CHAPTER 4
TRAFFIC OPERATIONS AND CAPACITY CONCEPTS

CONTENTS

LIST OF EXHIBITS

1. INTRODUCTION

OVERVIEW

The relationships between volume (flow rate), speed, and density are among the most fundamental in transportation engineering and can be used to describe traffic operations on any roadway. Similar principles apply to the pedestrian and transit modes, while bicycle speeds are primarily affected by facility grade and conditions, interactions with other modes, and bicyclist age and fitness level.

Capacity represents the maximum sustainable hourly flow rate at which persons or vehicles reasonably can be expected to traverse a point or a uniform segment of a lane or roadway during a given time period under prevailing roadway, environmental, traffic, and control conditions. Reasonable expectancy is the basis for defining capacity. A given system element's capacity is a flow rate that can be achieved repeatedly under the same prevailing conditions, as opposed to being the maximum flow rate that might ever be observed. Since the prevailing conditions (e.g., weather, mix of heavy vehicles) will vary within the day or from one day to the next, a system element's capacity at a given time will also vary.

CHAPTER ORGANIZATION

Chapter 4 describes how basic traffic operations relationships apply to the four travel modes covered by the *Highway Capacity Manual* (HCM).

Section 2 provides basic traffic operations relationships for the motorized vehicle mode, introduces the concept of travel time reliability, and describes additional parameters that can be used to describe aspects of traffic flow on interrupted- and uninterrupted-flow system elements. This section also provides capacity concepts for the motorized vehicle mode and describes three approaches for estimating traffic flow parameters.

Section 3 presents speed, flow, and density relationships for the pedestrian mode and capacity concepts for pedestrian circulation and queuing areas. Section 4 provides bicycle flow parameters and capacity concepts and describes the importance of stops and delay as measures of bicycle traffic operations. Finally, Section 5 describes the bus operations, bus vehicle, roadway infrastructure, traffic control, and passenger characteristics that influence bus speeds. The section also presents transit vehicle and person capacity concepts.

RELATED HCM CONTENT

Several of the operational performance measures presented in Chapter 4 (speed, delay, and density, in particular) are used in Chapter 5 to describe the quality of service provided by a roadway, or—in the case of the volume-to-capacity (demand-to-capacity) ratio—are used to define the threshold between Levels of Service (LOS) E and F.

Details of traffic operations and capacity relationships specific to a particular system element (for example, speed–flow curves for freeways) are provided in the "capacity concepts" subsections of the chapters in Volumes 2 and 3.

2. MOTORIZED VEHICLE MODE

A few basic parameters—volume, flow rate, speed, and density—can be used to describe traffic operations on any roadway. In the HCM, volume, flow rate, and speed are parameters common to both uninterrupted- and interrupted-flow facilities, but density applies primarily to uninterrupted flow. Some parameters related to flow rate, such as spacing and headway, are also used for both types of facilities. Other parameters, such as saturation flow and gap, are specific to interrupted flow.

BASIC MOTORIZED VEHICLE FLOW PARAMETERS

Volume and Flow Rate

Volume and flow rate are two measures that quantify the number of vehicles passing a point on a lane or roadway during a given time interval. These terms are defined as follows:

- *Volume*—the total number of vehicles passing over a given point or section of a lane or roadway during a given time interval; any time interval can be used, but volumes are typically expressed in terms of annual, daily, hourly, or subhourly periods.
- *Flow rate*—the equivalent hourly rate at which vehicles pass over a given point or section of a lane or roadway during a given time interval of less than 1 h, usually 15 min. This chapter focuses on flow rate and the variations in flow that can occur over the course of an hour.

Flow rate is the equivalent hourly volume that would occur if a subhourly flow was sustained for an entire hour.

There is a distinction between volume and flow rate. Volume is the number of vehicles observed or predicted to pass a point during a time interval. Flow rate represents the number of vehicles passing a point during a time interval less than 1 h, but expressed as an equivalent hourly rate. A flow rate is the number of vehicles observed in a subhourly period, divided by the time (in hours) of the observation. For example, a volume of 100 veh observed in a 15-min period implies a flow rate of 100 veh divided by 0.25 h, or 400 veh/h.

Observed volumes may reflect capacity constraints rather than true demand. Demand is usually the desired input to HCM analyses, although it is not always easy to determine.

Volume and flow rate are variables that help quantify *demand*, that is, the number of users (often expressed as the number of vehicles) who *desire* to use a given system element during a specific time period, typically 1 h or 15 min. Volume and flow rate also help quantify capacity, that is, the number of users who *can* use a given system element during a specific time period. As discussed in Chapter 3, Modal Characteristics, observed volumes may reflect upstream capacity constraints rather than the true demand that would exist without the presence of a bottleneck.

In many cases, demand volumes are the desired input to HCM analyses. (The analysis of traffic conditions downstream of a bottleneck that is not planned to be removed is an example of an exception.) When conditions are *undersaturated* (i.e., demand is less than capacity) and no upstream bottlenecks exist, demand volume at a location equivalent to the measured volume at that location can be assumed. Otherwise, ascertaining demand requires a count of undersaturated traffic upstream of a bottleneck (i.e., a count of arrival volume rather than departure volume) (1). When the queue from a bottleneck extends past the previous intersection or interchange, how much of the traffic approaching the end of the queue is actually destined for the bottleneck location may not be easy to determine. Furthermore, as illustrated in Chapter 3, demand patterns may change after a bottleneck is removed. Nevertheless, where bottlenecks exist, neglecting to use demand volumes as inputs to HCM methodologies will produce results that underestimate the presence and extent of congestion. In other words, using observed volumes instead of demand volumes will likely lead to inaccurate HCM results.

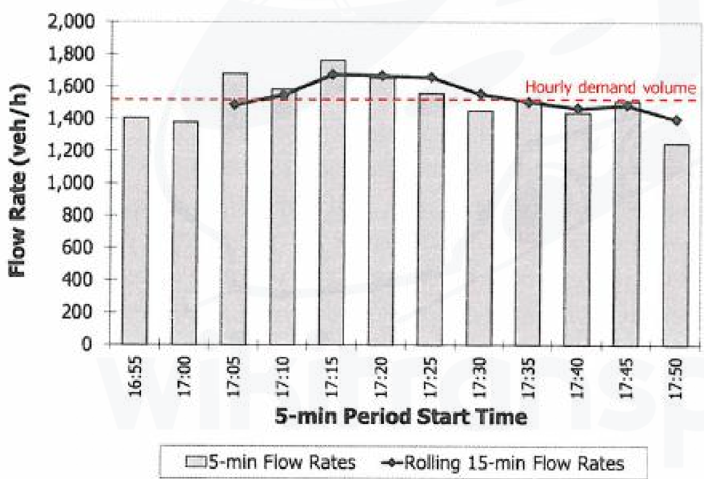
Using field-measured volumes as inputs to HCM methods without accounting for demand upstream of a bottleneck will lead to inaccurate results, such as demand-to-capacity ratios that can never exceed 1.

Subhourly Variations in Flow

Flow rates typically vary over the course of an hour. Exhibit 4-1 shows an example of the substantial short-term fluctuation in flow rate that can occur within an hour. Data from the approaches to an all-way STOP-controlled intersection are used. In this data set, the 5-min flow rate ranges from a low of 1,248 veh/h to a high of 1,764 veh/h, compared with a total peak hour entering volume of 1,516 veh. Designing the intersection to accommodate the peak hour volume would result in oversaturated conditions for a substantial portion of the hour.

Even when hourly volumes are less than a system element's capacity, flow rates within an hour may exceed capacity, creating oversaturated conditions.

Exhibit 4-1
Differences Between Short-Term Flow Rates and Hourly Demand Volumes



Note: SW 72nd Avenue at Dartmouth Street, Tigard, Oregon, 2008.

HCM analyses typically consider the peak 15 min of flow during the analysis hour. As illustrated in Exhibit 4-1, the use of a peak 15-min flow rate accommodates nearly all the variations in flow during the hour and therefore provides a good middle ground between designing for hourly volumes and designing for the most extreme 5-min flow rate.

Since inputs to HCM procedures are typically expressed in terms of hourly demands, the HCM uses the *peak hour factor* (PHF) to convert an hourly volume into a peak 15-min flow rate. Although traditionally called a "peak hour" factor, a PHF is applicable to any analysis hour, peak or off-peak. The PHF is the ratio of total hourly volume to the peak flow rate within the hour:

Peak hour factor (PHF) defined.

$$PHF = \frac{\text{hourly volume}}{\text{peak flow rate (within the hour)}}$$

Equation 4-1

If 15-min periods are used, the PHF may be computed by Equation 4-2:

Equation 4-2

$$PHF = \frac{V}{4 \times V_{15}}$$

where

PHF = peak hour factor,

V = hourly volume (veh/h), and

V_{15} = volume during the peak 15 min of the analysis hour (veh/15 min).

When the PHF is known, it can convert a peak hour volume to a peak flow rate, as in Equation 4-3:

Equation 4-3

$$v = \frac{V}{PHF}$$

where v is the flow rate for a peak 15-min period, expressed in vehicles per hour, and the other variables are as defined previously.

Equation 4-3 does not need to be used to estimate peak flow rates if traffic counts are available; however, the chosen count interval must identify the maximum 15-min flow period. Then the rate can be computed directly as 4 times the maximum 15-min count and the PHF would take the value 1.00.

Lower PHF values signify greater variability of flow, while higher values signify less flow variation within the hour. When hourly counts are used, the PHF can range from 1.00, indicating that the same demand occurs during each 15-min period of the hour, to a theoretical minimum of 0.25, indicating that the entire hourly demand occurs during the peak 15 min. PHFs in urban areas generally range between 0.80 and 0.98. PHFs over 0.95 are often indicative of high traffic volumes, sometimes with capacity constraints on flow during the peak hour. PHFs under 0.80 occur in locations with highly peaked demand, such as schools, factories with shift changes, and venues with scheduled events.

Speed

Although traffic volumes provide a method of quantifying capacity values, speed (or its reciprocal, *travel time rate*) is an important measure of the quality of the traffic service provided to the motorist. It helps define LOS for two-lane highways and urban streets.

Speed parameters.

Speed is defined as a rate of motion expressed as distance per unit of time, generally as miles per hour (mi/h). To characterize the speed of a traffic stream, a representative value must be used, because a broad distribution of individual speeds is observable in the traffic stream. Several speed parameters can be applied to a traffic stream. Among them are the following:

Average travel speed is a type of space mean speed.

- *Average travel speed*. The length of a roadway segment divided by the average travel time of vehicles traversing the segment, including all stopped delay times. It is a type of *space mean speed* because the average travel time weights the average by the time each vehicle spends in a defined roadway segment or space.

- *Time mean speed.* The arithmetic average of speeds of vehicles observed passing a point on a highway; also referred to as the *average spot speed.* The individual speeds of vehicles passing a point are recorded and averaged arithmetically. The time mean speed is always equal to or higher than the space mean speed. The two are equal only when the speeds of all vehicles in the traffic stream are equal.

A field-measured time mean speed will always be higher than the space mean speed, unless all vehicles in the traffic stream travel at the same speed, in which case the time mean speed will equal the space mean speed.

- *Free-flow speed.* The average speed of vehicles on a given segment, measured under low-volume conditions, when drivers are free to drive at their desired speed and are not constrained by the presence of other vehicles or downstream traffic control devices (i.e., traffic signals, roundabouts, or STOP signs).

Free-flow speed reflects drivers' desired speed, unconstrained by other vehicles or traffic control.

- *Average running speed.* A traffic stream measure based on the observation of travel times of vehicles traversing a section of highway of known length. It is the length of the segment divided by the average running time of vehicles that traverse the segment. Running time includes only time during which vehicles are in motion.

Average running speed only considers time spent in motion. It is also a type of space mean speed.

For most of the HCM procedures using speed as a service measure, average travel speed is the defining parameter. On uninterrupted-flow facilities operating with undersaturated flow, the average travel speed is equal to the average running speed.

Both time mean speed and space mean speed can be calculated from a sample of individual vehicle speeds. For example, three vehicles are recorded by a spot sensor (e.g., loop detectors, radar) with speeds of 30, 40, and 50 mi/h in the middle of a 1-mi roadway segment. The travel times for the same vehicles over the 1-mi segment are measured as 2.0 min, 1.5 min, and 1.2 min, respectively (i.e., by recording the times the vehicles enter and exit the segment). The time mean speed is 40 mi/h, calculated as (30 + 40 + 50 mi/h)/3. The space mean speed is 38.3 mi/h, calculated as (60 min/h) × [3/(2.0 + 1.5 + 1.2 min/mi)].

Space mean speed is recommended for HCM analyses. Speeds are best measured by observing travel times over a known length of highway. For uninterrupted-flow facilities operating in the range of stable flow, the length may be as short as several hundred feet for ease of observation.

Density

Density is the number of vehicles occupying a given length of a lane or roadway at a particular instant. For the computations in this manual, density is averaged over time and is usually expressed as vehicles per mile (veh/mi) or passenger cars per mile (pc/mi).

Computing density.

Measuring density directly in the field is difficult: it requires a vantage point for photographing, videotaping, or observing significant lengths of highway. However, density can be computed from the average travel speed and flow rate, which are measured more easily. Equation 4-4 is used for undersaturated traffic conditions.

$$\text{Density (veh/mi)} = \frac{\text{flow rate (veh/h)}}{\text{average travel speed (mi/h)}}$$

Equation 4-4

A highway segment with a flow rate of 1,000 veh/h and an average travel speed of 50 mi/h would have a density of (1,000 veh/h) / (50 mi/h) = 20 veh/mi.

Density is a critical parameter for uninterrupted-flow facilities because it characterizes the quality of traffic operations. It describes the proximity of vehicles to one another and reflects the freedom to maneuver within the traffic stream.

Roadway occupancy is frequently used as a surrogate for density in control systems because it is easier to measure (most often through equipment such as loop detectors). Occupancy in space is the proportion of roadway length covered by vehicles, and occupancy in time identifies the proportion of time a roadway cross section is occupied by vehicles. However, unless the length of vehicles is known precisely, the conversion from occupancy to density involves some error. A textbook (2) discusses derivation of occupancy and its relationship to density.

Headway and Spacing

Headway is the time between successive vehicles as they pass a point on a lane or roadway, measured from the same point on each vehicle. *Spacing* is the distance between successive vehicles in a traffic stream, measured from the same point on each vehicle (e.g., front bumper, front axle).

These characteristics are microscopic, because they relate to individual pairs of vehicles within the traffic stream. Within any traffic stream, both the spacing and the headway of individual vehicles are distributed over a range of values, generally related to the speed of the traffic stream and prevailing conditions. In the aggregate, these microscopic parameters relate to the macroscopic flow parameters of density and flow rate.

Relationships among density, speed and flow rate, and headway and spacing.

Spacing can be determined directly by measuring the distance between common points on successive vehicles at a particular instant. This generally requires costly aerial photographic techniques, so that spacing is usually derived from other direct measurements. Headway, in contrast, can be measured with stopwatch observations as vehicles pass a point on the roadway.

The density of a traffic stream is directly related to the average spacing between vehicles in the traffic stream:

Equation 4-5

$$\text{Density (veh/mi)} = \frac{\text{5,280 ft/mi}}{\text{average spacing (ft/veh)}}$$

The flow rate of a traffic stream is directly related to the average headway of vehicles in the traffic stream:

Equation 4-6

$$\text{Flow rate (veh/h)} = \frac{\text{3,600 s/h}}{\text{average headway (s/veh)}}$$

Finally, the relationship between average spacing and average headway in a traffic stream depends on speed. This relationship can be derived from the preceding two equations and the speed–flow–density relationship (Equation 4-4):

Equation 4-7

$$\text{Average headway (s/veh)} = \frac{\text{average spacing (ft/veh)}}{\text{average travel speed (ft/s)}}$$

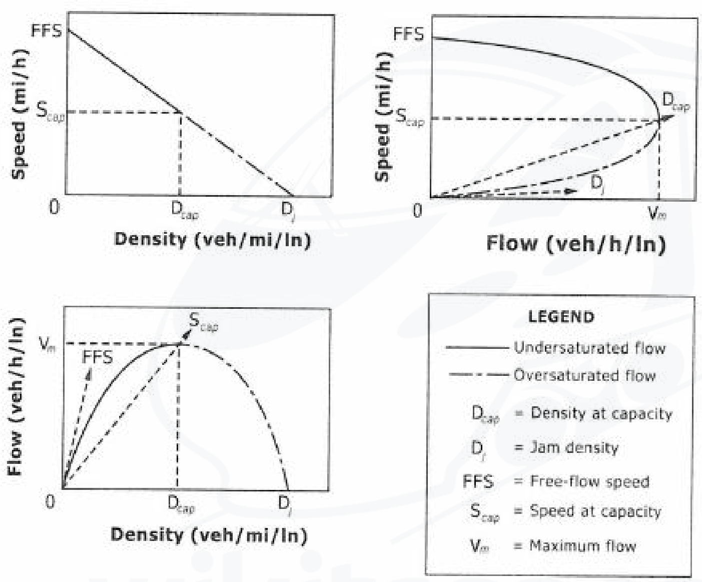
This relationship also holds for individual headways and spacings between pairs of vehicles. The speed used is that of the second vehicle in a pair.

Relationships Among Basic Parameters

Equation 4-4 cites the basic relationship among the three parameters, describing an uninterrupted traffic stream. Although Equation 4-4 allows for a given flow rate to occur in an infinite number of combinations of speed and density, additional relationships restrict the variety of flow conditions that can occur at a location.

Exhibit 4-2 shows a generalized, theoretical representation of these relationships, which are the basis for the capacity analysis of uninterrupted-flow facilities. The flow–density function is placed directly below the speed–density relationship because of their common horizontal scales, and the speed–flow function is placed next to the speed–density relationship because of their common vertical scales. The speed in all cases is *space mean speed*.

Exhibit 4-2
Generalized Relationships Among Speed, Density, and Flow Rate on Uninterrupted-Flow Facilities



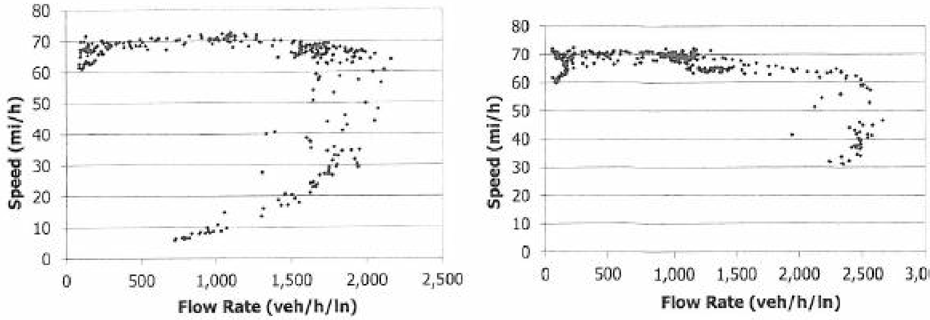
Source: Adapted from May (*2*).

Undersaturated, oversaturated, and queue discharge flow conditions were introduced in Section 5 of Chapter 2, Applications.

The form of these functions depends on the prevailing traffic and roadway conditions on the segment under study and on the segment length. Although the diagrams in Exhibit 4-2 show continuous curves, the full range of the functions is unlikely to appear at any particular location. Real-world data usually show discontinuities, with parts of the curves not present (*2*). Exhibit 4-3 shows that the real-world relationship between speed and undersaturated flow on freeways consists of a section of constant speed, followed by a section of declining speed until capacity is reached, unlike the idealized parabola shown in the speed–flow curve in Exhibit 4-2. Exhibit 4-3(a) shows a relatively complete curve, while Exhibit 4-3(b) has discontinuities. In addition, Exhibit 4-3 shows that a region of queue discharge flow exists between the two parts of the curves, where vehicles transition from oversaturated flow back to undersaturated flow after exiting a bottleneck.

Exhibit 4-3
Example Freeway Speed–Flow Data

Note that the real-world speed–flow curves in Exhibit 4-3 are not the idealized parabola indicated in Exhibit 4-2. The other relationships in Exhibit 4-2 therefore also have somewhat different shapes in the real world.



(a) I-405, Los Angeles, California

(b) I-805, San Diego, California

Source: Derived from California Department of Transportation data, 2008.

The curves of Exhibit 4-2 illustrate several significant details. A zero flow rate occurs under two different conditions. The first is when there are no vehicles on the segment—density is zero, and flow rate is zero. Speed is theoretical for this condition and would be selected by the first driver (presumably at a high value). This *free-flow speed* is represented by *FFS* in the graphs.

The second condition occurs when density becomes so high that all vehicles must stop—the speed and flow rate are zero because there is no movement and vehicles cannot pass a point on the roadway. The density at which all movement stops is called *jam density,* denoted by D_j in the diagrams.

Between these two extreme points, the dynamics of traffic flow produce a maximizing effect. As flow increases from zero, density also increases because more vehicles are on the roadway. When this happens, speed declines because of the interaction of vehicles. The decline is negligible at low and medium densities and flow rates and vehicles operate at the free-flow speed, as illustrated in Exhibit 4-3. As density increases, the generalized curves suggest that speed decreases significantly before capacity is achieved. Capacity is reached when the product of density and speed results in the maximum flow rate. This condition is shown as the speed at capacity S_{cap} (often called *critical speed*), density at capacity D_{cap} (sometimes referred to as *critical density*), and maximum flow v_m.

The slope of any ray drawn from the origin of the speed–flow curve represents the inverse of density, on the basis of Equation 4-4. Similarly, a ray in the flow–density graph represents speed. As examples, Exhibit 4-2 shows the average free-flow speed and speed at capacity, as well as optimum and jam densities. The three diagrams are redundant—if any one relationship is known, the other two are uniquely defined. The speed–density function is used mostly for theoretical work; the other two are used in this manual to define LOS for freeways and multilane highways.

Exhibit 4-2 shows that any flow rate other than capacity can occur under two conditions, one low density and high speed and the other high density and low speed. The high-density, low-speed side of the curves represents oversaturated flow. Sudden changes can occur in the state of traffic (i.e., in speed, density, and flow rate). LOS A through E are defined on the low-density, high-speed side of the curves, with the maximum-flow boundary of LOS E placed at capacity; in contrast, LOS F, which describes oversaturated and queue discharge traffic, is represented by the high-density, low-speed part of the curves.

TRAVEL TIME RELIABILITY

Sources of Travel Time Variability

The travel time experienced by a traveler on a given roadway facility varies from one trip to the next. The variation is a result of the following:

Travel time reliability is influenced by demand variations, weather, incidents, work zones, and special events, all of which can be modeled by HCM methods.

- *Recurring variations in demand,* by hour of day, day of week, and month of year;
- *Severe weather* (e.g., heavy rain, snow, poor visibility) that affects capacity and drivers' choice of free-flow speed;
- *Incidents* (e.g., crashes, stalls, debris) that affect capacity and drivers' choice of free-flow speed;
- *Work zones* that reduce capacity and (for longer-duration work) may influence demand; and
- *Special events* (e.g., major sporting events, large festivals or concerts) that produce temporary, intense traffic demands, which may be managed in part by changes in the facility's geometry or traffic control.

Reliability analysis accounts for nonrecurring traffic conditions and events that normally cannot be accounted for by the core HCM methods.

In contrast, the HCM's core freeway and urban street facility procedures (Chapters 10 and 16, respectively) describe the travel time of an average trip along a facility during a user-defined *analysis period,* typically the peak 15 min of a peak hour, under specific conditions (e.g., good weather, no incidents). Since this travel time is an average, conditions will be better at certain times of the day or on certain days during the year, because of lower-than-average traffic demands. There will also be days when travel will take much more time, because of incidents, severe weather, unusually high demand levels, or a combination.

Defining and Expressing Reliability

Travel time reliability quantifies the variation of travel time. It is defined by using the entire range of travel times for a given trip for a selected time period (for example, the weekday p.m. peak hour) and over a selected horizon (for example, a year). For the purpose of measuring reliability, a "trip" can occur on a specific facility or on a subset of the transportation network, or the definition can be broadened to include a traveler's initial origin and final destination. Measurement of travel time reliability requires a history of travel times sufficient to track travel time performance. When travel time measurements are taken over a long period (e.g., a year), a travel time distribution results (3).

A travel time distribution may be characterized in one of two ways. Both methods have useful applications and are valuable for understanding and describing reliability. They are as follows:

1. Measures of the *variability* in travel times that occur on a facility or a trip over the course of time, as expressed through metrics such as a 50th, 80th, or 95th percentile travel time; and

The travel time distribution can be characterized in terms of travel time variability or in terms of the success or failure of a given trip in meeting a target travel time.

Reliability is quantified from the distribution of travel times on a facility.

2. Measures of the reliability of facility travel times, such as the number of trips that *fail* or *succeed* in accordance with a predetermined performance standard, as expressed through metrics such as on-time performance or percent failure based on a target minimum speed or maximum travel time.

For convenience, the HCM uses the single term *reliability* for both the variability- and the reliability-based approaches to characterizing a facility's travel time distribution.

Similar approaches can be used to describe the variability in other HCM facility performance measures, including percentiles (e.g., 50th percentile speed) and the probability of achieving a particular LOS. For freeway facilities, distributions can be produced for such measures as facility speed, travel time, and average density. For urban streets, distributions can be produced for travel time, travel speed, and spatial stop rate, among others.

Performance Measures Derived from the Travel Time Distribution

Time-Based Reliability Measures

The travel time distribution can be used to derive a variety of performance measures that describe different aspects of reliability. Exhibit 4-4 illustrates a selection of time-based reliability performance measures that can be derived from the travel time distribution:

Planning time is the total *travel time required for an on-time arrival 95% of the time, while buffer time is the* extra *travel time beyond the average travel time required for an on-time arrival 95% of the time.*

- *Planning time,* the travel time a traveler would need to budget to ensure an on-time arrival 95% of the time;
- *Buffer time,* the extra travel time a traveler would need to budget, compared with the average travel time, to ensure an on-time arrival 95% of the time;
- *Misery time,* the average of the highest 5% of travel times (approximating a 97.5 percentile travel time), representing a near-worst-case condition;
- *On-time percentage,* a measure of success based on the percentage of trips that are made within a target travel time;
- *Percentage of trips exceeding a target maximum travel time,* a measure of failure;
- *Standard deviation,* the statistical measure of how much travel times vary from the average; and
- *Semi–standard deviation,* a statistical measure of travel time variance from the free-flow speed.

Time-based measures are useful for describing the reliability of individual facilities and trips but are difficult to use for comparisons.

In Exhibit 4-4, measures incorporating units of time appear as horizontal lines in the graph, while measures that are percentages of trips appear as areas underneath the travel time distribution. The former are useful for describing the reliability of individual facilities and trips, but they are difficult to compare across facilities or trips because facility and trip lengths vary. Percentage measures, on the other hand, can be compared across facilities and trips, as can index-based measures that are derived from time-based measures. These types of reliability measures are described next.

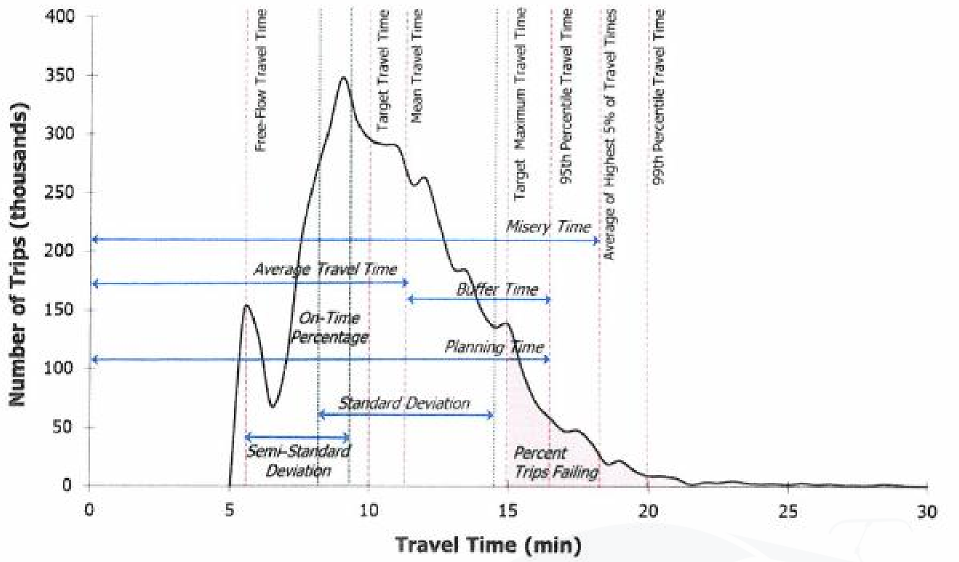


Exhibit 4-4
Derivation of Time-Based Reliability Performance Measures from the Travel Time Distribution

Source: Adapted from Zegeer et al. (*3*).

Index-Based Reliability Measures

To facilitate comparisons of different facilities or trips, travel time–based reliability measures can be converted into length-independent indices by dividing the base travel time measure by the free-flow travel time. Similarly, success and failure measures can be developed by comparing an index value with a target value. The following are examples:

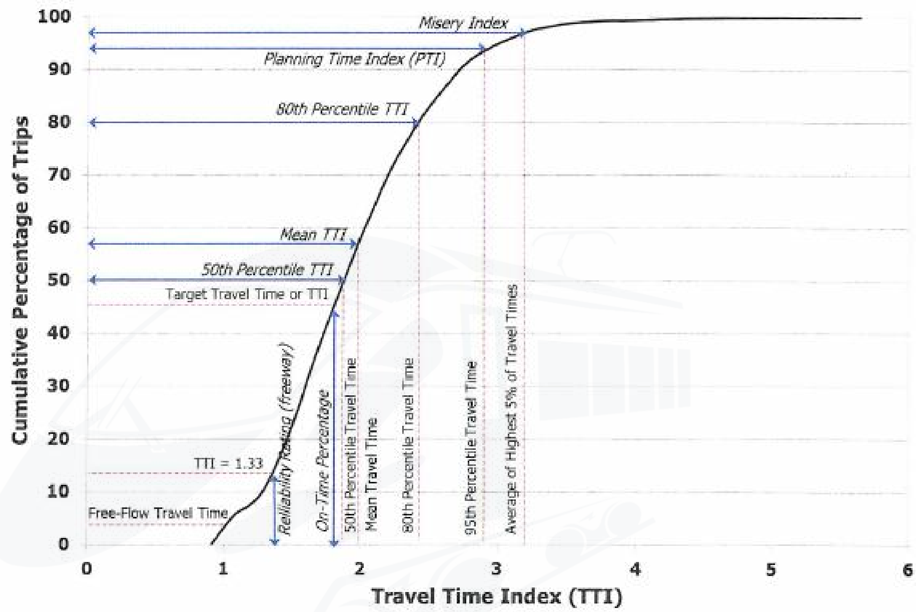
Index-based measures of reliability are independent of facility or trip length and thus are readily compared across facilities or trips.

- *Travel time index (TTI),* the average travel time on a facility divided by the travel time at free-flow speed; it can also be stated as a percentile travel time, as discussed below;
- *Planning time index (PTI),* the 95th percentile travel time divided by the free-flow travel time;
- *80th percentile TTI,* the 80th percentile travel time divided by the free-flow travel time; research indicates that this measure is more sensitive to operational changes than the PTI (*4*), which makes it useful for comparison and prioritization purposes;
- *50th percentile TTI,* the 50th percentile travel time divided by the free-flow travel time; its value will generally be slightly lower than the mean TTI due to the influence of rare, very long travel times in the travel time distribution;
- *Misery index,* the misery time divided by the free-flow travel time, a useful descriptor of near-worst-case conditions on rural facilities; and
- *Reliability rating,* the percentage of vehicle miles traveled experiencing a TTI less than 1.33 for freeways and 2.50 for urban streets; these thresholds approximate the points beyond which travel times become much more variable (unreliable).

The difference in threshold values for freeways and urban streets reflects differences in how free-flow speed is defined for these facilities.

Exhibit 4-5 illustrates a selection of index-based reliability measures. The same travel time distribution is used as in Exhibit 4-4, but travel times are converted to TTIs and the travel time distribution is plotted as a cumulative function. The mean travel time in this distribution happened to be exactly twice the free-flow travel time (i.e., a mean TTI of 2.00), but this result is coincidental. In this graph, index measure values are horizontal lines, while percentage measure values (e.g., on-time percentage, reliability rating) are vertical lines.

Exhibit 4-5
Derivation of Index-Based Reliability Performance Measures from the Travel Time Distribution



Other types of indices can be created by using a denominator other than free-flow travel time. For example, a *policy index* can be defined that is similar to the TTI but replaces free-flow speed with a target or "policy" speed, such as a desired minimum operating speed for the facility (typically chosen as a speed just above breakdown, thus providing maximum throughput).

The *buffer index* is the 95th percentile travel time divided by the average travel time. However, it is not recommended for tracking reliability trends over time because it is linked to two factors that can change: average and 95th percentile travel times. If one factor changes more in relation to the other, counterintuitive results can appear (*3, 4*).

ADDITIONAL UNINTERRUPTED-FLOW PARAMETERS

Headway

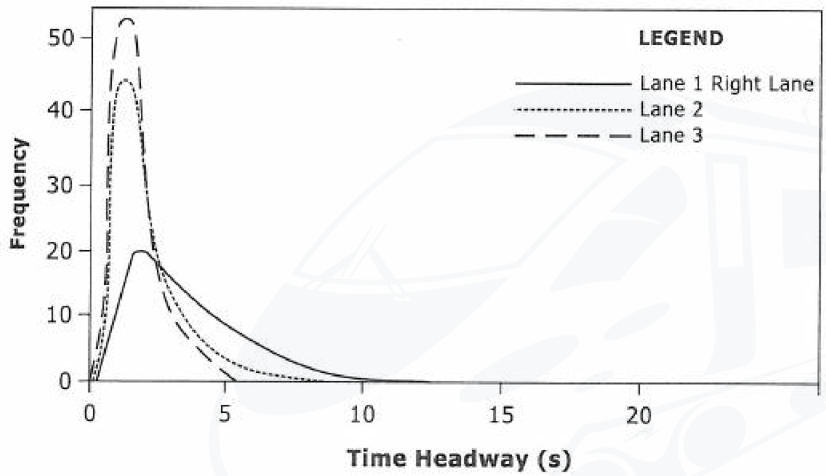
The average headway in a lane is the reciprocal of the flow rate. Thus, at a flow of 2,400 veh/h/ln, the average headway is (3,600 s/h) / (2,400 veh/h), or 1.5 s/veh. However, vehicles do not travel at constant headways. Vehicles tend to travel in groups (*platoons*), with varying headways between successive vehicles.

An example of the distribution of headways observed on the Long Island Expressway is shown in Exhibit 4-6. The headway distribution of Lane 3 is the most nearly uniform, as evidenced by the range of values and the high frequency of the modal value, which is the peak of the distribution curve. The distribution of Lane 2 is similar to that of Lane 3, with slightly greater scatter (range from 0.5 to 9.0 s). Lane 1 shows a much different pattern: it is more dispersed, with headways ranging from 0.5 to 12.0 s, and the frequency of the modal value is only about one-third of that for the other lanes. This indicates that the flow rate in the shoulder lane is usually lower than the flow rates in the adjacent lanes when the total flows on this segment are moderate to high.

Exhibit 4-6 shows relatively few headways smaller than 1.0 s. A vehicle traveling at 60 mi/h (88 ft/s) would have a spacing of 88 ft with a 1.0-s headway and only 44 ft with a 0.5-s headway. This effectively reduces the space between vehicles (rear bumper to front bumper) to only 25 to 30 ft. This spacing (also called *gap*) would be extremely difficult to maintain.

Headway *includes the vehicle length, while* gap *is the space between vehicles.*

Exhibit 4-6
Time Headway Distribution for Long Island Expressway



Source: Berry and Gandhi (*5*).

Drivers react to this intervehicle spacing, which they perceive directly, rather than to headway. Headway includes the length of the vehicle, which became smaller for passenger cars in the vehicle mix of the 1980s. In the 1990s and 2000s, because of the popularity of sport-utility vehicles, typical vehicle lengths increased. If drivers maintain the same intervehicle spacing and car lengths continue to increase, conceivably, decreases in capacity could result.

If traffic flow were truly random, small headways (less than 1.0 s) could theoretically occur. Several mathematical models have been developed that recognize the absence of small headways in most traffic streams (*6*).

Delay

Delay is the additional travel time experienced by a driver beyond that required to travel at a desired speed. The starting point for measuring delay for HCM purposes is the travel time at free-flow speed. However, it is also possible for reporting purposes to establish a maximum desired travel time, minimum travel speed, or minimum LOS from a transportation agency's point of view (e.g., a travel time for a segment or facility based on the speed at capacity) and to report a *threshold delay* as any additional travel time beyond the established threshold value.

There are several potential sources of delay on uninterrupted-flow facilities:

- *Traffic demand,* increasing levels of which cause drivers to reduce their speed from the free-flow speed because of increased vehicle interactions, as was illustrated in Exhibit 4-2 and Exhibit 4-3;
- *Incidents,* which can reduce the roadway capacity available to serve demand or simply cause drivers to slow down to observe what is happening (e.g., "rubbernecking");
- *Environmental conditions,* such as snow, heavy rain, or sun glare, that cause drivers to reduce their speed from the free-flow speed; and
- *Isolated control features,* such as manual toll collection, inspection stations, railroad grade crossings, or drawbridges on otherwise uninterrupted-flow facilities.

ADDITIONAL INTERRUPTED-FLOW PARAMETERS

Basic concepts for interrupted-flow facilities: intersection control, saturation flow rate, lost time, and queuing.

Interrupted flow can be more complex to analyze than uninterrupted flow because of the time dimension involved in allocating space to conflicting traffic streams. On an interrupted-flow facility, flow usually is dominated by points of fixed operation, such as traffic signals and STOP signs. These controls have different impacts on overall flow.

The operational state of traffic on an interrupted-flow facility is defined by the following measures:

- Volume and flow rate (discussed earlier in the chapter); and
- Control variables (signal, STOP, or YIELD control), which in turn influence
  - Saturation flow and departure headways,
  - Gaps available in the conflicting traffic streams, and
  - Control delay.

Signalized Intersection Flow

Saturation Flow

The most significant source of fixed interruptions on an interrupted-flow facility is traffic signals. A traffic signal periodically halts flow for each movement or set of movements. Movement on a given set of lanes is possible only for a portion of the total time, because the signal prohibits movement during some periods. Only the time during which the signal is effectively green is available for movement. For example, if one set of lanes at a signalized intersection receives a 30-s effective green time out of a 90-s total cycle, only 30/90 or one-third of total time is available for movement on the subject lanes. Thus, flow on the lanes can occur only for 20 min of each hour. If the lanes can accommodate a maximum flow rate of 1,500 veh/h with the signal green for a full hour, they can actually accommodate a total rate of flow of only 500 veh/h, since only one-third of each hour is available as green.

Impact of traffic signal control on maximum flow rate.

When the signal turns green, the dynamics of starting a stopped queue of vehicles must be considered. Exhibit 4-7 shows a queue of vehicles stopped at a signal. When the signal turns green, the queue begins to move. The headway between vehicles can be observed as the vehicles cross the stop line of the intersection. The first headway will be the elapsed time, in seconds, between the initiation of the green and the front wheels of the first vehicle crossing over the stop line. The second headway will be the elapsed time between the front bumpers (or wheels) of the first and second vehicles crossing over the stop line. Subsequent headways are measured similarly.

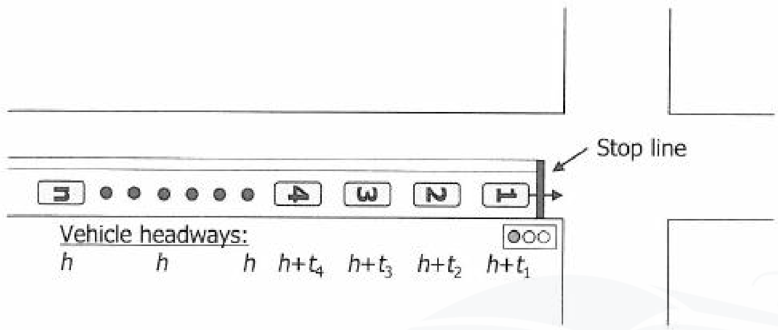


Exhibit 4-7
Acceleration Headways at a Signalized Intersection

The driver of the first vehicle in the queue must observe the signal change to green and react to the change by releasing the brake and accelerating through the intersection. As a result, the first headway will be comparatively long. The second vehicle in the queue follows a similar process, except that the reaction and acceleration period can occur while the first vehicle is beginning to move. The second vehicle will be moving faster than the first as it crosses the stop line, because it has a greater distance over which to accelerate. Its headway will generally be less than that of the first vehicle. The third and fourth vehicles follow a similar procedure, each achieving a slightly lower headway than the preceding vehicle. After four vehicles, the effect of the start-up reaction and acceleration has typically dissipated. Successive vehicles then move past the stop line at a more constant headway until the last vehicle in the original queue has passed the stop line.

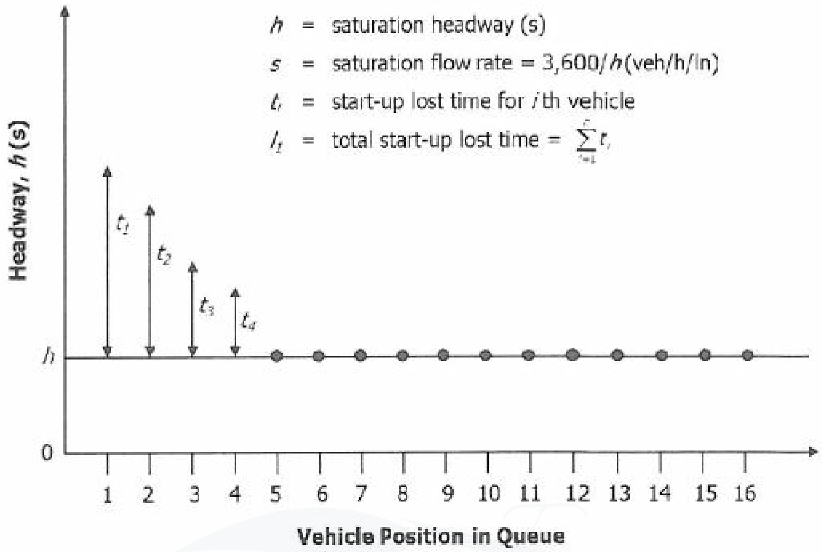
In Exhibit 4-7, this constant average headway, denoted as h, is achieved after four vehicles. The acceleration headways for the first four vehicles are, on the average, greater than h and are expressed as $h + t_i$, where t_i is the incremental headway for the ith vehicle due to the start-up reaction and acceleration. As i increases from 1 to 4, t_i decreases.

Exhibit 4-8 shows a conceptual plot of headways. The HCM recommends using the fifth vehicle following the beginning of a green as the starting point for saturation flow measurements.

The value h represents the *saturation headway*, estimated as the constant average headway between vehicles after the fourth vehicle in the queue and continuing until the last vehicle that was in the queue at the beginning of the green has cleared the intersection.

The reference point on the vehicle used to measure headways is typically the front bumper. Front axles are sometimes the reference point in studies utilizing tube counters to obtain the data.

Exhibit 4-8
Concept of Saturation Flow Rate and Lost Time



Saturation flow rate.

Saturation flow rate is defined as the flow rate per lane at which vehicles can pass through a signalized intersection. It is computed by Equation 4-8:

Equation 4-8

$$s = \frac{3{,}600}{h}$$

where s is the saturation flow rate (veh/h/ln) and h is the saturation headway (s).

The saturation flow rate is the number of vehicles per hour per lane that could pass through a signalized intersection if a green signal was displayed for the full hour, the flow of vehicles never stopped, and there were no large headways.

Lost Time

Each time a flow is stopped, it must start again, with the first four vehicles experiencing the start-up reaction and acceleration headways shown in Exhibit 4-7. In this exhibit, the first four vehicles in the queue encounter headways longer than the saturation headway, h. The increments, t_i, are called start-up lost times.

Total start-up lost time.

The *total start-up lost time* for the vehicles is the sum of the increments, as computed by using Equation 4-9.

Equation 4-9

$$l_1 = \sum_{i=1}^{n} t_i$$

where

l_1 = total start-up lost time (s),

t_i = lost time for ith vehicle in queue (s), and

n = last vehicle in queue.

Each stop of a stream of vehicles is another source of lost time. When one stream of vehicles stops, safety requires some clearance time before a conflicting stream of traffic is allowed to enter the intersection. The interval when no vehicles use the intersection is called *clearance lost time*, l_2. In practice, signal cycles provide for this clearance through change intervals, which can include

Clearance lost time.

yellow or red-clearance indications, or both. Drivers use the intersection during some portion of these intervals.

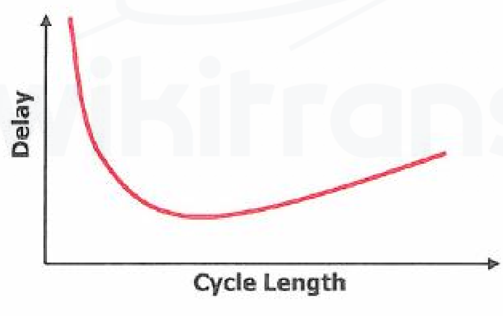
The relationship between saturation flow rate and lost times is critical. For any given lane or movement, vehicles use the intersection at the saturation flow rate for a period equal to the available green time plus the change interval minus the start-up and clearance lost times. Because lost time is experienced with each start and stop of a movement, the total amount of time lost over an hour is related to the signal timing. For example, if a signal has a 60-s cycle length, it will start and stop each movement 60 times per hour, and the total lost time per movement will be $60(l_1 + l_2)$.

Cycle Lengths

Lost time affects capacity and delay. As indicated by the relationship of cycle length to lost time, the capacity of an intersection increases as cycle length increases. However, the capacity increase can be offset somewhat by the observation that the saturation headway, *h*, can be longer when green times are long (e.g., greater than 50 s) (*7*). Capacity increases due to longer cycles are also often offset by the increase in delay that typically results from longer cycles, as discussed below. Other intersection features, such as turning lanes, can also offset the reduced capacity that results from short cycles. Longer cycles increase the number of vehicles in the queues and can cause the left-turn lane to overflow, reducing capacity by blocking the through lanes.

As indicated in Exhibit 4-9, there is a strong relationship between delay and cycle length. For every intersection there is a small range of cycle lengths that will result in the lowest average delay for motorists. Delay, however, is a complex variable affected by many variables besides cycle length.

Exhibit 4-9
Generalized Cycle Length and Delay Relationship



STOP- and YIELD-Controlled Intersection Flow

Two-Way STOP-Controlled Intersections

The driver on the minor street or the driver turning left from the major street at a two-way STOP-controlled intersection faces a specific task: selecting a gap in traffic through which to execute the desired movement. The term *gap* refers to the time interval (*time gap*) and corresponding distance for a given speed (*space gap*) between the major-street vehicles entering an unsignalized intersection, measured from back bumper to front bumper. The term *gap acceptance* describes the completion of a vehicle's movement into a gap.

The capacity of a minor-street approach depends on two factors:

- The distribution of available gaps in the major-street traffic stream, and
- The gap sizes required by drivers in other traffic streams to execute their desired movements.

Gap acceptance.

The distribution of available gaps in the major-street traffic stream depends on the total volume on the street, its directional distribution, the number of lanes on the major street, and the degree and type of platooning in the traffic stream. The gap sizes required by minor-movement drivers depend on the type of maneuver (left, through, right), the number of lanes on the major street, the speed of major-street traffic, sight distances, the length of time the minor-movement vehicle has been waiting, and driver characteristics (eyesight, reaction time, age, etc.).

Critical headway.

For ease of data collection, headways (e.g., front bumper to front bumper) are usually measured instead of gaps, since only half as much data are required (i.e., only front bumper positions need to be recorded, rather than both front and back bumper positions). The *critical headway* is the minimum time interval between the front bumpers of two successive vehicles in the major traffic stream that will allow the entry of one minor-street vehicle. When more than one minor-street vehicle uses one major-street gap, the time headway between the two minor-street vehicles is called *follow-up headway*. In general, the follow-up headway is shorter than the critical headway.

Roundabouts

The operation of roundabouts is similar to that of two-way STOP-controlled intersections. In roundabouts, however, entering drivers scan only one stream of traffic—the circulating stream—for an acceptable gap.

All-Way STOP-Controlled Intersections

At an all-way STOP-controlled intersection, all drivers must come to a complete stop. The decision to proceed is based in part on the rules of the road, which suggest that the driver on the right has the right-of-way, but it is also a function of the traffic condition on the other approaches. The departure headway for the subject approach is defined as the time between the departure of one vehicle and that of the next behind it. A departure headway is considered a saturation headway if the second vehicle stops behind the first at the stop line. If there is traffic on one approach only, vehicles can depart as rapidly as the drivers can safely accelerate into and clear the intersection. If traffic is present on other approaches, the saturation headway on the subject approach will increase, depending on the degree of conflict between vehicles.

Delay

As previously discussed in the section on uninterrupted-flow parameters, delay is the additional travel time experienced by a driver beyond that required to travel at a desired speed, and the starting point for measuring delay for HCM purposes is the travel time at free-flow speed.

Several types of delay are defined for interrupted-flow system elements, but *control delay*—the delay brought about by the presence of a traffic control device—is the principal HCM service measure for evaluating LOS at signalized and unsignalized intersections. Control delay includes delay when vehicles slow in advance of an intersection, time spent stopped on an intersection approach, time spent as vehicles move up in the queue, and time needed for vehicles to accelerate to their desired speed.

The following are other types of delay experienced on interrupted-flow roadways:

- *Traffic delay,* extra travel time resulting from the interaction of vehicles, causing drivers to reduce their speed below the free-flow speed;
- *Geometric delay,* extra travel time created by geometric features that cause drivers to reduce their speed (e.g., delay experienced where an arterial street makes a sharp turn, causing vehicles to slow, or the delay caused by the indirect route that through vehicles must take through a roundabout);
- *Incident delay,* the additional travel time experienced as a result of an incident, compared with the no-incident condition; and
- *Delay due to environmental conditions,* the additional travel time experienced due to severe weather conditions.

Transportation agencies may also choose to report a *threshold delay,* defined as the excess travel time that occurs beyond a defined speed or LOS established by norm (e.g., control delay exceeding LOS B, traffic operating at speeds less than 35 mi/h).

Number of Stops

Traffic control devices separate vehicles on conflicting paths by requiring one vehicle to stop or yield to the other. The stop causes delay and has an associated cost in terms of fuel consumption and wear on the vehicle. For this reason, information about stops incurred is useful in evaluating performance and calculating road user costs. This measure is typically expressed in terms of *stop rate,* which represents the count of stops divided by the number of vehicles served. Stop rate has units of stops per vehicle.

Stops are generally expected by motorists arriving at an intersection as a minor movement (e.g., a turn movement or a through movement on the minor street). However, through drivers do not expect to stop when they travel along a major street. Their expectation is that the signals will be coordinated to some degree such that they can arrive at each signal in succession while it is displaying a green indication for the through movement. For this reason, stop rate is a useful performance measure for evaluating coordinated signal systems.

Queuing

When demand exceeds capacity for a period of time or when an arrival headway is less than the service time (at the microscopic level) at a specific location, a queue forms (2). Queuing is both an important operational measure and a design consideration for an intersection and its vicinity. Queues that are longer than the available storage length can create several types of operational problems. A through-lane queue that extends past the entrance to a turn lane blocks access to the turn lane and keeps it from being used effectively. Similarly, a turn-lane queue overflow into a through lane interferes with the movement of through vehicles. Queues that extend upstream from an intersection can block access into and out of driveways and—in a worst case—can spill back into and block upstream intersections, causing side streets to begin to queue back.

Several queuing measures can be calculated, including the average queue length, the maximum back of queue, and the maximum probable queue (e.g., a 95th percentile queue).

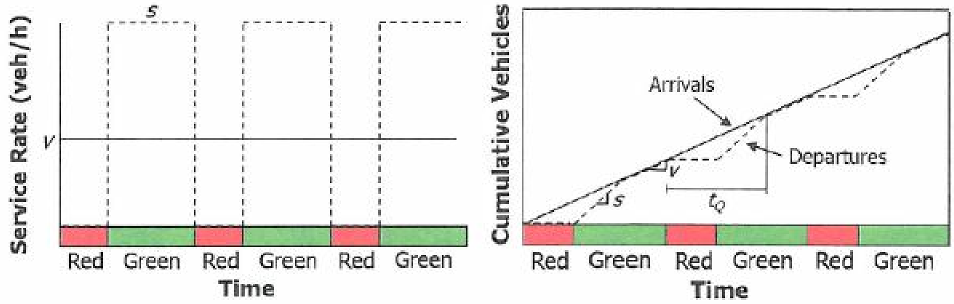
To predict the characteristics of a queuing system mathematically, the following system characteristics and parameters must be specified (*5*):

- Arrival pattern characteristics, including the average rate of arrival and the statistical distribution of time between arrivals;
- Service facility characteristics, including service-time average rates and the distribution and number of customers that can be served simultaneously or the number of channels available; and
- Queue discipline characteristics, such as the means of selecting which customer is next.

The arrival rate exceeds the service rate in oversaturated queues, while the arrival rate is less than the service rate in undersaturated queues. The length of an undersaturated queue can vary but will reach a steady state as more vehicles arrive. In contrast, the length of an oversaturated queue never reaches a steady state; it increases as more vehicles arrive until the arrival demand decreases.

An idealized undersaturated queue at a signalized intersection is shown in Exhibit 4-10. The exhibit assumes queuing on one approach at an intersection with two signal phases. In each cycle, the arrival demand (assumed to be constant in this ideal example) is less than the capacity of the approach, no vehicles wait longer than one cycle, and there is no overflow from one cycle to the next. Exhibit 4-10(a) specifies the arrival rate, v, in vehicles per hour; it is constant for the study period. The service rate, s, has two states: zero when the signal is effectively red and up to the saturation flow rate when the signal is effectively green. Note that the service rate is equal to the saturation flow rate only when there is a queue.

Exhibit 4-10
Idealized Queuing Diagram for a Two-Phase Signalized Intersection



(a) Arrival and Service Rates (b) Queue Lengths

Source: May (*2*).

Exhibit 4-10(b) diagrams cumulative vehicles over time. The horizontal line, v, in Exhibit 4-10(a) becomes the solid line in Exhibit 4-10(b), with the slope of the line equal to the arrival rate. Transferring the service rate from Exhibit 4-10(a) to Exhibit 4-10(b) creates a different graph. During the red period, the service rate is zero, so the service rate is shown as a horizontal dashed line in Exhibit 4-10(b). At the start of the green period, a queue is present, and the service rate is equal to the saturation flow rate. This forms a series of triangles, with the cumulative arrival line as the top side of each triangle and the cumulative service line forming the other two sides, illustrating that a steady state has been reached.

Each triangle represents the queue buildup and dissipation during one cycle length and can be analyzed to calculate the duration of the queue. It starts at the beginning of the red period and continues until the queue dissipates. Its value varies between the effective red time and the cycle length, and it is computed by using Equation 4-10:

$$vt_Q = s(t_Q - r) \text{ or } t_Q = \frac{sr}{s - v}$$

Equation 4-10

where

t_Q = time duration of queue (s),

v = mean arrival rate (veh/h),

s = mean service rate (veh/h), and

r = effective red time (s).

The queue length (i.e., the number of vehicles in the queue, as opposed to the location of the back of the queue) is represented by the vertical distance through the triangle. At the beginning of red, the queue length is zero. It increases to its maximum value at the end of the red period. Then the queue length decreases until the arrival line intersects the service line and the queue length equals zero.

The queuing characteristics can be modeled by varying the arrival rate, the service rate, and the timing plan. In real-life situations, arrival rates and service rates are continuously changing. These variations complicate the model, but the basic relationships do not change.

CAPACITY CONCEPTS

Definition of Capacity

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 a uniform section of a lane or roadway during a given time period under prevailing roadway, environmental, traffic, and control conditions.

Vehicle capacity is the maximum number of vehicles that can pass a given point during a specified period under prevailing roadway, traffic, and control conditions. This assumes that there is no influence from downstream traffic operation, such as queues backing into the analysis point.

Person capacity is the maximum number of persons that can pass a given point during a specified period under prevailing conditions. Person capacity is commonly used to evaluate public transit services, high-occupancy-vehicle lanes, and pedestrian facilities.

Prevailing roadway, environmental, traffic, and control conditions define capacity; these conditions should be reasonably uniform for any segment of a facility that is analyzed. Any change in the prevailing conditions changes a system element's capacity. Thus, an element's capacity can vary from one hour to the next or from one day to the next, as the prevailing conditions (e.g., weather, heavy vehicle percentage, presence or absence of a queue) vary.

Capacity is defined on the basis of reasonable expectancy.

Reasonable expectancy is the basis for defining capacity. That is, the stated capacity for a given system element is a flow rate that can be achieved repeatedly for peak periods of sufficient demand. Stated capacity values can be achieved on system elements with similar characteristics throughout North America. Capacity is not the absolute maximum flow rate observed on such a system element. The absolute maximum flow rate can vary from day to day and from location to location.

Persons per hour, passenger cars per hour, and vehicles per hour are measures that can define capacity, depending on the type of system element and the type of analysis. The concept of person flow is important in making strategic decisions about transportation modes in heavily traveled corridors and in defining the role of transit and high-occupancy-vehicle priority treatments. Person capacity and person flow weight each type of vehicle in the traffic stream by the number of occupants carried.

Base Conditions

Base conditions defined.

Many of the procedures in this manual provide a formula or simple tabular or graphic presentations for a set of specified standard conditions, which must be adjusted to account for prevailing conditions that do not match. These standard conditions are termed *base conditions.*

Base conditions assume good weather, good and dry pavement conditions, users who are familiar with the system element, and no impediments to traffic flow. Other more specific base conditions are identified in each methodological chapter in Volumes 2 and 3.

Prevailing conditions almost always differ from the base conditions.

In most capacity analyses, prevailing conditions differ from the base conditions (e.g., there are trucks in the traffic stream, lanes are narrow). As a result, computations of capacity, service flow rate, and LOS must include adjustments. Prevailing conditions are generally categorized as roadway, traffic, control, operations, or environment.

Roadway Conditions

Impact of roadway conditions.

Roadway conditions include geometric and other elements. In some cases, they influence the capacity of a system element; in others, they can affect a performance measure such as speed, but not the roadway's capacity or maximum flow rate.

Roadway factors include the following:

- Number of lanes,
- The type of system element and its land use environment,
- Lane widths,
- Shoulder widths and lateral clearances,
- Design speed,
- Horizontal and vertical alignments, and
- Availability of exclusive turn lanes at intersections.

The horizontal and vertical alignments of a highway depend on the design speed and the topography of the land on which it is constructed.

In general, as the severity of the terrain increases, capacity and service flow rates are reduced. This is significant for two-lane rural highways, where the severity of terrain can affect the operating capabilities of individual vehicles in the traffic stream and restrict opportunities for passing slow-moving vehicles.

Traffic Conditions

Traffic conditions that influence capacities and service levels include vehicle type, lane or directional distribution, and the driver population.

Vehicle Type

The entry of heavy vehicles—that is, vehicles other than passenger cars (a category that includes small trucks and vans)—into the traffic stream affects the number of vehicles that can be served. Heavy vehicles are vehicles that have more than four tires touching the pavement.

Trucks, buses, and recreational vehicles are the three groups of heavy vehicles addressed by the methods in this manual. As discussed in Chapter 3, Modal Characteristics, heavy vehicles adversely affect traffic in two ways:

- They are larger than passenger cars, so they occupy more roadway space and create larger time headways between vehicles.
- They have poorer operating capabilities than passenger cars, particularly with respect to acceleration, deceleration, and the ability to maintain speed on upgrades.

At signalized intersections, the larger headways produced by trucks decrease the saturation flow rate.

The second impact is more critical. The inability of heavy vehicles to keep pace with passenger cars in many situations creates large gaps in the traffic stream, which are difficult to fill by passing maneuvers. Queues may also develop behind a slow-moving heavy vehicle. The resulting inefficiencies in the use of roadway space cannot be completely overcome. This effect is particularly harmful on sustained, steep upgrades, where the difference in operating capabilities is most pronounced, and on two-lane highways, where passing requires use of the opposing travel lane.

Heavy vehicles also can affect downgrade operations, particularly when downgrades are steep enough to require operation in a low gear. In these cases, heavy vehicles must operate at slower speeds than do passenger cars, again forming gaps ahead and queues behind in the traffic stream.

Directional and Lane Distribution

Two traffic characteristics in addition to the vehicle type distribution affect capacity, service flow rates, and LOS: directional distribution and lane distribution. Directional distribution has a dramatic impact on two-lane rural highway operation, where optimal conditions are achieved when the amount of traffic is roughly equal in each direction. Capacity analyses for multilane highways focus on a single direction of flow. Nevertheless, each direction of the highway is usually designed to accommodate the peak flow rate in the peak direction. Typically, a.m. peak traffic occurs in one direction and p.m. peak traffic occurs in the opposite direction.

Lane distribution is another factor on multilane facilities. Traffic volumes are typically not distributed evenly between lanes, because of drivers pre-positioning themselves for downstream movements (e.g., left turns, exits), vehicle performance characteristics (e.g., heavy vehicles tending to keep right), and local traffic laws (e.g., left lane restricted to passing, trucks prohibited from the left lane), among other factors. The uneven distribution results in less efficient operations than if traffic was more evenly distributed.

Driver Population

It is generally accepted that driver populations who do not use a roadway on a regular basis display characteristics different from those of motorists who are familiar with the roadway. HCM methods allow the user to make an adjustment for driver population, for system elements where driver population has made a difference in the observed capacity. This adjustment is based on user judgment, and the HCM does not provide any quantitative means for determining it.

Control Conditions

For interrupted-flow facilities, the control of the time that specific traffic flows are allowed to move is critical to capacity, service flow rates, and LOS. The most critical type of control is the traffic signal. The type of control in use, signal phasing, allocation of green time, cycle length, and the relationship with adjacent control measures all affect operations.

STOP and YIELD signs also affect capacity, but in a less deterministic way. A traffic signal designates times when each movement is permitted; however, a STOP sign at a two-way STOP-controlled intersection only designates the right-of-way to the major street. Motorists traveling on the minor street must stop to find gaps in the major traffic flow. Therefore, the capacity of minor approaches depends on traffic conditions on the major street. An all-way STOP control requires drivers to stop and enter the intersection in rotation. Capacity and operational characteristics can vary widely, depending on the traffic demands on the various approaches.

Other types of controls and regulations can significantly affect capacity, service flow rates, and LOS. Restricted curb parking can increase the number of lanes available on a street or highway. Turn restrictions can eliminate conflicts at intersections, increasing capacity. Lane use controls can allocate roadway space to component movements and can create reversible lanes. One-way street routings can eliminate conflicts between left turns and opposing traffic.

Technology and Operations

Intelligent transportation systems.

Technological strategies, commonly known as intelligent transportation systems (ITS) strategies, aim to increase the safety and performance of roadway facilities. For this discussion, ITS includes any technology that allows drivers and traffic control system operators to gather and use real-time information to improve vehicle navigation, roadway system control, or both. Research on ITS has grown significantly but cannot be considered comprehensive in terms of evaluating ITS impacts on roadway capacity and quality of service.

Arterial ITS strategies that have been shown to improve vehicular throughput or reduce vehicular delay are adaptive signal control and traffic signal interconnection. A freeway ITS strategy, ramp metering, has improved mainline throughput and speed, while incident management techniques have reduced the time required to identify and clear incidents and thus minimized the time during which capacity is reduced as well as the associated delay. Variable freeway speed limits, combined with automated speed limit enforcement, also show promise but require additional study (*8*).

Other ITS strategies seek to shift demand to alternative routes or times, thus making better use of system capacity and reducing delay on individual facilities. Techniques include parking availability signs at the entrances to downtown areas, value pricing, variable message signs, highway advisory radio, integrated corridor management, real-time travel time and incident information provided to computers and mobile phones, and real-time in-vehicle navigation systems (*8*).

Other strategies for effectively operating roadways are not inherently based on technology, although they may be supported by technology. Examples include managed lanes and highway service patrols.

Specific impacts of technology and operations strategies on roadway capacity and performance are discussed in Chapter 37, ATDM: Supplemental, where research is available to document those impacts.

Environmental Conditions

A facility's capacity can be temporarily reduced by environmental conditions, such as heavy precipitation, adverse lighting conditions, or slippery road surfaces. A number of studies addressing the capacity-reducing effects of specific environmental conditions on freeways have been conducted. The results of these studies are presented in Chapter 10, Freeway Facilities Core Methodology. For interrupted-flow facilities, capacity reductions are reflected by reductions in the saturation flow rate during periods when precipitation is falling and when roadways are wet or covered by snow or ice.

ESTIMATION OF TRAFFIC FLOW PARAMETERS

Analyzing a roadway's performance involves assigning estimated values to traffic flow parameters as a function of either time or distance. There are three common approaches to estimating traffic flow parameters:

1. Deterministic models, such as those presented in the HCM;
2. Simulation models, which take a microscopic and stochastic approach to the representation of traffic flow; and
3. Field data observations, which attempt to measure the parameters directly by data collection and analysis.

All of these approaches can only produce estimates of the parameters of interest. Each approach involves assumptions and approximations. The three approaches are bound together by the common goal of representing field conditions accurately.

On the surface, field observations appear likely to produce the most accurate representation of traffic flow. However, quantitative observations of some traffic phenomena are difficult to produce in a consistent manner that avoids subjective interpretation. There are limits to the accuracy of human observation, and instrumentation of traffic flow data collection is not practical for routine field studies, except for very simple parameters such as flow rate. Field data observations require a level of effort that often exceeds the available resources. Modeling techniques have therefore been introduced as a practical, but approximate, method of estimating required parameters. It is important that modeling techniques be based on definitions and computations that are as consistent as possible with field observations and with each other.

Vehicle trajectories are the lowest common denominator for estimating traffic flow parameters.

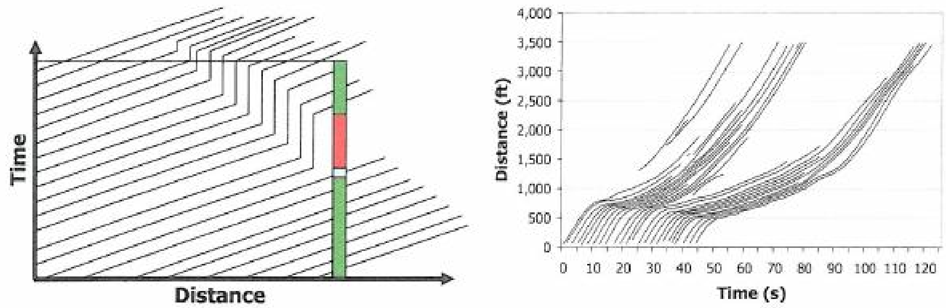
Vehicle time–space trajectories are recognized in the literature as the "lowest common denominator" for this purpose (9). Vehicle trajectories represent the "ground truth" that all measurement and analysis techniques attempt to represent. Microscopic simulation models create trajectories explicitly through algorithms that apply principles of traffic flow theory to the propagation of vehicles along a highway segment. Macroscopic deterministic models do not deal with trajectories at the same level of detail, but they attempt to produce an approximation of the results that would be obtained from trajectory analyses.

Field observations typically establish critical points along individual trajectories rather than complete trajectories.

With a few exceptions involving a significant research effort, field observations are not able to create complete trajectories. Instead, they attempt to establish critical points along individual trajectories. Because of its ability to create complete trajectories, simulation modeling may be viewed as a surrogate for field data collection through which the critical points on the trajectory may be established. Definition of the critical points in a manner that promotes compatibility between the analysis techniques is important.

Vehicle trajectories may be represented graphically or mathematically. The graphical representation shows the position of each vehicle in time and space as it traverses a length of the highway. Typical examples of vehicle trajectory plots are shown in Exhibit 4-11.

Exhibit 4-11
Typical Examples of Vehicle Trajectory Plots



(a) Interrupted Flow on a Signalized Approach (b) Uninterrupted Flow on a Freeway

Exhibit 4-11(a) depicts a classic queue accumulation and release at a signalized stop line. Exhibit 4-11(b) shows a typical freeway situation in which queuing and shock waves are caused entirely by vehicle interactions and not by traffic control devices.

Three characteristics of Exhibit 4-11 are not necessarily common to all time–space representations of vehicle trajectories:

1. Time may be shown on either the vertical or the horizontal axis. Note that Exhibit 4-11(a) shows time on the vertical axis, while Exhibit 4-11(b) shows time on the horizontal axis.
2. The angular shape of the interrupted-flow trajectory curves in Exhibit 4-11(a) does not represent the acceleration and deceleration in their true forms. This shape displays an approximation of the trajectory that is appropriate for some interpretations and inappropriate for others.
3. Both plots represent a single lane of operation in which each vehicle follows its leader according to established rules. Multilane trajectory plots differ from single-lane plots in two ways. First, the first-in, first-out queue discipline can be violated in multilane situations because of overtaking. In other words, a vehicle entering a link later than its leader could leave the link earlier. Graphically, this situation is represented by trajectory lines crossing each other. Second, some vehicles might change lanes. Lane changes cannot be represented in the Exhibit 4-11 plots because distance is shown as a one-dimensional scalar quantity. Because of these complexities, multilane trajectories are much harder to analyze.

While plots such as Exhibit 4-11 provide good visual insight into vehicle operations, they do not support quantitative assessments. To develop performance measures from vehicle trajectories, the trajectories must be represented mathematically rather than visually. A mathematical representation requires development of a set of properties that are associated with each vehicle at specific points in time and space. Because of the time-step formulation of most simulation models, time rather than distance is the preferred reference point.

The key to producing performance measures that are comparable among different estimation techniques is developing a set of definitions that enforce a consistent interpretation of the vehicle trajectories. The subject of trajectory-based definitions is treated in more detail in Chapter 7, Interpreting HCM and Alternative Tool Results, and in Chapter 36, Concepts: Supplemental.

3. PEDESTRIAN MODE

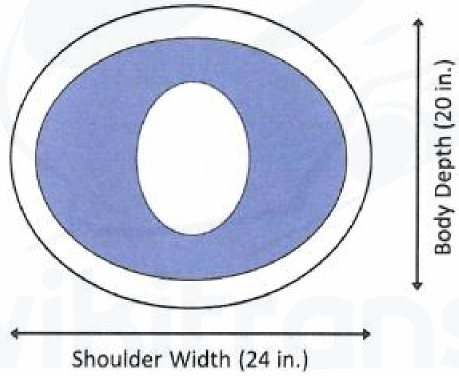
PEDESTRIAN CHARACTERISTICS

Pedestrian Space Requirements

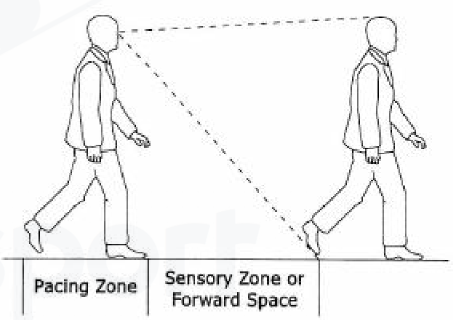
Pedestrian facility designers use body depth and shoulder breadth for minimum space standards, at least implicitly. A simplified body ellipse of 18 in. by 24 in., enclosing an area of 2.35 ft² and incorporating a heavily clothed 95th percentile male and his buffer area to other pedestrians, has been used as the basic space for a single pedestrian, on the basis of 1970s data (*10*). The body ellipse area represents the practical minimum space for standing pedestrians. More recent data, accounting for increases in the body size of the U.S. population since the 1970s, suggest that an extra 2 in. of body depth is required to provide an equivalent buffer area for a U.S. pedestrian in the 2010s. This larger body ellipse of 20 in. by 24 in. encloses an area of 2.6 ft² (*11*) and is shown in Exhibit 4-12(a).

In contrast to a standing pedestrian, a walking pedestrian requires a certain amount of forward space. This forward space is a critical dimension, since it determines the speed of the trip and the number of pedestrians able to pass a point in a given time period. The forward space in Exhibit 4-12(b) is categorized into a pacing zone and a sensory zone (*10*).

Exhibit 4-12
Pedestrian Body Ellipse for Standing Areas and Pedestrian Walking Space Requirement



(a) Pedestrian Body Ellipse



(b) Pedestrian Walking Space Requirement

Sources: Adapted from Fruin (*10*) and *TCRP Report 165: Transit Capacity and Quality of Service Manual*, 3rd edition (*11*).

Walking Speed

Factors affecting walking speed.

Pedestrian walking speed is highly dependent on the characteristics of the walking population. The proportion of elderly pedestrians (65 years old or more) and children in the population, as well as trip purpose, affects walking speed. A national study (*12*) found the average walking speed of younger (age 13–60) pedestrians crossing streets to be significantly different from that of older pedestrians (4.74 ft/s versus 4.25 ft/s, respectively). The 15th percentile speed, the speed used in the *Manual on Uniform Traffic Control Devices* (*13*) for timing the pedestrian clearance interval at traffic signals, was 3.03 ft/s for older pedestrians and 3.77 ft/s for younger pedestrians. Exhibit 4-13 shows these relationships.

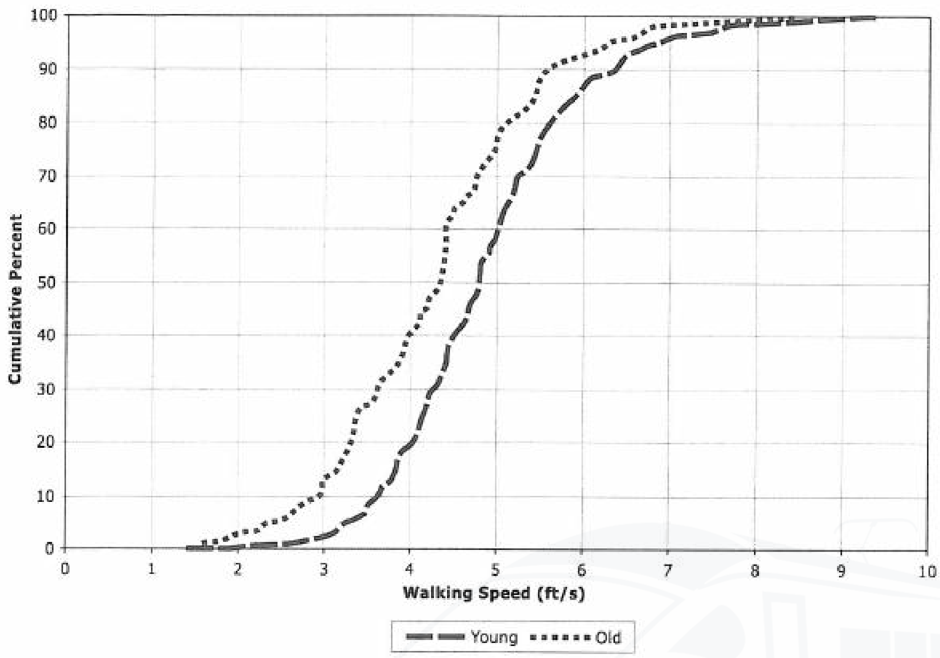


Exhibit 4-13
Observed Older and Younger Pedestrian Walking Speed Distribution at Unsignalized Intersections

Source: Adapted from *TCRP Report 112/NCHRP Report 562* (*12*).

Pedestrian Start-Up Time

At crosswalks located at signalized intersections, pedestrians may not step off the curb immediately when the WALK indication appears, in part because of perception–reaction time and in part to make sure that no vehicles have moved or are about to move into the crosswalk area. This hesitation is termed *pedestrian start-up time* and is used in evaluating pedestrian crosswalks at traffic signals.

PEDESTRIAN FLOW PARAMETERS

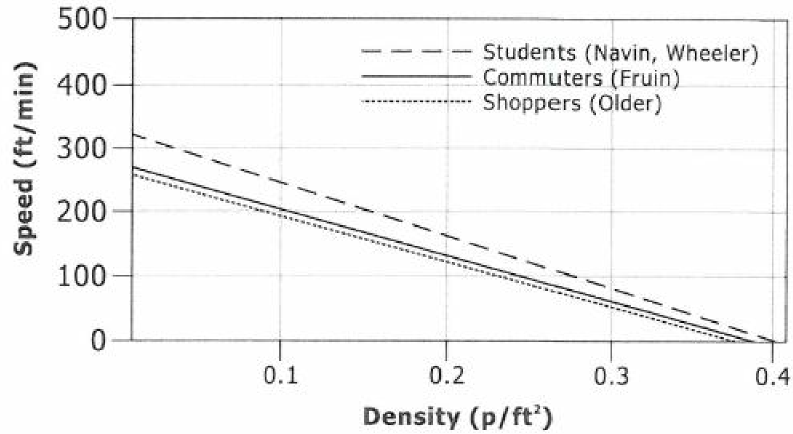
Speed, Flow, and Density Relationships

Speed–Density Relationships

The fundamental relationship between speed, density, and volume for directional pedestrian flow on facilities with no cross flows, where pedestrians are constrained to a fixed walkway width (because of walls or other barriers), is analogous to that for vehicular flow. As volume and density increase, pedestrian speed declines. As density increases and pedestrian space decreases, the degree of mobility afforded to the individual pedestrian declines, as does the average speed of the pedestrian stream.

Exhibit 4-14 shows the relationship between speed and density for three pedestrian classes.

Exhibit 4-14
Relationships Between Pedestrian Speed and Density



Source: Adapted from Pushkarev and Zupan (*14*).

Flow–Density Relationships

Similarities of pedestrian movement to vehicular traffic.

The relationship among density, speed, and directional flow for pedestrians is similar to that for vehicular traffic streams and is expressed in Equation 4-11:

Equation 4-11

$$v_{ped} = S_{ped} \times D_{ped}$$

where

v_{ped} = unit flow rate (p/min/ft),

S_{ped} = pedestrian speed (ft/min), and

D_{ped} = pedestrian density (p/ft²).

The flow variable in Equation 4-11 is the unit width flow, defined as the pedestrians per minute per unit width (e.g., foot) of walkway. An alternative, more useful, expression uses the reciprocal of density, or *space*:

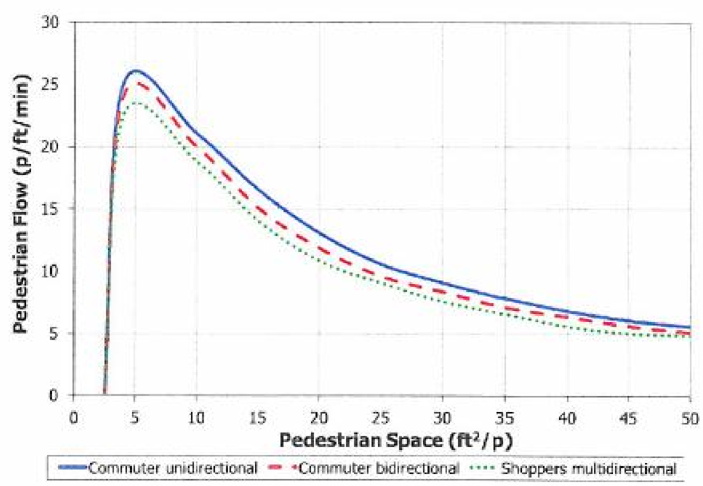
Equation 4-12

$$v_{ped} = \frac{S_{ped}}{M}$$

where M = pedestrian space (ft²/p).

The basic relationship between flow and space is illustrated in Exhibit 4-15:

Exhibit 4-15
Relationships Between Pedestrian Flow and Space



Source: Fruin (*10*).

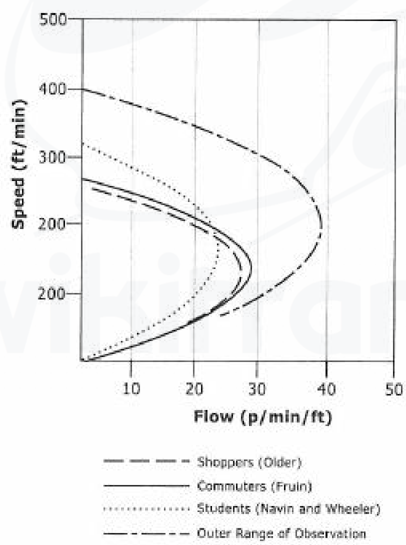
The conditions at maximum flow represent the capacity of the walkway facility. From Exhibit 4-15, it is apparent that all observations of maximum unit flow fall within a narrow range of density, with the average space per pedestrian varying between 5 and 9 ft^2/p. Even the outer range of these observations indicates that maximum flow occurs at this density, although the actual flow in this study is considerably higher than in the others. As space is reduced to less than 5 ft^2/p, the flow rate declines precipitously. All movement effectively stops at the minimum space allocation of 2 to 4 ft^2/p.

These relationships show that pedestrian traffic can be evaluated quantitatively by using basic concepts similar to those of vehicular traffic analysis. At flows near capacity, an average of 5 to 9 ft^2/p is required for each moving pedestrian. However, at this level of flow, the limited area available restricts pedestrian speed and freedom to maneuver.

Speed–Flow Relationships

Exhibit 4-16 illustrates the relationship between pedestrian speed and flow. These curves, similar to vehicle flow curves, show that when there are few pedestrians on a walkway (i.e., low flow levels), there is space available to choose higher walking speeds. As flow increases, speeds decline because of closer interactions among pedestrians. When a critical level of crowding occurs, movement becomes more difficult, and both flow and speed decline.

Exhibit 4-16
Relationships Between Pedestrian Speed and Flow

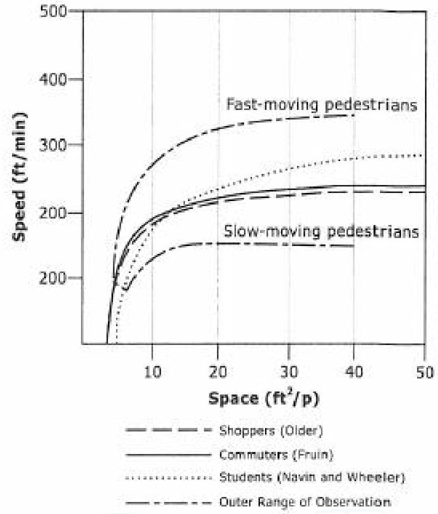


Source: Adapted from Pushkarev and Zupan (*14*).

Speed–Space Relationships

Exhibit 4-17 also confirms the relationships of walking speed and available space. The outer range of observations shown in Exhibit 4-17 indicates that at an average space of less than 15 ft^2/p, even the slowest pedestrians cannot achieve their desired walking speeds. Faster pedestrians, who walk at speeds of up to 350 ft/min, are not able to achieve that speed unless the average space is 40 ft^2/p or more.

Exhibit 4-17
Relationships Between Pedestrian Speed and Space



Source: Adapted from Pushkarev and Zupan (*14*).

Flow on Urban Sidewalks and Walkways

While the fundamental relationships described above hold for pedestrians on constrained facilities with linear flow (e.g., bridges and underground passageways), they are complicated on urban sidewalks and walkways by other factors. In particular, cross flows, stationary pedestrians, and the potential for spillover outside of the walkway affect pedestrian flows on these facilities. Quantitative research describing the effects of these factors on pedestrian flow is limited, but the effects are described qualitatively here.

Cross flows of pedestrians entering or exiting adjacent businesses, getting on or off buses at bus stops, or accessing street furniture are typical on most urban pedestrian facilities. Where pedestrian volumes are high, these cross flows will disrupt the speed–flow relationships described above, resulting in lower pedestrian speeds at equivalent flow rates. In addition, stationary pedestrians will be present on most urban pedestrian facilities as pedestrians stop within the walkway to talk, to look in store windows, or for other reasons. Stationary pedestrians reduce pedestrian flow by requiring pedestrians to maneuver around them and decreasing the available width of the walkway.

The furniture zone is the portion of the sidewalk dedicated to pedestrian amenities (e.g., benches) and is not intended to serve pedestrian flow.

Finally, in situations where pedestrians are not physically confined within the walkway, pedestrians will often choose to walk outside of the prescribed walking area (e.g., walk in the furniture zone or street) when high densities are reached. Thus, in practice, facilities will often break down, with pedestrians spilling over into the street, before the maximum flow rate shown in Exhibit 4-15 is reached.

The result of the combination of factors described above is that many pedestrian facilities will reach effective failure at densities far less than the facility's capacity. Analysis of pedestrian facilities should take into consideration local conditions, including the presence of destinations along the facility that contribute to cross-flows and stationary pedestrians, as well as opportunities for pedestrians to spill over onto adjacent facilities.

Pedestrian Type and Trip Purpose

The analysis of pedestrian flow is generally based on the mean, or average, walking speeds of groups of pedestrians. Within any group, or among groups, there can be considerable differences in flow characteristics due to trip purpose, adjacent land use, type of group, age, mobility, cognitive ability, and other factors.

Pedestrians going to and from work and using the same facilities day after day walk at higher speeds than do shoppers, as was shown in Exhibit 4-14. Older or very young persons tend to walk more slowly than do other groups. Shoppers not only tend to walk more slowly than do commuters but also can decrease the effective walkway width by stopping to window-shop and by carrying shopping bags. The analyst should adjust for pedestrian behavior that deviates from the regular patterns represented in the basic speed, volume, and density curves.

Influences of Pedestrians on Each Other

Photographic studies show that pedestrian movement on sidewalks is affected by other pedestrians, even when space is more than 40 ft^2/p. At 60 ft^2/p, pedestrians have been observed walking in a checkerboard pattern rather than directly behind or alongside each other. The same observations suggest the necessity of up to 100 ft^2/p before completely free movement occurs without conflicts, and that at 130 ft^2/p, individual pedestrians are no longer influenced by others (*15*). Bunching or platooning does not disappear until space is about 500 ft^2/p or higher.

Maintaining flow in the minor (opposing) direction.

Another issue is the ability to maintain flow in the minor direction on a sidewalk when it is opposed by a major pedestrian flow. For pedestrian streams of roughly equal flow in each direction, there is little reduction in the capacity of the walkway compared with one-way flow, because the directional streams tend to separate and occupy a proportional share of the walkway. However, if the directional split is 90% versus 10% and space is 10 ft^2/p, capacity reductions of about 15% have been observed. The reduction results from the minor flow using more than its proportionate share of the walkway.

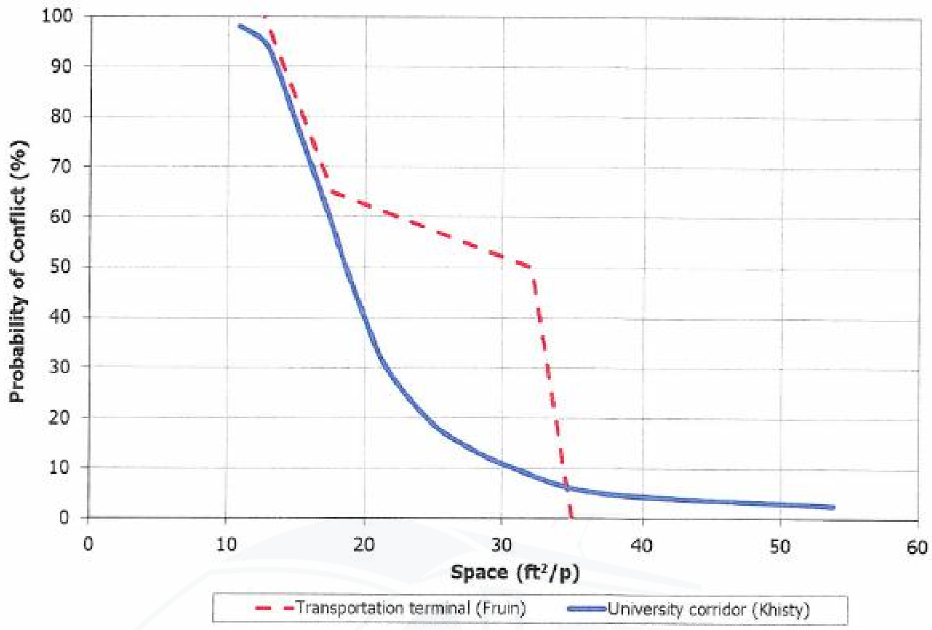
Opposing flows on stairways.

Similar but more severe effects are seen with stairways. In contrast to their behavior on a level surface, people tend to walk in lines or lanes in traversing stairs. A small reverse flow occupies one pedestrian lane (30 in.) of the stairway's width. For a stairway 60 in. (5 ft) wide, a small reverse flow could consume half its capacity (*11*).

Cross flows.

A pedestrian's ability to cross a pedestrian stream is impaired at space values less than 35 ft^2/p, as shown in Exhibit 4-18. Above that level, the probability of stopping or breaking the normal walking gait is nearly zero. Below 15 ft^2/p, almost every crossing movement encounters a conflict. Similarly, the ability to pass slower pedestrians is unimpaired above 35 ft^2/p, but it becomes progressively more difficult as space allocations drop to 18 ft^2/p, the point at which passing becomes virtually impossible (*10, 16*).

Exhibit 4-18
Probability of Conflict Within Pedestrian Cross Flows



Source: Adapted from Fruin (*10*) and Khisty (*16*).

Pedestrian Facility Characteristics

Effective Walkway Width

The lane concept used for highway analysis is frequently not applicable to pedestrian analysis, because studies have shown that pedestrians normally do not walk in organized lanes. The concept is meaningful, however, in the following situations:

- Determining how many pedestrians can walk abreast in a given walkway width—for example, in establishing the minimum sidewalk width that will permit two pedestrians to pass each other conveniently; and
- Determining the capacity of a stairway, since pedestrians will tend to organize into lanes on stairways.

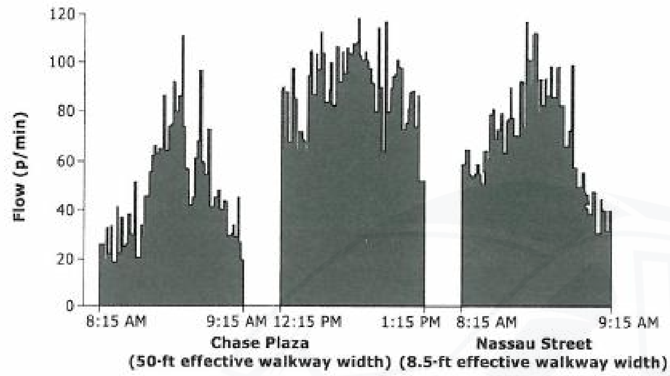
In other situations, the capacity of a pedestrian facility is directly related to the width of the facility. However, not all of the facility's width may be usable, because of obstructions and pedestrians' tendencies to shy away from curbs and building walls. The portion of a pedestrian facility's width that is used for pedestrian circulation is called the *effective width*. The degree to which single obstructions, such as poles, signs, and hydrants, influence pedestrian movement and reduce effective walkway width is not extensively documented. Although a single point of obstruction would not reduce the effective width of an entire walkway, it would affect the obstruction's immediate vicinity.

To avoid interference when two pedestrians pass each other, each should have at least 2.5 ft of walkway width (*14*). When pedestrians who know each other walk close together, each occupies an average width of 26 in., allowing considerable likelihood of contact due to body sway. Lateral spacing less than this occurs only in the most crowded situations.

Pedestrian Platoons

Average pedestrian flow rates are of limited usefulness unless reasonable time intervals are specified. Exhibit 4-19 illustrates that average flow rates can be misleading. The data shown are for two locations in New York City, but the pattern is generally characteristic of concentrated central business districts. Flows during a 1-min interval can be more than double the rate in another, particularly at relatively low flows. Even during the peak 15-min periods, the peak 1-min flow exceeded the average flow by at least 20% and sometimes up to 75%.

Exhibit 4-19
Minute-by-Minute Variations in Pedestrian Flow



Source: Adapted from Pushkarev and Zupan (*14*).

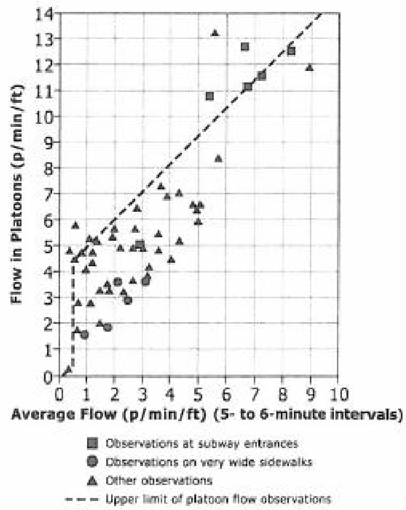
Depending on traffic patterns, a facility designed for average flow can afford a lower quality of flow for a portion of its pedestrian traffic. However, it is not prudent to design for extreme peak 1-min flows that occur only 1% or 2% of the time. A relevant time period should be determined through closer evaluation of the short-term fluctuations of pedestrian flow.

Short-term fluctuations are present in most unregulated pedestrian traffic flows because of the random arrivals of pedestrians. On sidewalks, these random fluctuations are exaggerated by the interruption of flow and queue formation caused by traffic signals. Transit facilities can add surges in demand by releasing large groups of pedestrians in short time intervals, followed by intervals during which no flow occurs. Until they disperse, pedestrians in these types of groups move together as a platoon (Exhibit 4-20). Platoons can also form when passing is impeded because of insufficient space and faster pedestrians must slow down behind slower-moving ones.

Exhibit 4-20
Platoon Flow on a Sidewalk

The scatter diagram shown in Exhibit 4-21 compares the platoon flow rate (i.e., the rate of flow within platoons of pedestrians) with the average flow rate for durations of 5 to 6 min. The dashed line approximates the upper limit of platoon flow observations.

Exhibit 4-21
Relationship Between Platoon Flow and Average Flow



Source: Adapted from Pushkarev and Zupan (*14*).

CAPACITY CONCEPTS

Pedestrian Circulation Facilities

Pedestrian capacity on facilities designed for pedestrian circulation is typically expressed in terms of *space* (square feet per pedestrian) or *unit flow* (pedestrians per minute per foot of walkway width). The relationship between space and flow was illustrated in Exhibit 4-15. Capacity occurs when the maximum flow rate is achieved. Typical values for pedestrian circulation facilities are as follows:

- Walkways with random flow, 23 p/min/ft;
- Walkways with platoon flow (average over 5 min), 18 p/min/ft;
- Cross-flow areas, 17 p/min/ft (sum of both flows); and
- Stairways (up direction), 15 p/min/ft.

As shown in Exhibit 4-16, average pedestrian speeds at capacity are about half the average speed obtained under less congested conditions. As a result, pedestrian circulation facilities are typically not designed for capacity but rather for a less congested condition that achieves lower pedestrian throughput but that provides pedestrians with greater opportunity to travel at their desired speed with minimal conflicts with other pedestrians. Moreover, as described above under "Flow on Urban Sidewalks and Walkways," pedestrian facilities often break down before maximum flow rates are achieved, as a result of pedestrian spillover outside of the walkway into the furniture zone or roadway.

Pedestrian Queuing Facilities

Pedestrian capacity on facilities designed for pedestrian queuing is expressed in terms of space (square feet per pedestrian). In a queuing area, the pedestrian stands temporarily while waiting to be served. In dense, standing crowds, there is little room to move, but limited circulation is possible as the average space per pedestrian increases. Queuing at or near capacity (2 to 3 ft²/p) typically occurs only in the most crowded elevators or transit vehicles. Queuing on sidewalks, waiting to cross at street corners, is more typically in the 3- to 6-ft²/p range, which is still crowded but provides some internal maneuverability.

4. BICYCLE MODE

BICYCLE FLOW PARAMETERS

Although bicyclists are not as regimented as vehicles, they tend to operate in distinct lanes of varying widths when space is available. The capacity of a bicycle facility depends on the number of effective lanes used by bicycles. Shared-lane facilities typically have only one effective lane, but segregated facilities such as bicycle lanes, shoulder bikeways, pathways, and cycle tracks may have more than one effective lane, depending on their width. When possible, an analysis of a facility should include a field evaluation of the number of effective lanes in use. When this is not possible, or when future facilities are planned, a standard width for an effective bicycle lane is 3.5 to 4 ft (*17, 18*). The American Association of State Highway and Transportation Officials recommends that off-street bicycle paths be 10 ft wide (*17*).

The effective bicycle lane width consists of the space used by a bicyclist while riding, plus shy distance to a passing bicyclist. It does not include shy distance to the curb and other elements that influence overall bicycle lane width.

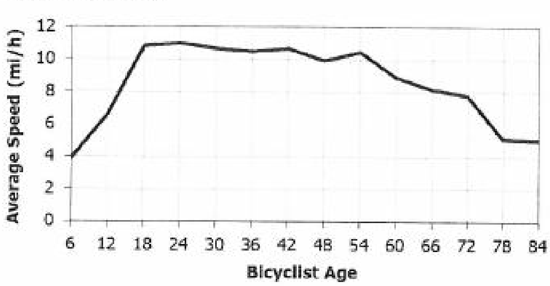
Research demonstrates that three-lane bicycle facilities operate more efficiently than two-lane bicycle facilities, affording considerably better quality of service to users (*19*). The improved efficiency is due primarily to increased opportunities for passing and for maneuvering around other bicyclists and pedestrians. This reinforces the value of determining the number of effective lanes as the principal input for analyzing a bicycle facility.

A study that compared mean bicycle speeds with bicycle flow rates over 5-min periods found at most a minor effect of flow rates on speed, for flow rates ranging from 50 to 1,500 bicycles/h. When the analysis focused on platoons of bicycles with headways less than 5 s, bicycle speeds trended slightly lower as flow rates increased (*20*).

Most bicyclists travel on facilities that are shared with automobiles. In these circumstances, bicycle flow is significantly affected by the characteristics of surrounding automobile flow. Bicyclists often must wait behind queues of automobiles. Even where bicyclists may pass such queues, they are often forced to slow because the available space in which to pass is too constrained to allow free-flow speeds to occur.

Data collected for more than 400 adult bicyclists riding on uninterrupted multiuse segments showed an average speed of 12.8 mi/h (*19*). However, the speed of an individual bicyclist varies considerably from this average on the basis of trail conditions, age, fitness level, and other factors. Exhibit 4-22 shows how bicyclist speed varies with age, on the basis of Danish data. Data are for typical bicyclists on flat terrain.

Exhibit 4-22
Age Effects on Bicyclist Speed



Source: Danish Road Directorate (*21*).

Flow rates of bicyclists usually vary over the course of an hour. As described above for automobiles, HCM analyses typically consider the peak 15 min of flow during the analysis hour. Because inputs to HCM procedures are typically expressed in terms of hourly demands, the HCM uses the PHF, shown by Equation 4-1, to convert an hourly volume into a peak 15-min flow rate. Data for bicycles on eight trails, recorded over three separate time periods for each trail, showed PHFs ranging from 0.70 to 0.99, with an average of 0.85 (*19*).

CAPACITY CONCEPTS

Because service quality deteriorates at flow levels well below capacity, the concept of capacity has little utility in the design and analysis of bicycle paths and other facilities. Capacity is rarely observed on bicycle facilities. Values for capacity, therefore, reflect sparse data, generally from Europe and generally extrapolated from flow rates over time periods substantially less than 1 h.

One study reported capacity values of 1,600 bicycles/h/ln for two-way bicycle facilities and 3,200 bicycles/h/ln for one-way facilities. Both values were for exclusive bicycle facilities operating under uninterrupted-flow conditions (*22*). Other studies have reported values in the range of 1,500 to 5,000 bicycles/h/ln for one-way uninterrupted-flow facilities (*19*).

Danish guidelines suggest that bicycle capacity is normally only relevant at signalized intersections in cities and that a rule of thumb for the capacity of a two-lane cycle track is 2,000 bicycles/h under interrupted-flow conditions (i.e., 1,000 bicycles/h/ln) (*23*). The HCM recommends a saturation flow rate of 2,000 bicycles/h/ln for a one-direction bicycle lane under interrupted-flow conditions, which is equivalent to a capacity of 1,000 bicycles/h/ln when the bicycle lane receives a green indication during 50% of the signal cycle.

DELAY

Delay is an important performance measure for bicyclists on interrupted-flow system elements. This is true because delay increases travel time and because the physical exertion required to accelerate a bicycle makes stopping or slowing undesirable and tiring. The difficulty involved in stopping and starting a bicycle often makes it appropriate to assess not only the control delay incurred by bicyclists but also the number of stops that bicyclists are required to make to traverse a facility. For example, a facility with STOP signs every several hundred feet will require bicyclists to stop frequently and thus will provide lower capacity and quality of service to users.

5. TRANSIT MODE

BUS SPEED PARAMETERS

Bus speeds on urban streets are influenced by the same factors that influence automobile speeds, particularly the delay caused by traffic signals and other forms of intersection control. As heavy vehicles, buses accelerate and decelerate more slowly than passenger cars. In addition, many bus-specific factors influence speed; these involve operations, vehicle, roadway, and passenger characteristics. These factors are described below.

Material in this section generally refers to buses but is also applicable to streetcars and light rail vehicles operating on urban streets, except where specifically stated otherwise.

Bus Operations

Stop Spacing

Unlike other urban street users, most transit vehicles (except for express buses) stop periodically so that passengers may board and alight. Each stop introduces up to seven forms of delay (*11*):

- *Deceleration delay,* as a bus slows down approaching a stop;
- *Bus stop failure,* which occurs when a bus arriving at a stop finds all loading areas occupied and must wait for space to become available;
- *Boarding lost time,* time spent waiting for passengers to travel from their waiting position at the bus stop to the bus door;
- *Passenger service time,* time for passenger loading, unloading, and fare payment, as well as time spent opening and closing the doors;
- *Traffic signal delay,* time spent waiting for a green light after serving passengers at a stop on the near side of an intersection (i.e., a *near-side* stop);
- *Reentry delay,* time spent waiting for a gap in traffic to leave the bus stop; and
- *Acceleration delay,* as a bus speeds up to its running speed on the street.

Increasing the stop spacing reduces the number of occurrences of these types of delay, which results in a net increase in speeds. (Passenger service times may increase, though, as passenger activity is concentrated at fewer stops.) Reported travel time savings due to stop consolidation have ranged from 4.4% to approximately 19% (*11*).

The ability to increase stop spacing depends on many factors, including the quality of the pedestrian network in the area, the locations of transit trip generators and transfer points, and driveway and curb parking locations (*11*).

Stop Location

Bus stop location affects bus speeds by influencing the amount of delay induced by other roadway users—particularly right-turning vehicles—on buses trying to access a bus stop. All other things being equal, far-side stops produce less delay than near-side stops, with the delay benefit increasing with increasing intersection volume-to-capacity ratio and increasing traffic signal cycle length.

However, other factors, such as those listed above for increasing stop spacing, must also be weighed when relocation of stops is considered (*24*).

Stopping Patterns

When a street is used by a high volume of buses, having all buses stop at the same set of stops can create bus congestion and slow down speeds. A *skip-stop* stopping pattern, under which buses are divided into groups that share a certain set of stops, can substantially improve overall bus speeds, as well as bus facility capacity, with the trade-off of making it more difficult for nonregular passengers to find their bus stop. *Platooning* occurs when buses travel together, like cars of a train, along a roadway. Platoons can be developed by traffic signals or can be deliberately formed through careful scheduling and field supervision, although the latter is rare in North America. Platooning minimizes bus passing activity and thus results in higher overall speeds (*11*).

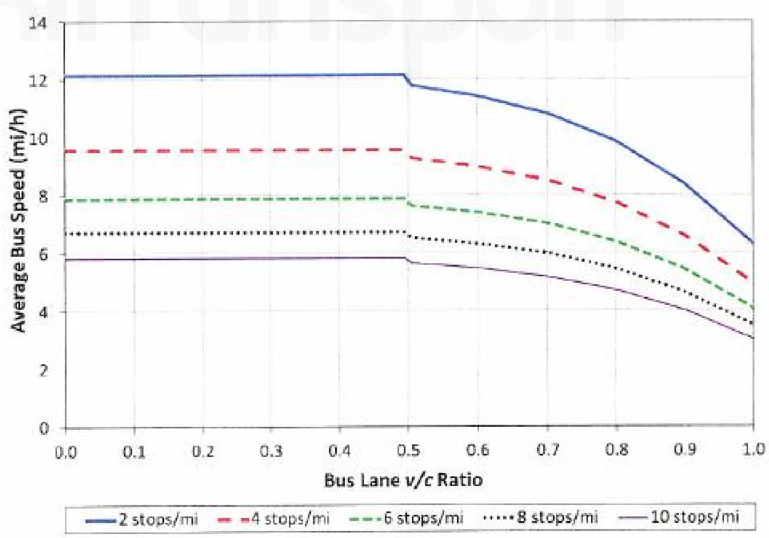
Fare Payment

The time required for passengers to pay a fare affects the passenger service time at stops. The average time needed to board a low-floor bus with no or prepaid (e.g., bus pass or free transfer) fare payment is 1.75 s/passenger. The various types of fare payment methods (e.g., cash, tickets, tokens, magnetic-stripe cards, smart cards) have service times associated with them that increase the service time by up to 3.25 s/passenger, on average, above the base level (*11*).

Service Planning and Scheduling

Bus speeds along an urban street decline when 50% or more of the hourly bus capacity is utilized, as illustrated in Exhibit 4-23. As the number of buses using a bus lane increases, there is a greater probability that one bus will delay other buses, either by using the remaining space at a bus stop or by requiring bus passing and weaving maneuvers. At a volume-to-capacity (v/c) ratio of 1.0, bus speeds are approximately half of those achievable at v/c ratios under 0.5 (*11*).

Exhibit 4-23
Illustrative Bus Speed Relationship to Bus Lane *v/c* Ratio



Source: *TCRP Report 165: Transit Capacity and Quality of Service Manual,* 3rd edition (*11*).
Notes: Assumes 30-s dwell times, 25-mi/h running speed, central business district bus lane with right-turn delays, and typical signal timing. *v/c* ratio = volume-to-capacity ratio.

Passenger Loads

On buses where demand exceeds seating capacity, causing some passengers to stand, more passenger service time (typically 0.5 s/passenger) is required at stops, because standing passengers must push toward the back of the bus to allow other passengers to board and because alighting passengers take longer to get to a door (*11*).

Vehicle Characteristics

Low-floor buses are in common use and eliminate the need for passengers to ascend and descend steps, which would otherwise typically add 0.5 s to each passenger's boarding or alighting time. Wide bus doors allow more passengers to board and alight simultaneously (*11*). Different types of buses have different acceleration characteristics, which influence the amount of acceleration delay incurred when a bus stops.

Roadway Infrastructure

Roadway infrastructure treatments are physical treatments designed to give transit vehicles a travel time advantage over motorized vehicle traffic or to avoid delays caused by other roadway users. The following are common infrastructure treatments used on urban streets:

Refer to TCRP Report 183 *(24) for illustrations and guidance on appropriate locations for transit preferential treatments.*

- *Exclusive bus lanes.* One or more lanes reserved for the full- or part-time use of buses. They restrict or eliminate interactions with other roadway users that slow down buses. With typical signal timing, bus lanes can provide a 1.0- to 1.8-min/mi speed benefit (*11*).
- *Queue jumps.* Short bus lane sections (often shared with a right-turn lane), in combination with an advance green indication for the lane, that allow buses to move past queues of cars at signals. They primarily provide a bus delay benefit at high intersection volume-to-capacity ratios (*24*).
- *Boarding islands.* A raised area within the roadway that allows buses to stop to serve passengers from an inside lane, thus avoiding delays associated with curb-lane travel (e.g., parking, deliveries, right-turning vehicles yielding to pedestrians) (*24*).
- *Curb extensions.* An extension of the sidewalk to the edge of the travel or bicycle lane (e.g., by removing on-street parking). Curb extensions eliminate reentry delay by allowing buses to stop in their travel lane. At the range of curb volumes appropriate for curb extensions (under 500 veh/h), they can save buses up to 4 s of delay per stop on average (*11*).

Traffic Operations

Traffic operations treatments are changes in the roadway's traffic control that are designed to give transit vehicles a travel time advantage over motorized vehicle traffic or to avoid delays caused by other roadway users. The following are common operations treatments used on urban streets:

- *Transit signal priority (TSP).* TSP modifies the traffic signal timing to reduce bus delay while maintaining signal coordination and overall traffic signal cycle length. Systems of intersections equipped with TSP have produced a wide range of results, from no change in corridor-level travel times up to an approximate 20% reduction in travel times. In general, bus travel time variability is reduced by TSP. The ability to obtain corridor-level reductions in travel times depends in part on whether bus schedules are changed to take advantage of TSP, as well as whether a bus is able to pass through the next downstream signal or simply arrives earlier on red (and thus obtains no net benefit) (*24*).

- *Movement restriction exemptions.* Buses are allowed to make movements at locations where other vehicles are not allowed to. This treatment allows buses to travel more direct routes; the time saved depends on the length of and the delay associated with the alternative route (*11*).
- *General traffic movement restrictions.* Motorized vehicles may be prohibited from making movements (e.g., left turns) during times of day when vehicles stopped to make turns would unduly delay other roadway users, including buses. There can also be associated safety and reliability benefits (*24*).
- *Parking restrictions.* Parking restrictions can be used to free roadway space for other uses, such as queue-jump lanes or part-time bus lanes, or to eliminate the traffic delays caused by high parking turnover. The impacts on adjacent land uses must be carefully considered, and regular enforcement is required to ensure that buses receive full benefit (*11*).

Passenger Characteristics

Passenger Distribution

The distribution of boarding passengers among bus stops affects the passenger service time of each stop. If passenger boardings are concentrated at one stop along a street, that street's bus capacity will be lower than if boardings were more evenly distributed. With a lower capacity, fewer scheduled buses in an hour will bring about bus interactions that affect bus speeds.

Strollers, Wheelchairs, and Bicycles

Passenger service times are longer for passengers with strollers or using wheelchairs, particularly with high-floor buses when a lift must be deployed. A passenger using a bicycle rack mounted to the bus will also cause service time to increase, except when other passengers are still being served after the bicycle has been secured. In many cases, these events are sufficiently infrequent to be indistinguishable from the normal variation in passenger demands and service times at a bus stop.

CAPACITY CONCEPTS

Differences Between Transit and Highway Capacity

Transit capacity is different from highway capacity: it deals with the movement of both people and vehicles, it depends on the size of the transit vehicles and how often they operate, and it reflects the interaction of passenger traffic and vehicle flow. Transit capacity depends on the operating policy of the transit agency, which specifies service frequencies and allowable passenger loadings. Accordingly, the traditional concepts applied to highway capacity must be adapted and broadened.

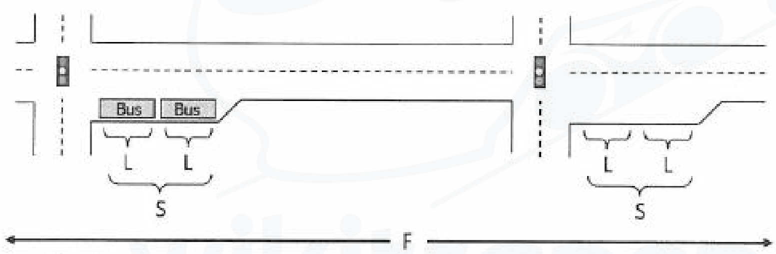
Two key characteristics differentiate transit from the automobile in terms of availability and capacity. First, automobiles have widespread access to roadway facilities, whereas transit service is available only in certain locations and during certain times. Second, roadway capacity is available 24 h/day once it is constructed, but transit passenger capacity is limited by the number of transit vehicles operated at a given time.

The HCM distinguishes between vehicle and person capacity. *Vehicle capacity* reflects the number of buses that pass a given location during a given time period and is thus most closely analogous to automobile capacity. *Person capacity* reflects the number of people that can be carried past a given location during a given time period under specified operating conditions, without unreasonable delay, hazard, or restriction, and with reasonable certainty.

Vehicle Capacity

Vehicle (bus) capacity is commonly determined for three locations along an urban street: individual loading areas (berths) at bus stops, individual bus stops, and an urban street facility, as illustrated in Exhibit 4-24. Each location directly influences the next. The vehicle capacity of a bus stop is controlled by the vehicle capacities of the loading areas, and the vehicle capacity of the urban street facility is controlled by the vehicle capacity of the critical stop within the facility.

Exhibit 4-24
Bus Loading Areas, Stops, and Facilities



L = loading area, S = bus stop, F = bus facility

Source: *TCRP Report 165: Transit Capacity and Quality of Service Manual,* 3rd edition (*11*).

Loading Area Capacity

The following are the main elements determining loading area capacity (*11*):

- *Dwell time,* the sum of passenger service time, boarding lost time, and the time required to open and close the bus doors.
- *Dwell time variability,* the difference in dwell times among different buses using the stop over the course of an hour.
- *Traffic signal timing,* affecting the proportion of time available in an hour for buses to enter (far-side) or exit (near-side) bus stops.
- *Failure rate,* a design input reflecting the desired probability that one bus will arrive at a bus stop only to find all loading areas already occupied. Capacity is improved with higher design failure rates, but speed and reliability suffer when buses must wait in the street to enter a stop.

- *Clearance time,* the sum of the time required for a bus to start up and travel its own length (freeing space for the next bus) and reentry delay.

Bus Stop Capacity

Bus stops consist of one or more loading areas. When a bus stop consists of a single loading area, its capacity is equivalent to the loading area capacity. However, when a bus stop consists of multiple loading areas, the number of loading areas and the design of the loading areas influence its capacity.

Effective loading areas.

Most on-street bus stops are *linear* bus stops, where the first bus to arrive occupies the first loading area, the second bus occupies the second loading area, and so on. Each additional linear loading area at a bus stop is less efficient than the one before it because buses stopped at one of the rear loading areas may block access to available loading areas in front of them.

Efficiency drops significantly above three loading areas. Efficiency is also affected by whether buses stop in or out of the travel lane and by whether platooning occurs (*11*).

Bus Facility Capacity

Bus facility capacity is constrained by the bus stop with the lowest capacity along the facility, or *critical stop*. This stop is usually the bus stop with the longest dwell time. However, a near-side stop at an intersection with high right-turning volumes (particularly in combination with high conflicting crosswalk volumes) or a stop before or after a signalized intersection approach with a short green time could also be the critical stop (*11*).

Person Capacity

For HCM analysis purposes, person capacity is typically calculated only at the facility level. It is determined by three main factors (*11*):

1. *Vehicle capacity,* which determines the maximum number of buses that can be scheduled to use the bus facility over the course of an hour;
2. *Agency policy,* which sets loading standards for buses and determines how frequently buses operate (which is usually less than the maximum possible frequency); and
3. *Passenger demand characteristics,* reflected by a PHF.

6. REFERENCES

Some of these references can be found in the Technical Reference Library in Volume 4.

1. Robertson, H. D., and J. E. Hummer. Volume Studies. In *Manual of Transportation Engineering Studies* (H. D. Robertson, J. E. Hummer, and D. C. Nelson, eds.), Institute of Transportation Engineers, Washington, D.C., 2000.
2. May, A. D., Jr. *Traffic Flow Fundamentals.* Prentice Hall, Englewood Cliffs, N.J., 1990.
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 into the* Highway Capacity Manual. SHRP 2 Report S2-L08-RW-1. Transportation Research Board of the National Academies, Washington, D.C., 2014.
4. Margiotta, R., T. Lomax, M. Hallenbeck, R. Dowling, A. Skabardonis, and S. Turner. *Analytical Procedures for Determining the Impacts of Reliability Mitigation Strategies.* SHRP 2 Report S2-L03-RR-1. Transportation Research Board of the National Academies, Washington, D.C., 2013.
5. Berry, D. S., and P. K. Gandhi. Headway Approach to Intersection Capacity. In *Highway Research Record 453,* Highway Research Board, National Research Council, Washington, D.C., 1973, pp. 56–60.
6. Gerlough, D. L., and M. J. Huber. *Special Report 165: Traffic Flow Theory.* Transportation Research Board, National Research Council, Washington, D.C., 1975.
7. Teply, S., D. I. Allingham, D. B. Richardson, B. W. Stephenson, and J. W. Gough. *Canadian Capacity Guide for Signalized Intersections,* 3rd ed. Institute of Transportation Engineers District 7, Canada, Feb. 2008.
8. Maccubbin, R. P., B. L. Staples, M. R. Mercer, F. Kabir, D. R. Abedon, and J. A. Bunch. *Intelligent Transportation Systems Benefits, Costs, and Lessons Learned: 2005 Update.* Report FHWA-OP-05-002. Federal Highway Administration, Washington, D.C., May 2005.
9. Dowling, R. *Traffic Analysis Toolbox Volume VI: Definition, Interpretation, and Calculation of Traffic Analysis Tools Measures of Effectiveness.* Federal Highway Administration, Washington, D.C., 2007.
10. Fruin, J. J. *Pedestrian Planning and Design,* rev. ed. Elevator World, Inc., Mobile, Ala., 1987.
11. Kittelson & Associates, Inc.; Parsons Brinckerhoff; KFH Group, Inc.; Texas A&M Transportation Institute; and Arup. *TCRP Report 165: Transit Capacity and Quality of Service Manual,* 3rd ed. Transportation Research Board of the National Academies, Washington, D.C., 2013.
12. Fitzpatrick, K., S. M. Turner, M. Brewer, P. J. Carlson, B. Ullman, N. D. Trout, E. S. Park, J. Whitacre, N. Lalani, and D. Lord. *TCRP Report 112/NCHRP Report 562: Improving Pedestrian Safety at Unsignalized Crossings.* Transportation Research Board of the National Academies, Washington, D.C., 2006.

13. *Manual on Uniform Traffic Control Devices.* Federal Highway Administration, Washington, D.C., 2009.

14. Pushkarev, B., and J. M. Zupan. *Urban Space for Pedestrians: A Report of the Regional Plan Association.* Massachusetts Institute of Technology Press, Cambridge, 1975.

15. Hall, E. T. *The Hidden Dimensions.* Doubleday and Co., Garden City, N.Y., 1966.

16. Khisty, C. J. Pedestrian Cross Flow Characteristics and Performance. *Environment and Behavior,* Vol. 17, No. 6, Nov. 1985, pp. 679–695.

17. *Guide for the Development of Bicycle Facilities.* American Association of State Highway and Transportation Officials, Washington, D.C., 1999.

18. *Urban Bikeway Design Guide,* 2nd ed. National Association of City Transportation Officials, New York, 2012.

19. Hummer, J. E., N. M. Rouphail, J. L. Toole, R. S. Patten, R. J. Schneider, J. S. Green, R. G. Hughes, and S. J. Fain. *Evaluation of Safety, Design, and Operation of Shared-Use Paths—Final Report.* Report FHWA-HRT-05-137. Federal Highway Administration, Washington, D.C., July 2006.

20. Botma, H., and H. Papendrecht. Traffic Operation of Bicycle Traffic. In *Transportation Research Record 1320,* Transportation Research Board, National Research Council, Washington, D.C., 1991, pp. 65–72.

21. *Collection of Cycle Concepts.* Vejdirektoratet (Danish Road Directorate), Copenhagen, 2000.

22. Botma, H. Method to Determine Level of Service for Bicycle Paths and Pedestrian–Bicycle Paths. In *Transportation Research Record 1502,* Transportation Research Board, National Research Council, Washington, D.C., 1995, pp. 38–44.

23. *Kapacitet og serviceniveau* (Capacity and Service Level). Vejdirektoratet—Vejregelrådet (Danish Road Directorate, Road Standards Committee), Copenhagen, May 2008.

24. Ryus, P., K. Laustsen, K. Blume, S. Beaird, and S. Langdon. *TCRP Report 183: A Guidebook on Transit-Supportive Roadway Strategies.* Transportation Research Board of the National Academies, Washington, D.C., 2016.