2.4.10.9 Some Aspects of Travel Time Modeling

As cost is not flow-dependent, it is a given model input. On the other hand, time is modeled by the help of flow-dependent functions and therefore represent a less certain input of the assignment model. But the quality of the forecasted volumes in toll projects depends largely on path time and related cost. Therefore the modeling of the flow-dependent time on links and nodes requires more attention than in ordinary planning projects without toll. The modeler has to consider especially the following aspects:

- For trips which originate or terminate outside the scope of the model, only a part of their path is covered by the model. Therefore it is not possible to evaluate the total travel time and thus it is not recommended to apply the same value of time distribution as for internal trips.
- Link and node flows that exceed capacity should be avoided. In this case capacity restraint functions do not produce realistic travel times. Especially if peak hours are modeled, the capacities of highly charged links need to be defined very carefully.

2.4.11 Dynamic User Equilibrium (DUE)

2.4.11.1 Introduction and Application Areas

The quantitative analysis of road network traffic performed through static assignment models yields the transport demand-supply equilibrium under the assumption of within-day stationarity. This implies that the relevant variables of the system (i.e. user flows, travel times, costs) are assumed to be constant over time within the reference period. Although static assignment models satisfactorily reproduce congestion effects on traffic flow and cost patterns, they do not allow to represent the variation over time of the demand flows (i.e. around the rush hour) and of the network performances (i.e. in presence of time varying tolls, lane usage, signal plans, link usage permission); most importantly, they cannot reproduce some important dynamic phenomena, such as the formation and dispersion of vehicle queues due to the temporary over-saturation of road sections, and the spillback, that is queues propagation towards upstream roads.

The Within-Day Dynamic Traffic Assignment (WDDTA) models are conceived to overcome this limit. Among them, the Dynamic User Equilibrium (DUE) model embedded within VISUM presents several new and unique features, which will be outlined in the following sections, yielding an algorithm highly efficient both in terms of memory usage and computing time. Thus, this model can be applied to large networks (hundreds of zones and up to one hundred thousand links and nodes) with long periods of analysis (possibly the entire day), and is particularly suitable for the following application fields:

- Simulation of heavily congested urban and extra urban networks, where oversaturation conditions and the back propagations of congestion among adjacent roads are present over a large part of the network for several hours each day;

- Simulation of networks with transient congestion effects, leading to route choice varying during the assignment period;
- Simulation of networks in presence of dynamic management and/or time varying access policies, such as time varying tolls, lane usage, signal plans, link usage permission;
- Simulation of incident effects and incident management;
- Simulation of evacuation plans, in particular when the maximum evacuation time is needed.

The aim of this chapter is to give a complete overview of the model underlying the Dynamic User Equilibrium method implemented in VISUM. However, in order to improve readability, any bibliographic reference is omitted, along with many analytic proofs. For those, and for a deeper insight into the model and/or the theories underlying it, the reader may refer to the bibliographic section, which includes all the scientific papers on which this model is based.

2.4.11.2 Overview of the Model

This model is aimed at solving the *Within-Day Dynamic Traffic Assignment* (WDDTA) on road networks addressing explicitly the simulation of queue spillovers. It is based on a macroscopic and continuous-time formulation *Dynamic User Equilibrium*.

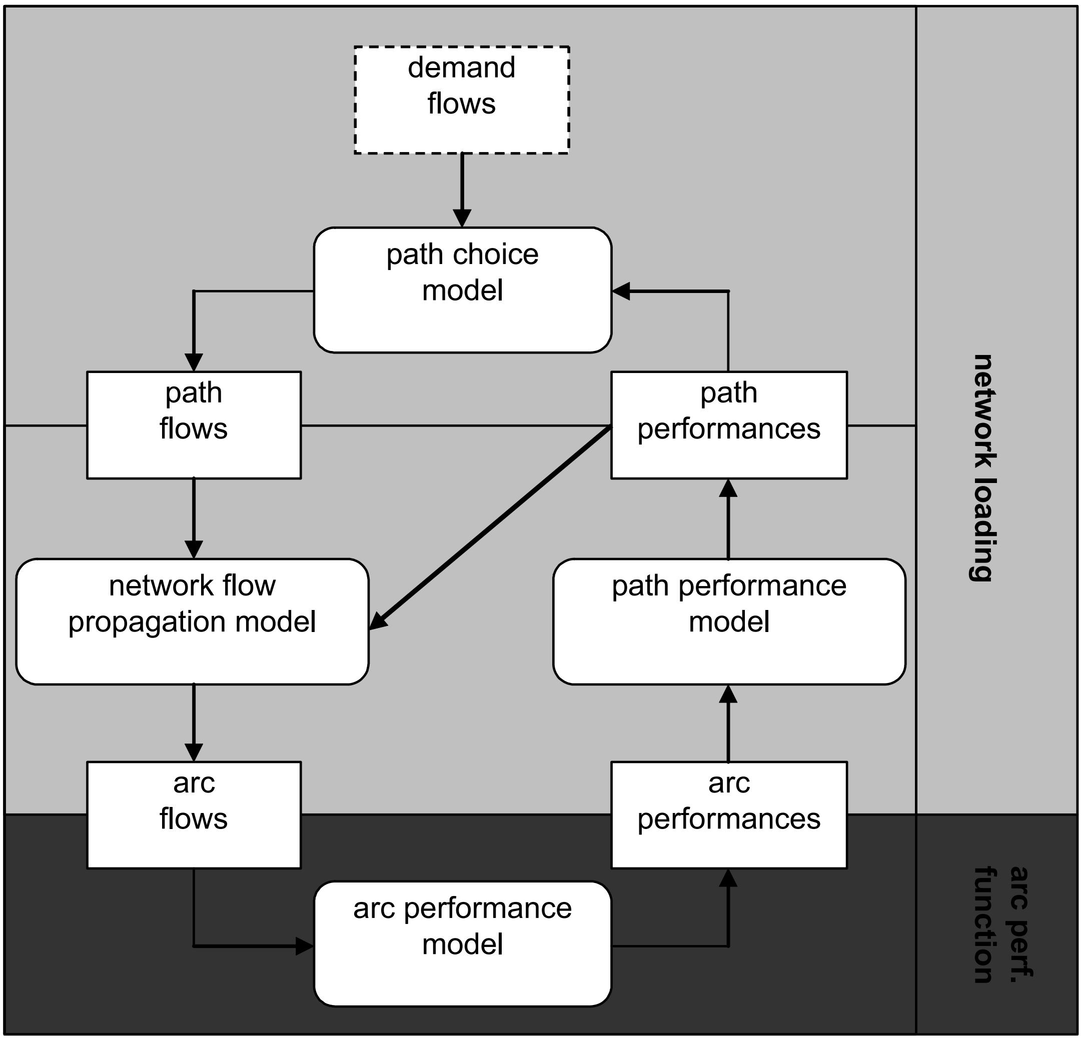

***Illustration 73**:* *Dynamic User Equilibrium Problem*

Apart from the temporal dimension, the main difference between the static and the dynamic user equilibrium relates to the consistency constraints between arc and path model variables. While in the static case these constraints involve only the spatial dimension of the system, in the dynamic case they concern the temporal dimension also. More specifically, for given path flows, the determination of the arc flows, which in the static case requires only the arc-path incidence matrix, in the dynamic case involves also the travel times on the network; that is, the network flow propagation model depends also on the path performances (diagonal arrow in **Illustration 73**).

The present formulation of the WDDTA has two key novelties compared to existing WDDTA methods:

1. Instead of a simulation approach, it adopts a temporal profile approach, where the value of a given variable of the problem (i.e., the variable temporal profile) is determined as a function of time for the entire period of analysis, based on the temporal profiles of the other variables of the problem, which are assumed to be fixed to their current value; this approach, conceptually depicted on the right hand side of **Illustration 74**, has an iterative nature, since each variable has to be recalculated until a convergence is achieved.

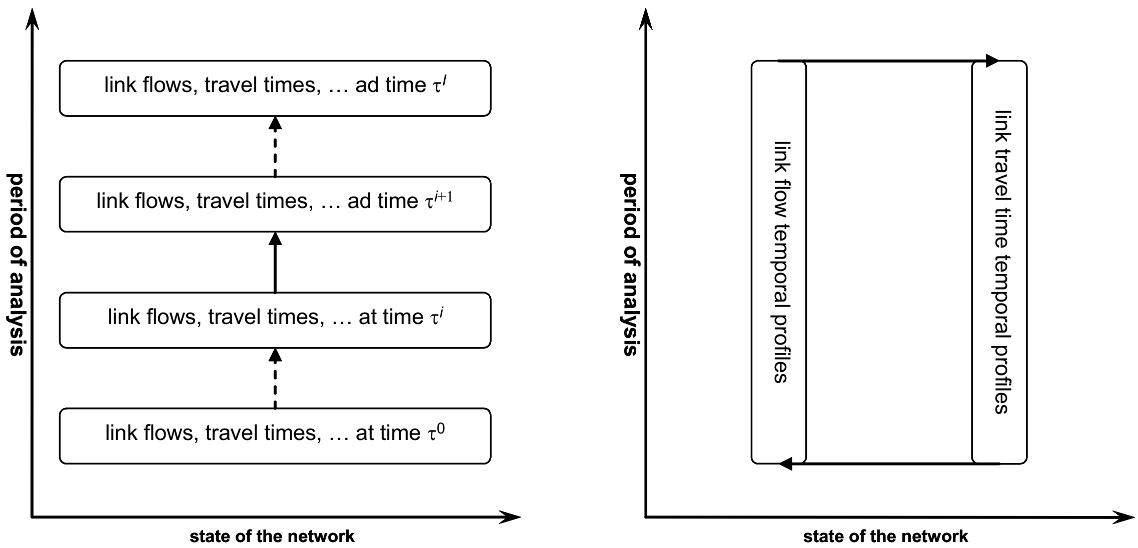

Illustration 74*: Temporal layer approach (left side) and temporal profile approach (right side) to the Continuous Dynamic Network Loading problem*

2. Spill-back can be modelled explicitly simply by switching between two alternative network performance models: without spillback, arc performance (the relationship between arc inflow and outflow profiles) depends only on the properties of that arc; with spillback, capacities upstream of bottlenecks are reduced so that arc storage capacities are not exceeded (**Illustration 75**).

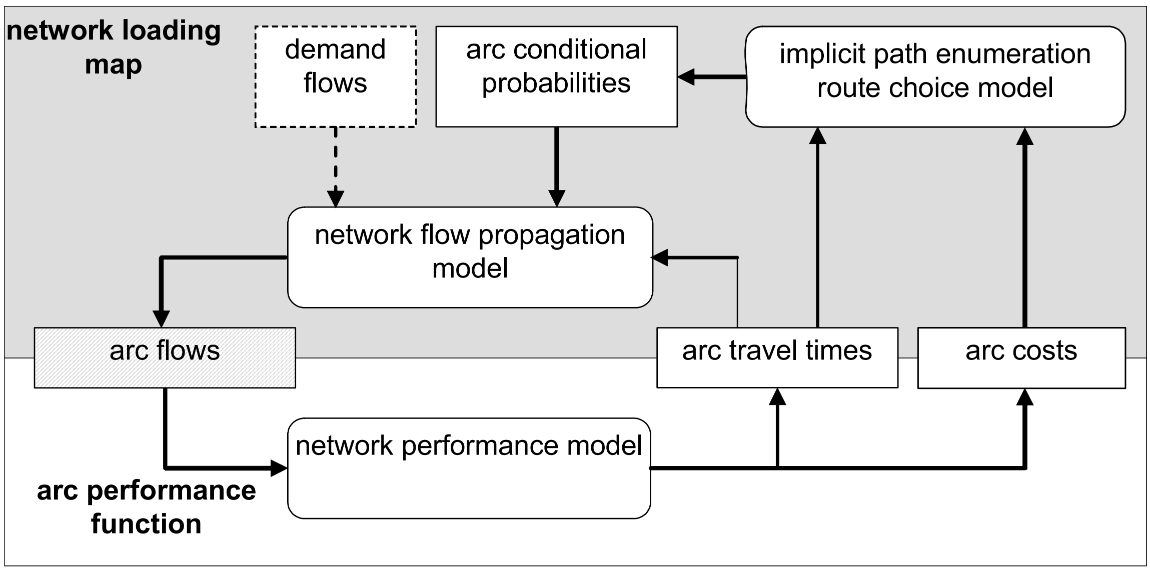

Illustration 75*: Scheme of the fixed point formulation for the WDDTA with spillback congestion*

3. The path choice model can adopt either a deterministic view where only objectively least-cost paths are loaded, or a Probit view where impedances are perturbed stochastically to reflect subjective user perceptions.

This approach presents several advantages:

- Consistency between path and link flows (network loading) is achieved in the same iteration as the equilibration between demand and supply, avoiding nested loops;
- An implicit path approach generates rational path probabilities without the need to enumerate all paths;
- As a major advantage of the temporal profile approach, the assignment period may be subdivided into long time intervals (typically 5-15 minutes), instead of a few seconds for the simulation approaches, saving computation time and memory. This allows overcoming the difficulty of solving WDDTA instances on large networks and long periods of analysis;
- The complexity of the algorithm is roughly equal to that of a static assignment multiplied by the number of (long) time intervals introduced.

For queue spillover modelling, the interaction among the flows on adjacent arcs is propagated in terms of time-varying arc exit capacities. The approach is then to reproduce the spillback phenomenon as a hypercritical flow state, either propagating backwards from the final section of an arc and reaching its initial section, or originating on the latter that reduces the capacities of the arcs belonging to its backward star and eventually influences their flow states.

The manual is organized as follows: In section 2.4.11.3, the main variables underlying the continuous model are introduced, along with some significant results of traffic flow theory underlying the presented models; the Network Performance Model and its submodels are addressed in section 2.4.11.4, while section 2.4.11.5 is devoted to the Network Loading Map; in section 2.4.11.6 the Dynamic User Equilibrium model is presented, both for the deterministic and Probit case, and a numeric example is presented and analyzed in section 2.4.11.7. Finally, the last section contains all the bibliographic references related to the model.

2.4.11.3 Mathematical Framework

As the analysis is carried out within a dynamic context, the model variables are temporal profiles, here represented as piecewise continuous functions of the time variable τ.

Users trips on the road network are modelled through a strongly connected oriented graph $G = (N, A)$, where N is the set of the nodes and $A \subseteq N \times N$ is the set of the arcs. Each link, turn, and connector in the VISUM network corresponds to an arc. Each network node and zone corresponds to a graph node.

Each arc a is identified by its initial node $TL(a)$, referred to as tail, and by its final node $HD(a)$, referred to as head; that is: $a = (TL(a), HD(a))$. Example: for an arc a representing a link, $TL(a)$ would correspond to its FromNode and $HD(a)$ to its ToNode. The forward and backward star of node $x \in N$ are denoted, respectively, $FS(x) = \{(x, y)\in A: x = TL(a)\}$ and $BS(o) = \{(x, y)\in A: y = HD(a)\}$. The zones constitute a subset $Z \subseteq N$ of nodes.

When travelling from an origin node $o\in N$ to a destination node $d\in Z$ users consider the set K_{od} of all the paths connecting o to d on G. We are interested in the many-to-one shortest path problem from each node $o\in N$ to a given destination $d\in Z$. Graph G is assumed to be strongly connected, so that K_{xd}, with $x\in N \neq d\in Z$, is non-empty.

Path topology is described through the following set notation:

$A(k)$ concatenated sequence of arcs constituting the path $k\in K_{od}$ from $o\in N$ to $d\in Z$.

With reference to the network flow pattern the following notation is adopted:

$D^{od}(\tau)$ demand flow of vehicles travelling from origin $o\in N$ to destination $d\in Z$ departing at time τ;

$f_a(\tau)$ flow of vehicles entering arc $a\in A$ at time τ,

$F_a(\tau)$ cumulative flow of vehicles entering arc $a\in A$ at time τ;

$u_a(\tau)$ outflow from arc $a\in A$ at time τ.

By definition, $$F_a(\tau) = \int_{-\infty}^{\tau} f_a(\sigma)\cdot d\sigma \quad (1)$$

For the calculation of network performance, travel times are introduced through entrance-exit functions, and the following notation is adopted:

$c_a(\tau)$ cost of travelling through arc $a\in A$ for vehicles entering it at time τ;

$t_a(\tau)$ exit time from arc $a\in A$ for vehicles entering it at time τ;

$t_a^{-1}(\tau)$ entry time on arc $a\in A$ for vehicles exiting it at time τ;

$C_k(\tau)$ cost of path $k \in K^{od}$ from $o\in N$ to $d\in Z$ for vehicles departing from node o at time τ;

$T_k(\tau)$ exit time from path $k \in K^{od}$ from $o\in N$ to $d\in Z$ for vehicles departing from o at time τ;

Due to the presence of time-varying costs, it may be convenient to wait at nodes in order to enter a given arc later. In the following, it is assumed that vehicles are not allowed to wait at nodes, but paths with cycles may result. However, the shortest paths include at most a finite number of cycles.

Since waiting at nodes is not allowed, the path exit time $T_k(\tau)$ is the sum of the travel times of its arcs $A(k)$, each of them referred to the instant when these vehicles enter the arc when travelling along the path. Moreover, assuming that path costs are additive with respect to arc costs, its cost $C_k(\tau)$ is the sum of the costs of its arcs $A(k)$, each of them referred to the time when they enter the arc when travelling along the path. The exit time and the cost of path k can then be retrieved, respectively, through the following recursive expressions:

$T_k(\tau) = T_h(t_a(\tau))$, (2)

$C_k(\tau) = c_a(\tau) + C_h(t_a(\tau))$, (3)

where $a = (o, x) \in A$ is the first arc of k and $h \in K_{xd}$ is the rest of path k (**Illustration 76**).

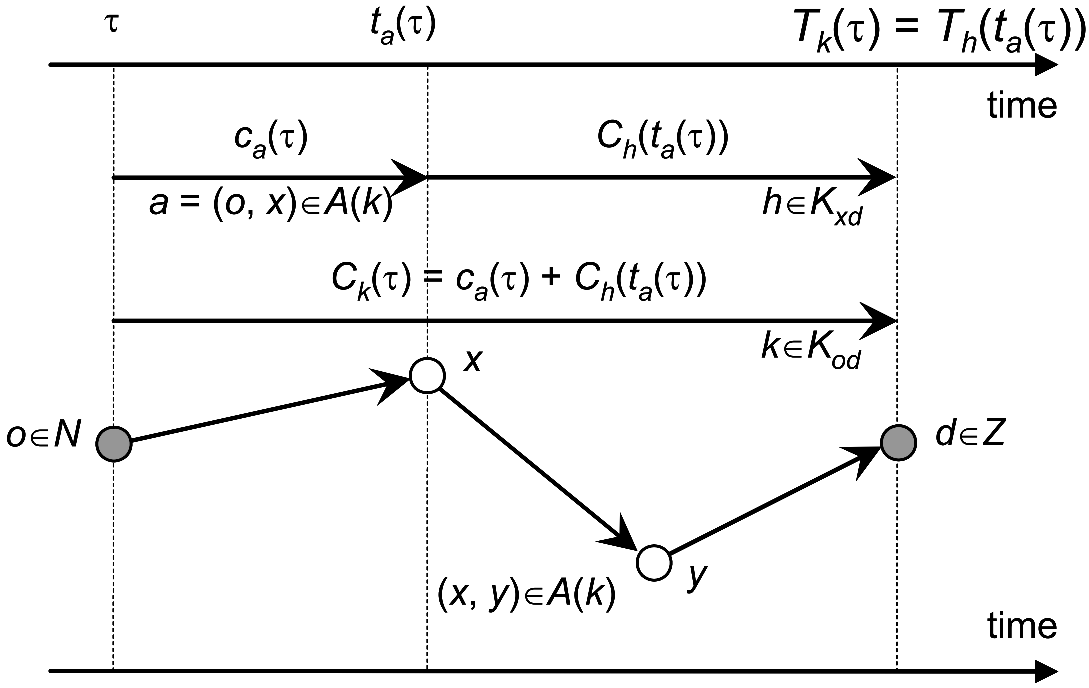

***Illustration 76**:* *Recursive expressions of path exit time, entrance time and cost*

The strict First In First Out (FIFO) rule holds when the following property is satisfied for each arc $a \in A$:

$t_a(\tau') > t_a(\tau)$, for any $\tau' > \tau$. (4)

The monotonicity expressed by (4) ensures that the temporal profiles of the arc exit times are invertible. Moreover, the FIFO rule applies also to the entrance times:

$t_{xy}^{-1}(\tau') > t_{xy}^{-1}(\tau)$, for any $\tau' > \tau$. (5)

Any arc $a \in A$ consists of a homogeneous channel with two bottlenecks located at the beginning and at the end. The flow states along the arc are determined on the basis of the *Simplified Theory of Kinematic Waves* (STKW), assuming the concave parabolic-trapezoidal fundamental diagram depicted in **Illustration 77**, expressing the vehicle flow $q_a(x,\tau)$ at a given section x of the arc and instant τ, as a function of the vehicle density $k_a(x,\tau)$ at the same section and instant. The arc is then characterized by:

L_a length of arc a;

Q_a capacity of the initial bottleneck and of the homogeneous channel associated to arc a, called in-capacity;

S_a capacity of the final bottleneck associated to arc a, simulating the average effect of capacity reductions at road intersections (i.e. due to the presence of traffic lights), called out-capacity; $S_a \leq Q_a$;

V_a maximum speed allowed on arc a, called free flow speed;

KJ_a maximum density on arc a called jam density;

W_a propagation speed of hypercritical flow states on arc a, called hypercritical kinematic wave speed.

Within this framework, for links the in-capacity corresponds to the physical mid-block capacity, whereas out-capacity reflects the bottleneck capacity imposed by the signal control or priority rules at the downstream junction. Exit connectors $(o, y) \in A: o \in Z, y \in N \setminus Z$ are arcs with infinite in-capacity, entry connectors $(x, d) \in A: x \in N \setminus Z, d \in Z\}$ are arcs with infinite out-capacity, while turns are represented by arcs having zero length and in-capacity equal to their out-capacity.

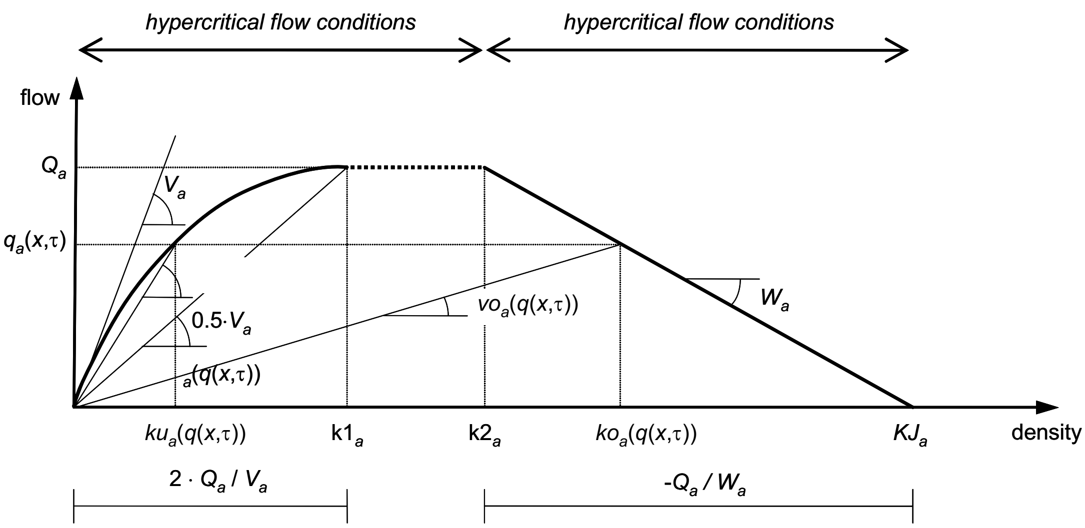

***Illustration 77**: The adopted parabolic-trapezoidal fundamental diagram, expressing the relation among vehicular flow, speed and density along a given arc*

With reference to **Illustration 77**, it is assumed that $k2_a \geq k1_a$, implying the following relation among the above parameters:

$$KJ_a \geq Q_a \cdot \left(\frac{2}{V_a} - \frac{1}{W_a} \right)$$

Based on the fundamental diagram, it is possible to identify two families of flow states:

► *hypocritical flow conditions,* corresponding to uncongested or slightly congested traffic; under these conditions, if vehicular density increases, the vehicular flow increases also;

- *hypercritical flow conditions,* corresponding to heavily congested traffic, where queues and "stop and go" phenomena occur; under these conditions, if vehicular density increases, the vehicular flow decreases.

Then, $ko_a(q)$ and $vo_a(q)$ express the density and the speed as functions of the flow in presence of hypercritical flow conditions, while $ku_a(q)$ and $vu_a(q)$ express the density and the speed as functions of the flow in presence of hypocritical flow conditions.

When modelling arcs with low speed limits, i.e. representing urban roads, it may be assumed that the vehicle speed under hypocritical flow conditions is constant and equal to the speed limit, until capacity is reached; in this case, the simpler trapezoidal fundamental diagram depicted in **Illustration 78** may be adopted, where, in order to guarantee $k2_a \geq k1_a$, the following relation must hold:

$$KJ_a \geq Q_a \cdot \left(\frac{1}{V_a} - \frac{1}{W_a} \right)$$

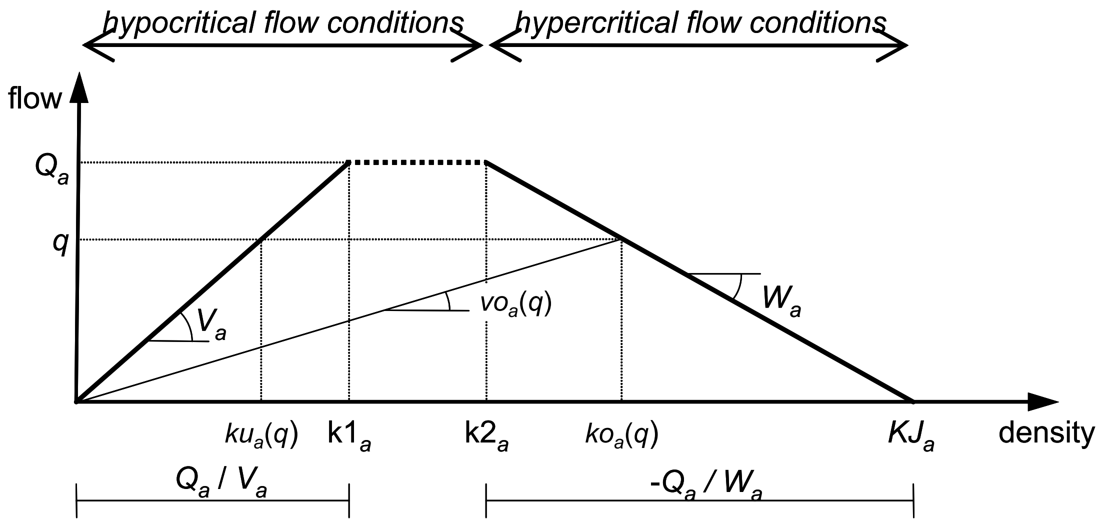

***Illustration 78**:* *The trapezoidal fundamental diagram suggested for urban links*

In order to implement the proposed models, the period of analysis $[0, \Theta]$ is divided into n time intervals identified by the sequence of instants $\tau = \{\tau 0, \ldots , \tau i, \ldots , \tau n\}$, with $\tau 0 = 0$, $\tau i < \tau j$ for any $0 \leq i < j \leq n$, and $\tau^n = \Theta$. For computational convenience, we introduce also an additional instant $\tau^{n+1} = \infty$.

In the following we approximate the temporal profile $g(\tau)$ of any variable through either a piecewise linear or a piecewise constant function, defined by the values $g^i = g\left(\tau^i\right)$ taken at each instant $\tau^i \in \tau$.. This way, any temporal profile $g(\tau)$ can be then represented numerically through the vector $g = (g0, \ldots , g^i , \ldots , g^n)$.

2.4.11.4 Network Performance Model

To represent the spillback phenomenon, we assume that each arc is characterized by two time-varying bottlenecks, one located at the beginning and the other one located at the end, called “entry capacity” and “exit capacity” respectively.

The entry capacity, bounded from above by the in-capacity, is meant to reproduce the effect of queues propagating backwards on the arc itself, which can reach the initial section and can thus induce spillback conditions on the upstream arcs. In this case the entry capacity is set to limit the current inflow at the value which keeps the number of vehicles on the arc equal to the storage capacity currently available. The latter is a function of the exit flow temporal profile, since the queue density along the arc changes dynamically in time and space accordingly with the STKW. Specifically, the space freed by vehicles exiting the arc at the head of the queue takes some time to become actually available at the tail of the queue, so that the jam density times the length is only the upper bound of the storage capacity, which can be reached only if the queue is not moving.

The exit capacity, bounded from above by the out-capacity, is meant to reproduce the effect of queue spillovers propagating backwards from the downstream arcs, which may generate hypercritical flow states on the arc itself. For given arc inflows, arc outflows and intersection priorities, which are here assumed proportional to the mid-block capacities, the exit capacities are obtained as a function of the entry capacities based on flow conservation at the node.

The network performance model is specified here as a circular chain of three models, namely the “exit flow and travel time model for time-varying capacities”, the “entry capacity model”, and the “exit capacity model”, which are solved iteratively. The three models, described separately in the following sections, are shown in context in **Illustration 79**.

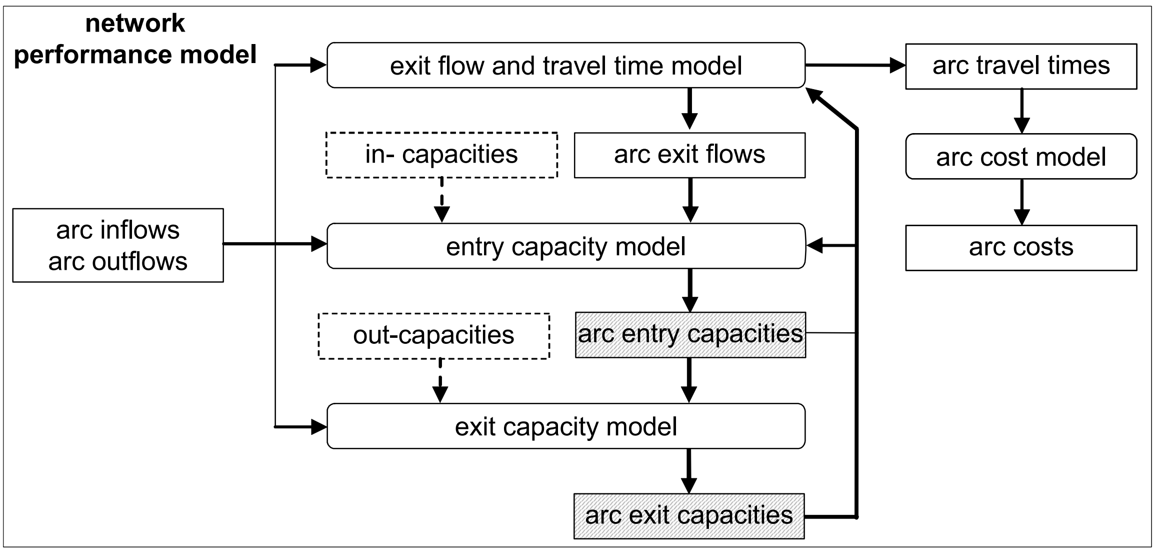

***Illustration 79**:* *Scheme of the fixed point formulation for the NPM*

Exit Flow and Travel Time Models for Time-varying Exit Capacity

Assuming the FIFO rule, i.e. no overtaking among vehicles can occur, we introduce in this section a link-based performance model for an arc with time-varying exit capacity, aimed at determining the "exit flow" temporal profile as a propagation of the inflow temporal profile to the arc final section, and then the corresponding travel time temporal profile.

Under the assumption that there is no capacity reduction at the end of arc $a \in A$, it is possible to express the *hypocritical exit time* $r_a(\tau)$, for a vehicle entering the arc at time τ, as a function of the previous portion of the inflow temporal profile that is the inflow $f_a(\sigma)$ at each instant $\sigma \le \tau$:

$$r_a(\tau) = r_a(f_a(\sigma) : \sigma \le \tau) . \quad (6)$$

Relation (6) will be specified in section *Hypocritical exit time model for a trapezoidal fundamental diagram* for the trapezoidal fundamental diagram (**Illustration 78**), and in section *Hypocritical exit time model for a parabolic fundamental diagram* for the parabolic fundamental diagram (**Illustration 77**).

When, instead, at the end of the arc there is a bottleneck with time-varying exit capacity $\psi_a(\sigma) \le S_a$ for each time σ, the problem of determining the overall exit time $t_a(\tau) \ge r_a(\tau)$ for a vehicle that enters the arc at time τ shall be addressed identifying firstly the *cumulative exit flow* temporal profile, whose value $E_a(\tau)$ at time τ is given by:

$$E_a(\tau) = min\left\{F_a\left(r_a^{-1}(\sigma)\right) + \psi_a(\tau) - \psi_a(\sigma) : \sigma \le \tau\right\} \quad (7)$$

where $\psi_a(\tau)$ denotes the *cumulative exit capacity* at time τ:

$$\Psi_a(\tau) = \int_{-\infty}^{\tau} \psi_a(\sigma) \cdot d\sigma \ , \quad (8)$$

i.e. between time points σ and τ, $\psi_a(\tau) - \psi_a(\sigma)$ vehicles can exit the arc.

The above expression (7) is based on the following specification of the FIFO rule, stating that the cumulative exit flow at the exit time $t_a(\tau)$ of a vehicle that enters the arc at τ is equal to the cumulative inflow at time τ, that is: $E_a(t_a(\tau)) = F_a(\tau)$.

Then, expression (7) can be explained as follows. If there is no queue at a given time τ, the travel time is equal to the hypocritical running time, so that, based on the FIFO rule (9), the cumulative exit flow is equal to the cumulative inflow at time $r_a^{-1}(\tau)$ when a vehicle that is leaving the arc at τ enters it. If a queue arises at time $\sigma < \tau$, from that instant until the queue vanishes the exit flow equals the exit capacity instead, and then, based on the FIFO rule, the cumulative exit flow $E_a(\tau)$ results from adding to the cumulative inflow at time $r_a^{-1}(\sigma)$ the integral of the exit capacity between σ

and τ, that is $\psi_a(\tau) - \psi_a(\sigma)$. Moreover, if there is no queue at time τ, the cumulative exit flow is the same as if the queue arises exactly at $\sigma = \tau$.

By definition, the exit flow $e_a(\tau)$ from arc a at time τ is:

$$e_a(\tau) = dE_a(r)/d\tau \qquad (10)$$

By construction, $e_a(\tau) \leq \psi_a(\tau)$ at any time τ and hypercritical exit flows occur whenever $e_a(\tau) = \psi_a(\tau)$.

Knowing the cumulative inflow and exit flow temporal profiles, the FIFO rule (9) yields an implicit expression for the arc exit time temporal profile. Once the cumulative exit flow temporal profile is known, the exit time temporal profile is calculated conventionally as:

$$t_a(\tau) = max\{r_a(\tau), min\{\sigma : E_a(\sigma) = F_a(\tau)\}\}. \qquad (11)$$

Illustration 80 depicts a graphical interpretation of equation (7), where the cumulative exit flow temporal profile $E_a(\tau)$ is the lower envelope of the following curves: a) the cumulative inflow $F_a(\tau)$ shifted forward in time by the hypocritical running time $r_a(\tau) - \tau$, thus yielding the temporal profile $F_a\left(r_a^{-1}(\tau)\right)$; this represents the rate at which vehicles entering the arc arrive at its end; b) for every time σ, the cumulative exit capacity temporal profile shifted vertically so that it goes through the point $\left(\sigma, F_a\left(r_a^{-1}(\sigma)\right)\right)$, this represents the rate at which vehicles can exit the arc following time σ. No queue is present when curve a) prevails. Queueing starts, when the cumulative exit flow curve falls below the time-shifted cumulative entry flow curve, i.e. more vehicles arrive at the final section of the arc than can exit; in the diagram, therefore, the queue arises at time σ' and vanishes at time σ''. In the same framework, the calculation of the exit time based on the cumulative inflow and exit flow temporal profiles is shown using thick arrows.

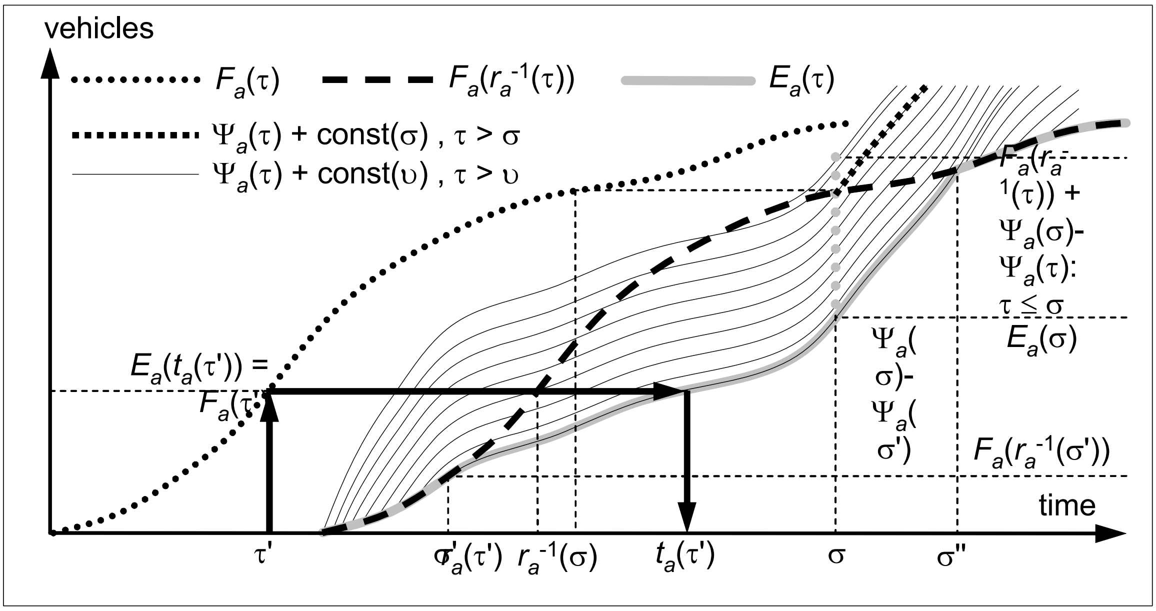

***Illustration 80**: Arc with time-varying capacity*

Hypocritical exit time model for a trapezoidal fundamental diagram

If the trapezoidal fundamental diagram is adopted to represent flow states on the arc, the hypocritical speed on the link is constant, and thus relation (6) is simply specified as:

$$r_a(\tau) = \tau + L_a / V_a . \qquad (12)$$

In this case, using (12) equation (7) can be made explicit as follows:

$$E_a(\tau) = min\{F_a(\sigma - L_{a/V_a}) + \psi_a(\tau) - \psi_a(\sigma) : \sigma \le \tau\}. \qquad (13)$$

Hypocritical exit time model for a parabolic fundamental diagram

If the parabolic fundamental diagram is adopted, the situation becomes more complicated because vehicles may travel at different speeds even at hypocritical densities. In the case where the arc inflow temporal profile is piecewise constant, the running link exit time can be determined at least approximately from the STKW. The general idea is to trace out the trajectory of a vehicle entering arc a at time τ, observing the different speeds it will encounter along the arc, and determining its exit time $t_a(\tau)$. Below we first explain the exact model. Since it can result in a large computational effort, we then replace it with a simpler model which averages traffic conditions and thus limits the number of different traffic situations encountered by any vehicle on the arc. Readers who would like to get a general feel of the model as a whole may just note the general idea and skip to the conclusion of this section.

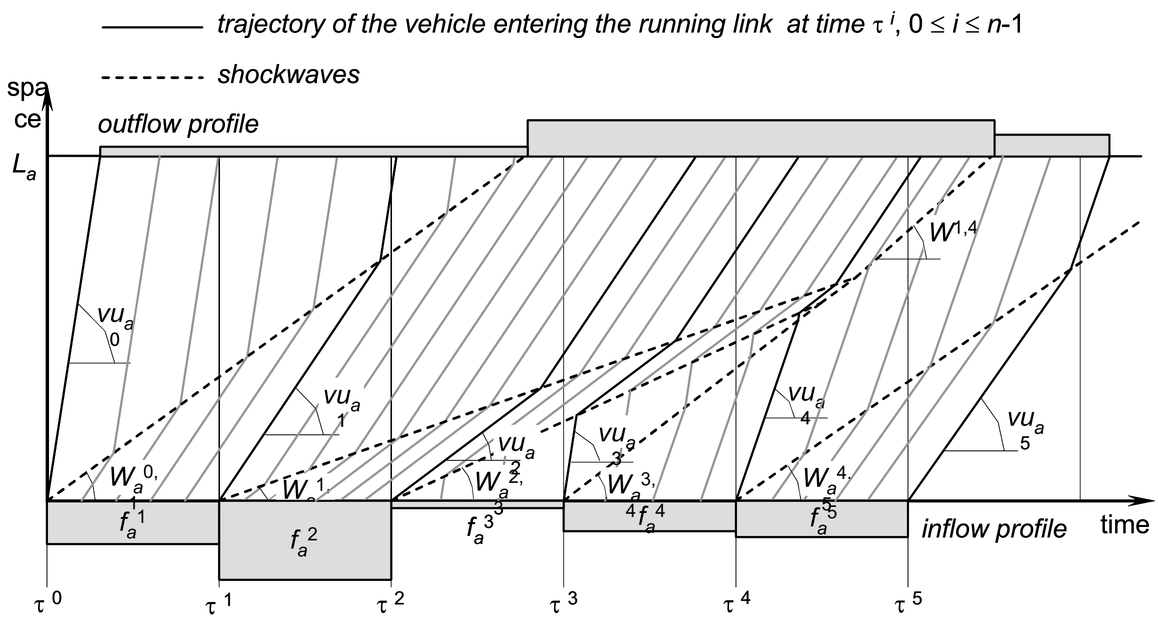

***Illustration 81**: Flow pattern given by the Simplified Theory of Kinematic Waves*

Based on the STKW, vehicles change their speeds instantaneously. As depicted in **Illustration 81**, when the inflow temporal profile is piecewise constant, vehicle trajectories are piecewise linear and the space-time plane comes out to be subdivided into flow regions characterized by homogeneous flow states and delimited by linear shock waves. The slope W_a^{ij} of the shockwave separating two hypocritical flow states $\Phi\left(f_a^i\right)$ and $\Phi\left(f_{a\,j}\right)$ is

$$W_a^{ij} = \frac{f_a^{\,j} - f_a^{\,i}}{ku_a(f_a^{\,j}) - ku_a(f_a^{\,i})} = vu_a(f_a^{\,i}) + vu_a(f_a^{\,j}) - V_a \ . \quad (14)$$

In theory, given a piece-wise constant inflow temporal profile, it is possible to determine the trajectory of a vehicle entering the arc at instant τ, and thus its hypocritical exit time $r_a(\tau)$. However, **Illustration 81** shows that it may be extremely cumbersome to determine these trajectories, in fact:

- many shockwaves may be active on the arc at the same time;
- shockwaves may be generated either at the initial section by flow discontinuities at times $\tau^i \;\; 0 \leq i \leq n\text{-}1$, or on any arc section at any time by shockwave intersections;
- a vehicle may cross many shockwaves while travelling on the arc, and all the crossing points have to be explicitly evaluated in order to determine its trajectory.

In order to overcome these difficulties, as depicted in **Illustration 82**, we assume that at each instant $r^i, 0 \leq i \leq n-1$, a fictitious shockwave is generated on the initial arc section separating the actual flow state $\phi\left(f_a^{\,i+1}\right)$ and the fictitious flow state corresponding to the average speed

$\lambda^i = L / \left(r_a{}^i - \tau^i \right)$ of the vehicle entered at instant τ^i. Fictitious shockwaves are very easy to deal with, in fact:

- they never meet each other, and thus are all generated on the running link initial section only at time $\tau^i, 0 \le i \le n-1$;
- each vehicle meets at the most the last generated fictitious shockwave, so that its trajectory is very easy to be determined.

Based on (14), the slope $W_a{}^i$ of the fictitious shockwave is:

$$W_a{}^i = \lambda^i + vu(f_a{}^{i+1}) - V_a \quad . \qquad (15)$$

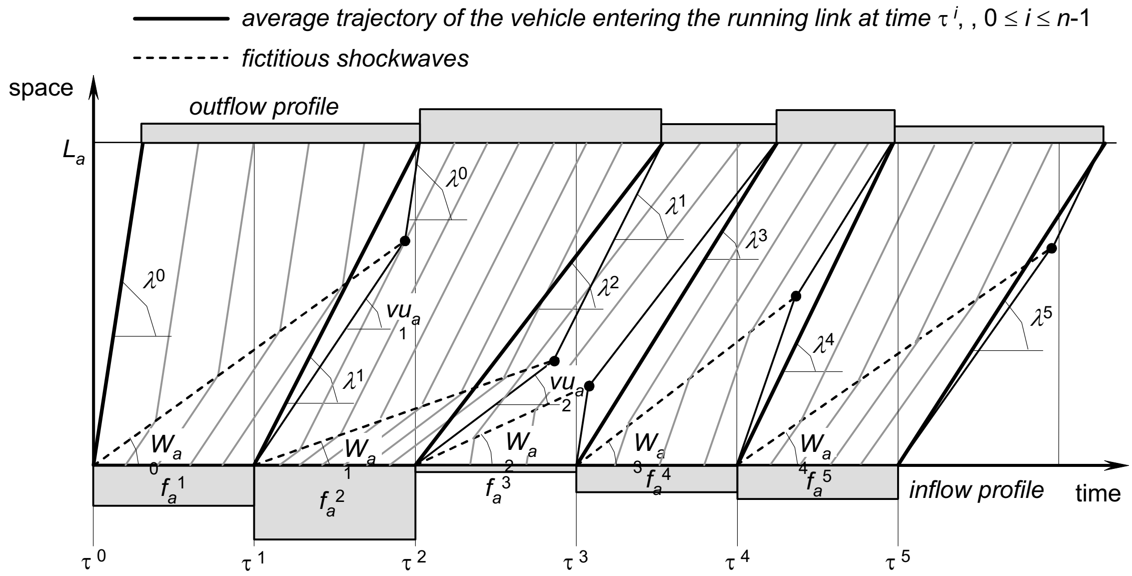

***Illustration 82**:* *Flow pattern given by the Averaged Kinematic Wave model.*

Note that the trajectory of a vehicle entering the running link at time $\tau \in \left(\tau^{i-1}, \tau^i \right]$ is directly influenced only by the average trajectory of the vehicle entered at time τ^{i-1}, which synthesizes the previous history of flow states.

The approximation introduced has little effect on the model efficacy. Moreover, it satisfies the FIFO rule, which is still ensured between the arc initial and final sections, while local violations that may occur within intermediate sections are of no interest.

Based on the above, the hypocritical running time $\tau_a{}^i = \tau_a(\tau)^i, 0 \le i \le n-1$, can be specified as follows:

a) if a vehicle entered at time τ^i does not meet the fictitious shockwave $W_a{}^{i-1}$ before the end of the arc, its hypocritical exit time is simply:

$$r_a{}^i = \tau^i + L_a / vu_a(f_a{}^i),$$

where f_a^i is the arc inflow during time interval $\left(\tau^{i-1}, \tau^i\right]$;

b) otherwise, its hypocritical exit time is determined on the basis of the two speeds it assumes before and after crossing the fictitious shockwave, that is:

$$r_a^{\ i} = \tau^i + \omega^i + \left(L_a - \omega^i \cdot vu(f_a^{\ i})\right) / \lambda^{i-1} ,$$

where ω^i is the vehicle travel time before it reaches the fictitious shockwave (**Illustration 83**):

$$\omega^i = (\tau^i - \tau^{i-1}) \cdot W_a^{\ i-1} / (vu(f_a^{\ i}) - W_a^{\ i-1}) \ .$$

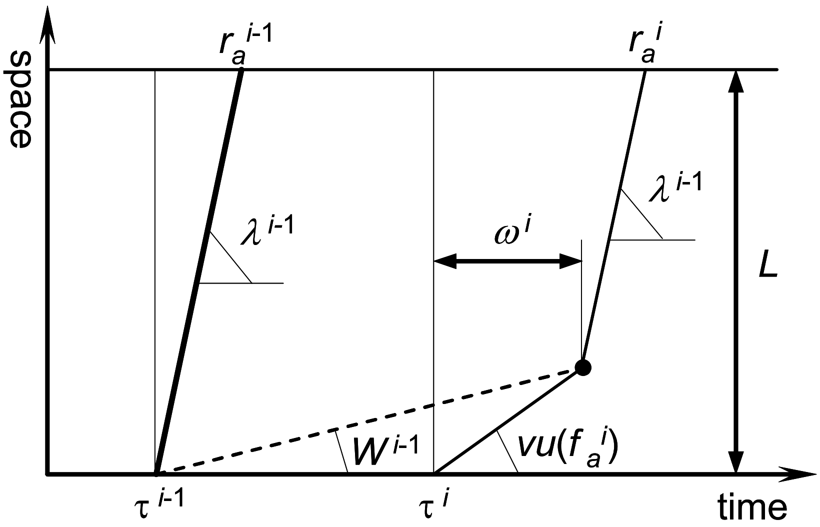

***Illustration 83**:* *Determination of the arc hypocritical exit time.*

Then, the hypocritical running time $r_a(\tau)$ specifying (6) is:

$$r_a(\tau) = r_a^{\ i} + \left(\tau - \tau^i\right) \bullet \left(r_a^{\ i+1} - r_a^{\ i}\right) / \left(\tau^{i+1} - \tau^i\right), \tau \in \left[\tau^i, \tau^{i+1}\right), 0 \le i \le n-1 \qquad (16).$$

Entry capacity model

In this section we propose a new approach to represent the effect on the entry capacity of queues that, generated on the arc final section by the exit capacity, reach the arc initial section, thus inducing spillback conditions. This part of the model is used only, if DUE is run with the spillback option activated. If the option is turned off, the storage capacity of an arc is assumed to be infinite, and the entry capacity of a link is never reduced below the in-capacity.

To help understand let us assume, for the moment, that the queue is uncompressible, i.e. only one hypercritical density exists. Then, the kinematic wave speed is infinitive – from either **Illustration 77** or **Illustration 78**, it is clear that $w_a = \infty$ when $KJ_a = k2_a$ – so that any hypercritical flow state occurring at the final section would back-propagate instantaneously.

This circumstance does not imply that the queue reaches the initial section instantaneously. There indeed, the exiting hypercritical flow state does not affect the entering hypocritical flow state until the arc has filled up completely, i.e. the cumulative number of vehicles that have entered the arc is smaller than the number of vehicles that have exited the arc plus the storage capacity. The latter in this case is constant in time and given by the arc length multiplied by the jam density. As soon as the queue exceeds the arc length, the entry capacity becomes equal to the exit capacity, i.e. all vehicles on the arc move as one rigid object.

Actually, hypercritical flow states may occur at different densities and their kinematic wave speed are not only quite lower than the vehicular free flow speed, implying that the delay affecting the backward translation in space from the final to the initial section of the flow states produced by the exit capacity is not negligible, but also somewhat different from each other, which generates a distortion in their forward translation in time. Notice that the fundamental diagrams adopted here, having both a linear hypercritical branch, are capable of representing the dominant delay effect but not the distortion effect, since all backward kinematic waves have the same slope.

The spillback effect on the entry capacity is here investigated by exploiting the analytical solution of the STKW.

The flow state occurring on an arc section is the result of the interaction among hypocritical flow states coming from upstream and hypercritical flow states coming from downstream. Specifically, on the initial section, the one flow state coming from upstream is the inflow, while the flow states coming from downstream are due to the exit capacity and can be determined by back-propagating the hypercritical portion of the cumulative exit flow temporal profile, thus yielding what we refer to as the “maximum cumulative inflow” temporal profile.

According to the Newell-Luke minimum principle, the flow state consistent with the spillback phenomenon occurring at the initial section is the one implying the lowest cumulative flow. Therefore, when the cumulative inflow equals or overcomes the maximum cumulative inflow, so that spillback actually occurs, the derivative of the latter temporal profile may be interpreted as an upper bound to the inflow. This permits to determine the proper value of the entry capacity that maintains the queue length equal to the arc length.

The instant $\upsilon_a(\tau)$ when the backward kinematic wave generated at time τ on the final section of arc $a \in A$ by the hypercritical exit flow $e_a(\tau) = \psi_a(\tau)$ would reach the initial section is given by:

$\upsilon_a(\tau) = \tau + L_a / w_a(e_a(\tau))$. (17)

By definition the points in time and space constituting the straight line trajectory produced by a kinematic wave are characterized by a same flow state. Moreover, **Illustration 84** shows that the number of vehicles encountered by the hypercritical wave relative to the exit flow q for any infinitesimal space ds travelled in the opposite direction is equal to the time interval ds $\bullet \left[1/v_a(q)+1/w_a(q)\right]$ multiplied by that flow. Therefore, integrating along the arc from the final to the initial section, we obtain the cumulative flow $H_a(\tau)$ that would be observed at time $\upsilon_a(\tau)$ in the initial section as:

$$H_a(\tau) = E_a(\tau)+e_a(\tau)\bullet L_a \bullet \left[1/v_a(e_a(\tau))+1/w_a(e_a(\tau))\right]. \qquad (18)$$

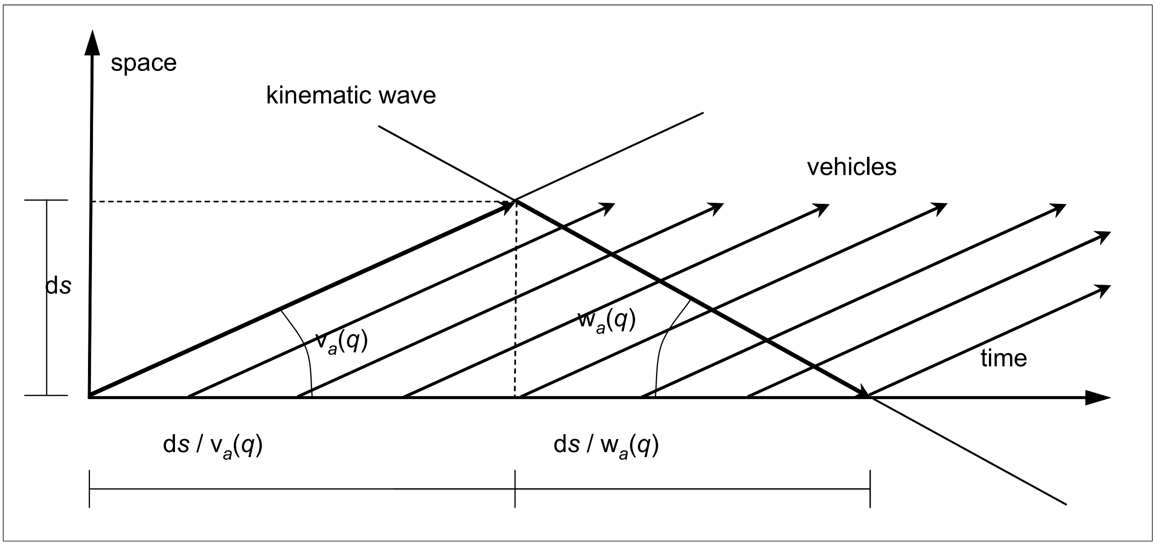

***Illustration 84**: Trajectories of a hypercritical kinematic wave and of the intersecting vehicles*

In the fundamental diagrams adopted here, the hypercritical branch is linear and therefore $\upsilon_a(\tau)$ is invertible. Indeed, since $w_a(q)=w_a$, based on (17) the time when $\upsilon_a(\sigma)=\tau$ is $\sigma=\tau-L_a/w_a$. Moreover, since $q/v_a(q)=KJ_a-q/w_a$, based on (18) we have: $H_a(\tau)=E_a(\tau)+L_a\bullet KJ_a$

Therefore, the *maximum cumulative inflow* $G_a(\tau)$ that could have entered the arc at time τ consistently with the exit flow pattern is given by:

$$G_a(\tau)=\begin{cases} E_a(\tau-L_a/w_a)+L_a\cdot KJ_a, & \text{if } e_a(\tau-L_a/w_a)=\psi_a(\tau-L_a/w_a); \\ \infty, & \text{otherwise.}\end{cases}$$

(19)

If at time τ the cumulative inflow $F_a(\tau)$ is equal or higher than the maximum cumulative inflow $G_a(\tau)$, so that spillback occurs at that instant, then the *entry capacity* $\mu_a(\tau)$ is given by the derivative $dG_a(\tau)/d\tau$ of the latter; otherwise, it is equal to the in-capacity Q_a. Differentiating $G_a(\tau)$ yields: $dG_a(\tau)/d\tau$ = $e_a(\tau-L_a/w_a)$; then, since $e_a(\tau-L_a/w_a)$ = $\psi_a(\tau-L_a/w_a)$, we get:

$$\mu_a(\tau)=\begin{cases} \psi_a(\tau-L_a/w_a), & \text{if } G_a(\tau)\le F_a(\tau); \\ Q_a, & \text{otherwise.}\end{cases} \qquad (20)$$

Illustration 85 shows how, based on equation (19), the maximum cumulative inflow temporal profile can be obtained graphically through a rigid translation (thick arrows) of the cumulative exit flow temporal profile for L_a / w_a in time and for $L_a \bullet KJ_a$ in value. Moreover, it points out that, when $G_a(\tau)$ is greater than $F_a(\tau)$, the queue is shorter than L_a and $\mu_a(\tau) = Q_a$, otherwise spillback occurs and $\mu_a(\tau) = \psi_a(\tau - L_a / w_a)$.

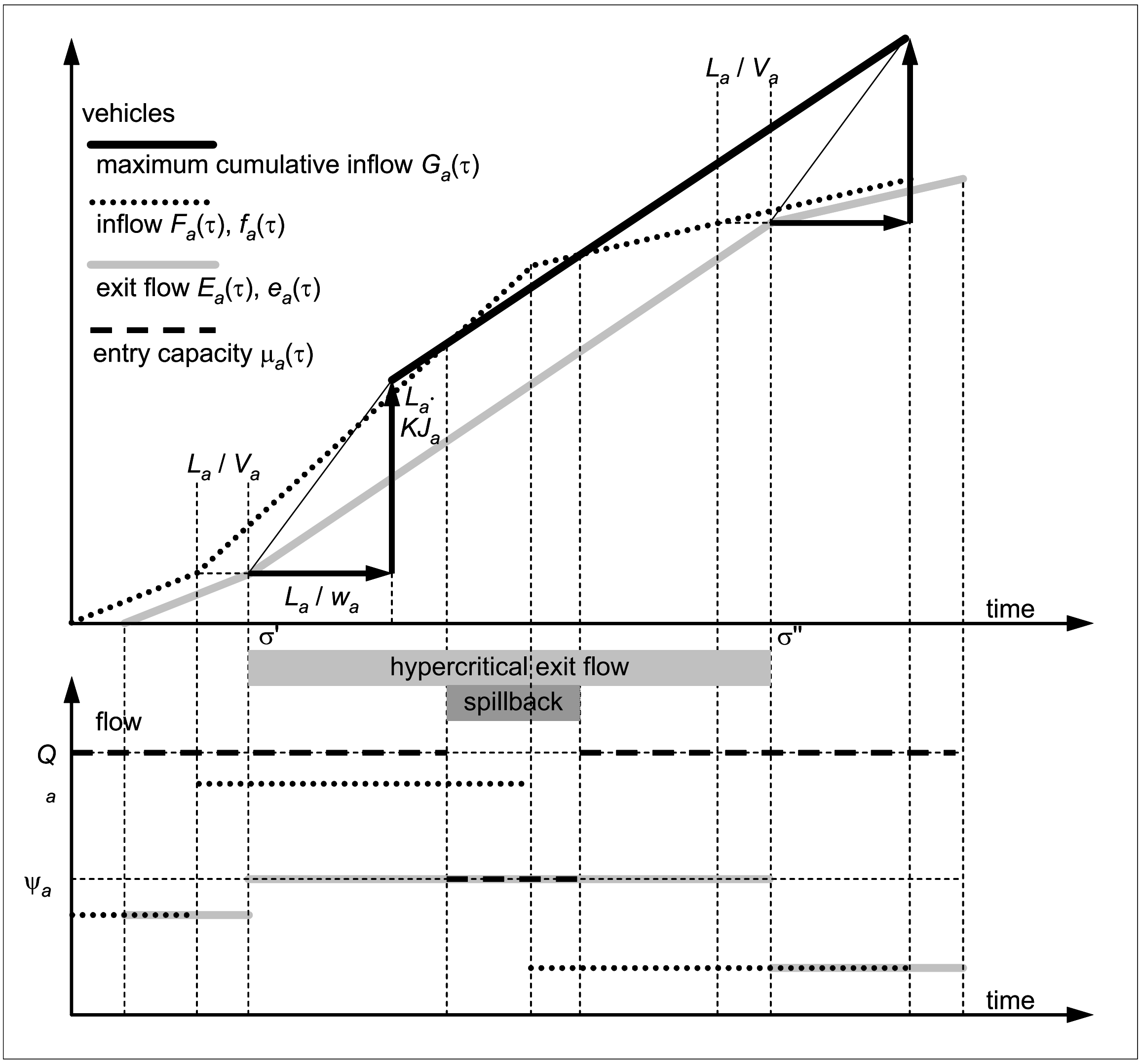

***Illustration 85**: Graphical determination of the entry capacity temporal profile in the case of triangular fundamental diagram, piecewise constant inflow, and constant exit capacity*

Exit Capacity Model

In this section we present a model to determine, for a given node, the exit capacities of the upstream arcs, on the basis of the entry capacities of the downstream arcs and of the local maneuver flows. In this model, only two typologies of nodes are allowed: “mergings” and “diversions”; in this case, in fact, this model can be expressed in terms of arc outflows and inflows. This is not a limitation: for VISUM networks a graph node will either connect one incoming link arc to several outgoing turn / connector arcs (in which case it is a diversion) or it will connect several incoming turn / connector arcs to one outgoing link arc (in which case it is a merging).

When considering a *merging* $x \in N$, that is an intersection with a singleton forward star, the problem is to split the entry capacity $\mu_b(\tau)$ of the arc $b = FS(x)$ available at time τ among the arcs belonging to its backward star, whose outflows compete to get through the intersection. In principle, we assume that the available capacity is partitioned proportionally to the out-capacity S_a of each arc $a \in BS(x)$. But this way it may happen that on some arc a the outflow $\mu_a(\tau)$ is lower than the share of entry capacity assigned to it, so that only a lesser portion of the latter is actually exploited. The rest of the entry capacity shall then be partitioned among the other arcs. Moreover, when no spillback phenomenon is active, the exit capacity $\psi_a(\tau)$ shall be set equal to the out-capacity S_a .

When considering a *diversion* $x \in N$, that is an intersection with a singleton backward star, the problem is to determine at time τ the most severe reduction to the outflow from the arc $a = BS(x)$ among those produced by the entry capacities of the arcs belonging to its forward star. Again, when no arc is spilling back, the exit capacity shall be set equal to the out-capacity. When only one arc $b \in FS(x)$ is spilling back, that is $f_b(\tau) \geq \mu_b(\tau)$, in order to ensure capacity conservation at the node while satisfying the FIFO rule applied to the vehicles exiting from arc a, the exit capacity $\psi_a(\tau)$ scaled by the share of vehicles turning on arc b is set equal to b's entry capacity: $\Psi_a(\tau) \bullet f_b(\tau) / u_a(\tau) = \mu_b(\tau)$. When more than one arc $b \in FS(x)$ is spilling back, the exit capacity is the most penalizing among the above values. On this basis, we have:

$$\psi_a(\tau) = min\{S_a ; \mu_b(\tau) \bullet u_a(\tau) / f_b(\tau) : b \in FS(x), f_b(\tau) \geq \mu_b(\tau)\} . \qquad (21)$$

Note that, in contrast with the models presented in the previous two sections, this model is spatially non-separable, because the exit capacities of all the arcs belonging to the backward star of a same node are determined jointly, and temporally separable, because all relations refer to a same instant.

It is assumed that vehicles do not occupy the intersection if they cannot cross it due to the presence of a queue on their successive arc, but wait until the necessary space becomes available. Indeed, this model is not capable of addressing the deterioration of performances due to a misusage of the intersection capacity.

Arc Cost Model

The cost for vehicles entering arc a at time τ is given by:

$$c_a(\tau) = \eta \bullet (t_a(\tau) - \tau) + m_a(\tau) \qquad (22)$$

where $m_a(\tau)$ is the *monetary cost,* while η is the *value of time.*

2.4.11.5 Network Loading

In this section we develop a formulation for the dynamic Network Loading Map with implicit path enumeration in the case of deterministic route choice model. To this end, we will firstly define and address the continuous dynamic shortest path problem, which lies at the heart of the route choice model.

Continuous Dynamic Shortest Path Problem

Contrary to the static case, in the dynamic context the shortest path problem involves explicitly the time dimension, since the costs of the arcs constituting a path are to be evaluated at different instants, consistently with the travel times experienced along the path, as induced by the recursive equation (3); then we will address the problem of finding the minimum cost $w_o{}^d(\tau)$ from each node $o \in N$ to a given destination $d \in Z$ for users departing at time τ:

$$w_o{}^d(\tau) = \min\{C_k(\tau) : k \in K_{od}\} \qquad (23)$$

It can be proved that the following dynamic version of the *Bellman relation* for each node $o \in N$ (**Illustration 86**) is equivalent to problem (23):

$$w_o{}^d(\tau) = \min\{\, C_{ox}(\tau) + w_x{}^d(t_{ox}(\tau)) \,\} : x \in FS(o) \,\}. \qquad (24)$$

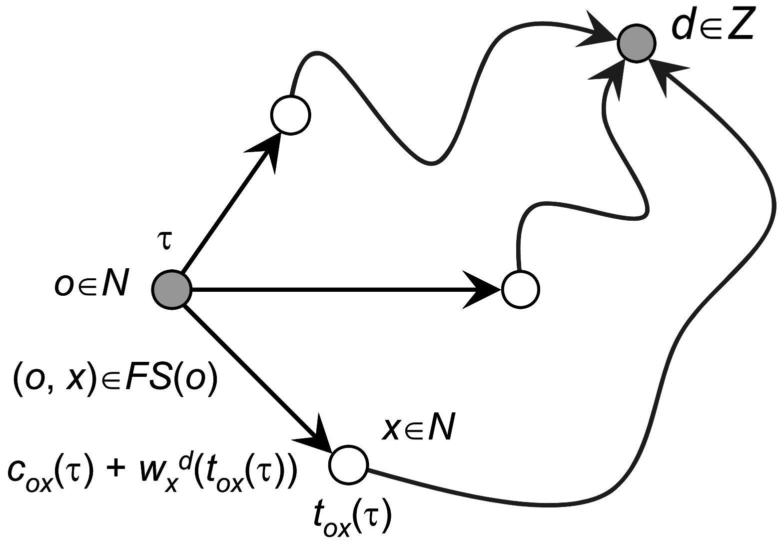

***Illustration 86**: Dynamic version of the Bellman relation.*

The set of Bellman relations (24) can be solved using a dynamic programming approach described below.

Path Choice and Network Flow Propagation Models

Under the assumption that users are perfectly informed rational decision-makers, the resulting behaviour is such that only shortest paths are utilized. The *deterministic* route choice model for users that travel between the origin $o \in N$ and the destination $d \in Z$ departing at time τ, can then be formulated through the following extension of the dynamic case of Wardrop's first principle:

- if path $k \in K_{od}$ is used, i.e., its choice probability $P_k(\tau)$ is positive, then its cost $C_k(\tau)$ is equal to the minimum cost $w_o{}^d(\tau)$ to travel between o and d departing at time τ;
- vice versa, if path k is unused, i.e., its choice probability is zero, then its cost may not be smaller than the minimum cost.

This can be formally expressed as follows:

$$P_k(\tau) \bullet \left[C_k(\tau) - w_o{}^d(\tau) \right] = 0. \qquad (25)$$

Moreover, the choice probabilities must be non-negative and amount to 1.

We now develop a formulation based on implicit path enumeration for the route choice model and for the corresponding network flow propagation model adopting the temporal-layer approach, where the temporal perspective is the exit time from the current node.

If the shortest paths from $o \in N$ to $d \in Z$ for users departing at time τ involve more than one arc exiting from an intermediate node x, then the *conditional probabilities* of these arcs at time τ for users directed to d could depend, in general, on the sub-path utilized from each o to x. Because of the additive nature of arc costs, we assume instead that the arc conditional probabilities at each node are equal for all users directed to the same destination regardless of the sub-path so far utilized.

Under this assumption, the choice probability $P_k(\tau)$ of a path $k \in K_{od}$ from $o \in N$ to $d \in Z$ for users departing at time τ is equal to the product of the conditional probabilities of its arcs $A(k)$, each of them referring to the time when these users enter the arc when travelling along the path. The choice probability of k can be then retrieved through the following recursive expression:

$$P_k(\tau) = p_{ox}{}^d(\tau) \bullet P_h(t_{ox}(\tau)), \qquad (26)$$

where (o, x) is the first arc of k and $h \in K_{xd}$ is the rest of path k.

On this basis it can be proved that the deterministic specification of the path choice probabilities defined by the dynamic Wardrop condition is equivalent to the specification of the arc conditional probabilities defined by the following system:

$$p_{ox}{}^d(\tau) \bullet \left[C_{ox}(\tau) + w_x{}^d(t_{ox}(\tau)) - w_o{}^d(\tau) \right] = 0 , \qquad (27)$$

$$\Sigma_{(o,x) \in FS(o)}\, p_{ox}{}^d(\tau) = 1 , \qquad (28)$$

$$p_{ox}{}^d(\tau) \geq 0 . \qquad (29)$$

Equation (27) states that users exiting at time τ from node $o \in N$ and directed to the destination $d \in Z$ may choose among the forward star $FS(o)$ only an arc (o, x) for which the cost $C_{ox}(\tau)$ plus the minimum cost $w_x{}^d(t_{ox}(\tau))$ to reach the destination once entered x at time $t_{ox}(\tau)$ is equal to the minimum cost $w_o{}^d(\tau)$.

The flow $f_{ox}{}^d(\tau)$ of vehicles directed to destination $d \in Z$ that enter the arc $(o, x) \in A$ at time τ is given by the arc conditional probability $p_{ox}{}^d(\tau)$ multiplied by the flow exiting from node o at time τ. The latter is given, in turn, by the sum of the outflow $u_{yo}{}^d(\tau)$ from each arc $(y, o) \in BS(o)$ entering o, and of the demand flow $D_{od}(\tau)$ from o to d. Then, we have:

$$f_{ox}{}^d(\tau) = p_{ox}{}^d(\tau) \bullet \left[D_{od}(\tau) + \Sigma_{(y,o) \in BS(o)} u_{yo}{}^d(\tau) \right]. \quad (30)$$

Applying the FIFO and vehicle conservation rules, the outflow from y at time τ can be expressed in terms of the inflow at o at time $t_{yo}{}^{-1}(\tau)$:

$$u_{yo}{}^d(\tau) = f_{yo}{}^d(t)_{yo}{}^{-1}(\tau) / \left[dt_{yo}(\tau) / d\tau \right], \quad (31)$$

where the weight $dt_{yo}(\tau)/d\tau$ stems from the fact that travel times vary over time, so that users exit from y at a certain rate and, in general, enter in o at a different rate, which is higher than the previous one, if the arc travel time is decreasing, and lower, otherwise.

The total inflow and outflow of arc $(o, x) \in A$ at time τ are then:

$$f_{ox}(\tau) = \Sigma_{d \in Z} f_{ox}{}^d(\tau); u_{ox}(\tau) = \Sigma_{d \in Z} u_{ox}{}^d(\tau) \quad (32)$$

2.4.11.6 Dynamic User Equilibrium Model

All the components of DTA have been introduced. Here we formulate the user equilibrium, where no user can reduce his perceived travel cost by unilaterally changing path, as a fixed point problem in the temporal profiles of the arc inflows and outflows.

The Deterministic Case

The formulation of DTA with implicit path enumeration yields the model depicted in **Illustration 87**.

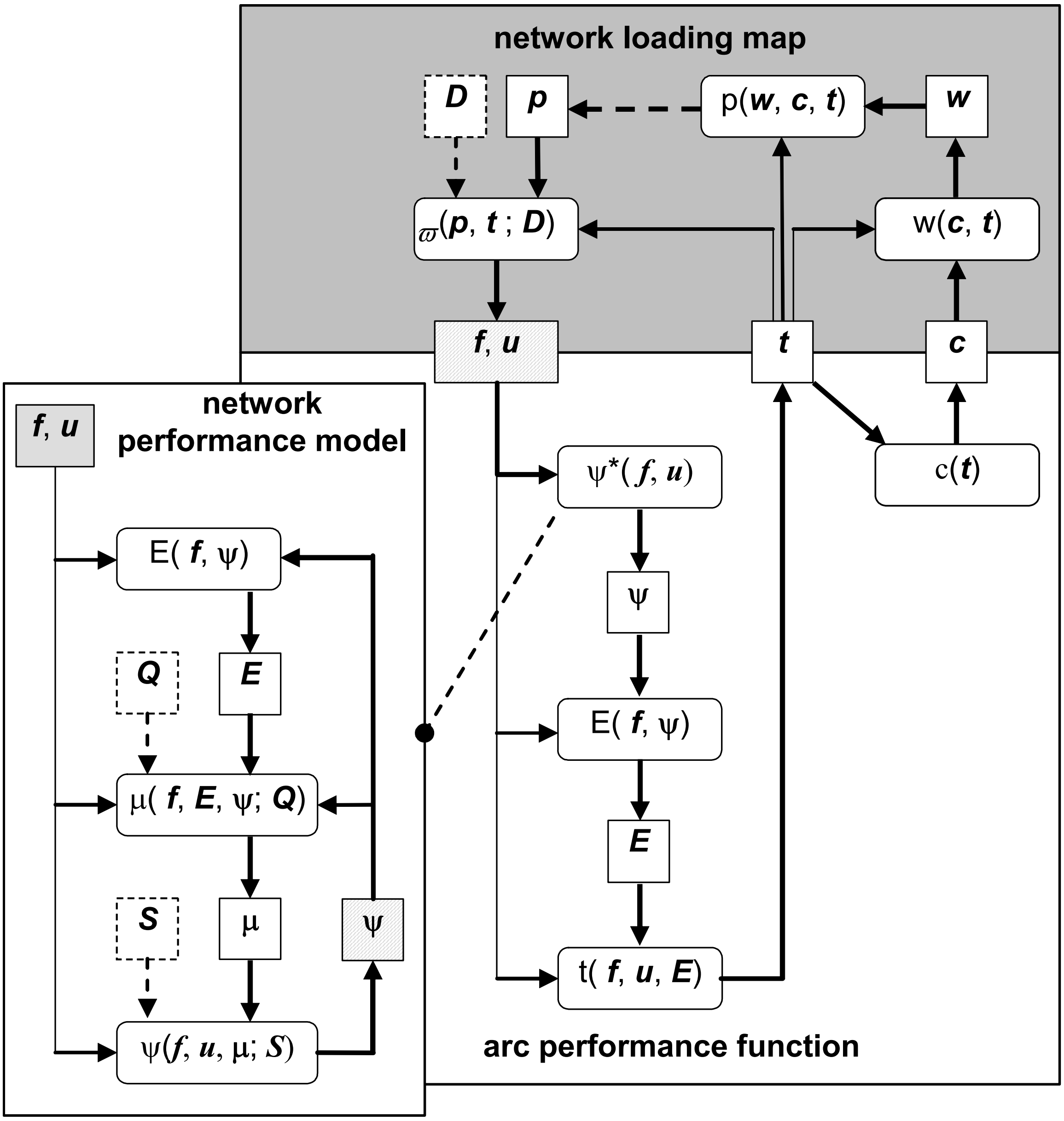

***Illustration 87**: Variables and models of the fixed point formulations for the NPM (left hand side) and for the DTA with spillback (right hand side)*

In analogy with the static case, the Network Loading Map (NLM) is a functional relation yielding, for given demand flows **D**, an arc flow pattern **f** consistent with the arc performances **t**, and **c**, through the deterministic route choice model p(w(**c**, **t**), **t**, **c**), and the network flow propagation model ω(**p**, **t**; **D**). The NLM has been formulated with implicit path enumeration by introducing the node minimum cost **w**, and the arc conditional probabilities **p**. Note that the dashed arrows indicate any solution of the corresponding choice map. In turn, the arc performance model yields the arc exit time pattern **t,** and the arc cost pattern **c**, consistent with the arc inflows **f** and arc outflows **u**.

The deterministic user equilibrium is formally given by the system of the NLM and of the arc performance model.

The Probit Case

In the Probit route choice model, which is based on the random utility theory, the arc costs perceived by users are not known with certainty and are thus regarded as independent random variables. We extend the Probit model to the dynamic case assuming that the arc cost $\hat{c}_a(\tau)$ of arc $a \in A$ perceived by users at time τ is equal to the sum of the arc cost $c_a(\tau)$ yielded by the arc performance model and of a time-varying random error, whose value at time τ is distributed as a normal variable. Its variance is assumed proportional, through a constant coefficient $\xi > 0$, to a time-varying cost term $\chi_a(\tau)$ > 0 independent of congestion.

The arc flow pattern resulting from the evaluation of the Probit NLM for given arc performances is obtained through the well-known Montecarlo method as follows:

a) Get a sample of H perceived arc cost patterns:

$\hat{c}_a{}^h(\tau) = c_a(\tau) + \psi_a{}^h \bullet [\zeta \bullet \chi_a(\tau)]^{0.5}$, in compact form: $\hat{c}^h = \hat{c}(c;\chi)$, (33)

where each $\psi_a{}^h$ is extracted from a standard normal variable N[0,1] and $h = 1, \ldots , H$..

b) For each perceived arc cost pattern of the sample, determine through the deterministic NLM a consistent arc inflow pattern.

c) Calculate the average of the resulting deterministic arc inflow patterns, thus obtaining an undistorted estimation of the Probit arc inflow pattern.

Note that, based on (33), the same outcome $\psi_a{}^h$ of the standard normal variable is used to perturb the whole temporal profile $\hat{c}_a{}^h(\tau)$. This is consistent with the behaviour of users, who perceive the arc cost temporal profile as a whole. On the contrary, the travel times that underlie the network flow propagation, are considered as constant throughout the simulation.

2.4.11.7 Example

In order to investigate the behavior of the proposed model and to show the effect of spillback on path choice, we analyze a simple example which presents intuitive solutions. It is located in folder EXAMPLES\DUE of your VISUM installation as BRAESS_WITHOUT_SPILLBACK.VER and BRAESS_WITH_SPILLBACK.VER.

We consider the Braess network depicted in **Illustration 88**; links have the characteristics reported in the corresponding table, and are all modelled with a parabolic-trapezoidal fundamental diagram. All link out-capacities are set equal to the corresponding in-capacities. The turn capacities are: $Q_{AC} = Q_{AE} = Q_{ED}$ = 2000 veh/h; $Q_{BD} = Q_{CF} = Q_{DF}$ = 1000 veh/h.

Link	L_a [km]	Q_a [veh/h]	V_a [km/h]	W_a [km/h]	1 / K_a [m]
A	0.4	2000	50	15	7.0
B	0.6	2000	50	15	7.0
C	0.6	2000	50	15	7.0
D	0.4	2000	50	15	7.0
E	0.4	2000	50	15	7.0
F	0.1	4000	50	15	3.5

The period of analysis is constituted by 100 intervals of 1 minute. We assume a constant demand for the first 33 minutes of simulation from node 1 to node 5 equal to D_{15} = 2300 veh/h.

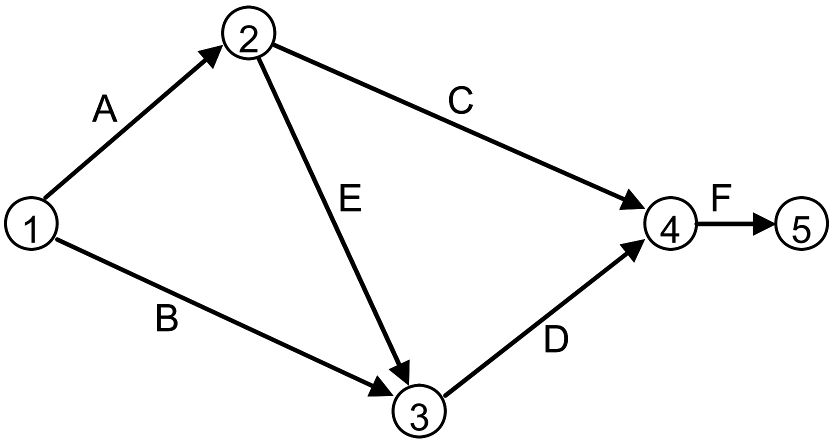

***Illustration 88**: The example network*

The outputs of two assignment runs, one without and the other with spillback congestion, are presented in **Illustration 89**. Without spillback, the congestion is evenly located only on turns CF and DF (which can be gathered observing turn travel times), so that on all the paths between node 1 and node 5 the queue is about equal, and path A-E-D-F has fewer users since it is clearly not convenient. With spillback, however, the queue propagates from turn CF to arc C and up to arc A, and from turn DF to arc D and up to arcs B and E. Moreover, the spillback effect is greater on arc B than on arc E because of the different capacities of turn ED and turn BD. Then, after an initial growth, the travel time on arc D remains constant, since congestion is propagated upward, while the travel time on arc B grows faster than the travel time on arc E, so that path A-E-D now becomes competitive, as it implies a longer route but a lower travel time, and flow on arc E increases from around 150 veh/h to 670 veh/h approximately.

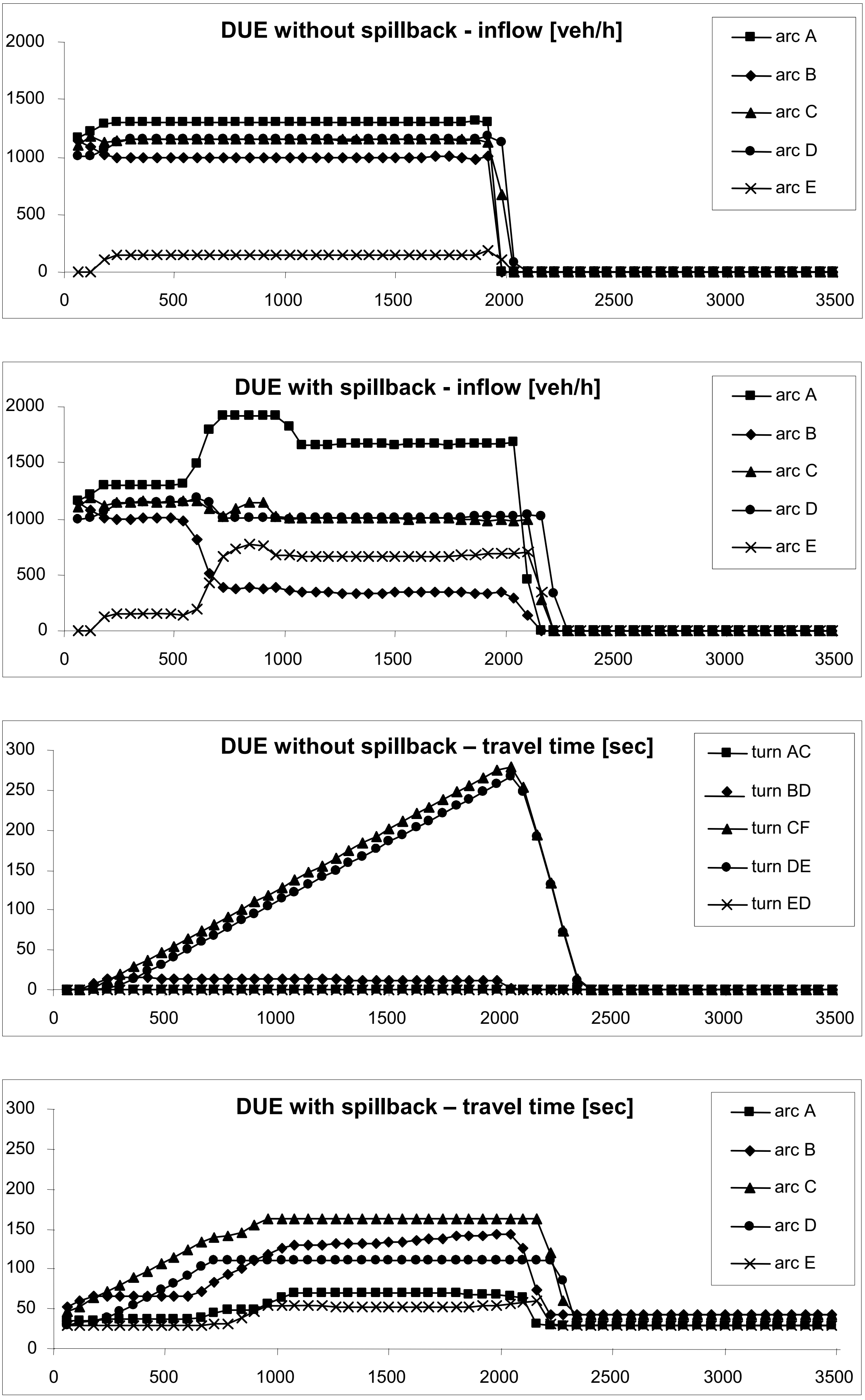

***Illustration 89**:* *Results of WDDTA without and with spillback*

References

Bellei G., Gentile G., Meschini L., Papola N.: A demand model with departure time choice for within-day dynamic traffic assignment. European Journal of Operational Research, in press, available online at: http://www.sciencedirect.com. (2005)

Bellei G., Gentile G., Papola N.: A within-day dynamic traffic assignment model for urban road networks. Transportation Research B 39, 1-29. (2005)

Gentile G., Meschini L., Papola N.: Macroscopic arc performance models with capacity constraints for within-day dynamic traffic assignment. Transportation Research B 39, 319-338. (2005)

Gentile G., Meschini L.: Fast heuristics for the continuous dynamic shortest path problem in traffic assignment. Submitted to the AIRO Winter 2005 special issue of the European Journal of Operational Research on Network Flows. (2006)

Gentile G., Meschini L., Papola N.: Spillback congestion in dynamic traffic assignment: a macroscopic flow model with time-varying bottlenecks. Accepted for publication in Transportation Research B. (2006)

2.4.12 Dynamic Stochastic Assignment

The dynamic stochastic assignment differs from all other PrT assignment procedures as a result of the explicit modelling of the time required to complete trips in the network. For dynamic stochastic assignment, capacity has to be set as an hourly value, not regarding the length of the time interval the demand is available for.

In contrast, all trips are completed in the case of static assignment procedures with no indication of the time required, capacities have to be specified according to the length of the time interval demand data is available for, and the volumes of all trips and the resultant impedances are superimposed upon each other at the individual network objects. Road-users subsequently have only to choose from a number of different routes for each journey. The departure time is irrelevant.

In the case of the dynamic assignment on the other hand, an assignment period T (e.g. 24 hours) is specified and divided up into time slices T_i of equal length (e.g. 15 minutes). Only the search for (alternative) routes for each journey is made with no reference to a specific time. As in the case of the static stochastic assignment, several shortest path searches are completed with network impedances that vary at random. All other operations explicitly include a time dimension.

From the entire demand and its temporal distribution curve, the portion with the desired departure time is determined for each time slice within this time interval. On the supply side, there are pairs to choose consisting of route and departure time slice, which, using PuT assignment terminology, are also called connections. The impedance of a connection is composed of its network impedance and the difference between the actual and desired