Guide to Traffic Management Part 2

Traffic Theory Concepts

Austroads

Guide to Traffic Management Part 2: Traffic Theory Concepts

Sydney 2020

Guide to Traffic Management Part 2: Traffic Theory Concepts

Edition 3.0 prepared by: David Green and Kenneth Lewis

Edition 3.0 project manager: Richard Delplace

Abstract

The Austroads *Guide to Traffic Management* has 13 Parts and provides comprehensive coverage of traffic management guidance for practitioners involved in traffic engineering, road design and road safety.

Guide to Traffic Management Part 2: Traffic Theory Concepts provides practitioners with the theoretical background necessary to appreciate the nature of traffic behaviour and to undertake analyses required in the development and assessment of both traffic management plans and road design proposals.

Keywords

Traffic analysis, traffic behaviour, traffic characteristics, traffic flow, traffic interactions, traffic theory, probability theory, headway distribution, queue, queuing, gap acceptance, car following, traffic bunches, bunching, overtaking, platoon dispersion, congestion management.

Edition 3.0 published April 2020

Edition 3.0 of the Guide has been updated with a new name and introduction to better reflect the content.

Edition 2.0 published October 2015

Edition 1.0 published July 2008

ISBN 978-1-925854-76-3

Austroads Project No. NTM6100 & NTM6118

Austroads Publication No. AGTM02-20

Pages 122

Publisher

Austroads Ltd.
Level 9, 287 Elizabeth Street
Sydney NSW 2000 Australia
Phone: +61 2 8265 3300

austroads@austroads.com.au
www.austroads.com.au

About Austroads

Austroads is the peak organisation of Australasian road transport and traffic agencies.

Austroads' purpose is to support our member organisations to deliver an improved Australasian road transport network. To succeed in this task, we undertake leading-edge road and transport research which underpins our input to policy development and published guidance on the design, construction and management of the road network and its associated infrastructure.

Austroads provides a collective approach that delivers value for money, encourages shared knowledge and drives consistency for road users.

Austroads is governed by a Board consisting of senior executive representatives from each of its eleven member organisations:

- Transport for NSW
- Department of Transport Victoria
- Queensland Department of Transport and Main Roads
- Main Roads Western Australia
- Department of Planning, Transport and Infrastructure South Australia
- Department of State Growth Tasmania
- Department of Infrastructure, Planning and Logistics Northern Territory
- Transport Canberra and City Services Directorate, Australian Capital Territory
- Department of Infrastructure, Transport, Regional Development and Communications
- Australian Local Government Association
- New Zealand Transport Agency.

Acknowledgements

Edition 1.0 prepared by David Bennett and project managed by John Erceg. Edition 2.0 prepared by: Clarissa Han, James Luk and Kaveh Bevrani and project managed by Dave Landmark.

Contents

Tables

Figures

1. Introduction

1.1. Purpose

The purpose of the *Austroads Guide to Traffic Management Part 2 (Traffic Theory Concepts)* is to provide an overview of the concepts used in traffic management in Australia and New Zealand.

Traffic theory is a complex area of study and it is not the intention of Part 2 to discuss it in detail. Instead, Part 2 aims to provide practitioners with the theoretical background necessary to appreciate the nature of traffic behaviour and to understand the theoretical concepts regarding traffic behaviour. Understanding traffic theory concepts can be used to develop strategies for the management of traffic.

The aim of this part is to encourage harmonisation of practice throughout Australia and New Zealand.

1.2. Intended User

The intended users are those practitioners who require a basic understanding of the traffic theory concepts used to analyse traffic by introducing the characteristics of traffic flow and the theories, models and statistical distributions used to describe many traffic phenomena. Part 2 provides an overview of the concepts that practitioners should consider when undertaking or commissioning traffic analysis.

1.3. How to Use

This part provides an overview of traffic theory concepts. It will help readers understand the concepts behind traffic flow. This is important as the traffic theory concepts underpin many treatments that are used when managing road corridors to achieve movement and place and network operation objectives for the road corridor network. The movement and place and network operation objectives for network management are outlined in Part 4.

This part should not be read in isolation with respect to understanding the full life cycle of implementing transport engineering treatments. Readers should understand the scope of this part and what is out of scope and when to refer to other parts of the Guide (refer to Table 1.1), other Guides, and documents external to Austroads. At the very least, before reading this part readers should access Part 1 to get an understanding of the *Guide to Traffic Management* and the context in which the guidance is provided.

The series provides guidance for good practice. Practices that differ from the Guide should be based on sound engineering judgement and be approved by the relevant road agency.

Within this part some key terms are used. Outlined below are the terms along with an explanation on the context that they should be view in:

- **Road:** Road refers to the road corridor. That is, the space between property boundaries located either side of the road and includes footpaths. This part provides guidance on the management of all travel modes permitted to travel within the road corridor.
- **Traffic:** While the term ‘traffic’ may be closely identified with vehicle traffic, it can also refer to any mode of traffic that may be used within the road corridor. This includes vehicle, freight, pedestrian, bicycle, bus and tram traffic etc. Therefore, traffic has the same meaning as transport in the context of this part. Where traffic is mentioned on its own and without identification of its type, it should be viewed as capturing all modes. Where it is specifically identified (e.g. bicycle traffic) it should be viewed in this form.
- **Transport:** Transport has the same meaning as traffic as it refers to any mode of transport that may be used within the road corridor. This includes general vehicles, freight, pedestrians, bicycles, buses and trams etc.

1.4. Scope

The *Austroads Guide to Traffic Management* series comprises 13 parts as outlined in Table 1.1. Part 2 is a 'theory and background' document and provides:

- an overview of the basic traffic variables and relationships and the random nature of traffic behaviour
- an overview of queuing and gap acceptance theory
- an overview of vehicle interactions in moving traffic and the principles underlying managed motorways.

Part 2 aims to give users a basic appreciation of traffic flow theory so that they can use this to understand traffic and how to manage it, and where they need to delve deeper into traffic flow theory.

It is noted that jurisdictions may have variations to practices outlined in this part. These are typically contained in supplements along with other complementary material. Readers should refer to such supplements to best understand jurisdictional practice.

Table 1.1: Parts of the Guide to Traffic Management

Part	Title	Type	Content
Part 1	Introduction to the Guide to Traffic Management	Theory and background	Part 1 provides background material to the Austroads *Guide to Traffic Management* through: • an introduction to the discipline of transport management • an outline of the Guide.
Part 2	Traffic Theory Concepts	Theory and background	Part 2 provides an outline of traffic theory concepts through: • an overview of the basic traffic variables and relationships and the random nature of traffic behaviour • an overview of queuing and gap acceptance theory • an overview of vehicle interactions in moving traffic and the principles underlying managed motorways.
Part 3	Transport Study and Analysis Methods	Theory and background	Part 3 provides details of the methods used for transport studies and analysis through: • describing an overall systems approach to transport studies, including statistical and sampling issues • presenting an overview of the concepts of capacity, level of service and degree of saturation and the factors which affect them • providing information and guidance on capacity analysis as applied to uninterrupted flow facilities, interrupted flow facilities and intersections, respectively.
Part 4	Network Management Strategies	Strategic	Part 4 provides an outline of the broad strategies for managing road corridor networks for the various road corridor users through: • an overview of what is network management • describing the 'movement and place' concept for road corridor management and the movement and place considerations for network types based on the user group (e.g. rural networks, bicycle networks, heavy vehicle networks and others) • an overview of the network operation planning process to be applied in road corridor management.
Part 5	Link Management	Operational road corridor management	Part 5 provides an outline of road corridor link management through: • an overview of network strategies that influence road corridor link management (this is also discussed in Part 4) • principles of access management • guidance on road corridor space allocation for general use and for specific road corridor user types • guidance on lane management principles • guidance on speed limit setting.

Part	Title	Type	Content
Part 6	Intersections, Interchanges and Crossings Management	Operational road corridor management	Part 6 provides an outline of road corridor intersection management through an overview of: • the type of intersections and guidance on how to select the appropriate intersection • roundabouts including guidance on their use, performance, road corridor space allocation and lane management, functional design and signalised roundabouts • signalised intersections including guidance on functional layout, road corridor space allocation and lane management • unsignalised intersections including transport controls, intersection capacity and flow and transport control devices • road interchanges including planning and route considerations, road corridor space allocation, interchange forms, and basic lane numbers and lane balance • rail crossings including types of protection, grade separated and at-grade crossings, path crossings, lighting and selection of treatments • pedestrian and cyclist crossings including mid-block crossings, bicycle treatment at intersections and intersections of paths.
Part 7	Activity Centre Transport Management	Operational road corridor management	Part 7 provides an outline of road corridor management within activity centres through an overview of: • principles and objectives of road corridor management in activity centres • techniques for transport management in activity centres • examples and typical issues associated with various types of activity centres.
Part 8	Local Street Management	Operational road corridor management	Part 8 provides an outline of how to manage local streets through the implementation of LATM. This is done through guidance on design considerations and an overview of: • the principles behind LATM • the LATM planning process and the steps to be taken • the objective decision process for LATM • considerations such as community participation and information, legal aspects and duty of care • types of LATM devices and guidance to aid selection.
Part 9	Transport Control Systems – Strategies and Operations	Operational road corridor management	Part 9 provides an outline of how to operate dynamic transport control systems through an overview of: • objectives and principles of transport operations • traffic operation services, measures and tools • systems and procedures for – operating traffic management centres – maintaining road corridor serviceability and safety through dynamic operations – operating signals, lane management systems, variable speed limits and other ITS for transport control – operating a smart motorway • performance indicators for operating arterial roads.

Part	Title	Type	Content
Part 10	Transport Control – Types of Devices	Operational road corridor management	Part 10 provides an outline of the transport control devices available to road agencies to manage the road corridor through an overview of: • principles and application of transport control devices • signage and marking schemes needed to manage the road corridor network in a holistic way • types of signs (both static and electronic) available and guidance on their use and management • electronic signs and guidance on their use and management but not going into the operation of the variable message signs (VMS) • types of pavement markings and guidance on their use and management • use of guide posts and delineators • traffic signals and guidance on their use and management but not going into the operation of the signals • traffic islands and guidance on their use and management • current and emerging devices that utilise direct communication with equipped vehicles and may be used for transport control purposes.
Part 11	Parking Management Techniques	Operational road corridor management	Part 11 provides guidance on parking management through an overview of: • principles of parking management • factors which influence parking demand • factors which impact on the supply of parking • parking policy framework • impacts of parking on the environment • key elements to consider in the management of off-street parking • key elements to consider in the management of on-street parking • issues impacting rural parking • considerations for park-and-ride facilities • devices used in the management and control of parking (e.g. signs and pavement markings) • duty of care and risk management required for parking management.
Part 12	Integrated Transport Assessments for Developments	Theory and background	Part 12 provides guidance on how to identify and assess the potential impacts of land developments on road corridor management through an overview of: • traffic impacts of developments and the need to assess them • how transport impact assessments fit into the overall planning regime • key considerations that need to be made to enable developments to function from a transport perspective • traffic impact assessments including how to conduct them • some of the other assessments beyond transport flow considerations that should be addressed (e.g. road corridor safety, infrastructure and pavement, and environmental).
Part 13	Safe System Approach to Transport Management	Theory and background	Part 13 provides guidance on how the Safe System philosophy links in with the other Guides through an overview of: • concepts required to achieve a safe road corridor • human factors and the need to consider these in the design and management of roads to achieve a Safe System • concepts behind road corridor safety engineering • key elements to be managed or applied in safety engineering of the road corridor.

1.5. Out of Scope

The subjects that are out of scope in this part are as follows:

- It does not delve into complex traffic theory. For detailed guidance on this, readers should refer to reference material addressing the subject matter which is referred to in this part.
- It does not provide guidance on transport management principles which are outlined in other parts as outlined in Table 1.1.
- It does not discuss design aspects; these are referred to in the *Guide to Road Design*.
- Road corridor safety is captured in the guidance covered throughout this Guide series. Further information on safety is provided in Part 13 and the *Guide to Road Safety*.
- It does not go into broader transport planning guidance. For this, readers are referred to the Australian Transport Assessment and Planning (ATAP) framework which provides a comprehensive framework for planning, assessing and developing transport systems and related initiatives. The ATAP is located at www.atap.gov.au.

2. Basic Traffic Variables and Relationships

2.1. Basic Descriptors of Traffic Flow

2.1.1. Volume

Volume (sometimes called 'flow' or 'flow rate' and here designated by the symbol 'q') is the number of vehicles per unit time passing a given point on a road. Volume may relate to a lane, a carriageway or a road and, in the case of a road, may include traffic in either one or both directions. In traffic flow analysis, a volume usually relates to only one direction of flow.

The unit of time used in relation to a volume may vary according to the application. Volumes expressed as vehicles per second (veh/s) or vehicles per hour (veh/h) are typically used in traffic flow analysis, while daily or annual volumes may be appropriate in other contexts, such as analyses of traffic growth over time.

2.1.2. Density

Density (also known as 'concentration') ('k') is the number of vehicles present within a unit length of lane, carriageway or road at a given instant of time. Density is usually expressed as vehicles per kilometre (veh/km) or, where appropriate for analysis purposes, vehicles per metre (veh/m).

2.1.3. Speed

Speed ('v') is the distance travelled by a vehicle per unit time and is typically expressed as either metres per second (m/s) or kilometres per hour (km/h).

The average speed of a stream of vehicles may be expressed as either the time mean speed or the space mean speed, which are defined below.

Time mean speed, v_t, is the arithmetic mean of the measured speeds of all vehicles passing a given point during a given time interval. Such individual measured speeds are called 'spot speeds'.

Space mean speed, v_s, is the arithmetic mean of the measured speeds of all vehicles within a given length of lane or carriageway, at a given instant of time.

In Section 2.3, it is shown that, for any given traffic stream, space mean speed can be estimated as the harmonic mean (the inverse of the mean of speed inverses as in Equation 2.5) of the same spot speeds whose arithmetic mean is the time mean speed.

2.1.4. Headway

A headway is the time interval separating the passing of a fixed point by two consecutive vehicles in a traffic stream. The average headway ('h') of the stream over a given time interval is the arithmetic mean of the series of headways occurring over that interval. Headway is usually expressed in units of seconds per vehicle (s/veh).

2.1.5. Spacing

A spacing is the distance between the fronts of two consecutive vehicles in a traffic stream at a given instant of time. The average spacing ('s') of the stream over a given length of lane or carriageway is the arithmetic mean of the individual spacings occurring over that length at that instant of time. Spacing is usually expressed in the units of metres per vehicle (m/veh).

2.1.6. Lane Occupancy

Lane occupancy is not one of the five principal traffic flow descriptors but can be very useful (see, for example, Section 7.5.2). It is the proportion of time, over a given time interval, that there is a vehicle present at a specified point in the lane. Given this definition, lane occupancy is a dimensionless measure.

While the definition applies to 'a specified point in the lane' over 'a given time interval', essentially the same lane occupancy will be experienced (for the same time interval) along any length of lane over which vehicles cannot enter or leave the lane. This will be true as long as the length is such that a vehicle's travel time over that length is short compared to the time interval used in calculating the occupancy.

It is worth noting that lane occupancy (average carriageway occupancy) is utilised in control systems for motorways. It is therefore important that an understanding of the importance of the accuracy of information captured to determine lane occupancy is understood.

2.2. Mathematical Relationships

2.2.1. Fundamental Relationships

As a consequence of the above definitions of the five principal traffic flow descriptors – volume, density, speed, headway and spacing – the following relationships exist between their average values for any traffic stream.

a. Volume and average headway are the inverses of each other.

$$q = 1 / h \text{ and } h = 1 / q \quad (2.1)$$

b. Density and average spacing are the inverses of each other.

$$k = 1 / s \text{ and } s = 1 / k \quad (2.2)$$

c. Volume is the product of density and (space mean) speed.

$$q = k . v \quad (2.3)$$

From these three fundamental relationships, various others may be directly derived, for example,

$$q = v / s \text{ and } s = h . v$$

2.3. Time and Space Mean Speed Relationships

In Section 2.1.3, it was noted that while the time mean speed is the arithmetic mean of a set of spot speeds measured for a traffic stream, the space mean speed of that traffic stream is the harmonic mean of the same set of spot speeds. This can be demonstrated as follows:

Let the values v_j, j = 1, . . . , N, be the spot speeds of N vehicles in a traffic stream, measured as they pass a given point during a unit time interval. By the definition given in Section 2.1.3, the time mean speed is the arithmetic mean of the spot speeds, that is:

$$v_t = \frac{\sum_{j=1}^{N} v_j}{N} \quad (2.4)$$

Now, without loss of generality, the stream of traffic can be considered to consist of n component sub-streams, i = 1, . . ., n, where sub-stream i has volume q_i and all vehicles in the sub-stream have the same speed, v_i. The density, k_i, of sub-stream i is then such that $q_i = v_i \, . \, k_i$.

In a unit length of road at a given instant, there are k_i vehicles from sub-stream i travelling at speed v_i, for all i, i = 1, . . . , n. Thus, by its definition in Section 2.1.3, the space mean speed is:

$$v_s = \frac{\sum_{i=1}^{n} k_i . v_i}{\sum_{i=1}^{n} k_i} = \frac{\sum_{i=1}^{n} q_i}{\sum_{i=1}^{n} \frac{q_i}{v_i}} = \frac{N}{\sum_{j=1}^{N} \frac{1}{v_j}} \qquad 2.5$$

This is the harmonic mean of the spot speeds.

Equation 2.5 also identifies the space mean speed as the length of a given section of lane or carriageway divided by the average travel time of vehicles in the traffic stream over that length. This can be seen by noting that, if that length is L, the final quotient in Equation 2.5 can be written equivalently as:

$$\frac{L}{\left(\sum_{j=1}^{N} \frac{L}{v_j}\right)/N}$$

For any given traffic stream, time mean speed is always greater than space mean speed except when all vehicles have exactly the same speed, in which case the two mean speeds are equal. This is further discussed, with the help of a numerical example, in Commentary 1.

[see Commentary 1]

Wardrop (1952) showed that, if σ_s^2 is the variance of space speeds (as defined in Commentary 1), an approximate relationship between the two mean speeds is:

$$V_t = V_s + \frac{\sigma_s^2}{v_s} \qquad 2.6$$

Finally, the correct average speed to use in Equation 2.3 is the space mean speed, rather than the time mean speed. This can be demonstrated by first observing the total volume of the traffic stream considered in Section 2.2.1 is:

$$q = \sum_{i=1}^{n} q_i \qquad 2.7$$

and the total density is:

$$k = \sum_{i=1}^{n} k_i = \sum_{i=1}^{n} \frac{q_i}{v_i} \qquad 2.8$$

Therefore, as shown in Equation 2.5,

$$\frac{q}{k} = \frac{\sum_{i=1}^{n} q_i}{\sum_{i=1}^{n} \frac{q_i}{v_i}} = V_s \qquad 2.9$$

That is, the quotient of q and k is the space mean speed.

2.4. Graphical Relationships for Uninterrupted Flow

Uninterrupted flow occurs in a traffic stream that is not delayed or interfered with by factors external to the traffic stream itself (such as intersections, pedestrian crossings etc.) but only by its own, internal, traffic interactions. In contrast, **interrupted flow** occurs when external factors have significant effects on the traffic flow.

It is instructive to examine the graphical relationships between the three principal traffic flow descriptors, volume, speed and density, for the case of uninterrupted flow, to examine the correspondence between the different graphical relationships and to interpret them in terms of traffic conditions on the road.

Figure 2.1 illustrates the relationship between (average) speed and volume. Feasible combinations of these two variables must lie within the region bounded by the vertical axis (zero minimum volume), the maximum feasible speed line, the minimum feasible average headway line (defining maximum feasible volume) and the maximum feasible density line (that maximum being the tangent of the angle θ). Typical observed speed-volume combinations lie on or close to the curve shown. The diagram is interpreted in terms of on-road traffic conditions later in this section.

Figure 2.1: Speed-volume relationship

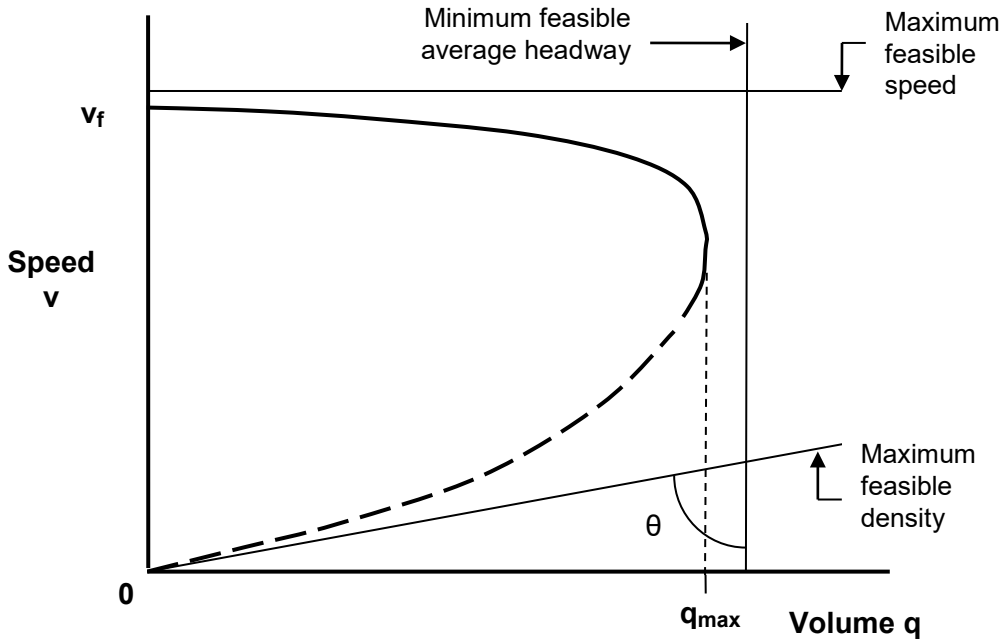


Figure 2.2 shows the typically observed relationship between (average) speed and (average) density for uninterrupted flow conditions. In this close-to-linear relationship, the speed steadily decreases from its maximum value, the **mean free speed**, v_f, when the density is at or close to zero, to zero speed at the maximum or 'jam' density, k_j. If a rectangle is drawn with one corner at the origin and the diagonally opposite corner at a point of interest, then its area is the volume corresponding to that point of interest. Interpretation of this figure in terms of on-road conditions also is provided below. The discussion in Commentary 2 of the implications if the speed-density relationship were in fact perfectly linear, provides useful insights.

[see Commentary 2]

Figure 2.2: Speed-density relationship

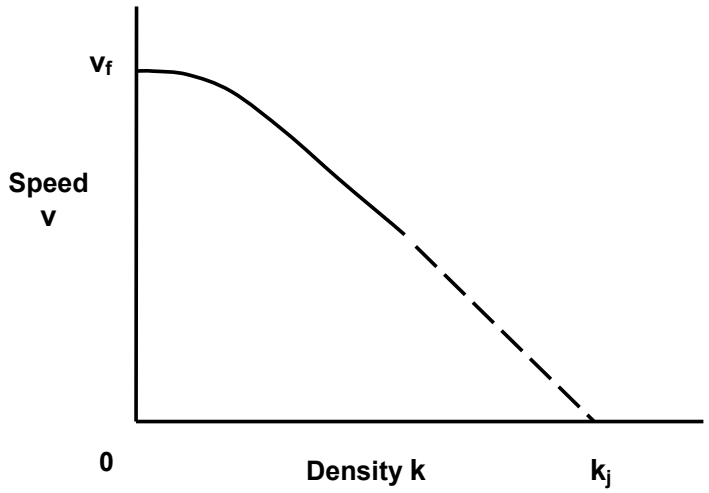
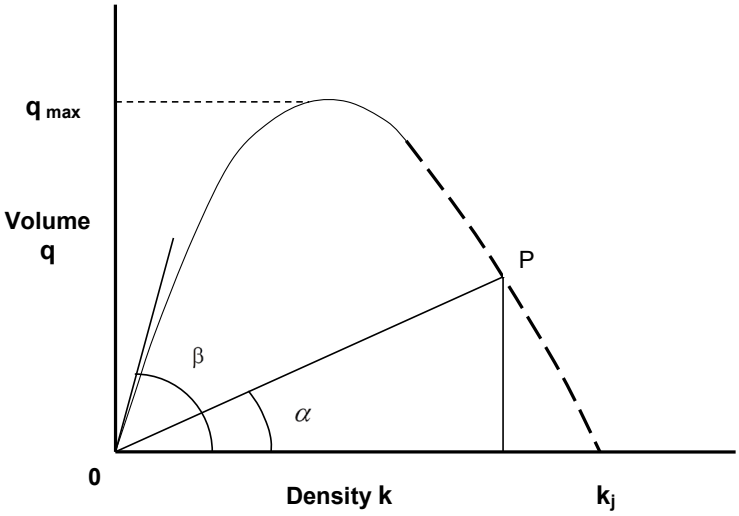


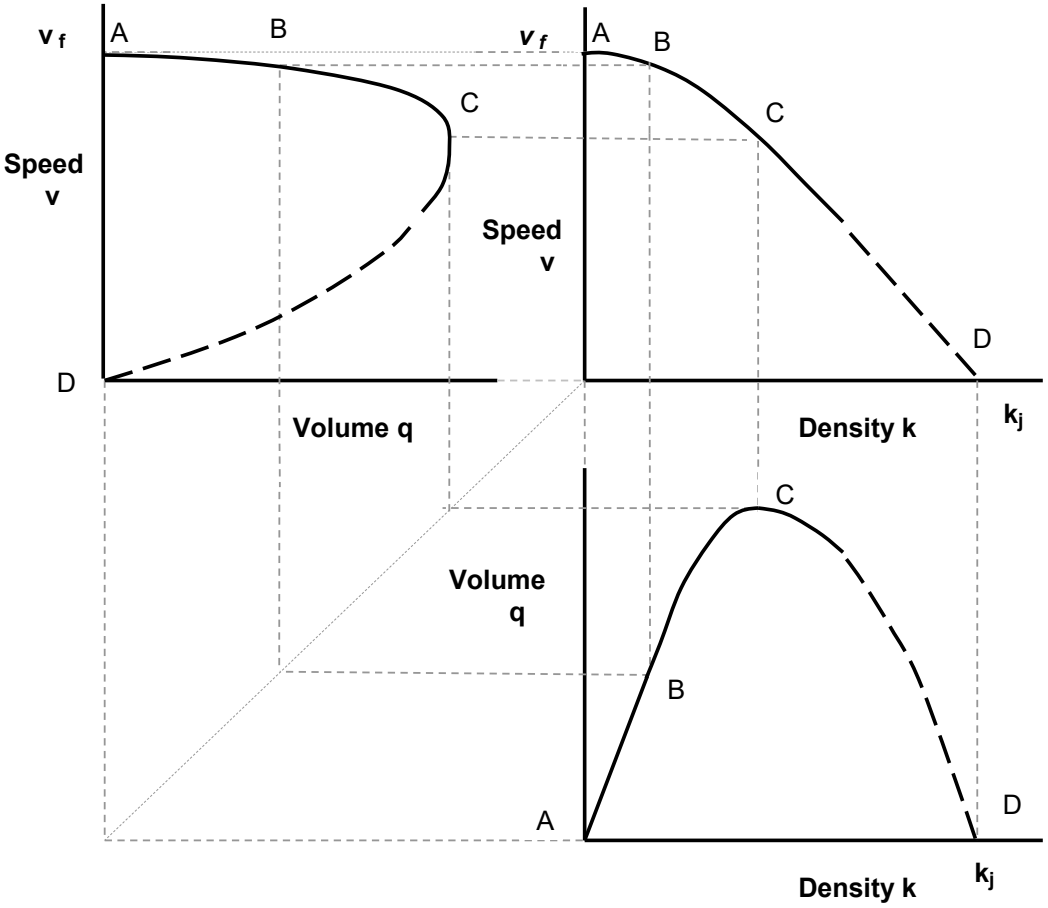
Figure 2.3 shows a typical volume-density relationship for uninterrupted flow as a curve that might be fitted to observed traffic data. As the (space mean) speed in these steady flow conditions is the quotient of volume and average density, the space mean speed corresponding to any point on the curve (such as point P) is represented by the slope, tan α, of a line from the origin to that point. The slope of the curve at the origin, tan β, is thus the mean free speed, V_f.

Figure 2.3: Volume-density relationship



The correspondence between Figure 2.1 to Figure 2.3 and their interpretation in terms of on-road conditions is best discussed by reference to Figure 2.4, which combines the relationships of Figure 2.1 to Figure 2.3 and indicates points representing the same traffic conditions, which are designated by the same letter in each part of the diagram.

Figure 2.4: Speed-volume-density relationships



At the points A in each part of Figure 2.4, density is close to zero, that is, there are very few vehicles on the road. Volume is also close to zero and there are no interactions between vehicles in the traffic stream to prevent drivers from travelling at their desired speeds, the average of which will be the mean free speed, v_f.

From A to the vicinity of B, traffic conditions can be described as 'free flow', in which each vehicle suffers very little restriction due to other traffic in the stream. Such restrictions start to become quite significant as the point B is passed. This could be considered the region of normal flow, in which drivers experience an increasing lack of freedom to manoeuvre (e.g. change lanes, change speed) but traffic nevertheless moves steadily at a reasonable speed, at least until conditions in the vicinity of C are reached.

As the point C is approached, traffic conditions become very unstable and substantial fluctuations in both speed and density can occur with very little change in volume. C is the point of maximum achievable volume and any further increases in density only decrease speed to such an extent that volume also decreases rapidly. Traffic is operating in the 'forced flow' region from C to D, the ultimate condition of which is reached at D, where volume is zero because the traffic is stationary at a maximum density ('bumper-to-bumper' condition) termed the **jam density**.

Thus, a driver would perceive excellent traffic conditions in the region of A, deteriorating gradually from A to B and towards C, and becoming poor to bad around C and from C to D.

2.5. Kinematic Wave Model

An extension of the fundamental relationships is to consider speed, flow and density as functions of time (t) and space (x), and the three parameters are not independent of one another. For example, flow is a function of density k, which is a function of time t. A model that considers the traffic process in time and space is the *kinematic* wave model of Lighthill and Whitham (1955), which is more suitable for high density conditions and therefore has its place in analysing flow breakdowns.

The kinematic model assumes that high density traffic will behave like a continuous fluid (hence also called a *continuum* model). Consider the flow in and out of a short length of road ∂x. The condition of *continuity* requires that if the density of vehicles has increased it must have been due to a difference in the amounts flowing in at one end and out at the other, or

$$\frac{\partial k}{\partial t}+\frac{\partial q}{\partial x}=0 \qquad 2.10$$

where

q is the flow (veh/h)

k is the density (veh/km)

x is distance (km)

t is time (h) to travel a distance of x km

With q as a function of density k, Lighthill and Whitham developed Equation 2.10 further into the LW model as follows:

$$\frac{\partial k}{\partial t}+\frac{\partial q}{\partial k}\frac{\partial k}{\partial x}=0 \qquad 2.11$$

Define below a *wave speed* U that represents the speed of waves carrying continuous changes of vehicle flow in a traffic stream:

$$U=\frac{\partial q}{\partial k} \qquad 2.12$$

then

$$\frac{\partial k}{\partial y}+U\frac{\partial k}{\partial x}=0 \qquad 2.13$$

Because q = v k from Equation 2.3, the wave speed:

$$U=\frac{\partial(vk)}{\partial k}$$

$$=v+k\,\frac{\partial v}{\partial k} \qquad 2.14$$

Because speed decreases with density, $\frac{\partial v}{\partial k}$ is always negative (Figure 2.4) and the wave speed U is therefore always less than the space mean speed v.

The relationship between space mean speed (v) and wave speed (U) are illustrated in the flow-density diagram in Figure 2.5, which also shows the shock wave speed (USW). The following observations can be made (Wohl & Martin 1967):

- At low densities when vehicle-to-vehicle interactions are minimal, $\frac{\partial v}{\partial k}$ is almost zero and the wave speed is similar to the space mean speed. The wave moves forward relative to the road.
- At the maximum flow and critical density, the wave is stationary. At densities higher than the critical density (k_c), the wave moves backward relative to the road.
- The wave speed changes with density according to Equation 2.14 and a traffic stream can have different densities on different sections of a freeway. A section of light traffic could follow a section of high density due to a decrease in lanes, an accident or on-ramp traffic. The wave in the low-density traffic moves forward (relative to the freeway) at a speed faster than the wave in the high density traffic.
- When the two waves meet, a new wave called a *shock wave* will be formed. All three waves move forward for the situation shown in Figure 2.5. The shock wave speed U_{SW} is given by:

$$U_{sw} = \frac{q_2 - q_1}{k_2 - k_1} \qquad 2.15$$

Figure 2.5: The relationship between vehicle speed, wave speed and shock wave speed

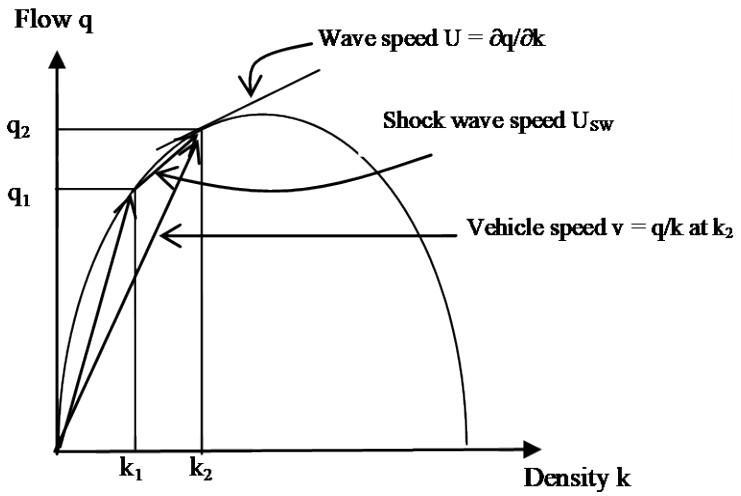
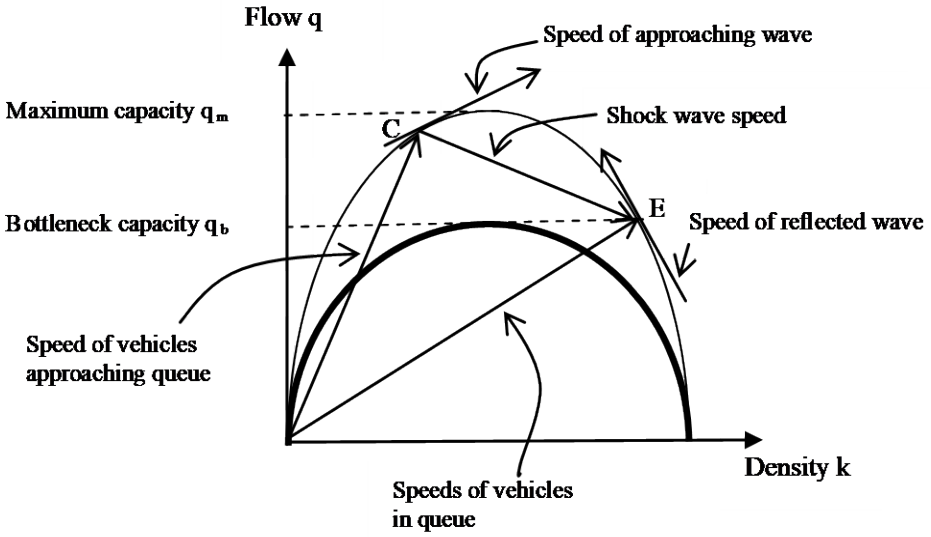


Figure 2.6 illustrates the case for a negative shock wave speed due to capacity decrease at a bottleneck (e.g. lane drop) on a freeway. Two fundamental diagrams are required. The inner diagram represents the characteristics of the bottleneck with capacity q_b less than the approach section. If the approach flow is larger than q_b, a complex queuing situation occurs at the entry to the bottleneck.

The density at the bottleneck entry suddenly increases from the density at C to the density at E in Figure 2.6. The wave speed at E is negative with respect to the freeway and will be reflected from the bottleneck back to the approach section. The reflected wave will meet the oncoming wave corresponding to the slope at C. A shock wave of negative speed relative to the freeway is formed. The effect of the bottleneck will be reflected along the entire approach section if the arrival flow remains constant, with a consequent loss of maintaining capacity flow (q_m).

Edie and Foote (1958) reported how shock waves were generated at an upgrade leading to the Holland Tunnel exit in New York. The shock waves propagated backward towards the tunnel entry with inefficient traffic flow. The solution was to control the entry of vehicles into the tunnel so that the entry flow did not exceed the capacity of the bottleneck section. The vehicles entered in short platoons of about 40 veh every 2 min with a 10 s gap between platoons.

Figure 2.6: Fundamental diagrams of a bottleneck section and the approach section



The kinematic model can be solved using the finite difference (or finite element) method and has continued to be an interesting area of research (see, e.g. Leo & Pretty 1992, Michalopolous 1988, Ngoduy, Hoogendoorn & Van Zuylen 2006, Papageorgiou 1983, Payne 1971). At the University of Queensland, Leo and Pretty were able to model the propagation of congested density upstream in a freeway lane drop situation. They also modelled the platoon movements in a pair of coordinated signals at very small, discrete levels of time (0.5 to 1 s) and space (about 15 m).

The LW model is a first order model with limitations such as (Papageorgiou 1998):

- Assume that vehicle speeds can change instantaneously, i.e. large values of acceleration and deceleration rates are assumed possible at a bottleneck (E in Figure 2.6).
- Predict that the tail-end of a platoon on arterial roads will speed up to catch up with the main platoon when it is more common to observe a dispersed tail-end.

Assume that outflow (q_b) at a freeway bottleneck is best achieved with some congestion at the bottleneck entry. This is equivalent to assuming that the outflow cannot be increased by avoiding mainline congestion, i.e. no control. The reality is that some control of a bottleneck (if possible) can improve throughput.

Second order LW models have been proposed (Daganzo 2006, Papageorgiou, Blosseville & Hadj-Salem 1990, Payne 1971, Schonhof and Helbing 2007) to overcome these limitations.

The use of Kinematic models in understanding freeway flow breakdowns is discussed further in Section 7.5.4.

3. The Stochastic Nature of Traffic Behaviour

3.1. Probabilistic Aspects of Traffic Flow

Traffic behaviour is influenced by a wide range of factors. Each vehicle on the road system is controlled by a driver whose individual decisions, on times of commencement of trips, routes to take, speeds at which to travel and many other things, determine where it is on the road network at any given time and what influence it may have on other road users. Equally, the decisions of others (such as pedestrians wishing to cross a road), the operation of traffic control devices (such as signals), weather and lighting conditions introduce further variation into the traffic situation faced by each road user. Given the variety of these and other similar factors, it is not surprising that probability theory should play a significant role in the description and analysis of traffic flows.

While many aspects of traffic behaviour may be stochastic in nature, they are not necessarily random. A wide variety of different statistical distributions may apply, for example, to the pattern of arrivals of vehicles at a particular location along a road.

If the location is distant from any external factor that influences traffic behaviour, such as a signalised intersection or a toll booth, then vehicle arrivals may well be random and statistical distributions appropriate to random behaviour would be applicable.

On the other hand, if the location at which vehicles are arriving is a short distance downstream from a signalised intersection, the pattern of arrivals would be far from random but would be likely to consist of periods of closely spaced arrivals of vehicles in ‘platoons’, separated by periods of much lighter traffic flow, which would be described by different statistical distributions. A further example is traffic on a freeway entrance ramp at a point downstream from a ramp metering device; such a device is designed to release vehicles at regular intervals, so that a uniform distribution of arrivals would apply at the downstream point.

The purposes of this section are, firstly, to provide information on some of the statistical distributions commonly employed in the development of traffic theory and in traffic analysis and, secondly, to explore their application to traffic headway distributions, that is, the patterns of arrivals of vehicles, pedestrians, cyclists and/or other road users at given points on the road, which is fundamental to many aspects of traffic theory.

3.2. Statistical Distributions in Traffic

Both discrete and continuous probability distributions may be used in describing traffic flows. Discrete distributions are those applicable to situations in which the item of interest can take only integer values, for example, the number of vehicles that will enter a car park through a gate during the next five minutes. Continuous distributions are those in which the item of interest can take both integer and non-integer values, for example, the time a vehicle will spend in the queue waiting to enter the car park via that gate.

In the following sections several distributions of both types are discussed.

3.2.1. The Binomial Distribution

The situation often used to introduce the binomial distribution is that of n independent trials or experiments, each of which can have only one of two possible outcomes, often designated as ‘success’ and ‘failure’. The probabilities of the outcomes known are the same for each trial and, because only the two outcomes are possible, necessarily add to one. A simple example of such a situation is a series of n tosses of a true coin, in which only the two outcomes ‘heads’ and ‘tails’ are possible, and each has a probability of 0.5.

A discrete probability distribution is appropriate to predict the number of ‘successes’ obtained in n trials or the number of ‘heads’ in n tosses of the coin.

The binomial frequency distribution can be written as:

$$b(x) = \binom{n}{x} p^x (1-p)^{n-x} \quad 3.1$$

where

$b(x)$ = the probability of exactly x successes in n trials ($x \leq n$)

p = the probability of a success in any one trial

$\binom{n}{x}$ = the binomial coefficient, equal to the number of different combinations of x items that can be formed from a group of n items $\frac{n!}{x!(n-x)!}$

The mean of the binomial frequency distribution, or the expected number of 'successes' in n trials if p is the probability of a 'success' in one trial, is:

$$E(x) = np \quad 3.2$$

And the variance of $_x$ is:

$$\sigma^2(x) = np(1-p) \quad 3.3$$

In traffic applications, the cumulative form of the binomial distribution is often useful. This can be written as:

$$B(X) = \sum_{x=0}^{X} b(x) \quad 3.4$$

where

$B(x)$ the probability of X or less successes in n trials

$b(x)$ is as defined in Equation 3.1

Example application

The situation of vehicles entering a car park through a particular gate (say Gate A) provides an example of how the binomial distribution might be applied in a traffic context. Assume that, at a certain time of day, the car park is being accessed by 300 veh/h and that 35% of these, on average, use Gate A. Assume also that the aim is to determine the probability of more than six vehicles using Gate A over a two minute period.

On average, 10 vehicles will access the car park over a two-minute period. These can be considered as 10 independent trials, in each of which there is a probability of 0.35 that the vehicle will use Gate A and a probability of 0.65 that it will use another gate. The probability that more than six of the 10 vehicles will use Gate A is 1 – B(6) where B(6) is calculated using Equation 3.4 with X = 6 and each b(x). x = 1,..., 6, calculated from Equation 3.1, with n = 10 and p = 0.35. It is found that there is a probability of 0.0260, or 2.6%, that more than six vehicles will use Gate A.

Note that the same answer can be obtained as the probability of three or less of the 10 vehicles using another gate, rather than Gate A, that is, B(3), calculated using Equation 3.4 with X = 3 and each b(x), x = 1,..., 3, being calculated from Equation 3.1, with n = 10 and p = 0.65.

3.2.2. The Poisson Distribution

Mathematically, the Poisson distribution is the limit of the binomial distribution as n approaches infinity and p approaches zero while the product n.p remains constant. The Poisson frequency distribution is:

$$p(x) = \frac{e^{-m} m^{x}}{x!} \qquad 3.5$$

where

p(x) = the probability of x occurrences of an event in a situation for which the expected number of occurrences is m

e = the base of Naperian logarithms

The Poisson distribution is discrete, in that x can take only integer values. However, m is not restricted to integer values.

As would be expected, the mean, or expected value, of the Poisson distribution is:

$$E(x) = m \qquad 3.6$$

Also, in keeping with its nature as the limit of the binomial distribution as p approaches zero and by comparison with Equation 3.3, the variance of the Poisson distribution is:

$$\sigma^2(x) = m \qquad 3.7$$

As is the case with the binomial distribution, the cumulative Poisson distribution is often useful in traffic theory and analysis. This cumulative form is:

$$P(X) = \sum_{x=0}^{X} p(x) \qquad 3.8$$

where

P(X) = the probability of x or less occurrences of an event where the expected number of occurrences is m

p(x) = is as defined in Equation 3.5

The most commonly quoted example of an application of the Poisson distribution in a road traffic context is the estimation of the probability that a certain number of vehicles will pass a particular point on a road during a given time period, given that vehicles arrive randomly at a known average rate. For example, if the average volume is 720 veh/h, the aim may be to estimate the probabilities that (a) no vehicles will pass and (b) three or more vehicles will pass during the next 10 s.

At a flow rate of 720 veh/h, the average number of vehicles passing in any 10 s interval is 2.0, which is thus the appropriate value for m in Equation 3.5. Applying Equation 3.5 for different values of x:

$$\mathrm{p}(0) = \mathrm{e}^{-2.0}(2.0)^0/0! = 0.1353$$

$$\mathrm{p}(1) = \mathrm{e}^{-2.0}(2.0)^1/1! = 0.2707$$

$$\mathrm{p}(2) = \mathrm{e}^{-2.0}(2.0)^2/2! = 0.2707$$

The first of these values indicates that the answer to question (a), the probability that no vehicles will pass during the next 10 s, is 13.53%.

Applying Equation 3.8 with X = 2, indicates that the probability that two or less vehicles will pass during the next 10 s is the sum of the three values above, i.e., 0.6767 or 67.67%. Thus the answer to question (b), the probability that three or more vehicles will pass, must be (100–67.67)% = 32.33%.

A different example of an application of the Poisson distribution in a road traffic context is in the analysis of crashes on a section of road that has an average of 19.5 crashes per year. Assume that there is negligible seasonal variation in crash likelihood and that, for crash response planning purposes, the probability of more than one crash occurring on this section of road in any one week of the year needs to be estimated.

The average number of crashes in a week is 19.5 / 52.18 = 0.3737. Applying Equation 3.5 with m = 0.3737, gives p(0) = 0.6882 and p(1) = 0.2572, so that, p(1) = 0.9454, by Equation 3.8. Therefore, the probability of more than one crash occurring in any one week of the year is estimated as 1–p(1) = 0.0546, or 5.46%.

3.2.3. Negative Binomial Distribution

The negative binomial distribution, like the binomial distribution, is applicable to situations in which there are two possible outcomes in each trial of a series of trials. These outcomes can be called ‘success’ and ‘failure’, with probabilities of occurrence in any one trial of p and 1–p respectively. The negative binomial distribution gives the probability that the k^{th} ‘success’ will occur in the n^{th} trial. This may be expressed as:

$$\Pr(\mathrm{n};\mathrm{k},\mathrm{p}) = \binom{\mathrm{n}-1}{\mathrm{k}-1}\mathrm{p}^{\mathrm{k}}(1-\mathrm{p})^{\mathrm{n}-\mathrm{k}} \qquad \mathrm{n} = \mathrm{k}, \mathrm{k}+1, \mathrm{k}+2, \ldots \qquad 3.9$$

3.2.4. Geometric Distribution

The geometric distribution is a special case of the negative binomial distribution. If x is the sequence number of a trial, the geometric distribution gives the probability that the first ‘success’ will occur in the n^{th} trial after a sequence of n–1 ‘failures’. This may be expressed as:

$$\Pr(\mathrm{n};1,\mathrm{p}) = \mathrm{p}(1-\mathrm{p})^{\mathrm{n}-1} \qquad \mathrm{n} = 1,2,3, \ldots \qquad 3.10$$

A simple example of an application of the geometric distribution is determining the probability that it will take exactly four tosses of a coin before the first ‘head’ comes up.

In a traffic context, an example might be the probability that the next three vehicles to access the car park considered in Section 3.2.1 will each enter by Gate A and the 4th will enter by another gate. This is evaluated using Equation 3.10 with p = 0.65 (i.e. 'failure' ≡ entry by Gate A), to give

$$\Pr(\mathrm{x} = 4) = (0.65)(0.35)^3 = 0.0279$$

The geometric distribution also has application in relation to queue lengths, as seen in Section 4.2.

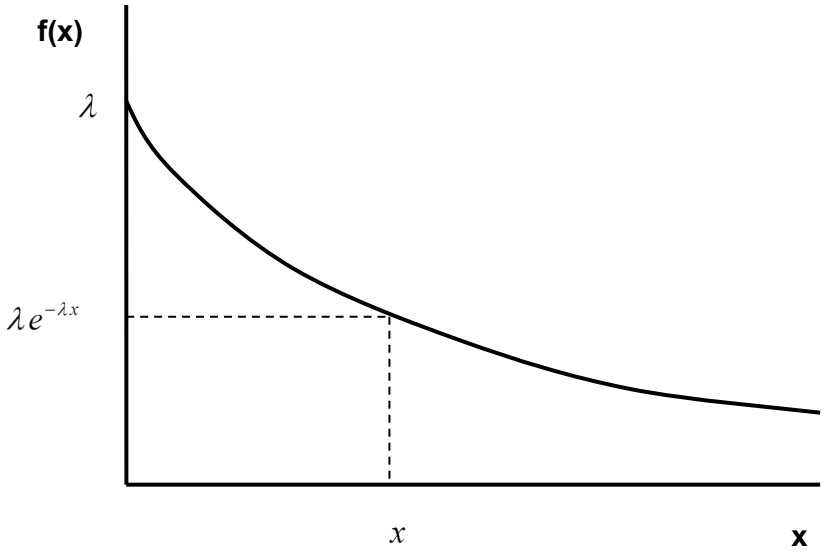
3.2.5. Negative Exponential Distribution

The negative exponential frequency distribution is illustrated in Figure 3.1. It is a continuous distribution and has the form:

$$\mathrm{f(x)} = \lambda \mathrm{e}^{-\lambda \mathrm{x}} \qquad 3.11$$

where λ is a known parameter.

Figure 3.1: Negative exponential frequency distribution



The mean of the negative exponential distribution, or the expected value of x is:

$$\mathrm{E(x)} = \frac{1}{\lambda} \qquad 3.12$$

and the variance is:

$$\sigma^2(\mathrm{x}) = \frac{1}{\lambda^2} \qquad 3.13$$

The negative exponential is the simplest relationship used to describe the distribution of headways in randomly arriving traffic, as is discussed in Section 3.3. The displaced negative exponential distribution, a variation allowing for a minimum possible headway in a lane of traffic, is also discussed in Section 3.3, as are other types of headway distributions.

3.2.6. Other Distributions

The Borel-Tanner distribution is a discrete distribution with particular application to consideration of bunch sizes in composite models of traffic flow, as discussed in Section 3.3.4.

A number of other discrete probability distributions are applicable to aspects of traffic analysis. For further information, the reader is referred to Homburger (1982) or texts such as Johnson, Miller and Freund (2011) or Walpole et al. (2011).

Other continuous distributions relevant to traffic include the normal distribution, Pearson distributions and the Erlang distribution. Some of these are briefly mentioned in Section 3.3, and references are provided.

3.3. Traffic Headway Distributions

Prior to discussing different models for headway distributions, it is observed that, conventionally, a headway between two consecutive vehicles is associated with the trailing vehicle. In other words, the headway for any vehicle is its time separation from the vehicle immediately ahead of it, rather than from the vehicle immediately behind (see, for example, Akcelik et al. 1999). Of course, this time separation is between the times at which the same point on each vehicle (typically the front) passes a given location on the road.

3.3.1. Random Arrivals – Negative Exponential Headways

The distribution of headways in a traffic stream is modelled according to the assumed pattern of vehicle arrivals, which is influenced by the traffic volume, the lane configuration of the road or carriageway involved, external influences on the traffic flow and other factors. One of the simpler and more widely applicable assumptions is that of random arrival of vehicles, for which the negative exponential headway distribution applies, as shown below.

Consider a one-way traffic stream with an average volume of q veh/s and random vehicle arrivals. The average number of vehicles arriving during a time interval of t seconds is qt and the probability of x vehicles arriving during any particular t second interval is given by the Poisson distribution (see Equation 3.5) as:

$$p(x) = \frac{e^{-qt}(qt)^x}{x!} \qquad 3.14$$

The probability of a headway greater than or equal to t is the probability of zero arrivals during a t second interval, that is:

$$Pr(h \geq t) = \frac{e^{-qt}(qt)^0}{0!} = e^{-qt} \qquad 3.15$$

And

$$Pr(h < t) = 1 - e^{-qt} \qquad 3.16$$

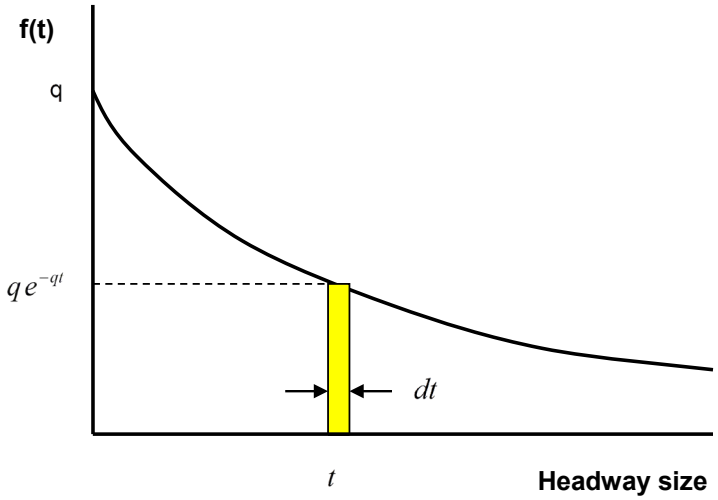
Equation 3.16 has the form of a cumulative probability distribution. The associated frequency distribution is found by differentiation to be:

$$f(t) = qe^{-qt} \qquad 3.17$$

This is the negative exponential distribution (see Equation 3.11, Section 3.2.5) with the parameter λ equal to q. Graphically, it has the same form as Figure 3.1, as is shown in Figure 3.2.

The mean of the negative exponential headway distribution is, of course, 1/q and the variance of the distribution is $1/q^2$, consistent with Equations 3.12 and 3.13.

Figure 3.2: Negative exponential headway distribution



The proportion of all headways that fall within the small interval between a given duration t and duration t + dt is the shaded area in Figure 3.2, that is:

$$\Pr(t \le h < t + dt) = qe^{-qt}.dt \qquad 3.18$$

Therefore, over a significant period H, containing qH headways, the number of headways of duration between t and t + dt is:

$$N(t \le h < t + dt) = qH.\,qe^{-qt}.dt \qquad 3.19$$

and, as dt is infinitesimally small, the total time spent in such headways is:

$$T(t \le h < t + dt) = t.qH.\,qe^{-qt}.dt \qquad 3.20$$

Hence, the total time spent in headways greater than or equal to t is given by:

$$T(h \ge t) = qH\int_t^\infty t.\,qe^{-qt}.dt$$

$$= H.\,e^{-qt}(1 + qt) \qquad 3.21$$

Substitution of t = 0 in Equation 3.21 produces the expected result that the total time spent in all headways is H.

It follows from Equation 3.21 that the **proportion** of time spent in headways greater than or equal to t is:

$$T(h \ge t)/_H = e^{-qt}(1 + qt) \qquad 3.22$$

The average duration of headways greater than or equal to t is the total time spent in such headways divided by the number of such headways, that is:

$$h_{av}(h \geq t) = \frac{H.\, qe^{-qt}(1+qt)}{qH.\, e^{-qt}}$$

$$= \frac{1}{q} + t \qquad 3.23$$

Substitution of t = 0 in Equation 3.23 produces the expected result that the average duration of all headways is 1/q.

In a derivation similar to that above, it is easily shown that the average duration of headways less than t is:

$$h_{av}(h < t) = \frac{1}{q} - \frac{te^{-qt}}{1-e^{-qt}} \qquad 3.24$$

Because the negative exponential headway distribution is derived from the assumption of random arrivals of vehicles, its applicability is restricted to lighter traffic flows, where there are few vehicle interactions to influence the travel behaviour of any individual vehicle. Further, it applies only to uninterrupted flow, that is, traffic flow that has not been affected by significant external influences, such as the presence of a signalised intersection shortly upstream, which can greatly affect arrival patterns. Finally, the negative exponential distribution allows headways right down to zero duration and therefore cannot correctly represent any situation in which the minimum feasible headway is greater than zero, such as traffic flow in a single lane.

Other types of headway distributions have been proposed to address these and similar issues and a number of these are briefly discussed in the following sections.

3.3.2. Equal Headways

The simplest headway distribution is based on the assumption of equal headways between successive vehicles in the traffic stream.

This distribution might be assumed, for example, a short distance downstream of a metering device on a single-lane entrance ramp to a freeway. The distance would be limited because the different acceleration behaviour of different vehicles would quickly introduce variation into the headways. The restriction to a single lane would be necessary because a metering device with more than one lane would release vehicles side-by-side, that is, separated by headways close to zero; combined with the headways equal to the release interval of the metering device, these would produce an overall headway distribution consisting of two sets of uniform headways.

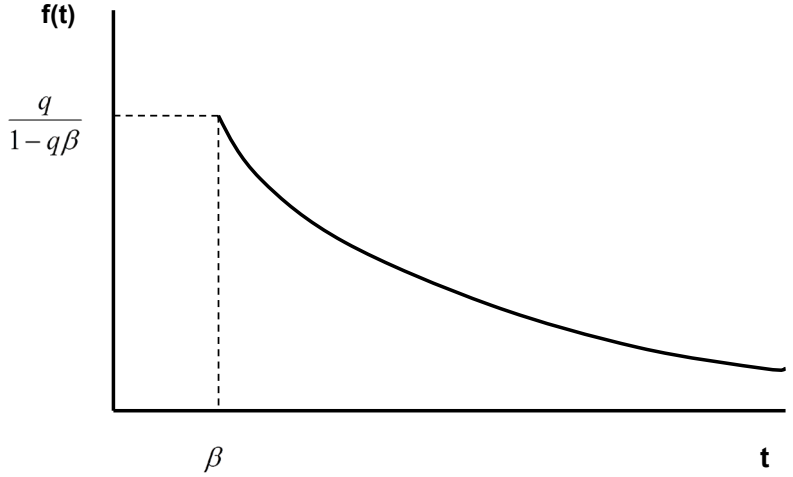
For a traffic stream with volume q, the variance of the basic uniform distribution is zero and the mean headway is the uniform size of each headway, that is:

$$E(h) = h = \frac{1}{q} \qquad 3.25$$

3.3.3. The Displaced Negative Exponential Distribution

Where vehicle arrivals are essentially random, but the minimum possible headway is greater than zero (which would apply, for example, to single lane flow with no overtaking), the displaced negative exponential distribution may be an appropriate representation of headways. As illustrated by Figure 3.3, the frequency distribution for the displaced negative exponential has the same shape as the negative exponential but is displaced to the right by an interval β, equal to the minimum possible headway.

Figure 3.3: Displaced negative exponential headway distribution



Mathematical manipulation of the basic negative exponential form to provide a unit area under the graph from $t = \beta$ to $t = \infty$ produces the displaced negative exponential distribution:

$$f(t) = \frac{q}{1-q\beta} e^{\frac{-q(t-\beta)}{1-q\beta}} \quad \text{for } t \geq \beta \qquad 3.26$$

Derivations paralleling those used above for the standard negative exponential lead to the following results.

The probability of a headway greater than or equal to t, for $t \geq \beta$, is:

$$Pr(h \geq t) = e^{\frac{-q(t-\beta)}{1-q\beta}} \quad \text{for } t \geq \beta \qquad 3.27$$

The average duration of headways greater than or equal to t, for $t \geq \beta$, is:

$$h_{av}(h \geq t) = \frac{1}{q} + t - \beta \quad \text{for } t \geq \beta \qquad 3.28$$

Note that Equation 3.28 gives an average headway of 1/q for all headways not less than β.

3.3.4. Composite Headway Distribution Models

In composite (or mixed) models of traffic headway distributions, the distinction is made between **free-flowing** vehicles and **restrained** vehicles, the latter being vehicles in the traffic stream forced (e.g. by limited lane changing or overtaking opportunities) to follow the vehicle immediately ahead. In such cases, a traffic **bunch** consists of either a single (free flowing) vehicle or a leading vehicle followed by one or more restrained vehicles.

Establishment of a composite headway distribution model requires specification of a enough the following components (among which there is some overlap):

- the proportion of restrained vehicles in the traffic stream (the remainder being free flowing)
- the distribution of bunch sizes
- the distribution of headways for restrained vehicles
- the distribution of headways for free-flowing vehicles (which, given that a single free flowing vehicle is defined as a bunch of size 1, is equivalent to the distribution of inter-bunch headways).

The double exponential model and the Borel-Tanner model are briefly discussed below, to illustrate how some of these components may be specified.

Double exponential distribution

The double exponential model combines a displaced negative exponential distribution for the headways of restrained vehicles with a standard negative exponential distribution for the headways of free-flowing vehicles. The double exponential headway distribution described here should not be confused with the Gumbel distribution, which sometimes is also referred to as the 'double exponential' distribution.

Because of the separation of the traffic stream into restrained and free-flowing sub-streams, it is simpler to state the model in terms of the average headway for each sub-stream, rather than in terms of the sub-stream volume. As volume q is the inverse of average headway $\overline{h}$ in any traffic stream, this simply means substituting $1/\overline{h}$ for q in relevant equations from Sections 3.3.1 and Section 3.3.3.

Let the proportion of restrained vehicles in the total traffic stream be denoted by θ and the minimum and average headways for such vehicles by β and $\overline{h_r}$ respectively. Then the displaced negative exponential distribution of restrained vehicle headways is such that:

$$\Pr(\mathrm{h} \geq \mathrm{t}) = \theta . \mathrm{e}^{\frac{-(\mathrm{t}-\beta)}{\overline{\mathrm{h_r}}-\beta}} \text{ for } \mathrm{t} \geq \beta$$

and

$$\Pr(\mathrm{h} \geq \mathrm{t}) = \theta \text{ for } \mathrm{t} \leq \beta \qquad 3.29$$

The proportion of free flowing vehicles in the total traffic stream is $(1-\theta)$ and, if their average headway is denoted by $\overline{\mathrm{h_f}}$, their standard negative exponential headway distribution is such that:

$$\Pr(\mathrm{h} \geq \mathrm{t}) = (1-\theta) . \mathrm{e}^{\frac{-\mathrm{t}}{\overline{\mathrm{h_f}}}} \text{ for } \mathrm{t} \geq 0 \qquad 3.30$$

Equations 3.29 and 3.30 together make up the double exponential headway distribution, one example of a distribution that may be applicable when bunching is likely to occur. For further details, see Salter and Hounsell (1996) and early investigations by Schuhl (1955) and Kell (1962).

A less complex headway model that may be used in the presence of bunching is the simple dichotomised distribution, in which within-bunch headways are each equal to a specified minimum value and inter-bunch (or free flowing vehicle) headways are greater than this value.

Borel-Tanner distribution

The Borel-Tanner distribution (Tanner 1961) gives a good representation of the distribution of bunch sizes in traffic on two-lane, two-way roads. It evaluates the probability of observing a bunch of $_{n}$ vehicles where the average bunch size is $_{m}$ vehicles as:

$$P_n = (n.\theta.e^{-\theta})^{n-1}.\frac{e^{-\theta}}{n!} \quad n = 0, 1, 2, \ldots \qquad 3.31$$

where $\theta = \frac{m-1}{m}$ = the proportion of restrained vehicles in the traffic stream.

Headway distributions for following vehicles within a bunch and for free-flowing vehicles are also required. For following vehicles, a distribution giving constant or close-to-constant headways (i.e. with small variance about an average headway) or a displaced negative exponential distribution might be assumed. For free-flowing vehicles, a negative exponential or some other distribution describing random or semi-random arrivals may be considered appropriate.

3.3.5. Other Headway Distributions

A variety of other headway distributions, both single and composite, have been developed to represent observed behaviour under different conditions of traffic flow.

Among the single distributions are the Pearson Type I (or beta) distribution and the Pearson Type III (or gamma) distribution (Ashton 1966). A particular case of the latter is the Erlang distribution in which the principal parameter is restricted to integer values. This has been shown to fit well with field observations of headways in traffic with essentially random arrivals and limited opportunities for overtaking, particularly in the representation of smaller headways. The generalised Poisson distribution is related to the gamma distribution in the same way that the Poisson distribution is related to the negative exponential distribution.

Among other composite models is Miller's travelling queue model (Miller 1961), in which randomly placed vehicles are moved backward in time, where necessary, to maintain a constant minimum headway. The model leads to a bunch size distribution that has shown close agreement with observed multi-lane flows.

For further information on headway distributions, the reader is referred to the references already mentioned and to general coverage of the topic in Drew (1968), Gipps (1984), Hoban (1984a and b), May (1990) and Salter and Hounsell (1996).

3.3.6. Matching Headway Distribution Type to Observed Behaviour

In many cases, a key decision for the traffic analyst is the selection of the type of headway distribution that is either:

a. most likely to correspond with the traffic situation under consideration

b. likely to best match a set of headways that has been observed in field measurements.

For the category (a) decisions, some guidance has been provided throughout the sections preceding of Section 3.3 and this is summarised in Table 3.1.

Table 3.1: Areas of application of different types of headway distribution

Type of headway distribution	Area(s) of application
Equal headways	A short distance downstream of a traffic metering device
Negative exponential	Wherever random arrivals of vehicles may occur, which usually is in situations of uninterrupted flow with low to medium traffic volumes and two or more lanes in the same direction of flow
Displaced negative exponential	Uninterrupted flow at medium to low traffic volumes, but with conditions (e.g. single lane flow) that place a lower limit on the possible size of headways
Double exponential	Headways in uninterrupted flow at medium to high traffic volumes with restricted overtaking opportunities, where bunching of traffic is likely to occur May also be applicable to platooned traffic in an interrupted flow situation (e.g. downstream from a signalised intersection)
Borel-Tanner	Describes bunch-size distributions for uninterrupted flow at medium to high traffic volumes with restricted overtaking opportunities, where bunching of traffic can occur
Miller's travelling queue model	Models both headway distributions and bunch size distributions for bunched traffic in an uninterrupted flow situation with limited overtaking opportunities or for platooned traffic downstream of a traffic interruption
Erlang	Uninterrupted flow with low to medium volumes (conditions for random arrivals) but with limited opportunities for overtaking

Category (b) decisions arise when field measurements of headways are available. In such cases, the characteristics of the field data can be examined analytically and/or graphically to identify types of theoretical distributions that most closely match the data. For any type of distribution so identified, further analysis can then be undertaken to evaluate the distribution parameters that provide the best fit with the observed data.

If an identified distribution is one of the standard types considered in the treatments of queuing theory and gap acceptance theory in Sections 4 to 6, the formulae presented in those sections can be applied directly to the analysis of traffic situations.

If, on the other hand, the identified distribution of headways is not one of those standard types, it will be necessary for the analyst to develop formulae similar to those in Sections 4 to 6, but based on the identified distribution. This can be done (perhaps with the aid of suitable mathematical or statistical advice) by following the methods employed in the derivations presented in Sections 4 to 6 and in their associated commentaries. Such a process is illustrated by Commentary 10, which derives gap acceptance formulae, using the same approaches as used in Commentaries 5 and 7 for negative exponential headways in major road traffic, but adapts the derivations to the case of a displaced negative exponential distribution of major road headways.

[see Commentary 5]

[see Commentary 6]

[see Commentary 7]

[see Commentary 10]

4. Queuing

4.1. Introduction and Definitions

In many traffic situations, particularly at intersections, conflicts between different traffic streams and/or fluctuations in flow can result in the formation of queues of vehicles. Queuing theory provides a way of analysing queue behaviour and predicting its consequences, including queue lengths and queuing delays.

The complete specification of a queuing system requires the values of the following five input characteristics:

- the distribution of arrivals, including the average arrival rate and the type of distribution, e.g. regular, random, Erlang, etc.
- whether the input source (i.e. the pool from which arrivals are drawn) is finite or infinite
- the queue discipline, i.e. the means of deciding the order in which queue members obtain service, which may be first-come-first-served, random, some priority system, etc.
- the channel configuration, which includes the number of separate queues, the number of service positions and whether queue members are served singly, in parallel or in series
- the distribution of service times for each service point, including the average service rate and the type of distribution.

Section 4.2, provides a brief outline of graphical representation and analysis of queuing situations, illustrated by a deterministic example in which both arrival and service behaviour are known with certainty.

Subsequent sections consider an elementary stochastic queuing system specification – that of a single channel, single server system with random arrivals, random service times and a first-come-first-served discipline. This is usually denoted as an [M/M/1] queuing system, the first M indicating random (Poisson) inputs, the second M indicating random (negative exponential) service and the 1 indicating the single-channel queue. For more complex systems, readers should refer to texts such as Cox and Smith (1961) and Drew (1968).

Two further, behavioural characteristics of a queuing system also require definition.

- **The state of a queuing system**. A queuing system is in state n if it contains exactly n queue members, including those in line and those in service.
- **Utilisation factor**. The utilisation factor, ρ, is the ratio of the average arrival rate, r, to the average service rate, s, that is:

$$\rho = \frac{r}{s} \qquad 4.1$$

(In the technical literature on queuing, some authors refer to ρ as the intensity of the queuing system.)

If ρ is less than one, queue length may vary over time due to random fluctuations, but the queuing system is stable and it is possible to calculate a time-independent probability of the queue being in a particular state. If ρ is greater than or equal to one, however, the queue length will increase with time.

As a final introductory comment, it is observed that analysis of queuing behaviour may involve either **deterministic** methods, in which the arrival time and service time for each unit in the queue are considered to be known with certainty, or **probabilistic** methods, which are appropriate when either or both of arrival times and service times vary stochastically.

4.2. Graphical Representation of Queues

Appreciation of traffic queuing behaviour often is assisted by a graphical representation of the situation under consideration. Such graphical representation is particularly well suited to deterministic queuing situations, in which both the vehicle arrival and the queue service patterns are known with certainty. As an example, consider the situation represented in Figure 4.1.

Figure 4.1: Graphical representation of queuing on a signalised intersection approach

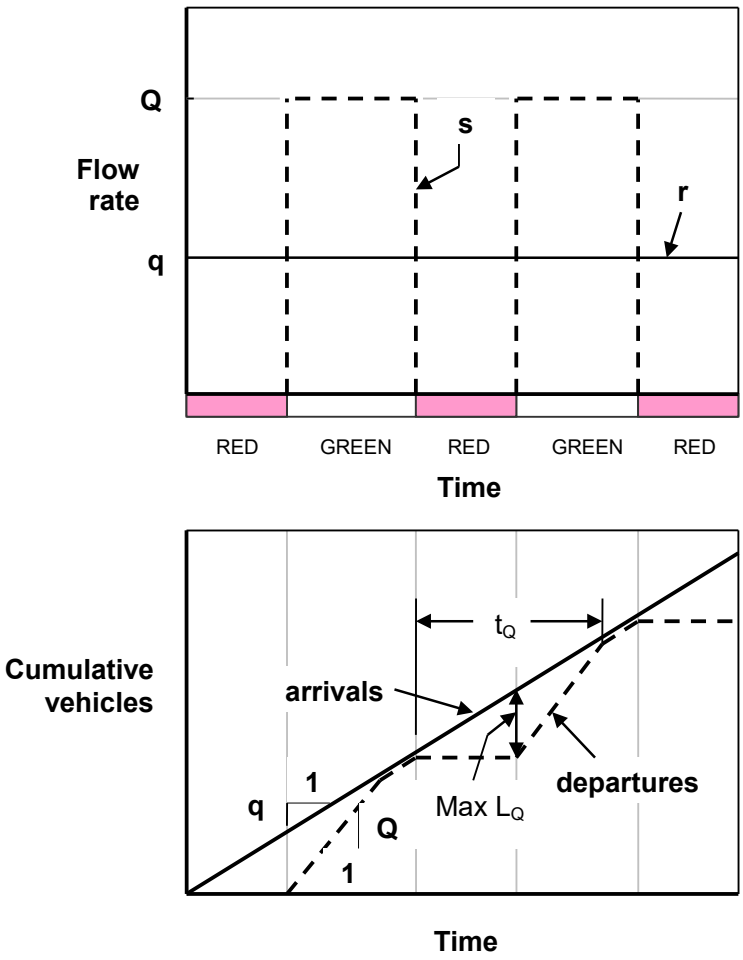


Figure 4.1 relates to traffic arriving on an approach to a signalised intersection. The upper graph plots flow rates against time and indicates an arrival rate, r, of vehicles at the tail of the queue at the intersection stop line that is constant, at q vehicles per unit time. It also shows that the 'service' provided to these vehicles (their passage through the intersection) can occur at a rate, s, of up to Q vehicles per unit time (the saturation flow rate) during the effective green time for this approach, but is zero during the effective red time.

The lower graph in Figure 4.1 plots cumulative vehicles arriving and departing on this approach against time. The solid line representing arrivals has slope q, equal to the arrival rate. The broken line representing vehicle departures has three different slopes, each equal to the corresponding rate of departures, as follows:

- during the effective red time for the approach, the slope is zero, that is, the departures line is horizontal
- from the beginning of the effective green, for as long as there is a queue at the stop line of the approach, the slope of the departures line is Q, the saturation flow rate
- toward the end of the effective green time, when the queue has dissipated, vehicles pass through the intersection at the same rate at which they arrive, so that the slope of the departures line is equal to the arrival rate, q.

In the lower graph of Figure 4.1 the vertical separation of the arrival and departure lines at any time represents the queue length at that time. For example, the maximum queue length, L_Q, is the vertical distance between the arrival and departure lines at the end of the effective red time.

Horizontal distances between the arrival and departure lines in the lower graph of Figure 4.1 represent the delay experienced by individual vehicles. For example, toward the end of the effective green time, when the arrival and departure lines are coincident, arriving vehicles experience no delay. The vehicle that arrives at the start of the effective red time experiences the maximum delay before departing at the start of the effective green time. The time for which a queue exists at the stop line (the queue duration) can be read from the graph as t_Q.

Figure 4.1 presents the simplest possible example, involving a constant arrival rate and service behaviour that is repeated exactly in every cycle of the traffic signals, but it illustrates how graphical methods can be applied to any deterministic queuing situation. Further development of graphical approaches can be found in May (1990).

4.3. Dynamic and Steady State Queuing

As noted in Section 4.1, the key characteristics that determine the performance of a queuing system are the average arrival rate, r, the average service rate, s, and their ratio, $\rho = r/s$, known as the utilisation factor (or, sometimes, the **intensity**) of the queuing system.

Typically, instantaneous arrival and service rates vary probabilistically from one moment to the next, but each of the **average** rates of arrival and service may be considered either to remain constant or to vary in some way, over an analysis period.

An example of variation of average rates is that over a two-hour peak period, the arrival rate of vehicles on a minor road at an unsignalised intersection is likely to increase from a low, off-peak level, build to a maximum peak rate, then steadily decrease toward the end of the peak. Over the same period, a similar pattern of growth and decay in the major road traffic would influence the number of acceptable gaps for minor road vehicles to enter the intersection, which, in turn, would first decrease the intersection's effective service rate for minor traffic to a minimum value, then steadily increase it again towards the end of the peak.

The **behaviour** of a queuing system, as reflected in queue lengths and delays, may vary over an analysis period for one of two reasons:

1. Over the period, there is variation in the average arrival rate and/or the average service rate.
2. Both rates remain constant over the analysis period but the average arrival rate equals or exceeds the average service rate (i.e. the utilisation factor, ρ, is greater than or equal to one).

Case (1) is what is normally termed **dynamic** queuing and applies to situations such as the growth and decay of traffic over a peak period, an example of which was discussed earlier in this subsection. In such situations, the analyst may be interested in the maximum queue lengths and delays that occur, given that vehicle arrival rates may approach or even exceed service rates over different parts of the peak.

Analysis of a dynamic queuing situation may assume deterministic or probabilistic behaviour but, in either case, the analysis period is normally divided into a succession of time intervals. Over each interval, the average arrival rate and the average service rate (and hence also the utilisation factor) can be considered to be constant, though they may vary from interval to interval. The analysis typically assumes a state of the system (as indicated by characteristics such as queue length) at the start of the first interval, then takes the state at the start of each succeeding interval to be the state at the end of the one before it.

If the queuing behaviour can be considered to be deterministic, the analysis process is relatively straightforward – for example, Figure 4.1 might represent the deterministic analysis of one interval in the examination of a dynamic queuing situation over a longer analysis period.

If the queuing behaviour is considered probabilistic, the analysis process can be much more complex, involving the application of probability theory to the analysis of each interval and producing a distribution of possible system states (rather than a single, known state) at each point of change from one interval to the next. Fortunately, however, approximation methods, such as the coordinate transformation method of Kimber and Hollis (1979), have been developed to allow the estimation of average queue lengths and average delays without the need for complex probabilistic calculations.

Case (2) above is dynamic (time dependent), even though average arrival and service rates are constant over time, because a utilisation factor greater than or equal to one results in queue lengths and delays growing steadily with time. This situation clearly cannot be sustained over a long period but could apply during a limited interval within a peak period.

Where average arrival and service rates each remain constant over an analysis period and, in addition, the utilisation factor is less than one, the queuing situation is said to be **steady state**. This means that average queue lengths and average delays will be constant over time and, in the case of probabilistic behaviour, it will be possible to derive probability distributions, which also will not vary with time, for aspects of queuing behaviour such as queue lengths and delays.

For a wide range of traffic engineering applications, analysis of steady state queues with randomly distributed arrivals and service times provides suitable guidance for road design and traffic management decisions. Section 4.4 addresses this type of analysis.

4.4. Steady State Queues with Random Arrivals and Service

4.4.1. Queue Lengths

This subsection provides the key formulae related to queue lengths in an [M/M/1] queuing system. The derivations of these formulae are provided in Commentary 3.

[see Commentary 3]

The key formulae are as follows.

The probability of the queue being empty – that is, having no units in service and no units waiting in a queue to be serviced – is:

$$P_0 = 1 - \rho \qquad 4.2$$

The probability that there are n units in the system, $n \geq 0$, including the unit in service (if any), is:

$$P_n = (1 - \rho)\rho^n \qquad 4.3$$

The expected number in the system is:

$$E(n) = \frac{\rho}{1 - \rho} = \frac{r}{s - r} \qquad 4.4$$

The probability of there being more than N items in the system is:

$$Pr(n > N) = \rho^{N+1} \qquad 4.5$$

The mean queue length, excluding the unit being serviced, is:

$$E(m) = \frac{\rho^2}{1-\rho} = \frac{\rho}{1-\rho} - \rho \quad 4.6$$

Thus,

$$E(m) = E(n).\rho = E(n) - \rho \quad 4.7$$

Given that $\rho < 1$, $E(m)$ is not (as might be expected) equal to E(n) – 1. This is because there is a finite probability that the system is empty, in which case n = m = 0.

Finally, the variance of the number of units in the system is:

$$\sigma^2(n) = \sum_{n=0}^{\infty} n^2 P_n - [E(n)]^2 = \frac{\rho}{(1-\rho)^2} \quad 4.8$$

4.4.2. Waiting Times in Queues

This subsection provides the key formulae related to delays experienced by units in an [M/M/1] queuing system. The derivations of these formulae are provided in Commentary 4.

[see Commentary 4]

The time spent by an individual in a queuing situation is made up of the time spent waiting in the queue until service is commenced (denoted by 'w') and the time spent being served.

The time spent waiting for service to be commenced will be zero if the queuing system is empty (i.e., no queue and no-one being served) when the individual arrives. The probability of a zero waiting time is thus the same as the probability of the system being empty, which is given by Equation 4.2. That is:

$$Pr(w = 0) = P_0 = 1 - \rho \quad 4.9$$

The probability of waiting for service to start for a time that is greater than zero but not greater than a particular time,w, is:

$$Pr(0 < wait \leq w) = \rho - \rho e^{-(s-r)w} \quad 4.10$$

And the probability of waiting for service to start for a time that is greater than w, is:

$$Pr(wait > w) = \rho e^{-(s-r)w} \quad 4.11$$

As would be expected, the probabilities in Equations 4.9, 4.10 and 4.11 sum to unity.

Over all arrivals, the average, or expected time spent waiting for service to start is:

$$E(w) = \frac{\rho}{s-r} = \frac{r}{s(s-r)} \quad 4.12$$

and the average waiting time over only those arrivals whose wait is non-zero is:

$$E(w \mid w > 0 = \frac{E(w)}{\rho} = \frac{1}{s - r} \quad 4.13$$

Let the total time that any unit spends in the system, including service time, be denoted by $_{\tau}$. Then the average, or expected total time spent in the system, over all arrivals, is:

$$E(\tau) = \frac{1}{s - r} \quad 4.14$$

Comparing this with Equation 4.12, as expected, it is found that:

$$E(\tau) = E(w) + \frac{1}{s} \quad 4.15$$

4.5. Example Application of Steady State Queuing Theory

The practical application of the theory outlined above for queues with random arrivals and service times is illustrated by the following example.

At a major sporting venue, patrons arrive in motor vehicles. At one of the entrances to the car parking area surrounding the venue, a single line of vehicles approaches the manually operated entry gate, where a cash payment is required to gain entry. The process of stopping at the payment point, offering payment, receiving change if necessary and departing the payment point averages 12 seconds per vehicle. Vehicles are arriving randomly at an average rate of 216 vehicles per hour. Appropriate design of the entry arrangements requires knowledge of the following values:

1. the queue storage length required prior to the payment point, assuming that the length provided must be adequate for at least 95% of the time
2. the proportion of vehicles that will have to wait in a queue before entering the payment point
3. the average total delay, including time spent in the payment process, to vehicles entering the parking area at this gate.

Value (1), the required queue storage length, is determined by application of Equation 4.5, as the aim to find the smallest value of N for which n, the number of vehicles in the queuing system, satisfies:

$$Pr(n > N) = \rho^{N+1} \leq 0.05$$

The average arrival rate of vehicles, r, is 216 veh/h and the average service time of 12 s/veh means that the average service rate, s, is 300 veh/h. Therefore, by Equation 4.1, ρ = r/s = 0.72 and the required value of N is identified as 9, by calculating:

$$Pr(n > 8) = (0.72)^9 = 0.052$$

and

$$Pr(n > 9) = (0.72)^{10} = 0.037$$

Given that n includes the vehicle occupying the payment point, space must be provided for eight vehicles to queue before the payment point. If a length of 6 m is allowed for each queued vehicle, a storage length of at least 48 m should be provided.

The following are other results related to queue lengths that may be of interest:

- The probability that no vehicles are present at any given instant is derived from Equation 4.2 as:

$$P_0 = 1 - \rho = 1 - 0.72 = 0.28 \text{ or } 28\%$$

- The probability of there being exactly 6 vehicles present, including that in service is given by Equation 4.3 as:

$$P_6 = (1 - \rho)\rho^6 = (1 - 0.72)(0.72)^6 = 0.0390 \text{ or } 3.9\%$$

- The mean number of vehicles present, including that in service, is given by Equation 4.4 as:

$$E(n) = \frac{\rho}{1 - \rho} = \frac{0.72}{0.28} = 2.57 \text{ vehicles}$$

- The mean number of vehicles present, excluding that in service, is given by Equation 4.6 as:

$$E(m) = \frac{\rho^2}{1 - \rho} = \frac{0.72^2}{0.28} = 1.85 \text{ vehicles}$$

  which is less than $E(n)$ by $\rho = 0.72$ vehicles, in accordance with Equation 4.7.

- The variance of the mean number of vehicles present, including that in service is given by Equation 4.8 as:

$$\sigma^2(n) = \frac{\rho}{(1 - \rho)^2} = \frac{0.72}{0.28^2} = 9.18 \text{ (vehicles)}^2$$

  which implies a standard deviation of 3.03 vehicles in the queue length distribution.

Value (2), the proportion of vehicles that will have to wait in a queue before entering the payment point, is obtained by noting that Equation 4.9 calculates that the probability that any arriving vehicle will **not** have to wait is 1 – ρ. Therefore, the probability that an arriving vehicle **will** have to wait is ρ, that is, 0.72. Hence 72% of vehicles will have to wait before entering the payment point.

Value (3), the average total delay to entering vehicles, including time spent in the payment process, is given by Equation 4.14 as:

$$E(\tau) = \frac{1}{s - r}$$

As observed above, the average service rate, s, is 300 veh/h, while the average arrival rate, r, is 216 veh/h. Each of these rates is converted to the units of veh/s by division by 3600. Therefore, the average total delay to entering vehicles is:

$$\frac{1}{s - r} = \frac{3600}{300 - 216} = 42.9 \frac{s}{veh}$$

This total delay includes the time spent in the payment process, which averages 12 s/veh. Therefore, the average waiting time in the queue before entering the payment point (see Equation 4.15) is 42.9 –12 = 30.9 s/veh.

The following are other results related to delays that may be of interest:

- The probability of a vehicle having to wait more than 20 s before entering the payment point is given by Equation 4.11 as:

$$\Pr(W > 20) = \rho e^{-(s-r)20} = 0.72e^{-\frac{(300-216)20}{3600}} = 0.4515 \text{ or } 45.15\%$$

- Overall arriving vehicles, the average waiting time before entering the payment point is also given by Equation 4.12 (consistent with the result obtained above) as:

$$E(w) = \frac{\rho}{s-r} = \frac{3600.(0.72)}{300-216} = 30.9 \text{ s}$$

- Over only those arriving vehicles that cannot immediately enter, the average waiting time before entering the payment point is given by Equation 4.13 as:

$$E(w \mid w > 0) = \frac{1}{s-r} = \frac{3600}{300-216} = 42.9 \text{ s}$$

4.6. Summary of Queuing Theory Formulae

Table 4.1 summarises the key formulae from elementary queuing theory.

Table 4.1: Summary of queuing theory formulae

Description	Eqn no.	Equation(1)
Utilisation factor (ratio of arrival and service rates)	4.1	$\rho = r/s$
Probability of the system being empty	4.2	$P_0 = 1 - \rho$
Probability of exactly n units in the system	4.3	$P_0 = (1-\rho)\,\rho^n$
Expected or average number of units in the system, including unit in service	4.4	$E(n) = \frac{\rho}{1-\rho} = \frac{r}{s-r}$
Probability of more than n units in the system	4.5	$\Pr(n > N) = \rho^{N+1}$
Expected or average number of units in the system, excluding unit in service	4.6	$E(m) = \frac{\rho^2}{1-\rho} = \frac{\rho}{1-\rho} - \rho$
Relationship between expected numbers in the system including and excluding unit in service	4.7	$E(m) = E(n).\ \rho = E(n) - \rho$
Variance of expected number in the system, including unit in service	4.8	$\sigma^2(n) = \frac{\rho}{(1-\rho)^2}$
Probability of a zero waiting time before start of service	4.9	$\Pr(w = 0) = P_0 = 1 - \rho$
Probability of wait greater than zero but not greater than w before start of service	4.10	$\Pr(0 < \text{wait} \le w) = \rho - \rho e^{-(s-r)w}$

Description	Eqn no.	Equation(1)
Probability of wait greater than w before start of service	4.11	$\Pr(\text{wait} > w) = \rho e^{-(s-r)w}$
Expected or average waiting time before start of service	4.12	$E(w) = \frac{\rho}{s-r} = \frac{r}{s(s-r)}$
Expected or average waiting time before start of service for those with a non-zero wait	4.13	$E(w \mid w > 0) = \frac{E(w)}{\rho} = \frac{1}{s-r}$
Expected or average total time in the system, including service time	4.14	$E(\tau) = \frac{1}{s-r}$
Relationship between expected total time in the system and expected waiting time before start of service	4.15	$E(\tau) = E(w) + \frac{1}{s}$

1 *r is the average arrival rate and s is the average service rate for units in the queue; their ratio, $\rho = r/s$, is known as the utilisation factor for the queue (see also Section 4.1 and Equation 4.1).*

5. Gap Acceptance

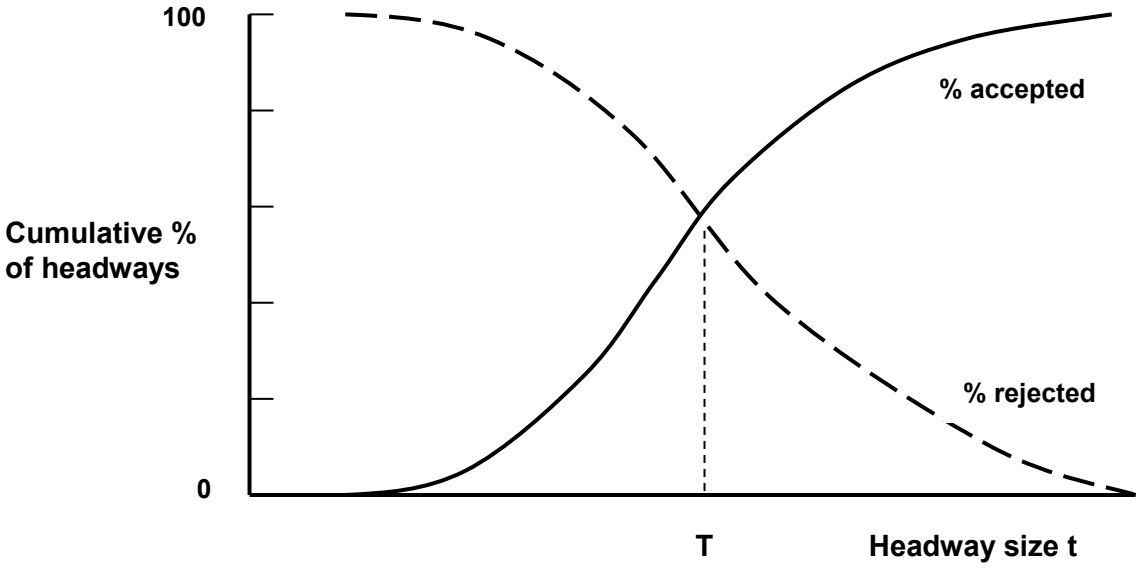
5.1. Introduction and Definitions

5.1.1. General Introduction

In many traffic situations, such as pedestrians or cyclists crossing a road, vehicles overtaking on undivided roads or side-road traffic entering major roads at unsignalised intersections, road users must wait for acceptable time gaps in the traffic stream to which they must give way before they can proceed. Such situations are examined using gap acceptance analysis.

The magnitude of the time interval considered acceptable to undertake a manoeuvre involving gap acceptance depends on the road geometry at the site, the characteristics of the traffic and the nature of the manoeuvre itself. Also, in the same gap acceptance situation, different people may be prepared to accept gaps of different sizes and even the same person, in the same situation, may be willing to accept smaller or larger gaps at different times. Typical observed behaviour for gap acceptance is shown in Figure 5.1. However, in the discussion of elementary gap acceptance analysis below, it is assumed that for a given manoeuvre in each situation, there is a single time gap which will be the minimum accepted by all drivers at all times. This is called the **critical** gap and it usually is identified as an average value from observed gap acceptances and rejections. For example, Raff and Hart (1950) proposed a method in which a diagram similar to Figure 5.1 is plotted from field observations and the critical gap is taken to be the gap 'T' corresponding to the intersection of the acceptance and rejection curves.

Figure 5.1: Typical gap acceptance behaviour



The other important factor in gap acceptance analysis is the headway distribution in the traffic stream which has priority, as this determines the frequencies with which gaps of different sizes occur. The discussion in Sections 5.2 and 5.3 limits itself to the simplest case of random arrivals, that is, a negative exponential headway distribution. Section 5.4 then considers formulae for the case of displaced negative exponential headways in the major traffic flow, as an illustration of how other headway distributions can be addressed.

5.1.2. Definitions

Table 5.1 provides definitions of key terms as they are used in the development and application of gap acceptance analysis.

Table 5.1: Definitions of gap acceptance terms

Term	Definition
Major and minor traffic streams	In a gap acceptance situation, the traffic stream that has priority is called the major traffic stream and the traffic stream that must yield right of way is called the minor traffic stream.
Unit (in a traffic stream)	An element of the traffic stream (vehicle, pedestrian, bicycle, etc.) moving as a single entity.
Major and minor roads	At a road intersection operating by gap acceptance, the road carrying the major traffic stream is called the major road and that carrying the minor traffic stream, the minor road.
Gap acceptance point	The point at which a minor traffic stream unit can commence its desired manoeuvre when a suitable gap presents itself (e.g. the arrival of a minor road vehicle at the stop line, or a pedestrian at the kerb).
Lag	Time interval between the arrival of a minor traffic stream unit at the gap acceptance point and the arrival of the next major traffic stream unit.
Gap	Time interval between the arrival of two consecutive major traffic stream vehicles (i.e. a major traffic stream headway), commencing after the arrival of a minor traffic stream member at the gap acceptance point.
Critical gap (critical lag)	The minimum gap (lag) acceptable to a minor traffic stream unit to perform a given manoeuvre in each gap acceptance situation. **(In the discussion in this section, it is assumed that the critical lag and the critical gap are of equal size for any given situation and are always the same for all road users.)**
Follow-up headway	The minimum additional duration of a major traffic stream gap (or lag) required to allow one additional minor traffic stream unit to follow the unit preceding it into the same manoeuvre, utilising the same gap (or lag).
Anti-block	That part of a sufficiently large gap or lag during which there remains a time at least as large as the critical gap before the end of the gap or lag, so that a minor traffic stream member could commence its desired manoeuvre.
Block	A time interval during which zero, one or more major stream traffic units may pass, while there always remains a time less than the critical gap before the arrival of the next major traffic stream unit, so that a minor traffic stream member could not commence its desired manoeuvre.

5.2. Principal Gap Acceptance Formulae

The principal formulae derived from gap acceptance theory for the case of random arrivals in both the major and minor traffic streams are presented in the following sections. The detailed derivations of formulae related to delays and to absorption capacities are provided in Commentaries 5 and 7 respectively.

5.2.1. Delays

The formulae related to delays experienced by minor flow traffic in gap acceptance situations with random arrivals in both the major and minor traffic streams are derived in Commentary 5. The key formulae are as follows.

[see Commentary 5]

The proportion of minor traffic stream units that will be delayed at the gap acceptance point (e.g. at the stop line) is:

$$\text{Proportion delayed} = 1 - e^{-qT} \quad 5.1$$

where

q = the volume of the conflicting major traffic stream, in veh/s

T = the size of the critical gap (or critical lag), in s/veh

An example, of the practical application of this formula to decisions on the provision of turning lanes at unsignalised intersections is discussed in Commentary 6.

[see Commentary 6]

The average delay at the gap acceptance point for all minor traffic stream units (including those that experience no delay) is:

$$d_{av}(d \geq 0) = \frac{1}{qe^{-qT}} - \frac{1}{q} - T\frac{s}{veh} \quad 5.2$$

The average delay at the gap acceptance point for those minor traffic stream units that do experience such delay is:

$$d_{av}(d > 0) = \frac{1}{qe^{-qT}} - \frac{T}{1 - e^{-qT}}\frac{s}{veh} \quad 5.3$$

Tanner (1962) provided an alternative method of calculating average delay that combines both gap acceptance and queuing theory in one formula. Details of this method are presented in Commentary 7.

[see Commentary 7]

5.2.2. Absorption Capacities

An important value for road traffic applications is the theoretical maximum rate at which minor traffic stream vehicles can cross or be absorbed into the major traffic stream in a gap acceptance situation. This maximum rate is known as the **theoretical absorption capacity**.

The formula for theoretical absorption capacity in gap acceptance situations with random arrivals in both the major and minor traffic streams is derived in Commentary 8.

[see Commentary 8]

This formula is as follows.

$$C = \frac{qe^{-qT}}{1 - e^{-qT_0}} \quad 5.4$$

where

q and T = defined in Section 5.2.1

T_0 = the follow-up headway, in s/veh

C = the theoretical absorption capacity in veh/s (multiply by 3600 for veh/h).

It is observed that the absorption capacity C is a theoretical upper limit, which may not be achieved in practice. It is common practice in traffic analysis to define a **practical absorption capacity** as

$$C_p = \alpha . C \quad 5.5$$

where the factor α is typically taken to be in the range 0.80 to 0.85.

5.2.3. Multi-lane Flows

The preceding development of gap acceptance results has implied a single-lane major traffic stream, in which minor traffic stream units seek acceptable gaps. If the major traffic stream comprises two or more lanes, travelling in the same or opposite directions, minor stream units can proceed with their desired manoeuvre only if an acceptable gap occurs simultaneously in each of the major stream lane flows that conflicts with that manoeuvre.

To examine how an analysis can accommodate this situation, consider the existence of two major stream lane flows with volumes q_1 and q_2, each with random arrivals (i.e. negative exponential headway distributions) and assume that a critical gap T is sought simultaneously in the two-lane flows. The individual probabilities of suitable gaps in each flow are:

$$\Pr(h \geq T) \text{ in } q_1 \text{ stream } = e^{-q_1 T} \quad 5.6a$$

And

$$\Pr(h \geq T) \text{ in } q_2 \text{ stream } = e^{-q_2 T} \quad 5.6b$$

If the two lane flows are assumed to be independent of each other, the probability of a suitable headway occurring simultaneously in both lanes is simply the product of the two probabilities in Equations 5.6a and 5.6b, so that:

$$\Pr\big((h \geq T) \text{ in both lanes}\big) = e^{-q_1 T} . e^{-q_2 T} = e^{-(q_1 + q_2)T} \quad 5.7$$

This is of the same form as the probabilities in Equations 5.6a and 5.6b, implying that, in gap acceptance analysis, a number of independent, separate, major traffic stream flows that conflict with a given minor stream manoeuvre can be treated as though they were a single flow with volume equal to the sum of the volumes of the individual flows.

5.3. More Complex Gap Acceptance Situations

5.3.1. Minor Road Approaches with Mixed Traffic

Equation 5.4, giving the theoretical absorption capacity in a gap acceptance situation, was derived assuming that the minor traffic stream units are homogeneous, in that each faces the same conflicting major traffic stream and each seeks to carry out the same manoeuvre, so that the same critical gap and follow-up headway applies to all.

In practice, this would be an unusual situation. From the minor approach to a T-intersection, for example, it is likely that some drivers will wish to turn left while others will wish to turn right. If the major road is carrying two-way traffic, this means that left turners from the minor road will need to give way only to traffic from their right, while right turners must give way to major road traffic in both directions. Not only will the conflicting major road volumes differ between left and right turners, but so will the required critical gaps and follow-up headways.

In addition, while an average value of a critical gap is considered adequate to represent the requirements of drivers of the same types of vehicles carrying out the same manoeuvre, analyses should recognise that average values may be significantly different for different types of vehicles. A large truck, for example, is likely to require a much larger gap to make a right turn than would a passenger car in the same situation.

The accommodation of these differences within the analysis process is achieved by considering the minor traffic stream to be divided into a number of homogeneous sub-streams, each consisting of the same type of vehicle seeking to carry out the same manoeuvre (e.g. trucks turning right). For each sub-stream, i, Equation 5.4 can then be used, with values of q, T and T_0 appropriate to that sub-stream, to calculate C_i, the theoretical absorption capacity that would apply to the approach lane if all its traffic belonged to sub-stream i.

Now if the total minor traffic stream consists of n such sub-streams and p_i is the proportion of the total minor traffic volume that is in sub-stream i, i = 1, ..., n, then the theoretical absorption capacity for the whole minor traffic stream is:

$$C_T = \frac{1}{\sum_1^n = 1 \frac{P_i}{C_i}} \qquad 5.8$$

This can be seen by noting that the overall capacity of the minor road approach lane in vehicles per unit time is the inverse of the average time interval between successive vehicles entering the intersection from that lane. This average time interval is the volume-weighted mean of the sub-stream average entry headways $1/C_i$, i = 1, …, n, that is, the denominator of the right-hand side of Equation 5.8.

5.3.2. Different Critical Gaps for Different Conflicting Major Flows

In some cases, a minor traffic stream driver may seek different sized gaps in the different major traffic flows that conflict with the desired movement at an intersection. A common case is that of a vehicle turning right from a minor road into a major road carrying traffic in both directions. From the stop line, the minor road vehicle must effectively cross the traffic flow coming from its right and then join the traffic flow from its left. The entering vehicle is clear of the flow from the right as soon as it has crossed that flow, whereas in joining the flow from the left, the entering vehicle faces the possibility of a rear-end collision with an approaching vehicle, even after the joining manoeuvre is completed. Hence, the driver of the minor road vehicle may seek to combine an acceptable gap in the flow from the right with a larger-sized gap in the flow from the left.

Typically, in such cases, a single follow-up headway size would apply, as this is principally concerned with the process of a following vehicle moving up to the point where the driver can assess the major road traffic coming from both directions, then decide whether to use the same gap as the vehicle ahead.

The theoretical absorption capacity for this situation is developed in Commentary 9 in a manner very similar to that in Section 5.2.2.

[see Commentary 9]

Assume that the total major traffic stream is made up of the traffic flow from the left, with volume q_L, and the flow from the right, with volume q_R. Assume also that a critical gap T_L in the major traffic flow from the left is the minimum that will allow one minor stream right-turner to cross that stream, and that a critical gap T_R in the major traffic flow from the right is the minimum that will allow one minor stream right-turner to join that stream. Finally, assume that an additional follow-up headway T_0 is sufficient to allow one additional minor stream vehicle to follow in undertaking the manoeuvre.

Then, the development in Commentary 9 shows that the theoretical maximum rate at which minor stream vehicles can turn right, that is, the theoretical absorption capacity, is:

$$C = \frac{(q_L + q_R)e^{-(q_L T_L + q_R T_R)}}{1 - e^{-(q_L + q_R)T_0}} \quad 5.9$$

While it is relatively rare for traffic analysts to consider differently sized critical gaps in the major traffic flows from opposite directions, this refinement may be appropriate in some circumstances. In any such case, Equation 5.9 defines the theoretical absorption capacity that should be used.

5.4. Formulae for Displaced Negative Exponential Headways in Major Traffic

Sections 5.2 and 5.3 have restricted their attention to the case of random arrivals – that is, a negative exponential distribution of headways – in the major traffic flow. As the type of headway distribution directly affects the pattern of acceptable gaps, it is instructive to consider how the key gap acceptance formulae are changed if a different type of distribution applies. As an illustration, the case of a displaced negative exponential distribution of headways in the major traffic flow is considered in Commentary 10 and the key formulae from that analysis are summarised below.

[see Commentary 10]

The probability density function for headway size, t, for a displaced negative exponential distribution with a minimum possible headway β is:

$$f(t) = \frac{q}{1 - q\beta} e^{\frac{-q(t-\beta)}{1-q\beta}} \text{ for } t \geq \beta \quad 5.10$$

and

$$f(t) = 0 \text{ for } t < \beta \quad 5.11$$

The probability of a headway of size t or greater (where $t \geq \beta$), in the major traffic stream, is:

$$Pr(\beta \leq t \leq h) = e^{\frac{-q(t-\beta)}{1-q\beta}} \quad 5.12$$

It follows from Equation 5.12 that, for a critical gap T, the proportion of minor road vehicles that suffer no delay at the stop line is:

$$\text{Proportion not delayed } = e^{\frac{-q(T-\beta)}{1-q\beta}} \quad 5.13$$

Conversely, the probability of a major traffic headway of size t or less (where $t \geq \beta$) is:

$$\Pr(\beta \leq h \leq t) = 1 - e^{\frac{-q(t-\beta)}{1-q\beta}} \quad 5.14$$

and, for a critical gap T, the proportion of minor road vehicles that are delayed at the stop line is:

$$\text{Proportion delayed} = 1 - e^{\frac{-q(t-\beta)}{1-q\beta}} \quad 5.15$$

The average duration of headways greater than or equal to t (where $t \geq \beta$) is:

$$h_{av}(\beta \leq t \leq h) = \frac{1}{q} + t - \beta \quad 5.16$$

and the average duration of headways less than or equal to t (where $t \geq \beta$) is:

$$h_{av}(\beta \leq h \leq t) = \frac{1}{q} - \frac{(t-\beta)e^{\frac{-q(t-\beta)}{1-q\beta}}}{1 - e^{\frac{-q(t-\beta)}{1-q\beta}}} \quad 5.17$$

Then the average delay experienced by all minor traffic stream units at the gap acceptance point is:

$$d_{av}(d \geq 0) = \frac{1}{qe^{\frac{-q(T-\beta)}{1-q\beta}}} - \frac{1}{q} - (T-\beta) \quad 5.18$$

The average delay at the gap acceptance point to only those minor traffic stream units that do experience such delay is:

$$d_{av}(d > 0) = \frac{1}{qe^{\frac{-q(T-\beta)}{1-q\beta}}} - \frac{(T-\beta)}{\left(1 - e^{\frac{-q(T-\beta)}{1-q\beta}}\right)} \quad 5.19$$

Finally, the theoretical absorption capacity for a minor traffic stream requiring a critical gap T and with a follow-up headway T_0, giving way to a major traffic stream of volume q, with displaced negative exponential headways each greater than or equal to β, is:

$$C = \frac{qe^{\frac{-q(T-\beta)}{1-q\beta}}}{1 - e^{\frac{-qT_0}{1-q\beta}}} \quad 5.20$$

Note that Equation 5.8 is also valid for displaced negative exponential distributions of major traffic headways. Recall that this equation addresses a minor traffic stream consisting of n sub-streams i, $i = 1,\ldots,n$, with p_i being the proportion of the minor traffic volume in sub-stream i and C_i being the theoretical absorption capacity for the approach lane if all its traffic belonged to sub-stream i. For this case, the theoretical absorption capacity for the whole minor traffic stream is:

$$C_T = \frac{1}{\sum_{i=1}^{n} \frac{p_i}{C_i}} \quad 5.8$$

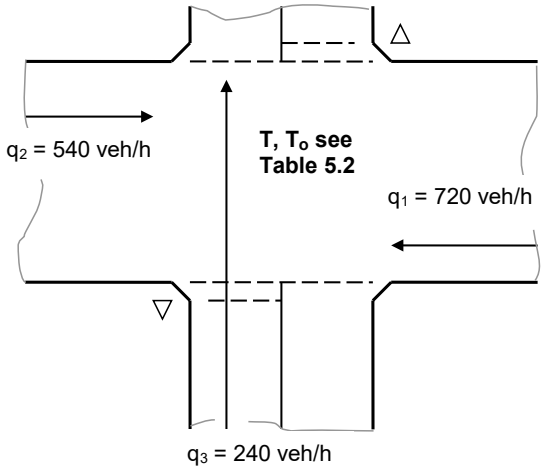
For displaced negative exponential major road headways, however, each C_i would be evaluated using Equation 5.20.

5.5. Example Applications of Gap Acceptance Analysis

5.5.1. Example 1 – Random Arrivals on Major Road, Mixed Minor Road Traffic

To illustrate the practical application of elementary gap acceptance analysis, consider the cross-intersection shown in Figure 5.2.

Figure 5.2: Example cross-intersection



The major road carries traffic flows of 540 veh/h from the left of drivers on the minor approach and 720 veh/h from their right, with random arrivals in both cases. At a particular time of day, the traffic on one single-lane minor road approach has a volume of 240 veh/h, of which 10% are trucks and the remainder cars. All the trucks turn right but, of the cars, 60% cross straight through, 25% turn left and 15% turn right. At this time of day, there is negligible traffic on the other minor road approach.

Required critical gaps and follow-up headways for right turning trucks and for through, left turning and right turning cars are as shown in Table 5.2.

Table 5.2: Critical gaps and follow-up headways

Minor traffic component	Proportion of total volume	Critical gap (s/veh)	Follow-up headway (s/veh)
Through cars	0.540	$T_L = T_R = 5.0$	2.5
Left turning cars	0.225	$T_R = 4.0$	2.0
Right turning cars	0.135	$T_L = 6.0$, $T_R = 5.0$	2.5
Right turning trucks	0.100	$T_L = 8.0$, $T_R = 7.0$	3.5

It is desired to estimate the practical capacity of the intersection for this approach and the average delay at the stop line for through cars.

The first step in estimating capacity is to identify the theoretical absorption capacity that would apply for this approach if its traffic was comprised entirely of vehicles in each of the 'Minor traffic component' categories in Table 5.2.

For the first two categories, only one critical gap applies, and Equation 5.4 can be used. For through cars, the opposing volume q is the sum of the volumes for major road traffic from the left and the right (540 + 720 = 1,260 veh/h or 0.35 veh/s), while for left turning cars, q consists only of the major road traffic from the right (720 veh/h = 0.20 veh/s). Hence, for through cars,

$$C = \frac{qe^{-qT}}{1 - e^{-qT_0}} = \frac{0.35e^{-(0.35x5)}}{1 - e^{-(0.35x2.5)}} = \frac{0.10430veh}{s} = \frac{375.5veh}{h}$$

and for left-turning cars,

$$C = \frac{qe^{-qT}}{1 - e^{-qT_0}} = \frac{0.2e^{-(0.2x4)}}{1 - e^{-(0.2x2)}} = \frac{0.27259veh}{s} = \frac{981.3veh}{h}$$

For the last two categories, different critical gaps apply for the two major traffic flow directions and Equation 5.9 must be applied. Hence, for right-turning cars,

$$C = \frac{(q_L + q_R)e^{-(q_L T_L + q_R T_R)}}{1 - e^{-(q_L + q_R)T_0}} = \frac{0.35e^{-(0.15x6 + 0.2x5)}}{1 - e^{-(0.35x2.5)}} = \frac{0.08977veh}{s} = \frac{323.2veh}{h}$$

and for right-turning trucks,

$$C = \frac{(q_L + q_R)e^{-(q_L T_L + q_R T_R)}}{1 - e^{-(q_L + q_R)T_0}} = \frac{0.35e^{-(0.15x8 + 0.2x7)}}{1 - e^{-(0.35x3.5)}} = \frac{0.03681veh}{s} = \frac{132.5veh}{h}$$

The overall theoretical absorption capacity for the mixed traffic can then be determined using Equation 5.8, as follows:

$$C_T = \frac{1}{\sum_{i=1}^{4} \frac{p_i}{C_i}} = \frac{1}{\left(\frac{0.54}{375.5} + \frac{0.225}{981.3} + \frac{0.135}{323.2} + \frac{0.1}{132.5}\right)} = \frac{352.1veh}{h}$$

The practical absorption capacity might be estimated as 80% of this figure, that is, 282 veh/h.

The average delay at the stop line for through cars is estimated by application of Equation 5.3, that is:

$$d_{av}(d \geq 0) = \frac{1}{qe^{-qT}} - \frac{1}{q} - T = \frac{1}{0.35e^{-(0.35x5)}} - \frac{1}{0.35} - 5 = \frac{8.58s}{veh}$$

5.5.2. Example 2 – Displaced Negative Exponential Headways on Major Road

Assume that at the intersection shown in Figure 5.2, all minor road traffic consists of through cars with gap acceptance characteristics as shown for that category in Table 5.2. For this situation, compare the average delays at the stop line and the theoretical absorption capacities that would apply if the major road volumes were as shown and the major road headway distributions were:

(a) Negative exponential.
(b) Displaced negative exponential, with minimum headway 1.5 s/veh.

For a negative exponential headway distribution on the major road

The average delay at the stop line over all through cars was calculated in Example 1, using Equation 5.2, with $q = \frac{1260\text{veh}}{\text{h}} = \frac{0.35\text{veh}}{\text{s}}$ and $T = \frac{5.0\text{s}}{\text{veh}}$, as 8.58 s/veh.

The average delay to only those through cars that are delayed is given by Equation 5.3, with the same values of q and T, as:

$$d_{av}(d > 0) = \frac{1}{0.35e^{-1.75}} - \frac{5.0}{(1 - e^{-1.75})} = \frac{10.39\text{s}}{\text{veh}}$$

The theoretical absorption capacity for through cars also was calculated in Example 1, using Equation 5.4, with $q = \frac{1260\text{veh}}{\text{h}} = \frac{0.35\text{veh}}{\text{s}}$, $T = \frac{5.0\text{s}}{\text{veh}}$ and $T_0 = \frac{2.5\text{s}}{\text{veh}}$, as 0.1043 veh/s or 375.5 veh/h.

For a displaced negative exponential headway distribution with minimum headway 1.5 s/veh

The average delay at the stop line over all through cars is given by Equation 5.18, with $q = \frac{1260\text{veh}}{\text{h}} = \frac{0.35\text{veh}}{\text{s}}$, $T = \frac{5.0\text{s}}{\text{veh}}$ and $\beta = \frac{1.5\text{s}}{\text{veh}}$, as:

$$d_{av}(d \geq 0) = \frac{1}{0.35e^{\frac{-0.35(5.0-1.5)}{(1-0.35x1.5)}}} - \frac{1}{0.35} - (5.0 - 1.5) = \frac{31.31\text{s}}{\text{veh}}$$

The average delay to only those through cars that are delayed is given by Equation 5.19, with the same values of q, T and β, as:

$$d_{av}(d > 0) = \frac{1}{0.35e^{\frac{-0.35(3.5)}{0.475}}} - \frac{(5.0-1.5)}{\left(1-e^{\frac{-0.35(3.5)}{0.475}}\right)} = \frac{33.88\text{s}}{\text{veh}}$$

The theoretical absorption capacity for through cars is obtained using Equation 5.20, with $q = 1260\ \text{veh/h} = 0.35\ \text{veh/s}$, $\beta = 1.5\ \text{s/veh}$, $T = 5.0\ \text{s/veh}$ and $T_0 = 2.5\ \text{s/veh}$, as:

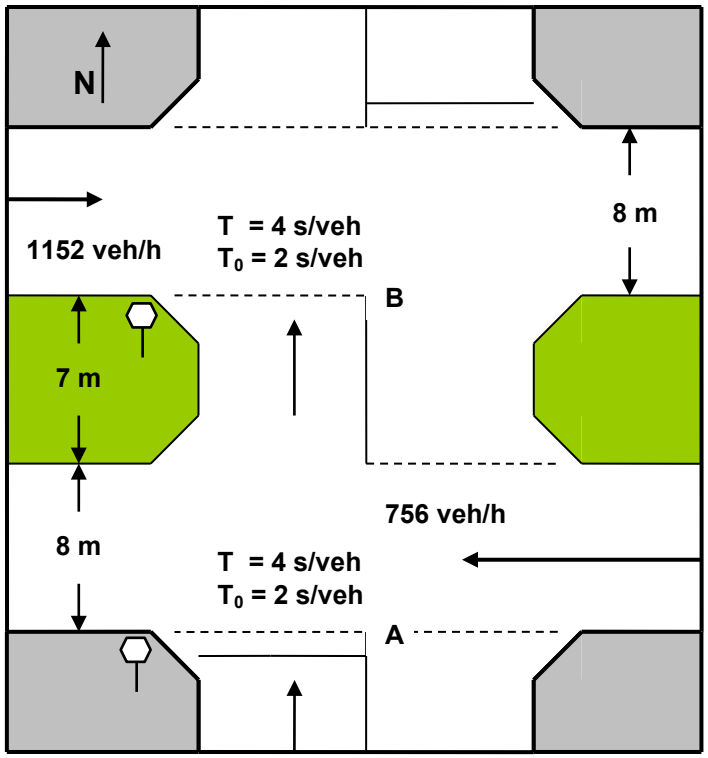
$$C = \frac{0.35e^{\frac{-0.35(3.5)}{0.475}}}{1 - e^{\frac{-0.35(2.5)}{0.475}}} = 0.03155\frac{\text{veh}}{\text{s}} = 113.6\ \frac{\text{veh}}{\text{h}}$$

The significant differences in delays and absorption capacities between the two cases are explained by the fact that with a negative exponential distribution, 17.4% of major road headways are larger than the critical gap, whereas with a displaced negative exponential distribution, this proportion is reduced to only 7.6%.

5.5.3. Example 3 – Staged Crossing

At an unsignalised intersection where the major road is divided, minor road traffic may be able to cross or turn right in two stages, crossing the first carriageway and then sheltering in the median gap before crossing or turning into the second carriageway. Figure 5.3 illustrates such a situation.

Figure 5.3: Staged crossing



At the unsignalised intersection shown in Figure 5.3, minor road vehicles must give way to traffic on each carriageway of the divided major road. The minor road traffic consists of passenger cars, so that the 7 m median width on the major road is enough to shelter one minor road vehicle waiting to cross the second carriageway. Traffic on the major road arrives randomly from both directions and has the volumes shown. Critical gaps and follow-up headways required by northbound traffic to cross each carriageway of the major road are also as shown. The task is to estimate the capacity of the intersection to accommodate northbound through traffic.

Start by considering the process involved in northbound traffic crossing the intersection. Because a vehicle waiting at the first stop line (A) cannot depart if the median storage space is occupied, the crossing process for any such vehicle consists of three components:

1. at stop line A, noting the departure of the preceding vehicle from stop line B
2. moving from stop line A to stop line B
3. waiting at stop line B for an opportunity to cross the second carriageway.

If there was no impediment to a northbound vehicle continuing on its way once it had crossed the first carriageway, the average delay at stop line A would be given by Equation 5.2, with $q = 756$ veh/h = 0.210 veh/s and $T = 40$ s/veh, as $d_{av} = 2.27$ s/veh. Therefore, once the median storage space is empty, a vehicle at stop line A can expect to wait 2.27 s before being able to depart.

The travel time between stop lines A and B is estimated as the time to cover a distance of 15 m at an average speed of (say) 12 km/h or 3.33 m/s, that is, 4.5 s.

This means that when a vehicle departs stop line B, it will take 2.27 + 4.5 = 6.77 s, on average, for the following vehicle to reach stop line B. This 6.77 s/veh can therefore be considered the effective follow-up headway (in lieu of the 2.0 s/veh shown in Figure 5.3) for northbound traffic crossing the second carriageway.

The theoretical capacity of the intersection for northbound through traffic can therefore be estimated from Equation 5.4 with q = 1152 veh/h = 0.320 veh/s, T = 40 s/veh and T_0 = 6.77 s/veh, giving a capacity of 0.1005 veh/s or 362 veh/h. The practical capacity may be estimated as 80 to 85% of this value, that is, 290 to 308 veh/h.

The average delay at stop line B is given by Equation 5.2, with q = 1152 veh/h = 0.320 veh/s and T = 40 s/veh, as d_{av} = 4.1145 s/veh. If the average time of 6.77 s/veh for a following vehicle to reach that stop line is added to this, a total of 10.88 s/veh is obtained. The inverse of this provides another estimate of the rate at which northbound vehicles depart from stop line B, this being 0.0919 veh/s or 331 veh/h. This value is close to the theoretical capacity evaluated in the preceding paragraph but is less, and not as good an estimate, because it makes no allowance for multiple vehicles crossing the second carriageway within the same gap, which may occur for gaps greater than 10.77 s/veh in the eastbound major road flow.

It is instructive to compare the above result with the capacity that would apply if the major road was undivided, so that crossing minor road vehicles would have to seek simultaneous gaps of sufficient size in both directions of major road traffic.

Given a major road width of 16 m (the sum of the two carriageway widths) and a total priority flow of 756 + 1152 = 1908 veh/h = 0.530 veh/s, a critical gap of 6 s/veh and a follow-up headway of 3 s/veh would be appropriate for the crossing manoeuvre. Substituting these values into Equation 5.4 produces a theoretical absorption capacity of 0.0277 veh/s or 98 veh/h, only 27% of the capacity with a staged crossing.

5.6. Summary of Basic Gap Acceptance Formulae

Table 5.3 summarises the key formulae from elementary gap acceptance theory, for both negative exponential and displaced negative exponential headway distributions in the major traffic stream.

Table 5.3: Summary of elementary gap acceptance formulae

Description	Eqn no.	Equation
For negative exponential distribution of major flow headways:		
Proportion of minor traffic stream units delayed at stop line or (for pedestrians crossing) kerb	5.1	$\text{Proportion delayed} = 1 - e^{-qT}$
Average delay at stop line or kerb over all minor stream units	5.2	$d_{av}(d \geq 0) = \frac{1}{qe^{-qT}} - \frac{1}{q} - T$
Average delay at stop line or kerb for those minor stream units that are delayed	5.3	$d_{av}(d > 0) = \frac{1}{qe^{-qT}} - \frac{T}{1 - e^{-qT}}$
Theoretical absorption capacity for basic gap acceptance situation	5.4	$C = \frac{qe^{-qT}}{1 - e^{-qT_0}}$
Theoretical absorption capacity for minor traffic manoeuvre requiring different critical gaps in major flows from left and right	5.9	$C = \frac{(q_L + q_R)e^{-(q_L T_L + q_R T_R)}}{1 - e^{-(q_L + q_R)T_0}}$
For displaced negative exponential distribution of major flow headways:		
Proportion of minor traffic stream units delayed at stop line or (for pedestrians crossing) kerb	5.15	$\text{Proportion delayed} = 1 - e^{\frac{-q(t-\beta)}{1-q\beta}}$
Average delay at stop line or kerb over all minor stream units	5.18	$d_{av}(d \geq 0) = \frac{1}{qe^{\frac{-q(T-\beta)}{1-q\beta}}} - \frac{1}{q} - (T - \beta)$

Description	Eqn no.	Equation
Average delay at stop line or kerb for those minor stream units that are delayed	5.19	$d_{av}(d > 0) = \frac{1}{qe^{\frac{-q(T-\beta)}{1-q\beta}}} - \frac{(T-\beta)}{\left(1 - e^{\frac{-q(T-\beta)}{1-q\beta}}\right)}$
Theoretical absorption capacity for basic gap acceptance situation	5.20	$C = \frac{qe^{\frac{-q(T-\beta)}{1-q\beta}}}{1 - e^{\frac{-qT_0}{1-q\beta}}}$
For either of the above distributions of major flow headways:		
Overall absorption capacity for a minor traffic stream consisting of i sub-streams with different gap acceptance behaviour	5.8	$C_T = \frac{1}{\sum_{i=1}^{n} \frac{p_i}{C_i}}$

6. Combined Gap Acceptance and Queueing Theory

6.1. Absorption Capacity as a Queuing Service Rate

Section 5.3.1 has considered the absorption capacity of an intersection approach lane where intersection entry is by gap acceptance and where the approach lane traffic is made up of component sub-streams that differ in their gap acceptance behaviour, due to differences in their combinations of vehicle type and desired manoeuvre at entry.

Equation 5.8 evaluates the overall absorption capacity, C_T, for the approach lane in terms of p_i, the proportion of the total minor traffic volume that is in sub-stream i, and C_i, the theoretical absorption capacity that would apply to the approach lane if all its traffic belonged to sub-stream i, $i = 1, \ldots, n$, as:

$$C_T = \frac{1}{\sum_{i=1}^{n} \frac{p_i}{C_i}} \qquad 5.8$$

The capacity C_T is an appropriate queuing service rate, s, to use in the calculation of values associated with queue length (Equations 4.3 to 4.7 inclusive) or with delay in reaching the head of the queue (Equations 4.10 to 4.14 inclusive). When considering vehicles in a particular sub-stream, however, different queuing service rates are appropriate for different components of delay.

Consider, for example, the total delay (from joining the tail of the queue to entering the intersection) for a vehicle in sub-stream i. This must be calculated using Equation 4.15, the right hand side of which has two components – the delay before reaching the head of the queue E(w), which can be calculated from Equation 4.12, and the delay in service, 1/s.

It is apparent that, in evaluating E(w), it is important to recognise that the delay experienced by any vehicle in progressing through the queue is governed by the behaviour of the queue as a whole, so that both the arrival rate, r, and the service rate, s, must take the values applicable to the whole minor traffic stream. For s, this value is C_T. When a vehicle in sub-stream i reaches the head of the queue, however, its in-service delay depends on its own, particular gap acceptance requirements and the appropriate service rate, s, is therefore C_i. The end result is that the total delay for a vehicle in sub-stream i must be calculated as:

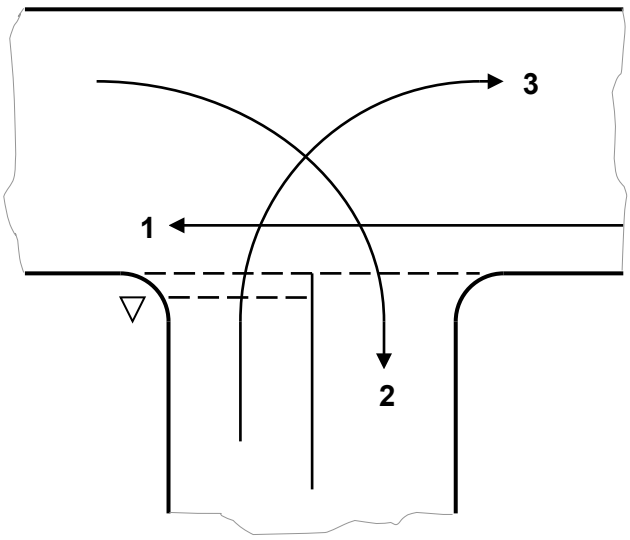
$$E(v_i) = \frac{r}{C_T(C_T - r)} + \frac{1}{C_i} \qquad 6.1$$

It is important to note that the in-service or head-of-queue delay for this situation must be calculated as the inverse of the queuing service rate (i.e. as 1/s or $1/C_i$), not as the stop line delay given by Equation 5.8. This is because the process of 'service' for any vehicle commences with the departure of the vehicle immediately ahead and includes movement up to the stop line, as well as waiting (if required) at the stop line for a suitable gap in the major flow.

6.2. Gap Acceptance with Multiple Levels of Priority

The consideration of elementary gap acceptance in Section 5 assumed a single major traffic stream (perhaps consisting of traffic from more than one direction) with priority over a single minor traffic stream. In practice, however, it is common to see multiple levels of priority, that is, situations in which a minor traffic stream must give way to a higher priority stream which, in turn, must give way to another stream of still higher priority. An example is illustrated by Figure 6.1, in which movement 1 has priority over both movements 2 and 3, and movement 2 has priority over movement 3.

Figure 6.1: Example T-intersection



The example situation in Figure 6.1, with random arrivals assumed for both streams 1 and 2, is used in the following to illustrate how queuing and gap acceptance theory can be combined to develop a means of evaluating absorption capacities for lower priority movements. The method presented was taken from Bennett (1996).

The theoretical absorption capacity for the second-priority movement 2 in Figure 6.1 is simply calculated from Equation 5.14 as:

$$C_2 = \frac{q_1 e^{-q_1 T_{(2)}}}{1 - e^{-q_1 T_{0(2)}}} \qquad 6.2$$

where

q_1 = the volume of traffic stream 1

T_2 and $T_{0(2)}$ = the critical gap and follow-up headway, respectively, for stream 2 vehicles crossing traffic stream 1

To calculate the theoretical absorption capacity for the third-priority movement 3, which must give way to both movements 1 and 2, observe that, for a stream 3 vehicle to be able to depart from the stop line, the following three conditions must apply simultaneously:

1. There is an adequate time gap before the arrival of the next stream 1 vehicle.
2. There is no queue of stream 2 vehicles waiting to turn into the minor road.
3. There is an adequate time gap before the arrival of the next stream 2 vehicle.

For random arrivals, the probabilities of these conditions applying at any time are the following:

$$\Pr(\text{Condition 1 applies}) = e^{-q_1 T_{(3)}} \quad 6.3$$

$$\Pr(\text{Condition 2 applies}) = P_{0(2)} = 1 - \frac{q_2}{C_2} \quad 6.4$$

$$\Pr(\text{Condition 3 applies}) = e^{-q_2 T_{(3)}} \quad 6.5$$

where

q_2 = the volume of traffic stream 2

$T_{(3)}$ = the critical gap for stream 3 vehicles entering the intersection

$P_{0(2)}$ the probability of the stream 2 queue being empty.

Assuming independence of the above three conditions, the probability that all will apply simultaneously is the product of the individual probabilities in Equations 6.3 to 6.5, that is:

$$\begin{aligned}\Pr(\text{stream 3 vehicle can depart}) &= e^{-(q_1+q_2)T_{(3)}} . P_{0(2)} \quad 6.6\\ &= e^{-(q_1+q_2)T_{(3)}} . e^{\ln P_{0(2)}}\\ &= e^{-\left[q_1+q_2-\frac{\ln P_{0(2)}}{T_{(3)}}\right]T_{(3)}}\end{aligned}$$

This probability is the same as the probability of a headway greater than or equal to $T_{(3)}$ in a traffic stream with random arrival of vehicles and volume qa, where:

$$q_a = q_1 + q_2 - \frac{\ln P_{0(2)}}{T_{(3)}} \quad 6.7$$

The situation can therefore be considered equivalent to a simple gap acceptance situation in which a minor traffic stream with volume q_3, critical gap $T_{(3)}$ and follow-up headway $T_{0(3)}$ must give way to a single major traffic stream of volume q_a, as defined by Equation 6.7. Thus, for example, the theoretical absorption capacity for stream 3 would be calculated as:

$$C_3 = \frac{q_a e^{-q_a T_{(3)}}}{1 - e^{-q_a T_{0(3)}}} \quad 6.8$$

The validity of the results obtained in the above example analysis has yet to be proven and this would probably be best accomplished by simulation of the assumed situation, over a range of traffic volumes and gap acceptance parameters. It is presented here as an illustration of how the combination of gap acceptance and queuing theory may be used to address more complex problems.

7. Vehicle Interactions in Moving Traffic

7.1. Overview

This section explores some of the theory developed to model interactions between vehicles within a moving stream of traffic. The nature of these interactions and the relative importance of each type depend most strongly on the characteristics of the facility on which the traffic is operating and on the traffic density, but other factors, such as the mix of vehicle types, also have an influence.

For example, on a rural, two-lane, two-way road where overtaking opportunities are limited, modelling of the formation of vehicle bunches and of the overtaking behaviour of drivers may be of primary concern. In contrast, on a multi-lane, urban freeway, theory related to flow breakdown and the propagation of shock waves through the traffic stream is likely to be most relevant to flow management.

The discussions in the following sections are presented as overviews of the general nature of different aspects of the theory, rather than as comprehensive coverage. For the reader who wishes to investigate any aspect more deeply, references are provided.

7.2. Car Following

Models describing how one vehicle follows another were first developed in the 1950s with the pioneering work of Reuschel (1950a and b) and Pipes (1953). This inspired a considerable volume of further work that continued into the 1960s in Japan (Kometani and Sasaki 1961) and in the United States (Chandler et al. 1958; Forbes et al. 1958; Forbes 1963; Herman et al. 1959; Herman and Potts1961; Herman and Rothery 1962, 1965) and established the basis for gradual continued development since.

7.2.1. Pipes' Model

The earliest theories are illustrated by the model developed by Pipes (1953), which simply stated that the minimum safe distance gap between a following vehicle and the vehicle ahead of it should be a constant multiple of the speed of the following vehicle. For ease of presentation of this model, the vehicle ahead is denoted as vehicle n and the one following as vehicle n+1.

As the minimum safe spacing, s_{min}, is the sum of the minimum safe distance gap plus the length, L_n, of the vehicle ahead, this led to the model:

$$S_{min} = K\,[V_{n+1}(t)] + L_n \qquad 7.1$$

where

K = a constant (in time units)

$v_i(t)$ = the speed of vehicle i at time t

The value of the constant K was derived from the common road safety advice to 'leave a clear gap of one car length between your vehicle and the vehicle ahead for each 10 miles per hour of your speed'. In metric units, an effective car length of 6 m and the equivalence of 10 miles per hour to 16 km/h or 4.44 m/s leads to a value K = 1.35 s and the model can be stated as:

$$S_{min} = 1.35\,[V_{n+1}(t)] + 6 \qquad 7.2$$

In practice, the value of the constant K was determined by field observation of minimum spacings and it was found that reasonable agreement between predicted and observed spacings was obtained for speeds between 20 and 60 km/h, using a value of K = 1.36 s in the model.

7.2.2. Forbes' Model

Forbes et al. (1958) produced exactly the same form of model as did Pipes by considering that the time gap between the front of the following vehicle and the rear of the vehicle in front should never be less than the following driver's reaction time, Δt. This meant that the minimum (time) headway, h_{min}, should be equal to the reaction time plus the time taken for the vehicle in front to travel its own length, that is:

$$h_{min} = \Delta t + \frac{L_n}{v_n(t)} \quad 7.3$$

Given that the general relationship between headway, spacing and speed is s = h.v (see Section 2.2.1), this can be written as:

$$S_{min} = \Delta t . v_n(t) + L_n \quad 7.4$$

which is the same as Pipes' model, with the constant K equal to the reaction time Δt.

7.2.3. General Motors' Models

The researchers associated with General Motors in the 1950s and 1960s produced a series of models (Chandler et al. 1958; Herman et al. 1959; Herman and Potts1961; Herman and Rothery 1962, 1965), all of which had the behavioural form:

$$\text{response} = f\ (\text{sensitivity}, \text{stimulus}) \quad 7.5$$

Over a number of years, a series of five models was developed, in each of which the response was considered to be the acceleration (deceleration if negative) of the following vehicle and the stimulus the relative speed between the leading and following vehicles. The differences between the models lay in the form of the sensitivity component, which was gradually developed to better explain and represent the results of concurrent field observations. All of the models included a reaction time Δt, by assuming that the stimulus (relative speed) at time t led to a response (acceleration) that occurred at time $t + \Delta t$.

The first model proposed a constant sensitivity term which was simply applied as a multiplicative factor to the relative speed, that is:

$$a_{n+1}(t + \Delta t) = \alpha_1 . [v_n(t) - v_{n+1}(t)] \quad 7.6$$

where

$a_i(t)$ = the acceleration of vehicle i at time t

α_1 = the constant sensitivity factor for Model 1

$v_i(t)$ = the speed of vehicle i at time t

The second and third models recognised the influence on sensitivity of the spacing between the vehicles – the second by using separate sensitivity factors for small and large spacings, which proved impractical, and the third by use of a sensitivity inversely proportional to the spacing, which produced the formulation:

$$a_{n+1}(t+\Delta t) = \frac{\alpha_3}{x_n(t) - x_{n+1}(t)} . [v_n(t) - v_{n+1}(t)] \quad 7.7$$

where

$\alpha 3$ = the constant factor for Model 3

xi(t) = the position of (the front of) vehicle i at time t

and all other terms are as previously defined.

The fourth model recognised that, just as sensitivity increased with decreasing spacing between the leading and following vehicles, it also increased with increasing speed. It therefore proposed a sensitivity factor directly proportional to the speed of the following vehicle and inversely proportional to the spacing, that is:

$$a_{n+1}(t+\Delta t) = \frac{\alpha_4 [v_{n+1}(t)]}{[x_n(t) - x_{n+1}(t)]} . [v_n(t) - v_{n+1}(t)] \quad 7.8$$

where

α_4 = the constant factor for Model 4

and all other terms are as previously defined.

Finally, a fifth, generalised model (of which Models 1 to 4 are special cases) was proposed. In this fifth model, the speed in the numerator and the spacing in the denominator of the sensitivity term were raised to exponents m and l respectively, giving:

$$a_{n+1}(t+\Delta t) = \frac{\alpha_5 [v_{n+1}(t)]^m}{[x_n(t) - x_{n+1}(t)]^l} . [v_n(t) - v_{n+1}(t)] \quad 7.9$$

where

m and l are (logically positive) exponents of speed and spacing, respectively, in the sensitivity term and all other terms are as previously defined.

Finally, it is noted that the independently proposed, macroscopic, Greenberg model of traffic flow, linking speed and density (Gazis et al. 1961) can be derived directly from the third General Motors microscopic model of car following. The Greenberg model can be stated as:

$$v = v_0 \ln\left(\frac{k_j}{k}\right) \qquad 7.10$$

where

v = speed of the traffic stream at density k

v_0 = 'optimum speed' of the traffic stream, i.e. speed at maximum volume

k_j = 'jam density' of the traffic stream, i.e. density when vehicles are bumper to bumper and stopped

k = density of the traffic stream at speed v

The derivation is not shown here but can be seen in May (1990), pp. 172–3.

Since the 1960s, there has been a steady further development of car following theory, examples of the outputs from which can be seen in Ahn et al. (2004), Aron (1988), Aycin and Benekohal (1998), Brackstone and McDonald (1999), Chakroborty and Kikuchi (1999), Dijker et al. (1998), Henn (1997), Hidas (1997), Lay (1998), Mehmood et al. (2003), Newell (2002), Ossen et al. (2006), Panwai and Dia (2005), Wang et al. (2005), and a wide range of other publications.

7.3. Traffic Bunches and Overtaking

7.3.1. General

While the terms 'bunch' and 'platoon', referring to a group of vehicles in a traffic stream, often are used interchangeably, their most common usage in Australasia (and that employed in this Guide) distinguishes between them as follows.

A bunch is a single-lane group of vehicles, associated with uninterrupted flow conditions and arising because of different desired travel speeds of different drivers and limitations on overtaking opportunities. Generally, a bunch consists of a lead vehicle, travelling at its own desired speed, closely followed by none, one or more other vehicles with equal or higher desired speeds. Under this definition, a bunch may consist of a single vehicle. The formation of a bunch of two or more vehicles is due to causes entirely internal to the traffic stream. The most common manifestation of traffic bunches is in traffic flow on two-lane two-way rural highways.

A platoon, on the other hand, may be a single-lane or multi-lane group of vehicles and is associated with interrupted flow conditions. A platoon is formed when traffic is stopped by an element external to the traffic stream, such as a red aspect at a signalised intersection. When the traffic is able to again proceed (the signal turns green) it moves away as a group or cluster of vehicles that is called a platoon.

This section is concerned with traffic bunches and the closely associated activity of overtaking, particularly where the overtaking manoeuvre makes use of the lane in which the opposing direction of traffic has priority. Traffic platoons are addressed in Section 7.4.

7.3.2. Traffic Bunches

The formation of traffic bunches has been addressed briefly in Section 3.3.4, which deals with composite models of headway distributions in a traffic stream. At that point, the above distinction between bunches and platoons was not made, as composite headway distribution models may be applied to both situations, but it is now appropriate to focus attention on bunches as defined above.

Section 3.3.4 noted that a composite headway distribution model may be defined by specifying a sufficient number of items from a list of characteristics of the traffic stream. One such sufficient set of characteristics is:

- the distribution of bunch sizes
- the distribution of headways for restrained vehicles
- the distribution of inter-bunch headways (i.e. the difference between the times at which the lead vehicles of two consecutive bunches pass a given point).

Bunch sizes

If the distribution of bunch sizes is known, the mean bunch size, m, also is known and it is possible to calculate θ, the proportion of vehicles in the traffic stream which are restrained (in the sense of being following vehicles in a bunch of size two or more) as:

$$\theta = \frac{\mathrm{m} - 1}{\mathrm{m}} \qquad 7.11$$

This is readily seen by considering a time period during which the number of vehicles passing a given point in the direction of interest is N. Given an average bunch size of m, these vehicles would be in N/m bunches, each consisting of exactly one free flowing vehicle, either alone or closely followed by other vehicles in a multi-vehicle bunch. Thus, there would be N/m free flowing vehicles and, hence, N – N/m vehicles following in bunches. The proportion of vehicles following in bunches would therefore be $\theta = \frac{N-N/m}{N} = 1 - \frac{1}{m} = \frac{m-1}{m}$.

A number of discrete probability distributions have been employed to represent the distribution of bunch sizes. Among these are the geometric distribution (Walpole et al. 2011), the Borel-Tanner distribution (Tanner 1961) and the two-parameter and one-parameter Miller distributions (Miller 1961).

The geometric distribution was introduced as a discrete distribution relevant to traffic theory in Section 3.2.3. In the context of its application to the distribution of bunch sizes in a traffic stream, it predicts the probability of observing a bunch of size n as:

$$\Pr(\mathrm{n}) = \theta^{\mathrm{n}-1}(1 - \theta) \qquad 7.12$$

Where θ is the proportion of vehicles in the traffic stream that are following in bunches, as previously defined. Given that θ can also be considered as the probability that any one vehicle in the traffic stream is a ‘restrained’ or ‘following’ vehicle, the right hand side of Equation 7.12 is seen to be the product of the probability that the lead vehicle of the observed bunch will be followed, first, by n–1 restrained vehicles (making up a bunch of size n), then by a free flowing vehicle (the lead vehicle of the next bunch).

The mean of the geometric distribution gives the average or expected bunch size as:

$$\mathrm{m} = \frac{1}{1 - \theta} \qquad 7.13$$

which is seen to be consistent with Equation 7.11.

The Borel-Tanner distribution (Tanner 1961) has already been discussed in Section 3.3.4 in the context of its application to bunch size distributions and is not considered further here.

Miller (1961) originally proposed a two-parameter model of bunch size distributions in which the probability of observing a bunch of exactly n vehicles is:

$$\Pr(n) = \frac{(a+1)(a+b+1)!\,(b+n-1)!}{b!\,(a+b+n+1)!} \qquad 7.14$$

where a and b are the two parameters, having values that provide the best fit of Equation 7.14 to observed data, while being related to the mean bunch size, m, by:

$$m = \frac{a+b+1}{a} \qquad 7.15$$

The probability of observing a single-vehicle bunch is then given by:

$$\Pr(1) = \frac{a+1}{a+b+2} \qquad 7.16$$

Miller (1961) then noted that the value of b often was small and proposed a one-parameter form of the distribution:

$$\Pr(n) = \frac{(a+1)(a+1)!\,(n-1)!}{(a+n+1)!} \qquad 7.17$$

with the parameter a fitted such that the mean bunch size, m, is given by:

$$m = \frac{a+1}{a} \qquad 7.18$$

Headway distributions within bunches

Given the relatively close following behind the lead vehicle in a multi-vehicle bunch and the fact that a minimum feasible headway must apply, a within-bunch headway distribution that provides for a small variance about a mean headway is generally appropriate. For some analyses, it may be sufficiently accurate to assume the extreme case of a uniform distribution, that is, equality of all within-bunch headways.

A more realistic representation would be provided by some other type of distribution (normal, log-normal or gamma, for example) with small variance and with limits that recognise both the minimum headway feasible for car-following and the maximum headway for which a vehicle can be considered to be following within a bunch rather than free flowing. However, the use of such a more realistic representation should be justified by the level of accuracy required by the application.

Headway distributions between bunches

Given that the headway for any vehicle is taken to be its time separation from the vehicle immediately ahead of it (see the start of Section 3.3) and that a single free flowing vehicle is taken to be a bunch of size 1 (Section 3.3.4), the headway for any free flowing vehicle is its time separation from the last (possibly the only) vehicle in the bunch immediately ahead. This same time separation is defined as the inter-bunch headway, so that the distribution of headways for free-flowing vehicles is the same characteristic as the distribution of headways between bunches.

Because these vehicles are free flowing, a headway distribution based on random arrivals (e.g. negative exponential, displaced negative exponential) is usually appropriate. For example, the discussion of the double exponential distribution in Section 3.3.4 notes that it utilises a standard negative exponential distribution of headways for free flowing vehicles (combining them with a displaced negative exponential distribution of headways for vehicles following in bunches).

7.3.3. Overtaking on Two-lane, Two-way Roads

The process of overtaking on a two-lane, two-way road, in which the overtaking vehicle utilises the opposing traffic lane, is one of the most complex tasks undertaken by drivers but is central to the operational efficiency of such roads. There have therefore been substantial quantities of both theoretical and empirical research on overtaking.

For traffic travelling in one direction on a two-lane, two-way road, the demand for overtaking generally arises because of the different desired travel speeds of different vehicles, so that vehicles catch up with others whose desired speeds are less and must then either reduce their speed to that of the vehicle ahead or overtake it. If this is the sole generator of overtaking demand, then the demand for overtaking would be the rate at which faster vehicles catch up with slower ones. Wardrop (1952) showed that if each driver has a constant desired speed and if desired speeds are normally distributed across drivers, this 'catch-up rate' is given by:

$$D_O = \frac{1}{\sqrt{\pi}} \frac{q^2 \sigma}{v_m^2} \qquad 7.19$$

where

D_0 = overtaking demand (overtakings/km/h)

q = traffic flow in the direction of travel (veh/h)

v_m = mean desired speed for traffic in this direction (km/h)

σ = standard deviation of distribution of desired speeds (km/h)

More recent research (Bar-Gera and Shinar 2005) indicates that the assumption that each driver has a fixed desired speed is an over-simplification. Rather, it appears that each driver's desired speed spans a range and the driver considers any vehicle ahead, travelling within that speed range, as a potential interference. This conclusion was drawn partly from the fact that some drivers were observed to substantially increase their own speed in order to overtake a vehicle that had been travelling a little faster than themselves. Nevertheless, Equation 7.19 is still considered a reasonable estimate of 'latent' demand for overtaking.

The level of overtaking demand in Equation 7.19 could be realised in practice only if drivers were always free to overtake whenever they wished. This is not the case on a two-lane, two-way road unless there is no traffic at all in the opposing direction and every driver is always able to be certain of that being so. Typically, however, there will be opposing traffic and a driver will be able to overtake only when:

- There is a sufficiently large gap in the opposing flow.
- The driving situation (particularly sight distance) is such that the driver wishing to overtake can be confident that there is a sufficiently large gap.

The supply of opportunities for overtaking can thus be viewed as an application of gap acceptance analogous to the unsignalised intersection applications discussed in Sections 5 and 6. In both, drivers are seeking suitable gaps in a conflicting traffic stream but, in the overtaking situation, those waiting for gaps, and the queues that form as a result, are moving. Also, there is evidence that overtaking drivers judge the size of available gaps by distance, rather than by their time duration (McLean 1989).

For traffic under steady state conditions in one direction on a length of two-lane, two-way road, equilibrium develops between demand for and supply of overtaking opportunities. Where overtaking opportunities are limited, faster vehicles tend to catch up to and queue behind slower vehicles. With increased queuing (bunching), the variance of speeds in the stream is reduced and, consequently, so is the catch-up rate and the demand for overtaking. Conversely, where there are more opportunities for overtaking, its occurrence will reduce the bunching, hence increasing the variance of speeds, the catch-up rate and the demand for overtaking. Consideration of this equilibrium led to a number of bunching/overtaking models (Kallberg 1981; Miller 1961, 1962, 1963) which are discussed in McLean (1989).

A second equilibrium exists between the bunching/overtaking occurring in two interacting streams of opposing traffic. Gipps (1974 and 1976) proposed a two-stream equilibrium model for the situation in which there were no external constraints (such as limited sight distance) on overtaking by vehicles in either of the opposing streams. He argued that the extent of bunching (bunch sizes, gaps between bunches) in the primary stream depends on the overtaking opportunities available, which depend on the gaps, and hence the bunching, in the opposing stream; further, a reciprocal relationship exists between opposing stream bunching and primary stream bunching. Building on the formulations of Miller (1961, 1962, 1963), Gipps (1974 and 1976) developed functions relating the mean bunch sizes in the two directions and demonstrated that more than one equilibrium state could exist for the same traffic conditions.

7.3.4. Bunching and Overtaking as Level of Service Measures

Hoban (1984a and b) suggested that the extent of traffic bunching occurring on a two-lane, two-way road is perhaps the best measure of level of service for the road, arguing that it is:

- Easy to measure and meaningful to drivers, engineers and road planners
- Applicable to both long and short road sections and takes account of upstream and downstream effects of particular road features
- Sensitive to the demand for overtaking as well as its supply.

Hoban (1984c) used traffic simulation to derive bunching criteria (proportion of vehicles following in bunches, proportion of journey time spent following and equivalent mean bunch size) that corresponded with the limiting traffic conditions for levels of service A to E, as defined by the 1965 Highway Capacity Manual (HRB) 1965.

Morrall and Werner (1990) noted that the 1985 Highway Capacity Manual (Transportation Research Board (TRB) 1985) had responded to recommendations such as Hoban's by introducing average percentage of time that vehicles are delayed while following in platoons ('bunches' in Australasian usage) because of their inability to pass ('overtake' in Australasian usage). Morrall and Werner (1990) used the TRARR simulation model (Hoban et al. 1985) to investigate the effects of road geometry and traffic characteristics on overtaking and per cent following. They then compared various level of service measures and procedures, including those in the 1965 and 1985 Highway Capacity Manuals, the per cent following count generated by simulation and the overtaking ratio (defined as the ratio of the achieved number of overtakings on a two-lane, two-way highway to the total number possible if the road had the same alignment but included continuous passing lanes). On the basis of their findings, they proposed that the overtaking ratio should be introduced as an additional measure of level of service to supplement those in the Highway Capacity Manual (TRB 1985).

Today, the level of service measures for two-lane, two-way roads included in the Highway Capacity Manual (TRB 2010) comprise per cent time spent following and average travel speed.

7.4. Platoon Dispersion

7.4.1. General

As noted in Section 7.3.1, a platoon, as used by Australasian traffic professionals, is a group of vehicles resulting from an interruption to traffic flow, such as the signal aspect turning red at a signalised intersection or crossing. When the green signal aspect appears, the vehicles that had been stopped move off in a group that we call a platoon.

If all the vehicles travelled at the same speed, the platoon would stay together as a tight group, as it moved along the road. In practice, however, there will be a distribution of speeds and the platoon will spread out, with faster vehicles moving further and further ahead of the centre of the group and slower vehicles dropping further and further behind. This spreading process is known as platoon dispersion and it is of interest because of its effect on efficiency of traffic flow, particularly in relation to coordination of traffic signals along a route or across a network. If dispersion did not occur and if downstream traffic signals were appropriately set, it would be possible for a much greater proportion of the platoon to move through the downstream signals without stopping than would be the case with the platoon dispersed.

Because of its importance to traffic management and control, there has been considerable research into platoon dispersion – see, for example, the historical review in Denney (1989). These studies have identified three principal ways of representing the dispersion process: in terms of kinematic wave theory, using diffusion theory and through recurrence modelling. These different approaches are discussed briefly in the following three sections.

7.4.2. Kinematic Wave Theory

Seddon (1971) extensively analysed platoon dispersion in terms of kinematic wave theory (Lighthill and Witham 1955), but found this approach unworkable because of its computational complexity and its inability to predict platoon behaviour once the platoon disperses to the point that vehicles are not interacting. Because of these reasons, Denney (1989) noted that kinematice wave theory for platoon dispersion has not been used in practical applications.

As mentioned, Leo and Pretty (1992) was successful in modelling the dispersion of platoons between to traffic signals using finite element method at a small discrete levels of time and space (Section 2.5).

7.4.3. Diffusion Theory

Pacey (1956) presented a simple, purely kinematic model of platoon dispersion, which considered the situation of normally distributed speeds in the traffic stream and argued that, in this case, the dispersion of a traffic platoon could be described by the dispersion in speeds. In developing the theory, Pacey also made the following simplifying (and somewhat unrealistic) assumptions:

- each vehicle travels at a constant speed
- vehicle speed is independent of position in the platoon
- there are no impediments to any one vehicle overtaking another.

In spite of these questionable assumptions, Pacey (1956) showed that the theory was quite successful in predicting flow profiles.

A presentation and extension of the theory by Grace and Potts (1964) starts with Pacey's assumptions, including that of normally distributed speeds, with mean m and standard deviation σ. It is further assumed that, at the start of the signal green phase (time t = 0), the traffic density k(x,0) is known, where x represents the position along the road from the signal stop line. The authors then extend the theory by noting that, with a change of variables, the traffic density at location x and time t can be obtained as the solution to the standard one-dimensional diffusion equation:

$$\frac{\partial k}{\partial t} + m\frac{\partial k}{\partial x} = \alpha^2 m^2 t\frac{\partial^2 k}{\partial x^2} \qquad 7.20$$

with the diffusion constant α being equal to σ/m, the coefficient of variation of the distribution of vehicle speeds in the platoon.

Seddon (1972a) presents the diffusion theory in terms of the traffic flows at the stop line and at a downstream location (where the level of dispersion is of interest) and of the vehicle travel times between those two points. Seddon notes that, if the distribution of vehicle speeds is assumed to be normal, it is possible to derive the distribution of travel times $g(\tau)d\tau$ between the two points. If the flow past the first point in the time interval t to t+dt is $q_1(t)dt$ then, of these vehicles, there will be $q_1(t)g(\tau)dtd\tau$ that will pass the downstream point at time $T = t + \tau$. Overall, therefore, the flow past the downstream point in the time interval T to T+dT will be:

$$q_2(T)dT = \int q_1(t)g(T-t)dtdT \qquad 7.21$$

the integration being over all values of t for which $q_1(t)$ exceeds zero.

In practical applications a more convenient expression of this dispersion relationship may be the discrete form:

$$q_2(j) = \sum_i q_1(i)g(j-i) \qquad 7.22$$

Where i and j are counts of discrete intervals of time at the first and second points respectively. In words, this expression says that the flow in the j^{th} interval at the second point is the sum, over all values of i, of the flow in the i^{th} interval at the first point multiplied by the probability of a travel time between the two points of j–i intervals.

Seddon (1972a) fitted the model to field data by selecting values of m and σ, the mean and standard deviation of platoon speeds to derive the travel time function $g(\tau)$ giving the best fit between observed and predicted arrival patter ns at the downstream point. He found the model then produced good predictions of arrival patterns at other downstream points. Denney (1989) and the references already noted in this section provide further detail for the interested reader.

7.4.4. Recurrence Model

Using data collected by others (Hillier and Rothery 1967), Robertson (1969) developed an empirical platoon dispersion model using a discrete iterative technique. This **recurrence model** has received wide application in the various versions of the TRANSYT network optimisation package, which takes account of platoon dispersion in selecting signal offsets to minimise total delay over a road network.

The recurrence model considers traffic flows over a series of equal, small time intervals at two locations on the road, x_1 (say, the stop line at a signalised intersection) and x_2 (a point some distance downstream from x_1 where the level of dispersion is of interest), relating flows at the two locations in different time intervals. Specifically, the recurrence relationship can be written as:

$$q_2(i+t) = F.q_1(i) + (1-F).q_2(i+t-1) \quad 7.23$$

where

$q_1(n)$ = the flow at location x_1 in time interval n

$q_2(n)$ = the flow at location x_2 in time interval n

i = a time interval counter

t = $\beta \cdot T$ where β = an empirical parameter < 1

and T = average undelayed travel time (cruise time) x_1 to x_2 (in time intervals)

with $\beta \cdot T$ representing the travel time of the head of a platoon

F = smoothing factor $= \frac{1}{1+\alpha t}$

where α = an empirical parameter

Robertson (1969) fitted this model to field data finding that values of $\alpha = 0.5$ and $\beta = 0.8$ gave the best predictions of flows at downstream points.

Seddon (1972b) shows that Equation 7.23 can be written equivalently as:

$$q_2(j) = \sum_{i=1}^{j-t} q_1(i) F(1-F)^{j-t-i} \quad 7.24$$

where j is a count of time intervals at the second point and j=i+t ; all other variables are as previously defined.

Seddon (1972b) points out the similarity between Equation 7.24 and Pacey's diffusion model as expressed in Equation 7.22, the only difference being the replacement of the travel time function $g(j-i)$ with the probability function $F(1-F)^{j-t-i}$, which will be recognised as a geometric distribution commonly used to represent the number of failures in a two-outcome trial before the first success. Here it is the probability that a vehicle passing the first point in the i^{th} interval will pass the second point in the j^{th} interval. Seddon (1972b) found good fits of the recurrence model to field data, but with values of the calibration parameters a little different from those suggested by Robertson (1969).

7.5. Congestion Management Theory

7.5.1. General

The purpose of this section is to overview aspects of traffic theory that form the basis of flow management on road facilities. This area of traffic management has been given increased attention by transport professionals since the 1980s, particularly in relation to freeways and other high standard facilities (Akcelik et al. 1999; Brilon 2000; May 1990).

Much of the theory underlying these developing approaches to flow management is not new but draws on the established relationships of traffic flow. However, using different ways of viewing key traffic characteristics, particularly density, researchers and traffic managers have been able to gain new insights into freeway performance and guidance on how that performance can be improved.

7.5.2. Flow Monitoring and Management

Density and occupancy

The three primary variables used to describe traffic flow are identified in Section 2 as volume (q), density (k) and speed (v), which, in aggregate terms, are related by q = k.v (Equation 2.3), in which the appropriate v is the space mean speed.

It has long been recognised that density is a fundamental measure of the level of service (LOS) being provided on a road at any particular time (e.g. HRB 1965) but, until relatively recently, the difficulties of field measurement of density led to the use of other LOS measures such as volume/capacity ratio. Historically, density (the number of vehicles in a unit length of lane or road) has been measured in the field by one of four methods, as follows:

- **Photographic techniques** measure density directly using photographs along a length of road, taken either from a fixed, high vantage point or from an aircraft. From the photographs, the number of vehicles in each length of road or lane that is of interest are counted and the density is obtained by dividing by the known length of road or lane.
- **Input-output counts** enable the number of vehicles in a road section to be updated from an initial, known number by adding counts of vehicles entering the section and subtracting counts of vehicles leaving. The passage detectors must be able to ensure accurate counts at both ends of the section and a means of regularly re-initialising the number of vehicles within the section is desirable. Such re-initialisation is difficult except in the situation of road sections with no intermediate entry or exit points and no lane changing, in which case the number of vehicles in each lane of the section can be obtained as the count of vehicles entering between the entry and exit of a specifically identified vehicle.
- **Speed-flow calculations** use point measurements of vehicles passing and individual vehicle speeds to calculate volume and space mean speed (the latter by means of Equation 2.5), then apply Equation 2.3 to determine density.
- **Occupancy measurement** became a viable means of determining density with the introduction of accurate presence detectors. Occupancy (at a given location over a (usually fairly short) period of time) is defined as the proportion of time for which the presence of a vehicle over the detector is recorded. Given that presence is recorded whenever any part of a vehicle length is over any part of the effective length of the detector, occupancy is related to the average spacing of vehicles by the relationship:

$$\mathrm{Occ} = \frac{\overline{\mathrm{L_V}} + \mathrm{L_D}}{\mathrm{s}} \qquad 7.25$$

where

Occ = occupancy, expressed as a proportion (veh.s/s)

$\overline{L_V}$ = average length of a vehicle (m)

$\mathrm{L_D}$ = effective length of detector (m)

s = average spacing of vehicles as defined in Section 2.1.5 (m/veh)

Then, given that density, k, is inversely related to spacing (see Equation 2.2) but is usually expressed in the units of veh/km rather than veh/m, density is obtained as:

$$k = \frac{1000}{s} = \frac{1000.\,Occ}{\overline{L_V} + L_D} = \frac{10.\,(\%Occ)}{\overline{L_V} + L_D} \qquad 7.26$$

Where k is density in veh/km, %Occ is occupancy expressed as a percentage, and all other variables are as previously defined.

Traffic density and per cent occupancy ranges corresponding to different levels of service are shown in Table 7.1 which has been adapted from May (1990) and Transportation Research Board (2010).

Table 7.1: Density and occupancy level of service indicators

Density pc/km/lane (1)	Per cent occupancy (2)	Level of service	Flow conditions	
0–7	0–5	A	Free flow operations	Uncongested flow conditions
7–11	5–7	B	Reasonably free flow operations	
11–16	7–11	C	Stable operations	
16–22	11–14	D	Bordering on unstable operations	
22–28	14–19	E	Extremely unstable flow operations	Near capacity flow conditions
28–54	19–36	F	Forced or breakdown operations	Congested flow conditions
> 54	> 36		Incident situation operations	

1 *Density in passenger car equivalents per kilometre per lane.*

2 *Assuming* $\overline{L_V} + L_D = 6.7m$.

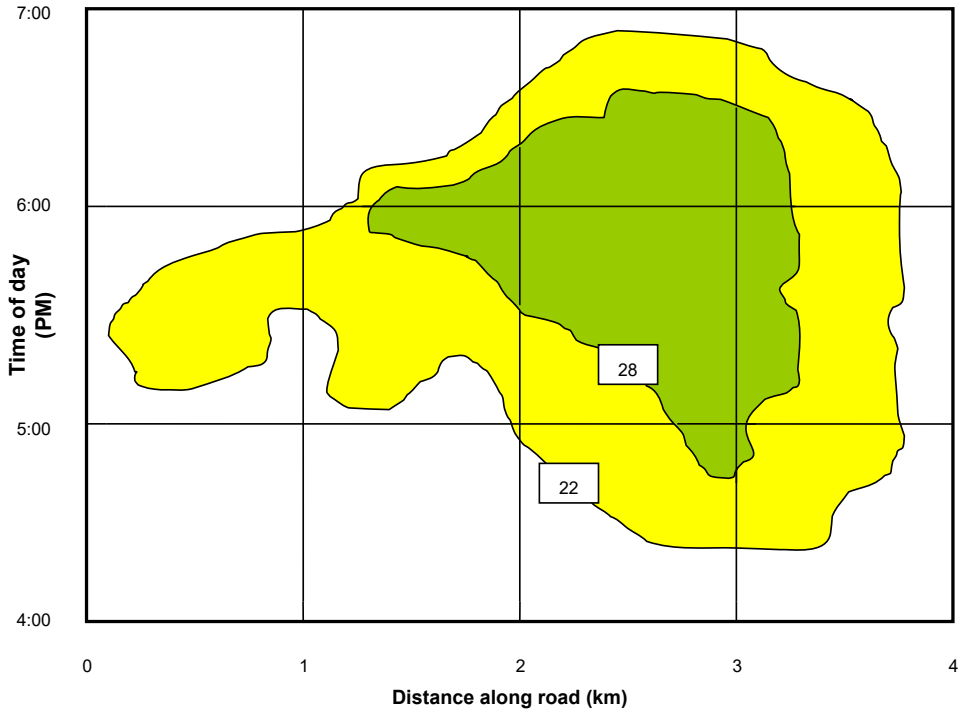
Density contour mapping

With the advent of reliable and accurate presence detectors, traffic densities derived from measured lane occupancies have become key indicators of traffic performance. Freeway managers, in particular, are using changes in density values over time at key locations on their facilities to monitor performance and provide guidance on interventions to improve performance.

The plotting of contours (lines of equal value) of density on a location-time plane has evolved as the most useful presentation of density data (and, similarly, data on other characteristics such as flow, speed, moving queues and gap availability) for the purposes of flow management.

A hypothetical example is shown in Figure 7.1, in which the contours separate areas of uncongested flow (k < 22 pc/km/lane, no shading), near capacity flow (22 < k < 28, light shading) and congested flow (k > 28, heavy shading), the values being consistent with Table 7.1. Some comments on this example are made following a brief introduction to shock waves in traffic.

Figure 7.1: Traffic density contours in the space-time domain



Shock waves in traffic

Shock waves are defined as 'boundary conditions in the space-time domain that denote a discontinuity in flow-density conditions' (May 1990). For flow management purposes, the boundary of most interest is that marking the discontinuity between uncongested and congested flow, which might correspond with a density of 28 pc/km/lane.

A shock wave is not simply a contour corresponding to a particular density value, however – it is an indication of a sudden change in flow conditions which may propagate through time and space. A simple example of a shock wave is free-flowing traffic forced to join the tail of a queue stopped by a red traffic signal. The discontinuity that occurs is the change from free-flowing traffic with medium density and speed to the jam conditions of a stationary queue – maximum density and zero speed. As time increases with the signal remaining red, the propagation of the wave is backwards in space (i.e. opposite to the travel direction) because, as the queue length increases, the point at which the change of flow conditions occurs moves further and further back. Such a shock wave would be classified as 'backward forming' – **backward** because it moves backward in space over time and **forming** because a greater extent of congestion is forming.

A change from red to green of the traffic signal in the preceding paragraph would result in a 'backward recovery' shock wave as vehicles further and further back in the queue successively start to move – **recovery** because the change is a decrease in the extent of the congestion and **backward** because the point at which the change is occurring progressively moves backward relative to the direction of travel.

As well as **backward**, shock waves may be **forward** or **stationary**, these designations indicating forward movement and lack of movement, respectively, of the shock wave over time. 'Forward forming' shock waves are relatively rare but 'forward recovery' shock waves occur frequently, for example, where a bottleneck has caused the formation of a slowly moving queue but, as the peak period ends and upstream demand decreases to less than the bottleneck capacity, the tail of the queue moves forward and the extent of congestion decreases.

Stationary shock waves are classified as either **frontal**, because they are at the front of the congested length of road (so that the change of flow conditions is from congested upstream of the shock wave location to uncongested downstream) or **rear**, because they are at the rear of the congestion (so that the reverse is the case). As a stationary shock wave does not move, it results in no change in the extent of congestion. A frontal stationary shock wave might occur, for example, at the downstream end of a bottleneck, while a rear stationary shock wave may be located upstream of the bottleneck when demand decreases to become equal to the bottleneck capacity, so that the length of the queue to enter the bottleneck does not change.

Flow management applications

The above discussion of shock waves is introductory only but, with a thorough knowledge of the topic, density contour maps can be interpreted to provide valuable insights into traffic performance, including identification of the real locations of bottlenecks, assessment of the consequences of congestion (in terms of delays and economic consequences) and the ability to distinguish between recurring and incident-generated congestion. While the process of interpretation is complex enough to require considerable experience, a little of what is involved can be appreciated by returning to the density contour map in Figure 7.1.

Considering the $k = 28$ contour, the boundary between uncongested and congested flow, the map first indicates the presence of a bottleneck a little beyond the 3 km location, which first causes a flow breakdown at approximately 4:45 pm, setting up a frontal stationary shockwave. The second effect of this bottleneck is a backward forming shockwave which takes approximately 75 minutes (from 4:45 to 6:00 pm) to move back 1.7 km to around the 1.3 km location. The velocity of this shock wave is thus -1.4 km/h (negative because it is moving backward). At 5:10 pm a bottleneck at location 3.3 km receives sufficient traffic to cause a second frontal stationary shock wave, which replaces that at 3.0 km and lasts at that location through until 6:30 pm. At around 6:00 pm, the approaching end of the peak results in reduction in upstream demand and the congestion begins to decrease, as indicated by the forward recovery shock wave from 1.3 km at 6:00 pm to 2.6 km at around 6:35 pm (shock wave velocity +2.2 km/h). The remainder of the peak congestion is reduced between 6:25 and 6:35 pm by the backward recovery shock wave moving back 0.7 km from the 3.3 km location to the 2.6 km location. The congestion ends at 6:35 pm.

Through similar interpretations of data, including that collected over longer periods of days or weeks, freeway managers are able to optimise the operation of ramp metering, assess the effects of lane additions and generally guide the design of interventions to improve flow performance. For further information, the reader is directed to the references in Section 7.5.1 and reports from Australasian road agencies actively involved in the area.

The following sections provide further discussions on some previously published traffic models that deal with the characteristics of flow breakdowns. These traffic state models classify traffic flow into different state regimes and provide different ways to illustrate congested freeway flow patterns (Austroads 2008; Han & Luk 2008).

7.5.3. Two Phase HCM Model

A conventional understanding of the formation of congested flow conditions is that a queue would form upstream of a bottleneck due to conditions such as lane drop, merge area, weaving section or upgrade. The trailing edge of the queue moves upstream at a rate depending on demand and capacity conditions. When the tail of this queue reaches any upstream location, freeway operation moves from the uncongested regime to the congested regime, at approximately the same flow.

The HCM 2010 (TRB 2010) and the 1986 and 2000 editions have advocated the need to consider maximum flows or capacities of a freeway segment in two regimes or phases. Two maximum flow rates can be identified as follows:

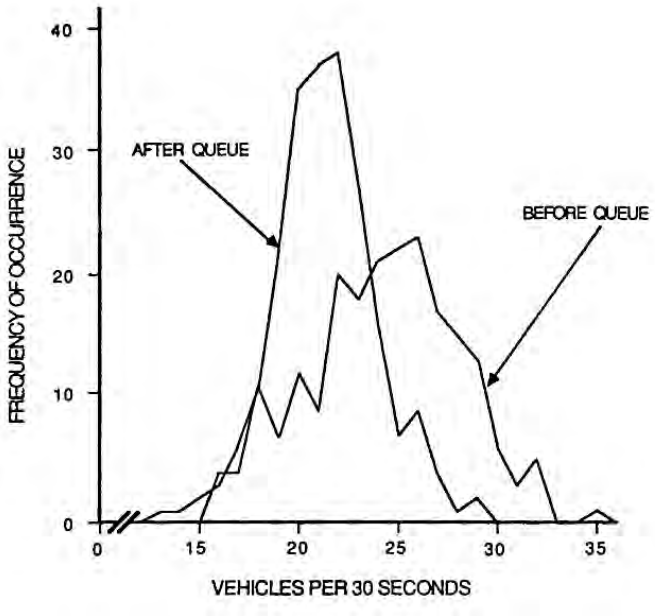
- Maximum flow when flow is stable – this is the maximum flow before the formation of a queue at a bottleneck, i.e. the maximum *pre-queue* flow.

- Maximum queue discharge flow – this is the maximum flow after a queue is formed and is associated with a speed drop and has been found to be less than the pre-queue maximum flow rate. A possible reason for this decrease in flow rate is *driver caution* – departures from a freeway queue require more care because drivers may not be aware of conditions downstream. This contrasts with a start-up queue at a signalised approach where maximum flow is achieved even though different vehicles have different acceleration rates.

There have been debates on where the maximum flows should be measured. Hall and Agyemang-Duah (1991) argued that the two phases are observable only if detectors are located at some distance *upstream* of a bottleneck, and that there is only one congested regime if they are *at a bottleneck*.

In a study of a bottleneck on a four-lane freeway near San Diego (Interstate 8), Banks (1990) measured the above two maximum flow rates. The frequency distribution polygons of the counts on the fast lane are shown in Figure 7.2. The results clearly showed that there is a statistically significant difference between the two flow rates.

Figure 7.2: Frequency distribution polygons of vehicle counts on the fast lane



Source: Banks (1990).

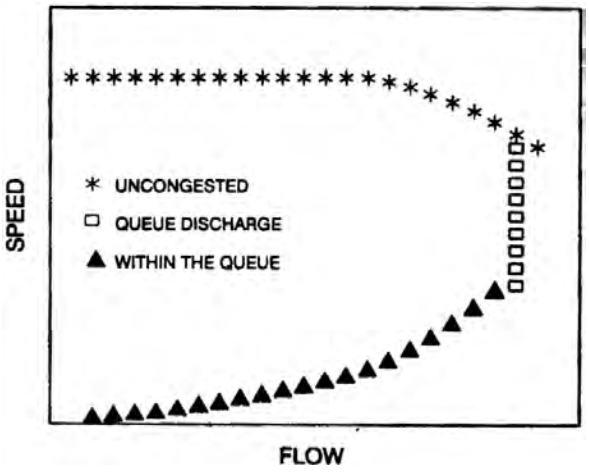
Hall, Hurdle and Banks (1992) finalised a speed-flow diagram as shown in Figure 7.3. The diagram overcomes the following issues:

- The parabolic shape in uncongested flow is no longer used; speed remains quite similar until the degree of saturation or volume/capacity ratio reaches 0.75.
- The queue discharge regime is included in the speed-flow diagram.
- Two maximum flow rates are used, one for the stable, pre-queue regime and another for the queue discharge rate (which is lower than the maximum pre-queue flow rate).

As mentioned, ramp metering is useful for reducing on-ramp flow so that the mainline demand is maintained at or just below capacity and therefore reduces the occurrences of flow breakdowns and also improves traffic conditions at the merge point.

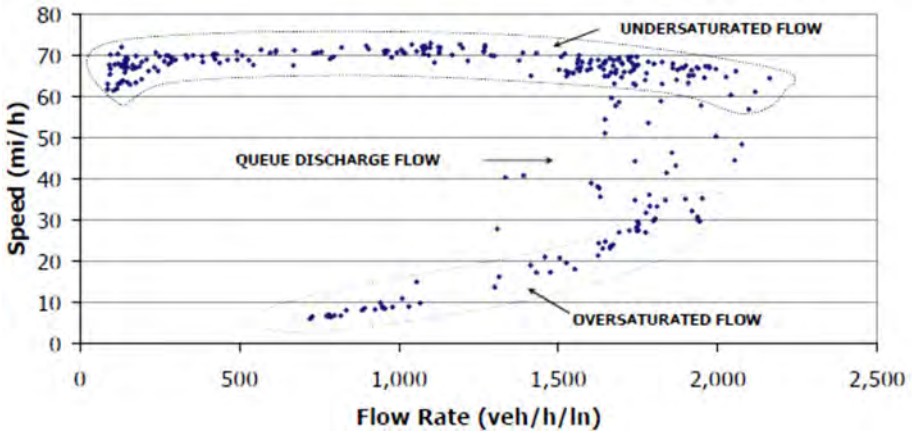
Hall, Hurdle and Banks (1992) also suggested that much more research is needed in understanding freeway congested flow. Figure 7.4 shows the speed-flow relationship for a freeway in the HCM (2010). Note that the queue discharge area covers a range of data. The queue discharge area may be represented by a vertical segment, as shown in Figure 7.3, recognising that the vertical segment is not really a speed-flow function, but is plotted on the graph without the location axis.

Figure 7.3: Generalised speed flow relation for a typical freeway segment



Source: Austroads (2008).

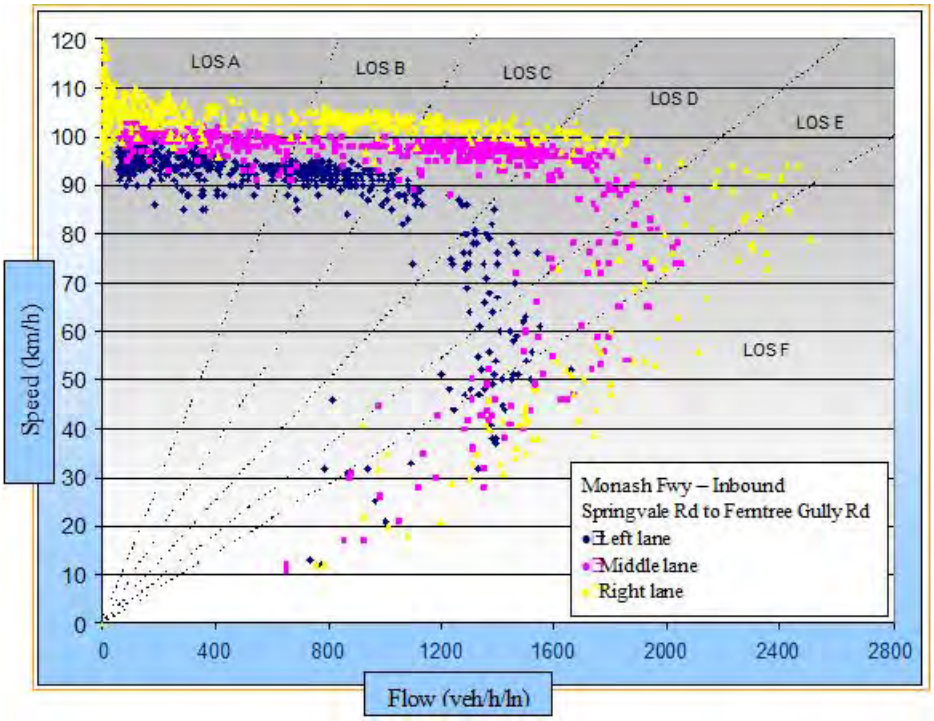
Figure 7.4: Speed-flow relationship for freeway in HCM (2010)



Source: Adapted from TRB (2010).

Figure 7.5 shows an example of a speed-flow relationship based on empirical data for each lane across a three-lane carriageway in Melbourne (VicRoads 2013). The flow breakdown at this location was initiated by an uncontrolled flow at an entry ramp merge. In the context of level of service (LOS), the flow breakdown generally occurs within LOS E where unstable flow leads to problems in sustaining the free flowing conditions. The changes to speed and flow rate are accompanied by increases in motorway lane occupancies (density values within LOS F).

Figure 7.5: An example of speed-flow diagram from a Melbourne freeway



Source: VicRoads (2013).

7.5.4. Three Phase Model

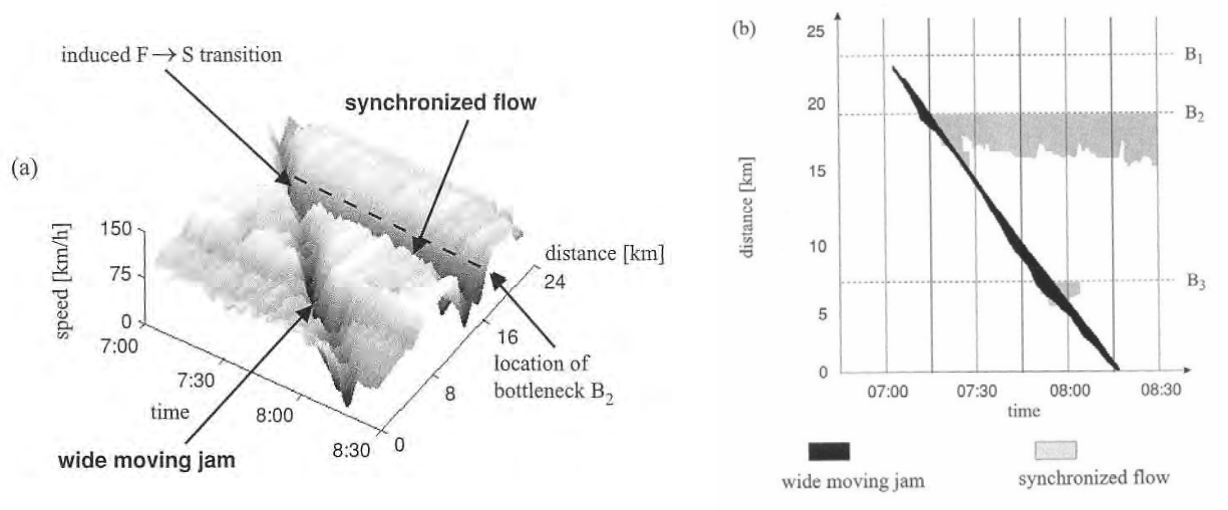
Three traffic phases

Kerner and Rehborn (1996) first proposed the classification of freeway traffic flow into three phases based on time series of flow, occupancy, and average speed. Kerner (2004) later completed the three-phase traffic theory based on earlier work. In the three-phase traffic theory, there are two traffic phases in congested traffic, *synchronised flow* and *wide moving jam,* defined as follows:

- A synchronised flow is a congested traffic state and the downstream front of this flow is often fixed at a freeway bottleneck. Within the downstream front of synchronised flow, vehicles accelerate from lower speeds in synchronised flow to higher speeds in free-flow.
- A wide moving jam is a moving jam that maintains the mean velocity of the downstream jam front, even when the jam propagates through any other traffic states or freeway bottlenecks.

The three traffic phases are therefore free-flow (F), synchronised flow (S) and wide moving jam (J). Figure 7.6 illustrates the traffic phase definition of synchronised flow and wide moving jams (Kerner 2004). The data in Figure 7.6 came from a section of Autobahn 5-South freeway near Frankfurt, Germany. There are three bottlenecks labelled as B1, B2 and B3. Average 1 min speed data in space and time is shown in (a). A two-dimensional graph of the same data with the free-flow phase in white, the synchronised flow phase in grey, and the wide moving jam phase in black is shown in (b).

Figure 7.6: Synchronised flow and wide moving jams in congested traffic

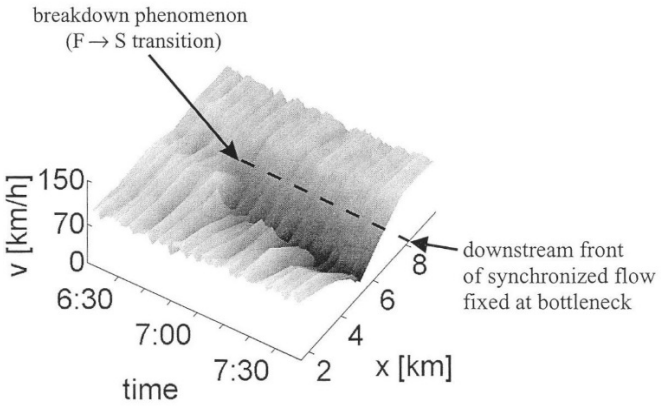


Source: Kerner (2004).

The three-phase traffic theory explains the complexity of traffic phenomena based on phase transitions among these three traffic phases. For example, transitions can be *spontaneous* F→S or *induced* F→S. In Kerner's three-phase theory, a transition from F→S is a flow breakdown (Kerner 2004; Kerner et. al. 2005).

An induced F→S transition is caused by a short-term external disturbance in traffic flow. This traffic flow can be related to the propagation of a moving spatio-temporal congested pattern that initially occurs at a different freeway location. Figure 7.6 (a) shows an example of induced F→S transition – the wide moving jam propagated through the bottleneck location B2 and induced the synchronised flow at this bottleneck. Figure 7.7 shows an example of spontaneous F→S transition. This breakdown phenomenon or F→S transition is caused by an internal local disturbance (e.g. an on-ramp bottleneck) in traffic flow. There are no external disturbances in traffic flow responsible for this phase transition.

Figure 7.7: An example of spontaneous F→S transition



Source: Kerner (2004).

The F→S transition or breakdown phenomenon usually occurs at the same freeway bottleneck. These bottlenecks are called *effectual* bottlenecks in Kerner's model. Examples of effectual bottlenecks are the bottleneck in Figure 7.7 and B1, B2, and B3 in Figure 7.6.

Based on different combinations of traffic phases, different congested patterns are formed. Kerner studied traffic flow on the A5 freeway over a large number of days and found that the spatio-temporal structure of congestion patterns exhibits predictable features. These features can be used to forecast freeway congestion and develop effective freeway control tools.

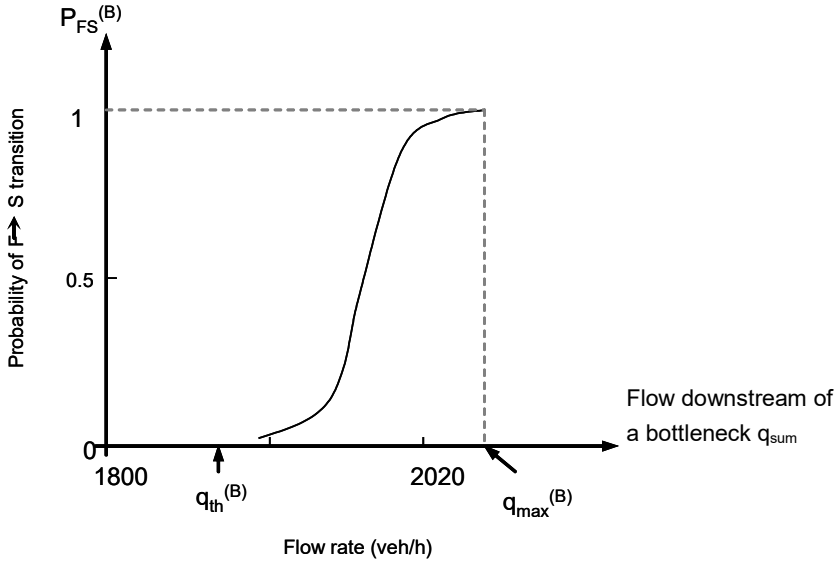
Lindgren (2005) also investigated a 30 km section of A5 freeway north of Frankfurt and found some similar traffic patterns that match Kerner's three traffic phases. In Lindgren's A5 freeway study, traffic flows were observed in which speeds across all lanes were notably lower than in free-flowing conditions, and they were more consistent across all lanes. This phenomenon was observed in congested flows upstream of the bottleneck following activation. This pattern matched Kerner's synchronised flow phase. Lindgren also revealed several occurrences of congested patterns in which a relatively short duration traffic disturbance travelled several kilometres upstream. This pattern matched Kerner's wide moving jam. Lindgren's study represented some of the first apparent independent validation of Kerner's traffic phase findings (Lindgren 2005; Lindgren et. al. 2006).

Empirical probabilistic nature of traffic breakdown

Kerner (2004 and 2007) found that the traffic breakdown exhibits a probabilistic nature. At a given flow rate, traffic breakdown at a freeway bottleneck can occur but it may not necessarily occur.

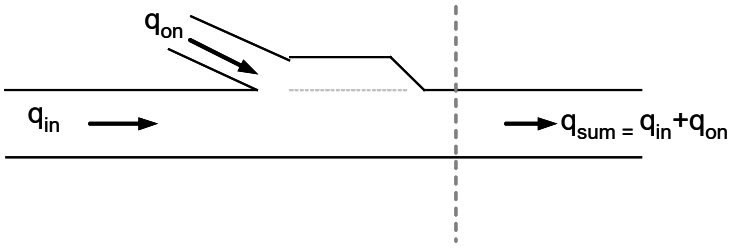
The probability for an F→S transition, i.e. a traffic breakdown, ($P_{FS}^{(B)}$) at a bottleneck is an increasing function of the flow *downstream* of the bottleneck q_{sum} as shown in Figure 7.8. q_{sum} is the sum of the flow on the on-ramp q_{on} and mainline upstream flow q_{in} (Figure 7.9). There is a threshold flow rate $q_{th}^{(B)}$ and a critical flow rate $q_{max}^{(B)}$. Regardless of free-flow control application there is a range when $q_{th}^{(B)} \leq q_{sum} \leq q_{max}^{(B)}$ within which traffic flow breakdowns can occur with probability $P_{FS}^{(B)} > 0$.

Figure 7.8: Probability of traffic breakdown



Source: Kerner (2007).

Figure 7.9: Traffic flow downstream of a bottleneck (q_{sum})



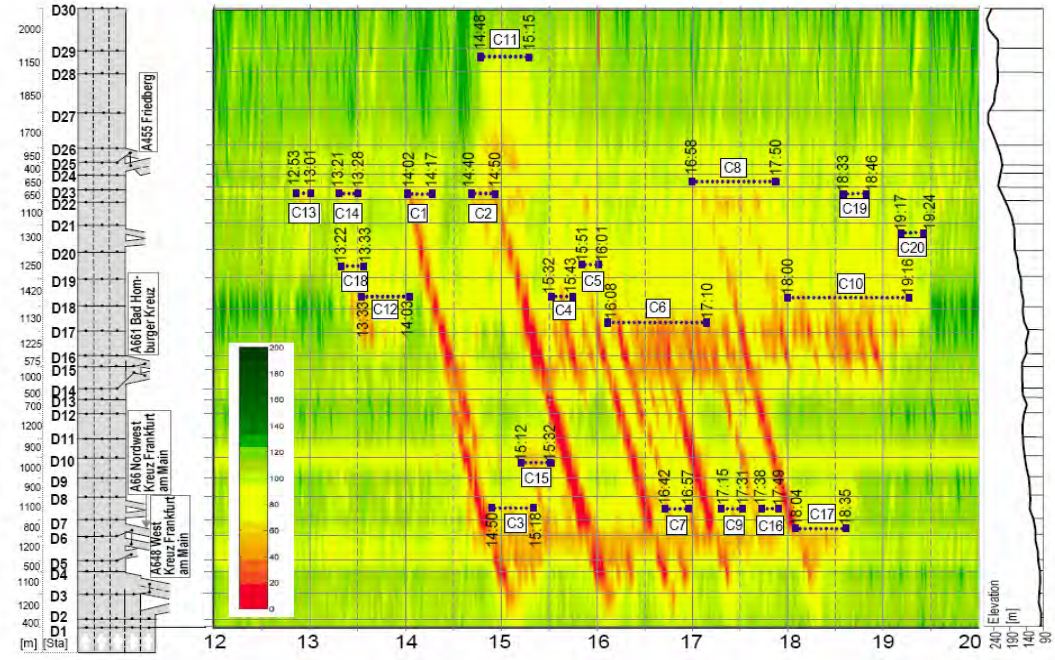
Source: Kerner (2007).

A flow breakdown, if due to a speed disturbance in free flow in the neighbourhood of a bottleneck, occurs only when the speed decreases below a critical speed. The critical speed depends on the q_{sum}. The smaller the q_{sum} , the lower the critical speed required for breakdown. The probability for traffic breakdown $P_{FS}^{(B)}$ is the probability of random critical speed disturbances appearing at the bottleneck. Disturbances with small amplitudes in free flow at the bottleneck do not lead to breakdown. However, if a random short-term speed disturbance in free flow at the bottleneck exceeds some critical values, traffic breakdown occurs.

Lindgren (2005) investigated the same section of A5 freeway traffic flows as Kerner did. Lindgren reviewed Kerner's work on three-phase models and found that Kerner's time series plots cannot show excess accumulation (queuing) between measurement locations resulting from bottleneck activation. Lindgren applied a *cumulative count curves technique* that was used to complement the three-phase models to observe transitions between free flows to queued conditions and identify time-dependent traffic features of bottlenecks. Figure 7.10 shows the bottlenecks in time and space identified by Lindgren.

Lindgren (2005) investigated 81 bottleneck activations and deactivations, where queued traffic prevailed upstream of each bottleneck and unqueued traffic was present downstream. Although Kerner suggested that traffic congestion can form and traffic can self-organise without a physical bottleneck, all 81 bottlenecks diagnosed in Lindgren's study were activated at a predictable location (e.g. merge, diverge, vertical curves) and appeared to be linked to particular triggers rather than to have occurred spontaneously.

Figure 7.10: A5 speed contour diagram in Lindgren's study – 1 min data



Source: Lindgren (2005).

7.5.5. Other Models of Flow Breakdowns

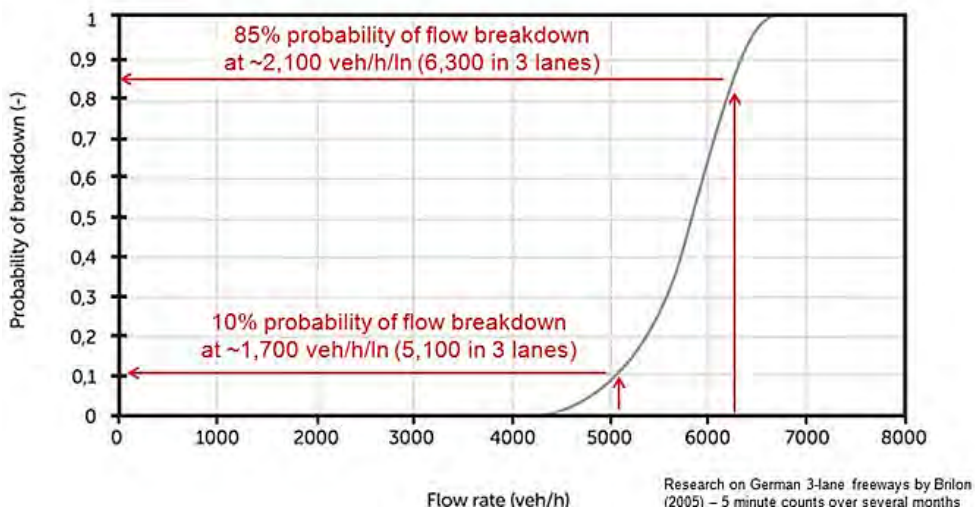
Stochastic concept of traffic capacity

Brilon, Geistefeldt and Regler (2005) studied 5 min data on the freeways around the city of Cologne, Germany and found that the concept of stochastic capacities seem to be more realistic and more useful than traditional use of single value capacity. Lorenz and Elefteriadou (2001) and Brilon and Geistefeldt (2009) also proposed that the capacity of a motorway facility is not so much deterministic, but rather a random variable and that breakdown probability can be related to traffic flow as shown in Figure 7.11. Their empirical analysis shows that the distribution of freeway capacity fits very well into a Weibull distribution (Figure 7.11). The overload probability (traffic breakdown) for a single bottleneck is equal to the capacity distribution function as shown in Figure 7.11. Examples of flow values and the likely flow breakdown probability at those flows are plotted on the graph. This finding is consistent with Kerner's analysis of the probabilistic nature of traffic breakdown (Section 7.5.4).

While it is likely that the shape of the probability curve may change for different motorways and traffic flow mix with heavy vehicles, the research can provide an indication of likely problems at different flow values. For example, at a flow of approximately 2100 veh/h/lane the curve indicates an 85% probability of flow breakdown. Similar characteristics and probabilities have been demonstrated by Main Roads Western Australia on the Mitchell Freeway in Perth.

The concept of randomness permits the demonstration of the capacity-reducing effect of wet road surfaces (–11%) and the capacity-increasing effect of traffic-adaptive variable speed limits.

Figure 7.11: Probability of flow breakdown



Source: Based on Brilon, Geistefeldt and Regler (2005).

The study by Brilon, Geistefeldt and Regler (2005) also showed that three traffic flow states exist in a freeway: fluent traffic state, congested traffic state and a transient state that occurs in each breakdown and recovery of traffic flow. These three states seem to match Kerner's three-phase theory but the definitions of the phases are slightly different.

The stochastic concept of capacity reveals that the optimum degree of saturation for a German freeway is around 90%. If the degree of saturation increases further, the risk of a breakdown becomes too high, so that the efficiency of freeway operation must be expected to be lower than a saturation of 90%.

Six traffic state model

Schonhof and Helbing (2007) investigated 1 min data for the same section on the A5 freeway as Kerner and Lindgren. They interpreted traffic flow by six states: free traffic (FT), pinned localised cluster (PLC), moving localised cluster (MLC), stop-and-go waves (SGW), oscillating congested traffic (OCT) and homogeneous congested traffic (HCT). The most frequent states at the investigated freeway were the PLC and OCT states. HCT occurs mainly after serious accidents with lane closures or during public holidays. An adaptive smoothing method was used to identify the different traffic states. This method interpolates and smoothes traffic data from successive freeway sections, taking into account the propagation speeds of perturbations in free and congested traffic.

Schonhof and Helbing (2007) found that the congested traffic states identified by this model were in good agreement with prediction of some second-order macroscopic traffic models and some microscopic car-following models.

Readers should refer to Austroads (2008), Austroads (2009a) and Han and Luk (2008) for further information regarding these flow breakdown models and their applications in the identification and analysis of freeway flow breakdowns.

8. Principles Underlying Managed Motorways

In recent years, road agencies have focussed on the management of motorways under congested flow conditions using e.g. ramp meterings and variable speed limit (VSL) signs. Austroads (2014) and VicRoads (2013) described the general principles underlying the use of these managed motorway tools.

8.1. Causes and Impacts of Flow Breakdowns

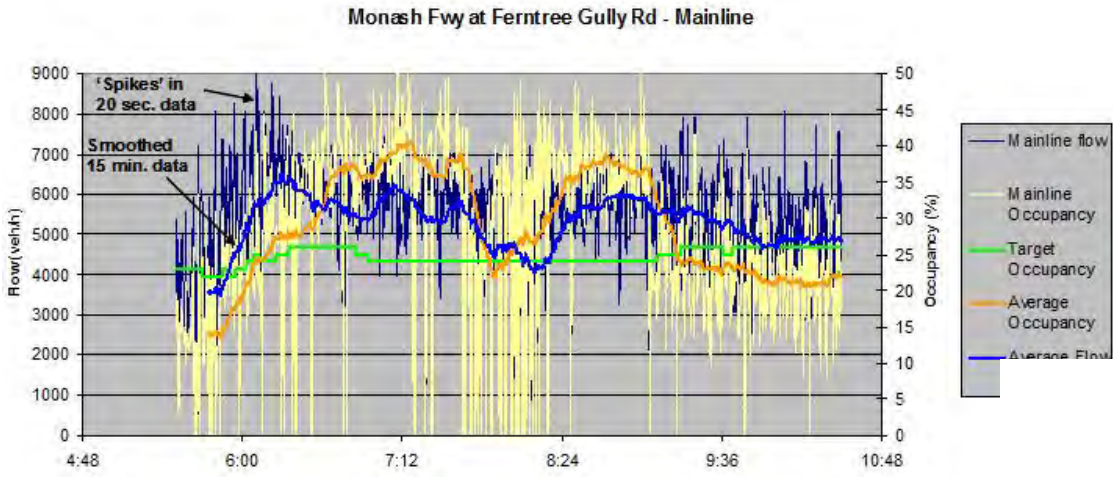
Traffic flow breakdown occurs within the section of a motorway where the flow first exceeds capacity and can be caused by recurrent or non-recurrent factors.

8.1.1. Bottlenecks

A bottleneck is a fixed location where the capacity is lower than the upstream capacity. Bottlenecks affect traffic flow capacity and have the potential to cause recurrent flow breakdown typically including:

- merging traffic from an entry ramp
- merging traffic at a lane drop, e.g. narrowing from four to three lanes
- high lane changing manoeuvres over a short distance – typically due to weaving prior to a high flow exit or prior to an increase in the number of lanes
- traffic queues at an exit ramp extending back to block the left lane of the motorway or causing traffic to slow down prior to exiting
- mainline locations where geometric features cause vehicles to slow down e.g. a steep upgrade, a tight radius curve, width restriction (real or perceived) or sight distance constraint.
- a lower speed limit
- speed differential between vehicles due to:
  - presence of trucks
  - random actions such as sudden braking following a driver's inattention
- short periods of very high density flow that are not sustainable. Examples of spikes in traffic flow are illustrated in Figure 8.1.

Figure 8.1: Example of high volume and density spikes in the traffic flow



Source: VicRoads (2013).

'Critical' bottlenecks are the locations along a section of motorway where recurrent flow breakdown usually occurs first, i.e. the location that first reaches capacity. These are typically at a lane drop or at an entry ramp merge with a combination of high mainline flow and a high entry flow. As flow breakdown at a bottleneck relates to an operational deficiency, recurring congestion is generally predictable and can be managed with appropriate control of flow. In some instances a solution may exist to correct a deficiency.

A 'potential' or 'latent' bottleneck becomes an 'active' bottleneck when flow breakdown occurs as a result of the flow exceeding capacity, i.e. the congestion is not the result of a shockwave that arrives from a downstream location.

It needs to be recognised that correction of a capacity deficiency at one location may move the point of critical capacity upstream or downstream to the next point of limiting capacity. As each point of capacity limitation is removed, the section of motorway should become more tolerant of flow variations up to the capacity limit along its length.

8.1.2. Non-recurrent Causes of Flow Breakdown

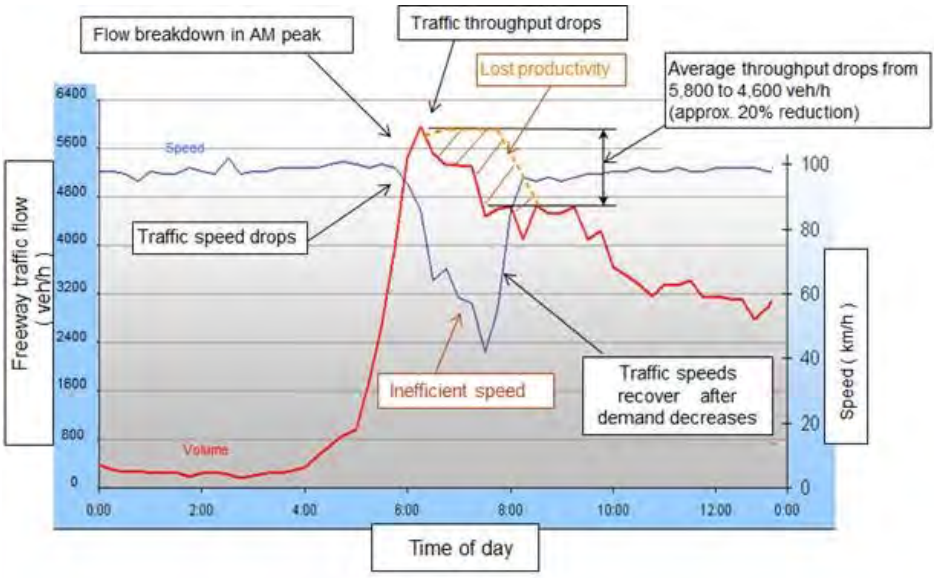
Non-recurrent traffic flow breakdown can also occur at any location on a motorway due to:

- an accident, object or other incident on the carriageway
- roadworks, including maintenance works
- driver behaviour that slows down the traffic flow such as:
  - 'rubber necking' to look at an incident
  - police presence or enforcement activity.

8.1.3. Effects of Flow Breakdown

Motorway traffic flow breakdown usually creates significant reductions in throughput and vehicle speeds and may result in substantial increases in travel time. During the period of flow breakdown, lane occupancy (density) rises as a result of reduced headway on the motorway. The reduction in throughput, which may average about 10–15%, represents under-utilisation of a high value facility and lost productivity. An example from a Melbourne freeway is shown in Figure 8.2.

Figure 8.2: Typical impacts of flow breakdown on traffic throughput and speed



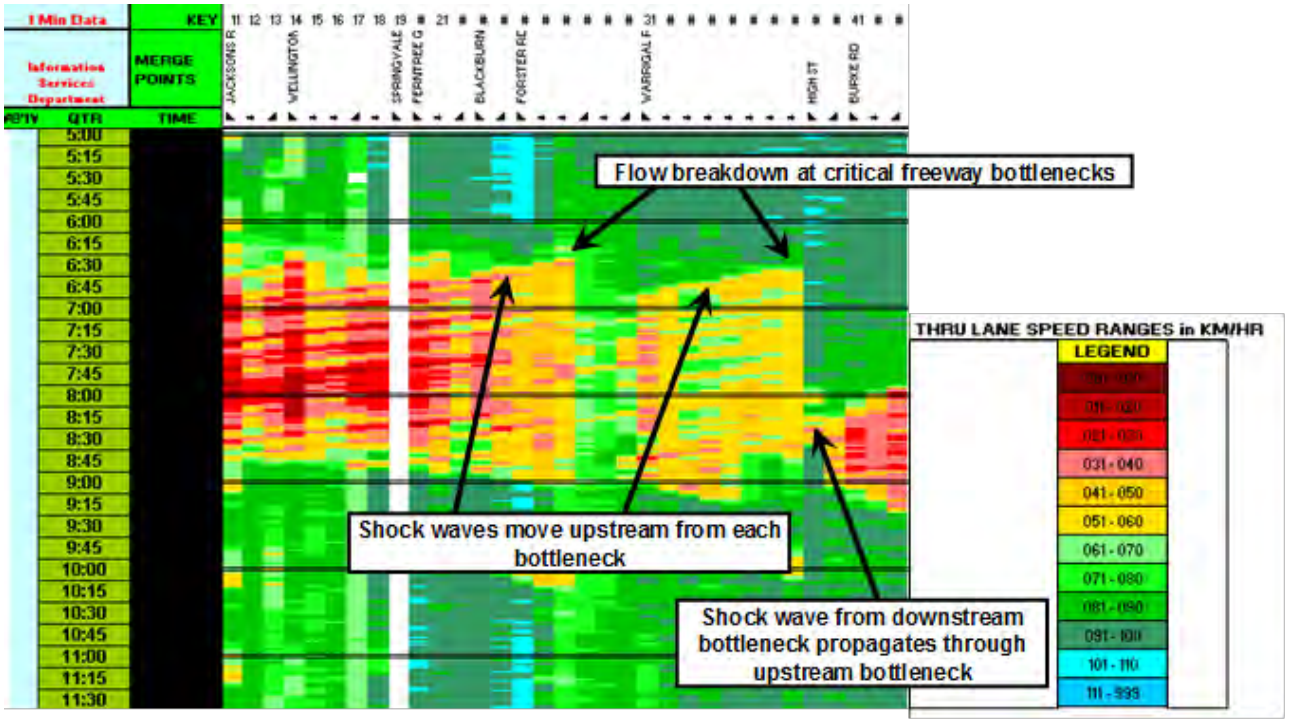
Source: VicRoads (2013).

After traffic flow breakdown occurs at a bottleneck, the congestion will result in slow speed travel at that location and loss of throughput, i.e. capacity flow is only reached for a relatively short time. The symptoms may be localised and remain at or near the bottleneck, or more usually, the congestion creates a moving queue with a shockwave that travels upstream from the initial location of flow breakdown, to affect the performance of an extended length of the motorway.

Figure 8.3 shows an example of freeway speed contour in the time and distance dimension. The flow breakdown and the resulting shockwave propagation upstream (i.e. backward shockwave) are shown by orange and red patches in the diagram. Typical characteristics within the shockwave area are:

- The lane occupancy will be high.
- Flow rates will typically be 10% to 20% lower than the maximum flow at the downstream bottleneck prior to breakdown.
- The speed will be low and variable as the shockwave moves upstream, i.e. stop-and-go waves are formed within the congested area.

Figure 8.3: Flow breakdown at bottlenecks and shockwave propagation

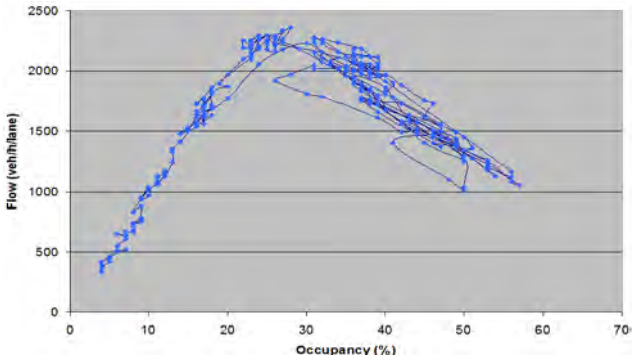


Source: VicRoads (2013).

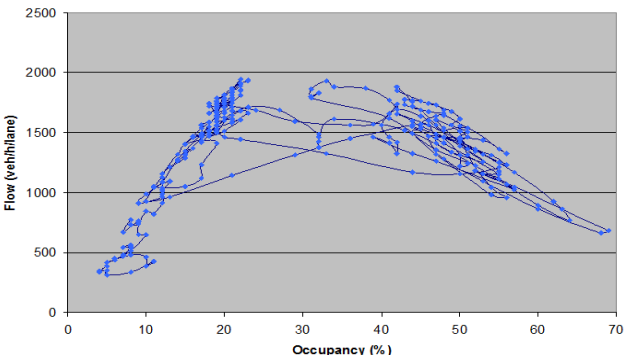
As the congestion moves in shockwaves from the point of initial flow breakdown, i.e. a critical bottleneck, the congestion at a particular upstream location may be the result of a bottleneck that is remote from the area under investigation. When investigating the cause of congestion at a particular point or when trying to identify the critical bottleneck along a length of motorway, it needs to be determined whether the data represents congestion from flow breakdown at that point or whether the congestion results from a downstream bottleneck, i.e. there is a need to differentiate between cause and symptom.

A data analysis related to the identification of a critical bottleneck location is shown in Figure 8.4. The example indicates that the flow at the bottleneck reached capacity and flow breakdown followed when demand exceeded capacity, while flows within the shockwave area were significantly lower than the capacity such that throughput is not optimised.

Figure 8.4: Flow-occupancy graphs of flow breakdown and shock wave



Flow-occupancy graph at a critical bottleneck(1)



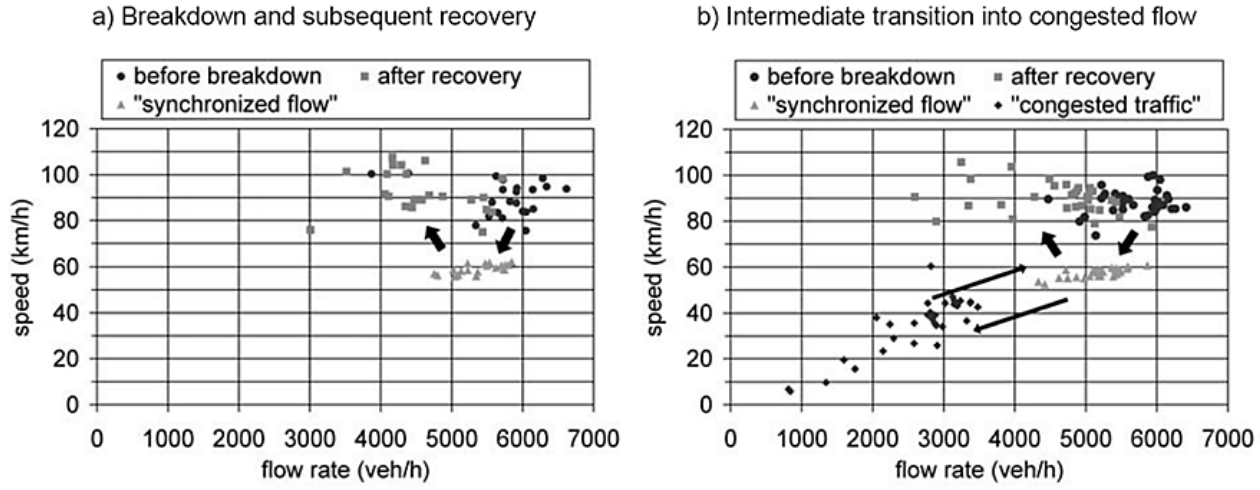
Flow-occupancy graph within the shockwave area upstream of a critical bottleneck(2)

1 Flow breakdown occurs at capacity (approx. 2200 veh/h/ln).
2 Maximum flow rate (approx. 1800 veh/h/ln) is lower than capacity.
Source: VicRoads (2013).

8.1.4. Recovery from Flow Breakdown

Whilst the mechanism for flow breakdown follows the general pattern described in the fundamental diagram, the recovery from flow breakdown follows a different phenomenon generally known as the hysteresis of traffic flow. As observed by Brilon, Geistefeldt and Regler (2005), after flow breakdown all recoveries to fluent traffic passed through synchronised flow (the transient state) and involved much lower traffic volumes than the preceding breakdown as shown in the examples in Figure 8.5.

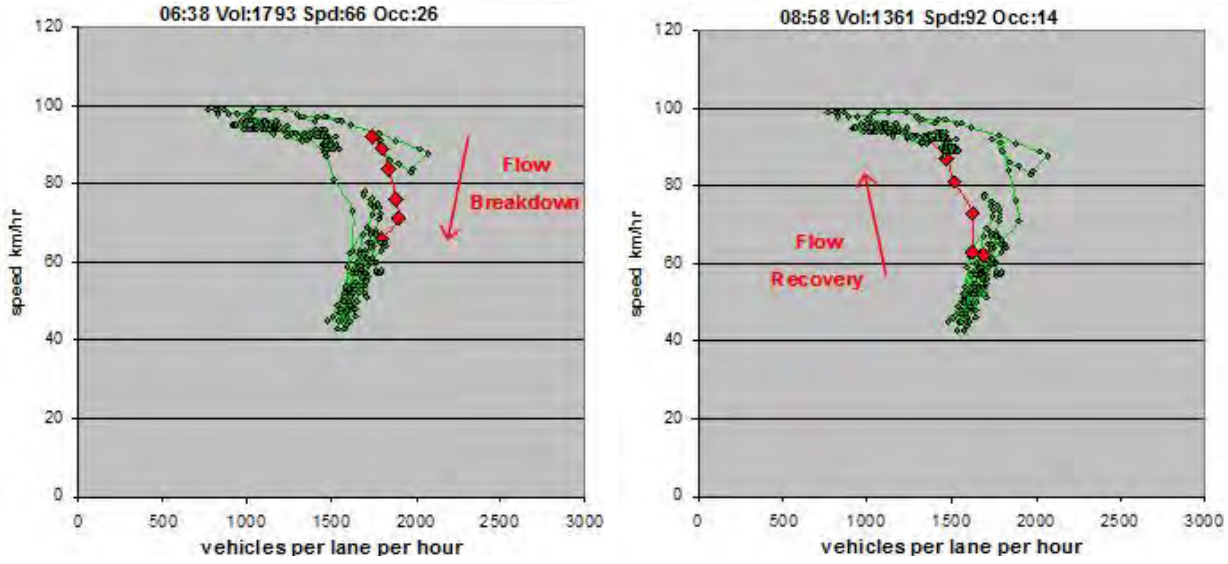
Figure 8.5: Two typical patterns of traffic dynamics during breakdown and recovery



Source: Brilon, Geistefeldt and Regler (2005).

Traffic flow breakdown and recovery observed on Melbourne's freeways exhibit similar characteristics. An example that shows the path of flow breakdown and recovery is provided in the speed-flow graphs in Figure 8.6. The lower flows on recovery illustrate that a motorway not only experiences lost productivity during flow breakdown but also throughout the flow recovery period.

Figure 8.6: Example of speed-flow diagram during flow breakdown and recovery



Flow breakdown **Flow recovery**

Source: VicRoads (2013).

8.2. Motorway Operational Capacity

8.2.1. Capacity of Segment

Capacity of a road segment, as determined for design purposes, is the maximum sustainable hourly rate at which persons or vehicles can reasonably be expected to traverse a point or uniform section of a lane or roadway during a given time period under the prevailing roadway, environmental, traffic and control conditions (TRB 2010).

For motorways, capacity could be expressed in passenger car equivalents (PCE) across all lanes. The concept of PCE is related to traffic behaviour due to the vehicle mix (i.e. presence of heavy vehicles) in the traffic flow. These factors include:

- physical space taken up by a large vehicle
- longer and more frequent gaps in front and behind heavy vehicles
- speed of vehicles in adjacent lanes and their spacing.

In motorway capacity analysis, heavy vehicles are converted into an equivalent number of passenger cars to achieve a consistent measure of flow.

In measuring the capacity, it is generally the maximum sustained 15 min flow rate, expressed in passenger cars per hour per lane (pc/h/ln), that can be accommodated by a uniform motorway segment under prevailing traffic and roadway conditions in one direction of flow. The flow rate measured over a short period is generally not sustained over a longer period. The ratio of maximum hourly volume to the maximum 15 minute flow rate expanded to an hourly volume is the peak hour factor (PHF). The PHF is a measure of traffic demand fluctuation within the peak hour and is typically up to 0.95 in high flow conditions.

Above three lanes, the capacity per lane drops with each additional lane added to a motorway. This phenomenon may be because the number of lane-changing conflict points increases with each additional lane. An additional factor may be that carriageways with four or more lanes have a greater mix of passive and aggressive drivers in their middle lanes resulting in greater uncertainty and friction within these lanes. *Austroads Guide to Traffic Management Part 3* (Austroads 2020a) provides further details on capacity of uninterrupted flow facilities.

Operational capacity is the actual real-time capacity for a road segment, which can vary depending on prevailing roadway, traffic and control conditions. These variable conditions include the percentage of heavy vehicles, driver population (passive or aggressive driving, familiar or unfamiliar with road), road geometry, road surface, time-of-day, weather and light. (Theoretical capacity for a road segment is an average capacity estimate over a period.)

Operational capacity, which can be either measured in total vehicles per hour or passenger car equivalents per hour, is particularly relevant to the control of managed motorways. For example, ramp signals maintain operational capacity while regulating inflow demand to prevent flow breakdown.

The German *Manual for the Design of Road Traffic Facilities* (RSRTA 2005) estimates mainline capacities for freeways with speed limits of 100 and 80 km/h on grades up to 2% as shown in Table 8.1.

Table 8.1: German freeway mainline capacities for two and three lane carriageways

Number of lanes	Flow	Capacity (veh/h)		
		Heavy vehicles		
		0%	10%	20%
3	Total flow	5800	5500	5200
	Flow/lane	1933	1833	1733
2	Total flow	4100	3900	3700
	Flow/lane	2050	1950	1850

Source: Based on German Manual for the Design of Road Traffic Facilities (RSRTA 2005).

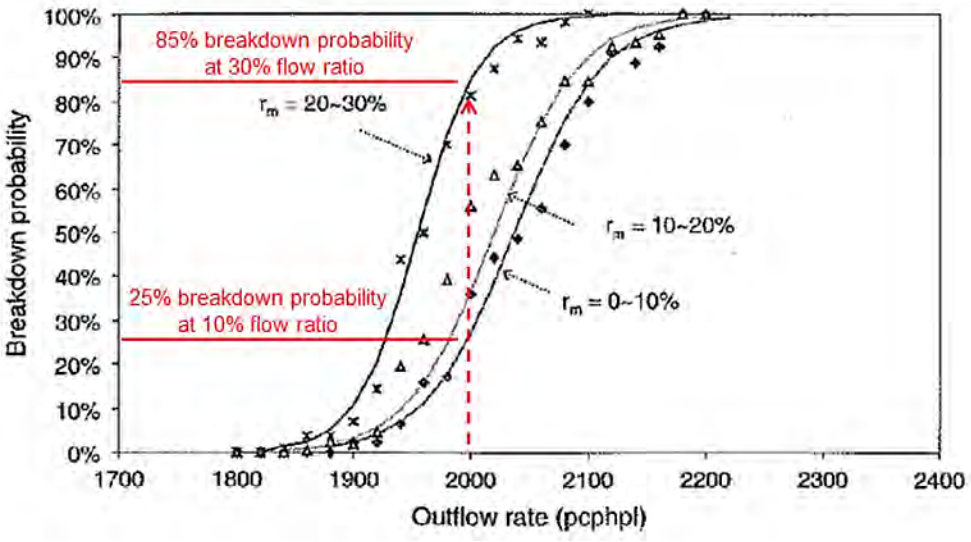
The UK *Design Manual for Roads and Bridges*, Volume 6, Section 2, Part 1, TD 22/6 (Highways Agency 2006) provides guidelines for the selection of entry ramp layouts for motorways of varying mainline and merging flows. The guidelines include a chart for the selection of an appropriate entry ramp layout. In regard to traffic flows (Chapter 3) it indicates:

> *For the purpose of designing grade-separated junctions and interchanges, the maximum flow per lane for motorways must be taken as 1800 vehicles per hour (vph). These flows do not represent the maximum hourly throughputs but flows greater than these will usually be associated with decreasing levels of service and safety.*

8.2.2. Capacity at Motorway Entry Ramp Merges

Contemporary traffic research has also provided insights in relation to the capacity of entry ramp merges. Research on Japanese freeways by Shawky and Nakamura (2007) indicated that an increasing ratio of entry ramp flow to outflow rates led to a higher breakdown probability, as shown in Figure 8.7. For example, for a flow of 2000 veh/h/ln, flow breakdown probability increased from approximately 25% with a 10% ratio of entry ramp flow to outflow, to a probability of 85% at a flow ratio of 30%.

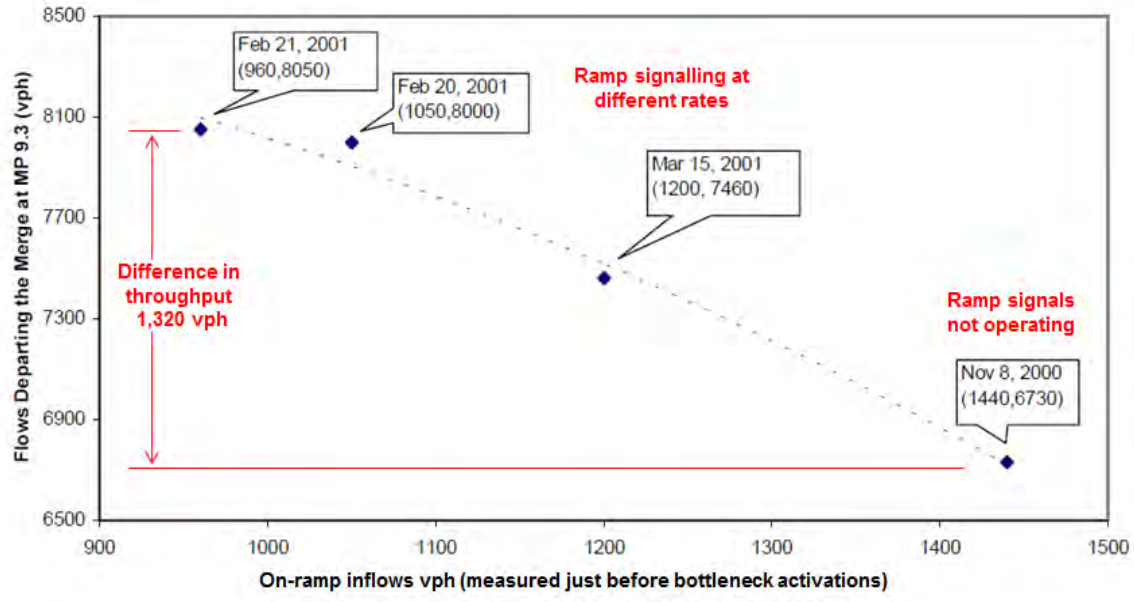
Figure 8.7: Observed and estimated breakdown probability at the Shibakoen ramp in Tokyo



Source: Shawky and Nakamura (2007).

When studying traffic at a freeway merge and the roles of ramp metering, Cassidy and Rudjanakanoknad (2002) found that as regulated entry ramp flows decrease through ramp metering, the capacity (throughput) departing the merge increases as shown in Figure 8.8.

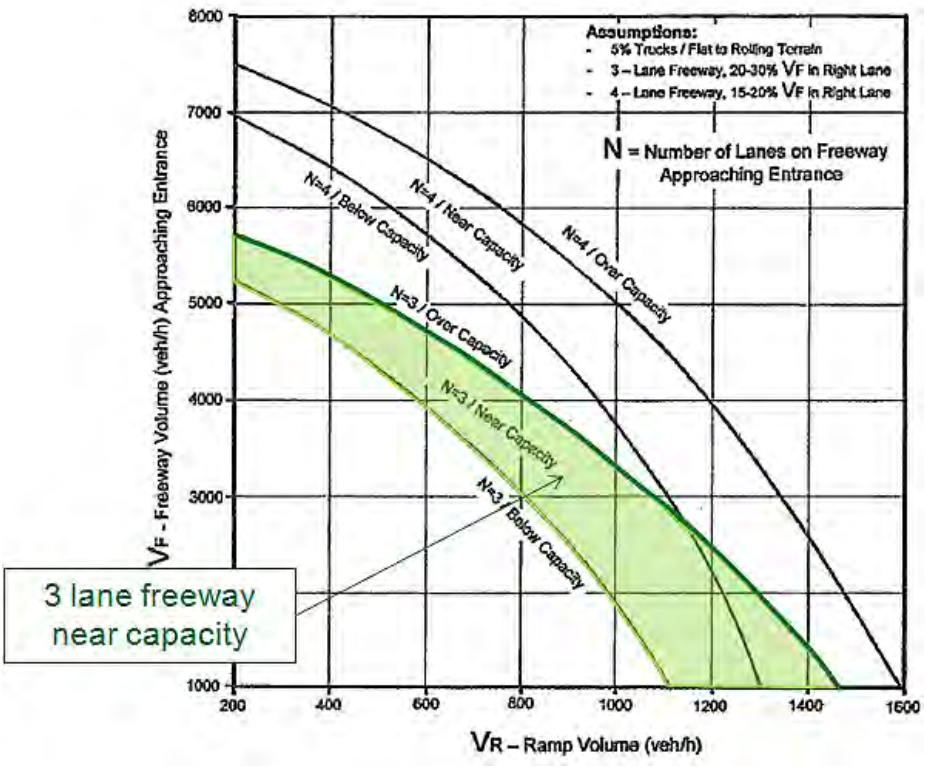
Figure 8.8: Freeway entry ramp capacity with increasing ramp flows



Source: Adapted from Cassidy and Rudjanakanoknad (2002).

The ITE *Freeway and Interchange Geometric Design Handbook* (Leisch and Mason 2006) provides guidelines for entry ramp capacity assessment. The merge capacity varies according to the upstream mainline flow and the entry ramp merging flow. For example, the mainline capacity is in the order of 2000 veh/h/ln with no entry ramp flow. The capacity reduces to approximately 1600 veh/h/ln with an entry ramp flow of 800 veh/h, i.e. approximately 20% capacity drop due to the merging traffic. The chart and summary of the entry ramp layouts for the various flows are shown in Figure 8.9 and explanations are in Table 8.2.

Figure 8.9: Entry ramp capacity assessment for single-lane merge ramps



Source: Adapted from Freeway and Interchange Geometric Design Handbook (Leisch & Mason 2006).

Table 8.2: Entry ramp capacity assessment for single-lane merge ramps

	
Procedure: (one-lane – taper or parallel) 1. Enter V_F and V_R and intersect. 2. Over, Near or Below Capacity (N = 3 or N = 4)? 3. If Near or Below – okay! 4. If Over: a. Consider single lane ramp add on freeway (capacity – 1900 v/h). b. Consider 2-lane entrance (see 2-lane procedure). c. Increase number of lanes on freeway. d. Add entrance lane as auxiliary lane. e. Consider two or more of above. f. Consider ramp metering.	**Procedure: (two-lane)** 1. Assign 60% of V_R to right lane (40% to left). 2. Auxiliary Lane added downstream. 3. Enter graph with V_F & 40% V_R and intersect. 4. Over, Near or Below Capacity (N = 3 or N = 4)? 5. If Near or Below – okay! 6. If Over: a. Consider 2-lane add ramp (capacity 3800 v/h) b. Consider ramp metering.

Notes:

Five-lane freeway subtract 22% from freeway volume approaching entrance and use N = 4.

Six-lane freeway subtract 35% from freeway volume approaching entrance and use N = 4.

Source: Based on Freeway and Interchange Geometric Design Handbook (Leisch & Mason 2006).

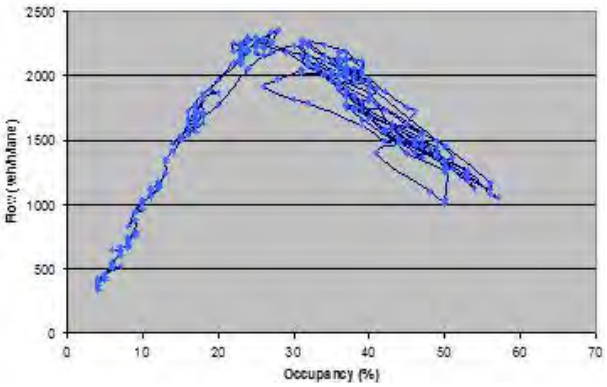
8.3. Merge Capacity for a Managed Motorway with Ramp Signals

A managed motorway with coordinated ramp signals has the ability to maintain optimal density and capacity by managing the carriageway occupancy and minimising flow breakdown.

The unmanaged flow results in flow breakdown, reduced throughput, reduced speed, congestion and lost productivity for the motorway. The managed flow results in optimum throughput and speed as the system controls and minimises the potential for flow breakdown as well as automating flow recovery when the flow nears the point of breakdown (Figure 8.10).

Figure 8.10: Example of unmanaged and managed motorway flow

Unmanaged motorway

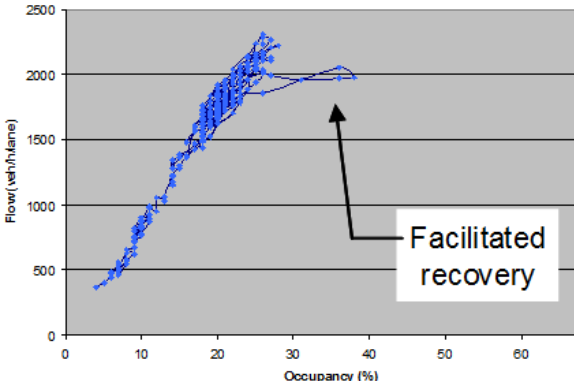


Flow breakdown occurs

Note:

- reduced throughput
- reduced speed
- congestion
- lost productivity.

Managed motorway



Flow breakdown avoided

Ramp signals with HERO control:

- prevent flow breakdown
- maintain optimum throughput
- maintain optimum speed
- facilitate flow recovery.

Source: VicRoads (2013).

While higher values can be achieved in practice, a value in the order of 2000 veh/h/ln (2100 pc/h/ln) is generally more sustainable over a range of conditions.

Optimum speeds and high capacity flow can only be achieved and maintained over a prolonged period by controlling density with coordinated ramp signals. Coordinated ramp metering will be more effective when the motorway is rid of localised geometric bottlenecks or any other localised issues causing bottlenecks, such as off-ramp overspill and short, narrow sections. Before implementing coordinated ramp metering or any other treatments to address corridor-long issues, localised treatments should first be applied to address localised congestion issues. Minor civil works or altering the signal phase timings at an exit ramp/arterial road intersection to allocate more green time to the ramp may be the first treatments applied where queue spill-back on the ramp reduces mainline capacity immediately upstream of the exit ramp.

Once localised bottlenecks have been addressed, typically, ramp merges are the critical capacity sections along a motorway. Thus, to prevent flow breakdown along a motorway corridor, it is important to maintain sufficient capacity through the ramp merges and other bottlenecks by regulating the in-flows from all ramps with coordinated ramp metering.

When designing motorway projects or upgrading existing motorways, operational capacity values should be used rather than theoretical values to gain an appropriate understanding of how the project will perform after construction and to ensure that adequate infrastructure is provided for the anticipated demands.

Based on VicRoads (2013) research and operational investigations, appropriate maximum capacity values for motorway design are:

- unmanaged motorways: 1800 pc/h/ln (typically 1700 veh/h/ln) which accepts a low risk of flow breakdown
- managed motorways: 2100 pc/h/ln (typically 2000 veh/h/ln) with well-designed infrastructure and a coordinated ramp metering system.

The above values may need to be adjusted for site-specific conditions which will impact motorway capacity, including road characteristics and vehicle mix. The ramp merges are usually the critical capacity sections which determine a motorway's maximum operational capacity.

Within the current Austroads Guides, content on ramp metering design and operation is spread across a number of parts of the Guide to Road Design and Guide to Traffic Management. Austroads (2014) provides more up-to-date best practices in ramp metering design and operation.

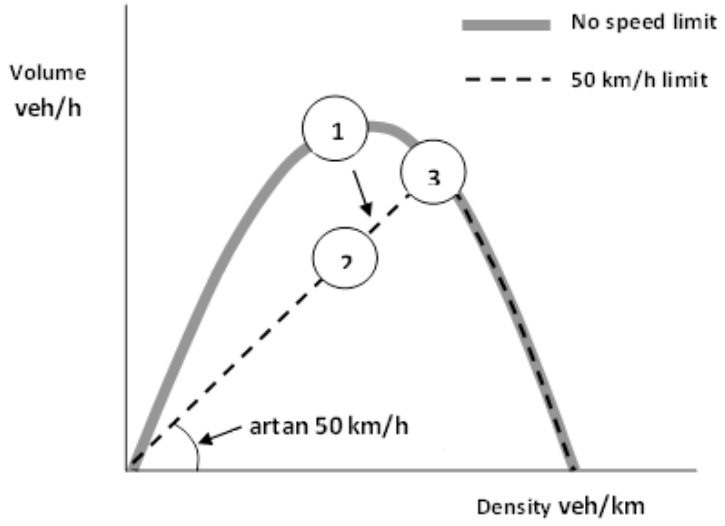
Note that ramp metering also provides other benefit such safety (Lee, Hellinga and Saccomanno 2006), driver behaviour (Wu, McDonald and Chatterjee 2007) and traffic diversion (Banks 2005). The interested reader is directed to these publications and also to the extensive work represented in their reference lists.

8.4. Theory Underlying Variable Speed Limits (VSL)

VSL are introduced to improve traffic flow efficiency, road safety or both. Hegyi, De Schutter and Hellendoorn (2005) reported two views on the use of speed limits. The first emphasises the homogenisation effect, whereas the second is more focused on the prevention of traffic breakdown. The idea of homogenisation is that speed limits reduce the speed differences between vehicles, which is expected to result in a higher (and safer) traffic flow. The approach typically uses speed limits that are close to, but above, the critical speed (the speed that corresponds to the maximal flow). However, Van den Hoogen and Smulders (1994) reported that the effect of homogenisation on freeway performance is negligible but that a positive safety effect can be expected. The traffic breakdown prevention approach focuses more on preventing high densities and it allows lower than critical speed limits. As opposed to the homogenisation approach, it can also resolve existing traffic jams.

Hegyi, De Schutter and Hellendoorn (2005) explained the mechanism of traffic breakdown prevention in terms of a change in the fundamental volume-density relationship as illustrated in Figure 8.11.

Figure 8.11: Change in volume-density relationship with speed limit reduction



Source: Hegyi, De Schutter and Hellendoorn (2005).

In the absence of a speed limit or with a high limit (say 100 km/h), the volume-density graph is the solid, gray-shaded, approximately parabolic curve shown in Figure 8.11, similar to the curve shown in Figure 2.3. Point 1 on this curve represents traffic conditions at close to capacity, where the slope of the curve is almost horizontal and flow breakdown (a switch of flow conditions to the unstable, right-hand side of the curve) is most likely to occur.

With the imposition of a 50 km/h speed limit, the initial part of the volume density curve is changed to the straight, broken line from the origin to Point 3, the slope of which is 50 km/h. The final part of the curve is unchanged, as shown by the broken line overlaying the thick, gray-shaded line from Point 3 down to the point of maximum density and zero volume. If traffic conditions corresponded to Point 1 prior to the speed limit reduction, they would change to somewhere between Points 2 and 3, decreasing volume and increasing density but, being on the up-sloping, initial part of the new volume-density relationship, providing more stable conditions and greatly reducing the likelihood of flow breakdown.

Hegyi, De Schutter and Hellendoorn (2005) noted particularly that main-stream speed limitation in the vicinity of a freeway on-ramp reduces the likelihood of flow breakdown because the reduction in volume illustrated in Figure 8.11 corresponds to larger (time) headways, or gaps, into which entering vehicles can merge. Hegyi, De Schutter and Hellendoorn (2005) also commented that that the coordinated use of VSL and ramp metering together produces traffic flow efficiency benefits far greater than the sum of the benefits from using each independently.

Abdel-Aty, Dilmore and Dhindsa (2006), Lee, Hellinga and Saccomanno (2006) and Lin, Kang and Chang (2004) primarily examined the safety benefits of VSL but also identified some flow efficiency benefits.

Austroads (2009b) further reviewed both overseas and Australian VSL practices and research work and found that most of the VSL installations had provided significant safety benefits. Although the application of VSL in some cases did not substantially contribute to improved traffic flow, a more homogeneous traffic situation would increase safety and therefore lead to less damage and loss of time for many drivers as well as improved reliability of the traffic system and a positive impact on the environment. Most of the evaluations also reported positive attitudes from drivers, better speed compliance with posted speed limits, reduced mean speed in the controlled area and less speed variance.

Therefore VSL is a useful tool for motorway control and as a signalling system for sections with unstable traffic flow and unsafe driving behaviour. Practitioners need to plan any evaluation before installation in order to collect before and after data for accurate comparison and assessment. Austroads (2009c) provided best practice recommendations on VSL design and operational principles.

Other than VSL and ramp metering, other managed motorway tools such as shoulder lane use, variable message signs, reversible lanes, automatic incident detection, driver information systems and lane use management systems are covered by various Austroads reports (e.g. Austroads 2008, 2009d and 2014). Readers should also refer to *AGTM Part 4: Network Management Strategies* (Austroads 2020b) and *Part 5: Link Management* (Austroads 2020c) for further information on the operation and application principles of these managed motorway tools.

References

Abdel-Aty, M, Dilmore, J & Dhindsa, A 2006, 'Evaluation of variable speed limits for real-time freeway safety improvement', *Accident Analysis & Prevention*, vol. 38, no. 2, pp. 335-45.

Ahn, S, Cassidy, MJ & Laval, J 2004, 'Verification of a simplified car-following theory', *Transportation Research Part B: Methodological*, vol. 38, no. 5, pp. 431-40.

Akcelik, R, Roper, R & Besley, M 1999, *Fundamental relationships for freeway traffic flows*, research report ARR 341, ARRB Transport Research, Vermont South, Vic.

Aron, M 1988, 'Car following in an urban network: simulation and experiments', *PTRC summer annual meeting, 16th, Bath, UK*, PTRC Education and Research Services, vol. P 307, pp. 27-39.

Ashton, WD 1966, *The theory of road traffic flow*, Methuen, London, UK.

Austroads 2008, *Freeway traffic flow under congested conditions: literature review*, AP-R318-08, Austroads, Sydney, NSW.

Austroads 2009a, *Freeway traffic flow under congested conditions: Monash freeway study*, AP-R334-09, Austroads, Sydney, NSW.

Austroads 2009b, *Best practice for variable speed limits: literature review*, AP-R342-09, Austroads, Sydney, NSW.

Austroads 2009c, *Best practice for variable speed limits: best practice recommendations*, AP-R344-09, Austroads, Sydney, NSW.

Austroads 2009d, *Freeway design parameters for fully managed operations*, AP-R341-09, Austroads, Sydney, NSW.

Austroads 2014, *Development of guide content on managed motorways*, AP-R464-14, Austroads, Sydney, NSW.

Austroads 2014, *Development of guide content on managed motorways*, AP-R464-14, Austroads, Sydney, NSW.

Austroads 2015a, *Guide to road design part 4C: interchanges*, AGRD4C-15, edn 2.0, Austroads, Sydney, NSW.

Austroads 2015b, *Guide to road design part 4B: roundabouts*, AGRD4B-15, edn 3.0, Austroads, Sydney, NSW.

Austroads 2017, *Guide to road design part 4A: unsignalised and signalised intersections*, edn, 3.0 AGRD04A-17 Austroads, Sydney, NSW.

Austroads 2020a, *Guide to traffic management part 3: transport study and analysis methods*, edn.4.0, AGTM03-20, Austroads, Sydney, NSW.

Austroads 2020b, *Guide to traffic management part 4: network management strategies*, edn.5.0, AGTM04-20, Austroads, Sydney, NSW.

Austroads 2020c, *Guide to traffic management part 5: link management*, edn.4.0, AGTM05-20, Austroads, Sydney, NSW.

Austroads 2020c, *Guide to traffic management part 6: intersections, interchanges and crossings management*, edn.4.0, AGTM06-20, Austroads, Sydney, NSW.

Aycin, MF & Benekohal, RF 1998, 'Linear acceleration car-following model development and validation', *Transportation Research Record*, no. 1644, pp. 10-9.

Banks, JH 1990, 'Flow processes at a freeway bottleneck', *Transportation Research Record*, no. 1287, pp. 20-8.

Banks, JH 2005, 'Metering ramps to divert traffic around bottlenecks: some elementary theory', *Transportation Research Record*, no. 1925, pp. 12-9.

Bar-Gera, H & Shinar, D 2005, 'The tendency of drivers to pass other vehicles', *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 8, no. 6, pp. 429-39.

Bennett, DW 1996, 'Elementary traffic flow theory and applications', in KW Ogden & SY Taylor (eds), *Traffic engineering and management*, Monash University, Department of Civil Engineering, Clayton, Vic, pp. 567-90.

Brackstone, M & McDonald, M 1999, 'Car-following: a historical review', *Transportation Research Part F: Traffic Psychology and Behaviour,* vol. 2, no. 4, pp. 181-96.

Brilon, W 2000, 'Traffic flow analysis beyond traditional methods', *Proceedings of the 4th international symposium on highway capacity*, TRB circular E-C018, Transportation Research Board, Washington, DC, USA, pp. 26-41.

Brilon, W & Geistefeldt, J 2009, 'Implications of the random capacity concept for freeways', *International symposium on freeway and tollway operations, 2nd, Honolulu, USA*, Transportation Research Board, Washington, DC, USA, 11 pp.

Brilon, W, Geistefeldt, J & Regler, M 2005, 'Reliability of freeway traffic flow: a stochastic concept of capacity', *International symposium on transportation and traffic theory, 16th, College Park, Maryland, USA*, Elsevier, Oxford, UK, pp.125-44.

Cassidy, M & Rudjanakanoknad, J 2002, *Study of traffic at a freeway merge and roles for ramp metering*, PATH working paper, UCB-ITS-PWP-2002-2, Institute of Transportation Studies, Berkeley, California, USA.

Chandler, RE, Herman, R & Montroll, EW 1958, 'Traffic dynamics: studies in car-following', *Operations Research*, vol. 6, no. 2, pp. 165-84.

Chakroborty, P & Kikuchi, S 1999, 'Evaluation of the General Motors based car-following models and a proposed fuzzy inference model', *Transportation Research Part C: Emerging Technologies*, vol. 7, no. 4, pp. 209-35.

Cox, DR & Smith, WL 1961, *Queues*, Methuen, London, UK.

Daganzo, ZF 2006, 'In traffic flow, cellular automata = kinematic waves', *Transportation Research*, vol. 40B, no. 5, pp. 396-403.

Denney, RW 1989, 'Traffic platoon dispersion modelling', *Journal of Transportation Engineering*, vol. 115, no. 2. pp. 193-207.

Dijker, T, Bovy, PHL & Vermijs, RGMM 1998, 'Car-following under congested conditions: empirical findings', *Transportation Research Record*, no. 1644, pp. 20-8.

Drew, DR 1968, *Traffic flow theory and control*, McGraw-Hill, New York, NY, USA.

Edie, LC & Foote, RS 1958, 'Traffic flow in tunnels', *Highway Research Board Proceedings*, vol. 37, pp. 334-44.

Forbes, TW 1963, 'Human factor considerations in traffic flow theory', *Highway Research Record 15*, pp. 60-6.

Forbes, TW, Zagorski, HJ, Holshouser, EL & Deterline, WA 1958, 'Measurement of driver reactions to tunnel conditions', *Highway Research Board Proceedings*, vol. 37, pp. 345-57.

Gazis, DC, Herman, R & Rothery, RW 1961, 'Nonlinear, follow-the-leader models for traffic flow', *Operations Research*, vol. 9, no. 4, pp. 545-67.

Gipps, PG 1974, 'Determination of equilibrium conditions for traffic on a two lane road', *Proceedings of the 6th international symposium on transportation and traffic theory, University of New South Wales, Sydney, NSW*, Elsevier, NY, USA, pp. 161-79.

Gipps, PG 1976, 'An abbreviated procedure for estimating equilibrium queue lengths in rural two lane traffic', *Transportation Science*, vol. 10, no. 4. pp. 337-47.

Gipps, PG (ed) 1984, *Traffic flow theory*, Department of Civil Engineering, Monash University, Clayton, Vic.

Grace, MJ & Potts, RB 1964, 'A theory of the diffusion of traffic platoons', *Operations Research*, vol. 12, no. 2, pp. 255-75.

Hall, FL & Agyemang-Duah, K 1991, 'Freeway capacity drop and the definition of capacity', *Transportation Research Record*, no. 1320, pp. 91-8.

Hall, FL, Hurdle, VF & Banks, JM 1992, 'Synthesis of recent work on the nature of speed-flow and flow-occupancy (or density) relationships on freeways', *Transportation Research Record*, no. 1365, pp. 12-7.

Han, C & Luk, JYK 2008, 'A review of freeway flow breakdown process', *ARRB conference, 23rd, 2008, Adelaide, South Australia*, ARRB Group, Vermont South, Vic, 12 pp.

Hegyi, A, De Schutter, B & Hellendoorn, H 2005, 'Model predictive control for optimal coordination of ramp metering and variable speed limits', *Transportation Research Part C: Emerging Technologies*, vol. 13, no. 3, pp. 185-209.

Henn, V 1997, 'A car following model using the techniques of fuzzy logic (in French)', *Recherche - Transports - Securite*, no. 54, pp. 15-25.

Herman, R, Montroll, EW, Potts, R & Rothery, RW 1959, 'Traffic dynamics: analysis of stability on car-following', *Operations Research*, vol. 7, no. 1, pp. 86-106.

Herman, R & Potts, RB 1961, 'Single-lane traffic theory and experiment', in R Herman (ed), *General Motors symposium on the theory of traffic flow, December 1959*, Van Nostrand, Princeton, NJ, pp. 120-46.

Herman, R & Rothery, RW 1962, 'Microscopic and macroscopic aspects of single-lane traffic flow', *Journal of the Operations Research Society of Japan*, vol. 5, no. 2, pp. 74-93.

Herman, R & Rothery, RW 1965, 'Analysis of experiments on single-lane bus flow', *Proceedings of the second international symposium on the theory of traffic flow, London, 1963,* OECD, Paris, France, pp. 1-11.

Hidas, P 1997, 'An urban car-following model based on desired spacing behaviour', *Conference of the Australian Institutes of Transport Research (CAITR), 19th, 1997, Melbourne, Vic*, Monash University, Institute of Transport Studies, Melbourne, Vic, 14 pp.

Highway Research Board 1965, *Highway capacity manual: Highway Research Board special report 87,* HRB, National Research Council, Washington, DC, USA.

Highways Agency 2006, *Design manual for roads and bridges: volume 6: section 2: part 1: layout of grade separated junctions TD 22/06*, Department for Transport, London, UK.

Hillier, JA & Rothery, RW 1967, 'The synchronisation of traffic signals for minimum delay', *Transportation Science*, vol. 1, no. 2, pp. 81-94.

Hoban, CJ 1984a, 'Bunching on two-lane rural roads', in Gipps, PG (ed), *Traffic flow theory*, Department of Civil Engineering, Monash University, Clayton, Vic, pp. 93-113.

Hoban, CJ 1984b, *Bunching on two lane roads*, AIR 359-14, Australian Road Research Board, Vermont South, Vic.

Hoban, CJ 1984c, 'Measuring quality of service on two lane rural roads', *Australian Road Research Board conference, 12th, 1984, Hobart, Tasmania*, Australian Road Research Board, Vermont South, Vic, pp. 117-31.

Hoban, CJ, Fawcett, GJ & Robinson, GK 1985, *A model for simulating traffic on two-lane rural roads: user guide and manual for TRARR version 3.0*, technical manual ATM10A, Australian Road Research Board, Vermont South, Vic.

Homburger, WS 1982, *Transportation and traffic engineering handbook,* 2nd edn, Institute of Transportation Engineers (ITE), Prentice-Hall, Englewood Cliffs, NJ, USA.

Johnson, RA, Miller I & Freund, J 2011, *Miller & Freund's probability and statistics for engineers*, 8th edn, Prentice Hall, Boston, MA, USA.

Kallberg, H 1981, *Overtakings and platoons on two-lane rural roads*, report 61, Technical Research Centre of Finland, Road and Traffic Laboratory, Espoo, Finland.

Kell, JH 1962, 'Analysing vehicular delay of intersections through simulation', *Highway Research Board Bulletin 356*, pp. 28-39.

Kerner, BS 2004, *The physics of traffic*, Springer, New York, NY, USA.

Kerner, BS 2007, 'On-ramp metering based on three-phase traffic theory: parts I, II & III', *Traffic Engineering and Control,* vol. 48, no. 1, pp. 28-35 (part I), vol. 48, no. 2, pp. 68-75 (part II) and vol. 48, no. 3, pp. 114-20 (part III).

Kerner, BS & Rehborn, H 1996, 'Experimental properties of complexity in traffic flow', *Physical Review E*, vol. 53, no. 5, pp. R4275-8.

Kerner, BS, Rehborn, H, Aleksic, M & Haug, A 2004, 'Recognition and tracking of spatio-temporal congested traffic patterns on freeways', *Transportation Research Part C,* vol. 12, no. 5, pp. 369-400.

Kerner, BS, Rehborn, H, Haug, A & Maiwald-Hiller, I 2005, 'Tracking of congested traffic patterns on freeways in California', *Traffic Engineering & Control,* vol. 46, no. 10, pp. 380-5.

Kimber, RM & Hollis, EM 1979, *Traffic queues and delays at road junctions*, laboratory report 909, Transport and Road Research Laboratory, Crowthorne, UK.

Kometani, E & Sasaki, T 1961, 'Car-following theory and stability limit of traffic volume', *Journal of the Operations Research Society of Japan*, vol. 3, no. 4, pp. 176-90.

Lay, MG 1998, 'Headways for car-following', *Highway Engineering in Australia*, vol. 30, no. 4, pp. 16, 18-9.

Lee, C, Hellinga, B & Saccomanno, F 2006, 'Evaluation of variable speed limits to improve traffic safety', *Transportation Research Part C: Emerging Technologies*, vol. 14, no. 3, pp. 213-28.

Leisch, JP & Mason, JM 2006, *Freeway and interchange: geometric design handbook*, Institute of Transportation Engineers, Washington, DC, USA.

Leo, CJ & Pretty, RL 1992, 'Numerical simulation of macroscopic continuum traffic models', *Transportation Research,* vol. 26B, no. 3, pp. 207-20.

Lighthill, MJ & Whitham, GB 1955, 'On kinematic waves II: a theory of traffic flow on long crowded roads', *Proceedings of the Royal Society of London*, The Royal Society, London, UK, series A, vol. 229, no. 1178, pp. 317-45.

Lin, PW, Kang, KP & Chang, GL 2004, 'Exploring the effectiveness of variable speed limit controls on highway work-zone operations', *Journal of Intelligent Transportation Systems*, vol. 8, no. 3, pp. 155-68.

Lindgren, RVF 2005, 'Analysis of flow features in queued traffic on a German freeway', PhD thesis, Portland State University, Portland, USA.

Lindgren, RV, Bertini, RL, Helbing, D & Schonhof, M 2006, 'Toward demonstrating predictability of bottleneck activation on German Autobahns', *Transportation Research Record*, no.1965, pp.12-22.

Lorenz, M & Elefteriadou, L 2001, 'Defining freeway capacity as function of breakdown probability', *Transportation Research Record*, no. 1776, pp. 43–51.

May, AD 1990, *Traffic flow fundamentals*, Prentice Hall, Englewood Cliffs, NJ, USA.

McLean, JR 1989, *Two-lane highway traffic operations: theory and practice*, Gordon and Breach Science, London, UK.

Mehmood, A, Saccomanno, F & Hellinga, B 2003, 'Application of system dynamics in car-following models', *Journal of Transportation Engineering*, vol. 129, no. 6, pp. 625-34.

Michalopolous, M 1988, 'A numerical methodology for analyzing traffic flow at congested surface streets', *ARRB conference, 14th, 1988, Canberra*, Australian Road Research Board, Vermont South, Vic, vol. 14, no. 2, pp. 173-85.

Miller, AJ 1961, 'A queueing model of road traffic flow', *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 23, no. 1, pp. 64-90.

Miller, AJ 1962, 'Road traffic flow considered as a stochastic process', *Mathematical Proceedings of the Cambridge Philosophical Society*, vol. 58, no. 2, pp. 312-25.

Miller, AJ 1963, 'Analysis of bunching in rural two-lane traffic', *Operations Research*, vol. 11, no. 2, pp. 236-47.

Morrall, JF & Werner, A 1990, 'Measuring level of service on two lane highways by overtakings', *Transportation Research Record*, no. 1287, pp. 62-9.

Newell, GF 2002, 'A simplified car-following theory: a lower order model', *Transportation Research Part B: Methodological*, vol. 36, no. 3, pp. 195–205.

Ngoduy, D, Hoogendoorn, SP & Van Zuylen, HJ 2006, 'Continuum traffic model for freeway with on- and off-ramp to explain different traffic congested states', *Transportation Research Record,* no. 1965, pp. 91-102.

Ossen, S, Hoogendoorn, SP & Gorte, BGH 2006, 'Interdriver differences in car-following: a vehicle trajectory-based study', *Transportation Research Record*, no.1965, pp. 121-9.

Pacey, GM 1956, 'The progress of a bunch of vehicles released from a traffic signal', research note RN/2665/GMPI, Road Research Laboratory, Crowthorne, UK.

Panwai, S & Dia, H 2005, 'Microscopic, macroscopic and stability analysis of reactive agent-based car following models', *Conference of the Australian Institutes of Transport Research (CAITR), 27th, 2005, Brisbane, Queensland*, Monash University, Institute of Transport Studies, Clayton, Vic, 26 pp.

Papageorgiou, M (ed) 1983, *Applications of automatic control concepts to traffic flow modeling and control*, Springer, Berlin, Germany.

Papageorgiou, M 1998, 'Some remarks on macroscopic traffic flow modelling', *Transportation Research*, vol. 32A, no. 5, pp. 323-9.

Papageorgiou, M, Blosseville, J-M & Hadj-Salem, H 1990, 'Modelling and real-time control of traffic flow on the southern part of Boulevard Peripherique in Paris: part I: modelling', *Transportation Research*, vol. 24A, no. 5, pp. 345-59.

Payne, HJ 1971, 'Models of freeway traffic and control', *Simulation Councils*, vol. 1, pp. 51-61.

Pipes, LA 1953, 'An operational analysis of traffic dynamics', *Journal of Applied Physics*, vol. 24, no. 3, pp. 274-87.

Raff, MS & Hart, JW 1950, *A volume warrant for urban stop signs*, The Eno Foundation for Highway Traffic Control, Saugatuck, CT, USA.

Rakha, H & Crowther, B 2002, 'Comparison of Greenshields, Pipes, and Van Aerde car-following and traffic stream models', *Transportation Research Record*, no. 1802, pp. 248-62.

Reuschel, A 1950a, 'Vehicle movements in a platoon (in German)', *Oesterreichischen Ingenieur-Archir*, vol. 4, pp. 193-215.

Reuschel, A 1950b, 'Vehicle movements in a platoon with uniform acceleration or deceleration of the lead vehicle (in German)', *Oesterreichischen Ingenieur-und Architechten-Vereines*, vol. 95, pp. 59-62, 73-7.

Robertson, DI 1969, *Transyt: a traffic network study tool*, report LR253, Road Research Laboratory, Crowthorne, UK.

Roads and Traffic Authority 1999, *Road design guide: section 4: intersections at-grade*, RTA, Sydney, NSW.

RSRTA 2005, 'Manual for the design of road traffic facilities', Road and Transportation Research Association, Cologne, Germany.

Salter, RJ & Hounsell, NB 1996, *Highway traffic analysis and design*, 3rd edn, Palgrave Macmillan, London, UK.

Schonhof, M & Helbing, D 2007, 'Empirical features of congested traffic states and their implications for traffic modelling', *Transportation Science*, vol. 41, no. 2, pp. 135-66.

Schuhl, A 1955, 'The probability theory applied to distribution of vehicles on two-lane highways', in Gerlough, DL & Schuhl, A, *Poisson and traffic*, The Eno Foundation for Highway Traffic Control, Saugatuck, CT, USA, pp. 59-75.

Seddon, PA 1971, 'Another look at platoon dispersion: 1. The kinematic wave theory', *Traffic Engineering and Control*, vol. 13, no. 8, pp. 332-6.

Seddon, PA 1972a, 'Another look at platoon dispersion: 2. The diffusion theory', *Traffic Engineering and Control*, vol. 13, no. 9, pp. 388-90.

Seddon, PA 1972b, 'Another look at platoon dispersion: 3. The recurrence relationship', *Traffic Engineering and Control*, vol. 13, no. 10, pp. 442-4.

Shawky, M & Nakamura, H 2007, 'Characteristics of breakdown phenomenon in merging sections of urban expressways in Japan', *Transportation Research Record*, no. 2012, pp. 11-9.

Sultan, B, Brackstone, N & McDonald, M 2004, 'Drivers' use of deceleration and acceleration information in car-following process', *Transportation Research Record*, no. 1883, pp. 31-9.

Tanner, JC 1961, 'Delays on a two-lane road', *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 23, no. 1, pp. 38-63.

Tanner, JC 1962, 'A theoretical analysis of delays at an uncontrolled intersection', *Biometrika*, vol. 49, no. 1/2, pp. 163-70.

Transportation Research Board 1985, *Highway capacity manual, special report 209*, TRB, National Research Council, Washington, DC, USA.

Transportation Research Board 2010, *Highway capacity manual*, 5th edn, TRB, Washington, DC, USA.

Van den Hoogen, E & Smulders, S 1994, 'Control by variable speed signs: results of the Dutch experiment', *Seventh international conference on road traffic monitoring and control, London, England*, IEE conference publication no. 391, pp. 145-9.

VicRoads 2013, *Managed freeways: freeway ramp signals handbook*, VicRoads, Kew, Vic.

Walpole, RE, Myers, RH, Myers, SL & Keying, EY 2011, *Probability and statistics for engineers and scientists*, 9th edn, Pearson Education, NJ, USA.

Wang, J, Liu, R & Montgomery, F 2005, 'Car-following model for motorway traffic', *Transportation Research Record*, no. 1934, pp. 33-42.

Wardrop, JG 1952, 'Road paper: some theoretical aspects of road traffic research', *Proceedings of the Institute of Civil Engineers*, vol. 1, no. 3, pp. 325-62.

Wohl, M & Martin, BV 1967, *Traffic system analysis: for engineers and planners*, McGraw-Hill, New York, NY, USA.

Wu, J, McDonald, M & Chatterjee, K 2007, 'A detailed evaluation of ramp metering impacts on driver behaviour', *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 10, no. 1, pp. 61-75.

Commentary 1

C1.1 Space Mean Speed and Time Mean Speed

For any traffic stream, space mean speed is always less than or equal to time mean speed because slower vehicles occupy any given segment of road for a longer period of time than faster vehicles, and therefore receive a greater weighting in the calculation of space mean speed than they do in the calculation of time mean speed. This is illustrated by the following simple numerical example:

Measured spot speeds at the start of a 1 km length of road record 50% of vehicles travelling at 30 km/h and 50% at 60 km/h. Hence, the time mean speed is v_t = 45 km/h.

Assume that all vehicles maintain constant speed. Then, travel times over the 1 km length are two minutes for those travelling at 30 km/h and one minute for those at 60 km/h. Assume 3 s headways at the measurement point, with every second vehicle travelling at 30 km/h and every other vehicle 60 km/h. Then, in any two minute period, 40 vehicles enter the 1 km segment, 20 travelling at 30 km/h and 20 at 60 km/h. At the end of the two minute period, all 20 travelling at 30 km/h are still within the segment but the first 10 of those at 60 km/h have already left it – that is, at this instant, there are 30 vehicles within the 1 km segment, 20 travelling at 30 km/h and 10 at 60 km/h. Thus the space mean speed is:

$$v_s = \frac{20(30) + 10(60)}{30} = \frac{1200}{30} = 40\,\frac{\text{km}}{\text{h}}$$

Note that this is the inverse of the average travel time over the 1 km segment, for all vehicles whose spot speeds are recorded at the measurement point. That is, 1.5 minutes/km is equivalent to 40 km/h.

Now, the variance of space speeds can be obtained by noting that, at any instant, 20 vehicles within the segment have a speed of 30 km/h and 10 have a speed of 60 km/h, that is, they have deviations from the 40 km/h space mean speed of -10 km/h and +20 km/h respectively. The variance of space speeds can thus be calculated as:

$$\sigma_s^2 = \frac{\sum_{i=1}^{N}(v_i - v_s)^2}{N-1} = \frac{20(-10)^2 + 10(+20)^2}{(30-1)} = 206.9\,\left(\frac{\text{km}}{\text{h}}\right)^2$$

Wardrop's relationship (Wardrop 1952) would then estimate the time mean speed as:

$$v_t = v_s + \frac{\sigma_s^2}{v_s} = 40 + \frac{206.9}{40} = 45.2\ \text{km/h}$$

which is close to the value of 45 km/h calculated as the arithmetic mean of the spot speeds.

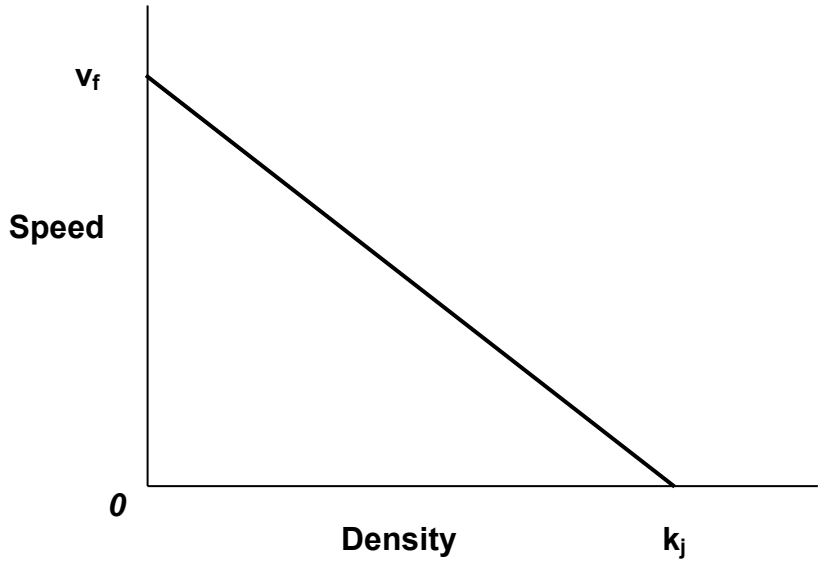
[Back to body text]

Commentary 2

C2.1 Implications of a Perfect Linear Speed-Density Relationship

Although a perfectly linear speed-density relationship for uninterrupted traffic flow (as illustrated in Figure C2. 1) is not observed in practice (see Figure 2.2), it is instructive to examine the implications of the relationship being perfectly linear.

Figure C2. 1: Perfectly linear speed-density relationship



In this case, the space mean speed of the traffic would decrease linearly from the mean free speed, v_f, at zero density, to zero speed at the jam density, k_j. That is, the speed-density relationship could be expressed mathematically as:

$$v = v_f - \left(\frac{v_f}{k_j}\right)k \qquad \text{C1}$$

which can be rearranged, with speed as the independent variable, in the form:

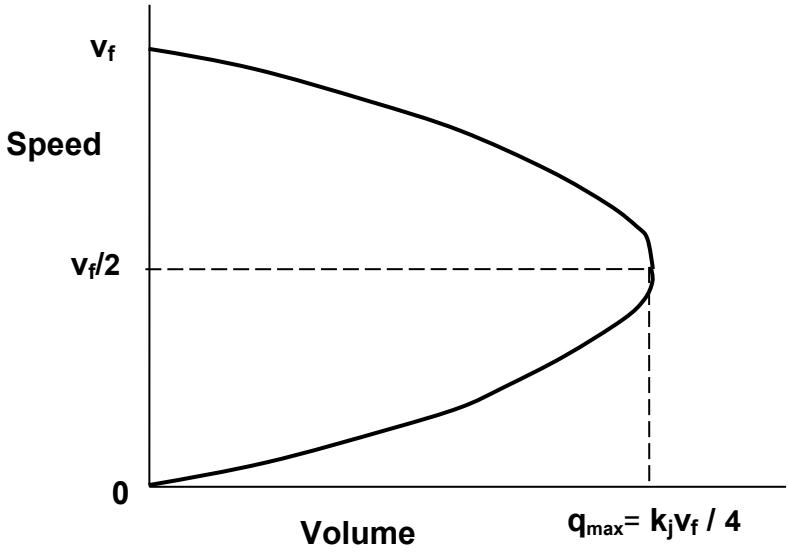
$$k = k_j - \left(\frac{k_j}{v_f}\right)v \qquad \text{C2}$$

The speed-volume relationship could then be derived by substituting the expression on the right hand side of Equation C2 in the fundamental traffic flow relationship q = k.v (see Equation 2.3) to give:

$$q = k_j v - \left(\frac{k_j}{v_f}\right)v^2 \qquad \text{C3}$$

Equation C.3 is a perfectly parabolic relationship, with zero volume at speeds of zero and v_f and a maximum volume of $q = (k_j v_f) / 4$ when $v = v_f / 2$. See Figure C2 2, which can be compared with the **approximately** parabolic relationship typically observed in practice, as shown in Figure 2.1.

Figure C2 1: Perfectly parabolic speed-volume relationship

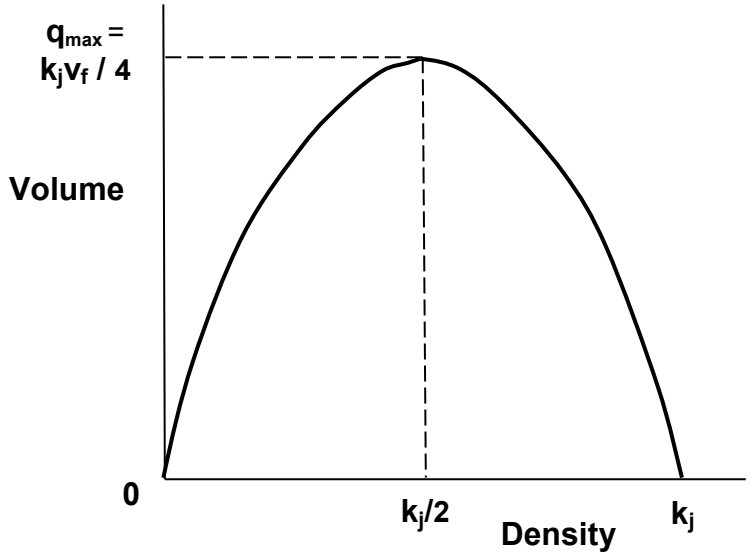


Similarly, the volume-density relationship corresponding to Figure C2. 1 could be derived by substituting the expression on the right hand side of Equation C1 in the fundamental relationship
q= k.v to give:

$$q = v_f k - \left(\frac{v_f}{k_j}\right) k^2 \qquad \text{C4}$$

Equation C4 is also a perfectly parabolic relationship, with zero volume at speeds of zero and k_j and maximum volume of $q = (k_j v_f) / 4$ at $k = k_j / 2$. See Figure C2 2, which can be compared with the **approximately** parabolic relationship typically observed in practice, as shown in Figure 2.3.

Figure C2 2: Perfectly parabolic volume-density relationship..



[Back to body text]

Commentary 3

C3.1 Derivations of Queue Length Formulae in Section 4.4.1

The elementary queuing system considered in Section 4.4.1 is an [M/M/1] system, that is, a single channel, single server system with random arrivals, random service times and a first-come-first-served discipline. Let the average arrival rate be r queue members per unit time and the average service rate be s queue members per unit time.

Let $P_n(t + \Delta t)$, $n > 0$, be the probability that the system is in state n at time $t + \Delta t$. If Δt is considered to be sufficiently small that only one event – an arrival or a departure (completion of service) – can occur during Δt, there are only three ways in which this state could be reached:

- The system is in state n at time t and there is no change during Δt.
- The system is in state n-1 at time t and there is one arrival during Δt.
- The system is in state n+1 at time t and there is one departure during Δt.

Now note the following probabilities:

$$\Pr(\text{one arrival during } \Delta t) = r.\Delta t \qquad \text{C5}$$

$$\Pr(\text{one departure during } \Delta t) = s.\Delta t \qquad \text{C6}$$

and, since no more than one event can occur during Δt,

$$\Pr(\text{no arrival during } \Delta t) = 1 - r.\Delta t \qquad \text{C7}$$

$$\Pr(\text{no departure during } \Delta t) = 1 - s.\Delta t \qquad \text{C8}$$

Then,

$$\begin{aligned} P_n(t + \Delta t) \quad = \quad & P_n(t) \text{ x } \Pr(\text{no arrival}) \text{ x } \Pr(\text{no departure}) \\ & + P_{n-1}(t) \text{ x } \Pr(\text{one arrival}) \text{ x } \Pr(\text{no departure}) \\ & + P_{n+1}(t) \text{ x } \Pr(\text{no arrival}) \text{ x } \Pr(\text{one departure}) \end{aligned} \qquad \text{C9}$$

Substituting the probabilities from Equations C5 to C8 into Equation C9 and ignoring all $(\Delta t)^2$ terms as negligibly small, leads to:

$$P_n(t + \Delta t) = P_n(t)[1 - (r+s).\Delta t] + P_{n-1}(t)[r.\Delta t] + P_{n+1}(t)[s.\Delta t] \qquad \text{C10}$$

Similarly, the special case of n = 0 is:

$$P_0(t + \Delta t) = P_0(t)[1 - r.\Delta t] + P_1(t)[s.\Delta t] \qquad \text{C11}$$

Taking the limit of $\{[P_n(t + \Delta t) - P_n(t)] / \Delta t\}$ as Δt approaches zero produces the derivatives of the above probabilities with respect to time, as follows:

$$P'_n(t) = -(r+s).P_n(t) + r.P_{n-1}(t) + s.P_{n+1}(t), \text{ for } n > 0 \qquad \text{C12}$$

And

$$P'_0(t) = -r.P_0(t) + s.P_1(t) \quad \text{C13}$$

For time-independent, steady-state conditions, these derivatives must be zero (substituting from Equation 4.1):

$$(1 + \rho).P_n = P_{n+1} + \rho.P_{n-1} \text{ for } n > 0 \quad \text{C14}$$

And

$$P_1 = \rho.P_0 \quad \text{C15}$$

where P_n is the steady-state probability of the system being in state n.

Letting n = 1 in Equation C14 and substituting for P_1 from Equation C15 leads to the result:

$$P_2 = \rho^2.P_0 \quad \text{C16}$$

and, in fact, it can be shown that the more general result applies:

$$P_n = \rho^n.P_0 \quad \text{C17}$$

Now observe that:

$$\sum_{n=0}^{\infty} P_n = 1 = P0\,(1 + \rho + \rho 2 + \rho 3 + \ldots.) = \frac{P_0}{1-\rho} \quad \text{C18}$$

So that

$$P_0 = 1 - \rho \quad \text{C19}$$

And

$$P_n = (1 - \rho)\,\rho^n \quad \text{C20}$$

The mean of this distribution is the expected number in the system and is calculated as:

$$E(n) = \sum_{n=N+1}^{\infty} n.\rho_n = (1-\rho) \sum_{n=N+1}^{\infty} n.\rho^n = (1-\rho)\,\frac{\rho}{(1-\rho)^2} = \frac{\rho}{1-\rho} = \frac{r}{s-r} \quad \text{C21}$$

The probability of there being more than N items in the queuing system is:

$$Pr(n > N) = \sum_{n=N+1}^{\infty} P_n\ 1 = (1-\rho) \sum_{n=N+1}^{\infty} \rho^n = \rho^{N+1} \quad \text{C22}$$

The mean queue length, excluding the unit being serviced, is determined as:

$$\mathrm{E(m)} = \sum_{\mathrm{n=1}}^{\infty} (n-1)\,\rho_{\mathrm{n}} = \frac{\rho^2}{1-\rho} = \frac{\rho}{1-\rho} - \rho \qquad \text{C23}$$

Thus,

$$\mathrm{E(m)} = \mathrm{E(n)}.\rho = \mathrm{E(n)} - \rho \qquad \text{C24}$$

Note that, as the steady-state case of $\rho < 1$ is being considered, E(m) is not (as might be expected) equal to E(n) – 1. This is because there is a finite probability that the system is empty, in which case n = m = 0.

The final result related to queue length is the variance of the number of units in the system, which can be shown to be:

$$\sigma^2(\mathrm{n}) = \sum_{\mathrm{n=0}}^{\infty} \mathrm{n}^2\ \rho_{\mathrm{n}} - [\mathrm{E(n)}]^2 = \frac{\rho}{(1-\rho)^2} \qquad \text{C25}$$

[Back to body text]

Commentary 4

C4.1 Derivations of Queue Delay Formulae in Section 4.4.2

The time spent by an individual in a queuing situation is made up of the time spent waiting in the queue until service is commenced (denoted by 'w') and the time spent being served.

The time spent waiting for service to be commenced will be zero if the queuing system is empty (i.e. no queue and no-one being served) when the individual arrives. The probability of a zero waiting time is thus the same as the probability of the system being empty, which is given by Equation 4.2 (or Equation C19). That is:

$$Pr(w = 0) = P_0 = 1 - \rho = \frac{\rho}{1-\rho} - \rho \qquad \text{C26}$$

If the queuing system is not empty when an individual arrives, the waiting time will be non-zero. If a probability frequency function f(w) is postulated for w, then the probability that the waiting time will be between w and w+dw can be stated as:

$$\mathrm{Pr(w < wait < w + dw) = f(w).dw} \qquad \text{C27}$$

The frequency distribution f(w) can be identified by observing that the joint probability that there will be n items in the system when an individual arrives and that the individual will have a waiting time before commencing service between w and w+dw, where dw is sufficiently small that it can accommodate only one event, is the product of the following three probabilities:

1. The probability, Pn, that the system is in state n when the individual arrives.
2. The probability of n – 1 units completing service during the time w. Given that the average service rate is s and therefore the average number of units served in time w is sw, this probability is given by the Poisson distribution as:

$$P_{n-1}(w) = \frac{(sw)^{n-1}e^{-sw}}{(n-1)!} \quad \text{C28}$$

3. The probability of the nth unit completing service during dw, which is:

$$P_1(dw) = s.dw \quad \text{C29}$$

Summing over all n > 0 gives

$$\begin{aligned} f(w).dw &= \sum_{n=1}^{\infty} P_n.P_{n-1}(w).P_1(dw) \\ &= \sum_{n=1}^{\infty} \rho^n(1-\rho)\frac{(sw)^{n-1}e^{-sw}}{(n-1)!}s.dw \\ &= \rho(s-r)e^{-sw}.dw.e^{rw} \end{aligned} \quad \text{C30}$$

Dividing through by dw,

$$f(w) = \rho(s-r)e^{-(s-r)w} \quad \text{C31}$$

The cumulative forms of this distribution are as follows:

$$\Pr(0 < \text{wait} \leq w) f(w) \int_0^w f(w).dw = \rho - \rho e^{-(s-r)w} \quad \text{C32}$$

And

$$\Pr(\text{wait} > w) = \int_w^{\infty} f(w).dw = \rho e^{-(s-r)w} \quad \text{C33}$$

As would be expected, these two probabilities and that in Equation C26 sum to unity.

The average, or expected waiting time before start of service, over all arrivals, is:

$$E(w) = 0(1-\rho) + \int_0^{\infty} w.f(w).dw = \frac{\rho}{s-r}\frac{r}{s(s-r)} \quad \text{C34}$$

And the average waiting time over only those arrivals whose wait is non-zero is:

$$\mathrm{E(w|w>0)} = \frac{\mathrm{E(w)}}{\rho} = \frac{1}{\mathrm{s-r}} \qquad \text{C35}$$

The total time spent in the system, including service time, is denoted by τ. If a probability frequency function $f(\tau)$ is postulated for τ, then:

$$\Pr(\tau < \text{total time } < \tau + \mathrm{d}\tau) = \mathrm{f}(\tau).\mathrm{d}\tau \text{ for } \tau > 0 \qquad \text{C36}$$

On arrival, an individual may find n units already in the system, where n = 0, 1, 2, …, $_\infty$. Given this, the frequency distribution $f(\tau)$ can be derived in a manner very similar to that used to derive the waiting time before start of service, w, by noting that:

$$\begin{aligned} f(\tau).\mathrm{d}\tau &= \sum_{\mathrm{n}=0}^{\infty} \mathrm{P_n}\,.\,\mathrm{P_n}(\tau).\,\mathrm{P_1}(\mathrm{d}\tau) \\ &= \sum_{\mathrm{n}=0}^{\infty} \rho^{\mathrm{n}}\,(1-\rho)\,\frac{(\mathrm{s}\tau)^{\mathrm{n}}\mathrm{e}^{-\mathrm{s}\tau}}{\mathrm{n!}}\mathrm{s}.\,\mathrm{d}\tau \\ &= \mathrm{s}(1-\rho)\mathrm{e}^{-\mathrm{s}\tau}.\,\mathrm{d}\tau.\,\mathrm{e}^{\mathrm{r}\tau} \end{aligned} \qquad \text{C37}$$

Dividing through by $\mathrm{d}\tau$ and rearranging:

$$\mathrm{f}(\tau) = (\mathrm{s}-\mathrm{r})\mathrm{e}^{-(\mathrm{s}-\mathrm{r})\tau} \qquad \text{C38}$$

The average, or expected total time in the system, over all arrivals, is the mean of this distribution, which is:

$$\mathrm{E}(\tau) = \int_0^{\infty} \tau.\,\mathrm{f}(\tau).\,\mathrm{d}\tau = \frac{1}{\mathrm{s}-\mathrm{r}} \qquad \text{C39}$$

Comparing this with Equation C34, as would be expected,

$$\mathrm{E}(\tau) = \mathrm{E}(\mathrm{w}) + \frac{1}{\mathrm{s}} \qquad \text{C40}$$

[Back to body text]

Commentary 5

C5.1 Derivations of Formulae in Section 5.2.1 for Delays to Minor Traffic in Gap Acceptance Situations

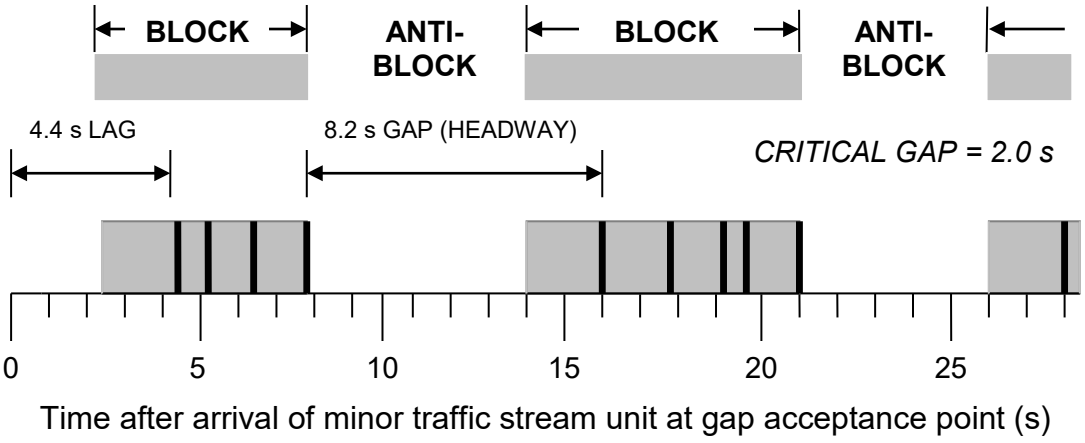
Figure C5 1 clarifies the definitions of gaps, lags, blocks and anti-blocks given in Table 5.1. In this figure, the arrival times (in seconds) at an unsignalised intersection of successive major road vehicles are shown as heavy lines rising vertically from a horizontal time axis. Time zero corresponds with the arrival of a minor road vehicle at its stop line at the intersection. All minor road vehicles are assumed to have a critical gap of 2.0 s.

On its arrival at time t = 0, the first minor road vehicle is faced with a lag (i.e. time before the arrival of the next major road vehicle) of 4.4 s. During the first 2.4 s of this lag, the minor road vehicle could (and probably would) depart, because a time not less than its critical gap of 2.0 s remains before the major road vehicle arrives.

A following minor road vehicle arriving at the stop line just after time t = 2.4 s could not depart, however, because it would be faced with a lag less than its critical gap of 2.0 s. Successive major road arrivals at times t = 5.2, 6.4 and 7.8 s mean that this minor road vehicle would then be faced with three gaps (headways between successive major road vehicles) of 0.8, 1.2 and 1.6 s, respectively. All these gaps are less than the minor road vehicle's critical gap, so it could not depart during this time. Time t = 7.8 s, however, marks the start of a gap of 8.2 s – much greater than the critical gap, and the minor road vehicle could then depart. In fact, any minor road vehicle arriving at the stop line during the interval t = 7.8 s and t = 14.0 s (2.0 s before the next major road arrival) could depart.

Thus, from the point of view of minor road drivers, the time at the stop line is divided into a series of intervals, which alternate between time when they are blocked from entering the intersection ('blocks') and time when they can enter the intersection ('anti-blocks'), as illustrated in Figure C5 1.

Figure C5 1: Major traffic stream arrivals, blocks and anti-blocks



By the definitions of blocks and anti-blocks, an anti-block can start only with the end of a block and vice versa. Therefore, over a significant period of time, the number of blocks and the number of anti-blocks will be equal.

Assuming random arrivals in the major traffic stream, its headway distribution will be negative exponential and the number of anti-blocks during a time period H will be the number of headways greater than or equal to the critical gap, T. Hence, if the volume of the major traffic stream is q:

$$\text{Number of anti-blocks } (= \text{number of blocks}),\ \mathrm{N} = \mathrm{Hqe}^{-\mathrm{qT}} \qquad \text{C41}$$

The average duration of anti-blocks is equal to the average duration of headways at least as large as the critical gap, T, minus the last interval T of such headways. Thus, drawing on Equation 3.23:

$$\text{Average duration of anti-blocks} = \left(\frac{1}{q} + T\right) - T = \frac{1}{q} \quad \text{C42}$$

Over the period H, the total time in anti-blocks is equal to the number of anti-blocks multiplied by their average duration, that is:

$$\text{Total time in anti-blocks} = Hqe^{-qT}.\frac{1}{q} = He^{-qT} \quad \text{C43}$$

The remainder of the period H must be spent in blocks, so that:

$$\text{Total time in blocks} = H(1 - e^{-qT}) \quad \text{C44}$$

The average duration of blocks can then be calculated as this total time divided by the number of blocks, that is:

$$\text{Average duration of blocks} = \frac{1-e^{-qT}}{qe^{-qT}} \quad \text{C45}$$

As minor traffic is arriving randomly, the proportion of minor traffic stream units that will be delayed at the gap acceptance point (e.g. at the stop line) is the proportion of time spent in blocks, that is:

$$\text{Proportion delayed} = 1 - e^{-qT} \quad \text{C46}$$

The average delay at the gap acceptance point for all minor traffic stream units can be derived by noting that the probability of any unit having to wait for n gaps, each less than T, before being able to proceed is given by the geometric distribution as:

$$P(n) = (1 - p)p^n \quad n = 0, 1, 2, ... \quad \text{C47}$$

where (drawing on Equation 3.16)

$$p = \Pr(\text{gap} < T) = 1 - e^{-qT} \quad \text{C48}$$

so that

$$P(n) = e^{-qT}(1 - e^{-qT})^n \quad \text{C49}$$

The expected number of gaps less that T for which a minor traffic stream unit has to wait before being able to proceed is then given by:

$$\begin{aligned} E(n) &= 0.P(0) + 1.P(1) + 2.P(2) + 3.P(3) + \quad \text{C50} \\ &= e^{-qT}(1 - e^{-qT}) + 2e^{-qT}(1 - e^{-qT})^2 + 3e^{-qT}(1 - e^{-qT})^3 + \\ &= \frac{1 - e^{-qT}}{e^{-qT}} \end{aligned}$$

Now the average duration of headways less than T in the major traffic stream is given by Equation 3.24 as:

$$h_{av}(h < T) = \frac{1}{q} - \frac{Te^{-qT}}{1 - e^{-qT}} \qquad \text{C51}$$

Then the average delay experienced by **all** minor traffic stream units at the gap acceptance point is:

$$d_{av}(d \geq 0) = E(n).[h_{av}(h < T)] \qquad \text{C51}$$

$$= \frac{(1 - e^{-qT})}{(e^{-qT})} \frac{(1 - e^{-qT} - qTe^{-qT})}{q(1 - e^{-qT})}$$

which simplifies to:

$$d_{av}(d \geq 0) = \frac{1}{qe^{-qT}} - \frac{1}{q} - T \qquad \text{C52}$$

The average delay at the gap acceptance **point to only those minor traffic stream units that do experience such delay** is obtained by dividing the average delay to all minor stream units (as in Equation C52) by the proportion experiencing non-zero delay (given by Equation C46) to give:

$$d_{av}(d \geq 0) = \frac{(1 - e^{-qT})}{(e^{-qT})} \frac{(1 - e^{-qT} - qTe^{-qT})}{q(1 - e^{-qT})} \frac{1}{(1 - e^{-qT})}$$

which simplifies to:

$$d_{av}(d \geq 0) = \frac{1}{qe^{-qT}} - \frac{T}{1 - e^{-qT}} \qquad \text{C53}$$

[Back to Section 3]
[Back to Section 5]

Commentary 6

Curve 2 in Figure 4.5.12 in RTA (1999) provides an example of the application of Equation 5.1 (or Equation C46). This curve is presented as the boundary separating the choice between rural intersection types 'AU' and 'CH', the latter of which includes provision of a sheltered turn lane for right-turning vehicles.

In RTA (1999), the equation to Curve 2 is provided as:

$$Q_{R=}\frac{D}{1-e^{-q_0 t_g}} \qquad \text{C54}$$

where

Q_R = the total volume turning right from the relevant approach (veh/h)

D = the number of right-turning vehicles per hour that will have to wait for a suitable gap (t_g) before being able to turn, rather than being able to turn immediately

q_0 = the traffic volume opposing the right turn (veh/s)

t_g = the minimum gap in opposing traffic that allows the right turn to be made (s/veh)

The criterion implied by Curve 2 is that provision of a sheltered turn lane should be considered if D or more of the Q_R vehicles per hour turning right are unable to do so immediately, but must wait until at least one opposing vehicle has passed before being able to turn. That is, that the proportion of right-turners delayed is:

$$\text{Proportion delayed} = \frac{D}{Q_R} \qquad \text{C55}$$

Then, applying Equation 5.1 (or C46), with $q = q_0$ and $T = t_g$, and rearranging, Equation C54 is obtained.

Note that in Figure 4.5.12 in RTA (1999), Curve 2 is drawn for values of D = 15 veh/h and t_g = 5 s/veh.

[Back to Section 3]
[Back to Section 5]

Commentary 7

C7.1 Average Delay

The following procedure provides an approximate and usually conservative guide for delay. The method is based on Tanner(1962). The average delay to minor stream vehicles entering a major stream may be calculated as follows:

- Determine the practical absorption capacity C_P of the minor stream approach as described in Section 5.2.2.
- Determine the number of minor stream approach lanes required. Usually the number of lanes required is equal to the next whole number greater than the total minor stream volume divided by C_P.
- Determine the volume per lane on the minor stream approach by dividing the total minor stream volume by the number of lanes on the approach.

The average delay to minor stream vehicles entering the major stream is then given by the following:

$$W_m = \frac{q_p e^{q_p t_f}\left[e^{q_p t_a} - q_p t_a - 1\right] + q_m e^{q_p t_a}\left[e^{q_p t_f} - q_p t_f - 1\right]}{q_p\left[q_p e^{q_p t_f} - q_m e^{q_p t_a}(e^{q_p t_f} - 1)\right]} \quad \text{C54}$$

where

W_m = average delay to minor stream vehicles (s/veh)

q_p = major stream volume (veh/s)

q_m = minor stream volume per lane (veh/s)

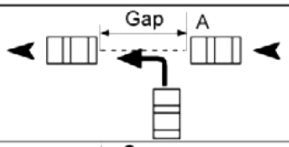
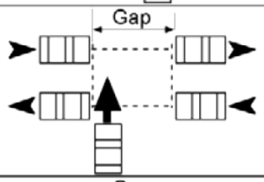
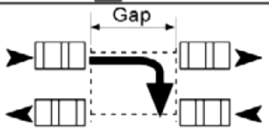
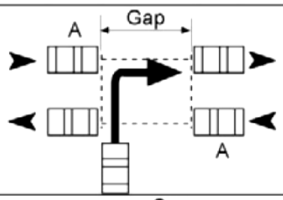
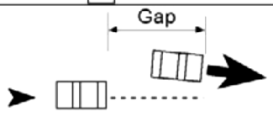
t_a = critical acceptance gap (s)

t_f follow-up headway (s)

The major stream volume can be the sum of more than one flow as illustrated in Figure C7 1. When the conflicting traffic stream arrivals at an intersection are randomly distributed, the capacity of the ‘major’ stream to absorb the ‘minor’ flow, along with the resulting delays, can be estimated from the graphs or formulas given in Figure C7 2. It is important to choose the right values of critical acceptance gap t_a and follow-up headway t_f to represent the situation being analysed. Suitable values are given in Table C7 1. Average delay to minor stream vehicles at unsignalised intersections can also be measured using data from Figure C7 3 to Figure C7 10.

Note that Figure C7 1 to Figure C7 10 were retrieved from the 2005 version of *Guide to Traffic Engineering Practice (GTEP) Part 5*, which is no longer available from Austroads and it is superseded by *Guide to Traffic Management Part 6* (Austroads 2020d) and *Guide to Road Design Parts 4A, 4B and 4C* (Austroads 2017, 2015b and 2015a).

Table C7 1: Minimum gap sight distance (MGSD)

Movement	Diagram	Description	t_a	t_f
Left hand turn		Not interfering with A Requiring A to slow	14-40 sec 5 sec	2-3 sec 2-3 sec
Crossing		Two lane/one way Three lane/one way Four lane/one way Two lane/two way Four lane/two way Six lane/two way	4 sec 6 sec 8 sec 5 sec 8 sec 8 sec	2 sec 3 sec 4 sec 3 sec 5 sec 5 sec
Right hand turn from major road		Across 1 lane Across 2 lanes Across 3 lanes	4 sec 5 sec 6 sec	2 sec 3 sec 4 sec
Right hand turn from minor road		Not interfering with A One way Two lane/two way Four lane/two way Six lane/two way	14-40 sec 3 sec 5 sec 8 sec 8 sec	3 sec 3 sec 3 sec 5 sec 5 sec
Merge		Acceleration lane	3 sec	2 sec

Note: t_a = critical acceptance gap
t_f = follow up headway

Figure C7 1: Example of major and minor flows

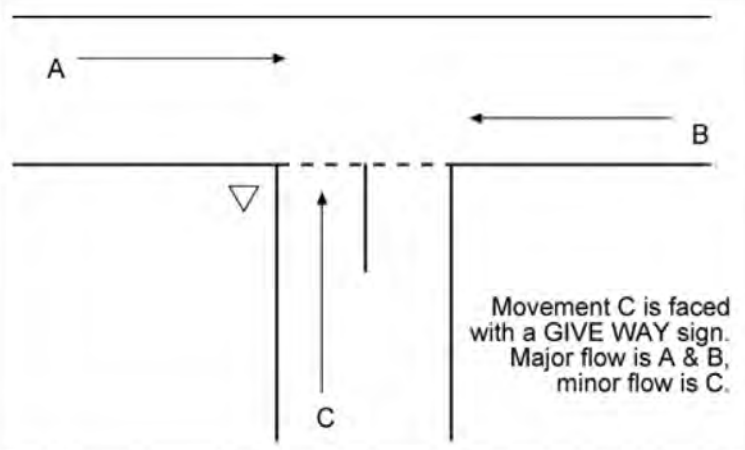
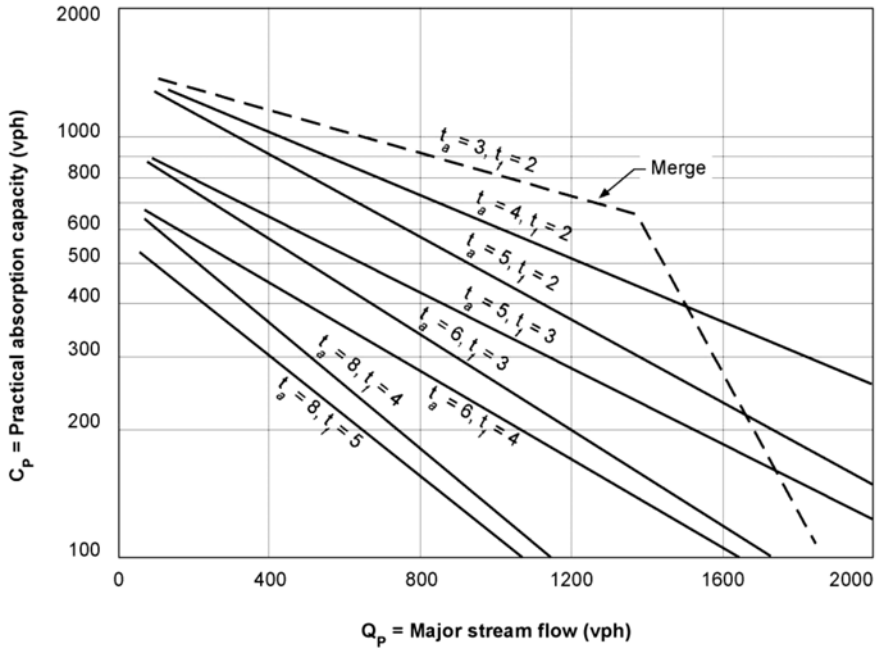


Figure C7 2: Unsignalised intersection (practical absorption capacity)



Formula:

$$C = \frac{q_p e^{-q_p t_a} \times 3{,}600}{1 - e^{-q_p t_f}}$$

Cp = 0.8 C

C = absorption capacity (vph)
C_p = practical absorption capacity (vph)
q_p = major (priority) stream flow rate (vps)

t_a = critical acceptance gap (sec)
t_f = follow up headway (sec)
e = constant (2.7183)

Figure C7 3: Average delay to minor stream vehicles at unsignalised intersections (t_a= 3 sec, t_f = 2 sec)

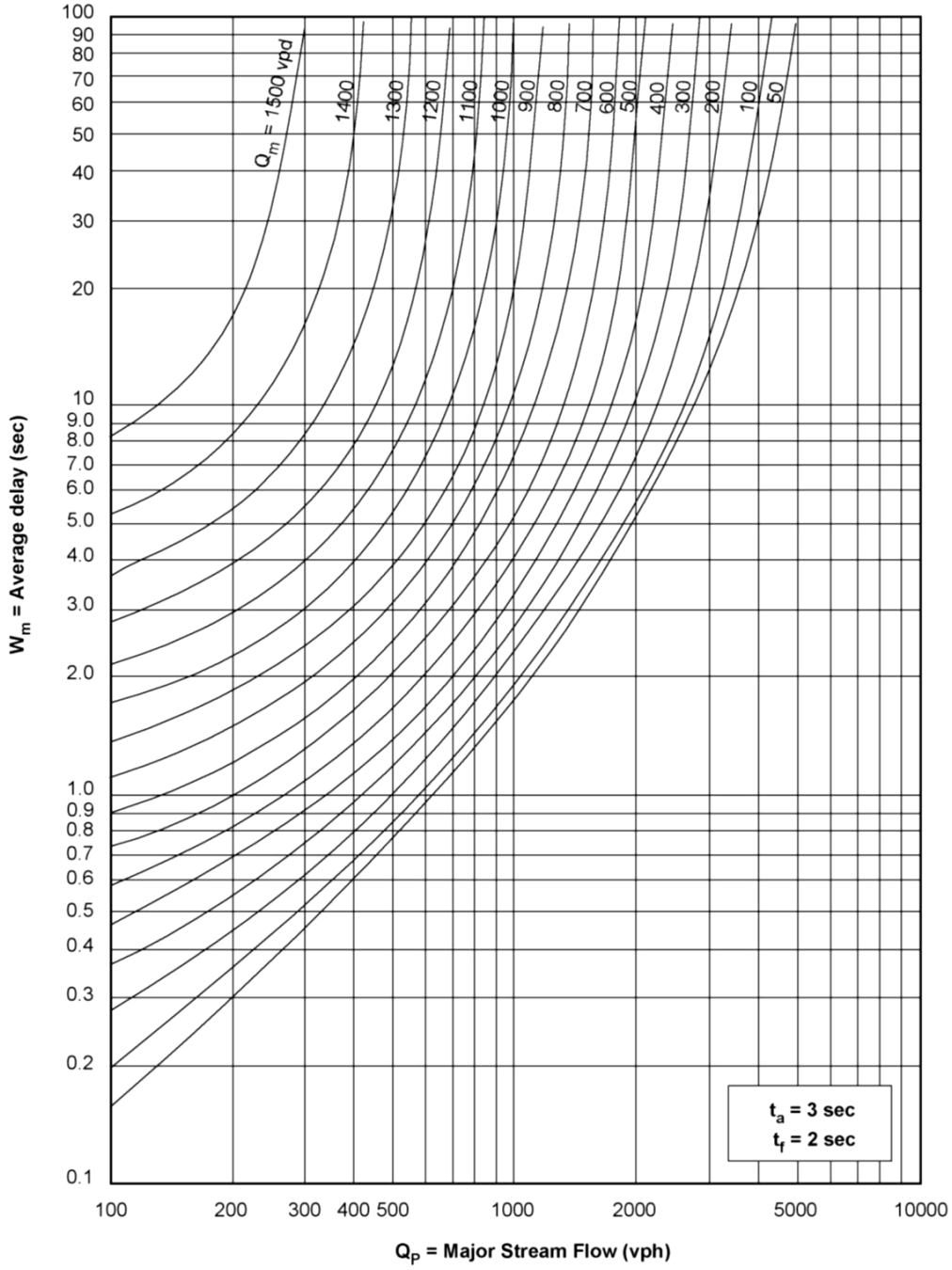


Figure C7 4: Average delay to minor stream vehicles at unsignalised intersections (t_a= 4 sec, t_f = 2 sec)

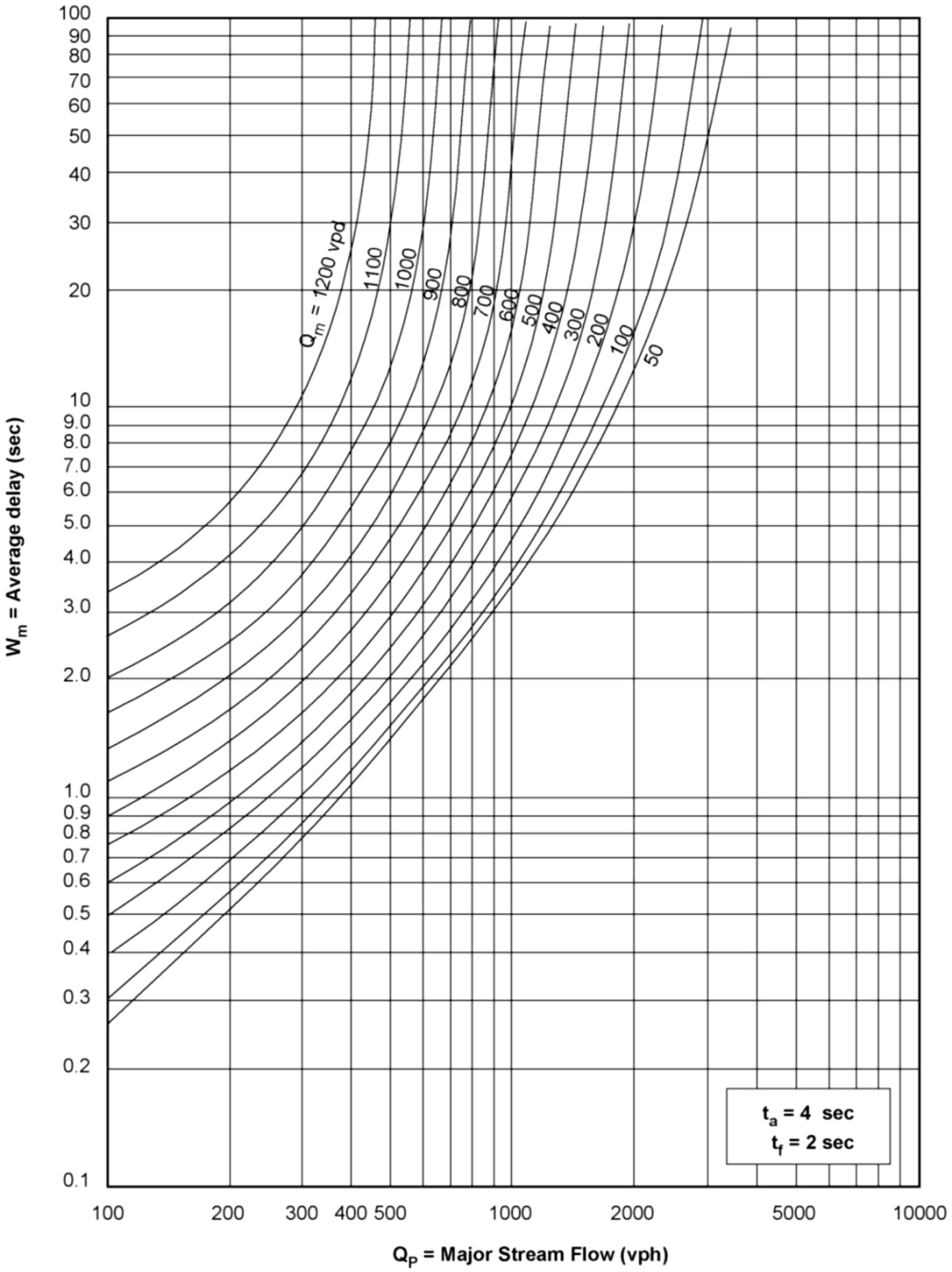


Figure C7 5: Average delay to minor stream vehicles at unsignalised intersections (t_a = 5 sec, t_f = 2 sec)

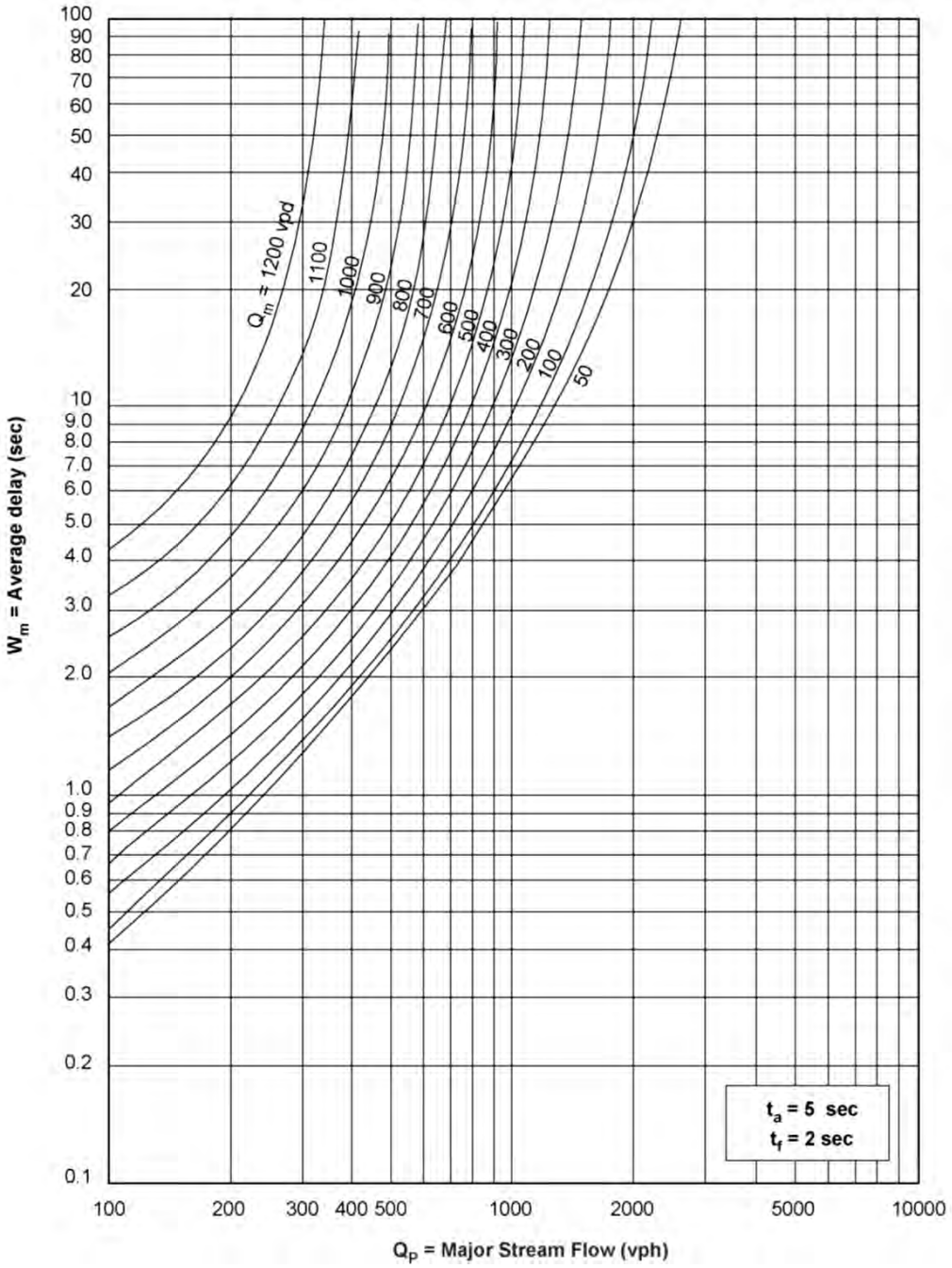


Figure C7 6: Average delay to minor stream vehicles at unsignalised intersections (t_a= 5 sec, t_f= 3 sec)

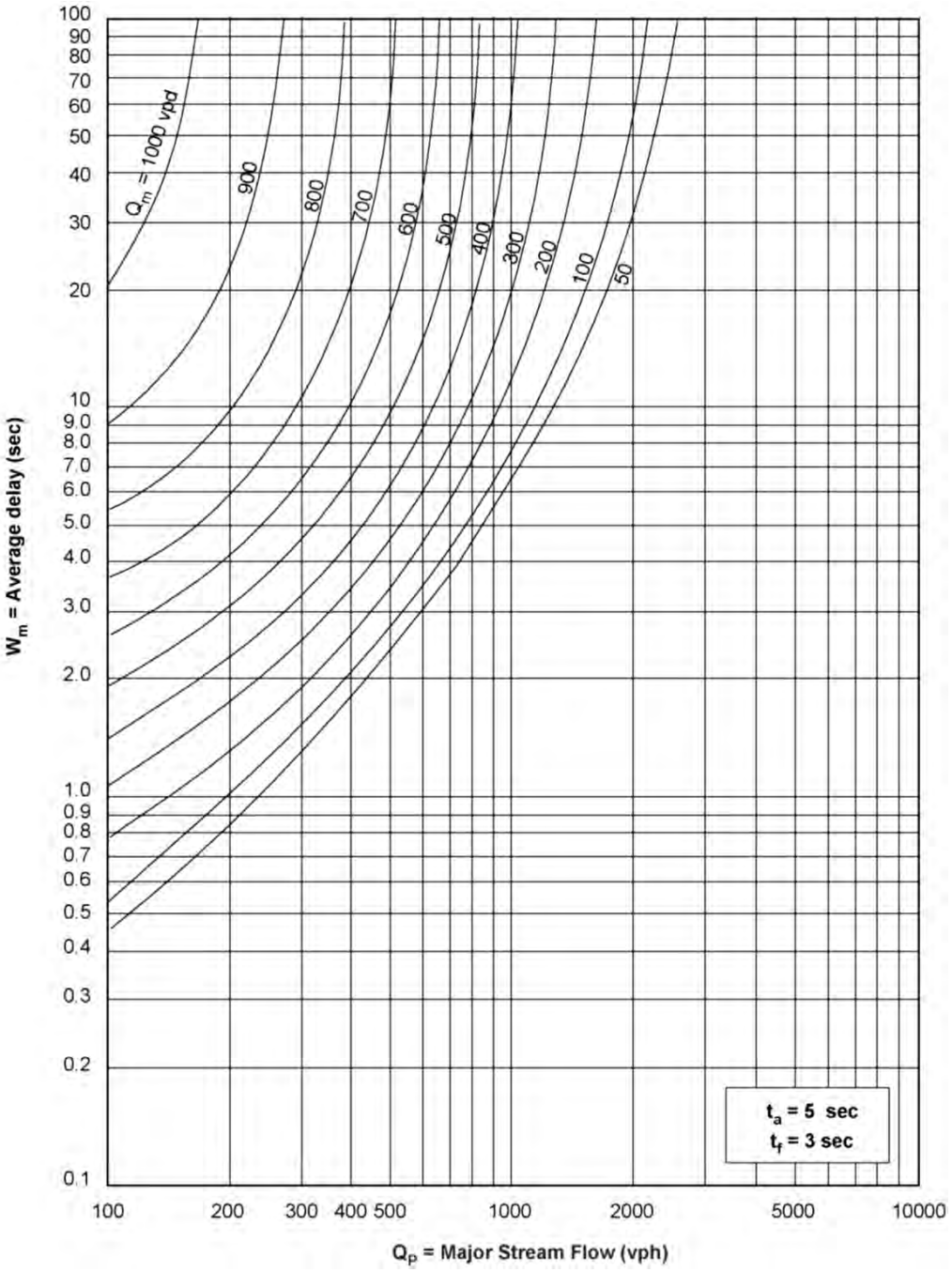


Figure C7 7: Average delay to minor stream vehicles at unsignalised intersections (t_a = 6 sec, t_f = 3 sec)

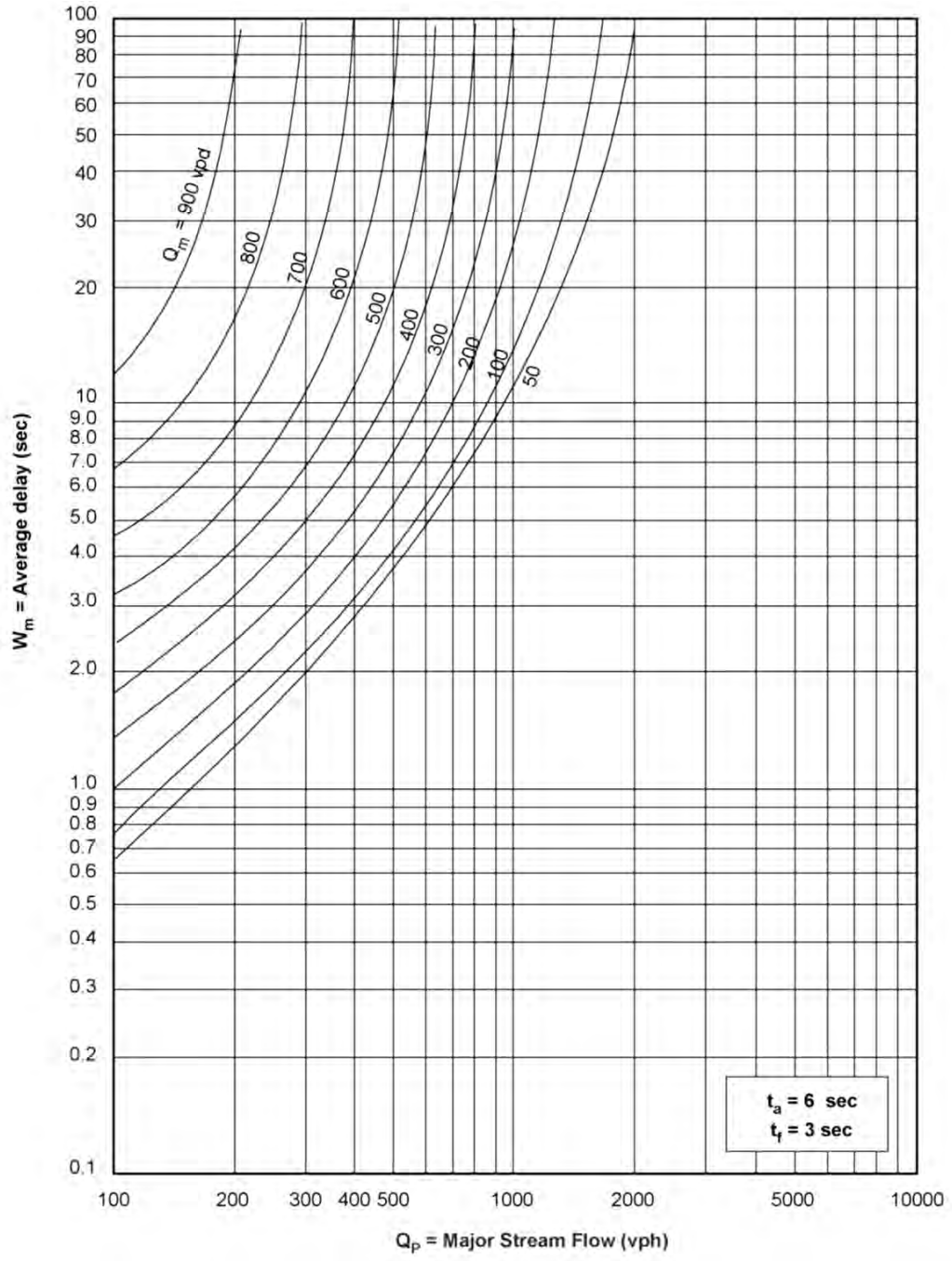


Figure C7 8: Average delay to minor stream vehicles at unsignalised intersections (t_a = 6 sec, t_f = 4 sec)

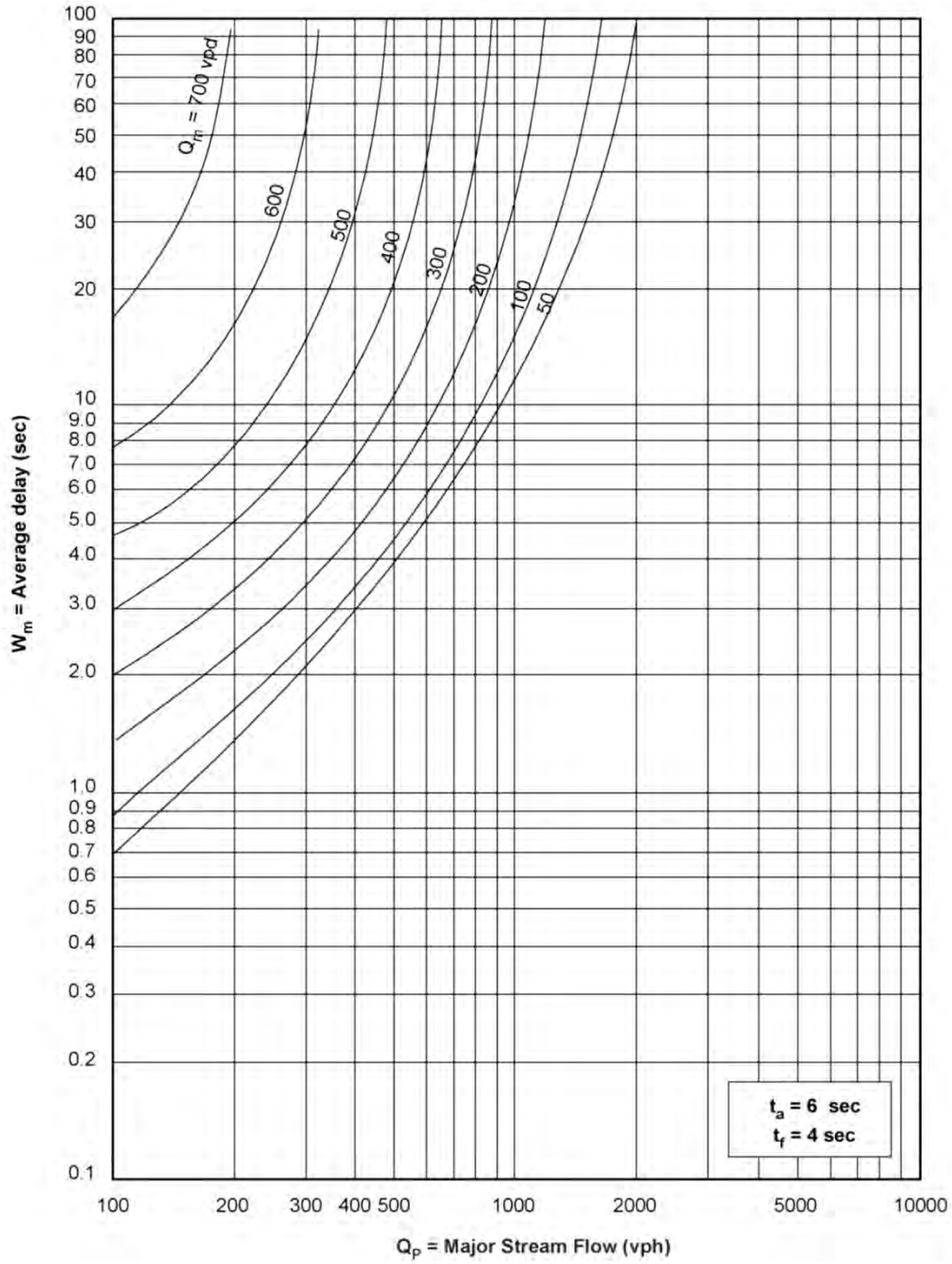


Figure C7 9: Average delay to minor stream vehicles at unsignalised intersections (t_a = 8 sec, t_f = 4 sec)

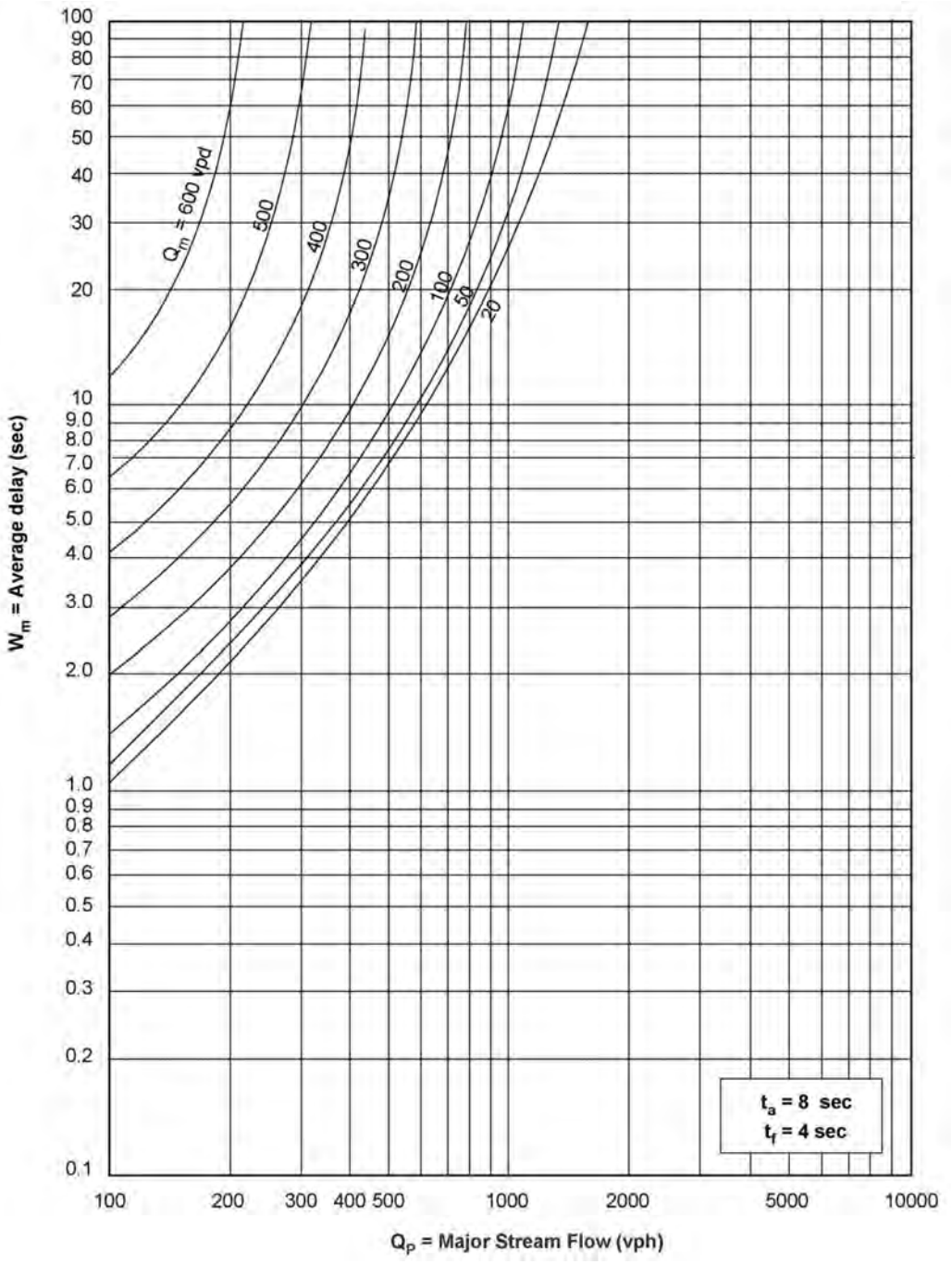
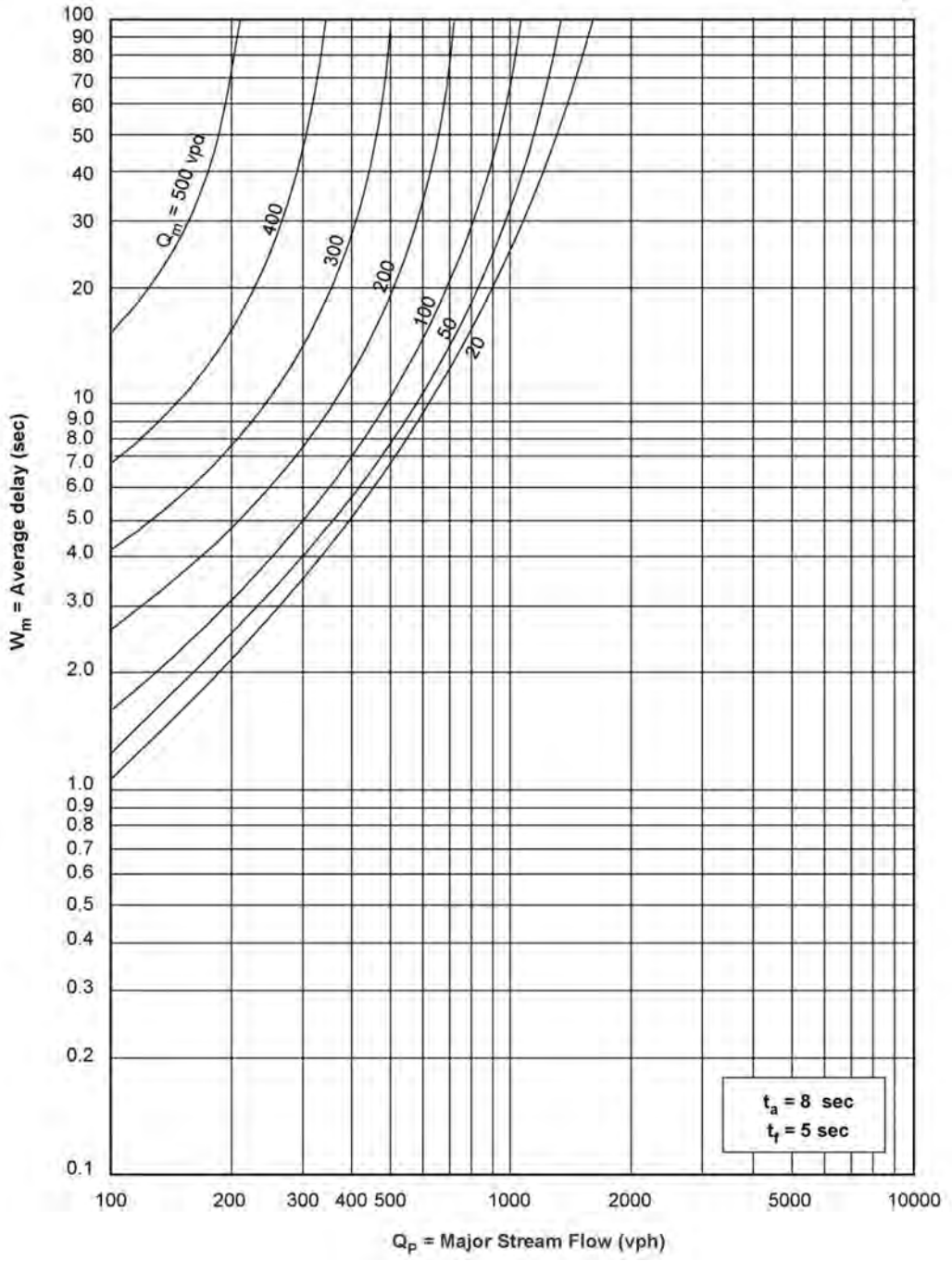


Figure C7 10: Average delay to minor stream vehicles at unsignalised intersections (t_a = 8 sec, t_f = 5 sec)



[Back to Section 3]
[Back to Section 5]

Commentary 8

C8.1 Derivation of Theoretical Absorption Capacity Formula in Section 5.2.2

To derive the theoretical absorption capacity, assume that a queue of minor traffic stream vehicles is waiting to cross or enter a major traffic stream that has volume q and a negative exponential headway distribution, and let t_i be the minimum gap required to allow i minor traffic stream vehicles to carry out the manoeuvre within the one gap, where i = 1, 2, 3,....

During a significant period of time, H , the number of major road headways greater than or equal to t_i, i = 1, 2, 3,..., is:

$$\text{No. of headways} \geq t_i = \text{H.qe}^{-\text{qt}_i} \qquad \text{C56}$$

Therefore, the number of major traffic stream headways that allow exactly i minor traffic stream vehicles to cross or enter the major stream, i = 1, 2, 3,..., is:

$$\text{n}_\text{i} = \text{H.}\left(\text{qe}^{-\text{qt}_\text{i}} - \text{qe}^{-\text{qt}_{\text{i}+1}}\right) \qquad \text{C57}$$

Hence, the total number of minor stream vehicles able to cross or join the major stream during the period H is:

$$\begin{aligned} \text{N} &= \sum_{\text{i}=1}^{\infty} \text{i.n}_\text{i} \qquad \text{C58} \\ &= \sum_{\text{i}=1}^{\infty} \text{i.Hq}\left(\text{e}^{-\text{qt}_\text{i}} - \text{e}^{-\text{qt}_{\text{i}+1}}\right) \\ &= \text{Hq}\sum_{\text{i}=1}^{\infty} \text{e}^{-\text{qt}_\text{i}} \end{aligned}$$

Taking the critical gap, T, as the minimum headway that will allow one minor stream vehicle to cross or join the major stream, assume that each additional time interval T_0 in the size of the headway is sufficient to allow one additional minor stream vehicle to follow in undertaking the manoeuvre. T_0 is known as the **follow-up headway** and the above assumption implies that:

$$\text{t}_\text{i} = \text{T} + (\text{i} - 1)\text{T}_0, \text{i} = 1,2,3,\ldots \qquad \text{C59}$$

Substituting from Equation C59 into Equation C58:

$$N = Hq\sum_{i=1}^{\infty} e^{-q(T+(i-1)T_0)} \quad \text{C60}$$

$$= Hqe^{-qT}\sum_{k=0}^{\infty}\left(e^{-qT_0}\right)^k$$

$$= \frac{Hqe^{-qT}}{1-e^{-qT_0}}$$

Thus the theoretical maximum rate at which minor stream vehicles can cross or join the major traffic stream, that is, the theoretical absorption capacity, is obtained as:

$$C = \frac{N}{H} = \frac{qe^{-qT}}{1-e^{-qT_0}} \quad \text{C61}$$

This is the relationship presented as Equation 5.4, in Section 5.2.2.

[Back to body text]

Commentary 9

At an intersection where gap acceptance applies, where major road traffic arrives randomly from each direction and where a right-turning vehicle from a minor road seeks different sized critical gaps in the traffic flows from the two different directions on the major road, the theoretical absorption capacity can be developed as follows:

Assume that the total major traffic stream is made up of the traffic flow from the left, with volume q_L, and the flow from the right, with volume q_R. Each gap or headway in the total traffic stream starts with the arrival of a vehicle from one direction and ends with the next arrival, which may be from the same or the opposite direction. Therefore, during a significant period of time H, the number of headways in the total major traffic stream will be:

$$\text{Total number of headways} = H(q_L + q_R) \quad \text{C62}$$

Now assume that a queue of minor traffic stream vehicles is waiting to turn right and must give way to this major traffic stream, which has a negative exponential headway distribution in each direction. Let the combination of a gap of at least $t_{L,i}$ in the major traffic from the left with a gap of at least $t_{R,i}$ in the major traffic from the right be the minimum condition required to allow i minor traffic stream vehicles to make the right turn within the one gap, where i = 1, 2, 3,....

The probability of the simultaneous occurrence of a headway greater than or equal to $t_{L,i}$ in the major road flow from the left and a headway greater than or equal to $t_{R,i}$ in the major road flow from the right is:

$$\Pr\left(h_L \geq t_{L,i} \mid h_R \geq t_{R,i}\right) = e^{-q_L t_{L,i}} . e^{-q_R t_{R,i}} = e^{-(q_L t_{L,i} + q_R t_{R,i})} \quad \text{C63}$$

Hence, during the significant period of time H, the number of major traffic stream headways large enough to allow at least i minor traffic stream vehicles to make the right turn, i = 1, 2, 3,... is:

$$\text{No. of headways } \left(h_L \geq t_{L,i} \mid h_R \geq t_{R,i}\right) = H(q_L + q_R)e^{-(q_L t_{L,i} + q_R t_{R,i})} \quad \text{C64}$$

Therefore, the number of major traffic stream headways that allow exactly i minor traffic stream vehicles to turn right, i = 1, 2, 3,... is:

$$n_i = H.(q_L + q_R).\left[e^{-(q_L t_{L,i} + q_R t_{R,i})} - e^{-(q_L t_{L,i+1} + q_R t_{R,i+1})}\right] \quad \text{C65}$$

Hence, the total number of minor stream vehicles able to turn right during the period H is:

$$\begin{aligned} N &= \sum_{i=1}^{\infty} i.n_i \\ &= H(q_L + q_R)\sum_{i=1}^{\infty} i.\left[e^{-(q_L t_{L,i} + q_R t_{R,i})} - e^{-(q_L t_{L,i+1} + q_R t_{R,i+1})}\right] \\ &= H(q_L + q_R)\sum_{i=1}^{\infty} e^{-(q_L t_{L,i} + q_R t_{R,i})} \end{aligned} \quad \text{C66}$$

Now assume that a critical gap T_L in the major traffic flow from the left is the minimum that will allow one minor stream right-turner to cross that stream, and that a critical gap T_R in the major traffic flow from the right is the minimum that will allow one minor stream right-turner to join that stream. Further, assume that an additional follow-up headway T_0 is sufficient to allow one additional minor stream vehicle to follow in undertaking the manoeuvre. This implies that:

$$t_{L,i} = T_L + (i-1)T_0 \text{ and } t_{R,i} = T_R + (i-1)T_0, i = 1,2,3,\ldots \quad \text{C67}$$

Substituting from Equations C67 into Equation C66:

$$\begin{aligned} N &= H(q_L + q_R)\sum_{i=1}^{\infty} e^{-(q_L T_L + q_L(i-1)T_0 + q_R T_R + q_R(i-1)T_0)} \\ &= H(q_L + q_R)e^{-(q_L T_L + q_R T_R)}\sum_{i=1}^{\infty} e^{-(q_L + q_R)(i-1)T_0} \\ &= H(q_L + q_R)e^{-(q_L T_L + q_R T_R)}\sum_{k=0}^{\infty}\left(e^{-(q_L + q_R)T_0}\right)^k \\ &= \frac{H(q_L + q_R)e^{-(q_L T_L + q_R T_R)}}{1 - e^{-(q_L + q_R)T_0}} \end{aligned} \quad \text{C68}$$

Thus the theoretical maximum rate at which minor stream vehicles can turn right, that is, the theoretical absorption capacity, is obtained as:

$$C = \frac{N}{H} = \frac{(q_L + q_R)e^{-(q_L T_L + q_R T_R)}}{1 - e^{-(q_L + q_R)T_0}} \quad \text{C69}$$

[Back to body text]

Commentary 10

C10.1 Derivation of Gap Acceptance Formulae for Displaced Negative Exponential Distribution of Headways in the Major Traffic Flow

(a) Probabilities of headways of given size

As discussed in Section 3.3.3, the displaced negative exponential headway distribution postulates a minimum possible headway of β s/veh and is characterised by the probability density function

$$f(t) = \frac{q}{1 - q\beta} e^{\frac{-q(t-\beta)}{1-q\beta}} \text{ for } t \geq \beta \quad \text{C70}$$

and

$$f(t) = 0 \text{ for } t \geq \beta \quad \text{C71}$$

By integration of the function in Equation C70 the probabilities for headway size, h are obtained:

$$Pr(\beta \leq t \leq h) = \int_t^{\infty} f(t).dt = e^{\frac{-q(t-\beta)}{1-q\beta}} \quad \text{C72}$$

and

$$Pr(\beta \leq h \leq t) = \int_{\beta}^{t} f(t).dt = 1 - e^{\frac{-q(t-\beta)}{1-q\beta}} \quad \text{C73}$$

and, of course, by definition:

$$Pr(h < \beta) = 0 \quad \text{C74}$$

Over a significant period of time, H, there will be qH headways in the major stream flow and it follows from Equations C72 to C74 that the number of headways, N, in each of the size ranges will be:

$$N(\beta \leq t \leq h) = qH.e^{\frac{-q(t-\beta)}{1-q\beta}} \quad \text{C75}$$

and

$$N(\beta \leq h \leq t) = qH.(1 - e^{\frac{-q(t-\beta)}{1-q\beta}}) \quad \text{C76}$$

$$N(h < \beta) = 0 \quad \text{C77}$$

(b) Average duration of headways within a given range

Over a period H, the number of headways of duration t to t+dt, where dt is infinitesimally small and $t \geq \beta$, will be:

$$N(t \leq h \leq t + dt) = qH.f(t).dt = qH.\frac{q}{1-q\beta}.e^{\frac{-q(t-\beta)}{1-q\beta}}.dt \quad \text{C78}$$

and the time spent in such headways will be:

$$T(t \leq h \leq t + dt) = t.N(t \leq h \leq t + dt) = t.qH.\frac{q}{1-q\beta}.e^{\frac{-q(t-\beta)}{1-q\beta}}.dt \quad \text{C79}$$

Therefore, the total time spent in headways ≥ t (where $t \geq \beta$) is:

$$T(\beta \leq t \leq h) = qH\int_t^{\infty} t.\frac{q}{1-q\beta}.e^{\frac{-q(t-\beta)}{1-q\beta}}.dt \quad \text{C80}$$

The integral here is of the form $\int t.ae^{-at+b}.dt$, for which the solution is $\left[-e^{-at+b}.\frac{(at+1)}{a^2}\right]$. In this case, $a = \frac{q}{1-q\beta}$ and $b = \frac{q\beta}{1-q\beta}$ and Equation C80 becomes:

$$T(\beta \leq t \leq h) = qH.e^{\frac{-q(t-\beta)}{1-q\beta}}.\left(\frac{1}{q} + t - \beta\right) \quad \text{C81}$$

The average duration of headways greater than or equal to t (where t ≥ β) can now be obtained as the total time spent in such headways (Equation C81) divided by the number of such headways (Equation C75), that is:

$$h_{av}(\beta \leq t \leq h) = \frac{1}{q} + t - \beta \quad \text{C82}$$

By a similar analysis, noting that the total time spent in headways ≤ t (where $t \geq \beta$) is H minus the time given by Equation C81and that the number of such headways is as in Equation C76, Equation C83 is derived:

$$h_{av}(\beta \leq h \leq t) = \frac{1}{q} - \frac{(t-\beta)e^{\frac{-q(t-\beta)}{1-q\beta}}}{1-e^{\frac{-q(t-\beta)}{1-q\beta}}} \quad \text{C83}$$

(c) Average delays to minor traffic

As in Commentary 5, the average delay at the gap acceptance point for all minor traffic stream units can be derived by noting that the probability of any unit having to wait for n gaps, each less than T, before being able to proceed is given by the geometric distribution as:

$$P(n) = (1-p)p^n \; n = 0, 1, 2, \ldots \quad \text{C84}$$

where (drawing on Equation C73)

$$p = Pr(\text{gap} < T) = 1 - e^{\frac{-q(T-\beta)}{1-q\beta}} \quad \text{C85}$$

so that

$$P(n) = e^{\frac{-q(T-\beta)}{1-q\beta}}(1 - e^{\frac{-q(T-\beta)}{1-q\beta}})^n \quad \text{C86}$$

The expected number of gaps less than T for which a minor traffic stream unit has to wait before being able to proceed is then given by:

$$E(n) = 0.P(0) + 1.P(1) + 2.P(2) + 3.P(3) + \ldots \quad \text{C87}$$

$$= e^{\frac{-q(T-\beta)}{1-q\beta}}(1 - e^{\frac{-q(T-\beta)}{1-q\beta}}) + 2e^{\frac{-q(T-\beta)}{1-q\beta}}(1 - e^{\frac{-q(T-\beta)}{1-q\beta}})^2 + 3e^{\frac{-q(T-\beta)}{1-q\beta}}(1 - e^{\frac{-q(T-\beta)}{1-q\beta}})^3 + \ldots$$

$$E(n) = \frac{1 - e^{\frac{-q(T-\beta)}{1-q\beta}}}{e^{\frac{-q(T-\beta)}{1-q\beta}}}$$

Now the average duration of headways less than T in the major traffic stream is given by Equation C83 as:

$$h_{av}(\beta \le h \le t) = \frac{1}{q} - \frac{(t-\beta)e^{\frac{-q(t-\beta)}{1-q\beta}}}{1 - e^{\frac{-q(t-\beta)}{1-q\beta}}} \quad \text{C87}$$

Then the average delay experienced by all minor traffic stream units at the gap acceptance point is:

$$d_{av}(d \ge 0) = E(n).[h_{av}(h < T)]$$

$$= \frac{(1 - e^{\frac{-q(T-\beta)}{1-q\beta}})}{e^{\frac{-q(T-\beta)}{1-q\beta}}} \frac{\left(1 - e^{\frac{-q(T-\beta)}{1-q\beta}} - q(T-\beta)e^{\frac{-q(T-\beta)}{1-q\beta}}\right)}{q.(1 - e^{\frac{-q(T-\beta)}{1-q\beta}})}$$

which simplifies to:

$$d_{av}(d \ge 0) = \frac{1}{qe^{\frac{-q(T-\beta)}{1-q\beta}}} - \frac{1}{q} - (T-\beta) \quad \text{C88}$$

The average delay at the gap acceptance point **to only those minor traffic stream units that do experience such delay** is obtained by dividing the average delay to all minor stream units (as in Equation C88) by the proportion experiencing non-zero delay (equal to the probability of a headway < T, as given by Equation C73) to give:

$$d_{av}(d>0)=\frac{(1-e^{\frac{-q(T-\beta)}{1-q\beta}})}{e^{\frac{-q(T-\beta)}{1-q\beta}}}\frac{\left(1-e^{\frac{-q(T-\beta)}{1-q\beta}}-q(T-\beta)e^{\frac{-q(T-\beta)}{1-q\beta}}\right)}{q.(1-e^{\frac{-q(T-\beta)}{1-q\beta}})}\frac{1}{(1-e^{\frac{-q(T-\beta)}{1-q\beta}})}$$

which simplifies to:

$$d_{av}(d>0)=\frac{1}{qe^{\frac{-q(T-\beta)}{1-q\beta}}}-\frac{(T-\beta)}{(1-e^{\frac{-q(T-\beta)}{1-q\beta}})} \quad \text{C89}$$

(d) Absorption capacity

To derive the theoretical absorption capacity, assume that a queue of minor traffic stream vehicles is waiting to cross or enter a major traffic stream that has volume q and a displaced negative exponential headway distribution, and let t_i be the minimum gap required to allow i minor traffic stream vehicles to carry out the manoeuvre within the one gap, where i = 1, 2, 3,...

During a significant period of time, H, the number of major road headways greater than or equal to t_i, i = 1, 2, 3,... is:

$$\text{No. of headways} \geq t_i = H.qe^{\frac{-q(t_i-\beta)}{1-q\beta}} \quad \text{C90}$$

Therefore, the number of major traffic stream headways that allow exactly i minor traffic stream vehicles to cross or enter the major stream, i = 1, 2, 3,... is:

$$n_i = H.(qe^{\frac{-q(t_i-\beta)}{1-q\beta}}-qe^{\frac{-q(t_{i+1}-\beta)}{1-q\beta}}) \quad \text{C91}$$

Hence, the total number of minor stream vehicles able to cross or join the major stream during the period H is:

$$\begin{aligned} N &= \sum_{i=1}^{\infty} i.n_i \\ &= \sum_{i=1}^{\infty} i.Hq(e^{\frac{-q(t_i-\beta)}{1-q\beta}}-e^{\frac{-q(t_{i+1}-\beta)}{1-q\beta}}) \\ &= Hq\sum_{i=1}^{\infty} e^{\frac{-q(t_i-\beta)}{1-q\beta}} \end{aligned} \quad \text{C92}$$

Taking the critical gap ,T, as the minimum headway that will allow one minor stream vehicle to cross or join the major stream, assume that each additional time interval T_0 in the size of the headway is sufficient to allow one additional minor stream vehicle to follow in undertaking the manoeuvre. T_0 is known as the follow-up headway and the above assumption implies that:

$$t_i = T + (i-1)T_0, i = 1,2,3,\ldots \quad \text{C93}$$

Substituting from Equation C93 into Equation C92, gives:

$$\begin{aligned} N &= Hq\sum_{i=1}^{\infty} e^{\frac{-q(T+(i-1)T_0-\beta)}{1-q\beta}} \\ &= Hqe^{\frac{-q(T-\beta)}{1-q\beta}}\sum_{k=0}^{\infty}\left(e^{\frac{-qT_0}{1-q\beta}}\right)^k \\ &= \frac{Hqe^{\frac{-q(T-\beta)}{1-q\beta}}}{1-e^{\frac{-qT_0}{1-q\beta}}} \end{aligned} \quad \text{C94}$$

Thus the theoretical maximum rate at which minor stream vehicles can cross or join the major traffic stream, that is, the theoretical absorption capacity, is obtained as:

$$C = \frac{N}{H} = \frac{qe^{\frac{-q(T-\beta)}{1-q\beta}}}{1-e^{\frac{-qT_0}{1-q\beta}}} \quad \text{C95}$$

[Back to Section 3]
[Back to Section 5]

Austroads' *Guide to Traffic Management* captures contemporary traffic management practice including emerging techniques and technologies, and relevant international experience. It provides valuable guidance to practitioners in managing road traffic efficiently, safely and economically.

Guide to Traffic Management Part 2: Traffic Theory Concepts provides an overview of the concepts used in traffic management in Australia and New Zealand. It helps practitioners to understand the theoretical concepts regarding traffic behaviour which can be used to develop strategies to manage traffic.

Guide to Traffic Management Part 2

Austroads

Austroads is the association of Australasian road and transport agencies.

www.austroads.com.au