Transportation Supply Models

Introduction
This lecture deals with the mathematical models simulating transportation supply systems. In broad terms a transportation supply model can be defined as a model, or rather a system of models, simulating the performances and flows resulting from user demand and the technical and organizational aspects of the physical transportation supply.
Transportation supply models combine traffic flow theory and network flow theory models. The former is used to analyze and simulate the performances of the main supply elements, the latter to represent the topological and functional structure of the system.

Traffic Flow Theory
Models derived from traffic flow theory simulate the effects of interactions between vehicles using the same transportation facility (or service) simultaneously. For simplicity's sake, the models presented refer to vehicle flow, although most of them can be applied to other types of users, such as trains, planes, and pedestrians. In the sections below we describe stationary uninterrupted flow models, followed by models of interrupted flow, derived from queuing theory.

Uninterrupted Flows
Multiple vehicles using the same facility may interact with each other and the effect of their interaction will increase with the number of vehicles. This phenomenon, called congestion, occurs in most transportation systems, generally worsening the overall performance of the facility, such as the mean speed or travel time. Indeed, a vehicle may be forced to move at less than its desired speed if it encounters a slower vehicle. The higher the number of vehicles on the infrastructure, the more likely this condition is to happen. This circumstance may also occur in transportation systems with scheduled services: the higher the number of vehicles on the infrastructure, the more likely out-of-schedule vehicles are to cause a delay to other vehicles.
In general, stochastic models may be used to characterize an interaction event that causes a delay in a probabilistic sense. For congested systems with continuous services, it is very often sufficient to adopt the aggregate deterministic models described below; they may be applied in areas far away from interruptions such as intersections and toll booths.

Fundamental Variables
Several variables can be observed in a traffic stream, that is, a sequence of cars moving along a road segment referred to as a link, a. In principle, although all variables should be related to link a, to simplify the notation the subscript a may be implied. The fundamental variables are as follows (see Fig.1).

τ The time at which the traffic is observed
L_a The length of the road segment corresponding to linking a
s A point along a link, or rather, its abscissa increasing (from a given origin, usually located at the beginning of the link) along the traffic direction ($s \in [0, L_a]$)
i An index denoting an observed vehicle
$v_i(s, \tau)$ The speed of vehicle i at time τ while traversing point (abscissa) s

Several variables can be defined for traffic observed at point s during the time interval [τ,τ + Δτ] (see Fig. 1) as follows.

$h_i(s)$ The headway between vehicles i and i − 1 crossing point s
m(s | τ, τ + Δτ) The number of vehicles traversing point s during the time interval [τ, τ + Δτ]
$\bar{h}(s) = \sum_{i=1,\ldots m} h_i(s)/m(s \mid \tau, \tau + \Delta\tau)$ The mean headway, among all vehicles crossing point s during the time interval [τ, τ + Δτ]
$\bar{v}(s) = \sum_{i=1,\ldots m} v_i(s)/m(s \mid \tau, \tau + \Delta\tau)$ The time mean speed, among all vehicles crossing point s during the time interval [τ, τ + Δτ]

Similarly, for traffic observed at time τ between points s and s+Δs, the following variables can be defined:

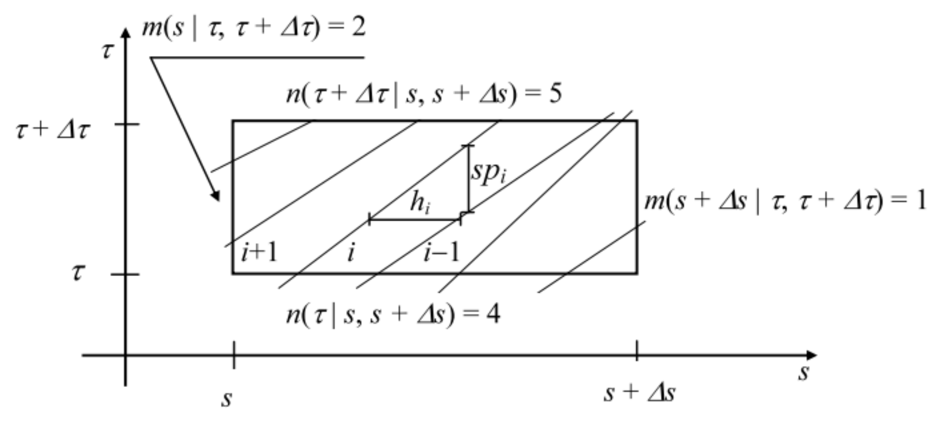

Fig. 1 Vehicle trajectories and traffic variables.

$sp_i(\tau)$ The spacing between vehicles i and i − 1 at time τ
n(τ | s, s + Δs) The number of vehicles at time τ between points s and s + Δs
$\overline{sp}(\tau) = \sum_{i=1,\ldots m} sp_i(\tau)/n(\tau \mid s, s + \Delta s)$ The mean spacing, among all vehicles between points s and s + Δs at time τ
$\overline{v_s}(\tau) = \sum_{i=1,\ldots m} v_i/n(\tau \mid s, s + \Delta s)$ The space mean speed, among all vehicles between points s and s + Δs at time τ

During the time interval [τ, τ + Δτ] between points s and s + Δs, a general flow conservation equation can be written:

$$\Delta n(s, s + \Delta s, \tau, \tau + \Delta\tau) + \Delta m(s, s + \Delta s, \tau, \tau + \Delta\tau) = \Delta z(s, s + \Delta s, \tau, \tau + \Delta\tau) \quad 1$$

Where:
$\Delta n(s, s + \Delta s, \tau, \tau + \Delta\tau) = n(\tau + \Delta\tau \mid s, s + \Delta s) - n(\tau \mid s, s + \Delta s)$ is the variation in the number of vehicles between points s and s + Δs during Δτ
$\Delta m(s, s + \Delta s, \tau, \tau + \Delta\tau) = m(s + \Delta s \mid \tau, \tau + \Delta\tau) - m(s \mid \tau, \tau + \Delta\tau)$ is the variation in the number of vehicles during time interval [τ,τ + Δτ] over space Δs
$\Delta z(s, s + \Delta s, \tau, \tau + \Delta\tau)$ is the number of entering minus exiting vehicles (if any) during the time interval [τ,τ + Δτ], due to entry/exit points (e.g., on/off ramps), between points s and s + Δs.

In the example of Fig. 1 no vehicles are entering/exiting in the segment Δs; then Δz = 0 (Δn is equal to 1 and Δm is equal to −1).

With the observed quantities two relevant variables, flow, and density can be introduced:

$f(s \mid \tau, \tau + \Delta\tau) = m(s \mid \tau, \tau + \Delta\tau)/\Delta\tau$ is the flow of vehicles crossing point s during the time interval [τ, τ + Δτ], measured in vehicles per unit of time.
$k(\tau \mid s, s + \Delta s) = n(\tau \mid s, s + \Delta s)/\Delta s$ is the density between points s and s + Δs at time τ, measured in vehicles per unit of length.

Flow and density are related to mean headway and mean spacing through the following relations.

$f(s \mid \tau, \tau + \Delta\tau) \cong 1/h(s)$
$k(\tau \mid s, s + \Delta s) \cong 1/sp(\tau)$

Note that if observations are perfectly synchronized with vehicles, the near-equality in the previous two equations becomes proper equality.

Moreover, if the general flow conservation equation (1) is divided by Δτ, the following equation is obtained.

$$\Delta n/\Delta\tau + \Delta f = \Delta e \qquad 2$$

Where
$\Delta f(s, s + \Delta s, \tau, \tau + \Delta\tau) = \Delta m(s, s + \Delta s, \tau, \tau + \Delta\tau)/\Delta\tau$ is the variation of the flow over space
$\Delta e(s, s + \Delta s, \tau, \tau + \Delta\tau) = \Delta z(s, s + \Delta s, \tau, \tau + \Delta\tau)/\Delta\tau$ is the (net) entering/ exiting flow

Finally, dividing by Δs, we obtain a further formulation of (1) that expresses the role of variation in density:

$$\Delta k/\Delta\tau + \Delta f/\Delta s = \Delta e/\Delta s \qquad 3$$

Where:
$\Delta k(s, s + \Delta s, \tau, \tau + \Delta\tau) = \Delta n(s, s + \Delta s, \tau, \tau + \Delta\tau)/\Delta s$ is the variation of the density over time.

Model Formulation

In this section, we describe several deterministic models developed under the assumption of stationarity, formally introduced below. In formulating such models it is assumed that a traffic stream (a discrete sequence of vehicles) is represented as a continuous (one-dimensional) fluid. Traffic flow is called stationary during a time interval [τ, τ +Δτ] between points s and s + Δs if the flow is (on average) independent of point s, and density is independent of time τ (other definitions are possible):

$f(s \mid \tau, \tau + \Delta\tau) = f$
$k(\tau \mid s, s + \Delta s) = k$

Note that this condition is chiefly theoretical and in practice can be observed only approximately for mean values in space or time. It is nevertheless useful in that it allows effective analysis of the phenomenon. In this case, the time mean speed is independent of location and the space mean speed is independent of time:

$\bar{v}_\tau(s) = \bar{v}_\tau$

$\bar{v}_s(\tau) = \bar{v}_s$

In the case of stationarity, both terms on the left side of the conservation equation (3) are identically null, anyhow other flow conservation conditions may be formulated. Hence, let n = k · Δs be the number, time-independent due to the assumption of stationarity, of vehicles on the stretch of road between cross-sections s and s + Δs, and let $\bar{v}_s$ be the space mean speed of these vehicles. The vehicle that at time τ is at the start of the stretch of road, cross-section s, will reach the end, cross-section s + Δs, on average at time τ + Δτ ′, with Δτ ′ = Δs/vs. Due to the assumption of stationarity, the number of vehicles crossing each cross-section during time Δτ is equal to f · Δτ. Thus the number of vehicles contained at time τ on section [s, s +Δs] is equal to the number of vehicles traversing cross-section s +Δs during the time interval [τ, τ +Δτ ′] (see Fig. 2); that is, kΔs = f Δτ ′ = f Δs/vs. Hence, under stationary conditions, flow, density, and space mean speed must satisfy the stationary flow conservation equation:

$$f = kv \qquad 4$$

Where
$v = \bar{v}_s$ is the space mean speed, simply called speed for further analysis of stationary conditions.

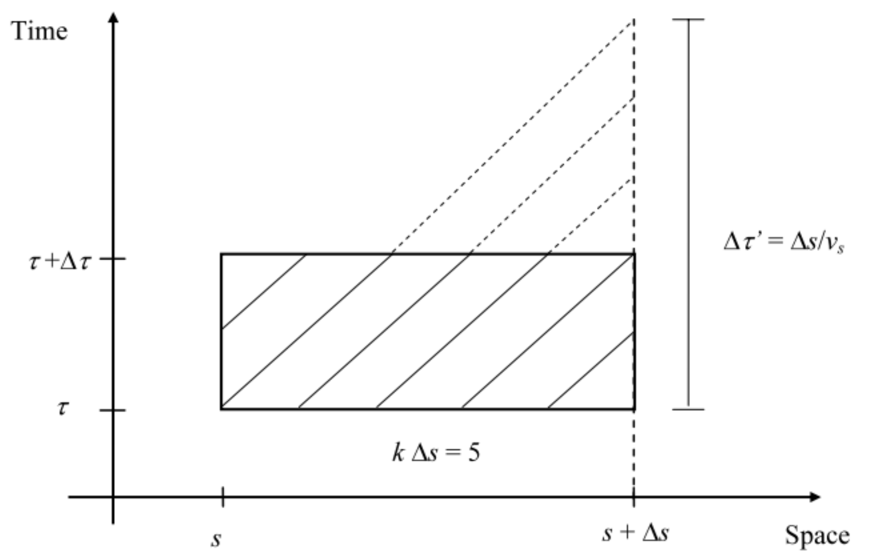

Fig. 2 Vehicle trajectories and traffic variables for stationary (deterministic) flows.

In stationary conditions, empirical relationships can be observed between each pair of variables: flow, density, and speed. In general, observations are rather scattered (see Fig. 3 for an example of a speed–flow empirical relationship) and various models may be adapted to describe such empirical relationships. These models are generally given the name fundamental diagram (of traffic flow) (see Fig. 4) and are specified by the following relations.

$$v = V(k) \quad (5)$$
$$f = f(k) \quad (6)$$
$$f = f(v) \quad (7)$$

Although only a model representation of empirical observations, this diagram permits some useful considerations to be made. It shows that flow may be zero under two conditions: when density is zero (no vehicles on the road) or when speed is zero (vehicles are not moving). The latter corresponds in reality to a stop-and-go condition.
In the first case, the speed assumes the theoretical maximum value, free-flow speed v_0, whereas in the second the density assumes the theoretical maximum value jam density, k_{jam}. Therefore, a traffic stream may be modeled through a partially compressible fluid, that is, a fluid that can be compressed up to a maximum value.

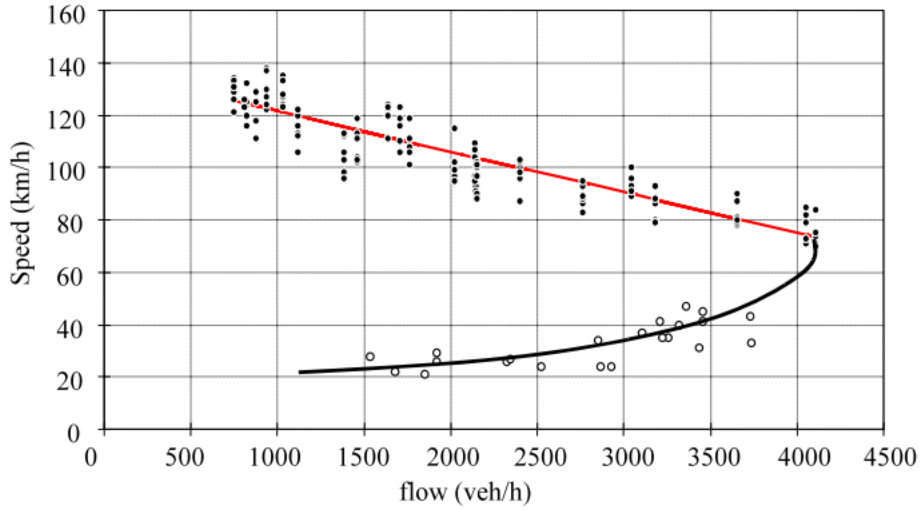

Fig.3 Relationship between speed and flow.

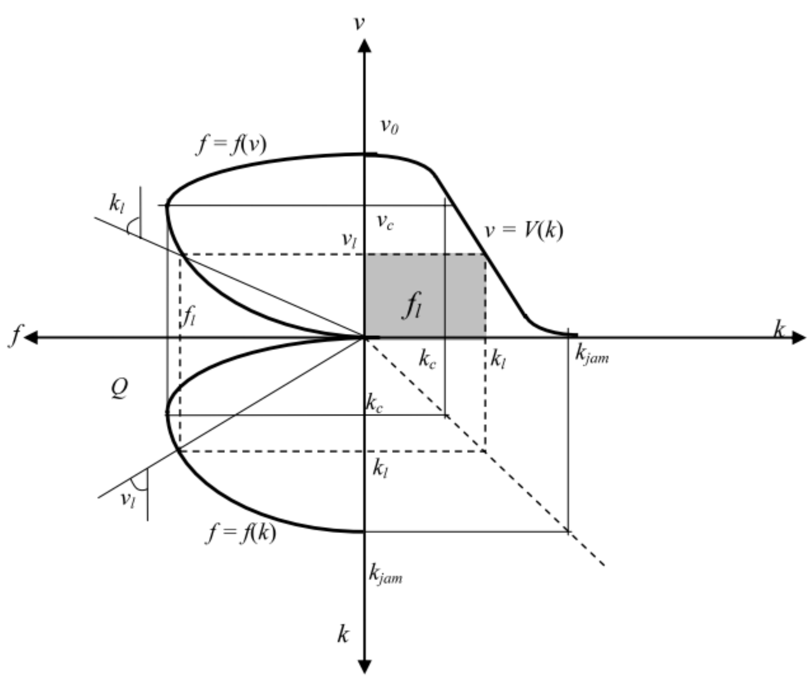

Fig. 4 Fundamental diagram of traffic flow.

The peak of the speed–flow (and density–flow) curve occurs at the theoretical maximum flow, capacity Q of the facility; the corresponding speed v_c and density k_c are referred to as the critical speed and the critical density. Thus any value of flow (except the capacity) may occur under two different conditions: low speed and high density and high speed and low density. The first condition represents an unstable state for the traffic stream, where any increase in density will cause a decrease in speed and thus in flow. This action produces another increase in density and so on until traffic becomes jammed. Conversely, the second condition is a stable state because any increase in density will cause a decrease in speed and an increase in flow. At capacity (or at critical speed or density) the stream is nonstable, this being a boundary condition between the other two. These results show that flow cannot be used as the unique parameter describing the state of a traffic stream; speed and density, instead, can univocally identify the prevailing traffic condition. For this reason, the relation v = V (k) is preferred to study traffic stream characteristics.
Mathematical formulations have been widely proposed for the fundamental diagram, based on the single regime or multi regimes functions. An example of a single regime function is Greenshields' linear model:

$$V(k) = v_0(1 - k/k_{jam})$$

Or Underwood's exponential model (useful for low densities):

$$V(k) = v_0 e^{-k/kc}.$$

An example of a multi-regime function is Greenberg's model:

$$V(k) = a_1 \ln(a_2/k) \text{ for } k>k_{min}$$

$$V(k) = a_1 \ln(a_2/k_{min}) \text{ for } k \le k_{min}$$

Where a_1, a_2 , and $k_{min} \le k_{jam}$ are constants to be calibrated.

Starting from the speed–density relationship, the flow–density relationship, f = f (k), may be easily derived by using the flow conservation equation under stationary conditions, or fundamental conservation equation (4):

$$f(k) = V(k)k$$

Greenshields' linear model yields:

$$f(k) = v_0(k - k^2/k_{jam})$$

In this case, the capacity is given by:

$$Q = v_0 k_{jam}/4$$

Moreover the flow–speed relationship can be obtained by introducing the inverse speed–density relationship: $k = V^{-1}(v)$, thus:

$f(v) = V(k = V^{-1}(v)) \cdot V^{-1}(v) = v \cdot V^{-1}(v)$

For example, Greenshields' linear model yields: $V^{-1}(v) = k_{jam}(1 - v/v_0)$ thus:

$f(v) = k_{jam}(v - v^2/v_0)$

In general, the flow–speed relationship may be inverted by only considering two different relationships, one in a stable regime, $v \in [v_c, v_o]$, and the other in an unstable regime, $v \in [0, v_c]$. Greenshield's linear model leads to:

$$v_{stable}(f) = \frac{v_0}{2}(1 + \sqrt{1 - \frac{4f}{v_0 k_{jam}}} = \frac{v_0}{2}(1 + \sqrt{1 - f/Q})$$
$$v_{unstable}(f) = \frac{v_0}{2}\sqrt{1 - f/Q})$$

In this particular case that one can assume the flow regime is always stable, concerning the relation

$v = v_{stable}(f)$ the corresponding relationship between travel time t and flow may be defined:

$t = t(f) = L/v_{stable}(f)$ 8

Queuing Models

The average delay experienced by vehicles that queue to cross a flow interruption point (intersections, toll barriers, merging sections, etc.) is affected by the number of vehicles waiting. This phenomenon may be analyzed with models derived from queuing theory, developed to simulate any waiting or user queue formation at a server (administrative counter, bank counter, etc.). The subject is treated below concerning generic users, at the same time highlighting the similarities with an uninterrupted flow.

Fundamental Variables

The main variables that describe queuing phenomena are:
τ The time at which the system is observed
τ_i The arrival time of user i $h_i = \tau_i - \tau_{i-1}$ The headway between successive users i and i −1 joining the queue at times τ_i and τ_{i-1}
$m_{IN}(\tau, \tau + \Delta\tau)$ Number of users joining the queue during $[\tau, \tau + \Delta\tau]$
$m_{OUT}(\tau, \tau + \Delta\tau)$ Number of users leaving the queue during $[\tau, \tau + \Delta\tau]$
$h(\tau, \tau + \Delta\tau) = \sum_{i=1,\ldots,m} h_i / m_{IN}(\tau, \tau + \Delta\tau)$ Mean headway between all vehicles joining the queue in the time interval $[\tau, \tau + \Delta\tau]$
$n(\tau)$ Number of users waiting to exit (queue length) at time τ Concerning observable quantities, flow variables may be introduced.

$u(\tau, \tau + \Delta\tau) = m_{IN}(\tau, \tau + \Delta\tau)/\Delta\tau$ arrival (entering) flow during $[\tau, \tau + \Delta\tau]$
$w(\tau, \tau + \Delta\tau) = m_{OUT}(\tau, \tau + \Delta\tau)/\Delta\tau$ exiting flow during $[\tau, \tau + \Delta\tau]$

Note that the main difference with the basic variables of running links is that space (s, Δs) is no longer explicitly referred to because it is irrelevant. Some of the above variables are shown in Fig. 5.

Concerning the service activity, let:

$t_{s,i}$ Be service time of user i
$t_s(\tau, \tau + \Delta\tau)$ Average service time among all users joining the queue in the time interval $[\tau, \tau + \Delta\tau]$
t_{wi} Total waiting time (pure waiting plus service time) of user i
$t_w(\tau, \tau + \Delta\tau)$ Average total waiting time among all users joining the queue in the time interval $[\tau, \tau + \Delta\tau]$

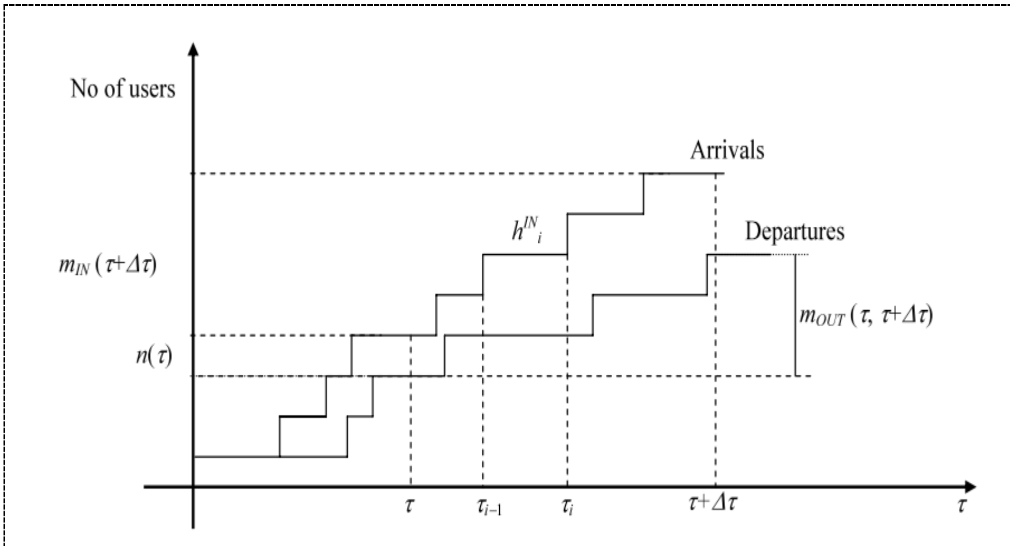

Fig. 5 Fundamental variables for queuing systems.

$Q(\tau, \tau + \Delta\tau) = 1/t_s(\tau, \tau + \Delta\tau)$ the (transversal) capacity or maximum exit flow, that is, the maximum number of users that may be served in the time unit, assumed constant during $[\tau, \tau + \Delta\tau]$ for simplicity's sake (otherwise Δτ can be redefined).

The capacity constraint on exiting flow is expressed by:

$$\omega \le Q$$

A general conservation equation, similar to (1) and (2) introduced for uninterrupted flow, holds in this case:

$n(\tau) + m_{IN}(\tau, \tau + \Delta\tau) = m_{OUT}(\tau, \tau + \Delta\tau) + n(\tau + \Delta\tau)$ 9

Moreover, dividing by $\Delta\tau$ we obtain:

$$\Delta n/\Delta\tau + w(\tau, \tau + \Delta\tau) - u(\tau, \tau + \Delta\tau) = 0 \quad 10$$

In the following subsection, we describe several deterministic models developed under the assumption that the headway between two consecutive vehicles and the service time is represented by deterministic variables. This is followed by a sub-section on stochastic models developed using random variables. In formulating such models, as in the case of uninterrupted flow models, we assume arrival at the queue is represented as a continuous (one-dimensional) fluid.

Deterministic Models

Deterministic models are based on the assumption that arrival and departure times are deterministic variables. According to the fluid approximation, the conservation equation (10) for $\Delta\tau \rightarrow 0$ becomes (see Fig. 6):

$$dn(\tau)/dt = u(\tau) - w(\tau)$$

Deterministic queuing systems can also be analyzed through the cumulative number of users that have arrived at the server by time τ, and the cumulative number of users that have departed from the server (leaving the queue) at time τ, as expressed by two functions termed arrival curve $A(\tau)$, and departure curve $D(\tau) \leq A(\tau)$, respectively; see Fig. 6. Queue length $n(\tau)$ at any time τ is given by:

$$n(\tau) = A(\tau) - D(\tau) \quad 11$$

Provided that the queue at time 0 is given by $n(0) = A(0) \geq 0$ with $D(0) = 0$. The arrival and departure functions are linked to entering and exiting users by the following relationships.

$$M_{IN}(\tau, \tau + \Delta\tau) = A(\tau + \Delta\tau) - A(\tau) \quad 12$$
$$m_{OUT}(\tau, \tau + \Delta\tau) = D(\tau + \Delta\tau) - D(\tau) \quad 13$$

The flow conservation equation (9) can also be obtained by subtracting member by member the relationships (12) and (13) and taking into account (11). The limit for $\Delta\tau \rightarrow 0$ of (12) and (13) leads to (see Fig. 7):

$$u(\tau) = dA(\tau)/d\tau$$

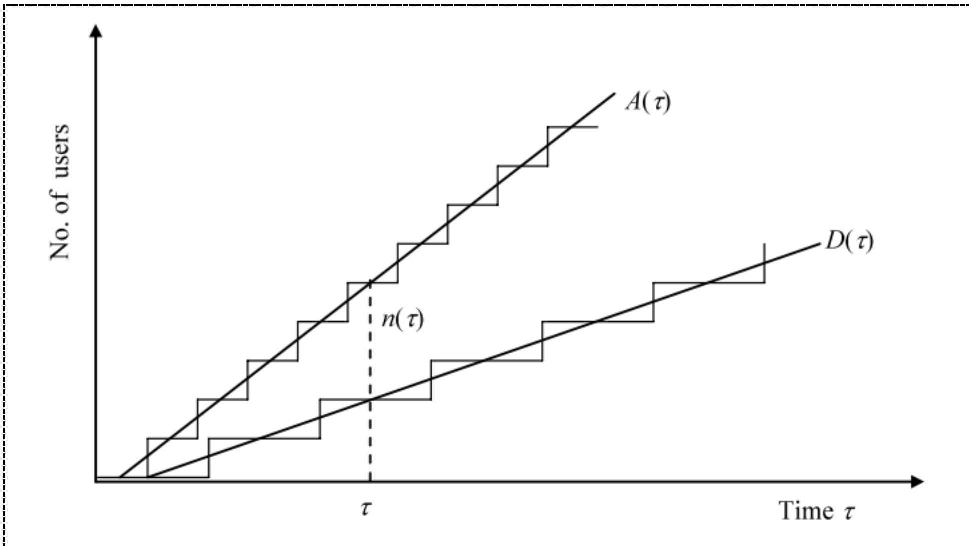

Fig. 6 Fluid approximation of deterministic queuing systems.

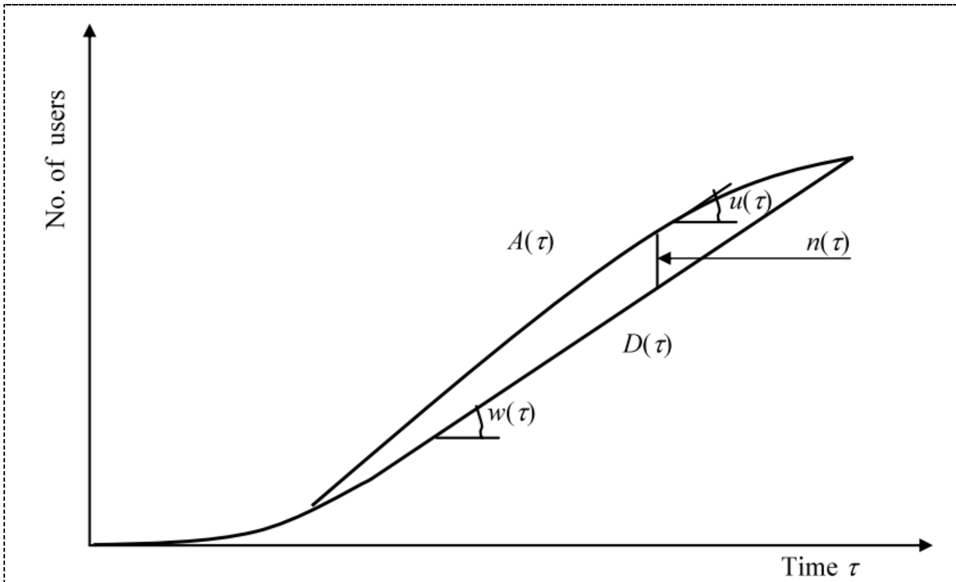

Fig. 7 Cumulative arrival and departure curves.

$w(\tau) = dD(\tau)/d\tau$

If during the time interval $[\tau_0, \tau_0 + \Delta\tau]$ the entering flow is constant over time, $u(\tau) = \bar{u}$, then the queuing system is named (flow-)stationary, and the arrival function $A(\tau)$ is linear with slope given by $\bar{u}$:

$A(\tau) = A(\tau_0) + \bar{u} \cdot (\tau - \tau_0) \quad \tau \in [\tau_0, \tau_0 + \Delta\tau]$

The exit flow may be equal to the entering flow $\bar{u}$,, or to the capacity Q as described below:

Undersaturation When the arrival flow is less than capacity ($\bar{u}$ < Q) the system is undersaturated. In this case, if there is a queue at time τ_0, its length decreases with time and vanishes after a time $\Delta\tau_0$ defined as (see Fig. 8)

$$\Delta\tau_0 = n(\tau_0)/(Q - \bar{u}) \qquad 14$$

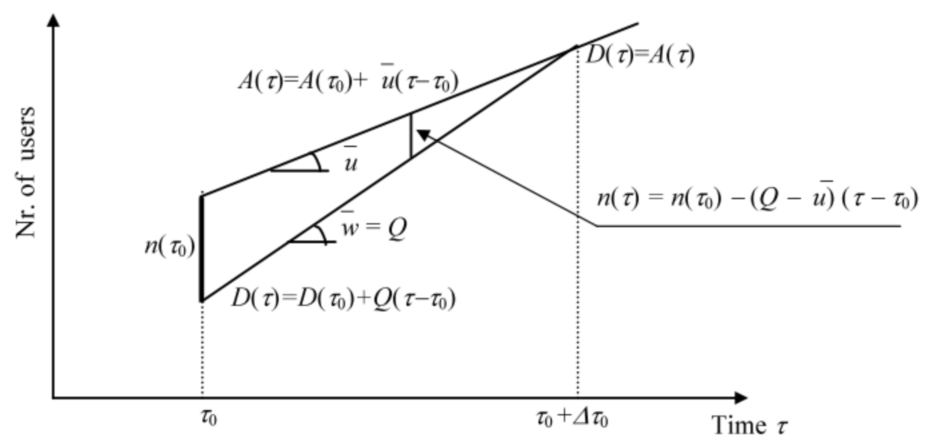

Fig. 8 Undersaturated queuing system.

Before time $\tau_0 + \Delta\tau_0$, the queue length is linearly decreasing with τ, and the exiting flow $\bar{\omega}$ is equal to capacity Q:

$$n(\tau) = n(\tau_0) - (Q - \bar{u})(\tau - \tau_0)$$

$$\bar{\omega} = Q \qquad 15$$

$$D(\tau) = D(\tau_0) + Q(\tau - \tau_0)$$

After time $\tau_0 + \Delta\tau_0$, the queue length is zero and the exiting flow $\bar{\omega}$ is equal to the arrival flow $\bar{u}$:

$$N(\tau_0 + \Delta\tau_0) = 0$$

$$\bar{\omega} = \bar{u} \qquad 16$$

$$D(\tau) = A(\tau) = A(\tau_0) + \bar{u}(\tau - \tau_0)$$

Oversaturation When the arrival flow rate is larger than capacity, $\bar{u} \geq Q$, the system is oversaturated. In this case queue length linearly increases with time τ and the exiting flow is equal to the capacity Q (see Fig. 9):

$n(\tau) = n(\tau_0) + (\bar{u} - Q)(\tau - \tau_0)$

$\bar{\omega} = Q$ 17

$D(\tau) = D(\tau_0) + Q(\tau - \tau_0)$

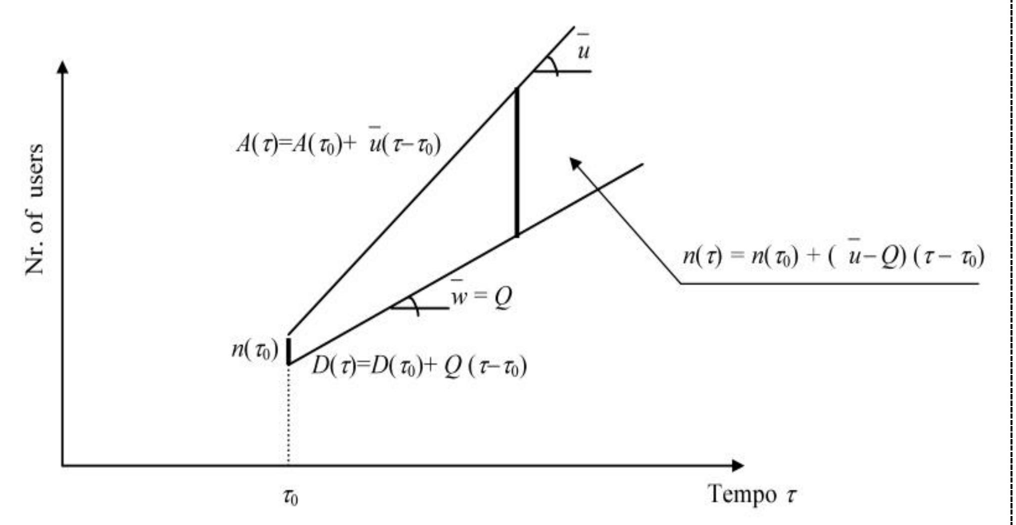

Fig. 9 Oversaturated queuing system.

General Condition By comparing (15) through (17) it is possible to formulate this general equation for calculating the queue length at generic time instant τ

$n(\tau) = \max[0, (n(\tau 0) + (\bar{u} - Q)(\tau - \tau_0))]$ 18

With the above results, any general case can be analyzed by modeling a sequence of periods during which arrival flow and capacity are constant. An important case is that of the queuing system at traffic lights which may be considered a sequence of undersaturated (green) and oversaturated (red) periods with zero capacity.
The delay can be defined as the time needed for a user to leave the system (passing the server), accounting for the time spent queuing (pure waiting). Thus the delay is the sum of two terms:

$tw = t_s + tw_q$

Where:

tw is the total delay.
$t_s = 1/Q$ is the average service time (time spent at the server).
tw_q is the queuing delay (time spent in the queue).

In undersaturated conditions ($\bar{u}$ < Q) if the queue length at the beginning of the period is zero (it remains equal to zero), the queuing delay is equal to zero, tw_q (u) = 0, and the total delay is equal to the average service time:

tw($\bar{u}$) = ts

In oversaturated conditions ($\bar{u} \geq$ Q), the queue length, and respective delay, would tend to infinity in the theoretical case of a stationary phenomenon lasting for an infinite time. In practice, however, oversaturated conditions last only for a finite period T. If the queue length is equal to zero at the beginning of the period, it will reach a value ($\bar{u}$ − Q) · T at the end of the period. Thus, the average queue over the whole period T is:

$\bar{n}$ = ($\bar{u}$ − Q) T/2

In this case, the average queuing delay is $\bar{x}$/Q, and the average total delay is (see Fig. 10):

tw ($\bar{u}$) = ts + ($\bar{u}$ − Q) T /2Q 19

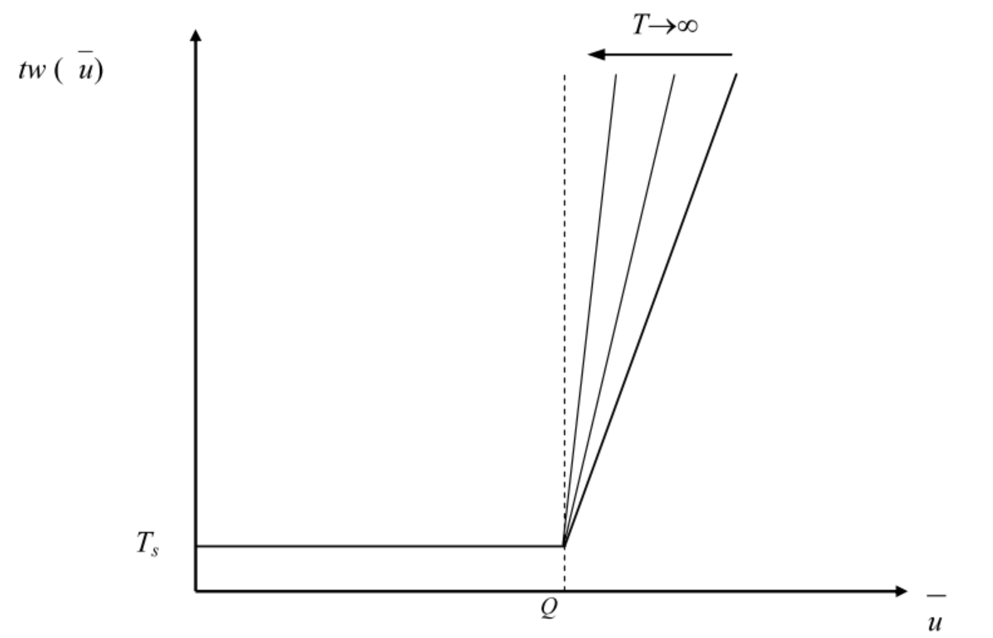

Fig. 10 Deterministic delay function at a server.

Stochastic Models

Stochastic models arise when the variables of the problem (e.g., user arrivals, service times of the server, etc.) cannot be assumed deterministic, due to the observed fluctuations, as is often the case, especially in transportation systems. If the system is undersaturated, it can be analyzed through (stochastic) queuing theory which includes the particular case of the deterministic models illustrated above. Some of the results of this theory are briefly reported below, without any claim to being exhaustive.

It is particularly necessary to specify the stochastic process describing the sequence of user arrivals (arrival pattern), the stochastic process describing the sequence of service times (service pattern), and the queue discipline. Arrival and service processes are usually assumed to be stationary renewal processes, in other words with stable characteristics in time that are independent of the past: that is, headways between successive arrivals and successive service times are independently distributed random variables with time-constant parameters. Let N be a random variable describing the queue length, and n is the realization of N. The characteristics of a queuing phenomenon can be redefined in the following concise notation:

a / b /c (d, e)

Where:

a denotes the type of arrival pattern, that is, the variable that describes time intervals between two successive arrivals:

D = Deterministic variable.
M = Negative exponential random variable.
E = Erlang random variable.
G = General distribution random variable.

b denotes the type of service pattern, such as a
c is the number of service channels: {1, 2,...}
d is the queue storage limit: $\{\infty, n_{max}\}$ or longitudinal capacity
e denotes the queuing discipline:

FIFO = First In–First Out (i.e., service in order of arrival).
LIFO = Last In–First Out (i.e., the last user is the first served).
SIRO = Service In Random Order.
HIFO = High In–First Out (i.e., the user with the maximum value of an indicator is the first served).

Fields d and e, if defined respectively by ∞ (no constraint on maximum queue length) and by FIFO, are generally omitted. In the following, we report the main results for the M/M/1 (∞, FIFO) and the M/G/1 (∞, FIFO) queuing systems, which are commonly used for simulating transportation facilities, such as signalized intersections.

Some definitions or notations differ from those traditionally adopted in dealing with queuing theory (the relative symbols are in brackets) so as to be consistent with those adopted above. The parameters defining the phenomenon are as follows.

u, (λ) The arrival rate or the expected value of the arrival flow.
$Q = 1/t_s$, (μ) the service rate (or capacity) of the system, the inverse of the expected service time.
u/Q, (ρ) The traffic intensity ratio or utilization factor.

n A value of the random variable N, the number of users present in the system, consisting of the number of users queuing plus the user present at the server, if any (the significance of the symbol n is thus slightly different).
t_w A value of the random variable TW, the time spent in the system or overall delay, consisting of queuing time plus service time.

(a) M/M/1 (∞, FIFO) Systems In undersaturated conditions (u/Q < 1):

$$E(N) = \frac{\frac{u}{Q}}{1-\frac{u}{Q}} = \frac{u}{Q-u} \qquad 20$$

$$VAR(N) = \frac{\frac{u}{Q}}{(1-\frac{u}{Q})^2}$$

According to Little's formula, the expected number of users in the system E [N] is the product of the average time in the system (expected value of delay) E [TW] multiplied by arrival rate u:

$$E[N] = u\,E[TW] \qquad 21$$

From which:

$$E(TW) = \frac{1}{Q-u} \qquad 22$$

The expected time spent in the queue $E[tw_q]$ (or queuing delay) is given by the difference between the expected delay E[tw] and the average service time $t_s = 1/Q$:

$$E[TW_q] = \frac{1}{Q-u} - \frac{1}{Q} = \frac{u}{Q(Q-u)} \qquad 23$$

According to Little's second formula, the expected value of the number of users in the queue E [Nq] is the product of the expected queuing delay E [TWq] multiplied by the arrival rate u:

$$E[N_q] = uE[TW_q] \qquad 24$$

And then:

$$E[N_q] = \frac{u^2}{Q(Q-u)} \qquad 25$$

(b) M/G/1 (∞, FIFO) Systems In this case the main results are the following.

$$E[N] = \frac{u}{Q}[1 + \frac{u}{2(Q-u)}]$$

$$E[TW] = \frac{1}{Q}[1 + \frac{u}{2(Q-u)}]$$

$$E[TW_q] = \frac{u}{2Q(Q-u)}$$