This paper is published in Networks and Spatial Economics, 3: (2003) 297-322


# The CONTRAM dynamic traffic assignment model

NICHOLAS B TAYLOR
*Transportation Division, TRL Limited*[1]
*Email: ntaylor@trl.co.uk*

*Abstract*

CONTRAM is a computer model of time-varying traffic in road networks, which takes as input the network definition and time-varying demand for travel between a set of origin and destination zones, and outputs the resulting network flows, routes and travel times. It combines a macroscopic time-sliced traffic model with disaggregate dynamic assignment of traffic, so is intermediate between macroscopic equilibrium and microscopic models. The paper details the methods used, including time-dependent queuing which plays a central role, and the treatment of network definition, user classes, road capacities, signals and coordination, vehicle emissions, Intelligent Transport Systems and research lines.

**Keywords:** assignment, congestion, ITS, networks, road, simulation, traffic

## 1. Introduction

CONTRAM (ref.) is a computer model of time-varying traffic in road networks, which takes as input the network definition and time-varying demand for travel between a set of origin and destination zones, and outputs the resulting network flows, routes and travel times. The program runs within a Windows graphical environment. The model falls into the category of dynamic traffic assignment (DTA) models. Its distinctive approach is to combine a form of microscopic simulation of traffic quanta, called 'packets' by analogy with communications networks, with a macroscopic time-dependent traffic model.

Application extends not only to a wide range of network scales, from city centres to the whole of the Netherlands (for example), but also to a range of problem types, including road scheme appraisal, simulation of complex temporal demand patterns at sporting venues and airports, and even modelling of pedestrian movements in railway concourses (full bibliography can be obtained from the author). By interpreting packets as analogous to individual vehicles, with behaviour beyond the optimum route seeking which leads to equilibrium solutions, it is possible to simulate response to Intelligent Transport Systems (ITS). CONTRAM is 'tactical' in the sense that it deals with the assignment of a given demand, not with how that demand is created. The process of travel forecasting by the 'four/five-step' procedure of generation, distribution, departure time choice, modal-split and assignment, requires methods which are beyond the scope of this paper.

This paper is organised as follows. Section 2 contains a summary of the model's principles. Section 3 touches on issues of dynamic assignment. Section 4 discusses data. Sections 5-8 focus on key methods including time-dependent queuing, and Section 9 addresses ITS. Finally Section 10 considers current research and future developments.

## 2. Principles of the CONTRAM model

Real traveller or vehicle histories are continuous in time. Actual journeys are characterised by events such as entering network links, stopping at junctions, departing from queues, receiving or acting on information, and arriving at a destination. Truly microscopic models simulate traffic by following each vehicle from moment to moment, or executing events in time sequence, but are computationally costly. In macroscopic modelling, the network state is described in terms of aggregate quantities such as the average traffic arrival rate on a link, but this can be a poor representation of dynamic conditions. Vehicle packet histories in CONTRAM are continuous in time, but interact only through an underlying time-sliced macroscopic traffic model, removing the need to model all events in strict sequence. This is sometimes called a 'mesoscopic' approach.

Time slices are a compromise between the relative simplicity of modelling with time-averaged values and the need to represent time variation. They also reflect most data collection practice which yields period counts. Time slices are chosen by the user, depending on the application, and range typically from 1 hour for slowly changing off-peak conditions, down to 5 minutes for highly dynamic on-line applications such as ITS modelling. A model can cover 24 hours (or more) and contain multiple peaks, as occur for example in airport traffic. The practical lower limit to resolution is set by the fact that CONTRAM cannot model explicit signal cycles and random variations in traffic.

Demand is input as a time-sliced matrix classified origin-destination (O-D) flows, which can be viewed as an approximation to an average daily profile, and could be derived from actual period counts or from a time-sliced demand model (see Figure 1).

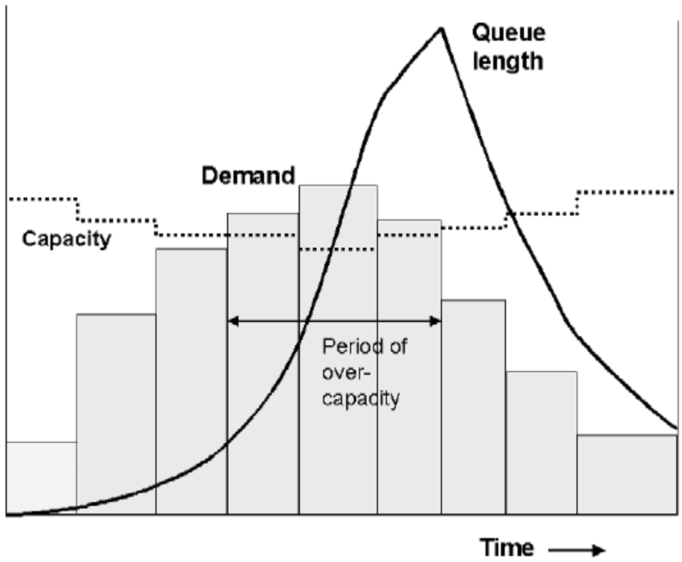


*Figure 1*. Typical demand, capacity and queue profiles and their relationship

CONTRAM has no elastic demand, but has been incorporated into a Variable Demand Matrix framework (Gordon *et al* 2003, Emmerson *et al* 2003), and there is associated Origin-Destination Matrix Estimation software (Hazelton and Gordon 2002).

A common feature of demand peaks is transient over-capacity, which cannot be handled adequately by a static model. Since a queue will continue to grow while demand exceeds capacity, the peak of queue length and delay tends to lag behind the peak of demand, and the greater the overloading of the network, the greater is this time lag (see Figure 1). This produces both a peak-spreading effect and a knock-on effect on later traffic. It is worth noting that queue length and delay are continuous functions, and change seamlessly through periods of over-capacity, an essential requirement for a dynamic model.

The demand is divided up into a stream of small ‘packets’, which are routed independently and assigned sequentially in journey start time order (an investigation of alternative packet assignment in parallel may be found in Greenwood and Taylor 1993). Iteration is needed because each packet can influence others starting their journeys earlier as well as later, and knowledge of the network state develops as if through day to day experience. Figure 2 illustrates this schematically. The ‘state’ of the network is to be thought of as being defined in space *and* time, rather than at a particular moment in time. Each packet’s path is calculated as a whole, with delays according to actual link arrival times, leaving a trace in the space-time description. Trip B is assigned computationally *after* trip A, but arrives in part of the congested centre of the network *before* A in traffic terms. The extra flow and queuing B contributes in Iteration $n$ causes additional delay to A in Iteration $n$+1.

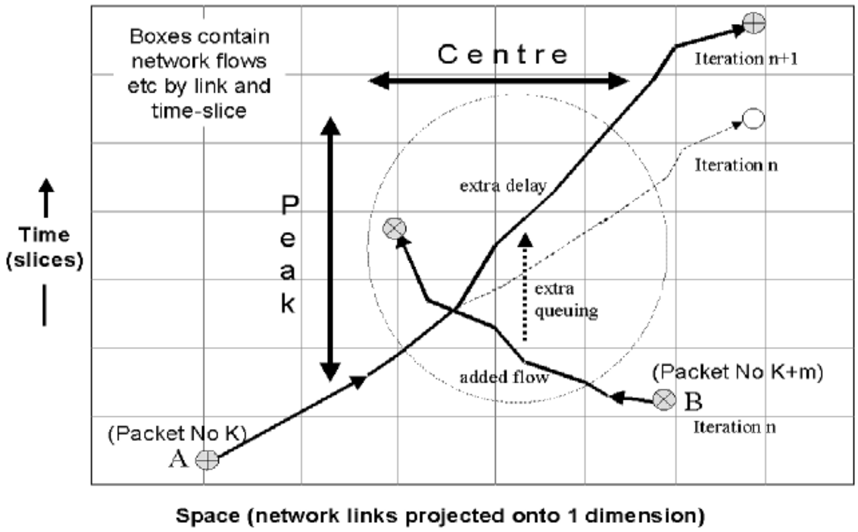


*Figure 2.* Schematic representation of interaction of packets over the traffic model

Each packet follows its minimum cost route in each iteration, given the 'current' network state. If the assignment converges, no packet can switch unilaterally to a route of lower time or cost, so in theory a dynamic Wardrop user equilibrium (DUE) applies, but there is no overall objective function to be minimised, so to that extent the model is 'heuristic'.

Network flows are accumulated, and capacities, queues and delays evaluated, mathematically within the aggregate time-dependent traffic model. Random variations in arrivals and departures can have a profound effect on queues and are accounted for within the queuing model itself (see Section 5). However, CONTRAM remains entirely deterministic, delivering a repeatable 'average' result for a given data set. In principle, any capacity/flow, speed/flow/delay, queuing model or cost function, expressible in closed-form or as an algorithm, could be employed. The methods used in practice are based on theoretical and empirical models many of which were developed at TRL.

## 3. Assignment and equilibrium

Minimum cost routes are found using a time-dependent link-based all-or-nothing tree-building method (Dijkstra 1959, Whiting and Hillier 1960). Because network conditions can change significantly even for small shifts in start time, a new route tree has to be built for each packet, and since only the minimum cost route to the actual destination is required, there is no advantage in using more efficient algorithms. However, computation time can be saved by guiding the algorithm to the destination using 'heuristic' reverse trees containing the *minimum possible* costs from all points to each destination (Taylor 1989). These are pre-built efficiently using D'Esopo's algorithm (Van Vliet 1977).

Each packet seeking its minimum cost route effectively has 'perfect information'. This may leave something to be desired from the point of view of realism, but it has a long history in the assessment of network or scheme performance, especially in congested networks, and provides a repeatable reference for developing models incorporating more realistic behavioural assumptions and ITS (see later in Sections 9 and 10). Cost is calculated link by link, based on the link length and travel time and other information (eg queuing time or toll) embodied in the generalised cost function defined for the particular user class to which the packet belongs (see later in section 4.2).

Packet re-assignment consists of first deducting the packet's flow from the network, and updating the network state, then finding its new route, loading it onto this, and updating the network again. When calculating the expected travel time along any link, the packet flow is temporarily added to the link flow to compensate for its finite size. This method of loading normally results in small changes provided packet sizes are not too big.

Running through the entire packet sequence constitutes one iteration of the model. Iterations are repeated a user-specified number of times, or until the model converges, or until a number of stability criteria prescribed in the data are satisfied. The latter include:

RMS change in link flows: $$\sqrt{\frac{1}{AS}\sum_{a=1}^{A}\sum_{s=1}^{S}\left(q_{as,n}-q_{as,n-1}\right)^2}$$

Average Absolute Difference (AAD): $$\frac{1}{AS}\sum_{a=1}^{A}\sum_{s=1}^{S}\left|q_{as,n}-q_{as,n-1}\right|$$

Percentage Relative AAD: $$\frac{100}{AS}\sum_{a=1}^{A}\sum_{s=1}^{S}\frac{\left|q_{as,n}-q_{as,n-1}\right|}{q_{as,n-1}}$$

where $q_{as}$ is the flow (vehicles) arriving on link $a$ ($\in$1..A) in time slice $s$ ($\in$1..S), and $n$ is iteration number. The percentage of links whose flows change by < 5% is also available as a termination criterion (<10% and <15% also monitored). A stability measure based on total perceived costs $C$ in the network is also evaluated:

Stability: $$100\frac{\left|C_n-C_{n-1}\right|}{C_{n-1}}$$

The most widely accepted measure of convergence is gap, which is a measure of the difference between actual and minimum route costs at any point in the assignment. However, extra iterations, in which no change of routes or network costs occurs, would be required to evaluate it, so it is not available as a convergence criterion. It is intended to report elsewhere research being conducted into splitting packets with variable flow proportions, where there is also the option to perform the extra iterations described.

## 4. Traffic description

### *4.1. Networks*

CONTRAM networks consist of conceptual nodes connected by unidirectional links. Special nodes and links act as origin/destination zones and connectors. Nodes are geometric points, while complex junctions such as roundabouts, or with dedicated turning lanes, are defined as sub-networks (see Figure 3).

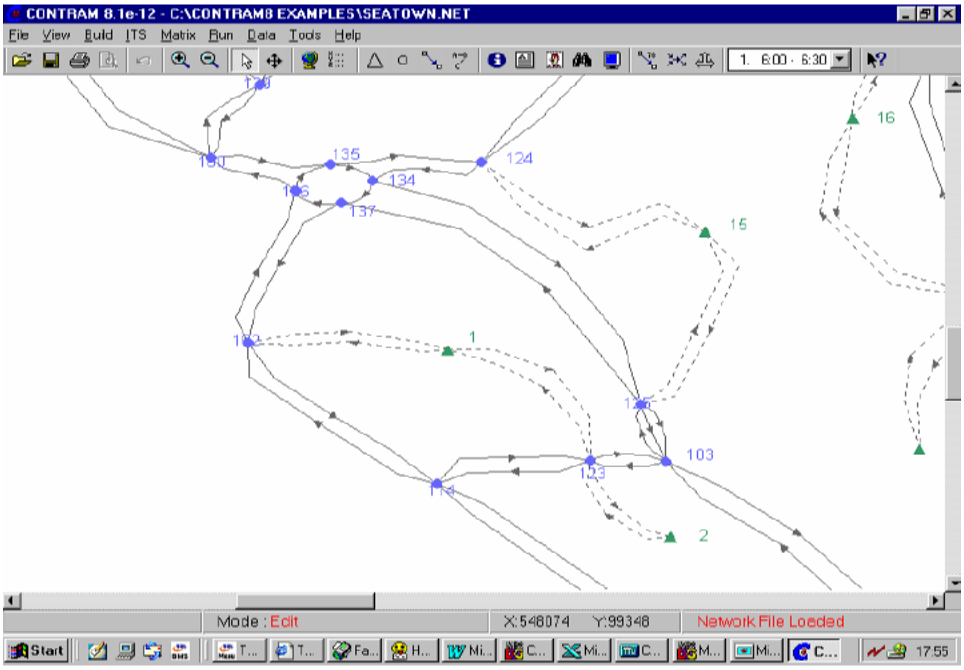


*Figure 3.* Sample of a network diagram

It is possible to have more than one link between a pair of nodes, so separate lanes and bus lanes can be represented, although weaving between lanes is not modelled.

Properties of links include:

- road length (specified independently of the coordinate separation of the nodes)
- number of lanes (lanes with independent properties can be defined as separate links)
- saturation flow or maximum capacity
- cruise speed or cruise time (zero time is allowed, eg for zone connectors)
- speed/flow function applying to the running length (if any)
- vehicle storage capacity for queuing[2]
- access and turn restrictions (if any)
- priority and capacity relationships involving other links (if applicable)
- signal timing information (if applicable)
- road category (jurisdiction code)

Internally, there is no distinction between 'real' links and zone connectors, so it is essential that notional zone connectors are coded to avoid having an unrealistic effect on the assignment. They are normally given short or zero lengths and travel times, and high capacities, so as not to affect route choice or journey costs. There is no 'road type' property as such. However, the 'category' (jurisdiction) code can be linked to a set of defaults, which apply to all roads in that category which do not have corresponding values defined explicitly. Category attributes include number of lanes, saturation flow, speed and speed/flow function, access restrictions and storage capacity.

### *4.2. User classes*

To represent different vehicle types and journey purposes, traffic in CONTRAM may be divided into up to 32 user classes. Each class has properties which include its PCU factor - the number of 'passenger car units' of capacity consumed by one unit of the class - the relative cruise speed factor of the class, and its generalised cost function given by:

$$C = aD + bT + cV^2 D + pT_q \qquad (4.1)$$

where $D$=distance travelled, $T$=total time spent, $V$=average speed and $T_q$ is time spent queuing, which is often valued more highly than time spent in motion. CONTRAM integrates all user classes fully. Pre-loading of public transport is not required, although a volume pre-load facility *is* available. Buses may be modelled as a separate class, assigned to fixed routes which may include links not available to other classes. CONTRAM's time resolution does not permit direct simulation of bus stops and bus priority. The delay buses cause can affect the routes chosen by other traffic, while bus journey times can be affected by traffic congestion. Other special classes, such as pedestrians, can also be modelled by appropriate use of speed and capacity.

Default class properties and generalised cost coefficients are currently provided for three vehicle classes representing light vehicles, buses and goods, but these have not been amended for some time, and it is advisable to refer to DMRB(ref.) for UK specific practice, or other appropriate source, for up to date values of time, operating cost etc.

Vehicles can be subject to tolls on certain links. Since these can vary by place, class and time they are defined in a special table. An additional global class attribute allows the cost of a toll perceived by a given user class to differ by a factor from its monetary cost to reflect, for example, indifference to tax. The monetary units are arbitrary, but the same cost units must be used for both tolls and generalised costs, and total network costs will be meaningful only if the same cost units apply to all classes.

*4.3. Traffic demand*

The demand data contain the time-dependent flow rates for each origin-destination-class and optionally fixed route combination, producing time sliced profiles similar to that in Figure 1. If a route is not specified then the demand is assigned to one or more routes according to minimum cost as defined by the generalised cost function for that class.

To create the packet stream, CONTRAM generates for each classified O-D movement a stream of more or less equal sized packets (see Figure 4) according to its demand profile. These are then interleaved and assigned in the order of their journey start times.

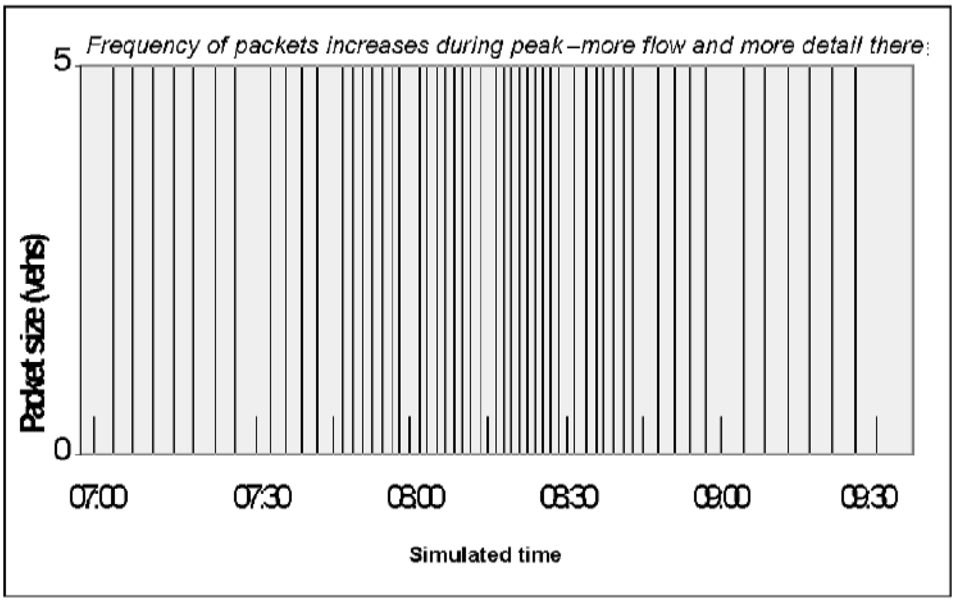


*Figure 4*. Representation of a packet stream

The packet size can be specified or calculated automatically. For each classified O-D movement the packets are all about the same size, so changes in flow rate between different time slices are reflected in changes in departure frequency. The formula for calculating the target packet size of an O-D movement is intended to balance size and number of packets, where $\{Q_i\}$ are the total volumes loaded on that O-D movement in each time slice, $n$ is the number of time slices with data, and $k$ is a user-supplied scaling factor (default value is √2 giving slightly larger packet size and shorter run time than value 1):

$$P = k \frac{\sum_{i=1}^{n} Q_i}{\sum_{i=1}^{n} \sqrt{Q_i}} \tag{4.2}$$

*4.4. Speed/flow functions*

Some assignment models use delay/flow functions, which can take various forms and represent the whole travel time along a link including the effects of queuing. An example is the power-function form used by SATURN (ref.), where delay is proportional to a user-definable power of the volume/capacity ratio, similar to the US Bureau of Public Roads function (FHWA 2000). This function, being continuous and integrable, lends itself both to calibration against exogenous delay/flow data and to incorporation in objective functions of the Beckman type.

However, the speed/flow functions used in CONTRAM play a more limited rôle, representing friction in uncongested conditions on motorways and other high-speed links, or as a proxy for the effect of junctions in simplified networks or ‘buffer networks’ which conduct traffic to and from an area to be simulated in detail. Currently, continuous, piecewise linear COBA (COst Benefit Analysis) functions are supported (COBA 2003) (see Figure 5). These are technically static functions which modify the cruise speed along a link in each time slice according to the average flow entering the link in that time slice. The ‘capacity’ point, $Q_C$, is used to define the last segment of the empirical relationship, but is *not* taken to be the capacity of the link for queuing purposes. Up to seven linear segments are supported, as well as user-class-dependence, and in the latest development version the choice of speed/flow function can be time-dependent to facilitate modelling tidal-flow systems, ramp metering, etc.

COBA speed/flow relationships are unsuitable for modelling congestion because they do not cover queuing conditions. Queuing on motorways is currently modelled in CONTRAM using an explicit bottleneck represented by the ‘stop line’ of a link. This gives the correct total delay which is essentially a function of the excess of demand over capacity, but is not entirely realistic where the capacity loss is due to flow breakdown, which often happens some way downstream of a merge. It also does not yield the correct density and physical extent of a queue, unless vehicle storage capacity is set appropriately. This is tied in with the distinction between 'vertical' and 'horizontal' queuing which will be referred to later in Section 5, but there is more to it because the dynamics of vehicles in large queues, especially where shock waves are present, extends beyond queuing theory into car-following behaviour.

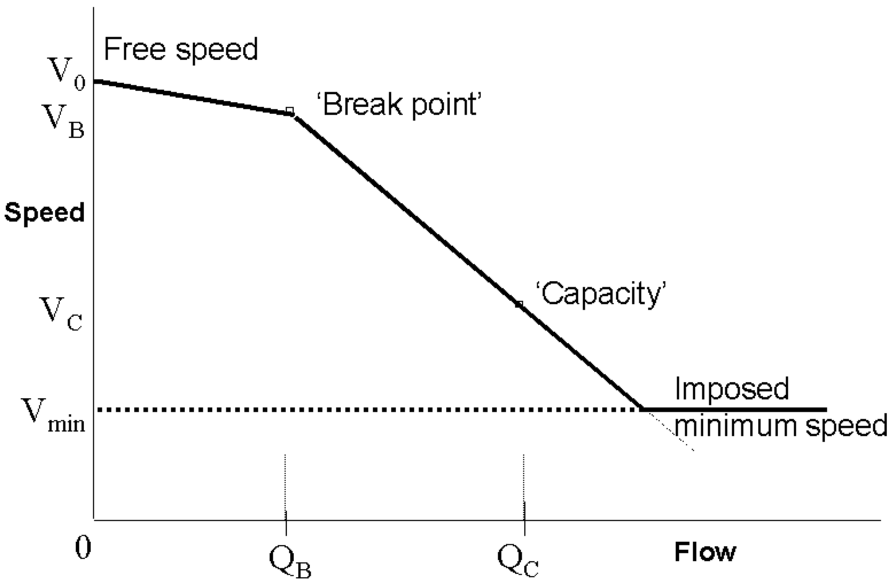


*Figure 5.* COBA uncongested speed/flow relationship

### *4.5. Model outputs*

CONTRAM aims to give a complete picture of the traffic in a network. Most results are derived from the macroscopic traffic model and relate to links, including the flows, average PCU factors, capacities, travel times, traffic intensities, average speeds and queues in each time slice. Flows are available disaggregated by both time slice and user class, and turning flows are available. Although individual packets are assigned 'all-or-nothing', when combined over a time slice several routes may be used. A skim of time-slice-dependent classified origin-destination movements provides flow rates, journey distances, times, costs and tolls incurred for each of the routes used. The paths and timings of individual packets can easily be extracted and re-cycled as baseline data for, say, studies of response to incidents and ITS. Fuel consumption and line-emissions (emissions per unit length of road) are predicted, based on a model with user-defined coefficients, enabling local pollution 'hot-spots' to be detected, or total pollution to be estimated using an external model of dispersion and extraneous sources. Summary results at the end of each iteration give information about convergence.

## 5. Time-dependent queues and delays

### *5.1. Overview*

At the heart of CONTRAM is time-dependent queuing theory. This provides closed-form solutions for cases where demand and capacity remain constant over a period of time, apart from random variations in arrivals and departures, as will be described in section 5.2. These solutions can be applied directly to demand and supply (capacity) which are held as average (constant) values within each of a sequence of time slices. The final queue in each time slice becomes the initial queue for the next. The resulting queues and delays are continuous (though not everywhere differentiable) functions of time.

FIFO (First In First Out) applies to a high degree to urban traffic where overtaking opportunities are limited, and also has a deeper meaning in relation to continuous dynamic assignment (eg Heydecker and Verlander 1999). In CONTRAM, FIFO is not obeyed absolutely due to the averaging of flows over each time slice. However, it is believed that the effect on overall results is small and can be minimised by using more, shorter time slices. The queues met by packets along their routes are calculated for their exact time of arrival (see earlier in Section 2), and the delays are calculated by tracing the discharge of the queue through future time slices, taking into account any future changes in capacity.

The general mathematical approach is that of Kimber and Hollis (1979) which draws on classical sources of queuing theory. Because of the simulation character of CONTRAM, its queue calculations require some extensions not found in, for example, TRL's junction models ARCADY, PICADY and OSCADY (Semmens 1985a,b, Burrow 1987), which also embody the time-dependent queuing model. Any queuing model requires a model of capacities, normally defined at stop lines in an urban model. The capacity relationships for priority and signal junctions developed by TRL exist in their most detailed form in the junction models, where details of geometry and the effects of movements sharing lanes or obscuring visibility are included. CONTRAM necessarily uses simplified versions.

### *5.2. Time-dependent queuing theory*

The time-dependent queue model is based on the 'divided' queue model of Kimber and Hollis (1979), so called because it divides queue growth and decay into a number of distinct regimes. For certain of these regimes the model is derived from the 'sheared' queue model developed by R Whiting (unpublished). This combines random queue processes, which are asymptotic to a finite steady-state equilibrium value, with over-saturated processes, which can grow indefinitely, as illustrated in Figure 6.

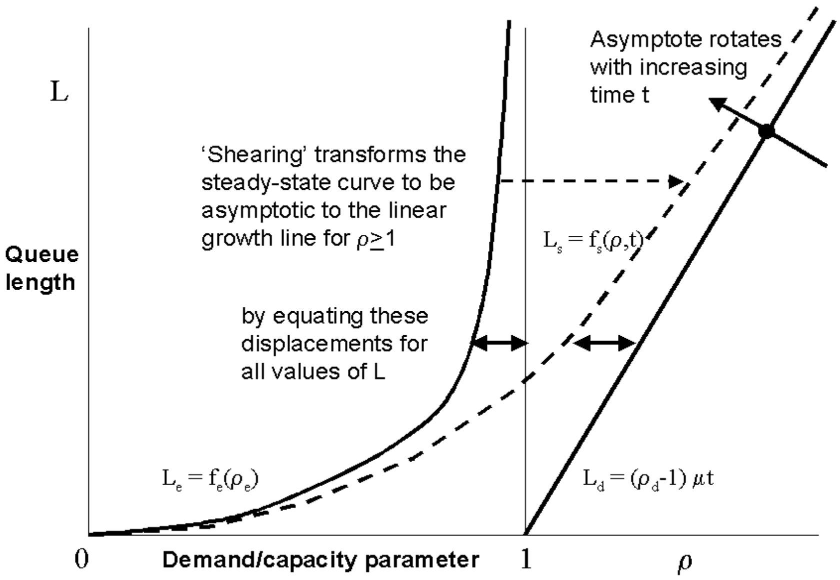


*Figure 6.* Shearing of queue functions

Equilibrium queues arise when average demand is less than average capacity because random fluctuations in demand and capacity can bring about queuing when demand temporarily exceeds capacity, but once any queue is discharged, this is not compensated by a temporary excess of capacity over demand. Deterministic queues arise due to the excess of average demand over average capacity, and simply represent those vehicles which cannot be discharged, nominally represented by (5.1) where $L_0$ is the queue at $t$=0, and $\rho$ is the traffic intensity, or ratio of demand to capacity $\mu$:

$$L_d(t) = L_0 + (\rho - 1)\mu t \tag{5.1}$$

Some types of queue have exact time-dependent mathematical descriptions, but these are very complicated (see eg Kimber and Hollis 1979). In the steady state, exact formulae at equilibrium can be obtained for certain types of queue in terms of the parameter $\rho$, limited to the range [0,1). The equilibrium queue $L_e$ (Figure 6, left) is of this type, and $L_e \to \infty$ as $\rho \to 1$. Its simplest form is that of the M/M/1 queue typical of give-way junctions, whose arrival and service time headways are exponentially distributed:

$$L_e = \frac{\rho}{1-\rho} \tag{5.2}$$

When $\rho$>1, the deterministic queue $L_d$ (right) grows indefinitely since capacity is unable to service demand. Shearing transforms the equilibrium queue to be asymptotic to the deterministic queue, resulting in a time-dependent model (broken graph) which covers the full range of $\rho$ and $t$, although it is no longer exact. The transformation equates the two displacements in Figure 6, which can be written in two ways:

$$\rho_d - \rho = 1 - \rho_e \qquad \text{or} \qquad \rho_d - 1 = \rho - \rho_e \tag{5.3}$$

The right hand form suggests a substitution in the bracketed term in equation (5.1). The quantity $\rho_e$ is by definition the inverse of the steady-state queue $L_e$ which by construction is equal to the deterministic queue $L_d$. Using the notation $x$ instead of $\rho_e$, and adding an initial queue $L_0$, results in a deterministic time-dependent formula which embodies the equilibrium queue model, given that $x$ is now expressible as some function $L_e^{-1}(L)$:

$$L = L_d(x,t) = L_0 + (\rho - x)\mu t \tag{5.4}$$

Here $\mu$ is the capacity, the product $\rho\mu$ is the rate of demand, and all parameters are assumed constant. The product $x\mu$ is the rate of throughput, and $x$ can be interpreted as the time-averaged mean[3] degree of saturation of the stop line, or utilisation of capacity.

A generalisation of the steady-state equilibrium queue formula, again using the variable $x$ rather than $\rho$, is used to embody the properties of the arrival and service processes:

$$L = L_e(x) = Ix + \frac{Cx^2}{1-x} \tag{5.5}$$

The coefficient $I$, known as the vehicle-in-service factor, is assumed to be 1 if vehicles have to slow down or stop, as at a priority junction, or 0 if they can move freely across the stop line, as during the green phase at a signal. The coefficient $C$ approximating the randomness of the system is expressed by Kimber and Daly (1986) as:

$$C = .5\left[\left(\frac{\sigma_a}{\tau_a}\right)^2 + \left(\frac{\sigma_s}{\tau_s}\right)^2\right] \tag{5.6}$$

where $\tau_i$, $\sigma_i$ are the mean and standard deviation of the headway distributions of the arrival and service ($i=a,s$) respectively. $C$ is 1.0 for a typical give-way with exponentially distributed arrival and service headways (ie M/M/1), and empirically 0.6 for a signal where the arrivals are random and the service quasi-uniform (approximately M/D/1)[4].

The solution of (5.4,5.5) is a quadratic, in general terms say:

$$L = L_s\left(t, L_0, \rho, \mu, I, C\right) \tag{5.7}$$

Although quite accurate for queues growing from zero, Kimber and Hollis (1979) find it to be less so for queues growing or decaying from a non-zero starting value[5,6]. Thus they divide the model into three regimes. In the first they eliminate the initial value $L_0$ in the growing queue by displacing time $t$ by the length of time $t_0$ which a queue starting from zero with the same constant parameters would require to reach $L_0$. The modified problem thus becomes (5.8) below, whose explicit solution is given by (5.9,5.10):

$$L = L_s\left(t + t_0, 0, \rho, \mu, I, C\right) \tag{5.8}$$

$$L = \frac{-b + \sqrt{b^2 - 4ac}}{2a} \qquad \text{(if } a \neq 0\text{)}$$

$$L = -\frac{c}{b} \qquad \text{(if } a = 0\text{)} \tag{5.9}$$

where:

$$\begin{aligned} a &= \mu(t + t_0) - (C - I) \\ b &= (1 - \rho)(\mu(t + t_0))^2 + (2\rho(C - I) + I)\mu(t + t_0) \\ c &= -\rho(\rho(C - I) + I)(\mu(t + t_0))^2 \\ t_0 &= L_s^{-1}\left(0, L_o, \rho, \mu, I, C\right) \end{aligned} \tag{5.10}$$

The inverse function $L_s^{-1}$ for the time displacement $t_0$ also gives rise to a quadratic. The 'back-projected' time $t_0$ is calculated from the change in queue length in the interval, ie $L_0$, divided by the average rate of change of queue length during that time, in capacity-normalised form $L_0' = dL/d(\mu t)$ evaluated at $t_0$, which is expressible in terms of $L_0$:

$$t_0 = \frac{L_0}{L'_0 \mu} \tag{5.11}$$

where:

$$L'_0 = \rho - \frac{L_0}{L_0 + I} \qquad \text{(if } C = I\text{)}$$

$$L'_0 = \rho - \frac{\left(-(L_0 + I) + \sqrt{(L_0 + I)^2 + 4L_0(C - I)}\right)}{2(C - I)} \qquad \text{(if } C \neq I\text{)} \tag{5.12}$$

The two remaining regimes apply to decaying queues. For a queue falling from an initial value less than $2L_e(\rho)$, the model is the mirror image of (5.8-12) with appropriate redefinition of the effective start time $t_0^*$, which becomes the time at which the queue length would have been equal to $2L_e$ (ie with $2L_e$-$L_0$ replacing $L_0$ in (5.11,5.12)) hence:

$$L = 2L_e - L_s\left(t + t_o^*, 0, \rho, \mu, I, C\right) \tag{5.13}$$

For a queue initially greater than $2L_e$, where demand must be less than capacity ($\rho$<1), as long as the queue remains $\geq 2L_e$ it is assumed to decay at a constant rate given by (5.12), after which the 'inverted' model takes over.

The divided queue model estimates *mean* queue length, which can be interpreted as the mean of a probability distribution, or as the mean value which would be observed at the same time each day if a traffic pattern with the same parameters were repeated day after day with random variation in arrivals and service. It has been validated by Kimber and Daly (1986), who also find that the envelope of observed values of an urban queue growing to equilibrium lies roughly between zero and twice the mean. The model describes a *vertical* queuing process taking place *at* the stop line, which does not take into account the physical space occupied by the queue. It is believed that this tends to underestimate queue length, since a growing queue intercepts arriving traffic so increasing the effective arrival rate. It is hoped that resources might be found to develop a practical horizontal queue model, and its natural extension to large queues on motorways where the dynamics of traffic within the queue also become important.

### *5.3. Phase queues*

The phase queue is the sawtooth component of the queue produced by a red/green signal. The phasing of the signal determines its capacity and this and the demand determine whether the queue stabilises or continues to grow. It is not possible to 'shear' the phase queue because it works on a different time scale from the random queue (5.4). Amber and red/amber periods are 'absorbed' into the effective green and red periods modelled, plus individual phase start and end delays and global 'lost time' displacements.

The model of phase queues is essentially that of Burrow (1987) (as implemented in OSCADY) with the addition of signal coordination modifications described later in section 6.4. Any queue remaining at the end of the green period is governed by the signal equilibrium queue described in the previous section. This will occur if $\rho \geq 1$ or for as long as an initial queue is able to maintain saturation. The maximum phase queue therefore occurs when the queue just discharges at the end of the green period. In this case the average lengths of the queue in the green and red periods are the same and depend only on the signal phasing, where $\lambda = g/c$ is the green proportion, $c$ is the cycle time (not to be confused with the randomness parameter $C$) and $\mu$ is the *capacity*, ie normally $gS/c$:

$$L_r = L_g = \frac{\mu(1-\lambda)c}{2} \tag{5.14}$$

When the flow is below saturation, which implies $\rho<1$, the average queue in the green period is less than that in the red period because it is able to discharge fully, so:

$$\begin{aligned} L_r &= \frac{\rho\mu(1-\lambda)c}{2} \\ L_g &= \frac{\rho^2\mu(1-\lambda)^2 c}{2(1-\lambda\rho)} \end{aligned} \tag{5.15}$$

If the degree of saturation in the previous time slice is $\geq 1$, the resulting queue will maintain the saturation of the stop line for a time:

$$t_d = \frac{L_0 - L_e(\rho)}{\mu(1-\rho)} \tag{5.16}$$

In this case the mean red and green phase queues in the current time slice are averages of (5.14) and (5.15) weighted by (5.16). The overall mean phase queue is the average of $L_r$ and $L_g$ weighted by the lengths of the red and green periods.

### *5.4. Queuing delay*

Most macroscopic traffic models need to calculate the total delay experienced by all travellers on each network link in a given time period. A packet-based model, in contrast, needs to calculate the delay to each individual packet. Although there are formulae for both quantities which are closely related to the sheared queue formula, they are not used partly for reasons of accuracy and partly because the capacity affecting a packet may not be constant. For *constant* capacity $\mu$, the mean time spent in a queue length $L$ is given by:

$$t_D = \frac{L}{x\mu} \tag{5.17}$$

Total delay to a vehicle, possibly accumulated over several time slices, is calculated by applying (5.17) iteratively to the queue remaining ahead of it at the start of each time slice, until clearance. The presence of $x$ in (5.17) reflects the fact that in general capacity is not fully utilised. For example, consider a steady-state queue, length $L$, generated by a flow rate of $\rho\mu$, persisting for an arbitrary period $T$. Total delay to all vehicles in the period is $LT$, and the number of vehicles arriving is $\rho\mu T$, so delay per arriving vehicle is $LT/\rho\mu T$, ie $L/\rho\mu$, not $L/\mu$. (5.17) is the generalisation of this to non-steady-state.

# 6. Other sources of delay

## *6.1. Speed/flow relationships*

COBA speed/flow functions have already been described in section 4.4. Their effect is to alter the average cruise speed along the running section of a link. Thus the effective delay is:

$$d_{SF} = D\left(\frac{1}{V_{SF}} - \frac{1}{V_0}\right) \tag{6.1}$$

where $D$ is the distance travelled and $V_0$ is the speed at zero flow. Queuing delay, as described in the previous section, is accounted separately, so if queuing delay is implicit in the speed/flow function care must be taken to ensure that the stop-line does not contribute any further delay, unless the specified link capacity is exceeded.

## *6.2. Geometric delay at junctions*

Geometric delay is the *extra* travel time which occurs when vehicles have to slow down to negotiate a turning movement. Semmens (1985a,b) and Burrow (1987) calculate this delay by working out the distance and time involved in decelerating before and accelerating after the junction, using deceleration and acceleration rates which depend on the speed change. The speed change in turn depends on the arriving and departing cruise speeds and the entry and exit speeds in the manoeuvre.

The basic model, where the junction approach, entry, exit and departure speeds are respectively $V_A$, $V_B$, $V_C$ and $V_D$, is:

| | | |
|---|---|---|
| Deceleration distance (m) | $d_{AB} = .5\ (V_A^2 - V_B^2)/a_{BA}$ | (6.2) |
| Acceleration distance (m) | $d_{CD} = .5\ (V_D^2 - V_C^2)/a_{CD}$ | |

where:

| | | |
|---|---|---|
| Deceleration rate (m/s$^2$)[7] | $a_{BA} = 1.06\ (V_A - V_B)/V_A + 0.23$ | (6.3) |
| Acceleration rate (m/s$^2$)[6] | $a_{CD} = 1.11\ (V_D - V_C)/V_D + 0.02$ | |

If an intersection is represented explicitly as a sub-network of links, and nodes are treated as points, it is necessary only to consider a single junction speed $V_J$, where $V_B = V_C = V_J$, and no distance is travelled within a node. Thus the geometric delay component is given by:

$$t_G = (V_A - V_J)/a_{BA} + (V_D - V_J)/a_{CD} - d_{AB}/V_A - d_{CD}/V_D \tag{6.4}$$

The speed $V_J$, or alternatively the whole delay $t_G$, associated with any turning movement can be entered explicitly in data. In general, although the above model presents no difficulties, deciding *automatically* what turning speed is appropriate to a given movement is not straightforward – in particular it is essential to identify correctly those movements which do *not* incur geometric delay, otherwise serious errors in routing could occur. Methods based on the geometry of junctions and the deflection of movements have been considered, but none is yet thought reliable enough to be incorporated in the model.

*6.2. Give-way junctions, including roundabouts*

Give-way junctions include priority junctions and roundabout entries, since roundabouts currently are modelled as sub-networks. At all types of give-way junction a simple model is adopted, in which capacity is a linear function of the opposing flows (see Figure 7):

$$\mu = S_I - \sum_{i=1}^{n} k_i q_i \tag{6.5}$$

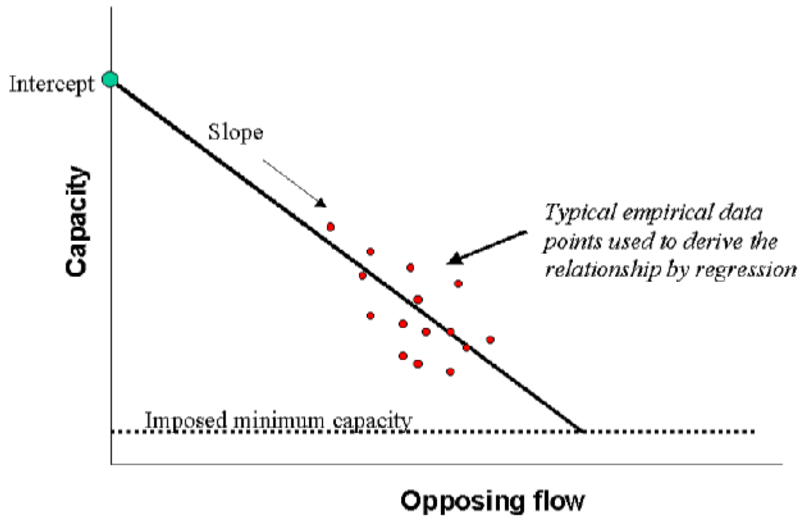


*Figure 7.* Linear relationship assumed for capacity versus flow at give-ways

where $S_I$ is the intercept capacity of the opposed link; $\{q_i\}$ are flows on the opposing links $[1..n]$; and $\{k_i\}$ are constants either determined empirically or calculated from geometry, possibly using the models described by Semmens (1985a,b). The intercept normally differs the saturation flow of the unopposed link, but defaults to the saturation flow specified for the link if no intercept is given.

### *6.3. Signals and signal plans*

Signal plans applies to whole junctions, and can be time-dependent or part-time. There is no formal association of signal plans with nodes (although signal plan may conveniently be identified with the nodes where they act). Each plan controls a specified set of approach links. On the continent of Europe, the term ‘plan’ often refers to the coordinated system of signals throughout an entire city, rather than at just one junction. Such a system *can* be modelled because in the above scheme there is no limit to the number of links controlled by each plan. Each signal plan has a cycle time, a number of stages, optional Webster and Cobbe (1966) split/cycle optimisation, and an offset relative to an arbitrary common zero which allows it to be coordinated with other plans. Fixed-time and some optimised plans will have stage lengths specified. Some signal models do not use the concept of a ‘stage’, recognising only movements or ‘phases’.
Stages are defined here as a logical framework upon which individual phases, with their start and end delays, are specified, so the stage lengths add up to the cycle time[8]. The number of stages in the plan must be consistent with that implied by signal arm (phase) data. The plan green time $g$ on each individual arm/phase is determined by the consecutive stages during which the signal is green, and the particular phase start and end delays specified in the data.

The capacity of a signal link is given by:

$$\mu = \frac{S(g - \delta_s + \delta_e)}{c} \qquad (6.6)$$

where $S$ is saturation flow, $g$ plan green time, which includes any phase delays, $c$ cycle time; $\delta_s$, $\delta_e$ are respectively global start and end displacements which allow for common vehicle starting, amber and red-amber effects (sometimes referred to as ‘lost time’ though this term is ambiguous and best avoided). Signal queues are calculated as separate random-and-oversaturation and phase components as described earlier in sections 5.2-5.3.

### *6.4. Signal coordination*

Signal coordination is allowed for by a simple green-wave model. It can occur only between signals whose cycle times are equal or related by a factor of two, and where the traffic stream is not interrupted by any intervening give-way junction which would add a random delay. The green wave proceeding downstream from a signal, which is assumed to extend uniformly across its green period, is mapped onto the green period of the next signal downstream, allowing for average transit time and the offset between the signals.

If, on average, a vehicle in the green wave leaving a signal always finds green at the next downstream signal then effectively it sees $g_{found}/c$ equal to 1, and by definition its coordination efficiency $e$=1. If, on average, a vehicle in the green wave finds green at the downstream signal in the same proportion as the actual green time, ie $g_{act}/c$, then by definition its coordination efficiency $e=0$. Thus the theoretical coordination efficiency is:

$$e^* = \frac{g_{found} - g_{act}}{c - g_{act}} \tag{6.7}$$

The practical coordination efficiency is modified by dispersion of the green wave which increases with travel time $t_{mm}$ (minutes), and by a user-supplied factor $f$ which can for example represent the effect of aging of fixed-time plans (Robertson and Hunt 1982), so:

$$e = \frac{fe^*}{1 + t_{mm}} \tag{6.8}$$

$e$ can be negative, for example if traffic enters the upstream signal from a minor arm, or in the direction opposite to that for which coordination is optimised. The coordination efficiency of a vehicle or packet affects not only its own probability of finding the next signal at green, but also, in aggregate, the length of queues there by altering the balance of flows arriving in the red and green phases. This may be approached by calculating the flow-weighted average coordination $E$ of the aggregate arriving traffic stream, and deriving modified effective arrival and capacity rates, as follows:

$$E = \frac{\sum_i q_i e_i}{\sum_i q_i} \qquad \text{over all packets } i\text{, size } q_i\text{, in the time slice} \tag{6.9}$$

$$\rho^* = \frac{(1 - E)\rho}{1 - E\rho} \tag{6.10}$$

$$\mu^* = (1 - E\rho)\mu \tag{6.11}$$

These transformed parameters are then used in the signal queue calculations, but are not correctly applicable to the random queue and delay calculations.

Although individual coordinated signals can be split-optimised provided their cycle times are unchanged, optimising a network of coordinated signals calls for a more complicated procedure, such as that implemented in TRANSYT (ref.).

## *6.5. Refinements to basic capacity models*

The basic approach to modelling signalised junctions, as described in sections 6.3-4, may be extended in a number of ways to deal with particular issues that arise in practice, for example:

- Flared approaches
- Opposed turns
- Part time signals

A method of allowing for flare based on Webster and Cobbe's (1966) method[9] uses the storage capacity $N_f$ and exit saturation flow $S_f$ of the flare as proxies for its geometrical properties. According to this simple model, flare at a signal affects both the effective saturation flow of the link as a whole, and the effective green time. The effective saturation flow is calculated for an extended traffic column and so is not simply the sum of the stop line saturation flows. If $S$ is the main link saturation flow, and $g$ the initial effective green time, the maximum additional discharge of the flare per cycle $N_q$, and the new effective combined values $S'$ and $g'$ are given by the following, where the parameter $\gamma$ is equal to .5 for a triangular shaped flare area:

$$N_q = \min\left(\gamma g S_f, N_f\right) \qquad (6.12)$$

$$S' = S + S_f\ (1-(N_q / N_f)^{\gamma}) \qquad (6.13)$$

$$g' = g + N_q / S' \qquad (6.14)$$

An opposed turn represents a signalised link which gives way to another signalised link, for which a model has been described by Burrow (1987). The weighted average saturation flow $S_W$ of the movement is calculated from the opposed and unopposed saturation flows $S_O$, $S_U$, and effective green times $g_O$, $g_U$ respectively:

$$S_W = \frac{S_O g_O + S_U g_U}{g_O + g_U} \qquad (6.15)$$

The unopposed saturation flow $S_U$ is that which would otherwise apply to the opposed link, but the opposed saturation flow is calculated empirically as:

$$S_O = \frac{S_U}{1 + \frac{1.5}{r_t} + f_p} + S_C \qquad (6.16)$$

where $r_t$ is the radius of the turn (assumed to be 12 metres in the absence of data), $f_p$ is a factor which takes account of the need to give way to the opposing traffic and $S_C$ is the 'clearance term', an extra amount of saturation flow which applies only when the opposed and opposing green phases end simultaneously. These terms are defined below:

$$f_p = \frac{12{x_O}^2}{1-(f_t x_O)^2\left(1+0.6(1-f_t N_s)\right)} \tag{6.17}$$

$$S_C = \frac{3600(N_s+1)(f_t x_O)^{0.2} p_{cu}}{g_O + g_U} \tag{6.18}$$

In (6.17-18) $x_O$ is the maximum degree of saturation of the opposing flows (subject to maximum of 1.0), $f_t$ is the proportion of traffic in the 'opposed' lane which is turning traffic, $N_s$ is number of vehicles which can be stored in the turn without blocking non-turning traffic, and $p_{cu}$ is the PCU factor applying to the turning traffic. If the greens do not overlap under the current plan the movement can be modelled as unopposed.

Part time signals are characterised by a signal plan which operates for a limited range of time slices. When the signal is inactive, a give-way process will apply if it has been defined. When the signal is operating, the normal signal model will apply. Part time opposed signalised movements are rare.

## 7. Blocking back of queues

The term blocking back refers to a queue building back from a bottleneck until it crosses the next junction upstream and affects links entering it, which may carry some traffic not destined to pass through the bottleneck. Once this occurs, the capacities and queuing processes at the 'blocked' upstream stop lines are no longer independent of those downstream. Under severe overload, queues can spread until much of the network is gridlocked.

Microscopic models in principle need not concern themselves about this process, simply feeding vehicles through when space becomes available ahead. Working through the traffic model alone – eg periodically checking the queue lengths and adjusting them - is thought to be both too slow-reacting and too difficult to make consistent with the packet loading. On the other hand, the feedback nature of the blocking back process means that local adjustments must be handled carefully to avoid the risk of instability.

Packet loading allows a compromise approach, illustrated by Figure 8. After each packet has been loaded, its route is examined in reverse. On each link D, it monitors the length of the queue which would occur if only the packets so far processed for the current time slice were to contribute to it. This is known as the 'current' queue, as distinct from the 'usual' queue which includes the rolling sum of all arrivals in the time slice. Once this queue reaches the storage capacity, it is assumed that no further vehicles can enter D during the time slice, so the *average* capacity of each upstream link U must equal the actual throughput of U up to that point. While the reduced capacity of U should deliver the correct final queue length on U, it does not represent the conditions faced by packets after the ‘onset’ of blocking back, since earlier packets have experienced a much higher capacity. Therefore, the later packets have to be held back artificially, so as not to depart from U until the next time slice.

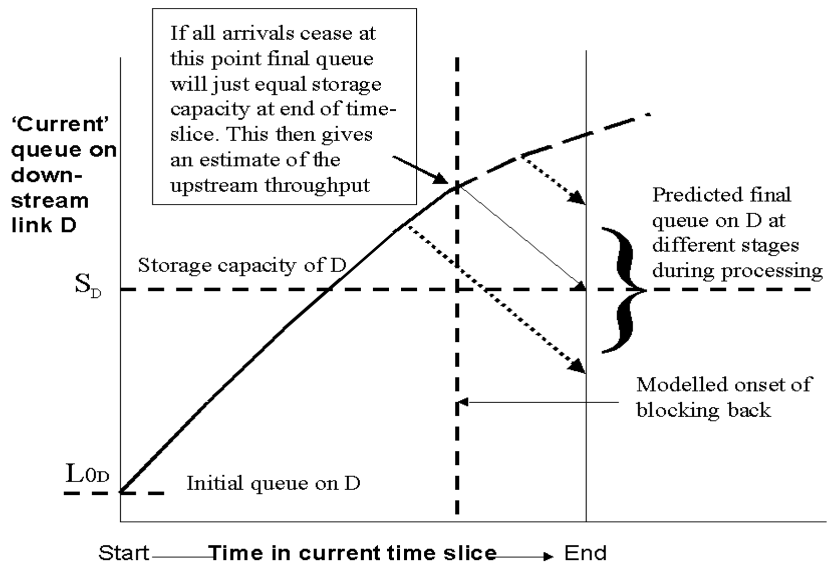


*Figure 8.* The method of calculating blocking back

Recently it has been found that this model overpredicts queues blocking back from short undersaturated links. The condition of the short queuing link D is characterised by $x_D$<1 in the notation of section 5.2. Due to the linkage of their queues, this will also apply to upstream links U, ie $x_U$<1. If the reduced capacity of U is assumed to equal its throughput, $x_U\mu_U$, then it will underestimate the true capacity $\mu_U$, so queues on U will be overestimated. The overestimation can be severe if it causes U to appear oversaturated when it is not.

The solution approach is essentially steady-state and probabilistic, although it avoids considering probability distributions explicitly. First, if $S_D$ is the maximum storage capacity of D then the mean queue length on D when blocking back is not $S_D$, but:

$$L_D \approx S_D x_D \tag{7.1}$$

since $x_D$ is roughly the fraction of time that a queue is present on D, and the shortness of D means that any queue has a high probability of filling it. Substituting this into (5.3) to get the equilibrium queue length on D alone implies that the randomness on D changes to:

$$C_D = \frac{(S_D - I)(1 - x_D)}{x_D} \tag{7.2}$$

This reduction in randomness reflects the fact that once the link is full it is only weakly affected by variations in the length of the upstream 'tail'. By treating the queue on U as an independent process with randomness $C_U$, such that $L_U$ and $L_D$ add up to the queue that could have been produced by (5.3), and assume that $x_U \approx x_D$ then:

$$C_U = C - C_D \tag{7.3}$$

In the case of U, there is no vehicle-in-service I, since the only real stop line is on D. Finally, the reduced capacity of U is calculated from the modelled throughput with an adjustment to reflect its incomplete utilisation:

$$\mu_{U(red)} = \frac{\theta_{U \to D}}{x_U} \tag{7.4}$$

Given (7.2-7.4) the queue lengths can be calculated in the usual way, link by link. Investigation of the blocking back model is still going on. Like that of horizontal and motorway queues referred to earlier on section 5.2 its further development will depend on resources becoming available.

## 8. Fuel consumption and emissions

CONTRAM calculates fuel consumption and line-emissions by link, user class and time slice, and also fuel consumption for the whole network by time slice. It supports several types of fuel/emissions function with user-defined coefficients and vehicle types, currently dependent on average link speed, including:

- MODEM – derived from EC DRIVE MODEM project outputs (Jost *et al* 1992)
- RR249 – as used in CONTRAM 5, including fuel used by cars while queuing
- SCOOT – extended functions for emissions only, as used in SCOOT (ref.)

These models refer back to work by Everall (1968), which showed that fuel consumption per unit distance tends to be minimum at some speed, rising both at low speeds due to idling losses and at high speeds due to increasing resistance to motion or reduced engine efficiency. This is represented by the formula (8.1) (illustrated by Figure 9) in which $V$ is average speed, $D$ is distance, and $Q$ is the quantity of the product considered.

$$Q = \left( a_o + \frac{a_{-1}}{V} + a_2 V^2 \right) D \tag{8.1}$$

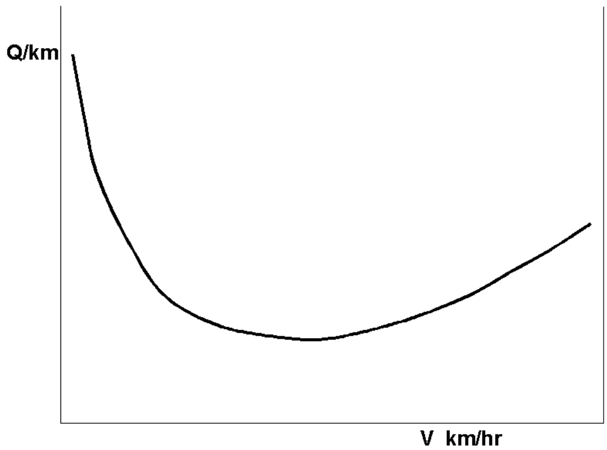


*Figure 9.* Form of Everall's function of average speed

These functions can be mixed provided that the units are consistent. The 'MODEM' functions are the simplest, and also the most complete because they estimate both fuel consumption and emissions, and are founded on the largest database. They have been derived by regressing (8.1) on tables given in Jost *et al* (1992), based in turn on driving cycles and measurements conducted during the DRIVE MODEM project. Their coefficients $\{a_i\}$ have been calculated for a range of vehicle types including:

- Pre-1985 petrol cars *non* catalyser (1400 cc, 1400-2000 cc and >2000 cc models)
- Pre-1990 petrol cars *non* catalyser (1400 cc, 1400-2000 cc and >2000 cc models)
- Petrol cars equipped with catalyser (1400 cc, 1400-2000 cc and >2000 cc models)
- Diesel cars (<2000 cc, and >2000 cc including trucks up to 3.5t)

In addition, information is available on the relative proportions of these vehicle types in within the total 'fleet'. CONTRAM allows these to be defined for each user class.

It is recognised that both the formulae and the fleet proportions need to be kept up to date. This is relatively easy to do for fleet proportions, but updating the emissions database is expensive and time consuming. The framework for these data in CONTRAM has been designed to be flexible so that new formulae, and even new types of function, can be introduced as and when they become available. For example, COBA 11 (ref.) vehicle operating cost and some of the new COPERT III (ref.) vehicle emissions formulae use an alternative form to (8.1) in which the 1/V term is replaced by a term in V. This option is supported by a development version of CONTRAM.

## 9. Intelligent Transport Systems

CONTRAM has supported ITS since the early 1990s. It is the assignment engine in the Highways Agency's (HA) MOLA (Motorway On-Line Advisor) software, for assessing alternative diversion strategies on the Kent motorway and trunk-road corridor in south-east England, linking London with the Channel ports and Tunnel (Harbord and Still 1997). Another variant, CONTRAM-I, was developed by Transportation Research Group (TRG) at Southampton University, under contract to TRL, to explore driver response to incidents in the absence of external information (van Vuren and Leonard 1994).

These functions have been added to the present release of CONTRAM thanks initially to funding from the Swedish National Road Administration, which is using the program for research and development of traffic management systems (Davidsson and Taylor 2003). An enhanced version will provide the assignment engine for Tactical MOLA which is being developed for the HA to take account of a wider range of impacts than just delay, and to model potential Active Traffic Management (ATM) measures.

Data for an 'ITS' run comprise four main parts:

- A set of baseline packet routes output by a previous non-incident run to simulate the 'typical' traffic pattern under normal conditions.
- An incident or other change to the network represented for example by a temporary reduction in capacity or possibly cruise speed on a link.
- A set of information sources defined by their type, location, period of activity, user selection criteria and content.
- A set of response rates specifying the proportion of packets of a given class which will respond to each type or individual source of information.

The table below shows types of information **currently** implemented.

| **Information type** | **Location** | **Available to** | **Content** |
|---|---|---|---|
| VMS[10] | Link | All classes | Diversion route |
| QUEUE | Everywhere | All classes | Self-diversion criteria |

Selection criteria and content are specific to the type of information and are given in keyword form to allow the flexibility to accommodate many different types of information. For example, a VMS, which must be located on a specific link, may have attributes such as ‘dest=904’, which means that only packets destined for zone number 904 are affected, and ‘route=102’, meaning that those which respond will take diversion route number 102, which may extend all the way to the destination, or may just by-pass the incident in which case the packet must find the rest of its way using standard minimum cost routing.

For QUEUE type ‘information’, which is not tied to particular sites, the model looks a specified distance ahead along a packet’s intended route (possibly extending over several links) and estimates the proportion of it occupied by a queue. If this exceeds the specified threshold the packet may seek an alternative route to its destination based on minimum uncongested journey time, a proxy for limited information about the network state.

In each case once the potential for response has been established, actual response depends upon the response rate specified. The model keeps a tally of the total volume of traffic of each class responding or not responding to a particular information source and allocates each packet to one or other set so as to stay as close as possible to the specified proportion.

The table below shows some types of information which might be supported in the **future**. These would require new or extended procedures to be implemented within the model.

| **Information type** | **Location** | **Available to** | **Message content** |
|---|---|---|---|
| RDS-TMC | Areas (cells) | Equipped users | Data on incidents or delays |
| DRG | Everywhere | Equipped users | Optimum/diversion routes |
| Pre-Trip | Origins | Proactive users | Recommended routes |

## 10. Conclusion and future developments

### *10.1. The contribution of CONTRAM*

CONTRAM represents an attempt to model what happens in the real world of traffic, by representing processes at an appropriate level of aggregation or detail, rather than trying to simulate every event or impose upon the world a theoretical regularity which it may not actually possess. Although its assignment and traffic models stem from theoretical principles, its packet modelling approach is not compatible with minimisation of a formal objective function. Nevertheless, in combination they provide a relatively efficient and flexible way to represent the real world and to focus detailed analysis where it can be most useful.

The model's structure places few limitations on the way processes like queuing, capacity, route choice and information could be modelled, and so in principle the best available validated methods, whether theoretical or empirical, can be adopted. Thus it should be possible to keep pace with most developments in the understanding and control of traffic, and driver behaviour and route choice.

### *10.2. Development of CONTRAM*

The UK government's Integrated Transport White Paper (ITWP 1998) quotes estimates of the cost of congestion from £7 billion to £15 billion per annum. Therefore even a small proportional reduction could yield huge benefits. Much of the recent development of CONTRAM has been funded by projects devoted to particular research or policy issues (see later in Acknowledgements). These aim to provide and test new functionality, not necessarily restricted to implementation by CONTRAM. However, as long as CONTRAM and other traffic models are defined as 'commercial products', but lack an effective 'market' mechanism to recover the value of congestion they could help to prevent, most general development has to be funded from licence income alone, which after subtracting maintenance overheads can be very limited. This can be a significant handicap to further development of the core modelling and computational methods.

### *10.3. Non-equilibrium modelling*

Aside from its ITS features, CONTRAM has been presented here as an equilibrium assignment model. Effects of pre-trip information, sensitivity to travel time variability, and longer term response to travel are technically outside the scope of assignment proper, except possibly stochastic assignment which includes perception errors in general terms. A progression of models could be envisaged as follows, with travel choice determined by:

(1) network state alone *(eg: current equilibrium assignment)*
(2) network state *plus* other information *(eg: current ITS model)*
(3) *perception of* network state *plus* other information[11]
(4) *perception of* network state *plus* other information *plus* learning from experience

At each step the travelling population is further disaggregated according to the information available to it and its behavioural characteristics, and the distinction is widened between the objective state of the network and the subjective perception of travellers, though these 'layers' could still be contained in a common macroscopic model framework. Objective-minimising macroscopic models would require functional integration of these diverse components, which could be hard to formulate, while microscopic approaches in full detail would require complex behavioural models, which could be hard to calibrate or verify. The mesoscopic approach appears well suited to dealing with moderately disaggregate problems.

**Acknowledgements**

In addition to internally funded development, significant contributions have been made by research projects devoted to specific methodological development and applications including:

- Congestion due to incidents UK Department for Transport (DfT)
- VMS/MOLA European Commission, English Highways Agency
- ITS enhancements Swedish National Road Administration
- Travel Time Variability DfT
- Variable Demand Modelling DfT
- Active Traffic Management English Highways Agency

Thanks are due to Ian Burrow for internal review and to two anonymous Referees for extensive comments. This paper is produced with the support of Research Director, TRL.

**References**

Burrow I J (1987). "*OSCADY: A computer program to model capacities, queues and delays at isolated traffic signal junctions*". TRL Report RR 105. TRL, Crowthorne, UK.

COBA (2003). "*Design Manual for Roads and Bridges. Vol 13 Economic assessment of roads schemes*". Section 1: The COBA Manual. HMSO.

COBA 11. http://www.roads.dft.gov.uk/roadnetwork/heta/highway/index2.htm

CONTRAM. **www.contram.com**. Includes bibliography of research, applications and related papers.

COPERT III. http://vergina.eng.auth.gr/mech/lat/copert/copert.htm

Davidsson F and Taylor N (2003). "ITS modelling in Sweden using CONTRAM". *Proc. ITS UK Smart Moving Conference,* Birmingham, April 2003.

Dijkstra E W (1959). "A note on two problems in connexion with graphs". *Numerische Mathematik 1, p269-271.*

DMRB. *Design Manual for Roads and Bridges*. Department for Transport, Gt. Marsham St, London.

Emmerson P. and N. Paulley (TRL), P. H. Bly, T. van Vuren (Mott Macdonald) and S. Porter (DfT). (2003). "Simplifying multi-stage modelling advice while retaining the essentials". In *Proc. European Transport Conference, Strasbourg, October 2003.*

Everall P F (1968). "*The effect of road and traffic conditions on fuel consumption*". TRL Report LR226. TRL, Crowthorne.

FHWA (2000). *Highway Capacity Manual*. US Federal Highway Administration, Washington DC.

Gordon A, White C and Porter S (2003). "Dynamic Integrated Assignment and Demand Modelling - DIADEM." *Paper prepared for DfT Variable Demand Modelling Seminar, July 2003*.

Greenwood D G and Taylor N B (1993) (University College London and TRL). "Parallelising the CONTRAM traffic assignment model". *Proc. ESM'93 Simulation Multiconference*, Lyon, June 1993.

Harbord B and Still P (1997) (Highways Agency and TRL). "MOLA - Development of a real time network model". *Proc. IEE Colloquium on strategic control of inter-urban road networks*, London, March 1997.

Hazelton M and Gordon A (2002) "Estimation of origin-destination trip matrices from link counts". *Proc. European Transport Conference 2002.*

Heydecker B and Verlander N (1998) (University College London). "Transient delay in oversaturated queues". *Proc. 3rd IMA International Conf. on Mathematics in Transport Planning and Control, Cardiff, 1-3 April 1998.*

Heydecker B and Verlander N (1999) (University College London). "Calculation of dynamic traffic equilibrium assignments". *Proc. Seminar F, European Transport Conference, 27-29 September 1999, vol P434, p79-91*. PTRC, London.

ITWP (1998). *Integrated Transport White Paper*. Department for Transport, London, July 1998. http://www.dft.gov.uk/itwp/paper/index.htm

Jost P, Hassel D, Weber F J and Sonnborn K-S (1992). "*Emission and fuel consumption modelling based on continuous measurements*". Deliverable No. 7 of DRIVE Project V1053 MODEM. TÜV Rheinland, in partnership with INRETS, TRL and CEDIA.

Kimber R M and Daly P (1986). "Time-dependent queueing at road junctions: observation and prediction". *Transportation Research Vol 20B, No 3.* Pergamon.

Kimber R M and Hollis E M (1979). "*Traffic queues and delays at road junctions*". TRL Report LR 909. TRL, Crowthorne, UK.

Leonard D R, Gower P and Taylor N B (1989). "*CONTRAM: structure of the model*". TRL Report RR 178. TRL, Crowthorne, UK.

Robertson D I and Hunt P B (1982). "A method of estimating the benefits of coordinating signals by TRANSYT and SCOOT". *Traffic Engineering and Control, November 1982.*

SATURN. **www.its.leeds.ac.uk/software/saturn**.

SCOOT. www.trlsoftware.co.uk/productSCOOT.htm

Semmens M C (1985a). "*ARCADY2: An enhanced program to model capacities, queues and delays at roundabouts*". TRL Report RR 35. TRL, Crowthorne, UK.

Semmens M C (1985b). "*PICADY2: An enhanced program to model capacities, queues and delays at major/minor priority junctions*". TRL Report RR 36. TRL, Crowthorne, UK.

Taylor N B (1989). "Speeding up quickest route assignment in CONTRAM with an heuristic algorithm". *Traffic Engineering and Control, February 1989.*

Taylor N B (1990). "*CONTRAM 5: an enhanced traffic assignment model*". TRL Report RR 249. TRL, Crowthorne, UK.

TRANSYT. **www.trlsoftware.co.uk/producttrans.htm**

Van Vliet D (1977). "D'Esopo: a forgotten tree-building algorithm". *Traffic Engineering and Control, Jul/Aug 1977.*

van Vuren T and Leonard D R (1994). "Urban congestion caused by incidents". *Traffic Engineering & Control, Jul/Aug 1994.*

Webster F V and Cobbe B M (1966). "Traffic signals". HMSO.

White C, Taylor N B and Hounsell N (1994). "CONTRAM - A computer suite for modelling road congestion". *Traffic Technology International '94. UK & International Press.*

Whiting P D and Hillier J A (1960). "A method of finding the shortest route through a road network". *Operational Research Quarterly, vol 11 nos 1-2, p37-40.*

## Notes

[1] TRL formerly known as the Transport Research Laboratory
[2] If not given, this is estimated from the link length, saturation flow and the length of road occupied by a stationary car (currently assumed to be 5.75m)
[3] Time-average is over the period $(0,t)$, mean is over many random trials.
[4] M and D stand for Markovian and Deterministic respectively. Numeral is the number of parallel servers.
[5] (5.4,5.5) suggest an alternative interpretation to Figure 6. In this, the variable average degree of saturation *at the stop line*, $x$, which can never exceed 1 regardless of the overall demand $\rho$, generates at time $t$ both a deterministic queue (5.4) and a numerically equal quasi-steady-state queue (5.5). Since the queue is *not* actually in a steady state, the solution cannot be exactly correct for finite $t$.
[6] Kimber and Hollis (1979) point out that the sheared model in some cases overestimates queue length for finite values of $t$. The topic of the accuracy and other properties of the queuing model are addressed in some detail by Heydecker and Verlander (1998).
[7] See Semmens (1985a,b)
[8] In Taylor (1990) 'stages' were defined separated by 'intergreens', so normally added up to less than the cycle time, but individual phase delays were not considered.
[9] Page 92 of Webster and Cobbe (1996) contains a misprint where a prime has been omitted in the equation for gain in effective green time for the case where the dischargeable queue is shorter than the flare.
[10] Variable Message Sign. This mode could also be used to simulate Variable Direction Signs.
[11] RGCONTRAM, developed by TRG to study route guidance, allowed 'distortion' of link costs (White *et al* 1994)