**Shockwave Theory**

**1. Shockwave**

Shockwaves are an important concept in traffic operational analysis. They can be used to describe the propagation of queues on a freeway, as well as their dissipation. Shockwaves are also used in the theory describing operations at signalized intersections, and thus find application across both uninterrupted and interrupted flow domains in traffic

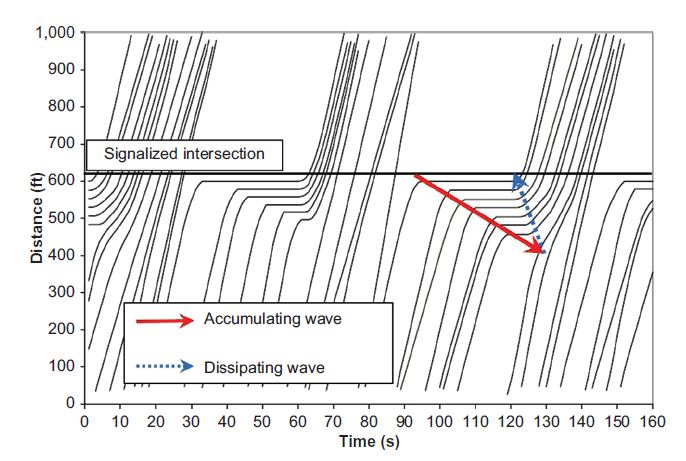
operations.

Shockwaves occur at the transition from one traffic state to another, say, free-flow to stopped, or from stopped back to flowing traffic. This application is common for freeway bottlenecks and incidents, which cause vehicles to drastically slow down or come to a full stop. On arterial streets, the same pattern is found at signalized intersection, where a red signal forces a transition from one traffic state (moving) to another (stopped). But shockwaves also occur between two moving regimes. For example, a slow moving tractor-trailer on a freeway may force a slower flow regime in its immediate proximity, distinguishing it from faster-speed, freely moving traffic in front or behind it. Another example would be a freeway section, in which vehicles tend to slow down (e.g., due to an upgrade, a bridge, a tunnel, sun glare, etc.), resulting in a changed traffic state and (depending on the demand level) congestion upstream of this section.

Figure below illustrates accumulating and dissipating waves at a signalized intersection, drawn on a time-space diagram of individual vehicle trajectories. As vehicles approach the (red) signal, they have to slow down until the signal turns green, and can then reaccelerate. The result of multiple stopping vehicles in sequence is a shockwave that describes the propagation of the back of the queue upstream of the signal.

Similarly, once the light turns green, another wave describes the rate at which vehicles discharge from congestion at the signal. It is very evident that the queue clears from the front (signal stop bar) toward the upstream end of the queue. This is also referred to as a front-clearing queue.

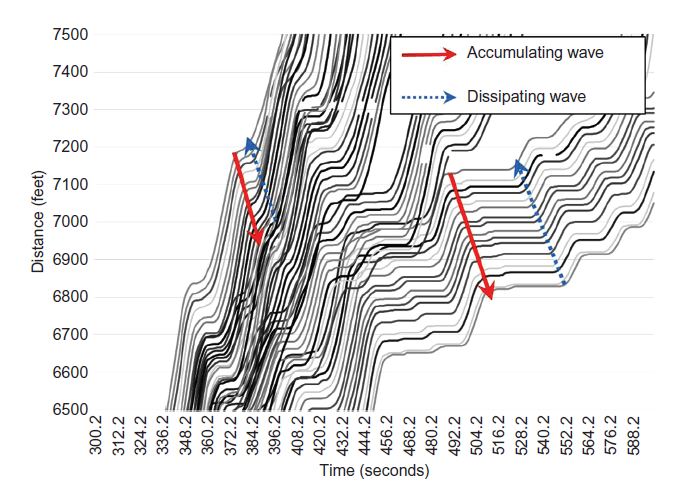
From the simple shockwave diagram in Figure below, the maximum back of queue is found mathematically by the intersection of the accumulating and dissipating waves. Further, the total delay of all vehicles in one cycle is described by the area of the triangle, with the average delay per vehicle being defined as the total area divided by the number of vehicles per cycle. This example is simplified, showing an undersaturated signal (the queue clears each cycle) and a random arrival profile. In reality, arrivals are expected to be at least partially platooned, resulting in multiple flow regimes that yield multiple accumulating waves, and an area of the delay region that is more complicated than a simple triangle. These concepts are discussed further in a later section.



**Figure:** Shockwaves at a signalized intersection.

For freeway operations, congestion often forms through the interaction of different drivers and vehicles in near-capacity situations. So rather than having a clear bottleneck (and as a result, clear congestion patterns), the turbulence in a heavy traffic stream on freeways is often more random

Sometimes, this results in congestion without apparent cause, a phenomenon that many freeway commuters are all too familiar with. The random variation of vehicle headways and random turbulence can result in breakdown, even without any apparent bottleneck, incident, or other “reason” for the congestion. Figure below illustrates this with a plot of simulated vehicle trajectories (space vs. time) on a congested freeway. Rather than having a clear bottleneck location, and clean resulting shockwaves as was the case in the signalized intersection example, the congested freeway shows multiple shockwaves resulting from turbulence and localized breakdown events. From an aerial view, this could be visualized as multiple waves of brake lights propagating upstream through the traffic stream.



**Figure**: Freeway trajectories and shockwaves.

The speed of the congestion shockwave then describes the speed at which congestion, or more precisely the back of queue, travels backward from the bottleneck location. Shockwaves are often readily observed visually through a wave of brake lights, an effect that is even more dramatic when observed from a bridge, tall building, or a helicopter. In fact,

the title page of this part of the book shows a freeway section with congestion on the far hill crest, which, over time, spills back to the position from where the picture is taken. Some isolated brake lights are already visible.

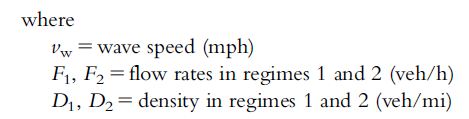
Mathematically, the speed of a shockwave can be calculated from the flow rates and densities of the upstream and downstream traffic states.

Recall the fundamental relationship of traffic flow, as:

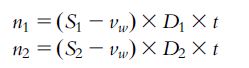
flow (F)=speed (S) x density (D)

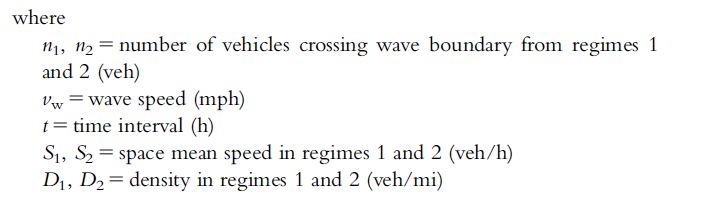
The speed of the wave between regimes 1 and 2 then is calculated as:





The number of vehicles crossing from regime 1 into regime 2 in a time interval, t, (or vice versa from regime 2 into regime 1) can be calculated as:





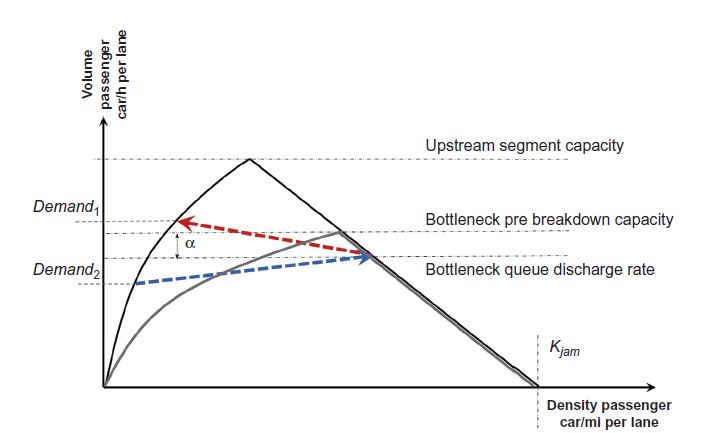
From equation of shockwave above it is evident that the shockwave speed can be either positive or negative. The sign of the shockwave speed goes with the direction of traffic, to where a positive shockwave speed travels downstream, while a negative shockwave speed means the wave is traveling upstream.

Most congestion wave speeds (from a bottleneck) are thus negative, as traffic accumulates and forms a queue upstream of the bottleneck, and the back of queue travels further upstream over time at the wave speed.

Dissipating shockwaves when congestion clears can either be positive or negative. A positive shockwave is one that occurs when demand at a fixed bottleneck location begins to drop after the peak congestion, and the queue therefore clears from the back. These “back-clearing” queues have a positive wave speed. On the other hand, “front-clearing” queues occur when the bottleneck itself is lifted, while the demand stays fixed. In this case, the queue clears from the front with a negative wave speed, meaning that the front of the queue travels upstream. Front-clearing queues are common on freeways when the bottleneck is due to an incident that at some point clears. Front-clearing queues are also common at signalized intersections, where the bottleneck is lifted once the signal turns green.

This is apparent in previous Figure, where both the forming and dissipating waves travel upstream from the bottleneck with a negative wave speed. In this case, the dissipating wave (lower dashed line) is faster than the forming wave (upper dashed line), which means the queue fully clears when the faster wave catches up with the slower one.

An illustration of shockwave theory applied to a freeway bottleneck is shown in Figure below. It shows a hypothetical case of an upstream segment (thin black line) and a downstream bottleneck with lower capacity (thick gray line). On the y-axis, the figure shows two demands: D1, which is a high demand above the bottleneck capacity that causes congestion, and D2, which is a lower demand below the bottleneck capacity that causes congestion to dissipate. The figure further distinguishes between the prebreakdown capacity and the queue discharge rate, recognizing a drop in throughput after a bottleneck is activated.

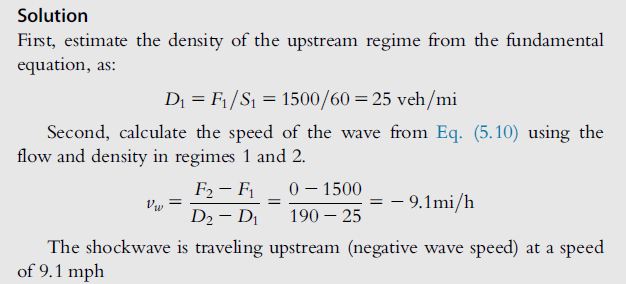


**Figure:** Illustration of shockwaves for freeway bottleneck.

In the figure, two shockwaves are illustrated. First the forming or accumulating wave that is shown as a light gray dashed arrow between the bottleneck queue discharge rate and the upstream segment (high) demand, D1. This shockwave moves against the direction of traffic and thus has a negative sign. The second wave is the dissipating wave, which is drawn as a dark gray dashed arrow between the (lower) demand D2 in the upstream segment, and the queue discharge rate. This wave travels in the direction of traffic and therefore has a positive sign.

**Example**

A freeway segment has a flow rate of 1500 veh/h (per lane) traveling at a speed of 60 mph. This traffic stream arrives at a crash location that results in a full closure of the facility, resulting in a queue at the jam density (maximum density) of 180 veh/mi (per lane). Estimate the speed at which the shockwave grows upstream of the crash location.



**2. Gap Acceptance**

Another fundamental theoretical concept in traffic operations is gap acceptance. Gap acceptance applies whenever the interaction of two traffic streams is not fully controlled through grade separation or signalization.

Common examples of where gap acceptance theory is applied are in evaluating the capacity and operations of a modern roundabout, a stop or yield-controlled intersection, or a permissive right- or left-turning movement at a signalized intersection. Gap acceptance theory equally applies to pedestrians trying to get across the street, or bicycles trying to cross or merge into a traffic stream from a side street.

The parameters describing gap acceptance are critically linked to driver behavior and Human Factors Principles. gap acceptance is linked to the level of comfort and safety perception of drivers. In other words, a more aggressive driver is likely to accept shorter gaps than a more conservative driver. Vehicle dynamics also play into gap acceptance, to where a slower vehicle (e.g., a truck or another heavy vehicle) is likely to require a longer gap than a small sports car. Similarly, a runner is likely to require a shorter gap than a pedestrian walking more slowly. When trying to describe gap acceptance for a traffic stream, this diversity of characteristics and dynamics attributes needs to be taken into account.

The two fundamental parameters in gap acceptance theory are:

(1) critical gap, and

(2) follow-up headway.

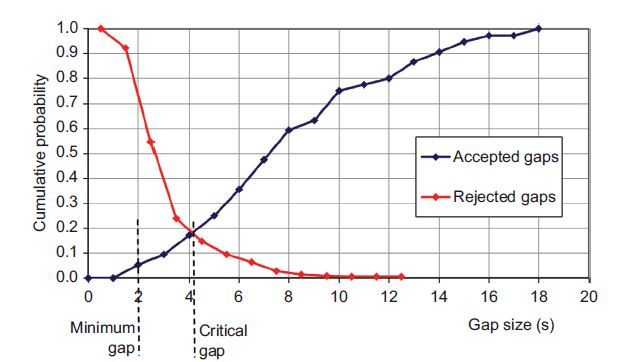
They are defined as follows:

1. Critical gap is the gap time in seconds at which drivers are equally likely to accept or reject the gap and enter the conflicting traffic stream. It is mathematically used to estimate the time needed for the first vehicle to enter the conflicting traffic stream.

2. Follow-up headway is the additional time beyond the critical gap needed for each additional vehicle to enter the conflicting traffic stream.

For example, if the critical gap (for the first vehicle) is 5 s, and the follow-up headway is 2 s, it is expected that two queued vehicles could enter in a gap of 7 s, three vehicles in a gap of 9 s, and so on.

The critical gap is different from the minimum gap needed for any vehicle to enter the traffic stream, although the two are commonly confused. The critical gap concept recognizes that the traffic stream is nonhomogeneous, and that different drivers have different risk thresholds. As such, a study measuring all accepted gaps would not find one value, but a distribution of values. Similarly, that study would find a distribution of rejected gaps. If plotting these two cumulative distributions, the critical gap is defined as the intercept of the two curves, or the gap size where drivers are equally likely to accept or reject the gap. This is illustrated graphically in Figure below based on a pedestrian gap acceptance study



**Figure:** Sample critical gap study using graphical method.

The figure above further shows the minimum gap (i.e., the smallest observed accepted gap), which is mathematically (and practically) very different from the critical gap. The illustration in Figure above is also referred to as the graphical method for estimating critical gap as described in (Schroeder et al., 2010). Other methods for estimating critical gap are a method based on maximum likelihood estimation (MLE), which estimates both a mean and standard deviation of the critical gap assuming a normal distribution, and the Ramsey-Routledge method, which can be used to estimate any distribution of critical gaps, including bimodal distributions (e.g., a pedestrian stream with half walkers and half runners). All three methods are described in detail in (Schroeder et al., 2010).

**3. Simple Queuing Theory**

Queuing theory finds frequent application in transportation operations to estimate queue lengths at intersections. Queue lengths are an important performance measure, as they directly reflect the experience of drivers, and can further be linked to important design considerations. For example, the estimated queue length of a left turn at a signalized intersection is a key piece of information to design the length of the left-turn pocket. Similarly, queue lengths are critically important for two closely spaced intersections to estimate the potential for the queue from one signal to spillback into the upstream signal, thereby potentially impacting its operations significantly.

Queue spillback is also a key consideration for freeway off-ramps, where the queue from the interchange ramp terminal should be contained within the off-ramp length if at all possible, and should not spill back onto the freeway, which would have big safety and operational implications.

Queuing theory has many different applications beyond transportation, and different queueing models exist to describe queue as a function of the number of channels approaching the queuing location, the number of servers able to process the queue, the arrival demands, the time needed to process each entity, and the priority order of processing entities. But while queuing theory can get quite complex, luckily most queues in transportation are simple first in/first out (FIFO) queues, and are often assumed to have fixed service times (on average).

In traffic operations, queues (and most other performance measures) can either be obtained deterministically (through the use of analytical equations), or can be simulated. The focus in the following sections is on the former, although simulation concepts are introduced in a later section.

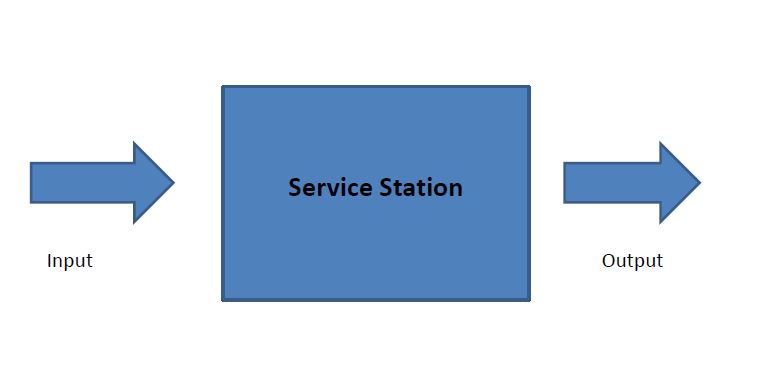
**3.1 Queuing Theory and Traffic Flow Analysis**

Purpose

To provide a means to estimate important measures of highway performance including

vehicle delay and traffic queue lengths. e.g. the required length of left turning bays.

**Queuing System**

****

**Assumptions in Queuing Models**

• Arrival Pattern.

• Departure Characteristics.

• Queue disciplines.

Two Possible Traffic Arrival Patterns or Distribution:

• Equal time intervals (derived from the assumption of uniform deterministic arrival)

• Exponentially distributed time intervals (derived from the assumption of Poisson distributed arrivals).

Assumptions for Vehicle Departure Characteristics or Distribution of Departure

• Distribution of the amount of time it takes a vehicle to depart on a particular service

center.

• The number of service stations or departure channels.

• Queuing Discipline.

• First –In- First- Out (FIFO), indicating that the first vehicle to arrive is the first vehicle to

depart.

• Last –In – First – Out (LIFO), indicating that the last vehicle to arrive is the first to depart.

**Queuing Models**

It is identified by 3 alphanumeric values:

• The first value indicates the arrival rate assumption.

• The second value gives the departure rate assumption.

• The third value indicates the number of departure channels.

D – is the traffic arrival and departure assumptions which is the uniform, deterministic distribution

M – is the traffic arrival and departure assumption which is exponential distribution or Markov Hence,

D/D/1 Model means the uniform, deterministic arrival and departure with one channel

D/M/1 Model means the uniform deterministic arrival and exponentially distributed departure with one channel.

**4 Queuing Models in Traffic Analysis**

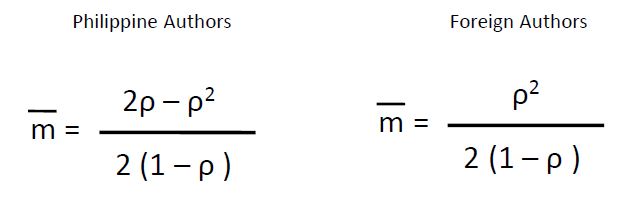
1. D/D/1 Queuing – simple system and could be graphically and mathematically solve.
2. M/D/1 Queuing for traffic intensity or density (ρ) that is less than 1 for the system to be stable.
3. M/M/1 Queuing for traffic intensity or density (ρ) that is less than 1 for the system to be stable.
4. M/M/N Queuing for traffic intensity or density (ρ) is greater than 1 and ρ/N (utilization factor) maybe greater than 1 for the system to be stable.

**1. D/D/1 Queuing**

Use Cumulative Plot Method

**2. M/D/1 Queuing**

a. Average length of Queues

****

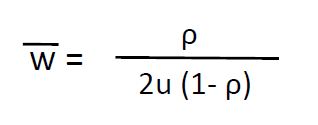
where:

m = average length of queues.

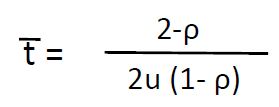
ρ = traffic density or utilization factor.

= λ / u (λ – arrival rate and u – departure rate).

b. Average waiting time in the system



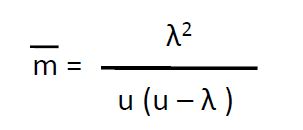
c. Average time spent in the system



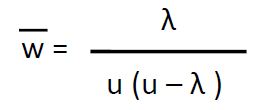
It is the summation of average queue waiting time and average departure time or service time.

**3. M/M/1 Queuing**

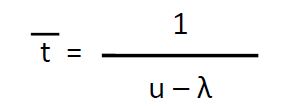
a. Average length of queue



b. Average waiting time in the system

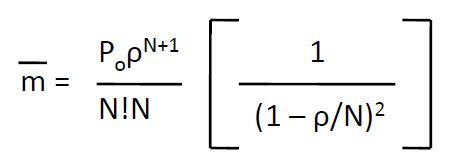


c. Average time spent in the system

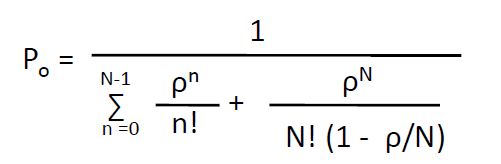


**4. M/M/N Queuing**

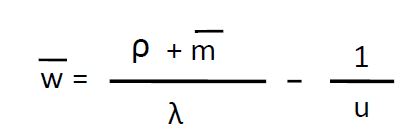
a. Average length of queue



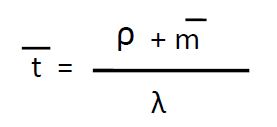
Where:



b. Average waiting time in the system



c. Average time spent in the system



**Probability of waiting in a Queue**

The probability of being in a queue, which is the probability that the number of vehicles in a system, n, is greater than the number of departure channels, N

