applies where the road user must use extra time to adjust speed and change lanes in heavy traffic due to a complex driving situation. Typical applications are warning signs for merging, lane drop, and similar situations. A perception–reaction time of 6.7 to 10.0 s is assumed plus 4.5 s for each required maneuver.

- Condition B: Deceleration to the listed advisory speed for the condition. Applies in cases where the road user must decelerate to a posted advisory speed to safely maneuver through the hazard. A 1.6 s perception–reaction time is assumed with a deceleration rate of 10 ft/s^2.

In all cases, sign visibility of 250 ft is assumed, based on sign-design standards.

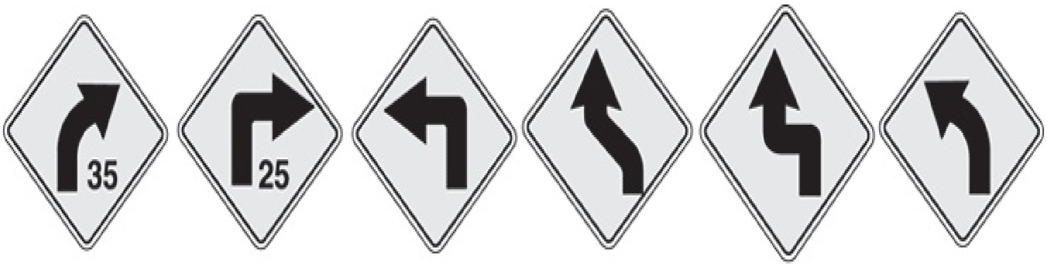
Supplementary panels indicating an advisory speed through the hazard are being replaced by speed indications directly on the warning sign itself. The advisory speed is the recommended safe speed through the hazardous area and is determined by an engineering study of the location. While no specific guideline is given, common practice is to include an advisory speed whenever the safe speed through the hazard is ≥ 10 mi/h less than the posted or statutory speed limit.

Warning signs are used to inform drivers of a variety of potentially hazardous circumstances, including the following:

- Changes in horizontal alignment
- Intersections
- Advance warning of control devices
- Converging traffic lanes
- Narrow roadways
- Changes in highway design
- Grades
- Roadway surface conditions
- Railroad crossings
- Entrances and crossings
- Miscellaneous

Figure 4.13 shows some sample warning signs from these categories.

Figure 4.13: A Sample of Warning Signs

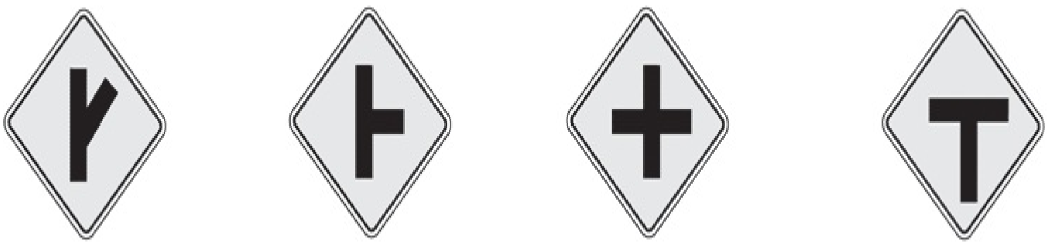


(a) Highway Alignment Warning Signs

4.3-2 Full Alternative Text

(b) Control Device Advance Warning Signs

4.3-2 Full Alternative Text



(c) Intersection Warning Signs

4.3-2 Full Alternative Text

(d) Vehicular and Non-Vehicular Crossing Warning Signs

4.3-2 Full Alternative Text

(Source: *Manual on Uniform Traffic Control Devices*, Federal Highway Administration, Washington, D.C., 2009, as revised through May 2012, Figs 2C-1, 2C-6, 2C-9, 2C-10, and 2C-11, pgs 109, 121, 127, 129, and 130.)

While not shown here, the MUTCD contains other warning signs in special sections of the manual related to work zones, school zones, and railroad crossings. The practitioner should consult these sections of the MUTCD directly for more specific information concerning these special situations.

4.3.3 Guide Signs

Guide signs provide information to road users concerning destinations, available services, and historical/recreational facilities. They serve a unique purpose in that drivers who are familiar or regular users of a route will generally not need to use them; they provide critical information, however, to unfamiliar road users. They serve a vital safety function: a confused driver approaching a junction or other decision point is a distinct hazard.

Guide signs are rectangular, with the long dimension horizontal, and have white lettering and borders. The background varies by the type of information contained on the sign. Directional or destination information is provided by signs with a green background; information on services is provided by signs with a blue background; cultural, historical, and/or recreational information is provided by signs with a brown background. Route markers, included in this category, have varying shapes and colors depending on the type and jurisdiction of the route.

The MUTCD provides guide-signing information for three types of facilities: conventional roads, freeways, and expressways. Guide signing is somewhat different from other types in that overuse is generally not a serious issue, unless it leads to confusion. Clarity and consistency of message are the most important aspects of guide signing. Several general principles may be applied:

- If a route services a number of destinations, the most important of these should be listed. Thus, a highway serving Philadelphia as well as several lesser suburbs would consistently list Philadelphia as the primary destination.

- No guide sign should list more than four destinations on a single sign. Up through 2000, the limitation was three, which should still be considered a practical maximum except in unusual circumstances. This, in conjunction with the first principle, makes the selection of priority destinations a critical part of effective guide signing.

- Where roadways have both a name and a route number, both should be indicated on the sign if space permits. In cases where only one may be listed, the route number takes precedence. Road maps and modern navigation systems show route numbers prominently, while not all facility names are included. Unfamiliar drivers are, therefore, more likely to know the route number than the facility name.

- Wherever possible, advance signing of important junctions should be given. This is more difficult on conventional highways, where junctions may be frequent and closely spaced. On freeways and expressways, this is critical, as high approach speeds make advance knowledge of upcoming junctions a significant safety issue.

- Confusion on the part of the driver must be avoided at all cost. Sign sequencing should be logical and should naturally lead the driver to the desired route selections. Overlapping sequences should be avoided wherever possible. Left-hand exits and other unusual junction features should be signed extremely carefully.

The size, placement, and lettering of guide signs vary considerably, and the manual gives information for numerous options. A number of site-specific conditions affect these design features, and there is more latitude and choice involved than for other types of highway signs. The MUTCD should be consulted directly for this information.

Route Markers

Figure 4.14 illustrates route markers that are used on all numbered routes. The signs have unique designs that signify the type of route involved:

Figure 4.14: Route Markers Illustrated

(Source: *Manual of Uniform Traffic Control Devices*, Federal Highway Administration, Washington, D.C., Dec. 2009, as revised through May 2012, Fig 2D-3, pg 143.)

Figure 4.14: Full Alternative Text

- Interstate highways have a unique shield shape, with red and blue background and white lettering. The same design is used for designated "business loops." Such loops are generally a major highway that is not part of the Interstate System but serves the business area of a city from an interchange on the Interstate System.

- U.S. route markers consist of black numerals on a white shield that is placed on a square sign with a black background.
- State route markers are designed by the individual states and, therefore, vary from state to state.
- County route markers follow a standard design, with yellow lettering on a blue background and a unique shape. The name of the county is placed on the route marker.
- Routes in national parks and/or national forests also have a unique shape and have white lettering on a brown background.

Route markers may be supplemented by a variety of panels indicating cardinal directions or other special purposes. Special purpose panels include JCT, ALT or ALTERNATE, BY-PASS, BUSINESS, TRUCK, TO, END, and TEMPORARY. Auxiliary panels match the colors of the marker they are supplementing.

Route markers are an essential part of directional guidance, particularly for numbered surface facilities. Such roads often go through local towns and developments, forming part of the local roadway system as they pass through. In many cases, a numbered route may combine with other numbered routes, and turns may be required as drivers navigate through developed areas following a numbered route. Various types of *route marker assemblies* are used to help guide drivers in these situations:

- Junction Assembly: Indicates an upcoming intersection with another numbered route.
- Advance Turn Assembly: Indicates that a turn must be made at an upcoming intersection to continue following the numbered route.
- Directional Assembly: Indicates required turning movements for route continuity at an intersection of numbered routes.
- Confirmation Assembly: Used after intersections to confirm to the driver that he/she is on the appropriate route.

- Trailblazer Assembly: Used on nonnumbered routes that lead to a numbered route. The use of these types of assemblies and other details concerning numbered route systems in the United States is covered in Chapter 31.

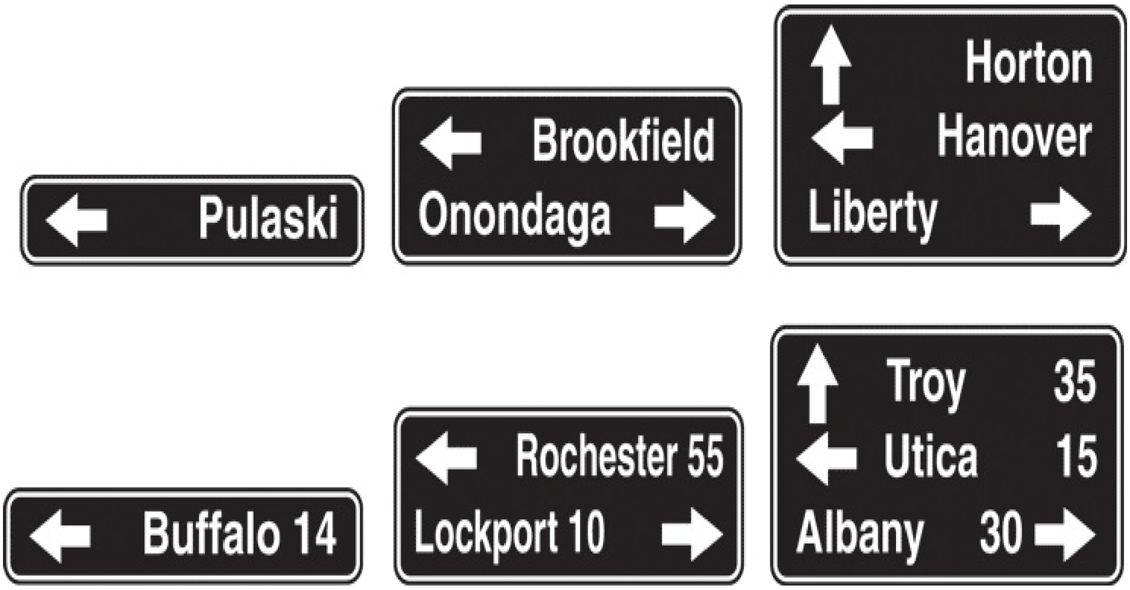
Destination Signs—Conventional Roads

Destination signs are used on conventional roadways to indicate the distance to critical destinations along the route and to mark key intersections or interchanges. On conventional roads, destination signs use an all-capital white legend on a green background. The distance in miles to the indicated destination may be indicated to the right of the destination.

Destination signs are generally used at intersections of U.S. or state numbered routes with Interstate, U.S., state numbered routes, or junctions forming part of a route to such a numbered route. Distance signs are usually placed on important routes leaving a municipality after a major junction with a numbered route.

Local street name signs are recommended for all suburban and urban junctions as well as for major intersections at rural locations. Local street name signs are categorized as conventional roadway destination signs. Figure 4.15 illustrates a selection of these signs.

Figure 4.15: Destination Signs for Conventional Roads



(Source: *Manual of Uniform Traffic Control Devices*, Federal Highway Administration, Washington, D.C., 2000, as revised through May 2012, Fig 2D-7, pg 155.)

Figure 4.15: Full Alternative Text

Destination Signs—Freeways and Expressways

Destination signs for freeways and expressways are similar, although there are different requirements for size and placement specified in the MUTCD. They differ from conventional road guide signs in a number of ways:

- Destinations are indicated in initial capitals and small letters.
- Numbered routes are indicated by inclusion of the appropriate marker type on the guide sign.
- Exit numbers are included as auxiliary panels located at the upper right or left corner of the guide sign (which indicates which side the exit is on).
- At major junctions, diagrammatic elements may be used on guide signs.

As for conventional roadways, distance signs are frequently used to indicate the mileage to critical destinations along the route. Every interchange and every significant at-grade intersection on an expressway is extensively signed with advance signing as well as with signing at the junction itself.

The distance between interchanges has a major impact on guide signing. Where interchanges are widely separated, advance guide signs can be placed as much as 5 or more miles from the interchange and may be repeated several times as the interchange is approached.

In urban and suburban situations, where interchanges are closely spaced, advanced signing is more difficult to achieve. Advance signing usually gives information only concerning the *next* interchange, to avoid confusion caused by overlapping signing sequences. The only exception to this is a distance sign indicating the distance to the next several interchanges. Thus, in urban and suburban areas with closely spaced interchanges, the advance sign for the next interchange is placed at the last off-ramp of the previous interchange.

A wide variety of sign types are used in freeway and expressway destination signing. A few of these are illustrated in Figure 4.16.

Figure 4.16: Sample Freeway and Expressway Destination Signs

(a) Advance Exit Sign

4.3-2 Full Alternative Text

(b) Gore Area Exit Sign

4.3-2 Full Alternative Text

(c) Pull-through Sign

4.3-2 Full Alternative Text

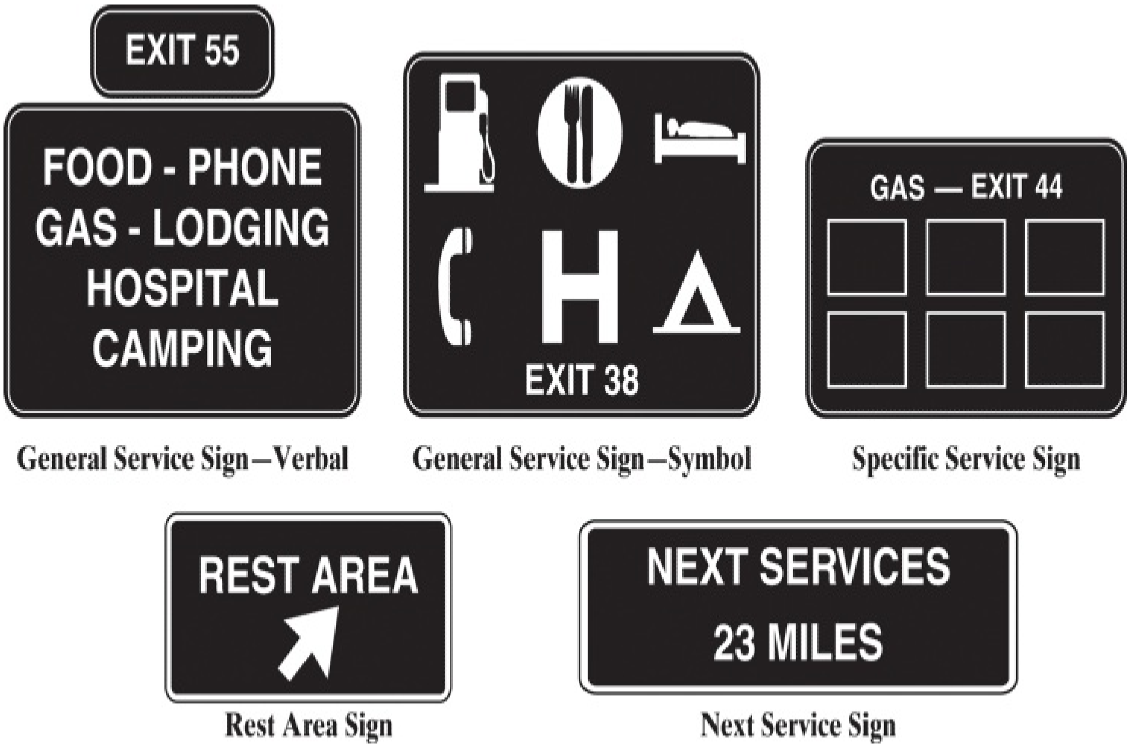
(Source: *Manual of Uniform Traffic Control Devices*, Federal Highway Administration, Washington, D.C., 2009, as revised through May 2012, Figs 2E-1, 2E-2, 2E-5, pgs 184, 196.)

Figure 4.16(a) shows a typical advance exit sign. These are placed at various distances from the interchange in accordance with the overall signing plan. The number and placement of advance exit sign are primarily dependent on interchange spacing. The gore area exit sign of Figure 4.16(b) is placed in the gore area, and is the last sign associated with a given ramp connection. Such signs are usually mounted on breakaway sign posts to avoid serious damage to vehicles straying into the gore area. The "pull-through" sign of Figure 4.16(c) is used primarily in urban or other areas with closely spaced interchanges. It is generally mounted on overhead supports next to the exit direction sign. It reinforces the direction for drivers intending to continue on the freeway.Chapter 31 contains a more detailed discussion of guide signing for freeways, expressways, and conventional roadways.

Service Guide Signs

Another important type of information drivers require is directions to a variety of traveler services. Drivers, particularly those who are unfamiliar with the area, need to be able to easily locate such services as fuel, food, lodging, medical assistance, and similar services. The MUTCD provides for a variety of signs, all using white legend and symbols on a blue background, to convey such information. In many cases, symbols are used to indicate the type of service available. On freeways, large signs using text messages may be used with exit number auxiliary panels. The maximum information is provided by freeway signs that indicate the actual brand names of available services (gas companies, restaurant names, etc.). Figure 4.17 illustrates some of the signs used to provide motorist service information.

Figure 4.17: Sample Service Guide Signs



(Source: *Manual on Uniform Traffic Control Devices*, Federal Highway Administration, Washington, D.C., 2009, as revised through May 2012, Figs 2I-2, 2I-3, 2I-5, 2J-1, pgs 302, 304, 307, 314.)

Figure 4.17: Full Alternative Text

There are a number of guidelines for specific service signing. No service is included that is more than 3 miles from the freeway interchange. No specific services are indicated where drivers cannot easily reenter the freeway at the interchange.

Specific services listed must also conform to a number of criteria regarding hours of operations and specific functions provided. All listed services must also be in compliance with all federal, state, and local laws and regulations concerning their operation. Consult the MUTCD directly for the details of these requirements.

Service guide signs on conventional highways are similar to those of Figure 4.17 but do not use exit numbers or auxiliary exit number panels.

Recreational and Cultural-Interest

Guide Signs

Information on historic, recreational, and/or cultural-interest areas or destinations is given on signs with white legend and/or symbols on a brown background. Symbols are used to depict the type of activity, but larger signs with word messages may be used as well. Figure 4.18 shows some examples of these signs. The MUTCD has introduced many acceptable symbols and should be consulted directly for illustrations of these.

Figure 4.18: Recreational and Cultural Destination Signs

(Source: *Manual of Uniform Traffic Control Devices*, Federal Highway Administration, Washington, D.C., 2009, as revised through May 2012, Figs 2M-1, 2M-2, pgs 333 and 334.)

Figure 4.18: Full Alternative Text

Mileposts

Mileposts are small 6 × 9 inch vertical white-on-green panels indicating the milage along the designated route. These are provided to allow the driver to estimate his/her progress along a route, and provide a location system for accident reporting and other emergencies that may occur along the route. Distance numbering is continuous within a state, with “zero” beginning at the south or west state lines or at the southern-most or western-most interchange at which the route begins. Where routes overlap, mileposts are continuous only for *one* of the routes. In such cases, the first milepost beyond the overlap should indicate the total milage traveled along the route that is *not* continuously numbered and posted. On some freeways, markers are placed every tenth of a mile for a more precise location system. Figure 4.19 illustrates typical mileposts.

Figure 4.19: Typical Mileposts

(Source: *Manual on Uniform Traffic Control Devices*, Federal Highway Administration, Washington, D.C., 2009, as revised through May 2012, Fig. 2H-2, pg 295.)

Figure 4.19: Full Alternative Text

4.4 Traffic Signals

The MUTCD defines nine types of traffic signals:

- Traffic control signals
- Pedestrian signals
- Emergency vehicle traffic control signals
- Traffic control signals for one-lane, two-way facilities
- Traffic control signals for freeway entrance ramps
- Traffic control signals for movable bridges
- Lane-use control signals
- Flashing beacons
- In-roadway lights

The most common of these is the traffic control signal, used at busy intersections to direct traffic to alternately stop and move.

4.4.1 Traffic Control Signals

Traffic signals are the most complicated form of traffic control devices available to traffic engineers. The MUTCD addresses the following:

- Physical standards for signal displays, including lens sizes, colors (specific pigments), arrangement of lenses within a single signal head, arrangement and placement of signal heads within an intersection, visibility requirements, etc.
- Definitions and meaning of the various indications authorized for use.
- Timing and sequence restrictions.

- Maintenance and operations criteria.

There are two important standards (requirements) regarding operation of traffic signals: (1) traffic control signals must be in operation at all times, and (2) STOP signs shall not be used in conjunction with a traffic control signal unless it is operating in the RED-flashing mode at all times.

No traffic signal should ever be “dark,” that is, showing no indications. This is particularly confusing to drivers and can result in accidents. Any accidents occurring while a signal is in the dark mode are the legal responsibility of the agency operating the signal in most states. When signals are inoperable, signal heads should be bagged or taken down to avoid such confusion. In power outages, police or other authorized agents should be used to direct traffic at all signalized locations.

The second principle relates to a common past practice—turning off signals at night and using STOP control during these hours. The problem is that during daytime hours, the driver may be confronted with a green signal *and* a STOP sign. This is extremely confusing and is no longer considered appropriate. The use of STOP signs in conjunction with permanent operation of a red flashing light is permissible, as the legal interpretation of a flashing red signal is the same as that of a STOP sign.

Signal Warrants

Traffic signals, when properly installed and operated at appropriate locations, provide a number of significant benefits:

- With appropriate physical designs, control measures, and signal timing, the capacity of critical intersection movements is increased.

- The frequency and severity of accidents are reduced for certain types of crashes, including right-angle, turn, and pedestrian accidents.

- When properly coordinated, signals can provide for nearly continuous movement of through traffic along an arterial at a designated speed under favorable traffic conditions.

- They provide for interruptions in heavy traffic streams to permit vehicular and pedestrian traffic to safely cross.

At the same time, misapplied or poorly designed signals can cause excessive delay, signal violations, increased accidents (particularly rear-end accidents), and drivers rerouting their trips to less appropriate routes.

The MUTCD provides very specific warrants for the use of traffic control signals. These warrants are far more detailed than those for other devices, due to their very high cost (relative to other control devices) and the negative impacts of their misapplication. Thus, the manual is clear that traffic control signals shall be installed only at locations where an engineering study has indicated that one or more of the specified warrants has been met and that application of signals will improve safety and/or capacity of the intersection. The manual goes further; if a study indicates that an existing signal is in place at a location that does not meet any of the warrants, it should be removed and replaced with a less severe form of control.

The MUTCD details nine different warrants, any one of which may indicate that installation of a traffic control signal is appropriate. Chapter 15 contains a detailed treatment of these warrants and their application as part of an overall process for determining the appropriate form of intersection control for any given situation.

Signal Indications

The MUTCD defines the meaning of each traffic control signal indication in great detail. The basic ideas in these definitions are summarized below:

- Green ball. A steady green circular indication allows vehicular traffic facing the ball to enter the intersection to travel straight through the intersection or to turn right or left, except when prohibited by lane-use controls or physical design. Turning vehicles must yield the right-of-way to opposing through vehicles and to pedestrians legally in a conflicting crosswalk. In the absence of pedestrian signals, pedestrians may proceed to cross the roadway within any legally marked or unmarked crosswalk.

- Yellow ball. The steady yellow circular indication is a transition between the green ball and the red ball indication. It warns drivers that the related green movement is being terminated or that a red indication will immediately follow. In most states, drivers are allowed to legally enter the intersection during a “yellow” display. Some states, however, only allow the drivers to enter on “yellow” if they can clear the intersection before the “yellow” terminates. This is very difficult for drivers, however, as they do not know when the “yellow” is timed to end. Where no pedestrian signals are in use, pedestrians may not begin crossing a street during the “yellow” indication.

- Red ball. The steady red circular indication requires all traffic (vehicular and pedestrian) facing it to stop at the STOP line, crosswalk line (if no STOP line exists), or at the conflicting pedestrian path (if no crosswalk or STOP line exists). All states allow right-turning traffic to proceed with caution after stopping, unless specifically prohibited by signing or statute. Some states allow left-turners from one one-way street turning into another to proceed with caution after stopping, but this is far from a universal statute.

- Flashing ball. A flashing “yellow” allows traffic to proceed with caution through the intersection. A flashing “red” has the same meaning as a STOP sign—the driver may proceed with caution after coming to a complete stop. Use of a flashing “green” is prohibited, and has no meaning.

- Arrow indications. Green, yellow, and red arrow indications have the same meanings as ball indications, except that they apply only to the movement designated by the arrow. A green left-turn arrow is only used to indicate a protected left turn (i.e., a left turn made on a green arrow will not encounter an opposing vehicular through movement). Such vehicles, however, may encounter pedestrians legally in the conflicting crosswalk and must yield to them. A green right-turn arrow is shown only when there are no pedestrians legally in the conflicting crosswalk. Yellow arrows warn drivers that the green arrow is about to terminate. The yellow arrow may be followed by a green ball indication where the protected left- and/or right-turning movement is followed by a permitted movement. A “permitted” left turn is made against an opposing vehicular flow. A “permitted” right turn is made against a conflicting pedestrian flow. It is followed by a red arrow (or red ball) where the movement must stop.

The MUTCD provides additional detailed discussion on how and when to apply various sequences and combinations of indications.

Signal Faces and Visibility Requirements

In general, a signal face should have three to five signal lenses, with some exceptions allowing for a sixth to be shown. Two lens sizes are provided for: 8-inch diameter and 12-inch lenses. The manual now mandates the use of 12-inch lenses for all *new* signal installations, except when used as a supplemental signal for pedestrian use only, or when used at very closely spaced intersections where visibility shields cannot be effectively used. Several other special-case uses of 8-inch lenses are also permitted—consult the MUTCD directly for these. Eight-inch lenses at existing installations may be kept in place for their useful service life. If replaced, they must be replaced with 12 inch lenses.

Table 4.2 shows the minimum visibility distances required for signal faces. Table 4.3 shows the minimum number of signal faces that must be provided for the major movement on each approach, where the speed limit or the 85th percentile speed is more than 45 mi/h. These apply even if the major movement is a turning movement. This requirement provides some measure of redundancy in case of an unexpected bulb failure.

Table 4.2: Minimum Sight Distances for Signal Faces

85th Percentile Speed (mph)	Minimum Sight Distance (ft)
20	175
25	215
30	270
35	325
40	390
45	460
50	540
55	625
60	715

Note: Distances in this table are derived from stopping sight distance plus an assumed queue length for shorter cycle lengths (60 to 75 s).

(Source: *Manual on Uniform Traffic Control Devices*, Federal Highway Administration, Washington, D.C., 2009, as revised through May 2012, Table 4D-1, pg 461.)

Table 4.2: Full Alternative Text

Table 4.3: Minimum Number of Signal Heads for Primary Movements

Number of through Lanes on Approach	Total Number of through Signal Faces for Approach*	Minimum Number of Over head-Mounted Primary through Signal Faces for Approach
1	2	1
2	2	1
3	3	2**
≥4	≥4	3**

* A minimum of two through signal faces is always required (see Section 4D.11 of the MUTCD).

These recommended numbers of through signal faces may be exceeded. Also, see cone of vision requirements otherwise indicated in Section 4D.13 of the MUTCD.

** If practical, all of the recommended number of primary through signal faces should be located overhead.

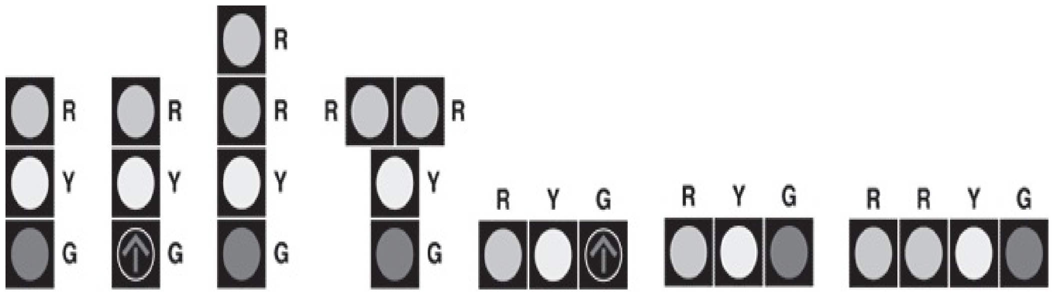
(Source: *Manual of Uniform Traffic Control Devices*, Federal Highway Administration, Washington, D.C., 2009, as revised through May 2012, Table 4D-2, pg 361.)

Table 4.3: Full Alternative Text

The arrangement of lenses on a signal face is also limited to approved options. In general, the red ball must be at the top of a vertical signal face or at the left of a horizontal signal face, followed by the yellow and green. Where arrow indications are on the same signal face as ball indications, they are located on the bottom of a vertical display or right of a horizontal display. Figure 4.20 shows a selection of the most commonly used lens arrangements. The MUTCD contains detailed discussion of the applicability of various signal face designs.

Figure 4.20: Typical Signal

Head Arrangements

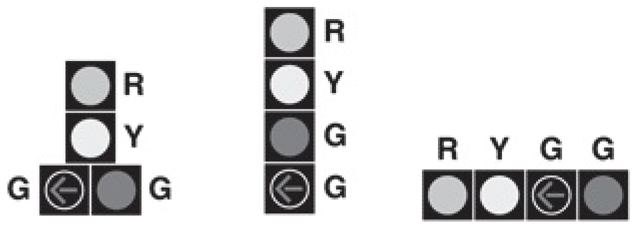


(a) Typical Signal Head Arrangements Where No Protected LTs Exist

4.4-4 Full Alternative Text

(b) Typical Signal Head Arrangements for Fully Protected LTs

4.4-4 Full Alternative Text



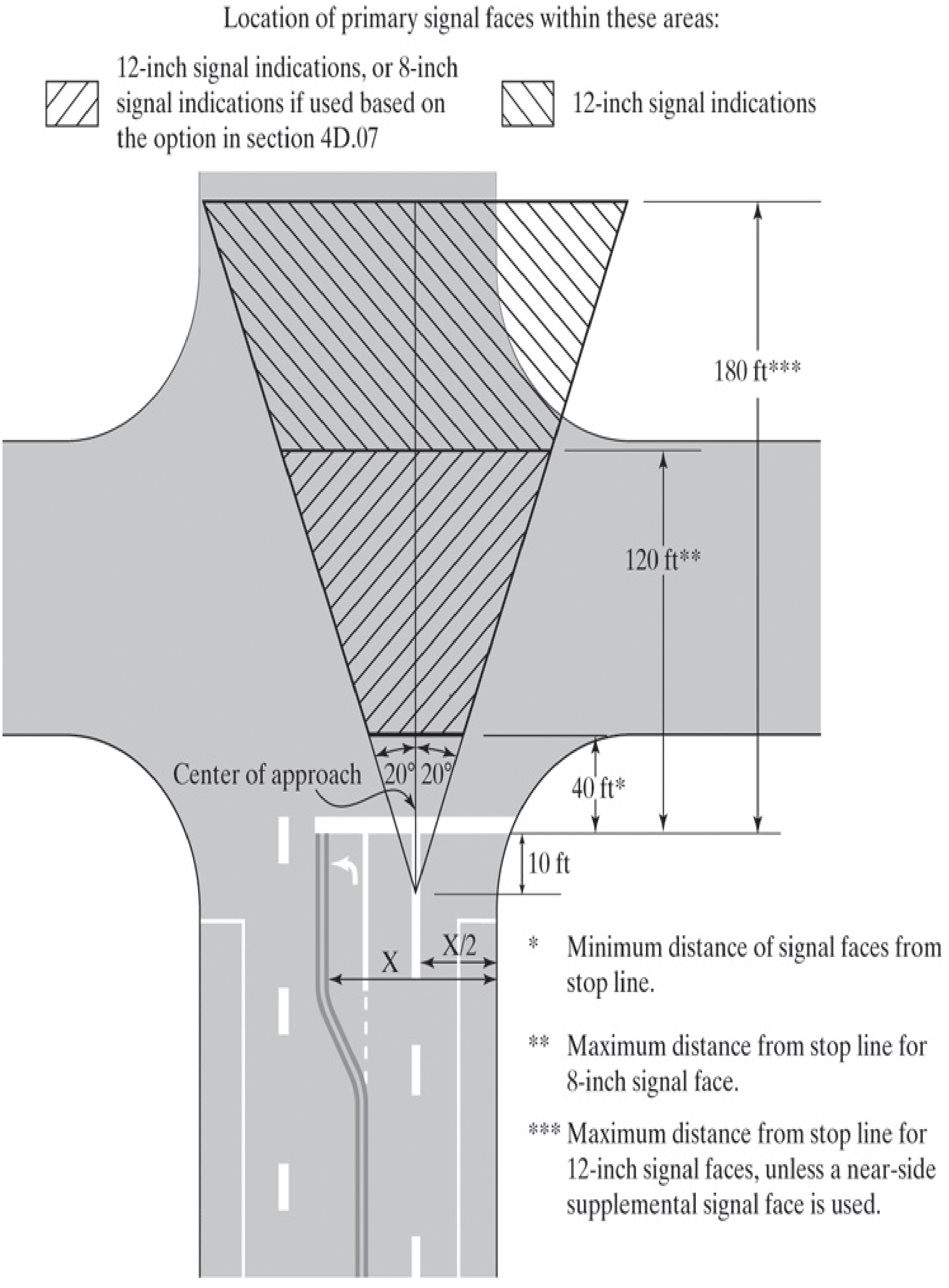
(c) Typical Signal Head Arrangements for Permitted or Protected/Permitted LTs

4.4-4 Full Alternative Text

(Source: *Manual on Uniform Traffic Control Devices*, Federal Highway Administration, Washington, D.C., 2009, as revised through May 2012, Figs 4D-2, 4D-8, and 4D-9, pgs 458, 469, and 470.)

Figure 4.21 shows the preferred placement of signal faces. At least one of the required signal faces for the major movement must be located between 40 and 150 ft of the STOP line, unless the physical design of the intersection prevents it. Horizontal placement should be within 20° of the centerline of the approach, facing straight ahead.

Figure 4.21: Horizontal Location of Signal Faces



Notes: 1. See Section 4D.11 for approaches with posted, statutory, or 85th percentile speeds of 45 mph or higher

2. See Section 4D.13 regarding location of signal faces that display a CIRCULAR GREEN signal indication for a permissive left-turn movement on approaches with an exclusive left-turn

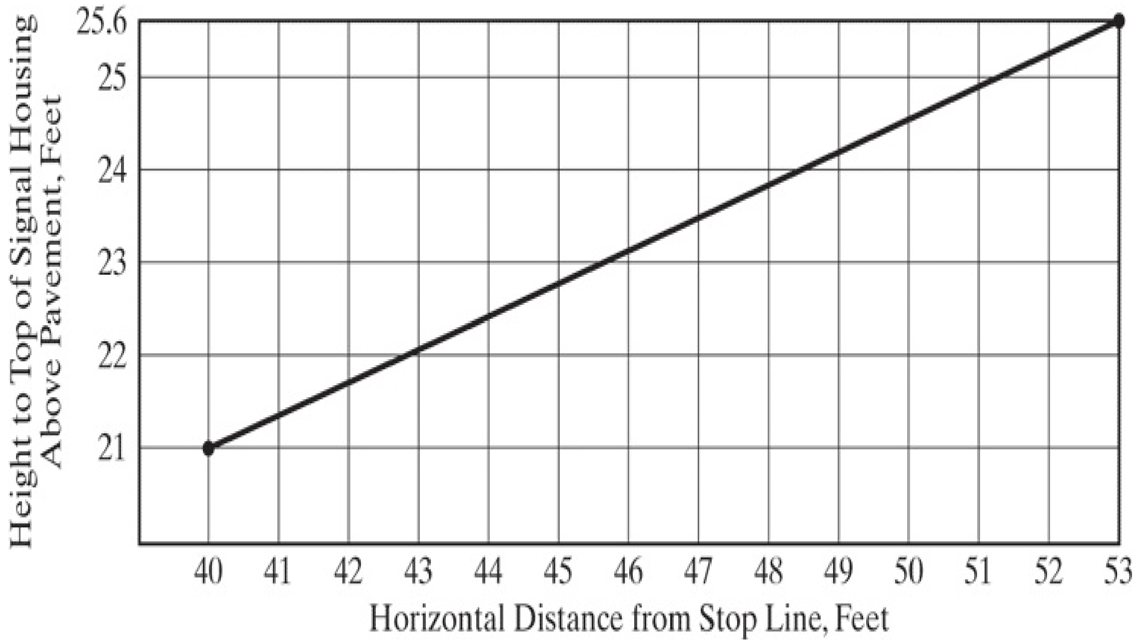
lane or lanes

(Source: *Manual of Uniform Traffic Control Devices*, Federal Highway Administration, Washington, D.C., 2009, AS REVISED THROUGH May 2012, Fig 4D-4, pg 463.)

Figure 4.21: Full Alternative Text

Figure 4.22 illustrates the standard for vertical placement of signal faces that are between 40 and 53 ft from the STOP line. The standard prescribes the maximum height of the top of the signal housing above the pavement.

Figure 4.22: Vertical Location of Signal Faces



(Source: *Manual of Uniform Traffic Control Devices*, Federal Highway Administration, Washington, D.C., 2009, as revised through May 2012, Fig 4D-5, pg 465.)

Figure 4.22: Full Alternative Text

Operational Restrictions

Continuous operation of traffic control signals is critical for safety. No signal face should ever be “dark” (i.e., with no lens illuminated). In cases where signalization is not deemed necessary at night, signals must be operated in the flashing mode (“yellow” for one street and “red” for the other). Signal operations must also be designed to allow flashing operation to be maintained even when the signal controller is undergoing maintenance or replacement.

When being installed, signal faces should be bagged and turned to make it obvious to drivers that they are not in operation. Signals should be made operational as soon as possible after installation—again, to minimize possible confusion to drivers.

Bulb maintenance is a critical part of safe signal operation, as a burned-out bulb can make a signal face appear to be “dark” during certain intervals. A regular bulb-replacement schedule must be maintained. It is common to replace signal bulbs regularly at about 75%–80% of their expected service life to avoid burn-out problems. Other malfunctions can lead to other nonstandard indications appearing, although most controllers are programmed to fall back to the flashing mode in the event of most malfunctions. Most signal agencies maintain a contract with a private maintenance organization that requires rapid response (in the order of 15–30 minutes) to any reported malfunction. The agency can also operate its own maintenance group under similar rules. Any accident occurring during a signal malfunction can lead to legal liability for the agency with jurisdiction.

4.4.2 Pedestrian Signals

The MUTCD now mandates the use of symbol pedestrian signals, which now must replace older “WALK” and “DON’T WALK” designs:

- Walking man (steady). The “WALK” indication is the image of a walking person in the color white. This indicates that it is permissible for a pedestrian to enter the crosswalk to begin crossing the street.
- Upraised hand (flashing). The “DON’T WALK” indication is an upraised hand in the color Portland orange. In the flashing mode, it indicates that no pedestrian may enter the crosswalk to begin crossing

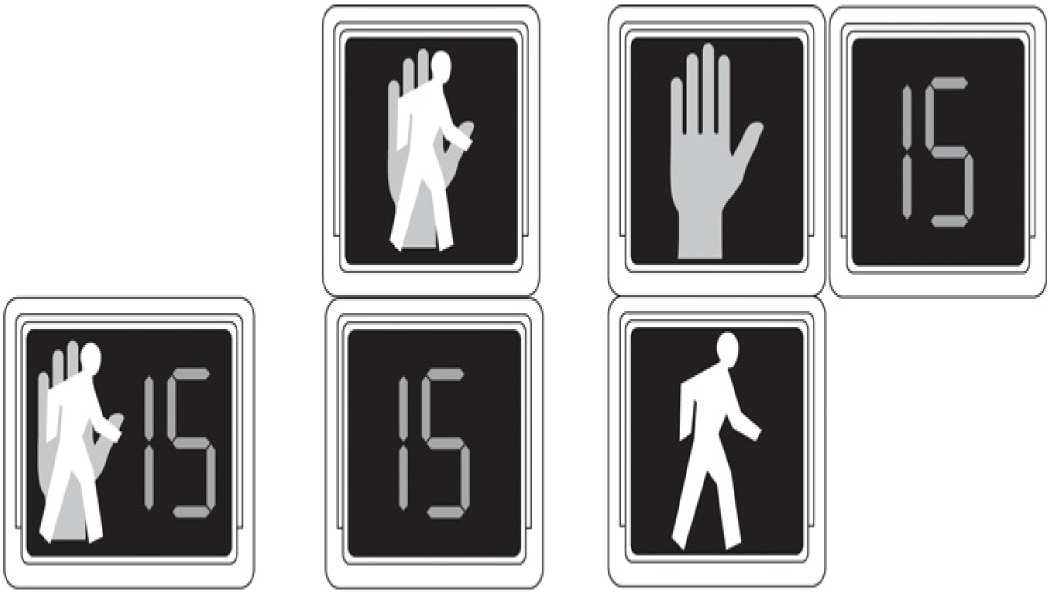
the street but that those already crossing may continue safely.

- Upraised hand (steady). In the steady mode, the upraised hand indicates that no pedestrian should begin crossing and that no pedestrian should still be in the crosswalk.

Through 2000, a flashing “WALK” indication was an option that could be used to indicate that right-turning vehicles may be conflicting with pedestrians legally in the crosswalk. The current manual does not permit a flashing WALKING MAN, effectively discontinuing this practice.

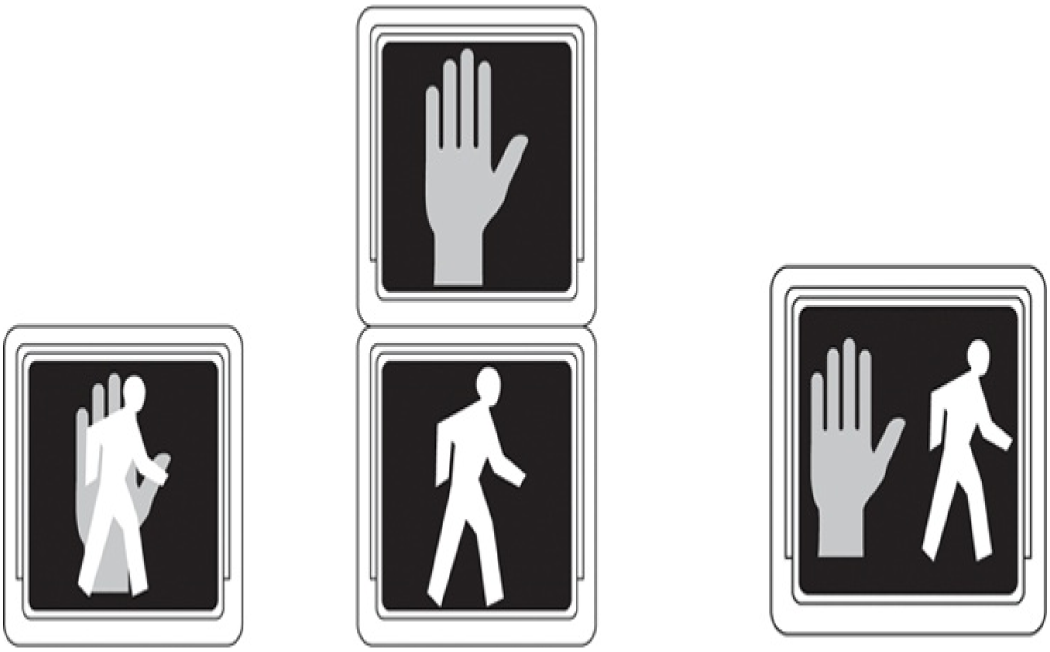
Figure 4.23 shows the new pedestrian signals. Note that both the UPRAISED HAND and WALKING MAN symbols should be shown in the form of a solid image. They may be located side by side on a single-section signal, may overlap on a single-section signal, or be arranged vertically on two-section signal. When not overlapping, the UPRAISED HAND is on the left, or on top in these displays. When not illuminated, neither symbol should be readily visible to pedestrians at the far end of the crosswalk.

Figure 4.23: Typical Pedestrian Signals



(a) Pedestrian Signals with Countdown Clock

4.4-4 Full Alternative Text



(b) Pedestrian Signals without Countdown Clock

4.4-4 Full Alternative Text

(Source: *Manual of Uniform Traffic Control Devices*, Federal Highway Administration, Washington, D.C., 2009, as revised through May 2012, Fig 4E-1, pg 496.)

Pedestrian signals may be employed with or without a “countdown clock.” The countdown clock shows pedestrians how many seconds are left before the “WALK” and “Flashing DON’T WALK” intervals will end. They are generally effective in minimizing the number of pedestrians who remain in the crosswalk when the “DON’T WALK” interval is active. Countdown clocks are most often used in conjunction with pre-timed vehicular signals, as knowledge of the end of each interval is necessary to determine how many seconds are left on the clock.

Chapters 19 and 20 discuss the use and application of pedestrian signals in the context of overall intersection control and operation. They include a discussion of when and where pedestrian signals are mandated as part of a signalization design.

4.4.3 Other Traffic Signals

The MUTCD provides specific criteria for the design, placement, and use of a number of other types of signals, including the following:

- Beacons
- In-roadway lights
- Lane-use control signals
- Ramp control signals (or ramp meters)

Beacons are generally used to identify a hazard or call attention to a critical control device, such as a speed limit sign, a STOP or YIELD sign, or a DO NOT ENTER sign. Lane-use control signals are used to control reversible lanes on bridges, in tunnels, and on streets and highways.

4.4.4 Traffic Signal Controllers

Modern traffic signal controllers are a complex combination of hardware and software that implements signal timing and ensures that signal indications operate consistently and continuously in accordance with the programmed signal-timing. Each signalized intersection has a controller dedicated to implementing the signal-timing plan at that intersection. In addition, master controllers coordinate the operation of many signals, allowing signals along an arterial or in a network to be coordinated to provide progressive movement and/or other arterial or network control policies.

Individual traffic controllers may operate in the *pre-timed* or *actuated* mode. In pre-timed operation, the sequence and timing of every signal indication is preset and is repeated in each signal cycle. In actuated operation, the sequence and timing of some or all of the green indications may change on a cycle-by-cycle basis in response to detected vehicular and pedestrian demand. Chapter 16 contains a complete discussion of modern signal hardware.

4.5 Special Types of Control

While not covered in this chapter, the MUTCD contains significant material covering special control situations, including the following:

- School zones
- Railroad crossings
- Construction and maintenance zones
- Pedestrian and bicycle controls

These situations invariably involve a combination of signing, markings, and/or signals for fully effective control. Consult the MUTCD directly for details on these and other applications not covered herein.

4.6 Closing Comments

This chapter has provided an introduction and overview to the design, placement, and use of traffic control devices. The MUTCD is not a stagnant document, and updates and revisions are constantly being issued. Thus, it is imperative that users consult the latest version of the manual, and all of its formal revisions. For convenience, the MUTCD can be accessed online. This is a convenient way of using the manual, as all updates and revisions are always included. Similarly, virtually every signal manufacturer has a web site that can be accessed to review detailed specifications and characteristics of controllers and other signal hardware and software.

References

- 1. *Manual of Uniform Traffic Control Devices*, Federal Highway Administration, U.S. Department of Transportation, Washington, D.C., 2009, as revised through May 2012.

- 2. Hawkins, H.G., "Evolution of the MUTCD: Early Standards for Traffic Control Devices," *ITE Journal*, Institute of Transportation Engineers, Washington, D.C., July 1992.

- 3. Hawkins, H.G., "Evolution of the MUTCD: Early Editions of the MUTCD," *ITE Journal*, Institute of Transportation Engineers, Washington, D.C., August 1992.

- 4. Hawkins, H.C., "Evolution of the MUTCD: The MUTCD since WWII," *ITE Journal*, Institute of Transportation Engineers, Washington, D.C., November 1992.

- 5. Hawkins, H.C., "Evolution of the MUTCD Mirrors American Progress since the 1920's," *Roads and Bridges*, Scranton Gillette, Communications Inc., Des Plaines, IL, July 1995.

Problems

1. 4-1. Define the following terms with respect to their meaning in the current edition of the MUTCD: standard, guideline, option, and support.

2. 4-2. Describe how color, shape, and legend are used to convey and reinforce messages given by traffic control devices.

3. 4-3. Why should overuse of regulatory and warning signs be avoided? Why is this not a problem with guide signs?

4. 4-4. How far from the point of a hazard should the following warning signs be placed?

    1. A “STOP ahead” warning sign on a road with a posted speed limit of 50 mi/h.

    2. A “curve ahead” warning sign with an advisory speed of 30 mi/h on a road with a posted speed limit of 45 mi/h.

    3. A “merge ahead” warning sign on a ramp with an 85th percentile speed of 35 mi/h.

5. 4-5. Select a one-mile stretch of freeway in your vicinity. Drive one direction of this facility with a friend or colleague. The passenger should count and note the number and type of traffic signs encountered. Are any of them confusing? Suggest improvements as appropriate. Comment on the overall quality of the signing in the test section.

6. 4-6. Select one signalized and one STOP or YIELD controlled intersection in your neighborhood. Note the placement of all devices at each intersection. Do they appear to meet MUTCD standards? Is visibility of all devices adequate? Comment on the effectiveness of traffic controls at each intersection.

Chapter 5 Traffic Stream Characteristics

Traffic streams are made up of individual drivers and vehicles interacting with each other and with the physical elements of the roadway and its general environment. Because both driver behavior and vehicle characteristics vary, individual vehicles within the traffic stream do not behave in exactly the same manner. Further, no two traffic streams will behave in exactly the same way, even in similar circumstances, because driver behavior varies with local characteristics and driving habits.

Dealing with traffic, therefore, involves an element of variability. A flow of water through channels and pipes of defined characteristics will behave in an entirely predictable fashion, in accord with the laws of hydraulics and fluid flow. A given flow of traffic through streets and highways of defined characteristics will vary with both time and location. Thus, the critical challenge of traffic engineering is to plan and design for a medium that is not predictable in exact terms—one that involves both physical constraints and the complex behavioral characteristics of human beings.

Fortunately, although exact characteristics vary, there is a reasonably consistent range of driver and, therefore, traffic stream behavior. Drivers on a highway designed for a safe speed of 60 mi/h may select speeds in a broad range (perhaps 45–65 mi/h); few, however, will travel at 80 mi/h or at 20 mi/h.

In describing traffic streams in quantitative terms, the purpose is to both understand the inherent variability in their characteristics and define normal ranges of behavior. To do so, key parameters must be defined and measured. Traffic engineers will analyze, evaluate, and ultimately plan improvements to traffic facilities based on such parameters and their knowledge of normal ranges of behavior.

This chapter focuses on the definition and description of the parameters most often used for this purpose and on the characteristics normally observed in traffic streams. These parameters are, in effect, the traffic engineer's measure of reality, and they constitute a language with which traffic streams are described and understood.

5.1 Types of Facilities

Traffic facilities are broadly separated into two principal categories:

- Uninterrupted flow
- Interrupted flow

Uninterrupted flow facilities have no external interruptions to the traffic stream. Pure uninterrupted flow exists primarily on freeways, where there are no intersections at grade, traffic signals, STOP or YIELD signs, or other interruptions external to the traffic stream itself. Because such facilities have full control of access, there are no intersections at grade, driveways, or any forms of direct access to abutting lands. Thus, the characteristics of the traffic stream are based solely on the interactions among vehicles and with the roadway and the general environment.

Although pure uninterrupted flow exists only on freeways, it can also exist on sections of surface highway, most often in rural areas, where there are long distances between fixed interruptions. Thus, uninterrupted flow may exist on some sections of rural two-lane highways and rural and suburban multilane highways. As a very general guideline, it is believed that uninterrupted flow can exist in situations where the distance from traffic signals and/or other significant fixed interruptions is more than 2 miles.

It should be remembered that the term “uninterrupted flow” refers to a type of facility, not the quality of operations on that facility. Thus, a freeway that experiences breakdowns and long delays during peak hours is still operating under uninterrupted flow. The causes for the breakdowns and delay are not external to the traffic stream but are caused entirely by the internal interactions within the traffic stream.

Interrupted flow facilities are those that incorporate fixed external interruptions into their design and operation. The most frequent and operationally significant external interruption is the traffic signal. The traffic signal alternatively starts and stops a given traffic stream, creating platoons of vehicles progressing down the facility. Other fixed interruptions include STOP and YIELD signs, unsignalized at-grade intersections, driveways, curb parking maneuvers, and other land-access operations. Virtually all urban surface streets and highways are interrupted flow facilities.

The major difference between uninterrupted and interrupted flow facilities is the impact of time. On uninterrupted facilities, the physical facility is available to drivers and vehicles at all times. On a given interrupted flow facility, movement is periodically barred by “red” signals. The signal timing, therefore, limits access to particular segments of the facility in time. Further, rather than a continuously moving traffic stream, at traffic signals, the traffic stream is periodically stopping and starting again.

Interrupted flow is, therefore, more complex than uninterrupted flow. Although many of the traffic flow parameters described in this chapter apply to both types of facilities, this chapter focuses primarily on the characteristics of uninterrupted flow. Many of these characteristics may also apply within a moving platoon of vehicles on an interrupted flow facility. Specific characteristics of traffic interruptions and their impact on flow are discussed in detail in Chapter 18.

5.2 Traffic Stream Parameters

Traffic stream parameters fall into two broad categories. *Macroscopic parameters* describe the traffic stream as a whole; *microscopic parameters* describe the behavior of individual vehicles or pairs of vehicles within the traffic stream.

The three principal macroscopic parameters that describe a traffic stream are (1) volume or rate of flow, (2) speed, and (3) density. Microscopic parameters include (1) the speed of individual vehicles, (2) headway, and (3) spacing.

5.2.1 Volume and Rate of Flow

Traffic volume is defined as the number of vehicles passing a point on a highway, or a given lane or direction of a highway, during a specified time interval. The unit of measurement for volume is simply "vehicles," although it is often expressed as "vehicles per unit time." Units of time used most often are "per day" or "per hour."

Daily volumes are used to establish trends over time, and for general planning purposes. Detailed design or control decisions require knowledge of hourly volumes for the peak hour(s) of the day.

Rates of flow are generally stated in units of "vehicles per hour," but represent flows that exist for periods of time less than one hour. A volume of 200 vehicles observed over a 15-minute period may be expressed as a rate of 200 × 4 = 800 vehicles/hour, even though 800 vehicles would not be observed if the full hour were counted. The 800 vehicles/hour becomes a rate of flow that exists for a 15-minute interval.

Daily Volumes

As noted, daily volumes are used to document annual trends in highway usage. Forecasts based upon observed trends can be used to help plan improved or new facilities to accommodate increasing demand.

There are four daily volume parameters that are widely used in traffic engineering:

- Average annual daily traffic (AADT). The average 24-hour volume at a given location over a full 365-day year; the number of vehicles passing a site in a year divided by 365 days (366 days in a leap year).
- Average annual weekday traffic (AAWT). The average 24-hour volume occurring on weekdays over a full 365-day year; the number of vehicles passing a site on weekdays in a year divided by the number of weekdays (usually 260).
- Average daily traffic (ADT). The average 24-hour volume at a given location over a defined time period less than one year; a common application is to measure an ADT for each month of the year.
- Average weekday traffic (AWT). The average 24-hour weekday volume at a given location over a defined time period less than one year; a common application is to measure an AWT for each month of the year.

All of these volumes are stated in terms of vehicles per day (veh/day). Daily volumes are generally not differentiated by direction or lane but are totals for an entire facility at the designated location.

Table 5.1 illustrates the compilation of these daily volumes based upon one year of count data at a sample location. The data in Table 5.1 are in a form that comes from a permanent count location (i.e., a location where automated detection of volume and transmittal of counts electronically to a central data bank is in place). AWT for each month is found by dividing the total monthly weekday volume by the number of weekdays in the month (column 5/column 2). The ADT is the total monthly volume divided by the number of days in the month (column 4/column 3) and AADT is the total observed volume for the year divided by 365 days/year. Average annual weekday traffic is the total observed volume on weekdays divided by 260 weekdays/year.

Table 5.1: Illustration of Daily

Volume Parameters

1. Month	2. No. of Weekdays in Month (days)	3. Total Days in Month (days)	4. Total Monthly Volume (veh)	5. Total Weekday Volume (veh)	6. AWT 5/2 (veh/day)	7. ADT 4/3 (veh/day)
Jan	22	31	425,000	208,000	9,455	13,710
Feb	20	28	410,000	220,000	11,000	14,643
Mar	22	31	385,000	185,000	8,409	12,419
Apr	22	30	400,000	200,000	9,091	13,333
May	21	31	450,000	215,000	10,238	14,516
Jun	22	30	500,000	230,000	10,455	16,667
Jul	23	31	580,000	260,000	11,304	18,710
Aug	21	31	570,000	260,000	12,381	18,387
Sep	22	30	490,000	205,000	9,318	16,333
Oct	22	31	420,000	190,000	8,636	13,548
Nov	21	30	415,000	200,000	9,524	13,833
Dec	22	31	400,000	210,000	9,545	12,903
Total	**260**	**365**	**5,445,000**	**2,583,000**	—	—

AADT=5,445,000/365=14,918 veh/dayAAWT=2,583,000/260=9,935

Table 5.1: Full Alternative Text

The sample data of Table 5.1 give a capsule description of the character of the facility on which it was measured. Note that ADTs are significantly higher than AWTs in each month. This suggests that the facility is serving a recreational or vacation area, with traffic strongly peaking on weekends. Also, both AWTs and ADTs are highest during the summer months, suggesting that the facility serves a warm-weather recreational/vacation area. Thus, if a detailed study were needed to provide data for an upgrading of this facility, the period to focus on would be weekends during the summer.

Because the AADT is substantially higher than the AAWT, a heavy weekend traffic load is indicated. If we focus solely on the weekends, the disparity is even greater. Consider the following:

- Total annual traffic on weekends = 5,445,000 − 2,583,000 = 2,862,000 vehicles
- Total number of weekend days in year = 365 − 260 = 105 weekend

days

- ADT on weekends = 2,862,000/105 = 27,257 veh/day.

The average traffic per day on weekends is almost 3× the AAWT! Any planning or design focusing on the AADT would grossly underestimate what is needed to accommodate this demand.

Hourly Volumes

Daily volumes, while useful for planning purposes, cannot be used alone for design or operational analysis purposes. Volume varies considerably over the 24 hours of the day, with periods of maximum flow occurring during the morning and evening commuter “rush hours.” The single hour of the day that has the highest hourly volume is referred to as the *peak hour*. The traffic volume within this hour is of greatest interest to traffic engineers for design and operational analysis usage. The peak-hour volume is generally stated as a *directional* volume (i.e., each direction of flow is counted separately).

Highway design and control must be designed to adequately serve the peak-hour traffic volume in the peak direction of flow. Since traffic going one way during the morning peak is usually going the opposite way during the evening peak, *both* sides of a facility must generally be designed to accommodate the peak directional flow during the peak hour. Where the directional disparity is significant, the concept of reversible lanes is sometimes useful. Washington, D.C., for example, makes extensive use of reversible lanes (direction changes by time of day) on its many wide boulevards and some of its freeways.

In design applications, peak-hour volumes are sometimes estimated from projections of the AADT. Traffic forecasts are most often cast in terms of AADTs based on documented trends and/or forecasting models. Because daily volumes, such as the AADT, are more stable than hourly volumes, projections can be more confidently made using them. AADTs are then converted to a peak-hour volume in the peak direction of flow. This is referred to as the “directional design hour volume” (DDHV), and is found using the following relationship:

$$DDHV = AADT \times K \times D \quad [5\text{-}1]$$

where:

K=proportion of daily traffic occurring duringthe peak hour (decimal), andD hour traffic traveling inthe peak direction of flow (decimal).

For design, the *K* factor often represents the proportion of AADT occurring during the *30th peak hour* of the year. If the 365 peak-hour volumes of the year at a given location are listed in descending order, the 30th peak hour is 30th on the list and represents a volume that is exceeded in only 29 hours of the year. For rural facilities, the 30th peak hour may have a significantly lower volume than the worst hour of the year, as critical peaks may occur only infrequently. In such cases, it is not considered economically feasible to invest large amounts of capital in providing additional capacity that will be used in only 29 hours of the year. In urban cases, where traffic is frequently at capacity levels during all daily commuter peaks, the 30th peak hour is often not substantially different from the highest peak hour of the year.

Factors *K* and *D* are based upon local or regional characteristics at existing locations. Most state highway departments, for example, continually monitor these proportions, and publish appropriate values for use in various areas of the state. The *K* factor decreases with increasing development density in the areas served by the facility. In high-density areas, substantial demand during off-peak periods exists. This effectively lowers the proportion of traffic occurring during the peak hour of the day. The volume generated by high-density development is generally larger than that generated by lower-density areas. Thus, it is important to remember that a high proportion of traffic occurring in the peak hour does not suggest that the peak-hour volume itself is large.

The *D* factor tends to be more variable and is influenced by a number of factors. Again, as development density increases, the *D* factor tends to decrease. As density increases, it is more likely to have substantial bidirectional demands. Radial routes (i.e., those serving movements into and out of central cities or other areas of activity) will have stronger directional distributions (higher *D* values) than those that are circumferential, (i.e., going around areas of central activity). Table 5.2 indicates general ranges for *K* and *D* factors. These are purely illustrative; specific data on these characteristics should be available from state or local highway agencies or should be locally calibrated before application.

Table 5.2: General Ranges for *K* and *D* Factors

Facility Type	Normal Range of Values	
	K Factor	*D* Factor
Rural	0.15–0.25	0.65–0.80
Suburban	0.12–0.15	0.55–0.65
Urban:		
Radial Route	0.07–0.12	0.55–0.60
Circumferential Route	0.07–0.12	0.50–0.55

Table 5.2: Full Alternative Text

Consider the case of a rural highway that has a 20-year forecast of AADT of 30,000 veh/day. Based upon the data of Table 5.2, what range of directional design hour volumes might be expected for this situation? Using the values of Table 5.2 for a rural highway, the *K* factor ranges from 0.15 to 0.25, and the *D* factor ranges from 0.65 to 0.80. The range of directional design hour volumes is, therefore:

$$DDHVLOW=30{,}000\times 0.15\times 0.65=2{,}925\text{ veh/h}DDHVHIGH=30{,}000\times 0.2$$

The expected range in DDHV is quite large under these criteria. Thus, determining appropriate values of *K* and *D* for the facility in question is critical in making such a forecast.

This simple illustration points out the difficulty in projecting future traffic demands accurately. Not only does volume change over time, but the basic characteristics of volume variation may change as well. Accurate projections require the identification of causative relationships that remain stable over time. Such relationships are difficult to discern in the complexity of observed travel behavior. Stability of these relationships over time cannot be guaranteed in any event, making demand forecasting an approximate process at best.

Subhourly Volumes and Rates of Flow

While hourly traffic volumes form the basis for many forms of traffic design and analysis, the variation of traffic within a given hour is also of considerable interest. The quality of traffic flow is often related to short-term fluctuations in traffic demand. A facility may have sufficient capacity to serve the peak-hour demand, but short-term peaks of flow within the hour may exceed capacity and create a breakdown.

Volumes observed for periods of less than one hour are generally expressed as equivalent hourly rates of flow. For example, 1,000 vehicles counted over a 15 minute interval could be expressed as 1,000 veh/0.25 h = 4,000 veh/h. The rate of flow of 4,000 veh/h is valid for the 15 minute period in which the volume of 1,000 veh was observed. Table 5.3 illustrates the difference between volumes and rates of flow.

Table 5.3: Illustration of Volumes and Rates of Flow

Time Interval	Volume for Time Interval (veh)	Rate of Flow for Time Interval (veh/h)
5:00–5:15 PM	1,000	1,000 / 0.25 = 4,000
5:15–5:30 PM	1,100	1,100 / 0.25 = 4,400
5:30–5:45 PM	1,200	1,200 / 0.25 = 4,800
5:45–6:00 PM	900	900 / 0.25 = 3,600
5:00–6:00 PM	$\Sigma = 4{,}200$	

Table 5.3: Full Alternative Text

The full hourly volume is the sum of the four 15-minute volume observations, or 4,200 veh/h. The rate of flow for each 15-minute interval is the volume observed for that interval divided by the 0.25 hours over which it was observed. In the worst period of time, 5:30–5:45 PM, the rate of flow is 4,800 veh/h. This is a *flow rate*, not a volume. The actual volume for the hour is only 4,200 veh/h.

Consider the situation that would exist if the capacity of the location in question were exactly 4,200 veh/h. While this is sufficient to handle the full-hour demand indicated in Table 5.3, the demand *rate of flow* during two of the 15-minute periods noted (5:15–5:30 PM and 5:30–5:45 PM) exceeds the capacity. The problem is that while demand may vary within a given hour, capacity is constant. In each 15-minute period, the capacity is 4,200/4 or 1,050 veh. Thus, within the peak hour shown, queues will develop in the half-hour period between 5:15 and 5:45 PM, during which the demand exceeds the capacity. Further, although demand is less than capacity in the first 15-minute period (5:00–5:15 PM), the unused capacity cannot be used in a later period. Table 5.4 compares the demand and capacity for each of the 15-minute intervals. The queue at the end of each period can be computed as the queue at the beginning of the period plus the arriving vehicles minus the departing vehicles.

Table 5.4: Queuing Analysis for the Data of Table 5.3

Time Interval	Arriving Vehicles (veh)	Departing Vehicles (veh)	Queue Size at End of Period (veh)
5:00–5:15 PM	1,000	1,050	0
5:15–5:30 PM	1,100	1,050	$0 + 1{,}100 - 1{,}050 = 50$
5:30–5:45 PM	1,200	1,050	$50 + 1{,}200 - 1{,}050 = 200$
5:45–6:00 PM	900	1,050	$200 + 900 - 1{,}050 = 50$

Table 5.4: Full Alternative Text

Even though the capacity of this segment over the full hour is equal to the peak-hour demand volume (4,200 veh/h), at the end of the hour, there remains a queue of 50 vehicles that has not been served. While this illustration shows that a queue exists for three out of four 15-minute periods within the peak hour, the dynamics of queue clearance may continue to negatively affect traffic for far longer.

Because of these types of impacts, it is often necessary to design facilities and analyze traffic conditions for a period of maximum rate of flow within the peak hour. For most practical purposes, 15 minutes is considered to be the minimum period of time over which traffic conditions are statistically stable. Although rates of flow can be computed for any period of time and researchers often use rates for periods of 1 to 5 minutes, rates of flow for shorter periods often represent transient conditions that defy consistent mathematical representations. In recent years, however, use of 5-minute rates of flow has increased, and there is some thought that these might be sufficiently stable for use in design and analysis, particularly on uninterrupted flow facilities. Despite this, most standard design and analysis practices continue to use the 15-minute interval as a base period.

The relationship between the hourly volume and the maximum rate of flow within the hour is defined by the *peak-hour factor*, as follows:

$$PHF=\frac{\text{hourly volume}}{\text{max. rate of flow}}$$

For standard 15-minute analysis period, this becomes:

$$PHF=\frac{V}{4\times V_{m15}} \quad [5\text{-}2]$$

where:

V = hourly volume, veh,
V_{m15} = maximum 15-minute volume within the hour, veh, and
PHF = peak-hour factor.

For the illustrative data in Tables 5.3 and 5.4:

$$PHF=\frac{4200}{4\times 1200}=0.875$$

The maximum possible value for the PHF is 1.00, which occurs when the volume in each interval is constant. For 15-minute periods, each would have a volume of exactly one-quarter of the full-hour volume. This indicates a condition in which there is virtually no variation of flow within the hour. The minimum value occurs when the entire hourly volume occurs in a single 15-minute interval. In this case, the PHF becomes 0.25, and represents the most extreme case of volume variation within the hour. In practical terms, the PHF generally varies between a low of 0.70 for rural and sparsely developed areas to approximately 0.98 in dense urban areas.

The PHF is descriptive of trip generation patterns and may apply to an area or portion of a street and highway system. When the value is known, it can be used to estimate a maximum flow rate within an hour based on the full-hour volume:

v=VPHF [5-3]

where:

v=maximum rate of flow within the hour, veh/h,V=hourly volume, veh/h, a hour factor.

This conversion is frequently used in the techniques and methodologies covered throughout this text.

5.2.2 Speed and Travel Time

Speed is the second macroscopic parameter describing the state of a traffic stream. Speed is defined as a rate of motion in distance per unit time. Travel time is the time taken to traverse a defined section of roadway. Speed and travel time are inversely related:

S=dt [5-4]

where:

S=speed, mi/h or ft/s,d=distance traversed, mi or ft, andt=travel time to trav

In a moving traffic stream, each vehicle travels at a different speed. Thus, the traffic stream does not have a single characteristic value, but rather a distribution of individual speeds. The traffic stream, taken as a whole, can be characterized using an average or typical speed.

There are two ways in which an average speed for a traffic stream can be computed:

- *Time mean speed* (TMS). The average speed of all vehicles passing a point on a highway or lane over some specified time period.
- *Space mean speed* (SMS). The average speed of all vehicles occupying a given section of highway or lane over some specified time period.

In essence, TMS is a point measure, whereas SMS describes a length of highway or lane. Figure 5.1 shows an example illustrating the differences between the two average speed measures.

Figure 5.1: Time Mean Speed and Space Mean Speed Illustrated

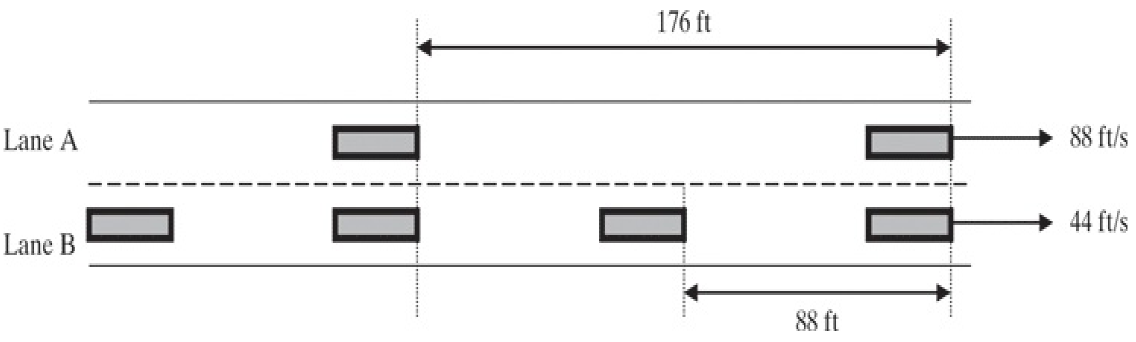


Figure 5.1: Full Alternative Text

To measure TMS, an observer would stand by the side of the road and record the speed of each vehicle as it passes. Given the speeds and the spacing shown in Figure 5.1, a vehicle will pass the observer in Lane A every 176 ft/88 ft/s = 2.0 s. Similarly, a vehicle will pass the observer in Lane B every 88 ft/44 ft/s = 2.0 s. Thus, as long as the traffic stream maintains the conditions shown, for every *n* vehicles traveling at 88 ft/s, the observer will also observe *n* vehicles traveling at 44 ft/s. The TMS may then be computed as:

TMS=88.0n + 44.0n2n=66.0ft/s

To measure SMS, an observer would need an elevated location from which the full extent of the section may be viewed. Again, however, as long as the traffic stream remains stable and uniform, as shown, there will be twice as many vehicles in Lane B as there are in Lane A. Therefore, the SMS is computed as:

$SMS=88\times n+44\times 2n3n=58.7mi/h$

In effect, SMS accounts for the fact that it takes a vehicle traveling at 44.0 ft/s twice as long to traverse the defined section as it does a vehicle traveling at 88.0 ft/s. The SMS weights slower vehicles more heavily in the average, based on the amount of time they occupy a highway section. Thus, the SMS is usually lower than the corresponding TMS, in which each vehicle is weighted equally. The two speed measures may theoretically be equal if all vehicles in the section are traveling at exactly the same speed.

Both the TMS and SMS may be computed from a series of measured travel times over a specified distance using the following relationships:

$TMS=\sum i(d/ti)n$ [5-5]
$SMS=d\sum i(ti/n)$ [5-6]

where:

TMS=time mean speed, ft/s,SMS=space mean speed, ft/s,d=distance traver

TMS is computed by finding each individual vehicle speed and taking a simple average of the results. *SMS* is computed by finding the average travel time for a vehicle to traverse the section and using the average travel time to compute a speed. Table 5.5 shows an example in the computation of TMS and SMS.

Table 5.5: Illustrative Computation of TMS and SMS

Vehicle No.	Distance *d* (ft)	Travel Time *t* (s)	Speed (ft/s)
1	1,000	18.0	1,000 / 18 = 55.6
2	1,000	20.0	1,000 / 20 = 50.0
3	1,000	22.0	1,000 / 22 = 45.5
4	1,000	19.0	1,000 / 19 = 52.6
5	1,000	20.0	1,000 / 20 = 50.0
6	1,000	20.0	1,000 / 20 = 50.0
Total	**6,000**	**119**	**303.7**
Average	**6,000/6 = 1,000**	**119/6 = 19.8**	**303.7/6 = 50.6**

TMS=50.6 ft/sSMS=1,000/19.8=50.4 ft/s

Table 5.5: Full Alternative Text

5.2.3 Density and Occupancy

Density

Density, the third primary measure of traffic stream characteristics, is defined as the number of vehicles occupying a given length of highway or lane, generally expressed as vehicles per mile or vehicles per mile per lane.

Density is difficult to measure directly, as an elevated vantage point from which the highway section under study may be observed is required. It is often computed from speed and flow rate measurements, as is discussed later in this chapter.

Density, however, is perhaps the most important of the three primary traffic stream parameters, because it is the measure most directly related to traffic demand. Demand does not occur as a rate of flow, even though traffic engineers use this parameter as the principal measure of demand. Traffic is generated from various land uses, injecting a number of vehicles into a confined roadway space. This process creates a density of vehicles. Drivers select speeds that are consistent with how close they are to other vehicles. The speed and density combine to give the observed rate of flow.

Density is also an important measure of the quality of traffic flow, as it is a measure of the proximity of other vehicles, a factor that influences freedom to maneuver and the psychological comfort of drivers.

Occupancy

Although density is difficult to measure directly, modern detectors can measure *occupancy*, which is a related parameter. Occupancy is defined as the proportion of time that a detector is "occupied," or covered, by a vehicle in a defined time period. Figure 5.2 illustrates density and occupancy.

Figure 5.2: Density and Occupancy Illustrated

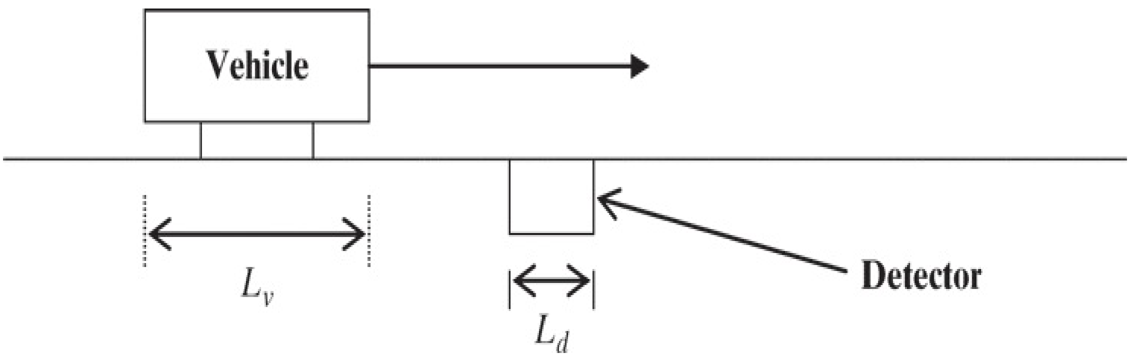


Figure 5.2: Full Alternative Text

In Figure 5.2, Lv is the average length of a vehicle (ft), while Ld is the length of the detector (which is normally a magnetic loop detector). If "occupancy" over a given detector is "*O*," then density may be computed as:

$D=5,280 \times OLv + Ld$ [5-7]

The lengths of the average vehicle and the detector are added, as the detector is generally activated as the front bumper engages the front boundary of the detector and is deactivated when the rear bumper clears the back boundary of the detector. Note that the measure of occupancy, *O*, is expressed as a decimal representing the portion of time that the detector is covered by a vehicle.

Consider a case in which a detector records an occupancy of 0.200 for a 15-minute analysis period. If the average length of a vehicle is 28 ft, and the detector is 3 ft long, what is the density?

$D=5,280 \times 0.20028 + 3=34.1$ veh/mi/ln

The occupancy is measured for a specific detector in a specific lane. Thus, the density estimated from occupancy is in units of vehicles per mile per lane. If there are adjacent detectors in additional lanes, the density in each lane may be summed to provide a density in veh/mi for a given direction of flow over several lanes.

5.2.4 Spacing and Headway: Microscopic Parameters

While flow, speed, and density represent macroscopic descriptors for the entire traffic stream, they can be related to microscopic parameters that describe individual vehicles within the traffic stream, or specific pairs of vehicles within the traffic stream.

Spacing

Spacing is defined as the distance between successive vehicles in a traffic lane, measured from some common reference point on the vehicles, such as the front bumper or front wheels. The *average* spacing in a traffic lane can be directly related to the density of the lane:

$D=5,280da$ [5-8]

where:

D=density, veh/mi/ln, andda=average spacing between vehicles in thelane,

Headway

Headway is defined as the time interval between successive vehicles as they pass a point along the lane, also measured between common reference points on the vehicles. The *average* headway in a lane is directly related to the rate of flow:

v=3,600ha [5-9]

where:

v=rate of flow, veh/h/ln, andha=average headway in the lane, s.

Use of Microscopic Measures

Microscopic measures are useful for many traffic analysis purposes. Because a spacing and/or a headway may be obtained for every pair of vehicles, the amount of data that can be collected in a short period of time is relatively large. A traffic stream with a volume of 1,000 veh over a 15-minute time period results in a *single* value of rate of flow, SMS, and density when observed. There would be, however, 1,000 headway and spacing measurements, assuming that all vehicle pairs were observed.

Use of microscopic measures also allows various vehicle types to be isolated in the traffic stream. Passenger car flows and densities, for example, could be derived from isolating spacing and headway for pairs of passenger cars following each other. Heavy vehicles could be similarly isolated and studied for their specific characteristics. There are some practical flaws in this approach. The nearby presence of heavy vehicles (even when not adjacent) might influence the behavior of individual passenger cars in the traffic stream.

Average speed can also be computed from headway and spacing measurements as:

S=daha [5-10]

where:

S = average speed, ft/s, d_a = average spacing, ft, and h_a = average headway, s.

Example

Traffic in a congested multilane highway lane is observed to have an average spacing of 200 ft, and an average headway of 3.8 s. Estimate the rate of flow, density and speed of traffic in this lane.

The solution is found using Equations 5-8 through 5-10:

$$v = \frac{3{,}600}{3.8} = 947 \text{ veh/h/ln} \quad D = \frac{5{,}280}{200} = 26.4 \text{ veh/mi/ln} \quad S = 0.68\ (200/3.8) = 3$$

5.3 Relationships among Flow Rate, Speed, and Density

The three macroscopic measures of the state of a given traffic stream—flow, speed, and density—are related as follows:

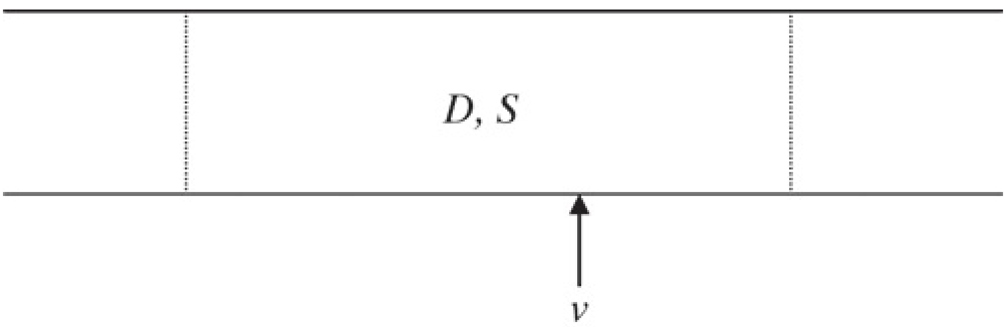
$$v = S \times D \quad [5\text{-}11]$$

where:

v = rate of flow, veh/h or veh/h/ln, S = space mean speed, mi/h, and D = density

SMS and density are measures that refer to a specific *section* of a lane or highway, while flow rate is a point measure. Figure 5.3 illustrates the relationship. The SMS and density measures must apply to the same defined section of roadway. Under stable flow conditions (i.e., the flow entering and leaving the section are the same; no queues are forming within the section), the rate of flow computed by Equation 5-11 applies to *any* point within the section. Where unstable operations exist (a queue is forming within the section), the computed flow rate represents an average for all points within the section.

Figure 5.3: Traffic Stream Parameters Illustrated



If a freeway lane were observed to have a SMS of 55 mi/h and a density of 25 veh/mi/ln, the flow rate in the lane could be estimated as:

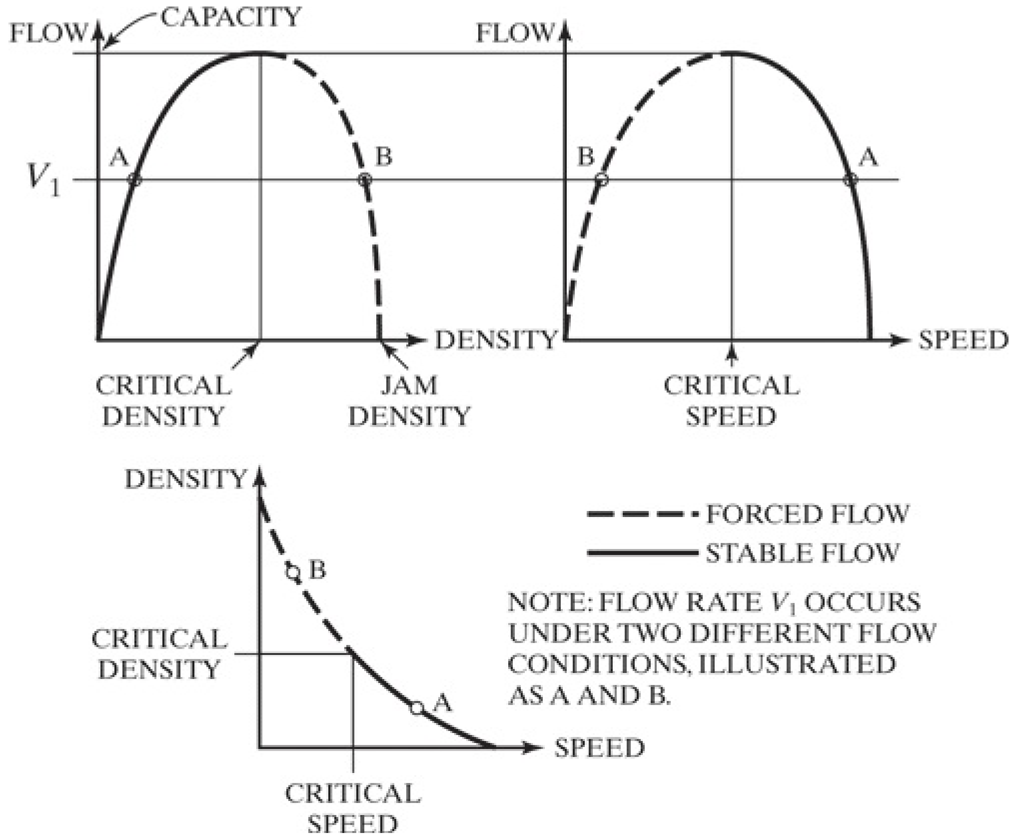
$v=55 \times 25=1,375$ veh/h/ln

As noted previously, this relationship is most often used to estimate density, which is difficult to measure directly, from measured values of flow rate and SMS. Consider a freeway lane with a measured SMS of 60 mi/h and a flow rate of 1,000 veh/h/ln. The density could be estimated from Equation 5-11 as:

$D=vS=1,00060=16.7$ veh/mi/ln

Equation 5-11 suggests that a given rate of flow (*v*) could be achieved by an infinite number of speed (*S*) and density (*D*) pairs having the same product. Thankfully, this is not what happens, as it would make the mathematical interpretation of traffic flow unintelligible. There are additional relationships between pairs of these variables that restrict the number of combinations that can and do occur in the field. Figure 5.4 illustrates the general form of these relationships. The exact shape and calibration of these relationships depend upon prevailing conditions, which vary from location to location and even over time at the same location.

Figure 5.4: Relationships among Speed, Flow, and Density



(Source: Reprinted with permission from Transportation Research Board, National Research Council, from *Highway Capacity Manual*, 3rd Edition, *Special Report 209*, pg 1–7, © 1994 by the National Academy of Sciences, Courtesy of the National Academies Press, Washington, D.C.)

Full Alternative Text

Note that a flow rate of “0 veh/h” occurs under two very different conditions. When there are no vehicles on the highway, flow is “0 veh/h” as no vehicles can be observed passing a point. Under this condition, speed is unmeasurable and is referred to as “free-flow speed,” a theoretical value that exists as a mathematical extension of the relationship between speed and flow (or speed and density). In practical terms, free-flow speed can be thought of as the speed a single vehicle could achieve when there are no other vehicles on the road and the motorist is driving as fast as is practicable given the geometry of the highway and its environmental surroundings.

A flow of “0 veh/h” also occurs when there are so many vehicles on the road that all motion stops. This occurs at a very high density, called the “jam density,” and no flow is observed, as no vehicle can pass a point to be counted when all vehicles are stopped.

Between these two extreme points on the relationships, there is a peaking characteristic. The peak of the flow-speed and flow-density curves is the maximum rate of flow, or the *capacity* of the roadway. Its value, like everything else about these relationships, depends upon the specific prevailing conditions at the time and location of the calibration measurements.

Operation at capacity, however, is very unstable. At capacity, with no usable gaps in the traffic stream, the slightest perturbation caused by an entering or lane-changing vehicle, or simply a driver hitting the brakes, causes a chain reaction that cannot be damped. The perturbation propagates upstream and continues until sufficient gaps in the traffic stream allow the event to be effectively dissipated.

The dashed portion of the curves represents *unstable* or *forced* flow. This effectively represents flow within a queue that has formed behind a breakdown location. A breakdown will occur at any point where the arriving flow rate exceeds the downstream capacity of the facility. Common points for such breakdowns include on-ramps on freeways, but accidents and incidents are also common, less predictable causes for the formation of queues. The solid line portion of the curves represents *stable* flow (i.e., moving traffic streams that can be maintained over a period of time).

Except for capacity flow, any flow rate may exist under two conditions:

1. A condition of relatively high speed and low density (on the stable portion of flow relationships)

2. A condition of relatively low speed and high density (on the unstable portion of flow relationships)

Obviously, traffic engineers would prefer to keep all facilities operating on the stable side of the curves.

Because a given volume or flow rate may occur under two very different sets of operating conditions, volume alone cannot completely describe flow conditions, nor can it be used as a measure of the quality of traffic flow. Values of speed and/or density, however, would define unique points on any of the relationships of Figure 5.4, and both describe aspects of quality that can be perceived by drivers and passengers.

The curves depicted in Figure 5.4 are generic. As noted, specific characteristics of such curves vary considerably depending upon the local prevailing conditions at the study site. While the speed–density relationship is the curve that most directly describes *driver behavior*—that is, drivers selecting an appropriate and safe speed for the densities they encounter—the curve that is most often calibrated is the speed–flow curve. This is because speeds and flow rates are more easily measured directly.

Figure 5.5 shows a generic speed–flow curve that represents the general form of such curves on modern freeway facilities (uninterrupted flow).

Figure 5.5: Speed–Flow Characteristics on a Modern Freeway

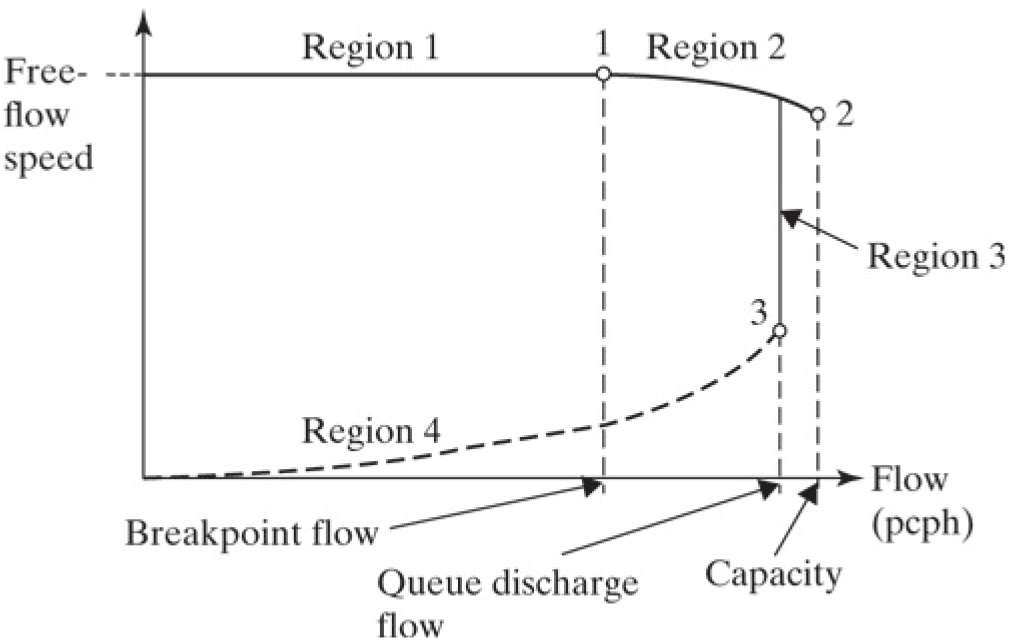


Figure 5.5: Full Alternative Text

Drivers on modern freeways (and other facilities as well) have become more aggressive over time. On freeways, this manifests itself in Region 1 of Figure 5.5. The free-flow speed is maintained over a broad range of flow rates, until a breakpoint is reached. In many cases, this breakpoint will be in the range of 1,000 to 1,600 pc/h/ln. Thus, average speeds on freeways are unaffected by flow levels until something in this range is reached. Beyond this breakpoint, speeds decline with further increases in flow rate. The decline, however, is not severe. Freeway capacities are often observed at average speeds of 50–60 mi/h. Thus, the decline in average speed when flow rates rise from the breakpoint to capacity may be as small as 5–10 mi/h. This is very different from what was observed in the 1950s and 1960s, when capacity usually occurred at average speeds of 30 mi/h.

Capacity is the flow rate at the end point of Region 2 of Figure 5.5. Any demand in excess of capacity forces a breakdown in flow and the formation of queues. Within the queue (after its formation), the speed–flow relationship is shown as Region 4. All flow in this range is highly unstable, with high short-term variability.

Region 3 is referred to as “queue discharge.” It is shown as a vertical line in Figure 5.5, but it is more of a broad range of points that approximates the line shown. It represents the average flow rate of vehicles leaving the queue. It is generally thought that the queue discharge flow rate is approximately 5% less than the capacity before breakdown, although a range of values have been observed in the field.

The generic curve explains what most drivers experience when a breakdown occurs: they are traveling at a relatively high speed, begin to slow down slightly, and suddenly hit the “brick wall” as a queue rapidly forms. Speeds drop precipitously from capacity (50–60 mi/h) to that experienced in the queue, which can be as low as 10–20 mi/h.

5.4 A Brief History of Mathematical Models of Freeway Flow—Traffic Flow Theory

Traffic flow theory is best defined as mathematical models that attempt to relate characteristics of traffic movement to each other and to underlying traffic parameters. The science of traffic flow theory formally began with the work of Bruce Greenshields and the Yale Bureau of Highway Traffic in the 1930s. The field continued to develop and plays an important role in traffic engineering.

Virtually every function in traffic engineering, from data collection and analysis to signal timing and to capacity and level of service analysis, utilizes analytical models of traffic behavior under a variety of underlying circumstances. These models, and their development and calibration, are the essence of traffic flow theory.

This section provides a very brief glimpse into this field. References [1] to [4] provide excellent sources of comprehensive material on modern traffic flow theory. Reference [1] in particular has chapters that address:

1. Introduction to traffic flow theory
2. Traffic stream characteristics
3. Human factors
4. Car following
5. Continuum flow models
6. Macroscopic flow models
7. Traffic impact models
8. Unsignalized intersections

9. Signalized intersections
10. Traffic simulation

All of the methodologies related to the analysis of freeway, multilane highway, and two-lane rural highway capacity and level of service analysis are based upon the fundamental relationships between the speed, flow, and density of an uninterrupted traffic stream, as described in this chapter. Chapter 6 of Reference [5] provides a comprehensive review of the development of speed–flow–density curves on uninterrupted flow facilities.

5.4.1 Historical Background

The earliest studies of uninterrupted flow characteristics and relationships were conducted by Bruce Greenshields [6]. His and other early studies focused on the relationship between the density and speed of an uninterrupted traffic stream. Greenshields postulated that the speed-density curve was linear.

Later, Ellis [7] investigated two- and three-segment linear curves with discontinuities. Greenberg [8] hypothesized a logarithmic curve for speed–density, whereas Underwood [9] used an exponential form. Edie [10] combined logarithmic and exponential forms for low- and high-density portions of the curve. Like Ellis, Edie's curves contained discontinuities. May [11] suggested using a bell-shaped curve.

Over the years, there have been many suggestions for mathematical descriptions of the relationships between speed, flow, and density on an uninterrupted flow facility. Although there have clearly been changes in driver behavior that influence the shape of these curves, there is no one form that will best fit data from all locations.

5.4.2 Deriving Speed–Flow and Density–Flow Curves from a Speed–Density Curve

Because Equation 5-11 is a fundamental relationship governing speed, flow, and density, once the relationship between speed and density is established, then speed–flow and speed–density curves are also fully determined.

Using Greenshields's simple linear speed–density curve as an example, consider the speed–density relationship shown in Figure 5.6. Two points of interest are the Y- and X-axis intercepts. The Y-intercept is 65.0 mi/h, and is called the "free-flow speed," that is, the speed that occurs when density (and therefore, flow) is zero. The X-intercept is 110 veh/mi/ln, the density at which all motion stops, making the speed zero. This is commonly called the "jam" density.

Figure 5.6: Sample Linear Speed–Density Relationship: Greenshields Model

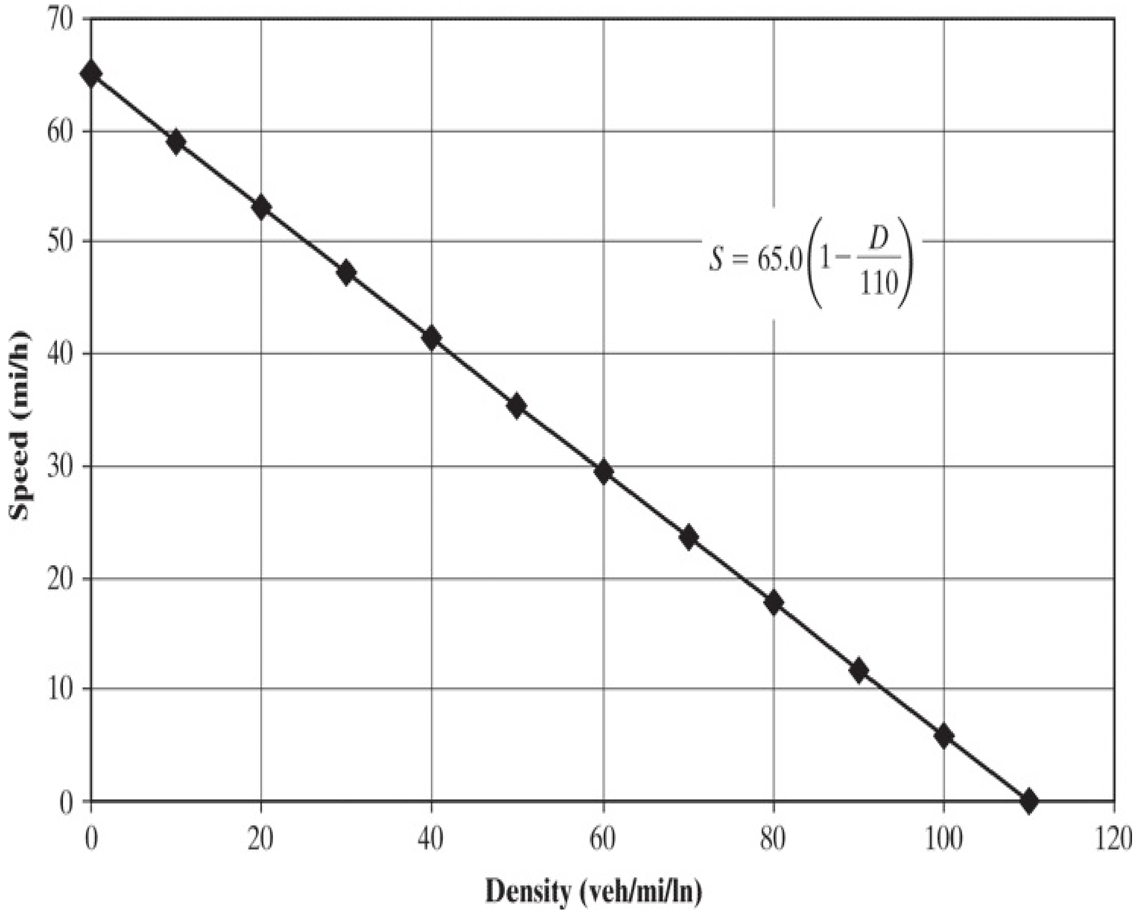


Figure 5.6: Full Alternative Text

Given the equation for speed versus density, and knowing that $v=S\times D$ (Equation 5-11) is always applicable, the relationship between flow and density is found by substituting $S=v/D$ in the speed–density equation. Then:

$$S=65.0\,\left(1-\frac{D}{110}\right)\quad \frac{v}{D}=65.0\,\left(1-\frac{D}{110}\right)\quad v=65\,D-0.59091\,D^2$$

The relationship between flow and speed is found by substituting $D=v/S$ in the speed-density equation. Then:

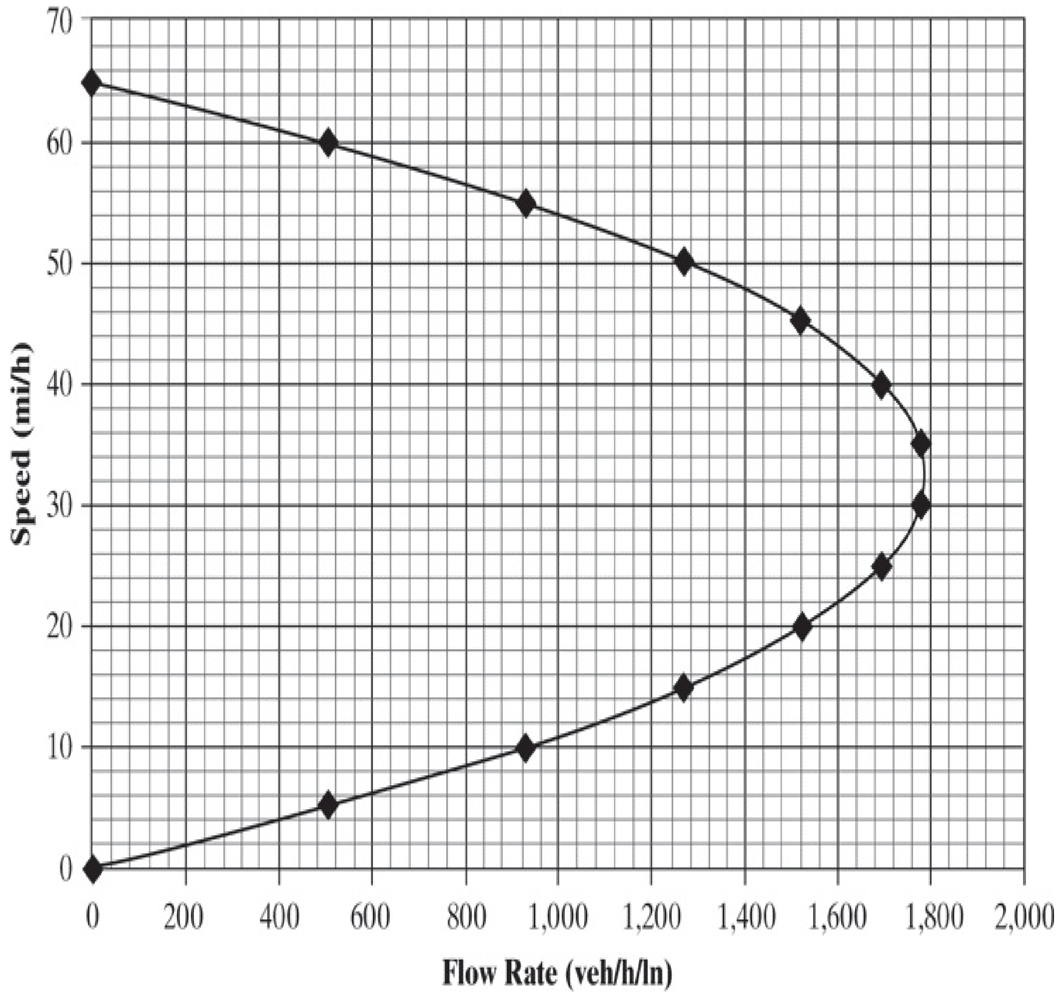
$$S=65.0\,\left(1-\frac{D}{110}\right)\quad S=65.0\,\left(1-\frac{v/S}{110}\right)\quad v=110\,S-1.6923\,S^2$$

As shown in Figure 5.7, both of these curves are parabolic.

Figure 5.7: Flow–Density and Speed–Flow Curves Resulting from a Linear Speed–Density Relationship: Greenshields Model

(a) Flow–Density Curve Resulting from Linear Speed–Density Relationship

5.4-6 Full Alternative Text



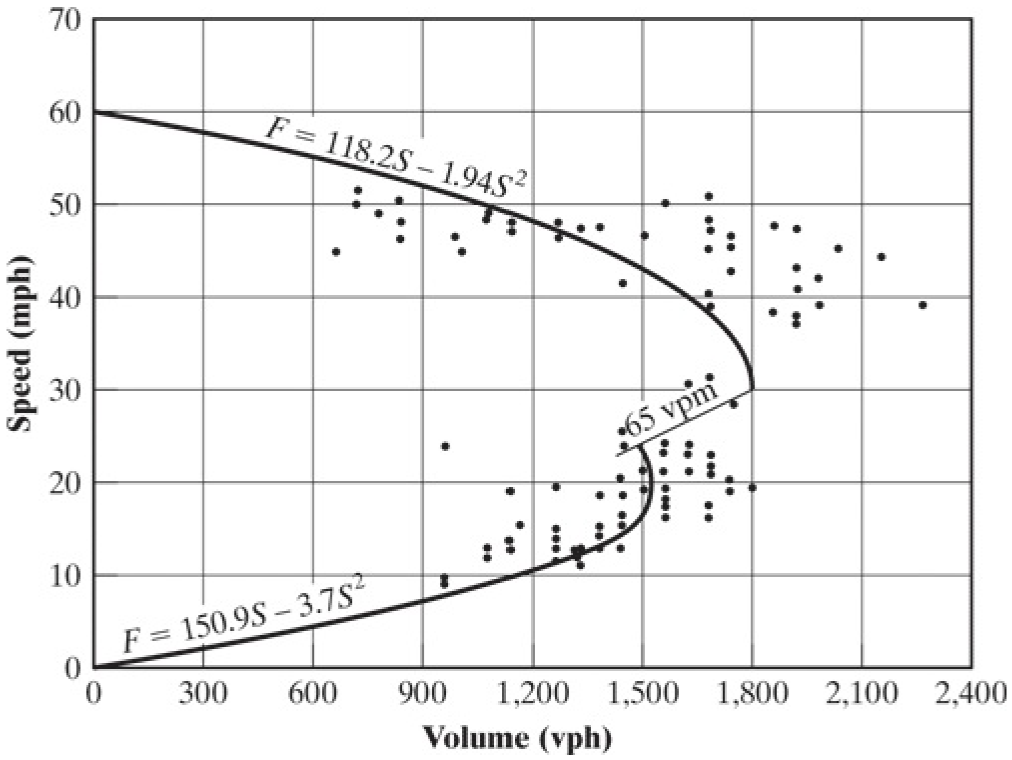
(b) Speed–Flow Curve Resulting from Linear Speed–Density Relationship

5.4-6 Full Alternative Text

Equation 5-11 (v=S×D) always applies. Therefore, calibrating any *one* of the relationships between *S* and *D*, *v* and *D*, or *v* and *S* defines all three relationships. Given one of the three, the other two may be algebraically derived.

Figure 5.8 illustrates a speed–flow curve resulting from a two-segment linear speed–density relationship. It results in *two* parabolas, one for each segment of the discontinuous speed–density curve. It is illustrated because of the discontinuity involved. From the speed–flow curve, it is clear that the discontinuity is near the peak of the curve(s)—that is, in the vicinity of capacity.

Figure 5.8: A Speed–Flow Curve with Discontinuity in the Vicinity of Capacity



(Source: Reprinted with permission from Transportation Research Board, National Research Council, J.S. Drake, J.L, Schofer, and A.D. May Jr., “A Statistical Analysis of Speed-Density Hypotheses,” *Transportation Research Record 154*, pg 78, 1967. © 1967 by the National Academy of Sciences, Courtesy of the National Academies Press, Washington, D.C.)

Figure 5.8: Full Alternative Text

The graph is from an old, but fascinating study in which various mathematical forms were fit to a set of data from the Eisenhower Expressway in Chicago in the early 1960s. While the data are not reflective of modern speed–flow behavior on freeways, the study determined that a discontinuous set of curves (not the one shown) best fit the data. Of interest is that Figure 5.7 seems to indicate that there are *two* capacities: one when approached from low speeds (unstable flow), and one when approached from high speeds (stable flow). This characteristic is quite complicated, and is discussed further in the sections that follow.

5.4.3 Determining Capacity from Speed–Flow–Density Relationships

Chapters 7 and 22 include detailed discussions of the concept of capacity and many of the nuances that it contains. One potential understanding of capacity is that capacity is the peak of a speed–flow or flow–density curve. From Figure 5.6, for example, it is clear that the "peak" of either curve occurs at a flow rate slightly less than 1,800 veh/h/ln (hard to read exact number from the scale shown). In Figure 5.7, there are two capacities. The "high" value is also about 1,800 veh/h/ln, which is on the high-speed or stable portion of the curve. The "low" value is approximately 1,550 veh/h/ln, which is on the low-speed or unstable side of the curve.

The capacity value can also be determined mathematically. Using the curves of Figure 5.6 as an example, it is necessary to determine the speed and density at which capacity occurs. In both cases, this occurs where the slope of the curve (or the first derivative of the curve) is zero.

For the flow–density curve:

$$v = 65.0\,D - 0.59091\,D^2 \qquad \frac{dv}{dD} = 0 = 65.0 - 1.18182\,D \qquad D = \frac{65.0}{1.18182} = 55.0\ \text{veh/h/ln}$$

For the speed–flow curve:

$$v = 110S - 1.6923\,S^2 \qquad \frac{dv}{dS} = 0 = 110 - 3.3846\,S \qquad S = \frac{110}{3.3846} = 32.5\ \text{mi/h}$$

The calculus and algebra confirm the obvious: for the linear model of Figures 5.5 and 5.6, capacity occurs when speed is exactly half the free-flow speed and when density is exactly half the jam density. The capacity is then found as the product of the speed and density at which it occurs, or:

$$c = S \times D = 32.5 \times 55.0 = 1{,}788\ \text{veh/h/ln}$$

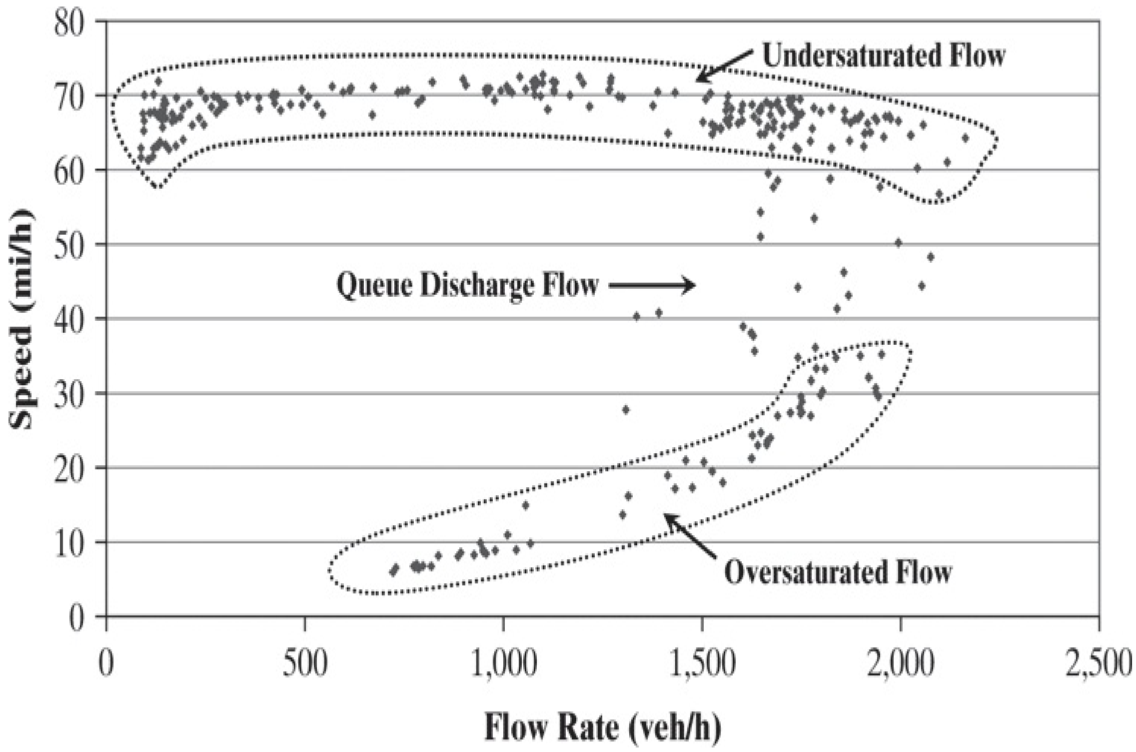
which confirms the observation from Figure 5.6 of "slightly less than 1,800 veh/h/ln."

5.4.4 Modern Uninterrupted Flow Characteristics

“Traffic flow theory” is really a misnomer. Traffic flow does not occur in theory. It occurs on real streets, highways, and freeways all over the world. The mathematical models that are developed by researchers are merely descriptions of driver behavior. Because of this, traffic flow theory is an evolving science. No model is ever static, because driver behavior changes over time. Nowhere is this clearer than in speed–flow–density relationships for uninterrupted flow.

The linear model of Greenshields and most of the other historic models discussed all have one common characteristic: speeds decline as flow rates increase. Drivers react to higher densities (which result in higher flows) by slowing down to maintain what they perceive to be safe operations. Modern uninterrupted flow, particularly on freeways, does not reflect this characteristic. In fact, drivers maintain high average speeds through a range of flow rates, and do not slow down until relatively high flow rates are reached. Figure 5.9 illustrates the general characteristics of uninterrupted flow on a modern freeway.

Figure 5.9: Speed–Flow Characteristics for a Modern Freeway



(Source: Reprinted with permission from *Highway Capacity Manual*, 6th Edition: *A Guide for Multimodal Mobility Analysis*, Transportation Research Board, © 2016 by the National Academy of Sciences, Courtesy of the National Academies Press, Washington, D.C.)

Figure 5.9: Full Alternative Text

Figure 5.9 shows three distinct ranges of data: (1) undersaturated (stable) flow, (2) queue discharge flow, and (3) oversaturated (unstable) flow. The speed throughout the undersaturated flow portion of the curve is remarkably stable. If a line were drawn through the center of these points, the speed would range from about 71 mi/h to a low of about 60 mi/h. Further, there seems to be no systematic decline in speeds with flow rate until a flow rate of approximately 1,200 to 1,300 veh/h/ln is reached. The capacity would be the peak of this portion of the curve, or approximately 2,200 veh/h/ln, which is achieved at an astonishingly high speed of approximately 60 mi/h.

Once capacity is reached and demand in fact exceeds capacity, a queue begins to form. The “queue discharge” portion of the curve reflects vehicles departing from the front of the queue. Such vehicles will begin to accelerate as they move downstream, assuming that no additional downstream congestion exists. The oversaturated portion of the curve is what exists within the queue that forms when demand exceeds capacity at a point.

While queue discharge rates vary widely at this site, their average is clearly *lower* than the capacity of the undersaturated portion of the curve. It is generally agreed that vehicles cannot depart the head of a queue at the same rate as they pass the same point under stable or undersaturated flow. Is this simply another explanation of the "two capacity" phenomenon of historic curves?

5.4.5 Calibrating a Speed–Flow–Density Relationship

How should data be collected to calibrate a speed–flow–density relationship for a specific uninterrupted flow segment? One of the problems involved in interpreting older studies is that it is not clear how or, more importantly, where data was collected.

While the speed–density relationship is the most descriptive of driver behavior, measuring density in the field is not always a simple task. Speed and flow rate or volume, however, are relatively simple traffic measurements. Most field studies, therefore, focus on calibration of the speed versus flow relationship, and derive the others.

If the capacity operation is to be observed, measurements must be taken near a point of frequent congestion. Most of these occur at on-ramps, where the arriving freeway and arriving on-ramp flows may regularly exceed the capacity of the downstream freeway segment. Under these conditions, queues may be expected to form on both the upstream freeway and the ramp roadway. Figure 5.10 illustrates a field setup for taking data to calibrate all regions of the curve.

Figure 5.10: Typical Setup for a Speed–Flow Calibration

Study

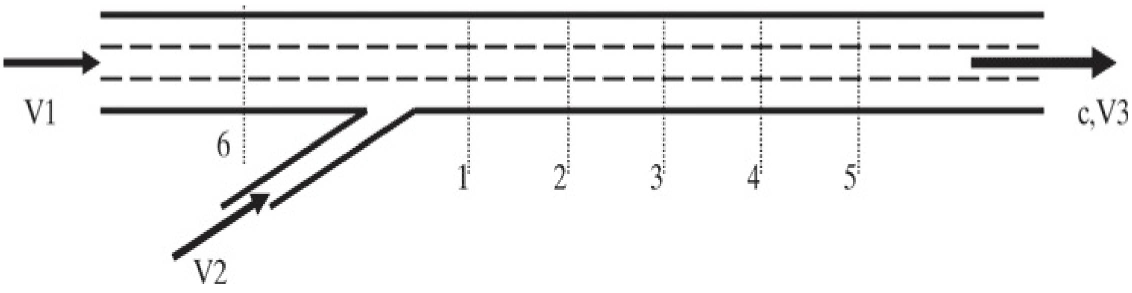


Figure 5.10: Full Alternative Text

Under stable flow, flow rates and speeds would be recorded at a point close to, but sufficiently downstream from, the merge for ramp vehicles to have accelerated to ambient speed. This is indicated as location 1. The measurements *must* take place downstream of the on-ramp, as V2 is part of the downstream demand.

Once queues start forming, stable flow no longer exists. Now, observation locations must shift. As unstable or oversaturated flow exists within the queue forming behind the on-ramp, observations must be made from within the queue, indicated here as location 6.

Downstream of the head of the queue, indicated here as locations 2 through 5 (perhaps including 1 depending upon its exact placement), discharging vehicles can be observed. Assuming no additional downstream congestion affecting the study area, the flow rate at these downstream locations will be fairly stable, whereas the speed increases as vehicles get further away from the head of the queue. Measurements at these locations can be combined to calibrate the "queue discharge" portion of the curve.

These are not simple observations. Care must be taken to avoid the observation of impacts of unseen downstream congestion. Capacity operations are most likely to exist in the last 15-minute intervals before the appearance of queues on the ramp and/or the freeway.

5.4.6 Curve Fitting

Once data have been collected, reduced, and recorded, a mathematical description of the data is sought. There are a variety of statistical tools available to accomplish this. Multiple linear and nonlinear regression techniques and software packages are used in the curve-fitting process.

Most of these tools define the “best” fit using an objective function, which the tool seeks to minimize. A common objective function is to minimize the sum of the squared differences between the actual data points and the curve that defines the relationship. The curve, in effect, represents predicted values of the independent variable (in this case speed), which are compared directly to the field-measured values to determine how “good” the fit is.

In some cases, there is not enough data available for formal regression analysis, or the spread of data is so broad as to complicate regression analysis. In such cases, graphic fits using the analyst’s best professional judgment are done to define the curve and determine its equation.

There are many statistics texts that provide detailed treatments of regression and multiple regression analysis. One of the most frequently used software packages for regression analysis is SPSS—Statistical Package for the Social Sciences. Despite its title, it is an excellent package and is used by many engineers who must do formal curve fitting. Another commonly used package is Statgraphics.

5.5 Characteristics of Interrupted Flow

The key feature of interrupted flow is the cyclical stopping and restarting of traffic streams at traffic signals, and at STOP or YIELD signs.

When traveling along a signalized street or arterial, platoons form, as groups of vehicles proceed in a manner that allows them to move continuously through a number of signals. Within platoons, many of the same characteristics as for uninterrupted traffic streams exist. It is the dynamics of starting and stopping groups of vehicles that adds complexity.

Fundamental concepts of interrupted flow are treated in Chapter 18, Principles of Intersection Signalization.

5.6 Closing Comments

This chapter has introduced the key macroscopic and microscopic parameters that are used to quantify and describe conditions within an uninterrupted traffic stream, and the fundamental relationships that govern them. Chapter 6 discusses the critical conceptual differences between volume (or rate of flow), capacity, and demand, all of which are generally quantified in the same units.

Like any engineering field, good traffic engineers must understand the medium with which they work. The medium for traffic engineers is traffic streams. Describing them in quantitatively precise terms is critical to the tasks of accommodating and controlling them in such a way as to provide for safe and efficient transportation for people and goods. Thus, the foundation of the profession lies in these descriptors.

References

- 1. Traffic Flow Theory – A State of the Art Report, Oak Ridge National Laboratory, Federal Highway Administration, Washington, D.C., 1999, as revised in 2001 (available on line at www.fhwa.gov).

- 2. “Traffic Flow Theory 2006,” Transportation Research Record 1965, Transportation Research Board, Washington, D.C., 2006 (22 papers).

- 3. “Traffic Flow Theory and Characteristics 2008,” Transportation Research Record 2088, Transportation Research Board, Washington, D.C., 2008 (23 papers).

- 4. “Traffic Flow Theory 2007,” Transportation Research Record 1999, Transportation Research Board, Washington, D.C., 2007 (22 papers).

- 5. Roess, R., and Prassas, E., The Highway Capacity Manual: A Conceptual and Research History, Vol 1, Uninterrupted Flow, Springer, Heidelberg, Germany, 2014.

- 6. Greenshields, B., “A Study of Highway Capacity,” Proceedings of the Highway Research Board, Vol. 14, Transportation Research Board, National Research Council, Washington, D.C., 1934.

- 7. Ellis, R., “Analysis of Linear Relationships in Speed-Density and Speed-Occupancy Curves,” Final Report, Northwestern University, Evanston, IL, December 1964.

- 8. Greenberg, H., “An Analysis of Traffic Flows,” Operations Research, Vol. 7, ORSA, Washington, D.C., 1959.

- 9. Underwood, R., “Speed, Volume, and Density Relationships,” Quality and Theory of Traffic Flow, Yale Bureau of Highway Traffic, Yale University, New Haven, CT, 1961.

- 10. Edie, L., “Car-Following and Steady-State Theory for Non-Congested Traffic,” Operations Research, Vol. 9, ORSA, Washington, D.C., 1961.

- 11. Duke, J., Schofer, J., and May Jr., A., “A Statistical Analysis of Speed-Density Hypotheses,” Highway Research Record 154, Transportation Research Board, National Research Council, Washington, D.C., 1967.

Problems

1. 5-1. A traffic stream (in a single lane) is observed to have average headways of 2.6 s/veh and average spacing of 235 ft between vehicles. For this lane:

    1. What is the flow rate?

    2. What is the density?

    3. What is the average speed?

2. 5-2. During the peak 15-minute period of the peak hour, the flow rate on a freeway (in one direction) is observed to be 5,600 veh/h. Determine the hourly volume if the peak-hour factor (PHF) is (a) 0.85, (b) 0.90, and (c) 0.95.

3. 5-3. The following traffic count data were taken from a permanent detector location on a major state highway.

1. Month	2. No. of Weekdays in Month (days)	3. Total Days in Month (days)	4. Total Monthly Volume (veh)	5. Total Weekday Volume (veh)
Jan	22	31	120,000	70,000
Feb	20	28	115,000	60,000
Mar	22	31	125,000	75,000
Apr	22	30	130,000	78,000
May	21	31	135,000	85,000
Jun	22	30	140,000	85,000
Jul	23	31	150,000	88,000
Aug	21	31	135,000	80,000
Sep	22	30	120,000	72,000
Oct	22	31	112,000	62,000
Nov	21	30	105,000	55,000
Dec	22	31	99,000	50,000

5.2-6 Full Alternative Text

From this data, determine (a) the AADT, (b) the ADT for each month, (c) the AAWT, and (d) the AWT for each month. From this information, what can be discerned about the character of the facility and the demand it serves?

4. 5-4. A freeway detector records an occupancy of 0.15 for a 15-minute period. If the detector is 6 ft long, and the average vehicle has a length of 20 ft, what is the density implied by this measurement?

5. 5-5. The following counts were taken on a major arterial during the evening peak period:

Time Period	Volume (veh)
4:00–4:15 PM	300
4:15–4:30 PM	325
4:30–4:45 PM	340
4:45–5:00 PM	360
5:00–5:15 PM	330
5:15–5:30 PM	310
5:30–5:45 PM	280
5:45–6:00 PM	240

5.2-7 Full Alternative Text

From this data, determine the following:

1. The peak hour
2. The peak-hour volume
3. The peak flow rate within the peak hour
4. The peak-hour factor (PHF)

6. 5-6. The flow rate on an arterial is 1,800 veh/h, evenly distributed over two lanes. If the average speed in these lanes is 40 mi/h, what is the density in veh/h/ln?

7. 5-7. The AADT for a section of suburban freeway is 150,000 veh/day. Assuming that this is an urban radial facility, what range of directional design hour volumes would be expected?

8. 5-8. The following travel times were measured for vehicles traversing a 2,000 ft segment of an arterial:

Vehicle	Travel Time (s)
1	40.5
2	44.2
3	41.7
4	47.3
5	46.5
6	41.9
7	43.0
8	47.0
9	42.6
10	43.3

5.2-8 Full Alternative Text

Determine the time mean speed (TMS) and space mean speed (SMS) for these vehicles.

9. 5-9. A peak-hour volume of 1,200 veh/h is observed on a freeway lane. What is the peak flow rate within this hour if the PHF is 0.87?

10. 5-10. The flow rate on an arterial lane is 1,300 veh/h. If the average speed in the same lane is 35 mi/h, what is the density?

11. 5-11. A study of speed–flow–density relationships at a particular freeway site has resulted in the following calibrated relationship:

    $S=71.2\,(1-D122)$

    1. Determine the free-flow speed and jam density for the relationship.
    2. Derive the equations for speed versus flow and flow versus density for this site.
    3. Determine the capacity of the segment mathematically.

Chapter 6 The Concepts of Demand, Volume, and Capacity

In Chapter 5, the fundamental parameters used to quantify traffic streams were defined and discussed. Traffic volume is most frequently used in traffic engineering, as it measures the quantity of traffic moving past a point or segment of a traffic facility. It is stated in units of vehicles per hour or vehicles per hour per lane for operational purposes.

Those units, however, are also used to quantify two other significant factors: the capacity of a roadway or lane and the traffic demand that exists for its use. While the measurement units are the same (all described as "veh/h"), the three are very different, and the traffic engineer must always be cognizant of the differences and the complex relationships among the three.

In simple terms, the three measures may be defined as follows:

- Traffic Demand: The number of vehicles that desire to pass a point or segment of a roadway in an hour, or expressed as an hourly rate in veh/h or persons/h.
- Traffic Volume: The number of vehicles that actually pass a point or segment of a roadway in an hour, or expressed as an hourly rate in veh/h or persons/h.
- Capacity: The maximum volume (or rate of flow) that a particular point or segment of a facility can accommodate in veh/h or persons/h.

Traffic demand reflects the travel desires of road users (or users of other transportation facilities). Capacity reflects the ability of existing or projected facilities to handle traffic. Traffic volume is what actually occurs at a point or on a segment of a facility.

When traffic demand is less than the capacity of facilities to accommodate it, observed volumes will be equal to the demand. However, when the capacity of facilities is insufficient to handle the traffic demand, observed volumes will be *less* than the demand, causing immediate and long-term changes in travel patterns.

6.1 When Capacity Constrains Demand

The principal constraint on demand is congestion. When roads are congested, drivers make a variety of alterations to their travel patterns to avoid the delay and stress that congestion causes. When presented with congestion, drivers and other travelers can change their intended travel by doing the following:

1. Diverting to other routes: Drivers and other travelers can select another route to go from A to B to avoid the congestion on the desired route.

2. Diverting travel to another time: Drivers and other travelers can choose to make their desired trip between A and B at a less congested time compared to their desired time of travel.

3. Diverting travel to a different destination: The desired trip between A and B can be diverted to a different destination, C. This is not possible for all trips. Work trips, for example, are controlled by the traveler's employment, which is generally at a fixed location. Shopping and other trips, however, can be made to alternative destinations, so that travelers may choose to go to a different shopping center when the path to the preferred one is congested.

4. Staying home: When congestion is severe enough, and widespread enough, some travelers will simply choose *not* to make a desired trip at all.

The problem is that actual field data can only measure volumes. Except in cases where congestion does not constrain demand, demand is very difficult to observe. For example, a volume study made at a location on a congested freeway would not reflect demand. True demand would be the sum of:

- actual volume passing the study location,

- volume using parallel or alternate routes that would pass the study location in the absence of congestion,
- volume passing the study location at a time different from that studied, which would have occurred during the study period in the absence of congestion,
- volume using alternative routes to alternative destinations that would pass the study location in the absence of congestion, and
- volume that would occur from trips that travelers have chosen not to make due to congestion.

The first is the only item that can be easily measured in the field. The 2nd, 3rd, and 4th items can be roughly estimated using a variety of techniques and complex field studies. The last is almost impossible to accurately assess, as there is no way to observe a trip that is not made.

Capacity is formally defined as the maximum rate of flow at which vehicles (or persons) can reasonably pass a point (or uniform segment) on a facility under prevailing conditions. Note that capacity is a *maximum*, and that it is expressed as a rate of flow for peak 15-minute time interval. Note also that it depends upon prevailing conditions which come in three different categories: (1) *roadway conditions*, which can change only with major re-design and construction; (2) *traffic conditions*, such as the presence of trucks, which can vary with time; and (3) *control conditions*, which can change only with changes in traffic control or regulation at the site.

Consider the following analogy: I have a 2-gallon bucket into which I pour 3 gallons of water. Two gallons, the size of the bucket, is the capacity. The demand is the 3 gallons of water that I try to pour into the 2-gallon bucket. The volume is what I can measure in the bucket when I finish pouring—which is limited to 2 gallons. In this case, the gallon that is *not* accommodated spills on the ground. Where excess traffic “spills” is much more complex.

6.2 Relationships among Demand, Volume (or Rate of Flow), and Capacity

There are many ways to depict the differences between volume, demand, and capacity. Consider the situation illustrated in Figure 6.1. It is a graph in which the demand flow rate is plotted and compared to the capacity of the roadway. Note that the flow rate scale is in veh/min, and that the timescale is in minutes.

Figure 6.1: Volume and Demand—an Illustration

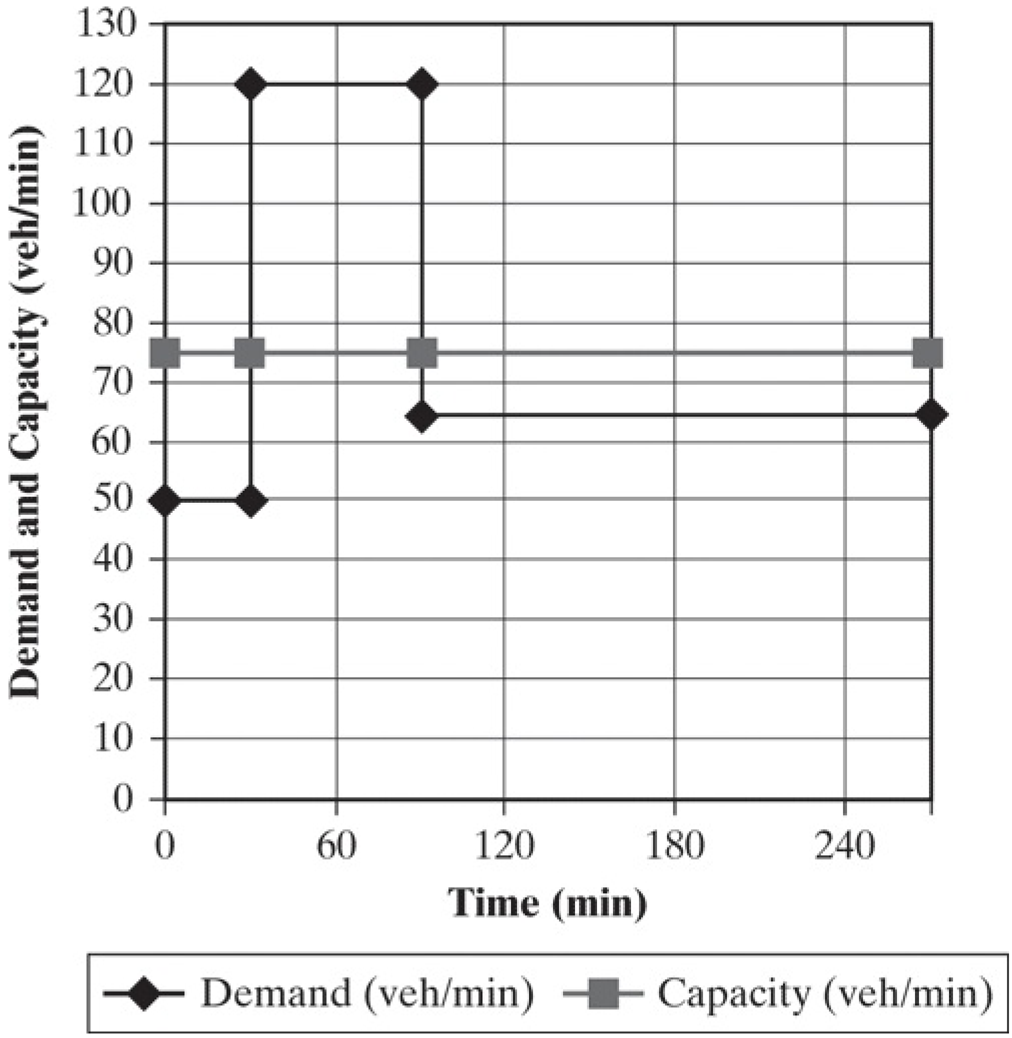


Figure 6.1: Full Alternative Text

The figure shows a case in which the capacity of the facility, 75 veh/min (or 75 × 60 = 4,500 veh/h) is exceeded by the demand from time 30 minutes to time 90 minutes. During this time, the demand is 120 veh/min (120 × 60 = 7,200 veh/h). After 90 minutes, the demand falls off to 65 veh/min (65 × 60 = 3,900 veh/h).

Several questions about this situation might be asked:

- How long does the queue get during the period when demand exceeds capacity?
- After 90 minutes, how long will it take the accumulated queue to dissipate?
- At this location, what would the observed volume be with respect to

time?

All of these can be answered by examining Figure 6.1. The plot is of a flow rate in veh/min versus time in minutes. The extent of the queue that forms during the period when demand exceeds capacity (from 30 to 90 minutes) is simply the area between the demand and capacity curves during that period. This area is:

$$Q=\left(120\ \frac{\text{veh}}{\text{min}}-75\ \frac{\text{veh}}{\text{min}}\right)\times(90\ \text{min}-30\ \text{min})=2{,}700\ \text{veh}$$

After 90 minutes, the demand drops to 65 veh/min, while the capacity remains 75 veh/min. Thus, the queue can be dissipated at a *rate* of 75 − 65 = 10 veh/min. Thus, to dissipate an accumulated queue of 2,700 vehicles would take 2,700/10 = 270 minutes—a period that starts at t = 90 minutes. The queue would clear at 360 min. This requires the assumption that the demand rate holds at 65 veh/min until t = 360 min, which is beyond the scale shown in Figure 6.1.

Essentially, in any case, the accumulation of a queue is the area between the demand and capacity curves for any period in which demand > capacity. The time to dissipate the queue is then found by determining the equivalent area between the demand and capacity curves immediately following the period in which demand > capacity. It is critical to note that neither curve must be static. Demand rates can and do change over time. Capacity need not be a constant either, as it depends upon traffic characteristics that can change, such as the presence of trucks in the traffic stream.

Figure 6.1 does not show a *volume* curve. First, there are questions involving *where* one would count volume, both before and after the breakdown occurring at 30 minutes. During the first 30 minutes, the volume curve would be exactly the same as the demand curve, as demand < capacity. Thereafter, volume—which would have to be counted as vehicles depart the queue—that is, downstream of the breakdown, would follow the capacity curve, discharging 75 veh/min. That would continue until the queue was dissipated (at 360 minutes), after which, the volume curve would once again follow the demand curve.

There is another way to present the information of Figure 6.1. Figure 6.2 shows a plot of the same data, but changes the Y-axis scale to “cumulative vehicles.” In this case, arriving vehicles (demand) is plotted and compared to departing vehicles (volume). Because the plot is of cumulative vehicles in both cases, the demand *flow rate* and the volume are the *slopes* of the respective lines for each.

Figure 6.2: Cumulative Arriving and Departing Vehicles for Figure 6.1

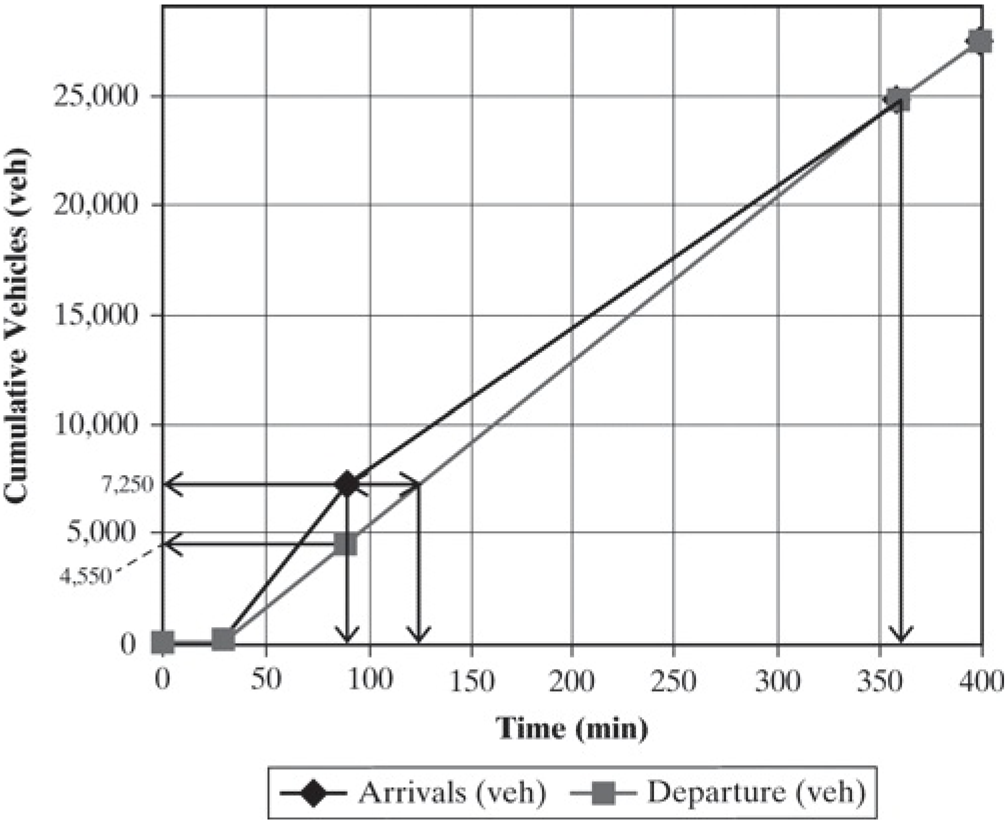


Figure 6.2: Full Alternative Text

Figure 6.1 plotted demand versus capacity. To complete the plot of Figure 6.2, the volume (in veh/min) must be determined as a function of time. Consider the following:

- Between 0 and 30 minutes, the demand < capacity. For this period of time, the volume (departure rate) will be equal to the arrival rate (demand).

- Between 30 and 90 minutes, demand > capacity. For this period of time, the volume (departure rate) will be equal to the capacity. It cannot be higher.
- Between 90 minutes and 360 minutes, the demand < capacity. However, queued vehicles do not completely dissipate until 360 minutes. Thus, the volume (departure rate) will be equal to the capacity during this period.
- After 360 minutes, the queue has dissipated, and demand < capacity. During this period, the volume (departure rate) will be equal to the demand (arrival rate).

Figure 6.2 shows the plot of cumulative arriving and departing vehicles, as described above.

The figure confirms that the maximum queue occurs at 90 minutes, and is 7,250 − 4,550 = 2,700 vehicles long, and that the queue dissipates at 360 minutes. However, the area between the curves has significance not seen in Figure 6.1. While a queue exists (30 minutes through 360 minutes), the area between the arrival and departure curves represents the total amount of delay (in vehicle-minutes) caused by the queuing at this location.

This area is comprised of two triangles: one between t = 30 and t = 90 minutes, and the other between t = 90 and t = 360 minutes. In both cases, the height of the triangles is the maximum queue of 2,700 vehicles. The bases are determined by the starting and ending times. Thus, the total delay to the 2,700 vehicles affected by the queue forming at this location is:

$$D=\frac{1}{2}(90-30)(2{,}700)+\frac{1}{2}(360-90)(2{,}700)$$
$$D=81{,}000+364{,}500=445{,}500\ \text{veh-min}$$

The total delay is 445,500 vehicle-minutes, or 445,500/60 = 7,425 vehicle-hours. This seems like a very extreme amount of delay. It is, however, spread out over all of the vehicles that arrive between t = 30 min and t = 360 min. Between 30 and 90 minutes, vehicles arrive at a rate of 120 veh/min. Thereafter, they arrive at a rate of 65 veh/min. Thus, the total number of arriving vehicles that are subject to delay is (60 × 120) + (270 × 65) = 24,750 vehicles. The average delay *per vehicle* is, therefore, 445,500/24,750 = 18 min/veh. This is not negligible, but not horrendous either.

From Figure 6.2, the maximum waiting time for an individual vehicle can also be found. This is the horizontal distance between the arrival and departure curves for any given vehicle. It is at a maximum value for the vehicle arriving exactly at 90 minutes, the time of the maximum queue size. In this case, from the figure, the waiting time for this vehicle is 130 − 90 = 40 minutes, which would have to be considered extreme, at least to the driver (and any passengers) of this vehicle.

It should be recognized that the situation shown in Figures 6.1 and 6.2 is highly idealized. Demand flow rates would not change from 50 veh/min to 125 veh/min in an instantaneous step-function, but would increase gradually over time. This would create a more complex geometry for areas, but the principles involved in determining the queue buildup and dissipation time would be the same.

The example shown in Figure 6.3, while still idealized, is more realistic. Note that the scales have been changed in Figure 6.3 to *hours* and *vehicles per hour*.

Figure 6.3: Demand and Capacity—Another Illustration

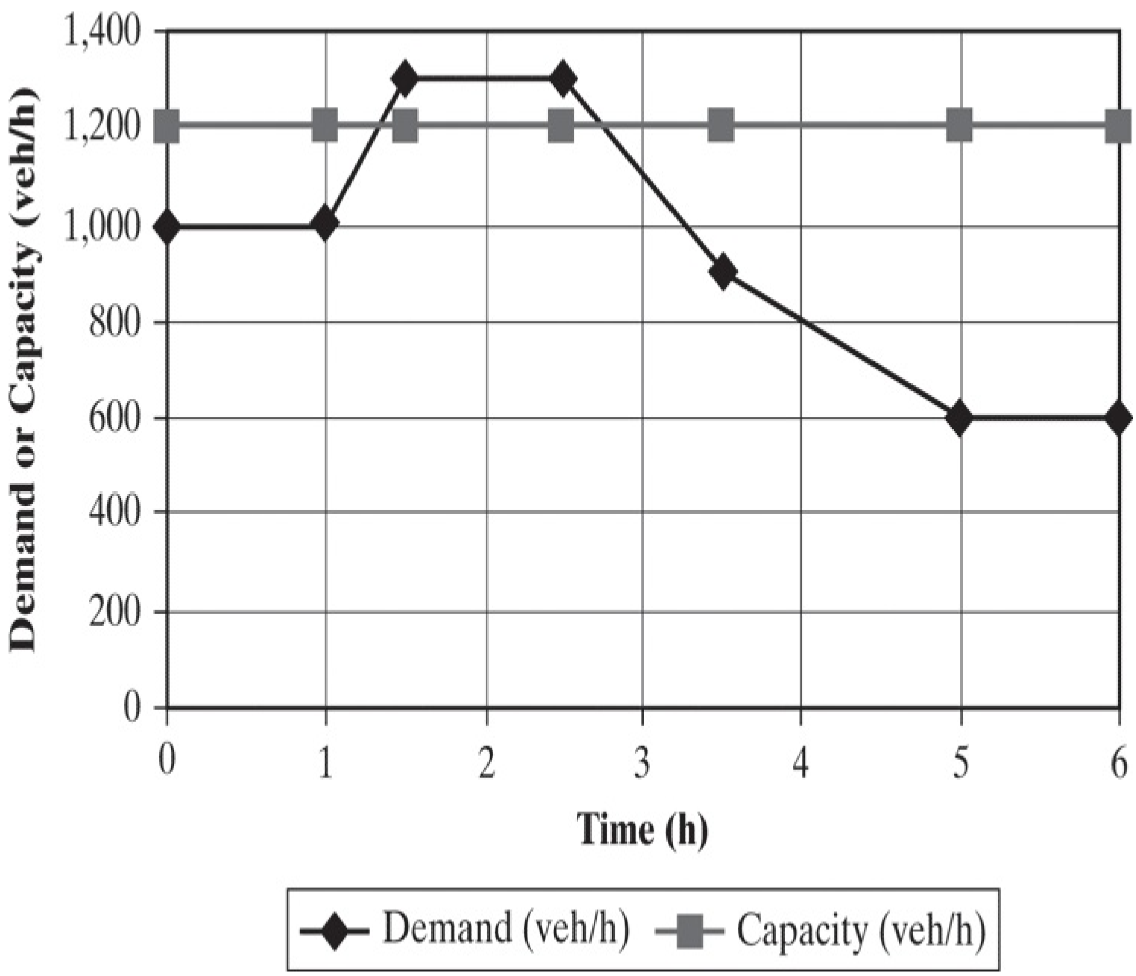


Figure 6.3: Full Alternative Text

Once again, the extent of the queue developing during the period when demand > capacity is the area between the two curves. Now, however, the points at which demand first exceeds capacity and last exceeds capacity must be precisely determined.

The demand rate increases from 1,000 veh/h to 1,300 veh/h uniformly between 1.0 and 1.5 h. It is increasing, therefore, at a rate of 300 veh/h/0.5 h, or 600 veh/h. How long does it take the demand rate to increase to the capacity of 1,200 veh/h? It would take 200/600 = 0.333 h to do so. Therefore, the point at which demand > capacity *begins* at 1.0 + 0.333 = 1.333 h.

Similarly, the demand rate decreases from 1,300 veh/h to 900 veh/h uniformly between 2.5 and 3.5 h. It is decreasing, therefore, at a rate of 400 veh/h/1 h, or 400 veh/h. How long does it take the demand rate to decrease back to the capacity of 1,200 veh/h? It would take 100/400 = 0.25 h. Thus, the point at which demand < capacity begins at 2.5 + 0.25 = 2.75

h.

The area between the demand and capacity curves during the period when demand > capacity is the sum of the areas of two triangles and a rectangle, as shown in Figure 6.4.

Figure 6.4: Areas on Figure 6.3 Illustrated

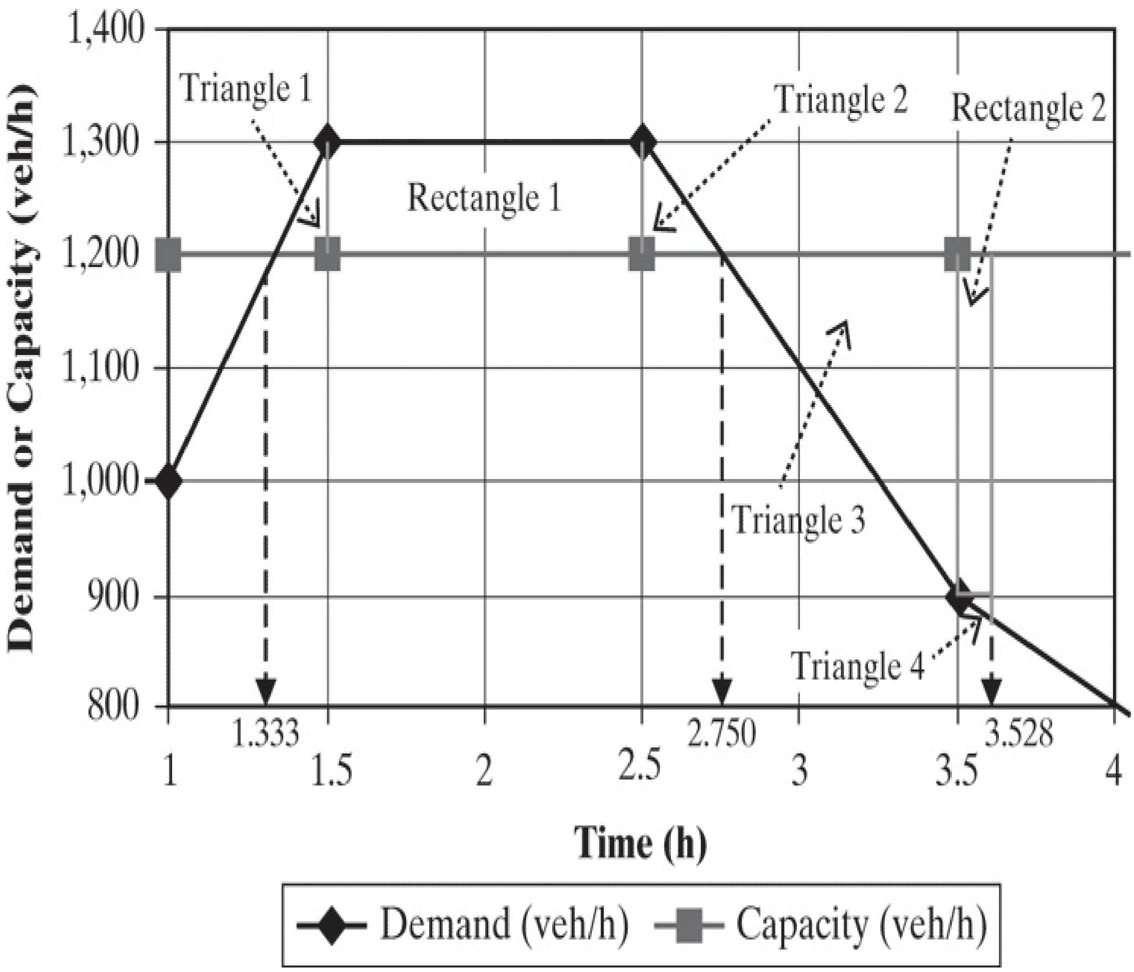


Figure 6.4: Full Alternative Text

- The first triangle occurs between t = 1.333 h and ends at t = 1.5 h. The area of a triangle is ½ the base × the height. The base is 1.500 − 1.333 = 0.167 h, and the height is 1,300 − 1,200 = 100 veh/h. The area under this triangle is 0.5 × 0.167 × 100 = 8.35 vehicles (Triangle 1, Figure 6.4).

- The rectangle occurs between t = 1.5 h and t = 2.5 h. The height of the rectangle is 1,300 −1,200 = 100 veh/h, and the area is (2.5−1.5) × 100 = 100 vehicles. (Rectangle 1, Figure 6.4).

- The last triangle occurs between t = 2.5 h and 2.75 h. The height of the triangle is again 100 veh/h. The area under the triangle is, therefore, 0.5 × 0.25 × 100 = 12.5 veh. (Triangle 2, Figure 6.4).

Therefore, the total size of the queue developing during the period when demand > capacity is 8.35 + 100 + 12.5 = 120.85, say 121 vehicles.

To find out how long it takes the queue to dissipate, we have to find a time where the area between the demand and capacity curves is 121 vehicles during the period immediately after the clearance of the queue, which occurs at 2.75 hours. Between 2.5 and 3.5 hours, the demand flow rate is decreasing at a rate of 400 veh/h. If the time to dissipate the queue is less than the 0.75 hour period between 2.75 h and 3.5 h, then we are looking for a triangle (Triangle 3, Figure 6.4) that has a base of t −2.75 and a height of 1,200 − 900 = 300 that has an area of 121 vehicles, or:

$$121=\frac{1}{2}(t-2.75)\times 300 \qquad 121=150(t-2.75) \qquad t-2.75=\frac{121}{150}=0.807 \qquad t=0.807+2.75=3.557\text{ h}$$

Because this is *more* than 2.5 hours, the change in demand flow rate that occurs at 3.5 h must be taken into account, and the task becomes more complex. The area between the curves in the triangle between 2.75 h and 3.5 h is 0.5 × 0.75 × (1,200 − 900) = 112.5 veh. This means that at 3.5 h, when the demand rate begins to change again, there is still a remaining queue of 121 − 112.5 = 8.5 veh left to clear. While this is an almost trivial amount, the computation will be completed to illustrate the process.

Between 3.5 h and 5 h, the demand rate decreases uniformly from 900 veh/h to 600 veh/h—a rate of decline of 300 veh/h/1.5 h = 200 veh/h. The queue will obviously clear a short time after 3.5 (there are only 8+ vehicles left in the queue to clear at this time).

We are now looking for an area of 8.5 veh between the demand and capacity curves between time t (when the queue clears) and 3.5 h. That area, however, is comprised of a rectangle (Rectangle 2, Figure 6.4) between t and 3.5 h with a height of 1,200 − 900 = 300 veh/h and a triangle (Triangle 4, Figure 6.4) between t and 3.5 h with a height that depends upon the value of time *t*. The value of the height of the triangle, in terms of t would be 200 × (*t* − 3.5). Time *t* may now be found as follows:

$8.5 = 300(t-3.5) + 0.5 \times (t-3.5) \times 200 \times (t-3.5)$ $8.5 = 300(t-3.5) + 100(t-3.5)^2$ $t = 3.528$ h.

6.3 The Formation of Queues and Their Impacts

In the previous sections, the issues of “congestion” and “breakdowns” were noted and discussed. A “breakdown” occurs when the volume of traffic arriving at a point exceeds the capacity of that point to discharge vehicles. This requires that the *capacity* of the facility (or facilities) immediately upstream of the breakdown location be *higher* than the capacity of the breakdown location itself. Thus, an overwhelming number of traffic breakdowns occur at junctions where the *capacity* of converging roadways exceeds the *capacity* of those roadways leaving the junction. Such locations are common throughout any traffic system. Typical situations include (but are not limited to) the following:

- On-Ramps: On limited-access facilities, or uninterrupted segments of surface facilities, traffic from freeway lanes *plus* traffic from the on-ramp must fit into downstream lanes on the facility—which is normally the same as the number of upstream lanes. For example, on a freeway with three lanes in one direction, a single-lane ramp and the three freeway lanes merge to form three downstream freeway lanes—four lanes of traffic merge into three.

- Major Merge Points: There are many points in a traffic system where two or more major roadways merge into a single roadway. At many of these, the number of arriving lanes exceeds the number of departing lanes.

- Intersections: At intersections, there are one or more arriving approaches and one or more departing approaches. Where the total number of arriving lanes exceeds the total number of departing lanes, the potential for a breakdown exists. Where intersections are signalized, arrivals on each approach are limited by upstream signals, which may permit more vehicles to arrive downstream than the downstream signal can accommodate.

Whenever more vehicles arrive at a point than can be discharged, a queue is formed. The queue grows until the rate of arrivals becomes less than the rate of departures.

Consider the case illustrated in Figure 6.5. It shows a classic case of a major merge point at which the capacity of the upstream roadways approaching the merge is higher than the capacity of the downstream facility. Figure 6.5 shows the capacity (*c*), the demand flow rate (*d*), and the actual flow rate (*v*) that would be observed for each roadway.

Figure 6.5: Demand, Capacity, and Volume at a Major Merge Point

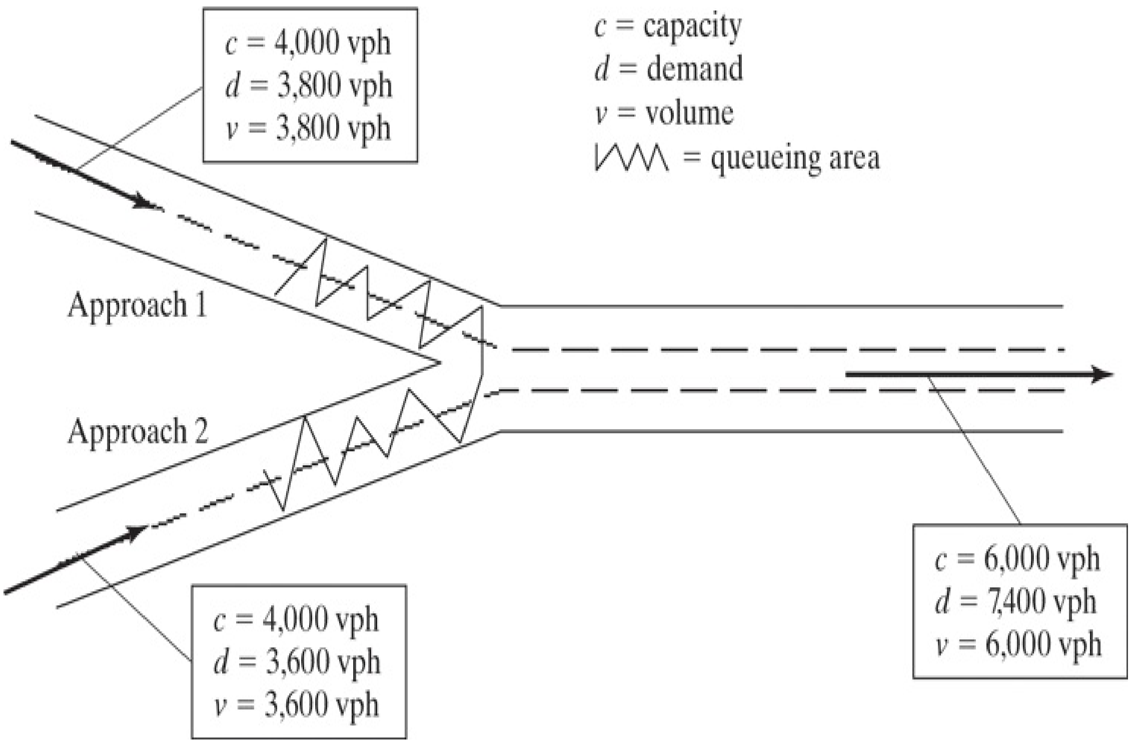


Figure 6.5: Full Alternative Text

From Figure 6.5, the following facts can be determined:

- On Approach 1, the capacity is 4,000 veh/h, and the demand flow rate is 3,800 veh/h. Capacity, therefore, is sufficient to accommodate the demand.

- On Approach 2, the capacity is 4,000 veh/h, and the demand flow rate is 3,600 veh/h. Capacity, therefore, is sufficient to accommodate the demand.

- On the downstream freeway, the capacity is 6,000 veh/h, but the demand flow rate consists of all vehicles arriving from Approaches 1 and 2, or 3,800 + 3,600 = 7,400 veh/h. This exceeds the capacity by 1,400 veh/h. Therefore, a queue will begin to form—and will build up at a rate of 1,400 veh/h for as long as the stated demand flow rates exist.

A key question is what volumes (or rates of flow) would be observed on the two arriving approaches, and on the downstream freeway. It must be noted that a queue is forming, starting at the merge point of the two approaches. The queue will propagate upstream (split between the two approaches) at a rate of 1,400 veh/h. Any volume (or rate of flow) counted from a location within the growing queue will be unstable. Arriving volumes can only be legitimately counted at a location upstream of the back of the queue (which is moving further backward over time).

Assuming count locations on the arriving approaches that are upstream of the propagating queue(s), the observed volumes on each approach would be equal to the demand, because the demand is *not* constrained by capacity.

A count taken immediately downstream of the merge point, however, would reflect the constraint of the downstream capacity on demand. The observed volume (or rate of flow) *could not* be higher than capacity, which equals 6,000 veh/h. Thus, the downstream volume (or rate of flow) would be observed to be 6,000 veh/h.

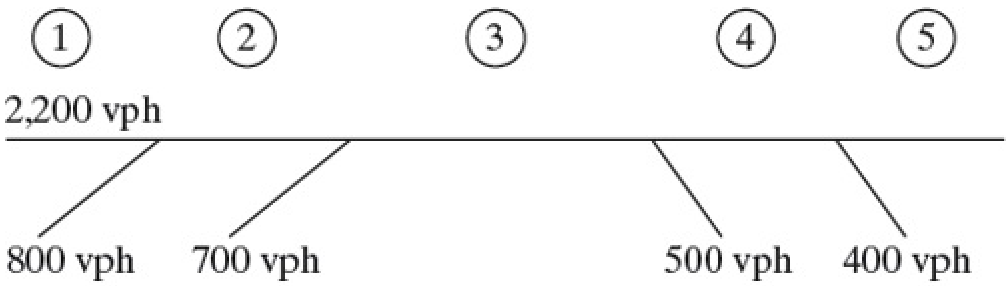
There is a subtle point here: Is the measured volume (rate of flow) of 6,000 veh/h a measure of capacity? In this case, given the defined values, yes. In general, however, discharge from a breakdown location may *not* reflect capacity. Capacity is actually defined as the maximum rate of flow that can be achieved under *stable* operating conditions. A queue is not stable.

In technical terms, the capacity would be measured as the downstream rate of flow *just before the breakdown occurs*. After the queue begins to form, the downstream flow rate reflects *queue discharge* conditions. It is generally thought that maximum queue discharge rates are somewhat less than stable capacity, although there are some field studies that contradict this. In practical terms, however, the maximum queue discharge rate may be more meaningful than capacity under stable conditions. The latter often reflects a transient condition that exists for only a short period of time.

6.4 Bottlenecks, Hidden Bottlenecks, and Demand Starvation

Consider the situation illustrated in Figure 6.6. It shows true demand flows (Figure 6.6a) and segment capacities (Figure 6.6b) for a section of freeway including two on-ramps and two off-ramps. This configuration creates four distinct segments of the freeway, labeled 1 through 4, as well as four ramps. Consider the volume (or rate of flow) that would be observed on the four freeway segments:

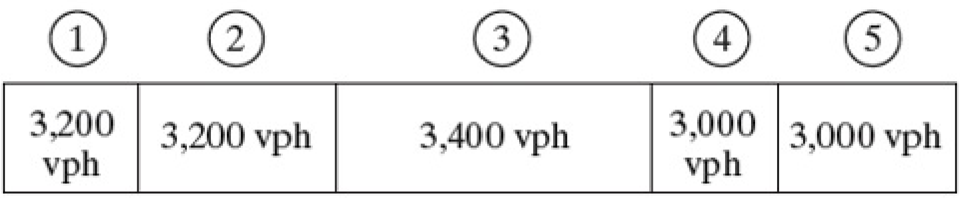
Figure 6.6: The Effect of Bottlenecks on Observed Volumes



(a) True Demand

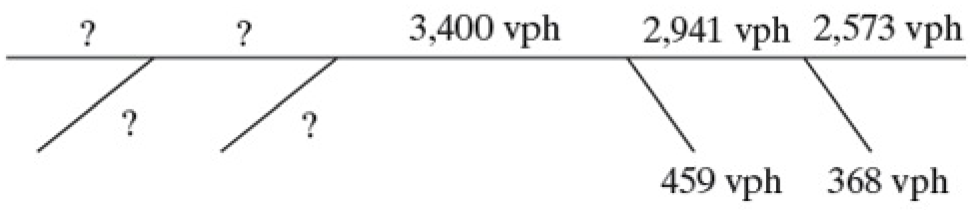
(a) True Demand

6.4-1 Full Alternative Text



(b) Segment Capacities

6.4-1 Full Alternative Text



(c) Observed Volumes

6.4-1 Full Alternative Text

- In Segment 1, the demand is 2,200 veh/h, and the capacity is 3,200 veh/h. Normally, the capacity would *not* constrain the demand.
- In Segment 2, the demand entering the segment is 2,200 + 800 = 3,000 veh/h, and the capacity is 3,200 veh/h. Once again, the capacity would *not* constrain the demand.
- In Segment 3, the demand entering the segment is 3,000 + 700 = 3,700 veh/h. The capacity of Segment 3, however, is only 3,400 veh/h. This is less than the demand, and *would* constrain the demand.
- In Segment 4, the demand is 3,700 − 500 = 3,200 veh/h, which is higher than the capacity of this segment, 3,000 veh/h. Once again, demand is constrained.
- In Segment 5, the demand is 3,200 − 400 = 2,800 veh/h, which is less than the capacity of 3,200 veh/h. Demand at this location would not be constrained.

What does all of this mean for the observed volumes (or flow rates, if counts were taken for less than a full hour)? The limiting segment is Segment 3, with a demand of 3,700 veh/h and a capacity of 3,400 veh/h. At the entry to this segment, a queue will begin to form, and will propagate upstream at a rate of 3,700 − 3,400 = 300 veh/h. Thus, the actual volumes observed in Segments 1 and 2 (where there is no constraint on demand) will depend upon *when* they are taken. If the counts occur *before* the queue backs up to and past the count location, they will be equal to the demand.

If, however, counts occur *after* the queue backs up to and past the count location, they will be unstable, and will represent movement within the queue.

In Segment 3, the observed volume will be 3,400 veh/h, which is equal to the capacity of the segment. No higher volume (or flow rate) can exist. The demand on Segment 4 is 3,200 veh/h. However, that depends upon 3,700 veh/h flowing through and out of Segment 3, which is not possible. Only 3,400 veh/h will approach the first off-ramp. If we assume that the distribution of vehicles between the first off-ramp and the freeway is the same as represented in the demand, then 3,200/3,700 = 0.865 of the vehicles approaching the off-ramp will continue on the freeway. Therefore, the volume observed in Segment 4 of the freeway is 3,400 × 0.865 = 2,941 veh/h (which is *less* than the capacity of this segment—3,000 veh/h).

Now, 2,941 veh/h are approaching the second off-ramp. Again, assuming the same split between the ramp and the freeway as in the demand, it is expected that 3,000/3,400 = 0.882 of the vehicles approaching the second off-ramp will continue on the freeway. Thus, the observed volume on Segment 5 of the freeway would be 0.882 × 2,941 = 2,593 veh/h.

6.4.1 The Hidden Bottleneck

The current volumes on Segments 4 and 5 of the freeway do not exceed the capacity of those segments. What would happen, however, if the bottleneck in Segment 3 were *improved* and the capacity of Segment 3 raised to 4,000 veh/h? Then, the true demand of 3,700 veh/h would flow through Segment 3 without constraint, and 3,200 veh/h would attempt to enter Segment 4. This, however, *does exceed* the capacity of Segment 4, creating a bottleneck. From observations of volume and queuing, Segment 4 appears to be operating acceptably in the original scenario. The potential bottleneck in Segment 4 was *hidden* by the fact that Segment 3 did not allow the true demand to proceed into Segment 4. Segment 4, therefore, represents a *hidden bottleneck*, obscured by the bottleneck in Segment 3. Fix Segment 3 without also fixing Segment 4, and the bottleneck simply moves downstream a bit. The point is this: Not every bottleneck or capacity constriction is obvious from field studies. Careful analysis needs to be conducted to identify locations such as this, so that they can be included in potential improvement projects.

6.4.2 Demand Starvation

What actually happens in Figure 6.6 is that true demand cannot reach Segment 4 to be observed as an actual volume. In effect, the upstream bottleneck in Segment 3 *starves* the demand to downstream points. Demand starvation is the cause of hidden bottlenecks, although demand starvation can occur in places where there is no hidden bottleneck as well.

The occurrence of hidden bottlenecks and demand starvation must be well understood. It explains, in very practical terms, why measured traffic volumes or flow rates do *not* necessarily represent true demand. As has been noted, true demand is very difficult to measure or estimate, and design and control measures must take into account the potential existence of substantial demand that is not directly observable in existing street and highway traffic.

6.5 Capacity versus Queue Discharge

It has been noted previously that *capacity* is the maximum rate of flow that can be sustained without experiencing a breakdown. It often occurs for very short periods of time, just before a breakdown occurs. After the breakdown, vehicles will leave the queue established at the breakdown location at the queue discharge rate. What is the queue discharge rate, and what impact does it have on breakdowns and queuing?

While capacity values are fairly well established, queue discharge rates are less stable, and no profession- wide criteria are available for its value. Studies have shown that it varies from a value equal to capacity to a value that is 5% to 10% *less* than capacity. A few studies actually resulted in measured queue discharge rates that were *higher* than capacity, but these have been rare occurrences. Nevertheless, the difference between the two is critical to how long it takes queues to clear *after* a blockage has been removed.

It is a common occurrence: A driver gets up in the morning, and the traffic report announces a stalled truck in one lane of a three-lane freeway segment on his route to work. The stall occurs at 6:00 AM. The traffic reporter happily announces that the blockage has been cleared at 6:30 AM. The driver heads to work, arriving at the site at 8:00 AM, and is stuck in a miles-long traffic jam. The traffic reporter must have made a mistake. But no, when the driver gets to the site of the reported blockage, there is nothing there, and from that point on, traffic moves freely. What just happened?

Consider two scenarios: (1) The capacity of the site is 2,000 veh/h/ln, but the queue discharge rate is only 1,800 veh/h/ln (a 10% drop). (2) The capacity of the site is 2,000 veh/h/ln, but the queue discharge rate is also 2,000 veh/h/ln (no drop). For both scenarios, the arriving flow rate at the location is 6,000 veh/h from 6:00 AM to 9:00 AM in the morning. From 9:00 AM to 10:00 AM, the arriving flow rate decreases to 5,000 veh/h. After 10 AM, the arriving flow rate decreases to 4,000 veh/h for the remainder of the day. Two questions will be examined: How large does the queue get under both scenarios, and how long does it take before the queue dissipates under both scenarios? Figure 6.7 illustrates the situation described.

Figure 6.7: Two Scenarios for a Breakdown

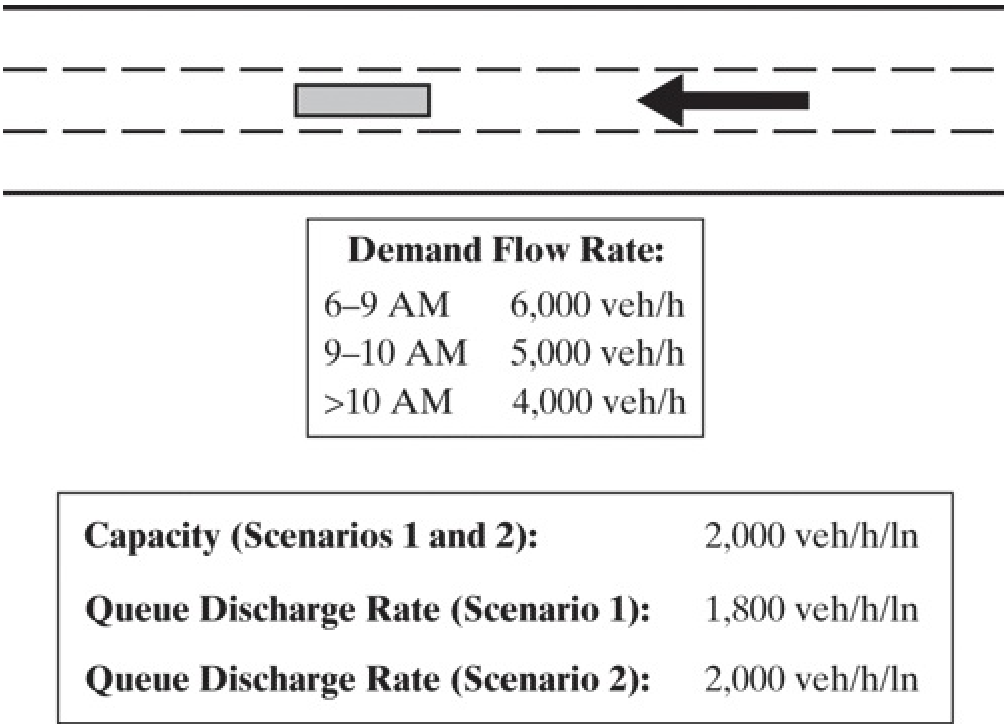


Figure 6.7: Full Alternative Text

There are two critical facts to consider: (1) Between 6:00 AM and 6:30 AM, there are only two lanes for moving traffic; at all other times (before and after), there are three moving lanes for traffic. (2) Because the demand is essentially equal to the capacity of the freeway before the breakdown occurs (6,000 veh/h arriving; capacity = 3 × 2,000 = 6,000 veh/h), when the breakdown does occur, a *queue is immediately established*. Therefore, between the time of the breakdown and the time that the queue is dissipated, the *queue departure rate* limits the number of vehicles that can move past the location.

Tables 6.1 and 6.2 illustrate the computation of the maximum extent of the queue, and the time required for it to fully dissipate under the two scenarios. In each case, the number of vehicles arriving at the blockage location is compared to the number of vehicles leaving it—limited by either the capacity or the queue discharge rate. In the first hour, half-hour time intervals are used, as the blockage exists for only half an hour. For all subsequent periods, one-hour time intervals are used.

Table 6.1: Queuing Analysis, Scenario 1

Time	Arrival Rate (veh/h)	Arrivals (veh)	Departure Rate (veh/h/ln)	Departures (veh)	Size of Queue (veh)
6:00–6:30 AM	6,000	½ × 6,000 = 3,000	1,800	½ × 2 × 1,800 = 1,800	3,000 − 1,800 = 1,200
6:30–7:00 AM	6,000	½ × 6,000 = 3,000	1,800	½ × 3 × 1,800 = 2,700	1,200 + 3,000 − 2,700 = 1,500
7:00–8:00 AM	6,000	6,000	1,800	3 × 1,800 = 5,400	1,500 + 6,000 − 5,400 = 2,100
8:00–9:00 AM	6,000	6,000	1,800	3 × 1,800 = 5,400	2,100 + 6,000 − 5,400 = 2,700
9:00–10:00 AM	5,000	5,000	1,800	3 × 1,800 = 5,400	2,700 + 5,000 − 5,400 = 2,100
After 10:00 AM	4,000	4,000	1,800	3 × 1,800 = 5,400	−1,600/h
Time to Dissipate Queue	After 10:00 AM, queue dissipates at a rate of 1,600 veh/h. There is a remaining queue at 10:00 AM of 2,100 veh. Thus, time to clear = 2,100/1,600 = 1.31 h. Queue clears at 10 + 1 h, 19 m = 11:19 AM				

Table 6.1: Full Alternative Text

Table 6.2: Queuing Analysis, Scenario 2

Time	Arrival Rate (veh/h)	Arrivals (veh)	Departure Rate (veh/h/ln)	Departures (veh)	Size of Queue (veh)
6:00–6:30 AM	6,000	½ × 6,000 = 3,000	2,000	½ × 2 × 2,000 = 2,000	3,000 − 2,000 = 1,000
6:30–7:00 AM	6,000	½ × 6,000 = 3,000	2,000	½ × 3 × 2,000 = 3,000	1,000 + 3,000 − 3,000 = 1,000
7:00–8:00 AM	6,000	6,000	2,000	3 × 2,000 = 6,000	1,000 + 6,000 − 6,000 = 1,000
8:00–9:00 AM	6,000	6,000	2,000	3 × 2,000 = 6,000	1,000 + 6,000 − 6,000= 1,000
9:00–10:00 AM	5,000	5,000	5,000	3 × 2,000 = 6,000	1,000 + 5,000 − 6,000=0
Time to dissipate queue	The queue dissipates at 10:00 AM.				

Table 6.2: Full Alternative Text

The analysis in Table 6.1 shows that the queue builds to a maximum length of 2,700 vehicles, which occurs at 9:00 AM. The queue finally clears at 11:19 AM, at which time, the original capacity of 2,000 veh/h/ln is regained. The poor driver who arrives at the site at 8:00 AM joins the end of a 2,100-vehicle queue, even though the blockage was gone at 6:30 AM! There are three lanes on the facility. Assuming that the 2,100 vehicles in queue are equally distributed, there are 700 veh/ln queued. Each vehicle would be expected to occupy a distance of between 30–40 ft (remember, vehicles are shuffling through the queue, not standing still), creating a queue length of about 700 × 35 = 24,500 ft, or 24,500/5,280 = 4.64 miles!

The analysis, however, changes radically if the queue discharge rate is the same as capacity, as seen in Table 6.2.

In this scenario, the queue reaches a maximum of 1,000 veh at 6:30 AM, and remains at that length until 9:00 AM. The entire queue dissipates during the hour between 9:00 AM and 10:00 AM, and is fully cleared at 10:00 AM. In this case, the driver arriving at 8:00 AM faces a queue of 1,000 veh, less than half the length that occurs in Scenario 1. Again, assuming an equal distribution of queue vehicles, the approximate length of the queue is a maximum of (1,000/3) × 35 = 11,667 ft, or 11,667/5,280 = 2.21 miles.

Obviously, the issue of the queue discharge rate is a critical one. Unfortunately, there is no national "standard" that can be established for it. Local and regional studies must be conducted to produce realistic values for various parts of the country to apply to specific analyses.

This entire illustration has used what is known as "deterministic queuing" analysis. It assumes that the queue forms at the point, that is, that queued vehicles are vertically stacked. It understates somewhat the formation of the queue and its size. This is because, in real terms, the back end of the queue is moving toward the arriving flow, which accelerates the arrival rate joining the queue. For many purposes, this is, however, a reasonable approach. Complete queuing analysis involves many factors that are difficult to precisely measure in the field.

6.6 Closing Comments

“Vehicles per hour” is a simple unit expressing the number of vehicles passing a point in an hour. Despite its simplicity, it is used to quantify three very different concepts: demand, capacity, and volume (or rate of flow).

Engineers must clearly understand the three concepts involved, and the critical differences between them. All critically relate to the traffic engineer’s profession, and to the understanding of current traffic characteristics and potential future impacts.

Problems

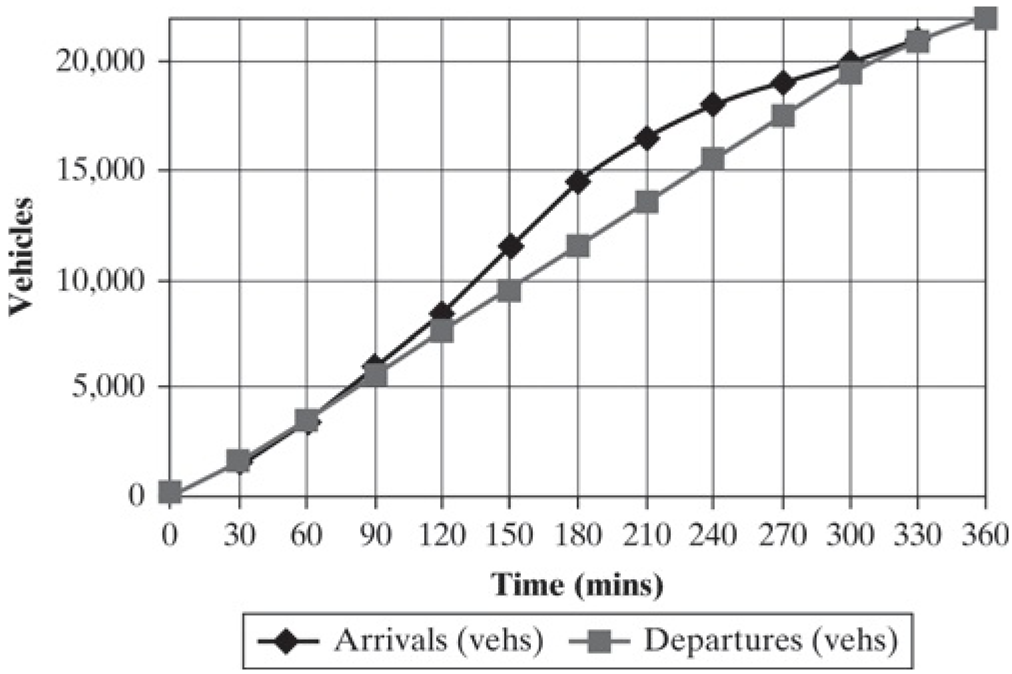
1. 6-1. A freeway bottleneck location (at a major on-ramp) has a capacity of 5,000 veh/h. During a typical morning peak period, the actual demand flow rate arriving at the bottleneck is as follows:

7:00 AM–8:00 AM	4,500 veh/h
8:00 AM–8:30 AM	5,400 veh/h
8:30 AM–9:30 AM	6,000 veh/h
9:30 AM–10:00 AM	5,000 veh/h
10:00 AM–11:00 AM	4,500 veh/h
After 11:00 AM	4,000 veh/h

    1. Draw a graph of demand versus capacity for the situation described.
    2. What is the size of the queue upstream of the bottleneck at 8:30 AM, 9:30 AM, 10:00 AM, and 11:00 AM?
    3. At what time does the queue dissipate?

2. 6-2. For the situation described in Problem 1, construct a plot of cumulative arrivals and departures versus time (in minutes), starting at 7:00 AM and ending at 1:00 PM., that is, *t* = 0.0 min at 7:00 AM and *t* = 240.0 min at 1:00 PM. From this plot:
    1. identify the maximum size of the queue that forms.
    2. identify the longest waiting time for a vehicle during the study time period.

3. 6-3. Consider the following plot of cumulative arriving and departing vehicles at a freeway bottleneck location:



6.1-3 Full Alternative Text

From this plot, determine the following:

1. What is the capacity of the bottleneck location?
2. What is the maximum size of the queue that develops?
3. What is the longest wait time that any vehicle experiences during the breakdown?

4. 6-4. A segment of eight-lane urban freeway experiences a breakdown that blocks two lanes completely for a period of one hour in the peak direction. The breakdown occurs at 9:00 AM and is cleared by 10:00 AM. The capacity of the segment is 2,100 veh/h/ln, with a maximum queue discharge rate of 1,800 veh/h/ln. Beginning at 9 AM, the traffic demand at this location is:

9:00 AM–10:00 AM	8,400 veh/h
10:00 AM–11:00 AM	8,000 veh/h
11:00 AM–12:00 Noon	7,000 veh/h
After 12:00 Noon	6,000 veh/h

1. What is the maximum queue size that will form in this scenario?
2. At what time does the stable capacity value resume?

Chapter 7 Level of Service and the *Highway Capacity Manual*: History and Fundamental Concepts

The U.S. standard for capacity and level of service analyses used throughout traffic engineering is the current edition of the *Highway Capacity Manual* (HCM), a publication of the Transportation Research Board (TRB) of the National Academy of Engineering. Its content is controlled by the Highway Capacity and Quality of Service Committee (HCQSC) of the TRB. The committee consists of 34 regular and special members appointed by the committee chair with TRB approval and over 100 additional professionals who participate through a series of subcommittees, each focused on particular parts of the manual.

The committee was formally established in 1944 by the then Highway Research Board. It followed on the work of the National Interregional Highway Committee (NIHC), which was appointed by President Franklin Delano Roosevelt in 1941, and completed its work in 1944. Its purpose was to study the nation's highway systems and make recommendations for new and/or improved facilities. This study was the first to recommend what would eventually become the Interstate System, and identified critical shortcomings in the nation's highways.

The committee was formed specifically to write the first edition of the HCM, and was staffed primarily by employees of the Bureau of Public Roads (BPR), and chaired by Olaf K. Normann, who had been a leading staff member to the NIHC. He had also pioneered early work in the measurement and quantification of traffic streams. He continued to chair the committee through its second edition in 1965, and died shortly before it was published. Throughout, he was supported by Powell Walker, another full-time employee of the BPR, who was the committee's first secretary. Other members of the committee, initially invited by Normann, came from a variety of public and private transportation agencies, including the Port of New York Authority, the New York City Department of Traffic, the California Department of Public Works, the Chicago Street and Traffic Commission, and several prominent transportation consultants of the time.

With the realization that rapid expansion of the highway network would soon occur, and was indeed already occurring, the manual was intended to provide a national standard for design and a degree of uniformity throughout the nation.

7.1 Uninterrupted and Interrupted Flow Facilities

From the beginning, the manual dealt with two very different categories of highway facilities, which were categorized as *uninterrupted* or *interrupted* flow. These terms are defined and discussed in detail in Chapter 6.

The first HCM dealt primarily with uninterrupted flow facilities, but included some information on signalized intersections. Subsequent editions added considerable detail on interrupted facilities. As will be seen, however, many of the fundamental concepts embodied in the HCM were developed with uninterrupted facilities in mind. Thus, in some cases, their application to interrupted flow facilities is somewhat less intuitive.

Table 7.1 shows the current types of segments or facilities that are treated in the current edition (the 6th edition) of the HCM.

Table 7.1: Uninterrupted and Interrupted Flow Facilities in the HCM

Uninterrupted Flow Facilities	Interrupted Flow Facilities
Basic freeway segments	Signalized intersections
Multilane highway segments	Two-way STOP-controlled intersections
Two-lane highway segments	Multiway STOP-controlled intersections
Freeway weaving segments	Roundabouts
Freeway merge segments	Interchanges
Freeway diverge segments	Urban streets
Freeway facilities	Urban arterials
	Bicycle facilities
	Pedestrian facilities

Source: Reprinted with permission from *Highway Capacity Manual, 6th Edition: A Guide for Multimodal Mobility Analysis*, Transportation Research Board, © 2016 by the National Academy of Sciences, Courtesy of the National Academies Press, Washington, D.C.

Table 7.1: Full Alternative Text

Many, but not all, of these analysis methodologies are covered in other chapters of this book. In general, this text includes material from the most recent edition of the HCM. The sixth edition of the HCM, with the subtitle *A Guide for Multimodal Mobility Analysis*, was published in late 2016.

7.2 A Brief Chronology of the *Highway Capacity Manual*

7.2.1 The 1950 Highway Capacity Manual

The first edition of the HCM [1] was jointly published by the U.S. Government Printing Office and the Highway Research Board. Because virtually all the work had been done by employees of the BPR, its director, Thomas H. MacDonald, insisted that it be published first as a series of articles in *Public Roads* [2,3].

It was a relatively short document of 147 pages, and provided basic design standards for various types of traffic facilities. Although some of the methodologies included could be used in analysis, its intended purpose was to provide design guidance and some national consistency in design standards.

It provided the first formal definition of *capacity*, which will be discussed later, and took initial steps in the consideration of quality of traffic flow. For its time, it was a critically important document that allowed designers to uniformly consider how big to make their facilities.

7.2.2 The 1965 Highway Capacity Manual

The Highway Capacity Committee was inactive for several years following the publication of the 1950HCM. It was reactivated in 1953 to begin work on a second edition of the manual. The 1950HCM spurred a great deal of interest and discussion, and as the use of the 1950HCM continued to increase, some of the gaps in existing knowledge became evident. O. K. Normann continued to chair the committee, but no longer worked full time on the effort.

Two major efforts were undertaken by the committee during this period. In 1954, a detailed survey of traffic officials across the United States resulted in the collection of detailed data on operations at 1,600 signalized intersections across the nation. These formed the basis for development of a more robust analysis methodology for signalized intersections. In 1957, the committee sponsored the publication of *Highway Research Bulletin 167* [4], which contained six foundational papers on highway capacity subjects covering a broad range of relevant research topics. Many have referred to this publication as the 1.5th edition of the HCM.

The second edition of the HCM was published with a 1965 date, although it first appeared in early 1966. The 1965HCM [5] marked a major step forward in the field, introducing a wide range of new material:

- The concept of *level of service* (LOS) was introduced and defined, and applied to a broad range of traffic facilities and segments.
- Greatly improved and expanded methodologies were developed for freeways, weaving segments, ramps, signalized intersections, and downtown streets, with some descriptive material addressing bus transit.
- Most methodologies, while still focusing on design, could be easily applied to analysis of existing or planned future facilities.

The size of the manual increased to 411 pages, and its use throughout the United States and the world increased dramatically.

7.2.3 The 1985 *Highway Capacity Manual*

Work on the 1985HCM began shortly after publication of the 1965HCM. Initial use of the 1965HCM in the profession revealed information gaps that still needed to be filled. The committee entered a new phase of its existence. With the death of O. K. Normann, the committee no longer could rely on the heavy assistance of BPR personnel. Indeed, the committee truly became a body of volunteer experts, all with full-time commitments elsewhere. A new paradigm emerged: the committee developed a research and development plan that was implemented through funded research efforts conducted by a variety of contracting agencies, including universities. Two funding agencies provided most of the support: the National Cooperative Highway Research Program (NCHRP) and the Federal Highway Administration (FHWA), which was the successor to the BPR.

The first funded efforts began in the early 1970s, and continued through the publication of the 1985HCM. Because publication of the HCM itself was delayed until 1985 (itself a delay from the initial 1983 target date), the committee released *Transportation Research Circular 212: Interim Materials on Highway Capacity* [6] in 1980 to allow the profession to conduct trial usage of proposed methodologies.

The 1985HCM [7] represented a major step forward in highway capacity and level of service analysis. At 506 pages, it introduced a number of new approaches:

- It was the first manual published in loose-leaf format. This was done with the expectation that page-by-page updates could be issued as needed by the committee. This proved to be an unworkable concept due to practical production considerations, and updates were actually issued as packages of revised chapters.

- Signalized intersection analysis adopted a critical movement approach for the first time. The approach, historically the basis for signal timing analysis, eliminated essential inconsistencies between signal timing and capacity analysis approaches.

- It was the first manual that included methodologies that were, in some cases, quite difficult to implement by hand. It was the first manual for which an accompanying software package, Highway Capacity Software (HCS), was developed. The software was initially developed by Polytechnic University under contract to FHWA, and later revised and maintained (to this day) by the McTrans Center of the University of Florida at Gainesville.

Because the loose-leaf format was intended to allow interim updates, the 1985HCM underwent two major updates in 1994 [8] and 1997 [9].

7.2.4 The 2000 *Highway Capacity Manual*

After the publication of the 1985HCM, the committee formally created a structure of subcommittees, and added members to each. Each subcommittee was focused on a particular chapter or issue, and invited interested professionals to join. This led to a large increase in manpower, while maintaining the workable size of the parent committee. It also shifted the review of specific research and methodology proposals from the parent committee to the subcommittees. The parent committee essentially reviewed and voted on recommendations from the various subcommittees. As the development of a fourth edition approached, the committee was in virtually continuous production mode as the pace and quantity of relevant research accelerated to a torrent.

The 2000HCM [10] was a significant change from earlier editions, and it tried to address the needs of an ever-growing number of different user groups. As a result, the 2000HCM grew to over 1,100 pages, and was the first manual in which virtually no member of the committee had read the entire document.

- It was the first HCM to focus heavily on planning applications.
- It was the first HCM to address analysis of multiple facilities in corridors and networks, although much of this material was descriptive.
- It was the first HCM to specifically address alternative tools, primarily simulation, and attempted to identify applications that were better suited to these. It also addressed the comparison between simulation outputs and those of the HCM.
- Because of a national effort to convert to metric standards, the 2000HCM was published in *two* versions, one in English units and the other in metric units.
- A CD-ROM with interactive elements was produced with the HCM.
- A number of older methodologies were replaced or significantly

updated with new research.

The complexity issue mushroomed with the 2000HCM. Some procedures were now virtually impossible to apply by hand, and the importance of the HCS package for implementation became paramount. The committee had to deal, for the first time, with the reality that the software *was* the manual for many users. The problem this presented was that the committee did not review the software, and could not certify that it actually reproduced the HCM procedures reliably. This problem remains unresolved to this date.

7.2.5 The 2010 *Highway Capacity Manual*

The ink was barely dry on the 2000HCM when preparation for the next full edition was under way. The pace of research continued to escalate, with nine major NCHRP and two major FHWA projects between 2000 and 2009. In addition, the committee had introduced a second publication for the first time in its history, the *Highway Capacity Manual Applications Guidebook* [11]. The guidebook presented major applications involving not only HCM methodologies, but additional tools as well in an integrated fashion.

The 2010HCM [12] presented a broad range of new material and applications, including, but not limited to the following:

- A new methodology for interchange ramp terminals over a wide range of interchange types.
- A new methodology for roundabouts.
- Introduction of multimodal analysis of some segments and facilities.
- A comprehensive set of default values for use in planning applications.
- A set of service volumes based upon average annual daily traffic for planning use.
- New material on the impacts of active traffic management approaches

to increase capacity and improve performance.

The committee also determined that the *Applications Guide* should be incorporated into the HCM itself, eliminating two separate documents. Further, a research library of research sources used in the development of the 2010 and previous manuals was accumulated. As a result, the 2010HCM ballooned to over 2,000 pages of material. To accommodate this, much material was moved to an electronic archive, and was not part of the printed document. All purchasers of the HCM received the printed document and a passcode to access the electronic material. Even the printed material was massive, however, and was published in loose-leaf form using three binders.

7.2.6 The 2016 *Highway Capacity Manual*

Shortly after the publication of the 2010HCM, the SHARP program sponsored several major research efforts focusing on travel time reliability and managed lanes. After spending several million dollars on these, the agency wanted it incorporated immediately into the HCM, and provided over another million dollars for production of an update to the 2010HCM [13].

The 2016HCM was originally envisioned as an update to the 2010HCM, not a full new edition. When the sheer volume of new material was realized, however, the committee decided to produce it as a full sixth edition. This after-the-fact action, however, meant that there was not the overall review of concepts, format, and internal consistency that normally precedes production of a full new edition. Thus, the 2016HCM looks more like an update to the 2010HCM than a completely redeveloped manual.

The 2016HCM adds new material on travel time reliability, managed lanes, and assessment of active traffic management strategies. Many other refinements to methodologies have also been developed and approved by the Committee.

7.3 The Concept of Capacity

How big is the bucket, and what can it hold? Capacity as a general term is relatively simple. A 5-gallon pail holds 5 gallons of fluid. What if the pail were a membrane? The capacity of the membrane might be dependent upon the characteristics of the fluid, primarily its weight and density. What if I could put 5 gallons of fluid in the pail, but it was then too heavy to be moved? There are many subtleties involved in defining capacity, particularly where traffic facilities are the subject.

The 2016HCM defined capacity as follows:

> The capacity of a system element is the maximum sustainable hourly flow rate at which persons or vehicles reasonably can be expected to traverse a point or uniform section of a lane or roadway during a given time period under prevailing roadway, environmental, traffic, and control conditions. [Ref *13*, pgs 4-12]

The definition is fundamentally the same as in previous editions going back to 1985. There have been some minor changes involving wordsmithing, but the concept is basically unchanged.

There are several key elements of the definition:

- Capacity is defined as a *maximum hourly rate*. For most cases, the rate used is for the peak 15 minutes of the analysis hour, usually, but not always, the peak hour. In any analysis, both the demand and capacity must be expressed in the same terms.

- Capacity may be defined in terms of *persons or vehicles*. This reflects the growing importance of transit and pedestrians, High-Occupancy Vehicle (HOV) lanes, and multimodal facilities, where person capacity may be more important than vehicle capacity.

- Capacity is defined for *prevailing conditions*. Roadway conditions include geometric characteristics such as number of lanes, lane and shoulder widths, and free-flow speed (FFS). Traffic conditions refer to the composition of the traffic stream in terms of passenger cars, trucks, buses, and recreational vehicles. For interrupted flow facilities, roadway and traffic conditions include *traffic controls* that prevail, primarily signalization. The 2016HCM adds environmental conditions to this list. While most analyses assume good weather and other conditions, the manual now includes methodologies to address the impacts of poor weather or other ancillary conditions. The important concept is that any change in the prevailing conditions results in a change in capacity.

- Capacity is defined for a *point or uniform segment* of a facility. A new segment must be created wherever and whenever one of the prevailing conditions changes. A uniform segment has consistent prevailing conditions throughout its length.

- Capacity refers to the somewhat vague concept of *reasonable expectancy*. The 2000HCM contains a definition of this concept:

> Reasonable expectancy is the basis for defining capacity. That is, the stated capacity for a given facility is a flow rate that can be achieved repeatedly for peak periods of sufficient demand. Stated capacity can be achieved on facilities with similar characteristics throughout North America. [Ref *10*, pg 2-2]

This is an important concept, as capacity includes some elements of stochastic variation over time and space. Capacity, therefore, is *not* the highest flow ever observed on any facility with consistent characteristics. Further, it is possible to observe actual flow rates in excess of the stated capacity on occasion.

7.3.1 Capacity of Uninterrupted Flow Facilities

Base capacity values for uninterrupted flow facilities in the 2016HCM are shown in Table 7.2. These values are stated in terms of pc/h/ln under more or less ideal conditions. For application, they would be converted to prevailing conditions in veh/h/ln.

Table 7.2: Basic Capacity Values for Uninterrupted Flow Facilities—2016HCM

Free-Flow Speed (mi/h)	Basic Capacity (pc/h/ln) for:		
	Freeways	Multilane Highways	Two-Lane Highways
All	NA	NA	3,200 pc/h (both dir) 1,700 pc/h (one dir)
≥75	2,400	NA	NA
70	2,400	2,300	NA
65	2,350	2,300	NA
60	2,300	2,200	NA
55	2,250	2,100	NA
50	NA	2,000	NA
45	NA	1,900	NA

Table 7.2: Full Alternative Text

For freeways and multilane highways, capacity depends upon the free-flow speed (FFS). FFS is the theoretical speed when flow is "0" on a speed–flow curve. In practical terms, actual average speeds remain at the FFS for a range of low flow rates. In general, up through a flow rate of 1,000 pc/h/ln, observed average speeds may be used as an estimate of the FFS.

Capacity decreases with decreasing FFS. Freeways generally exhibit FFS from 55 mi/h to 75 mi/h, while multilane highways tend to be in the 45 mi/h to 70 mi/h range. This is because multilane highways have more roadside frictional elements that tend to depress speeds. Also note that for any given FFS, a freeway has a higher capacity than a similar multilane highway, again due to reduced roadside frictions on freeways.

For two-lane rural highways, capacity is unrelated to FFS. The key element in the operation of these facilities is that passing maneuvers take place in the opposing traffic lane. Practical limits on the ability to pass have more to do with capacity than any other element.

7.3.2 Capacity of Interrupted Flow Facilities

The key element affecting capacity on interrupted flow facilities is signalization. On an uninterrupted flow facility, the pavement is available for use at all times. At a traffic signal, the pavement is available to various movements for only the portion of time that the signal is green. Thus, for interrupted flow facilities, capacity is limited both by the geometric conditions of the facility and by the timing of traffic signals.

Because signal timing varies considerably, interrupted flow facilities start with the concept of "saturation flow rate." Saturation flow rate is the maximum flow rate that can be moved, assuming that the signal is effectively green 100% of the time. In the 2016HCM, the base value of saturation flow rate for interrupted flow facilities is 1,900 *passenger cars per hour of green time per lane* (pc/hg/ln) for urban areas with a population of 250,000 or more; 1,750 pc/h/ln for smaller areas. Like the values of Table 7.2, this saturation flow rate is for ideal conditions which include standard lane widths, no grades, no heavy vehicles, no turning vehicles, and others.

The relationship between capacity on an interrupted flow facility and saturation flow rate is simply based upon the proportion of time that the signal is green:

$$c=s\left(\frac{g}{C}\right) \quad [7\text{-}1]$$

where:

c = capacity (pc/h or pc/h/ln), s = saturation flow rate (pc/hg or pc/hg/ln), C = cycle length (s), and G = effective green time (s).

Equation 7-1 applies specifically to a signalized intersection approach. In terms of capacity, however, it is the capacity of intersection approaches that control the capacity of an interrupted flow facility. Thus, the equation deals with values associated with a signalized intersection, such as cycle length, effective green time, and saturation flow rate.

Chapters 22 and 23 contain a more complete discussion of signalized intersections and their analysis. This includes detailed discussions of signalization parameters and issues.

7.4 The Concept of Level of Service

What is traffic like today? This is a seemingly simple question motorists ask themselves each day. In its purest form, level of service is a descriptive scale that describes the quality of traffic service on a variety of facility types.

Although capacity is an important concept, operating conditions at capacity are generally poor. Further, capacity operations tend to be unstable, as the smallest of perturbations within the traffic stream can cause a breakdown. Thus, the need for some form of quality scale that traffic engineers can use is quite important.

7.4.1 In the Beginning: The 1950HCM

The 1950HCM did not include levels of service either as a basic concept or as a definitive scale. However, capacity was structured in such a way as to include some consideration of operating quality. Three different levels of "capacity" were defined in 1950:

- Basic Capacity: The maximum number of passenger cars that can pass a given point on a lane or roadway during one hour under the most nearly ideal roadway and traffic conditions which can possible be attained. [Ref *1*, pg 6]
- Possible Capacity: The maximum number of vehicles that can pass a given point on a lane or roadway during one hour under prevailing roadway and traffic conditions. [Ref *1*, pg 6]
- Practical Capacity: The maximum number of vehicles that can pass a given point on a roadway or designated lane during one hour without the traffic density being so great as to cause unreasonable delay,

> hazard, or restriction to the driver's freedom to maneuver under prevailing roadway and traffic conditions. [Ref *1*, pg 7]

In terms of current usage, basic capacity would be referred to as "capacity under ideal or base conditions," possible capacity would be referred to as "capacity," and practical capacity would be the maximum flow consistent with a level of service C or D operation in current terms.

The 1950HCM did not introduce level of service formally, but it laid the conceptual groundwork in its definition of different types of "capacity."

7.4.2 Level of Service Concept Introduced: The 1965HCM

With much discussion and a significant amount of controversy, the Highway Capacity Committee determined that there should be only *one* value of capacity in the second edition. As a result, another mechanism would have to be introduced to measure the quality of traffic service provided for operating volumes less than capacity. Capacity was simply defined as the *possible capacity* of the 1950HCM.

Level of service (LOS) was created to fill this need. Conceptually, it was a six-letter scale, A through F, that provided a labeling scale for various ranges of service quality. Like school grades, A was the best and F (in some sense) represented failure. Other boundary points were more difficult to establish.

The 1965HCM defined level of service as follows:

> Level of service is a term which, broadly interpreted, denotes any one of an infinite number of combinations of operating conditions that may occur on a given lane or roadway when it is accommodating various traffic volumes. Level of service is a qualitative measure of the effect of a number of factors, which include speed and travel time, traffic interruptions, freedom to maneuver, safety, driving comfort and convenience, and operating cost. [Ref *5*, pg 7]

The factors included in the definition are all measures that drivers directly perceive and experience. Specifically *not* included is volume or rate of flow. Drivers within a traffic stream cannot perceive volume; it must be observed from a stationary point at the roadside as vehicles pass by the location. Nevertheless, the 1965HCM assumed that low flow rates were directly related to higher levels of service:

> From the viewpoint of the driver, low flow rates or volumes on a given lane or roadway provide higher levels of service than greater flow rates or volumes on the same roadway. Thus, the level of service for any particular lane or roadway varies inversely as some function of the flow or volume, or of the density. [Ref 5, pgs 7, 8]

Unfortunately, this statement somewhat contradicts the definition, and is actually incorrect. Low volumes or flow rates can reflect low densities and free-flowing traffic *or* very high densities and congestion. The existence of low flow rates alone does not indicate a good level of service.

While the specification of the various levels of service would vary over time, the general concept behind their meaning was documented in a paper by Wayne Kittelson [14]:

- LOS E was intended to indicate capacity operations.
- LOS D was intended to reflect maximum sustainable flow levels being observed nationally at the time. It was based primarily on studies conducted by Karl Moscowitz on California freeways.
- LOS C was intended to replace the concept of “practical capacity” for urban facilities.
- LOS B was intended to replace the concept of “practical capacity” for rural facilities.
- LOS A was an afterthought, and was included to represent high operating quality that might be expected by users of toll facilities. It was included at the recommendation of the New Jersey Turnpike authority.

The LOS A condition seems almost humorous today: Operators of modern toll facilities have neither the interest nor ability to provide free-flowing conditions at all hours of the day on toll facilities. At the time, it was thought that if drivers paid directly for the trip, they deserved the highest quality experience.

In point of fact, for many types of facilities, the 1965HCM did not have methodologies that allowed for the estimation of many operational parameters. For uninterrupted flow facilities, levels of service were based on *both* operating speed and the volume-to-capacity (v/c) ratio. The models used, however, virtually always resulted in the v/c ratio being the deciding factor.

For many types of facilities, volumes or v/c ratios were used to identify a general level of service that was only verbally defined. For these cases, no numerical definitions of level of service were provided.

7.4.3 LOS Develops: The 1985HCM and Its Updates

The third edition of the HCM made some subtle changes in the definition of the concept, although it is quite similar to the 1965 version. Specifically, "operating cost" is removed from the noted list of factors associated with LOS. The definition also incorporated a specific reference to *user perceptions*, emphasizing that LOS should deal with things directly perceived by motorists.

Of greater importance is that by the time its 1997 update was completed, *every* methodology for every facility type included levels of service related to numeric values of a perceivable quality measure.

It is also in the 1985HCM that LOS descriptions, and indeed capacity, moved away from full-hour volumes and LOS designations to values based upon the worst 15-minute interval within the analysis hour.

7.4.4 LOS Moves On: The 2000HCM

Despite a great deal of discussion and study concerning LOS, the 2000HCM did not significantly alter the fundamental concept. The term "safety" was removed from the concept definition, as this had never been even contemplated as a practical factor for inclusion. The reference to "user perceptions," added in 1985, was removed from the definition, although it is discussed more completely in subsequent paragraphs.

7.4.5 Introducing User Perception Indices: The 2010HCM

While the idea of user perceptions has always been present in the concept of level of service, it was historically handled by ensuring that, to the extent possible, defining service measures were directly discernible to drivers and other road users.

In the period between 2000 and 2010, a body of research was developed that allowed for the assessment of what users would actually think about a given situation. The research was based upon measuring user reactions to various hypothetical conditions that were simulated in instrumented test driving portals. A major NCHRP study produced valuable results for consideration [15].

The research led to some interesting anomalies. It showed that users do not, in general, perceive six different levels of service—in some cases the number is as low as two. It also showed that many users perceive factors not normally related to the traffic stream as very important to their perception of service quality. These included such aspects as landscaping and lighting. Pedestrians were more focused on the traffic environment with which they interacted rather than the dynamics of the pedestrian stream they were in. Motorists considered truck presence as being negative to their perceived trip quality.

After much discussion and debate, levels of service based upon a user perception index were introduced for pedestrians and bicyclists at intersections. The index was made available for auto-users at intersections, but was not used to determine LOS.

Table 7.3 shows the service measures used to determine level of service for each of the facility types included in the 2016HCM.

Table 7.3: Service Measures in the 2016HCM

Category	Segment Type	Service Measure Used
Vehicular interrupted flow	Urban streets	Average travel speed
	Signalized intersections	Control delay
	Two-way STOP-controlled intersections	Control delay
	All-way STOP-controlled intersections	Control delay
	Roundabouts	Control delay
	Interchange ramp terminals	Control delay
Vehicular uninterrupted flow	Two-lane highways	%Time spent following average travel speed
	Multilane highways	Density
	Basic freeway segments	Density
	Freeway weaving segments	Density
	Freeway merge and diverge segments	Density
	Freeway facilities	Average density
Other Road Users	Transit	See Note 1
	Pedestrians	User perception index
	Bicycles	User perception index

1. Transit services use several different measures to define LOS; see the Transit Capacity Manual.

Table 7.3: Full Alternative Text

7.4.6 The 2016HCM and Beyond

There are no significant changes to LOS and its specific definitions in the 2016HCM. There have been many discussions of LOS and its potential future, but any major conceptual changes will be delayed until the next, now seventh, full edition of the HCM, which is not even under consideration at this time.

For freeway and urban street systems, new measures, however, have been developed that are very different in approach from the more traditional LOS designations. For both, the issue of travel time reliability is of great interest to federal and state transportation agencies, as current federal funding requires it to be a central issue in decision making. The approach is quite different, as it tries to quantify travel time percentiles for a given route over a full year. Various travel time percentiles could be used as a measure of overall system effectiveness, while they would not in any way describe facility operations for a given 15-minute time interval or hour. Separate measures for trucks are also being developed, to allow more direct assessments of system effectiveness for goods movement.

Current discussions range from eliminating LOS in favor of sets of specific quality measures, to basing it on travel time reliability, and everything in between. Some of the structural problems with existing LOS criteria are discussed in the next section.

7.4.7 Structural Issues with Level of Service

Many of the difficulties involved in applying and interpreting levels of service involve its step-function nature. LOS does not define a uniform operating quality; rather it identifies a *range* of operating quality.

LOS was initially developed for two reasons: (1) to provide a descriptor of service quality where numerical prediction of operational outcomes was not possible and (2) to provide a simple language that could be used to explain complex situations to decision makers and the public, who are most often not engineers.

The first reason no longer exists, as all LOS values are associated with specific predictions of numerical measures. The second continues. The trade-off between simplicity of communication and the complexity of the actual situation is, at best, a very difficult balance.

Some specific difficulties with LOS are as follows:

1. Use of a Step-Function Descriptor: As noted, LOS identifies a range of operating conditions. The variables used to define LOS are,

however, continuous. For example, speed is a continuous variable. If LOS X is defined as a range between 40 mi/h and 50 mi/h, a small change in speed from 49 mi/h to 51 mi/h would result in a change in LOS. A much larger change between 41 mi/h and 49 mi/h, would not result in a change in LOS. The problem is that as a step-function, LOS can essentially overstate small changes in operating conditions while masking much larger ones.

2. Can LOS Relate Complexity?: Because current methodologies result in the prediction of multiple, and often complex, operating parameters and conditions, it is difficult for a single-letter LOS to adequately describe those conditions. On the other hand, because of such complex results, the need for a simplified mechanism for communication is also heightened. In any event, it is increasingly important that professionals use all numerical parameters available in their analyses, not just an LOS, which might be based on only one or some subset of the available parameters.

3. The Issue of Relativity: Should LOS X be the same in New York City as it is in Peoria? As originally conceived, engineers would select a LOS appropriate to their local situations. Over time, however, the easy association of LOS with school grades has led to a less flexible interpretation. LOS E is always bad, even in New York City, where most motorists would throw a party if it could be achieved during a typical peak hour throughout the system. The truth is that drivers will accept more congestion in major urban areas than they will in less-densely developed areas.

4. LOS in Law and Regulation: Many state and local agencies now refer to LOS directly in laws and regulations associated with development. Development fees may be assessed on the basis of how LOS is changed when a new development is added. Development proposals may be disallowed if they have too large an impact on LOS. The problem for planners and engineers, and especially for members of the Highway Capacity and Quality of Service Committee, is that each time level of service definitions and criteria are changed (as they are in each new manual or update), laws and regulations are, by association, also being changed. This is clearly beyond the scope of the committee’s mandate, but the committee has no control over what others do with its methodologies and approaches.

5. Aggregation of LOS: A problem that is now becoming critical is how LOS should be interpreted at various levels of aggregation. LOS has traditionally been used to describe operations at a point or in a uniform segment. It is now being extended to cover long lengths of a facility, and discussions have taken place over the potential use of LOS for corridors and networks. The inconsistency in the HCM is somewhat jarring: For signalized intersections, there is a clear recommendation against aggregating delays for all approaches to obtain an intersection-wide LOS designation. It would hide more critical operations on individual approaches or lane groups. On the other hand, an LOS may be applied to a 12-mile segment of a freeway.

What does LOS mean if it is applied to a large section of freeway if virtually no motorist drives the entire length of it? What is the relationship of LOS to user perceptions if users don’t experience the entire facility? These questions get worse if LOS is expanded to corridors and networks. This trend can lead to a conundrum: Imagine waking up to a traffic broadcast announcing that “Cleveland is operating at LOS D today.” What would you, as a motorist, do with this information? On the other hand, a city planner might want to know this and be able to use it.

For the time being, LOS is here to stay. There appears to be some momentum in the profession to make major alterations in the concept and its implementation in methodologies or to eliminate it in favor of multiple numeric measures. Whether this translates to a major change in the seventh edition of the HCM remains to be seen.

7.5 Service Volumes and Service Flow Rates

Closely associated with the concept of LOS is the concept of service volumes (SV) or service flow rates (SF). SF is defined as the maximum rate of flow that can reasonably be expected on a lane or roadway under prevailing roadway, traffic, and control conditions *while maintaining a particular level of service*. It is essentially the most traffic that can be accommodated at LOS X. Figure 7.1 illustrates the concept.

Figure 7.1: Illustration of Levels of Service and Service Flow Rates

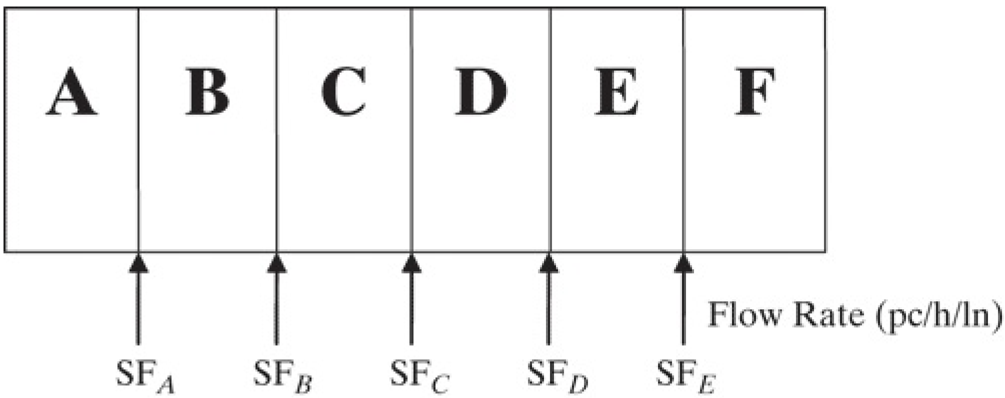


Figure 7.1: Full Alternative Text

The figure also highlights the issue of LOS as a step-function, that is, near the boundaries, a small change can signal a change in LOS, while a large change within a boundary does not. Note also that there are only five SFs, not six. LOS F represents unstable flow within a queue forming because arrivals exceed departures at a breakdown point.

For uninterrupted flow, the SF for LOS E, SF_E is synonymous with capacity. This relationship *does not* hold for many interrupted flow segments. In fact, it is difficult or impossible to precisely define SFs for some interrupted flow segments.

The term service volume (SV) is a vestige of earlier editions of the HCM in which capacity and level of service described conditions that existed over a full hour as opposed to the worst 15-minute interval within the hour. The relationship between SF and SV is the same as the relationship between an actual flow rate and an actual volume. They are related by the peak-hour factor (PHF), and SV may be computed from SF (or vice versa) as follows:

$SV_i = SF_i \times PHF$ [7.2]

where:

SV_i = service volume for LOS i, (veh/h), SF_i =
service flow rate for LOS i, (veh/h), and PHF = peak-hour factor.

7.6 The *v/c* Ratio and Its Use in Capacity Analysis

One of the most important measures resulting from a capacity and/or a level of service analysis is the *v/c* ratio—the ratio of the current or projected demand flow rate to the capacity of the facility. This ratio is used as a measure of the sufficiency of the existing or proposed capacity of the facility.

It is, of course, desirable that all facilities be designed to provide sufficient capacity to handle present and projected demands (i.e., that the *v/c* ratio be maintained at a value less than 1.00).

When estimating a *v/c* ratio, care must be taken to understand the origin of demand (*v*) and capacity (*c*) values. In existing cases, *true demand* consists of the actual arrival flow rate plus traffic that has diverted to other facilities, other times, or other destinations due to congestion limitations. If existing flow rate observations consist of *departing vehicles* from the study point, there is no guarantee that this represents true demand. On the other hand, a count of departing vehicles *cannot* exceed the actual capacity of the segment (you can't put 5 gallons in a 4-gallon jug!). Therefore, if a departing flow rate is compared to the capacity of a section, and a *v/c* ratio > 1.00 results, the conclusion must be that either (a) the counts were incorrect (too high) or (b) the capacity was underestimated. The latter is the usual culprit. Capacities are estimated using methodologies presented in the HCM. They are based upon national averages, and on the principle of "reasonable expectation." Actual capacity can be larger than the estimate produced by these methodologies.

When dealing with future situations and forecast demand flow rates, both the demand flow AND the capacity are estimates. A *v/c* ratio in excess of 1.00 implies that the forecast demand flow will exceed the estimated capacity of the facility.

When the true ratio of demand flow to capacity exceeds 1.00 (either in the present or the future), the implication is that queuing will occur and propagate upstream from the segment in question. The extent of queues and the time required to clear them depends on many conditions, including the time period over which $v/c > 1.00$ and by how much it exceeds 1.00. It depends upon the demand profile over time, as queues can start to dissipate only when demand flow decreases to levels less than the capacity of the segment. Further, queuing drivers tend to seek alternative routes to avoid congestion. Thus, the occurrence of $v/c > 1.00$ often causes a dynamic shift in demand patterns that could significantly impact operations in and around the subject segment.

In any event, the comparison of true demand flows to capacity is a principal objective of capacity and level of service analysis. Thus, in addition to level-of-service criteria, the v/c ratio is a major output of such an analysis.

7.7 Closing Comments

The HCM is one of the most basic and frequently used documents in the traffic engineering profession. The concepts of capacity, level of service, SFs, and v/c ratio are all important to understand, as they are the driving principles behind the design and development of many highway capacity analysis procedures.

References

- 1. *Highway Capacity Manual*, U.S. Government Printing Office and the Highway Research Board, Washington, D.C., 1950.
- 2. *Public Roads*, Vol. 25 (10), Bureau of Public Roads, Washington, D.C., October 1949.
- 3. *Public Roads*, Vol. 25 (11), Bureau of Public Roads, Washington, D.C., November 1949.
- 4. *Highway Research Bulletin 167* (6 papers), Transportation Research Board, Washington, D.C., 1957.
- 5. Reprinted with permission from *Highway Capacity Manual*, Special Report 87, Transportation Research Board, © 1965 by the National Academy of Sciences, Courtesy of the National Academies Press, Washington, D.C.
- 6. "Interim Materials on Highway Capacity," *Circular 212*, Transportation Research Board, Washington, D.C., 1980.
- 7. "Highway Capacity Manual," *Special Report 209*, Transportation Research Board, Washington, D.C., 1985.
- 8. "Highway Capacity Manual," *Special Report 209* (as updated), Transportation Research Board, Washington, D.C., 1994.
- 9. "Highway Capacity Manual," *Special Report 209* (as updated), Transportation Research Board, Washington, D.C., 1997.
- 10. Reprinted with permission from *Highway Capacity Manual*, Transportation Research Board, © 2000 by the National Academy of Sciences, Courtesy of the National Academies Press, Washington, D.C.
- 11. Kittelson, W., et al., *Highway Capacity Manual Applications Guidebook*, Transportation Research Board, Washington, D.C., 2003.

- 12. *Highway Capacity Manual*, Transportation Research Board, Washington, D.C., 2010.

- 13. Reprinted with permission from *Highway Capacity Manual*, Transportation Research Board, © 2010. by the National Academy of Sciences, Courtesy of the National Academies Press, Washington, D.C.

- 14. Kittelson, W., “Historical Overview of the Committee on Highway Capacity and Quality of Service,” *Transportation Research Circular E-C018: Proceedings of the Fourth International Conference on Highway Capacity,* Transportation Research Board, Washington, D.C., 2001.

- 15. Dowling, R., et al., “Multilmodal Level of Service Analysis for Urban Streets,” *Final Report*, NCHRP Project 3-70, Transportation Research Board, Washington, D.C., 2007.

Problems

1. 7-1. Explain the difference between “capacity” and a “service flow rate” for an uninterrupted flow facility.

2. 7-2.Explain the difference between the following terms:

    1. capacity under ideal conditions,

    2. capacity under prevailing conditions

3. 7-3. The following service flow rates (SFs) have been determined for a six-lane freeway in one direction:

    - SF_A = 3,000 veh/h
    - SF_B = 4,000 veh/h
    - SF_C = 4,800 veh/h
    - SF_D = 5,600 veh/h
    - SF_E = 6,300 veh/h

    1. What is the capacity of this segment?

    2. Determine the service volumes for this freeway, which has a PHF of 0.90.

4. 7-4. Explain the differences between freeways, multilane highways, and two-lane rural highways. What are the key differentiating

characteristics of each?

5. 7-5. A signalized intersection approach has a saturation flow rate of 1,200 pc/hg/ln. If the approach has three lanes, and the signal has a 40 s green time within a 70 s cycle length, what is the capacity of this approach?

Chapter 8 Intelligent Transportation Systems

For many decades, advances in computing power and the desire to improve traffic operations allowed traffic systems to become more "intelligent" in a number of ways:

- Central computers were used to control traffic signals in a number of pioneering cities, including San Jose CA, Overland Park KS, Toronto Canada, and New York City, NY.

- Computers were likewise used to meter traffic on freeway ramps and systems of freeway ramps.

- Researchers focused on improved control policies for traffic control, on surface streets and on freeways, in moderate to heavy to oversaturated conditions.

- Variable message signs were used to provide route advisory information to drivers, based upon traffic conditions detected with a variety of sensors.

By the late 1980s, the profession began discussing the overall approach as "intelligent vehicle highway systems" (IVHS). In 1991, IVHS America was formed as a public/private partnership to plan, promote, and coordinate development and deployment of IVHS in the United States. Other countries had comparable organizations and initiatives, with widespread activity in the United States, Canada, Japan, and Europe.

The terminology and the concept quickly evolved to an emphasis on *Intelligent Transportation Systems (ITS),* with a multimodal focus. Consistent with this, IVHS America became ITS America [1], which has grown to a major organization supporting the development of standards, providing a forum for professional dialog, and including a network of state-level ITS organizations. Comparable organizations exist in many countries, and *ITS International* [2] is a noteworthy publication. Key meetings include the ITS World Congress, ITS America Annual Meeting, Intertraffic, sessions at the Transportation Research Board (TRB) Annual Meeting, the International Road Federation (IRF) Meeting, and the Institute of Transportation Engineers (ITE) Annual Meeting.

The focus is not merely on moving vehicles more efficiently. It extends to providing routing information based upon real-time and historic information, traffic advisories, and finding and even scheduling parking. It also extends to multimodal transfer points, schedule coordination, information on current status and pending arrivals, and information by kiosk and smartphone apps for transit and for trip planning.

The investment in ITS has been truly substantial to date, and much space could be devoted to sample projects. But the authors believe that the reader is better served by indicating concepts and web resources, so that the most current information is used: up-to-date standards, rules and policies, and initiatives.

As substantial as the effort has been to date, the coming years (this was written in mid-2017) are likely to see an accelerated pace of ITS applications, and even of the concept of what constitutes a complete, integrated intelligent system. The forces at work include the following:

- For over a decade, the momentum has been building, key standards have been established, the infrastructure has been upgraded, and an experienced professional workforce has been developed.
- Manufacturers are embracing the perceived market forces for *connected and/or autonomous vehicles*; at the same time, they are embracing electric and hybrid vehicle expectations. Being innovative, early to market, and committed to such future vehicles has exploded as a competitive force.
- Agencies and government units are responding to the need for on-road testing, authorizations, and regulations.
- *Smartphones and a host of smartphone apps* have fundamentally changed how much information people *can* get, and much more they *expect* to get.
- Aided by a supportive communications infrastructure, there is a flood of information creating a *data-rich environment*. Routing

information, travel times, crowdsourced information, transit performance, taxi and for-hire vehicle data all contribute to future ITS designs.

8.1 An Overview

ITS has as its primary objective increasing system efficiency and safety, through the use of a set of tools, techniques, technology, standards, and best practices.

While attention is often drawn to the motoring public, ITS has always had a strong emphasis on commercial vehicle operations (CVO), freight, transit, and advanced traveler information systems. The emphasis on pedestrians and bicycles has grown, and safety is a keystone of ITS.

The USDOT ITS Joint Program Office has an excellent resource in its *ITS ePrimer* website [3]. It contains a set of 14 modules that provide comprehensive information on the range of topics shown in Table 8.1.

Table 8.1: The 14 Modules of the USDOT ePrimer Website

Module 1: Introduction to ITS
Module 2: Systems Engineering
Module 3: Transportation Management Systems
Module 4: Traffic Operations
Module 5: Personal Transportation
Module 6: Freight, Intermodal, and CVO
Module 7: Public Transportation
Module 8: Electronic Toll Collection and Pricing
Module 9: Supporting ITS Technologies
Module 10: Rural and Regional ITS Applications
Module 11: Sustainable Transportation
Module 12: Institutional Issues
Module 13: Connected Vehicles
Module 14: Emerging Issues

(Source: https://www.pcb.its.dot.gov/ePrimer.aspx.)

Table 8.1: Full Alternative Text

Table 8.2 shows the outline of Module 1, “Introduction to ITS,” as an illustration of the content of each of the modules (several modules have more detailed outlines). Among the topics listed, note the national ITS architecture and refer to Figure 8.1 displaying that architecture.

Table 8.2: The Table of Contents for ePrimer Module 1

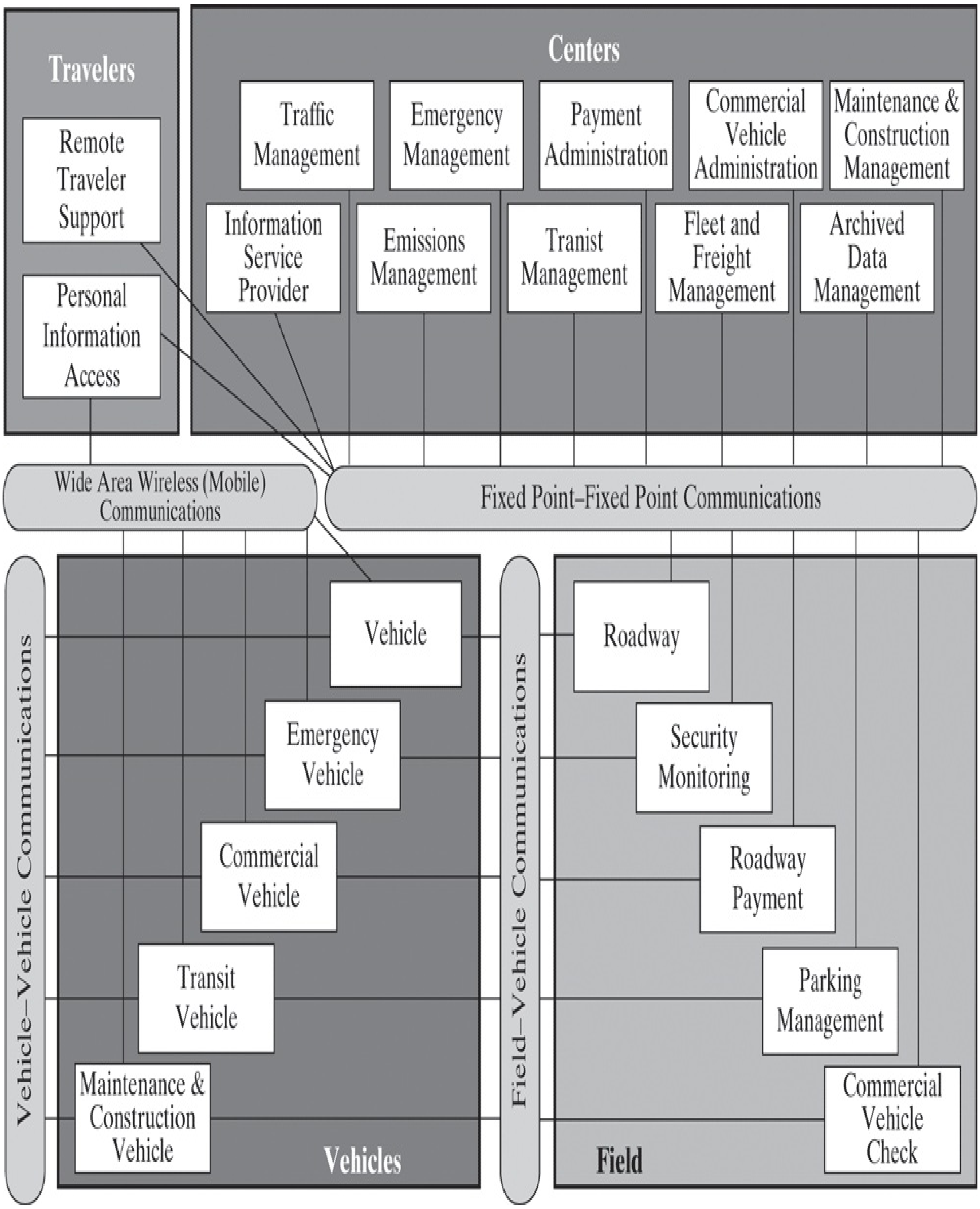
Introduction
ITS Overview
Transportation Challenges
Benefits of ITS
History of ITS
National and Regional ITS Architectures • Architectural Layers • Regional Architecture • National ITS Standards
Growth of ITS Deployment
Future Vision for ITS
ePrimer Overview
References

(Source: https://www.pcb.its.dot.gov/ePrimer.aspx, Module 1.)

Table 8.2: Full Alternative Text

Figure 8.1: National ITS

Architecture



(Source: https://www.pcb.its.dot.gov/ePrimer.aspx, Module 3.)

Figure 8.1: Full Alternative Text

The reader is strongly advised to (1) focus on key concepts that are part of the common dialog in ITS applications, including

- system architecture
- standards
- V diagram

and (2) gain a deeper knowledge by visiting [3] and inspecting each of the modules, so as to be aware of the content.

8.2 ITS Standards

An excellent source of information on current ITS standards is the USDOT website http://www.standards.its.dot.gov/ [4]. This provides access to the very detailed standards that have been developed and are current.

The web site notes that

> The ITS Standards Program has teamed with standards development organizations and public agencies to accelerate the development of open, non-proprietary communications interface standards [4].

To a very large extent, the key words are “open” and “nonproprietary.”

This site is the repository and path to such information as:

- existing ITS standards;
- ITS standards training: a self-paced series of 49+ modules (refer to Table 8.3 for an illustrative set) plus 21+ transit-related modules;
- international harmonization;
- deployment tools;
- lessons learned;
- fact sheets;
- deployment statistics and asset viewer;
- technical assistance and training opportunities.

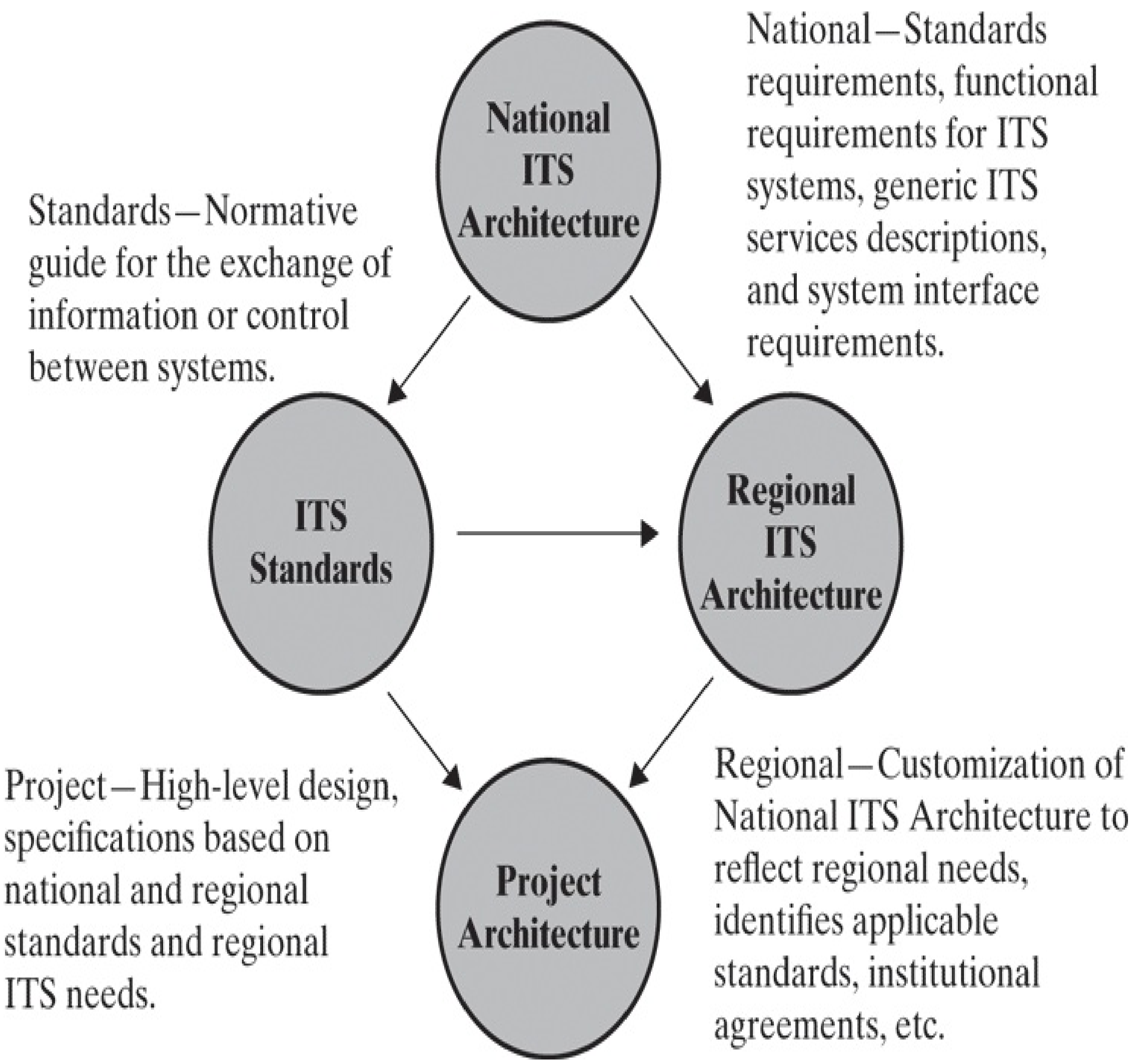
Table 8.3: Illustrative Training Modules from Reference [4]

Module #	Module Title	Module ID
5	Introduction to ITS Standards Testing	T101
6	Details on Acquiring Standards-based ITS Systems	A201
7	Identifying and Writing User Needs When ITS Standards Do Not Have SEP Content	A202
28	Building an ITS Infrastructure Based on the Advanced Transportation Controller (ATC) 5201 Standard—Part 1 of 2	A207a, Part 1 of 2
29	Building an ITS Infrastructure Based on the Advanced Transportation Controller (ATC) 5201 Standard—Part 2 of 2	A207b, Part 2 of 2
49	Applying Your Test Plan to a Transportation Sensor System (TSS) Based on the NTCIP 1209 Standard v02	T312
50	Applying Your Test Plan to Ramp Meter Control (RMC) Units Based on the NTCIP 1207 Standard v02	T309
51	Using the ISO TS 19091 Standard to Implement V2I Intersection Applications Introduction	CV271

Table 8.3: Full Alternative Text

The standards themselves are cited by their National Transportation Communications for ITS Protocol (NTCIP) numbers, and fit within an overall National ITS Architecture. Figure 8.2 shows the role of the standards relative to the systems architecture.

Figure 8.2: ITS Standards as Related to Architecture



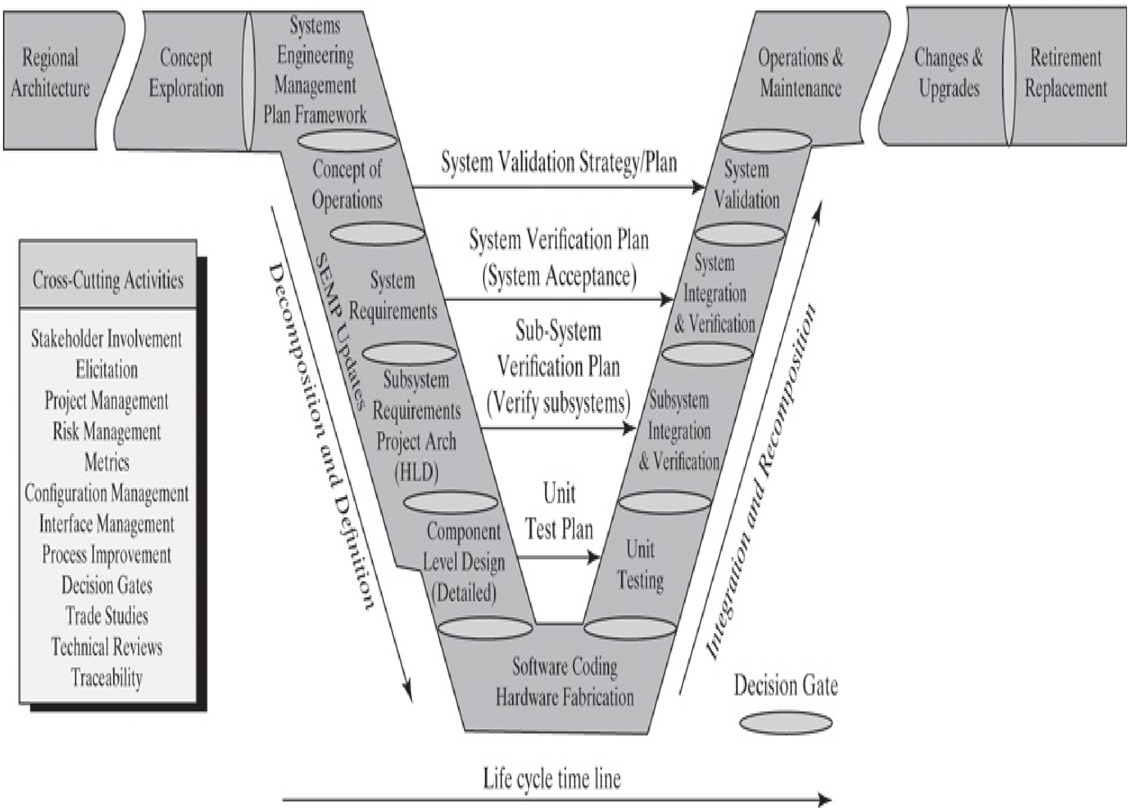
(Source: http://www.standards.its.dot.gov/)

Figure 8.2: Full Alternative Text

8.3 ITS Systems Engineering Process

Reference [3] notes that many process models have been developed for the systems engineering approach, but notes that the *V process model* shown in Figure 8.3 has gained wide acceptance in the standards community, and is the one chosen by USDOT [5, 6].

Figure 8.3: V Process Model, as USDOT SEP



(Source: https://www.pcb.its.dot.gov/ePrimer.aspx, Module 2.)

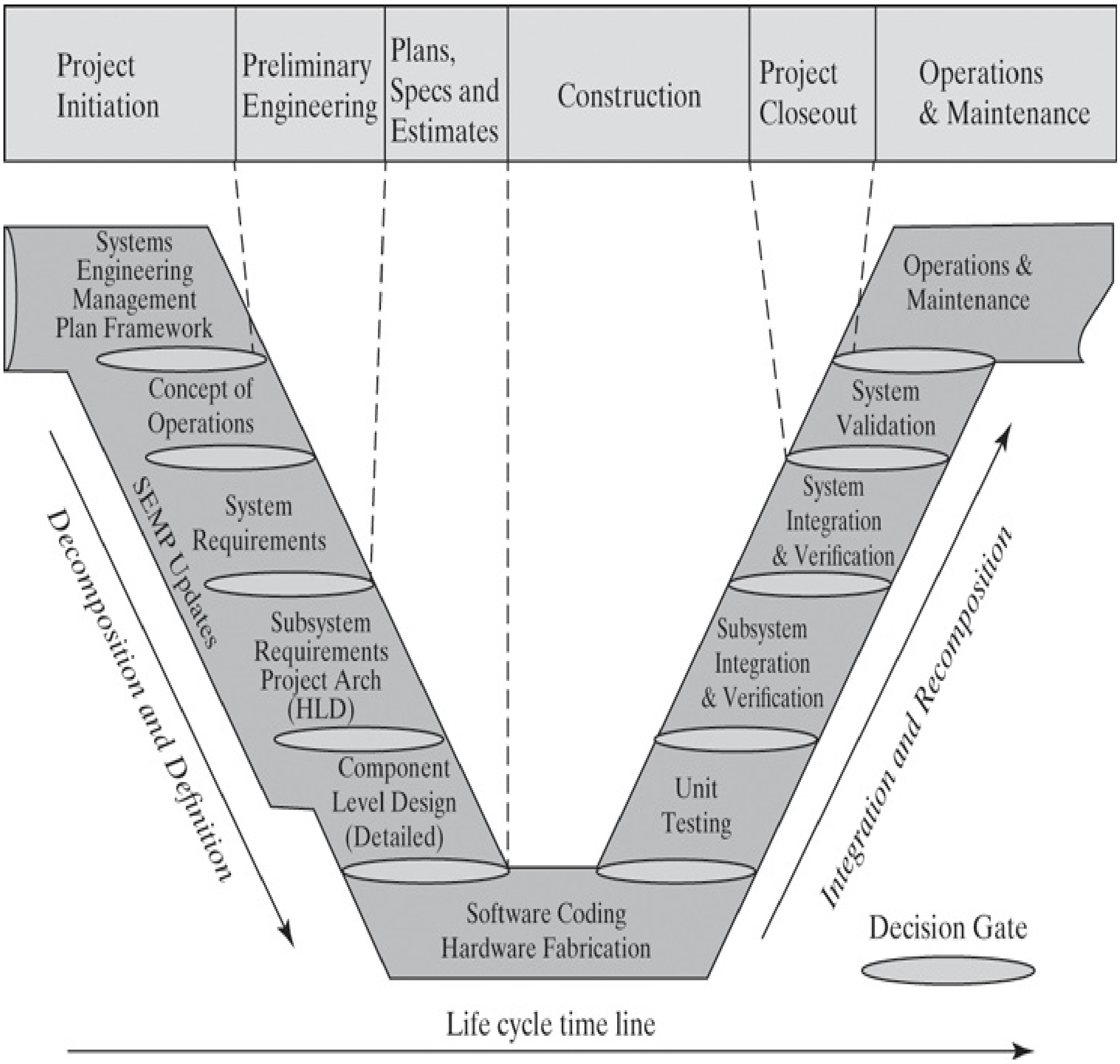
Figure 8.3: Full Alternative Text

The problems at the end of this particular chapter direct the reader to seek additional knowledge of a number of items, by going to the referenced websites, using the hyperlinks, and going into further detail.

Specifically, Reference [6] has hyperlinks for each step in the V process model, providing substantially more information. For the present purposes, it is sufficient to take note of the steps, the names of which tend to be self-explanatory. Reference [3] observes that as one moves from one step to the next, there may be iteration back to a prior step, particularly on the left side of the V.

Figure 8.4 shows the same V process model (or “V diagram”) as related to the traditional transportation development process.

Figure 8.4: Relation of the V Process Model to the Traditional Transportation Development Process



(Source: https://www.pcb.its.dot.gov/ePrimer.aspx, Module 2.)

Figure 8.4: Full Alternative Text

8.4 ITS-Related Commercial Routing and Delivery

Routing systems are of great importance to trucking and service vehicles. This is a large specialty market, with software available that can compute long-haul routing, urban routing, and provides answers that take into account the set of scheduled pickup and delivery points. Dynamic re-routing, as new pickups are added en route, is feasible.

Today, it is commonplace for package delivery services (FedEx, UPS, others) to offer real-time package tracking to customers on their web sites. Using bar-code scanning technology and wireless communication, packages are tracked in detail from origin to destination. At the delivery point, the driver uses a computer-based pad to record delivery time and often receiver's signature. This information is available in virtually real time to the sender.

Clearly, the package delivery services found a differentiating service feature that has rapidly been adopted in a highly competitive industry. What was special a few years ago has now become the expected standard of service.

The same data allow the service providers to obtain a wealth of data on the productivity (and down time) of their vehicles and drivers, and on the cost of delivery in various areas.

8.5 Sensing Traffic by Virtual and Other Detectors

As already mentioned, smartphones and other devices can provide such a wealth of data that the traffic engineer and manager must worry about being overwhelmed with data, and must plan for that day.

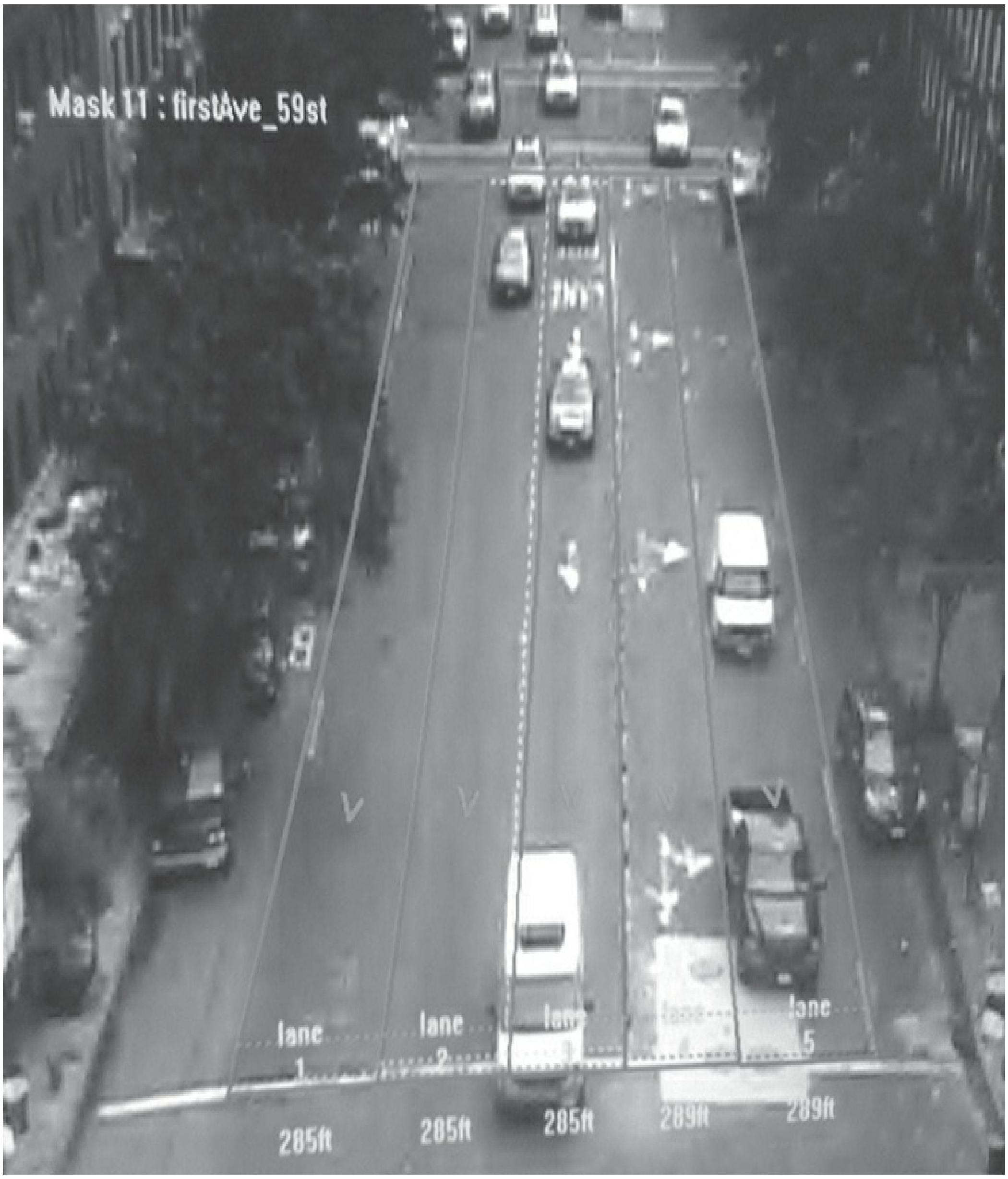
Indeed, electronic toll collection (ETC) tags such as E-ZPass (used by a number of states along the Northeast Corridor) do already provide the opportunity to sample travel times and estimate volumes, and ETC readers can be used for such purposes even where tolls are not collected.

But the day when these newer technologies can be the *primary* source of data has not yet arrived. Indeed, there are still a number of years in which other innovative and traditional sensors will be the primary source of traffic information. The mainstay of traffic detection for decades has been magnetic sensing of vehicles by a loop installed into the road surface. It is still important in many jurisdictions, but some have moved away from loops because of cuts made into the pavement during maintenance, or weather issues such as frost heave, or relative cost.

Three alternatives, coming into greater use, are as follows:

1. *Virtual detectors* generated in software, using a standard or infrared video camera to capture an image of the traffic and generate estimates of flow, speed, queue, and/or spatial occupancy. Refer to Figure 8.5 for an illustration. Using such a tool, the transportation professional can “locate” numerous “detectors” essentially by drawing them on top of the intersection image, and depending upon the software to process the data.

Figure 8.5: Software-Based Virtual Detector, Using Video Camera



(Source: NYC Department of Transportation.)

Figure 8.5: Full Alternative Text

2. *Microwave detectors*, used to identify flows and point occupancy. The detectors can be placed over specific lanes, or in a “side-fire” mode covering a number of lanes. In some applications, one detector can be used to cover several lanes and even both directions. Refer to Figure 8.6 for an illustration. Such detectors can be used in a cluster, with wireless data transmission from one point in the cluster.

3. *Wireless Detectors Imbedded in Pavement*, such as illustrated in Figure 8.7. Models are available for presence or count, and are

generally imbedded one per lane. The units are approximately 3 × 3 inches, 2 inches deep. Compared to loops, the manufacturer claims an easier install and less susceptibility to being broken than a loop with a much larger footprint. Figure 8.8 shows a travel time map generated from sets of such detectors, using software to identify the “signature” or profile of individual vehicles.

Figure 8.6: Illustration of Microwave Detectors

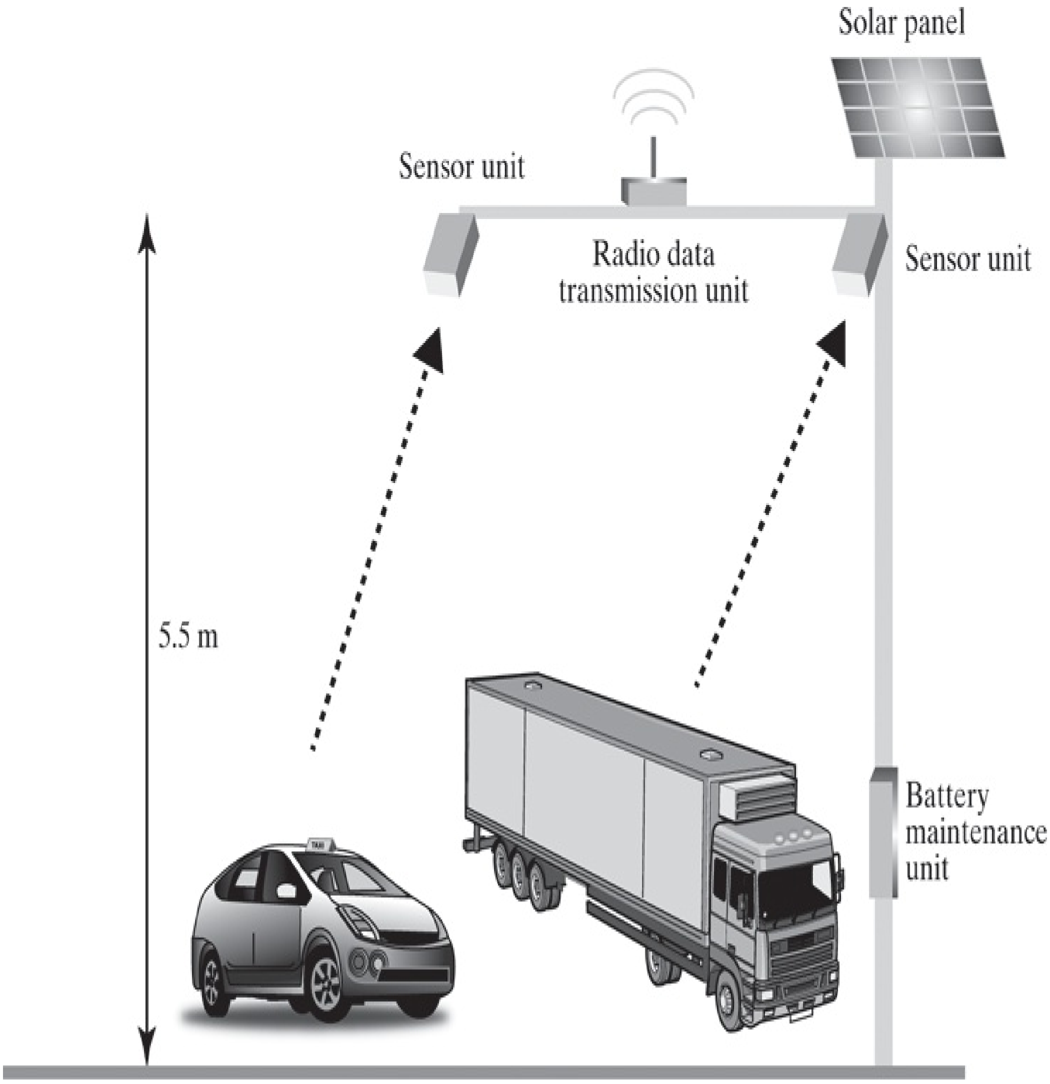
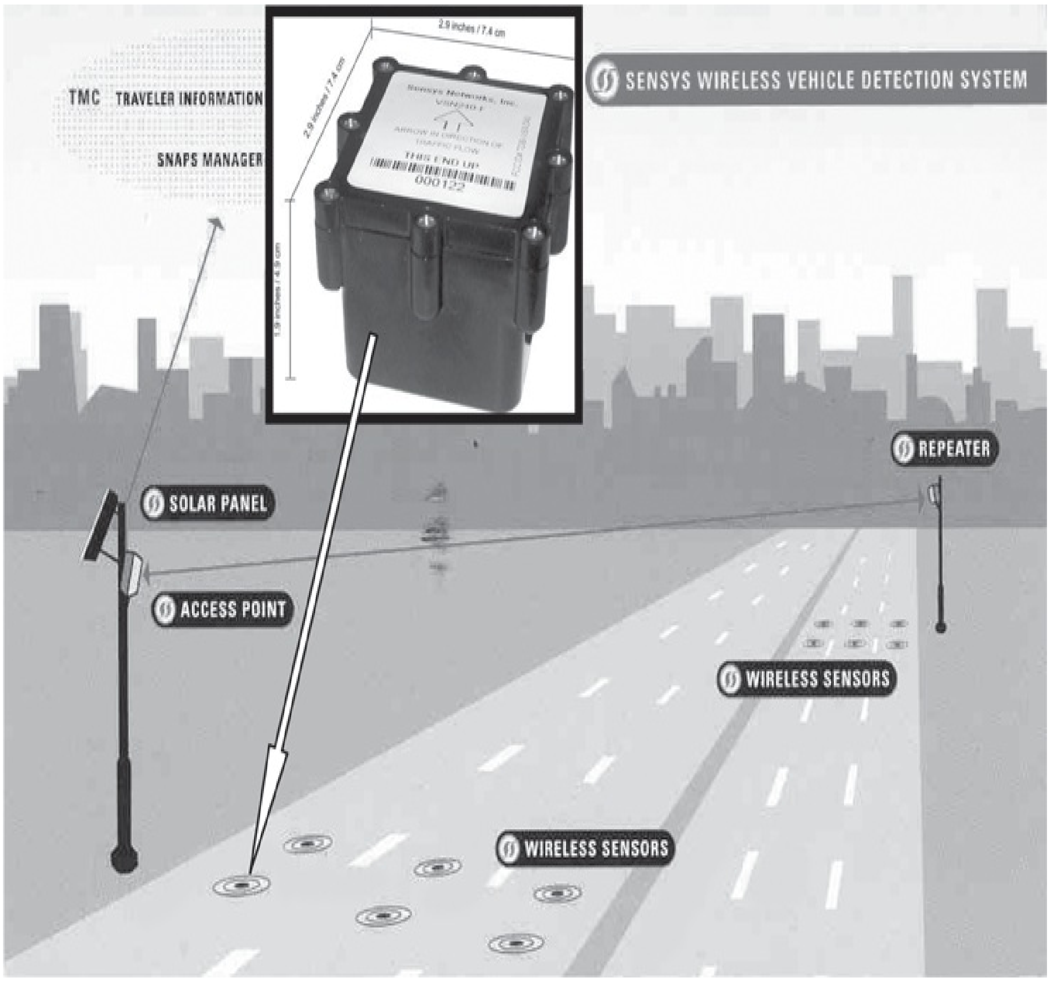


Figure 8.6: Full Alternative Text

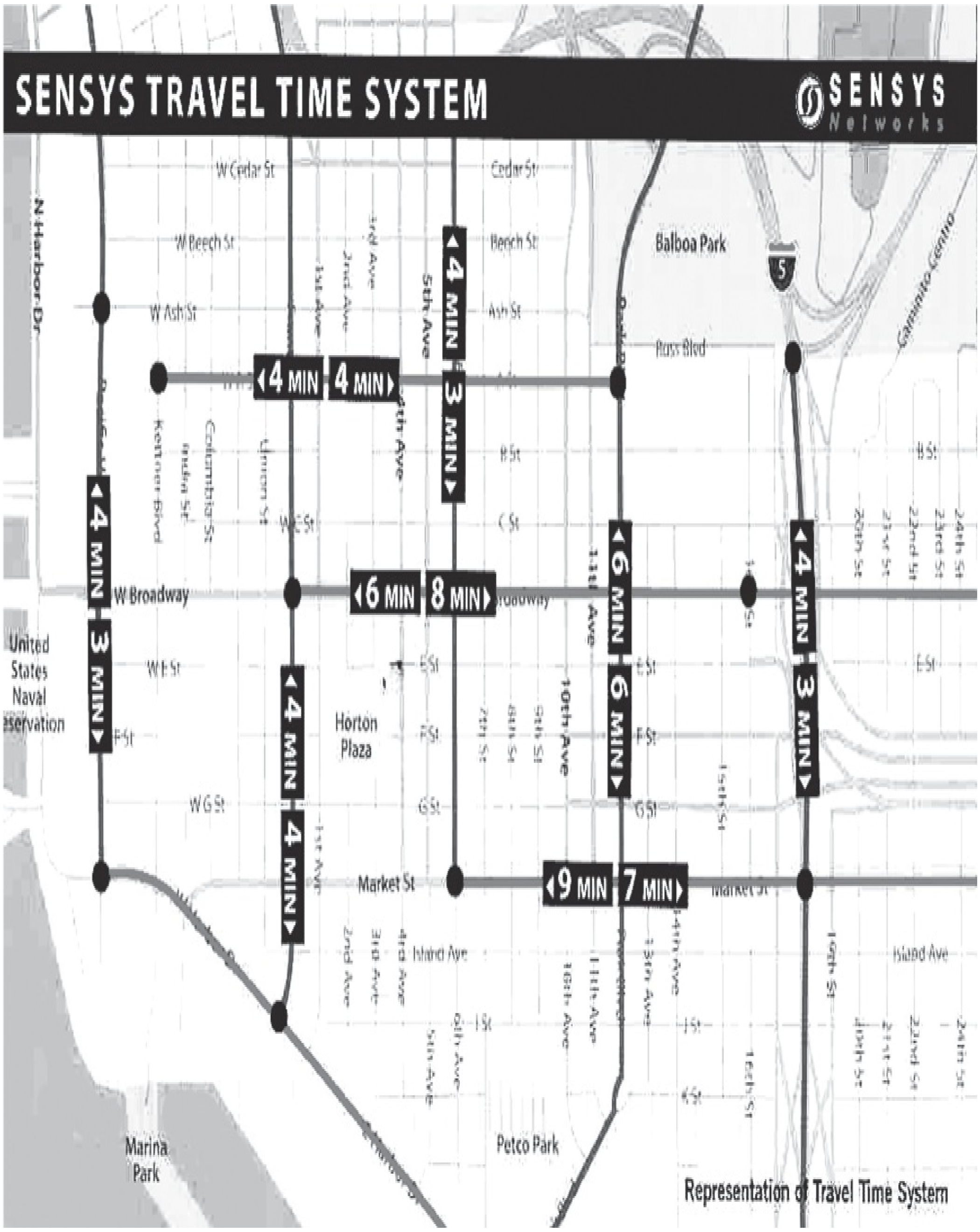
Figure 8.7: Wireless Detectors Imbedded into Road



(Source: Courtesy of Sensys Networks, Inc.)

Figure 8.7: Full Alternative Text

Figure 8.8: Travel Time Estimates from Sets of Figure 8.7 Detectors



(Source: Courtesy of Sensys Networks, Inc.)

Figure 8.8: Full Alternative Text

There are variations on these types, including a detector that uses a 360-degree video image for “area occupancy” detection.

The use of infrared imaging allows vehicles to be detected in a variety of weather conditions. Another variant is the use of sophisticated algorithms based upon coverage of the underlying pavement image allows data to be collected from stationary traffic as well as moving traffic.

8.6 Connected Vehicle Pilot Studies

Reference [4] lists a set of “Hot Topics” and a set of “Research Areas.” The Connected Vehicle is featured prominently on both lists.

Of special note at the time of this writing is the September 2016 award of three design/build/test pilot deployment projects by USDOT for three sites:

1. State of Wyoming—sections of Interstate 80 (I-80) using 75 roadside units, 400 instrumented fleet vehicles, and traveler information through the Wyoming 511 app and its commercial vehicle operator portal (CVOP). The WYDOT website notes that truck volumes on I-80 can reach 70% during seasonal peaks and extreme weather (blowing snow in winter, fog and high winds in summer).

2. New York City, NY—three distinct areas in the boroughs of Manhattan and Brooklyn: a 4-mile segment of the FDR Drive (a limited-access facility), 4 one-way corridors in Manhattan, and a 1.6 mile segment of Flatbush Avenue Brooklyn approaching the Manhattan Bridge. Some 5800 cabs, 1250 MTA buses, 400 commercial fleet delivery trucks, and 500 NYC vehicles will be equipped with connected vehicle technology. There will also be a focus on reducing vehicle–pedestrian conflicts using in-vehicle pedestrian warnings and personal devices equipping some 100 pedestrians.

3. Tampa-Hillsborough Expressway Authority Pilot—focus on rush-hour collision avoidance and wrong-way entry prevention on the Expressway; traffic flow optimization on several arterial segments, including pedestrian safety and bus priority on some; streetcar safety; and traffic management.

USDOT looks at these pilots as the “most complex and extensive deployment of integrated wireless in-vehicle, mobile device, and roadside technologies.”

8.7 Variable Pricing

In many areas, variable pricing of road usage has arrived, albeit in the form of discounts for using ETC.

Although congestion pricing (the earlier and more common term) has proven extremely controversial in many discussions, there are two realities that the reader will face in coming years:

1. Funding mechanisms for facility maintenance and operation need to be found. The most visible means has been a gasoline tax at point of purchase (so many cents per gallon), but this historic mechanism (a) is not sensitive to either time of day or total mileage traveled, if in the latter case one considers *relative* fuel efficiencies, (b) loses meaning if alternate fuels are used, including electricity, (c) generates less revenue if mileage efficiency improves, (d) is difficult to change for political reasons, even while (e) social justice issues might be exacerbated by any linkage of older vehicles, gas-powered vehicles, and household income skewing the tax burden.

2. ITS technology can not only make for safer and more efficient travel but can also make variable pricing much more feasible than in the past, because it is becoming pervasive. Indeed, some toll facilities no longer accept cash. And some non-toll facilities—and many surface streets—have smart sensors for ITS uses.

It can be argued that variable pricing is an entirely different subject from ITS, and the two should not be comingled. But it is wise for the reader to appreciate that two strong forces are at work, concurrently, and will influence the reader’s professional life: the need to equitably fund transportation facilities is pressing and a solution must be found, while at the same time the advance of technology and the benefits of ITS are putting in place the infrastructure that can make some collection methods both cost-effective and attractive. Indeed, one could argue that—putting aside ITS-specific infrastructure—the smartphone itself provides the needed infrastructure.

For further reading, refer to Module 8 of [3], which is titled “Electronic

Tolling and Pricing” and addresses pricing strategies, funding and financing, and value pricing (as well as congestion pricing).

8.8 Closing Comments

This chapter intentionally skips details of specific ITS systems, because (a) the field is moving rapidly and any “snapshot” of its present state is sure to be dated rapidly, perhaps even by the publication date of the text, and (b) the real issue is for the reader to be prepared to expand his/her view of providing transportation service in a highly competitive market in which computing, communications, and web services are being used in novel ways.

Furthermore, the evolving roles of public and private sectors—in some ways, structure vis-à-vis market responsiveness—should draw the reader’s attention. Today’s “right answer” can be swept away by what the enabling technologies make available.

There is another fundamental issue for the reader to consider: manufacturers need to devise products that are both more attractive and differentiated (at least in the short term, until the competition copies success). In that world, transportation data and information is *not* an end in its own right—the traditional view in our profession—but rather it is a product enhancement, or a service. Private sector forces may provide a data-rich environment for transportation professionals as a by-product of their own work, *and at an innovative pace driven by that work and its market.* And this pace can exceed the traditional pace of public sector planning and innovation, and the orderly process of standardization.

References

- 1. http://itsamerica.org/
- 2. http://www.ropl.com/magazines/its-international/
- 3. https://www.pcb.its.dot.gov/ePrimer.aspx
- 4. http://www.standards.its.dot.gov/
- 5. *Systems Engineering for ITS—an Introduction for Transportation Professionals*, USDOT, September 2007, http://ops.fhwa.dot.gov/publications/seitsguide/seguide.pdf
- 6. *Systems Engineering Guidebook for ITS*, Version 3.0, FHWA and Caltrans, November 2009, https://www.fhwa.dot.gov/cadiv/segb/

Problems

1. 8-1. Read the History of Intelligent Transportation Systems. www.its.dot.gov/index.htm, FHWA-JPO-16-329.

2. 8-2. Read Module 8 of [3], and prepare for a class discussion of the value of, and the potential need for, variable pricing related to transportation facility usage. If necessary, do additional web searching to be aware of past work on road usage fees based upon vehicle-miles traveled (VMT) rather than fixed tax per gallon of gasoline. Be prepared to contribute to a discussion, with supporting materials, on whether variable pricing by demand level is equitable or not.

3. 8-3. Do a web search on “ITS V2V” and then broaden your search to include V2I and I2V. Find and submit definitions of each, lists of related technologies, and graphics that clarify the concepts for a general audience. Be sure to identify sources.

4. 8-4. Go to Reference [6], and go through the hyperlinks for each step in the V process model, in sufficient detail that you fully understand the diagram in Figure 8.3. Summarize them in a table, with the first column being the step and the second being your summary of each (50 words or less, per step).

5. 8-5. Do a web search on “ITS probe vehicles,” and also look for any related discussion in the source cited in Problem 8-1. This chapter did not address networks of probe vehicles explicitly, because the authors took the view that whereas probe vehicles were once very relevant, the pervasiveness of smartphones, ETC devices, and such makes the need for a special network of probe vehicles a bit outdated. Either

support this view or argue against it, using and citing sources for your case.

6. 8-6. Consider that in a large region, you would like to obtain information on truck origins and destinations and the type of cargo carried (including empty or “deadhead” loads). What ITS infrastructure can be used for this purpose? Is it likely to be in place? Is a special survey needed, or special measurements? What are the most common classifications of commercial cargo, preferably in eight or fewer categories?

7. 8-7. Refer to the section on “Connected Vehicle Pilot Studies.” Search the literature, starting with the identified source(s), and write up a summary of the current progress or results of the three pilot studies cited. What lessons learned were listed, if any?

8. 8-8. Refer to the “virtual detector” discussion in the chapter, which addresses software (hopefully user-friendly) that can be used to place both point and area detectors on a video image. Write a summary of the current state of the art in such detectors, drawing on graphics and text from the web, properly sourced. Consider whether any existing technology can work with infra-red imaging, and what the limits of the technology are (weather, light, etc.), again properly sourced. Limit yourself to 5–8 pages.

Part II Traffic Studies and Programs

Chapter 9 Traffic Data Collection and Reduction Methodologies

The starting point for most traffic engineering is a comprehensive description of the current state of the streets and highways that comprise the system, current traffic demands on these facilities, and a projection of future demands.

This requires that information and data that can quantitatively describe the system and the travel demands on it be assembled. Given the size of the highway system and the reality that demands vary by time and location, assembling this information is a massive task. Nevertheless, data must be collected and reduced to some easily interpreted form for analysis.

Indeed, the data must be collected with a *prior* awareness of the expected uses and analyses of the data, so that it can be properly specified and acquired, and so that—in some scenarios—competing hypotheses are tested.

Further, the meaning of “traffic data” has expanded in recent years from a focus on motorized vehicles to a more holistic approach focusing on

- multiple modes (for example, motorized vehicles, pedestrians, bicycles, transit, and for-hire vehicles including traditional taxis and others);
- dedicated and/or shared use of space by various modes and user groups, often in the context of “complete streets” that focus on the mobility of, service to, and safety of various user groups and modes.

The modern traffic engineer cannot have a mindset limited to traffic counts, speeds, and maximizing vehicular capacity, although those elements continue to be important inputs. The focus has evolved to a balanced use of public space, respect for (and inclusion of) various modes and users, overall mobility, and safety. Meeting mobility needs has moved beyond “more vehicles, more roadway” to supporting the needs of a livable environment.

For design and for evaluation of operations, the first questions have to be "What is the stated or perceived need? How can it be met? How can it be improved? How can we measure whether the purpose has been achieved, and/or how the system is really operating?" The authors have too often seen the critique that "measurements show that such-and-such has not been achieved" when the design intent had never been to achieve that specific objective—except in the after-the-fact mind of the commentator.

Collection and reduction of traffic data cover a wide range of techniques and technologies from simple manual techniques (often aided by a variety of handheld or other devices for recording the data) to complex use of the ever-expanding technology of sensors, detectors, transmission, smartphone and tablet data apps, and third-party databases of travel times, origin and destinations, and other information.

This chapter provides a basic overview of data collection and reduction in traffic engineering. The technology applied, as noted, changes rapidly, and traffic engineers need to maintain current knowledge of this field.

Some of the basic references for the latest information on traffic data techniques and technologies include the following:

- The Federal Highway Administration's *Traffic Monitoring Guide* [1], which is updated approximately every two years. At this writing, the 2016 version is the latest edition.
- The Institute of Transportation Engineers' (ITE) *Traffic Engineering Handbook* [2], in its 7th edition at this writing;
- The ITE *Manual of Transportation Engineering Studies* [3], 2nd edition (2010), in conjunction with other references;
- The AASHTO *Highway Safety Manual* [4] with 2014 Supplement.

In addition, there are manufacturer pieces that can be found in *ITE Journal*, and through web searching. The web searches can be useful in tracing back to official agency practices and expectations for data, and well as properly sourced material. As with all web searches, some care must be exercised to not take all posted material as of equal value.

9.1 Sources of Data

There was a time when past studies were relatively inaccessible, lacked overall consistency, and were dated. To a large extent, this is no longer true: a number of states and other jurisdictions have user-friendly online databases of average daily traffic (ADT) and other information on major roads, of local studies done for maintenance and protection of traffic (MPT) during construction, data collected for traffic impact studies, and so forth.

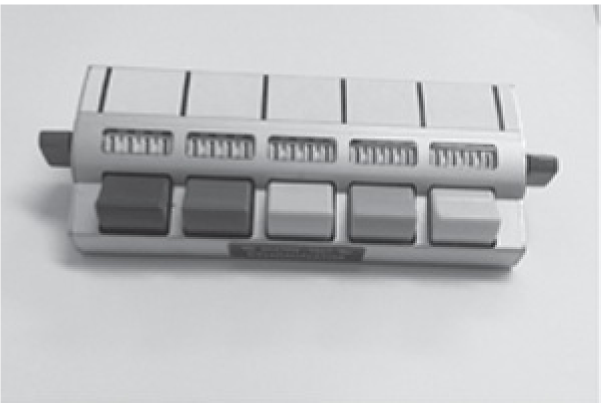
Likewise, such web sites may contain the specific requirements of a given agency on the amount of data to be collected for a specific study, including variables, classifications, durations (hours per day, number of days), and presentation formats.

The traffic engineer is expected to be aware of such existing data, draw upon it as needed, and place any new data in the context of what had been known to date.

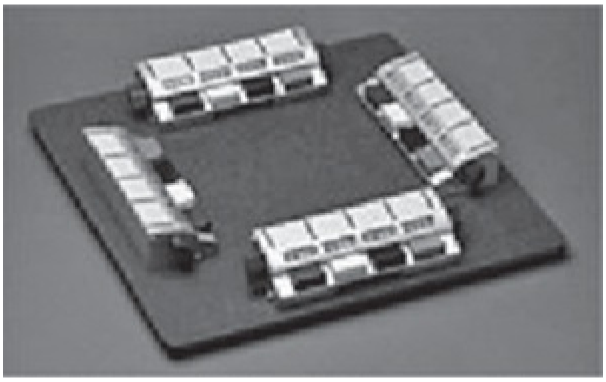
9.1.1 Traditional Approaches

The common image conjured upon by the words "traffic study" is a crew by the side of the road, literally counting vehicles, perhaps classified by turning movement and/or classified by vehicle type. The crews have been trained, they have proper vests and other safety equipment, and are using manual counters, such as shown in Figure 9.1.

Figure 9.1: Hand-Counters Illustrated



(a) Five-Button Hand Counter



(b) Counters Mounted on an Intersection Board

(Source: (a) Denominator Company Inc., (b) “Traffic Counts,” *Traffic Handbook*, Center for Transportation Research, Iowa State University, Fig 3-1, pg 3-2.)

The other traditional image is the roadside automatic traffic recorder (ATR), with tubes stretched across the traffic lanes. Refer to Figure 9.2.

Figure 9.2: Road Tubes with Portable Counter Illustrated

(Source: International Road Dynamics Inc, Saskatoon, SK, CANADA, www.irdinc.com.)

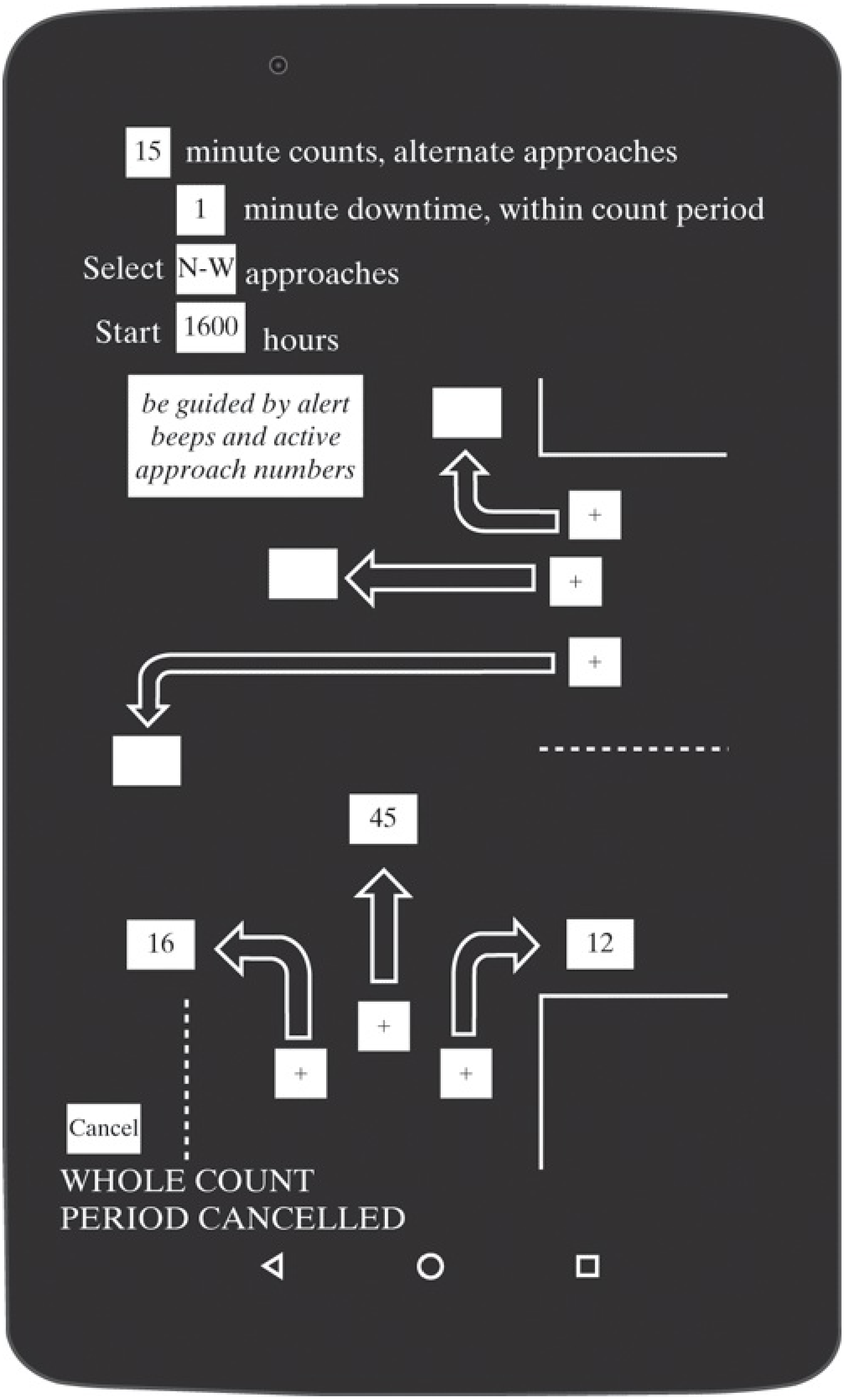
Both images are valid, and remain in common use. They are well-accepted, cost-effective in many applications, and consistent with past studies. Some evolution has taken place, in that an ATR no longer has a paper roll of printed numbers but rather a digital database, GPS location of the recording device, and the ability to download it remotely. But the underlying principles remain. For this introductory text, the present chapter pays some detailed attention to these “first principles” of data collection.

But the technology is changing rapidly, and the emphasis is on reliability, cost-effectiveness, and enhanced data quality by avoiding the potential for transcription errors. Crew training is always an issue, and the traffic engineer will find that the extent of the data collection effort, the training effort, and the technology, and even the number of reliable crew that can be assembled, will mean that “not one size fits all” in terms of which technique(s) to use to collect the needed data.

9.1.2 Changes in the Technology

It is now common that traffic count data, by movement or classification (that is, vehicle type), is done on a tablet computer, with data either wirelessly sent to a central database or downloaded after the collection period. The data formats are generally consistent with the required reporting and data storage formats. Transfer to Excel or other spreadsheets via “csv” export files are common. The data can be accompanied by GPS location of the collection point and time stamps from an internal clock. Refer to Figure 9.3 for an illustration.

Figure 9.3: Tablet App for Intersection Traffic Counts



(Source: Portable Studies by Afermas LLC, Sun Prairie, WI, www.portablestudies.com.)

Figure 9.3: Full Alternative Text

With regard to the sort of traffic counting traditionally done by ATRs, there has been a trend to two alternative technological approaches:

1. Side-fire or other radar detectors, allowing measurement to be taken without entering the moving lanes. Much of the discussion centers on ease of installation, accuracy of one approach over the other, calibrating the use of the newer technology, and the means to "bridge the gap" between two data acquisition technologies when long-term trends are watched over years and even decades. Availability of staff for installation and maintenance, comfort with traditional approaches, and ultimately cost-effectiveness are important considerations.

2. Point detectors affixed to the lanes, or imbedded into them, with wireless linkage to a local data collection point. The point detectors have long–lasting batteries in some renditions, or can be disposable in others. The local data collection point can be interrogated remotely, or uploaded on schedule. The same issues of comfort, cost-effectiveness, and accuracy are commonly considered.

Some jurisdictions have gone with one approach over the other; a number have stayed with the traditional ATRs, in part because there are higher priority issues drawing management's attention, the evaluation of a change can be an intensive effort, and the data arrives in digital form from any of the sources.

It is interesting that any move to one of the non-tube alternatives engenders another question, and sometimes confounds the decision-making (because it is addressed concurrently): Why not make the count location a permanent feature? For instance, one jurisdiction does an annual one-week screen-line count, traditionally using ATRs. When a different technology was considered, the immediate questions were as follows: Why not simply leave them there, and get continuous counts year-round, including seasonal trends? Given the long history of the annual ATR counts, will the newer technology provide accurate data, using the ATR as the baseline? Why consider the ATR as the baseline, given that well-installed side fire devices, ATRs, and manual counts all have some margin of error in measuring ground-truth traffic present? Even two or three crew members concurrently counting manually will obtain different results, for instance.

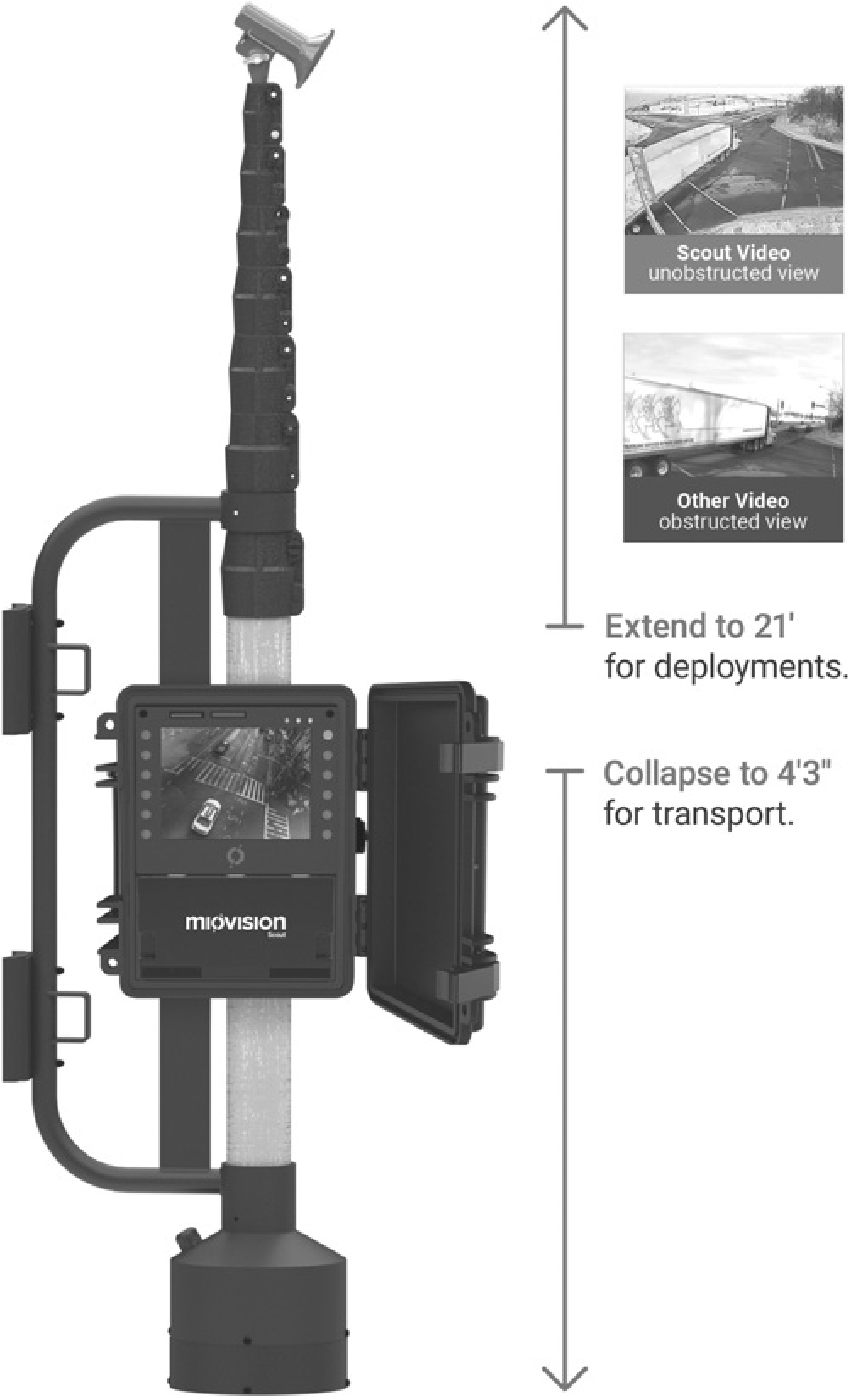
These questions must all be considered. The answers will vary by location, time, study purpose, and the specific technologies being considered for use.

9.1.3 Video-Based Measurements

Video records of traffic are very appealing, if the vantage point is high enough: the record is permanent, and can be revisited if necessary. The problems have to do with finding the vantage point, and essentially manual reduction from the video.

At the time this chapter was written, image processing of the video record has begun to take hold, and use of video cameras for full-intersection counts is becoming more common. Figure 9.4 shows one of the current technologies. The cabinet can be closed, the entire device can be affixed to a telephone pole or other secure fixture, and the telescoping extension extended to place the camera. The visible area is displayed to the installer. The manufacturer offers a data reduction service, as well as the devices.

Figure 9.4: Miovision Camera Setup for Field Work



(Source: Miovision Technologies, Kitchener, ON, CANADA, www.miovision.com.)

Figure 9.4: Full Alternative Text

9.1.4 Smartphones and Other Devices

The smartphone is an obvious platform for “tablet-based apps” that count traffic and do a host of specialty measurements that are easily written, checked, and wrapped in secure means of transmitting the data to a central location.

At the same time, the smartphone is itself a data point and a data source. While the owner can restrict access at some level, it is also true that many owners participate in “crowdsourcing” apps that provide them with navigation guidance, traffic conditions, and other information—and add them to the data driving those apps.

9.1.5 Existing Data

At this writing, it is commonplace that buses have GPS tracking, as do many truck fleets. In some cities, taxi fleets are equipped with GPS and already provide information on trip patterns, trip travel times, and potentially routing decisions. Such databases can provide “breadcrumb” trails on how taxis used the network, and can be related to traffic condition data from other sources, information never before routinely available.

On a network basis, scrubbed use of electronic toll collection (ETC) tag reader information can provide powerful renditions of network travel times, for (a) traffic studies, (b) public information, and (c) adaptive control decisions. Figure 9.5 shows such a rendering in Manhattan, NY, used for all three purposes.

Figure 9.5: Real-Time Travel Time Display in Manhattan,

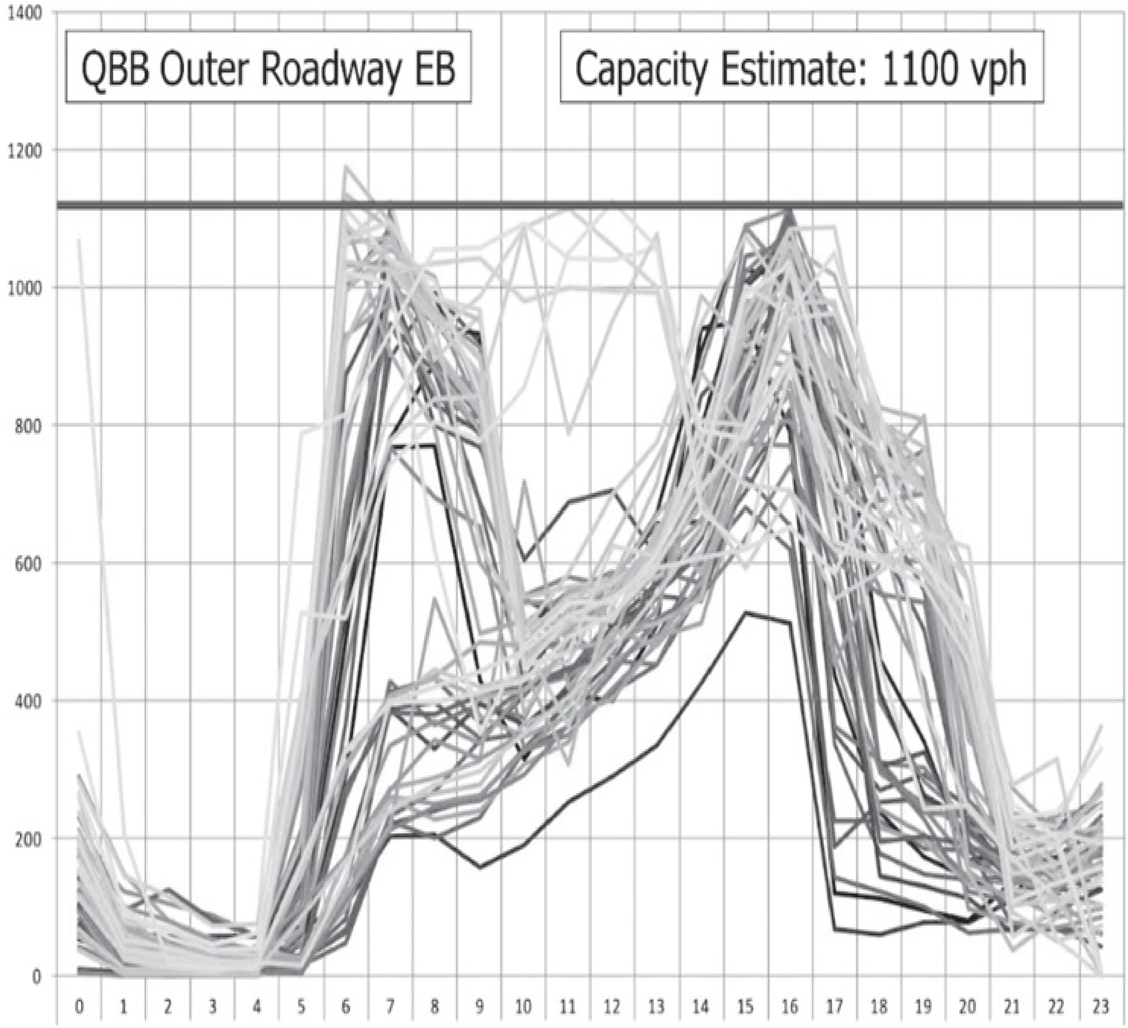
NY

Figure 9.5: Full Alternative Text

For many of those whose careers began before this edition of this text, the routine availability of GPS data and of network travel times is groundbreaking.

But even the relatively easy availability of public data files, for public information or for professional traffic studies, is another major advance. The data may have existed in paper reports archived and available for transcription, but the ready availability of electronic records—combined with powerful spreadsheet and graphics tools—provides for crisp insights. Figure 9.6 shows one such case: the issue at hand was determining the capacity of a given river crossing for the purposes of estimating arrival demand (or conversely, queuing upon departure). The definition of capacity includes a key phrase (italics added):

Figure 9.6: Capacity Estimated from Existing Data—Ed Koch Queensborough Bridge, NYC



(Source: KLD Engineering, P.C.)

Figure 9.6: Full Alternative Text

the maximum hourly rate at which persons or vehicles can be *reasonably expected* to traverse a point or a uniform segment of a lane or roadway during a given time period under prevailing roadway, traffic and control conditions

In Figure 9.6, some 57 days of data is displayed from permanent count stations. On the basis of that, the consensus was that the capacity could be estimated at 1,130 veh/h as that which “can be reasonably expected.” And on that finding, supported by existing data, the study conclusions and recommendations became clear.

9.1.6 Perspective

The authors are not going to be bold enough to believe or assert that the above discussion is comprehensive. New apps are being written every day, new generations of smartphones provide even more power, and the amount of data is no longer a “flow” but rather a “flood.”

The challenge is to decide what information is most relevant for which purpose. For instance,

- three smartphones in a single car may mean that there have to be corrections for *vehicle* counts. But at the same time, travel times over segments of a network can be attained more easily than ever before (again, with some allowances for double counting due to vehicle occupancy);
- at the same time, the relevant metric may not be a precise travel time (or even volume) but rather “bands” or thresholds that when left/crossed cause traffic signal plans to be changed;
- such travel time information can also allow mode estimates—buses travel differently than passenger cars, and have stop patterns that provide an insight into which mode the smartphone owner is using (and hence the travel performance of that mode).

The authors have spent careers that focused on various traffic data, with true origin-destination data simply an aspiration, or a deduction from other data, or from limited data. Smartphones and the endemic use of electronic control devices have opened the door to

- origin-destination information, at least from/to the network peripheries to locations within the network;

- routing decision information that can be linked to travel times and perhaps other factors;
- the ability to calibrate traffic study tools such as simulation models, based upon this knowledge.

The flip side of this is the rising concern over privacy. The discussion of privacy, of course, takes place in a much broader context. Many are uncomfortable with the amount of personal data collected by both government agencies and major social network, shopping, and general information purveyors.

Yes, there are serious privacy issues. The need to have both opt-out choices and data scrubbing even with opt-in must be addressed. But a new era of traffic data and traffic studies has arrived, and some specifics will simply outpace a text that is updated every few years.

9.2 The Connected Vehicle

In only the past few years, the “connected vehicle” has gone from a concept to a priority, driven by its potential and by market forces—auto manufacturers and new vendors are focused on a race to market, and feeding driver expectations.

The vehicle *communicating* with the environment, and leaving a breadcrumb trail of where it has been, what travel times it experienced, and what choices its routing algorithms have made, is a windfall of information. The vehicle *sensing* the presence of other vehicles, and of pedestrians and bicycles, creates new data on potential conflicts, avoidance maneuvers, warnings, emergency braking, and actual incidents.

For those entering the traffic profession, the challenge will be to make the flood of information manageable, define the next generation of studies, and make best use of the information to create a safer and more efficient traffic environment.

9.3 Applications of Traffic Data

Traffic engineers collect and reduce data for many reasons and applications:

- Managing the physical system. The physical traffic system includes a number of elements that must be monitored, including the roadway itself, traffic control devices, detectors and sensors, and light fixtures. Physical inventories must be maintained simply to know what is “out there.” Bulbs in traffic signals and lighting fixtures must not be burned out; traffic markings must be clearly legible; signs must be clean and visible; and so forth. Replacement and maintenance programs must be in place to ensure that all elements are in place, properly deployed, and safe.

- Establishing time trends. Traffic engineers need trend data to help identify future transportation needs. Traffic volume trends can identify areas and specific locations that can be expected to congest in the future. Accident data and statistics over time can identify core safety problems and site-specific situations that must be addressed and mitigated. Trend data allows the traffic engineer to anticipate problems and solve them *before* they actually occur.

- Understanding travel behavior. A good traffic engineer must understand how and why people (and goods) travel in order to provide an effective transportation system. Studies of how travelers make modal choices, trip-time decisions, and destination choices are critical to understanding the nature of traffic demand. Studies of parking and goods delivery characteristics provide information that allows efficient facilities to be provided for these activities.

- Calibrating basic relationships or parameters. Fundamental characteristics such as perception-reaction time, discharge headways at signalized intersections, headway and spacing relationships on freeways and other uninterrupted flow facilities, gap acceptance characteristics, and others must be properly quantified and calibrated to existing conditions. Such measures are incorporated into a variety of predictive and assessment models on which a great deal of traffic engineering is based.

- Assessing the effectiveness of improvements. When traffic improvements of any kind are implemented, follow-up studies are needed to confirm their effectiveness, and to allow for adjustments if all objectives are not met.

- Assessing potential impacts. An essential part of traffic engineering is the ability to predict and analyze the traffic impacts of new developments and to provide traffic input into air pollution models.

- Evaluating facility or system performance. All traffic facilities and systems must be periodically studied to determine whether they are delivering the intended quantity and quality of access and/or mobility service to the public.

Data and information from traffic studies provide the underpinning for all traffic planning, design, and analysis. If the data is not correct and valid, then any traffic engineering based on the data must be flawed. Some of the tasks involved in data collection and reduction are mundane. Data and information, however, are the foundation of traffic engineering. Without a good foundation, the structure will surely fall.

9.4 Types of Studies

It is simply infeasible to list all of the types of traffic studies that take place in modern traffic engineering. Eleven of the most common types of studies include the following:

1. Volume studies. Traffic counts are the most basic of studies, and are the primary measure of demand; virtually all aspects of traffic engineering require demand volume as an input, including planning, design, traffic control, traffic operations, detailed signal timing, and others.

2. Speed studies. Speed characteristics are strongly related to safety concerns and are needed to assess the viability of existing speed regulations and/or to set new ones on a rational basis.

3. Travel time studies. Travel times along sections of roadways constitute a major measure of quality of service to motorists and passengers, and also of relative congestion along the section. Many demand-forecasting and assessment models also require good and accurate travel times as a critical input.

4. Delay studies. Delay is a term that has many meanings, as will be discussed in later chapters. In essence, it is the part or parts of travel time that users find particularly annoying, such as stopping at a traffic signal or because of a midblock obstruction.

5. Density studies. Density is rarely directly observed. Some modern detectors can measure “occupancy,” which is directly related to density. Density is a major parameter describing quality of operations on uninterrupted flow facilities.

6. Accident and safety studies. As traffic safety is the primary responsibility of the traffic engineer, the focused study of accident characteristics, in terms of system-wide rates, relationships to causal factors, and at specific locations, is a critically important function.

7. Parking studies. These involve inventories of parking supply and a

variety of counting methodologies to determine accumulations and total parking demand. Interview studies also involve attitudinal factors to determine how and when parking facilities are used.

8. Goods movement and transit studies. Inventories of existing truck-loading facilities and transit systems are important descriptors of the transportation system. As these elements can be significant causes of congestion, proper planning and operational policies are a significant need.

9. Pedestrian studies. Pedestrians are an important part of the demand on transportation systems. Their characteristics in using crosswalks and signalized and unsignalized intersections constitute a required input for many types of analysis. Interview techniques can be used to assess behavioral patterns and to obtain more detailed information.

10. Calibration studies. Traffic engineering uses a variety of basic and not-so-basic models and relationships to describe and analyze traffic. Studies are needed to calibrate key values in models to ensure that they are reasonably representative of the conditions they claim to replicate.

11. Observation studies. Studies on the effectiveness of various traffic control devices are needed to assess how well controls have been designed and implemented. Observance rates are critical inputs to the evaluation of control measures.

This text includes a detailed treatment of volume studies, speed studies, travel time studies, delay studies, accident and safety studies, and parking studies. Others are mentioned but not treated in detail. The engineer is encouraged to consult other sources for detailed descriptions of study procedures and methodologies (for example [1-4]).

9.5 Manual Data Collection Methodologies

Notwithstanding all the modern technology available to the traffic engineer, some studies are best conducted manually. These studies tend to be for short duration and/or at highly focused locations. The use of automated equipment requires set-up and take-down effort, which may not be practical for short studies or spot observations. Certain types of information are difficult to obtain without manual observation. Vehicle occupancy, frequently of interest, is generally observed directly.

As has been noted, short time frames, short lead times, and the need for certain types of data lead to manual data collection approaches. The types of studies most often conducted manually are (a) traffic counts at a specific location or small number of such locations, usually when the time frame is less than 12 to 24 hours, (b) speed or travel time data at a focused location for short duration, (c) observance studies at specific locations for short durations, and (d) intersection delay studies of short duration.

9.5.1 Traffic Counting Applications

The most likely situation leading to a manual study is the traffic count. A traffic engineer is often faced with a problem requiring some detailed knowledge of existing traffic volumes. A signal timing problem requires peak-hour flow rates for the intersection, perhaps in the A.M. peak, the P.M. peak, and mid-day. Sending a few people out to conduct such a study can be easier than the logistics of other methods and lags in data processing.

The primary difficulty with hand-counters is that the data must be manually recorded in the field at periodic intervals. This disrupts the count. Obviously, while the observer writes down the register counts and resets them for the next period, uncounted vehicles pass by. In order to obtain continuous count information on a common basis, short breaks are introduced into the counting procedure. Such breaks must be systematic and uniform for all observers. The system revolves around the *counting period* for the study—the unit of time for which volumes are to be observed and recorded. Common counting periods are 5 minutes, 15 minutes, and 60 minutes, although other times can be, and occasionally must be, used.

The short breaks are generally arranged in one of two ways:

- A portion of each counting period is set aside for a short break.
- Every other counting period is used as a break.

In the first option, "x" out of every "y" minutes is counted. Thus, for 5-minute counting periods, an observer might count 4 out of every 5 minutes; for a 15-minute counting period, 12 or 13 minutes out of 15 minutes might be counted. To provide for a continuous count profile, the volume that is expected to occur during the short breaks must be estimated and added. This is generally done by assuming that the *rate of flow* during the missing minutes is the same as that during the actual count. Then:

$V_y = V_x (y/x)$ [9-1]

Where

V_y = volume for continuous counting period of "y" minutes (vehs), V_x = volume for discontinuous counts of "x" minutes (vehs), y = counting period (mins), and x = counting period minus short break (mins).

Consider a case where a manual counting survey is conducted by counting vehicles in 4 minutes out of every 5-minute period. If a count of 100 vehicles is obtained, what is the estimated count for the full 5-minute counting period? Using Equation 9-1:

$V_5 = 100(5/4) = 125$ vehs.

Such counts are reported as *full vehicles*, not fractions.

When alternate periods are used as breaks, full period counts are available in alternate counting periods. Missing counts are estimated using straight-line interpolation:

Vi=Vi−1+Vi+12 [9-2]

where

V i = volume in missing counting period “i” (vehs), V i−1 = volume in counting period “i−1” (vehs), and V i+1 = volume in counting period “i+1” (vehs).

Again, any estimated counts are rounded to the nearest whole vehicle.

In practice, it is often necessary to combine the two approaches. Consider the example shown in Table 9.1. In this case, a single observer is used to count two lanes of traffic on an urban arterial. Using hand counters, the observer alternates between lanes in each counting period. Thus, for each lane, alternating counts are obtained. Because the observer must also take short breaks to record data, only 4 minutes out of every 5 minutes are actually counted. Table 9.1 illustrates the three computations that would be involved in this study.

Table 9.1: Data from an Illustrative Volume Study

Period	Time (PM)	Actual Counts (4 minutes) vehs		Expanded Counts (x 5/4) vehs		Estimated Counts[1] vehs		Estimated Flow Rates veh/h	
		Lane 1	Lane 2	Lane 1	Lane 2	Lane 1	Lane 2	Lane 1	Lane 2
1	5:00	24		30.0		30	*43*	360	516
2	5:05		36		45.0	*33*	35	396	540
3	5:10	28		35.0		35	*37*	420	564
4	5:15		39		48.8	*36*	49	432	588
5	5:20	30		37.5		38	*54*	456	648
6	5:25		47		58.8	*41*	59	492	708
7	5:30	36		45.0		45	*61*	540	732
8	5:35		50		62.5	*44*	63	528	756
9	5:40	34		42.5		43	*61*	516	732
10	5:45		48		60.0	*46*	60	552	720
11	5:50	40		50.0		50	*59*	600	708
12	5:55		46		57.5	*55*	58	660	696
Total		**192**	**266**	**240.0**	**332.6**	**496**	**659**		
% in Lane		**41.9%**	**58.1%**	**41.9%**	**58.1%**	**42.9%**	**57.1%**		

1. *Italics* indicate an interpolated or extrapolated value.

Table 9.1: Full Alternative Text

Actual counts are shown, and represent 4-minute observations. These are expanded by a factor of 5/4 (Equation 9-1) to estimate volume in continuous 5-minute counting periods. At this point, each lane has counts for alternating counting periods. Missing counts are then interpolated (Equation 9-2) to estimate the count in each missing period.

The first period for lane 2 and the last period for lane 1 cannot be interpolated, as they constitute the first and last periods. The counts in these periods must be extrapolated, which is at best an approximate process.

All results are rounded to whole vehicles, but only in the final step. Thus, the "expanded counts" carry a decimal; the rounding is done with the "estimated counts," which include the expansion, interpolation, and extrapolation needed to complete the table. The "estimated counts" constitute 5-minute periods. The flow rate in each 5-minute period is computed by multiplying each count by 12 (twelve 5-minute periods in an hour).

9.5.2 Speed Study Applications

In Chapter 11 the analysis of speed data is fully discussed. Because of both the confidence bounds often needed in the results and the cost of collecting distinct data points in the desired time window, sample sizes for many speed studies are generally less than 100 vehicles.

Unless there is permanent detection equipment at the desired study location, many speed studies are conducted manually by one of two methods:

- Measurement of elapsed time over a short measured highway segment using a simple stopwatch.
- Direct measurement of speeds using either handheld or fixed-mounted radar meters.

When using a stopwatch to time vehicles as they traverse a short section of highway (often referred to as the “trap”), there are two potential sources of systematic error:

- parallax (viewing angle): the line of sight (if not 90 degrees) creates the appearance of a boundary crossing somewhat before it already occurs.
- manual operation of the stopwatch (or the timer on the smartphone).

Parallax error is systematic, and adjustments can remove its impact, as long as the viewing angle is known.

Figure 9.7 illustrates the parallax error. Normally, an observer would choose a location at which one boundary can be viewed exactly. Because of the angle of vision of the other boundary, the observer actually “sees” the vehicle cross the boundary before it actually does. The observer thinks that travel times over distance *d* are being measured, while the travel times actually represent distance d_{eff}.

Figure 9.7: Parallax Error

Illustrated

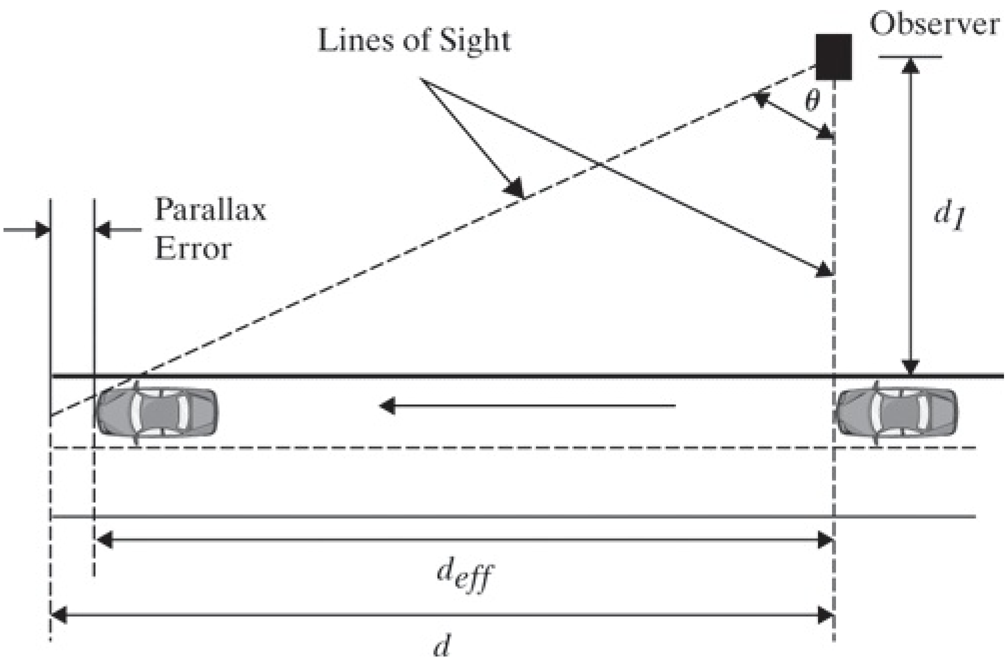


Figure 9.7: Full Alternative Text

Trigonometry may be used to adjust between *d*, the distance the observer thinks has been traversed, and d_{eff}, the distance that had actually been traversed:

Tan(θ) = d eff d 1 d eff = d 1 Tan(θ) S i = d 1 Tan(θ) t i [9-3]

where:

d 1 =
distance from observer to closest edge of observed vehicle, ft (taken at a 90°
d eff =
effective distance the subject vehicle has traveled when observed crossing th
d =
distance the observer "sees" the subject vehicle has traveled when observed
S i = speed of subject vehicle i, ft/s, t i =
travel time of subject vehicle to cross the study segment, s, and V =
angle of observation (degrees or radians).

Of course, it is always preferable to minimize the parallax error by positioning the observer at a location that would make it as small as possible – which generally results in moving further away from the roadside. In any event, the angle of observation (V) and the distance d_1 must be carefully measured in the field.

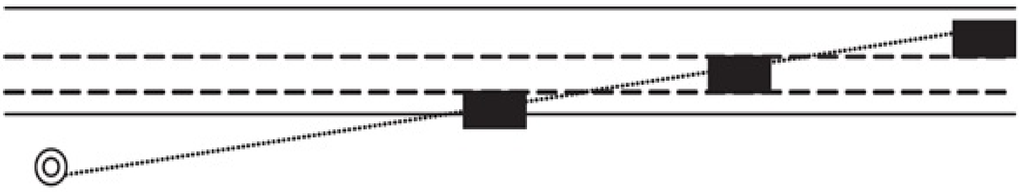
The random error due to manually starting and stopping the stopwatch or timer is the other common source of potential systematic error. To the extent that both actions systematically lag the actual event and are comparable, they do cancel out.

However, the inherent variability of the human must still be considered: If the randomness is a significant part of the travel time observed, it will affect the confidence bounds on the travel time estimates. Further, because the travel time is in the denominator of the individual speed computation, it can systematically skew the speed estimate.

Use of radar meters for speed measurement is the most common method to obtain speed data. Unfortunately, this method of measuring speeds is also associated with enforcement of speed laws. Thus, once the observation station is identified by drivers, everyone slows down, and the results do not reflect ambient driver behavior. Therefore, concealment of the device and the personnel involved in the study is important. Because such meters require a clear line of sight of oncoming vehicles, such concealment is often difficult to accomplish for any significant length of time. Figure 9.8 illustrates a common type of radar meter used to collect speed data.

Figure 9.8: Measuring Speeds with Radar Meters

(a) Handheld Radar Unit



(b) Resolving Multilane Traffic Streams with Radar

(Source: Used with permission from Alamy.)

9.5-2 Full Alternative Text

Radar meters use the Doppler Principle to measure speeds. A radar wave is reflected from a moving object (vehicle), and returns at a different frequency. The difference in emitted and reflected frequency is proportional to speed. Exact measurements require that the wave be in line with the approaching vehicle. Any angle creates a systematic error, much in the same way as parallax affects manual observations. As most radar measurements must be made from a roadside or overhead location, there is virtually always an angle involved. Most meters, however, have an adjustment that can be entered based upon the angle of deflection of the wave; results are automatically implemented and the correct speed is read directly.

As illustrated in Figure 9.8(b), using radar meters on multilane highways creates some interesting problems. Given a shallow angle (desirable for accuracy), the emitted wave may reflect off a vehicle in lane 1, 2, or 3. As shown in Figure 9.8(b), the longitudinal location of the vehicle when the wave is reflected can vary greatly depending upon what lane the vehicle is in. This sometimes makes it difficult to identify *which vehicle* has been observed.

As radar meters emit waves within the federally regulated range of frequencies, each radar meter must have a license from the Federal Communications Commission (FCC). Most radar meters are accurate to ±2 mi/h, and should not be used where great precision and accuracy are needed.

Manual speed studies are particularly interesting, as it is necessary to obtain a strictly random sample. Even the most experienced observer cannot measure the speed of *all* vehicles passing a point. Therefore, some common problems must be avoided: (a) avoid trying to measure the "fastest" and/or "slowest" vehicles, (b) where platooning of vehicles exist, remember that only the first driver in the platoon is actually choosing his/her speed, (c) truck or other heavy vehicle speeds should not be specifically targeted unless the study focuses on truck speeds. While it may seem contradictory, assuring a random sample generally requires a very rigid sampling process, such as "every nth vehicle in each lane."

9.5.3 Other Manual Study Applications

While vehicle counts and simple speed studies are the most often conducted using manual observations and recording of data, there are others, some of which are discussed in greater detail in subsequent chapters.

- Intersection delay studies: these are most often done by hand, given that they tend to focus on short-term peaks in demand.

- Travel time studies: these most often involve test vehicles moving within the traffic stream, with data observers/recorders marking appropriate times that various landmarks or other known locations are passed.

- Observance studies: these usually involve short-term observations of how well drivers are obeying a given regulation or control device; very common at STOP-controlled intersections.
- Parking and parking inventory studies: these often involve manual observation of parking supply and accumulation.

9.5.4 Staffing and Workforce Considerations

Manual studies must be adequately staffed. From a simple intersection count, which might involve 4–6 people, to complex network studies involving 80 or more people, staffing and training are always important components of manual studies.

For larger studies, it is very difficult to assemble a working group consisting of only experienced traffic personnel. The most common alternatives are to hire students from nearby colleges and/or use of temp agencies. In either case, it is critical that all personnel be carefully trained to understand the importance of their tasks, and exactly what is expected of them. Pre-designed field sheets should be designed to avoid mistakes, and should be carefully presented to all personnel. Such issues as communications among field personnel, use of any equipment, and how and where to provide results at the end of data collection must also be carefully planned.

For multiple counts, a real-time communication system is needed that connects all observers to a supervisor. Counts in such situations must all start and stop at exactly the same time; a single supervisor times the study and issues “start” and “stop” orders as appropriate.

For studies longer than 1 or 2 hours, sufficient personnel must be trained to allow for periodic relief of observers. There will always be “no shows” for any large study, so “extra” personnel are also needed to cover these eventualities. Practical limitations must also be recognized: a typical worker not experienced in traffic studies, for example, could be counted on to count and classify one heavy or two light movements, or to measure speeds in one lane.

9.6 Semi-Automated Studies Using Pneumatic Road Tubes and Similar Devices

A good number of traffic studies are conducted using a variety of portable traffic data collection/recording devices. The most common portable device used in traffic studies is the pneumatic road tube. A pneumatic road tube is a closed-end tube in which an air pressure is maintained. When stretched across a roadway, a vehicle (actually an axle) rolling over it creates an air pulse which travels through the tube, which is connected to some form of data capture device. Such tubes are most often used for traffic counting, but can also be used to measure speed.

While there are a variety of traffic counters available to use with road tubes, the most common types will record a total count at pre-set intervals, so that 5-, 10-, 15-, 60-minute counts can be automatically recorded.

Road tubes do not count vehicles—they count axles. A passenger car with two axles crosses the road tube registering a pulse *twice*. A tractor-trailer combination with multiple axles may cause as many as 5 or 6 actuations that will be recorded. To obtain an estimate of vehicle-counts, manual observations must be made to determine the *average number of axles per vehicle*. Obviously, if this had to be observed for the entire period of the study, there is no point in using road tubes. Representative samples do, however, have to be collected if reference data is not available elsewhere.

Table 9.2 illustrates a sample classification count. Vehicles are observed and classified by the number of axles on each. As shown, this data can be used to estimate the average number of axles on a vehicle at the study location.

Table 9.2: Sample Classification Count for a

Road-Tube Volume Study

Number of Axles	Number of Vehicles Observed	Total Axles for Observed Vehicles
2	400	$\times 2 = 800$
3	75	$\times 3 = 225$
4	25	$\times 4 = 100$
5	10	$\times 5 = 50$
6	5	$\times 6 = 30$
Total	**515**	**1,205**

Table 9.2: Full Alternative Text

The average number of axles per vehicle at this study location is $1205/515=2.34$ axles/vehicle . If the recording device shows a count of 7,000 axles for a given day, the estimated vehicle count would be $7{,}000/2.34=2{,}991$ vehicles. The complicating factor of missed vehicles due to concurrent actuations is not addressed in this example.

There are also a number of practical issues with road tubes:

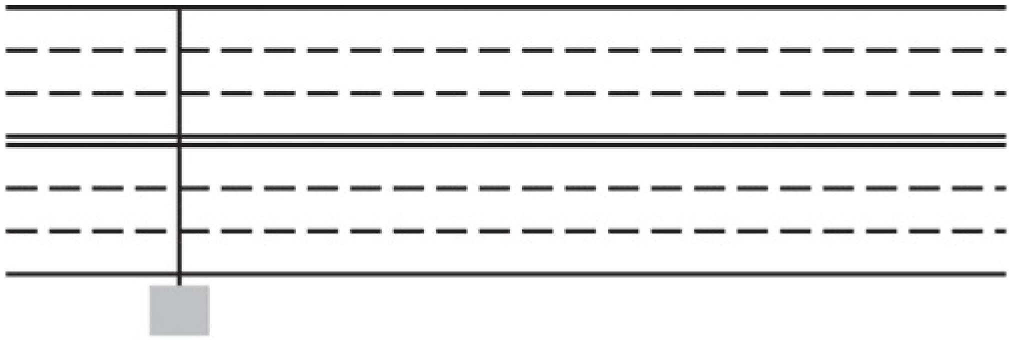
- If not tautly fastened to the pavement (usually accomplished with clamps and epoxy), the road tube can start a "whipping" action when vehicles continually traverse it, causing an eventual breakage. Once broken, the tube cannot relay data. Road tubes are also subject to vandalism. In either event, the tube should be inspected at regular intervals to ensure that it is still functioning.

- If the road tube is stretched across more than one lane, simultaneous actuations are possible, and the recorded counts may be low because of this. This problem is more severe as the number of lanes and the volume of traffic increases, but correction factors can be used or built in.

Figure 9.9 illustrates common set-ups for road tubes as used in traffic counting studies. Figure 9.9(a) shows a single tube across all lanes of a facility. In this configuration, a total two-way count of axles is obtained.

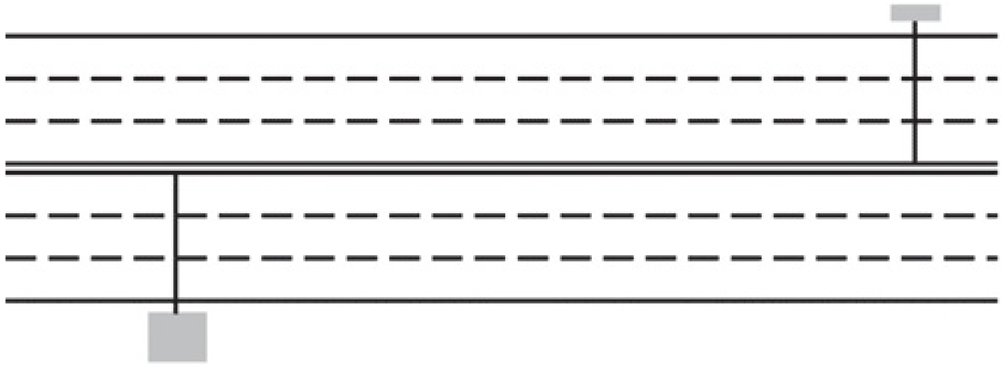
Figure 9.9(b) shows the most common technique—two road tubes set up to record axles in each direction separately. Figure 9.9(c) shows a less typical set-up in which lane counts can be deduced. The tubes must be close enough together that the number of lane-changes within the detection range is minimal.

Figure 9.9: Alternative Field Set-Ups for Road Tubes



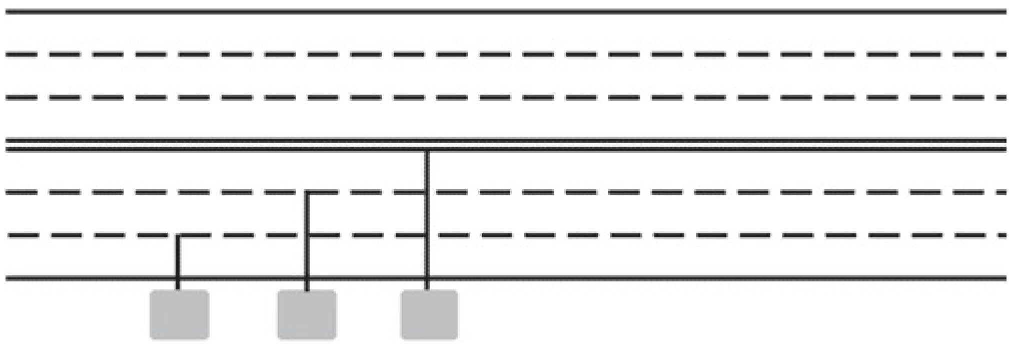
(a) Road Tube for Counting Total Two-Way Axle Counts

9.6-3 Full Alternative Text



(b) Road Tubes for Counting Directional Axle Counts

9.6-3 Full Alternative Text



(c) Road Tubes for Lane Axle Counts

9.6-3 Full Alternative Text

9.7 Permanent Detectors and Their Use

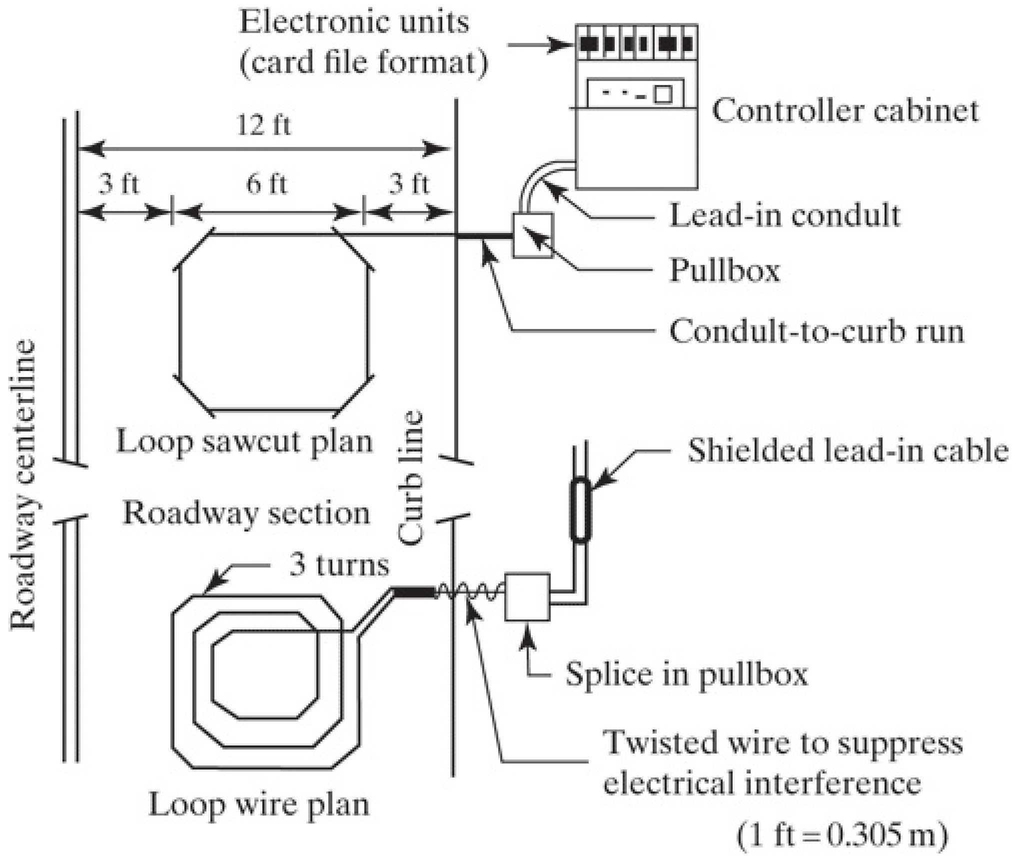
Rapid advances in traffic detector technology are rapidly changing the landscape for traffic studies. Detectors are used for all types of things, from data collection and transmission to the real-time operation of traffic signal systems. As intelligent transportation technology marches forward, there is great interest in real-time monitoring of traffic systems on a massive scale. This requires that the traffic system be instrumented with large numbers of permanent detectors and the capability to observe the data they provide in real time.

The *Traffic Detector Handbook* [5] provides an excellent overview of current detector and sensor technology. It classifies detectors and sensors into a number of broad categories based upon the type of technology used:

- Sound (acoustic)
- Opacity (optical, infrared, video image processing)
- Geomagnetism (magnetic sensors, magnetometers)
- Reflection of transmitted energy (infrared laser radar, ultrasonic, microwave radar)
- Electromagnetic induction (inductive loop detectors)
- Vibration (triboelectric, seismic, inertia switch sensors)

The most prevalent detector in use today is the inductive loop. When a metal object (vehicle) enters the field of the loop detector, inductance properties of the loop are reduced and sensed, thus recording the presence of the vehicle. The loop essentially creates an electromagnetic field that is disturbed when a metal object enters it. Induction loops require that a cut be made in the pavement surface, with one or more wire loops placed in the cut. A pull-wire connects the detector to a power source, which is then connected to a controller unit. Figure 9.10 shows a typical installation of an inductive loop detector.

Figure 9.10: Typical Installation of an Induction Loop Detector



(Source: *Traffic Detector Handbook*, 3rd Edition, Federal Highway Administration, Publication No. FHWA-HRT- 06-108, Washington, DC, 2006, Figure 1-4, pg 1-12.)

Figure 9.10: Full Alternative Text

The typical induction loop detector measures 6 ft × 6 ft and covers one lane. Multiple detectors are needed to monitor multiple lanes. Longer loops are available. These are often used in conjunction with actuated signals requiring a significant detection area. In some cases, a long detection area is provided by installing a series of 6 ft × 6 ft detectors in each lane.

Inductive loop detectors directly measure the presence and passage of vehicles. Other important measures, such as speed and density, can be deduced using calibrated algorithms, but the accuracy of these is often insufficient for research use.

There are a variety of permanent detectors that utilize the same Doppler Principle as handheld radar meters. They all rely on reflected energy from vehicles that can be detected and used to obtain a variety of traffic parameters. These include microwave radar meters, infrared sensors, laser radar meters, and ultrasonic detectors. The difference lies in the wavelength and frequency of energy that is emitted and sensed.

Perhaps the most exciting recent development in real-time traffic detection is the rapid advancement of video image processing (VIP) technologies. Cameras are installed (digital video) typically on a mast-arm, often at a signalized intersection location. The camera can be focused on a single lane, but can also be used to monitor multiple lanes. The system consists of the camera, a microprocessor to store and interpret images, and software that converts the images to traffic data. In essence, vehicles appear on a video image as a compressed package of pixels moving across the image. The system is calibrated to recognize the background image and account for changes due to ambient and artificial light and weather. Available software can discern and classify vehicles by length, count by lane, and provide speeds. VIPs are now being used to operate actuated signals in some locations, and it is almost certain that this technology will advance rapidly over the near and moderate term future.

9.8 Closing Comments

This chapter attempts to provide a broad overview of the complex subject of data collection and reduction for traffic engineering studies. Subsequent chapters discuss specific types of studies in greater detail, and present a more comprehensive picture of how specific types of data are analyzed, and how appropriate conclusions are drawn. The technology for collecting, storing and retrieval, and reduction of data continues to advance at a rapid pace, and the reader is encouraged to check the most current literature to get a more up-to-date view of the state-of-the-art.

References

- 1. *Traffic Monitoring Guide*, Federal Highway Administration, Washington, D.C., 2016.
- 2. Wolshen, B, and Pande, A. (Ed), *Traffic Engineering Handbook*, 7th Edition, John Wiley & Sons, Institute of Transportation Engineers, Washington, D.C., January 2016.
- 3. Schroeder, B., et al., *Manual of Transportation Engineering Studies*, 2nd Edition, Institute of Transportation Engineers, Washington, D.C., 2010.
- 4. *The Highway Safety Manual*, American Association of State Highway and Transportation Officials, Washington, D.C., 2010.
- 5. *Traffic Detector Handbook*, 3rd Edition, Federal Highway Administration, Publication No. FHWA-HRT-06-108, Washington, D.C., 2006.

Problems

1. 9-1. Traffic volumes on a four-lane freeway (two lanes in each direction) were counted manually from an overhead location, resulting in the data shown below. The desire was to obtain continuous 15-minute counts for each lane of the freeway for a two-hour period surrounding the morning peak hour.

Data for Problem 1

Time of Count (PM)	Eastbound		Westbound	
	Lane 1	Lane 2	Lane 1	Lane 2
4:00–4:12		360		310
4:15–4:27	350		285	
4:30–4:42		380		330
4:45–4:57	370		300	
5:00–5:12		370		340
5:15–5:27	345		280	
5:30–5:42		340		310
5:45–5:57	320		260	

Full Alternative Text

From the data shown, determine the following:

1. Continuous 15-minute volumes for each period and each lane.
2. The peak-hour, peak-hour volume, and PHF for each direction of flow, and for the freeway as a whole.
3. Directional flow rates during each 15-minute count period.

2. 9-2. A 24-hour count using a road tube at a rural highway location produces a count of 11,250 actuations. A representative sample count to classify vehicles resulted in the data shown below:

Data for Problem 2

Number of Axles per Vehicle	Number of Vehicles Observed
2	157
3	55
4	50
5	33
6	8

Full Alternative Text

Based upon this sample classification count, how many vehicles were observed during the 24-hour study? Note that this question ignores the effects of "double arrivals" that could lead to undercounts; there are usually correction factors taking that into account.

3. 9-3. A manual speed study is set up with the observer 50 ft (perpendicularly) from the roadway (d_1). The angle of observation to the far end of the speed trap is 70 degrees.

From this information determine:

1. The effective distance over which travel times are observed?
2. What is the speed of a vehicle with a travel time measured as 2.15 s?

Consult Figure 9.7 in considering this problem.

4. 9-4. The text makes the observation with regard to counting traffic that:

   > One jurisdiction does an annual one-week screenline count, traditionally using ATRs. When a different technology was considered, the immediate questions were: Why not simply leave them there, and get continuous counts year-round, including seasonal trends? Given the long history of the annual ATR counts, will the newer technology provide accurate data, using the ATR as the baseline? Why consider the ATR as the baseline, given that well-installed side fire devices, ATR's, and manual counts all have some margin of error in measuring ground-truth traffic present? Even 2 or 3 crew members concurrently counting manually will obtain different results, for instance.

   Write a position paper (not exceeding five pages) addressing each of these questions raised, and specify an evaluation plan by which a senior manager can make an informed decision. Add other issues you deem relevant, and address them. For the purpose of your paper, assume that the traditional counts have been inbound-traffic-only, 24 hours per day, 7 consecutive days (the 3rd week in September) on 40 surface streets, 2 or 3 lanes in each direction.

5. 9-5. Use a search engine to find whether your state and your city maintain a database of historic ADT and other data, available to the public. Use it to find available counts and other data within two miles of your college, office, or home. Summarize the data.

   Likewise, use a search engine to find whether traffic impact studies are available online and whether the supporting data is also available online, in the same state and city/local jurisdiction. If you need more information on "traffic impact studies," see Chapter 14 of this text or search online.

   For this question, do *not* "simplify" your work by making inquiries of the local transportation agencies. Use online searching only.

   Summarize in five pages or less.

6. 9-6. Visit the web site for Miovision^{TM}, and explore how it is used for intersection counts by movement, and also classified by vehicle type. Do a web search for comparable products. Look for technical articles on such technologies, particularly any comparative studies (to each other, or to more traditional ways of collecting the same data). Be careful to distinguish between technical articles and marketing materials. Search online for reports of the actual use of the technology, or of trends in its use. Summarize in five pages or less.

Chapter 10 Traffic Volume Studies and Characteristics

The most fundamental measure in traffic engineering is volume: how many vehicles are passing defined locations in the roadway system over time—particularly during the peak hour(s) of a typical day. There is virtually no decision concerning facility design or traffic control options that can be made without knowledge of existing and projected traffic volumes for the location(s) under study.

In Chapter 5, the concepts of volume and flow rate were introduced. There are four key variables that are related:

- Volume
- Rate of flow
- Demand
- Capacity

The units that describe numeric values of these parameters are all the same: vehicles per hour (veh/h) or passenger cars per hour (pc/h). They can also be cited in “per lane” terms. Chapter 6 contains a detailed discussion of the critical differences and relationships between and among these variables.

Techniques for collection and handling of volume (and other) traffic data were discussed in Chapter 9. This chapter presents techniques for statistical analysis of volume data, and the interpretation and presentation of study results. It also provides an overview of typical volume characteristics found on most highway systems.

10.1 Volume Characteristics

If traffic distributed itself uniformly among the 365×24=8,760 hours of the year, there is not a location in the nation that would experience congestion or significant delay. The problem for traffic engineers, of course, is that there are strong peaks during a typical day, caused primarily by commuters going to and from work. Depending upon the specific region and location, the peak hour of the day typically contains 10–15% of the 24-hour volume. In remote or rural areas, the percentage can go much higher, but the volumes are much lower in these surroundings.

The traffic engineer, therefore, must deal with the travel preferences of our society in planning, designing, and operating highway systems. In some dense urban areas, policies to induce spreading of the peak have been attempted, including the institution of flex-hours or days and/or variable pricing policies for toll and parking facilities. Nevertheless, the traffic engineer must still face the fundamental problem: traffic demand varies in time in ways that are quite inefficient. Demand varies by time of day, by day of the week, by month or season of the year, and in response to singular events (both planned and unplanned) such as construction detours, accidents or other incidents, and even severe weather. Modern intelligent transportation system (ITS) technologies will increasingly try to manage demand on a real-time basis by providing information on routes, current travel times, and related conditions directly to drivers. This is a rapidly growing technology sector, but its impacts have not yet been well documented.

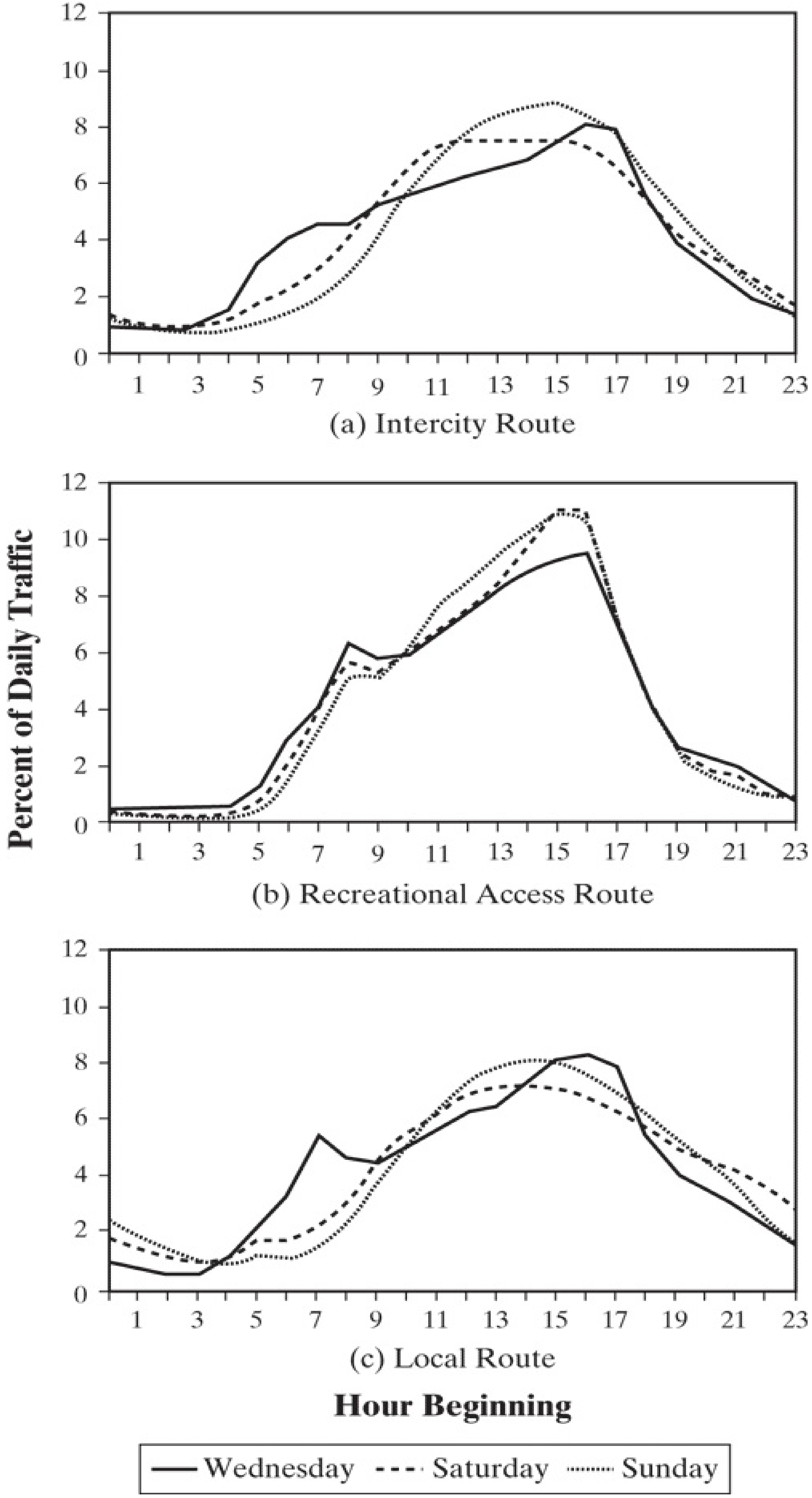
One of the many reasons for doing volume studies is to document these complex variation patterns and to evaluate the impact of ITS technologies and other measures on traffic demand.

10.1.1 Hourly Traffic Variation Patterns: The Phenomenon of the Peak Hour

When hourly traffic patterns are contemplated, we have been conditioned to think in terms of two “peak hours” of the day: morning and evening. Dominated by commuters going to work in the morning (usually between 7AM and 10AM) and returning in the evening (usually between 4PM and 7PM), these patterns tend to be repetitive and more predictable than other facets of traffic demand. This so-called typical pattern holds only for weekday travel, and modern evidence may suggest that this pattern is not as typical as we have been inclined to accept.

Figures 10.1 and 10.2 show hourly variation patterns documented in the *2016 Highway Capacity Manual* [1], for rural and urban roadways, respectively. Figure 10.1 uses data from the Washington State and Oregon Departments of Transportation; Figure 10.2 originates from a paper by McShane and Crowley [2].

Figure 10.1: Typical Hourly Variation Patterns on Rural Roadways

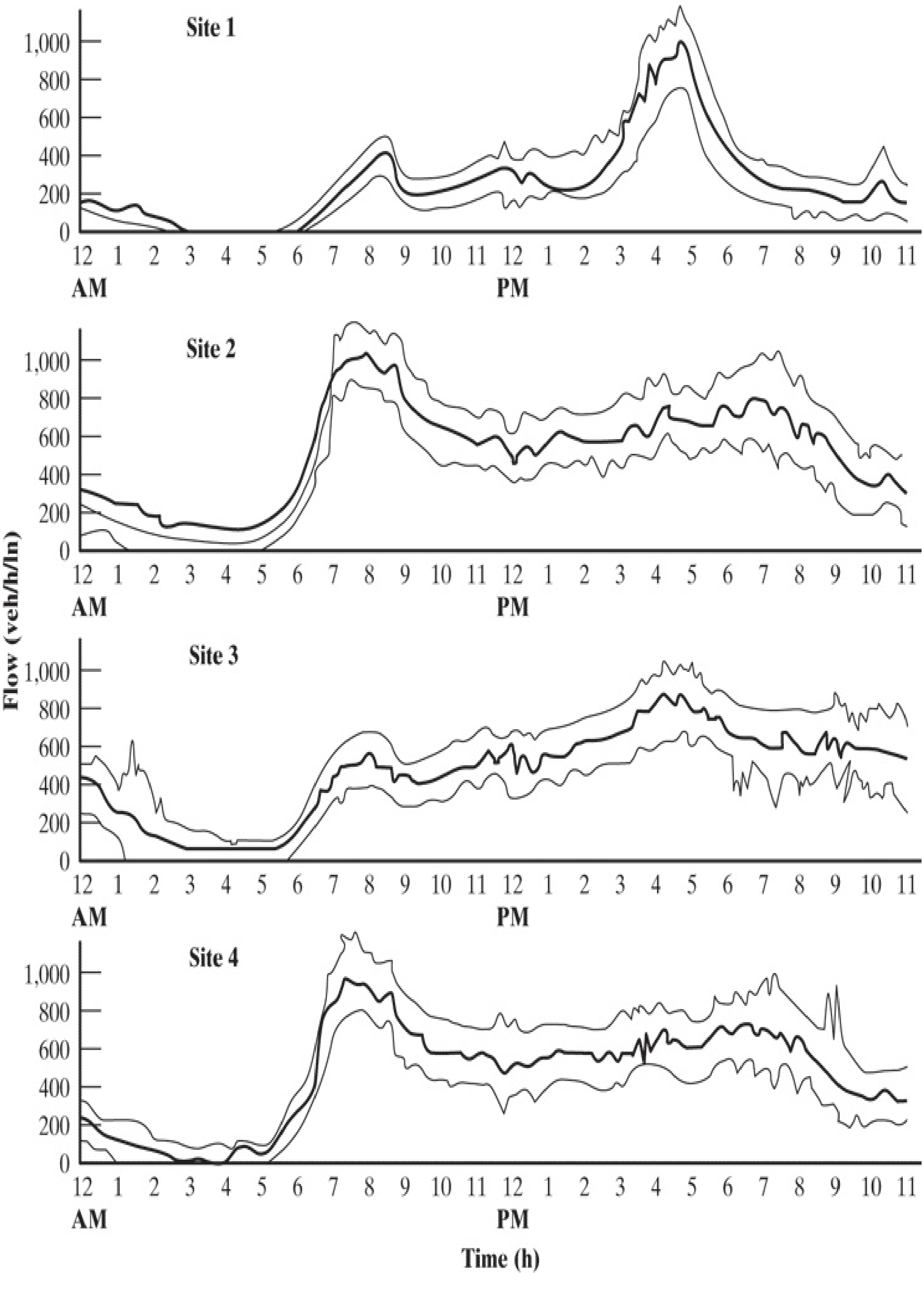


Percent of Daily Traffic
(a) Intercity Route
(b) Recreational Access Route
(c) Local Route
Hour Beginning
Wednesday
Saturday
Sunday

(Source: Reprinted with permission from *Highway Capacity Manual, 6th Edition: A Guide for Multimodal Mobility Analysis*, Transportation Research Board, © 2016 by the National Academy of Sciences, Courtesy of the National Academies Press, Washington, D.C.)

Figure 10.1: Full Alternative Text

Figure 10.2: Typical Hourly Variation Patterns on Urban Roadways



(Source: Reprinted with permission from *Highway Capacity Manual, 6th Edition: A Guide for Multimodal Mobility Analysis*, Transportation Research Board, © 2016 by the National Academy of Sciences, Courtesy of the National Academies Press, Washington, D.C.)

Figure 10.2: Full Alternative Text

In Figure 10.1, only the weekday (Wednesday) traffic shows AM and PM peaks—and the AM peaks are much smaller than the PM peaks. Intercity, recreational, and weekend traffic tend to show a single, more extended peak occurring in the PM, generally in early afternoon.

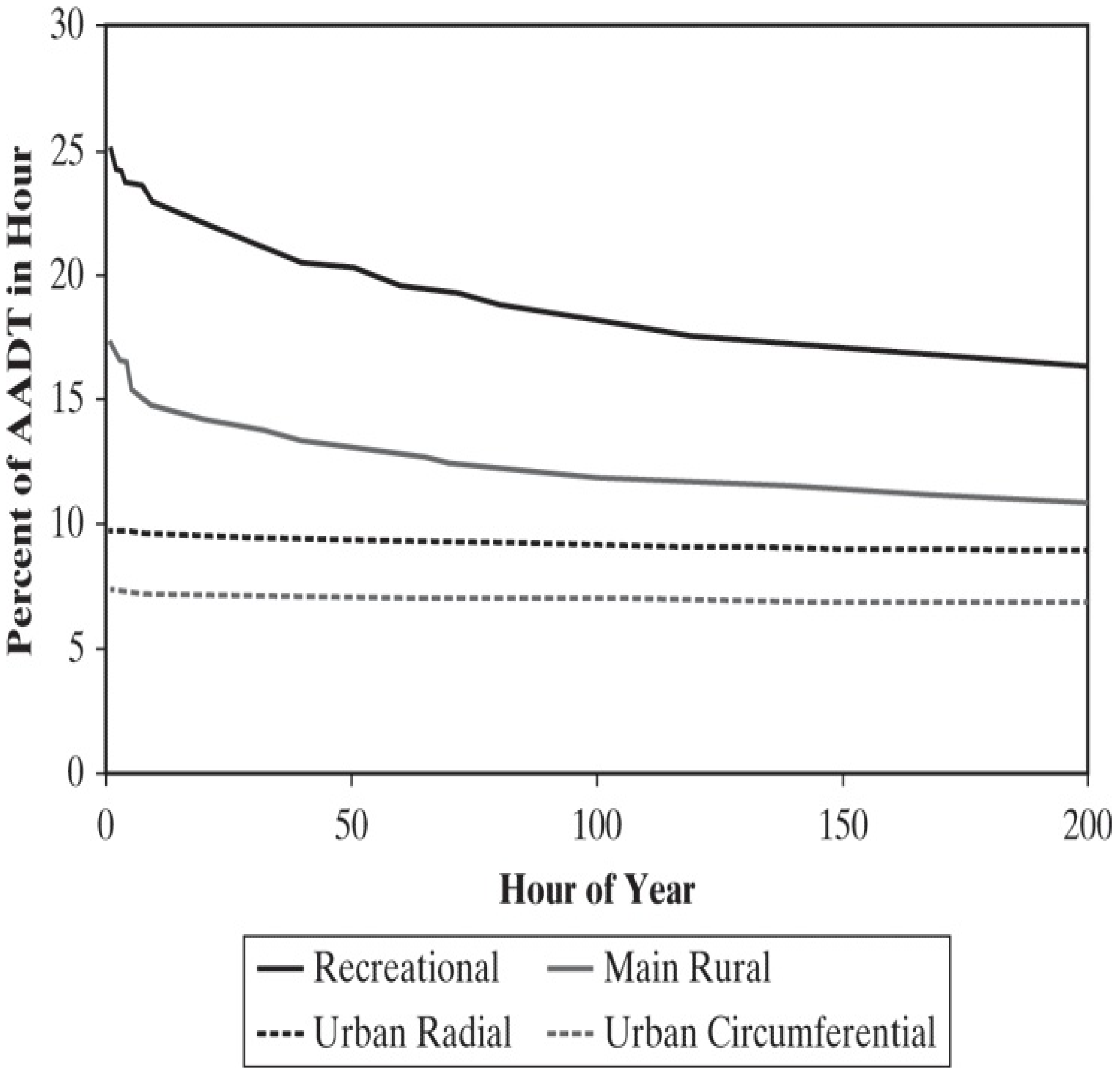
In Figure 10.2, weekday data from four urban sites are shown in a single direction. Sites 1 and 3 are in the opposite direction from Sites 2 and 4, but at the same location. Sites 1 and 2 and Sites 3 and 4 are only two blocks away from each other. Sites 2 and 4 show clear AM peaks, but traffic after the peak stays relatively high, and amazingly uniform for most of the day. Sites 1 and 3 (in the opposite direction) show only evening peaks. Site 3 also displays considerable off-peak traffic. Only Site 1 shows a strong PM peak with significantly less traffic during other portions of the day.

The absence of clear AM and PM peaks in many major urban areas is a spreading phenomenon. On one major facility, the Long Island Expressway (I-495) in New York, a recent study showed that, on a typical weekday, only one peak was discernible in traffic volume data—and it lasted for 10 to 12 hours per day. This characteristic is a direct result of system capacity constraints. Everyone who would like to drive during the normal peak hours cannot be accommodated. Because of this, individuals begin to make travel choices that allow them to increasingly travel during the “off-peak” hours. This process continues until off-peak periods are virtually impossible to separate from peak periods.

Figure 10.2 displays another interesting characteristic of note. The outer lines of each plot show the 95% confidence intervals for hourly volumes over the course of one year. Traffic engineers depend on the basic repeatability of peak-hour traffic demands. The variation in these volumes, however, is not insignificant. During the course of any given year, there are 365 peak hours at any location, one for each day of the year. The question for the traffic engineer is: Which one should be used for planning, design, and operations?

Figure 10.3 shows plots of peak-hour volumes (as a percentage of AADT) in decreasing order for a variety of facilities in Washington State. In all cases, there is clearly a “highest” peak hour of the year. The difference between this highest peak and the bulk of the year’s peak hours, however, depends upon the type of facility. The recreational route has greatest disparity.

Figure 10.3: Peak Hours as a Percentage of AADT



(Source: Reprinted with permission from *Guide for Multimodal Mobility Analysis*, Transportation Research Board, © 2016 by the National Academy of Sciences, Courtesy of the National Academies Press, Washington, D.C.)

Figure 10.3: Full Alternative Text

This is not unexpected, as traffic on such a route will tend to have enormous peaks during the appropriate season on weekends, with far less traffic on a “normal” day. The main rural route has less of a disparity, as at least some component of traffic consists of regular commuters. Urban roadways show far less of a gap between the highest hour and the bulk of peak hours.

It is interesting to examine the various peak hours for the types of facilities illustrated in Figure 10.3, which represents data from various facilities in Washington State. Table 10.1 tabulates the percentage of AADT occurring within designated peak hours for the facility types represented.

Table 10.1: Key Values from Figure 10.3

Type of Facility	Percent* of AADT Occurring in the ___ Peak Hour			
	1st	30th	100th	200th
Recreational	25%	22%	18%	15%
Main Rural	17%	14%	12%	11%
Urban Radial	10%	9%	9%	9%
Urban Circumferential	7%	7%	6%	6%

*Rounded to the nearest whole percentage.

Table 10.1: Full Alternative Text

The choice of which peak hour to use as a basis for planning, design, and operations is most critical for the recreational access route. In this case, the highest hour of the year carries 1.67 times as much traffic as the 200th peak hour of the year, and 1.14 times that of the 30th hour of the year. In the two urban cases, the highest hour of the year is only 1.11 and 1.17 times the 200th highest hour, respectively.

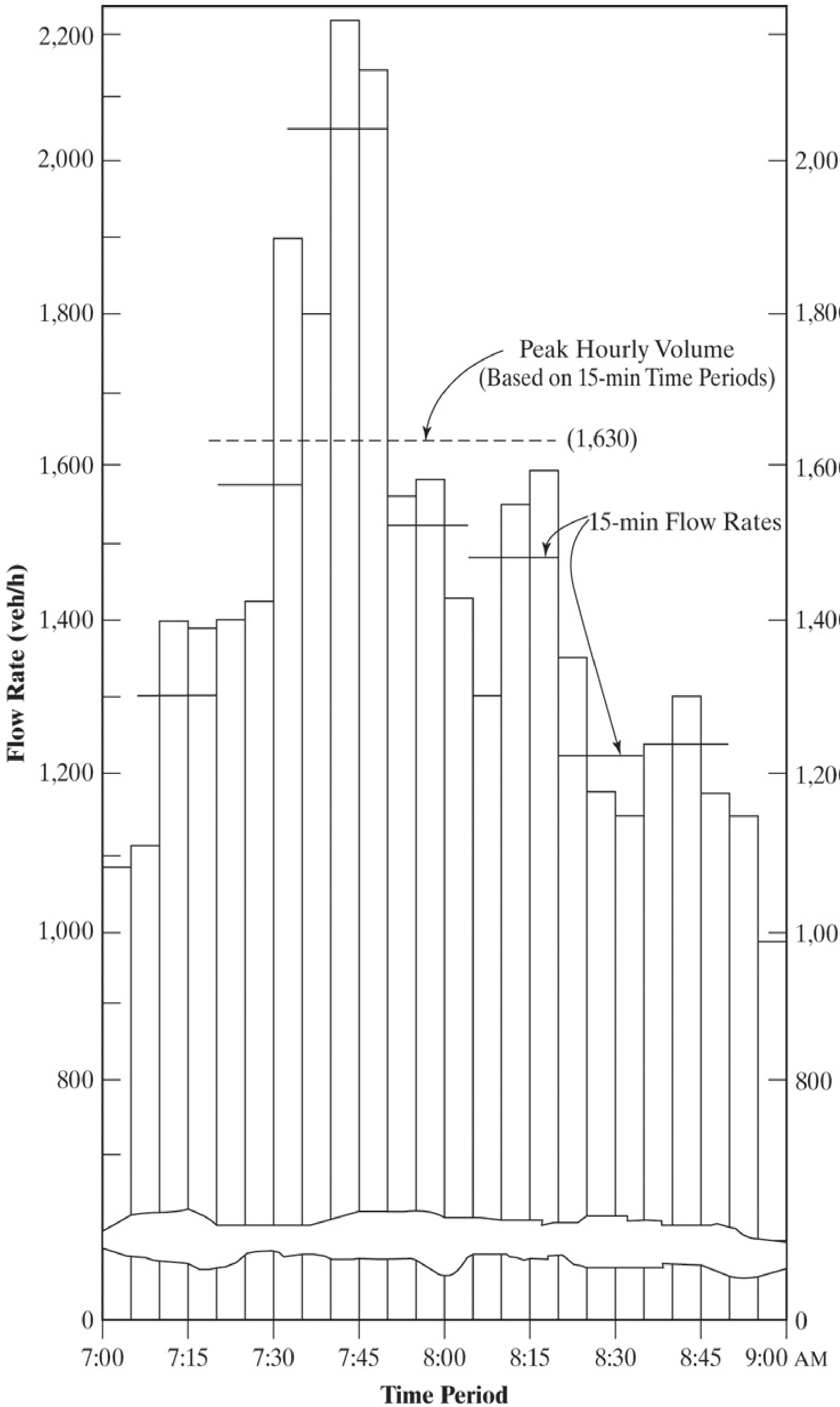
Historically, the 30th highest hour has been used in rural planning, design, and operations. There are two primary arguments for such a policy: (1) the target demand would be exceeded only 29 times per year and (2) the 30th peak hour generally marks a point where subsequent peak hours have similar volumes. The latter defines a point on many relationships where the curve begins to “flatten out,” a range of demands where it is deemed economic to invest in additional roadway capacity.

In urban settings, the choice of a design hour is far less clear and has far less impact. Typical design hours selected range from the 30th highest hour to the 100th highest hour. For the urban facilities of Figure 10.3, this choice represents a range of 9–10.0% of AADT for radial routes, and 6–7% for circumferential routes. With an AADT of 80,000 veh/day, for example, this range is a difference of 800 veh/h in demand. While this is not insignificant, given the stochastic elements in observed volumes, it is not enormous either.

10.1.2 Subhourly Variation Patterns: Flow Rates versus Volumes

In Chapter 5, it was noted that peaking of traffic flows within the peak hour often needed to be considered in design and operations. The peak-hour factor (PHF) was defined as a means of quantifying the difference between a maximum flow rate and the hourly volume within the peak hour. Figure 10.4 shows the difference among 5-minute, 15-minute, and peak-hourly flow rates from a freeway location in Minnesota.

Figure 10.4: Variations of Flow within the Peak Hour



2,200
2,000
1,800
1,600
1,400
1,200
1,000
800
0
Flow Rate (veh/h)
Peak Hourly Volume
(Based on 15-min Time Periods)
(1,630)
15-min Flow Rates
7:00
7:15
7:30
7:45
8:00
8:15
8:30
8:45
9:00 AM
Time Period

(Source: Reprinted with permission from *Highway Capacity Manual, 4th Edition*, Transportation Research Board, © 2000, by the National Academy of Sciences, Courtesy of the National Academies Press, Washington, D.C.)

Figure 10.4: Full Alternative Text

Flow rates can be measured for almost any period of time. For research purposes, periods from 1 to 5 minutes have frequently been used. Very small increments of time, however, become impractical at some point. In a 2-second interval, the range of volumes in a given lane would be limited to "0" or "1," and flow rates would be statistically meaningless.

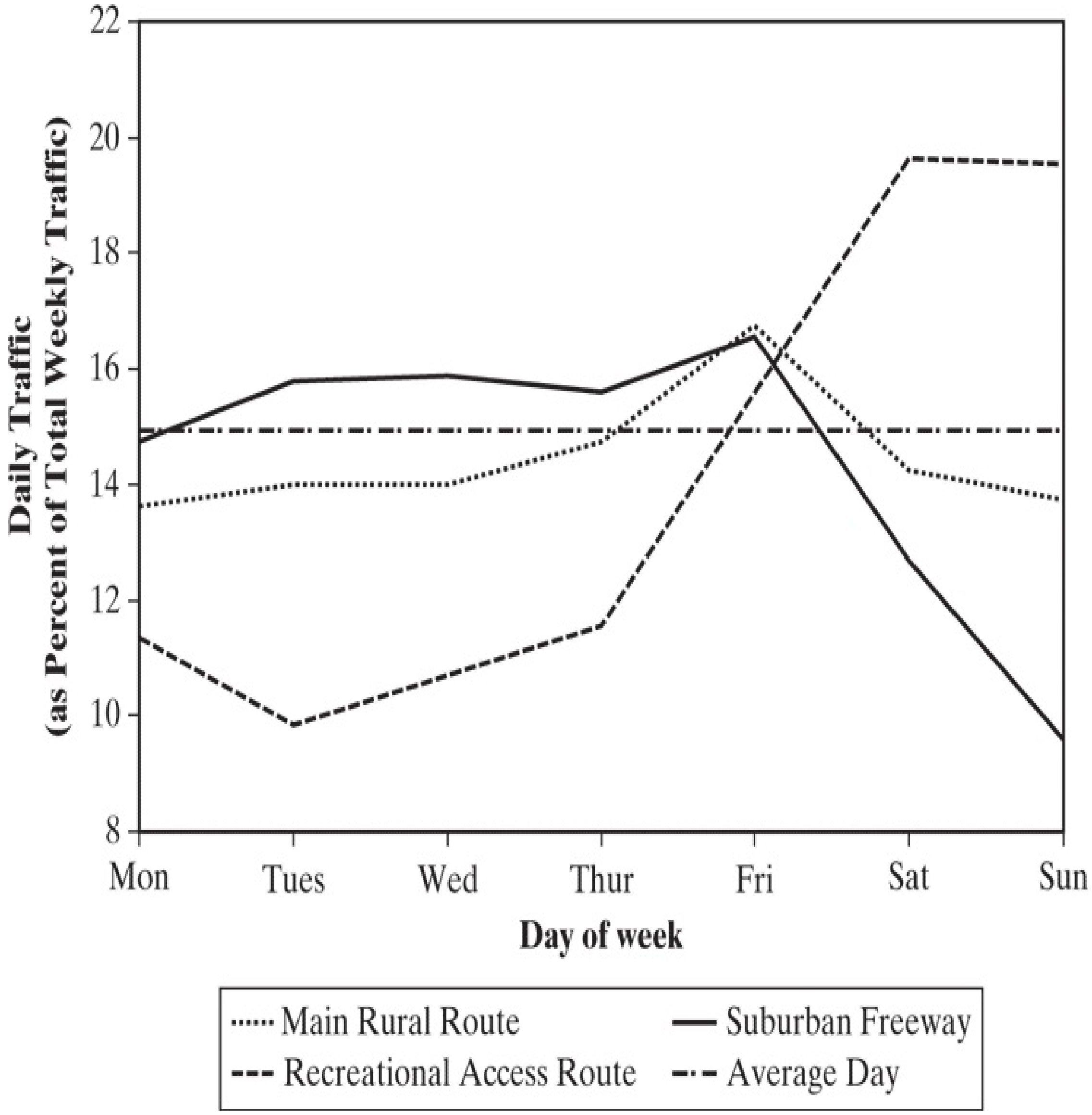
For most traffic engineering applications, 15 minutes is the standard time period used, primarily based on the belief that this is the shortest period of time over which flow rates are "statistically stable." "Statistically stable" implies that reasonable relationships can be calibrated among flow parameters, such as flow rate, speed, and density. In recent years, there is some thought that 5-minute flow rates might qualify as statistically stable, particularly on freeway facilities. Practice, however, continues to use 15 minutes as the standard period for flow rates.

The choice, however, has major implications. In Figure 10.4, the highest 5-minute rate of flow is 2,200 veh/h/ln; the highest 15-minute rate of flow is 2,050 veh/h/ln; the peak-hour volume is 1,630 veh/h/ln. Selecting a 15-minute base period for design and analysis means that, in this case, the demand flow rate (assuming no capacity constraints) would be 2,050 veh/h/ln. This value is 7% lower than the peak 5-minute flow rate and 20% higher than the peak-hour volume. In real design terms, these differences could translate into a design with one more or fewer lanes or differences in other geometric and control features. The use of 15-minute flow periods also implies that breakdowns of a shorter duration do not cause the kinds of instabilities that accompany breakdowns extending for 15 minutes or more.

10.1.3 Daily Variation Patterns

Traffic volumes also conform to daily variation patterns that are caused by the type of land uses and trip purposes served by the facility. Figure 10.5 illustrates some typical relationships.

Figure 10.5: Typical Daily Variations in Volume



(Source: Reprinted with permission from *A Guide for Multimodal Mobility Analysis*, Transportation Research Board, © 2016 by the National Academy of Sciences, Courtesy of the National Academies Press, Washington, D.C.)

Figure 10.5: Full Alternative Text

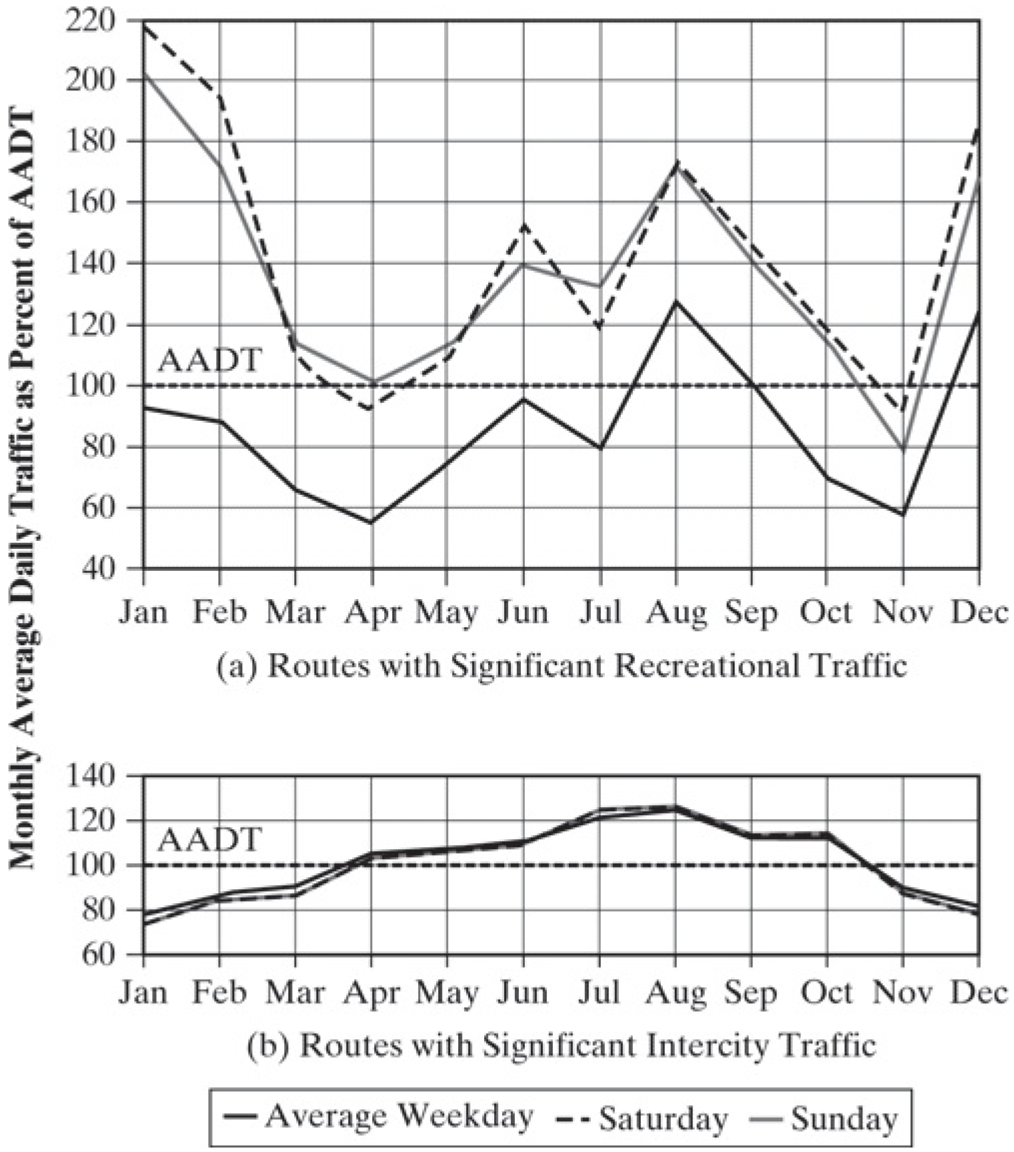
The recreational access route displays strong peaks on Saturdays and Sundays. This is a typical pattern for such routes, given the prevalence of recreational travel on weekends. Tuesdays through Thursdays have far less traffic demand. Friday, however, also carries heavier traffic than the average day, as weekend travelers get an early start. Monday is somewhat higher than other weekdays due to some vacationers returning after the weekend rather than on Sunday.

The suburban freeway obviously caters to commuters. Commuter trips are virtually a mirror image of recreational trips, with peaks occurring on weekdays and lower demand on weekends. The main rural route in this exhibit has a pattern similar to the recreational route, but with less variation between the weekdays and weekends. The route serves both recreational and commuter trips, and the mix tends to dampen the amount of variation observed.

10.1.4 Monthly or Seasonal Variation Patterns

Figures 10.6, 10.7, and 10.8 illustrate typical monthly volume variation patterns for various types of urban and rural facilities.

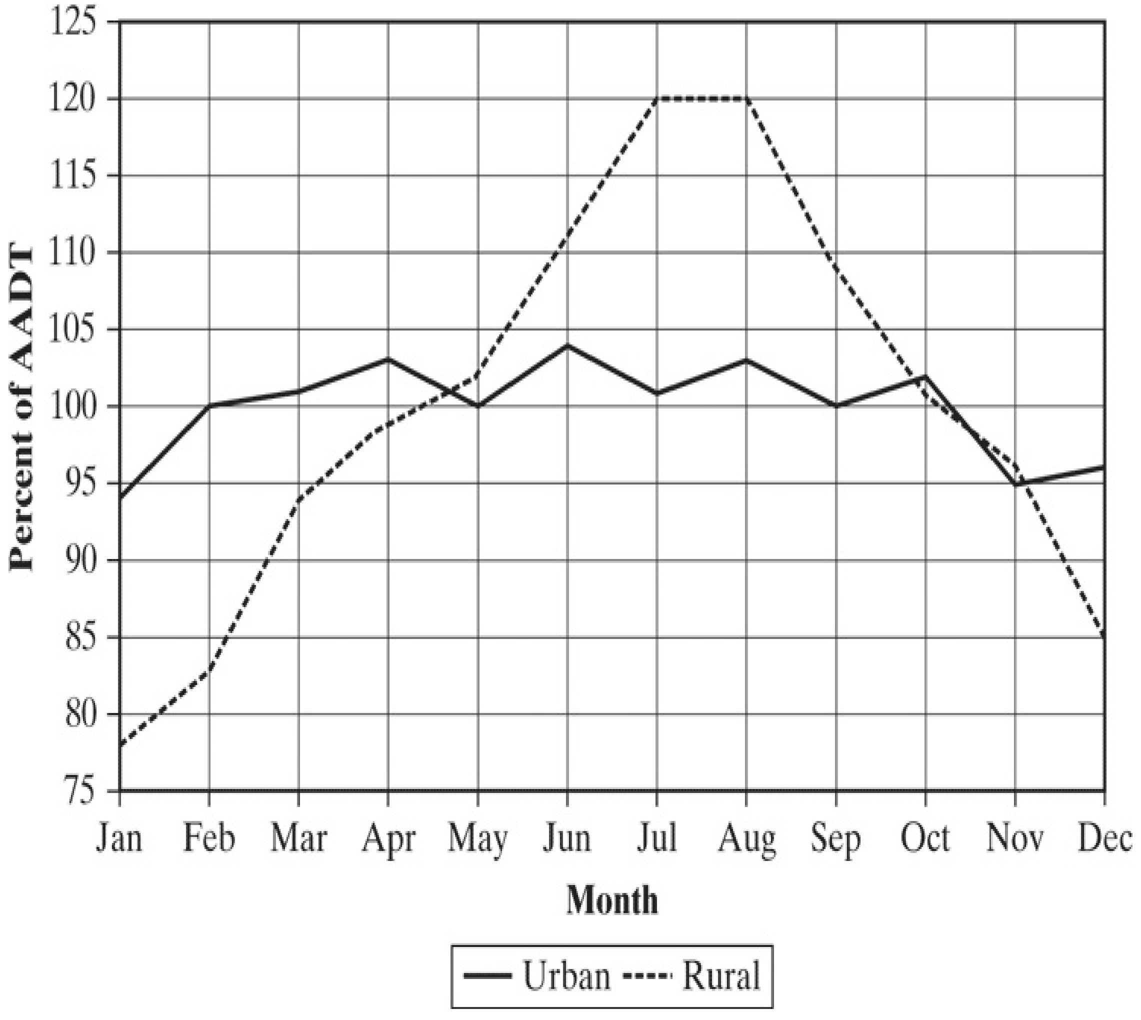
Figure 10.6: Monthly Variations on Typical Intercity and Recreational Highways



(Source: Reprinted with permission from *Highway Capacity Manual, 6th Edition: A Guide for Multimodal Mobility Analysis*, Transportation Research Board, © 2016 by the National Academy of Sciences, Courtesy of the National Academies Press, Washington, D.C.)

Figure 10.6: Full Alternative Text

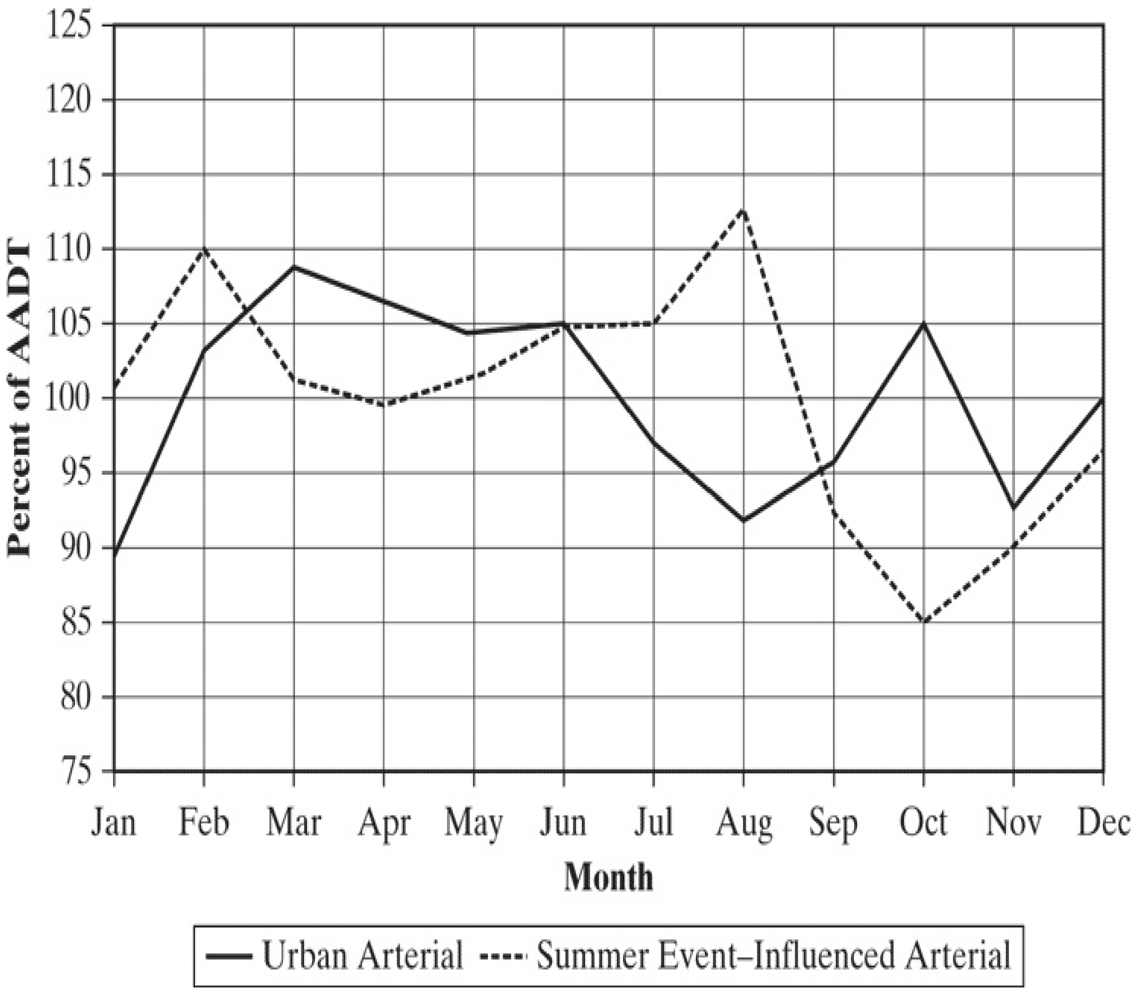
Figure 10.7: Typical Monthly Variation on an Interstate Highway



(Source: Reprinted with permission from *Highway Capacity Manual, 6th Edition: A Guide to Multimodal Mobility Analysis,* Transportation Research Board, © 2016 by the National Academy of Sciences, Courtesy of the National Academies Press, Washington, D.C.)

Figure 10.7: Full Alternative Text

Figure 10.8: Typical Monthly Variation on Urban Streets



(Source: Reprinted with permission from *A Guide to Multimodal Mobility Analysis*, Transportation Research Board, Washington D.C., 2016 © 2016, by the National Academy of Sciences, Courtesy of the National Academies Press, Washington, D.C.)

Figure 10.8: Full Alternative Text

- Figure 10.6(a) shows that typical recreational routes have similar variation patterns on weekdays and weekends, but that weekend volumes, as expected, are significantly higher than weekday volumes. As monthly peak periods occur both in winter and in summer, it is likely that a variety of recreational opportunities are served by the highways included in the data.

- Figure 10.6(b) shows that on intercity routes, there is virtually no difference between weekday and weekend volume patterns. Intercity routes often serve commuter, recreational, and other travel needs. The facilities included in the data clearly have their highest traffic volumes during the summer.

- The rural freeway in Figure 10.7 has a significant summer peak, indicating that it serves summer recreational destinations. The urban freeway has its lowest traffic volumes in December and January—winter months. This may reflect weather-induced decreases in demand and/or people taking vacations in warmer climates.

- Figure 10.8 is interesting. The “summer event influenced” streets show two clear peaks, one in the summer (the largest) and another in the winter. The urban arterial has low points in January and August. These patterns are likely due to unique characteristics of the destinations served by the facilities in the study. Urban street patterns depend greatly on local conditions, and it is difficult to identify truly “typical” conditions.

It might be expected that commuter routes would show a trend opposite to recreational routes (i.e., if recreational routes are peaking in the summer, then commuter routes should have less traffic during those periods). Figures 10.6, 10.7, and 10.8 do not reflect this. The problem is that few facilities are purely recreational or commuter; there is always some mix present. Further, much recreational travel is done by inhabitants of the region in question; the same motorists may be part of both the recreational and commuter demand during the same months. There are, however, some areas in which commuter traffic does clearly decline during summer recreational months. The distributions shown here are illustrative; different distributions are possible, and they do occur in other regions.

10.1.5 Some Final Thoughts on Volume Variation Patterns

One of the most difficult problems in traffic engineering is that we are continually planning and designing for a demand that represents a peak flow rate within a peak hour on a peak day during a peak season. When we are successful, the resulting facilities are underutilized most of the time.

It is only through the careful documentation of these variation patterns, however, that the traffic engineer can know the impact of this underutilization. Knowing the volume variation patterns governing a particular area or location is critical to finding appropriate design and control measures to optimize operations. It is also important to document these patterns so that estimates of an average annual daily traffic (AADT) can be discerned from data taken for much shorter time periods. It is simply impractical to count every location for a full year to determine AADT and related demand factors. Counts taken over a shorter period of time can, however, be adjusted to reflect a yearly average or a peak occurring during another part of the year, if the variation patterns are known and well documented. These concepts will be illustrated and applied in the sections that follow.

10.2 Intersection Volume Studies

There is no single location more complex in a traffic system than an at-grade intersection. At a typical 4-leg intersection, there are 12 separate movements—left, through, and right from each leg. If a count of intersection volumes is desired, with each movement classified by cars, taxis, trucks, and buses, each count period requires the observation of $12 \times 4 = 48$ separate pieces of data.

When intersections are counted manually (and they often are), observers must be positioned to properly see the movements they are counting. It is doubtful that an inexperienced counter could observe and classify more than one major or two minor movements simultaneously. For heavily used multilane approaches, it may be necessary to use separate observers for different lanes. In manual intersection studies, short-break and alternating-period approaches are almost always combined to reduce the number of observers needed. Rarely, however, can an intersection be counted with fewer than four observers, plus one crew chief to time count periods and breaks.

10.2.1 Arrival versus Departure Volumes: A Key Issue for Intersection Studies

At most intersections, volumes are counted as they depart the intersection. This is done both for convenience and because turning movements cannot be fully resolved until vehicles exit the intersection. Although this approach is fine where there is no capacity constraint (i.e., an unstable build-up of queues on the approach), it is not acceptable where demand exceeds the capacity of the approach. In such cases, it is necessary to observe *arrival* volumes, as these are a more accurate reflection of demand.

At signalized intersections, “unstable queue build-up” is detected when vehicles queued during a red interval are not fully cleared during the next green interval. At unsignalized intersections, “unstable queue build-up” can be identified by queues that become larger during each successive counting period.

Direct observation of arrival volumes at an intersection is difficult, as the queue is dynamic. As the queue grows and declines, the point of “arrival” changes. Therefore, the technique used to count arrival volumes is to count departure volumes and the number of queued vehicles at periodic intervals. For signalized approaches, the size of the queue would be recorded *at the beginning of each red phase*. This identifies the “residual queue” of vehicles that arrived during the previous signal cycle but were not serviced. For unsignalized approaches, the queue is counted at the end of each count period. When such an approach is followed, the arrival volume is estimated as:

$V_{ai} = V_{di} + N_{qi} - N_{q(i-1)}$ [10-1]

where:

V_{ai} = arrival volume during period i, veh, V_{di} =
departure volume during period i, veh, N_{qi} =

number of queued vehicles at the end of period i, veh, and N q(i−1) = number of queued vehicles at the end of period i−1, veh.

Estimates of arrival volume using this procedure identify only the localized arrival volume. This procedure *does not* identify diverted vehicles or the number of trips that were not made due to general congestion levels. Thus, although arrival volumes do represent localized demand, they do not measure diverted or repressed demand. Table 10.2 shows sample study data using this procedure to estimate arrival volumes.

Table 10.2: Estimating Arrival Volumes from Departure Counts: An Example

Time Period (PM)	Departure Count (vehs)	Queue Length (vehs)	Arrival Volume (vehs)
4:00–4:15	50	0	50
4:15–4:30	55	0	55
4:30–4:45	62	5	62 + 5 = 67
4:45–5:00	65	10	65 + 10 − 5 = 70
5:00–5:15	60	12	60 + 12 − 10 = 62
5:15–5:30	60	5	60 + 5 − 12 = 53
5:30–5:45	62	0	62 − 5 = 57
5:45–6:00	55	0	55
Total	**469**		**469**

Table 10.2: Full Alternative Text

Note that the study is set up so that the first and last count periods do not have residual queues. Also, the total departure and arrival count are the same, but the conversion from departures to arrivals causes a shift in the distribution of volumes by time period. Based on departure counts, the maximum 15-minute volume is 65 veh, or a flow rate of 65/0.25=260 veh/h. Using arrival counts, the maximum 15-minute volume is 70, or a flow rate of 70/0.25=280 veh/h. The difference is important, as the higher arrival flow rate (assuming that the study encompasses the peak period) represents a value that would be valid for use in planning, design, or operations.

10.2.2 Special Considerations for Signalized Intersections

At signalized intersections, count procedures are both simplified and more complicated at the same time. For manual observers, the signalized intersection simplifies counting, as not all movements are flowing at the same time. An observer who can normally count only one through movement at a time could actually count two such movements in the same count period by selecting, for example, the eastbound and northbound through movements. These two operate during different cycles of the signal.

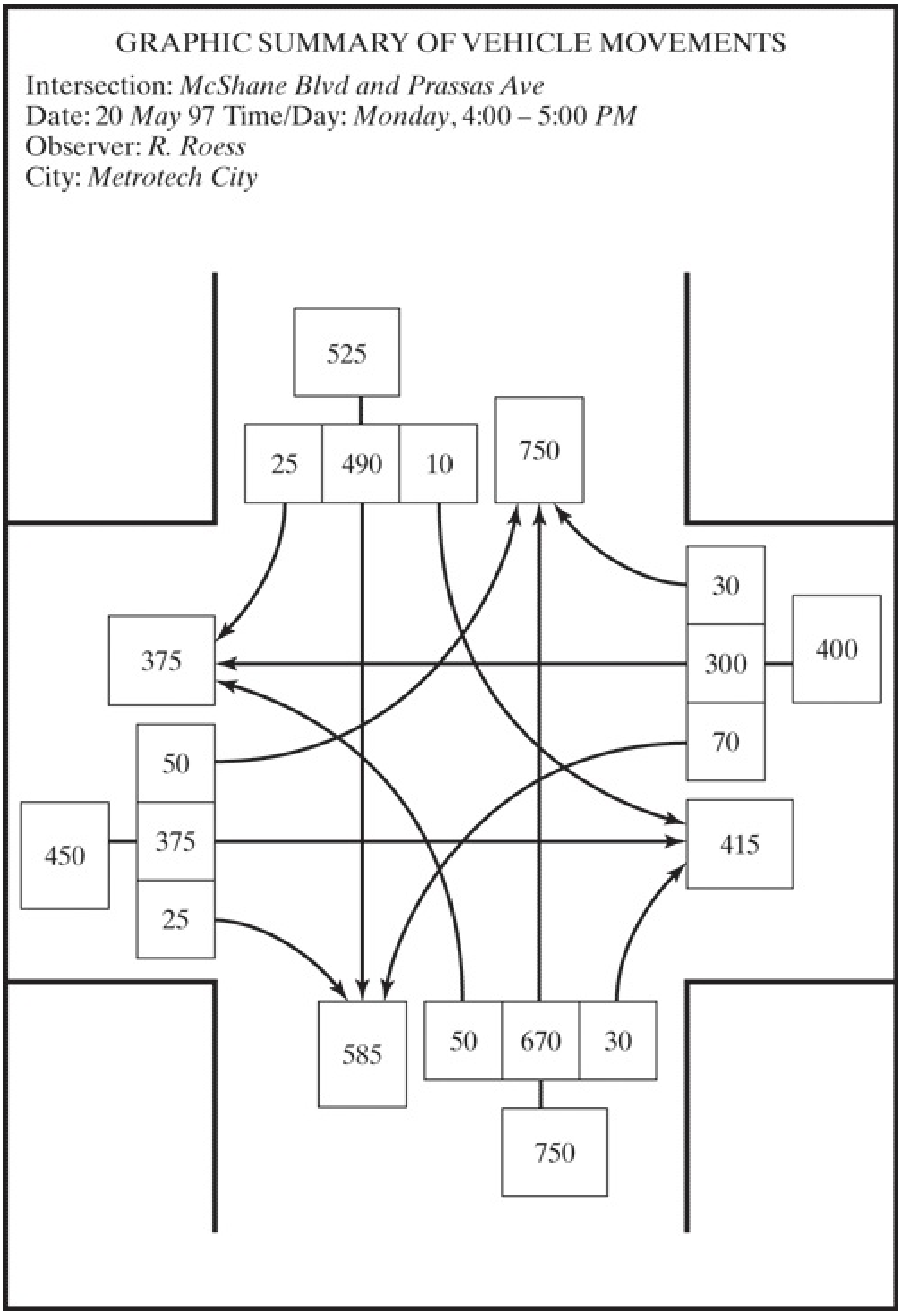
Count periods at signalized intersections, however, must be equal multiples of the cycle length. Further, actual counting times (exclusive of breaks) must also be equal multiples of the cycle length. This is to guarantee that all movements get the same number of green phases within a count period. Thus, for a 60-second signal cycle, a 4 out of 5-minute counting procedure may be employed. For a 90-second cycle, however, neither 4 nor 5 minutes are equal multiples of 90 seconds (1.5 minutes). For a 90-second cycle, a counting process of 12 out of 15 minutes would be appropriate, as would 4.5 out of 6 minutes.

Actuated signals present special problems, as both cycle lengths and green splits vary from cycle to cycle. Count periods are generally set to encompass a minimum of five signal cycles, using the maximum cycle length as a guide. The actual counting sequence is arbitrarily chosen to reflect this principle, but it is not possible to assure equal numbers of phases for each movement in each count period. This is not viewed as a major difficulty, as the premise of actuated signalization is that green times should be allocated proportionally to vehicle demands present during each cycle.

10.2.3 Presentation of Intersection Volume Data

Intersection volume data may be summarized and presented in a variety of ways. Simple tabular arrays can summarize counts for each count period by movement. Breakdowns by vehicle type are also most easily depicted in tables. More elaborate graphic presentations are most often prepared to depict peak-hour and/or full-day volumes. Figures 10.9 and 10.10 illustrate common forms for display of peak-hour or daily data. The first is a graphic intersection summary diagram that allows simple entry of data on a predesigned graphic form. The second is an intersection flow diagram in which the thickness of flow lines is based on relative volumes.

Figure 10.9: Graphic Intersection Summary Diagram



(Source: Used with permission from *Transportation and Traffic Engineering Handbook, 1st Edition*,Institute of Transportation Engineers, Washington DC, 1976, pg 410)

Figure 10.9: Full Alternative Text

Figure 10.10: An Intersection Flow Diagram

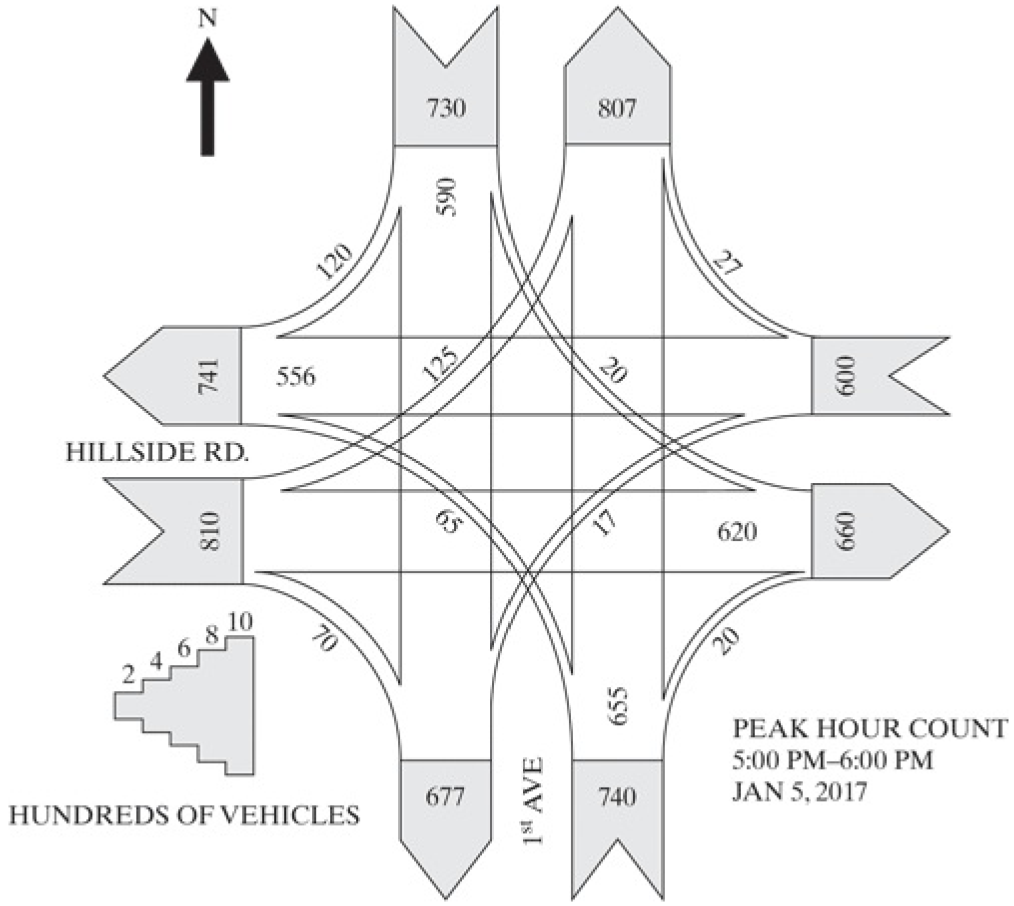


Figure 10.10: Full Alternative Text

10.3 Limited Network Volume Studies

Consider the following proposition: A volume study is to be made covering the period from 6:00 am to 12:00 midnight on the street network comprising midtown Manhattan (i.e., from 14th Street to 59th Street, 1st Avenue to 12th Avenue). Although this is a very big network, including over 500 street links and 500 intersections, it is not the entire city of New York, nor is it a statewide network.

Nevertheless, the size of the network is daunting for a simple reason: it is virtually impossible to acquire and train sufficient personnel to count all of these locations at the same time. Further, it would be impractically expensive to try and acquire sufficient portable counting equipment to do so. To conduct this study, it will be necessary to employ *sampling* techniques (i.e., not all locations within the study area will be counted at the same time, not even on the same day). Statistical manipulation based on these samples will be required to produce an hourly volume map of the network for each hour of the intended survey period, or for an average peak period.

Such “limited” networks exist in both small towns and large cities and around other major trip generators, such as airports, sports facilities, shopping malls, and other activity centers. Volume studies on such networks involve individual planning and some knowledge of basic characteristics, such as location of major generators and the nature of traffic on various facilities (local versus through users, for example). The establishment of a reasonable sampling methodology will require judgment based on such local familiarity.

Sampling procedures rely on the assumption that entire networks, or identifiable subportions of networks, have similar demand patterns in time. If these patterns can be measured at a few locations, the pattern can be superimposed on sample measurements from other locations in the network. To implement such a procedure, two types of counts are conducted:

- Control counts. Control counts are taken at selected representative locations to measure and quantify demand variation patterns in time. In general, control counts must be maintained continuously throughout the study period.
- Coverage counts. Coverage counts are taken at all locations for which data are needed. They are conducted as samples, with each location being counted for only a portion of the study period, in accordance with a pre-established sampling plan.

These types of counts and their use in volume analysis are discussed in the sections that follow.

10.3.1 Control Counts

Because control counts will be used to expand and adjust the results of coverage counts throughout the network under study, it is critical that representative control-count locations be properly selected. The hourly and daily variation patterns observed at a control count must be representative of a larger portion of the network if the sampling procedure is to be accurate and meaningful. It should be remembered that volume variation patterns are generated by land-use characteristics and by the type of traffic, particularly the percentages of through versus locally generated traffic in the traffic stream. With these principles in mind, there are some general guidelines that can be used in the selection of appropriate control-count locations:

1. There should be one control-count location for every 10 to 20 coverage-count locations to be sampled.
2. Different control-count locations should be established for each class of facility in the network—local streets, collectors, arterials, and the like—as different classes of facilities serve different mixes of through and local traffic.
3. Different control-count locations should be established for portions of the network with markedly different land-use characteristics.

These are only general guidelines. The engineer must exercise judgment and use his or her knowledge of the area under study to identify appropriate control-count locations.

10.3.2 Coverage Counts

All locations at which sample counts will be taken are called *coverage counts*. All coverage counts (and control counts as well) in a network study are taken at midblock locations to avoid the difficulty of separately recording turning movements. Each link of the network is counted at least once during the study period. Intersection turning movements may be approximately inferred from successive link volumes, and, when necessary, supplementary intersection counts can be taken. Counts at midblock locations allow for the use of portable automated counters, although the duration of some coverage counts may be too short to justify their use.

10.3.3 An Illustrative Study

The types of computations involved in expanding and adjusting sample network counts is best described by a simple example. Figure 10.11 shows one segment of a larger network that has been identified as having reasonably uniform traffic patterns in time. The network segment has seven links, one of which has been established as a control-count location. The other six links are coverage-count locations at which sample counts will be conducted. The various proposed study procedures all assume that there are only two field crews or automated counters that can be employed simultaneously in this segment of the network. A study procedure is needed to find the volume on each link of the network between 12:00 noon and 8:00 pm on a typical weekday. Three different approaches will be discussed. They are typical and not the only approaches that could be used. However, they illustrate all of the expansion and adjustment computations involved in such studies.

Figure 10.11: A Sample Network Volume Study

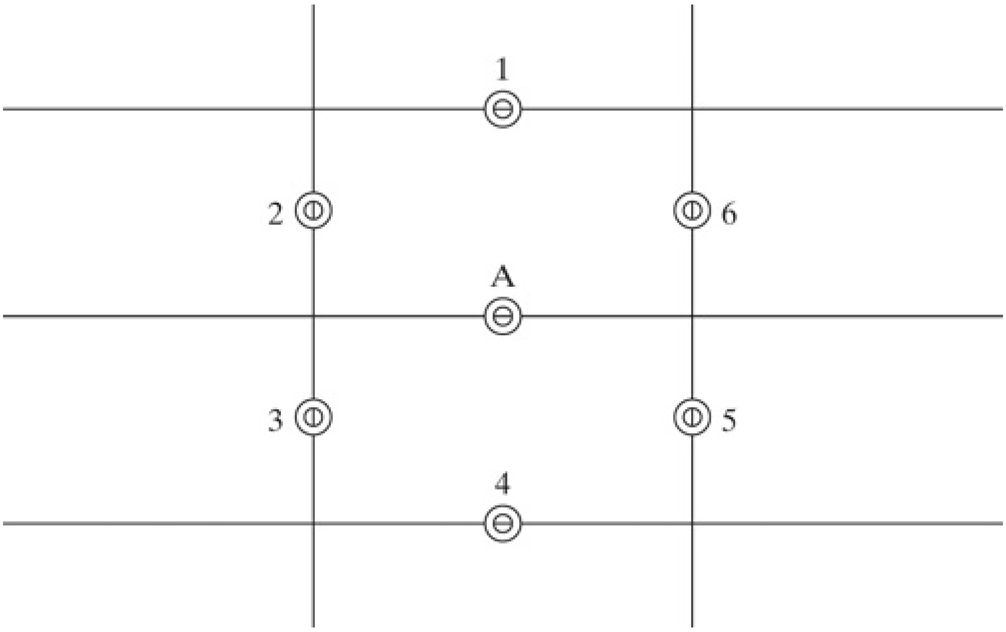


Figure 10.11: Full Alternative Text

Sample Problem 10-1: A One-Day Network Study Plan

It is possible to complete the study in a single day. One of the two available crews or set-ups would be used to count control Location A for the entire 8-hour period of the study. The second crew or set-up would be used to count each of Coverage Locations 1–6 for one hour. The sample data and analysis resulting from this approach are shown in Table 10.3.

Table 10.3: Data and Computations for a One-Day Network Volume Study

Control-Count Data Location A		Coverage-Count Data		
Time (PM)	Count (vehs)	Location	Time (PM)	Count (vehs)
12–1	825	1	12–1	840
1–2	811	2	1–2	625
2–3	912	3	2–3	600
3–4	975	4	4–5	390
4–5	1,056	5	5–6	1,215
5–6	1,153	6	6–7	1,440
6–7	938			
7–8	397			

(a) Data from a One-Day Study

10.3-3 Full Alternative Text

Time (PM)	Count (vehs)	Proportion of 8-Hour Total
12–1	825	825/7,067 = 0.117
1–2	811	811/7,067 = 0.115
2–3	912	912/7,067 = 0.129
3–4	975	975/7,067 = 0.138
4–5	1,056	1,056/7,067 = 0.149
5–6	1,153	1,153/7,067 = 0.163
6–7	938	938/7,067 = 0.133
7–8	397	397/7,067 = 0.056
Total	**7,067**	**1.000**

(b) Computation of Hourly Volume Proportions from Control-Count Data

10.3-4 Full Alternative Text

Location	Time (PM)	Count (vehs)	Estimated 8-Hr Volume (vehs)	Estimated Peak-Hour Volume (vehs)
1	12–1	840	840/0.117 = 7,179	×0.163 = 1,170
2	1–2	625	625/0.115 = 5,435	×0.163 = 886
3	2–3	600	600/0.129 = 4,651	×0.163 = 758
4	4–5	390	390/0.149 = 2,617	×0.163 = 427
5	5–6	1,215	1,215/0.163 = 7,454	×0.163 = 1,215
6	6–7	1,440	1,440/00.133 = 10,827	×0.163 = 1,765

(c) Expansion of Hourly Counts

10.3-5 Full Alternative Text

Note that full-hour data are shown. These data reflect expansion of actual counts that account for break periods. If machine counts were conducted, they would also reflect the conversion of axle counts to vehicle counts.

In Table 10.3(b), the control-count data are used to quantify the hourly variation pattern observed. It is now assumed that this pattern applies to all of coverage locations within the network. Thus, a count of 840 vehicles at Location 1 would represent 0.117 (or 11.7%) of the 8-hour total at this location. The 8-hour total can then be estimated as 840/0.117=7,179 vehicles. Moreover, the peak-hour volume can be estimated as 0.163×7,179=1,170 vehicles, as the hourly distribution shows that the highest volume hour contains 0.163 (or 16.3%) of the 8-hour volume. Note that this expansion of data results in estimates of 8-hour and peak-hour volumes at each of the seven count locations that represent *the day on which the counts were taken*. Daily and seasonal variations have not been eliminated by this study technique. Volumes for the entire network, however, have been estimated for common time periods.

Sample Problem 10-2: A Multiday Network Study Plan

In the one-day study approach, each coverage location was counted for one hour. Based on hourly variation patterns documented at the control location, these counts were expanded into 8-hour volume estimates. Hourly variation patterns, however, are not as stable as variations over larger periods of time. For this reason, it could be argued that a better approach would be to count each coverage location for a full eight hours.

Given the limitation to two simultaneous counts due to personnel and/or equipment, such a study would take place over six days. One crew would monitor the control location for the entire period of the study, while the second would count at one coverage location for eight hours on each of six days.

The data and computations associated with a six-day study are illustrated in Table 10.4. In this case, hourly patterns do not have to be documented, because each coverage location is counted for every hour of the study period. Unfortunately, the counts are spread over six days, over which volume may vary considerably at any given location. In this case, the control data are used to quantify the underlying *daily* variation pattern. The documented variation pattern is used to *adjust* the coverage counts.

Table 10.4: Data and Computations for a 6-Day Study Option

Control-Count Data Location A		Coverage-Count Data		
Day	8-Hour Count (vehs)	Coverage Location	Day	8-Hour Count (Vehs)
Monday 1	7,000	1	Monday 1	6,500
Tuesday	7,700	2	Tuesday	6,200
Wednesday	7,700	3	Wednesday	6,000
Thursday	8,400	4	Thursday	7,100
Friday	7,000	5	Friday	7,800
Monday 2	6,300	6	Monday 2	5,400

(a) Data for a Six-Day Study

10.3-7 Full Alternative Text

Day	8-Hour Count (vehs)	Adjustment Factor
Monday 1	7,000	7,350/7,000 = 1.05
Tuesday	7,700	7,350/7,700 = 0.95
Wednesday	7,700	7,350/7,700 = 0.95
Thursday	8,400	7,350/8,400 = 0.88
Friday	7,000	7,350/7,000 = 1.05
Monday 2	6,300	7,350/6,300 = 1.17
Total	44,100	
Average	44,100/6 = 7,350	

(b) Computation of Daily Adjustment Factors

10.3-8 Full Alternative Text

Station	Day	8-Hour Count (vehs)	Adjusted 8-Hour Count (vehs)
1	Monday 1	6,500	× 1.05 = 6,825
2	Tuesday	6,200	× 0.95 = 5,890
3	Wednesday	6,000	× 0.95 = 5,700
4	Thursday	7,100	× 0.88 = 6,248
5	Friday	7,800	× 1.05 = 8,190
6	Monday 2	5,400	× 1.17 = 6,318

(c) Adjustment of Coverage Counts

10.3-9 Full Alternative Text

Daily volume variations are quantified in terms of adjustment factors defined as follows: the volume for a given day multiplied by the factor yields a volume for the average day of the study period. Stated mathematically:

[&V_{a}|=|V_{i}F_{vi}&]

Va=ViFvi [10-2]

where:

V a = volume for the average day of the study period, veh, V i =

volume for day i, veh, and F vi = adjustment factor for day i.

Using data from the control location, at which the average volume will be known, adjustment factors for each day of the study may be computed as:

[&F_{vi}|=|V_{a}|sol|V_{i}&]

$F_{vi} = V_a / V_i$ [10-3]

where all terms are as previously defined. Factors for the sample study are calibrated in Table 10.4(b). Coverage counts are adjusted using Equation 10-2 in Table 10.4 (c).

The results represent the average 8-hour volumes for all locations for the six-day period of the study. Seasonal variations are not accounted for, nor are weekend days, which were excluded from the study.

Sample Problem 10-3: A Three-Day Study Plan

The first two approaches can be combined. If a one-day study is not deemed appropriate due to the estimation of 8-hour volumes based on 1-hour observations, and the six-day study is too expensive, a three-day study program can be devised in which each coverage location is counted for 4 hours on one of three days. The control location would have to be counted for the entire three-day study period; results would be used to calibrate the distribution of volume by 4-hour period and by day.

In this approach, 4-hour coverage counts must be (1) expanded to reflect the full 8-hour study period and (2) adjusted to reflect the average day of the 3-day study period. Table 10.5 illustrates the data and computations for the three-day study approach.

Table 10.5: Data and Computations for a 3-Day

Study Option

Time (PM)	Monday		Tuesday		Wednesday		Avg % of 8 Hours
	Count (vehs)	% of 8 Hours	Count (vehs)	% of 8 Hours	Count (vehs)	% of 8 Hours	
12–4	3,000	42.9%	3,200	42.7%	2,800	43.8%	43.1%
4–8	4,000	57.1%	4,300	57.3%	3,600	56.2%	56.9%
Total	**7,000**	**100.0%**	**7,500**	**100.0%**	**6,400**	**100.0%**	**100.0%**

(a) Control Data and Calibration of Hourly Variation Pattern

10.3-11 Full Alternative Text

Day	8-Hour Control-Count Location A (vehs)	Adjustment Factor
Monday	7,000	6,967/7,000 = 1.00
Tuesday	7,500	6,967/7,500 = 0.93
Wednesday	6,400	6,967/6,400 = 1.09
Total	**20,900**	
Average	20,900/3 = 6,967	

(b) Calibration of Daily Variation Factors

10.3-12 Full Alternative Text

Station	Day	Time (PM)	Count (vehs)	8-Hour Expanded Count (vehs)	8-Hour Adjusted Counts (vehs)
1	Monday	12–4	2,213	2,213/0.429 = 5,159	× 1.00 = 5,159
2	Monday	4–8	3,000	3,000/0.571 = 5,254	× 1.00 = 5,254
3	Tuesday	12–4	2,672	2,672/0.427 = 6,258	× 0.93 = 5,820
4	Tuesday	4–8	2,500	2,500/0.573 = 4,363	× 0.93 = 4,058
5	Wednesday	12–4	3,500	3,500/0.438 = 7,991	× 1.09 = 8,710
6	Wednesday	4–8	3,750	3,750/0.562 = 6,673	× 1.09 = 7,274

(c) Expansion and Adjustment of Coverage Counts

10.3-13 Full Alternative Text

Note that in expanding the 4-hour coverage counts to 8 hours, the proportional split of volume varied from day to day. The expansions used the proportion appropriate to the day of the count. As the variation was not great, it would have been equally justifiable to use the average hourly split for all three days.

Again, the results obtained represent the particular three-day period over which the counts were conducted. Volume variations involving other days of the week or seasonal factors are not considered.

The three approaches detailed in this section are illustrative. Expansion and adjustment of coverage counts based upon control observations can be organized in many different ways, covering any network size and study period. The selection of control locations involves much judgment, and the success of any particular study depends upon the quality of the judgment exercised in designing the study. The traffic engineer must design each study to achieve the particular information goals at hand.

10.3.4 Estimating Vehicle-Miles Traveled on a Network

One output of most limited-network volume studies is an estimate of the total vehicle-miles traveled (VMT) on the network during the period of interest. The estimate is done roughly by assuming that a vehicle counted on a link travels the entire length of the link. This is a reasonable assumption, as some vehicles traveling only a portion of a link will be counted while others will not, depending upon whether they cross the count location.

Sample Problem 10-4: Estimating Vehicle-Miles

Using the sample network of Sample Problems 10-1, 10-2, and 10-3, the 8-hour volume results of Table 10.5 and assuming all links are 0.25 miles long, Table 10.6 illustrates the estimation of VMT. In this case, the estimate is the average 8-hour VMT for the three days of the study. It cannot be expanded into an estimate of *annual* VMT without knowing more about daily and seasonal variation patterns throughout the year.

Table 10.6: Estimation of Vehicle-Miles Traveled on a Limited Network: An Example

Station	8-Hour Count (vehs)	Link Length (mi)	Link VMT (veh-miles)
A	6,967	0.25	1,741.75
1	5,159	0.25	1,289.75
2	5,254	0.25	1,313.50
3	5,820	0.25	1,455.00
4	4,058	0.25	1,014.50
5	8,710	0.25	2,177.50
6	7,274	0.25	1,818.50
Network Total			**10,810.50**

Table 10.6: Full Alternative Text

10.3.5 Display of Network Volume Results

As was the case with intersection volume studies, more detailed results of a limited network study are presented in tabular form, some of which have been illustrated herein. For peak hours or for daily total volumes, it is often convenient to provide a network flow map. This is similar to an intersection flow diagram in that the thickness of flow lines is proportional to the volume. An example of such a map is shown in Figure 10.12.

Figure 10.12: A Typical Network Flow Map

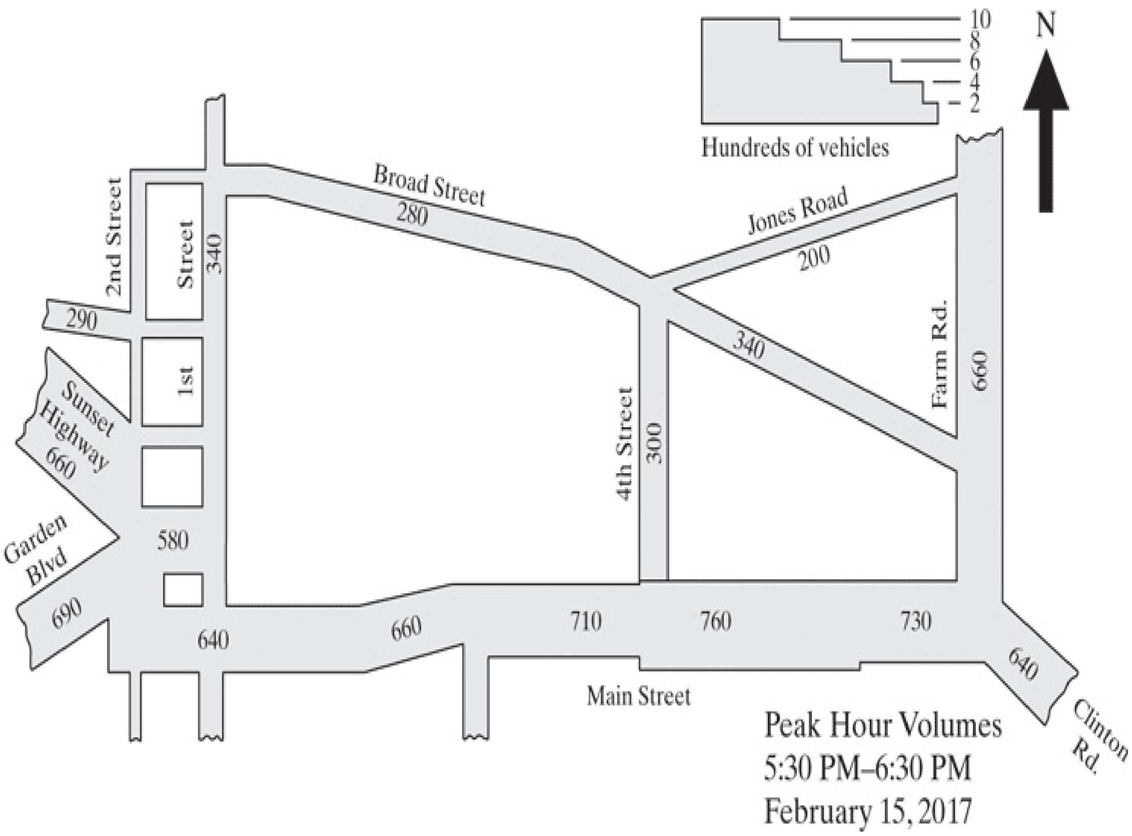


Figure 10.12 Full Alternative Text

10.3.6 Modern Alternatives

As discussed in Chapter 9, many major urban areas have substantial numbers of permanent sensors located throughout the street and highway system. In some cases, there will be sufficient numbers of detectors in a local network of interest to avoid a detailed manual study. The manipulation of data is basically the same as for manual studies, except that the fundamental collection task is automated.

Where there are some, but not sufficient, detectors in place, automated data can be used to supplement manually collected data and, in some cases, to check it through comparisons of automated and hand counts at a given common location.

10.4 Statewide Counting Programs

States generally have a special interest in observing trends in AADT, shifts within the ADT pattern, and VMT. These trends are used in statewide planning and for the programming of specific highway improvement projects. In recent years, there has been growing interest in person-miles traveled (PMT) and in statistics for other modes of transportation. Similar programs at the local and/or regional level are desirable for nonstate highway systems, although the cost is often prohibitive.

Following some general guidelines, as in Reference [3] for example, the state road system is divided into functional classifications. Within each classification, a pattern of control-count locations and coverage-count locations is established so that trends can be observed. Statewide programs are similar to limited network studies, except that the network involved is the entire state highway system and the time frame of the study is continuous (i.e., 365 days a year, every year).

Some general principles for statewide programs are as follows:

1. The objective of most statewide programs is to conduct a coverage count every year on every 2-mile segment of the state highway system, with the exception of low-volume roadways ($AADT < 100$ veh/day) . Low-volume roadways usually comprise about 50% of state system mileage and are classified as tertiary local roads.

2. The objective of coverage counts is to produce an annual estimate of AADT for each coverage location.

3. One control-count location is generally established for every 20 to 50 coverage-count locations, depending upon the characteristics of the region served. Criteria for establishing control locations are similar to those used for limited networks.

4. Control-count locations can be either *permanent counts* or *major or minor control counts*, which use representative samples. In both cases, control-count locations must monitor and calibrate daily

variation patterns and monthly or seasonal variation patterns for the full 365-day year.

5. All coverage counts are for a minimum period of 24 to 48 hours, eliminating the need to calibrate hourly variation patterns.

At permanent count locations, fixed detection equipment with data communications technology is used to provide a continuous flow of volume information. Major and minor control counts are generally made using portable counters and road tubes. Major control counts are generally made for one week during each month of the year. Minor control counts are generally made for one five-day (weekdays only) period in each season.

10.4.1 Sample Problems in Calibration of Daily and Monthly Variation Factors

Sample Problem 10-5: Calibrating Daily Variation Factors

The illustrative data in Table 10.7 are obtained from a permanent count location. At a permanent count location, data exist for all 52 weeks of the year (for 52 Sundays, 52 Mondays, 52 Tuesdays, etc.). (Note that in a 365-day year, one day will occur 53 times.)

Table 10.7: Calibration of Daily Variation Factors

Day	Yearly Average Volume for Day (vehs/day)	Daily Adjustment Factor DF
Monday	1820	1430/1820 = 0.79
Tuesday	1588	1430/1588 = 0.90
Wednesday	1406	1430/1406 = 1.02
Thursday	1300	1430/1300 = 1.10
Friday	1289	1430/1289 = 1.11
Saturday	1275	1430/1275 = 1.12
Sunday	1332	1430/1332 = 1.07
Total	**10,010**	
Estimated AADT	**1,430**	

Table 10.7: Full Alternative Text

Daily variation factors are calibrated based upon the average volumes observed during each day of the week. The base value for factor calibration is the average of the seven daily averages, which is a rough estimate of the AADT (but not exact, due to the 53rd piece of data for one day of the week). The factors can be plotted, as illustrated in Figure 10.13, and display a clear variation pattern that can be applied to coverage- count results.

Figure 10.13: Plot of Daily Variation Factors

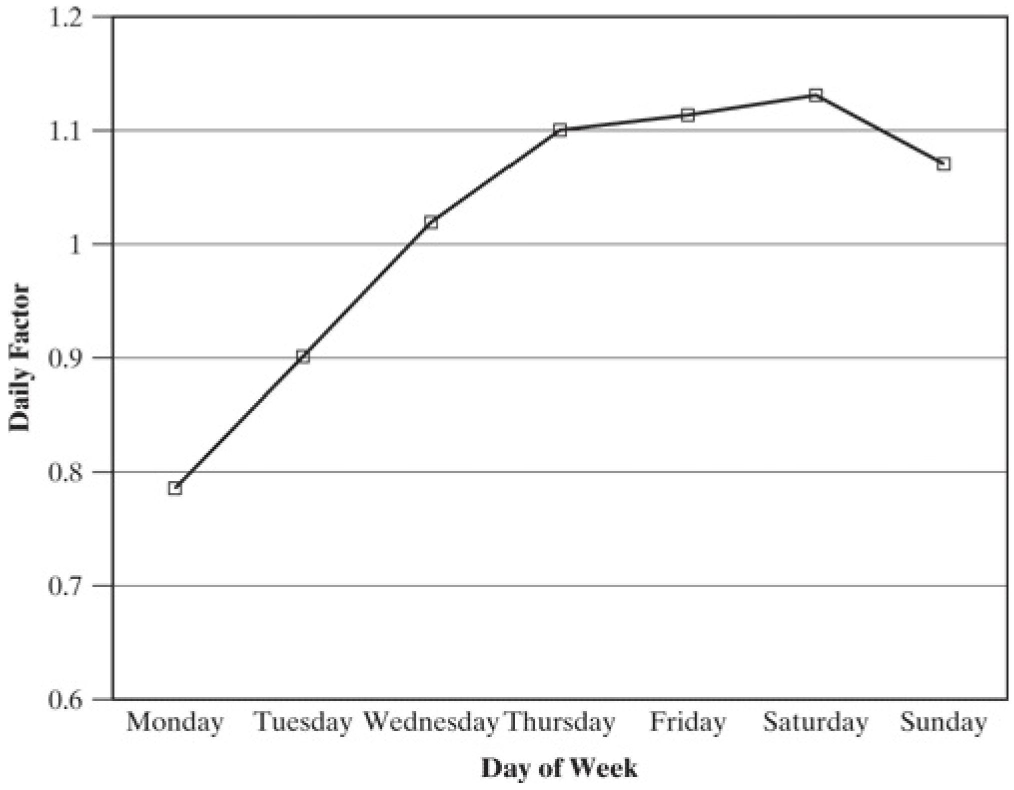


Figure 10.13: Full Alternative Text

Note that the sum of the seven daily adjustment factors *does not* add up to 7.00 (the actual total is 7.11). This is because of the way in which the factors are defined and computed. The daily averages are in the denominator of the calibration factors. In effect, the average factor is inverse to the average daily volume, so that the totals would not be expected to add to 7.00.

Daily adjustment factors can also be computed from the results of major and/or minor control counts. In a major control count, there would be 12 weeks of data, one week from each month of the year. The daily averages, rather than representing 52 weeks of data, reflect 12 representative weeks of data. The calibration computations, however, are exactly the same.

Sample Problem 10-6: Calibrating Monthly Variation Factors

Table 10.8 illustrates the calibration of monthly variation factors from permanent count data. The monthly factors are based upon monthly ADTs that have been observed at the permanent count location. Note that the sum of the 12 monthly variation patterns is not 12.00 (the actual sum is 12.29), as the monthly ADTs are in the denominator of the calibration.

Table 10.8: Calibration of Monthly Variation Factors

Month	Total Traffic (vehs)	ADT for Month (veh/day)	Monthly Factor (AADT/ADT)
January	19,840	/31 = 640	797/640 = 1.25
February	16,660	/28 = 595	797/595 = 1.34
March	21,235	/31 = 685	797/685 = 1.16
April	24,300	/30 = 810	797/810 = 0.98
May	25,885	/31 = 835	797/835 = 0.95
June	26,280	/30 = 876	797/876 = 0.91
July	27,652	/31 = 892	797/892 = 0.89
August	30,008	/31 = 968	797/968 = 0.82
September	28,620	/30 = 954	797/954 = 0.84
October	26,350	/31 = 850	797/850 = 0.94
November	22,290	/30 = 763	797/763 = 1.07
December	21,731	/31 = 701	797/701 = 1.14
Total	**290,851**	**AADT = 290,851/365 = 797 veh/day**	

Table 10.8: Full Alternative Text

Table 10.8 is based on permanent count data, such that the monthly ADTs are directly measured. One 7-day count in each month of the year would produce similar values, except that the ADT for each month would be estimated based on a single week of data, not the entire month.

Sample Problem 10-7: Estimating Monthly Variation Factors Based Upon 12 Weeks of Data

Consider monthly variation patterns based upon 12 weeks of data, i.e. one week in each month of the year is counted. The week, however, may or may not be representative of an average for the month.

In effect, an ADT for a given month is most likely to be observed in the middle of the month (i.e., the 14th to the 16th of any month). This statement is based upon the assumption that the volume trend within each month is unidirectional (i.e., volume grows throughout the month or declines throughout the month). Where a peak or low point exists within the month, this statement is not true.

Figure 10.14 illustrates a plot of 12 calibrated monthly variation factors, resulting from one-week counts in each month. The daily variation factors are plotted against the midpoint of the week in which the data for the month were taken.

Figure 10.14: Monthly Factor Calibrated from 12 Weeks of Data

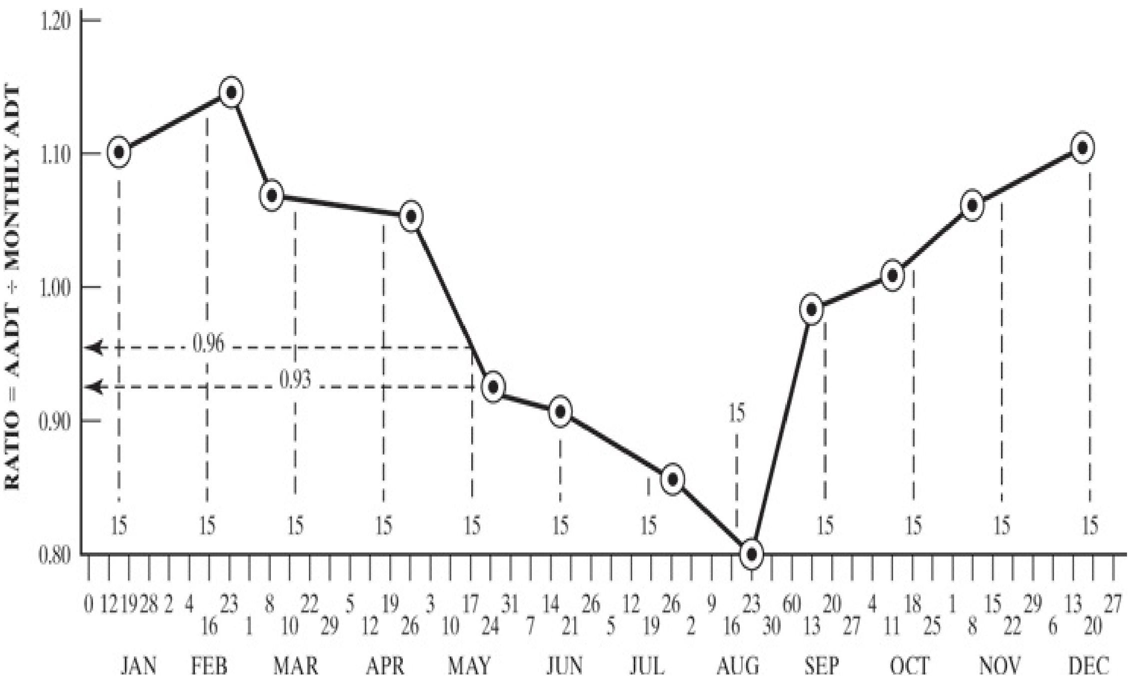


Figure 10.14: Full Alternative Text

This graph may now be entered at the middle of each month (the 15th), and adjusted factors read from the vertical axis. For example, in May the computed factor was 0.93, while the plot indicates that a factor computed for the middle of that month would have resulted in a factor of 0.96. Adjusting the factors in this manner results in a more representative computation based on monthly midpoints.

10.4.2 Grouping Data from Control-Count Locations

On state highway networks and systems, particularly in rural areas, it is possible for a broad region to have similar, if not the same, daily and/or monthly adjustment factors. In such regions, spatially contiguous control stations on the same classification of highway may be combined to form a single control group. The average factors for the group may then be applied over a wide area with similar variation patterns. In general, a statistical standard is applied to such groupings: Contiguous control counts on similar highway types may be grouped if the factors at the individual locations do not differ by more than [&|pminusns|0.10&] ±0.10 from the average for the group.

Sample Problem 10-8: Grouping Control Counts in Statewide Programs

Consider the example shown in Table 10.9. The daily variation factors for four consecutive control counts on a state highway have been calibrated as shown. It has been hypothesized that the four represent regions with similar daily variation patterns. Average factors have, therefore, been computed for the four grouped stations.

Table 10.9: A Trial Grouping

of Four Contiguous Control Stations Daily Variation Factors

Day	Daily Factor (DF) for Station Number: 1	2	3	4	Average Daily Factor DF
Monday	1.05	1.00	1.06	0.92	1.01
Tuesday	1.10	1.02	1.06	**0.89**	1.02
Wednesday	1.10	1.05	1.11	0.97	1.06
Thursday	1.06	1.06	1.03	1.00	1.04
Friday	1.01	1.03	1.00	0.91	0.98
Saturday	**0.85**	0.94	0.90	**1.21**	0.98
Sunday	0.83	0.90	0.84	1.10	0.92

Table 10.9: Full Alternative Text

The bold-faced factors indicate cases that violate the statistical rule for grouping (i.e., differences between these factors and the average for the group are more than ±0.10). This suggests that the proposed grouping is not appropriate. One might be tempted to remove Stations 1 and 4 from the group and combine only Stations 2 and 3. The proper technique, however, is to remove one station from the group at a time, as the resulting average factors will change. In this case, a cursory observation indicates that Station 4 does not really display a daily variation pattern similar to the others. This station has its peak traffic (DF<1.00) occurring during the week, while the other stations have their peak traffic on weekends. Thus, Station 4 is deleted from the proposed grouping and new averages are computed, as illustrated in Table 10.10.

Table 10.10: A Second Trial

Grouping of Control Stations Daily Variation Factors

Day	Daily Factor (DF) for Station:			Average Daily Factor (DF)
	1	2	3	
Monday	1.05	1.00	1.06	1.04
Tuesday	1.10	1.02	1.06	1.06
Wednesday	1.10	1.05	1.11	1.09
Thursday	1.06	1.06	1.03	1.05
Friday	1.01	1.03	1.00	1.01
Saturday	0.85	0.94	0.90	0.90
Sunday	0.83	0.90	0.84	0.86

Table 10.10: Full Alternative Text

Now, all factors at individual stations are within ±0.10 of the average for the group. This would be an appropriate grouping of control stations.

10.4.3 Using the Results

It should be noted that groups for daily factors and groups for monthly factors do not have to be the same. It is convenient if they are, however, and it is not at all unlikely that a set of stations grouped for one type of factor would also be appropriate for the other.

The state highway agency will use its counting program to generate basic trend data throughout the state. It will also generate, for contiguous portions of each state highway classification, a set of daily and monthly variation factors that can be applied to any coverage count within the influence area of the subject control grouping. An example of the type of data that would be made available is shown in Table 10.11.

Table 10.11: Typical Daily and Monthly Variation Factors for a Contiguous Area on a State Highway System

Daily Factors (DF)		Monthly Factors (MF)			
Day	Factor	Month	Factor	Month	Factor
Monday	1.072	January	1.215	July	0.913
Tuesday	1.121	February	1.191	August	0.882
Wednesday	1.108	March	1.100	September	0.884
Thursday	1.098	April	0.992	October	0.931
Friday	1.015	May	0.949	November	1.026
Saturday	0.899	June	0.918	December	1.114
Sunday	0.789				

Table 10.11: Full Alternative Text

Using these tables, any coverage count for a period of 24 hours or more can be converted to an estimate of the AADT using the following relationship:

$AADT=V24ij\times DFi\times MFj$ [10-4]

where:

AADT = average annual daily traffic, veh/day, V_{24ij} = 24-hr volume for day i in month j, veh, DF_i = daily adjustment factor for day i,and MF_j = monthly adjustment factor for month j.

Sample Problem 10-9: Predicting AADT from Statewide Adjustment Tables

Consider a coverage count taken at a location within the area represented by the factors of Table 10.11. A count of 1,000 vehicles was observed on a Tuesday in July. From Table 10.11, the daily factor for Tuesdays is 1.121, and the monthly factor for July is 0.913. Then:

$$AADT=1{,}000\times1.121\times0.913=1{,}023 \text{ vehs/day}$$

10.4.4 Estimating Annual Vehicle-Miles Traveled

Given estimates of AADT for every 2-mile segment of each category of roadway in the state system (excluding low- volume roads), estimates of annual VMT can be assembled. For each segment, the annual VMT is estimated as:

$$VMT_{365}=AADT\times L\times365 \quad [10\text{-}5]$$

where:

VMT_{365} = annual vehicle-miles traveled over the segment, AADT = average annual daily traffic, veh/day, and L = length of the segment, mi.

For any given roadway classification or system, the segment VMTs can be summed to give a regional or statewide total. The question of the precision or accuracy of such estimates is interesting, given that none of the low-volume roads are included and that a real statewide total would need to include inputs for all nonstate systems in the state. Regular counting programs at the local level are, in general, far less rigorous than state programs.

There are two other ways that are commonly used to estimate VMT:

- Use the number of registered vehicles with reported annual mileages, adjusting for out-of-state travel.
- Use fuel tax receipts by category of fuel (which relates to categories of vehicles), and estimate VMT using average fuel consumption ratings for different types of vehicles.

There is interest in improving statewide VMT estimating procedures, and a number of significant research efforts have been sponsored on this topic in recent years. There is also growing interest in nationwide PMT estimates, with appropriate modal categories.

10.5 Specialized Counting Studies

There are a number of instances in which simple counting of vehicles at a point, or at a series of points, is not sufficient to provide the information needed. Three principal examples of specialized counting techniques are (1) origin and destination counts, (2) cordon counts, and (3) screen-line counts.

10.5.1 Origin and Destination Counts

There are many instances in which normal point counts of vehicles must be supplemented with knowledge of the origins and destinations of the vehicles counted. In major regional planning applications, origin and destination studies involve massive home-interview efforts to establish regional travel patterns. In traffic applications, the scope of origin and destination counts are often more limited. Common applications include the following:

- Weaving-area studies
- Freeway studies
- Major activity center studies

Proper analysis of weaving-area operations requires that volume be broken down into two weaving and two nonweaving flows that are present. A total count is insufficient to evaluate performance. In freeway corridors, it is often important to know where vehicles enter and exit the freeway. Alternative routes, for example, cannot be accurately assessed without knowing the underlying pattern of origins and destinations. At major activity centers (sports facilities, airports, regional shopping centers, etc.), traffic planning of access and egress also requires knowledge of where vehicles are coming from when entering the development or going to when leaving the development.

Many ITS technologies hold great promise for providing detailed information on origins and destinations. Automated toll-collection systems can provide data on where vehicles enter and leave toll facilities. Automated license-plate reading technology is used in traffic enforcement and could be used to track vehicle paths through a traffic system. Although these technologies continue to advance rapidly, their use in traditional traffic data collection has been much slower due to the privacy issues that such use raises.

Historically, one of the first origin-destination count techniques was called a *lights-on study*. This method was often applied in weaving areas, where vehicles arriving on one leg could be asked to turn on their lights. With the advent of daytime running lights, this methodology is no longer viable.

Conventional traffic origin and destination counts rely primarily on one of three approaches:

- License-plate studies
- Postcard studies
- Interview studies

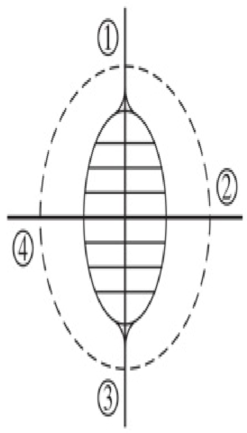
In a license-plate study, observers (or automated equipment) record the license plate numbers as they pass designated locations. This is a common method used to track freeway entries and exits at ramps. Postcard studies involve handing out color- or otherwise coded cards as vehicles enter the system under study and collecting them as vehicles leave. In both license-plate and postcard studies, the objective is to match up vehicles at their origin and at their destination. Interview studies involve stopping vehicles (with the approval and assistance of police), and asking a short series of questions concerning their trip, where it began, where it is going, and what route will be followed.

Major activity centers are more easily approached, as one end of the trip is known (everyone is at the activity center). Here, interviews are easier to conduct, and license-plate numbers of parked vehicles can be matched to home locations using data from the state Department of Motor Vehicles.

When attempting to match license-plate observations or postcards, sampling becomes a significant issue. If a sample of drivers is recorded at each entry and exit location, then the probability of finding matches is diminished considerably. If 50% of the entering vehicles at Exit 2 are observed, and 40% of the exiting vehicles at Exit 5 are observed, then the number of matches of vehicles traveling from Exit 2 to Exit 5 would be $0.50\times0.40=0.20$ or 20% . When such sampling techniques are used, separate counts of vehicles at all entry and exit points must be maintained to provide a means of expanding the sample data.

Consider the situation illustrated in Figure 10.15. It shows a small local downtown street network with four entry roadways and four exit roadways. Thus, there are $4\times4=16$ possible origin-destination pairs for vehicles accessing or traveling through the area. The data shown reflect both the observed origins and destinations (using license-plate samples) and the full-volume counts observed on each entry and exit leg.

Figure 10.15: Data from an Origin-Destination Count Using License Plate Matching



Destination Station	Origin Station				Row Sum T_j	Vol V_j
	1	2	3	4		
1	50	8	20	17	**95**	250
2	10	65	21	10	**106**	310
3	15	12	38	15	**80**	200
4	13	14	18	42	**87**	375
Col Sum T_i	**88**	**99**	**97**	**84**		
Volume V_i	210	200	325	400		**1,135**

Figure 10.15: Full Alternative Text

If the columns and rows are totaled, the sums should be equal to the observed total volumes, assuming that a 100% sample of license plates was obtained at each location. This is obviously not the case. Thus, the origin-destination volumes must be expanded to reflect the total number of vehicles counted. This can be done in either of two ways: (1) origin-destination cells can be expanded so that the row totals are correct (i.e., match the measured volume) or (2) origin- destination cells can be expanded so that the column totals are correct. Unfortunately, these two approaches will lead to two different sets of origin-destination volumes.

In practice, the average of the two approaches is adopted. This creates an iterative process, as the initial adjustment will still result in column and row totals that are not the same as the measured volumes. Iteration is continued until all row and column totals are within ±10% of the measured volumes.

The cell volumes, representing matched trips from Station *i* to Station *j*, are adjusted using factors based upon column closure and row closure:

TijN=Tij(N−1)(Fi+Fj2) [10-6]

where:

Fi=adjustment factor for origin i=Vi/Ti,Fj=adjustment factor for destination
−1)=number of trips from Station i to Station j after the (N
−1) iteration of the data (trips),Ti=sum of matched trips from Station i (trip

Sample Problem 10-10: Estimates from Sample Origin and Destination Data

Figure 10.15 shows sample data collected for an origin and destination study. As 100% samples were not collected at each origin and destination, an estimated O-D table must be constructed.

This will involve iteration, with the data of Figure 10.15 serving as the 0th iteration. Each adjustment cycle results in new values of Tij, Ti, Tj, Fi,and F j . The observed total volumes, of course, remain constant.

Table 10.12 shows the results of several iterations, with the final O-D counts accepted when all adjustment factors are greater than or equal to

0.90 or less than or equal to 1.10. In this case, the initial expansion of O-D counts was iterated twice to obtain the desired accuracy.

Table 10.12: Sample Expansion of Origin and Destination Data

Destination Station	Origin Station				T_j	V_j	F_j
	1	2	3	4			
1	50	8	20	17	**95**	250	2.63
2	10	65	21	10	**106**	310	2.92
3	15	12	38	15	**80**	200	2.50
4	13	14	18	42	**87**	375	4.31
T_i	**88**	**99**	**97**	**84**	**368**		
V_i	210	200	325	400		**1135**	
F_i	2.39	2.02	3.35	4.76			

(a) Field Data and Factors for Iteration 0

10.5-21 Full Alternative Text

Destination Station	Origin Station				T_j	V_j	F_j
	1	2	3	4			
1	125	19	60	63	**267**	250	0.94
2	27	161	66	38	**292**	310	1.06
3	37	27	111	54	**229**	200	0.87
4	44	44	69	191	**347**	375	1.08
T_i	**232**	**251**	**306**	**346**	**1135**		
V_i	210	200	325	400		**1135**	
F_i	0.90	0.80	1.06	1.16			

(b) Initial Expansion of O-D Matrix (Iteration 0)

10.5-22 Full Alternative Text

Destination Station	Origin Station 1	2	3	4	T_j	V_j	F_j
1	116	16	60	66	**257**	250	0.97
2	26	150	70	43	**288**	310	1.08
3	33	23	108	55	**218**	200	0.92
4	43	42	74	213	**372**	375	1.01
T_i	**217**	**230**	**311**	**376**	**1135**		
V_i	210	200	325	400		**1135**	
F_i	0.97	0.87	1.04	1.06			

(c) First Iteration of O-D Matrix

10.5-23 Full Alternative Text

Destination Station	Origin Station 1	2	3	4	T_j	V_j	F_j
1	112	15	60	67	**254**	250	0.98
2	27	145	74	46	**292**	310	1.06
3	31	20	105	55	**211**	200	0.95
4	43	39	76	221	**378**	375	0.99
T_i	**212**	**220**	**316**	**388**	**1135**		
V_i	210	200	325	400		**1135**	
F_i	0.99	0.91	1.03	1.03			

(d) Second Iteration of O-D Matrix

10.5-24 Full Alternative Text

10.5.2 Cordon Counts

A cordon is an imaginary boundary around a study area of interest. It is generally established to define a CBD or other major activity center where the accumulation of vehicles within the area is of great importance in traffic planning. Cordon volume studies require counting all streets and highways that cross the cordon, classifying the counts by direction and by 15- to 60-minute time intervals. In establishing the cordon, several principles should be followed:

- The cordoned area must be large enough to define the full area of interest, yet small enough so that accumulation estimates will be useful for parking and other traffic planning purposes.
- The cordon is established to cross all streets and highways at *midblock* locations, to avoid the complexity of establishing whether turning vehicles are entering or leaving the cordoned area.
- The cordon should be established to minimize the number of crossing points wherever possible. Natural or man-made barriers (e.g., rivers, railroads, limited access highways, and similar features) can be used as part of the cordon.
- Cordoned areas should have relatively uniform land use. Accumulation estimates are used to estimate street capacity and parking needs. Large cordons encompassing different land-use activities will not be focused enough for these purposes.

The accumulation of vehicles within a cordoned area is found by summarizing the total of all counts entering and leaving the area by time period. The cordon counts should begin at a time when the streets are virtually empty. As this condition is difficult to achieve, the study should start with an estimate of vehicles already within the cordon. This can be done by circulating through the area and counting parked and circulating vehicles encountered. Off-street parking facilities can be surveyed to estimate their overnight population.

Note that an estimate of parking and standing vehicles may *not* reflect true parking demand if supply is inadequate and many circulating vehicles are merely looking for a place to park. Also, demand discouraged from entering the cordoned area due to congestion is not evaluated by this study technique.

When all entry and exit counts are summed, the accumulation of vehicles within the cordoned area during any given period may be estimated as:

$$A_i = A_{i-1} + VE_i - VL_i \quad [10\text{-}7]$$

where:

A_i = accumulation for time period i, veh, A_{i-1} =

accumulation for time period i−1,veh, V Ei = total volume entering the cordoned area during time period i, veh, and V Li = total volume leaving the cordoned area during time period i, veh.

An example of a cordon volume study and the estimation of accumulation within the cordoned area is shown in Table 10.13. Figure 10.16 illustrates a typical presentation of accumulation data, whereas Figure 10.17 illustrates an interesting presentation of cordon-crossing information.

Table 10.13: Accumulation Computations for an Illustrative Cordon Study

Time	Vehicles Entering (vehs)	Vehicles Leaving (vehs)	Accumulation (vehs)
4:00–5:00 AM	–	–	250*
5:00–6:00 AM	100	20	250 + 100 − 20 = 330
6:00–7:00 AM	150	40	330 + 150 − 40 = 440
7:00–8:00 AM	200	40	440 + 200 − 40 = 600
8:00–9:00 AM	290	80	600 + 290 − 80 = 810
9:00–10:00 AM	350	120	810 + 350 − 120 = 1,040
10:00–11:00 AM	340	200	1,040 + 340 − 200 = 1,180
11:00–12:00 N	350	350	1,180 + 350 − 350 = 1,180
12:00–1:00 PM	260	300	1,180 + 260 − 300 = 1,140
1:00–2:00 PM	200	380	1,140 + 200 − 380 = 960
2:00–3:00 PM	180	420	960 + 180 − 420 = 720
3:00–4:00 PM	100	350	720 + 100 − 350 = 470
4:00–5:00 PM	120	320	120 − 320 = 270

*Estimated beginning accumulation.

Table 10.13: Full Alternative Text

Figure 10.16: Typical Presentation of Accumulation

Data Time

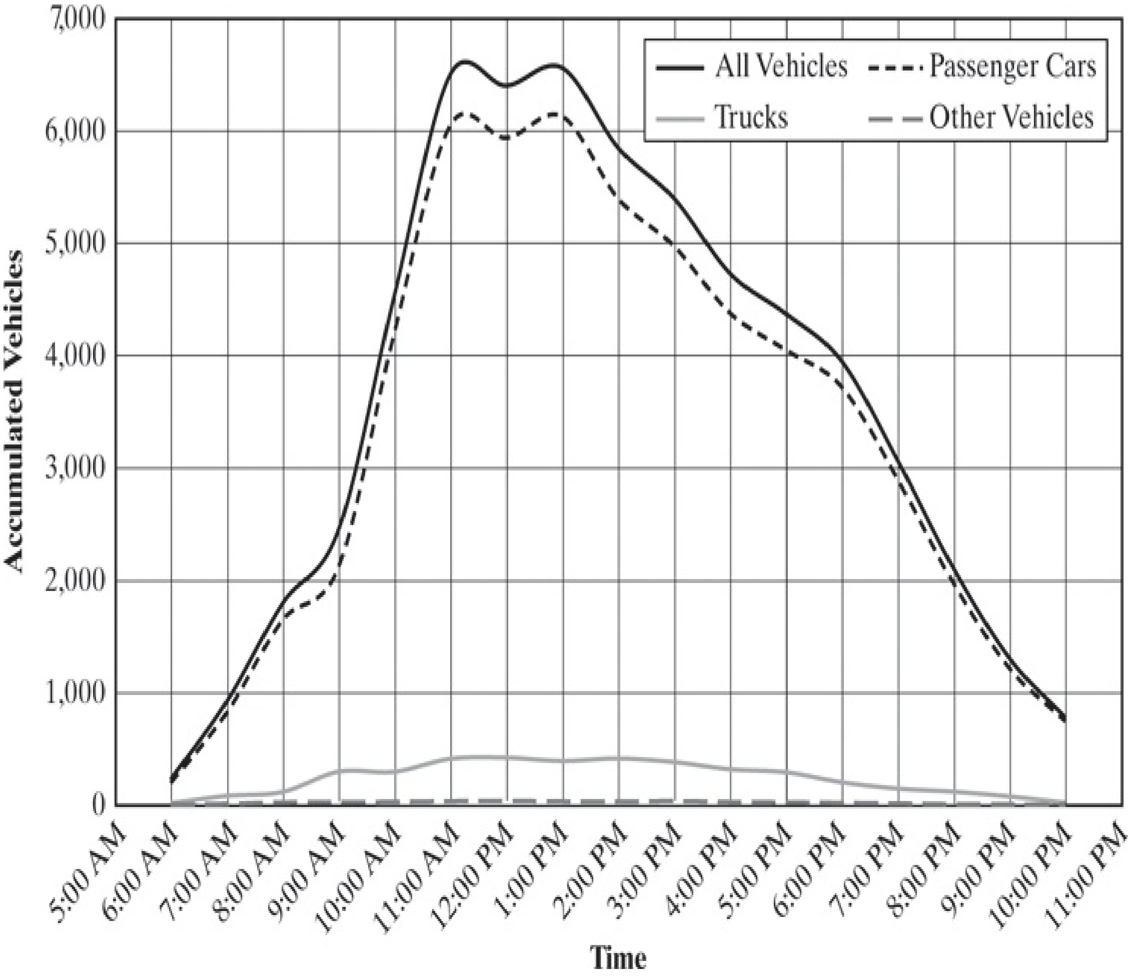


Figure 10.16: Full Alternative Text

Figure 10.17: Typical Presentation of Daily Cordon Crossings

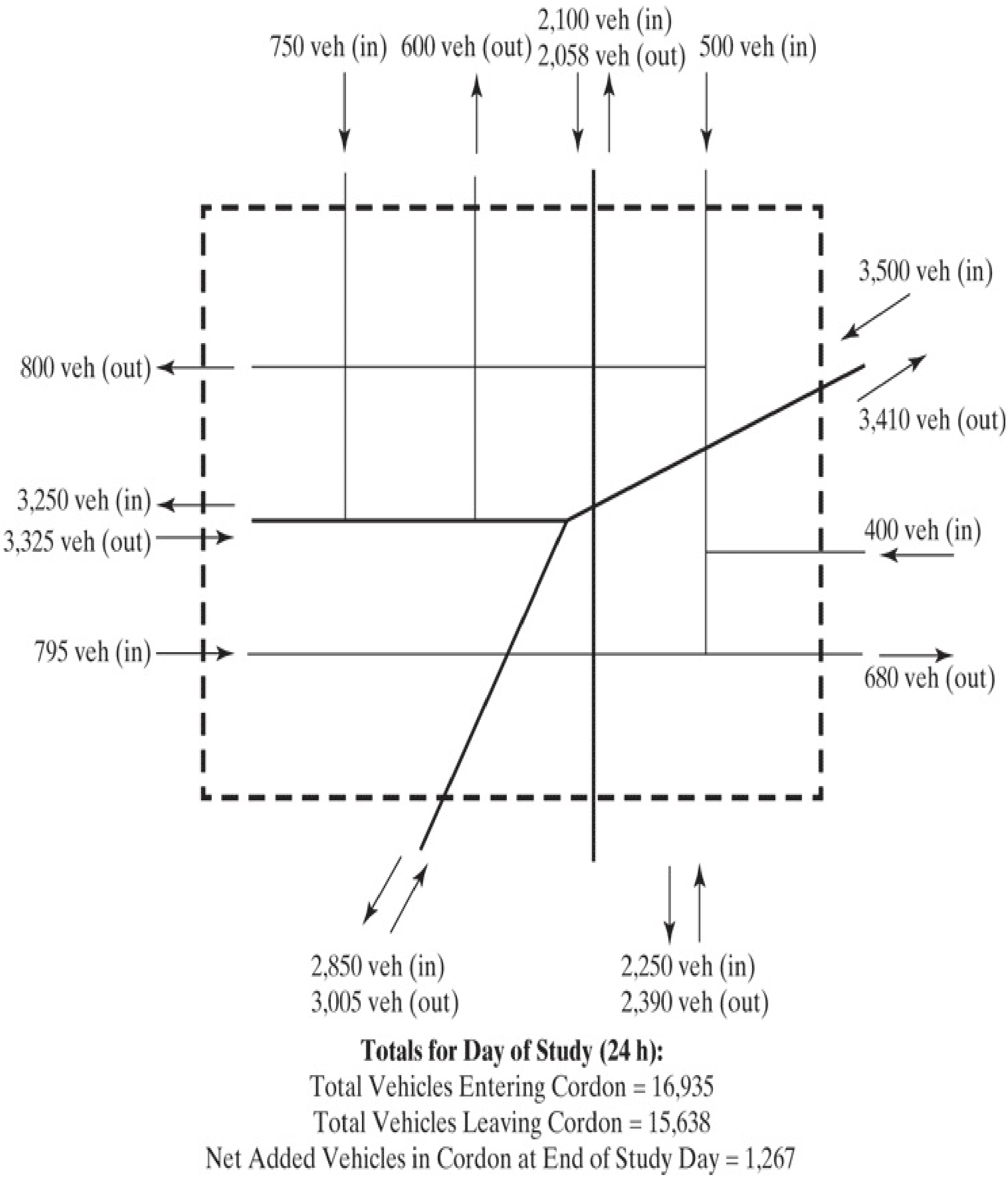


Figure 10.17: Full Alternative Text

10.5.3 Screen-Line Counts

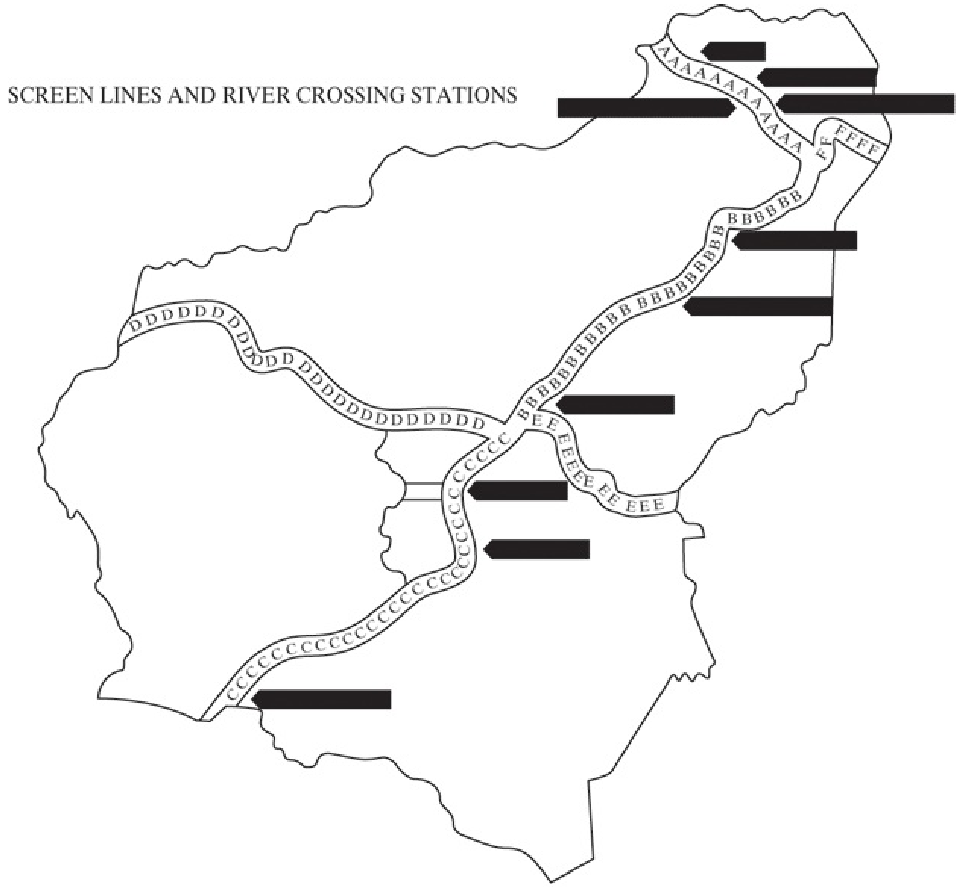
Screen-line counts and volume studies are generally conducted as part of a larger regional origin-destination study involving home interviews as the principal methodology. In such regional planning studies, home interview responses constitute a small but detailed sample that is used to estimate the number of trips per day (or some other specified time interval) between defined transportation zones that have been established within the study region.

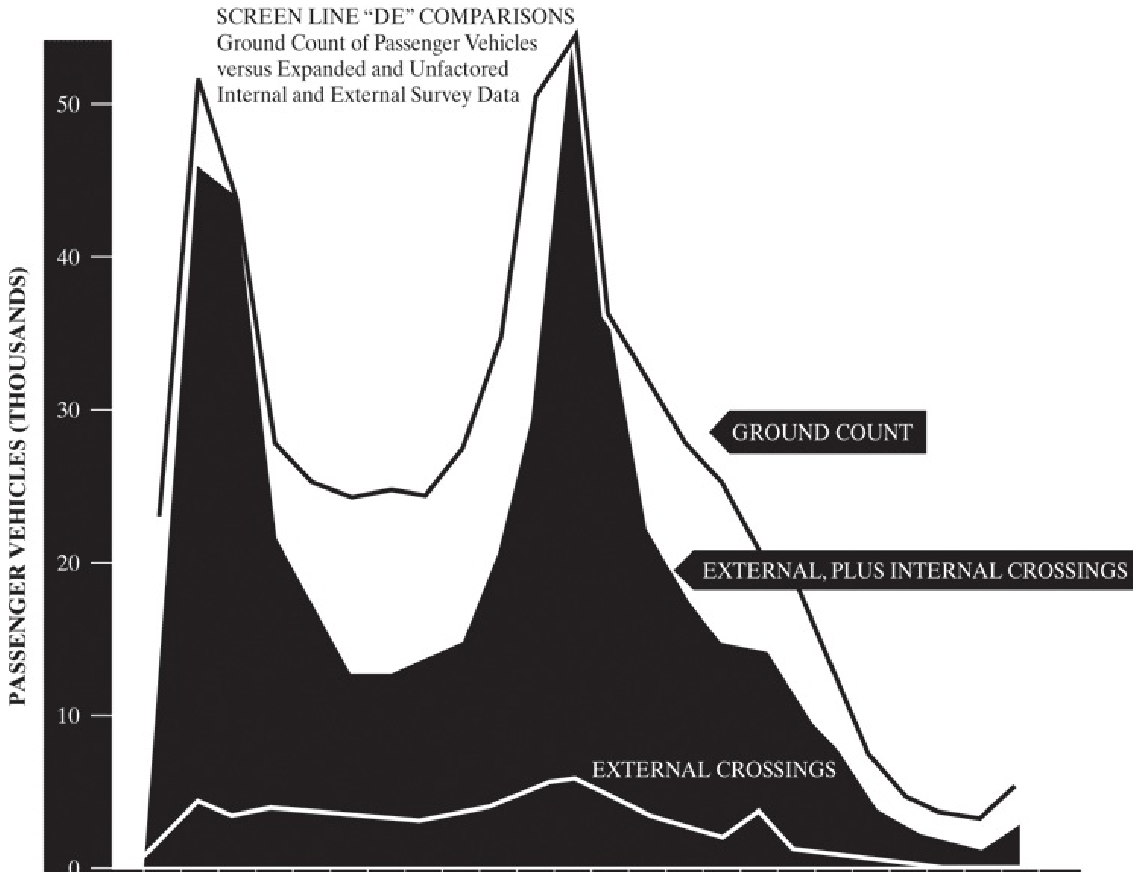
Because home interview samples are small and because additional data are used to estimate trip patterns for those passing through the study area or having only a single trip-end within the study area, it is necessary to use some form of field observations to check on the accuracy of predicted movements.

Screen lines are convenient barriers cutting through the study area with only a limited number of crossing points. Rivers, railroads, limited-access highways, and other features make good screen lines. The zone-to-zone trip estimates of a regional study can be summed in a way that yields the predicted number of trips across the screen line in a defined time period. A screen-line count can then be made to observe the actual number of crossings. The comparison of predicted versus observed crossings provides a means by which predicted zone-to-zone trips can be adjusted.

Figure 10.18 illustrates a study area for which two screen lines have been established. Predicted versus observed crossings are presented in graphic form. The ratio of observed to predicted crossings provides an adjustment factor that can be applied to all zonal trip combinations.

Figure 10.18: Illustration of a Screen Line Study


SCREEN LINES AND RIVER CROSSING STATIONS


SCREEN LINE "DE" COMPARISONS
Ground Count of Passenger Vehicles
versus Expanded and Unfactored
Internal and External Survey Data
PASSENGER VEHICLES (THOUSANDS)
50
40
30
20
10
0
GROUND COUNT
EXTERNAL, PLUS INTERNAL CROSSINGS
EXTERNAL CROSSINGS
06 AM
08
010
012 N
14
16
18
20
22
24
02
04
06
HOUR

(Source: Used with permission from Institute of Transportation Engineers Box, P, and Oppenlander, J, *Manual of Traffic Engineering Studies*, Institute of Transportation Engineers, Washington D.C., 1975, Figure 3-35, pg 43.)

Figure 10.18: Full Alternative Text

10.6 Closing Comments

The concept is simple: counting vehicles. As reviewed in this chapter, the process is not always simple, nor is the proper use of field results to obtain the desired statistics always straightforward. The field work of volume studies is relatively pedestrian but crucially important. Volume data is one of the primary bases for all traffic engineering analysis, planning, design, and operation.

Volume data must be accurately collected. They must be reduced to understandable forms, and properly analyzed to obtain the prescribed objective of the study. Data must then be presented clearly and unambiguously for use by traffic engineers and others involved in the planning and engineering process. No geometric or traffic control design can be effective if it is based on incorrect data related to traffic volumes and true demand. The importance, therefore, of performing volume studies properly cannot be understated.

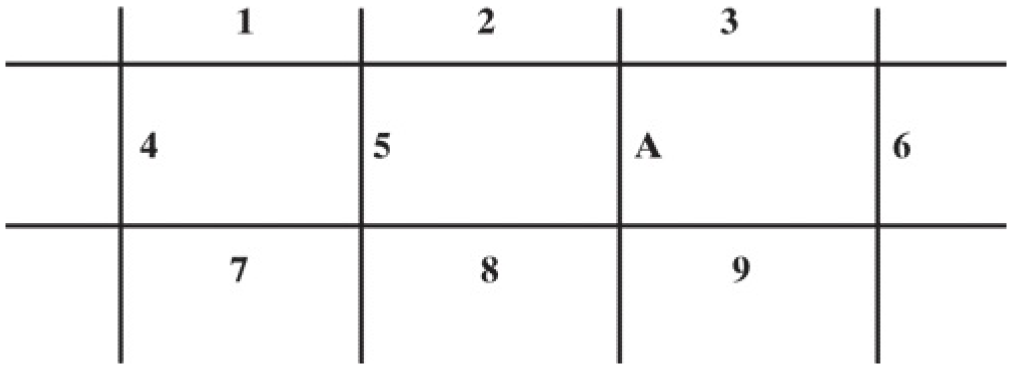
References

- 1. *Highway Capacity Manual, 6th Edition: A Guide for Multimodal Mobility Analysis*, Transportation Research Board, Washington, D.C., 2016.
- 2. McShane, W. and Crowley, K., “Regularity of Some Detector-Observed Arterial Traffic Volume Characteristics,” *Transportation Research Record 596*, Transportation Research Board, National Research Council, Washington, D.C., 1976.
- 3. *Traffic Monitoring Guide*, Federal Highway Administration, U.S. Department of Transportation, Washington D.C., 1985.

Problems

1. 10-1. A limited network counting study was conducted for the network shown below. Because only two sets of road tubes were available, the study was conducted over a period of several days, using Station A as a control location.

Network for Problem 10-1



Full Alternative Text

Using the data from the study, shown in the tables below, estimate the 12-hour volume (8:00 am to 8:00 pm) at each station for the average day of the study.

Axle Counts for Control Station A (Problem 10-1)

Day	Time Period		
	8:00–11:45	**12:00–3:45**	**4:00–7:45**
Monday	3,000	2,800	4,100
Tuesday	3,300	3,000	4,400
Wednesday	4,000	3,600	5,000

Full Alternative Text

Axle-Counts for Coverage Stations (Problem 10-1)

Station	Day	Time	Axle-Count
1	Monday	8:00–11:45	1,900
2	Monday	12:00–3:45	2,600
3	Monday	4:00–7:45	1,500
4	Tuesday	8:00–11:45	3,000
5	Tuesday	12:00–3:45	3,600
6	Tuesday	4:00–7:45	4,800
7	Wednesday	8:00–11:45	3,500
8	Wednesday	12:00–3:45	3,200
9	Wednesday	4:00–7:45	4,400

Full Alternative Text

Sample Vehicle Classification Count (Problem 10-1)

Vehicle Class	Vehicle Count
2-axle	1,100
3-axle	130
4-axle	40
5-axle	6

Full Alternative Text

2. 10-2. The following control counts were made at a state-maintained permanent count station. From the information given, calibrate the daily volume variation factors for this station:

Data for Problem 10-2

Day of Week	Average Annual Volume for Day
Sunday	3,500
Monday	4,400
Tuesday	4,200
Wednesday	4,300
Thursday	3,900
Friday	4,900
Saturday	3,100

Full Alternative Text

3. 10-3. What count period would you select for a volume study at an intersection with a signal cycle length of (a) 60 s, (b) 90 s, and (c) 120 s?

4. 10-4. The following control counts were made at an urban count station to develop daily and monthly variation factors. Calibrate these factors given the data shown below.

24-Hour Daily Volumes

First Week in Month of:	Day of Week						
	Mon	Tue	Wed	Thu	Fri	Sat	Sun
January	2000	2200	2250	2000	1800	1500	950
April	1900	2080	2110	1890	1750	1400	890
July	1700	1850	1900	1710	1580	1150	800
October	2100	2270	2300	2050	1800	1550	1010

Full Alternative Text

Standard Monthly Volumes

Third Week in Month of:	Average 24-Hour Count (vehs)
January	2250
February	2200
March	2000
April	2100
May	1950
June	1850
July	1800
August	1700
September	2000
October	2100
November	2150
December	2300

Full Alternative Text

5. 10-5. The four control stations shown have been regrouped for the purposes of calibrating daily variation factors. Is the grouping appropriate? If not, what would an appropriate grouping be? What are the combined daily variation factors for the appropriate group(s)? The stations are located sequentially along a state route.

Daily Variation Factors for Individual Stations

Station	Mon	Tue	Wed	Thu	Fri	Sat	Sun
1	1.04	1.00	0.96	1.08	1.17	0.90	0.80
2	1.12	1.07	0.97	1.06	1.02	0.87	0.82
3	0.97	0.99	0.89	1.01	0.86	1.01	1.06
4	1.01	1.00	1.01	1.09	1.10	0.85	0.85

Full Alternative Text

6. 10-6 Estimate the annual VMT for a section of the state highway system represented by the variation factors of Table 10.11. The following coverage counts are available for the locations within the section.

Coverage-Count Data

Station	Segment Length (Mi)	Coverage-Count Date	24-Hour Count (vehs)
1	3.0	Wed in March	9,120
2	2.7	Tue in September	10,255
3	2.5	Fri in August	16,060
4	4.6	Sun in May	21,858
5	1.8	Thu in December	9,508
6	1.6	Fri in January	11,344

Full Alternative Text

7. 10-7. The following origin and destination results were obtained from sample license plate observations at five locations. Expand and adjust the initial trip-table results to reflect the full population of vehicles during the study period.

Initial Origin and Destination Matches from Sample License- Plate Observations

Destination Station	Origin Station					Total Destination Count (vehs)
	1	2	3	4	5	
1	50	120	125	210	75	1,200
2	105	80	143	305	100	2,040
3	125	100	128	328	98	1,500
4	82	70	100	125	101	985
5	201	215	180	208	210	2,690
Total Origin Count (vehs)	**1,820**	**1,225**	**1,750**	**2,510**	**1,110**	**8,415**

Full Alternative Text

Chapter 11 Speed, Travel Time, and Delay Studies

11.1 Introduction

Speed, travel time, and delay are all related measures that are commonly used as indicators of performance for traffic facilities. All relate to a factor that is most directly experienced by motorists: How long does it take to get from A to B? Motorists have the obvious desire to complete their trip in the minimum time consistent with safety. The performance of a traffic facility is often described in terms of how well that objective is achieved.

In the *Highway Capacity Manual* [1], for example, average travel speed is used as a measure of effectiveness for arterials, for two-lane rural highways, and for more extensive facility evaluations. Control delay is the measure of effectiveness for signal and STOP-controlled intersections, and roundabouts. Whereas freeways use density as a primary measure of effectiveness, speed is an important component of the evaluation of freeway facility and system operation.

Thus, it is important that traffic engineers understand how to measure and interpret data on speed, travel time, and delay in ways that yield a basic understanding of the quality of operations on a facility, and in ways that directly relate to defined performance criteria. Speed is also an important factor in evaluating high-accident locations as well as in other safety-related investigations.

Speed is inversely related to travel time. The reasons and locations at which speeds or travel times would be measured are quite different. Speed measurements are most often taken at a point (or a short segment) of roadway under conditions of free flow. The intent is to determine the speeds that drivers select, unaffected by the existence of congestion. This information is used to determine general speed trends, to help determine reasonable speed limits, and to assess safety. Such studies are referred to as “spot speed studies,” because the focus is on a designated “spot” on a facility.

Travel time must be measured over a distance. While spot speeds can indeed be measured in terms of travel times over a short measured distance, generally less than 1,000 ft, most travel time measurements are made over a significant length of a facility. Such studies are generally done during times of congestion specifically to measure or quantify the extent and causes of congestion.

In general terms, delay is a portion of total travel time. It is a portion of travel time that is particularly identifiable and unusually annoying to the motorist. Delay along an arterial, for example, might include stopped time due to signals, midblock obstructions, or other causes of congestion.

At signal and STOP-controlled intersections, delay takes on more importance, as travel time is difficult to define for a point location. Unfortunately, delay at intersections, specifically signalized intersections, has many different definitions, and the traffic engineer must be careful to use measurements and criteria that relate to the same delay definition. Some of the most frequently used forms of intersection delay include the following:

- Stopped-time delay—the time a vehicle spends stopped waiting to proceed through a signal or STOP-controlled intersection.
- Approach delay—adds the delay due to deceleration to and acceleration from a stop to stopped time delay.
- Time-in-queue delay—the time between a vehicle joining the end of a queue at a signal or STOP-controlled intersection and the time it crosses the STOP line to proceed through the intersection.
- Control delay—the total delay at an intersection caused by a control device (either a signal or a STOP-sign), including both time-in-queue delay plus delays due to acceleration and deceleration.

Control delay was a term introduced in the 1985 *Highway Capacity Manual* [2], and it is used as the measure of effectiveness for signal, STOP-controlled intersections and roundabouts.

Along routes, another definition of *delay* may be applied: *travel-time delay* is the difference between the actual travel time traversing a section of highway and the driver's expected or desired travel time. It is more of a philosophic approach, as there are no clearly accurate methodologies for determining the expected travel time of a motorist over a given section of highway. For this reason, it is seldom used for assessing congestion along a highway segment.

Because speeds are generally studied at points under conditions of free flow and travel times and delays are generally studied along sections of roadway under congested conditions, the study techniques for each are quite different. Although sharing many similar elements, the analysis of data and the presentation of results also differ somewhat.

11.2 Spot Speed Studies

A spot speed study is conducted to determine the characteristics of speeds selected by drivers free of the limitations imposed by congestion. Thus, they are generally not conducted when volumes are in excess of 750–1,000 veh/h/ln on freeways or 500 veh/h/ln on other types of uninterrupted flow facilities.

Studies generally seek to measure an average and/or other representative speed, and to document the distribution of speeds that drivers select. Because individual drivers will choose to drive at different speeds, knowing the average speed, while quite useful, is insufficient to fully understand underlying characteristics.

There are, however, two different ways to define an *average* speed, and there are other statistics that could be useful relative to the spread of individual speeds around the mean. This will involve some statistical analysis of speed measurements, which are discussed and illustrated in this chapter.

11.2.1 Speed Definitions of Interest

When the speeds of individual vehicles are measured at a given spot or location, the result is a *distribution* of speeds, as no two vehicles will be traveling at exactly the same speed. The results of the study, therefore, must describe the observed distribution of speeds as clearly as possible. There are several key statistics that are used to describe spot speed distributions:

- Average or time mean speed: the average speed of all vehicles passing the study location during the period of the study, mi/h.
- Standard deviation: in simplistic terms, the standard deviation of speeds is the average difference between individual observed speeds and the average speed during the period of the study.
- 85th percentile speed: the speed below which 85% of the observed

vehicles travel, mi/h.

- Median speed: the speed that equally divides the distribution of spot speeds; 50% of observed speeds are higher than the median; 50% of observed speeds are lower than the median, mi/h.

- Pace: a 10 mi/h increment in speeds that encompasses the highest proportion of observed speeds (as compared with any other 10 mi/h increment).

The desired result of a spot speed study is to determine each of these measures and to determine an adequate mathematical description of the entire observed distribution.

11.2.2 Uses of Spot Speed Data

The results of spot speed studies are used for many different purposes by traffic engineers, including the following:

- Establishing the effectiveness of new or existing speed limits or enforcement practices

- Determining appropriate speed limits for application

- Establishing speed trends at the local, state, and national levels to assess the effectiveness of national policy on speed limits and enforcement

- Specific design applications such as determining appropriate sight distances, relationships between speed and highway alignment, and speed performance with respect to steepness and length of grades

- Specific control applications, such as the timing of “yellow” and “all red” intervals for traffic signals, proper placement of signs, and development of appropriate signal progressions

- Investigation of high-accident locations at which speed is suspected to be a contributing cause to the accident experience

This list is illustrative. It is not intended to be complete, as there are myriad situations that may require speed data for a complete analysis. Such studies are of significant importance and are among the tasks most commonly conducted by traffic engineers.

11.2.3 Collection of Spot Speed Data

The collection of spot speed and other data in the field is discussed in Chapter 9. Much speed data are collected using permanent detector locations. Loop detectors are the most common form used in these cases. Where no permanent detectors are in place at the desired study location, a variety of techniques and portable equipment can be deployed. Most such measurements, however, are made using hand-held or vehicle-mounted radar "guns" or detectors.

Because the individual observed speeds will be in the form of a distribution that will have to be mathematically described, individual speeds are arranged in the form of *frequencies* of observations within defined *speed groups*. Thus, the results are in terms of "this many speeds were observed between X and Y mi/h," rather than a listing of individual speeds. This is done as it will facilitate the statistical analyses and determinations that will be extracted from the data.

11.2.4 Analysis and Presentation of Spot Speed Data

The best way to present the analysis of typical spot speed data is by example. The discussions of this section will be illustrated using a comprehensive sample application throughout. Because of this, the standard Sample Problem format used in other chapters will not be used here.

Figure 11.1 represents a typical set of field data from a spot speed study taken at a location of interest on a major arterial. As noted, the data are summarized as frequencies of observances in predefined speed groups.

Figure 11.1: Field Data for an Illustrative Spot Speed Study

Sunny and Dry
Weather Conditions

June 26, 2017
Date and Time

45 mi/h
Posted Speed Limit

SPEED GROUP		OBSERVED VEHICLES	TOTAL
Lower limit (mi/h)	Upper limit (mi/h)		
>30	32		0
>32	34	𝍸	5
>34	36	𝍸	5
>36	38	𝍸 ||	7
>38	40	𝍸 𝍸 |||	13
>40	42	𝍸 𝍸 𝍸 𝍸 |	21
>42	44	𝍸 𝍸 𝍸 𝍸 𝍸 𝍸 |||	33
>44	46	𝍸 𝍸 𝍸 𝍸 𝍸 𝍸 𝍸 𝍸 𝍸 |	46
>46	48	𝍸 𝍸 𝍸 𝍸 𝍸 𝍸 𝍸 𝍸 𝍸 𝍸 𝍸 𝍸 𝍸	65
>48	50	𝍸 𝍸 𝍸 𝍸 𝍸 𝍸 𝍸 𝍸	40
>50	52	𝍸 𝍸 𝍸 𝍸 𝍸 𝍸 𝍸 |||	38
>52	54	𝍸 𝍸 𝍸 𝍸 𝍸 𝍸	30
>54	56	𝍸 𝍸 𝍸 𝍸 𝍸	25
>56	58	𝍸 𝍸 𝍸 |||	18
>58	60	𝍸 𝍸	10
>60	62	𝍸 ||	7
>62	64	|||	3
>64	66	||	2
>66	68		0

Roger P Roess
Name

Radar
Method of Collection

Figure 11.1: Full Alternative Text

As the observed speeds form a distribution, they will eventually be described in terms of a continuous distribution function. The mathematical characteristics of a continuous distribution do not allow for the description of the probability of any distinct value occurring—in a continuous function, one discrete speed is one value in a distribution with an infinite number of such values. In more practical terms, a continuous distribution cannot mathematically describe the occurrence of a speed of exactly 44.72 mi/h. It can, however, describe the occurrence of a speed in the range of 44.7 to 44.8 mi/h. Therefore, the statistical analysis of speed data is based upon the number of observed values within a set of defined speed ranges.

The data shown in Figure 11.1 use speed groups that are 2 mi/h in breadth. This is a practical value that is quite typical, although 1 mi/h groups are also used if the sample sizes are large enough. For statistical reasons that will be explained later, speed groups with a range of more than 5 mi/h are never used. The number of speed groups defined must relate to the expected range of the data, and to the number of speeds that will be observed and recorded. For example, defining 15 speed groups and collecting only 30 speeds would be illogical, as there would only be an average of 2 observations per group. In general, it is customary to collect 15–20 speeds for each defined speed group. This *does not* imply that each group would have 15–20 observations; rather that the total number of observations will be sufficient to define the underlying distribution and its characteristics.

Frequency Distribution Table

The first analysis step is to take the data of Figure 11.1 and put it into the form of a frequency distribution table, as illustrated in Table 11.1. This tabular array shows the total number of vehicles observed in each speed group. For the convenience of subsequent use, the table includes one speed group at each extreme for which no vehicles were observed. The “middle speed” (S) of the third column is taken as the midpoint value within the speed group. The use of this value will be discussed in a later section.

Table 11.1: Frequency

Distribution Table for Illustrative Spot Speed Study

Speed Group		Middle Speed, S (mi/h)	No. of Vehicles Observed in Group, n	% Frequency	Cum. % Frequency	nS	nS^2
Lower Limit (mi/h)	Upper Limit (mi/h)						
>30	32	31	0	0.00	0.00	0.00	0.00
>32	34	33	5	1.36	1.36	165.00	5,445.00
>34	36	35	5	1.36	2.72	175.00	6,125.00
>36	38	37	7	1.90	4.62	259.00	9,583.00
>38	40	39	13	3.53	8.15	507.00	19,773.00
>40	42	41	21	5.71	13.86	861.00	35,301.00
>42	44	43	33	8.97	22.83	1,419.00	61,017.00
>44	46	45	46	12.50	35.33	2,070.00	93,150.00
>46	48	47	65	17.66	52.99	3,055.00	143,585.00
>48	50	49	40	10.87	63.86	1,960.00	96,040.00
>50	52	51	38	10.33	74.18	1,938.00	98,838.00
>52	54	53	30	8.15	82.34	1,590.00	84,270.00
>54	56	55	25	6.79	89.13	1,375.00	75,625.00
>56	58	57	18	4.89	94.02	1,026.00	58,482.00
>58	60	59	10	2.72	96.74	590.00	34,810.00
>60	62	61	7	1.90	98.64	427.00	26,047.00
>62	64	63	3	0.82	99.46	189.00	11,907.00
>64	66	65	2	0.54	100.00	130.00	8,450.00
>66	68	67	0	0.00	100.00	0.00	0.00
			368	**100.00**		**17,736.00**	**868,448.00**

Table 11.1: Full Alternative Text

The fourth column of the table shows the number of vehicles observed in each speed group. This value is known as the *frequency* for the speed group. These values are taken directly from the field sheet of Figure 11.1.

In the fifth column, the percentage of total observations in each speed group is computed as:

$\% = 100 \frac{n_i}{N}$ [11-1]

where:

n_i=number of observations (frequency) in speedgroup, and$N$=total number

For the 38–40 mi/h speed group, there are 13 observations in a total sample of 368 speeds. Thus, the percent frequency is 100×(13/368)=3.53% for this group.

The cumulative percent frequency (cum %) is the percentage of vehicles traveling at or below the highest speed in the speed group:

$cum\% = 100(\sum_{1-x} n_i/N)$ [11-2]

where *x* is the consecutive number (starting with the lowest speed group) of the speed group for which the cumulative percent frequency is desired.

For the 38–40 mi/h speed group, the sum of the frequencies for all speed groups having a high-speed boundary of 40 mi/h or less is found as 5+5+7+13=30. The cumulative percent frequency is then 100(30/368)=8.15% for this speed group.

The last two columns of the frequency distribution table are simple multiplications that will be used in subsequent computations.

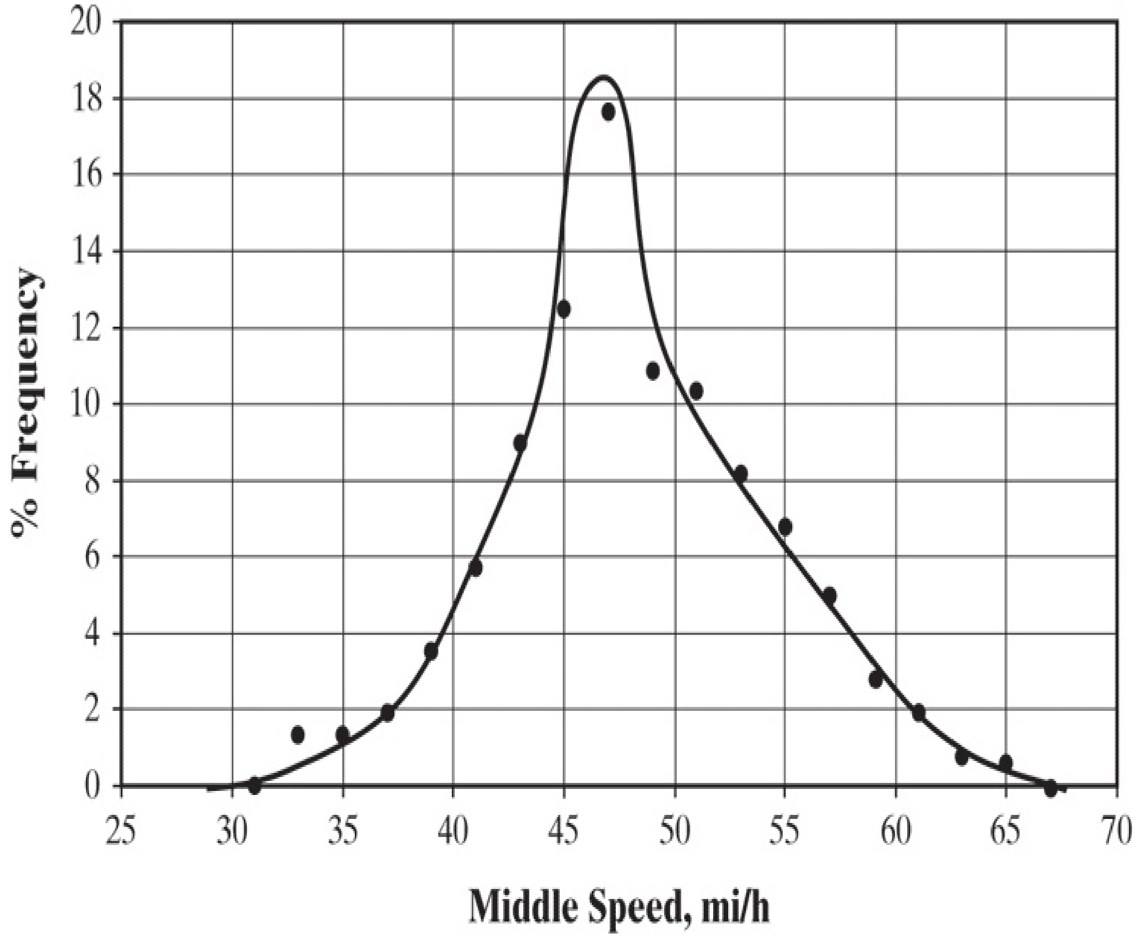
Frequency and Cumulative Frequency Distribution Curves

The data in Table 11.1 are then used to plot two curves that lend a visual impact to the information: (a) a frequency distribution curve and (b) a cumulative frequency distribution curve. These are illustrated in Figure 11.2, and are plotted as follows:

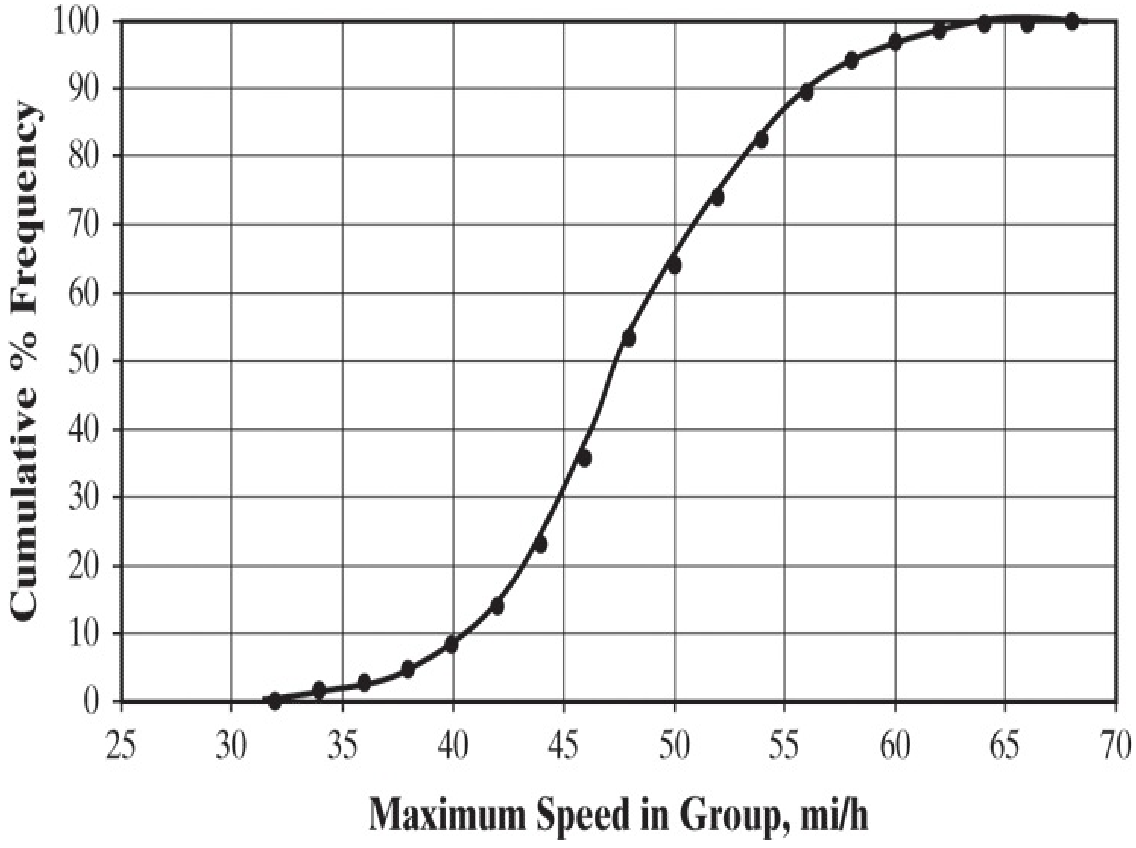
- Frequency distribution curve. For each speed group, the percent frequency of observations within the group is plotted versus the middle speed of the group (*S*).

- Cumulative frequency distribution curve. For each speed group, the percent cumulative frequency of observations is plotted versus the higher boundary of the speed group.

Figure 11.2: Frequency and Cumulative Frequency Curves for the Illustrative Problem



20
18
16
14
12
10
8
6
4
2
0
% Frequency
25
30
35
40
45
50
55
60
65
70
Middle Speed, mi/h



100
90
80
70
60
50
40
30
20
10
0
Cumulative % Frequency
25
30
35
40
45
50
55
60
65
70
Maximum Speed in Group, mi/h

Figure 11.2: Full Alternative Text

Note that the two frequencies are plotted versus *different* speeds. The middle speed is used for the frequency distribution curve. The cumulative frequency distribution curve, however, results in a very useful plot of speed versus the percentage of vehicles traveling at or below the designated speed. For this reason, the upper limit of the speed group is used as the plotting point.

In both cases, the plots are connected by a *smooth* curve that minimizes the total distance of points falling above the line and those falling below the line (on the vertical axis). A smooth curve is defined as one without any breaks in the slope of the curve. The “best fit” is done approximately (by eye). Some statistical packages will plot such a line automatically.

It is also convenient to plot the frequency distribution curve directly above the cumulative frequency distribution curve, using the same horizontal scale. This makes it easier to use the curves to graphically extract critical parameters.

Common Descriptive Statistics

Common descriptive statistics may be computed from the data in the frequency distribution table or determined graphically from the frequency and cumulative frequency distribution curves. These statistics are used to describe two important characteristics of the distribution:

- Central tendency: measures that describe the approximate middle or center of the distribution.
- Dispersion: measures that describe the extent to which data spread around the center of the distribution.

Measures of central tendency include the average or mean speed, the median speed, the modal speed, and the pace. Measures of dispersion include the 85th and 15th percentile speeds and the standard deviation.

Measures of Central Tendency:

Mean, Median, Mode, and Pace

The *average or mean speed* of a distribution is usually easily found as the sum of the observed values divided by the number of observations. In a spot speed study, however, individual values of speed may not be recorded; rather, the frequency of observations within defined speed groups is known. Computing the mean speed requires the assumption that *the average speed within a given speed group is the middle speed,* S, *of the group*. This is the reason that speed groups of more than 5 mi/h are never used. This assumption becomes less valid as the size of the speed groups increases. For 2 mi/h speed groups, as in the illustrative study, the assumption is usually quite good. If this assumption is made, the sum of all speeds in a given speed group may be computed as:

$n_i S_i$

where:

n_i=number of observations in speed group i, andS_i=middle speed of speed g

The sum of all speeds in the distribution may then be found by adding this product for all speed groups:

$\sum_i n_i S_i$

The mean or average speed is then computed as the sum divided by the number of observed speeds:

$\bar{x}=\frac{\sum_i n_i S_i}{N}$ [11-3]

where:

$\bar{x}$=average (time mean) speed for the sampleobservations, mi/h, andN=tota

For the illustrative study data presented in Figure 11.2 and Table 11.1, the average or mean speed is:

$\bar{x}=\frac{17{,}736}{368}=48.2\ \text{mi/h}$

where $\Sigma n_i S_i$ is the sum of the next-to-last column of the frequency distribution table of Table 11.1.

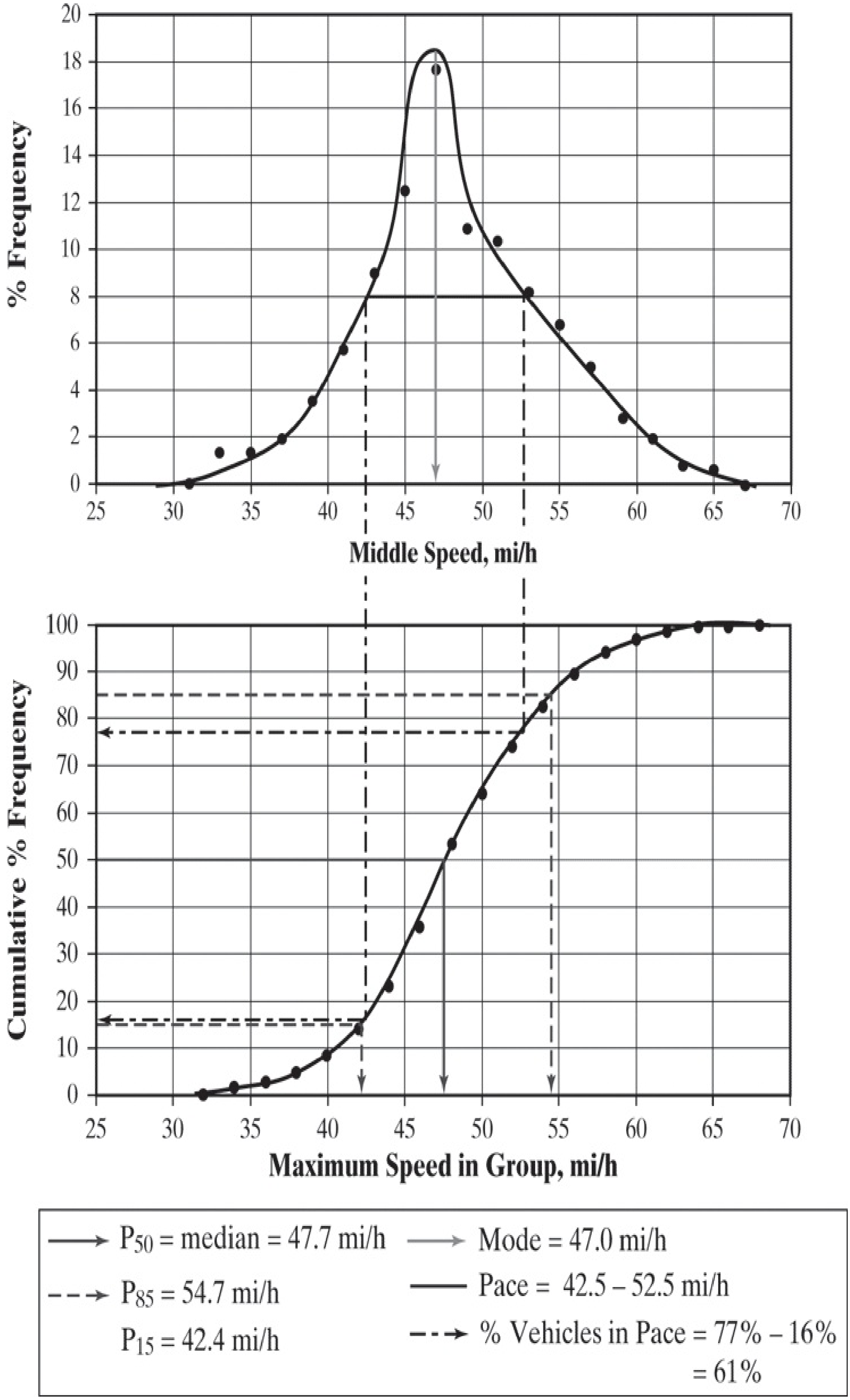
The *median speed* is defined as the speed that divides the distribution into equal parts (i.e., there are as many observations of speeds higher than the median as there are lower than the median). It is a positional value and is not affected by the absolute value of extreme observations.

The difference between the median and mean is best illustrated by example. Three speeds are observed: 30 mi/h, 40 mi/h, and 50 mi/h. Their average is (30+40+50)/3=40 mi/h. Their median is also 40 mi/h, as it equally divides the distribution, with one speed higher than 40 mi/h and one speed lower than 40 mi/h. Another three speeds are then observed: 30 mi/h, 40 mi/h, and 70 mi/h. Their average is (30+40+70)/3=46.7 mi/h. The median, however, is still 40 mi/h, with one speed higher and one speed lower than this observation. The mean is affected by the *magnitude* of the extreme observations; the median is affected only by the *number* of such observations.

As individual speeds have not been recorded in the illustrative study, the “middle value” is not easily determined from the tabular data of Table 11.1. It is easier to estimate the median graphically using the cumulative frequency distribution curve of Figure 11.2. By definition, the median equally divides the distribution. Therefore, 50% of all observed speeds should be less than the median.

This is exactly what the cumulative frequency distribution curve plots. If the curve is entered at 50% on the vertical axis, the median speed is found. This determination is illustrated, along with other descriptive variables, in Figure 11.3.

Figure 11.3: Graphic Determination of Key Variables Illustrated



% Frequency
Middle Speed, mi/h
Cumulative % Frequency
Maximum Speed in Group, mi/h
P50 = median = 47.7 mi/h
Mode = 47.0 mi/h
P85 = 54.7 mi/h
Pace = 42.5 – 52.5 mi/h
P15 = 42.4 mi/h
% Vehicles in Pace = 77% – 16%
= 61%

Figure 11.3: Full Alternative Text
$P_{50} = 47.7$ mi/h

where P_{50} is the median or 50th percentile speed.

The *pace* is a traffic engineering measure not commonly used for other statistical analyses. It is defined as *the 10 mi/h increment in speed in which the highest percentage of drivers is observed*. It is also found graphically using the frequency distribution curve of Figure 11.2. The solution recognizes that the area under the frequency distribution curve between any two speeds approximates the percentage of vehicles traveling between those two speeds, where the total area under the curve is 100%.

The pace is found as follows: A 10 mi/h template is scaled from the horizontal axis. Keeping this template horizontal, place an end on the lower left side of the curve and move slowly along the curve. When the right side of the template intersects the right side of the curve, the pace has been located. This procedure identifies the 10 mi/h increment that intersects the peak of the curve; this contains the most area and, therefore, the highest percentage of vehicles. The pace is shown in Figure 11.3 as 42.5 to 52.5 mi/h.

The mode is defined as the single value of speed that is most likely to occur. As no discrete values were recorded, the modal speed is also determined graphically from the frequency distribution curve. A vertical line is dropped from the peak of the curve, with the result found on the horizontal axis. For the illustrative study, the modal speed is 47.0 mi/h.

Measures of Dispersion

The most common statistical measure of dispersion in a distribution is the *standard deviation*. It is a measure of how far data spread around the mean value. In simple terms, the standard deviation is the average value of the difference between individual observations and the average value of those observations. Where discrete values of a variable are available, the equation for computing the standard deviation is:

$$s = \sqrt{\frac{\sum_i (x_i - \bar{x})^2}{N - 1}} \quad [11\text{-}4]$$

where:

s=the standard deviation,xi=speed measure i,x¯i=average speed, andN=nun

The difference between a given data point and the average is a direct measure of the magnitude of dispersion. These differences are squared to avoid positive and negative differences canceling, and summed for all data points. They are then divided by $N-1$. One statistical *degree of freedom* is lost because the mean of the distribution is known and used to compute the differences.

The principle of *degrees of freedom* can be explained by a simple example. If there are three numbers and it is known that the differences between the values and the mean for the first two are “3” and “2,” then the third or last difference must be “−5,” as the sum of all differences must be 0. Only the first “N − 1” observations of differences are statistically random.

Finally, a square root is taken of the results, as the values of the differences were squared to begin the computation.

Because discrete values of speed are not recorded, Equation 11-4 is modified to reflect group frequencies:

$$s=\sqrt{\frac{\Sigma n_i(S_i-\bar{x})^2}{N-1}}$$

which may be manipulated into a more convenient form, as follows:

$$s=\sqrt{\frac{\sum n_i S_i^2-N\bar{x}^2}{N-1}} \quad [11\text{-}5]$$

where all terms are as previously defined. This form is most convenient, as the first term is the sum of the last column of the frequency distribution table of Table 11.1. For the illustrative study, the standard deviation is:

$$s=\sqrt{\frac{868{,}448-368\,(48.22)^2}{368-1}}=6.14\ \text{mi/h}$$

Most observed speed distributions have standard deviations that are close to 5 mi/h, as this represents most driver behavior patterns reasonably well. Unlike averages and other central speeds, which vary widely from location to location, most speed studies yield similar standard deviations.

The *85th and 15th percentile speeds* give a general description of the high and low speeds observed by most reasonable drivers. It is generally thought that the upper and lower 15% of the distribution represents speeds that are either too fast or too slow for existing conditions. These values are found graphically from the cumulative frequency distribution curve, as shown in Figure 11.3. The curve is entered on the vertical axis at values of 85% and 15%. The respective speeds are found on the horizontal axis. For the illustrative study, these speeds are:

$$P_{85} = 54.7 \text{ mi/h} \quad P_{15} = 42.4 \text{ mi/h}$$

The 85th and 15th percentile speeds can be used to roughly estimate the standard deviation of the distribution, although this is not recommended when the data are available for a precise determination:

$$s_{est} = \frac{P_{85} - P_{15}}{2} \quad [11\text{-}6]$$

where all terms are as previously defined. For the illustrative spot speed study:

$$s_{est} = \frac{54.7 - 42.4}{2} = 6.15 \text{ mi/h}$$

In this case, the estimated value is very close to the actual computed value of 6.14 mi/h.

The 85th and 15th percentile speeds give insight to both the central tendency and dispersion of the distribution. As these values get closer to the mean, less dispersion exists and the stronger the central tendency of the distribution becomes.

The pace itself is a measure of the center of the distribution. The *percentage of vehicles traveling within the pace speeds* is a measure of both central tendency and dispersion. The smaller the percentage of vehicles traveling within the pace, the greater the degree of dispersion in the distribution.

The percentage of vehicles within the pace is found graphically using both frequency distribution and cumulative frequency distribution curves, as shown in Figure 11.3. The pace speeds were determined previously from the frequency distribution curves. Lines from these speeds are dropped vertically to the cumulative frequency distribution curve. The percentage of vehicles traveling at or below each of these speeds can then be determined from the vertical axis of the cumulative frequency distribution curve, as shown. Then:

% vehicles below 52.7mi/h=77.0% vehicles below 42.7 mi/h=16.0

Even though speeds between 32 and 66 mi/h were observed in this study, 61% of the vehicles traveled at speeds between 42.7 and 52.7 mi/h. The higher this percentage, the more centralized the data are—that is, tightly distributed around the mean. Lower percentages indicate greater dispersion. In this case, dispersion is somewhat more than expected. It is generally expected that about 70% of vehicles will travel within the pace. This correlates to the standard deviation, which is also somewhat higher than the expected 5 mi/h value, which also indicates greater than the usual dispersion.

11.2.5 Statistical Analysis of Spot Speed Data

All of the analysis presented to this point is done through simple arithmetic manipulation of the data themselves. To gain more insight and understanding of the data, it will be necessary to mathematically describe the data, and then use the mathematical description to conduct additional analysis.

Because the individual speed data form a distribution, the mathematical description of the data will be in the form of a *distribution function*. Such functions define the probability of an occurrence as the area under a designated portion of the curve. The area under the entire curve, by definition, must be 1.0, or 100%.

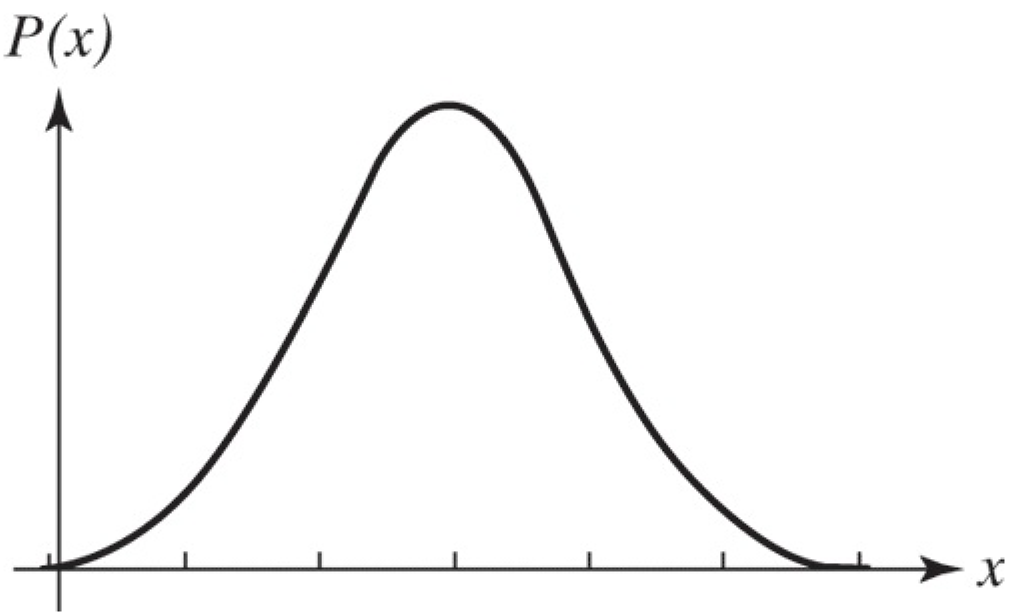
Like many human behavioral characteristics, speed data will most often be in the form of a *normal distribution*. The normal distribution has a strong central tendency around a mean value. As the value moves further away from the mean (on either side), the probability of its occurrence decreases. On a road with an average speed of 50 mi/h, there will be many speeds observed between 45 and 55 mi/h. There would be few speeds, however, observed at over 80 mi/h or below 20 mi/h.

The Normal Distribution and Its Characteristics

The *normal distribution* is one of the most commonly used in describing a wide variety of human behavioral characteristics. Figure 11.4 illustrates its general form.

$P(x)=12\pi\sigma e-(x-\mu)2/2\sigma 2$

Figure 11.4: General Form of the Normal Distribution Function



The value of the function, *P*(*x*), is entirely based upon the value of the variable *x*. All other symbols in the equation are, in fact, constants:

π=pi, or 3.14159,μ=true mean of the distribution,σ=true standard deviation

The theoretical normal distribution has a number of very interesting characteristics that will be useful in additional analyses.

- The distribution is perfectly symmetrical around the mean (μ).
- The distribution is asymptotic to the *x*-axis, that is, theoretical values of the statistic *x* extend from negative to positive infinity.
- The area under the curve denotes the probability of a value within the range for the area. The following key values are often referred to as follows:
    1. The probability of an occurrence between $\mu+\sigma$ and $\mu-\sigma=68.3\%$
    2. The probability of an occurrence between $\mu+2\sigma$ and $\mu-2\sigma$ is 95.5%.
    3. The probability of an occurrence between $\mu+3\sigma$ and $\mu+3\sigma$ is 97.7%.
    4. The probability of an occurrence between $\mu+1.96\sigma$ and $\mu-1.96\sigma$ is 95%.

The last value is of significance, as 95% is often used as a target probability to certify a key statistic as correct.

Another interesting characteristic is that for the mathematical normal distribution, the mean, the median, and the mode are all of the same value. This is because of the symmetry of values around the mean.

Figure 11.5 illustrates some of these key characteristics.

Figure 11.5: Key Characteristics of the Normal Distribution

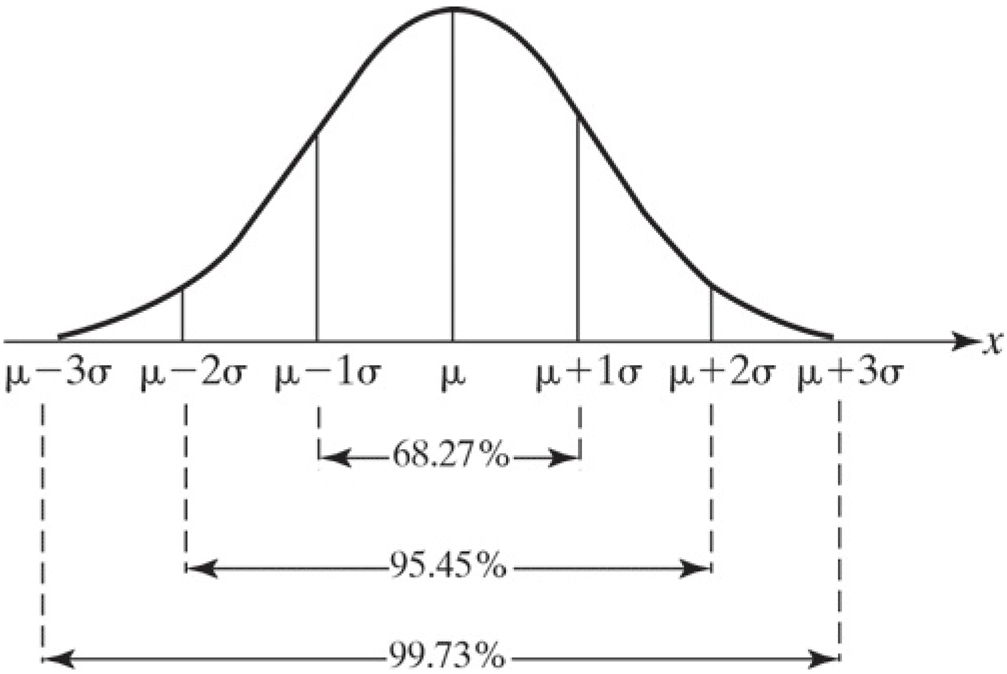


Figure 11.5: Full Alternative Text

The Standard Normal Distribution

A normal distribution is fully described by its mean (μ) and standard deviation (σ). The variance of the distribution is defined as the standard deviation squared (σ^2). Because these parameters fully define a normal distribution, a shorthand notation is commonly used:

x:N(μ,σ2)

which means *x* is a variable that is normally distributed with a mean of μ and a variance of σ2. For example, x:N(55,25) might signify a distribution of vehicle speeds with an average of 55 mi/h and a variance of 25—which implies a standard deviation of 5 mi/h.

To find probabilities of various ranges of values on a normal distribution, the distribution equation would have to be integrated. It would be useful if areas (probabilities) under the curve were tabulated for easy lookups. Unfortunately, there are an infinite number of different normal distributions, based on the possible values of mean and variance. However, tables do exist for the standard normal distribution, which has a defined mean of “0” and a defined standard deviation of “1.” It is noted as:

z:N(0,1)

where *z* is the designated statistic for this distribution. Table 11.2 shows a commonly used form of the standard normal distribution table.

Table 11.2 shows the area (probability) under the curve that is *less than* the value of *z*. The value of *z* is defined to the nearest tenth on the vertical axis, and to the nearest hundredth on the horizontal axis. The table then yields the probability of a value of *z* being less than or equal to the value entered.

For example, what is the probability of a value of *z* being less than or equal to 2.55 on the standard normal distribution? Table 11.2 is entered with 2.5 on the vertical axis and 0.05 on the horizontal axis. The value found is 0.9946. Thus, 0.9946 of all values on the standard normal distribution will be less than or equal to 2.55. The 0.9946 may be translated as 99.46%, of course.

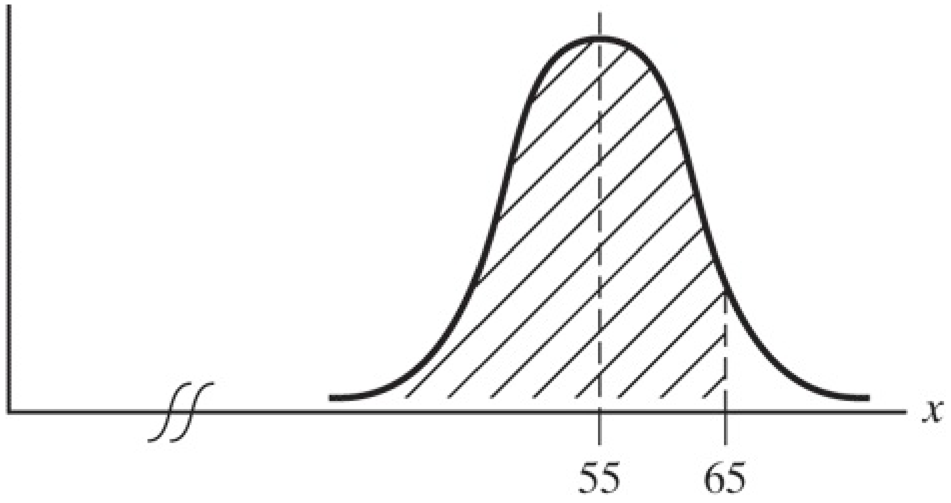
Various forms of the standard normal distribution exist in the literature. Some depict upper areas, that is, the probability of a value being *greater than or equal to* z. Some depict the area between +z and −z. Because the normal distribution is symmetric around the mean and the total area under the curve is 1.00, however, *all* of these can be discerned from Table 11.2.

For the previous example, if the probability of a value less than 2.55 is 0.9946, the probability of a value greater than 2.55 is 1−0.9946, or 0.0054. By symmetry, if the probability of a value greater than 2.55 is 0.0054, then the probability of a value less than −2.55 is also 0.0054. Then the probability of a value *between* +2.55 and −2.55 is 0.9946−0.0054=0.9892. Using this logic, any standard normal distribution table can be used to obtain any desired probability.

The measured distribution of speeds, however, is not going to be the standard normal distribution. The question is now: How do we convert a value of *x* on a distribution of x:N(μ,σ2) to an equivalent value of *z* on the standard normal distribution—*z*: *N*(0,1)? Figure 11.6 illustrates how this equivalence is established, using a sample normal distribution of *x*:[55,49] to find the probability of a speed being less than or equal to 65 mi/h.

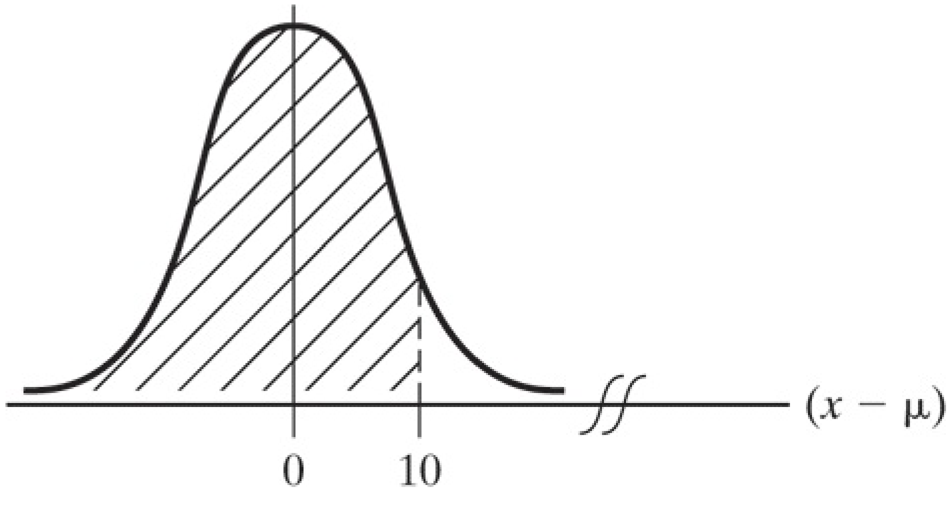
Figure 11.6: Shifting a

Normal Distribution to the Standard Normal Distribution



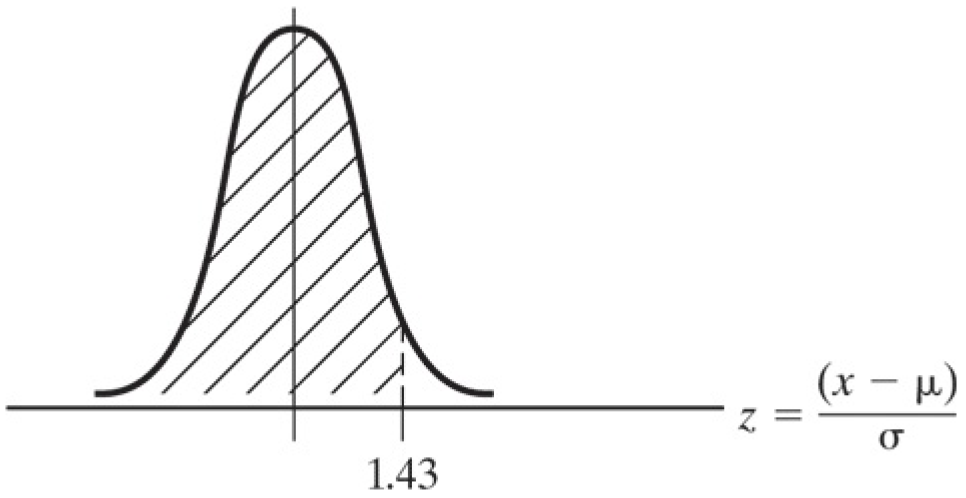
(a) The problem and normal distribution as specified

11.2-2 Full Alternative Text



(b) The axis translated to a zero mean

11.2-2 Full Alternative Text



(c) The axis scaled so that the *N*(0, 1) is used

11.2-2 Full Alternative Text

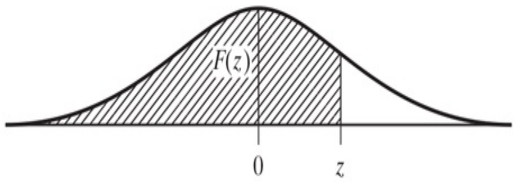
First, all values must be shifted to a mean of “0.” This is done by subtracting the mean value (55 mi/h in this case) from all values of *x*:

$z=x-\mu$

Then, the standard deviation of *x*, 7 mi/h in this case, must be shifted to a unit standard deviation of 1.00. This is done by dividing the difference between the mean and the data point by the standard deviation, σ. By doing this, every value of *x* on any arbitrary normal distribution can be converted to an equivalent value of *z* on the standard normal distribution:

$z=\frac{x-\mu}{\sigma}$ [11-7]

Table 11.2: The Standard Normal Distribution



$$F(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-z^2/2}\, dz$$

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
.4	.6554	.6591	.6628	.6661	.6700	.6736	.6772	.6808	.6844	.6879
.5	.6913	.6950	.6985	.7019	.7054	.7083	.7123	.7157	.7190	.7224
.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9432	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9658	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9716	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9812	.9826	.9830	.9834	.9838	.9842	.9846	.9854	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9937	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998

(Used with permission of Dover Publications, from E. L. Crow, F. A. Davis, and M. W. Maxfield, *Statistics Manual*. © 1960 Dover Publications.)

Table 11.2: Full Alternative Text

For the example, the equivalent value for 65 mi/h on a distribution of *x:* [55,49] on the standard normal distribution is computed as:

$$z=\frac{65-55}{7}=1.43$$

If the value of 1.43 is entered into Table 11.2, the probability of a value less than or equal to this is found to be 0.9236. Even though this value was found on the standard normal distribution, we can say that the probability of a speed of 65 mi/h or less on a normal distribution with a mean of 55 mi/h and a standard deviation of 7 mi/h is 92.36%.

Application: Tolerance and Confidence Intervals

When a spot speed study is conducted, a single value of the mean speed is computed. For the illustrative study of this chapter, the mean is 48.2 mi/h and the standard deviation is 6.14 mi/h, based upon a sample of 368 observations. In effect, this value, based upon a finite number of measured speeds, is being used to estimate the true mean of the underlying distribution of all vehicles traversing the site under uncongested conditions. The number of such vehicles, for all practical and statistical purposes, is infinite. The measured value of $\bar{x}$ is being used as an estimate for μ. The first statistical question that must be answered is: How good is this estimate?

Consider a classroom filled with 50 students. Because this is a finite sample, the average weight of class members could be absolutely determined: simply weigh each student and take the average. Perhaps that would take too long. We could select 10 people randomly and weigh only them, taking their average weight as an estimate of the average weight of the 50 students. In the extreme, we could randomly select 1 student, and use his or her weight as an estimate of the average weight of all 50 students. The last option wouldn’t be considered very accurate. In fact, if we are to use an estimate (i.e., not measure all 50 weights), the larger the sample size (10, 20, 30 students, etc.), the better the estimate of the average.

What if we considered the average speed of 48.2 mi/h to be part of a distribution of many average speeds of groups of 368 vehicles? From statistical theory, we know that if a distribution of individual values of speed is $x:N(\mu,\sigma2)$, then a distribution of average speeds ($x¯$) taken from the same population, with a constant sample size of 368 observations for each average speed, would be:

$xN¯:N(\mu,\sigma2N)$

where *N* is the sample size. In other words, the mean of the distribution would not change (think of the arithmetic, it can’t change). The variance, however, would become much smaller. If the average speed on a particular freeway segment is 60 mi/h, it is possible that we would find a few vehicles traveling at >80 mi/h. It is unlikely, however, that we would find 368 vehicles that have an *average* of >80 mi/h. Thus, a distribution of sample means will retain the same mean as the original distribution of single observations, but the variance (and standard deviation) will decrease as the sample size increases.

The standard deviation of a distribution of sample means is often referred to as the *standard error of the mean*, or:

$E=sN$ [11-8]

The standard error of the mean for our sample data is:

$E=6.14368=0.32$ mi/h

To obtain a statement regarding the relative accuracy of the sample mean (48.2 mi/h) as an estimate of the true mean of the infinite population of drivers in uncongested conditions, the known characteristics of the normal distribution are used. The probability of a value being within 1.96 standard deviations of the mean is 95%. The probability of a value being within 3.0 standard deviations of the mean is 99.7%. Our measured value of the mean, 48.2 mi/h, is now thought of as a single statistic from a distribution of sample means of 368 vehicles each. Thus:

$\bar{x} = u \pm 1.96\,E$ 95% of the time $\bar{x} = \mu \pm 3.0\,E$ 99.7% of the time

Translating for our data:

$48.2 = \mu \pm 1.96\,(0.32) = \mu$
$\pm 0.63 \mu = 48.2 \pm 0.63 = 47.57$ to 48.83 mi/h 95% of the time
$48.2 = \mu \pm 3.0\,(0.32) = \mu$
$\pm 0.96 \mu = 48.2 \pm 0.96 = 47.24$ to 49.16 mi/h 99.7% of the time

In English, this translates to the following statements:

- The true mean of the speed distribution lies between 47.57 and 48.83 mi/h with 95% *confidence.*
- The true mean of the speed distribution lies between 47.24 and 49.16 mi/h with 99.7% *confidence*.

The percentage value is called *confidence*. If we make a 95% confidence statement 100 times for 100 different studies, we can expect to be wrong 5 times. If we make a 99.7% statement 100 times, we can expect to be wrong 0.3 times—more reasonably, if we make such a statement 333 times, we can expect to be wrong once.

Obviously, as the confidence gets higher, the range gets bigger. Remember, that for a normal distribution, the result will be between $+\infty$ and $-\infty$ 100% of the time. Therefore, since we can never be 100% correct, the 95% confidence level is usually chosen, with occasional use of the 99.7% confidence level if fine accuracy is important.

The *tolerance* of an accuracy statement is the ± term, given the symbol *e*. It is related to the confidence level of the statement. Then:

for 95% confidence: $e = 1.96\,E = 1.96 sN$for 99.7% confidence: $e = 3.0\,E$ [11-9]

For our sample data, the value of *e* for 95% confidence was 0.63 mi/h; for 99.7% confidence, it was 0.96.

Application: Estimating the Sample Size

Although it is useful to know the confidence level and precision of a measured sample mean after the fact, it is more useful to determine what sample size is required to obtain a measurement that satisfies a predetermined precision and confidence level. Equation 11-9 can be solved for the sample size, *N*:

95% Confidence: N=1.962 s2e2=3.84 s2e299.7% Confidence: N=32 [11-10]

Consider the following problem: How many speeds must be collected to determine the true mean speed of the underlying distribution to within ±1.0 mi/h with 95% confidence? How do the results change if the tolerance is changed to ±0.5 mi/h and the confidence level to 99.7%?

The first problem is that the standard deviation of the distribution, *s*, is not known, as the study has not yet been conducted. Here, practical use is made of the knowledge that most speed distributions have standard deviations of approximately 5.0 mi/h. This value is assumed, and the results are shown in Table 11.3.

Table 11.3: Sample Size Computations Illustrated

Tolerance e (mi/h)	Confidence Level	
	95%	99.7%
1.0	$n = \frac{3.84(5)^2}{(1.0)^2} = 96$	$n = \frac{9.0(5)^2}{(1.0)^2} = 225$
0.5	$n = \frac{3.84(5)^2}{(0.5)^2} = 384$	$n = \frac{9.0(5)^2}{(0.5)^2} = 900$

Table 11.3: Full Alternative Text

A sample size of 96 speeds is required to achieve a tolerance of ±1.0 mi/h with 95% confidence. To achieve a tolerance of ±0.5 mi/h with 99.7% confidence, the required sample size must be almost 10 times greater. For most traffic engineering studies, a tolerance of ±1.0mi/h and a confidence level of 95% are quite sufficient.

Application: Before-and-After Spot Speed Studies

There are many situations in which existing speeds at a given location should be reduced. This occurs in situations where a high accident and/or accident severity rate is found to be related to excessive speed. It also arises where existing speed limits are being exceeded by an inordinate number of drivers.

There are many traffic engineering actions that can help reduce speeds, including lowered speed limits, stricter enforcement measures, warning signs, installation of rumble strips, and others. The major study issue, however, is to demonstrate that speeds have indeed been successfully reduced.

This is not an easy issue. Consider the following scenario: Assume that a new speed limit has been installed at a given location in an attempt to reduce the average speed by 5 mi/h. A speed study is conducted before implementing the reduced speed limit, and another is conducted several months after the new speed limit is in effect. Note that the “after” study is normally conducted after the new traffic engineering measures have been in effect for some time. This is done so that stable driver behavior is observed, rather than a transient response to something new. It is observed that the average speed of the “after” study is 3.5 mi/h less than the average speed of the “before” study. Statistically, there are two questions that must be answered:

- Is the observed reduction in average speeds real?
- Is the observed reduction in average speeds the intended 5 mi/h?

Although both questions appear to have obvious answers, they in fact do not. There are two reasons that a reduction in average speeds could have occurred: (1) the observed 3.5 mi/h reduction could occur because the new speed limit caused the true mean speed of the underlying distribution to be reduced, or (2) the observed 3.5 mi/h reduction could also occur because two different samples were selected from an underlying distribution that did not change. In statistical terms, the first is referred to as a *significant* reduction in speeds, whereas the latter is statistically *not significant* or *insignificant*.

The second question is equally tricky. Assuming that the observed 3.5 mi/h reduction in speeds is found to be statistically significant, it is necessary to determine whether the true mean speed of the underlying distribution has likely been reduced by 5 mi/h. Statistical testing will be required to answer both questions. Further, it will not be possible to answer either question with 100% certainty or confidence—95% is generally considered to be sufficient.

The concept of truth tables is used. The statistical tests for the significance of observed differences have four possible results: (1) the *actual* difference is significant, and the statistical test determines that it is *significant*; (2) the *actual* difference is not significant, and the statistical test determines that it is *not significant*; (3) the *actual* difference is significant and the statistical test determines that it is *not significant*; (4) the *actual* difference is not significant and the statistical test determines that it is *significant*. The first two outcomes result in an accurate assessment of the situation; the last two represent erroneous results. In statistical terms, outcome (4) is referred to as a Type I or α error, while outcome (3) is referred to as a Type II or β

error. The concept of the truth table is illustrated in Table 11.4.

Table 11.4: A Sample Truth Table

Actual Difference is:	Statistical Test Results in a Designation of:	
	Significant	**Not Significant**
Significant	Correct Result	Type I or α Error
Not Significant	Type II or β Error	Correct Result

Table 11.4: Full Alternative Text

In practical terms, the traffic engineer must avoid making a Type I error. In this case, it will appear that the problem (excessive speed) has been solved, when in fact it has not been solved. This may result in additional accidents, injuries, and/or deaths before the "truth" becomes apparent. If a Type II error is made, additional effort will be expended to entice lower speeds. Although this might involve additional expense, it is unlikely to lead to any negative safety results.

The statistical test applied to assess the significance of an observed reduction in mean speeds is the normal approximation. It is called an "approximation" because the distribution of differences in sample means (coming from the same underlying population) only approximates the normal distribution when the sample sizes for the "before" and "after" samples are ≥30. This will always be the case in properly conducted spot speed studies.

When two different sample means are observed, the observed difference is normally distributed when (a) the two samples come from the same underlying population and (b) the sample sizes are each ≥30. The difference is a statistic that is distributed as:

$(\bar{x}_1-\bar{x}_2):N(0,s_y^2)$

where:

sy=s12n1+s22n2 [11-11]

and:

sy=standard deviation of the difference in sample means,s1=standard devia

The normal approximation is applied by converting the observed reduction in mean speeds to a value of *z* on the standard normal distribution:

zd=(x¯1−x¯2)−0sy [11-12]

The standard normal distribution table of Table 11.2 is used to find the probability that a value equal to or less than zd occurs when both sample means are from the same underlying distribution. Then:

- If Prob ($z \leq zd$) ≥ 0.95, the observed reduction in speeds is *statistically significant*.
- If Prob ($z \leq zd$) < 0.95, the observed reduction in speeds is *not statistically significant*.

In the first case, it means that the observed difference in sample means would be exceeded less than 5% of the time, assuming that the two samples came from the same underlying distribution. Given that such a value was observed, this may be interpreted as being less than 5% probable that the observed difference came from the same underlying distribution and more than 95% probable that it resulted from a change in the underlying distribution.

Note that a *one-sided* test is conducted (i.e., we are testing the significance of an observed *reduction* in sample means, NOT an observed *difference* in sample means). If the observations revealed an increase in sample means, no statistical test is conducted, as it is obvious that the desired result was not achieved. Because of this, by convention, the higher speed value, which is usually the “before” sample, is labeled Sample 1, and the “after” sample as Sample 2.

If the observed reduction is found to be statistically significant, the second question can be entertained (i.e., Was the target speed reduction achieved?). This is done using only the results of the “after” distribution. Note that from the normal distribution characteristics, it is 95% probable that the true mean of the distribution is:

$\mu = \bar{x} \pm 1.96E$

If the target speed lies within this range, it can be stated that it was successfully achieved.

Consider the results of a before-and-after spot speed study, shown in Table 11.5, conducted to evaluate the effectiveness of a new speed limit intended to reduce the average speed at the location to 60 mi/h.

Table 11.5: Example Before-After Study of Speeds

Before Results		After Results
65.3 mi/h	$\bar{x}$	63.0 mi/h
5.0 mi/h	S	6.0 mi/h
50	N	60

Table 11.5: Full Alternative Text

A normal approximation test is conducted to determine whether the observed reduction in sample means is statistically significant:

The pooled standard deviation is computed using Equation 11-11:

$s_Y = 5.0250 + 6.0260 = 1.05 \text{mi/h}$

The z_d statistic is computed using Equation 11-12:

$z_d = (65.3 - 63.0) - 01.05 = 2.19$

From Table 11.2, the probability of z being less than or equal to 2.19 is found as:

Prob ($z \leq 2.19$) = 0.9857

As 98.57%>95%, the results indicate that the observed reduction in sample means was *statistically significant*.

Given these results, it is now possible to investigate whether or not the target speed of 60 mi/h was successfully achieved in the "after" sample. The 95% confidence interval for the "after" estimate of the true mean of the underlying distribution is:

$E=6/60=0.7746 \mu=63.0 \pm 1.96(0.7746) \mu=63.0 \pm 1.52 \mu=61.48 - 64.52$ mi/h

As the target speed of 60 mi/h does not lie in this range, it cannot be stated that it was successfully achieved.

In this case, while a significant reduction of speeds was achieved, it was not sufficient to achieve the target value of 60 mi/h. Additional study of the site would be undertaken and additional measures enacted to achieve additional speed reduction.

The 95% confidence criteria for certifying a significant reduction in observed speeds should be well understood. If a before-and-after study results in a confidence level of 94.5%, it would not be certified as statistically significant. This decision limits the probability of making a Type I error to less than 5%. When we state that the observed difference in mean speeds is not statistically significant in this case, however, it is 94.5% probable that we are making a Type II error. Before expending large amounts of funds on additional speed-reduction measures, a larger "after" speed sample should be taken to see whether or not 95% confidence can be achieved with an expanded data base.

Application: Testing for Normalcy Using the Chi-Square Goodness-of-Fit Test

Virtually all of the statistical analyses of this section start with the basic assumption that the speed distribution can be mathematically represented as normal. For completeness, it is therefore necessary to conduct a statistical test to confirm that this assumption is correct. The chi-square test is used to determine whether the difference between an observed distribution and its assumed mathematical form is significant. For grouped data, the chi-squared statistic is computed as:

$$\chi^2=\sum^{N_G}\frac{(n_i - f_i)^2}{f_i} \quad [11\text{-}13]$$

where:

χ^2=chi-square statistic, n_i=number of observations (frequency) in speed group i, f_i=t

Table 11.3 shows these computations for the illustrative spot speed study. Speed groups are already specified, and the observed frequencies are taken directly from the frequency distribution table (Table 11.1).

For convenience, the speed groups are listed from highest to lowest. This is to coordinate with the standard normal distribution table (Table 11.2), which gives probabilities of $z \leq z_d$. The upper limit of the highest group is adjusted to "infinity," as the theoretical normal distribution extends to both positive and negative infinity. The remaining columns of Table 11.6 focus on determining the theoretical frequencies, f_i, and on determining the final value of of χ^2.

Table 11.6: Chi-Square Test on Sample Data

Speed Group									
Upper Limit (mi/h)	Lower Limit (mi/h)	Observed Frequency n	Upper Limit on STD Normal zd	Prob. $z \leq zd$ (Table 10-2)	Prob. Of Occurrence in Group	Theoretical Frequency f	Combined Group n	Combined Group f	Chi-Square Group χ^2
∞	64	2	∞	1.0000	0.0051	1.8768			
64	62	3	2.57	0.9949	0.0071	2.6128	12	10.0832	0.3644
62	60	7	2.25	0.9878	0.0152	5.5936			
60	58	10	1.92	0.9726	0.0294	10.8192	10	10.8192	0.0620
58	56	18	1.60	0.9432	0.0452	16.6336	18	16.6336	0.1122
56	54	25	1.27	0.8980	0.0716	26.3488	25	26.3488	0.0690
54	52	30	0.94	0.8264	0.0940	34.592	30	34.592	0.6096
52	50	38	0.62	0.7324	0.1183	43.5344	38	43.5344	0.7036
50	48	40	0.29	0.6141	0.1261	46.4048	40	46.4048	0.8840
48	46	65	−0.03	0.4880	0.1286	47.3248	65	47.3248	6.6015
46	44	46	−0.36	0.3594	0.1111	40.8848	46	40.8848	0.6400
44	42	33	−0.68	0.2483	0.0921	33.8928	33	33.8928	0.0235
42	40	21	−1.01	0.1562	0.0661	24.3248	21	24.3248	0.4544
40	38	13	−1.34	0.0901	0.0416	15.3088	13	15.3088	0.3482
38	36	7	−1.66	0.0485	0.0252	9.2736	7	9.2736	0.5574
36	34	5	−1.99	0.0233	0.0129	4.7472	10	8.5744	0.2370
34	-∞	5	−2.31	0.0104	0.0104	3.8272			
					1.0000	**368.0000**	**368.0000**	**368.0000**	**11.6669**

Table 11.6: Full Alternative Text

The theoretical frequencies are the numbers of observations that would have occurred in the various speed groups *if the distribution were perfectly normal*. To find these values, the probability of an occurrence within each speed group must be determined from the standard normal table. This is done in columns 4 through 7 of Table 11.6, as follows:

1. The upper limit of each speed group (in mi/h) is converted to an

equivalent value of *z* on the standard normal distribution, using Equation 11-7. The computation for the upper limit of 64 mi/h is illustrated below:

$$z_{64} = \frac{64 - 48.2}{6.14} = 2.57$$

Note that the mean speed and standard deviation of the illustrative spot speed study are used in this computation.

2. Each computed value of *z* is now looked up on the standard normal table (Table 11.2). From this, the probability of $z \le z_d$ is found and entered into column 5 of Table 11.6.

3. Consider the 48–50 mi/h speed group in Table 11.6. From column 5, 0.6141 is the probability of a speed ≤50 mi/h occurring on a normal distribution; 0.4880 is the probability of a speed ≤48 mi/h occurring. Thus, the probability of an occurrence between 48 and 50 mi/h is 0.6141−0.4880=0.1261. The probabilities of column 6 are computed via sequential subtractions as shown here. The result is the probability of a speed being in any speed group, assuming a normal distribution.

4. The theoretical frequencies of column 7 are found by multiplying the sample size by the probability of an occurrence in that speed group. Fractional results are permitted for theoretical frequencies.

5. The chi-square test is valid only when all values of the theoretical frequencies are 5 or more. To achieve this, the first three and last two speed groups must be combined. The observed frequencies are similarly combined.

6. The value of chi-square for each speed group is computed as shown. The computation for the 40–42 mi/h speed group is illustrated here:

$$\chi^2 = \frac{(n_i - f_i)^2}{f_i} = \frac{(21 - 24.3248)^2}{24.3248} = 0.4544$$

These values are summed to yield the final value of χ^2 for the distribution, which is 11.6669.

To assess this result, a table of the chi-square distribution must be used. It is shown in Table 11.7. Probability values are shown on the horizontal axis of the table. The vertical axis shows *degrees of freedom*. For a chi-square distribution, the number of degrees of freedom is the number of data groups (after they are combined to yield theoretical frequencies of 5 or more), minus 3. Three degrees of freedom are lost because the computation of χ^2 requires that three characteristics of the measured distribution be known: the mean, the standard deviation, and the sample size. Thus, for the illustrative spot speed study, the number of degrees of freedom is 14–3=11.

Table 11.7: Upper Percentage Points on the Chi-Square Distribution

$$P(\chi^2) = \int_{\chi^2}^{\infty} \frac{1}{(f-2/2)12^{1/2}} (\chi^2)^{(f-2)/2_0 - \chi^2/2} \, d(\chi^2)$$

df	.995	.990	.975	.950	.900	.750	.500	.250	.100	.050	.025	.010	.005
1	3927×10^{-2}	1571×10^{-7}	9821×10^{-7}	3932×10^{-8}	0.01579	0.1015	0.4549	1.323	2.706	3.841	5.024	6.635	7.879
2	0.01003	0.02010	0.05064	0.1026	.2107	.5754	1.386	2.773	4.605	5.991	7.378	9.210	10.60
3	.07172	.1148	.2158	.3518	.5844	1.213	2.366	4.108	6.251	7.815	9.348	11.34	12.34
4	.2070	.2971	.4844	.7107	1.064	1.923	3.357	5.585	7.779	9.488	11.14	13.28	14.86
5	.4117	.5543	.8312	1.145	1.610	2.675	4.351	6.626	9.236	11.07	12.83	15.09	16.75
6	.6757	.8721	1.237	1.635	2.204	3.455	5.348	7.841	10.64	12.59	14.45	16.81	18.55
7	.9893	1.259	1.690	2.167	2.833	4.255	6.346	9.037	12.02	14.07	16.01	18.48	20.28
8	1.344	1.646	2.180	2.733	3.199	5.071	7.344	10.22	13.36	15.51	17.53	20.09	21.98
9	1.735	2.088	2.700	3.325	4.168	5.899	8.343	11.39	14.68	16.92	19.02	21.67	23.59
10	2.150	2.558	3.247	3.940	4.865	6.737	9.342	12.55	15.99	18.31	20.48	23.21	25.19
11	2.603	3.053	3.816	4.575	5.578	7.584	10.34	13.70	17.28	19.68	21.92	24.72	26.76
12	3.074	3.571	4.404	5.226	6.304	8.458	11.34	14.85	18.55	21.03	23.34	26.22	28.30
13	3.565	4.107	5.009	5.892	7.042	9.299	12.34	15.98	19.81	22.36	24.74	27.69	29.82
14	4.075	4.660	5.629	6.571	7.790	10.17	13.34	17.12	21.06	23.68	26.12	29.14	31.32
15	4.601	5.229	6.262	7.261	8.547	11.04	14.34	18.25	22.31	25.00	27.49	30.58	32.80
16	5.142	5.812	6.908	7.962	9.312	11.91	15.34	19.37	23.54	26.30	28.85	32.00	34.27
17	5.697	5.408	7.564	8.672	10.09	12.79	16.34	20.49	24.77	27.59	30.19	33.41	35.72
18	6.265	7.015	8.231	9.390	10.86	13.68	17.34	21.60	25.99	28.87	31.53	34.81	37.16
19	6.844	7.644	8.907	10.12	11.65	14.56	18.34	22.72	27.20	30.14	32.85	36.19	38.58
20	7.434	8.260	9.591	10.85	12.44	15.45	19.34	23.83	28.41	31.41	34.17	37.57	40.00
21	8.034	8.897	10.28	11.59	13.24	16.34	20.34	24.93	29.62	32.67	35.48	38.93	41.40
22	8.643	9.542	10.98	12.34	14.04	17.24	21.34	26.04	30.81	33.92	36.78	40.29	42.60
23	9.260	10.20	11.69	13.09	14.85	18.14	22.34	27.14	32.01	35.17	38.08	41.64	44.18
24	9.886	10.86	12.40	13.85	15.66	19.04	23.24	28.24	33.20	36.42	39.36	42.98	45.58
25	10.52	11.52	13.12	14.61	16.47	19.94	24.34	29.34	34.38	37.65	40.65	44.31	46.93
26	11.16	12.20	13.84	15.38	17.29	20.84	25.34	30.43	35.56	38.89	41.92	45.64	48.29
27	11.81	12.88	14.57	16.15	18.11	21.75	26.34	31.53	36.74	40.11	43.19	46.96	49.64
28	12.46	13.56	15.31	16.93	18.94	22.66	27.34	32.62	37.92	41.34	44.46	48.28	50.99
29	13.12	14.26	16.05	17.71	19.77	23.57	28.34	33.71	39.09	42.58	45.72	49.59	52.34
30	13.79	14.95	16.79	18.49	20.60	24.48	29.34	34.80	43.26	43.77	46.98	50.89	53.67
40	20.71	22.16	24.43	26.51	29.05	33.66	39.34	45.62	51.80	55.76	59.34	63.69	66.77
50	27.99	29.71	32.36	34.76	37.69	42.94	49.33	56.33	63.17	67.50	71.42	76.15	79.49
60	35.53	37.48	40.48	43.19	46.46	52.29	59.33	66.98	79.08	79.08	83.30	88.38	91.95
70	43.28	45.44	48.76	51.74	55.33	61.70	69.33	77.58	85.53	90.53	95.02	100.42	104.22
80	51.17	53.54	57.15	60.39	64.28	71.14	79.33	88.13	96.58	101.88	106.63	112.33	116.32
90	59.20	61.75	65.65	69.13	73.29	80.62	89.33	98.65	107.56	113.14	118.14	124.12	128.30
100	67.33	70.00	74.22	77.93	82.36	90.13	99.33	109.14	118.50	124.34	129.56	135.81	140.17
	−2.576	−2.326	1.960	−1.645	−1.28	−0.6745	0.0000	+0.6745	+1.282	+1.645	+1.960	+2.326	576

(Source: E. L. Crow, F. A. Davis, and M. W. Maxwell, *Statistics Manual*, Dover Publications, Mineola, NY, 1960.)

Table 11.7: Full Alternative Text

The values of χ^2 are shown in the body of Table 11.7. For the illustrative data, the value of χ^2 lies between the tabulated values of 10.34(Prob=0.50) and 13.70(Prob=0.25). Note also that the probabilities shown in the table represent the probability of a value being *greater* than or equal to χ^2. Interpolation is used to determine the precise probability level associated with a value of 11.6669 on a chi-square distribution with 11 degrees of freedom:

Probability	Chi-Square Value
0.50	10.34
P	11.6669
0.25	13.70

11.2-8 Full Alternative Text

$$P=0.25+(0.50-0.25)\left(\frac{13.70-11.6669}{13.70-10.34}\right)=0.40125$$

From this determination, it is 40.125% probable that a χ^2 value of 11.6669 or higher would exist if the distribution were statistically normal. The decision criteria are the same as for other statistical tests (i.e., to say that the data and the assumed mathematical description are *significantly different*, we must be 95% confident that this is true). For tables that yield a probability of a value *less than or equal to* the computed statistic, the probability must be 95% or more to certify a significant difference. This was the case in the normal approximation test. The corresponding decision point using a table with probabilities greater than or equal to the computed statistic is that the probability must be 5% *or less* to certify a significant statistical difference. In the case of the illustrative data, the probability of a value of 11.6669 or greater is 40.125%. This is quite a bit more than 5%. Thus, the data and the assumed mathematical description are *not significantly different*, and data's normalcy is successfully demonstrated.

A chi-square test is rarely conducted on spot speed results, since they are virtually always normal. If the data are seriously skewed, or take a shape obviously different from the normal distribution, this will be relatively obvious, and the test can be conducted. It is also possible to compare the data with other types of distributions. There are a number of distributions that have the same general shape as the normal distribution but have skews to the low or high end of the distribution. It is also possible that a given set of data can be reasonably described using a number of different distributions. This does not negate the validity of a normal description when it occurs. As long as speed data can be described as normal, all of the manipulations described herein are valid. If a speed distribution is found to be not normal, then other distributions can be used to describe it, and other statistical tests can be performed. These are not covered in this text, and the student is referred to standard statistics textbooks.

Applications: Other Statistical Tests

The normal approximation is the test generally used to establish a difference or reduction in speed. It requires that before-and-after sample sizes be 30 or more. Should one or both sample sizes not meet this requirement, the correct statistical test would use the *Students' t-Distribution.* This test is covered in virtually all standard statistical texts.

Note that in all of the applications discussed, the sample standard deviation, *s*, is used as an estimate of *σ*, the true mean of the population. The accuracy of this estimate can be evaluated using the *F-Distribution*, which is also covered in all standard statistical texts. In practical terms, the estimate is virtually always statistically adequate.

11.3 Travel-Time Studies

Travel-time studies involve significant lengths of a facility or group of facilities forming a route. Information on the travel time between key points within the study area is sought and is used to identify those segments in need of improvements. Travel-time studies are often coordinated with delay observations at points of congestion along the study route.

Travel-time information is used for many purposes, including the following:

- To identify problem locations on facilities by virtue of high travel times and/or delay.
- To measure arterial level of service, based on average travel speeds and travel times.
- To provide necessary input to traffic assignment models, which focus on link travel time as a key determinant of route selection.
- To provide travel-time data for economic evaluation of transportation improvements.
- To develop time contour maps and other depictions of traffic congestion in an area or region.

11.3.1 Field Study Techniques

Because travel time studies take place over significant lengths of highway, it is difficult to remotely observe the behavior of individual vehicles from outside the traffic stream. The general method of collecting data is to use a series of test vehicles driven in the traffic stream. An observer is riding with each test vehicle. The observer records the times at which key locations are passed and makes notes on any stopped delay times, and the causes of those delays where possible. As an alternative to carrying an observer with each vehicle, test vehicles can be equipped with automated recording devices.

To ensure consistency in test car results, drivers are generally instructed to use one of three specific driving techniques:

- Floating Car Technique: In the floating car technique, the driver seeks to maintain his/her position in the flow of vehicles. This is done in the following way: the driver will pass one vehicle for every vehicle that is observed passing the test car.

- Maximum Car Technique: In this approach, the driver is instructed to go as fast as possible, consistent with safety, without ever exceeding the design speed (or speed limit) of the roadway.

- Average Car Technique: In this case, the driver is instructed to approximate the average speed of vehicles (based upon driver judgment).

The floating car and average car techniques yield estimates of average travel times through the study segment. The maximum car technique yields a higher speed and lower travel times, sometimes approaching the 85th percentile speed of the distribution, which yields travel times at about the 15th percentile. Thus, it is critical that *all* test car drivers use the *same* driving technique.

The floating car method is very difficult to implement by drivers. It is virtually never used on multilane highways, where the numbers of passing vehicles may be large, requiring the driver to proceed aggressively to pass many vehicles as well. It can be an effective technique on two-lane facilities where passing maneuvers are rare.

While test car techniques are frequently used to gather travel time data, care must be taken to avoid using so many test cars that the travel time behavior of the traffic stream is fundamentally altered. Thus, sample sizes resulting from test-car studies are relatively low, and this can negatively impact the statistical accuracy and precision of the study results.

There are other techniques that can be used to collect travel time data without the use of test cars. One technique is to record license plate numbers and the time they are observed at key locations along a study route. Matching software is used to identify individual vehicles and their travel times between the points of observation. There are two key limitations to such an approach: (a) although total travel time between observation points is obtained, there is no detailed information on conditions or events in between, and (b) sampling can be difficult, given that it is almost impossible to observe all license plates of all vehicles at any given point.

The sampling issue is straightforward: assume that there are four observation points, and that at each point, 50% of all license plates are recorded. How many data sets will include the same vehicle at all observation points? The probability of a specific observation at one point is 0.50. The probability of observing the same vehicle at two observation points is $0.50 \times 0.50 = 0.25$. The probability of observing the same vehicle at three observation points is $0.50 \times 0.50 \times 0.50 = 0.125$. The probability of observing the same vehicle at all four observation points is $0.50 \times 0.50 \times 0.50 \times 0.50 = 0.0625$. Therefore, even if 50% of all vehicles are recorded at each point, the sample size yielding all four times is 6.25%. If more observation points are needed, then the sample size issue becomes even more difficult.

The license plate technique is most often used where there are only two observation points, perhaps three. In such cases, a reasonable matching sample can exceed what would be possible with test cars. Intermediate information along the route, however, would be lost.

In rare cases, an elevated vantage point will be available that allows real-time tracing of vehicles along a study segment of significant length. A number of observers would be needed, however, as a single observer will only be able to follow one vehicle at a time.

The sample data sheet of Table 11.8 is for a 7-mile section of Lincoln Highway, which is a major suburban multilane highway of six lanes. Checkpoints are defined in terms of mileposts. Intersections or other known geographic markers can also be used as identifiers. The elapsed stopwatch time to each checkpoint is noted. Section data refers to the distance between the previous checkpoint and the checkpoint noted. Thus, for the section labeled MP 16, the section data refer to the section between mileposts 16 and 17. The total stopped delay experienced in each section is noted, along with the number of stops. The “special notes” column contains the observer’s determination of the cause(s) of the delays noted. Section travel times are computed as the difference between cumulative times at successive checkpoints.

Table 11.8: A Sample Travel Time Field Sheet

Site: Lincoln Highway			Run No. 3			Start Location: Milepost 15.0
Recorder: William McShane			Date: Aug 10, 2002			Start Time: 5:00 PM

Checkpoint	Cum. Dist. Along Route (mi)	Cum. Trav. Time (min:sec)	Per Section			
			Stopped Delay (s)	No. of Stops	Section Travel Time (min:sec)	Special Notes
MP 16	1.0	1:35	0.0	0	1:35	
MP 17	2.0	3:05	0.0	0	1:30	
MP 18	3.0	5:50	42.6	3	2:45	Stops due to signals at: MP17.2 MP17.5 MP18.0
MP 19	4.0	7:50	46.0	4	2:00	Stops due to signal MP18.5 and double parked cars.
MP 20	5.0	9:03	0.0	0	1:13	
MP 21	6.0	10:45	6.0	1	1:42	Stop due to school bus.
MP 22	7.0	12:00	0.0	0	1:15	
Section Totals	**7.0**		**88.6**	**8**	**12:00**	

Table 11.8: Full Alternative Text

In this study, the segments ending in MP 18 and 19 display the highest delays and, therefore, the highest travel times. If this is consistently shown in *all* or *most* of the test runs, these sections would be subjected to more detailed study. As the delays are indicated as caused primarily by traffic control signals, their timing and coordination would be examined carefully to see if they can be improved. Double parking is also noted as a cause in one segment. Parking regulations would be reviewed, along with available legal parking supply, as would enforcement practices.

11.3.2 Travel Time Data along an Arterial: An Example in Statistical Analysis

Given the cost and logistics of travel-time studies (test cars, drivers, multiple runs, multiple days of study, etc.), there is a natural tendency to keep the number of observations, *N*, as small as possible. This case considers a hypothetical arterial on which the true mean running time is 196 s over a 3-mile section. The standard deviation of the travel time is 15 s. The distribution of running times is normal. Note that the discussion is, at this point, limited to *running times*. These do not include stopped delays encountered along the route and are not equivalent to *travel times*, as will be seen.

Given the normal distribution of travel times, the mean travel time for the section is 196 s, and 95% of all travel times would fall within $1.96(15)=29.4$ s of this value. Thus, the 95% confidence interval for travel times would be between $196–29.4=166.6$ s and $196+29.4=225.4$ s.

The speeds corresponding to these travel times (including the average) are:

$$S_1=\frac{3\ \text{mi}}{225.4\ \text{s}} \times 3600\ \frac{\text{s}}{\text{h}}=47.9\ \text{mi/h} \quad S_{av}=\frac{3\ \text{mi}}{196\ \text{s}} \times 3600\ \frac{\text{s}}{\text{h}}=55.1\ \text{mi/h} \quad S$$

Note that the average of the two 95% confidence interval limits is $(47.9 + 64.8)/2=56.4$ mi/h, NOT 55.1 mi/h. This discrepancy is due to the fact that the *travel times* are normally distributed and are therefore symmetric. The resulting running speed distribution is skewed. The distribution of speeds, which are inverse to travel times, cannot be normal if the travel times are normal. The 55.1 mi/h value is the appropriate average speed, based on the observed average travel time over the 3-mile study section.

So far, this discussion considers only the *running times* of test vehicles through the section. The actual *travel time* results of 20 test-car runs are illustrated in Figure 11.7.

Figure 11.7: Histogram of Hypothetical Travel Time Data for 20 Runs Over a 3-Mile Arterial Segment

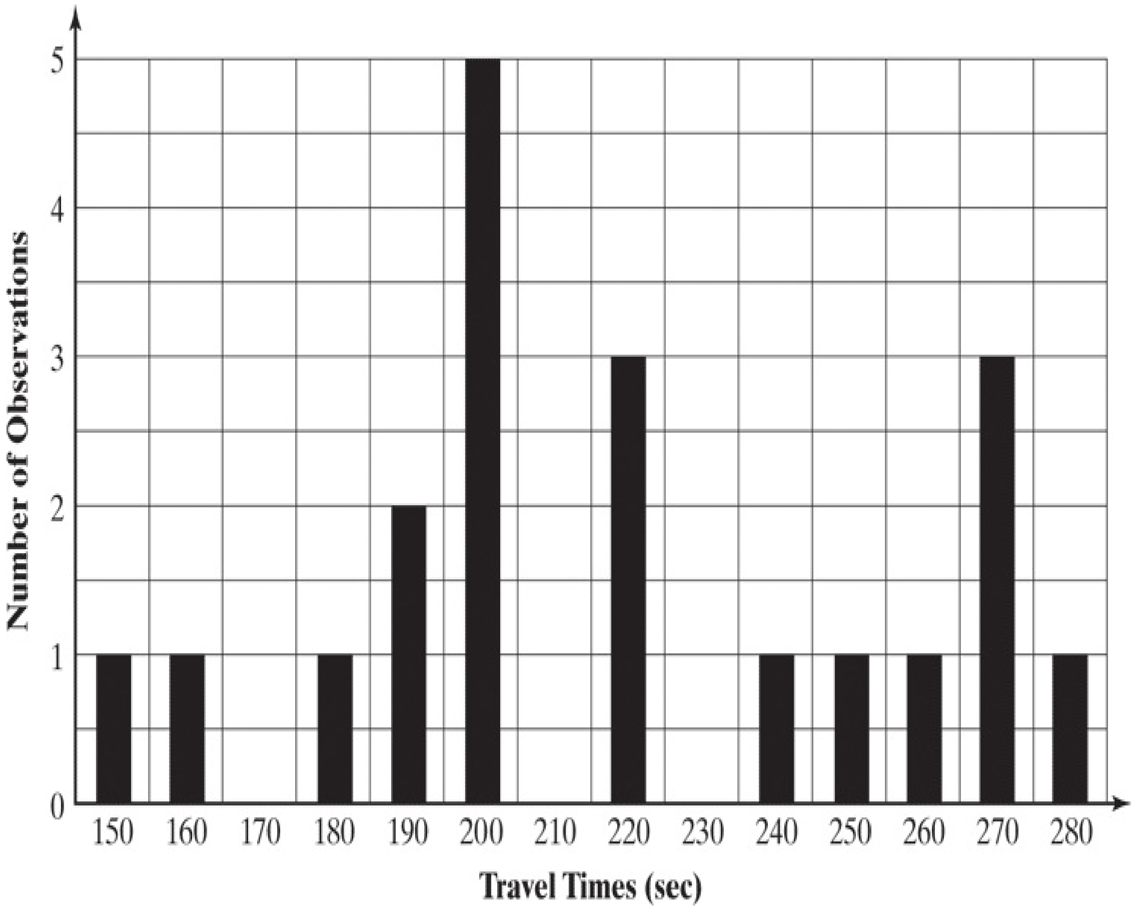


Figure 11.7: Full Alternative Text

This distribution does not look normal. In fact, it is not normal at all, as the total travel time represents the *sum* of running times (which are normally distributed) and stop time delay that follows another distribution entirely.

It is postulated that the distribution of stop time delay is as shown in Table 11.9.

Table 11.9: Postulated Distribution of Stop Time Delays

No. of Signal Stops	Probability of Occurrence	Duration of Stops
0	0.569	0s
1	0.300	40s
2	0.131	80s

Table 11.9: Full Alternative Text

Observations of the type illustrated in Table 11.8 result from the combination of random driver selection of running speeds and signal delay effects that follow the relationship specified above.

The actual mean travel time of the observations in Figure 11.7 is 218.5 s, with a standard deviation of 38.3 s. The 95% confidence limits on the average are:

$$218.5 \pm 1.96(38.3/20) = 218.5 \pm 16.79 \quad 201.71 - 235.29 \text{s}$$

The speeds associated with these average and limiting travel times are:

$$S1 = 3 \text{ mi} \quad 235.29 \text{s} \times 3600 \text{sh} = 45.9 \text{ mi/h} \quad Sav = 3 \text{ mi} \quad 218.5 \text{s} \times 3600 \text{sh} = 49.4 \text{ mi/h} \quad S$$

Another way of addressing the average travel time is to add the average running time (196 s) to the average delay time, which is computed from the probabilities noted above as:

$d_{av} = (0.569 \times 0) + (0.300 \times 40) + (0.131 \times 80) = 22.5 s$

The average travel time is then expected to be 196.0 + 22.5=218.5 s, which is the same average obtained from the histogram of measurements.

11.3.3 Overriding Default Values: Another Example of Statistical Analysis of Travel-Time Data

Figure 11.8 shows a default curve calibrated by a local highway jurisdiction for average travel speed along four-lane arterials within the jurisdiction. As with all “standard” values, the use of another value is always permissible as long as there are specific field measurements to justify replacing the standard value.

Figure 11.8: Default Curve Specified by Agency (Illustrative)

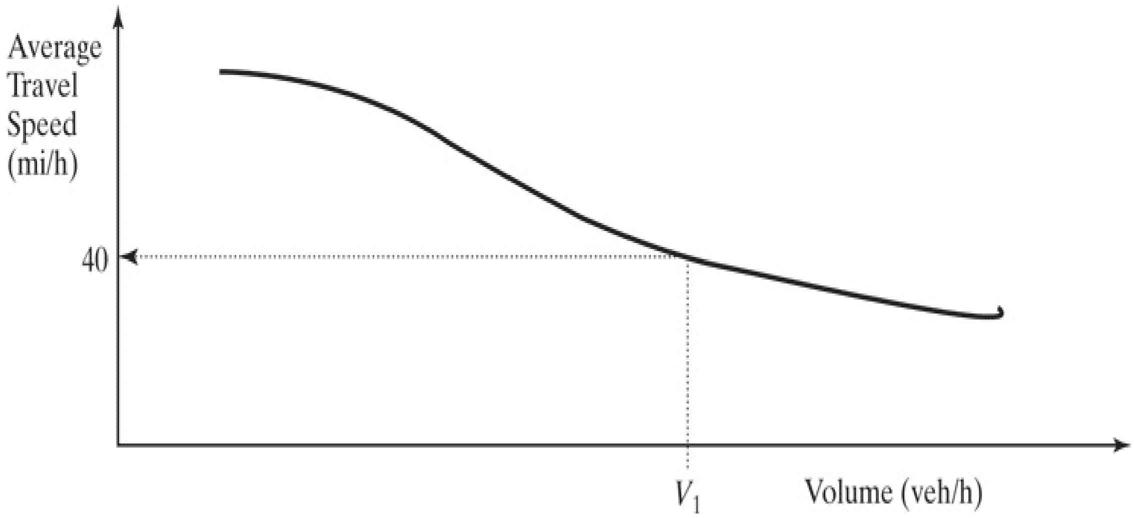


Figure 11.8: Full Alternative Text

Assume that a case exists in which the default value of travel speed for a given volume, V1, is 40 mi/h. Based on three travel-time runs over a 2-mile section, the measured average travel speed is 43 mi/h. The analysts would like to replace the standard value with the measured value. Is this appropriate?

The statistical issue is whether or not the observed 3 mi/h difference between the standard value and the measured value is *statistically significant*. As a practical matter (in this hypothetical case), practitioners generally believe that the standard values of Figure 11.8 are too low and that higher values are routinely observed. This suggests that a one-sided hypothesis test should be used.

Figure 11.9 shows a probable distribution of the random variable $Y=\Sigma t_i/N$, the estimator of the average travel time through the section. Based on the standard and measured average travel speeds, the corresponding travel times over a 2-mile section of the roadway are $(2/40) \times 3{,}600=180.0$ s and $(2/43) \times 3{,}600=167.4$ s. These two values are formulated, respectively, as the null and alternative hypotheses, as illustrated in Figure 11.9. The following points relate to Figure 11.9:

- Type I and Type II errors are equalized and set at 5% (0.05).
- From the standard normal table (Table 11.2), the value of z_d corresponding to Prob. $(z \le z_d)=0.95$ (corresponding to a one-sided test with Type I and II errors set at 5%) is 1.645.
- The difference between the null and alternate hypotheses is a travel time of $180.0-167.4=12.6$, noted as Δ.
- The standard deviation of travel times is known to be 28.0 s.

Figure 11.9: Testing the Default (Null Hypothesis) Against the Proposed Alternative Hypothesis

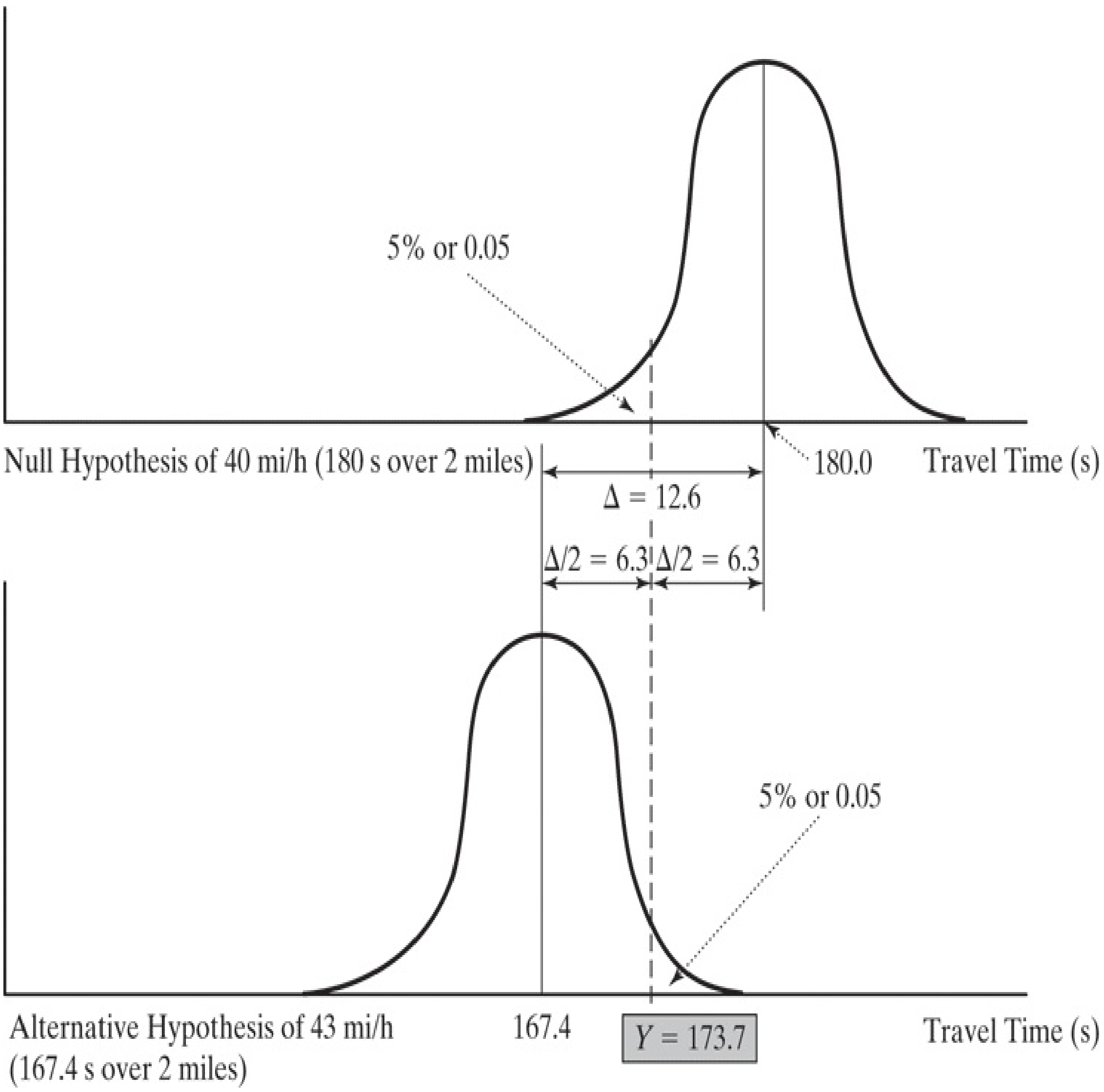


Figure 11.9: Full Alternative Text

From Figure 11.9, for the difference between the default and alternative hypotheses to be statistically significant, the value of Δ/2 must be equal to or larger than 1.645 times the standard error for travel times, or:

Δ/2≥1.645 (sN)6.3≥1.645 (283)=26.6

In this case, the difference is not significant, and the measured value of 43 mi/h cannot be accepted in place of the default value. This relationship can, of course, be solved for *N*:

N ≥ 8,486Δ2

using the known value of the standard deviation (28). Remember that Δ is stated in terms of the difference in *travel times* over the 2-mile test course, not the difference in average travel speeds. Table 11.10 shows the sample size requirements for accepting various alternative average travel speeds in place of the default value. For the alternative hypothesis of 43 mi/h to be accepted, a sample size of 8,486/(12.6)2=54 would have been required. However, as illustrated in Figure 11.9, had 54 samples been collected, the alternative hypothesis of 43 mi/h would have been accepted as long as the average travel time was less than 173.7 s (i.e., the average travel speed was greater than (2/173.7)×3,600=41.5 mi/h.Table 11.10 shows a number of different alternative hypotheses, along with the required sample sizes and decision points for each to be accepted.

Table 11.10: Required Sample Sizes and Decision Values for the Acceptance of Various Alternative Hypotheses

Default Value (Average Travel Speed) (mi/h)	Alternative Hypothesis (Average Travel Speed) (mi/h)	Required Sample Size N	Decision Point (Average Travel Speed) Y (mi/h)
40	42	≥115	41.0
40	43	≥54	41.4
40	44	≥32	41.9
40	45	≥22	42.4

Table 11.10: Full Alternative Text

While this problem illustrates some of the statistical analyses that can be applied to travel-time data, the reader should examine whether the study, as formulated, is appropriate. Should the Type II error be equalized with the Type I error? Does the existence of a default value imply that it should not? Should an alternative value higher than any measured value ever be accepted? (For example, should the alternative hypothesis of 43 mi/h be accepted if the average travel speed from a sample of 54 or more measurements is 41.6 mi/h, which is greater than the decision value of

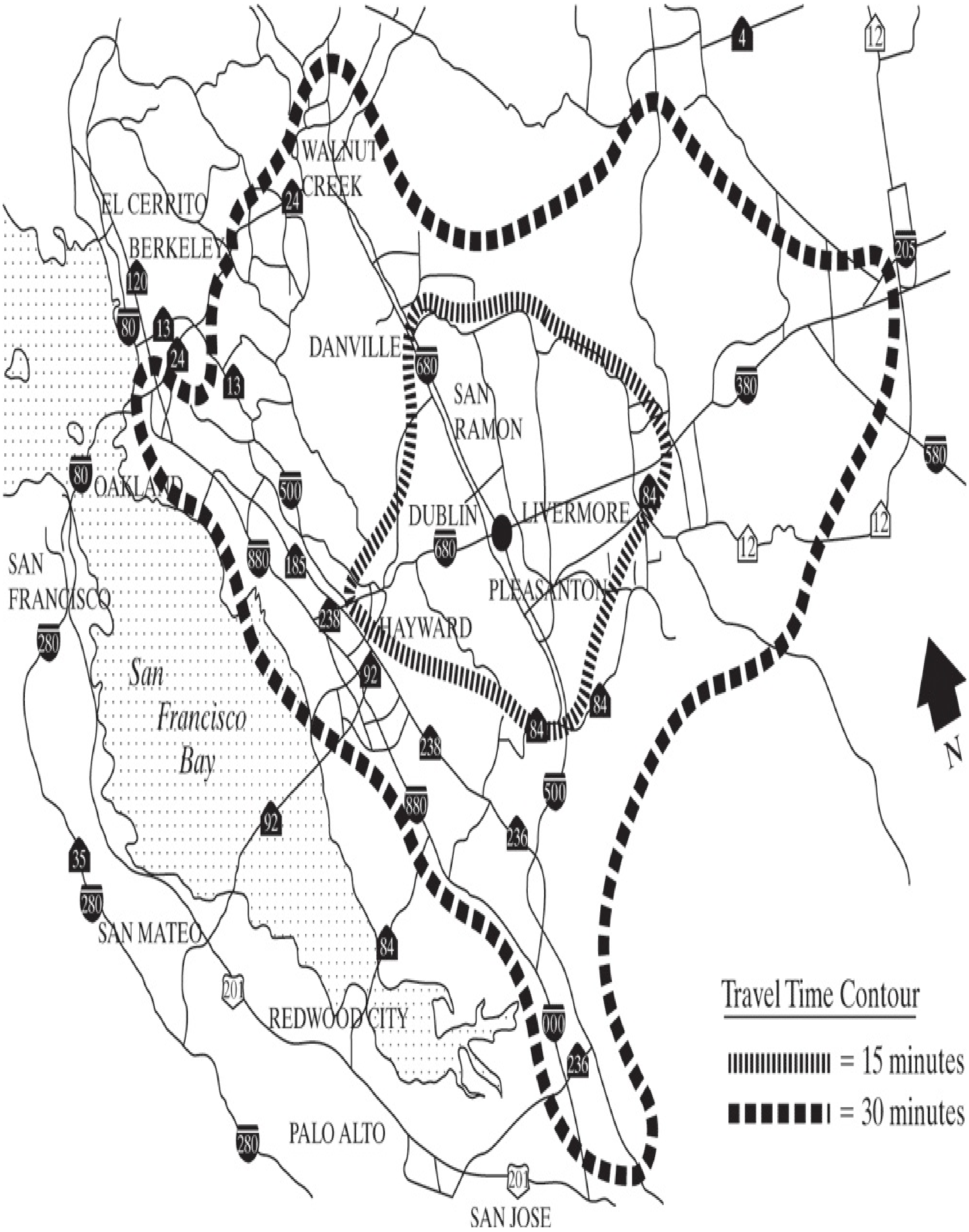
41.5 mi/h?)

Given the practical range of sample sizes for most travel time studies, it is very difficult to justify overriding default values for individual cases. However, a compendium of such cases—each with individually small sample sizes—can and should motivate an agency to review the default values and curves in use.

11.3.4 Travel-Time Displays

Travel-time data can be displayed in many interesting and informative ways. One method that is used for overall traffic planning in a region is the development of a travel-time contour map, of the type shown in Figure 11.10. Travel times along all major routes entering or leaving a central area are measured. Time contours are then plotted, usually in increments of 15 minutes. The shape of the contours gives an immediate visual assessment of corridor travel times in various directions. The closer together contour lines plot, the longer the travel time to progress any set distance. Such plots can be used for overall planning purposes and for identifying corridors and segments of the system that require improvement.

Figure 11.10: A Travel Time Contour Map



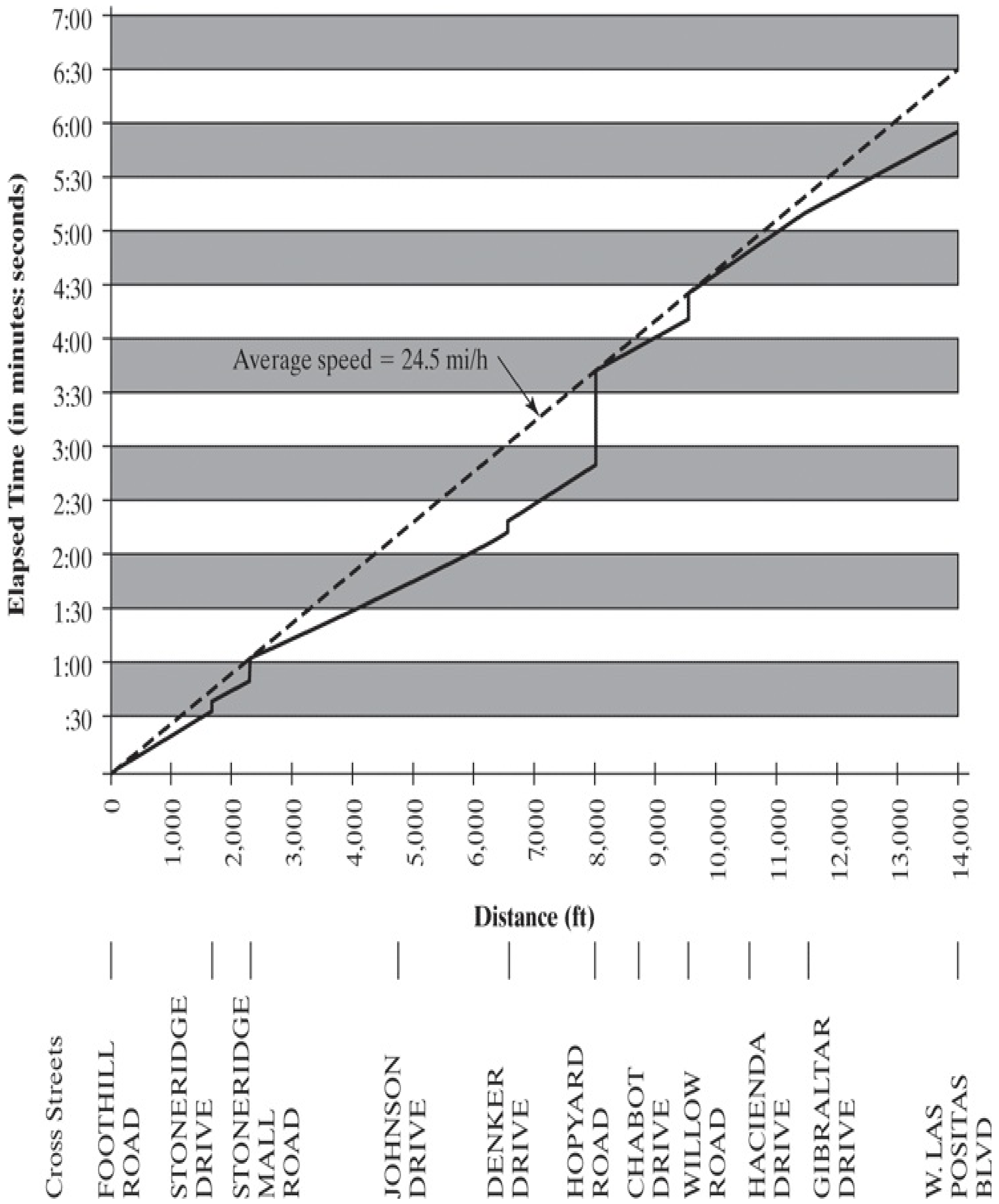
(Source: Used with permission of Prentice-Hall Inc, from Pline, J., Editor, *Traffic Engineering Handbook*, 4th Edition, Institute of Transportation Engineers, Washington, DC, 1992, Figure 3-6, pg 69.)

Figure 11.10: Full Alternative Text

Travel time along a route can be depicted in different ways as well. Figure

11.11 shows a plot of cumulative time along a route. The slope of the line in any given segment is speed (ft/s), and stopped delays are clearly indicated by vertical lines. Figure 11.12 shows average travel speeds plotted against distance. In both cases, problem areas are clearly indicated, and the traffic engineer can focus on those sections and locations experiencing the most congestion, as indicated by the highest travel times (or lowest average travel speeds).

Figure 11.11: A Plot of Elapsed Time vs. Distance

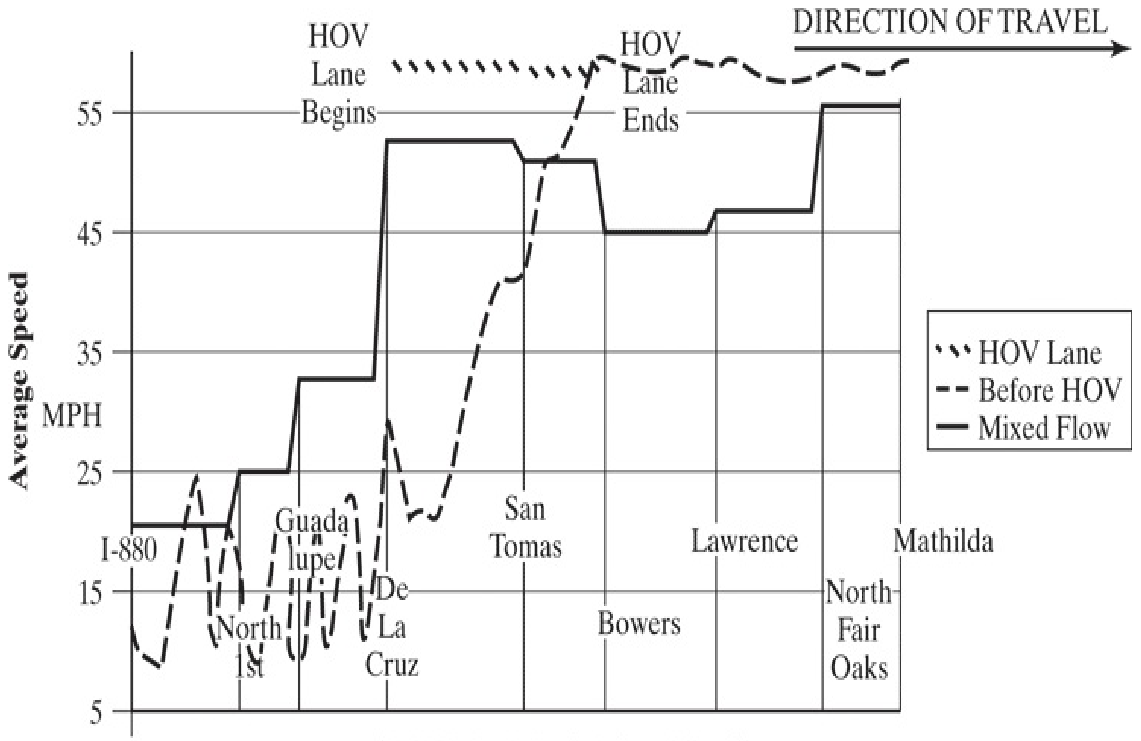


(Source: Used with permission of Prentice-Hall Inc, from Pline, J., Editor, *Traffic Engineering Handbook*, 4th Edition, Institute of Transportation Engineers, Washington, DC, 1992, Figure 3-7, pg 70.)

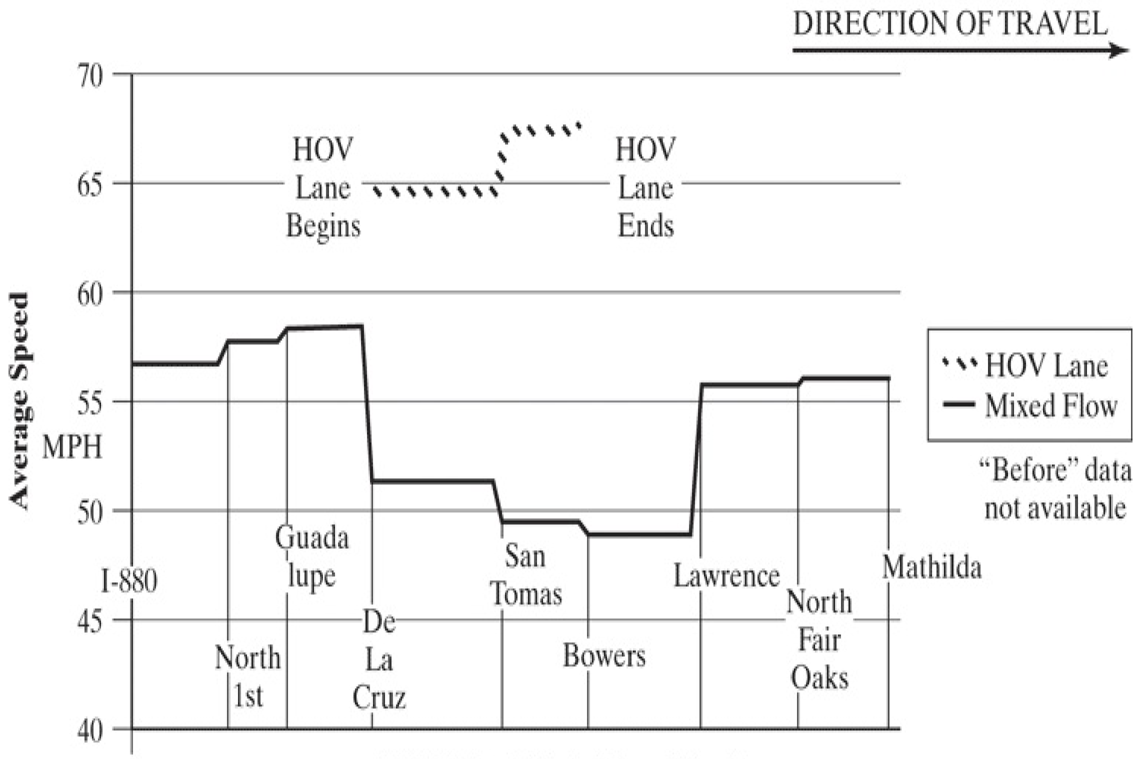
Figure 11.11: Full Alternative Text

Figure 11.12: Average Travel

Speeds Plotted vs. Segments of a Route



AM NB Peak Period Speed Profile



PM NB Peak Period Speed Profile

(Source: Used with permission of Prentice-Hall Inc, from Pline, J., Editor, *Traffic Engineering Handbook*, 4th Edition, Institute of Transportation Engineers, Washington, DC, 1992, Figure 3-8, pg 71.)

Figure 11.12: Full Alternative Text

11.4 Intersection Delay Studies

Some types of delay are measured as part of a travel time study by noting the location and duration of stopped periods during a test run. A complicating feature for all delay studies lies in the various definitions of delay, as reviewed earlier in the chapter. The measurement technique must conform to the delay definition.

Before 1997, the primary delay measure at intersections was stopped delay. Although no form of delay is easy to measure in the field, stopped delay was certainly the easiest. However, the current measure of effectiveness for signal and STOP-controlled intersections is *total control delay*. Control delay is best defined as time-in-queue delay plus time losses due to deceleration from and acceleration to ambient speed. The *Highway Capacity Manual* [1] defines a field measurement technique for control delay, using the field sheet shown in Figure 11.13.

Figure 11.13: Field Sheet for Signalized Intersection Delay Studies

INTERSECTION CONTROL DELAY WORKSHEET

General Information	Information
Analyst __________	Intersection __________
Agency or Company __________	Area Type ☐ CBD ☐ Other
Date Performed __________	Analysis Year __________
Analysis Time Period __________	

Input Initial Parameters

Number of Lanes, N __________	Total Vehicles Arriving V_T __________
	Stopped Vehicle Count V_{STOP} __________
Survey Count Interval I_s __________	Cycle Length D (s) __________

Input Field Data

Clock Time	Cycle Number	Number of Vehicles in Queue — Count Interval									
		1	2	3	4	5	6	7	8	9	10
Total											

(Source: Reprinted with permission from Used with permission of Transportation Research Board, *Highway Capacity Manual*, 4th Edition, © 2000 by the National Academy of Sciences, Courtesy of the National Academies Press, Washington, D.C.)

Figure 11.13: Full Alternative Text

The study methodology recommended in the *Highway Capacity Manual* is based on direct observation of vehicles-in-queue at frequent intervals and requires a minimum of two observers. The following should be noted:

1. The method is intended for undersaturated flow conditions and for cases where the maximum queue is about 20 to 25 vehicles.

2. The method does not directly measure acceleration–deceleration delay but uses an adjustment factor to estimate this component.

3. The method also uses an adjustment to correct for errors that are likely to occur in the sampling process.

4. Observers must make an estimate of free-flow speed before beginning a detailed survey. This is done by driving a vehicle through the intersection during periods when the light is green and there are no queues and/or by measuring approach speeds at a position where they are unaffected by the signal.

Actual measurements start at the beginning of the red phase of the subject lane group. There should be no overflow queue from the previous green phase when measurements start. The following tasks are performed by the two observers:

- Observer 1:

    - Keeps track of the end of standing queues for each cycle by observing the last vehicle in each lane that stops due to the signal. This count includes vehicles that arrive on green but stop or approach within one car length of queued vehicles that have not yet started to move.

    - At intervals between 10 s and 20 s, the number of vehicles in queue is recorded on the field sheet. The regular intervals for these observations should be an integral divisor of the cycle length. Vehicles in queue are those that are included in the queue of stopping vehicles (as previously defined) and have not yet exited the intersection. For through vehicles, “exiting the intersection” occurs when the rear wheels cross the STOP line; for turning vehicles, “exiting” occurs when the vehicle clears the opposing vehicular or pedestrian flow to which it must yield and begins to accelerate.

  - At the end of the survey period, vehicle-in-queue counts continue until all vehicles that entered the queue during the survey period have exited the intersection.

- Observer 2:

  - During the entire study period, separate counts are maintained of vehicles arriving during the survey period and of vehicles that stop one or more times during the survey period. Stopping vehicles are counted only once, regardless of how many times they stop.

For convenience, the survey period is defined as an integer number of cycles, although an arbitrary length of time (e.g., 15 minutes) could also be used and would be necessary where an actuated signal is involved.

Each column of the vehicle-in-queue counts is summed; the column sums are then added to yield the total vehicle-in-queue count for the study period. It is then assumed that the average time-in-queue for a counted vehicle is the time interval between counts. Then:

$TQ=(IS\times\sum ViqVT)\times 0.90$ [11-14]

where:

TQ=average time-in-queue, s/veh,IS=time interval between time-in-queue counts, s,$\sum Viq$=sum of all vehicle-in-queue counts, veh,VT=total number of vehicles arriving during thestudy pe

The adjustment factor (0.90) adjusts for errors that generally occur when this type of sampling technique is used. Such errors usually result in an overestimate of delay.

A further adjustment for acceleration–deceleration delay requires that two values be computed: (1) the average number of vehicles stopping per lane, per cycle, and (2) the proportion of vehicles arriving that actually stop. These are computed as:

VSLC=VSTOPNc×NL [11-15]

where:

VSLC=number of vehicles stopping per lane, percycle (veh/ln/cycle),VSTC
FVS=VSTOPVT [11-16]

where FVS is the fraction of vehicles stopping, and all other variables are as previously defined.

Using the number of stopping vehicles per lane, per cycle, and the measured free-flow speed for the approach in question, an adjustment factor is found from Table 11.11.

Table 11.11: Adjustment Factor for Acceleration/Deceleration Delay

Free-Flow Speed (mi/h)	Vehicles Stopping Per Lane, Per Cycle (V_{SLC})		
	≤7 vehs	8–19 vehs	20–30 vehs
≤37	+5	+2	−1
>37–45	+7	+4	+2
>45	+9	+7	+5

(Source: Reprinted with permission from Used with permission of Transportation Research Board, *Highway Capacity Manual*, 4th Edition, © 2000 by the National Academy of Sciences, Courtesy of the National Academies Press, Washington, D.C.)

Table 11.11: Full Alternative Text

The final estimate of control delay is then computed as:

$$d=TQ+(FVS\times CF) \quad [11\text{-}17]$$

where:

d=total control delay, s/veh, andCF=adjustment factor from Table 11.11.

Table 11.12 shows a facsimile of a field sheet, summarizing the data for a survey on a signalized intersection approach. The approach has two lanes, and the signal cycle length is 60 s. Ten cycles were surveyed, and the vehicle-in-queue count interval is 20 s.

Table 11.12: Sample Data for a Signalized Intersection Delay Study

Clock Time	Cycle Number	Number of Vehicles in Queue		
		+0 secs	+20 secs	+40 secs
5:00 PM	1	4	7	5
5:01 PM	2	6	6	5
5:02 PM	3	3	5	5
5:03 PM	4	2	6	4
5:04 PM	5	5	3	3
5:05 PM	6	5	4	5
5:06 PM	7	6	8	4
5:07 PM	8	3	4	3
5:08 PM	9	2	4	3
5:09 PM	10	4	3	5
	Total	**40**	**50**	**42**

$\sum Vqi=40+50+42=132$ vehs $VT=120$ vehs (observed) $VSTOP=75$ (obs

Table 11.12: Full Alternative Text

The average time-in-queue is computed using Equation 11-14:

$TQ=(20\times132120)\times0.90=19.8$ s/veh

To find the appropriate correction factor from Table 11.11, the number of vehicles stopping per lane per cycle is computed using Equation 11-15:

$VSLC=7510\times 2=3.75$ vehs

Using this and the measured free-flow speed of 35 mi/h, the correction factor is +5s. The control delay is now estimated using Equations 11-16 and 11-17:

FVS=75120=0.625d=19.8 + (0.625 × 5)=22.9 s/veh

A similar technique and field sheet can be used to measure stopped time delay as well. In this case, the interval counts include only vehicles stopped within the intersection queue area, not those moving within it. No adjustment for acceleration/deceleration delay would be added.

11.5 Closing Comments

Time is one of the key commodities that motorists and other travelers invest in getting from here to there. Travelers most often wish to minimize this investment by making their trips as short as possible. Travel-time and delay studies provide the traffic engineer with data concerning congestion, section travel times, and point delays. Through careful examination, the causes of congestion, excessive travel times, and delays can be determined and traffic engineering measures developed to ameliorate problems.

Speed is the inverse of travel time. While travelers wish to maximize the speed of their trip, they wish to do so consistent with safety. Speed data provide insight into many factors, including safety, and are used to help time traffic signals, set speed limits, and locate signs, and in a variety of other important traffic engineering activities.

References

- 1. *Highway Capacity Manual, 6th Edition: A Guide for Multimodal Mobility Analysis*, Transportation Research Board, National Science Foundation, Washington, D.C., 2016.

- 2. *Highway Capacity Manual, 3rd Edition, Special Report 209*, Transportation Research Board, National Science Foundation, Washington, D.C., 1985.

Problems

1. 11-1. Consider the spot speed data below, collected at a rural highway site under conditions of uncongested flow:

Speed Group (mi/h)	Number of Vehicles Observed (N)
15–20	0
20–25	4
25–30	9
30–35	18
35–40	35
40–45	42
45–50	32
50–55	20
55–60	9
60–65	0

11.2-13 Full Alternative Text

   1. Plot the frequency and cumulative frequency curves for this data.
   2. Determine the median speed, the modal speed, the pace, and the percentage of vehicles in the pact from the curves, and show how each was found.
   3. Compute the mean and standard deviation of the speed distribution.
   4. What are the confidence bounds of the estimate of the true mean speed of the distribution with 95% confidence? With 99.7% confidence?

5. Based on the results of this study, a second is to be conducted to achieve a tolerance of ±0.8 mi/h with 95% confidence. What sample size is required?

6. Can these data be adequately described as “normal”?

2. 11-2. A before-and-after speed study was conducted to determine the effectiveness of a series of rumble strips installed approaching a toll plaza to reduce approach speeds to 40 mi/h.

Item	Before Study	After Study
Average Speed	43.5 mi/h	40.8 mi/h
Standard Deviation	4.8 mi/h	5.3 mi/h
Sample Size	120	108

11.2-14 Full Alternative Text

1. Were the rumble strips effective in reducing average speeds at this location?

2. Were the rumble strips effective in reducing average speeds to 40 mi/h?

3. 11-3. The following data were collected during a delay study on a signalized intersection approach. The cycle length of the signal is 60 s.

Clock Time	Cycle Number	Number of Vehicles in Queue			
		+0 s	+15 s	+30 s	+45 s
9:00 AM	1	3	4	2	4
9:01 AM	2	1	2	3	3
9:02 AM	3	4	3	3	4
9:03 AM	4	2	3	3	4
9:04 AM	5	0	1	2	3

11.2-15 Full Alternative Text

Clock Time	Cycle Number	Number of Vehicles in Queue			
		+0 s	+15 s	+30 s	+45 s
9:05 AM	6	2	1	1	2
9:06 AM	7	4	3	4	3
9:07 AM	8	5	5	6	4
9:08 AM	9	2	3	4	3
9:09 AM	10	0	3	2	2
9:10 AM	11	1	2	3	1
9:11 AM	12	1	0	1	0
9:12 AM	13	2	2	1	2
9:13 AM	14	2	3	2	2
9:14 AM	15	4	3	3	3
$V_T = 435$ vehs $V_{STOP} = 305$ vehs FFS = 35 mi/h					

11.2-16 Full Alternative Text

Erin Blvd	Recorder: XYZ	Summary of 5 Runs		
Checkpoint Number	Cumulative Section Length (mi)	Cumulative Travel Time (min:sec)	Per Section	
			Delay (s)	No. of Stops
1	—	—	—	—
2	1.00	2:05	10	1
3	2.25	4:50	30	1
4	3.50	7:30	25	1
5	4.00	9:10	42	2
6	4.25	10:27	47	1
7	5.00	11:54	14	1

11.2-17 Full Alternative Text

1. Estimate the time spent in queue for the average vehicle.
2. Estimate the average control delay per vehicle on this approach.

4. 11-4. A series of travel time runs are to be made along an arterial section. Tabulate the number of runs required to estimate the overall average travel time with 95% confidence to within ± 2 min, ± 5 min, $\pm$ 10min, for standard deviations of 5, 10, and 15 minutes. Note that a 3×3 table of values is desired.

5. 11-5. The results of a travel time study are summarized in the table that follows. For these data:

   1. Tabulate and graphically present the results of the travel time and delay runs. Show the average travel speed and average running speed for each section.
   2. Note that the number of runs suggested in this problem (11-5) is not necessarily consistent with the results of Problem 11-3. Assuming that *each vehicle* makes five runs, how many test vehicles would be needed to achieve a tolerance of ± 3 mi/h with 95% confidence?

Chapter 12 Highway Traffic Safety: An Overview

12.1 Introduction

In the year 2016, 37,461 people were killed in motor vehicle traffic crashes on U.S. roads and highways in a total of over 6,300,000 police-reported crashes. As police-reported crashes are generally believed to make up only 50% of all crashes occurring, this implies a staggering total of over 12,000,000 crashes for the year. A more complete set of statistics for the year 2015 is shown in Table 12.1 [1]. Complete statistics for 2016 have not yet been released at this writing.

Table 12.1: National Highway Traffic Crash Statistics for 2015

Crash Type	Number of Deaths/Injuries	Number of Crashes*	Number of Vehicles Involved
Fatal	35,485	32,539	55,661
Injury	2,443,000	1,715,000	3,759,000
Property Damage Only (PDO)	NA	4,548,000	7,453,000

* Includes only crashes reported by police.

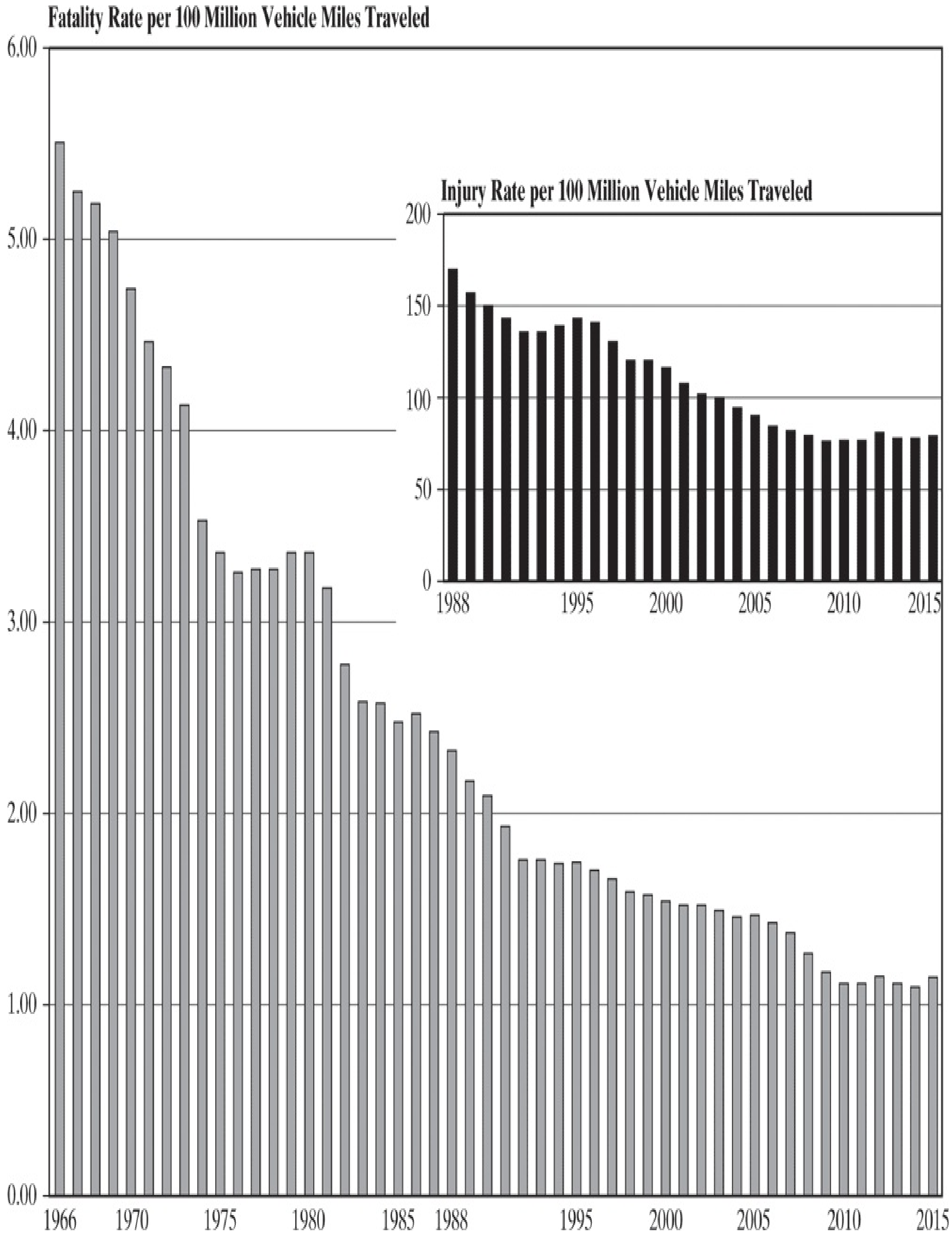
(Source: Compiled from *Traffic Safety Facts 2015* and *Quick Facts 2016*, National Highway Traffic Safety Administration, U.S.D.O.T, Washington, D.C., 2017.)

Table 12.1: Full Alternative Text

To fully appreciate these statistics, some context is needed: More people have been killed in highway crashes in the United States than in all of the wars in which the nation has been involved, from the Revolutionary War through Desert Storm.

Nevertheless, there has been a dramatic downtrend in the fatality *rate* in vehicle crashes, as shown in Figure 12.1. Although there has been an increase in total vehicle miles of travel over the same years, the total number of fatalities has not increased as dramatically, as shown in Figure 12.2 [2]. It is the decreased rate that draws one's attention.

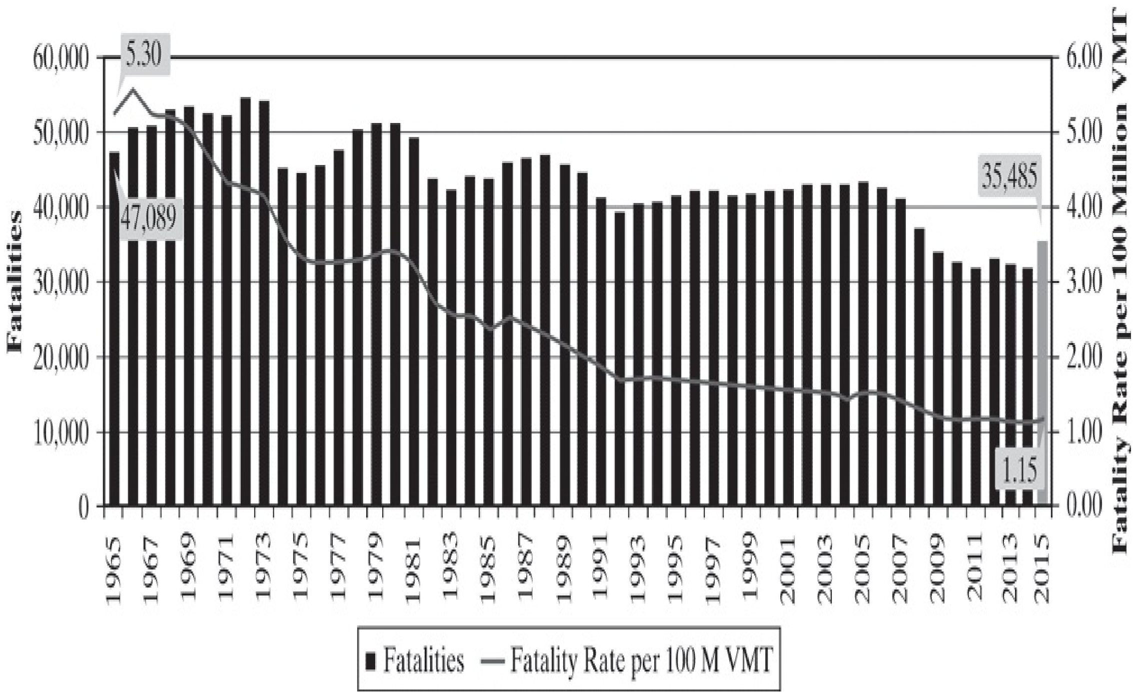
Figure 12.1: Motor Vehicle Fatality and Injury Rates per 100 Million Vehicle-Miles Traveled 1966–2015



(Source: *Traffic Safety Facts 2015*, National Highway Traffic Safety Administration, U.S. Department of Transportation, Washington, D.C., 2017, pg 20.)

Figure 12.1: Full Alternative Text

Figure 12.2: Motor Vehicle Fatalities and Fatality Rates, 1965–2015



(Source: “2015 Motor Vehicle Crashes: An Overview,” *Traffic Safety Facts – Research Note*, National Highway Traffic Safety Administration, U.S. Department of Transportation, Washington, D.C., August 2016, Figure 2.)

Figure 12.2: Full Alternative Text

The underlying factors for the decline are generally acknowledged to be as follows:

1. Vehicle design, driven by such improvements as seat belts, child seats, airbags (side as well as driver/passenger), crumple zones;

2. Roadway improvements, including barriers, signage and lighting, and basic design principles applied to new roads and to rehabilitation; and

3. Education and enforcement, leading to greater usage of seat belts,

common usage of child seats, graduated licenses, and behavior changes.

One interesting statistic is that from 1982 to 2015, the percentage of alcohol-impaired driving fatalities declined from 48% to 29% [1].

At this writing, there has been a noticeable increase in both the fatality rate and the number of fatalities from 2014 to 2016. Fatalities rose from 30,056 in 2014 to 32,485 in 2015 and 34,439 in 2016, an increase of 14.6% in two years. Over the same period, the fatality rate (deaths per 100 million vehicle miles) rose from 1.08 to 1.18, an increase of 9.3% [1,3].

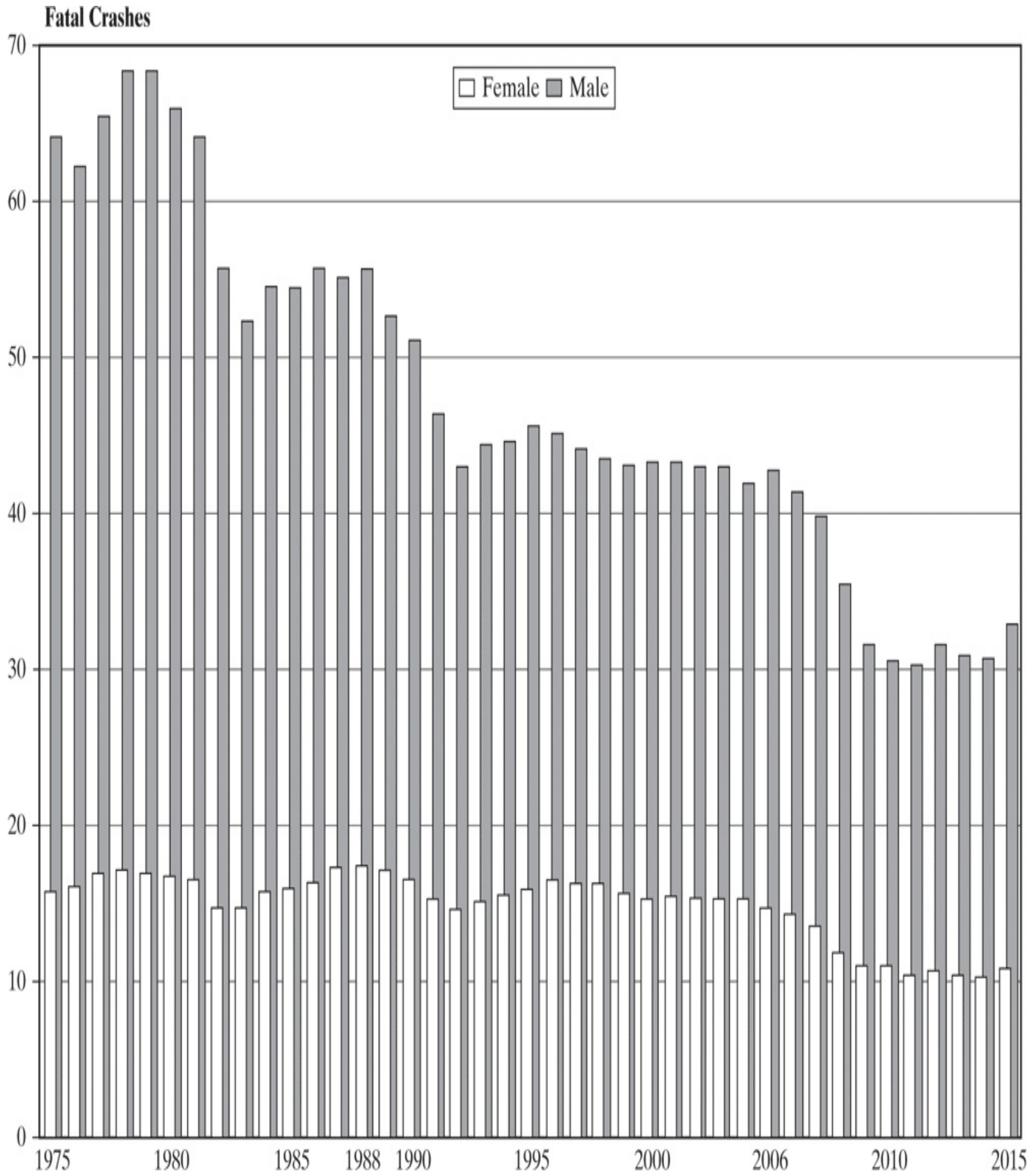
Is this an early warning sign of the effects of more distracted driving? Is it simply a statistical fluctuation? In 2018, this issue is being comprehensively researched to identify a cause for the sudden shift in a long-term trend.

It is however clear that much has been achieved over the decades and that a systematic focus on safety has had—and can continue to have—substantial benefits. The underlying approach has been fact-based decision making by the National Highway Traffic Safety Administration (NHTSA) and the Federal Motor Carrier Safety Administration (FMCSA), both units of United States Department of Transportation (USDOT).

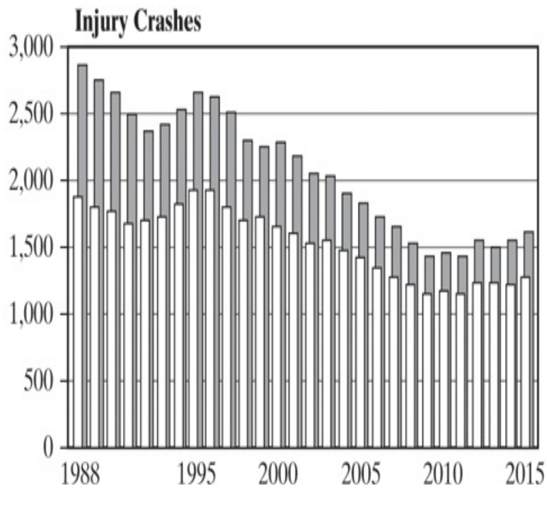
There have been long-term studies in NHTSA's Fatality Analysis Reporting System (FARS) and its National Automotive Sampling System that are included in its General Estimates System (GES) and its Crashworthiness Data System (CDS). GES was based upon a national sample of police accident reports; CDS was based upon passenger vehicle crashes, investigating injury mechanisms to identify potential improvements in vehicle design. The GES and CDS programs have been supplanted as part of NHTSA's Data Modernization Program, with the national sampling framework being redefined as part of that modernization. The current programs are the Crash Report Sampling System and the Crash Investigation Sampling System.

The available data indeed show overall downtrends, but still result in large numbers of fatalities (and serious injuries) because of the increase in vehicle miles traveled over the same years. And there are persistent problems to be addressed, traceable to driver characteristics and behavior. For instance, Figure 12.3 shows driver involvement rates by gender and crash severity. Clearly, there are gender-based differences.

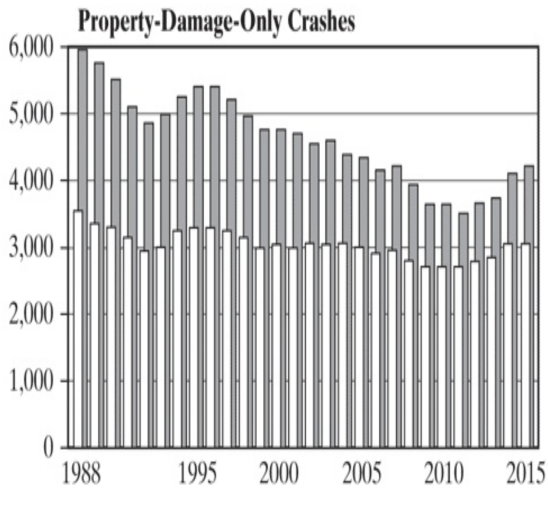
Figure 12.3: Driver Involvement Rates per 100,000 Licensed Drivers 16 Years and Older by Gender and Crash Severity, 1975–2015



Fatal Crashes
Female
Male
70
60
50
40
30
20
10
0
1975
1980
1985
1988
1990
1995
2000
2006
2010
2015



Injury Crashes
3,000
2,500
2,000
1,500
1,000
500
0
1988
1995
2000
2005
2010
2015



Property-Damage-Only Crashes
6,000
5,000
4,000
3,000
2,000
1,000
0
1988
1995
2000
2005
2010
2015

(Source: *Traffic Safety Facts 2015*, National Highway Traffic Safety Administration, U.S. Department of Transportation, Washington, D.C., 2017, pg 27.)

Figure 12.3: Full Alternative Text

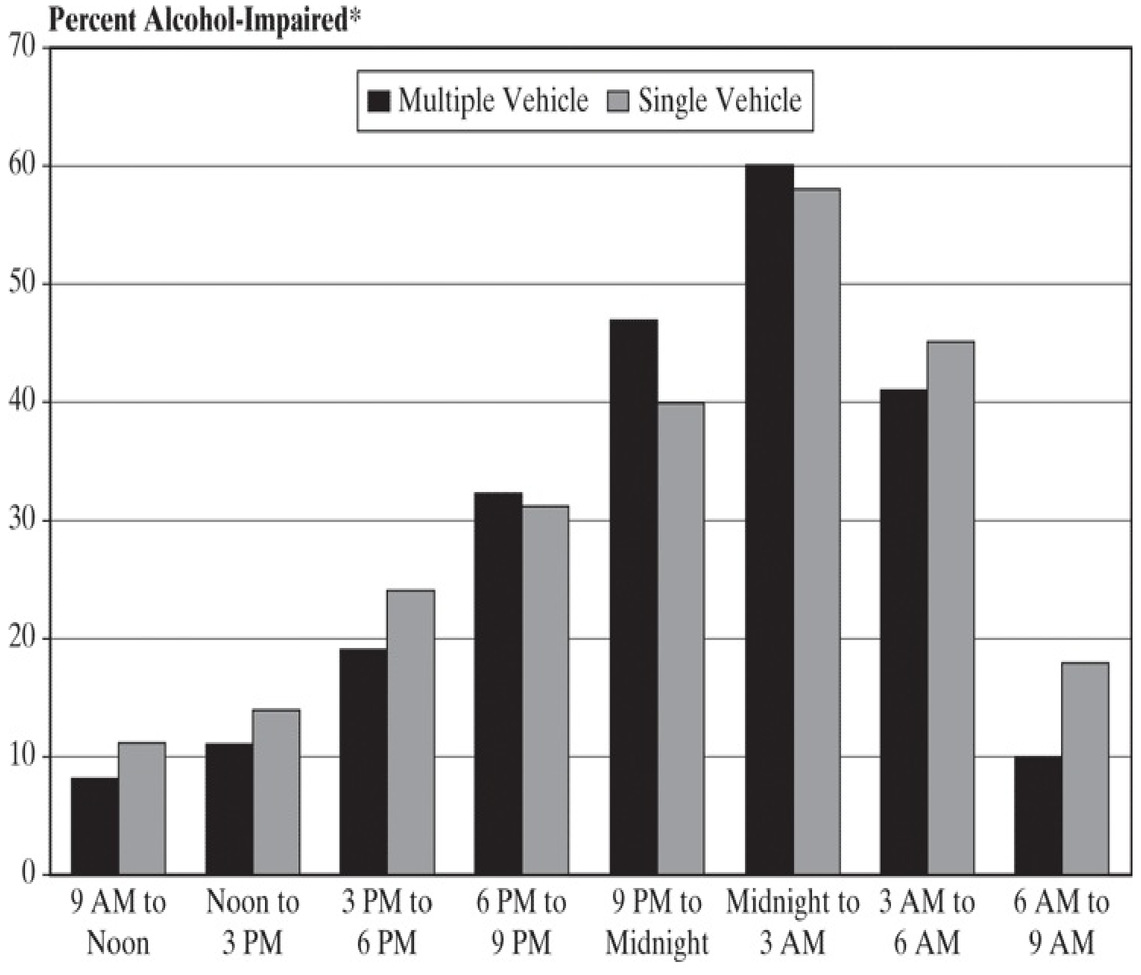
The data and the literature also show variations in crash patterns with respect to the following:

- Time of day, and weekday versus weekend
- Weather
- Light condition
- Roadway type, control type (if any), roadway condition
- Nature of encounter (angle, rear end, head-on; with fixed object; without fixed object)
- Vehicle type(s) involved
- Driver condition (alcohol, drugs, none) and behavior (distracted, not)

The data and the literature also show differences in the ability to capture certain data. For instance, on-scene detailed investigations tend to cite distracted driving factors more than standard police reports. Gaining insight from data requires some detailed insight into the design of the forms used, and the training (and priorities) of the collector. First responders, for instance, have multiple duties: control the scene for safe operations and for addressing injuries, mitigate risk, and collect available facts.

Figure 12.4 shows—for a range of time periods—the percentage of persons killed in alcohol-impaired driving crashes *for the given time period*. For those fatality situations between midnight and 3 AM, approximately 60% of the situations involved alcohol impairment. In the 9 AM to 3 PM time periods, it was about 10% in each case.

Figure 12.4: Percent of Persons Killed in Alcohol-Impaired Driving Crashes by Time of Day



(Source: *Traffic Safety Facts 2015*, National Highway Traffic Safety Administration, U.S. Department of Transportation, Washington, D.C., 2017, pg 111.)

Figure 12.4: Full Alternative Text

While these data lead to issues requiring detailed research, they do not necessarily give all of the information needed. For example, Figure 12.3 does not include information concerning how many vehicle-miles are driven by men and women, respectively. Figure 12.4 might merely reflect times of day people are most likely to be out drinking and driving.

12.2 Current and Emerging Priorities

First and foremost, we must set this section in context. One must never lose one's focus on the basics:

- The first section of this chapter provided an overview based upon long-term data programs and careful attention to the data collected. That attention led to mandates for improved vehicle designs, more attention to roadway design features, and education and enforcement priorities.
- Later sections will identify some of the basic tools and techniques that are used routinely in traffic engineering practice:
  - Crash diagrams;
  - Condition diagrams;
  - Metrics based upon exposure, rates, total numbers, actual events, and risk assessment; and
  - Related statistical analyses.

At the same time, there are clearly current themes of importance to both those entering the field and those practicing the profession. They are part of the dialog, and even the fabric, of modern traffic engineering. The state of the art and of the practice is evolving rapidly, and requires the awareness of the traffic engineer.

12.2.1 Crashes, Not Accidents

Although "accidents" is a word used in common parlance, the professional dialog is changing to an emphasis on "crashes," not accidents. The distinction is very fundamental: Accidents are things that inevitably happen and foster a dialog (and mind-set) on acceptable levels of such events. “Crashes” are events that can be avoided by design, technology, and programmatic efforts. One can more easily aspire to reduce crash events to zero, perhaps class by class, but zero nonetheless. This chapter will use “accident” in some cases for historical reasons (in context) or for certain terminology in common practice, such as “police accident reports.”

12.2.2 Pedestrians and Bicyclists

Historically, traffic safety has focused on harm to vehicle occupants by encounters with other vehicles, fixed objects, and nonfixed objects. In some discussions, pedestrians were listed as “nonfixed objects” with which vehicles collided. The harm to pedestrians was certainly acknowledged, but often recognizing that the total number of pedestrian fatalities was—relatively speaking—a fraction of vehicle-related fatalities. There has been a clear evolution to recognizing that street space serves multiple users and that each of the modes has its own rights to the space, including being safe from the other modes.

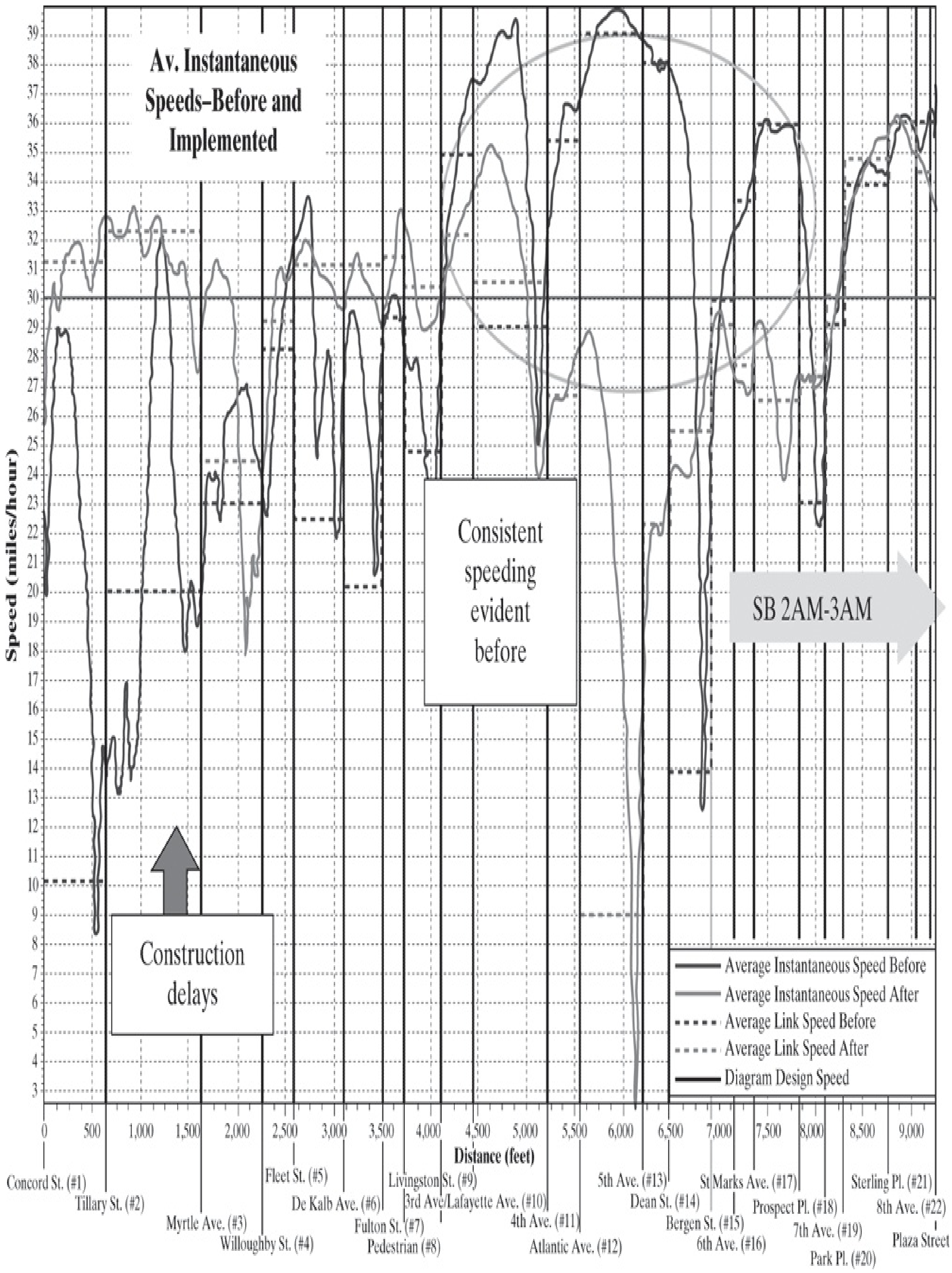
12.2.3 Traffic Calming

A modern traffic engineer is called upon to balance the modes and provide service to those desired by the public and by planners focused on the ambience of the urban environment. Mobility is not just moving more vehicles, but providing for multiple uses in balance. Quality of trip may take on more importance than speed of the trip. Safety emerges as an objective that should govern decisions more than in past years.

Urban space is designed for safe, concurrent movement of several travel modes, each with its own individual design features built into the urban environment. Consider an interesting case: Traffic signal progression is historically looked upon as a means of maximizing vehicle flow with minimum delay. But one jurisdiction looked at it differently, initially in the off-peak hours: Less green on the main street forces vehicle platooning; platoons govern speeding opportunities; less main street green means more pedestrian crossing time across wider main streets; and using progressions to *limit* distances that can be traveled nonstop can result in less speeding and keeping coherent platoons of vehicles. Figure 12.5 shows the before–

after speed profiles on one such arterial.

Figure 12.5: Speed Control by Signal Progression, An Example in Downtown Brooklyn, N.Y.



(Source: KLD Engineering, P.C.)

Figure 12.5: Full Alternative Text

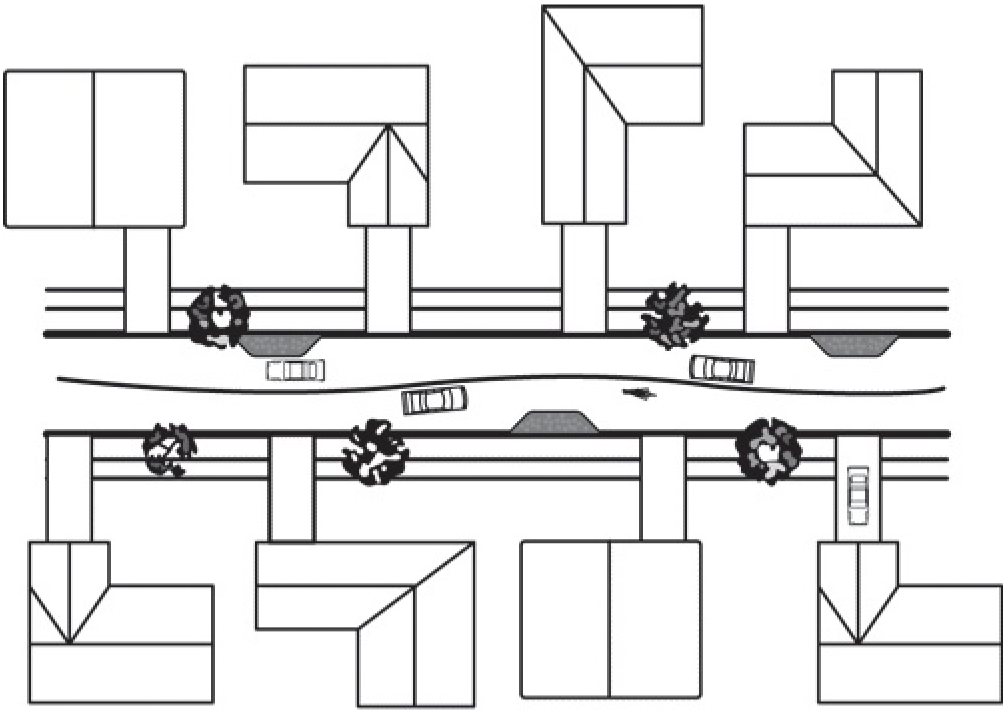
Many streets do not need all the vehicular capacity provided in earlier, standard designs. A network’s capacity is often controlled by a limited number of critical (or key) intersections; most intersections can cope with capacities suited to feeding the key locations. Such a mind-set opens the door to reallocating street space to other modes and to design features that shorten the pedestrian crossing distance and indeed better define the pedestrian space.

The most comprehensive source of information on traffic calming and traffic calming devices is the Federal Highway Administration's *Traffic Calming ePrimer*, available online. It is regularly updated to include the latest information. It describes a wide variety of devices and approaches involved in traffic calming, along with general guidelines for when, where, and how to implement these as part of an overall traffic calming program [4]. A partial list of devices used to implement traffic calming includes the following:

- Lateral shifts in alignment
- Chicanes
- Realigned intersections
- Traffic circles and roundabouts
- Speed bumps, humps, and tables
- Raised pedestrian crosswalks
- Chokers and other ways to narrow crossing distance for pedestrians
- Street closures and diverters
- Road diets (allocating roadway space to alternative modes, such as bicycles and transit)

A few of these are illustrated in Figure 12.6.

Figure 12.6: Traffic Calming Devices Illustrated



(a) Chicane Used to Slow Traffic

(a) Chicane Used to Slow Traffic

12.2-2 Full Alternative Text

(b) Chicane Illustrated

(b) Chicane Illustrated

12.2-2 Full Alternative Text

(c) Traffic Circle Illustrated

(c) Traffic Circle Illustrated

12.2-2 Full Alternative Text

(d) Curb Extension Reduces Crossing Distance

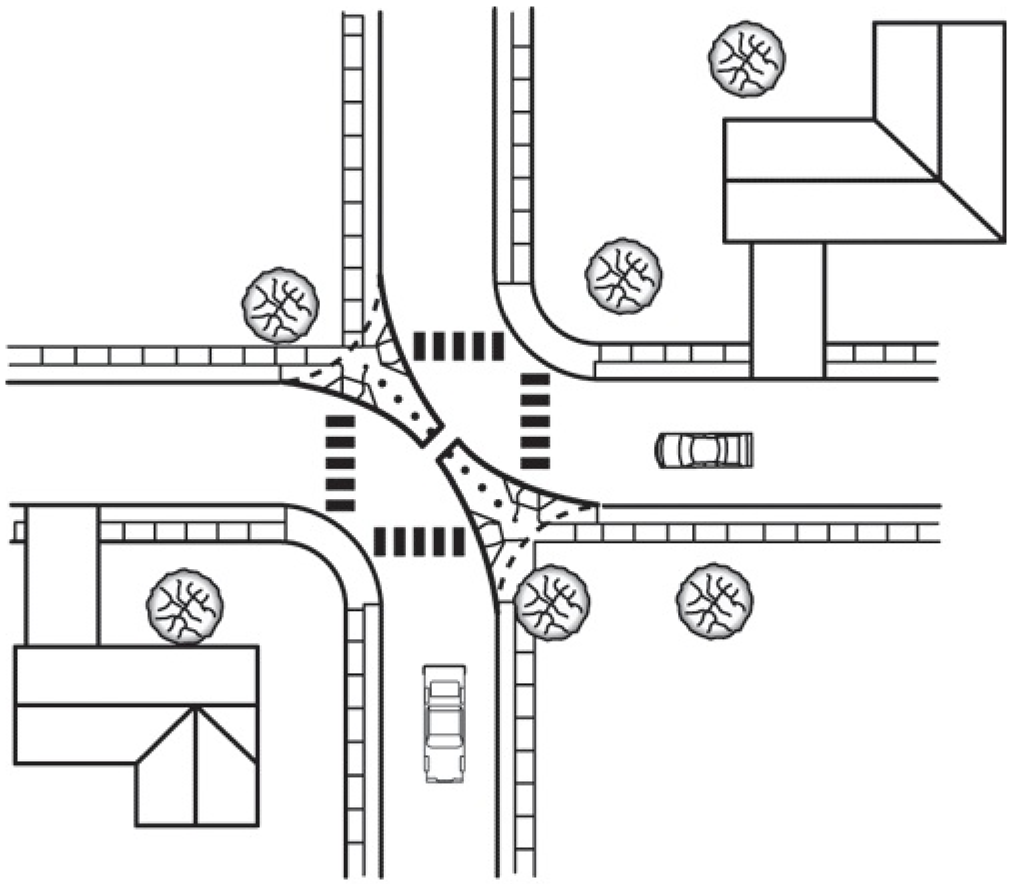
(d) Curb Extension Reduces Crossing Distance

12.2-2 Full Alternative Text

(e) Road Diet Illustrated

(e) Road Diet Illustrated

12.2-2 Full Alternative Text



(f) Diagonal Diverter

(f) Diagonal Diverter

Source: *Traffic Calming ePrimer*, Federal Highway Administration, U.S. Department of Transportation, Washington, D.C., 2017, Figures 3.5.1, 3.5.3, 3.7.4, 3.16.1, and 3.16.2.)

12.2-2 Full Alternative Text

Figures 12.6(a) and (b) illustrate chicanes, which are placed to force a vehicle to divert from a straight path and reduce speed. Figure 12.6(c) is a small traffic circle, which again forces drivers from a straight path and to negotiate the intersection at lower speed. In Figure 12.6(d), the extended curb shortens the pedestrian crossing distance, as well as emphasizes its presence to drivers. Figure 12.6(e) shows a reduction in vehicular lanes to accommodate a bicycle lane. Figure 12.6(f) illustrates a directional divider that prevents through movements at the intersection. These can be used to retrofit rectangular grid networks to provide traffic calming.

These devices help improve the travel environment for pedestrians and bicyclists. Because they also disrupt cars, trucks, and even buses, they should be carefully considered only as part of an overall traffic calming program where all impacts, positive and negative, can be considered.

12.2.4 Distracted Driving

There are clear prohibitions on texting while driving, as well as requirements for “hands free” cell phone use while driving. But there are questions as to whether these steps are enough, given the dramatic increases in high-profile incidents related to drivers being distracted. Human factors make it clear that there are limits to the number of concurrent tasks that can be done, and there are risk assessments that address the increase in task performance needs based upon traffic volume, weather, lighting, and other factors. At the same time, technology is providing a range of interruptions, as well as more opportunities for drivers to multitask and receive spoken messages.

12.2.5 Vision Zero

In 1997, Sweden’s parliament approved “Vision Zero,” a commitment to road traffic safety aimed at *no* fatalities or serious injuries. An overview of the international spread of the effort is provided in [5]. The Federal Highway Administration (FHWA) of the USDOT emphasizes a data-driven, interdisciplinary approach targeting areas for improvement, using proven countermeasures, and building on “the 4Es” (education, enforcement, engineering, and emergency medical/trauma services) [6]. The National Safety Council has partnered with NHTSA, FHWA, and FMCSA in a “Road to Zero” effort [7]. Many local jurisdictions have their own variants of the concept.

12.2.6 The Connected Vehicle

There is much attention to the advanced technology that allows vehicles to communicate with each other (V2V) and with the infrastructure (V2I) using a variety of sensors for early warnings, automatic responses, and crash avoidance. This extends to sensing pedestrians and other modes.

- In the prior edition of this text, the connected vehicle was a future concept. At this writing, market forces are driving vehicle manufacturers to accelerated research, partnering, and acquisitions, and a great focus on bringing products to market within a very few years.
- In the background, the federal agencies in the United States, ITS America (ITSA), and comparable organizations throughout the world have been addressing communication protocols and standards.
- Clearly, the connected vehicle concept is a means of improving safety and moving toward the vision-zero goal. But the pace of bringing real products to market is so rapid, and the competitive forces so strong, that the reader is well advised to read current literature, check market penetration, and forecasts. Forecasts recorded in this text will become dated even during the lifetime of the present edition of the text.

12.2.7 The Driverless Vehicle

This is closely related to the connected vehicle concept, but with the singular emphasis of the driving task being automated. There are experimental driverless vehicles on the road as of this writing, and various states are enacting legislation to allow testing. Auto manufacturers are focused on bringing products to market. Businesses are looking to the economics of goods movement over long and short distances, using driverless vehicles. The reader is well advised to follow trends and forecasts, and to factor into his/her thinking how such vehicles will affect safety goals, vision zero, and even road capacity.

12.2.8 Smartphone Apps

There is a rapidly emerging area of smartphone apps that provide navigation to drivers and to pedestrians and bus and train arrival times and schedules, and can provide warnings based upon V2I connected vehicle technology. At the same time, smartphones are looked upon as contributors to distracted driving and distracted walking. This is another area of attention for the traffic engineer.

12.2.9 Data-Rich Environment

Other chapters (for example, Chapter 9: Traffic Studies and Chapter 8: Intelligent Transportation Systems) make note of the virtual flood of information available to today's traffic engineer and to the public. There is virtual flood of information available to today's traffic engineer and to the public. These data can be harnessed for traffic safety studies and evaluations, and for detailed analysis. It can be used to prioritize areas needing attention. Above all, the truism that "a picture is worth a thousand words" is often quite true, given the amount of data. Data visualization becomes an especially important tool.

Figure 12.7 shows 2017 data (through May 31) on traffic-related fatalities for pedestrians and for vehicle occupants in New York City [8], using a web-based tool available on the city's Vision Zero website. Note that for this urban environment, the general national statistics on vehicle occupant vis-à-vis pedestrian fatalities are dramatically reversed: there are more pedestrian fatalities (41) than vehicle occupant fatalities (29) for the period shown.

Figure 12.7: New York City Pedestrian and Vehicle Occupant Fatalities in Traffic Crashes January 1–May 31, 1917

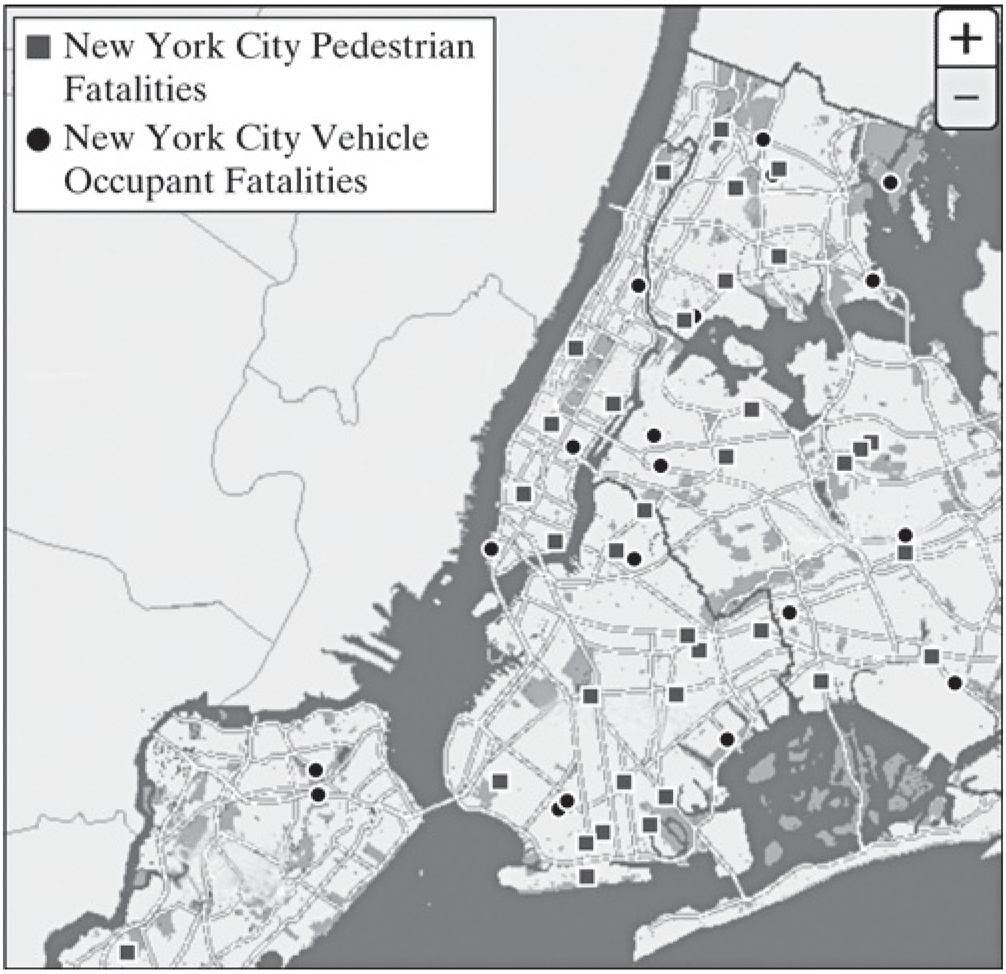


Figure 12.7: Full Alternative Text

12.3 The Highway Safety Manual

The *Highway Safety Manual* (HSM) [9,10] is the product of a massive effort that started in 1999 with a conference session at the Transportation Research Board (TRB) Annual Meeting (January 1999). It led to a workshop involving eight TRB committees and the FHWA to investigate the feasibility of producing such a document (December 1999). It was funded by a sequence of projects under the National Cooperative Highway Research Program (NCHRP) to develop the document, and resulted in its publication by the American Association of State Highway and Traffic Officials (AASHTO) in 2010. The document does not constitute a standard, but contains many useful guidelines. The 2nd edition is scheduled for release in 2019.

The emphasis in the HSM is on a science-based technical approach, tools for quantitative safety analyses, and predictive methods for estimating crash frequency and severity.

There are two terms that are used throughout the HSM and are now part of the vocabulary of traffic safety:

- *Safety performance functions* (SPFs) are equations for estimating expected average crash frequency as a function of traffic volume and roadway characteristics, generally stated for a standard base set of conditions.
- *Crash modification factors* (CMFs) are used to estimate the change in crash frequency or crash severity due to applying a specific treatment or to estimate the effect of conditions not in conformance with the base set used in estimating SPFs.

Table 12.2 shows an illustrative finding from the HSM, showing the potential effects of adding a median on a multilane road. Note that for an urban multilane arterial, the estimated CMF predicts a 22% decrease in all types of injuries, but a 9% increase in noninjury crashes. The professional considers such information in making informed decisions, with cost-effectiveness, priorities, and available funds as other factors.

Table 12.2: Illustrative Effects of Providing a Median on Multilane Road

Treatment	Setting (Road Type)	Traffic Volume	Accident Type (Severity)	CMF	Std. Error
Provide a median	Urban (Arterial Multilane)	Unspecified	All Types (Injury)	0.78	0.02
			All Types (Non-Injury)	1.09	0.02
	Rural (Multilane)		All Types (Injury)	0.88	0.03
			All Types (Non-Injury)	0.82	0.03

(Source: Adapted from (*Illustrative Effects of Providing a Median on Multilane Road*), (2011), by the American Association of State Highway and Transportation Officials, Washington, D.C. U.S.A)

Table 12.2: Full Alternative Text

The HSM can be used for all aspects of safety analysis, including the following project types:

1. Maintenance and operations: Improve safety at existing locations, selecting the best countermeasures
2. Project planning and preliminary engineering: Identify alternatives and evaluate for safety potential and cost-effectiveness
3. System planning: Identify needed projects and prioritize them
4. Design and construction: Implement projects

Methodologies are provided for the analyst to do the following:

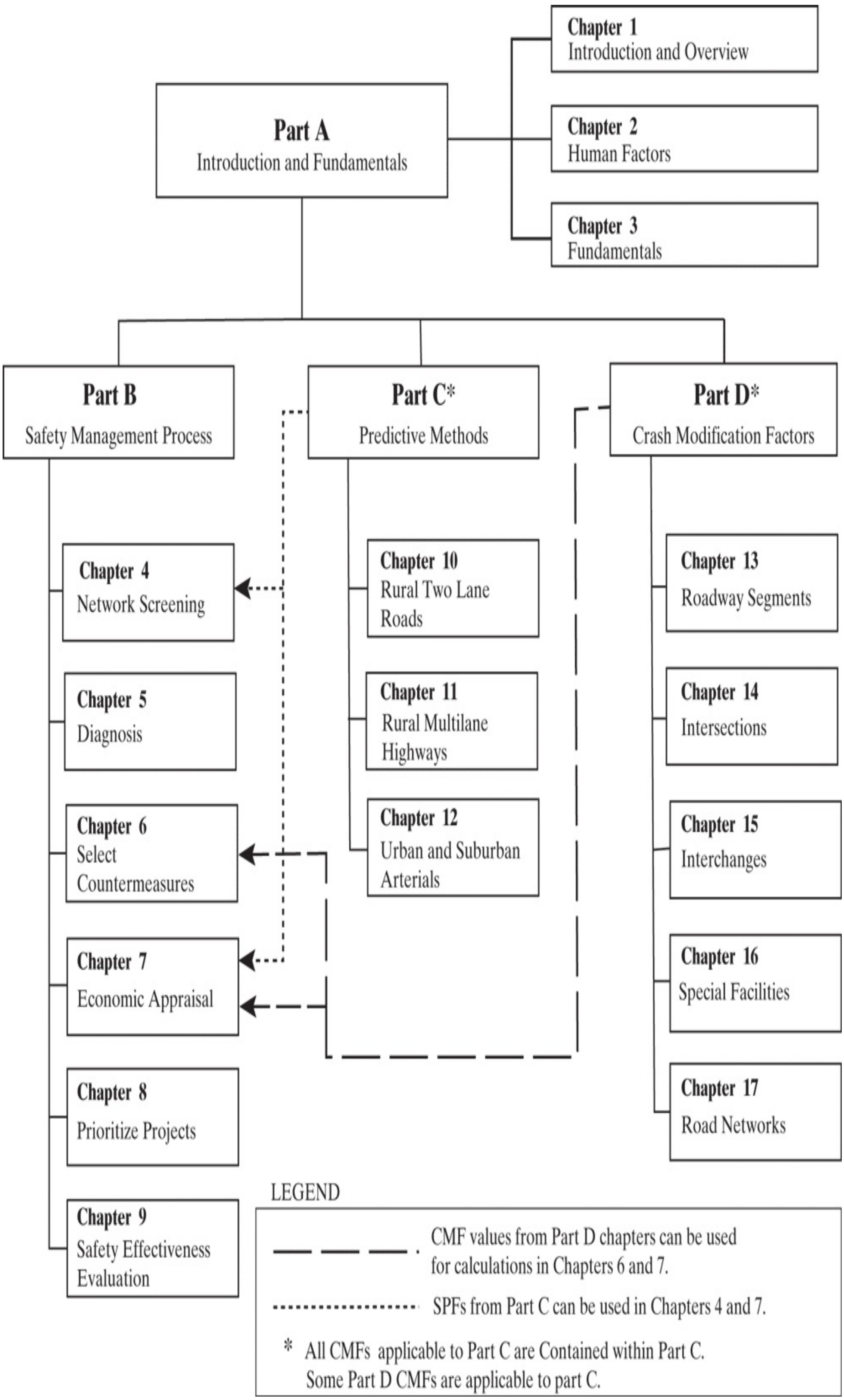
- Estimate the average crash frequency, with or without having historical crash data
- Estimate (or predict) the effectiveness of countermeasure(s)

- Evaluate the effectiveness of an implemented countermeasure

This text cannot possibly cover all of the information provided by the HSM to accomplish. Some examples, however, are included to illustrate the types of analysis that the HSM enables.

Figure 12.8 provides a flowchart for the content of the HSM.

Figure 12.8: Organization of the *Highway Safety Manual*



Part A
Introduction and Fundamentals
Chapter 1
Introduction and Overview
Chapter 2
Human Factors
Chapter 3
Fundamentals
Part B
Safety Management Process
Part C*
Predictive Methods
Part D*
Crash Modification Factors
Chapter 4
Network Screening
Chapter 5
Diagnosis
Chapter 6
Select Countermeasures
Chapter 7
Economic Appraisal
Chapter 8
Prioritize Projects
Chapter 9
Safety Effectiveness Evaluation
Chapter 10
Rural Two Lane Roads
Chapter 11
Rural Multilane Highways
Chapter 12
Urban and Suburban Arterials
Chapter 13
Roadway Segments
Chapter 14
Intersections
Chapter 15
Interchanges
Chapter 16
Special Facilities
Chapter 17
Road Networks
LEGEND
CMF values from Part D chapters can be used for calculations in Chapters 6 and 7.
SPFs from Part C can be used in Chapters 4 and 7.
* All CMFs applicable to Part C are Contained within Part C. Some Part D CMFs are applicable to part C.

(Source: Adapted from (*Highway Safety Manual*), (2010), by the American Association of State Highway and Transportation Officials, Washington, D.C. U.S.A)

Figure 12.8: Full Alternative Text

12.3.1 Steps for Performing an Analysis

This section will describe the general steps needed to perform a safety analysis of a specific site (roadway and/or intersections), as per the HSM.

1. Step 1. Gather data

   The first step is to gather all of the data needed to apply the safety prediction methods. A complementary publication to the HSM gives a detailed explanation of the data needed [11]. There are three categories of data needed: site characteristics, traffic volumes, and crash history data.

   Site characteristic data are needed for both roadway segments and intersections.

   Roadway data include area type, length of segment, detailed roadway cross-section description, horizontal and vertical alignment, driveways, roadside conditions, and lighting.

   Intersection data include a detailed geometry (e.g., lane types, width), type of traffic control (e.g., stop or signal, phasing, right-turn on red), enforcement devices (e.g., red light cameras), angle of intersection, sight distance, terrain, and lighting.

   More detailed descriptions of each of these general categories of data can be found in Ref [12].

For roadway segments, average annual daily traffic (AADT) volumes are needed. For intersections, average daily traffic volumes of the intersecting roads are needed.

Crash data needed include the following for each year of data: location, severity (fatal, injury, property damage only), distance from intersection, intersection- related or not intersection-related.

2. Step 2. Do a diagnosis of the location

   A diagnosis of the site is done in order to determine the types of crashes that occur and the safety issues existing at the specific site. The diagnosis should identify contributing causes to the crashes. It is recommended that three to five years of crash data should be analyzed.

   Detailed information is required for the following areas:

   1. Occurrence of crashes at the location in question
   2. Environmental and physical conditions existing at the location
   3. Traffic volumes

   The analysis of this information should identify the environmental and physical conditions that potentially or actually contribute to the observed occurrence of crashes.

   Environmental and physical conditions are established by a thorough site investigation conducted by appropriate field personnel. Two typical tools for displaying this information include

   1. Crash diagram
   2. Condition diagram

   Details on how to construct and present these key diagrams are given later in the chapter.

3. Step 3. Interpret data and select countermeasures

   This overview chapter cannot fully discuss and present all types of crash site analyses. The objective in analyzing collision, condition, and traffic data is straightforward: find contributing causes to the observed crashes among the design, control, operational, and environmental features. Doing so involves virtually all of the traffic engineer's knowledge, including statistical assessments, experience, and the application of professional judgment.

   The HSM lists the most common factors that are responsible for various crash types, and the NCHRP Report 500A, *Guidance for Implementation of the AASHTO Strategic Highway Safety Plan* [12], can be used as a reference for a more detailed discussion on contributing factors. The HSM additionally provides lists of countermeasures that may be appropriate based on the type of site, types of crashes, and safety concerns that were the result of the diagnosis in the previous step.

   Some discussion of effective countermeasures is included later in this chapter.

4. Step 4. Economic analysis of the countermeasures under consideration

   A benefit–cost or a cost-effectiveness analysis should be done to determine if implementation is economically justifiable. The HSM uses change in crash frequency and severity for quantifying the benefits of the countermeasures being considered and provides a predictive methodology for determining this change. There are many other project benefits, however, that the analyst may consider to evaluate the economic benefits of a countermeasure(s). The AASHTO Redbook, *A Manual of User Benefit Analysis for Highway and Bus-Transit Improvements*, is one reference that can provide guidance for evaluating other benefits. For example, there may be fuel consumption, noise, travel time, and/or other benefits to a particular countermeasure that may be important to consider.

5. Step 5. Evaluate safety effectiveness

   The safety effectiveness of specific countermeasure(s) is evaluated through the change in crash severity and frequency, as well as evaluating the changes in relation to how well funds are being used. Before–after crash statistics are compared. Care must be taken when using observed crash data to adjust for bias due to using a few years of data that may capture a random increase in crashes. The effect of using simple observational data without considering the randomness of crash data and how to avoid this error, called regression to the mean (RTM) bias, is discussed in more detail later in the chapter. But simply consider that a few years of low crashes may follow a few years of high crashes and vice versa routinely, without any changes being implemented. Thus, if your before data happen to be a random period of high-crash frequency, the change in crash frequency may not be due to the treatment.

   One of the effectiveness measures provided in the HSM is CMF. A CMF is the ratio of after data to before data. Within the HSM are tables of CMFs defined for a variety of countermeasures. Table 12.2, presented earlier, is a sample CMF table based on the effect of providing a median on multilane roads.

   The CMFs in the HSM were determined from the best available research. For most CMFs, standard errors are also provided to allow the analyst to calculate confidence bounds on the results.

   For countermeasures not included in the HSM, the CMF Clearinghouse is a continuously updated website database of countermeasures and accompanying CMFs [13].

12.3.2 System Planning

For system planning, additional steps are needed.

The first step would be to examine the network (network screening) to determine the high-crash locations at which the safety focus may most be needed. The HSM provides algorithms for screening the network to find the high-crash locations, but also the locations that have the most potential for improvement.

After the sites to be studied are decided upon and steps 1–5 are completed for each location, the HSM gives guidance on how to rank projects in order to prioritize implementation. Projects can be ranked on cost-effectiveness, incremental benefit–cost ratio, safety benefits, and reduced severity of crashes, among others.

12.3.3 The HSM Predictive Method for Calculating Predicted and Expected Average Crash Frequency

The HSM provides a methodology for estimating the predicted and expected average crash frequency in various situations. The predictive crash frequency method is used to estimate the average crash frequency when observed historical crash data are not available, for past or future years, and to improve the statistical reliability when using historical crash data. When historical crash data are available, the expected crash frequency is found by combining the observed historical crash data and the HSM predicted average crash frequency by weighting the values using an Empirical Bayes (EB) method. The EB method gives a more statistically reliable estimate of average crash frequency than the predicted or observed number alone.

To apply the predictive method, the roadway is divided into segments and intersections. Each segment and intersection is analyzed separately. If the analysis is of more than one intersection, the expected crash values of the segments and the intersections are added together.

The predicted average crash frequency (crashes/year) is found using regression models, called SPFs. The SPFs were calibrated for specific base conditions using data from similar sites. Inputs to the regression models include AADT, geometry, and control type, among others. The SPF predicted value is then adjusted for local conditions.

Unique SPF equations are provided for the following facilities:

- Rural two-lane highways
- Rural multilane highways
- Urban and suburban arterials

The value from the SPF is then adjusted for local conditions using CMFs and a calibration factor. CMFs adjust for site conditions that differ from the base conditions under which the SPFs were developed. The calibration factor adjusts for local conditions, such as climate, weather, and driver population.

The general form of the equation for the predicted average crash frequency is shown in Equation 12-1.

$$N_{pred} = N_{SPF} \times \prod_{i=1}^{n} CMF_i \times c \quad [12\text{-}1]$$

where:

N_{pred} = predicted average crash frequency, N_{SPF} = predicted crash frequency

12.3.4 An Overview of the HSM Models for Urban Intersections

The HSM contains detailed methodologies for all types of facilities and situations, and it would be impossible to even provide a meaningful overview herein. To illustrate how the general format is implemented, this section will provide some detail on predicting average crash frequency for an urban intersection.

Table 12.3 lists the input data needed for an urban intersection.

Table 12.3: Data Needs for Signalized Intersection *HSM* Analysis

Crash data history for five years, if available
All approach AADTs
Number of intersection legs
Type of intersection control
Phasing
Presence of RTOR (right turn on red)
Presence of red-light camera
Presence of left-turn lanes
Presence of right-turn lanes
Presence of lighting
Values for VEH-PED and VEH-BIKE Predictions
Pedestrian volumes
Bus stops within 1,000 ft of the intersection
Existence of schools within 1,000 ft of the intersection
Number of alcohol sales establishments within 1,000 sq ft of the intersection

Table 12.3: Full Alternative Text

The HSM provides methodologies to estimate crash frequencies for the following four types of intersections found on urban arterials:

- Three-leg intersections with STOP control on the minor approach (3ST)
- Three-leg signalized intersections (3SG)
- Four-leg intersections with STOP control on the minor-road

approaches (4ST)

- Four-leg signalized intersections (4SG)

For the purposes of this text, details are presented for four-leg signalized intersections (4SG) only. For other types of intersections, the HSM must be consulted directly.

Crash frequencies for all types of intersections are predicted for four types of crashes, each of which has unique models for application:

- Multivehicle crashes
- Single-vehicle crashes
- Vehicle–pedestrian crashes
- Vehicle–bicycle crashes

In addition, within each type of crash, total crashes, injury and fatality crashes, and property damage only (PDO) crashes are separately estimated.

SPFs were calibrated for the following base conditions for intersections on urban and suburban arterials:

- No left- or right-turn lanes
- Only permissive phasing
- Right-turn on red permitted
- No red-light cameras or other automated enforcement
- No bus stops, schools, or alcohol establishments within 1,000 feet of the intersection
- No intersection lighting

In general terms, the total predicted crash frequency at an urban intersection is estimated as:

$N_{pred,int}=c_i\ (N_{bi}+N_{pedi}+N_{bikei})$ [12-2]

where:

$N_{pred,int}$=predicted crash frequency for the sub-
ject intersection for the given year,c_i=calibration factor for intersections for
−pedestrian crashes forthe intersection for the given year, andN_{bikei}=SPF f
−pedestrian crashes forthe intersection for the given year.

Multivehicle Collisions at a 4SG Intersection

The SPF for multivehicle collisions at a 4SG intersection is estimated as:

$N_{bimv}=\exp\ [a+b\ \ln(AADT_{maj})+c\ \ln(AADT_{min})]$ [12-3]

where:

N_{bimv}=SPF (base) for multivehicle crashes forthe intersection for the giver

Table 12.4 shows the calibration coefficients for use in Equation 12-3.

Table 12.4: Calibration Coefficients for Equation 12-3 for a 4SG Intersection

Type of Crash	a	b	c
Total	−10.99	1.07	0.23
Injury & Fatal	−13.14	1.18	0.22
PDO	−11.02	1.02	0.24

(Source: Adapted from (*Highway Safety Manual*), (2010), by the American Association of State Highway and Transportation

Officials, Washington, D.C. U.S.A)

Table 12.4: Full Alternative Text

In general, both the total crash frequency and the separate frequencies of injury + fatal and PDO crashes would be estimated, all using Equation 12-3 with the appropriate regression coefficients taken from Table 12.4. When this is done, however, there is no guarantee that the sum of the injury + fatal and PDO crashes will add up to the predicted total crashes. Equations 12-4 and 12-5 are then used to adjust the injury + fatal and PDO crashes to ensure that they do add to the predicted total.

Nbimv(FI)=Nbimv(TOTAL) [N'bimv(FI)N'bimv(FI)+N'bimv(PDO)] [12-4]
Nbimv(PDO)=Nbimv(TOTAL)−Nbimv(FI) [12-5]

where:

FI=fatal + injury,PDO=property damage only, andPrime(′)=initial estimates

Single-Vehicle Collisions at a 4SG Intersection

Single-vehicle collisions are estimated using Equation 12-6, which is identical to Equation 12-3, except that it predicts crash frequency for single vehicles (N_{bisv}) and uses the regression coefficients of Table 12.5.

Table 12.5: Calibration Coefficients for Equation 12-6 for a 4SG Intersection

Type of Crash	a	b	c
Total	−10.21	0.68	0.27
Injury + Fatal	−9.25	0.43	0.29
PDO	−11.34	0.78	0.25

(Source: Adapted from (*Highway Safety Manual*, 1st Edition), (2010), by the American Association of State Highway and Transportation Officials, Washington, D.C. U.S.A)

Table 12.5: Full Alternative Text

$$N_{bisv}=\exp\left[a+b\ln(AADT_{maj})+c\ln(AADT_{min})\right] \quad [12\text{-}6]$$

where all variables have been previously defined.

As with multivehicle crashes, there is no guarantee that the sum of injury + fatal and PDO crashes adds up to the predicted total. Equations 12-4 and 12-5 are employed to correct for this, with single-vehicle values replacing multivehicle values.

Vehicle–Pedestrian Crashes at 4SG Intersections

The frequency of vehicle–pedestrian crashes is estimated as follows:

$$N_{pedbase}=\exp\left[a+b\times\ln(AADT_{total})+c\times\ln\left(\frac{AADT_{minor}}{AADT_{major}}\right)+d\times\ln \quad [12\text{-}7]$$

where:

$N_{pedbase}$=predicted average crash frequency for pedestrian −vehicle crashes, $AADT_{total}$=average annual daily traffic entering the inters

Coefficients for use with Equation 12-7 are shown in Table 12.6.

Table 12.6: Calibration Coefficients for Equation 12-7 for 4SG Intersections

Coefficient	Value
a	−9.53
b	0.40
c	0.26
d	0.45
e	0.04

(Source: Adapted from (*Highway Safety Manual*, 1st Edition), (2010), by the American Association of State Highway and Transportation Officials, Washington, D.C. U.S.A)

Table 12.6: Full Alternative Text

If precise counts of total pedestrian crossings at the intersection are not available, the default values shown in Table 12.7 may be used.

Table 12.7: Default Values for Pedestrian Volume for Use in Equation 12-7

General Level of Pedestrian Activity	Default Value for PedVol
high	3,200
medium – high	1,500
medium	700
medium – low	240
low	50

(Source: Adapted from (*Highway Safety Manual*, 1st Edition), (2010), by the American Association of State Highway and Transportation Officials, Washington, D.C. U.S.A.)

Table 12.7: Full Alternative Text

Vehicle–Bicycle Collisions at 4SG Intersections

The HSM is far less detailed concerning the estimation of vehicle–bicycle crashes for all types of facilities and locations than it is for multivehicle, single-vehicle, or vehicle– pedestrian crashes. The frequency of such collisions is based upon the SPFs for multivehicle and single-vehicle crashes, which would be estimated as discussed in previous sections.

In general, the frequency of vehicle–bicycle crashes is estimated as follows:

$$N_{bikei}=(N_{pred,mv}+N_{pred,sv})\times f_{bikei} \quad [12\text{-}8]$$

where:

N_{bikei}=predicted number of vehicle −bicycle crashesfor the intersection for the chosen year,$N_{pred,mv}$=predicte vehicle crash frequency(after all CMFs are applied), andf_{bike}=bike crash a

Crash Modification Factors for 4SG Intersections

There are six CMFs that affect multivehicle and single- vehicle crashes at an intersection, and three that affect vehicle–pedestrian accidents. Table 12.8 summarizes these.

Table 12.8: Crash Modification Factors for 4SG Intersections

Type of Crash	Crash Modification Factors Applied
Multi-Vehicle and Single-Vehicle	Intersection LT Lanes (CMF_{LT}) Intersection LT Signal Phasing (CMF_{SP}) Intersection RT Lanes (CMF_{RT}) Right-Turn-On-Red (CMF_{RTOR}) Lighting (CMF_{L}) Red-Light Cameras (CMF_{RLC})
Vehicle-Pedestrian Crashes	Bus Stops (CMF_{Bus}) Schools (CMF_{SCH}) Alcohol Sales Establishments (CMF_{ALC})

Table 12.8: Full Alternative Text

Table 12.9 summarizes CMFs for intersection left- and right-turn lanes. Table 12.10 shows CMFs for left-turn signal phasing.

Table 12.9: CMFs for Left- and Right-Turn Lanes at a 4SG Intersection

Type of Lane	CMF for Number of Approaches with Exclusive Turn Lane(s)			
	1	2	3	4
LT Lane(s)	0.90	0.81	0.73	0.66
RT Lane(s)	0.96	0.92	0.88	0.85

(Source: Adapted from (*Highway Safety Manual*, 1st Edition), (2010), by the American Association of State Highway and Transportation Officials, Washington, D.C. U.S.A)

Table 12.9: Full Alternative Text

Table 12.10: CMFs for Left-Turn Phasing at a 4SG Intersection

Type of LT Phasing	CMF for Number of Approaches w/LT Phasing			
	1	2	3	4
Permitted	1.00	1.00	1.00	1.00
Protected + Permitted, or Permitted + Protected	0.99	0.98	0.97	0.96
Protected	0.94	0.88	0.83	0.78

(Source: Adapted from (*Highway Safety Manual*, 1st Edition), (2010), by the American Association of State Highway and Transportation Officials, Washington, D.C. U.S.A)

Table 12.10: Full Alternative Text

The CMF for right-turn on red is given by Equation 12-9. The CMF for roadway lighting is given by Equation 12-10.

CMFRTOR=0.98nprohibited [12-9]

where $n_{prohibited}$ is the number of intersection approaches that prohibit right-turns on red.

CMFL=1−0.38 pnight [12-10]

where p_{night} is the proportion of base crashes that occurs at night, a value that is a constant of 0.235 for 4SG intersections.

The CMF for the existence of red-light cameras at an intersection is a bit more complicated than for other conditions. Equation 12-11 is used for the prediction.

CMFRLC=1−0.26 pra+0.18 pre [12-11]

where:

pra=proportion of multivehicle crashes that arerightangle collisions, andpre=proportion of multivehicle crashes that arerear-end collisions

The premise is that the presence of a red-light camera reduces right-angle collisions and increase rear-end collisions. Further, the adjustment only applies to multivehicle accidents, as neither type of crash can involve a single vehicle. Thus, for single-vehicle crashes, CMF_{RLC} is, by definition, 1.00.

The HSM provides many details that enable the analyst to estimate values of p_{ra} and p_{re}. For the purposes of this text, it is assumed that these values are known from intersection accident records.

Once the six applicable CMFs are estimated, the predicted number of multivehicle and single-vehicle crashes at the subject 4SG intersection is found as follows:

Npredi=Nbi×CMFLT×CMFSP×CMFRT×CMFRTOR×CMFL×CMFRLC [12-12]

where *i* is *mv* for multivehicle crashes and *sv* for single-vehicle crashes. All other terms have been previously defined.

Three conditions have been found to affect vehicle–pedestrian crashes: (1) the number of bus stops within 1,000 ft of the intersection, (2) the number of schools within 1,000 ft of the intersection, and (3) the number of stores selling alcohol within 1,000 ft of the intersection. These factors are summarized in Table 12.11.

Table 12.11: CMFs for Vehicle-Pedestrian Collisions at 4SG Intersections

Condition	Crash Modification Factor
Bus Stops within 1,000 ft of Intersection	
0	$CMF_{BS} = 1.00$
1 or 2	$CMF_{BS} = 2.78$
3 or more	$CMF_{BS} = 4.15$
Schools within 1,000 ft of Intersection	
0	$CMF_{SCH} = 1.00$
1 or more	$CMF_{SCH} = 1.35$
Stores Selling Alcohol within 1,000 ft of Intersection	
0	$CMF_{ALC} = 1.00$
1–8	$CMF_{ALC} = 1.12$
9 or more	$CMF_{ALC} = 1.56$

(Source: Adapted from (*Highway Safety Manual*, 1st Edition), (2010), by the American Association of State Highway and Transportation Officials, Washington, D.C. U.S.A)

Table 12.11: Full Alternative Text

The predicted number of vehicle–pedestrian crashes at the intersection may now be estimated as follows:

Npedpred=Npedi×CMFBS×CMFSCH×CMFALC [12-13]

where all terms are as previously defined.

With the estimation of predicted multivehicle and single-vehicle crashes, the number of vehicle–bicycle crashes can be determined using Equation 12-8, presented previously.

Putting It All Together

Equation 12-2, presented earlier, can now be applied to determine the total number of expected crashes at the 4SG intersection. The equation includes a local calibration factor, *c*, which is found by comparing a known year's actual crash numbers with that predicted by the HSM algorithms. The HSM gives detailed directions on how this should be done. For the purposes of this text, it is assumed that the value is known.

In the interests of time and space, we have not included detailed discussions of what the numbers and values presented imply. All values presented in the HSM are backed up by extensive research studies, and the manual includes much discussion of their import.

In reality, one suspects that an intersection with more than nine liquor stores within 1,000 ft probably has more to consider than just the accident consequences of this fact.

This text also does not cover how these predictions can be further modified to fit actual crash data. Again, the HSM contains much detail on this important factor.

Sample Problem 12-1: HSM Analysis of an Intersection

A four-leg signalized intersection on an urban arterial is to be evaluated for safety. No reliable crash history data are available. The HSM is used to predict the expected crash frequency. Known data for the intersection are summarized in Table 12.12.

Table 12.12: Data for Illustrative *HSM* Analysis of a 4SG Intersection

Condition	Value
Major Street AADT	34,000 veh/day
Minor Street AADT	16,000 veh/day
Red-Light Cameras	None
Number of Approaches with LT Lanes	2
Number of Approaches with RT Lanes	None
Number of Approaches with Exclusive LT Phasing	2
Number of Approaches with RTOR Prohibited	None
Intersection Lighting	Provided
Total Pedestrian Crossing Volume	540 peds/day
Maximum Number of Lanes Crossed by a Pedestrian	6 lanes
Number of Bus Stops within 1,000 ft of the Intersection	None
Number of Schools within 1,000 ft of the Intersection	1
Number of Liquor Stores within 1,000 ft of the Intersection	None
Local Calibration Factor	0.94

Table 12.12: Full Alternative Text

1. Step 1: Find SPFs for multivehicle and single-vehicle crashes and vehicle–pedestrian crashes (crashes/year), using Equation 12-3 and Table 12.4 for multivehicle crashes, Equation 12-6 and Table 12.5 for single-vehicle crashes, and Equation 12-7 and Table 12.6 for vehicle–pedestrian crashes:

    1. $N_{bmvTOTAL}=\exp[-10.99+1.07\times\ln(34{,}000)+0.23\times\ln(16{,}000)]=11.03$

    2. $N_{bmv(FI)}=\exp[-13.14+1.18\times\ln(34{,}000)+0.22\times\ln(16{,}000)]=3.68$

    3. $N_{bmv(PDO)}=\exp[-11.02+1.02\times\ln(34{,}000)+0.24\times$ l

    4. $N_{bsvTOTAL}=\exp[-10.21+0.68\times\ln(34{,}000)+0.27\times\ln(16{,}000)]=0.61$

    5. $N_{bsv(FI)}=\exp[-9.25+0.43\times\ln(34{,}000)+0.29\times\ln(16{,}000)]=0.14$

6. $N_{bsv(PDO)}=\exp[-11.34+0.79\times \ln(34{,}000)+0.25\times \ln(16{,}000)]=0.45$

7. $N_{pedbase}=\exp[-9.53+0.4 \times \ln(34{,}000+16{,}000)+0.26\times \ln(16{,}000/34{,}000)+0.45\times$
   $]=0.10$

2. Step 2: Adjust *mv* and *sv* SPF values so that the sum of FI and PDO crashes matches the base total (crashes/year), using Equations 12-4 and 12-5:

   $N_{bmv(FI)}=11.03\times 3.68 3.68+7.00=3.80 N_{bmv(PDO)}=11.0$

3. Step 3: Find total predicted average crash frequency for intersection (crashes/year)

   $N_{bmv}=3.80+7.23=11.03 N_{bsv}=0.14+0.46=0.61 N_{bInt}=11.$

   Note that these totals now agree with the predictions of total crashes in *Step 1*.

4. Step 4: Find CMFs.

   A perusal of Table 12.12 indicates that there are three conditions that do not conform to the base conditions for multivehicle and single-vehicle crashes: (a) there are two approaches with an LT lane, (b) there are two approaches with an exclusive LT phase (fully protected), and (c) the intersection is well-lighted.

   From Table 12.9, the CMF for LT lanes is 0.81, while the CMF for exclusive LT phasing is 0.88. The CMF for a lighted intersection is defined by Equation 12-10:

   $CMF_L=1-0.38\ p_{night} CMF_L=1-(0.38\times 0.235)=0.911$

   By definition, all other CMFs applied to multivehicle and single-vehicle crashes conform to base conditions, i.e., $CMF_i=1.00$ for these conditions. Then, applying Equation 12-12:

   $N_{pred i}=N_{bi}\times CMF_{LT}\times CMF_{SP}\times CMF_{RT}\times CMF_{RTOR}\times C$

A perusal of Table 12.12 also shows that there is only one CMF that would apply to the current case: a school is located within 1,000 ft of the intersection. From Table 12.11, the CMF_{SCH} is 1.35. Again, all other CMFs that might apply to vehicle–pedestrian crashes are, by definition, 1.00. Using Equation 12-13:

Npedpred=Npedi×CMFBS×CMFSCH×CMFALCNpedpr

5. Step 5: Find predicted average vehicle–bicycle crashes

   The number of vehicle–bicycle crashes is estimated using Equation 12-8:

   Nbikei=(Npred,mv+Npred,sv)×fbikeiNbikei=(7.162+0.396)×0.015=0.113 crashes/yr

   where 0.015 is the prescribed adjustment factor for 4SG intersections.

6. Step 6: Predict total crashes/yr for the intersection.

   The total number of crashes for the subject intersection is given by Equation 12-2:

   Npred,int=ci (Nbi+Npedi+Nbikei)Npred,int=0.94 (7.162+

   where 0.94 is the local calibration factor cited in Table 12.12, and N_{bi} is the sum of the multivehicle and single-vehicle crashes predicted (7.162 and 0.396, respectively).

12.3.5 The *HSM* Impact

While this text can only show a small sample of what the complex models of the HSM can do, it is impossible to understate the impact of this important document. It has assembled a massive database on highway traffic safety, and developed a set of predictive models that can estimate crash frequencies (crashes/yr) of a wide variety of facility types, including both point locations (intersections) and roadway sections.

Techniques are presented to allow localized calibrations through the use of local crash data. The resulting models not only allow the prediction of crash frequency for a set of given conditions but also evaluate how crash frequency would change under various improvement scenarios.

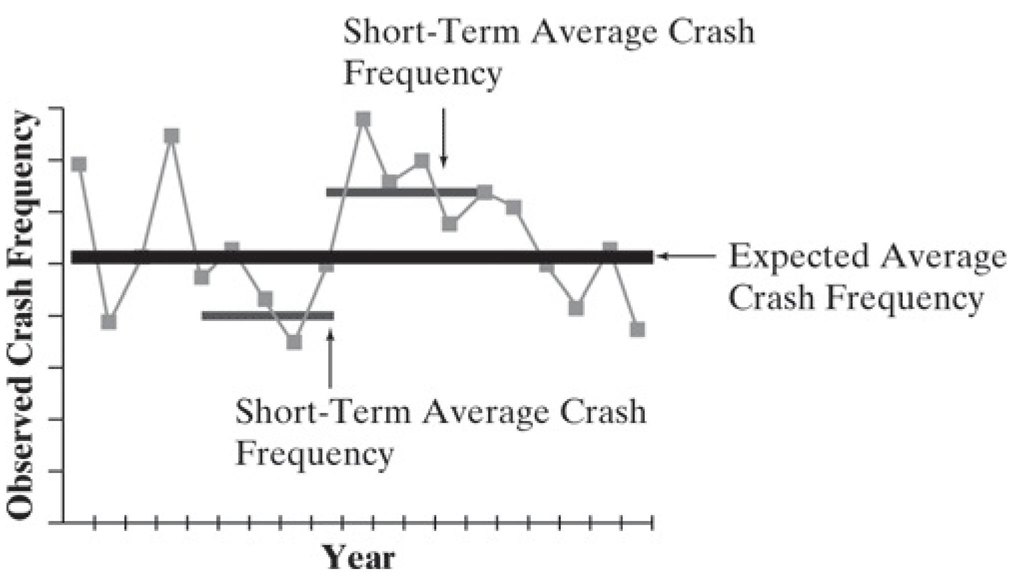
The HSM has dramatically changed how highway safety analysis is done, and provided extremely valuable predictive tools that drastically extend traffic engineers' ability to compare alternative design and control alternatives to improve crash frequency.

12.4 Historical Crash Data and Regression to the Mean

The number of crashes at any particular site is a random variable. The number fluctuates year to year, up and down, but tends to converge to a long-term average. This converging to an average over the long term is called RTM, which can lead to RTM bias when evaluating a treatment. RTM bias makes a treatment appear more effective than it actually is. Locations that had a high number of crashes over the past few years use a "before" treatment high-crash value that does not take into account the random increases in crashes. Thus, a simple before–after study that does not account for RTM may have RTM bias.

Figure 12.9 [9] shows the natural variability in crash frequency. One can see the difference between short-term average crash frequency and the expected average crash frequency.

Figure 12.9: Natural Variation in Crash Frequency

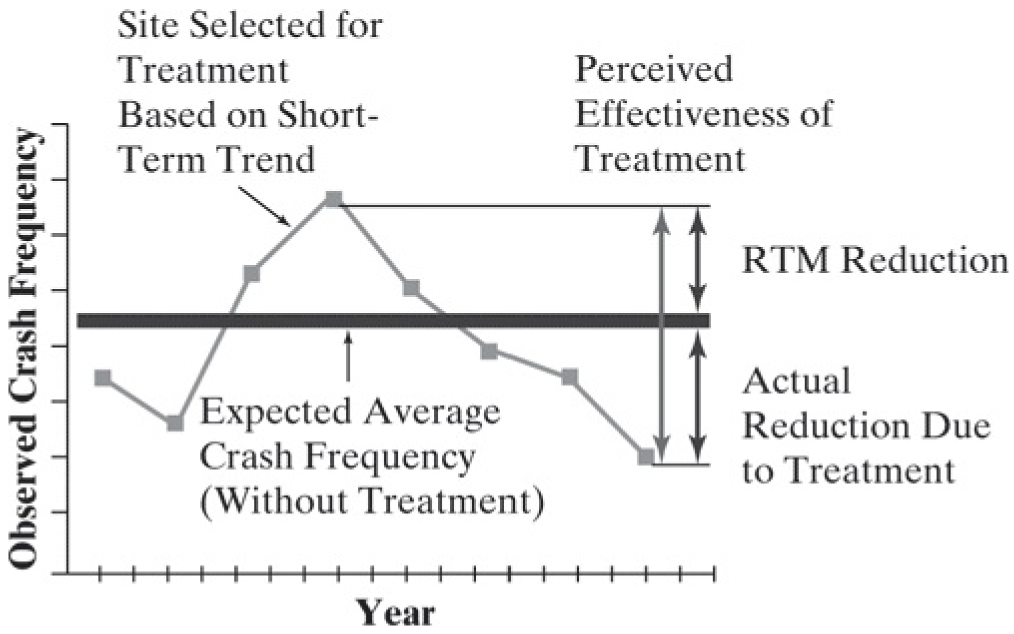


(Source: Adapted from (*Highway Safety Manual*, 1st Edition), (2010), by the American Association of State Highway and Transportation Officials, Washington, D.C. U.S.A.)

Figure 12.9: Full Alternative Text

Figure 12.10 shows the effect of not adjusting for RTM. The effectiveness of the treatment may be overestimated, and some or all of the change might have occurred without the treatment.

Figure 12.10: Effect of Not Adjusting for RTM



(Source: Adapted from (*Highway Safety Manual*, 1st Edition), (2010), by the American Association of State Highway and Transportation Officials, Washington, D.C. U.S.A)

Figure 12.10: Full Alternative Text

To avoid RTM errors, it is always better to work with a database that can establish a long-term expected average crash frequency for evaluation. Reliance on a single year of data, for example, runs the risk of introducing a possible substantial error due to RTM problems.

12.5 Effective Crash Countermeasures

The term “countermeasures” appears in the literature as actions taken to mitigate crash experience or severity. Some documents, such as the HSM, refer to “treatments” and the “crash effects” of treatments.

Volume 3 of the HSM addresses a range of treatments for various roads and conditions, and includes estimated CMFs in tabular form supported by text.

Another important document on countermeasures is *Countermeasures That Work* [14], sponsored by NHTSA. The current edition at this writing was the 8th edition (2016). New editions are scheduled for release on a biannual basis.

Countermeasures That Work is targeted to state highway safety offices (SHSOs). It explicitly notes that “It does not include countermeasures for which SHSOs have little or no authority or responsibility, or that cannot be supported under typical highway safety grant programs. For example, the guide does not include vehicle- or roadway- based solutions ... [or] ... countermeasures that are already in place in every State.”

The 8th edition focuses on nine areas:

1. Alcohol- and drug-impaired driving
2. Seat belts and child restraints
3. Speeding and speed management
4. Distracted and drowsy driving
5. Motorcycle safety
6. Young drivers
7. Older drivers

8. Pedestrians
9. Bicycles

Table 12.13 shows an illustration of the way in which *Countermeasures That Work* presents information. On countermeasures in chapters addressing each of the above topics, accompanied by a page or so on each countermeasure.

Table 12.13: Illustration of Information Presentation in *Countermeasures That Work* (Example: Alcohol-Related Crashes)

1. Deterrence: Laws

Countermeasure	Effectiveness	Cost	Use	Time
1.1 ALR/ALS	★★★★★	$$$	High	Medium
1.2 Open containers	★★★	$	High	Short
1.3 High-BAC sanctions	★★★	$	Medium	Short
1.4 BAC test refusal penalties	★★★	$	Unknown	Short
1.5 Alcohol-impaired driving law review	★★★	$$	Unknown	Medium

2. Deterrence: Enforcement

Countermeasure	Effectiveness	Cost	Use	Time
2.1 Publicized sobriety checkpoints	★★★★★	$$$	Medium	Short
2.2 High visibility saturation patrols	★★★★	$$	High	Short
2.3 Preliminary Breath Test devices (PBTs)†	★★★★	$$	High	Short
2.4 Passive alcohol sensors††	★★★★	$$	Unknown	Short
2.5 Integrated enforcement	★★★	$	Unknown	Short

†Proven for increasing arrests
††Proven for detecting impaired drivers

3. Deterrence: Prosecution and Adjudication

Countermeasure	Effectiveness	Cost	Use	Time
3.1 DWI courts†	★★★★	$$$	Low	Medium
3.2 Limits on diversion and plea agreements††	★★★★	$	Medium	Short
3.3 Court monitoring††	★★★	$	Low	Short
3.4 Sanctions	★★	Varies	Varies	Varies

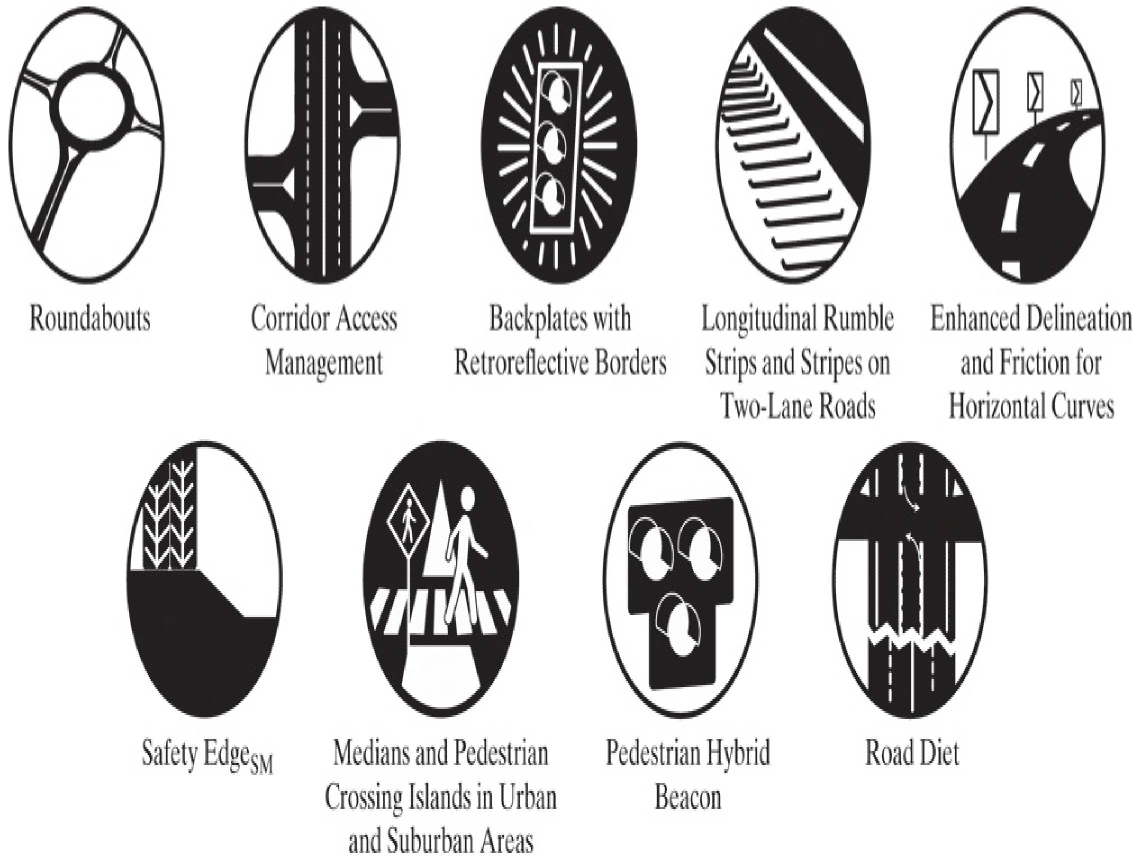
†Proven for reducing recidivism
††Proven for increasing convictions

(Source: Goodwin, A., et al, *Countermeasures That Work: A Highway Safety Countermeasure Guide for State Highway Safety Offices*, 8th Edition, National Highway Traffic Safety Administration, U.S. Department of Transportation, Washington, D.C., 2015, pgs 1–7.)

Table 12.13: Full Alternative Text

The FHWA Office of Safety lists nine countermeasure areas that have shown great effectiveness in safety improvement that are research-proven but not widely applied on a national basis, as *Proven Safety Countermeasures* [15]. Refer to Figure 12.11. Each of the buttons clicks through to additional information and resources. FHWA also cites another important source of information [16], which is the *Crash Modification Factors Clearinghouse*.

Figure 12.11: Underutilized Countermeasures Referenced by FHWA



(Source: https://safety.fhwa.dot.gov/provencountermeasures)

Figure 12.11: Full Alternative Text

Institute of Transportation Engineers (ITE) and other sources list countermeasures and probable areas of benefit. For instance, [17] is an ITE website listing signalization countermeasures at signalized intersections, geometric countermeasures at unsignalized intersections, and signs/markings/operational countermeasures related to one or both. Table 12.14 shows an illustration.

Table 12.14: Sample Countermeasures Drawn from an ITE Web Toolbox

		Potential Effectiveness (Percentage Reduction)							
Improvement Type(s)	Cost	Total Crashes	Right Angle Crashes	Left Turn Crashes	Rear-end Crashes	Sideswipe	Pedestrian	Red-Light Mauling	Older Driver
SIGNAL OPERATIONS IMPROVEMENTS									
Interconnect/Coordinate Traffic Signals; Optimization	Medium	15-17 [1]	25-38 [12]		•			• [2]	
Increase/Modify Clearance Intervals	Low	4-31 [1,9,10]	1-30 [1,9]		•			• [2]	
Improve Signal Timing (General)	Low	10-15 [1]	•	•		•		•	
Add Protected/Permissive LT Phase	Medium	4-10 [1,9]		40-64 [1,9]					
Use Green Arrow/Protected Left Turns/Movement Signal Phasing	Low	3 [9]		98 [9]					•
Use Split Phases	Low	25 [11]		•	•	•			
Use Leading Pedestrian Interval	Low						5 [8]		
Add Pedestrian Phase	Medium	23-25 [1]					7-60 [1,8]		
Add Left-Turn Phasing to an Existing Signal	Medium	23-48 [6,12]		63-70 [11]			5 [8]		
Provide Green Extension (Advance Detection)	Variable				•			•	
Install Signal Actuation	Variable				•	•			
Assume Slower Walking Speeds for Pedestrian Signal Timing	Low						•		•
Provide Advance Warning of Signal Changes at Rural Signalized Intersections	Medium	•	•		•			•	
Remove Signals from Late Night/Early Morning Flash	Low	29 [9]	80 [9]						
Consider Restricting Right-Turns-on-Red	Low						•		
Consider Installation of Pedestrian Countdown Signals (Incremental Cost)	Low						•		
Consider Installation of Animated Eye Signals (Incremental Cost)	Low						•		

(Note: Reference numbers in table refer to the source ITE document.)

(Source: http://library.ite.org/pub/326c7e9c-2354-d714-5181-4cc79fba5459)

[Table 12.14: Full Alternative Text]()

12.6 Approaches to Highway Safety

Improving highway safety involves consideration of three elements influencing traffic operations: the driver, the vehicle, and the roadway. *Safer Roads: A Guide to Road Safety Engineering* [18] cites five categories of strategies:

1. Exposure control
2. Crash prevention
3. Behavior modification
4. Injury control
5. Postinjury management

12.6.1 Exposure Control

Exposure control is common to both lists and involves strategies that reduce the number of vehicle-miles of travel by motorists.

Efforts to reduce auto use and travel cover a wide range of policy, planning, and design issues. Policies and practices that attempt to reduce auto use include the following:

- Diversion of travel to public transportation modes
- Substitution of telecommunications for travel
- Implementation of policies, taxes, and fees to discourage auto ownership and use
- Reorganization of land uses to minimize travel distances for various trip purposes

- Driver and vehicle restrictions through licensing and registration restrictions

Most of these strategies must take place over long time periods, and many require systemic physical changes to the urban infrastructure and behavioral changes in the traveling public. Some require massive investments (such as providing good public transportation alternatives and changing the urban land use structure), while others have not yet demonstrated the potential to affect large changes in travel behavior.

12.6.2 Crash Risk Control/Crash Prevention

Crash risk control and crash prevention are similar terms with a number of common features. They are not, however, the same. Crash prevention implies actions that reduce the number of crashes that occur for a given demand level. Crash risk control incorporates this, but also includes measures that reduce the *severity* of a crash when it occurs. Reduction of crash severity overlaps crash risk control and injury control strategies.

Crash prevention involves a number of policy measures, including driver and pedestrian training, removal of drivers with "bad" driving records (through the suspension or revocation of licenses), and provision of better highway designs and control devices that encourage good driving practices and minimize the occurrence of driver error.

Risk control, or reduction of severity, often involves the design and protection of roadside and median environments. Proper guardrail and/or impact- attenuating devices will reduce the impact energy transferred to the vehicle in a crash, and can direct the path of a vehicle away from objects or areas that would result in a more serious collision.

12.6.3 Behavior Modification

This category, separately listed in Reference [18], is an important component of strategies for crash prevention and exposure reduction.

Affecting mode choice is a major behavior modification action that is hard to successfully achieve. Often, this requires providing very high-class and convenient public transportation alternatives and implementing policies that make public transportation a much more attractive alternative than driving for commuter and other types of trips. This is an expensive process, often involving massive subsidies to keep the cost of public transportation reasonable, coupled with high parking and other fees associated with driving. Use of high-occupancy vehicle lanes and other restricted-use lanes to speed public transportation, providing a visual travel-time differential between public transportation and private automobiles, is another useful strategy.

If drivers and motorists cannot be successfully diverted to alternative modes, driver and pedestrian training programs are a common strategy for behavior modification. Many states offer insurance discounts if a basic driving safety course is completed every three years. There is, however, little statistical evidence that driver training has any measurable effect on crash prevention.

The final strategy in behavior modification is enforcement. This can be very effective, but it is also expensive. Speed limits will be more closely obeyed if enforcement is strict, and the fines for violations are expensive. In recent years, the use of automated systems for ticketing drivers who violate red lights have become quite popular. Automated speed enforcement is also possible with current technologies. The issues involved in automated enforcement are more legal than technical at present. While the license plate of a vehicle running a red light can be automatically recorded, it does not prove who is driving the vehicle. In most states, automated ticketing results in a fine, but does not include “points” on the owner’s license, since it cannot be proved that the owner was the driver at the time of the violation. Some modern technologies now also include photographing the driver’s face to address this situation.

12.6.4 Injury Control

Injury control focuses on crash survivability of occupants in a vehicular crash. This is primarily affected by better vehicle design that is generally “encouraged by an act of Congress.” Vehicle design features that have been implemented with improved crash survivability in mind include the following:

- Seat belts and shoulder harnesses, and laws requiring their use
- Child-restraint seats and systems, and laws requiring their use
- Anti-burst door locks
- Padded instrument panels
- Energy-absorbing steering posts and crumple zones
- Side door beams
- Air bags
- Head rests and restraints
- Shatterproof glass
- Forgiving interior fittings

12.6.5 Postinjury Management

Traffic fatalities tend to occur during three critical time periods:

- During the crash occurrence, or within minutes of it. Death is usually related to head or heart trauma or extreme loss of blood.
- Within one to two hours of the crash occurrence. In this period, death is usually due to the same causes noted above: head or heart trauma and/or loss of blood.
- Within 30 days of admission to the hospital. Death usually results from cessation of brain activity, organ failure, or infection.

About 50% of traffic fatalities occur in the first category, 35% in the second, and 15% in the third.

Deaths within one to two hours of a crash can be reduced by systems that ensure speedy emergency medical responses along with high-quality emergency care at the site and during transport to a hospital facility. Such systems involve speedy notification of emergency services, fast dispatch of appropriate equipment to the site, well-trained emergency medical technicians attending to immediate medical needs of victims, and well-staffed and equipped trauma centers at hospitals. Since survival often depends upon quickly stabilizing a victim at the crash site and speedy transport to a trauma center, communications and dispatch systems must be in place to respond to a variety of needs.

12.7 Commonly Used Crash Statistics and Analyses

12.7.1 Types of Statistics

Crash statistics generally address and describe one of three principal informational elements:

- Crash occurrence
- Crash involvements
- Crash severity

Crash occurrence relates to the numbers and types of crashes that occur, which are often described in terms of rates based on population or vehicle-miles traveled. *Crash involvement* concerns the numbers and types of vehicles and drivers involved in crashes, with population-based rates a very popular method of expression. *Crash severity* is generally dealt with by proxy: the numbers of fatalities and fatality rates are often used as a measure of the seriousness of crashes.

Statistics in these three categories can be stratified and analyzed in an almost infinite number of ways, depending upon the factors of interest to the analyst. Some common types of analyses include the following:

- Trends over time
- Stratification by highway type or geometric element
- Stratification by driver characteristics (gender, age)
- Stratification by contributing cause
- Stratification by crash type
- Stratification by environmental conditions

Such analyses allow the correlation of crash types with highway types and specific geometric elements, the identification of high-risk driver populations, quantifying the extent of Driving Under the Influence/Driving While Intoxicated (DUI/DWI) influence on crashes and fatalities, and other important determinations. Many of these factors can be addressed through policy or programmatic approaches. Changes in the design of guardrails have resulted from the correlation of crash and fatality rates with specific types of installations. Changes in the legal drinking age and in the legal definition of DUI/DWI have resulted partially from statistics showing the very high rate of involvement of this factor in fatal crashes. Improved federal requirements on vehicle safety features (air bags, seat belts and harnesses, energy-absorbing steering columns, padded dashboards) have occurred partially as a result of statistics linking these features to crash severity.

12.7.2 Crash Rates

Simple statistics citing total numbers of crashes, involvements, injuries, and/or deaths can be quite misleading, as they ignore the base from which they arise. An increase in the number of highway fatalities in a specific jurisdiction from one year to the next must be matched against population and vehicle-usage patterns to make any sense. For this reason, many crash statistics are presented in the form of rates.

Population-Based Crash Rates

Crash rates generally fall into one of two broad categories: *population-based* rates and *exposure-based* rates. Some common bases for population-based rates include the following:

- Area population
- Number of registered vehicles
- Number of licensed drivers
- Highway mileage

These values are relatively static (they do not change radically over short periods of time) and do not depend upon vehicle usage or the total amount of travel. They are useful in quantifying overall risk to individuals on a comparative basis. Numbers of registered vehicles and licensed drivers may also partially reflect usage.

Exposure-Based Crash Rates

Exposure-based rates attempt to measure the amount of travel as a surrogate for the individual's exposure to potential crash situations. The two most common bases for exposure-based rates are:

- Vehicle-miles traveled
- Vehicle-hours traveled

The two can vary widely depending upon the speed of travel, and comparisons based on mileage can yield different insights from those based on hours of exposure. For point locations, such as intersections, vehicle-miles or vehicle-hours have very little significance. Exposure rates for such cases are "event-based" using total volume passing through the point to define "events."

True "exposure" to risk involves a great deal more than just time or mileage. Exposure to vehicular or other conflicts that are susceptible to crash occurrence varies with many factors, including volume levels, roadside activity, intersection frequency, degree of access control, alignment, and many others. Data requirements make it difficult to quantify all of these factors in defining exposure. The traffic engineer should be cognizant of these and other factors when interpreting exposure-based crash rates.

Common Bases for Crash and Fatality Rates

In computing crash rates, numbers should be scaled to produce meaningful values. A fatality rate per mile of vehicle travel would yield numbers with many decimal places before the first significant digit, and would be difficult to conceptualize. The following list indicates commonly used forms for stating crash and fatality rates:

Population-based rates are generally stated according to the following:

- Fatalities, crashes, or involvements per 100,000 area population
- Fatalities, crashes, or involvements per 10,000 registered vehicles
- Fatalities, crashes, or involvements per 10,000 licensed drivers
- Fatalities, crashes, or involvements per 1,000 miles of highway

Exposure-based rates are generally stated according to the following:

- Fatalities, crashes, or involvements per 100,000,000 vehicle-miles traveled
- Fatalities, crashes, or involvements per 10,000,000 vehicle-hours traveled
- Fatalities, crashes, or involvements per 1,000,000 entering vehicles (for intersections only)

Sample Problem 12-2: Crash Statistics

The following are sample gross crash statistics for a relatively small urban jurisdiction in the Year 2017:

- Fatalities: 75
- Fatal crashes: 60
- Injury crashes: 300
- PDO crashes: 2,000

- Total involvements: 4,100
- Vehicle-miles traveled: 1,500,000,000
- Registered vehicles: 100,000
- Licensed drivers:150,000
- Area population: 300,000

In general terms, all rates are computed as:

Rate=Total×(ScaleBase) [12-14]

where:

Total=total number of crashes, involvements, orfatalities,Scale=scale of the miles traveled,Base=total base statistic for the period of the rate.

Using this formula, the following fatality rates can be computed using the sample data:

Rate 1=75 × (100,000300,000)=25 deaths per 100,000 populationRate 2=75 miles

Similar rates may also be computed for crashes and involvements but are not shown here.

Crash and fatality rates for a given county, city, or other jurisdiction should be compared against past years, as well as against state and national norms for the analysis year. Such rates may also be subdivided by highway type, driver age and gender groupings, time of day, and other useful breakdowns for analysis.

12.7.3 Severity Index

A widely used statistic for the description of relative crash severity is the severity index (SI), defined as the number of fatalities per crash. For the data of the previous example, there were 75 fatalities in a total of 2,360 crashes. This yields an SI of:

$SI=752360=0.0318$ deaths per accident

The SI is another statistic that should be compared with previous years and state and national norms, so that conclusions may be drawn with respect to the general severity of crashes in the subject jurisdiction.

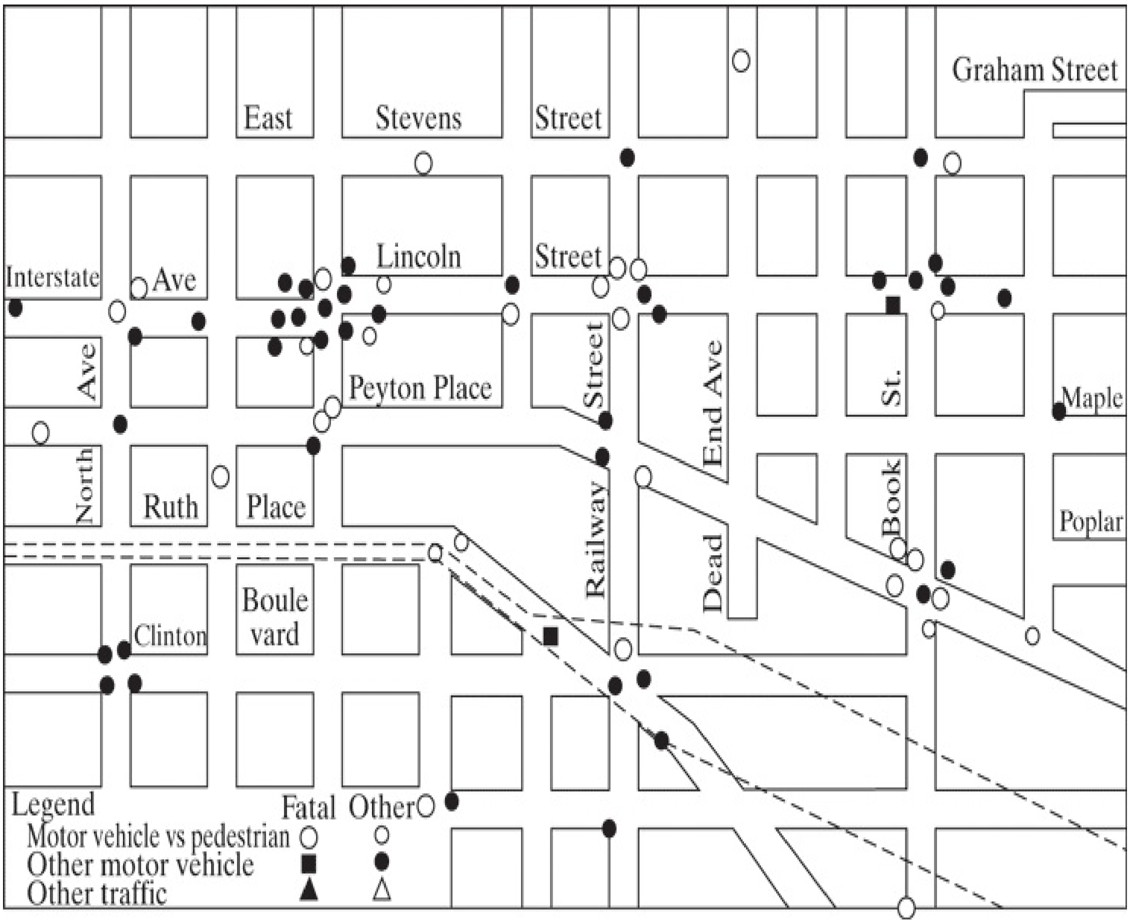
12.7.4 Identifying High-Accident Locations

Careful displays of crash statistics can tell a compelling story, identify critical trends, and spotlight specific problem areas. Care should be taken in the preparation of such displays to avoid misleading the reviewer; when the engineer reviews such information, he or she must analyze what the data say and (more importantly) what they *do not* say.

A primary function of a crash record system is to regularly identify locations with an unusually high rate of crashes and/or fatalities. Crash spot maps are a tool that can be used to assist in this task.

Figure 12.12 shows a sample crash spot map. Coded pins or markers are placed on a map. Color or shape codes are used to indicate the category and/or severity of the crash. Modern computer technology allows such maps to be electronically generated. To allow this, the system must contain a location code system sufficient to identify specific crash locations.

Figure 12.12: A Typical Accident Spot Map



(Source: Used with permission from *Institute of Transportation Engineers Manual of Traffic Engineering Studies*, Institute of Transportation Engineers, Washington D.C., 1994, pg 400)

Figure 12.12: Full Alternative Text

Computer record systems can also produce lists of crash locations ranked by either total number of crashes occurring or by defined crash or fatality rate. It is useful to examine both types of rankings, as they may yield significantly different results. Some locations with high crash numbers reflect high volumes and have a relatively low crash rate. Conversely, a small number of crashes occurring at a remote location with very little demand can produce a very high crash rate. While statistical rankings give the engineer a starting point, judgment must still be applied in the identification and selection of sites most in need of improvement during any given budget year.

One common approach to determining which locations require immediate attention is to identify those with crash rates that are significantly higher than the average for the jurisdiction under study. To say that the crash rate at a specific location is “significantly” higher than the average, only those locations with crash rates in the highest 5% of the (normal) distribution would be selected. In a one-tailed test, the value of *z* (on the standard normal distribution) for Prob (z) < 0.95 is 1.645. The actual value of *z* for a given crash location is computed as:

$$z = \frac{x_1 - \bar{x}}{s} \quad [12\text{-}15]$$

where:

x_1=accident rate at the location underconsideration,$\bar{x}$=average crash rate f tions within the jurisdiction under study.

If the value of *z* must be at least 1.645 for 95% confidence, the minimum crash rate that would be considered to be significantly higher than the average may be taken to be:

$$x_1 \geq 1.645s + \bar{x} \quad [12\text{-}16]$$

Locations with a higher crash rate than this value would be selected for specific study and remediation. It should be noted that in comparing average crash rates, similar locations should be grouped (i.e., crash rates for signalized intersections are compared to those for other signalized intersections; mid-block rates are compared to other mid-block rates).

Sample Problem 12-3: Identifying High-Accident Locations

Consider the following example: A major signalized intersection in a small city has a crash rate of 15.8 per 1,000,000 entering vehicles. The database for all signalized intersections in the jurisdiction indicates that the average crash rate is 12.1 per 1,000,000 entering vehicles, with a standard deviation of 2.5 per 1,000,000 entering vehicles. Should this intersection be singled out for study and remediation? Using Equation 12-15:

$$15.8 \geq (1.645 \times 2.5) + 12.1 = 16.2$$

For a 95% confidence level, the observed crash rate *does not* meet the criteria for designation as a significantly higher crash rate.

An important factor that tempers statistical identification of high-crash locations is the budget that can be applied to remediation projects in any given year. Ranking systems are important, as they can help set priorities. Priorities are necessary whenever funding is insufficient to address all locations identified as needing study and remediation. A jurisdiction may have 15 locations that are identified as having significantly higher crash rates than the average. However, if funding is available to address only eight of them in a given budget year, priorities must be established to select projects for implementation.

12.7.5 Before-and-After Crash Analysis

When a crash problem has been identified and an improvement implemented, the engineer must evaluate whether or not the remediation has been effective in reducing the number of crashes and/or fatalities. A before- and-after analysis must be conducted. The length of time considered before and after the improvement must be long enough to observe changes in crash occurrence. For most locations, periods ranging from three months to one year are used. The length of the "before" period and the "after" period is generally the same.

Consideration must be given to the "regression to the mean" as previously discussed: an atypically high-crash period might draw attention, but might also be just a short-term statistical fluctuation; improvements can be real, or can appear to be real simply by the situation returning toward its average.

Further, there can be a serious flaw in the way that most before-and-after crash analyses are conducted. There is generally a base assumption that any observed change in crash occurrence (or severity) is due to the corrective actions implemented. Because the time span involved in most studies is long, however, this may not be correct in any given case.

If possible, a control experiment or experiments should be established. These control experiments involve locations with similar crash experience that have not been treated with corrective measures. The controls establish the expected change in crash experience due to general environmental causes not influenced by corrective measures. For the subject location, the null hypothesis is that the change in crash experience is not significantly different from the change at observed control locations. While desirable from a statistical point of view, the establishment of control conditions is often a practical problem, requiring that some high-crash locations be left untreated during the period of the study. For this reason, many before–after studies of crashes are conducted without such control conditions.

12.8 Site Analysis

One of the most important tasks in traffic safety is the study and analysis of site-specific crash information to identify contributing causes and to develop site remediation measures that will lead to improved safety.

Once a location has been statistically identified as a “high-crash” location, detailed information is required in two principal areas:

1. Occurrence of crashes at the location in question
2. Environmental and physical conditions existing at the location

The analysis of this information must identify the environmental and physical conditions that potentially or actually contribute to the observed occurrence of crashes. Armed with such analyses, engineers may then develop countermeasures to alleviate the problem(s).

The best information on the occurrence of crashes is compiled by reviewing all crash reports for a given location over a specified study period. This can be done using computer crash records, but the most detailed data will be available from the actual police accident reports on file. Environmental and physical conditions are established by a thorough site investigation conducted by appropriate field personnel. Two primary graphical outputs are then prepared:

1. Crash diagram
2. Condition diagram

12.8.1 Crash Diagrams

A crash diagram is a schematic representation of all crashes occurring at a given location over a specified period. Depending upon the crash frequency, the “specified period” usually ranges from one to three years.

Each crash is represented by a set of arrows, one for each vehicle involved, which schematically represents the type of crash and directions of all vehicles. Arrows are generally labeled with codes indicating vehicle types, date and time of crash, and weather conditions.

The arrows are placed on a schematic (not-to-scale) drawing of the intersection with no interior details shown. One set of arrows represents one crash. It should be noted that arrows are not necessarily placed at the exact spot of the crash on the drawing. There could be several crashes that occurred at the same spot, but separate sets of arrows would be needed to depict them. Arrows illustrate the occurrence of the crash, and are placed as close to the actual spot of the crash as possible.

Figure 12.13 shows the standard symbols and codes used in the preparation of a typical crash diagram. Figure 12.14 shows an illustrative crash diagram for an intersection.

Figure 12.13: Symbols Used in Crash Diagrams

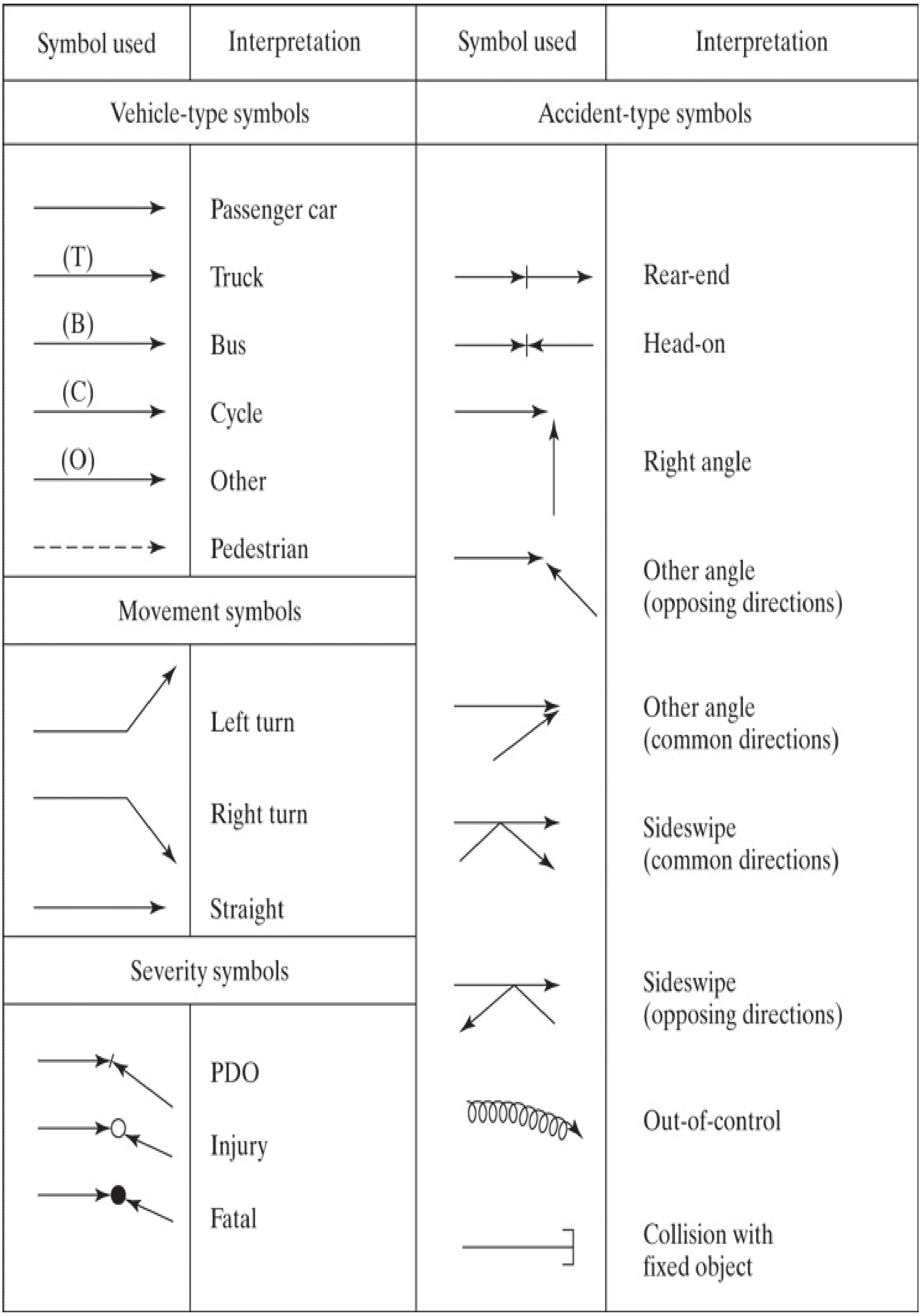
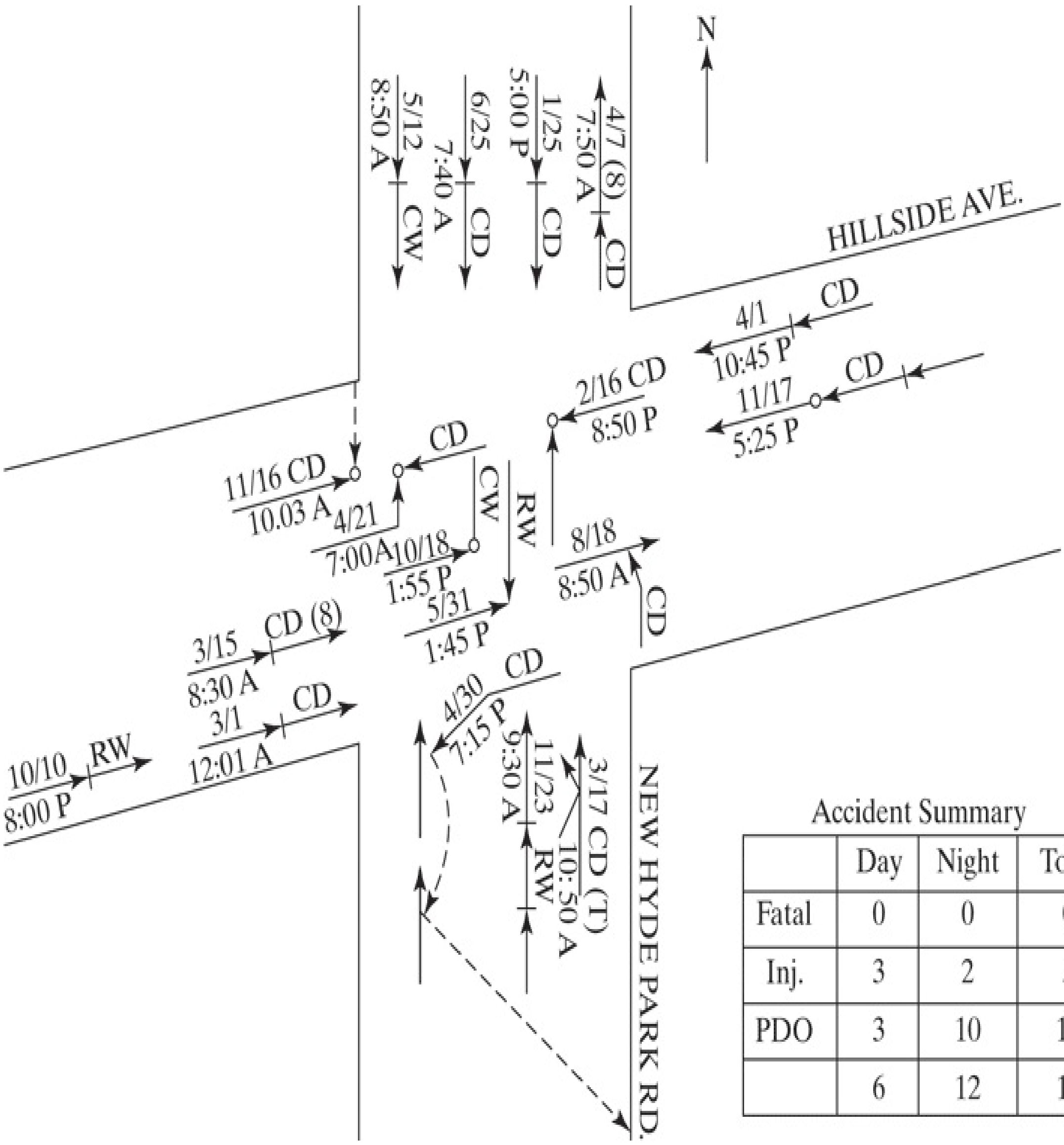


Figure 12.13: Full Alternative Text

Figure 12.14: An Illustrative

Crash Diagram



Accident Summary

	Day	Night	Total
Fatal	0	0	0
Inj.	3	2	5
PDO	3	10	13
	6	12	18

Figure 12.14: Full Alternative Text

The crash diagram provides a powerful visual record of crash occurrence over a significant period of time. In Figure 12.14, it is clear that the intersection has experienced primarily rear-end and right-angle collisions, with several injuries but no fatalities during the study period. Many of the crashes appear to be clustered at night. The diagram clearly points out these patterns, which now must be correlated to the physical and control characteristics of the site to determine contributing causes and appropriate corrective measures.

12.8.2 Condition Diagrams

A condition diagram describes all physical and environmental conditions at the crash site. The diagram must show all geometric features of the site, the location and description of all control devices (signs, signals, markings, lighting, etc.), and all relevant features of the roadside environment, such as the location of driveways, roadside objects, and land uses. The diagram must encompass a large enough area around the location to include all potentially relevant features. This may range from several hundred feet on intersection approaches to 0.25–0.50 miles on rural highway sections.

Figure 12.15 illustrates a condition diagram. It is for the same site and time period as the collision diagram of Figure 12.14. The diagram includes several hundred feet of each approach and shows all driveway locations and the commercial land uses they serve. Control details include signal locations and timing, location of all stop lines and crosswalks, and even the location of roadside trees, which could conceivably affect visibility of the signals.

Figure 12.15: Illustrative Condition Diagram

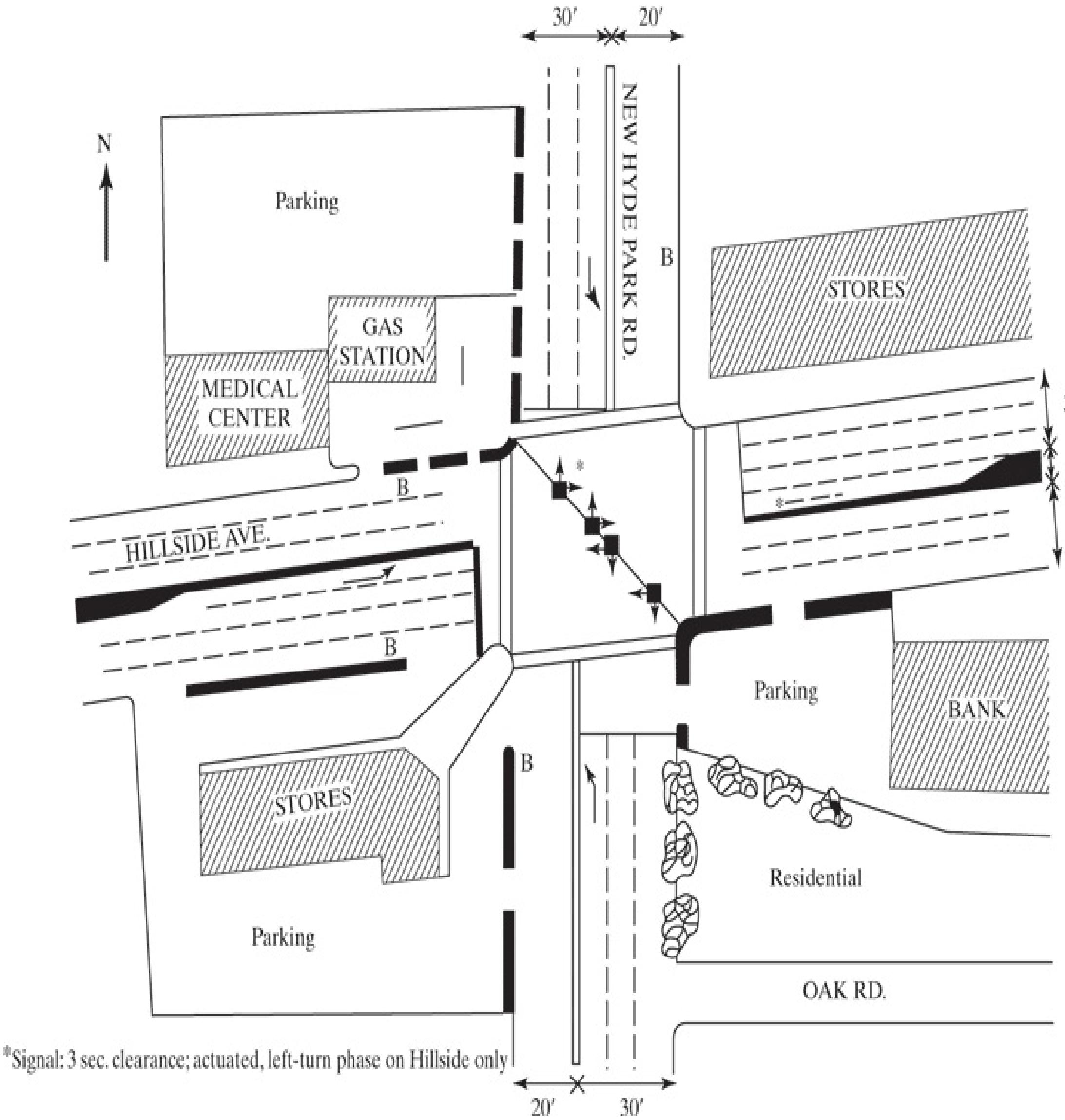


Figure 12.15: Full Alternative Text

12.8.3 Interpretation of Condition and Crash Diagrams

This brief overview chapter cannot fully discuss and present all types of crash site analyses. The objective in analyzing collision and condition diagrams is straightforward: find contributing causes to the observed crashes shown in the collision diagram among the design, control, operational, and environmental features summarized on the condition diagram. Doing so involves virtually all of the traffic engineer's knowledge, experience, and insight, and the application of professional judgment. A sample problem in analysis and interpretation follows, focusing on the data depicted in Figures 12.13 and 12.14.

Sample Problem 12-4: Interpreting Crash and Condition Diagrams

Crashes are generally grouped by type. Predominant types of crashes shown in Figure 12.14 are rear-end and right-angle collisions. For each type of crash, three questions should be asked:

1. What driver actions lead to the occurrence of such crashes?
2. What existing conditions at the site could contribute to drivers taking such actions?
3. What changes can be made to reduce the chances of such actions taking place?

Rear-end crashes occur when the lead vehicle stops suddenly or unexpectedly and/or when the trailing driver follows too closely for the prevailing speeds and environmental conditions. While “tailgating” by a following driver cannot be easily corrected by design or control measures, there are a number of factors evident in Figures 12.14 and 12.15 that may contribute to vehicles stopping suddenly or unexpectedly.

The condition diagram shows a number of driveways allowing access to and egress from the street at or near the intersection itself. Unexpected movements into or out of these driveways could cause mainline vehicles to stop suddenly. Because of these driveways, STOP lines are located well back from the sidewalk line, particularly in the northbound direction. Vehicles, therefore, are stopping at positions not normally expected, and following drivers may be surprised and unable to respond in time to avoid a collision. Potential corrective actions include closing some or all of these driveways and moving STOP lines closer to their normal positions.

Other potential causes of rear-end actions include signal timing (insufficient “yellow” and “all red” intervals), signal visibility (do trees block approaching drivers’ views?), and roadway lighting adequacy (given that most of the crashes occur at night).

Right-angle collisions indicate a breakdown in the right-of-way assignment by the signal. Signal visibility must be checked and the signal timing examined for reasonableness. Again, insufficient “yellow” and “all red” intervals could release vehicles before the competing vehicles have had time to clear the intersection. If the allocation of green is not reasonable, some drivers will “jump” the green or otherwise disregard it.

At this location, some of the causes compound each other. The setback of STOP lines to accommodate driveways, for example, lengthens the requirements for “all red” clearance intervals and, therefore, amplifies the effect of a shortfall in this factor.

This analysis is illustrative. The number of factors that can affect crash occurrence and/or severity at any given location is large indeed. A systematic approach, however, is needed if all relevant factors are to be identified and dealt with in an effective way. Traffic safety is not an isolated subject for study by traffic engineers. Rather, everything traffic engineers do is linked to a principal objective of safety. The importance of building safety into all traffic designs, control measures, and operational plans is emphasized throughout this text.

12.9 Closing Comments

This chapter provides a very general overview of the important subject of highway safety and crash studies. The subject is complex and covers a vast range of material. Everything the traffic engineer does, from field studies, to planning and design, to control and operations, is related to the provision of a safe system for vehicular travel. The traffic engineer is not alone in the focus on highway safety, as many other professionals, from urban planners to lawyers to public officials, also have an abiding interest in safe travel.

References

- 1. *Quick Facts 2016,* National Highway Safety Administration, U.S. Department of Transportation, Washington, D.C., 2017.
- 2. *Traffic Safety Facts Research Note: 2016 Motor Vehicle Crashes-Overview*, National Highway Safety Administration, U.S. Department of Transportation, Washington, D.C., October 2017.
- 3. *NSC Motor Vehicle Fatality Estimates*, National Safety Council, Chicago, IL, accessed on www.nsc.org/NewsDocuments/2017/12-month-estimate.
- 4. *Traffic Calming ePrimer*, Federal Highway Administration, U.S. Department of Transportation, Washington, D.C., 2017.
- 5. http://en/wikipedia.org/wiki/Vision_Zero.
- 6. http://safety.fhwa.dot.gov/tzd.
- 7. http://www.nsc.org/learn/NSC-initiatives/Pages/The-Road-to-Zero.aspx.
- 8. http://www.nycvzv.info.
- 9. *Highway Safety Manual*, 1st Edition, American Association of State Highway and Transportation Officials, Washington, D.C., 2010.
- 10. *Highway Safety Manual*, 1st Edition, Supplement, American Association of State Highway and Transportation Officials, Washington, D.C., 2014.
- 11. *Highway Safety Manual Data Needs Guide*, National Cooperative Highway Research Program, Transportation Research Board, Washington, D.C., June 2008.
- 12. *Guidance for Implementation of the AASHTO Strategic Highway Safety Plan*, NCHRP Report 500A, National Cooperative Highway Research Program, Transportation Research Board, Washington,

D.C., 2004.

- 13. *Crash Modification Factors Clearinghouse*, accessed at www.cmfclearinghouse.org.
- 14. Goodwin, A., et al, *Countermeasures That Work: A Highway Safety Countermeasures Guide for State Highway and Transportation Offices*, 8th Edition, National Highway and Transportation Administration, U.S. Department of Transportation, Washington, D.C., 2015.
- 15. http://www.safety.fhwa.dot.gov/provencountermeasures.
- 16. http://www.cmfclearinghouse.org.
- 17. http://library.ite.org/Pub/e26c7e9c-2354-d714-5181-4cc79fba5459.
- 18. Ogden, K.W., *Safer Roads: A Guide to Road Safety Engineering*, Avebury Technical, Brookfield, VT, 1996.

Problems

1. 12-1. Refer to Figures 12.1 and 12.2. Construct a graph of the number of fatalities expected if the fatality rate had stopped progressing downward in 1966, 1976, and 1986 respectively. Include the data of Figure 12.2 as a baseline. Comment.

2. 12-2. Early in this chapter it was said that 2014 to 2015 saw a noticeable "uptick" in both the fatality rate and the number of fatalities. At the time this chapter was written, various people viewed this as (a) a clear indication of the effects of distracted driving, including the spread of texting, cell phone usage, and other devices, (b) a statistical fluctuation in the pattern, or (c) other underlying factors not yet identified, including an increase in economic activity in a better economy. Given that you may well be reading this question a few years after that time period, use more recent data—and web searches of the literature—as the basis for commenting on the probable causality, if any. Try to be fact-based in your assessment, not merely speculative.

3. 12-3. Refer to Figure 12.4.

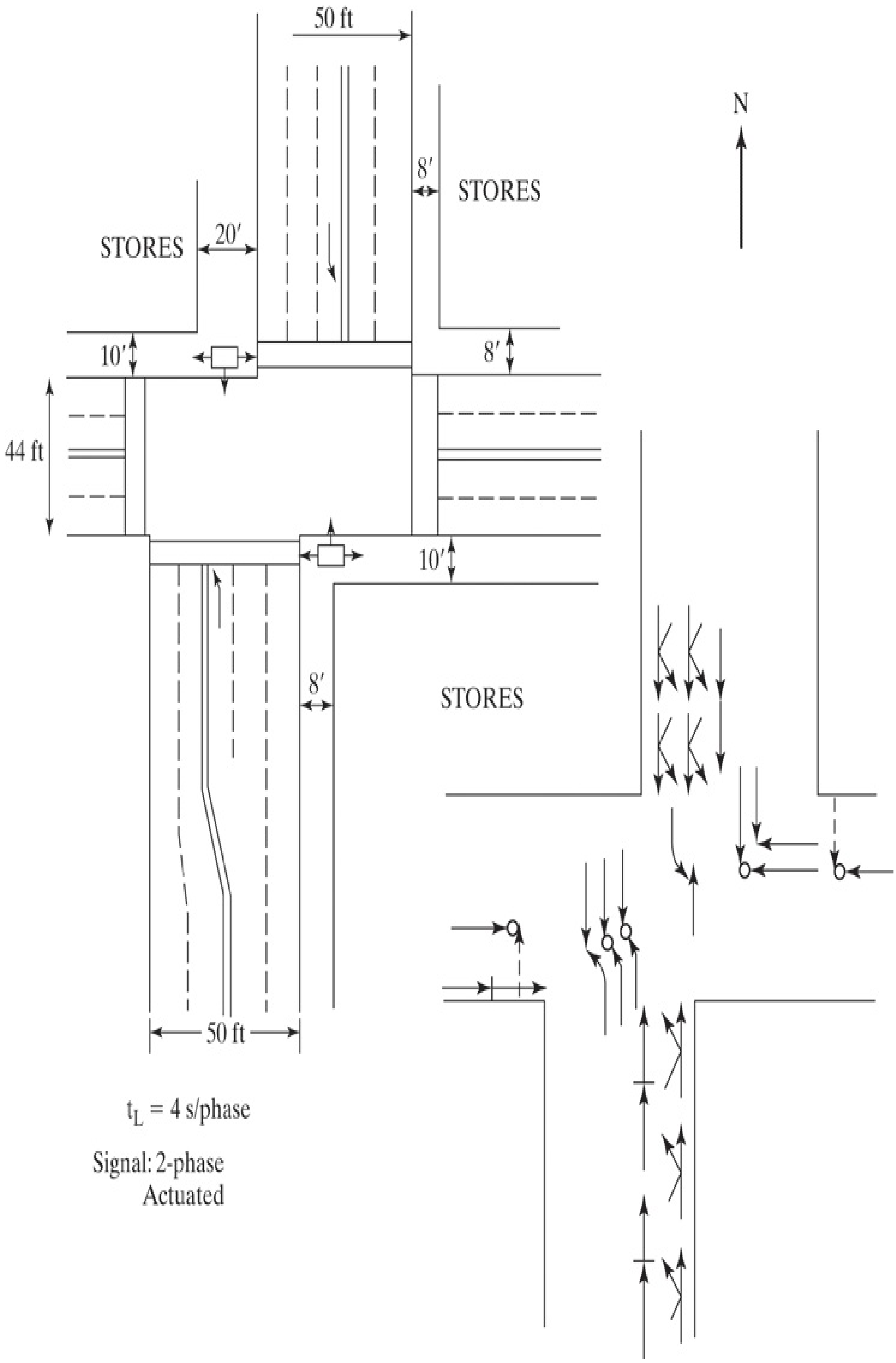
    1. Add the percentages for the single vehicle case. How is it possible that the percentage exceeds 100%? Is there a logical explanation, or is it an error? Hint: Read the title of the figure carefully, and the related words in the text. Explain it by a clear example using illustrative numbers to make your point, or explain why the graph is simply erroneous.

    2. If the graph is correct and reasonable, does it show the fatality rate by time of day (for instance, fatalities per million vehicle miles travelled), the distribution of fatalities over the day, or something else? What other data (if any) would you need to show the daily fatality rate by time of day, total and alcohol-related? What other data (if any) would you need to show the profile of fatalities by time of day, total and alcohol-related? Where would you seek this data (be specific as to the reference)?

4. 12-4. In this chapter, reference is made to “graduated licensing.” This refers to the practice adopted in a number of states of limiting the hours of operation or other features of drivers’ licenses, based upon age and/or other factors. Use the internet is to search for both current practices across the U.S., and conclusions about the safety effectiveness of the practice.
5. 12-5. Consider the following data for the year 2016 in a small suburban community:
    - Number of crashes 360
        - Fatal 10
        - Injury 36
        - PDO 314
    - Number of fatalities 15
    - Area population 50,000
    - Registered vehicles 35,000
    - Annual VMT 12,000,000
    - Average speed 30 mi/h

   Compute all relevant exposure- and population- based crash and fatality rates for this data. Compare these to national norms for the current year. (Hint: Use the Internet to locate current national norms.)
6. 12-6. Consider the collision and condition diagrams illustrated below. Discuss probable causes of the crashes observed. Recommend improvements, and illustrate them on a revised condition diagram.

Collision and Condition Diagram for Problem 12-6



Full Alternative Text

7. 12-7. For the intersection data provided below, estimate the annual number of crashes that would be expected at this 4-leg signalized intersection.

Condition	Value
Major Street AADT	60,000 veh/day
Minor Street AADT	25,000 veh/day
Red-Light Cameras	In place on all approaches
Number of Approaches with LT Lanes	4
Number of Approaches with RT Lanes	2
Number of Approaches with Exclusive LT Phasing	4
Number of Approaches with RTOR Prohibited	None
Intersection Lighting	Provided
Total Pedestrian Crossing Volume	1,800 peds/day
Maximum Number of Lanes Crossed by a Pedestrian	6 lanes
Number of Bus Stops within 1,000 ft of the Intersection	None
Number of Schools within 1,000 ft of the Intersection	1
Number of Liquor Stores within 1,000 ft of the Intersection	2
Local Calibration Factor	1.04

12.2-16 Full Alternative Text

Chapter 13 Parking: Characteristics, Studies, Programs, and Design

Almost every traveler starts and ends their trip as a pedestrian. For dense urban areas, many trips are accommodated by public transportation; some trips are completely made as pedestrians. For many less-dense urban areas, suburban and rural areas, the automobile remains the primary means by which trips are made.

With the exception of drive-through facilities now present at such varied destinations as banks and fast-food restaurants, auto travelers generally leave from their origins as pedestrians and enter their destinations as pedestrians. In terms of trips using private automobiles, the pedestrian portion of the trip starts and/or ends at a parking space.

At a residential trip end, private vehicles are accessed in private driveways, garages, on-street parking spaces, or nearby off-street lots or garages. At the other end of the trip, the location and nature of parking opportunities depends heavily on the land-use function and density as well as on a wide variety of public policy and planning issues.

For land to be productively used, it must be accessible. Auto accessibility is dependent on the supply, convenience, and cost of parking facilities. Major activity centers, from regional shopping malls to sports facilities to airports, rely on significant parking supply to provide site accessibility. Without such supply, these facilities could not operate profitably over a substantial period of time.

The economic survival of most activity centers, therefore, is directly related to parking and other forms of access. Parking supply must be balanced with other forms of access (public transportation), the traffic conditions created by such access, and the general environment of the activity center. Although economic viability is most directly related to the availability of parking, the environmental impacts of generated traffic may have negative effects as well.

This chapter attempts to provide an overview of issues related to parking. The coverage is not intended to be exhaustive, and the reader is encouraged to consult the available literature for more complete and detailed treatments of the subject. This chapter will address four key parking issues:

- Parking demand
- Parking studies
- Off-street parking facility design and location
- Parking management

Each of these is covered in the sections that follow.

13.1 Parking Demand

The key issue in parking is a determination of how many spaces are required for a particular development, and where they should be located. These requirements lead to locally based zoning regulations on minimum numbers of spaces that need to be provided when a development is built.

The need for parking spaces depends upon many factors, some of which are difficult to assess. The type and size of land use(s) in a development is a major factor, but so is the general density of the development environment and the amount and quality of public transportation access available.

13.1.1 Parking Generation

Just as land uses “generate” trips, they also “generate” a need for parking spaces. The most comprehensive source of parking-generation information is the Institute of Transportation Engineers’ *Parking Generation* [1]. The fourth edition of this basic reference was published in 2010, but updates are provided periodically, and the reader is encouraged to consult the latest edition directly for up-to-date criteria. Material in this text is based upon the fourth edition.

Parking generation relates the maximum observed number of occupied parking spaces to *one* underlying variable that is used as a surrogate for the size or activity level of the land use involved. Depending upon the land use involved, the most descriptive underlying variable is used, ranging from floor area to numbers of employees, or other relevant parameters. The underlying variables have been historically chosen based upon how well they statistically predict peak parking usage.

Reference [1] provides average parking-generation rates, ranges, and specific predictive algorithms for 106 different land uses. The data used to develop these values came primarily from single-use facilities in suburban areas with little or no public transportation access.

A summary of parking-generation rates and relationships, compiled from

Ref. [1], is shown in Table 13.1. Table 13.1 shows only a sample of the parking-generation data from *Parking Generation*. Data for many other uses is included in *Parking Generation*, but many categories are backed up with only small sample sizes. Even for those land uses included, the number of sites used to calibrate the values is not always significant, and the R2 values often connote significant variability in parking-generation rates.

Table 13.1: Typical Peak Parking-Generation Rates

Land Use	Avg Rate	Per X	Equation[2]	R^2	No. of Studies
Residential:					
Single-family detached housing	1.83	Dwelling unit	$3.2X - 8.6$	0.69	6
Condominium/Townhouse	1.38	Dwelling unit	$1.26X + 9$	0.95	12
Low/Mid-rise apartment (1–4 floors) *Weekday, Urban*	1.20	Dwelling unit	$0.92X + 4$	0.96	40
High-Rise Apartment (>4 floors)	1.37	Dwelling unit	$1.04X + 130$	0.85	7
Retail:					
Shopping center *Saturday-Dec.*	4.67	1,000 ft^2 GFA[1]	$4.60X + 115$	0.84	86

Shopping center *Saturday-Not Dec.*	2.87	1,000 ft^2 GFA[1]	3.38X − 116	0.98	26
Supermarket *Saturday-Suburban*	3.92	1,000 ft^2 GFA[1]	4.88X − 28	0.63	14
Supermarket *Weekday-Urban*	2.27	1,000 ft^2 GFA[1]	2.95X − 15	0.72	8
Home improvement superstore *Sat.*	3.19	1,000 ft^2 GFA[1]	Not available	Not available	40
Pharmacy w/Drive-thru window *Sat.*	2.18	1,000 ft^2 GFA[1]	Not available	Not available	16
Furniture store *Saturday*	1.04	1,000 ft^2 GFA[1]	0.87X + 4	0.90	7
Quality restaurant *Friday*	16.41	1,000 ft^2 GFA[1]	11.93X + 22	0.65	11
Family restaurant (high-turnover sit down *Saturday, Suburban*	13.50	1,000 ft^2 GFA[1]	Not available	Not available	13
Fast-food restaurant w/Drive-thru *Weekday*	9.98	1,000 ft^2 GFA[1]	Not available	Not available	27
Institutional:					
Elementary school	0.17	Student	Not available	Not available	5
Middle/Junior high school	0.09	Student	Not available	Not available	3
High school	0.23	Student	Not available	Not available	8
Community college	0.18	School population[2]	Not available	Not available	12
University/College *Suburban*	0.33	School population[2]	0.33X − 49	0.97	8
Church	8.37	1,000 ft^2 GFA[1]	Not available	Not available	12
Day care center *Weekday*	0.24	Student	0.30X − 5	0.72	39
Museum *Saturday*	1.32	1,000 ft^2 GFA[1]	Not available	Not available	4
Library *Weekday, Suburban*	2.62	1,000 ft^2 GFA[1]	1.48X + 27	0.63	7
Hospital *Weekday. Suburban*	0.81	Employee	0.69X + 253	0.74	15
Nursing home	0.98	1,000 ft^2 GFA[1]	0.54X + 17	0.77	13
Office:					
Office building *Weekday, Suburban*	2.84	1,000 ft^2 GFA[1]	2.51X + 26	0.91	176
Medical/Dental office building	3.20	1,000 ft^2 GFA[1]	0.34X − 13	0.91	86
Government office building	4.15	1,000 ft^2 GFA[1]	Not available	Not available	4
Lodging:					
Hotel *Saturday, Suburban*	1.20	Occupied room	Not available	Not available	4
Resort hotel	1.29	Occupied room	0.97X + 76	0.76	5
Recreational:					
Golf course *Saturday*	8.68	Hole	9.08X − 5	0.90	7
Bowling alley *Friday, Suburban*	5.02	Lane	6.63X − 35	0.97	4
Multiplex movie theater *Friday*	0.15	Seat	Not available	Not available	6
Ski (Snow) area *Saturday*	1.31	Acre	1.10X + 27	0.86	4
Health/Fitness club (gym)	5.27	1,000 ft^2 GFA[1]	3.35X + 38	0.60	25

[1]GFA = gross floor area

[2]School Population = students + employees + visitors

(Source: Used with permission from *Institute of Transportation Engineers Parking Generation*, 4th Edition, Institute of Transportation Engineers, Washington D.C., 2010)

Table 13.1: Full Alternative Text

Where practical, local data should be used to modify nationally representative rates. Many local planning agencies will have such data available, although the quality and timeliness will vary widely.

Note that in some cases, modifiers describing the setting (urban, suburban, and rural) of the land use and/or the day (weekday, Saturday, and Sunday) are shown. *Parking Generation* contains information for other time periods as well. In general, the peak settings and time periods were chosen for inclusion in Table 13.1.

Sample Problem 13-1: Parking Generation Estimation

Consider the case of a general office building, consisting of 50,000 sq ft of office space. What is the peak parking load expected to be at this facility? From Table 13.1 for office buildings, the average peak parking occupancy is 2.84 per thousand sq ft of building area, or in this case, $2.84\times50=142$ parking spaces. A more precise estimate might be obtained using the equation related to facility size:

$$P=2.51X+26=(2.51\times50)+26=151.5,\text{say }\ 152\ \text{ spaces}$$

This presents a modest range to the engineer—from 142 to 152 parking spaces needed. In this case, the table data were based upon a large sample size of 176 sites, and the regression coefficient (R^2) is strong—0.91. This prediction, therefore, may be considered to be relatively reliable.

In other cases, small sample sizes or weak R^2 values might lead the analyst to look for some local data for comparison purposes.

Table 13.1 contains information on three different types of restaurants—quality, family, and fast-food. The parking-generation rates for these vary considerably depending upon the day of the week, and the general setting in which the restaurant is found. Table 13.2 shows average parking-generation rates for restaurants per 1,000 ft^2 of gross floor area.

Table 13.2: Parking-Generation Rates per 1,000 ft^2 GFA for Restaurants

	Quality Restaurant	Family Restaurant	Fast-Food Restaurant
Mon–Thurs, Urban	10.60	5.55	9.98
Mon–Thurs, Suburban	10.60	10.60	9.98
Friday, Urban	16.41	5.55	9.98
Friday, Suburban	16.41	10.60	9.98
Saturday, Urban	16.40	NA	8.70
Saturday, Suburban	16.40	13.50	8.70

(Source: Used with permission from Institute of *Transportation Engineers Parking Generation*, 4th Edition, Institute of Transportation Engineers, Washington D.C., 2010.)

Table 13.2: Full Alternative Text

For general purposes, a “quality restaurant” is a sit-down facility catering to an adult population, which usually includes a bar. It can be a stand-alone facility or part of a regional or national chain. “Family restaurant” connotes a higher turnover rate, and a facility catering to families with or without children. Many are part of national chains (Applebee’s, Chili’s, Ruby Tuesday’s, Bob Evans, etc.), but they can be stand-alone or part of a local or regional chain as well. The rates in Table 13.2 are for such facilities without bars. “Fast-food restaurants” obviously represent very high turnover rates, and are often part of national chains, although they need not be. The rates in Table 13.2 are for fast-food facilities that have drive-thru windows. It should be noted that *Parking Generation* includes several additional restaurant categories.

For quality restaurants, Fridays and Saturdays represent the peak parking needs, which are approximately 60% higher than needs for other days of the week. Family restaurants show significantly different parking-generation rates for urban and suburban settings. It is precisely because of these variables that wherever possible local data on parking should be collected and reviewed to fine-tune published national averages.

It should be noted that these parking-generation rates are keyed on the maximum number of parking spaces used during peak periods. From the user perspective, a parking facility is often perceived as “full” when 95% of its spaces are filled. It is, therefore, common practice to infer that parking demand should include some allowance for between 5% and 10% empty spaces, even during peak demand periods [2].

13.1.2 Shared Parking

The parking-generation rates cited in Tables 13.1 and 13.2 reflect the parking needs of individual isolated facilities. Frequently, however, different facilities are in locations that can easily share parking spaces. The classic case is the shopping center, where several (or many) different stores and ancillary services (restaurants, banks, etc.) share parking. This offers obvious efficiencies, as a given space can serve multiple land uses, as long as the time of need is different.

The concept of shared parking simply means that a parking facility is used to satisfy the parking demand for multiple users or destinations [3, 4].

Consider an example: an individual living in an apartment house can pay a monthly fee for a parking space in an adjacent facility. There are two ways in which that can be accomplished. The individual can be given exclusive use of a particular numbered space or the individual can be given use of shared spaces, with a guarantee that one will always be available. The first option is the most inefficient—a separate space must be provided for each user, even if all users are never present at the same time. The second approach can lead to requiring fewer spaces to accommodate the demand, and can result in cheaper monthly fees to users.

There are many other scenarios in which parking can be shared. Land uses that generate peak demands on different days, in different seasons, and/or at different times of the day can share the use of the same parking spaces. Table 13.3 shows a sampling of various land uses that have peak parking demands on different days and times of the week.

Table 13.3: Typical Peak Parking Periods for Various Land Uses

Weekdays	Evenings	Weekends
• Banks and public services	• Auditoriums	• Religious institutions
• Offices and other employment centers	• Bars and dance halls	• Parks
• Park n' ride facilities	• Meeting halls	• Shops and malls
• Schools, daycare centers, and colleges	• Quality restaurants	
• Factories and distribution centers	• Theaters	
• Medical clinics	• Hotels	
• Professional services		

(Source: Used with permission from Victoria Transport Policy Institute Litman, T., *Parking Management Best Practices,* American Planning Association, Chicago IL, 2006, Table 5-1, pg 67)

Table 13.3: Full Alternative Text

Table 13.4 shows parking needs for various types of shopping centers. As shown, as the size and scale of the shopping center gets larger, the parking needs also increase (per 1,000 ft^2 GFA). This indicates that given the greater diversity of shops and services of larger shopping centers, the intensity of parking demand is also greater.

Table 13.4: Parking Needs for Shopping Centers per 1,000 ft^2 GFA

Type of Shopping Center	Parking Spaces/ 1,000 ft^2 GFA	Number of Study Sites
Strip	4.1	5
Neighborhood	4.7	8
Community	4.9	51
Regional	5.5	27
Super regional	5.1	22

(Source: Used with permission from Institute of *Transportation Engineers Parking Generation*, 4th Edition, Institute of Transportation Engineers, Washington D.C., 2010, pg 227)

Table 13.4: Full Alternative Text

For Table 13.4, the following definitions apply:

- Strip: <30,000 ft^2, anchored by a small business.
- Neighborhood: 30,000 to 100,000 ft^2, anchored by a supermarket and/or drug store.
- Community: 100,000 to 400,000 ft^2, anchored by general merchandise stores or discount retailer.
- Regional: 400,000 to 800,000 ft^2, anchored by a department store.
- Super Regional: > 800,000 ft^2, anchored by several department stores.

Other studies have produced even more detailed results. A 1998 study provides additional data on parking generation of shopping centers [5]. Over 400 shopping centers were surveyed, resulting in the establishment of recommended "parking ratios," the number of spaces provided per 1,000 sq ft of GLA. Centers were categorized by total size (in GLA), and by the percentage of total center GLA occupied by movie houses, restaurants, and other entertainment uses. The results are summarized in Table 13.5.

Table 13.5: Recommended Parking Ratios from a 1998 Study

Center Size (Total GLA)	Percent Usage by Movie Houses, Restaurants, and Other Entertainment				
	0%	**5%**	**10%**	**15%**	**20%**
0–399,999	4.00	4.00	4.00	4.15	4.30
400,000–419,999	4.00	4.00	4.00	4.15	4.30
420,000–439,999	4.06	4.06	4.06	4.21	4.36
440,000–459,999	4.11	4.11	4.11	4.26	4.41
460,000–479,999	4.17	4.17	4.17	4.32	4.47
480,000–499,999	4.22	4.22	4.22	4.37	4.52
500,000–519,999	4.28	4.28	4.28	4.43	4.58
520,000–539,999	4.33	4.33	4.33	4.48	4.63
540,000–559,999	4.39	4.39	4.39	4.54	4.69
560,000–579,000	4.44	4.44	4.44	4.59	4.74
580,000–599,999	4.50	4.50	4.50	4.65	4.80
600,000–2,500,000	4.50	4.50	4.50	4.65	4.80

(Source: Used with permission of Urban Land Institute, *Parking Requirements for Shopping Centers*, 2nd Edition, Washington, D.C., 1999, compiled from Appendix A, Recommended Parking Ratios.)

Table 13.5: Full Alternative Text

The guidelines were established such that the 20th peak parking hour of the year is accommodated (i.e., there are only 19 hours of the year when parking demand would exceed the recommended values). Parking demands accommodate both patrons and employees.

Where movie theaters, restaurants, and other entertainment facilities occupy more than 20% of the GLA, a more detailed shared parking approach is recommended. Parking requirements would be predicted for shopping facilities, and for movies, restaurants, and entertainment facilities separately. Local studies would be used to establish the amount of overlapping usage that might occur.

Sample Problem 13-2: Parking Generation for a Regional Shopping Center (1)

Consider the following case: a new regional shopping center with 1,000,000 sq ft of GLA is to be built. It is anticipated that about 15% of the GLA will be occupied by movie theaters, restaurants, or other entertainment facilities. How many parking spaces should be provided?

From Table 13.4, the center as described (a super-regional shopping center) would generate a need for 5.1 spaces per 1,000 ft^2 GLA:

$$P = 5.1 \times \frac{1,000,000}{1,000} = 5,100 \text{ parking spaces}$$

Using Table 13.5, 4.65 spaces per 1,000 ft^2 GFA would be needed, or:

$$P = 4.65 \times \frac{1,000,000}{1,000} = 4,650 \text{ parking spaces}$$

The data in Table 13.4 is newer, but Table 13.5 is more detailed. Both get in the same general ballpark, but some local or regional data should be examined to fine-tune the analysis, particularly since the difference in the two estimates is 450 parking spaces, a significant amount.

Reference [6] presents an even more detailed model for predicting peak parking needs. As the model is more detailed, additional input information is needed in order to apply it. Peak parking demand may be estimated as:

$P=N\times K\times R\times A\times prO$ [13-1]

where:

P=parking demand, spaces,N=size of activity measured in appropriate units
use parameters),K=portion of destinations that occur at any one time,R=per
destinations per day (or other timeperiod) per unit of activity,A=proportion

Sample Problem 13-3: Parking Generation for a Regional Shopping Center (2)

Consider the case of the same 1,000,000 ft^2 retail shopping center in the heart of a Central Business District (CBD). The following additional information has been collected:

- Approximately 40% of all shoppers are in the CBD for other reasons (pr=0.40).
- Approximately 70% of shoppers travel to the retail center by automobile (A=0.70).
- Approximate total activity at the center is estimated to be 45 person-destinations per 1,000 sq ft of gross leasable area, of which 20% occur during the peak parking accumulation period (R=45; K=0.20).
- The average auto occupancy of travelers to the shopping center is 1.5 persons per car (0=1.5).

As the unit of size is 1,000 sq ft of gross leasable area, N=1,000,000/1,000=1,000 for this illustration. The peak parking demand may now be estimated using Equation 13.1 as:

$$P = \frac{1{,}000 \times 45 \times 0.20 \times 0.70 \times 0.60}{1.5} = 2{,}520 \text{ parking spaces}$$

The result of Sample Problem 13-3 is considerably less than the results of Sample Problems 13-1 and 13-2. There are several reasons for this. The center is in a an urban area, with 40% of shoppers already there for other reasons (e.g., they work in the area), and 30% use public transportation or walk directly to the location. The more general data of Tables 13.4 and 13.5 assume that most people are arriving specifically to shop at the center, and all are arriving by car. The auto occupancy of 1.5 persons/vehicle is also somewhat higher than would be normally expected in a suburban setting.

Although this technique is analytically interesting, it requires that a number of estimates be made concerning parking activity. For the most part, these would be based on data from similar developments in the localized area or region or on nationwide activity information if no local information is available.

The point is that all estimates of parking demands, including those for shared parking require some knowledge of local and regional characteristics. Although national average data is a significant help, there is a great deal of variation in parking demand that is related to specific local characteristics.

13.1.3 Zoning Regulations

Control of parking supply for significant developments is generally maintained through zoning requirements. Local zoning regulations generally specify the minimum number of parking spaces that must be provided for developments of specified type and size. Zoning regulations also often specify needs for handicapped parking and set minimum standards for loading zones.

Most zoning regulations require that a specific number of parking spaces be provided for an individual facility, although shopping centers may be included as a "single facility." There are, however, options to individual requirements for each individual facility. Local policy can, and often should, encourage shared parking approaches.

Todd Litman, in *Parking Management Best Practices* [3, 4] suggests a number of approaches that can accomplish this:

- Agreements between various sites to share a single (or multiple) parking facilities within walking distance (to all the sites) can be encouraged. As noted, this works particularly well when the predominant land uses at the cooperating sites have peak demands at different times.

- Developers can be made to pay fees *in lieu* of building exclusive parking spaces for their buildings. These can be pooled to support construction of public parking lots or garages that serve multiple sites. Such public sites, however, must be provided and become the responsibility of a local agency. Often, a specific local agency is established to construct, maintain, and operate these public parking facilities.

- Businesses within a defined area can be placed in a downtown business improvement district, and taxed to provide funds for public parking facilities.

It is, of course, possible to employ several of these approaches simultaneously. When public parking facilities are provided, parking fees are generally charged to users to help defray the operating expenses of the parking agency.

The public parking approach makes the most sense in dense urban areas where it would be difficult to require that each development parcel accommodate both the functional building and adequate parking to handle the demand. In such areas, the density of development makes it possible for many destinations to be adequately served by a single parking facility.

In more suburban areas, where densities are far lower, a more traditional approach of requiring each development to provide adequate parking is generally followed.

Table 13.6 is a compilation of common zoning requirements governing parking for various land uses [2, 7]. In general, each jurisdiction (state, county, and local) will establish a full set of zoning requirements, which will include parking. For any specific area, local requirements in effect at the time of construction would be used. The recommended zoning requirements reflect the 85th percentile of peak demands observed, plus 5% to 10% cushion for unused parking spaces during peak demand hours.

Table 13.6: Example Zoning Requirements for Parking

Land Use	Recommended Zoning Requirement
Residential	
Single-family dwelling unit	2.00 per dwelling unit
Multifamily dwelling unit—Studio	1.25 per dwelling unit
Multifamily dwelling unit—1 bedroom	1.50 per dwelling unit
Multifamily dwelling unit—2 or more bedrooms	2.00 per dwelling unit
Accessary dwelling unit	1.00 per dwelling unit
Sleeping room	1.00 per dwelling unit + 2 for owners and managers
Elderly housing	0.50 per dwelling unit
Group and nursing homes	0.33 per resident
Day care center	1.00 per employee + 1.2 per licensed capacity enrollment + 1 drop-off space
Commercial lodgings	1.25 per sleeping room/unit + 10 per 1,000 ft^2 restaurant/lounge space + 20 per 1,000 ft^2 meeting room space + 30 per 1,000 ft^2 of exhibit/ballroom space
Hospital/Medical center	0.50 per employee + 0.33 per bed + 0.20 per outpatient + 0.25 medical staff + 1.00 per student/medical staff (medical centers only)
Retail	
General retail	3.30 per 1,000 ft^2 GFA
Convenience retail	4.00 per 1,000 ft^2 GFA
Service retail	2.40 per 1,000 ft^2 GFA
Hard goods retail	2.50 per 1,000 ft^2 of interior sales space + 1.50 per 1,000 ft^2 of interior storage and exterior display/storage space
Shopping center	See Table 11-7
Food and Beverage	
Fine dining	20.0 per 1,000 ft^2 gross leasable area (GLA)
Eating and drinking	25.0 per 1,000 ft^2 GLA
Family restaurant	12.0 per 1,000 ft^2 GLA
Fast food	10.0 per 1,000 ft^2 GLA (including kitchen, counter, waiting areas) + 0.50 per seat provided
Office and Business Services	
General business	3.60 per 1,000 ft^2 GFA up to 30,000 ft^2; thereafter 3.00 per 1,000 ft^2 GLA
Consumer service	4.00 per 1,000 ft^2 GFA up to 30,000 ft^2; thereafter 3.60 per 1,000 ft^2 GLA
Data processing/Telemarketing/Operations	7.00 per 1,000 ft^2 GFA up to 30,000 ft^2; thereafter 6.00 per 1,000 ft^2 GLA
Medical offices (not part of hospital or medical center)	6.00 per 1,000 ft^2 GFA up to 5,000 ft^2; thereafter 5.50 per 1,000 ft^2 GLA
Medical offices (part of hospital or medical center)	5.50 per 1,000 ft^2 GFA up to 5,000 ft^2; thereafter 5.00 per 1,000 ft^2 GLA
Industrial	2.00 per 1,000 ft^2 GFA + any required spaces for offices, sales, etc.
Educational	
Elementary or secondary	1.20 per classroom + 0.25 per student over driving age
College or university	To be established by the zoning administrator based on a study of parking needs

Cultural/Entertainment/Recreational	
Convention center	20.0 per 1,000 ft^2 GLA of exhibit, ballroom, and meeting space
Public assembly	0.25 per person at permitted capacity
Cinemas—Single screen	0.50 per seat
Cinemas—Up to 5 screens	0.33 per seat
Cinemas—> 5 screens	0.29 per seat
Theaters (live performance)	0.50 per seat
Arenas and stadiums	0.33 per seat
Recreation facilities	2.00 per player, or 0.33 per person at permitted capacity

(Source: Used with permission from Springer Science and Business Media Chrest A., et al, *Parking Structures: Planning, Design, Construction, Maintenance, and Repair*, 3rd Edition, Springer Science and Business, New York NY. 2001, Table 2-1, pgs 12 and 13 as adapted from *Recommended Zoning Ordinance Provisions for Off-Street Loading Space*, National Parking Association, Washington D.C., 1995)

Table 13.6: Full Alternative Text

Sample Problem 13-4: Zoning Requirements for a Regional Shopping Center

Consider the case of a regional shopping center of 500,000 ft^2 GFA, of which 450,000 ft^2 is leasable (GLA). Fifteen (15%) percent of the leasable space is used for restaurants and movie theaters.

The peak parking demand can be estimated from Table 13.4 or 13.5. From Table 13.4, a regional shopping center would generate a peak parking demand of 5.5 spaces per 1,000 ft^2 of GFA, or:

$$P = 5.5 \times \frac{500{,}000}{1{,}000} = 2{,}750 \text{ parking spaces}$$

From Table 13.5, 4.26 parking spaces per 1,000 ft^2 GLA would be needed, or:

$$P=4.26\times 450{,}0001{,}000=1{,}917\ \text{parking spaces}$$

From Table 13.7, a typical zoning statute would require 4.0+5(0.03)=4.15 parking spaces per 1,000 ft^2 GLA, or:

$$P=4.15\times 450{,}0001{,}000=1{,}868\ \text{parking spaces}$$

Table 13.7: Example Zoning Requirements for Shopping Center Parking (Spaces per 1,000 ft^2 GLA)

Shopping Center Size In ft^2 GLA	Percentage of GLA in Restaurant, Entertainment, and/or Cinema Space		
	0%–10%	11%–20%	>20%
<400,000	4.0		Shared parking
400,000–599,999	4.0–4.5 sliding scale[a]		Shared parking
≥ 600,000	4.5	4.5	Shared parking

[a]For each % above 10%, add 0.03 spaces per 1,000 ft^2 GLA.

(Source: Used with permission from Springer Science and Business Media Chrest A., et al, Parking Structures: Planning, Design, Construction, Maintenance, and Repair, 3rd Edition, Springer Science and Business, New York NY. 2001, Table 2-2, pg 16)

Table 13.7: Full Alternative Text

This is a considerable range of values. It helps, however, to understand the differences between the tables used in these estimates. Table 13.4, which produced 2,750 parking spaces, is an estimate of peak demand based upon general gross floor area (GFA). Table 13.7 is a zoning recommendation, which is based upon an 85th percentile parking peak—one that is exceeded 15 days per year (think Christmas holiday shopping). It also accounts for the partial shared parking effect of restaurants and movie theaters. This produces a more conservative value. Table 13.5, which cites recommended parking ratios is more like a zoning regulation than a parking demand estimate, and it also takes into account the effect of shared uses.

The data herein, as have been noted, reflect national averages and common practices. For any specific site, in this case a 500,000 ft^2 regional shopping center, local issues, and characteristics would have to be considered in reaching a final demand estimate, or to set a local zoning policy.

The recommended zoning requirements of Tables 13.6 and 13.7 would be significantly lower in urban areas with good transit access, captive walk-in patrons (people working or living in the immediate vicinity of the development), or organized car-pooling programs. In such areas, the modal split characteristics of users must be determined, and parking spaces may be reduced accordingly. Such a modal split estimate must consider local conditions, as this can vary widely. In a typical small urban community, transit may provide 10% to 15% of total access; in Manhattan (New York City), fewer than 5% of major midtown and downtown access is by private automobile.

13.1.4 Handicapped Parking Requirements

In any parking facility, handicapped spaces must be provided as required by federal and local laws and ordinances. Such standards affect both the number of spaces that must be required and their location. The Institute of Transportation Engineers [6] recommends the following minimum standards for provision of handicapped spaces:

- *Office*—0.02 spaces per 1,000 sq ft GFA
- *Bank*—1 to 2 spaces per bank

- *Restaurant*—0.30 spaces per 1,000 sq ft GFA
- *Retail* (< 500,000sq ft GFA)—0.075 spaces per 1,000 sq ft GFA
- *Retail* (≥500,000sq ft GFA)—0.060 spaces per 1,000 sq ft GFA

In all cases, there is an effective minimum of one handicapped space.

13.2 Parking Studies and Characteristics

There are a number of characteristics of parkers and parking that have a significant influence on planning. Critical to parking supply needs are the duration, accumulation, and proximity requirements of parkers. Duration and accumulation are related characteristics. If parking capacity is thought of in terms of “space-hours,” then vehicles parked for a longer duration consume more of that capacity than vehicles parked for only a short period. In any area, or at any specific facility, the goal is to provide enough parking spaces to accommodate the maximum accumulation on a typical day.

13.2.1 Proximity: How Far Will Parkers Walk?

Maximum walking distances that parkers will tolerate vary with trip purpose and urban area size. In general, tolerable walking distances are longer for work trips than for any other type of trip, perhaps because of the relatively long duration involved. Longer walking distances are tolerated for off-street parking spaces as opposed to on-street (or curb) parking spaces. As the urban area population increases, longer walking distances are experienced.

The willingness of parkers to walk certain distances to (or from) their destination to their car must be well understood, as it will have a significant influence over where parking capacity must be provided. Under any conditions, drivers tend to seek parking spaces as close as possible to their destination. Even in cities of large population (1,000,000–2,000,000), 75% of drivers park within 0.25 mile of their final destination.

Table 13.8 shows the distribution of walking distances between parking places and final destinations in urban areas. The distribution is based on studies in five different cities (Atlanta, Pittsburgh, Dallas, Denver, and

Seattle), as reported in Ref. [6].

Table 13.8 CBD Walking Distances to Parking Spaces

Distance		% Walking This Distance or Further	
Feet	Miles	Mean	Range
0	0	100	
250	0.05	70	60–80
500	0.10	50	40–60
750	0.14	35	25–45
1,000	0.19	27	17–37
1,500	0.28	16	8–24
2,000	0.38	10	5–15
3,000	0.57	4	0–8
4,000	0.76	3	0–6
5,000+	0.95+	1	0–2

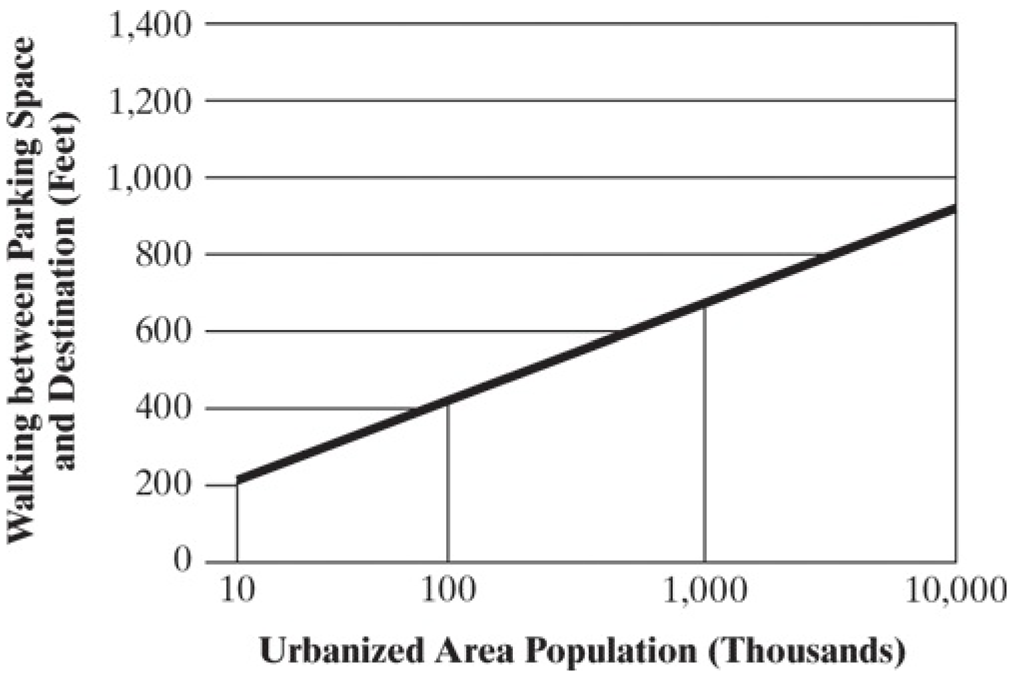
(Sources: Used with permission of Eno Foundation for Transportation, Weant, R., and Levinson, H., *Parking*, Westport, CT, 1990, Table 6-3, pg 98.)

Table 13.8 Full Alternative Text

As indicated in this table, parkers like to be close to their destination. One-half (50%) of all drivers park within 500 ft of their destination. Figure 13.1 shows average walking distances to and from parking spaces versus the total urban area population.

Figure 13.1 Average Walking Distance by Urbanized Area

Population



(Source: Used with permission of Eno Foundation for Transportation, Weant, R., and Levinson, H., *Parking*, Westport, CT, 1990, Figure 6.5, pg 98.)

Figure 13.1 Full Alternative Text

Again, this data emphasizes the need to place parking capacity in close proximity to the destination(s) served. Even in an urban region of over 10,000,000 population, the average walking distance to a parking place is approximately 900 ft.

Trip purpose and trip duration also affect the walking distances drivers are willing to accommodate. For shopping or other trips where things must be carried, shorter walking distances are sought. For short-term parking, such as to get a newspaper or a take-out order of food, short walking distances are also sought. Drivers will not walk 10 minutes if they are going to be parked for only 5 minutes. In locating parking capacity, general knowledge of parkers' characteristics is important, but local studies would provide a more accurate picture. In many cases, however, application of common sense and professional judgment is also an important component.

Table 13.9 shows the results of two more recent foreign studies [8, 9] which display similar walking characteristics.

Table 13.9 Walking Distances from Two Foreign Studies

Trip Purpose	Study 1 (Netherlands)	Study 2 (Indonesia)
Weekly shopping	587 ft	1,174 ft
Nonweekly shopping	1,230 ft	
Work	744 ft	847 ft
Business	NA	888 ft
Recreation/Social activities	670 ft	1,234 ft

Table 13.9 Full Alternative Text

Note that for both studies, walking tolerance was converted to average distances from a distribution (Netherlands) and walking time (Indonesia). For the latter, a walking speed of 4.0 ft/s was used.

In all cases, it is clear that parkers are loathe to walk great distances, with a range of 1000 to 1200 ft representing a practical outer limit. For many trip purposes, however, even these distances would be considered intolerable.

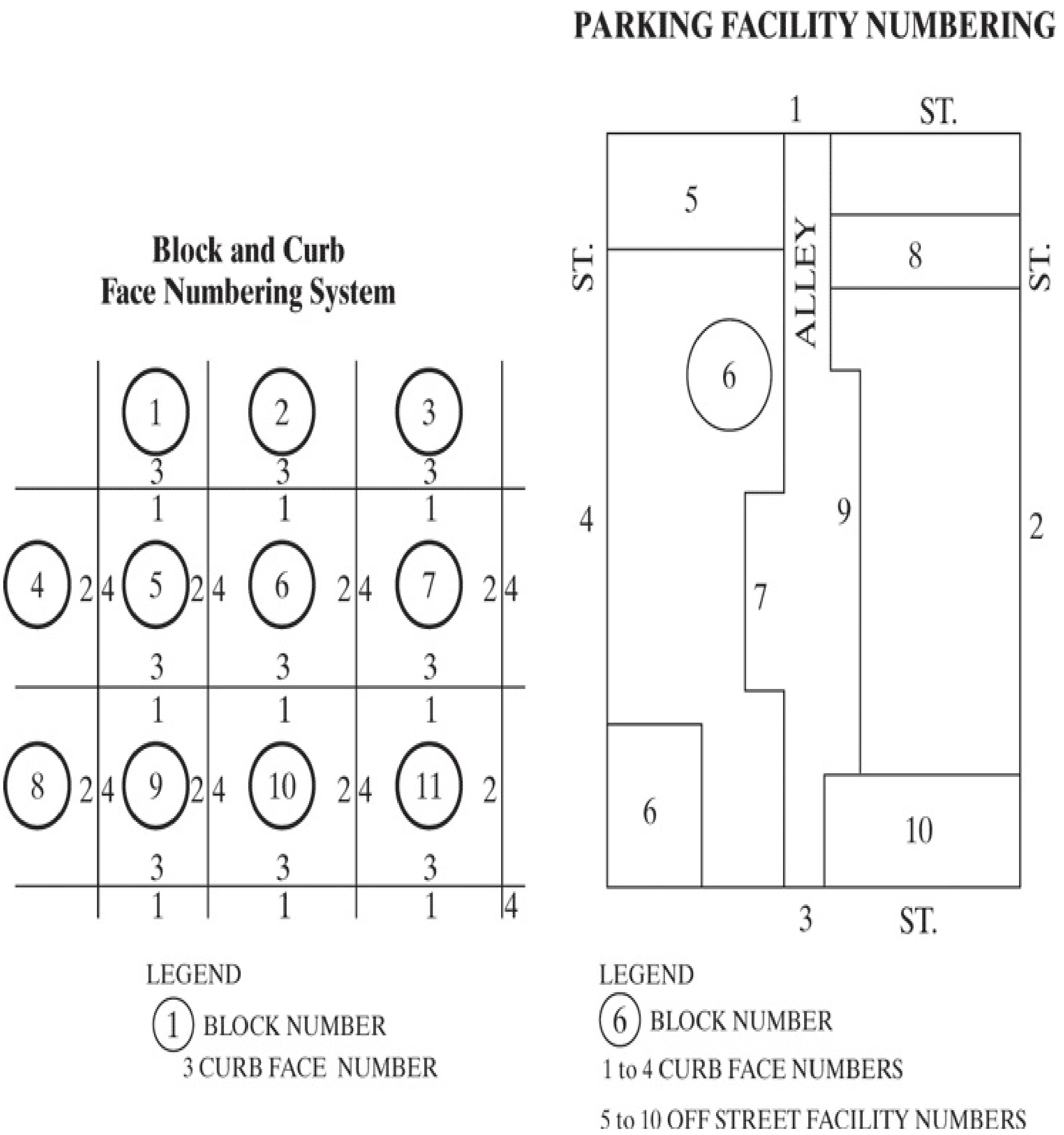
This is a subject about which local knowledge would be quite important. Interview studies would be employed to collect information locally. Parkers could be quickly questioned at their destination to determine how far they are walking. A few additional questions on the acceptability of their walking trip, and distance(s) they would be willing to walk under various scenarios would also be possible. As with most interview studies, the interview must be short and nonthreatening, and local police must be notified that it is being carried out. In some jurisdictions, a prior permit must be obtained before conducting such interview surveys.

13.2.2 Parking Inventories

One of the most important studies to be conducted in any overall assessment of parking needs is an inventory of existing parking supply. Such inventories include observations of the number of parking spaces and their location, time restrictions on use of parking spaces, and the type of parking facility (e.g., on-street, off-street lot, and off-street garage). Most parking inventory data are collected manually, with observers canvassing an area on foot, counting and noting curb spaces and applicable time restrictions, as well as recording the location, type, and capacity of off-street parking facilities. Use of intelligent transportation system technologies have begun to enhance the quantity of information available and the ease of accessing it. Some parking facilities have begun to use electronic tags (such as EZ Pass) to assess fees. Such a process, however, can also keep track of parking durations and accumulations on a real-time basis. Smart parking meters can provide the same types of information for curb parking spaces.

To facilitate the recording of parking locations, the study area is usually mapped and precoded in a systematic fashion. Figure 13-2 illustrates a simple coding system for blocks and block faces. Figure 13-3 illustrates the field sheets that would be used by observers.

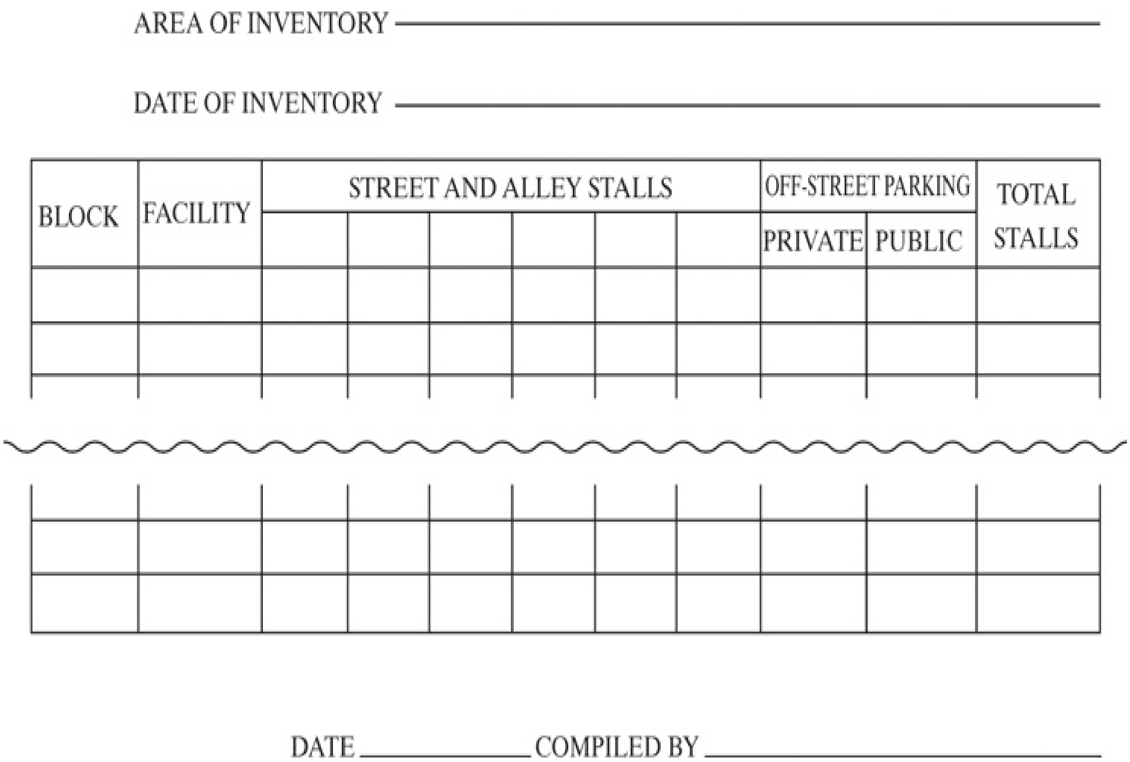
Figure 13.2: Illustrative System for Parking Location Coding



(Source: Used with permission of Institute of Transportation Engineers, Box. P. and Oppenlander, J., *Manual of Traffic Engineering Studies*, 4th Edition, Washington, D.C., 1976, Figures 10-1 and 10-2, pg 131.)

Figure 13.2: Full Alternative Text

Figure 13.3: A Parking Inventory Field Sheet



AREA OF INVENTORY ______

DATE OF INVENTORY ______

BLOCK	FACILITY	STREET AND ALLEY STALLS						OFF-STREET PARKING		TOTAL STALLS
								PRIVATE	PUBLIC	

DATE ______ COMPILED BY ______

(Source: Used with permission of Institute of Transportation Engineers, Box P. and Oppenlander, J., *Manual of Traffic Engineering Studies*, 4th Edition, Washington, D.C., 1976, Figure 10-3, pg 133.)

Figure 13.3: Full Alternative Text

Curb parking places are subdivided by parking restrictions and meter duration limits. Where several lines of a field sheet are needed for a given block face, a subtotal is prepared and shown. Where curb spaces are not clearly marked, curb lengths are used to estimate the number of available spaces, using the following guidelines:

- Parallel parking: 23 ft/stall
- Angle parking: 12.0 ft/stall
- 90-degree parking: 9.5 ft/stall

Although the parking inventory basically counts the number of spaces available during some period of interest—often the 8- to 11-hour business day—parking supply evaluations must take into account regulatory and time restrictions on those spaces and the average parking duration for the area. Total parking supply can be measured in terms of how many vehicles can be parked during the period of interest within the study area:

$$P=(\sum nNTD)\times F \quad [13\text{-}2]$$

where:

P=parking supply, vehs,N=number of spaces of a given type and timerestric tion are available during the study period, hrs,D=average parking duration c −values range from 0.85 to 0.95 and increase asaverage duration increases.

Sample Problem 13-5: Estimating Parking Supply

Consider an example in which an 11-hour study of an area revealed that there were 450 spaces available for the full 12 hours, 280 spaces available for 6 hours, 150 spaces available for 7 hours, and 100 spaces available for 5 hours. The average parking duration in the area was 1.4 hours. An efficiency factor of 0.90 will be used. Parking supply in this study area is computed as:

$$P=\{ [(450\times 12)+(280\times 6)+ (150\times 7)+(100\times 5)]1.4 \}\times 0.90=5{,}548 \text{ vehicles}$$

This result of Sample Problem 13-5 means that 5,548 vehicles could be parked in the study area over the 11-hour period of the study. It *does not* mean that all 5,548 vehicles could be parked at the same time. This analysis, however, requires that the average parking duration be known. Determining this important factor is discussed in the next section.

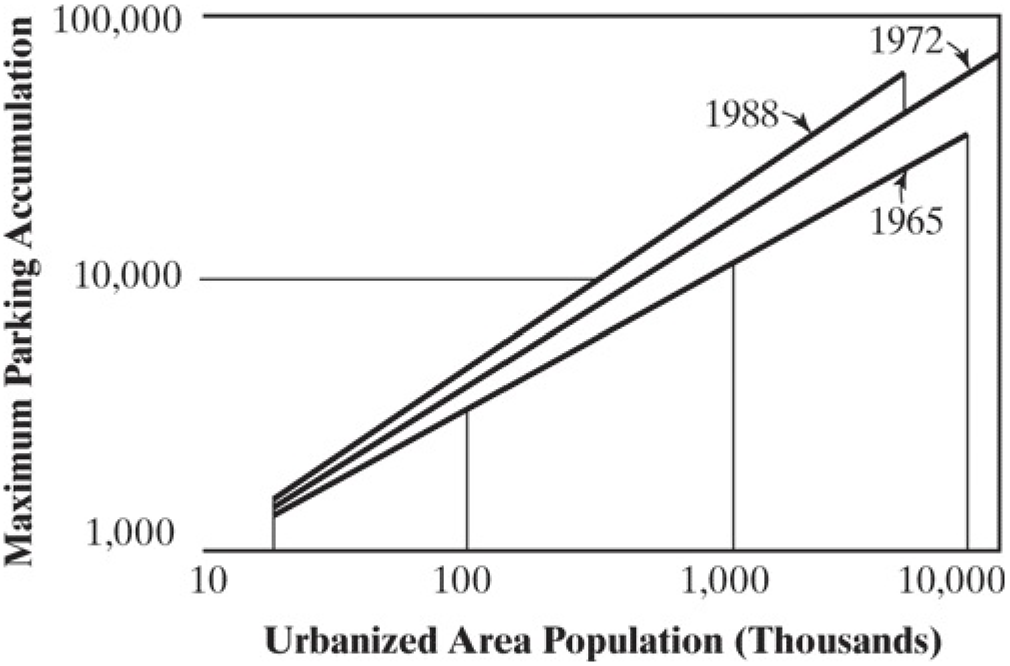
Inventory data can be displayed in tabular form, usually similar to that illustrated in Figure 13.3, or can be graphically displayed on coded maps. Maps provide a good overview, but cannot contain the detailed information provided in tabular summaries. Therefore, maps and other graphic displays are virtually always accompanied by tables.

13.2.3 Accumulation and Duration

Parking accumulation is defined as the total number of vehicles parked at any given time. Many parking studies seek to establish the distribution of parking accumulation over time to determine the peak accumulation and when it occurs. Of course, observed parking accumulations are constrained by parking supply; thus, parking demand that is constrained by lack of supply must be estimated using other means.

Nationwide studies have shown that parking accumulation in most cities has increased over time. Total accumulation in an urban area, however, is strongly related to the urbanized area population, as illustrated in Figure 13.4.

Figure 13.4: Parking Accumulation in Urbanized Areas by Population



(Source: Used with permission of Eno Foundation for Transportation, Weant, R., and Levinson, H., *Parking*, Westport, CT, 1990, Figure 6.8, pg 100.)

Figure 13.4: Full Alternative Text

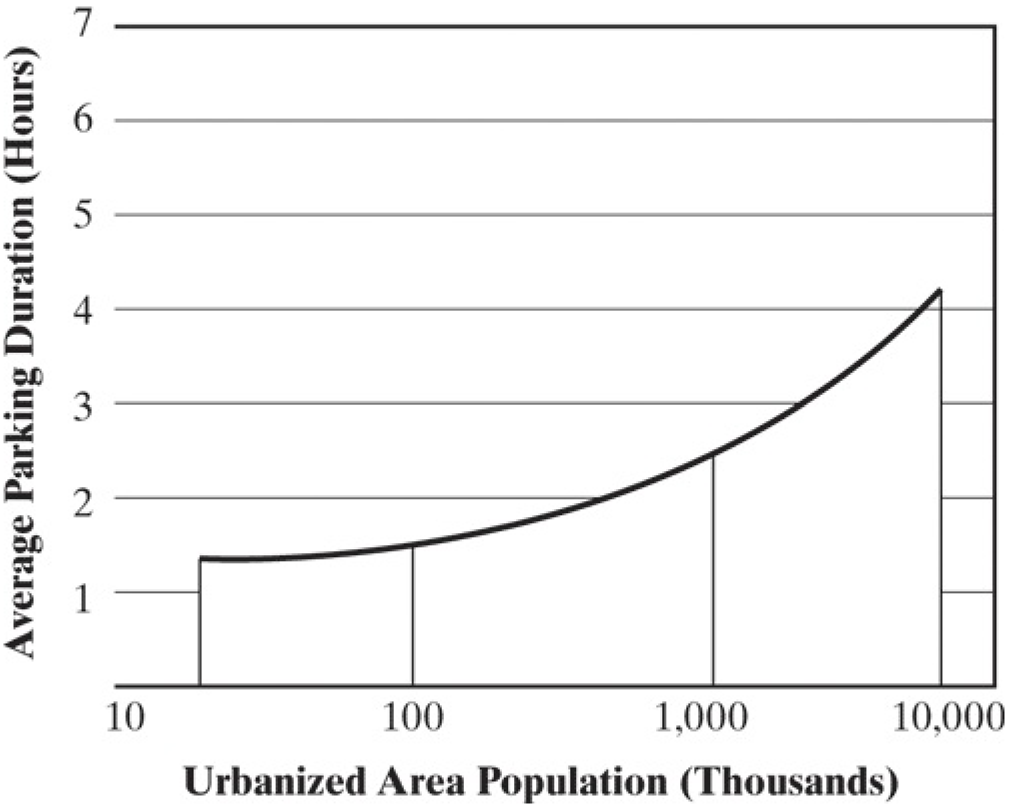
The data depicted in Figure 13.4 is quite old. Nevertheless, it represents the most comprehensive compilation of national data on the subject, and the trends it reveals are likely quite valid, although the specific accumulation values may be too old to be trusted. The following key features are revealed:

- As the population of an urbanized area increases, maximum parking accumulation also increases—an obvious expectation.
- Data for three years is depicted: maximum accumulations increase with decreasing age of the data. Over time, the maximum observed parking accumulations are increasing.

Parking duration is the length of time that individual vehicles remain parked. This characteristic is, therefore, a distribution of individual values, and both the distribution and the average value are of great interest.

Like parking accumulation, average parking durations are related to the size of the urban area, with average duration increasing with urban area population, as shown in Figure 13.5. Average duration also varies considerably with trip purpose, as indicated in Table 13.9, which is a compilation of information from several studies [10].

Figure 13.5: Parking Duration vs. Urbanized Area Population



(Source: Used with permission of Eno Foundation for Transportation, Weant, R., and Levinson, H., *Parking*, Westport, CT, 1990, Figure 6-4, pg 97.)

Figure 13.5: Full Alternative Text

Some of the data compiled in Table 13.10 is rather old, some collected in 1979. The times reported for work- related trips seem unreasonably short. A more recent study in the Netherlands [8] shows average work-related parking duration to be 6.08 h. Shopping-related and business-related parking duration were also longer than the values shown in Table 13.10: 1.83 h for shopping and 3.56 h for business.

Table 13.10: Parking Duration from Several Studies

Population	Shopping and Business	Work	Other	All Purposes
≤50,000	0.6 h	3.3 h	0.9 h	1.2 h
50,000–250,000	0.9 h	3.8 h	1.1 h	1.5 h
250,000–500,000	1.2 h	4.8 h	1.4 h	1.9 h
>500,000	1.5 h	5.2 h	1.6 h	2.6 h

(Source: Rastogi, R, *Validating Stated Parking Duration of Drivers in Kota City, India*, Paper No. 150, Indian Institute of Technology Roorkee, Uttarakhand, India, May 2014.)

Table 13.10: Full Alternative Text

Although generalized trends are interesting, it is obvious that durations vary widely from location to location. Thus, local studies of both parking duration and parking accumulation are important elements of an overall approach to the planning and operation of parking facilities.

The most commonly used technique for observing duration and accumulation characteristics of curb parking and surface parking lots is the recording of license plate numbers of parked vehicles. At regular intervals ranging from 10 to 30 minutes, an observer walks a particular route (usually up one block face and down the opposite block face), and records the license plate numbers of vehicles occupying each parking space. A typical field sheet is shown in Figure 13.6.

Figure 13.6: A License-Plate Parking Survey Sheet

LICENSE PLATE CHECK FIELD DATA SHEET

City________ Date 10 MAY 1978 Recorded by JONES Side of Street W

Street WRIGHT between 5th and 6th

Codes: 000 last three digits of license number. √ for repeat number from prior circuit _ for empty space

Space and Regulation	Time circuit begins											
	07	07^{30}	08	08^{30}	09	09^{30}	10	10^{30}	11	11^{30}	12^{00}	
5th												
X-WALK	–	–	–									
NPHC	–	–	–									
1HR M	–	713	√	√/TK								
" M	631	√	√	971								
" M	512	34L	√	019								
DRIVEWAY	–	–	–	–								
"	–	–	–	613								
1HR M	–	–	418	√								
" M	117	220	√	989								
" M	–	148	096	√								
FIRE HYD	–	–	–	–								
1HR M	042	–	216	√								
N PHC	–	–	–	774								
X-WALK	–	–	–	–								
6th												

(Source: Used with permission of Institute of Transportation Engineers, Box P. and Oppenlander, J., *Manual of Traffic Engineering Studies*, 4th Edition, Washington, D.C., 1976, Figure 10-6, pg 140.)

Figure 13.6: Full Alternative Text

Each defined parking space is listed on the field sheet prepared for the specific study, along with any time restrictions associated with it. A variety of special notations can be used to indicate a variety of circumstances, such as “T” for truck, “TK” for illegally parked and ticketed vehicle, and so on. One observer can be expected to observe up to

60 spaces every 15 minutes. Study areas, therefore, must be carefully mapped to allow planning of routes for complete data coverage.

Analysis of the data involves several summaries and computations that can be made using the field sheet information:

- Accumulation totals. Each column of each field sheet is summed to provide the total accumulation of parked vehicles within each time period on each observer's route.
- Duration distribution. By observing the license plate records of each space, vehicles can be classified as having been parked for one interval, two intervals, three intervals, etc. By examining each line of each field sheet, a duration distribution is created.
- Violations. The number of vehicles illegally parked, either because they occupy an illegal space or have exceeded the legal time restriction of a space, should be noted.

The average parking duration is computed as:

$D=\sum x(Nx \times X \times I)NT$ [13-3]

where:

D=average parking duration, h/veh,Nx=number of vehicles parked for x int

Another useful statistic is the parking turnover rate, *TR*. This rate indicates the number of parkers that, on average, use a parking stall over a period of 1 hour. It is computed as:

$TR=NTPS \times TS$ [13-4]

where:

TR=parking turnover rate, veh/stall/h,NT=total number of parked vehicles (

The average duration and turnover rate may be computed for each field sheet, for sectors of the study area, and/or for the study area as a whole. Table 13.11 shows a typical field sheet resulting from one observer's route. Table 13.12 shows how data from individual field sheets can be summarized to obtain areawide totals.

Table 13.11: Summary and Computations from a Typical Parking Survey Field Sheet

Pkg* Space	Time														
	8:00	8:30	9:00	9:30	10:00	10:30	11:00	11:30	12:00	12:30	1:00	1:30	2:00	2:30	3:00
1	-	-	861	√	√	-	136	-	140	√	-	-	201	√	√
2	470	√	380	-	-	412	307	-	900	√	√	√	√	-	070
3	-	211	√	√	√	400	√	√	-	-	666	-	855	999	-
4	175	√	√	500	√	222	-	-	616	√	√	√	√	√	-
5	333	-	-	380	√	√	420	√	707	-	-	-	-	-	-
hydrant	-	-	-	-	-	-	-	242TK	-	-	-	-	-	-	-
1-hr	-	-	484	√	909	-	811	√	√	158	√	√	685	√	-
1-hr	301	-	-	525	√	√	696	√	422	-	299	√	√	-	892
1-hr	-	675	895	√	√	703	√	819	-	401	√	√	288	-	412
1-hr	406	-	442	781	882	√	√	√	444	-	903	√	-	-	-
1-hr	-	-	115	√	618	√	818	√	√	906	√	-	-	893	√
2-hr	-	509	√	√	-	705	√	√	√	688	√	696	-	-	807
2-hr	-	-	214	√	√	√	209	-	248	√	797	√	√	√	√
2-hr	101	√	√	√	-	531	√	-	940	√	√	√	628	√	√
2-hr	-	392	√	√	√	251	√	772	-	835	√	√	√	-	-
Accum.	6	7	12	13	11	12	13	10	11	10	12	11	11	8	9

*All data for Block Face 61; timed spaces indicate parking meter limits; √ = same vehicle parked in space.

Table 13.11: Full Alternative Text

Table 13.12: Summary Data for an Entire Study Area Parking Survey

Block No.	Accumulation for Interval (1500 Total Stalls)														
	08:00	08:30	09:00	930	10:00	10:30	11:00	11:30	12:00	12:30	01:00	01:30	02:00	02:30	03:00
61	6	7	12	13	11	12	13	10	11	10	12	11	11	8	9
62	5	10	15	14	16	18	17	15	15	10	9	9	7	7	8
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
180	7	8	13	13	18	14	15	15	11	14	16	10	9	9	6
181	7	5	18	16	12	14	13	11	11	10	10	10	6	6	5
Total	806	900	1106	1285	1311	1300	1410	1309	1183	1002	920	935	970	726	694

(a) Summarizing Field Sheets for Accumulation Totals

13.2-12 Full Alternative Text

Block Face No.	Number of Intervals Parked					
	1	2	3	4	5	6
61	28	17	14	9	2	1
62	32	19	20	7	1	3
.	.	.	.	.	.	.
.	.	.	.	.	.	.
.	.	.	.	.	.	.
180	24	15	12	10	3	0
181	35	17	11	9	4	2
Total	**875**	**490**	**308**	**275**	**143**	**28**

(b) Summarizing Field Sheets for Duration Distribution

$\sum=2118$ total parkers observed

13.2-13 Full Alternative Text

Note that the survey includes only the study period. Thus, vehicles parked at 3:00 PM will have a duration that ends at that time, even though they may remain parked for an additional time period outside the study limits. For convenience, only the last three numbers of the license plates are recorded; in most states, the initial two or three letters/numbers represent a code indicating where the plate registration was issued. Thus, these letters/numbers are often repetitive on many plates. Sample Problem 13-6 illustrates how this data would be used to generate average duration and other statistics.

Sample Problem 13-6: Determining Key Parking Values from a Survey

The average duration for the study area, based on the summary of Table 13.12(b) is:

$D=$
$(875 \times 1 \times 0.5) + (490 \times 2 \times 0.5) + (308 \times 3 \times 0.5) + (275 \times 4 \times 0.5) + (143 \times 5 \times 0$

The turnover rate is:

$TR=21191500 \times 7=0.20$ veh/stall/h

The maximum observed accumulation occurs at 11:00 AM (from Table 13.12a), and is 1,410 vehicles, which represents use of $(1{,}410/1{,}500) \times 100=94\%$ of available spaces.

For off-street facilities, the study procedure is somewhat altered, with counts of the number of entering and departing vehicles recorded by 15-minute intervals. Accumulation estimates are based on a starting count of occupancy in the facility and the difference between entering and departing vehicles. A duration distribution for off-street facilities can also be obtained if the license-plate numbers of entering and departing vehicles are also recorded.

As noted earlier, accumulation and duration observations cannot reflect repressed demand due to inadequacies in the parking supply. Several findings, however, would serve to indicate that deficiencies exist:

- Large numbers of illegally parked vehicles
- Large numbers of vehicles parked unusually long distances from primary generators
- Maximum accumulations that occur for long periods of the day and/or where the maximum accumulation is virtually equal to the number of spaces legally available

Even these indications do not reflect trips either not made at all, or those diverted to other locations because of parking constraints. A cordon-count study may be used to estimate the total number of vehicles both parked and circulating within a study area, but trips not made are still not reflected in the results.

13.2.4 Other Types of Parking Studies

A number of other techniques can be used to gain information concerning parked vehicles and parkers. Origins of parked vehicles can be obtained by recording the license plate numbers of parked vehicles and petitioning the state motor vehicle agency for home addresses (which are assumed to be the origins). This technique, which requires special permission from state authorities, is frequently used at shopping centers, stadiums, and other large trip attractors.

Interviews of parkers are also useful and are most easily conducted at large trip attraction locations. Basic information on trip purpose, duration, distance walked, etc. can be obtained. In addition, however, attitudinal and background parker characteristic information can also be obtained to gain greater insight into how parking conditions affect users.

13.3 Design Aspects of Parking Facilities

Off-street parking facilities are provided as (1) surface lots or (2) parking garages. The latter may be above ground, below ground, or a combination of both.

13.3.1 Construction Costs

The construction costs of both surface lots and garages vary significantly depending upon location and specific site conditions. In general, surface lots are considerably cheaper than garages. Two of the most important factors involved that affect the cost of parking are:

- Design efficiency: How much area is used for each parking space. This can range from a low of 250 to 350 ft^2/space. It is obviously cheaper to provide less area than more. The specific value involves issues of predominant vehicle sizes, driver characteristics (older drivers may require larger spaces for convenience), and other factors. Efficiency is also affected by how much area is devoted to circulation, access, and egress.

- Type of construction: Surface lots are cheaper than above-ground structures which are cheaper than below-ground structures. Specific terrain and geographic characteristics also seriously affect construction costs.

Typical costs for construction of parking are shown in Table 13.13 [2].

Table 13.13: Typical Costs per Space for Parking Construction

Type of Facility	250 ft^2/Space	300 ft^2/Space	350 ft^2/Space
Surface lot	$1,250–$2,500	$1,500–$3,000	$1,750–$3,500
Above-ground garage	$5,000–$10,000	$6,000–$12,000	$7,000–$14,000
Below-ground garage	$12,500–$25,000	$15,000–$30,000	$17,500–$35,000

(Source: Used with permission from Chrest Springer Science and Business Media A., et al, *Parking Structures: Planning, Design, Construction, Maintenance, and Repair*, 3rd Edition, Springer Science and Business, New York NY, 2001, Table 2-5, pg 23)

Table 13.13: Full Alternative Text

Todd Litman [3] also cites parking costs, shown in Table 13.14, but only includes surface and above-ground structures. The figures are generally less than those of Table 13-13, but they assume rectangular sites, good soil conditions, and no special amenities in the parking facility.

Table 13.14: Typical Cost per Space for Parking Construction: Rectangular Sites

Lot/Garage Size	Small Site (30,000 ft^2)	Medium Site (60,000 ft^2)	Large Site (90,000 ft^2)
Area per Parking Space	350 ft^2	325 ft^2	315 ft^2
Surface parking	$1,838	$1,700	$1,654
Ground + 1 level	$7,258	$6,143	$5,705
Ground + 2 levels	$8,085	$6,767	$6,284
Ground + 3 levels	$8,407	$6,996	$6,491
Ground + 4 levels	$8,747	$7,269	$6,747
Ground + 5 levels	$8,973	$7,451	$6,918
Ground + 6 levels	$9,135	$7,581	$7,040
Ground + 7 levels	$9,256	$7,678	$7,132
Ground + 8 levels	$9,351	$7,754	$7,203

(Source: Used with permission from Victoria Transport Policy Institute Litman, T., *Parking Management Best Practices*, American Planning Association, Chicago IL, 2006, Table 4-1, pg 55)

Table 13.14: Full Alternative Text

From Tables 13.13 and 13.14, both of which are expressed in Year 2000 dollars, it is clear that construction costs for off-street parking vary with a wide range of variables, including:

- Area per space
- Type of parking (surface lot, above-ground structure, below-ground structure)
- Size of facility (overall size, number of levels)

In addition, the specifics of the site, including its shape, topography, and subsurface conditions are all of great importance, as are the structural design details, including materials used. The costs shown herein are merely illustrative, and, like all engineering projects, specific cost estimates must be prepared for each site and facility.

13.3.2 Basic Parking Dimensions

All parking dimension criteria are based upon the standard vehicle used as a template. For most parking purposes, the current template is a standard full-size car. Spaces for trucks or larger vehicles are generally incorporated as part of loading facilities that may (or may not be) located within the confines of the parking facility.

In the 1980s, the number of small or compact cars on the road spurred a movement to segregate "small car parking" from "large car parking" within parking facilities. Although this often led to the ability to pack more parking spaces into a given facility, it also made for inefficiencies when the mix of parkers was out of synch with the assumed distribution. However, small car sales have declined precipitously from their high points. In 1987, 52.1% of all car sales in the United States were small cars. By 1998, this had declined to 33.9% [11]. Although recent trends indicate slight recovery for small car sales, in 2016, the percent of small cars barely topped 20%. As a result of this, inclusion of differing size parking spaces is now virtually abandoned as a design feature.

Although the exact dimensions of a parking "design vehicle" vary from place to place, the width is generally 6 ft, 7 in with a length of approximately 17.0 ft. This is fairly large, and accommodates even large SUVs, such as the Ford Expedition.

Parking Stall Width

Parking stalls must be wide enough to encompass the vehicle and allow for door-opening clearance. The minimum door-opening clearance is 20 inches, but this should be increased to 24 to 27 inches where turnover rates are high. Only one door-opening clearance is provided per stall, as the parked vehicle and its adjacent neighbor can utilize *the same clearance space*.

Table 13.15 gives recommendations for parking stall width based upon turnover activity, which is strongly related to parking purpose [11].

Table 13.15: Recommended Parking Stall Width versus Parking Turnover

Typical Parking Characteristics	Parking Space Width
Low turnover (employees, students, etc.)	8 ft 3 in to 8 ft 6 in
Low-to-moderate turnover (offices, regional shopping center, long-term airport parking, etc.)	8 ft 6 in to 8 ft 9 in
Moderate-to-high turnover (community retail, medical, etc.)	8 ft 9 in to 9 ft

(Source: Used with permission from National Parking association *The Dimensions of Parking*, 5th Edition, National parking association and Urban Land Institute, 2010, Figure 7-2, pg 61.)

Table 13.15: Full Alternative Text

Reference [2] suggests a level of service approach to stall width decisions. Level of service A would have a width of 9 ft, LOS B, 8 ft 9 in, LOS C, 8 ft 6 in, and LOS D, 8 ft 3 in. Level of service D would be restricted to congested urban centers, like New York City, where drivers are happy “just to find a parking space.” Otherwise, the choice would be based primarily on turnover, which is quite similar to the recommendations of Table 13.14.

It should be noted that parking stall width is measured perpendicular to the stall boundary markings.

Parking Stall Length, Width, and Projections

Parking stall length is measured parallel to the parking angle. Stall length is based upon the length of the design vehicle plus a buffer for bumper extensions. In modern parking design, a uniform length of 18 ft is generally used.

The depth of a parking stall is the 90° projection of the design vehicle length and 6-inch bumper clearance. For a 90° parking stall, the length and depth of the stall are equivalent. For other-angle parking, the depth of the stall is smaller than the length. Parking space depth is often referred to as the vehicle projection (VP).

Table 13.16 shows the vehicle and width projections of an 18-ft long parking stall of various widths versus the angle of parking. Figure 13.7 illustrates various dimensions in Table 13.16.

Table 13.16: Length and Width Projections for Common Parking Stall Dimensions

θ (degrees)	Sin θ	SL (ft)	VP (ft)	WP (ft) for SW = 8.5 ft	WP (ft) for SW = 8.25 ft	WP (ft) for SW = 8.75 ft	WP (ft) for SW = 9 ft
45	0.7071	18	12.7	12.0	11.7	12.4	12.7
50	0.7660	18	13.8	11.1	10.8	11.4	11.7
55	0.8192	18	14.7	10.4	10.1	10.7	11.0
60	0.8660	18	15.6	9.8	9.5	10.1	10.4
65	0.9063	18	16.3	9.4	9.1	9.7	9.9
70	0.9397	18	16.9	9.0	8.8	9.3	9.6
75	0.9659	18	17.4	8.8	8.5	9.1	9.3
90	1.0000	18	18.0	8.5	8.3	8.8	9.0

Note: 8.5 ft = 8ft 6 in; 8.25 ft = 8 ft 3 in; 8.75 ft = 8 ft 9 in.

Table 13.16: Full Alternative Text

Figure 13.7: Parking Stall Dimensions Illustrated

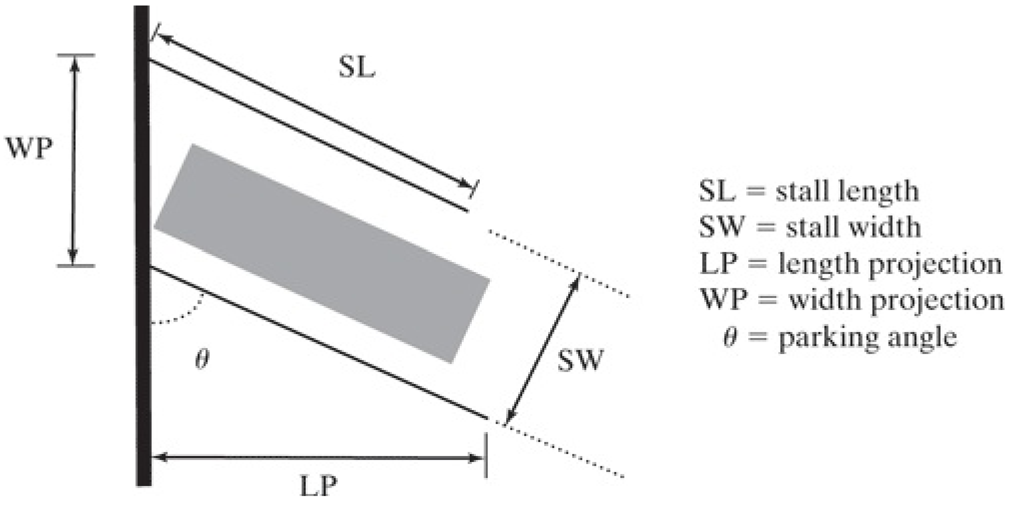


Figure 13.7: Full Alternative Text

The width and length projections of any parking stall design can be computed using trigonometry:

$LP = SL \times Sin(\theta)$ [13-5]
$WP = SWSin(\theta)$ [13-6]

Aisle Width

Aisles in parking lots must be sufficiently wide to allow drivers to safely and conveniently enter and leave parking stalls in a minimum number of maneuvers, usually one on entry and two on departure. As stalls become narrower, the aisles need to be a bit wider to achieve this. Aisles also carry circulating traffic and accommodate pedestrians walking to or from their vehicles. Aisle width depends upon the angle of parking and upon whether the aisle serves one-way or two-way traffic.

When angle parking is used, aisles are almost always one-way. Two-way aisles could conceivably be used with stalls angled in opposite directions on either side of the aisle, but this often creates awkward entry and exit maneuvers, and is generally avoided. Where 90° parking stalls are used, aisles may be one-way or two-way, although two-way aisles are most common for ease of circulation.

A summary of commonly used aisle widths in the United States is shown in Table 13.17. The table assumes one-way flow for angle parking and two-way flow for 90° parking [2, 6].

Table 13.17: Typical Parking Aisle Widths in the United States

Parking Angle (θ)	Range of Aisle Widths (ft)
45°	12.0–15.0
50°	12.5–16.0
55°	13.5–17.0
60°	14.0–18.0
65°	15.5–19.5
70°	17.0–20.5
75°	18.5–22.0
90°	24.0–26.0

Table 13.17: Full Alternative Text

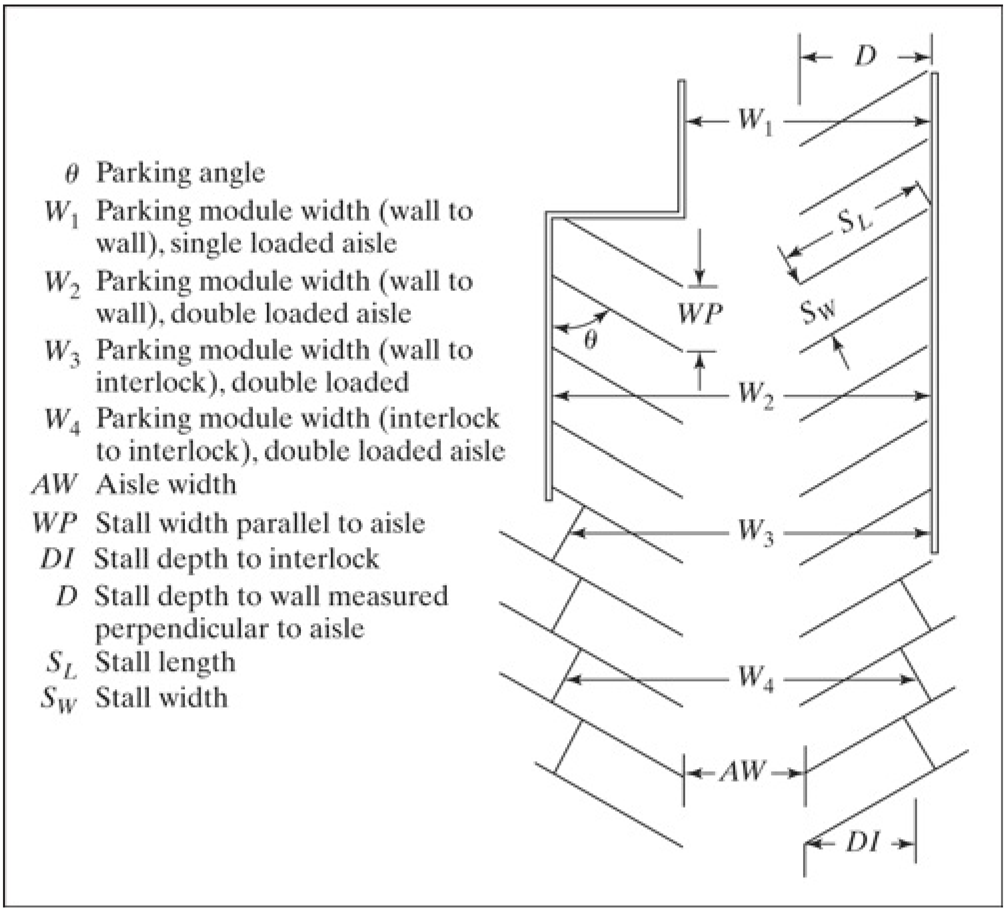
At shallower angles, the entry and exit maneuvers are relatively easy, and narrower lanes may be used. Thus, as the angle of parking increases, the typical aisle width also increases.

13.3.3 Parking Modules

A "parking module" refers to the basic layout of one aisle with a set of parking stalls on both sides of the aisle. There are many potential ways to lay out a parking module. For 90° stalls, two-way aisles are virtually always used, as vehicles may enter parking stalls conveniently from either approach direction. Where angle parking is used, vehicles may enter a stall in only one direction of travel and must depart in the same direction. In most cases, angle parking is arranged using one-way aisles, and stalls on both sides of the aisle are arranged to permit entries and exits from and to the same direction of travel. Angle stalls can also be arranged such that stalls on one side of the aisle are approached from the opposite direction as those on the other side of the aisle. In such cases, two-way aisles must be provided. Figure 13.8 defines the basic dimensions of a parking module.

Figure 13.8: Dimensional

Elements of Parking Modules



(Source: Used with permission of Institute of Transportation Engineers, *Traffic Engineering Handbook*, Prentice Hall 1965, Figure 7-1, pg 208.)

Figure 13.8: Full Alternative Text

Note that Figure 13.8 shows four different ways of laying out a module. One module width applies if both sets of stalls butt up against walls or other horizontal physical barriers. Another applies if both sets of stalls are “interlocked” (i.e., stalls interlock with those of the next adjacent parking module). A third applies if one set of stalls is against a wall, while the other is interlocked. Yet another module reflects only a single set of stalls against a wall.

Module width (W_2), where both sets of stalls are against a wall or other solid barrier, are generally the sum of two length projections plus the aisle width. Where only one row of parking exists on an aisle, the module width (W_1) is the sum of one length projection plus the aisle width. Where one or both aisles are interlocked, the module width (W_3 or W_5) may be reduced based upon the parking stall angle (θ), as shown in Table 13.18.

Table 13.18: Reduction in Module Widths for Interlocking Parking Stalls

Parking Stall Angle (θ)	Reduction for One Interlocking Row of Stalls	Reduction for Two Interlocking Rows of Stalls
45°	1.9 ft	3.8 ft
50°	1.7 ft	3.4 ft
55°	1.6 ft	3.2 ft
60°	1.4 ft	2.8 ft
65°	1.2 ft	2.4 ft
70°	1.0 ft	2.0 ft
75°	0.7 ft	1.4 ft
90°	0.0 ft	0.0 ft

Note: All values rounded to the nearest 0.1 ft.

Table 13.18: Full Alternative Text

The width of a parking module is determined as:

$W=n\times VP+AW-ri$ [13-7]

where:

W=width of parking module, ft,n=number of parking rows in module (1 or

Table 13.19 shows module widths for the configurations of Figure 13.8 (W_1, W_2, W_3, W_4). In each case, a mid-range aisle width (*AW*) from Table 13.16 is used.

Table 13.19: Parking Module Widths for Various Module Configurations

Parking Angle θ	Aisle Width Used AW (ft)	Vehicle Projection Used VP (ft)	Parking Module Width (ft)			
			W_1	W_2	W_3	W_4
45°	14	12.7	26.7	39.4	37.5	35.6
50°	15	13.8	28.8	42.6	40.9	39.2
55°	16	14.7	30.7	45.4	43.8	42.2
60°	17	15.6	32.6	48.2	46.8	45.4
65°	18	16.3	34.3	50.6	49.4	48.2
70°	19	16.9	35.9	52.8	51.8	50.8
75°	20	17.4	37.4	54.8	54.1	53.4
90°	24	18.0	42.0	60.0	60.0	60.0

Table 13.19: Full Alternative Text

The module width is an important dimension, as it allows a designer to assess how many modules (and therefore how many rows of parking stalls) can be fit into any given footprint. In doing so, it should be remembered that end aisles running perpendicular to the module must be provided to allow vehicles to move from one module to the next in their search for a parking space. In general, an end aisle of 29 ft is suggested for two-way circulation, and 17 ft for one-way circulation, not including any clearances required for structural elements [2].

13.3.4 Access and Egress

Off-street parking involves a number of access and egress alternatives. Obviously, the peak demand rates of entering and exiting vehicles must be estimated to do anything. The other major factors, however, generally involve how parking is to be controlled, and how fees (where applicable) are to be collected.

The base case would be one in which no parking fees are collected. In such a case, there is no need for manual or automated tracking of entries and exits, except for the purposes of determining when the lot is full—in which case prospective parkers need to be diverted. The number of entry/exit lanes needed would be entirely dependent on the demand and the physical capacity of the entry/exit lanes—which are generally limited by geometry.

Where fees are to be collected, there are a wide range of technologies and options now available that can be applied. Three major categories of systems can be deployed:

- Metered spaces; no monitoring at entry and exit points.

- Payment on departure: Generally, a ticket is dispensed upon entry (usually with a gate preventing entry until a ticket is given), and payment is made upon departure. Payment is often related to amount of time parked. In some cases, an automated toll tag (such as EZ Pass) may be used for both entry and exit, which greatly expands the capacity of both entry and exit lanes.

- Payment on foot: There are a number of systems that provide dispersed ticket dispensers throughout the facility, which are accessed after the vehicle is parked. Such ticket dispensers may require up-front payment (often based upon time parked); parkers must display the ticket through their windshields while parked. Some systems allow payment on departure, which also takes place at the dispersed dispensers on foot. The ticket is inserted, a fee is posted, and the parker then pays by inserting cash or a credit card. Such systems have many variable features that can be designed for specific situations, and generally avoid lengthy transactions on departure. In most cases, a paid ticket must be inserted at the exit point to prove payment.

Metered spaces are often used to favor high-priority parkers. In a shopping center, for example, the most convenient parking spaces are reserved for customers, with meters only allowing short- to moderate-term parkers.

Employees would be forced to use less convenient spaces to park for their work shifts, but would generally get a reduced rate for their trouble.

Table 13.20 shows typical design service rates (veh/h) that can be accommodated by various entry and exit systems [2, 6].

Table 13.20: Typical Design Service Rates of Entry/Exit Systems

Type of Entry/Exit System	Design Service Rate (veh/h)
Prepaid Frequent Parker (Entry or Exit)	
Insertion card	435
Proximity cart	600
Automatic vehicle ID	800
Pay per Use Vehicle Entry	
Push-button ticket dispenser	400
Auto spit ticket dispenser	450
Pay on entry flat fee—with gate	200
Pay on entry flat fee—no gate	300
Pay per Use Vehicle Exit	
Fixed fee to cashier—with gate	200
Fixed fee to cashier—no gate	250
Variable fee to cashier	150
Credit card with on-line check and signature required	115
Credit card with on-line check and no signature required	135
Validated for free parking	180
No Payment or Ticket Dispensed; Automated Entry Exit with EZ Pass or Similar Device	800

Notes: Virtually all pay-on-foot options involve no monitoring of entries or exits; use “no payment” category.

Values in table represent easy entry/exit geometries. If a sharp turn exists within 100 ft of entry/exit lane, these values may be reduced by 40%– 45%.

Table 13.20: Full Alternative Text

It should be noted that a single entry or exit lane may handle a variety of different types of transactions. Some estimate of the split would have to be made, and the design service rates proportionately adjusted.

The number of entry or exit lanes required for a given parking facility can be approximately estimated as:

$N=S\times RPHF\times u$ [13-8]

where:

N=number of entry or exit lanes needed,S=total number of parking spaces i

Table 13.21 shows general ranges for proportion of parking spaces with vehicles entering or leaving in the peak hour.

Table 13.21: Typical Peak Hour Volumes as a Proportion of Total Parking Spaces

Type of Activity	Peak Entries	Peak Departures
Residential	0.30–0.50	0.30–0.50
Hotel/Motel	0.30–0.60	0.30–0.50
Hospital visitors	0.40–0.60	0.50–0.75
Hospital employees	0.60–0.75	0.60–0.75
Central business district	0.40–0.60	0.40–0.60
Retail/Commercial	0.30–0.60	0.40–0.65
Airport—Short-term parking	0.70–0.90	0.70–0.90
Airport—Mid-term parking	0.90–1.00	0.90–1.00
Airport—Long-term parking	0.05–0.10	0.05–0.10

(Source: Used with permission from Eno Foundation for Transportation Weant, R.A., and Levinson, H.S. *Parking*, ENO Foundation for Transportation, Westport CT, 1990, Table 9-1, pg 185.)

Table 13.21: Full Alternative Text

As in the case of Table 13.19, a given parking lot may serve various types of activities, and values would have to be proportionally adjusted. The data of Table 13.20 shows a wide range, and is included only for illustrative purposes. Local data on these factors is essential if accurate predictions of peak entry and exit usage are to be made.

Sample Problem 13-7: Determining Entry/Exit Lanes

Just to illustrate the use of Equation 13.9, a short-term parking lot at an airport has 900 parking spaces. If the peak hour factor (*PHF*) is 0.85, how many entry and exit lanes would be needed? The entry system includes ½ automated tracking (EZ-Pass) and ½ push-button tickets dispensed. The exit system includes ½ automated tracking, ¼ variable cash collections, and ¼ credit card transactions with no signature required.

Table 13.21 indicates that a short-term airport parking facility will have peak arrivals and departures of 0.70 to 0.90 times the number of parking spaces. For this example, we will choose the midpoint, or 0.80.

From Table 13-20, automated entry and exit lanes can service 800 veh/h. Push-button entries can be processed at a rate of 400 veh/h. For a 50 to 50 mix, an entry rate of 600 veh/h would be used. On exit, automated exits can be serviced at 800 veh/h, variable cash transactions at 150 veh/h and credit cards (with no signature) can be processed at 135 veh/h. With a split of 0.50 – 0.25 – 0.25, the exit design service rate would be $0.50\times800+0.25\times150+0.25\times135=470.25$ veh/h, *say* 470 veh/h. Then:

$$N=\frac{S\times R}{PHF\times u} \quad N_{entry}=\frac{900\times0.80}{0.85\times600}=1.4\ \text{lanes} \quad N_{exit}=\frac{900\times0.80}{0.85\times}$$

Given that a partial lane cannot be built, two entry and two exit lanes (with booths and detectors) would be provided.

The design of entry and exit lanes to and from parking lots and garages is quite complex, as it involves many considerations not specifically treated herein. These include whether entries and exits are desired at more than one location, general traffic conditions on the access streets and surrounding intersections, locations of major parking generators, and others.

13.3.5 Parking Garages

Parking garages are subject to the same stall and module requirements as surface parking lots and have the same requirements for reservoir areas and circulation. The structure of a parking garage, however, presents additional constraints, such as building dimensions and the location of structural columns and other features. Ideal module and stall dimensions must sometimes be compromised to work around these structural features.

Parking garages, whether above or below ground, have the additional burden of providing vertical as well as horizontal circulation for vehicles. This involves a general design and layout that includes a ramp system, at least where self-parking is involved. Some smaller attendant-parking garages use elevators to move vertically, but this is a slow and often inefficient process.

Ramping systems fall into two general categories:

- Clearway systems. Ramps for interfloor circulation are completely separated from ramps providing entry and exit to and from the parking garage.
- Adjacent parking systems. Part or all of the ramp travel is performed on aisles that provide direct access to adjacent parking spaces.

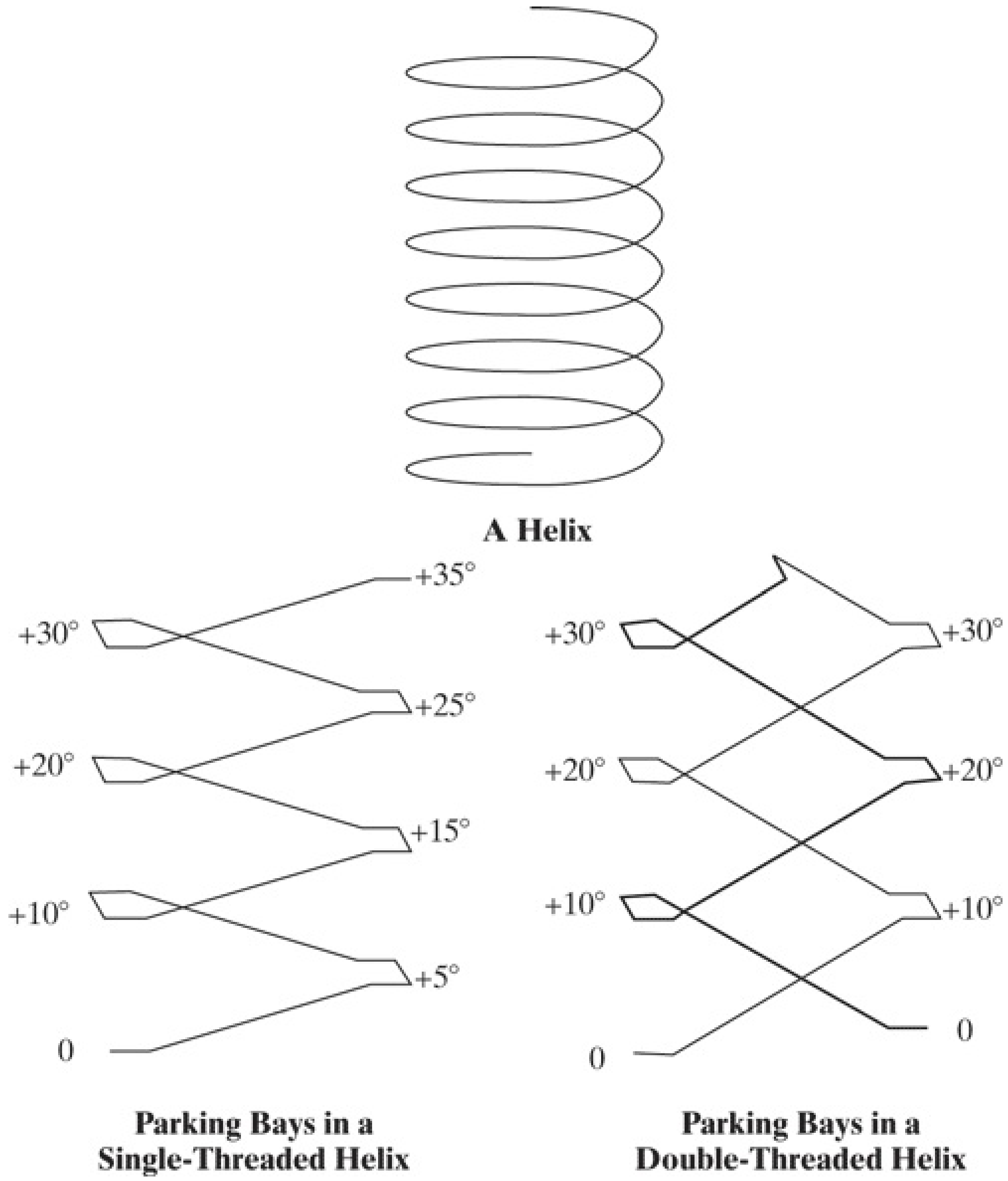
The former provides for easier and safer movement with minimum delays. Such systems, however, preempt a relatively large amount of potential parking space and are therefore usually used only in large facilities.

In some attendant-park garages and surface lots, mechanical stacking systems are used to increase the parking capacity of the facility. Mechanical systems are generally slow, however, and are most suited to longer-term parking durations, such as the full-day parking needs of working commuters, or overnight parking of residents.

There are, of course, many intricate details involved in the design and layout of parking garages and surface parking lots. This text covers only a few of the major considerations involved. The reader is advised to consult References [2, 4] and [11] directly for additional detail.

Figure 13.9 illustrates the fundamental geometry of vertical ramping systems in parking garages.

Figure 13.9: Basic Circulation Systems for Parking Garages



(Source: Used with permission from Springer Science and Business Media Chrest, A., et al, *Parking Structures: Planning, Design, Construction, Maintenance, and Repair*, Springer Science and Business, New York NY, 2001, Figure 3-2, pg 43)

Figure 13.9: Full Alternative Text

Parking garage circulation systems follow the general geometry of a helix, that is, a continuous system of ramps that leads upward or downward connecting the various levels of the garage to each other, and to entry and exit points.

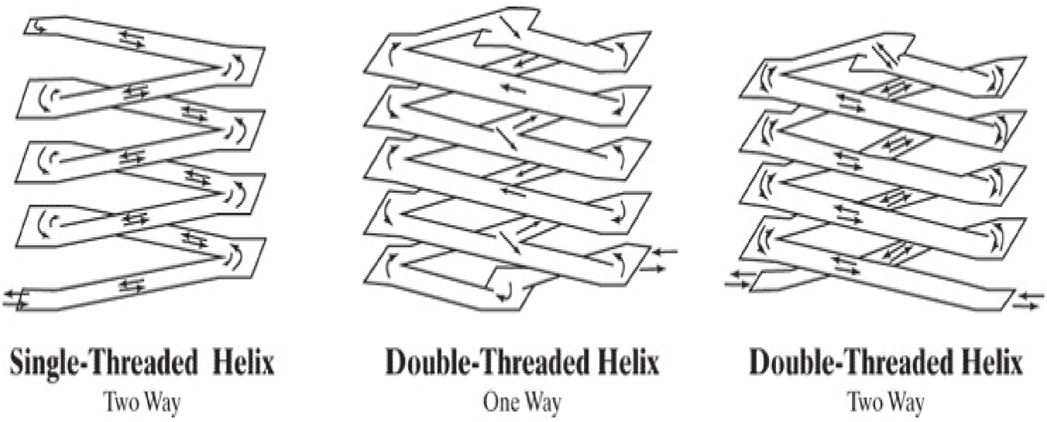
As noted previously, the ramps may be entirely separated from parking bays, or may be integrated, with parking spaces directly accessed from the ramp (or a combination of both). When ramps are external, they may actually follow the form of the curved helix. When parking is provided directly on ramps, the helix is formed by a succession of straight ramps, as shown in Figure 13.9.

In a single-threaded helix system, each helix (there may be more than one) provides a connection to every level of the parking facility. In a double-threaded helix, each ramp provides a connection to alternate parking levels. A minimum of two ramps are needed to provide access to all parking levels. In a single-threaded helix, the helix rises one floor in each revolution. In a double-threaded helix, the helix rises two floors in each revolution. There are, in rare cases, triple-threaded helix ramp systems, which require three separate ramps to access all levels, and the ramp rises three levels with each revolution. These cases, however, often require very long ramps and/or very steep grades, which make them difficult to implement except in the case of very large parking facilities.

Where an external curved helix is used, separate helixes are most often provided for traffic going up and coming down. It is possible to have a two-way circular helix, but the geometrics are inefficient, and most drivers are uncomfortable using them. When straight ramps are used, two-way flow (with appropriate ramp width) can be accommodated. Where one-way circulation is in place, separate helixes would be needed for up and down directions.

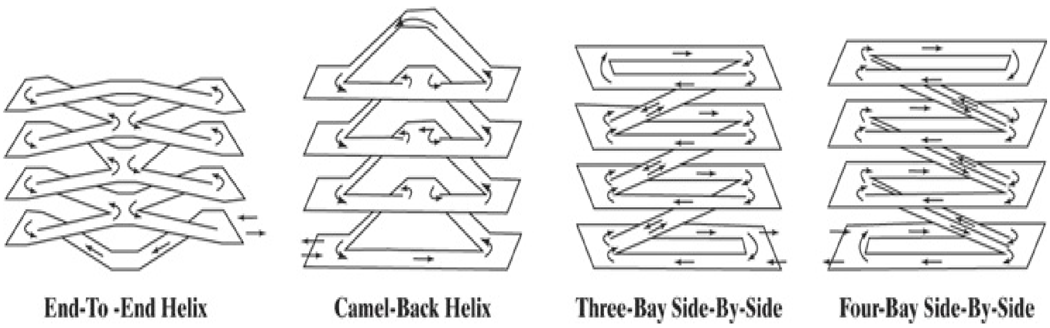
Helix ramp systems can be arranged in a wide variety of specific configurations, some of which are illustrated in Figure 13.10. The figure shows only some of the configurations that could be designed using the basic elements of a helix ramp system.

Figure 13.10: Illustrative Configurations for Garage Helix Ramp Systems



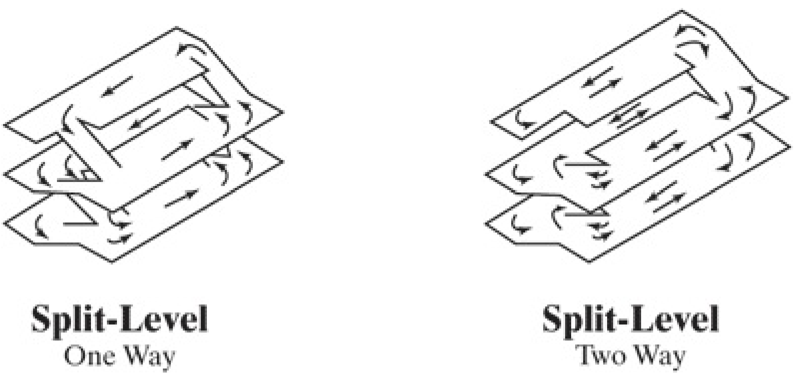
(a) Single- and Double-Threaded Helix Configurations

13.3-24 Full Alternative Text



(b) Single-Threaded Helix Options

13.3-24 Full Alternative Text



(c) Single-Threaded Split-Level Configurations

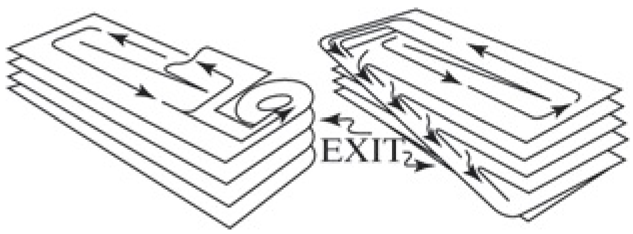
(Source: Used with permission from Springer Science and Business Media Chrest, A., et al, *Parking Structures: Planning, Design, Construction, Maintenance, and Repair*, 3rd Edition, Springer

Science and Business, New York NY, 2001, Figures 3-9, 3-10, and 3-11, pgs 61, 62)

13.3-24 Full Alternative Text

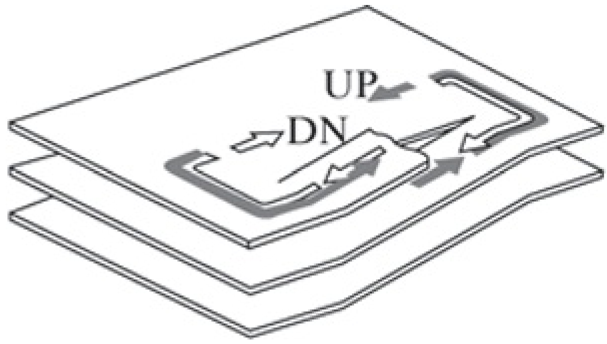
Figure 13.11 shows some additional options, illustrating combinations including sloped parking bays and level parking bays, as well as applications of clearway and adjacent ramps in garage circulation design.

Figure 13.11: Additional Illustrations of Parking Garage Ramp Systems



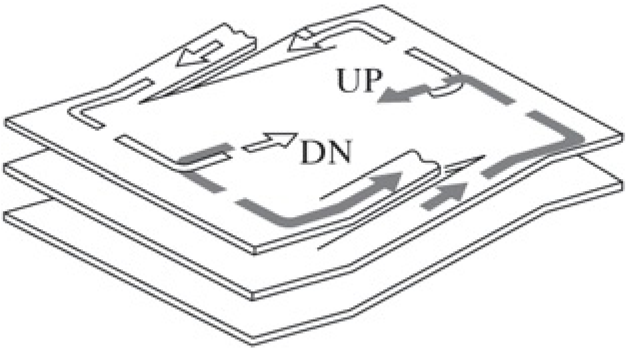
(a) Adjacent ramps for entering traffic; clearway ramps for exiting traffic.

13.3-24 Full Alternative Text



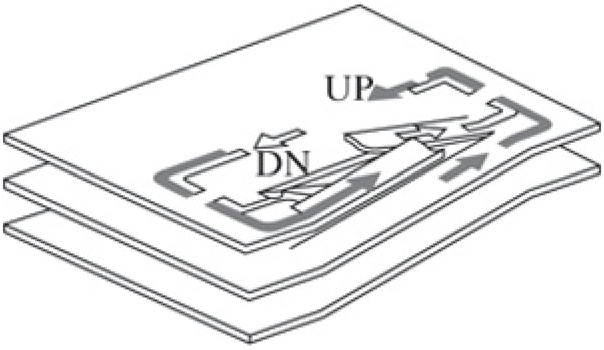
(b) Straight-ramp system with one ramp-well.

13.3-24 Full Alternative Text



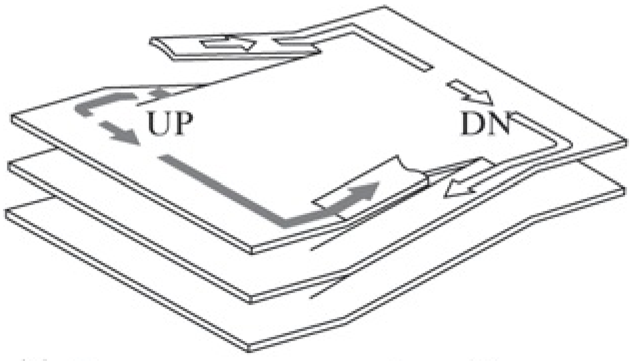
(c) Parallel straight ramp system with ramp-wells on opposing sides of the structure.

13.3-24 Full Alternative Text



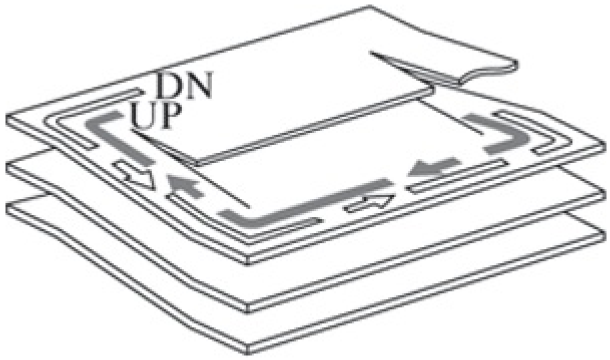
(d) Adjacent-parking type opposed straight-ramp system.

13.3-24 Full Alternative Text



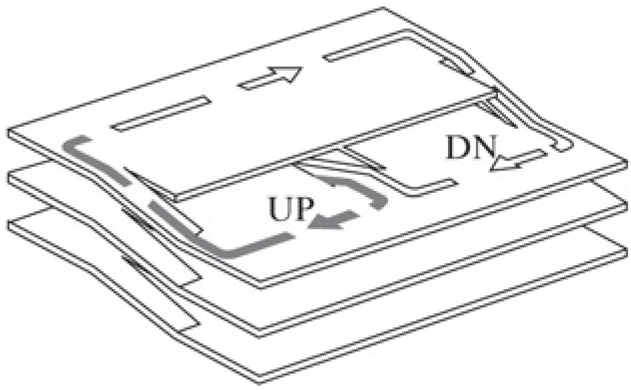
(e) Clearway type opposed straight-ramp system.

13.3-24 Full Alternative Text



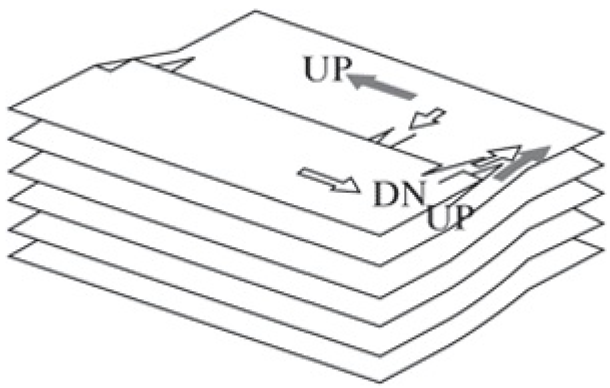
(f) Two-way staggered floor ramp system; ramps are placed at the ends of the garage to minimize turning conflicts.

13.3-24 Full Alternative Text



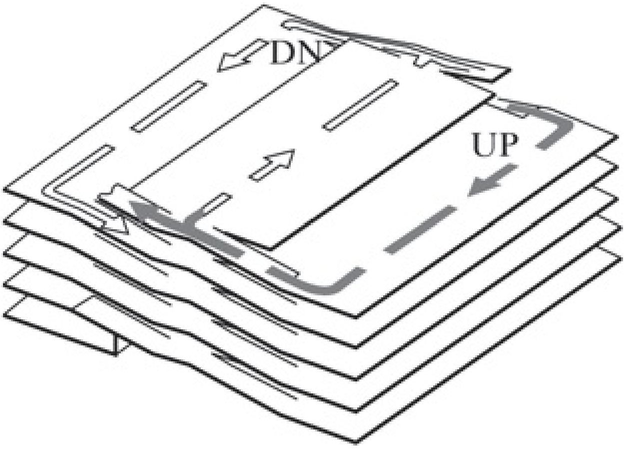
(g) Tandem staggered floor ramp system; ramps provide a clockwise circulation system.

13.3-24 Full Alternative Text



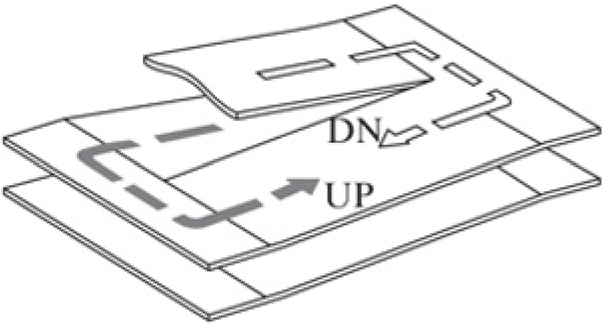
(h) This staggered floor ramp system provides parking on level floors and desirable one-way traffic flow.

13.3-24 Full Alternative Text



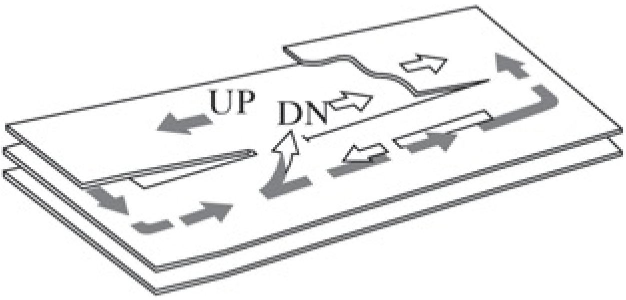
(i) Three-level staggered floor ramp system.

13.3-24 Full Alternative Text



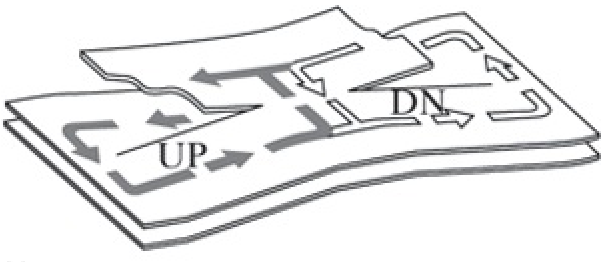
(j) Basic sloping floor concept.

13.3-24 Full Alternative Text



(k) Sloping-floor system with crossover ramp at midpoint.

13.3-24 Full Alternative Text



(l) Double sloping-floor system with midpoint crossover.

(Source: Used with permission from Eno Foundation for Transportation Weant, R.A., and Levinson, H.S., *Parking*, ENO Foundation for Transportation, Westport, CT, 1990, Figures 9.5–9.16, pgs 188–192)

13.3-24 Full Alternative Text

Although the fundamental concepts of garage circulation and ramp systems are relatively straightforward, the design of any particular facility can apply many elements and specific approaches to accomplish an effective parking facility. References [2], [5], [6], and [11] all contain additional detailed material on parking lot and garage design that should be consulted for specific applications.

13.4 Parking Programs, Policy, and Management

Every urban governmental unit must have a plan to deal effectively with parking needs and associated problems. Parking is often a controversial issue, as it is of vital concern to the business community in general and to particular businesses that are especially sensitive to parking. Further, parking has enormous financial impacts as well. In addition to the impact of parking on accessibility and the financial health of the community at large, parking facilities are expensive to build and to operate. On the flip side, revenues from parking fees are also enormous.

The public interest in parking falls within the government's general responsibility to protect the health, safety, and welfare of its citizens. Thus, the government has a responsibility to [6]:

- Establish parking program goals and objectives
- Develop policies and plans
- Establish program standards and performance criteria
- Establish zoning requirements for parking
- Regulate commercial parking
- Provide parking for specific public uses
- Manage and regulate on-street parking and loading
- Enforce laws, regulations, and codes concerning parking, and adjudicate offenses

There are a number of organizational approaches to effectively implement the public role. Parking can be placed under the authority of an existing department of the government. In small communities, where there is no professional traffic engineer or traffic department, a department of public works might be tasked with parking. In some cases, police departments have been given this responsibility (as an adjunct to their enforcement responsibilities), but this is not considered an optimal solution given that it will be subservient to the primary role of police departments. Where traffic departments exist, responsibility for parking can logically be placed there. In larger municipalities, separate departments can be established for parking. Parking boards may be created with appointed and/or elected members supervising the process. Because of the revenues and costs involved in parking, separate public parking authorities may also be established.

Parking facilities may be operated directly by governmental units or can be franchised to private operators. This is often a critical part of the process and may have a substantial impact on the net revenues from parking that find their way into the public coffers.

Parking policy varies widely depending upon local circumstances. In some major cities, parking supply is deliberately limited, and costs are deliberately kept high as a discouragement to driving. Such a policy works only where there is significant public transportation supply to maintain access to the city's businesses. Where parking is a major part of access, the planning, development, and operation of off-street parking facilities becomes a major issue. Private franchisees are often chosen to build, operate, and manage parking facilities. Although this generally provides a measure of expertise and relieves the government of the immediate need to finance and operate such facilities directly, the city must negotiate and assign a significant portion of parking revenues to the franchisee. Of course, parking lots and garages can be fully private, although such facilities are generally regulated.

Revenues are also earned from parking meter proceeds and from parking violations. Metering programs are implemented for two primary reasons: to regulate turnover rates and to earn revenue. The former is accomplished through time limits. These limits are established in conjunction with localized needs. Meters at a commuter rail station would, for example, have long-term time limits, as most people would be parking for a full working day. Parking spots near local businesses such as candy stores, barber shops, fast-food restaurants, florists, and similar uses would have relatively short-term time limits to encourage turnover and multiple users. Fees are set based on revenue needs and are influenced by general policy on encouragement or discouragement of parking.

No matter how the effort is organized and managed, parking programs must deal with the following elements:

1. Planning and policy. Overall objectives must be established and plans drafted to achieve them; general policy on parking must be set as part of the planning effort.

2. Curb management. Curb space must be allocated to curb parking, transit stops, taxi stands, loading areas, and other relevant uses; amounts and locations to be allocated must be set and the appropriate regulations implemented and signed.

3. Construction, maintenance, and operation of off-street parking facilities. Whether through private or governmental means, the construction of needed parking facilities must be encouraged and regulated; the financing of such facilities must be carefully planned so as to guarantee feasible operation while providing a revenue stream for the local government.

4. Enforcement. Parking and other curb-use regulations must be strictly enforced if they are to be effective; this task may be assigned to local police, or a separate parking violations bureau may be established; adjudication may also be accomplished through a separate traffic court system or through the regular local court system of the community.

To be most effective, parking policies should be integrated into an overall accessibility plan for central areas. Provision and/or improvement of public transportation services may mitigate some portion of parking demands while maintaining the fiscal viability of the city centers.

Parking is an essential part of accessibility to business, medical, retail, entertainment, and other critical functions. For many parts of the country, the automobile is the dominant form of access, thus the provision and management of parking becomes a critical function to the economic vitality of towns, cities, and their environs.

Todd Litman [3, 4] notes that paradigms for parking management have changed over time with the advance of technologies and approaches that can be associated with parking issues. Table 13.22 compares more traditional paradigms to more modern approaches.

Table 13.22: Parking Paradigm Shifts

Old Paradigm	New Paradigm
"Parking problem" means inadequate parking supply.	"Parking problem" can mean inadequate supply, inefficient management, inadequate user information, and other types of problems associated with parking facilities and activities.
More parking is better.	Too much parking is as harmful as too little.
Parking should generally be free. Whenever possible, parking facilities should be funded indirectly through building rents or leases.	As much as possible, users should pay directly for parking facilities.
Parking should be available on a first-come basis.	Parking should be managed to favor high-priority uses and encourage efficiency.
Parking requirements should be applied consistently without exception or variation.	Parking requirements should reflect each situation, and should be applied flexibly.
Traditional approaches should be favored. New approaches should be discouraged since they are unproved and not widely accepted.	Innovations should be encouraged, since even unsuccessful experiments often provide useful information.
Parking management should only be applied as a last resort where it would be too costly to increase supply.	Parking management programs should be widely applied to increase efficiency and prevent problems.
Transportation consists of driving. Dispersion of destinations (urban sprawl) is acceptable, or even desirable.	Driving is just one of many transport modes. Dispersed automobile-dependent land use patterns may be undesirable.

(Source: Used with permission from Victoria Transport Policy Institute Litman, T., *Parking Management Best Practices*, American Planning Association, Chicago, IL, 2006, Table 1-1, pg 7)

Table 13.22: Full Alternative Text

"Parking problems" means different things to different groups. To a motorist, a parking problem means he or she could not find a parking space (on-street or off-street) within a reasonable distance to his or her destination.

To residents, a “parking problem” may be the lack of adequate parking for their own cars, or the use of “their” parking supply by nonresidential users of nearby destinations. To developers, a “parking problem” could involve the costs imposed by zoning regulations to require parking, or the financial burdens of other local regulations. Everyone may find a “parking problem” in the aesthetics or other environmental impacts of parking facilities and the traffic they generate.

Because of these sometimes competing interests, local officials often find controversy when developing parking and related strategies. Motorists, residents, visitors, businesses, taxpayers, and others often present competing views, which local officials must consider in forming overall parking management strategies.

It falls to local planning boards and similar groups to work with the wide variety of constituencies involved to develop an effective parking management plan, along with other traffic plans, to allow an area to thrive both economically and socially.

The topic of parking management is only introduced here. Entire texts and other books have been written on the subject from many different points of view. The student is encouraged to consult the literature for additional, more detailed treatments.

13.5 Closing Comments

Without a place to park at both ends of a trip, the automobile would be a very ineffective transportation medium. Because our society relies so heavily on the private automobile for mobility and access, the subject of parking needs and the provision of adequate parking facilities is a critical element of the transportation system.

References

- 1. *Parking Generation*, 4th Edition, ssInstitute of Transportation Engineers, Washington, D.C., 2010.
- 2. Chrest, A., et al., *Parking Structures: Planning, Design, Construction, Maintenance, and Repair*, 3rd Edition, Springer Science and Business, New York, NY, 2001.
- 3. Litman, T., *Parking Management Best Practices*, American Planning Association, Chicago, IL, 2006.
- 4. Litman, T., *Parking Management: Strategies, Evaluation, and Planning*, Victoria, Transportation Policy Institute, Vancouver BC, CAN, September 2016.
- 5. *Parking Requirements for Shopping Centers*, 2nd Edition, Urban Land Institute, Washington, D.C., 1999.
- 6. Weant, R. and Levinson, H., *Parking*, Eno Foundation for Transportation, Westport, CT, 1990.
- 7. *Recommended Zoning Ordinance Provisions for Off-Street Loading Space*, National Parking Association, Washington, D.C., 1995.
- 8. van der Waerden, P. and Timmermans, H., “Car Drivers Characteristics and the Maximum Walking Distance Between Parking Facility and Final Destination,” *Journal of Transport and Land Use*, Vol 10, No 1, World Society for Transport and Land Use, Berkeley, CA, 2017.
- 9. Teknomo, K, and Kaznouri, H., “Parking Behavior in Central Business District: A Case Study of Surabay, Indonesia,” *EASTS Journal*, Vol 2, No 2, 1997.
- 10. Rastogi, R., “Validating Stated Parking Duration of Drivers in Kota City, India,” *Research Gate*, www.researchgate.net, May 2, 2014.

- 11. *The Dimensions of Parking*, 4th Edition, Urban Land Institute and National Parking Association, Washington, D.C., 2000.
- 12. *Guidelines for Parking Facility Design and Location: A Recommended Practice,* Institute of Transportation Engineers, Washington, D.C., April 1994.

Problems

1. 13-1. A high-rise apartment complex with 600 dwelling units is to be built. What is the expected peak parking demand for such a facility, assuming that it is in an area without significant transit access?

2. 13-2. A shopping center with 600,000 sq ft of gross leasable floor area is planned. It is expected that 10% of the floor area will be devoted to movie theaters and restaurants. What peak parking demand would be expected for such a development?

3. 13-3. Based on typical zoning regulations, what number of parking spaces should the developers of Problems 1 and 2 be asked to provide?

4. 13-4. A new office complex will house 2,000 back-office workers for the securities industries. Few external visitors are expected at this site. Each worker will account for 1.0 person-destinations per day. Of these, 85% are expected to occur during the peak hour. Only 7% of the workers will arrive by public transportation. Average car occupancy is 1.3. What peak parking demand can be expected at this facility?

5. 13-5. A parking study has found that the average parking duration in the city center is 35 minutes, and that the following spaces are available within the 14-hour study period (6:00 AM–8:00 PM) with a 90% efficiency factor. How many vehicles may be parked in the study area in one 14-hour day?

Number of Spaces	Time Available
100	6:00 AM–8:00 PM
150	12:00 Nn–8:00 PM
200	6:00 AM–12:00 Nn
300	8:00 AM–6:00 PM

13.2-25 Full Alternative Text

6. 13-6. Consider the license-plate data for a study period from 7:00 AM to 2:00 PM, which appear in the following table. For this data:

    1. Find the duration distribution and plot it as a bar chart.
    2. Plot the accumulation pattern.
    3. Compute the average parking duration.
    4. Summarize the overtime and parking violation rates.
    5. Compute the parking turnover rate.

    Is there a surplus or deficiency of parking supply on this block? How do you know this?

Parking Space	7:00	7:30	8:00	8:30	9:00	9:30	10:00	10:30	11:00	11:30	12:00	12:30	1:00	1:30	2:00
1 hr meter	100	√	-	150	√	√	246	385	-	691	√	√	-	810	√
1 hr	-	468	√	630	√	485	-	711	888	927	√	√	108	√	-
1 hr	848	911	√	√	221	747	922	√	-	787	√	452	√	-	289
1 hr	-	-	206	√	242	√	√	-	899	√	205	603	812	√	√
1 hr	-	-	566	665	√	333	848	√	999	-	720	-	802	√	-
1 hr	-	690	-	551	√	√	347	√	265	835	486	√	-	721	855
Hydrant	-	-	-	-	-	-	-	777	-	-	-	-	-	-	-
2 hr meter	-	-	940	√	√	505	608	√	√	√	121	123	√	-	880
2 hr	636	√	√	√	√	-	582	√	√	811	919	√	711	√	√
2 hr	-	399	√	√	401	904	√	√	789	√	556	√	√	√	232
2 hr	-	416	√	√	√	√	√	-	658	√	292	844	493	√	√
2 hr	188	√	√	-	665	558	√	√	√	213	√	-	779	√	√
2 hr	-	-	-	277	√	336	409	√	√	884	√	√	713	895	431
2 hr	-	-	837	√	√	418	575	√	952	√	√	√	√	-	762
2 hr	-	506	√	√	-	786	√	√	√	527	606	√	385	√	√
Hydrant	-	-	-	-	-	518	-	-	-	758	-	-	-	-	-
3 hr	-	079	√	√	√	√	√	√	√	-	441	√	611	√	√
3 hr	256	√	√	√	√	-	295	√	√	338	√	-	499	√	√
3 hr	-	-	848	√	√	√	√	√	-	933	√	√	√	√	√
Bus stop	-	-	-	-	-	740	142	-	-	-	-	-	-	-	-
Bus stop	-	-	-	-	-	915	-	-	-	-	-	-	-	-	-
Bus stop	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Bus stop	-	-	-	-	-	-	-	-	-	-	-	-	-	-	818
Bus stop	-	-	-	888	-	175	755	-	-	-	-	-	-	-	397

[13.2-26 Full Alternative Text](#)

Chapter 14 Traffic Impact Studies and Analyses

In the United States, landmark legislation regarding the environmental impact of Federal actions came into effect with the signing of the National Environmental Policy Act (NEPA) on January 1, 1970 [1].

NEPA's procedures apply to all agencies in the executive branch of the Federal government, and generally require an *environmental assessment* (EA) document that will result in a finding of no significant impact or an *environmental impact statement* (EIS) that includes a detailed process for its development, submission, review, and consequent decision-making. The legislation also established the Council on Environmental Quality (CEQ). See Ref. [2] for more information on CEQ and its role.

Over the years since 1970, the definition of a "major federal action" by an agency in the executive branch has come to include most things that the agency could prohibit or regulate [3]. This has come to the current state of the practice that a project is required to meet NEPA guidelines whenever a Federal agency provides any portion of the financing for the project, and sometimes when it simply reviews the project.

An EA or consequent EIS includes attention to a full range of potential environmental impacts, and certainly includes those due to traffic. Indeed, the traffic impact work is generally an important input to the assessment of noise and pollution impacts (due to the related mobile source emissions).

The individual states have generally passed their own environmental legislation, extending the range of needed environmental impact assessments, following a process akin to the Federal one. For instance, the State of New York has its State Environmental Quality Review Act (SEQR) [4].

Local governments generally have their own legislation and processes for actions taken at their own level. New York City has a CEQR process [5], and there is a full range of such legislation throughout the United States. It is imperative that a practicing professional be aware of the governing laws in a jurisdiction, including which level of government has purview on a given project and which agency will be the *lead agency* for the specific project or activity.

A traffic impact analysis (TIA) is a common element of both EA and EIS documents required by the relevant level of government, or can be required as a separate submittal by an agency that has jurisdiction. Despite its short form name, the TIA must have information on impact and proposed mitigation.

Different jurisdictions have their own guidance on what constitutes a significant impact that requires mitigation be evaluated and addressed. In some cases, it is a certain change in v/c ratio at an intersection, and/or level of service change on approaches, at intersections, and/or on arterials. Again, it is imperative that the practicing professional have knowledge of the specific requirements on state and local levels, and federal (if applicable).

Generally, all such legislation requires that the environmental impacts be identified and estimated using the current state of the art/state of the practice tools and methods, and that mitigation be investigated and proposed to the extent possible. It is *not* required that full mitigation be achieved, but rather that the impacts and effects be fully disclosed so that the relevant decision-maker as established by law can make a fully informed decision on whether the project is permitted to proceed. Indeed, the challenges to a decision tend to be on whether the process was followed rigorously, whether proper methodologies were used, and whether there was full disclosure of impacts. Provided with proper information and following an orderly process, the law explicitly vests the decision authority in a specific agency or designated position, *and the decision itself is not a valid basis for litigation*.[1]

[1]This last statement, indeed the entire paragraph, is drawn from both the law and the practice, but nuances can be better explained by the attorney on the team.

As a practical matter, a recommendation that there are impacts that cannot be fully mitigated will be the basis for lively discussion in the review process, and reviews tend to go smoother when full mitigation is feasible.

14.1 Scope of This Chapter

The reader should not expect to have mastered the ability to conduct full and complete traffic impact assessments after simply reading this chapter, or even this entire text.

Rather, this chapter is intended to focus the reader's attention on how information from the preceding chapters must be brought to bear in executing a traffic impact assessment, and on how the reader must use this knowledge to create design concepts that can mitigate impacts.

One of the authors has taught a project-based course centered around a traffic impact assessment, and has used the ITE *Transportation Impact Analyses for Site Development* [6] as a companion text for that course (it was a second course in a sequence, and also covered several chapters from this text, and built upon the chapters taught in the first course).

This chapter provides an overview of the process and techniques in the next two sections, and then provides two case studies that can be used as course projects or as the basis for discussion. The chapter does *not* provide total solutions to either of the case studies, and this is intentional: At this point, the learning is best done by meeting the challenges in a project-based experience, interacting with the instructor. (Issues are identified, and some guidance provided, but a definitive "correct solution" is absent, by design.)

14.2 An Overview of the Process

This section focuses on the process as shown in Figure 14.1. There are variations on this (and more comprehensive versions, for specific localities), but it can serve the purpose of this chapter. The boxes are numbered for convenience, and will be referred to as “steps” in this section.

Figure 14.1: One Rendering of the Process to be Followed in a TIA

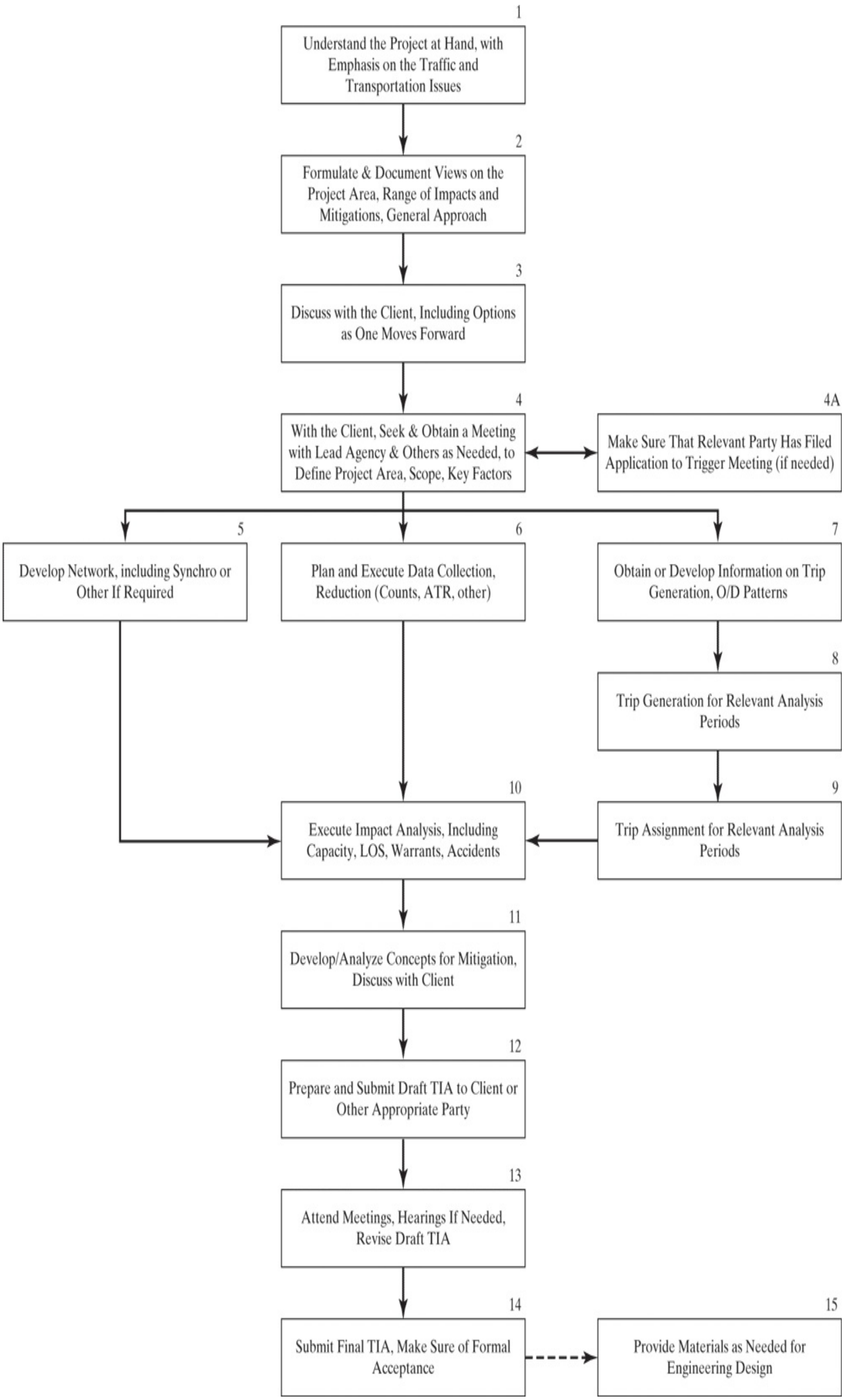

1
Understand the Project at Hand, with Emphasis on the Traffic and Transportation Issues
2
Formulate & Document Views on the Project Area, Range of Impacts and Mitigations, General Approach
3
Discuss with the Client, Including Options as One Moves Forward
4
With the Client, Seek & Obtain a Meeting with Lead Agency & Others as Needed, to Define Project Area, Scope, Key Factors
4A
Make Sure That Relevant Party Has Filed Application to Trigger Meeting (if needed)
5
Develop Network, including Synchro or Other If Required
6
Plan and Execute Data Collection, Reduction (Counts, ATR, other)
7
Obtain or Develop Information on Trip Generation, O/D Patterns
8
Trip Generation for Relevant Analysis Periods
9
Trip Assignment for Relevant Analysis Periods
10
Execute Impact Analysis, Including Capacity, LOS, Warrants, Accidents
11
Develop/Analyze Concepts for Mitigation, Discuss with Client
12
Prepare and Submit Draft TIA to Client or Other Appropriate Party
13
Attend Meetings, Hearings If Needed, Revise Draft TIA
14
Submit Final TIA, Make Sure of Formal Acceptance
15
Provide Materials as Needed for Engineering Design

Figure 14.1: Full Alternative Text

Step 1 appears rather obvious, but it is too common that both clients and other professionals have views and preferences that influence the traffic engineer's work, and are sometimes (a) oblivious to the needs (and limits) of the traffic component, (b) have competing priorities that would impose impractical limits on the traffic component, and (c) are insensitive to the effect of other decisions on the traffic component.

A few examples include projects in which:

- At an early stage, the client precluded increased vehicle occupancies as a mitigation or plan element because of *probable* resistance from the workforce, assuring themselves that the cognizant public agency would surely accommodate more vehicle-intensive solutions;

- In the midst of another project, after traffic circulation had been planned, the project architect casually moved 500,000 sq ft of one tower completely across the project to another tower because "it looked more balanced." (To that point, it had been part of the "dominant tower" and "signature building.") In response to a groan, the architect simply asked, "Oh, does that change anything for you?"

Of course, each professional specialty on a team brings their own special expertise, perspectives, and values. All of them however—including the traffic engineers—have to appreciate both the interactions and the synergies. One of the authors enjoyed learning about "shade studies" that determined what vegetation was feasible on a particular urban project, and how it was "obvious" to architects that the signature building entrance just had to be oriented to the south for another.

Step 2 is a working plan for the entire project and specifically preparation for Step 4. Important issues are:

- What do the local regulations and practices require in terms of hours/days of data, analysis tools, required methods [e.g., Highway Capacity Manual (HCM), critical lane volume (CLV), and Synchro], triggers for mitigation, and other?

- What do the local regulations and practices require in terms of site development as it affects traffic? This may include set-backs, buffer zones, mandated allowances for parking and transit, mandated emphasis on traffic calming within the site, and other.

- What days and periods within the day(s) are justified by the project or by local regulations and practices?
    - AM, Mid-Day, PM are commonly required
    - Weekend may be required for some developments
- What analysis periods are required?
    - The most common are the existing condition, the future no-build (FNB), and future build (FB).
    - On some projects, the period of construction is so large and/or so long that the peak of the construction period must also be analyzed.
- What exactly is the base case for analysis? Is it the future no build (FNB) with existing signalization or the FNB with optimized signals, or other?
- What are the local growth rates to be used, and what if any planned and approved major developments are there? Is there local guidance (i.e., a guidebook or set of tables) on trip generation rates or is the ITE *Trip Generation* publication [7] sufficient?
- What are the relevant road system and transit facilities? How far will concentrated traffic flow before it disperses into background levels? What intersections and other key points are affected and need analysis?
- What data exists, and in what form is it available (counts, ATR stations, accident data, and other)?

That is not to say that all these questions will be answered in Step 2, but rather that the engineer must get a handle on each of them, particularly with regard to the extent of the project area and the intersections/facilities to be affected.

Step 3 is important in that the client and/or their team (attorney, architect) must understand the difficulties that are likely to be encountered in the process and the need for a reasonable project area. If it is obvious to the traffic professional that the impact area is larger than the nearby intersection, it will also be obvious to the professionals who are reviewing the TIA. Some clients may like it to be smaller (sometimes much smaller), but unless they have time to iterate with the reviewing agency, both time and cost will dictate that they be made aware of reality rather early.

Of course, that is *not* to say that the defined project area (i.e., the impact area) *must* be large. There are many projects in which only a small number of key intersections are involved before the traffic distributes into the background levels. Smaller is better, but reasonable is best.

Step 4—the meeting with the cognizant lead agency and other relevant parties (e.g., state, counties, towns)—is a defining moment in the project. It is generally a formal step in the local process, needs an agenda, and must seek to arrive at a set of action items covering all points raised in Step 2, but especially:

- Mutual agreement on the defined impact area (the “project area”) for the analysis;

- Mutual agreement on the supporting data program (data to be collected, amounts, number of days, and so forth), and on the key intersections/facilities to be analyzed; and

- Clear understanding of local requirements on growth rates to be used, approved projects that need to be considered beyond the background growth rates, and standard practice documents to be used (many jurisdictions have publications or memoranda specifying these, including tools and techniques to be used).

To avoid wasted effort and awkwardness later, it is best—let us say vital—that the lead agency sign off on the agreed items. This may take the form of a letter accepting the minutes of the meeting, but in some cases is an email acknowledging the discussion and accepting the minutes. Verbal approvals are not really useful, if only because personnel change over the course of a project and because people invariably have slightly different recollections.

The traffic professional must be aware that in some jurisdictions, such formal meeting and agreement is simply not possible *unless the formal application to start the process has been filed* (Step 4A).[2] This application generally involves more issues than traffic (including timing issues known only to the client), and has to be filed by the appropriate party with the client's approval. The NYS SEQR is one such process.

[2]Some dialog has been known to take place on a conversational basis, but it is not reasonable for the traffic professional or the client to expect these to be binding.

After Step 4, the project tends to kick into high gear. A set of three major activities happen, somewhat concurrently:

1. The network is encoded into the analysis tool(s) to be used, whether they be spreadsheets or computer programs (Step 5); more is said on this in the next section;

2. The agreed data is collected and summarized (Step 6), and made ready for the analysis; and

3. The references for trip generation rates that have been agreed upon, whether based upon local practices or ITE [7] or information provided by the traffic professional and sourced, is documented (Step 7), used to establish number of trips generated (Step 8), and then distributed onto the network (Step 9) for each relevant time period (e.g., AM, Mid-Day, and PM) at each relevant stage (e.g., Existing, FNB, and FB).

The next step requires a good bit of careful work, but is somewhat anticlimatic, given the above—Step 10 is the actual execution of the analysis that is the heart of the "impact" section of the TIA.

Step 11 is the most creative and demanding part of the entire exercise, because design is a creative process as well as an orderly one: The traffic professionals must identify one or more mitigation plans that address the adverse impacts that become clear in the analysis work of Step 10.

The sets of solutions available in Step 11 (the mitigation plans) include:

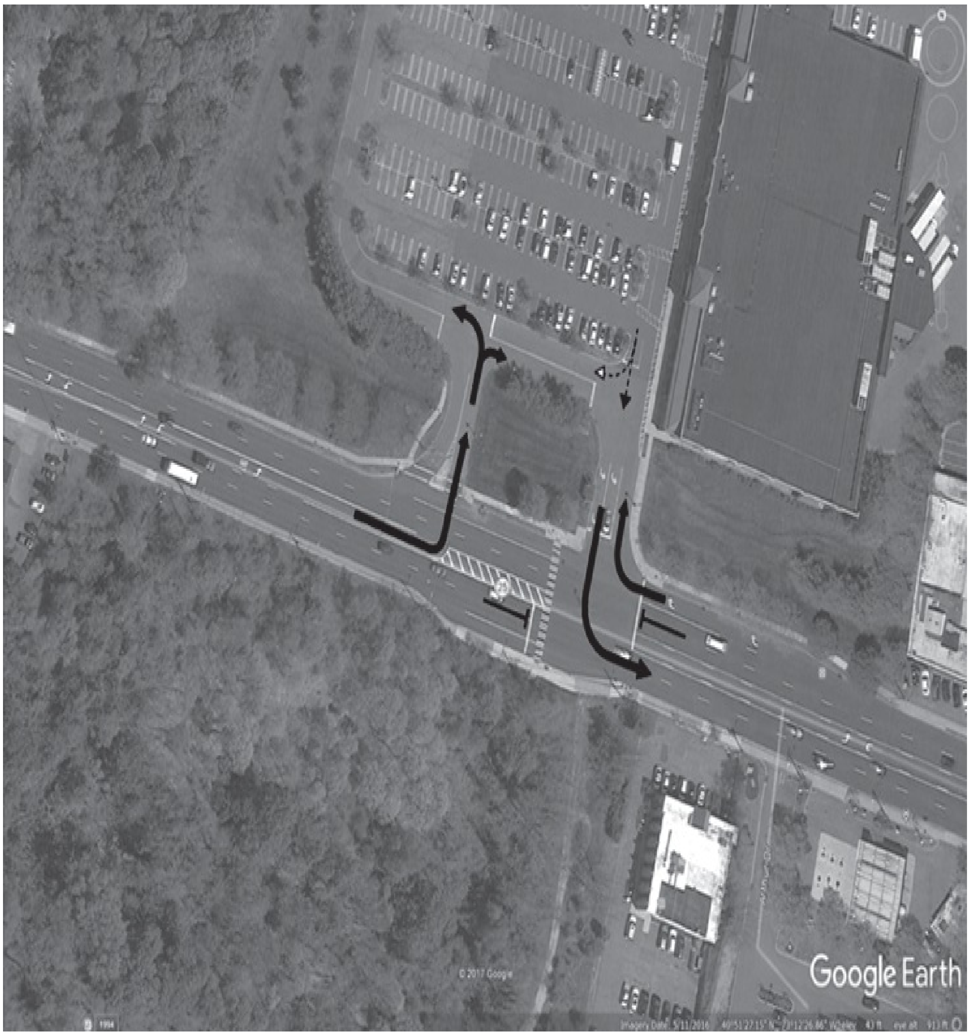
- Retiming of signals, including different phases and cycle lengths, as well as different offset plans;

- Addition of signals as warranted by increased traffic or other factors;
- Addition of driveways for the project as needed, consistent with local access management policies and maintenance of arterial flow and function. The driveways may include designs, one of which is illustrated in Figure 14.2. In that illustration, left turn conflicts are removed from the arterial and two signal phases suffice (left turn into the site is not signalized, but has good storage);

Figure 14.2: An Illustrative Innovative Design for Handling Turns

(a) The Objective in This Case Is to Remove Left Turn Conflicts from the Arterial

14.2-1 Full Alternative Text



(b) Left Turns on One Phase, Thru Traffic on Another Phase

(Source: Google Earth Image, with movement arrows superimposed.)

[14.2-1 Full Alternative Text]

- Addition of lanes approaching and departing from intersections, to increase throughput on specific approaches;
- Addition of new intersection(s) as needed, or of new lanes;
- Use of jug-handles and other solutions to reduce left turn activity;
- Programs to increase average vehicle occupancy; and
- Other traffic demand management solutions, such as shifting work

hours at the project site and/or sponsoring employee vanpooling and/or transit check programs.

For solutions in the last two bulleted items, it is incumbent upon the applicant (the client and their traffic professional) to make it clear how such policies and practices will *really* be put into effect. This will be expected by the TIA reviewers if the mitigation depends upon them, and approval may be contingent upon terms related to the proposed programs.

It is *imperative* in the view of the authors that the client be briefed in advance of the draft report about the mitigation options, and the related first-cut (i.e., rough cut) estimate of costs. This is for a very practical reason—the client will probably have to pay for some or all of the mitigation, and should know what the costs are likely to be.

At this stage (Step 11), there is likely to be some good discussions about the trade-offs amongst alternative mitigation approaches. These will often involve cost, ease of approval, and speed of the approval process (in some cases, the expression "time is money" is very apt, because of the overall project costs and schedule). In some cases, other factors known best to the client will arise, such as work rules for employees, as a cost factor.

Step 12 is the formal preparation of the draft TIA, internal reviews for quality control, client review and comment, and submittal to the lead agency by the appropriate party. This draft may include a CD containing the data files and the input streams for any computer programs used, as well as sample animations.

Step 13 is the review and approval process, which will surely involve meetings convened by the lead agency, may well involve public presentations and hearings, receipt and documentation of comments, and revisions to the TIA.

Step 14 is the submittal of the final TIA document, either as part of another document or a stand-alone document (depending upon the required process), leading (one would anticipate) to final, formal acceptance.

Step 15 is internal to the client or their team. The functional designs and traffic loads developed in the course of the TIA are an important input to the engineering design that must generally follow acceptance of the TIA by the cognizant agency.

A final note on time frames: this overall process is not instantaneous. All of the above steps can easily consume 6 to 12 months. Following approval of the mitigation plan in the TIA (for that is what acceptance means, as used above), the next steps are detailed engineering design, submittal of permit applications and related approvals, followed by construction. The construction period may be shut down in the winter months, and a Maintenance and Protection of Traffic Plan is generally required as part of the permit process. For sizable projects with a reasonable amount of mitigation work, these extra steps can add up to 15 to 18 months *after* the TIA acceptance. It is possible that the total process may move more rapidly, but that needs to be assessed.

14.3 Tools, Methods, and Metrics

This text has presented information on the state of the art and the state of the practice in traffic engineering, with a strong emphasis on the levels of service as defined in the *HCM*. Chapter 23 presents a critical movement analysis (CMA) approach to intersection capacity analysis. Some states still use such an approach instead of the HCM methodology, which is far more complex. Chapter 22 presents the HCM method.

But the reader *must* appreciate that *it is the local jurisdiction—usually at the state level—that determines the exact method to be used in that jurisdiction*. And some details of design practice (including the acceptance of some design concepts) are sometimes delegated to the local district or regional offices, so that variation within a state can be expected. When roads are solely within the control of a county or town, their rules and procedures may prevail. Therefore, knowledge of local practice and rules is *essential* to the practicing traffic professional.

Fortunately, these rules are usually easily obtained, and are posted on the official web sites of the state or local jurisdiction. Equally fortunate, the review process usually involves a *lead agency* that coordinates the information and needs.

At the same time, it is sometimes natural for counties to have different concerns and priorities than states, or for one region to have more precise rules and practices than its neighbor. Most often, the good will and professionalism of all concerned overcome potential problems, but there are protocols and practices to respect.

To consider the range of practices with regard to just intersection and arterial evaluation, the following is informative (and based in fact):

- Some require impact to be expressed in terms of *level of service (LOS) changes* by intersection or by lane group, and cite the HCS+ software [8] as the expected tool;

- Some want *both LOS and v/c ratio* changes to be reported, with the HCS+ software for single intersections and Synchro [9] for sets of intersections and arterials;

- Others specifically mandate a *CLV* methodology provided by the state (e.g., Maryland). The “not to exceed” CLV is 1,450 veh/h. The procedure is rooted in the method introduced in [10] as interim materials to an earlier edition of the *HCM*. Rather than being considered “dated,” it is re-emerging as an effective and efficient tool, and is the logical foundation for the treatment in Chapter 23 of this text;

- Several states accept *Synchro LOS results* as if they were as equally valid (and exactly the same) as the *HCM* results;

- A number of states require “*Synchro visualizations*” of the traffic conditions, although the actual visualizations are produced by a separate tool (the SimTraffic simulator [9]) that is sold as a companion to Synchro[3];

  [3]The two tools sometimes produce radically different results, particularly when intersection spillback and blockage is involved.

- Some states focus on the *intersections rather than the arterial*, primarily by silence on the arterial impacts (i.e., average travel speeds and arterial LOS); and

- At least one state had begun to focus on *arterials to the exclusion of intersections*, at least in the initial planning-level review. That has evolved to a more balanced view that includes arterial LOS *and* intersection LOS *and* intersection v/c ratio.

In terms of traffic visualizations, there are other tools that are commercially available and have merit, including VISSIM [11] and AIMSUN [12].

Related to the discussion of the critical movement analysis, consider the values of “maximum sum of critical movement volumes” that can be accommodated for various conditions. The computations are done using a lost time per phase of 4.0 seconds and a discharge headway of 1.9 sec/veh for passenger cars (consistent with the HCM saturation flow rate of 1,900 pc/h/ln). With 5% trucks, the discharge headway is changed to 2.0 sec/veh, and used in Table 14.1.

Table 14.1: Values of Maximum Sum of Critical Movement Volumes (veh/h), for Various Conditions, Including Cycle Length

$h =$	1.9 sec/veh,	0% trucks
$pce_{TRUCK} =$	2.0	5% trucks
$t_L =$	4.0 sec/phase	

$v/c = 1.00$

Cycle Length (sec)	Number of Phases		
	2	**3**	**4**
60	1564	1444	1323
70	1598	1495	1392
80	1624	1534	1444
90	1644	1564	1484
100	1660	1588	1516
110	1673	1608	1542
120	1684	1624	1564

$v/c = 0.90$

Cycle Length (sec)	2	3	4
60	1408	1299	1191
70	1438	1346	1253
80	1462	1380	1299
90	1480	1408	1335
100	1494	1429	1364
110	1506	1447	1388
120	1516	1462	1408

Table 14.1: Full Alternative Text

The upper set of values in Table 14.1 is based upon 100% utilization of the green by the vehicles. The lower table shows values for 90% utilization (v/c=0.90).

Some observations are in order, using the lower set:

- For a two-phase signal with a cycle length C = 80 sec, the number shown (namely, 1,462 veh/h) is comparable to the CLV upper limit cited above (namely, 1,450 veh/h);
- Each additional phase decreases the value by about 5%, using C = 80 sec as a reference condition;
- This is probably an overstatement, because added phases tend to imply longer cycle lengths, so a decrease of about 2.5% per added phase is a plausible rule of thumb;
- Within the lower set of values, starting with C = 80 sec as the reference, each increase of 10 sec of cycle length adds 1% to the displayed value. But one must remember discussions in this text that indicate the saturation flow becomes less efficient, so the nominal improvement is probably not significant; and
- Using the same starting point, each decrease of 10 seconds in cycle length loses 2% on the displayed value. But the main purpose of doing so would be queue management in congestion, a different priority.

Finally, this little exercise with Table 14.1 is interesting, but not dominant, for two distinct reasons: (a) the traffic professional is governed by the formal procedure adopted locally, not this exercise, in applying corrections, (b) changes to "maximum sum of critical movement volumes" is not identical to *capacity*, because the nominal losses cited above can be adjusted with shifts of green time (remember, v/c = 0.90 is used) to favor approaches with more lanes and thus more vehicles *and* because a measure such as added phases is generally taken to correct a problem that already degraded the base number.

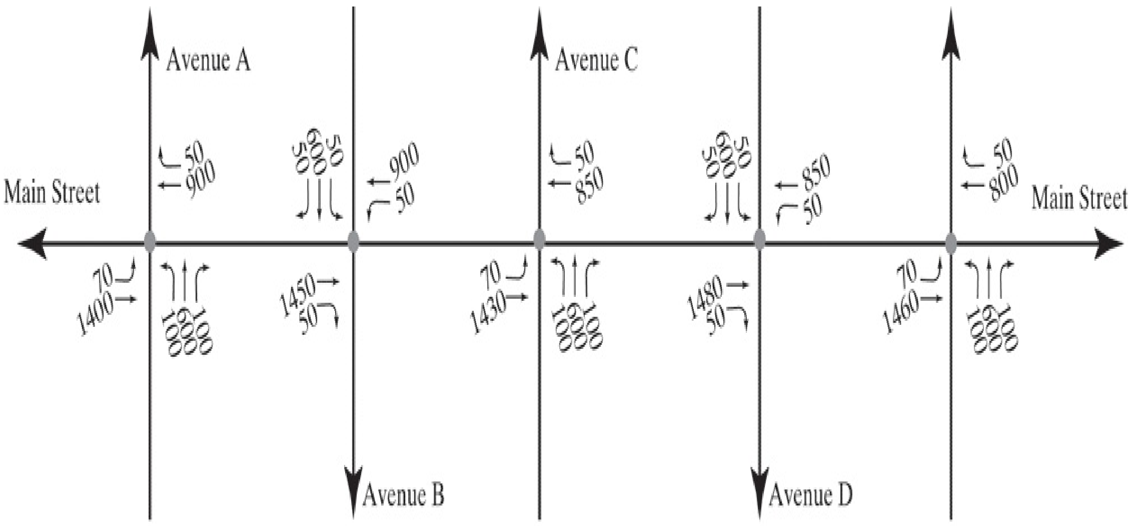
14.4 Case Study 1: Driveway Location

This case study only addresses one fragment of a TIA, namely the effect of driveway location on quality of flow along the arterial. The lessons to be learned by this exercise include:

1. The new driveway will add flows to both the NB and SB arterial flows, and do so “between main street platoons,” thus making coordination less effective;
2. The new driveway will take vehicles from the passing platoons, thus leaving holes in them and making them less cohesive, also making coordination less effective; and
3. The new driveway can totally disrupt the NB and/or SB green bandwidths if it is poorly placed within the block.

Refer to Figure 14.3. This case is also addressed in Chapter 21. Figure 14.4 shows one Synchro solution that can be used as a starting point.

Figure 14.3: Inputs for Case Study 1



All the intersections are spaced at 1,500 feet.
2 lanes per direction on Main St
Left turn bays on Main St, 250 ft, both directions
Right Turn on Red (RTOR) prohibited everywhere
60 ft/s free flow speed, Main Street
All avenues have 2 moving lanes, but are one-way streets
PHF = 0.91
Minimum ped crossing times = 17 sec across side streets, 30 sec Main St

Figure 14.3: Full Alternative Text

Figure 14.4: Sample Synchro Time-Space Diagram for Case Study 1

Time-Space Diagram – Main Street
Traffic Flow Diagram, 50th Percentile Flow and Green Times

6/1/2017

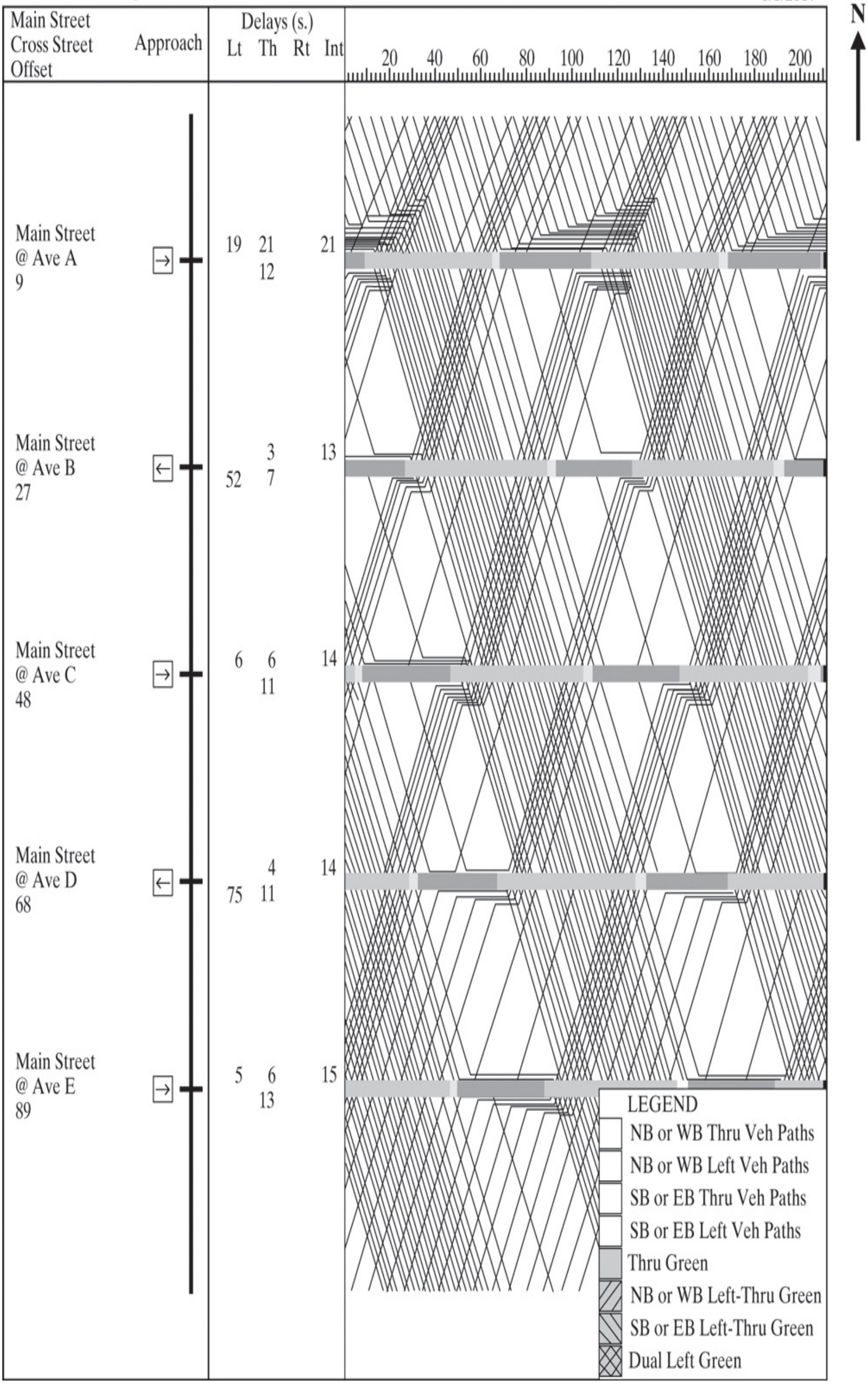


Figure 14.4: Full Alternative Text

A significant development is proposed for a site on the east side of Main Street (north is to the left), between Avenues B and C. The specifics are:

- The development will add “X” vehicles per hour to the Main Street traffic heading NB and “X” vehicles per hour to the Main Street traffic heading SB. Each of these flows will turn into the driveway(s) provided;
- At the same time, the development will generate “X” exiting vehicles per hour exiting and heading SB on Main Street and another “X” vehicles per hour and heading NB on Main Street; and
- The preferred driveway (pair) is to be the configuration shown in Figure 14.3. The signalized intersection is to be located “N” feet north of Avenue B.

Using Synchro and SIMTraffic, the reader/student is asked to analyze the impacts four scenarios, namely:

		N, feet	
		375	750
X, vph	100	Scenario 1	Scenario 3
	200	Scenario 2	Scenario 4

14.4-2 Full Alternative Text

It is not necessary to propose mitigation at this stage, unless the instructor makes that an addition part of the assignment.

The reader/student may choose to use VISSIM or AIMSUN for the visualizations; that is acceptable. The choice can be made based upon tools available to the college or to the student (limited student versions of some tools are available at low cost).

The reader/student should be aware that the emphasis is on *relative* impacts of the four scenarios. However, some of the tools may model some effects differently than others, and the signal optimization program (i.e., Synchro) may consider some queuing effects differently than the simulation models. The reader/student may have to “fine tune” the signal timing results. Also, some of the simulation models may give metrics such as arterial travel time more easily than others.

14.5 Case Study 2: Most Segments of a Traffic Impact Analysis

Case Study 2 is much more comprehensive than the first case study, but is less intense than a full traffic impact study, in that (a) the trip assignment paths are specified in detail, (b) the project area is defined, (c) the data is provided, and is balanced so that it is internally consistent,[4] and (d) applicable local rules are provided, in terms of requirements.

[4]In general, traffic counts provide numbers that simply do not add up. This may be due to parking lots or generators within individual links, but it can also be due to simple random errors in the field work. In the latter case, and assuming reasonable variations, the counts are then balanced by the analyst to reflect a more realistic snapshot.

Some readers/students may believe that the specified requirements on buffer zones, parking spaces per unit of activity, and other elements are very restrictive. They were however assembled from *real* requirements in *real* locations within the United States. There are locations that have most if not all of these requirements.

The only major embellishment is the requirement that the site allow for *both* (a) transit access as if 20% of the trips were using public transportation, *and* (b) parking that recognizes that 95% of the trips will arrive by auto. While this is unrealistic *in the short term* and might be viewed by some as a burden, the authors believe that it could represent good long-term planning on the part of the local jurisdiction. For instance, it provides a critical part of the enabling infrastructure that will lead to a future transit use of 20% (other parts include a bus route system that is sufficiently complete to enable the trips, and sufficiently frequent to make them attractive).[5]

[5]There was a time when fully accessible transportation was questioned on a cost-effectiveness basis, given that the full set of requirements did not exist for a meaningful number of trips. Those requirements included curb cuts, accessible entrance, accessible restrooms, legible signage, and then accessible buses and rail transit. Policy decisions to build this

infrastructure systematically over decades has resulted in accessible *systems* in a number of locales.

Some information is *not* provided. Actual trip generation rates not are provided, but can be obtained from Ref. [6] or [7] or from a web search that provides the reader/student with specific numbers used in certain jurisdictions. Spatial requirements for parking are not provided, but can be found or estimated as suggested in Problem 14-3 at the end of this chapter. Other needed information can be found in various sources by the motivated reader/student, or provided by the instructor.

14.5.1 The Project Area and the Existing Condition

Figure 14.5 shows the project area, including the two parcels that will be of interest in this case study. Table 14.2 provides details on the streets (number of lanes, etc.) and the available right-of-way.

Table 14.2: Details of Streets within the Project Area, Case Study 2

	R.O.W.	Lanes
SINCLAIR AVE	100 ft	2 12-ft lanes each direction; 10-ft shoulders; grass median, including 12-ft left turn bay, 150 ft long plus taper
EVERGREEN ST	80 ft	1 12-ft lane each direction; 10-ft shoulders; grass median, including 12-ft left turn bay, 150 ft long plus taper
WEEPING WILLOW	80 ft	1 12-ft lane each direction; 10-ft shoulders; grass median, including 12-ft left turn bay, 150 ft long plus taper
HAYES ST	80 ft	1 12-ft lane each direction; 10-ft shoulders; grass median, including 12-ft left turn bay, 150 ft long plus taper
N. PIONEER DR	60 ft	2 12-ft lanes WB, highway exit added as 3rd 12 ft lane (on left), then dropped as highway entrance; minimal shoulders
S. PIONEER DR	60 ft	2 12-ft lanes WB, highway exit added as 3rd 12 ft lane (on left), then dropped as highway entrance; minimal shoulders
SHADOW ROCK RD	80 ft	1 12-ft lane each direction; 10-ft shoulders; grass median, including 12-ft left turn bay, 150 ft long plus taper
WILLA CATHER HWY	Note 1	Note 2

Note 1 = All land between highway & service roads is part of the right of way;

Highway is presently 2 12-ft lanes each direction, 10 ft shoulders, 40 ft grass median;

Highway bridges over sinclair ave are a pair, each with 50 ft of roadway.

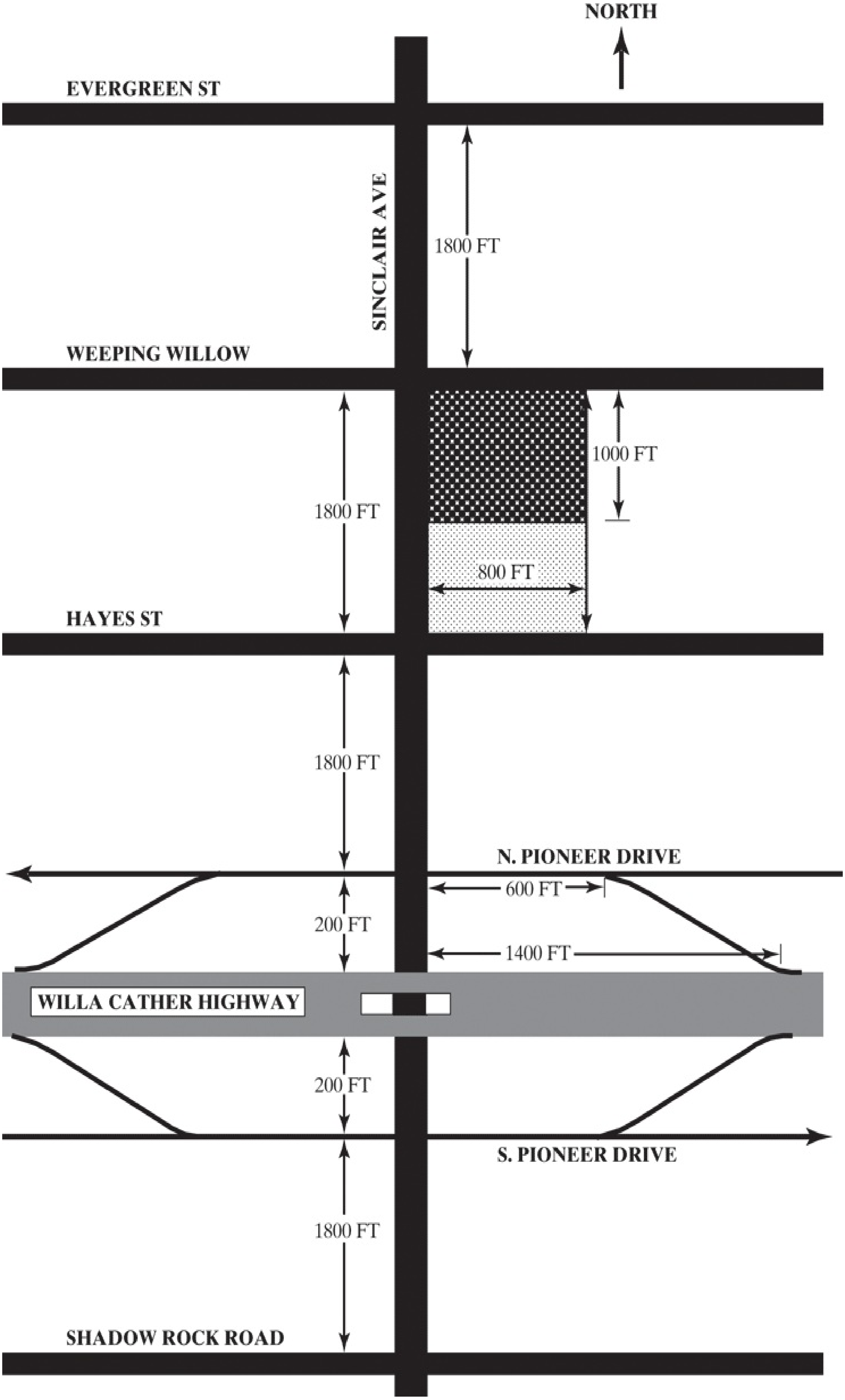
Note 2 = All ramps are single lanes, tapered; measurements shown on figure are to the gore area or the r.o.w. boundary.

Signal phasing = Protected lefts required, no permissive lefts allowed; lefts can be lead or lag; rtor generally permitted.

Distances = All distances shown along sinclair ave are measured to the r.o.w. boundaries, and do not include the r.o.w. itself.

Table 14.2 Full Alternative Text

Figure 14.5: Project Area for Case Study 2



NORTH
EVERGREEN ST
SINCLAIR AVE
1800 FT
WEEPING WILLOW
1800 FT
1000 FT
800 FT
HAYES ST
1800 FT
N. PIONEER DRIVE
600 FT
200 FT
1400 FT
WILLA CATHER HIGHWAY
200 FT
S. PIONEER DRIVE
1800 FT
SHADOW ROCK ROAD

Figure 14.5: Full Alternative Text

Table 14.3 provides the hourly *volumes* for four periods that may be relevant to the project at hand. Other important information, such as the PHF, is included in Table 14.3.

Table 14.3: Traffic Counts in the Project Area, Existing Condition, Case Study 2

PHF = 0.85 %TRUCKS = 5.0%

ALL VOLUMES IN HOURLY VPH EXISTING CONDITION, AM PEAK	NB			SB			EB			WB		
	L	T	R	L	T	R	L	T	R	L	T	R
SINCLAIR AVE & EVERGREEN ST	120	960	50	60	650	50	60	270	40	80	360	40
SINCLAIR AVE & WEEPING WILLOW	40	970	60	60	660	50	80	280	60	80	320	80
SINCLAIR AVE & HAYES ST	60	990	60	60	690	80	40	320	40	40	280	40
SINCLAIR AVE & N. PIONEER DR	90	1010			630	100				60	600	100
SINCLAIR AVE & S. PIONEER DR		1040	120	100	590		60	550	100			
SINCLAIR AVE & SHADOW ROCK RD	90	1060	60	60	590	60	40	320	40	60	340	60

WILLA CATHER HWY			
HOURLY	COUNT STATION AT EAST END OF SKETCH	5200	EB
		4850	WB
	AT RAMPS TO N. PIONEER	300	
	FROM N. PIONEER	250	
	TO S, PIONEER	270	
	FROM S. PIONEER	300	

EXISTING CONDITION, MIDDAY WEEKDAY	NB			SB			EB			WB		
	L	T	R	L	T	R	L	T	R	L	T	R
SINCLAIR AVE & EVERGREEN ST	84	672	35	42	455	35	42	189	28	56	252	28
SINCLAIR AVE & WEEPING WILLOW	28	679	42	42	462	35	56	196	42	56	224	56
SINCLAIR AVE & HAYES ST	42	693	42	42	483	56	28	224	28	28	196	28
SINCLAIR AVE & N. PIONEER DR	63	707			441	70				42	420	70
SINCLAIR AVE & S. PIONEER DR		728	84	70	413		42	385	70			
SINCLAIR AVE & SHADOW ROCK RD	63	742	42	42	413	42	28	224	28	42	238	42

WILLA CATHER HWY			
HOURLY	COUNT STATION AT EAST END OF SKETCH	3640	EB
		3395	WB
	AT RAMPS TO N. PIONEER	210	
	FROM N. PIONEER	175	
	TO S, PIONEER	189	
	FROM S. PIONEER	210	

EXISTING CONDITION, PM PEAK	NB			SB			EB			WB		
	L	T	R	L	T	R	L	T	R	L	T	R
SINCLAIR AVE & EVERGREEN ST	40	660	80	40	960	60	50	360	120	50	270	60
SINCLAIR AVE & WEEPING WILLOW	60	670	80	80	1030	80	50	320	40	60	280	60
SINCLAIR AVE & HAYES ST	40	670	40	40	970	40	80	280	60	60	320	60
SINCLAIR AVE & N. PIONEER DR	100	650			1010	60				120	550	100
SINCLAIR AVE & S. PIONEER DR		650	60	100	1070		100	600	90			
SINCLAIR AVE & SHADOW ROCK RD	40	590	60	60	1060	40	60	340	90	60	320	60

WILLA CATHER HWY			
HOURLY	COUNT STATION AT EAST END OF SKETCH	4850	EB
		5200	WB
	AT RAMPS TO N. PIONEER	300	
	FROM N. PIONEER	270	
	TO S, PIONEER	250	
	FROM S. PIONEER	300	

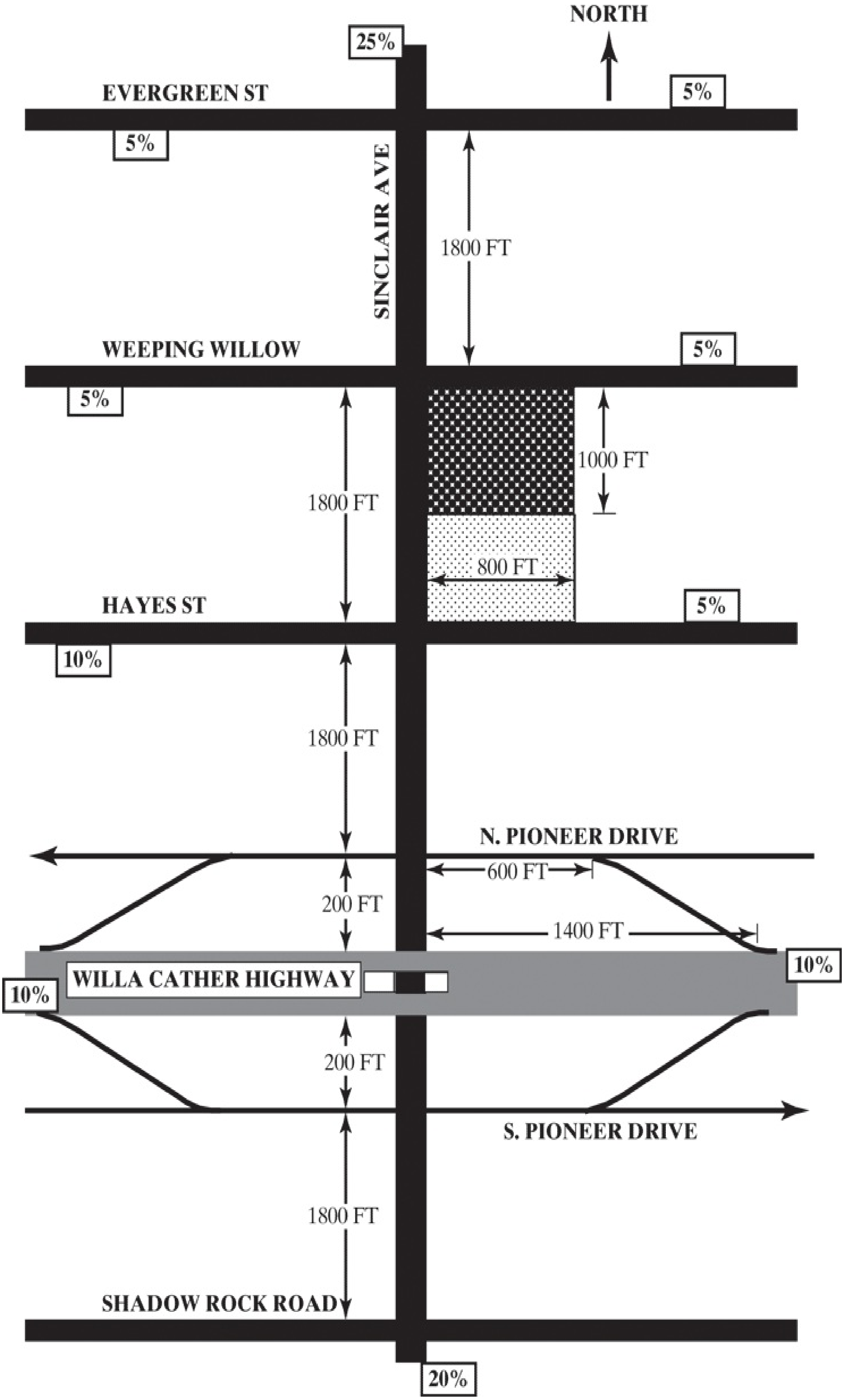
EXISTING CONDITION, SATURDAY	NB			SB			EB			WB		
	L	T	R	L	T	R	L	T	R	L	T	R
SINCLAIR AVE & EVERGREEN ST	120	960	50	60	650	50	60	270	40	80	360	40
SINCLAIR AVE & WEEPING WILLOW	40	970	60	60	660	50	80	280	60	80	320	80
SINCLAIR AVE & HAYES ST	60	990	60	60	690	80	40	320	40	40	280	40
SINCLAIR AVE & N. PIONEER DR	90	1010			630	100				60	300	100
SINCLAIR AVE & S. PIONEER DR		1040	120	100	590		60	250	100			
SINCLAIR AVE & SHADOW ROCK RD	90	1060	60	60	590	60	40	320	40	60	340	60

WILLA CATHER HWY			
HOURLY	COUNT STATION AT EAST END OF SKETCH	5200	EB
		4850	WB
	AT RAMPS TO N. PIONEER	300	
	FROM N. PIONEER	250	
	TO S, PIONEER	270	
	FROM S. PIONEER	300	

Table 14.3: Full Alternative Text

Figure 14.6 shows the sources of *new* traffic attracted to the developments at the site(s) of interest. Trips return from the site(s) to these destinations. Note that the *magnitude* of the traffic is *not* specified, either as arriving traffic or departing traffic in each of the four analysis periods (AM, Mid-Day, PM, and weekend).

Figure 14.6: Source of New Traffic Added because of the Development(s), Case Study



NORTH
25%
EVERGREEN ST
5%
5%
SINCLAIR AVE
1800 FT
WEEPING WILLOW
5%
5%
1000 FT
1800 FT
800 FT
HAYES ST
5%
10%
1800 FT
N. PIONEER DRIVE
600 FT
200 FT
1400 FT
10%
WILLA CATHER HIGHWAY
10%
200 FT
S. PIONEER DRIVE
1800 FT
SHADOW ROCK ROAD
20%

Figure 14.6: Full Alternative Text

Discussion Point 1: The reader/student must find an appropriate source of trip generation rates for the uses proposed (see next subsection). This includes consideration of both entering and departing traffic in each time period that is relevant.

Note that Figure 14.6 applies only to new traffic; it may be true (depending upon the land use at the site of interest) that some percentage of the traffic arriving at the site is drawn from existing traffic that passes the site. The practical implication is that the *existing* traffic that goes to the site does not add to the volumes for the purpose of impact assessment. It does however use parking that needs to be provided within the site, and this must be taken into account.

Discussion Point 2: Clearly, there is less impact (and therefore less mitigation) if a goodly percentage of the traffic using the site is diverted from traffic that would pass in any case. This point is made in some cases for gas stations and for breakfast shifts at fast food operations. But for other uses, the percentage is probably small. The reader/student must obtain information for the specific uses and/or argue the case.

14.5.2 Proposed Use(s) of the Two Site(s)

Refer to Figure 14.5. The plan is to develop the northern property as commercial space, specifically a suburban shopping mall, and the southern property as a multiplex theater. The multiplex is to have eight theaters, of which four will have 400 seats and four will have 200 seats. The shopping mall is to be built out to the limits of the local code (more on this below).

If it suits this plan, the two properties can be combined and considered as one.

Discussion Point 3: At the risk of getting ahead of ourselves, the local code requirements in the next subsection will require buffer zones between the two sites if they remain on the records as two distinct parcels. This may use space that could be dedicated to parking or other uses. However, there may be other issues of interest to the client (i.e., the owner of the parcels), such as future flexibility allowed by keeping them distinct or even some tax implications (none of the authors are knowledgeable in this area).

14.5.3 Local Code and Local Ordinance Requirements

Table 14.4 contains requirements that are extracted from various local codes, so that the reader/student will have a reasonable set of design considerations for this case study.

Table 14.4: Requirements from Local Ordinances

1. Each property must be at least 5 acres in this zoning, and have at least 300 ft of frontage on the main road. Building heights may not exceed 35 feet. The building(s) cannot cover more than 30% of the property.
2. There must be 100 ft of buffer facing all roads and other properties, except that the buffer may be reduced to 50 ft if the use on the other property is nonresidential. If the use on the other property is residential,[1] a 6 ft barrier wall must be constructed of stone or heavy timber and of appearance consistent with like construction in the jurisdiction.
3. For the shopping mall and like uses, five parking spaces per 1,000 sq ft of gross floor area (GFA) must be provided.
4. For the motion picture theater(s), one parking space per three seats must be provided.
5. The design must allow for up to 20% of the trips arriving by bus transit, for anticipated mode shifts over coming decades.[2] However, in assigning trips to modes, no more than 5% of the anticipated trips can be assigned to transit, given current realities.
6. Access management principles must be strongly considered and used to the maximum extent feasible, consistent with state requirements.[3]
7. Adequate loading bays for goods delivery and pickup must be provided, generally separated from the visitors to the site(s).
8. Entrances/exits for goods vehicles need to accommodate large tractor-trailers. All entrances/exits must accommodate a hook-and-ladder fire truck.
9. Internal circulation at any site shall be such that safe pedestrian and transit uses are encouraged, and that traffic-calming principles are used.
10. Mitigation plans are expected (a) if the v/c ratio on any approach at any existing intersection is increased by more than 0.03, (b) if the LOS on any approach at any existing intersection is decreased (made poorer) by one level of service, and (c) if the arterial LOS on any arterial is decreased (made poorer) by one level of service.
11. If additional intersection(s) is/are proposed, the intersection v/c and LOS at each such intersection must be comparable to nearby existing intersections *and* the arterial level of service cannot be decreased (made poorer).
12. The future no-build and future build traffic levels are to be taken as ten years from the present (i.e., existing) condition.
13. A section of the impact analysis shall address the construction phase. This section may be qualitative and descriptive, *if* no significant impact is anticipated to peak hour traffic, but the means for achieving this shall be described.

[1]The adjacent properties to the east are zoned commercial in this case study.

[2]Assume that the primary (sole) bus routes run north–south on the main arterial.

[3]Use the State of Florida guidelines or practices, if needed.

Full Alternative Text

14.5.4 Other Given Conditions

With regard to *basic traffic engineering*, the local practice dictates:

- Use a loss time of 4.0 seconds per phase;
- For saturation flow rates, use 1,900 pc/h/ln for the through lanes;
- All left turns at signalized intersections are to be protected; and
- RTOR at signalized intersections is prohibited, except if there is appropriate channelization in a separate lane, with at least YIELD sign control.

The *background traffic*, exclusive of these sites, is to be taken as 3.0% per year (and is to be compounded). No other planned or approved developments of note exist, or are assumed to be incorporated in the background growth.

The local *topography* has no significant elevation changes.

With regard to *tools and references*, the reader/student is expected to use *at least* the following tools, references, and practices in the course of this case study:

1. Signal timing (including alternate phase plans) for both existing and all future conditions is to be done using Synchro;
2. Given that the existing signal timings are not available (they usually are) *and* given that the local jurisdiction had scheduled signal retiming work in any case, “existing” signal timings are to be determined using Synchro optimization for the most suitable cycle length (more on this, in the next subsection);
3. Simulations and visualizations are to be done with a tool such as SIMTraffic, VISSIM, or AIMSUN;
4. Levels of service and v/c ratios for intersections and arterials[6] may be obtained from Synchro results, unless the simulations/visualizations indicate that there is a clear inconsistency;

[6]Remember that arterial LOS is based upon the average travel speed of the *through* vehicles.

5. Trip generation rates will be based upon [6] or [7] or other documented source;

6. Parking layout and internal circulation use relevant ITE or other documents; and or material in Chapter 13 of this text.

7. For this particular case study, this textbook may be used as a reference document.

If the traffic professional uses HCS+ or other such tool for capacity analysis or signal timing, that traffic professional must submit a reasoned argument on why these results are more relevant than Synchro, recognizing that this particular jurisdiction prefers and is accustomed to receiving Synchro results.

Note: The remaining subsections divide the traffic impact and mitigation work into a set of "elements" so that they can be easily assigned in parts, generating discussion and learning as the course progresses. The following schedule, used by one of the authors, is shown, merely as a suggestion. Periodic presentations by the student groups are to be encouraged, and a final presentation of the impact and proposed mitigation is an essential experience. (Work submitted at the intermediate due dates was evaluated on the basis of the learning experience, but the final comprehensive report and presentation was viewed with a higher level of expectation.)

Element	Due Date (in Weeks) Relative to the Week the Case Study Was Begun
1	1
2	2
3	4
4	5
5	6
6	7
7	9
8	10

14.5-6 Full Alternative Text

14.5.5 Element 1: System Cycle

Recommend a system cycle length along Sinclair Avenue, considering signal spacing, a reasonable vehicle speed, traffic volumes, and number of signal phases.

14.5.6 Element 2: The Developer's Favorite Access Plan

Consider the following hypothetical situation: The developer is very interested in a design that combines the two parcels, centers the development, and uses only two major driveways for public access, including transit. Refer to Figure 14.4. (Details of parking lot and internal circulation are not shown; they would remain for the traffic professional to work out, within the overall concept.)

Clearly, this creates two five-legged intersections that need a multiphase signal plan.

The reader/student is to evaluate the operational needs of these intersections and the required flow patterns given the arrival/departure suggested within Figure 14.7. Table 14.3 provides the existing flows, and the growth rate is known. For the added traffic, the reader/student may have to make some assumptions (just to get started) or use the trip generation references already cited.

Figure 14.7: A Concept of the Development, from the Client

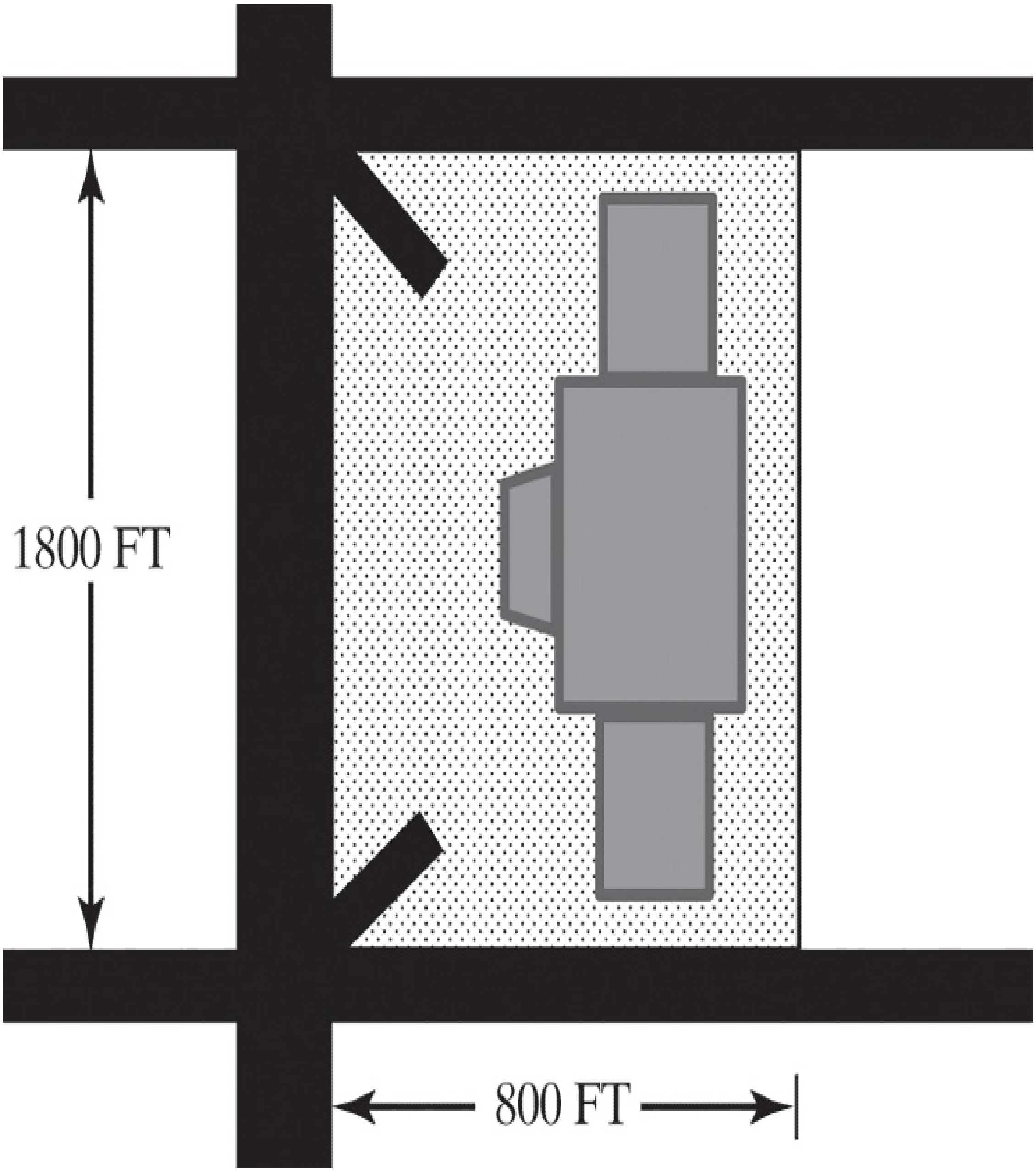


Figure 14.7: Full Alternative Text

Keep in mind that (a) the client *really* wants this “innovative” approach, and (b) you have to be the responsible professional, sometimes achieving what is desired by creative design and sometimes presenting a reality that serves the client well, while disappointing their initial notions.

Discussion Point 4: Yes, the authors have revealed that they are pessimistic about making the five-leg intersections work, given the flow rates and patterns. However, the obligation is for *you* to find a successful solution *or* to clearly explain to the client why the concept is not in their best interests.

Discussion Point 5: Remember that in *Element 1*, you recommended a system cycle length. That cycle length may be inconsistent with the values needed at the proposed five-leg intersections

The end product of the work on Element 2 should be a PowerPoint presentation explaining your findings, in terms that both your peers and a nonspecialist can understand. Use no more than five slides.

14.5.7 Element 3: Existing Conditions, Capacity, and LOS Analyses

Use Synchro (and other tools, as needed) to estimate the existing and FNB conditions at each intersection, and for each lane group at each intersection, and for the overall arterial.

Summarize the results in the format of Table 14.5. (The future build [FB] column will be completed in a later element of this work.)

Table 14.5: Format to Be Used in Element 3 Work

	v/c			LOS		
	Existing	FNB	FB w/o Mitigation	Existing	FNB	FB w/o Mitigation
Intersection						

Intersection	Lane Group						

Arterial		Average Travel Speed (mph)			LOS		
	NB						
	SB						

Note: FB not done in Element 3.

Table 14.5: Full Alternative Text

One table is needed for each time period of interest. If it is less than the four periods for which data has been provided, indicate why the period is irrelevant to the analysis, given the anticipated uses.

Discussion Point 6: Before this analysis is begun, the instructor will have to specify the system cycle length to be used by all teams in the course. This can be assigned after the submission and discussion of the Element 1 work.

Anticipate what traffic improvements will be needed due solely to the “FNB” traffic levels.

Discussion Point 7: It is logical that some operational problems will arise, due simply to the annual growth of traffic. In some cases, these would logically require some attention/action, independent of the added traffic due to the specific new development (i.e., the two sites at hand). It is good to know this in discussions with the local jurisdiction.

Given that you were not provided with existing signal timing, you may want to use critical movement analysis as a reasonable method of determining the initial signal timing and phasing. This can be refined by using Synchro.

Discussion Point 8: This work is part of this student assignment, but existing signal plans generally exist and are available from the local jurisdiction or state.

Discussion Point 9: In the same spirit of a limited student assignment, the application of the *traffic signal warrants* from the state’s MUTCD[7,8] is not part of this assignment. In some applications, there are unsignalized intersections that may need to be signalized, due to a combination of background growth and the project at hand. There may also be new intersections due to driveways, and these will have to be evaluated for signalization. The instructor may wish to require a warrant analysis for at least these new prospective warrants.

[7]Remember that the simple satisfaction of one or more traffic signal

warrants does not mandate the installation of a signal. Rather, the sense of the MUTCD is that a signal should not be installed *unless* one or more warrants are satisfied, and the satisfaction of warrants is simply the trigger that allows this evaluation (an engineering study involving many factors) to proceed.

[8]Some states specify that certain warrants are not used determining whether a signal is justified (e.g., the peak hour volume warrant).

Discussion Point 10: Again, in the same spirit of a limited student assignment, this particular case study does not require the reader/student to acquire and analyze the accident experience in the project area.

14.5.8 Element 4: Trip Generation

The reader/student has already been guided to the sources from which to obtain trip generation *rates*. These will often be expressed in terms of trips per hour for the peak hour and other hours. The "peak hour" for the development's traffic may however *not be the same as the general peak hour*. This depends upon the proposed use (e.g., supermarket, shopping center, multiplex cinema). The traffic professional must take care that the terminology used in the trip generation source refers to the traditional peak hour, or the peak activity hour (and day) of the specific use.

Likewise, the construction phase—if it has a noticeable impact on traffic levels at all—may actually define the peak hour for analysis.[9]

[9]For many projects, the construction phase workers will arrive and depart in off-peak hours (relative to the existing peaks), and not create intense loads. In some cases, materials delivery may need to be noted and considered.

In this Element 4, the reader/student is expected to generate estimates of the number and routes of the trips generated *based upon the desired maximum build-out of the site(s)*.

Discussion Point 11: The emphasis in the above sentence is intentional. Given local rules as given in Table 14.4, there are clearly limits imposed by available square footage of the "footprint" of the development. This does not necessarily mean that the rest of the site can support this maximum development. But it is a starting point.[10]

[10]The traffic professional must recognize that the developer and perhaps the architect may well focus on the maximum that can be done, and that it falls to the traffic professional to point out that the support functions also mandated by local rules (e.g., required parking spaces, internal circulation, space for transit, and for goods deliveries) actually impose constraints that limit the size of the development, by the arithmetic imposed.

Discussion Point 12: The local rule on building height may tempt the architect or the traffic professional to think in terms of a two-story shopping mall, with stores on both levels. However, if one is inclined to go down this path, it is important to discuss how many *successful* and *attractive* malls in the region are two-story designs. At the time of this writing, many malls are characterized by large open spaces, common areas, and generally one-story operation. That is not to say a two-story design cannot be found that would succeed, but rather that it should be a discussion point in class. Even with such a design, there are still likely to be significant open spaces and common areas that use up part of the footprint, leaving less gross floor area (GFA) than twice the footprint.

Estimate the required parking, consider the overall properties, and *start* forming opinions on whether the available space can accommodate the parking requirement.

Discussion Point 13: This is an exercise in arithmetic, but an important one. The underlying issues have been discussed above, and in related footnotes. Remember that local rules express parking requirements in terms of spaces per 1,000 sq ft of GFA or spaces per seats (in the theaters). GFA includes common areas, hallways, and such.

It is recognized that the parking layout is going to be a bit of a challenge for the reader/student, unless they are drawing on other knowledge (i.e., sources other than this textbook). Fortunately, such sources exist, in print or on the web. Templates based upon actual developments are also available from Google Earth or other tools.

Discussion Point 14: An area of lively discussion is sometimes the average vehicle occupancy for the different uses at the site. This is rendered moot by the local codes. At the same time, the numbers used in the codes implicitly assume average vehicle occupancies *and* percent trips by auto. As the years pass, the age of these implicit assumptions may be a basis for discussion, if more recent data proves inconsistent.

Determine whether combining the sites provides any advantages.

Discussion Point 15: This point has already been raised, but this is a logical point to remind the reader/student of it.

Likewise, take into account the special local requirement on building now to accommodate future transit usage of 20% of the trips.

Discussion Point 16: Note that this is a mandate, and is related to the long-term planning of the local jurisdiction. The spirit and intent is that this transit service be accommodated on-site, in attractive and efficient areas. Simply depending upon bus stops on the local arterial will *not* suffice in this submittal.

14.5.9 Element 5: Determine the Size of the Development, Trips Generated, and Internal Circulation

Continue the work begun in Element 4, with special attention to the needs of internal circulation, parking, transit, safe pedestrian travel, and space for goods vehicles. If the overall requirements dictate a smaller build-out than the maximum *nominally* allowed by the zoning, be prepared to address this.

Of necessity, internal circulation will depend upon driveway location(s) and on any special design features on the arterial. While these are nominally Element 6 of the project, there is an overlap.

14.5.10 Element 6: Driveway Locations, Special Arterial, and Intersection Design Features

Taking into account the mandate for access management to the extent feasible, develop recommendations on driveway locations, special arterial features, and intersections (present or proposed).

Consider the "lessons learned" in Case Study 1, with regard to driveway locations.

Remember the needs of goods delivery, which is addressed in the local rules. Consider the possibility (likelihood) of separate driveways for these deliveries, and remember that any "behind" the building has to allow sufficient space to turn and/or maneuver large trucks (some truck bay designs may affect this, also).

14.5.11 Element 7: Mitigation Measures

Return to Table 14.5. Complete the "Future Build" (FB) sections, assuming no mitigation. Highlight all entries that are determined to have an impact, in accord with local rules.

Develop ideas on mitigation in detail. The teams working on the project will need to be creative, while recognizing that improvements cost money and will probably be paid by the client.

A guiding principle sometimes overlooked at this stage is that the mitigation and related design is not simply intended to meet the minimum requirements of local rules, but also to assure that the development operates smoothly into the future and is/remains attractive to the public and the occupants of the businesses. It is sometimes useful for the traffic professional to have this discussion with the client and their team, and to develop a minimal “Plan A” and a “Plan B,” so that the client can see the costs and benefits of any enhancements the traffic professional believes can serve the overall success.

Add columns to Table 14.5 for Plans A and B. Prepare presentation materials. Be ready to engage in discussion on the plan(s) recommended.

Discussion Point 17: Whereas Plans A and B may be for internal discussion, with only one included in the Draft Final Report of Element 8, in this student assignment, both plans will be included in the Draft Final Report.

Discussion Point 18: Because this assignment will culminate in the presentation of the Draft Final Report, and not include the usual next round of agency review and hearings, culminating in a Final Report submittal, the word “Draft” is dropped in Element 8.

Discussion Point 19: The reader/student is reminded of the list of possible mitigation actions enumerated in Step 11, and is invited to add to the list as needed for this specific project.

14.5.12 Element 8: Final Report and Presentation

Each group will have 20 minutes to present their findings and recommendations. Business attire is required. The group need not have everyone have a speaking role (although there is merit to that), *but* the instructor may (i.e., will) direct questions to any group member, so that *all* group members have to be fully prepared.

A final report not exceeding 30 pages is required 24 hours prior to the class, sent by email to the instructor in PDF format. The PowerPoint slides for the presentation have to be sent at the same time.

14.6 Closing Comments

The purpose of this chapter has been to introduce the reader/student to the topic of traffic impact studies, including an overview of the process and emphasizing the need for creative design in meeting mitigation needs associated with a significant development. (Minor projects may result in a finding or no significant impact, although some estimates of future traffic load and a permitting process are still involved.)

Case Study 2 was used by one of the authors as the basis for a second course in traffic engineering. Reference [6] was specified as a required companion text in this mode. Early lectures covered other chapters of this textbook and Case Study 1. A presentation and a work session on computer tools was included. When Case Study 2 began, the class time was devoted to additional chapters of this textbook, discussion of each Element when it was initiated and when it was due, and ad hoc discussions based upon information requests from the students.

Case Study 2 can also be used as the basis of a few lectures in a course that is not project-based, with emphasis on the discussion points enumerated throughout.

References

- 1. http://www.epa.gov/oecaerth/basics/nepa.html
- 2. http://nepa.gov
- 3. http://en.wikipedia.org/wiki/National_Environmental_Policy_Act
- 4. http://www.dec.ny.gov/permits/357.html
- 5. http://www1.nyc.gov/site/oec/environmental-quality-review/environmental-quality-review.page
- 6. *Transportation Impact Analyses for Site Development: An ITE Recommended Practice*, Institute of Transportation Engineers, Washington, D.C., 2010.
- 7. *Trip Generation, 9th edition*, Institute of Transportation Engineers, Washington, D.C., 2012.
- 8. HCS+ Release 7.2.1, distributed by the McTrans Center, University of Florida, 2017; http://mctrans.ce.ufl.edu/hcs/
- 9. SYNCHRO Studio 10, Synchro plus SimTraffic and 3D Viewer, 2017. http://www.trafficware.com/synchro.html
- 10. “Interim Materials on Highway Capacity,” *Transportation Research Circular 212*, January 1980, Transportation Research Board, National Academy of Sciences, Washington, D.C.
- 11. VISSIM 9, PTV, 2017. http://vision-traffic.ptvgroup.com/en-us/products/ptv-vissim/
- 12. AIMSUN 8, TSS-Transport Simulation Systems, October 2017. http://www.aimsun.com

Problems

Note: The instructor may prefer that all problems in this chapter be done as group or team assignments, with the team not to exceed three members (perhaps four in a very large class).

1. 14-1. Do the analysis and impact assessment as specified in Case Study 1.

2. 14-2. For the higher value of “X” and the better value of “N” in Case Study 1 as determined in Problem 14-1, recommend any additional mitigation measures appropriate, and provide the supporting analysis.

3. 14-3. For Case Study 1 and the higher value of “X,” lay out a functional design of the parking lot, allowing for both the entering and departing values of “X,” and recognizing that they may be competing for internal roadway use and for parking spaces over the hour of analysis.

   If the student is not provided with information on the parking requirements and does not have access to a reference providing that information, the student should either (a) consult sources on the web, using a search engine, (b) use a tool such as Google Earth to “visit” a known suburban parking lot, and estimate the space per parked vehicle (taking into account space needed to travel to the parking spaces, and any separation between or at the end of rows of cars).

4. 14-4. For Case Study 2, execute and submit Element 1 in accord with the schedule in the chapter or the instructor’s specification. Submit an analysis for the group, not to exceed three pages.

5. 14-5. For Case Study 2, execute and submit Element 2 in accord with the schedule in the chapter or the instructor’s specification.

6. 14-6. For Case Study 2, execute and submit Element 3 in accord with the schedule in the chapter or the instructor’s specification.

7. 14-7. For Case Study 2, execute and submit Element 4 in accord with the schedule in the chapter or the instructor’s specification.

8. 14-8. For Case Study 2, execute and submit Element 5 in accord with the schedule in the chapter or the instructor’s specification.

9. 14-9. For Case Study 2, execute and submit Element 6 in accord with the schedule in the chapter or the instructor’s specification.

10. 14-10. For Case Study 2, execute and submit Element 7 in accord with the schedule in the chapter or the instructor’s specification.

11. 14-11. For Case Study 2, execute and submit Element 8 in accord with the schedule in the chapter or the instructor’s specification.

Part III Interrupted Flow Facilities: Design, Control, and Level of Service

Chapter 15 The Hierarchy of Intersection Control

The most complex individual locations within any street and highway system are at-grade intersections. At a typical intersection of two 2-way streets, there are 12 legal vehicular movements (left turn, through, and right-turn from four approaches) and four legal pedestrian crossing movements. As indicated in Figure 15.1, these movements create many potential conflicts where vehicles and/or pedestrian paths may try to occupy the same physical space at the same time.

Figure 15.1: Typical Conflicts at a Four-Leg Intersection

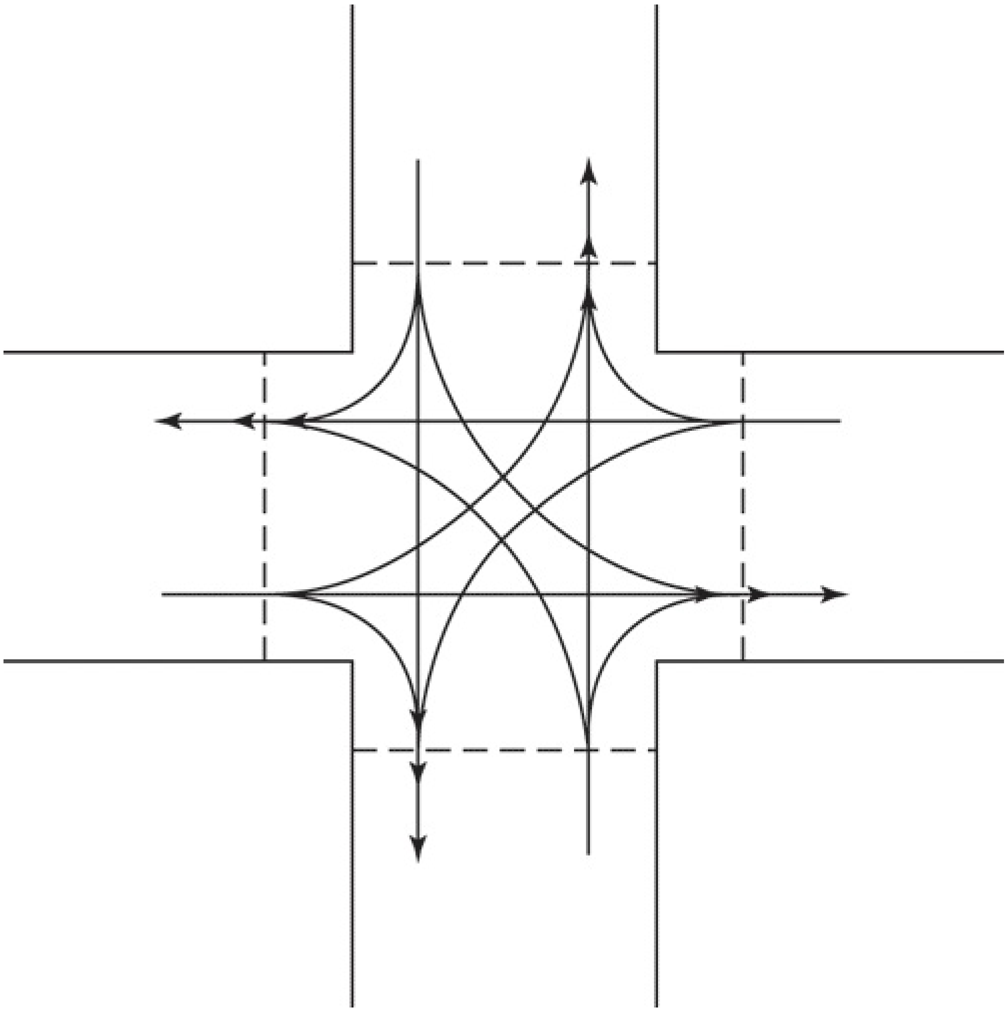


Figure 15.1: Full Alternative Text

As illustrated, there are a total of 16 potential vehicular crossing conflicts: four between through movements from the two streets; four between left-turning movements from the two streets, and eight between left-turning movements and through movements from the two streets. In addition, there are eight vehicular merge conflicts, as right- and left-turning vehicles merge into a through flow at the completion of their desired maneuver. Pedestrians add additional potential conflicts to the mix.

The critical task of the traffic engineer is to control and manage these conflicts in a manner that ensures safety and provides for efficient movement through the intersection for both motorists and pedestrians.

There are three basic levels of control that can be implemented at an intersection:

- Level I—Basic rules of the road;

- Level II—Direct assignment of right-of-way using YIELD or STOP signs; and
- Level III—Traffic signalization.

There are variations within each level of control as well. The selection of an appropriate level of control involves a determination of which (and how many) conflicts a driver should be able to perceive and avoid through the exercise of judgment. Where it is not reasonable to expect a driver to perceive and avoid a particular conflict, traffic controls must be imposed to assist the driver in doing so.

Two factors affect a driver's ability to avoid conflicts: (1) a driver must be able to see a potentially conflicting vehicle or pedestrian in time to implement an avoidance maneuver, and (2) the volume levels that exist must present reasonable opportunities for a safe maneuver to take place. The first involves considerations of sight distance and avoidance maneuvers, while the second involves an assessment of demand intensity, the complexity of potential conflicts that exist at a given intersection, and finally, the gaps available in major movements.

A rural intersection of two farm roads contains all of the potential conflicts illustrated in Figure 15.1. However, pedestrians are rare, and vehicular flows may be extremely low. There is a low probability of any two vehicles and/or pedestrians attempting to use a common physical point simultaneously. At the junction between two major urban arterials, the probability of vehicles or pedestrians on conflicting paths arriving simultaneously is quite high. The sections that follow discuss how a determination of an appropriate form of intersection control can be made, highlighting the important factors to consider in making such critical decisions.

15.1 Level I Control: Basic Rules of the Road

Basic rules of the road apply at any intersection where right-of-way is not explicitly assigned through the use of traffic signals, STOP, or YIELD signs. These rules are spelled out in each state's vehicle and traffic law, and drivers are expected to know them. At intersections, all states follow a similar format. In the absence of control devices, the driver on the left must yield to the driver on the right when the vehicle on the right is approaching in a manner that may create an impending hazard. In essence, the responsibility for avoiding a potential conflict is assigned to the vehicle on the left. Most state codes also specify that through vehicles have the right of way over turning vehicles at uncontrolled intersections. It is also almost universal that a pedestrian crossing a street legally has the right-of-way over all vehicles.

Operating under basic rules of the road does not imply that no control devices are in place at or in advance of the intersection, although that could be the case. Use of street-name signs, other guide signs, or advance intersection warning signs does not change the application of the basic rules. They may, however, be able to contribute to the safety of the operation by calling the driver's attention to the existence and location of the intersection.

In order to safely operate under basic rules of the road, drivers on conflicting approaches must be able to see each other in time to assess whether an "impending hazard" is present, and to take appropriate action to avoid an accident. Figure 15.2 illustrates a visibility triangle at a typical intersection. Sight distances must be analyzed to ensure that they are sufficient for drivers to judge and avoid conflicts.

Figure 15.2: Visibility Triangle at an Intersection

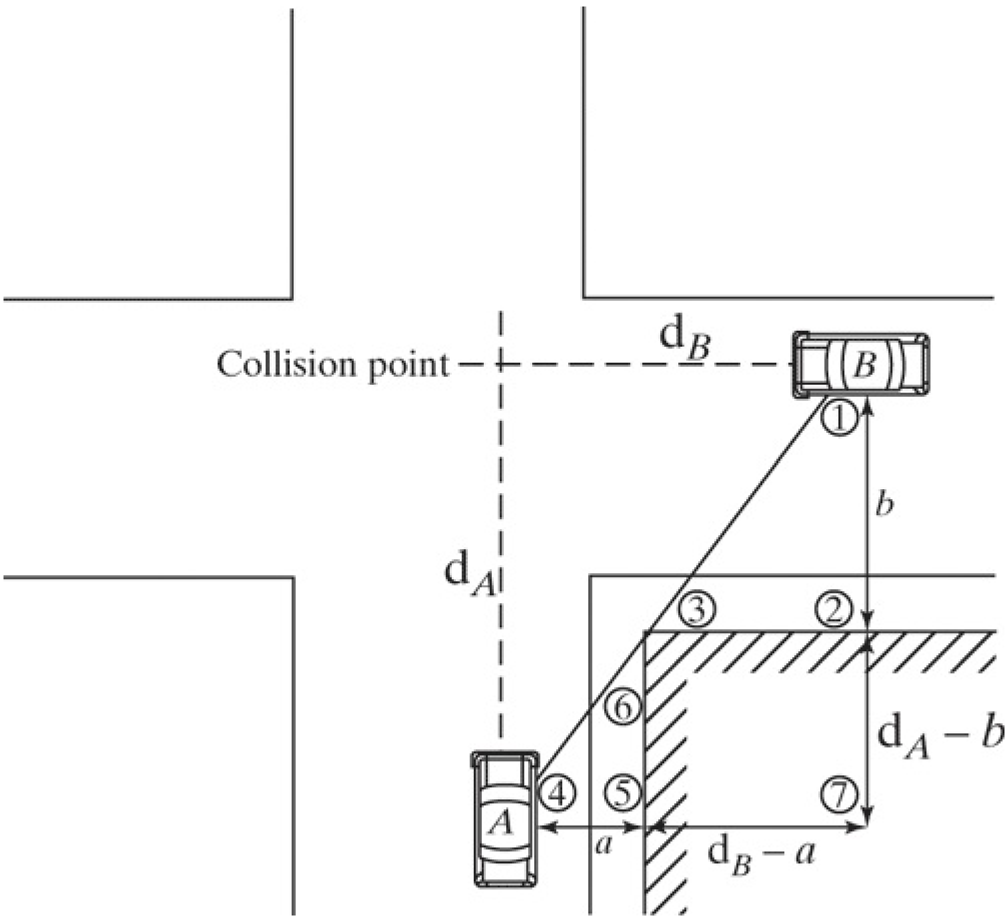


Figure 15.2: Full Alternative Text

At intersections, sight distances are normally limited by buildings or other sight-line obstructions located on or near the corners. There are, of course, four sight triangles at every intersection with four approaches. At the point where the drivers of both approaching vehicles first see each other, Vehicle A is located a distance of dA from the collision or conflict point, and Vehicle B is located a distance dB from the collision point. The sight triangle must be sufficiently large to ensure that at no time could two vehicles be on conflicting paths at distances and speeds that might lead to an accident, without sufficient time and distance being available for either driver to take evasive action.

Note that the sight line forms three similar triangles with sides of the sight obstruction: Δ 123,Δ 147, and Δ 645. Using the similarity of the triangles, a relationship between the critical distances in Figure 15.2 can be established:

bdb−a=dA−badb=adadA−b [15-1]

where:

da=distance from Vehicle A to the collisionpoint, ft;db=distance from Vehi

Thus, when the position of one vehicle is known, the position of the other when they first become visible to each other can be computed. The triangle is dynamic, and the position of one vehicle affects the position of the other when visibility is achieved.

The American Association of State Highway and Transportation Officials (AASHTO) suggests that to ensure safe operation with no control, both drivers should be able to stop before reaching the collision point when they first see each other. In other words, both dA and dB should be equal to or greater than the safe stopping distance at the points where visibility is established. AASHTO standards [1] suggest that a driver reaction time of 2.5 s be used in estimating safe stopping distance and that the 85th percentile speed of immediately approaching vehicles be used. AASHTO does suggest, however, that drivers slow from their midblock speeds when approaching uncontrolled intersections, and recommends use of an immediate approach speed that is assumed to be lower than the design speed of the facility. From Chapter 2, the ssafe stopping distance is given by:

dS=1.47 Si t+Si230 (0.348±0.01G) [15-2]

where:

ds=safe stopping distance, ft;Si=initial speed of vehicle, mi/h;G=grade, %;t

Using this equation, the following analysis steps may be used to test whether an intersection sight triangle meets these sight distance requirements:

1. Assume that Vehicle A is located one safe stopping distance from the collision point (i.e., dA=dS), using Equation 15-2. By convention, Vehicle A is generally selected as the vehicle on the *minor* street.

2. Using Equation 15-1, determine the location of Vehicle B when the drivers first see each other. This becomes the actual position of Vehicle B when visibility is established, dbACT.

3. Since the avoidance rule requires that both vehicles have one safe stopping distance available, the minimum requirement for d_b is the safe stopping distance for Vehicle B, computed using Equation 15-2. This becomes d_{bMIN}.

4. For the intersection to be safely operated under basic rules of the road (i.e., with no control), $d_{bACT} \geq d_{bMIN}$.

Historically, another approach to ensuring safe operation with no control has also been used. In this case, to avoid collision from the point at which visibility is established, *Vehicle A must travel 18 ft past the collision point in the same time that Vehicle B travels to a point 12 ft before the collision point*. This can be analytically expressed as:

$$\frac{d_a+18}{1.47\,S_a}=\frac{d_b-12}{1.47\,S_b} \qquad d_b=(d_a+18)\frac{S_b}{S_a}+12 \qquad [15\text{-}3]$$

where all variables are as previously defined. This, in effect, provides another means of estimating the minimum required distance, d_{bMIN}. In conjunction with the four-step analysis process outlined previously, it can also be used as a criterion to ensure safe operation.

At any intersection, all of the sight triangles must be checked and must be safe in order to implement basic rules of the road. If, for any of the sight triangles, $d_{bACT} < d_{bMIN}$, then operation with no control cannot be permitted. When this is the case, there are three potential remedies:

- Implement intersection control, using STOP- or YIELD-control, or traffic signals.

- Lower the speed limit on the major street to a point where sight distances are adequate.

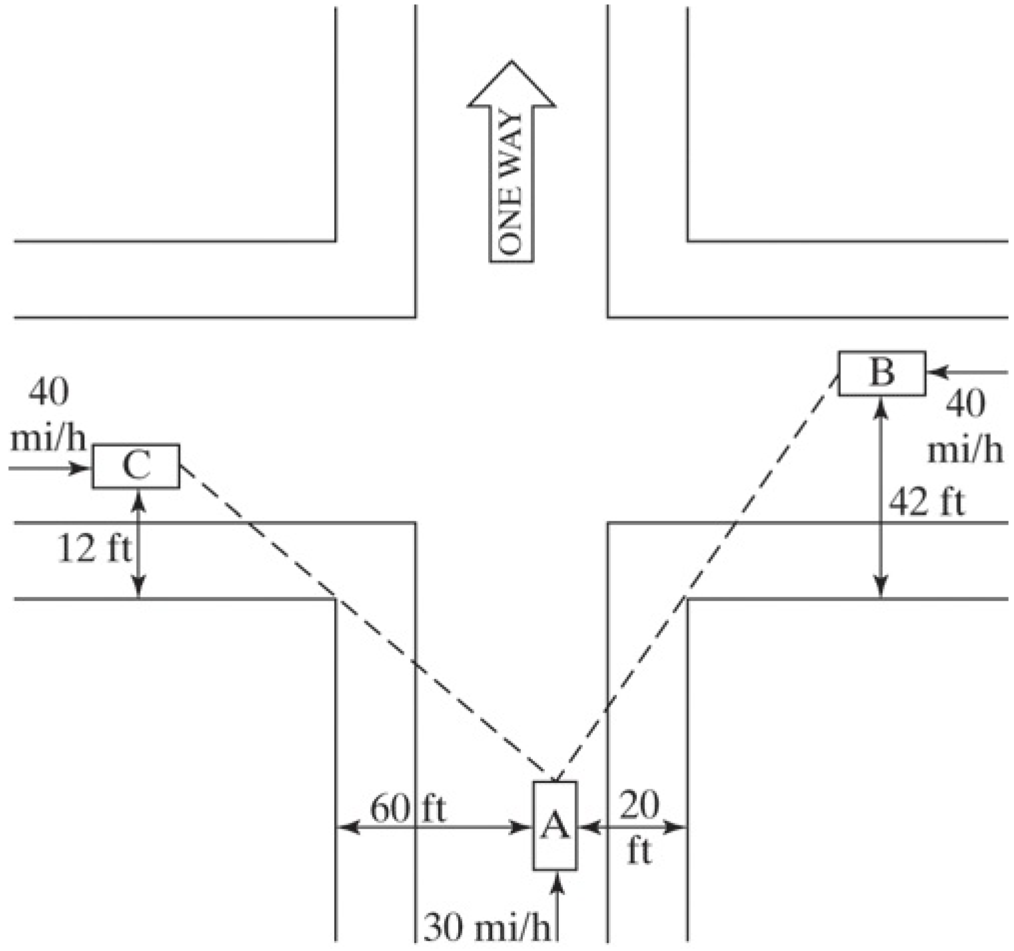
- Remove or reduce sight obstructions to provide adequate sight distances.

The first is the most common result. The exact form of control implemented would require consideration of warrants and other conditions, as discussed in subsequent portions of this chapter. The second approach is viable where sight distances at a series of uncontrolled intersections can be remedied by a reduced, but still reasonable speed limit. The latter depends upon the type of obstruction and ownership rights.

Sample Problem 15-1: Sight Distances at an Intersection

Consider the intersection illustrated in Figure 15.3. It shows an intersection of a one-way minor street and a two-way major street. In this case, there are two sight triangles that must be analyzed. The 85th percentile immediate approach speeds are shown.

Figure 15.3: Sample Problem: Intersection Sight Distance



(reaction time = 2.5 s, level grade)

Figure 15.3: Full Alternative Text

First, it is assumed that Vehicle A is one safe stopping distance from the collision point:

$$d_a = 1.47\times 30\times 2.5+\frac{30^2}{30(0.348+0)}=110.25+86.21=196.46\,ft$$

where 2.5 s is the standard driver reaction time used in safe stopping sight distance computations. Using [Equation 15-1], the actual position of Vehicle B when it is first visible to the driver of Vehicle A is found:

$$d_{bACT} = \frac{a\,d_a}{d_a - b} = \frac{20 \times 196.46}{196.46 - 42} = \frac{3929.2}{154.46} = 25.44 \text{ ft}$$

This must be compared with the minimum requirement for db estimated as either one safe stopping distance (Equation 15-2), or using Equation 15-3:

$$d_{bMIN} = 1.47 \times 40 \times 2.5 + \frac{40^2}{30(0.348+0)} = 147.00 + 153.26 = 300.26 \text{ft}$$

or:

$$d_{bMIN} = (196.46 + 18)\frac{40}{30} + 12 = 297.95 \text{ ft}$$

In this case, both of the minimum requirements are similar, and both are far larger than the actual distance of 25.44 ft. Thus, the sight triangle between Vehicles A and B fails to meet the criteria for safe operation under basic rules of the road.

Consider the actual meaning of the Sample Problem 15-1 results. Clearly, if Vehicle A is 196.46 ft away from the collision point when Vehicle B is only 25.44 ft away from it, they will not collide. Why, then, is this condition termed “unsafe”? It is unsafe because there could be a Vehicle B, further away than 25.44 ft, on a collision path with Vehicle A and the drivers would *not* be able to see each other.

Since the sight triangle between Vehicles A and B did not meet the sight-distance criteria, it is not necessary to check the sight triangle between vehicles A and C. Basic rules of the road may not be permitted at this intersection without reducing major street speeds or removing sight obstructions. This demonstrates that in this case YIELD or STOP signs should be imposed on the minor street as a minimum form of control.

Even if the intersection met the sight distance criteria, this does not mean that basic rules of the road should be applied to the intersection. Adequate sight distance is a *necessary*, but not *sufficient*, condition for adopting a “no-control” option. Traffic volumes or other conditions may make a higher level of control desirable or necessary.

15.2 Level II Control: YIELD and STOP Control

If a check of the intersection sight triangle indicates that it would not be safe to apply the basic rules of the road, then as a minimum, some form of level II control is often imposed. Even if sight distances are safe for operating under no control, there may be other reasons to implement a higher level of control as well. Usually, these would involve the intensity of traffic demand and the general complexity of the intersection environment.

The *Manual of Uniform Traffic Control Devices* (MUTCD) [2] gives some guidance as to conditions for which imposition of STOP or YIELD control is justified. Guidance is not very specific, however, and it requires the exercise of engineering judgment. At this writing, the current MUTCD is the 2009 edition, as amended through May 2012. The MUTCD is available online, and should always be checked for the latest versions and revisions.

The MUTCD gives some general advice for the imposition of *either* STOP signs *or* YIELD signs, under the category of *guidance*. As noted in Chapter 4, "guidance" covers recommended practices. While deviations are permitted, they should be well documented by an engineering study.

As seen in Table 15.1, applying this guidance requires exercise of considerable professional judgment, particularly for the first set of guidelines. The first condition simply addresses a situation in which the sight triangle is insufficient to provide for safety. STOP or YIELD signs may also be used to help establish a major or through road. If all unsignalized approaches to a major road are controlled by STOP or YIELD signs, through drivers have a clear right-of-way. The last condition addresses a situation in which virtually *all* intersections in an area or along an arterial are signalized. If a few isolated locations do not need to be signalized, then they should *at least* have STOP or YIELD signs, as drivers will be expecting some kind of definitive instructions.

Table 15.1: Guidance for

Using STOP or YIELD Control at an Intersection

YIELD or STOP signs should be used at an intersection if one or more of the following conditions exist:
A. An intersection of a less important road with a main road where application of the normal right-of-way rule would not be expected to provide reasonable compliance with the law; B. A street entering a designated through highway; and/or C. An unsignalized intersection within a signalized area.
In addition, YIELD or STOP signs should be used at the intersection of two minor streets or local roads where the intersection has three or more approaches and where one or more of the following conditions exist:
A. The combined vehicular, bicycle, and pedestrian volume entering the intersection from all approaches averages more than 2,000 units per day; B. The ability to see conflicting traffic is not sufficient to allow a road user to stop or yield in compliance with the normal right-of-way rule if such stopping or yielding is necessary, and/or; C. Crash records indicate that five or more crashes that involve failure to yield the right of way at the intersection under the normal right-of-way rule have been reported within a 3-year period; or that two or more such crashes have been reported within a 2-year period.

(Source: *Manual on Uniform Traffic Control Devices*, Federal Highway Administration, Washington, D.C., 2009, with revisions through 2012, pg 50, available at www.fhwa.com.)

Table 15.1: Full Alternative Text

The second set of guidelines provides some more definitive criteria for entering volumes and crash experience.

15.2.1 Two-Way Stop Control

The most common form of Level II control is the two-way STOP sign. In fact, such control may involve one or two STOP signs, depending upon the number of intersection approaches. It *is not* all-way STOP control, which is discussed later in this chapter.

Again under the heading of “guidance,” the MUTCD suggests several conditions under which the use of two-way STOP signs would be justified. This guidance is shown in Table 15.2.

Table 15.2: Guidance for Two-Way STOP Signs

At intersections where a full stop is not necessary at all times, consideration should first be given to using less restrictive measures, such as YIELD signs.
The use of STOP signs on the minor street approaches should be considered if engineering judgment indicates that a stop is always required because of one or more of the following conditions:
A. The vehicular traffic volumes on the through street or highway exceed 6,000 veh/day; B. A restricted view exists that requires road users on the minor street approach to stop in order to adequately observe conflicting traffic on the through street or highway; and/or C. Crash records indicate that three or more crashes that are susceptible to correction by installation of a STOP sign have been reported within a 12-month period, or that five or more such crashes have been reported within a 2-year period. Such crashes include right-angle collisions involving road users on the minor street approach failing to yield the right-of-way to traffic on the through street or highway.

(Source: *Manual of Uniform Traffic Control Devices*, Federal Highway Administration, Washington, D.C., 2009, with revisions through 2012, pg 52, available at www.fhwa.com.)

Table 15.2: Full Alternative Text

Item A establishes a reasonable level of major street traffic that would require use of a STOP sign to allow minor-street drivers to select an appropriate gap in a busy traffic stream. Item B merely restates the need for STOP (or YIELD) control where a sight triangle at the intersection is found to be inadequate. Item C establishes criteria for using a STOP sign to correct a perceived accident problem.

The MUTCD is somewhat more explicit in dealing with inappropriate uses of the STOP sign. Under the heading of a “standard” (i.e., a mandatory condition), STOP (or YIELD) signs *shall not* be installed at intersections where traffic control signals are installed and operating, except where signal operation is a flashing red at all times, or where a channelized right turn exists. This disallows a past practice in which some jurisdictions turned signals off at night, leaving STOP signs in place for the evening hours. During the day, however, an unfamiliar driver approaching a green signal with a STOP sign could become significantly confused. The manual also disallows the use of portable or part-time STOP signs except for emergency and temporary traffic control.

Under the heading of “guidance,” STOP signs *should not* be used for speed control, although this is frequently done on local streets designed in a straight grid pattern. In modern designs, street layout and geometric design would be used to discourage excessive speeds on local streets.

In general, STOP signs should be installed in a manner that minimizes the number of vehicles affected, which generally means installing them on the minor street.

AASHTO [1] also provides sight distance criteria for STOP-controlled intersections. A methodology based upon observed gap acceptance behavior of drivers at STOP-controlled intersections is used. A standard stop location is assumed for the minor street vehicle (Vehicle A in Figure 15.2). The distance to the collision point, d_a, has three components:

- Distance from the driver’s eye to the front of the vehicle (assumed to be 8 ft),
- Distance from the front of the vehicle to the curb line (assumed to be 10 ft), and
- Distance from the curb line to the center of the right-most travel lane approaching from the left, or from the curb line to the left-most travel lane approaching from the right.

Thus:

daSTOP=18+dcl [15-4]

where:

daSTOP=distance of Vehicle A on a STOP-controlledapproach from the collisionpoint, ft, anddcl=distance from the cu

The required sight distance for Vehicle B on the major street for STOP-controlled intersections is found as follows:

dbMIN=1.47 Smaj tg [15-5]

where:

dbMIN=minimum sight distance for VehicleB approaching on the major (u

Average gaps accepted are best observed in the field for the situation under study. In general, they range from 6.5 to 12.5 seconds depending upon the minor street movement and vehicle type, as well as some of the specific geometric conditions that exist.

For most STOP-controlled intersections, the design vehicle is the passenger car, and the criteria for left-turns are used, as they are the most restrictive. Trucks or combination vehicles are considered only when they make up a substantial proportion of the total traffic on the approach. Values for right-turn and through movements are used when no left-turn movement is present. For these typical conditions, AASHTO recommends the use of tg=7.5s.

Sample Problem 15-2: Sight Distance at a STOP-Controlled Intersection

Consider the case of a STOP-controlled approach at an intersection with a two-lane arterial with a design speed of 40 mi/h, as shown in Figure 15.4.

Figure 15.4: Sample Problem in STOP-Control Sight

Distance Requirements

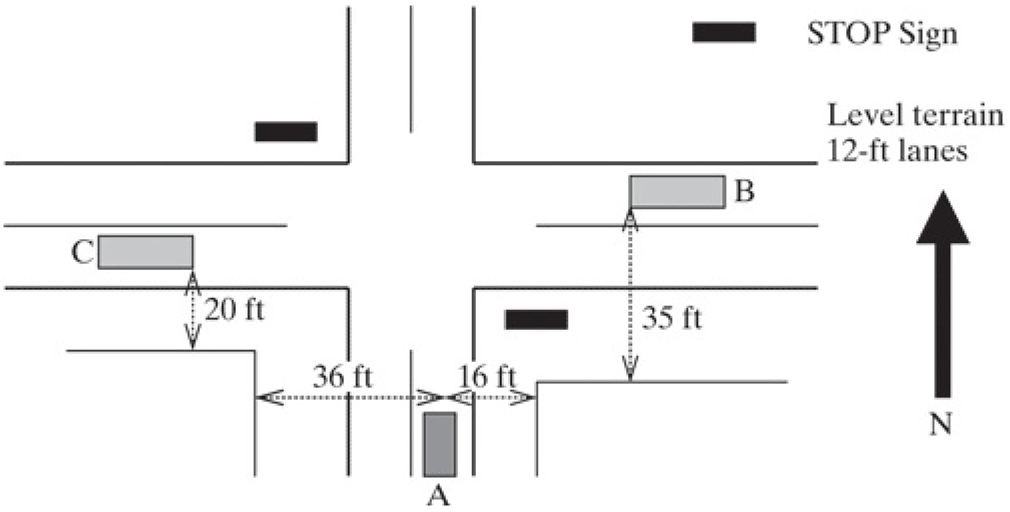


Figure 15.4: Full Alternative Text

Using [Equation 15-4], the position of the stopped vehicle on the minor approach can be determined:

$$daSTOP=18+dcldaSTOP\ (\text{from left})=18+6=24ftdaSTOP\ (\text{from right})=18+$$

The minimum sight distance requirement for Vehicles B and C is determined from [Equation 15-5], using a time gap (tg) of 7.5 seconds for typical conditions:

$$db,\ c\ MIN=1.47\ Smaj\ tg=1.47\times 40\times 7.5=441\ \text{ft}$$

Now the actual distance of Vehicles B and C from the collision point when visibility is established is determined using [Equation 15-1]:

$$db,\ c\ ACT=a\ dada$$
$$-bdbACT=16\times 3636-35=576\ \text{ft}>441\ \text{ftdcACT}=36\times 2424-20=216\ \text{ft}<441$$

In the case of major street Vehicle C approaching from the left, there is not sufficient sight distance to meet the criteria. The sight distance for Vehicle B approaching from the right meets the criteria. Note that it is possible for db,cACT to be negative. This would indicate that there was no sight obstruction from the direction analyzed.

Where the STOP-sign sight-distance criterion is not met, it is recommended that speed limits be reduced (with signs posted) to a level that would allow appropriate sight distance to the minor street. In the sample problem, the speed would have to be reduced to a point where dcMIN was 216 ft or less. Using Equation 15-5:

$$d_{cMIN} = 1.47\, S_{maj}\, t_g \quad 216 = 1.47 \times S_{maj} \times 7.5 \quad S_{maj} = \frac{216}{1.47 \times 7.5} = 19.6 \text{ mi/h}$$

The results of Sample Problem 15-2 indicate that a very low speed limit of approximately 20 mi/h would be necessary to have the intersection comply with sight distance requirements. Without extreme levels of enforcement, it is not likely that such a speed limit would be well observed. Removal or cutting back of sight obstructions is also a potential solution, but this is often impossible in developed areas, where buildings are the principal obstructions, as is the case here.

In the strictest sense, logic indicates that this intersection would have to be signalized, as sight distances cannot even accommodate STOP control. In practice, it is recognized that in such restricted cases, drivers will most often simply creep closer to the intersection line to obtain an adequate sight line. The assumed position of Vehicle A, in effect, would be compromised as drivers creep forward for visibility. Thus, it is likely that the STOP sign would be placed anyway, but any crashes at the intersection would force a reconsideration.

15.2.2 YIELD Control

A YIELD sign assigns right-of-way to the major uncontrolled street. It requires vehicles on the minor approach(es) to slow and yield the right-of-way to any major street vehicle approaching at a distance and speed that would present an impending hazard to the minor street vehicle if it entered the major street. Most state laws require that drivers on YIELD-controlled approaches slow to 8 to 10 mi/h before entering the major street.

Advice for YIELD control in the MUTCD is hardly definitive, and is given only under the heading of “options,” except for one relatively new mandatory usage. The advise is summarized in Table 15.3.

Table 15.3: Guidance and

Options for YIELD Signs

A YIELD sign *shall* be used to assign right-of-way at the entrance to a roundabout. YIELD signs at roundabouts *shall* be used to control the approach roadways and *shall not* be used to control the circulatory roadway.
YIELD signs may be installed:
A. On approaches to a through street or highway where conditions are such that a stop is not always required. B. At the second crossroad of a divided highway, where the median width at the intersection is 30 ft or greater. In this case, a STOP or YIELD sign may be installed at the entrance to the first roadway, and a YIELD sign may be installed at the entrance to the second roadway. C. On a channelized turn lane that is separated from the adjacent travel lane by an island, even if the adjacent lanes at the intersection are controlled by a highway traffic control signal or by a STOP sign. D. At an intersection where a special problem exists and where engineering judgment indicates the problem to be susceptible to correction by the use of YIELD signs. E. Facing the entering roadway for a merge-type movement if engineering judgment indicates that the control is needed because acceleration geometry and/or sign distance is not sufficient for merging traffic operation.

(Source: *Manual of Uniform Traffic Control Devices*, Federal Highway Administration, Washington, D.C., 2009, with revisions through 2012, pg 53, available at www.fhwa.com.)

Table 15.3: Full Alternative Text

The principal uses of the YIELD sign emanate from their mandatory use at roundabouts and Options B, C, and E. Option B is a common application where wide medians exist, and allow at least one crossing vehicle to be protected if a second pause is required to assess safety. Option C allows use of the YIELD sign to control channelized right turns at signalized and unsignalized intersections, and Option E allows their use at on-ramp or other merge situations. The latter is a frequent use in which adequate sight distance or geometry (i.e., inadequate length of the acceleration lane) make an uncontrolled merge potentially unsafe.

There has been some controversy over the use of YIELD signs at normal crossings. Because YIELD signs require drivers to slow down, the sight triangle may be analyzed using the legal reduced approach speed. In 2000, the Millennium Edition of the MUTCD required that sight distance sufficient for safety at the normal approach speed be present whenever a YIELD sign was used. This greatly discouraged their use at regular intersections. This prescription has been removed in the current edition.

With this latest change, the sight triangle for YIELD signs would be analyzed in the same way as for an uncontrolled intersection. The approach speed of controlled vehicles, however, would be that required by state law (8–10 mi/h).

15.2.3 Multiway Stop Control

Multiway STOP control, where all intersection approaches are controlled using STOP signs, remains a controversial form of control. Some agencies find it attractive, primarily as a safety measure. Others believe that the confusion that drivers often exhibit under this form of control negates any of the benefits it might provide.

MUTCD guidance and options with regard to multiway STOP control reflect this ongoing controversy. Multiway STOP control is most often used where there are significant conflicts between vehicles and pedestrians and/or bicyclists in all directions, and where vehicular demands on the intersecting roadways are approximately equal. Table 15.4 shows the guidance and options for multiway STOP control.

Table 15.4: Guidance and Options for Multiway STOP Signs

The following criteria should be considered in the engineering study for a multiway STOP sign:
A. Where traffic control signals are justified, the multiway STOP is an interim measure that can be installed quickly to control traffic while arrangements are being made for the installation of the traffic control signal.
B. Five or more reported crashes in a 12-month period that are susceptible to correction by a multiway STOP installation. Such crashes include right- and left-turn collisions as well as right-angle collisions.
C. Minimum volumes:
1. The vehicular volume entering the intersection from the major street approaches (total of both approaches) averages at least 300 veh/h for any 8 hours of an average day, and
2. The combined vehicular, pedestrian, and bicycle volume entering the intersection from the minor street approaches (total of both approaches) averages at least 200 units/h for the same 8 hours, with an average delay to minor-street vehicular traffic of at least 30 s/veh during the highest hour, but
3. If the 85th percentile approach speed of the major highway exceeds 40 mi/h, the minimum vehicular volume warrants are 70% of the above values.
D. Where no single criterion is satisfied, but where criteria B, C1, and C2 are all satisfied to 80% of the minimum values. Criterion C3 is excluded from this condition.
Other criteria that may be considered in an engineering study include:
A. The need to control left-turn conflicts;
B. The need to control pedestrian/vehicle conflicts near locations that generate high pedestrian volumes;
C. Locations where a road user, after stopping, cannot see conflicting traffic and is not able to negotiate the intersection unless conflicting cross traffic is also required to stop; and
D. An intersection of two residential collector (through) streets of similar design and operating characteristics where multiway stop control would improve traffic operational characteristics of the intersection.

(Source: *Manual on Uniform Traffic Control Devices*, Federal Highway Administration, Washington, D.C., 2009, with revisions through 2012, pg 52, available at www.fhwa.com.)

Table 15.4: Full Alternative Text

It should be noted that such control is generally implemented as a safety measure, as operations at such locations are often not very efficient. The current edition of the *Highway Capacity Manual*, [3] includes a methodology for analysis of the capacity and level of service provided by multiway STOP control.

15.3 Level III Control: Traffic Control Signals

The ultimate form of intersection control is the traffic signal. Because it alternately assigns right-of-way to specific movements, it can substantially reduce the number and nature of intersection conflicts as no other form of control can.

If drivers obey the signal, then driver judgment is not needed to avoid some of the most critical intersection conflicts. Imposition of traffic signal control does not, however, remove all conflicts from the realm of driver judgment. At two-phase signals, where all left-turns are made against an opposing vehicular flow, drivers must still evaluate and select gaps in opposing traffic through which to safely turn. At virtually all signals, some pedestrian-vehicle and bicycle-vehicle conflicts remain between legal movements, and driver vigilance and judgment are still required to avoid crashes. Nevertheless, drivers at signalized intersections do not have to negotiate the critical conflicts between crossing vehicle streams, and where exclusive left-turn phases are provided, critical conflicts between left turns and opposing through vehicles are also eliminated through signal control. This chapter deals with the issue of whether or not signal control is warranted or needed. Given that it is needed, Chapters 18 to 20 deal with the specifics of intersection signal design and timing.

While warrants and other criteria for STOP and YIELD signs are somewhat general in the MUTCD, warrants for signals are quite detailed. The cost involved in installation of traffic signals (e.g., power supply, signal controller, detectors, signal heads, and support structures, and other items) is considerably higher than for STOP or YIELD signs and can run into the hundreds of thousands of dollars, even millions, for complex intersections. Because of this, and because traffic signals introduce a fixed source of delay into the system, it is important that they not be overused; they should be installed only where no other solution or form of control would be effective in assuring safety and efficiency at the intersection.

15.3.1 Advantages of Traffic Signal Control

The MUTCD provides the following statements of the benefits of traffic signal control:

1. They provide for the orderly movement of traffic.
2. They increase the traffic-handling capacity of the intersection if proper physical layouts and control measures are used and if the signal timing is reviewed and updated on a regular basis (every 2 years) to ensure that it satisfies the current traffic demands.
3. They reduce the frequency and severity of certain types of crashes, especially right-angle collisions.
4. They are coordinated to provide for continuous or nearly continuous movement at a definite speed along a given route under favorable conditions.
5. They are used to interrupt heavy traffic at intervals to permit other traffic, vehicular or pedestrian, to cross.

These specific advantages address the primary reasons why a traffic signal would be installed: to increase capacity (thereby improving level of service), to improve safety, and to provide for orderly movement through a complex situation. Coordination of signals provides other benefits, but not all signals are necessarily coordinated.

15.3.2 Disadvantages of Traffic Signal Control

Capacity is increased by a well-designed signal at a well-designed intersection. Poor design of either the signalization or the geometry of the intersection can significantly reduce the benefits achieved or negate them entirely. Improperly designed traffic signals, or the placement of a signal where it is not justified, can lead to some of the following disadvantages, as detailed in the MUTCD:

1. Excessive delay.
2. Excessive disobedience of the signal indications.
3. Increased use of less adequate routes as road users attempt to avoid the traffic control signal.
4. Significant increases in the frequency of collisions (especially rear-end collisions).

The last item is of some interest. Even when they are properly installed and well-designed, traffic signal controls can lead to increases in rear-end accidents because of the cyclical stopping of traffic. Where safety is concerned, signals can reduce the number of right-angle, turning, and pedestrian/bicycle accidents; they might cause an increase in rear-end collisions (which tend to be less severe); they will have almost no impact on head-on or sideswipe accidents, or on single-vehicle accidents involving fixed objects.

Excessive delay can result from an improperly installed signal, but it can also occur if the signal timing is inappropriate. In general, excessive delay results from cycle lengths that are either too long or too short for the existing demands at the intersection. Further, drivers will tend to assume that a signal is broken if they experience an excessive wait, particularly when there is little or no demand occurring on the cross street.

15.3.3 Warrants for Traffic Signals

The MUTCD specifies nine different warrants that justify the installation of a traffic signal. The ninth is the most recent, covering the installation of a signal in coordination with a railroad crossing. Satisfying one or more of the criteria for signalization does *not* require or justify the installation of a signal. The manual *requires*, however, that a comprehensive engineering study be conducted to determine whether or not installation of a signal is justified. The study *must* include applicable factors reflected in the specified criteria, but could extend to other factors as well. On the other hand, traffic signal control *should not* be implemented if none of the criteria are met. The criteria provided, therefore, still require the exercise of engineering judgment. In the final analysis, if engineering studies and/or judgment indicate that signal installation *will not* improve the overall safety or operational efficiency at a candidate location, it should not be installed.

While offered only under the heading of an option, the MUTCD suggests that the following data be included in an engineering study of the need for a traffic signal [2009 MUTCD, page 437]:

1. “The number of vehicles entering the intersection from each approach during 12 hours of an average day. It is desirable that the hours selected contain the greatest percentage of the 24-hour traffic volume.”

2. “Vehicular volumes for each traffic movement, from each approach, classified by vehicle type (heavy trucks, passenger cars and light trucks, public-transit vehicles, and in some locations, bicycles), during each 15 minute period of the 2 hours in the morning and 2 hours in the afternoon during which total traffic entering the intersection is greatest.”

3. “Pedestrian volume counts on each crosswalk during the same periods as the vehicular counts in Item 2 above and during hours of highest pedestrian volume. Where young, elderly, and/or persons with physical or visual disabilities need special consideration, the pedestrians and their crossing times may be classified by general observation.”

4. “Information about nearby facilities and activity centers that serve the young, elderly, and/or persons with disabilities, including requests from persons with disabilities for accessible parking improvements at the location under study. These persons might not be adequately reflected in the pedestrian volume count if the absence of a signal restrains their mobility.”

5. “The posted or statutory speed limit or the 85th percentile speed on the uncontrolled approaches to the location.”

6. “A condition diagram showing details of the physical layout,

including such features as intersection geometrics, channelization, grades, sight distance restrictions, transit stops and routes, parking conditions, pavement markings, roadway lighting, driveways, nearby railroad crossings, distance to nearest traffic control signals, utility poles and fixtures, and adjacent land use.”

7. “A collision diagram showing crash experience by type, location, direction of movement, severity, weather, time of day, date, and day of week for at least one year.”

MUTCD also recommends collection of stopped-time delay data and queuing information at some locations where these are thought to be problems.

This data will allow the engineer to fully evaluate whether or not the intersection satisfies the requirements of one or more of the following warrants:

- Warrant 1: Eight-Hour Vehicular Volume
- Warrant 2: Four-Hour Vehicular Volume
- Warrant 3: Peak Hour
- Warrant 4: Pedestrian Volume
- Warrant 5: School Crossing
- Warrant 6: Coordinated Signal System
- Warrant 7: Crash Experience
- Warrant 8: Roadway Network
- Warrant 9: Intersection Near a Highway-Rail Crossing

It also provides a sufficient base for the exercise of engineering judgment in determining whether a traffic signal should be installed at the study location. Each of these warrants is presented and discussed in the sections that follow.

In most cases, an engineering study will include data from an existing location. In some cases, however, consideration of signalization relates to a future situation or design. In such cases, forecast demand volumes may be used to compare with some criteria in the warrants.

Warrant 1: Eight-Hour Vehicular Volume

The 8-hour vehicular volume warrant represents a merging of three different warrants in the pre-2000 MUTCD (old Warrants 1, 2, and 8). It addresses the need for signalization for conditions that exist over extended periods of the day (a minimum of 8 hours). Two of the most fundamental reasons for signalization are addressed:

- Heavy volumes on conflicting cross-movements that make it impractical for drivers to select gaps in an uninterrupted traffic stream through which to safely pass. This requirement is often referred to as the “minimum vehicular volume” condition (Condition A).

- Vehicular volumes on the major street are so heavy that no minor-street vehicle can safely pass through the major-street traffic stream without the aid of signals. This requirement is often referred to as the “interruption of continuous traffic” condition (Condition B).

Details of this warrant are shown in Table 15.5. The warrant is met when:

- Either Condition A or Condition B is met to the 100% level.

- Either Condition A or Condition B is met to the 70% level, where the intersection is located in an isolated community of population 10,000 or less, or where the major-street approach speed is 40 mi/h or higher.

- Both Conditions A and B are met to the 80% level.

Table 15.5: Warrant 1: Eight-Hour Vehicular Volume

Condition A—Minimum Vehicular Volume									
Number of lanes for moving traffic on each approach		Vehicles per hour on major street (total of both approaches)				Vehicles per hour on higher-volume minor-street approach (one direction only)			
Major Street	Minor Street	100%[a]	80%[b]	70%[c]	56%[d]	100%[a]	80%[b]	70%[c]	56%[d]
1	1	500	400	350	280	150	120	105	84
2 or more	1	600	480	420	336	150	120	105	84
2 or more	2 or more	600	480	420	336	200	160	140	112
1	2 or more	500	400	350	280	200	160	140	112
Condition B—Interruption of Continuous Traffic									
Number of lanes for moving traffic on each approach		Vehicles per hour on major street (total of both approaches)				Vehicles per hour on higher-volume minor-street approach (one direction only)			
Major Street	Minor Street	100%[a]	80%[b]	70%[c]	56%[d]	100%[a]	80%[b]	70%[c]	56%[d]
1	1	750	600	525	420	75	60	53	42
2 or more	1	900	720	630	504	75	60	53	42
2 or more	2 or more	900	720	630	504	100	80	70	56
1	2 or more	750	600	525	420	100	80	70	56

[a]Basic minimum hourly volume.

[b]Used for combination of Conditions A and B after adequate trial of other remedial measures.

[c]May be used when the major-street speed exceeds 40 mi/h or in an isolated community with a population of less than 10,000.

[d]May be used for combination of Conditions A and B after adequate trial of other remedial measures when the major-street speed exceeds 40 mi/h or in an isolated community with a population of less than 10,000.

(Source: *Manual on Uniform Traffic Control Devices*, Federal Highway Administration, Washington, D.C., 2009, revised through 2012, pg 436.)

Table 15.5: Full Alternative Text

Note that in applying these warrants, the major-street volume criteria are related to the total volume in both directions, while the minor-street volume criteria are applied to the highest volume in one direction. The volume criteria in Table 15.5 must be met for a minimum of 8 hours on a typical day. The 8 hours do not have to be consecutive, and often involve 4 hours around the morning peak and 4 hours around the evening peak. Major- and minor-street volumes must, however, be for the same 8 hours.

Either of the intersecting streets may be treated as the “major” approach, but the designation must be consistent for a given application. If the designation of the “major” street is not obvious, a warrant analysis can be conducted considering each as the “major” street in turn. While the designation of the major street may not be changed within any one analysis, the direction of peak one-way volume for the minor street need not be consistent.

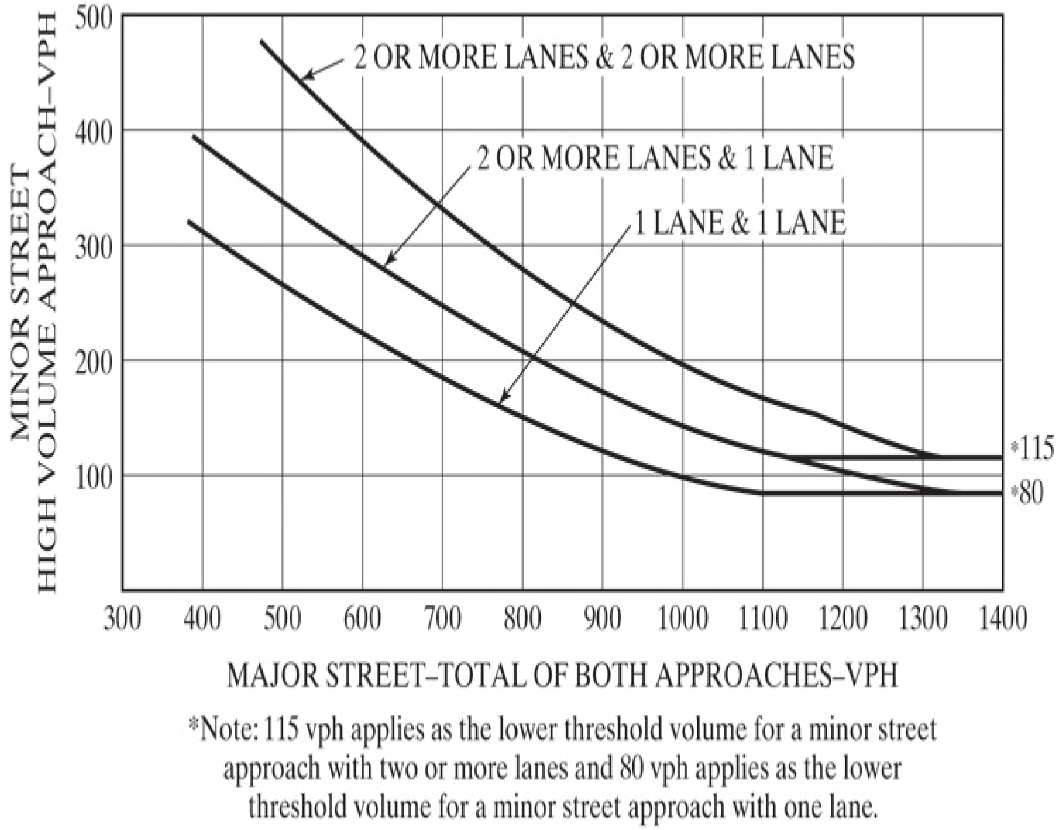
The 70% reduction allowed for rural communities of population 10,000 or less reflects the fact that drivers in small communities have little experience in driving under congested situations. They will require the guidance of traffic signal control at volume levels lower than those for drivers more used to driving in congested situations. The same reduction applies where the major-street speed limit is 40 mi/h or greater. As gap selection is more difficult through a higher-speed major-street flow, signals are justified at lower volumes.

The various elements of the 8-hour vehicular volume warrant are historically the oldest of the warrants, having been initially formulated and disseminated in the 1930s.

Warrant 2: Four-Hour Vehicular Volume

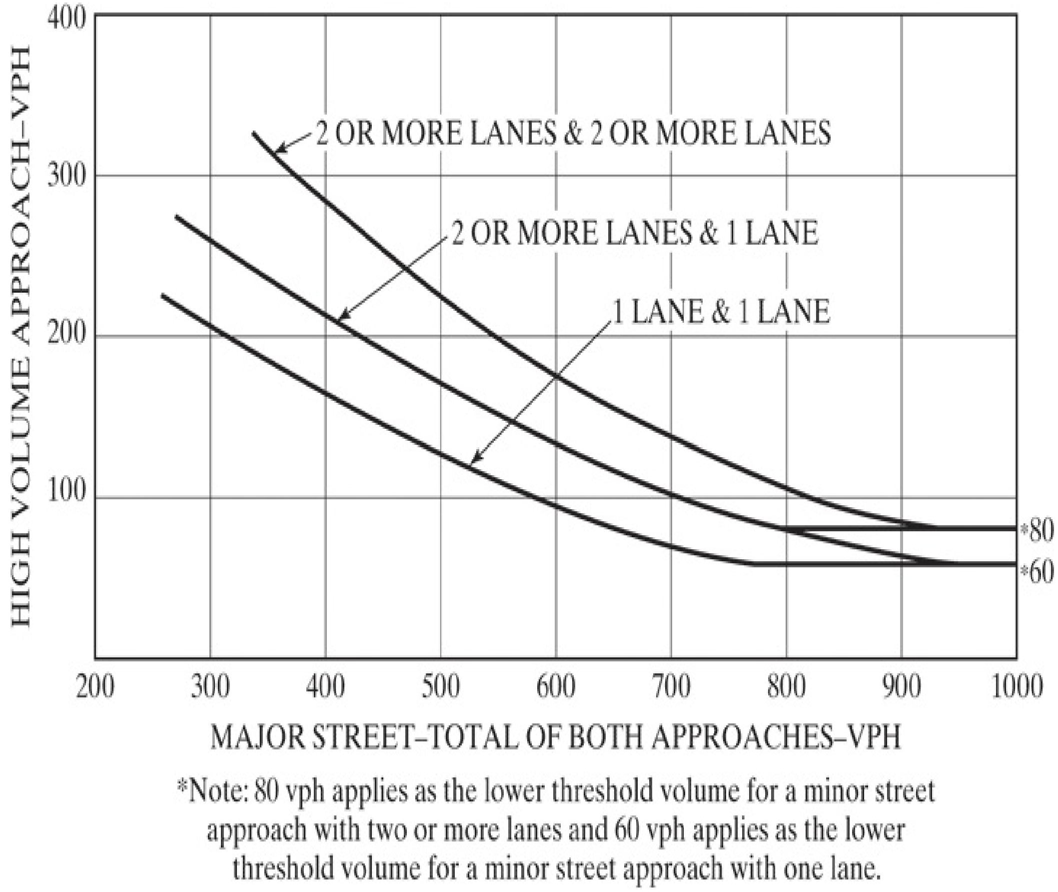
The 4-hour vehicular volume warrant was introduced in the 1970s to assist in the evaluation of situations where volume levels requiring signal control might exist for periods shorter than eight hours. Prior to the MUTCD Millennium Edition, this was old Warrant 9. Figure 15.5 shows the warrant, which is in the form of a continuous graph. Because this warrant is expressed as a continuous relationship between major and minor street volumes, it addresses a wide variety of conditions. Indeed, Conditions A and B of the 8-hour warrant represent two points in such a continuum for each configuration, but the older 8-hour warrant did not investigate or create criteria for the full range of potential conditions.

Figure 15.5: Warrant 2: Four-Hour Vehicular Volume



(a) Normal Conditions

15.3-6 Full Alternative Text



(b) Criteria for Small Communities (pop < 10,000) or High Major Street Approach Speed (≥ 40 mi/h)

(Sources: Federal Highway Administration, U.S. Department of Transportation, *Manual on Uniform Traffic Control Devices*, Washington DC, 2009, revised through 2012, Figures 4C-1 and 4C-2, pg 400)

15.3-6 Full Alternative Text

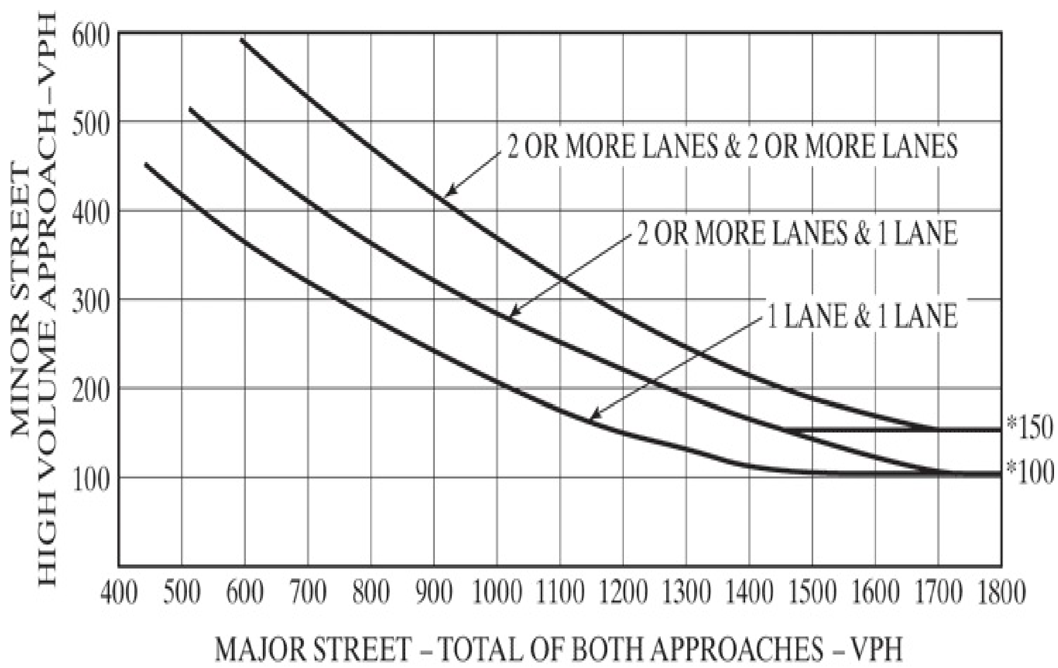
Figure 15.5(a) is the warrant for normal conditions, while Figure 15.5(b) reflects the 70% reduction applied to isolated small communities (with population less than 10,000), or where the major-street speed limit is 40 mi/h or higher. Because the 4-hour warrant represents a continuous set of conditions, there is no need to include an 80% reduction for two discrete conditions within the relationship.

To test the warrant, the two-way major-street volume is plotted against the highest one-way volume on the minor street for each hour of the study period. To meet the warrant, at least four hours must plot *above* the appropriate decision curve. The three curves represent intersections of (1) two streets with one lane in each direction, (2) one street with one lane in each direction with another having two or more lanes in each direction, and (3) two streets with more than one lane in each direction. In Case (2), the distinction between which intersecting street has one lane in each direction (major or minor) is no longer relevant, except for the footnotes.

Warrant 3: Peak Hour

Warrant 3 addresses two critical situations that might exist for only 1 hour of a typical day. The first is a volume condition, similar in form to Warrant 2, and shown in Figure 15.6 (old Warrant 11). The second is a delay warrant (old Warrant 10). If either condition is satisfied, the peak-hour warrant is met.

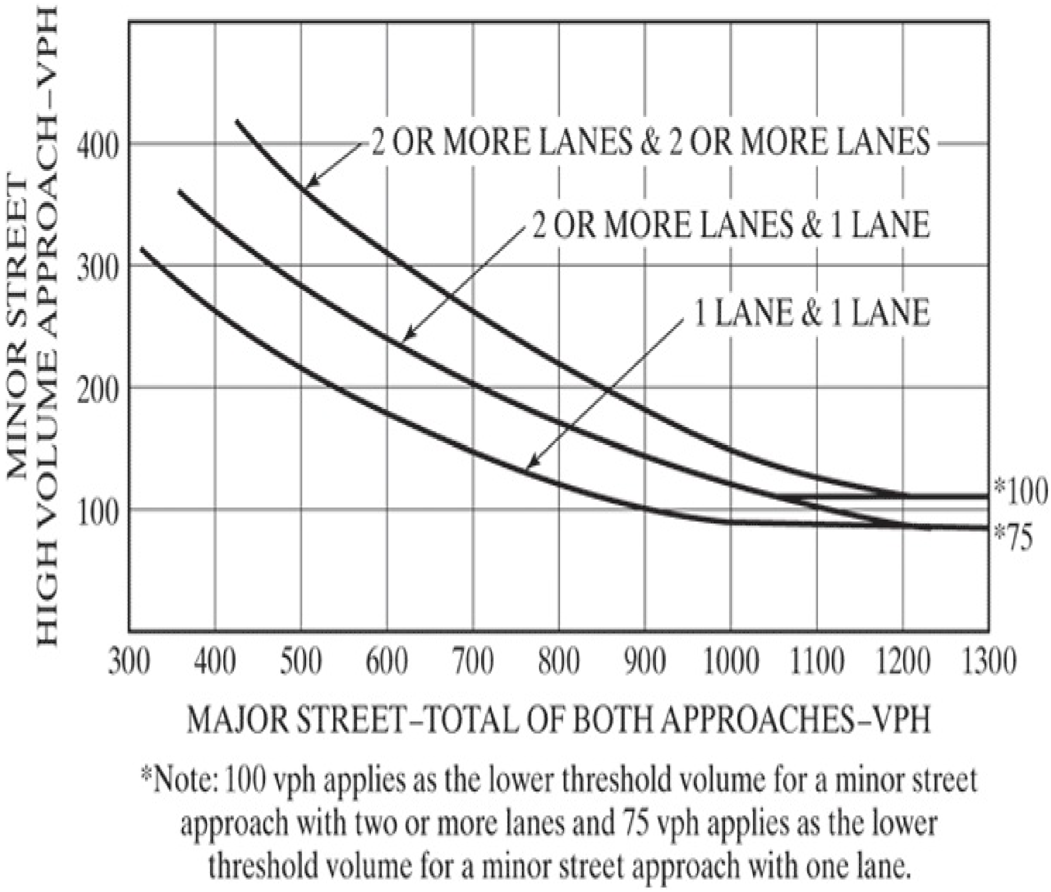
Figure 15.6: Warrant 3A: Peak-Hour Volume



*Note: 150 vph applies as the lower threshold volume for a minor street approach with two or more lanes and 100 vph applies as the lower threshold volume for a minor street approach with one lane.

(a) Normal Conditions

15.3-6 Full Alternative Text



(b) Criteria for Small Communities (Pop < 10,000) or High Major-Street Approach Speed (≥ 40 mi/h)

(Source: Federal Highway Administration, U.S. Department of Transportation, *Manual on Uniform Traffic Control Devices*, Washington DC, 2000, revised through 2012, Figures 4C-3 and 4C-4, pg 401)

15.3-6 Full Alternative Text

The volume portion of the warrant is implemented in the same manner as the 4-hour warrant. For each hour of the study, the two-way major street volume is plotted against the high single-direction volume on the minor street. For the Peak-Hour Volume Warrant, however, only one hour must plot above the appropriate decision line to meet the criteria. Criteria are given for normal conditions in Figure 15.6(a), and the 70% criteria for small isolated communities and high major-street speeds are shown in

Figure 15.6(b). The Peak-Hour Delay Warrant is summarized in Table 15.6.

Table 15.6: Warrant 3B: Peak-Hour Delay

The need for a traffic control signal shall be considered if an engineering study finds that all three of the following conditions exist for the same 1 hour (any four consecutive 15-minute periods) of an average day:

1. The total stopped-time delay experienced by traffic on one minor street approach (one direction only) controlled by a STOP sign equals or exceeds: 4 veh-hours for a one-lane approach; or 5 veh-hours for a two lane approach, and
2. The volume on the same minor street approach (one direction only) equals or exceeds 100 veh/h for one moving lane of traffic; or 150 veh/h for two moving lanes, and
3. The total entering volume serviced during the hour equals or exceeds 650 veh/h for intersections with three approaches, or 800 veh/h for intersections with four or more approaches.

(Source: *Manual of Uniform Traffic Control Devices*, Draft, Federal Highway Administration, Washington, D.C., 2009, revised through 2012, pg 439.)

Table 15.6: Full Alternative Text

It is important to recognize that the delay portion of Warrant 3 applies only to cases in which STOP control is already in effect for the minor street. Thus, delay during the peak hour is not a criterion that allows going from no control or YIELD control to signalization directly.

The MUTCD also emphasizes that the Peak-Hour Warrant should be applied only in special cases, such as office complexes, manufacturing plants, industrial complexes, or high-occupancy vehicle facilities that attract or discharge large numbers of vehicles over a short time.

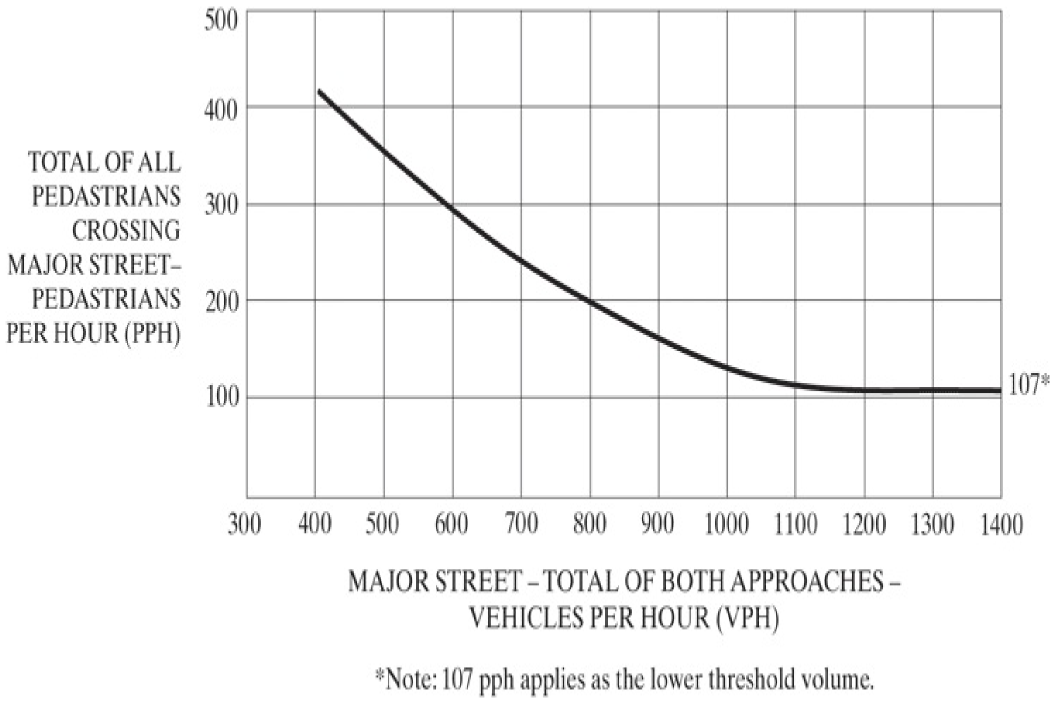
The MUTCD also recommends that when this is the *only* warrant that justifies the installation of a signal, it *should* normally be a traffic-actuated signal.

Warrant 4: Pedestrians

The Pedestrian Warrant addresses situations in which the need for signalization is the frequency of vehicle-pedestrian conflicts and the inability of pedestrians to avoid such conflicts due to the volume of traffic present. Signals may be placed under this warrant at mid-block locations, as well as at intersections.

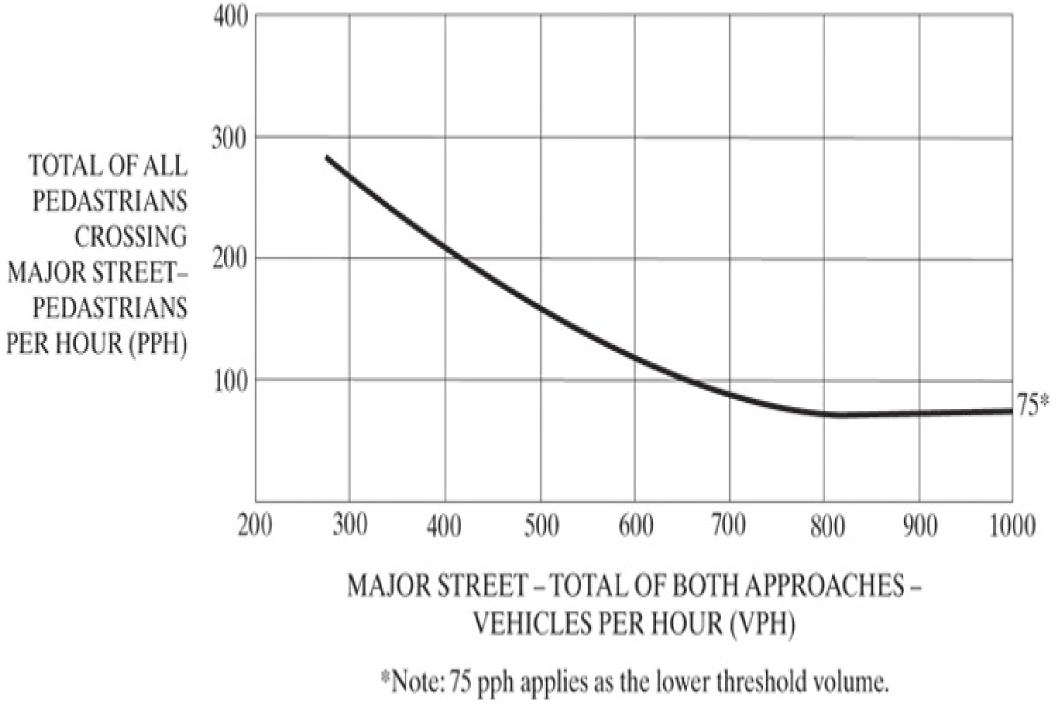
This warrant is met when any 4 hourly plots of total pedestrians crossing the major street and the total major street vehicular traffic falls over the line in Figure 15.7(a), or when any one similar hourly plot falls above the line in Figure 15.8(a). If the location is in a built-up area of a small community (pop<10,000) or where the posted or statutory speed limit, or the 85th percentile approach speed exceeds 35 mi/h, Figures 15.7(b) and, 15-8(b) may be used.

Figure 15.7: Four-Hour Pedestrian Warrant



(a) Normal Criteria

15.3-7 Full Alternative Text

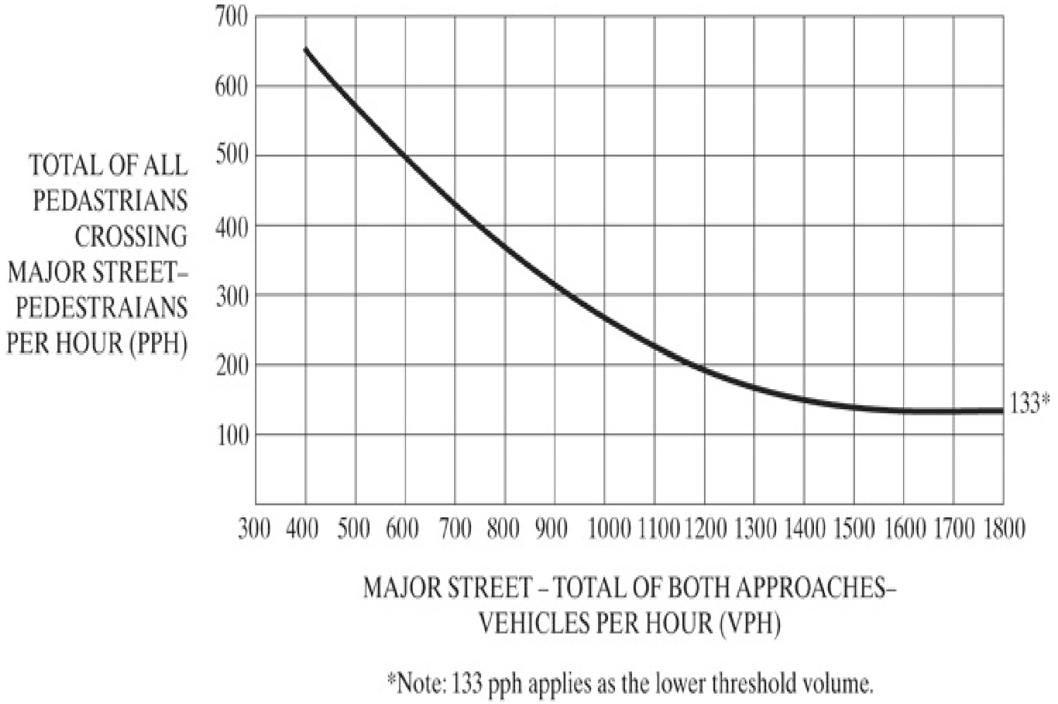


(b) Criteria for Small Communities (Pop < 10,000) or High Major Street Approach Speed (> 35 mi/h)

(Source: *Manual of Uniform Traffic Control Devices*, Federal Highway Administration, Washington, D.C., 2009, revised through 2012, Figures 4C-5 and 4C-6, pg 443.)

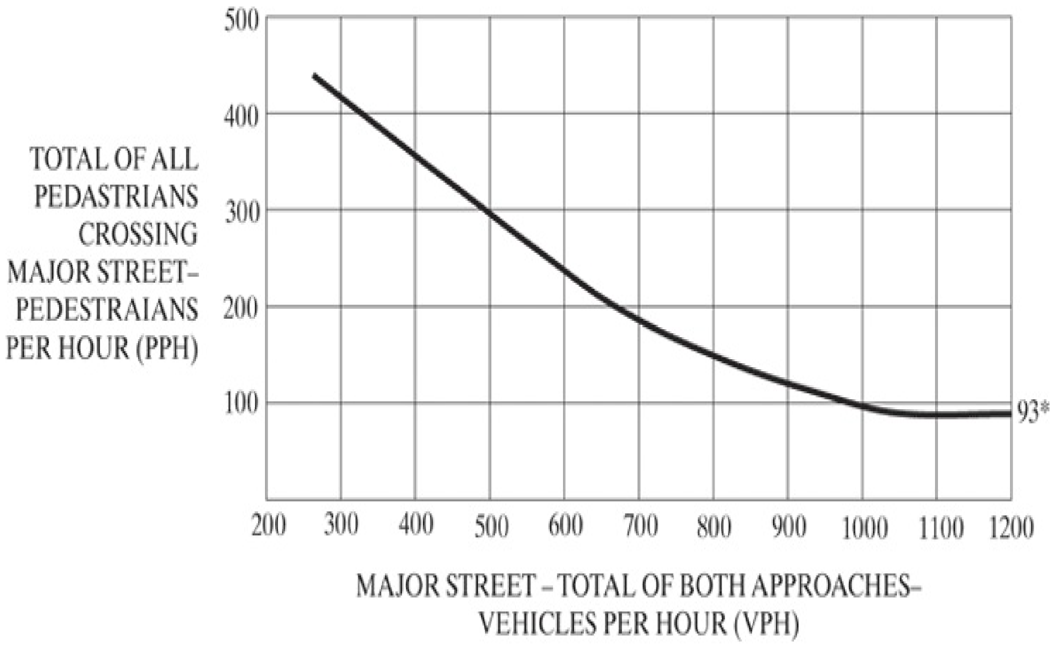
15.3-7 Full Alternative Text

Figure 15.8: Peak-Hour Pedestrian Warrant



(a) Normal Criteria

15.3-7 Full Alternative Text



(b) Criteria for Small Communities (Pop < 10,000) or High Major Street Approach Speed (> 35 mi/h)

(Source: *Manual of Uniform Traffic Control Devices*, Federal Highway Administration, Washington, D.C., 2009, revised through 2012, Figures 4C-7 and 4C-8, pg 447.)

15.3-7 Full Alternative Text

The figures address cases in which a steadier pedestrian flow over 4 hours requires signal control and the case in which a single peak hour has pedestrian- vehicle conflicts that must be signal controlled. The (b) figures apply the same 70% reduction in criteria that is used in conjunction with vehicular volume criteria in Warrants 1, 2, and 3. In this case, however, the threshold speed for applying the reduction is only 35 mi/h.

If a traffic signal is justified at an intersection by this warrant only, it will usually be at least a semiactuated signal (a full-actuated signal is also a possibility at an isolated intersection) with pedestrian push-buttons and signal heads for pedestrians crossing the major street. If it is within a coordinated signal system, it would also be coordinated into the system. If such a signal is located in mid-block, it will always be pedestrian-actuated, and parking and other sight restrictions should be eliminated within 20 ft of both sides of the crosswalk. Standard reinforcing markings and signs should also be provided.

If the intersection meets this warrant but also meets other vehicular warrants, any type of signal could be installed as appropriate to other conditions. Pedestrian signal heads would be required for major-street crossings. Pedestrian push-buttons would be installed unless the vehicular signal timing safely accommodates pedestrians in every signal cycle.

A signal *shall not* be implemented under this warrant if there is another signal within 300 ft of the location. Placement of a signal so close to another would only be permitted if did not disrupt progressive flow on the major street.

Pedestrian volume criteria may be reduced by as much as 50% if the 15th-percentile crossing speed is less than 3.5 mi/h, as might be the case where elderly, very young, or disabled pedestrians are present in significant percentages.

Warrant 5: School Crossing

This warrant is similar to the pedestrian warrant but is limited to application at designated school crossing locations, either at intersections or at mid-block locations. The warrant requires the study of available gaps to see whether they are “acceptable” for children to cross through. An acceptable gap would include the crossing time, buffer time, and an allowance for groups of children to start crossing the street. The frequency of acceptable gaps should be no less than one for each minute during which school children are crossing. The minimum number of children crossing the major street is 20 during the highest crossing hour.

Traffic signals are rarely implemented under this warrant. Children do not usually observe and obey signals regularly, particularly if they are very young. Thus, traffic signals would have to be augmented by crossing guards in most cases. Except in unusual circumstances involving a very heavily traveled major street, the crossing guard, perhaps augmented with STOP signs, would suffice under most circumstances without signalization. Where extremely high volumes of school children cross a very wide and heavily traveled major street, overpasses or underpasses should be provided with barriers preventing entry onto the street.

Warrant 6: Coordinated Signal System

Chapter 21 of this text addresses signal coordination and progression systems for arterials and networks. Maintaining platoons of vehicles moving through a "green wave" as they progress along an arterial is critical to the efficiency of coordinated signal systems. If the distance between two adjacent coordinated signals is too large, platoons begin to dissipate and the positive impact of the progression is sharply reduced. In such cases, the traffic engineer may place a signal at an intermediate intersection where it would not otherwise be warranted to reinforce the coordination scheme and to help maintain platoon coherence. The application of this warrant, shown in Table 15.7, should not result in signal spacing of less than 1,000 ft. Such signals, when placed, are often referred to as "spacer signals."

Table 15.7: Warrant 6: Coordinated Signal System

The need for a traffic control signal shall be considered if an engineering study finds that one of the following criteria is met:
1. On a one-way street or a street that has traffic predominantly in one direction, the adjacent traffic control signals are so far apart that they do not provide the necessary degree of vehicular platooning. 2. On a two-way street, adjacent traffic control signals do not provide the necessary degree of platooning and the proposed and adjacent traffic control signals will collectively provide a progressive operation.

(Source: Federal Highway Administration, US Dept. of Transportation, *Manual on Uniform Traffic Control Devices*, Washington DC, 2009, revised through 2012, pg 445)

Table 15.7: Full Alternative Text

The two criteria are similar, but not exactly the same. Inserting a signal in a one-way progression is always possible without damaging the progression. On a two-way street, it is not always possible to place a signal that will maintain the progression in both directions acceptably. This issue is discussed in greater detail in Chapter 21.

Warrant 7: Crash Experience

The Crash Experience Warrant addresses cases in which a traffic control signal would be installed to alleviate an observed high-accident occurrence at the intersection. The criteria are summarized in Table 15.8.

Table 15.8: Warrant 7: Crash Experience

The need for a traffic control signal shall be considered if an engineering study finds that all of the following criteria are met:
1. Adequate trial of alternatives with satisfactory observance and enforcement has failed to reduce the crash frequency, and 2. Five or more reported crashes of types susceptible to correction by a traffic control signal have occurred within a 12-month period, each involving a personal injury or property damage apparently exceeding the applicable requirements for a reportable crash, and 3. For each of any 8 hours of the day, vehicles per hour (vph) given in both of the 80% columns of Condition A (in Warrant 1) or the vph in both of the 80% columns of Condition B (in Warrant 1) exists on the major-street and the higher-volume minor-street approach, respectively, to the intersection, or the volume of pedestrian traffic is not less than 80% of the requirements specified in the Pedestrian Volume warrant. These major-street and minor-street volumes shall be for the same 8 hours. On the minor street, the higher volume shall not be required to be on the same approach during each of the 8 hours.

(Source: *Manual of Uniform Traffic Control Devices*, Draft, Federal Highway Administration, Washington, D.C., December 2009, revised through 2012, pg 445.)

Table 15.8: Full Alternative Text

The requirement for an adequate trial of alternative methods means that either YIELD or STOP control is already in place and properly enforced. These types of control can also address many of the same accident problems as signalization. Thus, a signal is justified only when these lesser measures have failed to adequately address the situation.

Crashes that are susceptible to correction by signalization include right-angle crashes, crashes involving turning vehicles from the two streets, and crashes between vehicles and pedestrians crossing the street on which the vehicle is traveling. Rear-end accidents are often increased with imposition of traffic signals (or STOP/YIELD signs), as some drivers may be induced to stop quickly or suddenly. Head-on and sideswipe collisions are not addressed by signalization; crashes between vehicles and fixed objects at corners are also not correctable through signalization.

Warrant 8: Roadway Network

This warrant addresses a developing situation, that is, a case in which present volumes would not justify signalization but where new development is expected to generate substantial traffic that would justify signalization. The MUTCD also allows other warrants to be applied based upon properly forecast vehicular and pedestrian volumes.

Large traffic generators, such as regional shopping centers, sports stadiums and arenas, and similar facilities are often built in areas that are sparsely populated and where existing roadways have light traffic. Such projects often require substantial roadway improvements that change the physical layout of the roadway network and create new or substantially enlarged intersections that will require signalization. Generally, the “existing” situation is irrelevant to the situation being assessed. The warrant is described in Table 15.9.

Table 15.9: Warrant 8: Roadway Network

The need for a traffic control signal shall be considered if an engineering study finds that the common intersection of two or more major routes meets one or both of the following criteria:
1. The intersection has a total existing, or immediately projected, entering volume of at least 1,000 veh/h during the peak hour of a typical weekday, and has 5-year projected traffic volumes, based upon an engineering study, that meet one or more of Warrants 1, 2 and 3 during an average weekday, or 2. The intersection has a total existing, or immediately projected, entering volume of at least 1,000 veh/h for each of any 5 hours of a non-normal business day (Saturday or Sunday).
A major route as used in this warrant shall have one or more of the following characteristics: 1. It is part of the street or highway system that serves as the principal roadway network for through traffic flow, or 2. It includes rural or suburban highways outside, entering, or traversing a city, or 3. It appears as a major route on an official plan, such as a major street plan in an urban area traffic and transportation study.

(Source: *Manual of Uniform Traffic Control Devices*, Draft, Federal Highway Administration, Washington, D.C., December 2009, revised through 2012, pg 445.)

Table 15.9: Full Alternative Text

"Immediately projected" generally refers to the traffic expected on day 1 of the opening of new facilities and/or traffic generators that create the need for signalization.

Warrant 9: Intersection Near a Highway-Rail Grade Crossing

This is a new warrant that was added to the 2009 MUTCD. It addresses a unique situation: an intersection that does not meet any other warrant for signalization, but is close enough to a highway-railroad crossing to present a hazard. Table 15.10 shows the detailed criteria for the warrant.

Table 15.10: Warrant 9: Intersection Near a Highway-Rail Grade Crossing

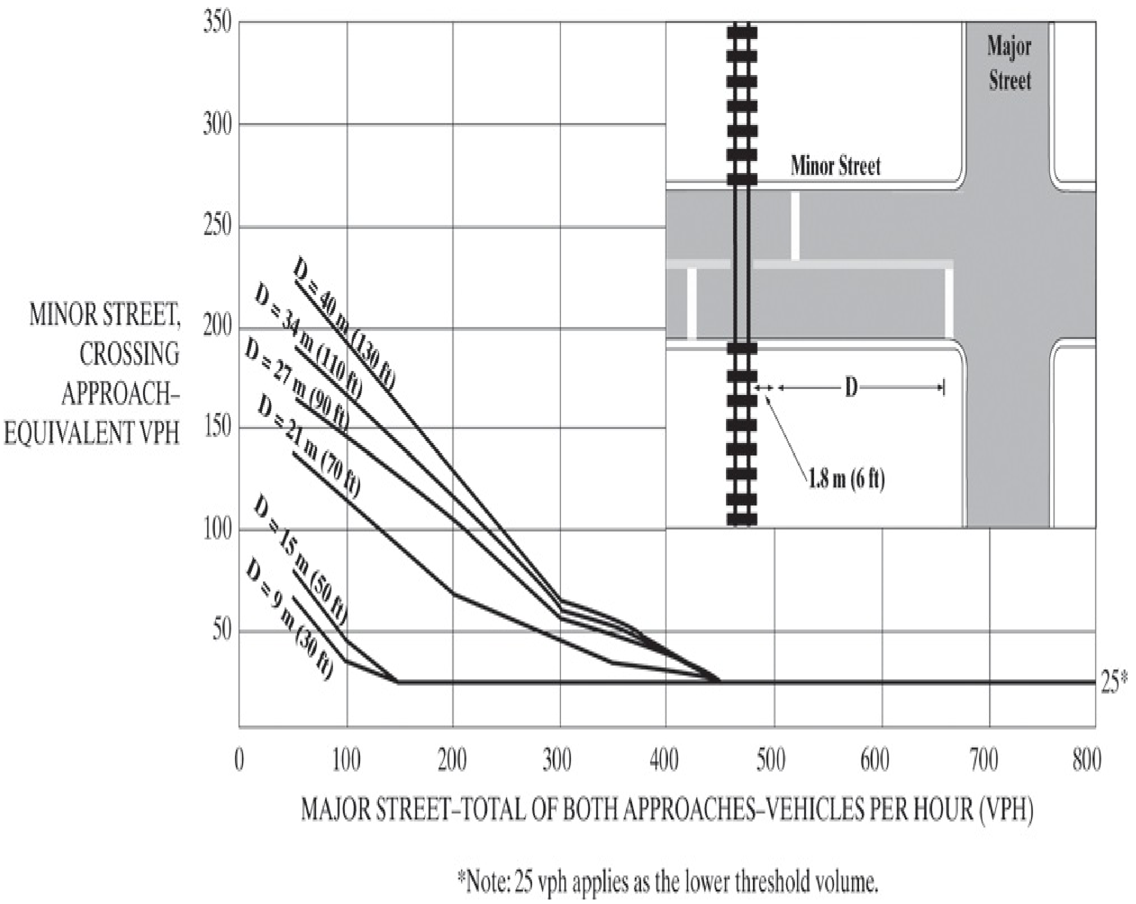
The need for a traffic control signal shall be considered if an engineering study finds that both of the following criteria are met:
1. A highway-rail grade crossing exists on an approach controlled by a STOP or YIELD sign and the center of the track nearest to the intersection is within 140 ft of the stop line on the approach, and 2. During the highest traffic volume hour during which trains use the crossing, the plotted point representing the vehicles per hour on the major street (total of both approaches) and the corresponding vehicles per hour on the minor-street approach that crosses the track (one direction only) falls above the applicable curve in Figure 14.9 or 14.10 for the existing combination of approach lanes over the track and distance D, which is the clear storage distance (between the grade crossing stop line and the near curb line of the major street).

(Source: *Manual of Uniform Traffic Control Devices*, Federal Highway Administration, Washington, D.C., 2009, revised through 2012, pg 446.)

Table 15.10: Full Alternative Text

The criteria are assessed using two figures and several adjustment factors. Figure 15.9 applies when there is only one lane approaching the intersection at the track-crossing location, and Figure 15.10 applies where there are two or more lanes approaching the track-crossing location.

Figure 15.9: Warrant 9: Railroad Crossings for One-Lane Approaches

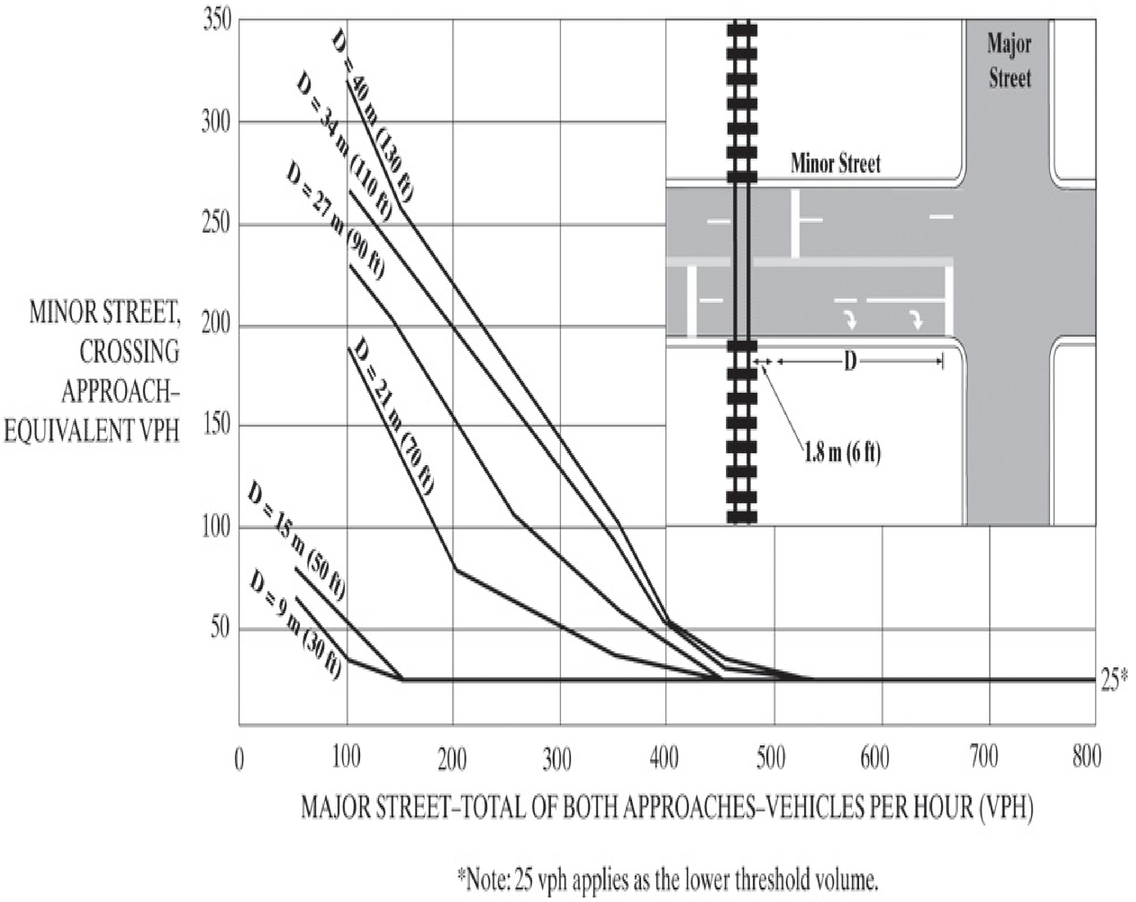


(Source: *Manual of Uniform Traffic Control Devices*, Federal Highway Administration, Washington, D.C., 2009, revised through 2012, Figure 4C-9, pg 447.)

Figure 15.9: Full Alternative Text

Figure 15.10: Warrant 9:

Railroad Crossings for Two- or More-Lane Approaches



(Source: *Manual of Uniform Traffic Control Devices*, Federal Highway Administration, Washington, D.C., 2009, revised through 2012, Figure 4C-10, pg 447.)

Figure 15.10: Full Alternative Text

The minor-street volume used in entering either Figure 15.9 or 15.10 may be multiplied by up to three adjustment factors: (1) an adjustment for train volume (Table 15.11), (2) an adjustment for presence of high- occupancy buses (Table 15.12), and (3) an adjustment for truck presence (Table 15.13). The base conditions for Figures 15.9 and 15.10 include four trains per day, no buses, and 10% trucks.

Table 15.11: Adjustment Factor for Train Frequency

Trains per Day	Adjustment Factor
1	0.67
2	0.91
3 to 5	1.00
6 to 8	1.18
9 to 11	1.25
12 or more	1.33

(Source: *Manual of Uniform Traffic control Devices*, Federal Highway Administration, Washington, D.C., 2009, revised through 2012, Table 4C-2, pg 448.)

Table 15.11: Full Alternative Text

Table 15.12: Adjustment Factor for High-Occupancy Buses

% of High-Occupancy Buses* on Minor-Street Approach	Adjustment Factor
0%	1.00
2%	1.09
4%	1.19
6% or more	1.32

*20 or more persons per bus.

(Source: *Manual of Uniform Traffic control Devices*, Federal Highway Administration, Washington, D.C., 2009, revised through 2012, Table 4C-3, pg 448.)

Table 15.12: Full Alternative Text

Table 15.13: Adjustment Factor for Tractor-Trailer Trucks

% of Tractor-Trailer Trucks on Minor-Street Approach	Adjustment Factor	
	D Less Than 70 ft	D of 70 ft or More
0%–2.5%	0.50	0.50
2.6%–7.5%	0.75	0.75
7.6%–12.5%	1.00	1.00
12.6%–17.5%	2.30	1.15
17.6%–22.5%	2.70	1.35
22.6%–27.5%	3.28	1.64
More than 27.5%	4.18	2.09

(Source: *Manual of Uniform Traffic Control Devices*, Federal Highway Administration, Washington, D.C., 2009, revised through 2012, Table 4C-4, pg 448.)

Table 15.13: Full Alternative Text

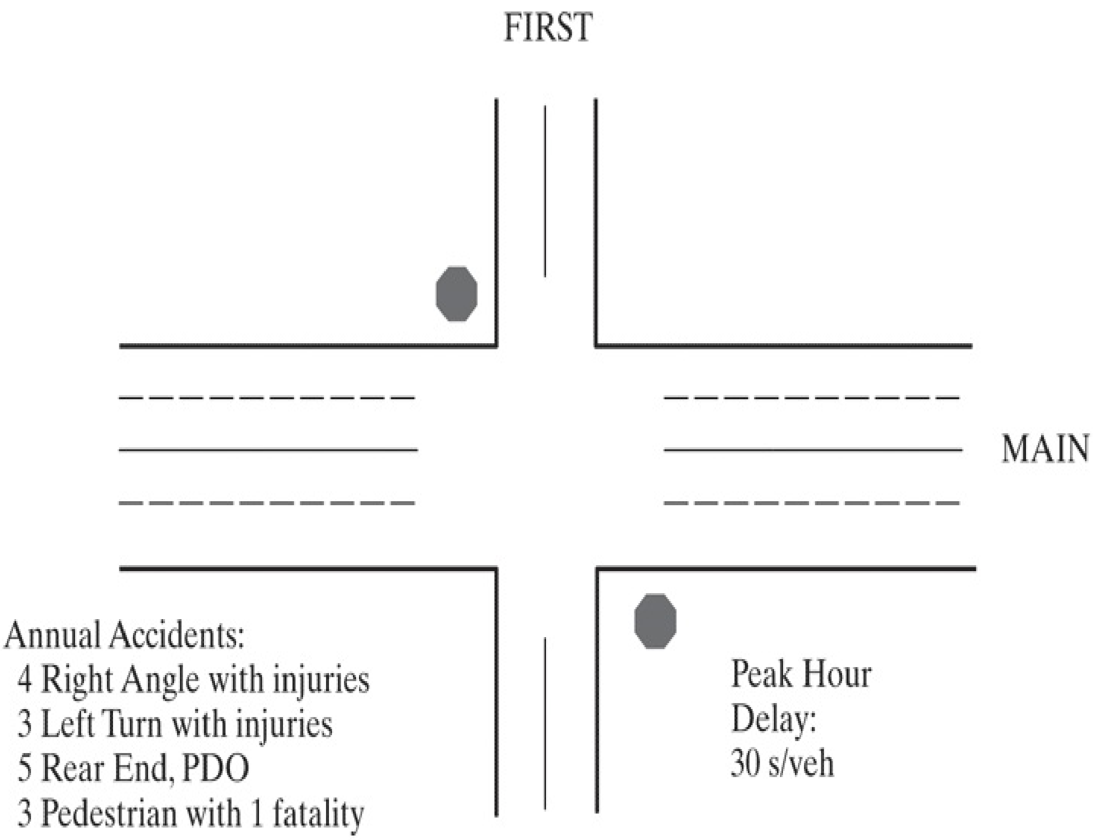
15.3.4 Summary

It is important to reiterate the basic meaning of these warrants. No signal should be placed without an engineering study showing that the criteria of at least one of the warrants are met. On the other hand, meeting one or more of these warrants *does not* necessitate signalization. Note that every warrant uses the language “The need for a traffic control signal *shall be considered*...” (emphasis added). While the “shall” is a mandatory standard, it calls only for consideration, not placement, of a traffic signal. The engineering study must also convince the traffic engineer that installation of a signal will improve the safety of the intersection, increase the capacity of the intersection, or improve the efficiency of operation at the intersection before the signal is installed. That is why the recommended information to be collected during an “engineering study” exceeds that needed to simply apply the nine warrants of the MUTCD. In the end, engineering judgment is called for, as is appropriate in any professional practice.

Sample Problem 15-3: Signal Warrant Analysis

Consider the intersection and related data shown in Figure 15.11.

Figure 15.11: Intersection and Data for Sample Problem in Signal Warrant



Time	Main Street Vol (veh/h)			First Ave Vol (veh/h)			Ped Vol (ped/h) Xing Main
	EB	WB	TOT	NB	SB	High Vol	
11 AM – 12	400	425	**825**	75	80	**80**	115
12 – 1 PM	450	465	**915**	85	85	**85**	120
1 – 2 PM	485	500	**985**	90	100	**100**	125
2 – 3 PM	525	525	**1050**	110	115	**115**	130
3 – 4 PM	515	525	**1040**	100	95	**100**	135
4 – 5 PM	540	550	**1090**	90	100	**100**	140
5 – 6 PM	550	580	**1130**	110	125	**125**	120
6 – 7 PM	545	525	**1070**	96	103	**103**	108
7 – 8 PM	505	506	**1011**	90	95	**95**	100
8 – 9 PM	485	490	**975**	85	75	**85**	90
9 – 10 PM	475	475	**950**	75	60	**75**	50
10 – 11 PM	400	410	**810**	50	55	**55**	25

Figure 15.11: Full Alternative Text

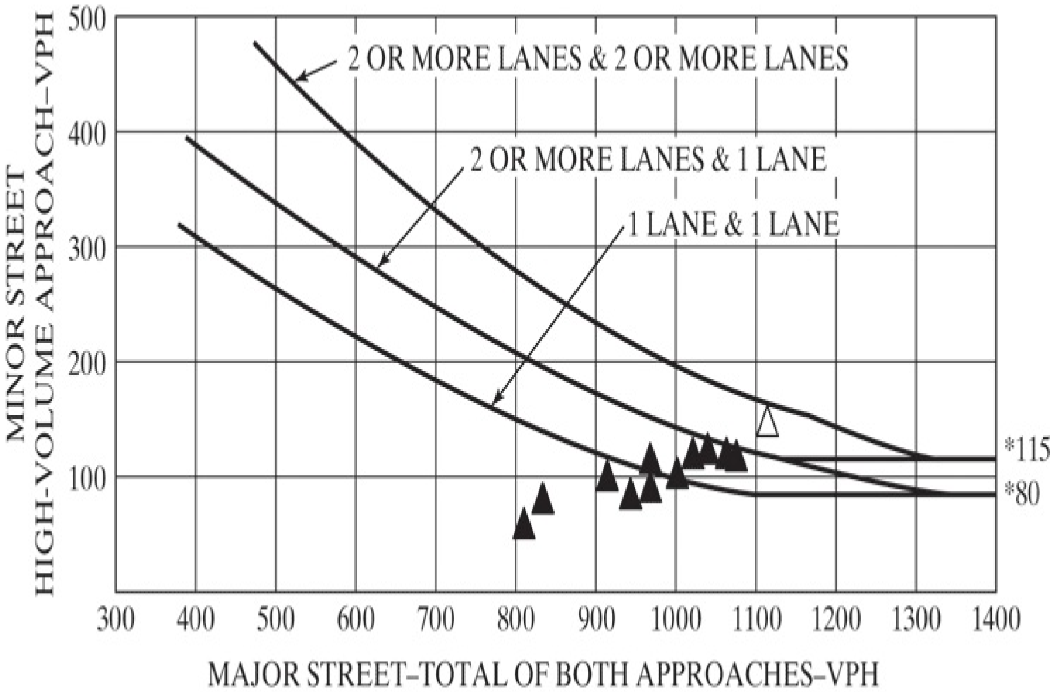
Note that the data is formatted in a way that is conducive to comparing with warrant criteria. Thus, a column adding the traffic in each direction on the major street is included, and a column listing the “high volume” in one direction on the minor street is also included. Pedestrian volumes are summarized for those crossing the major street, as this is the criterion used in the pedestrian warrant. As will be seen, not every warrant applies to every intersection, and data for some warrants is not provided.

Each of the applicable warrants are analyzed in turn.

- Warrant 1: There is no indication that the 70% reduction factor applies, so it is assumed that either Condition A or Condition B must be met at 100%, or both must be met at 80%. Condition A requires 600 veh/h in both directions on the multilane major street and 150 veh/h in the high-volume direction on the one-lane minor street. While all 12 hours shown in Figure 15.11 are greater than 600 veh/h, none have a one-way volume equal to or higher than 150 veh/h on the minor street. Condition A is not met. Condition B requires 900 veh/h on the major street (both directions) and 75 veh/h on the minor street (one direction). The 10 hours between 12:00 Noon and 10:00 PM meet the major-street criterion. The same 10 hours meet the minor-street criterion as well. Therefore, Condition B is met. As one condition is met at 100%, the consideration of whether both conditions are met at 80% is not necessary. *Warrant 1 is satisfied*.

- Warrant 2: Figure 15.12 shows the hourly volume data plotted against the four-hour warrant graph. The center decision curve (one street with multilane approaches, one with one-lane approaches) is used. Only one of the 12 hours of data is above the criterion. To meet the warrant, four are required. *The warrant is not met.*

Figure 15.12: Example Application of Warrant 2



*Note: 115 vph applies as the lower threshold volume for a minor street approach with two or more lanes and 80 vph applies as the lower threshold volume for a minor street approach with one lane.

Figure 15.12: Full Alternative Text

- Warrant 3: Figure 15.13 shows the hourly volume data plotted against the peak-hour volume warrant graph. Again, the center decision curve is used. None of the 12 hours of data is above the criterion. *The volume portion of this warrant is not met.*

Figure 15.13: Example Application of Warrant 3

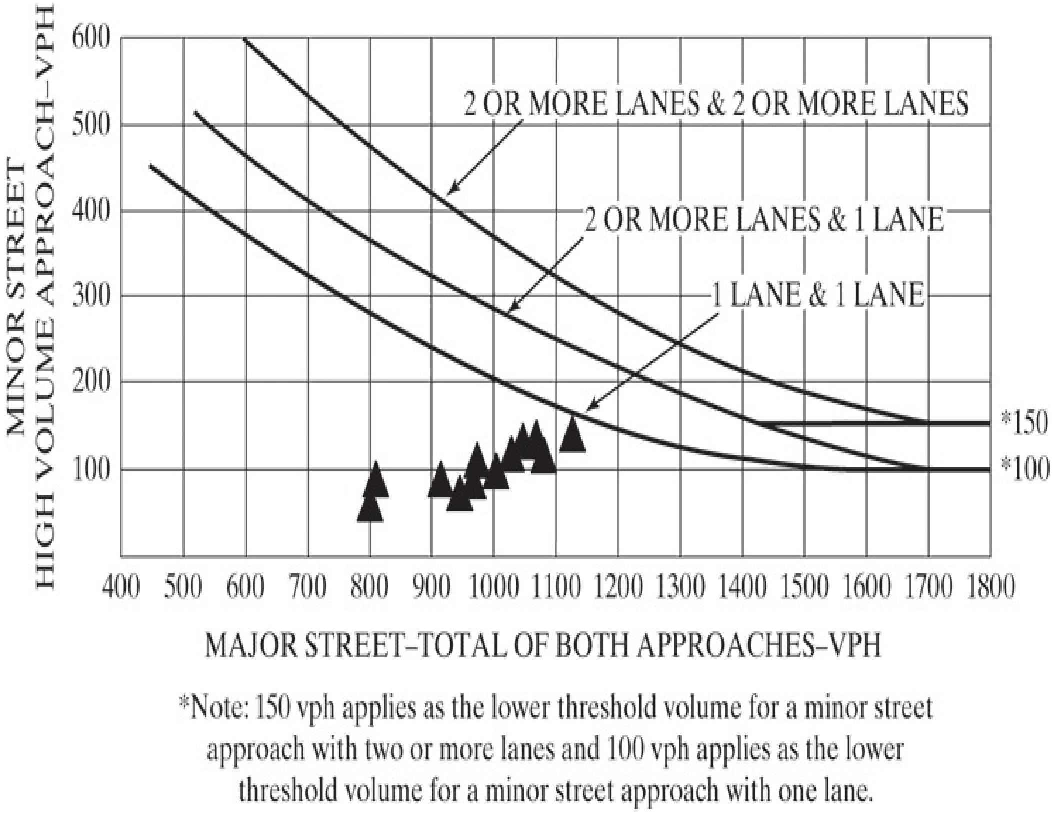
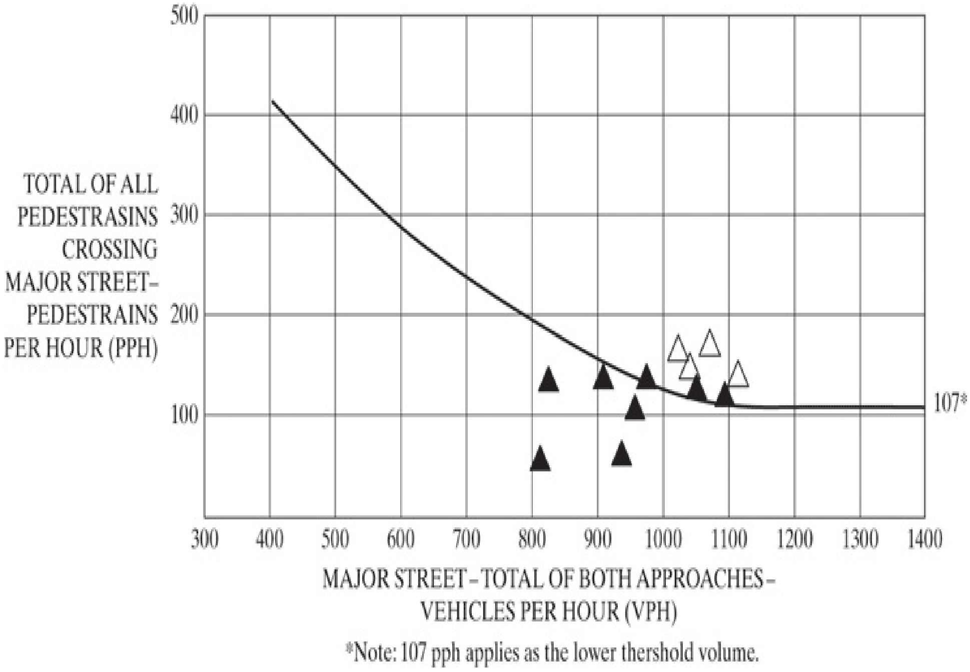


Figure 15.13: Full Alternative Text

The delay portion of the peak-hour warrant requires 4 veh-hours of delay in the high-volume direction on a STOP-controlled approach. The intersection data indicates that each vehicle experiences 30 s of delay. The peak one-direction volume is 125 veh/h, resulting in $125 \times 30 = 3{,}750$ veh-secs of aggregate delay, or $3{,}750/3{,}600 = 1.04$ veh-hrs of delay. This is less than that required by the warrant. *The delay portion of this warrant is not met*.

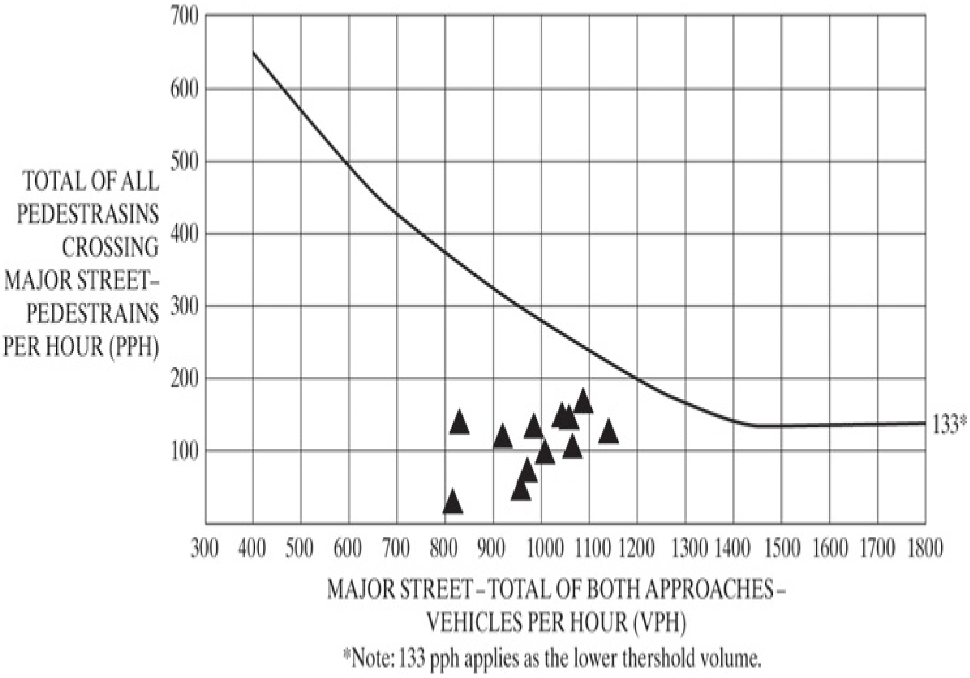
- Warrant 4: This warrant includes both a four-hour criterion and a peak-hour criterion, only one of which must be met to satisfy it. Figure 15.14 illustrates the solution.

Figure 15.14: Example Application of Warrant 4



(a) Four-Hour Pedestrian Warrant

15.3-14 Full Alternative Text



(b) Peak-Hour Pedestrian Warrant

15.3-14 Full Alternative Text

The 4-hour pedestrian warrant is met, while the peak-hour pedestrian warrant is not met. Because only one condition must be satisfied, *the* *pedestrian warrant is met.*

- Warrant 5: The school-crossing warrant does not apply. This is not a school crossing.

- Warrant 6: No information on signal progression is given, so this warrant cannot be applied.

- Warrant 7: The crash experience warrant has several criteria: Have lesser measures been tried? Yes, as the minor street is already STOP-controlled. Have five accidents susceptible to correction by signalization occurred in a 12-month period? Yes—four right-angle, three left-turn, and three pedestrian. Are the criteria for Warrants 1A or 1B met to the extent of 80%? Yes, Warrant 1B is met at 100%. *Therefore, the crash experience warrant is met.*

- Warrant 8: There is no information given concerning the roadway network, and the data reflects an existing situation. This warrant is not applicable in this case.

- Warrant 9: As this situation is not a highway-rail grade crossing location, this warrant does not apply.

In summary, a signal should be considered at this location, as the criteria for Warrants 1B (Interruption of Continuous Traffic), 4 (Pedestrians), and 7 (Crash Experience) are all met. Unless unusual circumstances are present, it would be reasonable to expect that the accident experience will improve with signalization, and it is, therefore, likely that one would be placed.

The fact that Warrant 1B is satisfied may suggest that a semiactuated signal be considered. In addition, Warrant 4 requires the use of pedestrian signals, at least for pedestrians crossing the major street. If a semiactuated signal is installed, it must have a pedestrian push-button (for pedestrians crossing the major street). The number of left turning accidents may also suggest consideration of protected left-turn phasing.

15.4 Closing Comments

In selecting an appropriate type of control for an intersection, the traffic engineer has many things to consider, including sight distances and warrants. In most cases, the objective is to provide the minimum level of control that will assure safe and efficient operations. In general, providing unneeded or excessive control leads to additional delay to drivers and passengers. With all of the analysis procedures and guidelines, however, engineering judgment is still required to make intelligent decisions. It is always useful to view the operation of existing intersections in the field in addition to reviewing study results before making recommendations on the best form of control.

References

- 1. *A Policy on Geometric Design of Highways and Streets*, 5th Edition, American Association of State Highway and Transportation Officials, Washington, D.C., 2004.

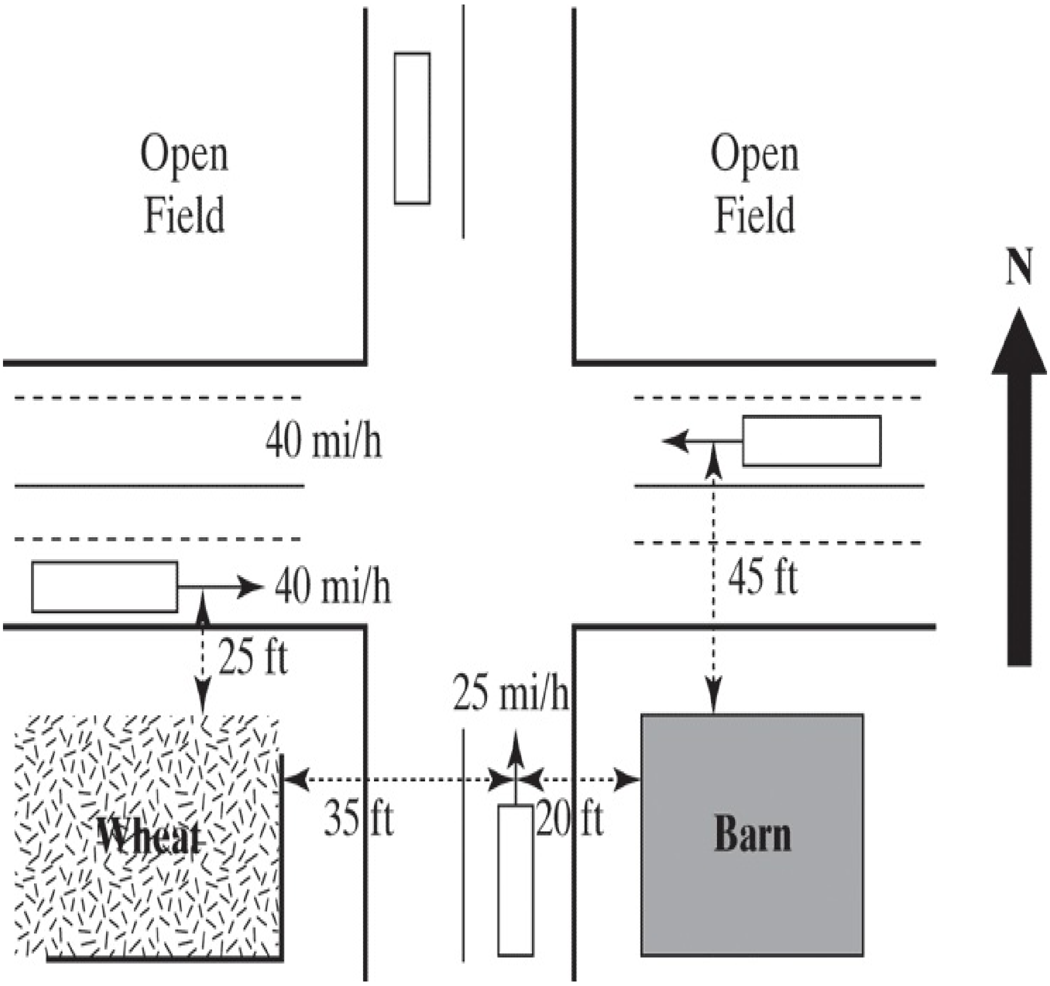
- 2. *Manual on Uniform Traffic Control Devices*, Federal Highway Administration, U.S. Department of Transportation, Washington, D.C., 2009, revised through 2012.

- 3. *Highway Capacity Manual, 6th Edition: A Guide for Multimodal Mobility Analysis*, Transportation Research Board, National Research Council, Washington, D.C., 2016.

Problems

1. 15-1. Consider the rural intersection shown on page 334. Can this intersection be safely operated under basic rules-of-the-road? If not, what type of control would you suggest? You may assume that a traffic signal is not warranted at this location. All grades are level, and the standard reaction time of 2.5 s may be used.

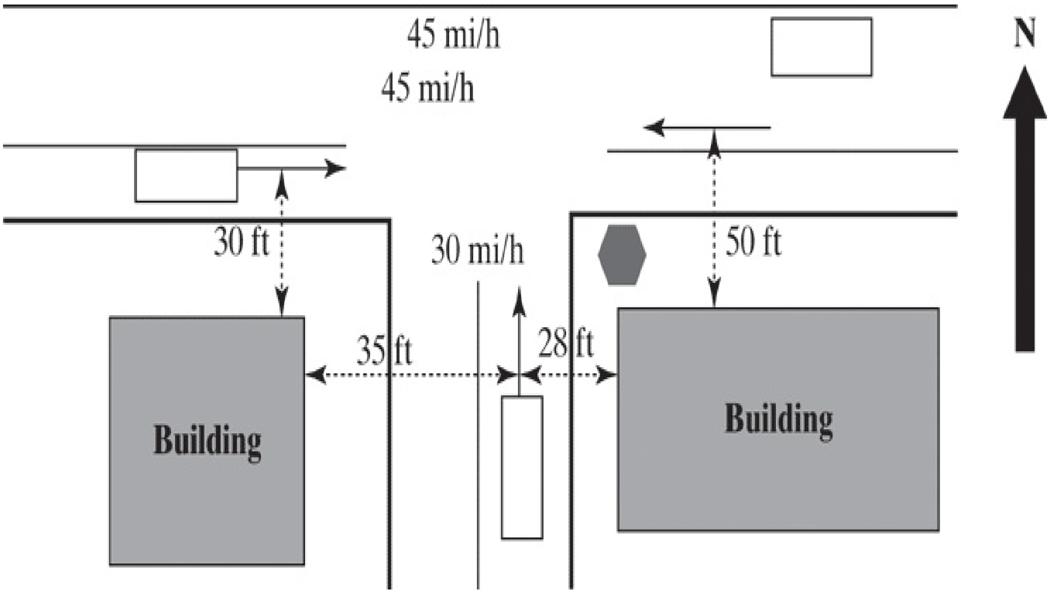
1. **Intersection for [Problem 15-1]**



[Full Alternative Text]

2. 15-2. For the STOP-controlled intersection on page 334, is the sight distance sufficient for safe operations? If not, what do you recommend? Again, you may assume that a traffic signal is not warranted at this location. All grades are level, and the standard reaction time of 2.5 s may be used. Lane widths on the E-W street are 12 ft.

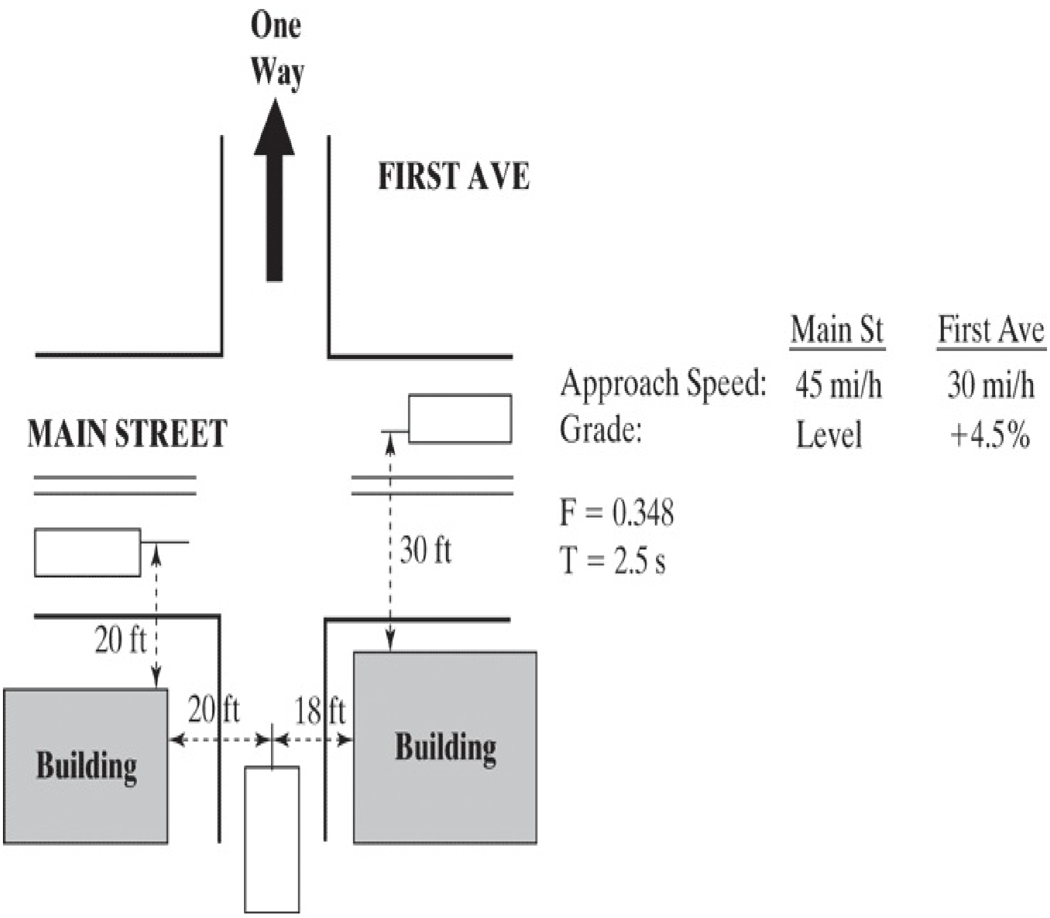
1. # Intersection for Problem 15-2



Full Alternative Text

2. 15-3. Determine whether the intersection shown on page 334 can be safely operated under basic rules of the road. If not, what form of control would you recommend, assuming that signalization is not warranted?

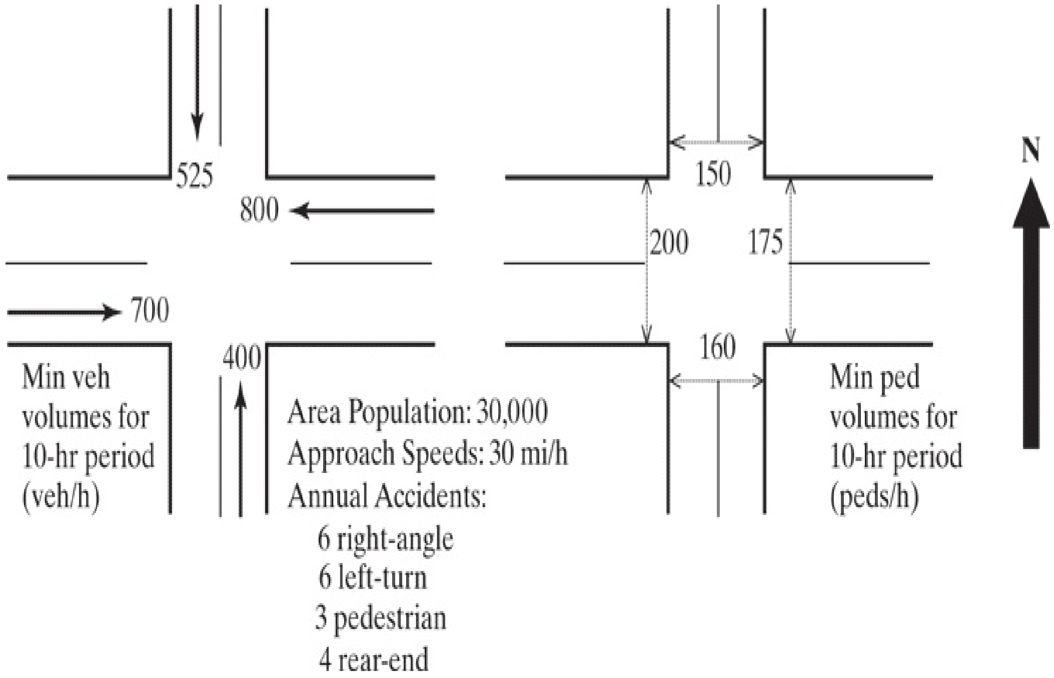
1. # Intersection for Problem 15-3



Full Alternative Text

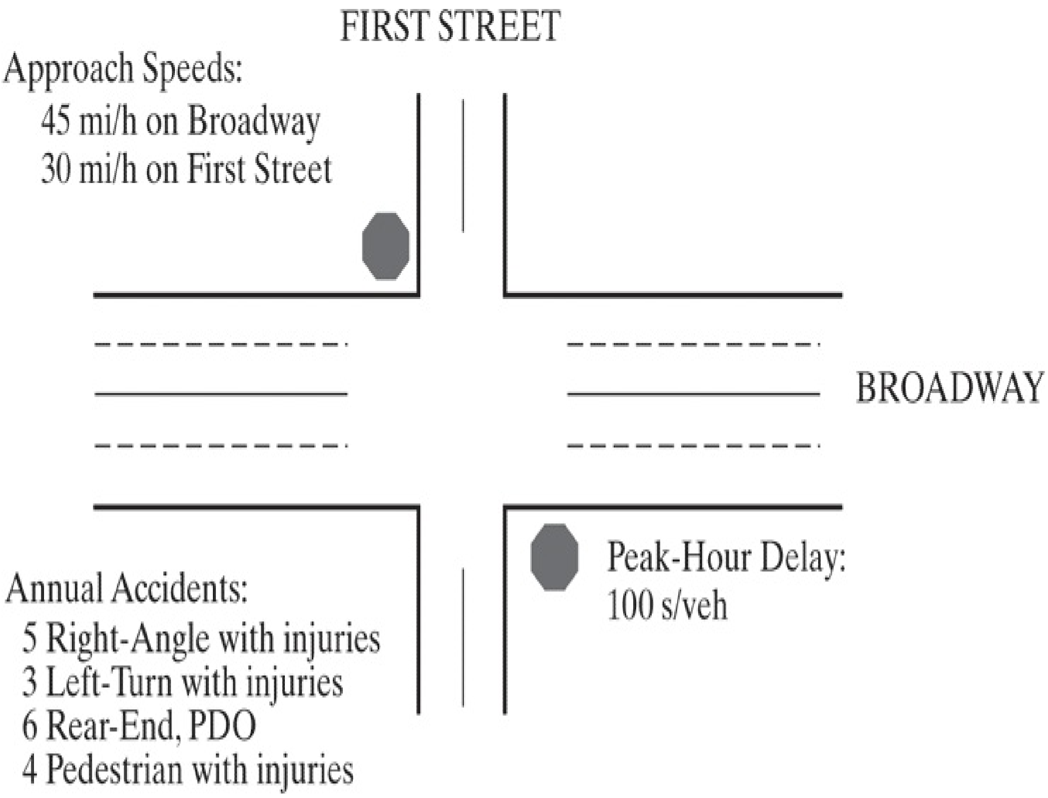
2. 15-4, 15-5,15-6,15-7. For each of the intersections shown below, conduct an analysis of signal warrants to determine whether or not a traffic signal should be used to control the intersection. For each warrant, determine whether the warrant is (a) met, (b) not met, or (c) not applicable, or insufficient information available for the determination. For each intersection, should a signal be installed? If so, can any insights be offered on the type of signal control to be implemented?

1. **Intersection for Problem 15-4**



Full Alternative Text

1. # Intersection for Problem 15-5



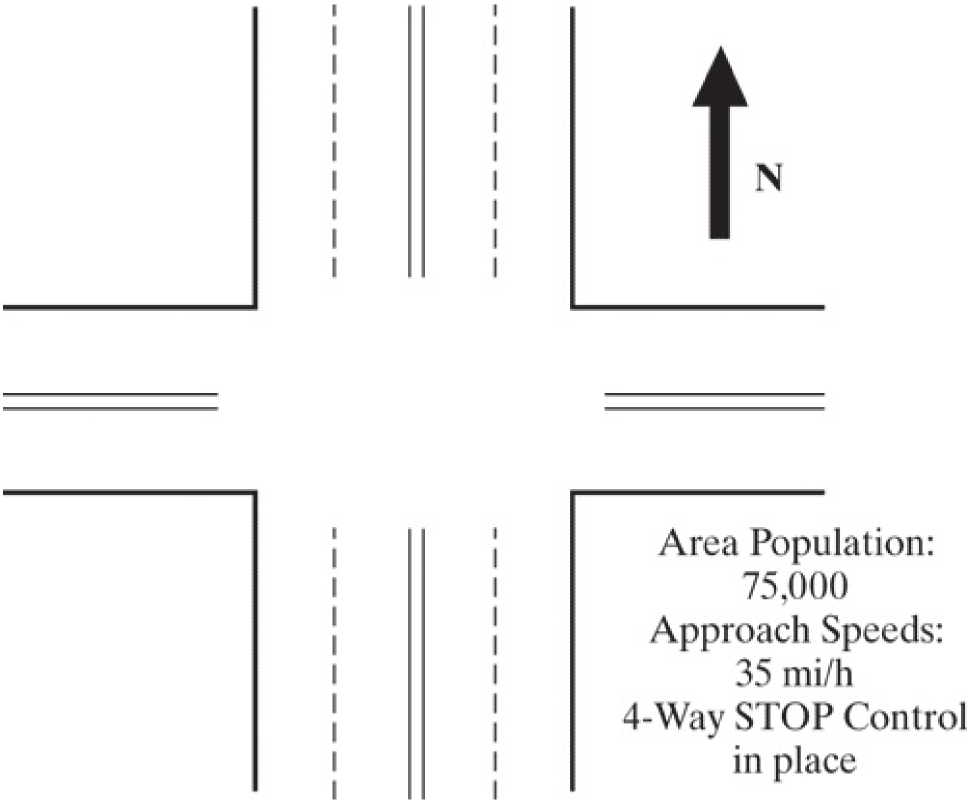
Full Alternative Text

Volume Data for Problem 15-5

Time	Volume on Broadway (veh/h)			Volume on First Street (veh/h)			Peds Xing Broadway (Peds/h)
	EB	WB	Total	NB	SB	High Vol	
10–11 AM	730	700	**1430**	300	400	**400**	**140**
11–12 AM	775	700	**1475**	300	400	**400**	**150**
12–01 PM	800	710	**1510**	315	410	**410**	**190**
01–02 PM	800	715	**1515**	325	420	**420**	**210**
02–03 PM	820	720	**1540**	350	450	**450**	**220**
03–04 PM	830	725	**1555**	360	450	**450**	**220**
04–05 PM	900	780	**1680**	400	480	**480**	**200**
05–06 PM	925	790	**1715**	410	520	**520**	**200**
06–07 PM	950	800	**1750**	375	510	**510**	**230**
07–08 PM	950	800	**1750**	350	480	**480**	**250**
08–09 PM	940	750	**1690**	320	420	**420**	**220**
09–10 PM	880	700	**1580**	306	400	**400**	**190**
10–11 PM	750	690	**1440**	295	390	**390**	**140**
11–12 PM	650	630	**1280**	260	380	**380**	**100**

Full Alternative Text

1. # Intersection for Problem 15-6

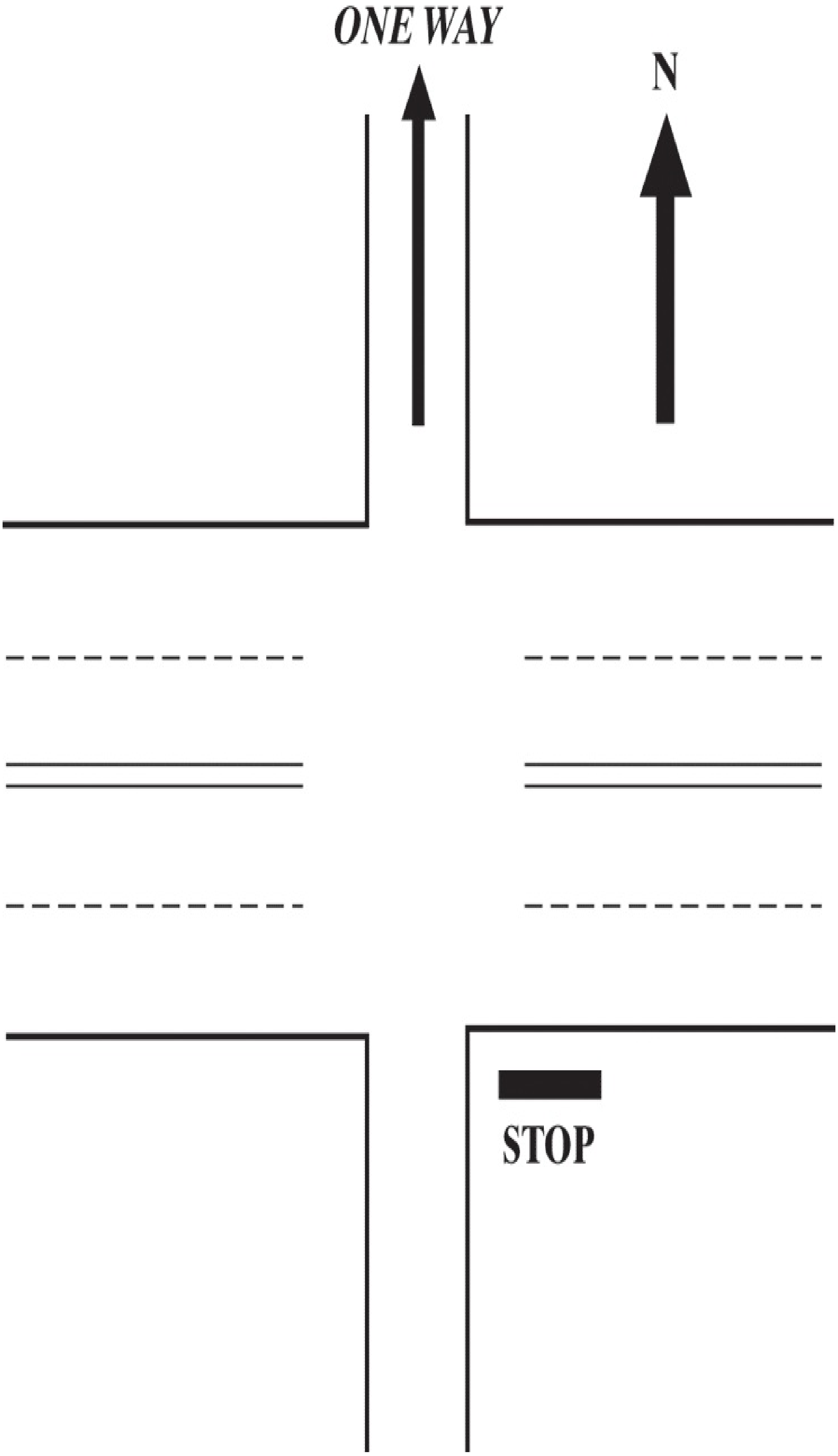


15.2-20 Full Alternative Text

	Volumes (veh/h)			
Hour	EB	WB	NB	SB
1	30	30	25	25
2	30	30	50	50
3	50	50	75	100
4	50	50	150	150
5	75	100	250	200
6	100	250	400	300
7	125	400	500	350
8	150	450	500	350
9	200	375	450	300
10	250	300	200	200
11	200	300	150	150
12	150	150	150	150
13	100	100	150	150
14	100	100	150	200
15	100	75	150	200
16	250	100	200	250
17	325	125	350	250
18	375	150	400	300
19	400	150	350	450
20	425	150	350	450
21	325	100	200	200
22	150	75	100	100
23	100	50	50	50
24	50	25	50	50

15.2-21 Full Alternative Text

1. **Intersection for Problem 15-7**



15.2-22 Full Alternative Text

Area Population: 40,000

Approach Speeds: 45 mi/h (E-W); 30 mi/h (N)

Accidents (Last 12 Months, Reportable):

- 8 right-angle
- 6 rear-end
- 3 left-turn
- 4 pedestrian

Delay to STOP-Controlled Vehicles: 72 s/veh

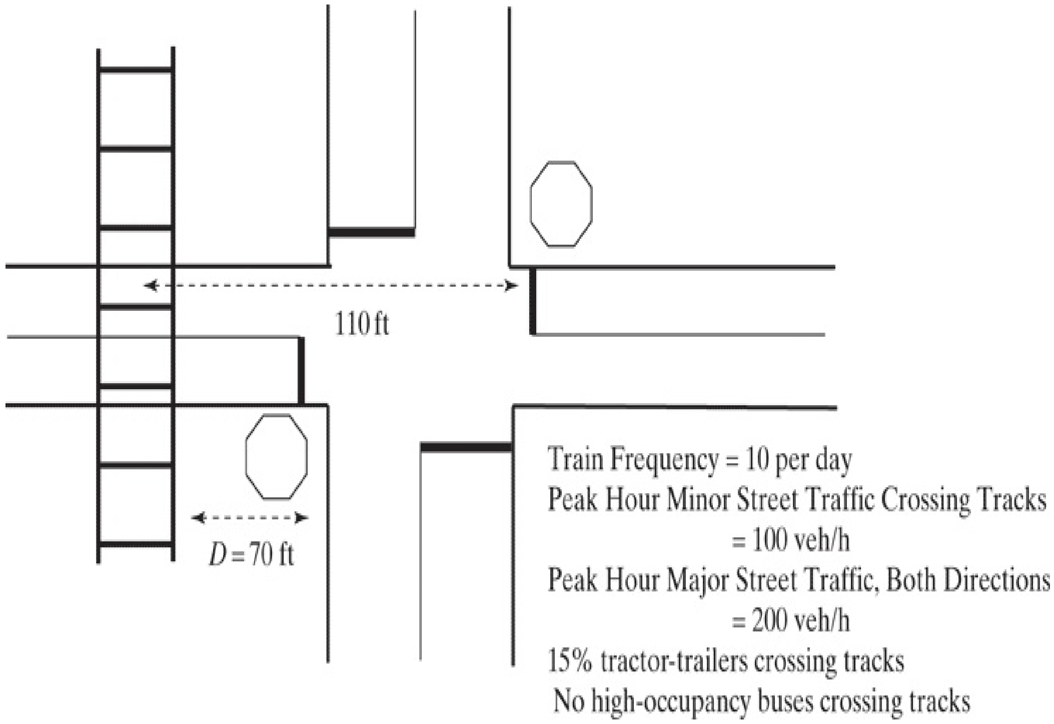
Volumes by Hour

Time Period	(Veh/h) Major St Total	(Veh/h) Minor St Total	(Peds/h) Peds Xing Major St
1–2 PM	800	100	200
2–3 PM	855	100	210
3–4 PM	1025	130	205
4–5 PM	1150	160	193
5–6 PM	1350	108	180
6–7 PM	1200	135	170
7–8 PM	1000	100	180
8–9 PM	975	85	200
9–10 PM	800	80	150
10–11 PM	900	80	100

15.2-23 Full Alternative Text

2. 15-8. The figure that follows illustrates a STOP-controlled intersection near a highway-railroad grade crossing. Should this intersection be signalized under the new Warrant 9, which applies to such situations?

1. **Intersection for Problem 15-8**



Full Alternative Text

Chapter 16 Traffic Signal Hardware

Signalizing an intersection, and in many cases coordinating that intersection along an arterial or in a network, is a significant engineering task. The traffic engineer is specifically involved in the function of the signal and the design of the control it imposes, as well as with the means of communicating the control information to the driver. The traffic engineer must also be aware of the complex and rapidly developing hardware that implements traffic signal control, even though the details of the technology are designed primarily by electrical and software engineers. This chapter provides an overview of hardware and technologies involved in signalization.

The reader is fortunate in that two of the three key references cited in this chapter are available for PDF download at no cost. The *Traffic Signal Manual: Second Edition* [1] was published in 2015 as National Cooperative Highway Research Program (NCHRP) Report 812, and can be downloaded at http://onlinepubs.trb.org/onlinepubs/nchrp/nchrp_rpt_812.pdf; it is also referred to as TSM2. The *Manual on Uniform Traffic Control Devices* [2] can be downloaded at https://mutcd.fhwa.dot.gov/. At this writing, the official version includes revisions 1 and 2, but the reader should confirm at the website just referenced that later editions or revisions have redefined the version current at the time of the reading.[1]

[1]As cited elsewhere in this text, the Federal *MUTCD* is a model document, with each state adopting its own version as its official document (it is expected to be in substantial compliance with the Federal model). Many simply adopt the Federal version as released, but some do add or edit. One state adopts the Federal version, but allows cities above a certain size to make their own decision.

The reader is *urged* to download these two documents, and make them part of his or her personal library. There is simply too much valuable information in them to provide more than an overview in this chapter. Indeed, some problems at the end of this chapter may require reference to one document or the other.

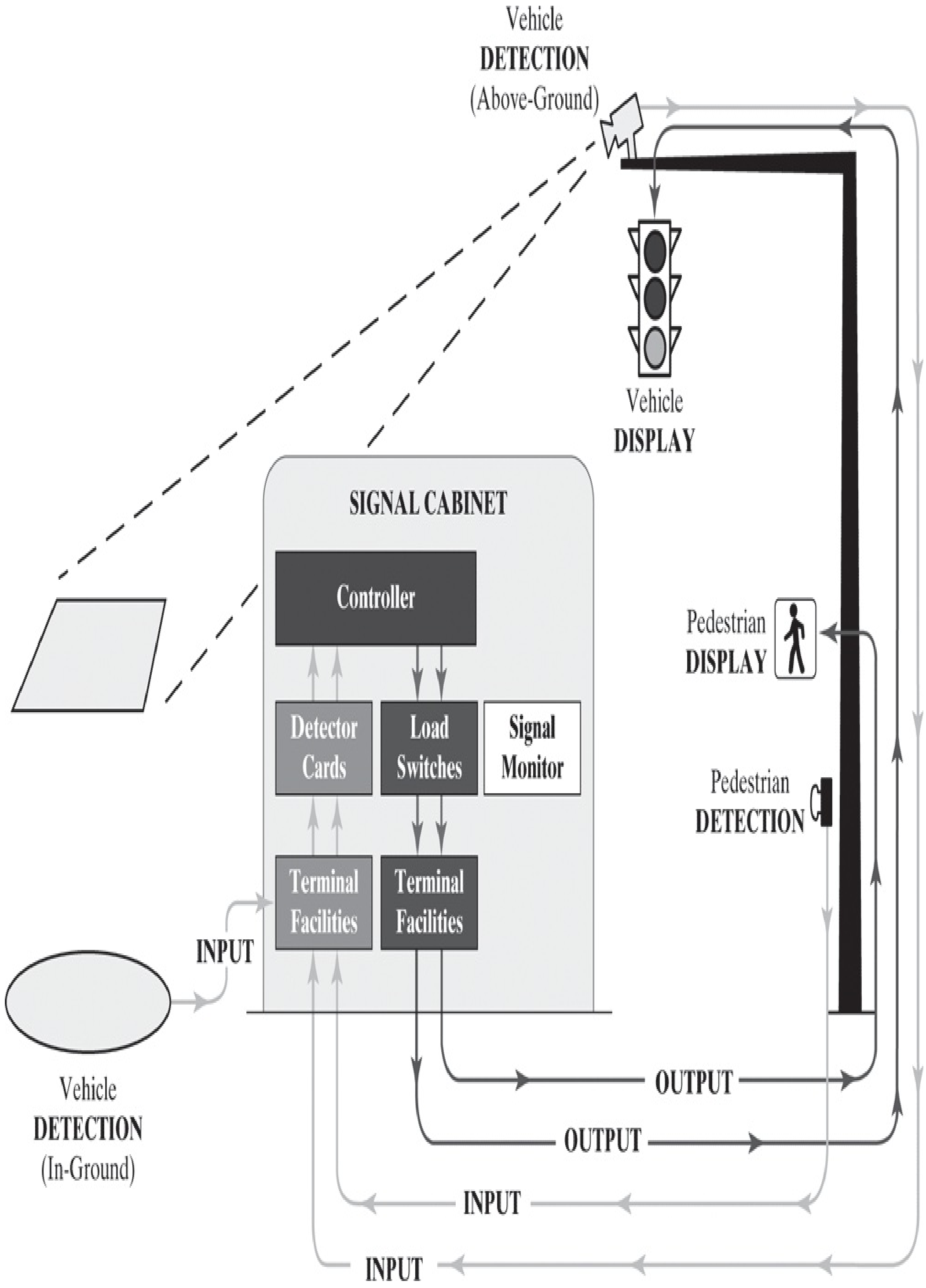
The third key reference is the *Highway Capacity Manual (HCM)* [3]. It is central to other chapters in this text, but for present purposes, it is simply listed because Ref. [1] refers to the *HCM* as the source of the movement and phase conventions shown in Ref. [1] and in this chapter.

16.1 Functional Layouts at a Signalized Intersection

Figure 16.1 shows a functional layout of equipment at a signalized intersection, and is well known to most people simply by their daily observation. The most visible parts are:

- The *vehicular signal displays*, which vary in number, placement, and complexity depending upon the intersection's geometry and traffic patterns. The *MUTCD* specifies the conventions in great detail;

Figure 16.1: Functional Depiction of Equipment at a Signalized Intersection



(Source: Reprinted with permission from *Traffic Signal Manual: Second Edition*, National Cooperative Highway Research Program (NCHRP) Report 812, Transportation Research Board, © 2015 by the National Academy of Sciences, Courtesy of the National Academies Press, Washington, D.C.)

Figure 16.1: Full Alternative Text

- The *pedestrian signal displays*, which provide comparable guidance to pedestrians by display of hand symbols in accord with the *MUTCD*. The use of countdown timers in concert with the hand symbols is common. Not all intersections have pedestrian signal displays, but in all cases the signal timing is to be designed to allow for pedestrian crossing times when pedestrians are present;

- *Bicyclist signal displays* are not shown in Figure 16.1, but are present for certain potential conflicts, such as through bicyclists and turning vehicles when a designated bikeway is present on the street. Generally, bicyclists are governed by the vehicular traffic displays;

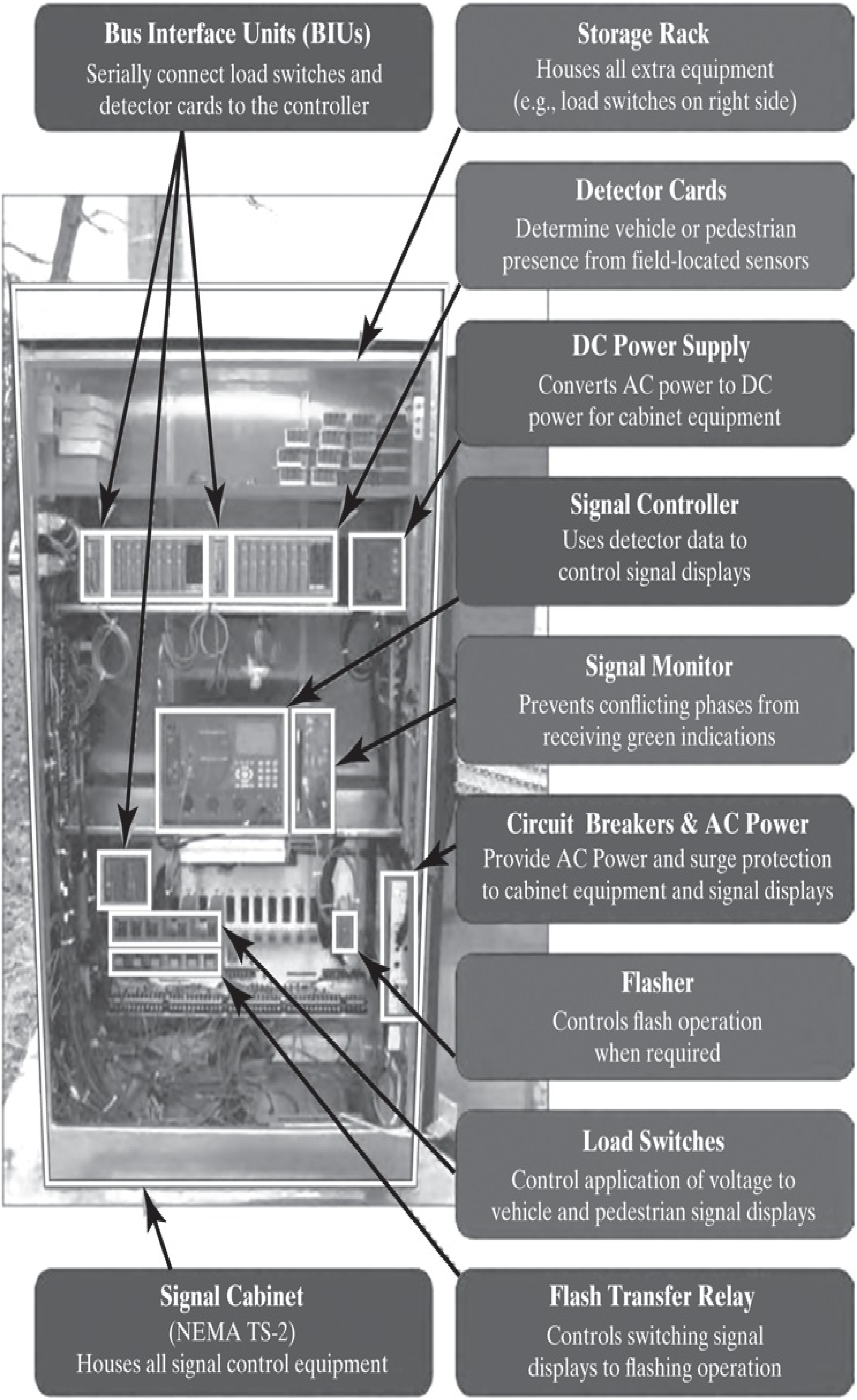
- *Detectors* to sense the presence of vehicles, pedestrians, and in some cases bicyclists. The vehicle detectors are either placed in the pavement (inductive loops of wire being the most common) or mounted above-ground (microwave and video being common, with multi-lens cameras that can place "virtual detectors" on all approaches by software becoming more common). Pedestrian detectors are most often push-buttons, although true zone sensors that detect pedestrians are being used at this writing;

- The *signal cabinet* that houses the actual traffic controller, the detector cards, the switches[2] that activate the signal lights, communications gear, uninterrupted power supply (UPS), and ancillary equipment.

  [2]Just as people still speak of "dialing a phone" that is digital and has no dial, terminology such as "switches" has to be understood to be electronic signals as well as mechanical switches. The phrase "activate the signal lights" can easily include sending the power to an LED display that has replaced actual signal lights seen through a colored lens.

The cabinet tends to be of standard sizes, and the display hardware (and some detector hardware) is standardized by the *MUTCD* and by connector standards. The advantages of this standardization include: (1) people can expect to see consistency in messaging, location, and appearance; (2) jurisdictions do not accidentally "lock themselves into" a limited number of vendors for cabinets and other equipment; and (3) nonfunctioning equipment can be swapped out more easily. Figure 16.2 shows a representative layout within a signal cabinet. Notice that the actual traffic signal controller can occupy only a small part of the cabinet.

Figure 16.2: Representative Layout within a Signal Cabinet (NEMA TS-2)

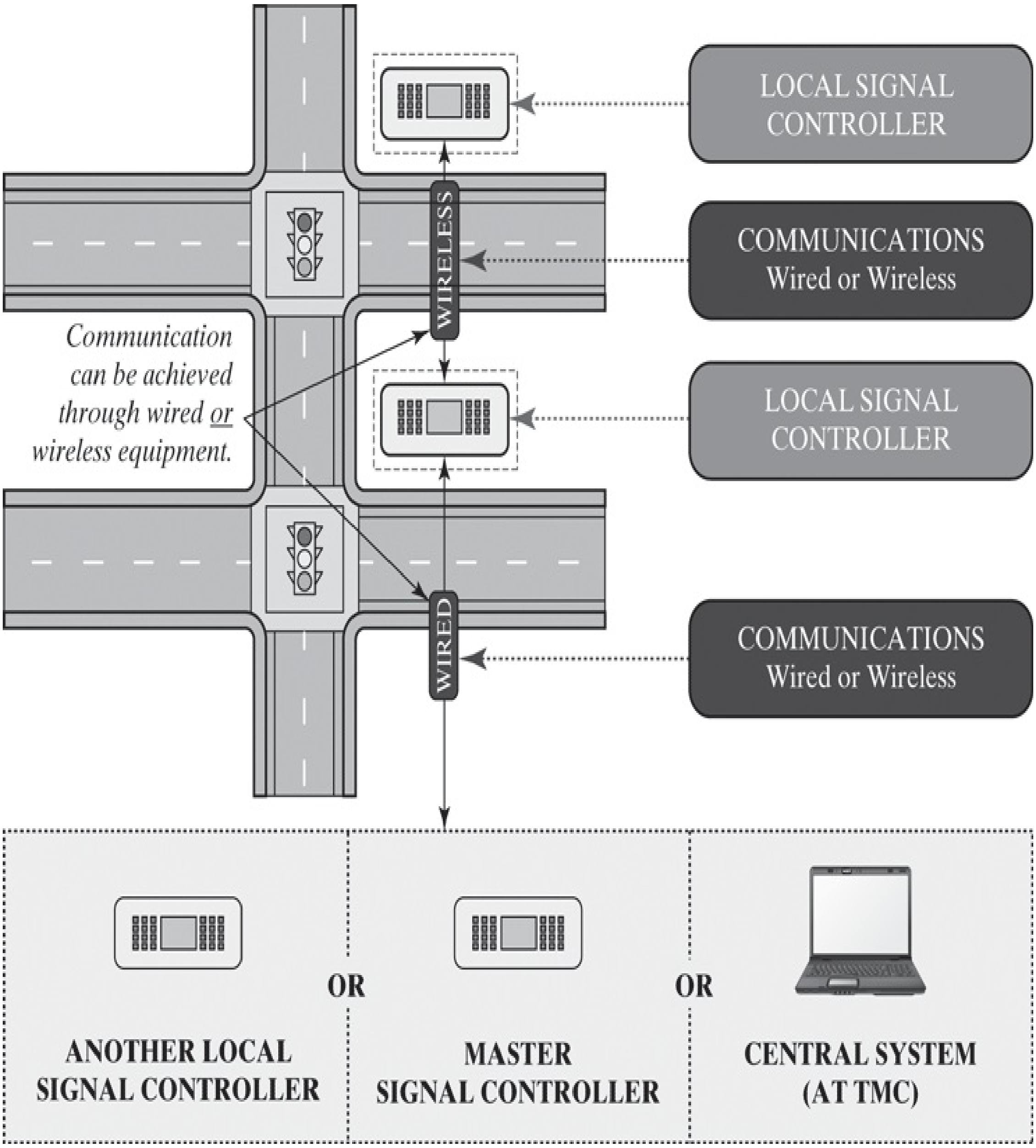

Bus Interface Units (BIUs)
Serially connect load switches and detector cards to the controller
Storage Rack
Houses all extra equipment (e.g., load switches on right side)
Detector Cards
Determine vehicle or pedestrian presence from field-located sensors
DC Power Supply
Converts AC power to DC power for cabinet equipment
Signal Controller
Uses detector data to control signal displays
Signal Monitor
Prevents conflicting phases from receiving green indications
Circuit Breakers & AC Power
Provide AC Power and surge protection to cabinet equipment and signal displays
Flasher
Controls flash operation when required
Load Switches
Control application of voltage to vehicle and pedestrian signal displays
Signal Cabinet
(NEMA TS-2)
Houses all signal control equipment
Flash Transfer Relay
Controls switching signal displays to flashing operation

(Source: Reprinted with permission from *Traffic Signal Manual: Second Edition*, National Cooperative Highway Research Program (NCHRP) Report 812, Transportation Research Board, © 2015 by the National Academy of Sciences, Courtesy of the National Academies Press, Washington, D.C.)

Figure 16.2: Full Alternative Text

While some intersections truly operate independent of all others (that is, “isolated” in terms of control and interaction), it is much more common that nearby signals interact. Figure 16.3 shows the concept, using either wired or wireless connectivity.

Figure 16.3: Interaction and Coordination amongst Traffic Signal Controllers



(Source: Reprinted with permission from *Traffic Signal Manual: Second Edition*, National Cooperative Highway Research Program (NCHRP) Report 812, Transportation Research Board, © 2015 by the National Academy of Sciences, Courtesy of the National Academies Press, Washington, D.C.)

Figure 16.3: Full Alternative Text

Chapter 21 addresses the development of the design of traffic signal coordination plans; the present chapter simply cites the hardware aspects of the coordination.

16.2 Some History

While there is some discussion among historians on the “first” traffic signal, there is some agreement that the first modern traffic signal was placed at the intersection of Euclid Avenue and 105th Street in Cleveland, Ohio on August 5, 1914. The signal included a single head with four faces, each containing a red and a green ball indication (no yellow). The “controller” was a booth located in the intersection in which a police officer manually changed the signal as appropriate.

For decades, most intersections were controlled by electro-mechanical controllers that worked as follows: a synchronous motor used the line frequency to achieve A fixed rotation speed, which was then passed through a “timing gear” that drove a “timing dial” at a speed determined by the timing gear ratio; this became the desired “cycle length” of the signal displays; a “timing dial” rotated once per cycle, and was divided into 100 slots; “timing keys” were placed to trigger a camshaft at certain times; individual plastic cams on the camshaft had a sequence of “teeth” that could be broken out, and were matched to a specific signal display (north-south green, north-south yellow, north-south red, and so forth); touching onto the cams were mechanical relays; when a cam rotated to a broken-tooth location because the camshaft rotated, the relay dropped into the space and the connection was made to send power to the associated signal display; when the same advanced and hit a non– broken-tooth location, the relay lifted and the electrical connection to the associated signal display was broken.

Of course, it was quite important that cams did not have break patterns that allowed conflicting movements to occur simultaneously (green in all directions, for instance). Manufacturers could provide controllers with what today would be called “default settings”: a set of cams broken to a simple signal pattern at a four-legged intersection. A web search on “electromechanical traffic controller” will result in finding a number of videos that show the process in detail.

One has to remember that such controllers predated today’s computers on a chip and much of today’s technology. They were common for most of the 20th century, and exist at some locations even at this writing.

As the need developed for coordination of traffic signals (for instance, main street green being turned on in a certain pattern along an arterial), schemes were developed to accommodate this with electromechanical controllers. This included the use of a “master controller” that sent signals to other connected controllers on when to initiate main street green.

Of necessity, electromechanical controllers displayed a fixed timing pattern, controlled by the timing dial (fixing the cycle length) and the location of the timing keys (fixing the duration of each signal pattern or “phase”). Because traffic patterns vary by time of day, cabinets sometimes contained “multi-dial” controllers, with predetermined switch times from one dial to another. Three-dial controllers were common, with one for AM, another for PM, and the third used for other periods.

So, time-of-day control was achieved with the technology of the era; master controllers could be programmed to provide more choices.

One of the earliest large computer controlled systems (New York City) simply used trigger signals from a central location—the Traffic Management Center (TMC), in today’s terminology—to *each* signal controller, to advance the camshaft. From this, both cycle length *and* phase durations could be varied, and would not be limited by the number of dials in a given controller cabinet. Of course, full use of such a system could result in adaptive control, except for the fixed phase sequences, *if* there were sufficiently reliable detectors in the field and *if* the computing power and the software existed.3

[3]As it developed, one of the greatest benefits of early computer control was gaining knowledge on whether the intersection equipment was actually working. The ability to have maintenance logs, dispatch crews, and be aware of when the signals came back on line was a true breakthrough.

Concurrent with the electromechanical controller, other controllers were developed circa the mid-20th century, each hardwired using the technology of the time, but allowing different types of flexibility:

- Semi-actuated controllers that let green rest on main street, but provided green to the side street based upon vehicles arriving at detectors (subject to main street having some minimum green and side street having some maximum green);

- Fully actuated controllers that had detectors on all approaches, and allocated green based upon arrivals, with comparable maximum-minimum concepts. Fully actuated controllers were used when the demand was comparable on two competing approaches and/or highly variable over a day on two competing approaches;

- Volume-density controllers that took into account the number of arrivals and the gap patterns in the arrivals (indicative of density), to provide more sophisticated, responsive control at an isolated intersection.

As technology advanced, so did the sophistication of controllers. For present purposes, suffice it to say that the profession gravitated to functional standards on what features a controller is to have, and which communications and interface protocols are to be used. *Today's controllers are designed for actuated control; pre-timed control can however be achieved by setting the minimum and maximum greens to force pre-timed operation.* Further, these controllers are designed to be part of integrated systems and to communicate as part of such systems.

16.3 Controller and Other Standards

Standardization is an invaluable practice in engineering design, construction, and many aspects of life. While a valid argument can be made that premature standardization can freeze development before a "best" approach is established, standardization does in fact provide for cost-efficiencies, interoperability, independence from being locked into a specific manufacturer, and even safety—imagine a world in which traffic signal colors have different meanings from country to country, or state to state, and/or the order of the colors on a signal head varies. The *MUTCD* [2] provides standards for traffic markings, signage, signal displays, and related matters—colors, patterns, sizes, placement, and so forth.

One of the authors would chat with a class about a home project, which would be very nicely completed with the addition of a single sheet of plywood in a highly visible location—with the dimensions 5 foot by 9 foot. The students were asked to go purchase such a sheet. After some polite nibbling at the subject, the students became animated, and the questioning turned to "What is wrong with you? Such a thing does not exist. Even if you could special-order it, the delay and the cost would not justify designing such a need into the project." And then the discussion of standards and standardization began.

In traffic engineering, standardization is not limited to the *MUTCD.* In the context of the present chapter, NEMA and NTCIP are keywords when one thinks of standards:

- NEMA is the abbreviation for the National Electrical Manufacturers Association [4], the largest association of electrical equipment manufacturers in the United States, founded in 1926. One of its key roles is establishing standards for a wide range of devices, including traffic signals, traffic signal cabinets, phasing diagrams, and hardware for dynamic message signs.
- NCTIP is the abbreviation for the National Transportation Communications for Intelligent Transportation Systems (ITS)

Protocol [5]. The focus is on communications devices, interconnection, and interfaces for ITS.

In practice, standards evolve through a consensus building process that includes key professional or trade organizations, facts from research, participation and reviews by relevant professional societies, and an approval process.

- For instance, the NTCIP is a joint standardization project that involves AASHTO, ITE, NEMA, and the USDOT's Office of the Assistant Secretary for Research and Technology. There is a substantial committee structure that considers relevant standards in subareas, submits them for comment, and moves them into the approval process.

- In another illustrative case, two of the authors were involved in the process that led to *MUTCD* Traffic Signal Warrant 9, Intersection Near a (Railroad) Grade Crossing. The process began with some states recommending that research on the subject be sponsored by the National Cooperative Highway Research Program (NCHRP), through the AASHTO process that sought such recommendations. The research was funded, and then awarded to an organization through NCHRP's process for requesting proposals to conduct the work. The cited two authors led the research team that was selected. In addition to conducting the research, it was vital to (a) report the results to a variety of interested parties, and (b) draft and redraft the recommended warrant, given the feedback. The interested parties included the NCHRP Project Advisory Panel, some Transportation Research Board committees, an ITE group on which several ITE committees were represented, and ultimately the National Committee on Uniform Control Devices (NCUTCD) [6]. After review, comment, and approval, the NCUTCD recommended the proposed warrant to the FHWA, which published it in the *Federal Register* as part of the public comment in the rule-making. It was then incorporated into the 2009 *MUTCD [2]* after comments were considered by the FHWA.

The point is that the process of standard building is intensive, involves consensus building, formal acceptances, and an overall structured process.

Rather than spend many pages on the history of evolution of standards in traffic signal controllers and related devices (for example, cabinets) or even in the enumeration of them, the reader is referred to References [4–6] for current standards. The list is extensive and evolving.

A few key points, at this writing:

1. NEMA maintains the TS 2 standard for traffic signal controllers and related equipment. The related equipment includes cabinets, detectors, electrical bus interface units, and load switches. The standard defines functionality, interfaces, electrical specifications, environmental endurance, and some physical specifications. Maximum dimensions are specified for the controller, but a manufacturer is free to make a unit of any smaller size (and any material or shape)—that meet the other requirements of the standard.

2. The Advanced Transportation Controller (ATC) family of standards is maintained by the cited consortium of NEMA, ITE, and AASHTO. The ATC 2070 standard specifies every detail of the controller hardware and internal subsystems, but not the application software functionality. It also specifies the form and function of a number of modules as well as the standard chassis and card cage.

3. NEMA reports in Ref. [7] that “The NEMA TS 2 standard and the Advanced Transportation Controller (ATC) standard are not mutually exclusive nor contradictory. A controller can meet both standards...” and “...you’ll find that the trend in traffic controllers is to adopt the ATC standard (for its added benefits) along with the NTCIP standards (for their benefits), while retaining their NEMA capabilities.... To ensure this multi-standard approach, the NEMA member companies are major contributors to both the ATC standards and the NTCIP.”

4. The Adaptive Traffic Control System (ATCS) or Adaptive Signal Control Technology (ASCT) systems[4] addressed in a later section of this chapter is now widely accepted for traffic control and for real-time signal coordination. There are a variety of such systems, so that the existence of standards such as the NCTIP protocols and other standards have great importance.

   [4]Yes, terminology evolves. The reader may find one of these descriptors to be rather dated, or might find both in use.

The reader, particularly those entering the profession, is encouraged to not only be familiar with the various organizations (TRB, ITE, and AASHTO) but to also become active in these organizations and their committees. The process begins by attending committee meetings as a guest, contributing knowledge gained by one's own work, expressing interest, and meeting obligations assigned. The individual not only contributes to the profession, but also becomes involved in the evolving state of the art and state of the practice, and extends one's own professional network.

This section has not addressed international standards or cooperation explicitly. However, that dialog exists through organizations such as TRB and ITS America, and is a focus of FHWA.

16.4 Common Terminology

Traffic movement at a signalized intersection is regulated by signal displays as defined in the *MUTCD [2]*, communicating by color (red, yellow, and green), indication (steady and flashing), and image (ball and arrow). One of the most recent additions is the flashing yellow arrow (FYA).

A concurrent set of displays is called a “signal phase.” With modern controller equipment that is designed for actuated operation, the following parameters would be set for each signal phase:

1. Minimum Green Time: The minimum amount of green time that must be allocated to a phase when it is initiated. It is determined based upon detector design, approach speeds, and other factors.

2. Maximum Green Time: The maximum amount of green time that can be allocated to a particular phase in a cycle. It begins timing when a “call” for service on a conflicting phase is received.

3. Passage Time: This feature really provides three functions: (a) it is the amount of time added to a green phase when additional actuations are noted within the green, (b) it defines the minimum gap between successive vehicles in a lane required to retain the green, and (c) it must be at least long enough to allow for a vehicle to travel from the detector to the intersection line at the approach speed.

4. Recall: The recall feature automatically places a call on a designated phase, regardless of whether demand is present or not. At a time when no vehicles are present, the recall directs the green to a designated phase. There are four forms of recall in use: (a) *Minimum recall* requires that the minimum green be allocated on the recalled phase. The recall is cancelled once the phase is initiated. (b) *Maximum recall* places a continuous call for service on the designated phase, causing the phase to allocate the maximum green time. (c) *Pedestrian recall* places a continuous call for service for pedestrians in the designated phase, and allocates the minimum safe pedestrian crossing time. (d) *Soft recall* places a call for service on the designated phase in the

absence of a call on conflicting phases. Most recall settings are implemented in the *minimum* or *soft* category. To implement a pre-timed signal plan, all phases are set to *maximum recall*.

5. Simultaneous Gap or Force Out: When engaged, this feature requires that all movements simultaneously served by a phase be terminated at the same time.

6. Dual Entry: When engaged, this feature requires that all movements that can be served by a phase receive the green simultaneously, even if demand is present on only one such movement.

While modern signal controllers are designed for actuated operation, there are still many situations in which pre-timed operations are preferable. Table 16.1 provides an overview of situations in which pre-timed and actuated control is used at individual intersections.

Table 16.1: Selection of Appropriate Signal Control Mode

	Pre-Timed		Actuated		
Type of Operation	**Isolated**	**Coordinated**	**Semi-Actuated**	**Fully Actuated**	**Coordinated**
Fixed Cycle Length	Yes	Yes	No	No	Yes
Conditions Where Applicable	Where detection is not available.	Where traffic is consistent, closely-spaced intersections, and where cross-street is consistent.	Where de-faulting to one movement is desirable, major road is posted < 40 mi/h, and cross road carries light traffic demand.	Where detection is provided on all approaches, isolated locations where posted speed is > 40 mi/h.	Arterial where traffic is heavy and adjacent intersections are nearby.
Example Application	Work zones.	Central business districts, inter-changes.	Highway opera-tions.	Locations without nearby signals, rural high-speed locations, inter-sections of two arterials.	Suburban arterial.
Key Benefit	Temporary ap-plication keeps signals opera-tional.	Predictable operations, lowest cost of equipment and maintenance.	Lower cost for highway maintenance.	Responsive to changing traffic patterns, efficient allocation of green time, reduced de-lay, and improved safety.	Lower arterial delay, potential reduction in delay for the system, depending on the settings.

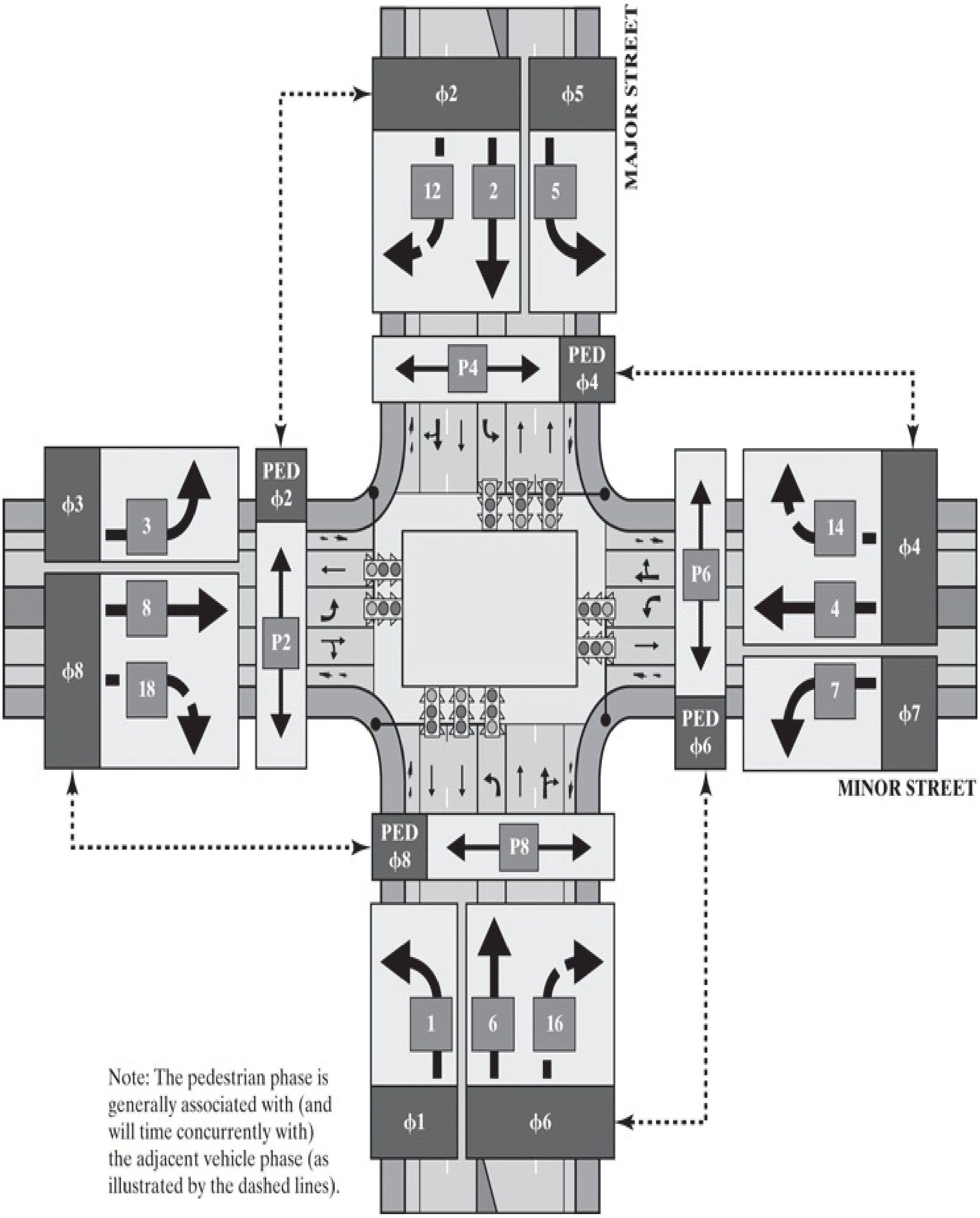
(Source: Kittelson and Associates, *Traffic Signal Timing Manual*, 1st Edition, Federal Highway Administration, Washington, D.C., June 2008, Table 5-1, pg 5-3.)

Table 16.1: Full Alternative Text

16.5 Convention for Numbering Movements and Phases

Figure 16.4 shows the defined labeling of each movement, shown as a solid arrow when protected and a dashed arrow when made concurrent with other, usually larger movements. A grouping of movements made concurrently is defined as a "Phase." The standard shorthand for a Phase is "ϕ," so that Phase 6 is also referred to as ϕ6, and includes Movements 6 and 16.

Figure 16.4: Convention for Labeling Movements and Phases, from the *HCM*



(Source: Reprinted with permission from *Traffic Signal Manual: Second Edition*, National Cooperative Highway Research Program (NCHRP) Report 812, Transportation Research Board, © 2015 by the National Academy of Sciences, Courtesy of the National Academies Press, Washington, D.C.)

Figure 16.4: Full Alternative Text

Other variations of Figure 16.4 are possible, depending upon whether movements are protected or permissive (for instance, left turns are sometimes permitted concurrent with the opposing green).

Note that Figure 16.4 shows the major street as the vertical one. Even though “north” is not indicated, this text will sometimes refer to “north” as the “vertical up” direction (for instance, Movement 6 might be called northbound, and the vertical called north-south, for the reader’s convenience).

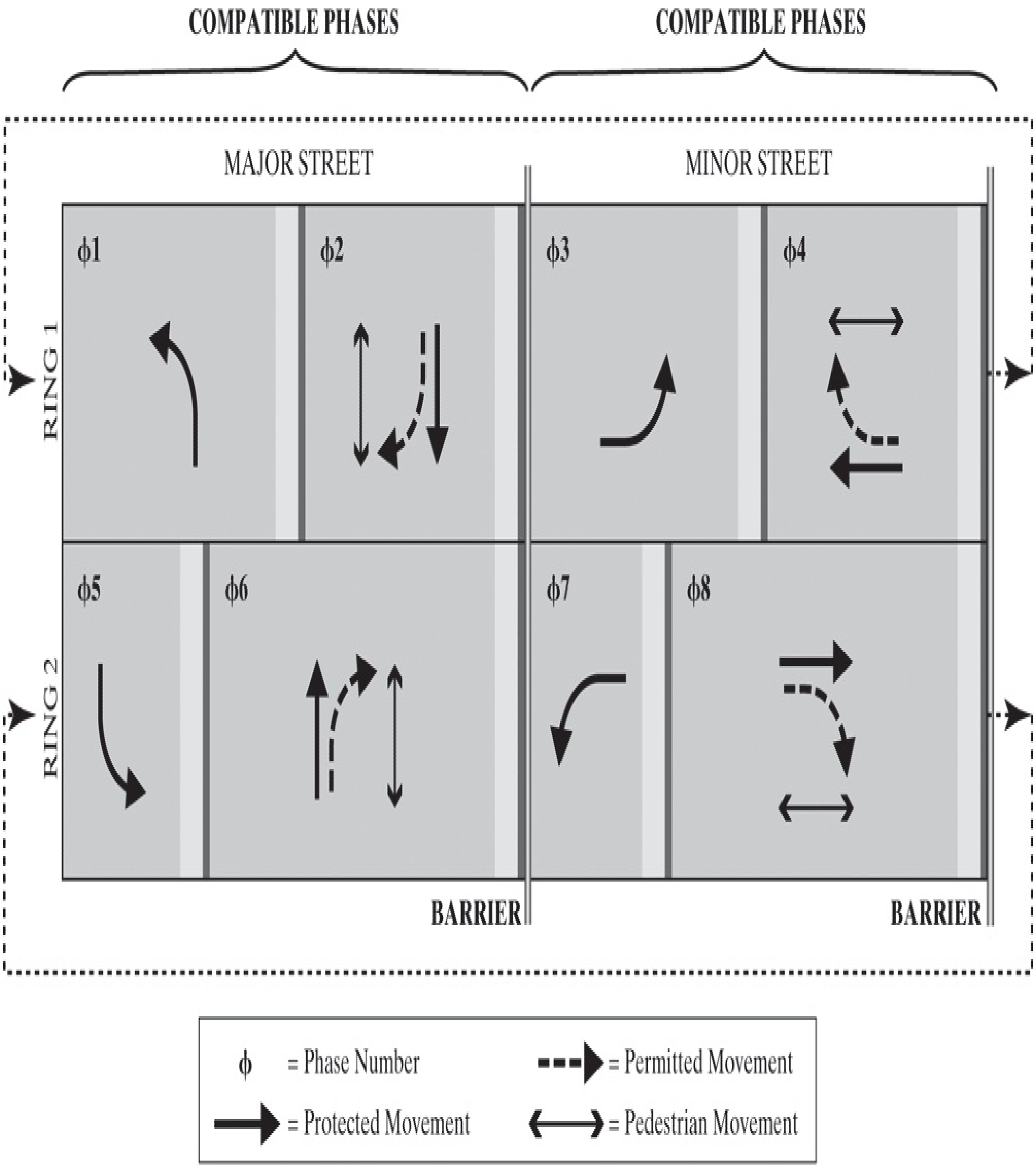
16.6 Ring-and-Barrier Diagram

Clearly, there are prohibited combinations in Figure 16.4. For instance, Movements 2 and 4 cannot be allowed at the same time.

A powerful tool for clarifying allowed combinations at a four-legged[5] intersection is the “ring and barrier diagram,” variations of which are shown in Figure 16.5.

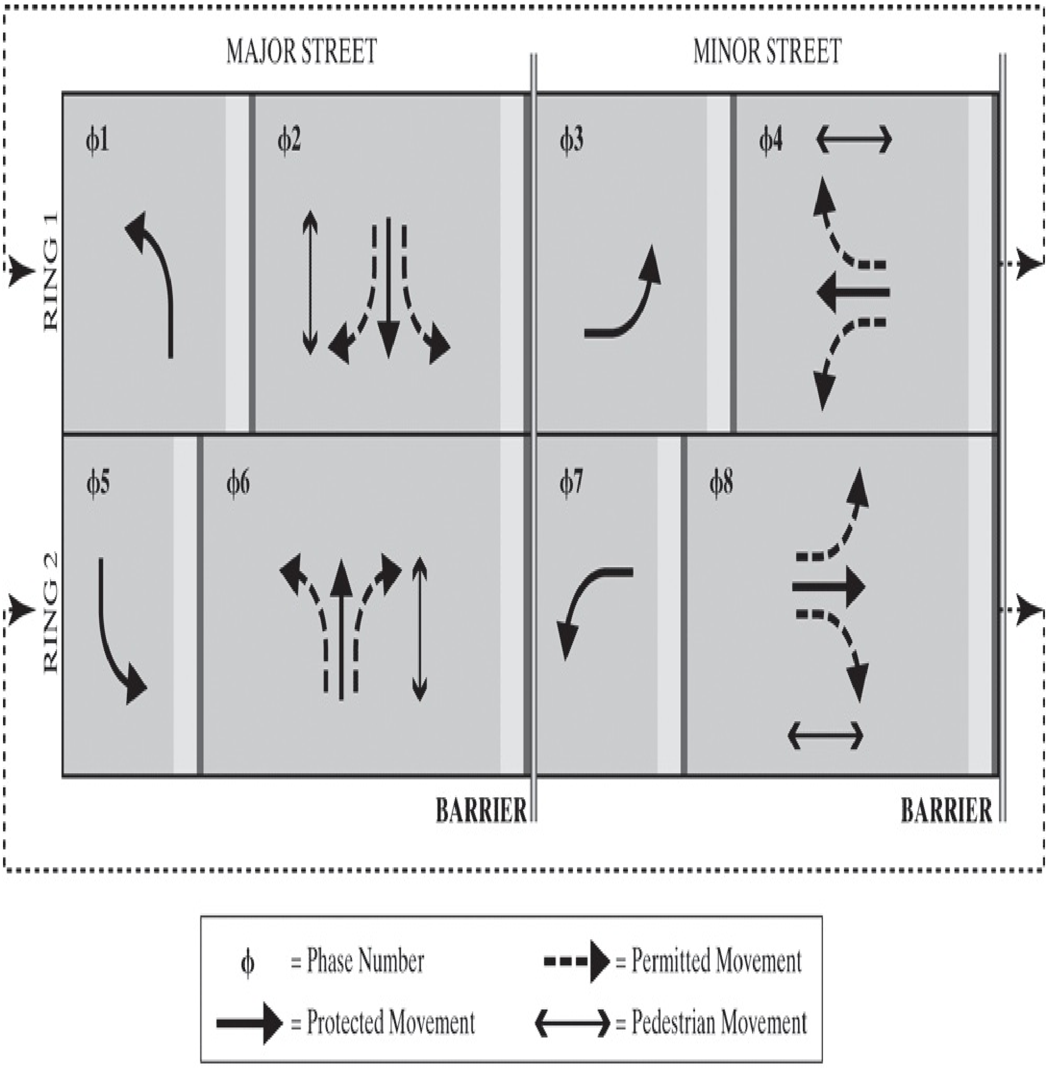
[5] Comparable diagrams are shown in [1] for three-legged intersections (“T” intersections) and intersections with five legs.

Figure 16.5: Basic Ring-and-Barrier Diagrams for Three Situations



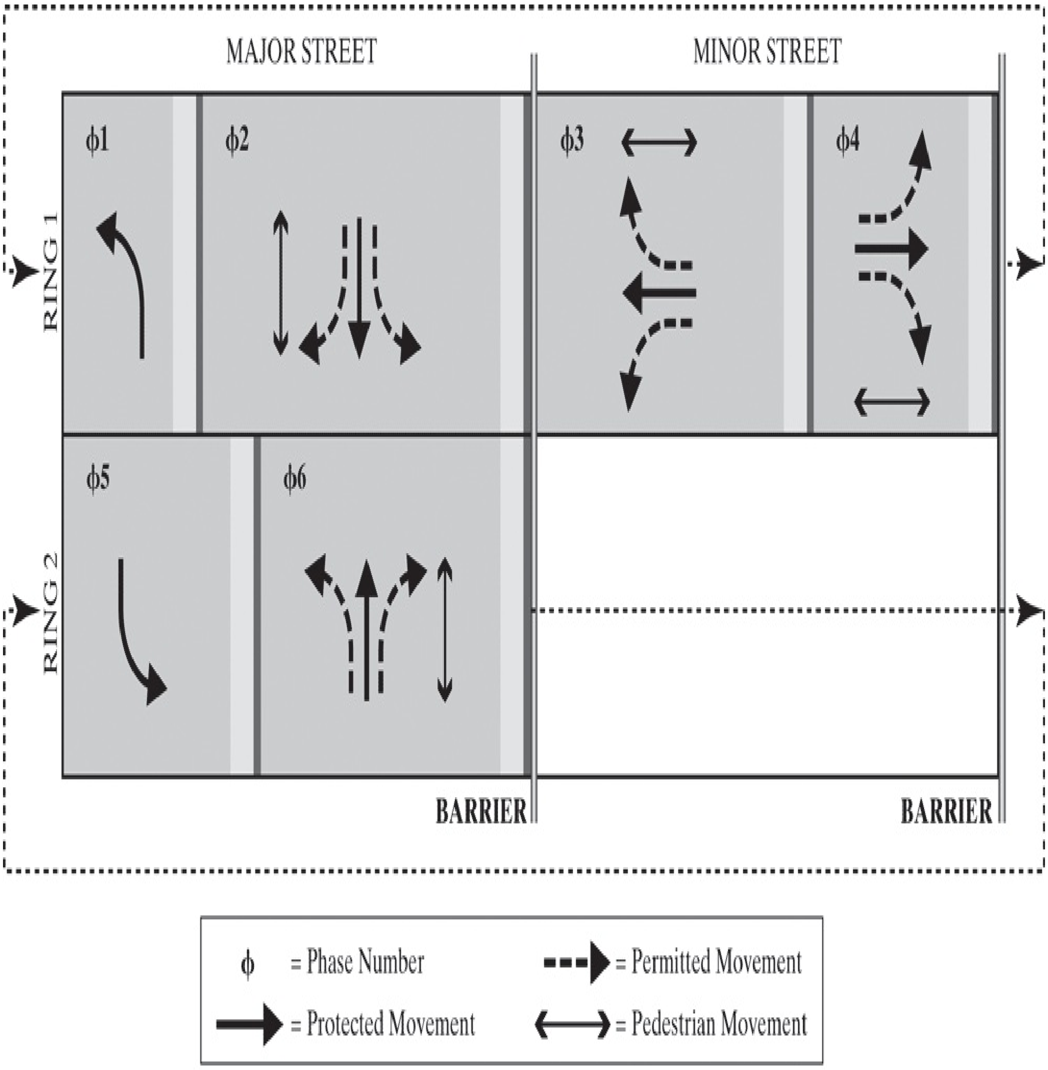
(a) Ring-and-Barrier Diagram for Protected Left Turn Phasing

[16.6-2 Full Alternative Text]



(b) Ring-and-Barrier Diagram for Protected-Permitted Left Turn Phasing

16.6-2 Full Alternative Text



(c) Ring-and-Barrier Diagram for Split Phasing

16.6-2 Full Alternative Text

(Source: Reprinted with permission from *Traffic Signal Manual: Second Edition*, National Cooperative Highway Research Program (NCHRP) Report 812, Transportation Research Board, © 2015 by the National Academy of Sciences, Courtesy of the National Academies Press, Washington, D.C.)

The key feature in Figure 16.5a (and the other parts) is the *barrier*: movements on one side of the barrier can *never* be allowed to be concurrent with movements on the other side of the barrier. For instance, a northbound left and an eastbound left cannot be allowed to be concurrent.

Note that Figure 16.5a is for an intersection at which *protected* left turns are required, and are shown as the *leading* movements in their respective directions. Be aware that some jurisdictions require protected left turns, and specify leading only for the lefts. Other jurisdictions allow protected-permitted left turns, and some have only permitted left turns (that is, they are made as the opportunity arises, across an active opposing thru movement). Refer to Figure 16.5b for an illustration of protected-permissive lefts, with leading lefts.

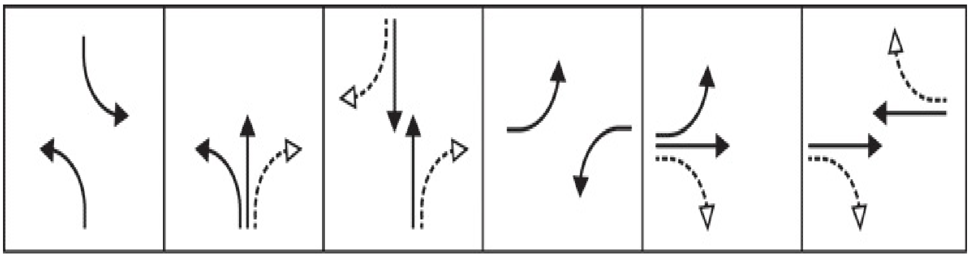
Returning to Figure 16.5a, the point has been made that the barriers separate movements that *cannot* be allowed to be concurrent. However, some movements *are* allowed concurrently. By convention, these are shown in two separate *rings* as on the top and bottom of Figure 16.5a (and the other parts). Each ring repeats continually, responding to traffic demand to govern durations.

Remember that Figure 16.5a is for protected leading left turns. While one can decipher what is happening at any given time by moving a straight-edge left to right on the diagram, it is sometimes easier to sketch the signal pattern as an observer in the field (and the drivers) would see the situation. This also helps as one learns to “read” the ring-and-barrier diagram.

For instance, for the display shown in Figure 16.5a, the actual sequence of signal displays will allow the movements shown in Figure 16.6b. However, it must be recognized that the actual display can vary, depending upon traffic demand:

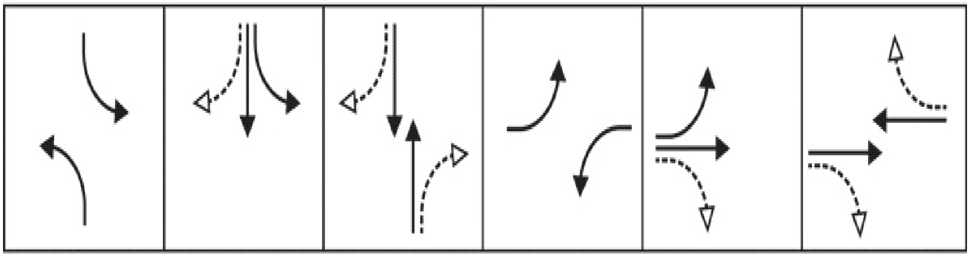
- Figure 16.6a shows that the northbound left has more demand than the southbound left. However, if the opposite were true, then Figure 16.6b would result as the display in the field;

Figure 16.6: Three Displays for Figure 16-5a, Depending upon Relative Demand



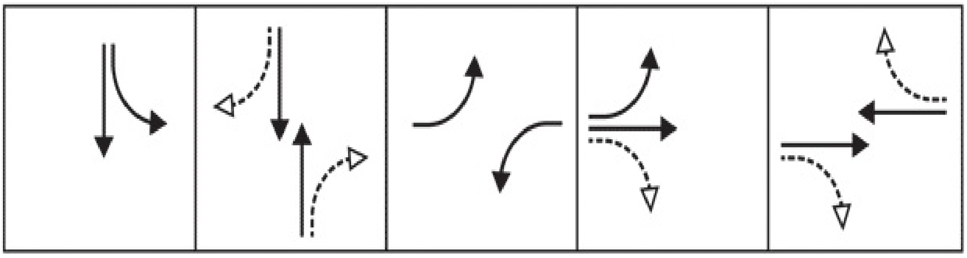
(a) Movement Durations of Figure 16-5

16.6-2 Full Alternative Text



(b) Southbound Left Greater than Northbound Left

16.6-2 Full Alternative Text



(c) No Northbound Left Demand

16.6-2 Full Alternative Text

- If *no* northbound left turns were detected, then Figure 16.6c could result.

Again, local practice and local regulation has to be considered: some jurisdictions require that all movements be allowed some minimum time, even when no demand is detected.

Figure 16.5b is quite similar to Figure 16.5a, except that after the protected part of the left turns, permissive operation is allowed.

Figure 16.5c is especially interesting, because it shows a case of “split phasing” in which one of the directions (east-west, in this case) has the two thru flows moving at different times. This is fairly rare, because certain rather special demand patterns have to be present to justify its use. Some jurisdictions simply do not allow split phasing, notably on state roads.

Figure 16.7 shows the signal displays for this particular *split phase* operation. For consistency with Figure 16.5c, the eastbound and westbound left turns are shown as permissive. However, in fact, they are running as protected because the opposing thru movement does not exist concurrently.

Figure 16.7: Signal Displays for the Split Phasing of Figure 16.5c

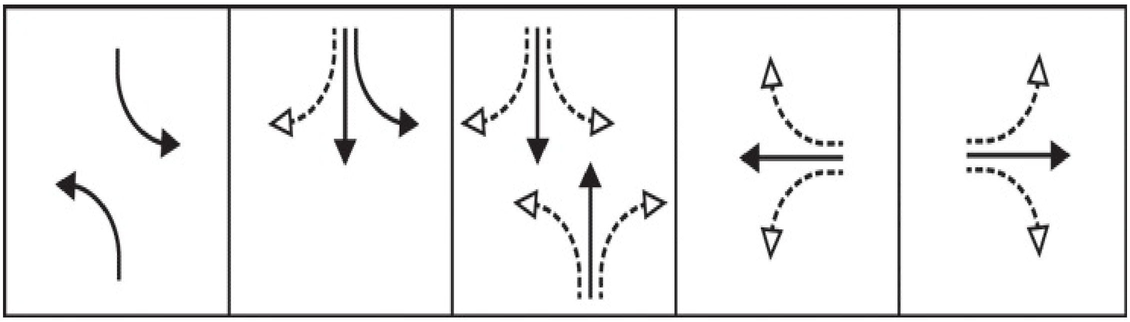


Figure 16.7: Full Alternative Text

Imagine an intersection with very heavy left turns in both the eastbound and westbound direction, comparable to the thru movements. While this is uncommon, it does occur. When one does the critical movement analysis as addressed in Chapter 23, the best solution may indeed be a split phase operation. Further, the demand pattern may well dictate two or more left turn lanes on at least one of the east–west approaches.

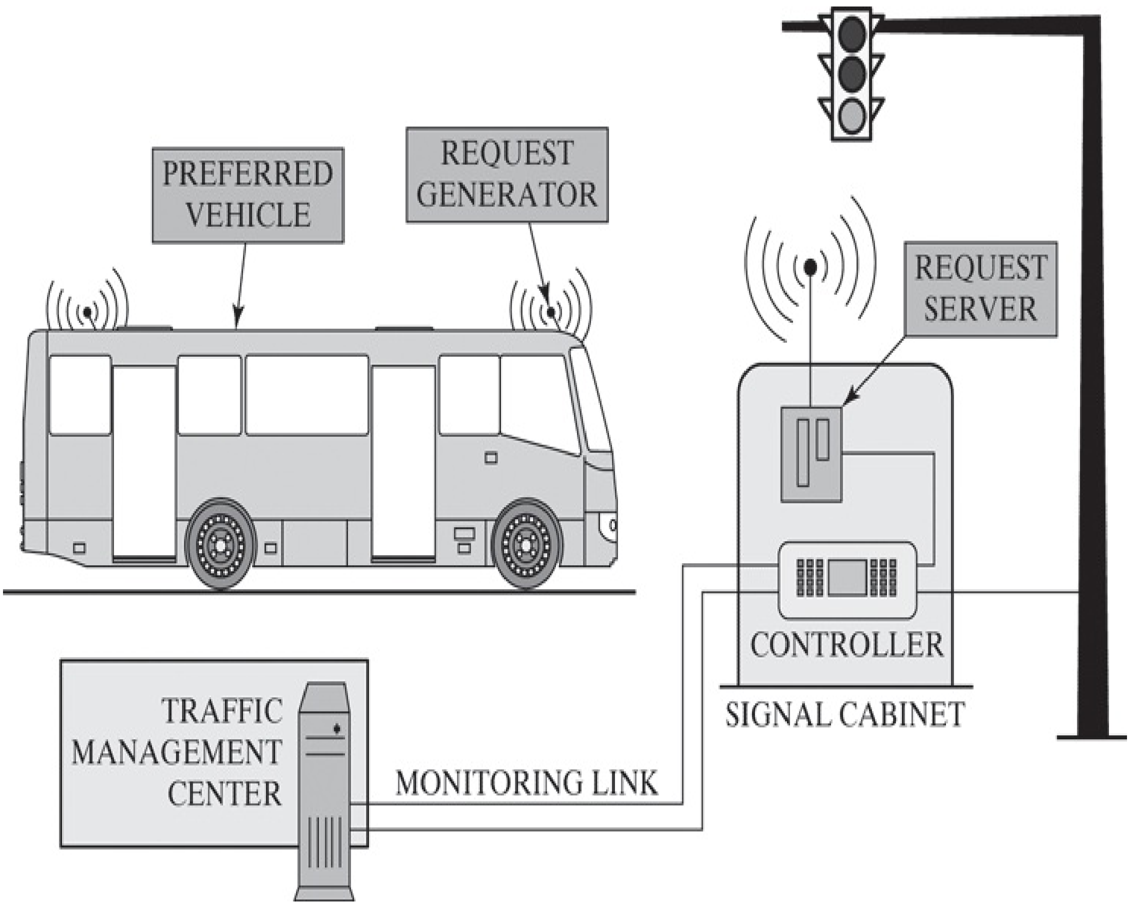
For additional material on ring-and-barrier diagrams for more complicated scenarios (including an additional leg into the intersection), see Ref. [1] and reflect upon the demand patterns that would lead to various implementations. Reference [1] also addresses cases in which traffic flowing in certain patterns through closely spaced intersections need special attention in designing the signal phasing.

There is also a very neat and simple case: For an intersection at which all north-south movements occur concurrently, followed by all east-west movements concurrently, the operation could be though of as a single ring system.

16.7 Preferential Treatment

Figure 16.8 shows an illustration of a transit vehicle being able to make a request to either initiate the green earlier than otherwise planned, or to extend the green. There is extensive literature on such systems, their merits, and results to date. It is quite important that the *design objective* be clearly defined, and that a relevant metric be used to assess the operation, in addition to other metrics that show effects on other components of the traffic stream. In some cases, the operation is justified in part by an overall system plan. For instance, the purpose of the preferential or priority treatment may be on-time arrival at an intermodal facility at which train or ferry departure schedules are fixed. In another case, the objective may be reducing the travel time of emergency service vehicles.

Figure 16.8: Communications Related to Bus Priority



(Source: Reprinted with permission from *Traffic Signal Manual: Second Edition*, National Cooperative Highway Research Program (NCHRP) Report 812, Transportation Research Board, © 2015 by the National Academy of Sciences, Courtesy of the National Academies Press, Washington, D.C.)

Figure 16.8: Full Alternative Text

Concepts of transit signal priority, bus priority systems, and others are addressed elsewhere in this text.

16.8 ASCT System Objectives

A topic that is sure to draw the attention of the reader, and particularly to influence the professional practice of many just entering the profession, is ASCT systems. These focus on using information from sensor systems (and perhaps other sources), data processing (and perhaps forecasting), to adaptively adjust signal timing to current needs. Reference [1] makes the distinction that "traffic responsive plan selection systems, which use predetermined timing plans" are a distinct category from ASCT.

In this classification, "traffic responsive plans" primarily draw on a library of predetermined plans, using sensor data (and perhaps history) to determine which library plan should be used in a given time period.

But there is a continuum, and complications. Some library-based plans restrict how often they may be changed. Others restrict some library plans to certain times of the day. Still others have no such restrictions. Some include fully adaptive elements at critical intersections or key locations, even with a library. Other systems are so extremely responsive that they can appear chaotic to drivers and to some traffic engineers, and need to be constrained more for a greater level of comfort.

Notwithstanding the above, there is a governing rule that applies to ASCT applications: Decide upon the true objective, design the system to achieve it, and then evaluate it according to the design objective while taking into account other impacts and metrics. Arriving at the "true objective" may be an iterative process and involve traffic professionals, planners, elected officials, and the public (as well as interested community and business groups). A challenge is always to have all involved remember the stated objectives.

Table 16.2 provides a good starting point for discussions: What does one hope to achieve with the system being considered? How does it fit into the schema of Table 16.2? How can one build on the starting point and arrive at consensus?

Table 16.2: ASCT System

Objectives

Objective	Description	Where Objective Is Typically Applied	How Objective Is Applied
Pipeline	Minimizing the number of stops experienced by a preferred movement on a critical route (from one end of a corridor to the other).	Linear arterial routes. Less commonly applied to progress heavy turning movements.	Maintaining large splits for the coordinated phases.
Smooth Flow	Prioritizing simultaneous bidirectional movements on a critical route in order to maximize the output.	Suburban arterials.	Maintenance of large splits for the coordinated phases.
Equitable Access	Providing sufficient arterial access to traffic generators along a corridor by placing increased emphasis on minor street demand.	Areas with significant left-turn and minor street demand (e.g., suburban retail shopping districts).	Appropriate application of split times to prevent long delays for minor movement (including pedestrians).
Manage Queues	Mitigation of queues and congestion caused by blocked intersections or movements. *Note: Most ASCT systems do not include features that specifically mitigate queues.*	Locations where queues block upstream intersections or movements.	Constraints on cycle and phase durations to ensure that large platoons progress at appropriate times. "Gating" to store vehicles at locations upstream of critical links may also be required.
Mitigate Oversaturation	Prevent, delay the onset, or limit the duration of oversaturated conditions. If oversaturated conditions do persist, clear overflow queues quickly.	Locations with oversaturated movement. *Note: ASCT systems can mitigate some oversaturation due to short-term capacity limitations, but are not a substitute for capacity improvements.*	Adjusting green time allocation for saturated phases.
Accommodate Long-Term Variability	Update signal timing more frequently than traditional systems and reduce deterioration of traffic operations over time.	Areas with changing traffic patterns, particularly growing communities.	While ASCT systems require some adjustments when major traffic pattern changes occur, they can adjust automatically to many changes.
Manage Events and Incidents	Manage surges in traffic (both planned and unplanned).	Areas with recurring planned special events (e.g., concerts, sporting events, or community activities).	Adjust signal timing parameters to match traffic surges over time. However, they generally cannot adjust rapidly to large surges.

(Source: Reprinted with permission from *Traffic Signal Manual: Second Edition*, National Cooperative Highway Research Program (NCHRP) Report 812, Transportation Research Board, © 2015 by the National Academy of Sciences, Courtesy of the National Academies Press, Washington, D.C.)

Table 16.2: Full Alternative Text

As an illustration, consider the following draft problem statement, intended to start a discussion rolling:

The primary area of concern is a set of north-south one-way arterials in an intensely developed urban area, around which one can draw a “box” that includes 15–20 intersections on each of the five arterials.

The objective is to keep traffic moving within the box, to enhance mobility. Reducing stops is more important than radical improvements in travel times through the box.

Another reality is that if traffic in the box “freezes up,” the consequent congestion will rapidly spread upstream along the arterials, first impacting the feeders to “the box” and then disrupting the east-west flows north and south of the box.

Some problems at the end of the chapter build upon this statement.

16.9 Sensors and Data Feeds

In prior editions of this text, this section would have been named “Detectors” and would have focused on the traditional devices—magnetic loops buried into the pavement, radar or ultrasonic devices mounted above ground, and related variations.

Today, the range of sensors that provide information to TMCs, to local controllers, and to the general public is much broader and their relative use is shifting continually. The range includes:

1. In-pavement sensors: The most traditional is the electromagnetic loop, cut into the pavement with the wiring leading back to the signal controller cabinet or to a roadside transmitter, operating on the change of inductance when vehicles enter the magnetic field generated by the loop. In some applications, the signal is based upon *presence*, so that the existence of a vehicle(s) in the field triggers an on/off (presence, absence) event. In other applications, the signal is based upon passage and vehicles might be counted. Magnetometers buried into the pavement are a variation on the concept, but with a smaller footprint and—more recently—with a wireless connection to a roadside device that can assemble the signals from a number of such devices. Another variation is “on pavement” devices that have low profiles and wireless connectivity, and do not need cuts into the pavement, reducing both installation cost and traffic disruption time.

2. Above-ground traditional sensors: A number of jurisdictions prefer to avoid in-pavement installations, because of digging by utilities (electric, gas, telecommunications), general repair work, and weather issues such as frost heave. In other cases, above-ground is simply easier or more practical, when they can be affixed to overpasses and such. However, roadside location on existing or special poles is common, as is placement on signal mast arms. Ultrasound, radar, and even infrared are all in use. Wireless or wired connections are both in use. Some devices have internal cameras that are used primarily for aiming the device upon installation, or checking on it in the field.

3. Cameras: Traditionally, cameras were used for qualitative assessment

of traffic, supplementing other detectors. The TMC operator would have the opportunity to see what is happening, rather than deducing it from the various traditional detectors. Advances in technology have led to using properly placed cameras to be used to employ software to count vehicles, estimate speeds, and detect occupancy.

4. Virtual detectors: A well placed camera can capture an image, and software allows the traffic staff to pin-point *multiple* “virtual” detectors throughout the field of view. These can be configured as point detectors or area coverage. The underlying mechanism is software that processes the image, segmenting it into portions designated by the traffic staff.

5. 360° cameras: This could be considered a variation on virtual detectors, but is singled out because cameras are now available that can be placed on or near the signal heads, view all directions, and provide quantitative data—including counts—for the several approaches. Similar devices can be used at roadside to collect traffic data that used to require field crews to be present.

6. ETC readers, bluetooth, and related: Electronic Toll Collection (ETC) tags can be read at any location, and are sometimes used for travel time estimation and relative volumes. They can also be used to estimate origin-destinations (O/D) and even routes within a network. For privacy reasons, it is important that any retained data be scrubbed of personal identifiable information (PII) and assigned randomly generated substitutions.

   Bluetooth devices can be read in the same way. Commercial trucking fleets, delivery services, buses, taxis, and others have GPS information available to fleet owners or relevant agencies. In principle, “breadcrumb” data on paths taken can be observed, with access to such data. Again, there are privacy concerns to be addressed. There are also proprietary issues to be addressed, because fleet operators are collecting data on their vehicles for competitive advantage in their own markets.

7. Smartphones and Smartphone Apps: Although smartphones could in principle be grouped into #6, the endemic presence of smartphones, the expanding versatility, and the existence of privately (and publicly) sourced “apps” put powerful tools in the hands of the general public.

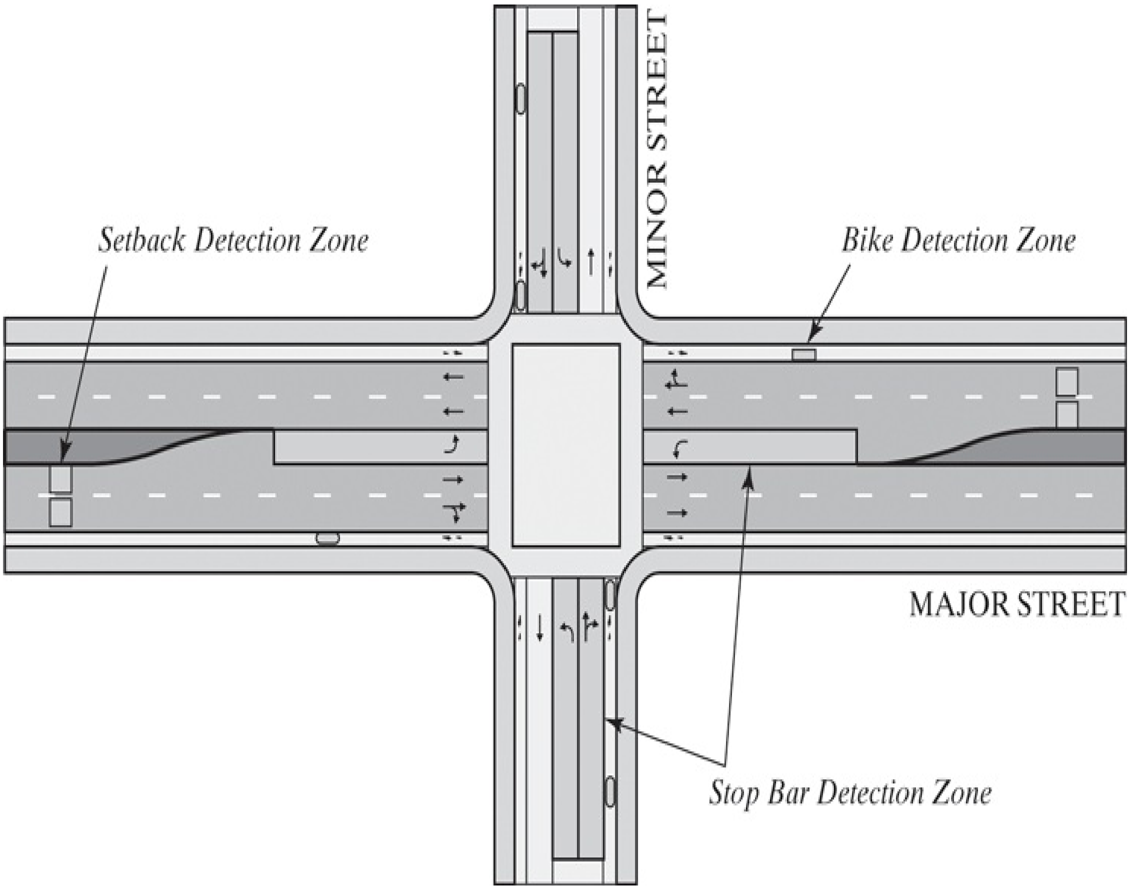
For instance, at the time of this writing, an app named Waze™ provides routing information based upon location data from its users (from which local travel times can be estimated), user reports of problems, and so forth. Another app allows the user to record routes taken, images of traffic, and time stamps—a complete digital record. Yet others alert drivers and pedestrians to potential conflicts.

8. Data feeds from other sources: The phrases “data fusion” and “big data” have become part of our vocabulary. Reflecting upon the previous items and considering how much more data is becoming available, we can only marvel at the opportunities (and challenges) to be encountered in coming years, with regard to sensing the state of the transportation system—passenger vehicles, trucks, bus transit, for-hire vehicles, pedestrians, and bicyclists.

A substantial issue in several of these “newer” technologies is how one meets the bandwidth demands of all that information. Perhaps the remedy lies not in substantially increased bandwidth, but rather in (a) being selective in what data or data summaries are sent back to the decision point, which might be the TMC, and/or (b) moving some of the decision-making to the local intersection, so that not everything has to be sent back to a TMC or equivalent.

Figure 16.9 shows an illustrative detector layout at an intersection. It was constructed for a discussion of traditional loop detectors, but the actual detectors could be a network of wireless magnetometers, virtual detectors placed onto camera images, and technologies not presently on the market. The long loops shown in some lanes could be implemented as a set of smaller loops, allowing queue lengths to be observed. Refer to Ref. [1] for further discussion.

Figure 16.9: Detectors Located at an Illustrative Intersection



(Source: Reprinted with permission from *Traffic Signal Manual: Second Edition*, National Cooperative Highway Research Program (NCHRP) Report 812, Transportation Research Board, © 2015 by the National Academy of Sciences, Courtesy of the National Academies Press, Washington, D.C.)

Figure 16.9: Full Alternative Text

16.10 Traffic Signal Display Hardware

The primary source of information is the MUTCD [2]. Chapter 4 of this text addresses some aspects of the standards by which communication with the driver and the travelling public take place—markings, signs, and signals.

Earlier in this chapter, Figure 16.1 showed the functional components of an intersection signal installation and Figure 16.2 showed a representative cabinet, opened to show its contents.

The *MUTCD* specifies the colors, displays, placement, number, size, and other requirements for the actual signal displays. NEMA standards address the electrical connections, the cabinet itself, and related subjects. NCTIP protocols are also relevant, notably for communications.

Figure 16.10 shows two signal heads in a field situation, one postmounted and the other mast-arm-mounted. The number of indications on a given head are dictated by the traffic movements built into the intersection signalization design, subject to the requirements of the *MUTCD*. There are other signal heads at the intersection, covering other approaches. For purpose of illustration, only two signal heads covering one approach are shown, and one for a different approach. In accord with the *MUTCD*, every approach will see two signal heads at this location.

Figure 16.10: Mast Arm and Post Mounted Signals Sharing a Common Structure

(Source: Photo courtesy of J Ulerio and R Roess.)

Figure 16.11 shows a more complicated (and wider) intersection, at which five signal heads are mounted on a single span wire.

Figure 16.11: An Illustration of Span Wire Signal Installation, with Five Signal Heads

(Source: Photo courtesy of J Ulerio and R Roess.)

The *MUTCD* itself can be downloaded (see Ref. [2] or the first page of this chapter), and course materials—and lectures—expanded to address in great detail the implementation of a signal timing plan in markings, signage, and signal hardware. For present purposes, a few notes call out some newer items:

- Solid “ball” indications (that is, not arrows) have historically been 8-inches or 12-inches in diameter. The current *MUTCD* however requires that all new installations use 12-inches. The smaller 8-inch ball indication may be used for some temporary purposes, and may be left in place on pre-existing installations until they must be replaced. However, in no case may a 12-inch green or yellow indication be used with an 8-inch red indication.

- The FYA was introduced into the current edition of the *MUTCD* for use as deemed appropriate, at locations with permissive left turns. The configuration of the signal head containing an FYA is specified precisely in the *MUTCD*. At this writing, there is a temporary authorization by FHWA for the use of an alternate signal head configuration, after application to FHWA.

- There are now nine traffic signal warrants.

The reader is reminded that the satisfaction of one or more traffic signal warrants does not mandate or justify the installation of a signal. Rather, it justifies an engineering study that will assess the need for a signal, using state of the practice and professional judgment. However, if no warrants are satisfied, signal installation is not considered. That is, satisfying one or more warrants is a necessary but not sufficient condition for a signal installation.

16.11 Traffic Signal Maintenance

Much attention is usually paid to instruction in the “creative” part of signal design, from the first study to initial design to detailed drawings to actual installation and even to producing a set of “as built” drawings of the intersection and its equipment. Attention then tends to shift to signal timing and retiming, and discussions of when signal-timing plans might become outdated.

However, this ignores a very important reality—field hardware has to be maintained, repaired, and replaced. It is vital to have a formal inventory of what equipment is where, and a regular program to inspect that it is working.[6]

[6]It was already noted in this chapter that one of the breakthrough advantages of early computer control systems was simply knowing when the equipment was not working. “Self-reports” by advanced controllers are now common.

The traffic signals exist to provide for public safety. Failed or missing equipment puts the public at risk. Some of the vital elements of a complete program are:

- A complete *inventory* of the installation, now usually done in a computer data base with a user-friendly interface. Such data bases can include digital photos of the field installation and access to the CAD drawings of the as-built plans;
- A regular plan of *inspection* of field installations, consistent with the resources available and the needs. In many cases, this now includes automated monitoring of detectors, signal controllers, and other equipment from a central location *as well as* a regular pattern of field inspections;
- Related *preventive maintenance* where appropriate, and logs of visits, actions, and results;
- A *reporting system* that can receive notifications of failed or damaged

(or missing) equipment and *act on the notification in a timely manner*. It is common for jurisdictions to have mandated response and repair times. Depending upon the nature of the problem, the requirement might vary from a few hours up to 24 hours, or perhaps longer (with interim measures taken). Failure to act not only puts the public at risk, but attaches liability to the agency responsible for the equipment; and

- A *training program* for staff and/or certification requirements for staff and/or contractors. Given the pace of technology, it is logical to expect that refresher training and new-equipment training be commonplace.

Even before getting to “maintenance” issues or even installation, *acceptance testing* of delivered equipment is common, in some cases including formal and intensive testing at an independent testing laboratory to ensure that specifications are met. Some jurisdictions will only purchase from preapproved lists of products that have met the jurisdiction’s requirements.

16.12 Closing Comments

To some extent, this chapter has blended a discussion of hardware with related standards, concepts built into the hardware (for instance, the ring-and-barrier diagram and the convention for labeling movements and phases), and typical installations. This is inevitable, because they are in fact intertwined.

Of necessity, one chapter cannot teach all that needs to be known. However, the references at the end of this chapter provide not just a roadmap but detailed knowledge of the state of the practice at the time of *this writing* (rather than publication date of this text). References [1] and [2] in particular are sizable, detailed (with fine multicolored illustrations), and free to download to one's own computer in PDF form.

References

- 1. *Traffic Signal Manual: Second Edition*, National Cooperative Highway Research Program (NCHRP) Report 812, Transportation Research Board, Washington, D.C., 2015. It can be downloaded at http://onlinepubs.trb.org/onlinepubs/nchrp/nchrp_rpt_812.pdf

- 2. *Manual of Uniform Traffic Control Devices*, Federal Highway Administration, Washington, D.C., 2009, as amended through May 2012. It can be downloaded at https://mutcd.fhwa.dot.gov/pdfs/2009r1r2/pdf_index.htm

- 3. *Highway Capacity Manual, 6th Edition: A Guide for Multimodal Mobility Analysis*, Transportation Research Board of the National Academies, Washington, D.C., 2016.

- 4. http://www.nema.org/About/pages/default.aspx

- 5. https://www.ntcip.org/

- 6. http://www.ncutcd.org/Pages/default.aspx

- 7. https://www.nema.org/Products/Pages/Transportation-Management-Systems-And-Associated-Control-Devices.aspx

- 8. Kittelson and Associates, *Traffic Signal Timing Manual*, 1st Edition, Federal Highway Administration, Washington, D.C., June 2008.

Problems

1. 16-1. In the United States, the *MUTCD* is not imposed by the Federal Government, but all states are expected to adopt Ref their own such manuals that are in substantial compliance with Ref. [2]. Does your own state[7] adopt Ref. [2] in its entirety, issue a supplement, or issue its own version? What are some key differences, if any?

   [7]If this text is being used outside the United States, change the question to “Does your own country or relevant subunit adopt ....”

2. 16-2. It is sometimes said that one of the advantages of ASCT systems and traffic responsive systems is that they adapt over longer periods of time to changing traffic patterns, thus keeping the traffic plan up to date with changing reality—thereby extending the time before a retiming is needed. Starting with Ref. [1] and its references, but including a web search if needed, comment on this assertion in a 3 to 5 page paper supported by such sources.

   Hint: There is divided opinion in the profession on this assertion, even if it appears logical.

3. 16-3. It is noted that “Further, the demand pattern may well dictate two or more left turn lanes on at least one of the east-west approaches. Focus on the words “two or more left turn lanes.”

   1. Are three dedicated left turn lanes (left turn only) allowed?

   2. If allowed, what are some representative installations and what is the experience with them? Particularly, does the third lane provide much added capacity to the left turn movement? Are there any special issues to consider?

   3. What guidelines exist, if any?

Do not base your answer on your own intuitive sense, but rather on a web search that will hopefully identify FHWA and/or various states, guidance and discussion.

4. 16-4. A draft problem statement is presented on page 351 to “start a discussion rolling.” Let’s do that.

   The suggestion is made that if the traffic in the box “freezes up” because it cannot process the demand, then the problem will spread upstream[8] and degrade operations there, perhaps most seriously.

   [8]Because some of the one-way arterials are southbound and others are northbound, “upstream” can include areas north *and* south of the box.

   1. Do you find this argument plausible and relevant? Why or why not?

   2. Assuming that you find it plausible and relevant (for the purpose of addressing this part “b” only, perhaps) does it follow that:

      - On a very good day, traffic flow in the box might be slow but manageable, while at the same time traffic flow is rather good both north and south of the box, *but*

      - On a bad day within the box, traffic control measures might be taken that improve operations within the box at the expense of traffic flow outside the box, *whereas*

      - On that same bad day, if no action were taken, traffic flow outside the box can be very adversely affected?

      Note that for residents, observers, and travelers outside the box, it might look like you are doing something good for others at their expense. How do you effectively make the case that “if we don’t take the action, it would have been even worse in your area”? How do you back up such a statement?

   3. Assuming that some traffic control measures are implemented when travel time on any given arterial segment with the box

degrades below a certain level, what metric(s) should be used to:

- Measure the benefit of action on traffic within the box, and
- Measure the benefit of action on traffic approaching and departing the box?

Hint: *Perhaps the real question is, how does one establish a baseline of the effects of* no action *on each of the cited groups?*

5. 16-5. One section of the chapter refers to a FYA, and the required signal head arrangement specified in the *MUTCD*.

    1. Go the Ref. [2], namely the *MUTCD*, and if necessary to the literature via web search and report on:

        - When a FYA can be considered. For instance, when a permissive left follows a protected left, or anytime there is a permissive left?
        - The exact displays to be used in the field

    2. The section also references an interim approval to use a different configuration in certain circumstances. Do a web search to find that information, and report on whether it is still interim, has been adopted, or has been discontinued. Report on results, if any are available.

6. 16-6. In your state (or relevant jurisdiction), what is the requirement on responding to a report of a signal failure? Does the mandated response time vary, depending upon the nature of the event?

Chapter 17 Fundamentals of Intersection Design and Layout

In Chapter 15, the selection of appropriate control measures for intersections was addressed. Whether signalized or unsignalized, the control measures implemented at an intersection must be synergistic with the design and layout of the intersection. In this chapter, an overview of several important intersection design features is provided. It is emphasized that this is only an overview, as the details of intersection design could be the subject of a textbook on its own.

The fundamentals treated here include techniques for determining the appropriate number and use of lanes at an intersection approach, channelization, right- and left-turn treatments, and special safety issues at intersections. There are a number of standard references for more detail on these and related subject areas, including the AASHTO *Policy on Geometric Design of Highways and Streets* [1], the *Manual on Uniform Traffic Control Devices* [2], the *Manual of Traffic Signal Design* [3], the *Traffic Detector Handbook* [4], and the *Highway Capacity Manual* [5].

Other aspects of intersection design are included in other chapters: Chapter 16 covers the placement of traffic signal hardware in an intersection, Chapter 25 covers unsignalized intersections, including STOP-controlled intersections and roundabouts, and Chapter 26 covers alternative (or distributed) intersection designs.

17.1 Intersection Design Objectives and Considerations

As in all aspects of traffic engineering, intersection design has two primary objectives: (1) to ensure safety for all users, including drivers, passengers, pedestrians, bicyclists, and others and (2) to promote efficient movement of all users (motorists, pedestrians, bicyclists, etc.) through the intersection. Achievement of both is not an easy task, as safety and efficiency are often competing rather than mutually reinforcing goals.

In developing an intersection design, AASHTO [1] recommends that the following elements be considered:

- Human factors
- Traffic considerations
- Physical elements
- Economic factors
- Functional intersection area

Human factors must be taken into account. Thus, intersection designs should accommodate reasonable approach speeds, user expectancy, decision and reaction times, and other user characteristics. Design should, for example, reinforce natural movement paths and trajectories, unless doing so presents a particular hazard.

Traffic considerations include provision of appropriate capacity for all user demands; the distribution of vehicle types and turning movements; approach speeds; and special requirements for transit vehicles, pedestrians, and bicyclists.

Physical elements include the nature of abutting properties, particularly traffic movements generated by these properties (parking, pedestrians, driveway movements, etc.). They also include the intersection angle, existence and location of traffic control devices, sight distances, and specific geometric characteristics, such as curb radii.

Economic factors include the cost of improvements (construction, operation, and maintenance), the effects of improvements on the value of abutting properties (whether used by the expanded right-of-way or not), and the effect of improvements on energy consumption.

Finally, intersection design must encompass the full functional intersection area. The operational intersection area includes approach areas that fully encompass deceleration and acceleration zones as well as queuing areas. The latter are particularly critical at signalized intersections.

17.2 A Basic Starting Point: Sizing the Intersection

One of the most critical aspects of intersection design is the determination of the number of lanes needed on each approach. This is not an exact science, as the result is affected by the type of control at the intersection, parking conditions and needs, availability of right-of-way, and a number of other factors that are not always directly under the control of the traffic engineer. Further, considerations of capacity, safety, and efficiency all influence the desirable number of lanes. As is the case in most design exercises, there is no one correct answer, and many alternatives may be available that provide for acceptable safety and operation.

17.2.1 Unsignalized Intersections

Unsignalized intersections may be operated under basic rules of the road (no control devices other than warning and guide signs), or under STOP or YIELD control. Roundabouts, which are a form of unsignalized intersection, are covered in Chapter 25.

When totally uncontrolled, intersection traffic volumes are generally light, and there is rarely a clear “major” street with significant volumes involved. In such cases, intersection areas do not often require more lanes than on the approaching roadway. Additional turning lanes are rarely provided. Where high speeds and/or visibility problems exist, channelization may be used in conjunction with warning signs to improve safety.

The conditions under which two-way (or one-way at a T-intersection or intersection of one-way roadways) STOP or YIELD control are appropriate are treated in Chapter 15. The existence of STOP- or YIELD-controlled approach(es), however, adds some new considerations into the design process:

- Should left-turn lanes be provided on the major street?
- Should right-turn lanes be provided on the major street?

- Should a right-turn lane (and/or left-turn lane) be provided on minor approaches?
- How many basic lanes does each minor approach require?

Most of these issues involve capacity considerations. For convenience, however, some general guidelines are presented herein.

When left turns are made from a mixed lane on the major street, there is the potential for unnecessary delay to through vehicles that must wait while left-turners find a gap in the opposing major-street traffic. The impact of major-street left turns on delay to all major-street approach traffic becomes noticeable when left turns exceed 150 vehs/h. This may be used as a general guideline indicating the probable need for a major-street left-turn lane, although a value as low as 100 vehs/h could be justified.

Right-turning vehicles from the major street do not have a major impact on the operation of STOP- or YIELD-controlled intersections. While they do not technically conflict with minor-street movements when they are made from shared lanes, they may impede some minor-street movements when drivers do not clearly signal that they are turning or approach the intersection at high speed. When major-street right turns are made from an exclusive lane, their intent to turn is more obvious to minor-street drivers. Right-turn lanes for major-street vehicles can be easily provided where on-street parking is permitted. In such situations, parking may be prohibited for 100 to 200 ft from the STOP line, thus creating a short right-turn lane.

Most STOP-controlled approaches have a single lane shared by all minor-street movements. Occasionally, two lanes are provided. Any approach with sufficient demand to require three lanes is probably inappropriate for STOP control. Approximate guidelines for the number of lanes required may be developed from the unsignalized intersection analysis methodology of the *Highway Capacity Manual*. Table 17.1 shows various combinations of minor-approach demand vs. total crossing traffic on the major street, along with guidelines as to whether one or two lanes would be needed. They are based on assumptions that (1) all major-street traffic is through traffic, (2) all minor-approach traffic is through traffic, and (3) various impedances and other non-ideal characteristics reduce the capacity of a lane to about 80% of its original value.

Table 17.1: Guidelines for Number of Lanes at STOP-Controlled Approaches[1]

Total Volume on Minor Approach (veh/h)	Total Volume on Major Street (veh/h)			
	500	1,000	1,500	2,000
100	1 lane	1 lane	1 lane	2 lanes
200	1 lane	1 lane	2 lanes	NA
300	1 lane	2 lanes	2 lanes	NA
400	1 lane	2 lanes	NA	NA
500	2 lanes	NA	NA	NA
600	2 lanes	NA	NA	NA
700	2 lanes	NA	NA	NA
800	2 lanes	NA	NA	NA

[1]Not including multiway STOP-controlled intersections.

NA = STOP control probably not appropriate for these volumes.

Table 17.1: Full Alternative Text

The other issue for consideration on minor STOP-controlled approaches is whether or not a right-turning lane should be provided. Because the right-turn movement at a STOP-controlled approach is much more efficient than crossing and left-turn movements, better operation can usually be accomplished by providing a right-turn lane. This is often as simple as banning parking within 200 ft of the STOP line, and it prevents right-turning drivers from being stuck in a queue when they could easily be executing their movements. Where a significant proportion of the minor-approach traffic is turning right (>20%), provision of a right-turning lane should always be considered.

Note that the lane criteria of Table 17.1 are approximate. Any finalized design should be subjected to detailed analysis using the appropriate procedures of the HCM 2016.

Consider the following example: a two-lane major roadway carries a volume of 800 veh/h, of which 10% turn left and 5% turn right at a local street. Both approaches on the local street are STOP-controlled and carry 150 veh/h, with 50 turning left and 50 turning right. Suggest an appropriate design for the intersection.

Given the relatively low volume of left turns (80/h) and right turns (40/h) on the major street, neither left- nor right-turn lanes would be required, although they could be provided if space is available. From Table 17.1, it appears that one lane would be sufficient for each of the minor-street approaches. The relatively heavy percentage of right turns (33%), however, suggests that a right-turn lane on each minor approach would be useful.

17.2.2 Signalized Intersections

Approximating the required size and layout of a signalized intersection involves many factors, including the demands on each lane group, the number of signal phases, and the signal cycle length.

Determining the appropriate number of lanes for each approach and lane group is not a simple design task. Like so many design tasks, there is no absolutely unique result, and many different combinations of physical design and signal timing can provide for a safe and efficient intersection.

The primary control on number of lanes is the *maximum sum of critical-lane volumes* that the intersection can support. This concept is more thoroughly discussed and illustrated in Chapter 18. The concept involves finding the single lane during a signal cycle that carries the most intense traffic—which means that it would be the one that consumes the most green time of all movements to process its demand. Each signal phase will have a critical lane volume, and the cycle length of the signal is set to accommodate the sum of these critical volumes for each phase in the signal plan. The equation governing the maximum sum of critical lane volumes is:

$$V_c = \frac{1}{h}\left[3{,}600 - N t_L \left(\frac{3{,}600}{C}\right)\right] \quad [17\text{-}1]$$

where:

V_c = maximum sum of critical-lane volumes, veh/h, h =
average saturation headway for prevailing conditions on the lane group or ap
N = number of phases in the cycle, t_L = lost time, s/phase, and C =
cycle length, s.

Table 17.2 gives approximate maximum sums of critical lane volumes for typical prevailing conditions. An average headway of 2.6 s/veh is used, along with a typical lost time per phase of 4.0 s. Maximum sums are tabulated for a number of combinations of *N* and *C*.

Table 17.2: Maximum Sums of Critical Lane Volumes for a Typical Signalized Intersection

Cycle Length (s)	No. of Phases		
	2	3	4
30	1,015	831	646
40	1,108	969	831
50	1,163	1,052	942
60	1,200	1,108	1,015
70	1,226	1,147	1,068
80	1,246	1,177	1,108
90	1,262	1,200	1,138
100	1,274	1,218	1,163
110	1,284	1,234	1,183
120	1,292	1,246	1,200

Table 17.2: Full Alternative Text

Consider the case of an intersection between two major arterials. Arterial 1 has a peak directional volume of 900 veh/h; Arterial 2 has a peak directional volume of 1,100 veh/h. Turning volumes are light, and a two-phase signal is anticipated. As a preliminary estimate, what number of lanes is needed to accommodate these volumes, and what range of cycle lengths might be appropriate?

From Table 17.2, the range of maximum sums of critical lane volumes is between 1,015 veh/h for a 30-s cycle length and 1,292 veh/h for a 120-s cycle length. The two critical volumes are given as 900 veh/h and 1,100 veh/h. If only one lane is provided for each, then the sum of critical-lane volumes is 900+1,100=2,000 veh/h, well outside the range of maximum values for reasonable cycle lengths. Table 17.3 shows a number of reasonable scenarios for the number of lanes on each critical approach along with the resulting sum of critical-lane volumes.

Table 17.3: Sum of Critical Lane Volumes (veh/h) for

Various Scenarios Sample Problem

No. of Lanes on Arterial 2	Critical Lane Volume for Arterial (veh/h)	No. of Lanes on Arterial[1]		
		1	2	3
		900/1 = 900	900/2 = 450	900/3 = 300
1	1,100/1 = 1,100	*2,000*	*1,550*	*1,400*
2	1,100/2 = 550	*1,450*	*1,000*[1]	*850*[1]
3	1,100/3 = 367	*1,267*[1]	*817*[1]	*667*[1]

[1]Acceptable lane plan with V_c acceptable at some cycle length.

Table 17.3: Full Alternative Text

With one lane on Arterial 1 and 3 lanes on Arterial 2, the sum of critical-lane volumes is 1,267 veh/h. From Table 17.2, this would be a workable solution with a cycle length over 100 s. With two lanes on each arterial, the sum of critical-lane volumes is 1,000 veh/h. This situation would be workable at any cycle length between 30 and 120 s. All other potentially workable scenarios in Table 17.3 could accommodate any cycle length between 30 and 120 s as well.

This type of analysis does not yield a final design or cycle length, as it is approximate. But it does give the traffic engineer a basic idea of where to start. In this case, providing two lanes on each arterial in the peak direction appears to be a reasonable solution. As peaks tend to be reciprocal (what goes one way in the morning comes back the opposite way in the evening), two lanes would also be provided for the off-peak directions on each arterial as well.

The signal timing should then be developed using the methodologies of Chapters 19 and 20. The final design and timing should then be subjected to analysis using the *Highway Capacity Manual* or some other appropriate analysis technique (see Chapters 22 and 23).

The number of anticipated phases is, of course, critical to a general analysis of this type. Suggested criteria for determining when protected left-turn phases are needed are given in Chapter 19. Because there is a critical-lane volume for *each* signal phase, a four-phase signal involves four critical-lane volumes, for example.

Exclusive left-turn lanes must be provided whenever a fully protected left-turn phase is used and is highly desirable when compound left-turn phasing (protected + permitted or vice-versa) is used.

17.3 Intersection Channelization

17.3.1 General Principles

Channelization can be provided through the use of painted markings or by installation of raised channelizing islands. The AASHTO *Policy on Geometric Design of Highways and Streets* [1] gives a number of reasons for considering channelization at an intersection:

- Vehicle paths may be confined so that no more than two paths cross at any one point.
- The angles at which merging, diverging, or weaving movements occur may be controlled.
- Pavement area may be reduced, decreasing the tendency to wander and narrowing the area of conflict between vehicle paths.
- Clearer indications of proper vehicle paths may be provided.
- Predominant movements may be given priority.
- Areas for pedestrian refuge may be provided.
- Separate storage lanes may be provided to permit turning vehicles to wait clear of through-traffic lanes.
- Space may be provided for the mounting of traffic control devices in more visible locations.
- Prohibited turns may be physically controlled.
- Vehicle speeds may be somewhat reduced.

The decision to channelize an intersection depends upon a number of factors, including the existence of sufficient right-of-way to accommodate an effective design. Factors such as terrain, visibility, demand, and cost also enter into the decision. Channelization supplements other control measures but can sometimes be used to simplify other elements of control.

17.3.2 Some Examples

It is difficult to discuss channelization in the abstract. A selection of examples illustrates the implementation of the principles noted previously.

Figure 17.1 shows the intersection of a major street (E–W) with a minor crossroad (N–S). A median island is provided on the major street. Partial channelization is provided for the SB right turn, and a left-turn lane is provided for the EB left turn. The two channelized turns are reciprocal, and the design reflects a situation in which these two turning movements are significant. The design illustrated minimizes the conflict between SB right turns and other movements and provides a storage lane for EB left turns, removing the conflict with EB through movements. The lack of any channelization for other turning movements suggests that they have light demand. The design does not provide for a great deal of pedestrian refuge, except for the wide median on the east leg of the intersection. This suggests that pedestrian volumes are relatively low at this location; if this is so, the crosswalk markings are optional. The channelization at this intersection would be appropriate for either an unsignalized or a signalized intersection.

Figure 17.1: A Four-Leg Intersection with Partial Channelization for SB-EB and EB-SB Movements

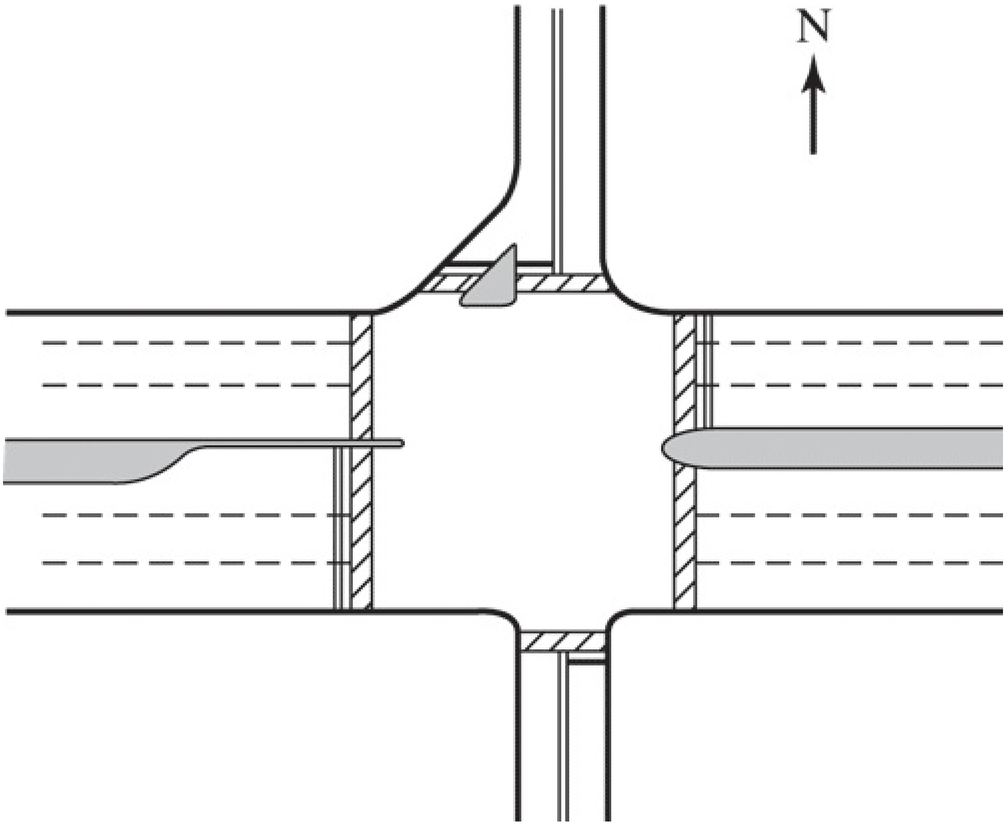


Figure 17.1: Full Alternative Text

Figure 17.2 shows a four-leg intersection with similar turning movements as in Figure 17.1. In this case, however, the SB-EB and EB-SB movements are far heavier, and require a more dramatic treatment. Here channelization is used to create two additional intersections to handle these dominant turns. Conflicts between the various turning movements are minimized in this design.

Figure 17.2: A Four-Leg Intersection Channelization for Major SB-EB and EB-SB Movements

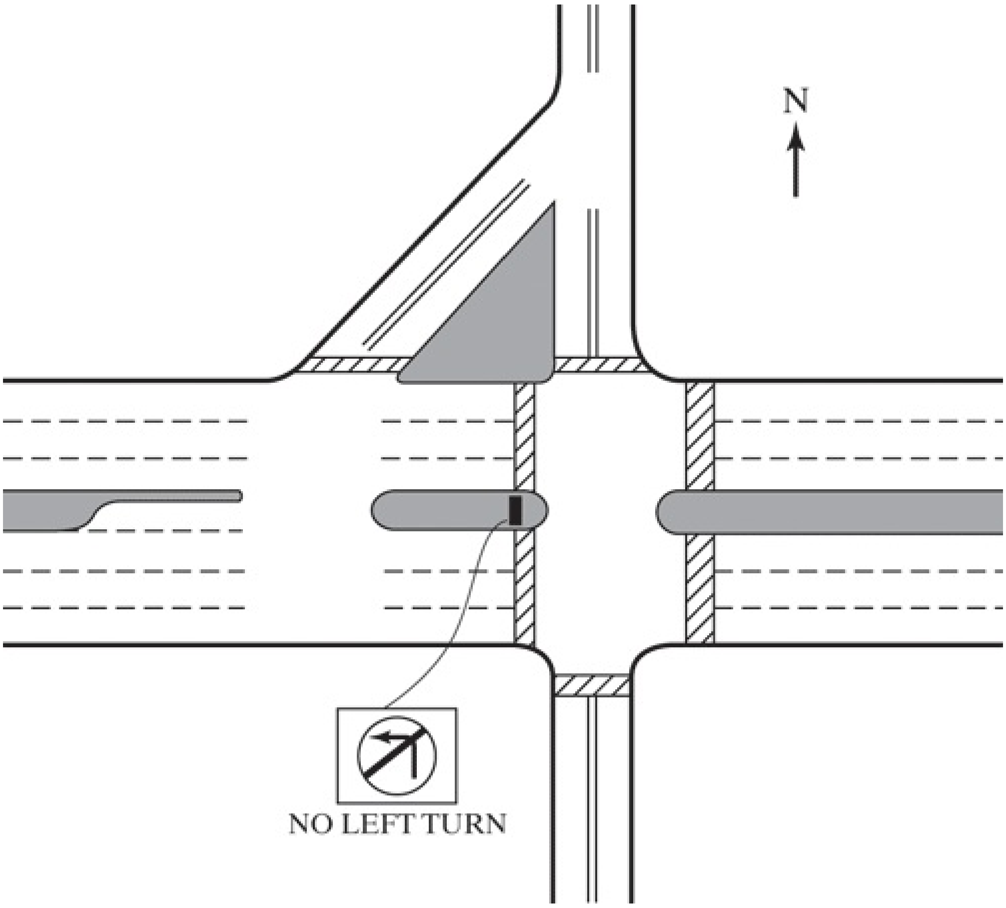


Figure 17.2: Full Alternative Text

Figure 17.3 is a similar four-leg intersection with far greater use of channelization. All right turns are channelized, and both major-street left-turning movements have an exclusive left-turn lane. This design addresses a situation in which turning movements are more dominant. Pedestrian refuge is provided only on the right-turn channelizing islands and this may be limited by the physical size of the islands. Again, the channelization scheme is appropriate for either signalized or unsignalized control.

Figure 17.3: A Four-Leg Intersection with Full Channelization of Right Turns

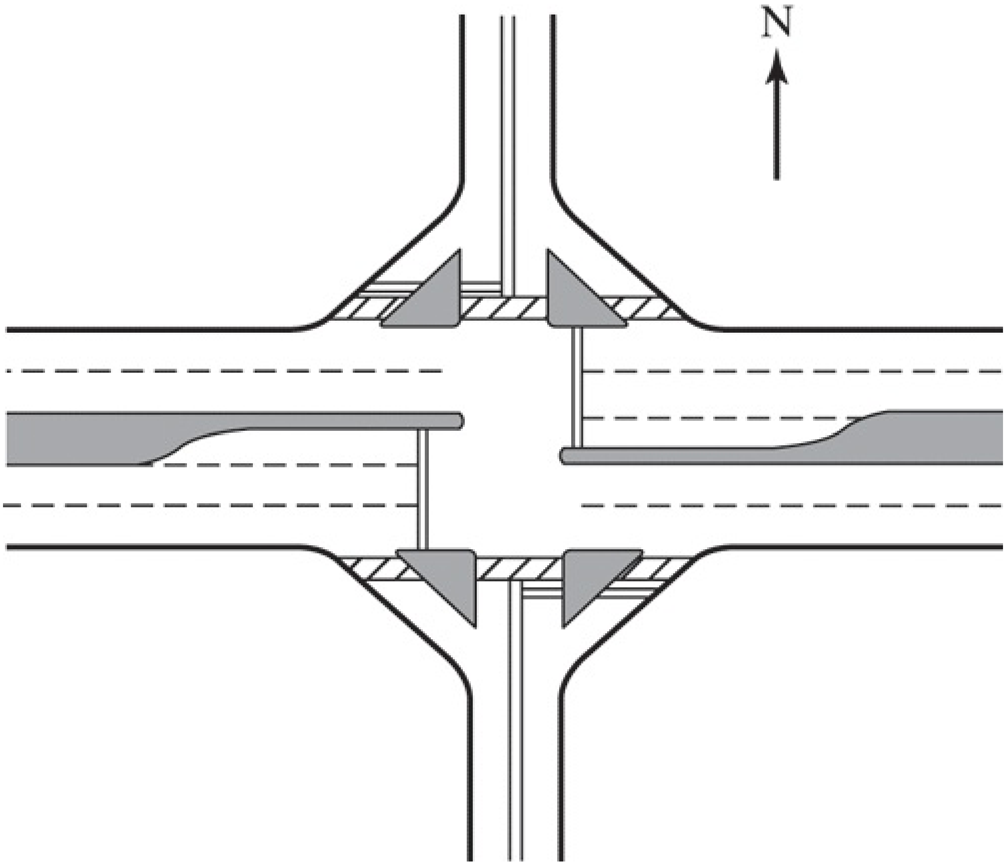


Figure 17.3: Full Alternative Text

Channelization is also a major design feature used in many new forms of alternative or distributed intersections. These are covered in some detail in Chapter 26.

17.3.3 Channelizing Right Turns

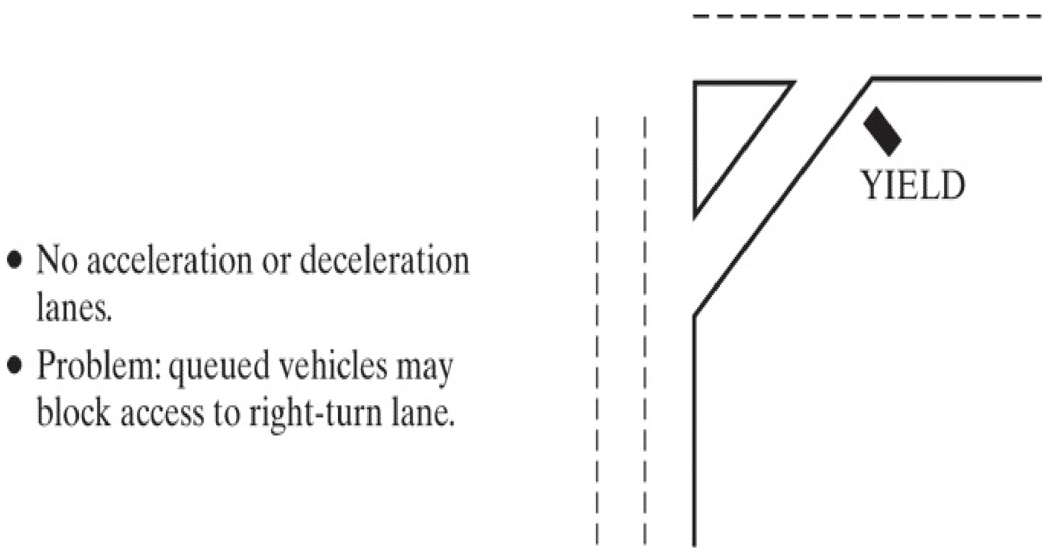
When space is available, it is virtually always desirable to provide a channelized path for right-turning vehicles. This is especially true at signalized intersections where such channelization accomplishes two major benefits:

- Where “right-turn on red” regulations are in effect, channelized right turns minimize the probability of a right-turning vehicle or vehicles being stuck behind a through vehicle in a shared lane.
- Where channelized, right turns can effectively be removed from the signalization design, as they would, in most cases, be controlled by a YIELD sign and would be permitted to move continuously.

The accomplishment of these benefits, however, depends upon some of the details of the channelization design.

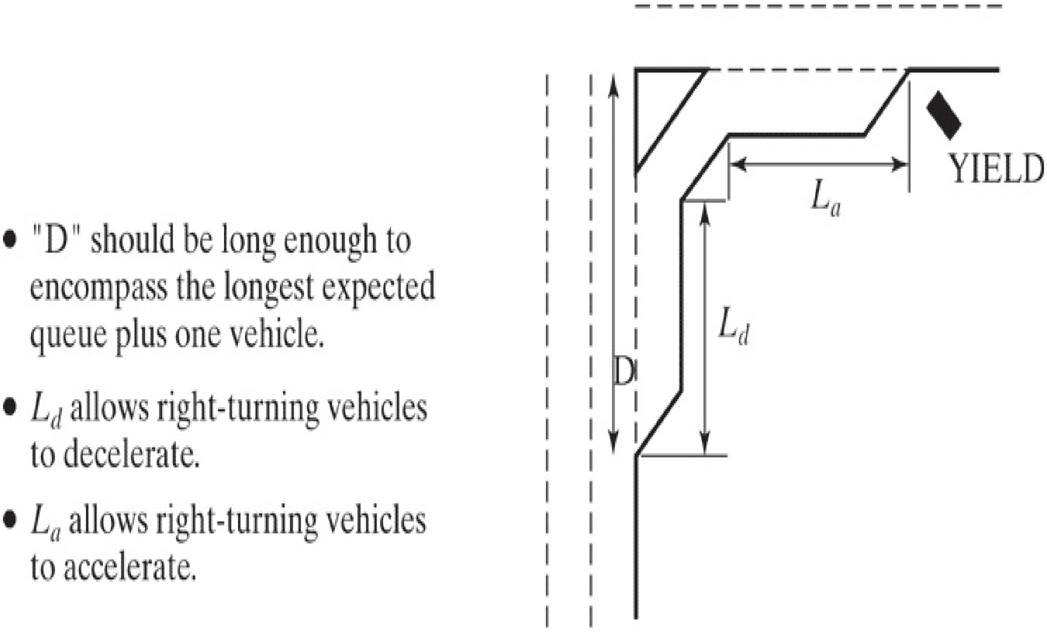
Figure 17.4 shows three different schemes for providing channelized right turns at an intersection. In Figure 17.4 (a), a simple channelizing triangle is provided. This design has limited benefits for two reasons: (1) through vehicles in the right lane may queue during the “red” signal phase, blocking access to the channelized right-turn lane and (2) high right-turn volumes may limit the utility of the right-hand lane to through vehicles during “green” phases.

Figure 17.4: Alternatives for Channelizing Right Turns



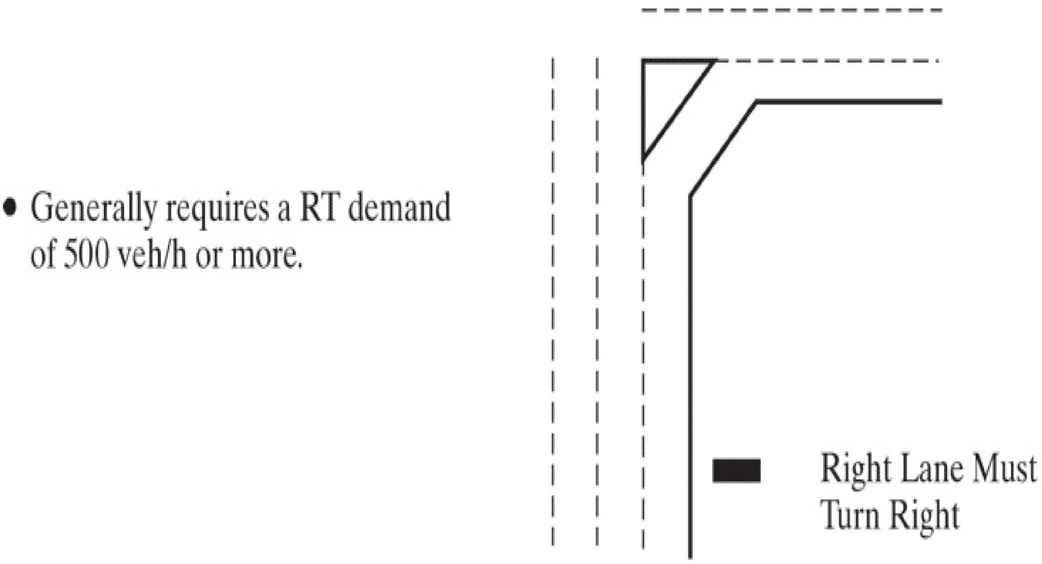
(a) Simple Channelized Right Turn

[17.3-4 Full Alternative Text]



(b) Channelized Right Turn with Acceleration and Deceleration Lanes

[17.3-4 Full Alternative Text]



(c) Channelized Right Turn with Lane Drop and Lane Addition

17.3-4 Full Alternative Text

In the second design (shown in Figure 17.4(b)), acceleration and deceleration lanes are added for the channelized right turn. If the lengths of the acceleration and deceleration lanes are sufficient, this design can avoid the problem of queues blocking access to the channelized right turn.

In the third design (Figure 17.4(c)), a very heavy right-turn movement can run continuously. A lane drop on the approach leg and a lane addition to the departure leg provide a continuous lane and an unopposed path for right-turning vehicles. This design requires unique situations in which the lane drop and lane addition are appropriate for the arterials involved. To be effective, the lane addition on the departure leg cannot be removed too close to the intersection. It should be carried for at least several thousand feet before it is dropped, if necessary.

Right-turn channelization can simplify intersection operations, particularly where the movement is significant. It can also make signalization more efficient, as channelized right turns, controlled by a YIELD sign, do not require green time to be served.

17.4 Special Situations at Intersections

This section deals with four unique intersection situations that require attention: (1) intersections with junction angles less than 60° or more than 120°, (2) T-intersections, (3) offset intersections, and (4) special treatments for heavy left-turn movements.

17.4.1 Intersections at Skewed Angles

Intersections, both signalized and unsignalized, work best when the angle of the intersection is 90°. Sight distances are easier to define, and drivers tend to expect intersections at right angles. Nevertheless, there are many situations in which the intersection angle is not 90°. Such angles may present special challenges to the traffic engineer, particularly when they are less than 60° or more than 120°. These occur relatively infrequently. Drivers are generally less familiar with their special characteristics, particularly vis-à-vis sight lines and distances.

Skewed-angled intersections are particularly hazardous when uncontrolled and combined with high intersection-approach speeds. Such cases generally occur in rural areas and involve primary state and/or county routes. The situation illustrated in Figure 17.5 provides an example.

Figure 17.5: A Skewed-Angle Rural Intersection

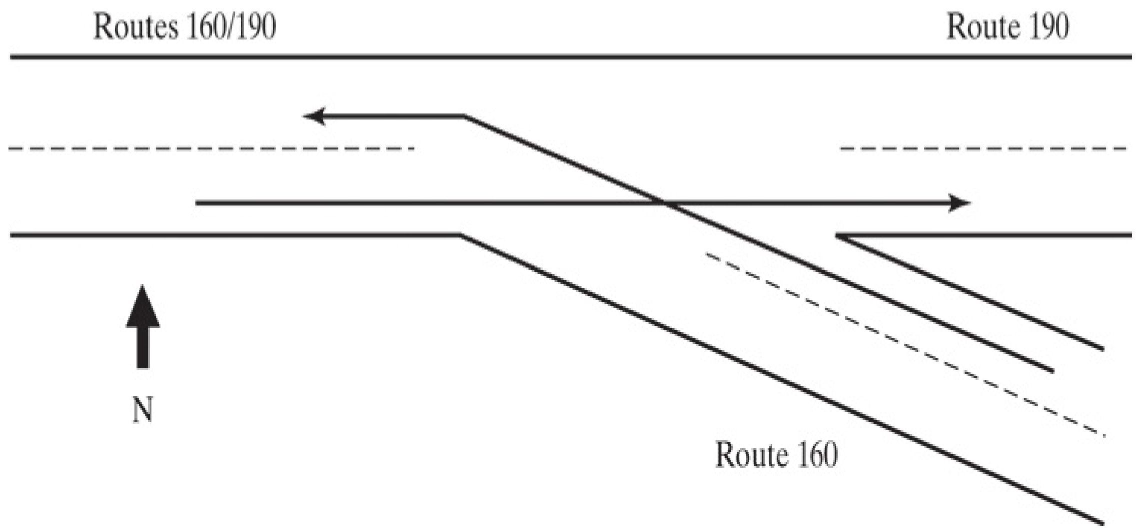


Figure 17.5: Full Alternative Text

The example is a rural junction of two-lane, high-speed arterials, Routes 160 and 190. Given relatively gentle terrain, low volumes, and the rural setting, speed limits of 50 mi/h are in effect on both facilities. Figure 17.5 also illustrates the two movements representing a particular hazard. The conflict between the WB movement on Route 160 and the EB movement on Route 190 is a significant safety hazard. At the junction shown, both roadways have similar designs. Thus, there is no visual cue to the driver indicating which route has precedence or right-of-way. Given that signalization is rarely justifiable in low-volume rural settings, other means must be considered to improve the safety of operations at the intersection.

The most direct means of improving the situation is to change the alignment of the intersection, making it clear which of the routes has the right-of-way. Figure 17.6 illustrates the two possible realignments. In the first case, Route 190 is given clear preference; vehicles arriving or departing on the east leg of Route 160 must go through a 90° intersection to complete their maneuver. In the second case, Route 160 is dominant, and those arriving or departing on the east leg of Route 190 go through the 90° intersection. In either case, the 90° intersection would be controlled using a STOP sign to clearly designate right-of-way.

Figure 17.6: Potential Realignment for Rural

Intersection

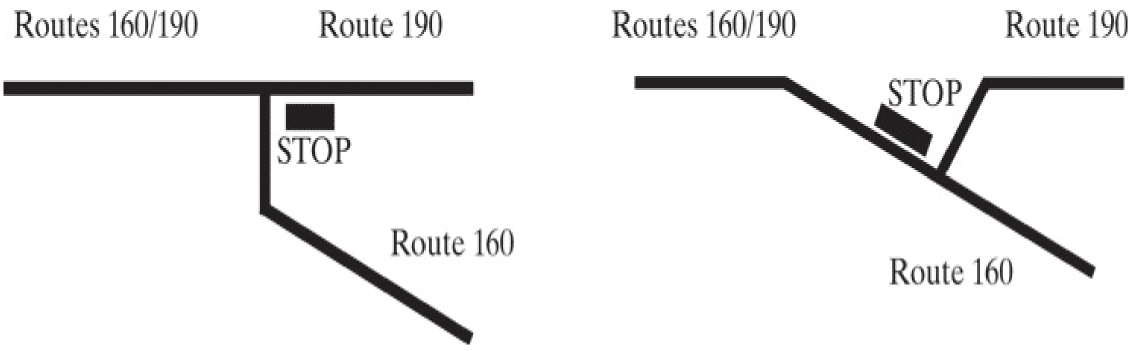


Figure 17.6: Full Alternative Text

While basic realignment is the best solution for high-speed odd-angle intersections, it requires that right-of-way be available to implement the change. Even in a rural setting, sufficient right-of-way to realign the intersection may not always be available. Other solutions can also be considered. Channelization can be used to better define the intersection movements, and control devices can be used to designate right-of-way. Figure 17.7 shows another potential design that requires less right-of-way than full realignment.

Figure 17.7: An Alternative Solution Using Channelization

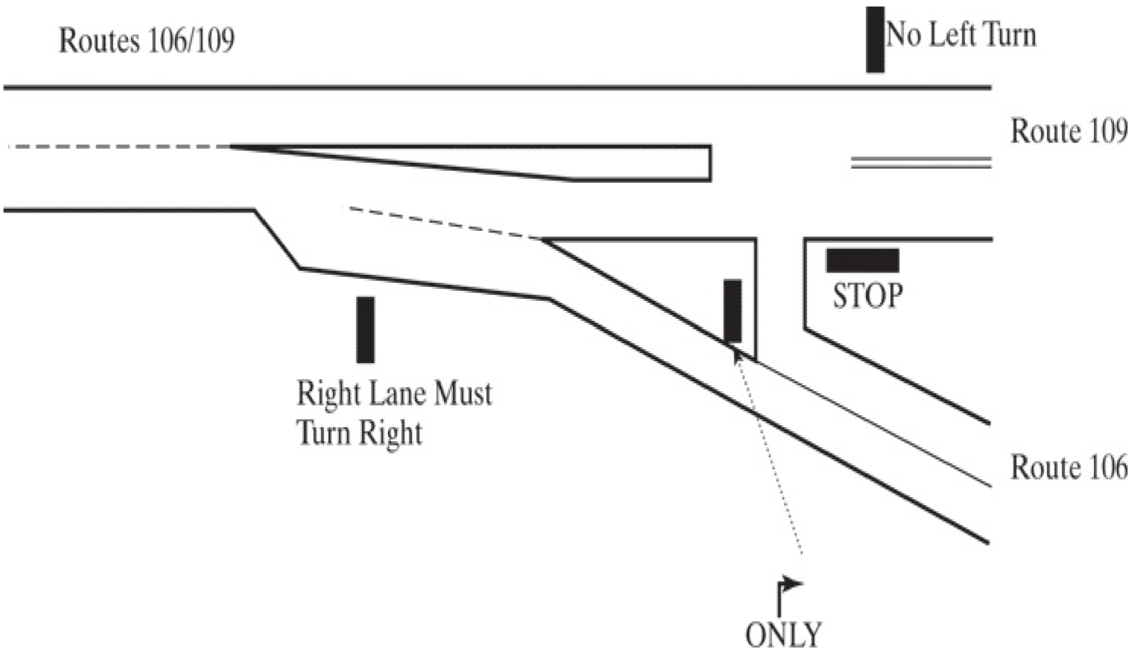


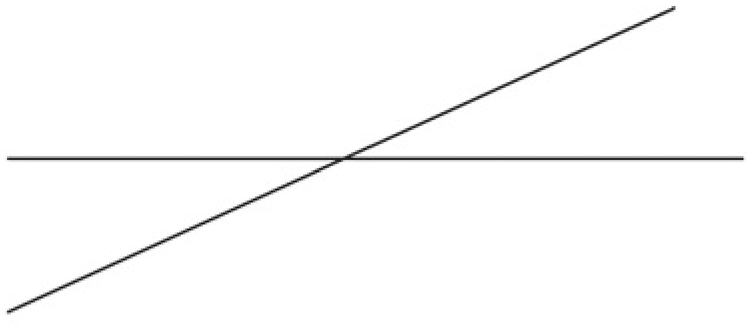
Figure 17.7: Full Alternative Text

In this case, only the WB movement on Route 106 was realigned. While this would still require some right-of-way, the amount needed is substantially less than for full realignment. Additional channelization is provided to separate EB movements on Routes 106 and 109. In addition to the regulatory signs indicated in Figure 17.7, warning and directional guide signs would be placed on all approaches to the intersection. In this solution, the WB left turn from Route 109 must be prohibited; an alternative route would have to be provided and appropriate guide signs designed and placed.

The junction illustrated is, in essence, a three-leg intersection. Skewed-angle four-leg intersections also occur in rural, suburban, and urban settings and present similar problems. Again, total realignment of such intersections is the most desirable solution. Figure 17.8 shows an intersection and the potential realignments that would eliminate the odd-angle junction. Where a four-leg intersection is involved, however, the realignment solution creates two separate intersections. Depending upon volumes and the general traffic environment of the intersection, the realignments proposed in Figure 17.8 could result in signalized or unsignalized intersections.

Figure 17.8: Realignment of

Four-Leg Odd-Angle Intersections

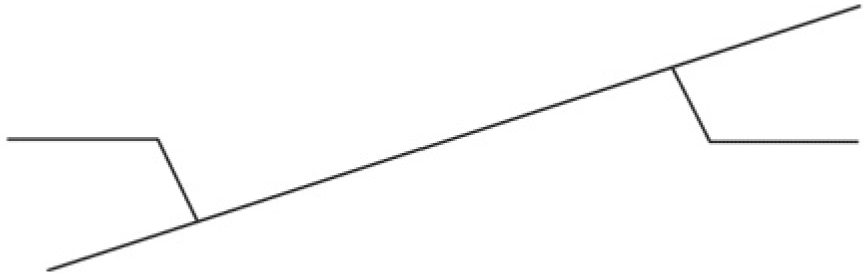


(a) Original Odd-Angle Intersection

17.4-4 Full Alternative Text

(b) Realignment #1

17.4-4 Full Alternative Text



(c) Realignment #2

17.4-4 Full Alternative Text

In urban and suburban settings, where right-of-way is a significant impediment to realigning intersection, signalization of the odd-angle intersection can be combined with channelization to achieve safe and efficient operations. Channelized right turns would be provided for acute-angle turns, and left-turn lanes (and signalization) would be provided as needed.

In extreme cases, where volumes and approach speeds present hazards that cannot be ameliorated through normal traffic engineering measures, consideration may be given to providing a full or partial interchange with the two main roadways grade-separated. Providing grade-separation would also involve some expansion of the traveled way, and overpasses in some suburban and urban surroundings may involve visual pollution and/or other negative environmental impacts.

17.4.2 T-Intersections: Opportunities for Creativity

In many ways, T-intersections are far simpler than traditional four-leg intersections. The typical four-leg intersection contains twelve vehicular movements and four crossing pedestrian movements. At a T-intersection, only six vehicular movements exist and there are only three crossing pedestrian movements. These are illustrated in Figure 17.9.

Figure 17.9: Simple T-Intersection Illustrated

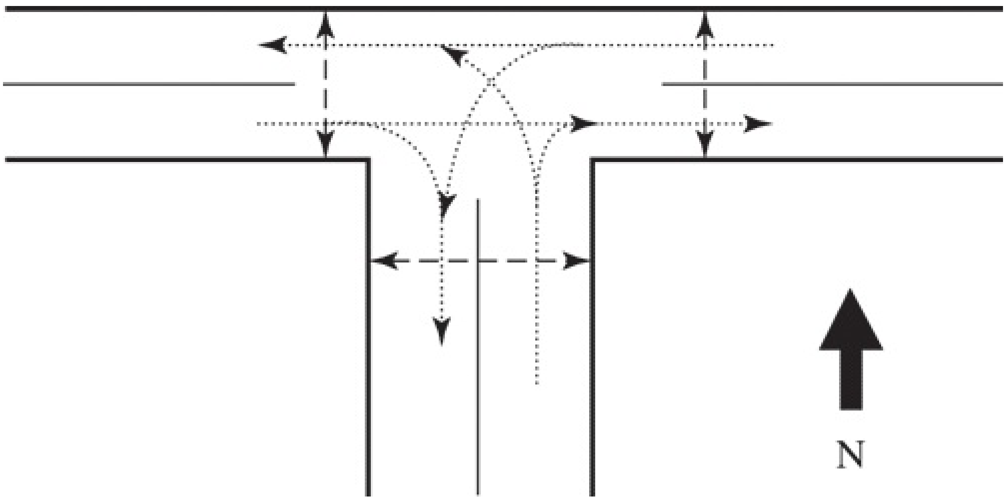


Figure 17.9: Full Alternative Text

Note that in the set of T-intersection vehicular movements, there is only one opposed left turn—the WB left-turn movement in this case. Because of this, conflicts are easier to manage, and signalization, when necessary, is easier to address.

Control options include all generally applicable alternatives for intersection control:

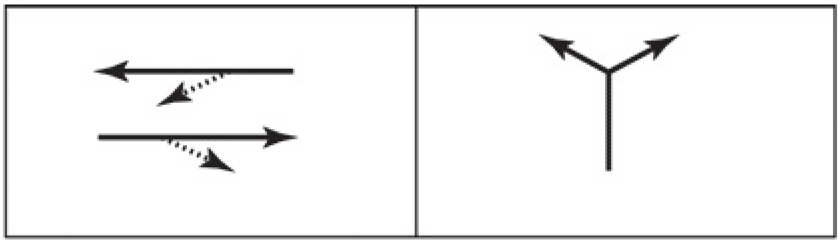
- Uncontrolled (warning and guide signs only)
- STOP or YIELD control
- Signal control

The intersection shown in Figure 17.9 has one lane for each approach. There are no channelized movements or left-turn lanes. If visibility is not appropriate for uncontrolled operation under basic rules of the road, then the options of STOP/YIELD control or signalization must be considered. The normal warrants would apply.

The T-intersection form, however, presents some relatively unique characteristics that influence how control is applied. STOP control is usually applied to the stem of the T-intersection, although it is possible to apply two-way STOP control to the cross street if movements into and out of the stem dominate.

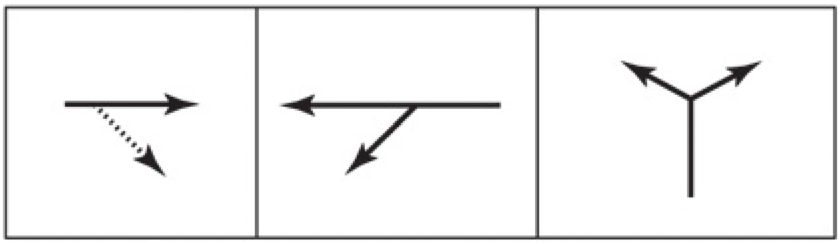
If needed, the form of signalization applied to the intersection of Figure 17.9 depends entirely on the need to protect the (WB) opposed left turn. A protected phase is normally suggested if the left-turn volume exceeds 200 veh/h or the cross-product of the left-turn volume and the opposing volume per lane exceeds 50,000. If left-turn protection is not needed, a simple two-phase signal plan is used. If the opposed left-turn must be protected and there is no left-turn lane available (as in Figure 17.9), a three-phase plan must be used. Figure 17.10 illustrates possible signal plans for the T-intersection of Figure 17.9. The three-phase plan is relatively inefficient, because a separate phase is needed for each of the three approaches.

Figure 17.10: Signalization Options for the T-Intersection of Figure 17.9



(a) A Two-Phase Signal Plan for the T-Intersection of Figure 17.9 (Permitted Left Turns)

17.4-4 Full Alternative Text



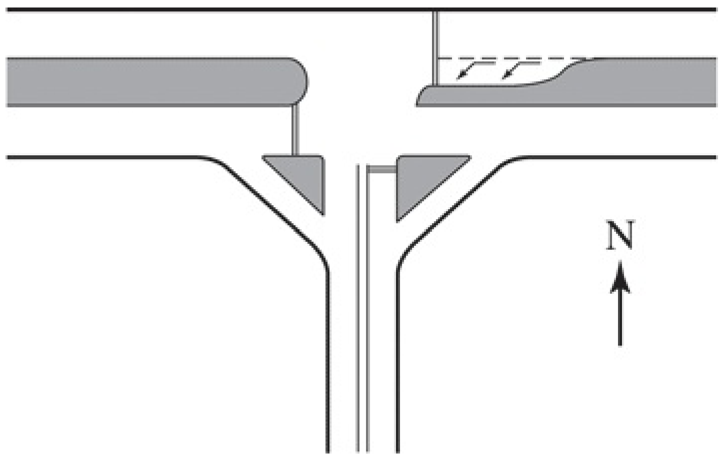
(b) A Three-Phase Signal Plan for the T-Intersection of Figure 17.9 (Protected Left Turns)

17.4-4 Full Alternative Text

Where a protected left-turn phase is desirable, the addition of an exclusive left-turn lane would simplify the signalization. Channelization and some additional right-of-way would be required to do this. Channelization can also be applied in other ways to simplify the overall operation and control of the intersection. Channelizing islands can be used to create separated right-turn paths for vehicles entering and leaving the stem via right turns. Such movements would be YIELD-controlled, regardless of the primary form of intersection control.

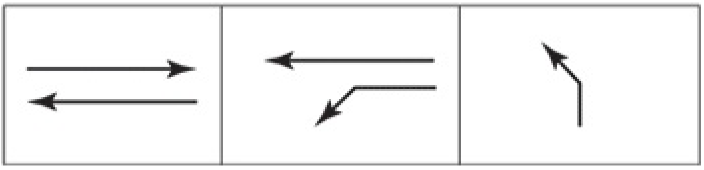
Figure 17.11 shows a T-intersection in which a left-turn lane is provided for the opposed left turn. Right turns are also channelized. Assuming that a signal with a protected left turn is needed at this location, the signal plan shown could be implemented. This plan is far more efficient than that of Figure 17.10, as EB and WB through flows can move simultaneously. Right turns move more or less continuously through the YIELD-controlled channelized turning roadways. The potential for queues to block access to the right-turn roadways, however, should be considered in timing the signal.

Figure 17.11: A Channelized T-Intersection with Improved Signalization



(a) A Channelized T-Intersection

17.4-4 Full Alternative Text



(b) Signal Plan with Protected Left Turn and YIELD-Controlled Right Turns

17.4-4 Full Alternative Text

Right turns can be completely eliminated from the signal plan if volumes are sufficient to allow lane drops or additions for the right-turning movements, as illustrated in Figure 17.12. Right turns into and out of the stem of the T-intersection become continuous movements.

Figure 17.12: T-Intersection and Signal Plan with Right-Turn Lane Drops and Lane Additions

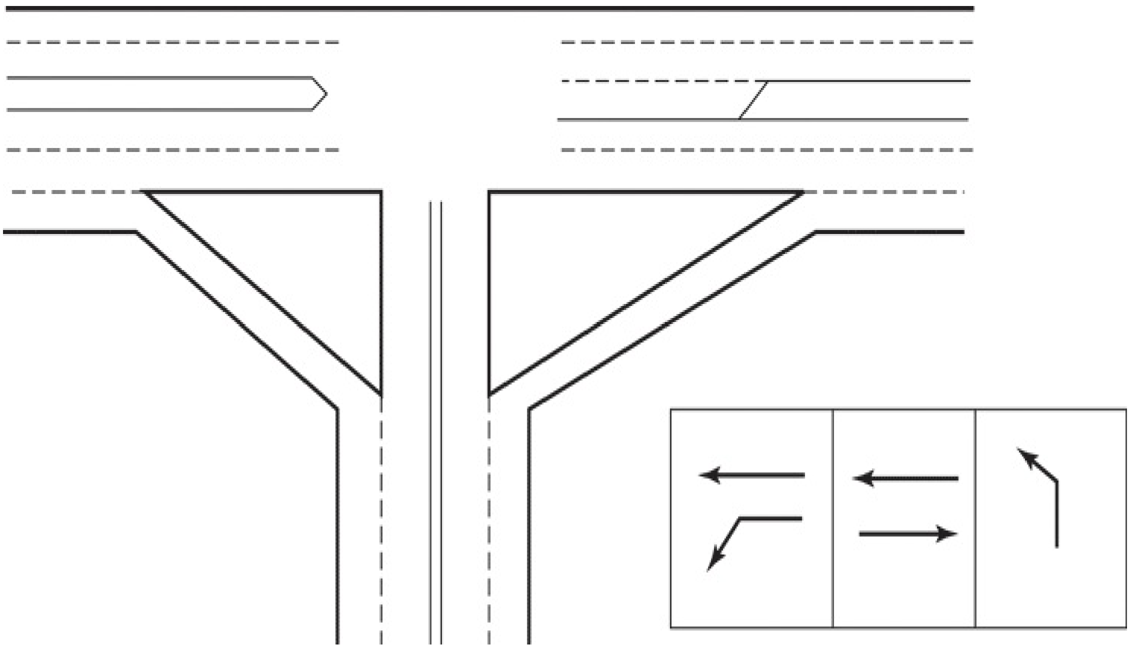
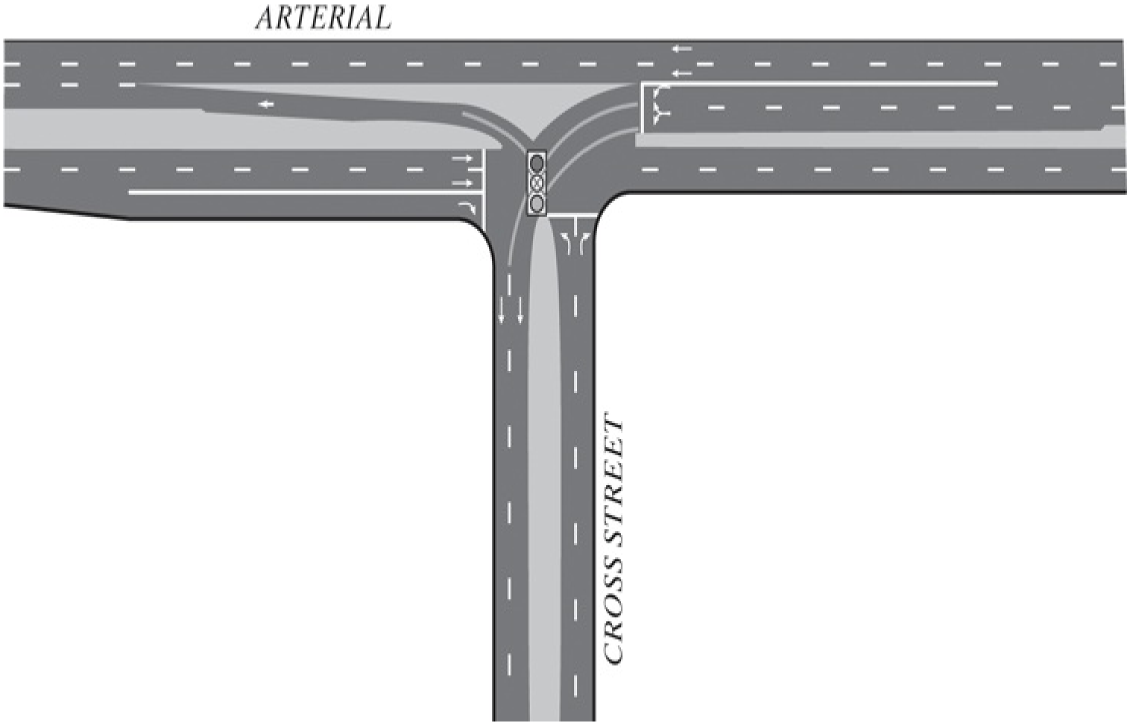


Figure 17.12: Full Alternative Text

In some cases, one through movement across the top of the T can be permitted to run continuously through a signal cycle. This is possible because this movement only presents merging and diverging conflicts with left turns into the stem and left turns out of the stem. These conflicts can be safely managed through the use of channelization, as shown in Figure 17.13.

Figure 17.13: T-Intersection with Continuous Green for One through Movement



(Source: Hughes, W., Jagannathon, R., Sengupta, D., and Human, J., "Alternative Intersections/Interchanges: Informational Report," *Publication Number FHWA-HRT-09-060*, Federal Highway Administration, Washington, D.C., 2009, Figure 150.)

Figure 17.13: Full Alternative Text

Such a design and signal timing is only feasible where there are virtually no pedestrians crossing the top of the T, or where overpasses or underpasses are provided for pedestrians, along with effective barriers to prevent them from crossing at grade.

17.4.3 Offset Intersections

One of the traffic engineer's most difficult problems is the safe operation of high-volume offset intersections. Figure 17.14 illustrates such an intersection with a modest right offset. In the case illustrated, the driver needs more sight distance (when compared with a perfectly aligned 90° intersection) to observe vehicles approaching from the right. The obstruction caused by the building becomes a more serious problem because of this. In addition to sight distance problems, the offset intersection distorts the normal trajectory of all movements, creating accident risks that do not exist at aligned intersections.

Figure 17.14: Offset Intersection with Sight Distance and Trajectory Problems

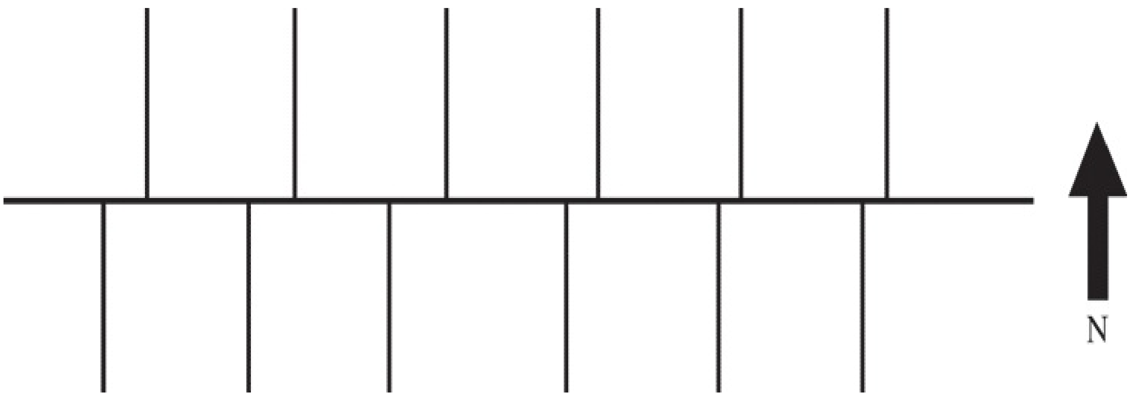
(Source: Photo courtesy of R. Roess and J. Ulerio.)

Offset intersections are rarely consciously designed. They are necessitated by a variety of situations, generally involving long-standing historic development patterns. Figure 17.14 illustrates a relatively common situation in which offset intersections occur.

In many older urban or suburban developments, zoning and other regulations were (and in some cases, still are) not particularly stringent. Additional development was considered to be an economic benefit because it added to the property tax base of the community involved. Firm control over the specific design of subdivision developments, therefore, is not always exercised by zoning boards and authorities.

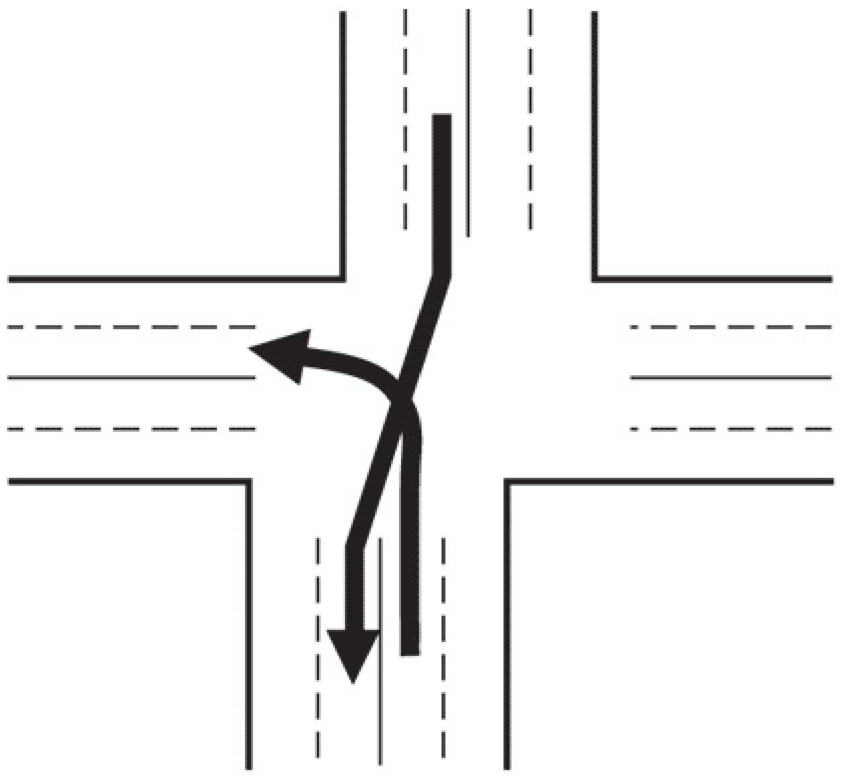
The situation depicted in Figure 17.15 occurs when Developer A obtains the land to the south of a major arterial and lays out a circulation system that will maximize the number of building lots that can be accommodated on the parcel. At a later time, Developer B obtains the rights to land north of the same arterial. Again, an internal layout that provides the maximum number of development parcels is selected. Without a strong planning board or other oversight group requiring it, there is no guarantee that opposing local streets will "line up." Offsets can and do occur frequently in such circumstances. In urban and suburban environments, it is rarely possible to acquire sufficient right-of-way to realign the intersections; therefore, other approaches to control and operation of such intersections must be considered.

Figure 17.15: A Common Situation for Offset Intersections



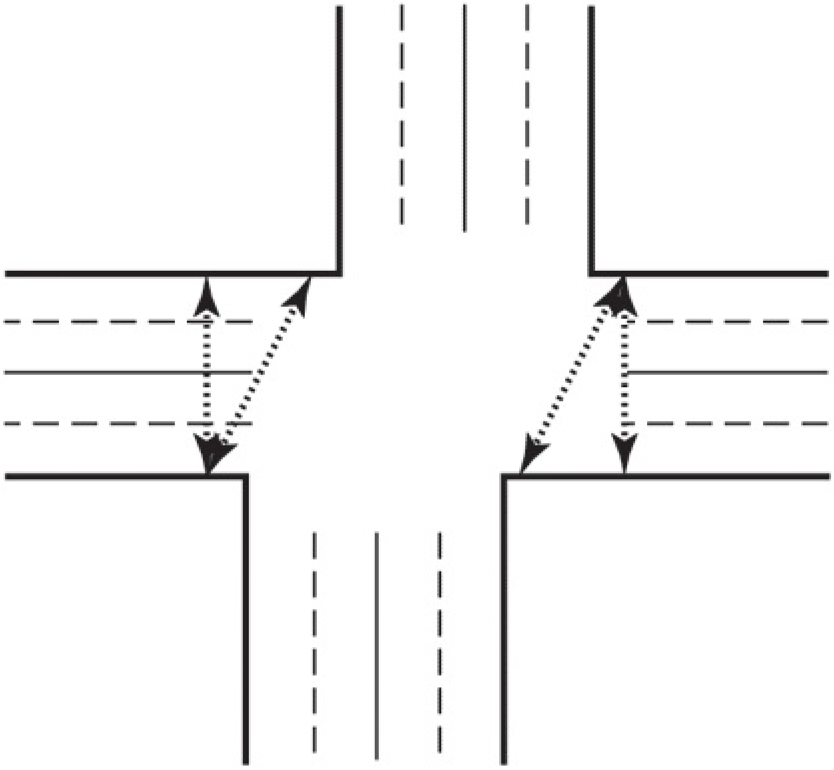
There are two major operational problems posed by a right-offset intersection, as illustrated in Figure 17.16.

Figure 17.16: Special Problems at Offset Intersections



(a) Left-Turn Trajectory Problems Illustrated

17.4-4 Full Alternative Text



(b) Pedestrian Path Problems Illustrated

17.4-4 Full Alternative Text

In Figure 17.16(a), the left-turn trajectories from the offset legs involve a high level of hazard. Unlike the situation with an aligned intersection, a vehicle turning left from either offset leg is in conflict with the opposing through vehicle almost immediately after crossing the STOP line. To avoid this conflict, left-turning vehicles must bear right as if they were going to go through to the opposite leg, beginning their left turns only when they are approximately halfway through the intersection. This, of course, is not a natural movement, and a high incidence of left-turn accidents often result at such intersections.

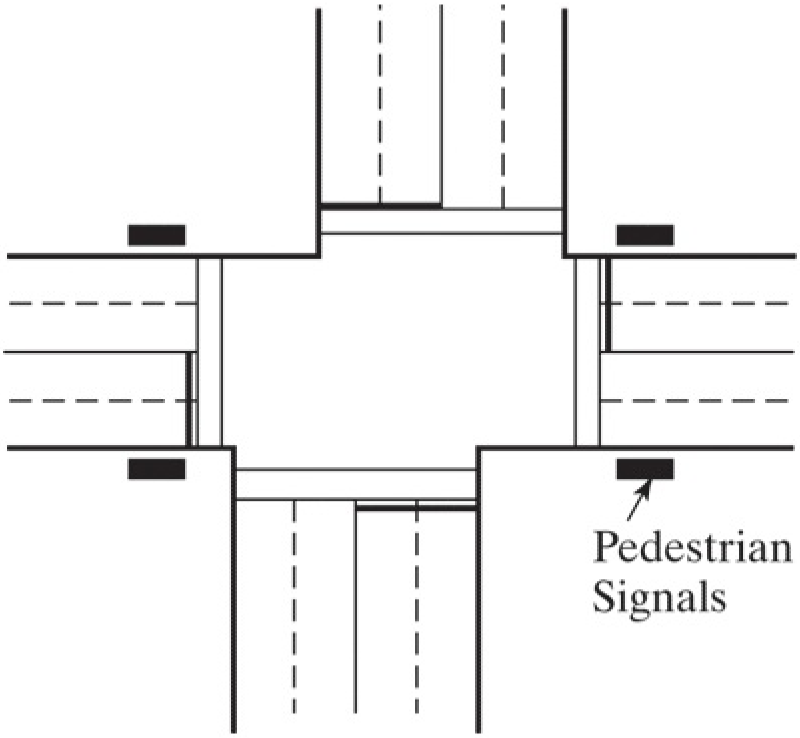
In Figure 17.16(b), the hazard to pedestrians crossing the aligned roadway is highlighted. Two paths are possible, and both are reasonably intuitive for pedestrians: They can cross from corner to corner, following an angled crossing path, or they can cross perpendicularly. The latter places one end of their crossing away from the street corner. Perpendicular crossings, however, minimize the crossing time and distance. On the other hand, right-turning vehicles encounter the pedestrian conflict at an unexpected location, after they have virtually completed their right turn. Diagonal crossings increase the exposure of pedestrians, but conflicts with right-turning vehicles are closer to the normal location.

Yet another special hazard at offset intersections, not clearly illustrated by Figure 17.16, is the heightened risk of sideswipe accidents as vehicles cross between the offset legs. Since the required angular path is not necessarily obvious, more vehicles will stray from their lane during the crossing.

There are, however, remedies that will minimize these additional hazards. Where the intersection is signalized, the left-turn conflict can be eliminated through the use of a fully protected left-turn phase in the direction of the offset. In this case, the left-turning vehicles will not be entering the intersection area at the same time as the opposing through vehicles. This requires, however, that one of the existing lanes be designated an exclusive turning lane, or that a left-turn lane can be added to each offset leg. If this is not possible, a more extreme remedy is to provide each of the offset legs with an exclusive signal phase. While this separates the left-turning vehicles from the opposing flows, it is an inefficient signal plan and can lead to four-phase signalization if left-turn phases are needed on the aligned arterial.

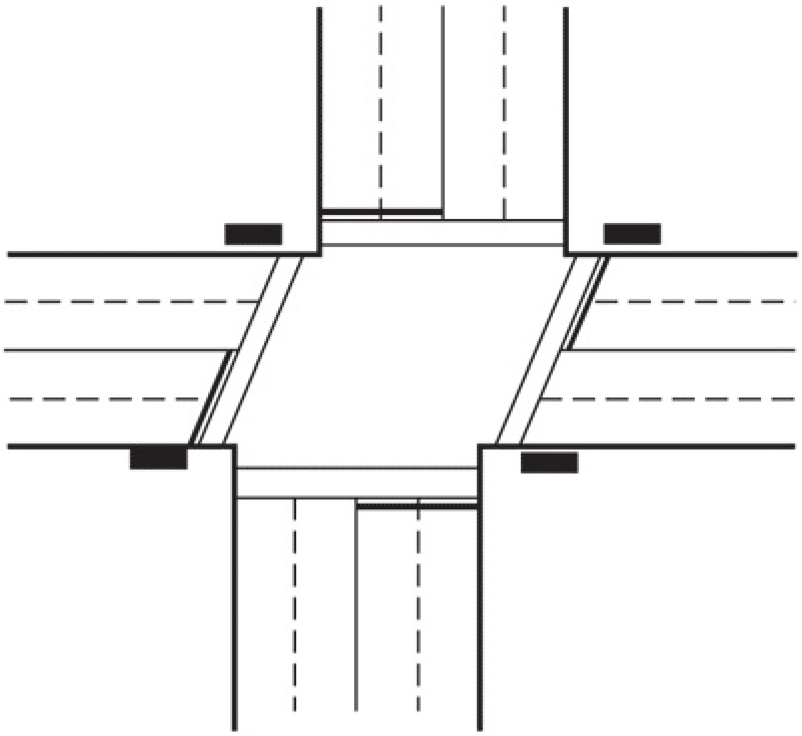
For pedestrian safety, it is absolutely necessary that the traffic engineer clearly designate the intended path they are to take. This is done through proper use of markings, signs, and pedestrian signals, as shown in Figure 17.17.

Figure 17.17: Signing, Markings, and Pedestrian Signals for a Right-Offset Intersection



(a) Markings for Perpendicular Pedestrian Crossings

17.4-4 Full Alternative Text



(b) Markings for Diagonal Pedestrian Crossings

17.4-4 Full Alternative Text

Crosswalk locations influence the location of STOP-lines and the position of pedestrian signals, which must be located in the line of sight (which is the walking path) of pedestrians. Vehicular signal timing is also influenced by the crossing paths implemented. Where perpendicular crossings are used, the distance between STOP-lines on the aligned street can be considerably longer than for diagonal crossings. This increases the length of the all-red interval for the aligned street and adds lost time to the signal cycle.

In extreme cases, where enforcement of perpendicular crossings becomes difficult, barriers can be placed at normal street corner locations, preventing pedestrians from entering the street at an inappropriate or unintended location.

To help vehicles follow appropriate paths through the offset intersection, dashed lane and centerline markings through the intersection may be added, as illustrated in Figure 17.18. The extended centerline marking would be yellow, and the lane lines would be white.

Figure 17.18: Dashed Lane and Centerline Through an Offset Intersection

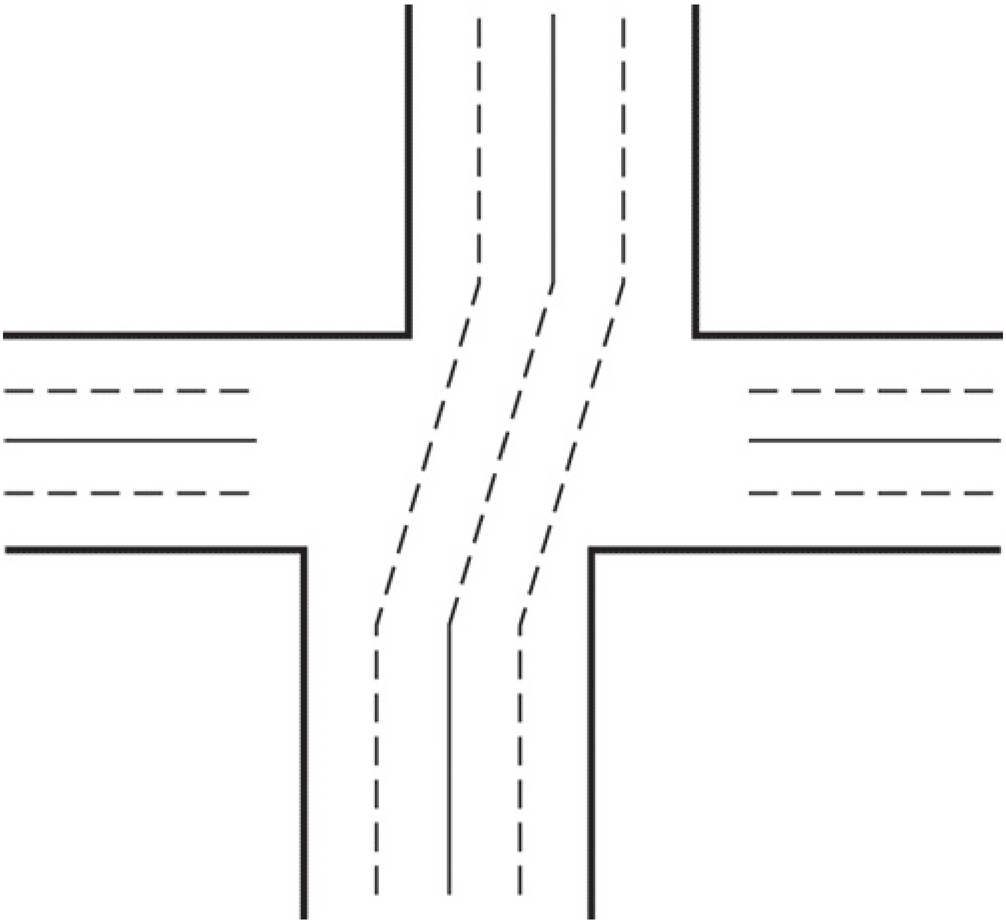


Figure 17.18: Full Alternative Text

Left-offset intersections share some of the same problems as right-offset intersections. The left-turn interaction with the opposing through flow is not as critical, however. The pedestrian–right-turn interaction is different, but potentially just as serious. Figure 17.19 illustrates.

Figure 17.19: Conflicts at a Left-Offset Intersection

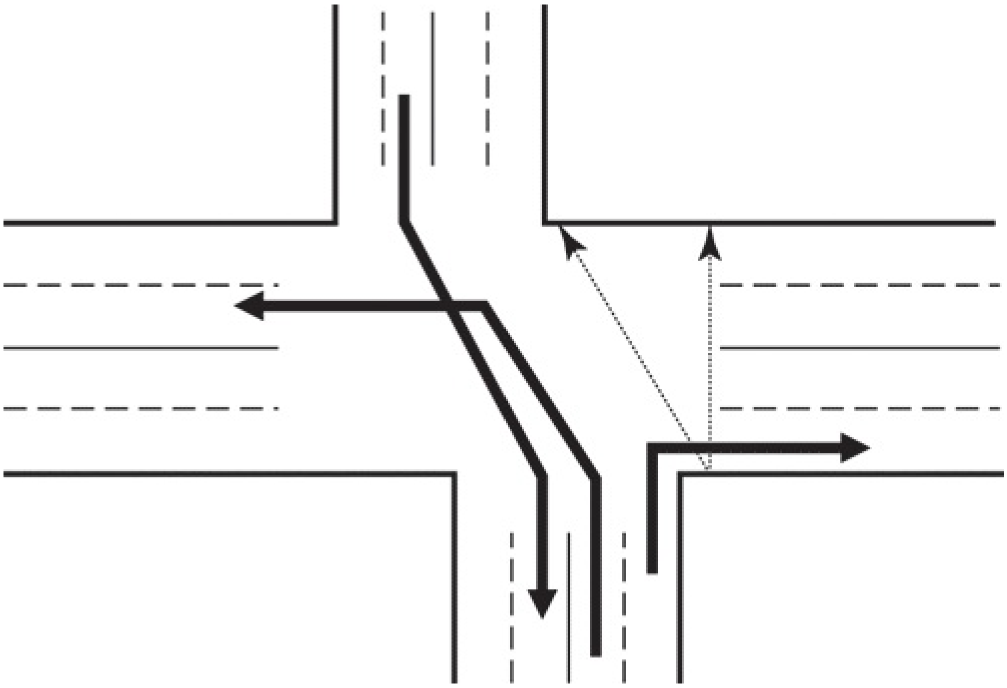


Figure 17.19: Full Alternative Text

The left-turn trajectory through the offset intersection is still quite different from an aligned intersection, but the left-turn movement does not thrust the vehicle immediately into the path of the oncoming through movement, as in a right-offset intersection. Sideswipe accidents are still a risk, and extended lane markings would be used to minimize this risk.

At a left-offset intersection, the diagonal pedestrian path is more difficult, as it brings the pedestrian into immediate conflict with right-turning vehicles more quickly than at a normally aligned intersection. For this reason, diagonal crossings are generally not recommended at left-offset intersections. The signing, marking, and signalization of perpendicular pedestrian crossings is similar to that used at a right-offset intersection.

When at all possible, offset intersections should be avoided. If sufficient right-of-way is available, basic realignment should be seriously considered. When confronted with such a situation however, the traffic engineering approaches discussed here can ameliorate some of the fundamental concerns associated with offset alignments. The traffic engineer should recognize that many of these measures will negatively affect capacity of the approaches due to the additional signal phases and longer lost times often involved. This is, however, a necessary price paid to optimize safety of intersection operation.

17.4.4 Special Treatments for Heavy Left-Turn Movements

Some of the most difficult intersection problems to solve involve heavy left-turn movements on major arterials. Accommodating such turns usually requires the addition of protected left-turn phasing, which often reduces the effective capacity to handle through movements. In some cases, adding an exclusive left-turn phase or phases is not practical, given the associated losses in through capacity.

Alternative treatments must be sought to handle such left-turn movements, with the objective of maintaining two-phase signalization at the intersection. Several design and control treatments are possible, including:

- Prohibition of left turns
- Provision of jug-handles
- Provision of at-grade loops and diamond ramps
- Provision of a continuous-flow (or diverted LT) intersection
- Provision of U-turn treatments

Prohibition of left turns is rarely a practical option for a heavy left-turn demand. Alternative paths would be needed to accommodate the demand for this movement, and diversion of a heavy flow onto an “around-the-block” or similar path often creates problems elsewhere.

All of the geometric approaches listed fall under the heading of “alternative” or “distributed” intersections, which are discussed in detail in Chapter 26.

17.5 Closing Comments

This chapter has provided an overview of several important elements of intersections design. It is not intended to be exhaustive, and the reader is encouraged to consult standard references for additional relevant topics and detail. As noted previously, additional aspects of intersection planning, design, and control are presented elsewhere in this text:

- Chapters 18–21 treat signalization of intersections in great detail.
- Chapter 16 covers traffic signal hardware, including controllers, detectors, street hardware, and its placement in the intersection.
- Chapter 25 covers unsignalized intersections, including roundabouts.
- Chapter 26 discusses interchanges and alternative intersections.

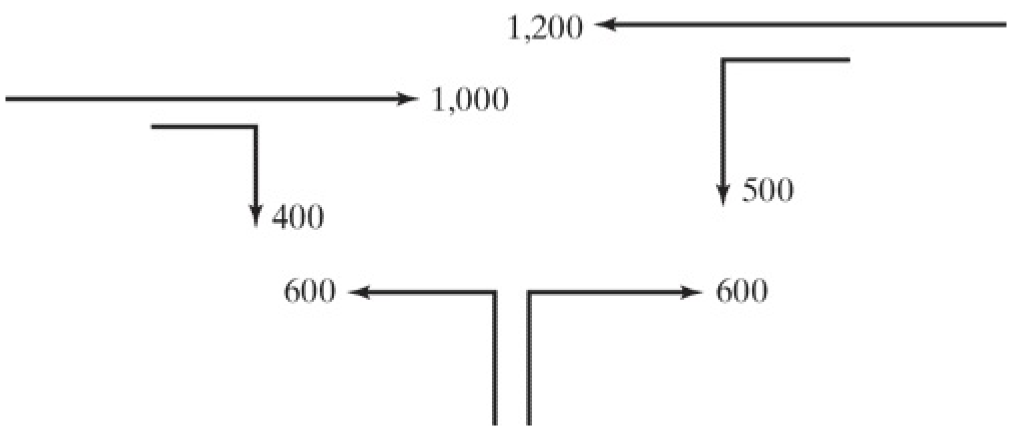
References

- 1. *A Policy on Geometric Design of Highways and Streets*, 6th Edition, American Association of State Highway and Transportation Officials, Washington, D.C., 2014.
- 2. *Manual of Uniform Traffic Control Devices*, Federal Highway Administration, U.S. Department of Transportation, 2009, as amended through May 2012.
- 3. Kell, J. and Fullerton, I., *Manual of Traffic Signal Design*, 2nd Edition, Institute of Transportation Engineers, Washington, D.C., 1991.
- 4. *Traffic Detector Handbook*, JHK & Associates, Institute of Transportation Engineers, Washington, D.C., n.d.
- 5. *Highway Capacity Manual, 6th Edition: A Guide for Multimodal Mobility Analysis*, Transportation Research Board, Washington, D.C., 2016.

Problems

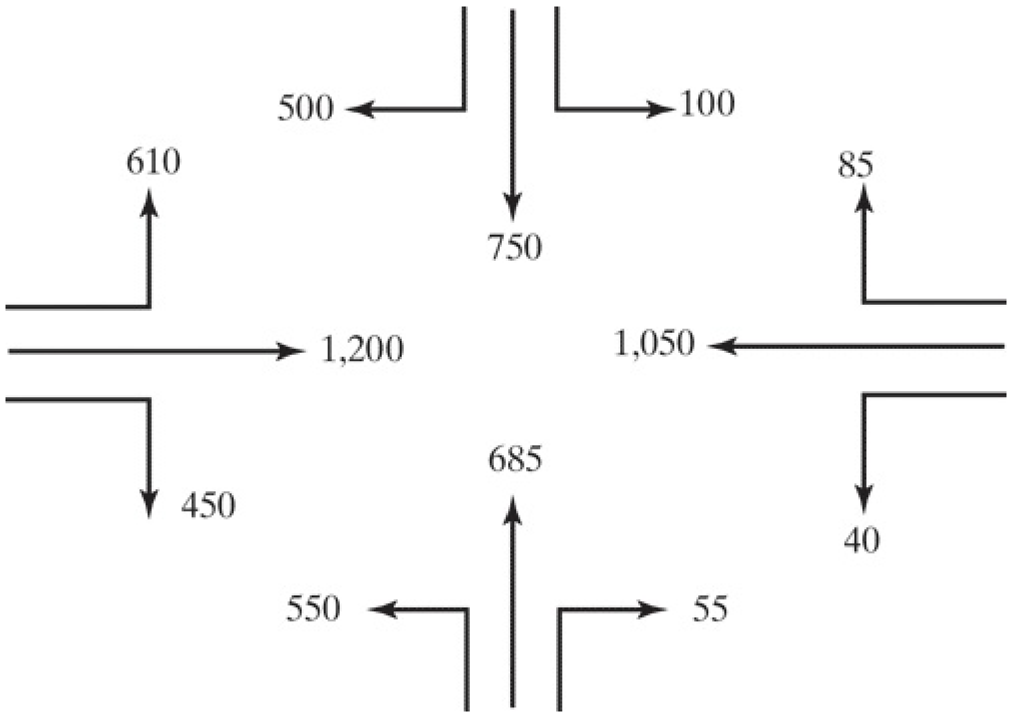
1. 17-1– 17-2. Each of the sets of demands shown in the figures below represent the forecast flows (already adjusted for PHF) expected at new intersections that are created as a result of large new developments. Assume that each intersection will be signalized. In each case, propose a design for the intersection.

Demands for Problem 17-1



Full Alternative Text

Demands for Problem 17-2



Full Alternative Text

2. 17-3. At T-intersections, what is required to use a signalization that provides a continuous green for the through movement across the top of the T?

3. 17-4. Why do offset intersections occur? What techniques can be used to ameliorate some of the obvious hazards that offset intersections present?

Chapter 18 Principles of Intersection Signalization

In Chapter 15, various options for intersection control were presented and discussed. Warrants for implementation of traffic control signals at an intersection, presented in the *Manual on Uniform Traffic Control Devices* [1], provide general and specific criteria for selection of an appropriate form of intersection control. At many intersections, the combination of traffic volumes, potential conflicts, overall safety of operation, efficiency of operation, and driver convenience will lead to a decision to install traffic control signals.

The operation of signalized intersections is often complex, involving competing vehicular and pedestrian movements. Appropriate methodologies for design and timing of signals and for the operational analysis of signalized intersections require that the behavior of drivers and pedestrians at a signalized intersection be modeled in a form that can be easily manipulated and optimized. This chapter discusses some of the fundamental operational characteristics at a signalized intersection and the ways in which they may be effectively represented analytically.

Subsequent chapters combine these and other analytic elements into practical applications for the design and timing of both pre-timed and actuated signals, the operational analysis of signalized intersections, and the coordination of signalized intersections to form a signal system.

This chapter focuses on four critical aspects of signalized intersection operation:

1. Discharge headways, saturation flow rates, and lost times
2. Allocation of time and the critical lane concept (referred to as the “time budget”)
3. The concept of left-turn (and right-turn) equivalency
4. Delay as a measure of service quality

There are other aspects of signalized intersection operation that are also important, and the *Highway Capacity Manual* [2] analysis model addresses many of them. These four, however, are central to understanding traffic behavior at signalized intersections.

18.1 Terms and Definitions

Traffic signals are complex devices that can operate in a variety of different modes. A number of key terms and definitions should be understood before pursuing a more substantive discussion.

18.1.1 Components of a Signal Cycle

The following terms describe portions and subportions of a signal cycle. The most fundamental unit in signal design and timing is the *cycle*, as defined below:

1. Cycle. A signal cycle is one complete rotation through all of the indications provided. In general, every legal vehicular movement receives a "green" indication during each cycle, although there are some exceptions to this rule.

2. Cycle length. The cycle length is the time (in seconds) that it takes to complete one full cycle of indications. It is given the symbol "*C*."

3. Interval. The interval is a period of time during which no signal indication changes. It is the smallest unit of time described within a signal cycle. There are several types of intervals within a signal cycle:

    1. Change interval. The change interval is the "yellow" indication for a given movement. It is part of the transition from "green" to "red," in which movements about to lose "green" are given a "yellow" signal, while all other movements have a "red" signal. It is timed to allow a vehicle that cannot safely stop when the "green" is withdrawn to enter the intersection legally. The change interval is given the symbol "yi" for movement(s) *i*.

    2. Clearance interval. The clearance interval is also part of the transition from "green" to "red" for a given set of movements. During the clearance interval, all movements have a "red"

signal. It is timed to allow a vehicle that legally enters the intersection on “yellow” to safely cross the intersection before conflicting flows are released. The clearance interval is given the symbol “ari” (for “all red”) for movement(s) *i*.

3. Green interval. Each movement usually has one green interval during the signal cycle. During a green interval, the movements permitted have a “green” light, while all other movements have a “red” light. The green interval is given the symbol “Gi” for movement(s) *i*.

4. Red interval. Each movement has a red interval during the signal cycle. All movements not permitted have a “red” light, while those permitted to move have a “green” light. In general, the red interval overlaps the green intervals for all other movements in the intersection. The red interval is given the symbol “Ri” for movement(s) *i*. Note that for a given movement or set of movements, the “red” signal is present during both the *clearance* (all red) and *red* intervals.

4. Phase. A signal phase consists of a green interval, plus the change and clearance intervals that follow it. It is a set of intervals that allows a designated movement or set of movements to flow and to be safely halted before release of a conflicting set of movements.

18.1.2 Types of Signal Operation

The traffic signals at an individual intersection can operate on a pre-timed basis or may be partially or fully actuated by arriving vehicles or pedestrians sensed by detectors.

1. Pre-timed operation. In pre-timed operation, the cycle length, phase sequence, and timing of each interval are constant. Each cycle of the signal follows the same predetermined plan. Modern signal controllers will allow different pre-timed settings to be established. An internal clock is used to activate the appropriate timing for each defined time period. In such cases, it is typical to have at least an AM peak, a PM peak, and an off-peak signal timing, but additional time periods may also be defined.

2. Semi-actuated operation. In semi-actuated operation, detectors are placed on the minor approach(es) to the intersection; there are no detectors on the major street. The light is green for the major street at all times except when a “call” or actuation is noted on one of the minor approaches. Then, subject to limitations such as a minimum major-street green, the green is transferred to the minor street. The green returns to the major street when the maximum minor-street green is reached or when the detector senses that there is no further demand on the minor street. Semi-actuated operation is often used where the primary reason for signalization is “interruption of continuous traffic,” as discussed in Chapter 15.

3. Full-actuated operation. In full-actuated operation, every lane of every approach must be monitored by a detector. Green time is allocated in accordance with information from detectors and programmed “rules,” established in the controller for capturing and retaining the green. In full- actuated operation, the cycle length, sequence of phases, and green time split may vary from cycle to cycle. Chapter 20 presents more detailed descriptions of actuated signal operation, along with a methodology for timing such signals.

In most urban and suburban settings, signalized intersections along arterials and in arterial networks are close enough to have a significant impact on adjacent signalized intersection operations. In such cases, it is common to coordinate signals into a signal system. When coordinated, such systems attempt to keep vehicles moving through sequences of individual signalized intersections without stopping for as long as possible. This is done by controlling the “offsets” between adjacent green signals, that is, the green at a downstream signal will initiate “x” seconds after its immediate upstream neighbor. Coordinated signal systems must operate on a common cycle length, as offsets cannot be maintained from cycle-to-cycle if cycle lengths vary at each intersection. Coordination is provided using a variety of technologies:

1. Master controllers: A “master controller” provides a linkage between a limited set of signals. Most such controllers can connect from 20 to 30 signals along an arterial or in a network. The master controller provides fixed settings for each offset between connected signals. Settings can be changed for defined periods of the day.

2. Computer control. In a computer-controlled system, the computer acts

as a “super-sized” master controller, coordinating the timings of a large number (hundreds) of signals. The computer selects or calculates an optimal coordination plan based on input from detectors placed throughout the system. In general, such selections are made only once in advance of an AM or PM peak period. The nature of a system transition from one timing plan to another is sufficiently disruptive to be avoided during peak-demand periods in a traditional system. Individual signals in a computer-controlled system generally operate in the pre-timed mode.

3. Adaptive traffic control systems (ATCS): Since the early 1990s, there has been rapid development and implementation of “adaptive” traffic control systems. In such systems, both individual intersection signal timings and offsets are continually modified in real time based upon advanced detection system inputs. In many cases, such systems use actuated controllers at individual intersections. Even though the system still requires a fixed cycle length (which can be changed periodically based upon detector input), the continuous reallocation of green within a fixed cycle length has been found to be useful in reducing delay and travel times. A critical part of adaptive traffic control systems is the underlying logic of software used to monitor the system and continually update timing patterns. There are a number of software systems in use, and the list of products is increasing each year.

Table 16.1 (Chapter 16) summarized the various types of signal controllers and provided guidance on their use. Prior to 1990, virtually all coordinated traffic signal systems on arterials and in networks used pre-timed signal controllers exclusively. Today, actuated controllers are regularly coordinated, although they lose one of their principal variable features—cycle length. To coordinate signals, cycle lengths must be common during any given time period, so that the offset between initiation of green at an upstream intersection and the adjacent downstream intersection is constant for every cycle. Pre-timed signals, because they are the cheapest to implement and maintain, are still a popular choice where demands are relatively constant throughout major periods of the day. Where demand levels (and relative demands for various movements) vary significantly during all times of the day, actuated signals are the most likely choice for use. Even when coordinated and using a constant cycle length, the allocation of green times amongst the defined phases can significantly reduce delay.

18.1.3 Treatment of Left Turns and Right Turns

The modeling of signalized intersection operation would be straightforward if left turns did not exist. Left turns at a signalized intersection can be handled in one of three ways:

1. Permitted left turns. A “permitted” left turn movement is one that is made across an opposing flow of vehicles. The driver is permitted to cross through the opposing flow, but must select an appropriate gap in the opposing traffic stream through which to turn. This is the most common form of left-turn phasing at signalized intersections, used where left-turn volumes are reasonable and where gaps in the opposing flow are adequate to accommodate left turns safely.

2. Protected left turns. A “protected” left turn movement is made without an opposing vehicular flow. The signal plan protects left-turning vehicles by stopping the opposing through movement. This requires that the left turns and the opposing through flow be accommodated in separate signal phases and leads to multiphase (more than two) signalization. In some cases, left turns are “protected” by geometry or regulation. Left turns from the stem of a T-intersection, for example, face no opposing flow, as there is no opposing approach to the intersection. Left turns from a one-way street similarly do not face an opposing vehicular flow.

3. Compound left turns. More complicated signal timing can be designed in which left turns are protected for a portion of the signal cycle and are permitted in another portion of the cycle. Protected and permitted portions of the cycle can be provided in any order. Such phasing is also referred to as *protected plus permitted* or *permitted plus protected*, depending upon the order of the sequence.

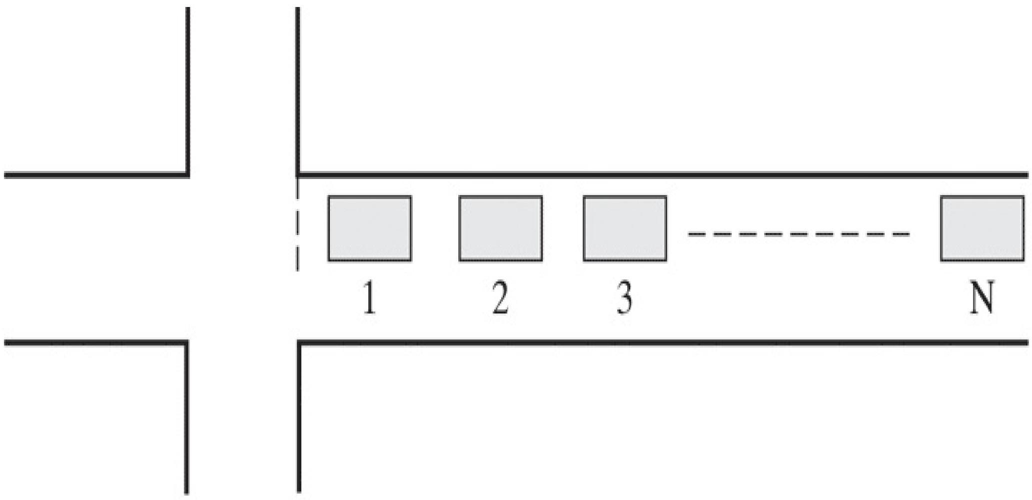
The permitted left turn movement is very complex. It involves the conflict between a left turn and an opposing through movement. The operation is affected by the left-turn flow rate and the opposing flow rate, the number of opposing lanes, whether left turns flow from an exclusive left-turn lane or from a shared lane, and the details of the signal timing. Modeling the interaction among these elements is a complicated process, one that often involves iterative elements.

The terms *protected* and *permitted* may also be applied to right turns. In this case, however, the conflict is between the right-turn vehicular movement and the pedestrian movement in the conflicting crosswalk. The vast majority of right turns at signalized intersections are handled on a permitted basis. Protected right turns generally occur at locations where there are overpasses or underpasses provided for pedestrians. At these locations, pedestrians are prohibited from making surface crossings; barriers are often required to enforce such a prohibition.

18.2 Discharge Headways, Saturation Flow, Lost Times, and Capacity

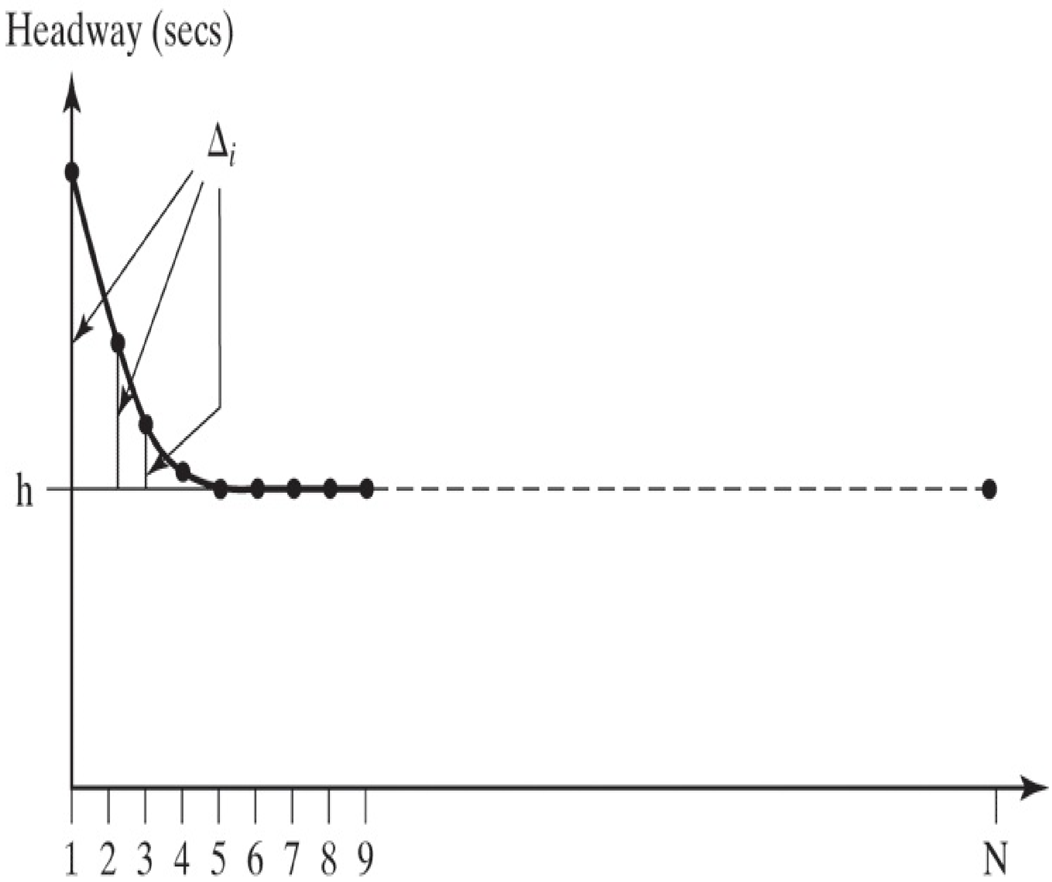
The fundamental element of a signalized intersection is the periodic stopping and restarting of the traffic stream. Figure 18.1 illustrates this process. When the light turns GREEN, there is a queue of stored vehicles that were stopped during the preceding RED phase, waiting to be discharged. As the queue of vehicles moves, headway measurements are taken as follows:

Figure 18.1: Flow Departing a Queue at a Signalized Intersection



(a) Vehicles in an Intersection Queue

18.2-1 Full Alternative Text



(b) Average Headways Departing Signal

18.2-1 Full Alternative Text

- The first headway is the time lapse between the initiation of the GREEN signal and the time that the front wheels of the first vehicle cross the intersection or stop line.
- The second headway is the time lapse between the time that the first vehicle's front wheels cross the intersection or stop line and the time that the second vehicle's front wheels cross the intersection or stop line.
- Subsequent headways are similarly measured.
- Only headways through the last vehicle in queue (at the initiation of the GREEN light) are considered to be operating under "saturated" conditions.

Note that depending on local practice, headways may be measured at the stop line, or at the intersection (or curb) line. Either is acceptable, and both are in common use. It is important to remember, however, that whichever reference point is used must be used throughout the signal timing process (such as in the determination of “yellow” and “all red” intervals), and must be used for all approaches.

If many queues of vehicles are observed at a given location and the average headway is plotted versus the queue position of the vehicle, a trend similar to that shown in Figure 18.1(b) emerges.

The first headway is relatively long. The first driver must go through the full perception-reaction sequence, move his or her foot from the brake to the accelerator, and accelerate through the intersection. The second headway is shorter, because the second driver can overlap the perception-reaction and acceleration process of the first driver. Each successive headway is a little bit smaller than the last. Eventually, the headways tend to level out. This generally occurs when queued vehicles have fully accelerated by the time they cross the stop line. At this point, a stable moving queue has been established.

18.2.1 Saturation Headway and Saturation Flow Rate

As noted, average headways will tend towards a constant value. In general, this occurs from the fourth or fifth headway position. The constant headway achieved is referred to as the *saturation headway*, as it is the average headway that can be achieved by a saturated, stable moving queue of vehicles passing through the signal. It is given the symbol “*h*,” in units of seconds/vehicle.

It is convenient to model behavior at a signalized intersection by assuming that every vehicle (in a given lane) consumes an average of “*h*” seconds of green time to enter the intersection. If every vehicle consumes “*h*” seconds of green time and if the signal were *always* green, then “*s*” vehicles per hour could enter the intersection. This is referred to as the *saturation flow rate*:

s=3,600h [18-1]

where:

s = saturation flow rate, vehicles per hour of greenper lane (veh/hg/ln), and
h = saturation headway, seconds/vehicle (s/veh).

If there are multiple lanes on an approach, the saturation headway and flow rate are usually not the same in all lanes. The saturation flow rate for the approach would then be the sum of the individual lane saturation flow rates that comprise the approach.

The saturation flow rate is, in effect, the capacity of the approach lane or lanes if they were available for use all of the time (i.e., if the signal were always GREEN). The signal, of course, is not always GREEN for any given movement. Thus, some mechanism (or model) for dealing with the cyclic starting and stopping of movements must be developed.

18.2.2 Start-Up Lost Time

The average headway per vehicle is actually greater than "*h*" seconds. The first several headways are, in fact, larger than "*h*" seconds, as illustrated in Figure 18.1(b). The first three or four headways involve additional time as drivers react to the GREEN signal and accelerate. The additional time involved in each of these initial headways (above and beyond "*h*" seconds) is noted by the symbol Δi (for headway *i*). These additional times are added, and are referred to as the *start-up lost time*:

$\ell 1=\sum i \Delta i$ [18-2]

where:

$\ell 1$ = start-up lost time, s/phase, and Δi =
incremental headway (above "h" seconds) for vehicle i, s

Thus, it is possible to model the amount of GREEN time required to discharge a queue of "*n*" vehicles as:

$Tn=\ell 1+nh$ [18-3]

where:

T_n =
GREEN time required to move queue of "n" vehicles through a signalized in
ℓ_1 = start-up lost time, s/phase, n = number of vehicles in queue, and h = saturation headway, s/veh.

While this particular model is not of great use, it does illustrate the basic concepts of saturation headway and start-up lost times. The start-up lost time is thought of as a period of time that is "lost" to vehicle use. Remaining GREEN time, however, may be assumed to be usable at a rate of *h* s/veh.

18.2.3 Clearance Lost Time

The start-up lost time occurs every time a queue of vehicles starts moving on a GREEN signal. There is also a lost time associated with stopping the queue at the end of the GREEN signal. This time is more difficult to observe in the field, as it requires that the standing queue of vehicles be large enough to consume *all* of the GREEN time provided. In such a situation, the clearance lost time, ℓ_2, is defined as the time interval between the last vehicle's front wheels crossing the intersection or stop line, and the initiation of the GREEN for the *next* phase. The clearance lost time occurs each time a flow of vehicles is stopped.

18.2.4 Total Lost Time and the Concept of Effective GREEN Time

If the start-up lost time occurs each time a queue starts to move and the clearance lost time occurs each time the flow of vehicles stops, then for each GREEN phase:

$t_L = \ell_1 + \ell_2$ [18-4]

where:

t_L =total lost time per phase, s/phase, and

All other variables are as previously defined.

The concept of lost times leads to the concept of *effective green time*. The actual signal goes through a sequence of intervals for each signal phase, as previously discussed.

- Green
- Yellow
- All-red
- Red

In terms of modeling, there are really only two time periods of interest: *effective green time* and *effective red time*. For any given set of movements, *effective green time* is the amount of time during which vehicles actually enter the intersection (at a rate of one vehicle every *h* seconds). The *effective red time* is the amount of time that they cannot enter the intersection. Effective green time is related to actual green time as follows:

$$g_i = G_i + y_i + ar_i - \ell_{1i} - \ell_{2i} \quad [18\text{-}5]$$

where:

g_i = effective green time for movement(s) i, s, G_i = actual green time for movement(s) i, s, y_i = yellow interval for movement(s) i, s, ar_i = all-red interval for movement(s) i, s, ℓ_{1i} = start-up lost time for movement(s) i, s, and ℓ_{2i} = clearance lost time for movement(s) i, s.

This model results in an effective green time that may be fully utilized by vehicles at the saturation flow rate (i.e., at an average headway of *h* s/veh).

It has been found that the start-up lost time, ℓ_1, is a relatively stable value at most intersections. A default value of 2.0 second/phase is often used where locally calibrated measurements are not available. The clearance lost time, ℓ_2, however, is generally dependent on the length of the yellow and all-red intervals for the subject phase. In general terms:

$$\ell_{2i} = y_i + ar_i - e \quad [18\text{-}6]$$

where:

ℓ_{2i} = clearance lost time for phase i, s/phase, y_i =
yellow interval for phase i, s, ar_i = all-red interval for phase i, s, e =
encroachment time, s.

The encroachment time, *e*, is the amount of time during the yellow and all-red that vehicles are observed to actually enter the intersection. Theoretically, one might expect this to be equal to the yellow interval, which is timed to allow vehicles that cannot stop to safely enter the intersection, but that is often not the case. Studies have found that the encroachment time is relatively constant under most conditions. A default value of 2.0 seconds is frequently used for this value where local measurements are not available.

If Equation 18-6 is substituted in Equation 18-5, the following results:

$$g_i = G_i - \ell_{1i} + e \quad [18\text{-}7]$$

where all terms have been previously defined. This equation also yields the following observation: When the default values for ℓ_{1i} and *e* are used (both are 2.0 seconds), effective green (g_i) and actual green (G_i) have the same value.

18.2.5 Capacity of an Intersection Lane or Lane Group

The saturation flow rate(s) represents the capacity of an intersection lane or lane group assuming that the light is always GREEN. The portion of real time that is effective green is defined by the “green ratio,” the ratio of the effective green time to the cycle length of the signal (*g*/*C*). The capacity of an intersection lane or lane group may then be computed as:

$c_i = s_i \left(\frac{g_i}{C} \right)$ [18-8]

where:

c_i = capacity of a lane or lane group i, veh/h, s_i =
saturation flow rate for a lane or lane group i, veh/hg, g_i =
effective green time for a lane or lane group i, s, and C =
signal cycle length, s.

Sample Problem 18-1: Concepts of Saturation Headway and Lost Time

Consider a given movement at a signalized intersection with the following known characteristics:

- Cycle length, C=60 s
- Green time, G=27 s
- Yellow plus all-red time, Y=y+ar=4 seconds
- Saturation headway, h=2.4 s/veh
- Start-up lost time, ℓ_1 =2.0 s

- Clearance lost time, $\ell_2 = 2.0$ seconds

For these characteristics, what is the capacity (per lane) for this movement?

The problem will be approached in two different ways. In the first, a ledger of time within the hour is created. Once the amount of time per hour used by vehicles at the saturation flow rate is established, capacity can be found by assuming that this time is used at a rate of one vehicle every *h* seconds. Since the characteristics stated are given on a *per phase* basis, these would have to be converted to a *per hour* basis. This is easily done knowing the number of signal cycles that occur within an hour. For a 60-second cycle, there are 3600/60=60 cycles within the hour. The subject movements will have one GREEN phase in each of these cycles. Then:

- Total Time in hour: 3,600 s
- RED time in hour: $(60–27–4)\times 60 = 1{,}740$ s
- Lost time in hour: $(2.0+2.0)\times 60 = 240$ s
- Remaining time in hour: $3{,}600–1{,}740–240 = 1{,}620$ s

The 1,620 remaining seconds of time in the hour represent the amount of time that can be used at a rate of one vehicle every *h* seconds, where h=2.4 s in this case. This number was calculated by deducting the periods during which no vehicles (in the subject movements) are effectively moving. These periods include the RED time as well as the start-up and clearance lost times in each signal cycle. The capacity of this movement may then be computed as:

$$\frac{1620}{2.4} = 675 \text{ veh/hg/ln}$$

A second approach to this problem utilizes Equations 18-6 and 18-8, with the following values:

$$s = \frac{3600}{2.4} = 1{,}500 \text{ veh/hg/ln} \quad g = 27+4-2-2 = 27 \text{ s} \quad c = 1{,}500\left(\frac{27}{60}\right) = 675 \text{ veh/hg/ln}$$

The two results are, as expected, the same. Capacity is found by isolating the effective green time available to the subject movements and by assuming that this time is used at the saturation flow rate (or headway).

18.2.6 Notable Studies on Saturation Headways, Flow Rates, and Lost Times

For purposes of illustrating basic concepts, subsequent sections of this chapter will assume that the value of saturation flow rate (or headway) is known. In reality, the saturation flow rate varies widely with a variety of prevailing conditions, including lane widths, heavy-vehicle presence, approach grades, parking conditions near the intersection, transit bus presence, vehicular and pedestrian flow rates, and other conditions.

The first significant studies of saturation flow were conducted by Bruce Greenshields in the 1940s [3]. His studies resulted in an average saturation flow rate of 1,714 veh/hg/ln and a start-up lost time of 3.7 s. The study, however, covered a variety of intersections with varying underlying characteristics. A later study in 1978 [4] reexamined the Greenshields hypothesis; it resulted in the same saturation flow rate (1,714 veh/hg/ln) but a lower start-up lost time of 1.1 s. The latter study had data from 175 intersections, covering a wide range of underlying characteristics.

A comprehensive study of saturation flow rates at intersections in five cities was conducted in 1987 and 1988 [5] to determine the effect of opposed left turns. It also produced, however, a good deal of data on saturation flow rates in general. Some of the results are summarized in Table 18.1.

Table 18.1: Saturation Flow Rates from a Nationwide Survey

Item	Single-Lane Approaches	Two-Lane Approaches
Number of Approaches	14	26
Number of 15-Minute Periods	101	156
Saturation Flow Rates		
Average	1,280 veh/hg/ln	1,337 veh/hg/ln
Minimum	636 veh/hg/ln	748 veh/hg/ln
Maximum	1,705 veh/hg/ln	1,969 veh/hg/ln
Saturation Headways		
Average	2.81 s/veh	2.69 s/veh
Minimum	2.11 s/veh	1.83 s/veh
Maximum	5.66 s/veh	4.81 s/veh

Table 18.1: Full Alternative Text

These results show generally lower saturation flow rates (and higher saturation headways) than previous studies. The data, however, reflect the impact of opposed left turns, truck presence, and a number of other “nonstandard” conditions, all of which have a significant impeding effect. The most remarkable result of this study, however, was the wide variation in measured saturation flow rates, both over time at the same site and from location to location. Even when underlying conditions remained fairly constant, the variation in observed saturation flow rates at a given location was as large as 20% to 25%. In a doctoral dissertation using the same data, Prassas demonstrated that saturation headways and flow rates have a significant stochastic component, making calibration of stable values difficult [6].

The study also isolated saturation flow rates for “ideal” conditions, which include all passenger cars, no turns, level grade, and 12-foot lanes. Even under these conditions, saturation flow rates varied from 1,240 to 2,092 pc/hg/ln for single-lane approaches, and from 1,668 to 2,361 pc/hg/ln for multilane approaches. The difference between observed saturation flow rates at single and multilane approaches is also interesting. Single-lane approaches have a number of unique characteristics that are addressed in the *Highway Capacity Manual* model for analysis of signalized intersections.

Current standards in the *Highway Capacity Manual* [1] use an ideal saturation flow rate of 1,900 pc/hg/ln for both single and multilane approaches (for cities with population >250,000, 1,750 pc/h/g elsewhere).

This ideal rate is then adjusted for a variety of prevailing conditions. The manual also provides default values for lost times. The default value for start-up lost time ($\ell 1$) is 2.0 s. For the clearance lost time ($\ell 2$), the default value varies with the “yellow” and “all-red” timings of the signal, as indicated in Equation 18-8.

18.3 The Critical-Lane and Time-Budget Concepts

In signal analysis and design, the “critical-lane” and “time budget” concepts are closely related. The time budget, in its simplest form, is the allocation of time to various vehicular and pedestrian movements at an intersection through signal control. Time is a constant: There are always 3,600 seconds in an hour, and all of them must be allocated. In any given hour, time is “budgeted” to legal vehicular and pedestrian movements and to lost times.

The “critical-lane” concept involves the identification of specific lane movements that will control the timing of a given signal phase. Consider the situation illustrated in Figure 18.2. A simple two-phase signal controls the intersection. Thus, all E–W movements are permitted during one phase, and all N–S movements are permitted in another phase. During each of these phases, there are four lanes of traffic (two in each direction) moving simultaneously. Demand is not evenly distributed among them; one of these lanes will have the most intense traffic demand. The signal must be timed to accommodate traffic in this lane—the “critical lane” for the phase.

Figure 18.2: Critical Lanes Illustrated

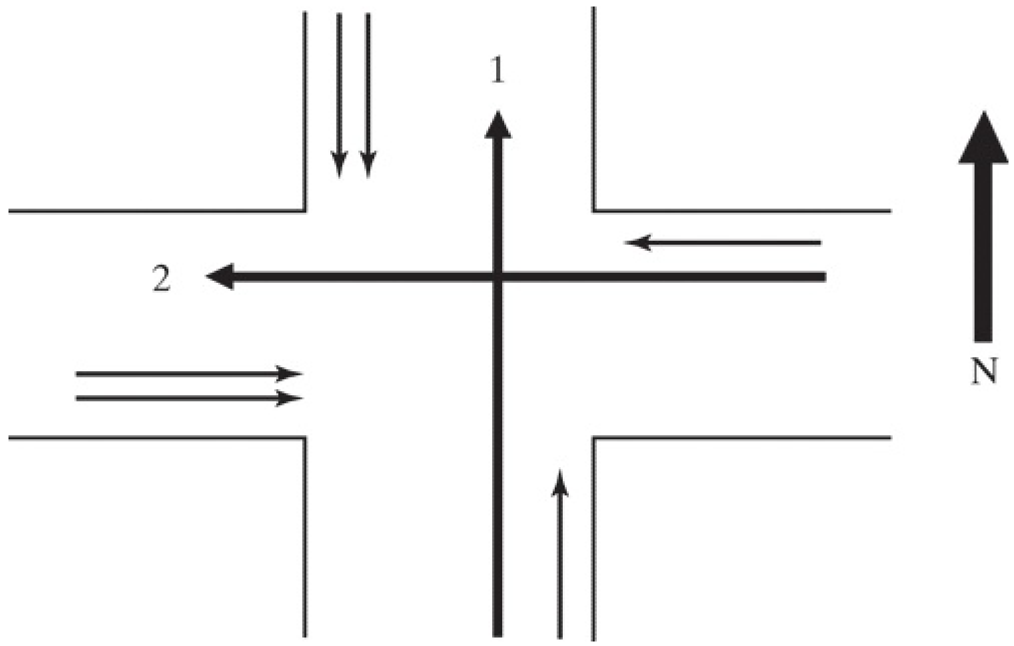


Figure 18.2: Full Alternative Text

In the illustration of Figure 18.2, the signal timing and design must accommodate the total demand flows in lanes 1 and 2. As these lanes have the most intense demand, if the signal accommodates them, all other lanes will be accommodated as well. Note that the critical lane is identified as the lane with the most *intense traffic demand*, not the lane with the highest volume. This is because there are many variables affecting traffic flow. A lane with many left-turning vehicles, for example, may require more time than an adjacent lane with no turning vehicles, but a higher volume. Determining the intensity of traffic demand in a lane involves accounting for prevailing conditions that may affect flow in that particular lane.

In establishing a time budget for the intersection of Figure 18.2, time would have to be allocated to four elements:

- Movement of vehicles in critical lane 1
- Movement of vehicles in critical lane 2
- Start-up and clearance lost times for vehicles in critical lane 1
- Start-up and clearance lost times for vehicles in critical lane 2

This can be thought of in the following way: lost times are not used by any vehicle. When deducted from total time, remaining time is effective green time and is allocated to critical-lane demands—in this case, in lanes 1 and 2. The total amount of effective green time, therefore, must be sufficient to accommodate the total demand in lanes 1 and 2 (the critical lanes). These critical demands must be accommodated one vehicle at a time, as they cannot move simultaneously.

The example of Figure 18.2 is a relatively simple case. In general, the following rules apply to the identification of critical lanes:

1. There is a critical lane and a critical-lane flow for each discrete signal phase provided.

2. Except for lost times, when no vehicles move, there must be *one* and *only one* critical lane moving during every second of effective green time in the signal cycle.

3. Where there are overlapping phases, the potential combination of lane flows yielding the highest sum of critical lane flows while preserving the requirement of item (b) identifies critical lanes.

Chapter 19 contains a detailed discussion of how to identify critical lanes for any signal timing and design.

18.3.1 The Maximum Sum of Critical-Lane Volumes: One View of Signalized Intersection Capacity

It is possible to consider the maximum possible sum of critical-lane volumes to be a general measure of the “capacity” of the intersection. This is not the same as the traditional view of capacity presented in the *Highway Capacity Manual*, but it is a useful concept to pursue.

By definition, each signal phase has one and only one critical lane. Except for lost times in the cycle, one critical lane is always moving. Lost times occur for each signal phase and represent time during which *no* vehicles in any lane are moving. The maximum sum of critical lane volumes may, therefore, be found by determining how much total lost time exists in the hour. The remaining time (total effective green time) may then be divided by the saturation headway.

To simplify this derivation, it is assumed that the total lost time per phase (tL) is a constant for all phases. Then, the total lost time per signal cycle is:

$$L=N\times t_L \quad [18\text{-}9]$$

where:

L = lost time per cycle, s/cycle; t_L = total lost time per phase (sum of $\ell_1 + \ell_2$), s/phase; and N = number of phases in the cycle

The total lost time in an hour depends upon the number of cycles occurring in the hour:

$$L_H=L\left(\frac{3600}{C}\right) \quad [18\text{-}10]$$

where:

L_H = lost time per hour, s/hr; L = lost time per cycle, s/cycle; and C = cycle length, s

The remaining time within the hour is devoted to effective green time for critical lane movements:

$$T_G = 3{,}600 - L_H \quad [18\text{-}11]$$

where:

T_G = total effective green time in the hour, s

This time may be used at a rate of one vehicle every *h* seconds, where *h* is the saturation headway:

$$V_c = \frac{T_G}{h} \quad [18\text{-}12]$$

where:

V_c = maximum sum of critical lane volumes, veh/h h = saturation headway, s/veh

Merging Equations 18-8 through 18-11, the following relationship emerges:

$$V_c = \frac{1}{h}\left[3{,}600 - N t_L \left(\frac{3{,}600}{C}\right)\right] \quad [18\text{-}13]$$

where all variables are as previously defined.

Sample Problem 18-2: Calculating the Maximum Sum of Critical Lane Volumes

Consider the example shown in Figure 18.2. If the signal at this location has two phases, a cycle length of 60 seconds, total lost times of 4 s/phase, and a saturation headway of 2.5 s/veh, what is the maximum allowable sum of critical lane flows?

The maximum sum of critical lane flows is found as:

$$V_c = \frac{1}{2.5}\left[3{,}600 - 2 \times 4 \times \left(\frac{3{,}600}{60}\right)\right] = 1{,}248 \text{ veh/h}$$

The equation indicates that there are 3,600/60=60 cycles in an hour. For each of these, 2×4=8 s of lost time is experienced, for a total of 8×60=480 s in the hour. The remaining 3,600−480=3,120 s may be used at a rate of one vehicle every 2.5 s.

If Equation 18-12 is plotted, an interesting relationship between the maximum sum of critical lane volumes (Vc), cycle length (*C*), and number of phases (*N*) may be observed, as illustrated in Figure 18.3.

Figure 18.3: Maximum Sum of Critical Lane Volumes Plotted

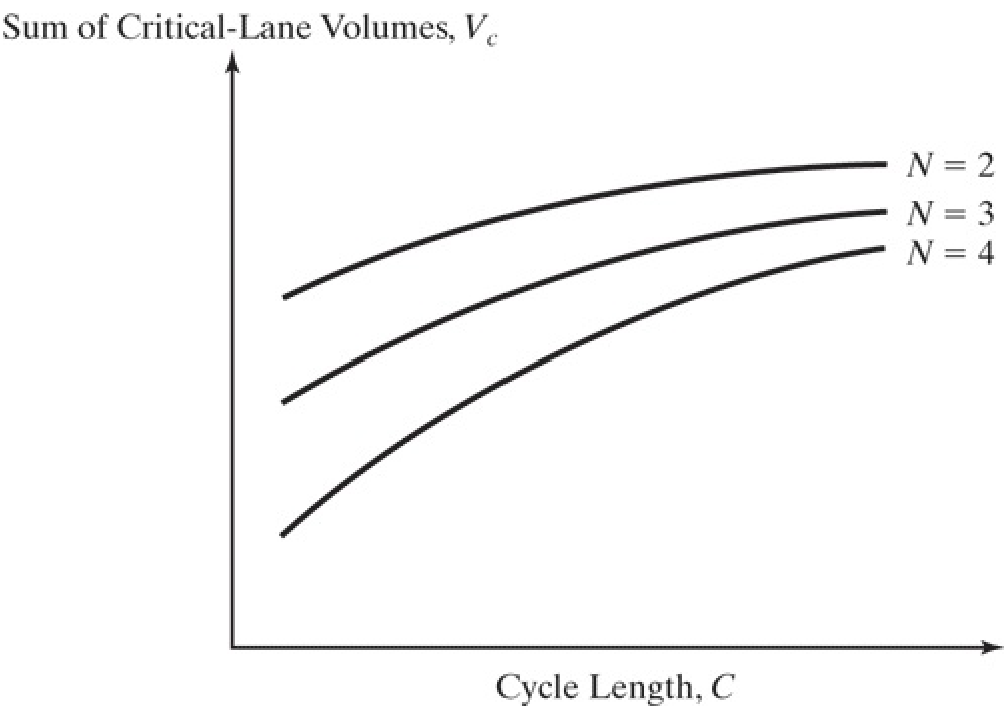


Figure 18.3: Full Alternative Text

As the cycle length increases, the “capacity” of the intersection also increases. This is because of lost times, which are constant per cycle. The longer the cycle length, the fewer cycles there are in an hour. This leads to less lost time in the hour, more effective green time in the hour, and a higher sum of critical-lane volumes. Note, however, that the relationship gets flatter as cycle length increases. As a general rule, increasing the cycle length may result in small increases in capacity. On the other hand, capacity can rarely be increased significantly by only increasing the cycle length. Other measures, such as adding lanes, are often also necessary.

Capacity also decreases as the number of phases increases. This is because for each phase, there is one full set of lost times in the cycle. Thus, a two-phase signal has only two sets of lost times in the cycle, while a three-phase signal has three.

These trends provide insight, but also raise an interesting question: Given these trends, it appears that all signals should have two phases and that the maximum practical cycle length should be used in all cases. After all, this combination would, apparently, yield the highest “capacity” for the intersection.

Using the maximum cycle length is not practical unless truly needed. Having a cycle length that is considerably longer than what is needed causes increases in delay to drivers and passengers. The increase in delay is because there will be times when vehicles on one approach are waiting for the green while there is no demand on conflicting approaches. Shorter cycle lengths yield less delay. Further, there is no incentive to maximize the cycle length. There will always be 3,600 seconds in the hour, and increasing the cycle length to accommodate increasing demand over time is quite simple, requiring only a resetting of the local signal controller. The shortest cycle length consistent with a v/c ratio in the range of 0.80 to 0.95 is generally used to produce optimal delays. Thus, the view of signal capacity is quite different from that of pavement capacity. When deciding on the number of lanes on a freeway (or on an intersection approach), it is desirable to build excess capacity (i.e., achieve a low v/c ratio). This is because once built, it is unlikely that engineers will get an opportunity to expand the facility for 20 or more years, and adjacent land development may make such expansion impossible. The 3,600 seconds in an hour, however, are immutable, and retiming the signal to allocate more of them to effective green time is a simple task requiring no field construction.

18.3.2 Finding an Appropriate Cycle Length

If it is assumed that the demands on an intersection are known and that the critical lanes can be identified, then Equation 18-12 could be solved using a known value of Vc to find a minimum acceptable cycle length:

$$C_{min} = \frac{N t_L}{1 - \left(\frac{V_c}{3{,}600/h}\right)} \quad [18\text{-}14]$$

Thus, if in the example of Figure 18.2, the actual sum of critical-lane volumes was determined to be 1,000 veh/h, the minimum feasible cycle length would be:

$$C_{min} = \frac{2 \times 4}{1 - \left(\frac{1{,}000}{3{,}600/2.5}\right)} = 26.2 \text{ s}$$

The cycle length could be reduced, in this case, from the given 60 to 30 seconds (the effective minimum cycle length used). This computation, however, assumes that the demand (Vc) is uniformly distributed throughout the hour and that every second of effective green time will be used. Neither of these assumptions is very practical. In general, signals would be timed for the flow rates occurring in the peak 15 minutes of the hour. Equation 18-13 could be modified by dividing Vc by a known peak-hour factor (PHF) to estimate the flow rate in the worst 15-minute period of the hour. Similarly, most signals would be timed to have somewhere between 80% and 95% of the available capacity actually used. Due to the normal stochastic variations in demand on a cycle-by-cycle and daily basis, some excess capacity must be provided to avoid failure of individual cycles or peak periods on a specific day. If demand, Vc, is also divided by the expected utilization of capacity (expressed in decimal form), then this is also accommodated. Introducing these changes transforms Equation 18-13 to:

$$C_{des} = \frac{N t_L}{1 - \left[\frac{1{,}000}{(3{,}600/h) \times PHF \times (v/c)}\right]} \quad [18\text{-}15]$$

where:

C_{des} = desirable cycle length, s; PHF = peak hour factor; and v/c = desired volume to capacity ratio

All other variables are as previously defined.

Sample Problem 18-3:

Determining a Desirable Cycle Length

Refer again to the intersection of Figure 18.1. If the PHF is 0.95 and it is desired to use no more than 90% of available capacity during the peak 15-minute period of the hour, what cycle length should be used? The desirable cycle length is found as:

$$C_{des} = \frac{2\times 4}{1-\left[\frac{1,000}{(3,600/2.5)\times 0.95\times 0.90}\right]} = \frac{8}{0.188} = 42.6\ \text{s}$$

In practical terms, this would lead to the use of a 45-second cycle length.

The relationship between a desirable cycle length, the sum of critical-lane volumes, and the target v/c ratio is quite interesting and is illustrated in Figure 18.4.

Figure 18.4: Desirable Cycle Length versus Sum of Critical Lane Volumes

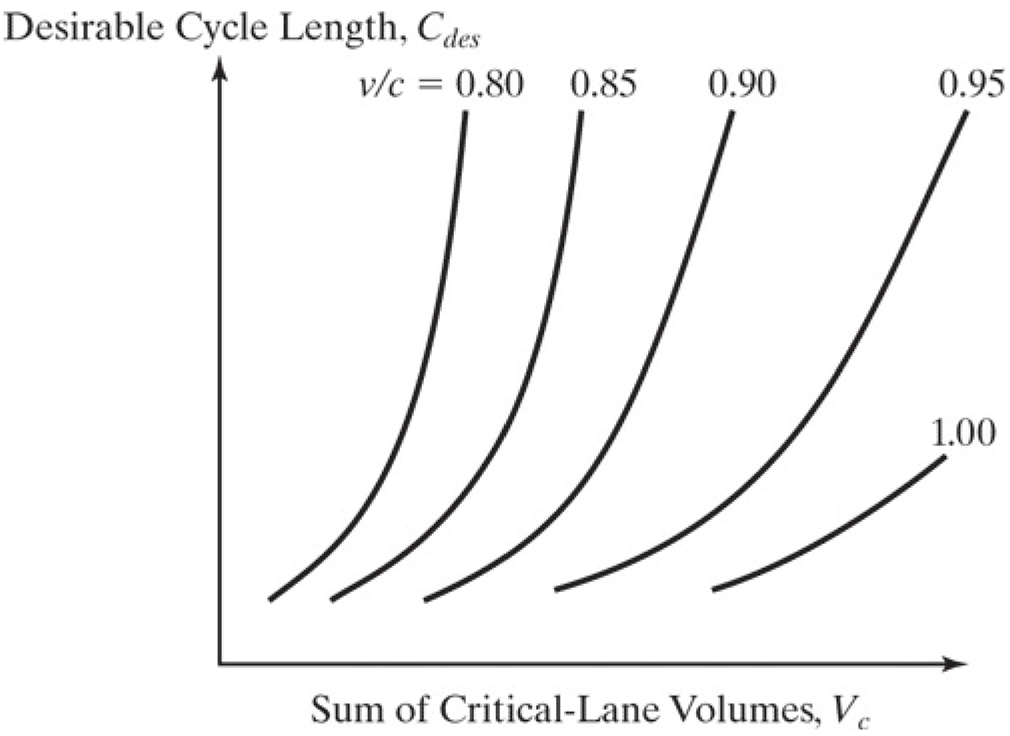


Figure 18.4: Full Alternative Text

Figure 18.4 illustrates a typical relationship for a specified number of phases, saturation headway, lost times, and peak-hour factor. If a vertical is drawn at any specified value of Vc (sum of critical lane volumes), it is clear that the resulting cycle length is very sensitive to the target v/c ratio. As the curves for each v/c ratio are eventually asymptotic to the vertical, it is not always possible to achieve a specified v/c ratio.

Sample Problem 18-4: Cycle Length vs. *v/c* Ratio

Consider the case of a three-phase signal, with $t_L=4$ s/phase, a saturation headway of 2.2 s/veh, a *PHF* of 0.90 and $V_c=1,200$ veh/h. Desirable cycle lengths will be computed for a range of target v/c ratios varying from 1.00 to 0.80.

$$C_{des} = \frac{3\times4}{1-\left[\frac{1,200}{(3,600/2.2)\times0.90\times1.00}\right]} = \frac{12}{0.1852} = 64.8 \Rightarrow 65 \text{ s}$$
$$C_{des} = \frac{3\times4}{1-\left[\frac{1,200}{(3,600/2.2)\times0.90\times0.95}\right]} = \frac{12}{0.1423} = 84.3 \Rightarrow 85 \text{ s}$$
$$C_{des} = \frac{3\times4}{1-\left[\frac{1,200}{(3,600/2.2)\times0.90\times0.90}\right]} = \frac{12}{0.0947} = 126.7 \Rightarrow 130 \text{ s}$$
$$C_{des} = \frac{3\times4}{1-\left[\frac{1,200}{(3,600/2.2)\times0.90\times0.85}\right]} = \frac{12}{0.0414}$$

$=289.9 \Rightarrow 290$ s $C_{des} = \frac{3 \times 4}{1-\left[\frac{1{,}200}{(3{,}600/2.2) \times 0.90 \times 0.80}\right]} = \frac{12}{-0.0185} = -648.6$ s

For this case, reasonable cycle lengths can provide target v/c ratios of 1.00 or 0.95. Achieving v/c ratios of 0.90 or 0.85 would require long cycle lengths beyond the practical limit of 120 s for pre-timed signals. The 130 s cycle needed to achieve a v/c ratio of 0.90 might be acceptable for an actuated signal location, or in some extreme cases warranting a longer pre-timed signal cycle. However, a v/c ratio of 0.80 cannot be achieved under any circumstances. The negative cycle length that results signifies that there is not enough time within the hour to accommodate the demand with the required green time plus the 12 s of lost time per cycle. In effect, more than 3,600 s would have to be available in the hour to accomplish this.

Sample Problem 18-5: Using the Time Budget and Critical Lane Concepts to Investigate Lane Requirements

Consider the intersection shown in Figure 18.5. The critical directional demands for this two-phase signal are shown with other key variables. Using the time-budget and critical-lane concepts, determine the number of lanes required for each of the critical movements and the minimum desirable cycle length that could be used.

Figure 18.5: Sample Problem Using the Time Budget and Critical Lane Concepts

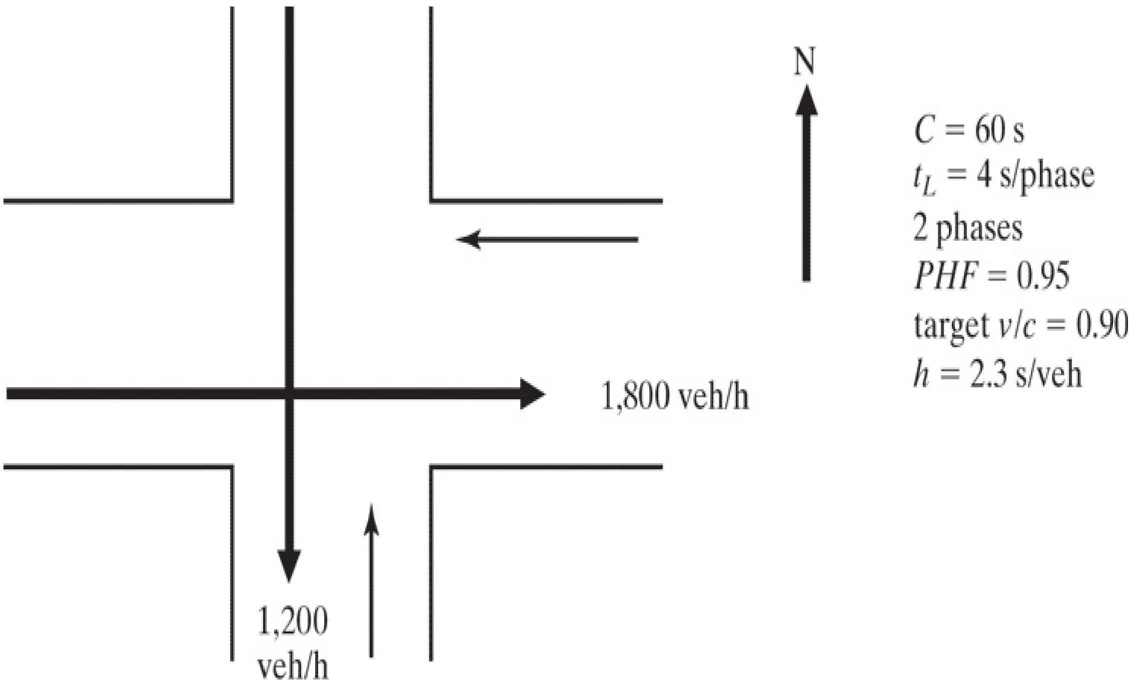


Figure 18.5: Full Alternative Text

Assuming that the initial specification of a 60-s cycle is correct and given the other specified conditions, the maximum sum of critical lanes that can be accommodated is computed using [Equation 18-12]:

$$V_c = \frac{1}{2.3}\left[3{,}600 - 2 \times 4 \times \left(\frac{3{,}600}{60}\right)\right] = 1{,}357 \text{ vehs/h}$$

The critical SB volume is 1,200 veh/h, and the critical EB volume is 1,800 veh/h. The number of lanes each must be divided into is now to be determined. Whatever combination is used, the sum of the critical-lane volumes for these two approaches must be below 1,357 veh/h. [Figure 18.6] shows a number of possible lane combinations and the resulting sum of critical lane volumes. As can be seen from the scenarios of [Figure 18.6], in order to have a sum of critical-lane volumes less than 1,357 veh/h, the SB approach must have at least two lanes, and the EB approach must have three lanes. Realizing that these demands probably reverse in the other peak hour (AM or PM), the N–S artery would probably require four lanes, and the E–W artery six lanes.

Figure 18.6: Possible Lane Scenarios for Sample Problem

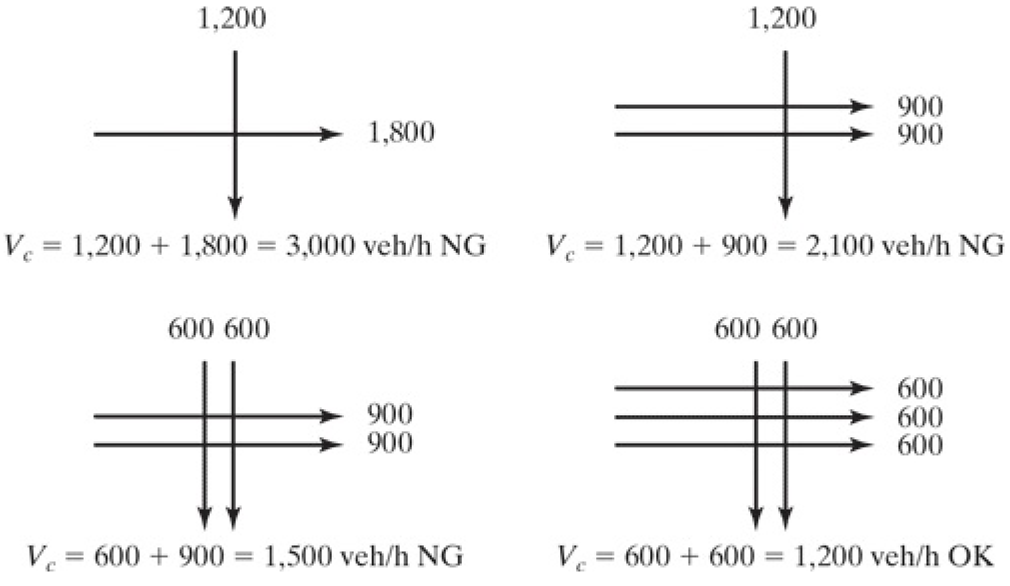


Figure 18.6: Full Alternative Text

If this scenario is built, Vc is only 1,200 veh/h. It is possible that the original cycle length of 60 s could be reduced. A minimum desirable cycle length may be computed from Equation 18-14:

$$C_{des} = \frac{2\times 4}{1-\left[\frac{1,200}{(3,600/2.3)\times 0.95\times 0.90}\right]} = 77.7 \Rightarrow 80 \text{ s}$$

The resulting cycle length is *larger* than the original 60 s because the equation takes both the *PHF* and target v/c ratios into account. Equation 18-12 for computing the maximum value of Vc does not; it assumed full use of capacity (v/c=1.00) and no peaking within the hour. In essence, the 2×3 lane design proposal should be combined with an 80 s cycle length to achieve the desired results.

Sample Problem 18-5 illustrates the critical relationship between number of lanes and cycle lengths. Clearly, there are other scenarios that would produce desirable results. Additional lanes could be provided in either direction, which would allow the use of a shorter cycle length. Unfortunately, for many cases, signal timing is considered with a fixed design already in place. Only where right-of-way is available or a new intersection is being constructed can major changes in the number of lanes be considered. Allocation of lanes to various movements is also a consideration. Optimal solutions are generally found more easily when the physical design and signalization can be treated in tandem.

Sample Problem 18-6: investigates another potential solution

If, in the problem of Figure 18.5, space limited both the EB and SB approaches to two lanes, the resulting Vc would be 1,500 veh/h. Would it be possible to accommodate this demand by lengthening the cycle length? Equation 18-14 is used:

$$C_{des} = 2\times 4\ 1-[\ 1{,}500\ (3{,}600/2.3)\times 0.95\times 0.90\ ] = 8\ -0.121 = -66.1\ \text{s NG}$$

The negative result indicates that there is no cycle length that can accommodate a Vc of 1,500 veh/h at this location.

18.4 The Concept of Left-Turn (and Right-Turn) Equivalency

The most difficult process to model at a signalized intersection is the left turn. Left turns are made in several different modes using different design elements. Left turns may be made from a lane shared with through vehicles (shared-lane operation) or from a lane dedicated to left-turning vehicles (exclusive-lane operation). Traffic signals may allow for permitted or protected left turns, or some combination of the two.

Whatever the case, however, a left-turning vehicle will consume more effective green time traversing the intersection than will a similar through vehicle. The most complex case is that of a permitted left turn made across an opposing vehicular flow from a shared lane. A left-turning vehicle in the shared lane must wait for an acceptable gap in the opposing flow. While waiting, the vehicle blocks the shared lane, and other vehicles (including through vehicles) in the lane are delayed behind it. Some vehicles will change lanes to avoid the delay, while others are unable to and must wait until the left-turner successfully completes the turn.

Many models of the signalized intersection account for this in terms of "through vehicle equivalents" (i.e., how many through vehicles would consume the same amount of effective green time traversing the stop-line as *one* left-turning vehicle?). Consider the situation depicted in Figure 18.7. If both the left lane and the right lane were observed, an equivalence similar to the following statement could be determined:

Figure 18.7: Sample Equivalence Observation on a Signalized Intersection Approach

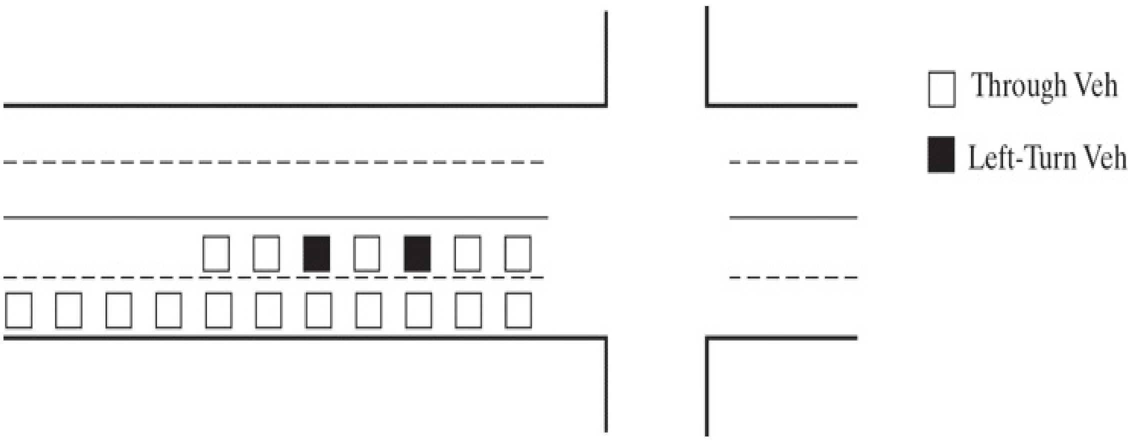


Figure 18.7: Full Alternative Text

> In the same amount of time, the left lane discharges five through vehicles and two left-turning vehicles, while the right lane discharges eleven through vehicles.

In terms of effective green time consumed, this observation means that eleven through vehicles are equivalent to five through vehicles plus two left turning vehicles. If the left-turn equivalent is defined as E_{LT}:

$$11 = 5 + 2E_{LT} \qquad E_{LT} = \frac{11-5}{2} = 3.0$$

It should be noted that this computation holds only for the prevailing characteristics of the approach during the observation period. The left-turn equivalent depends upon a number of factors, including how left turns are made (protected, permitted, compound), the opposing traffic flow, and the number of opposing lanes. Figure 18.8 illustrates the general form of the relationship for through vehicle equivalents of *permitted* left turns.

Figure 18.8: Relationship Among Left-Turn Equivalents, Opposing Flow, and Number of Opposing Lanes

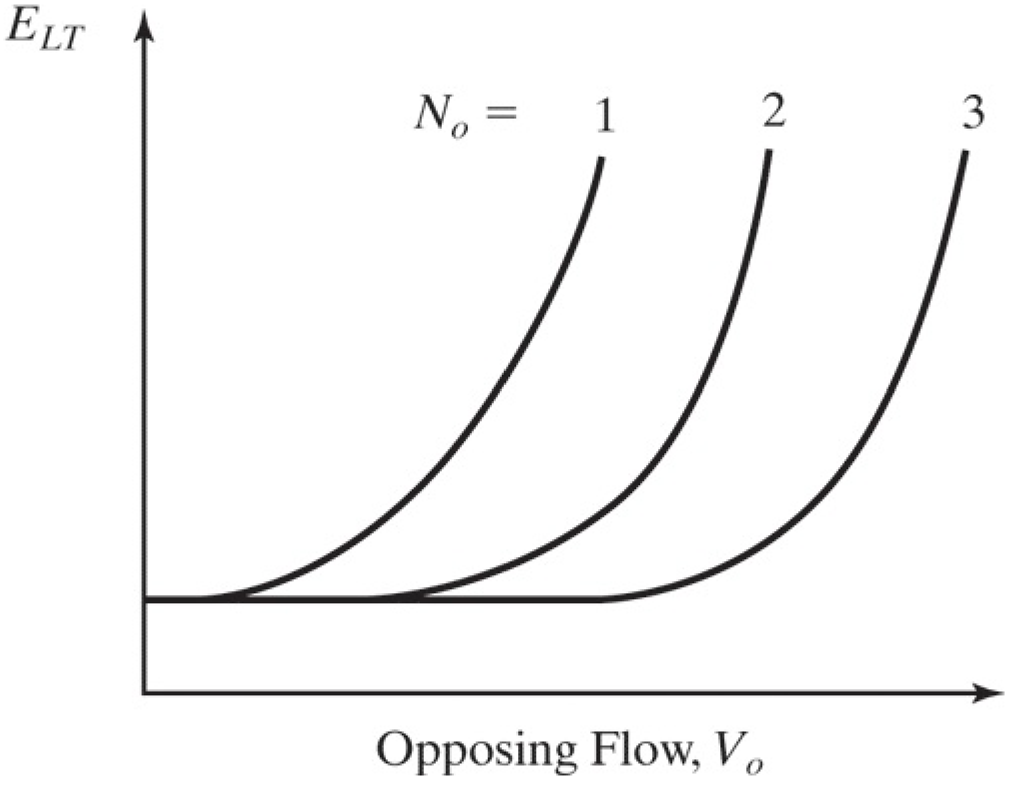


Figure 18.8: Full Alternative Text

The left-turn equivalent, ELT, increases as the opposing flow increases. For any given opposing flow, however, the equivalent decreases as the number of opposing lanes is increased from one to three. This latter relationship is not linear, as the task of selecting a gap through multilane opposing traffic is more difficult than selecting a gap through single-lane opposing traffic. Further, in a multilane traffic stream, vehicles do not pace each other side-by-side, and the gap distribution does not improve as much as the per-lane opposing flow decreases. To illustrate the use of left-turn equivalents in modeling, consider Sample Problem 18-7.

Sample Problem 18-7: Left-Turn Equivalency

An approach to a signalized intersection has two lanes, permitted left-turn phasing, 10% left-turning vehicles, and a left-turn equivalent of 5.0. The saturation headway for through vehicles is 2.0 s/veh. Determine the equivalent saturation flow rate and headway for all vehicles on this approach.

The left-turn equivalent may be interpreted to mean that each left-turning vehicle consumes 5.0 times the effective green time as a through vehicle. Thus, for the situation described, 10% of the traffic stream has a saturation headway of $2.0\times5.0=10.0$s/veh, while the remainder (90%) have a saturation headway of 2.0 s/veh. The average saturation headway for all vehicles is, therefore:

$h=(0.10\times10.0)+(0.90\times2.0)=2.80$ s/veh

This corresponds to a saturation flow rate of:

$s=3,6002.80=1,286$ veh/hg/ln

A number of models, including the *Highway Capacity Manual* approach, calibrate a multiplicative adjustment factor that converts an ideal (or through) saturation flow rate to a saturation flow rate for prevailing conditions:

$s_{prev} = s_{ideal} \times f_{LT}$ $f_{LT} = s_{prev}$ $s_{ideal} = (3,600/h_{prev})$ $(3,600/h_{ideal})$ $= h_{ideal}$ h_{prev} [18-16]

where:

s_{prev} = saturation flow rate under prevailing conditions, veh/hg/ln; s_{ideal} = saturation flow rate under ideal conditions, veh/hg/ln; f_{LT} = left-turn adjustment factor; h_{ideal} = saturation headway under ideal conditions, s/veh; and h_{prev} = saturation headway under prevailing conditions, s/veh

In effect, in the first solution, the prevailing headway, hprev, was computed as follows:

$$h_{prev} = (P_{LT} E_{LT} h_{ideal}) + [(1 - P_{LT}) h_{ideal}] \quad [18\text{-}17]$$

Combining Equations 18-15 and 18-16:

$$f_{LT} = \frac{h_{ideal}}{(P_{LT} E_{LT} h_{ideal}) + [(1 - P_{LT}) h_{ideal}]} \qquad f_{LT} = \frac{1}{P_{LT} E_{LT} + (1 - P_{LT})} = \frac{1}{1 + P_{LT}(E_{LT} - 1)} \quad [18\text{-}18]$$

Sample Problem 18-8: Defining the Left-Turn Adjustment Factor

Sample Problem 18-7 may now be solved using a left-turn adjustment factor. Note that the saturation headway under ideal conditions is 3,600/2.0=1,800 veh/hg/ln. Then:

$$f_{LT} = \frac{1}{1 + 0.10(5 - 1)} = 0.714 \qquad s_{prev} = 1800 \times 0.714 = 1{,}286 \text{ veh/hg/ln}$$

The result is the same as in Sample Problem 18-7.

It is important that the concept of left-turn equivalence be understood. Its use in multiplicative adjustment factors often obscures its intent and meaning. The fundamental concept, however, is unchanged—the equivalence is based on the fact that the effective green time consumed by a left-turning vehicle is E_{LT} times the effective green time consumed by a similar through vehicle.

A similar case can be made for describing the effects of right turns. Right turns are typically made through a conflicting pedestrian flow in the crosswalk to the immediate right of the approach. Like left turns, this interaction causes right turns to consume more effective green time than through movements. An equivalent, E_{RT}, is used to quantify these effects, and is used in the same manner as described for left-turn equivalents.

Signalized intersection and other traffic models use other types of equivalents as well. Heavy-vehicle and local-bus equivalents have similar meanings and result in similar equations. Some of these have been discussed in previous chapters, and others will be discussed in subsequent chapters.

18.5 Delay as a Measure of Effectiveness

Signalized intersections represent point locations within a surface street network. As point locations, the measures of operational quality or effectiveness used for highway sections are not relevant. Speed has no meaning at a point, and density requires a section of some length for measurement. A number of measures have been used to characterize the operational quality of a signalized intersection, the most common of which are:

- Delay
- Queuing
- Stops

These measures are all related. Delay refers to the amount of time consumed in traversing the intersection—the difference between the arrival time and the departure time, where these may be defined in a number of different ways. Queuing refers to the number of vehicles forced to queue behind the stop-line during a RED signal phase; common measures include the average queue length or a percentile queue length. Stops refer to the percentage or number of vehicles that must stop at the signal.

18.5.1 Types of Delay

The most common measure used to describe operational quality at a signalized intersection is delay, with queuing and/or stops often used as a secondary measure. While it is possible to measure delay in the field, it is a difficult process, and different observers may make judgments that could yield different results. For many purposes, it is, therefore, convenient to have a predictive model for the estimate of delay. Delay, however, can be quantified in many different ways. The most frequently used forms of delay are defined as follows:

1. Stopped-time delay. Stopped-time delay is defined as the time a vehicle is stopped in queue while waiting to pass through the intersection; average stopped-time delay is the average for all vehicles during a specified time period.

2. Approach delay. Approach delay includes stopped-time delay but adds the time loss due to deceleration from the approach speed to a stop and the time loss due to reacceleration back to the desired speed. Average approach delay is the average for all vehicles during a specified time period.

3. Time-in-queue delay. Time-in-queue delay is the total time from a vehicle joining an intersection queue to its discharge across the STOP line on departure. Again, average time-in-queue delay is the average for all vehicles during a specified time period.

4. Travel time delay. This is a more conceptual value. It is the difference between the driver's expected travel time through the intersection (or any roadway segment) and the actual time taken. Given the difficulty in establishing a "desired" travel time to traverse an intersection, this value is rarely used, other than as a philosophic concept.

5. Control delay. The concept of control delay was developed in the 1994 *Highway Capacity Manual*, and is included in the current HCM. It is the delay caused by a control device, either a traffic signal or a STOP-sign. It is approximately equal to time-in-queue delay plus the acceleration-deceleration delay component.

Figure 18.9 illustrates three of these delay types for a single vehicle approaching a RED signal.

Figure 18.9: Illustration of Delay Measures

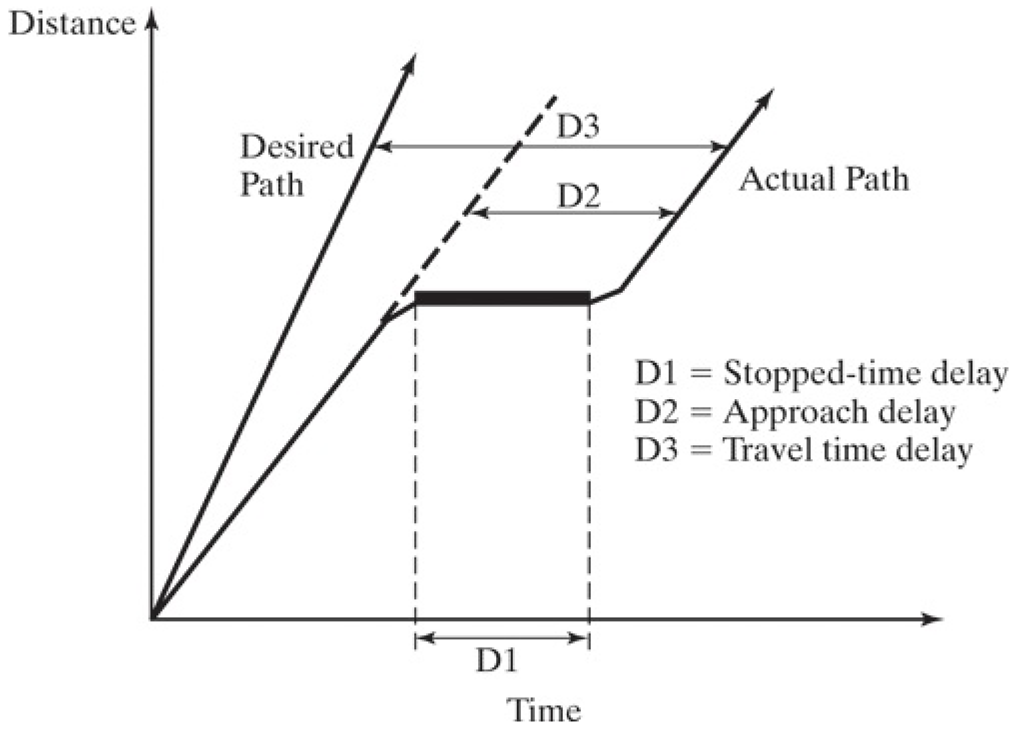


Figure 18.9: Full Alternative Text

Stopped-time delay for this vehicle includes only the time spent stopped at the signal. It begins when the vehicle is fully stopped and ends when the vehicle begins to accelerate. Approach delay includes additional time losses due to deceleration and acceleration. It is found by extending the velocity slope of the approaching vehicle as if no signal existed; the approach delay is the horizontal (time) difference between the hypothetical extension of the approaching velocity slope and the departure slope after full acceleration is achieved. Travel time delay is the difference in time between a hypothetical desired velocity line and the actual vehicle path. Time-in-queue delay cannot be effectively shown using one vehicle, as it involves joining and departing a queue of several vehicles.

Delay measures can be stated for a single vehicle, as an average for all vehicles over a specified time period, or as an aggregate total value for all vehicles over a specified time period. Aggregate delay is measured in total *vehicle-seconds*, *vehicle-minutes*, or *vehicle-hours* for all vehicles in the specified time interval. Average individual delay is generally stated in terms of s/veh for a specified time interval.

18.5.2 Basic Theoretical Models of

Delay

Virtually all analytic models of delay begin with a plot of cumulative vehicles arriving and departing vs. time at a given signal location. The time axis is divided into periods of effective green and effective red as illustrated in Figure 18.10.

Figure 18.10: Delay, Waiting Time, and Queue Length Illustrated

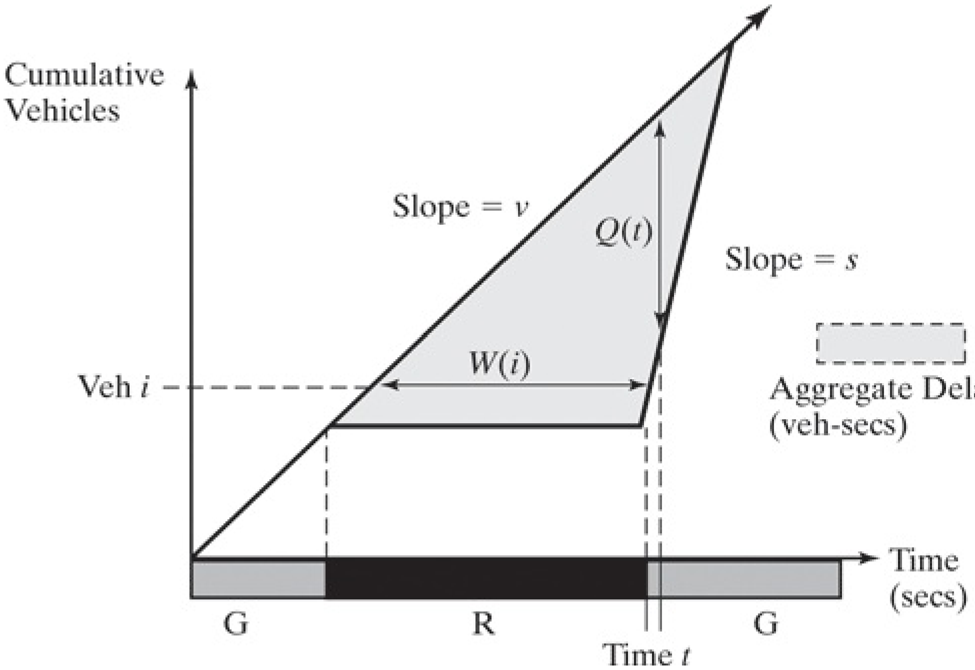


Figure 18.10: Full Alternative Text

Vehicles are assumed to arrive at a uniform rate of flow of v vehicles per unit time, seconds in this case. This is shown by the constant slope of the arrival curve. Uniform arrivals assume that the inter-vehicle arrival time between vehicles is a constant. Thus, if the arrival flow rate, v, is 1,800 vehs/h, then one vehicle arrives every 3,600/1,800=2.0s.

Assuming no pre-existing queue, vehicles arriving when the light is

GREEN continue through the intersection (i.e., the departure curve is the same as the arrival curve). When the light turns RED, however, vehicles continue to arrive, but none depart. Thus, the departure curve is parallel to the *x*-axis during the RED interval. When the next effective GREEN begins, vehicles queued during the RED interval depart from the intersection, now at the saturation flow rate, *s*, in veh/s. For stable operations, depicted here, the departure curve "catches up" with the arrival curve before the next RED interval begins (i.e., there is no residual or unserved queue left at the end of the effective GREEN).

This simple depiction of arrivals and departures at a signal allows the estimation of three critical parameters:

- The total time that any vehicle *i* spends waiting in the queue, $W(i)$, is given by the horizontal time-scale difference between the time of arrival and the time of departure.
- The total number of vehicles queued at any time *t*, $Q(t)$, is the vertical vehicle-scale difference between the number of vehicles that have arrived and the number of vehicles that have departed.
- The aggregate delay for all vehicles passing through the signal is the area between the arrival and departure curves (vehicles×time).

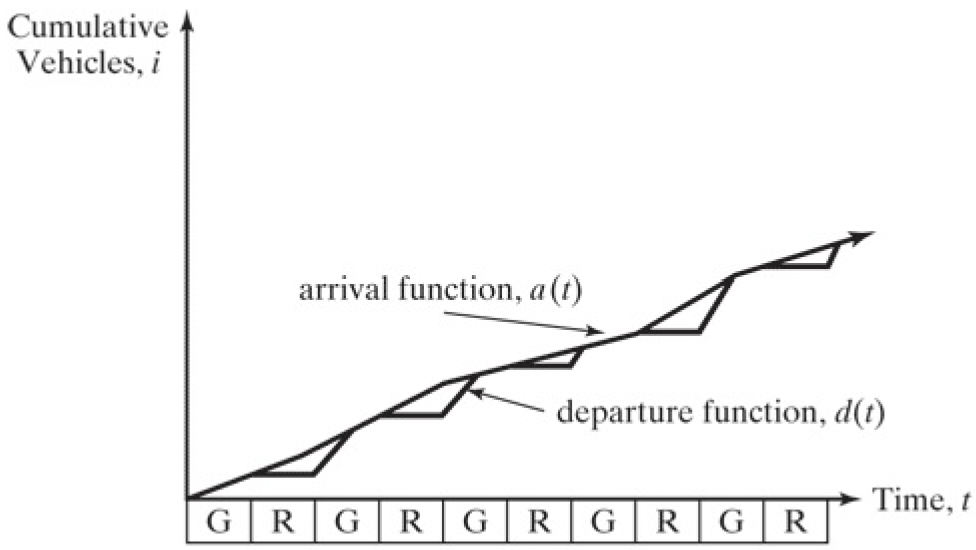
Note that since the plot illustrates vehicles arriving in queue and departing from queue, this model most closely represents what has been defined as *time-in-queue delay*. There are many simplifications that have been assumed, however, in constructing this simple depiction of delay. It is important to understand the two major simplifications:

- The assumption of a uniform arrival rate is a simplification. Even at a completely isolated location, actual arrivals would be *random* i.e., would have an average rate over time, but inter-vehicle arrival times would vary around an average rather than being constant. Within coordinated signal systems, however, vehicle arrivals are in platoons.
- It is assumed that the queue is building at a point location (as if vehicles were stacked on top of one another). In reality, as the queue grows, the rate at which vehicles arrive at its end is the arrival rate of vehicles (at a point), plus a component representing the backward growth of the queue in space.

Both of these can have a significant effect on actual results. Modern models account for the former in ways that will be discussed subsequently. The assumption of a “point queue,” however, is embedded in many modern applications.

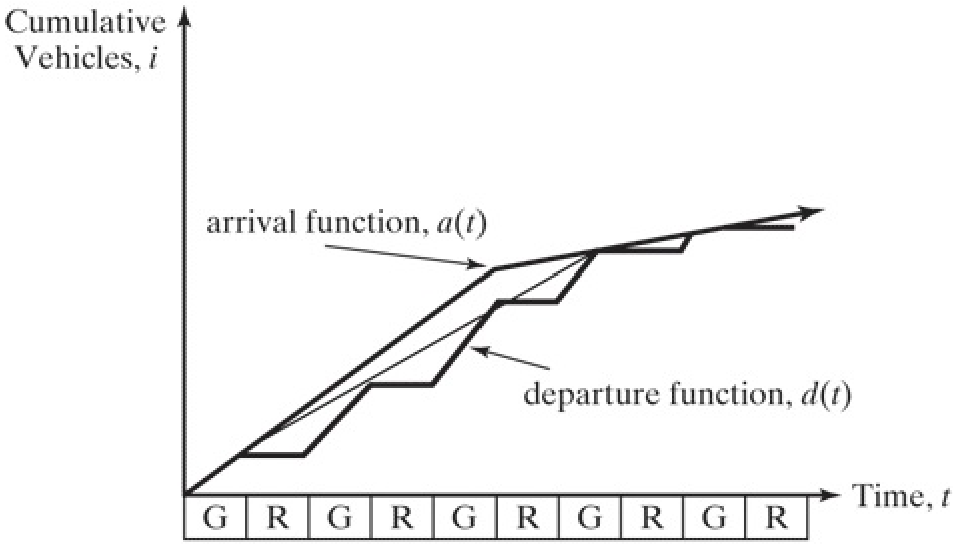
Figure 18.11 expands the range of Figure 18.10 to show a series of GREEN phases and depicts three different types of operation. It also allows for an arrival function, $a(t)$, that varies, while maintaining the departure function, $d(t)$, described previously.

Figure 18.11: Three Delay Scenarios



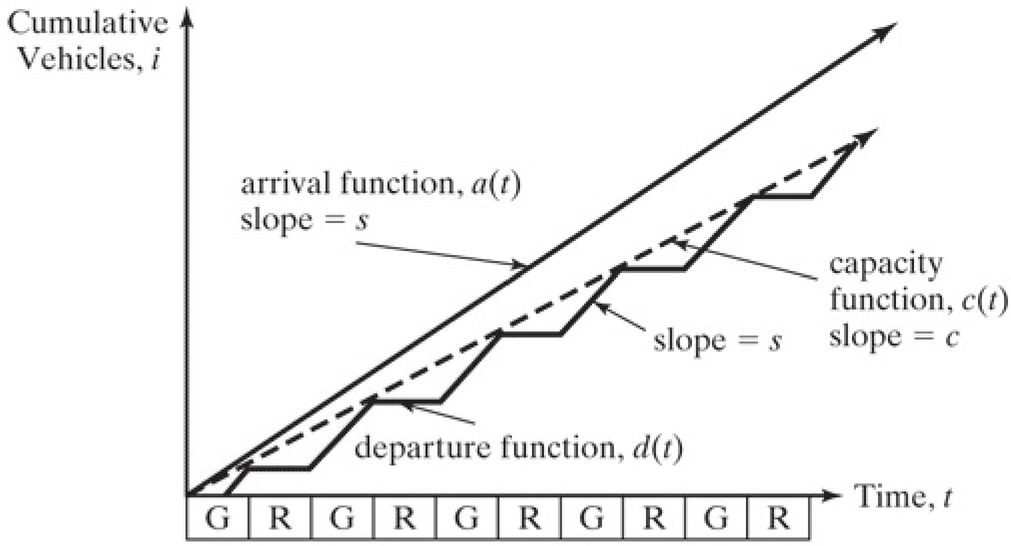
(a) Stable Flow

18.5-2 Full Alternative Text



(b) Individual Cycle Failures Within a Stable Operation

18.5-2 Full Alternative Text



(c) Demand Exceeds Capacity for a Significant Period

18.5-2 Full Alternative Text

(Source: Adapted with permission of Transportation Research Board, National Research Council, Washington, D.C., from V.F. Hurdle, “Signalized Intersection Delay Model: A Primer for the Uninitiated,” *Transportation Research Record 971*, 1984, pgs 97, 98.)

Figure 18.11(a) shows stable flow throughout the period depicted. No signal cycle “fails” (i.e., ends with some vehicles queued during the preceding RED unserved). During every GREEN phase, the departure function “catches up” with the arrival function. Total aggregate delay during this period is the total of all the triangular areas between the arrival and departure curves. This type of delay is often referred to as “uniform delay.”

In Figure 18.11(b), some of the signal phases “fail.” At the end of the second and third GREEN intervals, some vehicles are not served (i.e., they must wait for a second GREEN interval to depart the intersection). By the time the entire period ends, however, the departure function has “caught up” with the arrival function and there is no residual queue left unserved. This case represents a situation in which the overall period of analysis is stable (i.e., total demand does not exceed total capacity). Individual cycle failures within the period, however, have occurred. For these periods, there is a second component of delay in addition to uniform delay. It consists of the area between the arrival function and the dashed line, which represents the capacity of the intersection to discharge vehicles, and has the slope *c*. This type of delay is referred to as “overflow delay.”

Figure 18.11(c) shows the worst possible case: Every GREEN interval “fails” for a significant period of time, and the residual, or unserved, queue of vehicles continues to grow throughout the analysis period. In this case, the overflow delay component grows over time, quickly dwarfing the uniform delay component.

The latter case illustrates an important practical operational characteristic. When demand exceeds capacity ($v/c>1.00$), the delay depends upon *the length of time* that the condition exists. In Figure 18.11(b), the condition exists for only two phases. Thus, the queue and the resulting overflow delay is limited. In Figure 18.11(c), the condition exists for a long time, and the delay continues to grow throughout the oversaturated period.

Components of Delay

In analytic models for predicting delay, there are three distinct components of delay that may be identified:

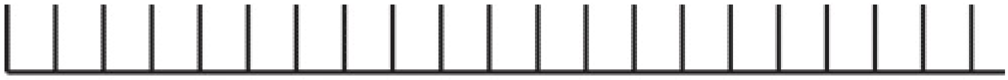
- Uniform delay is the delay based on an assumption of uniform arrivals and stable flow with no individual cycle failures.
- Random delay is the additional delay, above and beyond uniform delay, because flow is randomly distributed rather than uniform at isolated intersections.
- Overflow delay is the additional delay that occurs when the capacity of an individual phase or series of phases is less than the demand or arrival flow rate.

In addition, the delay impacts of platoon flow (rather than uniform or random) are treated as an adjustment to uniform delay. Many modern models combine the random and overflow delays into a single function, which is referred to as "overflow delay," even though it contains both components.

The differences between uniform, random, and platooned arrivals are illustrated in Figure 18.12. As noted, the analytic basis for most delay models is the assumption of uniform arrivals, which are depicted in Figure 18.12(a). Even at isolated intersections, however, arrivals would be random, as shown in Figure 18.12(b). With random arrivals, the underlying rate of arrivals is a constant, but the inter-arrival times are exponentially distributed around an average. In most urban and suburban cases, where a signalized intersection is likely to be part of a coordinated signal system, arrivals will be in organized platoons that move down the arterial in a cohesive group, as shown in Figure 18.12(c). The exact time that a platoon arrives at a downstream signal has an enormous potential effect on delay. A platoon of vehicles arriving at the beginning of the RED forces most vehicles to stop for the entire length of the RED phase. The same platoon of vehicles arriving at the beginning of the GREEN phase may flow through the intersection without any vehicles stopping. In both cases, the arrival flow, v, and the capacity of the intersection, *c*, are the same. The resulting delay, however, would vary significantly. The existence of platoon arrivals, therefore, necessitates a significant adjustment to models based on theoretically uniform or random flow.

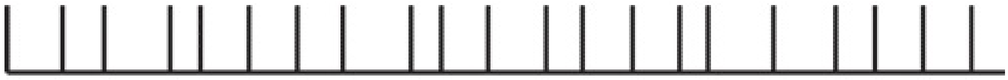
Figure 18.12: Arrival Patterns

Compared



(a) Uniform Arrivals

18.5-2 Full Alternative Text



(b) Random Arrivals

18.5-2 Full Alternative Text

(c) Reality = Platooned Arrivals – No Theoretical Solution Available

18.5-2 Full Alternative Text

Webster’s Uniform Delay Model

Virtually every model of delay starts with Webster’s model of uniform delay. Initially published in 1958 [7], this model begins with the simple illustration of delay depicted in Figure 18.13, with its assumptions of stable flow and a simple uniform arrival function. As noted previously, aggregate delay can be estimated as the area between the arrival and departure curves in the figure. Thus, Webster’s model for uniform delay is the area of the triangle formed by the arrival and departure functions. For clarity, this triangle is shown again in Figure 18.13.

Figure 18.13: Webster’s Uniform Delay Model

Illustrated

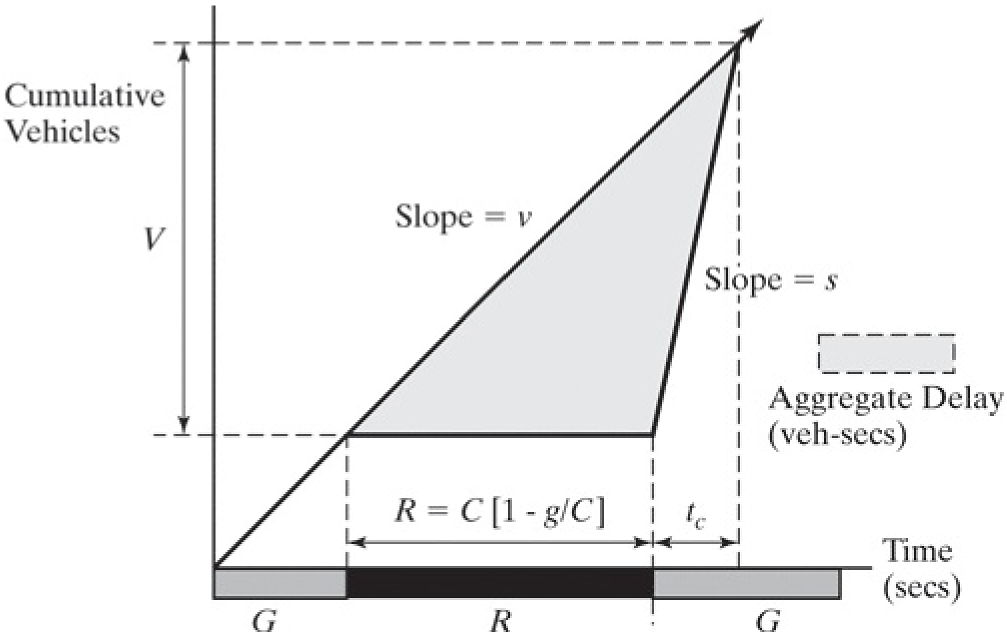


Figure 18.13: Full Alternative Text

The area of the aggregate delay triangle is simply one-half the base times the height, or:

$UDa=12RV$

where:

U D a = aggregate uniform delay, veh-secs; R =
length of the RED phase, s; and V = total vehicles in queue, vehs

By convention, traffic models are not developed in terms of RED time. Rather, they focus on GREEN time. Thus, Webster substitutes the following equivalence for the length of the RED phase:

$R=C[\ 1-(\ g\ C\ )\ ]$

where:

C = cycle length, s g = effective green time, s

In words, the RED time is the portion of the cycle length that is not effectively green.

The height of the triangle, *V*, is the total number of vehicles in the queue. In effect, it includes vehicles arriving during the RED phase, *R*, plus those that join the end of the queue while it is moving out of the intersection (i.e., during time tc, in Figure 18.13). Thus, determining the time it takes for the queue to clear, tc, is an important part of the model. This is done by setting the number of vehicles arriving during the period R+tc equal to the number of vehicles departing during the period tc, or:

$$v(R+t_c)=s\,t_c \qquad R+t_c=\left(\frac{s}{v}\right)t_c \qquad R=t_c\left[\left(\frac{s}{v}\right)-1\right] \qquad t_c=\frac{R}{\left[\left(\frac{s}{v}\right)-1\right]}$$

Then, substituting for tc:

$$V=v(R+t_c)=v\left[R+\frac{R}{\left(\frac{s}{v}-1\right)}\right]=R\left(\frac{vs}{s-v}\right)$$

and for *R*:

$$V=C\left[1-\left(\frac{g}{C}\right)\right]\left[\frac{vs}{s-v}\right]$$

Then, aggregate delay can be stated as:

$$UDa=\frac{1}{2}RV=\frac{1}{2}C^2\left[1-\left(\frac{g}{C}\right)\right]\left[\frac{vs}{s-v}\right] \quad [18\text{-}19]$$

where all variables are as previously defined.

Equation 18-18 estimates aggregate uniform delay in vehicle-seconds for one signal cycle. To get an estimate of average uniform delay per vehicle, the aggregate is divided by the number of vehicles arriving during the cycle, vC. Then:

$$UD=\frac{\frac{1}{2}C\left[1-\frac{g}{C}\right]^2}{\left[1-\frac{v}{s}\right]} \quad [18\text{-}20]$$

Another form of the equation uses the capacity, *c*, rather than the saturation flow rate, *s*. Noting that s=c/(g/C), the following form emerges:

$$UD=\frac{\frac{1}{2}C\left[1-\left(\frac{g}{C}\right)\right]^2}{\left[1-\left(\frac{g}{C}\right)\left(\frac{v}{c}\right)\right]}=\frac{0.50C\left[1-\left(\frac{g}{C}\right)\right]^2}{1-\left(\frac{g}{C}\right)X} \quad [18\text{-}21]$$

where:

UD = average uniform delay per vehicles, s/veh; C = cycle length, s; g = effective green time, s; v = arrival flow rate, veh/h; c =

capacity of intersection approach, veh/h; and X =
v/c ratio, or degree of saturation

This average includes the vehicles that arrive and depart on green, accruing no delay. This is appropriate. One of the objectives in signalizations is to minimize the number or proportion of vehicles that must stop. Any meaningful quality measure would have to include the positive impact of vehicles that are not delayed.

In Equation 18-20, it must be noted that the maximum value of *X* (the v/c ratio) is 1.00. As the uniform delay model assumes no overflow, the v/c ratio cannot be more than 1.00.

Modeling Random Delay

The uniform delay model assumes that arrivals are uniform and that no signal phases fail (i.e., that arrival flow is less than capacity during every signal cycle of the analysis period).

At isolated intersections, vehicle arrivals are more likely to be random. A number of stochastic models have been developed for this case, including those by Newall [8], Miller [9, 10], and Webster [7]. Such models assume that inter-vehicle arrival times are distributed according to the Poisson distribution, with an underlying average arrival rate of *v* vehicles/unit time. The models account for both the underlying randomness of arrivals and the fact that some individual cycles within a demand period with v/c<1.00 could fail due to this randomness. This additional delay is sometimes referred to as “overflow delay,” but it does not address situations in which v/c>1.00 for the entire analysis period. This text refers to additional delay due to randomness as “random delay,” *RD*, to distinguish it from true overflow delay when v/c>1.00. The most frequently used model for random delay is Webster’s formulation:

$RD=\frac{X^2}{2v(1-X)}$ [18-22]

where:

RD = average random delay per vehicle, s/veh X = v/c ratio

This formulation was found to somewhat overestimate delay, and Webster proposed that total delay (the sum of uniform and random delay) be estimated as:

$D=0.90(UD+RD)$ [18-23]

where:

D=sum of uniform and random delay

Modeling Overflow Delay

“Oversaturation” is used to describe extended time periods during which arriving vehicles exceed the capacity of the intersection approach to discharge vehicles. In such cases, queues grow, and overflow delay, in addition to uniform delay, accrues. As overflow delay accounts for the failure of an extended series of phases, it encompasses a portion of random delay as well.

Figure 18.14 illustrates a time period for which $v/c>1.00$. Again, as in the uniform delay model, it is assumed that the arrival function is uniform. During the period of oversaturation, delay consists of both uniform delay (in the triangles between the capacity and departure curves) and overflow delay (in the growing triangle between the arrival and capacity curves). The formula for the uniform delay component may be simplified in this case, as the v/c ratio (*X*) is the maximum value of 1.00 for the uniform delay component. Then:

$UD_o=0.50C[1-(gC)]21-(gC)X=0.50C[1-(gC)]21-(gC)1.00=0.50C[1-(gC$ [18-24]

Figure 18.14: An Oversaturated Period Illustrated

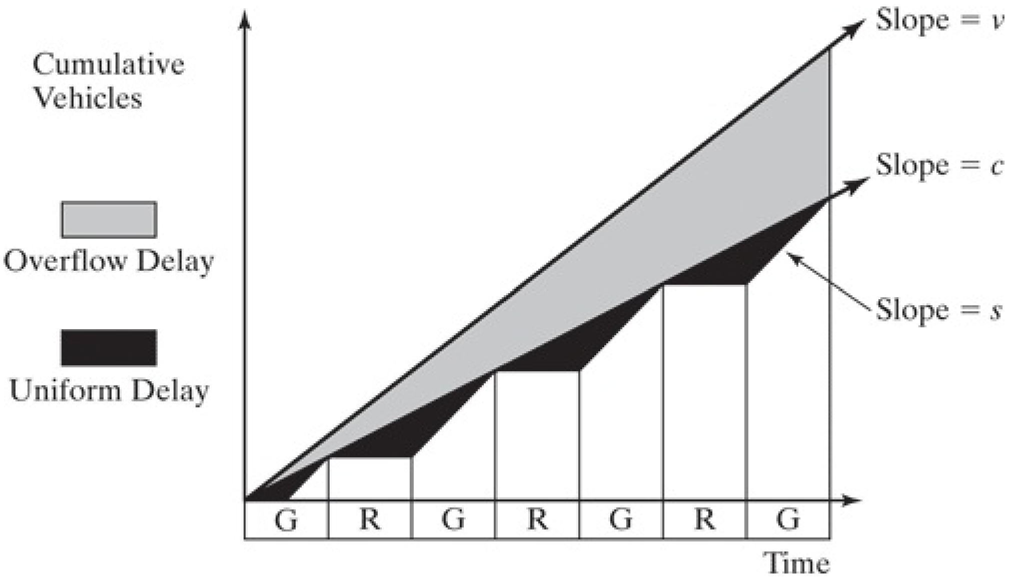


Figure 18.14: Full Alternative Text

To this, the overflow delay must be added. Figure 18.15 illustrates how the overflow delay is estimated. The aggregate and average overflow delay can be estimated as:

$$OD_a = \frac{1}{2}T(vT - cT) = \frac{T^2}{2}(v - c) \qquad OD = \frac{T}{2}[X - 1] \qquad [18\text{-}25]$$

Figure 18.15: Derivation of the Overflow Delay Formula

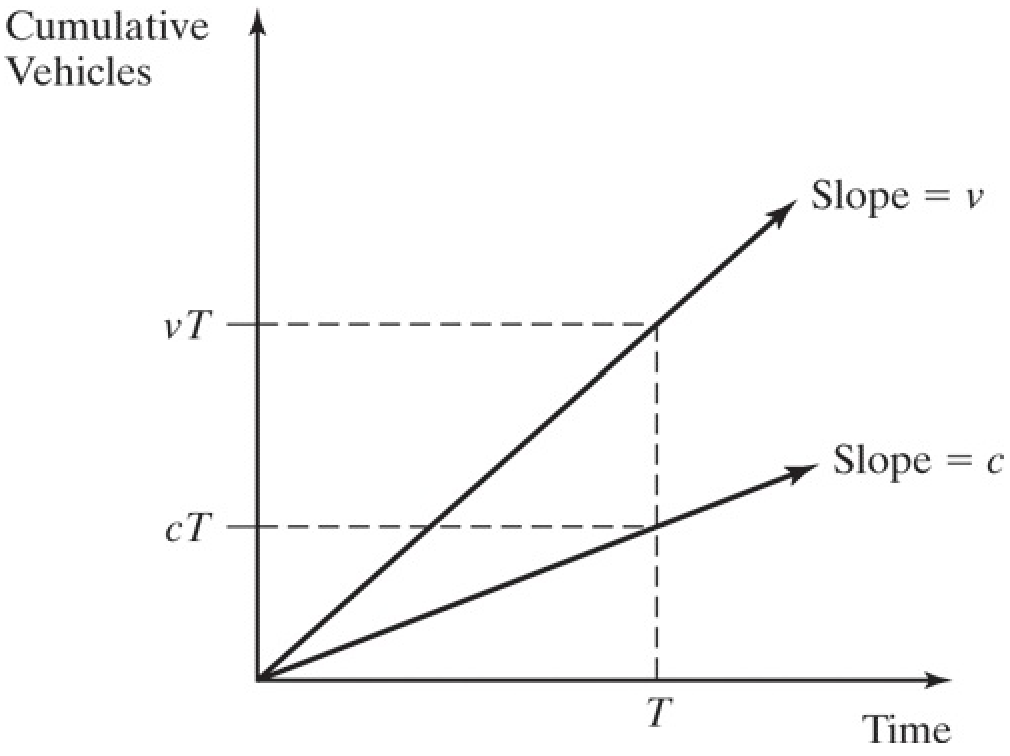


Figure 18.15: Full Alternative Text

where:

O D a = aggregate overflow delay, veh-secs OD =
average overflow delay per vehicle, s/veh

Other parameters are as previously defined.

In Equation 18-24, the average overflow delay is obtained by dividing the aggregate delay by *the number of vehicles discharged within time T, cT*. Unlike the formulation for uniform delay, where the number of vehicles arriving and the number of vehicles discharged during a cycle were the same, the overflow delay triangle includes vehicles that arrive within time *T*, but are not discharged within time *T*. The delay triangle, therefore, includes only the delay accrued by vehicles through time *T*, and excludes additional delay that vehicles still "stuck" in the queue will experience after time *T*.

Equation 18-24 may use any unit of time for "T." The resulting overflow delay, *OD*, will have the same units as specified for *T*, on a per-vehicle basis.

Equation 18-24 is time-dependent (i.e., the longer the period of oversaturation exists, the larger delay becomes). The predicted delay per vehicle is averaged over the entire period of oversaturation, *T*. This masks, however, a significant issue: vehicles arriving early during time *T* experience far less delay than vehicles arriving later during time *T*. A model for average overflow delay during a time period T1 through T2 may be developed, as illustrated in Figure 18.16. Note that the delay area formed is a trapezoid, not a triangle.

Figure 18.16: A Model for Overflow Delay Between Times T_1 and T_2

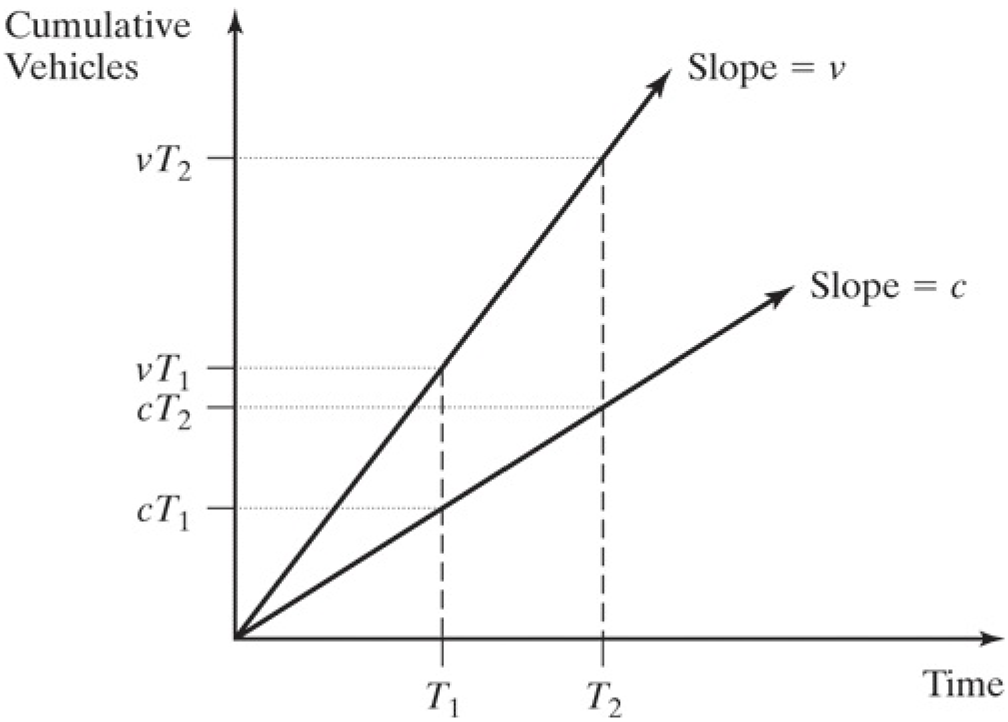


Figure 18.16: Full Alternative Text

The resulting model for average delay per vehicle during the time period T1 through T2 is:

OD=T1+T22(X−1) [18-26]

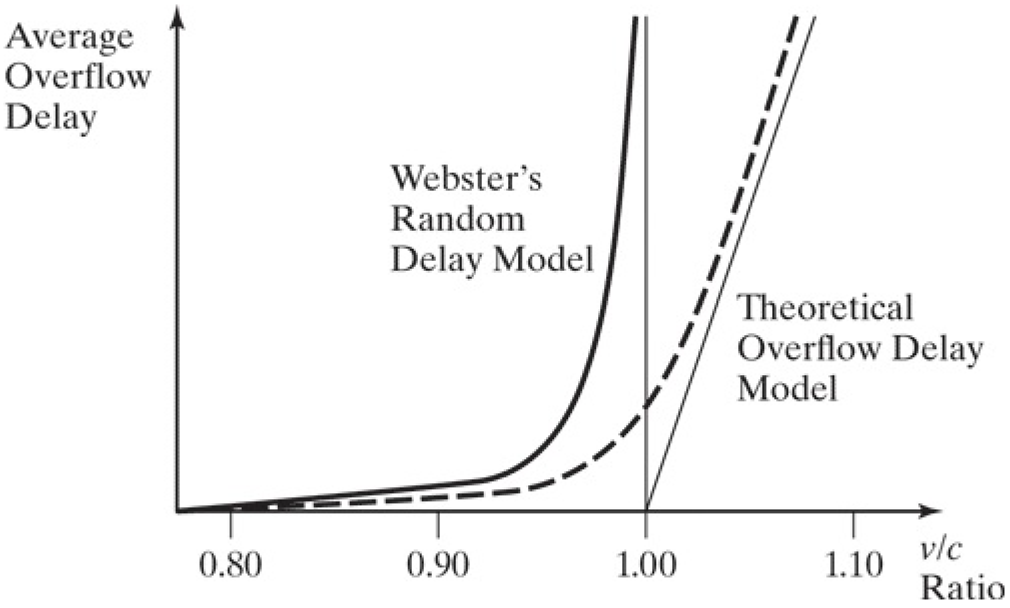
where all terms are as previously defined. Note that the trapezoidal shape of the delay area results in the T1+T2 formulation, emphasizing the growth of delay as the oversaturated condition continues over time. Also, this formulation predicts the average delay per vehicle that occurs during the specified interval, T1 through T2. Thus, delays to vehicles arriving before time T1 but discharging after T1 are included only to the extent of their delay within the specified times, not any delay they may have experienced in queue before T1. Similarly, vehicles discharging after T2 do have a delay component after T2 that is not included in the formulation.

The three varieties of delay—uniform, random, and overflow delay—can be modeled in relatively simple terms as long as simplifying assumptions are made in terms of arrival and discharge flows, and in the nature of the queuing that occurs, particularly during periods of oversaturation. The next section begins to consider some of the complications that arise from the direct use of these simplified models.

18.5.3 Inconsistencies in Random and Overflow Delay

Figure 18.17 illustrates a basic inconsistency in the random and overflow delay models previously discussed. The inconsistency occurs when the v/c ratio (*X*) is in the vicinity of 1.00. When the v/c ratio is below 1.00, a random delay model is used, as there is no “overflow” delay in this case. Webster’s random delay model (Equation 18-22), however, contains the term (1-*X*) in the denominator. Thus, as *X* approaches a value of 1.00, random delay increases asymptotically to an infinite value. When the v/c ratio (*X*) is greater than 1.00, an overflow delay model is applied. The overflow delay model of Equation 18-25, however, has an overflow delay of 0 when X=1.00, and increases uniformly with increasing values of *X* thereafter.

Figure 18.17: Random and Overflow Delay Models Compared



(Source: Adapted with permission of Transportation Research Board, National Research Council, Washington, D.C., from Hurdle, V.F. "Signalized Intersection Delay Model: A Primer for the Uninitiated, *Transportation Research Record 971*, 1984, pg 101.)

Figure 18.17: Full Alternative Text

Neither model is accurate in the immediate vicinity of v/c=1.00. Delay does not become infinite at v/c=1.00. There is no true "overflow" at v/c=1.00, although individual cycle failures due to random arrivals do occur. Similarly, the overflow model, with overflow delay=0.0 s/veh at v/c=1.00 is also unrealistic. The additional delay of individual cycle failures due to the randomness of arrivals is not reflected in this model.

In practical terms, most studies confirm that the uniform delay model is a sufficient predictive tool (except for the issue of platooned arrivals) when the v/c ratio is 0.85 or less. In this range, the true value of random delay is minuscule, and there is no overflow delay. Similarly, the simple theoretical overflow delay model (when added to uniform delay) is a reasonable predictor when v/c≥1.15 or so. The problem is that the most interesting cases fall in the intermediate range (0.85<v/c<1.15), for which neither model is adequate. Much of the more recent work in delay modeling involves attempts to bridge this gap, creating a model that closely follows the uniform delay model at low v/c ratios, and approaches the theoretical overflow delay model at high v/c ratios (≥1.15), producing "reasonable" delay estimates in between. Figure 18.17 illustrates this as the dashed line.

The most commonly used model for bridging this gap was developed by Akcelik for the Australian Road Research Board’s signalized intersection analysis procedure [11, 12]:

$OD = cT\ 4\ [\ (X-1) + (X-1)\ 2 + (\ 12(X - X_o)\ cT\ )\ ]\ X_o = 0.67 + (\ sg\ 600\ )$ [18-27]

where:

T = analysis period, h; X = v/c ratio; c = capacity, veh/h; s = saturation flow rate, veh/sg, (vehs per second of green); and g = effective green time, s

The only relatively recent study resulting in large amounts of delay measurements in the field was conducted by Reilly, *et al*. [13] in the early 1980s to calibrate a model for use in the 1985 edition of the *Highway Capacity Manual*. The study concluded that Equation 18-26 substantially overestimated field-measured values of delay and recommended that a factor of 0.50 be included in the model to adjust for this. The version of the delay equation that was included in the 1985 *Highway Capacity Manual* ultimately did not follow this recommendation, and included other empiric adjustments to the theoretical equation.

18.5.4 Delay Models in the HCM

The *Highway Capacity Manual* has used some form of delay as the measure of effectiveness since 1985. Many changes, however, have occurred over the years in the specific models and approaches taken to estimate delay, and indeed, in the specific delay measure used. In 1985, stopped-time delay was used directly. Uniform delay was estimated using a version of Webster’s Uniform Delay equation, and overflow delay was estimated using a version of Akcelik’s equation modified to fit a substantial delay data base. In 1994, stopped-time delay was abandoned in favor of control delay, and the equations were modified to reflect this. Many detailed changes were incorporated by 2000. In 2010, Webster’s Uniform Delay model was replaced by an Incremental Queue Accumulation model. This model breaks down the arrival and departure curves in very small units of time throughout the signal cycle. A version of Akcelik’s model was still used (although significantly revised) to estimate overflow delay. For 2016, some additional revisions have been made, but the basic approach remained the same.

In the final analysis, all delay modeling is based on the determination of the area between an arrival curve and a departure curve on a plot of cumulative vehicles vs. time. As the arrival and departure functions are permitted to become more complex and as rates are permitted to vary for various sub-parts of the signal cycle, the models become more complex as well.

18.5.5 Sample Problems in Delay Estimation

Sample Problem 18-9: Delay Estimation (1)

Consider the following situation: An intersection approach has an approach flow rate of 1,000 veh/h, a saturation flow rate of 2,800 veh/hg, a cycle length of 90 s, and a *g/C* ratio of 0.55. What average delay per vehicle is expected under these conditions?

Solution

To begin, the capacity and v/c ratio for the intersection approach must be computed. This will determine what model(s) are most appropriate for application in this case:

$$c=s\left(\frac{g}{C}\right)=2{,}800\times0.55=1{,}540\ \text{veh/h} \qquad v/c=X=\frac{1{,}000}{1{,}540}=0.649$$

As this is a relatively low value, the uniform delay equation (Equation 18-19) may be applied directly. There is little random delay at such a v/c ratio and no overflow delay to consider. Thus:

$$d=\left(\frac{C}{2}\right)\frac{\left[1-\left(\frac{g}{C}\right)\right]^2}{1-\left(\frac{v}{s}\right)}=\left(\frac{90}{2}\right)\frac{(1-0.55)^2}{1-\left(\frac{1{,}000}{2{,}800}\right)}=14.2\ \text{s/veh}$$

Note that this solution assumes that arrivals at the subject intersection approach are random. Platooning effects are not taken into account.

Sample Problem 18-10: Delay Estimation (2)

How would the above result change if the demand flow rate increased to 1,600 veh/h for a one-hour period?

Solution

In this case, the v/c ratio now changes to 1,600/1,540=1.039. This is in the difficult range of 0.85–1.15 for which neither the simple random flow model nor the simple overflow delay model are accurate. The Akcelik model of Equation 18-26 will be used. Total delay, however, includes both uniform delay and overflow delay. The uniform delay component when v/c>1.00 is given by Equation 18-23:

$$UD=0.50C\left[1-\left(\frac{g}{C}\right)\right]=0.50\times 90\times(1-0.55)=20.3\text{ s/veh}$$

Use of Akcelik's overflow delay model requires that the analysis period be selected or arbitrarily set. Using a one-hour time period, as specified, then:

$$OD=\frac{cT}{4}\left[(X-1)+\sqrt{(X-1)^2+\left(\frac{12(X-X_o)}{cT}\right)}\right]$$

$$X_o=0.67+\left(\frac{sg}{600}\right)=0.67+\left(\frac{0.778\times 49.5}{600}\right)=0.734$$

$$OD=\frac{1{,}540\times 1}{4}\left[(1.039-1)+\sqrt{(1.039-1)^2+\left(\frac{12\times(1.039-0.734)}{1540\times 1}\right)}\right]=39.1\text{ s/veh}$$

where:

$$g=0.55\times 90=49.5\text{ s}$$
$$s=2{,}800/3{,}600=0.778\text{ veh/sg}$$

In this case, even with the "overflow" quite small (approximately 4% of the demand flow), the additional average delay due to this overflow is considerable. The total expected delay in this situation is the sum of the uniform and overflow delay terms, or:

$$d=20.3+39.1=59.4\text{ s/veh}$$

Note that this computation, as in Sample Problem 18-9, assumes random arrivals on this intersection approach.

Sample Problem 18-11: Delay Estimation (3)

How would the result change if the demand flow rate increased to 1,900 veh/h over a two-hour period?

Solution

The v/c ratio in this case is now 1,900/1,540=1.23. In this range, the simple theoretical overflow model is an adequate predictor. As in Sample Problem 18-10, the Uniform Delay component must also be included; this computation is the same as in Sample Problem 18-10: 20.3 s/veg.

The overflow delay component may be estimated using the simple theoretical Equation 18-24:

$OD=T2(X-1)=7,2002(1.23-1)=828.0$ s/veh

As the period of oversaturation is given as two hours, and a result in seconds is desired, *T* is entered as $2\times 3,600=7,200$ s. The total delay experienced by the average motorist is the sum of uniform and overflow delay, or:

$d=20.3+828.0=848.3$ s/veh

This is a very large value but represents an average over the full two-hour period of oversaturation. Equation 18-25 may be used to examine the average delay to vehicles arriving in the first 15 minutes of oversaturation to those arriving in the last 15 minutes of oversaturation:

$O D$ first 15 $= T 1 + T 2 2 (X-1) = 0+900 2 (1.23-1)=103.5$ s/veh $O D$ last 15 $= 6,300+7,200 2 (1.23-1)=1,552.5$ s/veh

As previously noted, the delay experienced during periods of oversaturation is very much influenced by the length of time that oversaturated operations have prevailed. Total delay for each case would also include the 20.3 s/veh of uniform delay. As in Sample Problems 18-9 and 18-10, random arrivals are assumed.

These simple problems apply only the most basic theoretical delay models to illustrate fundamental approaches, and the importance of delay to assessing the quality of operations at a signalized intersection. More detailed analyses could be made to take into account incremental queue analysis and the impacts of progression on estimated delays.

18.6 Closing Comments

This chapter has reviewed four key concepts necessary to understand the operation of signalized intersections:

1. Saturation flow rate and lost times
2. The time budget and critical lanes
3. Left-turn (and right-turn) equivalency
4. Delay as a measure of effectiveness

These fundamental concepts are also the critical components of models of signalized intersection analysis. In Chapters 19 and 20, some of these concepts are implemented in a simple methodology for signal timing. In Chapter 22, all are used as parts of the HCM analysis procedure for signalized intersections.

References

- 1. *Manual of Uniform Traffic Control Devices*, Millennium Edition, Federal Highway Administration, U.S. Department of Transportation, Washington, D.C., 2003.

- 2. *Highway Capacity Manual*, 6th Edition: A Guide for Multimodal Mobility Analysis, Transportation Research Board, National Research Council, Washington, D.C., 2016.

- 3. Greenshields, B., "Traffic Performance at Intersections," *Yale Bureau Technical Report No. 1*, Yale University, New Haven, CT, 1947.

- 4. Kunzman, W., "Another Look at Signalized Intersection Capacity," *ITE Journal*, Institute of Transportation Engineers, Washington, D.C., August 1978.

- 5. Roess, R., et al., "Level of Service in Shared, Permissive Left-Turn Lane Groups," *Final Report*, FHWA Contract DTFH-87-C-0012, Transportation Training and Research Center, Polytechnic University, Brooklyn, NY, September 1989.

- 6. Prassas, E., "Modeling the Effects of Permissive Left Turns on Intersection Capacity," *Doctoral Dissertation*, Polytechnic University, Brooklyn, NY, December 1994.

- 7. Webster, F., "Traffic Signal Settings," *Road Research Paper No. 39*, Road Research Laboratory, Her Majesty's Stationery Office, London, UK, 1958.

- 8. Newall, G., "Approximation Methods for Queues with Application to the Fixed-Cycle Traffic Light," *SIAM Review*, Vol. 7, 1965.

- 9. Miller, A., "Settings for Fixed-Cycle Traffic Signals," *ARRB Bulletin 3*, Australian Road Research Board, Victoria, Australia, 1968.

- 10. Miller, A., "The Capacity of Signalized Intersections in

Australia," *ARRB Bulletin 3*, Australian Road Research Board, Victoria, Australia, 1968.

- 11. Akcelik, R., "Time-Dependent Expressions for Delay, Stop Rate, and Queue Lengths at Traffic Signals," *Report No. AIR 367-1*, Australian Road Research Board, Victoria, Australia, 1980.

- 12. Akcelik, R., "Traffic Signals: Capacity and Timing Analysis," *ARRB Report 123*, Australian Road Research Board, Victoria, Australia, March 1981.

- 13. Reilly, W. and Gardner, C., "Technique for Measuring Delay at Intersections," *Transportation Research Record 644*, Transportation Research Board, National Research Council, Washington, D.C., 1977.

Problems

1. 18-1. Consider the headway data shown in the following table. Data was taken from the center lane of a three-lane intersection approach for a total of 10 signal cycles. For the purposes of this analysis, the data may be considered to have been collected under ideal conditions.

    1. Plot the average headways vs. position in queue for the data shown. Sketch an approximate best-fit curve through the data.

    2. Using the approximate best-fit curve constructed in (a), determine the saturation headway and the start-up lost time for the data.

    3. What is the saturation flow rate for this data?

Data for Problem 1

Q Pos	Headways (s) for Cycle No.									
	1	2	3	4	5	6	7	8	9	10
1	3.6	3.7	3.5	3.6	3.4	3.3	3.6	3.7	3.5	3.7
2	2.6	3.0	2.4	2.6	2.2	2.2	2.7	2.8	2.7	2.8
3	2.0	2.4	2.0	2.1	1.8	2.0	2.4	2.4	2.3	2.4
4	1.7	2.0	2.0	2.0	1.7	2.0	2.0	2.0	2.0	2.0
5	1.6	1.9	2.0	1.9	1.7	1.8	1.9	2.1	2.0	1.8
6	1.7	1.8	1.9	1.9	1.6	1.7	1.8	1.8	1.8	1.7
7	1.7	1.8	1.8	1.8	1.7	1.7	1.7	1.9	1.9	1.9
8	1.6	1.7	1.8	1.8	1.7	1.7	1.8	1.8	1.9	1.9
9	x	1.8	x	1.6	x	1.7	1.7	x	x	1.7
10	x	1.7	x	1.8	x	x	1.7	x	X	1.7

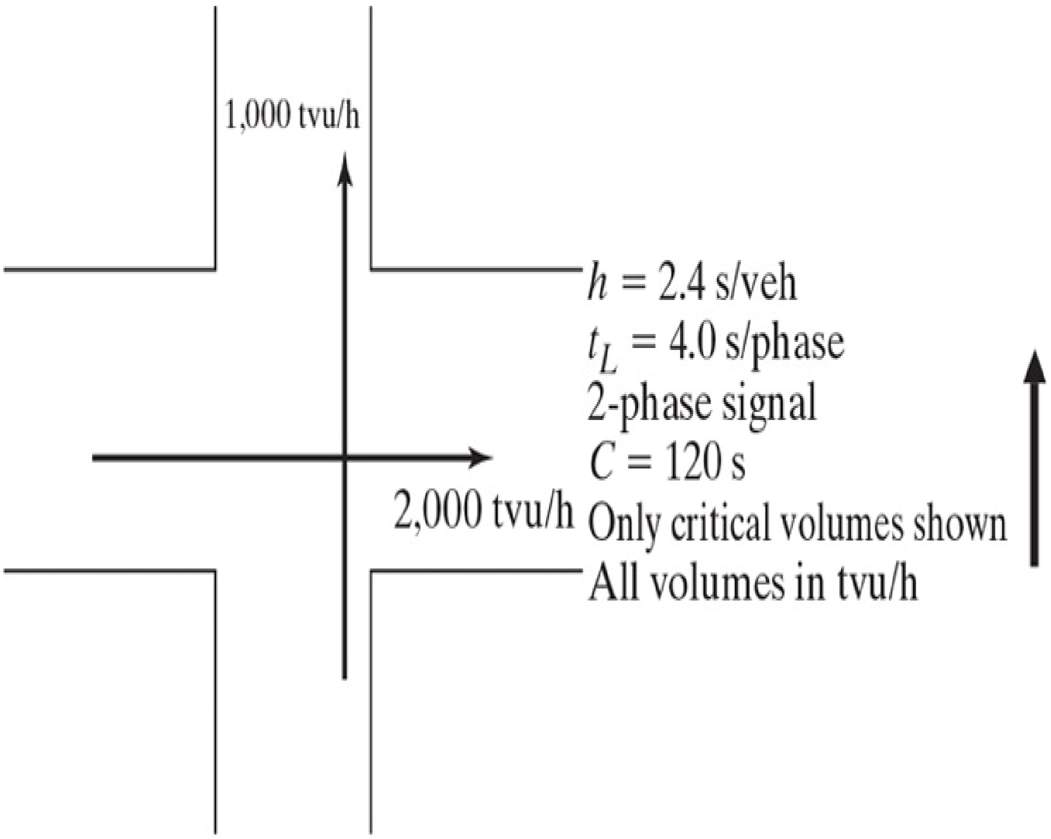
Full Alternative Text

2. 18-2. A signalized intersection approach has two lanes with no exclusive left- or right-turning lanes. The approach has a 50-second green out of a 90-second cycle. The yellow plus all-red intervals for the phase total 4.0 s. If the start-up lost time is 2.0 s/phase, the clearance lost time is 1.5 s/phase, and the saturation headway is 2.25 s/veh under prevailing conditions, what is the capacity of the intersection approach?

3. 18-3. What is the maximum sum of critical-lane volumes that may be served by an intersection having three phases, a cycle length of 90 s, a saturation headway of 2.2 s/veh, and a total lost time per phase of 4.0 s?

4. 18-4. For the intersection illustrated below, find the appropriate number of lanes for each lane group needed. Assume that all volumes shown have been converted to compatible "through-car equivalent" values for the conditions shown. Assume that critical volumes reverse in the other daily peak hour.

1. # Intersection for Problem 18.4



Full Alternative Text

2. 18-5. For the intersection of Problem 18.4, consider a case in which the E–W arterial has three lanes in each direction and the N–S arterial also has three lanes in each direction. For this case:

    1. What is the absolute minimum cycle length that could be used?

    2. What cycle length would be required to provide for a v/c ratio of 0.92 during the worst 15-minutes of the hour if the PHF is 0.98?

3. 18-6. At a signalized intersection, one lane is observed to discharge

40 through vehicles in the same time as the left lane discharges 10 through vehicles and 20 left-turning vehicles. For this case:

1. What is the through-vehicle equivalent, E_{LT}, for left-turning vehicles?
2. What is the left-turn adjustment factor, f_{LT}, for the case described?

4. 18-7. An intersection approach volume is 800 veh/h and includes 20% left turns with a through-vehicle equivalent of 2.5 tvu/left turn. What is the total equivalent through volume on the approach? Assume that all other vehicles on the approach have an equivalent of 1.00.

5. 18-8. An equation has been calibrated for the amount of time required to clear *N* vehicles through a given signal phase:

   $T=2.04+2.35\,N$

   1. What start-up lost time does this equation suggest exists?
   2. What saturation headway and saturation flow rate is implied by the equation?

6. 18-9. An intersection approach volume is 1,350 veh/h and includes 8% left turns with a through-vehicle equivalent of 2.7 tvu/left turn. What is the total equivalent through volume on the approach?

7. 18-10. An intersection approach has a demand volume of 500 veh/h, a saturation flow rate of 1,450 veh/hg, a cycle length of 80 s, and 50 s of effective green time. What average delay per vehicle is expected under these conditions?

8. 18-11. A signalized intersection approach operates at an effective ratio of 1.05 for a peak 30-minute period each evening. If the approach has a *g/C* ratio of 0.60 and the cycle length is 75 s:

    1. What is the average control delay for the entire 30-minute period?

    2. What is the average control delay during the first five minutes of the peak period?

9. 18-12. A signalized intersection approach experiences chronic oversaturation for a one-hour period each day. During this time, vehicles arrive at a rate of 2,000 veh/h. The saturation flow rate for the approach is 3,250 veh/hg, with a 100-second cycle length, and 55 s of effective green.

    1. What is the average control delay per vehicle for the full hour?

    2. What is the average control delay per vehicle for the first 15 minutes of the peak period?

    3. What is the average control delay per vehicle for the last 15 minutes of the peak hour?

Chapter 19 Fundamentals of Signal Timing and Design: Pre-timed Signals

A *pre-timed signal* is one in which all interval timings remain constant during each signal cycle. Thus, phase sequence, all green times, and the cycle length are fixed for the period of time over which the timing plan is applied. Yellow and all-red intervals are fixed as well, but they are also fixed in actuated controllers.

Modern signal controllers are fundamentally the same for both pre-timed and actuated control. Pre-timed control is achieved through controller settings that eliminate variable interval lengths. In previous times, older electro-mechanical signal controllers, often referred to as *three-dial controllers*, limited the number of different pre-timed signal plans that could be implemented to three. This allowed for the provision of an AM Peak timing, PM Peak timing, and an off-peak timing plan. Modern controllers allow as many as 20 different pre-timed patterns to be implemented. These plans are then tied to day and time settings that control when each of the patterns is in effect.

19.1 Introduction

Signal timing and design involve several important components, including the physical design and layout of the intersection itself. Physical design is treated in some detail in Chapter 17. This chapter focuses on the design and timing of traffic control signals.

These are the key steps involved in signal design and timing:

1. Development of a safe and effective phase plan and sequence
2. Determination of vehicular signal needs:
    1. Timing of “yellow” (change, yi) and “all-red” (clearance, ari) intervals for each signal phase
    2. Determination of the sum of critical-lane volumes (Vc)
    3. Determination of lost times per phase (tLi) and per cycle (*L*)
    4. Determination of an appropriate cycle length (*C*)
    5. Allocation of effective green time to the various phases defined in the phase plan—often referred to as “splitting” the green
3. Determination of pedestrian signal needs:
    1. Determine minimum pedestrian “green” times
    2. Check to see if vehicular greens meet minimum pedestrian needs
    3. If pedestrian needs are unmet by the vehicular signal timing, adjust timing and/or add pedestrian actuators to ensure pedestrian safety

Although most signal timings are developed for vehicles and checked for pedestrian needs, it is critical that signal-timing designs provide safety and relative efficiency for both. Approaches vary with relative vehicular and pedestrian flows, but every signal timing must consider and provide for the requirements of both groups.

Many aspects of signal timing are tied to the principles discussed in Chapter 18 and elsewhere in this text. The process, however, is not exact, nor is there often a single “right” design and timing for a traffic control signal. Thus, signal timing does involve judgmental elements and represents true engineering design in a most fundamental way.

All of the key elements of signal timing are discussed in some detail in this chapter, and various illustrations are offered. Note, however, that it is virtually impossible to develop a complete and final signal timing that will not be subject to subsequent fine tuning when the proposed design is analyzed using the HCM 2016 analysis model or some other analysis model or simulation. This is because no straightforward signal design and timing process can hope to include and fully address all of the potential complexities that may exist in any given situation. Thus, initial design and timing is often a starting point for analysis using a more complex model.

Chapters 22 and 23 of this text discuss two analysis models: one based on critical lane analysis, which is essentially signal timing in reverse (Chapter 23), and the second the HCM analysis model (Chapter 22). Chapter 23 is based on the planning-level model in the HCM 2016.

19.2 Development of a Signal Phase Plan

The most critical aspect of signal design and timing is the development of an appropriate phase plan. Once this is done, many other aspects of the signal timing can be analytically treated in a deterministic fashion. The phase plan and sequence involves the application of engineering judgment while applying a number of commonly used guidelines. In any given situation, there may be a number of feasible approaches that will work effectively.

19.2.1 Provisions for Left Turns: A Determining Factor

The single most important feature that drives the development of a phase plan is the treatment of left turns. As discussed in Chapter 18, left turns may be handled as permitted movements (with an opposing through flow), as protected movements (with the opposing vehicular through movement stopped), or as a combination of the two (compound phasing). The simplest signal phase plan has two phases, one for each of the crossing streets. In this plan, all left turns are permitted. Additional phases may be added to provide protection for some or all left turns, but additional phases add lost time to the cycle. Thus, the consideration of protection for left turns must weigh the inefficiency of adding phases and lost time to the cycle against the improved efficiency in operation of left-turning and other vehicles gained from that protection.

There are many guidelines that have been developed and used over the years that assist in making an initial determination of whether a particular left-turn movement needs to be fully or partially protected. Various guidelines have considered issues of left-turn volumes, the conflict between a left-turn volume and its opposing through movement, approach speeds of the opposing through movement, the number of left turn lanes, the number of opposing through lanes, left-turn visibility distances, left-turn accident occurrence, and other conditions. The guidelines presented here are examples. There are variations of these and others that are used by various traffic agencies across the United States.

The most frequently used reasons for implementing left-turn phasing involves two considerations of volume: the total left-turn volume, and the conflicting volume. If either of the two criteria below are exceeded, it is common practice to implement some form of left-turn protection:

$v_{LT} \geq 200$ veh/h [19-1]
$x_{prod} = v_{LT} \times (v_o N_o) \geq 50{,}000$ [19-2]

where:

v_{LT}=left-
turn flow rate, veh/h,v_o=opposing through movement flow rate,veh/h, andN

Equation 19-2 is often referred to as the "cross-product" rule. Various agencies may use different forms of this particular guideline.

The *Traffic Signal Timing Manual* [2] suggests a number of other conditions that should lead to some form of left-turn protection. These are summarized in Table 19.1.

Table 19.1: Additional Criteria for Potential Left-Turn Protection

Value	Protect If Value Is:
Number of LT lanes on the subject approach	≥ 2 lanes
Number of opposing through lanes	≥ 4 lanes
85th percentile speed of opposing traffic	>45 mi/h
Critical LT-related crash count	See Table 19.2
LT driver sight distance to oncoming traffic	See Table 19.3

Table 19.1: Full Alternative Text

Like the left-turn flow rate and cross-product criteria, these are not absolute guidelines. They provide assistance in deciding whether to include left-turn protection (either partial or full) for a particular left-turn movement in an initial signal phase plan. It is always possible that as the design and subsequent analysis takes place, some changes will be introduced.

Where the sight-distance criteria is the only reason for protecting a left-turn, consideration should first be given to improving the sight distance.

There are other considerations. Left-turn protection, for example, is rarely provided when left-turn flows are less than two vehicles per cycle. It is generally assumed that in the worst case, where opposing flows are so high that *no* left turns may filter through it, an average of two vehicles each cycle will wait in the intersection until the opposing flow is stopped and then complete their turns. Such vehicles are usually referred to as "sneakers."

Where a protected phase is needed for one left-turning movement, it is often convenient to provide one for the opposing left turn, even if it does not meet any of the normal guidelines. Sometimes, a left turn that does not meet any of the guidelines will present a particular problem that is revealed during the signal timing or in later analysis, and protection will be added.

In a typical four-leg intersection, there are four opposed left-turn movements. A decision on LT protection is made independently for each, based on the criteria discussed, and/or other local guidelines in effect. The overall pattern of protection is then considered as the phase plan is established.

Except for the LT-related crash criteria of Table 19.2, none of the guidelines help determine whether LT protection should be full or partial. The use of *compound phasing* (either protected + permitted or permitted + protected) is quite complex, and local policies and guidelines tend to govern these decisions. In general, full protection provides for maximum safety and clarity for motorists, as all potential LT conflicts with opposing through vehicles are eliminated. However, full protection generally adds to average delays and lengthens the cycle lengths needed for implementation.

Compound phasing usually produces less average delay than full protection, and may also increase the LT capacity for the controlled movement. It is, however, more complex for drivers, and does not eliminate all LT conflicts with opposing vehicles.

Table 19.2: Critical LT-Related Crash Count Criteria for LT Protection

Number of LT Movements on Subject Road	Period during Which Crashes Are Considered (years)	Critical LT-Related Crash Count	
		When Considering Full Protection	When Considering Partial Protection
1	1	≥ 6	≥ 4
1	2	≥ 11	≥ 6
1	3	≥ 14	≥ 7
2	1	≥ 11	≥ 6
2	2	≥ 18	≥ 9
2	3	≥ 26	≥ 13

(Source :Extracted from *Traffic Signal Timing Manual*, Federal Highway Administration, Washington, D.C., Figure 4-11, pg 4-13.)

Table 19.2: Full Alternative Text

Table 19.3: Minimum Sight Distance Criteria for LT Protection

Oncoming Traffic Speed Limit (mi/h)	Protect If Sight Distance to Oncoming Opposing Vehicles Is:
25	< 200 ft
30	< 240 ft
35	< 280 ft
40	< 320 ft
45	< 360 ft
50	< 400 ft
55	< 440 ft
60	< 480 ft

(Source: Extracted from *Traffic Signal Timing Manual*, Federal Highway Administration, Washington, D.C., Figure 4-11, pg 4-13.)

Table 19.3: Full Alternative Text

Some jurisdictions use compound phasing whenever possible because of the delay and capacity benefits; others avoid it almost completely to provide for maximum safety and clarity.

The *Traffic Engineering Handbook* [3] does offer some guidance on conditions that would normally dictate the use of full LT protection. Full protection is recommended when any *two* of the following criteria are met:

1. Left-turn flow rate is greater than 320 veh/h.
2. Opposing flow rate is greater than 1,100 veh/h.
3. Opposing speed limit is greater than or equal to 45 mi/h.
4. There are two or more left-turn lanes.

Additional guidelines are offered for critical combinations of conditions. Fully protected phasing is also recommended when any *one* of the following combinations exists:

1. There are three opposing traffic lanes, and the opposing speed is 45

mi/h or greater.

2. Left-turn flow rate is greater than 320 veh/h, and the percentage of heavy vehicles (in the LT flow) exceeds 2.5%.
3. The opposing flow rate exceeds 1,100 veh/h, and the percentage of left turns (in the subject movement) exceeds 2.5%.
4. Seven or more left-turn accidents have occurred within three years with compound phasing in operation.
5. The average stopped delay to left-turning traffic is acceptable for fully protected phasing, and the engineer judges that additional left-turn accidents would occur under the compound phasing option.

Item 5 obviously provides the traffic engineer with a great deal of latitude, and calls for the exercise of engineering judgment.

The guidelines discussed essentially indicate when compound phasing *should not* be implemented. They imply that compound phasing may be considered when left-turn protection is needed but none of these criteria are met.

In extreme cases, it may be necessary to ban left turns entirely. This must be done, however, with the utmost care. It is essential that alternative routes for vehicles wishing to turn left are available and that they do not unduly inconvenience the affected motorists. Further, the additional demands on alternative routes should not cause worse problems at nearby intersections. Special design treatments for left turns are also discussed in Chapters 25 and 26.

19.2.2 General Considerations in Signal Phasing

Several important considerations should be kept in mind when establishing a phase plan:

1. Phasing can be used to minimize accident risks by separating competing movements. A traffic signal always eliminates the basic

crossing conflicts present at intersections. Addition of left-turn protection can also eliminate some or all of the conflicts between left-turning movements and their opposing through movements. Additional phases generally lead to additional delay, which must be weighed against the safety and improved efficiency of protected left turns.

2. Although increasing the number of phases increases the total lost time in the cycle, the offsetting benefit is an increase in affected left-turn saturation flow rates.

3. All phase plans must be implemented in accordance with the standards and criteria of the MUTCD [4], and they must be accompanied by the necessary signs, markings, and signal hardware needed to identify appropriate lane usage.

4. The phase plan must be consistent with the intersection geometry, lane-use assignments, volumes and speeds, and pedestrian crossing requirements.

For example, it is not practical to provide a fully protected left-turn phase where there is no exclusive left-turn lane. If such phasing were implemented with a shared lane, the first vehicle in queue may be a through vehicle. When the protected left-turn green is initiated, the through vehicle blocks all left-turning vehicles from using the phase. Thus, protected left-turn phases *require* exclusive left-turn lanes for effective operation.

19.2.3 Phase and Ring Diagrams

A number of typical and a few not-so-typical phase plans are presented and discussed here. Signal phase plans are generally illustrated using *phase diagrams* and *ring diagrams*. In both cases, movements allowed during each phase are shown using arrows. Here, only those movements allowed in each phase are shown; in some of the literature, movements not allowed are also shown with a straight line at the head of the arrow, indicating that the movement is stopped during the subject phase. Figure 19.1 illustrates some of the basic conventions used in these diagrams.

Figure 19.1: Symbols Used in Phase and Ring Diagrams

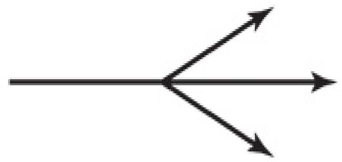
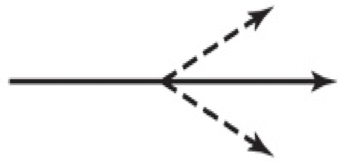
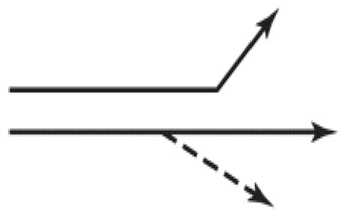
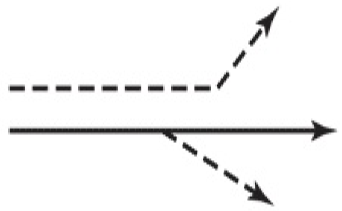
Through movement without turning movement.	
Through movement with protected right and left turns from shared lanes.	
Through movement with permitted right and left turns from shared lanes.	
Through movement with protected left turn from exclusive lane and permitted right turn from shared lane.	
Through movement with permitted left turn from exclusive lane and permitted right turn from shared lane.	

Figure 19.1: Full Alternative Text

A more complete definition and discussion of the use and interpretation of these symbols follows:

1. A solid arrow denotes a movement without opposition. All through movements are unopposed by definition. An unopposed left turn has no opposing through vehicular flow. An unopposed right turn has no opposing pedestrian movement in the crosswalk through which the right turn is made.

2. Opposed left- and or right-turn movements are shown as a dashed line.

3. Turning movements made from a shared lane(s) are shown as arrows connected to the through movement that shares the lane(s).

4. Turning movements from an exclusive lane(s) are shown as separate arrows, not connected to any through movement.

Although not shown in Figure 19.1, pedestrian paths may also be shown on phase or ring diagrams. They are generally shown as dotted lines with a double arrowhead, denoting movement in both directions in the crosswalk.

A *phase diagram* shows all movements being made in a given phase within a single block of the diagram. A *ring diagram* shows which movements are controlled by which "ring" on a signal controller. A "ring" of a controller generally controls one set of signal faces. The concept of rings originated with electro-mechanical controllers, which used timing cylinders to control the length of signal intervals. A "ring" was literally a portion of the cylinder that controlled traffic from one approach or lane group. Thus, although a phase involving two opposing through movements would be shown in one block of a phase diagram, each movement would be separately shown in a ring diagram. The ring diagram is more informative, particularly where overlapping phase sequences are involved. Chapter 16 describes signal hardware and the operation of signal controllers in more detail.

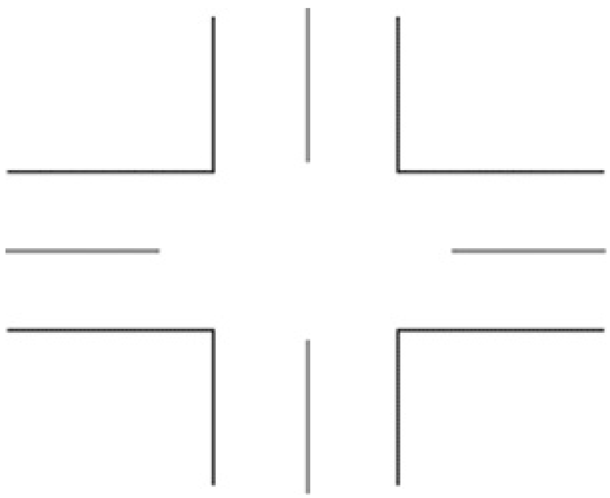
19.2.4 Common Phase Plans and Their Use

Simple two-phase signalization is the most common plan in use. If guidelines or professional judgments indicate the need to fully or partially protect one or more left-turn movements, a variety of options are available for doing so. The following sections illustrate and discuss the most common phase plans in general use.

Simple Two-Phase Signalization

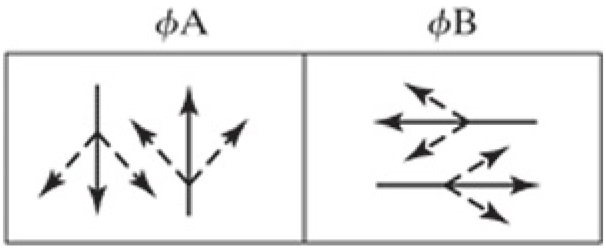
Figure 19.2 illustrates basic two-phase signalization. Each street receives one signal phase, and all left and right turns are made on a permitted basis. Exclusive lanes for left- and/or right-turning movements may be used but are not required for two-phase signalization. This form of signalization is appropriate where the mix of left turns and opposing through flows is such that no unreasonable delays or unsafe conditions are created by and/or for left-turners.

Figure 19.2: Illustration of a Two-Phase Signal



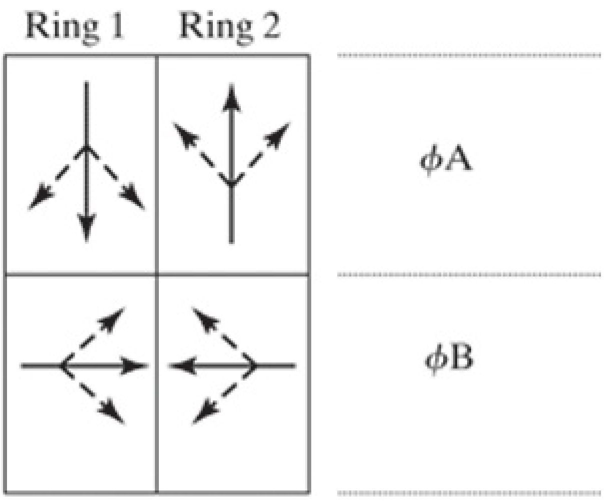
(a) Intersection Layout (exclusive LT/RT lanes optional)

19.2-4 Full Alternative Text



(b) Phase Diagram

19.2-4 Full Alternative Text



(c) Ring Diagram

19.2-4 Full Alternative Text

In this case, the phase diagram shows all N–S movements occurring in Phase A and all E–W movements occurring in Phase B. The ring diagram shows that in each phase, each set of directional movements is controlled by a separate ring of the signal controller. Because the basic signalization is relatively simple, both the phase and ring diagrams are quite similar, and both are relatively easy to interpret. This is not the case for more complex signal phase plans.

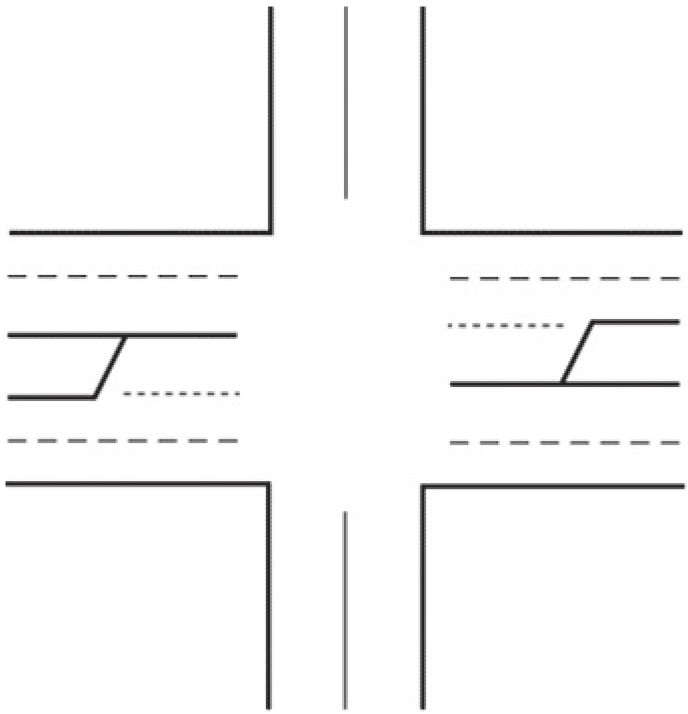
Note that all phase changes cut across both rings of the controllers, meaning that all transitions occur at the same times in both rings. These are referred to as “phase boundaries.” Also, it would make little difference which movements appear in which rings. The combination shown could be easily reversed without affecting the operation of the signal.

Exclusive Left-Turn Phasing

When a need for left-turn protection is indicated by guidelines or professional judgment, the simplest way to provide it is through the use of an exclusive left-turn phase(s). Two opposing left-turn movements are provided with a simultaneous and exclusive left-turn green, during which the two through movements on the subject street are stopped. An exclusive left-turn phase may be provided either before or after the through/right-turn phase for the subject street, although the most common practice is to provide it *before* the through phase. Because this is the most often-used sequence, drivers have become more comfortable with left-turn phases placed before the corresponding through phase.

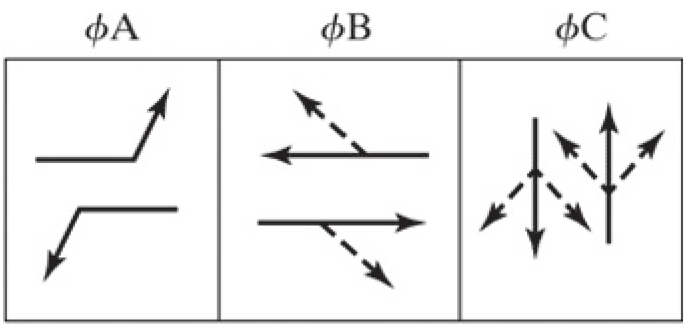
When an exclusive left-turn phase is used, an exclusive left-turn lane of sufficient length to accommodate expected queues must be provided. If an exclusive left-turn phase is implemented on one street and not the other, a three-phase signal plan emerges. Where an exclusive left-turn phase is implemented on both intersecting streets, a four-phase signal plan is formed. Figure 19.3 illustrates the use of an exclusive left-turn phase on the E-W street but not on the N–S street, where left turns are made on a permitted basis.

Figure 19.3: Exclusive Left-Turn Phase Illustrated



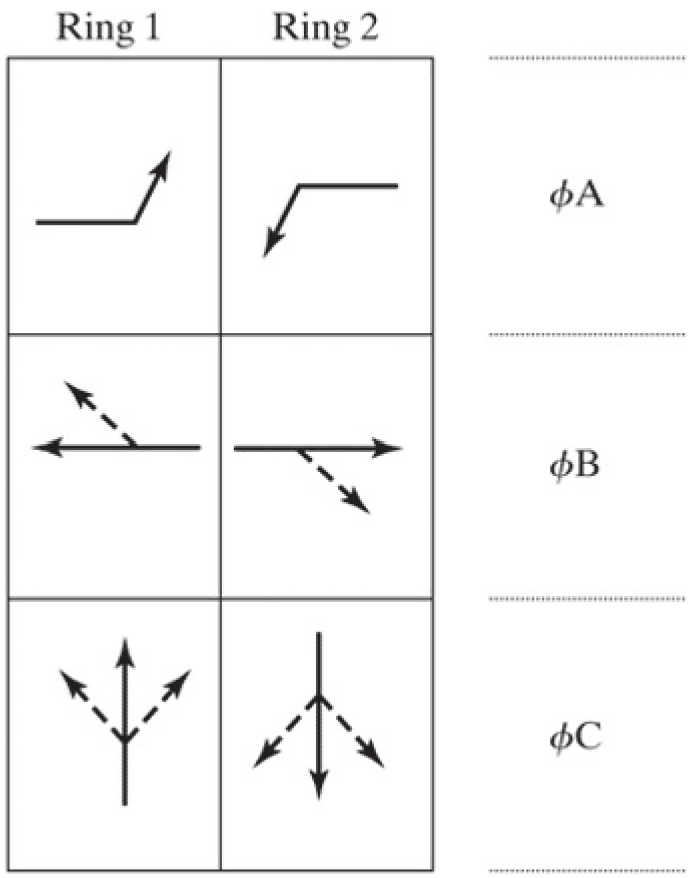
(a) Intersection Layout

19.2-4 Full Alternative Text



(b) Phase Diagram

19.2-4 Full Alternative Text



(c) Ring Diagram

19.2-4 Full Alternative Text

The phase plan of Figure 19.3 can be modified to provide for protected plus permitted left turns on the E-W street. This is done by adding a permitted left-turn movement to Phase B. In general, such compound phasing is used where the combination of left turns and opposing flows is so heavy that provision of fully protected phasing leads to undesirably long or unfeasible cycle lengths. Compound phasing is more difficult for drivers to comprehend and is more difficult to display.

Exclusive or protected left-turn movements are indicated by use of a green arrow. The arrow indication may be used only when there is no opposing through movement. In the case of a protected-plus-permitted compound phase, the green arrow is followed by a yellow arrow; the yellow arrow is then followed by a green ball indication during the permitted portion of the phase.

Leading and Lagging Green

Phases

When exclusive left-turn phases are used, a potential inefficiency exists. If the two left-turning movements have very different demand flow rates (on a per-lane basis), then providing them with protected left-turn phases of equal length assures that the smaller of the two left-turn movements will have excess green time that cannot be used.

Where this inefficiency leads to excessive or unfeasible cycle lengths and/or excessive delays, a phase plan in which opposing protected left-turn phases are separated should be considered. If a NB protected left-turn phase is separated from the SB protected left-turn phase, the two can be assigned different green times in accordance with their individual demand flow rates.

One historic approach to accomplishing this is referred to as “leading and lagging” green phases. A leading and lagging green sequence for a given street has three components:

- The leading green. Vehicles in one direction get the green while vehicles in the opposing direction are stopped. Thus, the left-turning movement in the direction of the “green” is protected.

- The overlapping through green. Left-turning vehicles in the initial green direction are stopped while through (and right-turning) vehicles in both directions are released. As an option, left turns may be allowed on a permitted basis in both directions during this portion of the phase, creating a compound phase plan.

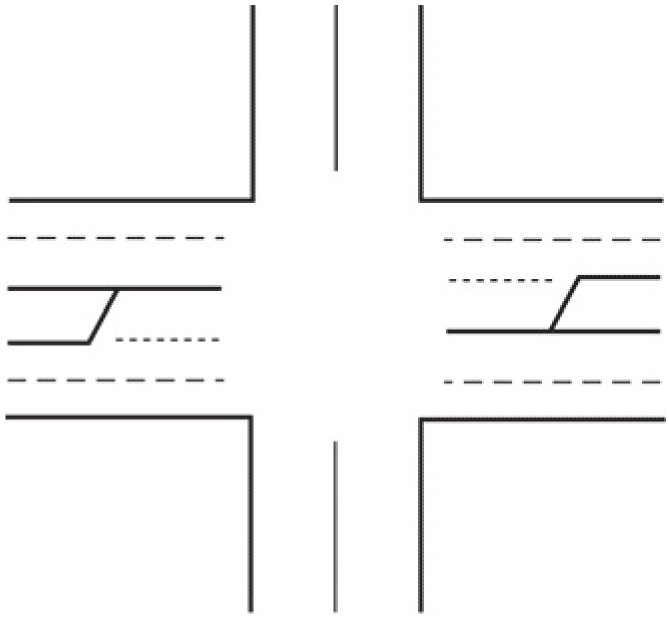
- The lagging green. Vehicles in the initial direction (all movements) are stopped while vehicles in the opposing direction continue to have the green. Because the opposing flow is stopped, left turns made during this part of the phase are protected.

The leading and lagging green sequence is no longer a standard phasing supported by the National Electronics Manufacturing Association (NEMA), which creates standards for signal controllers and other electronic devices. Such controllers, however, are still available, and this sequence is still used in some jurisdictions. The main issue with this type of phasing is the *lagging* green. Left-turners in the lagging movement get the green arrow *after* the opposing through movement is completed; modern drivers generally expect protected movements to occur *before* through movements.

At intersections where a one-way street crosses a two-way street, or when a T-intersection exists, there is only one opposed left-turn to be considered. Should it require protection, a *leading* green would be provided, but there would be no *lagging* green.

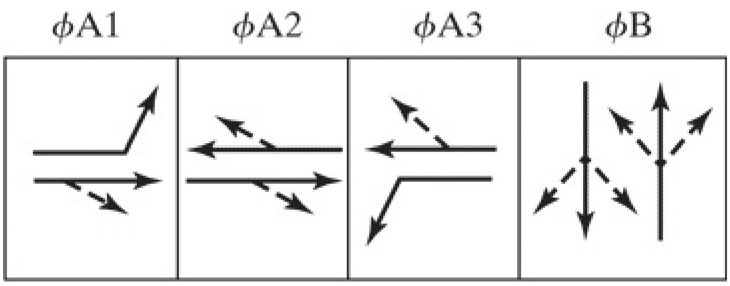
Figure 19.4 illustrates a leading and lagging green sequence in the E-W direction. A similar sequence can be used in the N–S direction as well. Again, an exclusive left-turn lane must be provided when a leading and lagging green is implemented.

Figure 19.4: Leading and Lagging Green Illustrated



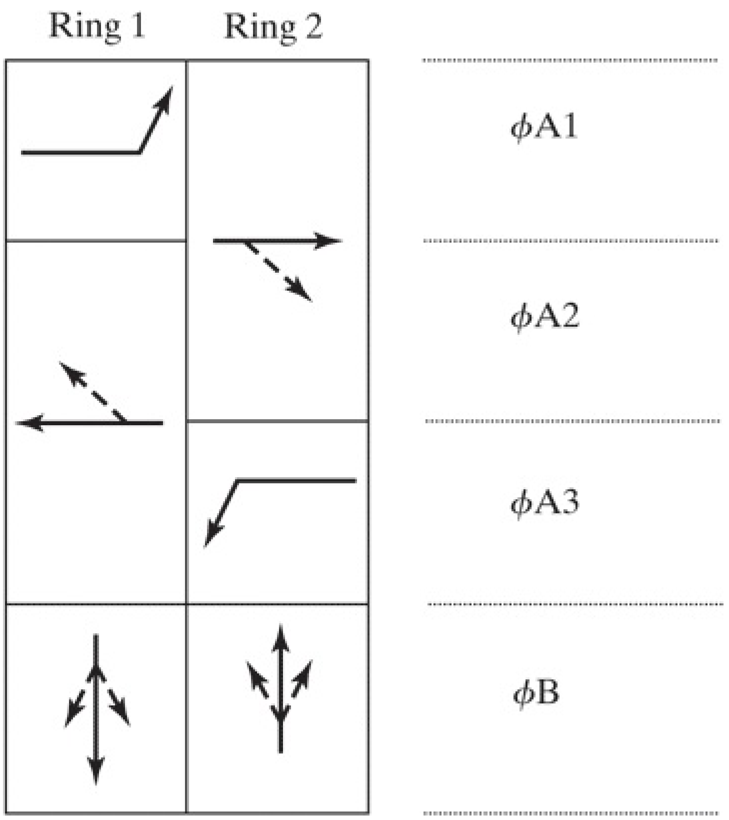
(a) Intersection Layout

19.2-4 Full Alternative Text



(b) Phase Diagram

19.2-4 Full Alternative Text



(c) Ring Diagram

19.2-4 Full Alternative Text

The leading and lagging green phase plan involves “overlapping” phases. The EB through is moving in Phases A1 and A2 while the WB through is moving in Phases A2 and A3. One critical question arises in this case: How many phases are there in this plan? It might be argued that there are *four* distinct phases: A1, A2, A3, and B. It might also be argued that Phases A1, A2, and A3 form a single overlapping phase and that the plan therefore involves only *two* phases. In fact, both answers are incorrect.

The ring diagram is critical in the analysis of overlapping phase plans. At the end of Phase A1, only Ring 1 goes through a transition, transferring the green from the EB left turn to the WB through and right-turn movements. At the end of Phase A2, only Ring 2 goes through a transition, transferring the green from the EB through and right-turn movements to the WB left turn. Each ring, therefore, goes through *three* transitions in a cycle. In effect, this is a *three*-phase signal plan. The ring diagram makes the difference between partial and full phase intervals clear, whereas the phase diagram can easily mask this important feature.

This distinction is critical to subsequent signal-timing computations. For each phase transition, a set of lost times (start-up plus clearance) is experienced. If the sum of the lost times per phase (tL) were 4.0 seconds per phase, then a two-phase signal would have 8.0 seconds of lost time per cycle (*L*), a three-phase signal would have 12.0 seconds of lost time per cycle, and a four-phase signal would have 16.0 seconds of lost time per cycle. The lost time per cycle has a dramatic effect on the required cycle length.

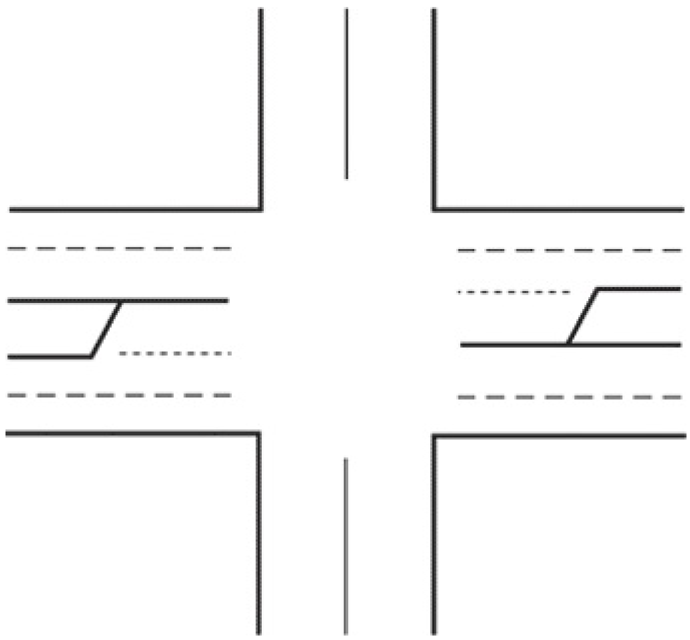
The ring diagram also makes the interval designations understandable. The E–W phases are labeled A1, A2, and A3, *not* A, B, and C. That is because there is only *one* phase boundary for the E–W movements—at the *end* of A3. Only when a full phase boundary exists, does the *letter* designation of the interval change.

Some form of left-turn protection could be added to the N–S street if needed. It would *not* have to be a leading and lagging green, however.

Exclusive Left-Turn Phase with Leading Green

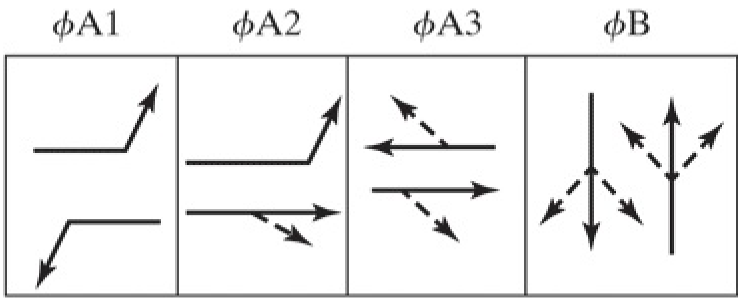
It was previously noted that NEMA does not have a set of controller specifications to implement a leading and lagging green phase plan. The NEMA standard phase sequence for providing unequal protected left-turn phases employs an exclusive left-turn phase followed by a leading green phase in the direction of the heaviest left-turn demand flow. In effect, this sequence provides the same benefits as the leading and lagging green, but it allows all protected left-turn movements to be made before the opposing through movements are released. Figure 19.5 illustrates such a phase plan for the E–W street.

Figure 19.5: Exclusive Left-Turn Phase Plus Leading Green Phase Illustrated



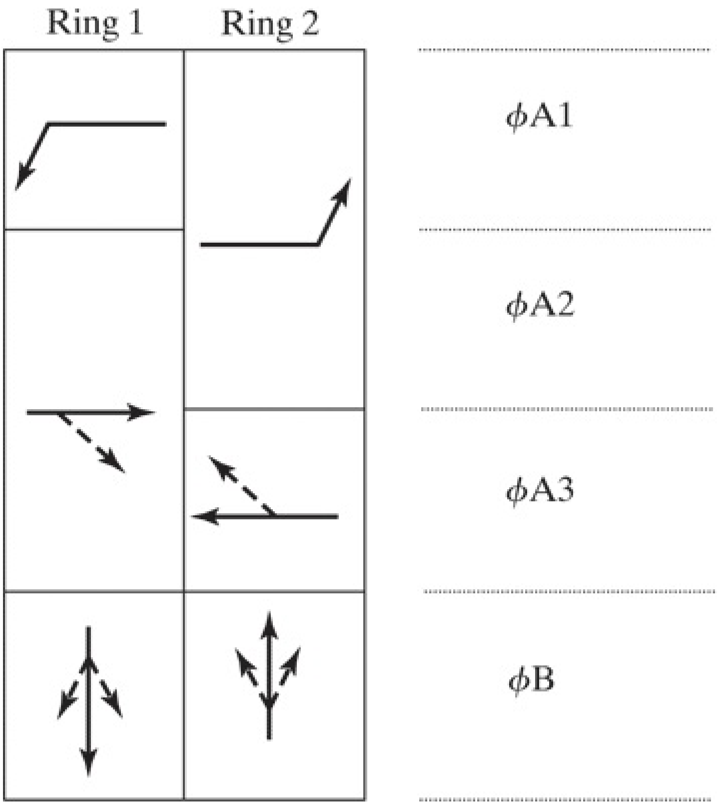
(a) Intersection Layout

19.2-4 Full Alternative Text



(b) Phase Diagram

19.2-4 Full Alternative Text



(c) Ring Diagram

19.2-4 Full Alternative Text

In the case illustrated, the EB left-turn movement receives the leading green as the heavier of the two left-turn demand flows. If the WB left turn required the leading green, this is easily accomplished by reversing the positions of the partial boundaries between Phases A1 and A2 and between A2 and A3.

Note there is a similarity between the leading and lagging green phase plan and the exclusive left-turn plus leading green phase plan. In both cases, the partial transition in each ring is between a protected left turn and the opposing through and right-turn movements, or vice versa. Virtually all overlapping phase sequences involve such transfers.

A compound phase can be implemented by allowing EB and WB permitted left-turn movements in Phase A3. In both cases, this creates a protected-plus-permitted phase sequence. However, it is critical to note that *one* of the left turns has an interruption between the protected and permitted portions of the phase. For this movement (generally the lower left turn flow), an additional lost time is introduced, as it literally starts and stops twice during the cycle. For this reason, NEMA phase plans often avoid compound phasing.

Similar phasing can also be implemented for the N–S street if needed, as long as an exclusive left-turn lane is provided.

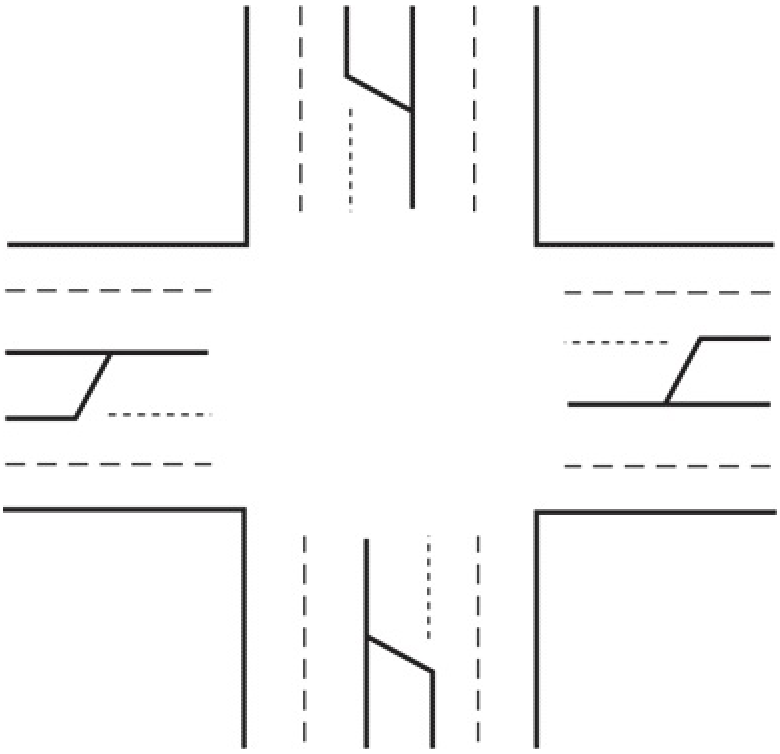
The issue of number of phases is also critical in this phase plan. The phase plan of Figure 19.5 involves *three* discrete phases and *three* phase transitions on each ring.

Eight-Phase Actuated Control

Any of the previous phase plans may be implemented in the pre-timed or actuated mode (with detectors present). However, actuated control offers the additional flexibility of skipping phases when no demand is detected. This is most often done for left-turn movements.

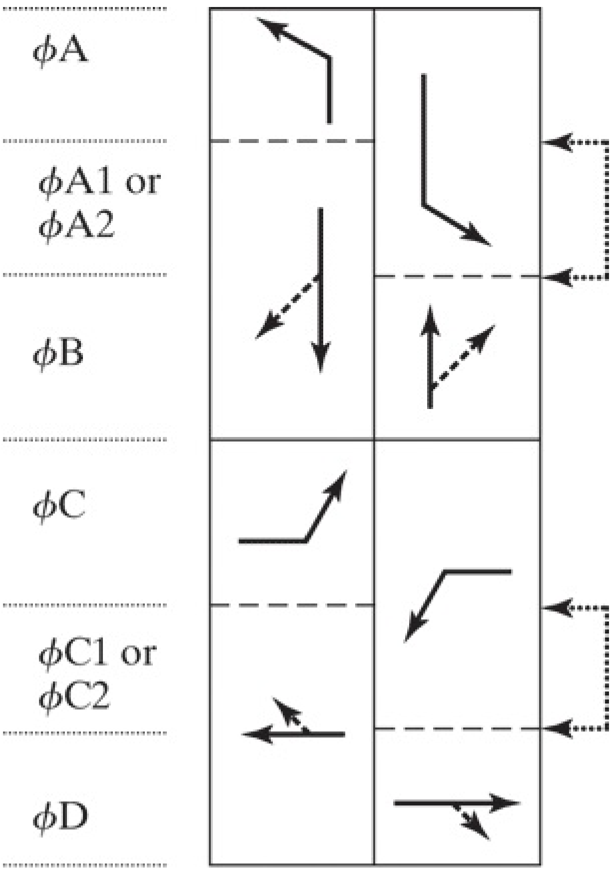
Protected left-turn phases may be skipped in any cycle where detectors indicate no left-turn demand. The most flexible controller follows the phase sequence of an exclusive left-turn phase plus a leading green. Figure 19.6 shows the actuated phase plan for such a controller.

Figure 19.6: "Quad-8" Actuated Phase Plan



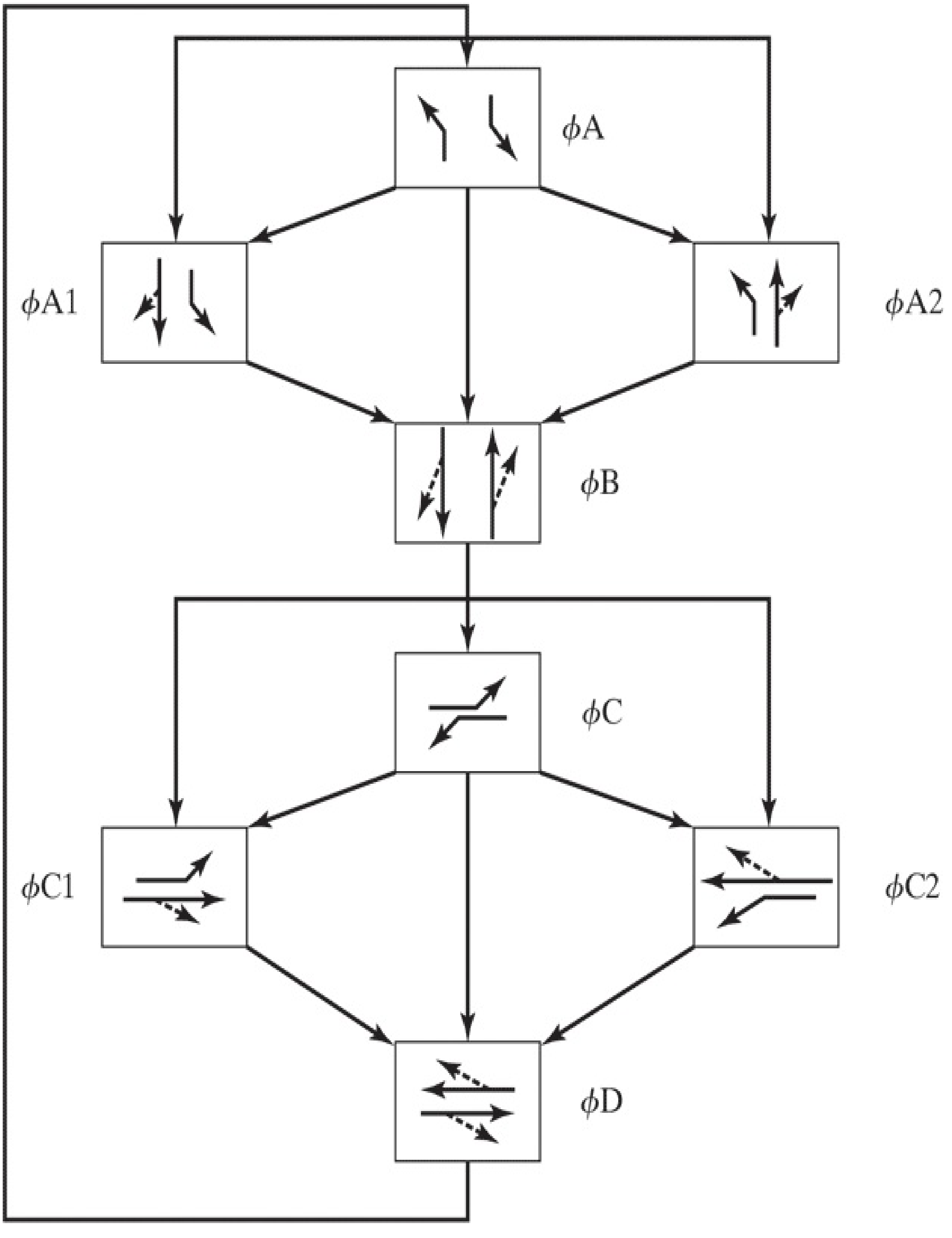
(a) Intersection Layout

19.2-4 Full Alternative Text



(b) Ring Diagram

19.2-4 Full Alternative Text



(c) Actuated Phase Diagram

19.2-4 Full Alternative Text

In this case, exclusive left-turn phases and leading greens are provided for both streets, and both streets have exclusive left-turn lanes as shown.

This type of actuated signalization provides for complete flexibility in both the phase sequence and in the timing of each phase. Each street may start its green phases in one of three ways, depending on demand:

- An exclusive left-turn phase in both directions if left-turn demand is present in both directions.
- A leading green phase (in the appropriate direction) if only one left-turn demand is present.
- A combined through and right-turn phase in both directions if no left-turn demand is present in either direction.

If the first option is selected, the next phase may be a leading green if one direction still has left-turn demand when the other has none, or a combined through and right-turn phase if both left-turn demands are simultaneously satisfied during the exclusive left-turn phase.

The ring diagram assumes that a full sequence requiring both the exclusive left-turn phase and the leading green phase (for one direction) are needed. The partial phase boundaries are shown as dashed lines because the relative position of these may switch from cycle to cycle depending on which left-turn demand flow is greater. If the entire sequence is needed, there are *four* phase transitions in either ring, making this (as a maximum) a *four-phase* signal plan. Thus, even though the controller defines eight potential phases, during any given cycle, a maximum of four phases may be activated.

Actuated control is generally used where signalized intersections are relatively isolated, or in modern signal systems where the cost of coordinating actuated signals is considered worthwhile. The type of flexibility provided by eight-phase actuated control is most effective where left-turn demands vary significantly over the course of the day.

19.2.5 Special Cases and Phase Plans

While the majority of intersections can be effectively signalized using the approaches presented in the previous section, there are many special cases that may arise that require more innovative phasing approaches. Some of these are discussed here, but it should be noted that other situations do exist that are not specifically addressed here.

The Exclusive Pedestrian Phase

Pedestrians are a critical factor to be addressed in any signalization. In some cases, large or dominant pedestrian flows require special attention.

Originally developed by New York City traffic engineer Henry Barnes in the 1960s for Manhattan, the exclusive pedestrian phase was implemented as a new approach to this situation. This type of phasing is often referred to as the "Barnes Dance."

Figure 19.7 illustrates this phasing. During the exclusive pedestrian phase, pedestrians are permitted to cross the intersection in any direction, including diagonally. All vehicular movements are stopped during the exclusive pedestrian phase. The exclusive pedestrian phase is virtually never used where more than two vehicular phases are needed.

Figure 19.7: An Exclusive Pedestrian Phase

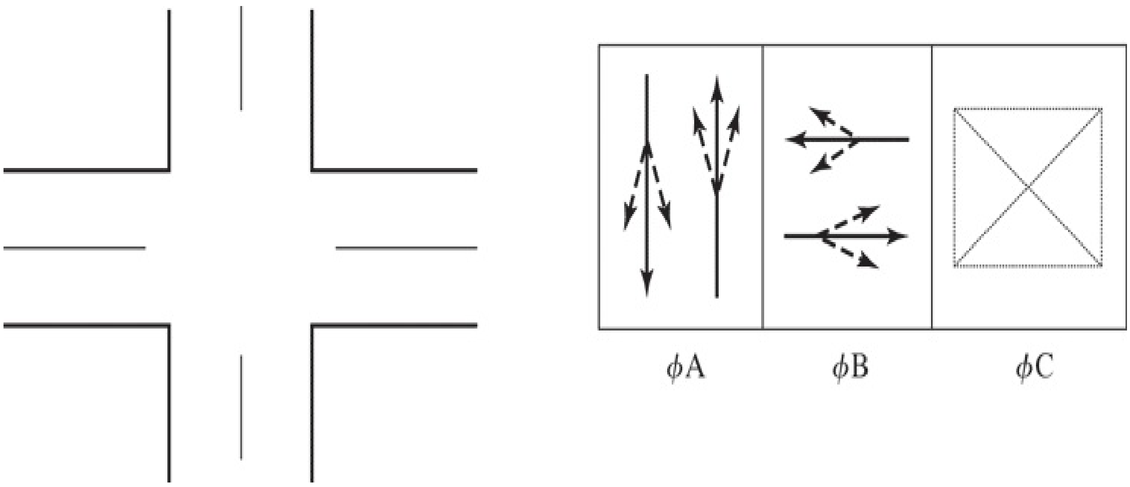


Figure 19.7: Full Alternative Text

The exclusive pedestrian phase has two principal benefits: pedestrian movements face no vehicular conflicts, and right-turn movements during vehicular phases have no pedestrian interference.

The exclusive pedestrian phase has several drawbacks. The primary problem is that the entire pedestrian phase must be treated as lost time in terms of the vehicular signalization. Delays to vehicles are substantially increased because of this, and vehicular capacity is significantly reduced.

The exclusive pedestrian phase never worked well in the city of its birth. Where extremely heavy pedestrian flows exist, such as in Manhattan, the issue of clearing them out of the intersection at the close of the pedestrian phase is a major enforcement problem. In New York, pedestrians occupied intersections for far longer periods than intended, and the negative impacts on vehicular movement were intolerable.

The exclusive pedestrian phase works best in small rural or suburban centers, where vehicular flows are not extremely high and where the volume of pedestrians is not likely to present a clearance problem at the end of the pedestrian phase. In such cases, it can provide additional safety for pedestrians in environments where drivers are not used to negotiating conflicts between vehicles and pedestrians.

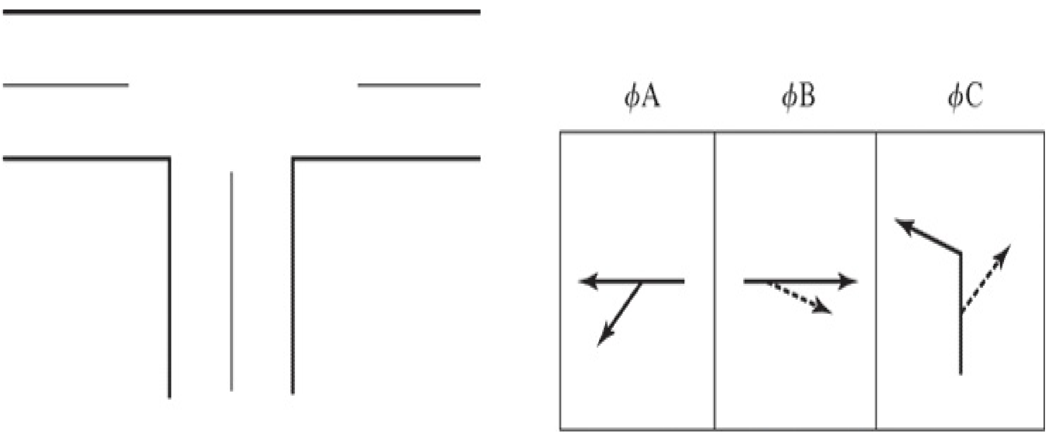
Unique Geometries and Signal Phasing

The typical intersection involves two streets intersecting at close to a 90° angle. Some intersections, however, involve atypical geometries that pose special problems for signal phase design. In some cases, geometry can actually help provide a more efficient signal phase plan. In all cases of atypical geometry, however, the signal phase plan must take into account the characteristics of the geometry.

T-intersections simplify signalization in that one set of vehicular movements is eliminated. There is generally only one opposed left turn at such intersections. Where that turn can be accommodated safely as a permitted left-turn, a simple two-phase signal results. Where the opposed left turn requires protection, the geometry can be used to develop some innovative and efficient phase plans.

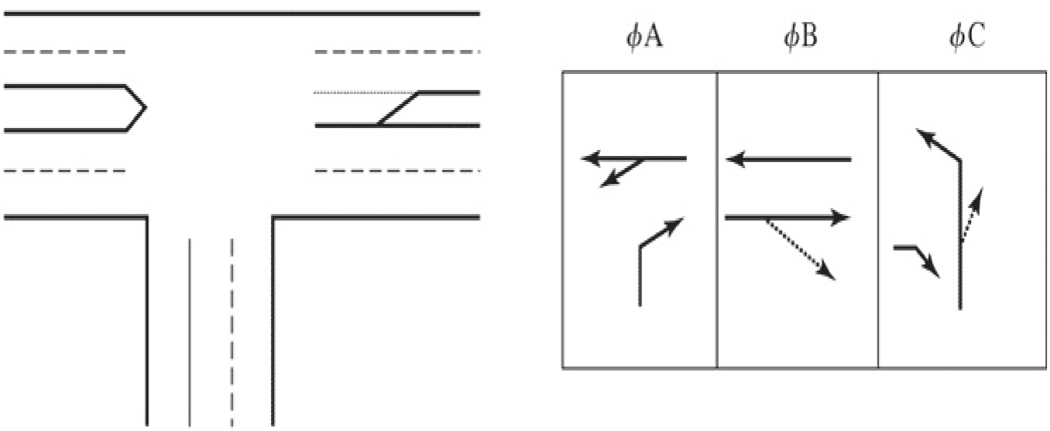
Figure 19.8 illustrates such a situation along with several candidate solutions. In Figure 19.8(a), there are no turning lanes provided. In such a case, providing the WB (opposed) left turn with a protected phase requires that each of the three approach legs have its own exclusive signal phase. Although achieving the required protected phasing for the opposed left turn, such phasing is not very efficient in that each movement uses only one of three phases. Delays to all vehicles tend to be longer than they would be if more efficient phasing could be implemented.

Figure 19.8: Signalization Options at a T-intersections



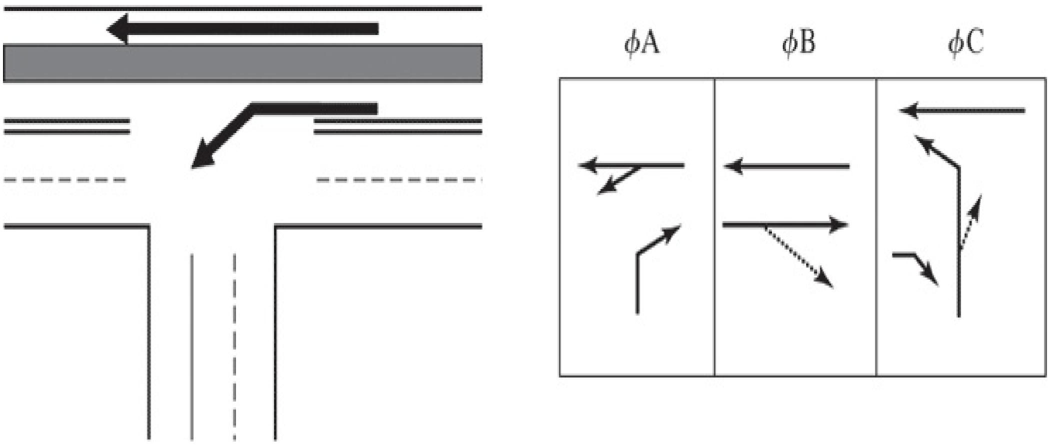
(a) T-Intersection, No LT Lane, Protected Phasing

19.2-4 Full Alternative Text



(b) T-Intersection, LT Lane, Protected Phasing

19.2-4 Full Alternative Text



(c) T-Intersection, Channelized through Movement

19.2-4 Full Alternative Text

If an exclusive left-turn lane is provided for the WB left-turn movement and if separate lanes for left and right turns are provided on the stem of the T, a more efficient phasing can be implemented. In this plan, the intersection geometry is used to allow several vehicular movements to use two of the three phases, including some overlaps between right turns from one street and selected movements from the other. This is illustrated in Figure 19.8(b).

If a left-turn lane for the WB left turn can be combined with a channelizing island separating the WB through movement from all other vehicle paths, a signalization can be adopted in which the WB through movement is never stopped. Figure 19.8(c) illustrates this approach. Note that this particular approach can be used only where there are no pedestrians present or where an overpass or underpass is provided for those crossing the E–W artery.

In each of the cases shown in Figure 19.8, a three-phase signal plan is used. Using geometry, however, additional movements can be added to each of the signal phases, improving the overall efficiency of the signalization. As the signal plan becomes more efficient, delays to drivers and passengers will be reduced, and the capacity for each movement will be increased.

Offset intersections present unique challenges, as the paths of vehicles traversing the intersection create markedly different conflict points between vehicular paths and between vehicle and pedestrian paths. Phase plans must be adjusted to account for this characteristic, which is often difficult for drivers to discern. Figure 19.9 illustrates an offset intersection.

Figure 19.9: An Offset Intersection

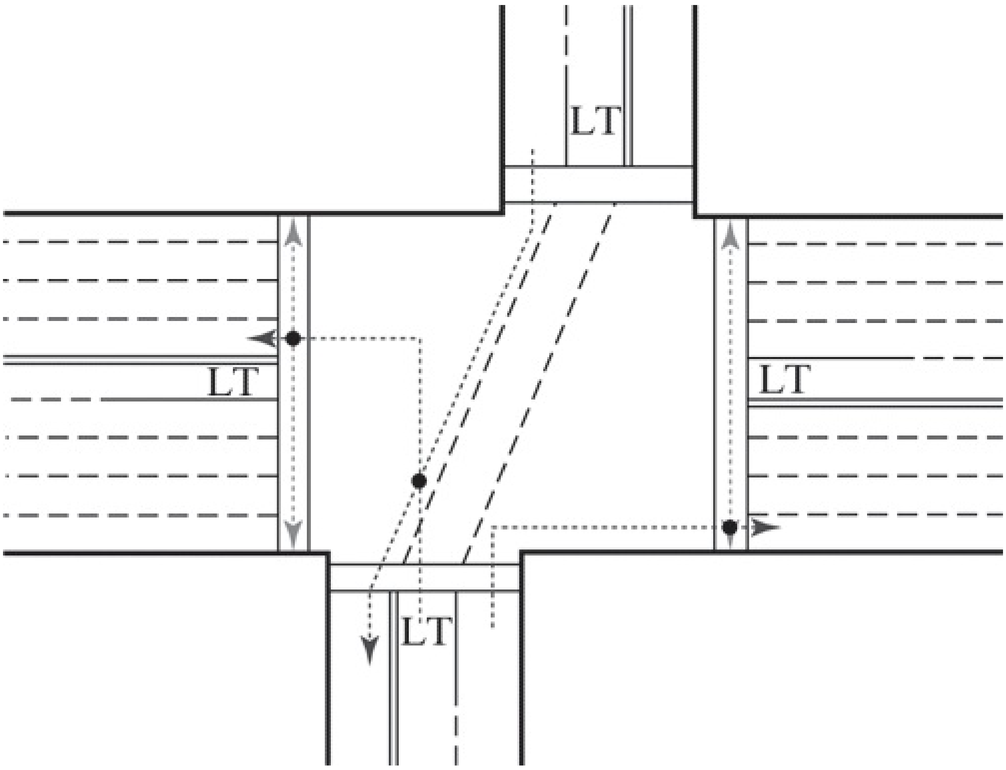


Figure 19.9: Full Alternative Text

There are three particular conflicts that occur at nontypical locations within the intersection, each of which can be extraordinarily dangerous:

- The NB left turn conflicts with the opposing through vehicles almost immediately on entering the intersection. This is because the left-turn trajectory moves immediately into conflict with opposing traffic, which is entering at an angle.
- Pedestrians crossing the E–W street (on either side) encounter vehicular conflicts at unusual locations—which neither the driver nor

the pedestrian are expecting.

The left-turn conflict is critical. While left-turning drivers *should* begin to cross the intersection by veering to the right (to avoid the opposing lane), this is not the normal path in typical intersections. Many, if not most, drivers will follow the typical path, which involves the danger of conflict in an unexpected location. The only way to avoid this conflict is to provide for a fully protected LT phase for the NB left turn, even if the normal guidelines for doing so are not met.

An exclusive NB LT phase (which would probably be accompanied by a SB LT phase for efficiency) would also avoid the unexpected LT—pedestrian conflict on the west crosswalk. The NB RT—pedestrian conflict is more difficult. RT drivers will encounter pedestrian conflicts after they have effectively completed their right turns. Pedestrians crossing with the light would normally not look for conflicting vehicles approaching directly from the left. Warning signs will help both drivers and pedestrians to be more vigilant. Another option would be to tilt the crosswalks to be parallel to the vehicular paths. This would place the RT—pedestrian conflict to a more normal location. This, however, also increases the amount of time that pedestrians need to cross the street, and will affect signal timing.

The seriousness of these problems involves many factors, including the actual offset distance for the intersection. Smaller offsets are easier to handle, and problems increase as the offset distance increases. It also depends on whether the offset is to the right or to the left. Figure 19.9 is a right-offset. A left-offset would change the relative trajectories, but would pose similar problems that would have to be addressed.

Multileg intersections (more than four legs) are a traffic engineer's worst nightmare. Although somewhat rare, these intersections do occur with sufficient frequency to present major problems in signal networks. In the worst case, an intersection could involve three two-way streets. Even if all left turns can be made on a permitted basis, three phases would result. For every opposed left turn that requires protection, another phase would be added. Thus, if all left turns needed protected phases at such a six-leg intersection, there would be six phases—a totally inefficient approach with very large lost times built into the cycle.

Where such situations arise, it is almost always necessary to simplify movements at the intersection. Making one or more of the intersection streets one-way would greatly simplify the number of legal movements that have to be signalized, but this is not always a practical solution. Banning some of the left turns can also help, but alternative paths need to be available and feasible.

Figure 19.10 illustrates a five-leg intersection, in this case, formed by an off-ramp from a limited access facility that feeds directly into a signalized intersection.

Figure 19.10: A Five-Leg Intersection

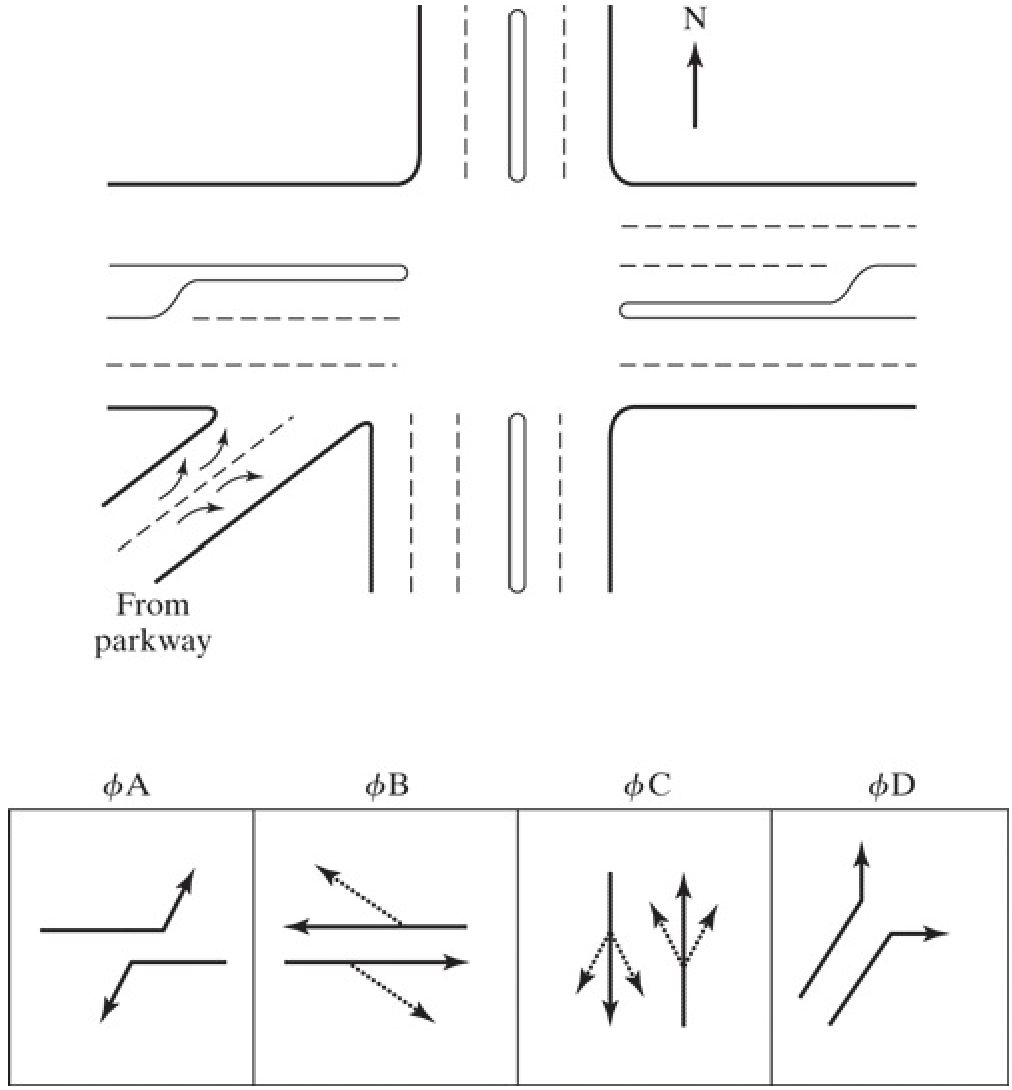


Figure 19.10: Full Alternative Text

In the example shown, a four-phase signal phase plan is needed to provide a protected left-turn phase for the E–W artery. Had the N–S artery required a protected left-turn phase as well (an exclusive LT lane would have to be provided), then a five-phase signalization could have resulted. Addition of a fifth, and, potentially, even a sixth phase creates inordinate amounts of lost time, increases delay, and reduces capacity to critical approaches and lane groups.

Wherever possible, design alternatives should be considered to eliminate five- and six-leg intersections. In the case illustrated in Figure 19.10, for example, redesign of the ramp to create another separate intersection should be considered. The ramp could be connected to either of the intersecting arteries in a T-intersection. The distance from the new intersection to the main intersection would be a critical feature and should be arranged to avoid queuing that would block egress from the ramp. It may be necessary to signalize the new intersection as well.

In Manhattan (New York City), Broadway created a major problem in traffic control. The street system in most of Manhattan is a perfect grid, with the distance between N–S avenues (uptown/downtown) a uniform 800 feet, and the distance between E and W streets (crosstown) an average of about 400 feet. Such a regular grid, particularly when combined with a one-way street system (initiated in the early 1960s), is relatively easy to signalize. Broadway, however, runs diagonally through the grid, creating a series of major multileg intersections involving three major intersecting arteries. Some of these “major” intersections include Times Square and Herald Square, and all involve major vehicular and major pedestrian flows.

To take advantage of the signalization benefits of a one-way, uniform grid street system, through flow on Broadway is banned at most of these intersections. This has effectively turned Broadway into a local street, with little through traffic. Through traffic is forced back onto the grid. Channelization is provided that forces vehicles approaching on Broadway to join either the avenue or the street, eliminating the need for multiphase signals. In addition, channelizing islands have been used to create unique pedestrian environments at these intersections.

Pure signalization solutions to multileg intersections are rarely efficient, but may become necessary. More radical changes in design are often considered in such cases:

- In Washington, D.C., a grid street pattern superimposed on L'Enfant's classic radial design results in numerous multileg intersections. In many cases, underpasses have been designed to take major artery through traffic under the intersection without signalization.
- Where sufficient land is available, roundabouts can be used to eliminate the need for signalization. There are limitations on the total traffic volume that can be handled by roundabouts, however.

Right-Turn Phasing

Although the use of protected left-turn phasing is common, the overwhelming majority of signalized intersections handle right turns on a permitted basis, mostly from shared lanes. Protected right-turn phasing is used only where the number of pedestrians is extremely high. Modern studies show that a pedestrian flow of 1,700 peds/h in a crosswalk can effectively block all right turns on green. Such a pedestrian flow is, however, extremely rare, and exists only in major urban city centers. Although use of a protected right-turn phase in such circumstances may help motorists, it may worsen pedestrian congestion on the street corner and on approaching sidewalks. In extreme cases, it is often useful to examine the feasibility of pedestrian overpasses or underpasses. These would generally be coordinated with barriers preventing pedestrians from entering the street at the corner. It should be noted, however, that pedestrian overpasses and/or underpasses are inconvenient for pedestrians and may pose security risks, primarily at night.

Compound right-turn phasing is usually implemented only in conjunction with an exclusive left-turn phase on the intersecting street. For example, NB and SB right turns may be without pedestrian interference during an EB and WB exclusive left-turn phase. Permitted right turns would then continue during the NB and SB through phase.

Exclusive right-turn lanes are useful where heavy right-turn movements exist, particularly where right-turn-on-red is permitted. Such lanes can be easily created on streets where curb parking is permitted. Parking may be prohibited within several hundred feet of the STOP line; the curb lane may then be used as an exclusive right-turn lane. Channelized right turns may also be provided. Channelized right turns are generally controlled by a YIELD sign and need not be included in the signalization plan. Chapter 17 contains a more detailed discussion of exclusive right-turn lanes and channelized right-turn treatments.

Right-Turn-on-Red

"Right-turn-on-red" (RTOR) was first permitted in California in 1937 only in conjunction with a sign authorizing the movement [5]. In recent years, virtually all states allow RTOR unless it is specifically prohibited by a sign. The federal government encouraged this approach in the 1970s by linking implementation of RTOR to receipt of federal-aid highway funds. In some urban areas, like New York City, right-turn-on-red is still generally prohibited. Signs indicating this general prohibition must be posted on all roadways entering the area. All RTOR laws require that the motorist stop before executing the right-turn movement on red.

When implemented using a shared right-turn through lane, the utility of RTOR is affected by the proportion of through vehicles using the lane. When a through vehicle reaches the STOP line, it blocks subsequent right-turners from using RTOR. Thus, provision of an exclusive right-turn lane greatly enhances the effectiveness of RTOR.

The major issues regarding RTOR continue to be (1) the delay savings to right-turning vehicles, and (2) the increased accident risk such movements cause. An ITE practice [6] states that the delay to an average right-turning vehicle is reduced by 9% in central business districts (CBDs), 31% in other urban areas, and 39% in rural areas. Another early study on the safety of RTOR [7] found that only 0.61% of all intersection accidents involved RTOR vehicles and that these accidents tended to be less severe than other intersection accidents.

Because there are potential safety issues involving RTOR, its use, and application should be carefully considered. The primary reasons for prohibiting RTOR include:

1. Restricted sight distance for right-turning motorist

2. High speed of conflicting through vehicles

3. High flow rates of conflicting through vehicles
4. High pedestrian flows in crosswalk directly in front of right-turning vehicles

Any of these conditions would make it difficult for drivers to discern and avoid conflicts during the RTOR maneuver.

19.2.6 Summary and Conclusion

The subject of phasing along with the selection of an appropriate phase plan is a critical part of effective intersection signalization. Although general criteria have been presented to assist in the design process, there are few firm standards. The traffic engineer must apply a knowledge and understanding of the various phasing options and how they affect other critical aspects of signalization, such as capacity and delay.

Phasing decisions are made for each approach on each of the intersection streets. It is possible, for example, for the E–W street to use an exclusive left-turn phase while the N–S street uses leading and lagging greens and compound phasing. The number of potential combinations for the intersection as a whole, therefore, is large.

The final signalization should also be analyzed using a comprehensive signalized intersection model or simulation. This allows for fine tuning of the signalization on a trial-and-error basis and for a wider range of alternatives to be quickly assessed.

19.3 Determining Vehicular Requirements for Signal Design and Timing

Once a candidate phase plan has been established, it is possible to establish the "timing" of the signal that would most effectively accommodate the vehicular demands present.

19.3.1 Change and Clearance Intervals

Despite not being intuitive, the timing process starts with the determination of *yellow* (change) and *all-red* (clearance) intervals. This is because other critical elements like cycle length and effective green times all require knowledge of lost times—which are fundamentally tied to the lengths of change and clearance intervals.

The terms "change" and "clearance" interval are used in a variety of ways in the literature. They refer to the *yellow* and *all-red* indications, respectively, that mark the transition from GREEN to RED in each signal phase. The *all-red* interval is a period during which all signal faces show a RED indication. The MUTCD specifically prohibits the use of a *yellow* indication to mark the transition from RED to GREEN, a practice common in many European countries.

The use of *yellow* and *all-red* intervals varies from state to state. A *yellow* interval is required in all states, but its legal meaning varies. A *permissive yellow law* means that a driver can enter the intersection during the entire *yellow* interval, and be in the intersection during the *all-red,* as long as he/she entered during the *yellow*. Where a *permissive yellow law* is in effect, an *all-red* interval must be provided.

A *restrictive yellow law* is more difficult to interpret, particularly because two different versions are in use across the country. In one case, a driver may only enter during the *yellow* if he/she can clear the intersection before the *yellow* ends. This makes it very difficult for the driver, as the end of the *yellow* is not known when the decision to enter the intersection is made. In an even more difficult version of the law, a driver may not enter the intersection on *yellow* unless it is unsafe or impossible to stop. The use of an *all-red* interval is optional when a restrictive yellow law is in effect.

Because of the variety of laws in effect, the MUTCD does not strictly require *all-red* intervals. The Institute of Transportation Engineers (ITE) recommends that both *yellow* and *all-red* intervals be used at all signals. ITE defines them as follows:

- Change interval (yellow). This interval allows a vehicle that is one safe stopping distance away from the STOP line when the GREEN is withdrawn to continue at the approach speed and enter the intersection legally on *yellow*. "Entering the intersection" is interpreted to be the front wheels crossing over the intersection curb line.

- Clearance interval (all-red). Assuming that a vehicle has just entered the intersection legally on *yellow*, the *all-red* must provide sufficient time for the vehicle to cross the intersection and clear its back bumper past the far curb line (or crosswalk line) before conflicting vehicles are given the GREEN.

The ITE recommends the following methodology for determining the length of the *yellow* or change interval [8]:

yi=t+1.47 S85i2×(a+32.2 Gi) [19-3]

where:

yi=length of the yellow interval for Phase i, s,t=driver reaction time, s (stan

Also, note that 32.2 ft/s^2 is the acceleration rate due to gravity. Downgrades have a negative value of *G*.

This equation was derived as the time required for a vehicle to traverse one safe stopping distance at its approach speed.

The ITE also recommends the following policy for determining the length of *all-red* clearance intervals [8]:

- For cases in which there is no or little pedestrian traffic:

  ari=w+L1.47 S15i [19-4]

- For cases in which significant pedestrian traffic exists:

ari=P+L1.47 S15i [19-5]

where:

ari=length of the all-
red phase for Phase i, s,w=width of the street being crossed, from curbto cu

ITE also allows for a third alternative where there are noticeable but not significant numbers of pedestrians present. It is generally preferable, however, to use one of the two interpretations above based upon whether or not pedestrians need to be considered in the timing of the *all-red* interval.

As noted previously, where *restrictive yellow laws* are in effect, the *all-red* interval is not required. When only a *yellow* interval is used, however, its length must include adequate time to clear the intersection. Thus, in these cases, the *yellow* interval would be the sum of Equations 19-3 and 19-4 or 19-5 as appropriate.

The choice between Equations 19-4 or 19-5 is illustrated in Figure 19.11. The choice is based upon the safety of pedestrians in each case.

Figure 19.11: The Influence of Pedestrians on Clearance (*All-Red*) Intervals

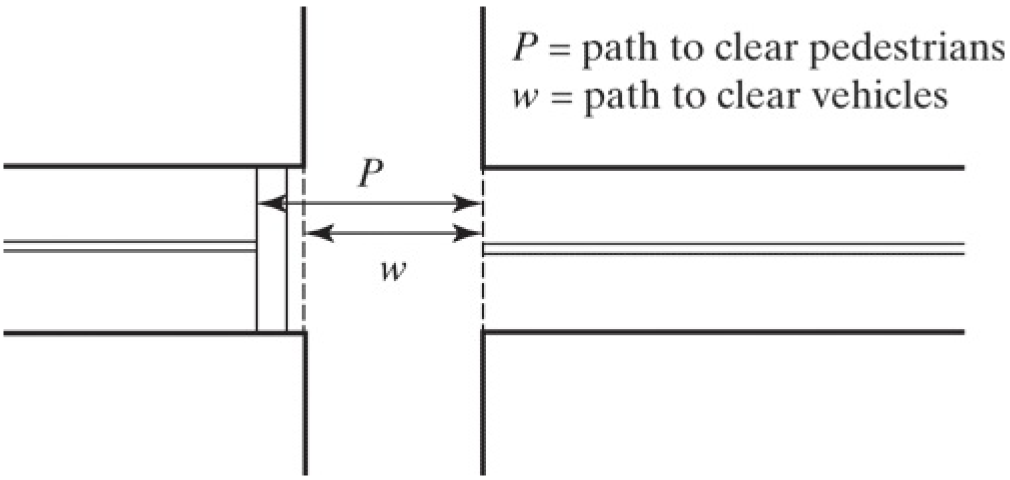


Figure 19.11: Full Alternative Text

When Equation 19-4 is used, the *all-red* interval allows the vehicle to cross the width of the roadway plus one car length during the interval. This guarantees that when the conflicting movements are released in the next *green* phase, that crossing vehicles will be out of the paths of conflicting through movements. Equation 19-5 allows the vehicle to cross to the far edge of the far crosswalk plus one car length within the *all-red* interval. This guarantees that when conflicting vehicles are released on the next *green* phase, they will be out of the paths of both conflicting vehicles and conflicting pedestrians.

Where small numbers of pedestrians are present, it is reasonable to expect them to avoid conflicts with crossing vehicles at the very end of the phase; where numbers are more significant, relying solely on pedestrian and driver judgment to avoid potential conflicts is probably not wise.

To provide for optimal safety, the equations for *yellow* and *all-red* intervals use different speeds: the 85th percentile and the 15th percentile, respectively. Because speed appears in the numerator of the *yellow* determination and in the denominator of the *all-red* determination, accommodating the majority of motorists safely requires the use of different percentiles. If only the average approach speed is known, the percentile speeds may be estimated as:

$$S_{85}=S_{av}+5 \quad S_{15}=S_{av}-5 \quad [19\text{-}6]$$

where:

S_{85} = 85th percentile speed, mi/h, S_{15} = 15th percentile speed, mi/h, and S_{av} =

Where approach speeds are not measured and the speed limit is used, both the *yellow* and *all-red* intervals will be determined using the same value of speed. This, however, is not a desirable practice.

Use of these ITE policies to determine *yellow* and *all-red* intervals assures that drivers will not be presented with a “dilemma zone,” which occurs when the combined length of the change and clearance intervals is not sufficient to allow a motorist who cannot safely stop when the *yellow* is initiated to cross through the intersection and out of conflicting vehicular and/or pedestrian paths before those flows are released. Where *yellow* and *all-red* phases are mistimed and a dilemma zone is created, agencies face possible liability for accidents that occur as a result.

Sample Problem 19-1: Determining the Length of the Change (Yellow) and Clearance (All-Red) Intervals

Compute the appropriate change and clearance intervals for a signalized intersection approach with the following characteristics:

- Average approach=35 mi/h
- Grade=−2.5%
- Distance from STOP line to far side of the most distant lane=48ft
- Distance from STOP line to far side of the most distant cross-walk=60 ft
- Standard vehicle length=20ft
- Reaction time=1.0s
- Deceleration rate=10ft/s2

- Some pedestrians present

To apply Equation 19-3 and 19-4/19-5, estimates of the 15th and 85th percentile speeds are needed. Using Equations 19-6:

$$S_{85} = 35 + 5 = 40 \text{ mi/h} \qquad S_{15} = 35 - 5 = 30 \text{ mi/h}$$

Using Equation 19-3, the length of the change or *yellow* interval should be:

$$y = t + \frac{1.47\, S_{85}}{2 \times (a + 32.2\, G)} \qquad y = 1.0 + \frac{1.47 \times 40}{2 \times [10 + (32.2 \times -0.025)]} = 1.0 + 3.2 = 4.2 \text{ s}$$

Equation 19-5 is used to compute the length of the clearance or *all-red* phase because there are some pedestrian flows present. This is a judgment call, as with a small number of pedestrians, Equation 19-4 might also be used. Erring on the side of caution, the length of the clearance interval is:

$$ar = \frac{P + L}{1.47\, S_{15}} \qquad ar = \frac{60 + 20}{1.47 \times 30} = 1.8 \text{ s}$$

19.3.2 Determining Lost Times

The *Highway Capacity Manual* [1] indicates that lost times vary with the length of the *yellow* and *all-red* phases in the signal timing. Thus, it is no longer appropriate to use a constant default value for lost times as was historically done in many signal timing methodologies. The HCM now recommends the use of the following default values for this determination:

- Start-up lost time, $\ell_1 = 2.0$ s/phase
- Motorist use of *yellow* and *all-red*, $e = 2.0$ s/phase

Using these default values, lost time per phase and lost time per cycle may be estimated as follows:

$$\ell_{2i} = y_i + ar_i - e \qquad tL_i = \ell_{1i} + \ell_{2i} \quad [19\text{-}7]$$

where:

ℓ_{1i} = start-up lost time for phase i, s, ℓ_{2i} = clearance lost time for phase i, s, tL_i = total los red clearance interval for phase i, s.

Sample Problem 19-2: Determining Lost Times

In Sample Problem 19-1, the *yellow* interval was computed as 4.2 seconds, and the *all-red* interval was found to be 1.8 seconds. Using the recommended default values for ℓ_1 and *e*, respectively, determine the lost times in the signal cycle.

The start-up lost time is the standard value of 2.0 s/phase. The clearance lost time is computed as:

$$\ell_2 = 4.2 + 1.8 - 2.0 = 4.0 \text{ s} \quad t_L = 2.0 + 4.0 = 6.0 \text{ s}$$

Note that when the HCM-recommended default values for ℓ_1 and *e* (both 2.0 s) are used, the lost time per phase, t_L is always equal to the sum of the *yellow* and *all-red* intervals, *Y*. Because the lost time for each phase may differ, based on different *yellow* and *all-red* intervals, the total lost time per cycle is merely the sum of lost times in each phase, or:

$$L = \sum_i^n t_{Li} \quad [19\text{-}8]$$

where:

L = total lost time per cycle, s
t_{Li} = total lost time for phase i, s, and
n = number

19.3.3 Determining the Sum of Critical-Lane Volumes

To estimate an appropriate cycle length and to split the cycle into appropriate green times for each phase, it is necessary to find the *critical-lane volume* for each discrete phase or portion of the cycle.

As discussed in Chapter 18, the *critical-lane volume* is the per-lane volume that controls the required length of a particular phase. For example, in the case of a simple two-phase signal, on a given phase the EB and WB flows move simultaneously. One of these per-lane volumes represents the most intense demand, and that is the one that will determine the appropriate length of the phase.

Making this determination is complicated by two factors:

- Simple volumes cannot be simply compared. Trucks require more time than passenger cars, left and right turns require more time than through vehicles, vehicles on a downgrade approach require less time than vehicles on a level or upgrade approach. Thus, *intensity* of demand is not measured accurately by simple volume.

- Where phase plans involve overlapping elements, the ring diagram must be carefully examined to determine which flows constitute *critical-lane volumes*.

Ideally, demand volumes would be converted to equivalents based on all of the traffic and roadway factors that might affect intensity. For initial signal timing, however, this is too complex a process. Demand volumes can, however, be converted to reflect the influence of the most significant factors affecting intensity: left and right turns. This is accomplished by converting all demand volumes to *equivalent through vehicle units* (tvu). Through vehicle equivalents for left and right turns are shown in Tables 19.4 and 19.5, respectively.

Table 19.4: Through-Vehicle Equivalents for Left-Turning Vehicles, E_{LT}

Opposing through Flow, V_o (veh/h)	Number of Opposing through Lanes		
	1	2	≥3
0	1.1	1.1	1.1
200	2.5	2.0	1.8
400	5.0	3.0	2.5
600	10.0*	5.0	4.0
800	13.0*	8.0	6.0
1,000	15.0*	13.0*	10.0*
≥ 1,200	15.0*	15.0*	15.0*
E_{LT} for all *protected* LTs = 1.05			

* For these situations, it is likely that all LTs are being made as "sneakers."

Table 19.4: Full Alternative Text

These values are actually a simplification of a more complex approach in the *Highway Capacity Manual* analysis model for signalized intersections, and they form an appropriate basis for signal timing and design. In using these tables, the following should be noted:

- Opposing volume, Vo, includes only the through volume on the opposing approach, in veh/h.
- Interpolation in Table 19.4 for opposing volume is appropriate, but values should be rounded to the nearest tenth.
- For right turns, the "conflicting crosswalk" is the crosswalk through which right-turning vehicles must pass.
- Pedestrian volumes indicated in Table 19.5 represent typical situations in moderate-sized communities. Pedestrian volumes in large cities, like New York, Chicago, or Boston, may be much higher, and the relative terms used (low, moderate, high, extreme) are not well correlated to such situations.

Table 19.5: Through-Vehicle Equivalents for Right-Turning Vehicles, E_{RT}

Category of Pedestrian Flow in Conflicting Crosswalk	Default Pedestrian Flow Rate (peds/h)	Equivalent
None	0	1.18
Low	50	1.21
Moderate	200	1.32
High	400	1.52
Extreme	800	2.14

Table 19.5: Full Alternative Text

- If exact pedestrian crosswalk counts are available, interpolation in Table 19.5 is permitted to the nearest 0.01. If only rough estimates of pedestrian activity are provided, interpolation is not recommended.

Once appropriate values for ELT and ERT have been selected, all right- and left-turn volumes must be converted to units of "through-vehicle equivalents." Subsequently, the demand intensity *per lane* is found for each approach or lane group.

VLTE=VLT × ELTVRTE=VRT × ERT [19-9]

where:

VLTE=equivalent LT volume in through-vehicleequivalents, tvu/h, andVRTE=equivalent RT volume in through-vehicleequivalents, tvu/h.

Other variables are as previously defined.

These equivalents are added to through vehicles that may be present in a given approach or lane group to find the total equivalent volume and equivalent volume per lane in each approach or lane group:

VEQ=VLTE+VTH+VRTEVEQL=VEQ/N [19-10]

where:

VEQ=total equivalent volume in a lane group orapproach, tvu/h,VEQL=tot

Finding the critical-lane volumes for the signal phase plan requires determining the critical path through the plan (i.e., the path that controls the signal timing). This is done by finding the path through the signal phase plan that results in the highest possible sum of critical-lane volumes. Because most signal plans involve two "rings," alternative paths must deal with two potential rings for each discrete portion of the phase plan. It must also be noted that the critical path may "switch" rings at any full phase barrier (i.e., a phase boundary that cuts through both rings). This process is best understood using a sample problem.

Sample Problem 19-3: Determining the Critical Path Through a Ring Diagram

Figure 19.12 shows a ring diagram for a signalization with overlapping phases. Lane volumes, VEQL, are shown for each movement in the phase diagram.

Figure 19.12: Determining Critical Lane Volume Illustrated

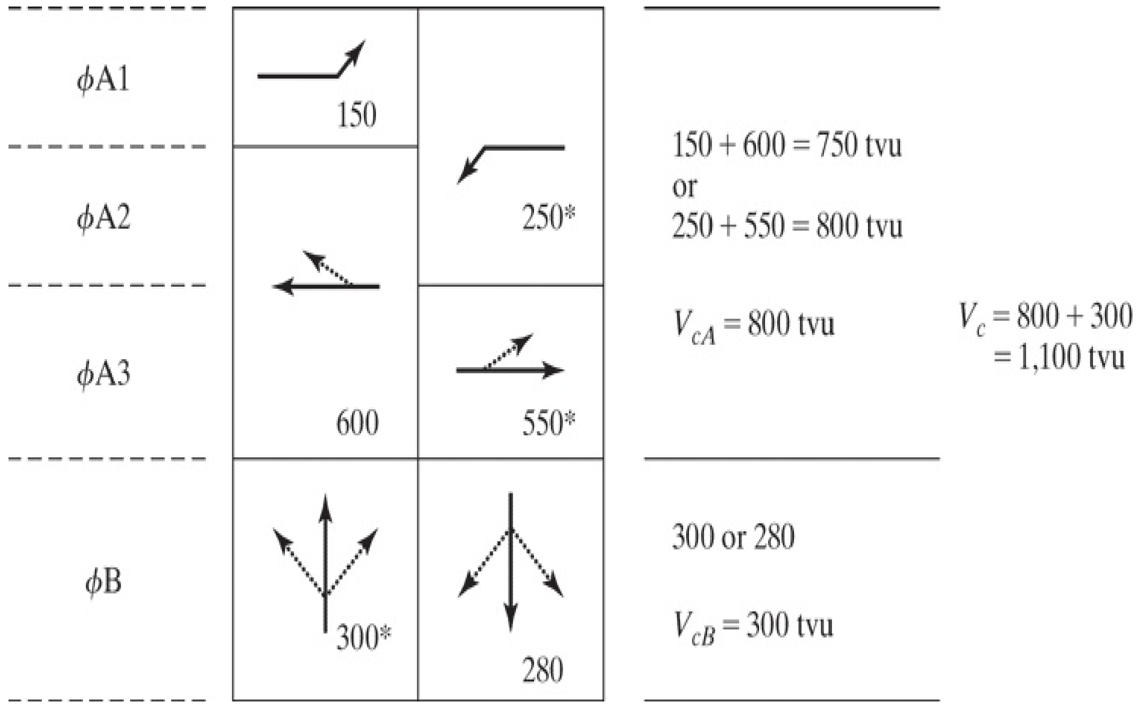


Figure 19.12: Full Alternative Text

To find the critical path, the controlling (maximum) equivalent volumes must be found for each portion of the cycle, working between full-phase transition boundaries. For the combined Phase A in Figure 19.12, the volumes that control the total length of A1, A2, and A3 are on Ring 1 or Ring 2. As shown, the maximum total comes from Ring 2 and yields a total critical-lane volume of 800 tvu. For Phase B, the choice is much simpler because there are no overlapping phases. Thus, the Ring 1 total of 300 tvu is identified as critical. The critical path through the cycle is now indicated by asterisks, and the sum of critical-lane volumes is $800 + 300 = 1{,}100$ tvus. In essence, if the intersection is thought of in terms of a number of vehicles in single lanes seeking to move through a single common conflicting point, the signal, in this case, must have a timing that is sufficient to handle 1,100 tvu through this point. The determination of critical-lane volumes is further illustrated in the complete signal-timing sample problems included in the last section of this chapter.

19.3.4 Determining the Desired Cycle Length

In Chapter 18, an equation describing the maximum sum of critical-lane volumes that could be handled by a signal was manipulated to find a desirable cycle length. That equation is used to find the desired cycle length, based on volumes (in tvu) and a default value for saturation flow rate. The default saturation flow rate, 1,700 tvu per hour of green, assumes typical conditions of lane width, heavy-vehicle presence, grades, parking, pedestrian and bicycle volumes, local buses, area type, lane utilization and other conditions. Common default values for saturation flow rate range from 1,500 to 1,700 in the literature, but these sometimes also account for typical left-turn and right-turn percentages as well. The method presented here makes these adjustments by converting demand to equivalent through-vehicle units.

When the default value for saturation flow rate is inserted into the relationship, the desired cycle length is computed as:

$$C_{des} = \frac{L}{1 - \left[\frac{V_c}{1700 \times PHF \times (v/c)}\right]} \quad [19\text{-}11]$$

where:

C_{des} = desirable cycle length, s,
L = total lost time per cycle, s,
PHF = peak-hour factor, and
(v/c) = target v/c ratio for the critical movements in the inters

Use of the peak-hour factor ensures that the signal timing is appropriate for the peak 15 minutes of the design hour. Target v/c ratios are generally in the range of 0.85 to 0.95. Very low values of v/c increase delays because vehicles are forced to wait while an underutilized green phase times out. Values of $v/c > 0.95$ indicate conditions in which frequent individual phase or cycle failures are possible, thereby increasing delay.

Sample Problem 19-4: Determining an Appropriate Cycle Length

Consider Sample Problem 19-3, illustrated previously in Figure 19.12. The sum of the critical-lane volumes for this case was shown to be 1,100 veh/h. What is the desirable cycle length for this three-phase signal if the total lost time per cycle is 4 s/phase × 3 phases/cycle = 12 s/cycle, the peak-hour factor is 0.92, and the target v/c ratio is 0.90? Using Equation 19-11:

$$C_{des}=121-\left[\frac{1,100}{1,700\times0.92\times0.90}\right]=121-0.781=54.8\ \text{s}$$

For a pre-timed signal, cycle lengths are generally implemented in 5-second increments between cycle lengths of 30 and 90 seconds, and in 10-second increments between 90 and 120 seconds. Thus, a 55-second cycle would be adopted in this case.

19.3.5 Splitting the Green

Once the cycle length is determined, the available *effective green time* in the cycle must be divided among the various signal phases. The available effective green time in the cycle is found by deducting the lost time per cycle from the cycle length:

$$g_{TOT}=C-L \quad [19\text{-}12]$$

where:

g_{TOT}=total effective green time in the cycle, s,and C, L are as previously c

The total effective green time is then allocated to the various phases or subphases of the signal plan in proportion to the critical lane volumes for each phase or subphase:

$$g_i=g_{TOT}\times\left(\frac{V_{ci}}{V_c}\right) \quad [19\text{-}13]$$

where:

g_i=effective green time for Phase i, s,V_{ci}=critical lane volume for Phase or

Sample Problem 19-5: Determining Green Times

Returning to the example of Sample Problems 19-3 and 19-4, illustrated in Figure 19.12, the situation is complicated somewhat by the presence of overlapping phases. For the critical path, the following critical-lane volumes were obtained:

- 250 veh/h/ln for the *sum* of Phases A1 and A2
- 550 veh/h/ln for Phase A3
- 300 veh/h/ln for Phase B

Remembering that the desired cycle length of 55 seconds contains 12 seconds of lost time, the total effective green time in the cycle may be computed using Equation 19-12:

$$g_{TOT}=55-12=43\text{ s}$$

Using Equation 19-13 and the critical-lane volumes just noted, the effective green times for the signal are estimated as:

$$g_{A1+A2}=43\times\left(\frac{250}{1100}\right)=9.8\text{ s}\quad g_{A3}=43\times\left(\frac{550}{1100}\right)=21.5\text{ s}\quad g_{B}=43\times(300\,110$$

The sum of these times (9.8 + 21.5 + 11.7) must equal 43.0 seconds, and it does. Together with the 12.0 seconds of lost time in the cycle, the 55-second cycle length is now fully allocated.

Because of the overlapping phases illustrated in this example, the signal timing is still not complete. The split between phases A1 and A2 must still be addressed. This can be done only by considering the noncritical Ring 1 for Phase A because this ring contains the transition between these two subphases. The total length of Phase A is 9.8 + 21.5 = 31.3 s. On the noncritical ring (Ring 1), critical-lane volumes are 150 for Phase A1 and 600 for the sum of Phases A2 and A3. Using these critical-lane volumes:

$$g_{A1}=31.3\times\left(\frac{150}{150+600}\right)=6.3\text{ s}$$

By implication, g_{A2} is now computed as the total length of Phase A, 31.3 seconds, minus the effective green times for Phases A1 and A3, both of which have now been determined (6.3 and 21.5 seconds, respectively). Thus:

$$g_{A2}=31.3-6.3-21.5=3.7\text{ s}$$

The signal timing is now complete except for the conversion of effective green times to actual green times.

$G_i = g_i + \ell_1 - e$ [19-14]

where:

G_i=actual green time for Phase i, s,g_i=effective green time for Phase i, s, ar red, s.

Because information on the timing of yellow and all-red phases was not provided for the example, this step cannot be completed. If, however, standard default values are in use for ℓ_1 and *e* (both 2.0 seconds), then the values of *G are the same as the values of* g.

Full signal-timing examples in the last section of this chapter will fully illustrate determination of actual green times.

As a general rule, very short phases should be avoided. In this case, the overlapping Phase A2 has an effective green time of only 3.7 seconds. This short overlap period may not provide sufficient efficiency to warrant the potential confusion of drivers. The short Phase A2, in this case, may be one argument in favor of simplifying the phase plan by using a common exclusive left-turn phase.

19.4 Determining Pedestrian Signal Requirements

To this point in the process, the signal design has considered vehicular requirements. Pedestrians, however, must also be accommodated by the signal timing. Problems arise because pedestrian requirements and vehicular requirements are often quite different. Consider the intersection of a wide major arterial and a small local collector. Vehicle demand on the major arterial is more intense than on the small collector, and the green split for vehicles would generally result in the arterial receiving a long green and the collector a relatively short green.

This, unfortunately, is exactly the opposite of what pedestrians would require. During the short collector green, pedestrians are crossing the wide arterial. During the long arterial green, pedestrians are crossing the narrower collector. In summary, pedestrians require a longer green during the shorter vehicular green, and a shorter green during the longer vehicular green.

Pedestrians require a minimum amount of time to begin to cross the street, and to safely complete the crossing. The minimum pedestrian green requirement is given by:

$G_{pi}=PW_{min\,i}+PC_i$ [19-15]

where:

G_{pi}=minimum pedestrian green time for Phase$i$, s.,$PW_{min\,i}$=minimum ped

The length of the pedestrian WALK interval (*PW*) depends upon the volume of pedestrians using the controlled crosswalk. Modern signal displays show a green outline of a walking man for this interval. The *Traffic Signal Timing Manual* [2] recommends the values shown in Table 19.6, based on a general description of pedestrian intensity.

Table 19.6: Recommended

Minimum Pedestrian walk Intervals (*pw*)

Conditions	WALK Interval Duration (s)
High pedestrian volume areas (e.g., schools, CBDs, sports venues)	10–15
Typical pedestrian volumes with longer cycle length (≥ 60 s)	7–10
Typical pedestrian volumes with shorter cycle length (< 60 s)	7
Negligible pedestrian volumes	4
Conditions where older pedestrians are present	Distance to center of road divided by 3.0 fps

(Source: *Traffic Signal Timing Manual*, Federal Highway Administration, Washington, D.C., Table 5-8, pg 5-15.)

Table 19.6: Full Alternative Text

The pedestrian clearance interval is computed as:

PCi=LSp [19-16]

where:

PCi=pedestrian clearance interval, s,L=length of the crosswalk, ft., andSp=

The standard walking speed used by many agencies is 4.0 ft/s. Support is growing for use of 3.5 ft/s, however, and some agencies are already using this value. Where older or handicapped pedestrians are prevalent, use of a speed as low as 3.0 ft/s is often appropriate.

Note that pedestrians have a green (WALK) interval and a clearance (Flashing DON’T WALK) interval. The latter is shown as a flashing Portland orange upraised hand. There is, however, no pedestrian change (or yellow) interval. This is because pedestrians can stop almost instantaneously, as opposed to vehicles, where stopping takes considerable time and distance to accomplish.

Depending on local policy, there are three general approaches to coordinating pedestrian and vehicular intervals. The policies differ on the treatment of pedestrians during the vehicular *yellow* and *all-red* intervals. The MUTCD [4] allows for pedestrians to complete their crossings during the vehicular *yellow* and *all-red* intervals, but not all agencies follow this policy. The three policies used are:

- The MUTCD policy: Pedestrians may complete their crossing during the vehicular *yellow* and *all-red* intervals. In this case, the pedestrian clearance interval *ends* when the *all-red* interval terminates. (Option 1)

- Pedestrians may complete their crossing during the vehicular *yellow* interval, but NOT during the vehicular *all-red* interval. In this case, the pedestrian clearance interval *ends* when the *yellow* interval terminates. (Option 2)

- Pedestrians must complete their crossing during the vehicular *green* interval. In this case, the pedestrian clearance interval *ends* when the *green* interval terminates. (Option 3)

The WALK interval may be longer than the minimum recommended for pedestrians. The actual length of the WALK interval depends upon which of the options for pedestrian clearance is in effect:

$PW_i = G_i + y_i + ar_i - PC_i$ (Option 1) $PW_i = G_i + y_i - PC_i$ (Option 2) $PW_i = G_i - PC_i$ (Option 3) [19-17]

where PW_i is the actual length of the pedestrian WALK interval, and all other variables are as previously defined.

For a signal timing to be viable for pedestrians, the minimum pedestrian crossing requirement, G_{pi}, in each phase must be compared with the time that pedestrians may be in the crosswalk, which varies with the policy in effect.

$G_{pi} \leq G_i + y_i + ar_i$ (Option 1) $G_{pi} \leq G_i + y_i$ (Option 2) $G_{pi} \leq G_i$ (Option 3) [19-18]

If the chosen condition is not met, pedestrians are not safely accommodated, and changes must be made to provide for their needs. Where the minimum pedestrian condition is not met in a given phase, two approaches may be taken:

- A pedestrian actuator may be provided. In this case, when pushed, the *next* green phase is lengthened to provide for the required green time. The additional green time is subtracted from other phases for a pre-timed signal to maintain the cycle length, which must remain constant. When pedestrian actuators are provided, pedestrian signals *must* be used.

- Retime the signal to provide the minimum pedestrian need in all cycles. This must be done in a manner that also maintains the vehicular balance of green times and results in a longer cycle length.

The first approach has limited utility. Where pedestrians are present in most cycles, it is reasonable to assume that the actuator will always be pushed, thus destroying the planned vehicular signal timing. In such cases, the approach should be to retime the signal to satisfy both vehicular and pedestrian needs in every cycle. Pedestrian actuators are useful in cases where pedestrians are relatively rare or where actuated signal controllers are used.

In the second case, the task is to provide the minimum pedestrian crossing time while maintaining the balance of effective green needed to accommodate vehicles.

Sample Problem 19-6: Rebalancing Green Times to Account for Pedestrians

Consider the case of the vehicular signal timing for a two-phase signal shown in Table 19.7. Minimum pedestrian needs are also shown for comparison.

Table 19.7: A Sample Signal Timing

Phase	Green Time, G (s)	Yellow Time, y (s)	All-Red Time, ar (s)	Lost Time, t_L (s)	Minimum Pedestrian Green, G_p (s)
A	40	3.0	2.0	5.0	20.0
B	15	3.0	2.0	5.0	30.0

Table 19.7: Full Alternative Text

In this case, Phase A serves a major arterial and thus has the longer vehicular green but the shorter pedestrian requirement. Phase B serves a minor cross-street but has the longer pedestrian requirement. Pedestrian requirements must be compared with the vehicular signal timing, using Equation 19-18. In this case, we will apply the most liberal policy, which allows pedestrians to be in the crosswalk during *G*, *y*, and *ar*.

$$G_{pi} \le G_i + y_i + ar_i \quad G_{pA} = 20.0 \le 40.0 + 3.0 + 2.0 = 45.0\ \text{s} \quad \text{OK} \quad G_{pB} = 30.0 \le 15.0 + 3.0 +$$

Note that because the lost time is equal to the sum of the *yellow* and *all-red* times, effective green, g, is equal to actual green, G, for both phases. Also note that the cycle length is, in this case 40 + 3 + 2 + 15 + 3 + 2 = 65 s.

If pedestrians are present in every cycle, use of pedestrian actuators would be disruptive to the vehicular signal timing. Therefore, the length of Phase B must be increased to provide for 30 seconds to accommodate pedestrian needs. As the *yellow* and *all-red* intervals are fixed, the green must be increased from 15.0 to 25.0 seconds to accomplish this. Then 25 + 3 + 2 = 30 s, which would meet pedestrian requirements.

If we increase the length of green for Phase B, we must also increase the length of green for Phase A to maintain the current balance between them (a ratio of 40/15). There are several ways to accomplish this arithmetically. This, however, is a pre-timed signal, and the resulting cycle length must be a multiple of 5 seconds (or 10 seconds for cycle lengths above 90 s). The simplest way to do this is to increase the cycle length by a factor of 25.0/15.0, the ratio of the required Phase B green to the actual Phase B green. Then:

$$C=65\times\left(\frac{25}{15}\right)=108.3\ \text{s},\ \ \text{SAY}\ \ 110\ \text{s}$$

There is now a total of 110 − 5 − 5 = 100 s of effective green time in the cycle, which, in this case, is also the amount of actual green time. It should be allocated in the original ratio of green times, 40 seconds for Phase A, and 15 seconds for Phase B:

$$G_A=100\times\left(\frac{40}{40+15}\right)=72.7\ \text{s} \quad G_B=100\times\left(\frac{15}{40+15}\right)=27.3\ \text{s}$$

This revised signal timing now meets all vehicular and pedestrian requirements. The “cost” of this, however, is a markedly longer cycle length, which would lead to some increase in delay to drivers and passengers. Pedestrian safety, however, can never be compromised.

Several key characteristics of the relationship between vehicular and pedestrian intervals are illustrated in Figure 19.13:

Figure 19.13: Pedestrian and Vehicular Intervals Illustrated

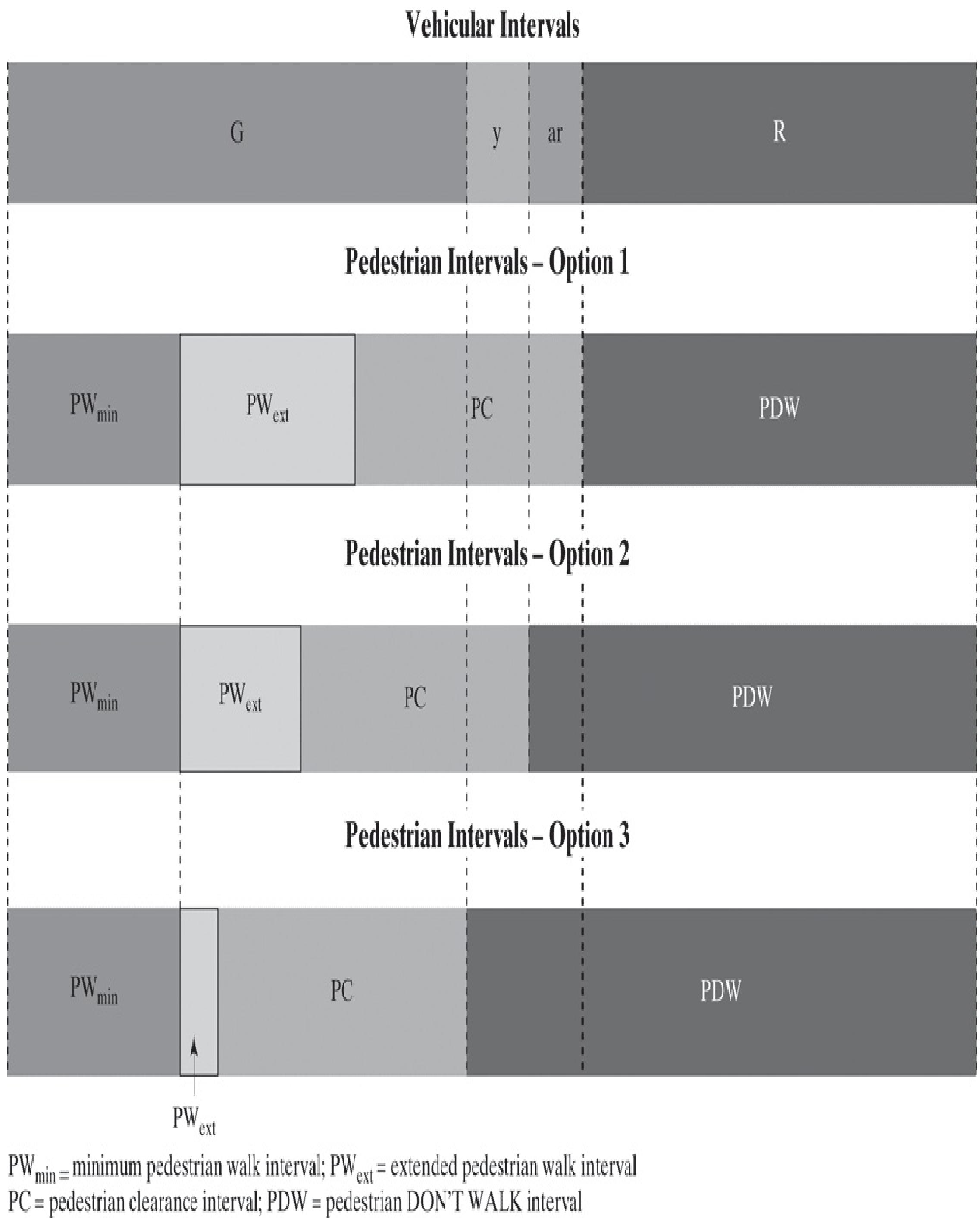


Figure 19.13: Full Alternative Text

- The pedestrian WALK interval begins when the vehicular green begins.
- The pedestrian DON'T WALK ends when the vehicular red ends.
- The pedestrian clearance interval is a constant based on the length of the crosswalk(s). The end of the pedestrian clearance interval depends

upon the option implemented. It occurs at the beginning of the vehicular red for Option 1, the beginning of the vehicular all-red for Option 2,and the beginning of the vehicular yellow in Option 3.

- The pedestrian WALK interval can be extended beyond the minimum requirement if the cycle length is long enough. The minimum value of the extension is “0.”

- Remember that there will be “yellow” between the green arrow for compound left turns, and the green ball as the phase transitions from protected to permitted (or vice versa). This yellow counts as green time for left turns.

19.5 Compound Signal Phasing

Although it is recommended that most initial signal timings avoid compound phasing (protected + permitted or permitted + protected), the methodology of this chapter can be easily adapted to do so, if desired. To estimate a compound phasing, the analyst will have to predetermine *how many* of the subject left turns will be made in the permitted portion of the phase, and *how many* will be made in the protected portion of the phase. Once this is done, timing can be estimated by adapting the methodology of this chapter. Essentially, the permitted and protected portions of the phase are treated as if they were separate phases.

19.6 Sample Signal Timing Problems

The procedures presented in this chapter will be illustrated in a series of signal-timing applications. The following steps should be followed

1. Develop a reasonable phase plan in accordance with the principles discussed herein. Use Equations 19-1 and 19-2, the criteria of Tables 19.1 through 19.3, and any applicable local agency guidelines to make an initial determination of whether left-turn movements need to be protected. Do not include compound phasing in preliminary signal timing; this may be tried as part of a more comprehensive intersection analysis later.

2. Convert all left-turn and right-turn movements to equivalent through vehicle units (tvu) using the equivalents of Tables 19.5 and 19.6, respectively.

3. Draw a ring diagram of the proposed phase plan, inserting lane volumes (in tvu) for each set of movements. Determine the critical path through the signal phasing as well as the sum of the critical-lane volumes (Vc) for the critical path.

4. Determine *yellow* and *all-red* intervals for each signal phase.

5. Determine lost times per cycle using Equations 19-7 and 19-8.

6. Determine the desirable cycle length, *C*, using Equation 19-11. For pre-timed signals, round up to reflect available controller cycle lengths. An appropriate *PHF* and reasonable target v/c ratio should be used.

7. Allocate the available effective green time within the cycle in proportion to the critical lane volumes for each portion of the phase plan.

8. Check pedestrian requirements and adjust signal timing as needed.

Sample Problem 19-7: A Simple Two-Phase Signal

Consider the intersection layout and demand volumes shown in Figure 19.14. It shows the intersection of two streets with one lane in each direction and relatively low turning volumes. Moderate pedestrian activity is present, and the *PHF* and target v/c ratio are specified.

Figure 19.14: Signal Timing Case 19-1

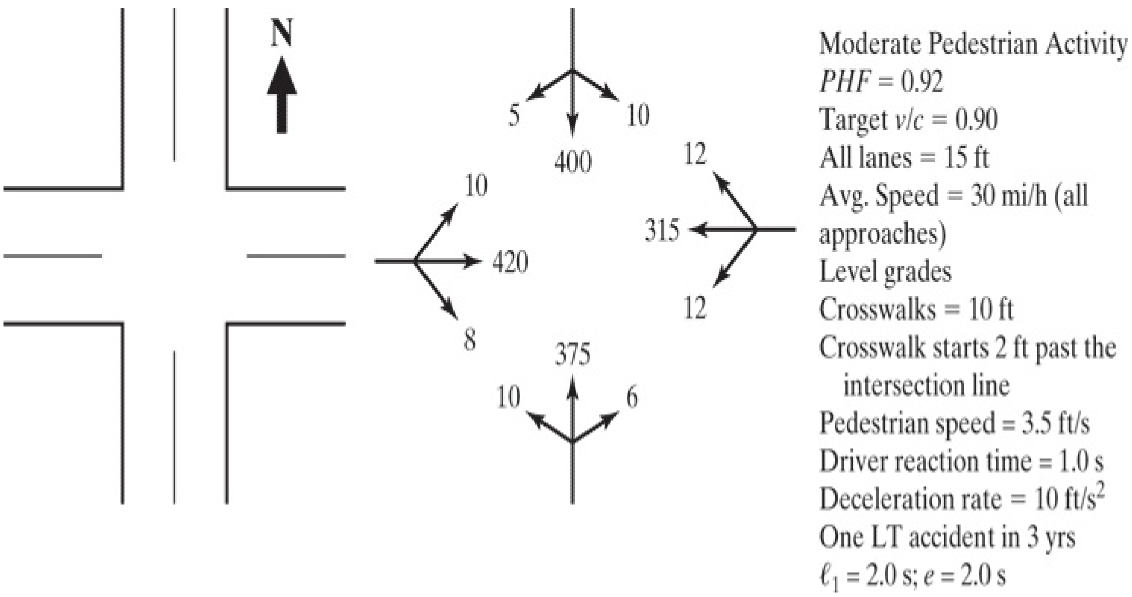


Figure 19.14: Full Alternative Text

Solution

1. Step 1: Develop a Phase Plan Given that there is only one lane for each approach, it is not possible to even consider including protected left turns in the phase plan. However, a check of the criteria of Equations 19-1 and 19-2 should be conducted.

Left-Turn Volumes (Equation 19-1):

$V_{LTEB}=10<200\ \ \text{veh/h}\quad \text{Not Met}V_{LTWB}=12<200\ \ \text{vel}$

Cross Products (Equation 19-2):

$xprod_{EB}=10\times(315/1)=3{,}150<50{,}000\ \text{Not Met}xprod_{WB}=$

None of the criteria of Tables 19.1 through 19.3 are met for any of the left turns at the intersection. Therefore, a simple two-phase signal plan is appropriate for this intersection under the conditions given.

2. Step 2: Convert Volumes to Through-Vehicle Equivalents The conversion of volumes to tvu is illustrated in Table 19.8. Equivalent values are taken from Tables 19.5 and 19.6, and they are interpolated for intermediate values of opposing volume. Note that all through vehicles are equivalent to 1.0 tvu.

Table 19.8: Conversion of Volumes to tvu for Sample Problem 19-1

Approach	Movement	Volume (Veh/h)	Equivalent (T19-5/6)	Volume (tvu/h)	Lane Group Vol (tvu/h)	No. of Lanes	Vol/Lane (tvu/h)
EB	L	10	*3.90*	39			
	T	420	1.00	420	470	1	470
	R	8	1.32	11			
WB	L	12	*5.50*	66			
	T	315	1.00	315	397	1	397
	R	12	1.32	16			
NB	L	10	5.00	50			
	T	375	1.00	375	433	1	433
	R	6	1.32	8			
SB	L	10	*4.70*	47			
	T	400	1.00	400	454	1	454
	R	5	1.32	7			

Note: *Italics* indicate an interpolated value in Table 19.5.

Table 19.8: Full Alternative Text

3. Step 3: Determine Critical-Lane Volumes The critical path through the signal phase plan is illustrated in Figure 19.15. As a two-phase signal, this is a relatively simple determination. For Phase A, either the EB or WB approach is critical. Because the EB approach has the higher lane volume, 470 tvu/h, this is the critical movement for Phase A. For Phase B, either the NB or SB approach is critical; SB has the higher lane volume (454 tvu/h), so this is the critical movement for Phase B. The sum of the critical-lane volumes, therefore, is $470 + 454 = 924$ tuv/h.

Figure 19.15:

Determination of Critical Lane Volumes Sample Problem 19-1

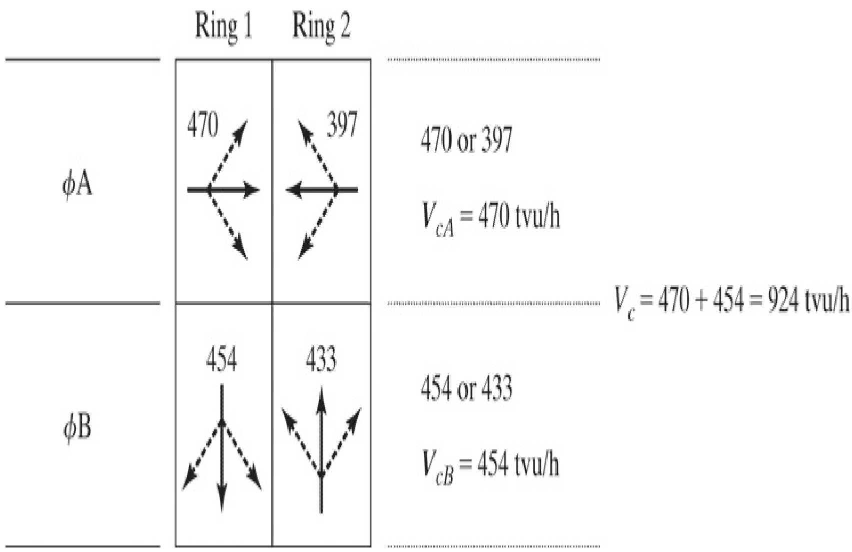


Figure 19.15: Full Alternative Text

4. Step 4: Determine Yellow and All-Red Intervals Yellow and *all-red* intervals are found using Equations 19-3 and 19-5, respectively. The average approach speed for all approaches is 30 mi/h. Thus the S85=30 + 5=35 mi/h, and the S15=30 − 5=25 mi/h. Because moderate numbers of pedestrians are present, the all-red interval will be computed using Equation 19-5, which allows vehicles to clear beyond the far crosswalk line. The distance to be crossed during the *all-red* clearance interval is the sum of two 15-ft lanes and a 10-ft crosswalk plus the 2-ft setback of the crosswalk, or *P = (15 × 2) + 10 + 2 = 42 ft*. Then:

   y=t+1.47×S852×(a+32.2G)yA,B=1.0+1.47×352×(10+32.:

   Because both streets have the same width, crosswalk width, and approach speed, the values of *y* and *ar* are the same for both Phases A and B of the signal.

5. Step 5: Determination of Lost Times Lost times are generally found using Equations 19-7 and 19-8. In this case, the recommended 2.0-second default values for start-up lost time (ℓ1) and extension of effective green into yellow and all-red (*e*) are used. When these two defaults are used, the total lost time is the same as the total yellow and all-red time, and effective green times are equal to actual green times. Thus:

   $L=3.6+1.7+3.6+1.7=10.6\ \text{s}$

   Note that there is one set of lost times for each discrete phase in the signal plan. In this case, there are two phases, two *yellow* intervals, and two *all-red* intervals.

6. Step 6: Determine the Desirable Cycle Length Equation 19-11 is used to determine the desirable cycle length:

   $C_{des}=L1-[\ V_c 1{,}700\times PHF\times(v/c)\ ] C_{des}=10.61-[\ 9241{,}700\times0.92\times0.90\ ]=10.61-0.656=30.8$, SAY 35 s

   As this is a pre-timed controller, a desirable cycle length of 35 or 40 seconds would be used. For the purposes of this signal timing case, the minimum value of 35 seconds will be used.

7. Step 7: Allocate Effective Green to Each Phase Given a 35-second cycle length with 10.6 seconds of lost time per cycle, the amount of effective green time to be allocated is 35.0 − 10.6 = 24.4 s (Equation 19-12). The allocation is done using Equation 19-13:

   $g_i=g_{TOT}\ (V_{ci}V_c) g_A=24.4\ (470924)=12.4\ \text{s} g_B=24.4\ (454$

   The cycle length may be checked as the total of effective green times plus the lost time per cycle, or 12.4 + 12.0 + 10.6 = 35 s. Effective green times may be converted to actual green times using Equation 19-14:

$$G_i = g_i + \ell_1 - e \quad G_A = 12.4 + 2.0 - 2.0 = 12.4\ \text{s} \quad G_B = 12.0 + 2.0 - 2.0 =$$

Again, note that when default values for start-up lost time (2.0 seconds) and extension of effective green into yellow and all-red (2.0 seconds) are used, the actual green time is numerically the same as effective green time.

8. Step 8: Check Pedestrian Requirements Equation 19-15 is used to compute the minimum pedestrian green requirement for each phase. Because both streets have equal width and equal crosswalk widths and because pedestrian traffic is "moderate" in all crosswalks, the requirements will be the same for each phase:

   $$G_p = PW_{min} + PC$$

   From Table 19.6, for "typical" pedestrian volumes and a short cycle length, a value of 7.0 s should be used for PW_{min}. The values of PC are computed using Equation 19-16:

   $$PC = \frac{L}{S_p} = \frac{30}{3.5} = 8.6\ \text{s}$$

   Therefore, each phase must accommodate a *minimum* pedestrian time of 7.0 + 8.6 = 15.6 s. The most restrictive pedestrian policy allows pedestrians to be in the crosswalk only during green intervals. The sufficiency of the vehicular signal timing for pedestrians depends upon which pedestrian option is in place. Table 19.9 illustrates.

Table 19.9: Pedestrian Safety Analysis for Sample

Problem 19-1

Phase	G (s)	y (s)	ar (s)	G_p (Option 1)	G_p (Option 2)	G_p (Option 3)
A	12.4	3.6	1.7	*17.7*	*16.0*	12.4
B	12.0	3.6	1.7	*17.3*	*15.6*	12.0

Note: *Italics* indicate values equal to or exceeding the minimum requirement of 15.6 seconds computed above.

Table 19.9: Full Alternative Text

As is seen from Table 19.9, the signal timing for vehicles is safe for pedestrians in both phases only if Options 1 or 2 are used. Option 1 allows pedestrians in the crosswalk during the *yellow* and *all-red* intervals; Option 2 allows pedestrians in the crosswalk during the *yellow* intervals. Option 3 allows pedestrians in the crosswalk only during green, and the green times in this case are not sufficient to provide for minimum needs. Thus, if pedestrian policy follows Options 1 or 2, the intersection is safe for pedestrians. If Option 3 is the local policy, however, neither phase is sufficient for pedestrian safety.

If Option 3 is the policy, then the green times of *both* Phases A and B need to be increased to a minimum of 15.6 seconds. Phase B, which has the shorter green, would control how much more time is needed. An increased cycle length would be adopted which accommodates an increase in G_B from 12.0 to 15.6 seconds, an increase of 15.6/12.0 = 1.3. Thus, the cycle length would be increased to 35× 1.3 = 45.5 s, which would be increased to 50 seconds for a pre-timed signal. The green time would be reallocated using Equation 19-13 as previously. The total lost time in the cycle remains 10.6 seconds, and the new $g_{TOT}=50.0-10.6=39.4$ s. Then:

$g_i = g_{TOT}\left(\frac{V_{ci}}{V_c}\right)$ $g_A = 39.4\left(\frac{470}{924}\right) = 20.0$ s $g_B = 39.4\ (454$

Both of the new green times are sufficient to accommodate pedestrians completely within the vehicular green. Further, the balance of vehicular greens has been maintained.

The bottom line is that if pedestrian Options 1 or 2 are locally deemed appropriate, the original signal timing with C = 35 s would be implemented. If Option 3 is in place, the larger cycle length of 50 seconds would be implemented.

If pedestrian signals are used (they would not be mandatory in this case), and if pedestrians were restricted to the green intervals (Option 3), the pedestrian clearance interval of 8.6 seconds would end when the green terminates. The actual WALK intervals would then be:

$PW_A = 20.0 - 8.6 = 13.4$ s $PW_B = 19.4 - 8.6 = 12.8$ s

Actual *PW* intervals could be computed for Options 1 and 2 as well, although both would use the 35-second cycle length as a base. For Option 1, the *PC* is indexed to the end of the *all-red* interval; for Option 2, the *PC* would be indexed to the end of the *yellow* interval.

Sample Problem 19-8: Intersection of Major Arterials

Figure 19.16 illustrates the intersection of two 4-lane arterials with significant demand volumes and exclusive left-turn lanes provided on each approach.

Figure 19.16: Sample Problem 19-8

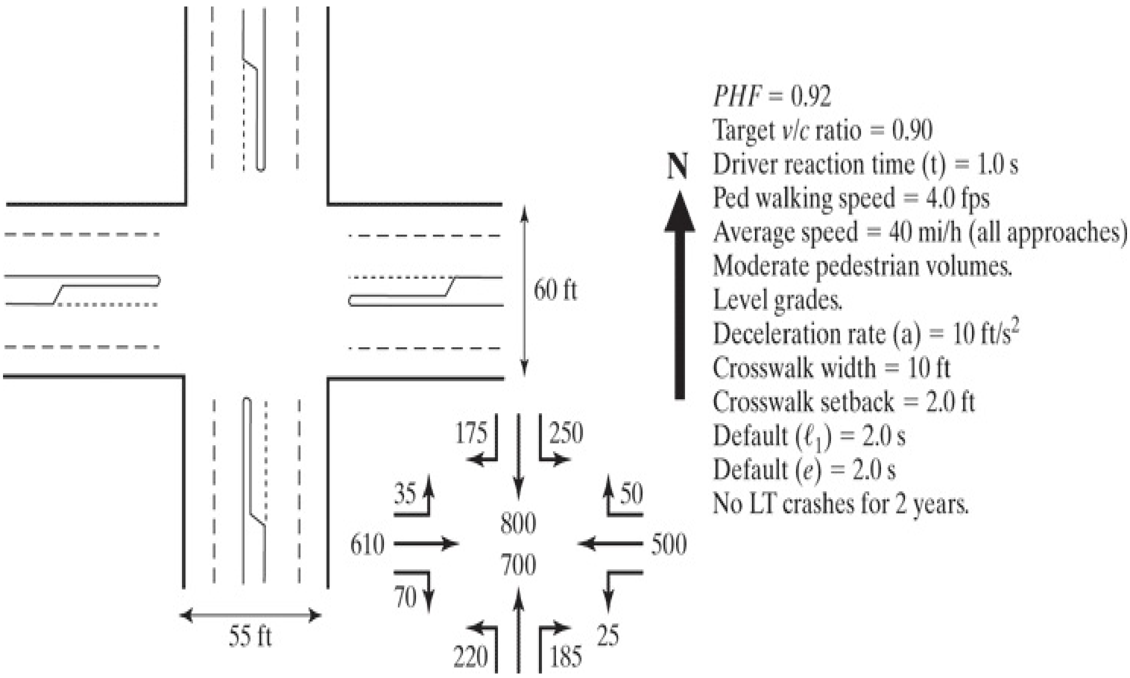


Figure 19.16: Full Alternative Text

1. Step 1: Develop a Phase Plan Each left-turn movement should be checked against the criteria of Equations 16-1 and 16-2 to determine whether or not it needs to be protected. The criteria of Tables 19.1 through 19.3 should be checked for this purpose to determine whether or not it needs to be protected.

EB left turn: $VLT=35<200$ veh/h

$xprod=35\times(500/2)=8{,}750<50{,}000$

No criteria of Tables 19.1 through 19.3 are met.

No protection needed.

WB left turn: $VLT=25<200$ veh/h

$xprod=25\times(601/2)=7{,}625<50{,}000$

No criteria of Tables 19.1 through 19.3 are met.

No protection needed.

NB left turn: $VLT=220>200$ veh/h

Protection needed.

SB left turn: $VLT=250>200$ veh/h

Protection needed.

Given that the NB and SB left turns require a protected phase, the next issue is how to provide it. The two opposing left-turn volumes, 220 veh/h (NB) and 250 veh/h (SB), are not numerically very different. Therefore, there appears to be little reason to separate the NB and SB protected phases. An exclusive left-turn phase will be used on the N–S arterial. A single phase using permitted left turns will be used on the E–W arterial. Unless difficulties arise with the signal timing, a compound phasing for the N–S arterial would not be considered.

2. Step 2: Convert Volumes to Through Vehicle Equivalents Through-vehicle equivalents are obtained from Tables 19.5 and 19.6 for left and right turns, respectively. The computations are illustrated in Table 19.10.

Table 19.10: Conversion of Volumes to tvu for Sample Problem 19-2

Approach	Movement	Volume (Veh/h)	Equivalent (T19-5/6)	Volume (tvu/h)	Lane Group Vol (tvu/h)	No. of Lanes	Vol/Lane (tvu/h)
EB	L	35	*4.00*	140	140	1	140
	T	610	1.00	610	702	2	351
	R	70	1.32	92			
WB	L	25	*5.15*	129	129	1	129
	T	500	1.00	500	566	2	283
	R	50	1.32	66			
NB	L	220	1.05	231	231	1	231
	T	700	1.00	700	944	2	472
	R	185	1.32	244			
SB	L	250	1.05	263	263	1	263
	T	800	1.00	800	1031	2	516
	R	175	1.32	231			

Note: *Italics* indicates a value interpolated in Table 19.5.

Table 19.10: Full Alternative Text

Note that exclusive LT lanes must be established as separate lane groups, with their demand volumes separately computed, as shown in Table 19.10. The equivalent for all protected left turns (Table 19.5) is 1.05.

3. Step 3: Determine Critical Lane Volumes As noted in Step 1, the signal phase plan includes an exclusive LT phase for the N–S artery and a single phase with permitted left turns for the E–W artery. Figure 19.17 illustrates this and the determination of critical-lane volumes.

Figure 19.17: Determination of Critical Lane Volumes—Sample Problem 19-8

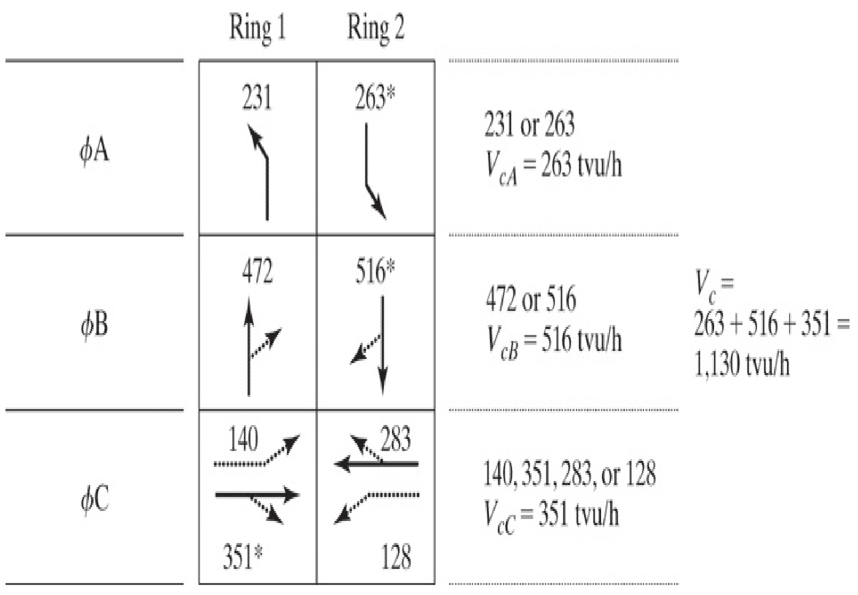


Figure 19.17: Full Alternative Text

Phase A is the exclusive N–S LT phase. The heaviest movement in the phase is 263 tvu/h for the SB left turn. In Phase B, the heavier movement is the SB through and right turn, with 516 tvu/h. In Phase C, both E–W left-turn lane groups and through/right-turn lane groups move at the same time. The heaviest movement is the EB TH/RT lanes, with 351 tvu/h. The sum of critical-lane volumes, Vc therefore, is 263 + 516 + 351=1,130 tvu/h.

Note that each “ring” handles two sets of movements in Phase C. This is possible, of course, because it is the same signal face that controls all movements in a given direction. The left-turn lane volume cannot be averaged with the through/right-turn movement because lane-use restrictions are involved. All left turns must be in the left-turn lane; none may be in the through/right-turn lanes.

4. Step 4: Determine Yellow and All-Red Intervals
   Equation 19-3 is used to determine the length of the *yellow* interval; Equation 19-5 is used to determine the length of the *all-red* interval. With an average speed of 40 mi/h on all approaches, the S_{85} = 40 + 5 = 45 mi/h; As the same average approach speed exists on all approaches, the *yellow* for all approach speeds is set using Equation 19-3:

   y=t+1.47 S852 (a+32.2G)yA,B,C=1.0+1.47×452 (10.0+32

   The *all-red* intervals will reflect the need to clear the full width of the street plus the width of the far crosswalk and the crosswalk setback, as there are “moderate” pedestrian flows (about 200 peds/h/xwalk). The width of the N-S street is 55 ft, and the width of the E–W street is 60 ft. The width of a crosswalk is 10 ft, and the crosswalk setback is 2 ft. During the N–S left-turn phase, it will be assumed that a vehicle must clear the entire width of the E–W artery. Thus, for Phase A, the width to be cleared (*P*) is 60 + 10 + 2 = 72 ft; for Phase B, it is also 72 ft; for Phase C, the distance to be cleared is 55 + 10 + 2 = 67 ft. Thus:

   ar=P+L1.47 S15arA,B=72+201.47×35=1.8 sarC=67+201.

   where 20 ft is the assumed length of a typical vehicle.

5. Step 5: Determination of Lost Times Remembering that where the default values for ℓ_1 and *e* are both 2.0

seconds that the lost time per phase, tL, is the same as the sum of the yellow plus all-red intervals. Because there are three phases, there are three *yellow* intervals and three *all-red* intervals that make up the lost time:

L=(4.3+1.8)+(4.3+1.8)+(4.3+1.7)=18.2 s

6. Step 6: Determine the Desirable Cycle Length The desirable cycle length is found using Equation 19-11:

   Cdes=L1−[Vc1700 PHF (v.c)]Cdes=18.21−[11301700×0.92×0.90]=92.4 s, SAY 100 s

   As this is a pre-timed signal controller, a cycle length of 100 seconds would be selected. Beyond 90 s, pre-timed cycle lengths are used in 10-second increments.

7. Step 7: Allocate Effective Green to Each Phase In a cycle length of 100 seconds, with 18.2 seconds of lost time per cycle, the amount of effective green time that must be allocated to the three phases is 100−18.2=81.8 s. Using Equation 19-13, the effective green time is allocated in proportion to the phase critical-lane volumes:

   gi=gTOT (VciVc)gA=81.8 (2631130)=19.0 sgB=81.8 (51

   Note that 0.1 s was added to g_A to ensure that the total green time allocated (19.0 + 37.4 + 25.4) was equal to 81.8 seconds of total effective green. As phase times are normally rounded to the nearest 0.1 seconds, it is sometimes necessary to add or subtract 0.1 seconds from a phase to account for a cumulative round-off error. Note also that when the default values for ℓ1 and *e* (both 2.0 seconds) are used, actual green times, *G*, equal effective green times, *g*.

8. Step 8: Check Pedestrian Requirements Pedestrian requirements are estimated using Equation 19-15. In this case, note that pedestrians will be permitted to cross the E–W artery only during Phase B. Pedestrians will cross the N–S artery during Phase C.

From Table 19.6, the minimum pedestrian green time for “typical” pedestrian volumes with a longer cycle length is between 7 and 10 seconds. This becomes a judgment call based upon specific site considerations. For this sample solution, a value of 8 seconds. will be used. Then:

$$G_{pi} = PW_{min\ i} + PC_i = PW_{min\ i} + \left(\frac{L_i}{S_p}\right)$$
$$G_{pB} = 8.0 + \left(\frac{60}{4.0}\right) = 23.0\ s$$
$$G_{pC} = 8.0 + \left(\frac{55}{4.0}\right) = 21.8\ s$$

As the actual vehicular green times are 37.4 seconds for Phase B and 25.4 seconds for Phase C, these minimum requirements are met regardless of which pedestrian option is in place.

With a left-turn phase, moderate pedestrian activity, and multilane approaches to be crossed, it is likely that pedestrian signals would be implemented. The actual length of the *PW* intervals, however, would depend upon the pedestrian option in place. Assuming that the most restrictive policy—Option 3—is used, the pedestrian clearance intervals (15 seconds for Phase B and 13.8 seconds for Phase B) *end* when the vehicular green is terminated. Then:

$$PW_B = 37.4 - 15.0 = 22.4\ s$$
$$PW_C = 25.4 - 13.8 = 11.6\ s$$

If Option 1 is in place, the length of the *PW* intervals would be increased by the sum of the *yellow* and *all-red* intervals. If Option 2 is in place, the length of the *PW* intervals would be increased by the length of the *yellow* interval.

Sample Problem 19-9: Another Junction of Major Arterials

Figure 19.18 illustrates another junction of major arterials. In this case, the E–W artery has three through lanes, plus an exclusive LT lane and an exclusive RT lane in each direction. In effect, each movement on the E–W artery has its own lane group. The N–S artery has two lanes in each direction, with no exclusive LT or RT lanes. The number of pedestrians at this intersection is negligible, but occasional pedestrians will be present.

Figure 19.18: Sample Problem 19-9

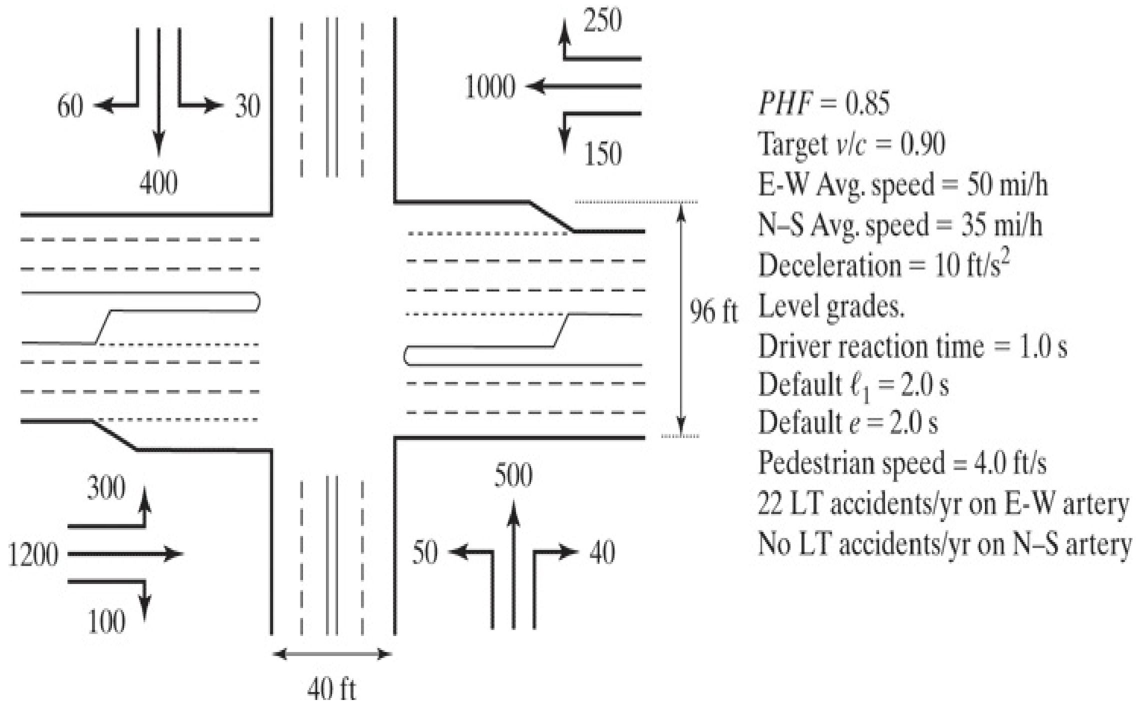


Figure 19.18: Full Alternative Text

1. Step 1: Develop a Phase Plan Phasing is determined by the need for left-turn protection. Using the criteria of Equations 19-1 and 19-2, and the criteria of Tables 19.1 through 19.3, each left turn movement is examined.

    - EB: $V_{LT}=300$ veh/h >200 veh/h

    Protected phase needed.

    - WB: $V_{LT}=150$ veh/h <200 veh/h

      $x_{prod}=150\times(1200/3)=60{,}000>50{,}000$

Protected phase needed.

- NB: $V_{LT}=50 \text{ veh/h} < 200 \text{ veh/h}$

  $\text{xprod} = 50 \times (400/2) = 10{,}000 < 50{,}000$

  No criteria of Tables 19.1 through 19.3 are met.

No protection needed.

- SB: $V_{LT}=30 \text{ veh/h} < 200 \text{ veh/h}$

  $\text{xprod} = 30 \times (500/2) = 7{,}500 < 50{,}000$

  No criteria of Tables 19.1 through 19.3 are met.

No protection needed.

It should be noted that both the EB and WB approaches also meet two additional criteria for protection from Tables 19.1 and 19.2. The 85th percentile speed ($50+5=55$ mi/h) exceeds 45 mi/h, and the number of LT crashes exceeds 11/yr. The latter is sufficient to mandate *full* protection for these turns.

It is fortunate that protected LTs are *not* needed on the N–S artery. Had protected phasing been required for the NB and SB approaches, the lack of an exclusive LT lane on these approaches would have resulted in either the need to create one, or a very inefficient phase plan.

The E–W approaches have LT lanes, and protected left turns are needed on both approaches. Because the LT volumes EB and WB are very different (300 veh/h vs. 150 veh/h), a phase plan that splits the protected LT phases would be advisable. A NEMA phase plan, using an exclusive LT phase followed by a leading green for the EB direction, will be employed for the E–W artery.

2. Step 2: Convert Volumes to Through-Vehicle Equivalents Tables 19.5 and 19.6 are used to find through-vehicle equivalents for left- and right-turn

volumes, respectively. “Negligible” pedestrian activity will be interpreted as “low” pedestrian activity in Table 19.6. Conversion computations are illustrated in Table 19.11. Note that the EB and WB approaches have a separate lane group for each movement, whereas the NB and SB approaches have a single lane group serving all movements from shared lanes.

Table 19.11: Conversion of Volumes to tvu for Sample Problem 19-3

Approach	Movement	Volume (Veh/h)	Equivalent (T19-5/6)	Volume (tvu/h)	Lane Group Vol (tvu/h)	No. of Lanes	Vol/Lane (tvu/h)
EB	L	300	1.05	315	315	1	315
	T	1200	1.00	1200	1200	3	400
	R	100	1.21	121	121	1	121
WB	L	150	1.05	158	158	1	158
	T	1000	1.00	1000	1000	3	334
	R	250	1.21	303	303	1	303
NB	L	50	3.00	150			
	T	500	1.00	500	698	2	349
	R	40	1.21	48			
SB	L	30	*4.00*	120			
	T	400	1.00	400	593	2	297
	R	60	1.21	73			

Italics indicate values interpolated from Table

19.5.

Table 19.11: Full Alternative Text

3. Step 3: Determine Critical-Lane Volumes Figure 19.19 shows a ring diagram for the phase plan discussed in Step 1 and illustrates the selection of the critical-lane volumes.

Figure 19.19: Determination of Critical Lane Volumes—Sample Problem 19-9

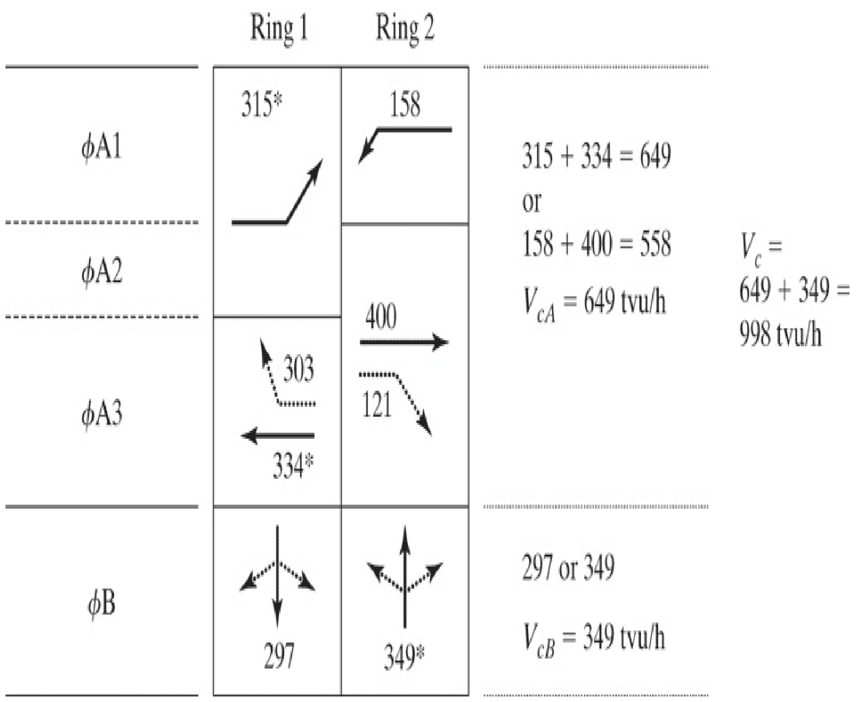


Figure 19.19: Full Alternative Text

The phasing involves overlaps. For the combined Phase A, the critical path is down Ring 1, which has a sum of critical-lane volumes of 649 tvu/h. For Phase B, the choice is simpler because there are no overlapping phases. Ring 2, serving the NB approach, has a critical-lane volume of 349 tvu/h. The sum of all critical-lane volumes (Vc) is 649 + 349=998 tvu/h.

Note also that overlapping phases have a unique characteristic. In this Sample Problem, for overlapping Phase A, the largest left-turn movement is EB and the largest through movement is EB as well. Because of this, the overlapping phase plan will yield a smaller sum of critical lane volumes than one using an exclusive left-turn phase for both left-turn movements. Had the largest left-turn and through movements been from opposing approaches (an infrequent occurrence), the sum of critical-lane volumes would be the same for the overlapping sequence and for a single exclusive LT phase. In other words, little is gained by using overlapping phases where a left turn and its opposing through (through plus right turn) movement are the larger movements.

4. Step 4: Determine Yellow and All-Red Intervals
   Equation 19-3 is used to determine the appropriate length of the *yellow* change intervals. Note that the signal design is a *three*-phase signal and there are three transitions in the cycle. Because of the overlapping sequence, the transition at the end of the protected EB/WB left turns occurs at different times on Ring 1 and Ring 2. For simplicity, it is assumed that left-turning vehicles from the EB and WB approaches cross the entire width of the N–S artery. *All-red* intervals are determined using Equation 19-4 because there are few pedestrians present.

   Percentile speeds are estimated from the measured average approach speeds given:

   S85EW=50+5=55 mi/hS15EW=50−5=45 mi/hS85NS=35

   Then:

   y=t+1.47S852(a+32.2G)yA1,A2,A3=1.0+1.47×552(10+3 where 20 ft is the assumed average length of a typical vehicle.

5. Step 5: Determination of Lost Times Because the problem statement specifies the default values of 2.0 seconds each for start-up lost time and extension of effective green into *yellow* and *all-red* intervals, the total lost time in each phase, tL is equal to the sum of the *yellow* and *all-red* intervals. Note also that the critical path for the signal has three phases. There are three sets of *yellow* and *all-red* times that will be used to compute total lost time, two associated with the E–W street and one associated with the N–S street. Therefore:

   $$L=(y_{A1/A2}+ar_{A1/A2})+(y_{A3}+ar_{A3})+(y_B+ar_B)$$
   $$L=(5.0+0.9)+(5.0+0.9)+(3.9+2.6)=18.3\ \text{s}$$

   Note from Figure 19.19 that the first phase transition occurs at the end of Phase A1, but only on Ring 2. A similar transition occurs at the end of Phase A2, but only on Ring 1. The two other transitions, at the end of Phases A3 and B, occur on both rings.

6. Step 6: Determine the Desirable Cycle Length The desirable cycle length is found using Equation 19-11:

   $$C_{des}=\frac{L}{1-\left[\frac{V_c}{1700\ \text{PHF}\ (v/c)}\right]}$$
   $$C_{des}=\frac{18.3}{1-\left[\frac{998}{1700\times0.85\times0.90}\right]}=\frac{18.3}{1-0.767}=78.5\ \text{s SAY}\ 80\ \text{s}$$

   As this is a pre-timed controller, a cycle length of 80 seconds would be selected.

7. Step 7: Allocate Effective Green to Each Phase A signal cycle of 80.0 seconds with 18.3 seconds of lost time has 80.0 − 18.3 = 61.7 s of effective green time to allocate in accordance with Equation 19-13. Note that in allocating green to the critical path, Phases A1 and A2 are treated as a single segment. Subsequently, the location of the Ring 2 transition between Phases A1 and A2 will have to be established.

gi=gTOT (VciVc)gA1/A2=61.7 (315998)=19.5 sgA3=61

The specific lengths of Phases A1 and A2 are determined by fixing the Ring 2 transition between them. This requires consideration of the noncritical path through combined Phase A, which occurs on Ring 2. The total length of combined Phase A is the sum of gA1/A2+gA3=19.5+20.6=40.1 s. The Ring 2 transition is based on the relative values of the lane volumes for Phase A1 and the combined Phase A2/A3, or:

gA1=40.1 (158158+400)=11.4 s

By implication, Phase A2 is the total length of combined Phase A minus the length of Phase A1 and Phase A3, or:

gA2=40.1−11.4−20.6=8.1 s

Now, the signal has been completely timed for vehicular needs. With the assumption of default values for ℓ1(2.0 s) and *e* (2.0 s), actual green times are equal to effective green times computed previously.

8. Step 8: Pedestrian Requirements Although there are few pedestrians at this intersection, they do exist, and the signal timing must be checked to see if they are safely accommodated.

   From Table 19.6, with negligible pedestrian volumes, the minimum *PW* interval is 4.0 s. The *PC* intervals are dependent upon street widths being crossed. Note that pedestrians cross the N–S street during Phase A3, while pedestrians cross the E–W street during Phase B. Then:

   PC=(LSp)PCA3=(404.0)=10.0 sPCB=(964)=24.0 s

   and:

   Gpi=PWi+PCiGA3=4.0+10.0=14.0 sGB=4.0+24.0 =28.0

   The actual green times are 20.6 s for Phase A3, which exceeds the minimum needed, and 21.6 s for Phase B, which is not sufficient. Because there are few pedestrians, an Option 1 policy would be appropriate. If the *yellow* and *all-red* intervals for Phase B are added to the green, the time for pedestrians to be in the crosswalk is extended to 21.6 + 3.9 + 2.6 = 28.1 s, which would be sufficient, although just barely.

As a comparison, if an Option 3 policy is in effect, the Phase B green would have to be increased to 28.0 seconds. To do this, the cycle length would be increased to 80.0 (28.0/21.6) = 103.7 s, which, in practical terms, calls for a 110-second cycle length. The greens would then be reassigned using the same process illustrated previously, but with the larger cycle length.

Because pedestrians are “negligible,” this might be a case where pedestrian actuators are considered for a pre-timed signal. This would not be done, however, without a clear idea of exactly how many pedestrians are likely to show up during peak periods.

Sample Problem 19-10: A T-Intersection

Figure 19.20 illustrates a typical T-intersection, with exclusive lanes for various movements as shown. Note that there is only one opposed left turn in the WB direction.

Figure 19.20: Sample Problem 19-10

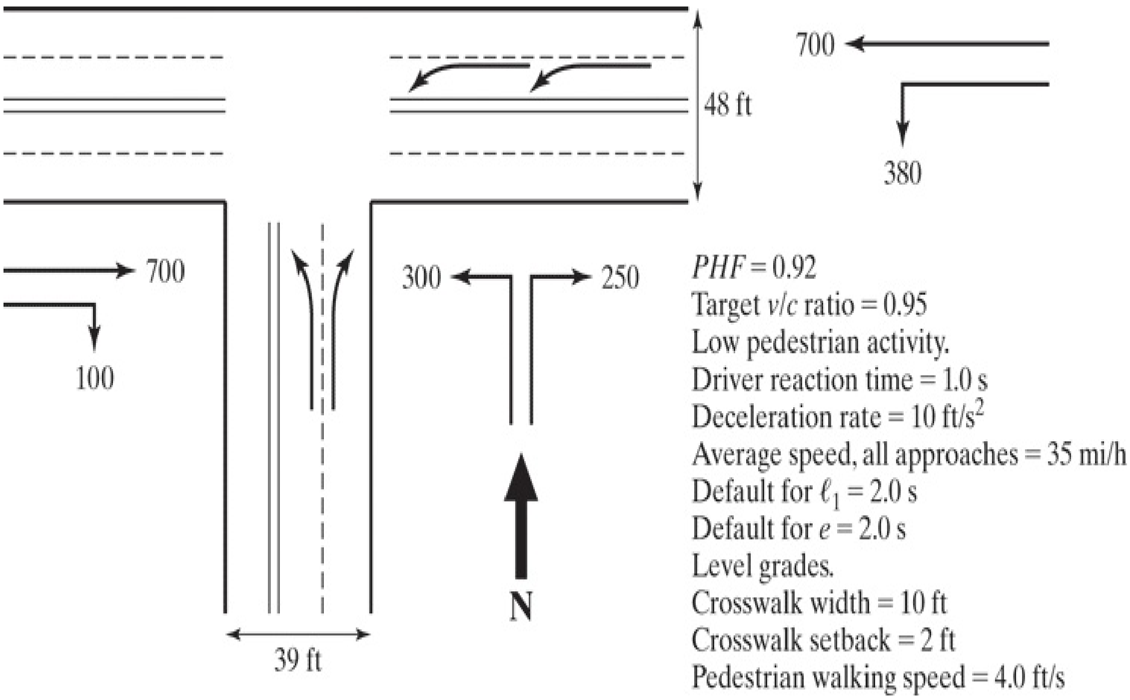


Figure 19.20: Full Alternative Text

1. Step 1: Develop a Phase Plan In this case, there is only one opposed left turn to check for the need of a protected phase. As the WB left turn>200 veh/h, it should be provided with a protected left-turn phase. There is no EB or SB left turn, and the NB left turn is unopposed, due to the geometry. The standard way of providing for the necessary phasing would be to use a leading WB green with no lagging EB green.

2. Step 2: Convert Volumes to Through-Vehicle Equivalents Table 19.12 shows the conversion of volumes to through vehicle equivalents, using the equivalent values given in Tables 19.5 and 19.6 for left and right turns, respectively.

Table 19.12: Conversion of Volumes to tvu for

Sample Problem 19-4

Approach	Movement	Volume (Veh/h)	Equivalent (T19-5/6)	Volume (tvu/h)	Lane Group Vol (tvu/h)	No. of Lanes	Vol/Lane (tvu/h)
EB	T	700	1.00	700	821	2	411
	R	100	1.21	121			
WB	L	380	1.05	399	399	1	399
	T	700	1.00	700	700	1	700
NB	L	300	1.10	330	330	1	330
	R	250	1.21	303	303	1	303

Table 19.12: Full Alternative Text

Note that the NB left turn is *not* strictly a protected turn. There are three possible ways to treat this movement, each of which yields a different equivalent value:

- The WB LT could be considered to be the same as a protected LT, leading to an equivalent of 1.05 in Table 19.5.
- The WB LT could be considered to be a permitted LT with Vo=0 veh/h,, leading to an equivalent of 1.1 in Table 19.5.
- The WB LT could be considered as analogous to a right turn, as the principle “opposing” flow would be the pedestrians in the left crosswalk. The equivalent would be selected from Table 19.6 (with low pedestrian flow), leading to an equivalent of 1.21.

Local agency policy would govern which is used. For illustration purposes, the middle option is used here.

3. Step 3: Determine Critical-Lane Volumes Figure 19.21 shows the ring diagram for the phasing described in Step

1 and illustrates the determination of the sum of critical-lane volumes.

Figure 19.21: Determination of Critical Lane Volumes—Sample Problem 19-10

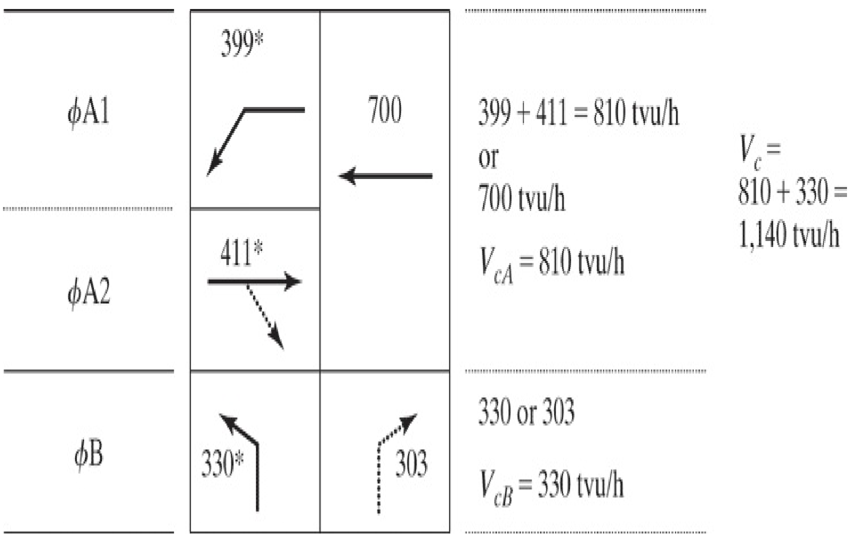


Figure 19.21: Full Alternative Text

In this case, the selection of the critical path through combined Phase A is interesting. Ring 1 goes through two phases; Ring 2 goes through only one. In this case, the critical path goes through Ring 1 and has a total of three phases. Had the Phase A critical path been through Ring 2, the signal would have only two critical phases. In such cases, the highest critical-lane volume total *does not alone determine the critical path*. Because one path has an additional phase and, therefore, an additional set of lost times, it could possibly be critical even if it has the lower total critical-lane volume. In such a case, the cycle length would be computed using *each* path, and the one yielding the largest desirable cycle length would be critical. In this case, the path yielding three phases has the highest sum of critical-lane volumes, so only one cycle length will have to be computed.

4. Step 4: Determine Yellow and All-Red Intervals Both *yellow* and *all-red* intervals for both streets will be computed using Equations 19-3 and 19-4 (low pedestrian activity) and the average speed of 35 mi/h for both streets. For Phases A1 and A2, it will be assumed that both the left-turn and through movements from the E–W street cross the entire 39-ft width of the N–S street. Similarly, in Phase B, it will be assumed that both movements cross the entire 48-ft width of the E–W street. Then:

   $$y = t + \frac{1.47\, S_{85}}{2(a + 32.2G)} \qquad y_{A1,A2,B} = 1.0 + \frac{1.47 \times (35+5)}{2\,(1\ldots}$$

5. Step 5: Determination of Lost Times Once again, 2.0-second default values are used for start-up lost time ($\ell 1$) and extension of effective green into yellow and all-red (e), so that the total lost time for each phase is equal to the sum of the yellow plus all-red intervals:

   $$L = (y_{A1} + ar_{A1}) + (y_{A2} + ar_{A2}) + (y_B + ar_B) \qquad L = (3.9+1.3) + (3.9+1.3) + (3.9+1.5) = 15.8 \text{ s}$$

6. Step 6: Determine the Desirable Cycle Length Equation 19-11 is once again used to determine the desirable cycle length, using the sum of critical-lane volumes, 1,140 tvu/h:

   $$C_{des} = \frac{L}{1 - \left[\dfrac{V_c}{1700 \times PHF \times (v/c)}\right]} \qquad C_{des} = \frac{15.8}{1 - \left[\dfrac{1140}{1700 \times 0.92 \times 0.95}\right]} = \frac{15.8}{1 - 0.767} = 67.8 \text{ s}, \ \text{SAY } 70 \text{ s}$$

7. Step 7: Allocate Effective Green to Each Phase The available effective green time for this signal is 70.0 − 15.8 = 54.2 s. It is allocated in proportion to the critical-lane volumes for each phase:

$g_i = g_{TOT}\left(\frac{V_{ci}}{V_c}\right)$ $g_{A1} = 54.2\left(\frac{399}{1140}\right) = 19.0\ s$ $g_{A2} = 54.2\ ($

Because the usual defaults for ℓ_1 and *e* are used, actual green times are numerically equal to effective green times.

8. Step 8: Check Pedestrian Requirements Although there is low pedestrian activity at this intersection, pedestrians must still be safely accommodated by the signal phasing. It will be assumed that pedestrians cross the N–S street only during Phase A2 and that pedestrians crossing the E–W street will use Phase B.

   From Table 19.6, with "negligible" pedestrian volume, the minimum *PW* interval is 4.0 seconds. The *PC intervals* depend upon the length of the crosswalk used, and the pedestrian walking speed:

   $PC = \frac{L}{S_P}$ $PC_{A2} = \frac{39}{4.0} = 9.8\ s$ $PC_B = \frac{48}{4.0} = 12.0\ s$

   The minimum pedestrian green times are, therefore:

   $G_{pi} = PW_i + PC_i$ $G_{pA2} = 4.0 + 9.8 = 13.8\ s$ $G_{pB} = 4.0 + 12.0 = 16.0$

   The actual vehicular green times are 19.5 seconds for Phase A2 (which is sufficient) and 15.7 seconds for Phase B, which is just short of the minimum pedestrian green. Pedestrian policy once again enters the decision. Under Option 1, both the *yellow* and *all-red* intervals could be used by pedestrians, bringing the amount of time available in Phase B to 15.7 + 3.9 + 1.5 = 21.1 s, which would be sufficient to handle pedestrian needs. Under Option 2, only the *yellow* interval may be added, yielding 15.7 + 3.9 = 19.6 s, which is also sufficient.

   Only under Option 3, where pedestrians are permitted in the crosswalk only during the *green* interval, the time does not meet pedestrian requirements. If this is the policy, then the cycle length would have to be increased by a factor of 16.0/15.7 = 1.02, yielding a cycle length of 70.0 × 1.02 = 71.4 s. For a pre-timed signal, the cycle length would be pushed to 75 seconds, with greens reallocated as previously.

To summarize, for Options 1 and 2, the vehicular signal timing is sufficient for pedestrian needs. For Option 3, the cycle length would be increased to 75 seconds, and the greens reallocated.

References

- 1. *Highway Capacity Manual, 6th Edition: A Guide for Multimodal Mobility Analysis*, Transportation Research Board, Washington, D.C., 2016.
- 2. Kittelson and Associates, *Traffic Signal Timing Manual*, 1st Edition, Federal Highway Administration, Washington, D.C., 2008.
- 3. Pusey, R., and Butzer, G., "Traffic Control Signals," *Traffic Engineering Handbook*, 5th Edition, Institute of Transportation Engineers, Washington, D.C., 2000.
- 4. *Manual of Uniform Traffic Control Devices*, Federal Highway Administration, Washington, D.C., 2009, as updated through 2012.
- 5. McGee, H., and Warren, D., "Right Turn on Red," *Public Roads*, U.S. Department of Transportation, Washington, D.C., June 1976.
- 6. "Driver Behavior at RTOR Locations," ITE Technical Committee 4M-20, *ITE Journal*, Institute of Transportation Engineers, Washington, D.C., April 1992.
- 7. McGee, H., "Accident Experience with Right Turn on Red," *Transportation Research Record 644*, Transportation Research Board, Washington, D.C., 1977.
- 8. "Recommended Practice: Determining Vehicle Change Intervals," ITE Technical Committee 4A-16, *ITE Journal*, Institute of Transportation Engineers, Washington, D.C., May 1985.

Problems

1. 19-1. Signals are being installed and timed for three intersections along a major rural arterial. Some relevant data is available for the arterial at the three intersections, as shown in the table below.

Intersection	Direction	V_{LT} (veh/h)	V_o (veh/h)	No. of Opposing Lanes	Avg Speed of Opposing Traffic	No. of LT Crashes in 3 Yrs	No. of LT Lanes Provided
1	EB	100	1,200	3	50 mi/h	5	1
	WB	120	1,400	3	50 mi/h	8	1
2	EB	250	1,800	3	40 mi/h	8	2
	WB	220	2,000	3	40 mi/h	10	2
3	EB	85	1,000	3	50 mi/h	3	1
	WB	90	1,000	3	50 mi/h	4	1

19.2-13 Full Alternative Text

It may be assumed that sight distances to left-turning vehicles are adequate in all cases. Which movements would likely require protected phasing for LTs, and why?

2. 19-2. What change and clearance intervals are recommended for an intersection with an average approach speed of 35 mi/h, a grade of −2%, a cross-street width of 50 ft, and 10-ft crosswalks with a 2-ft setback from the curb? Assume a standard vehicle length of 20 ft, a driver reaction time of 1.0 seconds, a deceleration rate of 10 ft/s^2, and significant pedestrian movements.

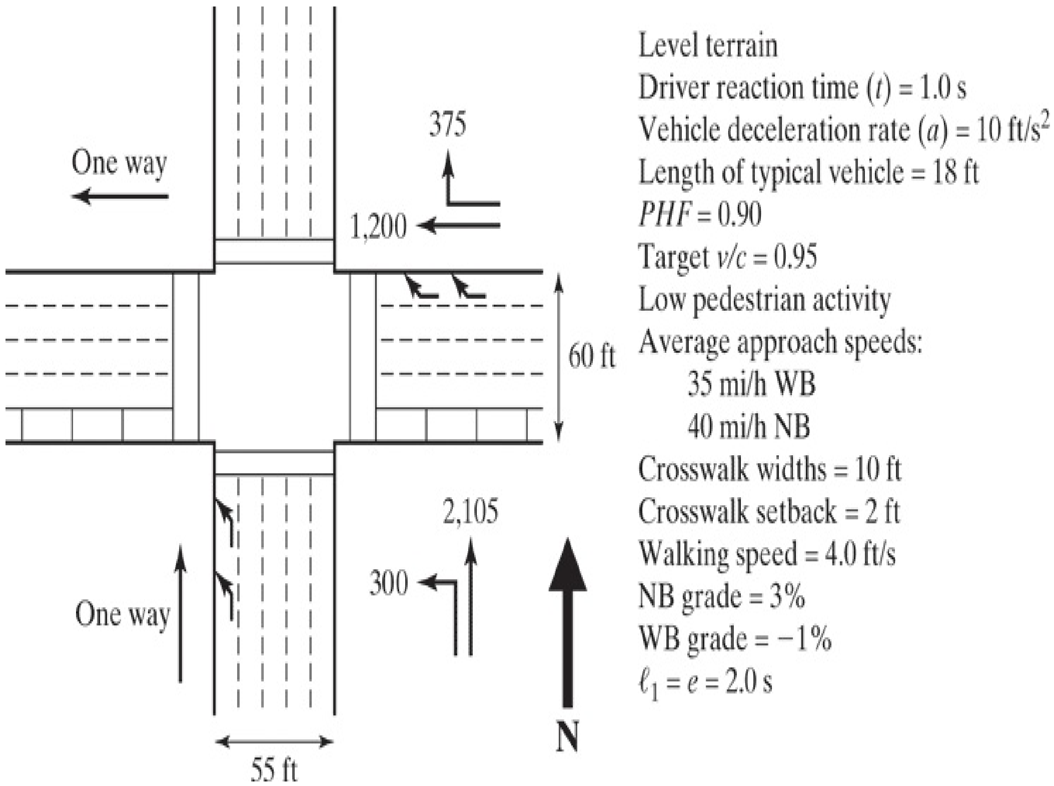
3. 19-3. An analysis of pedestrian needs at a signalized intersection is

undertaken. Important parameters concerning pedestrian needs and the existing vehicular signal timing are given in the table here. Are pedestrians safely accommodated by this signal timing? If not, what signal timing should be implemented? Assume that the standard default values for start-up lost time and extension of effective green into *yellow* and *all-red* (2.0 seconds each) are in effect.

Phase	G (s)	y (s)	ar (s)	G_{pi} (s)
A (local collector)	21.5	3.0	1.5	30.0
B (major arterial)	60.0	3.0	1.0	15.0

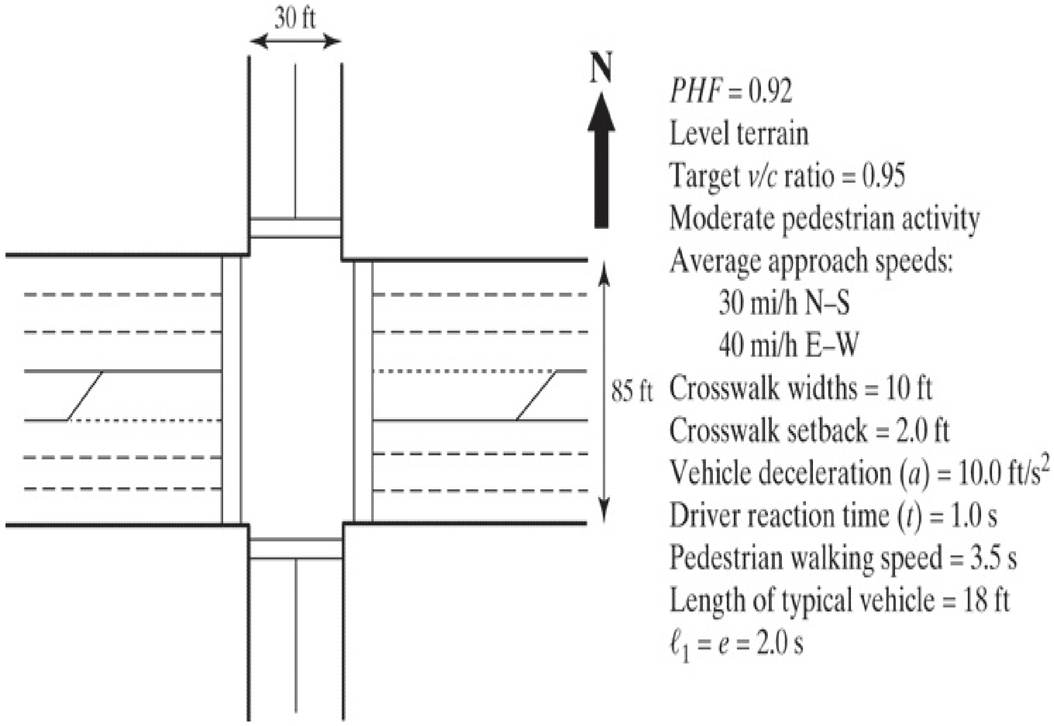
19.2-14 Full Alternative Text

4. 19-4. Develop a signal timing for the intersection shown below.



19.2-15 Full Alternative Text

5. 19-5. Develop a signal timing for the intersection shown below.



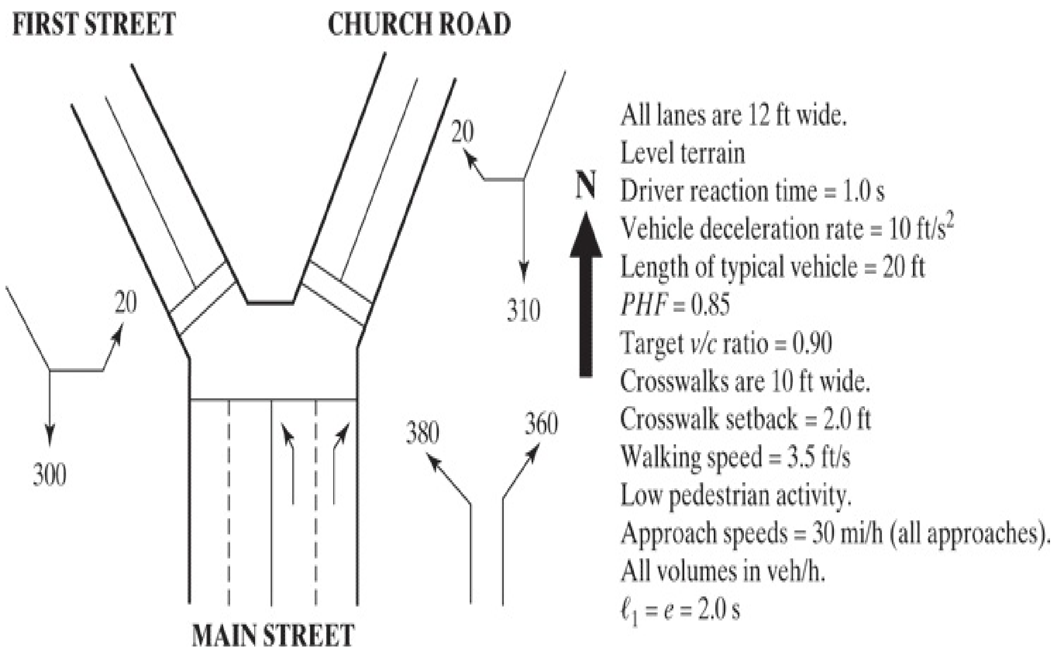
19.2-16 Full Alternative Text

Demand Volumes for Problem 16-4 (veh/h)

Approach	Left	Through	Right
EB	200	800	120
WB	160	1,050	100
NB	10	420	10
SB	12	400	8

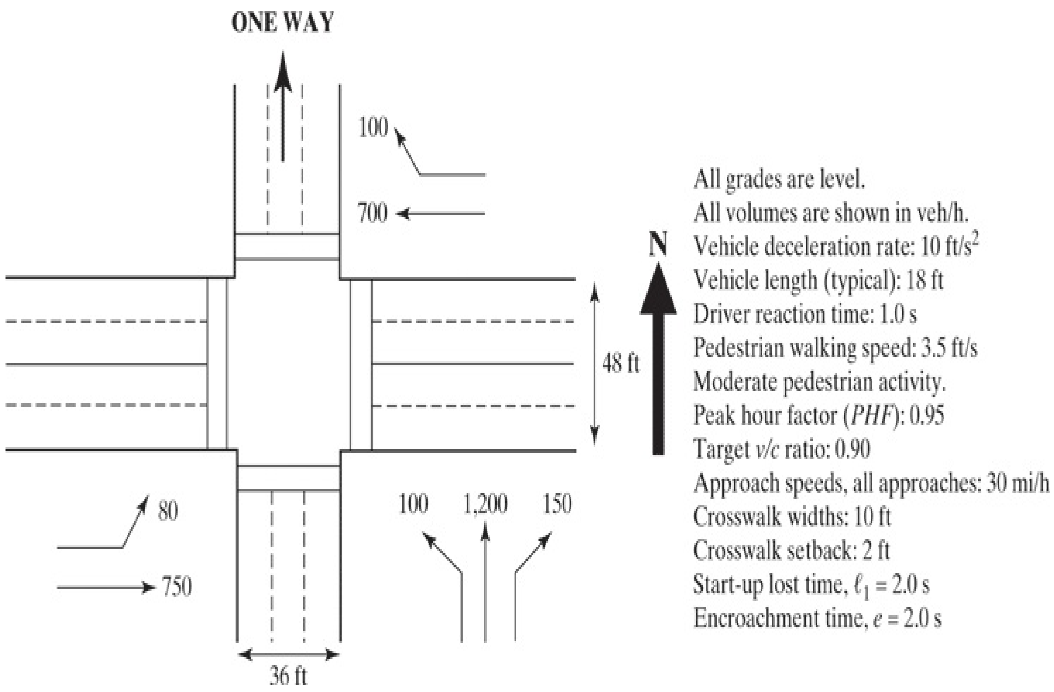
Full Alternative Text

6. 19-6. Develop a signal timing for the intersection shown below:



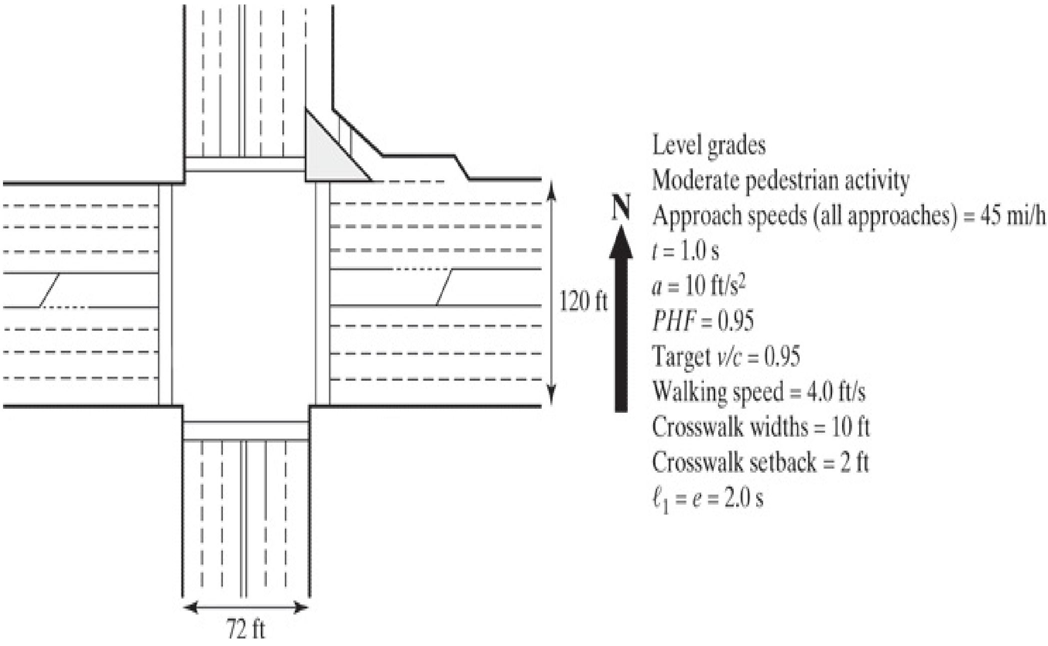
19.2-18 Full Alternative Text

7. 19-7. Develop a signal timing for the intersection shown below.



19.2-19 Full Alternative Text

8. 19-8. Develop a signal timing and design for the major intersection of two busy suburban arterials shown below.



Volumes for Problem 19-8

Approach	LT (veh/h)	TH (veh/h)	RT (veh/h)
EB	500	1600	75
WB	120	1200	350
NB	100	1000	35
SB	90	900	40

Chapter 20 Fundamentals of Signal Timing and Design: Actuated Signals

When pre-timed signal timing is employed, the phase sequence, cycle length, and all interval times are uniform and constant from cycle to cycle. Pre-timed control can provide for several predetermined time periods during which different timings may be applied. During any one period, however, each signal cycle is an exact replica of every other signal cycle.

Actuated control uses information on current demands and operations, obtained from detectors within the intersection, to alter one or more aspects of the signal timing on a cycle-by-cycle basis. Actuated controllers may be programmed to accommodate:

- Variable phase sequences (e.g., optional protected LT phases)
- Variable green times for each phase
- Variable cycle length, caused by variable green times

Such variability allows the signal to allocate green time based on current demands and operations. Pre-timed signals are timed to accommodate average demand flows during a peak 15-minute period. Even within that period, however, demands vary on a cycle-by-cycle basis. Thus, it is, at least conceptually, more efficient to have signal timing vary in the same way.

Consider the situation illustrated in Figure 20.1. Five consecutive cycles are shown, including the capacity and demand during each. Note that over the five cycles shown, the signal has the capacity to discharge 50 vehicles and that total demand during the five cycles is also 50 vehicles. Thus, over the five cycles shown, total demand is equal to total capacity.

Figure 20.1: Effects of a

Variable Demand at a Traffic Signal

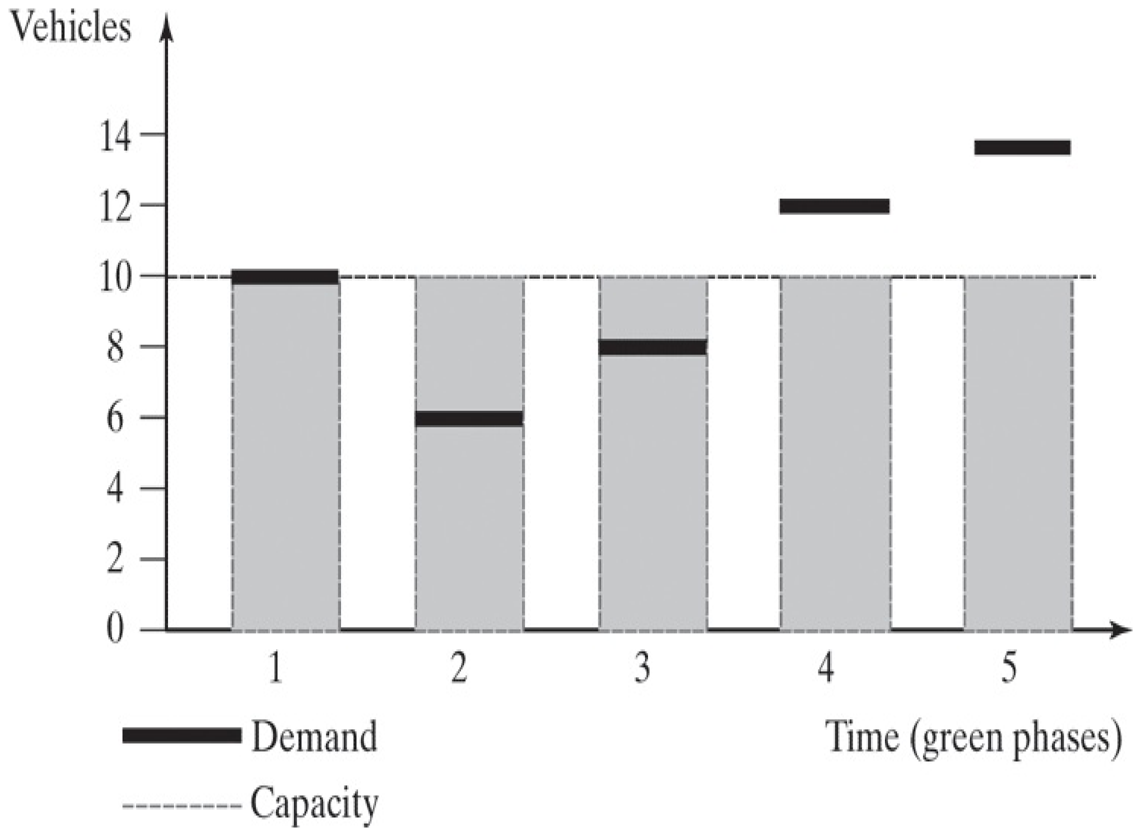


Figure 20.1: Full Alternative Text

Actual operations over the five cycles, however, result in a queue of unserved vehicles with pre-timed operation. In the first cycle, 10 vehicles arrive and 10 vehicles are discharged. In the second, six vehicles arrive and six are discharged. In the third cycle, eight vehicles arrive and eight are discharged. Note that from the second and third cycles, there is unused capacity for an additional six vehicles. In cycle 4, 12 vehicles arrive and only ten are discharged, leaving a queue of two unserved vehicles. In cycle 5, 14 vehicles arrive and only ten are discharged, leaving an additional 4 unserved vehicles. Thus, at the end of the five cycles, there is an unserved queue of six vehicles. This occurs despite the fact that over the entire period, the demand is equal to the capacity.

The difficulty with pre-timed operation is that the unused capacity of six vehicles in cycles 2 and 3 may not be used by excess vehicles arriving in cycles 4 and 5. If the signal had been a properly timed actuated signal, the green in cycles 2 and 3 could have been terminated when no demand was present and additional green time could have been added to cycles 4 and 5 to accommodate a higher number of vehicles. The ability of the signal timing to respond to short-term variations in arrival demand makes the overall signal operation more efficient. Even if the total amount of green time allocated over the five cycles illustrated did not change, the ability to "save" unused green time from cycles 2 and 3 to increase green time in cycles 4 and 5 would significantly reduce delay and avoid or reduce a residual queue of unserved vehicles at the end of the five-cycle period.

Another major benefit of actuated signal timing is that a single programmed timing pattern can flex to handle varying demand periods throughout the day, including peak and off-peak periods and changes in the balance of movements.

If the advantages of allowing signal timing to vary on a cycle-by-cycle basis are significant, why aren't all signalized intersections actuated? The principal issue is coordination of signal systems and cost. To effectively coordinate a network of signals to provide for progressive movement of vehicles through the system, all signals must operate on a uniform cycle length. Thus, where signals must be interconnected for progressive movement, the cycle length cannot be permitted to vary at different intersections.

Actuated signal control is often used at isolated signalized intersections, usually a minimum of 2.0 miles from the nearest adjacent signal. Over the past two decades, however, the use of actuated signal controllers in coordinated signal systems has greatly increased. In such systems, the cycle length must be kept constant, but it can be changed at intervals as short as 15 minutes, and the allocation of green time within the cycle may change on a cycle-by-cycle basis.

Cost remains an issue, as actuated signals involve far more street hardware (detectors and communications), and therefore are significantly more expensive to implement than pre-timed signals.

20.1 Types of Actuated Control

There are two basic types of actuated control:

1. Semiactuated control. This form of control is used where a small side street intersects with a major arterial or collector. This type of control should be considered whenever Warrant 1B is the principal reason justifying signalization. Semiactuated signals are almost always two-phase, with all turns being made on a permitted basis. Detectors are placed only on the side street. The green is on the major street at all times unless a “call” on the side street is noted. The number and duration of side-street greens is limited by the signal timing and can be restricted to times that do not interfere with progressive signal-timing patterns along the collector or arterial.

2. Full-actuated control. In full-actuated operation, all lanes of all approaches are monitored by detectors. The phase sequence, green allocations, and cycle length are all subject to variation. This form of control is effective for both two-phase and multiphase operations and can accommodate optional phases.

Most actuated controllers have additional features that can be implemented if their use is appropriate, such as variable minimum green times, variable passage times, and priority vehicle controls, which are discussed later in this chapter. These options are often referred to as “volume-density” features.

Computer-controlled signal systems do not necessarily constitute actuated control at individual intersections, although actuated control often exists. In such systems, the computer plays the role of a large master controller, establishing and maintaining offsets for progression throughout a network or series of arterials with either pre-timed or actuated signals.

20.2 Detectors and Detection

The hardware for detection of vehicles is advancing rapidly. Pressure-plate detectors, popular in the 1970s and 1980s, are rarely used in modern traffic engineering. Most detectors rely on creating or observing changes in magnetic or electromagnetic fields, which occur when a metallic object (a vehicle) passes through such a field.

In general, there are two types of detection systems used:

- Passage or Point Detection: The passage of a vehicle over a short detector creates a pulse. The number of such pulses is not remembered or stored, but the existence of a pulse indicates that at least one vehicle (in the approach and lane of the detector) requires service.

- Presence or Area Detection: The presence of a vehicle within a detection zone creates a continuous pulse. The pulse begins when the vehicle enters the detection zone and terminates when the vehicle leaves it. Therefore, the detector(s) can discern the number of vehicles stored in the detection area.

Detectors are linked to signal controllers in one of two modes: *locked* or *unlocked*. When "locked," a detector actuation creates a continuous call for service until the green is initiated for the subject lane(s). Passage or point detectors are always linked in the locked mode. When "unlocked," a call for service is initiated when a vehicle enters the detection zone, and is terminated when it leaves the detection zone. Presence or area detectors are almost always linked in the unlocked mode, unless the front end of the detection zone is more than 2 to 3 ft from the STOP line.

Presence or area detection provides a great deal more information to the signal controller, and allows greater flexibility in designing the signalization. It also allows permissive movements, like RTOR, to clear out of the detection zone without leaving an active call for service.

The *Traffic Engineering Handbook* [1] and the *Traffic Signal Timing Manual* [2] contain descriptions of the principal types of detectors that use the technology of magnetic or electromagnetic fields:

- Inductive loop. A loop assembly is installed in the pavement, usually by saw-cutting through the existing pavement. The loop is laid into the saw cut in a variety of shapes, including square, rectangle, trapezoid, or circle. The saw cut is refilled with an epoxy sealant. The loop is connected to a low-grade electrical source, creating an electromagnetic field that is disturbed whenever a metallic object (vehicle) moves across it. This is the most common type of detector in use today.

- Microloop. This is a small cylindrical passive transducer that senses changes in the vertical component of the earth's magnetic field and converts them into electronically discernible signals. The sensor is cylindrical, about 2.5 inches in length and 0.75 inches in diameter. The probe is placed in a hole drilled in the roadway surface.

- Magnetic. These detectors measure changes in the concentration of lines of flux in the earth's magnetic field and convert such changes to an electronically discernible signal. The sensor unit contains a small coil of wire that is placed below the roadway surface.

For all of the magnetic class of detectors, one or more detectors must be used in each lane of each approach. A disadvantage is that all must be placed in or below the pavement. In areas where pavement condition is a serious issue, these detectors could become damaged or inoperable.

Most induction loops measure approximately 6 ft sq, or are circular with a diameter of about 6 ft. Single loops in a lane are used to implement passage or point detection. Presence or area detection may be implemented using a single long loop (approximately 6 ft by 30–40 ft), or multiple loops (of approximately 6 ft in length) in each lane.

Another class of detector uses *sonic* or *ultrasonic* waves that can be emitted from an overhead or elevated roadside location. Such detectors rely on the echoes from reflected waves (ultrasonic) or on the Doppler principle of changes in reflected frequency when waves reflect back from a moving object (sonic). The emitted wave spreads in a cone-like shape and can, therefore, cover more than one lane with a single detector unit, depending upon its exact placement.

There are other detector types in use, such as radar, optical, and even older pressure-plate systems. The vast majority of detectors are of the types described.

A rapidly emerging technology is video imaging, in which real-time video of an intersection approach or other traffic location is combined with computerized pattern-recognition software. Virtual detectors are defined within the video screen, and software is programmed to note changes in pixel intensity at the virtual detector location. Now in common use for data and remote observation, such detection systems have only recently been employed to operate signals in real time.

The timing of an actuated signal is very much influenced by the type of detection in place. The use of presence detectors has greatly increased in recent years, but both point and presence detectors are in common use throughout the United States.

20.3 Actuated Control Features and Operation

Actuated signal controllers are manufactured in accordance with one of two standards. The most common is that of the National Electronic Manufacturer's Association (NEMA). NEMA standards specify all features, functions, and timing intervals, and timing software is provided as a built–in feature of the hardware (often referred to as "firmware"). The second set of standards is for the Type 170 class of controllers, used primarily by the California Department of Transportation and the New York State Department of Transportation. Type 170 controllers do not come with built-in software, which is generally available through third-party vendors. While NEMA software cannot be modified by an agency, Type 170 software can be modified. U.S. manufacturers of signal controllers include Control Technologies, Eagle, Econolite, Kentronics, Naztec, and others. Most manufacturers maintain current web sites, and students are urged to consult them for the most up-to-date descriptions of hardware, software, and functions.

20.3.1 Actuated Controller Features

Regardless of the controller type, virtually all actuated controllers offer the same basic functions, although the methodology for implementing them may vary by type and manufacturer. For each actuated phase, the following basic features must be set on the controller:

1. Minimum Green Time (G_{min}). Each actuated phase has a minimum green time, which serves as the smallest amount of green time that may be allocated to a phase when it is initiated, s.

2. Passage Time (PT). This time actually serves three different purposes: (1) It represents the maximum gap between actuations at a single detector required to retain the green. (2) It is the amount of time

added to the green phase when an additional actuation is received within the unit extension. (3) It must be of sufficient length to allow a vehicle to travel from the detector to the STOP line.

3. Maximum Green Time (G_{max}). Each phase has a maximum green time that limits the length of a green phase, even if there are continued actuations that would normally retain the green. The "maximum green time" begins when there is a "call" (or detector actuation) on a competing phase.

4. Recall Settings. Each actuated phase has a number of recall settings. The recall settings determine what happens to the signal when there is no demand or a system failure.

5. Yellow and All-Red Intervals. Yellow and all-red intervals provide for safe transition from "green" to "red." They are fixed times and are not subject to variation, even in an actuated controller. They are found in the same manner as for pre-timed signals (refer to Chapter 19).

6. Pedestrian WALK ("Walking Man"), Clearance ("Flashing Up-raised Hand"), and DON'T WALK ("Up-raised Hand") intervals. Pedestrian intervals must also be set. With actuated signals, however, the total length of the GREEN is not known. Thus, pedestrian intervals are set in accordance with the minimum green time for each phase. Pedestrian push buttons are often, but not always, needed to ensure adequate crossing times.

Volume-density features add several other functions. They are generally used at intersections with high approach speeds (≥45 mi/h), and in conjunction with presence detectors. In addition to the normal features of any actuated controller, the volume-density features include two important functions:

1. Variable Minimum Green. Because presence detectors are capable of "remembering" the number of queued vehicles, the minimum green time may be varied to reflect the number of queued vehicles that must be served on the next "green" interval.

2. Gap Reduction: Using standard functions, the unit or vehicle extension is a constant value. Volume density features allow the

minimum gap required to retain the green to be reduced over time. Doing this makes it more difficult to retain the green on a particular phase as the phase gets longer. Implementing the gap-reduction feature usually involves identifying four different measures:

1. Initial passage time: PT_1 (s) (maximum value)
2. Final passage time, PT_2 (s) (minimum value)
3. Time into the green that gap reduction begins, t1 (s)
4. Time into the green that gap reduction ends, t2 (s)

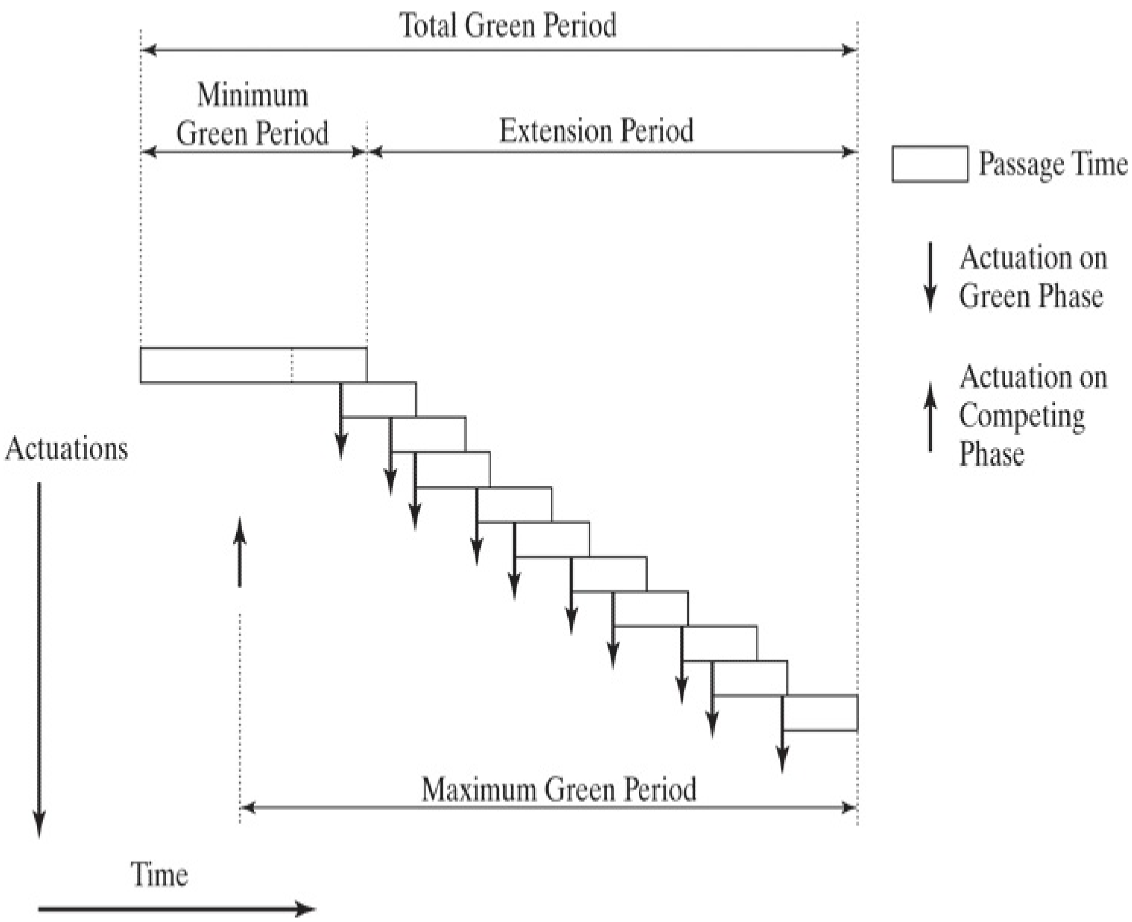
Time t1 begins when a “call” on a competing phase is noted.

Some controllers contain additional features that may be implemented. Those noted here, however, are common to virtually all controllers and controller types.

20.3.2 Actuated Controller Operation

Figure 20.2 illustrates the operation of an actuated phase based on the three critical settings: minimum green, maximum green, and the passage time.

Figure 20.2: Operation of an Actuated Phase



(Source: *Traffic Detector Handbook*, 2nd Edition, JHK & Associates, Tucson, AZ, Institute of Transportation Engineers, pg 66.)

Figure 20.2: Full Alternative Text

When the green is initiated for a phase, it will be *at least* as long as the minimum green period, Gmin. The controller divides the minimum green into an initial portion and a portion equal to one passage time. If an additional “call” is received during the initial portion of the minimum green, no time is added to the phase, as there is sufficient time within the minimum green to cross the STOP line (yellow and all-red intervals take care of clearing the intersection). If a “call” is received during the last *PT* seconds of the minimum green, *PT* seconds of green are added to the phase. Thereafter, every time an additional “call” is received during a unit extension of *PT* seconds, an additional period of *PT* seconds is added to the green.

Note that the additional periods of *PT* seconds are added *from the time of the actuation or “call.”* They are *not* added to the end of the previous unit extension, as this would accumulate unused green times within each unit extension and include them in the total “green” period.

The “green” is terminated in one of two ways:

1. A unit extension of *PT* seconds expires without an additional actuation. Such a termination is commonly referred as a “gap out.”

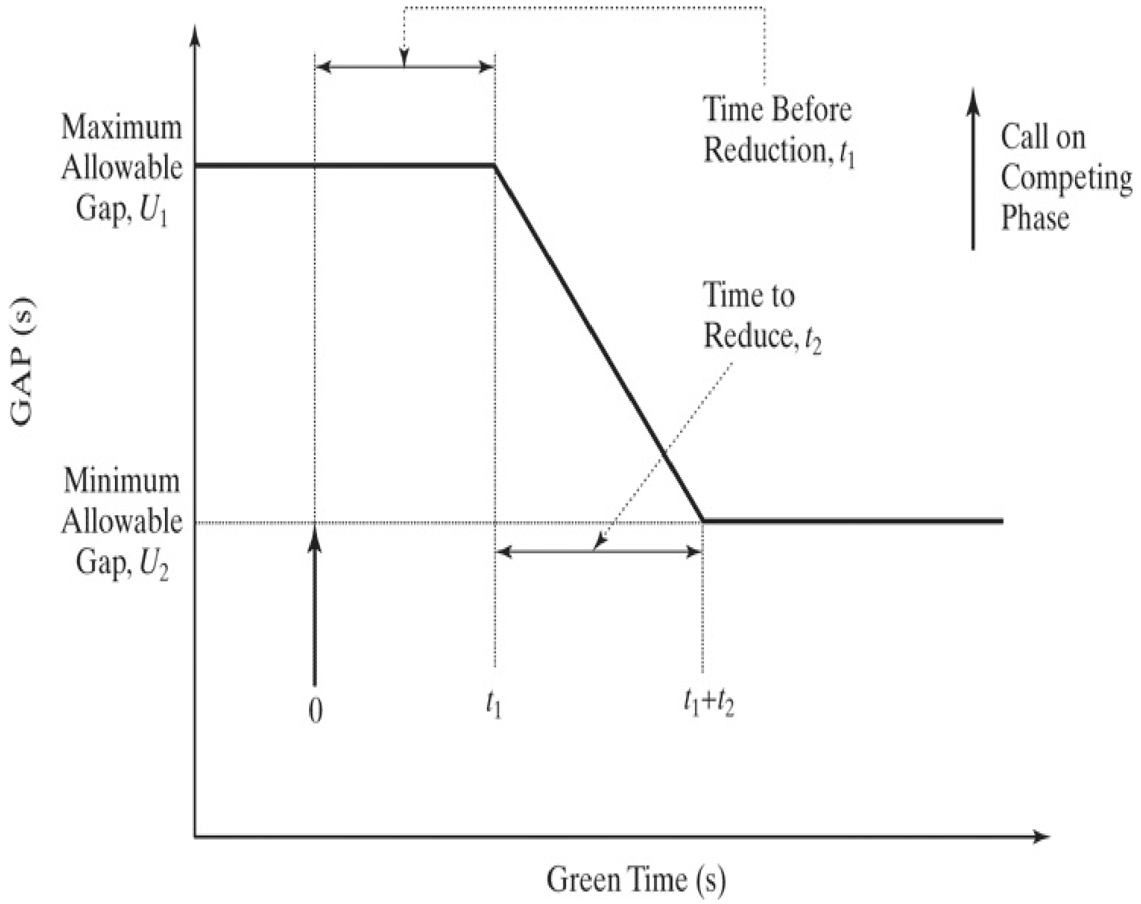
2. The maximum green is reached. Such a termination is referred to as a “max out.” The maximum green begins timing out when a “call” on a competing phase is noted. During the most congested periods of flow, however, it may be assumed that demand exists more or less continuously on all phases. The maximum green, therefore, begins timing out at the beginning of the green period in such a situation.

Assuming that demand exists continuously on all phases, the green period would be limited to a range of Gmin to Gmax. During periods of light flow, with no demand on a competing phase, the length of any green period can be unlimited, depending upon the setting of the recall functions.

In most situations, parallel lanes on an approach operate in parallel with each other. For example, in a three-lane approach, there will be three detectors (one for each lane). If *any* of the three lanes receives an additional “call” within *PT* seconds, the green will be extended. Where multiple detectors are connected in series, using a single lead-in cable, gaps may reflect a lead vehicle crossing one detector and a following vehicle crossing another. While this type of operation is less desirable, it is less expensive to install and is, therefore, used frequently.

Figure 20.3 illustrates the operation of the “gap-reduction” feature on actuated signal controllers. Note the four critical times that must be set on the controller. Depending upon the manufacturer and model selected, there are a number of different protocols for implementing these four times.

Figure 20.3: Gap-Reduction Feature on Volume-Density Controllers



(Source: *Traffic Detector Handbook*, 2nd Edition, JHK & Associates, Tucson AZ, Institute of Transportation Engineers, pg 68.)

Figure 20.3: Full Alternative Text

20.4 Actuated Signal Timing and Design

In an actuated signal design, the traffic engineer does not provide an exact signal timing. Rather, a phase plan is established, and minima and maxima are set, along with programmed rules for determining the green period between limiting values based on vehicle actuations at detectors.

20.4.1 Phase Plans

Phase plans are established using the same types of considerations as for pre-timed signals (see Chapter 19). The primary difference is the flexibility in phase sequencing offered by actuated controllers.

Protected left-turn phases may be installed at lower left-turn flow rates, as these phases may be skipped during any cycle in which no left-turn demand is present. There are no precise guidelines for minimum left-turn demands and/or cross-products for actuated signals, so the engineer has considerable flexibility in determining an optimum phase plan.

20.4.2 Minimum Green Times

Minimum green times must be set for each phase in an actuated signalization, including the nonactuated phase of a semiactuated controller. The minimum green timing on an actuated phase is based on the type and location of detectors.

Point Detection

Point detectors only provide an indication that a “call” has been received on the subject phase. The number of calls experienced and/or serviced is not retained. Thus, if a point detector is located *d* feet from the STOP line, it must be assumed that a queue of vehicles fully occupies the distance *d*. The minimum green time, therefore, must be long enough to clear a queue of vehicles fully occupying the distance *d*, or:

$$G_{min\,i} = \ell_1 + 2.0 \times \text{Int}\left[\frac{d}{25}\right] \quad [20\text{-}1]$$

where:

$G_{min\,i}$ = minimum green time for Phase i, s, ℓ_1 = start-up lost time, s, d = distance between detector and STOP line, ft, 25 = assumed head-to-head spacing between vehicles in queue, ft, and 2 = assumed headway between vehicles, s.

The integer function requires that the value of *d*/25 be rounded to the next highest integer value. In essence, it requires that a vehicle straddling the detector be serviced within the minimum green period. Various agencies will set the value of ℓ_1 based on local policy. Values between 2.0 and 4.0 seconds are most often used. The *Traffic Signal Timing Manual* [2] recommends a value of 3.0 seconds be used. It is also possible to use a value other than 2.0 seconds for headway, although this is the most frequently used value in modern practice.

Presence Detection

Where presence detectors are in use, the minimum green time can be variable, based on the number of vehicles sensed in the queue when the green is initiated. In general:

$$G_{min} = \ell_1 + 2n \quad [20\text{-}2]$$

Where:

ℓ_1 = start-up lost time, s, and n = number of vehicles stored in the detection area, vehs.

This is true, however, only if the front edge of the detector rests on (or very near—within 2 ft) of the STOP line. If the front edge of the detector is further away from the STOP line, the minimum green must assume that the distance between the front edge of the detector and the STOP line is full. Equation 20-1 is used, with *d* equal to the distance between the front edge of the detector and the STOP line, and the minimum green becomes:

$Gmin=\ell 1+2.0 \text{ Int } [d25]+2.0n$ [20-3]

Driver Expectation

Equations 20-1 through 20-3 establish minimum green times based upon the operational requirements of the detector design and location. While it is general practice to make minimum green times as small as possible, experience has shown that drivers have basic expectations for minimum green times in various situations. If minimum greens are unexpectedly short, there could be an increase in rear-end collisions, for example.

The *Traffic Signal Timing Manual* [2] provides guidelines for driver expectation regarding minimum green times. These are shown in Table 20.1.

Table 20.1: Typical values for driver expected minimum green times

Phase Type	Facility Type	Minimum Green to Satisfy Driver Expectancy (s)
Through	Major arterial, speed limit $>$ 40 mi/h	10–15
	Major arterial, speed limit $\leq$ 40 mi/h	7–15
	Minor arterial	4–10
	Collector, local street, driveway	2–10
Left turn	All	2–5

(Source: Kittelson and Associates, *Traffic Signal Timing Handbook*, 1st Edition, Federal Highway Administration, Washington, D.C., June 2008, Table 5-3, pg 5–8.)

Table 20.1: Full Alternative Text

Driver expectations for minimum green times should be considered in all cases, but must be weighed against the potential of unused green time during periods of low flow during off-hours of the day.

20.4.3 Passage Time

As noted previously, the passage time serves three different purposes. In terms of signal operation, it serves as both the minimum allowable gap to retain a green signal and as the amount of green time added when an additional actuation is detected within the minimum allowable gap.

The passage time is selected with three criteria in mind:

- The passage time should be long enough such that a subsequent vehicle operating in dense traffic at a safe headway will be able to retain a green signal (assuming the maximum green has not yet been reached).
- The passage time should not be so long that straggling vehicles may retain the green or that excessive time is added to the green (beyond what one vehicle reasonably requires to cross the STOP line on green).
- The passage time should not be so long that it allows the green to be extended to the maximum on a regular basis.

The passage time has a minimum value based upon the location of the point detector, or the front edge of a presence detector. The passage time must be at least large enough to allow a vehicle traveling at the 15th percentile approach speed to traverse the distance between the detector (or front edge of the detector) to the STOP line, or:

$PT_{min}=\frac{d}{1.47\ S_{15}}$ [20-4]

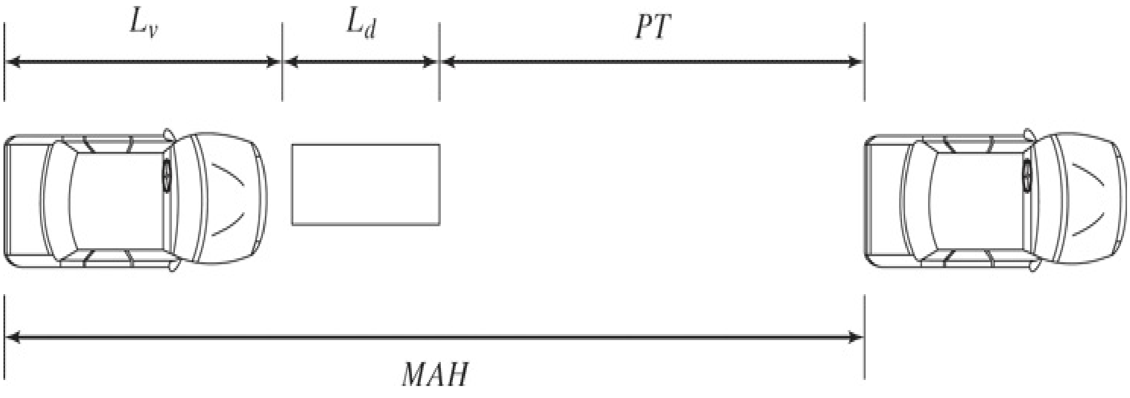
where S_{15} is the 15th percentile approach speed, which may be estimated as the average approach speed minus 5 mi/h.

Presence Detection

Figure 20.4 illustrates the relationship between key variables at a presence detector. The key variable is the *maximum allowable headway* (MAH) or gap that will retain the green for a detector in a single lane. The illustration can be used to derive the following equation:

$$PT = MAH - \frac{L_v + L_d}{1.47\, S_a} \quad [20\text{-}5]$$

Figure 20.4: Relationship Between Passage Time, Detector and Vehicle Length, and Maximum Allowable Gap



(Source: Kittelson and Associates, *Traffic Signal Timing Manual*, 1st Edition, Federal Highway Administration, Washington, D.C., June 2008, Figure 5-4, pg 5–5.)

Figure 20.4: Full Alternative Text

where:

PT = passage time, s, MAH = maximum allowable headway, s, S_a = average approach speed, mi/h, L_v =
length of vehicle, ft (use default value of 20 ft), and L_d =
length of the detection zone, ft.

In Figure 20.4, the passage time, *PT*, is measured as the time that the detector is unoccupied. This is the setting on the controller that will determine whether the green is retained or not. Thus, if the maximum allowable headway is set, the passage time will be found by subtracting the amount of time it takes a vehicle traveling at the average approach speed to traverse both the length of the vehicle and the length of the detector.

Maximum allowable headways used are generally in the range of 2.0 to 4.0 seconds. Larger values tend to result in high delays. General practice is to use MAH=3.0 s, where the gap reduction feature is not in use, and 4.0 (for the maximum *MAH*) when the gap reduction feature is in use.

Point Detection

For point detection, the length of the detector is essentially “0” ft. Because a crossing vehicle registers only a pulse of 0.10 to 0.15 seconds duration, the length of the vehicle is irrelevant. Thus, for point detection, *PT* is equal to the maximum allowable headway, *MAH*.

20.4.4 Detector Location

The minimum green time and the detector location are mathematically linked. Where presence detectors exist, the front edge of the detector is generally within 2 ft of the STOP line, often on it. Equation 20-2 describes the relationship, which produces a variable minimum green time based upon the number of vehicles stored within the detection area.

Where point detectors are used, or where presence detectors are more than 2 ft from the STOP line, the point detector (or front edge of the presence detector) is located to produce a preselected minimum green time.

Because many actuated signals are at locations where demands are quite low during off-peak periods, there is often the desire to keep minimum green times as low as possible, thus minimizing the waiting period for a vehicle on a competing phase when there is no demand on the subject phase. A practical minimum limit on the minimum green time is the assumed start-up lost time, $\ell 1+2.0$ s. This is the amount of time needed to process a single vehicle; it ranges between 4.0 and 6.0 seconds, depending upon the assumed start-up lost time.

When this strategy is used, Equation 20-1 is used to compute the appropriate detector location for the selected minimum green.

Sample Problem 20-1: Estimating Detector Location

A minimum green on an approach to an actuated signal is to be set at 6.0 seconds, with an assumed start-up lost time of 4.0 seconds. How far may the detector be located from the STOP line?

From Equation 20-1:

G min = 6.0=4.0+2.0 Int [d 25] Int [d 25] = 6.0−4.0 2.0 =1.0

Due to the integer function, the detector may be located anywhere between 0.1 and 25.0 ft from the STOP line. Note that where presence detectors are used, the location refers to the *front* of the detector.

There are practical limitations on the placement of detectors that must be observed: the detector(s) must be placed such that no vehicle can arrive at the STOP line without having crossed the detector. In practical terms, this means that no detector can be placed where a vehicle can enter the traffic stream from driveway or curb parking space located between the detector and the STOP line. In many urban and suburban settings, this requires that the detector be located quite close to the STOP line.

Presence detectors are more flexible, in that they can detect vehicles entering the detection area from the side. Thus, it is only the location of the *front* of the area detector that is limited as described previously.

In practical terms, point detectors are most often located close to the STOP line. Longer setbacks require longer minimum green times, which often results in wasted green time during off-peak periods of low flow. Where setbacks longer than 30 to 40 ft are desired, modern practice almost always uses presence detectors to avoid this problem.

20.4.5 Yellow and All-Red Intervals

Yellow and all-red intervals are determined in the same fashion as for pre-timed signals:

y=t+1.47 S852 (a+32.2G) [20-6]
ar=w+L1.47 S15 or P+L1.47 S15 [20-7]

where:

y = yellow interval, s, ar = all red interval, s, S 85 =
85th percentile speed, mi/h, S 15 = 15th percentile speed, mi/h, a =
deceleration rate (10 ft/s 2 , default value), t =
driver reaction time (1.0 s, default value), G =
grade (expressed as a decimal), w = width of street being crossed, ft, and P = distance from near curb to far side of far crosswalk, ft

As in the case of pre-timed signals, yellow and all-red times must be known to determine the total lost time in the cycle, *L*, which is needed to determine maximum green times. The relationships between yellow and all red times and lost times are repeated here for convenience:

L=∑itLitLi=ℓ1i+ℓ2iℓ2i=yi+ari−etLi=ℓ1i+yi+ari−e [20-8]

where:

L = total lost time in the cycle, s/cycle, t Li = total lost time for Phase i, s, ℓ 1i = start-
up lost time for Phase i, s (measured value, or 2.0 seconds default value), ℓ 2i = clearance lost time for Phase i, s, e =
encroachment of effective green into yellow and all-
red periods for Phase i, s (measured value, or 2.0 seconds default value), y

i = yellow interval for Phase i, s, and a r i = all red interval for Phase i, s

Note that when the default values for ℓ1 and *e* are used, the total lost time per cycle, *L*, is equal to the sum of the yellow and all-red phases associated with critical movements in the cycle, and that effective green, *g*, is equal to actual green, *G*.

20.4.6 Maximum Green Times and the Critical Cycle

The “critical cycle” for a full-actuated signal is one in which each phase reaches its maximum green time. For semi-actuated signals, the “critical cycle” involves the maximum green time for the side street and the minimum green time for the major street, which has no detectors and no maximum green time.

Maximum green times for actuated phases and/or the minimum green time for the major street with semiactuated signalization are found by determining a cycle length and initial green split based on average demands during the peak analysis period. The method is the same as that used for determining cycle lengths and green times for a pre-timed signal:

Ci=L1−[Vc1700×PHF×(v/c)] [20-9]

where:

C i = initial cycle length, s, V c = sum of critical lane volumes, veh/h, PHF = peak hour factor, and v/c = desired v/c ratio to be achieved

Because the objective in actuated signalization is to have little unused green time during peak periods, the v/c ratio chosen in this determination is taken to be 0.95 or higher in most applications.

Knowing the cycle length, green times are then determined as:

gi=(C−L)×(VciVc) [20-10]

where:

g_i = effective green time for Phase i, s, and V_{ci} =
critical lane volume for Phase i, veh/h

All other variables are as previously defined.

Note that for actuated signals, the initial cycle length, C_i, is *not* rounded to the next highest 5-second increment (or 10-second increment), as was done for pre-timed signals. The computed cycle length, to the nearest 0.1 second, is used directly.

These computations result in a cycle length and green times that would accommodate the average cycle demands in the peak 15 minutes of the analysis hour. They are not, however, sufficient to handle perturbations occurring during the peak 15-minute demand period when individual cycle demands exceed the capacity of the cycle. Thus, to provide enough flexibility in the controller to adequately service peak cycle-by-cycle demands during the analysis period, green times determined from Equation 20-10 are multiplied by a factor of between 1.25 and 1.50. The results would then become the maximum green times for each phase and/or the minimum green time for a major street at a semiactuated signal.

The "critical cycle length" is then equal to the sum of the actual maximum green times (and/or the minimum green time for a major street at a semiactuated location) plus yellow and all-red transitions.

$$C_c=\sum_N(G_i+y_i+ar_i) \quad [20\text{-}11]$$

where:

C_c = critical cycle length, s, and G_i =
actual maximum green time for actuated Phase i, or actual minimum green

All other terms as previously defined.

The timing of an actuated signal involves a number of practical considerations that may override the results of the computations as described. Particularly at a semiactuated signal location with low side-street demands, the maximum green, Gmax, may compute to a value that is less than the minimum green, Gmin. Although a rarer occurrence, this could happen on a given phase at full-actuated location as well, particularly where protected left-turn phases are involved. In such cases, the Gmax is judgmentally set as Gmin+nU, where *n* is the maximum number of vehicles to be served during a single green phase and *U* is the unit extension. The value of *n* is usually taken to be two to four vehicles per cycle. Then, to maintain an appropriate balance between all phases, values of Gmax for other phases must then be adjusted to maintain a ratio equal to the balance of critical-lane volumes for each phase.

20.4.7 Pedestrian Requirements for Actuated Signals

As for pre-timed signals, pedestrians require the following amount of time to safely cross a street:

G pi = P W min i +P C i P C i = L S p [20-12]

where:

G pi = minimum green time required for pedestrians in Phase i, s, P W min i =
minimum pedestrian WALK interval for Phase i, s (Table 19.6, Ch 19), P C i = pedestrian clearance interval for Phase i, s, L =
length of the crosswalk, ft, and S p = walking speed of pedestrians, ft/s

The safety of pedestrians must be checked, and is dependent upon which of the following policies is in effect:

- Option 1: Pedestrians may be in the crosswalk during *G*, *y*, and *ar* intervals.
- Option 2: Pedestrians may be in the crosswalk during *G*, and *y* intervals.
- Option 3: Pedestrians may be in the crosswalk during *G* intervals only.

Pedestrians are safe when Equation 20-13 is satisfied:

Gpi(Option 1)≤Gmin i+yi+ariGpi (Option 2)≤Gmin i+yiGpi (Option 3)≤Gm
[20-13]

where:

G pi = minimum time required for pedestrian safety for Phase i, s, and G min i = minimum vehicular green, Phase i, s

All other variables as previously defined.

Note that with actuated signals, pedestrian requirements must be compared to the *minimum* vehicular green time, as it cannot be assured that any green phase will exceed this value.

With pre-timed signals, when safe crossing is not assured, either the cycle length must be increased to accommodate both pedestrians and vehicles or pedestrian-actuated push-button and pedestrian signals must be installed. To avoid upsetting the balance of vehicular greens, the most common practice is to increase the cycle length.

With actuated signals, safe crossing based on minimum green times is most often *not* provided. Increasing minimum greens to accommodate pedestrians during every cycle is not an option, as that would create inefficiencies for vehicles that the actuated signal was installed to avoid. Thus, whenever the minimum green *does not* provide safe crossing, a pedestrian push button is almost always installed, and pedestrian signals must be used.

In such cases, the pedestrian signal rests on a DON'T WALK indication. When the pedestrian push button is actuated, *on the next green phase* the minimum green time is increased to satisfy the requirements of Equation 20-13. Pedestrians, unfortunately, often do not understand how pedestrian signals work. They often anticipate that they will get a WALK indication immediately after pushing the pedestrian button. This is *not* the case, as the expanded pedestrian time is not provided until the *next* cycle. For this reason, actuated signals in intense pedestrian environments with heavy pedestrian crossing flows are often avoided where possible.

20.4.8 Dual Entry Feature

The dual entry feature, when engaged, calls phases that can begin concurrently, even when only one of the phases is called. For example, if the NB and SB through phases generally run concurrently, then a call for service on either one would initiate both phases. When the feature is not engaged, only the phase that is “called” will run. Common practice is to engage this feature for pairs of phases that would commonly be expected to run simultaneously.

It is usually *not* engaged for left turn movements, where the option to provide a protected phase in only one direction (or in neither direction) is common.

20.4.9 Simultaneous Force-Off Feature

This feature is engaged when concurrent movements on opposing approaches *must* end at the same time. Thus, when transferring the green to the other street, all greens on the conflicting street must terminate at the same time. For opposing protected left-turn movements, it is possible to develop a phase plan in which one left turn is terminated before the other. In such cases, this feature is *not* engaged.

20.4.10 Recall Features

Recall settings determine what the signal controller will do when there is no demand present on one or more approach lane groups. A recall setting can be “off” or “on” for each phase; when “on,” there are several different recall options:

- Minimum Recall: When engaged, the minimum recall feature places a “call” on the designated phase, even if there is no demand present. If there is no demand anywhere in the intersection, this would force the signal to initiate a green on the designated phase for a duration of at least the minimum green. Common practice is to engage the minimum recall feature on major street through movements.

- Maximum Recall: When engaged, the feature causes a continuous “call” on the designated phase, even if there is no demand present. It forces every designated green phase to extend to its maximum green. This feature is not often used, but is appropriate if vehicle detection is out of service, or if a pre-timed operation is desired.

- Pedestrian Recall: When engaged, the controller places a continuous call for service on the designated phase. It forces the minimum pedestrian green to be implemented during every green phase. It is most often used where pedestrian actuation buttons are out of service, or during time periods when there are large numbers of pedestrians present.

- Soft Recall: When engaged, the controller places a “call” on the designated phase when there are no competing calls present. This is most commonly used for major street through phases during light demand periods, to ensure that the signal rests on the main through movements, particularly at noncoordinated signals.

The minimum recall is the most frequently used recall setting. If all phases were set to minimum recall when there was no demand present, the signal would cycle through its phases using minimum green times. If all recall features were “off” during such a period, the green would stay on the last phase to have received a call, and would not move until a call on another phase was received. In the common usage, the recall would be set only on the major street through movement, guaranteeing that the green would reside on the major street through phases in the absence of any demand.

20.5 Sample Problems in Actuated Signal Design and Timing

Timing of an actuated signal is less definitive than for pre-timed signals and calls for more judgment to be exercised by the engineer. In any given instance, it is possible that several different signal timings and designs would work acceptably. Some of the considerations involved are best illustrated in sample problems.

Sample Problem 20-2: A Semiactuated Signal Timing

Figure 20.5 shows an intersection that will be signalized using a semiactuated controller. For convenience, the demand volumes shown have already been converted to through vehicle units (tvu). This conversion is the same as for pre-timed signals (see Chapter 19).

Figure 20.5: Intersection for Sample Problem 20-2: A Semiactuated Signal

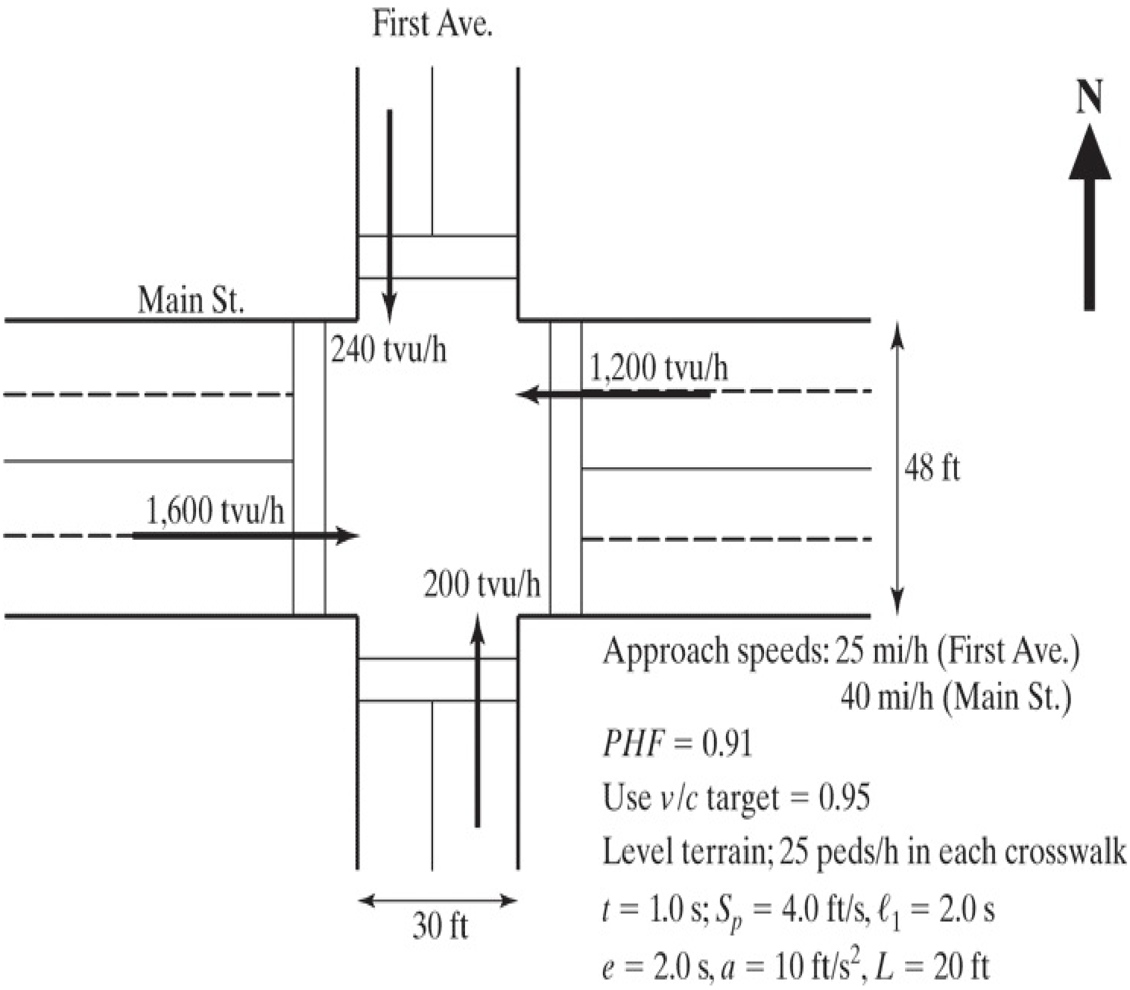


Figure 20.5: Full Alternative Text

1. Step 1: Phasing

   As this is a semiactuated signal, there are only two phases, as follows:

   - Phase A—All First Avenue movements (minor street)

- Phase B—All Main Street movements (major street)

2. Step 2: Minimum Green Time and Detector Location

   For a semiactuated signal, only the side-street phase is actuated and only side-street approaches have detectors. Point detectors are virtually always used in such cases, and are used here. For semiactuated signals, the objective is generally to provide only the amount of green time necessary to clear side-street vehicles, with as little unused green time as possible. Therefore, the minimum green time for First Avenue should be as low as possible. Using a start-up lost time of 4.0 seconds, the minimum green time that could be allocated would be 6.0 seconds. If G_{min} is set at 6.0 seconds, then the detector placement is determined by solving for *d* in Equation 20-1:

   $$G_{min} = 6.0 = 4.0 + 2.0\,\mathrm{Int}\left(\frac{d}{25}\right) \quad \mathrm{Int}\left(\frac{d}{25}\right) = \frac{6.0-4.0}{2.0} = 1.0$$

   The detector would be placed anywhere between 0.1 and 25.0 ft from the STOP line. It must be placed such that no vehicle can enter the approach without traversing the detector.

3. Step 3: Passage Time

   For point detectors, the passage time is equal to be the maximum allowable headway (*MAH*). The recommended value is 3.0 seconds. This must be greater than the passage time from the detector to the STOP line, assuming the maximum setback of 25.0 ft, or:

   $$PT_{min} = \frac{d}{1.47 \times S_{15}} = \frac{25}{1.47 \times (25-5)} = 0.85\ \mathrm{s} < 3.0\ \mathrm{s\ OK}$$

   The 3.0-second unit extension is safe and will be implemented.

4. Step 4: Sum of Critical-Lane Volumes

All of the demand volumes of Figure 20.5 have already been converted into through vehicle equivalents (tvu). The SB movement has a higher volume than the NB movement (both approaches have one lane). Thus, the critical-lane volume for Phase A is 240 tvu/h/ln. The EB volume of 1,600 tvu/h is critical for Phase 2, but is divided into two lanes. Thus, the critical-lane volume for Phase B is 1600/2=800 tvu/h/ln. The sum of the critical-lane volumes, Vc, is 240+800=1,040 tvu/h/ln.

5. Step 5: Yellow and All-Red Times

   To determine other signal timing parameters, an initial cycle length must be selected. This requires, however, that all lost times within the cycle be known, which requires that the *yellow* and *all-red* intervals be established. *Yellow* intervals for each phase are estimated using Equation 20-6, while all-red intervals are estimated using Equation 20-7. Average approach speeds are given for Main Street and First Avenue. The 85th percentile speed may be estimated as 5 mi/h more than the average, while the 15th percentile speed is estimated as 5 mi/h less than the average.

   y = t+ 1.47 S 85 2(a+32.2G) y A = 1.0+ 1.47(25+5) 2(10+32.2×0) =3.2 s y B = 1.0+ 1.47(40+5) 2(10+32.2×0) =4.3 s ar = w+L 1.47 S 15 a r A = 48+20 1.47(25−5) =2.3 s a r B = 30+20 1.47(40−5) =1.0 s

   Note that *all-red* times use *w* rather than *P*, as the number of pedestrians (25/h) is relatively small.

6. Step 6: Lost Time Per Cycle

   With default values of 2.0 seconds used for both ℓ_1 and *e*, the lost time per cycle is equal to the sum of the yellow and all-red times in the cycle, or:

   L=3.2+2.3+4.3+1.0=10.8 s/cycle

7. Step 7: Maximum Green (Phase A) and Minimum Green

(Phase B)

As a semi-actuated signal, the critical cycle is composed of the maximum green for the side street (First Avenue, Phase A), the minimum green for the major street (Main Street, Phase B), and the *yellow* and *all-red* intervals from each. The initial cycle length is estimated using Equation 20-9:

$$C_i = \frac{L}{1-\left[\frac{V_c}{1700\ PHF\ (v/c)}\right]} = \frac{10.8}{1-\left[\frac{1040}{1700\times0.92\times0.9\ldots}\right.}$$

For a semiactuated signal, this value does not have to be rounded. Green splits based on this cycle length are determined using Equation 20-10:

$$g_i = (C-L)\times\left(\frac{V_{ci}}{V_c}\right) \quad g_A = (36.0-10.8)\times\left(\frac{240}{1040}\right)=5.8\text{ s} \quad g_B = (36.0-10.8)\times\left(\frac{800}{1040}\right)=19.4\text{ s}$$

Effective green times and actual green times are equal, given the default values of 2.0 seconds for both ℓ_1 and *e*. Standard practice establishes the maximum green for the minor street and the minimum green for the major street as 1.50 times the above values, or:

$$G_{maxA} = 5.8\times1.5=8.7\text{ s} \quad G_{minB} = 19.4\times1.5=29.1\text{ s}$$

The G_{maxA} of 8.7 seconds is larger than the G_{minA} of 6.0 seconds established earlier, although not greatly so. At 8.7 seconds, a maximum of two, perhaps three vehicles could pass through a single side-street green. If this flexibility is not considered to be sufficient, a larger G_{maxA} would be set, and the G_{minB} adjusted to maintain the same ratio of green times as in the original timing.

8. Step 8: Pedestrian Requirements

Pedestrians cross the minor street during Phase B, and the major street during Phase A. The pedestrian crossing requirement must be compared to the minimum green (Option 3), the minimum green plus *yellow* (Option 2), or the minimum green plus *yellow* and *all-red* intervals

(Option 1). Then:

G pi = P W min i +P C i P C i = L S p

From Table 19.6 (Chapter 19), the minimum pedestrian WALK interval is 4.0 seconds, where the pedestrian volume of 25/h may be considered to be “negligible.”

Then:

G pA = 4.0+ 48 4 =16.0 s G pB = 4.0+ 30 4 =11.5 s

For Phase B, pedestrians will be safe during every cycle, as the G_{minB} of 29.1 seconds is larger than the minimum need of 11.5 seconds, without using any *yellow* or *all-red* time. For Phase A, however, pedestrians need 16.0 seconds, but G_{minA} is only 6.0 seconds. Even with the *yellow* and *all-red* times added (Option 3), this only provides 6.0+3.2+2.3=11.5 seconds.

Because this is a semiactuated signal, increasing the minimum green on every phase is not practical. A pedestrian push-button on the minor street is required in any event, as the green is always on the major street when there are no vehicles “calling” for service on the minor street. A pedestrian arriving when no vehicles are present will require a push-button regardless of whether or not the minimum green times are adequate. In this case, they are not, and the push-button not only provides a “call” for service on the minor street, but will implement (for that cycle), a minimum green, which depends upon the pedestrian policy in place.

If Option 1 is in place, the minimum green on pedestrian actuation would be 16.0−3.2−2.3=10.5 seconds. If Option 2 is used, only *yellow* time may be deducted, and the minimum vehicular green on pedestrian actuation would be 16.0−3.2=12.8 seconds. If Option 3 is in place, pedestrians may use *only* the green time, and 16.0 seconds would be allocated upon a pedestrian “call”.

As the minimum green provides more than enough time for safe crossings of the minor street during Phase B, no pedestrian push button is needed and pedestrian signals are optional. Push-buttons and pedestrian signals are required, however, for major street crossings during Phase A.

9. Step 9: Dual Entry, Simultaneous Force Off, and Recall

   Because of the two-phase plan, opposing approaches must both get the green and lose the green at the same time. Thus, the dual entry and simultaneous force off switches will be “on” for both phases.

   Recall switches would be set to cause the green to rest on the major street unless there is a call on the minor street. In practical terms, either a *minimum recall* or a *soft recall* would be established for Phase B.

Sample Problem 20-3: A Variation on Sample Problem 20-2

What would change if the side-street critical-demand volume were only 85 veh/h, instead of the 240 veh/h of Sample Problem 20-2?

None of the details of detector placement would change, nor would the length of the yellow and all-red phases. Thus, the following values have already been determined:

- $G_{minA} = 6.0$ s
- $y_A = 3.2$ s
- $ar_A = 2.3$ s
- $y_B = 4.3$ s
- $ar_B = 1.0$ s

- L=10.8 s/cycle

The only significant change is in the critical lane volumes. For the minor street (First Avenue, Phase 1), the critical volume is now 85 veh/h. The major street (Main Street, Phase 2) critical-lane volume is unchanged: 1600/2=800 tvu/h/ln. The sum of critical-lane volumes is, therefore, 800+85=885 tvu/h/ln. This will change the initial cycle length and the values of G_{maxA} and G_{minB} that result from it:

Ci=10.81−[8851700×0.92×0.95]=10.81−0.596=26.7 sGA=(26.7−10.8)×(85885)=1.5 sGB=(26.7−10.8)×(800885)=14.4 s sGmaxA=1.5×1.5=2.2 sGmaxB=14.4×1.5=2

This timing is not reasonable, as G_{maxA} (2.2 s) is less than G_{minA} (6.0 s). An alternative way of establishing a reasonable G_{maxA} must be found. The minimum time for G_A is 6.0 seconds, which, given the start-up lost time, might admit one vehicle. How many vehicles (at most) do we wish to allow during Phase A? As a minor street with low volumes, we might add enough green time to accommodate one to three additional vehicles. At approximately 2.0 s/veh, this would add from 2.0 to 6.0 seconds to the minimum green time of 6.0 seconds, moving G_{maxA} to between 8.0 and 12.0 seconds. For illustrative purposes, we will make G_{maxA} 8.0 seconds.

Now, the value of G_{minB} must be increased to maintain the same balance with G_{maxA} as in the original timing. Then:

G minB =21.6(8.0 2.2)=78.5 s

With a smaller side-street demand, the balance of green times is significantly changed, with the major street receiving a significantly longer Gmin and the side street receiving a somewhat smaller Gmax than in the original solution (Sample Problem 20-2).

Sample Problem 20-4: Two Major Arterials with Full-Actuation

An isolated suburban intersection of two major arterials is to be signalized using a full-actuated controller. Presence detection is to be used, except for LT lanes, which will be monitored by point detectors. The intersection is shown in Figure 20.6, and all volumes have already been converted to tvu for convenience. Left-turn slots of 250 ft in length are provided for each approach. The tvu conversions assume that a protected left-turn phase will be provided for all approaches. For all through approaches, 60-ft presence detectors are provided. For LT lanes, a point detector is located 4 ft from the STOP line.

Figure 20.6: Actuated Signal Timing Sample Problem 4: Full-actuated Control

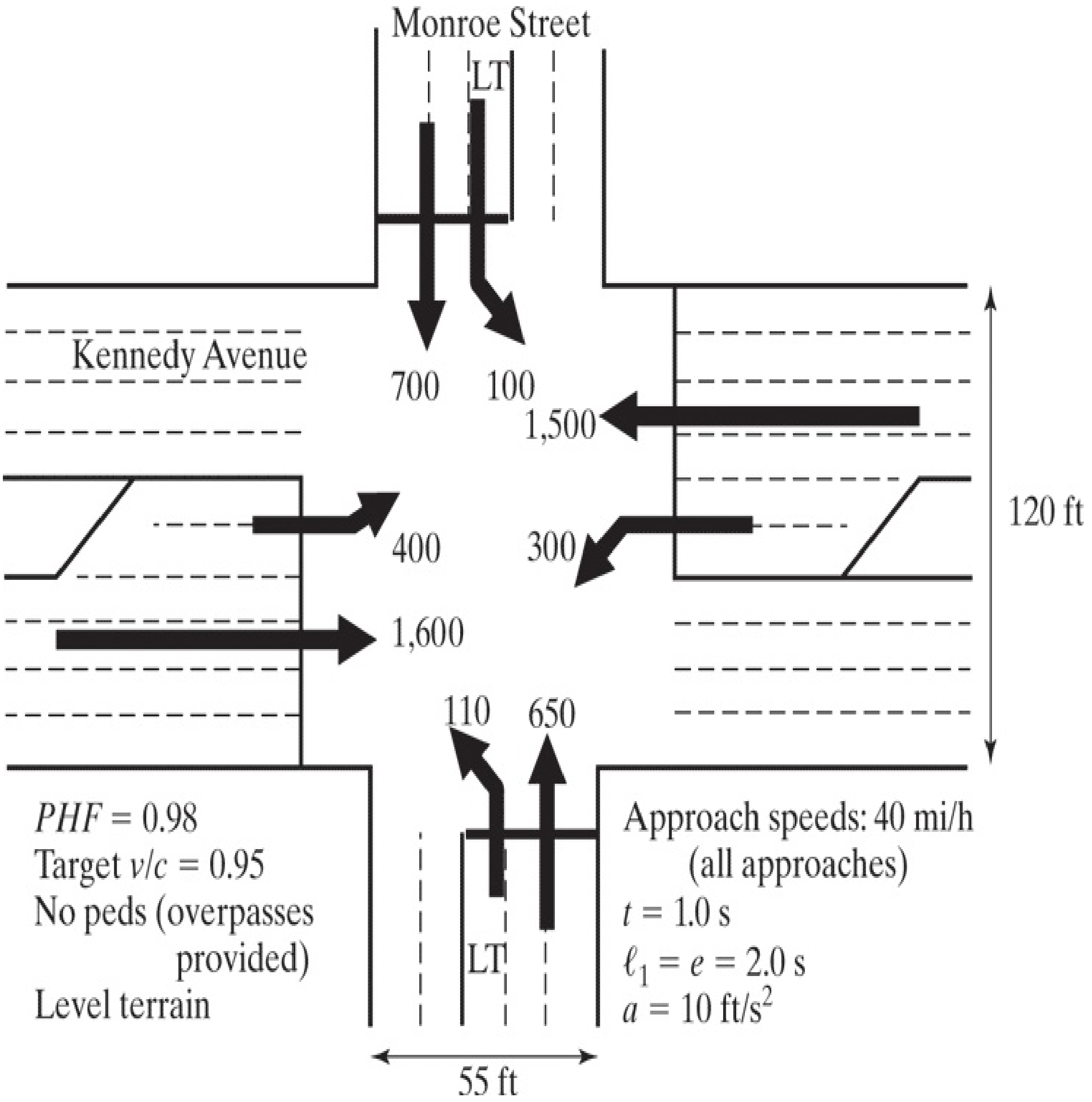


Figure 20.6: Full Alternative Text

1. Step 1: Phasing

   The problem statement indicates that protected left-turn phasing will be implemented on all approaches. The Monroe Street LTs do not meet any of the normal criteria for protection under pre-timed signalization. This is an example of actuated signalization allowing more flexibility in providing LT protection for a wider variety of demands. Note that Kennedy Avenue has double left-turn lanes in each direction and that Monroe Street has a single left-turn lane in each direction.

   At a heavily utilized intersection such as this, quad-eight phasing would be desirable. Each street would have an exclusive LT phase followed by a leading green in the direction of heavier LT flow and a TH/RT phase. As indicated in Chapter 19, such phasing provides much flexibility in that LT phasing is always optional and can be skipped in any cycle in which no LT demand is noted. The resulting signalization has a maximum of four phases in any given cycle and a minimum of two. It is treated as a *four-phase* signal, as this option leads to the maximum lost times.

Quad-eight phasing involves overlaps (again, see Chapter 19) that would be taken into account if this were a pre-timed signal. As an actuated signal, the worst-case cycle, however, would occur when there are no overlap periods. This would occur when the LT flow in opposing directions are equal. Thus, the signal timing will be considered as if this were a simple four-phase operation without overlaps. The controller, however, will allow one protected LT to be terminated before the opposing protected LT, creating a leading green phase. It would also allow one protected LT to begin without the other if no demand is present. Finally, if there are no LTs in either direction, no LT phases would be implemented in that cycle. On the controller, dual entry and simultaneous force off switches are set to accomplish this.

The four phases are:

- Phase A—Protected LT for Kennedy Avenue
- Phase B—TH/RT for Kennedy Avenue
- Phase C—Protected LT for Monroe Street
- Phase D—TH/RT for Monroe Street

2. Step 2: Passage Times

As no gap reduction will be in use at this intersection, the maximum allowable headway, *MAH* is 3.0 seconds. This will be the same for all approaches, including left-turn phases.

The LT phases—A and C—use point detectors. For these cases, the *PT* is equal to the *MAH*, or 3.0 seconds. The through/right turn phases—B and D—use 60-ft presence detectors. The *PT* for these cases is found from Equation 20-4:

PT = MAH− L v + L d 1.47 S a PT = 3.0− 60+20 1.47×40 =1.6 s

3. Step 3: Minimum Green Times and Detector Placement

   The detector design has been specified. The minimum green time for area detection is variable, based on the number of vehicles sensed within the detection area when the green is initiated. The value can vary from the time needed to service one waiting vehicle to the time needed to service *Int* (60/25)=3 vehicles. Using Equation 20-2, the range of minimum green times can be established for Phases B and D.

   G min = ℓ 1 +2n G min,low = 2.0+(2×1)=4.0 s G min,high = 2.0+(2×3)=8.0 s

   Phases A and C use point detectors located 4 ft from the STOP line. The minimum green time for these phases is given by Equation 20-1:

   Gmin=ℓ1+2.0 Int (d25)Gmin=2.0+2.0Int (425)=4.0 s

   The minimum green times are summarized below:

   G minA/C = 4.0 s G minB/D = 4.0−8.0 s

4. Step 4: Critical-Lane Volumes

   As the volumes given have already been converted to tvu, critical-lane volumes for each phase are easily identified:

- Phase A (Kennedy Ave, LT)=400/2=200 tvu/h/ln
- Phase B
  (Kennedy Ave, TH/RT)=1600/4=400 tvu/h/ln
- Phase C (Monroe St, LT)=110/1=110 tvu/h/ln
- Phase D (Monroe St, TH/RT)=700/2=350 tvu/h/ln
- V c =200+400+110+350=1,060 tvu/h/ln

5. Step 5: Yellow and All-Red Times

   Yellow times are found using Equation 20-6; all-red times are found using Equation 20-7. With a 40-mi/h average approach speed for all movements, the S_{85} may be estimated as 40+5=45 mi/h, and the S_{15} may be estimated as 40−5=35 mi/h. As there are no pedestrians, *all-red* intervals will use *w* the width being crossed. The standard length of a vehicle is taken as 20 ft. Then:

   y = t+ 1.47 S 85 2 (a+32.2G) y A,B,C,D = 1.0+ 1.47×45 2 (10+32.2×0) =4.3 s ar = w+L 1.47 S 15 a r A,B = 55+20 1.47×35 =1.5 s a r C,D = 120+20 1.47×35 =2.7 s

6. Step 6: Total Lost Time

   There are four phases in the worst-case cycle. As the standard defaults are used for ℓ1 and *e*, the total lost time is equal to the sum of the yellow and all-red intervals in the cycle:

   L=(4.3+1.5)+(4.3+1.5)+(4.3+2.7)+(4.3+2.7)=25.6 s

7. Step 7: Maximum Green Times and the Critical Cycle

   The initial cycle length for determining maximum green times is determined by Equation 20-9:

   Ci=25.61−[10601700×0.98×0.95]=25.61−0.670=77.6 s

Green times are found using Equation 20-10:

$$g_i = g_{TOT}\left(\frac{V_{ci}}{V_c}\right)$$
$$g_{TOT} = 77.6-25.6=52.0 \text{ s}$$
$$g_A = 52.0\left(\frac{200}{1060}\right)=9.8 \text{ s}$$
$$g_B = 52.0\left(\frac{400}{1060}\right)=19.6 \text{ s}$$
$$g_C = 52.0\left(\frac{110}{1060}\right)=5.4 \text{ s}$$
$$g_D = 52.0\left(\frac{350}{1060}\right)=17.2 \text{ s}$$

Because standard defaults are in place, the effective green, *g*, is equal to the actual green, *G*. Maximum green times are found by multiplying the average green times above by 1.50:

$$G_{maxA} = 9.8\times1.5=14.6 \text{ s}$$
$$G_{maxB} = 19.6\times1.5=29.4 \text{ s}$$
$$G_{maxC} = 5.4\times1.5=8.1 \text{ s}$$
$$G_{maxD} = 17.2\times1.5=25.8 \text{ s}$$

The critical cycle length is the sum of the maximum greens plus the lost time per cycle, or:

$$C_c=14.6+29.4+8.1+25.8+25.6=103.2 \text{ s}$$

All of the maximum green times are compatible with the minimum green times computed earlier.

8. Step 8: Pedestrians

   As overpasses are provided for pedestrians, there are no at-grade crossings permitted, and no pedestrian checks are required for this signalization.

9. Step 9: Dual Entry, Simultaneous Force Off, and Recall Settings

   This actuated signal is to operate with optional LT phases that can (a) begin at the same time and end at different times, (b) begin in one direction and not the other, or (c) be skipped entirely. Because leading green phases are possible, the through movements may or may not start at the same time. The through movements, however, *must* end at the same time, as the green is then transferred to the conflicting street. There are no lagging greens.

   Figure 20.7 shows a ring diagram for the signal, with the settings for dual entry, simultaneous force off, and recall indicated.

Figure 20.7: Ring Diagram for Actuated Signal Timing Sample Problem 4

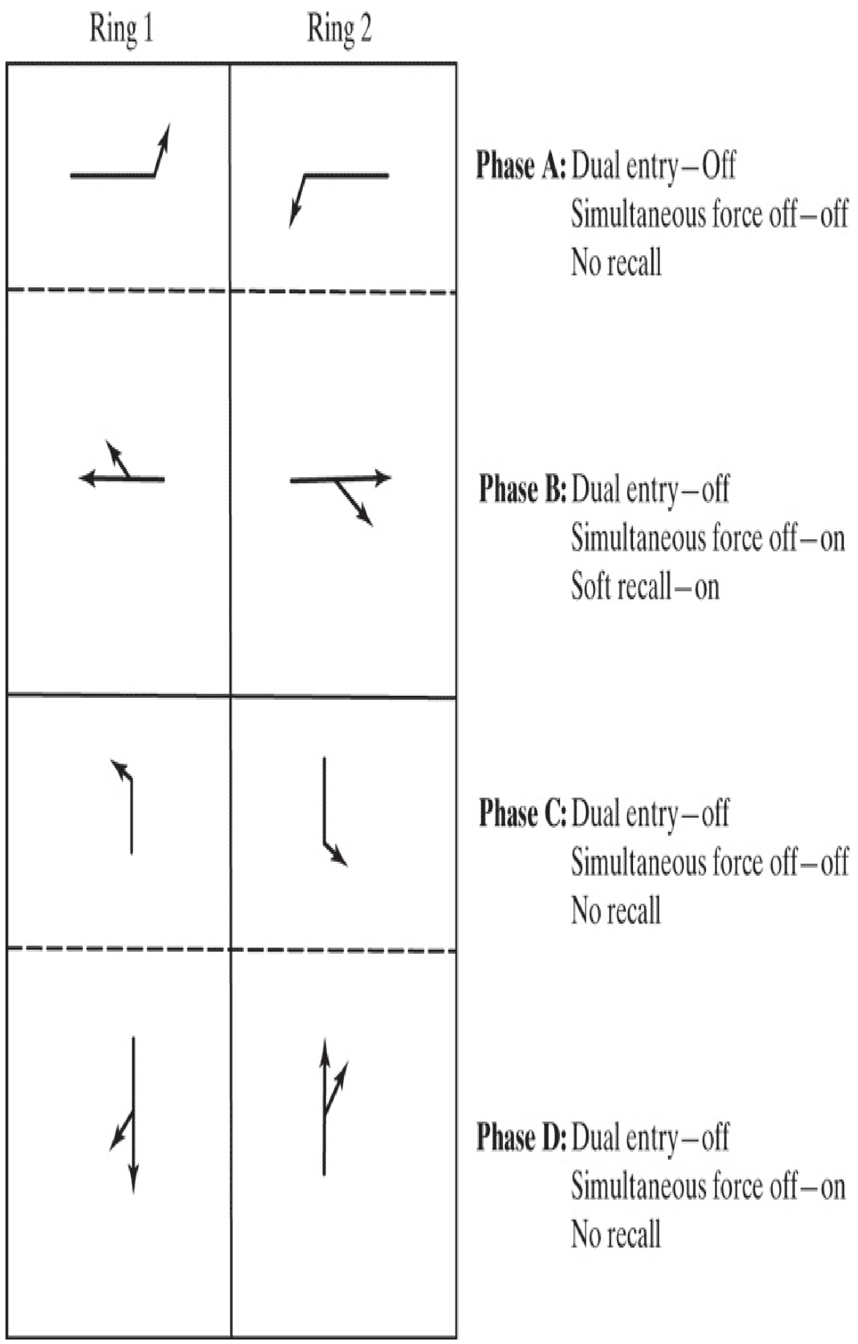


Figure 20.7: Full Alternative Text

References

- 1. Pusey, R.S. and Butzer, G.L, “Traffic Control Signals,” *Traffic Engineering Handbook*, 5th Edition, Institute of Transportation Engineers, Washington, D.C., 1999, pp. 453–528.
- 2. Kittelson and Associates, *Traffic Signal Timing Handbook*, Federal Highway Administration, Washington, D.C., June 2008.

Problems

Unless otherwise noted, use the following default values for each of the following actuated signal timing problems:

- Driver reaction time, t=1.0 s
- Vehicle deceleration rate, a=10 ft/s
- Length of a vehicle=L=20 ft
- Start-up lost time, ℓ 1 =2.0 s
- Encroachment time, e=2.0 s
- Level terrain
- Low pedestrian activity at all locations (50 peds/h each cross walk)
- PHF=0.90
- Target v/c ratio for actuated widths=12 ft
- Lane signals=0.95
- Crosswalk widths=10 ft, with 2 ft setback
- Pedestrian crossing speed=4.0 ft
- All volumes in veh/h

If any other assumptions are necessary, specifically indicate them as part of the answer.

1. 20-1. A semiactuated signal is to be installed and timed for the location shown. Because of light side-street demand, a short minimum green of 6.0 seconds is desired. Point detectors will be used. For the conditions shown:

    1. How far from the STOP line should the side-street detectors be

located?

2. Recommend a passage time.

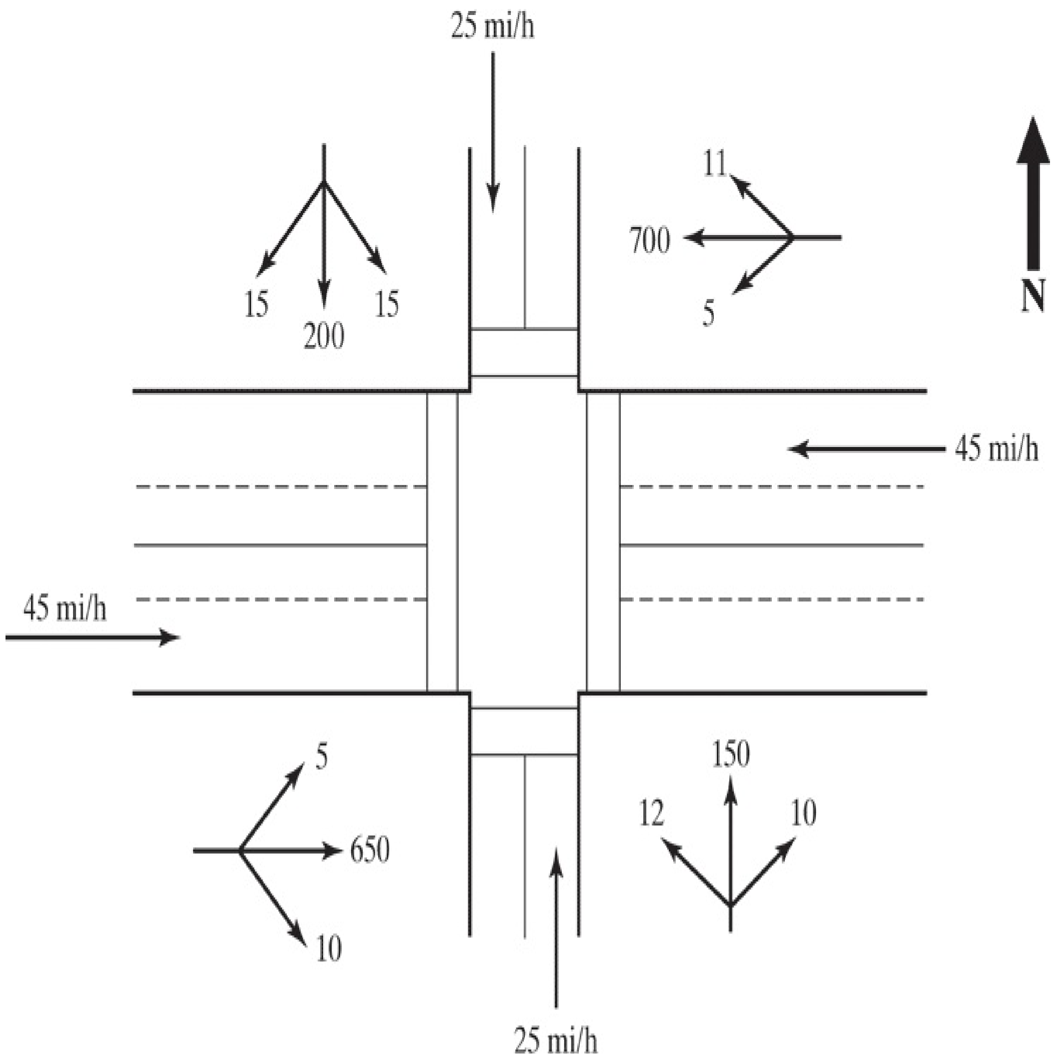
3. Compute yellow and all-red times.

4. Recommend a maximum side-street green and a minimum main street green.

5. What is the critical cycle length?

6. Are pedestrian signals and/or push buttons required for crossing the main street? The side street?

1. # Intersection for Problem 20-1



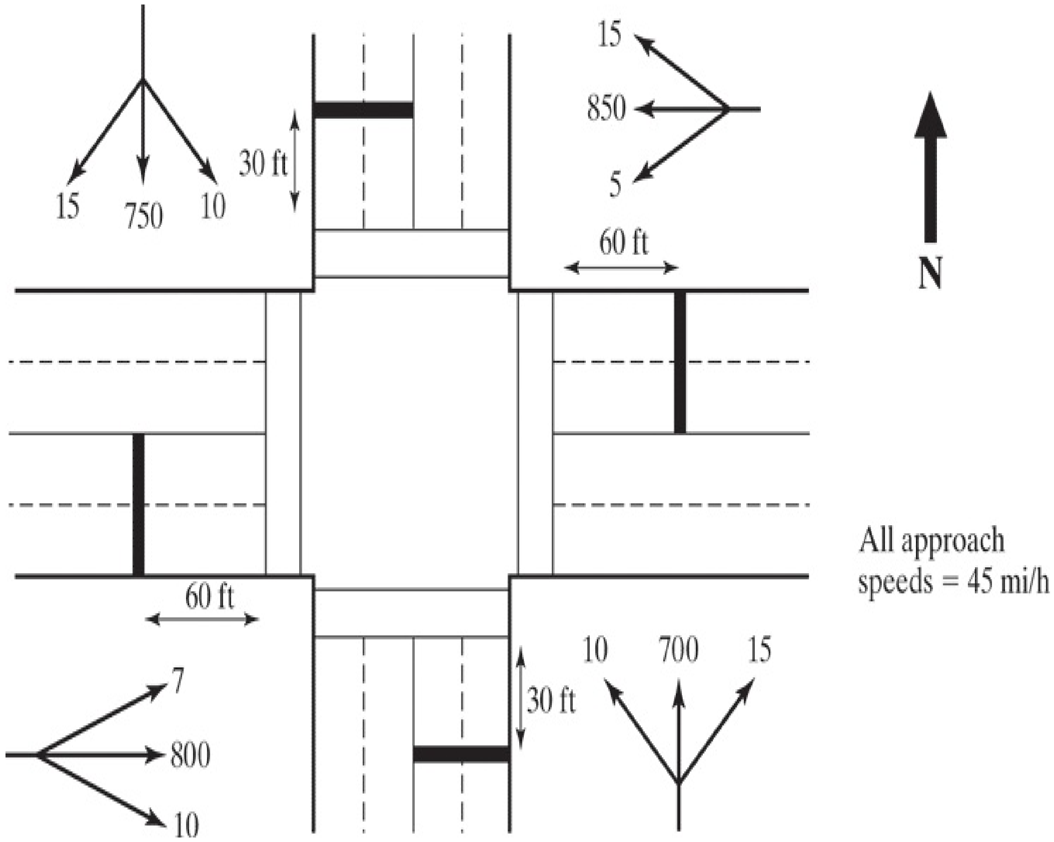
Full Alternative Text

2. 20-2. A full-actuated controller must be retimed at the intersection shown. Detector locations are fixed from a previous installation and cannot be moved. For the conditions shown:

   1. Recommend a suitable phase plan for the signal.
   2. What minimum greens should be used?

3. Recommend a passage time.
4. Compute yellow and all-red times.
5. Recommend maximum green times for each phase.
6. What is the critical cycle length?
7. Are pedestrian signals and/or push buttons needed for any phases?

1. **Intersection for Problem 20-2**



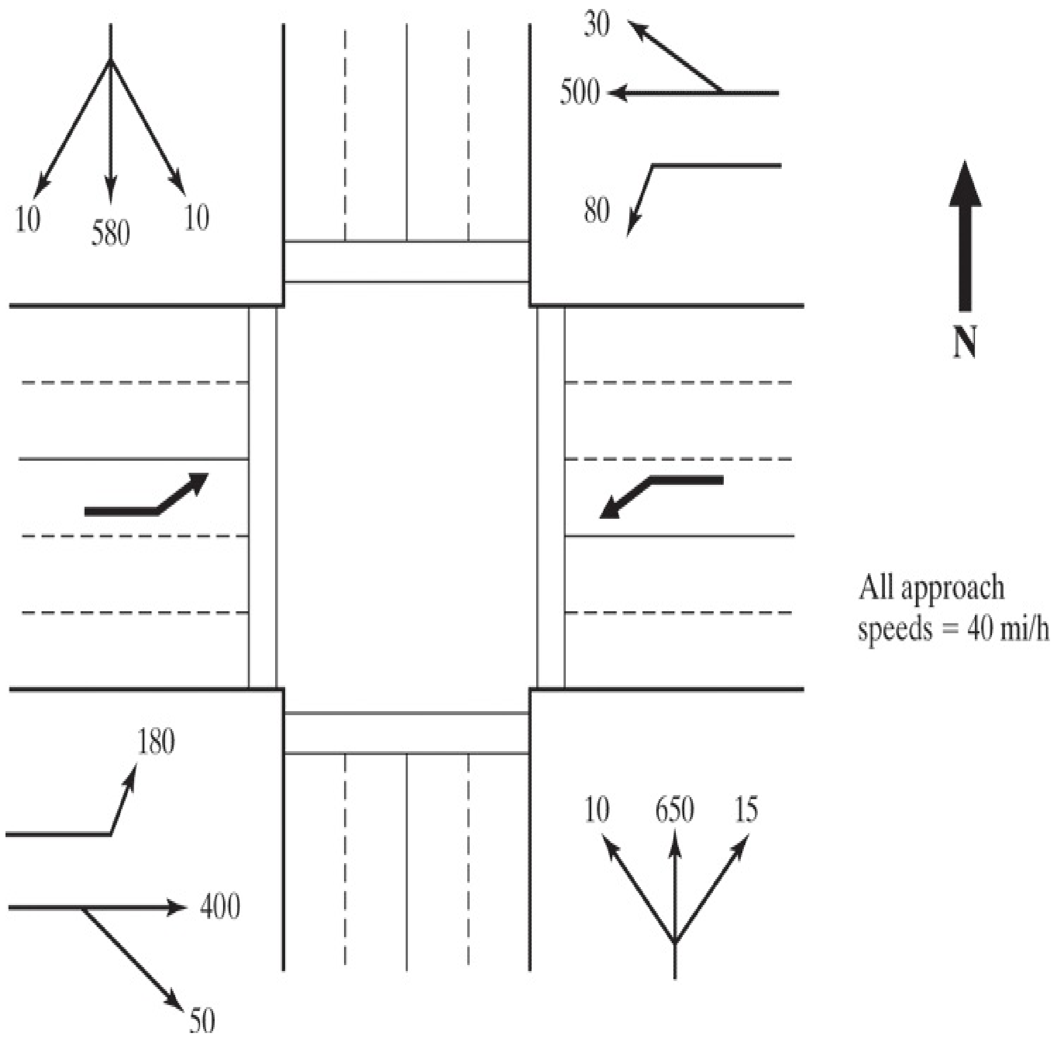
Full Alternative Text

2. 20-3. A full-actuated signal is to be installed at the major intersection shown. 40-ft presence detectors with their front edges located at the STOP line will be used. For this location and the conditions shown:

    1. Recommend a suitable phase plan for the signal.
    2. Recommend minimum green times.
    3. Recommend a passage time.
    4. Compute yellow and all-red times.

5. Recommend maximum green times for each phase.
6. What is the critical cycle length?
7. How are the dual entry and simultaneous force-off switches set?
8. Are pedestrian signals and/or push buttons needed for any phases?

1. **Intersection for Problem 20-3**



Full Alternative Text

Chapter 21 Signal Coordination for Arterials and Networks

In situations where signals are close enough together, it is desirable to time the signals so that vehicles arrive at the downstream intersection in platoons. It serves no purpose to have drivers held at one signal watching wasted green at a downstream signal, only to arrive there just as the signal turns red. To accomplish having the vehicles arrive at the downstream intersection in platoons, during the green phase, it is necessary to coordinate the green times so that vehicles may move efficiently through a *set* of signals in order to minimize stops and delay.

In some cases, two signals are so closely spaced that they should be considered to be one signal. In other cases, the signals are so far apart that they may be considered as isolated intersections. Vehicles released from a signal often maintain their grouping for over 2,000 feet. Common practice is to coordinate signals less than one half mile apart on major streets and highways [1]. Signals over one-half mile apart may be coordinated if the platoons can be maintained.

21.1 A Key Requirement: A Common Cycle Length

In coordinated systems, all signals must have the same cycle length. This is necessary to ensure that the beginning of green occurs at the same time relative to the green at the upstream and downstream intersections. There are some exceptions, where a critical intersection has such a high volume that it may require a double cycle length, for instance, but this is done rarely and only when no other solution is feasible. When this is done, the cycle length must be a multiple of the common cycle length.

21.2 The Time-Space Diagram

The time-space diagram is a plot of signal indications as a function of time for two or more signals. The diagram is scaled with respect to distance, so that one may easily plot vehicle positions as a function of time. Figure 21.1 is a time-space diagram for two intersections. Standard conventions are used in the figure: The green time indication is shown by a blank line and red by a thick solid line. This figure illustrates the path (trajectory) that a vehicle takes as time passes. At t=t1, the first signal turns green. The vehicle passes through the intersection at the beginning of green and moves down the street. It reaches the second intersection at some time t=t2. In Figure 21.1, the vehicle arrives downstream at the start of green and passes through the downstream intersection without stopping. The time difference between the first vehicle that can pass through the entire system without stopping and the last vehicle that can pass through without stopping at a given speed is defined as the bandwidth, measured in seconds.

Figure 21.1: Illustrative Vehicle Trajectory on a Time-Space Diagram

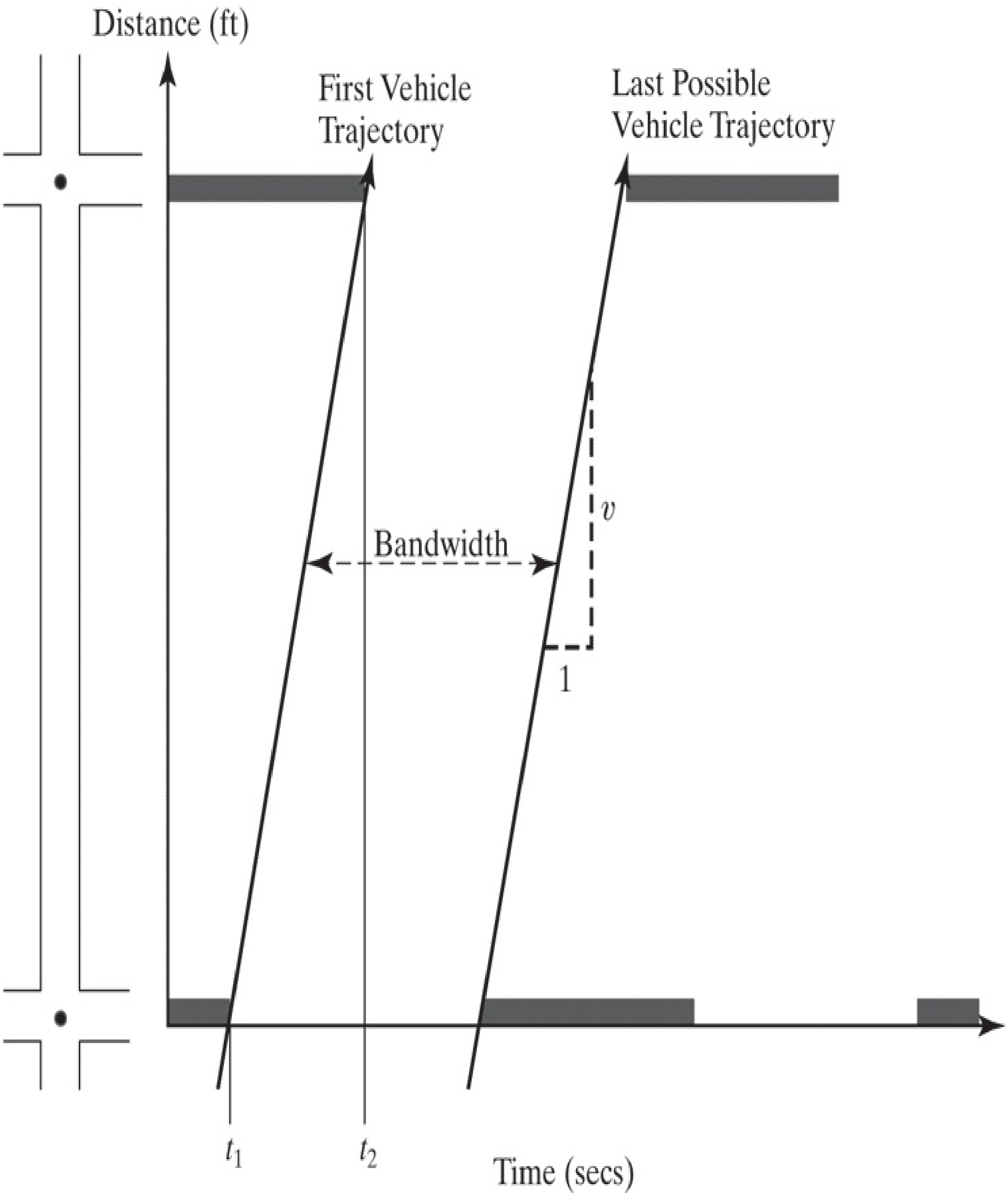


Figure 21.1: Full Alternative Text

The difference between the two green initiation times (i.e., the difference between the time when the upstream intersection turns green and the downstream intersection turns green), is referred to as the *signal offset*, or simply the *offset*. In Figure 21.1, the offset is defined as t2 minus t1. Offset is usually expressed as a positive number between zero and the cycle length. The offset can be defined relative to the previous intersection or relative to a master clock zero. The master clock is a background timing device for the controller logic. Each intersection's initiation of main street green would be referenced to zero on the master clock (which runs a full cycle equal to the cycle length).

There are other definitions of offset used in practice. Some signal hardware uses “offset” defined in terms of red initiation, rather than green; other hardware uses the end of green as the reference point. Some hardware uses offset in seconds; other hardware uses offset as a percentage of the cycle length.

Sample Problem 21-1

If the cycle length in Figure 21.1 is 60 seconds, and t1=10 seconds relative to the master clock zero, and t2 is 30 seconds after t1, what is the offset at the downstream intersection relative to the master clock and relative to the previous intersection?

Solution

Offset at downstream intersection:

Offset Relative to Master Clock = 10+30=40 sOffset Relative to Previous I

21.3 Ideal Offsets

The “ideal offset” is defined as the offset such that the first vehicle of a platoon arrives at the downstream signal just as it turns green. It is usually assumed that the platoon was moving as it went through the upstream intersection. If so, the ideal offset is given by:

$t_{ideal}=L/S$ [21-1]

where:

t_{ideal}=ideal offset, s,L=distance between signalizedintersections, ft, and$S$=a

If the vehicle was stopped at the upstream intersection and had to accelerate after some initial start-up delay, the ideal offset could add the start-up time at the first intersection, which would typically add 2 to 4 seconds.

In general, the start-up time would be included only at the *first* of a series of signals to be coordinated, and often not at all. Usually, this will reflect the ideal offset desired for maximum bandwidth, minimum delay, and minimum stops. Even if the vehicle is stopped at the first intersection, it will be moving through most of the system.

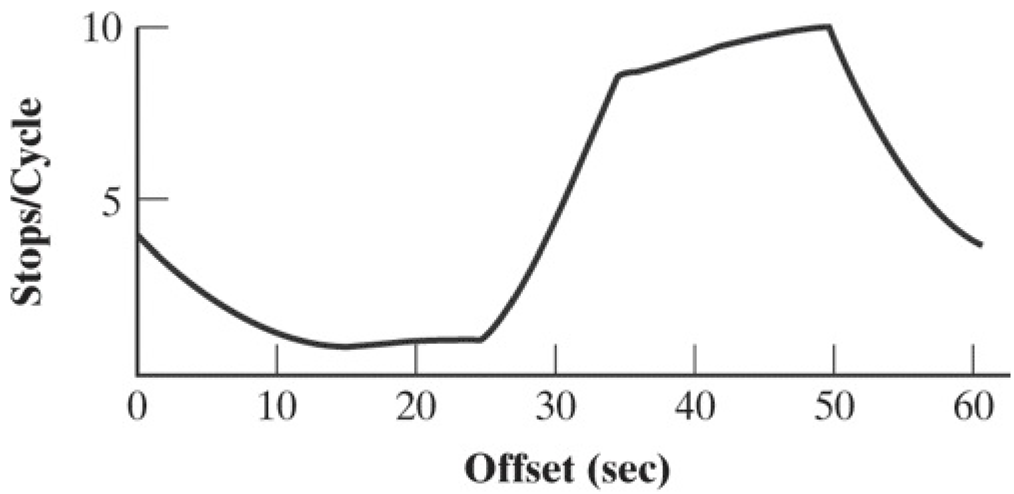
Figure 21.1 also illustrates the concept of *bandwidth*. Bandwidth is the amount of green time that can be used by a continuously moving platoon of vehicles through a series of intersections. In Figure 21.1, the bandwidth is the entire green time at both intersections, because several key conditions exist:

- The green times at both intersections are the same.
- The ideal offset is illustrated.
- There are only two intersections.

In most cases, the bandwidth will be less, perhaps significantly so, than the full green time.

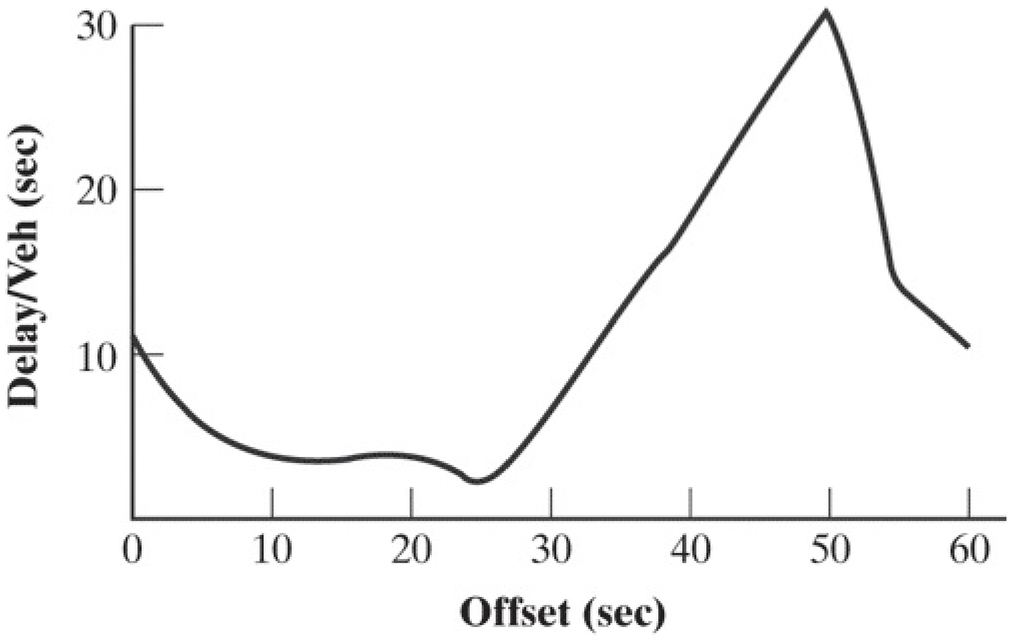
Figure 21.2 illustrates the effect of offset on stops and delay for a platoon of vehicles leaving one intersection and passing through another. In this example, a 25-second offset is ideal, as it produces the minimum delay and the minimum number of stops. The effect of allowing a poor offset to exist is clearly indicated: delay can climb to 30 seconds per vehicle and the stops to 10 per cycle. Note that the penalty for deviating from the ideal offset is usually not equal in positive and negative deviations. An offset of $(25+10)=35$s causes much more harm than an offset of $(25-10)=15$s, although both are 10 seconds from the ideal offset. Figure 21.2 is illustrative, as each situation would have similar but different characteristics.

Figure 21.2: Illustration of the Effects of Offset on Stops and Delay



(a) Stops

21.3-1 Full Alternative Text



(b) Delay

600 ft block

600 vph in two lanes

All through traffic

Free speed 24 mph

50-50 split

60 sec cycle length

21.3-1 Full Alternative Text

21.4 Signal Progression on One-Way Streets

Signal progression on a one-way street is relatively simple. For the purpose of this section, it will be assumed that a cycle length has been chosen and that the green allocation at each signal has been previously determined.

21.4.1 Determining Ideal Offsets

Consider the one-way arterial shown in Figure 21.3, with the link lengths indicated. Assuming no vehicles are queued at the signals, the ideal offsets can be determined if the platoon speed is known. For the purpose of illustration, a desired platoon speed of 60 ft/s will be used. The cycle length is 60 seconds, and the effective green time at each intersection is 50% of the cycle length, or 30 seconds. Ideal offsets are computed using Equation 21-1, and are listed in Table 21.1.

Table 21.1: Ideal Offsets for the Case Study

Signal	Ideal Offset (Relative to Previous Intersection)	Offset Relative to Master Clock
6	$= 1{,}800/60 = 30$ s	40 s
5	$= 600/60 = 10$ s	10 s
4	$= 1{,}200/60 = 20$ s	60 s $= 0$ s
3	$= 1{,}200/60 = 20$ s	40 s
2	$= 1{,}200/60 = 20$ s	20 s
1	0	0 s

Table 21.1: Full Alternative Text

Figure 21.3: Case Study in Progression of One-Way Street

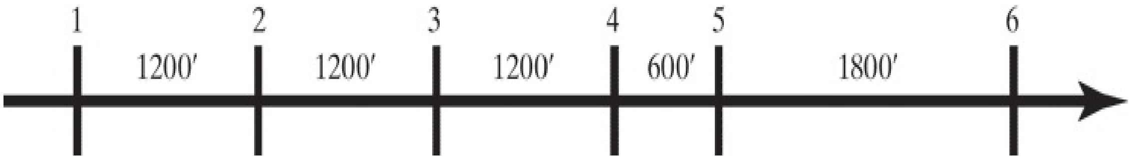


Figure 21.3: Full Alternative Text

In order to see the pattern that results, the time-space diagram should be constructed according to the following rules:

1. The vertical should be scaled so as to accommodate the dimensions of the arterial, and the horizontal so as to accommodate at least three to four cycle lengths.

2. The beginning intersection (Number 1, in this case) should be scaled first, with main street green (MSG) initiation at t=0, followed by periods of green and red (yellow may be shown for precision). See Point 1 in Figure 21.4.

Figure 21.4: Time-Space Diagram for the Case Study

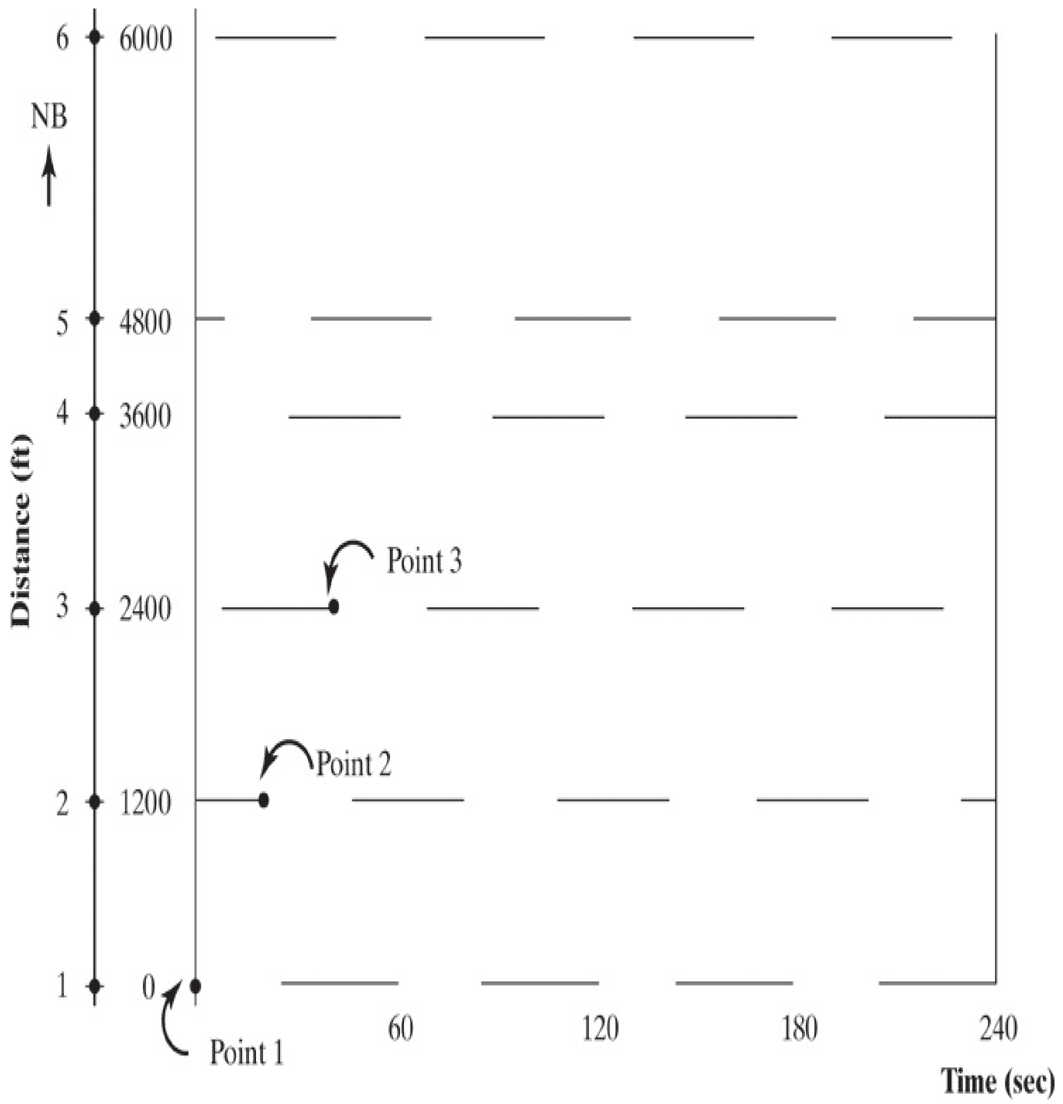


Figure 21.4: Full Alternative Text

3. The main street green (or other offset position, if MSG is not used) of the next downstream signal should be located next, relative to t=0 and at the proper distance from the first intersection. With this point located (Point 2 in Figure 21.4), fill in the periods of effective green and red for this signal.

4. Repeat the procedure for all other intersections, working one at a time.

Figure 21.4 has some interesting features that can be explored with the aid of Figure 21.5. Note, there is a window of green, which looks like a “green wave” that is visible to a stationary observer at Signal 1. The signals turn green in order, corresponding to the planned speed of the platoon, and give the visual effect of a wave of green opening before the driver. The speed at which this green wave travels down the arterial is called the progression speed. In this example, the progression speed is 60 ft/s. If a vehicle (or platoon) were to travel at 60 ft/s, it would arrive at each of the signals just as it turn greens; this is indicated by the solid trajectory lines in Figure 21.5.

Figure 21.5: Vehicle Trajectory of “Green Wave” in Progressed Movement of Figure 21.4

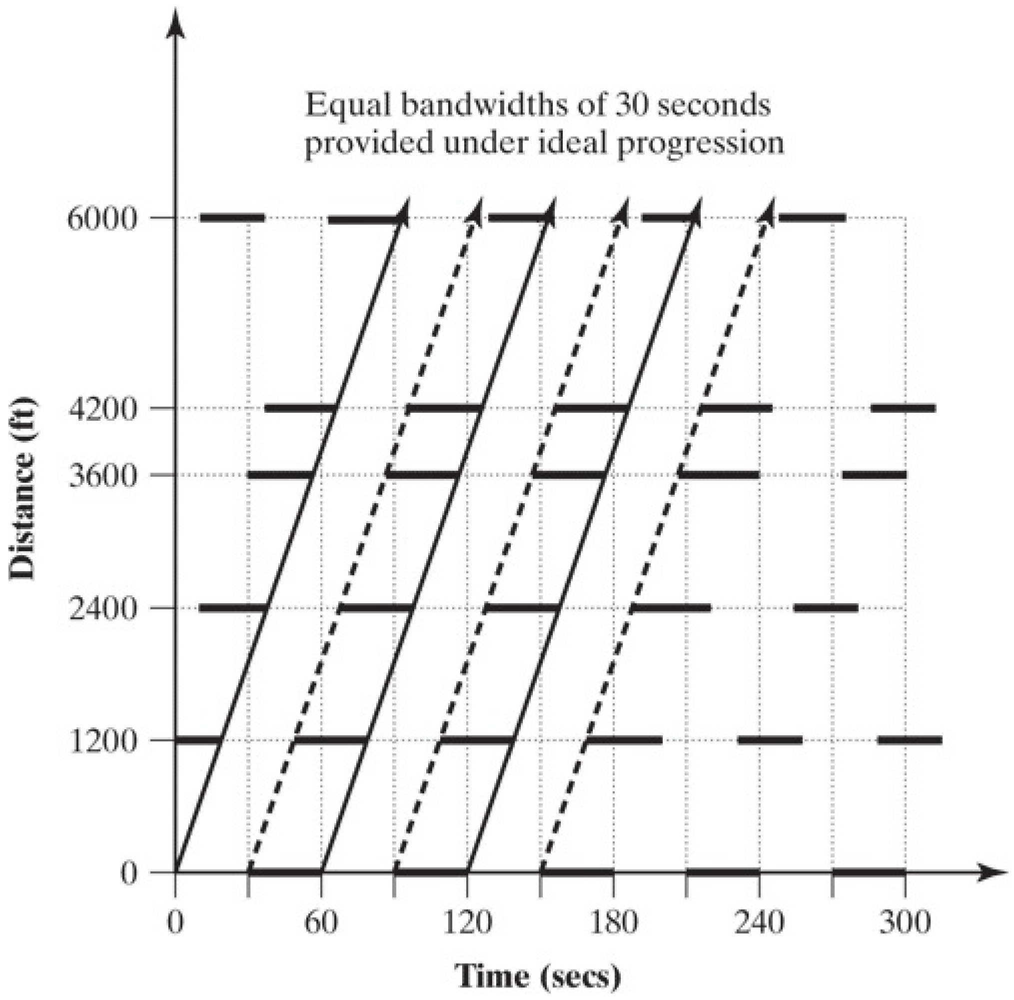


Figure 21.5: Full Alternative Text

Note that the “window” of green in Figure 21.5 has its end indicated by the dotted trajectory line; this is also the trajectory of the *last* vehicle that could travel through the progression without stopping at 60 ft/s. This “window” is the bandwidth, as defined earlier. In this case, it equals the green time (30 seconds) because all signals have the same green time and have ideal offsets.

21.4.2 Bandwidth Efficiency

The efficiency of a bandwidth is defined as the ratio of the bandwidth to the cycle length, expressed as a percentage:

$EFFBW=(BWC)\times100\%$ [21-2]

where:

EFFBW=bandwidth efficiency, %,BW=bandwidth, s, andC=cycle length, s

A bandwidth efficiency of 40% to 55% is considered good. Note that although the cycle ength must be the same (or a multiple) at all signals being coordinated, the splits at each intersection may be different. The bandwidth is limited by the minimum green in the direction of interest. The bandwidth efficiency of the arterial in Figure 21.5 is 50%.

21.4.3 Bandwidth Capacity

In terms of vehicles that can be put through the system without stopping, the *bandwidth capacity* is the number of vehicles that can pass through a defined series of signals without stopping. The bandwidth capacity can be described by the following equation:

cBW=3,600 × BW ×LC × h [21-3]

where:

cBW=bandwidth capacity, veh/h,BW=bandwidth, s,L=number of through l

Equation 21-3 does not contain any factors to account for nonuniform lane utilization and is intended only to indicate some limit beyond which the offset plan will degrade, certainly resulting in stopping and internal queuing. It should also be noted that bandwidth capacity is *not* the same as lane group capacity. Where the bandwidth is less than the full green time, there is additional lane group capacity outside of the bandwidth.

In the arterial of Figure 21.5, the bandwidth can carry 30/2=15 vehicles per lane per cycle in a nonstop path through the defined system, assuming that the saturation headway is 2.0 s/veh. Thus, the arterial can handle 15 veh/cycle×1 cycle/60 s×3,600 s/h=900 veh/h/ln very efficiently if they are organized into 15-vehicle platoons when they travel through this system.

21.4.4 Potential Problems

Consider what would happen if the actual speed of vehicle platoons in the case study was 50 ft/s, instead of the 60 ft/s anticipated. The green wave would still progress at 60 ft/s, but the platoon arrivals would lag behind it. The effect of this on bandwidth is enormous, as shown in Figure 21.6. Only a small window now exists for a platoon of vehicles to continuously flow through all six signals without stopping.

Figure 21.6: The Effect of a 50 ft/s Platoon Speed on Progression in Case Study

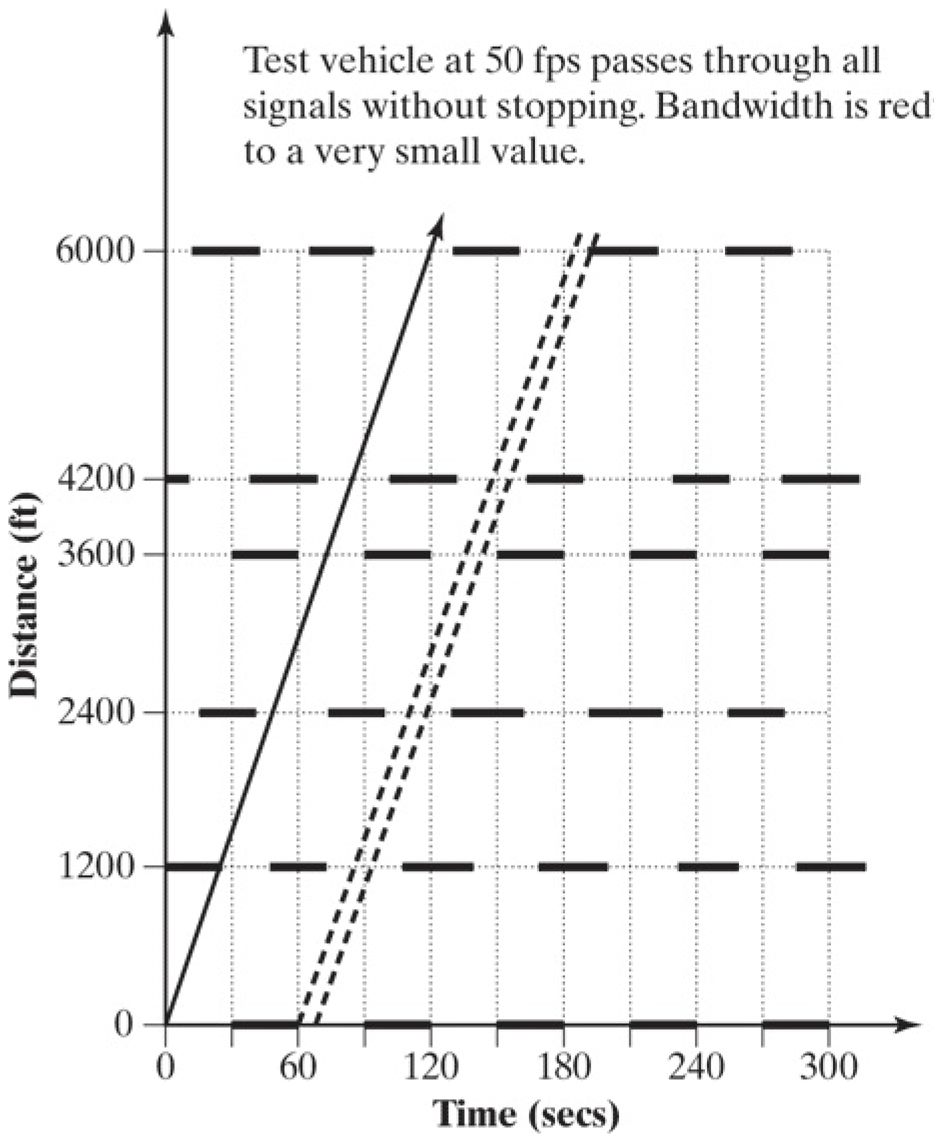


Figure 21.6: Full Alternative Text

Figure 21.7 shows the effect of the vehicle traveling faster than anticipated, 70 ft/s in this illustration. In this case, the vehicles arrive a little too early and are delayed; some stops will have to be made to allow the “green wave” to catch up to the platoon. The effect on bandwidth is not as severe as in Figure 21.6. In this case, the bandwidth impact of *underestimating* the platoon speed (60 ft/s instead of 70 ft/s) is not as severe as the consequences of *overestimating* the platoon speed (60 ft/s instead of 50 ft/s).

Figure 21.7: The Effect of a 70 ft/s Platoon Speed on Progression

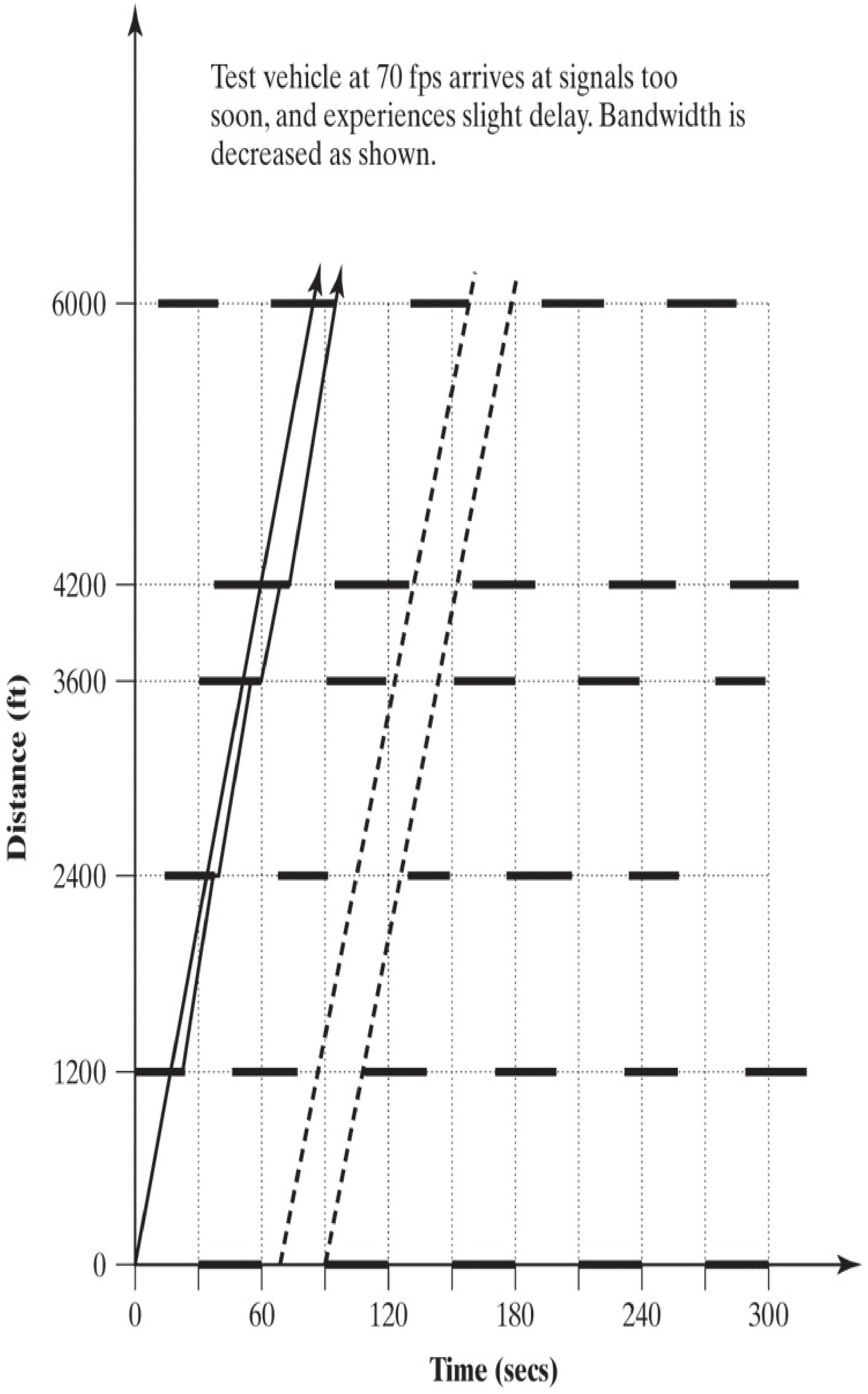

Test vehicle at 70 fps arrives at signals too soon, and experiences slight delay. Bandwidth is decreased as shown.
Distance (ft)
6000
4200
3600
2400
1200
0
0
60
120
180
240
300
Time (secs)

Figure 21.7: Full Alternative Text

21.4.5 The Effect of Queued Vehicles at Signals

To this point, it has been assumed that there is no queue standing at the downstream intersection when the platoon from the upstream signal arrives. This is generally not a reasonable assumption. Vehicles that enter the traffic stream between platoons will progress to the downstream signal, which will often be “red.” They form a queue that partially blocks the progress of the arriving platoon. These vehicles may include stragglers from the last platoon, vehicles that turned into the block from unsignalized intersections or driveways, or vehicles that came out of parking lots or parking spots. The ideal offset must be adjusted to allow for these vehicles, so as to avoid unnecessary stops. The situation without such an adjustment is depicted in Figure 21.8, where it can be seen that the arriving platoon is delayed behind the queued vehicles as the queued vehicles begin to accelerate through the intersection.

Figure 21.8: The Effect of Queued Vehicles at a Signal

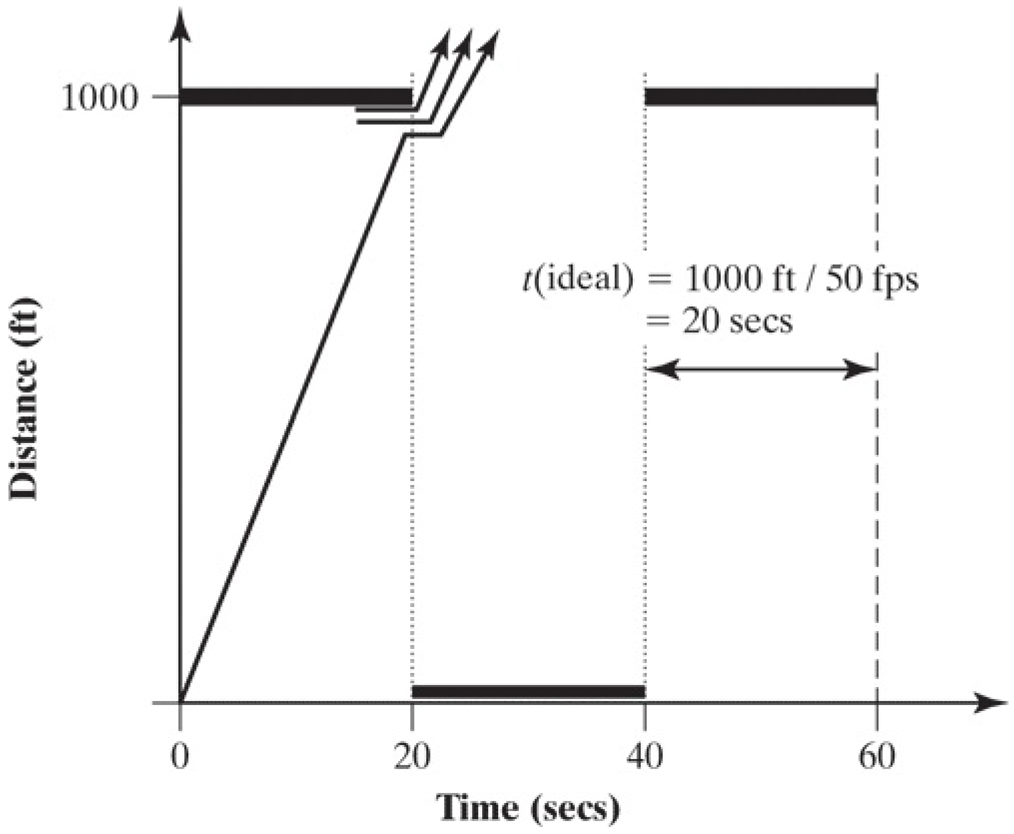


Figure 21.8: Full Alternative Text

To adjust for the queued vehicles, the ideal offset is adjusted as follows:

tadj=LS − (Qh + ℓ1) [21-4]

where:

tadj=adjusted ideal offset, s,L=distance between signals, ft,S=speed, ft/s,Q=
up lost time, s.

The lost time is counted only at the first downstream intersection, at most. If the vehicle(s) from the preceding intersection were themselves stationary, their startup causes a shift that automatically takes care of the startup at subsequent intersections.

Offsets can be adjusted to allow for queue clearance before the arrival of a platoon from the upstream intersection. Figure 21.9 shows the situation for use of the modified ideal offset equation.

Figure 21.9: Adjustment in

Offset to Accommodate Queued Vehicles

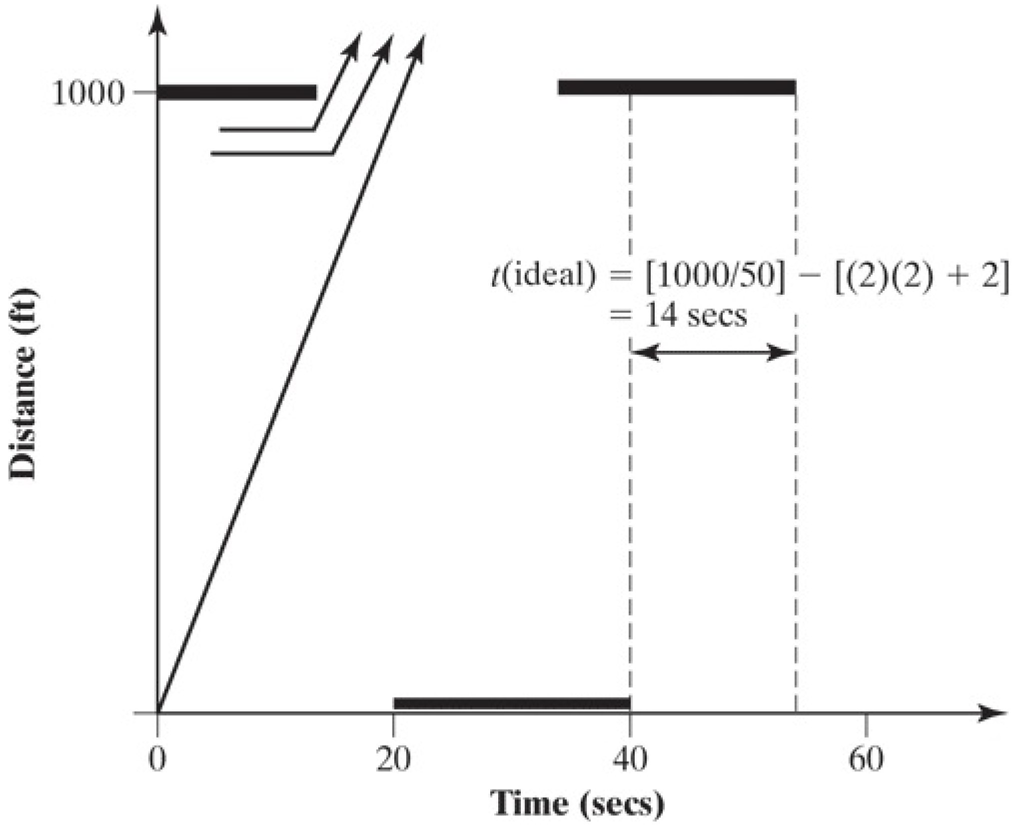
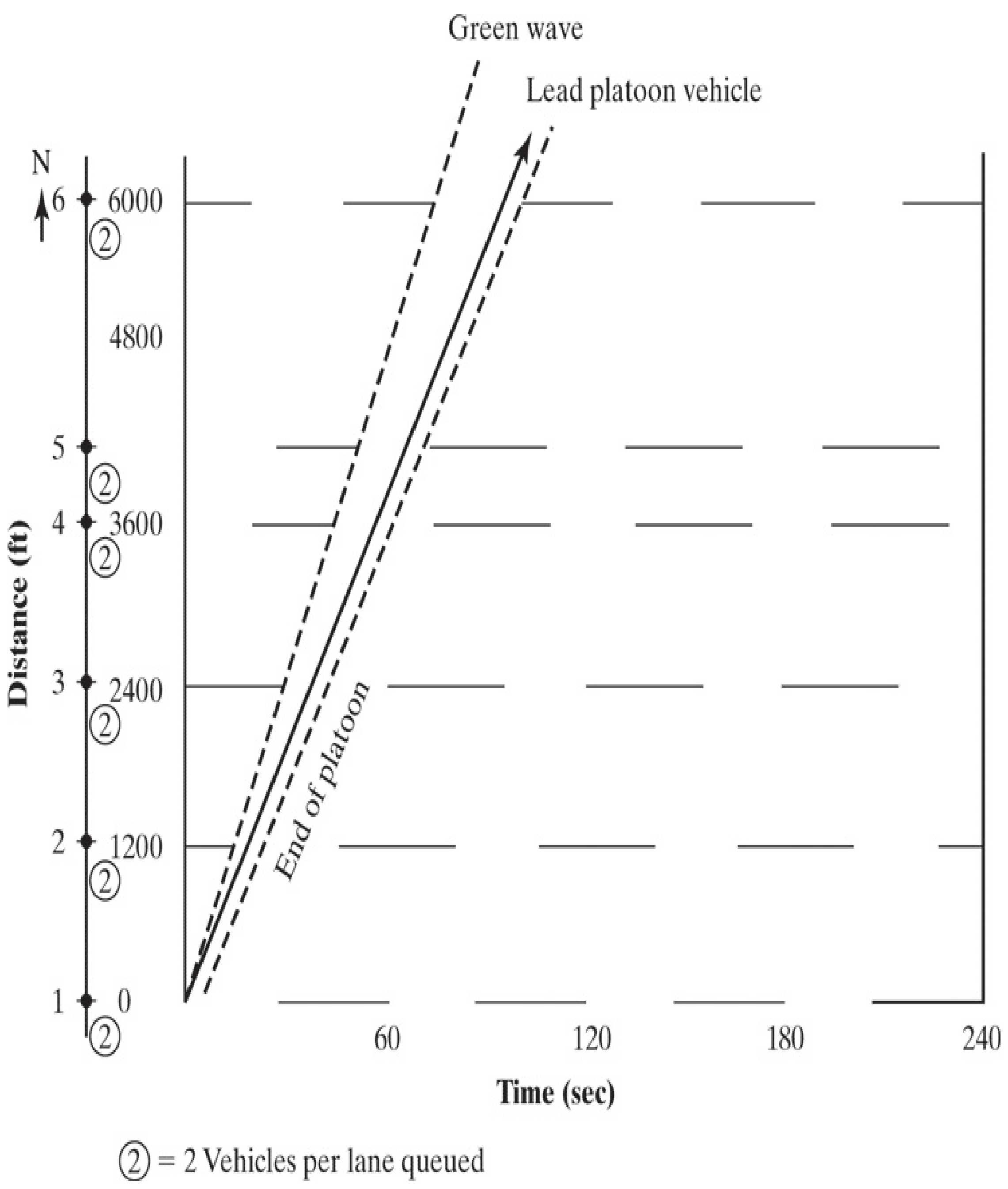


Figure 21.9: Full Alternative Text

Figure 21.10 shows the time-space diagram for the case study of Figure 21.5, given queues of two vehicles per lane in all links. Note that the arriving vehicle platoon has smooth flow, and the lead vehicle has 60 ft/s travel speed. The visual image of the “green wave,” however, is much faster, due to the need to clear the queues in advance of the arriving platoon.

Figure 21.10: Effect of Queue Clearance on Progression Speed



[Figure 21.10: Full Alternative Text](#)

The "green wave," or the progression speed as it is more properly called, is traveling at varying speeds as it moves down the arterial. The "green wave" will appear to move ahead of the platoon, clearing queued vehicles in advance of it. The progression speed can be computed for each link as shown in Table 21.2.

Table 21.2: Progression

Speeds in Figure 21.10

Signal	Link Offset (Relative to Previous Intersection)	Speed of Progression, ft/s
Signal 5 $\rightarrow$ 6	$= 1{,}800/60 - 4 = 26$ s	$1{,}800/26 = 69.2$
Signal 4 $\rightarrow$ 5	$= 600/60 - 4 = 6$ s	$600/6 = 100$
Signal 3 $\rightarrow$ 4	$= 1{,}200/60 - 4 = 16$ s	$1{,}200/16 = 75$
Signal 2 $\rightarrow$ 3	$= 1{,}200/60 - 4 = 16$ s	$1{,}200/16 = 75$
Signal 1 $\rightarrow$ 2	$= 1{,}200/60 - (4 + 2) = 14$ s	$1{,}200/14 = 85.7$
1	$= 0$ s	n/a

Table 21.2: Full Alternative Text

It should be noted, however, that the bandwidth and, therefore, the bandwidth capacity is now much smaller. Thus, by clearing out the queue in advance of the platoon, more of the green time is used by queued vehicles and less is available to the moving platoon.

The preceding discussion assumes that the queue is known at each signal. In fact, this is not an easy number to know. However, if we know that there is a queue and know its approximate size, the link offset can be set better than by pretending that no queue exists.

Consider the sources of the queued vehicles:

- Vehicles turning in from upstream side streets during their green (which is main-street red).
- Vehicles leaving parking garages or spaces.
- Stragglers from previous platoons.

There can be great cycle-to-cycle variation in the actual queue size, although the average queue size may be estimated. Even at that, queue estimation is a difficult and expensive task. Even the act of adjusting the offsets can influence the queue size. For instance, the arrival pattern of the vehicles from the side streets may be altered. Queue estimation is therefore a significant task in practical terms.

21.5 Signal Progression for Two-Way Streets and Networks

The task of progressing traffic on a one-way street has been shown to be relatively straightforward. To highlight the essence of the problem on a two-way street, assume that the arterial shown in Figure 21.5 is a two-way street rather than a one-way street. Figure 21.11 shows the trajectory of a *southbound* vehicle on this arterial. The first vehicle is just fortunate enough not to be stopped until Signal 2, but is then stopped again for Signal 1, for a total of two stops and 40 seconds of delay. There is no bandwidth, meaning that it is not possible to have a vehicle platoon pass along the arterial nonstop.

Figure 21.11: Case Study: The Southbound Result of a Northbound Progression

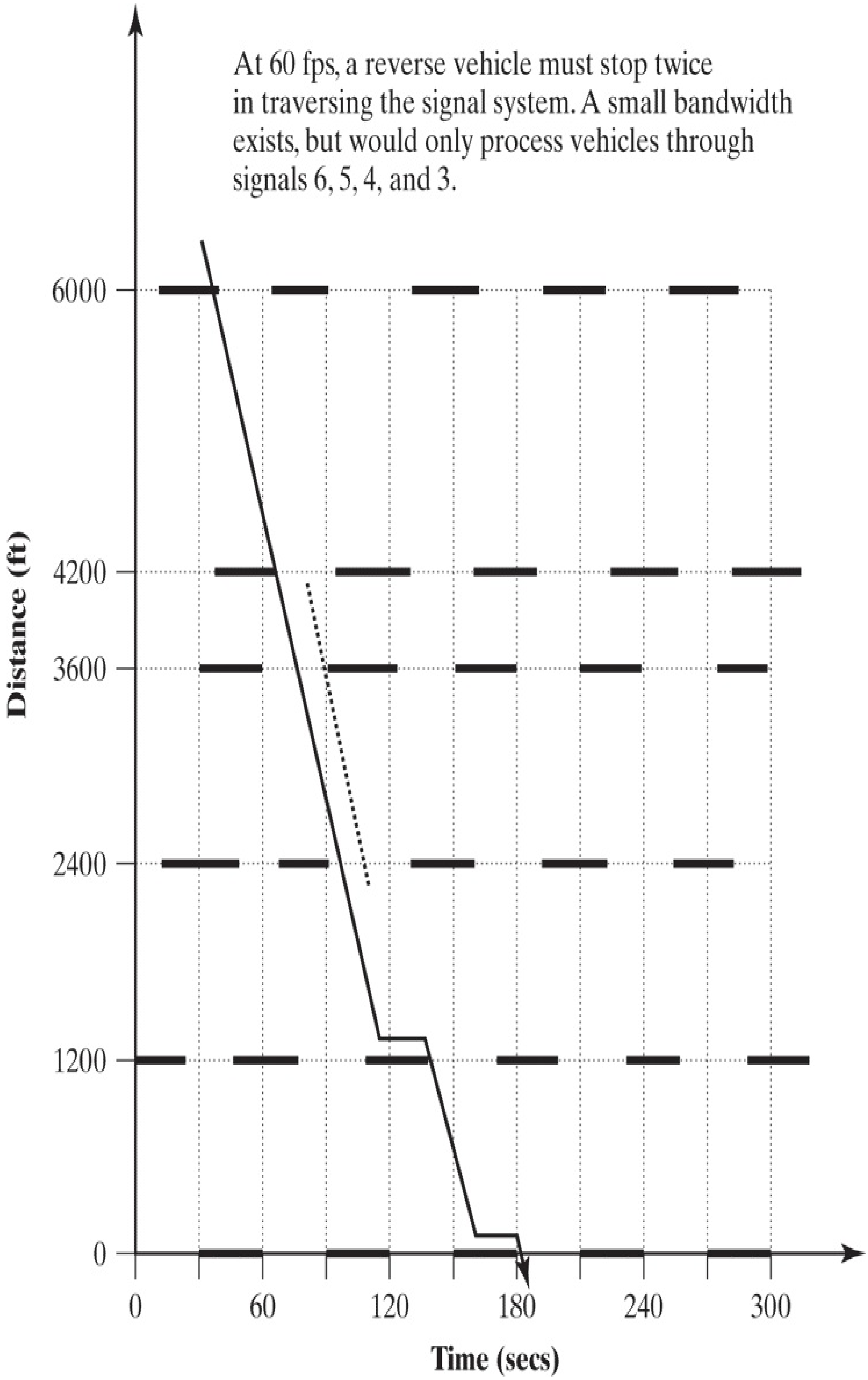

At 60 fps, a reverse vehicle must stop twice in traversing the signal system. A small bandwidth exists, but would only process vehicles through signals 6, 5, 4, and 3.
Distance (ft)
6000
4200
3600
2400
1200
0
0
60
120
180
240
300
Time (secs)

Figure 21.11: Full Alternative Text

Of course, if the offsets or the travel times had been different, it might have been possible to have a southbound bandwidth through all six signals.

Figure 21.12 illustrates the bandwidths for another example signal-timing plan. The northbound efficiency can be estimated as $(17/60)\times100\%=28.3\%$, and the northbound bandwidth capacity is $17/2\times1/60\times3{,}600=510$ veh/h/ln (assuming a 2 second headway). The southbound bandwidth is obviously terrible—there is no bandwidth through the defined system. The northbound efficiency is only 28.3%. This system is badly in need of retiming, at least on the basis of the bandwidth objective. Just looking at the time-space diagram, one might imagine sliding the pattern at Signal 4 to the right and the pattern at Signal 1 to the left, allowing some coordination for the southbound vehicles. However, if the per lane northbound demand volume is equal to or less than 510 veh/h/ln and if the flows are well organized (and if there is no internal queue development), the system will operate well in the northbound direction, even though better timing plans might be obtained.

Figure 21.12: Bandwidths on a Time-Space Diagram

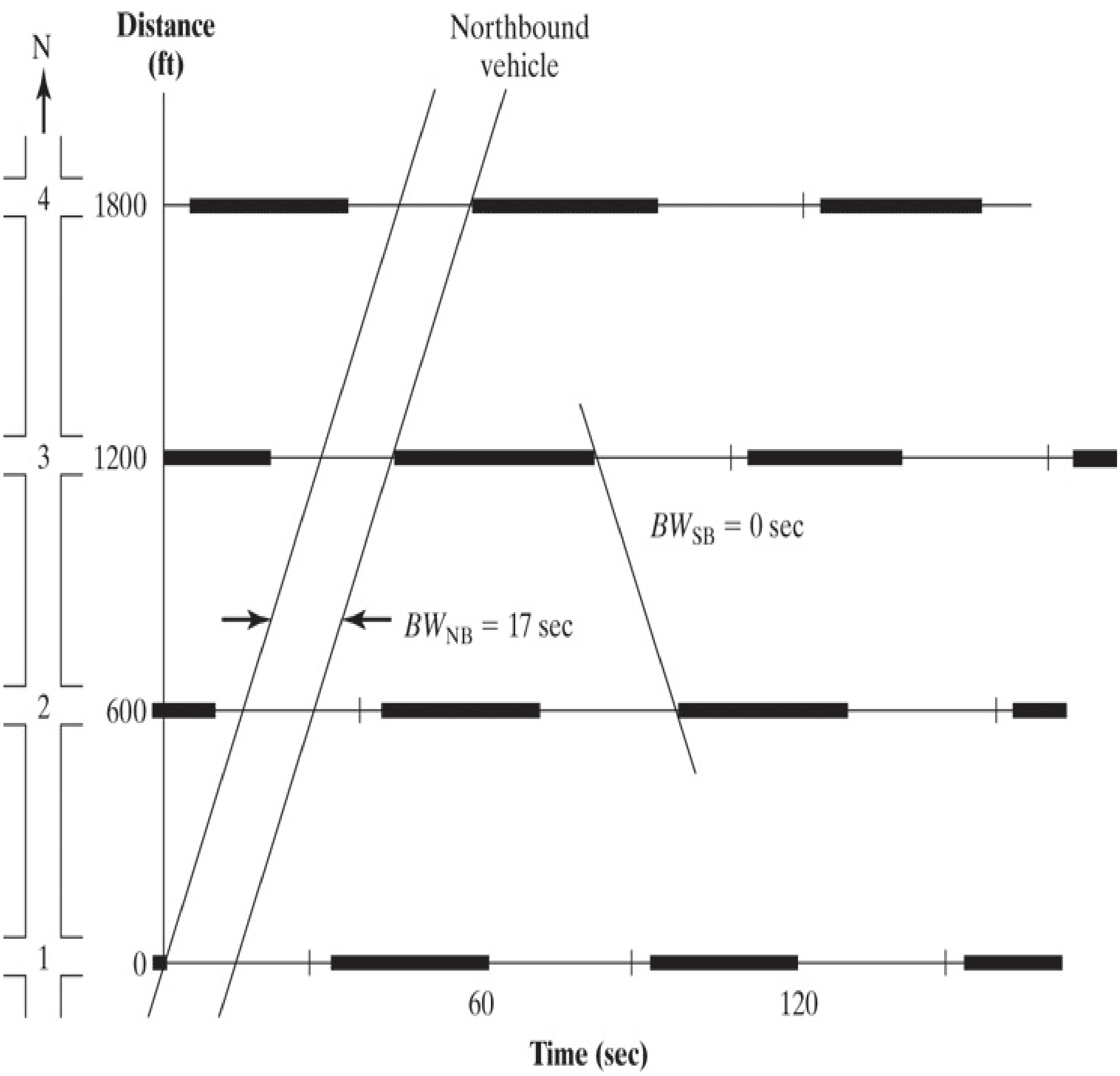


Figure 21.12: Full Alternative Text

The bandwidth concept is very popular in traffic engineering practice because the windows of green are easy visual images for both working professionals and public presentations. The most significant shortcoming of designing offset plans to maximize bandwidths is that internal queues are often overlooked in the bandwidth approach.

There are computer programs for determining offsets that look for maximum bandwidth solutions that go beyond the historical formulations, such as PASSER [2] and Tru-Traffic [3]. Other optimization programs search for offsets that minimize delays and stops, such Transyt-7F [4] and Synchro [5].

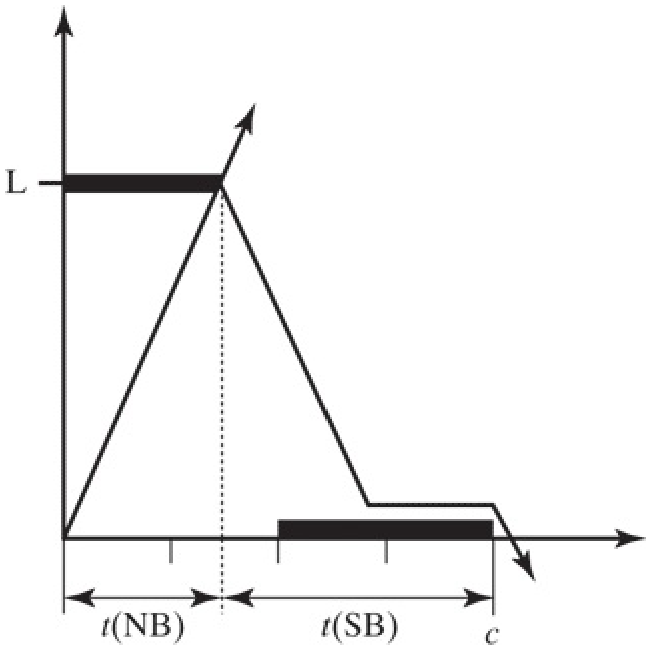
21.5.1 Offsets on a Two-Way

Street

In the case study shown in Figure 21.11 (which was timed for a perfect progression in the northbound direction, but ignored the southbound direction) if any offset were changed to accommodate the southbound vehicles, then the northbound bandwidth would suffer. For instance, if the northbound offset at Signal 2 were decreased by 20 seconds, then the pattern at that signal would shift to the left by 20 seconds, resulting in a “window” of green of only 10 seconds on the northbound, rather than the 30 seconds in the original display of Figure 21.5.

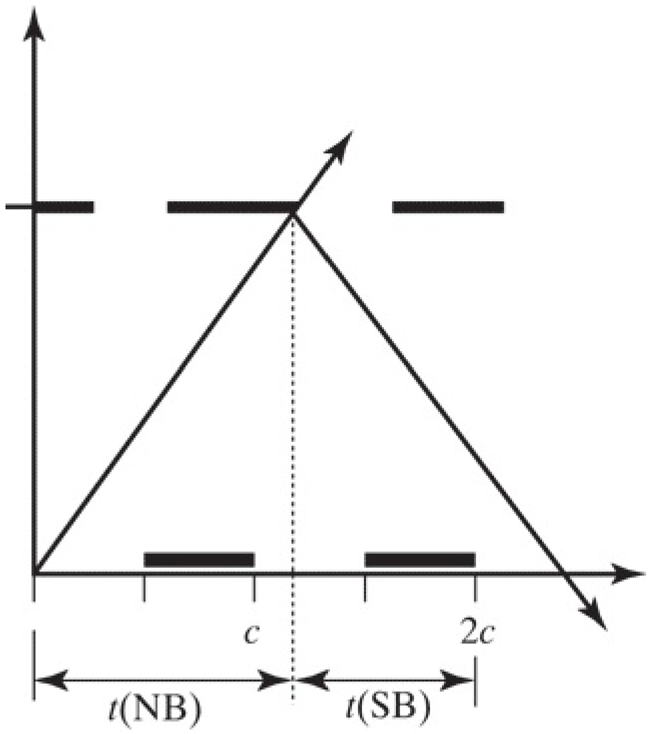
The fact that the offsets on a two-way street are interrelated presents one of the most fundamental problems of signal optimization. Note that inspection of a typical time-space diagram yields the obvious conclusion that the offsets in two directions add to one cycle length, shown in Figure 21.13(a). However, for longer blocks, the offsets might add to two (or more) cycle lengths, shown in Figure 21.13(b).

Figure 21.13: Offsets on a Two-Way Street Are Not Independent



(a) Offsets Add to One Cycle Length

21.5-3 Full Alternative Text



(b) Offsets Add to Two Cycle Lengths

21.5-3 Full Alternative Text

Figure 21.13 illustrates both actual offsets and travel times, which are not necessarily the same. While the engineer might desire the ideal offset to be the same as the travel times, this is not always the case. Once the offset is specified in one direction, it is automatically set in the other. The general expression for the two offsets in a link on a two-way street can be written as:

$t_{1i} + t_{2i} = nC$ [21-5]

where:

t_{1i}=offset in direction 1 (link i), s,t_{2i}=offset in direction 2 (link i), s,n=integ

To have n=1 (Figure 21.13(a)), $t_{1i} \leq C$; to have n=2 (Figure 21.13(b)), $C < t_{1i} \leq 2C$.

Any actual offset can be expressed as the desired "ideal" offset, plus an "error" or "discrepancy" term:

$t_{actual(i,j)} = t_{ideal(i,j)} + e_{ij}$ [21-6]

where *j* represents the direction and *i* represents the link. In a number of signal optimization programs that are used for two-way arterials, the objective is to minimize some function of the discrepancies between the actual and ideal offsets.

21.5.2 Network Closure

The relative difficulty of finding progressions on a two-way street, as compared with a one-way street, might lead one to conclude that the best approach is to establish a system of one-way streets to avoid the problem. A one-way street system has a number of advantages, not the least of which is elimination of left turns against opposing traffic. One-way streets simplify network signalization, but they do not eliminate closure problems, and they carry other practical disadvantages, such as longer trip lengths.

Figure 21.14 illustrates network closure requirements. In any set of four signals, offsets may be set on three legs in one direction. Setting three offsets, however, fixes the timing of all four signals. Thus, setting three offsets fixes the fourth.

Figure 21.14: Network Closure Illustrated

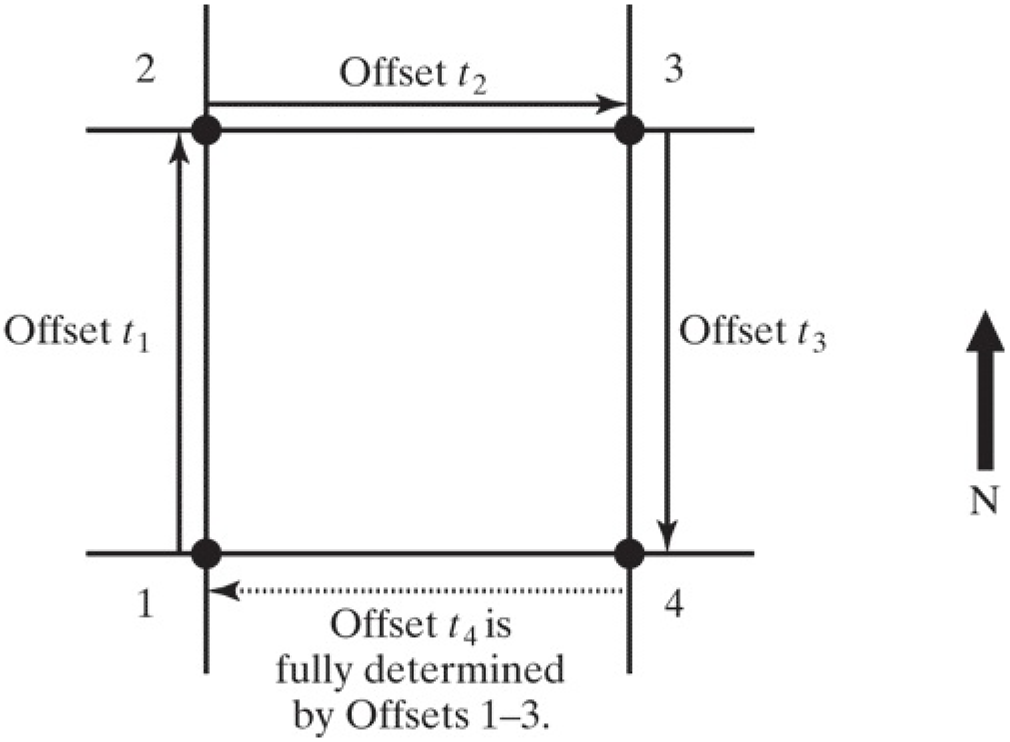


Figure 21.14: Full Alternative Text

Figure 21.15 extends this to a grid of one-way streets, in which offsets on all of the north-south streets are independently specified. The specification of one east-west street then “locks in” all other east-west offsets. Note that the key feature is that an open tree of one-way links can be completely independently set and that it is the closing or “closure” of the open tree that presents constraints on some of the links.

Figure 21.15: Impact of Closure on a Rectangular Street Grid

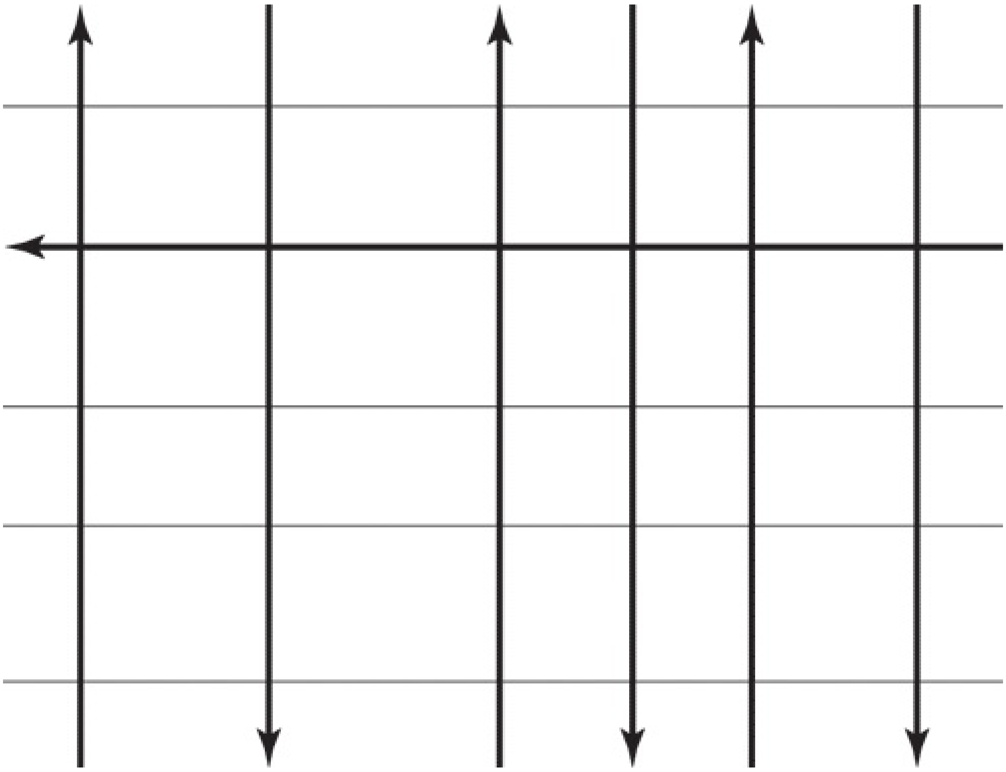


Figure 21.15: Full Alternative Text

To develop the constraint equation, refer to Figure 21.14 and walk through the following steps, keying to the green in all steps:

1. Step 1. Begin at Intersection 1 and consider the green initiation to be time t=0.

2. Step 2. Move to Intersection 2, noting that the link offset t1 specifies the time of green initiation at this intersection, relative to its upstream neighbor. Thus, green starts at Intersection 2 facing northbound at:

   t=0 + t1.

3. Step 3. Recognizing that the westbound vehicles get released after the N–S green is finished, green begins at Intersection 2 facing west at:

   t=0 + t1 + gNS,2

4. Step 4. Moving to Intersection 3, the link offset specifies the time of green initiation at Intersection 3 relative to

Intersection 2. Thus, the green begins at Intersection 3, facing west at:

$$t=0 + t_1 + g_{NS,2} + t_2$$

5. Step 5. Similar to Step 3, the green begins at Intersection 3, but facing south, after the E–W green is finished at time:

   $$t=0 + t_1 + g_{NS,2} + t_2 + g_{EW,3}$$

6. Step 6. Moving to Intersection 4, the green begins in the southbound direction after the offset t_3 is added:

   $$t=0+t_1+g_{NS,2}+t_2+g_{EW,3}+t_3$$

7. Step 7. Turning at Intersection 4, it is the NS green that is added to be at the start of green facing east:

   $$t=0+t_1+g_{NS,2}+t_2+g_{EW,3}+t_3+g_{NS,4}$$

8. Step 8. Moving to Intersection 1, it is t_4 that is relevant to be at the start of green facing east:

   $$t=0+t_1+g_{NS,2}+t_2+g_{EW,3}+t_3+g_{NS,4}+t_4$$

9. Step 9. Turning at Intersection 1, green will begin in the north direction after the EW green finishes:

   $$t=0+t_1+g_{NS,2}+t_2+g_{EW,3}+t_3+g_{NS,4}+t_4+g_{EW,1}$$

   This will bring us back to where we started. Thus, this is either $t=0$ or a multiple of the cycle length.

The following relationship results:

$$nC=0 + t_1 + g_{NS,2} + t_2 + g_{EW,3} + t_3+ g_{NS,4} + t_4 + g_{EW,1} \quad [21\text{-}7]$$

where the only caution is that the *g* values should include the change and clearance intervals.

The interrelationships stated in Equation 21-7 are constraints on freely setting all offsets. In these equations, one can trade off between green allocations and offsets. To get a better offset in Link 4, one can adjust the splits as well as the other offsets.

While it is sometimes necessary to consider networks in their entirety, it is common traffic engineering practice to decompose networks into noninterlocking arterials whenever possible. Figure 21.16 illustrates this process.

Figure 21.16: Decomposing a Network into Noninterlocking Arterial Segments

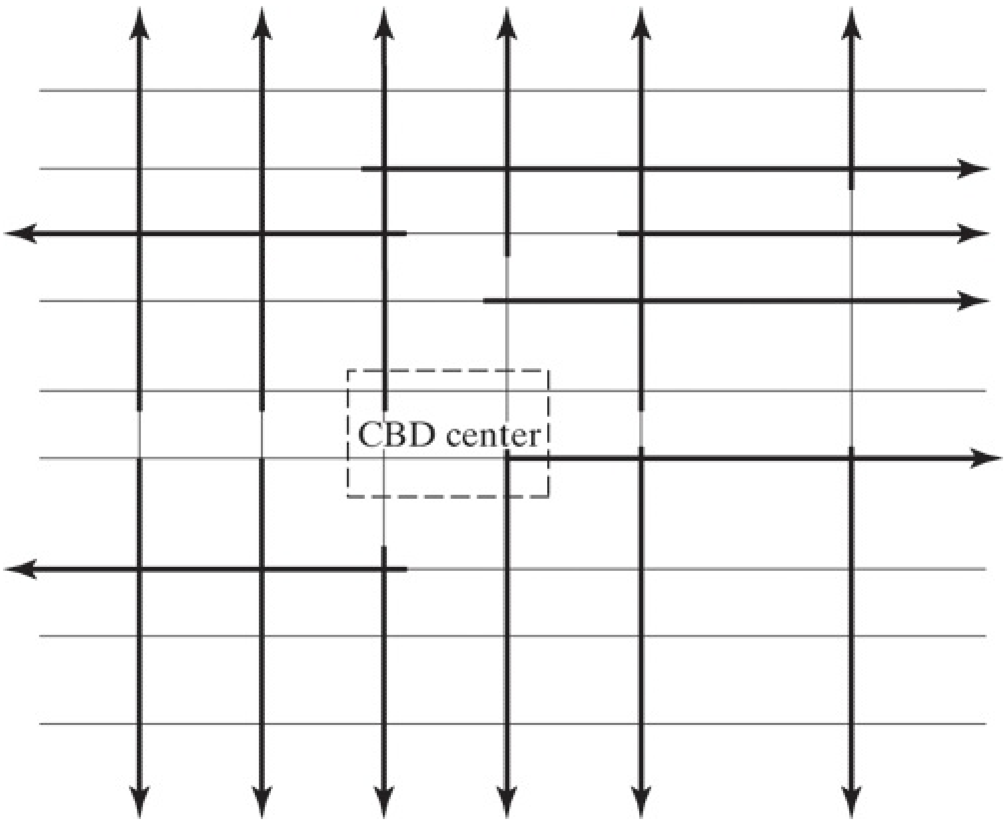


Figure 21.16: Full Alternative Text

Decomposition works well where a clear center of activity can be identified and where few vehicles are expected to pass through the center without stopping (or starting) at or near the center. As the discontinuity in all of the progressions occur in and directly around the identified center, large volumes passing through can create significant problems in such a scheme.

In summary, if offsets are set in one direction on a two-way street, then the reverse direction is fixed. In a network, you can set any “open tree” of links, but links that close the tree already have their offsets specified.

21.5.3 Finding Compromise Solutions

The engineer often wishes to design for maximum bandwidth in one direction, subject to some relation between bandwidths in the two directions. Sometimes, one direction is completely ignored. Much more commonly, the bandwidths in the two directions are designed to be in the same ratio as the flows in the two directions.

As mentioned earlier, there are computer programs that do the computations for maximum bandwidth that are commonly used by traffic engineers. Thus, it is not worthwhile to present an elaborate manual technique herein. However, to get a feel for the basic technique and trade-offs, a small “by-hand” example will be shown.

Refer to Figure 21.17, which shows four signals and decent progression in both directions. For purposes of illustration, assume it is given that a signal with 50:50 split must be located midway between Intersections 2 and 3. Figure 21.18 shows the possible effect of inserting the new signal into the system. It would appear that there is no way to include this signal without destroying one or the other bandwidth, or cutting both in half.

Figure 21.17: Case Study: Four Intersections with Good Two-Way Progression

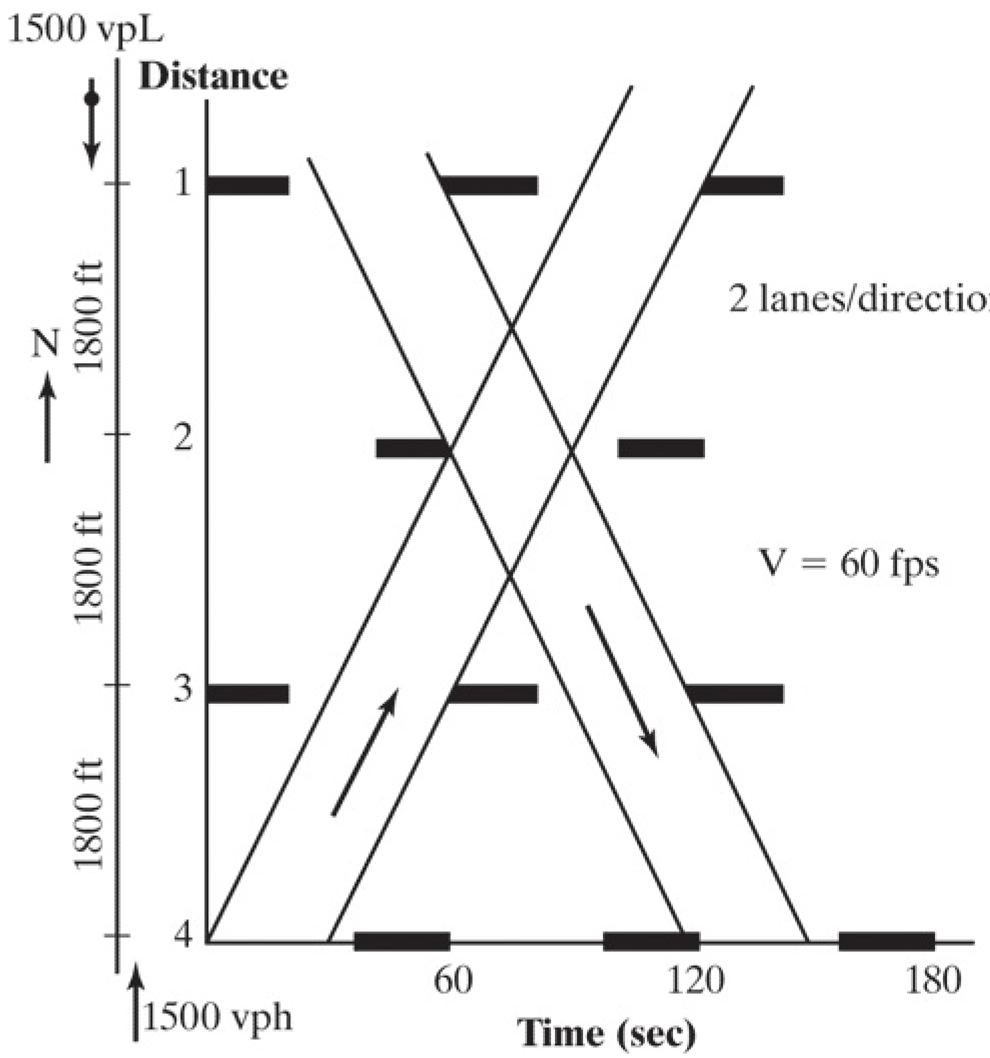


Figure 21.17: Full Alternative Text

Figure 21.18: Case Study: The Effect of Inserting a New Signal into the System

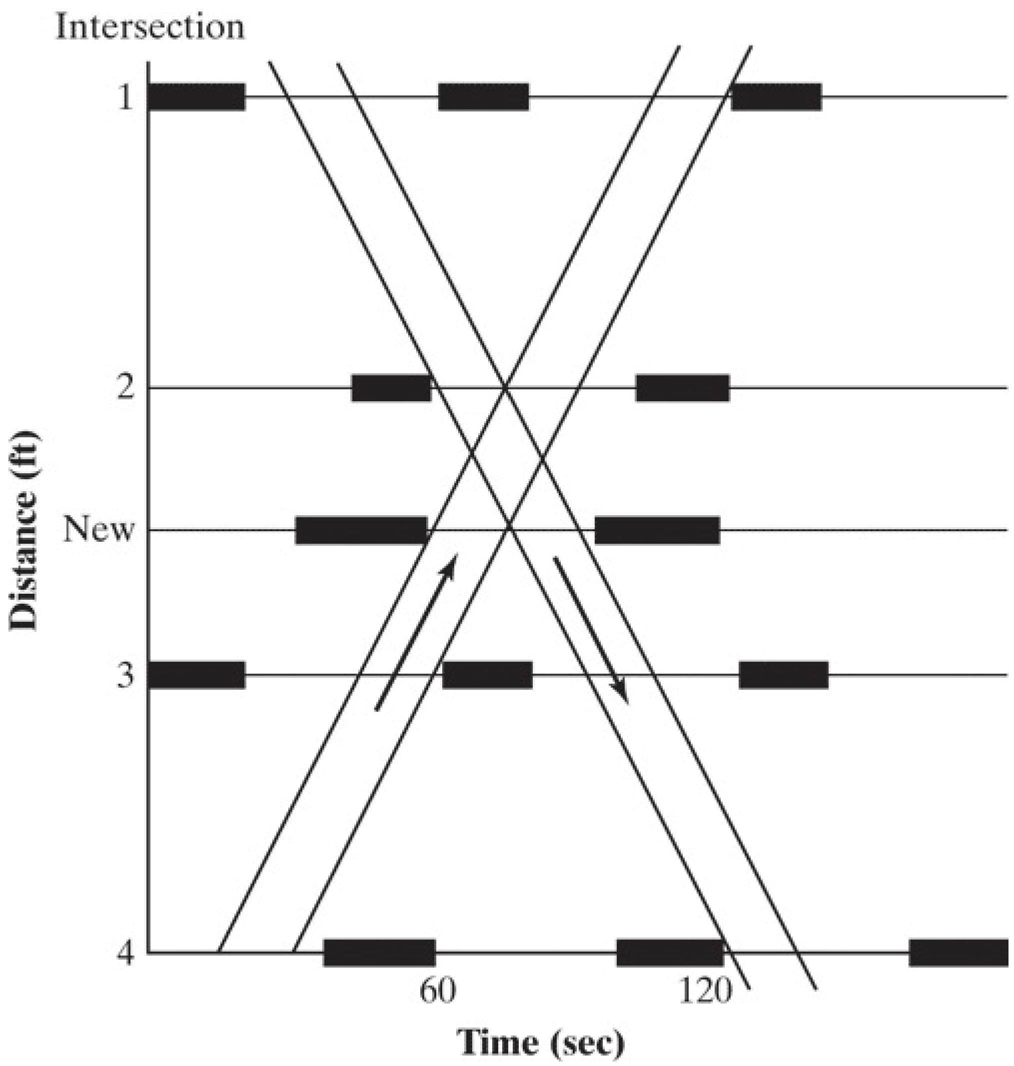


Figure 21.18: Full Alternative Text

To solve this problem, the engineer must move the offsets around until a more satisfactory timing plan develops. A change in cycle length may even be required.

Note that in Figure 21.17, the northbound vehicle takes 3,600 ft/60 ft/s = 60 s to travel from Intersection 4 to Intersection 2, or—given C=60s—one cycle length. If the cycle length had been C=120s, the vehicle would have arrived at Intersection 2 at $C/2$, or one-half the cycle length. If we try the 120-second cycle length, a solution presents itself.

Figure 21.19 shows one solution to the problem, for C=120s, which has a 40-second bandwidth in both directions for an efficiency of 33%. The 40-second bandwidth can handle (40/2.0)=20 vehicles per lane per cycle. Thus, if the demand volume is greater than 3,600(40)(2)/(2.0 × 120)=1,200veh/h, it will not be possible to process the vehicles nonstop through the system.

Figure 21.19: Case Study: Solution with C = 120 s

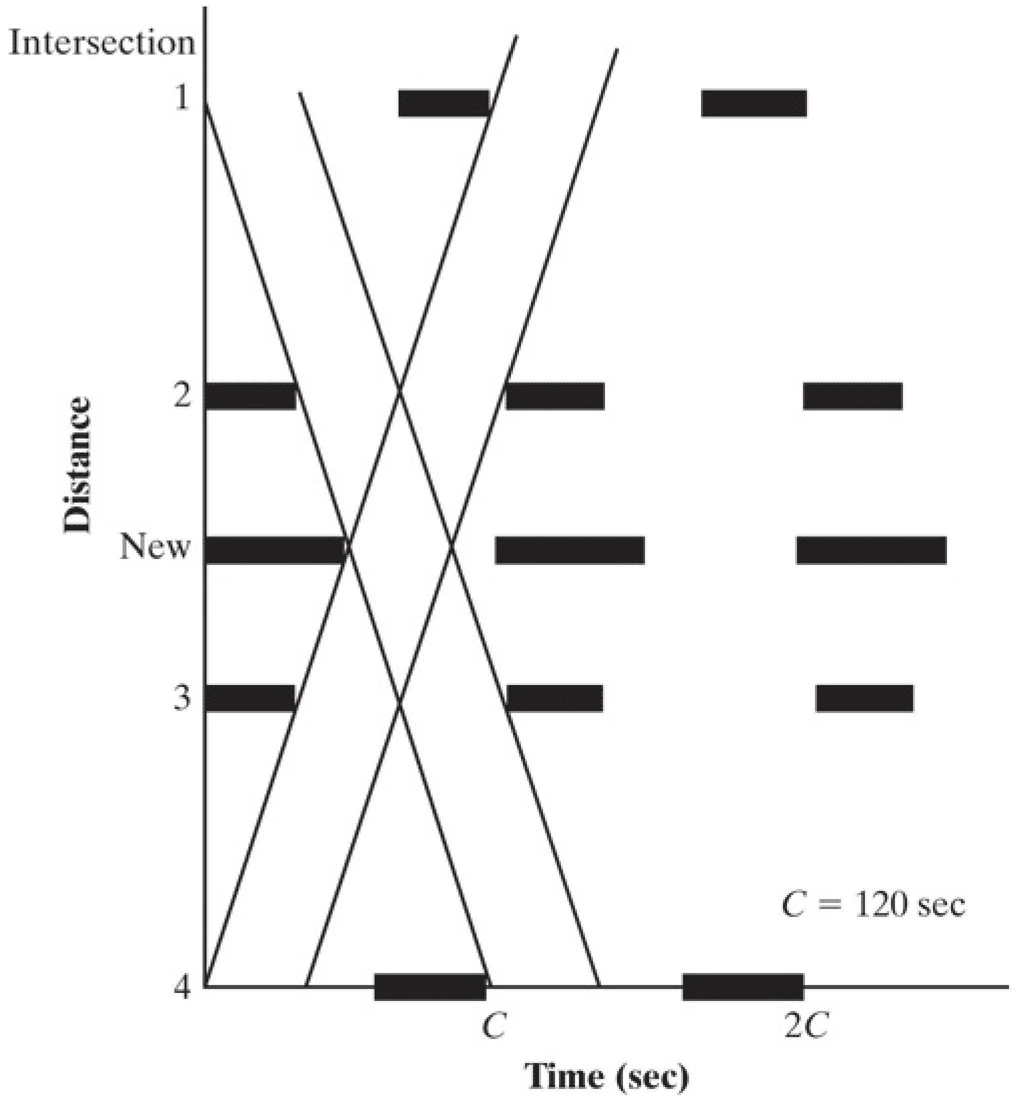


Figure 21.19: Full Alternative Text

As indicated in the original information (see Figure 21.17), the northbound demand is 1,500 veh/h. Thus, there will be some difficulty in the form of excess vehicles in the platoon. They can enter the system but cannot pass Signal 2 nonstop. They will be "chopped off" the end of the platoon and will be queued vehicles in the next cycle. They will be released in the early part of the cycle and arrive at Signal 1 at the beginning of red. Figure

21.20 illustrates this, showing that these vehicles then disturb the next northbound through platoon.

Figure 21.20: Case Study: Effect of Platoons with Demand Volume 1,500 veh/h

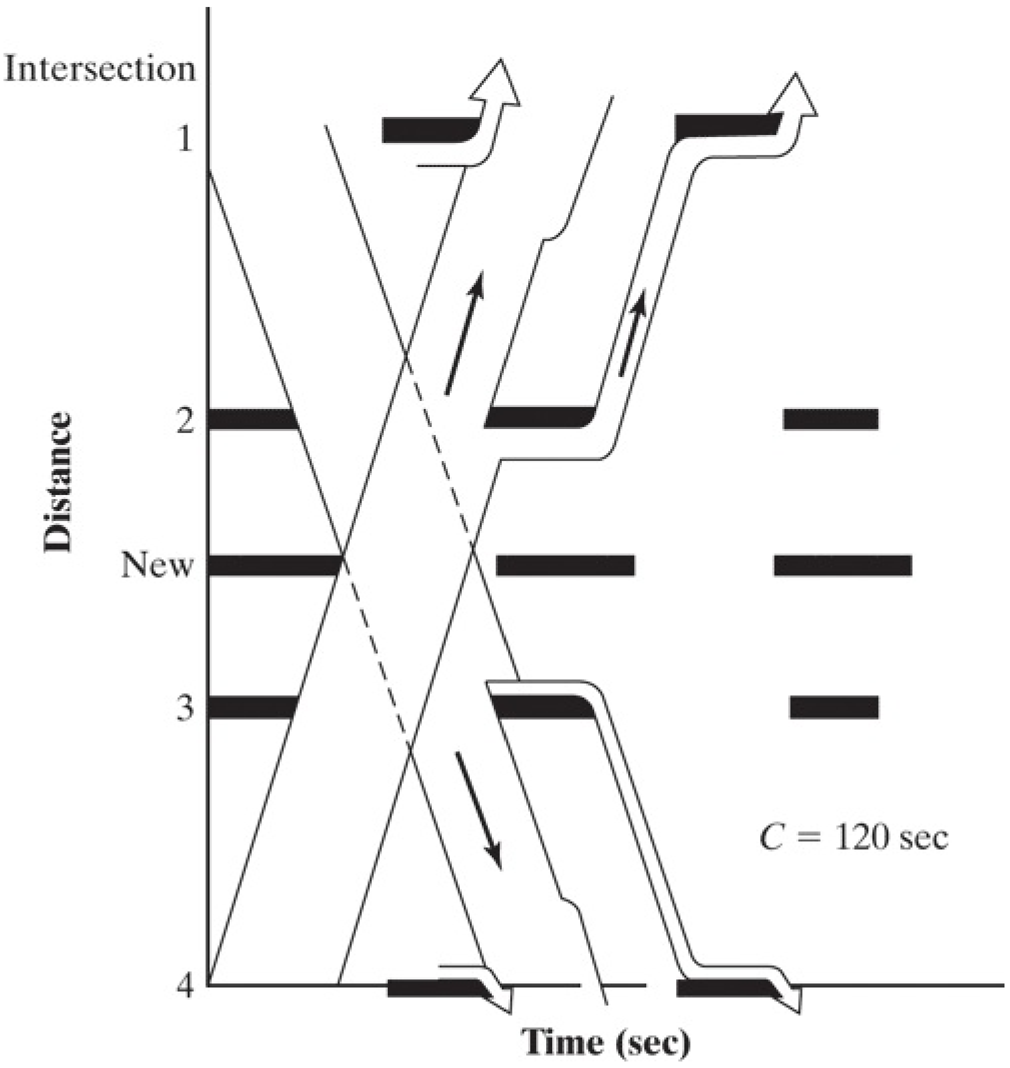


Figure 21.20: Full Alternative Text

Figure 21.20 illustrates the limitation of the bandwidth approach when internal queuing arises, disrupting the bandwidth. The figure also shows the southbound platoon pattern, suggesting that a demand of exactly 1,200 veh/h might give rise to minor problems of the same sort at Signals 3 and

4.

If one were to continue a trial-and-error attempt at a good solution, it should be noted that:

- If the green initiation at Intersection 1 comes earlier in order to help the main northbound platoon avoid the queued vehicles, the southbound platoon is released sooner and gets stopped or disrupted at Intersection 2.
- Likewise, shifting the green at Intersection 2 cannot help the northbound progression without harming the southbound progression.
- Shifting the green at Intersection 3 cannot help the southbound progression without harming the northbound progression.
- Some green can be taken from the side street and given to the main street.

This illustration showed insights that can be gained by simple inspection of a time-space diagram, using the concepts of bandwidth, efficiency, and an upper bound on demand volume that can be handled nonstop.

21.6 Types of Progression

21.6.1 Progression Terminology

A *simple progression* is the name given to the progression in which all signals are set so that a vehicle released from the first intersection will arrive at all downstream intersections just as the signals at those intersections initiate green. That is, each offset is the ideal offset, set by Equation 21-4 with zero queue. Of necessity, simple progressions are effective only on one-way streets or on two-way streets on which the reverse flow is small or neglected.

Because the simple progression results in a green wave that advances with the vehicles, it is often called a *forward progression*, taking its name from the visual image of the advance of the green down the street.

It may happen that the simple progression is revised two or more times in a day so as to conform to the direction of the major flow or to the flow level (since the desired platoon speed can vary with traffic demand). In this case, the scheme may be referred to as a *flexible progression*.

Under certain circumstances, the internal queues are sufficiently large that the ideal offset is negative; that is, the downstream signal must turn green before the upstream signal to allow sufficient time for the queue to start moving before the arrival of the platoon. Figure 21.21 has link lengths of 600 ft, platoon speeds of 60 ft/s, and internal queues averaging seven vehicles per lane at each intersection. The visual image of such a pattern is of the green marching upstream toward the drivers in the platoon. Thus, it is referred to as a *reverse progression*. Figure 21.21 also illustrates one of the unfortunate realities of so many internal queued vehicles: The platoon’s lead vehicle only gets to Signal 4 before encountering a red indication. As the platoon passes Signal 3, there are only 12 seconds of green to accommodate it, resulting in all vehicles beyond the sixth (i.e., 12/2=6) being cut off at Signal 3.

Figure 21.21: Illustration of Reverse Progression

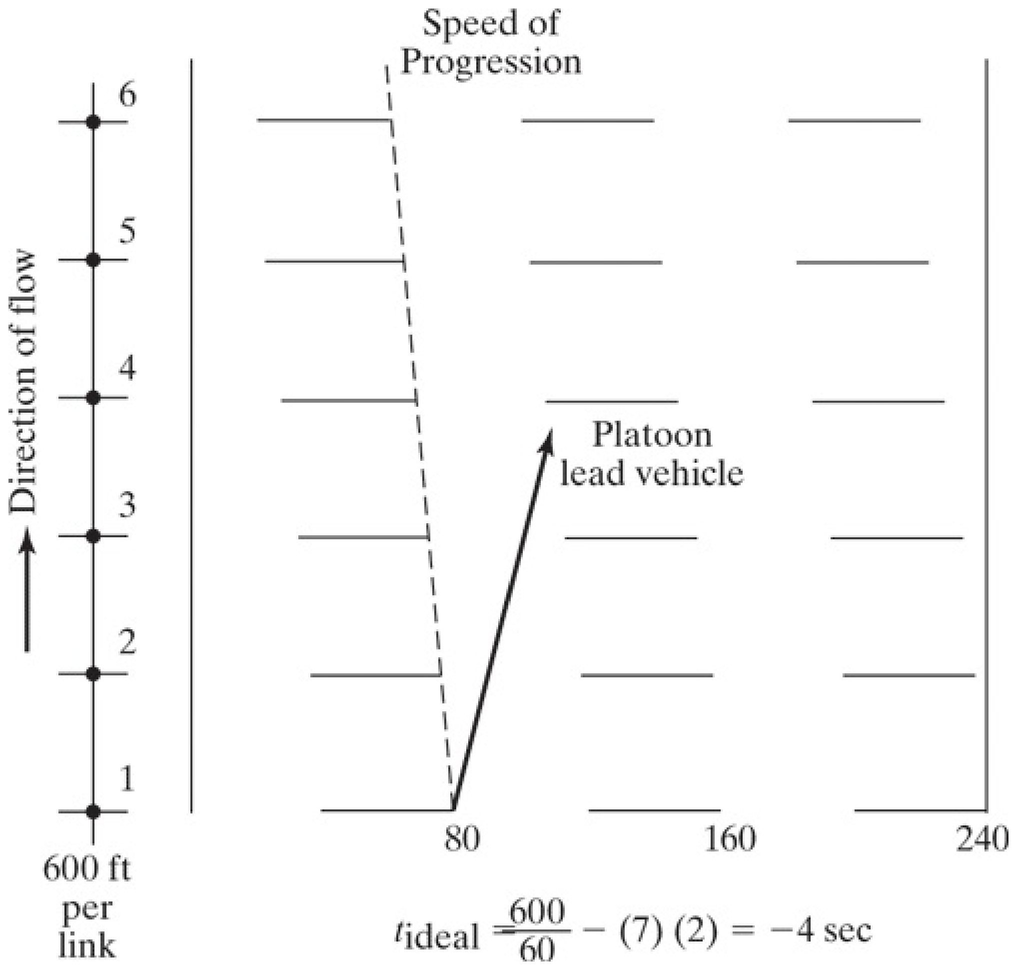


Figure 21.21: Full Alternative Text

In the next several sections, common progression systems that can work extremely effectively on two-way arterials and streets are presented. As will be seen, these systems rely on having uniform block lengths and an appropriate relationship among block length, progression speed, and cycle length.

21.6.2 The Alternating Progression

For certain uniform block lengths and all intersections with a 50:50 split of effective green time, it is possible to select a feasible cycle length such that:

$\frac{C}{2}=\frac{L}{S}$ [21-8]

where:

C=cycle length, s,L=block length, ft, andS=platoon speed, ft/s.

In this situation, the progression of Figure 21.22 can be obtained. There is no limit to the number of signals that may be included in the progression. The name for this pattern is derived from the “alternate” appearance of the signal displays: As the observer at Signal 1 looks downstream, the signals alternate—red, green, red, green, and so forth.

Figure 21.22: Alternate Progression

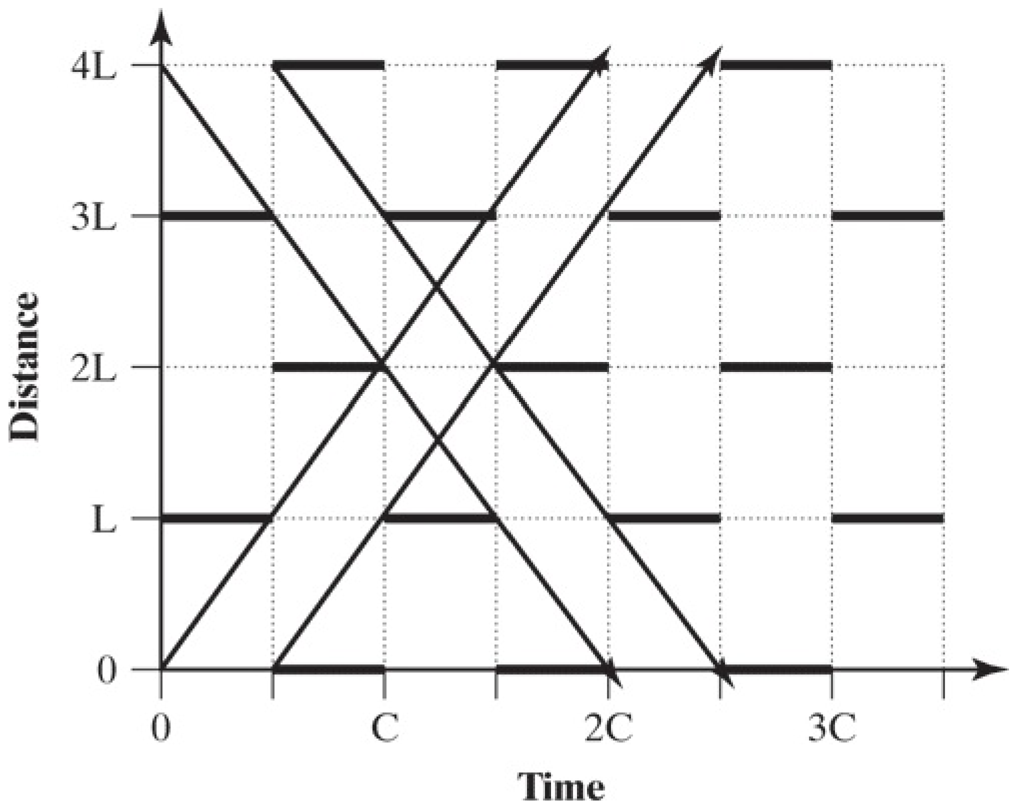


Figure 21.22: Full Alternative Text

The key to Equation 21-8 is that the ideal offset in either direction (with zero internal queues) is *L/S*. That is, the travel time to each platoon is exactly one-half the cycle length, so that the two travel times add up to the cycle length.

The efficiency of an alternate system is 50% in each direction, because all of the green is used in each direction. The bandwidth capacity for an alternating progression is found using Equation 21-3 and noting that the bandwidth, *BW*, is equal to one-half the cycle length, *C*. If a saturation headway of 2.0 s/veh is assumed, then:

$$CBW=3,600 \times BW \times Lh \times C=3,600 \times 0.5C \times L2.0 \times C=900L$$

where all terms are as previously defined. This is an approximation based on the assumed saturation headway of 2.0 s/veh.

Note that if the splits are not 50:50 at some signals, then (1) if the split favors the main street, it simply represents excess green, suited for accommodating miscellaneous vehicles, and (2) if the split favors the side street, the bandwidths are reduced.

21.6.3 The Double-Alternating Progression

For certain uniform block lengths with 50:50 splits, it is not possible to satisfy Equation 21-8, but it is possible to select a feasible cycle length such that:

$$C4=LS \quad [21\text{-}9]$$

In this situation, the progression illustrated in Figure 21.23 can be obtained.

Figure 21.23: Double Alternate Progression

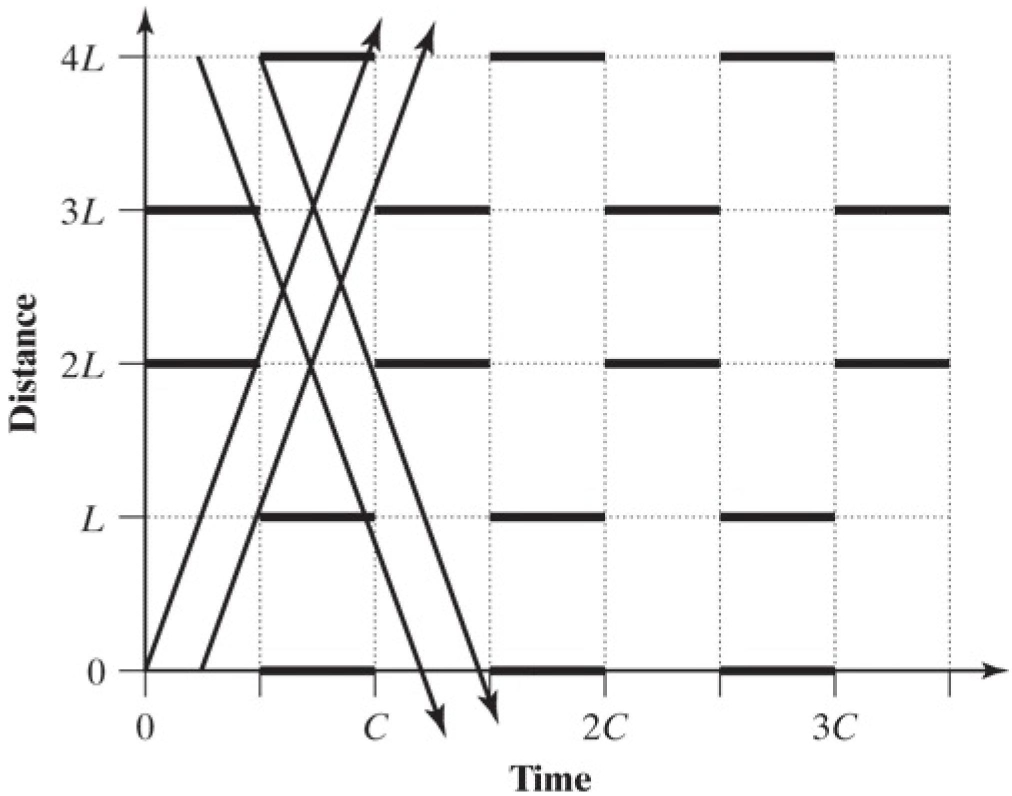


Figure 21.23: Full Alternative Text

The key is that the ideal offset in either direction (with zero internal queues) over *two* blocks is one-half of a cycle length, so that two such travel times (one in each direction) add up to a cycle length. There is no limit to the number of signals that can be involved in this system, just as there was no limit with the alternate system.

The name of the pattern is derived from the “double alternate” appearance of the signal displays—that is, as the observer at Signal 1 looks downstream, the signals alternate in pairs: green, green, red, red, green, green, red, red, and so forth.

The efficiency of the double alternate signal system is 25% in each direction, because only half of the green is used in each direction. The upper limit on the bandwidth capacity may be approximated by assuming a 2.0 s/veh saturation headway and noting that the *BW* is one-quarter of *C*.

As with the alternate system, if the splits are not 50:50 at some signals, then (1) if the split favors the main street, it simply represents excess green, suited for accommodating miscellaneous vehicles, and (2) if the split favors the side street, the bandwidths are reduced.

$CBW=3{,}600 \times BW \times Lh \times C=3{,}600 \times 0.25C \times L2.0 \times C=450L$

21.6.4 The Simultaneous Progression

For very closely spaced signals or for rather high vehicle speeds, it may be best to have all the signals turn green at the same time. This is called a simultaneous system, since all the signals turn green simultaneously. Figure 21.24 illustrates a simultaneous progression.

Figure 21.24: Simultaneous Progression

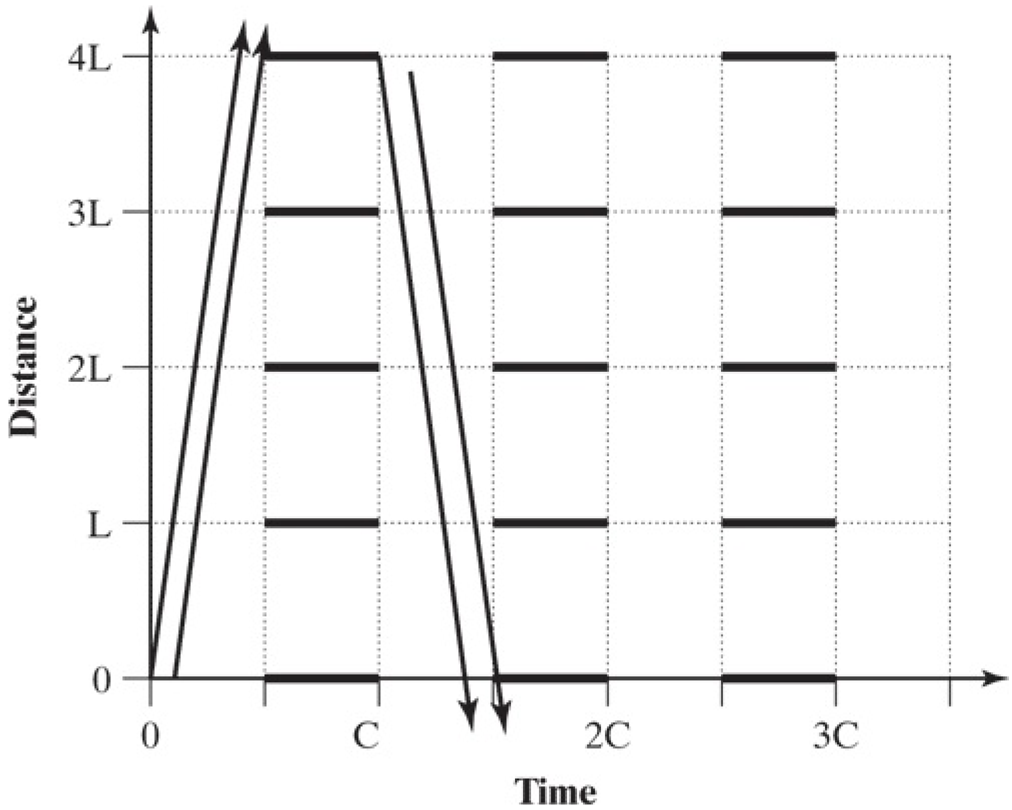


Figure 21.24: Full Alternative Text

The efficiency of a simultaneous system depends on the number of signals involved. For N signals:

$$EFF(\%)=[12-(N-1)\times LS\times C]\times 100\% \quad [21\text{-}10]$$

For four signals with $L=400\text{ft}$, $C=80\text{s}$, and $S=45\text{ft/s}$, the efficiency is 16.7%. For the same number of signals with $L=200\text{ft}$, it is 33.3%.

Simultaneous systems are advantageous only under a limited number of special circumstances. The foremost of these special circumstances is very short block lengths. The simultaneous system has an additional advantage, however, that is not at all clear from a bandwidth analysis: under very heavy flow conditions, it forestalls breakdown and spillback. This is so because (1) it allows for vehicle clearance time at the downstream intersection where queues inevitably exist during heavy flow, and (2) it cuts platoons off in a way that generally prevents blockage of intersections. This works to the advantage of cross traffic.

21.6.5 Insights Regarding the Importance of Signal Spacing and Cycle Length

It is now clear that:

- All progressions have their roots in the desire for ideal offsets.
- For certain combinations of cycle length, block length, and platoon speed, some very satisfactory two-way progressions can be implemented.
- Other progressions can be designed to suit individual cases, using the concept of ideal offset and queue clearance, trial-and-error bandwidth-based approaches, or computer-based algorithms.

A logical first step in approaching a system is simply to ride the system and inspect it. As you sit at one signal, do you see the downstream signal green but with no vehicles being processed? Do you arrive at signals that have standing queues but were not timed to get them moving before your platoon arrived? Do you arrive on the red at some signals? Is the flow in the other direction significant, or is the traffic really a one-way pattern, even if the streets are two way?

It is very useful to sketch out how much of the system can be thought of as an "open tree" of one-way links. This can be done with a local map and an appreciation of the traffic flow patterns. Referring to Figure 21.16, a distinction should be made among:

- Streets that are one way
- Streets that can be treated as one way, due to the actual or desired flow patterns
- Streets that must be treated as two way
- Larger grids in which streets (one-way and two-way) interact because they form unavoidable "closed trees" and are each important in that they cannot be ignored for the sake of establishing a "master grid" that is an open tree
- Smaller grids in which the issue is not coordination but rather local land access and circulation, so that they can be treated differently (downtown grids may well fall into the latter category, at least in some cases)

The next most important issue is the cycle length dictated by the signal spacing and platoon speed. Attention must focus on the combination of cycle length, block length, and platoon speed, as shown earlier in this chapter.

Figure 21.25 shows the three progressions of the preceding sections—alternating, double alternating, and simultaneous—on the same scale. The basic "message" is that as the average signal spacing decreases, the type of progression best suited to the task changes.

Figure 21.25: Comparison of Scales on Which Standard Patterns are Used

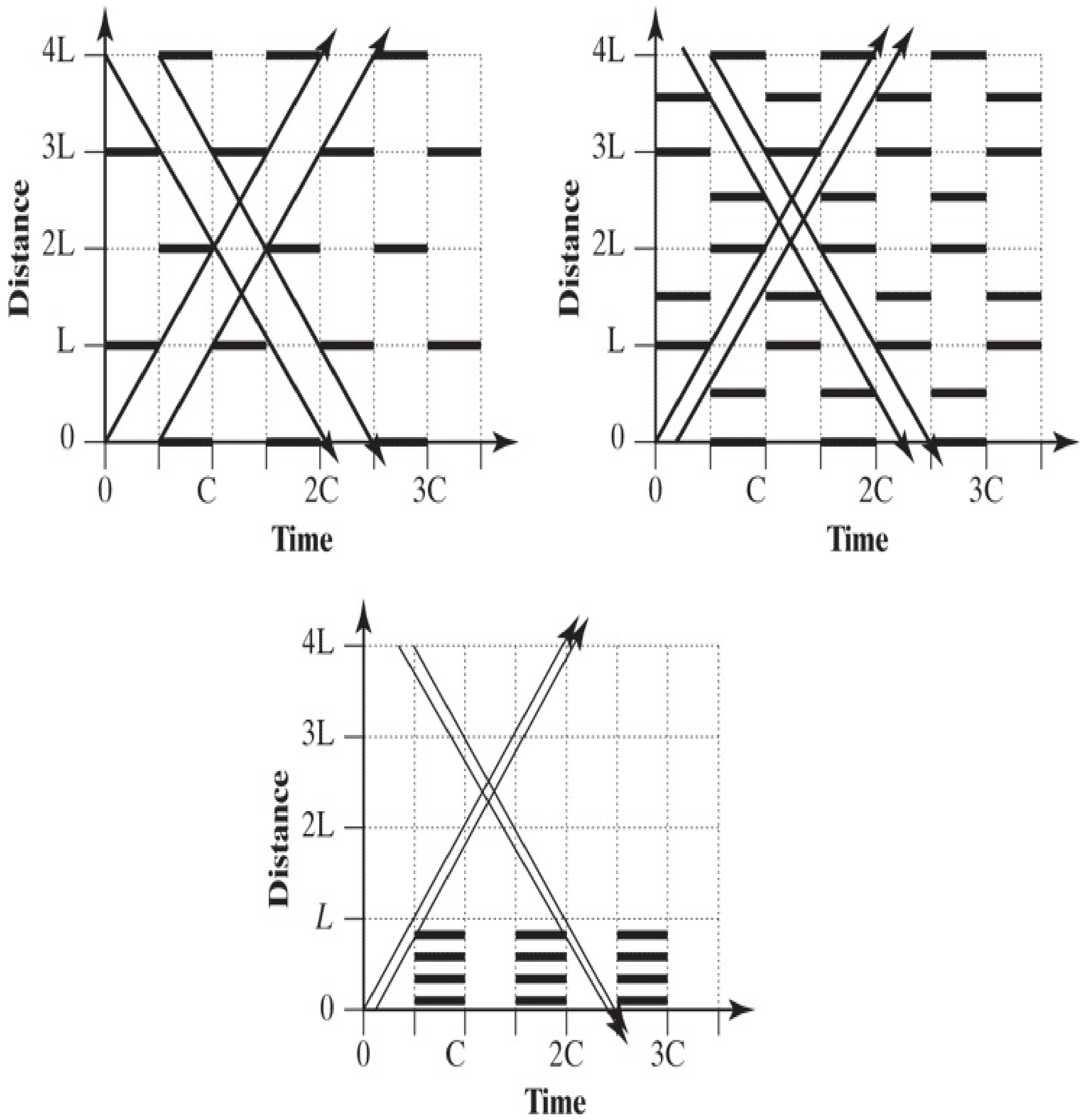


Figure 21.25: Full Alternative Text

Figure 21.26 illustrates a hypothetical arterial that comes from a low-density suburban environment with a larger signal spacing, into the outlying area of a city, and finally passes through one of the city's CBDs. As the arterial changes, the progression used may also be changed, to suit the dimensions.

Figure 21.26: Hypothetical Use of Several Patterns Along the Same Arterial

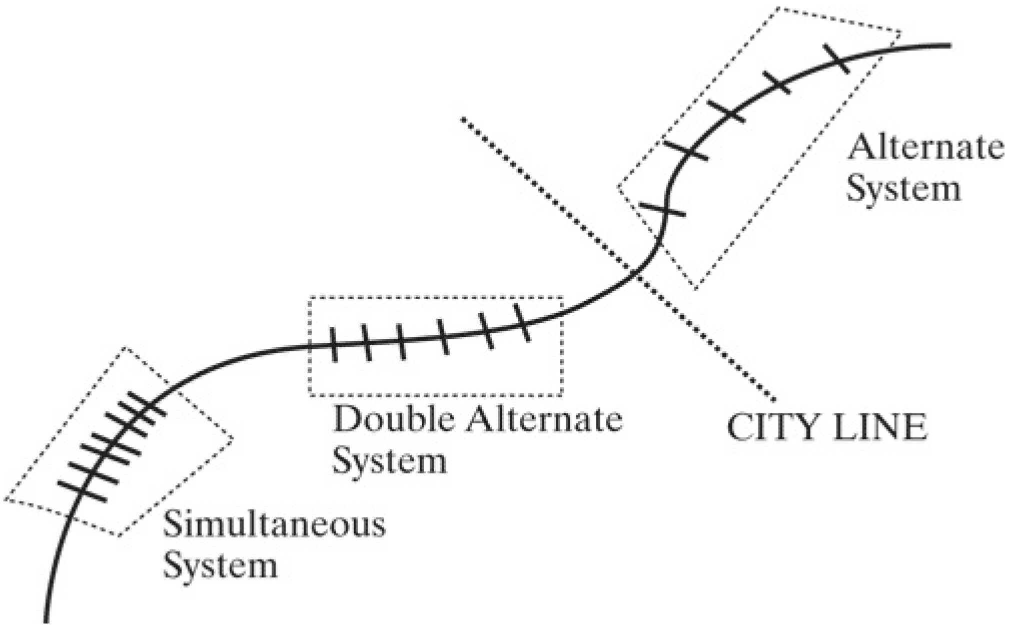


Figure 21.26: Full Alternative Text

Note that the basic lesson here is that a system can sometimes be best handled by breaking it up into several smaller systems. This can be done with good effect on even smaller systems, such as ten consecutive signals, of which a contiguous six are spaced uniformly and the other four also uniformly, but at different block lengths. Note that to the extent that block lengths do not exist perfectly uniformly, these plans can serve as a basis from which adaptations can be made. Note also that the suitability of the cycle length has been significant. It is often amazing how often the cycle length is poorly set for system purposes.

21.7 Software for Signal Progression Design

Various software packages are available for determining progressions. Two commonly used programs are illustrated: TruTraffic and Synchro. TruTraffic is based on the bandwidth approach. Bandwidth-based solutions are less data intensive, very suitable in many applications, and easy to manage, but do not take into account internal queuing. Synchro focuses on delay- or stop-optimized signal settings and thus take into account internal queuing.

21.7.1 TruTraffic

Bandwidth-based solutions are an important tool particularly for off-peak periods when demand is relatively small and/or when flow is highly directional. Bandwidth solutions can also be used to move platoons along relatively long arterials in a set of bandwidths, with the breaks between bandwidths occurring where it is logical or suitable to stop and re-form platoons—for instance, just upstream of a set of closely spaced signals, so that the platoons that might overflow the short block spacings are not stopped in that section of the arterial. Bandwidth solutions are also used effectively to discourage speeding, encourage adherence to the speed limit, and identify green that can be allocated to increased pedestrian walking times [9].

Figure 21.27 shows the TruTraffic output for the arterial addressed in Figure 21.11, a two-way arterial, on which we wish to achieve equal bandwidth (16 seconds) in the two directions if at all possible. Because of the way in which output is printed out (generally, on long paper), the directions are reversed from prior time-space diagrams. Arrows have been added to emphasize the directions in which time and distance increase. From the figure, note that the bandwidths are shown in seconds, and partial bandwidths are shown when breaks are required. In such cases, the message is that the platooned vehicles can only travel so many blocks without stopping.

Figure 21.27: A Solution for Equal Bandwidths in Both Directions for the Arterial in [Figure 21.11]

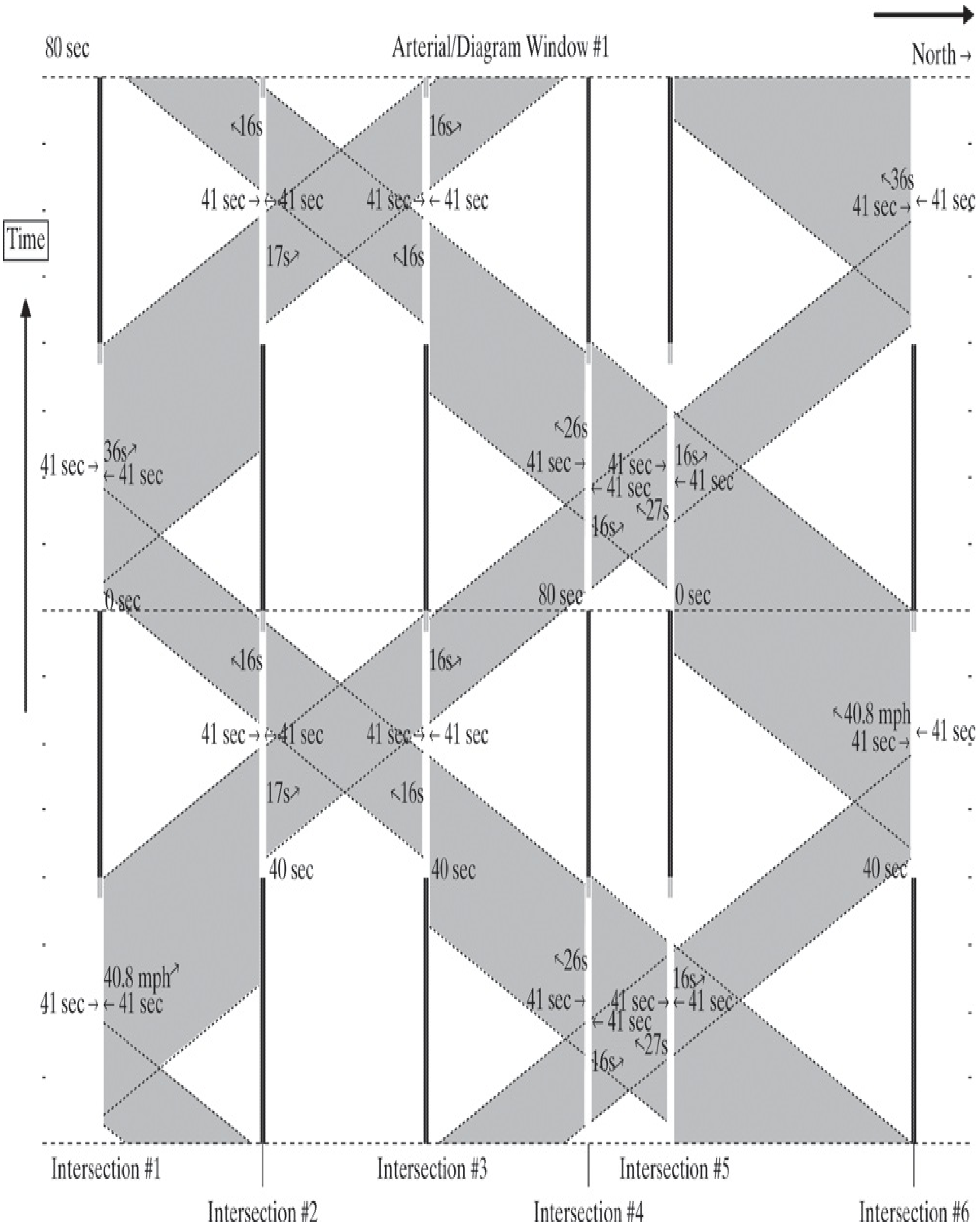
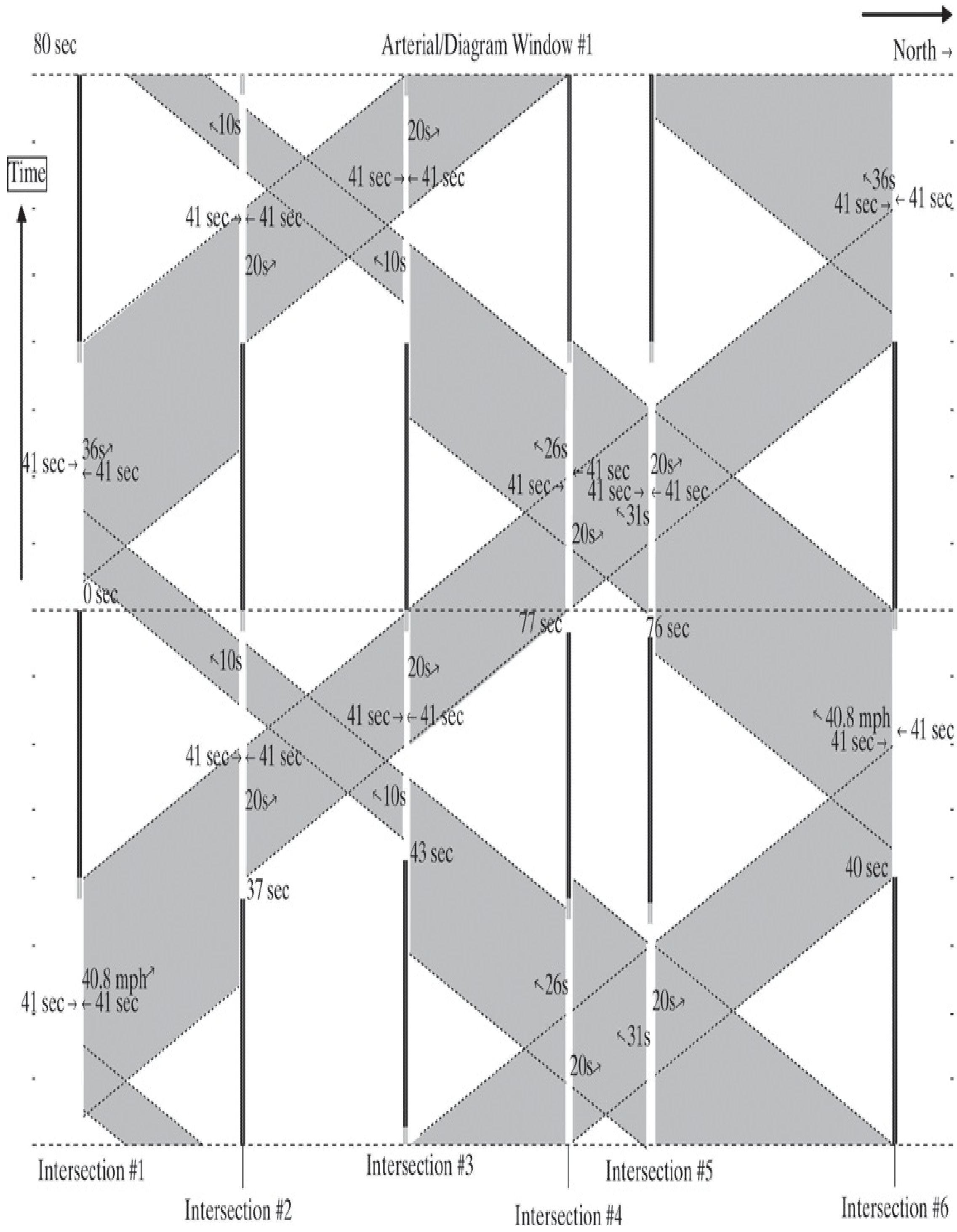


Figure 21.27: Full Alternative Text

Figure 21.28 shows the same arterial with the target to get a NB bandwidth of at least 20 seconds, with “as good as possible” in the SB direction. Notice that in order to use TruTraffic, details such as demand volumes, pedestrians, percent heavy vehicles, etc. are not needed.

Figure 21.28: A Solution for the Arterial in Figure 21.5 as a Two-Way Arterial



[Figure 21.28: Full Alternative Text](#)

21.7.2 Synchro

Synchro optimizes an arterial by trying to minimize stops and delay. Unlike TruTraffic, the details at each intersection are required inputs.

Consider the inputs as shown in Figure 21.29. The East/West street (Main Street) is the direction to be coordinated.

Figure 21.29: Inputs for an Illustrative Synchro Case

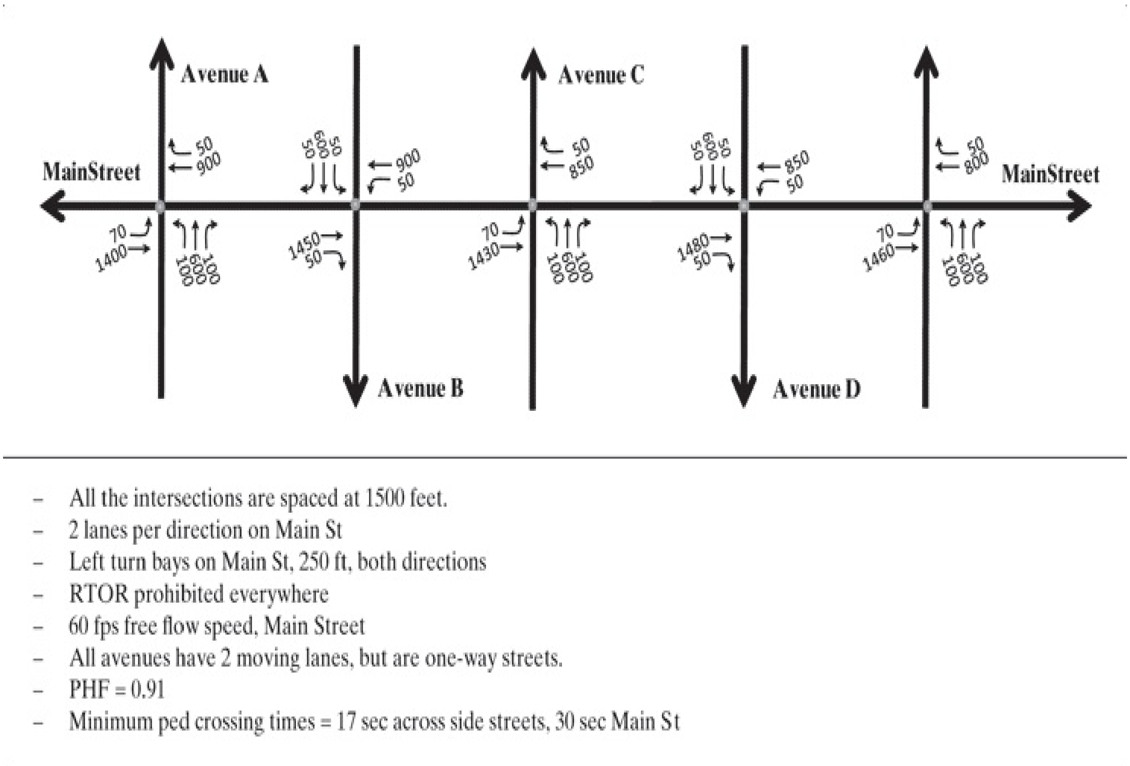


Figure 21.29: Full Alternative Text

Figure 21.30 shows an illustrative Synchro solution for a cycle length of 100 seconds. The EB direction (from top to bottom of the time space diagram) has a bandwidth going through all four intersections, with a queue only at the start of the arterial (intersection with Avenue A). In the westbound direction (from bottom to top of the time space diagram, there is no bandwidth through all four intersections. Vehicles in queue at Avenue E can move through Avenue D without stopping, but then have to stop again at Avenue C and then again at Avenue A. At Avenue A, the eastbound vehicles experience more delay than the westbound vehicles. Once the queue clears, eastbound vehicles can travel through the remaining intersections without stopping.

Figure 21.30: Sample Synchro Time-Space Diagram for Illustrative Case

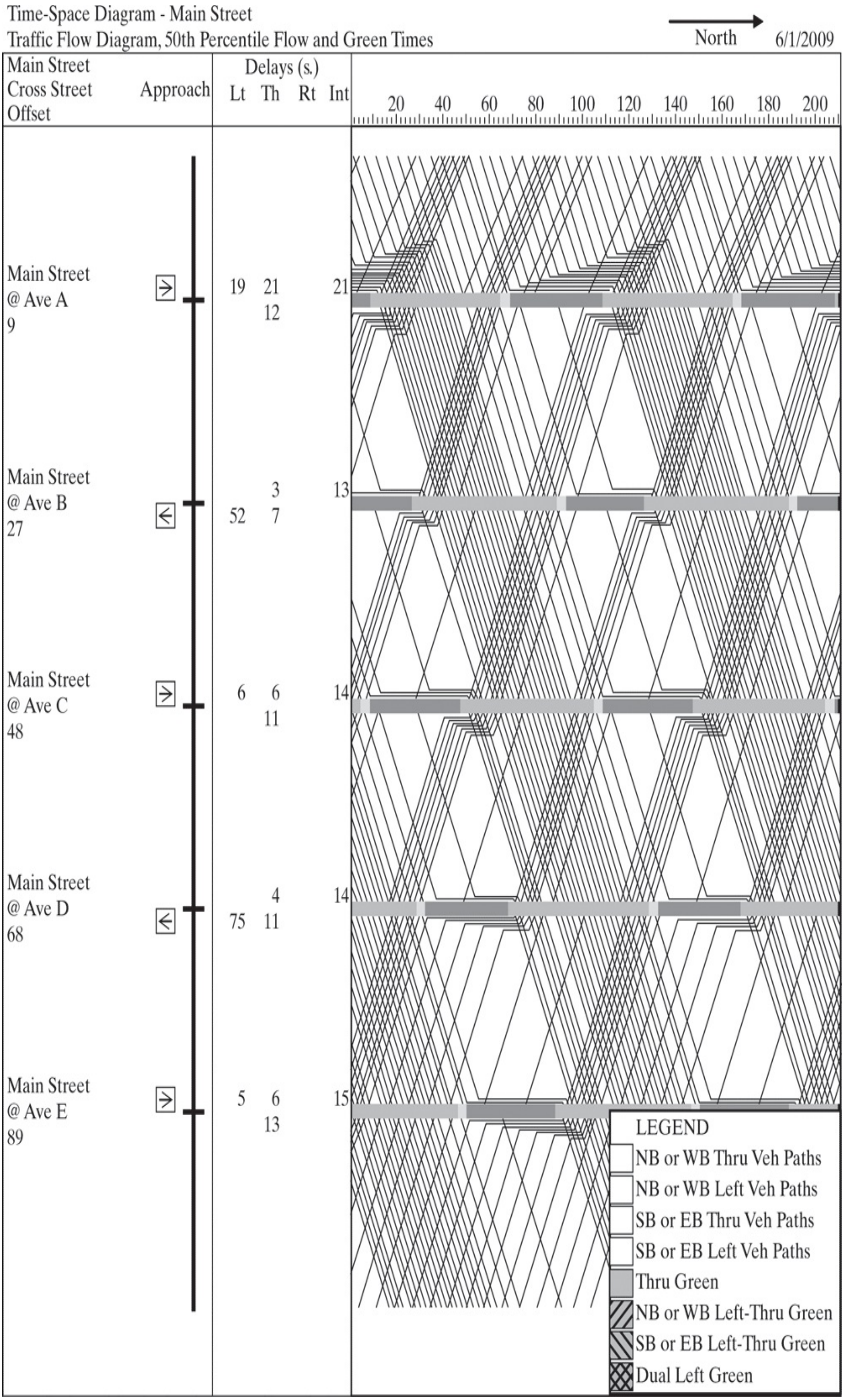

Time-Space Diagram - Main Street
Traffic Flow Diagram, 50th Percentile Flow and Green Times
North
6/1/2009
Main Street
Cross Street
Offset
Approach
Delays (s.)
Lt Th Rt Int
20 40 60 80 100 120 140 160 180 200
Main Street
@ Ave A
9
19 21 21
12
Main Street
@ Ave B
27
3 13
52 7
Main Street
@ Ave C
48
6 6 14
11
Main Street
@ Ave D
68
4 14
75 11
Main Street
@ Ave E
89
5 6 15
13
LEGEND
NB or WB Thru Veh Paths
NB or WB Left Veh Paths
SB or EB Thru Veh Paths
SB or EB Left Veh Paths
Thru Green
NB or WB Left-Thru Green
SB or EB Left-Thru Green
Dual Left Green

Figure 21.30: Full Alternative Text

21.8 Coordination of Signals for Oversaturated Networks

It is well recognized that an oversaturated traffic environment is fundamentally different from an undersaturated environment. In undersaturated networks, capacity is adequate and queue lengths are generally well contained within individual approaches. On the other hand, the oversaturated environment is characterized by an excess of demand relative to capacity ($v/c>1.00$) and thus unstable queues that tend to expand over time with the potential of physically blocking intersections (blockage, spillback), thus slowing queue discharge rates and, in effect, reducing capacity when it is most needed. Control policies for oversaturated networks therefore focus on maintaining and fully exploiting capacity to maximize productivity (vehicle throughput) of the system by controlling the inherent instability of queue growth.

The general approaches to oversaturation conditions are described in the next sections of the chapter. References [6] to [11] give a more complete discussion of timing oversaturated environments.

21.8.1 System Objectives for Oversaturated Conditions

When networks are congested, the explicit objectives change from minimize delay and stops to:

- Maximize system throughput. This is the primary objective. It is achieved by (a) avoiding queue spillback, which blocks intersections and wastes green time; (b) avoiding starvation, the tardy arrival of traffic at the stop-bar, which wastes green time; and (c) managing queue formation to yield the highest service rate across the stop-bar.
- Fully utilize storage capacity. This objective seeks to confine congested conditions to a limited area by managing queue formation

in the context of a “feed forward” system.

- Provide equitable service. This objective seeks to allocate service to cross-street traffic and to left-turners so that all travelers are serviced adequately and the imperative of traffic safety is observed.

Because intersection blockage can so degrade the network, its removal must be the prime objective of the traffic engineer. The overall approach can be stated in a logical set of steps [8]:

- Address the root causes of congestion—first, foremost, and continually.
- Update the signalization, for poor signalization is frequently the cause of what looks like an incurable problem.
- If the problem persists, use novel signalization to minimize the impact and spatial extent of the extreme congestion.
- Provide more space by use of turn bays and parking restrictions.
- Consider both prohibitions and enforcement realistically—do they represent futile effort that will only transfer the problem?
- Take other available steps, such as right-turn-on-red, recognizing that the benefits will generally not be as significant as either signalization or more space.
- Develop site-specific evaluations where there are conflicting goals, such as providing local parking versus moving traffic, when the decision is ambiguous. Explicitly consider the solution in terms of economics.

The last step was intended (for instance) to focus the debate on the use of space, by quantifying the effects and tradeoffs—for example, use for good delivery, bus lane, or parking.

21.8.2 Metering Plans

One short-term demand management strategy is *metering*. Note that the metering concept does not explicitly minimize delay and stops, but rather manages the queue formation in a manner that maximizes the productivity of the congested system. Three forms of metering can be applied within a congested traffic environment:

1. internal,
2. external, and
3. release.

Internal metering refers to the use of control strategies within a congested network so as to influence the distribution of vehicles arriving at or departing from a critical location. The vehicles involved are stored on links that are part of the congested system, so as to eliminate or significantly limit the occurrence of either upstream or downstream intersection blockage.

Figures 21.31 and 21.32 show situations in which internal metering might be used. In Figure 21.31(a) the volume being discharged at intersections upstream of a *CI* is controlled, thus creating a "moving storage" situation on the upstream links. In Figure 21.31(b) the turn-in flow from cross-streets is limited, thus preserving the arterial for its through flow. In Figure 21.32 there is metering in the face of a backup from "outside."

Figure 21.31: Internal Metering Used to Limit Volume Arriving at Critical Location

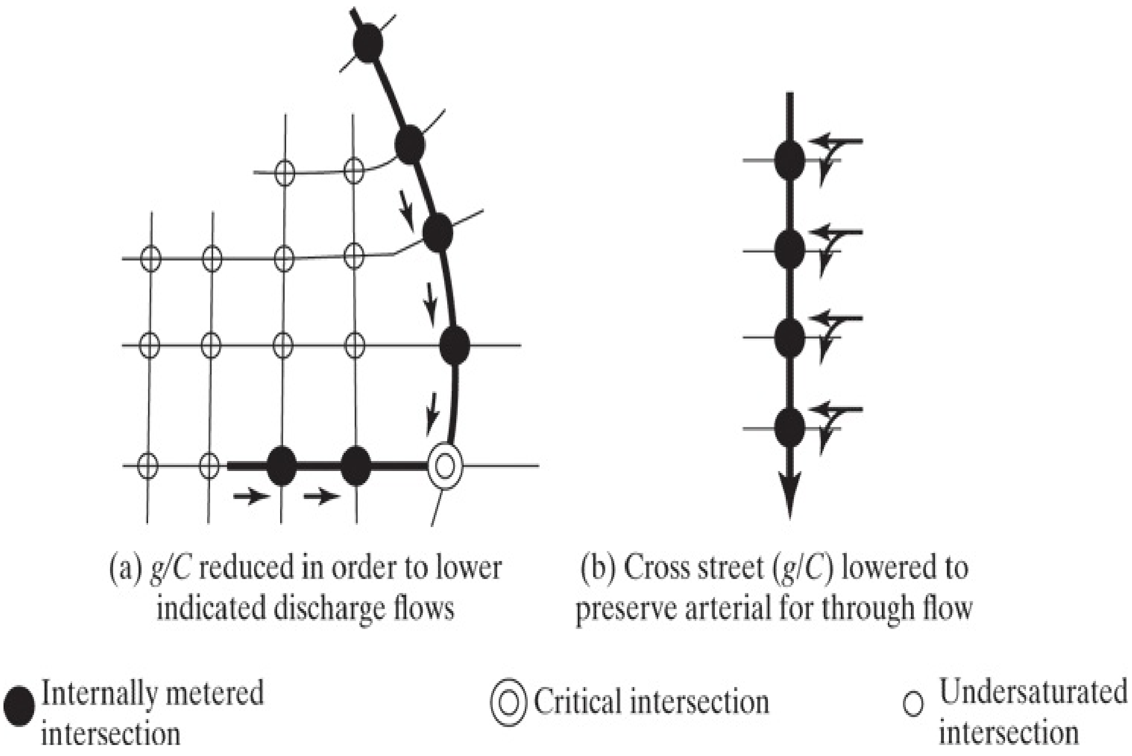


Figure 21.31: Full Alternative Text

Figure 21.32: Application of Internal Metering in the Face of a Backup from "Outside"

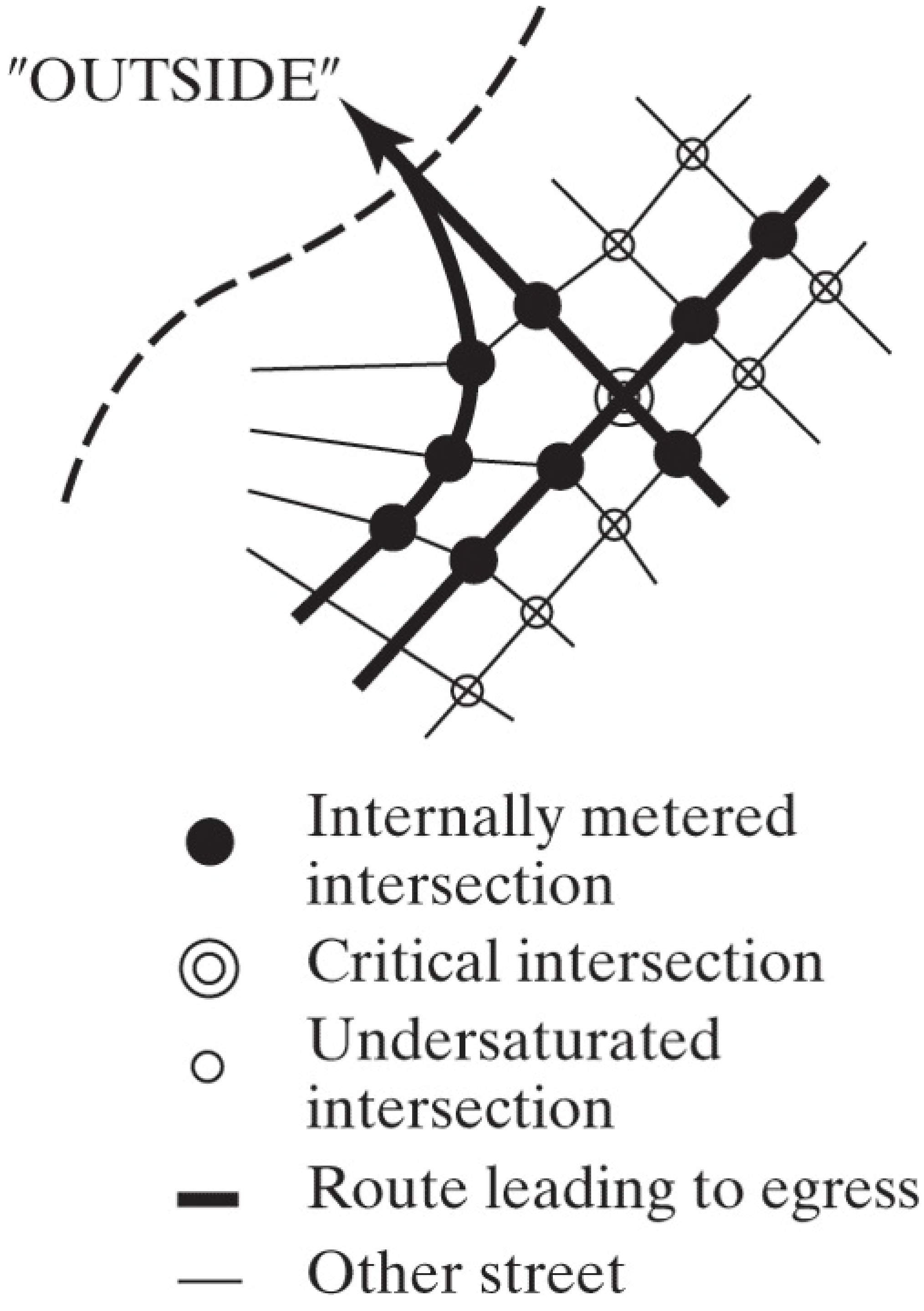


Figure 21.32: Full Alternative Text

External metering refers to the control of the major access points to the defined system, so that inflow rates into the system are limited if the system is already too congested (or in danger of becoming so).

External metering is convenient conceptually, because the storage problem belongs to “somebody else” outside the system. However, there may be limits to how much metering can be done without creating major problems in the “other” areas. Figure 21.33 shows a network with metering at the access points.

Figure 21.33: Illustration of External Metering

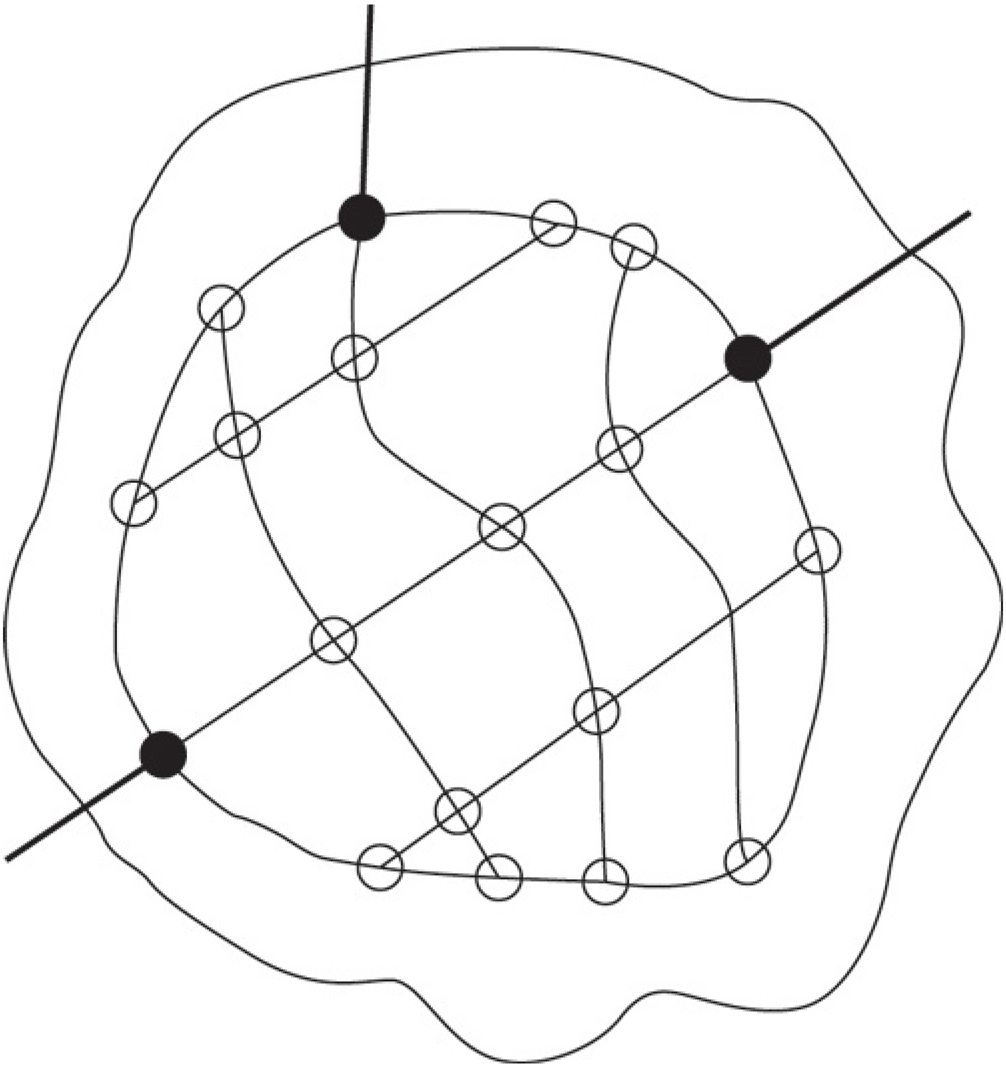


Figure 21.33: Full Alternative Text

As a practical matter, there must be a limited number of major access points (such as river crossings, a downtown surrounded by water on three sides, a system that receives traffic from a limited number of radial arterials, and so forth). Without effective control of access, the control points can potentially be bypassed by drivers selecting alternative routes.

Release metering refers to cases where vehicles are stored in such locations as parking garages and lots, from which their release can be controlled (at least in principle). The fact that they are stored “off street” also frees the traffic engineer of the need to worry about their storage and their spillback potential.

Release metering can be used at shopping centers, mega-centers, major construction sites, and other concentrations. While there are practical problems with public (and property-owner) acceptance, this could even be —and has been—a developer strategy to lower the facility’s discharge rates so that adverse impacts are avoided (since traffic impacts generally need to be mitigated at a cost to the developer). Such strategies are of particular interest when the associated roadway system is distributing traffic to egress routes or along heavily congested arterials.

21.8.3 Signal Remedies

It is difficult to overstate how often the basic problem in oversaturated networks is poor signalization. Once the signalization is improved through reasonably short cycle lengths, proper offsets (including queue clearance), and proper splits, other options may not be needed.

Rapid Adjustment to Splits

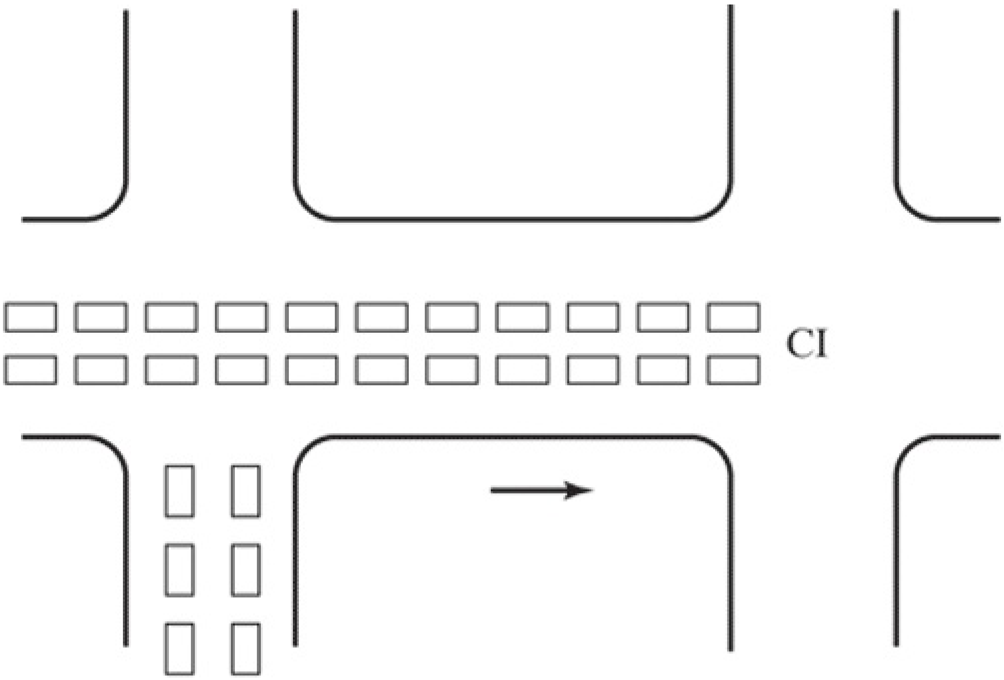
Rapid adjustment to splits is used to meet short-term relative demand changes (i.e., in competing directions). There are locations such as college entrances that have short bursts of inflows followed by short bursts of outflows (both in the order of 15–25 minutes), directly related to their class schedules. Control in such cases must adapt to the rapidly changing demand, in order to avoid precipitating oversaturation that can promulgate and perpetuate itself.

Equity Offsets

Offsets on an arterial are usually set to move vehicles smoothly along the arterial, as is logical. Unfortunately, as the queue length approaches the block length, such progressions lose meaning, for it is quite unlikely that both the queue *and* the arriving vehicles will be processed at the downstream intersection, and thus the arrivals will be stopped in any case. At the same time, the cross street traffic at the upstream intersection is probably poorly served because of intersection blockages. Equity offsets are used to avoid these problems.

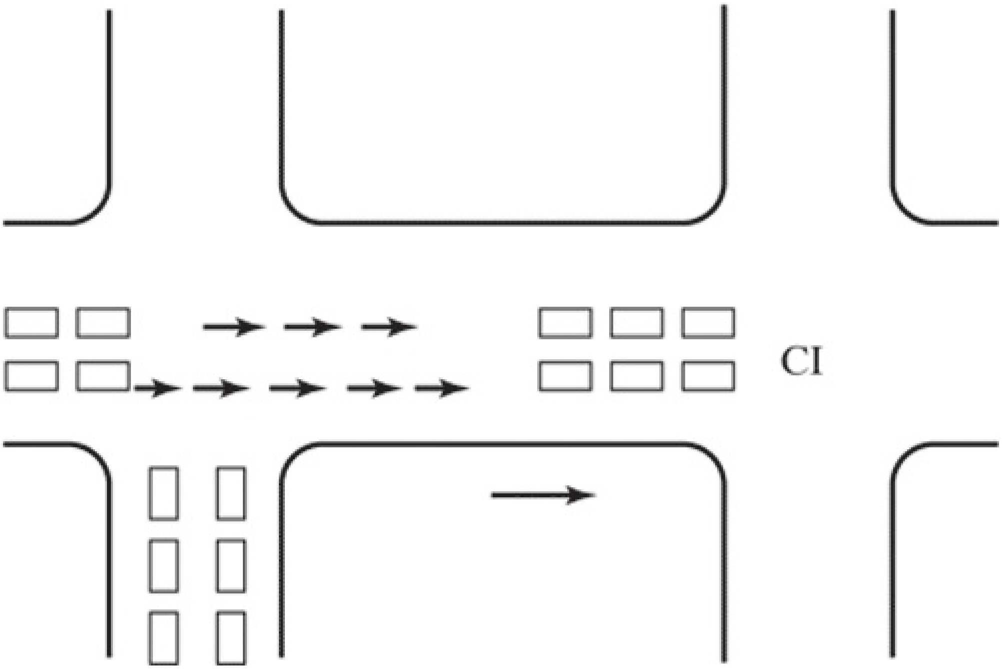
Consider the following case, illustrated in Figure 21.34. Allow the congested arterial to have its green at the upstream intersection until its vehicles just begin to move, then switch the signal, so that these vehicles flush out the intersection, but no new vehicles continue to enter. This gives the cross street traffic an opportunity to pass through a clear intersection.

Figure 21.34: Concept of Equity Offsets to Clear Side Streets



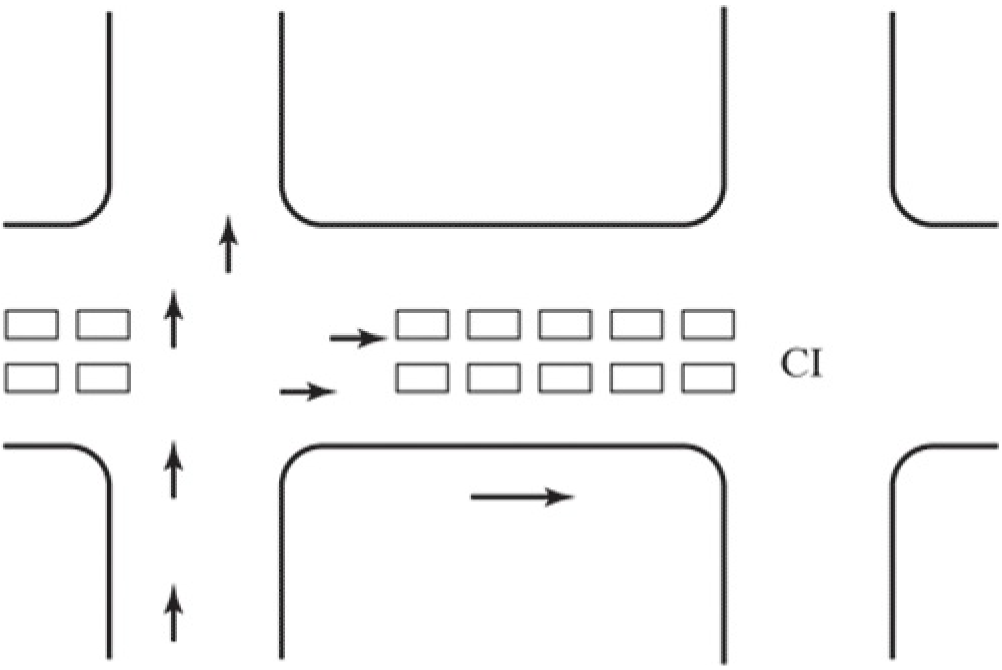
(a) Main-street vehicles block upstream

21.8-3 Full Alternative Text



(b) As main-street vehicles just begin to clear, give green to side street

[21.8-3 Full Alternative Text]



(c) Cross street are thus allowed to discharge

[21.8-3 Full Alternative Text]

This concept, defined as *equity offset*, can be translated into the equation:

tequity=giC−LSacc [21-11]

where gi is the *upstream* main street (i.e., the congested intersection) green fraction and Sacc is the speed of the “acceleration wave” shown in Figure 21.35.

Figure 21.35: Equity Offset to Benefit Cross Street

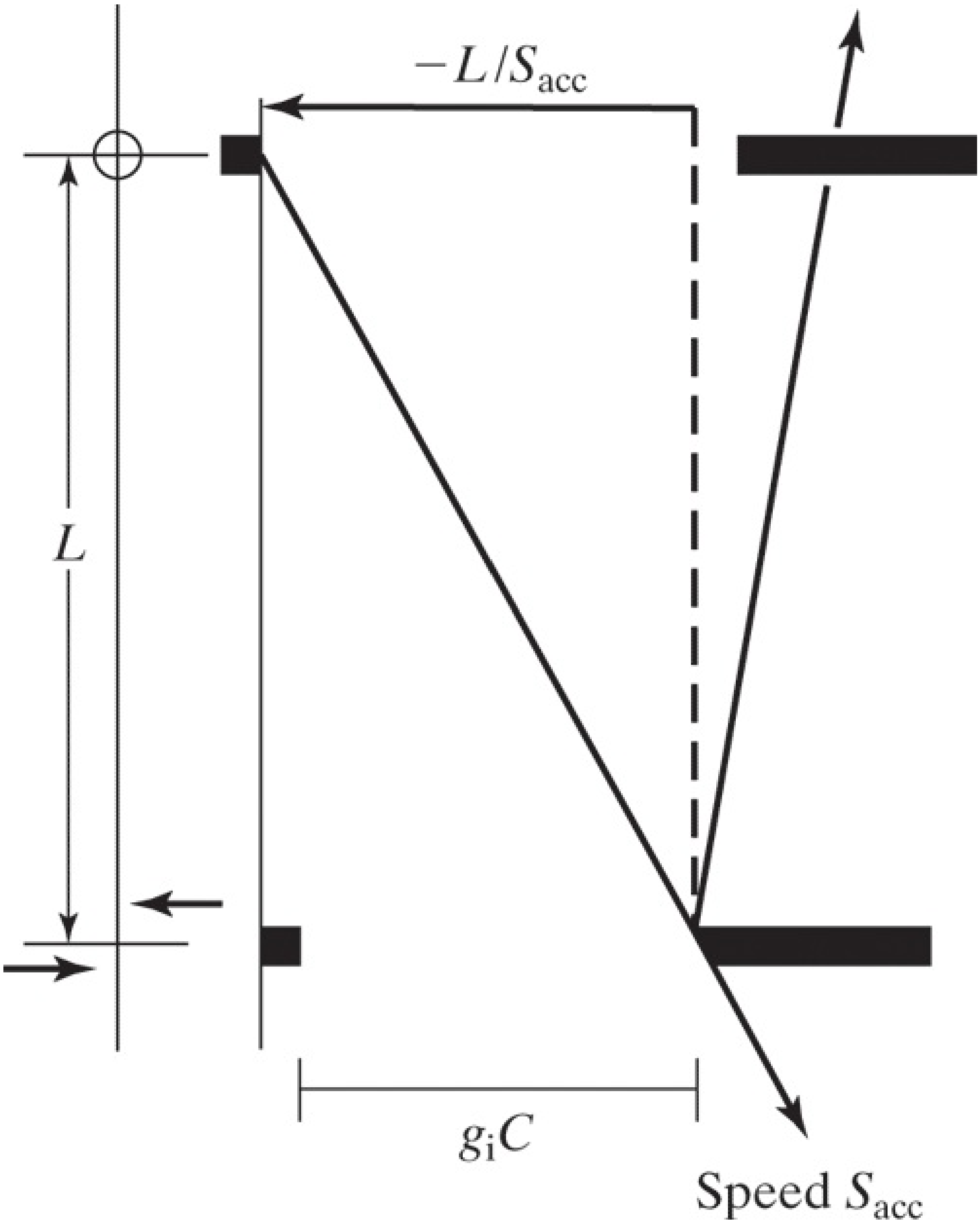


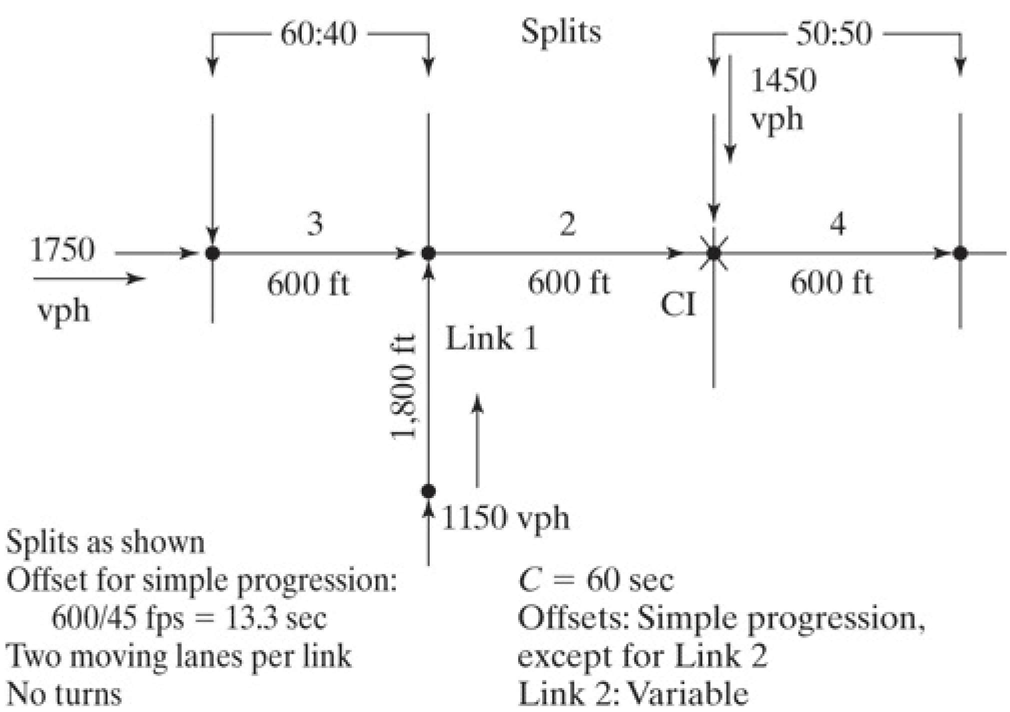
Figure 21.35: Full Alternative Text

A typical value for the acceleration wave is 16 ft/s. It is clear from Figure 21.35 that equity offset causes the upstream signal to go red just when "normal" offsets would have caused it to switch to green in this particular case. This is not surprising, for the purpose is different—equity offsets are intended to be fair (i.e., equitable) to cross-street traffic.

Simulation tests using a microscopic simulation model have shown the value of using equity offsets: Congestion does not spread as fast as otherwise and may not affect the cross streets at all.

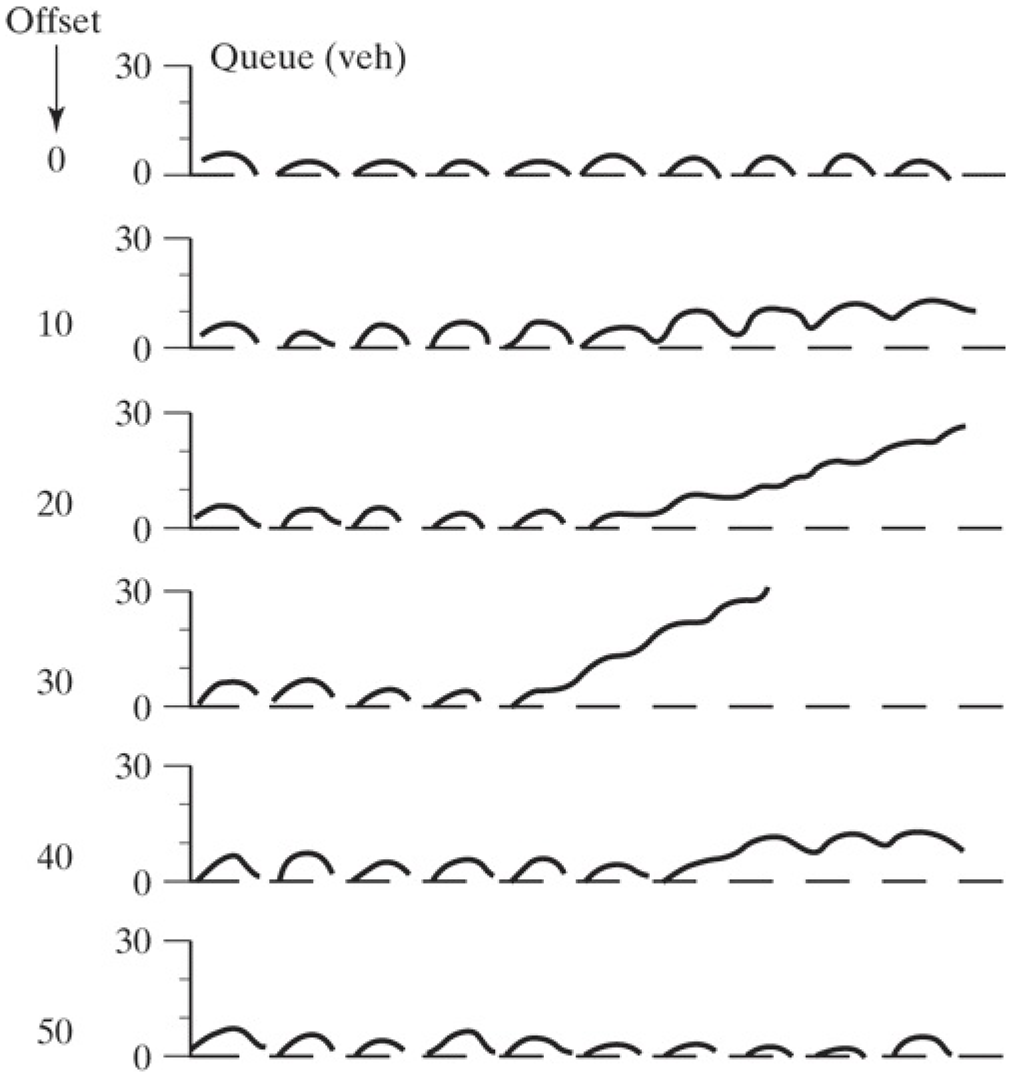
Figure 21.36(a) shows a test network used to test the equity offset concept. Link 2 is upstream of the CI. For the demands and signal splits shown it is likely to accumulate vehicles, with spillback into its upstream intersection likely. If this occurs, the discharge from Link 1 will be blocked and its queue will grow. In the extreme, congestion will spread.

Figure 21.36: Equity Offsets Avoid Side-Street Congestion, Despite Spillbacks



(a) Test network

[21.8-3 Full Alternative Text]



(b) Side-street queue (Link 1)

[21.8-3 Full Alternative Text]

The equity offset is computed as:

$$t_{equity}=(0.60)(60)-\frac{600}{16}=-1.5\ \text{s}$$

using Equation 21-11. Compare this to the ideal offset for progressing a movement. At 25 ft per vehicle, the full link can hold $600/25=24$ vehicles. At a platoon speed of 50 ft/s, computing the ideal offset adjusted for the queue would have yielded $t_{ideal}=(600/50)-(24)(2)=-36$ s.. Of course, progressed movement is a silly objective when 24 vehicles are queued.

Figure 21.36(b) shows the side-street queue (i.e., the Link 1 queue) as a function of the main-street offset. Note that an offset of –36 seconds is the same as an offset of +24 seconds when C=60 seconds. The figure shows the best result for allowing the side street to clear is when the equity offset (offset $=-1.5$ seconds) is in effect, and, in this case, the worst results when the queue-adjusted “ideal offset” (offset $=24$ seconds) would have been in effect.

The aforementioned discussion assumes that the cross-street traffic does not turn into space opened on the congested arterial. If a significant number of cross-street vehicles do turn into the arterial, a modification in the offset is appropriate to ensure that the upstream traffic on the congested arterial also has its fair share.

The equity offset concept has been used to keep side-street flows moving when an arterial backs up from a CI. It may also be used to keep an arterial functioning when the cross streets back up across the arterial from their critical intersections.

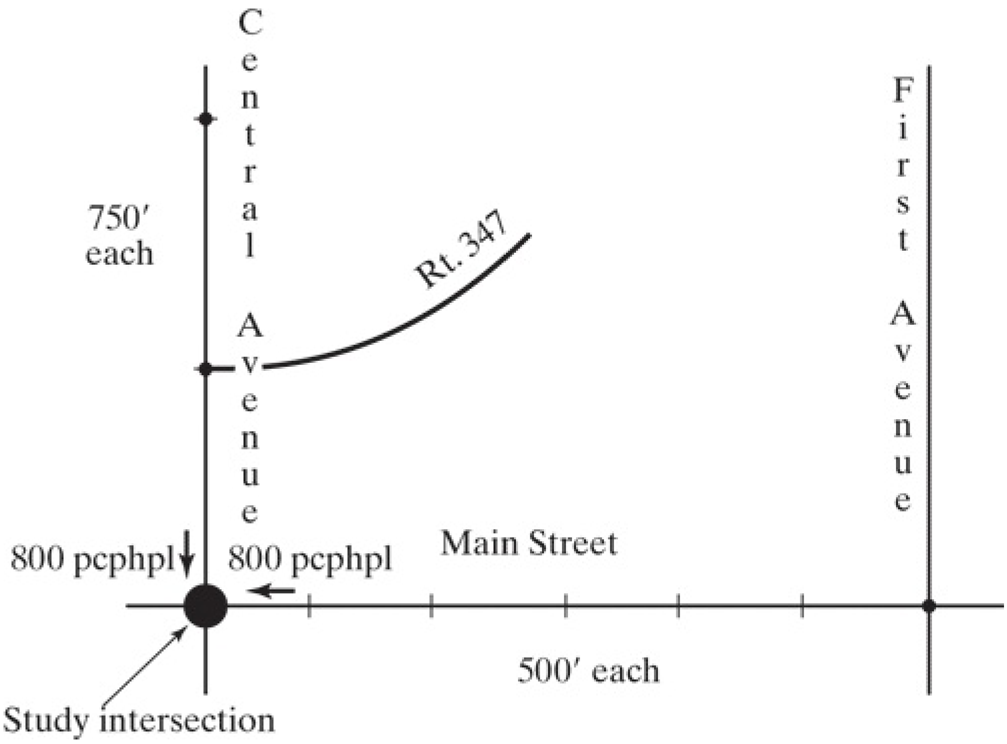
Phase Reservice

The term “phase reservice” refers to servicing important phases more than once in a cycle, by going back to them, generally to the disbenefit of side street movements on other phases. The technique is used for clearing queues on protected lefts and saturated approaches, but generally requires that there are undersaturated phases at the intersection, so that one can “catch up” with servicing them on a future cycle if necessary. Phase reservice can aid the basic objectives of maximizing throughput and queue management. It does require that both drivers and pedestrians become familiar with this sort of operation, so that all concerned are aware that the “normal” sequence of phases cannot be counted upon.

Imbalanced Split

For congested flow, the standard rule of allocating the available green in proportion to the relative demands could be used, but it does not address an important problem. Consider the illustration of Figure 21.37. If the prime concern is to avoid impacting Route 347 and First Ave. (but with little concern for the minor streets in between, if any), it is not reasonable to use a 50:50 split.

Figure 21.37: An Illustration of Split Determination



(Source: *Traffic Signal Timing Manual*, 1st Edition, Federal Highway Administration, Washington, D.C., June 2008, Figure 6-2.)

Figure 21.37: Full Alternative Text

Considering that the relative storage available is 750 ft in one direction and 3,000 ft in the other, and we wish neither to be adversely affected, the impact could be delayed for the longest time by causing the excess-vehicle queue to grow in proportion to their available storage. The two critical-lane discharge flows fi would have to be set such that:

d1−f1d2−f2=L1L2 [21-12]

and:

f1+f2=CAP [21-13]

where di is the demand (veh/h/ln), Li is the storage and *CAP* is the sum of the critical-lane flows.

For the illustrative problem, using CAP=1,550 veh/h/ln, equations 21-12 and 21-13 result in f1=954 veh/h/ln and f2=759 veh/h/ln, where direction 1 is the shorter distance. This is a 56:44 split.

Note that in the extreme, if only one direction has a cross route that should not be impacted, much of the green could be given to that direction (other than some minimum for other phases) in order to achieve that end.

Pedestrian Push Buttons

In areas in which there is relatively little pedestrian traffic, satisfying the pedestrian minimum crossing times on all approaches may lead to phases that waste green time and to longer-than-necessary cycle lengths. In such cases, traffic engineers consider invoking the minimum pedestrian crossing time only when there is an actual pedestrian actuation of a pedestrian button.

21.8.4 Why Shorter Cycle Lengths are Important

As the cycle length increases, so do the lengths of the stored queue and the length of the discharged platoons, which then arrive at downstream intersections that may have shorter cycle lengths and cannot be stored or processed easily. Thus, the likelihood of intersection blockage increases, with substantial adverse impacts on system capacity. This is particularly acute when short link lengths are involved. Consider that each cycle nominally discharges $v_i\ C/3{,}600$ vehicles. If each vehicle requires *D* ft of storage space, the length of the downstream link in a congested environment (assuming the downstream signal can process the queue in one cycle, but that it will be forced to stop) would have to be:

$$L \geq (v1C3,600) \times D \quad [21\text{-}14]$$

where *L* is the available downstream space in feet. This “available” space may be the full link length or some lower value, perhaps 150 ft less than the true length (to keep the queue away from the discharging intersection, or to allow for turn-ins).

Equation 21-13 may be rearranged as:

$$C \leq (LD)(3,600vi) \quad [21\text{-}15]$$

The vi in this case is the discharge volume per downstream lane, which may differ from the demand volume, particularly at the fringes of the “system” being considered. Refer to Figure 21.38 for an illustration of this relationship. It can be seen from the figure, that only rather high flows (≥ 800 veh/h/ln) and short blocks (L/D=10) will create very severe limits on the cycle length. However, these are just the situations of most interest for conditions of extreme congestion. *Note that the discharge volume downstream, vi, depends upon the upstream demand and (g/C) allocation,* and that this analysis really has to be carried along the arterial.

Figure 21.38: Maximum Cycle Length as a Function of Block Length

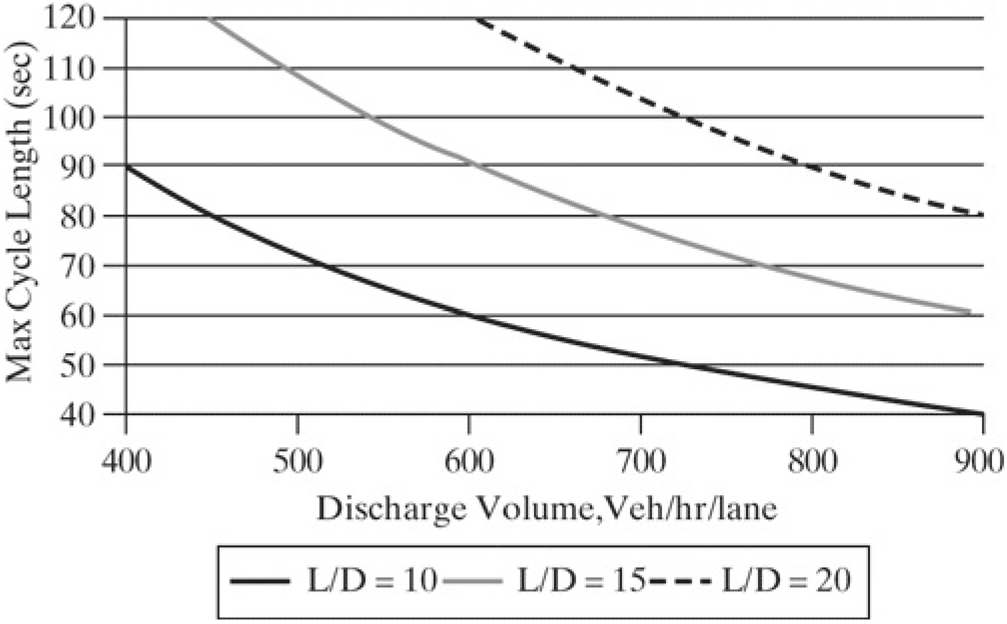


Figure 21.38: Full Alternative Text

The important lesson from Figure 21.38 is that shorter cycle lengths are not only good but *necessary* to manage the size of queues arriving in downstream links.

This analysis assumes that the downstream link can itself discharge the arriving queue in one cycle. In order to achieve this (for instance, at a critical intersection), it may be necessary to allow the downstream capability to determine the upstream discharge, which would have to be achieved by reassigning green time there (to minor movements) or imposing an all-red (i.e., metering). Failing this, the downstream queues will grow and other measures will be needed to avoid spillback.

21.8.5 Summary of Oversaturated Conditions

The problem of congestion and saturation is widespread and is not often approached in any consistent manner. Definite measures can be taken, but preventive action addressing the root causes must be given a high priority. Among the possible measures, those relating to signalization generally can have the greatest impact. The nonsignal remedies are in no way to be minimized, particularly those that provide space, whether for direct productivity increases or for removing impediments to the principal flow.

References

- 1. *Traffic Signal Timing Manual*, 1st Edition, Federal Highway Administration, Washington, D.C., June 2008.
- 2. *PASSER II-90 Microcomputer User's Guide*, distributed by the McTrans Center, Gainsville, Fl, 1991.
- 3. Tru-Traffic 10.0, www.tsppd.com
- 4. *TRANSYT-7F Users Guide, Methodology for Optimizing Signal Timing* (MOST) Volume 4, USDOT, Federal Highway Administration, Washington, D.C., 1991.
- 5. *Synchro Users Guide*, Trafficware LTD, www.trafficware.com
- 6. McShane, W.R., et al. "Traffic Control in Oversaturated Street Networks," National Cooperative Highway Research Program, (NCHRP) Report 194, Transportation Research Board, National Research Council, Washington DC, 1978.
- 7. Hajbabaie, A. et al., "Traffic Signal Coordination and Queue Management in Oversaturated Intersection," *Final Report*, Purdue University, 2011.
- 8. *Signal Timing Under Saturated Conditions, Literature Review*, Federal Highway Administration, Washington DC, 2009.
- 9. Lieberman, E.B., Chang, J., Prassas, E.S., "Formulation of Real-Time Control Policy for Oversaturated Arterials," *Transportation Research Record 1727*, TRB, National Research Council, Washington DC, 2000.
- 10. "Signal Timing for Congested Conditions," ITE Professional Development Program, October 11, 2007 webinar, presented by Woody Hood, Maryland DOT
- 11. TRB Traffic Signal Systems Committee (TRB Committee ANB25) website, http://www.signalsystems.org.vt.edu/

Problems

1. 21-1. Two signals are spaced at 1,000 ft on an urban arterial. It is desired to establish the offset between these two signals, considering only the primary flow in one direction. The desired progression speed is 40 mph. The cycle length is 60 seconds. Saturation headway may be taken as 2.0 s/veh and the start-up lost time as 2.0 seconds.

    1. What is the ideal offset between the two intersections, assuming that vehicles arriving at the upstream intersection are already in a progression (i.e., a moving platoon), at the initiation of the green?

    2. What is the ideal offset between the two intersections, assuming that the upstream signal is the first in the progression (i.e., vehicles are starting from a standing queue)?

    3. What is the ideal offset, assuming that an average queue of three vehicles per lane is expected at the downstream intersection at the initiation of the green? Assume the base conditions of part (a).

    4. Consider the offset of part (a). What is the resulting offset in the opposite (off-peak) direction? What impact will this have on traffic traveling in the opposite direction?

    5. Consider the offset of part (a). If the progression speed were improperly estimated and the actual desired speed of drivers was 45 mph, what impact would this have on the primary direction progression?

2. 21-2. Consider the time-space diagram on page 496 for this problem. For the signals shown:

    1. What is the NB progression speed?

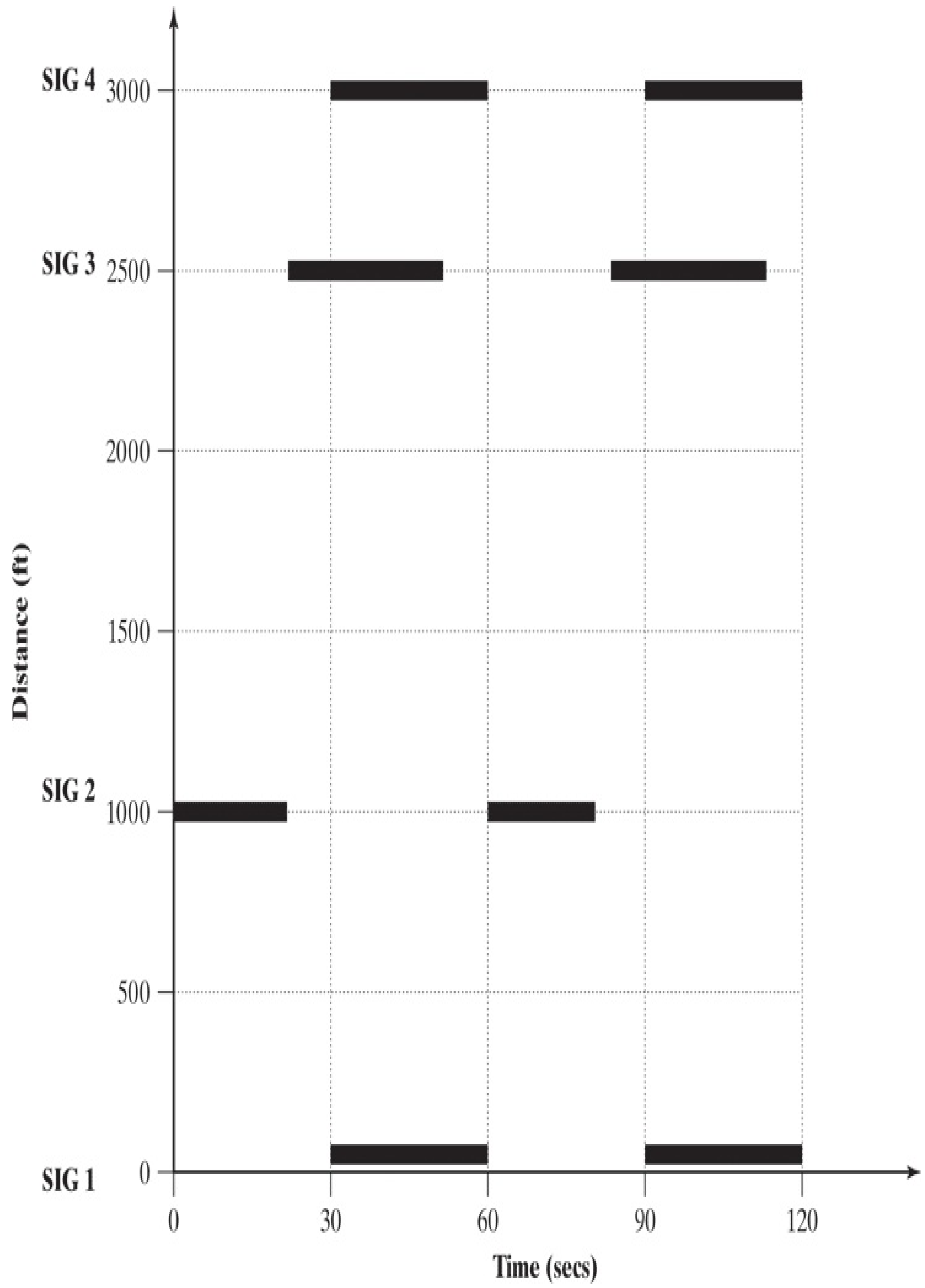
    2. What is the NB bandwidth and bandwidth capacity? Assume a

saturation headway of 2.0 s/veh.

3. What is the bandwidth in the SB direction for the same desired speed as the NB progression speed? What is the SB bandwidth capacity for this situation?

4. A new development introduces a major driveway, which must be signalized between Intersections 2 and 3. It requires 15 seconds of green out of the 60-second system cycle length. Assuming that you had complete flexibility as to the exact location of the new driveway, where would you place it? Why?

Time Space Diagram for Problem 21-2



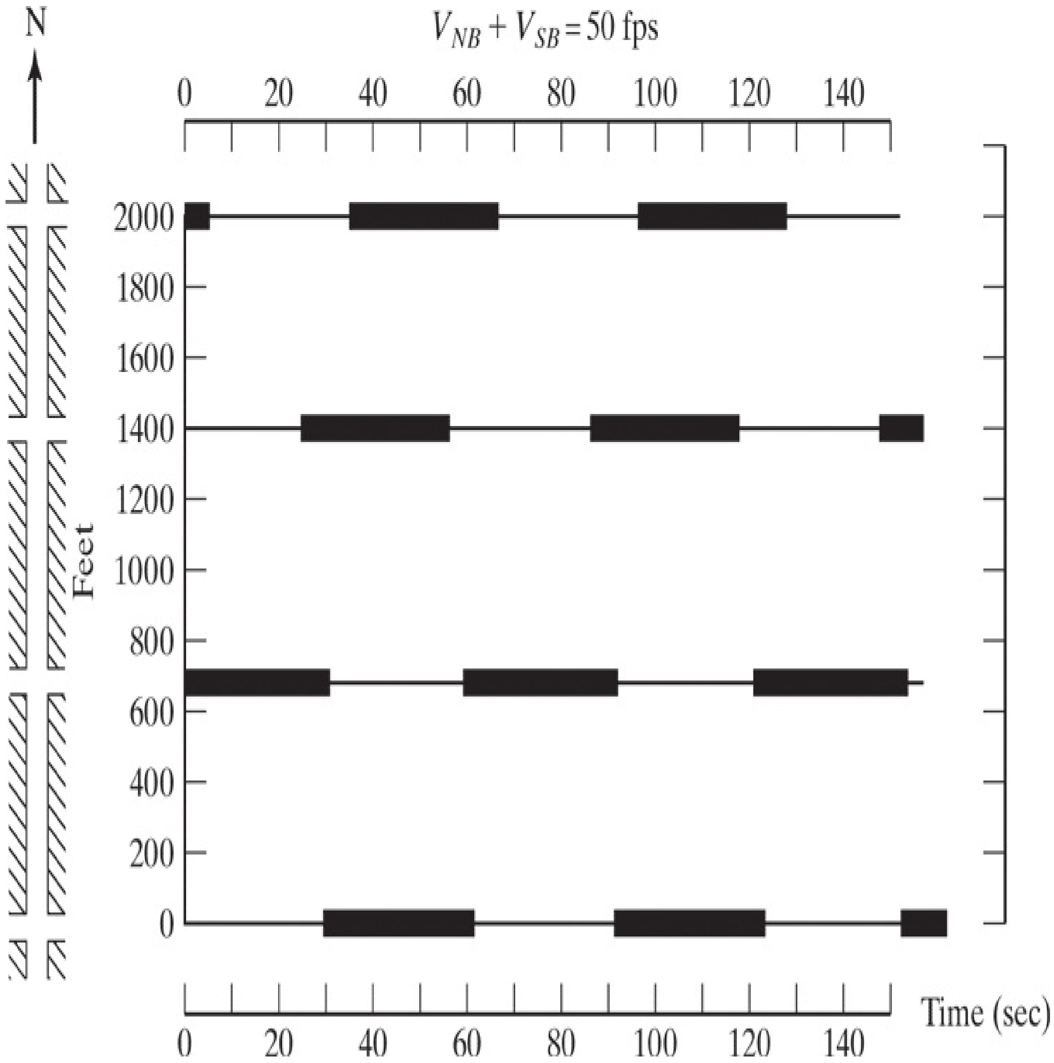
Full Alternative Text

3. 21-3. A downtown grid has equal block lengths of 750 ft along its primary arterial. It is desired to provide for a progression speed of 30 mph, providing equal service to traffic in both directions along the arterial.

   Would you suggest an alternating or a double-alternating progression scheme? Why?

4. 21-4. Refer to the time-space diagram on page 497. Trace the lead NB vehicle through the system. Do the same for the lead SB vehicle. Use a platoon speed of 50 ft/s. Estimate the number of stops and the seconds of delay for each of these vehicles.

Time-space Diagram for Problem 21-4

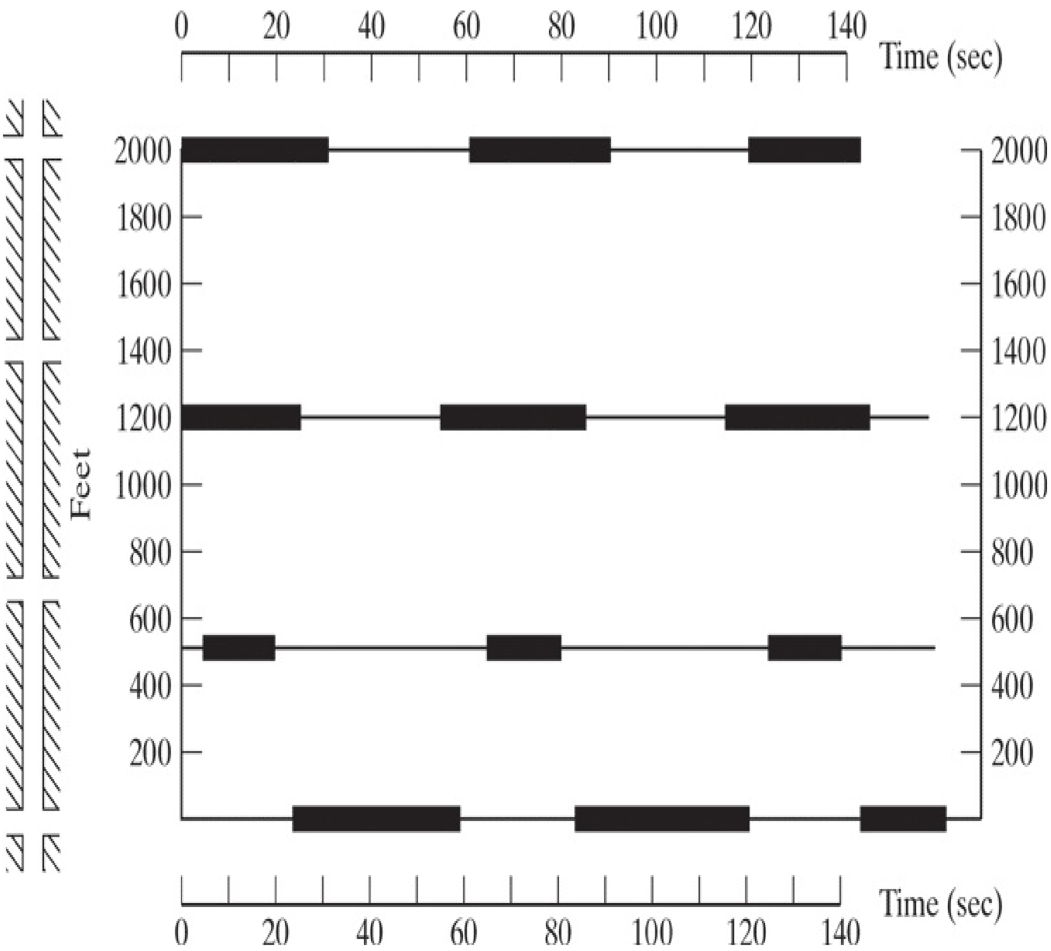


Full Alternative Text

5. 21-5. Refer to the time-space diagram on page 497. Find the NB and

the SB bandwidths (in seconds). Determine the efficiency of the system in each direction and the bandwidth capacity. There are three lanes in each direction. The progression speed is 50 ft/s.

Time-space Diagram for Problem 21-5



Full Alternative Text

6. 21-6.

    1. If vehicles are traveling at 60 fps on a suburban road and the signals are 2,400 ft apart, what cycle length would you recommend? What offset would you recommend?

    2. If an unsignalized intersection is to be inserted at 1,200 ft

(midway), what would you recommend? Is there a better location?

7. 21-7.

   1. Construct a time-space diagram for the following information and estimate the northbound bandwidth and efficiency for platoons going at 50 ft/s.

Data for Problem 21-7

Signal No.	Offset (s)	Cycle length	Split (MSG first)
6	52	60	50:50
5	36	60	60:40
4	20	60	60:40
3	52	60	60:40
2	24	60	50:50
1	0	60	60:40

   Full Alternative Text

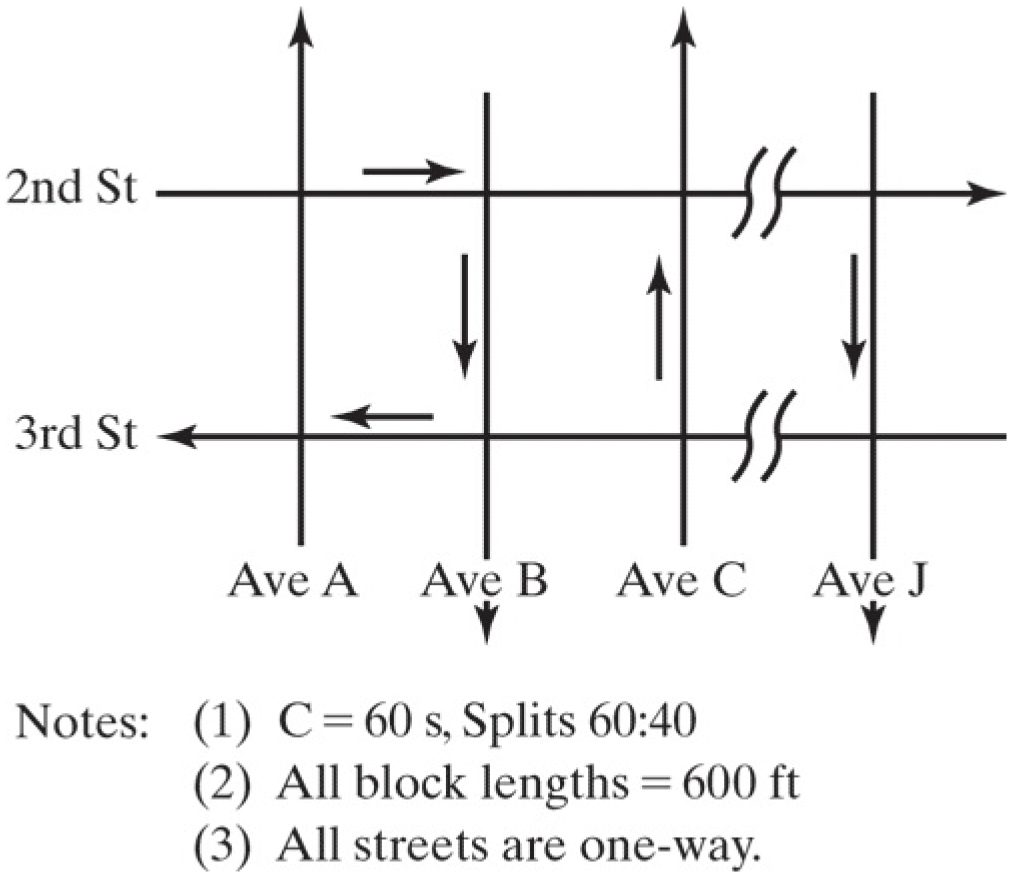
   All of the offsets are relative to the Master Clock zero. All signals are two-phase. There are two lanes in each direction. All block lengths are 1,200 ft.

   2. Estimate the number of platooned vehicles that can be handled nonstop northbound and southbound.

8. 21-8. Refer to figure. The 2nd street is eastbound with offsets of+ 15 s between successive signals. The 3rd street is westbound with offsets of + 10 s between successive signals. Avenue A is northbound. The

offset between the intersections of Avenue A and 3rd Street and Avenue A and 2nd street is 20 seconds. Given this information, find the offsets along Avenues B through J. The directions alternate, and all splits are 60:40, with the 60 on the main streets (2nd and 3rd streets). Cycle length = 60 s.

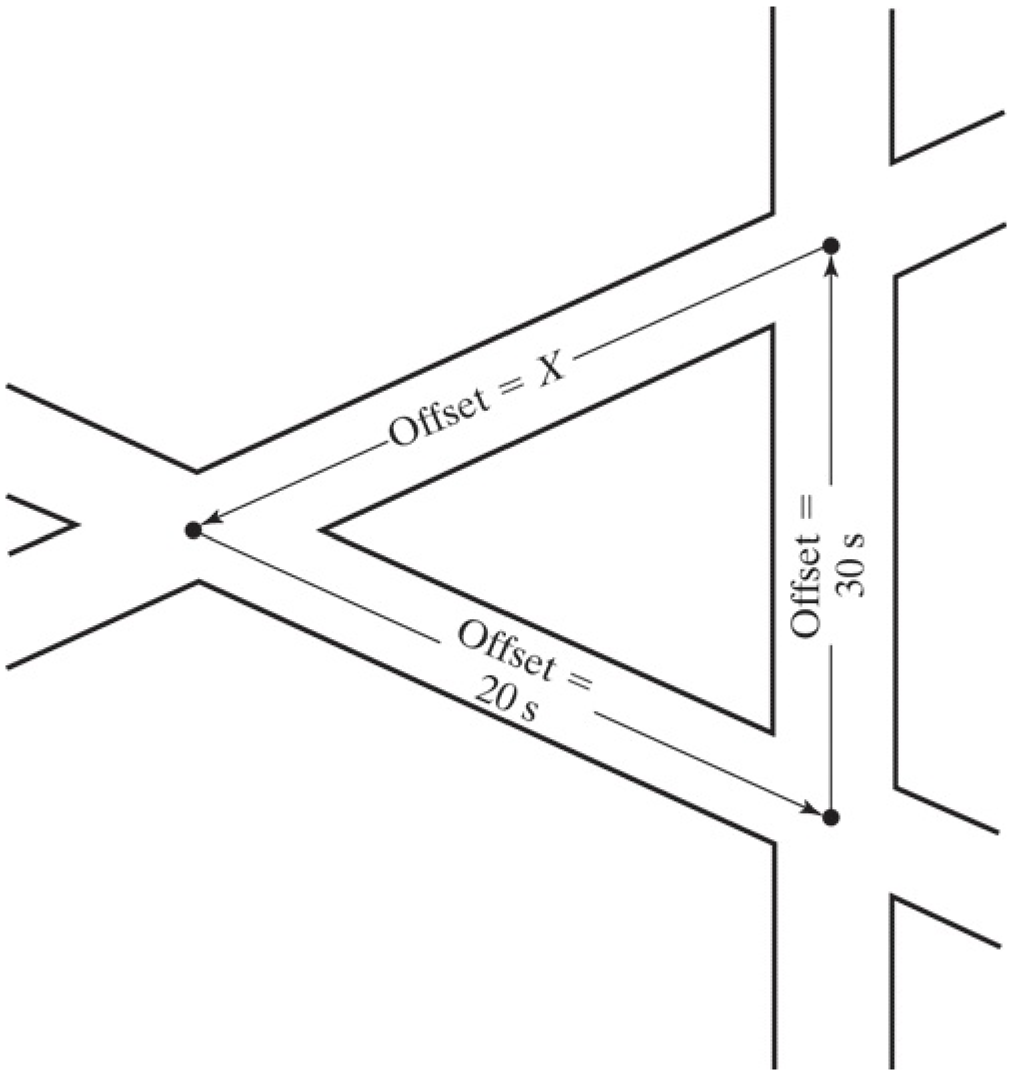
Network for Problem 21-8



Full Alternative Text

9. 21-9. Refer to figure. Find the unknown offset *X*. The cycle length is 80 seconds. The splits are 50:50.

Network for Problem 21-9



Full Alternative Text

Chapter 22 Capacity and Level of Service Analysis: Signalized Intersections—The HCM Method

The signalized intersection is the most complex location in any traffic system. In Chapter 18, some simple models of critical operational characteristics were presented. In Chapters 19 and 20, these were applied to create a simple signal-timing methodology for pre-timed and actuated signals, respectively.

Analysis of a signalized intersection is conceptually the reverse of signal timing. In signal timing, signal settings are devised to provide adequate capacity to handle known demand flows. In analysis, signal timing is known, and the capacity is estimated. Theoretically, one process should be the reverse of the other, using a single model.

The signal-timing methodology involved a number of simplifying assumptions, among which was a default value for saturation flow rate that reflected "typical" conditions. A complete analysis of any signalized intersection requires use of a more complex model that addresses all of the many variables affecting intersection operations as well as some of the more intricate interactions among component flows.

The most frequently used model for analysis of signalized intersections in the United States is the model contained in the *Highway Capacity Manual* (HCM) [1]. This model first appeared in the 1985 edition of the HCM and has been revised and updated in subsequent editions (1994, 1997, 2000, 2010, and 2016). The model has become increasingly complex, involving several iterative elements. It has become almost impossible, except for the simplest problems, to do complete solutions using this model by hand. Its implementation, therefore, has been primarily through computer software that replicates the model.

Beginning with the 2010 edition of the HCM, the model was presented to handle an actuated controller at the signalized intersection. The modeling approach is, however, still basically for pre-timed control, with an iterative shell that calculates the average signal timing of the actuated controller for the given time period. The 2016 HCM methodology analyzes pre-timed signals by incorporating settings on the assumed actuated controller that make it operate in the pre-timed mode.

The first section of this chapter will focus on the analysis of pre-timed signals, as it is more straightforward to present the basic modeling theory for these. In the second section, the chapter provides a general overview of what the actuated signal analysis methodology actually entails.

There are other models that are in use elsewhere. The SIDRA [2,3] model and associated computer package were developed for use in Australia by the Australian Road Research Board. Some of its elements, particularly in delay estimation, have been adapted and applied in the HCM. A Canadian model also exists [4]. All of these models are “deterministic” analytic models. In deterministic models, the same input data produce the same result each time the model is applied.

There are also a number of simulation models that may be used to analyze individual intersections, as well as networks. Some of the most frequently used are SIMTRAFFIC, VISSIM, AIMSUN, and CORSIM. Simulation introduces stochastic elements. This means that the same input parameters do not produce the same results each time the simulation is run. Because of this characteristic, most stochastic simulation models are run multiple times with a given set of input data, and average results of those runs are used in evaluating the intersection.

This chapter attempts to describe the overall concepts and some of the details of the HCM methodology. Some of the complexities are noted without going into full detail, and the student is encouraged to consult the HCM directly for additional information on such subjects.

Part I Analysis of Pre-timed Signalized Intersections

22.1 Fundamental Concepts

There are six fundamental concepts used in the HCM 2016 signalized intersection analysis methodology that should be understood before considering any of the details of the model:

- The lane group concept
- The v/s ratio as a measure of demand
- Capacity and saturation flow rate concepts
- Level-of-service criteria and concepts
- Effective green time and lost time concepts
- Analysis time period

22.1.1 The Lane Group Concept

The signal-timing methodology of Chapter 19 relied on *critical lanes* and *critical-lane volumes*. In the HCM model, the total demand in a group of lanes is used, that is, instead of identifying a set of “critical lanes,” a set of “critical-lane groups” is identified. A lane group can be a single lane or set of lanes within which there are no lane-use restrictions impeding driver selection of lanes.

Where several lanes operate in equilibrium (i.e., where there are no lane-use restrictions impeding the drivers selection of which lane to use), the *lane group* is treated as a single entity.

Not all methodologies do this. Both the Australian and Canadian models focus on individual lanes, taking into account unequal use of lanes. The HCM also accounts for unequal use of lanes through a process of adjustments to saturation flow rate.

22.1.2 The *v/s* Ratio as a Measure of Demand

In Chapter 19, a signal timing methodology was based on conversion of demand volumes to “through-vehicle equivalents.” This allowed volumes with markedly different percentages of right- and left-turning vehicles to be directly compared in the determination of “critical lanes.” Other conditions that might affect the equivalency of volumes (heavy vehicles, grades, parking conditions, etc.) were not taken into consideration.

In the HCM model, demand flow rates are not converted. They are stated as “veh/h” under prevailing conditions. Without conversion to some common base, demand flow rates cannot be compared directly. Instead of converting demand flow rates, the HCM model applies all adjustments to the saturation flow rate. As a result, the methodology yields saturation flow rates and capacities that are defined in terms of prevailing conditions.

The HCM model includes adjustments for a wide variety of prevailing conditions, including the presence of left- and right-turning vehicles, heavy vehicles, parking and others that will be discussed later in the chapter. These are then compared with demand volumes that reflect the same prevailing conditions.

To obtain a single parameter that will allow the intensity of demand in each lane group to be compared directly, the demand flow rate, v, is divided by the saturation flow rate, *s*, to form the “flow ratio,” v/s. Since the prevailing conditions in each lane group are reflected in both the flow rate and the saturation flow rate values, this dimensionless number may be used to represent the magnitude of the demand in each lane group.

22.1.3 Capacity and Saturation Flow Rate Concepts

The HCM model does not produce a value for the capacity of the intersection. Rather, each lane group is considered separately, and a capacity for each is estimated.

Why not simply add all of the lane group capacities to find the capacity of the intersection as a whole? Doing so would ignore the fact that traffic demand does not reach its peak on all approaches at the same time. Unless the demand split on each of the lane groups matched the split of capacities, it would be impossible to successfully accommodate a total demand equal to a capacity so defined. In effect, the "capacity" of the intersection as a whole is not a useful or relevant concept. The intent of signalization is to allocate sufficient time to various lane groups and movements to accommodate demand. Capacity is provided to specific movements to accommodate movement demands.

The concepts of saturation flow rate, capacity, and volume-to-capacity ratio (v/c) are all interrelated in the HCM analysis model.

Saturation Flow Rates

In Chapters 19 and 20, it was assumed that the saturation headway or saturation flow rate reflecting prevailing conditions was known. A key part of the HCM model is a methodology for estimating the saturation flow rate of any lane group based on known prevailing traffic parameters. The algorithm takes the form:

$s_i = s_o N \prod_i f_i$ [22-1]

where:

S_i=saturation flow rate of lane group i underprevailing conditions, veh/hg,S

The HCM now provides 12 adjustment factors covering a wide variety of potential prevailing conditions. Each adjustment factor involves a separate model, some of which are quite complex. These are described in detail later in the chapter.

Note that the algorithm includes multiplication of the base saturation flow rate by the number of lanes in the lane group, *N*. This produces a *total* saturation flow rate for the lane group in question.

Capacity of a Lane Group

The relationship between saturation flow rates and capacities is basically the same as that presented in Chapter 18. The saturation flow rate is an estimate of the capacity of a lane group if the signal were green 100% of the time. In fact, the signal is only effectively green for a portion of the time. Thus:

$c_i=s_i(g_iC)$ [22-2]

where:

c_i=capacity of lane group i, veh/h,s_i=saturation flow rate of lane group i, ve

This formulation may not be used, however, for lane groups with permitted phasing. Such lane groups have other mechanisms that affect capacity, which will be described later.

The *v/c* Ratio

In signal analysis, the v/c ratio is often referred to as the "degree of saturation" and given the symbol "*X*." This is convenient, as the term "v/c" appears in many equations that can be more simply expressed using a single variable, *X*.

The v/c ratio, or degree of saturation, is a principal output measure from the analysis of a signalized intersection. It is a measure of the sufficiency of available capacity to handle existing or projected demands. Obviously, cases in which $v/c>1.00$ indicate a shortage of capacity to handle the demand. Care must be taken, however, in interpreting such cases, depending on how the v/c value was determined.

Capacity, which is difficult to directly observe in the field, is most often estimated using Equation 22-2. Measured demands are usually a result of counts of *departure flows*. Departure flows are counted because it is easier to classify them by movement as they depart the intersection. True demand, however, must be based on *arrival flows*. There are several different scenarios in which a v/c ratio can be in excess of 1.00:

- A departure count is compared to the capacity estimate using Equation 22-2. It is theoretically impossible to count a departure flow that is in excess of the true capacity of the lane group. In this case, obtaining a v/c ratio estimate > 1.00 (assuming the departure counts are accurate) represents an *underestimate* of the capacity of the lane group. The estimated saturation flow rate resulting from the HCM model is lower than the actual value being achieved.

- An arrival count is compared to an estimate of capacity using Equation 22-2. In this case, the arrival count (assuming it is accurate and complete) represents existing demand. A v/c ratio in excess of 1.00 indicates that queuing is likely to occur. If queues are, in fact, not observed, this is another indication that the capacity has been *underestimated* by the model. Note that arrival counts do not capture such demand elements as traffic diverted to other routes or repressed demand.

- A forecast future demand is compared to an estimated capacity using Equation 22-2. In this case, the forecast demand is always an arrival demand flow, and a v/c ratio in excess of 1.00 indicates that queuing is likely to occur, based upon the estimated value of capacity.

The key in all cases is that capacity is an *estimate* based upon nationally observed norms and averages. In any given case, the actual capacity can be either higher or lower than the estimate. In fact, actual capacity has stochastic elements and will vary over time at any given location and over space at different, but similar, locations.

Analytically, the v/c ratio for any given lane group is found directly by dividing the demand flow rate by the capacity. Another expression can, however, be derived by inserting Equation 22-2 for capacity:

Xi=vici=(v/s)i(g/C)i [22-3]

where:

Xi=degree of saturation (v/c ratio) for lanegroup i,vi=demand flow rate for (v/s)i=flow ratio for lane group i, and(g/C)i=green ratio for lane group i.

Since demands are eventually expressed as v/s ratios in the HCM model, the latter form of the equation is sometimes convenient for understanding the relationships.

Although the HCM does not define a capacity for the entire intersection, it does define a *critical* v/c *ratio* for the intersection. It is defined as the sum of the demands on critical lane groups divided by the sum of the capacities of critical lane groups, or:

$$X_C = \frac{\sum_i v_{ci}}{\sum_i \left(s_{ci} \times \frac{g_{ci}}{C}\right)} = \frac{\sum_i (v/s)_{ci}}{\sum_i (g_{ci}/C)} \quad [22\text{-}4]$$

where:

X_c=critical v/c ratio for the intersection,v_{ci}=demand flow rate for critical la

The term $\sum(g_{ci}/C)$ is the total proportion of the cycle length that is effectively green for all critical-lane groups. Since the definition of a critical-lane group is that one and only one such lane group must be moving during all phases, the only time a critical movement is *not* moving is during the lost times of the cycle. Thus, $\sum(g_{ci}/C)$ may also be expressed as:

$$\frac{C-L}{C}$$

where *L* is the total lost time per cycle. Inserting this into Equation 22-4 yields:

$$X_C = \frac{\sum_i (v/s)_{ci}}{\left(\frac{C-L}{C}\right)} = \sum_i (v/s)_{ci} \times \left(\frac{C}{C-L}\right) \quad [22\text{-}5]$$

As the value of X_c varies with cycle length, it is difficult to apply to future cases in which the exact signal timing may not be known. Thus, for analysis purposes, the HCM defines a value of X_c based on the maximum feasible cycle length, which results in the minimum feasible value of X_c. For pre-timed signals, the maximum feasible cycle length is usually taken to be 120 s, but this is sometimes exceeded in special situations. For actuated signals, longer cycle lengths are not as rare, and 150 s is usually used as a practical maximum. Equation 22-5 then becomes:

$$X_{Cmin} = \sum_i (v/s)_{ci} \times \left(\frac{C_{max}}{C_{max}-L}\right) \quad [22\text{-}6]$$

where:

X_{cmin}=minimum feasible v/c ratio, andC_{max}=maximum feasible cycle len

The latter value is more useful in comparing future alternatives, particularly physical design scenarios. The cycle length is assumed to be the maximum and is, in effect, held constant for all cases compared. Use of the maximum cycle length gives a view of the “best” critical v/c ratio achievable through signal timing, given the physical design and the phase plan specified.

It should be noted that editions of the HCM beyond 1997 have abandoned this concept. However, it remains a useful one for situations in which exact signal timings are not known.

The critical v/c ratio, X_c, is an important indicator of capacity sufficiency in analysis. If X_c is ≤ 1.00, then the proposed physical design, cycle length, and phase plan are sufficient to handle all critical demands. This *does not* mean that all lane groups will operate at $X_i \leq 1.00$. It does, however, indicate that all critical lane groups can achieve $X_i \leq 1.00$ by reallocating the green time within the existing cycle and phase plan. When $X_c > 1.00$, sufficient capacity may be provided only by taking one or more of the following actions:

- Increasing the cycle length
- Devising a more efficient phase plan
- Adding a lane or lanes to one or more critical lane groups

Increasing cycle length can add small amounts of capacity, as the lost time per hour is diminished. Devising a more efficient phase plan generally means considering additional left-turn protection or making a fully protected left turn a protected plus permitted left turn. It may also mean consideration of more complex phasing such as leading and lagging greens and/or exclusive LT phases followed by a leading green in the direction of heaviest left-turn flow. Chapters 19 and 20 contain full discussions of various phasing options.

In many cases, significant capacity shortfalls can be remedied only by adding one or more lanes to critical-lane groups. This increases the saturation flow rate and capacity of these lane groups, while the demand is constant.

22.1.4 Level-of-Service Concepts

and Criteria

Level of service is defined in the HCM in terms of *total control delay* per vehicle in a lane group. "Total control delay" is basically *time in queue* delay, as defined in Chapter 11, plus acceleration-deceleration delay. Level-of-service criteria are shown in Table 22.1.

Table 22.1: Level of Service Criteria for Signalized Intersections

Control Delay (s/veh)	LOS when $v/c \leq 1.0$	LOS when $v/c > 1.0$
≤10	A	F
>10–20	B	F
>20–35	C	F
>35–55	D	F
>55–80	E	F
>80	F	F

(Source: Reprinted with permission from *Highway Capacity Manual, 6th Edition: A Guide for Multimodal Mobility Analysis*, Transportation Research Board, the National Academy of Sciences, Courtesy of the National Academies Press, Washington, D.C., 2006.)

Table 22.1: Full Alternative Text

Note that *any* lane group operating at a v/c ratio greater than 1.00 is also labeled as LOS F. In effect, any signalized intersection lane group that has an average delay greater than 80 s/veh *OR* a v/c>1.00 is operating at LOS F.

As delay is difficult to measure in the field and because it cannot be measured for future situations, delay is estimated using analytic models, some of which were discussed in Chapter 18. Delay is not a simple measure, however, and varies (in order of importance) with the following measures:

- Quality of progression
- Cycle length
- Green time
- v/c ratio

Because of this, level-of-service results must be carefully considered. It is possible, for example, to obtain a result in which delay is greater than 80 s/veh (LOS F) while the v/c ratio is less than 1.00. Thus, at a signalized intersection, LOS F does not necessarily imply that there is a capacity deficiency. Such a result is relatively common for short phases (such as LT phases) in a long cycle length, or where the green splits are grossly out of sync with demands.

The reverse, however, does not lead to confusion. If v/c>1.00 for a short time—one 15-minute interval, for example—delay could be less than 80 s/veh, but must still be labeled LOS F.

Understanding the results of a signalized intersection analysis will require consideration of *both* the level of service and the v/c ratio for each lane group. Only then can the results be understood in terms of the sufficiency of the capacity provided and of the acceptability of delays experienced by road users.

22.1.5 Effective Green Time and Lost Time Concepts

The relationship between effective green times and lost times is discussed in detail in Chapter 18. In terms of capacity analysis, any given movement has effective green time, gi, and effective red time, ri. Figure 22.1

illustrates how these values are related to actual green, yellow, and red times in the HCM.

Figure 22.1: Effective Green Times and Lost Times in the HCM Model

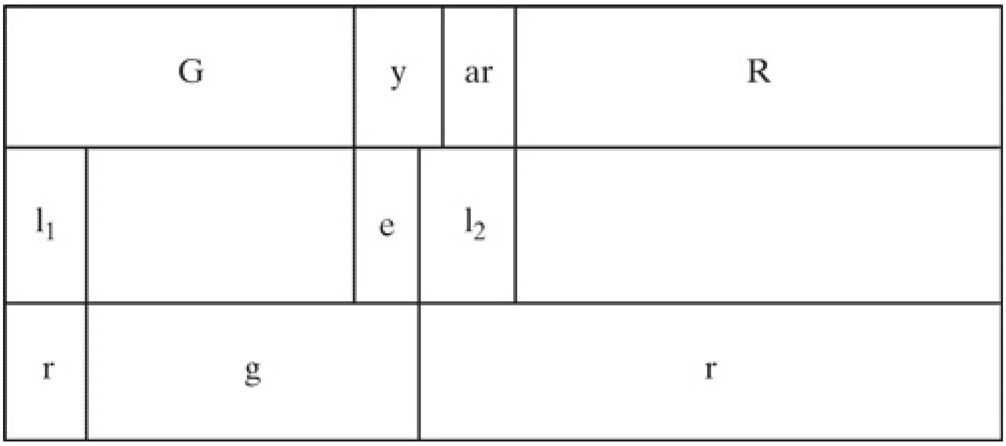


Figure 22.1: Full Alternative Text

Effective green and red times may be found as follows:

$g_i = G_i + Y_i + ar_i - \ell_1 - \ell_2 g_i = G_i + e - \ell_1$ $r_i = C - g_i$ [22-7]

where:

g_i=effective green time, phase i, s,G_i=actual green time, phase i, s,y_i=yello
up lost time, s/phase,l_2=clearance lost time, s/phase,e=extension of effectiv

Where there are overlapping phases, care must be taken in the application of lost times. When a movement continues into a subsequent phase, start-up lost time is assessed in the phase where the movement begins, while clearance lost time is assessed in the phase where the movement ends. For such a movement, there is NO lost time at the boundary between the phases serving the subject movement.

The case of a leading and lagging green phase with protected plus permitted left turns is illustrated in Figure 22.2. Eastbound (EB) movements begin in Phase 1a and continue in Phase 1b. The start-up lost time is only applied in Phase 1a and clearance lost time is applied at the end of Phase 1b. Westbound (WB) movements, however, begin in Phase 1b and have their start-up lost time applied there. Thus, in Phase 1b, the start-up lost time for WB movements is part of the effective green time for EB movements. As no movements begin in Phase 1c, no start-up lost times are assessed here, but there is clearance lost time applied for the WB movements. All northbound/southbound (NB/SB) movements flow in Phase 2; their lost times are assessed in this phase. Essentially, three sets of lost times are applied over four subphases. As effective green times affect capacity and delay, it is important that a systematic way of properly accounting for lost times be followed.

Figure 22.2: Lost Times Applied to Overlapping Phases

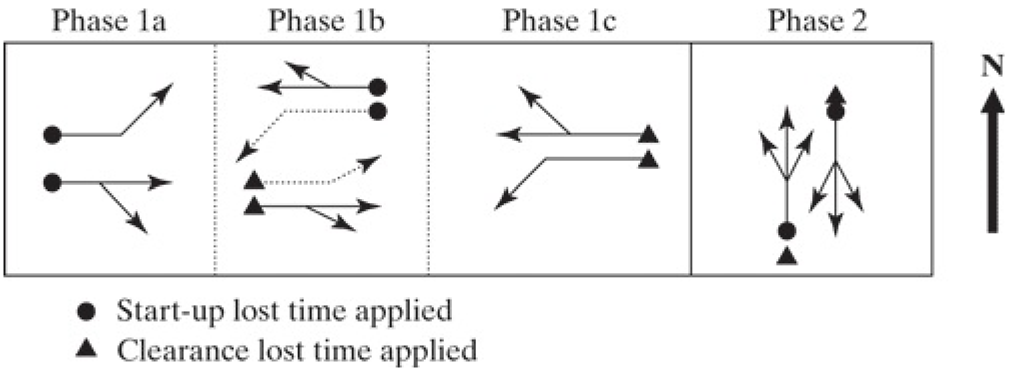


Figure 22.2: Full Alternative Text

22.1.6 Analysis Time Period

The basic time period for analysis recommended by the HCM remains a peak 15-minute period within the analysis hour, which is most often (but need not be) one of the peak hours of the day. Beginning with the HCM 2000, and continuing in 2010 and 2016, the HCM provides for some flexibility in this regard, recognizing that delay is particularly sensitive to the analysis period, especially when oversaturation exists.

There are three basic time options for analysis:

1. The peak 15 minute within the analysis hour
2. The full 60 minute analysis hour
3. Sequential 15 minute periods for an analysis period of one hour or greater

The first option is appropriate in cases where no oversaturation exists (i.e., no lane groups have v/c>1.00). This focuses attention on the worst period within the analysis hour, where 15 minute remains the shortest period during which stable flows are thought to exist. The second option allows for an analysis of average conditions over the full analysis hour. It could, however, mask shorter periods during which v/c>1.00, even though the full hour has sufficient capacity.

The third option is the most comprehensive. It requires, however, that demand flows be measured or predicted in 15-minute time increments, which is often difficult. It allows, however, for the most accurate analysis of oversaturation conditions. The initial 15-minute period of analysis would be selected such that all lane groups operate with v/c≤1.00, and would end in a 15-minute period occurring after all queues have been dissipated. During each 15-minute period, residual queues of unserved vehicles would be estimated and would be used to estimate additional delay due to a queue existing at the start of an analysis period in the subsequent interval. In this way, the impact of residual queues on delay and level of service in each successive period can be estimated.

22.2 Model Structure for Pre-timed Signals

The basic structure of the HCM model for signalized intersections is relatively straightforward and includes many of the conceptual treatments presented in Chapter 18. The model becomes more complex, however, when permitted or compound left turns are involved and/or when a movement can choose between two lane groups, for example, when an approach has a shared left/through lane and a through-only lane. This results in algorithms with many variables and several iterative aspects. It becomes particularly complex when actuated signals are analyzed because the average signal timing for the studied time period must be found. For simplicity and clarity of presentation, the model is first discussed within terms of its application to pre-timed signals.

In this section, the conceptual building blocks of the HCM procedure will be described and illustrated. The fundamental approach is presented without diversions into some of the more lengthy detailed calculations, such as for lane groups that serve more than one movement and actuated signals. Subsequent sections of the chapter will address these details. The structure of the HCM model is illustrated in Figure 22.3.

Figure 22.3: Flow Chart for Signal Analysis (Pre-timed Signals) – 2016 HCM

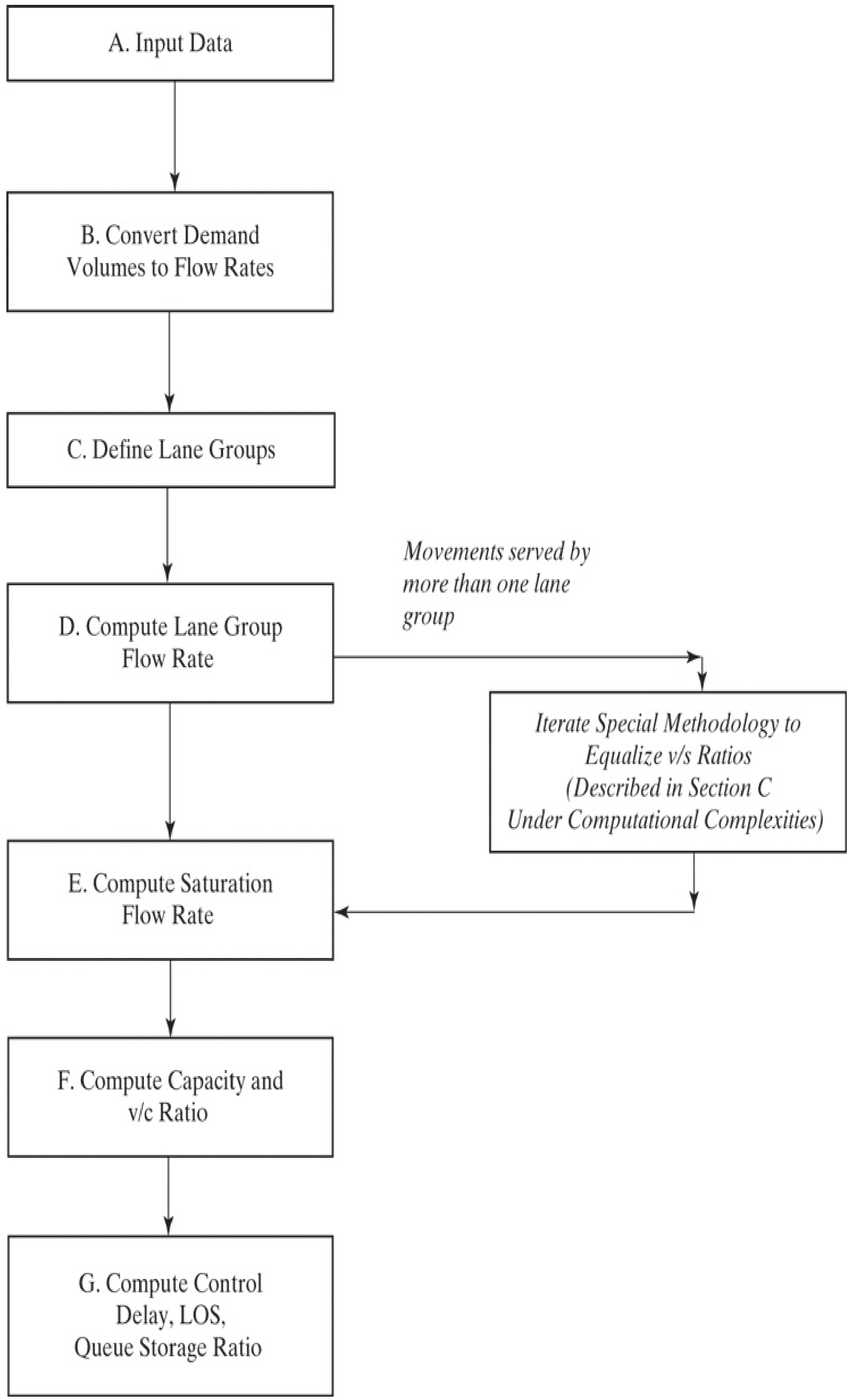

A. Input Data
B. Convert Demand Volumes to Flow Rates
C. Define Lane Groups
D. Compute Lane Group Flow Rate
Movements served by more than one lane group
Iterate Special Methodology to Equalize v/s Ratios (Described in Section C Under Computational Complexities)
E. Compute Saturation Flow Rate
F. Compute Capacity and v/c Ratio
G. Compute Control Delay, LOS, Queue Storage Ratio

Figure 22.3: Full Alternative Text

22.3 Computational Steps in the Model

As depicted in Figure 22.3, the analysis steps are described below.

22.3.1 Define Input Data

Before any of the analysis steps can be undertaken, all input data must be provided. The data include complete descriptions of traffic characteristics, roadway geometry, and signalization. For pre-timed signals, the signal cycle and all of its intervals must be specified. For actuated signals, controller settings are specified, and the methodology directly incorporates these settings into computing the average signal timing for the given conditions.

Table 22.2 summarizes all of the input data needed to conduct a full analysis of a signalized intersection. Most of the variables included in Table 22.2 have been previously defined. Others require some additional definition or discussion. The HCM also provides recommendations for default values that may be used in cases where field data on a particular characteristic is not available. Caution should be exercised in using these, as the accuracy of v/c, delay, and level of service predictions is influenced.

Table 22.2: Data Requirements for Signalized Intersection Analysis – Pre-timed Signals

Type of Condition	Parameter	Default Value
Geometric Conditions	Number of Lanes, N	Must be provided
	Lane Width, W (ft)	12 ft
	Number of Receiving Lanes	Must be provided
	Grade, G (%)	Flat 0%, Moderate 3%, Steep 6%
	Existence of LT or RT Lanes	Must be provided
	Length of Storage Bay for LT or RT lane (ft)	Must be provided
	Parking Conditions (Yes/No)	Must be provided
Traffic Conditions	Demand Volume (or flow rate) by Movement, V (vph)	Must be provided
	RTOR Flow Rate, (veh/h)	0
	Base Saturation Flow Rate, s_o (pc/hg/ln)	1,900 Metro pop $\geq$ 250 K, 1,750 otherwise
	Percent Heavy Vehicles, P_T (%)	3%
	Pedestrian Flow in Conflicting Crosswalk, v_p (peds/h)	Must be provided
	Local Buses Stopping at Intersection, N_B (buses/h)	12 CBD, 2 Non-CBD
	Parking Activity, N_m (maneuvers/h)	See Table 22-4.
	Proportion of Vehicles Arriving on Green, P	Provided if known
	Arrival Type, needed if P unknown	3
	Speed Limit (mi/h)	Must be provided
	Bicycle Flow Rate, v_{bic} (bicycles/h)	Must be provided
	Peak Hour Factor, PHF	Hourly data and 15-min analysis period: Total entering vol $\geq$ 1,000 vph: 0.92 Total entering vol $<$ 1,000 vph: 0.90 Otherwise: 1.00
Signalization Conditions	Cycle Length, C (s)	Must be provided
	Phase Plan	Must be provided
	Green Time per phase, G (s)	Major street through: 50 s; Minor through: 30 s; Left-turn phase: 20 s
	Yellow change + red clearance	4 s
	Pedestrian Push-Button per phase (Yes/No)	Must be provided
	Minimum Pedestrian Green per phase, G_p (s)	Based on 3.5 ft/s walking speed
Other	Analysis Period Duration (h)	0.25 h
	Area Type	Must be provided

(Source: Reprinted with permission from *Highway Capacity Manual, 6th Edition: A Guide for Multimodal Mobility Analysis*, Transportation Research Board, the National Academy of Sciences, Courtesy of the National Academies Press, Washington, D.C., 2016.)

Table 22.2: Full Alternative Text

Geometric Conditions

A full description of the intersection must be specified, as listed in Table 22.2. *Parking conditions* must be specified for each approach. For a typical two-way street, each approach either has curb parking or not. On a one-way street approach, parking may exist on the right and/or left side. For the purposes of the intersection, curb parking is noted only if it exists within 250 feet of the STOP line of the approach. Most of the other geometric conditions that must be specified are commonly used variables that have been defined elsewhere in this text.

Where wide lanes exist, some observation of their use should be made. Lanes of 18–20 feet often become two lanes under intense demand, particularly if it is a curb lane that could be used as a through lane plus a narrow RT lane. The analyst should try to characterize lanes as they would be used, not necessarily as they are striped. In most cases, however, these would be the same.

Traffic Conditions

There are a number of interesting variables that are used to describe the traffic conditions. The *Proportion of Vehicles Arriving on Green, P*, would be input only if known from field measurements. The proportion of vehicles arriving on green has a major impact on delay predictions but does not have significant influence on other portions of the methodology. If P is not known, then a general description of *arrival type* may be input, which will be used to estimate the value of P.

“Arrival type” is used to describe the quality of progression for vehicles arriving on each approach. There are six defined arrival types, 1 through 6, with AT 1 representing the worst progression quality and AT 6 representing the best progression quality. Definitions are given in Table 22.3.

Table 22.3: Arrival Types Defined

Arrival Type	Typical Signal Spacing (ft)	Conditions Under Which Arrival Type is Likely to Occur
1	≤1,600	Coordinated operation on a 2-way street where the subject direction does not receive good progression.
2	>1,600–3,200	A less extreme version of Arrival Type 1.
3	>3,200	Isolated signals or widely-spaced coordinated signals.
4	>1,600–3,200	Coordinated operation on a 2-way street where the subject direction receives good progression.
5	≤1,600	Coordinated operation on a 2-way street where the subject direction receives good progression.
6	≤800	Coordinated operation on a 1-way street in dense networks and CBDs.

(Source: Reprinted with permission from *Highway Capacity Manual, 6th Edition: A Guide for Multimodal Mobility Analysis*, Transportation Research Board, the National Academy of Sciences, Courtesy of the National Academies Press, Washington, D.C., 2016.)

Table 22.3: Full Alternative Text

The proportion of vehicles arriving on green is calculated from *AT* using Equation 22-8.

$P=AT\times(g/C)3$ [22-8]

where:

P=proportion of vehicles arriving on green,decimal,g=effective green time

If no field data are available, the descriptions shown in Table 22.3 may be used with knowledge of signal spacing to make a rough default estimate of arrival type.

Given the significant impact arrival type can have on delay estimates, it is important that a common arrival type be used when comparing different intersection designs and signal timings. High delays should not be simply dismissed or mitigated by assuming an improved progression quality.

Pedestrian flows in conflicting crosswalks must be specified for a signalized intersection analysis. The “conflicting crosswalk” is the crosswalk that right-turning vehicles turn through.

Parking activity is measured in terms of the number of parking maneuvers per hour into and out of parking spaces (Nm) located within 250 feet of the STOP line on the approach in question. Parking affects the adjacent lane group. Parking on a one-way approach may occur on both sides of the approach.

Movements into and out of parking spaces have additional negative impacts on operations over the frictional impacts of traveling in the lane adjacent to the parking lane. This is due to the lane adjacent to the parking lane being disrupted for some finite amount of time each time such a maneuver takes place. Parking activity should be observed in the field, but it is often not readily available. The HCM recommends the use of the default values shown in Table 22.4 in such cases, which assume 25-foot parking spaces and 80% occupancy.

Table 22.4: Recommended Default Values for Parking Activity

Street Type	Number of Parking Spaces within 250 ft of STOP Line	Parking Time Limit (h)	Turnover Rate (vph)	Maneuvers per Hour (N_m): Recommended Default Value
Two-Way	10	1	1.0	16
		2	0.5	8
One-Way	20	1	1.0	32
		2	0.5	16

(Source: Reprinted with permission from *Highway Capacity Manual, 6th Edition: A Guide for Multimodal Mobility Analysis*, Transportation Research Board, National Academy of Sciences, Courtesy of the National Academies Press, Washington, D.C., 2016.)

Table 22.4: Full Alternative Text

A *local bus* is defined as one that stops within the confines of the intersection (250 feet from stop line), either on the near side or far side of the intersection, to pick up and/or discharge passengers. A local bus that does not stop at the intersection is included as a heavy vehicle in the heavy vehicle percentage.

As noted previously, use of default values should be avoided whenever local field data or projections are available. For each default value used in lieu of specific data for the intersection, the accuracy of predicted operating conditions becomes less reliable.

Signalization Conditions

The analysis methodology uses an algorithm to determine the adequacy of pedestrian crossing times. It is similar to the approach in chapter 19 in principle, but somewhat different in specifics.

$Gp=3.2+0.27\ Npeds+(LSp)$ for $WE \le 10$ ft$Gp=3.2+2.7\ (NpedsWE)+(LSp)$ for $WE > 10$ ft [22-9]

where:

Gp=minimum green time for safe pedestriancrossings, s,L=length of the cr

For pre-timed signals, the HCM checks green times against these minimum values and issues a warning statement if pedestrian crossings are unsafe under these guidelines. Analysis, however, may continue whether or not the minimum pedestrian green condition is met by the signalization or not.

It should also be noted that many local and state agencies have their own policies on what constitutes safe crossings for pedestrians.

22.3.2 Convert Demand Volumes to Demand Flow Rates

It is most desirable to specify the demand flow rates by using actual counts from field data or projections with peak-hour factor (PHF)=1.0. If, however, demand is specified as hourly volumes, then they must be converted to flow rates. The PHF is calculated for the entire intersection using Equation 22-10 and demand flow rate is found using Equation 22-11.

PHF=v604v15 [22-10]
v=VPHF [22-11]

where:

PHF=peak-hour factor,v60=full-
hour volume of all vehicles enteringthe intersection (all approaches), veh,v
min volume of all vehiclesentering the intersection (all appro-
aches), veh,v=demand flow rate, veh/h, andV=demand volume, veh/h.

22.3.3 Define Lane Groups

There are six different types of lane groups than can exist on an intersection approach:

- Single-lane approach in which all three movements (LT, TH, RT) are made from one lane.

- Exclusive LT lane or lanes; multiple LT lanes form a single lane group.
- Exclusive RT lane or lanes; multiple RT lanes form a single lane group.
- Exclusive TH lanes (no turns from these lanes) form a single lane group.
- Shared LT/TH lane.
- Shared RT/TH lane.

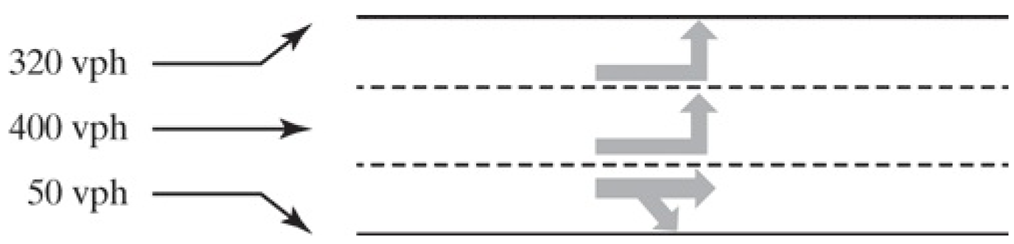
Each of these lane groups are analyzed separately, with saturation flow rate, capacity, and delay calculated for each.

22.3.4 Assign Demand Flow Rate

When a movement is served by only one lane group, assigning demand flow rate to that lane group is straight forward. The sample problems below illustrate the procedure under these conditions.

Sample Problem 22-1: Assigning Demand Flow Rates (1)

Define lane groups and assign lane group flow rate for the following approach. PHF=1.0



22.4-5 Full Alternative Text

Table 22.5 illustrates the demand flow assignment.

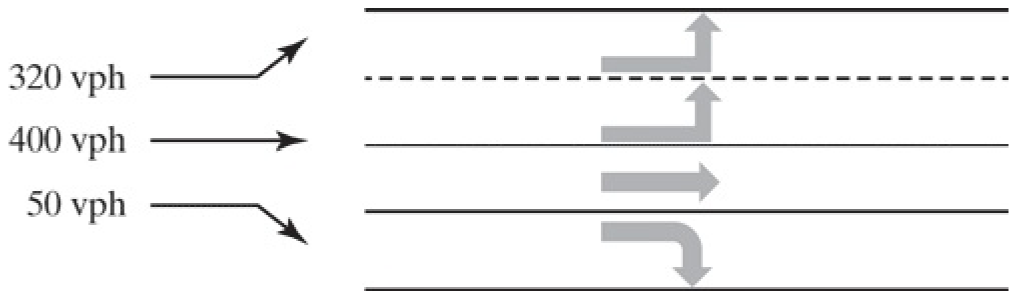
Table 22.5: Demand Flow Assignment for Sample Problem 22-1

Movement	Movement Flow Rate	Lane Group	Number of Lanes in Lane Group	Lane Group Flow Rate
LT	320	LT	2	320
TH	400	RT/TH	1	450
RT	50			

Table 22.5: Full Alternative Text

Sample Problem 22-2: Assigning Demand Flow Rates (2)

If the same approach had an exclusive right-turn bay and a through-only lane, what would be the lane group flows?



22.4-7 Full Alternative Text

Table 22.6 illustrates the demand flow assignment.

Table 22.6: Demand Flow

Assignment for Sample Problem 22-2

Movement	Movement Flow Rate	Lane Group	Number of Lanes in Lane Group	Lane Group Flow Rate
LT	320	LT	2	320
TH	400	TH	1	400
RT	50	RT	1	50

Table 22.6: Full Alternative Text

These Sample Problems were straightforward cases where the vehicle does not need to choose between lane groups. However, one of the complexities in HCM 2016 is the process for assigning demand flow rates to lane groups when a movement is served by two lane groups, such as an approach with a left/through shared and a through/right shared lane. Each lane is a separate lane group and the through movement can choose which lane group to enter. The HCM presents an iterative method for dividing the subject volume between two lane groups by equalizing the v/s ratios of the lane groups. This method is discussed later in the chapter.

As for many variables, the HCM states that field-measured values for lane group demand flow rates are preferred over the estimation procedure, and the method allows lane group demand volumes or flow rates to be specified as input.

22.3.5 Estimating the Saturation Flow Rate for Each Lane Group

Among the most detailed computations in the HCM is the estimation of the saturation flow rate for each of the defined lane groups in the intersection. The saturation flow rate is computed as:

s=soNfwfHVfHVgfpfbbfafLUfRTfLTfRpbfLpbfwzfmsfsp [22-12]

where:

s=saturation flow rate for the lane group,veh/h,so=base saturation flow rate
≥ 250,000, 1,900 pc/hg/ln;otherwise 1,750 pc/hg/ln, or locally cali-
brated value),N=number of lanes in the lane group, andfi=adjustment facto

Adjustment for Lane Width

While the standard lane width at a signalized intersection is 12 feet, research [5] has indicated that as long as lane width is between 10 and 12.9 feet, there is no impact on saturation flow rate or capacity. Thus, there are only three values for the lane width adjustment factor:

fw=0.96Lane width < 10 feetfw=1.0010 feet ≤ Lane width
≤ 12.9 feetfw=1.04Lane width > 12.9 feet

Lane widths under 10 feet are not recommended.

Adjustment for Heavy Vehicles and Grade

The adjustment for heavy vehicles and grades accounts for the difference in operation of a truck versus a passenger car as well as the effect of a grade on vehicle operation. For downgrades (G<0%), Equation 22-13 is used. For all other cases, Equation 22-14 is used.

fHVg=100−0.79PHV−2.07G100 [22-13]
fHVg=100−0.78PHV−0.31G2100 [22-14]

where:

fHV=heavy vehicle adjustment factor,PHV=percentage of heavy vehicles i

A “heavy vehicle” is any vehicle with more than four wheels touching the ground during normal operations. Heavy vehicles are not segmented into separate classes. Thus, they include trucks, recreational vehicles, and buses not stopping within the confines of the intersection. Buses that do stop within the confines of the intersection are treated as a separate class of vehicles: local buses.

Also note that in this equation, PHV is stated as a percentage. In other equations, it is usually stated in terms of a decimal proportion.

Adjustment for Parking Conditions

The parking adjustment factor involves two variables: parking conditions and movements, and the number of lanes in the lane group. If there is no parking adjacent to the lane group, the factor is, by definition, 1.00. If there is parking adjacent to the lane group, the impact on the lane directly adjacent to the parking lane is a 10% loss of capacity due to the frictional impact of parked vehicles, plus 18 seconds of blockage for each movement into or out of a parking space within 250 feet of the STOP line. Thus, the impact on an adjacent lane is:

$P=0.90-(18Nm3,600)$

where:

P=adjustment factor applied only to the laneadjacent to parking lane, andN

It is then assumed that the adjustment to additional lanes in the lane group is 1.00 (unaffected), or:

$fp=(N-1)+PN$

where:

N=number of lanes in the lane group

These two expressions are combined to yield the final equation for the parking adjustment factor:

$fp=N-0.1-(18Nm3,600)N\geq 0.05$ [22-15]

There are several external limitations on this equation:

- $0 \leq N_m \leq 180; \text{if} N_m > 180$, use 180 mvts/h
- $f_p(\text{min}) = 0.05$
- $f_p(\text{no parking}) = 1.00$

On a single-lane one-way street with parking on both sides, Nm is the total number of right- and left-parking maneuvers.

Adjustment for Local Bus Blockage

The local bus blockage factor accounts for the impact of local buses stopping to pick up and/or discharge passengers at a near-side or far-side bus stop within 250 feet of the near or far STOP line. Again, the primary impact is on the lane in which the bus stops (or the lane adjacent in cases where an off-line bus stop is provided). It assumes that each bus blocks the lane for 14.4 seconds of green time. Thus:

$$B = 1.0 - (14.4 N_B 3{,}600)$$

where:

B = adjustment factor applied only to the lane blocked by local buses, and N_B

As in the case of the parking adjustment factor, the impact on other lanes in the lane group is assumed to be nil, so that an adjustment of 1.00 would be applied. Then:

$$f_{bb} = (N - 1) + BN$$

where:

N = number of lanes in the lane group.

Combining these equations yields:

$f_{bb}=\frac{N-\left(\frac{14.4 N_B}{3{,}600}\right)}{N} \geq 0.05$ [22-16]

There are also several limitations on the use of this equation:

- $0 \leq N_B \leq 250$; if $N_B > 250$, use 250 b/h
- $f_{bb}(min)=0.05$

If the bus stop involved is a terminal location and/or layover point, field studies may be necessary to determine how much green time each bus blocks. The value of 14.4 seconds in Equation 22-16 could then be replaced by a field-measured value in such cases.

Adjustment for Type of Area

As noted previously, signalized intersection locations are characterized as "CBD" (central business district) or "Other," with the adjustment based on this classification:

- CBD location: $f_a=0.90$
- Other location: $f_a=1.00$

This adjustment accounts for the generally more complex driving environment of central business districts and the extra caution that drivers often exercise in such environments. Judgment should be exercised in applying this adjustment. Not all central business districts will have such complex environments that headways would be increased for that reason alone. Some non-CBD locations may have a combination of local environmental factors, which would make the application of this adjustment advisable.

Adjustment for Lane Utilization

The adjustment for lane utilization accounts for unequal use of lanes in a lane group with more than one exclusive lane. Where demand volumes can be observed on a lane-by-lane basis, the adjustment factor may be directly computed as:

fLU=vgvg1N [22-17]

where:

vg=demand flow rate for the lane group,veh/h,vg1=demand flow rate for th

When applied in this fashion, the factor adjusts the saturation flow rate downward so that the resulting v/c ratios and delays represent, in effect, conditions in the worst lane of the lane group. The HCM states that a lane utilization factor of 1.00 can be used "when uniform traffic distribution" can be assumed.

Default values for the lane utilization factor are shown in Table 22.7.

Table 22.7: Recommended Default Values for Lane Utilization

Movements Group	Number of Lanes in Movement Group (In)	Traffic in Most Heavily Traveled Lane (%)	Lane Utilization Adjustment Factor f_{LU}
Exclusive through	1	100.0	1.000
	2	52.5	0.952
	3[a]	36.7	0.908
Exclusive left turn	1	100.0	1.000
	2[a]	51.5	0.971
Exclusive right turn	1	100.0	1.000
	2[a]	56.5	0.885

(Source: Reprinted with permission from *Highway Capacity Manual, 6th Edition: A Guide for Multimodal Mobility Analysis*, Transportation Research Board, the National Academy of Sciences, Courtesy of the National Academies Press, Washington, D.C., 2016.)

Note: [a] If a movement group has more lanes than shown in this exhibit, it is recommended that field surveys be conducted or the smallest fLU value shown for that type of movement group be used.

Table 22.7: Full Alternative Text

Adjustment for Right Turns

The right-turn adjustment factor accounts for the fact that such vehicles have longer saturation headways than through vehicles, as they are turning on a tight radius requiring reduced speed and greater caution. This factor does not account for pedestrian interference with right-turning vehicles. Right turns occur under three different scenarios:

- From an exclusive RT lane
- From a shared lane
- From a single-lane approach

For right turns from an exclusive RT lane, the adjustment factor for right turns (fRT) is a constant 0.85, computed from:

fRT=1ER [22-18]

where:

ER=through-car equivalent for a right turningvehicle = 1.18.

For right turns from a shared TH/RT lane or a single-lane approach, the right-turn through-car equivalent, ER, is adjusted by the right-turn pedestrian and bicycle interference factor. The adjustment factor for right turns (fRT) is then computed from:

fRT=11+PR(ERfRpb−1) [22-19]

where:

PR=proportion of right turners in the shared lanefRpb=right-turn pedestrian and bicycleinterference factor

Because the adjustment for pedestrian and bicycle interference is incorporated into the shared-lane permissive phasing right-turn factor, it is removed from the factors that are multiplied times the base saturation flow rate in the calculation of adjusted saturation flow rate of Equation 22-12. The calculations for finding the fRpb factor are shown later in this section.

Adjustment for Left Turns

There are six basic situations in which left turns may be made:

- Case 1: Exclusive LT lane with protected phasing
- Case 2: Exclusive LT lane with permitted phasing
- Case 3: Exclusive LT lane with compound phasing
- Case 4: Shared lane with protected phasing
- Case 5: Shared lane with permitted phasing
- Case 6: Shared lane with compound phasing

All of these options are frequently encountered in the field, with the exception of Case 4, which exists primarily on one-way streets with no opposing flows.

As was the case for right turns, left-turning vehicles have a lower saturation flow rate than through vehicles due to the fact that they are executing a turning maneuver on a restricted radius.

For left turns from an exclusive LT lane with protected phasing, the adjustment factor for left turns (fLT) is a constant 0.95 computed from:

fLT=1EL [22-20]

where:

EL=through-car equivalent for a right turningvehicle = 1.05.

The reduction in saturation flow rate for protected left-turning vehicles is less than that for protected right-turning vehicles, as the radius of curvature involved in the maneuver is greater, that is, right-turning vehicles have a sharper turn than left-turning vehicles. (Compare $fLT=0.95$ to $fRT=0.85$.)

The left-turn adjustment factor (fLT) for protected left turns made from a shared lane is computed as:

$fLT=11+PL(EL-1)$ [22-21]

Left turns with permitted phasing have a saturation flow rate calculated from Equation 22-22.

$sp=voe-votc/36001-e-votf/3600$ [22-22]

where:

vo=opposing volume,tc=critical headway (or critical gap) is theminimum ti
up headway is the time betweenthe departure of one vehicle and the next(=:

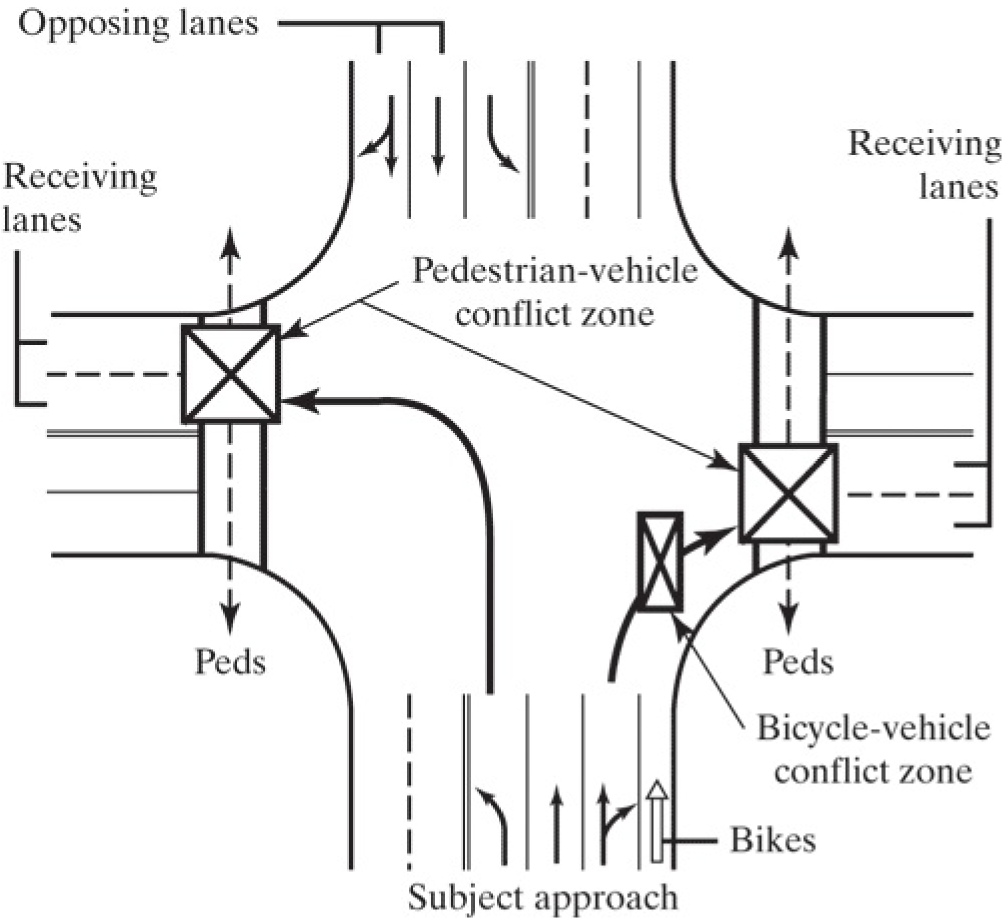
Left turns made on a one-way street turning into a one-way street are analyzed as right turns, since the turning radius is consistent with a right-turning movement.

Adjustments for Pedestrian and Bicycle Interference with Turning Vehicles

These two adjustment factors were added to the HCM in 2000 to account for the interference of both pedestrians and bicycles with right- and left-turning vehicles at a signalized intersection. The factor for interference with left turns, fLpb, only considers the interaction of left turns with pedestrians and not bicycles (even though the labeling of the factor suggests otherwise). The exception to this is on a one-way street intersecting with one-way street, when left-turns are analyzed as if they were right turns.

Figure 22.4 illustrates the conflicts between turning vehicles, pedestrians, and bicycles.

Figure 22.4: Pedestrian and Bicycle Interference with Turning Vehicles



(Source: Reprinted with permission from *Highway Capacity Manual, 6th Edition: A Guide for Multimodal Mobility Analysis*, Transportation Research Board, National Academy of Sciences, Courtesy of the National Academies Press, Washington, D.C., 2016.)

Figure 22.4: Full Alternative Text

Right-turning vehicles and left-turning vehicles from a one-way street encounter pedestrian and bicycle (right turns only) interference virtually immediately upon starting their maneuver. Left-turning vehicles from a two-way street with permitted phasing will encounter pedestrian interference after the opposing queue clears.

The basic modeling approach is to estimate the proportion of time that the pedestrian-vehicle and bicycle-vehicle conflict areas are blocked to vehicles (because they are occupied by pedestrians and/or bicyclists, who have the right of way). Only when these conflict zones are unblocked can vehicles travel through them.

The following computational steps are followed:

1. Step 1: Estimate Pedestrian Flow Rate during Green Phase (Left and Right Turns)

   This is the actual pedestrian demand flow rate during the green phase. The rate is adjusted to reflect the fact that pedestrians are moving only during the green phase of the signal.

   vpedg=vped(Cgped)≤5,000 [22-23]

   where:

   vpedg=flow rate for pedestrians during thegreen phase, pe

   The value of gped is taken to be the sum of the pedestrian walk and clearance intervals where they exist, or the length of the vehicular effective green where no pedestrian signals exist. Note that *two* values of this parameter are computed, one for pedestrians in the crosswalk in conflict with right turns and one for pedestrians in conflict with left turns.

   If the left turns are made in permitted or protected-permitted phasing, gped is substituted with the effective green of the permitted phase.

2. Step 2: Estimate the Average Pedestrian Occupancy in the Conflict Zone (Left and Right Turns)

“Occupancy” measures represent the proportion of green time during which pedestrians and/or bicycles are present in a particular area for which the measure is defined. The occupancy of pedestrians in the conflict area of the crosswalk is estimated as:

$OCCpedg=vpedg2,000 \text{ for } vpedg \le 1,000 OCCpedg=0.40+ (vpedg10,000) \le 0.90 \text{ for } 1,000<vpedg \le 5,000$ [22-24]

where:

OCCpedg=occupancy of the pedestrian conflictarea by pe

All other variables are as previously defined.

The first equation assumes that each pedestrian blocks the crosswalk conflict area for approximately 1.8 seconds, considering walking speeds and the likelihood of parallel crossings. At higher demand flows, the likelihood of parallel crossings is much higher, and each additional pedestrian blocks the crosswalk conflict area for another 0.36 seconds.

If the left turns have permitted or protected-permitted phasing, then the left turns must wait until the opposing queue clears before making a turn. The average pedestrian occupancy calculated above must be adjusted for this reduced time for conflicts between left-turn vehicles and pedestrians, using Equation 22-25.

$OCCpedu=OCCpedg(1-0.5gqgped)$ [22-25]

where:

OCCpedu=pedestrian occupancy after opposingqueue cle

Calculating gq is described later in the chapter.

3. Step 3: Estimate the Bicycle Flow Rate during the Green Phase (Right Turns Only)

   The bicycle flow rate during the green phase is found in the same way that the pedestrian flow rate during green was estimated:

$$v_{bicg}=v_{bic}\left(\frac{C}{g}\right)\leq 1900 \quad [22\text{-}26]$$

where:

v_{bicg}=bicycle flow rate during the greenphase, bic/hg,v_{bi}

4. Step 4: Estimate the Average Bicycle Occupancy in the Conflict Zone (Right Turns Only)

   Bicycle occupancy may then be estimated as:

   $$OCC_{bicg}=0.02+\left(\frac{v_{bicg}}{2700}\right) \quad [22\text{-}27]$$

   where:

   OCC_{bicg}=occupancy of the conflict area bybicycles.

   All other variables are as previously defined.

5. Step 5: Estimate the Relevant Conflict Zone Occupancy (Left and Right Turns)

   The occupancies computed in Steps 2 and 3 treat each element (pedestrians and bicycles) separately. Further, it is assumed that turning vehicles are present to block during all portions of the green phase.

   Equation 22-28 is used to estimate the conflict zone occupancy for right-turns with no bicycles present, and for left-turns from a one-way street:

   $$OCC_r=\left(\frac{g_{ped}}{g}\right)OCC_{pedg} \quad [22\text{-}28]$$

   where:

   OCC_r=occupancy of the conflict zone.

   For *right-turning vehicles*, where both pedestrian and bicycle flows exist, Equation 22-29 is used because the two interfering flows overlap, and simply adding the two occupancy values would result in too great an adjustment. The overlapping impact of pedestrians and bicycles on right-turning vehicles is quantified using Equation 22-29.

OCCr=(gpedgOCCpedg)+OCCbicg −(gpedgOCCpedgOCCbicg) [22-29]

where all terms have been previously defined.

The relevant conflict zone for permitted left turns is dependent on gaps available in the opposing volume after the opposing queue clears. It is found using Equation 22-30.

OCCr=(gped−gqgp−gq)(OCCpedu)e−5.00vo/3600 [22-30]

where:

gp=permitted green time, s, andvo=opposing demand flov

6. Step 6: Estimate the Unblocked Portion of the Phase, ApbT (Right and Left Turns)

   Once the occupancies of the conflict zone have been established, the unblocked time for conflict zones (one for right turns, one for left turns) are computed as follows:

   ApbT=1−OCCr   ifNrec=NturnApbT=1−0.6OCCr [22-31]

   where:

   Nrec=number of receiving lanes (lanes intowhich right- o turning movementsare made), andNturn=number of turnir turning movementsare made).

   All other variables are as previously defined.

Where the number of receiving lanes is equal to the number of turning lanes, drivers have virtually no ability to avoid the conflict area. Where the number of receiving lanes exceeds the number of turning lanes, drivers can maneuver around pedestrians and bicyclists to a limited extent.

7. Step 7: Determine Adjustment Factors

   The adjustment factors for interference of pedestrians and/or bicyclists with left-turn and right-turn movements can be determined as follows:

Adjustment Factor for Pedestrian and Bicycle Interference with Right Turns (fRpb)

- If there are no conflicting pedestrians or bicyclists, fRpb=1.0.
- If right turns are made as a protected phase, fRpb=1.0.
- If right turns are made with permitted phasing, fRpb=ApbT.
- If right turns are made with compound phasing, then fRpb=ApbT for the permitted portion of the phase and fRpb=1.0 for the protected portion of the phase.

Adjustment Factor for Pedestrian Interference

with Left Turns (f_{Lpb})

- If there are no conflicting pedestrians, $f_{Lpb}=1.0$.
- If left turns are made as a protected phase, $f_{Lpb}=1.0$.
- If left turns are made from an exclusive lane on a one-way street, $f_{Lpb}=A_{pbT}$.
- If left turns are made as a permitted phase, $f_{Lpb}=A_{pbT}$.
- If left turns are made from an exclusive lane on a one-way street with compound phasing, then $f_{Lpb}=A_{pbT}$ for the permitted portion of the phase and $f_{Lpb}=1.0$ for the protected portion of the phase.

Adjustment for the Presence of a Work Zone

The adjustment for the presence of a work zone accounts for closure of one or more lanes due to the presence of a work zone within 250 feet upstream of the stop line. The factor is calculated as follows:

$f_{wz}=0.858 f_{wid} f_{reduce} \leq 1.0$ [22-32]
$f_{wid}=11-0.0057(a_w-12)$ [22-33]
$f_{reduce}=11-0.0402(n_o-n_{wz})$ [22-34]

where:

f_{wid}=adjustment factor for approach width,where approac

The work zone presence factor computed is applied to *all* lane groups on the approach.

Adjustments for Downstream Lane Blockage and Sustained Spillback

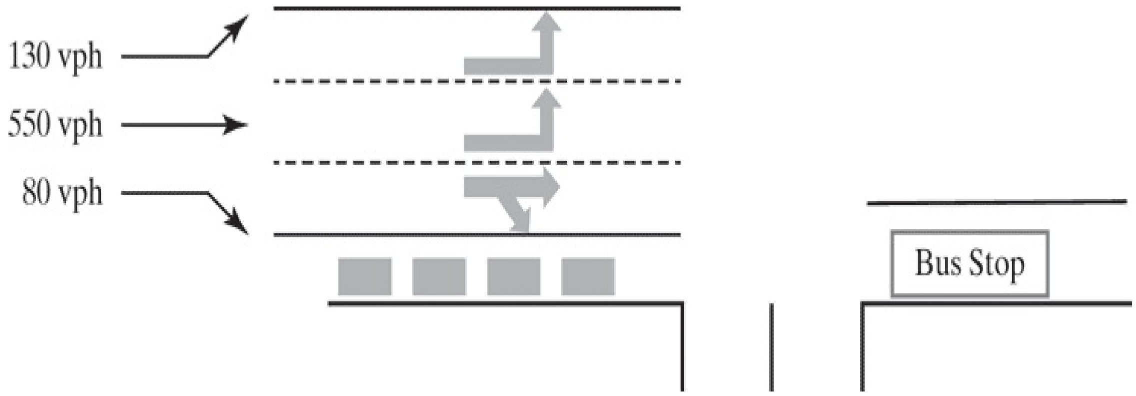
These last two adjustments to saturation flow rate are based on HCM procedures for an urban street segment analysis. They are rarely used when analyzing a single intersection, and are not discussed here.

Summary and Sample Problems

The estimation of the saturation flow rate for each defined analysis lane group can require many calculations. It is important to understand the basic relationships and concepts of each factor.

Sample Problem 22-3: Estimating Saturation Flow Rate (1)

Find the saturation flow rate for lane groups on the EB approach shown below. Cycle Length=60sec;g=gped=35sec



22.4-10 Full Alternative Text

- Protected left-turn phasing
- 200 ped per hour in the conflicting crosswalk
- 11-foot left-turn lanes
- 13-foot TH/RT lane
- 3% heavy vehicles
- 1% grade
- Parking on right curb with no vehicles moving in or out of parking spots during analysis period
- Local bus stop on far-side of the subject approach with 15 buses stopping per hour
- Non-CBD location
- No bicycles

Solution

There are two lane groups to examine on the subject approach: the two left-turn lanes form one lane group and the through/right-turn lane is a separate lane group.

Adjustment factors are found from the equations, tables, and discussions presented previously, as follows:

fwLT=1.00; fwTR=1.04fHVg=100−0.78PHV
−0.31G100=100−(0.78×3)−(0.31×1)100=0.974fpTR=N
−0.1−(18 Nm3,600)N=1−0.1−(18×03,600)1=0.90fbbTR=N
−(14.4NB3,600)N=1−(14.4 ×153,600)1=0.94fa=1.0 (non-
CBD location)fLULT=0.971 (Table 22.7); fLURT=1.000

Determining the pedestrian/bicycle interference factor involves several steps:

vpedg=vped(Cgped)=100(9035)=515peds/hg(Eq23−17)
OCCpedg=vpedg2000=5152000=0.257 (Eq 24−18)OCCr=gpedg×OCCpec
fRT=11+PR ((ER/fRpb)−1)=11+(80/630)
((1.18/0.85)−1)=0.953   fLT=0.95 (Excl LT w/Protected Phase)

All other factors do not apply to this case and are, by definition, 1.0. Note that because the EB TH/RT pedestrian-bicycle interference factor is incorporated into the right-turn factor, the value for fRpb for that lane group is removed from the table. Results for the lane group saturation flow rate are illustrated in Table 22-8.

Table 22.8: Computation of Saturation Flow Rates for Sample Problem 22-3

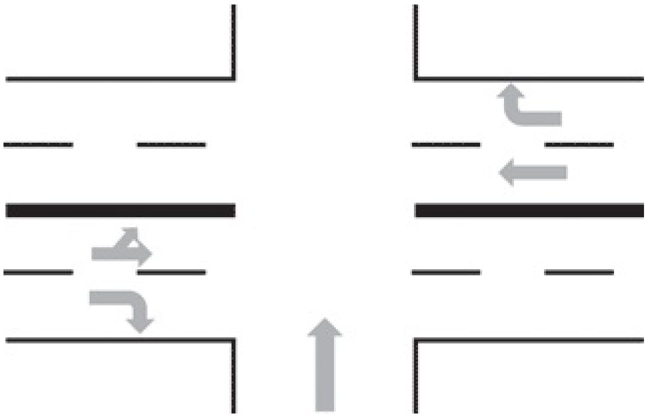
Lane Group	s_o	f_w	f_{HVg}	f_p	f_{bb}	f_a	f_{LU}	f_{RT}	f_{LT}	f_{Lpb}	f_{Rpb}	s (veh/h/l)	s (veh/h)
EB LT	1900	1.00	0.974	1.00	1.00	1.00	0.971	1.00	0.95	1.00	1.00	1707	3414
EB TH/RT	1900	1.04	0.974	0.90	0.94	1.00	1.00	0.953	1.00	1.00	–	1551	1551

Table 22.8: Full Alternative Text

Sample Problem 22-4: Estimating

Saturation Flow Rate (2)

Find the saturation flow rate for the *EB* lane groups on the intersection shown below.



22.4-12 Full Alternative Text

- Demand flows for EB and WB approaches: 300 TH, 50 RT
- Permitted left-turn phasing
- No parking, no buses, no ped, no bicycles
- 12-foot lanes
- 5% heavy vehicles
- CBD location

Solution

There are two lane groups on the subject EB approach: a shared LT/TH lane and a right-turn lane.

For the shared lane, find the saturation flow rate for a permitted left-turn phase, using Equation 22-22.

$$s_p=\frac{350e^{-350 \times 4.5/3600}}{1-e^{-350 \times 2.5/3600}}=1047\text{vph}$$

The adjustment to saturation flow rate for heavy vehicles is found using Equation 22-14.

fHVg=100−0.78×5−0.31×02100=0.96

- Adjustment for a CBD location, fa=0.90
- Adjustment for an exclusive right-turn lane, fRT=0.85

Results for the EB lane groups are illustrated in Table 22.9.

Table 22.9: Computation of Saturation Flow Rates for Sample Problem 22-4

Lane Group	s_o	f_{HVg}	f_a	f_{RT}	*s (veh/h/ln)*	*s (veh/h)*
LT/TH	1047	0.96	0.9	1.00	904	904
RT	1900	0.96	0.9	0.85	1395	1395

Table 22.9: Full Alternative Text

22.3.6 Determine Lane Group Capacities and *v/c* Ratios

At this point in the analysis, lane groups have been established, demand flow rates, v, for each lane group defined, and saturation flow rates, *s*, for each lane group have been estimated. As a result, both the demand and saturation flow rates for each lane group have been adjusted to reflect the same prevailing conditions. The ratio of v to *s* for each lane group can be computed and can be used as the variable indicating the relative demand intensity on each lane group.

Several important analytic steps may now be accomplished:

1. The v/s ratio for each lane group is computed.

2. Relative v/s ratios are used to identify the critical lane groups in the phase plan.

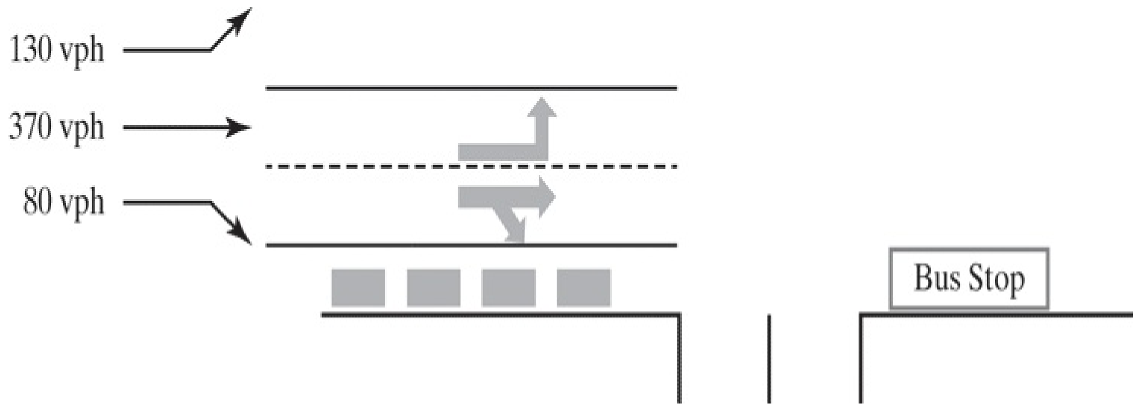
3. Lane group capacities are computed, using Equation 22-2: ci=si(giC). This equation cannot be used for left-turn lane groups with permitted or protected-permitted phasing. Capacity, when there is permitted left-turn phasing, is affected by other factors, such as finding gaps in the opposing direction’s traffic flow. The methodology for permitted left-turn phasing is described later in the chapter.

4. Lane group v/c ratios are computed (Equation 22-3): Xi=vici=(v/s)i(g/C)i

5. The critical v/c ratio for the intersection is computed (Equation 22-5): XC=∑i(v/s)ci×(CC−L)

Sample Problem 22-5: Estimating Capacity and v/c Ratio

For the values shown below, find g/C ratio, capacity, and v/c ratio for the LT and TH/RT lane groups.



22.4-14 Full Alternative Text

- Cycle length=90 seconds

- Leading left-turn phase=20 seconds

- Through/right phase=35 seconds
- l1=2 sec,e=3 sec,Y=4 seconds
- sLT=1707 veh/hg;sTH/RT=1,354 veh/hg
- 200 pedestrians per hour
- 3.5 feet/sec walking speed
- 68 feet crossing length

Check if this phase timing is enough to serve the pedestrians.

Solution

1. Check that timing meets minimum time needed for pedestrians:

   npeds=2003600×90=5peds/cycleGp=3.2+0.27 ×5+(683.5)=24s

   TH/RT phase is 35 seconds, so minimum is met.

2. Calculate effective green time:

   LT lane group: 20+4−2−(4−3)=21 secondsTH/RT lane group: 35+4−2

   The results for the east bound lane groups are shown in Table 22-10.

Table 22.10: Calculations for Sample Problem 22-5

Lane Group	s (veh/h)	v (veh/h)	v/s	g/C	$c(veh/h) = s \times g/C$	$v/c = v/c$
LT	1707	130	0.08	0.23	393	0.33
TH/RT	1354	450	0.33	0.40	542	0.83

Table 22.10: Full Alternative Text

3. Table 22.11 shows the results for all other approaches and lane groups at the intersection. From these, find the critical v/c ratio for the intersection, X_c. The NB and SB phases occur concurrently.

Table 22.11: Results for Sample Problem 22-5

Lane Group	*v/s*	*g/C*	*C*	*v/c*
EB LT	0.08	0.23	322	0.33
EB TH/RT	0.33	0.36	481	0.83
WB LT	0.11	0.19	797	0.45
WB TH/RT	0.26	0.36	541	0.64
NB LTR	0.18	0.36	281	0.71
SB LTR	0.12	0.36	262	0.48

Table 22.11: Full Alternative Text

In order to calculate the critical v/c ratio, identification of critical lane groups is needed. In Chapter 16, critical lanes were identified by finding the critical path through the signal ring diagram that resulted in the highest sum of critical-lane volumes, Vc. The method here is exactly the same, except that instead of adding critical-lane volumes, critical v/s ratios are added.

The critical movements for this intersection are those with the higher v/s ratio in each phase. For the left-turn phase, the WB left has the higher v/s ratio. For the EB/WB TH/RT phase, the EB movement is highest. For the NB/SB phase, the NB movements are highest. Thus the sum of the critical v/s ratios is:

WB LT+EB TH/RT+NB LTR=0.11+0.33+0.18=0.62Lost time/phase=l1+(Y−e)=2+(4–3)=3s/phaseTotal lost time, L=3×3=9s

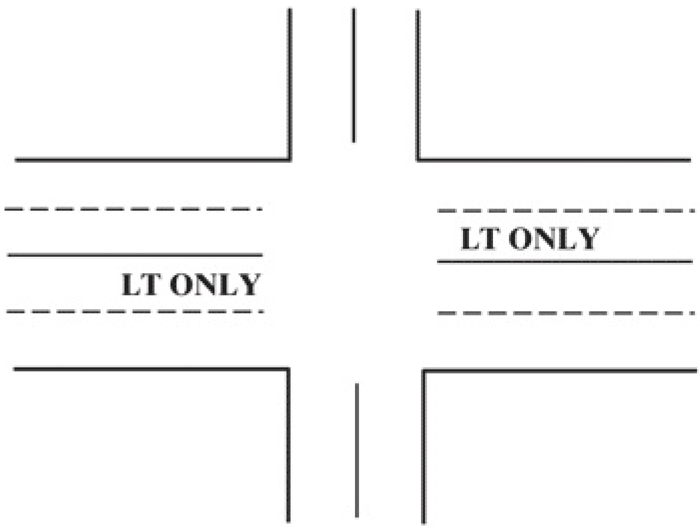
Critical v/c, Xc:

$XC=\sum i(v/s)ci\times(CC-L)=0.62\times9090-9=0.69$

Sample Problem 22-6: Determining Critical Movements

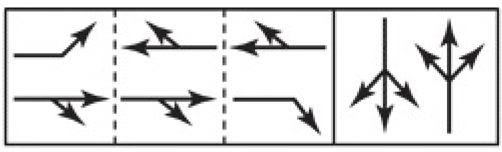
For the intersection shown in Figure 22.5, *find the critical movements*.

Figure 22.5: Determining Critical Lane Groups Using v/s Ratios



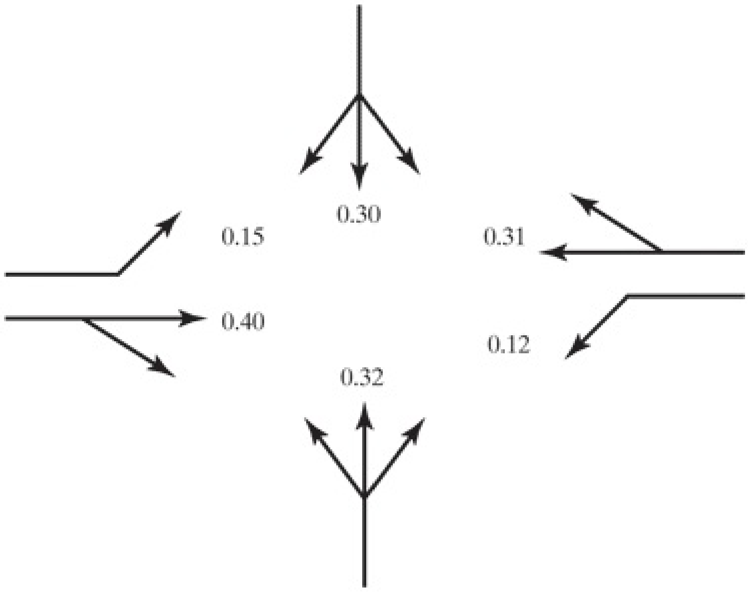
(a) Geometry

22.4-17 Full Alternative Text



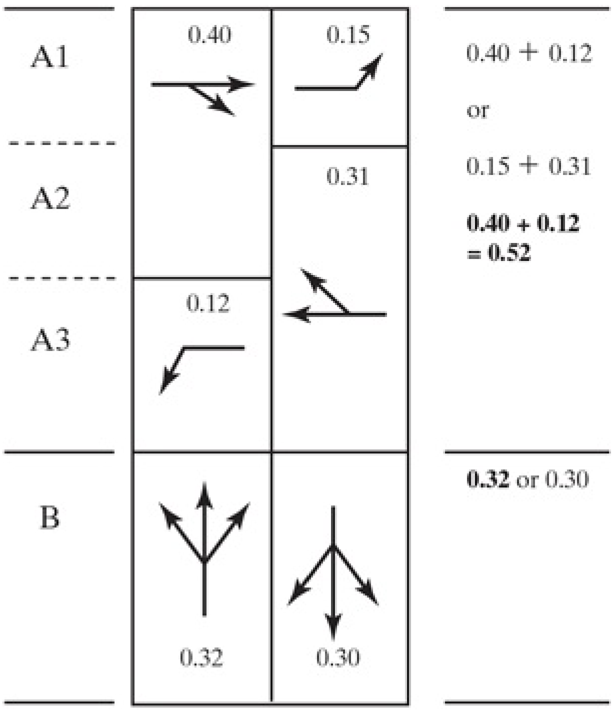
(b) Phase Diagram

22.4-17 Full Alternative Text



(c) Lane Group v/s Ratios

[22.4-17 Full Alternative Text]



$$\Sigma(v/s)_{ci} = 0.52 + 0.32 = 0.84$$

(d) Ring Diagram

[22.4-17 Full Alternative Text]

The illustration shows a signal with leading and lagging green phases on the E–W arterial and a single phase for the N–S arterial.

Solution

The critical path through Phases A1 through A3 is determined by which ring has the highest sum of v/s ratios. In this case, for the A phases, the left ring has the highest total, yielding a sum of v/s ratios of 0.52.

The critical movement in Phase B is a straightforward comparison of the two rings, which have concurrent phases. The highest total again is on the left ring, with a v/s of 0.32.

Thus, the critical path through the signal is entirely along the left ring, and the sum of critical lane v/s ratios is $0.52+0.32=0.84$.

22.3.7 Estimating Delay and Level of Service

Levels of service are based on control delay, as discussed previously. Specific criteria were given in Table 22.1. In the analysis of capacity, values of the v/c ratio for each lane group will have been established. Using these results, and other signalization information, the delay for each lane group may be computed as:

$$d=d_1+d_2+d_3 \quad [22\text{-}35]$$

where:

d=average control delay per vehicle, s/veh,d_1=average uniform delay per v

Uniform Delay

As discussed in Chapter 18, uniform delay can be obtained using Webster's uniform delay equation, Equation 22-36. This equation is only accurate for exclusive lanes that are served with protected-only phase.

$d1=0.5C[1-gC]21-[\min(1,X)]\times gC$ [22-36]

where:

C=cycle length, s,g=effective green time for lane group, s, and$X$=v/c ratio f

Effect of Progression

Webster’s uniform-delay equation assumes that arrivals are uniform over time. In fact, arrivals are at best random and are most often platooned as a result of coordinated signal systems. The quality of signal coordination or progression can have an immense impact on delay.

Consider the following situation: An approach to a signalized intersection is allocated 30 seconds of effective green out of a 60 seconds cycle. A platoon of 15 vehicles at exactly 2.0 seconds headways is approaching the intersection. Note that the 15 vehicles will exactly consume 30 seconds of effective green time available (15 × 2.0=30). Thus, for this signal cycle, the v/c ratio is 1.0.

With perfect progression provided, the platoon arrives at the signal just as the light turns green. The 15 vehicles proceed through the intersection with no delay to any vehicle. In the worst possible case, however, the platoon arrives just as the light turns *red*. The entire platoon stops for 30 seconds, with every vehicle experiencing virtually the entire 30 seconds of delay. When the green is initiated, the platoon fully clears the intersection. In both cases, the v/c ratio is 1.0 for the cycle. The delay, however, could vary from 0 s/veh to almost 30 s/veh, dependent solely upon when the platoon arrives (i.e., the quality of the progression).

Thus when using this equation, it must be adjusted by a progression factor, which accounts for the effect of progression. This is accomplished by adjusting the uniform delay calculated in Equation 22-36 by the progression factor found with Equation 22-37.

$PF=1-P1-gC\times 1-y1-\min(1,X)P\times[1+y1-PC/g1-g/C]$ $y=\min(1,X)\times g/C$ [22-37]

where:

PF=progression factor,y=flow ratio, v/s, andP=proportion of vehicles arrivi (AT3)(gC)

Sample Problem 22-7: Estimating Uniform Delay

Find uniform delay for one through-only lane with the following characteristics:

- $C=90$ seconds
- $g=40$ seconds
- $v=700$ veh/h
- $s=1800$ veh/h/ln
- Arrival type=4

Solution

- $g/C=40/90=0.44$
- $v/s=700/1800=0.389$
- $X=0.389/0.44=0.875$
- $P=(4/3)\times 0.444=0.59$

Calculate *PF*:

PF=1−0.591−0.44×1−0.3891−0.875×0.59×[1+0.3891−0.59×90/401−40/90]=0.716

Calculate d_1:

d1=0.5C [1−gC]21−[min(1,X)]×gC×PF=0.5×90 (1−0.44)21−0.875×0.44×0

The Incremental Queue Accumulation Approach for Calculating Uniform Delay, d_1

Webster's equation can only accurately predict uniform delay of a lane group that serves only one movement with protected-only phasing, such as a through-only lane group or a left-turn or right-turn exclusive lane group served by protected-only phasing. For all other lane groups, uniform delay is calculated using a queue accumulation polygon with the incremental queue accumulation (IQA) methodology, as documented in References [6] and [7]. The IQA method does not limit the shape of the queue accumulation diagram to a simple triangle, but rather can describe all types of complex phasing. The area of the resulting polygon is then found by breaking the polygon into component triangles and trapezoids for which the area can be found.

When finding uniform delay using IQA, the effect of progression is built into the methodology. To do this, it is necessary to separate the arrival flow rate of the lane group into the rate arriving on effective green, and the rate arriving on effective red. The parameter "P" (proportion of vehicles arriving on green) is either estimated from field measurements or estimated using Equation 22-38 for a known arrival type.

P=(AT3)(gC) [22-38]

The analysis begins by creating a queue accumulation polygon in which arrival and departure curves are constructed for the effective red portion of the phase and the effective green portion of the phase. This construction most often begins with effective red, because that is the time when, for an unsaturated analysis, there is no queue.

Note that each polygon deals with three queues:

- q1=size of queue at the beginning of the effective red phase, resulting from unserved demand on the previous red phase(s), veh, usually zero for unsaturated intersections,

- q_2=size of the queue at the end of the effective red phase (and beginning of effective green phase), veh, and

- q_3=size of the queue at the end of the effective green phase (and beginning of the next effective red phase), veh.

In a series of phases, q_3 at the end of the first computational cycle becomes q_1 for the beginning of the next.

The following steps determine the uniform delay using IQA.

1. The arrival rate during the effective red is given by:

   $$v_r=(1-P)\frac{V C}{r} \quad [22\text{-}39]$$

   where:

   v_r=arrival flow rate during effectivered, veh/h,P=proportion of vehicle

   It is assumed that the queue at the beginning of effective red is zero, $q_1=0$.

2. The queue at the end of the effective red time is found by:

   $$q_2=q_1+\left(\frac{v_r-s}{3600}\right)\times r \quad q_2\geq 0 \quad [22\text{-}40]$$

   where:

   q_1=queue at the beginning of effec-
   tive red, veh,q_2=queue at the end of effectivered, veh,v_r=average arriv

3. The uniform delay during the effective red time can then be found using Equation 22-41:

   $$d_r=r\times\left(\frac{q_1+q_2}{2}\right) \quad [22\text{-}41]$$

   where:

   d_r=uniform delay during effectivered, s.

   Note that this *does not include* delay due to an initial queue, nor delay to vehicles in a residual queue.

4. The same steps must be repeated to find the uniform delay during effective green time. The starting point, however, is the queue at the end of the effective red time, q2. The arrival rate during the effective green time is found by:

   vg=VP(gC)=VPCg [22-42]

   where:

   vg=average arrival rate duringeffective green, veh/h,g=effective green

5. Find the queue at the end of effective green time.

   q3=q2+(vg−s3600)×g [22-43]

6. For an undersaturated condition, the queue calculated with Equation 22-43 will most often be negative because the queue will clear before the end of the effective green time. After the queue clears, arriving vehicles experience no delay; thus, in order to find delay during effective green, it is necessary to find the time when the queue is fully dissipated, that is, when q3 is equal to “0.” This is the time when the number of vehicles in the queue is equal to the number of departures. The number of vehicles in the queue consists of the queue at the end of effective red plus vehicles that join the queue before it is fully processed.

   To find the time where q3=0, the queue at the start of effective green plus the number of vehicles joining the queue during the time until q3=0 is set equal to the number of departures:

   q2+(vg3600)×Δt2=(s3600)×Δt2 [22-44]

   where:

   Δt2=time during effective greenwhen the queue reaches 0.

   The left side of Equation 22-44 represents the vehicles to be processed during effective green (equals the queue at the end of effective red, q2, plus the vehicles that arrive during the effective green and join the queue which is cleared at time Δt2). The right side of Equation 22-44 represents the number of departures during the portion of effective green, $\Delta t2$, until the queue=0. The departure rate during effective green is the adjusted saturation flow rate, *s*, for the lane group. Thus, the number of departures until the queue is zero is *s/3600* × $\Delta t2$. Solving Equation 22-45 for $\Delta t2$:

$$\Delta t2=\frac{3600q2}{s-vg} \quad [22\text{-}45]$$

The delay during time $\Delta t2$ is found using Equation 22-46:

$$dg=\Delta t2\times\left(\frac{q2+q3}{2}\right) \quad [22\text{-}46]$$

where: dg=Uniform delay during the effective green time, s.

All other terms are as previously defined.

7. Uniform delay (s/veh) is then the sum of the uniform delays incurred during the effective red and effective green phases divided by the total number of vehicle arrivals in the cycle:

$$d1=\frac{(dr+dg)}{(q2+na)} \quad [22\text{-}47]$$

where:

na=number of vehicle arrivals duringeffective green $=\frac{vg}{3600}\times g$.

All other terms are as previously defined.

Note that uniform delay is the average delay experienced by all vehicles in the cycle, including those that arrive after the queue clears and experience no delay.

For a more complicated phasing, these same steps are repeated for each component of the total polygon. The components of the polygon are determined by separating the phase into polygons where the arrival rate and departure rate are constant. Arrival rates may change due to the effect of platooning. Departure rates will change due to the phase going from red to green, of course, but also may change when the green time is used by different movement, such as when there is permitted phasing for left-turn vehicles. In cases of protected plus permitted (or permitted plus protected) phasing, the saturation flow rate is different for the permitted and protected portions of the phase. Therefore, while the shape of the polygons can get very complicated, the basic principles are relatively simple. References 8 and 9 provide additional discussion of IQA.

Sample Problem 22-8: Using IQA to Determine Uniform Delay

Find uniform delay, using IQA, for the same through-only lane group as in Sample Problem 22-7*. The characteristics are repeated here for convenience.*

- $C=90$ seconds
- $G=40$ seconds
- $l1=2$ seconds
- $e=2$ seconds
- $v=700$ vph
- $s=1800$ vphpl
- Arrival type=4

Solution

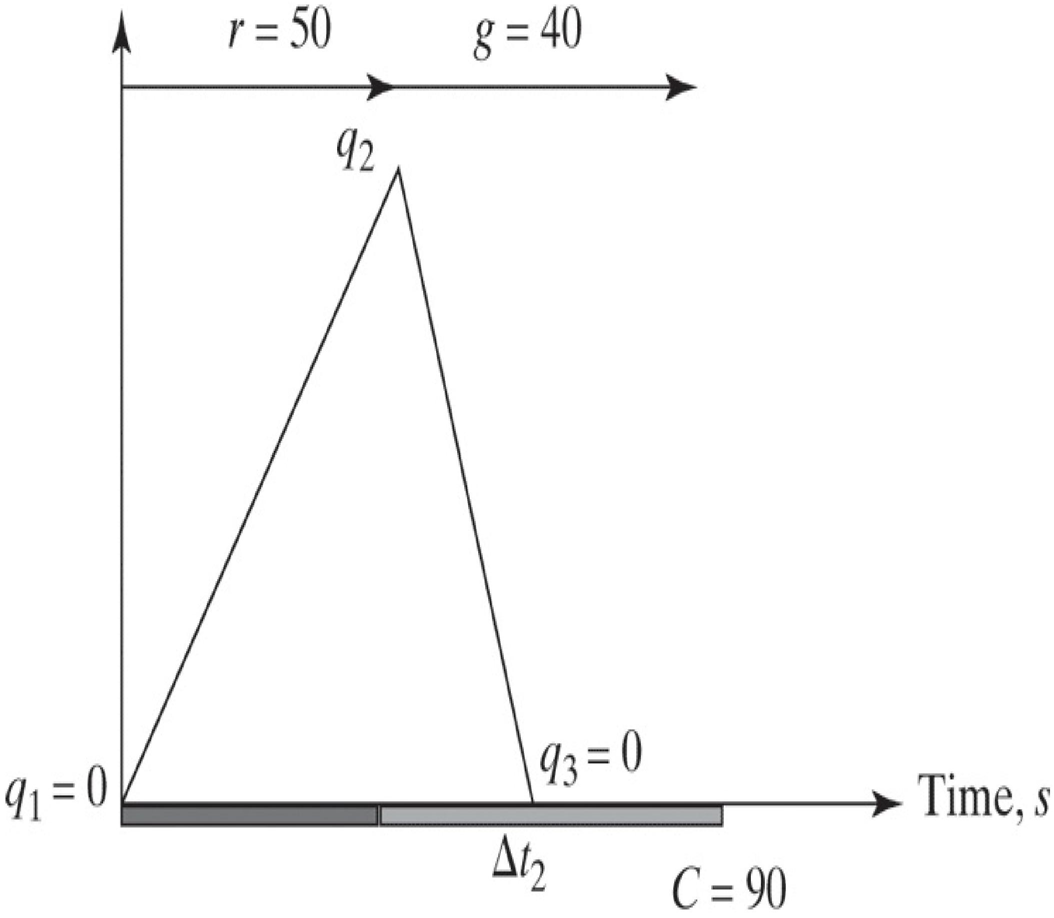
1. Calculate effective green time:

   $g=G-l1+e=40-2+2=40$ s

2. Calculate effective red time:

   r=90–40=50 seconds

3. Draw queue accumulation diagram:



22.4-17 Full Alternative Text

4. Calculate proportion of vehicles arriving on green:

   $P=(AT3)(gC)=(43)(4090)=0.59$

5. Calculate arrival rate on red:

   $vr=(1-P)VCr=(1-0.59)700\times9050=515$ veh/h

6. Calculate queue at end of effective red, assuming no queue at start of effective red:

   $q2=q1+(vr-s3600)\times r=0+(515-03600)\times50=7.16$ veh

7. Calculate delay during effective red:

   $dr=r\times(q1+q22)=50\times(0+7.162)=178.9$ veh−s

8. Calculate arrival rate on green:

$$Vg=VPCg=700\times 0.59\times 9040=931 \text{ veh/h}$$

9. Calculate queue at end of effective green:

   $$q3=q2+(vg-s3600)\times g=7.16+(931-18003600)\times 40=-2.5 \text{ veh}$$

   This is negative because the queue clears before the end of effective green.

10. Calculate time during effective green when the queue clears:

    $$\Delta t2=3600q2s-Vg=3600\times 7.161800-931=29.6 \text{ s}$$

11. Calculate delay during effective green:

    $$dg=\Delta t2\times(q2+q32)=29.6\times(7.16+02)=106.1 \text{ veh}-\text{s}$$

12. Calculate the total number of arrivals during effective green:

    $$na=vg3600\times g=9313600\times 40=10.3 \text{ veh}$$

13. Calculate average uniform delay:

    $$d1=(dr+dg)(q2+na)=(178.9+106.1)(7.16+10.3)=16.28 \text{ s/veh}$$

For this simple case of a lane group that only serves one movement in protected mode, the answer will be the same as in Sample Problem 22-7 (16.28 seconds). The beauty of the IQA method, however, is that it is not limited to a simple triangle, where there is only one arrival rate and one departure rate throughout the cycle. It can also model delay for all types of phasing plans and lane group types. Examples of this are discussed later in the chapter.

There are times when there is a residual queue at the end of the effective green. In this case, the procedure is iterated, with the residual queue used as the queue at start of the effective red, instead of zero. The steps are iterated until end queue equals start queue.

Incremental Delay

The incremental-delay equation is based on Akcelik's equation (see Chapter 18) and includes two types of delay. Because vehicles do not arrive uniformly as is assumed in the d1 term, the d2 term includes delay caused by vehicles arriving randomly cycle by cycle, creating the possibility that some cycles will be oversaturated (cycle failure). The d2 term also includes delay from overflow (oversaturation) when the entire analysis period is oversaturated, that is, v/cratio>1.00. Incremental delay is estimated as:

d2=900T×[(X−1)+(X−1)2+(8kIXcT)] [22-48]

where:

T=analysis time period, h,X=v/c ratio for lane group,c=capacity of lane gro
timed controllers), andI=upstream filtering/metering adjustmentfactor (I = 1

For pre-timed controllers, or unactuated movements in a semi-actuated controller, the k factor is always 0.50. The upstream filtering/metering adjustment factor is only used in arterial analyses. A value of 1.00 is assumed for all analyses of individual intersections.

Sample Problem 22-9: Determining Incremental Delay

For the same problem in Sample Problem 22-7 and 22-8, find the incremental delay.

Solution

d2=900×0.25×[(0.875−1)+(0.875−1)2+
(8×.5×1×0.8751800×0.44×0.25)]=12.8 s/veh

Initial Queue Delay

Initial queue delay, d3, represents added delay due to a queue that was not totally processed in the previous time period, that is, it is the average unmet demand from the previous time period. The existence of an initial queue will affect the uniform delay calculated previously, which will need to be adjusted. The equations for calculating initial queue and adjusting uniform delay are not included herein, and the reader is referred to HCM Chapter 19.

Control Delay and Level of Service

Control delay, *d*, is the sum of the uniform delay, incremental delay, and initial queue delay. This value is then used in Table 22.1 to assign level of service.

Sample Problem 22-10: Determining Total Control Delay and LOS

Find control delay and level of service when d1=16.28, d2=12.8, and d3=0.0 (all in seconds).

Solution

$d=d1+d2+d3=16.28+12.8+0.00=29.08$ s/veh

From Table 22.1, the LOS=C.

Movement Groups

The HCM combines certain lane groups into what is called a movement group to present results. Movement groups are used in the analysis of an urban street. Exclusive turning lane groups are lane groups and also movement groups. Shared lane groups and TH-only lane groups are combined into one movement group for purposes of presenting the results. Table 22.12 shows some possible combinations of movement groups.

Table 22.12: Movement Groups and Lane Groups

Lane Group	Number of Lanes in Lane Group	Movement Group	Number of Lanes in Movement Group
LT	1	LT	1
TH	2	TH/RT	3
RT	1		
LT	1	LTR	2
TR	1		

Table 22.12: Full Alternative Text

Capacity, v/c ratio, delay, and level of service may also be reported for movement groups. Movement groups are used more regularly in the urban streets methodology.

Aggregating Delay

To aggregate delays, the average delays are weighted by the number of vehicles experiencing the delays. The HCM allows delays to be aggregated to the approach and to the overall intersection, as follows:

dA=∑idivi∑ vidI=∑AdAvA∑AvA [22-49]

where:

di=total control delay per vehicle, lane groupi, s/veh,dA=total control delay

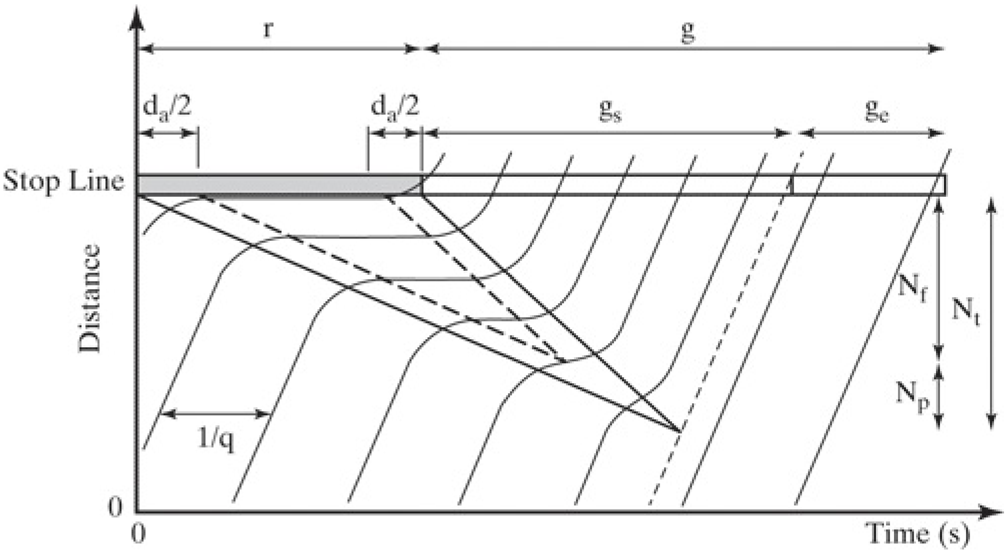
Levels of service may then be applied to lane groups, movement groups, approaches, and the overall intersection.

22.3.8 Estimating Queue Service Ratio

The queue storage ratio is the ratio of the storage space used by the average maximum back of queue during the analysis period to the total storage length available. Vehicles in queue include the vehicles that arrive on red as well as those vehicles that join the queue after the initiation of green and fully stop. Vehicles that join the queue, slow down, but do not fully stop are not included in the back of queue calculations. Figure 22.6 depicts vehicles arriving and traveling through the intersection.

Each line represents a single vehicle and speed is the slope of the trajectory. Thus vehicles that come to a full stop are those with a horizontal line (speed=0) in their trajectory. The first five vehicles make a full stop at the intersection, as seen from the horizontal section of their trajectory. The sixth vehicle experiences only a partial stop, that is, it does experience delay while decelerating and accelerating. The back of queue for this cycle, therefore, is five vehicles. The average back of queue over the analysis period is used to calculate the queue storage ratio. Back of queue consists of three components, as shown in Equation 22-50.

Figure 22.6: Time-Space Diagram of Vehicle Trajectory at Intersection Approach



(Source: Reprinted with permission from *Highway Capacity Manual, 6th Edition: A Guide for Multimodal Mobility Analysis*, the National Academy of Sciences, Courtesy of the National Academies Press, Washington, D.C., Exhibit 31-20, pgs 31–64.)

Figure 22.6: Full Alternative Text

$Q=Q1+Q2+Q3$ [22-50]

where:

Q=back of queue, veh/ln,$Q1$=first-term back of queue, veh/ln,$Q2$=second-term back of queue, veh/ln, and$Q3$=third-term back of queue, veh/ln.

First-term back of queue is that depicted in Figure 22.6. Second-term back of queue calculates the effect of random cycle failures on back of queue and also calculates the effect of sustained oversaturation (demand greater than capacity during the analysis period). The third-term back of queue calculates the effect that an initial queue has on back of queue.

The equations for calculating the three back of queue terms are not presented here, and the reader is referred to the 2016 HCM for these details.

Queue storage ratio is calculated with Equation 22-51.

Rq=LhQLa [22-51]
Lh=Lpc(1−0.01PHV)+0.01LHVPHV

where:

Rq=queue storage ratio,La=available queue storage length, ft/ln,Q=back of car lane length=25 feet,LHV=stored heavy-
vehicle lanelength=45 feet, andPHV=percent heavy vehicle, %.

If the queue storage ratio is less than 1.0, then blockage of the available queue storage will not occur. Blockage of the available queue storage length will occur when the queue storage ration is greater than or equal to 1.0.

22.4 Interpreting the Results of Signalized Intersection Analysis

At the completion of the HCM analysis procedure, the traffic engineer has the following results available for review:

- *v/c* ratios (*X*) for each movement group
- Delays and levels of service for each approach
- Delay for the overall intersection

All of these results must be considered to obtain a complete overview of predicted operating conditions in the signalized intersection and to get an idea of how to address any problems revealed by the analysis.

The v/c ratio and delay values are *not* strongly linked, and a number of interesting combinations can arise. The v/c ratio for any lane group, however, represents an absolute prediction of the sufficiency of the capacity provided to that group. Further, the critical v/c ratio represents an absolute prediction of the total sufficiency of capacity in all critical lane groups. The following scenarios may arise:

- Scenario 1: $X_c \leq 1.00$; *all* $X_i \leq 1.00$. These results indicate that there are no capacity deficiencies in any lane group. If there are no initial queues, then there will be no residual queues in any lane group at the end of the analysis period. The analyst may wish to consider the balance of *X* values among the various lane groups, particularly the critical lane groups. It is often a policy to provide balanced *X* ratios for all critical lane groups. This is best accomplished when all critical lane groups have $X_i \approx X_c$.
- Scenario 2: $X_c \leq 1.00$; *some* $X_i > 1.00$. As long as $X_c \leq 1.00$, all demands can be handled within the phase plan, cycle length, and physical design provided. All X_i values may be reduced to values less than 1.00 by reallocation of green time from lane groups with lower X_i values to those with $X_i > 1.00$. A suggested procedure for

reallocation of green time is presented later in this chapter.

- Scenario 3: Xc>1.00; *some or all* Xi>1.00. In this case, sufficient capacity can be provided to all critical-lane groups only by changing the phase plan, cycle length, and/or physical design of the intersection. Improving the efficiency of the phase plan involves considering protected left-turn phasing where none exists or protected plus permitted phasing where fully protected phasing exists. This may have big benefits, depending on the magnitude of left-turn demands. Increasing the cycle length will add small amounts of capacity. This may not be practical if the cycle length is already long or where the capacity deficiencies are significant. Adding lanes to critical-lane groups will have the biggest impact on capacity and may allow for more effective lane use allocations.

Delays must also be carefully considered, but should be tempered by an understanding of local conditions. Level-of-service designations are based on delay criteria, but acceptability of various delay levels may vary by location. For example, drivers in a small rural CBD will not accept the delay levels that drivers in a big city will.

As noted earlier, LOS F may exist where v/c ratios are less than 1.00. This situation may imply a poorly timed signal (retiming should be considered), or it may reflect a short protected turning phase in a relatively long cycle length. The latter may not be easily remedied; indeed, long delay to a relatively minor movement at a busy urban or suburban intersection is sometimes intentional.

Aggregate levels of service for approaches—and particularly for the intersection—may mask problems in one or more movement groups. Individual movement group delays and levels of service should always be reported and must be considered with aggregate measures. This is often a serious problem when consultants or other engineers report only the approach delays and level of service.

Where movement group delays vary widely, some reallocation of green time may help balance the situation. However, when changing the allocation of green time to achieve better balance in lane group delays, the impact of the reallocation on v/c ratios must be watched carefully.

22.5 Methodological Complexities

Previous sections of this chapter have dealt with portions of the HCM model for analysis of pre-timed signalized intersections that have simple phase plans and can be calculated manually. In this section of the chapter, some of the more complex models will be discussed. Some elements will not be completely detailed, and for these cases, the reader should consult the HCM directly for fuller descriptions.

The following aspects of the model are addressed:

- Modeling delay and capacity for permitted left turns
- Analysis of compound phasing
- Movements served by more than one lane group

Analysis of actuated signals is covered in Part II of this chapter.

22.5.1 Modeling Delay and Capacity for Permitted Left Turns

The modeling of permitted left turns must account for the complex interactions between permitted left turns and the opposing flow of vehicles. These interactions involve several discrete time intervals within a green phase that must be separately addressed.

Figure 22.7 illustrates these portions of the green phase. It shows a subject approach (WB approach) with an opposing EB flow. When the green phase is initiated, vehicles on both approaches begin to move. Vehicles from the standing queue on the opposing approach move through the intersection *with no gaps*, at the saturation flow rate, sop. Thus, *no* left turn from the subject approach may proceed during the time it takes this opposing queue of vehicles to clear the intersection. If a left-turning vehicle arrives in the subject approach during this time, it must wait, *blocking the left-most lane*, until the opposing queue has cleared. After the opposing queue has cleared, left turns from the subject approach are made through gaps in the now unsaturated opposing flow. The rate at which they can be made as well as their impact on the operation of the subject approach is dependent on the number of left turns and the magnitude and lane distribution of the opposing flow.

Figure 22.7: Portions of the Green Phase Illustrated

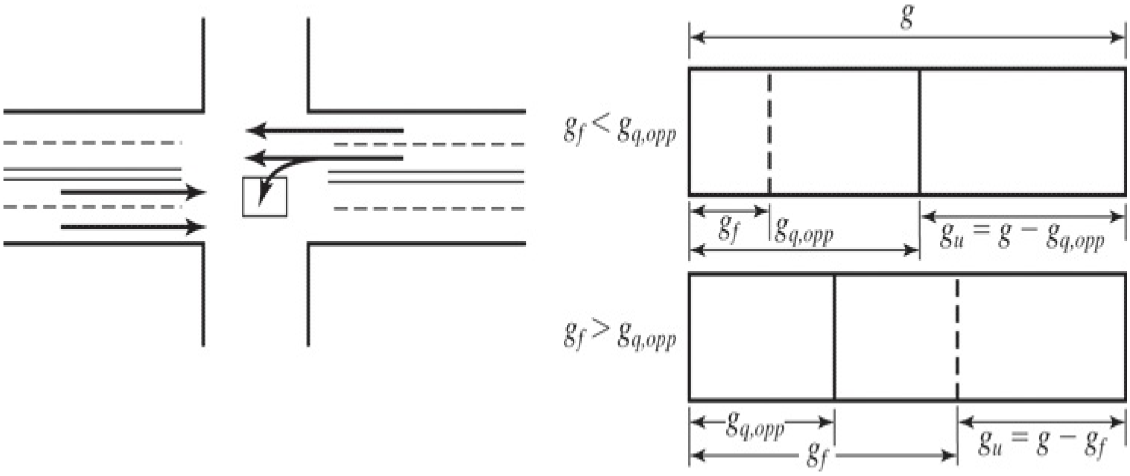


Figure 22.7: Full Alternative Text

Another fundamental concept is that left-turning vehicles have no impact on the operation of the subject approach *until the first left-turning vehicle arrives*. There are three distinct portions of the green phase that may be defined as follows:

gq,opp=average amount of green time required for aqueue of standing vehi
turning vehicle arrives on the subjectapproachgu=average amount of time a
turningvehicle in the subject direction

Figure 22.7 illustrates the relationship between these key variables. The value of gu depends upon the relative values of gf and gq,opp. If the first left-turn vehicle arrives before the opposing queue clears, then the unsaturated time occurs after the opposing queue clears. If the first left-turn vehicle arrives after the opposing queue clears, then the unsaturated time occurs after the first left arrives.

$$g_u = g - g_{q,opp} \quad \text{for } g_{q,opp} \geq g_f \qquad g_u = g - g_f \quad \text{for } g_{q,opp} < g_f \quad [22\text{-}52]$$

where:

g=total effective green for the phase, s.

When defined in this fashion, g_u represents the actual time (per phase) that left turns may filter through an unsaturated opposing flow.

Lastly, left turns can move during the clearance lost time as sneakers.

Permitted left turns are analyzed using the IQA method. The arrival and departure rates are found during each portion of the green phase, using the above definitions of critical portions of the green phase. Thus the model for left turns must consider what type of left-turn operation is taking place at various times within a given green phase. There are five separate portions of the signal timing that can be identified:

- Interval 1−r: red phase
- Interval 2−g_f: green time before first left-turning vehicle arrives
- Interval 3−g_{diff}: time between clearance of opposing queue and arrival of first left-turning vehicle on the subject approach
- Interval 4−g_u: green time during which left turns are made through an opposing unsaturated flow
- Interval 5−“sneakers”

Interval 1: *r*

As done in the protected phase model, start at the beginning of effective red time, assume the queue is zero, and calculate the queue at the end of the red phase.

Interval 2: g_f

Before the first left-turning vehicle arrives in the subject approach, left-turning vehicles have no impact on the operation of the left lane. Thus, during this period, vehicles arrive at vg, the arrival rate on green, and depart at the saturation flow rate, *s*, the saturation flow rate for the lane without left turners. Where permitted left turns are made from an exclusive LT lane, gf is 0.0, as the first vehicle in queue is, by definition, a left-turning vehicle.

The algorithm for estimating gf depends on whether the subject approach is a multilane or a single-lane approach. Departure behavior of queues is somewhat altered when there are left- and right-turning vehicles in the same queue.

gf=Ge−(0.86LTC0.717)−ℓ1Shared Multilane Approachgf=Ge −(0.882LTC0.629)−ℓ1Shared Single-LaneApproachgf=0Exclusive Left-turn Lane [22-53]

Equations 22-53 are subject to a minimum value of zero and a maximum value as calculated in Equation 22-54

gf,max=(1−PL)0.5PL(1−[1−PL]0.5gp)−l1 [22-54]

where:

gf=portion of green phase before the arrivalof the first left-
turning vehicle on thesubject approach, s,G=actual green time for lane grou
up lost time for the permitted phase, s,LTC=left turns per cycle, veh/cycle[
vLT×C/3,600
],gf,max=maximum value of gf,PL=proportion of left turns in the sharedlan

Interval 3: gdiff

The time between opposing queue clearance time and the arrival of the first left turn in the subject direction is gdiff=gq,opp−gf. If the first left-turning vehicle on the subject approach arrives before the opposing queue clears (gq>gf), the vehicle must wait, blocking the left lane during this interval. No vehicle can move in the left lane while the left-turner waits. Therefore, the saturation flow rate is 0.00. Where gf≥gq, this time period does not exist.

When the *opposing approach* has a single lane, a unique situation for left-turners arises. Left-turning vehicles located within the opposing standing queue will create gaps in the opposing queue as it clears. Left-turners on the subject approach may make use of these gaps to execute their turns. Thus, when the opposing approach has only one lane, some left turns from the subject approach can be made during the time period (gdiff) and the saturation flow rate will be adjusted using a through-vehicle equivalence of EL2 (defined in Equation 22-57).

Interval 4: gu

This is the period during which left turns from the subject approach filter through an unsaturated opposing flow. During this period of time, the saturation flow rate will be adjusted using a through-vehicle equivalence value of EL1 (defined in Equation 22-59) to reflect the impedance of the opposing flow.

In order to calculate gu, the opposing queue clearance time must first be found. Equation 22-55 is used to find gq,opp, the time when the *opposing* queue clears, calculated for the lane with the longest queue clearance time. Thus, for each opposing lane, Equation 22-55 is found and the largest value is then used in Equation 22-56 to find gu.

gq,opp=vr,opp×rso−vg,opp [22-55]

where:

gq,opp=average time for opposing standingqueue to clear the intersection, s
gu,sub=Gperm−l1,opp−gq,opp−l1,sub+esub [22-56]

where:

gu,sub=unsaturated green time for the subjectapproach, s,Gperm=actual gre
up lost time for the opposingapproach,l1,sub=start-
up lost time for the subjectapproach, andesub=extension of effective green

Interval 5: “sneakers”

During the ending or clearance lost time, l2, left turns can move as sneakers. The number of sneakers depends upon the proportion of left turns in the lane group, PL. The number of sneakers per cycle is estimated as (1+PL).

Queue Accumulation Polygon for Permitted Left Turns

Consider the queue accumulation polygon for a shared left-through lane with permissive-only phasing, shown in Figure 22.8. The queue will then grow again because no vehicles will depart until the opposing queue clears at time gq (Point Qq). After the opposing queue clears, the remaining green time left is the unsaturated green time, gu. If the intersection is not fully saturated nor oversaturated, the subject queue will clear during the unsaturated time, at Point Qp.

Figure 22.8: Queue Accumulation Polygon for a Shared Lane with Permitted Left Turns

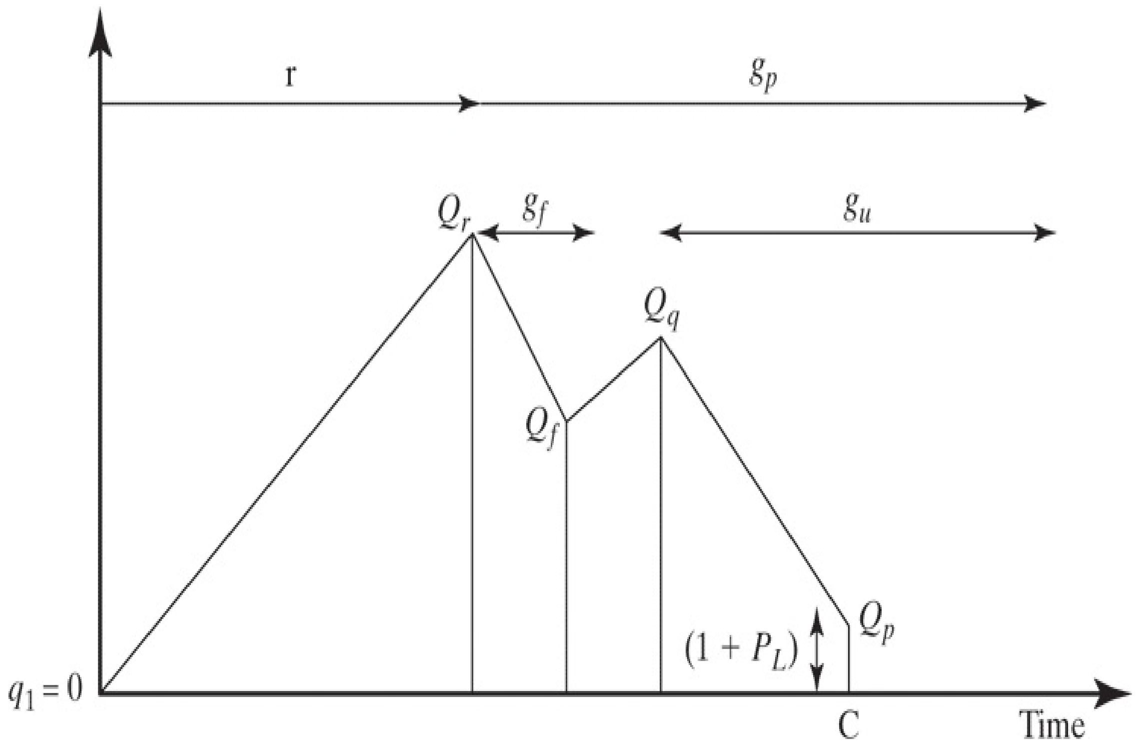


Figure 22.8: Full Alternative Text

Saturation Flow Rate for the Five Intervals of a Permitted Left-Turn Phase

- Interval 1: *r* red time. When the phase is effectively red, vehicles continue to arrive, but the departure rate is zero and the queue grows to Qr.
- Interval 2: gf, the time before the first left-turning vehicle arrives. When effective green begins, the queue begins to dissipate until the first left-turn vehicle arrives at time gf and blocks the lane. The saturation flow rate during this period is the adjusted through saturation flow rate, sTH,.

  The queue at the end of gf is Qf.
- Interval 3: gdiff. During gdiff=gq,opp−gf≥0, two situations are possible:

1. If the opposing approach has more than one lane, the saturation flow rate is zero.

2. If the opposing approach has only one lane, subject left turns can be made through gaps in the opposing queue created by opposing left-turning vehicles. The saturation flow rate during gdiff is:

   s=sTH×11+PL((EL2/fLpb)−1)

   where:

   EL2=through vehicle equivalent for aleft turn made during perioc −gf.

   The value of the equivalent, EL2, is determined using a probabilistic model that considers how long a left-turning vehicle in the subject approach would have to wait for a left-turning vehicle in the opposing approach to open a gap in the opposing traffic stream.

   EL2=1−PTHonPLTo [22-57]

   where:

   PTHo=proportion of through vehiclesin the opposing single-laneapproach,PLTo=proportion of left turns in theopposing singl laneapproach, andn=maximum number ofopposing vehicles in tir −gf, is estimatedas 0.278×(gq −gf), where0.278 roughly estimatessaturation flow rate ofopposin

- Interval 4: gu, the time in which left turns filter through an unsaturated opposing flow.

  1. During gu, the saturation flow rate is:

     s=sTH×11+PL((EL1/fLpb)−1) [22-58]

     where:

     EL1=through vehicle equivalent of avehicle executing a left turnc turningvehicles in the lane fromwhich left turns are made.

The through-car equivalent of a left-turn vehicle during gu is found using Equation 22-58. It is the ratio of the permitted left-turn saturation flow rate to the base saturation flow rate.

EL1=sospsp=voe−votc/3,6001−e−votfh/3,600 [22-59]

where:

so=base saturation flow rate,sp=saturation flow rate of apermittec up headway = 2.5 s.

- Interval 5: Sneakers

  During the clearance lost time, sneakers may proceed at a rate of:

  s=(3,600/C)×ns

  where:

  ns=Expected number of sneakersper cycle=(1+PL).

Using the portions of the permitted green, the arrival rates, and saturation flow rates, it is now possible to calculate capacity and uniform delay.

Sample Problem 22-11: A Left/Through Lane with Permitted Phasing

Find uniform delay and capacity for a shared left/through lane with permitted-only phasing, with the following characteristics:

- 90 s cycle length
- 45 s green phase
- Arrival type=2
- l1=e=2s

- 80 lefts and 300 through vehicles, vph
- Vopp=500 vphpl (one opposing through-only lane with arrival type 4)
- sTH=1750 vphpl

Solution for uniform delay

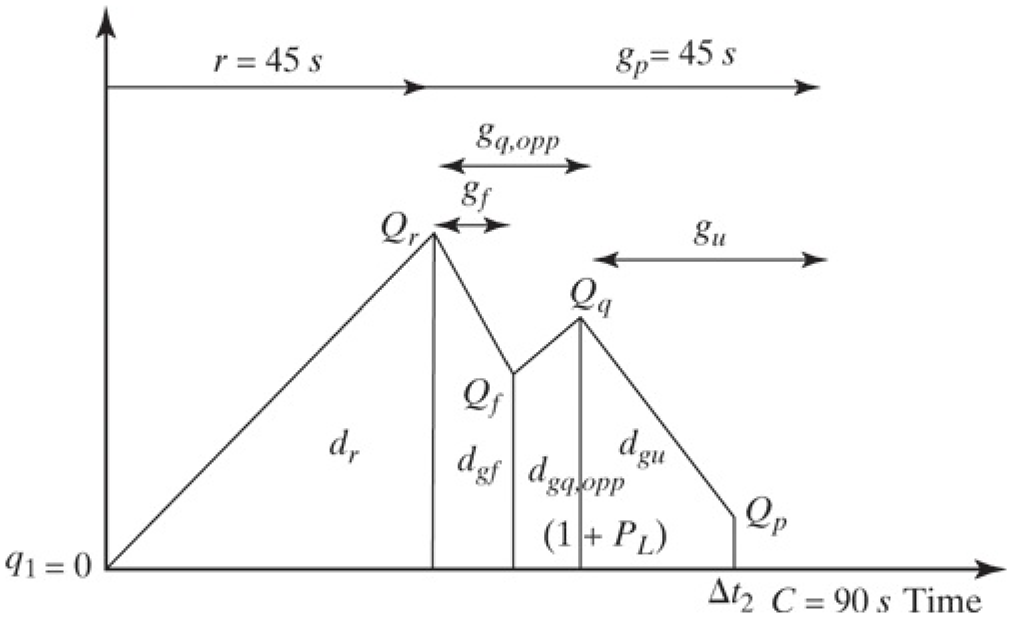
1. Calculate effective green time:

   $g=G-l1+e=45-2+2=45$ s

2. Calculate effective red time:

   $r=90-45=45$ s

3. Draw queue accumulation diagram.

   22.6-19 Full Alternative Text

4. Calculate proportion of vehicles arriving on green:

   P=(AT3)(gC)=(23)(4590)=0.33

5. Calculate arrival rate on red:

$vr=(1-P)VCr=(1-0.33)380\times9045=505$ veh/h

6. Calculate queue at end of effective red, assuming no queue at start of effective red:

   $Qr=q1+(vr-s3600)\times r=0+(505-03600)\times45=6.3$ veh

7. Calculate delay during effective red:

   $dr=r\times(q1+Qr2)=45\times(0+6.32)=142.1$ veh−s

8. Calculate arrival rate on green:

   $Vg=VPCg=380\times0.33\times9045=255$ veh/h

9. Calculate time until arrival of first left-turning vehicle:

   $LTC=vLT\times C3{,}600=80\times903{,}600=2gf=Ge-(0.86LTC0.717)-l1=45e-(0.86\times20.717)-2 \quad =8.95$ s

10. Calculate queue at end of gf:

    $Qf=Qr+(vg-sTH3600)\times gf=6.3+(255-17503600)\times 8.95=2.6$ veh

11. Calculate delay during gf:

    $dgf=gf\times(Qr+Qf2)=8.95\times(6.3+2.62)=39.9$ veh−s

12. Calculate time until opposing queue clears (Arrival type=4):

    $vr,opp=(1-P)\ V\ Cr=(1-0.66)\ 500\times9045=333veh/hgq,opp=vr,opp\times rso-vg,opp=333\times451750-667=13.9$

13. Calculate subject queue at the end of $g_{q,opp}$:

    $Qq=Qf+(vg-s3600)\times gq,opp=2.6+(255-03600)\times13.9=9.3$ veh

14. Calculate delay during $g_{q,opp}$:

    $dgq,opp=gq,opp\times(Qf+Qq2)=13.9\times(2.6+9.32)=82.6$ veh−s

15. Calculate unsaturated green time, gu:

$gu=gp-gq,opp=45-13.9=31.2$ s

16. Calculate saturation flow rate during gu:

    $sp=voe-votc/3{,}6001-e-votfh/3{,}600=500e-500\times4.5/3{,}6001-e$
    $-500\times2.5/3{,}600=912$ veh/h$EL1=sosp=1750912=1.92PL=80380=0.21s$
    $-1)=1750\times11+0.21((1.92/1)-1)=1467$vphpl

17. Calculate queue at end of gu:

    $Qgu=Qq+(vg-s3600)\times gu=9.3+(255-14673600)\times31.2=-1.16$ veh

    **This is negative because the queue clears before the end of the unsaturated green time.*

18. Calculate time during gu when the queue clears:

    $\Delta t2=3600\ Qqs-Vg=3600\times9.31750-255=22.5$ s

19. Calculate delay during unsaturated green, gu:

    $dgu=\Delta t2\times(qgq,opp+q\Lambda t22)=22.5\times(9.3+02)=104.8$ veh−sec

20. Calculate the total number of arrivals during effective green:

    $na=vg3600\times g=2553600\times45=3.2$ veh

21. Calculate average uniform delay:

    $d1=(dr+dgf+dgq,opp+dgu)(Qr+na)=(142.1+39.9+82.6+104.8)$
    $(6.3+3.2)=38.9$ s/veh

Solution for capacity of the lane group

Vehicles depart during gf and gu. Given that we know the effective green time and saturation flow rate of both these portions of the green, we can calculate the capacity of each portion using $c=s\times(g/C)$. Additionally, some vehicles depart as sneakers. Total capacity of the lane group is then the sum of these three amounts.

1. Calculate capacity during gf:

   $cgf=sgf\times(gfC)=1750\times(8.9590)=174$ vph

2. Calculate capacity during gu:

   $cgu=sgu\times(guC)=1467\times(31.290)=508$ vph

3. Calculate capacity of sneakers:

   $csneakers=3600\times(1+PL)C=3600\times(1+0.21)90=48$ vph

4. Calculate total capacity of the lane group:

   $c=cgf+cgu+csneakers=174+508+48=730$ vph

22.5.2 Modeling Compound Phasing

Protected plus permitted or permitted plus protected phasing is the most complex aspect of signalized intersection operations to model analytically. The approach to estimating saturation flow rates, capacities, and delays, however, is the same as shown above but with a more complicated polygon.

In terms of saturation flow rates and capacities, the general approach taken in the HCM is straightforward. The protected and permitted portions of the phase are separated, with saturation flow rates and capacities computed separately for each portion of the phase. The appropriate green times are associated with each portion of the phase. For example, the protected portion of a left-turn phase is analyzed as if it were a fully protected phase, while the permitted portion of the phase is analyzed as if it were a fully permitted phase, using the left-turn model described in the previous section.

In analyzing the permitted portion of the phase, however, the algorithms used to predict gf,gq, and gu must be modified to reflect the fact that the timing of the permitted phase does not necessarily start at the beginning of the green phase. In compound phasing, the values of gf,gq, and gu must be altered to reflect this. For example, gf is the time *within the permitted phase* to the arrival of the first left-turning vehicle in the subject phase. It is indexed to the beginning of green on the subject approach. If the approach has a protected plus permitted phase, the predicted value of gf would be relative to the start of the green—which is the beginning of the *protected* portion of the phase. The value needed must be indexed to the start of the *permitted* portion of the green, which requires an adjustment.

Depending upon the order and type of compound phasing in place, there are many different scenarios requiring different adjustments to the prediction of gf,gq, and gu. All of these are detailed in the HCM, and the reader is referred to Chapter 31 of the HCM.

22.5.3 Movements Served by More than One Lane Group

If a movement group consists of two or more lane groups, at least one movement will have to choose which lane group it will enter. For example, if an approach has a left/through lane and a through/right lane, as shown in Figure 22.9, a through vehicle may choose to enter either lane. Drivers will try to choose the lane that they believe will minimize their service time.

Figure 22.9: An Approach with Two Lane Groups and Three Movements

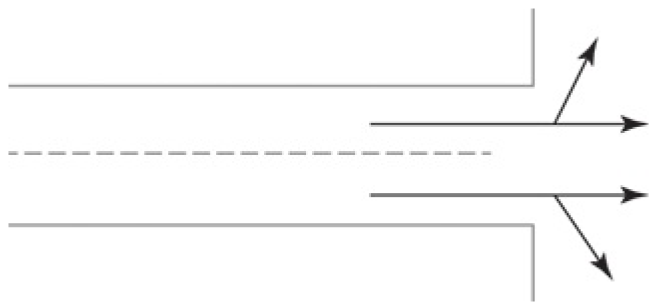


Figure 22.9: Full Alternative Text

The method in the HCM used to estimate the volume in each lane group involves equalizing the v/s ratios of the lane groups involved.

Although all the equations used in this methodology are not detailed here, a very general description of the process follows.

1. Step 1. Start with an initial estimate of demand flow rate in shared lane.

2. Step 2. Compute flow rate in exclusive lane group.

3. Step 3. Compute the proportion of turns in the shared-lane lane group.

4. Step 4. Compute lane-group saturation flow rate. Saturation flow rate estimated as part of this methodology is somewhat different than saturation flow rate computed in other parts of the analysis. When equalizing v/s ratios, the left- and right-turn equivalency factors which may be needed (EL, ER, EL1, EL2) are modified to account for the probability of a lane change.

5. Step 5. Compute flow ratio for the subject movement group, using Equation 22-60.

   y*=∑i=1nviNi∑i=1nsiNi [22-60]

   where:

   y*=flow ratio for subject movement group,vi=flow rate in

6. Step 6. Revise exclusive lane-group flow rates using Equation 22-61.

   vi=siy* [22-61]

   where:

   vi=demand flow rate in lane group i, andsi=saturation flow

7. Step 7. Calculate the shared-lane lane-group flow rate by subtracting the volume in the exclusive lane from the total volume being divided between lane groups in the movement group, using Equation 22-62.

   $v_{sh}=v_t-v_e$ [22-62]

   where:

   v_{sh}=flow rate in shared lane,v_t=total volume that may cho

8. Step 8. Compare revised demand flow rates computed in steps 6 and 7 with flow rates used initially. If they differ by more than 0.1 vph, then iterate these steps using the flow rates computed in Steps 6 and 7 as the initial flow rates. Iterate until the difference between initial and end flow rates are less than 0.1 vph.

Part II Analysis of Actuated Signals

The HCM provides a detailed model for actuated signals for estimating the average signal timing during the analysis period, given the controller and detector parameters. It is an algorithm that requires iterations and is not possible to compute manually.

Determining actuated phase duration starts by assuming the phase durations are equal to the maximum green times entered. Using the volumes and settings input, new phase durations are calculated and compared to the starting values. If the ending and starting phase times are not equivalent, then the calculations begin again, starting with the final phase times of the previous iteration.

An actuated phase consists of five intervals, as shown in Equation 22-63.

Dp=l1+gq+ge+y+ar [22-63]

where:

Dp=duration of an actuated phase, s,l1=start-up lost time, s,gq=queue clearance time, s,ge=green extension time, s, time way, MAH, and the maximum greentime,y=yellow change interval, s, anda red clearance interval, s.

Inputs that are needed for a fully actuated-control analysis, in addition to the inputs provided in Table 22.2 for a pre-timed analysis, are shown in Table 22.13.

Table 22.13: Additional Data Requirements for a Fully Actuated Signal Analysis

Type of Condition	Parameter	Default Value
Actuated Control Parameters	Passage Time	2.0 s (presence detector)
	Maximum Green Time	Major street through mvmt: 50 s Minor street through mvmt: 30 s Left-turn mvmt: 20 s
	Minimum Green Time	Major street through mvmt: 10 s Minor street through mvmt: 8 s Left-turn mvmt: 6 s
	Phase recall	No recall
	Dual Entry	Not enabled (i.e., use single entry)
	Simultaneous Gap-out	Enabled
	Stop line detector length	40 ft, presence detection mode

Table 22.13: Full Alternative Text

The equations used to calculate average phase times are not detailed in this text and the reader is directed to Chapter 31 of the 2016 HCM for a complete description of the methodology.

Part III Calibration Issues

The HCM model is based on a default base saturation flow rate of 1,900 pc/hg/ln or 1,750 pc/hg/ln (based upon metropolitan area population). This value is adjusted by up to 11 adjustment factors to predict a prevailing saturation flow rate for a lane group. The HCM provides guidance on the measurement of the prevailing saturation flow rate, s. Although it allows for substituting a locally calibrated value of the base rate, so, it does not provide a means for doing so. It also does not provide a procedure for measuring lost times in the field.

It is also useful to quickly review how the calibration of adjustment factors of various types may be addressed, even if this is impractical in many cases. A study procedure for measuring delays in the field is detailed in Chapter 9.

22.6 Measuring Prevailing Saturation Flow Rates

As defined in Chapter 18, saturation flow rate is the maximum average rate at which vehicles in a standing queue may pass through green phase, after start-up lost times have been dissipated. It is measured on a lane- by-lane basis through observations of headways as vehicles pass over the stop line of the intersection approach. The first headway begins when the green is initiated and ends when the first vehicle in queue crosses the stop line (front wheels). The second headway begins when the first vehicle (front wheels) crosses the stop line and ends when the second vehicle in queue (front wheels) crosses the stop line. Subsequent headways are similarly measured.

The HCM suggests that, for most cases, the first four headways include an element of lost time and, thus, are not included in saturation flow rate observations. Saturation headways, therefore, begin with the fifth headway in queue and end when the last vehicle in the standing queue crosses the stop line (again, front wheels). Subsequent headways do not necessarily represent saturation flow.

22.7 Measuring Base Saturation Flow Rates

The base saturation flow rate assumes a set of “ideal” conditions that include 12-foot lanes, no heavy vehicles, no turning vehicles, no local buses, level terrain, and non-CBD location, among others. It is usually impossible to find a location that has all of these conditions.

In calibrating a base saturation flow rate, a location is sought with near ideal physical conditions. An approach with three or more lanes is recommended, as the middle lane can provide for observations without the influence of turning movements. Heavy vehicles cannot be avoided, but sites that have few heavy vehicles provide the best data. Even where data are observed under near ideal physical conditions, *all headways* observed after the first heavy vehicle must be discarded when considering the base rate.

22.8 Measuring Start-Up Lost Time

If the first four headways contain a component of start-up lost time, then these headways can be used to measure the start-up lost time. If a saturation headway for the data has been established as *h* s/veh, then the lost time component in each of the first four headways is (hi−h), where hi is the total of observed headways for vehicles 1–4 in queue. The start-up lost time is the sum of these increments. Both saturation flow rate and start-up lost time are observed for a given lane during each signal cycle. The calibrated value for use in analysis would be the average of these observations.

Start-up lost time under base conditions can be observed as well by choosing a location and lane that conforms to the base conditions for geometrics with no turning vehicles and by eliminating consideration of any headways observed after the arrival of the first heavy vehicle.

Sample Problem 22-12: Measuring Saturation Flow Rates and Start-Up Lost Times

The application of these principles is best illustrated through example. Table 22.14 shows data for six signal cycles of a center lane of a three-lane approach (no turning vehicles) that is geometrically ideal. In general, calibration would involve more cycles and several locations. To keep the illustration to a reasonable size, however, the limited data of Table 22.14 will be used.

Table 22.14: Example in Measuring Saturation Flow

Rate and Start-Up Lost Time

Queue Position	Observed Headways (s) in Cycle No. ___						Sum of Sat Hdwys	No. of Sat Hdwys
	1	2	3	4	5	6		
1	3.5	2.9	3.9	4.2H	2.9	3.2	0.0	0
2	3.2	3.0	3.3	3.6	3.5H	3.0	0.0	0
3	2.6	2.3	2.4	3.2H	2.7	2.5	0.0	0
4	2.8H	2.2	2.4	2.5	2.1	2.9H	0.0	0
5	2.5	2.3	*2.1*	2.1	2.2	2.5	13.7	6
6	2.3	*2.1*	*2.4*	2.2	2.0	2.3	13.3	6
7	3.2H	*2.0*	*2.4*	2.4	2.2	2.3	14.5	6
8	2.5	*1.9*	2.2	2.3	2.4	2.0	13.3	6
9	4.5	2.9H	2.7H	1.9	2.2	2.4	12.1	5
10	6.0	2.5	2.4	2.3	2.7H	2.1	12.0	5
11		2.8H	4.0	2.2	2.4	2.0	9.4	4
12		2.5	7.0	2.9H	5.0	2.3	7.7	3
13		5.0		4.1		6.0	0.0	0
14		7.5					0.0	0
15							0.0	0
Sum							**96.0**	**41**

Notes: H=heavy vehicle.

Single underline: beginning of saturation headways.

Double underline: end of standing queue clearance; end of saturation headways.

Italics: saturation headway under base conditions.

Table 22.14: Full Alternative Text

Note that saturation conditions are said to exist only between the fifth headway and the headway of the last vehicle present in the standing queue when the signal turns green. Only the headways occurring between these limits can be used to calibrate saturation flow rate. The first four headways in each queue will be used subsequently to establish the start-up lost time.

The saturation headway for the lane in question is the average of all observed headways representing saturated conditions. As seen in Table 22.14, there are 41 observed saturation headways totaling 96.0 seconds.

From this data, the average saturation headway (under prevailing conditions) at this location is:

$h=96.041=2.34$ s/veh

From this, the saturation flow rate for this lane may be computed as:

$s=3{,}6002.34=1{,}538$ veh/hg/ln

If a lane group had more than one lane, the saturation headways and flow rates would be separately measured for each lane. The saturation flow rate for the lane group is then the sum of the saturation flow rates for each lane.

Measuring the base saturation flow rate for this location involves eliminating the impact of heavy vehicles, assuming that all other features of the lane conform to base conditions. As the heavy vehicles may conceivably influence the behavior of any vehicle in queue behind it, the only headways that can be used for such a calibration are those before the arrival of the first heavy vehicle. Again, saturation headways begin only with the fifth headway. Looking at Table 22.15, there are only eight headways that qualify as saturation headways occurring before the arrival of the first heavy vehicle:

- Headways 5–8 of Cycle 2
- Headways 5–8 of Cycle 3

Table 22.15: Calibration of Start-Up Lost Time from

Table 22.14 Data

Position In Queue	Observed Headway (s) for Cycle No. ____						Avg h (s)	h_{avg} − 2.175 (s)
	1	2	3	4	5	6		
1	2.5	2.9	3.9	H	2.9	3.2	3.080	0.905
2	3.2	3.0	3.3	H	H	3.0	3.125	0.950
3	2.6	2.3	2.4	H	H	2.5	2.450	0.275
4	H	2.2	2.4	H	H	H	2.300	0.125
Sum								**2.255**

H = headway occurring after arrival of first heavy vehicle.

Table 22.15: Full Alternative Text

The sum of these eight headways is 17.4 seconds, and the base saturation headway and flow rate may be computed as:

ho=17.48=2.175 s / vehso=3,6002.175=1,655 pc / hg / ln

Start-up lost time is evaluated relative to the base saturation headway. It is calibrated using the first four headways in each queue, as these contain a component of start-up lost time in addition to the base saturation headway. As the lost time is relative to base conditions, however, only headways occurring before the arrival of the first heavy vehicle can be used. The average headway for each of the first four positions in queue is determined from the remaining measurements. The component of start-up lost time in each of the first four queue positions is then taken as (hi−ho).

This computation is shown in Table 22.15, which eliminates all headways occurring after the arrival of the first heavy vehicle. The start-up lost time for this lane is 2.255 s/cycle.

Where more than one lane exists in the lane group, the start-up lost time would be separately calibrated for each lane. The start-up lost time for the lane group would be the *average* of these values.

Clearly, for actual calibration, more data would be needed and should involve a number of different sites. The theory and manipulation of the data to determine actual and base saturation flow rates, however, do not change with the amount of data available.

22.9 Calibrating Adjustment Factors

Of the 11 adjustment factors applied to the base saturation flow rate in the HCM model, some are quite complex and would require major research studies for local calibration. Included in this group are the left-turn and right-turn adjustment factors and the pedestrian/bicycle interference adjustment factors. A number of the adjustment factors are relatively straightforward and would not be difficult to calibrate locally, at least on a theoretical basis. It may always be difficult to find appropriate sites with the desired characteristics for calibration. Three adjustment factors involve only a single variable:

- Lane width (12-feet base condition)
- Grade (0% base condition)
- Area type (non-CBD base condition)

Two additional factors involve two variables:

- Parking (no parking base condition)
- Local bus blockage (no buses base condition)

The heavy vehicle/grade factor involves a number of considerations, and the lane utilization factor should be locally measured in any event and is found using Equation 22-16 or default values.

Calibration of all of these factors involves the controlled observation of saturation headways under conditions in which only one variable does not conform to base conditions. By definition, an adjustment factor converts a base saturation flow rate to one representing a specific prevailing condition, or:

s=sofi [22-64]

where fi is the adjustment factor for condition *i*. Thus, by definition, the adjustment factor must be calibrated as:

fi=sso=(3,600/h)(3,600/ho)=hoh [22-65]

where all terms have been previously defined.

For example, to calibrate a set of lane-width adjustment factors, a number of saturation headways would have to be determined at sites representing different lane widths but where all other underlying characteristics conformed to base conditions.

Sample Problem 22-13: Calibrating Adjustment Factors

The following data were obtained for various lane widths. Calibrate lane-width adjustment factors for this data.

h10=2.6 s / vehh11=2.4 s / vehh12=2.1 s / veh (base conditions)h13=2.0 s / v

Adjustment factors for the various observed lane widths could then be calibrated using Equation 22-56:

fw10=2.12.6=0.808fw11=2.12.4=0.875fw12=2.12.1=1.000fw13=2.12.0=1.

Adjustment factors for lanes wider than 12 feet are greater than 1.000, indicating that saturation flow rates increase from the base value for wide lanes (>12ft). For lanes narrower than 12 ft, the adjustment factor is less than 1.000.

The results of Sample Problem 22-13 are somewhat different from the 2016 HCM, which uses a factor of 1.00 for all lane widths between 10.0 and 12.0 feet.

Similar types of calibration can be done for any of the simpler adjustments. If a substantial database of headway measurements can be achieved for any given factor, regression analysis may be used to determine an appropriate relationship that describes the factors.

Calibrating heavy vehicle factors (or passenger car equivalents for heavy vehicles) is a bit more complicated and is illustrated in Sample Problem 22-14.

Sample Problem 22-14: Heavy Vehicle Adjustment

Refer to the sample problem for calibration of prevailing and base saturation flow rates. In this case, all conditions conformed to the base, except for the presence of heavy vehicles. Of the 41 observed saturation headways, 6 were heavy vehicles, representing a population of $(6/41)\times100=14.63\%$. *Calibrate the appropriate heavy vehicle/grade adjustment factor*.

The actual adjustment factor for this case is easily calibrated. The base saturation headway was calibrated to be 2.175 s/veh, while the prevailing saturation flow rate (representing all base conditions, except for heavy vehicle presence) was 2.34 s/veh. The adjustment factor is:

$$f_{HV}=\frac{2.175}{2.34}=0.929$$

This calibration, however, is only good for 14.63% heavy vehicles on the existing grade (unknown). Additional observations at times and locations with varying heavy vehicle presence would be required to generate a more complete relationship.

There is another way to look at the situation that produces a more generic calibration. If all 41 headways had been passenger cars, the sum of the headways would have been $41\times2.175=89.18$ s. In fact, the sum of the 41 headways was 96.0 s. Therefore, the six heavy vehicles caused $96.00-89.18=6.82$ s of additional time consumption due to their presence. If *all* of the additional time consumed is assigned to the six heavy vehicles, each heavy vehicle accounted for $6.82/6=1.137$ s of additional headway time. If the base saturation headway is 2.175 s/veh, the saturation headway for a heavy vehicle would be $2.175+1.137=3.312$ s/veh. Thus, one heavy vehicle consumes as much headway time as $3.312/2.175=1.523$ passenger cars. This is, in effect, the passenger car equivalent for this case, E_{HV}. This can be converted to an adjustment factor using Equation 22-14:

$f_{HV}=11+0.146(1.523-1)=0.929$

It should be noted that the 2016 HCM does not explicitly use a value of E_{HV} in the determination of heavy-vehicle/grade adjustment factors.

22.10 Normalizing Signalized Intersection Analysis

In many cases, it will be difficult or too expensive to calibrate individual factors involved in signalized intersection analysis. Nevertheless, in some cases, it will be clear that the results of HCM analysis are not correct for local conditions. This occurs when the results of analysis are compared to field measurements and obvious differences arise.

It is possible to “normalize” the HCM procedure by observing departure volumes on fully saturated, signalized intersection approaches—conditions that connote capacity operation. Consider the following example.

Sample Problem 22-15: Normalizing an HCM Analysis

Consider the case of a three-lane intersection approach with a 30 s effective green phase in a 60 s cycle. Assume further that the product of all 11 adjustment factors that apply to the prevailing conditions is 0.80. How could an analyst normalize field observations that differ from the computed results?

Solution

The computed results are:

$s=s_oNF=1,900 \times 3 \times 0.80=4,560 \text{ veh / h} gc=4,560 \times (3060)=2,280 \text{ veh / h}$

This is the predicted capacity of the lane group using the HCM model. Despite this result, field observations measured a peak 15-min departure flow rate from this lane group (under fully saturated conditions) of 2,400 veh/h.

The measured value represents a field calibration of the actual capacity of the lane group, as it was observed under fully saturated conditions. As it is more than the estimated value, the conclusion must be that the estimated value using the HCM model is too low. The difficulty is that it may be too low for many different reasons:

- The base saturation flow rate of 1,900 pc/hg/pl is too low.
- One or more adjustment factors is too low.
- The product of 11 adjustment factors is not an accurate prediction of the *combination* of prevailing conditions existing in the lane group.

All of this assumes that the measured value was accurately observed. The latter point is a significant difficulty with the methodology. Calibration studies for adjustment factors focus on isolated impacts of a single condition. Is the impact of 20% heavy vehicles in an 11-foot lane on a 5% upgrade the same as the product of the three appropriate adjustments, $fHV \times fw \times fg$? This premise, particularly where there are 11 separately calibrated adjustments, has never been adequately tested using field data.

The local traffic engineer does not have the resources to check the accuracy of each factor involved in the HCM model, let alone the algorithms used to generate the estimate of capacity.

On the other hand, the value of the base saturation flow rate may be adjusted to reflect the field measured value of capacity. The measured capacity value is first converted to an equivalent value of prevailing saturation flow rate for the lane group:

$s=c(g/C)=2,4000.50=4,800$ veh/hg

Using Equation 22-12, with the product of all adjustment factors of 0.80, the base saturation flow rate may be normalized:

$so=sNF=4,8003\times0.80=2,000$ pc/hg/ln

This normalized value may now be used in subsequent analyses concerning the subject intersection. If several such “normalizing” studies at various locations reveal a common areawide value, it may be more broadly applied.

It must be remembered, however, that this process does *not* mean that the actual base saturation flow rate is 2,000 pc/hg/ln, and it is assumed that the value of saturation flow rate is independent of g/C. If this value were observed directly, it might be quite different. It reflects, however, an adjusted value that normalizes the entire model for a number of underlying local conditions that renders some base values used in the model inaccurate.

Sample Problem 22-15 illustrates a “normalization” process for a single intersection or intersection approach. The 2016 HCM recommends that local calibrations be done for the local area as a whole. See Chapter 30 of the 2016 HCM for guidance on this process. The normalization illustrated here, however, can be a used to adjust differences between measured and predicted capacity values at individual locations.

Part IV Closing Comments

The HCM model for analysis of signalized intersections is complex and incorporates many submodels and many algorithms, some of which are not detailed in this chapter. The presentation herein focuses on key conceptual and methodological content, leaving the student to directly consult the HCM for additional details. Despite the great complexity of the HCM approach, it results from a relatively straightforward model concept to handle the myriad different conditions that could exist at a signalized intersection.

References

- 1. *Highway Capacity Manual, 6th Edition: A Guide for Multimodal Mobility Analysis*, Transportation Research Board, National Research Council, Washington, D.C., 2016.

- 2. Akcelik, R., "SIDRA for the Highway Capacity Manual," *Compendium of Papers*, 60th Annual Meeting of the ITE, Institute of Transportation Engineers, Washington, D.C., 1990.

- 3. Akcelik, R., *SIDRA 4.1 User's Guide*, Australian Road Research Board, Australia, Aug. 1995.

- 4. Teply, S., *Canadian Capacity Guide for Signalized Intersections*, 2nd Edition, Institute of Transportation Engineers, District 7—Canada, June 1995.

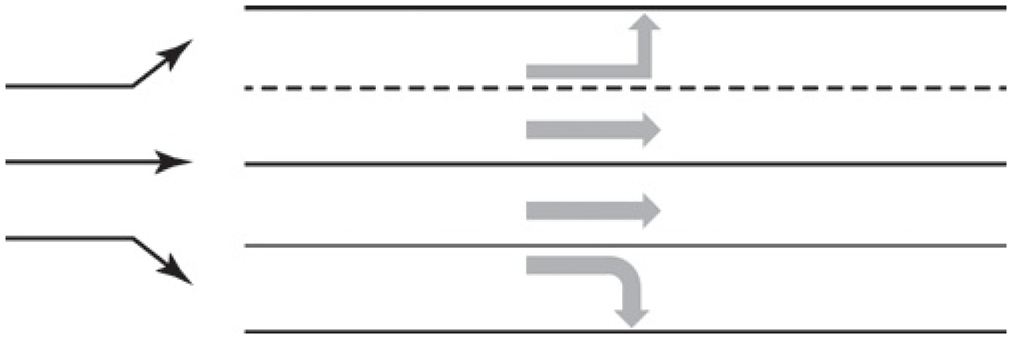
- 5. Zegeer, J.D., *Field Validation of Intersection Capacity Factors*, Transportation Research Record 1091, Transportation Research Board. 1986.

- 6. Strong, D., Rouphail, N., and Courage, K., "New Calculation Method for Existing and Extended HCM Delay Estimation Procedures," Proceedings 85th Annual meeting TRB, Washington, D.C., Jan 2006.

- 7. Strong, D. and Rouphail, N., "Incorporating the Effects of Traffic Signal Progression into the Proposed Incremental Queue Accumulation (IQA) Method," Proceedings 85th Annual TRB Meeting, Washington, D.C., Jan 2006.

Problems

1. 22-1. The eastbound approach shown below is a one-way street intersecting with a two-way street (two lanes per direction). Find the saturation flow rate and capacity of the eastbound approach, which has the following characteristics:



   - 60 s effective green time and ped time in a 100 s cycle
   - Four 11-foot lanes
   - 10% heavy vehicles
   - 3% upgrade
   - Parking on one side with 15 mvts/h within 250 feet of the stop line
   - 20 local buses/h
   - 8% right turns from an exclusive RT lane
   - 12% left turns from an exclusive LT lane
   - 100 peds/h in each crosswalk
   - No bicycles
   - A CBD location

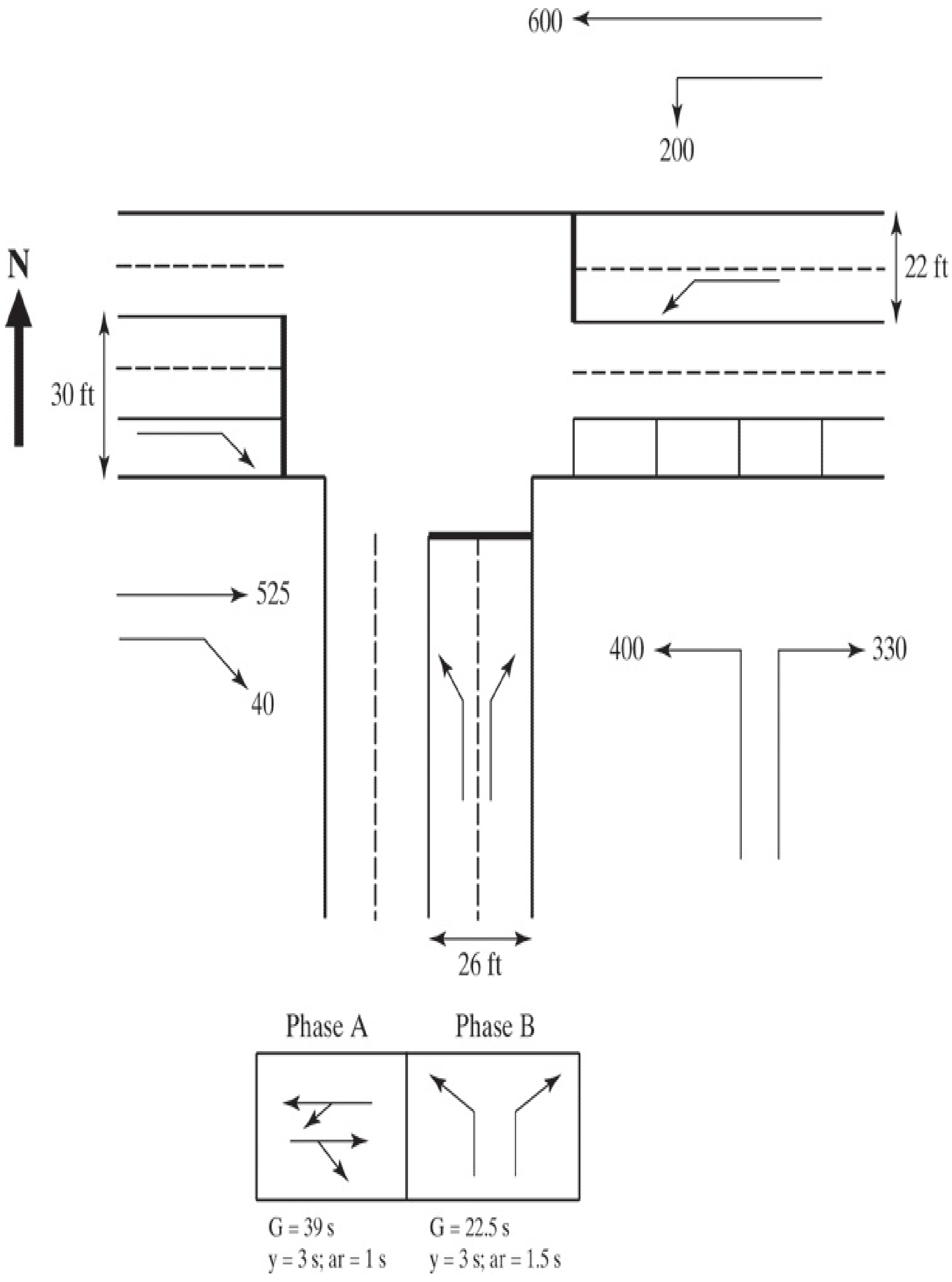
2. 22-2. Referring to the intersection shown on page 543, calculate v/c ratios and Xc. Saturation flow rates for each lane group are given in the table below. PHF=0.92.l1=2s.e=2s.

Saturation Flow Rates for Problem 22-2

Lane Group	N (lanes)	s (veh/hg/ln)
EB TH	2	1680
EB RT	1	1500
WB LT	1	774
WB TH	1	1697
NB LT	1	1676
NB RT	1	1500

Full Alternative Text

1. # Intersection for Problem 22-2



Full Alternative Text

2. 22-3. For the westbound approach only in the intersection of Problem 22-2, compute delay and LOS. The arrival type for WB is 5; EB is 2.

3. 22-4. Given the lane group delay computations in the table below, find the LOS for each lane group, and the delay and LOS for each approach and the overall intersection delay and LOS. Is the intersection working well? Why or why not?

Table for Problem 22-4

Item	Lane Group					
	EB T	EB R	WB LT	WB TH	NB LT	NB RT
C	80	80	80	80	80	80
X	0.941	0.159	0.359	0.627	0.960	0.886
d_1	**17.3**	**13.4**	**0.6**	**0**	**4.35**	**4.21**
T	0.25	0.25	0.25	0.25	0.25	0.25
c	607	271	605	1040	453	405
d_2	**24.5**	**1.2**	**1.7**	**2.9**	**33.5**	**23.7**
d_3	**0**	**0**	**0**	**0**	**0**	**0**
d	**41.8**	**14.6**	**2.2**	**2.9**	**37.8**	**27.9**

Full Alternative Text

4. 22-5. Using the data in the table below, determine the prevailing and base saturation headway and saturation flow rate.

Vehicle in Queue	Cycle 1	Cycle 2	Cycle 3	Cycle 4	Cycle 5
1	2.8	2.9	3.0	3.1	2.7
2	2.6	2.6	2.5	3.5H	2.6
3	3.9L	2.3	2.2	2.9	2.5
4	10.2H	2.1	2.0	2.5	2.0
5	8.7	4.0L	1.9	2.2	1.9
6	3.0	9.9L	2.2	2.0	1.9
7	2.9	9.8	2.9H	1.9	3.6H
8	5.0	3.3	2.6	1.8	9.0
9	7.1	2.8	2.1	7.0	4.0
10	9.0	2.2	4.0	8.0	4.9
11		1.9	5.0		9.0
12		5.5			
13		4.0			

Notes: H = heavy vehicle; L = left turn; Underline = last vehicle in standing queue.

22.2-26 Full Alternative Text

Chapter 23 Planning-Level Analysis of Signalized Intersections

In Chapter 22, a detailed, complex methodology for analyzing a signalized intersection was presented. In this chapter, a relatively simple analysis methodology is presented. This method is the planning method for signalized intersections in the 2016 *Highway Capacity Manual* (HCM) [1] and it is based on critical movement analysis (CMA).

In 1980, the Transportation Research Board (TRB) issued a set of preliminary capacity and level of service (LOS) analysis methodologies (*Interim Materials on Highway Capacity*) in advance of the anticipated 1985 HCM [2]. It included a relatively straightforward method for analysis of signalized intersections called “Critical Movement Analysis” [3]. The methodology could be easily implemented by hand in a reasonable amount of time, and based LOS determinations on the “sum of critical lane flows.” Although general delay estimates for each LOS for the intersection as a whole were provided by tabulation, there was no attempt to estimate average delays for individual movements or approaches.

The earliest work on critical lane analysis of signalized intersections goes back to the original concepts of saturation headways and lost time, developed by Bruce Greenshields. Its application to analysis originally appeared in 1961 in a paper by Capelle and Pinell [4]. Messer and Fambro [5] produced a definitive methodology in 1977, which was adapted for inclusion in the *Interim Materials*. By the time the 1985 HCM was published, however, Messer had built the basic procedure into a far more complex analysis methodology. Subsequent revisions have served to further complicate the approach.

As a result, there are traffic agencies that still (at this writing) use the methodology of the *Interim Materials* to analyze signalized intersections, including the California Department of Transportation. Because of the great complexity of the HCM approach, there are frequent calls for a return to a simpler approach.

23.1 The TRB Circular 212 Methodology

The methodology of TRB Circular 212 (*Interim Materials on Highway Capacity*) provided two levels of analysis: planning and operations/design. The first was done entirely in units of mixed veh/h with few adjustments. Average or “typical” conditions are assumed. The planning approach was as close to a “back of the envelope” approach as could be accomplished. The most complex signalized intersections could be analyzed in minutes. The operations and design model provided more detailed adjustments for heavy vehicle presence, local bus presence, lane width, parking conditions, and other prevailing conditions. Even with these adjustments included, the most complex intersection could be analyzed by hand in less than 15 minutes.

Unlike the HCM operational method presented in Chapter 22, critical lane or CMA applies all adjustment factors to the demand volume or flow rate. Thus, volumes in veh/h are inflated by adjustment factors to reflect “through passenger car units (tpc),” as is at least partially done for signal timing.

While the Circular 212 methodologies continue to see some usage, this chapter presents the two-level planning methodology from the 2016 HCM, which is based upon similar concepts but adds some additional details that allow for an analysis that is sensitive to more underlying conditions.

23.2 The 2016 HCM Planning Methodology

While similar in concept to the Circular 212 methodology, the 2016 HCM planning methodology adds several features:

- Additional adjustment factors covering a broader range of prevailing conditions have been added.
- Theoretical delay models are applied to estimate delay and LOS.
- While the HCM detailed analysis procedure applies all adjustments to saturation flow rate, this methodology applies all adjustments to demand flow rates, as in previous CMA approaches.
- The methodology is divided into two distinct parts.
  - Part I is a CMA methodology that requires a minimum amount of data. The output of Part I is the Intersection *v/c* ratio, X_c, which is based on the critical movements (as defined in Chapter 22), and a general description of the intersection's relationship to capacity (under, near, or over).
  - Part II is a delay and LOS methodology. The user may end the analysis after Part I if delay and LOS are not desired.

23.2.1 Part I of the Methodology

1. Step 1: Specify Input Data

   For Part I of the methodology, the following inputs are needed:

   - Complete information on number of lanes and lane use (exclusive or shared)
   - Hourly demand flow rates (in veh/h for prevailing conditions) for each movement

   The following additional optional data may be entered or defaults are provided:

   - Data specifying the characteristics of each movement's demand, including: percent of heavy vehicles, parking conditions and activity, and pedestrian flows. If movement characteristics are unknown, defaults provided may be used.
   - Base saturation flow rate
   - Cycle length
   - Effective green times (required when there is protected/permissive left-turn phasing).

   Part II of the methodology requires quality of progression as an additional input.

2. Step 2. Define Left-Turn Treatment and Phase Sequence

   If the left-turn operational mode is known, it is entered directly into the methodology. If unknown, the following guidelines may be used. Note that these guidelines differ from those used in signal timing given in Chapters 19 and 20.

There are three tests to determine whether a left-turn movement should be a protected phase:

1. The left-turn volume is greater than or equal to 240 veh/h.

2. The cross-product of left-turn volume and the total opposing through volume is greater than or equal to the thresholds shown in Table 23.1.

3. If there is more than one exclusive left-turn lane on the approach.

Table 23.1: Left-Turn Phase Check for Cross-Product Thresholds

Number of Opposing Lanes	Threshold of Cross-Product
1	≥50,000
2	≥90,000
3	≥110,000

Table 23.1: Full Alternative Text

If any one of the three tests is met, a protected left turn should be provided, regardless of the other two tests. There are other considerations for choosing to provide left-turn protection that are not an explicit part of the planning methodology, which are described in Chapter 19.

The planning methodology can *only* be used with the following types of left-turn phasing:

- Protected-only left turns, including opposing left turns moving together, lead-lag protected phasing, or split phasing
- Permitted-only left-turn phasing
- Protected-permitted left-turn phasing (Note: For protected/permitted phasing (also called compound phasing), the effective green times for each portion of the phase must be specified.)

The method cannot be used when only one opposing left turn is protected. If one left turn requires protection and the opposing left does not, the methodology assumes both are protected.

3. Step 3. Define Lane Groups

   Movements are analyzed in lane groups, which are defined differently than in the operational methodology. There are two types of lane groups:

   - Exclusive turning lanes: Left-turn only or right-turn only lanes are separate lane group(s).
   - All remaining lanes are combined, including all through-only lanes and shared lanes.

   Note that a shared lane may behave as an exclusive-only turn lane when the turning flow rate is high or impeded by the opposing movement. In this case, the shared lane should be defined as a de facto exclusive turn lane.

4. Step 4. Convert Demand Volumes under Prevailing Conditions to Demand Flow Rates in Through Passenger Car Units under Base Conditions

   As noted previously, all CMA methodologies make adjustments on the demand side of the equation as opposed to the capacity side of the equation as in the HCM operational methodology. The conversion is accomplished using Equation 23-1. Table 23.2 shows the values for each of the equivalency factors in Equation 23-1.

vadj=VEHVEPHFELTERTEpELUEother [23-1]

Table 23.2: Adjustment Factors to Convert to Through Passenger Car Equivalents

<table>
<tr><th>Characteristic Adjusted</th><th>Equivalency</th></tr>
<tr><td>Adjustment for Heavy Vehicles, E_{HV}
The number of through cars displaced (or equivalent) to one heavy vehicle, $E_T = 2$</td><td>$E_{HV} = 1 + 0.01P_{HV}(E_T - 1)$
Where:
P_{HV} = Percent heavy vehicles
Default $P_{HV} = 3\%$</td></tr>
<tr><td>Adjustment for Peak Hour Factor, E_{PHF}
Adjust volumes to account for fluctuations within the peak hour, to analyze the peak 15 minutes within the peak hour</td><td>$E_{PHF} = 1/PHF$
Default $PHF = 0.92$</td></tr>
<tr><td>Adjustment for Left Turns, E_{LT}</td><td>Protected-only Phasing $E_{LT} = 1.05$
Permitted-only Phasing
<table>
<tr><th>Opposing Volume, veh/h = opposing through + right turners</th><th>E_{LT}</th></tr>
<tr><td>< 200</td><td>1.1</td></tr>
<tr><td>200 – 599</td><td>2.0</td></tr>
<tr><td>600 – 799</td><td>3.0</td></tr>
<tr><td>800 – 999</td><td>4.0</td></tr>
<tr><td>≥1000</td><td>5.0</td></tr>
</table></td></tr>
<tr><td>Adjustment for Right Turns, E_{RT}
(Dependent upon pedestrian activity)</td><td>Protected-Permissive Left-turn Phasing: See Step 4a
<table>
<tr><th>Level of Pedestrian Activity</th><th>Pedestrian Volume</th><th>E_{RT}</th></tr>
<tr><td>None or Low</td><td>0–199</td><td>1.2</td></tr>
<tr><td>Moderate</td><td>200–399</td><td>1.3</td></tr>
<tr><td>High</td><td>400–799</td><td>1.5</td></tr>
<tr><td>Extreme</td><td>≥ 800</td><td>2.1</td></tr>
</table></td></tr>
<tr><td>Adjustment for Parking Activity, E_p</td><td>
<table>
<tr><th>Level of Parking Activity</th><th>Number of Lanes in Lane Group</th><th>E_p</th></tr>
<tr><td>No Parking</td><td></td><td>1.00</td></tr>
<tr><td>Parking Allowed in Adjacent Lane Group</td><td>1</td><td>1.20</td></tr>
<tr><td></td><td>2</td><td>1.10</td></tr>
<tr><td></td><td>3</td><td>1.05</td></tr>
</table></td></tr>
</table>

[23.2-2 Full Alternative Text]

Characteristic Adjusted	Equivalency		
Adjustment for Lane Utilization, E_{LU}	Lane Group	Number of Lanes in Lane Group	E_{LU}
	Through or Shared	1	1.00
		2	1.05
		≥3	1.10
	Exclusive Left	1	1.00
		≥2	1.03
	Exclusive Right	1	1.00
		≥2	1.13
Adjustment for Other Conditions, E_{other}	User decision to adjust for factors not accounted for by other factors above, such as, buses, grade, or area type		

[23.2-3 Full Alternative Text]

where:

vadj=movement flow rate as equivalentthrough passeng
hour factor equivalency factor, ELT=left-
turn equivalency factor, ERT=right-
turn equivalency factor, Ep=parking activity equivalenc

5. Step 4a. Finding Left-Turn Equivalency Factor for Protected/Permitted Left Turns

   The case of compound left turns, that is, a phasing plan that has both protected and permitted left-turn portions, is more complex. The HCM methodology of Chapter 22 treats the two portions of a compound phase separately. This CMA method uses a simpler approach. A single left-turn equivalent is defined, which is applied across the entire compound phase, treating it as a single time period. It weights the individual equivalents for the protected and permitted portions of the phase in proportion to the effective green times of each portion, as shown in Equation 23-2. Note that only for compound left-turn phasing must the green times be known for Part I of the methodology.

$ELTC=(ELTPT*gLTPT)+$
$(ELTPM*gLTPM)gLTPT+gLTPM$ [23-2]

where:

ELTC=left-
turn equivalent for compound LTphasing,ELTPT=left-
turn equivalent for protected portionof the compound LT
turn equivalent for permitted portionof the compound LT

6. Step 5. Assign Flow Rates into Lane Groups

   The adjusted volumes calculated in Step 4 are assigned to the appropriate lane group and then divided by the number of lanes in that lane group to find the flow rate per lane in each lane group, using Equation 23-3.

   $vi=vadj,iNi$ [23-3]

   where:

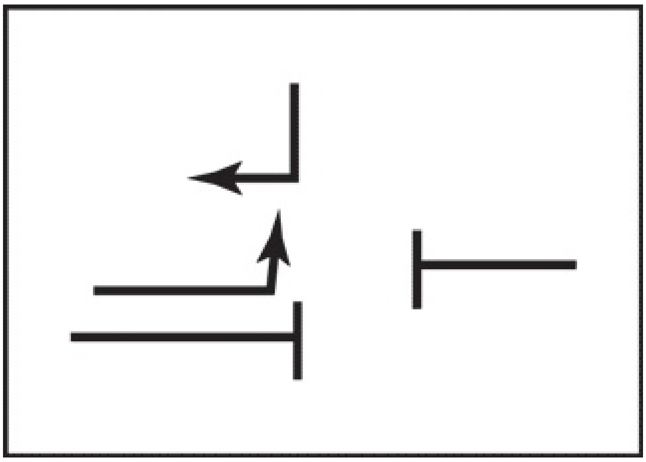
   vi=flow rate per lane for lane group i, tpc/h/ln,vadj,i=equi

   Before continuing to Step 6, check for a de facto turn lane. Either a shared left/through lane or a shared through/right lane can behave as a de facto exclusive turn lane. It can be assumed that a shared lane behaves as a de facto exclusive turn lane if the flow rate of the turning traffic (veh/h) is greater than the through-car equivalent adjusted flow rate in the shared lane (tpc/h/ln).

7. Step 6. Finding Critical Lane Groups

   The critical lane groups that control the signal timing must be determined. The process used to find the critical lane groups is basically the same as that described in Chapter 19. However, an additional dynamic is taken into consideration. When right-turning vehicles are allowed to proceed during a cross-street's protected left-turn phase, as shown in Figure 23.1, the right-turn flow rate is reduced by the number that could occur during the left-turn phase. In Figure 23.1, the southbound right-turn flow rate proceeds at the same time as the eastbound protected left-turn movement. The southbound flow rate is thus reduced by the same flow rate as the eastbound left-turn flow rate.

Figure 23.1: Phase with Exclusive Right Turn Permitted During Cross-Street Protected Left Turn



Once the critical path is determined, the sum of the critical lane flow rates (in tpc/h) is calculated.

8. Step 7. Calculate Cycle Length

If the cycle length is known, proceed to Step 8. If the cycle length is unknown, then local policies or pedestrian needs should be considered. Without any of this information, it is recommended that the cycle length is set, by assuming 30 seconds is needed for each critical phase.

9. Step 8. Calculate Intersection Capacity

   Since CMA applies all conversions to the demand side of the equation, capacity is not adjusted, and is expressed in tpc/h/ln under equivalent base conditions. The base saturation flow rate is entered by the analyst as an input, often using the default values from the HCM:

   - 1900 pc/hg/ln for metropolitan areas with a population ≥250,000, or
   - 1750 pc/hg/ln for populations ≥250,000.

   Intersection capacity is then found using Equations 23-4 and 23-5.

   cI=soC−LC [23-4]
   L=nc×lp [23-5]

   where:

   cI=Intersection capacity, tpc/h/ln,so=Base saturation flow

   Note that the intersection capacity is stated as a maximum sum of critical lane volumes. The operational method (Chapter 22), produces capacities for each lane group in veh/h.

10. Step 9. Calculate Intersection Volume-to-Capacity Ratio, X_c

    The *v/c* ratio for the intersection as a whole is calculated from the sum of the flow rates of the critical lanes, using Equation 23-6.

Xc=∑ivcicI [23-6]

where:

Xc=critical volume-to-
capacity ratio for theintersection,∑ivci=sum of the critical
fromEquation 23-14 in this case, the maximum
sum of critical lane flow rates).

11. Step 10. Define Intersection Sufficiency

Intersection sufficiency is defined as the ability of the intersection to process the given demand. The intersection sufficiency is based on the critical v/c ratio for the intersection, X_c, as either over, near, or under capacity. Table 23.3 gives the definitions of intersection sufficiency based on the X_c.

1. # Table 23.3: Capacity Assessment Levels and Descriptions for Intersection Sufficiency

Critical v/c Ratio X_c	Capacity Assessment	Description
<0.85	Under	All demand may be accommodated; low to moderate delays
0.85–0.98	Near	Demand is near capacity and some lane groups require more than one cycle to clear the intersection; all demand, however, may be accommodated with the analysis period; moderate to high delays
>0.98	Over	Demand may require multiple cycle lengths to clear; delays are high and queues are long

(Source: *Highway Capacity Manual, 6th Edition: A Guide for Multimodal Mobility Analysis*, Transportation Research Board, Washington, D.C., 2016, modified from Exhibit 31-37.)

Table 23.3: Full Alternative Text

Phase II continues with Steps 11–13, and results in determination of lane group delays and levels of service.

23.2.2 Part II of the Methodology

1. Step 11. Calculate Effective Green Times

   If the signal timing is known, then skip this step and proceed to Step 12. If the signal timing is unknown, effective green time is found, as described in Chapter 19, by dividing the available effective green time of the cycle among the critical phases. The available green time in the cycle is found by deducting the total lost time per cycle, *L*, from the cycle length, *C*, using Equation 23-7.

   gTOT=C − L [23-7]

   where:

   gTOT=total effective green time in the cycle, s,C=cycle le

   The total effective green time is then allocated to the various critical phases, using Equation 23-8.

   gci=gTOT×VciVc [23-8]

   where:

   gci=effective green time for critical lane i, s,vci=flow rate

   If there is protected/permissive phasing, the noncritical lane groups may be assigned effective green times based on their relation to the critical lane groups, as described in Chapter 19.

2. Step 12. Capacity and v/c Ratios

   The capacity and *v/c* ratio is calculated for each lane group, using Equations 23-9 and 23-10. In Equation 23-9, the base saturation flow rate (usually defaulted as 1,900 tpc/hg/ln for metropolitan areas with population ≥250,000; 1,750 pc/hg/ln otherwise) is multiplied by the *g/C* ratio to account for the fact that the lane group does not have a full hour of green time, and then multiplied by the number of lanes to get the capacity of the lane group.

ci=soNigiC [23-9]
Xi=Nivici [23-10]

where:

ci=capacity of lane group "i," tpc/h,so=base saturation flo

The intersection capacity and intersection *v/c* ratio is computed using Equations 23-11 and 23-12.

csum=so∑i=1ncpgc,iC [23-11]
Xc=Vccsum [23-12]

where:

csum=Intersection capacity per lane, tpc/h/ln, andXc=Inte

3. Step 13. Delay and Level of Service

   The basic flaw in the CMA method of Circular 212 is that it does not predict delay values, but rather assumes that, given estimated sums of critical lane volumes, certain delay ranges would hold.

   The problem is that the sum of critical lane flows, or indeed *v/c* ratios, does not correlate well to delay. Low *v/c* values imply large amounts of unused green time, which *increases* delay. A low *v/c* ratio may be the *problem*, and is rarely the *solution* to the problem. It is, therefore, sometimes necessary to make an estimate of delay to appropriately assess the operation of a signalized intersection.

   The HCM methodology for estimating delays has become more complex over the years. A more simplistic approach is taken here using simple theoretical equations as a base and a simplified adjustment for progression quality. Additionally, the approach taken here also assumes no preexisting queues at the beginning of the analysis period.

Given these simplifications, delay is computed as:

di=d1i PF+d2id1i=0.5 C [1−(giC)]21−[min (1,Xi)*(giC)]d2i=225 [(Xi−1)+(Xi−1)2+16 Xici] [23-13]

where:

di=average delay per vehicle in lane group i, s/veh,d1i=av

All other variables are as previously defined.

The uniform delay term, d_1, is Webster’s Equation, in the form used in the 2000 HCM. The overflow delay term is from Akcelik’s Equation in Reference [3], in the form still used in the 2016 HCM, with certain assumed values:

- Analysis period=0.25 h
- Adjustment for type of controller=0.50
- Adjustment for upstream filtering=1.0

Recommended progression adjustment factors are shown in Table 23.4. They have been simplified from the 2000 HCM to define three general categories of progression: good, random, and poor. For good progression, platoons arrive during the green interval, and/or most vehicles arrive on green. Random arrivals describes situations in which a signal is not coordinated, or isolated. The proportion of vehicles arriving on green is approximately equal to the *g/C* ratio for the lane group. In poor progression, platoons arrive during the red interval and/or most vehicles arrive on red.

Table 23.4:

Progression Adjustment Factors

Quality of Progression	Progression Factor, *PF*
Good	0.70
Random	1.00
Poor	1.25

Table 23.4: Full Alternative Text

Although not shown here, delays may be aggregated for approaches, and for the intersection as a whole. The process is the same as that used in the operational HCM model of Chapter 22. Delay values may be compared to the criteria in Table 23.5 to determine levels of service for lane groups, approaches, and/or the intersection as a whole. If any *v/c* ratio is greater than 1.0, then it is designated as LOS F, no matter what the delay value.

1. **Table 23.5: Levels of Service for Signalized Intersection Lane Groups and Approaches**

Level of Service	Delay (s/veh)
A	≤ 10
B	$>10-20$
C	$>20-35$
D	$>35-55$
E	$>55-80$
F	>80

(Source: Reprinted with permission from *Highway Capacity Manual, 6th Edition: A Guide for Multimodal Mobility Analysis,* National Academy of Sciences, Courtesy of National Academies Press, Washington, D.C., Exhibit 31-20, pgs 31–64.)

Table 23.5: Full Alternative Text

Sample Problem 23-1: Using the 2016 HCM Planning Level Intersection Analysis Methodology

1. Step 1: Specify Input Data

   Figure 23.2 shows the intersection of a two two-way arterials. The E/W approaches have two lanes in each direction, a shared left-through lane and a shared through-right lane. The N/S approaches have four lanes per direction, two through-only lanes, one shared through-right lane, and an exclusive left-turn bay. All relevant information is shown in the figure.

Figure 23.2: Figure for Sample Problem

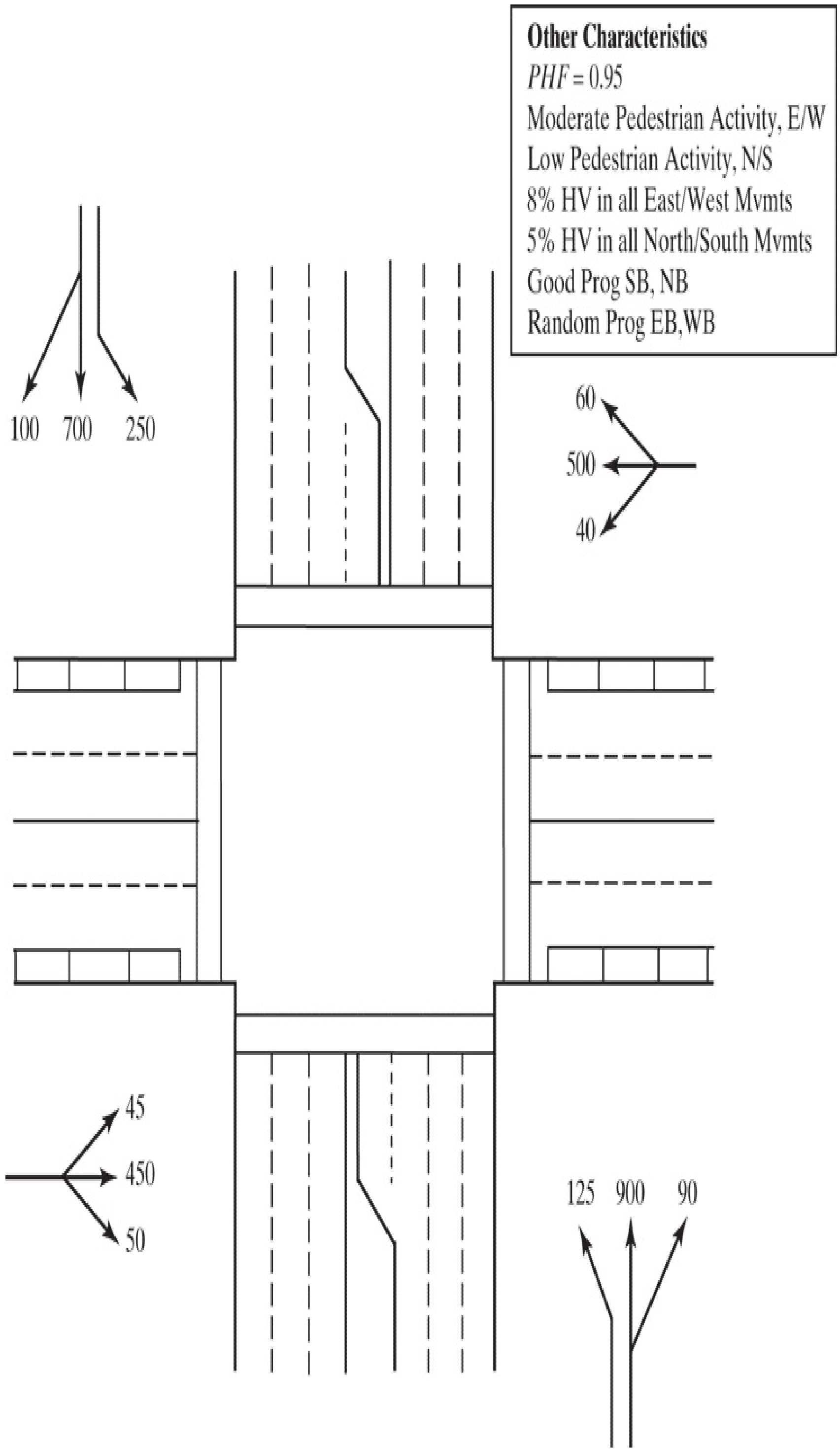


Figure 23.2: Full Alternative Text

2. Step 2. Define Left-Turn Treatment and Phase Sequence

   Left turns are checked to determine whether the intersection will accommodate them without an exclusive left-turn phase. There are three left-turn checks.

- Check #1: Is there more than one exclusive left-turn lane on any of the approaches?
- Check #2: Are there more than 240 left turns?
- Check #3: Compare the cross-product of the left-turn volume and opposing through volume to the values in Table 23.1.

Table 23.6 gives the results of the left-turn checks. It can be seen that based on the three checks, only the southbound approach requires a protected left-turn phase. However, the planning methodology does not handle cases where opposing lefts do not both have a protected phase. Thus in this case, the northbound approach will be analyzed with a protected only phase, moving concurrently with the southbound left-turning vehicles. Note that only one of the checks needs to be met in order to require a protected left-turn phase.

Table 23.6: Left-turn Check Results

Left-turn Check Number	Eastbound	Westbound	Northbound	Southbound
Check #1	No	No	No	No
Check #2	No	No	No	Yes
Cross-Product	45 × 500 = 22,500	40 × 450 = 18,000	125 × 700 = 87,500	250 × 900 = 225,000
Threshold	>90,000	>90,000	>110,000	>110,000
Check #3	No	No	No	Yes
Final Decision	Permitted	Permitted	Protected (based on SB lefts)	Protected

Table 23.6: Full Alternative Text

3. Step 3. Define Lane Groups

   Lane groups are either exclusive-only turn lanes by themselves or shared lanes analyzed together with through-only lanes. The eastbound and westbound approaches are each analyzed as one lane group, left/through/right (LTR) lane group. The northbound and southbound approaches are analyzed as two lane groups. The exclusive left-turn lane is one lane group and all other lanes are combined together into the second lane group.

4. Step 4. Convert Demand Volumes under Prevailing Conditions to Demand Flow Rates in Through Passenger Car Units under Base Conditions

   Converting demand volumes in vehicles per hour into demand flow rates in through passenger car units per hour is accomplished using the equivalency values found in Table 23.2. For this case, equivalency values are found for heavy vehicles, peak-hour factor, right turns, left turns, and lane utilization. Demand flow rate is the product of the demand volume and each of the equivalency values, using Equation 23.1. Table 23.7 shows the results.

Table 23.7: Adjustment of Demand Volumes

	Eastbound			Westbound			Northbound			Southbound		
	L	TH	R	L	TH	R	L	TH	R	L	TH	R
Demand Volume,veh/h	45	450	50	40	500	60	125	900	90	250	700	100
E_{HV}	1.08	1.08	1.08	1.08	1.08	1.08	1.05	1.05	1.05	1.05	1.05	1.05
E_{PHF}	1.05	1.05	1.05	1.05	1.05	1.05	1.05	1.05	1.05	1.05	1.05	1.05
E_{RT}	1	1	1.3	1	1	1.3	1	1	1.2	1	1	1.2
E_{LT}	2	1	1	2	1	1	1.05	1	1	1.05	1	1
E_{P}	1	1.1	1.1	1	1.1	1.1	1	1	1	1	1	1
E_{LU}	1.05	1.05	1.05	1.05	1.05	1.05	1	1.1	1.1	1	1.1	1.1
E_{Other}	1	1	1	1	1	1	1	1	1	1	1	1
Adjusted Flow Rate, tcu/hr	107	589	85	95	655	102	145	1094	131	290	851	146

[23.2-9 Full Alternative Text](#)

Lane Group Flow Rate, tcu/hr		782			852		145	1226		290	997	
Lane Group Flow Rate per Lane, tcu/hr/ln		391			426		145	409		290	332	
Critical Lanes					X			X		X		

[23.2-10 Full Alternative Text](#)

5. Step 5. Assign Flow Rates into Lane Groups

   The adjusted flow rates are assigned into lane groups, and the lane group flow rate per lane is calculated, using [Equation 23-3](#).

6. Step 6. Finding Critical Lanes

   For each phase, the critical lane is the lane with the highest flow rate. This lane controls the length of time required for a particular phase.

   The eastbound and westbound movements proceed into the intersection at the same time, in a single phase and single lane group each direction. Comparing the eastbound LTR lane group flow rate per lane (391 tcu/h/ln) to the westbound LTR lane group flow rate per lane (426 tpc/h/ln), the higher of the two is the critical lane. In this case, it is the westbound LTR lane group.

The northbound and southbound movements are serviced in two phases. The northbound and southbound left turns are serviced in a protected left-turn only phase. The through and right-turn movements are serviced together in a separate phase. For the protected left-turn phase, the northbound left-turn lane flow rate (145 tcu/h/ln) is compared with the southbound left-turn lane flow rate (290 tcu/h/ln). The southbound left-turn lane flow rate is the higher of the two, and thus is the critical lane. The NB and SB through and right-turning vehicles are serviced in a separate lane group and a separate phase. The northbound through/right flow rate (409 tcu/h/ln) is higher than the southbound through/right flow rate (332 tcu/h/ln) and thus is the critical lane. Table 23.7 summarizes all of the computations of Steps 4, 5, and 6.

7. Step 7. Calculate Cycle Length

   Since the cycle length is not known, it is estimated by assuming that each critical phase requires 30 seconds. In this example, there are three critical phases and the cycle length is estimated to be 90 seconds.

8. Step 8. Calculate Intersection Capacity

   Using a base saturation flow rate of 1,900 pc/hg/ln and a lost time per phase of 4 sec/phase, intersection capacity is found using Equations 23-4 and 23-5.

   $$L = n_c \times l_p = 3 \times 4 = 12 \qquad c_I = s_o \frac{C - L}{C} = 1900 \frac{90 - 12}{90} = 1647 \text{ tcu/h/ln}$$

   where all variables are as defined previously.

9. Step 9. Calculate Intersection Volume-to-Capacity Ratio,

X_c

Equation 23-6 is used to calculate the v/c ratio for the intersection as a whole with all variables as previously defined.

Xc=∑ivcicI=426+409+2901647=11251647=0.68

10. Step 10. Define Intersection Sufficiency

    Intersection sufficiency is defined by the relationship of the sum of the critical flow rates to the intersection capacity. From Table 23.3, because Xc<0.85, this intersection is defined as "*Under* capacity."

    This is the end of the first part of the planning methodology. The analyst may stop here or continue in order to calculate delay and LOS.

11. Step 11. Calculate Effective Green Times

    The available effective green time in the cycle must be divided among the signal phases. Available green time is found using Equation 23-7. Effective green time per phase is found using Equation 23-8. The calculations are shown in the following equations with all variables as previously defined.

    gTOT=C−L=90−12=78 s
    gci=gTOT×VciVcgE/W=78×4261125=30 s/vehg=NSL=7

    where:

    gE/W=effective green time for the eastbound/westbound
    turn phase, andgNSTR=effective green time for the northb

12. Step 12. Capacity and v/c Ratios

    The capacity and *v/c* ratio for each lane group are found using Equations 23-9 and 23-10. Since CMA applies all conversions to the demand side of the equation, capacity is unadjusted, and is expressed in pc/h under equivalent base conditions. The base saturation flow rate used is 1,900 pc/hg/ln, and capacity is thus multiplied by the *g/C* ratio and the number of lanes.

ci=soNigiC   cEB=cWB=1900×2×3090=1267   cNE

The lane group *v/c* ratio is then calculated, as shown in Table 23.8. Intersection capacity and intersection *v/c* ratio are calculated as follows.

csum=so∑i=1ncpgc,iC=1900×30+20+2890=1647Xc=Vc

Table 23.8: Results of Steps 12 and 13

Lane Group	EB LTR	WB LTR	NB Left	NB TR	SB Left	SB TR
Lane Group Flow Rate, tpc/h	782	852	145	1226	290	997
Lane Group Flow Rate per Lane, tpc/h/ln	391	426	145	409	290	332
Critical Lanes		X		X	X	
Lane Group Capacity	1267	1267	422	1773	422	1773
v/c ratio	0.62	0.67	0.34	0.69	0.69	0.56
Uniform Delay, d_1 s/veh	25.2	25.8	29.5	27.2	32.1	25.9
Progression Factor	1.00	1.00	0.70	0.70	0.70	0.70
Incremental Delay, d_2, s/veh	2.3	2.9	2.2	2.2	8.8	1.3
Control Delay, s/veh	27.4	28.7	22.8	21.3	31.3	19.4
Lane Group LOS	C	C	C	C	C	B
Approach Delay, s/veh	27.4	28.7	21.5		22.1	
Approach LOS	C	C	C		C	
Intersection Delay, s/veh	24.1					
Intersection LOS	C					
Intersection Capacity, tpc/h/ln	1647		Intersection v/c		0.68	

Table 23.8: Full Alternative Text

13. Step 13. Delay and Level of Service

Calculations for delay are made using Equations 23-13. The progression factor for the eastbound and westbound direction is 1.00, since arrivals are random. For the northbound and southbound directions, the progression factor used is for "good" progression and is found in Table 23-4 to be 0.70. Table 23-5 is then used to lookup level of Service. The results of the delay and LOS calculations are shown in Table 23.8.

23.3 Closing Comments

The 2016 HCM planning methodology presented in this chapter is similar in methodology to the HCM model for pre-timed signals, but with many defaults and much simpler signal timing, phasing, and delay calculations. Lane groups here are the movement groups of the operational method.

Although relatively simple and straightforward, the method is not quite at the “back-of-the-envelope” level, but can be done by hand in a reasonable amount of time (under 10 to 15 minutes, depending if timing is known or not). Spreadsheets can be programmed to fully automate the process. Because of its straightforward approach, it is more easily understood than the operational HCM methodology, which is a “black box” to many users.

It is noted, however, that both the HCM operational method and this planning methodology are fundamentally two *predictions*. The *v/c* output of both methods is based upon the capacity estimate, which directly flows from the saturation flow rate prediction. It is has been some time since a substantial data base of both saturation flow rates and delays has been collected and compared to any model.

References

- 1. *Highway Capacity Manual, 6th Edition: A Guide for Multimodal Mobility Analysis*, Transportation Research Board, Washington, D.C., 2016.

- 2. *Highway Capacity Manual*, Special Report 209, Transportation Research Board, Washington, D.C., 1985.

- 3. *Interim Materials on Highway Capacity*, Circular 212, Transportation Research Board, Washington, D.C., 1980.

- 4. Capelle, D.G. and Pinell, C., “Capacity Study of Signalized Diamond Interchanges,” Highway Research Bulletin 291, Transportation Research Board, Washington, D.C., 1961.

- 5. Messer, C.J. and Fambro, D.B., “A New Critical Lane Analysis for Intersection Design,” 56th Annual Meeting of the Transportation Research Board, Washington, D.C., January 1977.

- 6. *Traffic Engineering Handbook*, ITE, John Wiley and Sons, 2016.

Problems

1. 23-1 Analyze the intersection shown in the figure below. The following information is available for this intersection: (i) progression quality = good (EB); poor (WB), and random (NB); (ii) 5% HV in all movements; (iii) 300 ped/h crossing in each crosswalk of the main street, 400 ped/h crossing the side street; (iv) peak-hour factor=0.88. The timing is known for this intersection and is shown in the figure.

1. **Intersection for Problem 23-1**

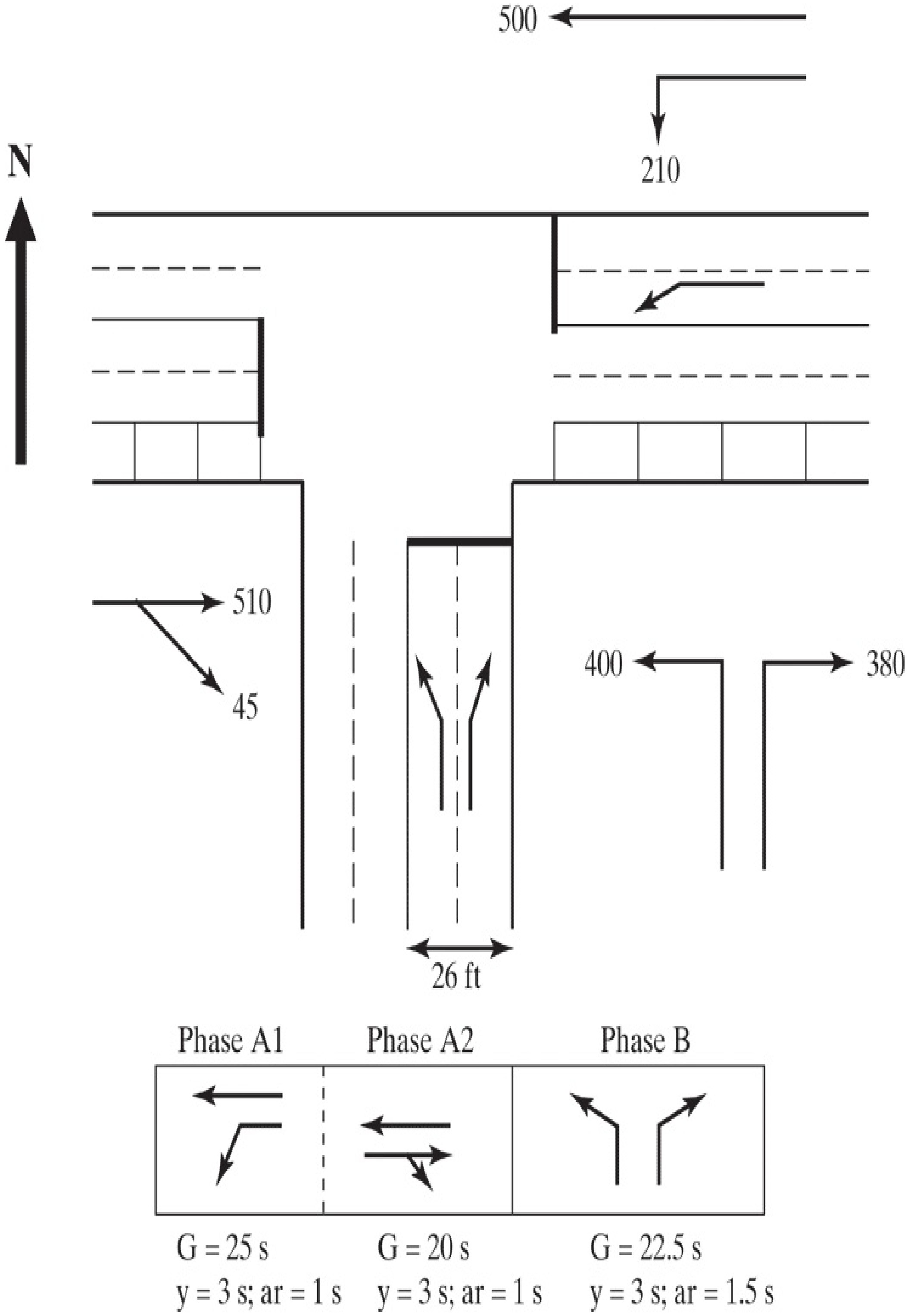
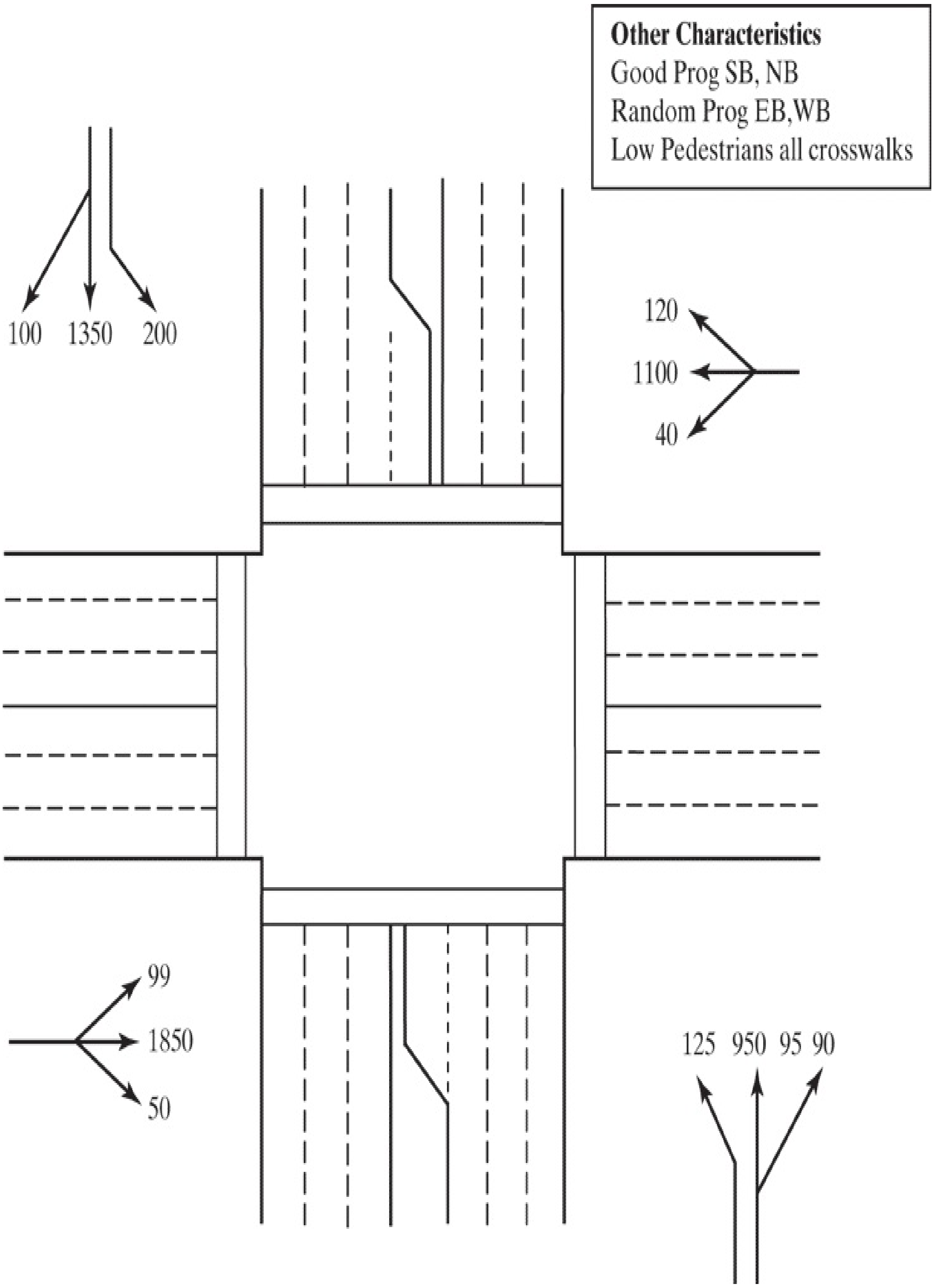


Figure 23-1 Full Alternative Text

2. 23-2 Analyze the intersection in the figure on page 558 for intersection sufficiency. Minimal data are available so defaults for the optional data should be used.

1. # Intersection for Problem 23-2



[Full Alternative Text]

Chapter 24 Urban Streets and Arterials: Complete Streets and Level of Service

Complete streets are multimodal streets; they are streets that are designed to ensure that all users have the ability to make safe trips, comfortable trips, and efficient trips. Ensuring that users have equal access to transportation, creates more livable cities and communities. Complete streets provide a mix of transportation alternatives that consider all modes of transportation equitably. Equity is an important element of a complete street. It means that vehicles (private and public), pedestrians, and cyclists, regardless of age and/or ability, have equal access to safe transportation. Smart Growth America [1] describes a complete street as a street that makes crossing the street, walking to shops, walking to transit stations, and bicycling to work easy and safe, as well allowing buses to run on time. A complete street allows all individuals to have equitable access to transportation regardless of where they live, their income level, and/or their disabilities.

Accomplishing these goals means considering the street in a different manner than transportation planners and agencies have in the past. Streets were generally designed first for automobiles, with transit and pedestrians a distant second, and no thought for bicycles. Today, the priorities for the transportation system have changed, and many cities and communities have adopted the complete street philosophy along with a vision zero program. The goal of vision zero is zero fatalities. This is accomplished with proper design and law enforcement. Vision zero planning began in Sweden and has now been adopted by many cities in the United States and being considered by many more [2], but vision zero is only one aspect of complete streets. Complete streets do promote safety but have other benefits as well.

The benefits of complete streets include the following [3, 4, 5]:

- Safer streets

- Equitable mobility
- Promoting exercise and thus better health
- Economic development. Complete streets have been shown to revitalize community economies by providing better access to shops, restaurants, and workplaces. This raises property values and encourages more private investment that provides more jobs.
- Connected networks
- Lowering transportation costs
- Improving air quality
- Improved landscaping and lighting, which encourage walking and biking.

24.1 Designing Urban Streets

There are many manuals that provide guidance for designing multimodal streets that are safe for all users. The National Association of City Transportation Officials Urban Street Design Guide [*6*], the Pedestrian [7] and Bicycle [8] Safety Guides, and the Pedestrian and Bicycle Information Center [9], all provide guidance on planning and designing a complete street. Individual states, cities, and localities create their own guide books for formulating policies that work best for their specific area. New Jersey [10], Tennessee [11], and Boston [12] are just a few of the localities that have developed their own guidelines and policy manuals.

Figure 24.1 shows an example of complete street design [12], with four modes sharing the street: pedestrians, bicycles, motor vehicles, and transit.

Figure 24.1: Example Complete Street Design

(Source: *Boston Complete Streets Guidelines*, Boston Transportation Departments, Boston MA, 2013, Figure 13, available at www.bostoncompletestreets.org.)

Elements are needed to provide safe and effective mobility for pedestrians, bicycles, and transit on streets and at intersections. Some of the measures available for making streets safer for each user are discussed in the following sections. Choosing the best measures will vary by location (context-sensitive planning) based on elements such as land use, crash data, desired objective, inventory of current facility, and importantly, discussions with stakeholders (neighborhood residents, businesses affected, and various divisions within the local, city, state transportation agency involved, for example). An important element, particularly for the safety of pedestrians and bicycles on neighborhood streets, is managing vehicle speeds (traffic calming).

There are more countermeasures than can be enumerated in this textbook, but some of the more commonly used ones are discussed in the following sections. The guide books previously mentioned [6]-[12] as well as the Federal Highway Administration's publications on Proven Safety Countermeasures [13] and Road Diet Guide [14] provide a more complete list. The Pedestrian [7] and Bicycle [8] Safety Guides also provide estimated costs of countermeasures.

24.1.1 Pedestrians

Some of the more common recommendations for creating a safer environment for pedestrians are as follows:

- Continuous sidewalks that are separated from motor vehicle traffic by a barrier or buffer of at least 4 feet, with paved shoulders and wide enough for wheel chairs
- Good lighting so that pedestrians feel safer
- Neighborhood slow zones, which use traffic calming measures in residential areas to reduce speeds
- Leading pedestrian intervals, which initiate the pedestrian walk interval some seconds before the vehicle signal changes to green. This allows the pedestrians to get a head start into the crosswalk and thus be more visible to the turning vehicle
- Senior zones, which create an environment for slower moving pedestrians to be safer doing normal movements
- Pedestrian medians for crossing wide streets
- Daylighting, which removes parking spaces close to the corner for greater visibility
- Curb extensions (to shorten the crossing distance for the pedestrian)

24.1.2 Bicycles

Some of the more common recommendations for creating a safer environment for bicycles are as follows:

- Bike lanes (preferably separated from motor vehicles by some type of barrier or shoulder)
- Bicycle signals
- Bicycle boulevards (streets having shared bicycle and vehicle space)
- Improved road surface
- More bike parking stations
- Horizontal bike grates instead of vertical because tires can get caught in a vertical grate.

24.1.3 Transit

Common measures to create a safer and more efficient environment for transit include the following:

- Comfortable transit stops, with seating and protection from weather
- Better lighting at transit stops
- Easy and safe access to the transit stop and onto the transit vehicle
- Transit signal priority (TSP), which creates faster trips for buses; TSP allows communication between buses, the traffic management center, and signal controllers, which can either extend the green time to allow the bus through the intersection without stopping or have an early return to green to reduce delay to the bus.

24.1.4 Traffic Calming

Traffic calming measures improve the quality of service for pedestrians and bicycles. Some common measures used to reduce the speed of vehicles include the following:

- Speed bumps (to slow vehicle speed)
- Road diets, which reduce vehicle-moving lanes to add space for bike lanes, wider sidewalks, and so on
- Chicanes, which are a midblock speed control that diverts the motor vehicle path from a straight line to reduce speed, as shown in Figure 24.2

Figure 24.2: A Chicane Used in Traffic Calming

(Source: http://www.pedbikesafe.org/PEDSAFE/countermeasures_detail.cfm?CM_NUM=33)

- Speed cameras with signs notifying drivers of their existence

- Street furniture and landscaping (more clearly separates pedestrians and vehicles)
- Enforcement

Table 24.1 presents some of the benefits and disbenefits of traffic calming measures [14].

Table 24.1: Effects Reported by Practitioners

Road Diet Feature	Primary/Intended Impacts	Secondary/Unintended Impacts	
		Positive	Negative
Bike Lanes	• Increased mobility and safety for bicyclists and higher bicycle volumes. • Increased comfort levels for bicyclists due to separation from vehicles.	• Increased property values.	• Could reduce parking depending on design.
Fewer Travel Lanes	• Reallocate space for other uses.	• Pedestrian crossings are easier, less complex. • Can make finding a gap easier for cross traffic. • Allows for wider travel lanes.	• Mail trucks and transit vehicles may block traffic when stopped. • May reduce capacity. • In some jurisdictions, maintenance funding is tied to number of lane-miles, so reducing the number of lanes can have a negative influence on maintenance budgets. • Similarly, some federal funds may be reduced. • If travel lanes are widened, can encourage increased speeds
Two-Way Left-Turn Lane	• Provide dedicated left-turn lane.	• Makes efficient use of limited roadway area.	• Could be difficult for drivers to access left-turn lane if demand for left turns is too high.
Pedestrian Refuge Island	• Increased safety and mobility for pedestrians.	• Makes pedestrian crossings safer and easier. • Prevents illegal use of TWLTL to pass slower traffic or access an upstream turn lane.	• May cause issues with snow removal. • Can effectively increase congestion by preventing illegal maneuvers.
Buffers (grass, concrete medians, plastic delineators)	• Provide barriers and space between travel modes.	• Increases comfort level for bicyclists by providing separation from vehicles. • Barriers can prevent a vehicle from entering a lane reserved for other modes.	• Grass and delineator barriers will necessitate ongoing maintenance.

(Source: Knapp, et al, *A Road Diet Guide*, Federal Highway Administration, Washington D.C., November 2014, Table 2, pg 1.)

Table 24.1: Full Alternative Text

24.2 Level of Service Analysis of a Multimodal Street Segment

The *Highway Capacity Manual* (HCM) [15] provides methods for evaluating an urban street for four modes that can be found on it. A level of service (LOS) is calculated for each of the four modes: vehicles, pedestrians, bicycles, and transit. These levels can be compared to determine the quality of service for each of the modes on a facility, but they may not be combined together to get one LOS.

The calculations for each mode are not detailed in this chapter, as they are complex and lengthy. Rather, the general concepts behind determining LOS for each mode are described. The reader is referred to HCM Chapters 16, 18, 29, and 30 for a full description of the calculations for each mode.

This chapter discusses the HCM methodology as it applies to the analysis of vehicles, pedestrians, bicycles, and transit on a single segment of an urban street. A street segment consists of a link and the downstream intersection, which is usually a signalized intersection. The HCM also provides a methodology for considering a series of adjacent street segments. This type of analysis is considered a facility analysis.

24.2.1 Vehicle Methodology

The methodology for determining the LOS on a segment for motorized vehicles involves calculating the average travel time on the link and the v/c ratio at the downstream intersection for the through movement. Only through movements are considered because the main purpose of an arterial is to move vehicles along the facility. The v/c ratio at the downstream intersection is found using the signalized intersection methodology described in Chapter 22. The number of inputs needed for the vehicle analysis methodology is quite extensive and is shown in Table 24.2.

Table 24.2: Required Input

Data for a Vehicle LOS Analysis on an Urban Street Segment

Required Data and Units	Potential Data Source(s)	Suggested Default Value
	Traffic Characteristics Data	
Demand flow rate by movement group at boundary intersection (veh/h)	Field data, past counts	Must be provided
Access point flow rate by movement group (veh/h)	Field data, past counts	See discussion in text
Midsegment flow rate (veh/h)	Field data, past counts	Estimate using demand flow rate at downstream boundary intersection
	Geometric Data	
Number of lanes by movement group at boundary intersection.	Field data, arterial photo	Must be provided
Upstream intersection width (ft)	Field data, arterial photo	Must be provided
Segment approach turn bay length at boundary intersection (ft)	Field data, arterial photo	Must be provided
Number of midsegment through lanes.	Field data, arterial photo	Must be provided

24.2-2 Full Alternative Text

Number of lanes at access point – segment approach	Field data, arterial photo	(a) Number of through lanes on approach = number of midsegment through lanes (b) No right-turn lanes (c) If median present, one left-turn lane per approach; otherwise, no left-turn lanes
Number of lanes at access point – access point approach	Field data, arterial photo	One left-turn and one right-turn lane
Segment approach turn bay length at access points (ft)	Field data, arterial photo	40% of access point spacing, where spacing = $2 \times (5280)/D_a$ in feet, but not more than 300 ft nor less than 50 ft
Segment length (ft)	Field data, arterial photo	Must be provided
Restrictive median length (ft)	Field data, arterial photo	Must be provided
Portion of segment with curb (decimal)	Field data, arterial photo	1.0 (curb present on both sides of segment)
Number of access point approaches	Field data, arterial photo	See discussion in text
Proportion of segment with on-street parking (decimal)	Field data	Must be provided
	Other Data	
Analysis period duration (h)	Set by analyst	0.25 h
Speed limit (mi/h)	Field data, road inventory	Must be provided
	Performance Measure Data	
Through control delay at boundary intersection (s/veh)	HCM method output	Must be provided
Through stopped vehicles at boundary intersection (veh)	HCM method output	Must be provided
2^{nd}- and 3^{rd}- term back-of-queue for through movement at boundary intersection (veh/ln)	HCM method output	Must be provided
Capacity by movement group at boundary intersection (veh/h)	HCM method output	Must be provided
Midsegment delay (s/veh)	Field data	0.0 s/veh
Midsegment stops (stops/veh)	Field data	0.0 stops/veh

24.2-3 Full Alternative Text

(Source: Reprinted with permission from *Highway Capacity Manual, 6th Edition – A Guide for Multimodal Mobility Analysis*, Transportation Research Board, the National Academy of Sciences, Courtesy of the National Academies Press, Washington, D.C., 2016.)

The thresholds for LOS on an urban street segment are different for different base free flow speeds. Base free flow speed is the freely chosen speed when flow rate is low and thus not interfering with the driver's choice of speed when the segment is long. By using base free flow speed to vary LOS thresholds recognizes that different types of arterials are judged differently by the user. On an arterial with a base free flow speed of 25 mi/h versus a base free flow speed of 50 mi/h, the user would not expect the same travel speed.

The LOS thresholds for vehicles are shown in Table 24.3. LOS is based on average travel speed, and describes the performance of the urban street in operational terms rather than a user-perceived quality of the service. For this reason, a separate LOS score is calculated that better represents user perception of quality of service.

Table 24.3: HCM Criteria for Vehicle LOS on an Arterial Segment (Average Travel Speed, mi/h)

LOS	Base Free Flow Speed (mi/h)							v/c ratio of Downstream Boundary Intersection
	55	50	45	40	35	30	25	
A	>44	>40	>36	>32	>28	>24	>20	≤1
B	>37	>34	>30	>27	>23	>20	>17	≤1
C	>28	>25	>23	>20	>18	>15	>13	≤1
D	>22	>20	>18	>16	>14	>12	>10	≤1
E	>17	>15	>14	>12	>11	>9	>8	≤1
F	≤17	≤15	≤14	≤12	≤11	≤9	≤8	>1

(Source: Reprinted with permission from *Highway Capacity Manual, 6th Edition – A Guide for Multimodal Mobility Analysis*, Transportation Research Board, the National Academy of Sciences, Courtesy of the National Academies Press, Washington, D.C., 2016.)

Table 24.3: Full Alternative Text

Research has shown that drivers perceive the quality of their trip primarily based on the number of stops they must make [16]. Thus a separate methodology calculates an automobile traveler perception score of 1 to 6 (A to F) that is based on the probability of a driver rating the trip a certain LOS. The probabilities are calculated from regression equations with independent variables of stops/mile and the proportion of left-turn lanes or bays on the segment (at any unsignalized intersections and a downstream signalized intersection).

24.2.2 Pedestrian Methodology

The HCM urban street segment pedestrian methodology evaluates the quality of service for pedestrians at the intersection, on the link (not including the intersection), and on the segment as a whole (link plus intersection). The inputs required for a pedestrian analysis are shown in Table 24.4.

Table 24.4: HCM Inputs for a Pedestrian Analysis

Required Data and Units	Potential Data Source(s)	Suggested Default Value
	Traffic Characteristics	
Midsegment motorized vehicle flow rate* (veh/h)	Field data, past counts, forecasts	Must be provided
Midsegment pedestrian flow rate (ped/h)	Field data, past counts	Must be provided
Proportion of on-street parking occupied (decimal)	Field data	0.50 (if parking lane present)
	Geometric Design	
Downstream intersection width* (ft)	Field data, arterial photo	Must be provided
Segment length* (ft)	Field data, arterial photo	Must be provided
Number of midsegment through lanes*	Field data, arterial photo	Must be provided
Outside through lane width (ft)	Field data, arterial photo	12 ft
Bicycle lane width (ft)	Field data, arterial photo	5.0 ft (if provided)
Paved outside shoulder width (ft)	Field data, arterial photo	Must be provided
Striped parking lane width (ft)	Field data, arterial photo	8.0 ft (if provided)
Curb presence (yes/no)	Field data, arterial photo	Must be provided
Sidewalk presence (yes/no)	Field data, arterial photo	Must be provided

24.2-6 Full Alternative Text

Total walkway width (ft)	Field data, arterial photo	9.0 ft (business/office uses) 11.0 ft (residential/industrial uses)
Effective width of fixed objects (ft)	Field data	2.0 ft inside, 2.0 ft outside (business/office uses) 0.0 ft inside, 0.0 ft outside (residential/industrial uses)
Buffer width (ft)	Field data, arterial photo	0.0 ft (business/office uses) 6.0 ft (residential/industrial uses)
Spacing of objects in buffer (ft)	Field data, arterial photo	Must be provided
	Other Data	
Distance to nearest signal-controlled crossing (ft)	Field data, arterial photo	One-third the distance between signal-controlled crossings that bracket the segment
Legality of midsegment pedestrian crossing (legal/illegal)	Field data, local traffic laws	Must be provided
Proportion of sidewalk adjacent to window, building, and fence (decimal)	Field data	0.0 (non-CBD area) 0.5 building, 0.5 window (CBD)
	Performance Measures	
Motorized vehicle midsegment running speed* (mi/h)	HCM method output	Must be provided
Pedestrian delay at boundary intersection (s/ped)	HCM method output	Must be provided
Pedestrian delay at midsegment signalized crossing (s/ped)	HCM method output	20 s/ped (if present)
Pedestrian delay at uncontrolled crossing (s/ped)	HCM method output	Must be provided
Pedestrian LOS score for intersection (decimal)	HCM method output	Must be provided

24.2-7 Full Alternative Text

*Also used or calculated by the motorized vehicle methodology.

(Source: Reprinted with permission from *Highway Capacity Manual, 6th Edition – A Guide for Multimodal Mobility Analysis*, Transportation Research Board, the National Academy of Sciences, Courtesy of the National Academies Press, Washington, D.C., 2016.)

Intersection pedestrian LOS is based on a score that predicts pedestrian perceived quality of service based on pedestrian delay and vehicle demand and speed. Although not part of the LOS calculation, performance measures for circulation area at the corner and on the crosswalk are also calculated.

The link pedestrian LOS score for the segment is computed based on calculations of average pedestrian free flow walking speed, which is the average speed of pedestrians walking when not interfered by other pedestrians, average pedestrian space, and pedestrian delay at the intersection.

The link score is computed based on sidewalk width, vehicle speed and volume, effective width of outside through lane, bicycle lane, shoulder width, and buffer width. An important factor influencing link score is the separation of the pedestrians from the vehicles and how fast those vehicles are traveling on the link.

A total segment pedestrian LOS is based on the average pedestrian space (ft^2/ped), pedestrian intersection score, the pedestrian link score, and a roadway crossing difficulty factor. Table 24.5 shows the thresholds for pedestrian LOS.

Table 24.5: Pedestrian LOS Score

Pedestrian LOS Score	LOS by Avwerage Pedestrian Space (ft^2/ped)					
	>60	>40 – 60	>24 – 40	>15 – 24	>8.0 – 15*	≤8.0*
≤2.00	A	B	C	D	E	F
>2.00–2.75	B	B	C	D	E	F
>2.75–3.50	C	C	C	D	E	F
>3.50–4.25	D	D	D	D	E	F
>4.25–5.00	E	E	E	E	E	F
>5.00	F	F	F	F	F	F

*In cross-flow situations, the LOS E/F threshold is 13 ft 3 /ped. Chapter 4 (of the 2016 HCM) describes the concept of "cross flow" and situations where it should be considered.

(Source: Reprinted with permission from *A Guide for Multimodal Mobility Analysis*, Transportation Research Board, the National Academy of Sciences, Courtesy of the National Academies Press, Washington, D.C., 2016.)

Table 24.5: Full Alternative Text

24.2.3 Bicycle Methodology

The HCM bicycle performance methodology evaluates the quality of service for bicycles at the downstream intersection, on the link, and on the segment as a whole (link and downstream intersection). Table 24.6 shows the inputs needed for a bicycle LOS analysis.

Table 24.6: Inputs Needed for a Bicycle Analysis

Required Data and Units	Potential Data Source(s)	Suggested Default Value
Traffic Characteristics		
Midsegment motorized vehicle flow rate[1] (veh/h)	Field data, past counts, forecasts	Must be provided
Heavy vehicle percentage (%)	Field data, past counts	Must be provided
Proportion of on-street parking occupied (decimal)	Field data	0.50 (if parking lane present)
Geometric Design		
Segment length[1] (ft)	Field data, aerial photo	Must be provided
Number of midsegment through lanes[1]	Field data, aerial photo	Must be provided
Outside through lane width[2] (ft)	Field data, aerial photo	12 ft

24.2-10 Full Alternative Text

Bicycle lane width[3] (ft)	Field data, aerial photo	5.0 ft (if provided)
Paved outside shoulder width[3] (ft)	Field data, aerial photo	Must be provided
Striped parking lane width (ft)	Field data, aerial photo	Must be provided
Median type (divided/undivided)	Field data, aerial photo	Must be provided
Curb presence (yes/no)	Field data	Must be provided
Number of access point approaches	Field data, aerial photo	See discussion in text
Other Data		
Pavement condition[4] (FHWA 5-point scale)	Field data, pavement condition inventory	3.5 (good)
Performance Measures		
Motorized vehicle midsegment running speed[1] (mi/h)	HCM method output	Must be provided
Bicycle delay at boundary intersection (s/bicycle)	HCM method output	Must be provided
Bicycle LOS score at boundary intersection (decimal)	HCM method output	Must be provided

24.2-11 Full Alternative Text

[1] Also used or calculated by the motorized vehicle methodology.

[2] High sensitivity (±2 LOS letters) of LOS to the choice of default value.

[3] Moderate sensitivity (±1 LOS letter) of LOS to the choice of default value.

[4] Sensitivity reflects pavement conditions 2-5. Very poor pavement (i.e., 1) typically results in LOS F, regardless of other input values.

(Source: Reprinted with permission from *Highway Capacity Manual, 6th Edition – A Guide for Multimodal Mobility Analysis*, Transportation Research Board, the National Academy of Sciences, Courtesy of the National Academies Press, Washington, D.C., 2016.)

The intersection LOS score for bicycles is computed based on bicycle delay at the intersection, capacity of the bicycle lane, buffer from traffic of the bike lane, and whether there is a bike lane or not.

Some of the influential factors in determining the link bicycle score are bicycle travel speed, vehicle demand flow rate and running speed, percent heavy vehicles, pavement condition, and the effective width of the outside through lane. Vehicle demand flow, vehicle running speed, percent heavy vehicles, and the buffer distance from the vehicles influence how safe the bicyclists feel. The buffer distance is represented by the effective width of the outside, which includes width of a bicycle lane, shoulder, and parking lane.

The segment bicycle score is calculated with a regression equation using the independent variables score of the intersection, the link, a factor for conflicts at unsignalized intersections, and the number of access points on the segment. Table 24.7 shows the LOS thresholds for a bicycle analysis.

Table 24.7: Level of Service

for a Bicycle Analysis

LOS	LOS Score
A	≤2.00
B	>2.00–2.75
C	>2.75–3.50
D	>3.50–4.25
E	>4.25–5.00
F	>5.00

(Source: Reprinted with permission from *Highway Capacity Manual, 6th Edition – A Guide for Multimodal Mobility Analysis*, Transportation Research Board, the National Academy of Sciences, Courtesy of the National Academies Press, Washington, D.C., 2016.)

Table 24.7: Full Alternative Text

24.2.4 Transit Methodology

The HCM transit performance methodology evaluates the quality of service of public transit operating on the street segment. A score is computed based on the transit vehicle running speed on the link, the delay at the intersection, the overall segment travel speed, a score for the perceived waiting time and perceived travel time (called the transit wait-ride score), and the pedestrian link LOS. Table 24.8 shows the inputs required for a transit analysis.

Table 24.8: HCM Inputs for a Transit Analysis

Required Data and Units	Potential Source(s)	Suggested Default Value
	Traffic Characteristics	
Dwell time (s)	Field data, AVL data[1]	60 s (downtown stop, transit center, major on-line transfer point, major park-and-ride) 30 s (major outlying stop) 15 s (typical outlying stop)
Excess wait time (min)	Field data, AVL data[1]	See discussion in text
Passenger trip length (mi)	National Transit Database	3.7 mi
Time frequency (veh/h)	Transit schedules	Must be provided
Passenger load factor (pass/seat)	Field data, APC data[2]	0.80 pass/seat
	Geometric Data	
Segment length[3] (ft)	Field data, aerial photo	Must be provided
	Other Data	
CBD of 5 million plus metro area (yes/no)	Census data	Must be provided
Traffic signal green-to-cycle length ratio (decimal)	Field data or HCM method output	Must be provided (if present)
Traffic signal cycle length (s)	Field data or HCM method output	Must be provided (if present)
Proportion of transit stops with shelters (decimal)	Field data or transit facility inventory	Must be provided
Proportion of transit stops with benches (decimal)	Field data or transit facility inventory	Must be provided
	Performance Measures	
Motorized vehicle running time[3] (s)	HCM method output	Must be provided
Pedestrian LOS score for link (decimal)	HCM method output	Must be provided
Reentry delay (s/veh)	HCM method output	Must be provided
Roundabout volume-to-capacity ratio (decimal)	HCM method output	Must be provided (if present)

[1] AVL = automatic vehicle location.

[2] APC = automatic passenger counter.

[3] Also used in the motorized vehicle methodology.

(Source: Reprinted with permission from *Highway Capacity Manual, 6th Edition – A Guide for Multimodal Mobility Analysis*, Transportation Research Board, the National Academy of Sciences, Courtesy of the National Academies Press, Washington, D.C., 2016.)

Table 24.8: Full Alternative Text

Transit vehicle running time includes the time it takes to travel the segment without stopping as well as the delay time caused by stopping to pick up and discharge passengers. The delay time caused by stopping includes acceleration/deceleration, serving the passengers, and merging back into traffic.

The transit wait-ride score tries to capture how the passenger perceives the waiting time and the trip travel time. Some of the factors that influence this perception are benches and shelters at the stops, passengers per seat on the vehicle, headway between transit vehicles, and extra wait time due to late arrivals.

Table 24.9 shows the LOS thresholds for transit analysis. Note that these thresholds are the same as for a bicycle analysis.

Table 24.9: Level of Service for a Transit Analysis

LOS	LOS Score
A	≤2.00
B	>2.00–2.75
C	>2.75–3.50
D	>3.50–4.25
E	>4.25–5.00
F	>5.00

(Source: Reprinted with permission from *Highway Capacity Manual, 6th Edition – A Guide for Multimodal Mobility Analysis*, Transportation Research Board, the National Academy of Sciences, Courtesy of the National Academies Press, Washington, D.C., 2016.)

Table 24.9: Full Alternative Text

24.2.5 Summary

While the 2016 HCM allows for the analysis of LOS on various user groups on an urban street, it does not comprise a unified multimodal LOS methodology. Indeed, it might not be appropriate for the HCM to do so. The relative importance of the various modes in any given situation depends upon many things, including the relative mix of modal users, issues of accessibility and mobility, and local priorities. These would not be expected to be the same throughout the nation or the world.

Current LOS methodologies provide an operations- based LOS for vehicles, while other modes (pedestrians, bicycles, transit) are based upon estimated user-perception scales. A user-perception scale is provided for vehicles, but may not be used to determine LOS.

24.3 Facility Level of Service Analysis

Facility LOS combines the results of the individual segments by mode to compute a facility LOS for each mode. Results calculated for each segment by mode are combined by calculating weighted averages for each of the measures of effectiveness calculated for the segment. LOS thresholds for the facility are the same as for the segment, defined in Tables 24.3, 24.5, 24.7, and 24.9 for vehicles, pedestrians, bicycles, and transit, respectively.

24.4 Closing Comments

Complete street design and philosophy have taken effect in over 1,100 localities in the United States and are growing [17]. This is because street function has moved beyond simply providing vehicle access and throughput: Modes other than automobiles are now considered as important when considering the purpose and quality of service of a facility. It is important to look at the LOS experienced by all users. Comparing levels of service for each of the modes allows the analyst to determine how changes to one mode may affect other modes.

References

- 1. Smart Growth America, https://smartgrowthamerica.org/program/national-complete-streets-coalition/
- 2. Vision Zero Network, http://visionzeronetwork.org/resources/vision-zero-cities/
- 3. Complete Streets Fundamentals, https://smartgrowthamerica.org/tag/complete-streets-fundamentals/
- 4. Benefits of Complete Streets, University of Delaware Institute of Public Administration, http://www.ipa.udel.edu/healthyDEtoolkit/completestreets/benefits.html
- 5. National League of Cities, Sustainable Cities Institute, http://www.sustainablecitiesinstitute.org/topics/transportation/complete-streets
- 6. Urban Street Design Guide, National Association of City Transportation Officials, https://nacto.org/publication/urban-street-design-guide/street-design-elements/
- 7. Zegeer, C., et al., PEDSAFE, Pedestrian Safety Guide and Countermeasure Selection System, 2013, http://www.pedbikesafe.org/pedsafe/guide_background.cfm
- 8. Zegeer, C., et al., BikeSafe, Bicycle Safety Guide and Countermeasure Selection System, http://www.pedbikesafe.org/bikesafe/
- 9. Pedestrian and Bicycle Information Center, www.pedbikeinfo.org/planning/facilities.cfm
- 10. Making Complete Streets a Reality, New Jersey DOT, http://www.state.nj.us/transportation/eng/completestreets/pdf/cspolicydevelopmentguide2012.pdf
- 11. Complete Streets Design Guidelines, Tennessee DOT, http://

www.knoxtrans.org/plans/complete_streets/guidelines.pdf

- 12. Boston Complete Streets Guidelines, 2013 http://bostoncompletestreets.org/pdf/2013/BCS_Guidelines_LowRes.pdf

- 13. Office of Safety Proven Safety Countermeasures, USDOT Federal Highway Administration, https://safety.fhwa.dot.gov/provencountermeasures/

- 14. FHWA Road Diet Informational Guide, 2014, https://safety.fhwa.dot.gov/road_diets/info_guide/

- 15. *Highway Capacity Manual, 6th Edition: A Guide for Multimodal Mobility Analysis*, Transportation Research Board, Washington, D.C., 2016.

- 16. Vandehey, M. et al, *Measuring and Predicting the Performance of Automobiles Traffic on Urban Streets*, Final Report *NCHRP* Project 3-79, Kittelson and Associates, Portland OR.

- 17. Complete Streets Policies Nationwide, Smart Growth America, https://smartgrowthamerica.org/program/national-complete-streets-coalition/policy-development/policy-atlas/

Problems

1. 24-1. Discuss how urban street design has changed and the benefits of these changes.

2. 24-2. What are the factors that affect level of service for pedestrians and why are these factors important?

3. 24-3. What are the factors that affect level of service for bicycles and why are these factors important?

4. 24-4. What are the factors that affect level of service for transit vehicles and why are these factors important?

5. 24-5. How is level of service defined for vehicles? Why is there a separate measure for quality of service for vehicles and how is it different than the measure for LOS?

.

Chapter 25 Unsignalized Intersections and Roundabouts

Unsignalized intersections is a term that covers four fundamental types of intersections, each of which is not controlled in part or wholly by a traffic control signal. These four general classes of intersections are as follows:

- Uncontrolled intersections
- Two-way STOP-controlled (TWSC) intersections (including some YIELD-controlled intersections)
- Multiway STOP-controlled intersections
- Roundabouts

Completely "uncontrolled" intersections do not have any control devices that designate the right-of-way to any identified movements. Right-of-way follows the general right-of-way laws in effect, which are uniform among the 50 states: The driver on the left must give way to any vehicle approaching on the right that is close enough to impose an impending hazard. Also, through vehicles have the right-of-way over conflicting turning vehicles, again, when they are close enough to impose an impending hazard. Through these uniform driving laws, it is the responsibility of the driver on the left to avoid dangerous conflicts with vehicles on the right, and turning drivers must avoid dangerous conflicts with through vehicles. "Uncontrolled" intersections may have warning or guide signs in place, as these do not affect right-of-way regulations.

Fully uncontrolled intersections are not treated in any detail in this chapter. They generally exist in low-volume situations, and do not normally involve high delays or significant accident risks. The *Manual of Uniform Traffic Control Devices* (MUTCD) [1] contains warrants that assist traffic engineers in selecting an appropriate level of intersections control. These, as well as sight distance considerations, are covered in detail in Chapter 15.

TWSC intersections include a number of possible configurations, including STOP signs on both approaches of a two-way minor street, STOP signs on the stem of T-intersections, and STOP signs on one approach on a minor one-way street. YIELD signs can also be placed in these configurations and would be included in this general category.

Multiway or all-way STOP-controlled (AWSC) intersections have STOP signs on all approaches to the intersection. YIELD signs are not permitted in such configurations.

Roundabouts have been used in Europe for many decades. The popularity and use of roundabouts in the United States have steadily risen over the last 20 years, and they are often thought of as an alternative to signalization of an intersection. A roundabout is fundamentally different from a traffic circle, in that on a roundabout, circulating vehicles have the right-of-way over entering vehicles, even though they are on the left. YIELD signs are used on all approaches to emphasize this hierarchy. In older traffic circles, normal Right-of-way (ROW) rules apply, meaning that circulating traffic would give way to entering vehicles, which are always on the right. Except in very low-volume applications, traffic circles are not normally used, although they continue to exist.

This chapter provides an overview of TWSC intersections, AWSC intersections, and roundabouts. Design aspects and capacities of these intersections are the primary focus.

Part I Two Way Stop-Controlled Intersections

TWSC intersections are a common form of unsignalized intersections. In earlier times, such intersections were not thought to be a major issue, either in design or in capacity analysis. The 1965 *Highway Capacity Manual* (HCM) [2] stated that:

> In a sense, then, capacities and the larger service volumes of unsignalized intersections may be considered of academic interest only: in practice, by the time such levels are reached at important intersections, signals ordinarily will be installed. [Ref. 2, pg 155]

The 1965 HCM provided a simple analysis procedure for TWSC intersections by assuming they were signalized with simple two-phase operation.

The first appearance of a specific methodology for analysis of TWSC intersections appeared in *Interim Materials on Highway Capacity* [3], which was published in 1980. The methodology presented was based upon a European publication [4], which was itself based upon a German publication [5]. The methodology has been continuously updated with each successive edition of the HCM, but the fundamental structure of the model remains unchanged.

25.1 TWSC Intersection Operation: A Fundamental Modeling Approach

Drivers seeking to cross a major street from a controlled leg of a TWSC intersection are basically performing a gap acceptance maneuver. The driver of the stopped vehicle must observe gaps in the major street traffic stream and select a gap through which to safely complete the desired maneuver. There are three basic variables that control such maneuvers:

- Distribution of Gaps in the Conflicting Traffic Stream: The actual distribution of gaps in the conflicting traffic stream through which the controlled vehicle must pass.
- Critical Gap or Critical Headway, t_c: The average size gap that a controlled driver will choose to pass through. While much of the literature refers to this value as the critical gap, the HCM now refers to it as the critical headway.
- Follow-Up Time, t_f: The average time for a second driver to use a gap after the first driver departs, assuming that the gap is large enough to accommodate two or more vehicles.

The prediction of these three characteristics is central to the analytic methodologies applied to TWSC intersections.

The issues are complicated by the fact that there are several different movements at a TWSC intersection that are seeking gaps, and the traffic stream through which they must navigate differs for each movement. Major street left turns and U-turns and minor street right-turn, through, and left-turn maneuvers all seek gaps. Some of these movements conflict with each other. When a gap appears, and more than one vehicle is seeking to pass through the gap, there is a strict priority order in which gaps are used:

1. Left turns from the major street

2. Right turns from the minor street
3. U-turns from the major street
4. Through movements from the minor street
5. Left turns from the minor street

Where major street U-turns are not separately identified, they are assumed to be part of the major street left-turn movement.

If, for example, a left-turning vehicle from the major street and a through vehicle from the minor street were waiting, and a gap appeared, the major street left-turner would use it first. The through vehicle on the minor street would be forced to wait for another gap—assuming that the initial gap was not large enough to accommodate both sequentially. The analysis structure treats this priority system as completely rigid—no exceptions permitted. Although this assumption is good in most situations, it should be noted that aggressive drivers sometimes do not adhere to it, particularly when some congestion is present and delays are long.

The analysis methodology for TWSC intersections has several basic steps, each executed in strict order of the priority rank of the controlled movement:

1. Express Demand in Flow Rates for Each Movement: Like any analysis, the basic data on demand volumes must be provided. If only full peak-hour volumes are available, they are converted to flow rates as follows:

   vi=ViPHF [25-1]

   where:

   v i =demand flow rate for Movement i,veh/h, v i =demand volume for Movement i, veh/h, and PHF =peak-hour factor.

   A single PHF for the entire intersection is used. This conversion, however, assumes that all movements at the intersection peak at the same time, which is generally not true. It is preferable to collect volumes for all movements in common 15 min increments, so that the period(s) with the worst demand conditions can be identified directly. When this is done, no adjustment for PHF is used.

It is noted that there is no conversion for heavy vehicles, and that the entire analysis methodology is conducted in mixed vehicles per hour.

2. Determine Conflicting Flow Rates for Each Movement: Each subject movement seeking gaps does so through a different conflicting traffic stream:

    - Left turns and U-turns from the major street seek gaps through the opposing major street through and right-turn movements

    - Right turns from the minor street seek gaps in the right-most lane of the major street

    - Through movements from the minor street seek gaps through all major street movements

    - Left turns from the minor street seek gaps through all major street movements and the opposing minor street through movement

3. Determine Critical Gaps (Headways) and Follow-Up Times: Different subject movements will require different gap sizes through which to make their desired maneuvers, depending upon the complexity of the maneuver.

4. Determine Potential Capacities: Potential capacity assumes that each subject movement has (a) an exclusive lane from which to operate and (b) full use of all available gaps—that is, no higher priority movements present.

5. Determine Movement Capacities: Potential capacities are modified to account for the presence of higher priority movements seeking gaps. This process is called “impedance.”

6. Determine Shared-Lane Capacities: Movement capacities are further modified to account for sharing of lanes between controlled movements.

7. Determine Delay and Level of Service: The output of the analysis is a determination of average control delay for each movement, lane, and approach in the intersection. Delay is then used to determine level of service (LOS). Full intersection delay may be computed, but no intersection-level LOS is assigned, as many vehicles will experience “0” delay, which would make the average for the intersection as a whole relatively meaningless.

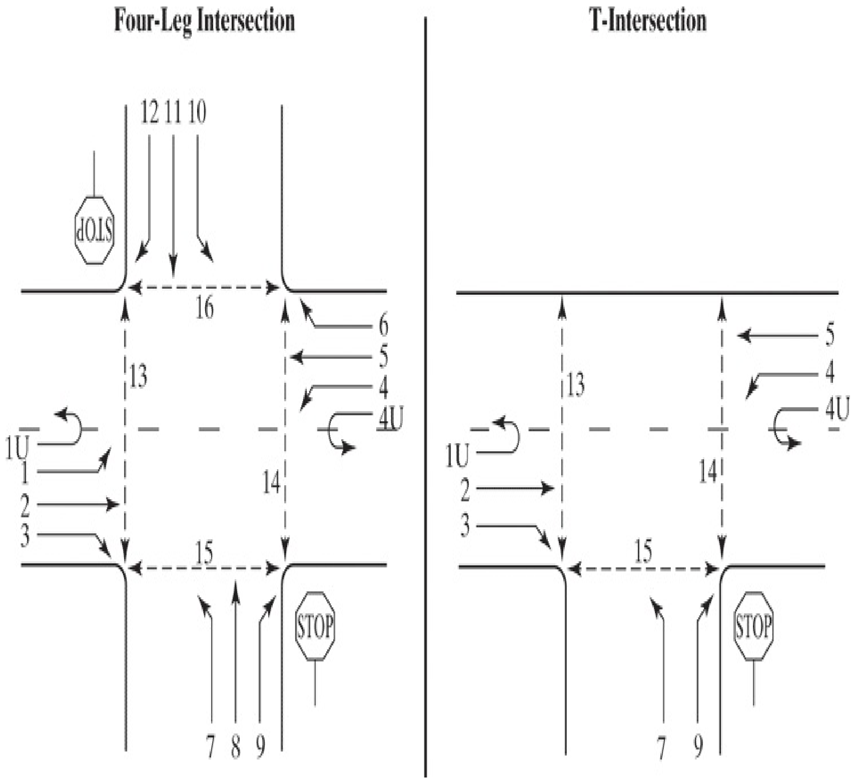
Subsequent sections will discuss and illustrate each of these steps in greater detail.

25.2 Computational Steps in TWSC Intersection Analysis

1. Step 1: Express Demand as Flow Rates during a Peak 15-Minute Analysis Period

   Demand volumes should be established using field data for the intersection(s) of interest. Because the analysis methodology uses a rigid priority of movements using gaps, a standard movement numbering scheme is used, as illustrated in Figure 25.1.

Figure 25.1: Movement Numbering Scheme for TWSC Intersections



(Source: Reprinted with permission from *Highway Capacity Manual, 6th Edition – A Guide for Multimodal Mobility Analysis*, Transportation Research Board, the National Academy of Sciences, Courtesy of the National Academies Press, Washington, D.C., 2016.)

Figure 25.1: Full Alternative Text

Note that major street movements are labeled 1 through 6, while minor street movements are labeled 7 through 12. Pedestrian movements are labeled 13 through 16. The mix of movements is, of course, much simpler at a T-intersection than at a typical four-leg intersection. Also, the numbering scheme has nothing to do with directional orientation. Strict adherence to the numbering protocol will simplify subsequent computations.

As noted previously, if demand volumes are based upon hourly data, they should be converted to flow rates using Equation 25-1. If they have been entered directly as flow rates, then no conversion is necessary.

2. Step 2: Determine Conflicting Flow Rates for Each Movement

   Conflicting flow rates must be established for each subject movement seeking gaps. The conflicting flow rate, however, depends upon the movement seeking gaps.

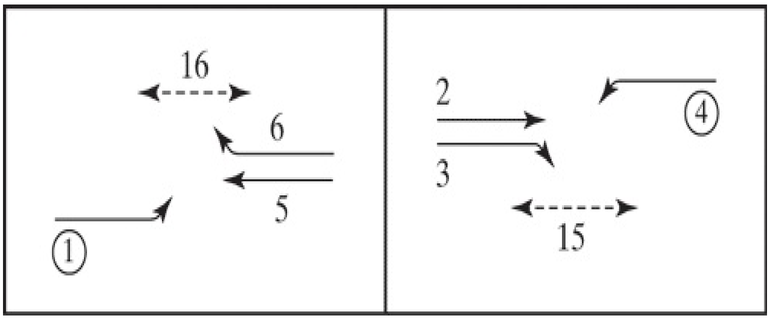
   Two movements—the through and left-turn movements from the minor street—may face one of two scenarios. If there is no median on the major street large enough to store at least one vehicle, these two movements proceed in *one stage*, that is, they cross all conflicting flows in a single movement. However, if a median exists that can safely store one or more vehicles, minor street through and left-turn movements can execute their desired maneuver in *two stages*. In this case, the driver crosses the roadway from the left, and stops or pauses again in the median to select a gap in traffic from the right. The estimation of conflicting flow rates is slightly different for *one-stage* and *two-stage* maneuvers for these movements.

   The estimation of conflicting flow rates proceeds in the order of movement priorities defined earlier. Priority 1 movements face no conflicting flows, and therefore have no conflicting flow rate. Priority 2, 3, and 4 movements face increasingly difficult conflicts.

   Figure 25.2 illustrates the conflicting movements faced by each subject movement. The illustrations are used to develop the estimating equations for conflicting flow rates that are shown in Table 25.1.

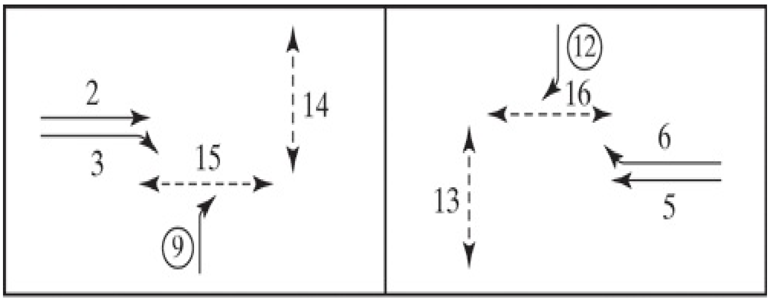
Figure 25.2: Conflicting Flows at TWSC Intersections

Illustrated



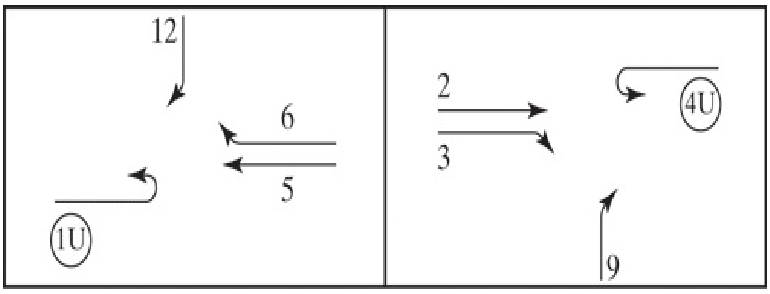
(a) Conflicting Flows for Major Street LTs (Mvts 1,4)

25.3-1 Full Alternative Text



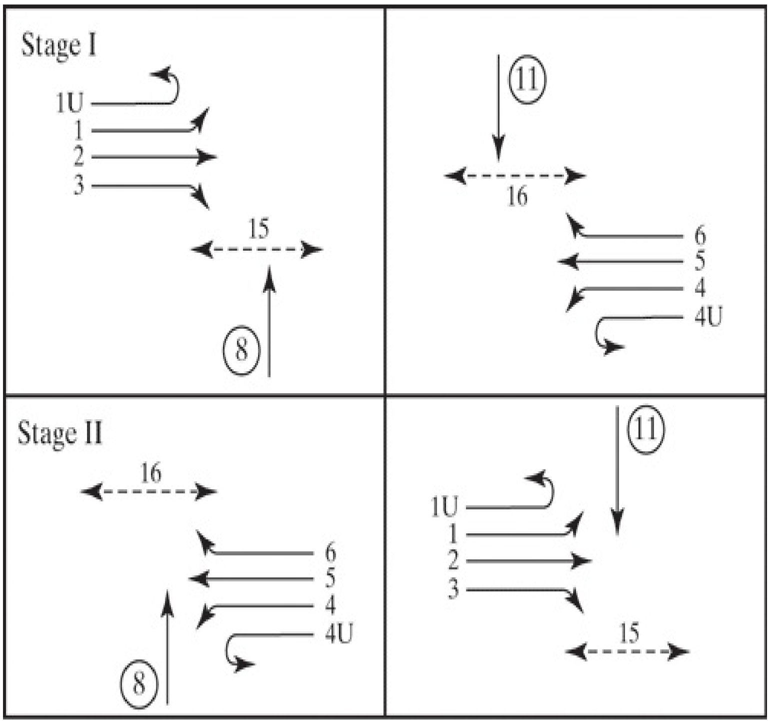
(b) Conflicting Flows for Minor Street RTs (Mvts 9,12)

25.3-1 Full Alternative Text



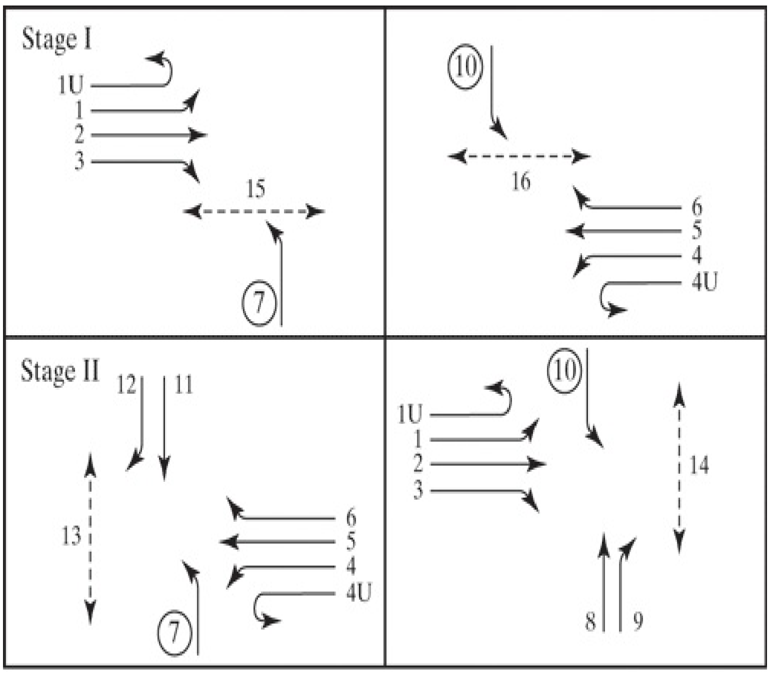
(c) Conflicting Flows for Major Street U-Turns (Mvts 1U.4U)

25.3-1 Full Alternative Text



(d) Conflicting Flows for Minor Street THs (Mvts 8,11)

[25.3-1 Full Alternative Text]



(e) Conflicting Flows for Minor Street LTs (Mvts 7,10)

[25.3-1 Full Alternative Text]

(Source: Reprinted with permission from *Highway Capacity Manual, 6th Edition – A Guide for Multimodal Mobility Analysis*, Transportation Research Board, the National Academy of Sciences, Courtesy of the National Academies Press, Washington, D.C., 2016.)

Table 25.1: Conflicting Volume Equations for TWSC Intersections

Mvt Nos.	Movements	Equations
1,4	Major St LTs	$v_{c1} = v_5 + v_6 + v_{16}$ $v_{c4} = v_2 + v_3 + v_{15}$
9,12	Minor St RTs (Onto 2-Lane Major Street)	$v_{c9} = v_2 + 0.5v_3 + v_{14} + v_{15}$ $v_{c12} = v_5 + 0.5v_6 + v_{13} + v_{16}$
9,12	Minor St RTs (Onto 4- and 6-Lane Major Streets)	$v_{c9} = 0.5v_2 + 0.5v_3 + v_{14} + v_{15}$ $v_{c12} = 0.5v_5 + 0.5v_6 + v_{13} + v_{16}$
1U,4U	Major St U-Turns (2-Lane Major Street)	No data exists for this case; treat U-turns as part of Major Street LTs
1U,4U	Major St U-Turns (4-Lane Major Street)	$v_{c1U} = v_5 + v_6$ $v_{c4U} = v_2 + v_3$
1U,4U	Major St U-Turns (6-Lane Major Street)	$v_{c1U} = 0.73v_5 + 0.73v_6$ $v_{c4U} = 0.73v_2 + 0.73v_3$
8,11	Minor St Through*	*Stage I* $v_{c8I} = 2(v_1 + v_{1U}) + v_2 + 0.5v_3 + v_{15}$ $v_{c12I} = 2(v_4 + v_{4U}) + v_5 + 0.5v_6 + v_{16}$ *Stage II* $v_{c8II} = 2(v_4 + v_{4U}) + v_5 + v_6 + v_{16}$ $v_{c12II} = 2(v_1 + v_{1U}) + v_2 + v_3 + v_{15}$
7,10	Minor Street LTs* (2-Lane Major Street)	*Stage I* $v_{c7I} = 2v_1 + v_2 + 0.5v_3 + v_{15}$ $v_{c10I} = 2v_4 + v_5 + 0.5v_6 + v_{16}$ *Stage II* $v_{c7II} = 2v_4 + v_5 + 0.5v_6 + 0.5v_{12} + 0.5v_{11} + v_{13}$ $v_{c10II} = 2v_1 + v_2 + 0.5v_3 + 0.5v_9 + 0.5v_8 + v_{14}$
7,10	Minor Street LTs* (4-Lane Major Street)	*Stage I* $v_{c7I} = 2(v_1 + v_{1U}) + v_2 + 0.5v_3 + v_{15}$ $v_{c10I} = 2(v_4 + v_{4U}) + v_5 + 0.5v_6 + v_{16}$ *Stage II* $v_{c7II} = 2(v_4 + v_{4U}) + 0.5v_5 + 0.5v_{11} + v_{13}$ $v_{c10II} = 2(v_1 + v_{1U}) + 0.5v_2 + 0.5v_8 + v_{14}$
7,10	Minor Street LTs* (6-Lane Major Street)	*Stage I* $v_{c7I} = 2(v_1 + v_{1U}) + v_2 + 0.5v_3 + v_{15}$ $v_{c10I} = 2(v_4 + v_{4U}) + v_5 + 0.5v_6 + v_{16}$ *Stage II* $v_{c7II} = 2(v_4 + v_{4U}) + 0.4v_5 + 0.5v_{11} + v_{13}$ $v_{c10II} = 2(v_1 + v_{1U}) + 0.4v_2 + 0.5v_8 + v_{14}$

* Where only one stage exists, add the conflicts for Stages I and II to find the conflicting flow rate.

Table 25.1: Full Alternative Text

It should be noted that Figure 25.2 (d) and (e) show two-stage gap acceptance equations. In cases where only one stage exists, the total conflicting flow rate is the sum of that shown for Stage I and Stage II.

There is a great deal of detailed information held within these equations, based upon insights gained from field studies and validations over the years. For example, approaching right turns from the major street do not technically conflict with other movements. They are, however, included in some equations, as such vehicles may not use directional signals, and minor street drivers may think that they are going to impose a conflict. However, where there are designated right-turn lanes on the major street, these vehicles may be dropped from the conflicting flow rate computations of Table 25.1.

The right turn from the minor street technically only conflicts with the through flow in the approaching right-hand lane. Thus, for four-lane and six-lane major streets, these flows are somewhat discounted in the computation of conflicting flow rate.

Lastly, when a minor street right-turn movement has an exclusive lane physically separated by a triangular island, and is separately controlled with a STOP or YIELD sign, it may be eliminated from conflicting flow rate computations.

3. Step 3: Determine Critical Gaps (Headways) and Follow-Up Times

   The critical gap (t_{ci}) and follow-up time (t_{fi}) must be determined for each movement seeking gaps in a conflicting traffic stream, that is, major street left turns and U-turns, minor street right turns, minor street through movements, and minor street left turns. Equations are used to estimate these values:

$t_{ci}=t_{cbase}+f_{cHV}P_{HV}+f_{cG}G-f_{3LT}$ [25-2]
$t_{fi}=t_{fbase}+f_{fHV}P_{HV}$ [25-3]

where:

t_{ci} =crictical gap for Movement i, s, t_{fi} =follow-up time for Movement i, s, t_{cbase} =base critical gap, s (Table 25.2), t_{fbase} =base follow-up time, s (Table 25.2), f_{cHV} =adjustment factor on critical gap for heavy vehicles (Tab
f_{fHV} =adjustment factor on follow-up time for heavy vehicles (Table 25.3), f_{cG} =adjustment factor on critical gap for grades (Table 25.3), f_{3LT} =adjustment factor for intersection geometry (Table 25.3), P_{HV} =proportion of heavy vehicles expressed as a decimal, anc
G = grade, expressed as a percentage.

Table 25.2: Base Values of Critical Gap and Follow-Up Time for TWSC Intersections

Vehicle Movement	Base Critical Gap, t_{ci} (s)			Base Follow-Up Time, t_{fi} (s)		
	2-Lane Major Streets	4-Lane Major Streets	6-Lane Major Streets	2-Lane Major Streets	4-Lane Major Streets	6-Lane Major Streets
LT from Major Street	4.1	4.1	5.3	2.2	2.2	3.1
U-Turn from Major Street	NA	6.4 (wide)* 6.9 (narrow)*	5.6	NA	2.5 (wide)* 3.1 (narrow)*	2.3
RT from Minor Street	6.2	6.9	7.1	3.3	3.3	3.9
Minor Street TH Traffic	1 Stage: 6.5** 2 Stages: I: 5.5** II 5.5**			4.0	4.0	4.0
LT from Minor Street	1 Stage: 7.1 2 Stages: I-6.1 II-6.1	1 Stage: 7.5 2 Stages: I-6.5 II-6.5	1 Stage: 6.4 2 Stages: I-7.3 II-6.7	3.5	3.5	3.8

NA=not available; treat U-turns as major street LTs.

* Narrow U-turns have a median nose width<21 feet; wide U-turns have a median nose width≥21 feet.

** Values are rough estimate only for 6-lane major streets; use with caution.

(Source: Reprinted with permission from *Highway Capacity Manual, 6th Edition – A Guide for Multimodal Mobility Analysis*, Transportation Research Board, the National Academy of Sciences, Courtesy of the National

Academies Press, Washington, D.C., 2016.)

Table 25.2: Full Alternative Text

Table 25.3: Adjustment Factors for Critical Gap and Follow-Up Time for TWSC Intersections

Adjustment Factor	Value(s)
f_{cHV}	1.0 for major streets with one lane in each direction 2.0 for major streets with two or three lanes in each direction
f_{fHV}	0.9 for major streets with one lane in each direction 1.0 for major streets with two or three lanes in each direction
f_{cG}	0.1 for Movements 9 and 12 0.2 for Movements 7,8,10, and 11
f_{3LT}	0.7 for minor street LTs at 3-leg intersections; 0.0 otherwise

Table 25.3: Full Alternative Text

4. Step 4: Compute Potential Capacities

    Potential capacity assumes that each subject movement operates out of an exclusive lane or lanes, and that there are no higher-priority movements present to utilize gaps. Thus, *all* available gaps of sufficient length are assumed to be used by the subject movement. Subsequent computational steps will adjust for the impedance of higher-priority movements and shared lanes where they exist.

The equation for potential capacity is based upon theoretical gap acceptance theory, and is based upon conflicting flow rates (estimated in Step 2) and critical gaps and follow-up times (estimated in Step 3):

cpi=vci [e−vci tci/36001−e−vci tfi/3600] [25-4]

where:

c pi =potential capacity of Movement i, veh/h, v fi =conflicting flow rate for Movement i, veh/h, t ci =critical gap for Movement i, s , and t fi =follow-up time for Movement i, s.

This equation assumes that all conflicting flows are randomly distributed, that is, there is no platooning in conflicting flows. This is often not true. A methodology was initially introduced in the 1985 HCM [6] to adjust for platooning. It was general, and a very rough estimate at best. In subsequent manuals, the methodology was made increasingly complex. In the 2016 HCM [7], it is virtually impossible to discern how the adjustment is made. Various chapters refer to other chapters, which subsequently refer the reader someplace else. The bottom line is that the current methodology is basically undocumented in the HCM. Its use is incorporated into computational software, but is a virtual black box.

In 2015, a paper by M. Kyte et al. [8] demonstrated that applying the current adjustment for platoon flow did not meaningfully or logically change the results, and questioned whether it should be included at all. Since the adjustment is not well documented and is of questionable value, it is not included here.

5. Step 5: Determine Movement Capacities

Movement capacities are the first modification of potential capacities: movement capacities take into account the *impedance* created by the presence of higher-priority movements, which will consume some portion of the available gaps, leaving them unavailable to lower-priority movements.

Each vehicular movement (Ranks 2 through 4) is potentially impeded by higher-priority movements. They can be impeded by higher-priority vehicular movements, as well as by pedestrian movements. The 2016 HCM takes the position that pedestrian impedances are generally quite small, and may be ignored in most cases. It still, however, provides the opportunity to account for pedestrian impedance when the analyst deems it necessary. This text will present the methodology for treating both vehicular and pedestrian impedances, with the understanding that the latter are now optional.

Table 25.4 summarizes the potential impedances that may exist for each subject movement at a TWSC intersection.

Table 25.4: Potential Impedances at TWSC Intersections

Subject Vehicular Flow	Impeding Vehicular Flows	Impeding Pedestrian Flows
1 (Major St LT)	None	16
4 (Major St LT)	None	15
1U (Major St U-Turn)	12	None
4U (Major Street U-Turn)	9	None
7 (Minor St LT)	1,4,11,12	15,13
8 (Minor St TH)	1,4	15,16
9 (Minor St RT)	None	15,14
10 (Minor St LT)	1,4,8,9	16,14
11 (Minor St TH)	1,4	15,16
12 (Minor St RT)	None	16,13

Table 25.4: Full Alternative Text

Obviously, the issue of impedance is quite complex. In the case of minor street left-turn movements, there can be up to four impeding vehicular flows, and two impeding pedestrian flows. An *impedance factor* (P_y) is found for *each* impeding flow, y. The potential capacity of each movement is then modified to a movement capacity through the use of these factors.

Theoretically, the determination of impedance factors is relatively straightforward. For an impeding vehicular movement:

$$P_y = 1 - \frac{v_y}{c_{my}} \quad [25\text{-}5]$$

where:

P_y = impedance factor for impeding Movement y, v_y = dema

For example, if an impeding movement has a demand value of 50 veh/h and a movement capacity of 150 veh/h, then it would utilize 50/150=0.333 of the available gaps. That would leave 1−0.333, or 0.667 of the gaps still available to lower-priority movements.

A slight problem exists: We are using impedance factors to estimate movement capacities, but we need movement capacities to compute the impedance factors. Because of this, computations for movement capacity *must* proceed from higher-priority movements to lower-priority movements.

For pedestrian impedances:

$P_j=1-v_j\ (wS_p)3600$ [25-6]

where:

P_j =impedance factor for impeding pedestrian flow j, v_j =impeding pedestrian flow rate for impeding pedestrian flo
w =lane width impeded vehicles are entering, ft, and S_p =walking speed, ft/s.

It should be noted that this equation also assumes that pedestrian flows are more or less random. Pedestrian flows may be in platoons. The 2016 HCM provides a methodology to adjust for this, using v_y as the number of pedestrian “groups” or platoons crossing per hour.

The final adjustment factor applied to each potential capacity is computed as:

$f_i=\prod\prod_{y=0-4}^{j=0-2}P_yP_j$ [25-7]

where f_i is the impedance adjustment factor for subject movement i, and all other variables as previously defined. Then:

$c_{mi}=c_{pi}f_i$ [25-8]

where all terms are as previously defined.

Depending upon the type of subject movement, the number of impeding vehicular flows may vary between 0 and 4, and the number of impeding pedestrian flows between 0 and 2.

As noted, the estimation of movement capacities must proceed from the higher-priority movements to the lower. The order followed is summarized in the sections that follow, along with any special adjustment that may be necessary. Note that Rank 1 movements, through and right-turn movements from the major street, are not impeded by any vehicular or pedestrian movement. The methodology does not compute either potential or movement capacities for these.

Table 25.5 lists subject movements in the order in which movement capacities would be computed, along with the specific impedance factors that would be used. These represent initial computations. Some will be subject to additional modifications, which are explained below.

Table 25.5: Movement Capacity Computations before Adjustments (TWSC Intersections)

$c_{mi} = c_{pi} f_i$

Movement	Mvt No.	$f_i =$
Major Street Left Turn	1	P^*_{16}
Major Street Left Turn	4	P^*_{15}
Minor Street Right Turn	9	$P^*_{14} x P^*_{15}$
Minor Street Right Turn	12	$P^*_{13} x P^*_{16}$
Major Street U-Turn	1U	P_{12}
Major Street U-Turn	4U	P_9
Minor Street Through	8	$P_1 x P_4 x P^*_{15} x P^*_{16}$
Minor Street Through	11	$P_1 x P_4 x P^*_{15} x P^*_{16}$
Minor Street Left Turn	7	$P_1 x P_4 x P_{11} x P_{12} x P^*_{15} x P^*_{13}$
Minor Street Left Turn	10	$P_1 x P_4 x P_8 x P_9 x P^*_{14} x P^*_{16}$

cmi=cpi fi

*Optional impedance of a pedestrian movement; may be taken as 1.00 if experience indicates that pedestrians have little impact on vehicular operations.

Table 25.5: Full Alternative Text

There are four situations in which the equations of Table 25.5 must be modified:

1. When major street left turns are made from a lane shared with major street through vehicles, or from a short exclusive lane from which overflow queues are expected, the impedance of these movements must be modified.

2. For all cases of one-stage minor street left turns, some of the impedance factors overlap and must be adjusted to avoid overestimating their negative

effect.

3. When minor street through movements are made in a two-stage process, the impedances must be considered on each stage separately, and a total movement capacity estimated.

4. When minor street left-turn movements are made in a two-stage process, the impedances must be considered on each stage separately, and a total movement capacity estimated.

Each of these situations is discussed below.

Modification for Major Street Left Turns (From Shared Lanes or Short Exclusive Lanes)

All movements that are impeded by left turns from the major street assume that there is an exclusive lane long enough to accommodate all queued vehicles in this movement. Where this is not the case, the impedance factors P_1 and P_4 must be replaced by modified values P ′1 and P′4:

P′1=1−(1−P1)[1+Xm1nL+11−Xm1(nL+1)]Xm1=v2s2+v3s3 [25-9]
P′4=1−(1−P4)[1+Xm4nL+11−Xm4(nL+1)]Xm4=v5s5+v6s6 [25-10]

where:

P ′ 1 , P ′ 4
=adjusted impedance facotrs for Movements 1 and 4, P 1 , P 4
=unadjusted impedance factors for Movements 1 and 4, X m1 =v/c ratio for shared lane serving Movements 1 and 2, X m4
=v/c ratio for shared lane serving Movements 5 and 6, v i
=demand flow rate for Movement i, veh/h, s i
=saturation flow rate for Movement i, veh/hg, and n L
=number of vehicles that can be fully stored in the major s

turn lane.

In these equations, the value of P′ quickly approaches *P* as n_L gets larger. In the special case where *no* exclusive left-turn lane exists on the major street, n_L becomes “0,” and Equations 25-9 and 25-10 simplify to:

P'1/4=1−1−P1/41−Xm1/4 [25-11]

Implementing these adjustments requires the saturation flow rates for major street through and right-turn movements be known. This can be measured in the field, but the 2016 HCM recommends default values for use in most analyses:

s2,5=1,800 veh/hg/lns3,6=1,500 veh/hg/ln

Modification for Minor Street Left Turns: All One- Stage Turns

One-stage minor street left turns are potentially impeded by four higher priority vehicular movements and two pedestrian movements. Specifically, in terms of vehicular impedances, Movement 7 is impeded by Movements 1, 4, 11, and 12. Movement 10 is impeded by Movements 1, 4, 8, and 9. Because there are interdependencies among these impeding movements, it is likely that simply multiplying all of the applicable individual impedance factors would result in overstating the total impedance effects. In each case, the product of the individual impedance factors for the major street left turns and the opposing minor street through movement is adjusted. For Movement 7, let P″=P1 x P4 x P11 and for Movement 10, let P″=P1 x P4 x P8. Then:

P1/4/11 or P1/4/8=0.65 P″−P″P″+3+0.6 P″ [25-12]

These combined impedance factors replace the individual impedance factors in computing the total impedance adjustment for Movements 7 and 10:

f7=P1/4/11 x P12 x P15* x P13*f10=P1/4/8 x P9 x P14* x
[25-13]

where all terms are as previously defined. The asterisk indicates optional pedestrian impedance factors.

Modification for Two-Stage Minor Street Through Movements (Movements 8 and 11) and Two-Stage Minor Street Left Turns (Movements 7 and 10)

Two-stage movements from the minor street (through and left-turn movements) occur where the major street is divided by a median that is wide enough to provide storage for one or more vehicles. In such cases, drivers will execute two separate maneuvers—one that traverses the first roadway, and another that traverses the second roadway. Such a situation can affect Movements 7, 8, 10, and 11.

In each case, the process is the same:

- Estimate the movement capacity assuming that it is a one-stage movement. This process utilizes the total conflicting flow rate and all of the potentially impeding vehicular flows, as indicated in Table 25.5.
- Estimate the movement capacity separately for Stage I and Stage II of the subject maneuver. This is done using the separate conflicting flow rates for each stage, and only the impedances affecting that stage.

This process will yield three results:

cm1i=Movement i capacity, assuming that it is a one-stage Movement, veh/h,cIi=capacity of Stage I of Movem

Determining these three values of capacity for each two-stage movement will require finding corresponding values of potential capacity, based upon different conflicting flow rates. Converting potential capacities to movement capacities requires that the normal impeding flows be categorized based upon which stage of a two-stage movement they impact. Table 25.6 shows the impeding movements that would be considered in each stage of a two-stage maneuver.

Table 25.6: Impedances in Two-Stage Movements

Subject Movement	Impeding Flows – Stage I	Impeding Flows – Stage II
Minor Street TH – Movement 8	1, 15*	4, 16*
Minor Street TH – Movement 11	4, 16*	1, 15*
Minor Street LT – Movement 7	1, 15*	4, 11, 12, 13*
Minor Street LT – Movement 10	4, 16*	1, 8, 9, 14*

* Optional pedestrian impedance; may be set to 1.00 if pedestrian flows are thought to have little impact on vehicular flows.

Table 25.6: Full Alternative Text

Once movement capacities have been established (a) assuming a one-stage movement, (b) for Stage I of a two-stage movement, and (c) for Stage II of a two-stage movement, the total capacity for each subject movement may be determined. Two intermediate variables are computed:

a=1−0.32 e−1.3 nm   nm>0 [25-13]
yi=cIi−cm1icIIi−vL−cm1i [25-14]

where:

a =adjustment for two-phase movements, y i
=intermediate variable for movement i, v L
=major street left-turn and U-turn rate, veh/h (either v 1 + v 1u or v 4 + v 4U), and n m
=number of vehicles that can be stored in median; if n m is 0, a one-
stage movement exists, and this procedure is not used.

All other variables as previously defined.

Then:

For y≠1: cmi=ayi(nm+1)−1 [yi (yinm−1) (cIIi−vL) +(yi −1) cm1i]
For y=1: cmi=anm−1 [nm(cIIi−vL)+cm1i] [25-15]

where all variables have been previously defined.

At the end of this step, the adjustment of capacities due to gaps being utilized by higher-priority movements has been completed. The next step will address the issue of shared lanes.

6. Step 6: Determine Shared-Lane Capacities

There are two cases in which multiple movements may share a single lane. An exclusive left-turn lane on the major street may be shared by left-turners and those making a U-turn. Minor street lanes may be shared by two or even three movements. The shared-lane capacity in such cases is computed as:

cSHx=∑ivi∑i(vicmi) [25-16]

where:

c SHx =capacity of shared-lane x, veh/h v i
=flow rate for Movement i that is sharing lane x, veh/h, and
c mi
=Movement capacity of Movement i that is sharing lane x.

This equation is very simple conceptually. Consider a single-lane approach that serves all three minor street movements (LT, TH, RT). The flow rates and movement capacities for these lanes are:

Movement	Flow Rate, v	Movement Capacity, c_m	v/c_m
LT	100 veh/h	400 veh/h	0.250
TH	200 veh/h	500 veh/h	0.400
RT	50 veh/h	800 veh/h	0.063

25.3-7 Full Alternative Text

Essentially, the LT uses 0.250 of its movement capacity, the TH movement uses 0.400 of its movement capacity, and the RT uses 0.063 of its movement capacity. If all three movements are combined in a single lane, it is logical to assume that 0.250+0.400+0.063=0.713 of its capacity would be used. Thus, the total lane flow rate of 100+200+50=350 veh/h is equivalent to 0.713 of the lane's capacity. The capacity is, therefore, 350/0.713=491 veh/h (rounded to the nearest whole number).

Once shared-lane capacities are estimated, delays for each lane can be estimated.

Note that the HCM also includes a somewhat convoluted adjustment for *flared lanes* on a minor street approach. These are cases in which right-turning vehicles can use the flare as if it were a short exclusive RT lane. Consult the manual directly for this modification.

7. Step 7: Estimate Delay for Rank 2, 3, and 4 Movements

   Delay may be estimated for each lane from which a Rank 2, 3, or 4 movement takes place. The estimating equation is:

   dx=3.600cmx+900 T [vxcmx−1+(vxcmx−1)2+(3,600cmx) (vxcmx)450 T]+5 [25-17]

   where:

   d x =contol delay per vehicle in lane x, s/veh, v x =demand flow rate in lane x, veh/h, c mx =movement capacity of lane x, veh/h, T =analysis time period, h (normally 0.25h −15 minutes), and 5 =assumed acceleration/deceleration delay, s/veh.

   The subscript *x* may refer to a lane exclusively serving movement *i*, or may refer to a shared lane that is handling two or three different movements. In the latter case, both the demand flow rate and movement capacity are for the shared lane.

   The HCM also provides models to predict delays to Rank 1 movements (TH and RT on the major street), which may occur when major street LTs operate from a shared lane. However, it notes that these are most often negligible. Consult the HCM directly for this methodology.

   Once the control delay is estimated for each lane, averages may be computed for approaches with more than one lane:

   dA=∑i(di vi)∑ivi [25-18]

where all variables are as previously defined.

An average delay for the entire intersection can also be computed based upon the average delays on each approach (Equation 25-18). In most cases, delay to all Rank 1 movements is taken to be “0” in computing this average. As noted previously, the HCM contains a methodology to estimate delays to Rank 1 movements, but it is not often used.

Once average control delays have been computed for each lane (containing Rank 2, 3, and/or 4 movements), and again averaged for each approach, LOS may be applied to each lane and approach using the criteria of Table 25.7.

Table 25.7: Level of Service Criteria for TWSC Intersections

Control Delay (s/veh)	LOS by Volume-to-Capacity Ratio	
	$v/c \leq 1.00$	$v/c > 1.00$
0 – 10	A	F
>10 – 15	B	F
>15 – 25	C	F
>25 – 35	D	F
>35 – 50	E	F
>50	F	F

(Source: Reprinted with permission from *Highway Capacity Manual, 6th Edition – A Guide for Multimodal Mobility Analysis*, Transportation Research Board, the National

Academy of Sciences, Courtesy of the National Academies Press, Washington, D.C., 2016.)

Table 25.7: Full Alternative Text

The HCM does not recommend assigning a LOS to the overall intersection, as the large number of major street vehicles with no delay will distort the average. More important are the delay and LOS of the STOP-controlled lanes and approaches.

25.3 Interpreting Results

The HCM methodology for TWSC intersections is extremely detailed, and includes a number of complicated modifications, some of which are not covered in this text. At the end of the process, however, the primary question is simply “Can this TWSC intersection work acceptably or not?” Other than adding lanes to controlled approaches and/or reallocating them among the controlled movements, there is little that can be done to improve a TWSC intersection.

If it doesn’t work, other control options need to be considered:

- Signalization is a possibility, but signal warrants should be carefully considered. Signal warrants are covered in Chapter 15. A methodology for analysis of signalized intersections is covered in Chapters 22 and 23.
- Roundabouts are also a potential control option. A methodology for design and analysis of these is presented later in this chapter.
- Other geometric options under the general heading of “alternative intersections” may also be considered. See Chapter 26 for a general overview of these.

It should also be noted that “failure” of a TWSC intersection does not necessarily result in long queues and exorbitant delays. LOS F indicates that the controlled approaches do not operate well in accordance with the strict priority scheme assumed by the methodology. Failure may be exhibited by many controlled vehicles selecting gaps in conflicting traffic that are too small for safety, and higher accident rates may result. Major street vehicles may be forced to give way to minor street vehicles in such situations.

Sample Problem 24-1: Analysis of a TWSC Intersection

Consider the TWSC intersection shown in Figure 25.3. Analyze the intersection and determine the average control delays and LOS for each of the lanes and approaches that use gaps in a conflicting traffic stream.

Figure 25.3: Sample Problem in TWSC Analysis

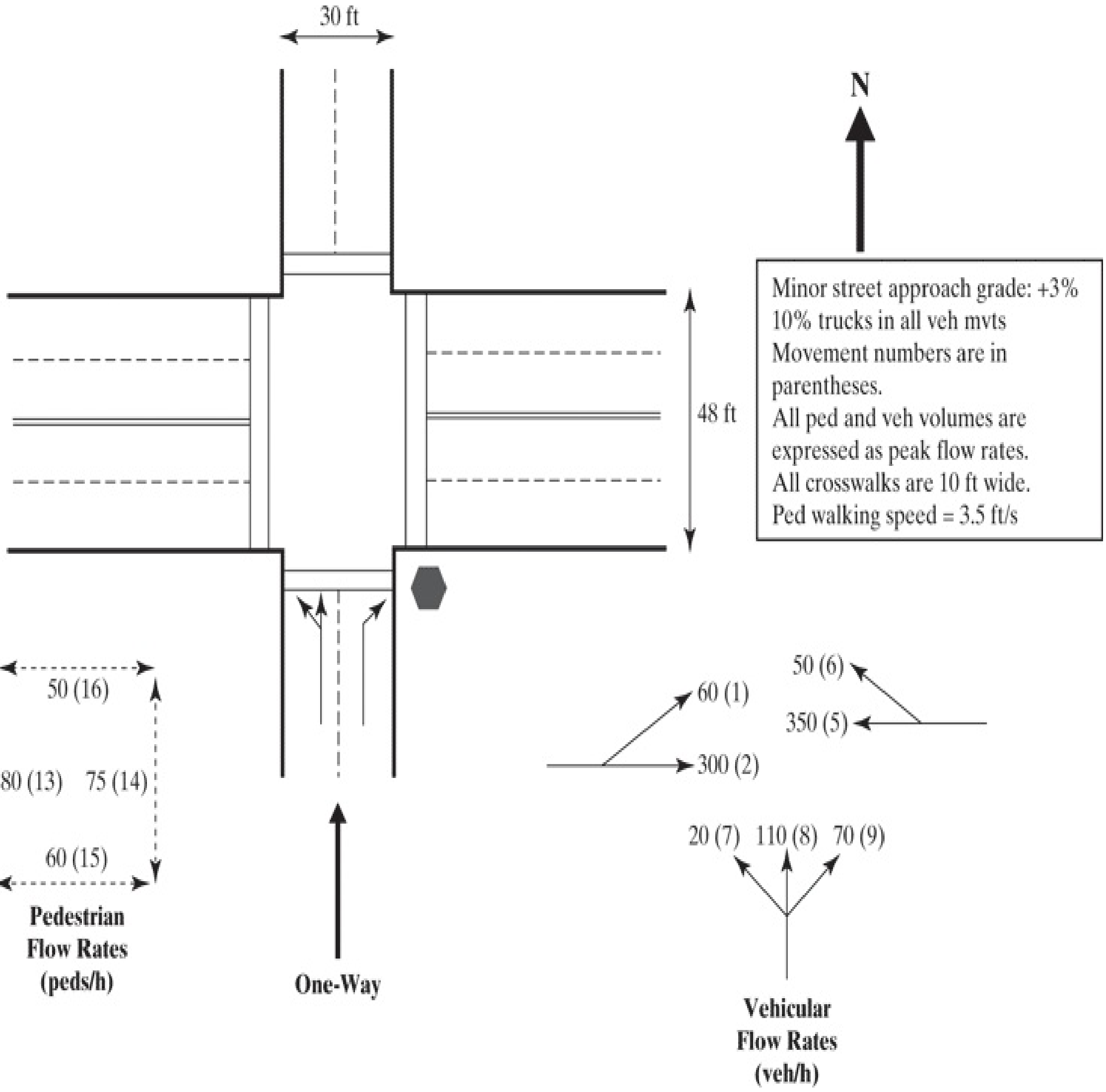


Figure 25.3: Full Alternative Text

The intersection shown is a two-way major street with a one-way minor street. The minor street, therefore, has only one controlled approach that must be analyzed. Because of the one-way street, major street left turns only exist in the EB direction, and are made from a shared lane. There are no U-turns at the intersection.

The solution will follow the steps outlined in the chapter.

1. Step 1: Express Demand as Flow Rates during a Peak 15-Minute Analysis Period

   This has already been done in the problem statement. Note also that movement numbers in accordance with Figure 25.1 are shown in parentheses to make the use of equations more straightforward. In any analysis of TWSC intersections, the movement numbers should be clearly assigned to avoid confusion.

2. Step 2: Determine the Conflicting Flow Rates for Each Movement

   Conflicting flows are illustrated in Figure 25.2, and equations for their computation are taken from Table 25.2. Note that because this is a one-way minor street, some of the minor street movements do not exist—that is, their value is 0 veh/h.

   For the major street LT (Movement 1):

   $v_{c1}=v_5+v_6+v_{16}=350+50+50=450$ conflicts/h

   For the minor street RT (Movement 9) onto a four-lane major street:

   $v_{c9}=0.5v_2+05v_3+v_{14}+v_{15}=(0.5\times300)+(0.5\times0)+75+60=285$ conflicts/h

   For the minor street TH (Movement 8), which is a one-stage movement:

   $v_{c8I}=2(v_1+v_{1U})+v_2+0.5v_3+v_{15}=2(60+0)+300+(0.5\times0)+60=480$ conflicts/h$v_{c8II}=2(v_4+v_{4U})+v_5+v_6+v_{16}$

   For the minor street LT (Movement 7) which is a one-stage movement with a four-lane major street:

vc7I=2(v1+v1U)+v2+0.5v3+v15=2(60+0)+300+(0.5×0)+60=480 conflicts/hvc7II=2(v4+v4U)+0.5v5+0.5v
(0.5×350)+
(0.5×50)+80=255 conflicts/hvc7=480+255=735 conflicts/

Note that in this case, the minor street LT has a lower conflicting flow rate than the minor street TH movement. This is primarily due to the fact that LTs are assumed to be made into the left lane of the major street, and thereby avoid ½ of the WB major street through movement.

3. Step 3: Determine Critical Gaps (Headways) and Follow-Up Times

   Critical gaps are computed using Equation 25-2, while follow-up times use Equation 25-3. Base critical gaps are found in Table 25.2, and adjustment factors are found in Table 25.3.

   Critical gaps are computed as:

   tci=tcbase+fcHVPHV+fcGG-f3LTtc1=4.1+(2.0×0.10)+(0×0)-0.0=4.3 stc7=7.5+(2×0.10)+(0.2×3)-0.7=7.6 stc8=6.5+(2×0.10)+(0.2×3)-0.0=7.3 stc9=6.9+(2×0.10)+(0.1×3)-0.0=7.4 s

   Follow-up times are computed as:

   tfi=tfbase+ffHVPHVtf1=2.2+(1.0×0.10)=2.4 stf7=3.5+(1.0×0.10)=3.6 stf8=4.0+(1.0×0.10)=4.1 stf9=3.3+(1.0×0.10)=3.4 s

4. Step 4: Compute Potential Capacities

   Potential capacities are computed using Equation 25-4:

   cpi=vci [e−vcitci/36001−e−vcitfi/3600]

   A spreadsheet should be used to implement this equation to avoid errors. The results are shown below:

   cp1=450 [e-450×4.3/36001-e-

450×2.4/3600]=1,014 veh/hcp9=285 [e-285×7.4/36001-e-285×3.4/3600]=672 veh/hcp8=930 [e-930×7.2/36001-e-930×4.1/3600]=216 veh/hcp7=735 [e-735×7.6/36001-e-735×3.6/3600]=305 veh/h

5. Step 5: Determine Movement Capacities

   This is the most complex step in the process because it must follow the strict priority regimen. Movement capacities require impedance factors for higher-priority movements, which cannot be estimated until the movement capacity of the higher-priority movements is known. Thus, computations start with Rank 2 movements and follow on to Rank 2 and Rank 4 movements. Rank 2 movements face only pedestrian impedances, which can be computed directly. Thus, movement capacity computations will proceed in the following order:

   - Major Street LT (Movement 1)
   - Minor Street RT (Movement 9)
   - Minor Street TH (Movement 8)
   - Minor Street LT (Movement 7)

   Because this problem is a one-way minor street, some of the impedances will be simplified, as some impeding vehicular flows do not exist in this case. From Table 25.4, these movements face the following potential impeding flows:

   - The major street LT (Movement 1) is impeded only by pedestrian flow 16.
   - The minor street RT (Movement 9) is impeded only by pedestrian flows 15 and 16.
   - The minor street TH (Movement 8) is impeded by vehicular flow 1 (vehicular flow 4 does not exist) and pedestrian flows 15 and 16.

- The minor street LT (Movement 7) is impeded by vehicular flow 1 (flows 4, 11, and 12 do not exist) and pedestrian flows 15 and 13.

Because so many potentially impeding movements do not exist, computation of impedance factors (P_y) is relatively simple. There is only one modification or adjustment that will be needed—the impedance factor for the major street LT (Movement 1) must be adjusted to account for its operation out of a shared lane on the major street.

Impedance factors for pedestrian movements 13, 15, and 16 are needed. They are computed using Equation 25-6:

$$P_j = 1 - \frac{v_j\left(\frac{w}{S_p}\right)}{3600} \qquad P_{13} = 1 - \frac{80\left(\frac{10}{3.5}\right)}{3600} = 0.9365 \qquad P_{15} = 1 - \frac{60\left(\frac{10}{3.5}\right)}{3600} = 0.9524 \qquad P_{16} = 1 = \frac{50\left(\frac{10}{3.5}\right)}{3600} = 0.9603$$

Movement capacities are estimated using Equations 25-7 and 25-8:

$$f_i = \prod_{y,j} P_y P_j \qquad c_{mi} = c_{pi} f_i$$

We must proceed in order of the priority of movements. Movement 1, the minor street LT, is impeded by pedestrian flow 16. Therefore:

$$c_{m1} = 1{,}014 \times 0.9603 = 974 \text{ veh/h}$$

Movement 9, the minor street RT, is impeded by pedestrian flows 15 and 16. Therefore:

$$c_{m9} = 672 \times 0.9524 \times 0.9603 = 615 \text{ veh/h}$$

Movements 7 and 8, the minor street LTs and TH flows, are impeded by pedestrian flows and vehicular movement 1. The impedance factor for vehicular movement 1 must, therefore, be computed. The initial impedance factor (P_1) is computed using Equation 25-5:

$$P_y = 1 - \frac{v_y}{c_{my}}$$

$P1=1-60974=0.9384$

Because the major street LT (Movement 1) shares a lane with through vehicles, this initial impedance factor must be adjusted using Equations 25-11 and 25-9:

$P'1=1-1-P11-Xm1$
$Xm1=v2s2+v3s3$

Default values are used for s_2 (1,800 veh/hg/ln) and s_3 (1,500 veh/hg/ln). Then:

$Xm1=3001800+01500=0.1667$
$P'1=1-1-0.93841-0.1667=1-0.06160.8333=0.9261$

Now, all of the relevant impedance factors are known, and the movement capacities for Movements 8 and 7 can be computed:

$cm8=cp8P'1P15P16=216\times0.9261\times0.9524\times0.9603=183$ veh/h
$cm7=cp7P'1P13P15=305\times0.9261\times0.9365\times0.9524=252$ veh/h

6. Step 6: Determine Shared-Lane Capacities

   In this sample problem, the minor street LT and TH flows share a single lane. The shared-lane capacity of this lane is given by Equation 25-16:

   $cSH7,8=v7+v8(v7cm7)+(v8cm8)=20+110(20252)+(110183)=1300.0794+0.6011=1300.6805=191$ veh/h

   The minor street RT operates out of an exclusive lane, and there are no major street U-turns. Thus, there are no other shared-lane situations to analyze.

7. Step 7: Estimate Delay for Rank 2, 3, and 4 Movements

   Delay is computed using Equation 25-17, which is applied to each lane:

dx=3600cmx+900 T [vccmx−1+(vxcmx−1)2+(3600cmx) (vxcmx)450 T]+5

The length of the analysis period, *T*, will be taken as 15 minutes, entered as 0.25 h in the equation. Because of its frequent use in the equation, it is valuable to compute the *v/c* ratio for each of the lanes under study:

- v1/cm1=60/974=0.0616
- v7/8/cSH7,8=(20+110)/191=0.6806
- v9/cm9=70/615=0.1138

An average delay for the minor street approach, which has two lanes, may also be computed using Equation 25-18:

Levels of service may now be assigned using the criteria of Table 25.7. Table 25.8 shows the estimated delays and resulting LOS for each lane and approach.

Then:

d1=3600974-1+(900×0.25) [(0.0616-1)+(0.0616-1)2+(3600974) 0.0616450×0.25]+5d1=2.575+225 [-.9384+(-0.9384)2+0.2277112.5]+5=7.8 s/vehd7/8=3600191-1+(900×0.25) [0.6806-1+(0.6806-1)2+(3600191) 0.6806450×0.25]+5d7/8=17.848+225 [-0.3194 (-0.3194)2+12.8281112.5]+5=55.5 s/vehd9=3600615+(900×0.25) [(0.1138-1)+(0.1138-1)2+(3600615) 0.1138450×0.25]+5d9=5.854+225 [-0.8862+0.
d7/8/9=(55.8×130)+(11.6×70)130+70=8066200=40.3 s/veh

Table 25.8: Delays and Levels of Service for Sample Problem

Lane/Approach	Movement(s)	Control Delay (s/veh)	LOS (Table 25-7)
Major Street LT	1	7.8	A
Minor Street Left Lane	7,8	55.5	F
Minor Street Right Lane	9	11.6	B
Minor Street Approach	7,8,9	40.3	E

Table 25.8: Full Alternative Text

The minor street left turns and through vehicles, which share a lane, are experiencing delays that are so long they must be classified as LOS F. Even when averaged with the minor street RT lane, the delays are still quite long, and the LOS (E) is not good.

There are a few minor adjustments that might be tried to improve things. First, allowing minor street through vehicles to use either the left or right lanes would be wise, but would not be expected to improve LOS beyond E. The LOS of F might, however, be mitigated in this fashion. If there were room to provide a third lane on the minor street, that would bring a larger improvement, as each movement would have its own lane. This is possible given the 30-foot street width, which would allow for three 30-foot lanes. If about six more feet of right-of-way were available, three 12-foot lanes would be preferable.

While not shown here, analysis of the three-lane minor street option would be fairly easy, as it would use most of the results of the initial analysis. There would be no shared lane to analyze, and the delay for each minor street movement, now in its own lane, would be computed. A roundabout option might also be considered, and could be analyzed using the methodology presented later in the chapter.

Signalization would not immediately be a clear option, as the volumes do not approach volume warrants for installing a signal. If a high number of accidents were occurring, it might trip the *crash warrant* for signalization.

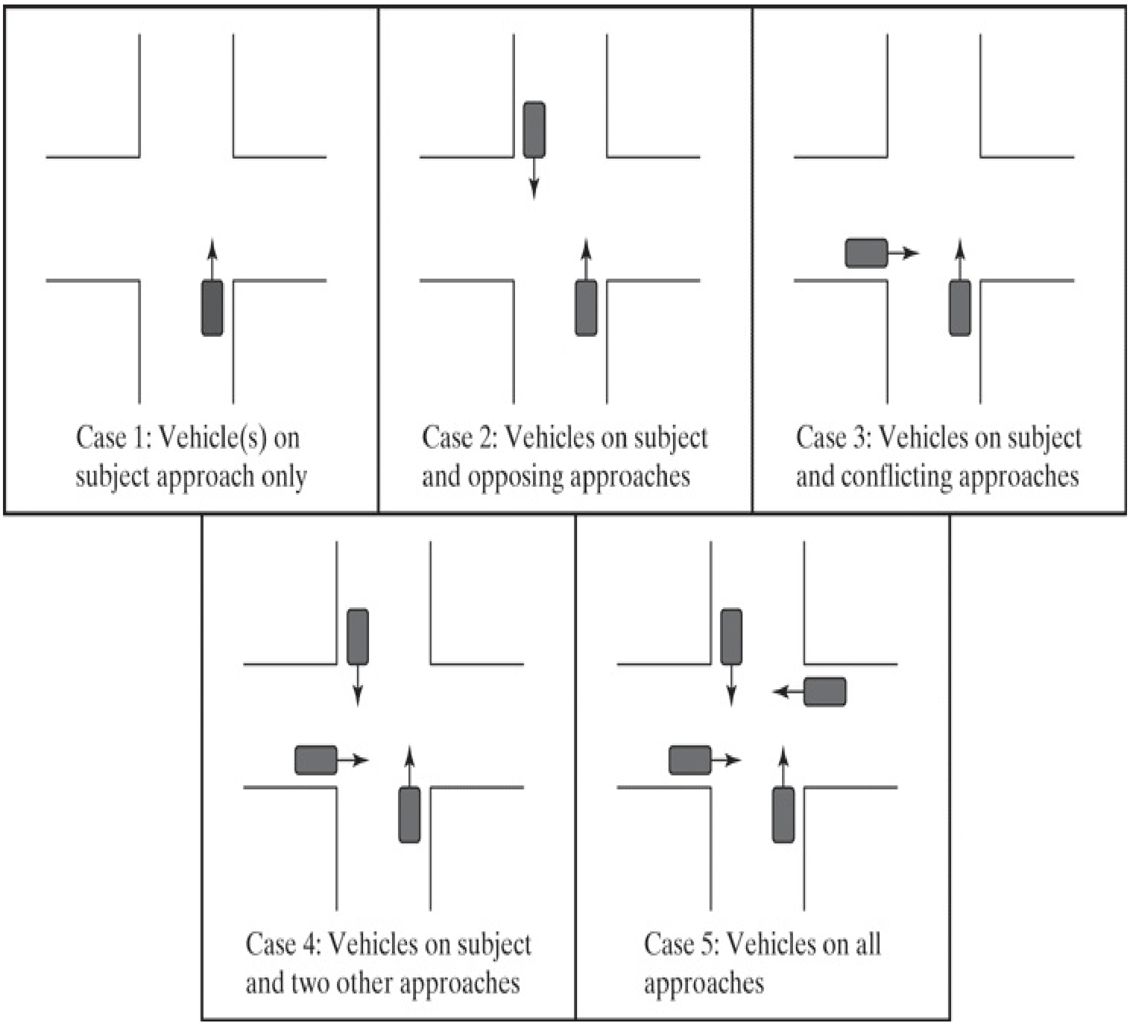
Part II All-Way STOP-Controlled Intersections

The HCM analysis methodology for AWSC intersections is cumbersome, and involves multiple (in some cases over 100) iterations to achieve a final result. This is because there are numerous scenarios that a driver stopped at such an intersection can face, and each one involves different conflicts that the driver must negotiate to complete the desired maneuver.

The methodology focuses on estimating the capacity of *each* approach lane to the AWSC intersection, while holding the mix of demands on opposing and conflicting approaches constant. The approach being analyzed is referred to as the *subject approach*. The *opposing approach* is the approach directly opposite the subject approach. A *conflicting approach* is one that handles vehicles coming from the right or left of the subject approach.

Figure 25.4 illustrates the number of situations that a driver stopped on the subject approach might face.

Figure 25.4: Various Scenarios Faced by a Driver on the Subject Approach at a AWSC Intersection



(Source: Reprinted with permission from *Highway Capacity Manual, 6th Edition – A Guide for Multimodal Mobility Analysis*, Transportation Research Board, the National Academy of Sciences, Courtesy of the National Academies Press, Washington, D.C., 2016.)

Figure 25.4: Full Alternative Text

As indicated in Figure 25.4, vehicles face one of five scenarios at an AWSC intersection, each of which presents a different set of potential conflicts and driving decisions:

- Case 1: Vehicles only on the subject approach. No vehicles are on the opposing or conflicting approaches.

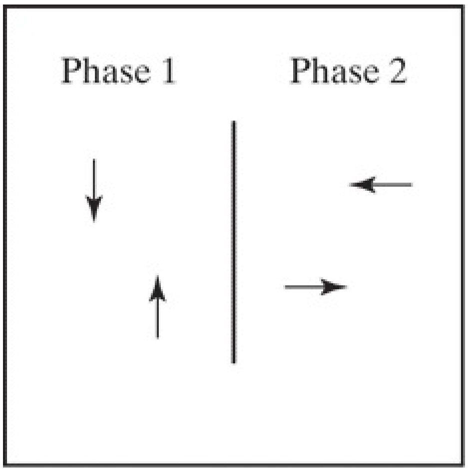
- Case 2: Vehicles are on the subject and opposing approaches only. No vehicles are on the conflicting approaches.

- Case 3: Vehicles are on the subject and one of the conflicting approaches. No vehicles are on the opposing or other conflicting approach.
- Case 4: Vehicles are on the subject approach and two of the three other opposing and conflicting approaches. Either the opposing or one of the conflicting approaches has no vehicles.
- Case 5: Vehicles are on all of the approaches—subject, opposing, and both conflicting approaches.

Obviously, as the situation moves from Case 1 to higher cases, the complexity facing the subject driver becomes progressively more difficult. The degree of complexity is further complicated by the mix of movements on opposing and conflicting approaches. Adding to the complexity is the fact that right-of-way priorities at AWSC intersections are not clearly defined, and drivers are generally not sure of the many “who goes first” decisions they may have to make.

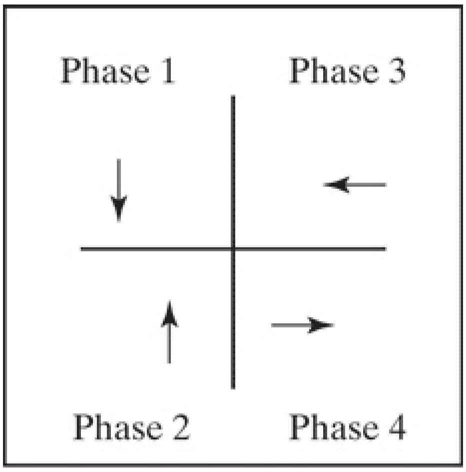
Field observations suggest that AWSC intersections operate in a quasi-signalized mode, depending upon the basic geometry. For AWSC intersections with one lane on each approach, drivers tend to behave as if a two-phase signal were in place, with vehicles alternating between the E–W and N–S streets. Where AWSC intersections have two lanes on each approach, drivers tend to behave as if a four-phase signal were in place, with the right-of-way rotating for each approach individually in a clockwise fashion. These regimes are illustrated in Figure 25.5.

Figure 25.5: Operating Regimes for AWSC Intersections



(a) 2-Phase Regime (One-Lane Approaches)

25.5-10 Full Alternative Text



(b) 4-Phase Regime (Multilane Approaches)

25.5-10 Full Alternative Text

(Source: Reprinted with permission from *Highway Capacity Manual, 6th Edition – A Guide for Multimodal Mobility Analysis*, Transportation Research Board, the National Academy of Sciences, Courtesy of the National Academies Press, Washington, D.C., 2016.)

It should be noted that three-lane approaches are rarely STOP-controlled, so they are not covered by this methodology. Any approach with more than three lanes should *never* be controlled with a STOP-sign.

The analysis methodology is based upon three primary variables:

- Saturation headway (h_{si}): time between successive vehicles departing an approach for Case *i*, assuming there is a continuous queue of vehicles present (s/veh),

- Departure headway (h_d): average time between successive vehicles on an approach departing, taking into account the probabilities of each potential case being present, and

- Service time (t_s): average time spent in the first queue position waiting to depart; it is equal to the departure headway minus the amount of time it takes to move up from the second queue position into the first (called *move-up time, m*) (s/veh).

The analytic construct becomes complex, however, because the saturation headway is different for every potential case. The departure headway must take into account the probabilities of each of the potential cases occurring over time. Probabilities are not, however, limited to five—one for each case. Where multiple lanes exist on an approach, probabilities are needed for *each lane of each approach* for every possible scenario that exists. In fact, at an AWSC intersection with all approaches having two lanes, there are 64 different scenarios that might exist, each with a different saturation headway and each with a different probability.

The capacity of each lane of each approach at an AWSC intersection is computed as part of this methodology. It is, however, computationally very complex, and involves two levels of iterations (each level can involve high numbers of iterations):

- Departure headways (h_d) depend upon the degree of utilization of opposing and conflicting lanes, which is dependent upon their departure headways. Computations start with an assumed departure headway for all lanes and continue iterating until all departure headways computed are within ±0.1 of the assumed value. For each iteration, the computed headways for the previous iteration are assumed. In some cases, closure may take more than 100 iterations.

- Capacity for each lane of each approach is also iterative (once the departure headways are computed). Capacities are found by increasing subject lane flow rates until the degree of utilization reaches a value of 1.00.

Obviously, actual computations will rely on software, such as the Highway Capacity Software package. The process will be defined in the sections that follow, and a simple sample problem (two one-way streets with one lane each) will be used to illustrate it.

25.4 Computational Steps

1. Step 1: Convert Movement Demands to Flow Rates and Determine Lane Flow Rates

   Just as was done for TWSC intersections, volumes used for analysis are stated as flow rates in mixed veh/h. It is best for demands to be field measured in 15-minute increments, which would allow the selection of the worst period(s) for analysis. In such cases, flow rates would be calculated from field data. Where only hourly demand volumes are known, they are converted to flow rates using a single PHF for the entire intersection (see Equation 25-1).

   If there are multiple lanes on an approach, demand flow rates must be assigned to each lane. In general, all left turns are assigned to the left lane, all right turns are assigned to the right lane, and through movements are evenly divided among available lanes.

2. Step 2: Determine Geometry Group for the Intersection

   Many of the parameters used vary with the specific geometry of the AWSC intersection, based upon the number of lanes on each approach. Table 25.9 is used to determine the geometry group for the intersection under study.

Table 25.9: Determination of Geometry Group for AWSC Intersections

No. of Lanes				
Intersection Configuration	Subject Approach	Opposing Approach	Conflicting Approaches[a]	Geometry Group
Four leg or T	1	0 or 1	1	1
Four leg or T	1	0 or 1	2	2
Four leg or T	1	2	1	3a/4a
T	1	2	2	3b
Four leg	1	2	2	4b
Four leg or T	1	0 or 1	3	5
	1	3	1	
	2	0, 1, or 2	1 or 2	
	3	0 or 1	1	
	3	0 or 1	2 or 3	
	3	2 or 3	1	
Four leg or T	1	3	2	6
	1	2	3	
	1	3	3	
	2	3	1, 2, or 3	
	2	0, 1, 2, or 3	3	
	3	2 or 3	2 or 3	

Note: [a] If the number of lanes on conflicting approaches is different, the higher of the two should be used.

(Source: Reprinted with permission from *Highway Capacity Manual, 6th Edition – A Guide for Multimodal Mobility Analysis*, Transportation Research Board, the National Academy of Sciences, Courtesy of the National Academies Press, Washington, D.C., 2016.)

Table 25.9: Full Alternative Text

3. Step 3: Determine Saturation Headways for Each Conflict Scenario

   The saturation headway (h_{si}) must be established for each scenario *i*. As noted, for an AWSC intersection with four two-lane approaches, there are 64 such scenarios that are possible. Each saturation headway is computed as:

   hsi=hbasei+hadj [25-19]

   where:

   h si =saturation headway for scenario i, s/veh, h basei =base saturation headway for scenerio i, s/veh, and h adj =adjustments to the base saturation headway, s/veh.

   Base saturation headways are given in Table 25.10.

Table 25.10: Base Saturation Headways for AWSC Intersections

DOC Case	No. of Veh.[a]	Base Saturation Headway (h_{basei}) for Geometry Group (s/veh):							
		1	2	3a	3b	4a	4b	5	6
1	0	3.9	3.9	4.0	4.3	4.0	4.5	4.5	4.5
2	1 2 ≥3	4.7	4.7	4.8	5.1	4.8	5.3	6.4 7.2	6.0 6.8 7.4
3	1 2 ≥3	5.8	5.8	5.9	6.2	5.9	6.4	6.4 7.2	6.6 7.3 7.8
4	2 3 4 ≥5	7.0	7.0	7.1	7.4	7.1	7.6	7.6 7.8 9.0	8.1 8.7 9.6 12.3
5	3 4 5 ≥6	9.6	9.6	9.7	10.0	9.7	10.2	9.7 9.7 10.0 11.5	10.0 11.1 11.4 13.3

Note: [a] Number of vehicles on opposing and conflicting approaches.

(Source: Reprinted with permission from *Highway Capacity Manual, 6th Edition – A Guide for Multimodal Mobility Analysis,* Transportation Research Board, the National Academy of Sciences, Courtesy of the National Academies Press, Washington, D.C., 2016.)

Table 25.10: Full Alternative Text

The adjustment to base saturation headway is computed as:

$h_{adj}=h_{LT}P_{LT}+h_{RT}P_{RT}+h_{HV}P_{HV}$ [25-20]

where:

h_{adj}=adjustment to base saturation headway, s/veh,h_i=adju

Adjustment factors (h_i) are given in Table 25.11.

Table 25.11: Adjustment Factors to Base Saturation Headway

Adjustment For:	Saturation Headway Adjustments for Geometry Group:							
	1	2	3a	3b	4a	4b	5	6
LT	0.2	0.2	0.2	0.2	0.2	0.2	0.5	0.5
RT	−0.6	−0.6	−0.6	−0.6	−0.6	−0.6	−0.7	−0.7
HV	1.7	1.7	1.7	1.7	1.7	1.7	1.7	1.7

(Source: Reprinted with permission from *Highway Capacity Manual, 6th Edition – A Guide for Multimodal Mobility Analysis*, Transportation Research Board, the National Academy of Sciences, Courtesy of the National Academies Press, Washington, D.C., 2016.)

Table 25.11: Full Alternative Text

Of course, the complexity in this process is recognizing how many scenarios exist in any given case, and then determining the proper values to select from Tables 25.10 and 25.11 for each.

In the worst case, where all approaches have two lanes, there are a total of six lanes on the opposing and two conflicting approaches. The scenarios must cover every possibility of a vehicle being present (or not) on each of those lanes. In essence, the concept is simple: For each opposing and conflicting lane, there is either 0 vehicles present, or 1 vehicle present.

For six opposing and conflicting lanes, how many potential scenarios might a vehicle on the subject approach face? Consider the following:

- For DOC 1 (no vehicles on either the opposing or conflicting approaches), there is only one possible scenario. All six opposing and conflicting lanes are empty.

- For DOC 2 (vehicles on the opposing approach, none on conflicting approaches), there are three possible scenarios—one vehicle in lane 1, one vehicle in lane 2, or vehicles in both lanes 1 and 2 of the opposing approach.

- For DOC 3 (vehicles on one conflicting approach only), there are the same three scenarios as for DOC 2, but for each of two conflicting approaches—thus, there are six possible scenarios.

- For DOC 4 (vehicles on two opposing and conflicting approaches, but one such approach remains empty), the situation is a bit more complex. There are scenarios involving two, three, or four vehicles being present on these approaches. There are only three scenarios for four vehicles being present on two of the three opposing and conflicting approaches—for each, one of these approaches remains empty, and there are only three of them. If three opposing/conflicting vehicles are present (on two approaches), there are twelve potential scenarios—there must be two vehicles on one approach, and one vehicle on another. There are only three possibilities to have two vehicles on one approach. For each of these, there are four other lanes for the third vehicle to reside. Thus, there are 3×4=12 scenarios for three vehicles. If only two vehicles are present, we have the reverse of the case where three vehicles are present—one of the approaches must be empty, and there are only three ways that can occur. For each of those cases, one vehicle must be on each of the other approaches, and there are two lanes in each approach, breeding 2×2=4 possibilities. Thus, the total number of scenarios for two vehicles is 3×4=12. Summarizing, for DOC 4, there are 3 scenarios for four opposing/conflicting vehicles to be present, 12 scenarios for three opposing/conflicting vehicles to be present, and 12 scenarios for two opposing/conflicting vehicles to be present. The total is 27 scenarios for DOC 4.

- For DOC 5 (vehicles present on all opposing/conflicting approaches), the situation becomes even more complex. Options for this to occur can involve as few as three vehicles (one on each opposing/conflicting approach) to six vehicles (all six opposing/conflicting lanes have a vehicle). There is only one option for six vehicles to be present. If five vehicles are present, one of the six opposing/conflicting lanes is empty. There are six options for this. If four vehicles are present, two opposing/conflicting approaches will have one vehicle, and the other will have two. There are three options for the approach with two vehicles. For each one of these options, there are four possible combinations of lane occupancies in the approaches with only one vehicle. This leads to 3×4 or 12 scenarios for four vehicles to be present. For three vehicles to be present, one must be on each approach, and there are two options in each case, leading to 2×2×2=8 scenarios. The total number of scenarios for DOC 5 is 1+6+12+8=27.

Thus, the total number of scenarios for an AWSC intersection with two lanes on each approach is 64. Table 25.12 summarizes these 64 possible scenarios. The methodology requires that a saturation headway be estimated for *each* of these scenarios—and there are 64 scenarios for each of the four intersection approaches, for a total of 256 saturation headways that would have to be estimated.

1. # Table 25.12: Lane Occupancy Scenarios for an AWSC Intersection with 2 Lanes on Each Approach

i	DOC Case	No. of Vehicles	Opposing Approach		Conflicting Left Approach		Conflicting Right Approach	
			Lane 1	Lane 2	Lane 1	Lane 2	Lane 1	Lane 2
1	1	0	0	0	0	0	0	0
2	2	1	1	0	0	0	0	0
3			0	1	0	0	0	0
4		2	1	1	0	0	0	0
5	3	1	0	0	1	0	0	0
6			0	0	0	1	0	0
7			0	0	0	0	1	0
8			0	0	0	0	0	1
9		2	0	0	1	1	0	0
10			0	0	0	0	1	1
11	4	2	0	0	0	1	0	1
12			0	0	1	0	0	1
13			0	0	1	0	1	0
14			0	0	0	1	1	0
15			0	1	0	1	0	0
16			1	0	1	0	0	0
17			0	1	0	0	1	0
18			1	0	0	1	0	0
19			0	1	1	0	0	0
20			0	1	0	0	0	1
21			1	0	0	0	1	0
22			1	0	0	0	0	1
23		3	0	0	0	1	1	1
24			0	0	1	1	0	1
25			0	0	1	1	1	0
26			1	0	1	1	0	0
27			1	1	1	0	0	0
28			1	1	0	0	1	0
29			1	1	0	0	0	1
30			0	1	1	1	0	0
31			1	0	0	0	1	1
32			0	0	1	0	1	1
33			1	1	0	1	0	0
34			0	1	0	0	1	1
35		4	1	1	0	0	1	1
36			0	0	1	1	1	1
37			1	1	1	1	0	0

25.6-13 Full Alternative Text

38	5	3	0	1	0	1	0	1
39			1	0	0	1	1	0
40			0	1	1	0	1	0
41			0	1	0	1	1	0
42			0	1	1	0	0	1
43			1	0	1	0	0	1
44			1	0	0	1	0	1
45			1	0	1	0	1	0
46		4	1	0	0	1	1	1
47			0	1	1	1	1	0
48			0	1	1	1	0	1
49			1	0	1	0	1	1
50			1	0	1	1	1	0
51			0	1	0	1	1	1
52			1	1	1	0	0	1
53			1	0	1	1	0	1
54			0	1	1	0	1	1
55			1	1	0	1	1	0
56			1	1	0	1	0	1
57			1	1	1	0	1	0
58		5	1	0	1	1	1	1
59			1	1	0	1	1	1
60			1	1	1	0	1	1
61			0	1	1	1	1	1
62			1	1	1	1	1	0
63			1	1	1	1	0	1
64		6	1	1	1	1	1	1

25.6-14 Full Alternative Text

Notes: DOC=degree-of-conflict; No. of Vehicles=total number of vehicles on opposing and conflicting approaches.

(Source: Reprinted with permission from *Highway Capacity Manual, 6th Edition – A Guide for Multimodal Mobility Analysis*, Transportation Research Board, the National Academy of Sciences, Courtesy of the National Academies Press, Washington, D.C., 2016.)

Of course, the situation for other geometries is simpler. In the case of an AWSC intersection with two one-way streets and one lane per approach, there are only two scenarios to be considered: either the one conflicting lane has a vehicle, or it doesn’t. There is no opposing approach, and there is no second conflicting approach. Each subject approach, therefore, has only two scenarios, and a total of four saturation headways would have to be estimated.

For the most normal case of an AWSC intersection with four approaches of one lane, the situation is also significantly simpler. Without going through all of the logic, there are 11 scenarios for each of four approaches.

Obviously, there are other geometric situations that could exist: T-intersections with three approaches, intersections in which some approaches have one lane and others have two, and so on. For each one, the possible scenarios would have to be established.

In any event, once scenarios have been established, the saturation headway for each of them must be estimated using Equations 25-19 and 25-20 with Tables 25.10 and 25.11. Consider the following example:

What is the saturation headway for an AWSC intersection with four two-lane approaches for

Scenario 42 (Table 25.12)? The subject approach has 10% right turns and 8% left turns, and contains 5% heavy vehicles.

From Table 25.9, for two lanes on the subject approach, two lanes on the opposing approach, and two lanes on each conflicting approach, the geometry group is five. Scenario 42 is a DOC 5 case with three vehicles present on opposing/conflicting lanes.

From Table 25.10, the base saturation headway (h_{si}) is 9.7 s. From Table 25.11, the adjustment factors for this case are 0.5 for LT, −0.7 for RT, and 1.7 for HV. Then:

hadj=hLTPLT+hRTPRT+hHVPHVhadj=(0.5×0.08)+(-0.7×0.10)+(1.7×0.05)=0.055hs42=hbase42+hadj=9.7+0.055=9.755 s

While each estimation of a saturation headway for a scenario is relatively straightforward, there are many such estimates to be made. Thus, the computational task is difficult and cumbersome, and best done using software.

2. Step 4: *Determine Departure Headways (*h_d*)*

The departure headway, at least conceptually, is estimated as:

hdx=∑i=1NP′i hsi [25-21]

where:

h dx =departure headway for lane x, s/veh, P ′ i =adjusted probability that scenario i will occur, h si =saturation headway for scenario i, s/veh, and N =number of scenarios that may occur(2−64).

Essentially, the departure headway is estimated as the sum of the saturation headways for potential scenarios multiplied by the probability of that scenario occurring.

To implement this equation, the probability of each potential scenario occurring must be estimated. The problem is that the probabilities are dependent upon the degree of saturation (X_j) in the opposing and conflicting approaches—which are in turn dependent upon the departure headways in those lanes. Thus, the entire process becomes iterative, and continues until the *initially assumed* value of departure headway is within 0.1 of the computed value. The process will be presented and illustrated in a very simple example at the end of Part II.

The process begins by computing an initial value of occupancy for each opposing and conflicting lane:

Xj=vjhdj3600 [25-22]

where:

X j =degree of saturation for opposing/conflicting lane j, v
j
=demand flow rate in opposing/conflicting lane j, veh/h, ar
h dj
=departure headway for opposing/conflicting lane j, s/veh.

Iterations begin by assuming a departure headway of 3.2 seconds for all opposing and conflicting lanes. In subsequent iterations, the departure headways resulting from the previous iteration are used as a starting point. Iterations continue until all values of h_{dj} are within ±0.1 of the value assumed at the beginning of the iteration.

Before any iterations can begin, however, the probability of each possible scenario occurring must be estimated. For each defined scenario, various opposing/conflicting lanes either have one vehicle present, or zero vehicles present. The probability of a lane *j* being occupied is X_j. The probability of a lane *j* being unoccupied is (1−Xj). Thus, the probability of any given scenario existing is the product of the probability of any given lane being occupied or not occupied.

Consider Scenario 22 of Table 25.12. In this case of an AWSC intersection with four two-lane approaches, there is a vehicle in lane 1 of the opposing approach, a vehicle in lane 2 of the right conflicting approach, and no vehicles on all other lanes: lane 2 opposing, lanes 1 and 2 left conflicting, and lane 1 right conflicting. To simplify, let us assume that the degree of saturation is 0.25 for all six opposing and conflicting lanes. The probability of Scenario 22 occurring is the product of the probability that lane 1 opposing and lane 2 right conflicting is occupied, while all other lanes are unoccupied, or:

$P_{22}=$
$(X_{O1})\times(1-X_{O2})\times(1-X_{CL1})\times(1-X_{CL2})\times(1-X_{CR1})\times(X_{C}$

where:

$O1$ =opposing lane 1, $O2$ =opposing lane 2, $CL1$ =conflicting left lane 1, $CL2$ =conflicting left lane 2, $CR1$ =conflicting right lane 1, and $CR2$ =conflicting right lane 2.

There are six opposing/conflicting lanes. Therefore, there are six probabilities that each lane is occupied or not (as specified by the scenario). Then, for Scenario 22:

$P_{22}=0.25\times(1-0.25)\times(1-0.25)\times(1-0.25)\times(1-0.25)\times0.25=($

In general terms:

$P_i=\prod_{j=1}^{n}P(a_j)$ [25-23]

where:

P_i =probability of scenario i occuring, a_j =indicator of whether opposing/conflicting lane j is occupi
$a_j=1$ when occupied, $a_j=0$ when unoccupied), j =opposing/conflicting lane, and n =number of opposing/conflicting lanes (2−6)

Equation 25-23 assumes that there is no correlation among the various probabilities for possible scenarios. Actually, there is some serial dependence of each probability related to the previous probability computation. The HCM presents a methodology to adjust for these dependencies.

First, the possible scenarios are grouped by DOC. In Table 25.12, DOC 1 includes only Scenario 1; DOC 2 includes Scenarios 2–4; DOC 3 includes Scenarios 5–10, DOC 4 includes Scenarios 11–37, and DOC 5 includes Scenarios 38–64. Obviously, with less complex geometries, fewer scenarios would exist, and the groupings would change.

The probability of each DOC case occurring is computed as the sum of the probabilities of the scenarios that produce that DOC:

PDOCz=∑i=xyPi [25-24]

where:

P DOCz =probability that DOC z exists (z=1−5), P i =probability that scenario i exists, x =first scenario in DOC z, and y =last scenario in DOC z.

For example, in Table 25.12, P_{DOC3} would be the sum of the probabilities (P_i) for Scenarios 5 through 10.

Then, adjustment factors are computed for each DOC and applied to all scenario probabilities within that DOC:

AdjPDOC1=0.01 [PDOC2+2PDOC3+3PDOC5]/1AdjPD PDOC2]/3AdjPDOC3=0.01 [PDOC4+2PDOC5-3PDOC3]/6AdjPDOC4=0.01 [PDOC5-6PDOC4]/27AdjPDOC5=-0.01 [10PDOC5]/27 [25-25]

where $AdjP_{DOCz}$ is the adjustment to be applied to *all* scenario probabilities within DOC *z* and all other terms are as previously defined. Then, adjusted probabilities for each scenario are computed as:

P′i=P+iAdjPDOCz [25-26]

where all terms are as previously defined.

At the end of this obviously complex computational process, the departure headway, h_d, will have been estimated for each opposing and/or conflicting lane at the AWSC intersection.

3. Step 5: Determine the Capacity of Each Lane of Each Approach to the AWSC Intersection

   The capacity of each approach lane is now computed. Each subject approach lane must be considered in turn. Capacity is defined as the maximum throughput the lane can sustain, considering the demand flow rates of each of the opposing/conflicting lanes to be fixed.

   The problem again is that doing this requires a complex iterative approach. The degree of saturation (X_i) for the subject lane has been computed as part of the determination of departure headways in Step 4. If the value is below 1.00, the demand flow rate on the subject lane is increased. If it is above 1.00, the demand flow rate is decreased. The objective is to iterate the process until the degree of saturation is exactly 1.00. Step 4 is now rejoined with a new demand flow rate on the subject lane, and the demands on all opposing/conflicting lanes held constant. This is done for each subject lane in turn until the degree of saturation reaches 1.00, which defines its capacity.

   Again, due to the complexity of computations, software is virtually always used for this process.

4. Step 6: Determine Control Delay and Level of Service for Each Lane, Approach, and the Intersection as a Whole

Average control delay is computed for each lane of the AWSC intersection. It is then averaged to obtain similar values for each approach, and again for the intersection as a whole. Levels of service are directly related to average control delay, and are the same as those for TWSC intersections, shown in Table 25.7.

Average control delay for each lane is estimated as:

dx=tsx+900T [(Xx−1)+(Xx−1)2+hdxXx450T]+5 [25-27]

where:

d x =control delay for lane x, s/veh, t sx
=service time for lane x, s/veh(t sx = h dx
−m) where m=2.0 s/veh for geometry groups 1−4, and 2.3
T =length of the analysis period, h, X x
=degree of saturation for lane x (X x = v x h dx
/3600), and h dx =departure headway for lane x, s/veh.

Average delays for each approach and the intersection as a whole are weighted by the demand flow rate on each lane and approach. These computations are similar to those conducted for TWSC intersections, except that for AWSC intersections, the intersection average is more meaningful, in that all lanes are controlled.

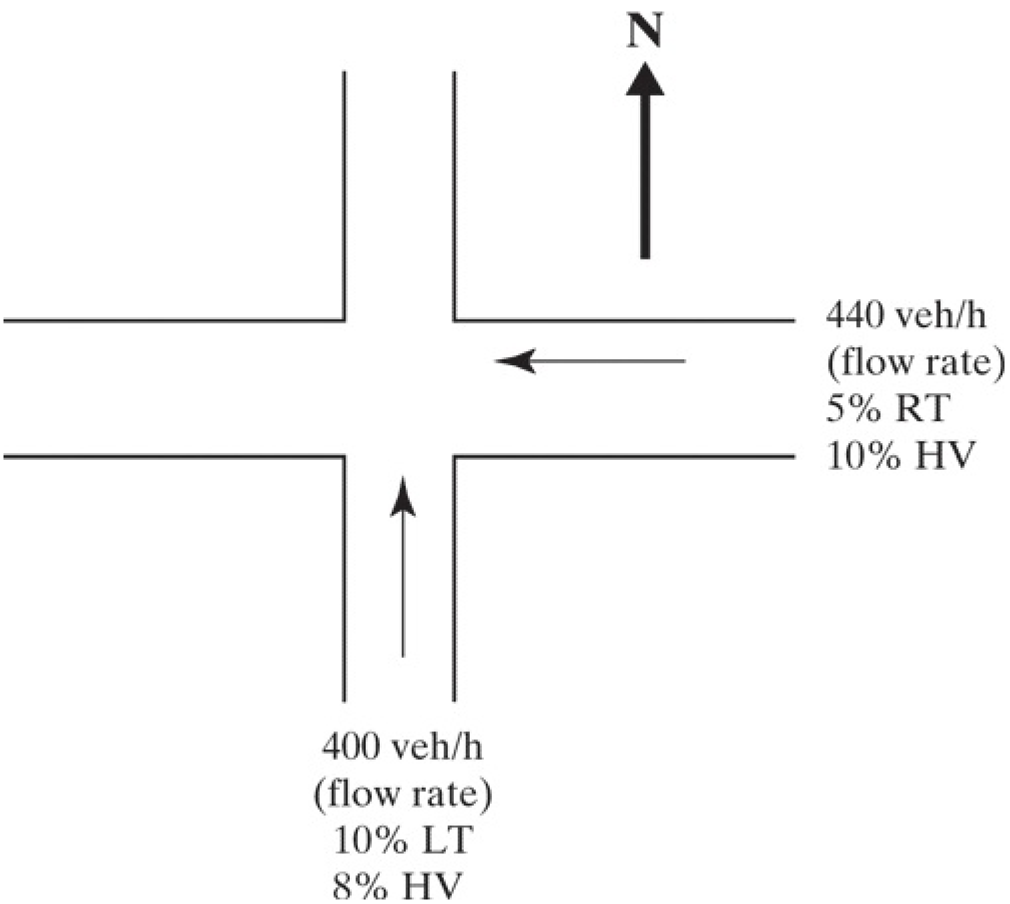
25.5 Comment

There is an obvious question related to this methodology: Is the complexity of the procedure warranted for AWSC intersections? AWSC intersections generally do not serve high volumes, so the fundamental issue is will they work or not in a given situation. It can be argued that the detail of individual lane conditions and specific delays may simply be overanalysis of a relatively straightforward form of control. This is always the issue in operational analysis: How much detail is needed to answer the fundamental questions that are needed? Yet, even though this process can be implemented using software, it is important that professionals using such software understand the underlying principles that are applied. Unwarranted complexity tends to obscure understanding.

Sample Problem 25-2: Analysis of an AWSC Intersection

Only the simplest of problems can be adequately documented by hand. Consider the intersection shown in Figure 25.6. It shows an AWSC intersection of two one-way streets, with each approach having one lane.

Figure 25.6: Sample Problem for AWSC Analysis



(Source: Reprinted with permission from *Highway Capacity Manual, 6th Edition – A Guide for Multimodal Mobility Analysis*, Transportation Research Board, the National Academy of Sciences, Courtesy of the National Academies Press, Washington, D.C., 2016.)

Figure 25.6: Full Alternative Text

Note that demands are already stated as flow rates; no PHF conversions will be necessary. Essentially, Step 1 of the solution is already completed in the problem statement.

Before computations begin, it is important to establish the number of scenarios that exist for the NB and WB approaches, each of which will be treated as the subject approach in turn. In each case, there are only two scenarios to address: one in which the conflicting approach is occupied, and one in which it is not. Table 25.13 describes these four scenarios.

Table 25.13: Scenarios for

Sample Problem in AWSC Analysis

Subject Approach	Scenario	Vehicles in Conflicting Lane	Geometry Group (Table 25-9)	DOC
WB	1	0	1	1
	2	1	1	3
NB	1	0	1	1
	2	1	1	3

Table 25.13: Full Alternative Text

1. Step 2: Determine the Intersection Geometry Group

   As indicated in Table 25.13, Table 25.9 indicates that this intersection may be classified in Geometry Group 1.

2. Step 3: Determine Saturation Headways for Each Scenario

   For both subject approaches, there are a total of four scenarios, two for each subject approach. Saturation headways for each scenario are estimated using Equation 25-19 and 25-20:

   hsi=hbasei+hadjhadj=hLTPLT+hRTPLT+hHVPLT

   Base saturation headways (h_{basei}) are obtained from Table 25.10. Adjustment factors (h_j) are found in Table 25.11. Values for P_{LT}, P_{RT}, and P_{HV} are given in the problem statement, but must be expressed as a decimal for use. From Table 25.10, base saturation headways for Scenario 1 (for both approaches) is 3.9 s/veh (DOC 1, Group 1). For Scenario 2 (for both approaches), the value is 5.8 s/veh (DOC 3, Group 1, 1 vehicle present on conflicting approach). From Table 25.11, $h_{LT}=0.2$, $h_{RT}=-0.6$, and $h_{HV}=1.7$. From the problem statement, $P_{LT}=0.10$ (NB) and 0.00 (WB), $P_{RT}=0.00$ (NB) and 0.05 (WB). The proportion of heavy vehicles, $P_{HV}=0.08$ (NB) and 0.10 (WB). Then:

hadjNB=(0.2×0.10)-(0.6×0.00)+(1.7×0.08)=0.156hadjWB=(0.2×0.00)-(0.6×0.05)+(1.7×0.10)=0.140hsNB1=3.9+0.156=4.056 s/vehhsNB2=

3. Step 4: Determine the Departure Headway for Each Approach

   In this case, there are two scenarios for each subject approach (NB, WB). The departure headways depend upon the degree of saturation (X) for the conflicting approach in each case. The process is iterative, but starts with an assumption that all values of h_d are 3.2 s/veh. Then, using Equation 25-22:

   Xj=vjhdj3600XNB=400×3.23600=0.356XWB=440×3.23

   These values set the probabilities that the conflicting lane is empty or occupied. For the NB approach, it is 0.356 probable that the lane is occupied, and (1−0.356)=0.644 probable that it is empty. For the WB approach, it is 0.391 probable that the lane is occupied, and (1−0.391)=0.609 probable that it is empty. The “empty” states exist for both Scenarios 1, while the “occupied” states exist for both Scenarios 2. Since there is only one conflicting lane to consider in each case, there are no multiple probabilities to multiply. Thus, the probabilities that each scenario exists are as follows:

   PNB1=0.609PNB2=0.391PWB1=0.644PWB2=0.356

   Note that the WB approach is the “conflicting approach” for the NB subject approach, and vice-versa.

   Equations 25-25 must be used to adjust these computations. The probability that DOC 1 exists is the probability of each Scenario 1—the conflicting lane is empty. The probability that DOC 3 exists is the probability of each Scenario 2—one conflicting lane is occupied. The probability of all other DOCs (2, 4, and 5) is 0.0, as none of these can occur.

The adjustment to initial scenario probabilities is estimated using Equations 25-25:

$$AdjPDOC1=0.01\ [PDOC2+2PDOC3+3PDOC4]/1 AdjPD$$
$$NB=0.01\ [0.0+(2\times 0.391)+(3\times 0)]/1=0.00782\ s/veh AdjPDOC1\text{-}WB=0.01\ [0.0+(2\times 0.356)+(3\times 0)]/1=0.00712\ s/veh AdjPDOC3=0.01\ [PDOC4+2PDC$$
$$3PDOC3]/6 AdjPDOC3\text{-}NB=0.01\ [0.0+(2\times 0.0)-(3\times 0.391)]/6=-0.0020\ s/veh AdjPDOC3\text{-}WB=0.01\ [0.0+(2\times 0.0)-(3\times 0.356)]/6=-0.0018\ s/veh$$

Adjustments for DOC 1 are applied to Scenarios 1, while adjustments for DOC 3 are applied to Scenarios 2. Then:

$$P'i=Pi+AdjPi P'NB1=0.609+0.00782=0.6182 P'NB2=0.391-0.00200=0.3890 P'WB1=0.644+0.00712=0.6511 P'WB2=0.356-0.0018=0.3542$$

Departure headways may now be computed using Equation 25-21:

$$hd=\sum iP'ihsi$$
$$hNB=(0.6182\times 4.056)+(0.3890\times 5.956)=2.507+2.317=4.802\ s/veh hWB=(0.6511\times 4.040)+(0.3542\times 5.940)=2.630+2.104=4.734\ s/veh$$

In general, final departure headways are rounded to the nearest 0.1 s/veh. In this case, the NB approach rounds to 4.8 seconds and WB approach rounds to a departure headway of 4.7 s.

This, however, is not the final result. The computed values (4.8 and 4.7s/veh) are quite different from the initial assumed value of h_d (3.2 s/veh). The result must now be iterated until the assumed and computed values agree to within ±0.1. Each successive iteration begins with the results from the previous iteration. For the NB and WB subject approaches, the results of these iterations (each of which follows the same steps as the initial computation) are shown in Table 25.14. As can be seen, closure occurs two iterations past the initial computation.

Table 25.14: Iterated Solutions for Departure Headway for the Sample AWSC Intersection

	Northbound Solution			Westbound Solution		
Iteration No.:	**1**	**2**	**3**	**1**	**2**	**3**
Demand Flow, NB	400	400	400	400	400	400
Demand Flow, EB	440	440	440	440	440	440
Initial h_d	**3.2**	**4.8**	**5.1**	**3.2**	**4.7**	**4.1**
h_{base} (*Scenario* 1)	3.9	3.9	3.9	3.9	3.9	3.9
h_{base} (*Scenario* 2)	5.8	5.8	5.8	5.8	5.8	5.8
h_{LT}	0.20	0.20	0.20	0.20	0.20	0.20
h_{RT}	−0.60	−0.60	−0.60	−0.60	−0.60	−0.60
h_{HV}	1.70	1.70	1.70	1.70	1.70	1.70
P_{LT}	0.10	0.10	0.10	0.00	0.00	0.00
P_{RT}	0.00	0.00	0.00	0.05	0.05	0.05
P_{HV}	0.08	0.08	0.08	0.10	0.10	0.10
h_{adj}	0.156	0.156	0.156	0.140	0.140	0.140
h_s (*Scenario* 1)	4.056	4.056	4.056	4.040	4.040	4.040
h_s (*Scenario* 2)	5.956	5.956	5.956	5.940	5.940	5.940
X_j	0.391	0.579	0.507	0.356	0.535	0.565
P (*Scenario* 1)	0.609	0.421	0.493	0.644	0.465	0.435
P (*Scenario* 2)	0.391	0.579	0.507	0.356	0.535	0.565
P (*DOC* 1)	0.609	0.421	0.493	0.644	0.465	0.435
P (*DOC* 3)	0.391	0.579	0.507	0.356	0.535	0.565
AdjP (*DOC* 1)	0.008	0.008	0.010	0.007	0.011	0.011
AdjP (*DOC* 3)	−0.002	−0.017	−0.015	−0.002	−0.161	−0.170
P' (*Scenario* 1)	0.617	0.430	0.503	0.652	0.475	0.446
P' (*Scenario* 2)	0.389	0.561	0.492	0.354	0.375	0.396
h_d	**4.8**	**5.1**	**5.0**	**4.7**	**4.1**	**4.2**

Table 25.14: Full Alternative Text

The final result is that the departure headway (h_d) for the

NB approach is 5.1 s/veh, and the departure headway for the WB approach is 4.1 s/veh.

4. Step 5: Determine the Capacity of Controlled Approaches

    At first glance, the answer to this question appears to be obvious: If the departure headways for the NB and WB approaches are 5.1 and 4.1 s/veh, respectively, the capacity of each approach should be:

    $$c_{NB} = \frac{3600}{5.1} = 706 \text{ veh/h} \quad c_{WB} = \frac{3600}{4.1} = 878 \text{ veh/h}$$

    Unfortunately, that approach does not take into account the interdependence of the two approaches. With a higher volume on the subject approach (capacity), the departure headways of *both* approaches are affected. Thus, the problem is once again iterative.

    Table 25.14 shows the results of the three iterations required to determine the departure headways for the two approaches. Now, the demand flow rate in each approach (separately) is incrementally increased—while keeping the demand on the conflicting approach constant. For each demand flow rate, a new set of iterations are needed to produce a departure headway. The demand flow rate on the subject approach is increased until the resulting degree of saturation (vhd/3600) becomes 1.000. This is now an iteration of individual solutions, each of which is itself iterative. Obviously, we cannot show all of these computations. Suffice it to say that each iteration produces a table like Table 25.14, and iterations continue until the degree of saturation reaches 1.000.

    For the sample problem, the following capacities are determined in this way:

    $$c_{NB} = 722 \text{ veh/h} \quad c_{WB} = 863 \text{ veh/h}$$

    In this case, the final capacities of the two controlled approaches are somewhat larger than the simple approach suggested. This will be the case in most solutions.

5. Step 6: Determine Control Delay and LOS for Each Approach

   [Equation 25-27](#) is used to estimate the average control delay on each approach:

   $$d_x = t_{sx} + 900T\left[(X_x-1)+\sqrt{(X_x-1)^2+\frac{h_{dx}X_x}{450T}}\right]+5$$

   In each case, values are taken from the third iteration of the solution in [Table 25.14](#). Remember that the service time, t_{sx}, for each case is the departure headway minus the move-up time, which has a default value of 2.0 s/veh for Geometry Group 1. Then:

   $$t_{sNB}=4.5–2.0=2.5 \text{ s/veh} \quad t_{sWB}=4.5–2.0=2.5 \text{ s/veh} \quad T=0.25\text{h}$$

   and:

   $$d_{NB}=2.5+900\times0.25\left[(0.567-1)+\sqrt{(0.567-1)^2+\frac{5.1\times0.567}{450\times0.25}}\right]+5=13.4 \text{ s/veh} \quad d_{WB}=2.5+900\times0.\ldots1)+\sqrt{(0.501-1)^2+\frac{4.1\times0.501}{450\times0.25}}\right]+5=12.0 \text{ s/veh}$$

   From [Table 25.7](#), both approaches operate at LOS B. This is clearly acceptable, and the AWSC intersection is expected to operate well.

Part III Roundabouts

The roundabout is a relatively new form of intersection for the United States, but one that is growing rapidly over time. In 1995, there were approximately a dozen true roundabouts in the United States; by 2016, the number was approaching 5,000, with no slowing in sight. As was noted in the introduction, roundabouts are not the same as traffic circles or rotaries, which have existed throughout the United States for many years. In traffic circles and rotaries, entering vehicles have the right-of-way over circulating vehicles; in roundabouts, circulating vehicles have the right-of-way, and all entry roadways are controlled with YIELD signs and markings.

The first known traffic circle in the United States was built in 1905 in New York City—Columbus Circle. It continues to operate today, but is assisted by traffic signals. Because traffic circles give entering vehicles the right-of-way, high-speed entries are encouraged. By 1950, it had become apparent that traffic circles were experiencing high crash rates and congestion. The modern roundabout was developed in the United Kingdom, where in 1966 a mandatory “give way” rule was enacted for all traffic circles—requiring that entering traffic give way to circulating traffic. The benefits of such operation were soon apparent, and adoption throughout Europe, other continents, and (finally) the United States followed [9].

The principal benefits of roundabouts over other forms of intersection control involve primarily safety. Compared to signalized intersections, benefits also include less delay to vehicles, improved safety for pedestrians and bicyclists (as well as motorists), and sometimes higher capacities. A study of 11 U.S. intersections where signals or other controls were replaced with roundabouts, a reduction of 37% in total accidents, 51% in injury/fatal accidents, and 29% in “property damage only” (PDO) accidents resulted [9].

While there is general agreement that roundabouts generally improve safety and reduce delays, they remain controversial, as many U.S. drivers have difficulty negotiating them. There are also counterexamples in which the general benefits have not been effectively realized.

One of the most controversial roundabouts in the nation was installed at the intersection of Jacaranda Boulevard and Venice Avenue in Venice, Florida, in 2011. It almost immediately became the high-crash intersection in all of Sarasota County. As a signalized intersection, a total of 11 accidents occurred in 2008 and 2009. After its installation, there were 52 accidents in 2012, 57 in 2013, and 50 in 2014. There have been ancillary impacts as well. As a signalized intersection in 2007, 60,000 vehicles/day traversed the intersection; since 2011, traffic through the intersection has decreased to 33,000 vehicles/day [10, 11]. Drivers now try to find alternative routes to avoid traversing the roundabout, increasing volumes at other nearby intersections.

The intersection has many unique aspects that undoubtedly contribute to this experience. Venice has a large senior population, with an average age of 67. The intersection feeds a major interchange to I-75, which generates high truck demand. Because of the proximity to Gulf beach areas, there are many unfamiliar drivers using the intersection. In 2015 and 2016, significant design modifications were introduced at this roundabout to mitigate some of the accident problems. At this writing, no new crash statistics are available to assess the success of the changes.

Roundabouts will continue to grow in usage, as most installations do achieve the anticipated benefits. It is wise to consider all of the characteristics of a particular location, as some cases can run counter to the general trends.

25.6 Types of Roundabouts and General Characteristics

The *AASHTO Green Book* [12] defines three categories of roundabouts: mini-roundabouts, single-lane roundabouts, and multilane roundabouts. It should be noted, that for this text, “multilane” roundabouts consist of two circulating lanes. There is little experience with roundabouts having more than two circulating lanes. Table 25.15 summarizes the key characteristics of each of these types of roundabouts.

Table 25.15: Key Characteristics of Roundabouts

Characteristic	Mini-Roundabout	Single-Lane Roundabout	Multilane Roundabout
Maximum Entry Design Speed	15−20 mi/h	20−25 mi/h	25−30 mi/h
Maximum Number of Entering Lane Per Approach	1	1	2+
Typical Inscribed Circle Diameter	45−90 ft	90−150 ft	140−250 ft
Central Island Treatment	Mountable	Raised	Raised
Typical Daily Volumes on 4-Leg Roundabout	0−15,000 veh/day	0−20,000 veh/day	20,000+ veh/day

(Source: Adapted from *A Policy on Geometric Design of Highways and Streets*, 6th Edition, American Association of State Highway and Transportation Officials, Washington, D.C., 2011.)

Table 25.15: Full Alternative Text

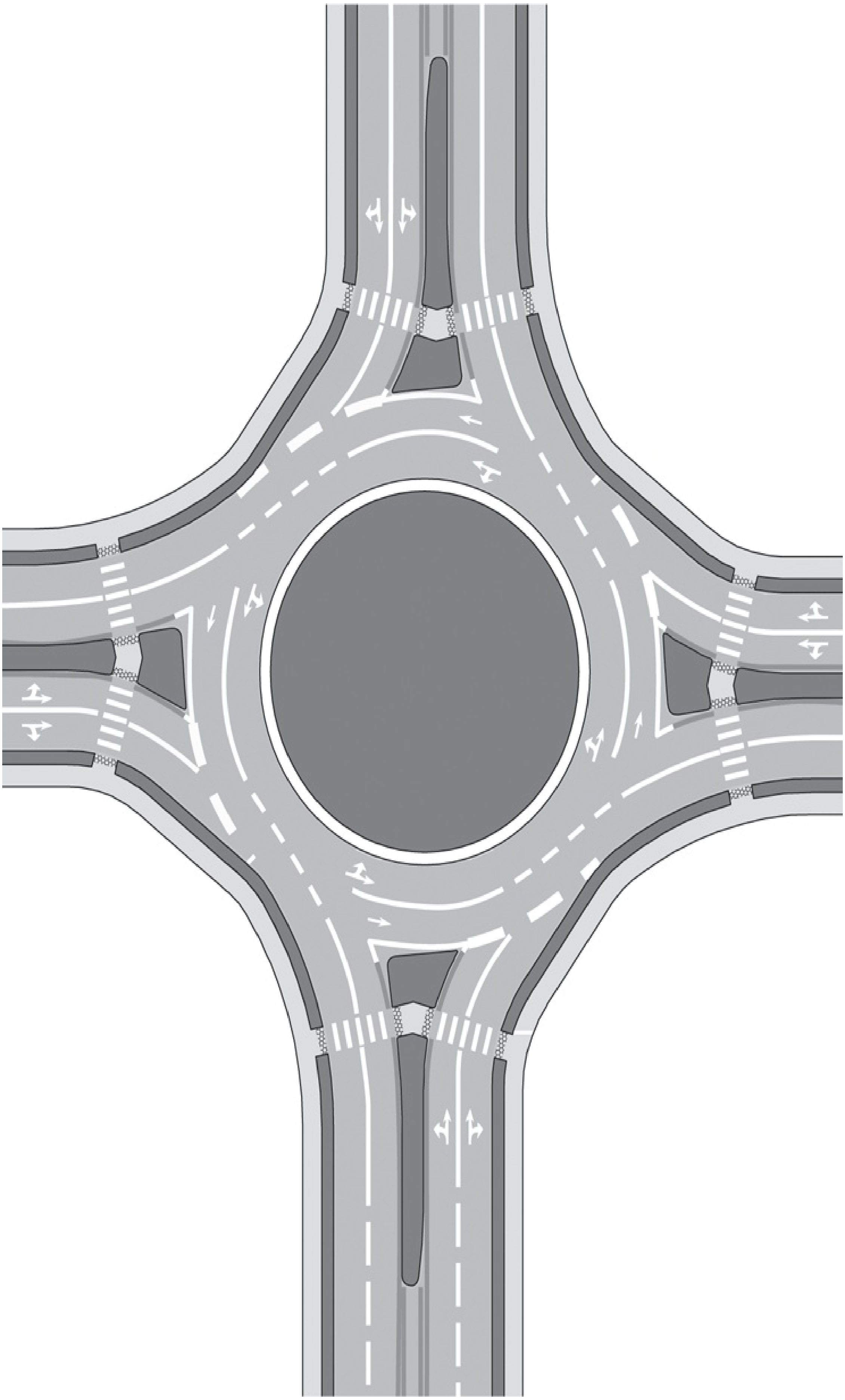
Note that all forms are designed for relatively low-speed entries, which is one of the major reasons that safety is generally improved. Mini-roundabouts are relatively small, and are generally used at intersections of local roads in residential areas. Mini-roundabouts are limited to one lane circulation. For major intersections with high demands (>20,000 veh/day), the multilane roundabout is the option most often selected. Details of capacity and operations, however, are related to peak-hour flows, which must be carefully examined.

25.7 Signing and Marking for Roundabouts

Critical to the safe and efficient operation of roundabouts is proper signing and marking. It is of the utmost importance that drivers recognize what they must do to execute their desired maneuver through the roundabout.

The MUTCD [1] contains many examples of the application of signing and marking to roundabouts. The most complex situation involves two-lane roundabouts with two-lane entry roadways. A typical set of markings is illustrated in Figure 25.7.

Figure 25.7: Typical Markings for a two-Lane Roundabout with two-Lane Entries



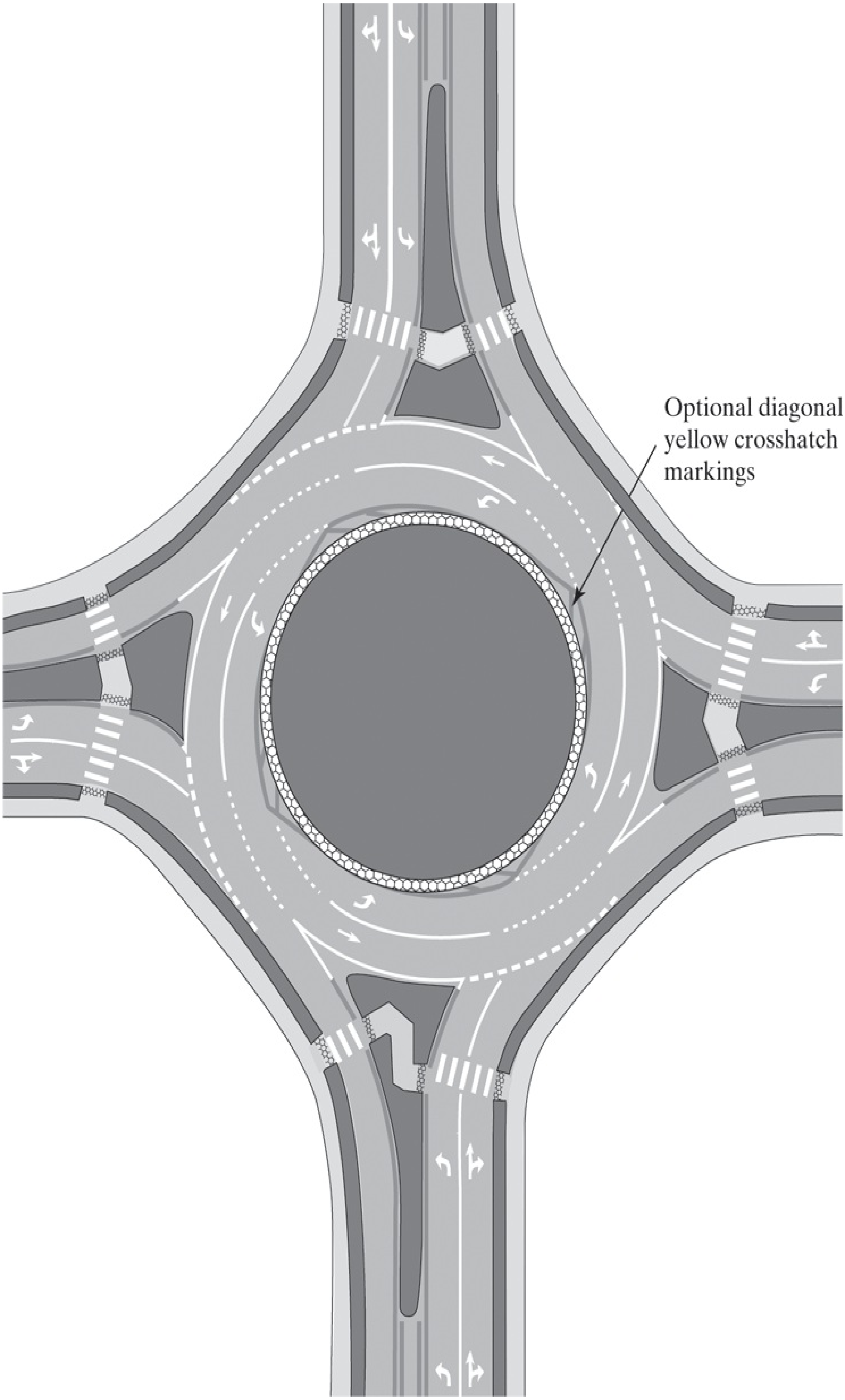
(Source: *Manual of Uniform Traffic Control Devices (MUTCD)*, Federal Highway Administration, Washington, D.C., 2009, as amended through May 2012, Figure #C-6, pg 404.)

Figure 25.7: Full Alternative Text

In Figure 25.7, two-lane exits are also permitted. This presents drivers who are entering the roundabout with the possibility that they may have to cross the path of exiting vehicles in *both* circulating lanes. In two-lane roundabouts, this is always the most significant conflict, and one that drivers have the most difficulty in executing. The short broken lane line designates the only areas in which drivers may cross the circulating lane line.

There are many options in roundabout design. A typical roundabout can have one or two circulating lanes, one- or two-lane entry roadways (or some combination of both), and one- or two-lane exit roadways (or some combination of both). Figure 25.8 shows a two-lane roundabout with all two-lane entry roadways, but one-lane exit roadways. Some of the crossing conflicts are eliminated, but in this scheme, drivers must know what lane they have to be in at all times, and where they have to enter their desired lane.

Figure 25.8: Typical Markings for a Two-Lane Roundabout with two-Lane Entries and one-Lane Exits


Optional diagonal
yellow crosshatch
markings

Note: The marking configuration shown on this figure requires U-turning drivers to change lanes within the circulatory roadway.

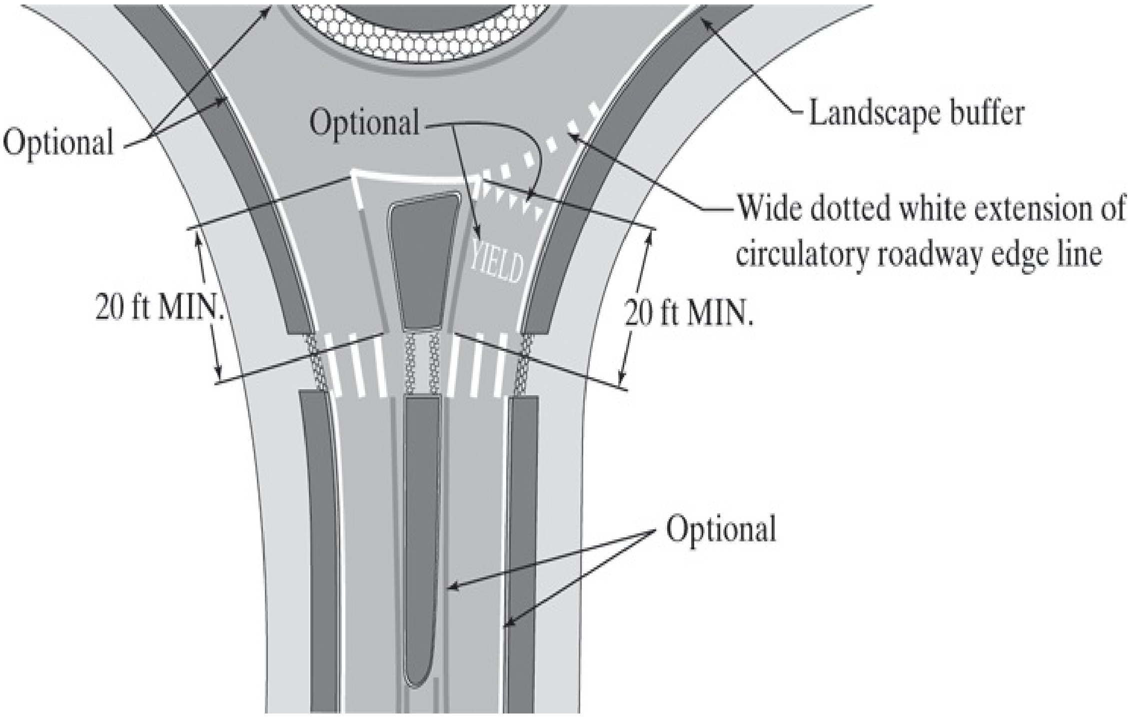
(Source: *Manual of Uniform Traffic Control Devices (MUTCD)*, Federal Highway Administration, Washington, D.C., 2009, as amended through May 2012, Figure 3C-5, pg 405.)

Figure 25.8: Full Alternative Text

In this case, vehicles in the right circulating lane must exit at each exit roadway. To continue, they must move into the left lane, which becomes the right lane after passing the exit roadway. This reduces conflicts somewhat, but requires greater vigilance on the part of the driver to be in the appropriate lane at all times.

Figure 25.9 shows the typical markings for an approach roadway to a roundabout. The word YIELD is often included with the standard YIELD line to emphasize that entering traffic must yield to circulating traffic. These markings, of course, supplement a YIELD sign, which must be posted on each entering roadway. Crosswalks are a minimum of 20 feet from the edge of the circulating roadway, so that pedestrian conflicts do not occur at the same time (and in the same place) that drivers are making a lane or exit decision.

Figure 25.9: Typical Markings for an Entry Roadway to a Roundabout



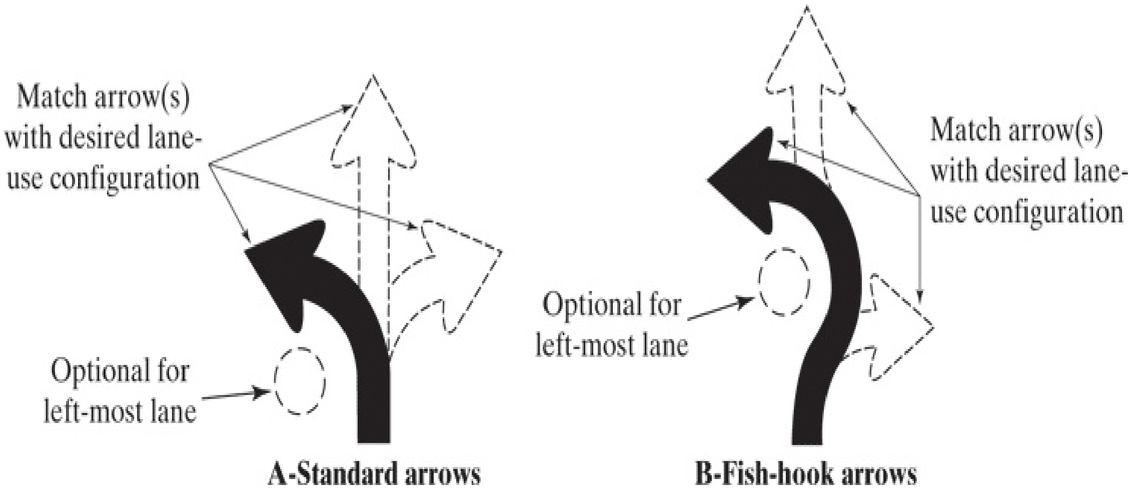
(Source: *Manual of Uniform Traffic Control Devices (MUTCD)*, Federal Highway Administration, Washington, D.C., 2009, as amended through May 2012, Figure 3C-1, pg 399.)

Figure 25.9: Full Alternative Text

The MUTCD contains numerous additional illustrations and discussions of roundabout markings, which may be accessed directly. Markings are extremely important to the safe and efficient operation of roundabouts. Drivers need clear and concise guidance on what paths to follow to achieve their desired maneuvers.

Two types of signs are critical to safe and efficient operation of roundabouts: YIELD signs and lane use control signs. As noted previously, the MUTCD *requires* that YIELD signs be posted on each roadway entering the roundabout. Lane use control signs are also very important. As the roundabout is approached, lane use control signs should clearly indicate how approaching vehicles should complete each of the four potential maneuvers: U-turn, left turn, through, and right turn. This is accomplished by posting one or more lane use control signs for each entering lane. Two-lane entries may have separate signs for each lane, or may show use of both lanes on one sign. These signs are typical regulatory signs using graphic arrows to indicate the proper use of each lane. The arrow symbols used are shown in Figure 25.10.

Figure 25.10: Lane Use Control Arrows for Use on Roundabout Approaches



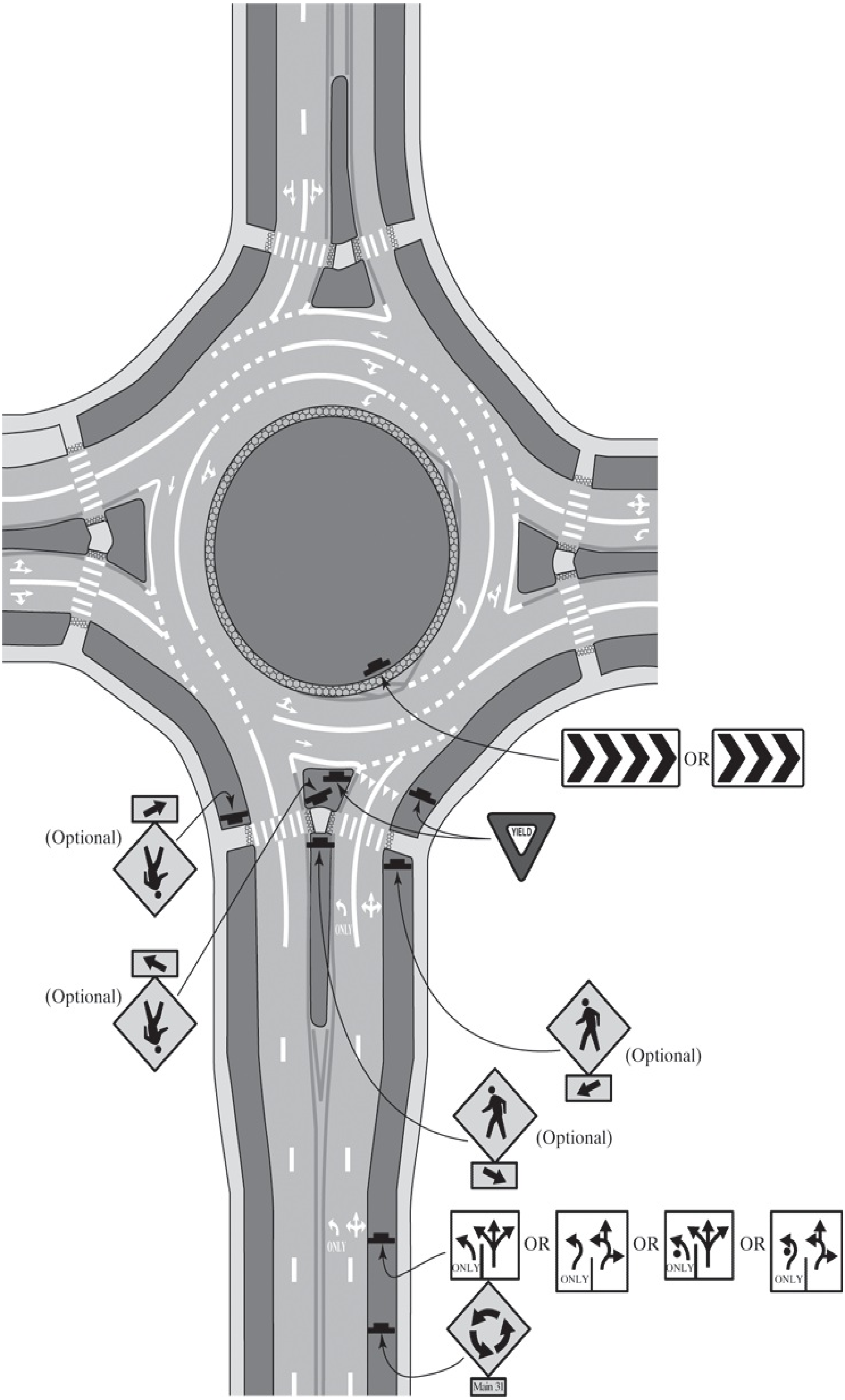
(Source: *Manual of Uniform Traffic Control Devices (MUTCD)*, Federal Highway Administration, Washington, D.C., 2009, as amended through May 2012, Figure 2B-5, pg 62.)

Figure 25.10: Full Alternative Text

At each crosswalk, a pedestrian crossing warning sign should also be used. The MUTCD classifies these as optional signs, but wherever there are more than occasional pedestrians, it is wise to include them.

Figure 25.11 depicts the typical signing for one roundabout approach with two lanes. The specifics of signing, of course, must be coordinated with the specifics of the roundabout geometry.

Figure 25.11: Typical Signing for a Two-Lane Roundabout Entry Roadway


OR
YIELD
(Optional)
(Optional)
(Optional)
(Optional)
ONLY
ONLY
OR
OR
OR
Main 31

Notes

1. Signs shown for only one leg
2. See Section 2D.38 (of the MUTCD) for guide signs at roundabouts
3. See Chapter 3C (of the MUTCD) for markings at roundabouts

(Source: *Manual of Uniform Traffic Control Devices (MUTCD)*, Federal Highway Administration, Washington, D.C., 2009, as amended through May 2012, Figure 2B-23, pg 87.)

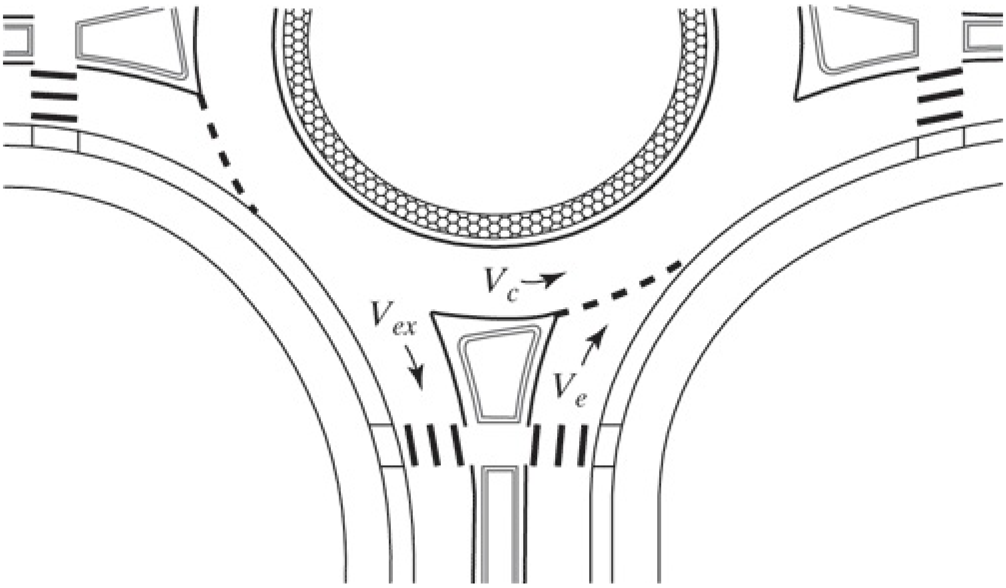
Figure 25.11: Full Alternative Text

Because of the complexity of roundabouts, and the significant number of alternative geometries that may exist, it is impossible to cover all signing and marking possibilities here. The key point is that the geometry, marking, and signing must be carefully coordinated and clearly indicated, allowing drivers to safely achieve their desired maneuver.

25.8 Capacity and Level of Service Analysis of Roundabouts

The 2016 HCM [7] provides a detailed methodology for estimating the capacity and LOS of roundabouts. It covers one- and two-lane roundabouts with one- and two-lane entry and exit roadways. It *does not* cover three-lane roundabouts or entry/exit roadways. It focuses on each entering roadway, and leads to an estimate of the capacity of each entry lane and the average control delay to vehicles. LOS is based upon control delay. The procedure also checks the capacity of each exit roadway to ensure that queuing onto the roundabout itself is not occurring. Figure 25.12 illustrates the three demand parameters that are analyzed for each roadway approach.

Figure 25.12: Critical Demand Flows for Analysis of Roundabout Approaches



(Source: Reprinted with permission from *Highway Capacity Manual, 6th Edition – A Guide for Multimodal Mobility*

Analysis, Transportation Research Board, the National Academy of Sciences, Courtesy of the National Academies Press, Washington, D.C., 2016.)

Figure 25.12: Full Alternative Text

The methodology is conducted in *passenger car equivalent* (pce) units. It, therefore, includes an adjustment of demand flows to reflect heavy vehicle presence. It also works with demand flow rates representing a peak 15-minute period. This may or may not involve an adjustment for the PHF, depending upon whether or not demand flows were established as hourly volumes or actual 15-minute flow rates.

The most important issue is determining the circulating flow rate (v_c) and exiting flow rate (v_{ex}) at each junction. These are calculated from the desired maneuvers of entering drivers from each approach. For example, assuming that the approach shown in Figure 25.12 is the NB approach, the circulating traffic at that point would include the following movements: WB U-Turn, SB LT, SB U-Turn, EB-TH, the EB-LT, and the EB U-Turn. The exiting flow rate would include: EB RT, SB TH, WB LT, and the NB U-Turn. Table 25.16 shows the equations used to calculate the circulating and exiting flow rates at each approach of a four-leg roundabout.

Table 25.16: Calculating Circulating and Exiting Flow Rates at Each Approach of a 4-Leg Roundabout

Approach	Circulating Flow Rate, v_c	Exiting Flow Rate, v_{ex}
NB	$v_{cNB} = v_{WBU} + v_{SBL} + v_{SBU} + v_{EBT} + v_{EBL} + v_{EBU}$	$v_{exNB} = v_{EBR} + v_{SBT} + v_{WBL} + v_{NBU}$
SB	$v_{cSB} = v_{EBU} + v_{NBL} + v_{NBU} + v_{WBT} + v_{WBL} + v_{WBU}$	$v_{exSB} = v_{WBR} + v_{NBT} + v_{EBL} + v_{SBU}$
EB	$v_{cEB} = v_{NBU} + v_{WBL} + v_{WBU} + v_{SBT} + v_{SBL} + v_{SBU}$	$v_{exEB} = v_{SBR} + v_{WBT} + v_{NBL} + v_{EBU}$
WB	$v_{cWB} = v_{SBU} + v_{EBL} + v_{EBU} + v_{NBT} + v_{NBL} + v_{NBU}$	$v_{exWB} = v_{NBR} + v_{EBT} + v_{SBL} + v_{WBU}$

Table 25.16: Full Alternative Text

Note that all of these computations would be carried out using flow rates in pce/h—that is, *after* conversions for heavy vehicle presence and peaking have been made.

Computational procedures to determine approach roadway capacity and LOS are presented in the steps that follow.

1. Step 1: Convert Movement Demand Volumes to Flow Rates

   As was the case for AWSC intersections, hourly demand volumes are converted to flow rates representing the worst 15 minutes in the hour using Equation 25-1:

   $$v_i=\frac{V_i}{PHF}$$

   where all terms are as previously defined.

   It is preferable to measure flow rates directly. This requires that demands be observed and recorded in 15-minute intervals. The worst combination(s) of flow rates would determine the choice of the analysis period(s). If only hourly demands were recorded, then all are divided by a single average PHF for the intersection. This computation is a worst case, assuming that all movements peak in exactly the same 15 minutes.

2. Step 2: Convert Demand Flow Rates to Passenger Car Equivalents

   The roundabout methodology does account for heavy vehicle presence, but in a very general way. Passenger car equivalents for *all* heavy vehicles are 2.0 pce/heavy vehicle. Then:

   $$v_{pce}=\frac{v}{f_{HV}} \quad f_{HV}=\frac{1}{1+P_T(E_T-1)}=\frac{1}{1+P_T(2-1)}=\frac{1}{1+P_T} \quad [25\text{-}28]$$

   where:

v =demand flow rate, veh/h, v pce
=demand flow rate, pc/h, E T
=passenger car equivalent for heavy vehicles, 2 pce/heavey
P T
=proportion of heavy vehicles in the traffic stream, decima

For general purposes, any vehicle with more than four wheels on the ground would be considered a “heavy vehicle.” This would include most trucks (except for four-wheel pick-up trucks), buses, cars with trailers, and recreational vehicles.

3. Step 3: Determine Circulating and Exiting Flow Rates

   Using the equations of Table 25.16, all circulating and exiting flow rates would be computed for all approaches to the roundabout.

4. Step 4: Determine Entry Flow Rates by Lane

   If there is only one entry lane, then all entry flows (L, TH, R) are assigned to that lane. If there are two entry lanes, the entry demand flow must be assigned to the appropriate lane. In general, left turns are assumed to enter in the left lane, right turns are assumed to enter in the right lane, and through vehicles are distributed to the two lanes in accordance with observed or assumed patterns. There are, however, five different distributions that might occur:

   1. L/TR: In some cases, there are so many left turns that the left lane becomes a de facto left-turn lane; in this case all through and right turns are in the right lane. This condition occurs when $v_U+v_L>v_T+v_R$.

   2. LT/R: In some cases there are so many right turns that the right lane becomes a de facto right-turn lane; in this case all through vehicles and left turns are in the left lane. This condition occurs when $v_R>v_U+v_T+v_L$.

3. LT/TR: This is the usual situation. Neither a de facto left-turn lane nor right-turn lane exists. Through vehicles are assigned to a lane in accordance with field observations or assumptions. As a default condition, through vehicles may be assigned to equalize the total lane flows in both lanes.

4. L/LTR: This is a fairly rare case that may occur where the left turn demand is significantly greater than the through and right-turn demand. Field observations are preferred, but as a default, left turns can be assigned to equalize total flow in both lanes.

5. LTR/R: This is another fairly rare case that may occur where the right turn demand is significantly greater than the through and left-turn demand. Field observations are preferred, but as a default, right turns can be assigned to equalize total flow in both lanes.

It should be noted that some roundabouts include physically channelized right-turn bypass lanes. In such cases, right turns using the bypass lane are *not included* in the flow rates entering the roundabout.

5. Step 5: Determine the Capacity of Each Entry Lane and Each Bypass Lane as Appropriate in Passenger Car Equivalents

The capacity of each entry is estimated using a series of equations shown in Table 25.17. The equation used depends upon:

Table 25.17: Equations for Estimating the

Capacity of Roundabout Entry Roadways

Number of Entry Lanes	One-Lane Roundabout	Two-Lane Roundabout
1	$c = 1{,}130e^{(-0.001\,v_c)}$	$c = 1{,}130e^{(-0.0007\,v_c)}$
2	$c_R = 1{,}130e^{(-0.001\,v_c)}$ $c_L = 1{,}130e^{(-0.001\,v_c)}$	$c_R = 1{,}130e^{(-0.0007\,v_c)}$ $c_L = 1{,}130e^{(-0.00075\,v_c)}$

c=capacity of single entry lane;
cR=capacity of right entry lane (two
−lane entry); cL=capacity of left entry lane (two
−lane entry);
vc=circulating traffic flow. All units pce/h.

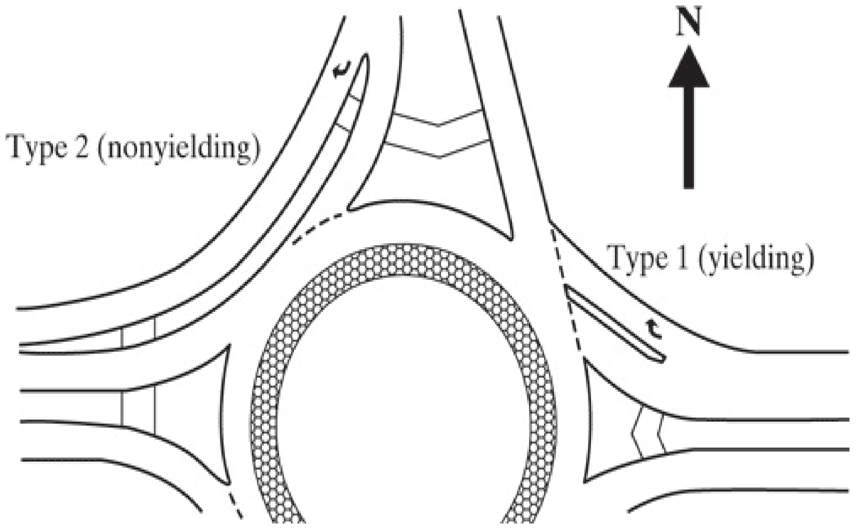
Table 25.17: Full Alternative Text

- The number of entry lanes (1, 2), and
- The number of circulating lanes (1, 2).

The capacity of any existing right-turn bypass lanes must also be estimated. There are two types of bypass lanes that could exist: nonyielding and yielding. Figure 25.13 illustrates these types of bypass lanes.

Figure 25.13: Yielding vs. Non-Yielding Right-Turn Bypass Lanes at a

Roundabout



(Source: Reprinted with permission from *Highway Capacity Manual, 6th Edition – A Guide for Multimodal Mobility Analysis,* Transportation Research Board, the National Academy of Sciences, Courtesy of the National Academies Press, Washington, D.C., 2016.)

Figure 25.13: Full Alternative Text

For nonyielding bypass lanes, right turns merge into the exiting traffic stream well downstream of the roundabout exit itself. Yielding lanes merge into traffic as it exits the roundabout. Because of proximity, these right turns would still be controlled using a YIELD sign. The capacity of nonyielding right-turn bypass lanes has not been studied in the United States; thus, there are no estimating equations included for this type of lane. Its capacity is expected to be high, as the merge is between two streams moving at similar speeds. For analysis purposes, the total downstream traffic flow (after the merge) should be considered in conjunction with the geometry and traffic control of the downstream arterial.

For yielding bypass lanes, capacity is based upon the number of exiting lanes fed by the roundabout and bypass lane. Then:

cbypass=1,130 e(-0.001 vex) for 1 exit lanecbypass=1,13(
[25-29]

where:

c bypasss =capacity of the bypass lane, pce/h, and v ex =demand flow of exiting traffic, pce/h.

Pedestrians may have an impact on the capacity of entry roadways. In most cases, pedestrians walk through queued vehicles waiting to enter the roundabout, and have little impact on operations. In some cases, with significant pedestrian flows and relatively low vehicular flow, the effect can be more significant. The HCM provides an adjustment factor (f_{ped}) that can be applied to approach lane capacities to estimate the impact. Table 25.18 provides equations to estimate this factor.

Table 25.18: Adjustment Factor for Pedestrian Interference at Roundabouts

Case	Equation
Single-Lane Entry Roadways	
$v_c > 881$ pce/h	$f_{ped} = 1.00$
$v_c \leq 881$ pce/h and $n_{ped} \leq 101$ ped/h	$f_{ped} = 1 - 0.000137 n_{ped}$
$v_c \leq 881$ pce/h and $n_{ped} > 101$ ped/h	$f_{ped} = \dfrac{1{,}119.5 - 0.715 v_c - 0.644 n_{ped} + 0.00073 v_c n_{ped}}{1{,}086.6 - 0.654 v_c}$
Two-Lane Entry Roadways	
$n_{ped} < 100$ ped/h	$f_{ped} = 1 - \dfrac{n_{ped}}{100}\left[1 - \dfrac{1{,}260.6 - 0.329 v_c - 38.1}{1{,}380 - 0.5 v_c}\right]$ $f_{ped} \leq 1.00$
$n_{ped} \geq 100$ ped/h	$f_{ped} = 1 - \dfrac{n_{ped}}{100}\left[1 - \dfrac{1{,}260.6 - 0.329 v_c - 0.381 n_{ped}}{1{,}380 - 0.5 v_c}\right]$ $f_{ped} \leq 1.00$

nped=number of crossing pedestrians per hour in conflicting crosswalk.

Table 25.18: Full Alternative Text

There is little information available regarding the capacity of exit lanes themselves. The Federal Highway Administration recommends that 1,200 veh/h be used as rough default capacity for an exit lane. Exit flows in excess of this may indicate the need for a two-lane exit roadway.

The capacity that is computed using these equations is in units of pce/h. Normally, these would be converted to capacities in veh/h, using the appropriate heavy-vehicle adjustment factor, f_{HV}. The HCM does recommend this, and provides a formula for use of weighted-average adjustment factors where heavy vehicle presence varies for each movement using the roundabout. For this text, we will use capacity in pce/h, and compare these capacities to demand flow rates, which have already been converted to pce/h.

6. Step 6: Estimate the Average Control Delay in Each Approach Lane and Determine LOS

   Average control delay for each approach lane is estimated as:

   d=3600c+900 T [(X−1)+(X−1)2+(3600c) X450 T]+[5×min (X,1)] [25-30]

   where:

   d =average control delay, s/veh, c
   =capacity of approach lane, pce/h, X
   =v/c for approach lane, and T
   =analysis period (default=0.25 h).

   As for TWSC and AWSC intersections, delay is averaged over the lanes of each approach to obtain approach delay, and may be averaged over all entry lanes to obtain a total average delay. Note that the latter does *not* consider delay to vehicles while circulating in the roundabout.

   LOS is assigned using the same criteria used for TWSC and AWSC intersections, which is shown in Table 25.1.

Sample Problem 25-3: Analysis of a Roundabout

It is proposed to replace an existing intersection of Main Street (a four-lane arterial) and Franklin Road (a two-lane arterial) with a one-lane or two-lane roundabout. The intersection is currently controlled using STOP signs on Franklin Road, and significant delays and queuing have been observed. Determine the entry capacities of a one-lane and two-lane roundabout for this situation, and compare the resulting LOS. Which (if either) of these options should be implemented? The current intersection is shown in Figure 25.14.

Figure 25.14: Sample Problem in Analysis of Roundabouts

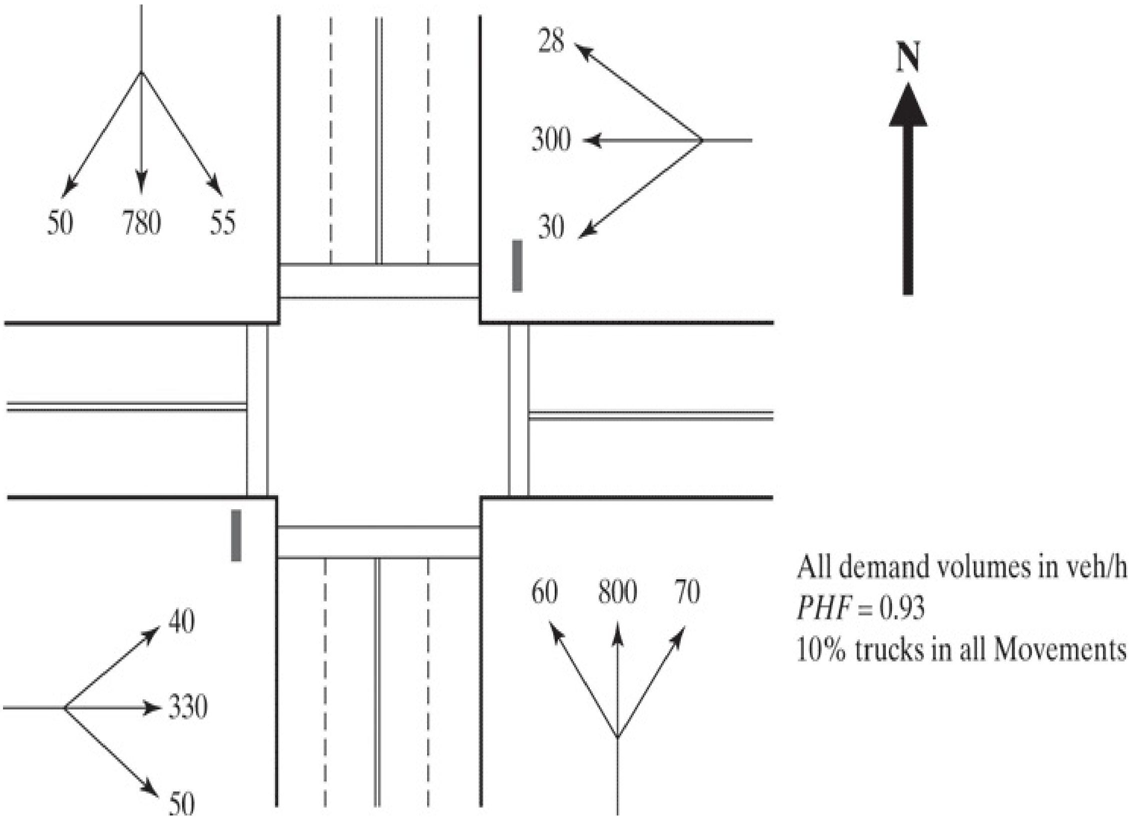


Figure 25.14: Full Alternative Text

The following assumptions are made to facilitate this analysis: (a) the two roundabout options have completely one-lane or completely two-lane circulation; (b) the number of entering lanes will be the same as the number of lanes entering the current intersection.

1. Steps 1 and 2: Convert Demand Volumes to Flow Rates

in pce/h

These two steps are easily combined. All movements share a common PHF and a common truck presence of 10%. Combining the two computations:

$v=VPHF\times fHV$

where, using Equation 25-28:

$fHV=11+PT=11+0.10=0.909$

Using the NB through volume as an example:

$vNB=8000.93\times 0.909=946$ pce/h

Table 25.19 summarizes all of the demand flow rates in pce/h.

Table 25.19: Conversion of Demand Volumes to Flow Rates in pce/h

Approach	Movement	V (veh/h)	PHF	f_{HV}	v (pce/h)*
NB	L	60	0.93	0.909	71
	TH	800	0.93	0.909	946
	R	70	0.93	0.909	83
SB	L	55	0.93	0.909	65
	TH	780	0.93	0.909	923
	R	50	0.93	0.909	59
EB	L	40	0.93	0.909	47
	TH	330	0.93	0.909	390
	R	50	0.93	0.909	59
WB	L	30	0.93	0.909	35
	TH	300	0.93	0.909	355
	R	28	0.93	0.909	33

* rounded to the nearest integer

Table 25.19: Full Alternative Text

2. Step 3: Determine Circulating and Exiting Flow Rates for Each Approach

   Using the equations in Table 25.16, the circulating (v_c) and exiting (v_{ex}) flow rates for each approach may be established. Note that the values entered are in pce/h from Table 25.8, and that the results are in units of pce/h. There are no U-turns indicated in the problem statement. Table 25.20 illustrates these computations.

Table 25.20: Circulating and Exit Flow Rates for Each

Approach (pce/h) for the Sample Problem

Approach	Circulating Flow Rate, v_c	Exiting Flow Rate, v_{ex}
NB	$v_{cNB} = v_{WBU} + v_{SBL} + v_{SBU} + v_{EBT} + v_{EBL} + v_{EBU}$ $v_{cNB} = 0 + 65 + 0 + 390 + 47 + 0 = 502$	$v_{exNB} = v_{EBR} + v_{SBT} + v_{WBL} + v_{NBU}$ $v_{exNB} = 59 + 923 + 35 + 0 = 1{,}017$
SB	$v_{cSB} = v_{EBU} + v_{NBL} + v_{NBU} + v_{WBT} + v_{WBL} + v_{WBU}$ $v_{cSB} = 0 + 71 + 0 + 355 + 35 + 0 = 461$	$v_{exSB} = v_{WBR} + v_{NBT} + v_{EBL} + v_{SBU}$ $v_{exSB} = 33 + 946 + 47 + 0 = 1{,}026$
EB	$v_{cEB} = v_{NBU} + v_{WBL} + v_{WBU} + v_{SBT} + v_{SBL} + v_{SBU}$ $v_{cEB} = 0 + 35 + 0 + 923 + 65 + 0 = 1{,}023$	$v_{exEB} = v_{SBR} + v_{WBT} + v_{NBL} + v_{EBU}$ $v_{exEB} = 59 + 355 + 71 + 0 = 485$
WB	$v_{cWB} = v_{SBU} + v_{EBL} + v_{EBU} + v_{NBT} + v_{NBL} + v_{NBU}$ $v_{cWB} = 0 + 47 + 0 + 946 + 71 + 0 = 1{,}064$	$v_{exWB} = v_{NBR} + v_{EBT} + v_{SBL} + v_{WBU}$ $v_{exWB} = 83 + 390 + 65 + 0 = 538$

Table 25.20: Full Alternative Text

3. Step 4: Determine Entry Flow Rates by Lane

   The EB and WB approaches have only one lane, so all entering flow is in that lane. The NB and SB approaches have two lanes, and the demand must be allocated to each lane.

   The first issue is to check whether a de facto left- or right-turn lane exists on these approaches. A de facto left-turn lane exists if vU+vL>vT+vR. Then:

   - NB Approach: 0+71>946+83? No.
   - SB Approach: 0+65>923+59? No.

   Therefore, no de facto left-turn lane exists on either approach.

   A de facto right-turn lane exists if vR>vU+vT+vL. Then:

   - NB Approach: 83>0+946+71? No.

- SB Approach: 59>0+923+65? No.

Therefore, no de facto right-turn lane exists on either approach.

The normal default condition is that all left-turning vehicles (and U-turns) are in the left lane, all right-turning vehicles are in the right lane, and through vehicles are allocated to equalize the total flow in each lane. Then:

- NB Approach: $v_{NB}=71+946+83=1{,}100$ pce/h total

  $v_{NBL}=71+479=550$ pce/h$v_{NBR}=83+467=550$ pce/h

- SB Approach: $v_{SB}=65+923+59=1{,}047$ pce/h total

  $v_{SBL}=65+458=523$ pce/h$v_{SBR}=59+465=524$ pce/h

4. Step 5: Determine the Capacity of Each Entry Lane and Bypass Lane

   Note that no bypass lanes exist in either proposed solution. There will be two capacities computed for each entry lane—one for a one-lane roundabout and the other for a two-lane roundabout. As there are six entry lanes, there will be 12 capacities computed using the equations of Table 25.17. Each will use the appropriate circulating flow rate as computed in Table 25.20. As a sample, the capacity of the EB approach lane to a one-lane roundabout is computed as:

   $c_{EB}=1{,}130\ e^{(-0.001\ v_{cEB})}=1{,}130\ e^{(-0.001\times 1{,}023)}=522$

   Table 25.21 summarizes the results of these capacity computations for all cases.

Table 25.21: Capacities of Entry

Lanes for Sample Problem

Approach	Lane	v_c	$c(1)$*	$c(2)$**
NB	L	502	684	795
	R	502	684	775
SB	L	461	713	818
	R	461	713	800
EB		1023	406	552
WB		1064	390	537

* entry capacity for a one-lane roundabout

** entry capacity for a two-lane roundabout

Table 25.21: Full Alternative Text

None of these capacities are subject to further modification. No pedestrian flows are indicated in the problem statement.

The capacity of exit lanes is roughly estimated to be in the range of 1,200 pce/h. None of the exiting flows for this example (for either a one-lane or two-lane roundabout) exceeds this value. The NB and SB approaches have exiting flows that are near 1,200 pce/h, but these approaches also have two exiting lanes to handle the traffic.

The degree of saturation (X) for each approach lane should be computed to assess whether adequate capacity is available for each approach lane. This is done in Table 25.22.

Table 25.22: Comparison of Demand Flows and Capacity for Entry Lanes Sample Problem

Approach Lane	Demand Flow (pce/h)	One-lane Roundabout		Two-lane Roundabout	
		Capacity (pce/h)	$X = v/c$	Capacity (pce/h)	$X = v/c$
NBL	550	684	0.804	795	0.692
NBR	550	684	0.804	775	0.710
SBL	523	713	0.734	818	0.639
SBR	524	713	0.735	800	0.655
EB	496	406	**1.222**	552	0.899
WB	423	390	**1.085**	537	0.788

Table 25.22: Full Alternative Text

It is clear from Table 25.22 that the one-lane roundabout is not an option. The EB and WB approaches fail, with demand flows in excess of the available capacity. The two-lane roundabout is at least minimally viable, but the *v/c* ratios for EB and WB approaches are still quite high.

At this point, the one-lane roundabout option is eliminated. The LOS of entry lanes to a two-lane roundabout must be examined.

5. Step 6: Determine the Control Delay and LOS for Each Approach Lane

The control delay for each approaching lane is estimated using Equation 25-29. The computation for the EB approach lane is shown, with all results summarized in Table 25.23.

Table 25.23: Control Delay and LOS for Entry Lanes

Approach Lane	Capacity (pce/h)	X	Control Delay d (s/veh)	LOS (Table 25-7)
NBL	795	0.692	17.5	C
NBR	775	0.710	18.7	C
SBL	818	0.639	15.0	C
SBR	800	0.655	15.9	C
EB	552	0.899	44.5	E
WB	537	0.788	31.2	D

Table 25.23: Full Alternative Text

d=3600c+900 T [(X-1)+(X-1)2+(3600c) X450 T]+[5×min (X,1)]dEB=3600552+900×0.25 [(0.899-1)+(0.899-1)2+(3600552)×0.899450×0.25]+(5×0.899)dEB=6.52+225 [(-0.101)+(-0.101)2+0.052]+4.50dEB=6.52+33.44+4.50=44.46 s/ve

All NB and SB lanes operate at LOS C. The lane delays could be averaged to determine an approach delay and LOS, but the results are similar in both cases, and the LOS for the NB and SB approaches would still be C. The EB approach operates at LOS E, while the WB approach is at LOS D. A weighted average (by demand flow) delay for the entire roundabout may be computed as:

d=(550×17.5)+(550×18.7)+(523×15.0)+(524×15.9)+(496*44.5)+(423×31.2)(550+550+523+524+496+423)d=23.3 s/veh (LOS C)

While it can be said that the entire roundabout operates at LOS C, this masks the fact that the EB approach is operating at 89.9% of capacity in LOS E. The WB approach is also experiencing some difficulty at LOS D.

What should be the conclusion? It is clear that the existing TWSC intersection is probably not appropriate for the level of demand that is present. An analysis of the intersection as a TWSC intersection would provide additional insight.

As a roundabout, the high *v/c* ratios (*X*) on the EB and WB approaches push delays to high levels with resulting poor LOS. The most obvious solution is to add a lane to each of these approaches, making each a two-lane approach. This will significantly increase the capacity of these approaches, with concomitant decreases in delay and better LOS.

The intersection could also be considered for signalization. A capacity and LOS analysis of this option would be required to consider this.

25.9 Closing Comments

This chapter has presented material on three major types of “unsignalized” intersections: TWSC intersections, AWSC intersections, and roundabouts. TWSC and AWSC intersections provide levels of right-of-way control where no control is clearly inappropriate, but where signalization is not warranted. Roundabouts are somewhat broader in their application, and are in some cases appropriate substitutes for a signalized intersection, depending upon local conditions.

Chapter 26 provides an overview of what are generally classified as “alternative intersections.” These represent more complex designs that may be used in conjunction with traffic signals or other types of control. The geometry is used primarily to simplify the mix of movements—most often simplifying the way left turns are accommodated—to make the overall operation safer and more efficient.

References

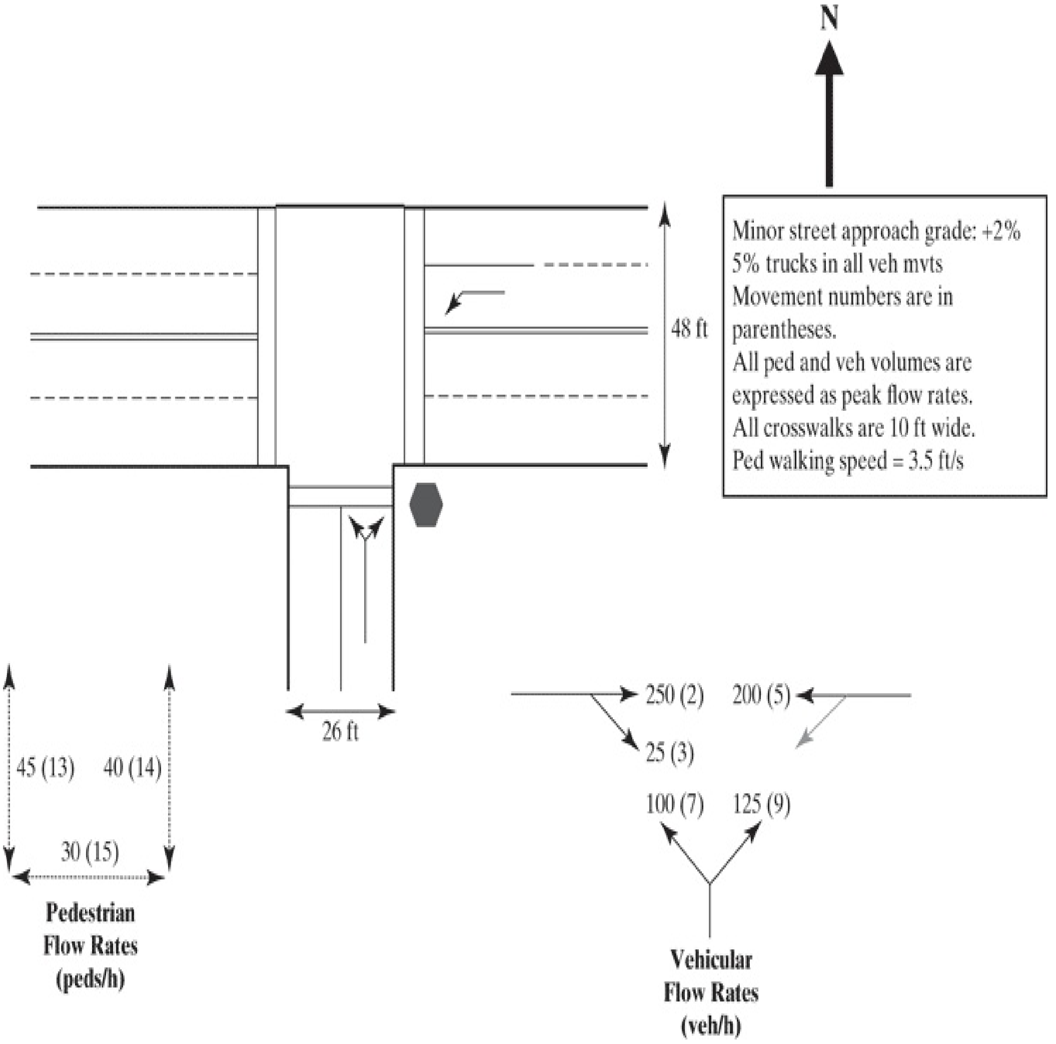
- 1. *Manual of Uniform Traffic Control Devices*, Federal Highway Administration, Washington, D.C., 2009, as amended through May 2012.
- 2. Reprinted with permission from *Highway Capacity Manual, 6th Edition – A Guide for Multimodal Mobility Analysis*, Transportation Research Board, the National Academy of Sciences, Courtesy of the National Academies Press, Washington, D.C., 2016.
- 3. “Interim Materials on Highway Capacity,” *Circular 212*, Transportation Research Board, Washington, D.C., 1980.
- 4. *Capacity of At-Grade Junctions*, Organization for Economic Co-operation and Development, Paris, France, 1974.
- 5. “Merkblatt for Lichtsignalanlagen an Landstrassen, Ausgabe 1972,” *Forschungshasellschaft fur das Strassebwesen*, Koln, Germany, 1972.
- 6. “Highway Capacity Manual,” *Special Report 209*, Transportation Research Board, Washington, D.C., 1985.
- 7. *Highway Capacity Manual, 6th Edition: A Guide for Multimodal Mobility Analysis*, Transportation Research Board, Washington, D.C., 2016.
- 8. Kyte, M., Enas, A., and Troutbeck, R., “The Effect of Upstream Signals on the Capacity of a Downstream TWSC Intersection, *Transportation Research Record 2483*, Transportation Research Board, Washington, D.C., 2015.
- 9. Robinson, B, et al, *Roundabouts: An Informational Guide*, Federal Highway Administration, Contract No. DTFH61-97-R-00038, Washington, D.C., June 2000.
- 10. Womack, C., “Go Around-Go Around,” *Sarasota Herald Tribune*, Sarasota FL, August 16, 2015.

- 11. Kimel, E., “Safety Changes Eyed for Venice Roundabout,” *Sarasota Herald Tribune*, Sarasota FL, March 14, 2016.

- 12. *A Policy on Geometric Design of Highways and Streets*, 6th Edition, American Association of State Highway and Transportation Officials, Washington, D.C., 2011.

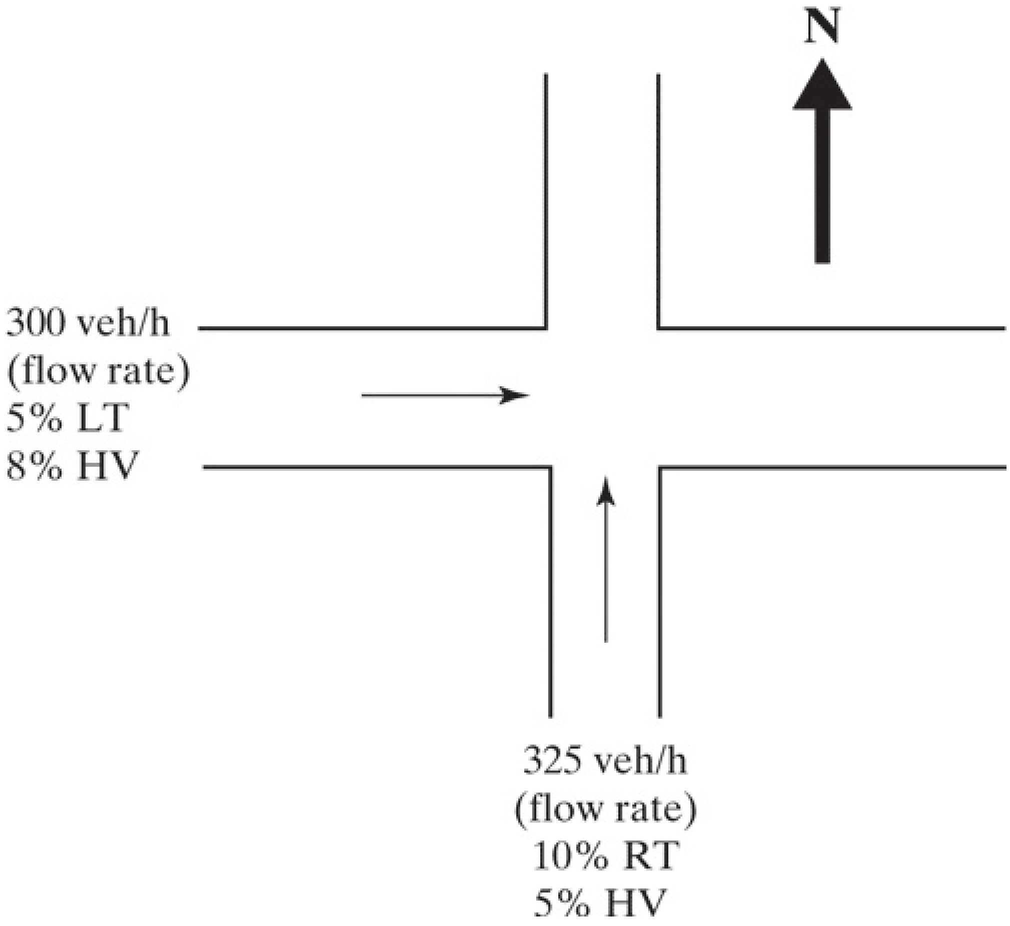
Problems

1. 25-1 Determine the capacity and level of service for the TWSC intersection illustrated below.



   [25.2-27 Full Alternative Text]

2. 25-2 At what level of service would the following AWSC intersection be expected to operate?

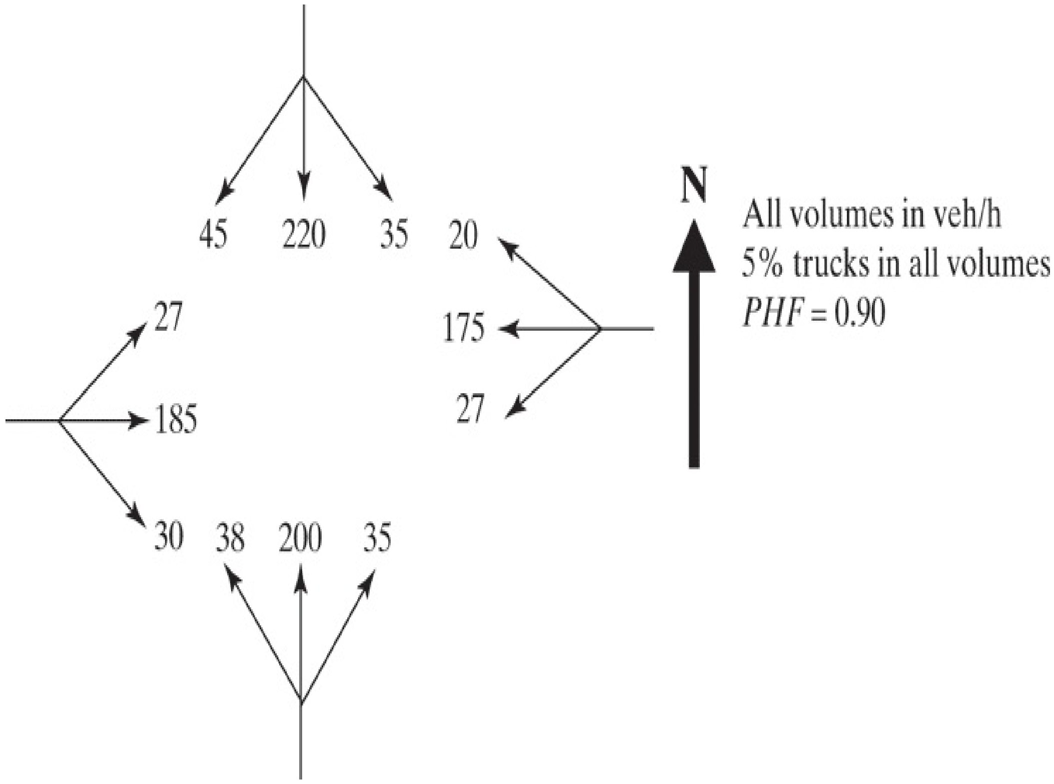


25.2-28 Full Alternative Text

3. 25-3 A one-lane roundabout with four single-lane entry and four single-lane exit roadways is planned to replace a busy intersection. There are few pedestrians and bicycles, so these can essentially be eliminated from the analysis—although the specific design would have to provide adequate safety for them.

   For the demands specified below, will the capacity of the single-lane roundabout be sufficient? At what level of service would the roundabout be expected to operate?

   Without drawing the entire roundabout, what key control devices and safety features would you incorporate into the roundabout design?



[25.2-29 Full Alternative Text]

Chapter 26 Interchanges and Alternative Intersections

A number of chapters of this text deal with the complex subject of intersections, including their design, analysis, and control. Chapters 18–20 deal with the fundamentals of signalized intersections and the timing of traffic signals. Chapter 17 deals with basic geometric elements of intersection design. Chapter 25 presents material on unsignalized intersections, including two-way and all-way STOP controlled intersections and roundabouts. Chapter 23 discusses a simplified critical movement analysis methodology for operational analysis of signalized intersections, while Chapter 22 provides an overview of the 2016 HCM methodology for signalized intersection analysis.

This chapter focuses on interchanges and what are often referred to as “alternative intersections.” An *interchange* is defined as a grade-separated intersection, with many or all turning movement using ramps. Often, one or both of the facilities is a freeway or expressway, but interchanges can exist on any type of highway where an intersection is grade-separated.

Alternative intersections are at-grade intersections with unique designs intended to reduce turning conflicts and/or simplify signalization. They may involve the creation of ancillary intersections and the design of some can spread over a considerable distance.

The basic objective of both interchanges and alternative intersections is to efficiently handle significant turning movements that would otherwise lead to complex signalizations and excessive delay. In effect, alternative geometries are created to reduce turning conflicts.

This chapter provides an overview of the basic geometries in current use, and how the geometries effect traffic flows.

Capacity and level of service analysis of these types of facilities relies on individual methodologies for each portion of an intersection, with the addition of extra travel time that may be experienced due to the geometric design. Although not presented in detail, an outline of this approach is also presented.

There are three major sources of information on interchanges and alternative intersections:

1. The 2016 *Highway Capacity Manual* (HCM) [1] provides detailed information on the analysis of interchanges and alternative intersections that involve one or more at-grade junctions, which may be signalized or unsignalized. The HCM *does not* specifically treat interchanges that do not involve controlled at-grade intersections. Ramp connections involving merging and diverging movements are treated by applying HCM analysis procedures for ramp junctions, which are covered in Chapter 30 of this text.

2. *A Policy on Geometric Design of Streets and Highways*, 6th Edition [2], AASHTO (also known as the AASHTO “Green Book”) details geometric criteria for various types of interchanges and alternative intersections.

3. The FHWA has published an informational report on interchanges and alternative intersections that is available online [3]. It contains a variety of information, including evaluation approaches on these facilities.

These sources provide a comprehensive overview of the subject. Other sources are available that address specific aspects of design and analysis of such facilities.

26.1 Interchanges

26.1.1 Types of Interchanges

There are many possible geometric configurations that grade-separated interchanges may take. Unfortunately, not all sources categorize them in the same way. In general, the broad categories of grade-separated interchanges are as follows:

- Full cloverleaf (CLO): In a CLO, each turning movement uses a separate ramp, with all such movements accomplished through merging and diverging maneuvers. Left-turn movements are handled on “loop ramps.”

- Partial cloverleaf (PARCLO): In a PARCLO, at least one of the left-turning movements does not have a loop ramp. A PARCLO is generally described as one quadrant, two quadrants, or three quadrants, depending upon how many loop ramps are provided. The left-turn movements that *are not* served by loop ramps are served by some form of at-grade intersection.

- Diamond (DIAMOND): Diamond interchanges have one ramp in each quadrant, which handles both right and left turns (and in some configurations, through movements) on the same ramp. Such interchanges always involve at least one surface street or arterial, and ramp connections to these may be controlled by signals, STOP signs, or YIELD signs.

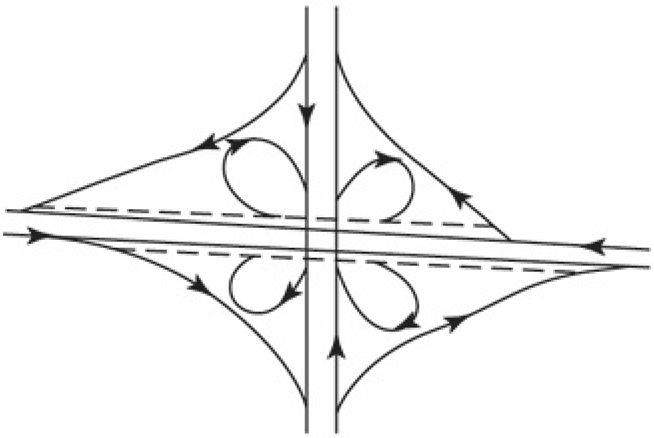
- Directional (DIRECT): In a directional interchange, all right and left turns are handled by ramps that directly turn right or left, that is, there are no loop ramps.

- Single-point urban interchange (SPUI): Popular in urban areas, where right-of-way tends to be seriously restricted, ramps are designed to funnel all left turns through a single signalized intersection with efficient signal timing.

- Diverging diamond interchange (DDI): With traditional diamond interchanges, large left-turn flows generally create operational problems and delays. This relatively new approach provides crossover intersections between the ramp–street connections, essentially allowing all turns (right and left) to be made without opposing flows. This allows for more efficient two-phase signalization.
- One quadrant (ONEQUAD): This type of interchange is used only where two arterials or streets are involved. All movements are handled on a single ramp that is controlled by a signal, STOP sign, or YIELD sign on each arterial.

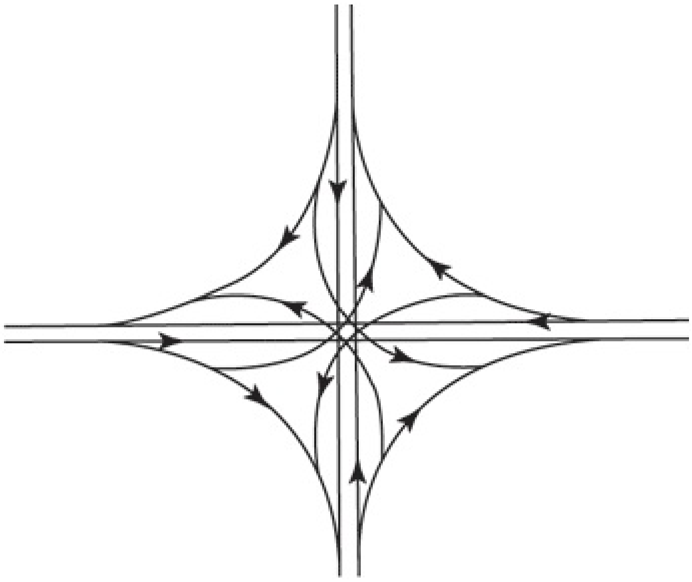
Figure 26.1 illustrates a CLO and direct interchange. The key feature in all of these is that all mainline connections are ramp merges or diverges. These, as noted previously, may be analyzed using merge and diverge methodologies, and do not fall within the purview of the HCM “interchange methodology.”

Figure 26.1: Full Cloverleaf and Directional Interchanges Illustrated



(a) Full Cloverleaf Interchange

26.1-1 Full Alternative Text



(b) Full Directional Interchange

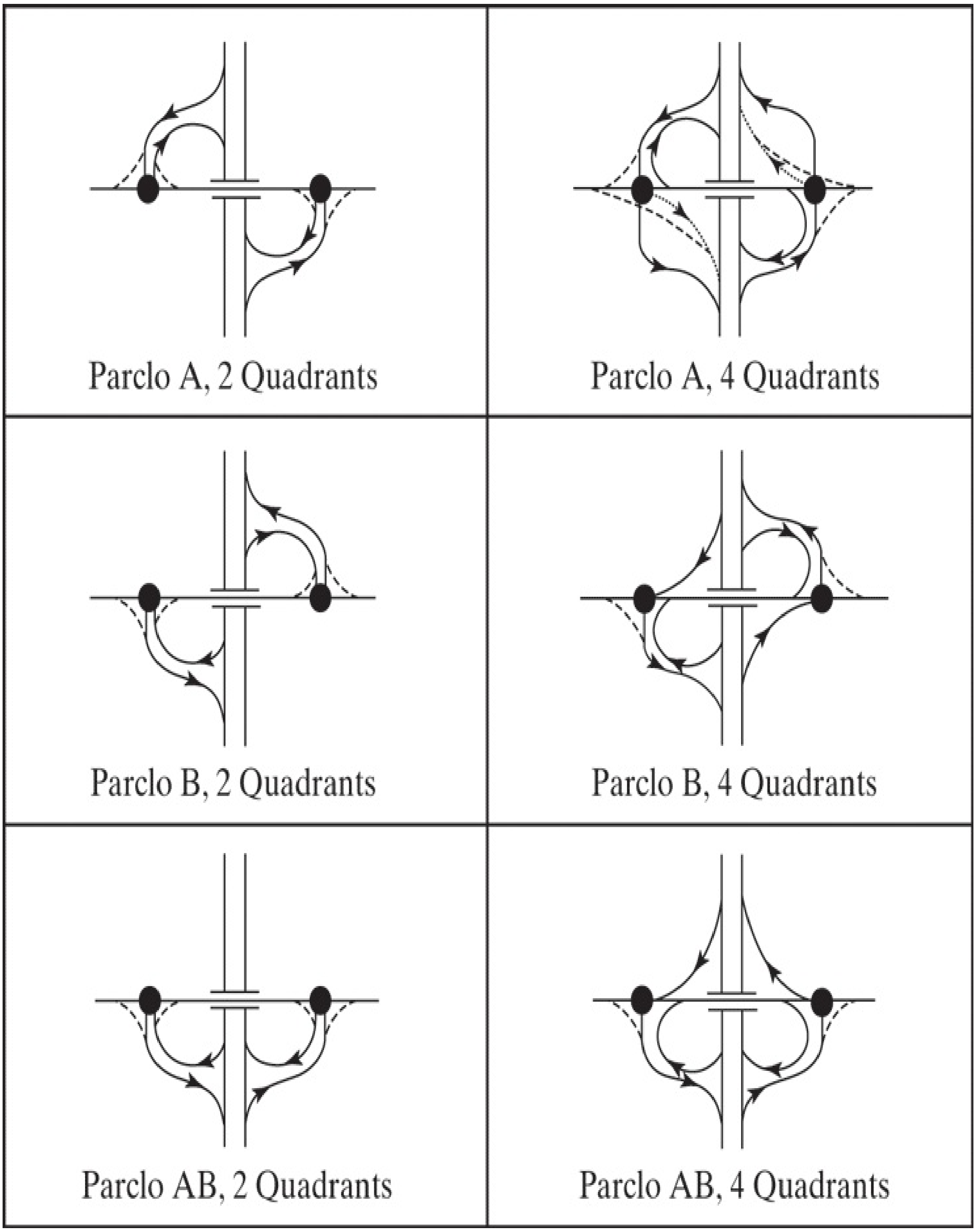
26.1-1 Full Alternative Text

(Source: Adapted from *A Policy on Geometric Design of Highways and Streets*, American Association of State Highway and Transportation Officials, Washington, D.C., 2011.)

Cloverleaf interchanges require a considerable amount of right-of-way, whereas directional interchanges involve four-level structures, which are quite expensive and present a visually imposing barrier. The amount of land required for CLO interchanges is largely dependent on the radius of curvature used on the loop ramps. More gentle radii are beneficial to operations, but significantly increase the footprint of the interchange.

Figure 26.2 shows a variety of PARCLO interchanges. Note that each (for the examples shown) involves two at-grade interchanges that would be controlled by signals, STOP signs, or YIELD signs.

Figure 26.2: Partial Cloverleaf Interchanges Illustrated



(Sources: Reprinted with permission from *Highway Capacity Manual, 6th Edition: A Guide for Multimodal Mobility Analysis*, Transportation Research Board, National Academy of Sciences, Courtesy of the National Academies Press, Washington, D.C., 2016.)

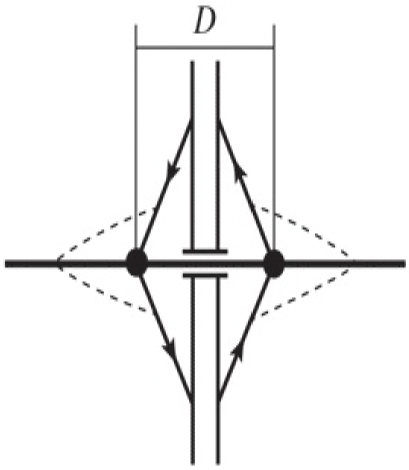
Figure 26.2: Full Alternative Text

The nomenclature for PARCLOs is interesting: In PARCLO A configurations, both loop ramps are located *before* the crossroad. In this configuration, they serve traffic entering the freeway or major arterial. All ramps are in two quadrants. Nonloop ramps can be added to the two other quadrants to simplify turning movements. In PARCLO B configurations, both loop ramps are located *after* the crossroad, that is, they handle traffic exiting the freeway or major arterial. Again, nonloop ramps can be added in the other two quadrants to simplify turning movements. In PARCLO AB configurations, both loop ramps are on the same side of the crossroad, meaning that one will handle traffic *entering* the freeway or major arterial, while the other handles *exiting* traffic. Again, direct ramps can be added to the other quadrants to simplify turning movements.

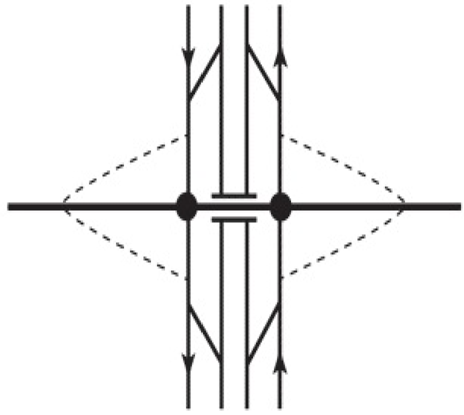
All of the PARCLOs shown in Figure 26.2 have two loop ramps. PARCLOs can also come in configurations with one loop ramp or as many as three.

Figure 26.3 illustrates various forms of diamond interchanges. The diamond interchange is quite prevalent, particularly in low-density areas, because of its simplicity, lower construction costs, and generally efficient use of right-of-way. Various forms, however, are also frequently used in urban areas.

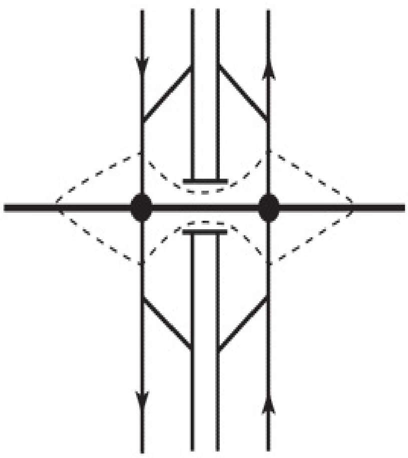
Figure 26.3: Diamond Interchanges Illustrated



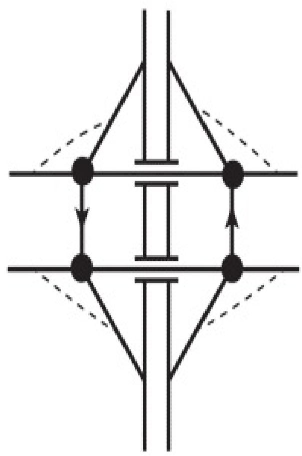
Conventional, $D > 800$ ft
Compressed, $D = 400 - 800$ ft
Tight Urban, $D < 400$ ft



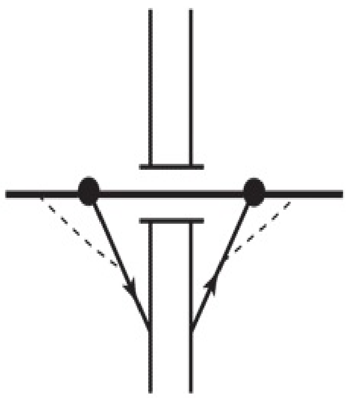
Diamond with continuous
frontage roads



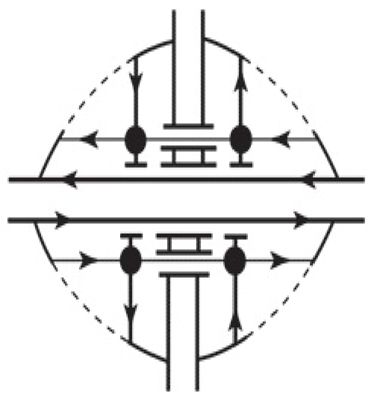
Diamond with continuous frontage
roads and U-turn lanes



Split diamond interchange
(crossroads, one- or two-way)



Partial (half) diamond
interchange



Three-level diamond
interchange

(Source: Reprinted with permission from *Highway Capacity Manual, 6th Edition: A Guide for Multimodal Mobility Analysis*, Transportation Research Board, National Academy of Sciences, Courtesy of the National Academies Press, Washington, D.C., 2016.)

Figure 26.3: Full Alternative Text

The conventional diamond interchange uses four straight ramps, forming two surface intersections with the connecting street or arterial. These intersections may be signalized or controlled by STOP or YIELD signs. “Compressed” and “tight urban” diamond interchanges have the same basic geometry, but are classified based upon the distance between the two surface intersections that are formed. The latter two types are more typical of suburban or urban locations.

In many urban and suburban areas, continuous service roads are provided for the freeway. In these cases, the diamond ramps connect to these service roads. The surface intersections formed are more complex, in that the service roads are part of the local street network and carry sometimes significant volumes that are not specifically associated with movements using interchange ramps. In such cases, designated U-turn roadways may or may not be provided.

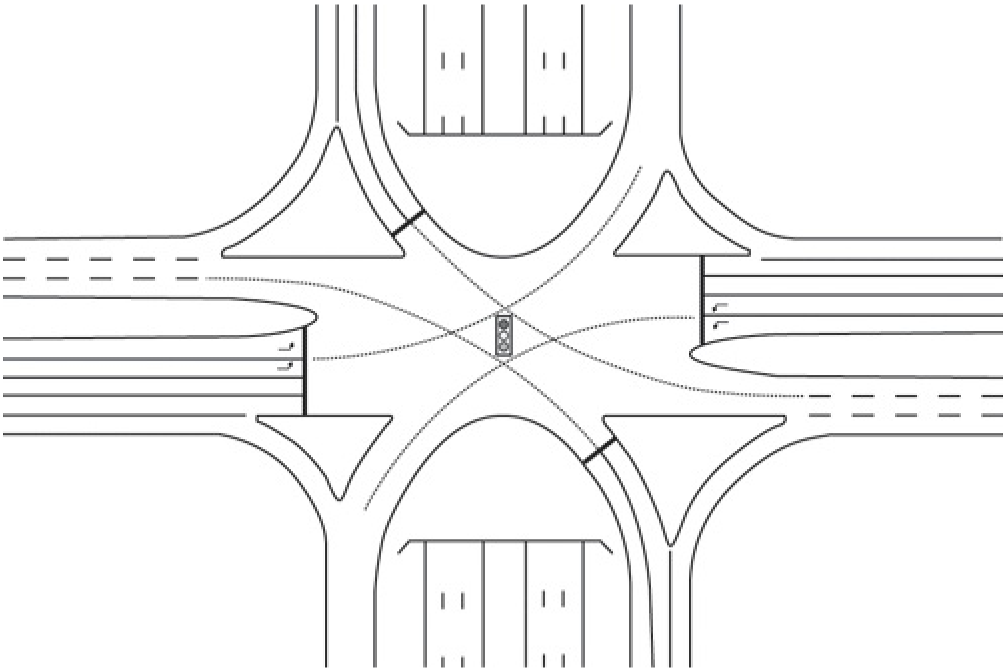
Split diamond interchanges utilize two generally closely spaced cross streets (which may be one-way or two-way) to separate some of the turning movements. In these cases, four surface intersections are created, and a short service road connecting the two is provided.

In a three-level diamond, which is a relatively rare configuration, a short set of lanes from the freeway is provided to handle all ramp-freeway connections.

Figure 26.4 shows a SPUI. Such interchanges are generally used in urban or suburban settings. The major facility is most often a freeway, but other types of major facilities are possible, as long as an overpass or underpass is provided for the major facility. The intersecting major street may be on the lower or upper level of the interchange.

Figure 26.4: A Single-Point

Urban Interchange Illustrated



(Source: Reprinted with permission from "A Comparison of the Operations of Single-Point and Tight Urban Diamond Interchanges," *Transportation Research Record 1847*, Transportation Research Board, National Academy of Sciences, Courtesy of the National Academies Press, Washington, D.C., 2003.)

Figure 26.4: Full Alternative Text

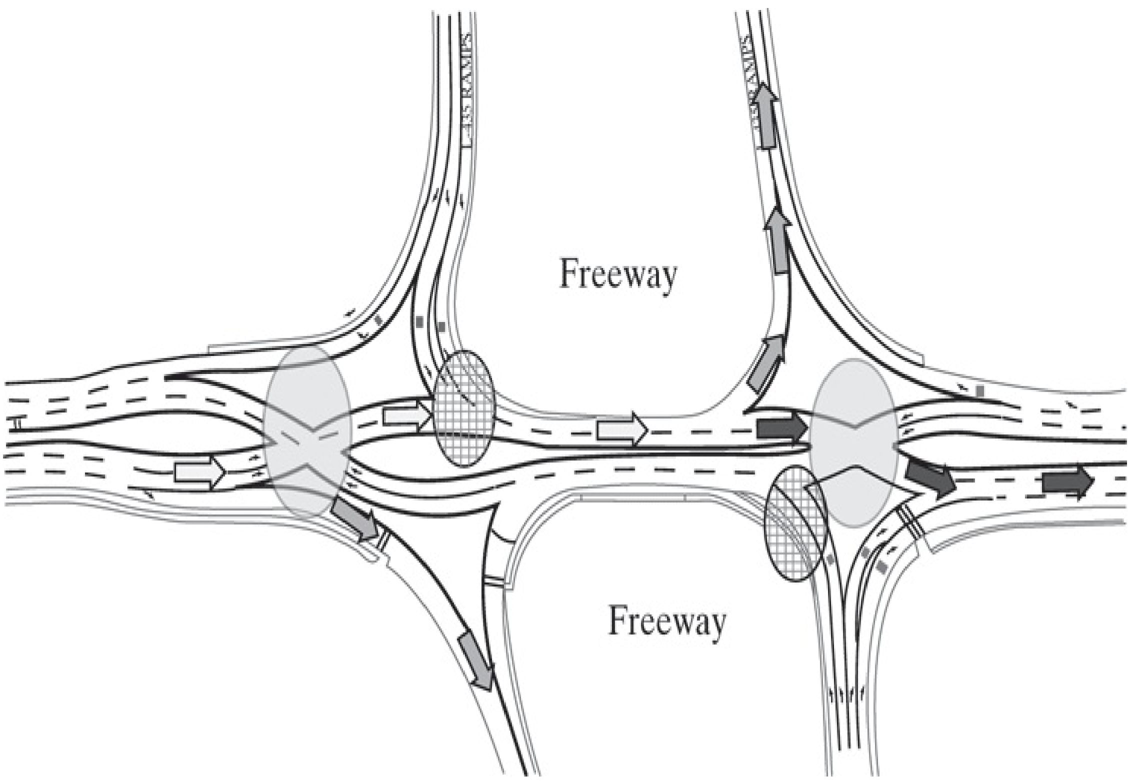
Right-turn movements onto and off of the major facility are handled using ramp merges or diverges. Ramps entering the surface street are generally controlled using a YIELD sign. All left turns are funneled through a signalized intersection, most often using a three-phase timing plan: one phase for left turns exiting the major facility, one phase for left turns entering the facility, and one phase for through movements on the surface street. The design avoids having two closely spaced intersections that would normally have to be coordinated if a typical tight urban diamond interchange were used. An excellent comparison of SPUIs versus tight urban diamond interchanges by Jones and Selinger is available in the literature [4]. A general overview of the design and operation of SPUIs was prepared by Messer et al. [5].

The DDI is one of the newer developments in interchange design that is rapidly seeing increased use across the country. These interchanges are sometimes referred to as double crossover diamond interchanges.

Traditional diamond interchanges become operationally inefficient when large volumes of left-turning vehicles are present. On the surface street, queues of left-turning vehicles accumulate during RED phases, sometimes encroaching on the upstream intersection. The two surface intersections, where large left-turn volumes exist, often involve complex multiphase signalization, which can lead to long cycle lengths, which in turn exacerbates queuing between the intersections.

The DDI ameliorates these problems by providing for two crossovers. In effect, between the two intersections, directional flows are on the *left* instead of the *right*. Figure 26.5 illustrates how this is done.

Figure 26.5: Diverging Diamond Interchange Illustrated



(Source: Hughes, W., Jagannathon, R., Sengupta, D., and Human, J., "Alternative Intersections/Interchanges: Informational Report," *Publication Number FHWA-HRT-09-060*, Federal Highway Administration, Washington, D.C., 2009, Figure 154.)

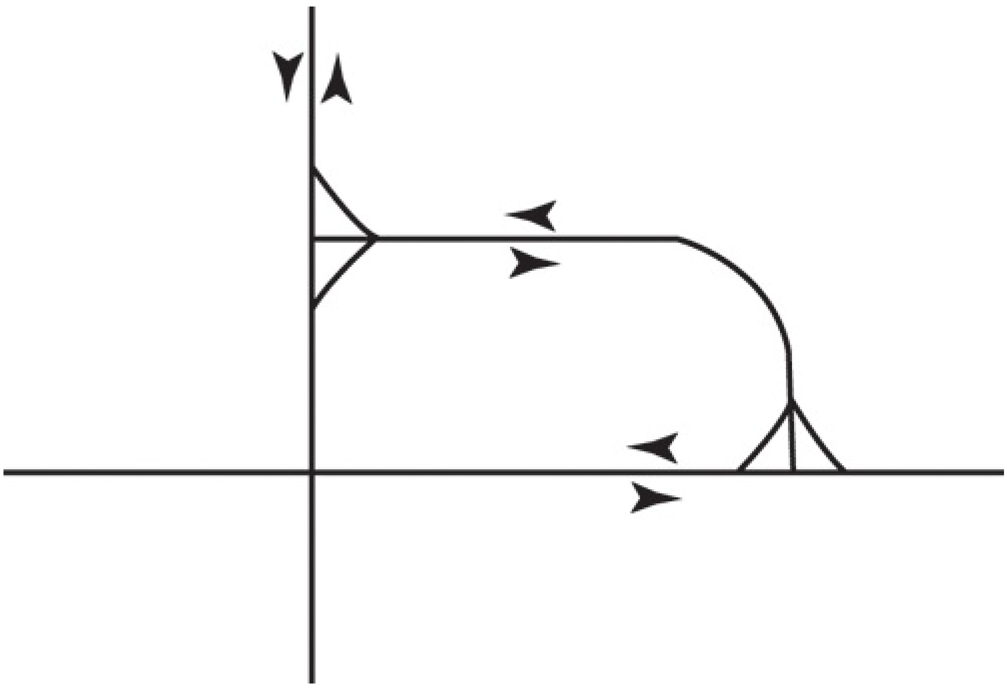
Figure 26.5: Full Alternative Text

Through the use of the crossovers, all turning movements become unopposed, which translates to not requiring a separate turning phase to safely accommodate them. Two surface intersections remain, but can be safely timed using an efficient overlapping phase system.

The first known DDI in the United States opened in Springfield, Missouri, on June 22, 2009 [2]. By the end of 2015, there were 60 such interchanges up and running, and more than 40 more in planning and design stages. Virtually all of the U.S. DDIs were constructed as replacements for existing traditional diamond interchanges that were experiencing severe congestion, delays, and accidents.

The single-quadrant interchange is most often used in low-density or rural areas where all movements can be handled efficiently with a single ramp accommodating both directions of flow. This type of interchange can only be used where both highways are surface arterials or streets, and would be classified as an "interchange" only where the two intersection facilities are grade-separated. Figure 26.6 illustrates a one-quadrant interchange.

Figure 26.6: A One-Quadrant Interchange Illustrated



(Source: Adapted from *A Policy on Geometric Design of Highways and Streets*, American Association of State Highway and Transportation Officials, Washington, D.C., 2011.)

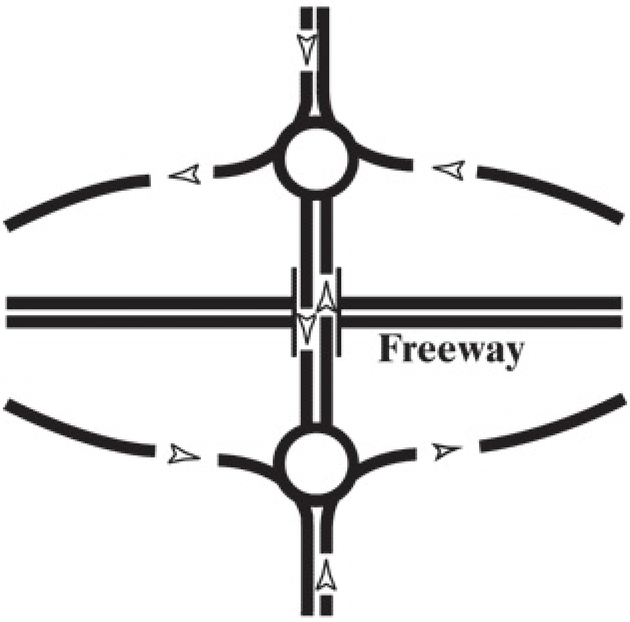
Figure 26.6: Full Alternative Text

The two intersections created may be controlled by signals, STOP signs, or YIELD signs. The one-quadrant approach has limited applicability, as it does not serve large turning volumes efficiently, both intersecting streets must be surface facilities, and densities should be relatively low. The benefits of the approach include minimizing structural costs, and consuming a minimum of right-of-way.

26.1.2 Interchanges with Roundabouts

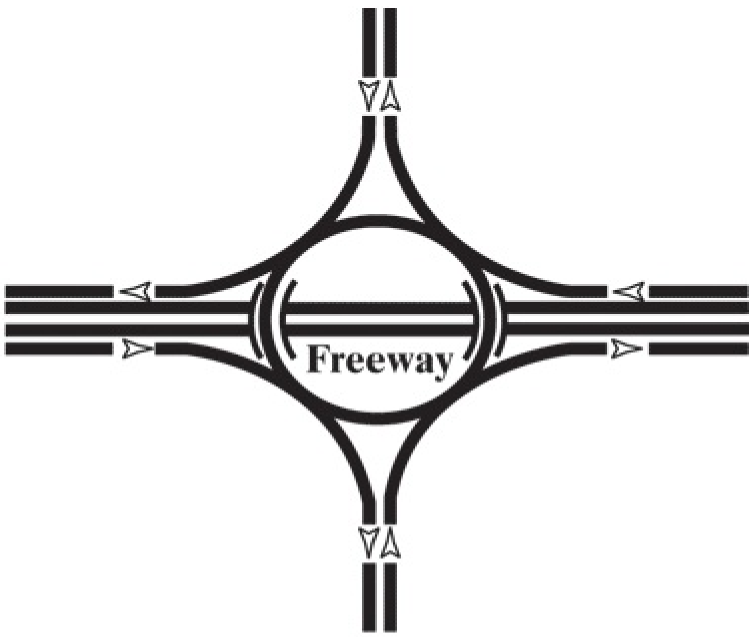
As the use of roundabouts has increased over the years, a number of innovative interchange designs have been developed that include one or more roundabouts as part of the overall interchange. Figure 26.7 illustrates the two major designs that have developed.

Figure 26.7: Interchanges Using Roundabouts Illustrated



(a) Roundabout Ramp Terminals

26.1-1 Full Alternative Text



(b) Single-Point Roundabout Interchange

26.1-1 Full Alternative Text

(Source: Reprinted with permission from *Highway Capacity Manual, 6th Edition: A Guide for Multimodal Mobility Analysis*, Transportation Research Board, National Academy of Sciences, Courtesy of the National Academies Press, Washington, D.C., 2016.)

In Figure 26.7(a), two roundabouts replace the two intersections formed by the ramp–street junctions. Given the simplified number of entry and exit points from these roundabouts (two entries and two exits), the shape of the roundabouts is sometimes altered to form the shape of a "teardrop," by elongating the roundabouts along the axis of the surface street.

In Figure 26.7 (b), the roundabout forms part of the on- and off-ramp movements. Where demand volumes are appropriate for a roundabout, this can be a very efficient approach.

In both cases, each individual roundabout would have to be operationally evaluated using the methodology presented in Chapter 25.

26.1.3 Impacts of Interchange Type and Selection of an Appropriate Interchange

There are many factors that must be considered in selecting an appropriate type of interchange for any given situation. These include, but are not limited to, the following:

- Availability and cost of right-of-way
- Structural costs
- Aesthetic considerations
- Social considerations

- Operational efficiency
- Safety

Cost issues, as noted, are related to acquisition of adequate right-of-way and structures. Some forms of interchanges involve complex multilevel structures, whereas others involve a simple overpass or underpass separating the two intersection facilities. Use of structures, however, can help minimize the utilization of land, and can be used to increase operational efficiency of the interchange.

The single biggest impact on operations is to turning movements and how they are made, particularly left-turning movements. At any interchange between two facilities, there are four possible origins and four possible destinations. Figure 26.8 illustrates this concept.

Figure 26.8: Origins and Destinations for an Interchange (4-Leg)

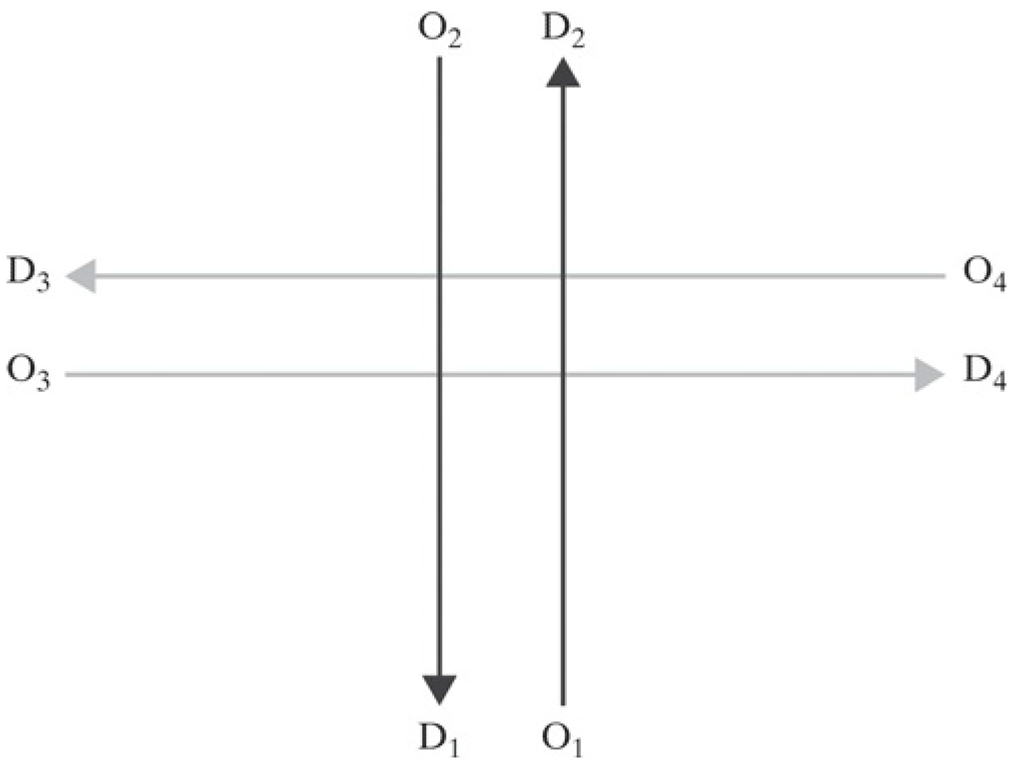
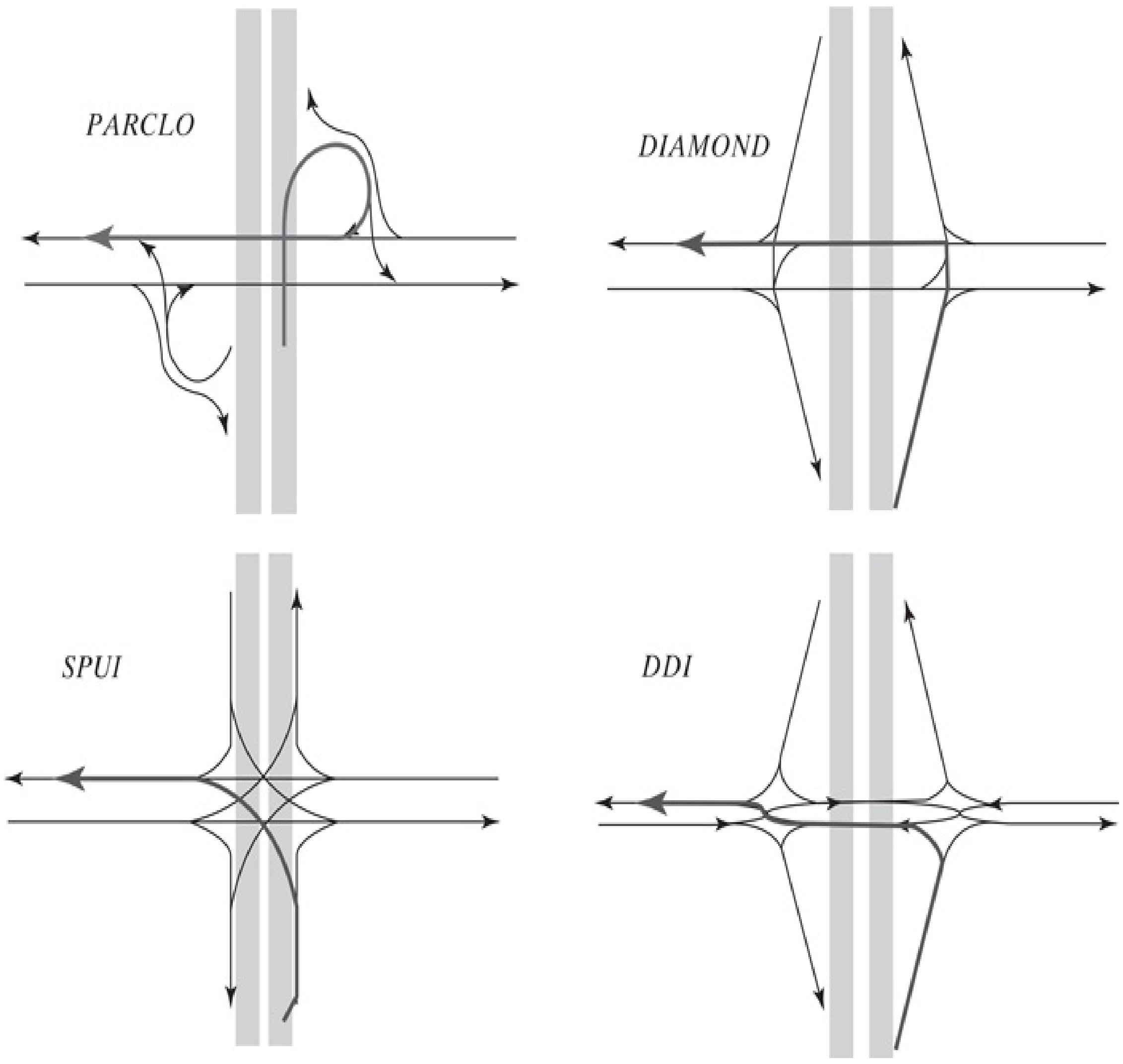


Figure 26.8: Full Alternative Text

If U-turns are considered, there are 4×4=16 origin–destination movements that are possible. Four are U-turns, four are through movements, four are right turns, and four are left turns. Depending upon the type of interchange selected, the turning movements may take very different paths. Through movements are, for the most part, unaffected by the type of interchange in place.

Figure 26.9 illustrates how four different types of interchanges affect the left-turn movement O_1 to D_3.

Figure 26.9: Effect of Interchange Type on Turning Paths Illustrated



(Source: Reprinted with permission from Base figure from *Highway Capacity Manual, 6th Edition: A Guide for Multimodal Mobility Analysis*, Transportation Research Board, National Academy of Sciences, Courtesy of the National Academies Press, Washington, D.C., 2016.)

Figure 26.9: Full Alternative Text

In the PARCLO shown, the subject left turn is made using a loop ramp, which merges into the destination traffic stream. These vehicles then travel as a through movement at one intersection, which may or may not be signalized. In the DIAMOND configuration, the same movement makes a left-turn through one intersection, and travels as a through movement through the second intersection. Either intersection may be signalized or otherwise controlled. In the SPUI, the left turn travels through only one intersection, which is often signalized. In the DDI, the subject movement merges into the destination traffic stream, and then goes through the crossover intersection as a through movement. The crossover intersection and the merge may be signalized.

Similar illustrations can be provided for other interchange configurations as well. The key point, however, is that when considering interchange types, the path of each movement must be carefully considered in conjunction with alternative forms of intersection control to arrive at an optimal approach.

The HCM [1] contains a detailed procedure for selection of an appropriate interchange type based upon information generally available during the design and planning stages of development. Chapter 34 of the HCM: “Interchange Ramp Terminals—Supplemental” should be consulted directly for this methodology. It should be noted, however, that the method only addresses interchanges involving signalized intersections. It does not consider interchanges that do not have signals, or those that involve roundabouts.

26.2 Alternative Intersections

In addition to traditional roundabouts, there are four major categories of alternative intersections in general usage. These include the following:

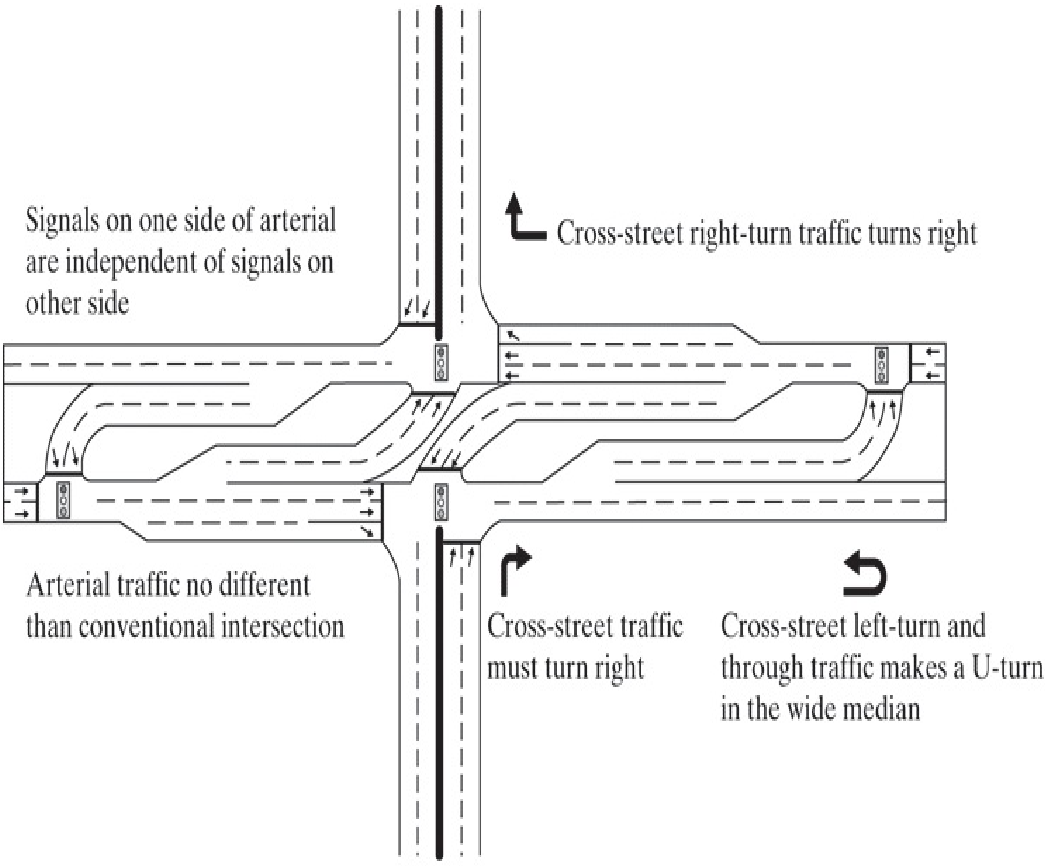
- Restricted crossing U-turn intersections (RCUT)
- Median U-turn intersections (MUT)
- Displaced left-turn intersections (DLT)
- Quadrant or jug-handle intersections

All of these are surface intersections without overpasses or underpasses. All turning movements are made at-grade. These intersections all seek to simplify turning movements and conflicts at the intersection, and do so by rerouting left turns and (in some cases) minor street through movements.

26.2.1 Restricted Crossing U-Turn Intersections (RCUT)

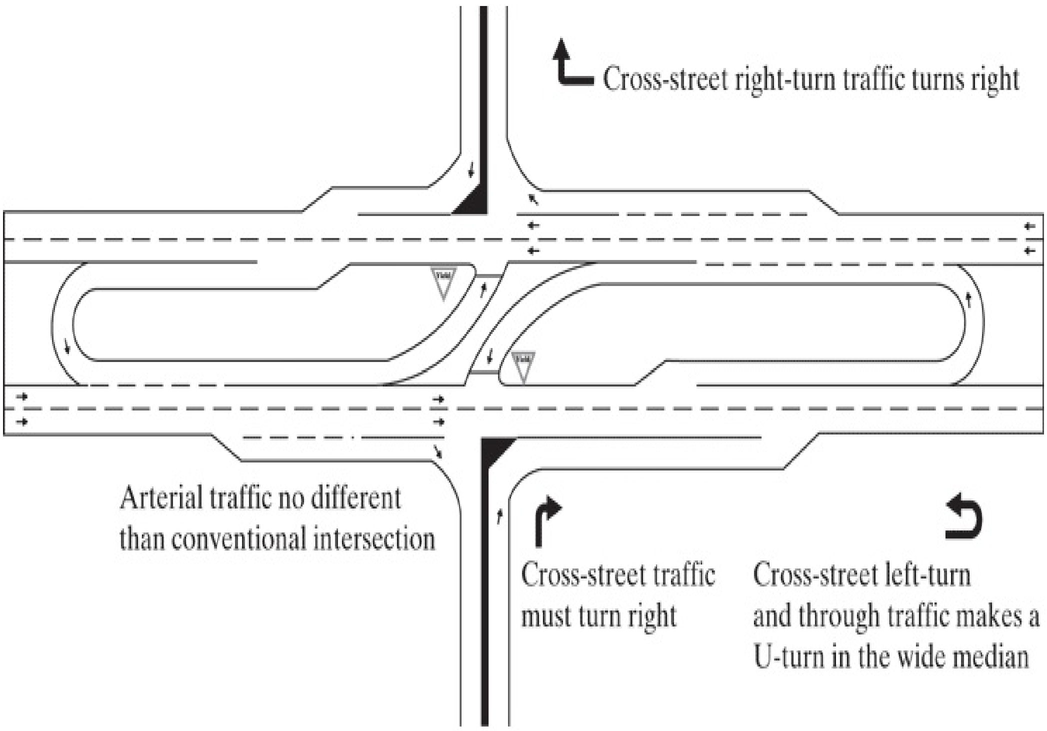
At a RCUT intersection, minor street through and left-turn movements are rerouted through a U-turn lane on the major street. Such intersections can operate with or without signals. Figure 26.10 illustrates two types of RCUT intersection, one with signalization and one without.

Figure 26.10: Restricted Crossing U-Turn Intersections Illustrated



(a) Signalized RCUT Intersection

26.2-1 Full Alternative Text



(b) RCUT Intersection Using Merges/Diverges

26.2-1 Full Alternative Text

(Source: Reprinted with permission from *Highway Capacity Manual, 6th Edition: A Guide for Multimodal Mobility Analysis,* Transportation Research Board, National Academy of Sciences, Courtesy of the National Academies Press, Washington, D.C., 2016.)

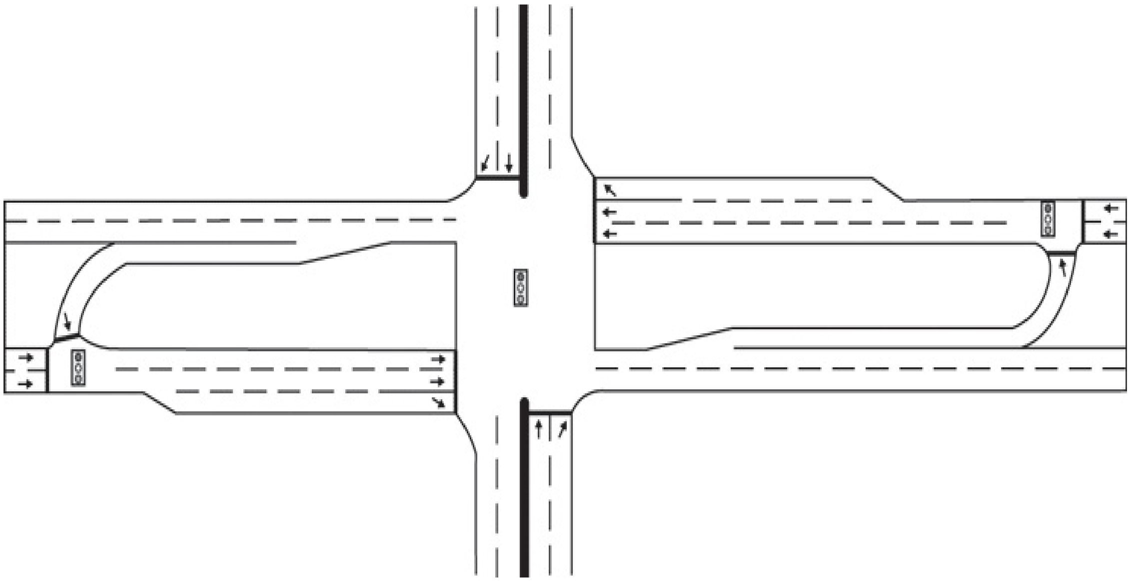
In Figure 26.10(a), both the main intersection and the U-turn roadways are controlled by traffic signals. In some cases, only the main intersection would by signalized. In Figure 26.10(b), all movements are achieved by higher-speed merging or diverging maneuvers. This type of configuration works best when the side-street crossing traffic is light, and most minor street vehicles are turning right or left.

Obviously, the first requirement for operating such an intersection is a relatively wide median on the major street. In Figure 26.10(a), there are double turning lanes in each direction that occupy median space. Thus, a minimum of $4\times12=48$ feet would be needed—and that is without a buffer between the opposing turning lanes. Something in the 50–60 feet range would generally be required. As noted, the minor street through flow should be low to avoid congestion caused by large numbers of vehicles making U-turns. The distance between the point where LT or minor street TH vehicles enter the major street and the start of the U-turn lane must be sufficient to allow for safe and efficient lane-changing for these vehicles. Additional detailed information on RCUT intersections can be found in Refs. [1, 2, 3, 6, 7, 8].

26.2.2 Median U-Turn Intersections (MUT)

In many ways, MUTs are similar to RCUTs. The primary difference is that no through movements are diverted. At an MUT, left turns from the minor and/or major street are diverted to a U-turn lane downstream of the intersection. Figure 26.11 illustrates an MUT with traffic signals controlling the U-turn lanes and the main intersection.

Figure 26.11: MUT Intersection with Signalization Illustrated



(Source: Reprinted with permission from *Highway Capacity Manual, 6th Edition: A Guide for Multimodal Mobility Analysis*, Transportation Research Board, National Academy of Sciences, Courtesy of the National Academies Press, Washington, D.C., 2016.)

Figure 26.11: Full Alternative Text

When signals are used to control the U-turn lanes of either an MUT or an RCUT, timing must avoid queues that spill back into the through lanes. An MUT can be created without signalization of the U-turn lanes, depending upon existing traffic conditions.

26.2.3 Displaced Left-Turn Intersections

Displaced left-turn intersections have been in use since the 1990s, and were originally referred to as “continuous flow intersections.” The concept is relatively simple: to avoid a heavy opposed LT movement at an intersection, left turns are shifted to the left side of the intersection at an upstream location. At the main intersection, these turns are then unopposed, and can be handled in the same phase as through movements. Figure 26.12 illustrates a DLT.

Figure 26.12: A Displaced Left Turn or Continuous Flow Intersection Illustrated

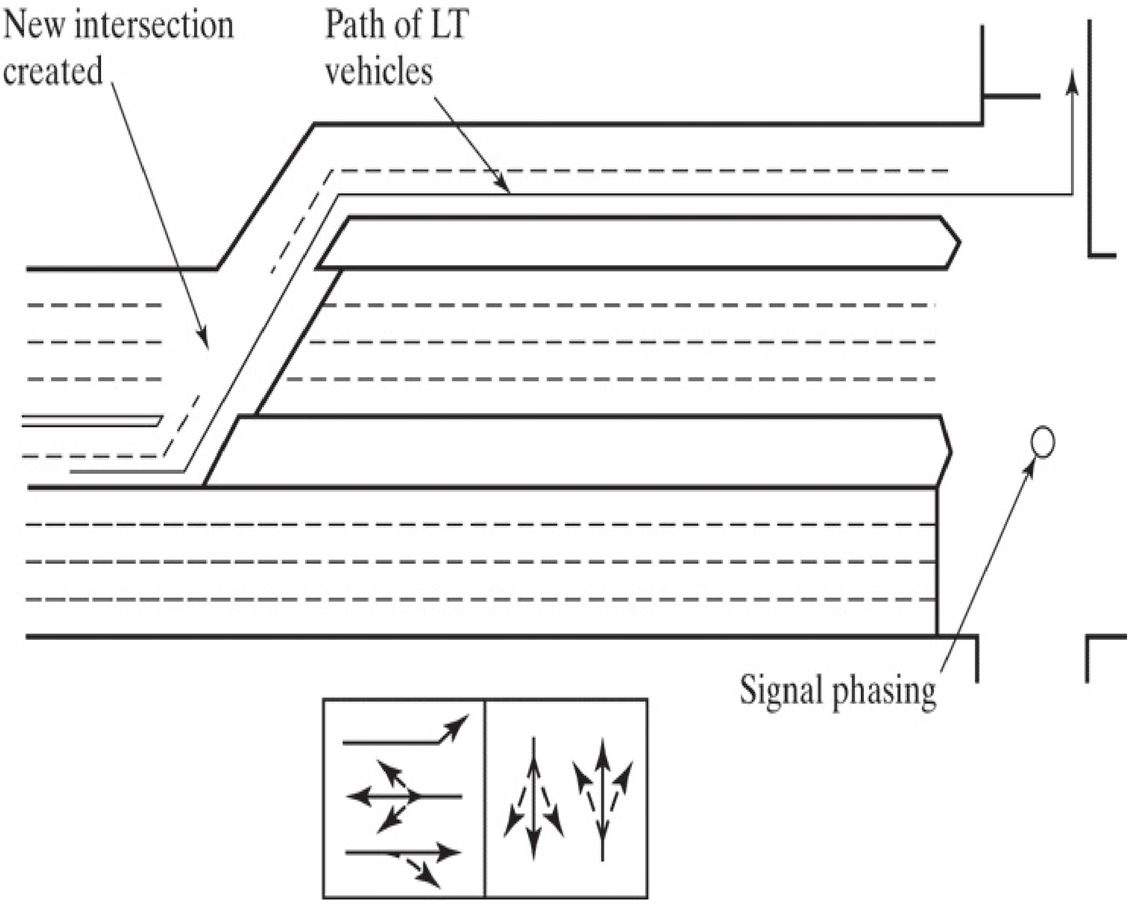


Figure 26.12: Full Alternative Text

As indicated in Figure 26.12, these intersections can often be signalized using a simple two-phase timing plan. A second intersection at the crossover point is, however, also created. It may or may not require signalization, depending upon traffic demands. If a signal is required, again, it would be a simple two-phase timing plan.

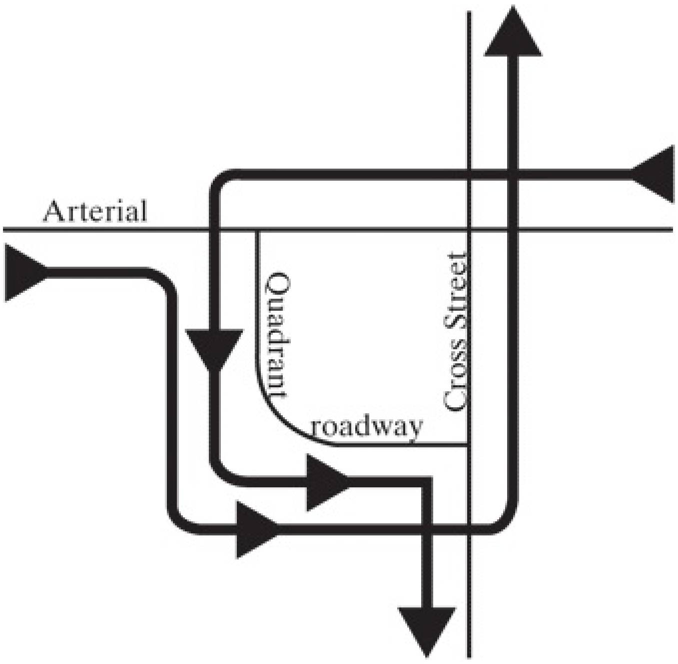
The setback of the crossover intersection is critical. Queues from the main intersection cannot be permitted to spill back and block crossover maneuvers. This is easier to do when the crossover is signalized, as common traffic law makes “blocking the intersection” illegal. While Figure 26.12 shows two left-turn lanes after the crossover, one may be sufficient. For exceedingly heavy flows, three left-turns lanes could be provided.

Figure 26.12 shows a DLT in one quadrant of the intersection. It is possible to design all four quadrants in this manner, as long as sufficient right-of-way is provided.

26.2.4 Quadrant and Jug-Handle Intersections

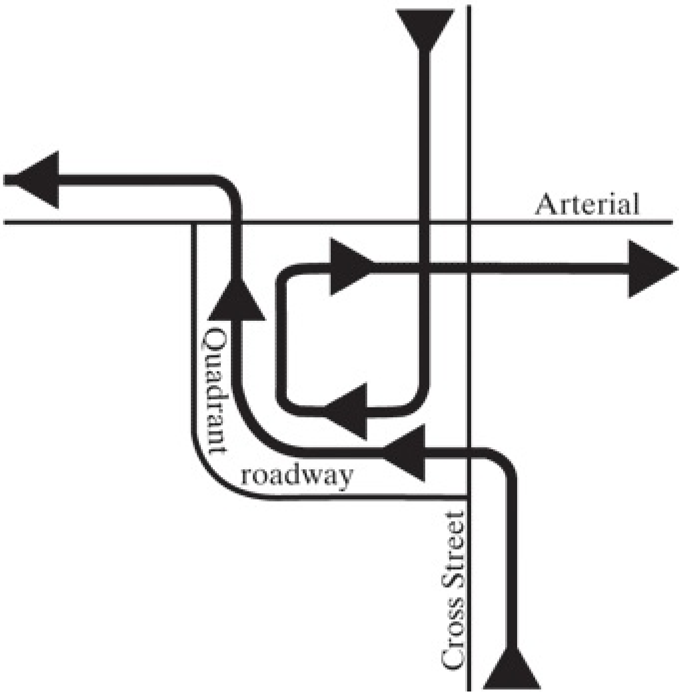
Quadrant and jug-handle intersections are hardly a new development. The jug-handle form has been used extensively, particularly in New Jersey for over 50 years. The concept is quite simple: all left turns are handled by a surface connecting roadway located in one or more quadrants of the intersection. A single-quadrant connection can handle all left-turning movements, but it must be a two-way connector, and it will form two additional intersections, one with each of the intersection streets/arteries. Figures 26.13–16 illustrate several forms of quadrant intersections.

Figure 26.13: Quadrant Intersections in One Quadrant



(a) Left turn pattern from the arterial

26.2-1 Full Alternative Text



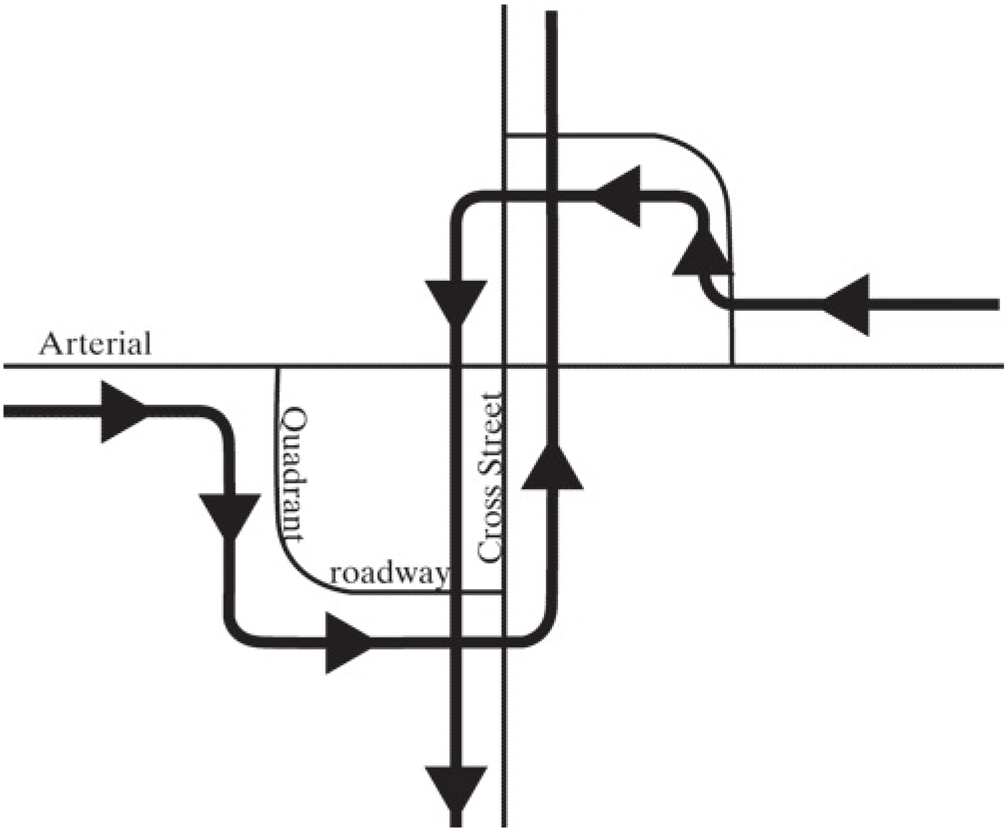
(b) Left turn pattern from the cross street

26.2-1 Full Alternative Text

(Source: Hughes, W., Jagannathan, R., Segupta, D., and

Hummer, J., “Alternative Intersections/Interchanges: Informational Report,” *Publication Number FHWA-HRT-09-060*, Federal Highway Administration, Washington, D.C., 2009, Figure 126.)

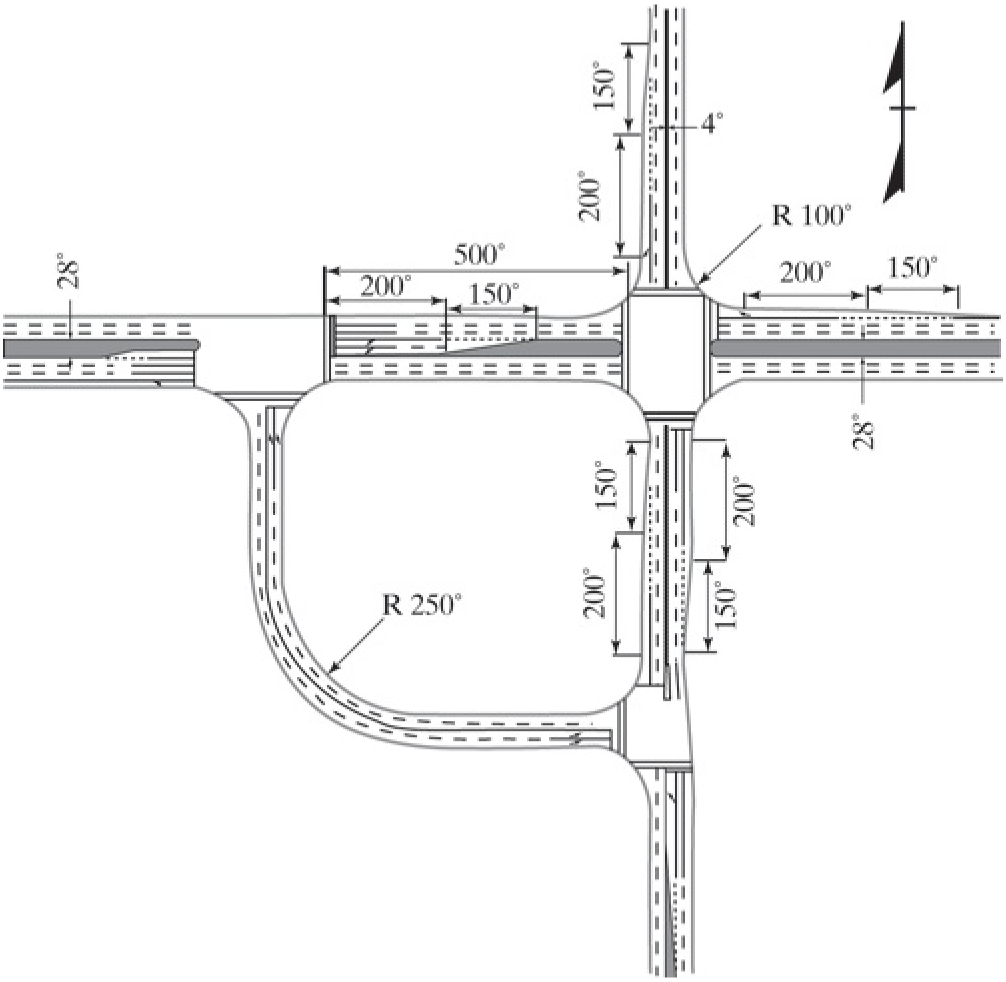
Figure 26.14: Quadrant Intersection in Two Quadrants



(Source: Hughes, W., Jagannathan, R., Segupta, D., and Hummer, J., “Alternative Intersections/Interchanges: Informational Report,” *Publication Number FHWA-HRT-09-060*, Federal Highway Administration, Washington, D.C., 2009, Figure 127.)

Figure 26.14: Full Alternative Text

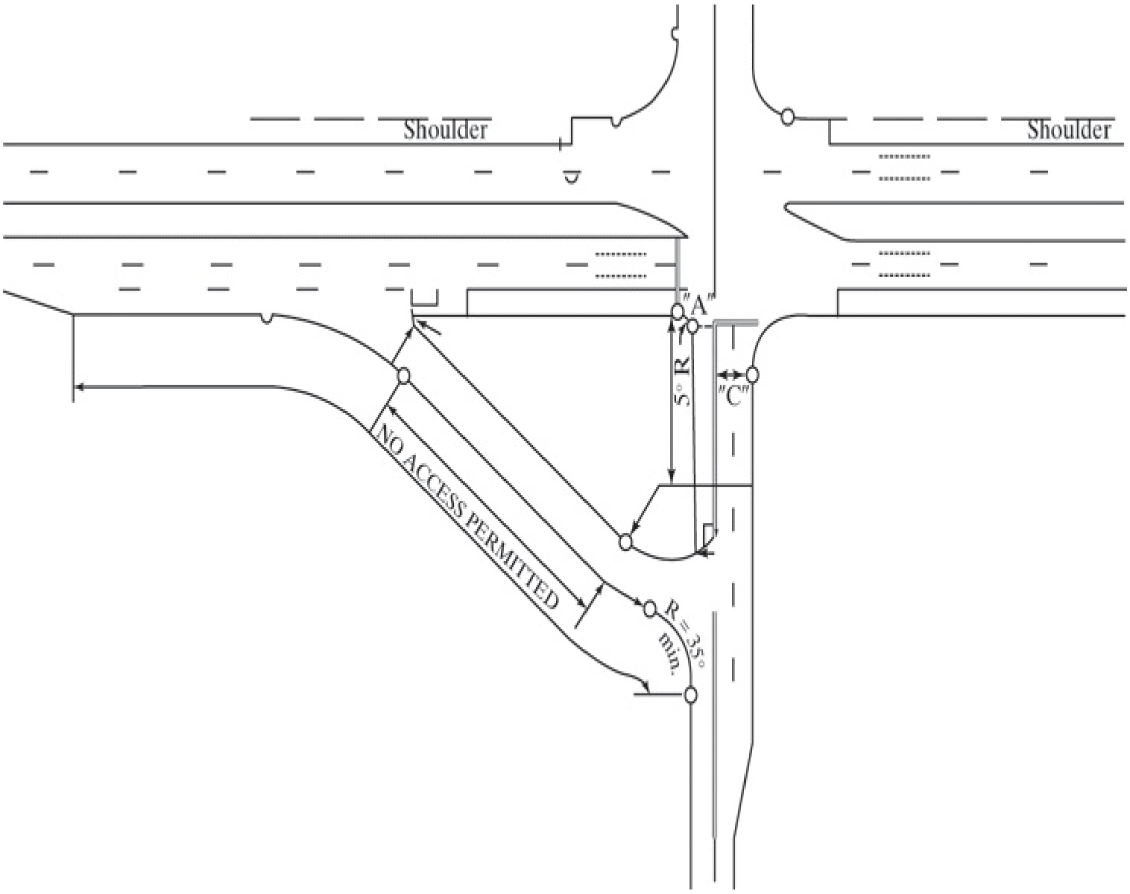
Figure 26.15: Illustrative Design for a Single Quadrant Intersection with a Two-Way Connecting Roadway



(Source: Hughes, W., Jagannathan, R., Segupta, D., and Hummer, J., “Alternative Intersections/Interchanges: Informational Report,” *Publication Number FHWA-HRT-09-060*, Federal Highway Administration, Washington, D.C., 2009, Figure 128.)

Figure 26.15: Full Alternative Text

Figure 26.16: A Typical Jug-Handle Intersection



(Source: Hughes, W., Jagannathan, R., Segupta, D., and Hummer, J., “Alternative Intersections/Interchanges: Informational Report,” *Publication Number FHWA-HRT-09-060*, Federal Highway Administration, Washington, D.C., 2009, Figure 143.)

Figure 26.16: Full Alternative Text

Note that only left turns from the arterial are shown in Figure 26.14. If the two connecting roadways are bi-directional, left turns from the cross street would also be handled. In all quadrant intersections, however, diverted left turns still have to make at least one left turn to either enter and/or exit the connecting roadway. This, however, removes them from the primary intersection, which can then be timed more efficiently.

Figure 26.15 shows an illustrative design of a two-way connecting roadway at a single quadrant intersection. The primary intersection is signalized, and, in many cases, two-way connecting roadway intersections may also have to be signalized. Queuing is a critical issue. Queues from the main intersection cannot be permitted to spill back into and block the connecting roadway intersections. A minimum spacing of 500 feet between the main and connecting roadway intersections is generally used, and longer setbacks are desirable where right-of-way exists. At the connecting roadway intersections, green time for the connecting roadways is kept to a minimum, with timing favoring through movements. This also helps to minimize queues.

A typical jug-handle design is illustrated in Figure 26.16. The jug-handle is defined (by the New Jersey Department of Transportation) as an at-grade ramp that allows left turns to be diverted. In most cases, jug handles are one-way connections, and most often only left turns from the major street are affected.

26.2.5 Left-Turn Management

The intent of alternative intersections is to simplify complex intersection movements and conflicts, particularly those involving left turns. In all cases, alternative geometries are used to remove all or some left turns (and/or U-turns) from the primary intersection. Doing so, however, often creates additional signalized intersections.

The benefit is that the main intersection in such configurations can often be controlled with a simple two-phase signal. The objective is to design the alternative intersection so that all of the signalized elements can be effectively controlled with two phases.

Queuing from the main intersection into new intersections with connecting roadways must be considered as part of the planning and design process, as well as in the timing of any traffic signals involved.

Left turns are always the most difficult elements to accommodate at signalized and unsignalized intersections. They typically involve more potential conflicts than other movements, and their presence can lead to complex multiphase signalization that lengthens cycle lengths, increases delay, and increases the size of queues that develop during RED intervals. In essence, alternative intersections seek to replace complex signalization with innovative geometry. The cost of doing this, however, is that some vehicles are diverted from their desired path and will experience additional travel time in following the indirect paths prescribed by the alternative intersection design.

26.3 Level of Service Analysis

The HCM [1] provides detailed level of service analysis procedures for *distributed intersections*, which include interchanges and alternative intersections that include one or more intersections controlled by traffic signals or STOP signs. The procedures are complex and very detailed, but they do not include interchanges that entail only ramp merge and diverge maneuvers. While the HCM includes a methodology for analysis of merge and diverge junctions, there is no methodology for incorporating these into an overall interchange analysis.

This chapter does not detail the entire analysis procedure for interchanges and alternative intersections, but does explain the basic concept and its key elements.

26.3.1 A Framework for Level of Service Analysis of Interchanges and Alternative Intersections

Level of service analysis of interchanges and alternative intersections is based upon the *experienced travel time* (ETT) of individual movements through the interchange or alternative intersection. In general, the ETT for a movement includes all delays experienced along its path due to control devices (signals, STOP signs), and extra travel time due to redirection of a movement's path from what would exist at a simple surface intersection.

$ETT=\sum d_i+\sum EDTT$ [26-1]

where:

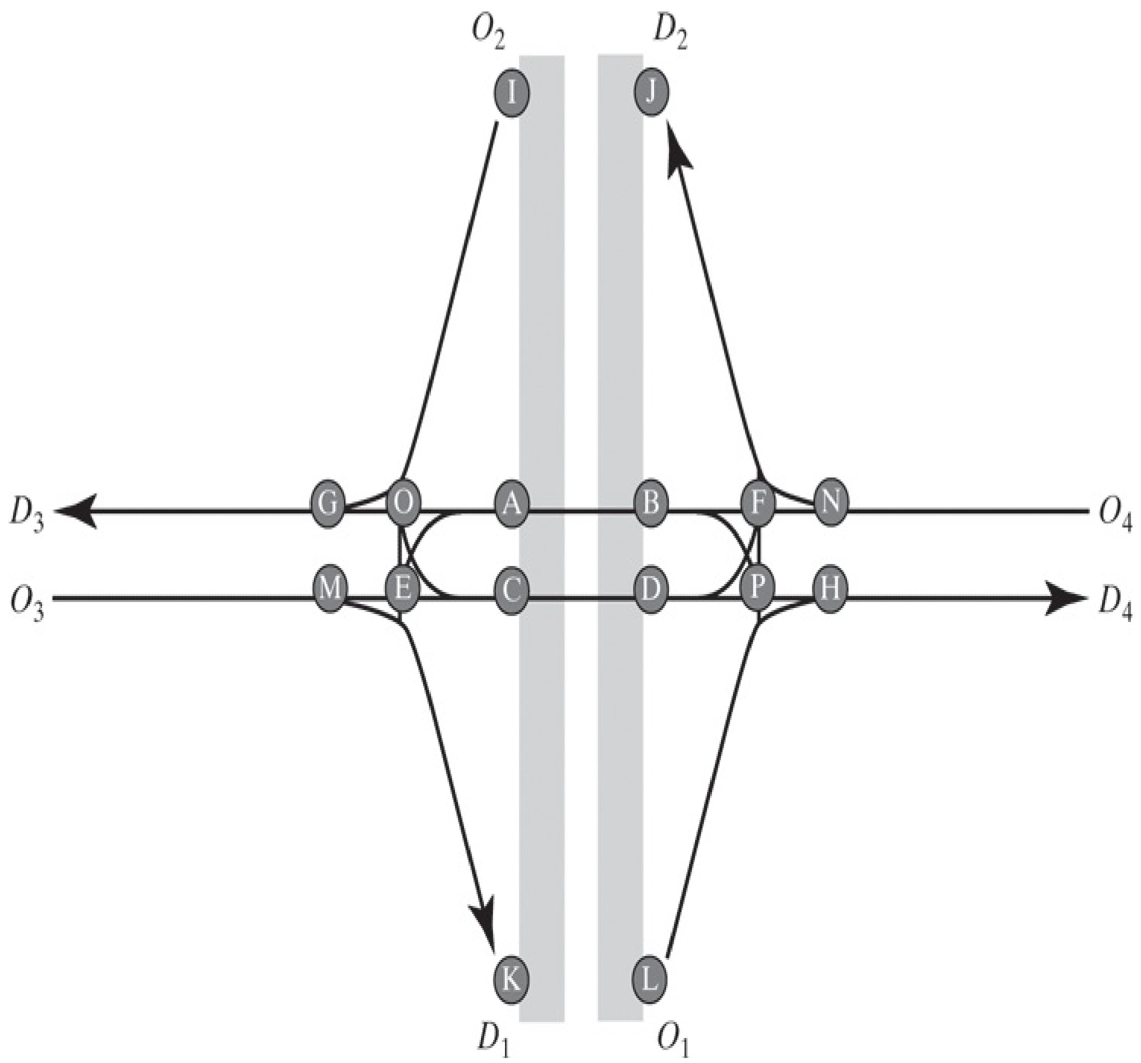
$ETT=$experienced travel time (s/veh),$d_i=$control delay for each junction i tha

The EDTT is evaluated based upon travel times between the origin and destination points compared to the hypothetical travel time that would be experienced if the movement were made at a 90-degree angle at a surface intersection. As delays are included separately, these travel times are evaluated at the free-flow speed. In most cases, through movements do not experience EDTT. Some designs, however (such as an RCUT intersection), do divert through movements, and would have an EDTT.

26.3.2 Extra-Distance Travel Time

In some cases, the EDTT is at least theoretically obvious. In many cases, however, it is not, and care must be taken in its evaluation. Consider the diamond interchange illustrated in Figure 26.17.

Figure 26.17: A Diamond Interchange Illustrating *EDTT*



(Source: Reprinted with permission from *Highway Capacity Manual, 6th Edition: A Guide for Multimodal Mobility Analysis*, Transportation Research Board, National Academy of Sciences, Courtesy of the National Academies Press, Washington, D.C., 2016.)

Figure 26.17: Full Alternative Text

Consider the left-turn movement from O_1 to D_3:

- The actual path of this left turn is L, P, B, A, O, G.
- If a hypothetical surface intersection were in place (the freeway would be at-grade), the actual path would be L, B, A, O, G.
- The last portion of both paths is the same. Therefore, the travel time involved in path A, O, G is irrelevant and can be ignored.

- Then, the EDTT for this movement would be:

EDTT1,3=TTLBP−TTLB [26-2]

where:

EDTT1,3=extra distance travel time formovement 1−3 (s/veh),TTLBP=trav
flow speed (s/veh), andTTLB=hypothetical travel time along pathLB if a 90-
degree surface intersectionwere in place, also at free-flowspeed (s/veh).

Obviously, to evaluate this parameter, information on the distances involved between these points and the free-flow speed would have to be available. As an example, if the distance LPB were 1,000 feet, and the free-flow speed was 25 mi/h, the TT_{LPB} would be 1,000/(1.47×25)=27.2 s. The hypothetical path LB would be shorter, as a simple 90-degree left turn would be made to accomplish the path. For this example, assume that the distance is 700 feet. The free-flow speed might be somewhat lower due to the dynamics of a 90-degree left turn—assume 20 mi/h. Then, TT_{LB} would equal 700/(1.47×20)=23.8 s. The EDTT for movement 1–3 is then 27.2−23.8=3.4 s/veh.

In a sense, the EDTT can be thought of as "geometric delay," that is, delay caused by a geometry, which forces the movement into a longer travel path.

An EDTT can also be negative when a geometry provides a shorter travel path for a given movement. Consider the right turn movement O_1 to D_4.

- The actual path for movement 1–4 is L,H.
- If a 90-degree surface intersection were in place, the path would be L, D, P, H.
- The EDTT for movement 1–4 is TTLH–TTLDPH.

In this case, TT_{LH} will be *less than* TT_{LDPH}. If the distance from L to H for the diamond interchange is 950 feet with the free-flow speed 25 mi/h, TT_{LH} would be 950/(1.47×25)=25.9 s/veh. If a hypothetical surface intersection were in place, the distance between L and H would be longer, and the free-flow speed would be somewhat less, given that a 90-degree right turn would be required. If the distance is 1,100 feet and the free-flow speed is 20 mi/h, then TT_{LDPH} would equal 1,100/(1.47×20)=37.4 s/veh. The $EDTT_{1,4}$ is then 25.9−37.4=−11.5 s/veh.

In the case of the right turn, the diamond ramp moves vehicles closer to their desired destination than if a surface intersection was in place. The geometry *reduces* travel time for this movement.

This is the essence of ETT as a service measure for level of service at interchanges and alternative intersections. Geometric delay is added to control delay at each junction a movement passes through. Various interchange and alternative intersection forms influence both of these components: the specific geometry alters travel distances for many movements, and the number of junctions through which each movement must pass is also potentially altered. In the case of signals, various geometries produce different sets of signalized intersection movements, which will have a significant impact on control delays at each junction.

The criteria for level of service determination are given in Table 26.1. The criteria are similar to, but not the same as, those for signalized intersections:

- The criteria for interchanges represent considerable higher thresholds than for signalized intersections. This reflects the reality that, in many instances, movements may have to traverse more than one intersection to complete their maneuver.

- The criteria for alternative intersections are exactly the same as for signalized intersections. This reflects the fact that the primary junction is still a single intersection. Even where alternative designs create additional intersections, which may also be signalized, performance should match levels required of a typical signalized intersection.

- The criteria for interchanges with roundabouts are similar to, but not the same as, signalized intersections. Since roundabouts are intended to reduce delays by eliminating the need for signals, the maximum delay permitted is somewhat less than for signalized intersections.

Table 26.1: Levels of Service for Interchanges and Alternative Intersections

ETT (s/veh)	Condition		
	$v/c \le 1$ and $R_Q \le 1$ for Every Lane Group	$v/c > 1$ for Any Lane Group	$R_Q > 1$ for Any Lane Group
≤15	A	F	F
>15–30	B	F	F
>30–55	C	F	F
>55–85	D	F	F
>85–120	E	F	F
>120	F	F	F

(a) LOS Criteria for Interchanges with Signalized Intersections

26.3-1 Full Alternative Text

ETT (s/veh)	Condition		
	$v/c \leq 1$ and $R_Q \leq 1$ for Every Lane Group	$v/c > 1$ for Any Lane Group	$R_Q > 1$ for Any Lane Group
≤10	A	F	F
>10–20	B	F	F
>20–35	C	F	F
>35–55*E*	D	F	F
>55–80	E	F	F
>80	F	F	F

(b) LOS Criteria for Alternative Intersections

26.3-2 Full Alternative Text

ETT (s/veh)	Condition		
	$v/c \leq 1$ and $R_Q \leq 1$ for All Approaches	$v/c > 1$ for Any Approach	$R_Q > 1$ for Any Approach
≤15	A	F	F
>15–25	B	F	F
>25–35	C	F	F
>35–50	D	F	F
>50–75	E	F	F
>75	F	F	F

(c) LOS Criteria for Interchanges with Roundabouts

26.3-3 Full Alternative Text

(Source: Reprinted with permission from *Highway Capacity Manual, 6th Edition: A Guide for Multimodal Mobility Analysis*, Transportation Research Board, National Academy of Sciences, Courtesy of the National Academies Press, Washington, D.C., 2016.)

Note that any case in which a lane-group junction operates at a v/c>1.00, LOS F is assigned to *all* movements that move through that junction and lane group, including links for the movement that might otherwise be characterized as a better LOS.

Similarly, if any lane-group link operates with a queue ratio (R_Q) of greater than 1.00, then all links for that movement are characterized as LOS F. The queue ratio is the number of vehicles queued on the link under described demands divided by the maximum number of queued vehicles permitted by physical limitations.

In all cases, the criteria apply to each movement (O–D pair) in the interchange or alternative intersection. Averages, weighted by demand flow rate, can be computed for approaches and for the interchange or intersection as a whole to evaluate overall levels of service. User caution is advised, however, as the breakdown of a single element can influence operations across a much larger area.

26.3.3 Interchanges: Changes in Saturation Flow Rate Estimation

For alternative intersection geometries, there are no changes to saturation flow rate computations associated with signalized elements of the intersection. That is, the saturation flow rate for each lane group at each signalized element follows the normal procedure for signalized intersections, as described in Chapter 22.

For interchanges with signalized elements, however, this is not necessarily true. There are several unique aspects of signalized interchanges that require modifications to the usual procedures for estimating saturation flow rates. For lane groups at a signalized interchange, the saturation flow rate is estimated as:

$$s = s_o\ N\ f_w\ f_{HVg}\ f_p\ f_a\ f_{RT}\ f_{LT}\ f_{Rpb}\ f_{Lpb}\ f_v\ f_{LU}\ f_{DDI} \quad [26\text{-}3]$$

Of the eleven adjustment factors (f_i), two are unique to certain types of signalized interchanges and three others are modified from their normal signalized intersection use at some interchanges. The terms of Equation 26-3 are defined as follows:

- s=saturation flow rate for a lane group, (veh/hg),
- s_o=base saturation flow rate for a lane group, (pc/hg/ln), obtained from Chapter 22,
- N=number of lanes in the lane group,
- f_w=adjustment factor for lane width, obtained from Chapter 22,
- f_{HVg}=adjustment factor for heavy vehicles and grade, obtained from Chapter 22,
- f_p=adjustment factor for parking activity, obtained from Chapter 22,
- f_a=adjustment factor for area type, obtained from Chapter 22,
- f_{RT}=adjustment factor for right turns, obtained from Chapter 22 or modified as indicated herein,
- f_{LT}=adjustment factor for left turns, obtained from Chapter 22 or modified as indicated herein,
- f_{Rpb}=adjustment factor for pedestrian and bicycle interference with right turns, obtained from Chapter 22,
- f_{Lpb}=adjustment factor for pedestrian and bicycle interference with left turns, obtained from Chapter 22,
- f_v=adjustment factor for traffic pressure at signalized interchanges,
- f_{LU}=adjustment factor for lane utilization, obtained from Chapter 22 or modified as indicated herein, and,
- f_{DDI}=adjustment factor for *DDI* interchanges.

Adjustment Factor for Traffic Pressure, f_v

The influence of traffic pressure on saturation flow rates have been demonstrated in the context of signalized intersections [9]. This adjustment reflects observations that during periods of high demand and congestion, drivers become more aggressive. This is reflected in higher than normal saturation flow rates for affected lane groups. Although many traffic engineers believe that this phenomenon exists at all signalized intersections, it has not been adequately documented outside of interchange environments. The factor is computed as follows:

$$f_v = 11.07 - P \times \min(v'_i, 30) \quad [26\text{-}4]$$

where:

f_v = adjustment factor for traffic pressure, P = traffic pressure parameter (0.00672 for left turns, 0.00486 for through movements and right turns), and v'_i = traffic flow rate movement i, expressed as veh/cycle/ln.

Where a lane group is shared by left turns and through/right turn movements, the factor is taken as the weighted average based on relative flow rates of the various movements. Note, however, that the parameter v'_i reflects the total flow rate in the lane group.

Left turns are more affected by traffic pressure than through vehicles or right turns. The factor can be greater than, equal to, or less than 1.00, depending upon the applicable flow rate.

Consider the example of a two-lane left turn lane group serving a demand of 2,200 veh/h, controlled by a signal with a 60 s cycle length. What would the appropriate value of f_v be for this lane group?

Equation 26-4 is used with $P=0.00672$ (left turns). With a 60 s cycle length, there are $3600/60=60$ cycles in the hour. Given two lanes, the value of v'_i is $2{,}200/(60\times2)=18.3$, say 18 vehicles/cycle/ln. Then:

$$f_v = 11.07 - 0.00672 \times 18 = 11.07 - 0.12 = 1.05$$

In this case, traffic is sufficiently intense to have a positive impact on the saturation flow rate that results. Equation 26-4 is applied to *all* interchange approach lane groups at a signalized interchange.

Modification of Lane Utilization Adjustment Factors, f_{LU}

Special models have been developed for the lane-utilization adjustment for external arterial approaches to signalized interchanges with *two* signalized junctions on the arterial. The types of interchanges that have this property are diamond interchanges, some PARCLO interchanges, and DDI interchanges. For all other approaches and types of interchanges, f_{LU} is computed as described in Chapter 22.

In general, f_{LU} is computed as follows:

fLU=1PLmax N [26-5]

where:

fLU=adjustment for lane utilization,PLmax=proportion of approach flow in

The HCM provides equations for the estimation of lane distribution at external arterial approaches to a two-intersection interchange. There are two equations, one for diamond and PARCLO interchanges and the other for DDI interchanges. For diamond and PARCLO interchanges:

PLi=1N+a1 (vRvL+vT+vR)+a2 (vLvL+vT+vR) +a3 (D×vL106) [26-6]

where:

PLi=proportion of traffic in lane i of the externalarterial approach.N=numbe
0 if there is an exclusive RT lane on the external approach
), veh/h,vL=O–D demand flow rate traveling through the firstintersection an

Equation 26-6 is valid only for D<800 feet; otherwise use f_{LU} as described in Chapter 22.

Coefficient a_i depends upon the type of interchange (only those with two signalized intersections are included), as shown in Table 26.2.

Table 26.2: Coefficients a_i for Equation 26-6 for Diamond and PARCLO Interchanges

Interchange Type	Number of Lanes in Lane Group	Leftmost Lane (L_1)			Rightmost Lane (L_n)		
		a_1	a_2	a_3	a_1	a_2	a_3
Diamond	2	−0.154	0.187	−0.181	-	-	-
	3	−0.245	0.465	0	0.609	−0.326	0
	4	−0.328	0.684	0	0.640	−0.233	0
Parclo A-2Q	2	0	−0.527	0	-	-	-
	3	0	−0.363	0	0	0.605	0
	4	0	−0.257	0	0	0.747	0
Parclo B-2Q,B-4Q, AB-4Q (WB)	2	0.387	−0.344	0	-	-	-
	3	0.559	−0.218	0	−0.429	0.695	0
	4	0.643	−0.103	0	−0.359	0.794	0
Parclo A-4Q, AB-2Q (EB), AB-4Q (EB)	2	−0.306	−0.484	0	-	-	-
	3	−0.333	−0.289	0	0.579	0.428	0
	4	−0.233	−0.237	0	0.703	0.641	0
Parclo AB-2Q (WB)	2	0.468	0	0	-	-	-
	3	0.735	0	0	−0.308	0	0
	4	0.768	0	0	−0.202	0	0

(Source: Reprinted with permission from *Highway Capacity Manual, 6th Edition: A Guide for Multimodal Mobility Analysis*, Transportation Research Board, National Academy of Sciences, Courtesy of the National Academies Press, Washington, D.C., 2016.)

Table 26.2: Full Alternative Text

Note that coefficients are provided only for the leftmost lane and the rightmost lane of the lane group. If there are only two lanes in the group, then coefficients are provided only for the leftmost lane. Assuming that the leftmost lane is lane *l* and lane *n* is the rightmost lane, all lane use proportions are found as follows:

For N=2: PL2=1−PL1For N=3: PL2=1−PL1−PL3For N=4: PL2=PL3=1−
[26-7]

This guarantees that the proportion of vehicles in *all* lanes of the lane group adds up to 1.00. For use in Equation 26-5, P_{Lmax} is the proportion of traffic in the high-usage lane, as indicated by Equation 26-7.

Consider a three-lane external approach (no exclusive lanes) to a diamond interchange with the following characteristics:

- Total approach flow rate=2,000 veh/h,
- *D*, the distance between intersections, is 600 feet, and
- Of the external approaching flow in the lane group, 30% will turn left at the second intersection, 0% will turn right at the second intersection (there are no right turns at the second intersection of a diamond interchange), and 70% will continue through the second intersection.

What lane-use adjustment factor should be used for this external approach lane group?

Equation 26-6 must be used to find the proportion of traffic in the leftmost (lane 1) and rightmost (lane 3) of the lane group.

PLi=1N+a1 (vRvL+vT+vR)+a2 (vLvL+vT+vR)+a3 (D×vL106)

For the subject problem:

N=3 lanesvL=2,000×0.30=600 veh/hvR=2,000×0.00=0 veh/hvT=2,000×0.7

Coefficient a_i is drawn from Table 26.2 for a diamond interchange with a three-lane external approach, and is shown in Table 26.3.

Table 26.3: Coefficients for Example Calculation

Lane	a_1	a_2	a_3
Leftmost (Lane 1)	−0.245	0.465	0.000
Rightmost (Lane 3)	0.609	−0.326	0.000

Table 26.3: Full Alternative Text

Then:

PL1=13-
0.245 (0600+1,400+0)+0.465 (600600+1,400+0)+0.000 (600×600106)=0.1
0.193-0.235=0.572

From this forecast distribution, it is clear that PLmax=0.572 (lane 2), and the lane utilization adjustment factor can be computed from Equation 26-5:

fLU=1PLmax N=10.572×3=0.583

The HCM also provides models to estimate f_{LU} at DDIs based upon a research study [10]. The equation for lane distribution (on external arterial approaches) is:

PLiDDI=a1 LTDR+a2 [26-8]

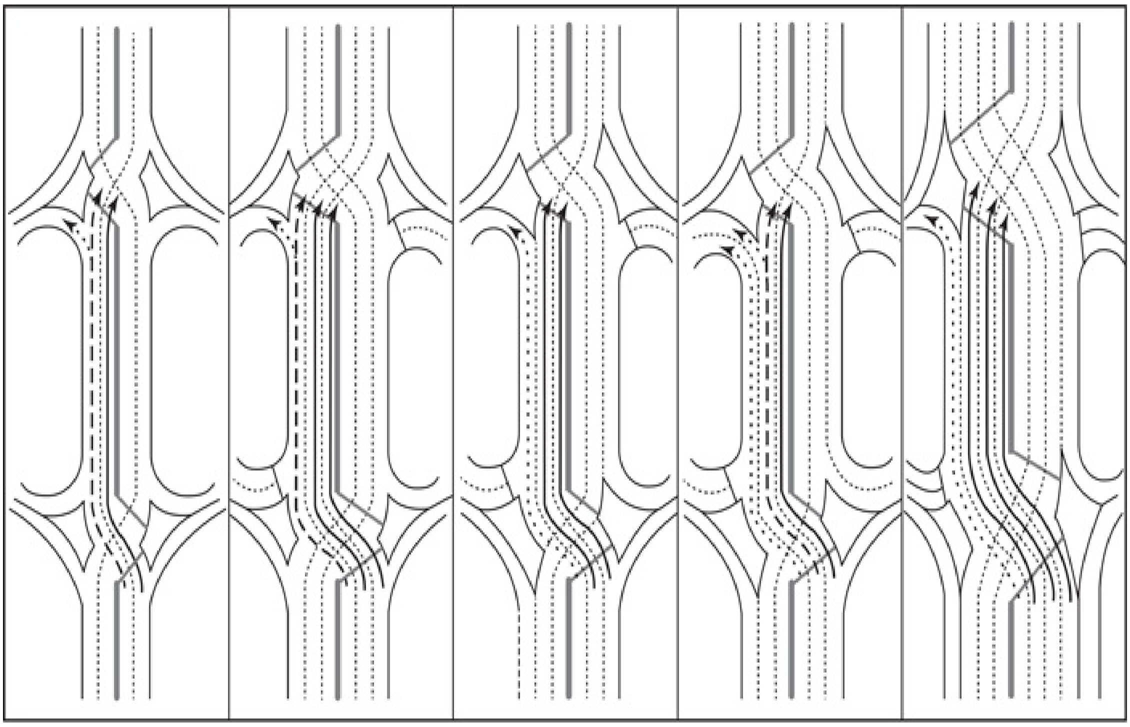
where:

PLiDDI=proportion of lane group flow in lane i at a DDI,ai=coefficients (Se
turn demand ratio; left-
turn demand atsecond intersection divided by the total lanegroup demand flo

Note that, as in the case of diamond and PARCLO interchanges, the subject lane group is the external arterial approach, and that left turns are made at the second, or downstream intersection.

Figure 26.18 shows five different lane-use configurations that occur in DDIs, and the coefficient a_i that would be used in each case in Equation 26-8.

Figure 26.18: Coefficients a_i for DDIs



2-Lane Shared LT Lane	3-Lane Shared LT Lane	3-Lane Exclusive LT Lane	3-Lane Exclusive and Shared LT Lane	4-Lane Exclusive LT Lane

Lane Configuration	Regime	Lane	a_1	a_2
2-lane shared	I (LTDR ≤ 0.35)	Left	0.2129	0.5250
	II (LTDR > 0.35)	Left	0.5386	0.4110
3-lane shared	I-1 (LTDR ≤ 0.13)	Middle	−0.1831	0.3863
	I-2 (0.13 < LTDR ≤ 0.43)	Leftmost	0.2245	0.3336
	II (LTDR > 0.43)	Leftmost	0.6460	0.1523
3-lane exclusive	I (LTDR ≤ 0.33)	Middle	−0.5983	0.5237
	II (LTDR > 0.33)	Leftmost	0.9695	0.0096
3-lane exclusive with middle shared lane	I (LTDR ≤ 0.50)	Middle	−0.2884	0.5626
	II (LTDR > 0.50)	Leftmost	0.4903	0.1761
4-lane exclusive	I (LTDR ≤ 0.35)	Center-left	−0.5432	0.5095
	II (LTDR > 0.35)	Leftmost	0.9286	−0.0071

(Source: Reprinted with permission from *Guide for Multimodal Mobility Analysis*, Transportation Research Board, National Academy of Sciences, Courtesy of the National Academies Press, Washington, D.C., 2016.)

Figure 26.18: Full Alternative Text

Figure 26.18 only predicts the lane occupancy in *one* lane for each case. The table was calibrated such that this lane *is* the maximum flow lane, and occupancies for the other lanes are not needed.

Adjustment Factor for DDIs (f_{DDI})

Research has demonstrated that the model for diamond interchanges overestimates saturation flow rates for DDIs [11]. A constant adjustment factor of 0.913 is applied to all approaches at the two crossover intersections of a DDI.

Adjustment to f_{LT} and f_{RT} for Turning Radius at Interchanges

Saturation flow rates at interchanges are affected by the turning radius of right- and left-turn movements [12]. Normally applicable adjustment factors for right (f_{RT}) and left turns (f_{LT}) are, therefore, modified when applied to signalized interchanges.

An adjustment that accounts for the effects of turning radius is computed as:

fR=11+(5.61R) [26-9]

where:

fR=intermediate adjustment factor accounting for theeffects of turning radiu

This intermediate adjustment factor is then used to modify f_{LT} and f_{RT} as described in Chapter 22. Table 26.4 shows the modifications.

Table 26.4: Modifications to Left-Turn and Right-Turn Adjustment Factors at Signalized Interchanges

Type of Turning Movement	Left Turns	Right Turns
Protected Exclusive Turning Lane	$f_{LT} = f_R$	$f_{RT} = f_R$
Protected Shared Turning Lane	$f_{LT} = \dfrac{1}{1 + P_{LT}\left(\dfrac{1}{f_R} - 1\right)}$	$f_{LT} = \dfrac{1}{1 + P_{RT}\left(\dfrac{1}{f_R} - 1\right)}$
Permitted Exclusive or Shared Turning Lane	$f_{LT} = f_{LT}$ No modification.	$f_{RT} = f_{RT}$ No modification.

Table 26.4: Full Alternative Text

26.3.4 Interchanges: Other Modifications to Signalized Intersection Analysis

The HCM methodology for signalized interchange analysis incorporates a variety of additional modifications to the normal signalized intersection model (Chapter 22) that address the impact of interchange operations on the incorporated individual signals.

These involve primarily modifications to lost time estimates, which impacts effective green times. These modifications, which are not detailed here, involve three basic conditions:

- Discharge from some approaches can be blocked or hindered by the existence of a queue from the downstream intersection.
- Green time at the downstream intersection may not be fully utilized if demand is constrained by the upstream signal (demand starvation).
- Overlapping signal sequences can impact effective green times at DDIs.

Consult the HCM [1] directly for the details of these adjustments.

26.4 Closing Comments

This chapter has presented an overview of interchanges and alternative intersections that are in common use. These represent geometric design alternatives that can assist in simplifying the control of individual movements, particularly where they include traffic signals. The general process for level of service analysis has been discussed, although the HCM must be consulted directly for documentation of all components of these complex methodologies.

Few interchanges or alternative intersections will be analyzed by hand, and either software implementing the HCM or a variety of applicable simulation packages will generally be used for this purpose.

Nevertheless, traffic engineers should be aware of the basic issues involved in interchange and alternative intersection design so that they can reasonably consider these alternatives where a need exists.

References

- 1. *Highway Capacity Manual, 6th Edition: A Guide for Multimodal Mobility Analysis*, Transportation Research Board, Washington, D.C., 2016.

- 2. *A Policy on Geometric Design of Highways and Streets*, 6th Edition, American Association of State Highway and Traffic Officials, Washington, D.C., 2011.

- 3. Hughes, W., Jagannathan, R., Sengupta, D., and Hummer, J., “Alternative Intersections/Interchanges: Informational Report,” *Publication Number FHWA-HRT-090060*, Federal Highway Administration, Washington, D.C., April 2009.

- 4. Jones, E., and Selinger, M., “A Comparison of the Operations of Single Point and Tight Urban Diamond Interchanges,” *Transportation Research Record 1847*, Transportation Research Board, Washington, D.C., 2003.

- 5. Messer, C., Bonneson, J., Anderson, S., and McFarland, W., “Single Point Urban Interchange Design and Operations Analysis,” *NCHRP Synthesis 345*, Transportation Research Board, Washington, D.C., 1991.

- 6. Hummer, J., Ray, B., Daleiden, A., Jenior, P., and Knudsen, J., “Restricted Crossing U-Turn Informational Guide,” *Publication Number FHWA-SA-14-070*, Federal Highway Administration, Washington, D.C., 2014.

- 7. Hummer, J., “Proposed MUT and RCUT HCM Procedure Inputs, U-Turn Crossover Saturation Flow Adjustment Factor,” *Draft Technical Memo*, Wayne State University, Detroit MI, June 21, 2014.

- 8. Hummer, J., Blue, V., Cate, J., and Stephenson, R., “Taking Advantage of the Flexibility Offered by Unconventional Designs,” *ITE Journal*, Vol. 82, No. 9, Institute of Transportation Engineers, Washington, D.C., September 2012.

- 9. Elefteriadou, L., Fang, C., Roess, R., Prassas, E., Yeon, X., Kandyli, K., Wang, H., and Mason, J., “Capacity and Quality of Service of Interchange Ramp Terminals,” *Final Report*, NCHRP Project 3-60, Pennsylvania State University, University Park PA, March 2005.

- 10. Yeom, C., Schroeder, B., Cunningham, C., Vaughan, C., Rouphail, N., and Hummer, J., “Lane Utilization at Two-Lane Arterial Approaches to Double Crossover Diamond Interchanges,” *Transportation Research Record 2461*, Transportation Research Board, Washington, D.C., 2014.

- *11. EDC-2 Intersection and Interchange Geometrics Web-Site*, Federal Highway Administration, Washington, D.C., www.fhwa.dot.gov/everydaycounts/edctwo/2012 .

- 12. Messer, C. and Bonneson, J., “Capacity of Interchange Ramp Terminals,” *Final Report*, NCHRP Project 3-47, Texas A & M Research Foundation, College Station TX, April 1997.

Problems

1. 26-1. At an urban interchange between a freeway and major arterial, major movements involve heavy left-turning flows from both freeway directions. If existing buildings make it difficult to acquire significant increases to the right-of-way, what type of interchange would you recommend, and why?

2. 26-2. Discuss the advantages of a split diamond interchange with two one-way arterials as opposed to a standard diamond interchange with a single two-way arterial.

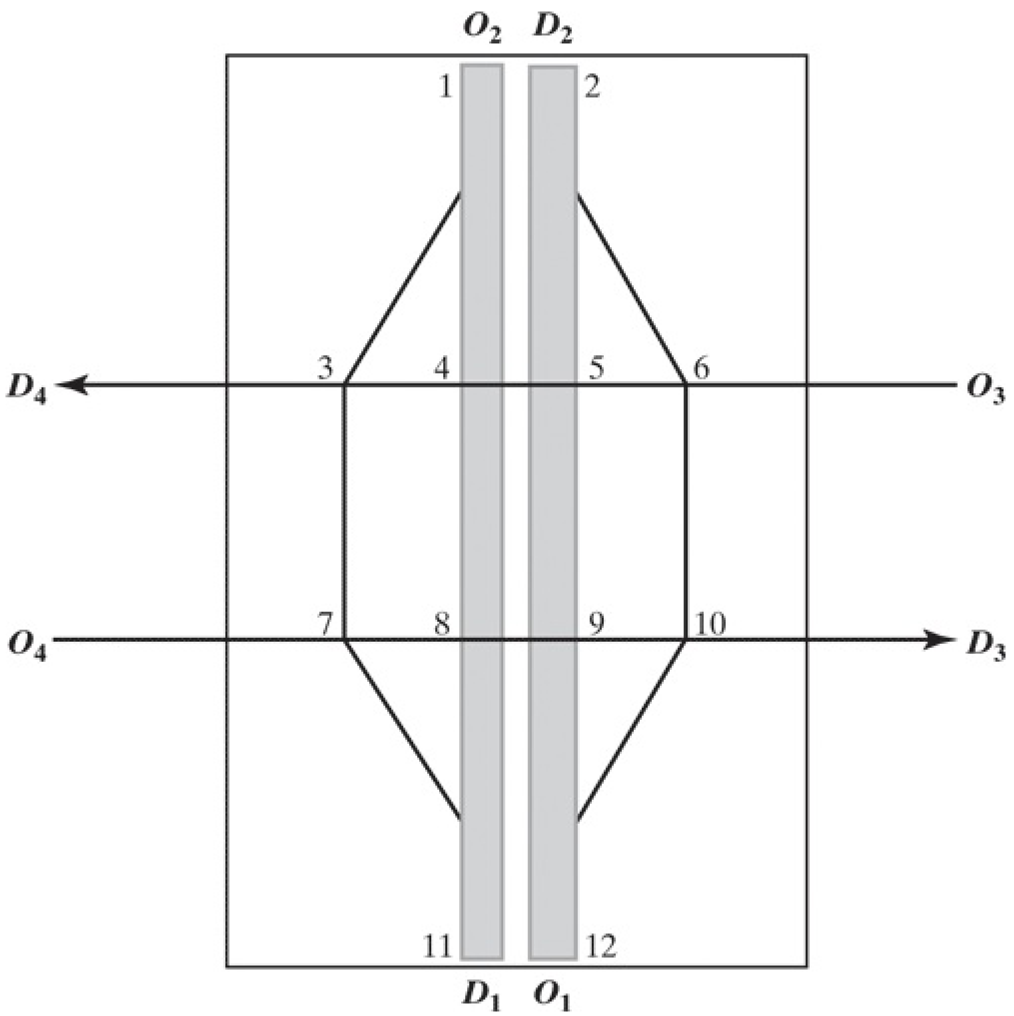
3. 26-3. Consider the following split diamond interchange with two one-way arterials:

   Relevant link lengths are as follows:

   - Ramps 1 to 3, 7 to 11, 6 to 2, and 12 to 10=1,000 feet
   - Links 3 to 7, 4 to 8, 9 to 5, and 10 to 6=1,200 feet
   - Links 6 to 5, 4 to 3, 7 to 8, and 9 to 10=200 feet
   - Freeway links 1 to 4, 2 to 5, 7 to 11, and 9 to 12=980 feet
   - Arterial links 5 to 4 and 8 to 9=240 feet

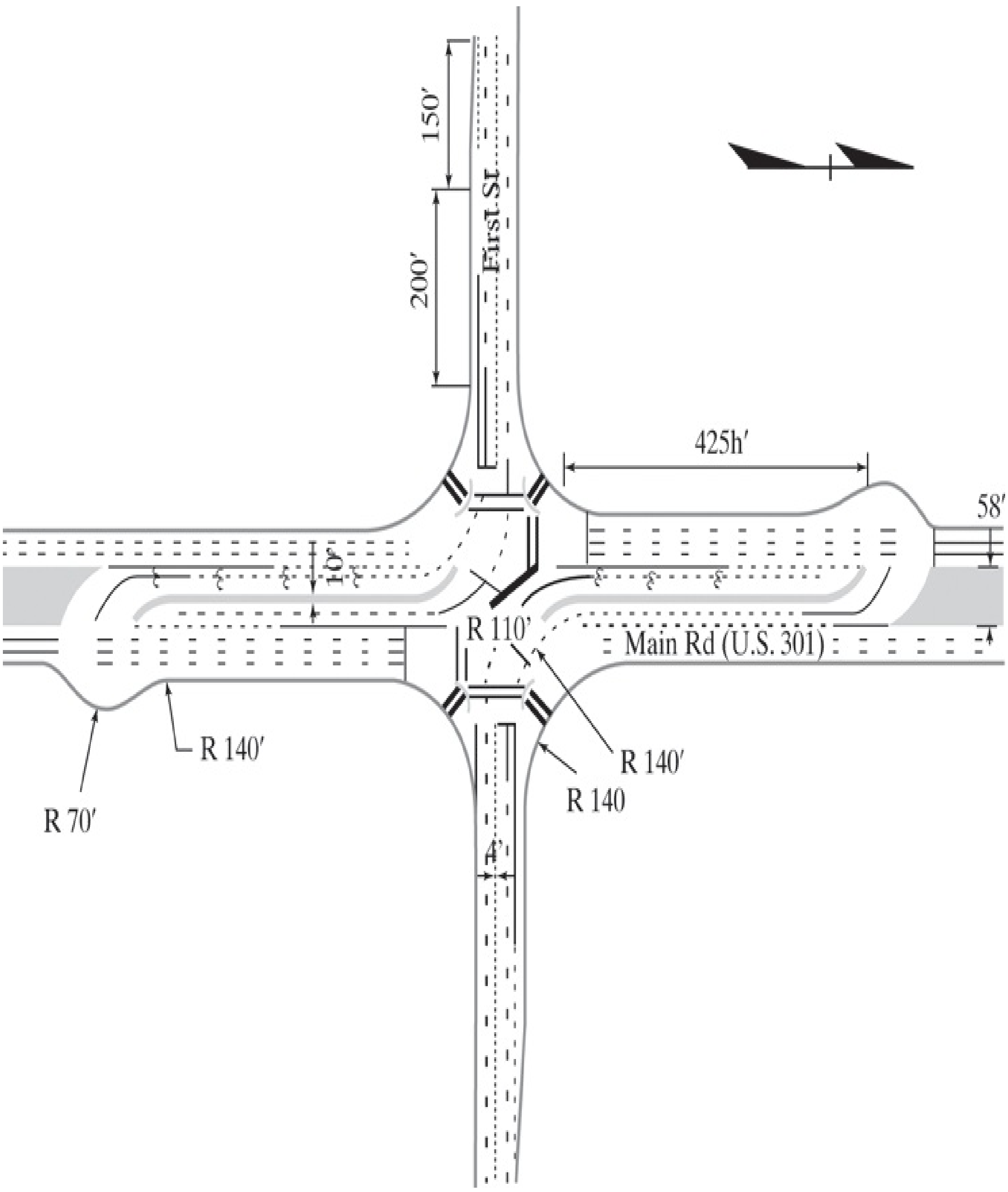
   Field measurements indicate that all movements through this interchange experience 24.6 s/veh of control delay at *each* signalized intersection they pass through. Field measurements also indicate that free-flow speeds on ramps is 30 mi/h, and on the arterial is 40 mi/h.

   What is the level of service provided to Movements O_1 to D_4 and O_1 to D_3?



26.2-8 Full Alternative Text

4. 26-4. Consider the following RCUT intersection:



(Source: “Restricted Crossing U-Turn Intersection,” *TechBrief*, Federal Highway Administration, Washington, D.C., October 2009, Figure 2.)

[26.2-9 Full Alternative Text]

In addition to the information shown on the diagram, the following is known:

- The primary intersection is signalized. Average control delay for first street right-turning vehicles 22.0 s/veh. Average delay for main street through vehicles is 10.0 s/veh, and for right-turning vehicles, 11.5 s/veh.

- The arterial is 112 feet wide, curb to curb.
- U-turn lanes are not signalized. Average travel time for a vehicle to complete the U-turn is 8.2 s/veh.
- Free-flow speed is 30 mi/h on First Street and 45 mi/h on the main street.

What is the level of service provided to first street through and left-turning vehicles?

5. 26-5. A two-lane external approach to a two-intersection PARCLO B-2Q interchange has a demand flow rate of 1,300 veh/h that will continue to the second intersection. Of these, 40% will turn left at the second intersection, and 60% will continue through the second intersection. The distance between the two intersections is 850 feet. What lane utilization adjustment factor would be applied to the external intersection?

6. 26-6. A DDI has three lanes, of which the left lane is exclusively for left turns at the internal crossover. The total approach demand at the external crossover is 2,800 veh/h. Of this, 1,200 veh/h are turning left at the internal crossover. What lane utilization adjustment factor would be applied to the external crossover approach?

7. 26-7. An off ramp in a SPUI is heavily used, and has a peak flow rate of 1,000 veh/h. Right turns and left turns have separate lanes. Right turns are controlled with a YIELD sign. Left turns are, of course, signalized, and are made on a radius of 250 feet. What left-turn adjustment factor should be applied to the signalized left-turn lane?

Part IV Uninterrupted Flow Facilities

Design, Control, and Level of Service

Chapter 27 An Overview of Geometric Design of Roadways

27.1 Introduction to Highway Design Elements

Highways are complex physical structures involving compacted soil, sub-base layers of aggregate, pavements, drainage structures, bridge structures, and other physical elements. From an operational viewpoint, however, it is the geometric characteristics of the roadway that primarily influence traffic flow and operations. Three main elements define the geometry of a highway section:

- Horizontal alignment
- Vertical alignment
- Cross-sectional elements

Virtually all standard practices in geometric highway design are specified by the American Association of State Highway and Transportation Officials (AASHTO) in the current version of *A Policy on Geometric Design of Highways and Streets* [1]. The latest edition of this key reference (at this writing) was published in 2011, and is often referred to as the "Green Book," due to the predominant color of its cover.

This chapter presents an overview of the subject of roadway geometry. While it includes important material, the AASHTO policy should be consulted directly for greater detail.

27.1.1 Horizontal Alignment

The horizontal alignment refers to the plan view of the highway. The horizontal alignment includes tangent sections and the horizontal curves and other transition elements that join them.

Highway design is generally initiated by laying out a set of tangents on topographical and development maps of the service area. Selection of an appropriate route and specific location of these tangent lines involves many considerations and is a complex task. Some of the more important considerations include the following:

- Forecast demand volumes, with known or projected origin-destination patterns
- Patterns of development
- Topography
- Natural barriers
- Subsurface conditions
- Drainage patterns
- Economic considerations
- Environmental considerations
- Social considerations

The first two items deal with anticipated demand on the facility and the specific origins and destinations that are to be served. The next four are important engineering factors that must be considered. The last three are critically important. Cost is always an important factor, but it must be compared with quantifiable benefits.

Environmental impact statements are required of virtually all highway projects, and much effort is put into providing remedies for unavoidable negative impacts on the environment. Social considerations are also important and cover a wide range of issues. It is particularly important that highways be built in ways that do not disrupt local communities, either by dividing them, enticing unwanted development, or causing particularly damaging environmental impacts. Although this text does not deal in detail with this complex process of decision making, the reader should be aware of its existence and of the influence it has on highway programs in the United States.

27.1.2 Vertical Alignment

Vertical alignment refers to the design of the facility in the profile view. Straight grades are connected by vertical curves, which provide for transition between adjacent grades. The “grade” refers to the longitudinal slope of the facility, expressed as “feet of rise or fall” per “longitudinal foot” of roadway length. As a dimensionless value, the grade may be expressed either as a decimal or as a percentage (by multiplying the decimal by 100).

In vertical design, attempts are made to conform to the topography, wherever possible, to reduce the need for costly excavations and landfills as well as to maintain aesthetics. Primary design criteria for vertical curves include the following:

- Provision of adequate sight distance at all points along the profile
- Provision of adequate drainage
- Maintenance of comfortable operations
- Maintenance of reasonable aesthetics

The specifics of vertical design usually follow from the horizontal route layout and specific horizontal design. The horizontal layout, however, is often modified or established in part to minimize problems in the vertical design.

27.1.3 Cross-Sectional Elements

The third physical dimension, or view, of a highway that must be designed is the cross-section. The cross-section is a cut across the plane of the highway. Within the cross-section, such elements as lane widths, superelevation (cross-slope), medians, shoulders, drainage, embankments (or cut sections), and similar features are established. As the cross-section may vary along the length of a given facility, cross-sections are generally designed every 100 feet along the facility length and at any other locations that form a transition or change in the cross-sectional characteristics of the facility.

27.1.4 Surveying and Stationing

In the field, route surveyors define the geometry of a highway by “staking” out the horizontal and vertical position of the route every 100 feet. The cross-section is similarly staked out every 100 feet.

While this text does not deal with the details of route surveying, it is useful to understand the conventions of “stationing” which are used in the process. Stationing of a new or reconstructed route is generally initiated at the western or northern end of the project. “Stations” are established every 100 ft, and are given the notation xxx+yy. Values of *xxx* indicate the number of hundreds of feet of the location from the origin point. The *yy* values indicate intermediate distances of less than 100 ft.

Regular stations are established every 100 ft, and are numbered 0+00,100+00,200+00, etc. Various elements of the highway are “staked” by surveyors at these stations. If key points of transition occur between full stations, they are also staked and would be given a notation such as 1200 + 52, which signifies a location 1,252 ft. from the origin. This notation is used to describe points along a horizontal or vertical alignment in subsequent sections of this chapter.

Reference [2] is a text in route surveying, which can be consulted for more detailed information on the subject. A number of other texts on the subject are also available.

27.2 Horizontal Alignment of Highways

The horizontal alignment of a highway is its path in a plan view of the surrounding terrain. Horizontal curves have critical geometric properties and characteristics that should be well understood. They are also subject to standard design criteria established by AASHTO, or by local and state highway agencies

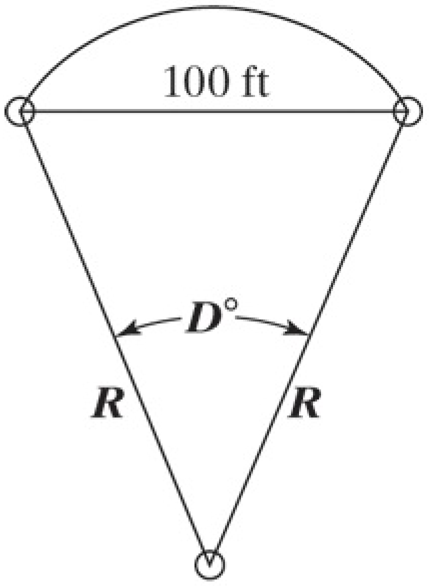
27.2.1 Quantifying the Severity of Horizontal Curves: Radius and Degree of Curvature

All horizontal curves are circular, that is, formed by an arc with a constant radius. Compound horizontal curves may be formed by consecutive horizontal curves with different radii. On high-speed, high-type facilities, a horizontal curve and a tangent (straight) segment are often joined by a spiral transition curve, which is a curve with a varying radius that starts at ∞ at the connection to the tangent segment, and ends at the radius of the circular curve at the connection to the curve.

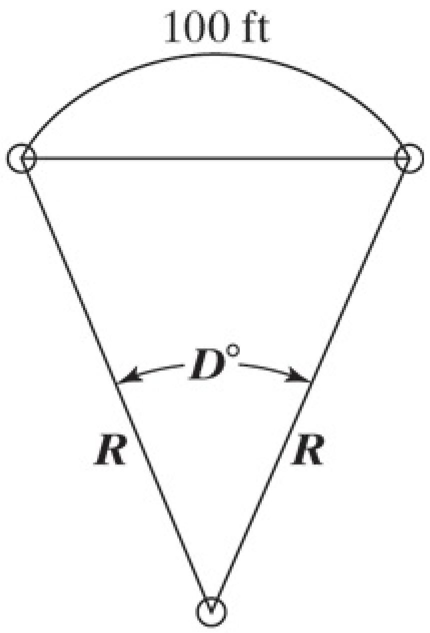
The severity of a circular horizontal curve is measured by the *radius* or by the *degree of curvature*, which is a related measure. Degree of curvature is most often used, as higher values depict sharper, or more severe, curves. Conversely, larger radii depict less severe curves.

Figure 27.1 illustrates two ways of defining degree of curvature. The *chord definition* is illustrated in Figure 27.1 (a). The degree of curvature is defined as the central angle subtending a 100 ft chord on the circular curve. The *arc definition* is illustrated in Figure 27.1 (b), and is the most frequently used. In this definition, the degree of curvature is defined as the central angle subtending a 100 ft arc.

Figure 27.1: Definitions of Degree of Curvature



(a) Chord Definition



(b) Arc Definition

Using the arc definition, it is possible to derive the relationship between the radius (R) and the degree of curvature (D). The ratio of the circumference of the circle to 360° is set equal to the ratio of 100 ft to $D°$.

Then:

$$\frac{2\pi R}{360}=\frac{100}{D} \quad D=\frac{360\times 100}{2\pi R}=\frac{18,000}{\pi R}$$

Noting that $\pi=3.141592654$ ..., then:

$$D=\frac{18,000}{3.141592654R}=\frac{5,729.58}{R} \quad [27\text{-}1]$$

where:

D=degree of curvature, and R=radius of curvature, ft.

Thus, for example, a circular curve with a radius of 2,000 ft has a degree of curvature of:

$$D=\frac{5,729.58}{2,000}=2.915^{o}$$

It should be noted that for up to 4° curves, there is little difference between the arc and the chord definition of degree of curvature. This text, however, will use only the arc definition illustrated in Figure 27.1 (b).

27.2.2 Review of Trigonometric Functions

The geometry of horizontal curves is described mathematically using trigonometric functions. A brief review of these functions is included as a refresher for those who may not have used trigonometry for some time. Figure 27.2 illustrates a right triangle, from which the definitions of trigonometric functions are drawn.

Figure 27.2: Trigonometric Functions Illustrated

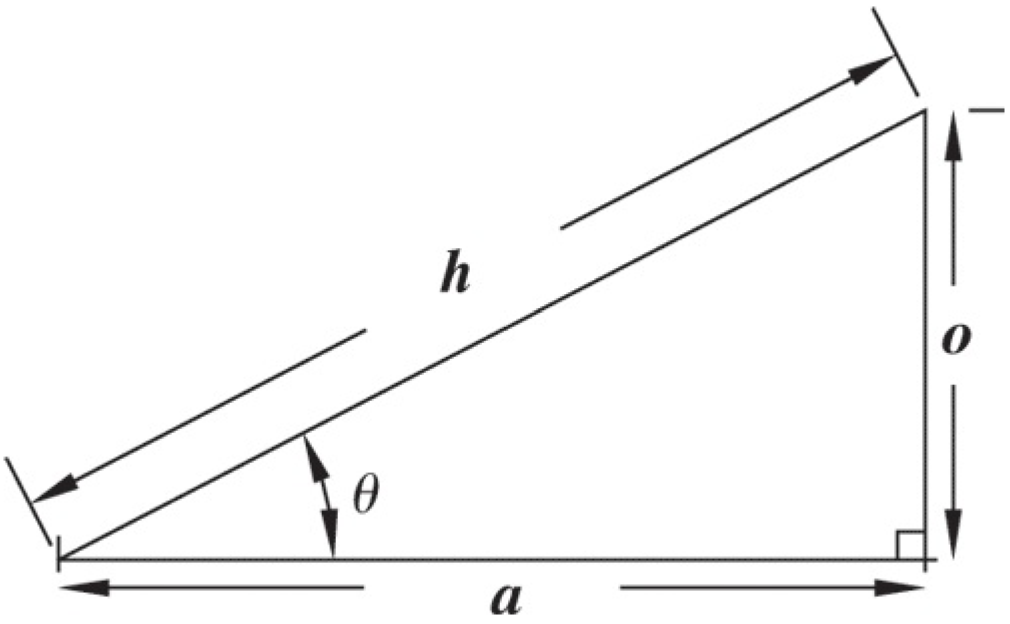


Figure 27.2: Full Alternative Text

In Figure 27.2:

- “*o*” is the length of the opposite leg of the right triangle,
- “*a*” is the length of the adjacent leg of the right triangle, and
- “*h*” is the length of the hypotenuse of the right triangle.

The units of length must be consistent in all cases.

Using the legs of the right triangle, the following trigonometric functions are defined:

- Sine θ=o/h
- Cosine θ=a/h
- Tangent θ=o/a

From these primary functions, several derivative functions are also defined:

- Cosecant θ=1/Sine θ=h/o
- Secant θ=1/Cosine θ=h/a

- Cotangent $\theta = 1/\text{Tangent } \theta = a/o$
- Exsecant $\theta = \text{Secant } \theta - 1$
- Versine $\theta = 1 - \text{Cosine } \theta$

Trigonometric functions are tabulated in many mathematics texts and are generally included on most calculators and in virtually all spreadsheet software. When using spreadsheet software or calculators, the user must determine whether angles are entered in *degrees* or *radians*. In a full circle, there are 2π radians and 360°. Thus, one radian is equal to $360/2(3.141592654) = 57.3°$.

27.2.3 Critical Characteristics of Horizontal Curves

Figure 27.3 depicts a circular horizontal curve connecting two tangent lines. The following points are defined:

Figure 27.3: Characteristics of Circular Horizontal Highway Curves

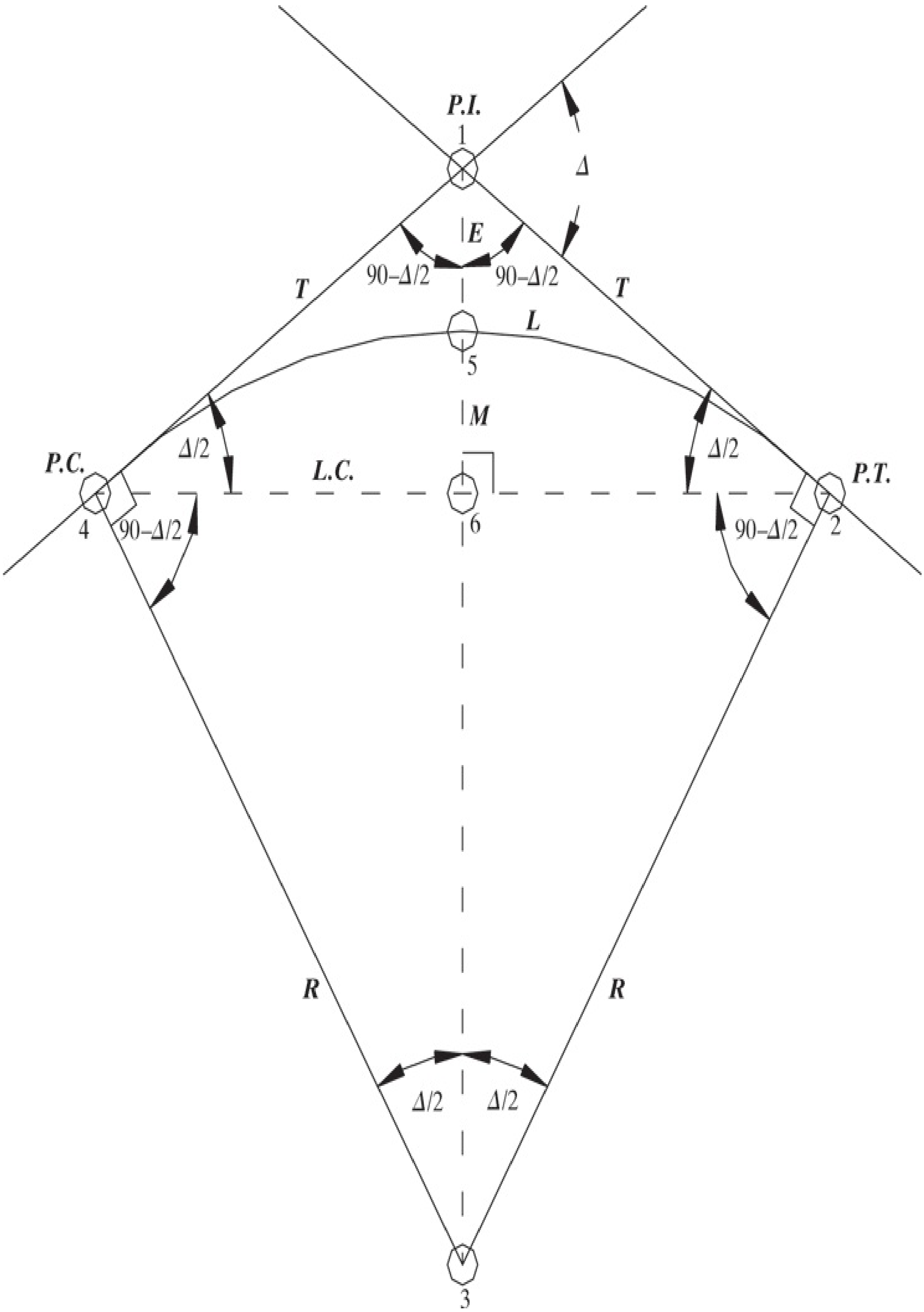


Figure 27.3: Full Alternative Text

- P.I.=point of intersection; point at which the two extended tangent lines meet,

- P.C.=point of curvature; the point at which the circular horizontal curve begins,
- P.T.=point of tangency; the point at which the circular horizontal curve ends,
- T=tangent length; length from the *P.C.* to the *P.I.* and from the *P.I.* to the *P.T.*, ft.,
- E=external distance from point 5 to the *P.I.* in Figure 27.3, ft.,
- M=middle ordinate distance from point 5 to point 6 in Figure 27.3, ft.,
- LC=length of the long chord, from the *P.C.* to the *P.T.*, ft.,
- Δ=angle of the curve, sometimes referred to as the angle of deflection, degrees, and
- R=radius of the circular curve, ft.

A number of geometric characteristics of the circular curve are of interest in deriving important relationships:

- Radii join tangent lines at right (90°) angles at the *P.C.* and *P. T.*
- A line drawn from the *P.I.* to the center of the circular curve bisects $\angle 412$ and $\angle 432$ (numbers refer to Figure 27.3).
- $\angle 412$ equals $180 - \Delta$; thus, $\angle 413$ and $\angle 312$ must be half this, or $90 - \Delta/2$ as shown in Figure 27.3.
- Triangle 412 is an isosceles triangle. Thus, $\angle 142 = \angle 124$, and the sum of these, plus $\angle 412$ $(180 - \Delta)$, must be 180°. Therefore, $\angle 142 = \angle 124 = \Delta/2$.
- $\angle 346$ and $\angle 326$ must be $90 - \Delta/2$, and the central angle, $\angle 432$, is equal to Δ.
- The long chord (*L.C.*) and the line from point 1 to point 3 meet at a right (90°) angle.

Given the characteristics shown in Figure 27.3, some of the key relationships for horizontal curves may be derived. The derivations are not shown here, but exist in every surveying and many geometric design textbooks and manuals. The results are summarized in Table 27.1.

Table 27.1: Determining Key Values for Circular Roadway Curves

Variable	Definition	Equation
T	Length of the Tangent	$T = R \text{ Tan } (\Delta/2)$
L	Length of the Curve	$L = 100\ (\Delta/D)$
M	Length of the Middle Ordinate	$M = R\,[1 - \text{Cos } (\Delta/2)]$
E	Length of the External Ordinate	$E = R\left[\left(\dfrac{1}{\text{Cos } (\Delta/2)}\right) - 1\right]$
LC	Length of the Long Chord	$LC = 2R \text{ Sin } (\Delta/2)$

Δ, D in degrees; *T*, *R*, *L*, *M*, *E* must be in compatible units, usually ft. or meters.

Table 27.1: Full Alternative Text

Sample Problem 27-1: An Example Illustrating the Characteristics of a Horizontal Curve

Two tangent lines meet at Station 3200+15. The radius of curvature is 1,200 ft, and the angle of deflection is 14°. Find the length of the curve (*L*), the stations for the *P.C.* and *P.T.*, and all other relevant characteristics of the curve (*LC*, *M*, *E*). Figure 27.4 illustrates the case.

Figure 27.4: Example Circular Roadway Curve Illustrated (Figure not to scale)

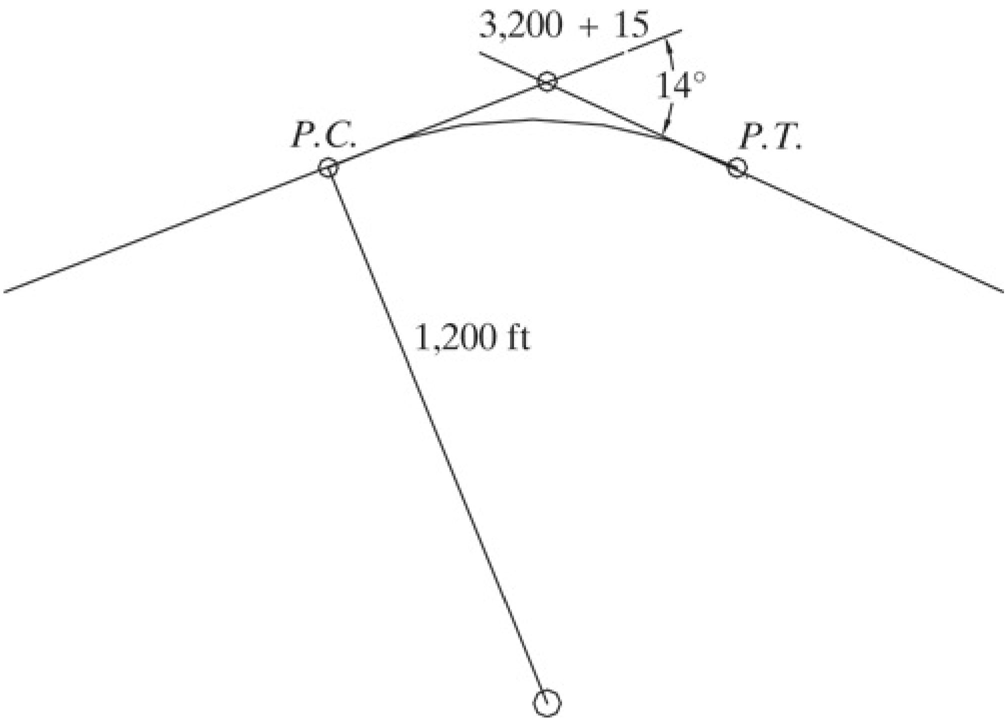


Figure 27.4: Full Alternative Text

Using Equation 27-1 and the equations in Table 27.1, all of the key measures for the curve of Figure 27.4 may be found:

$D=5,729.58R=5,729.581,200=4.77°L=100\ (\Delta/D)=100(14/4.77)=293.5\ \text{ft}T=$

$M=R\ [\ 1-\text{Cos}\ (\Delta 2)\ ]=1,200\ [\ 1-\text{Cos}\ (142)\ ]=1,200\ [1-0.9925]=9.0\ \text{ft}E=R\ [\ 1\text{Cos}\ (\Delta 2)-1\ ]=1,200\ [\ 1\text{Cos}\ (142)-1\ ]=1,200\ [\ 10.9925-1$

$]=9.1$ ft$LC=2R\sin(\Delta/2)=2\times1,200\sin(14/2)=2\times1,200\times0.1219=292.6$ ft

Obviously, this is a fairly short curve, due to a small angle of deflection (14°). The question also asked for station designations for the *P.C.* and *P.T.* The computed characteristics are used to find these values. The station of the *P.I.* is given as 3,200+15, which indicates that it is 3,215 ft from the beginning of the project.

The *P.C.* is found as the *P.I.*-*T*. This is 3,215−147.4=3,067.6, which is station 3,000+67.7. The station of the *P.T.* is found as the P.C.+L. This is 3,067.7+293.5=3,361.2 which is station 3,300+61.2. All station units are in feet.

27.2.4 Superelevation of Horizontal Curves

Most highway curves are "superelevated," or banked, to assist drivers in resisting the effects of centripetal force. Superelevation is quantified as a decimal or percentage, computed as follows:

$$e=\frac{(\text{total rise in pavement from edge to edge})}{(\text{width of pavement})}$$

As noted in Chapter 3, the two factors that keep a vehicle on a highway curve are side friction between the tires and the pavement, and the horizontal element of support provided by a banked or "superelevated" pavement. The speed of a vehicle and the radius of curvature are related to the superelevation rate (*e*) and the coefficient of side friction (*f*), by the equation:

$$R=\frac{S^2}{15(e+f)} \quad [27\text{-}2]$$

where:

R=radius of curvature, ft,S=speed of the vehicle, mi/h,e=rate of superelevat expressed as a decimal), and$f$=coefficient of side friction.

In design, these values become limits: *S* is the design speed for the facility; *e* is the maximum rate of superelevation permitted; and *f* is a design value of the coefficient of side friction representing tires in reasonable condition on a wet pavement. The resulting value of *R* is the minimum radius of curvature permitted for these conditions.

Maximum Superelevation Rates

AASHTO [1] recommends the use of maximum superelevation rates between 4% and 12%. For design purposes, only increments of 2% are used. Maximum rates adopted vary from region to region based upon factors such as climate, terrain, development density, and frequency of slow-moving vehicles. Some of the practical considerations involved in setting this range, and for selection of an appropriate rate include the following:

- Twelve percent (12%) is the maximum superelevation rate in use. Drivers feel uncomfortable on sections with higher rates, and driver effort to maintain lateral position is high when speeds are reduced on such curves. Some jurisdictions use 10% as a maximum practical limit.

- Where snow and ice are prevalent, a maximum value of 8% is generally used. Many agencies use this as an upper limit regardless, due to the effect of rain or mud on highways.

- In urban areas, where speeds may be reduced frequently due to congestion, maximum rates of 4% to 6% are often used.

- On low-speed urban streets or at intersections, superelevation may be eliminated.

It should be noted that on open highway sections, there is generally a minimum superelevation maintained, even on straight sections. This is to provide for cross-drainage of water to the appropriate roadside(s) where sewers or drainage ditches are present for longitudinal drainage. This minimum rate is usually in the range of 1.5% for high-type surfaces and 2.0% to 2.5% for low-type surfaces.

Side-Friction Factors (Coefficients

of Side Friction, *f*)

Design values of the side-friction factor vary with design speed. Design values represent wet pavements and tires in reasonable but not top condition. Values also represent frictional forces that can be comfortably achieved; they do not represent, for example, the maximum side friction that is achieved the instant before skidding.

Table 27.2 illustrates commonly used side friction factors (*f*). Consult Ref 1 directly for a more thorough discussion of side friction factors. Actual side friction factors vary with a number of variables, including the superelevation rate.

Table 27.2: Side Friction Factors Used in Design

Design Speed (mi/h)	*f*	Design Speed (mi/h)	*f*	Design Speed (mi/h)	*f*
10	0.38	35	0.17	60	0.12
15	0.32	40	0.16	65	0.11
20	0.26	45	0.15	70	0.10
25	0.23	50	0.14	75	0.09
30	0.20	55	0.13	80	0.08

Table 27.2: Full Alternative Text

Determining Design Values of Superelevation

Once a maximum superelevation rate and a design speed are set, the minimum radius of curvature can be found using Equation 27-2. This can be expressed as a maximum degree of curvature using Equation 27-1.

Consider a roadway with a design speed of 60 mi/h, for which a maximum superelevation rate of 0.06 has been selected. What are the minimum radius of curvature and/or maximum degree of curve that can be included on this facility?

For a design speed of 60 mi/h, Table 27.2 indicates a design value for the coefficient of side friction (*f*) of 0.120. Then:

Rmin=S215 (emax+fdes)=60215 (0.06+0.12)=1,333.3 ftDmax=5,729.58Rn

While this limits the degree of curvature to a maximum of 4.3° for the facility, it does not determine the appropriate rate of superelevation for degrees of curvature less than 4.3° (or a radius greater than 1,333.3 ft). The actual rate of superelevation for any curve with less than the maximum degree of curvature (or more than the minimum radius) is found by solving Equation 27-2 for *e* using the design speed for *S* and the appropriate design value of *f*. Then:

e=S215 R−fdes [27-3]

For the highway described above, what superelevation rate would be used for a curve with a radius of 1,500 ft? Using Equation 27-3:

e=60215×1,500−0.12=0.04

Thus, while the maximum superelevation rate for this facility was set at 0.06, a superelevation rate of 0.04 would be used for a curve with a radius of 1,500 ft, which is *larger* than the minimum radius for the design constraints specified for the facility. AASHTO standards [1] contain many curves and tables yielding results of such analyses for various specified constraints, for ease of use in design.

Achieving Superelevation

The transition from a tangent section with a normal superelevation for drainage to a superelevated horizontal curve occurs in two stages:

- Tangent Runoff: The outside lane of the curve must have a transition from the normal drainage superelevation to a level or flat condition prior to being rotated to the full superelevation for the horizontal curve. The length of this transition is called the tangent runoff and is noted as L_t.

- Superelevation Runoff: Once a flat cross-section is achieved for the outside lane of the curve, it must be rotated (with other lanes) to the full superelevation rate of the horizontal curve. The length of this transition is called the superelevation runoff and is noted as L_s.

For most undivided highways, rotation is around the centerline of the roadway, although rotation can also be accomplished around the inside or outside edge of the roadway as well. For divided highways, each directional roadway is separately rotated, usually around the inside or outside edge of the roadway.

Figure 27.5 illustrates the rotation of undivided two-lane, four-lane, and six-lane highways around the centerline, although the slopes shown are exaggerated for clarity. The rotation is accomplished in three steps:

Figure 27.5: Achieving Superelevation by Rotation around a Centerline

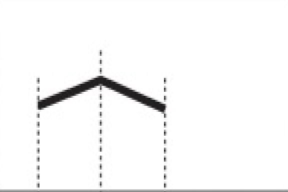
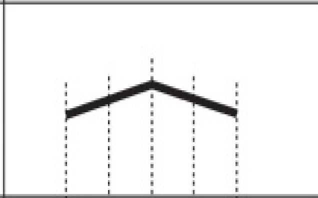
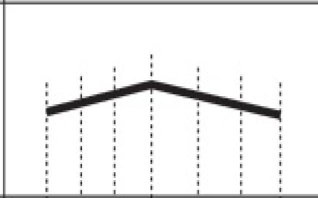
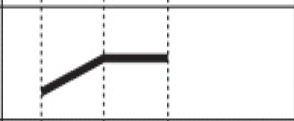
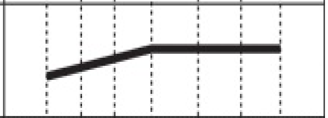
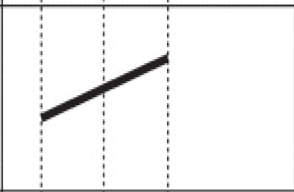
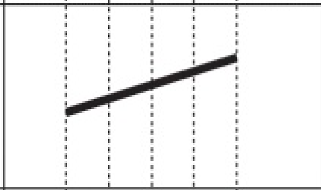
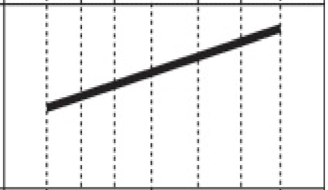
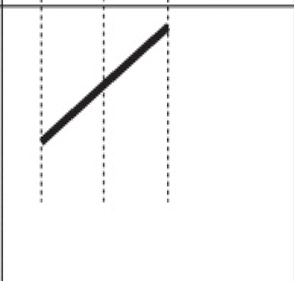
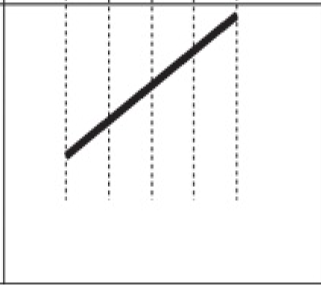
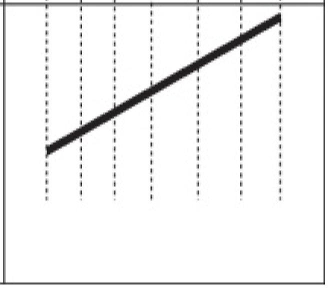
Position	One Lane Rotated	Two Lanes Rotated	Three Lanes Rotated
Normal drainage cross-slope.			
Outside lane rotated to flat.			
Outside lane rotated to normal cross-slope of inside lanes.			
Full superelevation of horizontal curve.			

Figure 27.5: Full Alternative Text

1. Step 1: The outside lane(s) are rotated from their normal cross-slope to a flat condition.

2. Step 2: The outside lane(s) are rotated from the flat position until they equal the normal cross-slope of the inside lanes.

3. Step 3: All lanes are rotated from the condition of step 2 to the full superelevation of the horizontal curve.

The tangent runoff is the distance taken to accomplish step 1, while the superelevation runoff is the distance taken to accomplish steps 2 and 3. The tangent and superelevation runoffs are, of course, implemented for the transition from tangent to horizontal curve and for the reverse transition from horizontal curve back to tangent.

In effect, the transition from a normal cross-slope to a fully superelevated section is accomplished by creating a grade differential between the rotation axis and the pavement edge lines. To achieve safe and comfortable operations, there are limitations on how much of a differential may be accommodated.

The recommended minimum length of superelevation runoff is given as:

Ls=w n ed bwΔm [27-4]

where:

Ls=minimum length of superelevationrunoff, ft,w=width of a lane, ft,n=nu

AASHTO-recommended values for the maximum relative gradient, Δ, are shown in Table 27.3. The adjustment factor, bw, depends upon the number of lanes being rotated, as shown in Table 27.4.

Table 27.3: Maximum Relative Gradients (Δ) for Superelevation Runoff

Design Speed (mi/h)	Maximum Relative Gradient (decimal)	Design Speed (mi/h)	Maximum Relative Gradient (decimal)
15	0.0078	50	0.0050
20	0.0074	55	0.0047
25	0.0070	60	0.0045
30	0.0066	65	0.0043
35	0.0062	70	0.0040
40	0.0058	75	0.0038
45	0.0054	80	0.0035

(Source: Adapted from *A Policy on Geometric Design of Highways and Streets*, 6th Edition, American Association of State Highway and Transportation Officials, Washington, D.C., 2011.)

Table 27.3: Full Alternative Text

Table 27.4: Adjustment Factor (b_w) for Number of Lanes Rotated

Number of Lanes Rotated	Adjustment Factor, b_w
1	1.00
2	0.75
3	0.67

(Source: Adapted from *A Policy on Geometric Design of Highways and Streets*, 6th Edition, American Association of State Highway and Transportation Officials, Washington, D.C., 2011.)

Table 27.4: Full Alternative Text

Consider the example of a four-lane highway, with a superelevation rate of 0.04 achieved by rotating two 12 ft lanes around the centerline. The design speed of the highway is 60 mi/h. What is the appropriate minimum length of superelevation runoff? From Table 27.3, the maximum relative gradient for 60 mi/h is 0.0045; from Table 27.4, the adjustment factor for rotating two lanes is 0.75. Thus:

$$L_s=\frac{w\ n\ e_d\ b_w}{\Delta_w}=\frac{12\times2\times0.04\times0.75}{0.0045}=160\ \text{ft}$$

Note that while it is a four-lane cross-section being rotated, $n=2$, as rotation is around the centerline. Where separate pavements on a divided highway are rotated around an edge, the full number of lanes on the pavement would be used.

The length of the tangent runoff is related to the length of the superelevation runoff, as follows:

$$L_t=\frac{e_{NC}}{e_d}L_s \quad [27\text{-}5]$$

where:

Lt=length of tangent runoff, ft,Ls=length of superelevation runoff, ft,eNC=
slope, decimal, anded=design superelevation rate, decimal.

If, in the previous example, the normal drainage cross-slope was 0.01, then the length of the tangent runoff would be:

Lt=0.010.04 160=40 ft

The total transition length between the normal cross-section to the fully superelevated cross-section is the sum of the superelevation and tangent runoffs, or (in this example) 160+40=200 ft.

To provide drivers with the most comfortable operation, from 60% to 80% of the total runoff is achieved on the tangent section, with the remaining runoff achieved on the horizontal curve. The large majority of agencies use a constant value of 67% of total runoff on the tangent.

Where a spiral transition curve (see next section) is used between the tangent and horizontal curves, the superelevation is achieved entirely on the spiral. Where possible, the tangent and superelevation runoff should be accomplished on the spiral.

27.2.5 Spiral Transition Curves

While not impossible, it is difficult for drivers to travel immediately from a tangent section to a circular curve with a constant radius. A spiral transition curve begins with a tangent (degree of curve, D=0) and gradually and uniformly increases the degree of curvature (decreases the radius) until the intended circular degree of curve is reached.

Use of a spiral transition provides for a number of benefits:

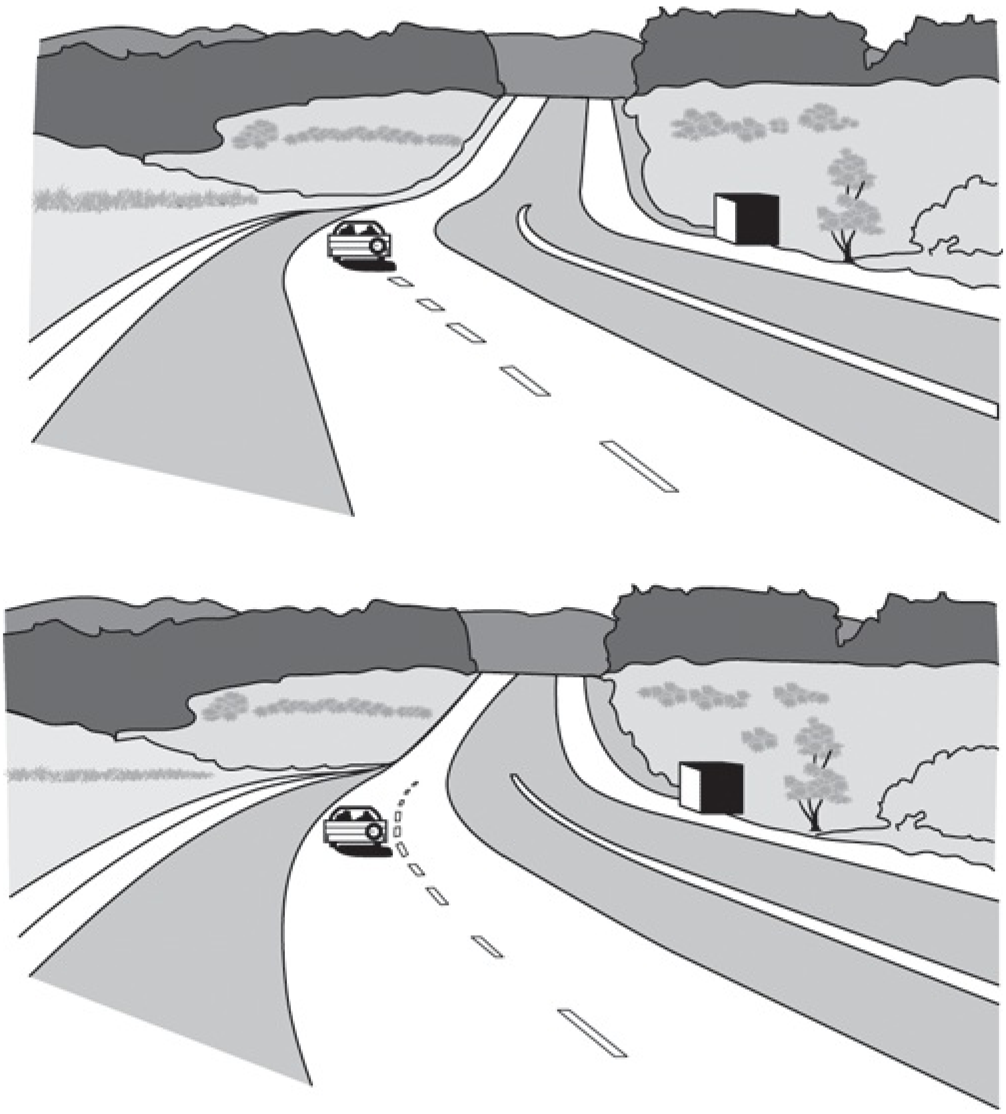
- Provides an easy path for drivers to follow: Centrifugal and centripetal forces are increased gradually.
- Provides a desirable arrangement for superelevation runoff
- Provides a desirable arrangement for pavement widening on curves

(often done to accommodate off-tracking of commercial vehicles)

- Enhances highway appearance

The latter is illustrated in Figure 27.6, where the visual impact of a spiral transition curve is obvious. Spiral transition curves are not always used, as construction is difficult and construction cost is generally higher than for a simple circular curve. They are recommended for high-volume situations where degree of curvature exceeds 3°. The geometric characteristics of spiral transition curves are complex; they are illustrated in Figure 27.7.

Figure 27.6: The Visual Impact of a Spiral Transition Curve



(Source: Used with permission of Yale University Press, C. Tunnard and B. Pushkarev, *Manmade America*, New Haven CT, 1963.)

Figure 27.6: Full Alternative Text

Figure 27.7: Geometry of Spiral Transition Curves

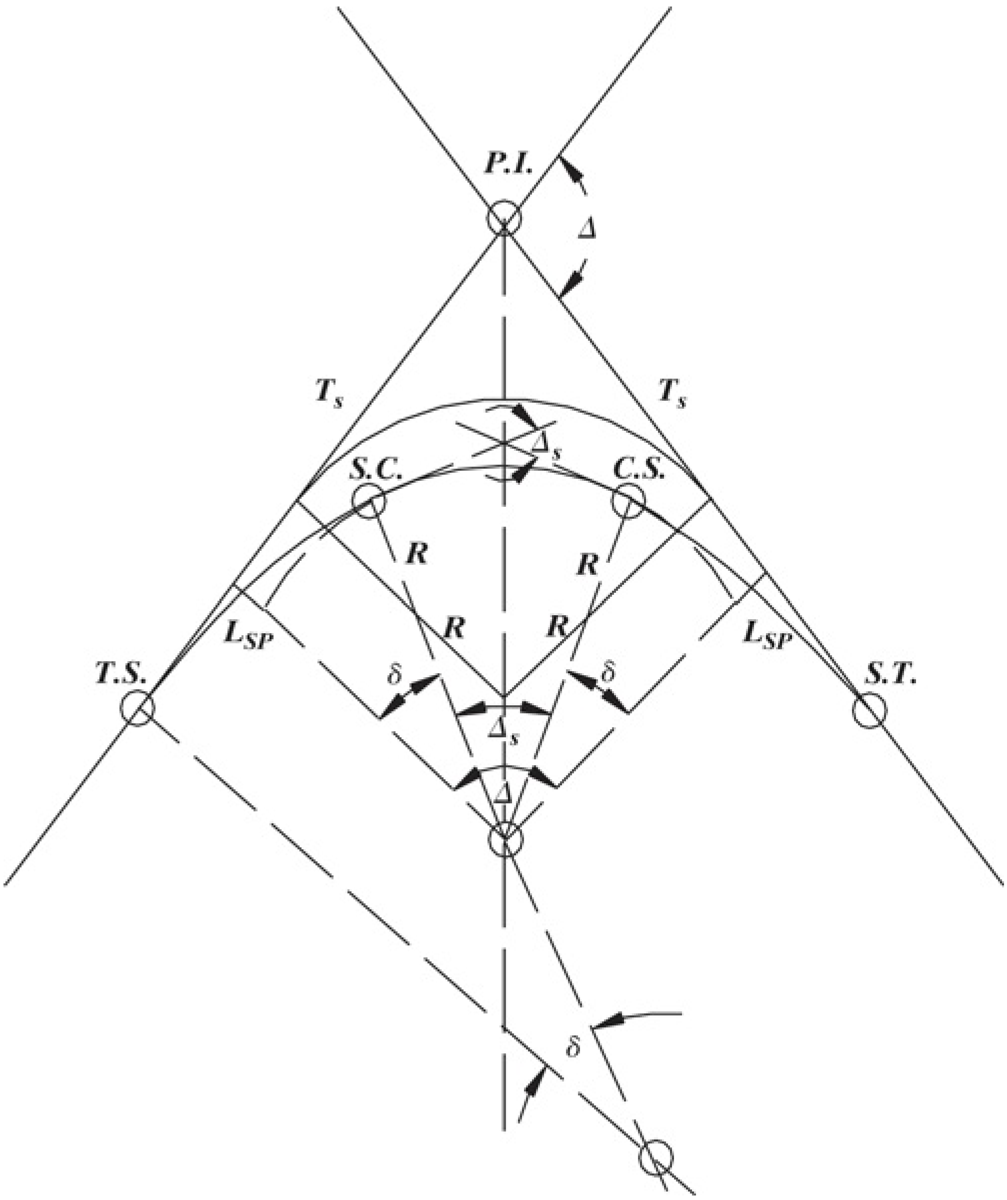


Figure 27.7: Full Alternative Text

The key variables in Figure 27.7 are defined as:

T.S.=transition station from tangent to spiral,S.C.=transition station from sp

Without going through the very detailed derivations for many of the terms included in Figure 27.7, some of the key relationships are shown in Table 27.5.

Table 27.5: Key Variables

Defining Spiral Transition Curves

Variable	Definition	Equation
$L_{SP\min}$	Minimum Length of Spiral	The lesser of: $L_{SP\min} = \sqrt{15.84\,R}$; $L_{SP\min} = 0.7875\,\frac{S^3}{R}$
$L_{SP\max}$	Maximum Length of Spiral	$L_{SP\max} = \sqrt{79.2\,R}$
δ	Spiral Angle of Deflection	$\delta = \frac{L_{SP}\,D}{200}$
Δ_s	Circular Angle of Deflection with Spiral	$\Delta_s = \Delta - 2\delta$
T_s	Length of Tangent Distance	$T_s = \left[R\,\mathrm{Tan}\left(\frac{\Delta}{2}\right)\right] + \left[\left(R\,\mathrm{Cos}\,(\delta) - R + \frac{L_{SP}^2}{6\,R}\right) \times \mathrm{Tan}\left(\frac{\Delta}{2}\right)\right] + \left[L_{SP} - R\,\mathrm{Sin}\,(\delta)\right]$

Note: All variables as previously defined.

Table 27.5: Full Alternative Text

The key determination is the appropriate length of the spiral curve, L_{SP}, as other variables depend on this starting point. Based upon various studies of safe and comfortable operation of vehicles on horizontal curves, the common approach is to establish minimum and maximum values. Assuming that the value falls within the minimum and maximum limits, one approach is to make the length of the spiral equal to the sum of the tangent runoff (L_t) and the spiral runoff (L_s). Another approach is to make the length of the spiral equal to the spiral runoff alone. If these or other approaches fall outside the minimum and maximum recommended lengths, the following actions are recommended:

- If the desired length is < the minimum value, use the minimum spiral length.
- If the desired length is > the maximum value, use the maximum spiral length.

AASHTO does provide recommended values for use in design, as shown in Table 27.6.

Table 27.6: AASHTO Recommended Lengths for Spiral Transition Curves

Design Speed (mi/h)	Length of Spiral (ft)	Design Speed (mi/h)	Length of Spiral (ft)
15	44	50	147
20	59	55	161
25	74	60	176
30	88	65	191
35	103	70	205
40	117	75	220
45	132	80	235

(Source: Adapted from *A Policy on Geometric Design of Highways and Streets*, 6th Edition, American Association of State Highway and Transportation Officials, Washington, D.C., 2011.)

Table 27.6: Full Alternative Text

Sample Problem 27-2: A Spiral

Transition Curve

Consider the following situation: A 4° curve is to be designed on a highway with two 12 ft lanes and a design speed of 60 mi/h. A maximum superelevation rate of 0.06 has been established, and the appropriate side-friction factor for 60 mi/h is found from Table 27.2 as 0.120. The normal drainage cross-slope on the tangent is 0.01. Spiral transition curves are to be used. Determine the length of the spiral and the appropriate stations for the *T.S.*, *S.C.*, *C.S.*, and *S.T.* The angle of deflection for the original tangents is 38°, and the *P.I.* is at station 1,100+62. The segment has a two-lane cross-section.

The radius of curvature for the circular portion of the curve is found from the degree of curvature as (Equation 27-1):

R=5,729.58D=5,729.584=1,432.4 ft

For a 60 mi/h design speed, Table 27.6 recommends a 191 ft spiral curve. This must be checked to see that it conforms to the minimum and maximum requirements, using the equations in Table 27.5:

LSPmin=15.84 R=15.84× 1,432.4=22,689.2=150.6 ft,   ORLSPmin=0.7875

The recommended 191 ft spiral transition curve lies within the minimum and maximum recommended values for the curve as described. It will, therefore, be used.

The desirable length of the spiral can also be determined as the length of the superelevation runoff, or the length of the superelevation plus tangent runoffs. For a 60 mi/h design speed and a radius of 1,432.4 ft, the superelevation rate is found using Equation 27-3:

e=S215 R−f=60215×1,432.4−0.12=0.048

The length of the superelevation and tangent runoffs are computed from Equations 27-4 and 27-5, respectively. For 60 mi/h, the design value of Δ is 0.0045 (Table 27.3). The adjustment factor for two lanes being rotated is 0.75 (Table 27.4). Then:

Ls=w n ed bwΔw=12×2×0.048×0.750.0045=0.8640.0045=192 ftLt=eNCed

The recommended length of the curve, 191 ft, conforms closely with the length of the superelevation runoff. If the policy is that the spiral should encompass all of the runoff, that is, 192+40=232.0 ft, then this longer length would be adopted, as it is still less than the maximum value computed previously. We will proceed using the recommended value of 191 ft.

Remaining values are computed using the equations detailed in Tables 27.1 and 27.5.

The angle of deflection for the spiral is:

$\delta=LSP\ D200=191\times4200=3.8o$

The angle of deflection for the circular portion of the curve is:

$\Delta s=\Delta-2\delta=38-(2\times3.8)=30.4o$

The length of the circular portion of the curve, Lc, is:

$Lc=100\Delta sD=100\ 30.44=760.0\ \text{ft}$

The spiral tangent distance, T_s, is:

$Ts=[\ R\ Tan\ (\Delta2)\ ]+[\ (R\ Cos\ (\delta)-R+LSP26\ R)\times Tan\ (\Delta2)\ ]+[\ LSP-R\ Sin\ (\delta)\ ]Ts=[\ 1432.4\ Tan\ (382)\ ]+[\ (1432.4\ Cos\ (3.8)-1432.4+19126\times1432.4)\times Tan\ (382)\ ]+[\ 191-1432.4\ Sin\ (3.8)\ ]Ts=[\ 1432.4\times0.3443\ ]+[\ (1432.4\times0.9999-1432.4+4.2447)\times0.3443\ ]+[\ 191-1432.4\times0.00663\ ]Ts=493.2+1.4+181.5=676.1\ \text{ft}$

From these results, the curve may now be stationed:

$T.S.=P.I.-Ts=1{,}162-676.1=485.9\ \text{ or }\ \text{Station}\ \ 400+85.9S.C.=T.S.+LSP=485.9+191$

27.2.6 Sight Distance on Horizontal Curves

One of the most fundamental design criteria for all highway facilities is that a minimum sight distance equal to the safe stopping distance must be provided at every point along the roadway.

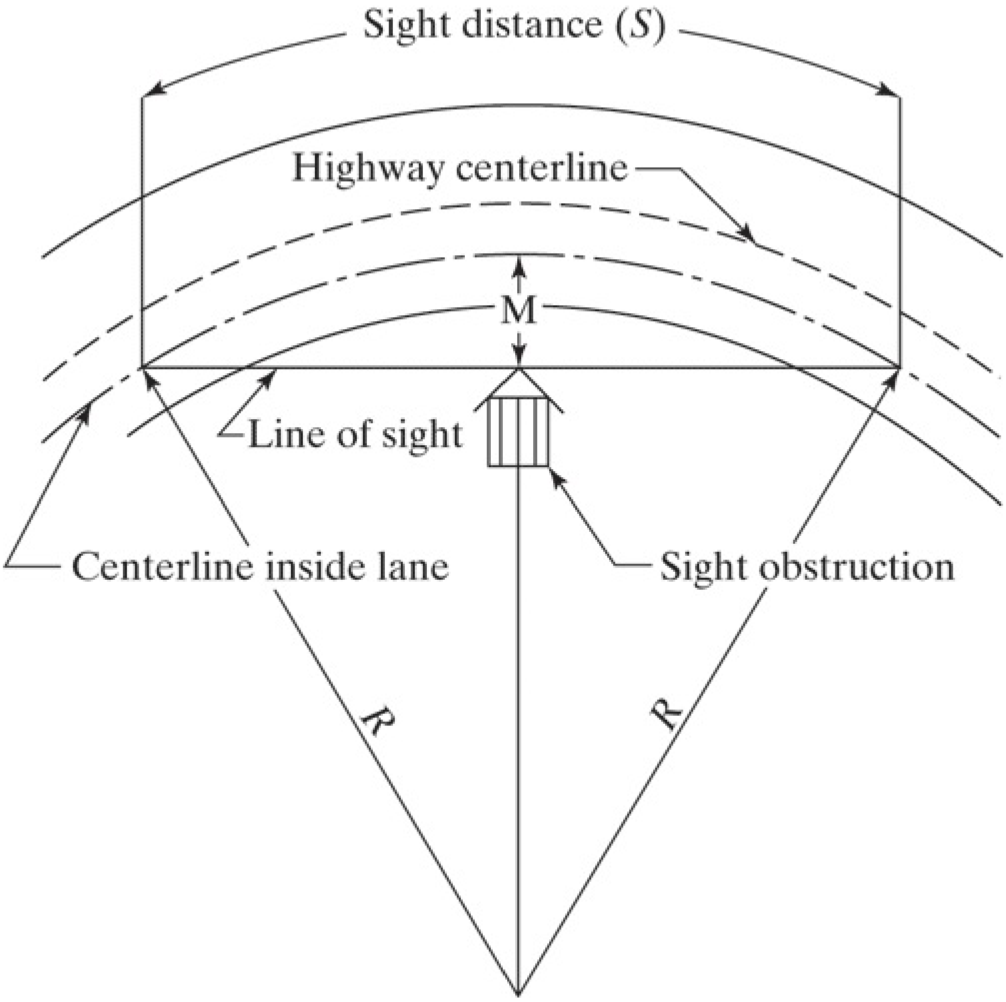
On horizontal curves, sight distance is limited by roadside objects (on the inside of the curve) that block drivers' line of sight. Roadside objects such as buildings, trees, and natural barriers disrupt motorists' sight lines. Figure 27.8 illustrates a sight restriction on a horizontal curve.

Figure 27.8: A Sight Distance Restriction on a Horizontal Curve

(Photo courtesy of J. Ulerio and R. Roess)

Figure 27.9 illustrates the effect of horizontal curves on sight distance. Sight distance is measured along the arc of the roadway, using the centerline of the inside travel lane. The middle ordinate, *M*, is the distance from the centerline of the inside lane to the nearest roadside sight blockage. In its current criteria, AASHTO refers to the middle ordinate as the "horizontal sight line offset (HSO)." For consistency, this text continues to use *M* for this value.

Figure 27.9: Sight Restrictions on Horizontal Curves



(Source: Adapted from *A Policy on Geometric Design of Highways and Streets*, 4th Edition, American Association of State Highway and Transportation Officials, Washington, D.C., 2001.)

Figure 27.9: Full Alternative Text

The formula for the middle ordinate was given in Table 27.1 as:

$M=R\ \ [1-\mathrm{Cos}\ \ (\Delta 2)]$

The length of the circular curve has also been defined in Table 27.1. In this case, however, the length of the curve is set equal to the required stopping sight distance. Then:

$L=ds=100\ (\Delta D)\Delta=ds\ D100$

Substituting in the equation for *M*:

$M=R\ \ [1-\mathrm{Cos}\ (ds\ D200)]$

The equation can be expressed uniformly using either the degree of curvature, *D*, or the radius of curvature, *R*:

$M=5{,}729.28D\ \ [\ 1-\mathrm{Cos}\ (ds\ D200)\ ]\quad M=R\ \ [\ 1-\mathrm{Cos}\ \ (28.65\ dsR)\ ]$ [27-6]

Remember (see Chapter 3) that the safe stopping distance used in each of these equations may be computed as:

$ds=1.47\ S\ t+S230\ (0.348\pm G)$

where:

ds=safe stopping distance, ft,S=design speed, mi/h,t=reaction time, s (
2.5 s is the AASHTO standard for highway stops
), andG=grade, expressed as a decimal.

Sample Problem 27-3: Sight Distance on Horizontal Curve

Consider the following situation: A 6° curve (measured at the centerline of the inside lane) is being designed for a highway with a design speed of 70 mi/h. The grade is level, and driver reaction time will be taken as 2.5 seconds, the AASHTO standard for highway braking reaction. What is the closest any roadside object may be placed to the centerline of the inside lane of the roadway?

The safe stopping distance, d_s, is computed as:

ds=1.47×70×2.5+70230 (0.348+0.000)=257.3+469.3=726.6 ft

The minimum clearance at the roadside is given by the middle ordinate for a sight distance of 726.6 ft:

M=5,729.58D [1−Cos (ds D200)]M=5,279.586 [1−Cos (726.6×6200)]M=879.93 [1−Cos (21.798)]M=879.93 [1−0.9285]=62.9 ft

Thus, for this curve, no objects or other sight blockages on the inside roadside may be closer than 62.9 ft to the centerline of the inside lane.

27.2.7 Compound Horizontal Curves

A compound horizontal curve consists of two or more consecutive horizontal curves in a single direction with different radii. Figure 27.10 illustrates such a curve.

Figure 27.10: Compound Horizontal Curve Illustrated

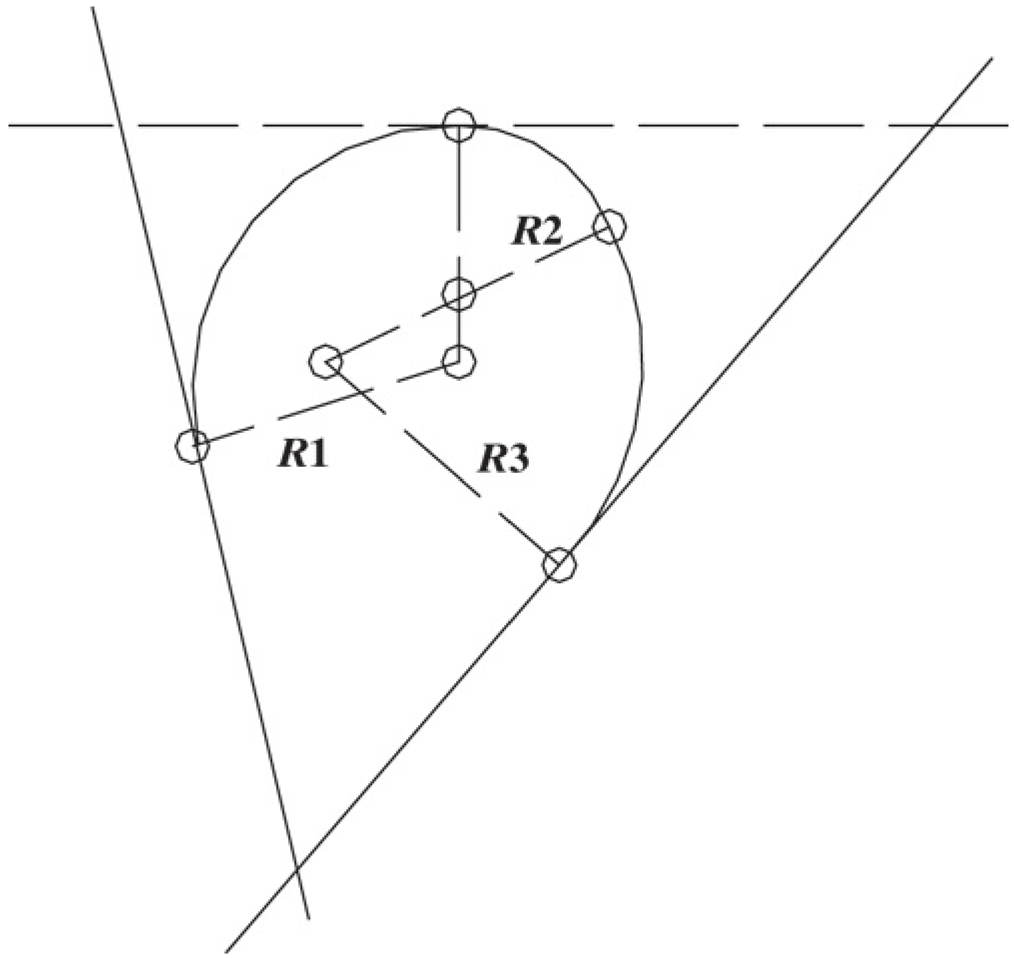


Figure 27.10: Full Alternative Text

Some general criteria for such curves include the following:

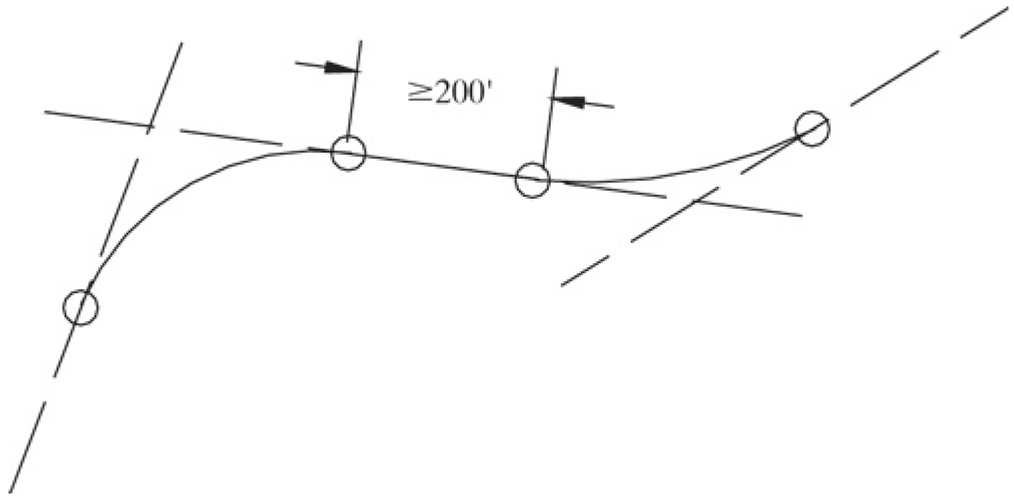
- Use of compound curves should be limited to cases in which physical conditions require it.
- Whenever two consecutive curves are connected on a highway segment, the larger radii should not be more than 1.5 times the smaller. A similar criterion is that the degrees of curvature should not differ by more than 5°.
- Whenever two consecutive curves in the same direction are separated by a short tangent (≤200 ft) they should be combined in a compound curve.
- A compound curve is merely a series of simple horizontal curves subject to the same criteria as isolated horizontal curves.

- AASHTO relaxes some of these criteria for compound curves on ramps.

27.2.8 Reverse Horizontal Curves

A reverse curve consists of two consecutive horizontal curves in opposite directions. Such a curve is illustrated in Figure 27.11. Two horizontal curves in opposite directions should always be separated by a tangent of at least 200 ft. Use of spiral transition curves is a significant assist to drivers negotiating reverse curves.

Figure 27.11: A Reverse Horizontal Curve Illustrated



27.3 Vertical Alignment of Highways

The vertical alignment of a highway is the *profile* design of the facility in the vertical plane. The vertical alignment is composed of a series of vertical tangents connected by vertical curves. Vertical curves are in the shape of a *parabola*. This provides for a natural transition from a tangent to a curved section as part of the curve characteristics. Therefore, there is no need to investigate or provide transition curves, such as the spiral for horizontal curves.

The longitudinal slope of a highway is called the *grade*. It is generally stated as a percentage.

In vertical design, attempts are made to conform to the topography wherever possible to reduce the need for costly excavations and landfills as well as to maintain aesthetics. Primary design criteria for vertical curves include the following:

- Provision of adequate sight distance at all points along the profile
- Provision of adequate drainage
- Maintenance of comfortable operations
- Maintenance of reasonable aesthetics

27.3.1 Grades

Vertical tangents are characterized by their longitudinal slope, or grade. When expressed as a percentage, the grade indicates the relative rise (or fall) of the facility in the longitudinal direction as a percentage of the length of the section under study. Thus, a 4% grade of 2,000 ft involves a vertical rise of $2000 \times 4/100 = 80$ ft. Upgrades have positive slopes and percent grades, while downgrades have negative slopes and percent grades. Note that while grades are traditionally stated in percent, in most of the previous equations of this text, grade (*G*) is expressed as a decimal. In the equations of this section involving grade, the value is expressed as percent.

Maximum recommended grades for use in design depend upon the type of facility, the terrain in which it is built, and the design speed. AASHTO recommends maximum grades for a wide variety of facilities and situations. A summary of some of these is shown in Table 27.7.

Table 27.7: Maximum Grade (Percentage) Criteria—Current Practice

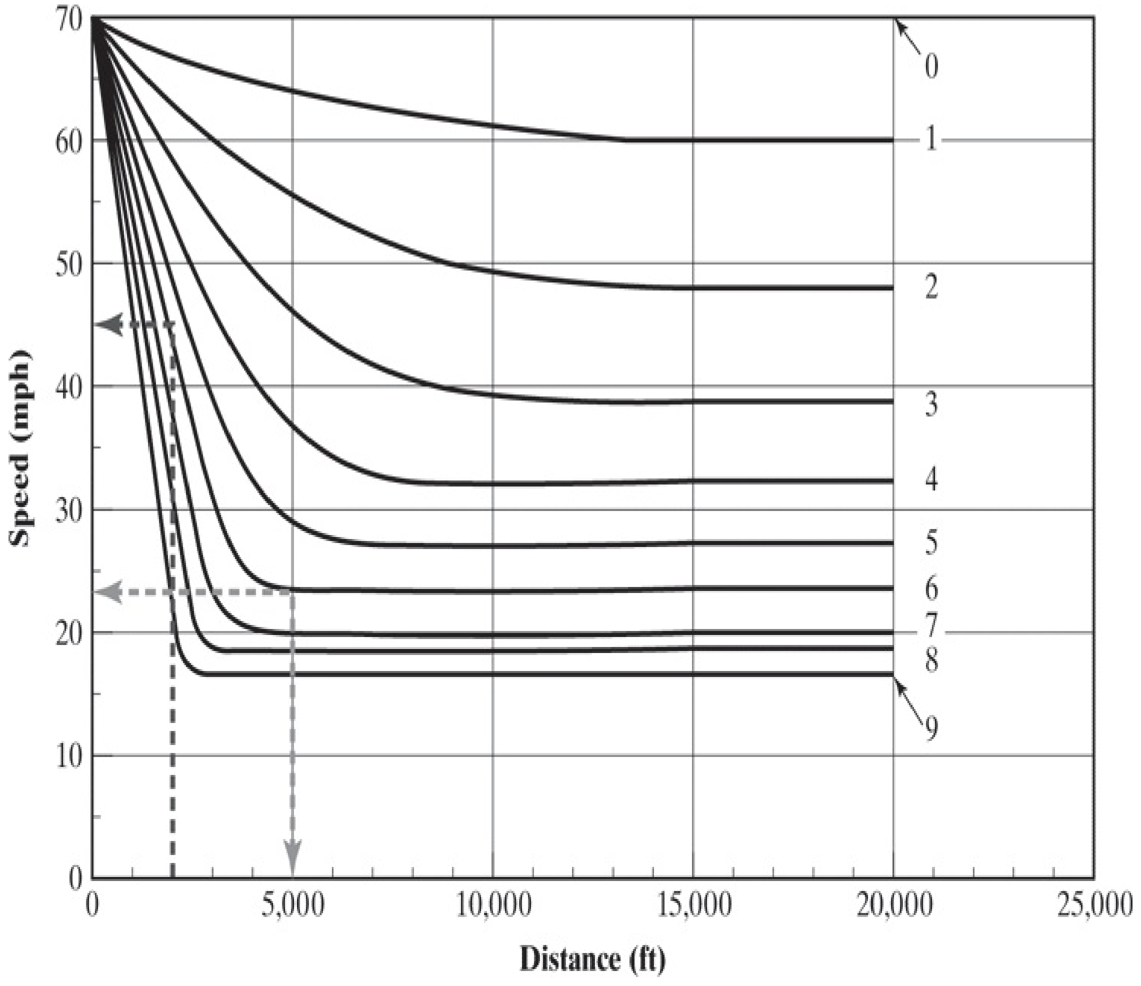
Type of Facility	Terrain	Design Speed (mi/h)						
		20	30	40	50	60	70	80
Local Rural Roads	Level	8	7	7	8	5	NA	NA
	Rolling	11	10	10	8	6	NA	NA
	Mountainous	16	14	13	10	NA	NA	NA
Local Residential Streets	All	15	15	NA	NA	NA	NA	NA
Rural Collectors	Level	7	7	7	6	5	NA	NA
	Rolling	10	9	8	7	6	NA	NA
	Mountainous	12	10	10	9	8	NA	NA
Urban Collectors	Level	9	9	9	7	6	NA	NA
	Rolling	12	11	10	8	7	NA	NA
	Mountainous	14	12	12	10	9	NA	NA
Rural Arterials	Level	NA	NA	5	4	3	3	3
	Rolling	NA	NA	6	5	4	4	4
	Mountainous	NA	NA	8	7	6	5	5
Urban Arterials	Level	NA	8	7	6	5	NA	NA
	Rolling	NA	9	8	7	6	NA	NA
	Mountainous	NA	11	10	9	8	NA	NA
Urban and Rural Freeways	Level	NA	NA	NA	4	3	3	3
	Rolling	NA	NA	NA	5	4	4	4
	Mountainous	NA	NA	NA	6	6	5	NA

NA=not applicable; indicates that this combination of facility, terrain, and design speed is not recommended.

Table 27.7: Full Alternative Text

The principal operational impact of a grade is that trucks will be forced to slow down as they progress up the grade. This creates gaps in the traffic stream that cannot be effectively filled by simple passing maneuvers. Figure 27.12 illustrates the effect of upgrades on the operation of trucks with a weight-to-horsepower ratio of 200 lb/hp, which is considered to be operationally typical of heavy trucks on most highways. It depicts deceleration behavior with an assumed entry speed of 70 mi/h.

Figure 27.12: Deceleration of Typical Trucks (200 lbs/hp) on Upgrades



(Source: Adapted from *A Policy on Geometric Design of Highways and Streets*, 6th Edition, American Association of State Highway and Transportation Officials, Washington, D.C., 2011.)

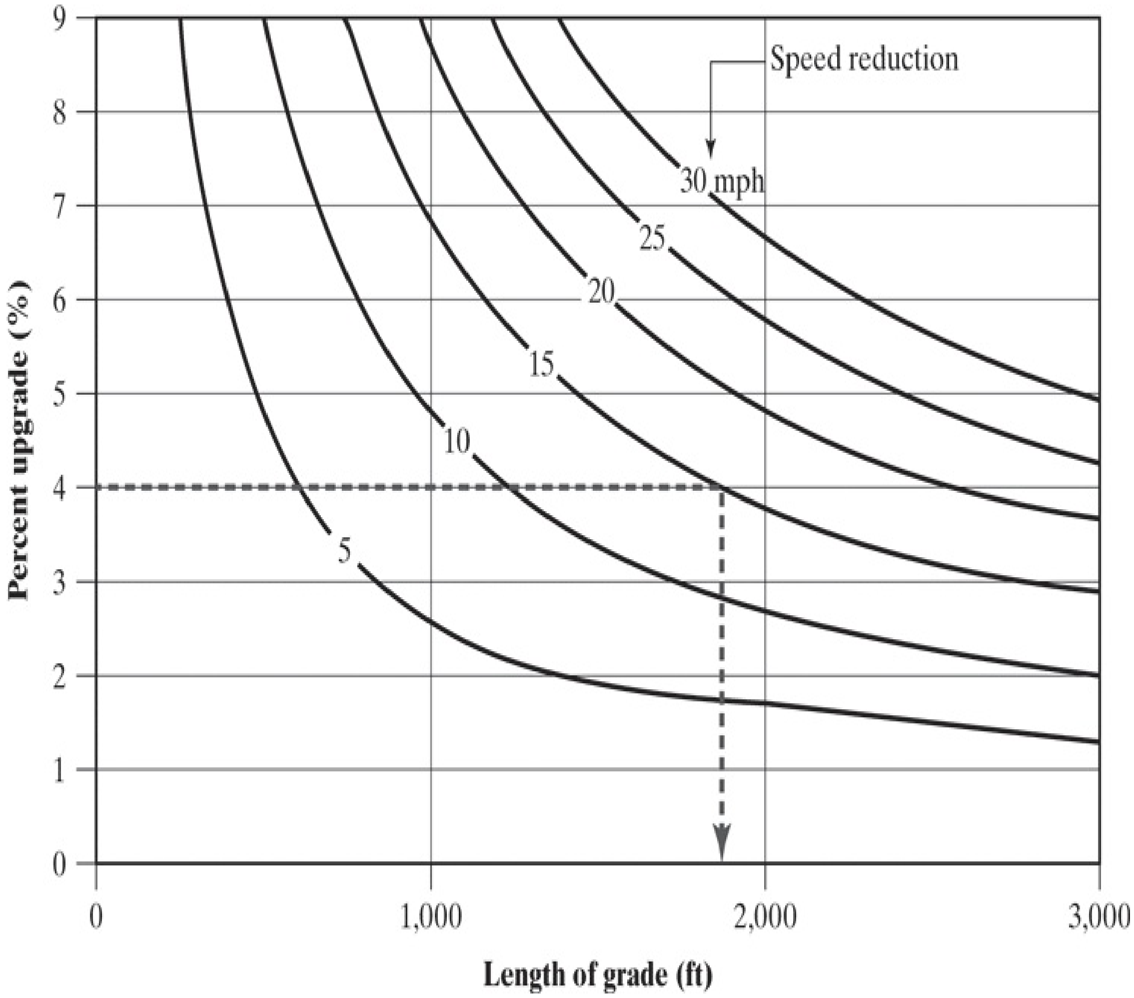
Figure 27.12: Full Alternative Text

Because of the operation of trucks on grades, simple maximum grade criteria are not sufficient for design. An example is shown in Figure 27.12. Trucks entering an upgrade with an assumed speed of 70 mi/h begin to slow. The length of the upgrade determines the extent of deceleration. For example, a truck entering a 6% upgrade at 70 mi/h slows to 45 mi/h after 2,000 ft. Eventually, the truck reaches its "crawl speed." The crawl speed is that constant speed that the truck can maintain for any length of grade (of the given steepness). Using the same example, a truck on a 6% upgrade has a crawl speed of 23 mi/h that is reached after approximately 5,000 ft.

Thus, the interference of trucks with general highway operations is related not only to the steepness of the grade but to its length as well. For most design purposes, grades should not be longer than the "critical length." For grades entered at 70 mi/h, the critical length is generally defined as the length at which the speed of trucks is 15 mi/h less than their speed upon entering the grade. When trucks enter an upgrade from slower speed, a speed reduction of 10 mi/h may be used to define the critical length of grade.

Figure 27.13 shows the relationship between length of grade, percentage grade, and speed reduction for 200 lb/hp trucks entering a grade at 70 mi/h. These curves can be used to determine critical length of grade. It should be noted that terrain may make it impossible to limit grades to the critical length or shorter.

Figure 27.13: Critical Lengths of Grade for a Typical Truck (200 lbs/hp)



(Source: Adapted from *A Policy on Geometric Design of Highways and Streets*, 6th Edition, American Association of State Highway and Transportation Officials, Washington, D.C., 2011.)

Figure 27.13: Full Alternative Text

Sample Problem 27-4: An Example of a Vertical Grade

Consider the following situation: A rural freeway in rolling terrain has a design speed of 60 mi/h. What is the longest and steepest grade that should be included on the facility?

From Table 27.7, for a freeway facility with a design speed of 60 mi/h in rolling terrain, the maximum allowable grade is 4%. Entering Figure 27.13 with 4% on the vertical axis, moving to the “15 mi/h” curve, the critical length of grade is seen to be approximately 1,900 ft.

Again, it must be emphasized that terrain sometimes makes it impossible to consistently follow maximum grade design criteria. This is particularly true for desirable maximum grade lengths. Where the terrain is rising for significant distances, the profile of the roadway must do so as well. It is, however, true that grades longer than the critical length will generally operate poorly, and the addition of a climbing lane may be warranted in such situations.

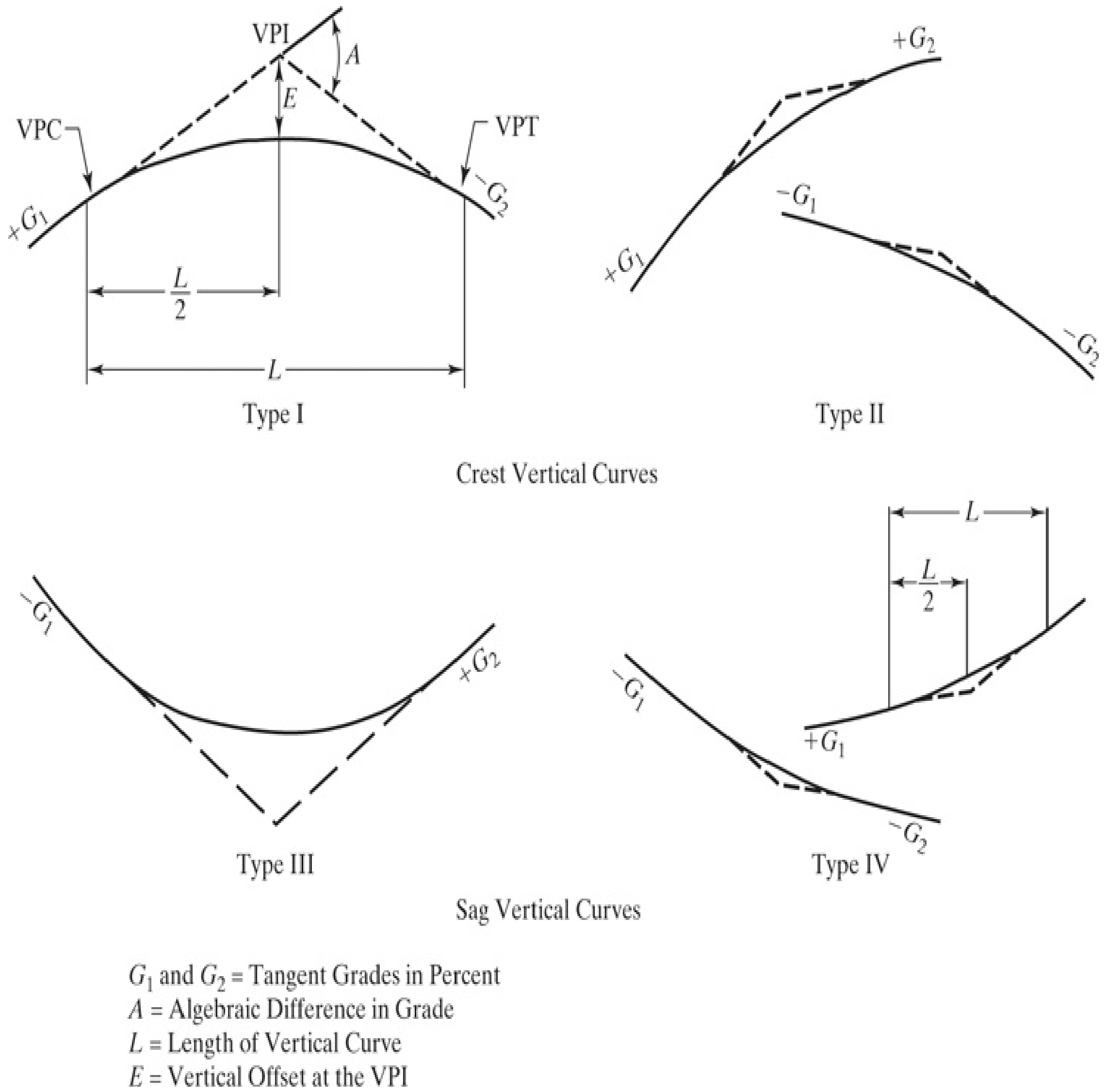
27.3.2 Geometric Characteristics of Vertical Curves

As noted previously, vertical curves are in the shape of a parabola. In general, there are two types of vertical curves:

- Crest vertical curves
- Sag vertical curves

For crest vertical curves, the entry tangent grade is greater than the exit tangent grade. While traveling along a crest vertical curve, the grade is constantly declining. For sag vertical curves, the opposite is true: the entry tangent grade is lower than the exit tangent grade, and while traveling along the curve, the grade is constantly increasing. Figure 27.14 illustrates the various types of vertical curves.

Figure 27.14: Vertical Curves Illustrated



(Source: Adapted from *A Policy on Geometric Design of Highways and Streets*, 6th Edition, American Association of State Highway and Transportation Officials, Washington, D.C., 2011.)

Figure 27.14: Full Alternative Text

The terms used in Figure 27.14 are defined as:

VPI=vertical point of intersectionVPC=vertical point of curvatureVPT=ver

Length, and all stationing, on a vertical curve is measured in the plan view, (i.e., along a level axis). Another useful variable is defined as follows:

$A=|G2-G1|$ [27-7]

where *A* is the algebraic change in grade, %.

As noted previously, the geometric shape of a vertical curve is parabolic. The general form of a parabola is:

$$y=ax^2+bx+c$$

For the purposes of describing a vertical curve, let:

- $y=Y_x$=the elevation of the roadway at a point "x" from the *VPC*, ft,
- x=distance from the *VPC*, ft, and
- $c=Y_o$=elevation of the *VPC*, which occurs where $x=0$, ft.

Then:

$$Y_x=ax^2+bx+Y_o$$

Also, consider that the slope of the curve at any point *x* is the first derivative of this equation, or:

$$\frac{dy}{dz}=2ax+b$$

When $x=0$, the slope is equal to the entry grade, G_1. Note that in these equations, grades *G* must be expressed as a decimal. Therefore, % grades are divided by 100. Thus:

$$\frac{dy}{dx}=\frac{G_1}{100}=2a(0)+b \qquad b=\frac{G_1}{100}$$

Also, the second derivative of the equation is equal to the rate of change in slope along the grade, or:

$$\frac{d^2y}{dx^2}=2a=\frac{\left(\frac{G_2-G_1}{100}\right)}{L} \qquad a=\frac{G_2-G_1}{200L}$$

Thus, the final form of the equation for a vertical curve is given as:

$$Y_x=\left(\frac{G_2-G_1}{200L}\right)x^2+\left(\frac{G_1}{100}\right)x+Y_o \quad [27\text{-}8]$$

The location of the high point (on a crest vertical curve) or the low point (on a sag vertical curve) is at a point where the slope (or first derivative) is equal to "zero." Taking the first derivative of the final curve:

$$\frac{dY_x}{dx}=0=\left(\frac{G_2-G_1}{100L}\right)x+\frac{G_1}{100} \quad x_{high,\,low}=\frac{-G_1L}{G_2-G_1} \quad [27\text{-}9]$$

In all of these equations, care must be taken to address the sign of the grade. A negative grade has a minus (−) sign that must be accounted for in the equation. Double negatives become positives in the equation. If both grades are negative (downgrades), the low point on the curve is the *VPT* and the high point is the *VPC*. If both grades are positive (upgrades), the low point on the curve is the *VPC* and the high point is the *VPT*.

Sample Problem 27-5: A Vertical Curve

Consider the following situation: A vertical curve of 600 ft connects a +4% grade to a −2% grade. The elevation of the *VPC* is 1,250 ft. Find the elevation of the *PVI*, the high point on the curve, and the *VPT*.

The elevation of the *VPI* is found from the elevation of the *VPC*, the approach grade, and the length of the vertical curve. The *VPI* is located on the extension of the approach grade at a point $\frac{1}{2}L$ into the curve, or:

$$Y_{VPI}=Y_{VPC}+\left(\frac{G_1}{100}\right)\left(\frac{L}{2}\right)=1{,}250+\left(\frac{4}{100}\right)\left(\frac{600}{2}\right)=1{,}250+12=1{,}262 \text{ ft}$$

Note that because the grade is stated as a percentage, it must be divided by 100 (converted to a decimal) for use in this calculation.

Following the format of Equation 27-8, the equation for this particular vertical curve is:

$$Y_x=\left(\frac{-2-4}{200\times 600}\right)x^2+\left(\frac{4}{100}\right)x+1{,}250 \qquad Y_x=-0.0005\,x^2+0.04\,x+1{,}250$$

The elevation of the *VPT* is the elevation of the curve at the end of its length of 600 ft, or where $x=600$:

$$Y_{PVT}=-0.00005\,(600^2)+0.04\,(600)+1{,}250 \qquad Y_{PVT}=-18+24+1{,}250=1{,}256 \text{ ft}$$

The high point of the curve occurs at a point where:

$$x_{high}=\frac{-G_1\,L}{G_2-G_1}=\frac{-4\,(600)}{-2-4}=\frac{-2{,}400}{-6}=400 \text{ ft}$$

Then:

$Y_{high}=-0.00005\ (4002)+0.04\ (400)+1{,}250=-8+16+1{,}250=1{,}258$ ft

27.3.3 Sight Distance on Vertical Curves

The minimum length of vertical curve is governed by sight-distance considerations. On vertical curves, sight distance is measured from an assumed eye height of 3.5 ft and an object height of 2.0 ft. Figure 27.15 shows a situation in which sight distance is limited by vertical curvature.

Figure 27.15: Sight Restriction Due to Vertical Curvature

(Source: Photo courtesy of J. Ulerio and R. Roess)

Crest Vertical Curves

For crest vertical curves, the daylight sight line controls minimum length of vertical curves. Equations for the minimum length of a crest vertical curve are shown in Table 27.8. Equations depend upon whether the stopping sight distance, d_s, is greater than or less than the length of the vertical curve, *L*. The two equations yield equal results when ds=L.

Table 27.8: Determining the

Minimum Length of a Vertical Curve

	Crest Vertical Curves	Sag Vertical Curves
$d_s > L$	$L = 2d_s - \dfrac{2{,}158}{A}$	$L = 2d_s - \dfrac{400 + 3.5d_s}{A}$
$d_s < L$	$L = \dfrac{A\,d_s^2}{2{,}158}$	$L = \dfrac{Ad_s^2}{400 + 3.5d_s}$

ds=safe stopping distance, ft L=length of vertical curve, ft

A=absolute value of the algebraic difference in grade, %=|G2−G1|.

Table 27.8: Full Alternative Text

Sag Vertical Curves

For sag vertical curves, the sight distance is limited by the headlamp range during nighttime driving conditions. Again, two equations result, depending upon whether d_s greater than or less than the length of the vertical curve, *L*. These equations are also shown in Table 27.8. As with crest vertical curves, the two equations yield equal results when ds=L.

While the sag equations are technically based upon headlight range at night, safety dictates that this distance be at least equal to the safe stopping distance, d_s. Therefore, the safe stopping distance is generally used in the equations.

Sample Problem 27-6: Sight

Distance on a Vertical Curve

Consider the following situation: What is the minimum length of vertical curve that must be provided to connect a +5% grade with a +2% grade on a highway with a design speed of 60 mi/h? Driver reaction time is the AASHTO standard of 2.5 s for simple highway stopping reactions.

This vertical curve is a *crest* vertical curve, as the departure grade is less than the approach grade. The safe stopping distance is computed assuming that the vehicle is on a 2% upgrade. This is a conservative assumption, in that stopping distance is lower on a 5% upgrade than on a 2% upgrade. This results in a “worst-case” stopping distance of:

$$ds=1.47\ S\ t+S230\ (0.348+G)=$$
$$(1.47\times 60\times 2.5)+60230\times 0.368=220.5+326.1=546.6\text{ ft}$$

Rounding off this number, a stopping sight distance requirement of 547 ft will be used. The first computation is made assuming that the stopping sight distance is *more* than the resulting length of curve; if the result is that $L>ds$, a recomputation will be necessary. From Table 27.8:

$$L=2\ ds-2{,}158A=2\ (547)-(2{,}158|\ 2-5\ |)=1{,}094-719.3=364.7,\ \text{ say }\ 365\text{ ft}$$

In this case, $ds>L$, as assumed, and the 365 ft is taken as the result.

For sag vertical curves, AASHTO [1] gives alternative criteria for minimum length based on (a) driver comfort, and (b) situations where an overpass interrupts the headbeams at night. It also provides equations like those in Table 27.8 for passing sight distance.

27.3.4 Some Design Guidelines for Vertical Curves

AASHTO gives a number of common-sense guidelines for the design of highway profiles, which are summarized below:

1. A smooth grade line with gradual changes is preferred to a line with

numerous breaks and short grades.

2. Profiles should avoid the “roller-coaster” appearance, as well as “hidden dips” in the alignment.

3. Undulating grade lines involving substantial lengths of momentum (down) grades should be carefully evaluated with respect to operation of trucks.

4. Broken-back grade lines (two consecutive vertical curves in the same direction separated by a short tangent section) should be avoided wherever possible.

5. On long grades, it may be preferable to place the steepest grades at the bottom, lightening the grade on the ascent. If this is difficult, short sections of lighter grades should be inserted periodically to aid operations.

6. Where at-grade intersections occur on roadway sections with moderate to steep grades, the grade should be reduced or flattened through the intersection area.

7. Sag vertical curves in cuts should be avoided unless adequate drainage is provided.

27.4 Cross-Sectional Elements of Highways

The cross-section of a highway includes a number of elements critical to the design of the facility. The cross-section view of a highway is a 90° cut across the facility from roadside to roadside. The cross-section includes the following features:

- Travel lanes
- Shoulders
- Side slopes
- Curbs
- Medians and median barriers
- Guardrails
- Drainage channels

General design practice is to specify the cross-section at each station (i.e., at points 100 ft apart and at intermediate points where a change in the cross-sectional design occurs). The important cross-sectional features are briefly discussed in the sections that follow.

27.4.1 Travel Lanes and Pavement

Paved travel lanes provide the space that moving (and sometimes parked) vehicles occupy during normal operations. The standard width of a travel lane is 12 ft (metric standard is 3.6 m), although narrower lanes are permitted when necessary. The minimum recommended lane width is 9 ft. Lanes wider than 12 ft are sometimes provided on curves to account for the off-tracking of the rear wheels of large trucks. Narrow lanes will have a negative impact on the capacity of the roadway and on operations [3]. In general, 9-ft and 10-ft lanes should be avoided wherever possible. Nine-foot (9-ft) lanes are acceptable only on low-volume, low-speed rural or residential roadways, and 10 ft lanes are acceptable only on low-speed facilities.

All pavements have a cross-slope that is provided (a) to provide adequate drainage and (b) to provide superelevation on curves. For high-type pavements (portland cement concrete, asphaltic concrete), normal drainage cross-slopes range from 1.5% to 2.0%. On low-type pavements (penetration surfaces, compacted earth, etc.), the range of drainage cross-slopes is between 2% and 6%.

How the drainage cross-slope is developed depends upon the type of highway and the design of other drainage facilities. A pavement can be drained to *both* sides of the roadway or to one side. Where water is drained to both sides of the pavement, there must be drainage ditches or culverts and pipes on both sides of the pavement. In some cases, water drained to the roadside is simply absorbed into the earth; studies testing whether the soil is adequate to handle maximum expected water loads must be conducted before adopting this approach. Where more than one lane is drained to one side of the roadway, each successive lane should have a cross-slope that is 0.5% steeper than the previous lane. Figure 27.16 illustrates a typical cross-slope for a four-lane pavement.

Figure 27.16: Typical Highway Cross-Slope for Drainage

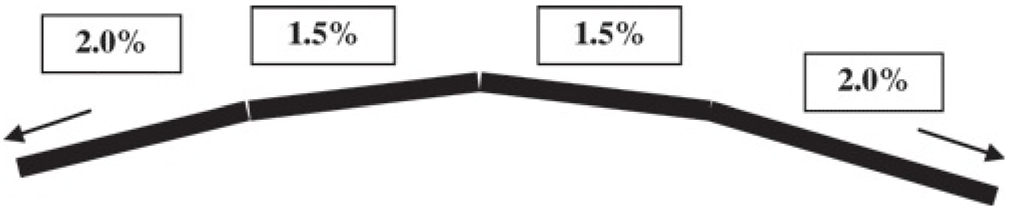


Figure 27.16: Full Alternative Text

On superelevated sections, cross-slopes are usually sufficient for drainage purposes, and a slope differential between adjacent lanes is not needed.

Superelevated sections, of course, must drain to the inside of the horizontal curve, and the design of drainage facilities must accommodate this.

27.4.2 Shoulders

AASHTO defines shoulders in the following way: “A shoulder is the portion of the roadway contiguous with the traveled way that accommodates stopped vehicles, emergency use, and lateral support of sub-base, base, and surface courses (of the roadway structure)” [Ref [1], pg 4–8]. Shoulders vary widely in both size and physical appearance. Normally, the shoulder width ranges from 2 ft to 12 ft. Most shoulders are “stabilized” (i.e., treated with some kind of material that provides a reasonable surface for vehicles). This can range from a fully paved shoulder to shoulders stabilized with penetration or stone surfaces or simply grass over compacted earth. For safety, it is critical that the joint between the traveled way and the shoulder be well maintained.

Shoulders are generally considered necessary on rural highways serving a significant mobility function, on all freeways, and on some types of urban highways. In these cases, a minimum width of 10 ft is generally used, as this provides for stopped vehicles to be about 2 ft clear of the traveled way. The narrowest 2 ft shoulders should be used only for the lowest classifications of highways. Even in these cases, 8 ft is considered desirable.

Shoulders serve a variety of functions, including the following:

- Providing a refuge for stalled or temporarily stopped vehicles
- Providing a buffer for accident recovery
- Contributing to driving ease and driver confidence
- Increasing sight distance on horizontal curves
- Improving capacity and operations on most highways
- Provision of space for maintenance operations and equipment
- Provision of space for snow removal and storage

- Provision of lateral clearance for signs, guardrails, and other roadside objects
- Improved drainage on a traveled way
- Provision of structural support for the roadbed

Reference [4] provides an excellent study of the use of roadway shoulders.

Table 27.9 shows recommended cross-slopes for shoulders, based on the type of surface. No shoulder should have a cross-slope of more than 7:1, as the probability of rollover is greatly increased for vehicles entering a more steeply sloped shoulder.

Table 27.9: Recommended Cross-Slopes for Shoulders

Type of Surface	Recommended Cross-Slope (%)
Bituminous	2.0–6.0
Gravel or stone	4.0–6.0
Turf	6.0–8.0

Table 27.9: Full Alternative Text

27.4.3 Side-Slopes for Cuts and Embankments

Where roadways are located in cut sections or on embankments, side-slopes must be carefully designed to provide for safe operation. In urban areas, sufficient right-of-way is generally not available to provide for natural side-slopes, and retaining walls are frequently used.

Where natural side-slopes are provided, the following limitations must be considered:

- A 3:1 side-slope is the maximum for safe operation of maintenance and mowing equipment.
- A 4:1 side-slope is the maximum desirable for accident safety. Barriers should be used to prevent vehicles from entering a side-slope area with a steeper slope.
- A 2:1 side-slope is the maximum on which grass can be grown, and only then in good climates.
- A 6:1 side-slope is the maximum that is structurally stable for where sandy soils are predominate.

Table 27.10 shows recommended side-slopes for various terrains and heights of cut and/or fill.

Table 27.10: Recommended Side-Slopes for Cut and Fill Sections

Height of Cut Or Fill (ft)	Terrain		
	Level or Rolling	Moderately Steep	Steep
0−4	6:1	4:1	4:1
4−10	4:1	3:1	2:1
10−15	3:1	2.5:1	1.75:1*
15−20	2:1	2:1	1.5:1*
>20	2:1	1.5:1*	1.5:1*

*Avoid where soils are subject to erosion.

Table 27.10: Full Alternative Text

27.4.4 Guardrail

One of the most important features of any cross-section design is the use and placement of guardrail. “Guardrail” is intended to prevent vehicles from entering a dangerous area of the roadside or median during an accident or intended action.

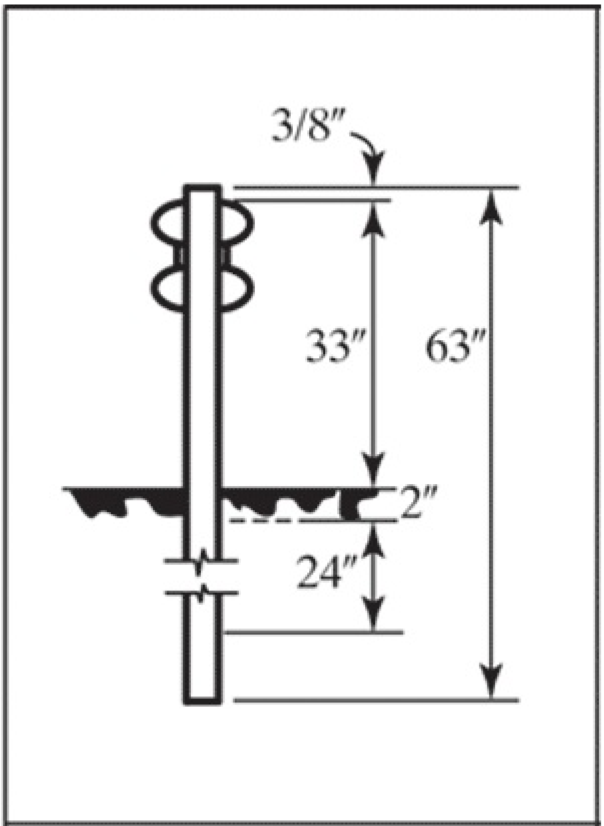
Roadside guardrail is provided to prevent vehicles from entering a cross-slope steeper than 4:1, or from colliding with roadside objects such as trees, culverts, lighting standards, sign posts, and so on. Once a vehicle hits a section of guardrail, the physical design also guides the vehicle into a safer trajectory, usually in the direction of traffic flow.

Median guardrail is primarily provided to prevent vehicles from encroaching into the opposing lane(s) of traffic. It also prevents vehicles from colliding with median objects. The need for median guardrail depends upon the design of the median itself. If the median is 20 ft or wider *and* if there are no dangerous objects in the median, guardrail is usually not provided, and the median is not curbed. Wide medians can effectively serve as accident recovery areas for encroaching drivers.

Narrower medians generally require some type of barrier, as the potential for encroaching vehicles to cross the entire median and enter the opposing traffic lanes is significant.

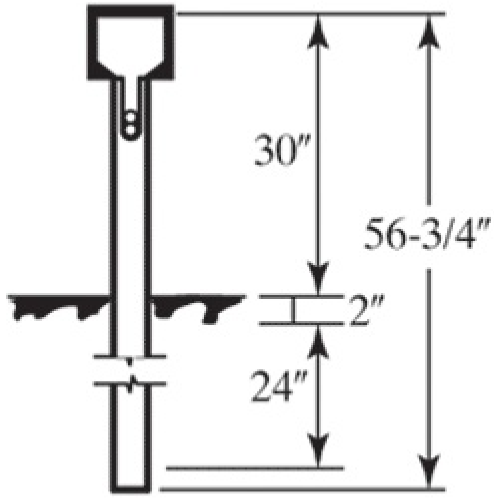
Figure 27.17 illustrates common types of guardrail in current use.

Figure 27.17: Common Types of Median and Roadside Barriers



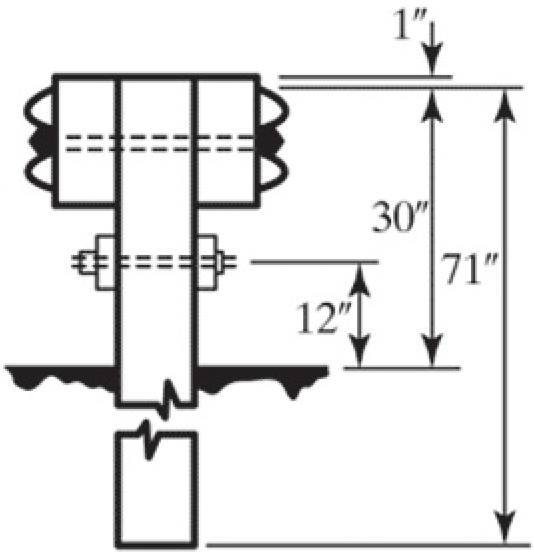
“W” Section (Steel Weak Post)

27.4-11 Full Alternative Text



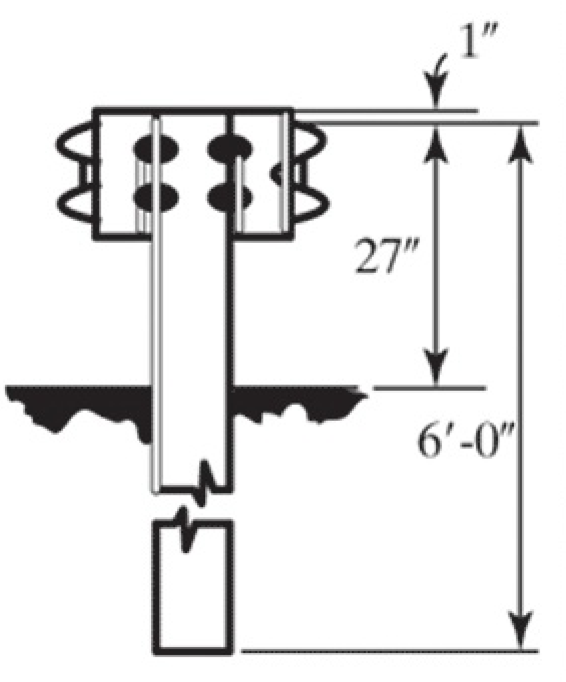
Box Beam

27.4-11 Full Alternative Text



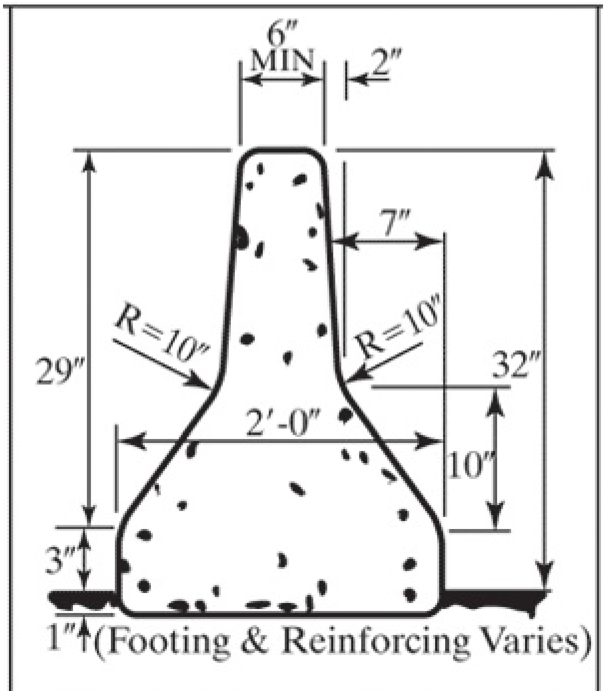
Blocked-Out "W" Beam (Wood Posts)

27.4-11 Full Alternative Text



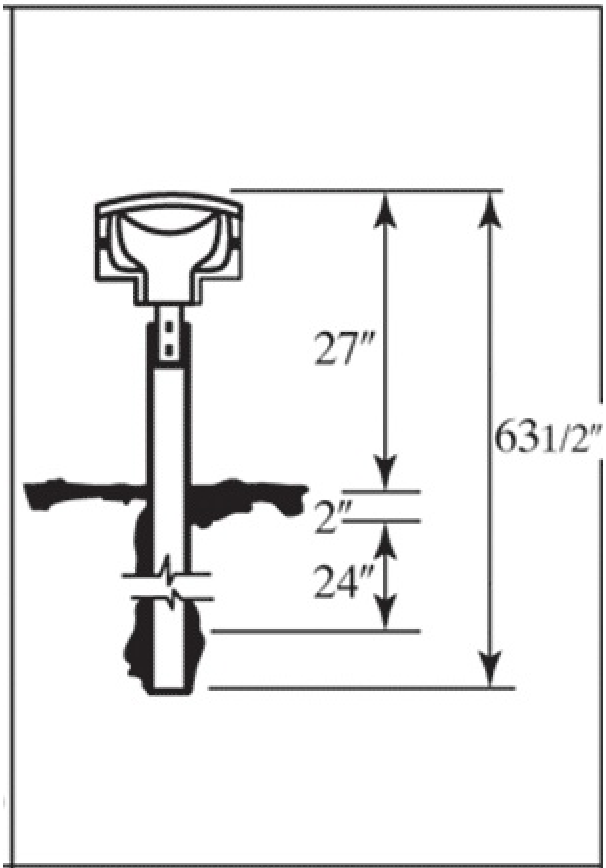
Blocked-Out "W" Beam (Steel Posts)

27.4-11 Full Alternative Text



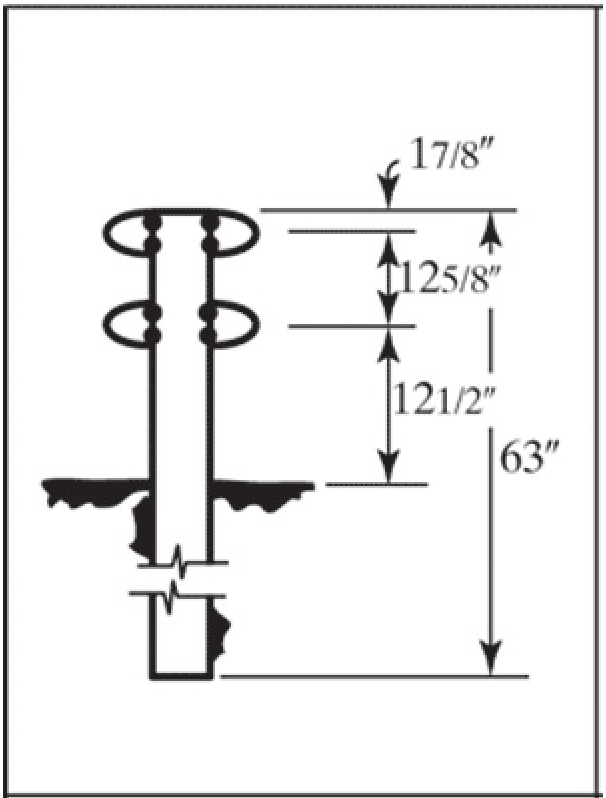
Concrete Median Barrier

27.4-11 Full Alternative Text



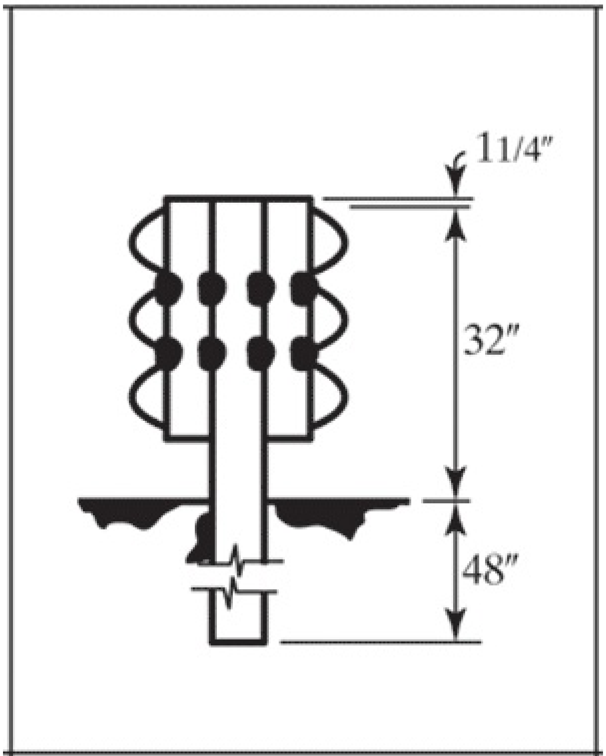
Aluminum Strong Beam

27.4-11 Full Alternative Text



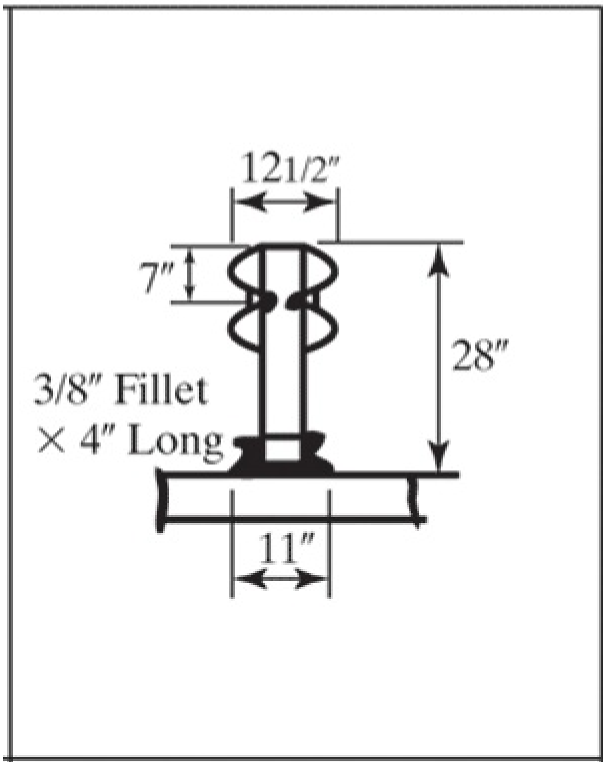
Aluminum Balanced Beam

27.4-11 Full Alternative Text



Blocked-Out Thrie Beam (Steel Post)

27.4-11 Full Alternative Text



Metal Beam Guard Fence (Steel) “Break-away” Post)

27.4-11 Full Alternative Text

(Source: Adapted from *A Policy on Geometric Design of Highways and Streets*, 6th Edition, American Association of State Highway and Transportation Officials, Washington, D.C., 2011.)

The major differences in the various designs are the flexibility of guardrail upon impact and the strength of the barrier in preventing a vehicle from crossing through the barrier.

The box-beam design, for example, is quite flexible. Upon collision, several posts of the box-beam will give way, allowing the beam to flex as much as 10 to 12 ft. The colliding vehicle is gently straightened and guided back toward the travel lane over a length of the guardrail. Obviously, this type of guardrail is not useful in narrow medians, as it could well deflect into the opposing traffic lanes.

The most inflexible design is the concrete median or roadside barrier. These blocks are almost immovable, and it is virtually impossible to crash through them. Thus, they are used in narrow roadway medians (particularly on urban freeways), and on roadsides where virtually no deflection would be safe. On collision with such a barrier, the vehicle is straightened out almost immediately, and the friction of the vehicle against the barrier brings it to a stop.

The details of guardrail design are critical. End treatments must be carefully done. A vehicle colliding with a blunt end of a guardrail section is in extreme danger. Thus, most W-beam and box-beam guardrails are bent away from the traveled way, with their ends buried in the roadside. Even with this done, vehicles can (with some difficulty) hit the buried end and “ramp up” the guardrail with one or more wheels. Various impact attenuating devices can also be used to protect the ends of such barriers. Concrete barriers have sloped ends, but are usually protected by impact-attenuating devices, such as sand or water barrels or mechanical attenuators.

Connection of guardrail to bridge railings and abutments is also important. As most guardrails deflect, they cannot be isolated from fixed objects, as they could conceivably “guide” a vehicle into a dangerous collision with such an object. Thus, where guardrails meet bridge railings or abutments, they are anchored onto the railing or abutment itself to ensure that encroaching vehicles are guided away from the object.

27.5 Closing Comments

This chapter has provided a brief overview of the critical functional and geometric characteristics of highways. There are many more details involved in highway geometry than those illustrated herein. The current AASHTO standard—*A Policy on Geometric Design of Highways and Streets*—should be consulted directly for a more detailed presentation of specific design practices and policies.

References

- 1. *A Policy on Geometric Design of Highways and Streets*, 6th Edition, American Association of State Highway and Transportation Officials, Washington, D.C., 2011.
- 2. Kavanagh, B.F., *Surveying With Construction Applications*, 4th Edition, Prentice-Hall, Upper Saddle River, NJ, 2001.
- 3. *Highway Capacity Manual, 6th Edition: A Multimodal Guide for Mobility Analysis*, National Research Council, Transportation Research Board, Washington, D.C., 2016.

Problems

1. 27-1 The point of intersection (*P.I.*) of two tangent lines is Station 11,500+66. The radius of curvature is 1,000 feet, and the angle of deflection is 60°. Find the length of the curve, the stations for the *P.C.* and *P.T.*, and all other relevant characteristics of the curve (*LC, M, E*).

2. 27-2 A 6° curve is to be designed on a highway with a design speed of 60 mi/h. Spiral transition curves are to be used. Determine the length of the spiral and the appropriate stations for the *T.S., S.C., C.S.*, and *S.T.* The angle of deflection for the original tangents is 40°, and the *P.I.* is at station 15,100+26. The segment consists of two 12-ft lanes.

3. 27-3 A 5° curve (measured at the centerline of the inside lane) is being designed for a highway with a design speed of 65 mi/h. The curve is on a 2% upgrade, and driver reaction time may be taken as 2.5 seconds. What is the closest any roadside object may be placed to the centerline of the inside lane of the roadway while maintaining adequate stopping sight distance?

4. 27-4 What is the appropriate superelevation rate for a curve with a 1,200-ft radius on highway with a design speed of 60 mi/h? The maximum design superelevation is 6% for this highway.

5. 27-5 What length of superelevation runoff should be used to achieve a superelevation rate of 10%? The design speed is 70 mi/h, and a three-lane cross-section (12 ft lanes) is under consideration. Superelevation will be achieved by rotating all three lanes around the inside edge of the pavement.

6. 27-6 Find the maximum allowable grade and critical length of grade for each of the following facilities:

    1. A rural freeway in mountainous terrain with a design speed of 60 mi/h

    2. A rural arterial in rolling terrain with a design speed of 45 mi/h

    3. An urban arterial in level terrain with a design speed of 40 mi/h

7. 27-7 A vertical curve of 1,000 ft is designed to connect a grade of +4% to a grade of −5%. The *V.P.I.* is located at station 1,500 + 55 and has a known elevation of 500 ft. Find the following:

    1. The station of the *V.P.C.* and the *V.P.T.*

    2. The elevation of the *V.P.C.* and the *V.P.T.*

    3. The elevation of points along the vertical curve at 100-ft intervals

    4. The location and elevation of the high point on the curve

8. 27-8 Find the minimum length of curve for the following scenarios:

Entry Grade	Exit Grade	Design Speed	Reaction Time
3%	8%	45 mi/h	2.5 s
−4%	2%	65 mi/h	2.5 s
0%	−3%	70 mi/h	2.5 s

27.2-11 Full Alternative Text

9. 27-9 A vertical curve is to be designed to connect a −4% grade to a +1% grade on a facility with a design speed of 70 mi/h. For economic reasons, a minimum-length curve will be provided. A driver-reaction time of 2.5 seconds may be used in sight distance determinations. The *VPI* of the curve is at station 5,100 + 22 and has an elevation of 1,285 ft. Find the station and elevation of the *VPC* and *VPT*, the high point of the curve, and at 100-ft intervals along the curve.

Chapter 28 Capacity and Level of Service Analysis: Basic Freeway and Multilane Highway Segments

The procedures in this chapter cover the capacity and level of service analysis of multilane highway segments under uninterrupted flow. These include freeway segments outside the influence of turbulence areas (ramps, weaving segments) and multilane highway segments far enough away from the nearest traffic signal (approximately 2 miles) to be considered as uninterrupted flow facilities.

28.1 Facility Types Included

Freeways are the only types of facilities providing pure uninterrupted flow. All entries to and exits from freeways are made using ramps designed to allow such movements to occur without interruption to the freeway traffic stream. There are no at-grade intersections (either signalized or unsignalized), no driveway access, and no parking permitted within the right-of-way. Full control of access is provided. Freeways are generally classified by the total number of lanes provided in both directions, for example, a six-lane freeway has three lanes in each direction. Common categories are four-, six-, and eight-lane freeways, although some freeway sections in major urban areas may have ten or more lanes in specific segments.

Multilane surface facilities should be classified and analyzed as urban streets (arterials) if signal spacing is less than 2 miles. Uninterrupted flow can exist on multilane facilities where the segment is at least 2 miles away from the nearest signal.

Multilane highway segments are classified by the number of lanes and the type of median treatment provided. Surface multilane facilities generally consist of four- or six-lane alignments. They can be *undivided* (i.e., having no median but with a double-solid yellow marking separating the two directions of flow), or *divided*, with a physical median separating the two directions of flow. In suburban areas, a third median treatment is also used: the two-way left-turn lane. This treatment requires an alignment with an odd number of lanes—most commonly three, five, or seven. The center lane is used as a continuous left-turn lane for both directions of flow.

The median treatment of a surface multilane highway can have a significant impact on operations. A physical median prevents mid-block left turns across the median except at locations where a break in the median barrier is provided. Mid-block left turns can be made at any point on an undivided alignment. Where a two-way left-turn lane is provided, mid-block left turns are permitted without restriction, but vehicles waiting to turn do so in the special lane and do not unduly restrict through vehicles.

In terms of capacity analysis procedures, both basic freeway sections and multilane highways are categorized by the *free-flow speed* (FFS). By definition, the FFS is the speed intercept when flow is “zero” on a calibrated speed-flow curve. In practical terms, it is the average speed of the traffic stream when flow rates are less than approximately 1,000 veh/h/ln.

Figure 28.1 illustrates some common freeway and multilane alignments.

Figure 28.1: Typical Freeway and Multilane Highway Alignments

(a) A Typical 8-Lane Freeway

(b) A Divided Multilane Rural Highway

(c) A Divided Multilane Suburban Highway

(d) An Undivided Multilane Suburban Highway

(e) An Multilane Highway w/TWLTL

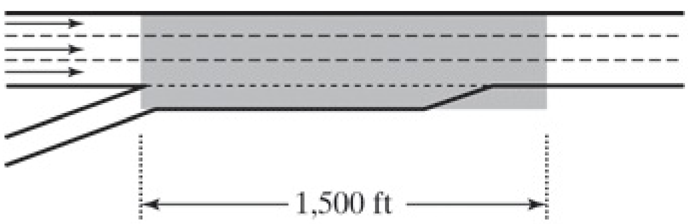
(f) An Undivided Multilane Rural Highway

(Source: Photo (a) courtesy of J. Ulerio; (b),(c),(d),(f) Used with permission of Transportation Research Board, National Research Council, “Highway Capacity Manual,” *Special Report 209*, 1994, Illustrations 7-1 through 7-4, pg 7–3; (e) Used with permission of Transportation Research Board, National Research Council, *Highway Capacity Manual*, Dec 2000, Illustration 12-8, pg 12–6.)

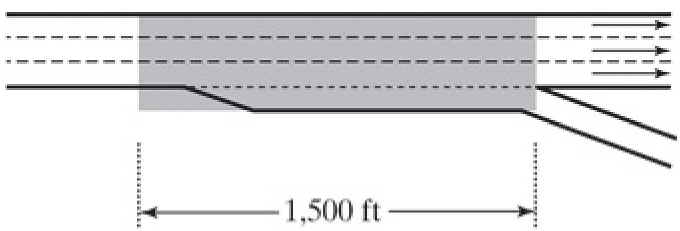
28.2 Segment Types on Freeways and Some Multilane Highways

Freeways are comprised of four types of segments. Three are areas of high turbulence within the traffic stream, and are treated by separate methodologies discussed in Chapters 29 and 30 of this text. *Weaving segments* involve traffic streams that effectively cross each-others path over a length of highway without the aid of traffic control devices (other than guide or warning signs). *Merge segments* occur at on-ramps or other segments where two or more separate flows are merging to form a single traffic stream. *Diverge segments* occur at off-ramps or other segments where a single traffic stream separates into two or more separate flows. Figure 28.2 illustrates merge, diverge, and weaving segments and their operational influence areas.

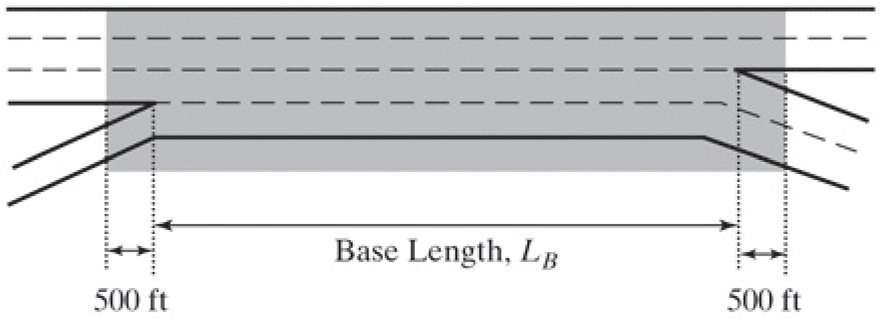
Figure 28.2: Types of Segments on Freeways and Some Multilane Highways



(a) Merge Segment



(b) Diverge Segment



(Source: Reprinted with permission from *Highway Capacity Manual, 6th Edition: A Guide for Multimodal Mobility Analysis*, Transportation Research Board, National Academy of Sciences, Courtesy of the National Academies Press, Washington, D.C., 2016.)

The influence area for merge segments extends 1,500 feet downstream of the merge point. For diverge segments, the influence area extends 1,500 feet upstream of the diverge point. For weaving segments, the influence area extends 500 feet upstream and downstream of the beginning and end of the segment.

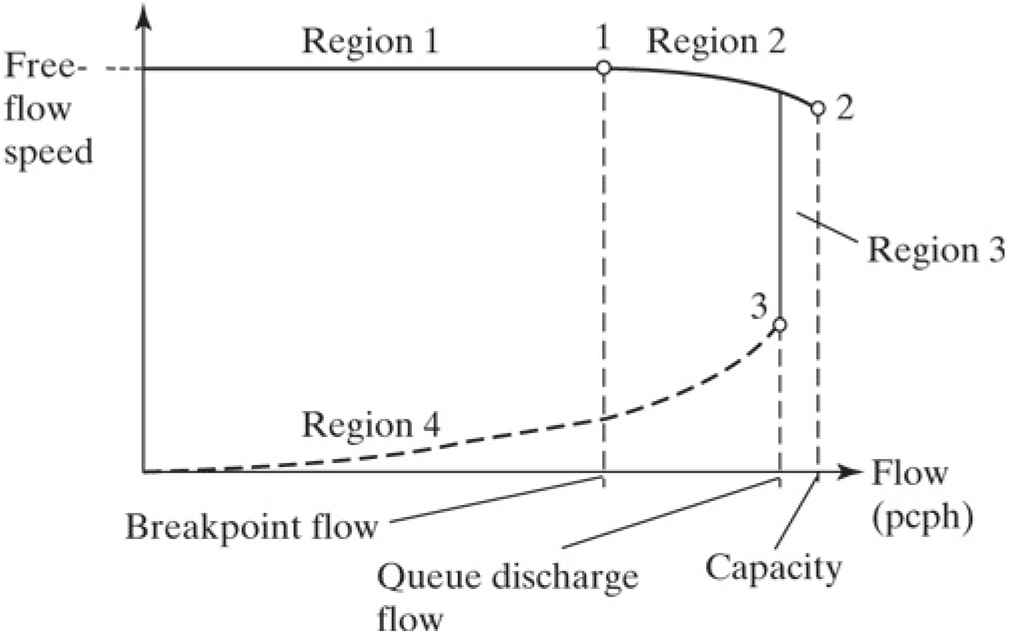
Basic freeway segments are all parts of the freeway that are *not* part of a merge, diverge, or weaving influence area. It should be noted that merge, diverge, and weaving segments may also exist on multilane highways.

28.3 Generic Speed-Flow Characteristics on Freeways and Multilane Highways

Capacity analysis procedures for basic freeway segments and multilane highways are based upon calibrated speed-flow curves for segments with various free-flow speeds, operating under *base conditions*. There is only one base condition for such segments in the 2016 *Highway Capacity Manual* (HCM): The traffic stream consists of only passenger cars. All other elements, such as lane widths and lateral clearances, driver population, ramp or roadside access density and others, are taken into account in establishing the base speed-flow curve for analysis in any given case.

Figure 28.3 shows the generic form of speed-flow curves used in the methodology. The figure shows a broad range of flow rates over which average speeds remain constant—at the FFS. Modern drivers maintain high average speeds at relatively high rates of flow on freeways and multilane highways. In most cases, flow rates have no impact on average speeds for flow rates up to 1,000 pc/h/ln, and in some cases considerably higher.

Figure 28.3: Generic Speed-Flow Curve for Freeways and Multilane Highways

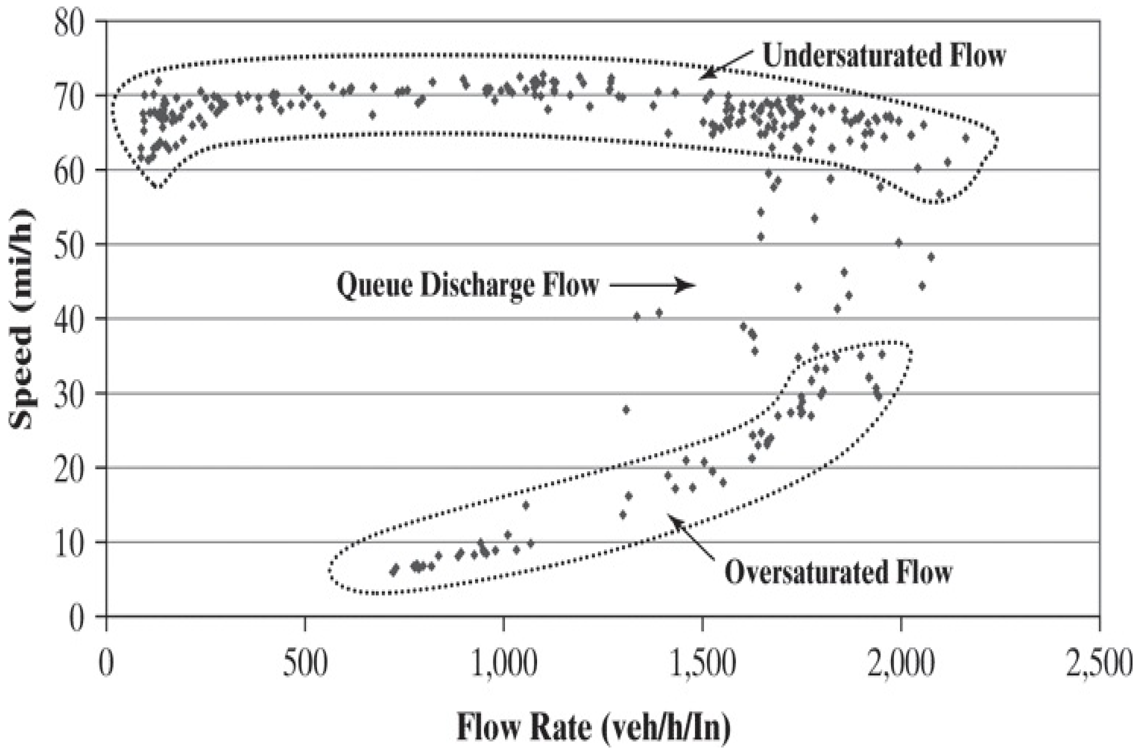


[Figure 28.3: Full Alternative Text]

Region 1 is that portion of the curve where the FFS prevails. Region 2 defines that portion of the curve where increasing flows result in decreasing speeds. Point 2 represents operation at capacity, which is highly unstable. Additional demand flows beyond capacity cause a breakdown, and the development of a queue. Region 3 represents the departure flow from the queue, or *queue discharge flow*. In most cases, it is *less* than capacity, with default values in the range of 5% to 10% less than capacity in common use. Region 4 represents unstable flow *within* the queue.

Figure 28.4 shows real speed-flow data from I-405 (Los Angeles) from 2004. It clearly shows three distinct regions of the curve: undersaturated flow (before capacity is reached), queue discharge flow, and oversaturated flow within the queue.

Figure 28.4: Speed-Flow Data from California



(Source: Reprinted with permission from *Highway Capacity Manual, 6th Edition: A Guide for Multimodal Mobility Analysis*, Transportation Research Board, National Academy of Sciences, Courtesy of the National Academies Press, Washington, D.C., 2016.)

Figure 28.4: Full Alternative Text

Note that queue discharge flow is not a clear-cut value, but rather a broad range of points. An average value is normally adopted, but discharge rates can vary substantially over time.

28.4 Levels of Service for Freeways and Multilane Highways

For freeways and multilane highways, the measure of effectiveness used to define levels of service is *density*. The use of density, rather than speed, is based primarily on the shape of the speed-flow relationships depicted in Figures 28.3 and 28.4. Because average speed remains constant through most of the range of flows and because the total difference between FFS and the speed at capacity is relatively small, defining five level-of-service boundaries based on speed would be problematic.

If flow rates vary while speeds remain relatively stable, then density must be varying throughout the range of flows, given the basic relationship that $v = S \times D$. Further, density describes the proximity of vehicles to each other, which is the principal determinant of freedom to maneuver. Thus, it is an appropriate descriptor of service quality.

For uninterrupted flow facilities, the density boundary between levels of service E and F is defined as the density at which capacity occurs. For both freeways and multilane highways, the density at capacity is approximately 45 pc/mi/ln.

Other level-of-service boundaries are set judgmentally by the HCQSC to provide reasonable ranges of both density and service flow rates. Table 28.1 shows the defined level-of-service criteria for basic freeway sections and multilane highways, which are the same.

Table 28.1: Level of Service Criteria for Basic Freeway Segments and Multilane Highways

Level of Service	Density Range for Basic Freeway Sections (pc/mi/ln)	Density Range for Multilane Highways (pc/mi/ln)
A	$\geq 0 \leq 11$	$\geq 0 \leq 11$
B	$> 11 \leq 18$	$> 11 \leq 18$
C	$> 18 \leq 26$	$> 18 \leq 26$
D	$> 26 \leq 35$	$> 26 \leq 35$
E	$> 35 \leq 45$	$> 35 \leq 45$
F	Demand Exceeds Capacity >45	Demand Exceeds Capacity >45

Table 28.1: Full Alternative Text

The general operating conditions for these levels of service can be described as follows:

- Level of service A is intended to describe free-flow operations. At these low densities, the operation of each vehicle is not greatly influenced by the presence of others. Speeds are not affected by flow in this level of service, and operation is at the FFS. Lane changing, merging, and diverging maneuvers are easily accomplished, as many large gaps in lane flow exist. Short-duration lane blockages may cause the level of service to deteriorate somewhat, but do not cause significant disruption to flow. Average spacing between vehicles is a minimum of 480 ft, or approximately 24 car lengths, at this level of service.

- At level of service B, drivers begin to respond to the existence of other vehicles in the traffic stream, although operation is still at the FFS. Maneuvering within the traffic stream is still relatively easy, but drivers must be more vigilant in searching for gaps in lane flows. The traffic stream still has sufficient gaps to dampen the impact of most minor lane disruptions. Average spacing is a minimum of 293 feet, or approximately 15 car lengths.

- At level of service C, the presence of other vehicles begins to restrict maneuverability within the traffic stream. Operations remain at the FFS, but drivers now need to adjust their course to find gaps they can use to pass or merge. A significant increase in driver vigilance is required at this level. Although there are still sufficient gaps in the traffic stream to dampen the impact of minor lane blockages, any significant blockage could lead to breakdown and queuing. Average spacing is a minimum of 203 feet, or approximately 10 car lengths.

- Level of service D is the range in which average speeds begin to decline with increasing flows. Density deteriorates more quickly with flow in this range. At level of service D, breakdowns can occur quickly in response to small increases in flow. Maneuvering within the traffic stream is now quite difficult, and drivers often have to search for gaps for some time before successfully passing or merging. The ability of the traffic stream to dampen the impact of even minor lane disruptions is severely restricted, and most such blockages result in queue formation unless removed very quickly. Average spacing is a minimum of 151 feet, or approximately seven car lengths.

- Level of service E represents operation in the vicinity of capacity. The maximum density limit of level of service E is capacity operation. For such an operation there are few or no usable gaps in the traffic stream, and any perturbation caused by lane-changing or merging maneuvers will create a shock wave in the traffic stream. Even the smallest lane disruptions may cause extensive queuing. Maneuvering within the traffic stream is now very difficult, as other vehicles must give way to accommodate a lane-changing or merging vehicle. The average spacing is a minimum of 117 feet, or approximately six car lengths.

- Level of service F describes operation within the queue that forms upstream of a breakdown point. Such breakdowns may be caused by accidents or incidents, or may occur at locations where arrival demand exceeds the capacity of the section on a regular basis. Actual operating conditions vary widely, and are subject to short-term perturbations. As vehicles “shuffle” through the queue, there are times when they are standing still, and times when they move briskly for short distances. Level of service *F* is also used to describe the point of the breakdown, where demand flow (v) exceeds capacity (*c*). In reality, operation at the point of the breakdown is usually good, as vehicles discharge from the queue. Nevertheless, it is insufficient capacity at the point of breakdown that causes the queue, and level of service F provides an appropriate descriptor for this condition.

Note that in Table 28.1, LOS F is identified when the density is higher than 45 pc/mi/ln or when “demand exceeds capacity,” that is, the *v/c* ratio>1.00. That is because no density can be predicted for such cases. Thus, in analysis, LOS F exists when demand exceeds capacity, and the density will be higher than 45 pc/h/ln.

28.5 Base Speed-Flow Curves

In the 2016 HCM, the discrete speed-flow curves of the 2010 HCM were replaced with a procedure in which a base FFS for a given segment is developed for use. This accommodates the reality that FFS, which characterizes each curve, is a continuous variable, not a series of discrete options. It also reflects the development of several new approaches to adjustments implemented in the 2016 HCM.

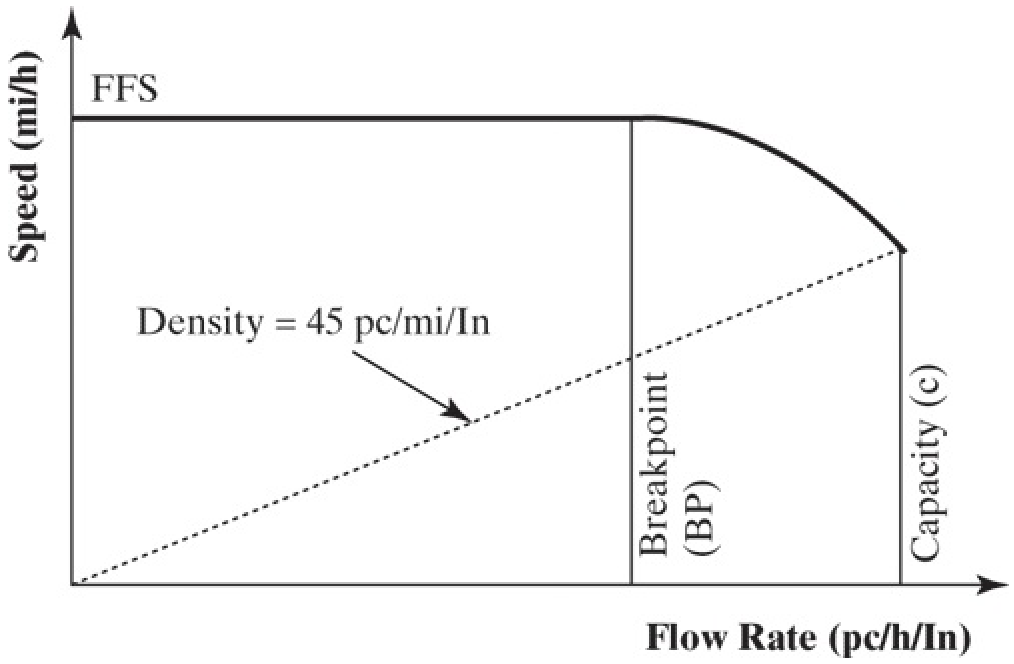
In essence, however, the 2016 HCM is still based upon the calibrated curves in the 2010 HCM, as no new data were collected or analyzed. Several problems with the calibration of the 2010 curves, therefore, still remain [2]:

- The multilane highway speed-flow curves have not been calibrated with new data for over 25 years.
- At the direction of the Highway Capacity and Quality of Service Committee (HCQSC), freeway speed-flow curves were "adjusted" to guarantee that service flow rates on multilane highways were always lower than on freeways with the same FFS.
- While the 2010 curves showed that the constant-speed portion was smaller than previous editions, particularly at high values of FFS, the actual data showed that they should have been even smaller.
- The arbitrary adjustments made suggest that at LOS C and D, service flow rates are being overpredicted by the curves.

All of these problems were transferred to the 2016 curves. Figure 28.5 shows the format of the speed-flow curves used in the 2016 HCM.

Figure 28.5: Base Form of Speed-Flow Curves for Freeways and Multilane

Highways



(Source: Reprinted with permission from *Highway Capacity Manual, 6th Edition: A Guide for Multimodal Mobility Analysis*, Transportation Research Board, National Academy of Sciences, Courtesy of the National Academies Press, Washington, D.C., 2016.)

Figure 28.5: Full Alternative Text

Developing a specific curve for an analysis segment involves determining the FFS, the breakpoint flow rate (BP), and the capacity (c). All other elements of the curve follow a standard form.

28.5.1 Base Equation for Speed-Flow Curves

The equation for all speed-flow curves has two portions: (a) for values of flow rate less than or equal to the breakpoint flow (BP), speed is a constant —the FFS; (b) for flow between the breakpoint and capacity (c), a curve of standard format is applied. The form of the base speed-flow curve is as follows:

S=FFSadjvp≤BPS=FFSadj − [(FFSadj−cadj45) (vp−BP)a(cadj−BP)a]vp>BP [28-1]

where:

FFSadj=adjusted free-
flow speed, mi/h,cadj=adjusted capacity, pc/h/ln,BP=breakpoint flow rate,

Table 28.2 shows how these parameters are determined.

Table 28.2: Determining Parameters for a Specific Speed-Flow Curve

Parameter	Basic Freeway Segment	Multilane Highway Segment
FFS (mi/h)	Measure or Predict[1]	Measure or Predict[1]
FFS_{adj} (mi/h)	FFS × SAF	FFS
SAF[2]	See Factors[2]	No Adjustment Permitted
c (pc/h/ln)	2,200 + 10 (FFS-50) c ≤ 2,400 pc/h/ln and 55 ≤ FFS ≤ 75	1,900 + 20(FFS-45) c ≤ 2,300 pc/h/ln and 45 ≤ FFS ≤ 70
c_{adj} (pc/h/ln)	c × CAF	c
CAF[3]	See Factors[3]	No Adjustment Permitted
BP	[1000 + 40 × (75-FFS_{adj})] × CAF[2]	1,400 pc/h/ln
a	2.00	1.31

1. Methodology for prediction of FFS is discussed in a subsequent section of this chapter.

2. SAF=speed adjustment factor; accounts for impacts of poor weather, incidents, work zones, and driver population; SAF discussed in a subsequent section of this chapter.

3. CAF=capacity adjustment factor; accounts for impacts of poor weather, incidents, work zones, and driver population; CAF discussed in a subsequent section of this chapter.

(Source: Modified from *Highway Capacity Manual, 6th Edition: A Guide for Multimodal Mobility Analysis*, Transportation Research Board, Washington, D.C., 2016, Exhibit 12-6, pg 12-10.)

Table 28.2: Full Alternative Text

There are some clear differences between freeways and multilane highways in the base curves. The biggest difference is that the breakpoint for freeways is variable, based upon the FFS and the SAF. For multilane highways, all curves break at 1,400 pc/h/ln. This may represent a real difference in the operating characteristics of these facilities, but more likely occurs due to the age of the multilane calibrations compared to those for freeways, which use far more recent data.

In general, speed adjustment factors (SAF) and capacity adjustment factors (CAF) may be applied to basic freeway segments, but may not be applied to multilane highways, as these effects (weather, incidents, work zones, driver population) have not been specifically studied for this class of highway. Determination of the SAF and CAF is discussed in a later section of this chapter.

28.5.2 Measuring or Predicting the Free-Flow Speed

The free-flow speed of a facility is best determined by field measurement. Given the shape of speed-flow relationships for freeways and multilane highways, an average speed measured when flow is less than or equal to 1,000 veh/h/ln may be taken to represent the FFS.

When new facilities or redesigned facilities are under consideration, it is not possible to measure FFS. Even for existing facilities, the time and cost of conducting field studies may not be warranted.

For such cases, models have been developed that allow the analyst to estimate the FFS based upon characteristics of the segment under study.

Estimating FFS for Freeways

The FFS of a freeway can be estimated as:

$FFS = 75.4 - f_{LW} - f_{LC} - 3.22\ TRD^{0.84}$ [28-2]

where:

FFS = free-flow speed of the freeway, mi/h, f_{LW} = adjustment for lane width, mi/h, f_{LC}=
side lateralclearance, mi/h, and TRD = total ramp density, ramps/mi.

The base condition for lane width is an average width of 12 feet or greater. For narrower lanes, the base FFS is reduced by the factors shown in Table 28.3.

Table 28.3: Adjustment to Free-Flow Speed for Lane Width on a Freeway

Lane Width (ft)	Reduction in Free-Flow Speed, f_{LW} (mi/h)
$\geq$ 12	0.0
11	1.9
10	6.6

(Source: Reprinted with permission from Transportation Research Board, National Research Council, *Highway Capacity Manual*, National Academy of Sciences, Courtesy of the National Academies Press, Washington, D.C., 2000.)

Table 28.3: Full Alternative Text

The base lateral clearance is 6 feet or greater on the right side and 2 feet or greater on the median, or left, side of the basic freeway section. Adjustments for right-side lateral clearances less than 6 feet are given in Table 28.4. There are no adjustments provided for median clearances less than 2 feet, as such conditions are considered rare.

Table 28.4: Adjustment to Free-Flow Speed for Lateral Clearance on a Freeway

Right Shoulder Lateral Clearance (ft)	Reduction in Free-Flow Speed, f_{LC} (mi/h)			
	Lanes in One Direction			
	2	**3**	**4**	**≥ 5**
≥ 6	0.0	0.0	0.0	0.0
5	0.6	0.4	0.2	0.1
4	1.2	0.8	0.4	0.2
3	1.8	1.2	0.6	0.3
2	2.4	1.6	0.8	0.4
1	2.0	2.0	1.0	0.5
0	3.6	2.4	1.2	0.6

(Source: Reprinted with permission from Transportation Research Board, National Research Council, *Highway Capacity Manual*, National Academy of Sciences, Courtesy of the National Academies Press, Washington, D.C., 2000.)

Table 28.4: Full Alternative Text

Care should be taken in assessing whether an “obstruction” exists on the right side of the freeway. Obstructions may be continuous, such as a guardrail or retaining wall, or they may be periodic, such as light supports and bridge abutments. In some cases, drivers may become accustomed to some obstructions, and the impact of these on FFSs may be minimal.

Right-side obstructions primarily influence driver behavior in the right lane. Drivers “shy away” from such obstructions, moving further to the left in the lane. Drivers in adjacent lanes may also shift somewhat to the left in response to vehicle placements in the right lane. The overall effect is to cause vehicles to travel closer to each other laterally than would normally be the case, thus making flow less efficient. This is the same effect as for narrow lanes. Since the primary impact is on the right lane, the total impact on FFS declines as the number of lanes increases.

Total ramp density is the total number of on-ramps and off-ramps within ±3 miles of the midpoint of the study segment, divided by 6 miles. Ramp density is a surrogate measure that relates to the intensity of land use activity in the vicinity of the study segment. In practical terms, drivers will drive at lower speeds where there are frequent on- and off-ramps creating turbulence in the traffic stream.

Estimating FFS for Multilane Highways

The FFS for a multilane highway may be estimated as:

$$FFS = BFFS - f_{LW} - f_{LC} - f_M - f_A \quad [28\text{-}3]$$

where:

FFS=speed of the multilane highway, mi/h,BFFS=base free-flow speed (as discussedbelow), mi/h,fLW=adjustment for lane width, mi/h

There is not a great deal of guidance in the HCM as to the base FFS for use in Equation 28-2. The design speed may be used as a reasonable surrogate, if it is known. The speed limit can be used to develop a rough estimate: BFFS can be roughly estimated as the speed limit+7 mi/h for speed limits less than 45 mi/h, or the speed limit+5 mi/h for speed limits over 45 mi/h. In the complete absence of any other information, a default value of 60 mi/h may be used as a last resort.

The base lane width for multilane highways is 12 ft, as was the case for freeways. For narrower lanes, the FFS is reduced by the values shown in Table 28.4. This adjustment is the same for multilane highways as for freeways.

For multilane highways, the lateral clearance adjustment is based on the *total lateral clearance*, which is the sum of the lateral clearances on the right side of the roadway and on the left (median) side of the roadway. Although this seems like a simple concept, there are some details that must be observed:

- A lateral clearance of 6 ft is the base condition. Thus, no right- or left-side lateral clearance is ever taken to be greater than 6 ft, even if greater clearance physically exists. Thus, the base total lateral clearance is 12 ft (6 ft for the right side, 6 ft for the left or median side).

- For an undivided multilane highway, there is no left-or median-side lateral clearance. However, there is a separate adjustment taken for type of median, including the undivided case. To avoid double-counting the impact of an undivided highway, the left or median lateral clearance on an undivided highway is assumed to be 6 ft.

- For multilane highways with two-way left-turn lanes, the left or median lateral clearance is also taken as 6 ft.

- For a divided multilane highway, the left- or median-side lateral clearance may be based on the location of a median barrier, periodic objects (light standards, abutments, etc.) in the median, or the distance to the opposing traffic lane. As noted previously, the maximum value is 6 ft.

The adjustments to FFS for total lateral clearance on a multilane highway are shown in Table 28.5.

Table 28.5: Adjustment to

Free-Flow Speed for Total Lateral Clearance on a Multilane Highway

4-Lane Multilane Highways		6-Lane Multilane Highways	
Total Lateral Clearance (ft)	Reduction in Free-Flow Speed, f_{LC} (mi/h)	Total Lateral Clearance (ft)	Reduction in Free-Flow Speed, f_{LC} (mi/h)
≥12	0.0	≥12	0.0
10	0.4	10	0.4
8	0.9	8	0.9
6	1.3	6	1.3
4	1.8	4	1.7
2	3.6	2	2.8
0	5.4	0	3.9

(Source: Reprinted with permission from Transportation Research Board, National Research Council, *Highway Capacity Manual*, National Academy of Sciences, Courtesy of the National Academies Press, Washington, D.C., 2000.)

Table 28.5: Full Alternative Text

The median-type adjustment is shown in Table 28.6. A reduction of 1.6 mi/h is made for undivided configurations, whereas divided multilane highways, or multilane highways with two-way left-turn lanes, represent base conditions.

Table 28.6: Adjustment to

Free-Flow Speed for Median Type on Multilane Highways

Median Type	Reduction in Free-Flow Speed, f_M (mi/h)
Undivided	1.6
TWLTL	0.0
Divided	0.0

(Source: Reprinted with permission from Transportation Research Board, National Research Council, *Highway Capacity Manual*, National Academy of Sciences, Courtesy of the National Academies Press, Washington, D.C., 2000.)

Table 28.6: Full Alternative Text

A critical adjustment to base FFS is related to access-point density. Access-point density is the average number of unsignalized driveways or roadways per mile that provide access to the multilane highway on *the right* side of the roadway (for the subject direction of traffic).

Driveways or other entrances with little traffic, that, for other reasons, do not affect driver behavior should not be included in the access-point density. Adjustments are shown in Table 28.7.

Table 28.7: Adjustment to Free-Flow Speed for Access-Point Density on a Multilane Highway

Access Density (Access Points/Mi)	Reduction in Free-Flow Speed, f_A (mi/h)
0	0.0
10	2.5
20	5.0
30	7.5
≥40	10.0

(Source: Reprinted with permission from Transportation Research Board, National Research Council, *Highway Capacity Manual*, National Academy of Sciences, Courtesy of the National Academies Press, Washington, D.C., 2000.)

Table 28.7: Full Alternative Text

Sample Problem 28-1: Determing the Free-Flow Speed

A basic freeway segment and a multilane highway segment are described in Table 28.8. From the information given, estimate the FFS for each segment.

Table 28.8: Sample Problem in Determining *FFS*

Parameter	Freeway Segment	Multilane Highway Segment
Number of Lanes	6	4
Lane Width	12 ft	11 ft
Lateral Clearance, Right Side	4 ft	2 ft
Alignment	NA	Undivided
Base Free-Flow Speed	NA	65.0 mi/h
Ramps/Mile	4	NA
Access Points Per Mile	NA	20

Table 28.8: Full Alternative Text

Equations 28-2 and 28-3 are used to estimate FFS for freeways and multilane highway segments, respectively. The FFS for the basic freeway segment is computed as:

FFS=75.4−fLW−fLC−3.22 TRD0.84

where:

fLW=0.0 mi/h (Table 28.3, 12 ft lanes),fLC=0.8 mi/h (Table 28.4, 4 ft clea

Then:

FFS=75.4−0.0−0.9−3.22 (40.84)=64.3 mi/h

The FFS for the multilane highway segment is computed as:

FFS=BFFS−fLW−fLC−fM−fA

where:

fLW=1.9 mi/h (Table 28.3, 11 ft lanes),fLC=0.9 mi/h (
Table 28.5, 2+6=8 ft total lateral clearance, 4 lanes
),fM=1.6 mi/h (Table 28.6, undivided),fA=5.0 mi/h (Table 28.7, 20 access p

Then:

$FFS = 65.0 - 1.9 - 0.9 - 1.6 - 5.0 = 55.6$ mi/h

28.5.3 Capacity Adjustment Factors and Speed Adjustment Factors

The vast majority of capacity or level of service analyses are conducted for assumed conditions which include the following:

- Good weather
- No traffic incidents or accidents
- No work zones
- Motorists who are regular and familiar users of the facility

Over the years, there has been considerable interest in assessing the impact of such conditions when one or more of these assumed conditions do not exist. The 2016 HCM introduces the CAF and the SAF to allow this.

The normal capacity value can be multiplied by a CAF to reflect the impact of adverse weather, lane blockages due to incidents or accidents, work zones, and motorist populations that are not regular users of the facility in question. Similarly, the FFS (or a base speed estimate) can be adjusted by a SAF to reflect impact of these situations on speed.

There are four components that impact the applicable CAF in any given case, and three components that impact the applicable SAF. Table 28.9 illustrates these.

Table 28.9: Factors Considered in *CAF* and *SAF*

Condition	CAF	SAF
Weather	Yes	Yes
Incidents	Yes	No
Driver Population	Yes	Yes
Work Zones	Yes	Yes

Table 28.9: Full Alternative Text

Note that an adjustment for incidents is only recommended for capacity. No speed adjustment is permitted to be made for incidents, due to a lack of research on the subject.

It is possible to apply CAFs and SAFs for more than one factor in any given situation, as indicated in Equations 28-4:

$$CAF = CAFW \times CAFI \times CAFDP \times CAFWZ$$
$$SAF = SAFW \times SAFDP \times SAFWZ \quad [28\text{-}4]$$

where:

CAFW=capacity adjustment factor foradverse weather,SAFW=speed adjus
standarddriver population,SAFDP=speed adjustment factor for non-
standarddriver population,CAFWZ=capacity adjustment factor for workzor

While the use of multiple factors is permitted, it should be done with great caution, as each of these adjustment factors was researched and calibrated in isolation. No field calibrations have considered whether the impacts of several are fully multiplicative, as there may be overlapping operational effects. In other words, using these factors to estimate a capacity and FFS in a work zone with an incident on a snowy day with weekend drivers is probably not a good idea.

As indicated in Table 28.9, the CAF is used to modify the estimated capacity under base conditions, and the *SAF* is used to modify the FFS. Both, in turn, affect the base equation for the speed-flow curve in any given application.

Adjustments for Inclement

Weather

Tables 28.10 and 28.11 show *CAFs* and *SAFs* for various forms of inclement weather. They are based upon a comprehensive NCHRP study of default values for use in capacity analysis [3].

Table 28.10: Capacity Adjustment Factors (*CAF*) for Inclement Weather

Weather	Definition of Conditions	*CAF* for Unadjusted *FFS* of: (mi/h)				
		55	60	65	70	75
Medium Rain	>0.10–0.25 in/h	0.94	0.93	0.92	0.91	0.90
Heavy Rain	>0.25 in/h	0.89	0.88	0.86	0.84	0.82
Light Snow	>0.00–0.05 in/h	0.97	0.96	0.96	0.95	0.95
Light-Medium Snow	0.05–0.10 in/h	0.95	0.94	0.92	0.90	0.88
Medium-Heavy Snow	>0.10–0.50 in/h	0.93	0.91	0.90	0.88	0.87
Heavy Snow	>0.50 in/h	0.80	0.78	0.76	0.74	0.72
Severe Cold	<–4°F	0.93	0.92	0.92	0.91	0.90
Low Visibility	0.50–0.99 mi	0.90	0.90	0.90	0.90	0.90
Very Low Visibility	0.25–0.49 mi	0.88	0.88	0.88	0.88	0.88
Minimal Visibility	<0.25 mi	0.90	0.90	0.90	0.90	0.90
Normal Weather	None of the Above	1.00	1.00	1.00	1.00	1.00

(Source: Reprinted with permission from *Highway Capacity Manual, 6th Edition: A Guide for Multimodal Mobility Analysis*, Transportation Research Board, National Academy of Sciences, Courtesy of the National Academies Press, Washington, D.C., 2016.)

Table 28.10: Full Alternative Text

Table 28.11: Speed Adjustment Factors (*SAF*) for Inclement Weather

Weather	Definition of Conditions	*SAF* for Unadjusted *FFS* of: (mi/h)				
		55	60	65	70	75
Medium Rain	>0.10 − 0.25 in/h	0.96	0.95	0.94	0.93	0.93
Heavy Rain	>0.25 in/h	0.94	0.93	0.93	0.92	0.91
Light Snow	>0.00 − 0.05 in/h	0.94	0.92	0.89	0.87	0.84
Light-Medium Snow	0.05 − 0.10 in/h	0.92	0.90	0.88	0.86	0.83
Medium-Heavy Snow	>0.10 − 0.50 in/h	0.90	0.88	0.86	0.84	0.82
Heavy Snow	>0.50 in/h	0.88	0.86	0.85	0.83	0.81
Severe Cold	< − 4°F	0.95	0.95	0.94	0.94	0.92
Low Visibility	0.50 − 0.99 mi	0.96	0.95	0.94	0.94	0.93
Very Low Visibility	0.25 − 0.49 mi	0.95	0.94	0.93	0.92	0.91
Minimal Visibility	<0.25 mi	0.95	0.94	0.93	0.92	0.91
Normal Weather	None of the Above	1.00	1.00	1.00	1.00	1.00

(Source: Reprinted with permission from *Highway Capacity Manual, 6th Edition: A Guide for Multimodal Mobility Analysis*, Transportation Research Board, National Academy of Sciences, Courtesy of the National Academies Press, Washington, D.C., 2016.)

Table 28.11: Full Alternative Text

CAF for Traffic Incidents

The CAF for the occurrence of traffic incidents is shown in Table 28.12. Values are based on an NCHRP study of default values [3].

Table 28.12: *CAF* for the Effect of Traffic Incidents

Number of Lanes	No Incidents	*CAF* for Incidents with the Following Effects				
		Shoulder Closed	1 Lane Closed	2 Lanes Closed	3 Lanes Closed	4 Lanes Closed
2	1.00	0.81	0.70	NA	NA	NA
3	1.00	0.83	0.74	0.51	NA	NA
4	1.00	0.85	0.77	0.50	0.52	NA
5	1.00	0.87	0.81	0.67	0.50	0.50
6	1.00	0.89	0.85	0.75	0.52	0.52
7	1.00	0.89	0.88	0.80	0.63	0.63
8	1.00	0.89	0.89	0.84	0.66	0.66

(Source: Reprinted with permission from *Highway Capacity Manual, 6th Edition: A Guide for Multimodal Mobility Analysis*, Transportation Research Board, National Academy of Sciences, Courtesy of the National Academies Press, Washington, D.C., 2016.)

Table 28.12: Full Alternative Text

It should be noted that the CAFs in Table 28.12 apply only to those lanes that remain open during the incident. For example, a four-lane freeway segment in one direction with two lanes closed due to a traffic incident, has only two lanes open. From Table 28.12, the remaining two lanes will carry only 0.50 of their normal capacity. The two lanes blocked by the incident can process no vehicles.

Adjustments for a Non-Standard

Driver Population

An adjustment for driver population has been in the HCM since 1985 in one form or another. Prior to the 2016 HCM, it was applied to demand volumes. The 2016 HCM applies adjustments to both FFS and capacity for this characteristic.

Standard analysis procedures assume a regular driver population that is familiar with the facility and its surrounding environment, that is, primarily commuters. An adjustment has always been available to account for situations in which weekend or recreational traffic might be the dominant demand problem. Table 28.13 shows the CAF and SAF for various driver populations.

Table 28.13: *CAF* and *SAF* for Non-Standard Driver Populations

Driver Familiarity Classification	*CAF*	*SAF*
All Familiar Drivers	1.0000	1.000
Mostly Familiar Drivers	0.968	0.975
Balanced Mix of Familiar and Unfamiliar Drivers	0.939	0.950
Mostly Unfamiliar Drivers	0.898	0.913
All Unfamiliar Drivers	0.852	0.863

(Source: Reprinted with permission from *Highway Capacity Manual, 6th Edition: A Guide for Multimodal Mobility Analysis*, Transportation Research Board, National Academy of Sciences, Courtesy of the National Academies Press, Washington, D.C., 2016.)

Table 28.13: Full Alternative Text

Adjustment factors for driver population were algebraically developed from the range of driver population adjustments to demand flows given in previous editions of the HCM. The definition of intermediate ranges is entirely the result of judgmental interpolation. It should also be noted that there is no substantial research data to support these specific values.

Adjustments for Work Zones

The approach to work zone adjustments is quite different from those for other categories of CAFs and SAFs. Instead, a complete methodology to predict the capacity and FFS in a work zone is provided. CAFs and SAFs are determined from the results of those analyses. The methodology is based upon a national study of work zone operations [4, 5].

Both the capacity and FFS of a work zone are based upon the lane closure severity index (LSCI), which is defined as follows:

$$LCSI=\frac{1}{OR\times N_o} \quad [28\text{-}5]$$

where:

$OR=\text{open ratio}, N_o/N, N=\text{number of lanes in normal operation(without worl}$

Table 28.14 shows the normal range of LCSI values.

Table 28.14: Values of the Lane Closure Severity Index

Lanes in Normal Operation (N)	Open Lanes Through Work Zone (N_o)	Open Ratio (OR)	$LCSI$
3	3	1.00	0.33
2	2	1.00	0.50
4	3	0.75	0.44
3	2	0.67	0.75
4	2	0.50	1.00
2	1	0.50	2.00
3	1	0.33	3.00
4	1	0.25	4.00

(Source: Reprinted with permission from *Highway Capacity Manual, 6th Edition: A Guide for Multimodal Mobility Analysis*, Transportation Research Board, National Academy of Sciences, Courtesy of the National Academies Press, Washington, D.C., 2016.)

Table 28.14: Full Alternative Text

Work zone capacity is based upon a prediction of the queue discharge rate from the work zone (QDR_{WZ}). Research found that it was extremely difficult to actually observe pre-breakdown capacity in a work zone. On the other hand, queue discharge when a breakdown was already in place was relatively easy to observe. The queue discharge rate is estimated as:

QDRWZ=2,093−(154 LCSI)−(194 fBR)−−(179 fAT)+(9 fLAT)−(59 fDN) [28-6]

where:

QDRWZ=queue discharge rate from work zone (pc/h/open lane),LCSI=lan
0 forconcrete or other hard barriers; 1 for cones or other soft barriers
),fAT=adjustment for area type (0 for urban, 1 for rural
),fLAT=adjustment for lateral clearance tobarrier (0–12 ft), andfDN=adjust

0 for day, 1 for night).

Once the queue discharge rate is estimated, the capacity of the work zone is estimated as:

cWZ=QDRWZ (100100−aWZ) [28-7]

where:

αWZ=percentage reduction from capacity toQDR, andcWZ=capacity of the

The value of *αWZ* should be locally calibrated. If no calibration is possible, a default value of 13.4% may be used. Essentially, the default value assumes that the *QDR* for a work zone is 13.4% less than its capacity.

Once the capacity of the work zone is estimated, the CAF_{WZ} is computed as:

CAFWZ=cWZc [28-8]

where *c* is the base capacity (pc/h/ln) for the freeway, based upon its *FFS*.

The FFS in a work zone is estimated as:

FFSWZ=9.95+(33.49 fSR)+(0.53 SLWZ)−(5.60 LCSI)−(3.84 fBR) −(1.71 fDN)−(8.7 TRD) [28-9]

where:

SLWZ=speed limit in the work zone (mi/h),fSR=speed limit ratio (SL/SLW

Again, once the FFS_{WZ} is estimated, the SAF_{WZ} is computed as:

SAFWZ=FFSWZFFS [28-10]

where the FFS is the free-flow speed of the freeway outside of the work zone.

Because of the approach taken, it is rare that a CAF_{WZ} or a SAF_{WZ} would be applied as such. Rather, the capacity and FFS of the work zone would be directly computed from the methodology described, with the work zone being treated as a separate segment for analysis.

Sample Problem 28-2: Work Zone Analysis

Consider the case of long-term work zone on a freeway. The freeway has eight lanes (four lanes in each direction), two of which are closed due to major maintenance operations. The freeway itself has a FFS of 70 mi/h outside of the work zone. The work zone is delineated by concrete barriers, which are located 0 feet from the edge of travel lanes. The work zone is in a generally rural area, and the critical period for operations is during the day. The speed limit through the work zone is 45 mi/h, while the speed limit for the freeway is 70 mi/h. The total ramp density on the freeway is 3 ramps/mile. For these conditions, estimate the capacity and the FFS of the work zone.

From Table 28.14, the LCSI for this case is 1.00 (four lanes, two open). Equations 28-6, 28-7, and 28-9 are now applied to estimate the capacity and FFS for the work zone. The queue discharge rate from the work zone is computed as:

$$QDR_{WZ}=2{,}093-(154\ LCSI)-(194\ f_{BR})+(9\ f_{LAT})-(179\ f_{AT})-(59\ f_{DN})$$

where:

$$LCSI=1.00\ (\text{Table 28.14}), f_{BR}=0\ (\text{concrete barrier}), f_{LAT}=0\ \text{ft (given)}, f_{AT}=$$

Then:

$$QDR_{WZ}=2{,}093-(154\times1)-(194\times0)+(9\times0)-$$
$$-(9\times0)-(179\times1)-(59\times0)=1{,}760\ \text{pc/h/open lane}$$

Equation 28-7 is then used to estimate the capacity of the work zone:

$$c_{WZ}=QDR_{WZ}\ (100100-a_{WZ})$$

Using the default value for a_{WZ} of 13.4%, the capacity of the work zone is:

$$c_{WZ}=1{,}760\times(100100-13.4)=2{,}032\ \text{pc/h/open lane}$$

Because there are two open lanes in the work zone, it will be able to handle a total of $2{,}032\times 2=4{,}064$ pc/h. Note that this capacity is still stated in pc/h, and that it would have to be converted to veh/h to account for the impact of trucks using the methodology discussed later in this chapter.

Equation 28-9 is used to estimate the FFS of the work zone:

$$FFS_{WZ}=9.95+(33.49\ f_{SR})+(0.53\ SL_{WZ})-(5.60\ LCSI)-(3.84\ f_{BR})-(1.71\ f_{DN})-(8.7\ TRD)$$

where:

$$f_{SR}=70/45=1.56\quad SL_{WZ}=45\text{ mi/h (given)},\quad LCSI=1.00\ \text{((Table 28.14)},\quad f_{BR}=0\ \text{f}$$

Then:

$$FFS_{WZ}=9.95+(33.49\times 1.56)+(0.53\times 45)-(5.60\times 1)-(3.84\times 0)-(1.71\times 0)-(8.7\times 3)=55.3\text{ mi/h}$$

Note that, in this case, the FFS_{WZ} is higher than the speed limit in the work zone.

A Final Word on CAFs and SAFs

As initially noted, most capacity and level of service analyses are conducted assuming “normal” conditions, that is, good weather, no incidents, no work zones, and a typical driver population with regular familiar users of a facility. The existence of these factors, however, allows analysts to consider the likely impacts of periodic or long-term disruptions to those normal conditions.

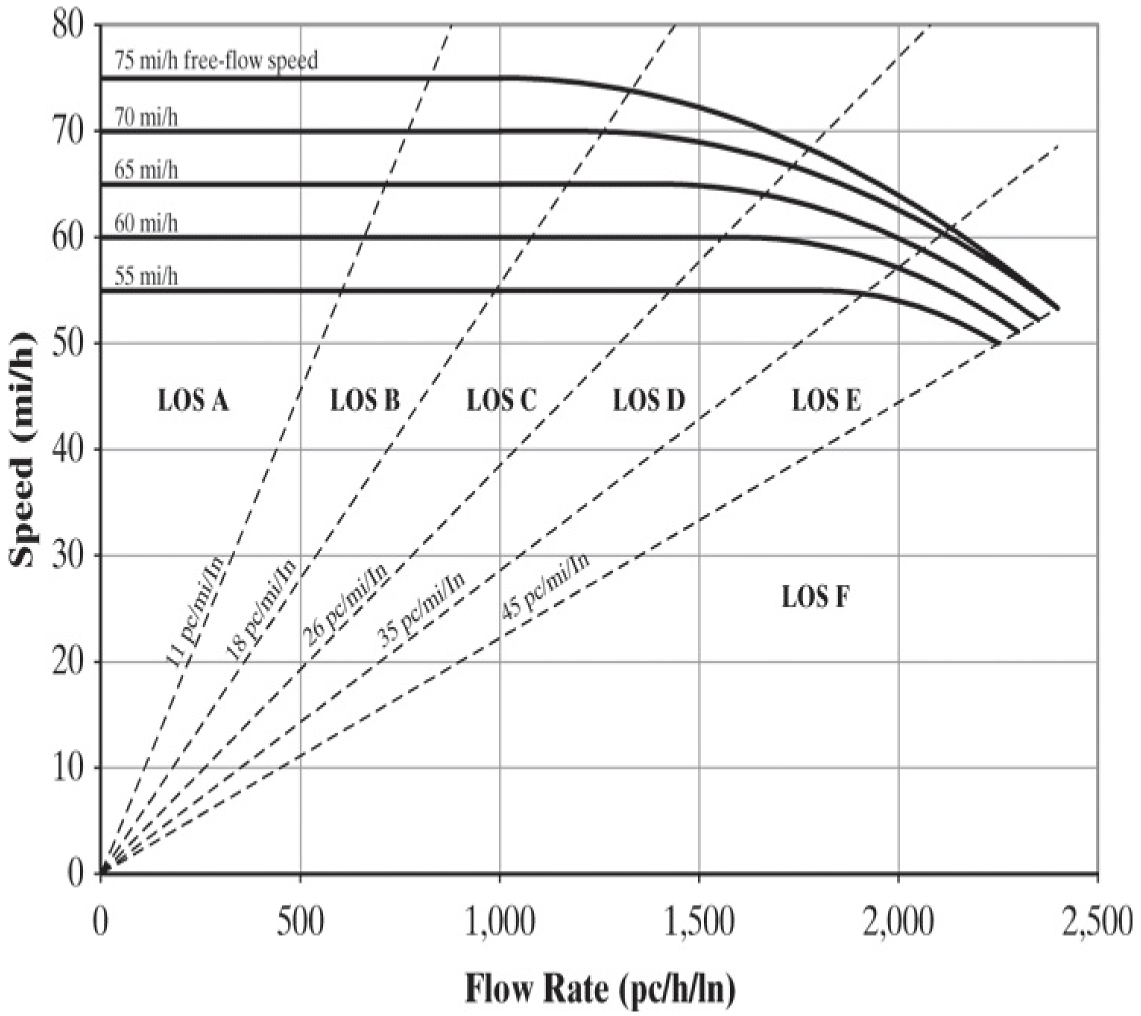
These factors can be applied to freeways, but they are not intended to be used with multilane highway methodologies. This was a judgment of the HCQSC, recognizing that virtually all of the research and data behind the factors came from freeways. Logically, one would expect to find similar impacts on multilane highways, but none have been calibrated to date.

Finally, these factors are presented here as part of the analysis methodology for basic freeway segments. As will be seen in subsequent chapters, they may also be applied to weaving, merging, and diverging segments on freeways.

28.5.4 Sample Curves for a Selection of Free-Flow Speeds

The 2016 HCM provides base curves for selected values of FFS: 55, 60, 65, 65, 70, and 75 mi/h for freeways, and 45, 50, 55, 60, 65, and 70 mi/h for multilane highways. These are shown in Figures 28.6 and 28.7.

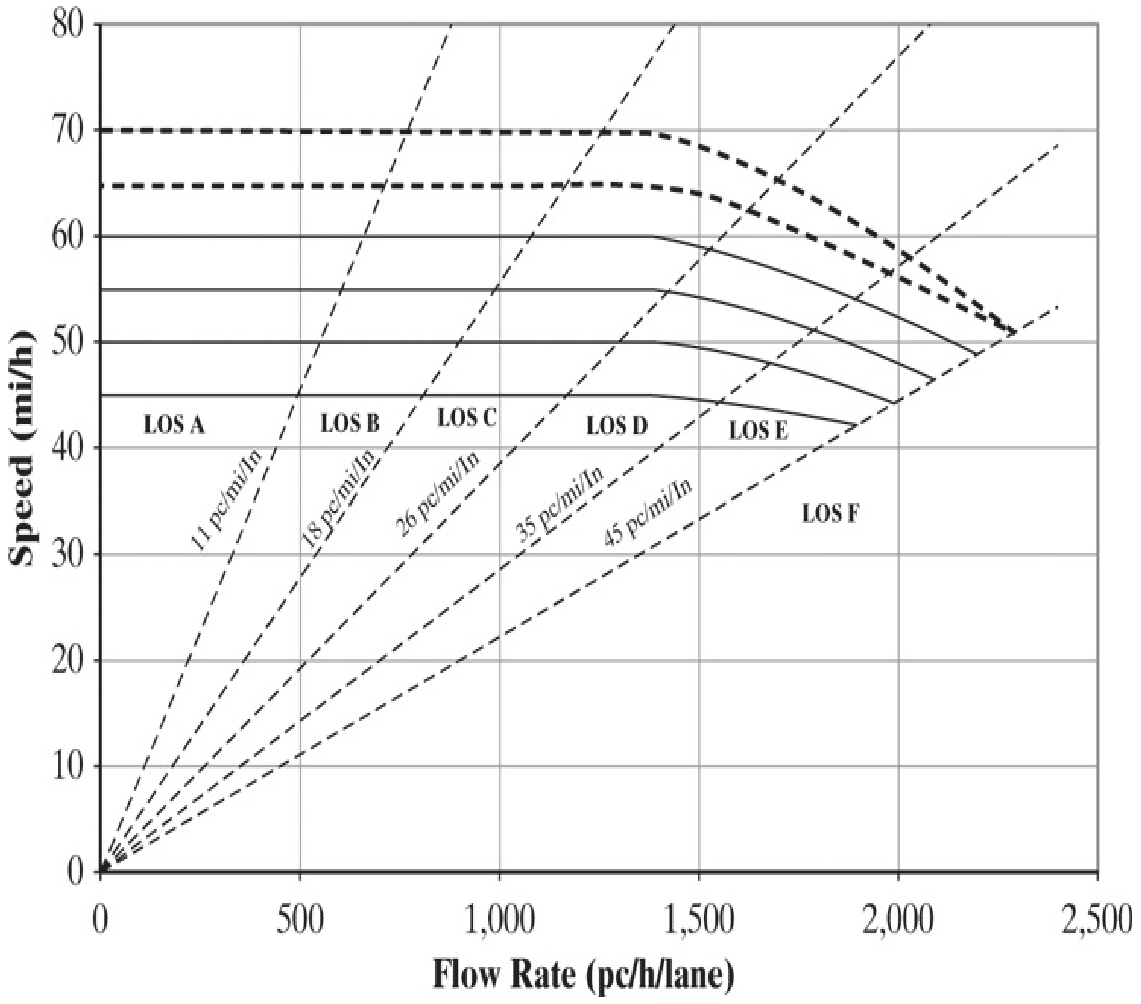
Figure 28.6: Example Base Speed-Flow Curves for Freeways



(Source: Reprinted with permission from *Highway Capacity Manual, 6th Edition: A Guide for Multimodal Mobility Analysis*, Transportation Research Board, National Academy of Sciences, Courtesy of the National Academies Press, Washington, D.C., 2016.)

Figure 28.6: Full Alternative Text

Figure 28.7: Example Base Speed-Flow Curves for Multilane Highways



(Source: ReprinSth permission from *Highway Capacity Manual, 6th Edition: A Guide for Multimodal Mobility Analysis*, Transportation Research Board, National Academy of Sciences, Courtesy of the National Academies Press, Washington, D.C., 2016.)

Figure 28.7: Full Alternative Text

These curves also show the levels of service. LOS boundaries are defined by density (pc/mi/ln), as indicated in Table 28.1. On a speed-flow plot, the density boundaries represent uniform slopes from the origin (remember that D=v/S). The point at which each LOS boundary crosses each speed-flow curve defines the maximum flow rate that can be accommodated with operating conditions consistent with the defined LOS. These are referred to as “maximum service flow rates,” and are shown in Tables 28.15 and 28.16 for freeways and multilane highways, respectively.

Table 28.15: Maximum Service Flow Rates (*MSF*) for Freeways (pc/h/ln)

FFS (mi/h)	Level of Service				
	A	B	C	D	E
75	820	1,310	1,750	2,110	2,400
70	770	1,250	1,690	2,080	2,400
65	710	1,170	1,630	2,030	2,350
60	660	1,080	1,560	2,010	2,300
55	600	990	1,430	1,900	2,250

Note: All values rounded to the nearest 10 pc/h/ln.

(Source: Reprinted with permission from *Highway Capacity Manual, 6th Edition: A Guide for Multimodal Mobility Analysis*, Transportation Research Board, National Academy of Sciences, Courtesy of the National Academies Press, Washington, D.C., 2016.)

Table 28.15: Full Alternative Text

Table 28.16: Maximum Service Flow Rates (*MSF*) for Multilane Highways (pc/h/ln)

FFS (mi/h)	Level of Service				
	A	B	C	D	E
70	760	1,260	1,700	2,020	2,300
65	710	1,170	1,630	2,000	2,250
60	660	1,080	1,550	1,980	2,200
55	600	990	1,430	1,850	2,100
50	550	900	1,300	1,710	2,000
45	290	810	1,170	1,550	1,900

Note:All values rounded to the nearest 10 pc/h/ln.

(Source: Reprinted with permission from *Highway Capacity Manual, 6th Edition: A Guide for Multimodal Mobility Analysis*, Transportation Research Board, National Academy of Sciences, Courtesy of the National Academies Press, Washington, D.C., 2016.)

Table 28.16: Full Alternative Text

Note that maximum service flow rates for any given LOS and FFS are *always* higher on freeways than on multilane highways. Historically, this has not always been the case in previous editions of the HCM. On a policy level, a suggestion that capacity (and by implication, operations) would improve by turning a freeway into a surface multilane highway, would be thoroughly illogical.

While these curves are presented for information and insight, the HCM methodology assumes that professionals will develop a specific speed-flow curve for the prevailing conditions at hand, working with an exact estimation of the FFS that applies.

28.6 Applications of Base Curves to Capacity and LOS Analysis of Freeways and Multilane Highways

In the previous section, the process for developing a speed-flow curve for some prevailing conditions for a specific freeway or multilane highway segment was described. Development of a base curve includes consideration of the following prevailing conditions:

- Lane widths
- Lateral clearance(s)
- Total ramp or access point density
- Type of median (multilane highways)
- Driver population

While the development of a base curve will address these conditions, there are two major remaining prevailing conditions that are *not* included in the base curves: heavy vehicle presence and peak hour factor (PHF). Base curves are developed with a flow rate scale in units of pc/h/ln. Prevailing conditions include demand volumes in veh/h that include heavy vehicles such as trucks, buses, and recreational vehicles. Their operating characteristics are not the same as passenger cars, and they can have a serious deleterious impact on both capacity and speed. The adjustment for heavy vehicles, however, is made on the demand side of the equation through the use of an adjustment factor (f_{HV}) that converts a traffic stream in pc/h to an equivalent one in veh/h (or vice-versa).

There are, however, three different types of applications within which this conversion will take place:

- Operational analysis

- Design analysis
- Service flow rate and service volume analysis

In addition to these, the HCM defines "planning analysis." This, however, consists of beginning the analysis with an AADT as a demand input, rather than a peak-hour volume. Planning analysis begins with a conversion of an AADT to a directional design-hour volume (DDHV) using the traditional procedure as described in Chapter 5.

28.6.1 Operational Analysis

The most common form of analysis is *operational analysis*. In this form of analysis, all traffic, roadway, and control conditions are defined for an existing or projected highway section, and the expected level of service and operating parameters are determined.

The basic approach is to convert the existing or forecast demand volumes to an equivalent flow rate under ideal conditions:

$v_p = \frac{V}{PHF \times f_{HV} \times N}$ [28-11]

where:

vp=flow rate per lane under equivalent idealconditions (pc/h/ln),V=demand
hour factor,fHV=adjustment factor for presence of heavyvehicles, andN=nu

The result is used in one of two ways. The computed value of v_p may be entered into the equation for the base speed-flow curve developed for the segment under study. It will be in the form of Equation 28-1, and will result in a prediction of average speed for the flow rate entered. Density may then be computed as $D=v_p/S$, and compared to the LOS criteria in Table 28.1. Alternatively, the base speed-flow curve developed can be entered graphically with a value of v_p to determine the average speed and LOS. Both techniques will be illustrated in the sample problems at the end of this chapter.

28.6.2 Design Analysis

In design analysis, an existing or forecast demand volume is used to determine the number of lanes needed to provide for a specified level of service. The number of lanes may be computed as:

$N_i = \frac{DDHV}{PHF \times MSF_i \times f_{HV}}$ [28-12]

where:

N_i=number of lanes (in one direction) requiredto provide for level of servic hour volume (veh/h),MSF_i=maximum service flow rate for LOS i fromTab

Equation 28-12, however, will almost always result in a fractional answer. If the equation indicates that 3.1 lanes are needed to provide LOS *i*, then four lanes will have to be provided. Because of this, it is often more convenient to compute the service flow rate and service volume for the desired level of service for a range of reasonable values of *N* (usually two, three, four, and possibly five lanes). Then the demand volume or flow rate can be compared to the results for a simpler determination of the required number of lanes.

28.6.3 Service Flow Rate and Service Volume Analysis

It is often useful to determine the service flow rates and service volumes for the various levels of service under prevailing conditions. Various demand levels may then be compared to these estimates for a speedy determination of expected level of service. The service flow rate for a given level of service is computed as:

$SF_i = MSF_i \times N \times f_{HV}$ [28-13]

where:

SF_i=service flow rate for level of service "i" (veh/h),MSF_i=maximum servi

The maximum service flow rates for each level of service, MSF_i, are taken from Table 28.15 (for freeways) and Table 28.16 (for multilane highways). The tables, however, only provide values for *FFS* in even 5 mi/h increments. If the segment under study has an intermediate *FFS*, a plot of the speed-flow curve for the segment, with LOS lines, must be constructed to obtain values.

Service flow rates are stated in terms of peak flows within the peak hour, usually for a 15-minute analysis period. It is often convenient to convert service flow rates to service volumes over the full peak hour. This is done using the peak-hour factor:

$SV_i = SF_i \times PHF$ [28-14]

where:

SV_i=service volume over a full peak hour for level of service "i" (veh/h).

All other variables as previously defined.

28.7 The Heavy Vehicle Adjustment Factor and Related Issues

The common feature of operational analysis, design analysis, and service volume analysis is the need to determine an adjustment factor that accounts for the presence of heavy vehicles in the traffic stream (f_{HV}). While in appearance, the adjustment is similar to those in previous editions of the HCM, the 2016 HCM is substantially different in its approach to heavy vehicles.

Previous HCMs have dealt with three different categories of heavy vehicles: trucks, buses, and recreational vehicles. The 2016 HCM deals with only two categories: single-unit truck (SUT) and tractor-trailer combinations (TT). All buses and recreational vehicles are now classified as SUTs. Further, the 2016 HCM does not treat the two truck categories separately, but rather looks at the traffic stream in terms of the *mix* of trucks that might be present.

28.7.1 Passenger Car Equivalents

The heavy vehicle adjustment factor (f_{HV}) is actually an algebraic manipulation of passenger-car equivalents (E_{HV}). The passenger car equivalent is a concept that proposes that for every heavy vehicle present in the traffic stream, E_{HV} passenger cars are displaced.

Consider the following situation: A traffic stream contains 15% (0.15) heavy vehicles, each of which displaces 3.0 (E_{HV}) passenger cars from the traffic stream. If the traffic stream volume is 3,000 veh/h, what is the volume stated as passenger car equivalents? The calculation is relatively simple. Of the 3,000 veh/h, 15% or 450 veh/h are heavy vehicles. The remainder, or 2,550 veh/h, are passenger cars. Each heavy vehicle is the equivalent of 3.0 passenger cars. Therefore, the equivalent passenger-car traffic stream is:

450×3.0=1,3502,550×1.0=2,550_ 3,900 pc/h

The prevailing traffic stream of 3,000 veh/h operates as if it were 3,900 pc/h.

The heavy vehicle adjustment factor is designed to convert a traffic stream in pc/h to one in veh/h under prevailing conditions, or vveh/h=vpc/h×fHV. The heavy vehicle adjustment factor, therefore, is calibrated as fHV=vveh/h/vpc/h. Calibration, however, requires that equivalent flow rates in veh/h and pc/h be defined. This requires extensive effort, and can be done in many different ways, depending upon the desired outcome.

The relationship between f_{HV} and E_{HV}, however, is relatively straightforward. The small example above can be used to illustrate it. By definition, the heavy vehicle adjustment factor is computed as:

fHV=vveh/hvpc/h [28-15]

For the example above, this yields fHV=3,000/3,900=0.76923. Using Equation 28-15 as a starting point, we can substitute an equation for the process used to compute *v* in pc/h. Then:

fHV=vveh/h[vveh/h×PHV×EHV]+[vveh/h×(1−PHV)×1]

where:

PHV=proportion of heavy vehicles, given as 0.15,andEHV=heavy vehicle
fHV=3,000[3,000×0.15×3.0]+
[3,000×(1−0.15)×1]=3,0001,350+2,550=0.76923

We could now use this factor to make the original computation, that is, knowing the flow rate in pc/h, we could now convert it back to veh/h, as:

vveh/h=3,900×0.76923=3,000 veh/h

The equation for f_{HV} could then be simplified algebraically to yield the general equation used throughout the HCM:

fHV=11+PHV (EHV−1) [28-16]

where all terms have been previously defined.

A number of different approaches have been used over the years to define the equivalence of a mixed traffic stream to one with only passenger cars:

- In the 1965 HCM, Powell Walker developed a methodology based on the relative numbers of passing maneuvers of passenger cars passing trucks, and passenger cars passing other passenger cars. The methodology was not based on observed passing behavior, but on the passing implications of measured speed distributions for trucks and passenger cars. The method was developed for two-lane rural highways, but was applied to multilane facilities as well. The calibration was never documented. Years later, Werner documented the methodology using more modern data from Canadian two-lane highways [6].

- A modification of the Walker method used relative delay caused by one class of vehicles to other vehicles in the traffic stream.

- For the 1985 HCM, Krammas and Crowley [7] calibrated equivalents for level terrain based on relative observed headways in the traffic stream for passenger cars following passenger cars, passenger cars following trucks, trucks following passenger cars, and trucks following trucks. The method was applied to equivalents on general terrain segments.

- For the 1985 HCM, Linzer, Roess, and McShane [8] used simulation outputs to determine equivalent v/c ratios in mixed and passenger car only streams, which were used to determine equivalents for trucks on specific grades.

- For 2000, Webster and Elefteriadou [9] used new simulations to develop revised passenger car equivalents, modifying earlier work based upon economic equivalence, that is, the cost of pavement generated by truck usage vs. passenger car usage.

These are just a few of the ideas that have been used to define equivalence and calibrate passenger car equivalence for trucks. Others include defining equivalence based upon the speed or density of the traffic stream (mixed flow versus passenger car only flow). There are flaws in all of these approaches. The concept of equivalence is not simple or straightforward, and it is virtually impossible to develop an approach that does not have some deficiencies.

For the 2016 HCM, an entirely new modeling approach and concept has been developed. Using the kinematic relationships governing truck and passenger car performance on grades and more gentle terrains, speed-flow curves for passenger car only traffic streams and mixed traffic streams can be developed and compared. For 2016, equivalents are based upon capacity, that is, use of the equivalents estimates a mixed flow capacity from a passenger car only capacity. Capacity is, of course, a single point on a speed-flow curve, and equivalents calibrated by comparing capacities are different from those that would arise from considering a range of points in the speed-flow continuum. This approach is documented in the 2016 HCM for the first time. At the time of publication of this text, research papers on the calibration of this approach have not yet been published.

Another difference between 2016 and previous methodologies is that previous HCMs focused on the speed-density behavior of the traffic stream at the *end* of a sustained grade (as this is where truck performance would be the worst); the new approach considers the space mean speed of trucks and passenger cars over the full length of the grade. The latter is obviously faster than the former.

The 2016 HCM presents passenger car equivalents for heavy vehicles under a variety of situations. A single heavy-vehicle equivalent is used, and tables have been prepared for traffic streams with various percentages of heavy vehicles, with a mix of trucks made up of:

- 70% SUT, 30% TT (default mix)
- 50% SUT, 50% TT (equal mix)
- 30% SUT, 70% TT (heavy mix)

Further, passenger car equivalents have been calibrated for:

- Extended segments of general terrain (level or rolling), and
- Specific grades.

The procedures presented are generally sufficient for the purposes of determining level of service. In some cases, where there are many trucks in conjunction with long and/or steep grades, the average speeds predicted by the methodology may be too high. For these cases, the 2016 HCM contains a detailed model for directly predicting average (space mean) speeds of the mixed traffic stream. This latter model is not included in this text. The 2016 HCM [1] should be consulted directly for these cases.

28.7.2 Passenger Car Equivalents for General Terrain Segments

Typically, a long section of roadway may be considered to be a general terrain segment if no one grade of 3% or greater is longer than 0.25 miles, and no one grade of less than 3% is longer than 0.50 miles. There are two categories of general terrain segment defined:

- Level terrain: Level terrain consists of short grades, generally less than 2% in severity. The combination of horizontal and vertical alignment permits trucks and other heavy vehicles to maintain the same speed as passenger cars in the traffic stream.

- Rolling terrain: Rolling terrain is any combination of horizontal and vertical alignment that causes trucks and other heavy vehicles to reduce their speeds substantially below those of passenger cars, but does not force heavy vehicles to operate at crawl speed for a significant distance. Crawl speed is defined as the minimum speed that a heavy vehicle can sustain on a specified grade, regardless of its length.

That category of *mountainous terrain*, present in previous editions of the HCM, has been eliminated. Any series of grades constituting mountainous terrain should be analyzed as a series of specific grades.

Passenger car equivalents for trucks do not depend upon the mix of trucks, and are shown in Table 28.17.

Table 28.17: Passenger Car

Equivalents for Heavy Vehicles in General Terrain Segments

Passenger Car Equivalent	Type of Terrain	
	Level	Rolling
E_{HV}	2.0	3.0

(Source: Reprinted with permission from *Highway Capacity Manual, 6th Edition: A Guide for Multimodal Mobility Analysis*, Transportation Research Board, National Academy of Sciences, Courtesy of the National Academies Press, Washington, D.C., 2016.)

Table 28.17: Full Alternative Text

28.7.3 Passenger Car Equivalents for Specific Grades

The 2016 HCM provides passenger car equivalents (E_{HV}) for heavy vehicles on specific grades. Table 28.18 shows equivalents for a typical mix of heavy vehicles in the traffic stream: 70% SUTs and 30% TTs. Where the exact mix of heavy vehicles is unknown, this would be used as the default mix. Remember that buses and recreational vehicles are included in the category of SUT. Research shows that the weight-to-horsepower ratio for SUT is an average of 65 lb/hp, while the ratio for TT is an average of 130 lb/hp.

Table 28.18: Passenger Car Equivalents for a Typical Mix

of Heavy Vehicles on Specific Grades

% Grade	Length (mi)	Percentage of Trucks (Including Buses & RVs)							
		2%	4%	5%	6%	8%	10%	15%	≥20%
≤0	ALL	2.39	2.18	2.12	2.07	2.01	1.96	1.89	1.85
2	0.125	2.67	2.32	2.23	2.17	2.08	2.03	1.95	1.89
	0.375	3.63	2.82	2.64	2.52	2.35	2.25	2.10	2.02
	0.625	4.12	3.08	2.85	2.69	2.49	2,36	2.18	2.08
	0.875	4.37	3.21	2.96	2.78	2.56	2.42	2.22	2.11
	1.250	4.53	3.29	3.02	2.84	2.60	2.45	2.24	2.13
	≥1.500	4.58	3.31	3.04	2.86	2.61	2.46	2.46	2.14
2.5	0.125	2.75	2.36	2.27	2.20	2.11	2.04	1.95	1.90
	0.375	4.01	3.02	2.80	2.65	2.46	2.33	2.15	2.06
	0.625	4.66	3.35	3.08	2.88	2.64	2.48	2.36	2.15
	0.875	4.99	3.52	3.21	3.00	2.73	2.56	2.32	2.19
	1.250	5.20	3.64	3.30	3.08	2.79	2.60	2.35	2.22
	≥1.500	5.26	3.67	3.33	3.10	2.80	2.62	2.36	2.23
3.5	0.125	2.93	2.45	2.34	2.26	2.16	2.09	1.98	1.92
	0.375	4.86	3.46	3.16	2.96	2.69	2.53	2.30	2.18
	0.625	5.88	3.99	3.59	3.32	2.98	2.76	2.46	2.31
	0.875	6.40	4.75	3.81	3.51	3.12	2.88	2.55	2.38
	1.250	6.74	5.15	3.96	3.63	3.21	2.96	2.60	2.42
	≥1.500	6.83	5.27	3.99	3.66	3.24	2.98	2.62	2.44
4.5	0.125	3.13	2.56	2.43	2.21	2.13	2.01	2.01	1.95
	0.375	5.88	3.99	3.59	2.98	2.76	2.46	2.46	2.31
	0.625	7.35	4.75	4.22	3.39	3.10	2.71	2.71	2.51
	0.875	8.11	5.15	4.54	3.60	3.27	2.83	2.83	2.61
	≥1.000	8.33	5.27	4.63	3.66	3.33	2.87	2.87	2.64
5.5	0.125	3.27	2.69	2.53	2.42	2.28	2.19	2.95	1.98
	0.375	7.09	4.62	4.11	3.76	3.31	3.04	2.66	2.47
	0.625	9.13	5.68	4.97	4.49	3.88	3.51	3.00	2.74
	0.875	10.21	6.24	5.43	4.88	4.18	3.76	3.18	2.89
	≥1.000	10.52	6.41	5.57	5.00	4.27	3.83	3.24	2.93
6	0.125	3.51	2.76	2.59	2.47	2.32	2.22	2.08	2.00
	0.375	7.78	4.98	4.40	4.01	3.51	3.20	2.78	2.56
	0.625	10.17	6.23	5.42	4.87	4.17	3.75	3.18	2.88
	0.875	11.43	6.88	5.95	5.32	4.53	4.04	3.39	3.06
	≥1.000	11.81	7.08	6.11	5.46	4.64	4.13	3.45	3.11

(Source: Reprinted with permission from *Highway Capacity Manual, 6th Edition: A Guide for Multimodal Mobility Analysis*, Transportation Research Board, National Academy of Sciences, Courtesy of the National Academies Press, Washington, D.C., 2016.)

Table 28.18: Full Alternative Text

Table 28.19 shows a sample of passenger car equivalents for truck mixes that include more TT than the typical mix. The alternative mixes include 50% SUT/50% TT, and 30% SUT/70% TT. The former is more likely to occur in urban areas, while the latter is more likely to occur in rural areas.

Table 28.19: Passenger Car Equivalents for Atypical Mixes of Heavy Vehicles on Specific Grades

% Grade	Length (mi)	30% SUT – 70% TT (Heavy Mix)				50% SUT – 50% TT (Equal Mix)			
		Percentage of Trucks (Including Buses & RVs)							
		2%	**5%**	**10%**	**15%**	**2%**	**5%**	**10%**	**15%**
≤0	ALL	2.62	2.30	2.12	2.04	2.67	2.31	2.11	2.02
2	0.125	2.62	2.30	2.12	2.04	2.67	2.31	2.11	2.02
	0.375	3.76	2.78	2.38	2.22	3.76	2.77	2.36	2.20
	0.625	4.47	3.08	2.54	2.34	4.32	3.01	2.49	2.29
	0.875	4.80	3.22	2.61	2.39	4.57	3.11	2.55	2.33
	1.250	5.00	3.30	2.66	2.42	4.71	3.17	2.58	2.36
	≥1.500	5.04	3.32	2.67	2.43	4.74	3.19	2.59	2.36
2.5	0.125	2.62	2.30	2.12	2.04	2.67	2.31	2.11	2.02
	0.375	4.11	2.93	2.46	2.28	4.10	2.92	2.44	2.26
	0.625	5.04	3.32	2.67	2.43	4.84	3.23	2.61	2.38
	0.875	5.48	3.51	2.77	2.50	5.17	3.37	2.69	2.43
	1.250	5.73	3.61	2.83	2.54	5.36	3.45	2.73	2.47
	≥1.500	5.80	3.64	2.84	2.55	5.40	3.47	2.74	2.47
3.5	0.125	2.62	2.30	2.12	2.04	2.67	2.31	2.11	2.02
	0.375	4.88	2.93	2.63	2.41	4.89	3.25	2.62	2.39
	0.625	6.34	3.32	2.97	2.64	6.05	3.75	2.89	2.58
	0.875	7.03	3.51	3.12	2.76	6.58	3.97	3.01	2.67
	1.250	7.44	3.61	3.22	2.82	6.88	4.10	3.09	2.72
	≥1.500	7.53	3.64	3.24	2.84	6.95	4.13	3.10	2.73
4.5	0.125	2.62	2.30	2.12	2.04	2.67	2.31	2.11	2.02
	0.375	5.80	3.64	2.84	2.55	5.83	3.65	2.84	2.55
	0.625	7.90	4.53	3.32	2.90	7.53	4.38	3.24	2.83
	0.875	8.91	4.96	3.56	3.07	8.32	4.72	3.42	2.97
	≥1.000	9.19	5.08	3.62	3.11	8.53	4.81	3.47	3.00
5.5	0.125	2.62	2.30	2.12	2.04	2.67	2.31	2.11	2.02
	0.375	6.87	4.10	3.09	2.73	6.97	4.14	3.11	2.74
	0.625	9.78	5.33	3.76	3.21	9.37	5.16	3.67	3.14
	0.875	11.20	5.94	4.09	3.45	10.49	5.65	3.93	3.34
	≥1.000	11.60	6.11	4.18	3.51	10.80	5.78	4.01	3.39
6	0.125	2.62	2.30	2.12	2.04	2.67	2.31	2.11	2.02
	0.375	7.48	4.36	3.23	2.73	7.64	4.43	3.26	2.85
	0.625	10.87	5.79	4.01	3.39	10.45	5.63	3.92	3.33
	0.875	12.54	6.51	4.40	3.67	11.78	6.20	4.24	3.56
	≥1.000	13.02	6.71	4.51	3.75	12.15	6.36	4.85	3.62

(Source: Reprinted with permission from Excerpts from *Highway Capacity Manual, 6th Edition: A Guide for Multimodal Mobility Analysis*, Transportation Research Board, National Academy of Sciences, Courtesy of the National Academies Press, Washington, D.C., 2016.)

Table 28.19: Full Alternative Text

Note that there are no separate tables for downgrades. All downgrades are evaluated using the passenger car equivalents for the grade category of ≤0%.

28.7.4 Composite Grades

The passenger car equivalents given in Tables 28.18 and 28.19 are based on a constant grade of known length. In most situations, however, highway alignment leads to composite grades, that is, a series of upgrades and/or downgrades of varying steepness. In such cases, an equivalent uniform grade must be used to enter Tables 28.18 or 28.19. A simple technique for finding the equivalent uniform grade is the *average grade technique*. The average grade is defined as:

$$GAV=\left(\frac{\text{Total Rise}}{\text{Total Length}}\right)\times 100 \quad [28\text{-}17]$$

where:

GAV=average grade (%),Total Rise=total elevation difference from thebeg

Sample Problem 28-3: Composite Grade Computation

Consider a composite grade of 3,000 ft of 4% grade followed by 1,000 ft of 2% grade. What is the composite grade?

Rise (4% grade)=0.04×3,000=120.0 ftRise (2% grade)=0.02×1,000=20.0 ft

Then:

$GAV=(1404,000)\times 100=3.5\%$

A passenger car equivalent for a 3.5% grade 4,000 feet long would be found in the tables provided.

The difficulty with composite grades is that this approach is an approximation. The operation of trucks and their effect on the traffic stream are dependent on the exact profile of the highway segment. The average grade technique is a reasonable approximation when the total length of the composite grade is 4,000 feet or less, OR when no single grade is steeper than 4%.

In previous editions of the HCM, an alternative approach for situations involving longer or steeper composite grades was provided. It was based on finding a constant percent grade that would lead to the same speed for trucks at the end of the grade. As the current model for mixed flow operation no longer depends on the speed of trucks at the end of the grade, this approach is no longer viable. For 2016, when longer or steeper composite grades exist, users must apply the mixed flow model directly, and not use passenger car equivalent values in the tables. The detailed mixed flow model is not treated in this text, but may be found in Chapters 25 and 26 of the 2016 HCM [1].

28.7.5 The Heavy Vehicle Adjustment Factor (f_{HV})

Once the appropriate passenger car equivalent (E_{HV}) is found, the heavy vehicle adjustment factor is computed using Equation 28-16, described previously.

28.8 Sample Problems

Sample Problem 28-4: Operational Analysis of an Older Urban Freeway

A segment of an old freeway in a large urban area has the following characteristics:

- Four lanes (two lanes in each direction)
- 10 ft lane widths
- Lateral obstructions at the roadside 0 feet from the pavement edge
- Total ramp density=5 ramps/mile
- Rolling terrain
- Trucks (typical mix)=2%
- Peak-hour factor=0.95

The current peak-hour demand in the peak direction is 3,000 veh/h. At what level of service is the freeway expected to operate during peak periods?

Solution

1. Step 1: *Estimate the free-flow speed of the freeway*

   The first step in any operational analysis is to determine the FFS for the roadway segment under study. As this is a freeway segment, the FFS is estimated using Equation

28-2:

$$FFS = 75.4 - f_{LW} - f_{LC} - 3.22\ TRD^{0.84}$$

where:

$f_{LW} = 6.6$ mi/h (Table 28.3, 10-ft lanes), $f_{LC} = 3.6$ mi/h (Table 28.4, 0-ft clearance, 2 lanes), $TRD = 5$ ramps/mi (given).

Then:

$$FFS = 75.4 - 6.6 - 3.6 - 3.22\ (5^{0.84}) = 52.8 \text{ mi/h}$$

2. Step 2: *Determine the speed-flow equation for the segment*

   The speed-flow equation is of the form indicated in Equation 28-1:

   $$S = FFS_{adj} \quad v_p \le BP \qquad S = FFS_{adj} - \left[\frac{\left(FFS_{adj} - \frac{c_{adj}}{45}\right)(v_p - BP)^a}{(c_{adj} - BP)^a} \right] \quad v_p > BP$$

   All of the parameters needed to define this curve are computed using Table 28.2. Note that the analysis does not indicate that issues of weather, incidents, driver population, or work zones exist. Thus, the CAF and the SAF may be taken to be 1.00. From Table 28.2:

   $$c = 2{,}200 + 10\ (FFS - 50) = 2{,}200 + 10\ (52.8 - 50) = 2{,}228 \text{ pc/h/ln}$$

   $$BP = 1{,}000 + 40\ (75 - FFS_{adj})\ CAF^2 = 1{,}000 + 40\ (75 - 52.8)\ 1^2$$

   Then:

   $$S = 52.8 \quad v_p \le 1{,}888 \qquad S = 52.8 - \left[\frac{\left(52.8 - \frac{2{,}228}{45}\right)(v_p - 1{,}888)^2}{(2{,}228 - 1{,}888)^2} \right] \quad v_p > 1{,}888$$

3. Step 3: *Convert the demand volume to a flow rate in passenger car equivalents*

The demand in veh/h must be converted to a flow rate in passenger car equivalents so that it may be used in the calibrated speed-flow equation. The conversion is accomplished using Equation 28-10:

vp=VPHF×N×fHV

where:

V=3,000 veh/h (given),PHF=0.95 (given),N=2 (4-lane freeway), andfHV=computed using Equation 28-16.fHV=11+PHV (EHV−1)

where:

PHV=0.02 (given), andEHV=3.0 (Table 28.17, rolling terr

Then:

fHV=11+0.02 (3−1)=0.962vp=3,0000.95×2×0.962=1,641

4. Step 4: *Find the speed and density of the traffic stream, and determine the level of service*

   The calibrated speed-flow equation of Step 2 is used to determine the speed of the traffic stream. Because the demand flow of 1,641 pc/h/ln<1,888 pc/h/ln, the expected speed of the traffic stream will be the FFS, or 52.8 mi/h.

   The density of the traffic stream is computed as:

   D=vpS=1,64152.8=31.1 pc/mi/ln

   From Table 28.1, this density results in operation at LOS D (26–35 pc/mi/ln).

   The results show that the freeway segment is expected to operate at LOS D, which would be generally acceptable in a large urban area. The demand volume of 3,000 veh/h for a two-lane segment of freeway is not a particularly high one. The poor operations experienced are largely because of the deficient geometrics (lane width, lateral clearance) of the freeway segment.

Sample Problem 28-5: Operational Analysis of a Multilane Highway Segment

A four-lane multilane highway segment with a full median carries a peak-hour volume of 2,600 veh/h in the heaviest direction. It has the following additional characteristics:

- 12 ft lanes
- 4 ft clearance at the roadside and in the median
- 10 access points/mile
- 10% trucks, standard mix
- Peak-hour factor=0.88

The segment in question is on a sustained 3.5% grade, 1.25 miles in length. The base FFS may be taken as 65 mi/h. What is the expected level of service on the upgrade, and on the downgrade?

1. Step 1: *Estimate the free-flow speed of the multilane highway*

   The FFS of a multilane highway is estimated using Equation 28-3:

   FFS=BFFS−fLW−fLC−fM−fA

   where:

   BFFS=65.0 mi/h (given),fLW=0.0 mi/h (Table 28.3, 12 ft l
   Table 28.4, 4+4=8 ft total lateral clearance
   ),fM=0.0 mi/h (Table 28.6, median), andfA=2.5 mi/h (

Table 28.7, 10 access points/mi).

Then:

$$FFS = 65.0 - 0.0 - 0.9 - 0.0 - 2.5 = 61.6 \text{ mi/h}$$

2. Step 2: *Determine the speed-flow equation for the segment*

   Once again, the speed-flow curve will be of the form of [Equation 28-1](#):

   $$S = FFS_{adj} \quad v_p \le BP$$
   $$S = FFS_{adj} - \left[\frac{\left(FFS_{adj} - \frac{c_{adj}}{45}\right)(v_p - BP)^a}{(c_{adj} - BP)^a}\right] \quad v_p > BP$$

   The key values for the equation are selected or computed from [Table 28.2](#). Note that for multilane highways, no CAFs or SAFs may be applied. Then:

   $$c = 1{,}900 + 20\,(FFS - 45) = 1{,}900 + 20\,(61.6 - 45) = 2{,}232 \text{ pc/h/ln}$$
   $$BP = 1{,}400 \text{ pc/h/ln}$$

   Then:

   $$S = 61.6 \quad v_p \le 1{,}400$$
   $$S = 61.6 - \left[\frac{\left(61.6 - \frac{2{,}232}{45}\right)(v_p - 1{,}400)^{1.31}}{(2{,}232 - 1{,}400)^{1.31}}\right] \quad v_p > 1{,}400$$

3. Step 3: *Convert the demand volume to a flow rate in passenger car equivalents*

   The conversion of the demand volume to a flow rate in passenger car equivalents is accomplished using [Equation 28-11](#):

   $$v_p = \frac{V}{PHF \times N \times f_{HV}}$$

   where:

   $$V = 2{,}600 \text{ veh/h (given)}, PHF = 0.88 \text{ (given)}, N = 2 \text{ lanes (one}$$
   $$f_{HV} = \frac{1}{1 + P_{HV}(E_{HV} - 1)}$$

   where:

$P_{HV}=0.10$ (given), and $E_{HV}=2.96$ (Table 28.18, 3.5% grade, 1.25 mi, 10% trucks).

Then:

$f_{HV}=\frac{1}{1+0.10\,(2.96-1)}=0.836$ $v_p=\frac{2,600}{0.88\times2\times0.836}=1,7$

4. Step 4: *Find the speed and density of the traffic stream, and determine the level of service*

   Using the equation calibrated for this segment, the value of v_p is entered to find the estimated average speed of the traffic speed. Note that since v_p is greater than the breakpoint of 1,400 pc/h/ln, the curvilinear portion of the equation is used:

   $$S=61.6-\left[\left(\frac{61.6-2,232}{45}\right)\frac{(1,767-1,400)^{1.31}}{(2,232-1,400)^{1.31}}\right]=57.5\text{ mi/h}=57.5\text{ mi/h}$$

   The density may now be computed as:

   $$D=\frac{v_p}{S}=\frac{1,767}{57.5}=30.7\text{ pc/mi/ln}$$

From Table 28.1, this is level of service D (26–35 pc/mi/ln). The segment, therefore, operates relatively poorly. At the same time, the speed and density of the traffic stream are well within the stable range. Whether or not this is acceptable depends upon the specifics of the location, such as the development environment, safety records, and user and resident views.

Sample Problem 28-6: A Design Application

A new freeway is being designed through a rural area. The DDHV has been forecasted to be 2,700 veh/h during the peak hour (one direction). The following conditions are expected to prevail:

- PHF=0.85

- Familiar users of the facility
- 15% trucks, heavy mix (30% SUT, 70% TT)
- TRD=0.50 ramps/mile

A long segment of the facility is on level terrain, but one 2-mile segment is on a sustained 4.5% grade. If the objective is to provide for LOS C operation during peak periods, with LOS D an absolute minimum, how many lanes must be provided?

The example calls for the determination of the number of required lanes on three distinct segments of the facility: (a) a long level terrain segment, (b) a 2-mile, 4.5% upgrade, and (c) a 2-mile 4.5% downgrade.

1. Step 1: *Estimate the free-flow speed of the freeway*

   The FFS is estimated using Equation 28-2. Because this is a design application, it is assumed that lane widths and lateral clearances will be standard, that is, 12 ft and 6 ft, respectively. There would then be no adjustments for these features.

   FFS=75.4−fLW−fLC
   −3.22 TRD0.84FFS=75.4−0.0−0.0−3.22 (0.50.84)=74.8 m

2. Step 2: *Determine the speed-flow curve for the segments*

   The speed-flow curve will once again be of the form of Equation 28-1. Table 28.2 is used to compute or determine the key parameters. Note that design is always conducted for good weather, no incidents, and no work zones. The problem statement also specifies that users are familiar with the segments. Thus, there are no CAFs or SAFs to be applied. Then:

   c=2,200+10 (FFS
   −50)=2,200+10×(74.8−50)=2,448 pc/h/ln>2,400 pc/h/ln
   [1,000+40 (75−FFSadj)] CAF2=
   [1,000+40(75−74.8)]×12=1,008 pc/h/lna=2.00

Then:

S=74.8vp≤1,008S=74.8 − [(74.8−2,40045) (vp −1,008)2.00(2,400−1,008)2.00]vp>1,008S=74.8−[21.5 (vp−1,008)21,937,664]vp>1,008

3. Step 3: *Determine the maximum service flow rate for LOS C and LOS D for the segment*

   From Table 28.1, LOS C has a maximum density of 26 pc/mi/ln, and LOS D has a maximum density of 35 pc/mi/ln. What is needed is the speed-flow point on the curve for the segment that produces these values of density. This is best done by plotting the curve previously calibrated, and constructing slopes from the origin representing 26 pc/mi/ln and 35 pc/mi/ln. The curve is best plotted using a spreadsheet, and is shown in Figure 28.8.

Figure 28.8: Graphic Solution for MSF_C and MSF_D for Sample Problem 28-3

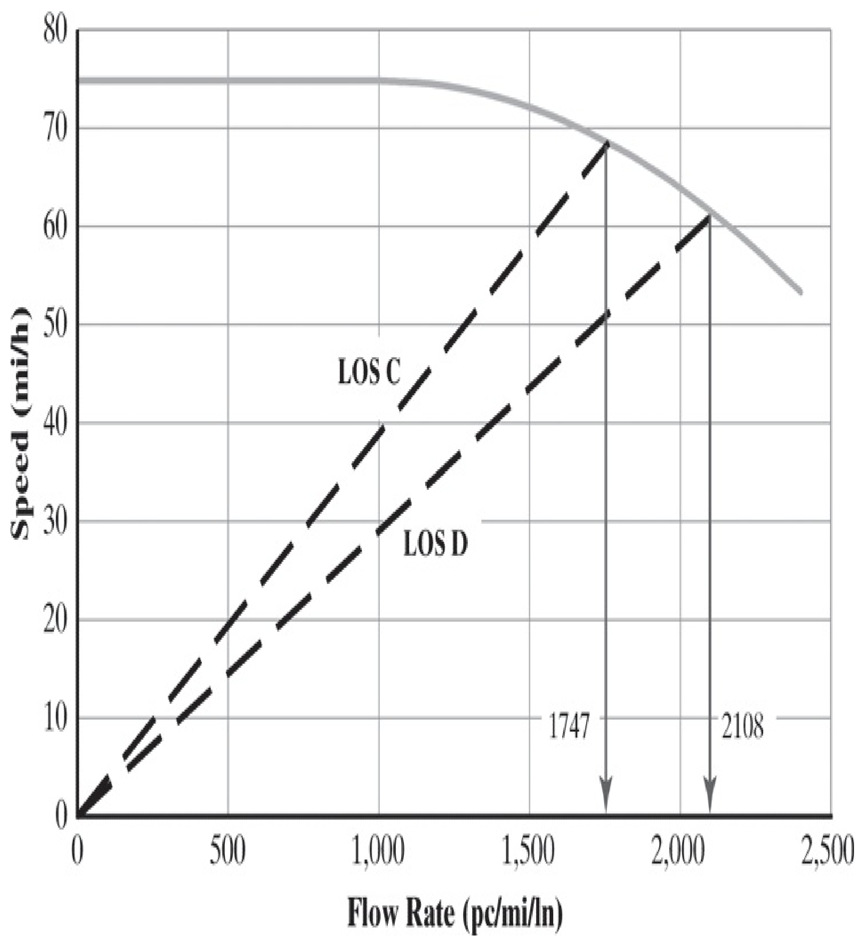


Figure 28.8: Full Alternative Text

From [Figure 28.8], the following values are determined:

- MSFC=1,630 pc/h/ln, and
- MSFD=2,120 pc/h/ln.

Approximate values could also have been obtained by interpolating in [Table 28.15]. The interpolation in this case is minor, given that the FFS is very close to 75 mi/h, which is included in the table. In this case, the scale of the figure makes interpolation a better option, and it was used to check the values determined from the figure.

4. Step 4: *Determine the heavy vehicle adjustment factor (f_{HV})*

   [Equation 28-12] is used to determine the number of lanes needed to deliver a target LOS. The equation, however, includes an adjustment for the PHF (which is given), and the heavy vehicle adjustment factor, which is not. In fact, because we are considering a level terrain segment, a 4.5% upgrade and a 4.5% downgrade, there may be three different values of this factor.

The heavy vehicle adjustment factor is computed using Equation 28-15:

fHV=11+PHV (EHV−1)

There are three different values for E_{HV}:

- EHV (level terrain)=2.0 (Table 28.7),
- EHV (upgrade)=3.11 (Table 28.9, 4.5%, 2 mi, 15% trucks, 70% TT-Heavy Mix)
- EHV (downgrade)=2.04 (Table 28.9, <0%, 2 mi, 15% trucks, 70% TT-heavy Mix)

Then:

fHV,Level=11+0.15 (2.0−1)=0.870fHV,upgrade=11+0.15

5. Step 5: *Find the number of lanes needed to provide LOS C or LOS D*

   Equation 28-12 is used to compute the required number of lanes. We have two target levels of service to consider, and three different segments:

   Ni=DDHVPHF×MSFi×fHV
   NC, level=27000.85×1747×0.870=2.09 lanesNC, upgrade

Note that the results are all fractional. Only full lanes can be built. To provide for LOS C, all three segments require that three lanes (in each direction) be provided. If we are willing to live with LOS D, all three segments could suffice with two lanes (in each direction).

There are obviously enormous economic consequences to the ultimate decision to build a six-lane freeway instead of a four-lane freeway.

Construction costs are higher, and a wider right-of-way would have to be purchased. Growth expectations would be a critical factor in the decision. If it is expected that significant growth in traffic will occur, it is likely that the six-lane option would be selected.

This problem illustrates some of the real-world considerations that enter into decision-making. The HCM analysis *does not* provide a final answer; rather, it shows what would be needed to provide for various levels of service. Costs, environmental impacts, social impacts, growth patterns and other considerations need to be included in sorting out the options available.

Sample Problem 28-7: Evaluating Growth

A six-lane urban freeway has a measured FFS of 65 mi/h, and generally rolling terrain. Traffic consists of 10% trucks (standard mix). Other standard conditions apply, so there are no CAFs or SAFs to be applied. The PHF on the facility is 0.92. Current traffic volume is 3,600 veh/h (in one direction), and it is expected to grow at a rate of 6% per year for the next 20 years.

What is the current LOS on the facility, and what levels can be expected in 5 years? 10 years? 15 years? 20 years?

General Approach

These questions could be answered by doing five separate operational analyses of the type illustrated in Sample Problems 28-1 and 28-2. That would be somewhat cumbersome. There is an easier approach: Determine the service flow rates (*SF*) and service volumes (*SV*) for the five stable levels of service, and compare the actual volumes in 0, 5,10, 15, and 20 years directly to the results to determine the LOS.

1. Step 1: *Determine the heavy vehicle adjustment factor*

The heavy vehicle adjustment factor (f_{HV}) must be used in all service flow rate and service volume computations. The heavy vehicle adjustment factor is computed using Equation 28-15:

fHV=11+PHV (EHV−1)

where:

PHV=0.10 (given), andEHV=3.0 (Table 28.17, rolling ter

Then:

fHV=11+0.10 (3−1)=0.833

2. Step 2: *Compute service flow rates for LOS A–E*

   Service flow rates are computed using Equation 28-13:

   SFi=MSFi×N×fHV

   where:

   N=3 (given), andfHV=0.833 (computed in Step 1).

   Maximum service flow rate for a 65-mi/h FFS are obtained from Table 28.15: LOS A=710 pc/h/ln, LOS B=1,170 pc/h/ln, LOS C=1,630 pc/h/ln, LOS D=2,030 pc/h/ln, LOS E=2,350 pc/h/ln. Then:

   SFA=710×3×0.833=1,774 veh/hSFB=1,170×3×0.833=2,9

3. Step 3: *Compute service volumes for LOS A–E*

   Because the current demand is stated as a full-hour volume, not a flow rate, it cannot be directly compared to service flow rates. Either the demand must be converted to a flow rate or each service flow rate must be converted to a service volume. Both do so by using the PHF. Service volumes are computed as:

   SVA=1,774×0.92=1,632 veh/hSVB=2,924×0.92=2,690 v

4. Step 4: *Determine the target year levels of service*

   The current volume is given as 3,600 veh/h. Comparing to the service volumes computed in Step 3, this would result in LOS C.

   Traffic is expected to grow at 6% per year. Thus, determining the LOS in 5, 10, 15, and 20 years requires that the expected demand volumes for those years be computed as:

   $V_N = V_0 \times 1.06^N V_5 = 3600 \times 1.06^5 = 4{,}818 \text{ veh/h} \ \text{(LOS E)} V_1$

   These results are a bit scary. This highway will fail by the 10th year, and most likely sooner. Expressed as a volume, the capacity (SV_E) for this facility is 5,403 veh/h. To find out when the facility fails exactly, leave the number of years unknown, and set the demand volume to 5,403 veh/h:

   $5{,}503 = 3600 \times 1.06^N \qquad N = 7.3 \text{ years}$

   Of course, operations will deteriorate badly well before this. The current LOS is C, but within five years, it has deteriorated to E. A plan to deal with the extreme problem of growth traffic on this facility must be started immediately. The full-hour capacity of 5,503 veh/h for three lanes suggests an hourly capacity of $5503/3 = 1{,}834$ veh/h/ln. With demand at 11,546 vehicles per full hour in 20 years, $11{,}546/1{,}834 = 6.3$ lanes would be needed in each direction! This is obviously not a practical solution. Innovative alternatives would have to be developed:

   - Can the growth rate be reduced by better zoning controls?
   - Are there public transportation options that could be considered?
   - Can improvements to alternate facilities be made to

handle some of the traffic?

  - Can work hours be adjusted to spread the demand over a greater number of hours of the day?

Once again, the HCM analysis provides us with very useful information. It does not solve the problem, however. We must use the insight yielded by the analysis to help develop reasonable solutions to a rapidly advancing traffic problem.

Sample Problem 28-8: Use of a CAF and SAF

How would the result to Sample Problem 28-1 change if we wished to focus on operations during a bad weather event: moderate (medium) rain? This would be something that would be done if the facility were located in an area where moderate rainfall is a common occurrence, thus justifying its consideration in evaluating operations and implementing designs.

1. Step 1: *Determine the applicable CAF and SAF*

   The CAF for moderate (medium) rain is obtained from Table 28.10. The SAF is obtained from Table 28.11. Both depend upon the base (unadjusted) FFS of the facility, which was estimated in Sample Problem 28-1 as 52.8 mi/h. Then:

   - $\text{CAF (medium rain)}=0.94$ (Table 28.10), and
   - $\text{SAF (medium rain)}=0.96$ (Table 28.11).

   The CAF and SAF are used to modify the capacity (*c*) and FFS estimated for the facility operating in good weather. Both were computed in Sample Problem 28-4. FFS, as noted, was found to be 52.8 mi/h. and the capacity (*c*) was found to be 2,228 pc/h/ln. These are now modified using the CAF and SAF:

   $$FFSadj=FFS\times SAF=52.8\times 0.96=50.7\text{ mi/h}cadj=c\times CAF=2,$$

2. Step 2: *Develop adjusted speed-flow equation for the segment*

   The base form of the equation remains that of Equation 28-1. However, the adjusted values of FFS_{adj} and c_{adj} must be used. In addition, the value of the breakpoint in the curve is also affected, as indicated in Table 28.2. Then:

   $$BP = 1,000 + 40\,(75 - FFS_{adj})\,CAF^2 = 1,000 + 40\,(75 - 50.7)\,0.$$

   Then, the speed-flow equation becomes:

   $$S = 50.7 \quad v_p \le 1,859$$
   $$S = 50.7 - \left[\left(50.7 - \frac{2,094}{45}\right) \frac{(v_p - 1,859)^{2.00}}{(2,094 - 1,859)^{2.00}} \right] \quad v_p > 1,859$$
   $$S = 50.7 - \left[\frac{4.2 \times (v_p - 1,859)^2}{55,225} \right] \quad v_p > 1,859$$

3. Step 3: *Determine the speed, density, and LOS of the traffic stream*

   The appropriate value of v_p was computed in Sample Problem 28-1 as 1,641 pc/h/ln. Placing this in the equation for the segment, it is seen that the speed will be the FFS_{adj} or 50.7 mi/h, as 1,641 pc/h/ln is less than the breakpoint of 1,859 pc/h/ln.

   The density may now be computed as:

   $$D = \frac{v_p}{S} = \frac{1641}{50.7} = 32.4 \text{ pc/mi/ln}$$

From Table 28.1, this is LOS D, which is the same as in the original problem. Thus, the existence of persistent moderate rainfall, the LOS of the facility is not meaningfully changed, although the speed is a bit lower, and the density a bit higher than when the weather is good.

28.9 Closing Comments

The 2016 HCM introduces some new approaches to the analysis of basic freeway segments and multilane highways. Instead of relying on a set of standard speed-flow curves, the new approach focuses on developing a specific speed-flow curve that applies to the segment under study.

For freeways, it provides additional flexibility to deal with adverse weather, traffic incidents, work zones, and nonstandard driver populations. This flexibility is achieved through the use of CAF and SAF. For freeways, these adjustments also carry through to methodologies for the analysis of weaving, merging, and diverging segments.

References

- 1. *Highway Capacity Manual, 6th Edition: A Guide for Multimodal Mobility Analysis*, Transportation Research Board, Washington, D.C., 2016.

- 2. Roess, R.P., "Speed-Flow Curves for the Highway Capacity Manual 2010," *Transportation Research Record 2257*, Transportation Research Board, Washington, D.C., 2011.

- 3. Zegeer, J., J. Bonneson, R. Dowling, P. Ryus, M. Vandehey, W. Kittelson, N. Rouphail, B. Schroeder, A. Hajbabaie, B. Aghdashi, T. Chase, S. Sajjadi, R. Margiotta, and L. Elefteriadou. *Incorporating Travel Time Reliability in the Highway Capacity Manual*. SHRP 2 Report S2-L08-RW-1. Transportation Research Board, Washington, D.C., 2014.

- 4. Yeom, C., A. Hajbabaie, B. Schroeder, C. Vaughan, X. Xuan, and N. Rouphail. *Innovative Work Zone Capacity Models from Nationwide Field and Archival Sources.* Presented at the 94th Annual Meeting of the Transportation Research Board, Washington, D.C., 2014.

- 5. Hajbabaie, A., C. Yeom, N. Rouphail, W. Rasdorf, and B. Schroeder. *Freeway Work Zone Free-Flow Speed Prediction from Multi-State Sensor Data.* Presented at the 94th Annual Meeting of the Transportation Research Board, Washington, D.C., 2014.

- 6. Werner, A., "Effect of Recreational Vehicles on Highway Capacity," *M.S. Thesis*, University of Calgary, Calgary, Canada, April 1974.

- 7. Krammas, R., and Crowley, K., "Passenger Car Equivalents for Trucks on Level Freeway Segments," *Transportation Research Record 1194*, Transportation Research Board, Washington, D.C., 1988.

- 8. Linzer, E., Roess, R., and McShane, W., "Effect of Trucks, Buses, and Recreational Vehicles on Freeway Capacity and Service Volume," *Transportation Research Record 699*, Transportation Research Board, Washington, D.C., 1979.

- 9. Elefteriadou, L., Torbic, D., and Webster, N., "Development of Passenger Car Equivalents for Freeways, Two-Lane Highways, and Arterials," *Transportation Research Record 1572*, Transportation Research Board, Washington, D.C., 1997.

Problems

1. 28-1 Estimate the free-flow speed of a four-lane undivided multilane highway having the following characteristics:
    - Base free-flow speed=60 mi/h
    - Average lane width=11 ft
    - Lateral clearances=3 ft at both roadsides
    - Access point density=15/mi on each side of the roadway
    - Good weather, no incidents, no work zones, and regular users of the facility may be assumed.
2. 28-2 Estimate the free-flow speed of a six-lane suburban freeway with 12 ft lanes, a right-side lateral clearance of 2 ft, and a ramp density of 3.5 ramps/mi. Normal conditions, that is, good weather, no incidents, no work zones, and regular users of the facility, may be assumed.
3. 28-3 What average grade would be used to analyze a segment with 1,000 ft of 2% upgrade followed by 1,500 ft of 3% upgrade?
4. 28-4 A freeway operating in generally rolling terrain has a traffic composition of 15% trucks (standard mix). If the observed peak hour demand is 3,200 veh/h (in one direction), what is the equivalent volume in passenger car equivalents?
5. 28-5 Find the upgrade and downgrade service flow rates and service

volumes for an eight-lane freeway with the following characteristics:

- 11 ft lanes
- 2 ft right-side lateral clearance
- 4.2 ramps/mi
- 3% trucks (standard mix)
- Good weather, no incidents, no work zones, and regular drivers
- Peak-hour factor=0.92

The segment in question is on a sustained grade of 3.5%, 1.5 miles in length.

6. 28-6 An existing six-lane divided multilane highway with a field-measured free-flow speed of 45 mi/h serves a peak-hour volume of 4,000 veh/h, with 10% trucks (50% SUT, 50% TT). The PHF is 0.88. The highway has generally rolling terrain. What is the likely level of service for this segment? Good weather, no incidents, no work zones, and regular drivers may be assumed.

7. 28-7 A long segment of suburban freeway is to be designed on level terrain. The level segment, however, is followed by a 4.5% grade, 2.0 miles in length. If the DDHV is 2,500 veh/h with 15% trucks (standard mix), how many lanes will be needed on the (a) upgrade, (b) downgrade, (c) level terrain segment to provide for level of service C? Lane widths and lateral clearances may be assumed to be 12 ft and 6 ft, respectively. Ramp density is expected to be 1.0 ramps/mi. The PHF is 0.92. Good weather, no incidents, no work zones, and regular users of the facility may be assumed.

8. 28-8 An older urban four-lane freeway has the following characteristics:

- 11 ft lanes
- No lateral clearance (0 ft)
- 4.5 ramps/mi
- 5% trucks (standard mix)
- Peak-hour factor=0.90
- Rolling terrain

The present peak-hour demand on the facility is 2,100 veh/h, and anticipated growth is expected to be 3% per year. What is the present LOS? What is the expected LOS in 5 years? 10 years? 20 years? To avoid breakdown (LOS F), when will substantial improvements be needed to this facility, and/or alternative routes created? Good weather, no incidents, no work zones, and regular drivers may be assumed.

9. 28-9 A six-lane recreational freeway has the following characteristics:
    - 12 ft lanes
    - 6 ft lateral clearances
    - 2 ramps/mi
    - 10% trucks (standard mix)
    - Peak-hour factor=0.95
    - Level terrain
    - Peak-hour demand=4,000 veh/h

    Because the facility serves recreational users to a nearby ski area, the peak period occurs on weekends, and light to moderate snow is the normal weather pattern. If virtually all users of the facility are unfamiliar with the roadway, what is the expected LOS during the peak period?

10. 28-10 A work zone on a rural six-lane freeway has one lane closed for an extended period of time. The work zone is protected by concrete barriers, which are set immediately at the edge of the travel lanes (lateral clearance=0 ft). The speed limit on the freeway is 70 mi/h, but through the work zone it is 50 mi/h. The total ramp density in the area is 1 ramp/mile, and the critical period of interest is during the daytime. What is the expected FFS and capacity of this work zone?

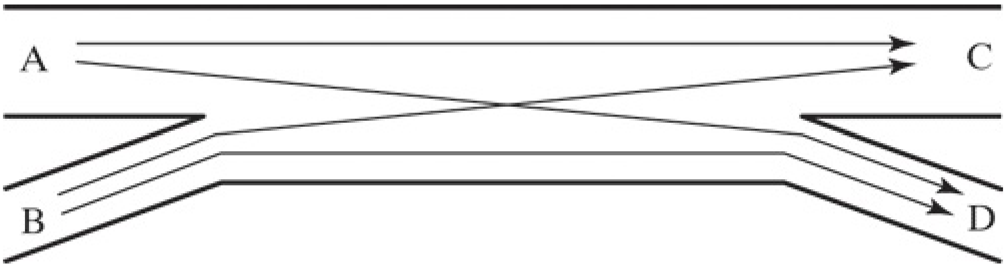
Chapter 29 Capacity and Level of Service Analysis: Weaving Segments on Freeways and Multilane Highways

In Chapter 28, capacity and level-of-service analysis approaches for basic freeway and multilane highway sections were presented and illustrated. Segments of such facilities that accommodate weaving, merging, and/or diverging maneuvers, however, experience additional turbulence as a result of these movements. This additional turbulence in the traffic stream results in operations that cannot be simply analyzed using basic segment techniques.

While there are no generally accepted measures of “turbulence” in the traffic stream, the most distinguishing characteristic of weaving, merging, and diverging segments is the additional lane-changing these maneuvers cause. Other elements of turbulence include the need for greater vigilance on the part of drivers, more frequent changes in speed, and average speeds that may be somewhat lower than on similar basic sections.

Figure 29.1 illustrates the basic maneuvers involved in weaving segments. *Weaving* occurs when one movement must cross the path of another along a length of facility without the aid of signals or other control devices, with the exception of guide and/or warning signs. Such situations are created when a merge area is closely followed by a diverge area. The flow entering on the left leg of the merge and leaving on the right leg of the diverge must cross the path of the flow entering on the right leg of the merge and leaving on the left leg at the diverge point. Depending upon the specific geometry of the segment, these maneuvers may require lane changes to be successfully completed. Further, other vehicles in the segment (i.e., those that do not weave from one side of the roadway to the other) may make additional lane changes to avoid concentrated areas of turbulence within the segment.

Figure 29.1: Movements in a Weaving Segment



(Source: Reprinted with permission from *Highway Capacity Manual, 6th Edition: A Guide for Multimodal Mobility Analysis*, Transportation Research Board, National Academy of Sciences, Courtesy of the National Academies Press, Washington, D.C., 2016.)

Figure 29.1: Full Alternative Text

Movements A-D and B-C must cross each other's path between the entry and exit gore areas. These are referred to as *weaving movements*. Movements A-C and B-D do not cross each other, but may be involved in additional lane-changing to avoid turbulence. These are referred to as *non-weaving* or *outer movements.*

The difference between weaving and separate merging and diverging movements is unclear at best. Weaving occurs when a merge segment is "closely followed" by a diverge segment. The exact meaning of "closely followed" is not well defined. The *Highway Capacity Manual* (HCM) 2000 [1] employed a uniform 2,500 ft length as the maximum for weaving operations; recent research [2] indicates that this length is variable. At some point, however, the merge and diverge ends of the weaving segment are far enough apart to operate independently. In such cases, the merge and diverge segments are treated separately (see Chapter 30), and the segment between them is treated as a basic freeway or multilane highway segment.

Even where the distance between a merge and diverge is less than the maximum, the classification of the movement depends upon the details of the configuration. For example, a one-lane, right-hand, on-ramp followed by a one-lane, right-hand, off-ramp is considered a weaving section only if the two are connected by a continuous auxiliary lane. If the on-ramp and off-ramp have separate, discontinuous acceleration and deceleration lanes, they are treated as isolated merge and diverge areas, respectively, independent of the distance between them. The 1965 HCM [3] recognized weaving movements over distances up to 8,000 feet, but this was based on a single data point, and lengths greater than 2,500 feet were subsequently removed from consideration as weaving areas.

Even though the nature of lane-changing and other turbulence factors is similar in weaving, merging, and diverging segments, the methodologies for analysis of weaving segments are different from those for merging and diverging segments, and are treated in separate chapters of this text. This is primarily an accident of research history, as conceptually, similar characteristics occur in all cases. Research efforts on these subjects have been done at different times using different data bases, as mandated by sponsoring agencies. Beginning with the 2010 HCM, considerable effort was invested to make the approaches more consistent, particularly in terms of level-of-service measures and criteria.

HCM methodologies for weaving segments are calibrated for freeways. These methods can be applied, with caution, to multilane highways with uncontrolled weaving operations, but must generally be considered more approximate in these cases.

29.1 Level of Service Criteria for Weaving Segments

The measure of effectiveness for weaving segments is density. This is consistent with freeway and multilane highway methodologies. Level of service criteria are shown in Table 29.1. Note that different criteria are specified for weaving segments on freeways and on multilane highways. Boundary conditions for multilane highways are set at somewhat higher densities than for freeways, reflecting users' lower expectations on multilane highways. This is somewhat inconsistent with the criteria for basic sections, which are the same for freeways and multilane highways.

Table 29.1: Level of Service Criteria for Weaving Segment

Level of Service	Density Range (pc/mi/ln) for:	
	Freeways	Multilane Highways
A	0–10	0–12
B	>10−20	>12–24
C	>20−28	>24−32
D	>28−35	>32−36
E	>35−43	>36−40
F	>43 or demand exceeds capacity	>40 or demand exceeds capacity

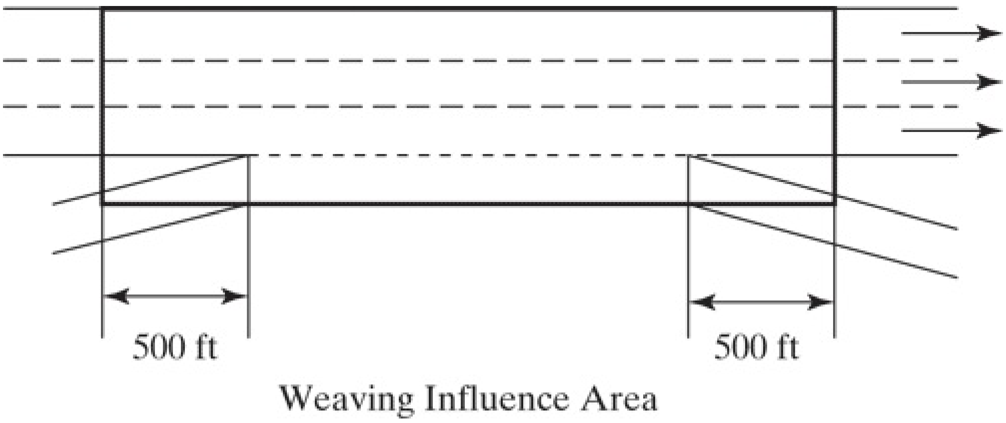
(Source: Reprinted with permission from *Highway Capacity Manual, 6th Edition: A Guide for Multimodal Mobility Analysis*, Transportation Research Board, National Academy of Sciences, Courtesy of the National Academies Press, Washington, D.C., 2016.)

Table 29.1: Full Alternative Text

For weaving segments, LOS F occurs under two conditions: (a) when demand exceeds capacity of the segment, that is, when v/c>1.00, or (b) when the density exceeds the maximum value for freeways or multilane highways, as appropriate. It is possible, in some cases, therefore, to have a situation described as LOS F where the *v/c* ratio is less than 1.00. This should, however, not occur frequently.

The level of service applies over the influence area of the weaving segment, which is defined as the basic length from merge point to diverge point, plus 500 ft upstream and downstream, as illustrated in Figure 29.2.

Figure 29.2: Weaving Influence Area Illustrated



(Source: Reprinted with permission from Transportation Research Board, National Research Council, modified from Highway Capacity Manual, National Academy of Sciences, Courtesy of the National Academies Press, Washington, D.C., 2000.)

Figure 29.2: Full Alternative Text

29.2 Converting Demand Volumes to Flow Rates in pc/h

Procedures for weaving analysis rely on algorithms calibrated in terms of demand flow rates in passenger car units for base or ideal conditions. Thus, component demand volumes must be converted to flow rates in pc/h before proceeding to use the methodology.

The conversion is accomplished using Equation 29-1:

$$v_i = \frac{V_i}{PHF \times f_{HV}} \quad [29\text{-}1]$$

where:

v_i = flow rate for movement i, pc/h, V_i = full-hour volume for movement i, veh/h, PHF = peak hour factor, and f_{HV} = heavy

The heavy vehicle factor is the same one used for basic freeway and multilane highway segments. It is found using the methods and values presented in Chapter 28.

29.3 A Brief History of the Development of Weaving Segment Methodologies

Weaving areas have been the subject of a great deal of research since the late 1960s, yet many features of current procedures continue to rely, at least partially, on judgment. This is primarily due to the great difficulty and cost of collecting comprehensive data on weaving operations. Weaving areas cover significant lengths and generally require videotaping from elevated vantage points, often using aircraft, or time-linked separate observation of entry and exit terminals and visual matching of vehicles. Further, there are a large number of variables affecting weaving operations, and, therefore, a large number of sites reflecting these variables would be needed to provide a statistically desirable database.

The first research study leading up to the third edition 1985 HCM focused on weaving areas [4]. This was unfortunate, as basic section models would be revised later, causing judgmental modification in weaving models for consistency. It relied on 48 sets of data collected by the then Bureau of Public Roads in the late 1960s and an additional 12 sets collected specifically for the study. The methodology that resulted was complex and iterative. It was later modified as part of a study of all freeway-related methodologies [5] in the late 1970s. In 1980, a set of interim analysis procedures was published by the Transportation Research Board [6], which included the modified weaving analysis procedure. It also contained an independently developed methodology that often produced substantially different results. The latter methodology was documented in a subsequent study [7]. To resolve the differences between these two methodologies, another study was conducted in the early 1980s, using a new data base consisting of 10 sites [8]. This study produced yet a third methodology, substantially different from the first two. As the publication date of the 1985 HCM approached, the three methodologies were judgmentally merged, using the ten 1980s sites for general validation purposes [9]. A number of studies throughout the 1980s and 1990s continued to examine the various weaving approaches, with no common consensus emerging [10–13].

It was, therefore, no surprise that a new study, relying on some new data but primarily on simulation, was commissioned as part of the research for the 2000 HCM [14]. Unfortunately, the simulation approach was not particularly successful, and it yielded a number of trends that were judged (by the Highway Capacity and Quality of Service Committee of the Transportation Research Board) to be counterintuitive. The method of the 2000 HCM resulted from a further judgmental modification of earlier procedures [15].

The weaving segment analysis methodology presented in this text resulted from National Cooperative Highway Research Program Project 3-75, *Analysis of Freeway Weaving Sections* [2]. This procedure was developed for inclusion in the 2010 HCM, and was formally approved by the HCQSC at its 2009 Summer Meeting. With minor changes, the methodology remains in place in the 2016 HCM.

29.4 Component Flows in a Weaving Area

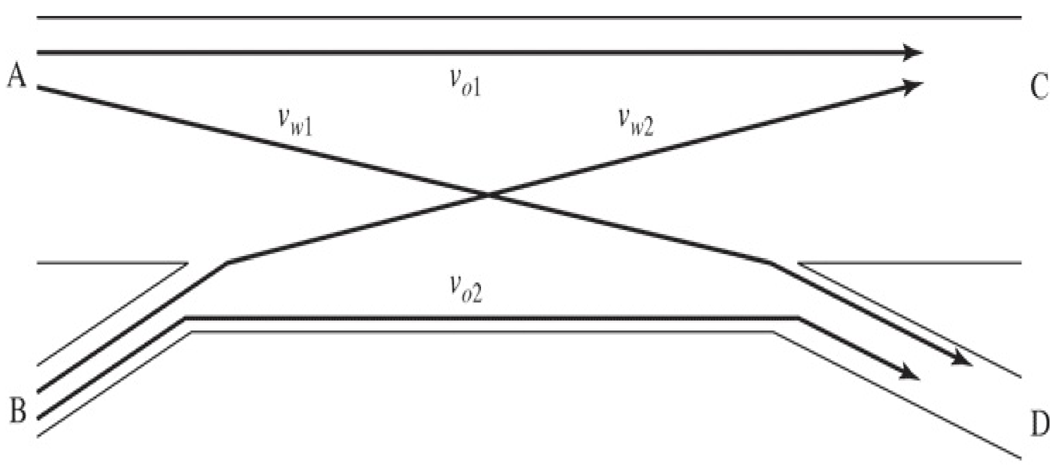
In a typical weaving area, there are four component flows that may exist. By definition, the two that cross each other's path are called *weaving flows,* while those that do not are called *nonweaving or outer flows,* as indicated in Figure 29.1 previously.

To simplify the description of flows in a weaving segment, a simple notation has been developed over the years: weaving flows use the subscript "w," while outer or nonweaving flows use the subscript "o." The larger of the two outer or weaving flows is given a second subscript "1," while the smaller uses the subscript "2." Thus:

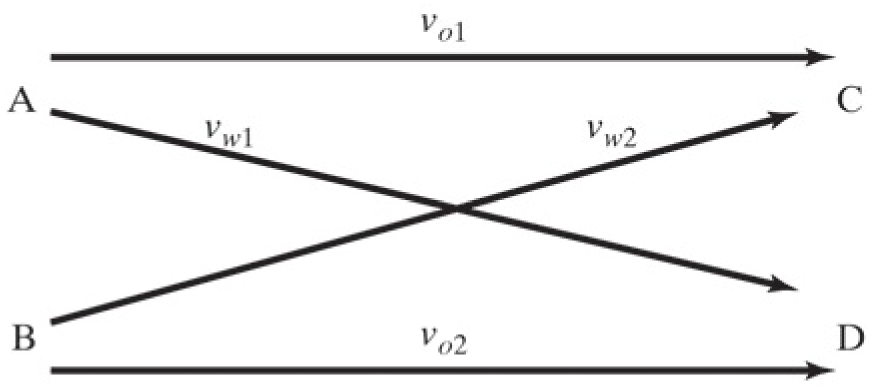
- vo1=the larger outer flow, pc/h, equivalentbase conditions,
- vo2=the smaller outer flow, pc/h, equivalentbase conditions,
- vw1=the larger weaving flow, pc/h, equivalentbase conditions, and
- vw2=the smaller weaving flow, pc/h, equivalentbase conditions.

Figure 29.3 (a) shows an example of how these designations are used, and also illustrates a weaving diagram. Figure 29.3 (b) is called the *weaving diagram*. In block form, it shows the weaving and nonweaving flows and their relative positions on the roadway. By convention, it is always drawn with traffic moving from left to right. It is a convenient form to illustrate the component flows in a consistent way for analysis. Other critical variables, used in analysis algorithms, may be computed from the base variables described:

Figure 29.3: Component Flows in a Weaving Segment and the Weaving Diagram



29.4-2 Full Alternative Text



29.4-2 Full Alternative Text

- vW=total weaving flow rate, pc/h=vw1+vw2
- vNW=total non-weaving flow rate, pc/h=vo1+vo2
- v=total flow rate in weaving segment, pc/h=vW+vNW
- VR=volume ratio=vW/v
- R=weaving ratio=vw2/vW

29.5 Critical Geometric Variables Describing a Weaving Segment

Three geometric variables have a significant effect on the quality of weaving operations:

- Lane configuration
- Length of the weaving area, ft
- Width (number of lanes) in the weaving area

Each of these has an impact on the amount of lane-changing that must or may occur, and the intensity of that lane-changing.

29.5.1 Lane Configuration

Lane configuration refers to the manner in which entry and exit legs “connect” with each other. This is a critical characteristic, as it ultimately determines how many lane changes *must* be made by weaving vehicles to successfully complete their weaving maneuver. These are mandatory lane changes, as they must be made for the weaving vehicle to get from its entry leg to its desired exit leg. By definition, these lane changes must be made within the weaving section.

There are many lane configurations that may exist based upon the number and location of entry and exit lanes, and the number of lanes within the weaving segment. Weaving segments are categorized in two ways:

- One-sided versus two-sided weaving segments
- Ramp-weave versus major weaving segments

In a one-sided weaving segment, weaving movements are substantially restricted to lanes on one side of the facility, usually (but not always) the right side. In two-sided weaving sections, at least one of the weaving movements must use lanes on both sides of the facility. Weaving turbulence in one-sided segments is more localized, whereas in two-sided segments, it may spread across most or all lanes of the facility. In more specific terms, the following definitions apply:

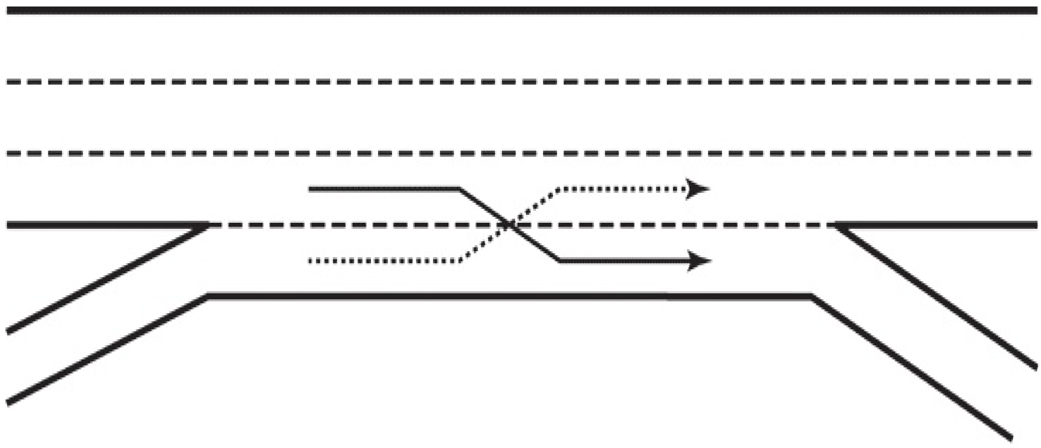
- A *one-sided weaving segment* is one in which no weaving maneuver requires more than two lane changes.

- A *two-sided weaving segment* is one in which one weaving maneuver requires three or more lane changes, or one in which a one-lane on-ramp on one side of the facility is closely followed by a one-lane off-ramp on the other side of the facility.

The ramp-weave segment is very common, and has a standard characteristic: a one-lane on-ramp is followed by a one-lane off-ramp (on the *same* side of the facility) and are connected by a continuous auxiliary lane. In major weaving segments, at least three of the entry and exit legs have more than one lane. In ramp-weaves, ramp roadways generally have design speeds that are lower, sometimes significantly, than that of the main facility. Because of this, on-ramp and off-ramp vehicles are most often accelerating or decelerating as they traverse the weaving segment.

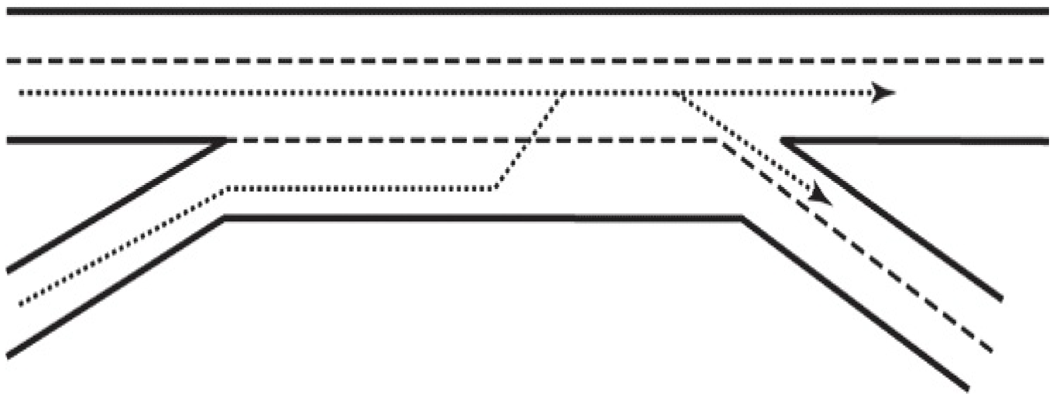
In major weave segments, entry and exit legs are often designed to standards that are closer to those of the main facility. Consequently, there is less acceleration and deceleration within the segment than for ramp-weaves. Figure 29.4 illustrates some of these characteristics.

Figure 29.4: Weaving Segment Configurations Illustrated



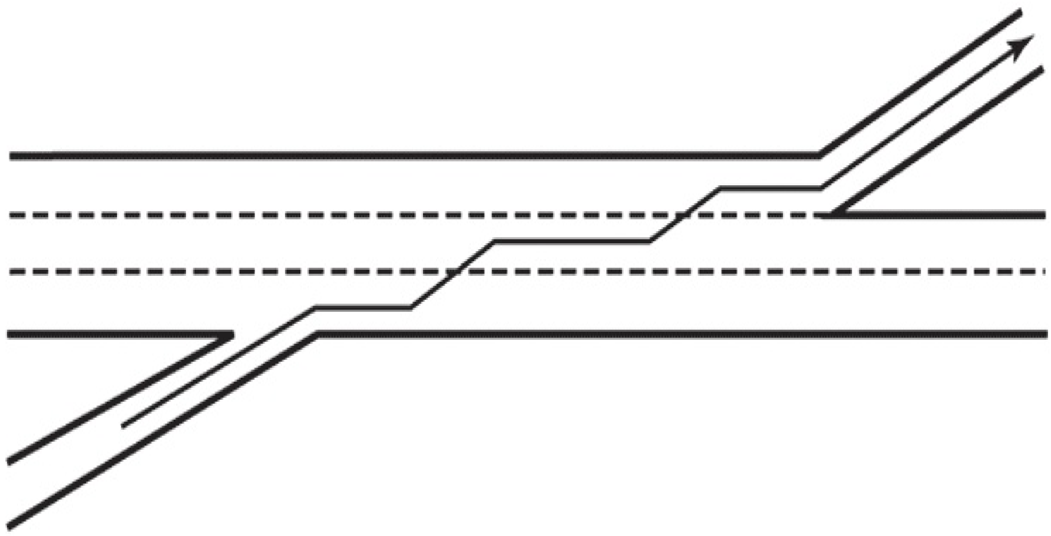
(a) One-Sided Ramp-Weave

29.5-2 Full Alternative Text



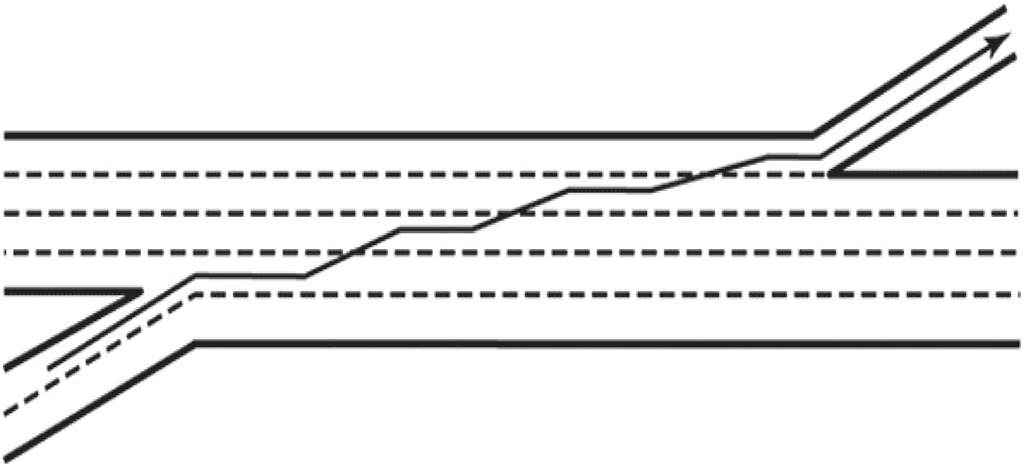
(b) One-Sided Major Weave

29.5-2 Full Alternative Text



(c) Two-Sided Weaving with Single-Lane Ramps

29.5-2 Full Alternative Text



(d) Two-Sided Weaving with Three Lane Changes

29.5-2 Full Alternative Text

(Source: Reprinted with permission from *Highway Capacity Manual, 6th Edition: A Guide for Multimodal Mobility Analysis*, Transportation Research Board, National Academy of Sciences, Courtesy of the National Academies Press, Washington, D.C., 2016.)

Figure 29.4 (a) shows a one-sided ramp-weave segment. Created by an on-ramp followed by an off-ramp connected by a continuous auxiliary lane, every weaving vehicle must make at least one lane change: on-ramp vehicles from the auxiliary lane to the right lane of the facility, off-ramp vehicles from the right lane of the facility to the auxiliary lane. As both ramps are on the right side of the freeway, these lane changes are somewhat restricted to one side of the facility. Figure 29.4 (b) is a major weave segment, as three of the four entry and exit legs have two or more lanes. Once again, however, the focus of weaving lane changes is on one side (the right) of the facility. Figure 29.4(c) is the most common two-sided configuration. In this case, a left-side on-ramp is closely followed by a right-side off-ramp; the reverse arrangement produces a similar configuration. Ramp-to-ramp vehicles must cross the entire facility, and will occupy every lane within the segment for some period of time. Figure 29.4(d) is a major weave, again because three entry and exit lanes have two or more lanes. It is clearly also a two-sided configuration, as ramp-to-ramp vehicles again must cross most of the lanes of the facility, making at least three lane changes.

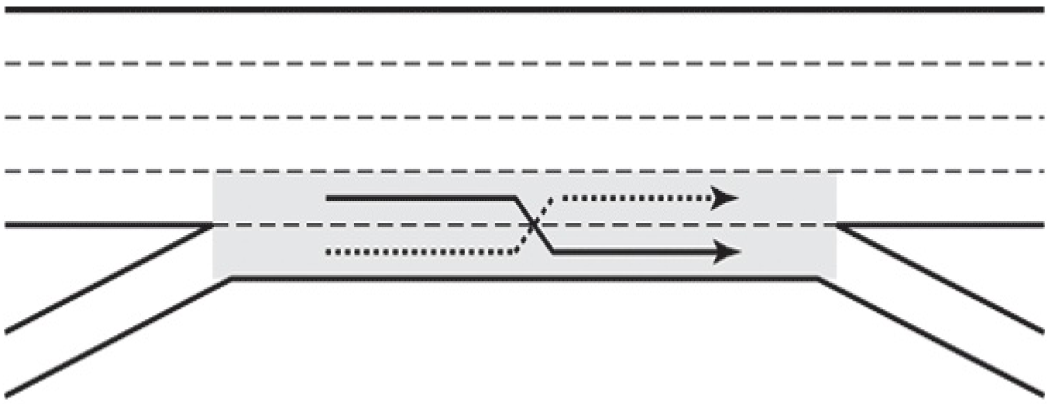
Numerical Characteristics of One-Sided Weaving Configurations

Three numerical descriptors have been defined that quantify the key element of configuration. It is noted that these definitions apply only to one-sided weaving segments, in which the ramp-to-facility and facility-to-ramp movements are the weaving movements:

LCRF=minimum number of lane changes thata ramp-to-
facility weaving vehiclemust make to successfully complete theramp-to-facility movement,LCFR=minimum number of lane changes thata facility-to-ramp weaving vehicle mustmake to successfully complete the facility-to-
ramp movement, andNWV=number of lanes from which a weavingmaneuv

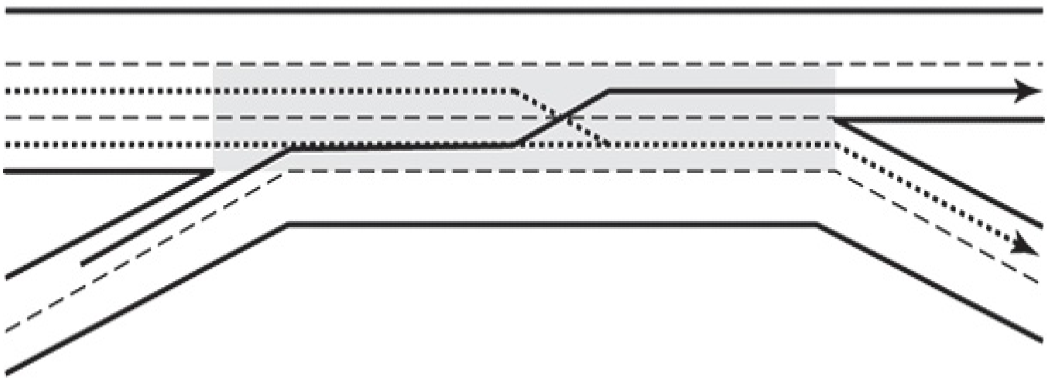
Figure 29.5 illustrates these critical parameters. The values of LC_{RF} and LC_{FR} are determined by assuming that every weaving vehicle enters the section in the lane closest to its desired exit, and leaves the section in the lane closest to its entry. In Figure 29.5 (a), all ramp-to-facility vehicles enter in the auxiliary lane and leave in the right-most lane of the facility. Facility-to-ramp vehicles enter in the right-most lane of the facility and leave in the auxiliary lane. Each vehicle in both flows must make one lane change to successfully complete their desired maneuver. For this case, LCRF=LCFR=1. Any weaving vehicle entering or leaving on a facility lane that is NOT the right-most lane would have to make two or more lane changes. Thus, the only lanes in which weaving may be accomplished with a single lane change are the auxiliary lane and the right lane of the facility, that is, NWV=2.

Figure 29.5: Configuration Parameters Illustrated



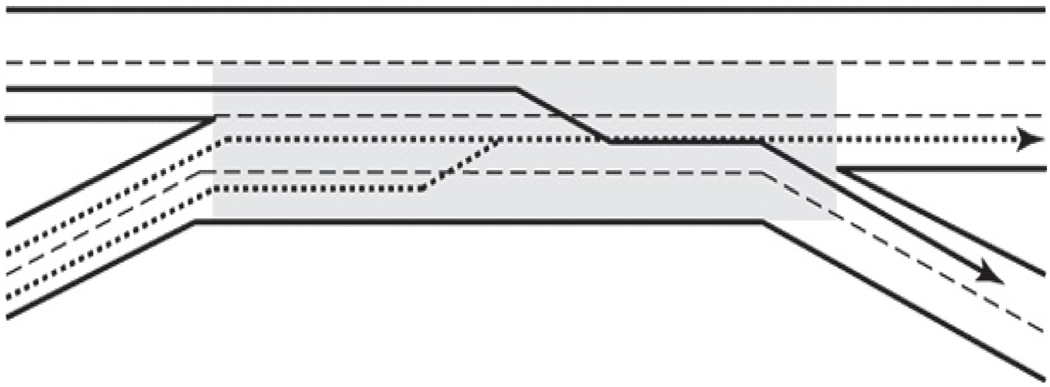
(a) A Five-Lane Ramp-Weave Segment

29.5-2 Full Alternative Text



(b) A Four-Lane Major Weave Segment (No Lane Balance)

29.5-2 Full Alternative Text



(c) A Four-Lane Major Weave Segment (With Lane Balance)

29.5-2 Full Alternative Text

(Source: Reprinted with permission from *Highway Capacity Manual, 6th Edition: A Guide for Multimodal Mobility Analysis*, Transportation Research Board, National Academy of Sciences, Courtesy of the National Academies Press, Washington, D.C., 2016.)

Figure 29.5 (b) is a major weaving segment. Vehicles weaving from right to left are assumed to enter on the left lane of the on-ramp and leave on the right lane of the left exit leg. The configuration requires that one lane change be made to do this, that is, LCRF=1. Weaving vehicles moving from the left leg to the right leg have a simpler task. A vehicle entering on the right lane of the left entry leg and leaving on the left lane of the right entry leg can do so without making any lane changes. This occurs because two entry leg lanes merge into a single lane. In this case, LCFR=0. As shown by the dotted line in Figure 29.5 (b), a left-to-right weaving vehicle may also enter in the second lane of the left leg and leave in the left lane of the right leg by making a single lane change. Because of this, weaving vehicles may enter the segment on either of the two middle lanes and weave with no more than one lane change, that is, NWV=2.

Figure 29.5(c) is also a major weave section. Its most distinctive characteristic occurs at the exit gore area: lane balance. *Lane balance* exists at an exit gore when the number of lanes leaving the exit gore is *one more* than the number of lanes entering it. In this case, four lanes approach the exit gore, but five lanes depart it. One approaching lane splits to two at the exit. This provides for great flexibility in use of that lane. Vehicles approaching in that lane may access either exit leg without making a lane change. Vehicles entering on the left lane of the right entry leg and exiting on the right lane of the left exit leg can do so without making a lane change, that is, LCRF=0. Vehicles entering on the right lane of the left leg and leaving on the left lane of the right leg may do so by making a single lane change, that is, LCFR=1. As shown by the dotted line in Figure 29.5 (c), vehicles may also enter on the right lane of the right leg and leave on the right lane of the right leg by making a single lane change. Thus, weaving vehicles may enter any of the three right-most lanes of the weaving segment and successfully complete their desired maneuvers with no more than one lane change, that is, NWV=3.

In terms of one-sided weaving segments, values of LC_{FR} and LC_{RF} are normally 0 or 1. In some cases, a value of 2 is also possible. The value of N_{WV} can be either 2 or 3; no other values are possible.

Numerical Characteristics of Two-Sided Weaving Configurations

In two-sided configurations, ramp-to-facility and facility- to-ramp movements are NOT the weaving flows. In such configurations, the ramp-to-ramp vehicles weave across facility-to-facility vehicles. Although the through vehicles on the facility actually weave in such sections, they are the dominant movement, and do not have to make any lane changes in the segment.

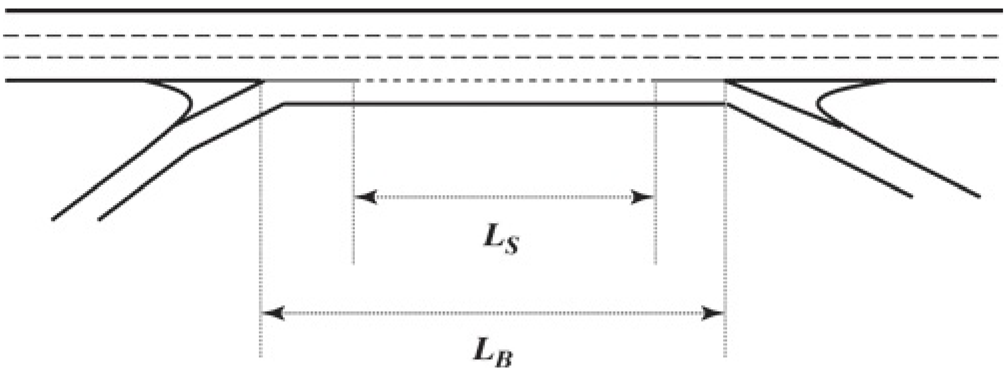
Therefore, in a two-sided weaving configuration, only the ramp-to-ramp vehicles are considered to be "weaving." The minimum number of lane changes needed to successfully move from ramp to ramp is the key characteristic: LC_{RR}. In both Figures 29.4(b) and (c), shown previously, this value is 3. By definition, in all two-sided weaving segments, N_{WV} is set to "0."

29.5.2 Length of the Weaving Area

While configuration has a tremendous impact on the number of lane changes that must be made within the confines of the weaving area, the length of the section is a critical determinant of the *intensity* of lane-changing within the section. As all of the *required* lane changes must take place between the entry and exit gores of the weaving area, the length of the section controls the intensity of lane-changing. If 1,000 lane changes must be made within the weaving area, then the intensity of those lane changes will be half as high if the section length is 1,000 feet as compared with 500 feet.

Figure 29.6 shows two potential ways in which the length of a weaving segment could be measured. Both of these represent changes from the definition of length used in the 1965 through 2000 editions of the HCM.

Figure 29.6: Measuring the Length of a Weaving Segment



(Source: Reprinted with permission from *Highway Capacity Manual, 6th Edition: A Guide for Multimodal Mobility Analysis*, Transportation Research Board, National Academy of Sciences, Courtesy of the National Academies Press, Washington, D.C., 2016.)

Figure 29.6: Full Alternative Text

These lengths are defined as:

LS=Short length (ft); the distance between theend points of any barrier mar hibit or discourage lane-
changing.LB=Base length (ft); the distance betweenpoints in the respective

While logic would indicate that the base length would be the best measure, all of the algorithms calibrated for this methodology produced significantly better results when the short length was used. Therefore, the methodology uses the short length as the input parameter in all elements. This is not to suggest that there is no lane-changing over a barrier line in a weaving segment. Lane changes can generally be observed over barrier lines and, indeed, even painted gore areas. Such barrier markings do, however, act as a partial deterrent, and the majority of lane changes do take place over the dashed line.

In some cases, barrier markings are not used, and the two lengths are the same. If an analysis of a future situation is conducted, the appropriate length should be based upon local or agency policy regarding the marking of weaving segments. Where even that is not available, a default value (based upon the data base used in developing the methodology) may be used in which LS=0.77×LB.

29.5.3 Width of a Weaving Area

The total width of the weaving area is measured as the total number of lanes available for all flows, *N*. The width of the section has an impact on the total number of lane changes that drivers can choose to make.

29.6 Computational Procedures for Weaving Area Analysis

The computational procedures for weaving areas are most easily used in the operational analysis mode (i.e., all geometric and traffic conditions are specified), and the analysis results in a determination of level of service and weaving segment capacity. The steps in the procedure are illustrated by the flow chart in Figure 29.7.

Figure 29.7: Flow Chart for Weaving Segment Methodology

STEP 1: INPUT
Specify geometry, weaving and non-weaving demand volumes, free-flow speed of the section.

STEP 2: VOLUME ADJUSTMENT
Adjust demand volumes to reflect the peak hour factor (PHF), heavy vehicle presence, and driver population.

STEP 3: DETERMINE CONFIGURATION CHARACTERISTICS
Determine the lane-change characteristics that define the effects of configuration.

STEP 4: DETERMINE THE MAXIMUM WEAVING LENGTH
Estimate the maximum length for weaving operations under the conditions specified.

If length exceeds the maximum:
TREAT AS SEPARATE RAMP JUNCTIONS

If length ≤ the maximum:
CONTINUE

STEP 5: DETERMINE THE CAPACITY OF THE WEAVING SECTION
Estimate weaving section capacity and the v/c ratio for the existing or projected demand flow rates.

If v/c > 1.00:
LEVEL OF SERVICE = F

If v/c ≤ 1.00:
CONTINUE

STEP 6: DETERMINE LANE-CHANGING RATES
Estimate the rate at which weaving and non-weaving vehicles make lane changes.

STEP 7: DETERMINE THE AVERAGE SPEED OF WEAVING AND NON-WEAVING VEHICLES
Estimate the average speed of weaving and non-weaving vehicles in the weaving section; compute the space mean speed of all vehicles in the weaving section.

STEP 8: DETERMINE THE LEVEL OF SERVICE
Convert the space mean speed to the density of the weaving section. Compare results to level of service criteria and assign the appropriate level of service.

(Source: Reprinted with permission from *Highway Capacity Manual, 6th Edition: A Guide for Multimodal Mobility Analysis*, Transportation Research Board, National Academy of Sciences, Courtesy of the National Academies Press, Washington, D.C., 2016.)

Figure 29.7: Full Alternative Text

As with most analysis methodologies, the first step is always to specify the segment under study and its demand flows. For an existing case, these will be based on measured characteristics. For future cases, the geometry would be based upon a proposed plan or design, and the demand flows (and characteristics) would be based upon forecasts. Where not all information is available, default values may be used; these can be based upon regional or agency policies, or on national recommendations. Such recommendations are included in the HCM 2016.

The second step has already been discussed. All demand volumes must be converted to flow rates in pc/h for equivalent ideal conditions. This is done using Equation 29-1 and adjustment factors from Chapter 28 for basic freeway and multilane highway segments.

The remainder of the methodology is based upon four types of models:

- Algorithms to predict the total rate of lane-changing taking place in the weaving segment. This includes both required and optional lane changes made by weaving vehicles and lane changes made by nonweaving vehicles. The total rate of lane-changing is a measure of turbulence, and reflects both demand flow rates and configuration characteristics.

- Algorithms to predict the average speed of weaving and nonweaving vehicles within the weaving segment, given stable operations, that is, NOT LOS F.

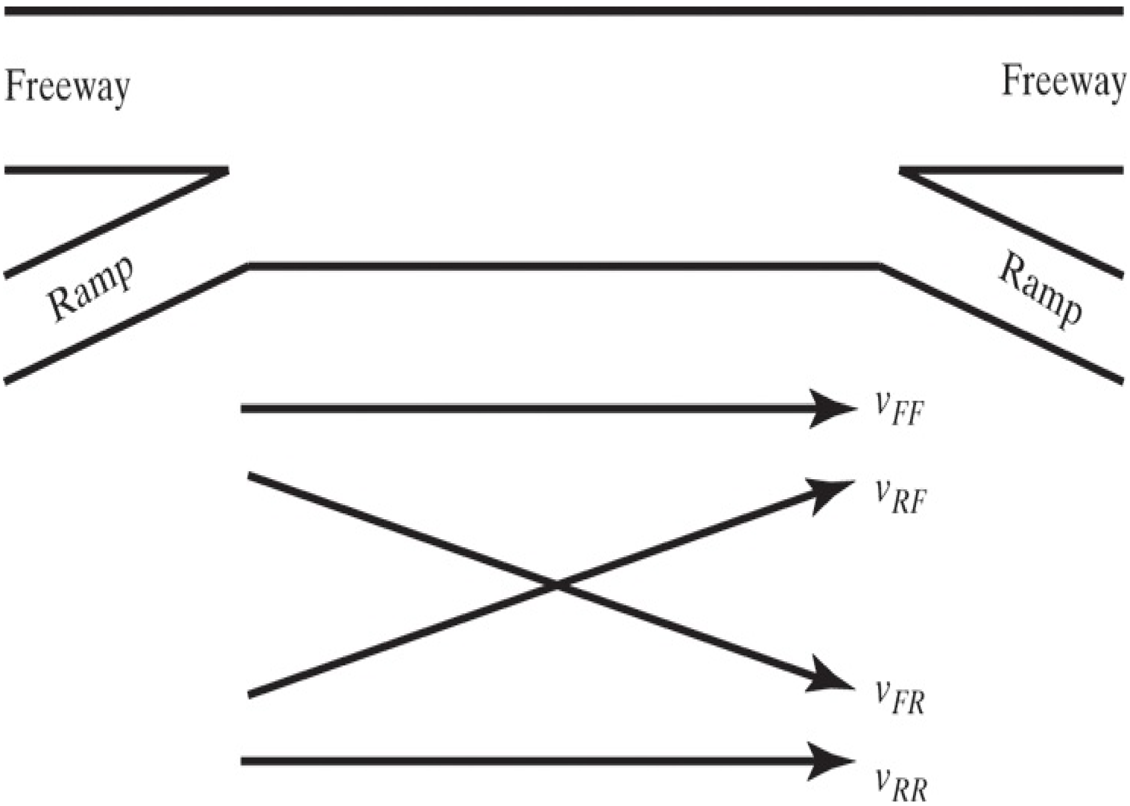
- Algorithms to predict the capacity of the weaving segment under both ideal and prevailing conditions.

- An algorithm to estimate the maximum length at which weaving operations exist. Longer segments, even if an apparent weaving configuration exists, operate as if the merge and diverge operations were separate. In such cases, the entry and exit gore areas are separately analyzed using the merge and diverge methodologies presented in Chapter 30.

29.6.1 Parameters Used in Weaving Computations

There are a very large number of variables that are used as input to, output from, or intermediate values in the overall methodology. It is convenient to define them in one place, rather than spread them across the chapter. Figure 29.8 illustrates and defines variables used in the analysis of one-sided weaving segments. Figure 29.9 does so for two-sided weaving segments. As discussed, the basic definition of weaving and nonweaving flows is different in one-sided and two-sided segments, and this influences several portions of the methodology.

Figure 29.8: Weaving Variables for One-Sided Weaving



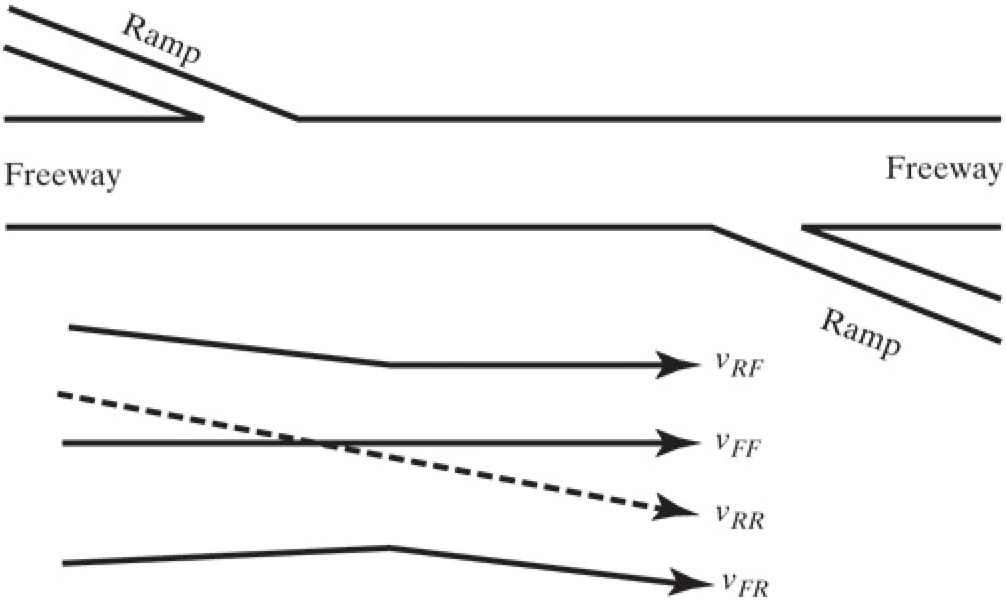
(Source: Reprinted with permission from *Highway Capacity Manual, 6th Edition: A Guide for Multimodal Mobility Analysis*, Transportation Research Board, National Academy of Sciences, Courtesy of the National Academies Press, Washington, D.C., 2016.)

vFF=freeway-to-
freeway demand flow rate in the weaving segment in passenger cars p
to-
freeway demand flow rate in the weaving segment (pc/h)vFR=freewa
to-
ramp demand flow rate in the weaving segment (pc/h)vRR=ramp-
to-
ramp demand flow rate in the weaving segment (pc/h)vW=weaving de
weaving demand flow rate in the weaving segment (pc/h), vNW=vFF+
weaving vehicles within the weaving segment (mi/h)S=average speed
flow speed of the weaving segment (mi/h)D=average density of all ve
ramp to the freewayLCFR=minimum number of lane changes that mu
rampLCMIN=minimum rate of lane-
changing that must exist for all weaving vehicles to complete their we
(LCFR×vFR)LCW=total
−segment (lc/h)LCNM=total rate of lane-changing by non-

weaving vehicles within the weaving segment (lc/h)LCALL=total rate
changing of all vehicles within the weaving Segment (lc/h), LCALL=

Figure 29.8: Full Alternative Text

Figure 29.9: Weaving Variables for Two-Sided Weaving Segment



(Source: Reprinted with permission from *Highway Capacity Manual, 6th Edition: A Guide for Multimodal Mobility Analysis*, Transportation Research Board, National Academy of Sciences, Courtesy of the National Academies Press, Washington, D.C., 2016.)

vW=total weaving demand flow rate within the weaving segment (pc/l
weaving demand flow rate within the weaving segment (pc/h), vNW=
to-
ramp vehicle to complete aweaving maneuverLCMIN=minimum rate
changing that must exist for all weaving vehicles to complete their we

Figure 29.9: Full Alternative Text

All other variables as defined in Figure 29.7

29.6.2 Volume Adjustment (Step 2)

Equation 29-1, presented previously, is used to convert all component demand volumes to demand flow rates under equivalent base (or ideal) conditions.

29.6.3 Determining Configuration Characteristics (Step 3)

Two parameters quantify the impact of configuration on lane-changing. One of these is the number of lanes *from which* weaving maneuvers can be completed with no more than one lane change, N_{WV}, which has been previously discussed and defined. The second is LC_{MIN}. This is defined as the minimum rate at which weaving vehicles must change lanes to successfully complete *all* weaving maneuvers in lane changes per hour (lc/h). It is easily determined from the values of LC_{FR}, LC_{RF}, and LC_{RR}, which have also been defined previously:

For *one-sided* weaving segments:

LCMIN=(LCFR×vFR)+(LCRF×vRF) [29-2]

For *two-sided* weaving segments:

LCMIIN=LCRR×vRR [29-3]

where all variables are as defined in Figures 29.8 and 29.9.

LC_{MIN} effectively quantifies the hourly rate of lane changes that *must* be made by all weaving vehicles to successfully reach their desired destinations. It is NOT the total lane-changing rate in the segment, which is determined later in the methodology. Total lane-changing includes optional lane changes made by weaving vehicles and all lane changes made by nonweaving vehicles. The importance of LC_{MIN} is that it is primarily a function of the configuration, which forces all of these lane changes to be made *within the confines of the weaving segment*. Optional lane changes, whether made by weaving or non-weaving vehicles, can be made within the weaving segment, but could just as easily be made upstream or downstream of the weaving section.

29.6.4 Determining the Maximum Weaving Length (Step 4)

"Weaving" implies that vehicles involved in such maneuvers are using the length of the segment to complete their maneuvers. When the length of the segment is long enough, however, merging at the entry gore and diverging at the exit gore are physically separate, and weaving does not exist. Analytically, such cases are treated as separate merge and diverge segments, with the potential for some length of basic facility between them.

Defining the "maximum length," however, can be accomplished using two different interpretations. In general terms, it is the length at which weaving turbulence no longer has an impact on operations in or capacity of the segment. Unfortunately, basing the maximum length on operational equivalence to basic facilities results in far longer distances compared to basing it on capacity equivalence. If the operational definition were used, however, the resulting capacities of the weaving segments could be significantly higher than the capacities of similar basic facility segments. Therefore, the methodology bases the determination of maximum weaving length on capacity equivalence. The following regression equation has been calibrated:

$$LMAX=[5{,}728\ (1+VR)1.6]-[1{,}566\ NWV] \quad [29\text{-}4]$$

where all variables are as defined in Figures 29.8 and 29.9.

The model indicates that the maximum weaving length increases as VR, the volume ratio, increases. This is quite logical, as when more of the total traffic is weaving, the impact of weaving is expected to extend over a longer distance. The maximum weaving length decreases with N_{WV}. This variable can only be 2 or 3, (except for two-sided weaves, where N_{WV} is defined as “0”) and represents the number of lanes from which a weaving maneuver can be completed with one or fewer lane changes. Given the same flow and split of weaving vehicles, there will be fewer lane changes in a segment in which N_{WV} is 3 than in one in which the value is 2.

Once estimated, the actual weaving length of the segment under study (short length definition) must be compared to the maximum:

- If LMAX≥LS, continue the analysis using the weaving methodology.
- If LMAX<LS, use merge and diverge analysis methodologies presented in Chapter 30.

It should be noted that Equation 29-4 was calibrated for freeways. Its application to weaving segments on multilane highways and C-D roadways is highly approximate.

29.6.5 Determine the Capacity of the Weaving Segment (Step 5)

The methodology calls for determining capacity before investigating operating parameters and level of service. This is because the models used in estimating densities and speeds within the weaving segment are only valid for cases in which flow is stable, that is, LOS is NOT F. LOS F exists when the demand flow exceeds the capacity of the segment. Logically, then, capacity must be known to determine if stable flow exists; only then can valid estimates of density and speed be made.

There are two situations in which breakdown is expected in a weaving segment:

- Breakdown of a weaving section is expected when the total demand flow exceeds the total capacity of the segment. In practice, this breakdown occurs when a density of 43 pc/mi/ln is reached in the freeway weaving segment (40 pc/mi/ln for multilane highways).
- Breakdown of a weaving section is expected when the total weaving flow rate exceeds the capacity of the segment to handle weaving

flows. The following criteria define the maximum weaving flow rates (total, both weaving flows) that can be accommodated in a weaving segment:

- 2,400 pc/h when NWV=two lanes
- 3,500 pc/h when NWV=three lanes

Capacity of a Weaving Segment Based Upon Breakdown Density

The breakdown density of 43 pc/mi/ln (for freeways) is a logical extension of the calibrated breakdown density on basic freeway segments—45 pc/h/ln. Given the additional turbulence present in weaving segments, it is logical to assume that breakdown would occur at a lower density. Further, the research behind this methodology [2] found no stable operations at higher densities. Fortunately, the methodology does not require trial-and-error computations until the breakdown density is reached. A relatively straightforward regression relationship was calibrated that estimates the capacity at which this density occurs.

Because of turbulence in the weaving segment, and the fact that some weaving segment lanes cannot be used to full advantage due to the existing split between component flows, the capacity controlled by a density of 43 pc/h/ln must be less than the capacity of a lane on a basic facility segment with the same free-flow speed as the weaving segment. Therefore, the algorithm for estimating this capacity is essentially a deduction from the basic facility capacity:

cIWL=cIFL−[438.2 (1+VR)1.6]+[0.0765 LS]+[119.8 NWV] [29-5]

where:

cIWL=capacity per lane of the weaving sectionunder ideal conditions, (pc/l
ment with the same free-flow speed asthe weaving segment (pc/h/ln).

All other variables as previously defined.

Values of basic facility capacity under ideal conditions are taken from

Chapter 28, but are repeated in Table 29.2 for convenience.

Table 29.2: Basic Facility Capacity Values (c_{IFL}) for Use in Equation 29-5

Freeways		Multilane Highways and C–D Roadways	
FFS (mi/h)	Capacity (pc/h/ln)	FFS (mi/h)	Capacity (pc/h/ln)
≥75	2,400	≥65	2,300
70	2,400	60	2,200
65	2,350	55	2,100
60	2,300	50	2,000
55	2,250	45	1,900

Table 29.2: Full Alternative Text

The weaving segment capacity per lane under ideal conditions must now be converted to a total capacity for the weaving segment under prevailing conditions:

cW1=cIWL N fHV [29-6]

where:

cW1=capacity of the weaving section basedupon breakdown density, veh/h

All other variables as previously defined.

Capacity of a Weaving Segment Based upon Maximum Weaving

Flow Rates

It is possible for the split among component flows to be such that the number of weaving vehicles reaches its capacity *before* the density of the entire weaving segment reaches 43 pc/h/ln. In these cases, the effective control on the capacity of the segment is the limiting values of weaving flow rate noted earlier. Because the proportion of weaving vehicles is a traffic characteristic of the demand (i.e., fixed for any given analysis), weaving turbulence can cause a breakdown while there is still "capacity" available for nonweaving vehicles. In this type of breakdown, on-ramp vehicles queue on the ramp, while off-ramp vehicles queue on the approaching facility segment. Freer flow may exist in the most distant outside lane(s). Capacity of the weaving segment based upon maximum weaving flow rates is found as:

cIW=2,400VR for NWV=2cIW=3,500VR for NWV=3 [29-7]

where:

cIW=capacity of the weaving segment underideal conditions (pc/h).

All other variables as previously defined.

Note that unlike Equation 29-5, which defined weaving capacity under ideal conditions *per lane*, Equation 29-7 defines the total capacity of the weaving segment under ideal conditions. This, of course, must be converted to prevailing conditions:

cW2=cIW N fHV [29-8]

where:

cW2=capacity of a weaving segment basedupon maximum weaving flow ra

All other variables as previously defined.

Final Capacity of the Weaving Segment and the v/c Ratio

As there are two controls on capacity of a weaving segment, the actual capacity is based upon the smallest of the two values computed in Equations 29-6 and 29-8:

$$c_W = \min(c_{W1}, c_{W2}) \quad [29\text{-}9]$$

In cases where a capacity adjustment factor is appropriate for the facility, the capacity of the weaving segment is adjusted as:

$$c_{Wadj} = c_W \, CAF \quad [29\text{-}10]$$

The effective demand-to-capacity ratio is simply the ratio of the total demand flow to the estimated capacity. At this point in the methodology, the demand flow rate, *v*, is expressed in pc/h under equivalent ideal conditions, while capacity, c_W, is expressed in veh/h under prevailing conditions. Thus, to find the appropriate ratio, one must be converted so that both are stated in the same terms:

$$v/c = \frac{v \times f_{HV}}{c_{Wadj}} \quad [29\text{-}11]$$

where all terms have been previously defined.

Final Assessment of Capacity

If the *v/c* ratio exceeds 1.00, LOS F is automatically assigned, and all computations cease. If the *v/c* ratio is less than or equal to 1.00, computations continue to find speed and density within the weaving segment.

29.6.6 Determining Total Lane-Changing Rates within the Weaving Segment (Step 6)

There are three types of lane-changing maneuvers that exist within a weaving segment:

- *Required* lane changes made by *weaving vehicles*: These lane changes must be made to successfully complete a weaving maneuver. They represent the *absolute minimum lane-changing rate* that can exist in the weaving section for the defined demands. By definition, these lane changes must be made within the confines of the weaving segment. This has been discussed previously, and the rate for such lane changes is defined as LC_{MIN}, and was determined in Step 3 of the methodology.

- *Optional* lane changes made by *weaving vehicles*: These involve lane changes by weaving vehicles that choose to enter the segment on a lane *that is not the closest to their desired destination*, and/or leave the segment on a lane *that is not the closest to their entry leg*. Such entries and exits require additional lane changes to be made within the weaving segment, and act to increase turbulence.

- *Optional* lane changes made by *nonweaving vehicles*: Nonweaving vehicles are never required to make lane changes within a weaving segment. They may, however, *choose* to make lane changes to avoid perceived turbulence.

While LC_{MIN} is known based upon the segment configuration and component demand flow rates, the last two categories of optional lane-changing are estimated based upon regression equations developed in Ref. 16. Total lane-changing rates are separately determined for weaving vehicles and nonweaving vehicles.

Total Lane-Changing Rate for Weaving Vehicles

The total lane-changing rate for weaving vehicles in a weaving segment is estimated as:

LCW=LCMIN + 0.39 [(LS−300)0.5 N2 (1+ID)0.8] [29-12]

where:

LCW=total lane-changing rate for weavingvehicles within a weaving segment, lc/h,N=numb

Other variables as previously defined.

The term LS−300 is interesting. It suggests that for a segment shorter than 300 ft, weaving vehicles *do not make any optional lane changes*. As the second term of the equation cannot be negative (LC_W can never be less than LC_{MIN}), for all weaving lengths less than 300 ft (a hopefully very rare event), L_S must be set at 300 ft.

The equation is logical in its form. As length increases, weaving vehicles have more distance and time to make optional lane changes. As the number of lanes, *N*, increases, there are more possible lane changes that can be made.

The interchange density (ID) is unique to weaving analysis. While ID was used in HCM 2000 to predict the FFS of a basic freeway segment, in HCM 2010 and HCM 2016, ID was replaced by total ramp density. The weaving methodology, however, was calibrated before this change was made. A higher ID yields more lane-changing as weaving vehicles align themselves as a result of upstream or downstream turbulence.

For weaving sections, ID considers a facility segment 3 miles upstream and downstream of the middle of the weaving segment. The weaving segment itself counts as *one* interchange within this 6-mile range.

When applying Equation 29-12 to a weaving segment on a multilane highway, ID is replaced by the density of roadside access points in the analysis direction. *Only* significant unsignalized access points should be considered in this density. Application of Equation 29-12 to multilane highway weaving segments is highly approximate.

Total Lane-Changing Rate for Nonweaving Vehicles

Because no nonweaving vehicle *must* make a lane change within the weaving segment, all such lane changes are optional. This makes them far more difficult to predict than weaving vehicle lane changes, which are tied to the configuration of the weaving segment and the demand flow rates. The methodology has *two* basic equations that are used to estimate nonweaving lane-changing rates:

$$LC_{NW1} = (0.206 v_{NW}) + (0.542 L_S) - (192.6\ N) \qquad LC_{NW2} = 2{,}135 + 0.223\,(v_{NW} - 2{,}000) \quad [29\text{-}13]$$

where:

LC_{NW1} = first estimate, nonweaving vehicle lane-changing rate, lc/h, and LC_{NW2} = second estimate, nonweaving vehicle lane-changing rate, lc/h.

All other variables as previously defined.

The first equation covers the majority of situations. It presents a logical set of trends. As nonweaving flow increases, nonweaving lane-changing also increases. As the length of the segment increases, nonweaving lane-changing increases, as such vehicles have more distance and time to make such movements. Nonweaving lane-changing *decreases* as the number of lanes in the weaving segment increases. This is less obvious. As the width of the weaving segment increases, nonweaving vehicles have a better opportunity to segregate from weaving vehicles in outer lanes. This would tend to decrease their desire to make lane changes out of these lanes. The first equation has an arbitrary minimum of “0.”

The two equations, unfortunately, are *very* discontinuous. Therefore, it is critical to have a methodology that provides for smooth transitions from one equation to the other without distorting the results. This is done using a lane-changing index, I_{NW}:

$$I_{NW} = \frac{L_S\, ID\, v_{NW}}{10{,}000} \quad [29\text{-}14]$$

where all variables are as previously defined. The origin of this index is to explain when the second Equation 29-13 is used. It applies to cases in which long lengths, high interchange densities, and/or high nonweaving flows conspire to create far more lane-changing among such vehicles than normally expected. In calibrating these algorithms [16], the first equation applies to cases in which $I_{NW} \le 1{,}300$. The second applies to cases in which $I_{NW} \ge 1{,}950$. For values in between, a straight-line interpolation of the two equations is used. Thus:

LCNW=LCNW1 INW≤1300 or LCNW1>LCNW2LCNW=LCNW2 INW (LCNW2−LCNW1) (INW−1300650) 1950>INW>1300 [29-15]

Total Lane-Changing in a Weaving Segment

The total lane-changing rate in any weaving segment is simply the sum of the lane-changing rate for weaving vehicles and the lane-changing rate for non-weaving vehicles:

LCALL=LCW+LCNW [29-16]

where:

LCALL=total lane-changing rate in a weavingsegment, lc/h.

All other variables as previously defined.

29.6.7 Determining the Average Speed of Vehicles within a Weaving Segment (Step 7)

The heart of the methodology for weaving segments is the estimation of average speeds within the weaving segment. The average speeds of weaving and nonweaving vehicles are estimated separately, as they are affected by different factors and can be quite different in some cases. Estimated speeds, together with known demand flow rates, will yield a density estimate, which is used to determine the level of service. Thus, while speed is a secondary performance measure for weaving segments, it must be computed to obtain an estimate of density—the primary measure of effectiveness used for weaving segments.

Average Speed of Weaving

Vehicles

The general algorithm for prediction of the average speed of weaving vehicles in a weaving segment is basically the same as that in HCM 2000:

SW=SMIN+(Smax−SMIN1+W) [29-17]

where:

SW=average speed of weaving vehicles, mi/h,SMIN=minimum average sp

The maximum speed of weaving vehicles is the free-flow speed (perhaps adjusted by the SAF) of the facility. The minimum average speed is set at 15 mi/h. The weaving intensity factor, *W*, is found as:

W=0.226 (LCALLLS)0.789 [29-18]

where all variables are as previously defined. The term LC_{ALL}/L_S is essentially a measure of lane-changing intensity over length—total lane changes per foot of weaving segment length. Thus, lane-changing behavior becomes the primary measure of weaving intensity. Then:

SW=15+(FFS×SAF−151+W) [29-19]

where all terms are as previously defined.

The term (1+W) is used instead of *W* because *W* can be less than or greater than 1.00. Dividing by a number that can be less than or more than 1.00 creates inconsistent arithmetic results. The (1+W) ensures that all denominators are more than 1.00, and that as *W* increases, speed decreases.

Average Speed of Nonweaving Vehicles

The average speed of nonweaving vehicles is treated as a reduction from the free-flow speed according to the following algorithm:

SNW=FFS×SAF−(0.0072 LCMIN)−(0.0048 vN) [29-20]

where all terms are as previously defined.

Nonweaving speed obviously decreases as *v/N* increases. More surprising is the appearance of LC_{MIN} in the equation. As this is a regression equation, its appearance is as a measure of weaving turbulence. For nonweaving speeds, it was a stronger statistical predictor than other measures, such as *W* or LC_{ALL}.

Average Speed of All Vehicles

Given estimates of both average speed of weaving vehicles and average speed of nonweaving vehicles, a space mean speed for all vehicles may be computed as:

S=vW+vNW(vWSW)+(vNWSNW) [29-21]

where all variables are as previously defined.

29.6.8 Determining Density and Level of Service in a Weaving Segment (Step 8)

The final computation in the analysis of weaving segments is the conversion of average speed and demand flow rate into an estimate of density, from which level of service is determined using Table 29.1.

D=(vN)S [29-22]

where *D* is the density in pc/mi/ln and all other variables are as previously defined.

The methodology results in estimating both an average speed and an average density of all vehicles within the weaving segment, and a determination of the prevailing level of service given the geometric characteristics of the segment and the demand characteristics. The capacity of the weaving segment is also determined for the prevailing conditions specified. This information provides for significant insight into the expected operational characteristics of the segment, as well as insight into existing or potential problems

29.7 Sample Problems in Weaving Segment Analysis

Sample Problem 29-1: Analysis of a Ramp-Weave Area

Figure 29.10 illustrates a typical ramp-weave section on a six-lane freeway (three lanes in each direction). The analysis is to determine the expected level of service and capacity for the prevailing conditions shown.

Figure 29.10: Ramp-Weave Segment for Sample Problem 29-1

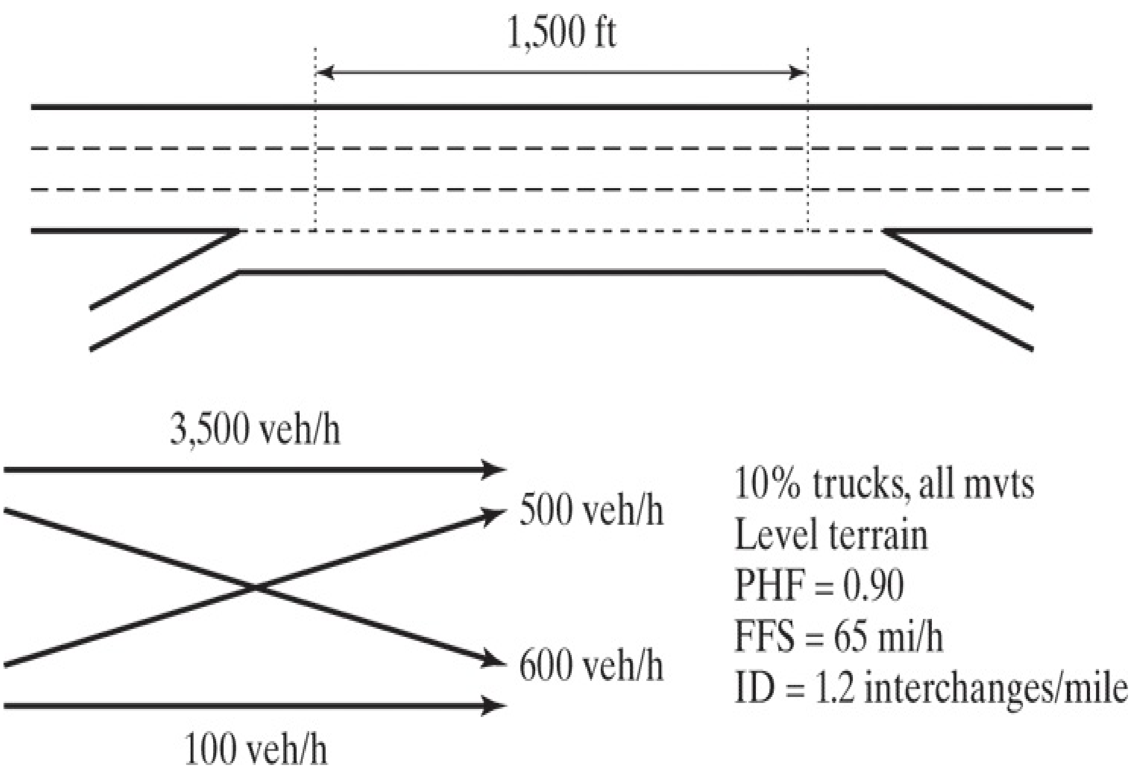


Figure 29.10: Full Alternative Text

Solution:

1. Steps 1, 2: Convert All Demand Volumes to Flow Rates in pc/h under Equivalent Base Conditions

   Each of the component demand volumes is converted to a demand flow rate in pc/h under equivalent base conditions using Equation 29-1:

   $v_i = \frac{V_i}{PHF \times f_{HV}}$

   where:

   $PHF = 0.9$ (given)

   The heavy-vehicle factor, f_{HV}, is computed using Equation 28-15 and a value of E_{HV} selected from Table 28.10 (both in Chapter 28) for trucks on level terrain ($E_{HV} = 2$). The proportion of trucks, P_{HV}, is given as 10% or 0.10. Then:

   $f_{HV} = \frac{1}{1 + P_{HV}(E_{HV} - 1)} = \frac{1}{1 + 0.10(2 - 1)} = 0.909$

   and:

   $v_{o1} = \frac{3500}{0.90 \times 0.909} = 4{,}278 \text{ pc/h} \quad v_{o2} = \frac{100}{0.90 \times 0.909} = 12$

   Other critical variables used in the analysis may now be computed and/or summarized:

   $v_W = v_{w1} + v_{w2} = 733 + 611 = 1{,}344 \text{ pc/h} \quad v_{NW} = v_{o1} + v_{o2} = 427$

2. Step 3: Determine Configuration Characteristics

   The two critical numeric variables that define configuration are N_{WV}, the number of lanes from which a weaving movement can be successfully executed with no more than one lane change, and LC_{MIN}, the minimum number of lane changes that must be made by all weaving vehicles to successfully complete their maneuvers.

The number of weaving lanes, N_{WV}, is determined by perusing the site drawing (Figure 29.10) and comparing it to the illustration of Figure 29.5. As a ramp-weave, the value of N_{WV} is 2. The value of LC_{MIN} is found from Equation 29-2:

LCMIN=(LCFR×vFR)+(LCRF×vRF)

where:

LCFR=minimum number of lane changes for a freeway-to-
ramp vehicle needed to successfully execute aweaving ma
to-
ramp demand flow rate,pc/h = vw1= 733 pc/h,LCRF=min
to-
freeway vehicle needed to successfully executea weaving
to-freeway demand flow rate,pc/h = vw2 = 611 pc/h.

Then:

LCMIN=(1×733)+(1×611)=1,344 lc/h

3. Step 4: Determine the Maximum Weaving Length

The maximum length for which this segment may be considered to be a "weaving segment" is estimated using Equation 29-4:

LMAX=[5728 (1+VR)1.6]−[1566 NWL]LMAX=
[5728 (1+0.233)1.6]−[1566×2]LMAX=
[8,008]−[3,132]=4,876 ft

As the actual length of the segment, 1,500 ft, is far less than this maximum, the segment is operating as a weaving segment, and the analysis may continue.

4. Step 5: Determine the Capacity of the Weaving Segment

   The capacity of the weaving segment can be determined by overall operation at a density of 43 pc/h/ln, the density at which it is believed breakdown occurs in weaving segments, or on the capacity of the segment to handle weaving flows. The former is estimated using Equation 29-5. It is based upon the per-lane capacity of a basic freeway segment with a 65 mi/h free-flow speed, which is 2,350 pc/h/ln (Table 29.2):

   $$c_{IWL} = c_{IFL} - [438.2\,(1+VR)^{1.6}] + [0.0765\,L_S] + [119.9\,N_{WV}]$$
   $$c_{IWL} = 2{,}350 - [438.2\,(1+0.233)^{1.6}] + [0.0765 \times 1500] + [119.9 \times 2]$$
   $$c_{IWL} = 2{,}350 - 612.7 + 114.8 + 239.6 = 2{,}092 \text{ pc/h/ln}$$

   This value must be converted to a capacity under prevailing conditions using Equation 29-6:

   $$c_{W1} = c_{IWL}\,N\,f_{HV}$$
   $$c_{W1} = 2092 \times 4 \times 0.909 = 7{,}607 \text{ veh/h}$$

   The capacity based upon maximum weaving demand flow rate, based upon $N_{WV}=2$, is estimated using Equation 29-7:

   $$c_{IW} = \frac{2400}{VR} = \frac{2400}{0.233} = 10{,}300 \text{ pc/h}$$

   This value must also be converted to prevailing conditions using Equation 29-8:

   $$c_{w2} = c_{IW}\,f_{HV} = 10{,}300 \times 0.909 = 9{,}363 \text{ veh/h}$$

   The limiting capacity is obviously based upon the density condition, that is, 7,607 veh/h. As with any capacity, this is defined in terms of a maximum demand flow rate that the segment can accommodate without breakdown. This must be compared with the demand flow rate, also under prevailing conditions. The total demand flow rate, *v*, was computed previously (from given volumes) as 5,744 pc/h. This is already a flow rate, but must be converted to a flow rate in mixed veh/h:

$$v = v_{pc/h} \times f_{HV} = 5744 \times 0.909 = 5{,}221 \text{ veh/h}$$

As the demand flow rate is less than the capacity of the segment ($v/c = 5221/7607 = 0.686$), operations will be stable, and LOS F *does not* exist in the segment. The analysis may move forward to estimate density, level of service, and speed within the segment.

5. Step 6: Determine Lane-Changing Rates

   In order to estimate speed and density in the weaving segment, total lane-changing rates within the segment must be estimated. Lane-changing rates for weaving and non-weaving vehicles are separately estimated. The lane-changing rate for weaving vehicles is computed using Equation 29-11:

$$LC_W = LC_{MIN} + 0.39\left[(L_S - 300)^{0.5} N^2 (1 + ID)^{0.8}\right]$$
$$LC_W = 1344 + 0.39\left[(1500 - 300)^{0.5}\, 4^2\, (1 + 1.2)^{0.8}\right]$$
$$LC_W = 1244 + 0.39\,(34.64 \times 16 \times 1.88) = 1334 + 406.4 = 1{,}740$$

   The lane-changing rate for nonweaving vehicles is obtained from Equations 29-13. Use of these equations requires that the non-weaving lane-change index be determined, as defined in Equation 29-14:

$$I_{NW} = \frac{L_S \; ID \; v_{nw}}{10{,}000} = \frac{1500 \times 1.2 \times 4400}{10{,}000} = 792$$

   For this value, the Equation 29-13 for $I_{NW} \le 1{,}300$ is used:

$$LC_{NW} = (0.206\; v_{NW}) + (0.542\; L_S) - (192.6\; N)$$
$$LC_{NW} = (0.206 \times 4400) + (0.542 \times 1500) - (192.6 \times 4)$$
$$LC_{NW} = 906.4 + 813.0 - 770.4 = 949$$

   The total lane-changing rate in the segment is the sum of the weaving vehicle rate and the non-weaving vehicle rate, or:

$$LC_{ALL} = LC_W + LC_{NW} = 1740 + 949 = 2{,}689 \text{ lc/h}$$

6. Step 7: Determine the Average Speed of Weaving and Nonweaving Vehicles

   The average speed of weaving vehicles in the weaving segment is estimated using Equations 29-18 and 29-19. Equation 29-18 is used to find the weaving intensity factor, *W*:

   W=0.226 (LCALLLS)0.789=0.226 (2,6891,500)0.789=0.

   Then:

   SW=15+(FFS×SAF−151+W)=15+(65×1−151+0.3582)=51.8 mi/h

   The average speed of nonweaving vehicles in the weaving segment is estimated using Equation 29-20:

   SNW=FFS×CAF−(0.0072 LCMIN)−(0.0048 vN)SNW=65×1−(0.0072×1344)−(0.0048×5744

   These results indicate that weaving vehicles are actually traveling somewhat faster than nonweaving vehicles within the weaving segment. While unusual for ramp-weaving segments, this is entirely possible given the dominance of the through freeway flow in the segment. Nonweaving vehicles may be crowding into the two outer freeway lanes to avoid the weaving turbulence, and may therefore experience slightly lower speeds (and higher densities) than weaving vehicles.

   The average speed of all vehicles in the segment is computed from Equation 29-21:

   S=vW+vNW(vWSW)+(vNWSNW)=1344+4400(134451.8)+(440048.4)=574425.94+90.91=49.1 mi/h

7. Step 8: Determine Density and Level of Service in the Weaving Segment

The average density in the weaving segment is computed from Equation 29-23:

$$D=(vN)S=(57444)49.1=29.2\ \text{pc/mi/ln}$$

From Table 29.1, this is LOS D, but is very close to the LOS C/D boundary of 28 pc/h/ln.

Discussion

This ramp-weave segment is operating acceptably in the better portion of LOS D. The capacity of the segment is 7,607 veh/h (as a flow rate) and the demand flow rate is 5,221 veh/h. Demand flow could increase by $7607-5221=2{,}386$ veh/h, or 45.7%, before reaching capacity.

Sample Problem 29-2: Analysis of a Major Weaving Area

The freeway weaving area shown in Figure 29.11 is to be analyzed to determine the expected level of service for the conditions shown and the capacity of the weaving area. For convenience, all demand volumes have already been converted to flow rates in pc/h under equivalent base conditions. For information purposes, the following values were used to make these conversions:

Figure 29.11: Weaving Segment for Sample Problem 29-2

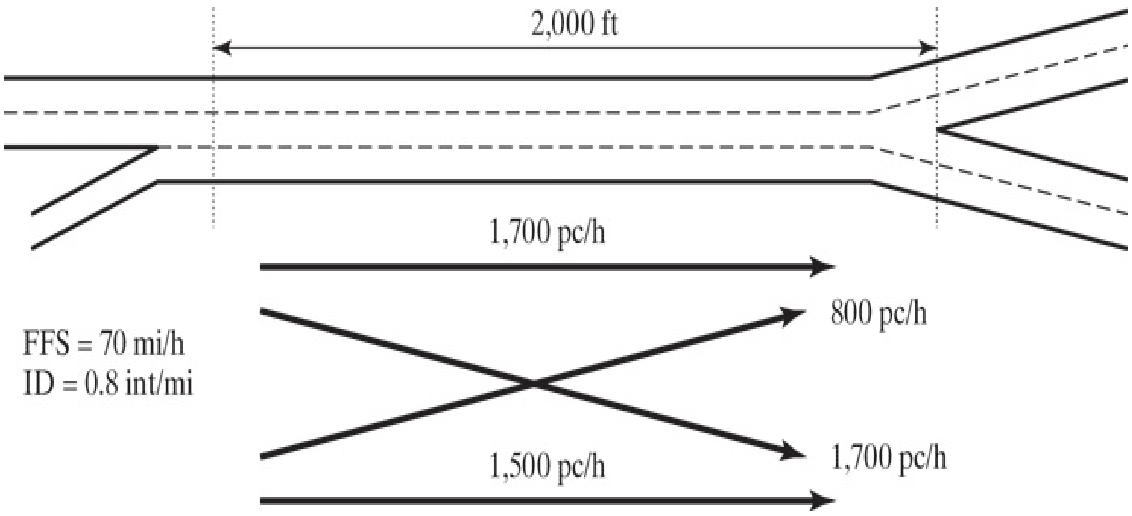


Figure 29.11: Full Alternative Text

- $PHF=0.95$
- $f_{HV}=0.93$

Solution:

1. Steps 1, 2: Convert All Demand Volumes to Flow Rates in pc/h under Equivalent Base Conditions

   As all demands are specified as flow rates in pc/h under equivalent base conditions, no further conversion of these is necessary. Key analysis variables are summarized below:

   $v_W=800+1{,}700=2{,}500 \text{ pc/h} v_{NW}=1{,}700+1{,}500=3{,}200 \text{ pc/}$

   Note that this is a major weaving configuration. The weave from left to right can be made with no lane changes ($LC_{FR}=0$), while the weave from right to left requires one lane change ($LC_{RF}=1$). Successful weaving maneuvers can be made from any of the three lanes in the segment with no more than one lane change, that is, $N_{WV}=3$.

2. Step 3: Determine Configuration Characteristics

One of the configuration characteristics, N_{WV}, was determined to be 3. The second configuration characteristic needed is LC_{MIN}, as determined by Equation 29-2:

$$LC_{MIN} = (LC_{FR}\ v_{FR}) + (LC_{RF}\ v_{RF})$$
$$LC_{MIN} = (0 \times 1{,}700) + (1 \times 800) = 800 \text{ lc/h}$$

3. Step 4: Determine the Maximum Weaving Length

   The maximum weaving length is estimated using Equation 29-4:

   $$L_{MAX} = [5{,}728\ (1+VR)^{1.6}] - [1{,}566\ N_{WV}]$$
   $$L_{MAX} = [5{,}728\ (1+0.439)^{1.6}] - [1{,}544 \times 3] = 5{,}556 \text{ ft}$$

   As the actual length of the weaving segment is only 2,000 feet, it falls within this limit, and the analysis of the segment as a weaving segment may continue.

4. Step 5: Determine the Capacity of the Weaving Segment

   In order to determine whether stable operations prevail, the capacity of the weaving segment must be determined. Capacity may be determined in two ways. It may be limited by a breakdown density of 43 pc/h/ln, or by a maximum weaving flow rate the segment can accommodate. Capacity, as determined by a breakdown density of 43 pc/h/ln, is estimated using Equation 29-5:

   $$c_{IWL} = c_{IFL} - [438.2\ (1+VR)^{1.6}] + [0.0765\ L_S] + [119.8\ N_{WV}]$$
   $$c_{IWL} = 2400 - [438.2\ (1+0.439)^{1.6}] + [0.0765 \times 2000] + [119.8 \times 3]$$
   $$c_{IWL} = 2400 - 784.5 + 153 + 359.4 = 2{,}128 \text{ pc/h/ln}$$

   This capacity is stated in terms of pc/h/ln under equivalent ideal conditions. As there are three lanes in the segment, the total capacity of the weaving segment is $2128 \times 3 = 6{,}384$ pc/h. This value is still for equivalent ideal conditions. It could be converted to veh/h under prevailing conditions by multiplying by the appropriate

f_{HV} value. This is not necessary in this case, as the demand flow rates are already stated in equivalent ideal terms and may be directly compared to capacity.

The capacity of the segment, as limited by maximum weaving flow rate, is estimated using Equation 29-8 for NWV=3:

cIW=3,500VR=3,5000.439=7,973 pc/h

As this value is larger than capacity limited by density, the smaller value is used. The capacity of the weaving segment is 6,384 pc/h under equivalent ideal conditions.

As the total demand flow rate is 5,700 pc/h, capacity is sufficient. The *v/c* ratio is 5,700/6,384=0.893, which means that demand flows are quite near capacity. A 10.7% increase in demand would create a LOS F situation. Because operations are still in the stable zone (though barely), we can continue the analysis to find the LOS.

5. Step 6: Determine Lane-Changing Rates

   In order to estimate speeds in the weaving segments, and then density and level of service, the total lane-changing rate within the weaving segment must be determined using Equations 29-12 through 29-16. Lane-changing rates are separately estimated for weaving and nonweaving vehicles. Equation 29-12 is used to estimate the lane-changing rate for weaving vehicles:

   LCW=LCMIN+0.39 [(LS −300)0.5 N2 (1+ID)0.8]LCW=800+0.39 [(2000−300)0.

   Equation 29-13 is used to estimate the rate of lane-changing among nonweaving vehicles. There are two equations presented, and the lane-changing index must be computed to interpret the results from these equations. The index is computed using Equation 29-14:

INW=LS ID vNW10,000=2000×0.8×320010,000=512<1

Because the index is less than 1,300, the *first* Equation 29-13 is used:

LCNW=(0.206 vNW)+(0.542 LS)−(192 N)LCNW=(0.206×3200)+(0.542×2000)−(192×3)LCNW=659.2+1,084.0−576.0=1,1

The total lane-changing rate in the weaving segment is the sum of the rates for weaving and nonweaving vehicles:

LCALL=512+1167=1,679 lc/h

6. Step 7: Determine the Average Speeds of Weaving and Nonweaving Vehicles

   The average speed of weaving vehicles within the weaving segment is estimated using Equations 29-18 and 29-19. Equation 29-18 determines the weaving intensity factor, *W*:

   W=0.226 (LCALLLS)0.789=0.226 (16792000)0.789=0.

   Then:

   SW=15+(FFS×1−151+W)=15+(70×1−151+0.197)=60.9 mi/h

   The average speed of nonweaving vehicles in the segment is estimated using Equation 29-20:

   SNW=FFS−(0.0072×LCMIN)−(0.0048×vN)SNW=70−(0.0072×800)−(0.0048×1900)=5

   In this case, nonweaving vehicles will be traveling over 5 mi/h slower than weaving vehicles. This is not unexpected. Weaving vehicles dominate this segment (3,200 pc/h vs. 2,500 pc/h) and the configuration favors weaving vehicles.

The average speed of all vehicles is computed using Equation 29-21:

S=vW+vNW(vWSW)+(vNWSNW)=2500+3200(250055.1)+(320060.9)=570045.37+52.55=58.2 mi/h

7. Step 8: Determine the Density and Level of Service

   Density in the weaving segment is computed using Equation 29-22:

   D=(vN)S=190058.2=32.6 pc/mi/h

From Table 29.2, this is level of service D.

Discussion

The weaving segment is currently operating stably in LOS D, but not far from the LOS E boundary. While speeds appear to be acceptable, the demand is almost 90% of the capacity, and there is little room for growth in demand at this location. The bottom line is that with virtually any traffic growth, this segment will reach capacity. Operations will deteriorate rapidly with demand growth. Even if only ambient growth is expected (as opposed to growth caused by new development), the segment should be looked at immediately for potential improvements.

References

- 1. *Highway Capacity Manual*, Transportation Research Board, Washington, D.C., 2000.

- 2. Roess, R., et al, “Analysis of Freeway Weaving Sections,” *Final Report*, National Cooperative Highway Research Program Project 3-75, Polytechnic University, Brooklyn NY, 11201, January 2008.

- 3. “Highway Capacity Manual,” *Special Report 87*, Transportation Research Board, Washington, D.C., 1965.

- 4. Pignataro, L., *et al*., “Weaving Areas—Design and Analysis,” *NCHRP Report 159*, Transportation Research Board, Washington, D.C., 1975.

- 5. Roess, R., *et al*., “Freeway Capacity Analysis Procedures,” *Final Report*, Project No. DOT-FH-11-9336, Polytechnic University, Brooklyn, NY, 1978.

- 6. “Interim Procedures on Highway Capacity,” *Circular 212*, Transportation Research Board, Washington, D.C., 1980.

- 7. Leisch, J, “Completion of Procedures for Analysis and Design of Traffic Weaving Areas,” *Final Report*, Vols 1 and 2, U.S. Department of Transportation, Federal Highway Administration, Washington, D.C., 1983.

- 8. Reilly W., *et al*., “Weaving Analysis Procedures for the New Highway Capacity Manual,” *Technical Report*, JHK & Associates, Tucson, AZ, 1983.

- 9. Roess, R., “Development of Weaving Area Analysis Procedures for the 1985 Highway Capacity Manual,” *Transportation Research Record 1112*, Transportation Research Board, Washington, D.C., 1987.

- 10. Fazio, J., “Development and Testing of a Weaving Operational Design and Analysis Procedure,” *Master’s Thesis*, University of

Illinois at Chicago, Chicago IL, 1985.

- 11. Cassidy, M. and May, A., Jr., “Proposed Analytic Technique for Estimating Capacity and Level of Service of Major Freeway Weaving Sections,” *Transportation Research Record 1320,* Transportation Research Board, Washington, D.C., 1992.

- 12. Ostrom, B., *et al*., “Suggested Procedures for Analyzing Freeway Weaving Sections,” *Transportation Research Record 1398*, Transportation Research Board, Washington, D.C., 1993.

- 13. Windover, J. and May, A., Jr., “Revisions to Level D Methodology of Analyzing Freeway Ramp-Weaving Sections,” *Transportation Research Record 1457*, Transportation Research Board, Washington, D.C., 1995.

- 14. “Weaving Zones,” *Draft Report*, NCHRP Project 3-55(5), Viggen Corporation, Sterling, VA, 1998.

- 15. Roess, R. and Ulerio, J., “Weaving Area Analysis in the HCM 2000,” *Transportation Research Record*, Transportation Research Board, Washington, D.C., 2000.

- 16. Roess, R. and Ulerio, J., “Capacity of Ramp- Freeway Junctions,” *Final Report*, Polytechnic University, Brooklyn, NY, 1993.

Problems

1. 29-1. Consider the pair of ramps shown in Figure 29.12. It may be assumed that there is no ramp-to-ramp flow.

Figure 29.12: Weaving Segment for Problem 29-1

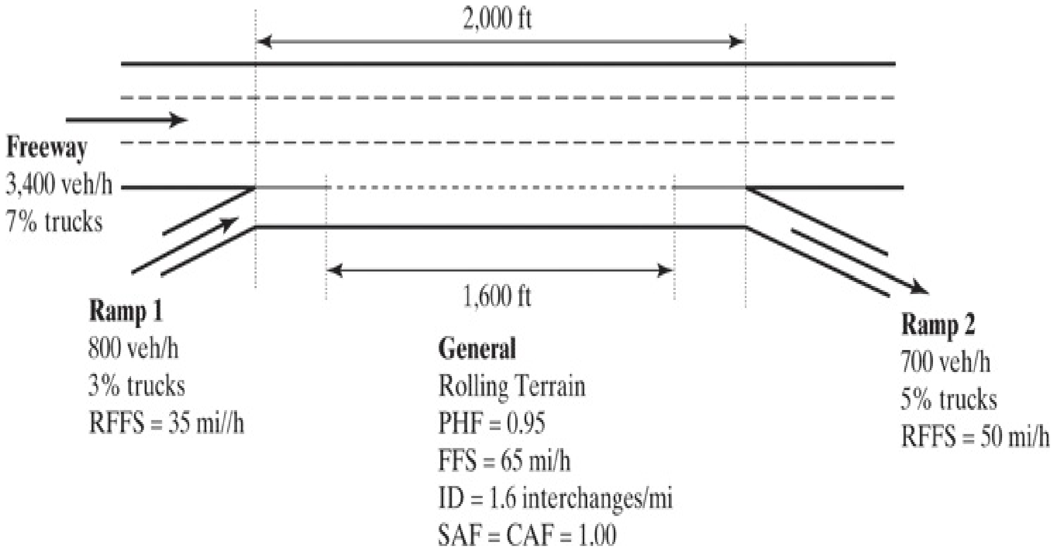


Figure 29.12: Full Alternative Text

Given the existing demand volumes and other prevailing conditions, at what level of service is this section expected to operate? If problems exist, which elements appear to be causing the difficulty?

2. 29-2. Consider the weaving area in Figure 29.13. All demands are shown as flow rates in pc/h under equivalent base conditions.

Figure 29.13: Weaving Segment for Problem 29-2

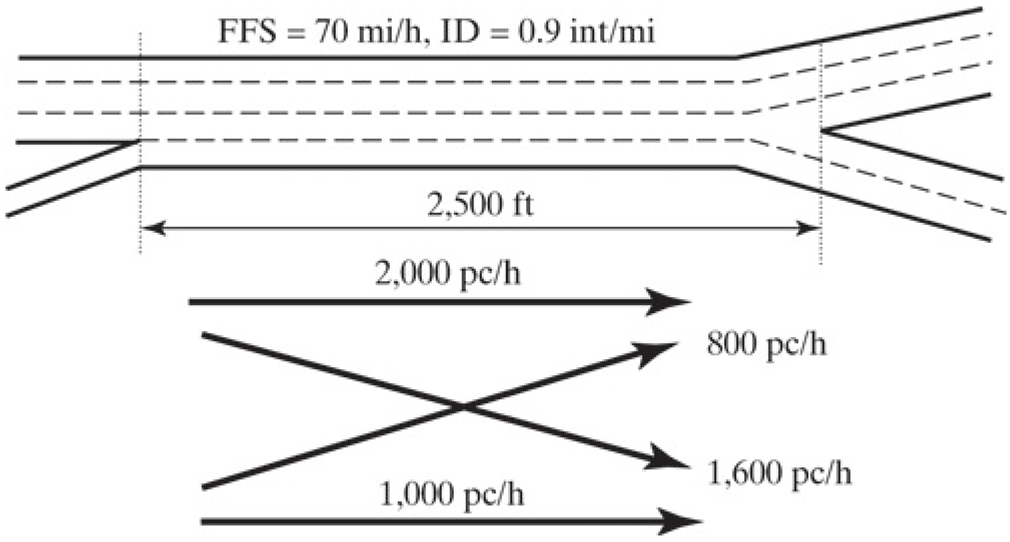
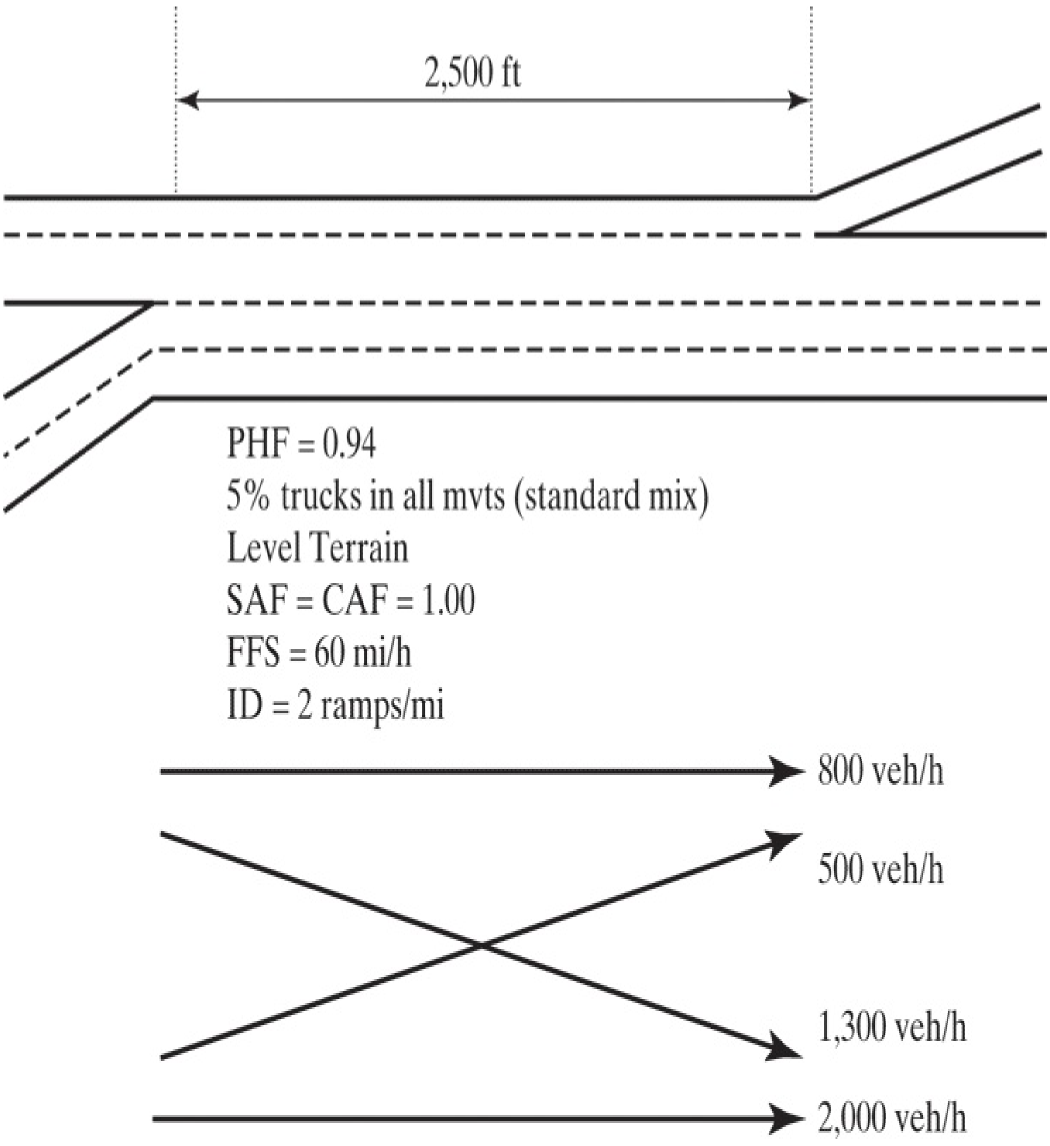


Figure 29.13: Full Alternative Text

1. Describe the critical characteristics of the segment.
2. What is the expected level of service for these conditions?
3. What is the capacity of the weaving section under equivalent ideal conditions?
4. If all demands include 10% trucks in rolling terrain, and all drivers are assumed to be familiar with the facility, and the PHF=0.92, what is the capacity of the segment under prevailing conditions?

3. 29-3. Consider the weaving segment shown in Figure 29.14. For the demands shown, what is the expected level of service? What is the capacity of the weaving segment? What improvements, if any, would you recommend? You may assume that standard conditions apply, and the SAF=CAF=1.00.

Figure 29.14: Weaving Segment for Problem 29-3



.

Figure 29.14: Full Alternative Text

4. 29-4. The weaving segment shown in Figure 29.15 is located on a C-D roadway as part of a freeway interchange. For the demands shown, find the expected level of service and capacity of the segment. What improvements, if any, would you recommend? Standard conditions apply, that is, SAF=CAF=1.00.

Figure 29.15: Weaving Segment (C-D Roadway)

for Problem 29-4

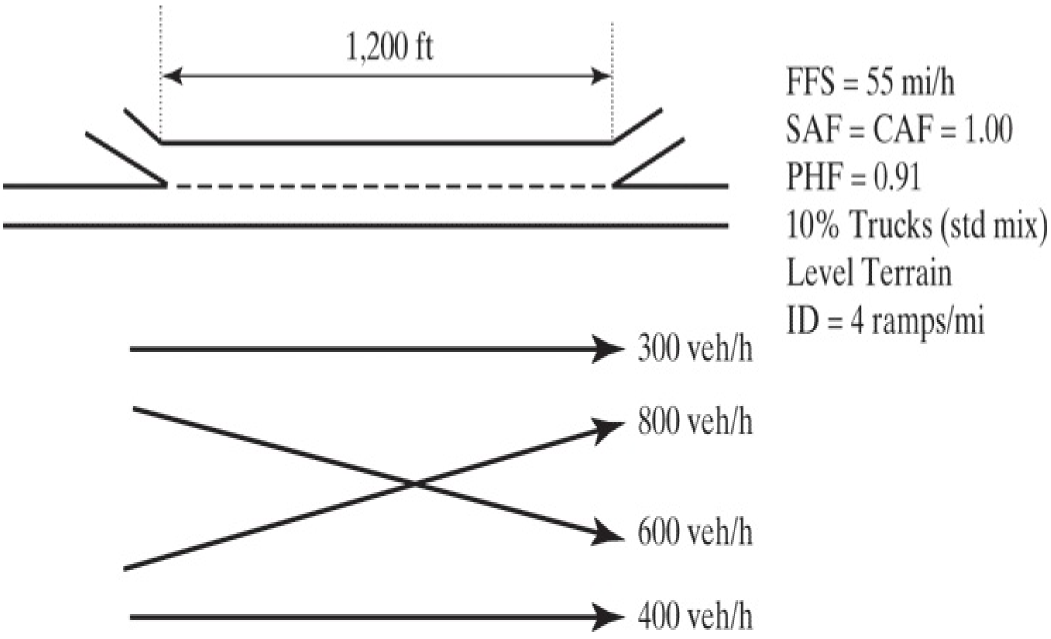


Figure 29.15: Full Alternative Text

Chapter 30 Capacity and Level of Service Analysis: Merge and Diverge Segments on Freeways and Multilane Highways

In Chapter 28, methodologies were presented to analyze basic segments on freeways and multilane highways. Chapter 29 focused on the analysis of weaving segments on these facilities, and featured a methodology that specifically considered the characteristics of turbulence, and how that turbulence affected traffic operations in them. This chapter focuses on two additional types of turbulence areas on freeways and multilane highways: merging and diverging segments. Such segments often involve on-ramps and/or off-ramps, but merging and diverging movements may also occur at major merge and diverge points in the absence of ramps.

The procedures presented in Chapter 29 for weaving segments specifically quantified “turbulence” in terms of the lane-changing behavior of weaving and nonweaving vehicles traversing the segment. For merge and diverge segments, current methodologies do not specifically account for lane-changing. Rather, models predict the macroscopic outcomes caused by turbulence in terms of speeds and densities. No specific predictions of lane-changing activity, however, are produced. This fundamental difference is primarily the result of how and when the original studies for these methodologies occurred. The methodologies of this chapter resulted from a 1993 national study of ramp operations [1]. The basic research on weaving segments took place over a decade later, and benefited from more sophisticated data collection and reduction technologies than were available in 1993.

30.1 Level-of-Service Criteria

The measure of effectiveness for merging, and diverging, segments is density. This is consistent with freeway and multilane highway methodologies. Level-of-service criteria are shown in Table 30.1.

Table 30.1: Level of Service Criteria for Merge and Diverge Segments

Level of Service	Density (pc/mi/ln)
A	0–10
B	> 10–20
C	>20–28
D	> 28–35
E	> 35
F	*Demand Exceeds Capacity*

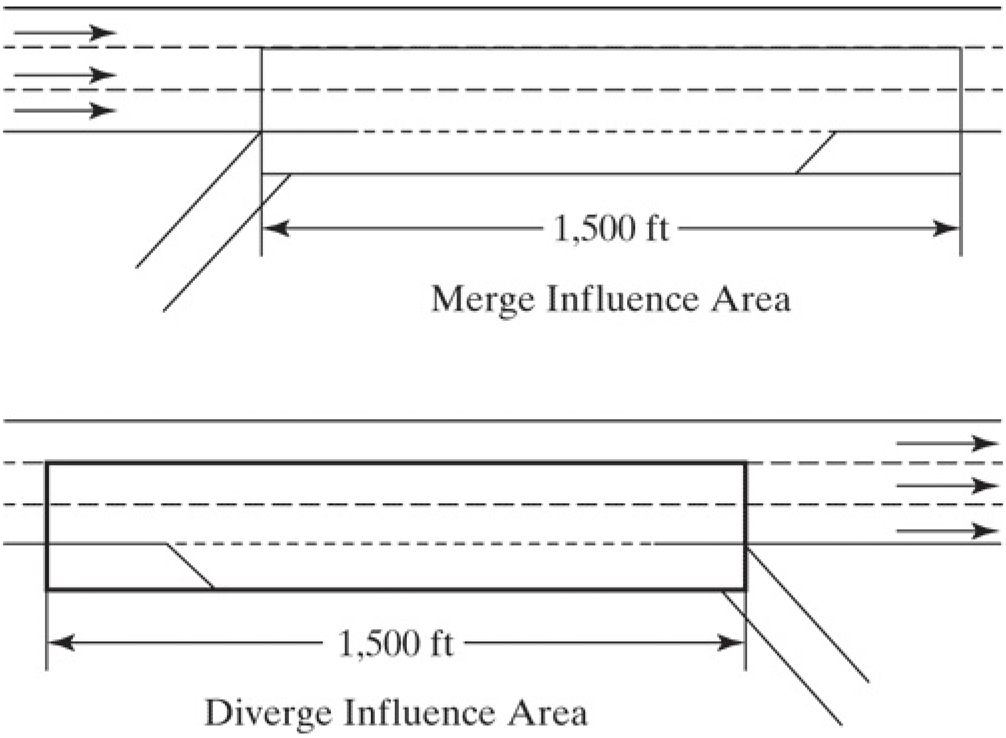
(Source: Reprinted with permission from *Highway Capacity Manual, 6th Edition: A Guide for Multimodal Mobility Analysis*, Transportation Research Board, National Academy of Sciences, Courtesy of the National Academies Press, Washington, D.C., 2016.)

Table 30.1: Full Alternative Text

As was the case for weaving segments, LOS F occurs when demand exceeds capacity of the merge or diverge segment, that is, when v/c>1.00. The limit of LOS E is defined as the capacity of the segment.

For merge and diverge areas, predicted densities reflect the “merge/diverge influence area," which consists of lanes 1 and 2 (right and next-to-right lanes of the freeway) and the acceleration or deceleration lane for a distance 1,500 feet upstream of a diverge or 1,500 feet downstream of a merge. These influence areas are illustrated in Figure 30.1.

Figure 30.1: Influence Areas for Merge and Diverge Segments



(Source: Reprinted with permission from *Highway Capacity Manual, 6th Edition: A Guide for Multimodal Mobility Analysis*, Transportation Research Board, National Academy of Sciences, Courtesy of the National Academies Press, Washington, D.C., 2000.)

Figure 30.1: Full Alternative Text

Note that Figure 30.1 focuses on merge and diverge segments created by one-lane, right-hand on- and off-ramps. The base methodology presented in this chapter is calibrated for these cases. Other merge and diverge configurations are treated as special cases, and will involve modifications to the base methodology.

In some cases, these definitions cause overlaps of more than one influence area. For example, if an on-ramp is followed by an off-ramp less than 3,000 feet away, the two 1,500 feet influence areas will at least partially overlap. In such cases, the worst density or level of service is applied to the overlap area. Other overlaps between ramp and weaving segments and/or basic segments are similarly treated: The worst applicable operating prediction applies.

30.2 Converting Demand Volumes

Procedures for merge and diverge segments rely on algorithms calibrated in terms of demand flow rates in passenger car units for base or ideal conditions. Thus, as in the weaving segment procedures of Chapter 29, volumes are converted as:

$v_i = \frac{V_i}{PHF \times f_{HV}}$ [30-1]

where:

v_i=flow rate for component “i,” pc/h,V_i=demand volume for component “i”
hour factor, andf_{HV}=heavy-vehicle adjustment factor.

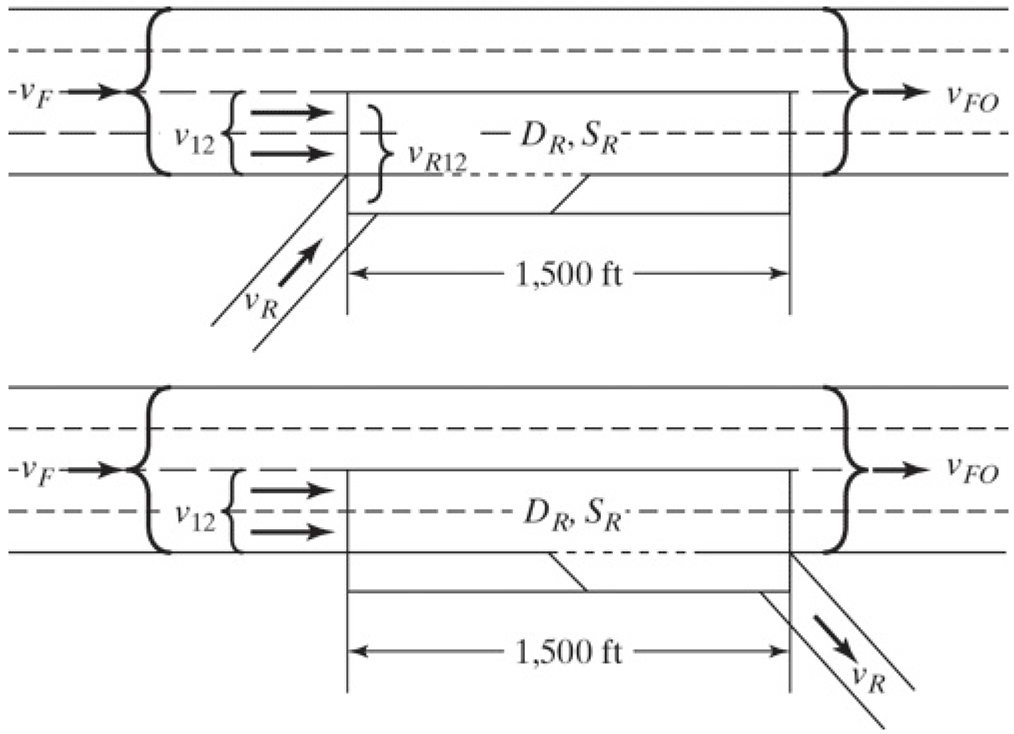
The heavy-vehicle adjustment factor is the same one used for basic freeway and multilane highway segments. They are found using the methods and values presented in Chapter 28.

30.3 Fundamental Variables Involved in Merge and Diverge Segment Analysis

As illustrated in Figure 30.1, analysis procedures for merge and diverge areas focus on the merge or diverge influence area that encompasses lanes 1 and 2 (shoulder and adjacent) freeway lanes and the acceleration or deceleration lane for a distance of 1,500 feet upstream of a diverge point or 1,500 downstream of a merge point.

Analysis procedures provide algorithms for estimating the density in these influence areas. Estimated densities are compared to the criteria of Table 30.1 to establish the level of service. Because the analysis of merge and diverge areas focuses on influence areas including only the two right-most lanes of the freeway, a critical step in the methodology is the estimation of the lane distribution of traffic immediately upstream of the merge or diverge. Specifically, a determination of the approaching demand flow remaining in lanes 1 and 2 immediately upstream of the merge or diverge is required. Figure 30.2 shows the key variables involved in the analysis methodology.

Figure 30.2: Critical Variables in Merge and Diverge Analysis



(Source: Reprinted with permission from *Highway Capacity Manual, 6th Edition: A Guide for Multimodal Mobility Analysis*, Transportation Research Board, National Academy of Sciences, Courtesy of the National Academies Press, Washington, D.C., 2016.)

Figure 30.2: Full Alternative Text

The variables included in Figure 30.2 are defined as follows:

- v_F= freeway demand flow rate immediately upstream of merge or diverge junction, in pc/h under equivalent base conditions,
- v_{12}= freeway demand flow rate in lanes 1 and 2 of the freeway immediately upstream of the merge or diverge junctions, in pc/h under equivalent base conditions,
- v_R= ramp demand flow rate, in pc/h under equivalent base conditions,
- v_{R12}= total demand flow rate entering a merge influence area, $v_R + v_{12}$, in pc/h under equivalent base conditions,
- v_{FO}= total outbound demand flow continuing downstream on the

freeway, pc/h under equivalent base conditions,

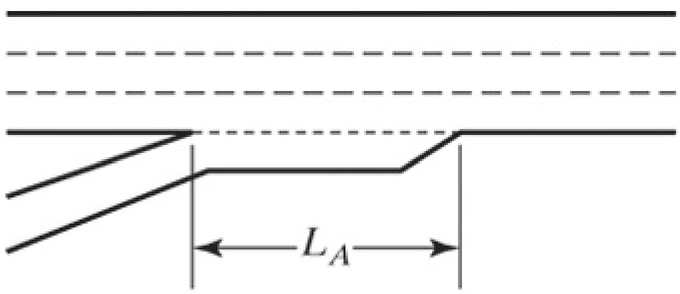
- DR= average density in the ramp influence area, pc/mi/ln, and
- SR= space mean speed of all vehicles in the ramp influence area, mi/h.

Other than the standard geometric characteristics of the facility that are used to determine its free-flow speed and adjustments to convert demand volumes in veh/h to pc/h under equivalent base conditions (Equation 30-1), there are two specific geometric variables of importance in merge and diverge analysis:

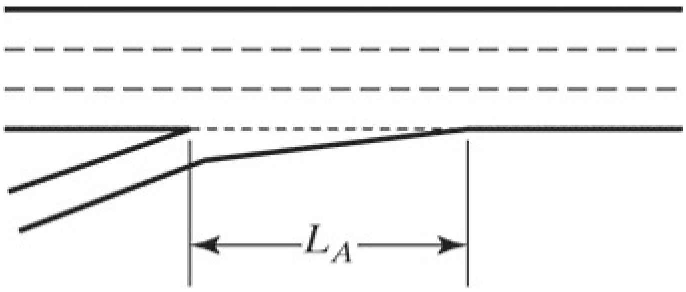
- L_a or Ld=length of the acceleration or deceleration lane, ft, and
- RFFS=free-flow speed of the ramp, mi/h.

The length of the acceleration or deceleration lane is measured from the point at which the ramp lane and lane 1 of the main facility touch to the point at which the acceleration or deceleration lane begins or ends. This definition includes the taper portion of the acceleration or deceleration lane and is the same for both parallel and tapered lanes. Figure 30.3 illustrates the measurement of length of acceleration and deceleration lanes.

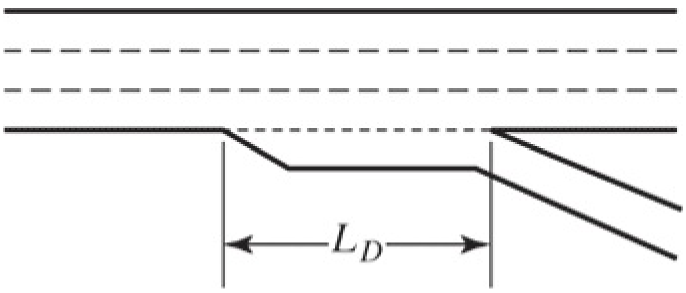
Figure 30.3: Measuring the Length of Acceleration and Deceleration Lanes



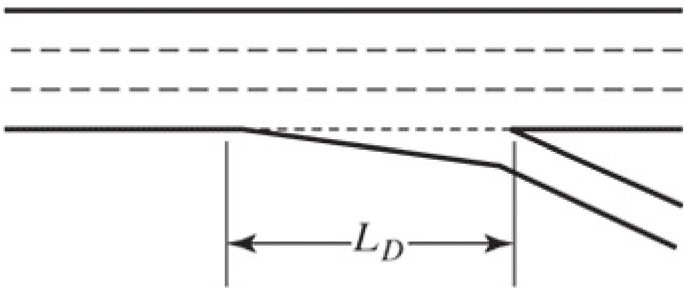
(a) Parallel Acceleration Lane



(b) Tapered Acceleration Lane



(c) Parallel Deceleration Lane



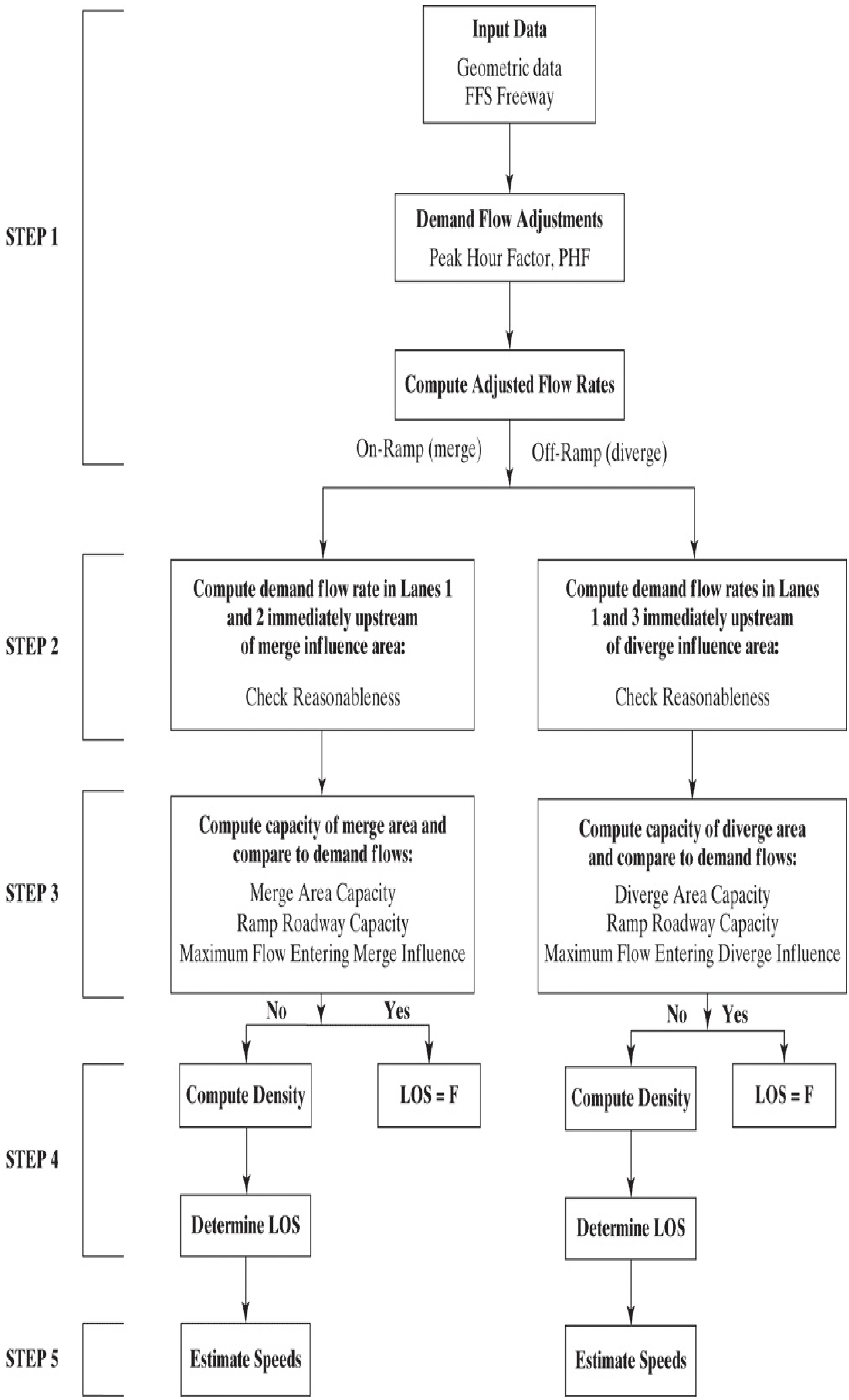
(d) Tapered Deceleration Lane

The free-flow speed of the ramp is best observed in the field but may be estimated as the design speed of the most restrictive element of the ramp. Many ramps include compound horizontal curves or a number of separate horizontal or vertical curves. The free-flow speed is generally controlled by the design speed (or maximum safe operating speed) of the most severe of these.

30.4 Computational Procedures for Merge and Diverge Segments

Figure 30.4 is a flow chart of the analysis methodology for merge or diverge junctions. It illustrates the following five fundamental steps:

Figure 30.4: Flow Chart for Analysis of Ramp-Facility Junctions


STEP 1
STEP 2
STEP 3
STEP 4
STEP 5
Input Data
Geometric data
FFS Freeway
Demand Flow Adjustments
Peak Hour Factor, PHF
Compute Adjusted Flow Rates
On-Ramp (merge)
Off-Ramp (diverge)
Compute demand flow rate in Lanes 1 and 2 immediately upstream of merge influence area:
Check Reasonableness
Compute demand flow rates in Lanes 1 and 3 immediately upstream of diverge influence area:
Check Reasonableness
Compute capacity of merge area and compare to demand flows:
Merge Area Capacity
Ramp Roadway Capacity
Maximum Flow Entering Merge Influence
Compute capacity of diverge area and compare to demand flows:
Diverge Area Capacity
Ramp Roadway Capacity
Maximum Flow Entering Diverge Influence
No
Yes
No
Yes
Compute Density
LOS = F
Compute Density
LOS = F
Determine LOS
Determine LOS
Estimate Speeds
Estimate Speeds

(Source: Reprinted with permission from *Highway Capacity Manual, 6th Edition: A Guide for Multimodal Mobility Analysis*, Transportation Research Board, National Academy of Sciences, Courtesy of the National Academies Press, Washington, D.C., 2016.)

Figure 30.4: Full Alternative Text

1. Specify all traffic and roadway data for the junction to be analyzed: peak-hour demands, PHF, traffic composition, driver population, and geometric details of the site, including the free-flow speed for the facility and for the ramp. Convert all demand volumes to flow rates in pc/h under equivalent base conditions using Equation 30-1.

2. Determine the demand flow in lanes 1 and 2 of the facility immediately upstream of the merge or diverge junction using the appropriate algorithm as specified.

3. Determine whether the demand flow exceeds the capacity of any critical element of the junction. Where demand exceeds capacity, level of service F is assigned and the analysis is complete.

4. If operation is determined to be stable, determine the density of all vehicles within the ramp influence area. Table 30.1 is then used to determine level of service based on the density in the ramp influence area.

5. If the operation is determined to be stable, determine the speed of all vehicles within the ramp influence area and across all facility lanes as secondary measures of performance.

Once all input characteristics of the merge or diverge junction are specified and all demand volumes have been converted to flow rates in pc/h under equivalent base conditions, remaining parts of the methodology may be completed.

It should be noted that the base methodology for merge and diverge segments is based upon single-lane, right-hand on- and off-ramps. There are many other types of configuration, including multilane on- and off-ramp junctions, left-hand ramps, major merge and diverge segments, and ramps on five-lane (one direction) facility segments. These are handled as “special case” which are treated later in this chapter. These cases involve logical modifications to the base methodology for each case. In few cases, substantial databases were available to calibrate these modifications. In most cases, the modifications are based upon theoretical models and informed judgment of the Highway Capacity and Quality of Service Committee (HCQSC).

30.4.1 Estimating Demand Flow Rates in Lanes 1 and 2 (Step 2)

The starting point for analysis is the determination of demand flow rates in lanes 1 and 2 (the two right-most lanes) of the facility immediately upstream of the merge or diverge junction. This is done using a series of regression-based algorithms developed as part of a nationwide study of ramp-freeway junctions [1].

Basic Algorithms

For merge areas, the flow rate remaining in lanes 1 and 2 immediately upstream of the junction is computed simply as a proportion of the total approaching facility flow:

$v12=vF \times PFM$ [30-2]

where:

PFM=proportion of approaching vehiclesremaining in lanes 1 and 2 immed

The value of PFM varies with the number of lanes on the facility, demand flow levels, the proximity of adjacent ramps (in some cases), the length of the acceleration lane (in some cases), and the free-flow speed of the ramp (in some cases.)

The general approach to estimating the demand flow rate in lanes 1 and 2 immediately upstream of a diverge is somewhat different from the one used for merge areas. This is because all of the off-ramp traffic is assumed to be in lanes 1 and 2 at this point. Thus, the flow in lanes 1 and 2 is taken as the off-ramp flow plus a proportion of the through traffic on the facility.

v12=vR + (vF − vR)PFD [30-3]

where:

PFD=proportion of through freeway vehiclesremaining in lanes 1 and 2 imı
ramp demand flow rate, pc/h, andvF=demand flow rate for approachingfree

Depending upon the specific case, the value of P_{FD} varies with the freeway flow rate, the ramp flow rate, and (in some cases) with activity on adjacent upstream and downstream ramps.

Determining Values of $\mathrm{P_{FM}}$ and $\mathrm{P_{FD}}$

A master table for determining the appropriate value of P_{FM} and/or P_{FD} is shown in Table 30.2. Table 30.3 follows with selection criteria for the various equations presented in Table 30.2.

Table 30.2: Equations for Determining the Value of P_{FM} and P_{FD}

No. of Freeway Lanes*	$P_{FM} =$	Eqn No.	$P_{FD} =$	Eqn No.
4	1.000	30-4	1.000	30-10
6	$0.5775 + 0.000028L_a$	30-5	$0.760 - 0.000025v_F - 0.000046v_R$	30-11
	$0.7289 - 0.0000135(v_F + v_R) - 0.003296RFFS + 0.000063L_{UP}$	30-6	$0.717 - 0.000039v_F + 0.604(v_u/L_{UP})$**	30-12
	$0.5487 + 0.2628(v_D/L_{DN})$	30-7	$0.616 - 0.000021v_F + 0.124(v_D/L_{DN})$	30-13
8	$0.2178 - 0.000125v_R + 0.01115(L_a/RFFS)$ if $v_F/RFFS \leq 72$	30-8	0.436	30-14
	$0.2178 + 0.000125v_R$ if $v_F/RFFS > 72$	30-9		

*Total number. Lanes per direction is half value shown, that is, 8-lane freeway has 4 lanes in each direction.

vD=flow rate on adjacent downstream ramp (pc/h); vU=flow rate on adjacent upstream ramp (pc/h); LUP=distance to upstream adjacent ramp (ft); LDN=distance to downstream adjacent ramp (ft); all other variables as previously defined.

**Equation applies only when v_u/L_UP is less than or equal to 0.20; if not, use Eqn 30-11 instead.

Table 30.2: Full Alternative Text

Table 30.3: Selecting the Appropriate Equation from Table 30.2 for 6-Lane Freeways

Adjacent Upstream Ramp	*Subject Ramp*	Adjacent Downstream Ramp	Use Equation(s)
None		None	30-5
None		On	30-5
None		Off	30-7 or 30-5
On		None	30-5
Off	**On**	None	30-6 or 30-5
On		On	30-5
On		Off	30-7 or 30-5
Off		On	30-6 or 30-5
Off		Off	30-6, 30-7, or 30-5
None		None	30-11
None		On	30-11
None		Off	30-13 or 30-11
On	**Off**	None	30-12 or 30-11
Off		None	30-11
On		On	30-12 or 30-11
On		Off	30-12, 30-13, or 30-11

Table 30.3: Full Alternative Text

The selection of an appropriate equation to compute an applicable value of P_{FM} and/or P_{FD}, at first glance, appears to be quite complex. The equations reflect a number of critical characteristics:

- For four-lane freeways (two lanes in each direction), the equation is trivial, as only lanes 1 and 2 exist, and all demand must be in them. The value of either P_{FM} or P_{FD} is, therefore, 1.00.

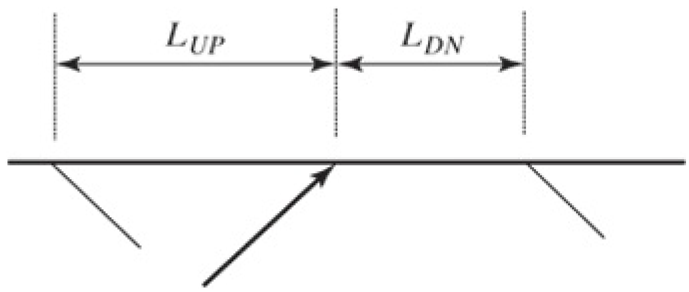
- For six- and eight-lane facilities, it is believed that the flow remaining in lanes 1 and 2 is dependent upon the distance to and flow rate on adjacent upstream and downstream ramps. A driver entering the facility on a nearby upstream on-ramp is more likely to remain in lanes 1 and 2 if the distance between ramps is insufficient to allow the driver to make two lane changes to reach outer lanes. Likewise, a driver knowing he or she has to exit at a nearby downstream off-ramp is more likely to move into lanes 1 and 2 than a driver proceeding downstream on the main facility. Although these are logical expectations, the database on ramp junctions was sufficient to establish these relationships for only a few scenarios on six-lane freeways.

- For six-lane freeways, it was possible to model the impact of adjacent ramps in four cases: (a) the impact of an adjacent upstream off-ramp on a subject on-ramp; (b) the impact of an adjacent downstream off-ramp on a subject on-ramp; (c) the impact of an adjacent upstream on-ramp on a subject off-ramp; and (d) the impact of an adjacent downstream off-ramp on a subject off-ramp. It *was not* possible to define the impact of upstream or downstream on-ramps on a subject on-ramp, or of upstream off-ramps or downstream on-ramps on a subject off-ramp.

- For eight-lane freeways, the database for the methodology did not allow a statistically definitive statement of the effects of adjacent ramps. Therefore, for eight-lane freeways, no such equations are provided.

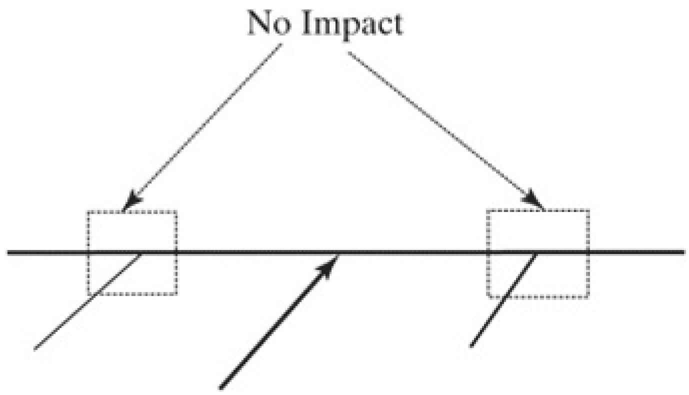
Table 30.2 can be used directly for any merge or diverge analysis on a four-lane or eight-lane freeway, as only one equation applies for each case. Table 30.3 is needed to select the appropriate equation for any merge or diverge analysis on a six-lane freeway. Figure 30.5 further illustrates the choices that need to be made for six-lane freeways.

Figure 30.5: Impact of Adjacent Ramps on 6-Lane Freeways Illustrated

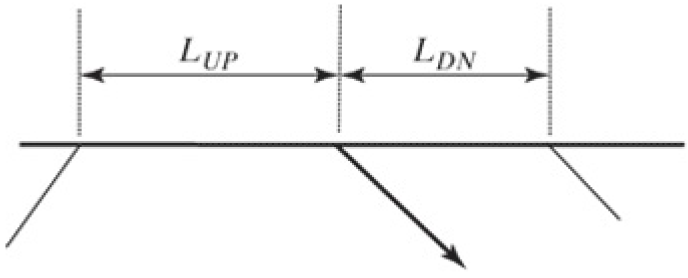


(a) On-Ramp w/ Adj. Upstream Off-Ramp and Adj.

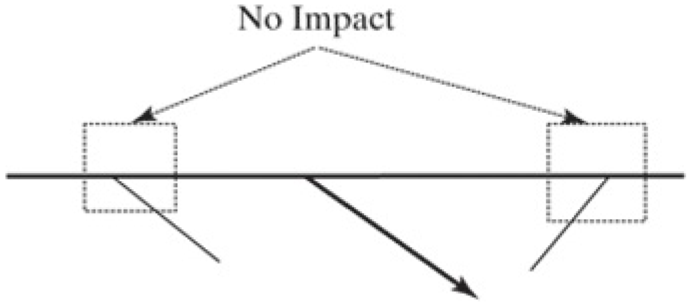
Downstream Off-Ramp



(b) On-Ramp w/ Adj. Upstream On-Ramp and Adj. Downstream On-Ramp



(c) Off-Ramp w/ Adj. Upstream On-Ramp and Adj. Downstream Off-Ramp



(d) Off-Ramp w/ Adj. Upstream Off-Ramp and Adj. Downstream On-Ramp

Figures 30.5(a) and (c) depict the configurations for which equations taking into account adjacent upstream and/or downstream ramps are provided. Figures 30.5(b) and (d) depict configurations in which there are no equations that account for the impact of adjacent upstream and/or downstream ramps. This does not mean to imply that there are no such impacts for configurations lacking specific equations, nor, for that matter, for ramps on eight-lane freeways. It simply means that the research used to calibrate these equations did not contain sufficient data to statistically define these impacts.

The Equivalence Distance on Six-Lane Freeways

Even for configurations illustrated in Figure 30.5(a) and (c), it is not certain that a specific impact of adjacent ramp traffic can be anticipated. For these cases, the distance to the adjacent upstream (L_{UP}) or adjacent downstream (L_{DN}) is critical. At some distance, the ramps are so far apart that the impact of adjacent ramp traffic is no longer measurable.

For on-ramps on six-lane freeways, the equation for isolated ramps is Equation 30-5, while Equation 30-11 is used for isolated off-ramps. These equations also serve as defaults for all other configurations where no other equation properly applies.

Table 30.3 shows that for some configurations, more than one equation may apply. For an on-ramp with an adjacent upstream off-ramp, Equation 30-6 or 30-5 is indicated. The appropriate equation for this and other configurations where a choice is indicated depends upon the equivalence distance, L_{EQ}, between the subject and adjacent ramps. The equivalence distance is defined as the distance at which the two equations yield the same value of P_{FM} or P_{FD}.

Because there are four equations in Table 30.2 that consider the specific impacts of adjacent upstream or downstream ramps, there are four equations for the equivalence distance, L_{EQ}, as shown in Table 30.4.

Table 30.4: Equations for Equivalence Distance on 6-Lane Freeways

Configuration	Selecting Between Equations	Equation	Equation Number
On-Ramp w/Upstream Off-Ramp	30-5 30-6	$L_{EQ} = 0.214(v_F + v_R) + 0.444\, L_a + 53.32\, RFFS - 2{,}403$	30-15
On-Ramp w/Downstream Off-Ramp	30-5 30-7	$L_{EQ} = \dfrac{v_D}{0.1096 + 0.000107\, L_a}$	30-16
Off-Ramp w/Upstream On-Ramp	30-11 30-12	$L_{EQ} = \dfrac{v_U}{0.071 + 0.000023\, v_F - 0.000076\, v_R}$	30-17
Off-Ramp w/Downstream Off-Ramp	30-11 30-13	$L_{EQ} = \dfrac{v_D}{1.15 - 0.000032\, v_F - 0.000369\, v_R}$	30-18

Table 30.4: Full Alternative Text

In each case, the selection of the appropriate equation is based upon the comparison of the actual distance between the ramps and the equivalence distance:

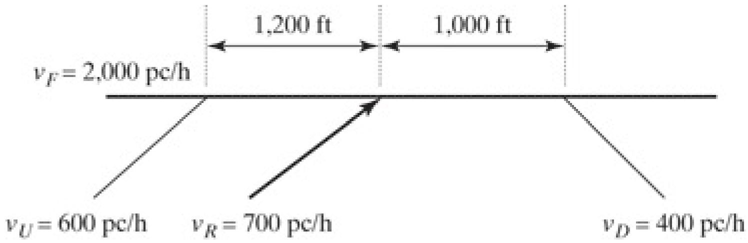
- If L_{UP} or LDN≥LEQ, use the base equation for isolated ramps (30-5 for on-ramps, 30-11 for off-ramps).
- If L_{UP} or LDN<LEQ, use the configuration specific equation (30-6 or 30-7 for on-ramps, 30-12 or 30-13 for off-ramps).

It is also possible that a subject ramp may be in the range affected by *both* an upstream adjacent ramp *and* a downstream adjacent ramp. These procedures do not allow simultaneous consideration of three-ramp sequences. Thus, two solutions are necessary: one in conjunction with the adjacent upstream ramp and one in conjunction with the adjacent downstream ramp. The solution that produces the highest value of P_{FM} or P_{FD} is the one that is used.

Sample Problem 30-1: Selecting Equations for P_{FM} or P_{FD}

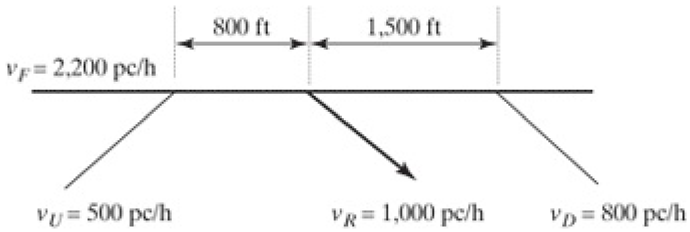
The process of selecting an appropriate equation is best explained by example. Consider the two ramp sequences illustrated in Figure 30.6. Both sequences are on six-lane freeways. The length of the acceleration or deceleration lanes on all ramps is 400 feet.

Figure 30.6: Two Example Ramp Sequences



(a) On-Ramp Sequence

30.4-5 Full Alternative Text



(b) Off-Ramp Sequence

30.4-5 Full Alternative Text

For ramp sequence (a), an on-ramp has an upstream adjacent on-ramp and a downstream adjacent off-ramp. From Table 30.3, the upstream on-ramp does not affect the subject ramp, and the general default Equation 30-5 is used to compute P_{FM}. The downstream off-ramp, however, may affect the subject ramp: Equation 30-7 or 30-5 may apply. The determination must be based upon the equivalence distance, L_{EQ}. From Table 30.4, the equivalence distance is found using Equation 30-16:

$$L_{EQ}=\frac{v_D}{0.1096+0.000107\,L_a}=\frac{400}{0.1096+(0.0000107\times 400)}=3{,}512\text{ft}>1{,}000\text{ft}$$

Because the actual distance to the downstream ramp is *less* than the equivalence distance, the special equation accounting for the downstream ramp is used: Equation 30-7.

Therefore, the flow rate remaining in lanes 1 and 2 immediately upstream of the subject off-ramp is determined by *both* Equations 30-5 and 30-7. The equation yielding the highest value of P_{FM} determines the result.

In the second sequence (Figure 30.6(b)) the off-ramp has an adjacent upstream on-ramp and an adjacent downstream off-ramp. From Table 30.3, considering the upstream ramp, the flow remaining in lanes 1 and 2 can be estimated using Equation 30-11 or 30-12. Again, the selection between the two depends on the equivalence distance. From Table 30.4, Equation 30-17 is used to estimate the equivalence distance:

$$L_{EQ}=\frac{v_U}{0.071+0.000023\,v_F-0.000076\,v_R}=\frac{500}{0.071+(0.000023\times 2200)-(0.000076\times 1000)}=10{,}870\text{ ft}>800\text{ ft}$$

Because the actual distance to the upstream on-ramp is 800 feet, considerably less than the equivalence distance, the equation that includes the impact of the upstream ramp is used: Equation 30-12.

If the downstream off-ramp is considered, the value of P_{FD} may be computed using either Equation 30-11 or 30-13. Once again, the equivalence distance must be examined. From Table 30.4, Equation 30-18 is used to find the equivalence distance in this case:

$$L_{EQ}=\frac{v_D}{1.15-0.000032\,v_F-0.000369\,v_R}$$

Here, it is necessary to be careful about what values are used in this equation. Several are straightforward: v_D is given as 800 pc/h; v_R is given as 1,000 pc/h. The freeway flow rate, v_F, is taken at a point just upstream of the subject ramp. For this case, it includes the 2,200 pc/h entering on the freeway *plus* the 500 pc/h entering on the upstream on-ramp. At the point just upstream of the subject ramp, $v_F=2200+500=2{,}700$ pc/h. Then:

$$L_{EQ}=\frac{800}{1.15-(0.000032\times 2{,}700)-(0.000369\times 1{,}000)}=1{,}152\text{ ft}<1{,}500\text{ft}$$

Because the actual distance is *more* than the equivalence distance, the general default Equation 30-11 is used. We once again have two potential solutions. Equation 30-12 is used when the upstream on-ramp is considered. Equation 30-11 is used when the downstream off-ramp is considered. The result used will be the higher of the two values of P_{FD}.

Obviously, these solutions sometimes become very detailed. To find the appropriate values of P_{FM} or P_{FD}, the correct equations must be selected. In this case, each of the subject ramps involved two potential solutions, and worst prediction was the result taken.

More complete sample problems toward the end of the chapter will further illustrate how these selections are made, and how those selections affect the results of merge and diverge segment analyses.

Computing v_{12} Immediately Upstream of a Subject Ramp

Once the appropriate value(s) of P_{FM} and/or P_{FD} are computed, Equations 30-2 (merge segments) and 30-3 (diverge segments) are applied directly to estimate flow in lanes 1 and 2 of the freeway.

Checking the “Reasonableness” of Results

Once the flow rate for lanes 1 and 2 have been predicted, it is necessary to subject the results to a “reasonableness” check. Because the algorithms used are regression-based, results can occasionally lead to illogical lane distributions. This can occur where site conditions are near or outside the boundaries of the calibration database used in the regression. There are two conditions that the estimated lane distribution must meet:

- Average flow rate in the outer lanes may not exceed 2,700 pc/h/ln.
- Average flow rate in the outer lanes may not be more than 1.5 times the average flow rate in lanes 1 and 2.

Obviously, the size of the freeway determines the number of outer lanes. For four-lane freeways (two lanes in each direction), there are no outer lanes, and all vehicles approach in lanes 1 and 2. For six-lane freeways (three lanes in each direction), there is one outer lane (lane 3). For eight-lane freeways (four lanes in each direction), there are two outer lanes (lanes 3 and 4).

If either or both of these criteria are violated by the predicted lane distribution, the flow rate in lanes 1 and 2 must be adjusted to accommodate these limits. If the average flow rate in outer lanes exceeds 2,700 pc/h/ln, it is set at 2,700 pc/h/ln, and the flow rate in lanes 1 and 2 is recomputed as:

v12=vF−2700 NO [30-19]

where N_O is the number of outer lanes. If the average flow rate in outer lanes exceeds 1.5 times the average flow rate in lanes 1 and 2, the outer lane flow is set at 1.5 times the average flow in lanes 1 and 2, and the flow rate in lanes 1 and 2 is recomputed as:

For NO=1: v12=vF1.75For NO=2: v12=vF2.50For NO>2v12=2 vF(1.5 N [30-20]

In cases where both limitations are violated, the revision that meets *both* criteria is used.

30.4.2 Estimating the Capacity of the Merge or Diverge Segment (Step 3)

The analysis procedure for merge and diverge areas determines whether the segment in question has failed (LOS=F) based upon a comparison of demand flow rates to critical capacity values.

In general, the basic capacity of the facility is not affected by merging or diverging activities. Because of this, the basic facility capacity must be checked immediately upstream and/or downstream of the merge or diverge. Ramp roadway capacities must also be examined for adequacy. When demand flows exceed any of these capacities, a failure is expected, and the level of service is determined to be F.

The total flow entering the ramp influence area is also checked. Although a maximum desirable value is set for this flow, exceeding it does not imply level of service F if no other capacity value is exceeded. In cases where only this maximum is violated, expectations are that service quality will be less than that predicted by the methodology. Capacity values are given in Table 30.5.

Table 30.5: Capacity Values for Ramp Checkpoints

FFS (mi/h)	Capacity of Upstream/Downstream Freeway Segment[a]				Max. Desirable Flow Rate (v_{R12}) Entering Merge Influence Area[b]	Max. Desirable Flow Rate (v_{12}) Entering Diverge Influence Area[b]
	No. of Lanes in One Direction					
	2	3	4	>4		
≥70	4,800	7,200	9,600	2,400/ln	4,600	4,400
65	4,700	7,050	9,400	2,350/ln	4,600	4,400
60	4,600	6,900	9,200	2,300/ln	4,600	4,400
55	4,500	6,750	9,000	2,250/ln	4,600	4,400

[a]Demand in excess of these capacities results in LOS F.

[b]Demand in excess of these values alone does not result in LOS F; operations may be worse than predicted by this methodology.

Table 30.5: Full Alternative Text

FFS (mi/h)	Capacity of Upstream/Downstream Multilane Highway or C-D Segment[a] No. of Lanes in One Direction: 2	3	>3	Max. Desirable Flow Rate (v_{R12}) Entering Merge Influence Area[b]	Max. Desirable Flow Rate (v_{12}) Entering Diverge Influence Area[b]
≥60	4,400	6,600	2,200/ln	4,600	4,400
55	4,200	6,300	2,100/ln	4,600	4,400
50	4,000	6,000	2,000/ln	4,600	4,400
45	3,800	5,700	1,900/ln	4,600	4,400

[a]Demand in excess of these capacities results in LOS F.

[b]Demand in excess of these values alone does not result in LOS F; operations may be worse than predicted by this methodology.

30.4-6 Full Alternative Text

Ramp FFS S_{FR} (mi/h)	Capacity of Ramp Roadway: Single-Lane Ramps	Two-Lane Ramps
>50	2,200	4,400
>40–50	2,100	4,200
>30–40	2,000	4,000
≥20–30	1,900	3,800
<20	1,800	3,600

Note: Capacity of a ramp roadway does not ensure an equal capacity at its freeway or other high-speed junction. Junction capacity must be checked against criteria in this table.

(Source: Reprinted with permission from *Highway Capacity Manual, 6th Edition: A Guide for Multimodal Mobility Analysis*, Transportation Research Board, National Academy of Sciences, Courtesy of the National Academies Press, Washington, D.C., 2016.)

30.4-7 Full Alternative Text

The freeway and multilane capacity values shown are the same as those for basic freeway sections used in Chapter 28. They are repeated here for convenience. Note that in Chapter 28, capacities were shown as *per lane* values. Table 30.5 multiplies these by the appropriate number of lanes (in one direction), again, for ease of use. Other values shown in Table 30.5 may be approximately applied to merging or diverging multilane highway segments.

The specific checkpoints that should be compared to the capacity criteria of Table 30.5 may be summarized as follows:

- For merge areas, the maximum facility flow occurs downstream of the merge. Thus, the facility capacity is compared with the downstream facility flow (vFO=vF + vR).
- For diverge areas, the maximum facility flow occurs upstream of the diverge. Thus, the facility capacity is compared to the approaching upstream facility flow, vF.
- Where lanes are added or dropped at a merge or diverge, both the upstream (vF) and downstream (vFO) facility flows must be compared to capacity criteria.
- For merge areas, the flow entering the ramp influence area is vR12=v12 + vR. This sum is compared to the maximum desirable flow indicated in Table 30.5.
- For diverge areas, the flow entering the ramp influence area is v12, as the off-ramp flow is already included. It is compared directly with the maximum desirable flow indicated in Table 30.5.
- All ramp flows, vR, must be checked against the ramp capacities given in Table 30.5.

The ramp capacity check is most important for diverge areas. Diverge segments rarely fail unless the capacity of one of the diverging legs is exceeded by the demand flow. This is most likely to happen on the off-ramp. It should also be noted that the capacities shown in Table 30.5 for two-lane ramps may be quite misleading. They refer to the ramp roadway itself, not to the junction with the main facility. There is no evidence, for example, that a two-lane on-ramp junction can accommodate any greater flow than a one-lane junction. It is unlikely that a two-lane on-ramp can handle more than 2,250 to 2,400 pc/h through the merge area. For higher on-ramp demands, a two-lane on-ramp would have to be combined with a lane addition at the facility junction.

The capacities in Table 30.5 are given in pc/h for ideal or base conditions. They can, of course, be converted to veh/h using the heavy-vehicle adjustment factor, as in Chapters 28 and 29. However, since all demand volumes have already been converted to flow rates in pc/h, it is convenient to directly compare the converted demand flow rates to the ideal capacities determined from Table 30.5. It should also be noted that a capacity adjustment factor (CAF) may be applied to any of the capacities of Table 30.5 (see Chapter 28 for a full discussion of these).

If the upstream freeway flow, downstream freeway flow, or ramp flow exceeds capacity, then LOS F is assigned to the merge or diverge segment, and the analysis ends. If demand flows are less than capacities for each of these elements, the analysis continues to determine the appropriate level of service.

30.4.3 Determining Density and Level of Service in the Ramp Influence Area (Step 4)

If all facility and ramp capacity checks indicate that stable flow prevails in the merge or diverge area, the density in the ramp influence area may be estimated using Equation 30-21 for merge areas and Equation 30-22 for diverge areas:

$$D_R = 5.475 + 0.00734 v_R + 0.0078 v_{12} - 0.00627 L_a \quad [30\text{-}21]$$
$$D_R = 4.252 + 0.0086 v_{12} - 0.009 L_d \quad [30\text{-}22]$$

where all variables have been previously defined. In both cases, the density in the ramp influence area is dependent upon the flow entering it (v_R and v_{12} for merge areas and v_{12} for diverge areas), and the length of the acceleration or deceleration lane. The density computed by Equation 30-21 or 30-22 is directly compared to the criteria of Table 30.1 to determine the expected level of service.

30.4.4 Determining Expected Speed Measures (Step 5)

Although it is not a measure of effectiveness, and the determination of an expected speed is not required to estimate density (as was the case for weaving areas), it is often convenient to have an average speed as an additional measure or as an input to system analyses. Because speed behavior in the vicinity of ramps (1,500 ft segment encompassing the ramp influence area) is different from basic sections, three algorithms are provided for merge areas and three for diverge areas as follows:

- Estimation algorithm for average speed within the ramp influence area, which includes lanes 1 and 2 and the acceleration or deceleration lane within the 1,500 ft length of the influence area.
- Estimation algorithm for average speed in outer lanes (where they exist) within the 1,500 ft boundaries of the ramp influence area.
- Algorithm for combining the above into an average space mean speed across all lanes within the 1,500 ft boundaries of the ramp influence area.

Table 30.6 summarizes these algorithms for merge areas, and Table 30.7 summarizes them for diverge areas.

Table 30.6: Estimating Speeds in Merge Segments

Average Speed In	Estimation Equation	
Ramp Influence Area	$S_R = FFS \times SAF - (FFS \times SAF - 42)\, M_S$ $M_S = 0.321 + 0.0039\, e^{(v_{R12}/1000)} - 0.002\,(L_a \times RFFS \times SAF/1000)$	[Eqns 30-29]
Outer Lanes	$S_o = FFS \times SAF \quad v_{oa} < 500\ pc/h$ $S_o = FFS \times SAF - 0.0036\,(v_{oa} - 500) \quad v_{oa} = 500 - 2{,}300\ pc/h$ $S_o = FFS \times SAF - 6.53 - 0.0006\,(v_{oa} - 2300) \quad v_{oa} > 2{,}300\ pc/h$	[Eqns 30-30]
All Lanes	$S = \dfrac{v_{R12} + v_{oa}N_o}{\left(\dfrac{v_{R12}}{S_R}\right) + \left(\dfrac{v_{oa}N_o}{S_o}\right)}$	[Eqn 30-31]

Table 30.6: Full Alternative Text

Table 30.7: Estimating Speeds in Diverge Segments

Average Speed In	Estimation Equation	
Ramp Influence Area	$S_R = FFS \times SAF - (FFS \times SAF - 42)\, D_S$ $D_S = 0.883 + 0.00009\, v_{12} - 0.013\, RFFS \times SAF$	[Eqns 30-32]
Outer Lanes	$S_o = 1.097FFS \times SAF \quad v_{oa} < 1{,}000\ pc/h$ $S_o = 1.097FFS \times SAF - 0.0039\,(v_{oa} - 1000) \quad v_{oa} \geq 1{,}000\ pc/h$	[Eqns 30-33]
All Lanes	$S = \dfrac{v_{12} + v_{oa}N_o}{\left(\dfrac{v_{12}}{S_R}\right) + \left(\dfrac{v_{oa}N_o}{S_o}\right)}$	[Eqn 30-34]

Table 30.7: Full Alternative Text

Most of the variables in Tables 30.6 and 30.7 are as previously defined. For the variables appearing here for the first time:

SR=average speed of vehicles within the rampinfluence area, mi/h,So=aver

It should be noted that where there are only two lanes on the freeway mainline (a four-lane freeway), then all vehicles are in the ramp influence area, and only the first equations in Tables 30.6 and 30.7 are needed.

30.4.5 Final Comments on the Base Procedure

As noted at the beginning of the discussion of computational procedures, the methodology presented applies directly only to the base case of a single-lane on- or off-ramp on the right-hand side of a freeway or multilane highway. Because there are many merge and diverge configurations that do not conform to these conditions, there are a variety of “special cases” which are discussed in the next section.

30.5 Special Cases in Merge and Diverge Analysis

Merge and diverge analysis procedures were calibrated primarily for single-lane, right-hand on- and off-ramps. Modifications have been developed so that a broad range of merge and diverge geometries can be analyzed using these procedures. These “special applications” include the following:

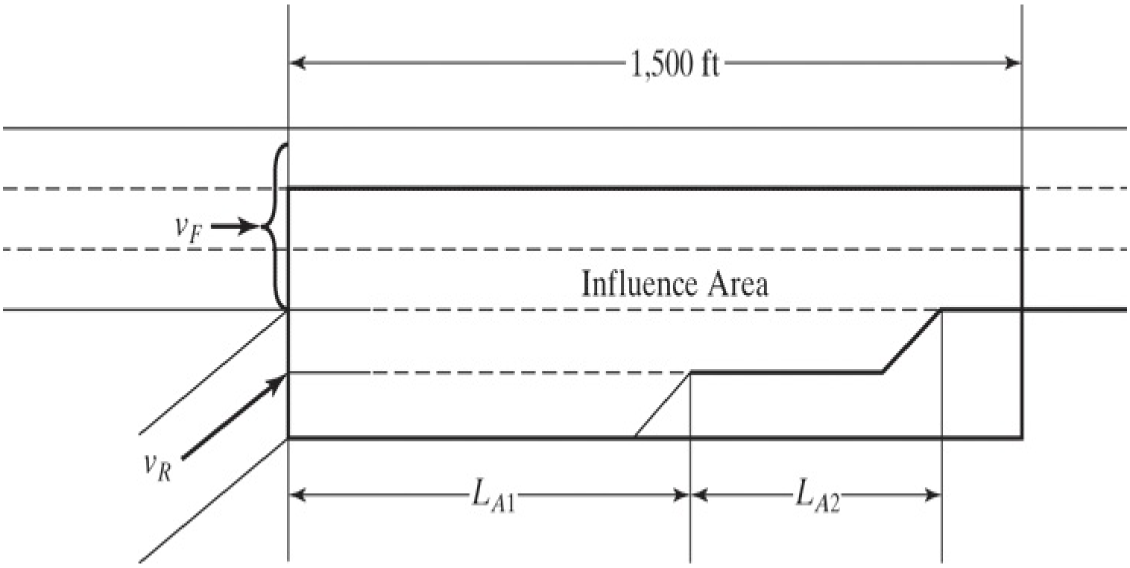
- Two-lane, right-hand on- and off-ramps
- On- and off-ramps on five-lane (one direction) freeway sections
- One-lane, left-hand on- and off-ramps
- Major merge and diverge areas
- Lane drops and lane additions

The category of “major merge and diverge areas” is an extremely broad one that encompasses virtually any merge or diverge configuration that is not the base case or one of the other special applications listed.

30.5.1 Two-Lane On-Ramps

Figure 30.7 illustrates the typical geometry of a two-lane on-ramp. Two lanes join the freeway at the merge point. There are, in effect, two acceleration lanes. First, the right ramp lane merges into the left ramp lane; subsequently, the left ramp lane merges into the right freeway lane. The lengths of these two acceleration lanes are as shown in the figure below.

Figure 30.7: Typical Two-Lane On-Ramp



(Source: Reprinted with permission from Transportation Research Board, National Research Council, *Highway Capacity Manual*, National Academy of Sciences, Courtesy of the National Academies Press, Washington, D.C., 2000.)

Figure 30.7: Full Alternative Text

The general procedure for on-ramps is modified in two ways. When estimating the demand flow in lanes 1 and 2 immediately upstream of the on-ramp (v12), the standard equation is used:

$v_{12}=v_F \times P_{FM}$

However, instead of using the standard equations to find PFM, the following values are used:

- PFM=1.000 for four-lane freeways
- PFM=0.555 for six-lane freeways
- PFM=0.209 for eight-lane freeways

In addition, in the density equation, the length of the acceleration lane is replaced by an effective length that considers both lanes of the two-lane merge area:

$L_{aEFF}=2L_{A1}+L_{A2}$ [30-35]

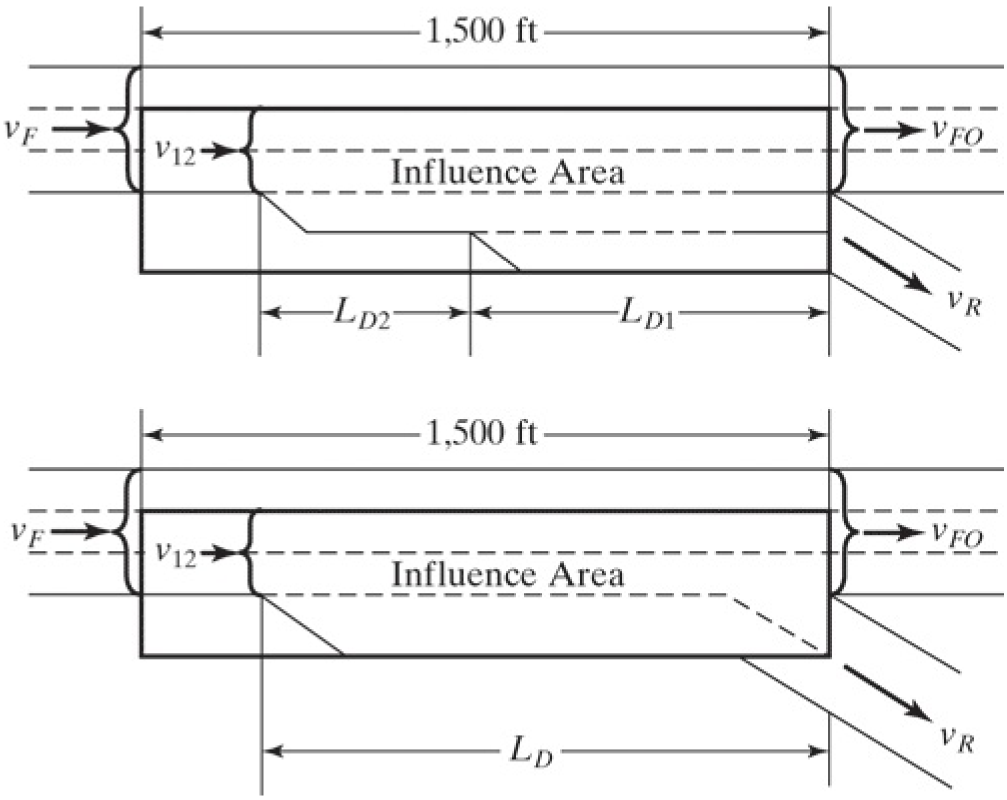
where LA1 and LA2 are defined in Figure 30.7.

Occasionally, a two-lane on-ramp will be used at a location where one or two lanes are being added to the downstream freeway section. Depending on the details of such merge areas, they could be treated as *lane additions* or as *major merge areas*.

30.5.2 Two-Lane Off-Ramps

Figure 30.8 illustrates two common geometries used with two-lane off-ramps. The first is a mirror image of a typical two-lane on-ramp junction, with two deceleration lanes provided. The second provides a single deceleration lane, with the left-hand ramp lane originating at the diverge point without a separate deceleration lane.

Figure 30.8: Typical Geometries for Two-Lane Off-Ramps



(Source: Reprinted with permission from Transportation Research Board, National Research Council, *Highway Capacity Manual*, National Academy of Sciences, Courtesy of the National Academies Press, Washington, D.C., 2000.)

Figure 30.8: Full Alternative Text

As was the case with two-lane on-ramps, the standard procedures are applied to the analysis of two-lane off-ramps with two modifications. In the standard equation,

$$v_{12} = v_R + (v_F - v_R)P_{FD}$$

the following values are used for P_{FD}:

- $P_{FD} = 1.000$ for four-lane freeways
- $P_{FD} = 0.450$ for six-lane freeways
- $P_{FD} = 0.260$ for eight-lane freeways

Also, the length of the acceleration lane in the density equation is replaced with an effective length, computed as follows:

$LdEFF=2LD1+LD2$ [30-36]

where LD1 and LD2 are defined in Figure 30.8. This modification is applied only in the case of the geometry shown in the first part of Figure 30.8. Where there is only one deceleration lane, it is used without modification.

30.5.3 On- and Off-Ramps on Five-Lane Freeway Segments (One Direction)

In some areas of the country, freeway sections with five lanes in a single direction are not uncommon. The procedure for analyzing right-hand ramps on such sections is relatively simple: An estimate of the demand flow in lane 5 (the left-most lane) of the section is made. This is deducted from the total approaching freeway flow; the remaining flow is in the right four lanes of the section. Once this deduction is made, the section can be analyzed as if it were a ramp on an eight-lane freeway (four lanes in one direction). Table 30.8 gives simple algorithms for determining the flow in lane 5 (v5). Then:

$v4EFF=vF-v5$ [30-37]

Table 30.8: Estimating Demand Flow in Lane 5 of a Five-Lane Freeway Section

On-Ramps		Off-Ramps	
v_F (pc/h)	v_5 (pc/h)	v_F (pc/h)	v_5 (pc/h)
$\geq 8{,}500$	2,500	$\geq 7{,}000$	$0.200v_F$
7,500–8,499	$0.295v_F$	5,500–6,999	$0.150v_F$
6,500–7,499	$0.270v_F$	4,000–5,499	$0.100v_F$
5,500–6,499	$0.240v_F$	$< 4{,}000$	0
$< 5{,}500$	$0.220v_F$		

(Source: Reprinted with permission from Transportation Research Board, National Research Council, *Highway Capacity Manual*, National Academy of Sciences, Courtesy of the National Academies Press, Washington, D.C., 2000.)

Table 30.8: Full Alternative Text

and the remainder of the problem is analyzed using v4EFF as the approaching freeway flow on a four-lane (one direction) freeway section.

Obviously, a similar approach could be taken where a ramp exists on a one-direction freeway segment with *more* than five lanes (some do exist, although rare). However, no calibrated methodology for estimating flow in the "outer lanes" exists, and estimates would have to be based on local field observations.

30.5.4 Left-Hand On- and Off-Ramps

Left-hand on- and off-ramps are found, with varying frequency, in most parts of the nation. A technique for modifying analysis procedures for application to left-hand ramps was developed in the 1970s by Leisch [2]. The technique follows the following steps:

- Estimate v12 for the prevailing conditions as if the ramp were on the right-hand side of the freeway.

- To estimate the traffic remaining in the two left-most lanes of the freeway (v12 for a four-lane freeway, v23 for a six-lane freeway, v34 for an eight-lane freeway), multiply the result by the appropriate factor selected from Table 30.9.

Table 30.9: Conversion of v12 Estimates for Left-Hand Ramps

$$v_{xy} = v_{12} \times f_{LH}$$

Adjustment Factor, f_{LH}		
To Estimate:	**For On-Ramps**	**For Off-Ramps**
v_{12} on four-lane freeways (2 lanes ea. dir.)	1.00	1.00
v_{23} on six-lane freeways (3 lanes ea. dir.)	1.12	1.05
v_{34} on eight-lane freeways (4 lanes ea. dir.)	1.20	1.10

Table 30.9: Full Alternative Text

- Using the demand flow in the two left-most freeway lanes instead of v12, check capacities and estimate density in the ramp influence area without further modification to the methodology.

- Speed algorithms should be viewed as only very rough estimates for left-hand ramps. Speed predictions for “outer lanes” may not be applied.

30.5.5 Lane Additions and Lane Drops

Many merge and diverge junctions involve the addition of a lane (at a merge area) or the deletion of a lane (at a diverge area). In general, these areas are relatively straightforward to analyze, applying the following general principles:

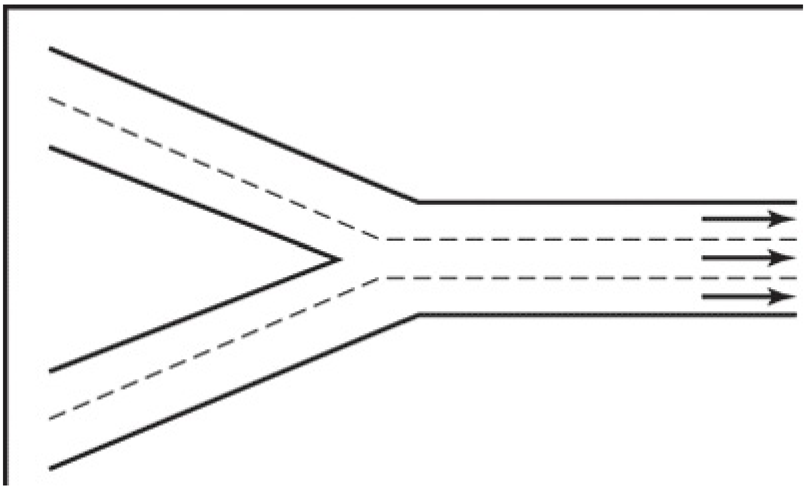
- Where a single-lane ramp adds a lane (at a merge) or deletes a lane (at a diverge), the capacity of the ramp is determined by its free-flow speed, and it is analyzed as a ramp roadway using the criteria of Table 30.5. Level-of-service criteria for basic freeway sections are applied to upstream and downstream freeway segments, which will have a different number of lanes.
- Where a two-lane ramp results in a lane addition or a lane deletion, it is treated as a major merge or diverge area. The techniques described in the next section are applied.

30.5.6 Major Merge and Diverge Areas

A major merge area is formed when two multilane roadways join to form a single freeway or multilane highway segment. A major diverge area occurs when a freeway or multilane highway segment splits into two multilane downstream roadways. These multilane merge and diverge situations may be part of major freeway interchanges or may involve significant multilane ramp connections to surface streets. The typical characteristic of these roadways are that they are often designed to accommodate relatively high speeds, which somewhat changes the dynamics of merge and diverge operations.

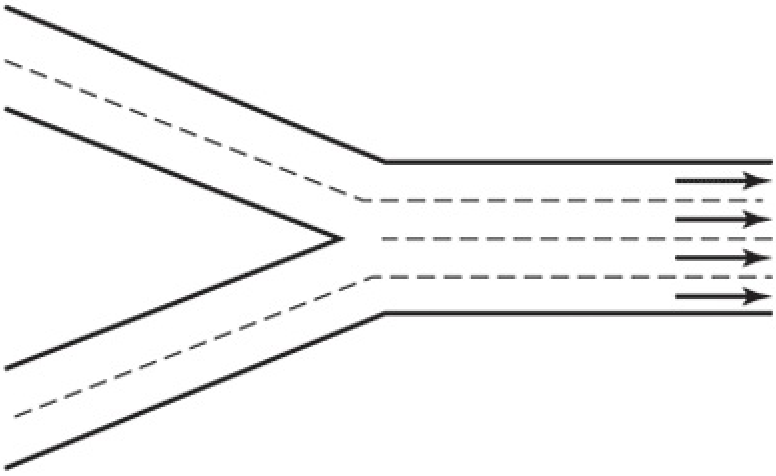
At a major merge area, a lane may be dropped, or the number of lanes in the downstream section may be the same as the total approaching the merge. Similarly, at a diverge area, a lane may be added, or the total lanes leaving the diverge area may be equal to the number on the approaching facility segment. Figure 30.9 illustrates these configurations.

Figure 30.9: Major Merge and Diverge Areas



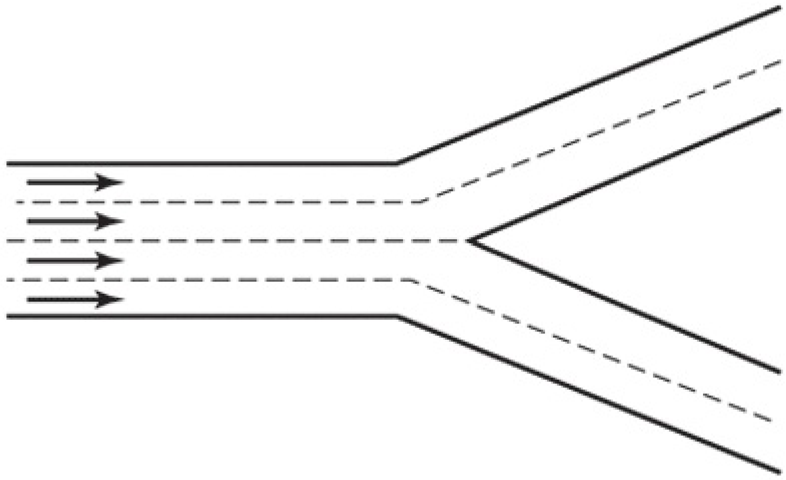
(a) Major Merge Area With Lane Drop

30.5-12 Full Alternative Text



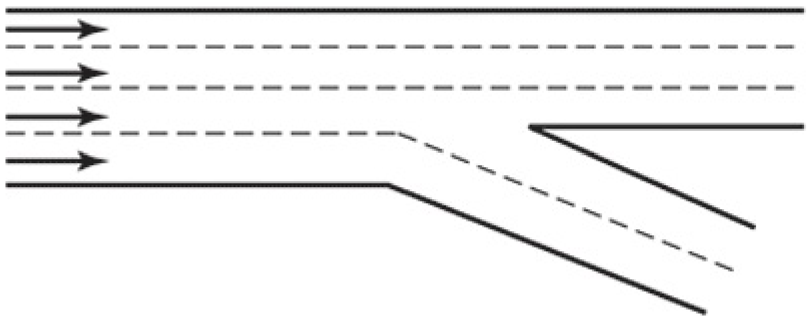
(b) Major Merge Area Without Lane Drop

30.5-12 Full Alternative Text



(c) Major Diverge Area Without Lane Addition

30.5-12 Full Alternative Text



(d) Major Diverge Area With Lane Addition

30.5-12 Full Alternative Text

(Source: Reprinted with permission from Transportation Research Board, National Research Council, *Highway Capacity Manual*, National Academy of Sciences, Courtesy of the National Academies Press, Washington, D.C., 2000.)

The analysis of major merge and diverge areas is generally limited to an examination of the demand-capacity balance of approaching and departing facility segments. No level-of-service criteria are applied.

For major diverge areas, an algorithm has been developed to roughly estimate the density across all approaching freeway lanes for a segment 1,500 feet upstream of the diverge:

$D=0.0109\ (v_{FN})$ [30-38]

where:

D=density across all freeway lanes, from diverge to apoint 1,500 ft upstrea

This is an approximation at best and is generally not used to assign a level of service to the diverge area.

30.6 Closing Comments

The methodology used in the HCM 2016 for merge and diverge segments is fundamentally unchanged from the version originally presented in the HCM 2000, with the exception that the ability to apply a CAF and speed adjustment factor (SAF) has been added. These factors are presented and discussed in Chapter 28.

The methodology focuses on two of the three potential components of a ramp: the ramp-freeway junction and the ramp roadway. Where a ramp connects two freeways (or multilane highways), there would be two ramp-freeway junctions.

The methodology *does not* treat the ramp-street junction. Such junctions are usually signalized intersections, or are controlled by STOP or YIELD signs. The appropriate methodology for analyzing such junctions must be applied. Where an off-ramp-street junction is expected to fail (i.e., LOS F), queues will develop on the ramp that may well back up into the ramp-freeway junction, which would not, in such a case, operate as anticipated by the methodology of this chapter. When an on-ramp-street junction fails, demand entering the ramp would be constrained, and operations of the ramp-freeway junction may look better than anticipated by the methodology of this chapter.

In any event, the methodology discussed is very detailed, and is primarily based on regression analysis of moderately large nationwide databases. It is important to follow the methodology carefully to avoid missing key steps in the process.

30.7 Sample Problems in Merging and Diverging Analysis

Sample Problem 30-2: Analysis of an Isolated On-Ramp

An on-ramp to a busy eight-lane urban freeway is illustrated in Figure 30.10. An analysis of this merge area is to determine the likely level of service under the prevailing conditions shown.

Figure 30.10: On-Ramp Merge Segment for Sample Problem 30-1

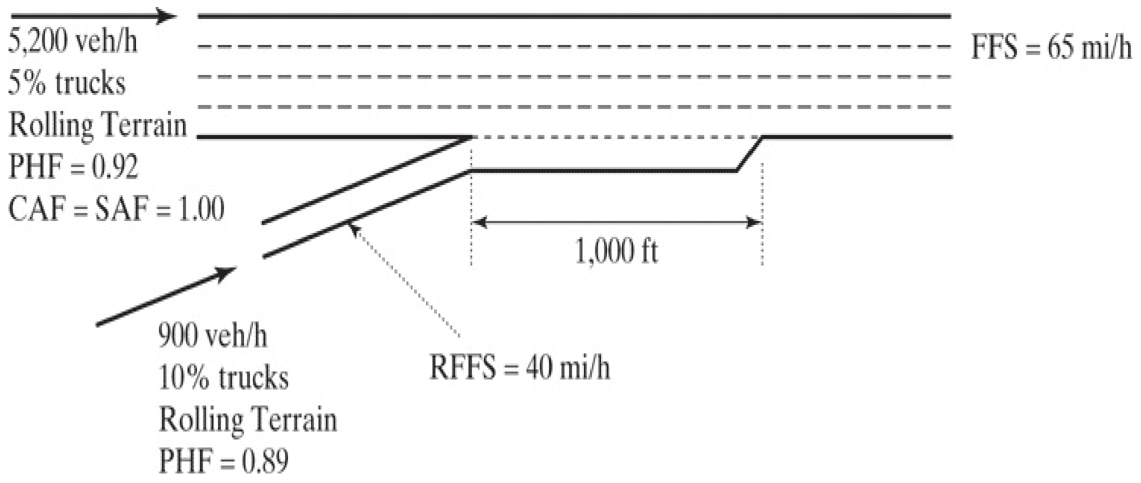


Figure 30.10: Full Alternative Text

Solution

1. Step 1: Convert All Demand Volumes to Flow Rates in pc/h under Equivalent Ideal Conditions

   The freeway and ramp flows approaching the merge area must be converted to flow rates in pc/h under equivalent base conditions using Equation 30-1. In this case, note that the truck percentages and PHF are different for the two. From Chapter 28, the passenger car equivalent for trucks (E_{HV}) in rolling terrain is 3.0.

   For the ramp demand flow:

   fHV=11+PHV (EHV −1)=11+0.10 (3−1)=0.833vR=VRPHF×fHV=9000.89×0.8

   For the freeway demand flow:

   fHV=11+PHV (EHV −1)=11+0.05 (3−1)=0.909vF=VFPHF×fHV=5,2000.92×0

2. Step 2: Determine the Demand Flow Remaining in Lanes 1 and 2 Immediately Upstream of the Merge

   Table 30.2 gives values of PFM, the proportion of freeway vehicles remaining in lanes 1 and 2 immediately upstream of a merge. For an eight-lane freeway (four lanes in each direction), Equation 30-8 or 30-9 is used to estimate P_{FM}, depending upon the value of $v_F/RFFS$ which is 6,218/40=155.45. As this is more than 72, Equation 30-9 is used. Then:

   PFM=0.2178−0.000125vR=0.2178−(0.000125×1,214)=0.

   This prediction must be checked for “reasonableness.” The average flow rate in lanes 1 and 2 is 410/2=205 pc/h —very low by any judgment. This leaves 6,218−410=5,808 pc/h in the two outer lanes (lanes 3 and 4), or 5808/2=2,904 pc/h/ln. This violates the maximum reasonable limit of 2,700 pc/h/ln. It also violates the 1.5 rule: 2819>1.5×205=308 pc/h/ln. In this case, the 1.5 rule is violated by a great deal. The expected flow rate in lanes 1 and 2, therefore, must be revised in accordance with Equation 30-20:

$$v_{12} = \frac{v_F}{2.50} = \frac{6,218}{2.50} = 2,487 \text{ pc/h}$$

With this value for v_{12}, the outer lanes would carry $6,218 - 2,487 = 3,732$ pc/h, or $3,731/2 = 1,857$ pc/h/ln, which now satisfies both “reasonableness” criteria. The example will move forward using this value.

3. Step 3: Check Capacity of Merge Area and Compare to Demand Flows

   To determine whether the section will fail (LOS F), the capacity values of Table 30.5 must be consulted. For a merge section, the critical capacity check is on the downstream freeway section, where:

   $$v_{FO} = v_F + v_R = 6,218 + 1,214 = 7,432 \text{ pc/h}$$

   From Table 30.5, the capacity of a four-lane freeway section is 9,400 pc/h when the FFS is 65 mi/h. As $9,400 > 7,432$, no failure is expected due to total downstream flow.

   The capacity of a one-lane ramp with a free-flow speed of 40 mi/h must also be checked. From Table 30.5, such a ramp has a capacity of 2,000 pc/h. As this is greater than the ramp demand flow of 1,214 pc/h, this element will not fail either.

   Total flow entering the merge influence area is:

   $$v_{R12} = v_R + v_{12} = 1,214 + 2,487 = 3,701 \text{ pc/h}$$

   As the maximum desirable entering flow for single-lane merge area is 4,600 pc/h, this element is also acceptable.

4. Step 4: Estimate Density and Level of Service in the Ramp Influence Area

As stable operations are expected, Equation 30-21 is used to estimate the density in the ramp influence area:

$$DR=5.475+0.00734\ vR+0.0078\ v12-0.00627\ LaDR=5.47$$
$$(0.00734\times1{,}214)+$$
$$(0.0078\times2{,}487)-(0.00627\times1{,}000)DR=5.475+8.91+19.40-$$

From the criteria in Table 30.1, this is LOS C, but close to the LOS D boundary of 28 pc/mi/ln.

5. Step 5: Estimate Speed Parameters

   Although not used to determine level of service, the algorithms of Table 30.6 may be used to estimate speed parameters of interest:

$$MS=0.321+0.0039\ e(vR12/1000)-0.002\ (La\times RFFS\times SAI$$
$$[\ 0.0039\ e(3701/1000)\ ]-[\ 0.002\ (1000\times40\times1.00/1000)$$
$$]MS=0.321+0.158-0.080=0.399SR=FFS\times SAF$$
$$-(FFS\times SAF$$
$$-42)\ MS=65\times1-(65\times1-42)\times0.399=59.8\ mi/hSO=FFS\times S$$
$$-0.0036\ (vOA-500)=65\times1-[\ 0.0036\ (1{,}857-500)$$
$$]=60.1\ mi/h$$

   The average speed in the ramp influence area is 59.8 mi/h, while the average speed in outer lanes is 60.1 mi/h. The average speed of all vehicles is:

$$S=vR12+vOA\ \ NO(vR12SR)+(vOA\ \ NOSO)=3701+$$
$$(1857\times2)(370159.8)+$$
$$(1857\times260.1)=7{,}41561.9+61.8=59.9\ mi/h$$

Discussion

Several additional items may be of interest. The lane distribution of the incoming freeway flow (vF) was checked for reasonableness and adjusted accordingly. In this case, it was estimated that 2,487 pc/h use lanes 1 and 2, while 6,218−2,487=3,731 pc/h use lanes 3 and 4. This is not unexpected, given the large ramp flow (1,214 pc/h) entering at the on-ramp.

It is also useful to check the LOS on the downstream basic freeway section. It carries a total of 7,240 pc/h in four lanes, or 1,810 pc/h/ln. Using the standard speed-flow curve for FFS=65 (see Chapter 28), this is LOS D.

What does this mean, considering that the LOS for the ramp influence area is determined to be C? It means that the total freeway flow is the determining element in overall level of service. This is as it should be, as it is always undesirable to have minor movements (in this case, the on-ramp), controlling the overall operation of the facility.

Sample Problem 30-3: Analysis of a Sequence of Freeway Ramps

Figure 30.11 shows a series of three ramps on a six-lane freeway (three lanes in each direction). All three ramps are to be analyzed to determine the level of service expected under the prevailing conditions shown.

Figure 30.11: Ramp Sequence for Sample Problem 30-2

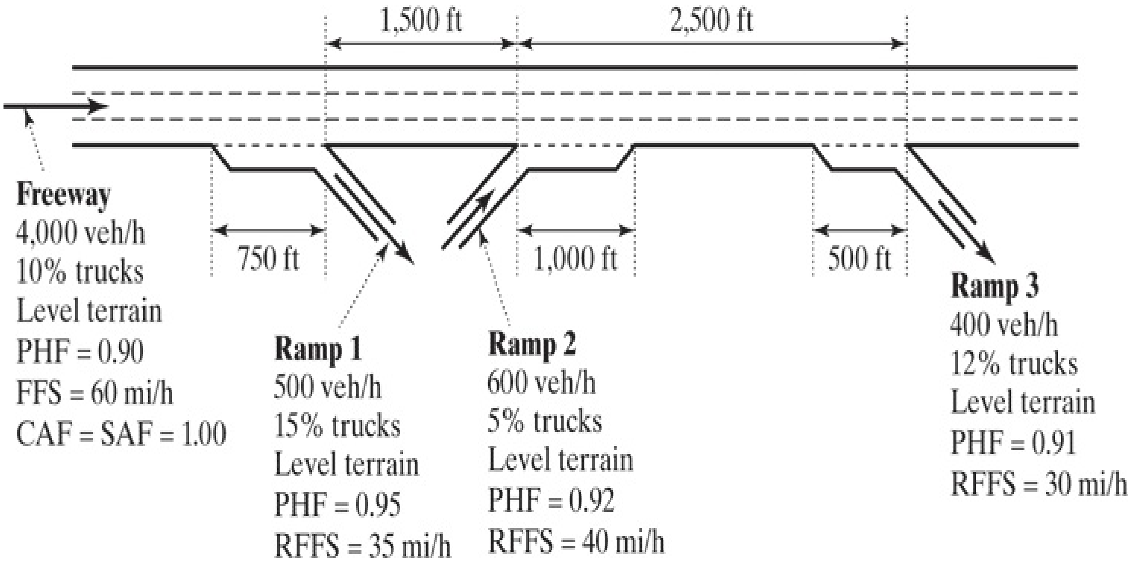


Figure 30.11: Full Alternative Text

This example has a number of elements. The ramp sequence is OFF-ON-OFF. Because the ramps are on a 6-lane freeway, it is possible that operations at one will be affected by the presence of the others. For simplicity, the example is solved by treating all three ramps simultaneously during each step of the computations.

Solution

1. Step 1: Convert All Demand Volumes to Flow Rates in pc/h under Equivalent Ideal Conditions

   Before applying any of the models for ramp analysis, all demand volumes must be converted to flow rates in pc/h under equivalent base conditions. This is done using Equation 30-1. Peak-hour factors for each movement are given, as are truck percentages. The heavy-vehicle factor is computed using E_{HV} values from Chapter 28. For level terrain (which prevails on all elements), EHV=2 for all movements. There are no CAFs or SAFs that apply, that is, their values are 1.00.

   While the E_{HV} is the same for all elements, the proportion of trucks in each demand volume is different. Thus, there will be four different values for the heavy vehicle adjustment factor, f_{HV}. The computations for the four adjustment factors are shown in Table 30.10.

Table 30.10: Computation of Heavy Vehicle Adjustment Factors

for Sample Problem 30-2

$$f_{HV} = \frac{1}{1 + P_{HV}(E_{HV} - 1)}$$

Element	P_{HV}	E_{HV}	f_{HV}
Freeway	0.10	2	$\frac{1}{1 + 0.10\,(2 - 1)} = 0.909$
Ramp 1	0.15	2	$\frac{1}{1 + 0.15(2 - 1)} = 0.870$
Ramp 2	0.05	2	$\frac{1}{1 + 0.05(2 - 1)} = 0.952$
Ramps 3	0.12	2	$\frac{1}{1 + 0.12(2 - 1)} = 0.893$

Table 30.10: Full Alternative Text

The conversion of demand volumes to flow rates in pc/h is accomplished using Equation 30-1. These computations are illustrated in Table 30.11. Note that all peak-hour factors are specified in the problem statement.

Table 30.11: Computation of Demand Flow Rates in pc/h for Sample

Problem 30-2

$v_i = \dfrac{V_i}{PHF \times f_{HV}}$				
Element	**Volume (veh/h)**	**Peak Hour Factor (*PHF*)**	f_{HV}	**Flow Rate in pc/h**
Freeway	4,000	0.90	0.909	$\dfrac{4000}{0.90 \times 0.909} = 4{,}889\ pc/h$
Ramp 1	500	0.95	0.870	$\dfrac{500}{0.95 \times 0.870} = 605\ pc/h$
Ramp 2	600	0.92	0.952	$\dfrac{600}{0.92 \times 0.952} = 685\ pc/h$
Ramp 3	400	0.91	0.893	$\dfrac{400}{0.91 \times 0.893} = 492\ pc/h$

Table 30.11: Full Alternative Text

2. Step 2: Determine the Flow in Lanes 1 and 2 Immediately Upstream of Each Ramp in the Sequence

   This is the most interesting part of the problem. The appropriate equation from Table 30.2 must be chosen to compute v_{12} for each ramp. This will at least involve consideration of the impact of the upstream and downstream adjacent ramp in each case, and the use of the selection table (Table 30.3).

   - Ramp 1: The first ramp is part of a three-ramp sequence that can be described as None-OFF-On (no upstream adjacent ramp; an adjacent downstream on-ramp). Using Table 30.2, for a six-lane freeway and the sequence indicated, Equation 30-11 should be used to determine v12.

$v_{12(1)} = v_{R1} +$
$(v_{F1} - v_{R1})\ P_{FD}$ $P_{FD} = 0.760 - 0.000025 v_F$
$- 0.000046 v_R$ $P_{FD} = 0.760 - (0.000025 \times 4{,}889) - (0.000$
$(4{,}889 - 605) \times 0.610 = 2{,}613$ pc/h

The resulting lane distribution must be checked for reasonableness. A six-lane freeway has only one outer lane, which would carry $4{,}889 - 2{,}613 = 2{,}276$ pc/h. This is less than 2,700 pc/h; it is *not* less than $1.5 \times (2{,}613/2) = 1{,}960$ pc/h. The predicted lane distribution is *not* reasonable, and must be adjusted using Equation 30-20:

$$v_{12(1)} = \frac{v_{F1}}{1.75} = \frac{4{,}889}{1.75} = 2{,}794 \text{ pc/h}$$

This is the value that will be used in subsequent calculation.

- Ramp 2: The second ramp is an on-ramp that can be described as part of an Off-ON-Off sequence. From Table 30.2, there are three potential equations that might apply: Equation 30-6, which considers the effect of the upstream off-ramp; Equation 30-7, which considers the effect of the downstream off-ramp; or Equation 30-5, which treats the ramp as if it were isolated. It is even possible that two of these apply, in which case the equation yielding the larger v_{12} estimate is used. To determine which of these apply requires the use of the equivalence distances, computed using the equations in Table 30.4.

  In considering whether the impact of the upstream off-ramp must be considered, Equation 30-15 is used:

  $$L_{EQ} = 0.214(v_R + v_F) + 0.444 L_a + 52.32 RFFS - 2{,}403$$

  Note that for Ramp 2, the approaching freeway flow is the beginning freeway flow minus the off-ramp flow at Ramp 1:

$v_{F2}=v_{F1}-v_{R1}=4{,}889-605=4{,}284\ \text{pc/h}$

Thus:

$L_{EQ}=0.214\,(685+4{,}284)+(0.444\times1{,}000)+(52.32\times40)-2{,}403 \quad L_{EQ}=1{,}063+444+2{,}093-2{,}403=1,$

As the actual distance to the upstream ramp is 1,500 ft>1,195 ft, the impact of the upstream off-ramp should not be considered, and Equation 30-5 is used.

To determine whether or not the effect of the downstream off-ramp must be considered, Equation 30-16 is used to compute LEQ:

$L_{EQ}=\frac{v_D}{0.1096+0.000107\,L_a}=\frac{492}{0.1096+(0.000107\times1{,}000)}=\frac{492}{0.2166}=2{,}271\ \text{ft}$

The actual distance to the downstream off-ramp is 2,500 ft>2,271 ft. Thus, the impact of the downstream off-ramp is also not considered, and Equation 30-5 is used. Through the determination of these equivalence distances, it is seen that Ramp 2 may be considered to be an isolated ramp. Only one—Equation 30-5—applies to the estimation of $v_{12(2)}$.

$v_{12(2)}=v_{F2}\times P_{FM} \quad P_{FM}=0.5775+0.000028L_a \quad P_{FM}=0$
$(0.000028\times1{,}000)=0.6055 \quad v_{12(2)}=4{,}284\times0.6055=2{,}5$

This distribution must also be tested for reasonableness. The outer lane carries 4,284−2,594=1,690 pc/h<2,700 pc/h. It also carries less than 1.5×(2,594/2)=1,946 pc/h. Therefore, the predicted lane distribution is reasonable, and will be used.

- Ramp 3: The third ramp is now considered as part of an On-OFF-None sequence. From Table 30.3, Equation 30-11 or 30-12 is used. To determine which is the appropriate one for application, Equation 30-16 (see Table 30.6) is used to compute LEQ. In applying this equation, note that vF3 includes the on-ramp flow from Ramp 2. Thus:

$$vF3=vF2+vR2=4{,}284+685=4{,}969\ \text{pc/h}LEQ=vU0.07$$
$$-0.000076vRLEQ=6850.071+$$
$$(0.000023\times 4{,}969)-(0.000076\times 492)=4{,}628\ \text{f}$$

As the actual distance to the upstream on-ramp is only 2,500 ft<4,628 ft, Equation 30-12 is used to consider the impact of Ramp 2 on lane distribution at Ramp 3:

$$v12(3)=vR3+$$
$$(vF3-vR3)\ PFDPFD=0.717-0.000039vF+0.604(vUI$$
$$(4{,}969-492)\times 0.688=3{,}572\ \text{pc/h}$$

Again, the predicted lane distribution should be checked for reasonableness. The outer lane flow is 4,969−3,572=1,397 pc/h/ln<2,700 pc/h/ln. It is also less than 1.5×(3,572/2)=2,681 pc/h/ln. Therefore, the distribution is reasonable, and will be used.

Summarizing the results for v12 immediately upstream of each of the three ramps:

$$v12(1)=2{,}794\ \text{pc/h}v12(2)=2{,}594\ \text{pc/h}v12(3)=3{,}572$$

3. Step 3: Check Capacities

   The capacities and limiting values of Table 30.5 must now be checked to see whether operations are stable or whether level of service F exists. The freeway flow check is made between Ramps 2 and 3, as this is the point where total freeway flow is greatest (vF3). These checks are performed in Table 30.12. Remember that the freeway FFS is 60 mi/h.

Table 30.12: Capacity

Checks for Sample Problem 30-2

Item	Demand Flow (pc/h)	Capacity (pc/h) (Table 28.5)
v_{F3}	4,969	6,900 (FFS = 60 mi/h)
$v_{12(1)}$	2,794	4,400
$v_{R12(2)}$	2,594 + 685 = 3,279	4,600
$v_{12(3)}$	3,572	4,400
v_{R1}	605	2,000 ($RFFS$ = 35 mi/h)
v_{R2}	685	2,000 ($RFFS$ = 40 mi/h)
v_{R3}	492	1,900 ($RFFS$ = 30 mi/h)

Table 30.12: Full Alternative Text

None of the demand flows exceed the capacities or limiting values of Table 30.5. Thus, stable operation is expected throughout the section.

4. Step 4: Determine Densities and Levels of Service in Each Ramp Influence Area

   The density in the ramp influence area is estimated using Equation 30-21 for on-ramps and 30-22 for off-ramps:

   DR1=4.252+0.0086v12(1)−0.009Ld(1)DR1=4.252+
   (0.0086×2,794)−(0.009×750)=21.53 pc/mi/lnDR2=5.475
   (0.0073×685)+
   (0.0078×2,594)−(0.0062×1,000)=24.51 pc/mi/lnDR3=4.2
   (0.0086×3,572)−(0.009×500)=30.47 pc/mi/ln

   From Table 30.1, Ramp 1 operates at LOS C, Ramp 2 at LOS C, and Ramp 3 at LOS D.

5. Step 5: Determine Speeds for Each Ramp

   As was done in Sample Problem 30-1, the algorithms of Tables 30.6 and 30.7 may be used to estimate space mean speeds within each ramp influence area and across all freeway lanes within the 1,500 ft range of each ramp influence area. Because of the length of these computations, they are not shown here. Each would follow the sequence illustrated in Sample Problem 30-1. The results, however, are shown in Table 30.13.

Table 30.13: Speed Results for Sample Problem 30-2

Item	Ramp 1	Ramp 2	Ramp 3
Average Speed in Ramp Influence Area, S_R (mi/h)	47.8	53.4	45.4
Average Speed in Outer Lanes S_o (mi/h)	61.5	64.3	63.1
Average Speed in All Lanes S (mi/h)	52.8	56.3	49.2

Table 30.13: Full Alternative Text

Discussion

Note that the ramp influence areas of Ramps 2 and 3 overlap for a distance of 500 ft (1,500+1,500=3,000 ft). For this overlapping segment, the influence area having the highest density and lowest LOS would be used. In this case, Ramp 3 has the worst LOS–D. That would govern the overlap area.

Again, it is interesting to check the basic freeway level of service associated with the controlling (or largest) total freeway flow, which occurs between Ramps 2 and 3. The demand flow per lane for this segment is 4,969/3=1,656 pc/h/ln. From Chapter 28 and an FFS of 60 mi/h, the level of service is found to be level of service D. This is compared with the Ramp 3 LOS, which is also D. Thus, the operation of the freeway as a whole and ramp sequence are somewhat in balance, a desirable condition.

References

- 1. Roess, R, and Ulerio, J., “Capacity of Ramp-Freeway Junctions,” *Final Report*, Polytechnic University, Brooklyn, NY 1993.

- 2. Leisch, J., *Capacity Analysis Techniques for Design and Operation of Freeway Facilities*, Federal Highway Administration, U.S. Department of Transportation, Washington, D.C., 1974.

Problems

1. 30-1. Consider the pair of ramps shown in Figure 30.12. It may be assumed that there is no ramp-to-ramp flow.

Figure 30.12: Segment for Problem 30-1

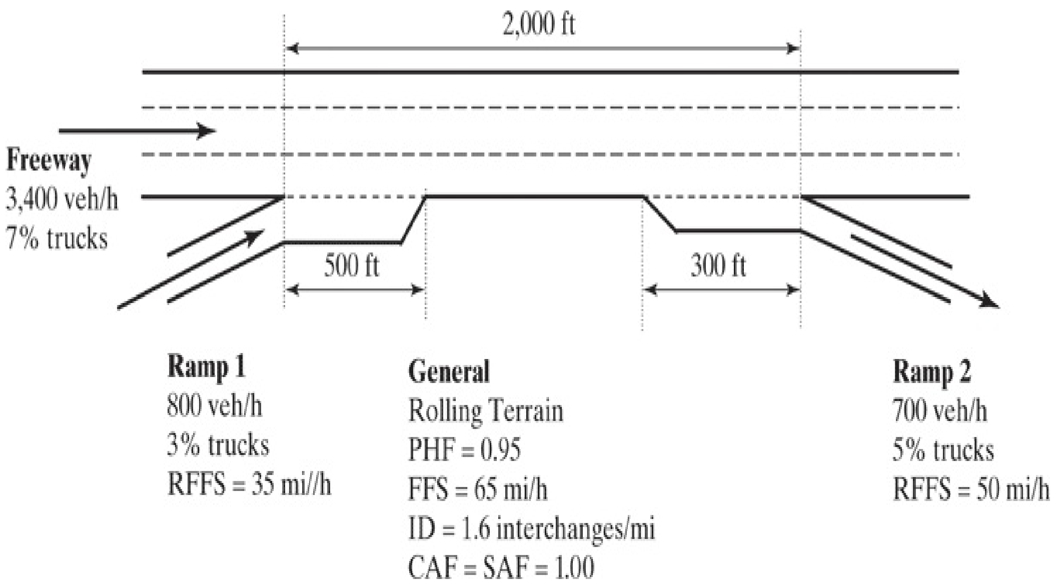


Figure 30.12: Full Alternative Text

   1. Given the existing demand volumes and other prevailing conditions, at what level of service is this section expected to operate? If problems exist, which elements appear to be causing the difficulty?

   2. Note that this segment is identical to that of Problem 29-1, except that in 29-1, the two ramps are connected by a continuous auxiliary lane, forming a ramp-weave segment. Compare the solutions to the weaving segment to part (a), which is a ramp sequence. Which would you recommend, and why?

2. 30-2. Consider the on-ramp shown in Figure 30.13. There are no conditions for which a CAF or SAF would apply.

Figure 30.13: Segment for Problem 30-2

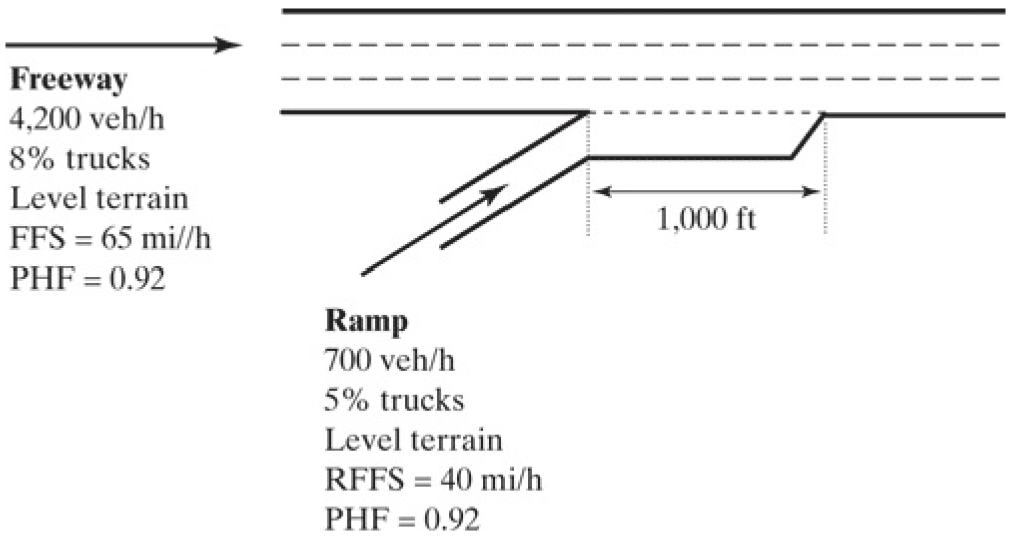


Figure 30.13: Full Alternative Text

1. At what level of service would the merge area be expected to operate?
2. A new development nearby opens and increases the on-ramp volume to 1,000 veh/h. How does this affect the level of service?

3. 30-3. Figure 30.14 illustrates two consecutive ramps on an older freeway. It may be assumed that there is a ramp-to-ramp flow of 150 veh/h.

Figure 30.14: Segment for Problem 30-3

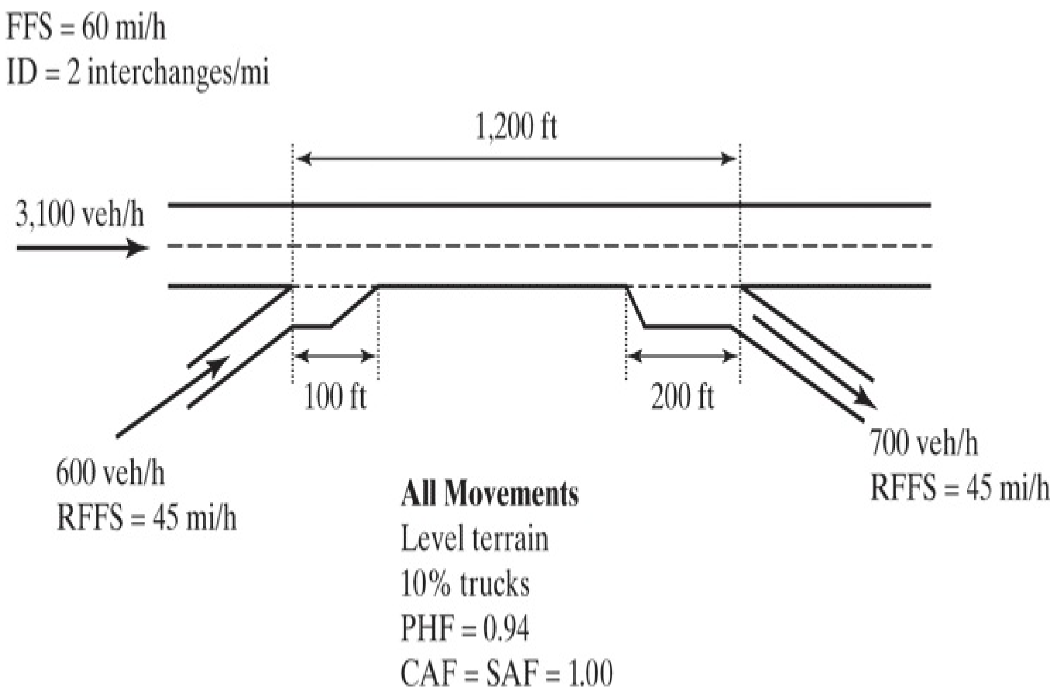


Figure 30.14: Full Alternative Text

1. What is the expected level of service for the conditions shown?
2. Several improvement plans are under consideration:
    1. Connect the two ramps with a continuous auxiliary lane, forming a weaving segment.
    2. Add a third lane to the freeway and extend the length of acceleration and deceleration lanes to 300 feet.
    3. Provide a lane addition at the on-ramp that continues past the off-ramp on the downstream freeway section. The off-ramp deceleration lane remains 200 feet long.

Which of these three improvements would you recommend? Why? Justify your answer.

Chapter 31 Operation and Analysis of Freeways and Highways

This chapter covers a wide range of topics related to the overall operation and analysis of freeways (in urban and rural environments) and other types of rural highway facilities.

The *2016 Highway Capacity Manual* (HCM) [1] now includes a complex methodology for analysis of freeway facilities, and the ability to use this methodology to evaluate active transportation and demand management (ATDM) strategies and managed lanes. Although these are far too complex to describe in detail in this text, a brief overview will be included here.

This chapter provides overviews of the following subjects related to freeway and rural highway facilities:

1. Markings for freeways and rural highways
2. Signing for freeways and highways
3. Setting speed limits for freeways and highways
4. Managed lanes on freeways
5. ATDM strategies for freeways, including ramp metering
6. The 2016 HCM methodology for freeway facility evaluation

The treatment of each will be relatively brief, and appropriate references are provided for those interested in greater detail on these issues.

31.1 Traffic Markings on Freeways and Rural Highways

Traffic markings on freeways and rural highways include lane lines, edge markings, and specialized markings for on- and off-ramp gore areas. In addition, where at-grade intersections occur on rural highways, intersection markings, covered in Chapter 4, would also be used. On rural two-lane highways, centerline markings, in conjunction with signs, are used to designate passing and nonpassing zones. In all cases, specific marking standards and guidelines are found in the current version of the *Manual on Uniform Traffic Control Devices* (MUTCD). At this writing, the current version (which is found on-line through the FHWA website) is the 2009 edition, as amended through 2012 [2].

31.1.1 Freeway Mainline Markings

Figure 31.1 illustrates typical mainline freeway markings on a freeway. Lane lines are provided to delineate proper lateral placement for vehicles. The lane line is a dashed white line, with dimensions and spacing as defined in the MUTCD. Edge markings are mandated for all freeway segments. Right edge markings are solid single white lines, whereas left edge markings are a solid single yellow line.

Figure 31.1: Freeway Mainline Markings Illustrated

(Photo courtesy of R. Roess and J. Ulerio)

31.1.2 Rural Highway Markings

Marking conventions for rural highways vary according to the specific configuration in place, demand volumes (Average Annual Daily Traffic, AADT), and other factors. As always, the MUTCD [2] is the primary reference to consult for the most current criteria.

Centerlines

Centerlines are critical markings on all types of highways, as they separate opposing directions of flow. Keeping opposing flows clearly separated is a safety element of utmost importance. Nevertheless, centerline markings are not required on all roadways, particularly on low-volume roadways in rural areas. General guidelines in the MUTCD include the following:

- Centerline markings *shall* (mandatory standard) be placed on all paved two-way streets or highways that have three or more traffic lanes.
- Centerline markings *should* (guidance) be placed on all rural arterials

and collectors that have a traveled width of more than 18 ft and an AADT of 3,000 veh/day or greater.

Centerline markings are often *not* used on narrow two-lane alignments of 18 ft or less. On such narrow roadways, most drivers will operate their vehicles in a position that covers the centerline, moving to the edge only when an opposing vehicle approaches. Use of a centerline on such narrow roadways would create two 9 ft (or less) lanes. Such lane widths are no longer recommended, but still exist on many tertiary local rural roads in harsh terrain. Such roadway widths should be avoided where any appreciable traffic demand exists.

Edge Markings

The MUTCD provides the following criteria:

- Edge markings *shall* (mandatory standard) be placed on paved streets or highways with the following characteristics: freeways, expressways, or rural highways with a traveled way of 20 ft or more in width and an AADT of 6,000 veh/day or more.

- Edge markings *should* (guidance) be placed on rural highways with a traveled way of 20 ft or more in width and an AADT of 3,000 veh/day or more.

Centerline Markings to Control Passing on Two-Lane Rural Highways

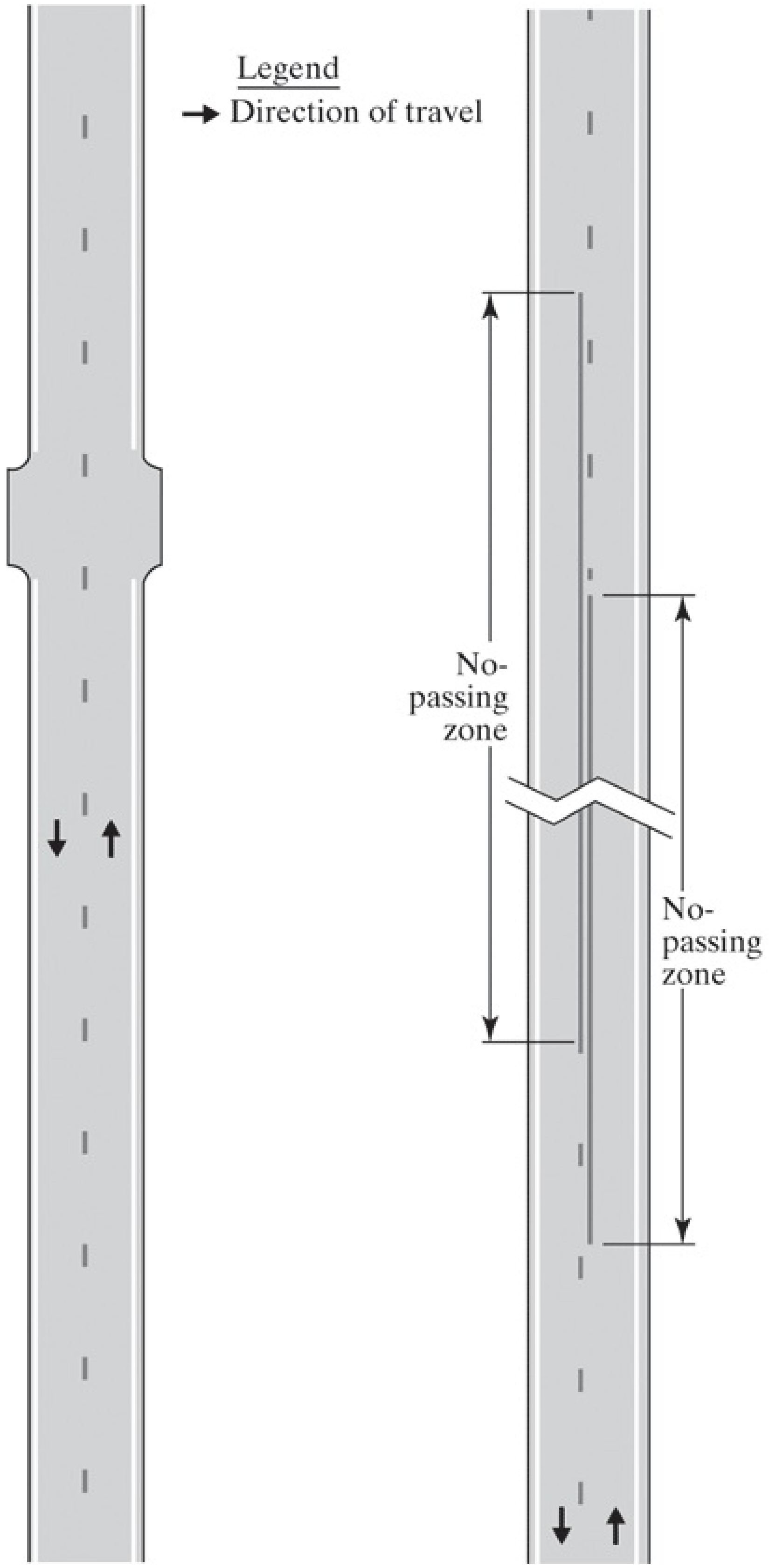
The unique characteristic of two-lane rural highways is that the passing maneuver takes place in the opposing travel lane. This creates a potentially very hazardous situation for drivers if they do not receive specific information from markings and signs about when and where such maneuvers may be safely attempted. This unique characteristic also means that when volumes are high, traffic in one direction interacts with and affects traffic in the opposing direction.

Figure 31.2 illustrates the typical markings used to demark safe and unsafe passing zones on a two-lane rural highway.

Figure 31.2: Typical Markings for Passing Control on Two-Lane, Two-Way Rural Highways

A-Typical two-lane, two-way marking with passing permitted in both directions

B-Typical two-lane, two-way marking with no-passing zones



(Source: *Manual of Uniform Traffic Control Devices*, Federal Highway Administration, Washington, D.C., 2009, as updated through May 2012, Figure 3B-1, pg 350.)

Figure 31.2: Full Alternative Text

A single yellow dashed centerline signifies that passing in either direction is permitted. A double solid yellow centerline signifies that no passing from either direction is permitted. A single solid and a single dashed yellow centerline signify that passing is permitted from the side with the dashed line and prohibited from the direction with the solid line.

The decision on whether passing is to be permitted or not at any location is based upon the concept of *passing sight distance*. The passing maneuver involves use of the opposing traffic lane and is, therefore, essentially very dangerous. The necessary sight distance for a safe passing maneuver involves four distinct distance elements:

- The distance traversed during perception and reaction, when the driver decides to execute a passing maneuver, and the beginning of the maneuver when the passing vehicle encroaches on the opposing lane.
- The distance traveled by the passing vehicle while occupying the left (or opposing) lane.
- The distance between the passing vehicle and a potentially approaching vehicle in the opposite direction when the passing vehicle returns to its travel lane.
- A minimum distance buffer between the passing and opposing vehicle for safety and comfort of both drivers.

The driver of the passing vehicle must be able to see the opposing lane for a distance equal to all of these elements if passing is to be permitted. Distance computations are complex, and depend upon a number of assumed values, including the speed of the passing, passed, and approaching opposing vehicle, acceleration rates, and other factors.

The MUTCD provides criteria for minimum passing sight distances. Whenever sight distances on a two-lane, two-way highway fall short of these criteria, a “no passing” zone *must* be established. The criteria are shown in Table 31.1.

Table 31.1: MUTCD Passing Sight Distance Requirements for Two-Lane, Two-Way Rural Highways

85th Percentile Speed, or Speed Limit (mi/h)	Minimum Passing Sight Distance (ft)
25	450
30	500
35	550
40	600
45	700
50	800
55	900
60	1,000
65	1,100
70	1,200

(Source: *Manual of Uniform Traffic Control Devices*, Federal Highway Administration, Washington, D.C., 2009, as updated through May 2012, Table 3B-1, pg 352.)

Table 31.1: Full Alternative Text

It should be noted that while the MUTCD mandates no passing zones when the criteria of Table 31.1 are not met, it *does not* mandate that passing be permitted when the criteria are met. There are many ways to estimate safe passing sight distance, many of which yield larger values than the criteria of Table 31.1. Agencies may prohibit passing in accordance with their own sight distance requirements when they are more stringent than those shown in Table 31.1. They may not, however, apply *less* stringent criteria.

It should also be noted that special pennant-shaped “No Passing” signs *must* be placed at the beginning of every “No Passing” zone. This is critical, as pavement markings often become difficult to see during inclement weather (and invisible during snow), and may also be worn over time.

Special Markings for Three-Lane Highways

Three-lane rural highway alignments are not uncommon. They exist in areas where two-lane highways would present capacity restrictions, but where either availability of right-of-way or cost makes provision of a full four-lane alignment impractical.

Such alignments were frequently used in the 1940s and 1950s. In their original form, the three-lane highway had one lane for the exclusive use of traffic in each direction, and a center passing lane that could be used by vehicles in either direction. This, however, proved to be a very dangerous arrangement. When passing on a two-lane highway, drivers are immediately aware of the danger involved in entering the opposing traffic lane to execute their maneuver. On three-lane alignments, drivers were less likely to consider that an opposing passing vehicle might be using the center lane at the same time. The original three-lane alignments experienced very high accident rates, with accompanying high fatality rates.

Current practice is to stripe two lanes in one direction and one lane in the other. Yellow markings clearly separate the two directions. In many cases, the direction receiving two lanes is reversed periodically, to give drivers in both directions reasonable opportunities to pass. Special transition markings are used at transfer locations.

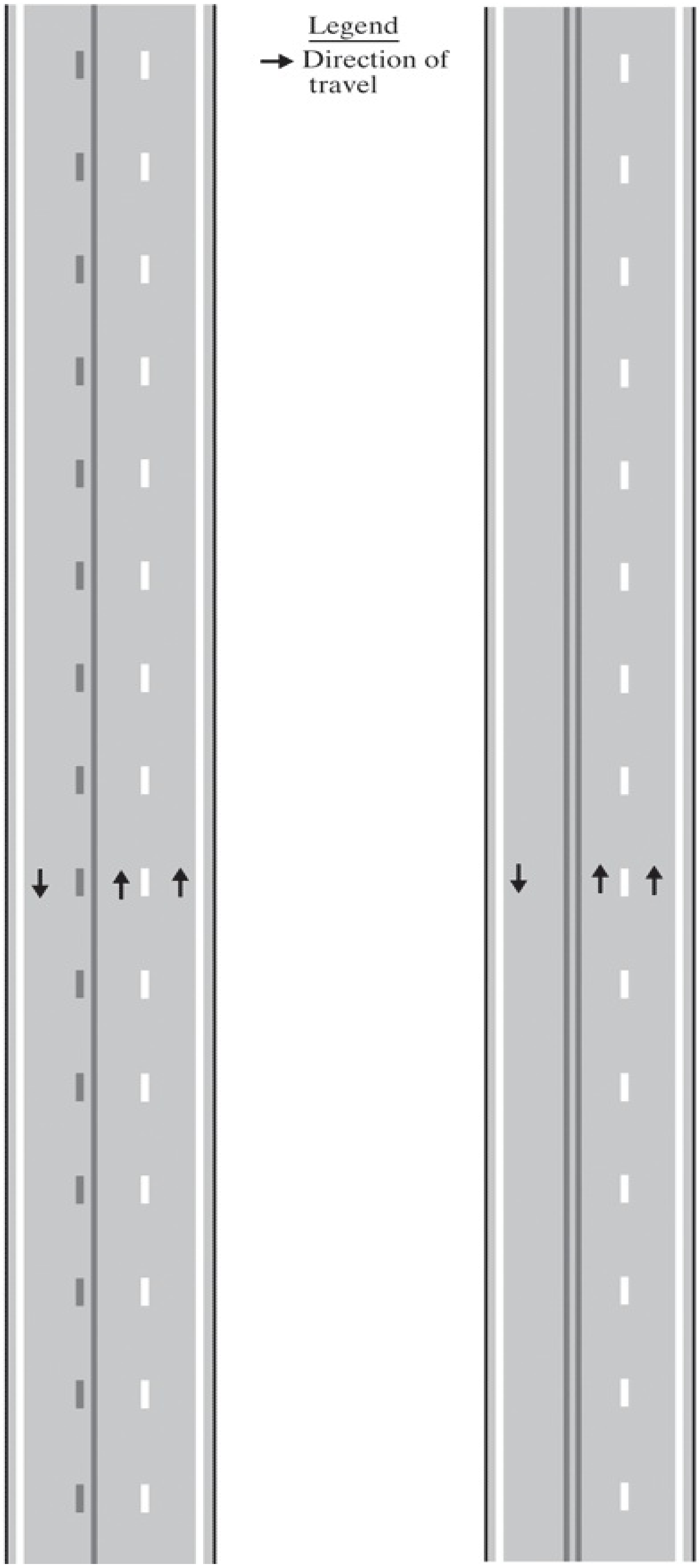
Figure 31.3 shows typical three-lane highway marking patterns, while

Figure 31.4 illustrates transition markings where the direction of the center lane is being reversed.

Figure 31.3: Typical Markings for Three-Lane Rural Highways

A-Typical three-lane, two-way marking with passing permitted in single-lane direction

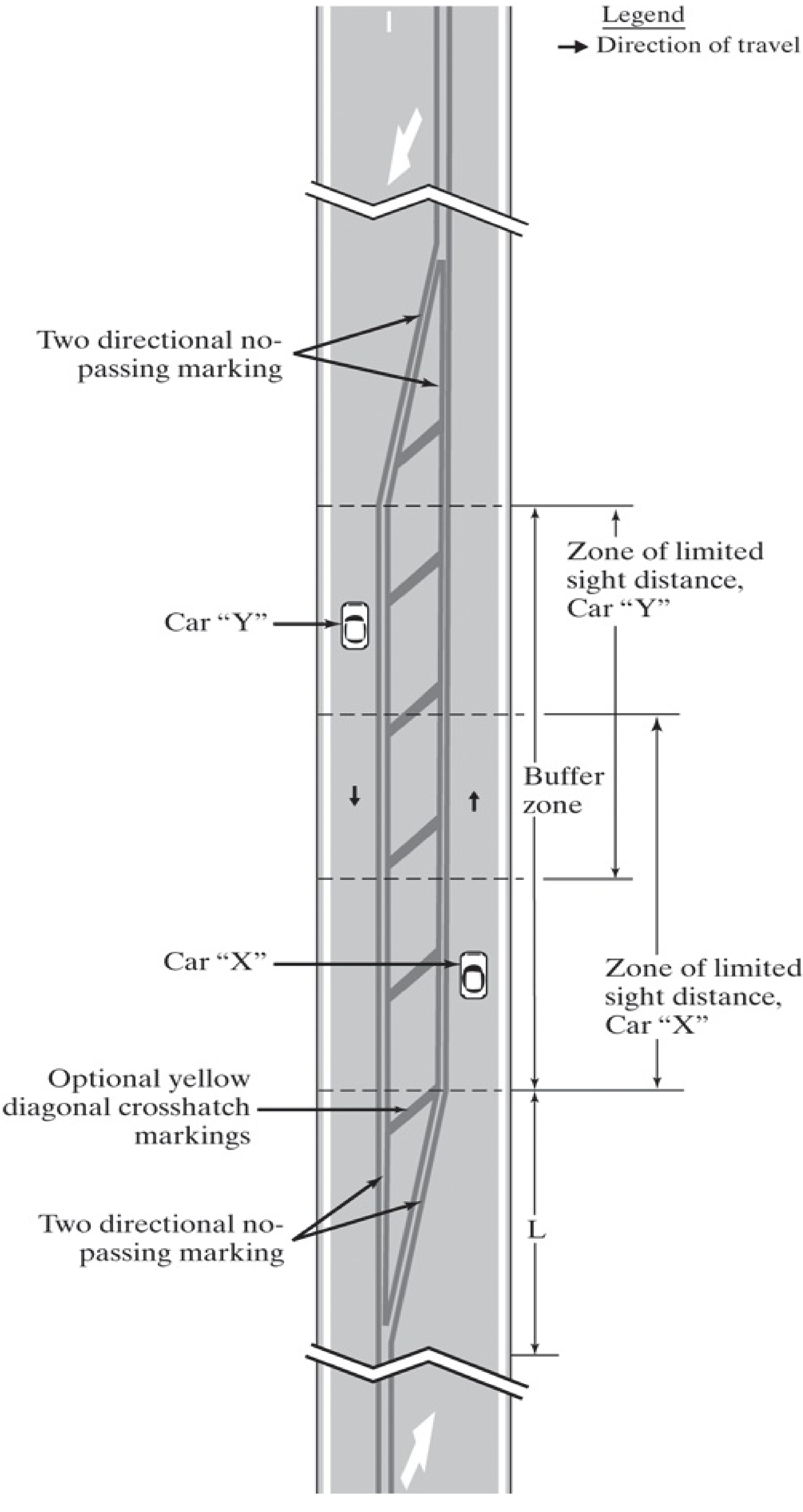
B-Typical three-lane, two-way marking with passing prohibited in single-lane direction



(Source: *Manual of Uniform Traffic Control Devices*, Federal Highway Administration, Washington, D.C., 2009, as amended through 2012, Figure 3B-3, pg 352.)

Figure 31.3: Full Alternative Text

Figure 31.4: Transition Markings for Three-Lane Rural Highways


Legend
Direction of travel
Two directional no-passing marking
Zone of limited sight distance, Car "Y"
Car "Y"
Buffer zone
Car "X"
Zone of limited sight distance, Car "X"
Optional yellow diagonal crosshatch markings
Two directional no-passing marking
L

(Source: *Manual of Uniform Traffic Control Devices*, Federal Highway Administration, Washington, D.C., 2009, as updated through 2012, Figure 3B-5, pg 355.)

Figure 31.4: Full Alternative Text

Note that in Figure 31.3A, passing from the single-lane direction is permitted. The yellow markings, however, clearly indicate that the driver would be entering an opposing lane. Overall, this marking pattern has proven to be generally safe. Where passing opportunities can be provided for both directions relatively frequently, the marking pattern in Figure 31.3B is preferred.

The "buffer zone" in Figure 31.4 must be a minimum of 40 ft in length. The length of the tapered portions of the marking is dependent upon the posted speed limit. Where the speed limit is ≤45 mi/h, the length of the taper is:

L=WS [31-1]

Where the speed limit is <45 mi/h, the length of the taper is:

L=WS260 [31-2]

Where:
L=length of the taper, ft,W=width of the center lane, ft, andS=85th percenti

As a general rule, the minimum length of tapers should be 100 ft in urban areas and 200 ft in rural areas. Buffer zones should be a minimum of 50 ft in length.

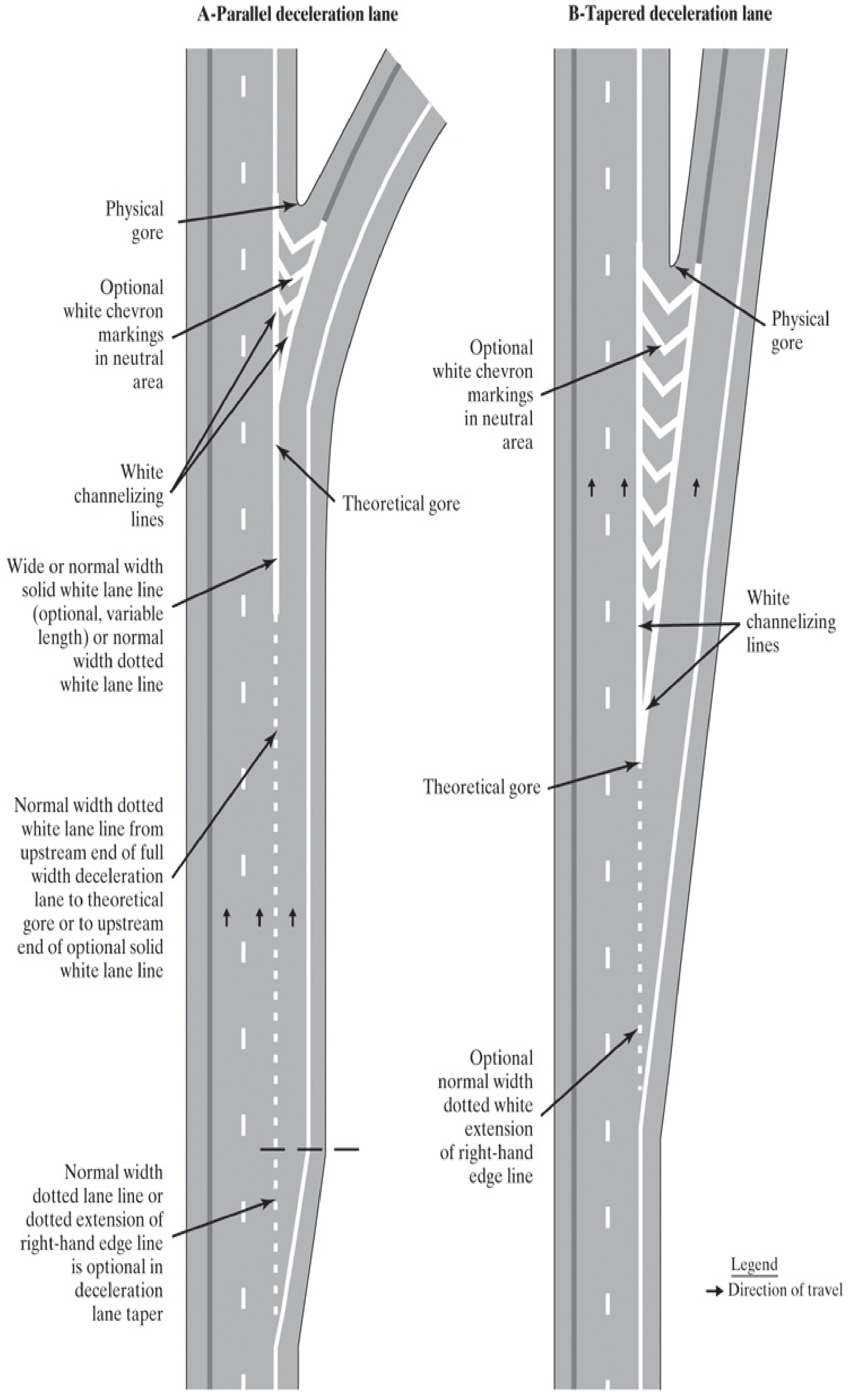
The "zones of limited sight distance" are based upon the sight distances given in Table 31.1, or in accordance with local policy.

31.1.3 Ramp Junction Markings

Ramp junctions occur at all freeway interchanges and on all types of rural highways at locations where grade-separated interchanges are provided. Although not common, some grade-separated interchanges also occur in urban areas on arterials.

Figure 31.5 shows typical markings used for off-ramps; Figure 31.6 shows typical markings used for on-ramps.

Figure 31.5: Typical Off-Ramp Markings

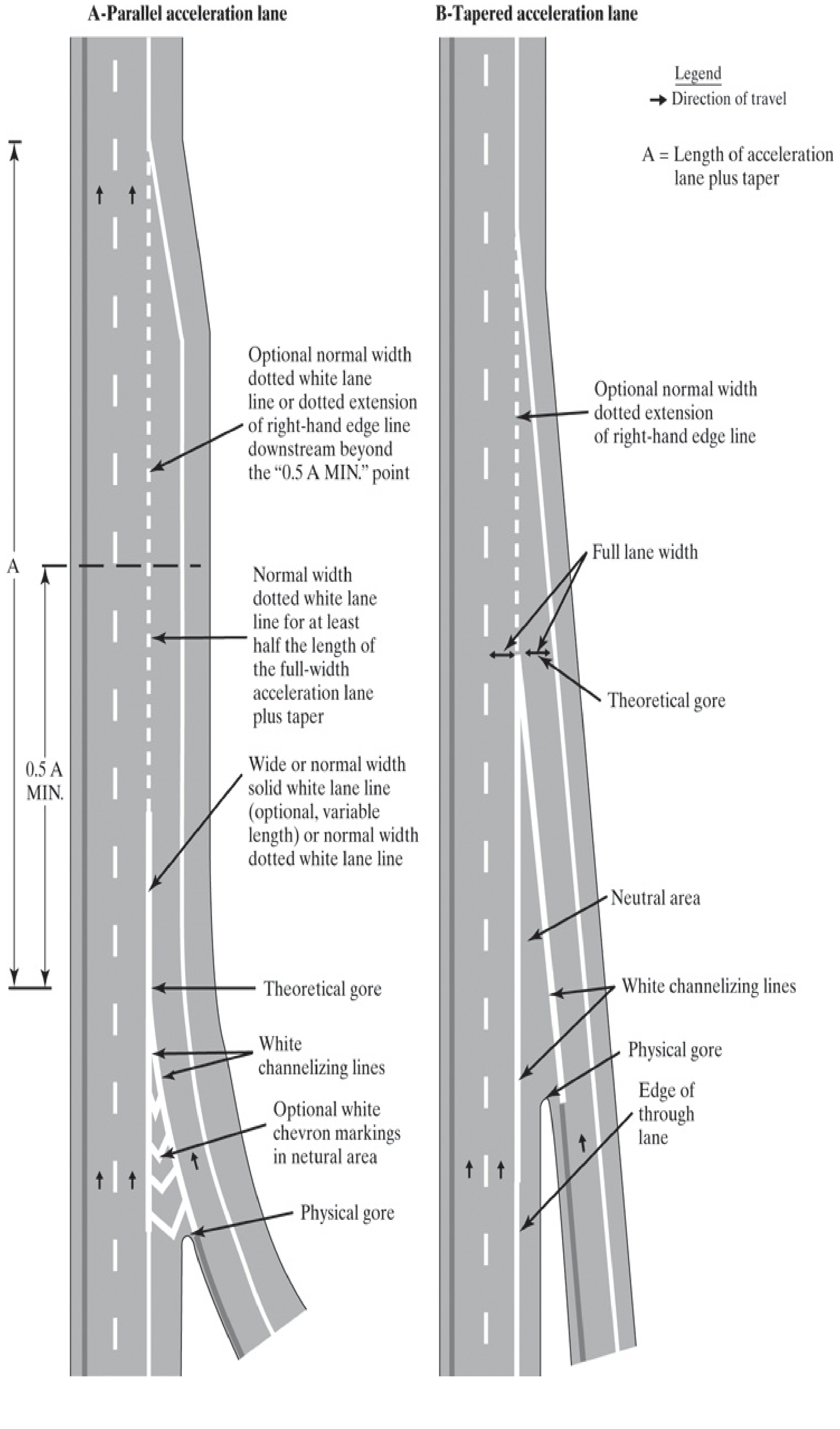


A-Parallel deceleration lane
Physical gore
Optional white chevron markings in neutral area
White channelizing lines
Theoretical gore
Wide or normal width solid white lane line (optional, variable length) or normal width dotted white lane line
Normal width dotted white lane line from upstream end of full width deceleration lane to theoretical gore or to upstream end of optional solid white lane line
Normal width dotted lane line or dotted extension of right-hand edge line is optional in deceleration lane taper
B-Tapered deceleration lane
Physical gore
Optional white chevron markings in neutral area
White channelizing lines
Theoretical gore
Optional normal width dotted white extension of right-hand edge line
Legend
Direction of travel

(Source: *Manual of Uniform Traffic Control Devices*, Federal Highway Administration, Washington, D.C., 2009, as updated through 2012, Figure 3B-8, pg 358.)

Figure 31.5: Full Alternative Text

Figure 31.6: Typical On-Ramp Markings


A-Parallel acceleration lane
B-Tapered acceleration lane
Legend
Direction of travel
A = Length of acceleration lane plus taper
A
0.5 A MIN.
Optional normal width dotted white lane line or dotted extension of right-hand edge line downstream beyond the "0.5 A MIN." point
Normal width dotted white lane line for at least half the length of the full-width acceleration lane plus taper
Wide or normal width solid white lane line (optional, variable length) or normal width dotted white lane line
Theoretical gore
White channelizing lines
Optional white chevron markings in netural area
Physical gore
Optional normal width dotted extension of right-hand edge line
Full lane width
Theoretical gore
Neutral area
White channelizing lines
Physical gore
Edge of through lane

(Source: *Manual of Uniform Traffic Control Devices*, Federal Highway Administration, Washington, DC., 2009, as updated through 2012, Figure 3B-9, pg 360.)

Figure 31.6: Full Alternative Text

Two basic design types are used:

- Parallel acceleration or deceleration lanes
- Tapered acceleration or deceleration lanes

The parallel acceleration/deceleration design is the more common of the two. It affords a merging or diverging driver the ability to select the exact point of the merge or diverge depending upon traffic conditions. Taper designs direct vehicles to a smaller zone in which all merging or diverging maneuvers take place.

As shown in Figures 31.5 and 31.6, a dotted line is used to demark the boundary between the freeway (or highway) lane and the ramp lane. For parallel designs, a solid channelizing line may be used for some distance from the gore area. This is an optional, but frequently used, marking. The channelizing line, in such cases, generally extends about 25%–30% of the distance between the gore area and the beginning/end of the tapered portion of the ramp lane.

The gore area itself is delineated with channelizing lines. The interior of the gore area is often marked with chevron markings at off-ramps (see Figure 31.5). The chevron markings are positioned—with the point facing the approaching driver—to visually guide the driver who encroaches onto the gore area back into the appropriate lane (either the ramp or right lane of the freeway or highway).

31.2 Signing for Freeways and Rural Highways

The majority of signs on freeways and rural highways are to provide directional or route guidance to drivers. Other signs include regulatory signs (primarily speed limits) and warning signs of various types. The use of such signs is discussed in some detail in Chapter 4.

Additional discussion is provided here for route numbering systems and displays, and on conventions for posting directional guidance.

31.2.1 Reference Location Posts

Reference location posts provide a location system along highways on which they are installed. Formerly (and still often) referred to as “mileposts,” they indicate the number of miles along a highway from a designated terminus.

They provide an effective location system for accidents and emergencies, and are frequently used as the basis for exit numbering on freeways and other rural highways.

The numbering system is contiguous within a state. Each state has its own numbering system, so that the numbering sequence begins anew at a state boundary. By convention, mile “0” is located at:

- For north–south highways: the southern state boundary or southern-most point on a highway beginning within a state; and
- For east–west highways: the western state boundary or westernmost point on a highway beginning within a state.

Cardinal directions used in highway designations recognize only two axes: north–south or east–west. Each highway is classified based upon the general direction of the route within the state, as defined by its endpoints.

- If a straight line connecting the endpoints of a highway forms a line that is between 45° and 135° from the horizontal, it is classified as a north–south highway.
- If the straight line so defined forms an angle of <45˚ or >135˚, it is classified as an east–west highway.

Note that a north–south highway may contain individual sections that are oriented in the east–west direction and vice-versa. The cardinal direction designation refers only to the overall orientation of the facility.

The MUTCD *mandates* that reference points be placed on all freeway facilities and on expressways that are “located on a route where there is reference post continuity.” They *may* also be placed on all other classes of rural highways.

Reference posts are placed every mile along the route. Intermediate or enhanced reference posts may be placed every 0.10 mile. Figure 31.7 illustrates mileposts and tenth-mileposts.

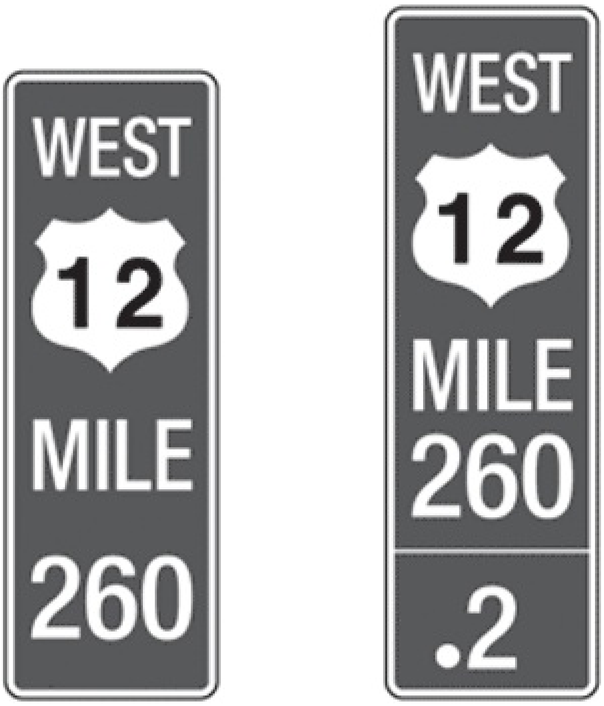
Figure 31.7: Reference Location Posts

(a) Example Mileposts

31.2-2 Full Alternative Text

(b) Example Tenth-Mileposts

31.2-2 Full Alternative Text



(c) Enhanced Milepost and Tenth-Mileposts

(Source: *Manual on Uniform Traffic Control Devices*, Federal Highway Administration, Washington, D.C., 2009, as amended through 2012, Figures 2H-2, 2H-3, and 2H-4, pgs 295 and 296.)

31.2-2 Full Alternative Text

31.2.2 Route Numbering Systems and Route Signs

There are four numbered highway systems within the United States:

- The Eisenhower National System of Interstate and Defense Highways (the Interstate System)
- The U.S. Route System
- State Highway Systems
- County Road Systems

The Interstate and U.S. highway systems are numbered by AASHTO, based upon recommendations from individual state highway departments in accordance with published policies [3,4]. State and county road systems are numbered by the agency with jurisdiction in accordance with standards and criteria established by each state.

The oldest numbered highway system in the country is the U.S. Route System. The system originally resulted from a series of meetings between the American Association of State Highway Officials (AASHO, the forerunner of AASHTO) and representatives of state highway agencies. The meetings took place between 1923 and 1927, culminating in the establishment of the U.S. Route System on November 11, 1926 [5].

Before the U.S. Route System, a loose network of named national routes (such as the Lincoln Highway) existed, with each named route sponsored by private organizations and motorists clubs. The initial U.S. Route System was intended to replace these with a more orderly system, and was expected to encompass approximately 50,000 miles of rural highways. By the time the system was inaugurated, it contained more than 75,000 miles of roadways.

The U.S. numbering systems followed these guidelines:

- Principal north–south routes were assigned one- or two-digit numbers ending in “1.”
- North–south routes of secondary importance were assigned one- or two-digit numbers ending in “5.”
- Transcontinental and principal east–west routes were assigned numbers in multiples of 10.
- Numbers of all principal and secondary routes were to be in numerical order from east to west and from north to south.
- Branch routes were assigned three-digit numbers with the last two representing the principal route to which they connected.

Some of these conventions were adopted or modified for the Interstate System route designations:

- All primary east–west routes have one- or two-digit even numbers.
- All primary north–south routes have one- or two-digit odd numbers.
- All branch routes have three-digit numbers, with the last two indicating the primary route to which they connect.

The last feature leads to multiple routes having the same three-digit route number. Interstate routes I-495 and I-695, for example, exist in several different places, but all connect to I-95, the principal north–south interstate on the east coast.

Figure 31.8 illustrates the east coast interstate system between New York and Washington, DC.

Figure 31.8: The Interstate System Serving the Northeast Between New York and Washington D.C.

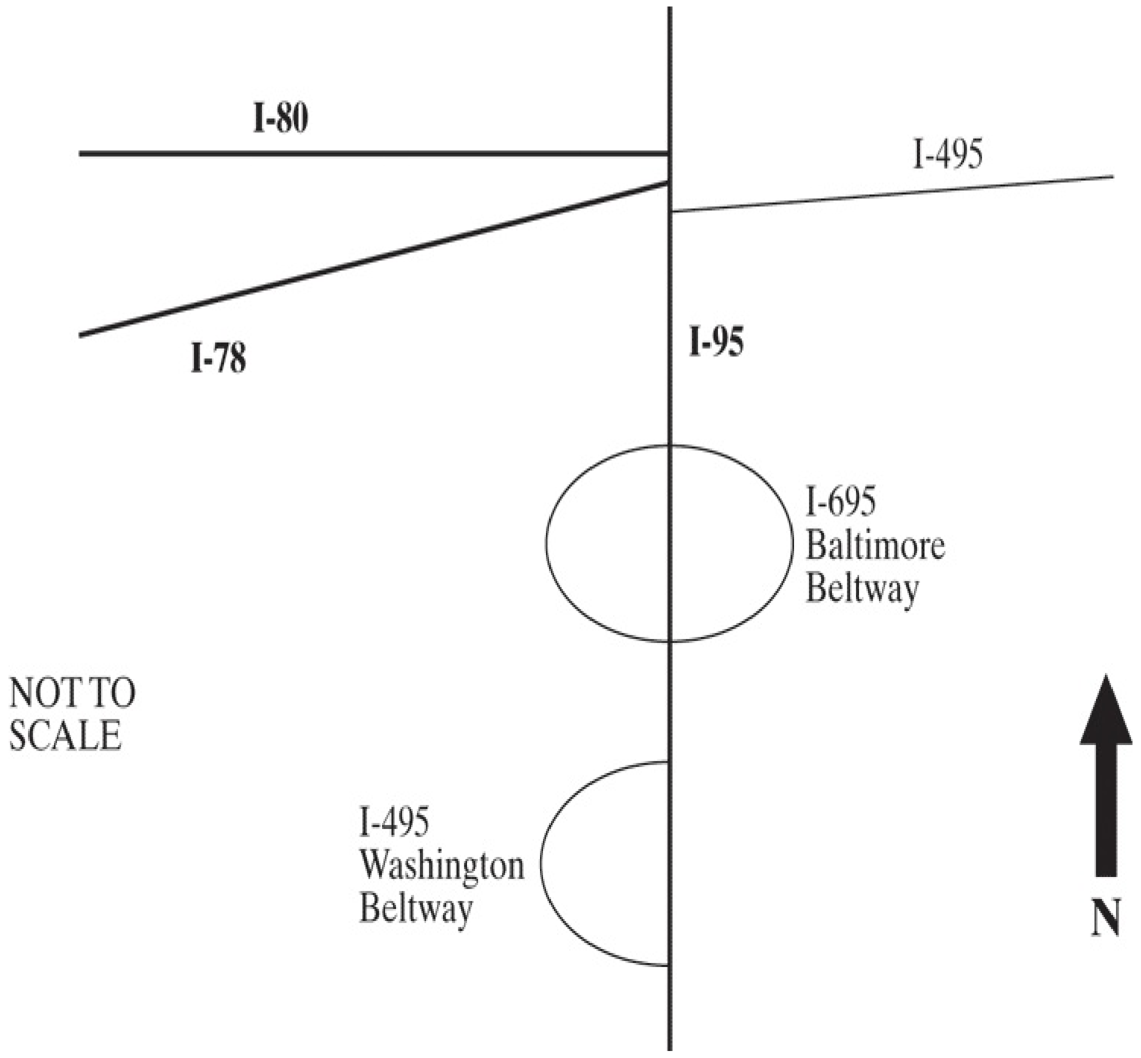


Figure 31.8: Full Alternative Text

Numbered routes are identified by the appropriate shield bearing the route number, with an auxiliary panel indicating the cardinal direction of the route. Standard shield designs are illustrated in Figure 31.9. The Interstate and U.S. systems each have a standard design used throughout the country. Each state has a uniquely designed shield used within it. County road shields are the same throughout the nation, but the name of the county appears as part of the shield design.

Figure 31.9: Route Marker Shields

(Source: *Manual on Uniform Traffic Control Devices*, Federal Highway Administration, Washington, D.C, 2009, as updated through 2012, Figure 2D-3, pg 143.)

Figure 31.9: Full Alternative Text

When numbered routes converge, both route numbers are signed using the appropriate shields. All route shields are posted at common locations. Because cardinal destinations define the general direction of the route, it is possible to have a given segment of highway with multiple route numbers and different cardinal directions.

For example, a section of the New York State Thruway (a north–south route, I-87) is convergent with a section of the Cross-Westchester Expressway (an east–west route, I-287). In one direction, this section is both “I-87 North” and “I-287 West.”

31.2.3 Interchange Numbering

Systems

On freeways and some expressways, interchanges are numbered using one of two approaches:

- Milepost Numbering: The exit number is the milepost number closest to the interchange.
- Sequential Numbering: Exits are sequentially numbered, with Exit 1 beginning at the westernmost or southernmost interchange within the state.

Milepost numbering is now the preferred system according to the MUTCD, and many, if not most, states have converted from sequential interchange numbering to the milepost system. Milepost numbering provides the driver with more information. With a known exit number and the nearest milepost, a driver can estimate the distance to his/her desired destination. Milepost numbering has another distinct advantage: new interchanges can be added to the facility without disrupting other exit numbers.

Where routes converge, mileposts and exit numbers are continuous for only one route. In terms of hierarchy, Interstate routes take precedence over all other systems, followed (in order) by U.S. routes, state routes, and county routes. Where two routes of equal precedence converge, the primary highway will take precedence over a secondary highway. Where two routes have exactly the same precedence, the highway agency with jurisdiction makes the decision on which mileposts and exit numbers will be continuous.

It should be noted that all route and exit numbering systems can be incorporated into guide signs to provide the driver with additional information.

31.2.4 Route Sign Assemblies

A route sign assembly is any posting of a single or multiple route number signs. Where numbered routes converge, diverge, or intersect, the proper design and display of route sign assemblies is a critical element of directional guidance. Figure 31.10 illustrates a typical case of numbered routes converging at the entrance to a town and diverging at the exit. Drivers must be given clear directions on how to follow a given route through the town, as well as where to turn to access any one of the intersecting routes.

Figure 31.10: The Importance of Route Sign Assemblies

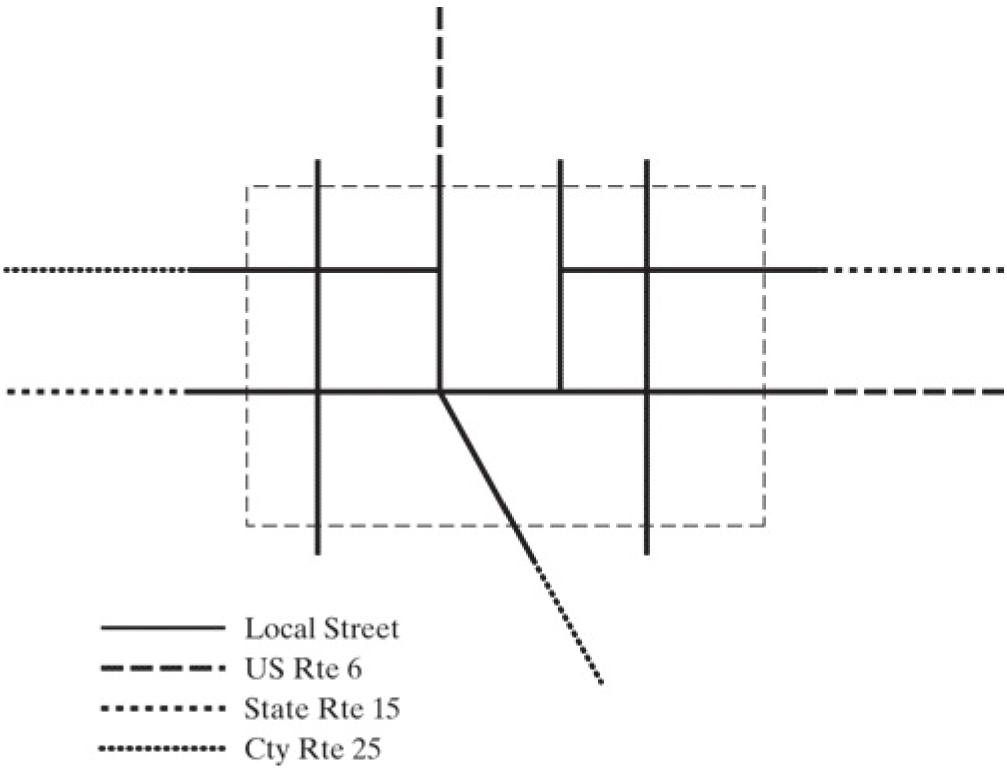


Figure 31.10: Full Alternative Text

Figure 31.10 is typical of many rural communities. Numbered highways and roads provide access to the community, but also merge into the local street system as they pass through. It may be necessary for a driver to make multiple turns going through the community to follow his/her desired route.

The MUTCD defines five types of route sign assemblies:

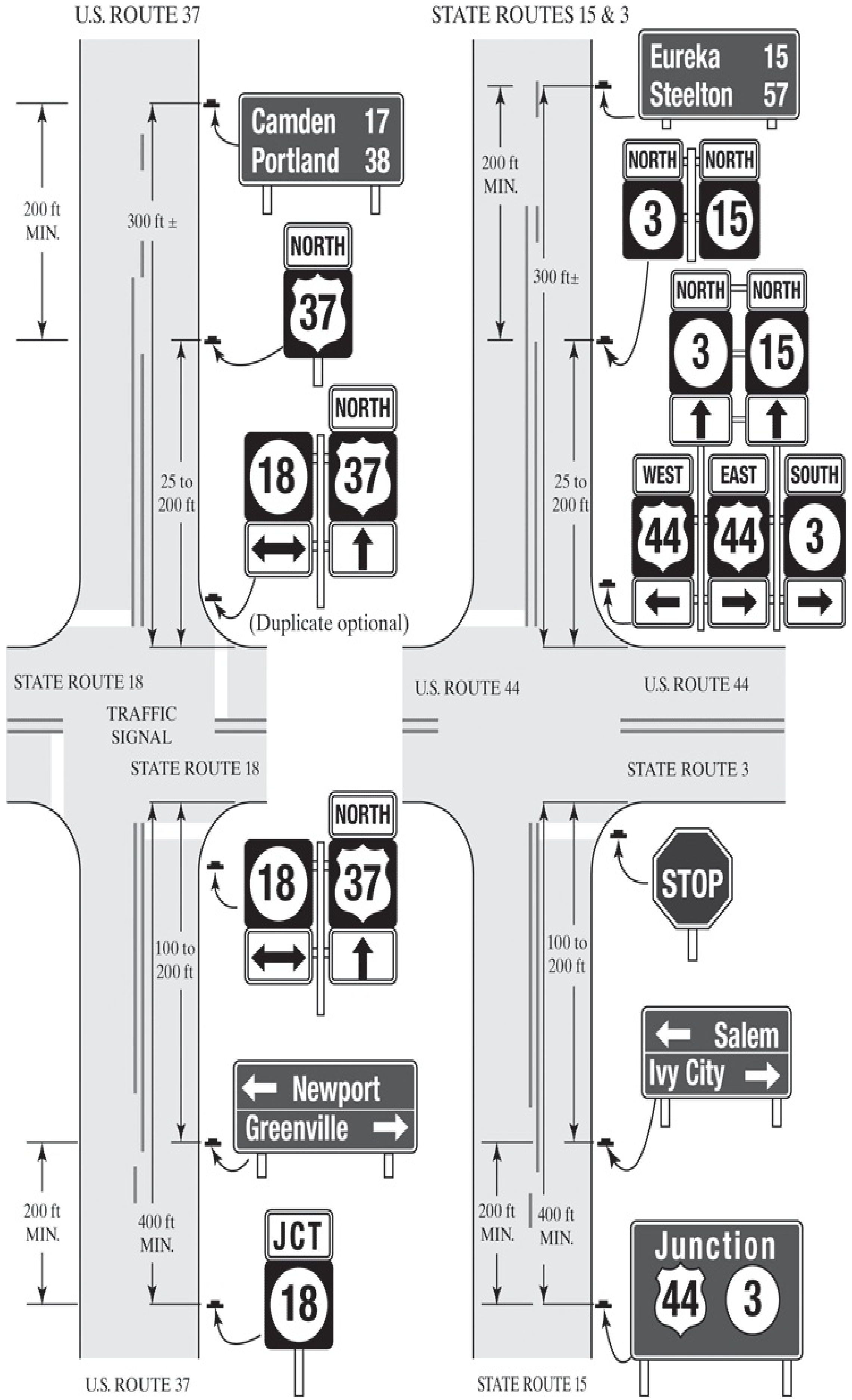
- Junction Assembly: Used to indicate an upcoming intersection with another numbered route(s).

- Advance Route Turn Assembly: Used to indicate that a turn must be made at an upcoming intersection to remain on the indicated route.
- Directional Assembly: Used to indicate required turn movements for route continuity at an intersection of numbered routes, as well as at the beginning or end of numbered routes.
- Confirming or Reassurance Assemblies: Used after motorists have passed through an intersection of numbered routes. Within a short distance, such an assembly assures the motorist that he or she is on the intended route.
- Trailblazer Assemblies: Used on non-numbered routes that lead to a numbered route. The “To” auxiliary panel is used in conjunction with the route shield of the numbered route.

The MUTCD gives relatively precise guidelines on the exact placement and arrangement of route sign assemblies. It should be consulted directly for these details.

Figure 31.11 shows two typical examples of the use of route sign assemblies. Both show signing in only one direction. Each approach to the intersection would have similar signing.

Figure 31.11: Typical Use of Route Sign Assemblies


U.S. ROUTE 37
STATE ROUTES 15 & 3
Camden 17
Portland 38
200 ft MIN.
300 ft ±
NORTH
37
25 to 200 ft
NORTH
18
37
(Duplicate optional)
Eureka 15
Steelton 57
200 ft MIN.
NORTH NORTH
3 15
300 ft±
NORTH NORTH
3 15
25 to 200 ft
WEST EAST SOUTH
44 44 3
STATE ROUTE 18
TRAFFIC SIGNAL
STATE ROUTE 18
U.S. ROUTE 44
U.S. ROUTE 44
STATE ROUTE 3
NORTH
18
37
100 to 200 ft
Newport
Greenville
200 ft MIN.
400 ft MIN.
JCT
18
STOP
100 to 200 ft
Salem
Ivy City
200 ft MIN.
400 ft MIN.
Junction
44 3
U.S. ROUTE 37
STATE ROUTE 15

(Source: *Manual of Uniform Traffic Control Devices*, Federal Highway Administration, Washington, D.C., 2009, as updated through 2012, Figure 2D-6, pg 149.)

Figure 31.11: Full Alternative Text

Both examples are for drivers approaching from the south. Assemblies are encountered in the following order:

- The first assembly in both cases is the junction assembly. Note that two different styles are used in the examples. The first example uses the more prevalent display, and is simpler because there is only one intersecting route. The second example shows a larger sign, and is used to emphasize that the approaching junction will be with two numbered routes. Either design may be used.
- The next signs encountered are typical directional guidance signs (for conventional roads), and are not classified as route sign assemblies.
- The third signs encountered are directional assemblies. The standard location for these is on the far side of the intersection. In one of the illustrations, a duplicate is provided on the near side of the intersection.
- Once the intersection is crossed, a confirming assembly is posted within 200 ft of the intersection, to assure drivers that they are indeed on their desired route.
- The last signs shown are destination distance signs, which are optional, and are not classified as route sign assemblies.

Note that in neither case is an advance turn assembly used, as drivers *do not* have to make a turn at the intersection to follow the same numbered route.

The MUTCD contains numerous other examples of route sign assemblies and their use, and should be consulted for additional guidance.

31.2.5 Freeway and Expressway

Guide Signing

Freeways and most expressways have numbered exits and mileposts, and guide signing is keyed to those features. As noted in Chapter 4, guide signs are rectangular, with the long dimension horizontal. Backgrounds are color-coded based upon the type of guidance being offered:

- Green: general directional guidance.
- Blue: service information.
- Brown: historical/recreational locations.

Directional guide signs, however, make up the majority of these signs, and are the most important in assuring safe and unconfused operation of vehicles. A typical directional guidance sign is illustrated in Figure 31.12. In this case, it is a sign that would be posted right at the exit.

Figure 31.12: A Typical Freeway Directional Guide Sign

(Source: *Manual of Uniform Traffic Control Devices*, Federal Highway Administration, Washington, D.C., 2009, as amended through 2012, Figure 2E-6, pg 220.)

Figure 31.12: Full Alternative Text

The sign provides a great deal of information:

- The ramp leads to U.S. Route 56 West (the U.S. shield is used).
- The primary destination reached by selecting this route is the city or town of Utopia.
- The exit number is 211A. Assuming that this is a milepost-numbered exit, the exit is located at milepost 211. Exit 211A refers to the ramp leading to U.S. 56 West; there is doubtless an Exit 211B that would lead to U.S. 56 East.
- The exit tab is located on the right side of the sign, indicating that this is a right-hand exit. For left-hand exits, the tab would be located on the left side of the sign.

In general, drivers should be given as much advance warning of interchanges and destinations as possible. The cardinal rule is "a confused driver is a dangerous driver." However, applying this approach leads to very different guide signing in urban and rural environments.

In rural areas, advance signing is much easier to accomplish because there are long distances between interchanges. Figure 31.13 shows a typical signing sequence for rural areas.

Figure 31.13: A Typical Guide Sign Sequence in a Rural Area

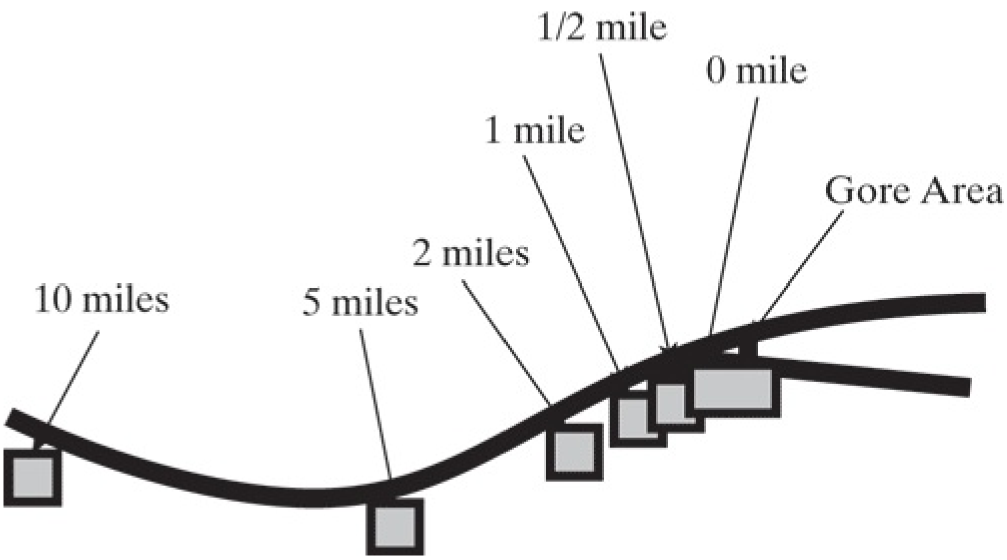


Figure 31.13: Full Alternative Text

The first directional guide sign can be as far as 10 miles away, assuming that there are no other exits between the sign and the subject interchange. If a 10-mile advance sign is placed, the typical sequence would have additional signs placed at 5 miles, 2 miles, 1 mile, and 12 mile, and at the ramp itself. Where distance between ramps permits, the first advance sign should be *at least* 2 miles from the exit. Where distances between ramps are *less than* 2 miles, the first advance sign would be placed shortly beyond the location of the previous interchange. This is because placing specific advance signs for an exit are very confusing if they are placed “out of sequence.”

At the point of the exit ramp, a large sign of the type shown in Figure 31.12 is placed. The sign is generally mounted on a sign support located on the right side (of a right-hand ramp), cantilevered out over, or just short of, the gore area. It is also typical to place a small sign of the type shown in Figure 31.14 in the gore area, always mounted on a break-away sign support for safety.

Figure 31.14: Gore Area Exit Sign with Speed Warning Panel

(Source: *Manual on Uniform Traffic Control Devices*, Federal Highway Administration, Washington, D.C., 2009, as amended through 2012, Figure 2E-28, pg 222.)

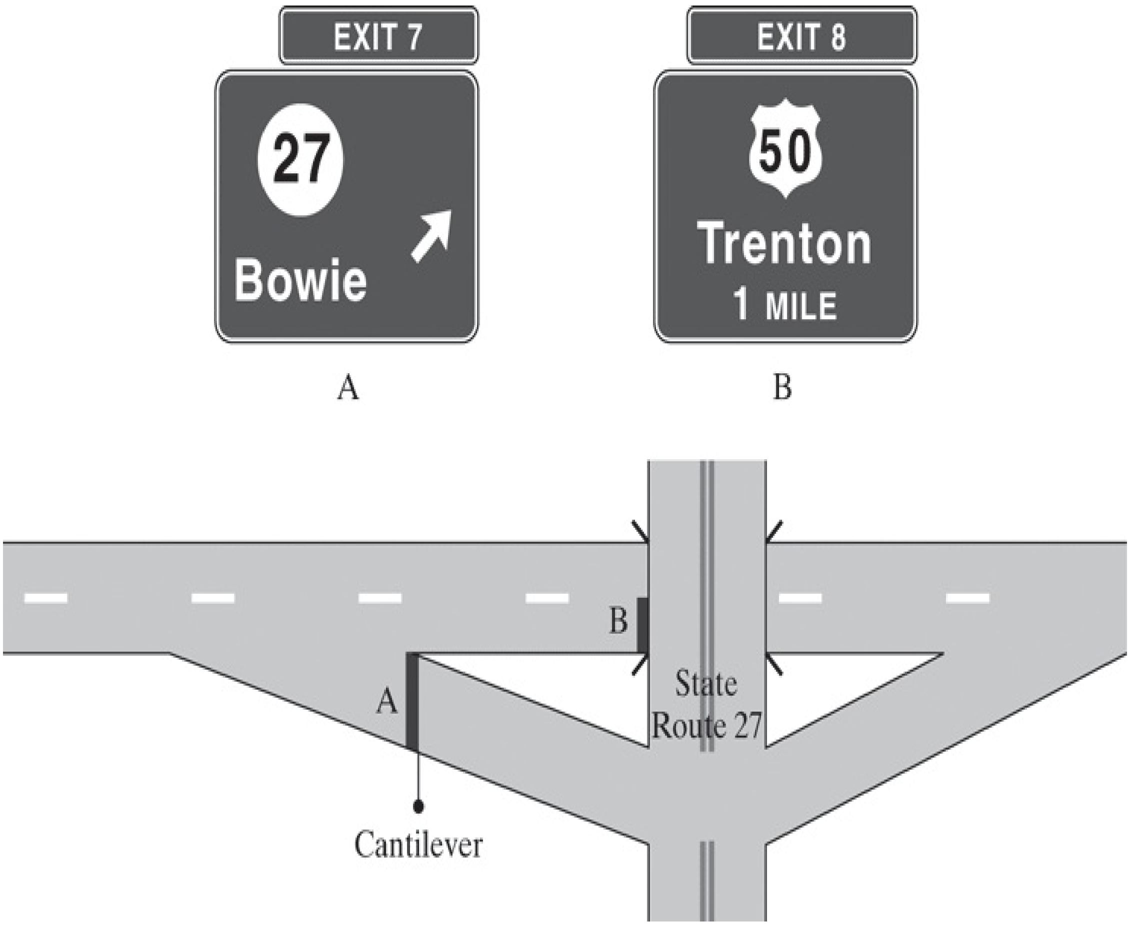
Figure 31.14: Full Alternative Text

The exit sign of Figure 31.14 may be used with or without the speed warning panel. In general, if the safe speed on the ramp is more than 10–15 mi/h slower than the main roadway, use of a warning panel is appropriate.

In urban areas, guide signing must be done very carefully. Because interchanges tend to be closely spaced, there is little opportunity to place advance exit signs. In many cases, the first, and perhaps only, advance exit sign must be placed near the location of the previous interchange. To avoid confusion at the point of the previous exit, a technique known as “sign spreading” is used to separate the signs at the previous exit from the advance sign for the upcoming exit. Figure 31.15 illustrates this concept.

Figure 31.15: Illustration of Sign Spreading at an Urban

Interchange



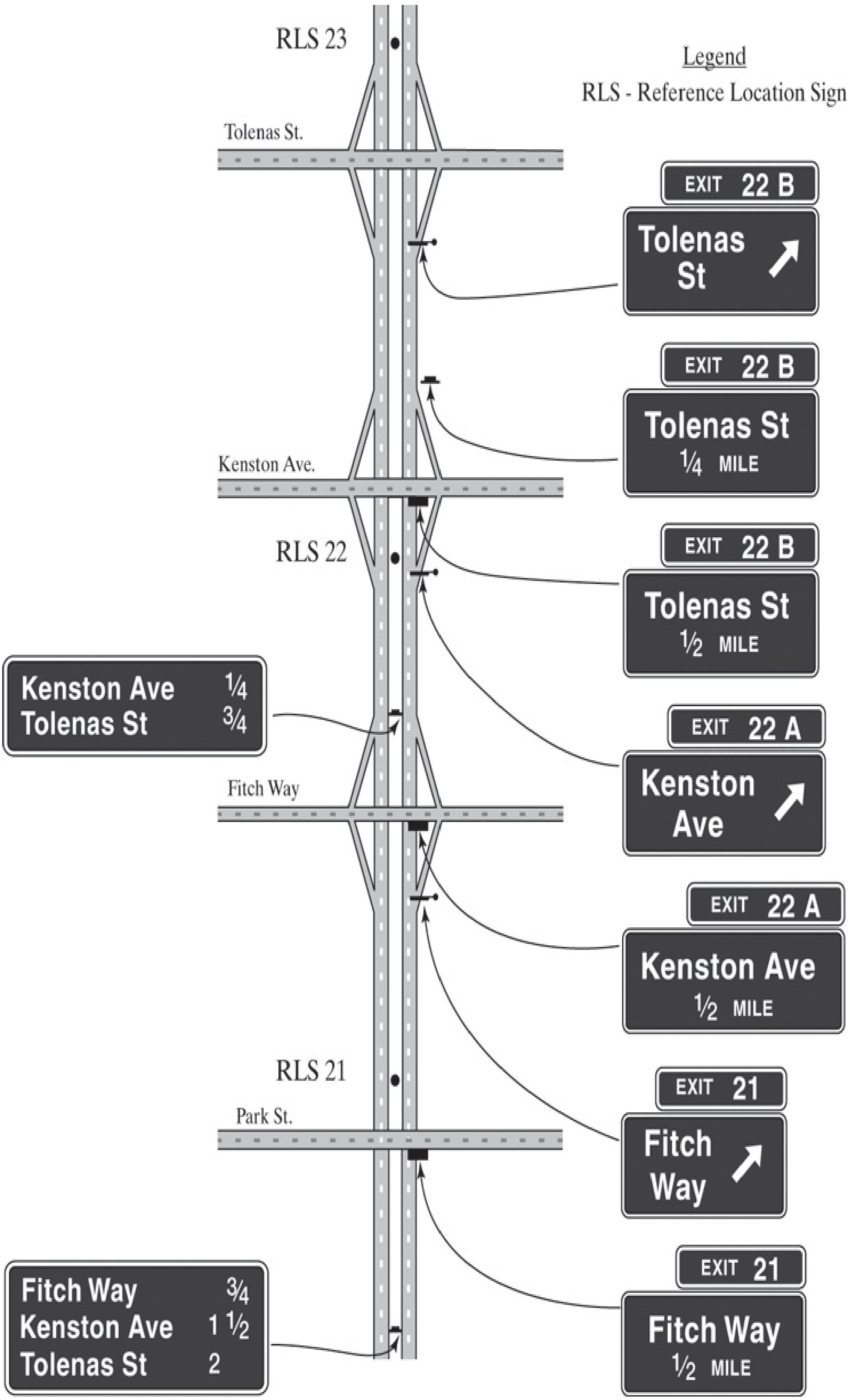
(Source: *Manual on Uniform Traffic Control Devices*, Federal Highway Administration, Washington, D.C., 2009, as amended through 2012, Figure 2E-1, pg 184.)

Figure 31.15: Full Alternative Text

Note that the Exit 7 sign is placed on an overhead cantilevered support over the gore area. The advance sign for Exit 8 is *not* placed at the same location, but is on another overhead support (in this case an overpass), a short distance beyond Exit 7.

Figure 31.16 shows the guide signing for a series of closely spaced exits on a section of urban freeway. Distances are such that only one advance sign is used for each exit, and each is displayed using the sign spreading technique. Sequential exit numbers are used, possibly because there are more than one exit that would be closest to a single milepost. Exits numbered 22A and 22B are separate interchanges. It is likely that the A and B designations were made necessary when an additional exit was introduced on the facility.

Figure 31.16: Guide Signing for Closely-Spaced Urban Interchanges


Legend
RLS - Reference Location Sign
RLS 23
Tolenas St.
EXIT 22 B
Tolenas St
EXIT 22 B
Tolenas St 1/4 MILE
Kenston Ave.
RLS 22
EXIT 22 B
Tolenas St 1/2 MILE
Kenston Ave 1/4
Tolenas St 3/4
EXIT 22 A
Kenston Ave
Fitch Way
EXIT 22 A
Kenston Ave 1/2 MILE
RLS 21
Park St.
EXIT 21
Fitch Way
EXIT 21
Fitch Way 1/2 MILE
Fitch Way 3/4
Kenston Ave 1 1/2
Tolenas St 2

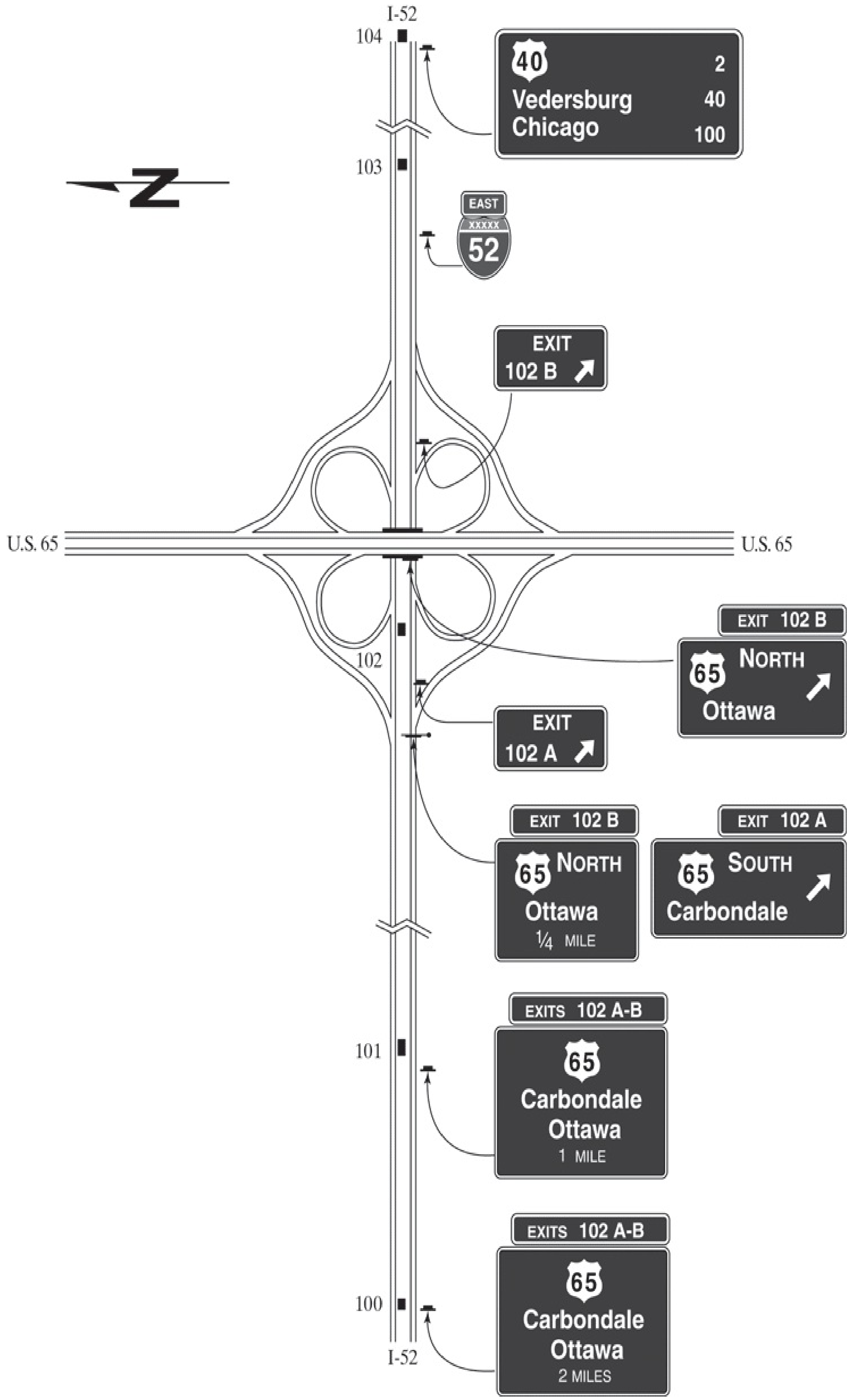
(Source: *Manual on Uniform Traffic Control Devices*, Federal Highway Administration, Washington, D.C., 2009, as amended through 2012, Figure 3E30, pg 224.)

Figure 31.16: Full Alternative Text

Note that there are additional signs, called "exit sequence" signs, which indicate the distance to multiple upcoming interchanges. Note that no exit numbers are included on these, to avoid any confusion in the numbering sequence. Such signs are used to provide additional advance warning of upcoming exits, without introducing confusing overlapping number sequences.

Figure 31.17 illustrates guide signing for a typical cloverleaf interchange between two freeways. The figure shows only *one quadrant* of the interchange (for NB vehicles). Similar signing would be needed for SB, EB, and WB approaches. Complex interchange signing can become very expensive. The one quadrant shown in Figure 31.17 involves placement of *eight* signs, some of which would require costly sign support structures. Signing the full interchange would require approximately 32 signs, with numerous sign support structures.

Figure 31.17: Guide Signing for One Quadrant of a Typical Cloverleaf Interchange

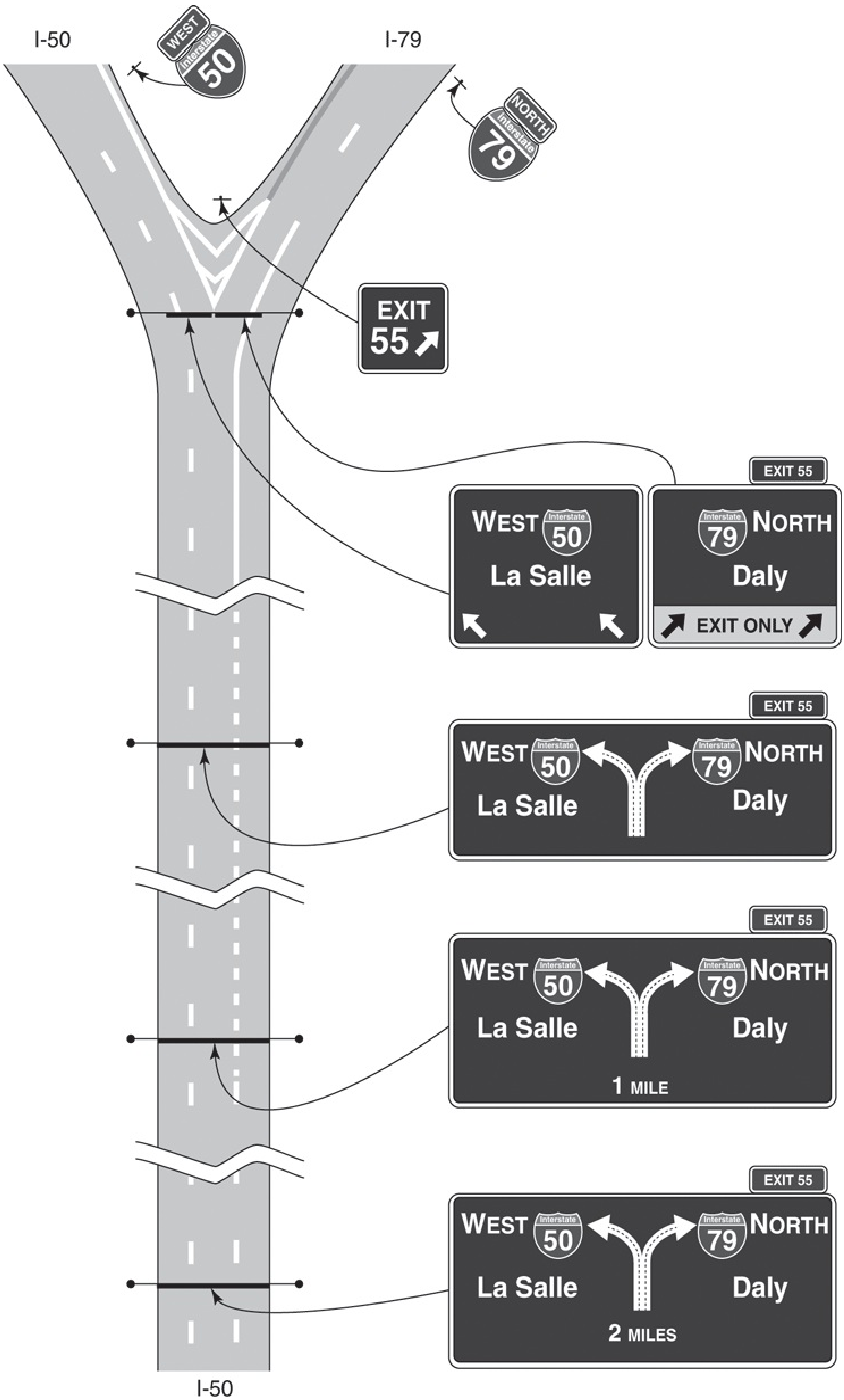

I-52
104
40
2
Vedersburg
40
Chicago
100
103
EAST
52
EXIT
102 B
U.S. 65
U.S. 65
EXIT 102 B
65
NORTH
Ottawa
102
EXIT
102 A
EXIT 102 B
65
NORTH
Ottawa
1/4 MILE
EXIT 102 A
65
SOUTH
Carbondale
EXITS 102 A-B
65
Carbondale
Ottawa
1 MILE
101
EXITS 102 A-B
65
Carbondale
Ottawa
2 MILES
100
I-52

(Source: *Manual of Uniform Traffic Control Devices*, Federal Highway Administration, Washington, D.C., 2009, as amended through 2012, Figure 2E-36, pg 231.)

Figure 31.17: Full Alternative Text

Figure 31.18 illustrates the signing for a major diverge of two interstate routes—I-50 West and I-79 South. Advance signs are placed at 2 miles, 1 mile, and ½ miles. These signs are diagrammatic, and clearly show how the three approach lanes split at the diverge point. This allows approaching drivers to move into the appropriate lane well in advance of the gore area.

Figure 31.18: Diagrammatic Signing for a Major Diverge of Interstate Routes



I-50
I-79
WEST Interstate 50
NORTH Interstate 79
EXIT 55
EXIT 55
WEST 50 La Salle
NORTH 79 Daly
EXIT ONLY
EXIT 55
WEST 50 La Salle
NORTH 79 Daly
EXIT 55
WEST 50 La Salle
NORTH 79 Daly
1 MILE
EXIT 55
WEST 50 La Salle
NORTH 79 Daly
2 MILES
I-50

(Source: *Manual of Uniform Traffic Control Devices*, Federal Highway Administration, Washington, D.C., 2009, as amended through 2012, Figure 2E-10, pg 202.)

Figure 31.18: Full Alternative Text

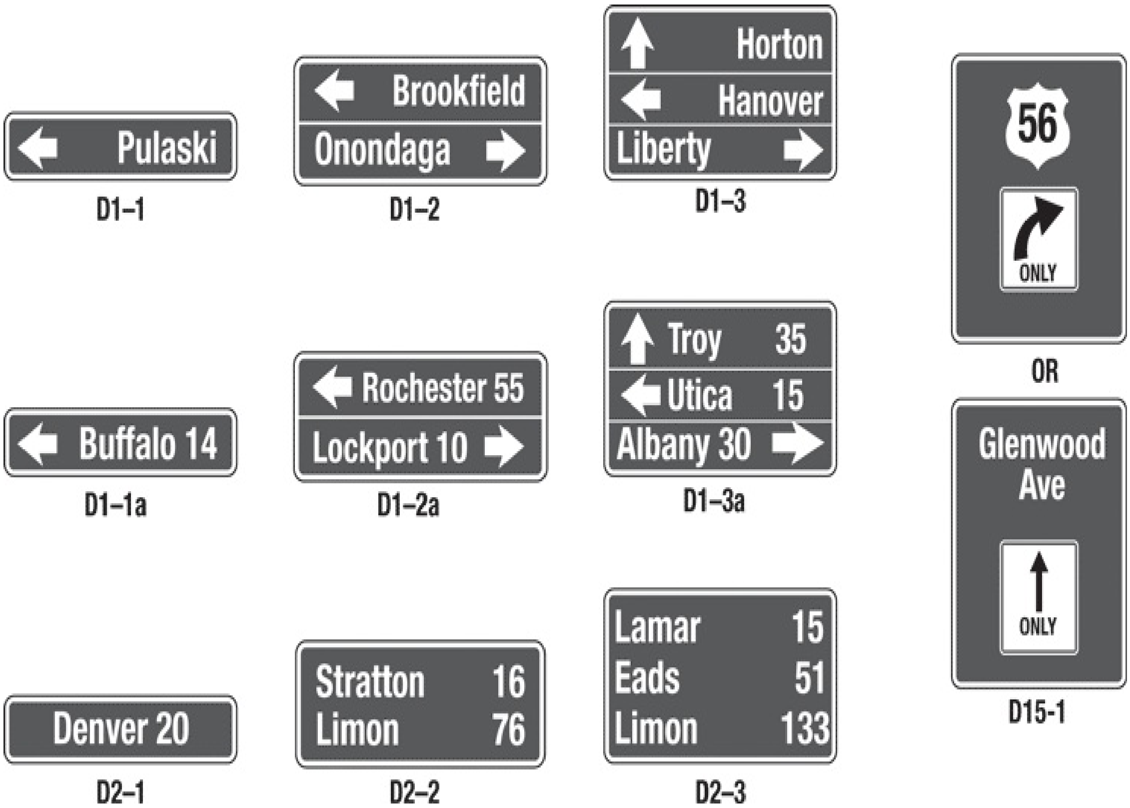
The last sign at the diverge point uses arrows to indicate which two lanes can be used for each numbered route.

The general principle for destination guide signing is to keep the message as simple and easy-to-understand as possible. Consult Chapter 4 for additional guidelines and illustrations of the application of service information and historical/recreational signing.

31.2.6 Guide Signs for Conventional Roads

Guide signing for roads other than freeways and expressways primarily consists of route sign assemblies, as previously discussed, and destination signing. Numbered routes may or may not be involved on such facilities. Where numbered routes are not in play, destination names become the primary means of conveying information. Advance destination signs are generally placed at least 200 ft from an intersection, with confirming destination signs located shortly after passing through an intersection. A sample of guide signs for conventional roads is shown in Figure 31.19.

Figure 31.19: Destination and Distance Signs for Use on Conventional Roadways



(Source: *Manual on Uniform Traffic Control Devices*, Federal Highway Administration, Washington, D.C., 2009, as amended through 2012, Figure 2D-7, pg 155.)

Figure 31.19: Full Alternative Text

31.2.7 Warning Signs on Rural Highways

Consult Chapter 4 for a complete discussion of warning signs and their application. On freeways and expressways, the most frequently used warning signs are for animal crossings, unexpected changes in alignment, or changes in speed limits. On conventional rural roads, a full range of warning signs would be applied as needed.

31.3 Establishing and Posting of Speed Limits on Rural Roads

Chapter 4 contains a brief discussion of speed limits in general, as well as of signing protocols. On freeways, expressways, and other rural highways, speed limits are generally of the linear type, that is, applying to a specified section of a designated highway. In some local rural communities, an area speed limit may be in effect for local roads.

The MUTCD requires the posting of linear speed limit signs at:

- Points of change from one speed limit to another, and
- Beyond major intersections and at other locations where it is necessary to remind drivers of the speed limit that is applicable.

In practical terms, the latter requirement is generally interpreted to mean speed limit signs should be located at points of change, and within 1,000 ft of major entry locations. “Major” entry locations would include all on-ramps on freeways and expressways, and significant at-grade intersections on other rural highways.

Where the state statutory speed limit is in effect, signs should be periodically posted reminding drivers of this fact. Signs would be posted according to the same rules cited above for linear speed limits (technically, a state speed limit is an area speed limit). A general guideline is that in the absence of major entry points, speed limit signs should be placed at least every mile. At borders between states, signs indicating the statutory speed limit of the state being entered would be placed and/or any change in linear speed limit.

The setting of an appropriate speed limit for a freeway, expressway, or rural highway calls for much judgment to be exercised. The general philosophy applied to setting speed limits is that the majority of drivers are not suicidal. They generally will, with no controls imposed, tend to select a range of speeds that is safe for the conditions that exist. Using this approach, speed limits are often set at the 85th percentile speed of free-flowing traffic, rounded to the nearest 5 mi/h.

The traffic engineer, however, must also take into account other factors that might make it prudent to set a lower speed limit. Most of these would involve conditions that drivers might have difficulty discerning, including the following:

- Design speed of the facility section (*no* speed limit should ever exceed the design speed of the facility)
- Details of the roadway geometry, including sight distances
- Roadside development intensity and roadside environment
- Accident experience
- Observed pace speeds—the 10 mi/h speed increment with the highest percentage of drivers

While speed limits below the 85th percentile speed may be required for safety, it should be recognized that such speed limits are more difficult to enforce. Enforcement vigilance is usually required to obtain significant observance.

There are a number of types of speed limits that may be applied. In addition to the primary speed limit, additional speed limits may be applied:

- Truck speed limits
- Night speed limits
- Minimum speed limits

Truck speed limits only apply to trucks (as defined in each state’s vehicle and traffic code). They are generally introduced in situations where the operation of trucks at the primary speed limit involves safety issues. This generally occurs where there are many long downhill segments that could lead to “runaway” trucks if they are operating at the primary speed limit. It may also occur where there are significant numbers of roadside entry points. In such cases, the stopping distance for trucks at the primary speed limit may not be sufficient for safety.

Night speed limits are frequently used in harsh terrain, where reduced nighttime visibility would make it unsafe to drive at the primary speed limit. The night speed limit sign has white reflectorized lettering on a black background.

Minimum speed limits are employed to reduce the variability of individual vehicle speeds within the traffic stream. They are generally applied to freeways, and are infrequently used on other types of highways. Minimum speed limits are very difficult to enforce, and where heavy traffic exists either due to demand or due to accidents/incidents, all speeds may be reduced to a level below the posted minimum.

All applicable speed limit signs should be posted at the same locations. There is a general caveat that no more than three speed limits should be in effect on any given segment of highway. See Chapter 4 for illustrations of speed limit signs and their posting.

31.4 Managed Lanes on Freeways

Managed lanes [6, 7] on freeways have become a common tool for operational agencies in urban and suburban areas where freeway congestion has become a systemic problem. Managed lanes are created by constructing additional lanes, or reconfiguring existing lanes where flow can be monitored and actively managed. In general, there are three types of lanes that are in common use:

- High-occupancy vehicle (HOV) lanes: Lane usage is limited to vehicles carrying a minimum of “x” people. At a minimum, two people would be required, but limits as high as three or four people are not uncommon. In some cases, the number of people required for entry into the lane can be varied by time of day. Such lanes may or may not admit buses and/or taxis (with passengers).

- High-occupancy toll (HOT) lanes: Lane usage is permitted to vehicles carrying a minimum of “x” people (as in the case of HOV lanes), but other vehicles are permitted to use the lane for a fee or toll. Such lanes are common where simple HOV lanes do not generate significant usage. They are also a means of generating revenue.

- Express toll (ET) lanes: Lane usage is by fee or toll; there are no restrictions on occupancy. These lanes are revenue-generating, and rely on the fact that some drivers are willing to pay for a faster trip. Tolls are generally variable depending upon time of day and/or traffic conditions.

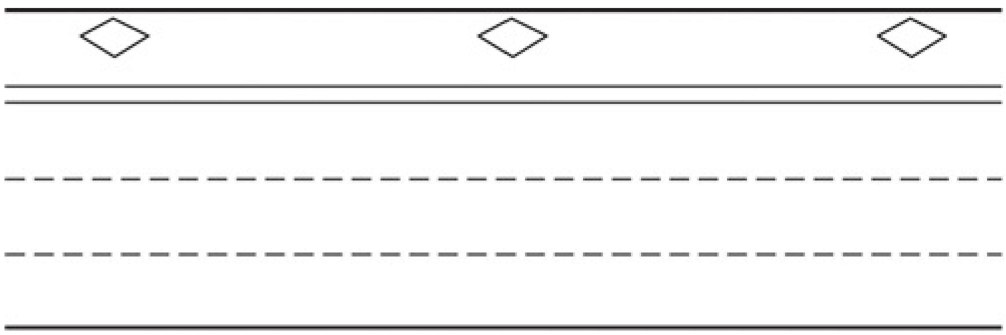
In all of these cases, the lanes must be set up and managed in such a way to produce conditions in which use of the restricted lane(s) provides the driver with a faster and less stressful driving environment.

In terms of physical design, there are a wide range of options available. The simplest design is a single or double lane segregated from normal use lanes by traffic markings, which could be a simple double barrier line, or a somewhat wider painted area (of from 2–5 ft). Other designs feature physical barriers between the managed and general use lanes.

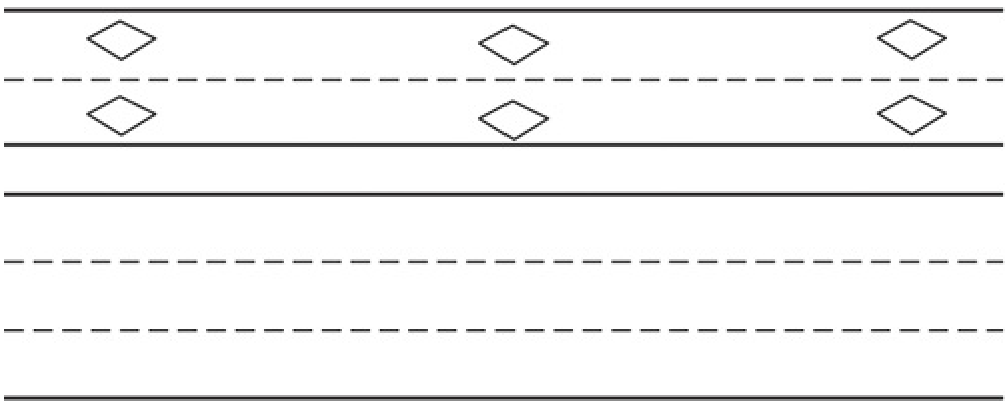
In all cases, points of interchange where drivers can enter or leave the managed lane are limited to designated areas. Where simple barrier lines are used to segregate lanes, a zone is periodically striped that allows for lane changes into and out of the managed lane. In other areas with wider barriers, slip ramps must be provided—sometimes with simple markings and others with physical boundaries.

Figure 31.20 shows some of the typical types of separations between managed and general use lanes currently in use. Figure 31.21 shows various designs for access and egress to and from managed lanes.

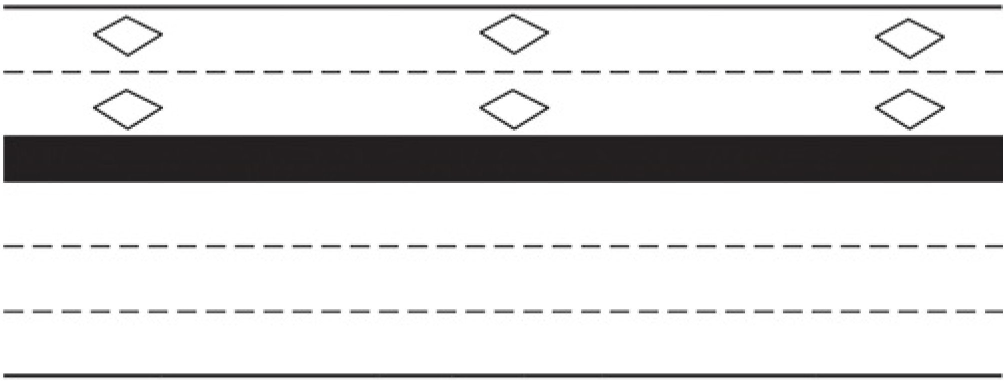
Figure 31.20: Types of Managed Lane Design



(a) Single Managed Lane Segregated by Barrier Markings



(b) Two Managed Lanes Segregated by Wide Paint Markings

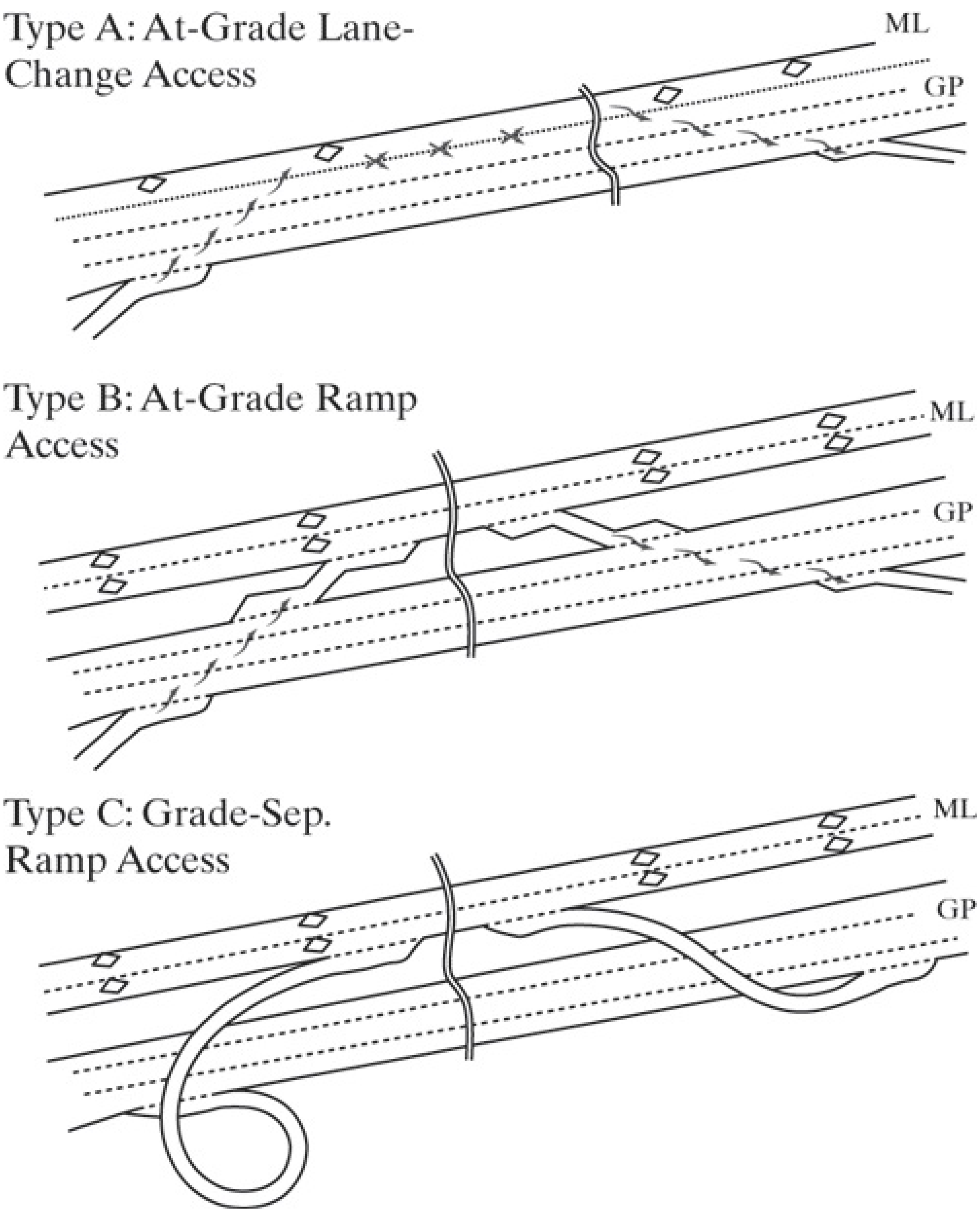


(c) Two Managed Lanes Segregated by Physical Barrier

(Source: Reprinted with permission from *Analysis of Managed Lanes on Freeway Facilities*, NCHRP Web-Only Document 191, National Cooperative Highway Research Program, Transportation Research Board, National Academy of Sciences, Courtesy of the National Academies Press, Washington, D.C., 2014.)

31.4-2 Full Alternative Text

Figure 31.21: Access and Egress From Managed Lanes Illustrated



(Source: Reprinted with permission from *Analysis of Managed Lanes on Freeway Facilities*, NCHRP Web-Only Document 191, National Cooperative Highway Research Program, Transportation Research Board, National Academy of Sciences, Courtesy of the National Academies Press, Washington, D.C., 2014.)

Figure 31.21: Full Alternative Text

Managed lanes are actively monitored for traffic conditions, and technologies must be provided to collect fees or tolls, and to ensure that ineligible vehicles are not using the lane. Common technologies involve drivers obtaining coded window stickers that can be read at high speeds as vehicles pass under overhead sensors. License plate cameras are also frequently deployed to identify vehicles that do not have the proper stickers for lane use.

For simple HOV lanes, such technology is not required, and simple physical observation is used to enforce regulations for use. License plate cameras are, however, often used and can be set to also monitor the number of persons in a vehicle.

31.5 Active Transportation and Demand Management Strategies

ATDM strategies cover a broad range of measures that can be applied to freeways and expressways, as well as to urban street systems. They are generally part of a system- or facility-wide management effort to control congestion and operations. This section provides a general overview of ATDM as applied to freeways, expressways, and other rural highways.

For freeway and expressway facilities, the following types of ATDM strategies are seeing increased use, particularly in suburban and urban areas:

- Managed lanes (see previous section)
- Dynamic ramp metering
- Dynamic lane and shoulder use controls
- Dynamic speed limits
- Queue warning systems
- Dynamic pricing strategies
- Dynamic traveler information systems

As discussed previously, *managed lanes*, including HOV, HOT, and express toll lanes, are often implemented in conjunction with *dynamic pricing strategies*. In such cases, fees or tolls for managed lane usage can be dynamically varied by time of day or day of week, as well as in response to real-time monitoring of traffic conditions.

Ramp metering is a technology that has been explored and tested many times over the past 20 years. In the beginning, ramp metering rates were set to allow “x” vehicles per hour to enter the freeway. Modern systems use *dynamic ramp metering*, in which rates at each ramp are set based upon real-time traffic condition information.

Ramp metering is a complex subject in and of itself. A Federal Highway Administration (FHWA) handbook [8] provides a useful overview of this subject. Ramp metering has a simple objective: limit and control entries into the freeway or expressway traffic stream to avoid the creation of merge bottlenecks. Key operational elements include a dynamic metering rate (how many veh/h will be permitted to enter at any given time), and an evaluation of the impacts of vehicles diverted from the controlled ramp. The latter is critical. Vehicles not permitted onto the freeway at a metered location go somewhere else—the next on-ramp, adjacent arterials, local streets, and so on. Some of the impacts of this diversion may require remediation.

Dynamic speed control utilizes variable speed limit signs, and attempts to make the traffic stream travel at a more uniform speed consistent with higher through-puts. To limit the range of individual speeds in the traffic stream, different speed limits may be applied to different lanes.

The simplest form of *dynamic lane use control* is allowing the use of paved shoulders as a travel lane during periods of congestion. Variable message signing is needed, and care must be taken to clearly mark ramp junctions and lanes to avoid dangerous conflicts.

Queue warning systems monitor downstream traffic conditions at closely spaced periodic intervals, and use software to determine when a downstream queue is forming, and its approximate size at any given time. Advance information is then relayed to drivers using variable message signs.

Traveler information systems provide real-time reporting of traffic conditions, and may or may not provide alternative route guidance where conditions warrant.

31.6 Analysis of Freeway Facilities

In the 2000 HCM, a methodology was introduced to allow analysis of long contiguous sections of a freeway facility that may include individual basic freeway segments, weaving segments, merge segments, and diverge segments. The 2000 HCM, however, lacked a great deal of computational detail, and the methodology was not included in the Highway Software Capacity software. Because of this, the methodology was largely ignored.

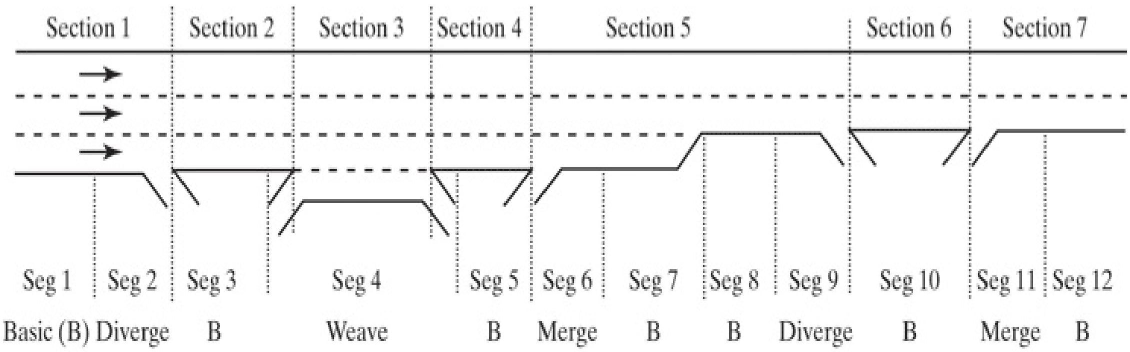
The method was greatly expanded in the 2010 HCM, and software was developed for its implementation. It was further expanded for the 2016 HCM, and a new HCM chapter was developed to detail its application to a reliability analysis of various configurations, including managed lanes and ATDM strategies. The methodology is far too complex to completely describe herein, but an overview of the procedural structure is worthwhile. The primary software for implementation of these methodologies is the most current version of FREEVAL.

Any facility analysis must begin with a definition of the length of freeway involved, together with all of the relevant geometric and traffic data. Because the overall methodology relies on the analysis of individual segments, the freeway facility must be divided into sections, and then segments for analysis.

Division into sections is relatively straightforward using the definitions of weaving, merging, and diverging sections contained in Chapter 28. Any section not falling into these categories is a basic freeway section.

For analysis, however, sections must be further broken into segments. In addition to boundaries between sections, segments must be established at all points where a change in geometric or traffic conditions occurs. Because the influence areas of ramp junctions extend 1,500 ft downstream of an of an on-ramp and 1,500 ft upstream of an off-ramp, longer acceleration or deceleration lanes must be treated as basic freeway segments outside the influence area. These would include the additional lane (the accel or decel lane). Separate basic freeway segments would be established *before* the beginning of the deceleration lane or *after the* end of the deceleration lane. A basic freeway section might have to be broken up into separate segments if there are changes in the geometry, such as an extended upgrade or downgrade, or change in general terrain from level to rolling (or vice-versa), or lanes are added or subtracted. An illustration of the process for segmenting a freeway facility is shown in Figure 31.22.

Figure 31.22: Segments and Time-Space Definition for Freeway Facility Analysis



Analysis Period	Seg 1	Seg 2	Seg 3	Seg 4	Seg 5	Seg 6	Seg 7	Seg 8	Seg 9	Seg 10	Seg 11	Seg 12
1												
2												
3												
4												
5												
6												
7												
8												

(Source: Reprinted with permission from *Highway Capacity Manual, 6th Edition: A Guide for Multimodal Mobility Analysis*, Transportation Research Board, National Academy of Sciences, Courtesy of the National Academies Press, Washington, D.C., 2016.)

Figure 31.22: Full Alternative Text

Figure 31.22 also illustrates the analysis structure of the freeway facility methodology. In general, a time-space analysis area is established, using 15-minute time periods, and the individual segments established. Longer time periods are possible, but results would lack the detail that use of standard 15-minute time blocks would provide.

The complex methodology begins by analyzing individual cells within the time-space analysis area. Where operations are found to be stable (i.e., *not* LOS F), the results are entered. Where any cell, however, operates at LOS F, a complex series of models traces the impact of the breakdown through both time and space, altering the results in adjacent and nearby cells as necessary. For this reason, the time-space area must be defined such that there are no LOS F results at the boundaries—that is, the first or last row, the first or last column. The analysis space must contain any failure and its propagation to cells not on any boundary.

Demand flow rates must be entered for each cell of the analysis area. The results of analysis will be speed and LOS predictions for all segments and time periods within the space, including the time and spatial impacts of any breakdowns. An overall LOS is also provided for each 15-minute period across the facility, based upon an average density—except that where any cell is at LOS F, all cells for that time period are labeled LOS F.

Chapter 11 of the 2016 HCM contains detailed descriptions of how to apply the freeway facility methodology to estimate travel time reliability over a full year. The methodology includes the ability to measure the impacts of weather and incidents based upon regional and national averages. If local data on these characteristics are available, they can be entered—obviously a long and tedious process. Although reliability covers a full year, it can be limited to a defined period of the day over the analysis year.

Travel time reliability is measured primarily in terms of the travel time index (TTI). The TTI is the percentile travel time from one end of the defined facility to the other divided by the free-flow travel time from one end of the defined facility to the other. Thus, there are several versions of TTI based upon which percentile travel time is used. Common percentiles for use are the 95th percentile, the 80th percentile, and the 50th percentile. For example, a facility reliability analysis produced the following travel times:

$TT_{95}=45$ min $TT_{80}=41$ min $TT_{50}=34$ min Free-flow travel time $=30$ min

Then:

TTI95=4530=1.50TTI80=4130=1.37TTI50=3430=1.13

Obviously, the higher the TTI value, the less “reliable” travel time is over a full year during the designated time of the day covered by the analysis.

There has been much discussion of travel time reliability in the profession. Nevertheless, federal agencies are now using it as a measure of overall effectiveness of the nation’s highway programs, leading to extensive research into methods for estimating the value.

There have also been questions as to how relevant reliability over a full year really is. Most motorists, when questioned, really would like to know what probabilities exist for travel times for a specific trip (covering many facilities) on a particular day and time.

References

- 1. *The Highway Capacity Manual, 6th Edition: A Guide for Multimodal Mobility Analysis*, Transportation Research Board, Washington, D.C., 2016.

- 2. *Manual of Uniform Traffic Control Devices*, Federal Highway Administration, Washington, D.C, 2009, as amended through 2012.

- 3. “Purpose and Policy in the Establishment and Development of United States Numbered Highways,” American Association of State Highway and Transportation Officials, Washington, D.C., revised September 15, 1970.

- 4. “In the Establishment of a Marking System of Routes Comprising the National System of Interstate and Defense Highways,” American Association of State Highway and Transportation Officials, Washington, D.C., August 14, 1954, revised August 10, 1973.

- 5. Weingroff, P.E., “From Names to Numbers: The Origins of the U.S. Numbered Highway System,” *AASHTO Quarterly*, American Association of State Highway and Transportation Officials, Washington, D.C., Spring 1997.

- 6. Wang, Y., et al “Analysis of Managed Lanes on Freeways,” *NCHRP Web-Only Document 191*, National Cooperative Highway Research Program, Transportation Research Board, Washington, D.C.

- 7. *Guide for Highway Capacity and Operations Analysis of Active Transportation and Demand Management Strategies*, Federal Highway Administration, Washington, D.C., 2014.

- 8. *Ramp Management and Control Handbook*, Federal Highway Administration, Washington, D.C., January 2006.

Problems

1. 31-1 A three-lane rural highway has 12-foot lanes and a speed limit of 55 mi/h. There is no passing permitted in the direction with a single lane. What is the minimum length of the transition and buffer markings at a location where the center-lane direction is to be switched?

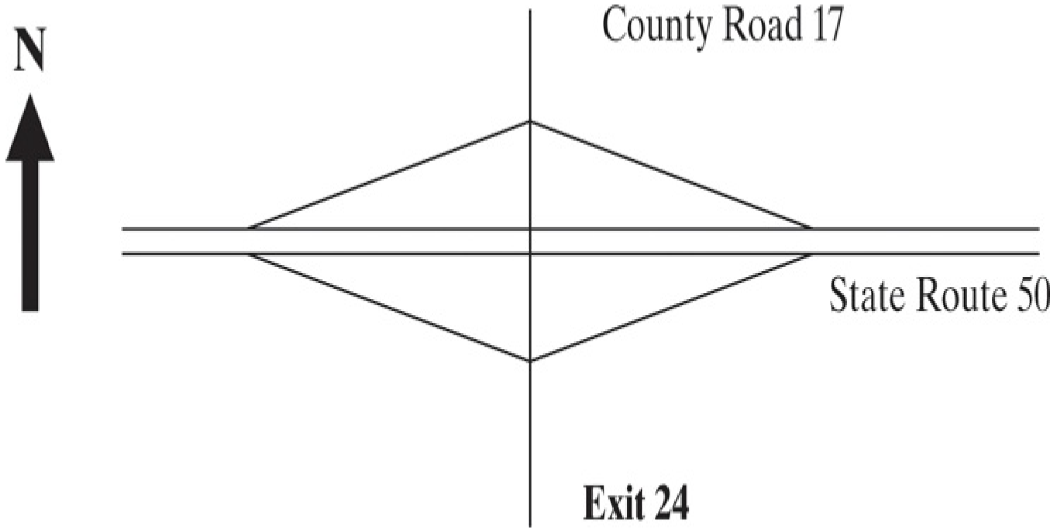
2. 31-2 An expressway in a suburban area has a design speed of 65 mi/h and an 85th percentile speed of 72 mi/h. It is experiencing a high accident rate compared to similar highways in the same jurisdiction. What speed limit would you recommend? What additional information would you like to have before making such a recommendation?

3. 31-3 This class project should be assigned to groups with a minimum of two persons in each group.

   Select a 5-mile section of freeway or rural highway in your area. Survey the section in both directions, making note of all traffic signs and markings that exist. Evaluate the effectiveness of these signs and markings, and suggest improvements that might result in better communication with motorists. Write a report on your findings, including photographs where appropriate to illustrate your comments.

4. 31-4 The figure below illustrates a diamond interchange between a state-numbered freeway and a county road. The diamond–county road intersections are STOP-controlled. Indicate what guide signs and route signs you would place, specifying their location. Prepare a rough sketch of each sign to indicate its precise content. Note that there are no other exits on State Route 50 within 25 miles of this location.

Interchange for Problem 31-4



Full Alternative Text

5. 31-5 List the types of managed lanes currently in use on U.S. freeways, along with a brief description of what each is intended to accomplish.

6. 31-6 List the types of ATDM strategies currently in use, and briefly describe their purpose.

7. 31-7 A reliability study along a 15-mile section of suburban freeway reveals the following travel times:

   $TT_{95} = 33.0$ minutes $TT_{80} = 27.0$ minutes $TT_{50} = 15.0$ minutes Free-Flow Speed: 70 mi/h

   From this information, compute the relevant travel time indices, and comment on the meaning of the results.

Index

Contents

List of Illustrations

Landmarks

77. 67
78. 68
79. 69
80. 70
81. 71
82. 72
83. 73
84. 74
85. 75
86. 76
87. 77
88. 78
89. 79
90. 80
91. 81
92. 82
93. 83
94. 84
95. 85
96. 86
97. 87
98. 88
99. 89
100. 90
101. 91
102. 92
103. 93
104. 94
105. 95
106. 96
107. 97
108. 98
109. 99
110. 100
111. 101
112. 102
113. 103
114. 104
115. 105
116. 106

117. 107
118. 108
119. 109
120. 110
121. 111
122. 112
123. 113
124. 114
125. 115
126. 116
127. 117
128. 118
129. 119
130. 120
131. 121
132. 122
133. 123
134. 124
135. 125
136. 126
137. 127
138. 128
139. 129
140. 130
141. 131
142. 132
143. 133
144. 134
145. 135
146. 136
147. 137
148. 138
149. 139
150. 140
151. 141
152. 142
153. 143
154. 144
155. 145
156. 146

157. 147
158. 148
159. 149
160. 150
161. 151
162. 152
163. 153
164. 154
165. 155
166. 156
167. 157
168. 158
169. 159
170. 160
171. 161
172. 162
173. 163
174. 164
175. 165
176. 166
177. 167
178. 168
179. 169
180. 170
181. 171
182. 172
183. 173
184. 174
185. 175
186. 176
187. 177
188. 178
189. 179
190. 180
191. 181
192. 182
193. 183
194. 184
195. 185
196. 186

197. 187
198. 188
199. 189
200. 190
201. 191
202. 192
203. 193
204. 194
205. 195
206. 196
207. 197
208. 198
209. 199
210. 200
211. 201
212. 202
213. 203
214. 204
215. 205
216. 206
217. 207
218. 208
219. 209
220. 210
221. 211
222. 212
223. 213
224. 214
225. 215
226. 216
227. 217
228. 218
229. 219
230. 220
231. 221
232. 222
233. 223
234. 224
235. 225
236. 226

277. 267
278. 268
279. 269
280. 270
281. 271
282. 272
283. 273
284. 274
285. 275
286. 276
287. 277
288. 278
289. 279
290. 280
291. 281
292. 282
293. 283
294. 284
295. 285
296. 286
297. 287
298. 288
299. 289
300. 290
301. 291
302. 292
303. 293
304. 294
305. 295
306. 296
307. 297
308. 298
309. 299
310. 300
311. 301
312. 302
313. 303
314. 304
315. 305
316. 306

The following statistics are shown for 2015:

Statistic: 2015 value

Miles of public roadway: 4.19 million

Vehicle-miles traveled: 3.11 trillion

Total population of the United States: 321 million

Licensed drivers: 218 million

Registered vehicles: 256 million

Fatalities: 35,485

The table lists revenue in billions of dollars, the sources and the percentage of total, as follows:

Revenue; Source; Percent of total

41.2; State and local motor fuel taxes; 26.9

28.0; Federal motor fuel and other excise taxes; 18.3

23.2; State license fees; 15.2

12.7; Tolls and other local user fees; 8.3

105.1; Subtotal road-user taxes; 68.7

30.0; State and local general fund allocations; 19.6

18.0; Federal general fund allocations and deficit financing; 11.7

48.0; Subtotal general funds; 31.3

153.1; TOTAL; 100.0

Horizontal axis plots years from 1920 to 2015 in increments of 5. Vertical axis on left plots public road mileage, measured in millions between 3.0 and 9.0 in increments of 1.0 while vertical axis on right plots vehicle miles of travel (V M T), measured in trillions between 0.0 and 3.5 in increments of 0.5. The graph plots three lines. An upward sloping line with a slight slope represents public road mileage, marking 3 million in 1920 to slightly more than 4 million in 2015. A second line representing vehicle miles of travel is plotted with a steep upward slope, marking around 0.1 trillion miles in 1920 to 3.2 trillion miles in 2015. A third line is drawn between 1980 and 2015, representing lane miles that marks between 3 trillion miles to 3.4 trillion miles.

The approximate values for various years are as follows:

Year: Public road mileage; Vehicle miles of travel; Lane miles

1920: 3.05; 0.05; no data

1930: 3.2; 0.15; no data

1940: 3.25; 0.30; no data

1950: 3.3; 0.40; no data

1965: 3.75; 0.70; no data

1975: 3.85; 1.3; no data

1980: 3.85; 1.55; 2.7

1985: 3.80; 1.75; 2.85

1990: 3.80; 2.0; 2.90

2000: 3.90; 2.5; 3.0

2005: 4.0; 2.7; 3.1

2015: 4.25; 3.0; 3.4