ALTERNATIVE METHODS FOR DEVELOPING HORIZONTAL/SPATIAL QUEUEING MODELS IN ROAD TRAFFIC

Hugo Pietrantonio, D.Sc. (Corresponding Author)
Lecturer, Traffic Engineering
Department of Transportation Engineering
Escola Politécnica – University of São Paulo, Brazil
Edifício de Engenharia Civil, Cidade Universitária
São Paulo/SP, CEP 05508-900, Brazil
Email: hpietr@usp.br
Phone: +55-11-3091-5492; Fax: +55-11-3091-5716

Abstract: This paper discusses current models proposed for predicting queue evolution in road traffic using horizontal (or spatial) queueing models, contrasting them to vertical (or stacked) queueing models and proposing several revisions that embodies shock waves and consistent balancing conditions (in a form that renders several relations in closed-form expressions). The results are derived from those obtained from the standard horizontal queueing model (conventional macroscopic traffic flow model) and shown to reproduce these calculations (at least for first order models, i.e. LWR kinematic model and its shock waves) or allow the use of simple formulas that can make them applicable with usual data (simple data if accepting some loss of precision, but smaller than in current formulas). The proposal is also extended to the analysis with multiple periods, prediction of queueing delays and the measurement of demand using queue measurements. Finally, the concluding section highlights proposals that can be adopted for revision of current formulas based on vertical queueing models (as those of the HCM), identifying features that deserve further study.

Key words: road traffic, queueing model, horizontal queues.
ALTERNATIVE METHODS FOR DEVELOPING HORIZONTAL/SPATIAL
QUEUEING MODELS IN ROAD TRAFFIC

Hugo Pietrantonio, D.Sc. (Corresponding Author)
Lecturer, Traffic Engineering
Department of Transportation Engineering
Escola Politécnica – University of São Paulo, Brazil

1. INTRODUCTION

After more than 75 years from the birth of modern traffic flow theory, several models used in the theory
and practice of traffic engineering are based on vertical (or stacked) queueing models, even if this feature is
not properly set in their presentation. As a major example, recent versions of the HCM (TRB, 2000, 2011) that
surely can be received as a major contribution to the development of Traffic Engineering practice worldwide,
adopts such model forms in several expressions (mostly those related to the estimation of delay at intersections)
without acknowledging its roots in vertical queueing models. By its turn, horizontal (or spatial) queueing
models are not distinguished and the application of its concepts and methods are only being introduced in a very
slow pace, if done, and mixed with inconsistent views.

Among the new concepts and methods in recent versions of the HCM, the promotion of the multimodal
evaluation of road infrastructure is especially welcome. One should also foresee the importance of stressing the
user point of view in evaluating the quality of the traffic operation, even if yet mixing the old view related to
Level of Service measures (LoS as supply related evaluation) and the new view based on Quality of Service
measures (QoS representing the user evaluation as LoS scores) in a clumsy way. And also the importance of
introducing the stop rate (and the spatial stop rate) as a fundamental performance measure (a variable that surely
has its place in methods for evaluating environmental effects as seen in recent proposals for prediction of traffic
noise emission but similarly applicable to the prediction of other effects as air pollution and fuel consumption).
Several other minor enhancements to traffic analysis methods were proposed or similarly developed, perhaps in
a partial way.

In general, each proposed innovation introduced in a version of the HCM opens a strong stream of
research intended to improve its formulation in a definitive way. However, several traditional features keep
being repeated even if there is a clear need for revision. This paper is related to a small addition to the basic
concepts that is taken as a point deserving better treatment.

Dealing with concepts related to vertical and horizontal queueing models, in the more recent versions of
the HCM the concept of flow rate is distinguished from the related concept of demand flow rate (TRB, 2011,
Ch.4; previously introduced in TRB, 2000, Ch.7) for road traffic. The distinction signs the clear recognition
that traffic demand flow can be unmet in some situations, generating queues of unserved users, and that traffic flow can be fed not only by traffic demand flow but by pre-existing queues as well. It relates to current developments of Transportation Planning methods, in which conventional static traffic assignment procedures of the past were being challenged by the requirement of dynamic traffic assignment procedures (not to mention the interaction with traffic control). It also relates to the introduction of the recent versions of analysis methods with over-demand and multiple periods, in which the dynamics linking subsequent periods recognizes the differences of the traffic demand flow (the inflow) and the traffic flow (the outflow) and its effects on queue evolution along periods (by some sort of balancing formula). But, as argued ahead, some concepts remain unclear even in the recent versions of the HCM (only traffic flow rate is defined; demand flow rate is just alluded; measurement of traffic volume and flow rate is detailed; measurement of demand flow rate is not discussed) and its treatment unsatisfactory (even if introducing formulas that accept over-demand and proposing analysis options that consider multiple periods and the effect of pre-existing queues).

This paper tries to make the point that this small addition displays the tip of an iceberg that merits a better picture: the need for revising most usual formulas for queueing (and also for delay) by replacing conventional queueing models by the more realistic horizontal queueing models. As argued in the following, the vertical queueing models deliver results that are inconsistent with the spatial evolution of queue lengths and to any approach that recognizes this feature (not just the horizontal queueing models here presented but also macroscopic or microscopic traffic simulation models that explicitly represent the space requirements of vehicles in road traffic). Otherwise, horizontal queueing model can usually deliver viable replacements to conventional queueing and delay formulas that are consistent with the spatial evolution of queues as observed in the field or evaluated by the other modeling approaches that represent vehicle space requirements in road traffic. These alternative formulas will be derived and discussed as well as compared to the similar (but more complex) analysis of a simple capacity bottleneck with the first order macroscopic traffic model.

The initial motivation for the revision of vertical queueing models comes from field observation, where any careful analyst is able to conclude that conventional formulas are imprecise. The analysis of the queue dynamic with a first order macroscopic traffic model provides a better consistent alternative (as shown in Hurdle, Son, 2001), even if not so simple, and is a textbook problem for continuous (uninterrupted) and discontinuous (interrupted) flow conditions (see May, 1990, Ch.11). However, this analysis can provide insights to other options for improving most operational models of traffic in intersections (currently based on the vertical queueing models) that were not fully explored. One of the oldest references to a revision with the motivation of introducing more realistic formulas can be found in a NCHRP report of the 70’s (Curry, Anderson, 1972, App.B), where a horizontal queueing model is related to the quoted analysis of a first order macroscopic traffic flow model for continuous (uninterrupted) flow conditions. The same macroscopic
modeling approach was also previously applied to the analysis of discontinuous (interrupted) flow conditions at intersections (e.g. since the seminal paper of Lighthill, Whitham, of 1955, as recovered in Rorbech, 1968 and Stephanopoulos, Michalopoulos, 1979) but the results were cumbersome to apply and could not be translated into simplified analytical formulas in the spirit of that traditionally used in the HCM (even more when long queues are generated by sustained over-demand and simultaneous pairs of stopping and starting waves from interruptions are superposed along the queue length). The best effort with this aim seems to be that generated from the related research at the TTI about traffic signals on ramp interchanges (Messer, Bonneson, 1997, App.D), producing a clear understanding of the involved relations but not able to translate them into simple formulas as those proposed in this paper (the collaboration of Rouphail, Akçelik, 1992, can also be noted).

More recently, horizontal queueing models based on a first order macroscopic traffic model for arterials were also proposed as the work horse in estimating queues and delay due to work zones (as can be exemplified by Ramezani, Benekohal, 2012) and arterial travel time distributions from mobile data (a research stream that seems to be initiated by Hofleitner, Herring, Bayen, 2010).

The next section highlights the differences between vertical and horizontal queueing models and discusses the kind of inconsistencies that appears in approaches based on the vertical version (as the HCM), using the comparison to the corresponding analysis based on the kinematic waves generated by capacity bottlenecks or interrupted flow conditions at intersections, including the extension of formulas to the analysis methods with multiple periods, advocated with much more emphasis for over-demand conditions in the HCM 2010. Then alternative formulas for predicting the average queue at intersection approaches in the spirit of the HCM but based on the horizontal queueing models are presented and it is shown that formulas based on the horizontal version are more precise and consistent in the prediction of the dynamic of queue evolution, without increasing data needs or analysis tasks in a significant way (as demonstrated by illustrating the practical application of the proposed revision to a real case study on a heavily saturated signalized intersection). Finally, in the concluding section, the paper suggests points that can be considered for the revision of HCM’s formulas with horizontal queueing models and identifies features judged to be promising for further study.

Before proceeding, a special note should be made on matters of style. First, an idiosyncratic form is used for numbering formulas. As several expressions are related to the same variables, the same number is kept for all formulas that refer to the same variable even if appearing several times along the text, but then distinguished by a companying letter in each occurrence. Also, in some expressions, several formulas are chained for the same variables and the number is appended by a list of letters, one for each version of the formula. As example, consider that variable X is related to formula (1) and appears once as $X = a$ and ahead as $X = b/c = d,e$. In the first appearance, the formula for X is referred as (1a). In the second appearance, the formulas for X are referred as (1b,c), meaning that formula (1b) is $X = b/c$ and formula (1c) is $X = d,e$. After
a while, this simple referencing scheme should be felt as clear and natural (but, more important, conveniently
short and yet precise). Second, a minor idiosyncrasy also appears in the way some superscripts or subscripts are
used. This is mainly related to time variables (but can also occurs for space variables) for which a superscript T
means that the variable (perhaps as average) refers to any instant along the period T (i.e. for any $t \in T$) while a
subscript T means that the variable refers to the (instantaneous) value at the end of the period T (i.e. just for
$t = T$). Finally, as a mean for providing a comparative view, each figure groups several pieces of information
distinguished by letters, displayed in a logical order. Unfortunately, this recourse also means that reference to
different parts of each figure are spread along the discussion and will require some backward and forward
reading effort for checking the relation of text and figure (an inconvenient cost for a convenient overall picture).
2. CLARIFYING DIFFERENCES BETWEEN VERTICAL AND HORIZONTAL QUEUEING MODELS

In road traffic, the horizontal (or spatial) dimension of queues can not be overlooked. It is very common, at least in more trafficked networks, that queues of vehicle can take some kilometers or miles along major roads during peak periods (merging and splitting by forming ramifications commanded by the structure of the road network and its traffic circulation). This observation means that simple book-keeping identities (as presented in ch.1 of Newell, 1971, or Gross et alii, 2008, or Gautham, 2011) that calculate queueing evolution by balancing arrivals and departures can not be directly used for traffic queues (at least not without further thought and adjustment).

The error made can be related to the usual definition of traffic flow as a point measure (i.e. as the flow of traffic measured when passing at a static section of the road) as exemplified in Figure 1. Traffic flow rate, quoting the corresponding definition in the HCM 2010 (TRB, 2011, pg.4-2), is (the hourly rate of) vehicles that “pass over a given point or section of a lane or roadway during a given time interval” (of less than 1h, usually 15min), where the inessential features of the definition were kept in the parenthesis (the same definition, if expressed in number of vehicles instead of a rate is referred to as both traffic volume or traffic count). Traffic demand flow is described as the number of vehicle occupants or drivers (usually expressed as the number of transport or traffic units or as a hourly rate) “who desire to use a given system element during a specific time period” (typically 1h or 15min) and “… that would exist without the presence of a bottleneck”, again keeping the inessential features of the definition in the parenthesis or outside the quotation marks.

The terminology is not well set here. Traffic flow as referenced in the HCM (as rate or volume) can be more clearly referred to as served traffic demand (that part of the traffic demand that could flow through the observation section of the road element). Traffic demand flow as referenced in the HCM (as rate or volume too) is not clearly explained. Surely it is referring to a traffic demand for flow (in a wider sense, traffic has other demands as parking for vehicles and crossing for pedestrians). However, the expression “desire to use a given system element” is too loose as reference to the qualification of flow demand. Someone that desires to use the element can kept its plan and remain in the queue of the capacity bottleneck but, alternatively, can also change the path (or period) of travel due the knowledge of the bottleneck congestion. As referred here, only the first part of the unserved demand is kept in the definition (due to the intention to develop queueing models) and this traffic demand could be more clearly referred to as manifest traffic demand (the other parts would be referred to as deviated demand, retimed demand and so on, depending on the way it changed its manifestation).

The concept of manifest demand is usual in some other areas, meaning that the demand kept its choice of travel mode, path and period, even if not immediately served. Ahead, further qualification will be made by
distinguishing the local manifest demand (that passed all other capacity bottlenecks but the one at the current element) from some adjusted manifest demand (that can embody the demand that was previously retained by some other bottlenecks). The concepts of inflow and outflow (or feeding flow and leaving flow) are usually equivalent to traffic demand flow and traffic flow as local manifest and served demand, also represented in the dynamic assignment output of (modern) Transport Planning models. Otherwise, the concept of travel desire (as volume or rate, in transport or traffic units) from static assignment output of Transport Planning models is a fully adjusted manifest demand (with no retention due to capacity bottlenecks even if considering the increase in travel time generated from congestion effects, taking travel time, route, mode and O/D as given choices).

2.1 Vertical and Horizontal Models of Queueing During the Peak Period

The concept of traffic flow rate is traditional and its measure can be made in any section A as the outflow from the link can be evaluated as the flow through A in a given period T as $q_A^T = \frac{N_{AT}}{T}$, where $N_{AT}$ is the traffic flow volume passing through A during period T and $q_A^T$ is the corresponding average traffic flow rate. As can be seen from Figure 1a, considering a link a=(B;A) with length $L$ and $N_f$ lanes, a queue originating from the node A (downstream) and accumulating $n_t$ vehicles at time t (all lanes) reaches the average length $z_t = \frac{n_t}{N_f} \ell_v$, where $\ell_v = c_v + d_v$ is the length occupied by each vehicle of the queue in a traffic lane, summing the vehicle length $c_v$ and the distance $d_v$ to the vehicle ahead (in the HCM 2010 version there is an explicit recommendation to use $\ell_v = \ell_{LV} \cdot p_{LV} + \ell_{HV} \cdot p_{HV}$, with $p_{LV}, p_{HV}$ being the proportion of light and heavy vehicles in the traffic flow and $\ell_{LV} = 25ft \approx 7.6m, \ell_{HV} = 45ft \approx 13.7m$, reasonable values for stopped queues; however, in more general terms, it should weight vehicle spacings for different flow conditions for queueing, i.e. at least stopped queues and free discharging queues, as discussed ahead).

The traffic demand flow rate would be measured at any section that can represent the traffic that is going to use link a=(B,A) after entering the link at B but before passing through A or being constrained by any bottleneck effect generated from A as if (then) manifesting as demand flow rate at A. Based on the context of Figure 1a, despite not being clearly stated in the HCM, most views would accept that the concept of traffic demand flow rate can be measured in any section before the queue and then translated in time so as to be referred to period T at section A. If traffic flow rate and demand flow rate are constant during period T, the (unconstrained) demand flow rate will be constant $Q_A^T = Q_{wa}^T$ and the queue length will grow linearly (at a
constant propagation speed). The assumption of constancy for $Q_A^T$ solves the problem of translating its value in time (to be more discussed ahead) but the length and movement of the queue can not be similarly overlooked.

**Peak Period: T**

![Diagram of Peak Period]

**Queue Evolution**

$$Q_A^T > q_s^T = q_A^T$$

$$z_t = \frac{n_t \ell_v}{N_r}$$

**a. Queue Formation in the Peak Period**

**Next Period: Tn**

**Queue Evolution**

**Time Translation**

*Case I: $Q_{A_n}^{T_n} < q_n^T = q_n^T$*  

$$t_{t_n} = t_p$$

$$n_{AT} = K_n^{T_n} z_{AT}$$

$$z_{AT} (n_{AT})$$

*Case II: $q_n^T = q_n^T < q_{nA}^{T_n}$*  

$$t_{t_n} = t_q + t_m$$

$$n_{AM} = K_n^{T_n} z_{AM}$$

$$z_{AM} (n_{AM})$$

**b. Queue Dissipation in the Next Period**

Figure 1. Basic Variables for the Horizontal (Spatial) Queueing Models
These conceptual and practical distinctions can clearly be understood by first deriving the equations from the vertical queueing model for a simple capacity bottleneck at the end of link (B,A) under over-demand conditions (i.e. inflow $Q_A^T > C_A^T$ and outflow $q_A^T \leq C_A^T$) with a peak period of duration $T$. The estimated queue at instant $t$ from the vertical queueing model is denoted by $\tilde{n}_1$ (the apparent queue evaluated by balancing arrival and departure flows, to be distinguished from the real queue $n_1$ to be estimated with the horizontal queueing model in the next step). Then, initially ignoring pre-existing queues at the beginning of the analysis period, the vertical queueing model estimate for the queue generated by the capacity bottleneck at A is

$$\tilde{n}_A = (Q_A^T - q_A^T)t$$ (1)

that can be translated to a (vertical queueing model measure for the) queue length

$$Z_{At} = \frac{\tilde{n}_A}{N_v} \ell_v$$ (2)

where, as before, $\ell_v$ is the space occupied in a lane by each vehicle of the queue and $N_v$ is the number of lanes in the link (B,A), as conventionally used in HCM formulas for signalized or unsignalized intersections. A very simple outflow model, for short $q_A^T = \min\{Q_A^T, C_A^T\}$ (better as $q_A^T = \min\{Q_A^T, C_A^T\}$ if $\tilde{n}_A = 0$ and otherwise $q_A^T = C_A^T$, or $q_A^T = C_{sA}^T$ if distinguishing a saturated capacity, for $\tilde{n}_A > 0$), is also usual in the HCM tradition.

From these formulas it is clear that the queue generated by the capacity bottleneck at A is evaluated as the imbalance between manifest traffic demand flow and served traffic flow as would be measured at A. Of course, with over-demand at A, the manifest traffic demand flow would never be measured unless the capacity bottleneck would be absent (perhaps by eliminating the constraints that limits its capacity). However, it can be measured if the traffic flow rates that accrue to A can be observed before reaching the queue (or as explained ahead). But then, by considering the physical extent of the queue, one is forced to admit that a vehicle join the queue some time before the instant when it would pass through the section A in the absence of the queue (and counted in $Q_A^T$). So being, formulas (1) and (2) are inconsistent and not valid for field observations.

The simplest version of the horizontal queueing model estimate is obtained by recognizing that the correction to the vertical queueing model estimate is the addition of the number of vehicles that join the queue because it is horizontal but would not be in the queue if it were vertical. This correction is the number of vehicles that would be moving with normal conditions along the (real) queue length $z_{At}$ to A if it were not horizontal, that is $\delta n_{At} = Q_A^T \delta t$ adding the arrivals at the end of the queue during $\delta t = \frac{Z_{At}}{V_u^T}$, where $V_u^T$ is the...
normal (unconstrained by queueing) speed before reaching the capacity bottleneck on link \( a=(B,A) \) (as if flowing in the absence of the queue). Then, the correction is \( n_{At} = \tilde{n}_{At} + \delta n_{At} = \tilde{n}_{At} + \frac{Q_{A}^{T} \cdot n_{At} \cdot \ell_{v}}{V_{u}^{T} \cdot N_{f}} \) or the

**horizontal queueing model** estimate is

\[
\begin{align*}
n_{At} & \approx \frac{\tilde{n}_{At}}{1 - \frac{Q_{A}^{T} \cdot \ell_{v}}{V_{u}^{T} \cdot N_{f}}} = \frac{\tilde{n}_{At}}{1 - \frac{Q_{A}^{T}}{V_{u}^{T}} / N_{f} / \ell_{v}}
\end{align*}
\]

where \( f_{Qm}^{T} = \frac{1}{f_{Qm}} \) with \( f_{Qm}^{T} \equiv 1 - \frac{Q_{A}^{T}}{V_{u}^{T}} / N_{f} / \ell_{v} \) (a correction that is also valid for the queue evolution \( \Delta n_{A}^{T}, \Delta \tilde{n}_{A}^{T} \) if the involved factors are constant in \( T \)). Note that the traffic conditions \( Q_{A}^{T}, V_{u}^{T} \) correspond to the arrival at the end of the queue during the interval \( \delta t \) (as if \( \tilde{n}_{A}^{T} \) is the pre-existent queue before the last interval \( \delta t \) of \( T \), no matter how it was built-in but estimated with a vertical queueing model in \( T \), e.g. using a fluid flow/diffusion approximation or a more elaborated stochastic queueing model, meaning that only the horizontal correction is missing).

This formula was shortly recognized (e.g. in 4) as being consistent to the similar result gathered by analyzing the simple capacity bottleneck with the first order macroscopic traffic model by tracking its shock wave during queue formation, when (and where) the normal flow \( q_{u}^{T} = Q_{A}^{T} \) outside the queue (with speed and density \( V_{u}^{T}, K_{u}^{T} = \frac{Q_{A}^{T}}{V_{u}^{T}} \)) meets the congested flow \( q_{n}^{T} = q_{A}^{T} \) inside the queue (with speed and density \( V_{n}^{T}, K_{n}^{T} = \frac{q_{n}^{T}}{V_{n}^{T}} \)). The propagation speed of shock waves in the first order macroscopic traffic model is generically given by \( w = \pm \frac{\Delta q}{\Delta K} \) (positive if forward), that corresponds to the condition of traffic flow conservation at the queue frontier when moving at the shock speed \( w \) (if preferred, the shock speed formula as positive, meaning forward waves by convention, can take care of the sign of \( w \), then meaning the reverse backward direction with a negative value). For the queue formation, the speed of the congestion wave is

\[
w_{c}^{T} = -\frac{Q_{A}^{T} - q_{A}^{T}}{K_{u}^{T} - K_{n}^{T}} \] (moving backward as \( Q_{A}^{T} > q_{A}^{T} \) and \( K_{u}^{T} < K_{n}^{T} \)) that produce, for instant \( t \in [0; T] \) in the period \( T = [0; T] \), a queue length of size \( z_{At} = w_{c}^{T} \cdot t \) (backward of \( A \)) with \( n_{At} = z_{At} \cdot K_{n}^{T} \) vehicles, delivering
The shock speeds reveal measured at static section the queue is moving with speed for compact queues, corrections are around 10% but more than 50% can be observed for other situations.

\[ K_n^T = \frac{N_f}{\ell_v} \]

reproducing the previous result if \( K_n^T = Q_\Lambda^T / V_u^T \). Note that (3) implies that

\[ K_n^T = \frac{n_f}{z_T} = \frac{N_f}{\ell_v} \]

and the condition translates into the implicit supposition that the average spacing occupied by vehicles in a lane of the queue is \( \bar{e}_{n_f}^T = \ell_v \). As the value of \( \ell_v \) is taken by assumption, the condition could always be satisfied if the value of \( \bar{e}_{n_f} \) is known (what is not true unless the traffic operating conditions inside the queue is known). Otherwise, (3) is an approximation for (4). As example, the saturation flow of a normal traffic lane usually reaches 1800v/h/ln at 40 to 60km/h, meaning a spacing of 22 to 33m (of course, an upper limit to \( \bar{e}_{n_f} \) for \( K_n^T \) at 45 to 30v/km/ln and showing that a compact queue (conventionally supposed as \( \ell_v \) of 6 to 8m or \( K_n^T \) of 167 to 125v/km/ln) occurs only for stopped queues (as those generated by red times at signals) or for hard bottlenecks (that command low flows and speeds in their queues). Values for the correction factor \( n/\bar{n} \) against \( Q_\Lambda /N_f \) and \( V_\Lambda \) for \( \ell_v = 8m;15m;25m \) are shown in Table A1 (ahead).

For compact queues, corrections are around 10% but more than 50% can be observed for other situations.

The same formula has another revealing interpretation: if the back or front of queue is moving with speed \( w \), queue balancing can not be calculated using static flow measures (May, 1990, pg.323). If the front of the queue is moving with speed \( w_f \) and flowing \( q_f = K_f . V_f \) or \( q_{nf} = K_{nf} . V_{nf} \) (on each side of wave as measured at static sections), the wave speed is such that the moving section sees a relative flow

\[ q_{nf} = K_f \left( \dot{V}_f + w_f \right) = K_{nf} \left( \dot{V}_{nf} + w_f \right). \]

Similarly for the back of queue if moving with speed \( w_b \) and flowing \( q_b = K_b . V_b \) or \( q_{nb} = K_{nb} . V_{nb} \) but seeing a relative flow \( \dot{q}_{nb} = K_b \left( \dot{V}_b + w_b \right) = K_{nb} \left( \dot{V}_{nb} + w_b \right). \)

The shock speeds reveal the spatial evolution of the front and the back of the queue and deliver the correct queue balancing relation as \( n_{Af} = \left( q_{Af} - \dot{q}_{Af}^T \right) t \). In the simple capacity bottleneck, \( w_f = 0 \), \( q_f = q_{Af}^T \) and

\[ w_b = -w_c^T \]

\[ q_b = Q_A^T \]

with \( q_{nf} = q_{nb} = q_n = q_A^T \), \( w_c^T = -Q_{A_f}^T - q_{A_f}^T, K_{u}^T = Q_{A_f}^T / V_{u}^T, K_{c}^T = Q_{A_f}^T / V_{c}^T, \]

\[ \dot{Q}_A^T = K_{u}^T \left( V_{u}^T + w_{c}^T \right) = K_{c}^T \left( V_{c}^T + w_{c}^T \right), \]

\[ \dot{Q}_A^T = Q_A^T, \]

delivering

\[ n_{Af} = \left( K_{u}^T \left( V_{u}^T - \frac{Q_{A_f}^T - q_{A_f}^T}{K_{u}^T - K_{c}^T} \right) - q_A^T \right) t = \left( Q_{A_f}^T - q_{A_f}^T - K_{u}^T \right) t = \left( Q_{A_f}^T - q_{A_f}^T \right) t = \frac{Q_{A_f}^T - q_{A_f}^T}{1 - K_{u}^T / K_{c}^T} \]
formulas (4a,b). This alternative view shows another path for building a horizontal queueing model. Even if wave speeds are not clearly used in the previous approach (4a,b,c), the implied values can be obtained by imposing $n_{AT} = K_n^T \cdot w_c^T \cdot T$ (after obtaining $n_{AT}$ and using the corresponding value of $K_n^T$), if needed.

2.2 Distinct Cases of Queue Dissipation During the Next Period

The basic model can also be analyzed after the over-demand period (during the next period of duration $T_n$). There are two possible solutions for the peaking problem in the next period: a decrease of demand in the back of queue (Case I) or an increase of flow in the front of queue (Case II). For the simple capacity bottleneck problem, the analysis with the first order macroscopic traffic model is again a convenient benchmark, given the queue length $z_{AT}$ at the end of peak period $T$ and checking the effect of a change in traffic flow demand to $Q^{TN}_A$ and of traffic flow discharge to $q^{TN}_A$ (that can predict Case I or Case II). The general patterns of queue evolution are displayed in Figure 1b (mixed cases are also possible but not shown).

Before proceeding, the detailed notation used for referring to queues should be clarified. In addition to the symbols used for the vertical (stacked) queue $\bar{n}$ (n tilde), the horizontal (spatial) queue $n$ (real n) can follow some alternative concepts and measures. As previously introduced, the real queue is considered as a restrained queue (meaning a queue with constrained flow $q_n < S$, smaller than the saturation flow, either stopped with $V_n = 0$ and $K_n = K_f$ at jam density or flowing with $V_n < V_s$ and $K_n > K_s$). This concept excludes from $n$ the free discharging queue (meaning a “queue” flowing $q_n = S$, with no constraint ahead, with $V_n = V_s$ and $K_n = K_s$). The symbol $\hat{n}$ (n cap) is reserved for the free discharging queue measure (in general, not treated as queue, unless occurring mixed with restrained queues). The alternative concept of extended queue (with mixed states, encompassing restrained and free discharging queues) measures the real number of vehicles up to a reference section (A), represented the symbol $\hat{n}$ (n hat) instead. If referring to free flowing vehicles outside queues in balancing relations, the symbol used is $\hat{n}$ (a real number of vehicles not queueing). Conventional measures are eventually undefined and should be treated with care (e.g. those in the HCM are usually taken as extended queues but field values are currently related to slow moving vehicles, with $V_n \leq 5\text{mph} \approx 8\text{km/h}$, and better described as standing queues, if asking for a specific term; the concept of moving queue is avoided by defining intermediate moving states as congested for $V \leq 2/3 \cdot V_{\text{target}}$ or severely congested for $V \leq 1/3 \cdot V_{\text{target}}$; see TRB, 2011, pg.24-20,21). Finally, queue reduction (as vehicles taken from queue) is represented with the symbol $\bar{n}$ (n uncap) and sometimes combined with the previous description.
of state (e.g. using $\tilde{n}$ for vehicles taken from a free discharging “queue”). This degree of precision in notation is felt to be a valuable recourse for the clarity in the following discussion.

The **vertical queueing model estimate for the queue dissipation time** $\tilde{t}_s$ does not depend on the traffic situation (Case I or II) as long as $Q_A^{tn}$ and $q_A^{tn}$ are known and $Q_A^{tn} < q_A^{tn}$, being

$$\tilde{t}_s = \frac{\tilde{n}_{AT}}{q_A^{tn} - Q_A^{tn}}$$

what can be translated into the vertical queueing model estimate of queue lengths by using the hints given in Figure 1b. The dissipation has, however, a distinct pattern even in this simplified view.

For Case I, the maximum extent of queue from A corresponds to $\tilde{n}_{AT}(\tilde{z}_{AT})$ and decreases to zero during $\tilde{t}_s$ (for an intermediate position, the dissipation time can be linearly interpolated between $\tilde{n}_{AT}(\tilde{z}_{AT})$ and zero as long as the traffic conditions are constant during the dissipation time). Then the maximum queue is $\tilde{n}_{AM} = \tilde{n}_{AT}$ and its length is $\tilde{z}_{AM} = \tilde{z}_{AT}$ for this case (at the end of the previous peak period).

For Case II, the picture in Figure 1b shows an increase of the queue extent during dissipation (i.e. during the simultaneous reduction of queue size). An usual (but inadequate) measure for the maximum extent of queue from A estimated using the vertical queueing model corresponds to

$$\tilde{t}_s = \frac{\tilde{n}_{AT}}{q_A^{tn} - Q_A^{tn}} \quad \text{and} \quad (6a,b,c)$$

implying $\tilde{t}_s = \frac{\tilde{n}_{AT}}{q_A^{tn} - Q_A^{tn}}$ and a (vertical queueing model measure for the) of the maximum queue length as

$$\tilde{z}_{AM} = \tilde{z}_{AT} + \frac{Q_A^{tn} \tilde{t}_s}{N_v} \ell_v = \frac{\tilde{n}_{AM}}{N_v} \ell_v = \frac{\tilde{z}_{AT}}{1 - Q_A^{tn}/q_A^{tn}}$$

marking the position where the queue generated from A during the peak is dissipated (as effectively the queue would be null at $\tilde{t}_s$ and the full length $\tilde{z}_{AM}$ would be discharging the flow $q_A^{tn} > Q_A^{tn}$).

It is easy to realize that this usual measure is inconsistent with the general picture drawn from horizontal queueing models, unless $q_n^T = 0$ (the case of a stopped queue). Otherwise, a better measure for the maximum extent of queue from A estimated using the vertical queueing model corresponds to

$$\tilde{n}_{AM} = \tilde{n}_{AT} + (Q_A^{tn} - q_n^T) \tilde{t}_s = \left(\tilde{q}_A^{tn} - q_n^T\right) \tilde{t}_s = \frac{\tilde{n}_{AT}}{(q_A^{tn} - Q_A^{tn})(q_A^{tn} - q_n^T)}$$

(6d,e,f)
implying $\tilde{t}_s = \frac{\tilde{n}_{AT}}{q_A^T - Q_A^T}$ as well but a better estimate of (vertical queueing model measure for the) maximum queue length as

$$Z_{AM} = \tilde{Z}_{AT} + \tilde{Z}_{ts} = \frac{(\tilde{n}_{AT} + \tilde{n}_{ts})}{N_{\ell}}$$

$$(7d,e,f)$$

where $\tilde{n}_{ts} = (Q_A^T - q_n^T)\tilde{t}_s$ corresponds to the further extent of queue propagation during dissipation, with the same interpretation as before (effectively marking the position where the queue would be null at $\tilde{t}_s$ and the end of the full length $Z_{AM}$ that would be discharging the flow $q_A^T > Q_A^T$).

At an intermediate position, as long as the traffic conditions are constant during the dissipation time, the dissipation time can be linearly interpolated both between $\tilde{n}_{AT}(\tilde{Z}_{AT})$ or $\tilde{n}_{AM}(\tilde{Z}_{AM})$ and zero (for decrease at the front or increase at the back of queue as well as for the effective number of vehicles in queue, the difference of these interpolated values). However, the new image is not fully clear too.

It should be clear that the measures and their meaning are built by analogy to the picture grasped from the horizontal queueing model, and having no justification in the vertical model version. The distinction of the dissipation patterns is also based on the picture drawn by the horizontal queueing model and is shown ahead to be revealing of the large error (mainly for Case II) embodied into the queue length estimation with the vertical queueing model. This is a very important point as this simple model is widely used in the HCM traditional procedure for construction of QAP-Queue Accumulation Polygons and also in the “horizontal” queueing model of the TRANSYT 7F (since its Release 11). Then, these measures have to be corrected too.

The change of the traffic flow demand and traffic flow discharge in the next period will generate a shock wave in the back of queue with speed $w_b^T = \frac{q_n^T - Q_A^T}{K_n^T - K_u^T}$ and a shockwave in the front of queue with speed $w_f^T = \frac{q_A^T - q_n^T}{K_s^T - K_n^T}$ (as $q_n^T = q_n^T$ in the queue, initially during $T_n$, at least until some internal shock wave reaches an intermediate point in the queue). A generic evaluation can be tried, but not only the patterns of queue evolution are qualitatively distinct (as before) as the dissipation times for Cases I or II may differ too.

In Case I, if demand decreases in the next period with $Q_A^T = Q_{uA}^T < q_n^T$ in the back of queue and $q_A^T \approx q_A^T$ in the front of queue, the first particular case occurs. A queue shrinking wave is predicted (i.e. a shock wave for traffic recovery that moves forward by immediately reducing the queue size and extent) with
20 \[ w_{r}^{Tn} = \frac{q_{n}^{T} - Q_{A}^{Tn}}{K_{n}^{T} - K_{u}^{Tn}} \] (using \( q_{n}^{Tn} = q_{A}^{T} \), forward as \( Q_{A}^{Tn} < q_{A}^{T} \) and \( K_{u}^{Tn} < K_{n}^{T} \)). In general, Case I corresponds to dissipation of structural (recurrent) congestion where \( q_{A}^{T} = S_{A}^{T} \) and \( q_{n}^{Tn} = S_{n}^{Tn} \) are the (free, meaning unimpeded ahead) queue discharging flows at the structural capacity bottleneck (usually generated by some fixed physical or operational feature). The dissipation as recovery time (by the queue shrinking to A) is

\[
T_{p} = \frac{Z_{AT}}{w_{r}^{Tn}} = \frac{w_{c}^{T}}{w_{r}^{Tn}} \cdot T = \frac{T}{w_{r}^{Tn}/w_{c}^{T}} \tag{8a,b,c}
\]
as the front of queue is stationary (\( w_{r}^{Tn} = 0 \)). Then the flow returns to normal operation at A with \( q_{A}^{Tn} = Q_{A}^{Tn} \) after \( T_{p} \), the recovery time (in general, \( q_{n}^{Tn} \) can also change in the next period and a perturbation wave can be generated from the front of queue for propagating this secondary change, inside the queue, up to its end).

Note that the recovery wave can be used to define the number of vehicles reduced from the queue \( \bar{n}_{Rt} \) as \( \bar{n}_{Rt} = K_{n}^{T} (z_{AT} - z_{Rt}) = K_{n}^{T} w_{r} \cdot t \), then delivering

\[
\bar{n}_{Rt} = \frac{\bar{n}_{Rt}}{f_{Qn}^{Tn}} = \frac{(q_{n}^{T} - Q_{A}^{Tn}) t}{1 - K_{n}^{T} / K_{n}^{T}} = \frac{(q_{n}^{T} - Q_{A}^{Tn}) t}{1 - Q_{A}^{Tn} / V_{u}^{Tn}} \tag{9a,b,c}
\]
where \( f_{q} = \frac{1}{1 - K_{n}^{T} / K_{n}^{T}} = 1 - \frac{Q_{A}^{Tn} / V_{u}^{Tn}}{N_{x} / \ell_{x}} \), in part similarly to (3a,b) and (4a,b). Then the recovery time \( T_{p} \) can be evaluated by the condition \( n_{AT} = \bar{n}_{R(t)} \) (delivering the same result

\[
T_{p} = \frac{(Q_{A}^{T} - q_{A}^{T}) / f_{Qn}^{Tn}}{(Q_{n}^{T} - Q_{A}^{T}) / f_{Qn}^{Tn}} \cdot T, \quad q_{n}^{T} = q_{A}^{T} \). The approach is clearly legitimate because the recovery wave is taking vehicles from the queue (where density is \( K_{n}^{T} \)) and delivers \( n_{t} = n_{AT} - \bar{n}_{Rt} \) for \( t \in [0; t_{p}] \) in period

\[
T_{p} = [T; T + t_{p}] . \text{ If } n_{AT} \text{ is known (from the previous period), then } T_{p} = \frac{n_{AT}}{(q_{n}^{T} - Q_{A}^{T}) / f_{Qn}^{Tn}} \text{ is obtained. Wave speeds can be obtained by imposing } \bar{n}_{R(t)} = K_{n}^{T} w_{r} \cdot T_{p} = n_{AT} \text{, if needed.}
\]

In Case II, if capacity increases in the next period with \( Q_{A}^{T} \sim Q_{A}^{Tn} \) in the back of queue but \( q_{A}^{T} > q_{A}^{T} \) in the front of queue (with \( q_{n}^{Tn} > Q_{A}^{Tn} \)), the propagation speed \( w_{c}^{Tn} \) at the back of queue is maintained while a dissipation (queue discharging) wave starts with speed \( w_{s}^{Tn} = -\frac{q_{n}^{Tn} - q_{A}^{T}}{K_{s}^{Tn} - K_{A}^{T}} \) from the front of queue (backward
if \( q_{A}^{Tn} > q_{A}^{T} \) and \( K_{S}^{Tn} < K_{A}^{T} \), using \( q_{n}^{Tn} = q_{n}^{T} = q_{A}^{T} \). In general, Case II corresponds to dissipation of incidental (non-recurrent) congestion where \( q_{A}^{T} = S_{A}^{T} \) is the (reduced) free queue discharging capacity generated by an incident and \( q_{A}^{Tn} = S_{n}^{Tn} \) is the free queue discharging capacity after the removal of the incident (it also corresponds to traffic interruption and free queue flow discharging for discontinuous/interrupted flow conditions or to the easing of operational bottlenecks due reduction of merging or weaving after the peak). If demand does not change, the congestion wave speed remains constant (\( w_{b}^{Tn} = w_{c}^{Tn} = w_{c}^{T} \)) at the back of the queue (at least up to the meeting of some merging or diverging section during queue propagation). The queue will decrease in size and length as long as the dissipation wave is faster than the congestion wave (i.e.

\[ w_{s}^{Tn} > w_{c}^{Tn} \], a condition that can usually be translated into \( Q_{A}^{Tn} < q_{A}^{Tn} \) (even more as \( K_{S}^{Tn} > K_{u}^{Tn} \)).

In this case, the dissipation time can be obtained by setting the condition for the dissipation wave to reach the congestion wave, at a position \( M \) further away from \( A \) than \( z_{AT} \) with \( z_{AM} = z_{AT} + w_{c}^{Tn} t_{q} = w_{s}^{Tn} t_{q} \), delivering the dissipation as the **reaching time**

\[
t_{q} = \frac{z_{AT}}{w_{s}^{Tn} - w_{c}^{Tn}} = \frac{w_{c}^{T}}{w_{s}^{Tn} - w_{c}^{Tn}} = \frac{T}{w_{s}^{Tn}/w_{c}^{T} - w_{c}^{Tn}/w_{c}^{T}} \tag{10a,b,c}
\]

(where the notation \( w_{c}^{T}, w_{c}^{Tn} \) permits different congestion wave speeds, as \( Q_{A}^{Tn} \) can also change even with \( Q_{A}^{Tn} \sim Q_{A}^{T} \)). These formulas \( 10a,b,c \) include the previous formulas \( 8a,b,c \) as \( w_{f}^{Tn} = 0 \) (stationary front of queue) and \( w_{b}^{Tn} = -w_{f}^{Tn} \) (forward instead of backward wave at the back of queue) in Case I.

The point \( M \) is the maximum extent of queue at \( z_{AM} = w_{s}^{Tn} t_{q} = \frac{z_{AT}}{1 - w_{c}^{Tn}/w_{s}^{Tn}} \), the position where the queue is dissipated. At position \( M \), a normalization wave starts at \( t_{q} \) with wave speed \( w_{m}^{Tn} = \frac{q_{A}^{Tn} - q_{A}^{Tn}}{K_{S}^{Tn} - K_{u}^{Tn}} \) that returns the road to the normal condition at position \( A \) (a flow drop \( q_{A}^{Tn} \) to \( Q_{A}^{Tn} \)) after the **normalization time**

\[
t_{m} = \frac{z_{AM}}{w_{m}^{Tn}} = \frac{z_{AT}}{w_{m}^{Tn} (1 - w_{c}^{Tn}/w_{s}^{Tn})} = \frac{T}{w_{m}^{Tn}/w_{c}^{Tn} (1 - w_{c}^{Tn}/w_{s}^{Tn})} \tag{11a,b,c}
\]

(queue discharging finishes when the normalization wave reaches \( A \), meaning that the dissipation time as seen at \( M \) with \( t_{s}^{M} = t_{q} \) occurs before the dissipation time as seen at \( A \) with \( t_{s}^{A} = t_{q} + t_{m} \)). As
\[ z_{AM} = w_{s}^{tn} \cdot t_q = w_{m}^{tn} \cdot t_m \], the total dissipation time \( t_i = t_q + t_m \) can be split as \( t_q = t_i \cdot \frac{w_m}{w_s + w_m} \) and \( t_m = t_i \cdot \frac{w_s}{w_s + w_m} \) (with a switched part ratio of speeds, a rule that will be shown to be useful ahead).

Note that, knowing the end of peak queue \( n_{AT} \), the reference queue (as if measured from the reference section A, keeping queue density) can be obtained incrementally by using \( n_{A(T+1)} = n_{AT} + n_{St} \) with

\[ n_{St} = \frac{\bar{n}_{St}}{f_{Qn}^{tn}} = \frac{(Q_A^{Tn} - q_{n}^{Tn})t}{f_{Qn}^{tn}} \] and, knowing \( t_q \), the maximum reference queue \( n_{AM} \) (related to the maximum queue extent \( z_{AM} \)) can be obtained as \( n_{AM} = n_{AT} + n_{St(q)} = \frac{\bar{n}_{AT}}{f_{Qn}^{tn}} + \frac{\bar{n}_{St(q)}}{f_{Qn}^{tn}} = \left( \frac{Q_A^{T} - q_{n}^{T}T}{f_{Qn}^{tn}} \right) + \left( \frac{Q_A^{Tn} - q_{n}^{Tn}}{f_{Qn}^{tn}} \right) t_q \)

(adding queue build up by the congestion wave \( w_{c}^{tn} \) during \( t_q \)), as generically

\[ n_{A(T+1)} = n_{AT} + n_{St} = \frac{\bar{n}_{AT}}{f_{Qn}^{tn}} + \frac{\bar{n}_{St}}{f_{Qn}^{tn}} = \left( \frac{Q_A^{T} - q_{n}^{T}T}{f_{Qn}^{tn}} \right) + \left( \frac{Q_A^{Tn} - q_{n}^{Tn}}{f_{Qn}^{tn}} \right) t_q \] (3d,e,f)

with \( q_{n}^{Tn} = q_{n}^{T} \) for \( t \in [0; t_q] \) in period \( T_q = [T; T + t_q] \). In a similar way, the dissipation wave can be used to define the number of vehicles that left the queue from the front \( \bar{n}_{St} = K_{n}^{T} \cdot z_{St} = K_{n}^{T} \cdot w_{s} \cdot t \), delivering

\[ \bar{n}_{St} = \frac{\bar{n}_{St}}{f_{Qn}^{tn}} = \frac{(q_{A}^{Tn} - q_{n}^{Tn})t}{1 - K_{s}^{Tn} / K_{n}^{T}} \geq 1 - \frac{q_{A}^{Tn} / V_{s}^{tn} \cdot N_{f} / \ell_{v}}{1 - K_{s}^{Tn} / K_{n}^{T}} \] (9d,e,f)

where \( \bar{f}_{es}^{tn} = \frac{1}{f_{Qn}^{tn}} \) with \( f_{Qn}^{tn} = 1 - \frac{K_{s}^{Tn} / K_{n}^{T}}{1 - \frac{q_{A}^{Tn} / V_{s}^{tn} \cdot N_{f} / \ell_{v}}{1 - K_{s}^{Tn} / K_{n}^{T}}} \) similarly to (3a,b) and (4a,b). The approach is clearly legitimate because the shock wave \( w_{s}^{tn} \) is taking vehicles from the front of the queue (where density is \( K_{n}^{T} \)).

As the shock wave \( w_{c}^{tn} \) keeps adding vehicles to the end of the queue, then \( n_{1} = n_{A(T+1)} - \bar{n}_{St} \) in period \( T_q \). If \( n_{AM} \) is known, then \( t_q = \frac{n_{AM}}{(q_{A}^{Tn} - q_{n}^{Tn})/f_{Qn}^{tn}} \) is obtained. Otherwise, if \( n_{AT} \) is known (from the previous period), then \( t_q = \frac{n_{AT}}{(q_{A}^{Tn} - q_{n}^{Tn})/f_{Qn}^{tn} - (Q_A^{Tn} - q_{n}^{Tn})/f_{Qn}^{tn}} \) and \( n_{AM} = n_{AT} + \frac{(Q_A^{Tn} - q_{n}^{Tn})t_q}{f_{Qn}^{tn}} \) or
1. \( n_{AM} = \frac{(q^T_n - q^T_n) t_q}{f^T_{Sn}} \) corresponding to the position M reached at \( t_q \) (a “fictitious” queue as \( n_{iq} = 0 \) and there is no more vehicles present in queue). The implicit wave speeds can be obtained using

2. \( n_{AM} = n_{A(T+t_q)} = n_{AT} + K_n w_c^{Tn} t_q \) or \( n_{AM} = \tilde{n}_{St(q)} = K_n w_s^{Tn} t_q \), as before, if needed.

3. The same approach can be extended to the normalization wave and its extent by defining \( \tilde{n}_{Mn} \) as

4. \( \tilde{n}_{Mn} = K_n^T z_{Mn} = K_n^T w_m^{Tn} t \) and deriving the number of vehicles reduced from the “fictitious” queue by the normalization wave as \( \tilde{n}_{Mn} = \frac{(q^T_n - Q^T_n) t}{K^T_s / K^T_n - K^T_u / K^T_n} = \frac{(q^T_n - Q^T_n) t}{f^T_{Qn} - f^T_{Sn}} = \frac{\tilde{n}_{Mn}}{f^T_{Qn} - f^T_{Sn}} \) in period \( T_m = [T + t_q; T + t_q + t_m] \) (i.e. now the correction factor is \( \tilde{f}^T_{cm} = \frac{1}{f^T_{Qn} - f^T_{Sn}} \). Then, if \( n_{AM} \) is known,

5. \( t_m = \frac{n_{AM}}{(q^T_n - Q^T_n) / (f^T_{Qn} - f^T_{Sn})} \) is obtained from \( \tilde{n}_{Mn} = n_{AM} \) (as

6. \( n_{AM} = K_n^T z_{AM} = K_n^T w_m^{Tn} = K_n^T w_{m}^{Tn} t_m \), Note that the right answers were gathered by supposing a

7. “fictitious” queue that was never present between A and M.

8. However, the picture can be drawn in a better (more realistic) way. Even if there is no queue present from M to A at \( t_q \), the number of vehicles observed between M and A (in free queue discharging flow) is

9. evaluated as \( \tilde{n}_{AM} = K_n^{Tn} z_{AM} = K_n^{Tn} w_s^{Tn} \cdot t_q = K_n^{Tn} w_m^{Tn} \cdot t_m \) and is a better measure of the effective number of vehicles between M and A as a free discharging “queue” (as it is not queueing). Using the previously

10. developed formulas, the discharging “queue” is \( \tilde{n}_{St} = \frac{(q^T_n - q^T_n) t}{K^T_s / K^T_n - 1} = \tilde{n}_{St} \) with \( \tilde{f}^T_{sn} = \frac{1}{f^T_{Sn}} \) where

11. \( \tilde{f}^T_{Sn} = \frac{K^T_s}{K^T_s - 1} \geq \frac{N_{\ell} / \ell \nu}{q^T_n / V_s^{Tn}} - 1 \) is the complementary conversion factor (for \( \tilde{n}_{St} = \frac{K^T_s}{K^T_n} \cdot n_{St} \) with

12. \( f^T_{Sn} = \frac{\tilde{f}^T_{Sn}}{1 - \tilde{f}^T_{Sn}} \) or \( \tilde{f}^T_{Sn} = \frac{1}{f^T_{Sn}} = \frac{1 - \tilde{f}^T_{Sn}}{f^T_{Sn}} \). Again, the dissipation time \( t_q \) (obtained by the condition \( n_{iq} = 0 \))

13. also satisfies \( \tilde{n}_{AM} = \tilde{n}_{St(q)} \) with \( \tilde{n}_{AM} = \frac{K^T_s}{K^T_n} \cdot n_{AM} \) and \( t_q = \frac{\tilde{n}_{AM}}{(q^T_n - q^T_n) / \tilde{f}^T_{Sn}} \) (as before, by noting that

14. \( \tilde{f}^T_{Sn} = \frac{K^T_s}{K^T_s f^T_{Sn}} \), showing again that results are simple to reproduce for any assumed constant traffic density).
Ahead, procedures that can handle different densities are presented (by balancing arrivals and departures but considering intermediate densities of vehicles along road segments, instead of tracking wave speeds).

The measure of the number of vehicles reduced from the free discharging “queue” during normalizing can be defined as \( \bar{n}_{M_t} \) from M to A and evaluated using \( \bar{n}_{M_t} = \frac{\bar{q}_s^T - \bar{Q}_A^T}{1 - \frac{K_u^T}{K_s^T}}\frac{t}{f_{QS}^T} = \frac{1}{f_{QS}^T} - \frac{f_{Sn}^T}{f_{Sn}^T - f_{Sn}^T} \) and the number of vehicles in the free discharging “queue” from A can be defined and evaluated as \( \bar{n}_{A_t} = \bar{n}_{AM} - \bar{n}_{M_t} \), going to zero in \( T_m \) (similarly, returning to normal traffic conditions). The normalization time can be evaluated by the condition \( \bar{t}_{M_t} = \bar{t}_{AM} \) for known \( \bar{n}_{AM} \), delivering the previous formulas for \( t_m \) and recovering the wave speed by \( \bar{n}_{AM} = \frac{K_s^T}{K_n^T} \bar{n}_{AM} \) or \( \bar{n}_{M_t} = \frac{K_s^T}{K_n^T} \bar{n}_{M_t} \). If the maximum extent of the free discharging “queue” is known, then \( t_m = \frac{\bar{q}_s^T - \bar{q}_A^T}{f_{QS}^T} \) (as before, by noting that \( f_{QS}^T = f_{Sn}^T \), showing that results are simple to reproduce as long as a constant traffic density can be assumed and appropriate factors are used to convert wave speeds into vehicle numbers).

Note that, similarly, the number of vehicles in free queue discharging flow added between S and A (as the free discharging “queue”) can be defined as \( \bar{n}_{St} = \frac{K_s^T}{K_n^T} \bar{n}_{AM} \) or \( \bar{n}_{M_t} = \frac{K_s^T}{K_n^T} \bar{n}_{M_t} \). If the maximum extent of the free discharging “queue” is known, then \( t_m = \frac{\bar{q}_s^T - \bar{q}_A^T}{\bar{q}_s^T} \) (as before, by noting that \( f_{QS}^T = f_{Sn}^T \), showing that results are simple to reproduce as long as a constant traffic density can be assumed and appropriate factors are used to convert wave speeds into vehicle numbers).

Note also that, even if the discussion is mostly related to Case II, it applies to Case I as then \( K_s^T = K_n^T = \frac{K_s^T}{K_n^T} \) and \( \bar{n}_t \) is \( n_t \) trivially. Then a common description can be developed to both basic cases just setting \( t_q = 0 \) for Case I and rewriting \( t_m \) as \( t_p \) also for Case II (as done in the summary Tables 1 and 2 ahead). Alternatively, the generic description can set \( n_{AM} = 0 \) for Case I (resulting from \( w_c^T = w_r^T \) forward and \( w_s^T = 0 \)), with \( t_m = 0 \) as a side effect, but losing too much of the qualitative difference between cases.
Explicitly or not, the conditions for the particular cases have to be checked so as to select appropriate
formulas (at least for the maximum queue extent $z_{AM}$, the dissipation time $t_q$, and related variables) or
adjusting vertical queueing model estimates. As discussed, these conditions should consider $Q_{TA}^{Tn}$, $q_{TA}^{Tn}$ and also
$q_{in}^T$ (from the queue generated in the over-demand period). At least for pure cases (I and II), the previous
discussion suggest that Case I can again be subsumed by Case II. For example, both can be treated by defining
$t_r = t_q + t_p$ and $t_p = t_m$ (equivalently, $w_r = w_m$) for Case II and assuming artificially for Case I that
$t_q = 0$ and $t_p = t_r$, meaning that the maximum queue extent occurs at $M$ in $t_q^M = t_q$ and queue clearing
occurs at $A$ in $t_s^A = t_s^M + t_p$, distinguishing queue dissipation at $M$ in $t_q^M = t_q$ from dissipation and
clearing at $A$ in $t_r^A$ (if referred to the beginning of the peak period, maximum queue extent and queue
clearing occur in $t_{sM} = T + t_q$ and $t_{sA} = t_{sM} + t_p$). But, for horizontal queueing models, the explicit
distinction of Case I and Case II seems to be the best option and the procedure can also check for mixed cases
(e.g. verifying both $Q_{TA}^{Tn} < q_{in}^T$ and $q_{TA}^{Tn} > q_{in}^T$ or $q_{TA}^{Tn} < q_{in}^T$, all for $q_{TA}^{Tn} > Q_{TA}^T$).

In a wider theoretical view, these representations of the dissipation and normalization phases are crude
(as the queue discharge occurs with acceleration of vehicles and platoon dispersion, both phenomena that a first
order macroscopic traffic models are not adequate to address). During dissipation, vehicles from the front queue
will shortly accelerate to a larger speed (and platoon dispersion will occur). During normalization, the normal
demand will meet the dissipation flow after the acceleration phase and the picture is even worse. No first order
traffic model can handle both these phenomena appropriately (indeed, acceleration can be treated in an ad hoc
manner as in the HCM’s algorithm for freeway systems, by imposing a vehicle acceleration model). But for
both phenomena, the previous representation is taken as sufficiently insightful here (e.g. compared to Mongeot,
Lesort, 2000).

For the here proposed methods, the simplification remains with a problem: to decide on a representative
value of traffic conditions from free queue discharging. As vehicles are accelerating during part of the time at
least, speeds and densities will be changing even ignoring platoon dispersion. The practical answer (i.e. that
obtained from field data) is usually representative of the bottleneck section. In general, field data will
 correspond to the observation of saturation flow with queue discharging at the bottleneck section or during
green at the stop line of signalized intersections (outside the few seconds corresponding to lost time). In the
current view, this condition is taken as the limit of the forced flow regime for unimpeded conditions ahead but,
for long green times at least, acceleration can easily produce traffic conditions corresponding to normal flow
instead (as the concept of free queue discharging requires that traffic flow can collectively accelerate ahead).
2.3 Relating Vertical and Horizontal Queue Dissipation Time Estimates

The main point here is the comparison of results to those obtained with vertical queueing models (as those implicit in the conventional HCM formulas for intersections). Differences in queueing were previously
commented. The following discussion is related to differences in dissipation time. Usually $\tilde{t}_q \neq t_q \neq t_f$ and the discrepancies between formula (5) and (8a,b,c), (10a,b,c) plus (11a,b,c) are not easy to fix.

The horizontal queueing model estimate for the dissipation time at $M$ (the relevant variable for estimating the queue evolution, e.g. by building a QAP-Queue Accumulation Polygon, meaning that free queue discharging during the normalization time is not taken as queueing) is

$$t_q = \frac{z_{AM}}{w_s} = \frac{z_{AT}}{w_s - w_c} = \frac{\bar{n}_{AT}}{(K_n - K_u)} \text{with} \quad w_c = \frac{Q_n - q_n}{K_n - K_u} \text{and} \quad w_s = \frac{Q_n - q_n}{K_n - K_s} \text{using}$$

$$z_{AT} = \frac{n_{AT}}{K_n} \text{and} \quad n_{AT} = \frac{\bar{n}_{AT}}{1 - K_u / K_n} \text{or} \quad t_q = \frac{\bar{n}_{AT}}{(1 - K_u / K_n)} \left( \frac{q_n - q_n}{1 - K_u / K_n} \right) = \frac{\bar{n}_{AT}}{(1 - K_u / K_n)} \frac{Q_n - q_n}{1 - K_u / K_n} = \frac{\bar{n}_{AT}}{(1 - K_u / K_n)}$$

(10d,e)

Both delivers $t_q = \frac{\bar{n}_{AT}}{(1 - K_u / K_n)}$ and the conversion formula

$$t_q = \frac{\bar{n}_{AT}}{(1 - K_u / K_n)} \left( \frac{Q_n - q_n}{1 - K_u / K_n} \right) = \frac{\bar{n}_{AT}}{(1 - K_u / K_n)} \frac{Q_n - q_n}{1 - K_u / K_n} \text{with} \quad n_{AT} = \frac{\bar{n}_{AT}}{(1 - K_u / K_n)}$$

$$f_{Qn} = \frac{1 - K_u}{K_n}, \quad f_{Tn} = \frac{1 - K_S}{K_n}, \quad f_{Sn} = \frac{1 - K_n}{K_n}, \quad \text{also} \quad K_u = \frac{Q_n}{V_u}, K_u = \frac{Q_n}{V_u}$$

unconstrained flows (with unconstrained speeds) and $K_n, K_S$ are traffic densities at queueing and queue discharging flows (with $K_n = K_n$ also during queue dissipation). The additive composition (as implied by

$$n_{AM} = n_{A(T+q)} \) is replaced by $n_{AM} = \frac{n_{AT}}{1 - w_c / w_s}$ (as implied by $z_{AM} = \frac{z_{AT}}{1 - w_c / w_s}$) but the dissipation time $t_q$ can also be obtained by using

$$n_{AM} = n_{A(T+q)} = n_{AT} + n_{S(T)} = \bar{n}_{S(T)}, \quad \text{producing the same result.}$$

Anyway, it delivers no simple formula by requiring detailed knowledge of traffic conditions that determines wave speeds (but delivering a viable estimate for queues at intersections, as further simplification can be obtained by using $K_n = \frac{N}{f_{\nu}}$, at least for stopped queues, and good approximations can be
developed if queue discharging conditions are known, e.g. if \( q_{A}^{T_{n}} = S_{A}^{T_{n}} \) at saturation flow with its known speed). Values for the correction factor \( t_{q}/\tilde{t}_{s} \) against \( Q_{A}/N_{\ell} \) and \( V_{A} = V_{uA} \) for

\[ q_{A}/N_{\ell} = 2000; 1800; 1500 \text{ v/h per lane are shown in Table A3 (ahead), for stopped and moving queues (} q_{n}/N_{\ell} = 0 \text{ or } 1000 \text{ v/h per lane). Differences are clearly significant for all combinations (differences are less significant when } V_{s} \text{ approaches } V_{A}. \]

Correspondingly, \( t_{m} = \frac{Z_{m}^{T_{n}}}{w_{m}^{T_{n}}} = \frac{Z_{m}^{T_{n}}}{w_{m}^{T_{n}}(1-w_{c}^{T_{n}}/w_{s}^{T_{n}})} \) with \( w_{m}^{T_{n}} = \frac{q_{A}^{T_{n}}-q_{A}^{T_{n}}}{K_{u}^{T_{n}}-K_{u}^{T_{n}}} \) can be restated, by defining fictitiously \( n_{AM} = K_{u}^{T_{n}}Z_{AM} \) (as the real queue at M is zero when the dissipation wave catch the congestion wave) or using \( Z_{AT} = \frac{n_{AT}}{K_{n}^{T_{n}}} = \frac{n_{AT}}{K_{n}^{T_{n}}-K_{u}^{T_{n}}} \) again, as \( t_{m} = \frac{n_{AM}}{K_{n}^{T_{n}}}, \frac{w_{m}^{T_{n}}}{w_{m}^{T_{n}}}, \frac{w_{m}^{T_{n}}}{w_{s}^{T_{n}}-w_{c}^{T_{n}}} \) or

\[ t_{m} = \frac{n_{AM}}{(q_{A}^{T_{n}}-q_{A}^{T_{n}})/(f_{Q_{n}}^{T_{n}}-f_{S_{n}}^{T_{n}})} = \frac{\tilde{t}_{s}}{(q_{A}^{T_{n}}-q_{A}^{T_{n}})/(f_{Q_{n}}^{T_{n}}-f_{S_{n}}^{T_{n}})} \]

\[ (11d,e) \]

for \( t_{m} = f_{m}\tilde{t}_{s} \) (as \( \tilde{t}_{s} = \frac{n_{AT}}{q_{A}^{T_{n}}-Q_{A}^{T_{n}}} \)) where \( f_{m} = \frac{(f_{Q_{n}}^{T_{n}}-f_{S_{n}}^{T_{n}})/(f_{S_{n}}^{T_{n}}^{T_{n}}-Q_{A}^{T_{n}}^{T_{n}})/(f_{Q_{n}}^{T_{n}}/f_{Q_{n}}^{T_{n}})}{f_{Q_{n}}^{T_{n}}/f_{Q_{n}}^{T_{n}}}, \)

\[ f_{Q_{n}}^{T_{n}} = 1 - \frac{K_{u}^{T_{n}}}{K_{n}^{T_{n}}}, f_{S_{n}}^{T_{n}} = 1 - \frac{K_{u}^{T_{n}}}{K_{n}^{T_{n}}}, f_{S_{n}}^{T_{n}} = 1 - \frac{K_{u}^{T_{n}}}{K_{n}^{T_{n}}}, \]

\[ f_{Q_{n}}^{T_{n}} = 1 - \frac{K_{u}^{T_{n}}}{K_{n}^{T_{n}}}, f_{S_{n}}^{T_{n}} = 1 - \frac{K_{u}^{T_{n}}}{K_{n}^{T_{n}}} \]

Better defining \( n_{AM} = K_{S}^{T_{n}}Z_{AM} \), delivers no simpler result. The same result is also obtained by using \( \tilde{n}_{M_{(im)}} = n_{AM} \) (or \( n_{AM} = n_{AT} + n_{S_{(ii)}} \)) with \( \tilde{n}_{AM} = n_{AT} + n_{S_{(ii)}} \) and the previous formula for the dissipation time \( t_{q} \) (of course, the simpler results previously discussed are also obtained if \( n_{AM} \), or \( \tilde{n}_{AM} \) is known).

As before, the implied values of wave speeds are consistent. For the peak period \( T \), wave speeds can be obtained by imposing \( Z_{AT} = w_{c}^{T_{n}} \cdot T \) (after obtaining \( Z_{AT} \)). Even better, as can easily be checked, balancing in
any interval \( T_0 \subseteq T \) as \( 0 \leq t_0 \leq T \), and segment \( z_0 = w_c^T t_0 \), with \( 0 \leq z_0 \leq z_{AT} \), delivers the wave speed \( w_c^T \) (for the next period \( T_n \), however, the analysis in \( t_\ell \) should consider \( t_q \) and \( t_m \) separately).

Another path is, however, more interesting. The horizontal queueing model estimate for the dissipation time at \( A \), combining (10a) and (11b) in \( t_\ell = t_q + t_m \), is \( t_\ell = \frac{Z_{AT}}{w_s - w_c} + \frac{Z_{AT}}{w_m (1 - w_c / w_s)} \) that translates to \( t_\ell = \frac{Z_{AT}}{w_m / w_s (w_s - w_c)} = \frac{Z_{AT}}{w_m / w_s (w_s - w_c)} = \frac{Z_{AT}}{w_m / w_s (w_s - w_c)} \) or

\[
\begin{align*}
  t_\ell &= \frac{n_{AT}}{K_n^T (q_A^T - Q_A)} (w_s - w_c) + \frac{Q_s}{K_n^T (q_A^T - Q_A)} (w_s + w_c) \\
  &= \frac{K_n^T - K_u^T}{K_n^T (q_A^T - Q_A)} (w_s + w_c) \\
  &= \frac{K_n^T - K_u^T}{K_n^T (q_A^T - Q_A)} (w_s + w_c) \\
  &= \frac{K_n^T - K_u^T}{K_n^T (q_A^T - Q_A)} (w_s + w_c),
\end{align*}
\]

again using \( Z_{AT} = \frac{n_{AT}}{K_n^T} \) and

\[
  n_{AT} = \frac{\tilde{n}_{AT}}{1 - K_u^T / K_n^T},
\]

but now surprisingly (as by adding 10e and 11e) delivering

\[
  t_\ell = \left( \frac{q_A^T - Q_A^T}{1 - K_u^T / K_n^T} \right) \tilde{t}_s = \left( \frac{f_{Q_n}}{f_{Q_n}} \right) (a \text{ remarkably simple relation revealing that the alternative expressions for } t_\ell, \tilde{t}_s \text{ are similar in content but } t_\ell = \tilde{t}_s \text{ only if } Q_A^T = Q_A^T). \]

Note that the same relation applies for \( t_p \) in Case I as (8a) implies \( t_p = \frac{n_{AT}}{K_n^T} \) and directly delivers

\[
  t_p = \left( \frac{q_A^T - Q_A^T}{1 - K_u^T / K_n^T} \right) \tilde{t}_s = \left( \frac{f_{Q_n}}{f_{Q_n}} \right)
\]

(then providing another justification for using \( t_\ell = t_p \) and \( t_q = 0 \) in Case I).

The simple relation can also be obtained by balancing flows during the periods \( T \) and \( T_n \) considering traffic densities along the segments between section \( A \) and any section at \( z \) before (unaffected by) the queue (with unconstrained flows). For the peak period \( T \) along \( Z_{AT} \), the traffic density is \( K_u^T \) before queueing, the increase in density (the queue) for the peak period \( T \) is \( K_n^T - K_u^T \) and should be filled by the flow imbalance during peak as \( K_u^T Z_{AT} + Q_A^T T = K_n^T Z_{AT} + Q_A^T T \) or \((K_n^T - K_u^T) Z_{AT} = (Q_A^T - Q_A^T) T\), that reproduces formula (4a,b), as \( n_{AT} = Z_{AT} K_n^T \), and reveals the real meaning of \( \tilde{n}_{AT} = (Q_A^T - Q_A^T) T \) (see Yi, Tiao, Zhan, 2008). For the next period \( T_n \) along \( Z_{AM} \), the traffic density is \( K_u^T \) after queueing (i.e. after \( T_\ell \)) and the decrease in
density from queue after the peak period in $T_n$ is $K_n^T - K_{un}^T$ along $z_{AT}$ only and should be fed by the flow imbalance during the dissipation and clearing period as

$$K_n^Tz_{AT} + K_{un}^T \left( z_{AM} - z_{AT} \right) + Q_A^T \cdot t_q = K_n^Tz_{AM} + Q_A^T \cdot t_q \text{ or } \left( K_n^T - K_{un}^T \right)z_{AT} = (q_A^T - Q_A^T)t_q.$$  Using both delivers formulas (12a,b), showing another path for deriving horizontal queueing model formulas (whatever the dynamics along $z$, if ignoring the time translation of periods).

In general, $t_q \equiv \tilde{t}_s$ is a good approximation for the dissipation time at $A$ (the variable that explains the bottleneck capacity) because in general $K_n^T$ is reasonably greater than $K_u^T, K_{un}^T$ (then dominating both the numerator and the denominator of the conversion ratio) but better formulas are

$$t_q = \frac{\tilde{t}_s \cdot f_{Qn}^T}{f_{Qn}^T} \approx \frac{N_{qs} - Q_{Aq}^T}{\ell_v} \cdot \frac{V_{un}^T}{V_u^T}$$

(12d,e)

with

$$f_{Qn}^T = \frac{K_n^T - K_{un}^T}{K_n^T - K_{un}^T} \approx \frac{N_{qs} / \ell_v - Q_{Aq}^T / V_u^T}{N_{qs} / \ell_v - Q_{Aq}^T / V_u^T}.$$  Obtained with similar approximations as those embodied into formulas 3a,b). If demand is constant, as when queueing is caused by incidents or interruptions, then $t_q = \tilde{t}_s$.

Values for the correction factor $t_q / \tilde{t}_s$ against $Q_{Aq}^T / N_{qs}$ and $V_A$ for $\ell_v = 8m, 15m, 25m$ are shown in Table A2 (ahead). Differences are rarely over 15%, unless flow and spacing (or demand drop) are both large.

A similar (not so simple) relation can be developed for $t_q$ or $t_m$ by balancing flows between unaffected sections up to the dissipation of queues (when the traffic density between $M$ and $A$ is $K_{un}^T$, before clearing when the traffic density is $K_{un}^T$). Up to dissipation of queues along $z_{AM}$, the increase in density is $K_{sn}^T - K_{un}^T$ should be filled by the flow imbalance during peak and dissipation time as

$$\left( K_{sn}^T - K_{un}^T \right)z_{AM} = (Q_{Aq}^T - q_{Aq}^T)T - (q_{Aq}^T - Q_{Aq}^T)t_q.$$  Then up to clearing along $z_{AM}$, the corresponding decrease in density is $K_{sn}^T - K_{un}^T$ and should be fed by the flow imbalance during the clearing period as

$$\left( K_{sn}^T - K_{un}^T \right)z_{AM} = (Q_{Aq}^T - q_{Aq}^T)t_m.$$  Knowing $t_q = t_q + t_m$, the components can be obtained, delivering

$$t_q = \frac{Q_{Aq}^T - q_{Aq}^T}{q_{Aq}^T - Q_{Aq}^T}T - \frac{K_{sn}^T - K_{un}^T}{K_{sn}^T - K_{un}^T}t_q \approx \tilde{t}_q - t_q \cdot f_{Su}^T / f_{Su}^T \frac{1 - \left( K_{sn}^T - K_{un}^T \right) / \left( K_{sn}^T - K_{un}^T \right)}{1 - f_{Su}^T / f_{Su}^T}$$

(10f,g)

and/or


\[ t_m = \frac{\frac{q^T \text{T}}{q^T_A - q^T_A \text{T}}}{1 - \frac{K_S^T - K_u^T}{K^T - K_u^T}} = \frac{t_r - t_q}{1 - \frac{f^T_{su}}{\frac{f^T_{su}}{f^T}}}, \quad (11f,g) \]

with

\[ \frac{f^T_{su}}{f^T_{su}} = \frac{K_S^T - K_u^T}{K^T - K_u^T} = \frac{q^T}{V^T - q^T_A} = \frac{q^T}{V^T - q^T_A \text{T}} = \frac{f^T_{su}}{f^T_{su}} \]

(where, again, further simplification can be obtained by introducing known traffic flow variables to estimate traffic densities at queue discharging). If demand is constant, as when queuing is caused by incidents or interruptions, the previous relation just implies trivially that \( t_r = t_q + t_m \). However, the switched part rate of speeds rule can always be used, delivering

\[ t_q = t_r \cdot \frac{w_m}{w_s + w_m} = t_r \cdot \frac{\left(q^T - q^T_A\right) / \left(f^T_{su} - f^T_{su}\right)}{\left(q^T_A - q^T\right) / \left(f^T_{su}/f^T_{su}\right)} \]

(10h,i)

(equivalent to the previously obtained formulas) and

\[ t_m = t_r \cdot \frac{w_s}{w_s + w_m} = t_r \cdot \frac{\left(q^T_A - q^T\right) / f^T_{su}}{\left(q^T_A - q^T\right) / f^T_{su} + \left(q^T_A - q^T\right) / f^T_{su}} \]

(11h,i)

(also warranting \( t_r = t_q + t_m \)). For Case I, \( t_q = 0 \) (as implied in \( t_r = t_r \) with \( z_{AM} = z_{AT} \)). For Case II, the maximum extent of queue can be obtained as \( z_{AM} = w_s^T \cdot t_q \) or \( z_{AM} = w_m^T \cdot t_m \) (or also any of the previous additive expressions linking \( z_{AM} \) to \( z_{AT} \) as \( z_{AM} = z_{AT} + w_m^T \cdot t_q \)) and translated into vehicles as a reference queue \( n_{AM} = K_n^T \cdot z_{AM} \) (a “fictitious” queue with density \( K_n^T = K_n^T \)) or an extended “queue”

\[ \hat{n}_{AM} = \frac{K_n^T}{K_n^T} \cdot n_{AM} \]

(the better measure of the effective number of vehicles as a mixed state queue \( \hat{n}_r = \hat{n}_r + n_t \)

taking \( z_{AM} \) at \( t_q \), if not accepting the null restrained queue \( n_{T+q} = 0 \) (or \( n_t = 0 \), \( t_T \in T_m \) in the strict concept.

Again, the implied values of wave speeds can be obtained by using any interval \( T_1 \subseteq T_q \) as

\[ 0 \leq t_1 \leq t_q \text{ and segment } z_1 = w_c^T \cdot t_1 \text{ with } z_{AT} \leq z_1 \leq z_{AM} \text{ for wave speed } w_c^T, \text{ any interval } T_2 \subseteq T_q \text{ as} \]

\[ 0 \leq t_2 \leq t_q \text{ and segment } z_2 = w_s^T \cdot t_2 \text{ with } 0 \leq z_2 \leq z_{AT} \text{ for wave speed } w_s^T, \text{ and any interval } T_3 \subseteq T_m \text{ as} \]

\[ 0 \leq t_3 \leq t_m \text{ and segment } z_3 = w_m^T \cdot t_3 \text{ with } 0 \leq z_3 \leq z_{AM} \text{ for wave speed } w_m^T, \text{ from the known } z - t \text{ window inflow, outflow and traffic densities.} \]

Further, the above relations imply a (global) **consistency constraint on the wave speeds** (for Case II) in the next period \( T_n \) that can be summarized in
\[
\frac{w_s T_n + w_c T_n}{w_s - w_c} = \frac{K_n T_n - K_u T_n}{K_S - K_u T_n}
\] (13a)

The key to obtain the previously quoted surprising simplification from as \( t_\ell = t_q + t_m \) in the next period, with

\[\begin{align*}
K_n T_n &= K_n T \quad \text{and} \quad q_n T_n = q_n T \quad \text{(otherwise independent of T and } z_{AT}) \text{, or} \\
\frac{(q_A T_n - q_n T_n)/(f_{Qn} T_n + (q_A T_n - Q_A T_n)/(f_{Qn} T_n - f_{sn} T_n))}{(q_A T_n - q_n T_n)/(f_{Qn} T_n - (Q_A T_n - q_n T_n)/(f_{Qn} T_n))} &= \frac{f_{Qn} T_n - f_{sn} T_n}{f_{Qn} T_n - f_{sn} T_n} 
\end{align*}\] (13b)

The key to reduce (10e) and (11e) to \( t_\ell = t_q + t_m \). These relations are implicit in (12b,c) and comes from the

balancing on links for arrivals and departures along \( Z_{AM} \) in \( t_\ell = t_q + t_m \) and \( t_m \). For the next period \( T_n \), the

balancing in \( T_\ell \) delivers \( K_n T_n z_{AT} + K_u T_n (z_{AM} - z_{AT}) + Q_A T_n t_\ell = K_u T_n z_{AM} + q_{AT} T_n t_\ell \) or

\[\begin{align*}
(K_n T_n - K_u T_n) z_{AT} &= (q_A T_n - Q_A T_n) t_\ell \quad \text{and balancing in } T_m \text{ delivers } K_S T_n z_{AM} + Q_A T_n t_m = K_u T_n z_{AM} + q_{AT} T_n t_m \quad \text{or} \\
(K_S T_n - K_u T_n) z_{AM} &= (q_A T_n - Q_A T_n) t_m \quad \text{producing } \frac{K_u T_n - K_u T_n}{K_S T_n - K_u T_n} = \frac{t_q + t_m}{t_m} \quad \text{and then (13a) by using} \\
z_{AM} &= \frac{z_{AT}}{1 - w_c T_n / w_s T_n}, \quad t_q = \frac{z_{AM}}{w_s T_n} \quad \text{and} \quad t_m = \frac{z_{AM}}{w_m T_n} \quad \text{(otherwise 12b,c implies 13a,b for } t_\ell = t_q + t_m \). \text{ Note also} \\
\end{align*}\]

An alternative approach for relating vertical and horizontal queueing model estimates can also be

developed by defining a vertical queueing model variable for \( t_m \) using \( \tilde{t}_m = \frac{z_{AM}}{V_u T_n} = \frac{\tilde{n}_{AM}}{V_u T_n} K_n T_n \) (as a lower

bound of normalization time, since a normalization wave speed \( w_m = V_u T_n \) corresponds to \( V_s = V_u T_n \) in free

queue discharging, both upper limits in horizontal queueing models). Then the previous formula for \( \tilde{t}_q \) will be

better attributed to \( \tilde{t}_q \), delivering \( \tilde{t}_q = \tilde{t}_r - \tilde{t}_m \) at the end of the queue. Following the vertical queueing

model assumptions, \( \tilde{n}_{AM} = \tilde{n}_{AT} + (Q_n T_n - q_n T_n) \tilde{t}_q \) will then delivers \( \tilde{t}_m = \frac{\tilde{n}_{AT} + (Q_n T_n - q_n T_n) \tilde{t}_q}{(Q_n T_n - q_n T_n) + V_u T_n K_n T_n} \) and

\[\begin{align*}
\tilde{t}_q &= \frac{V_u T_n K_n T_n \tilde{t}_r - \tilde{n}_{AT}}{(Q_n T_n - q_n T_n) + V_u T_n K_n T_n} \quad \text{(meaning that } \tilde{t}_q < \tilde{t}_r \quad \text{and warranting } \tilde{t}_r = \tilde{t}_q + \tilde{t}_m \equiv t_\ell = t_q + t_m \text{ in the final} \\
\end{align*}\]

result). Nevertheless, any estimate of this variable based on a vertical queueing model is artificial and, in some

sense, fictitious (even if compatible to a limit condition of horizontal queueing models).
Finally, considering the previous discussion about flows relative to the moving frontiers of queues, further alternative horizontal queueing model formulas for \( t_q \) and \( t_m \) (and also \( t_i = t_q + t_m \) ) are

\[
t_q = \frac{\hat{n}_{AT}}{q_A^{Tn} - \hat{Q}_A^{Tn}}
\]

(10j)

\[
t_m = \frac{\hat{n}_{AM}}{q_M^{Tn} - \hat{Q}_M^{Tn}}
\]

(11j)

where \( \hat{q}_A^{Tn}, \hat{Q}_A^{Tn} \) are the flows departing from and arriving to the queue and \( \hat{q}_M^{Tn}, \hat{Q}_M^{Tn} \) are the flows departing from and arriving to the discharging “queue” (as seen by the moving front and back waves).

Noting that \( \hat{n}_{AT} = n_{AT} \) with \( n_{AT} = K_n^T z_{AT} \) and using \( \hat{q}_A^{Tn} = K_n^T (V_n^T + w_{Tn}^s) \hat{Q}_A^{Tn} = K_n^T (V_n^T + w_{cTn}^T) \), as seen from the queue operating conditions, it can be seen that formula (10j) reproduces formula (10a).

Alternatively, using \( \hat{Q}_A^{Tn} = K_u^{Tn} (V_u^n + w_{cTn}^T) = Q_A^{Tn} + K_u^{Tn} w_{cTn}^T \) for the inflow to initial queue and \( \hat{q}_A^{Tn} = K_s^{Tn} (V_s^T + w_{Tn}^s) = q_A^{Tn} + K_s^{Tn} w_{Tn}^s \) for the outflow from the initial queue, formula (10j) can be translated to

\[
t_q = \frac{n_{AT}}{(q_A^{Tn} - \hat{Q}_A^{Tn}) + (K_s^{Tn} w_{Tn}^s - K_u^{Tn} w_{cTn}^T)}
\]

(delivering the same result as implied by the flow conservation condition as observed traveling with the wave speeds).

The key to the related estimate of formula (11j) is (again) to understand that the (real) number of vehicles \( \hat{n}_{AM} = \hat{n}_{AM} \) between M and A has to be evaluated using \( K_s^{Tn} \) (the density in front of M with free queue discharging). Then, using \( \hat{n}_{AM} = K_S^{Tn} z_{AM} \) and \( \hat{q}_M^{Tn} = q_A^{Tn} = K_S^{Tn} V_{sTn}^T, \hat{Q}_M^{Tn} = K_S^{Tn} (V_{sTn}^T - w_{mTn}^T) \) (as the front of the discharging queue is fixed at A), it can be seen that formula (11j) reproduces formula (11a).

Alternatively using \( \hat{Q}_M^{Tn} = K_u^{Tn} (V_u^{Tn} - w_{mTn}^T) = Q_A^{Tn} - K_u^{Tn} w_{mTn}^T \) for the inflow to the discharging queue, the formula can be translated to

\[
t_m = \frac{\hat{n}_{AM}}{(q_A^{Tn} - \hat{Q}_A^{Tn}) + K_u^{Tn} w_{mTn}^T}
\]

(as there is no change on the static outflow condition). These formulas are, however, related to the artificial view of a queue between M and A at \( t_m \), that does not exist.

A major distinction for considering Cases I or II comes from the fact that queueing occurs at A during \( t_p \) for Case I, after T, but not for Case II, that has free discharging flow after T at A. If queueing produces some upstream blocking effect (as in converging or diverging junctions), the same distinction have to be considered at intermediate upstream points (similarly for free flow discharging). Queueing
at A will occur during \( T + t_p \) (and blocking time will decrease linearly in \( z \), as long as traffic variables remain constant, up to 0 with \( z_{AM} = z_{AT} \) for Case I but only during \( T \) for Case II (blocking time will also decrease linearly in \( z \), as long as traffic variables remain constant, up to 0 at \( z_{AM} > z_{AT} \). Similarly, free flow discharging time during \( t_f \) at A will decrease linearly in \( z \), as long as traffic variables remain constant, up to 0 with \( z_{AM} = z_{AT} \). In the same vein, at a given time \( t \), the average density between A and the end of the queue during its dissipation is \( K^T_n \) for Case I but not for Case II, at least after \( T \). During dissipation, for Case II, the limits of the queue can be obtained as

\[
z_t = z_{A(T+1)} - z_{St} \quad \text{where } z_{A(T+1)} \quad \text{and } z_{St} \quad \text{can be linearly interpolated as } z_{A(T+1)} = z_{AM} \frac{T + t}{T + t_q} \quad \text{and}
\]

\[
z_{St} = z_{AM} \frac{t}{t_q}, \quad \text{delivering the average queue density as } K_z = K^T_n \frac{z_t}{z_{A(T+1)}} + K^{Tn}_s \frac{z_{St}}{z_{A(T+1)}} \quad \text{in } z_{A(T+1)},
\]

for \( t \leq t_q \) (and \( K^{Tn}_s \) in \( z_{St} \) after \( t_q \), up to \( t_f \), corresponding to the free flow discharging zone).

Another variable can be introduced for representing maximum queueing time \( t_{mn} \) (that occurs at A for the pure cases), so as to permit a generic form for prediction of queue blocking time \( t_{nz} \) (with stopped queue) at a position \( z \) (also for free flow discharging time \( t_{Sz} \) using \( t_{ms} \)), assuming \( t_{nz} = 0 \) at \( z_m \) and linear variation while \( 0 < t < t_{mn} \) as

\[
t_{nz} = t_{mn} f_z \quad \quad (14)
\]

with \( f_z = 1 - z/z_m \), where \( z_m = z_{AT} \) and \( t_{mn} = T + t_p \) for Case I or \( z_m = z_{AM} \) and \( t_{mn} = T \) for Case II (also \( t_{Sz} = t_{ms} f_z \), with \( t_{ms} = 0 \) for Case I and \( t_{ms} = t_f \) for Case II, while \( T < t < T + t_f \)). Similarly, the average density between A and the end of queue (changing midway from \( K^T_n \) to \( K^{Tn}_s \)) at time \( t \) is

\[
K_z = K^T_n - (K^T_n - K^{Tn}_s) f_{Kt} \quad \quad (15)
\]

where \( f_{Kt} = \frac{t/(T + t)}{t_q/(T + t_q)} \) for Case II while \( 0 < t < t_q \) (otherwise \( f_{Kt} = 1 \) and \( K_z = K^T_n \) in the queue during \( T \) for Cases I or II and also \( 0 < t < t_p \) in Case I; the expressions assume no queueing when

\[
K_z = K^{Tn}_s \quad \text{for } t_q < t < t_f \quad \text{in Case II, during normalization, meaning that } q^{Tn}_n = S^{Tn}_n \quad \text{along the road}.
\]
Note that the previous discussion sticks to the idea that queue discharge freely from the section A (as when A is the primary bottleneck). As previously noted, most relations are valid when congestion is solved by the partial relief of the bottleneck at A (meaning that its flow increases but remains restrained by a harder bottleneck ahead; then A is a secondary bottleneck operating under forced flow conditions imposed by the primary, i.e. harder, bottleneck ahead). However, under this condition, $q_{A}^{T_n} < S_{A}^{T_n}$ in A (correspondingly, $V_{s}^{T_n} < V_{s}^{T_n}$ and $K_{s}^{T_n} > K_{s}^{T_n}$) and the relief condition is also the operation in the queue commanded by the primary bottleneck ahead. In the forced flow regime, some ad hoc relations are usually accepted for the average traffic conditions (usually a rough description of a widely unstable and varying regime of stop and go traffic) as the simplified linear relation $K_s = K_J - (K_J - K_s)(q_s/S)$, with $K_s = S/V_s$ for free queue discharging flow (i.e. $q_s = S$ with $V_s = V_s$), that implies $V_s = q_s/(K_s - (K_J - K_s)(q_s/S))$ for the corresponding speeds in the forced flow regime, a relation that predict the traffic conditions at A for forced flow with $q_s \leq S$ (i.e. for the queueing flow imposed by the primary bottleneck) and permits to conform the application of HQMs to the boundaries of road elements (a condition for their articulation with node models as well). As a trivial detail, note that the operation under forced flow adds another steady state traffic condition with $q = q_s$ (the other being the normal operation flowing the full demand with $q = Q$ along the link). Both steady state traffic conditions can be easily checked and should be treated directly (FOTMs and HQMs are intrinsically transient).

Finally, it is interesting to discuss the implications of the horizontal queueing model to obtain a better understanding valid and more general forms of the widely employed balancing conditions embodied in most model formulations. The simple form of balancing can not be applied to a section but to a link where feeding conditions are undisturbed by queueing constraints (i.e. with normal flow conditions at the entry section B that heads to the bottleneck section A, meaning that all the dynamics of queue formation and dissipation occurs inside the link length L). The balancing conditions applied to a link $a=(B,A)$ then implies that

$$n_{LT} = \int_{0}^{T} (q_{inl}[t] - q_{outl}[t])dt = \int_{0}^{T} (Q_{B}[t] - q_{A}[t])dt = \int_{0}^{T} (Q_{A}[t - \tau^* - \tau] - q_{\Lambda}[t])dt,$$

by assuming $n_{L0} = 0$ and using the (unknown up to the end of T) time translation of flow between B and A by the travel time $\tau^* [t] = L/\bar{V}_{L}[t]$ for vehicles passing A at time t and getting an average (endogenous) speed $\bar{V}_{L}[t]$ in travelling to B (meaning that $n_{LT} = \int_{T - \tau}^{T} Q_{A}[t]dt$, as well, for vehicles that entered link but have not passed A up to the end of T). Of course, $Q_{B}[t], q_{\Lambda}[t]$ (or $q_{inl}[t], q_{outl}[t]$) are not independent of traffic flow conditions (as related to several forms of metering or spillback effects) but this is not the point to be discussed.
here. Setting \( n_{LT} = x_{PLT} + n_{qLT} \), by splitting the total number of vehicles in queueing and non-queueing states as

\[ n_{qT}, x_{PT}, \text{ and restating } \tau^*[t] \text{ as } \tau^*[t] = (L - z[z^*[t]])/V_{sl} + z[z^*[t]]/V_{sT} = \tau^*_a[t] + \tau^*_T[t], \]

also splitting the travel time into the running (unconstrained) time and the time in queue, the implied relation for the simplest case (of queueing up to the stop line) is 

\[ n_{qLT} = \int_{T-c_a[t]}^{T} Q_A[t] \, dt \]

for queue, satisfying 

\[ n_{qLT} = K_n^T z_{AT} \]

(as \( x_{PLT} = \int_{T-c_a[t]}^{T} Q_A[t] \, dt \), satisfying \( x_{PLT} = K_n^T (L - z_T) \), for arriving vehicles before queueing). However, the distinction of Cases I and II points to the need of an even more detailed description as 

\[ n_{LT} = x_{PT} + n_{qLT} + n_{sLT} \]

by also distinguishing vehicles leaving the queue \( n_{sT} \) but yet heading to A (usually as free queue discharging flow), then splitting \( \tau^*[t] \) as 

\[ \tau^*[t] = (L - z[z^*[t]])/V_{sl} + (z_b[z^*[t]] - z_l[z_l^*[t]])/V_{sT} + z_l[z_l^*[t]]/V_{sT} = \tau^*_a[t] + \tau^*_T[t] \]

where 

\[ \tau^*_T[t] = \tau^*_T[t] + \tau^*_s[t] \]

(on the extended queue) and implying 

\[ n_{qLT} = \int_{T-c_a[t]}^{T} Q_A[t] \, dt \]

for queue, satisfying 

\[ n_{qLT} = K_n^T z_{AT}, \text{ and } n_{sLT} = \int_{T-c_a[t]}^{T} Q_A[t] \, dt \]

for the free discharging queue, satisfying 

\[ n_{sLT} = K_s^T z_{AT} \]

(as \( x_{PLT} = \int_{T-c_a[t]}^{T} Q_A[t] \, dt \), satisfying \( x_{PLT} = K_n^T (L - z_T) \) with \( z_T = z_{nT} + z_{sT} \), for arriving vehicles before queueing). Note also that variables marked with asterisks (*) and the profile of the back and front of the queue are solutions to the overall problem analysis (complicating their evaluation). Nevertheless, the treatment of a general time-varying setting is not considered here, meaning that traffic conditions are taken as uniform in each time period (only the translation of time periods remain an additional complicating problem to be dealt ahead).

In any case, all these relations are clearly distinct from (and more complex than) the common form of flow balancing condition applied when using the vertical queueing model as

\[ \tilde{n}_{AT} = \tilde{n}_{QT} = \int_0^T (q_{inA}[t] - q_{outA}[t]) \, dt = \int_0^T (Q_A[t] - q_A[t]) \, dt \]

(assuming \( \tilde{n}_{q0} = 0 \)). A more general relation, corresponding to the horizontal queueing model, is 

\[ n_{AT} = n_{qT} = \int_0^T (Q_A[t] - q_A[t]) \, dt \]

using point measures but having to track the back and front of queue (as discussed before). These relations are needed if trying to develop queue estimates from the input-output approach (e.g. as advocated by Daganzo and his partners, from Daganzo, 1983), including further understanding of queueing patterns (as Cases I or II). Without further adjustments, the conventional form of input-output analysis is not sufficient for developing consistent measures of queueing and queueing delay (at least when there are other sources of delay on road elements).
3. PRACTICES IN THE FORMULATION OF HORIZONTAL QUEUEING MODELS

Summarizing what was presented up to this point (see Table 1), it was shown that horizontal queueing models can be derived from four fully equivalent ways (that deliver the same result under similar conditions):

A. by tracking shock waves based on first order macroscopic traffic flow analysis models;
B. by introducing corrective factors into vertical queueing model estimates as \( n_{At} \), \( n_{St} \), \( \bar{n}_{St} \), and \( \bar{n}_{Mt} \);
C. by using flow balancing conditions on fixed road segments taking into account their traffic densities;
D. by using conventional flow balancing conditions for moving queues as seen at shock wave frontiers;
(from which the method B will be taken as the simpler version in the next section).

It was shown that there is no inconsistence between conventional (vertical) queueing analysis and traditional (horizontal) shock-wave analysis but those generated by the differences between vertical and horizontal queueing model implicit assumptions. Of course, critics to first order traffic flow analysis models can request other model formulations but, as long as the new models represent the space requirements of vehicles, it is expected that the same equivalence relation can be found for the more general traffic flow models, under similar assumptions.

All these versions require data on traffic densities, speeds and/or flows during queue discharging (some require data on wave speeds). Apparently, no better approach can be developed using these alternative views. Formulas (12a,c) and (10d,e) will be the preferred options used to develop approximate formulas for queueing at intersections in the next section (formulas 12d,e and 10f,g,i are also simple to apply).

The relation between model estimates can be checked in Table A1, A2, and A3 (Appendix A) for representative parameters of approaches at intersections (the approximate formulas of the next section will have to adopt assumptions that can deliver the approximate values of wave speeds so as to insert the appropriate correction; fortunately, this is a viable option for the usual situation of approaches at intersections, as exemplified by the assumptions taken ahead).

Further detailing on the corrective method B is summarized in Table 2. Shortly, the same steps can be carried-out using alternative concepts and measures for HQMs: the reference queue \( n_{At} \) (an equivalent queue of vehicles filling the same queue extent from A, assuming an uniform queueing density \( K_n^T \)), the restrained queue \( n_{At} \) (the effective number of vehicles queueing, i.e. either stopped or flowing with \( q_n^T < S_A^T \)), or the extended queue \( n_{At} \) (the effective number of vehicles filling the queue extent from A, as a mixed queue \( \hat{n}_{At} = \hat{n}_{At} + n_t \), i.e. either as restrained queue \( n_t \) or as free queue discharging flow “queue” \( \hat{n}_{At} \)). In Table 2, dissipation and normalization times also are expressed with different formulas but just for presenting alternative options (not related to queueing measures and that can be freely interchanged as desired, all derived from 10a and 11a).
PERIOD \ METHOD A: wave tracking B: correction formulas C: fixed segment D: moving queue

Peak Period (T): $Q_A^T > q_A^T$
-queue build-up

$w_c = \frac{Q_A^T - q_A^T}{K_n - K_u}, K_u = \frac{Q_A^T}{V_u}, K_n = \frac{N_T}{\varepsilon_n}$
$w_c = \frac{Q_A^T - q_A^T}{V_u} \cdot T_n$

$T_n = \frac{Q_A^T}{V_u} \cdot T_n$
$\hat{n}_A = n_A = \frac{\tilde{n}_A}{f_{Qn}}$
$z_A = \frac{w_c}{T_n} \cdot T_n$
$\hat{n}_A = n_A = \frac{\tilde{n}_A}{f_{Qn}}$
$z_A = w_c \cdot T_n$

$T_n = \frac{Q_A^T}{V_u} \cdot T_n$
$\hat{n}_A = n_A = \frac{\tilde{n}_A}{f_{Qn}}$
$z_A = w_c \cdot T_n$

Next Period (Tn): $Q_A^T < q_A^T$
-queue dissipation (Tq in Tn)

$t_q = \frac{z_A}{w_s - w_c} = \frac{AM}{W_s - W_c}$
$z_A = w_s \cdot t_q = \frac{AM}{W_s}$
$n_A = K_u \cdot z_A$
$\hat{n}_A = n_A = \tilde{n}_A$

$z_A = w_s \cdot t_q = \frac{AM}{W_s}$
$n_A = n_A = \tilde{n}_A$
$z_A = \frac{n_A}{N_f} \cdot \varepsilon_n$

-traffic recovery (Tp/Tm in Tn)

if $q_A^T < q_n$ or after $t_q$

$w_r = \frac{Q_A^T - Q_A^T}{K_S - K_u} \cdot f_{Qn} = l - \frac{Q_A^T}{V_u}$
$z_A = \frac{Q_A^T}{V_u} \cdot f_{Qn} = \frac{Q_A^T}{V_u}$
$\hat{n}_A = \frac{\tilde{n}_A}{f_{Qn}}$
$K_u = K_u \cdot z_A$

$z_A = \frac{Q_A^T}{V_u} \cdot f_{Qn}$
$\hat{n}_A = \frac{\tilde{n}_A}{f_{Qn}}$
$z_A = \frac{n_A}{N_f} \cdot \varepsilon_n$

Summary of Formulas Derived from the (Equivalent) Alternative Approaches for Developing Horizontal (Spatial) Queueing Models.
### Table 2 – Summary of (Equivalent) Alternative Corrective Formulas (Method B) Derived from the Horizontal (Spatial) Queuing Model.

<table>
<thead>
<tr>
<th>PERIOD</th>
<th>METHOD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Peak Period (T): $Q_A^T &gt; q_A^T$</td>
<td>VQM ($\bar{n}$)</td>
</tr>
<tr>
<td>$f_{Tn}^q = 1 - \frac{Q_A^T}{V_n} \frac{T_n}{\bar{n}}$</td>
<td>$\bar{n}_A = (Q_A^T - q_A^T) t, t \in T$</td>
</tr>
<tr>
<td>$\bar{z}<em>{AT} = \frac{\bar{n}</em>{AT} \bar{z}_n}{N_f}$</td>
<td>$z_{AT} = \frac{\bar{n}_{AT} \bar{z}_n}{N_f}$</td>
</tr>
</tbody>
</table>

Next Period (Tn): $Q_A^T < q_A^T$, $q_n^T = q_A^T$ (recovery time $t_p$; dissipation time $t_q = 0$) | Queue relief (Tp in Tn) | $f_{Tn}^q = 1 - \frac{Q_A^T}{V_n} \frac{T_n}{\bar{n}}$ |
| $\bar{n}_R = (Q_A^T - q_A^T) t, t \in Tn$ | $\bar{n}_{AT+1} = \bar{n}_{AT} - \bar{n}_R, t \in Tn$ | $\bar{n}_{AT+1} = \bar{n}_{AT} - \bar{n}_R, t \in Tn$ | $\bar{n}_{AT+1} = \bar{n}_{AT} - \bar{n}_R, t \in Tn$ | $\bar{n}_{AT+1} = \bar{n}_{AT} - \bar{n}_R, t \in Tn$ |
| $\bar{z}_{AM} = \frac{2}{Tn} $, $t_\phi = \frac{\bar{n}_{AM}}{q_A^T - Q_A^T}$ | $\bar{n}_{AM} = \frac{q_A^T - Q_A^T}{q_A^T - q_A^T}$ | $\bar{n}_{AM} = \frac{q_A^T - Q_A^T}{q_A^T - q_A^T}$ | $\bar{n}_{AM} = \frac{q_A^T - Q_A^T}{q_A^T - q_A^T}$ |
| $n_{AT+1} = n_{AT} + n_{AT} - \bar{n}_R, t \in Tn$ | $t_q = \bar{t}_q = \frac{\bar{n}_{AT}}{q_A^T - Q_A^T}$ | $t_q = \bar{t}_q = \frac{\bar{n}_{AT}}{q_A^T - Q_A^T}$ | $t_q = \bar{t}_q = \frac{\bar{n}_{AT}}{q_A^T - Q_A^T}$ |

Next Period (Tn): $Q_A^T < q_A^T$, $q_n^T < q_A^T$ (otherwise $t_q = 0$) | Queue dissipation (Tq in Tn) | $f_{Tn}^q = 1 - \frac{Q_A^T}{V_n} \frac{T_n}{\bar{n}}$ |
| $f_{Tn}^q = 1 - \frac{Q_A^T}{V_n} \frac{T_n}{\bar{n}}$ | $\bar{n}_{AT} = \frac{q_A^T - Q_A^T}{q_A^T} t, t \in Tn$ | $n_{AT+1} = n_{AT} + n_{AT} - \bar{n}_R, t \in Tn$ | $\bar{n}_{AM} = \frac{q_A^T - Q_A^T}{q_A^T} t, t \in Tn$ |
| $\bar{z}_{AM} = \frac{\bar{n}_{AM} \bar{z}_n}{N_f}$, $t_\phi = \frac{\bar{n}_{AM}}{q_A^T - Q_A^T}$ | $n_{AM} = n_{AT} + n_{AT} - \bar{n}_R, t \in Tn$ | $n_{AM} = n_{AT} + n_{AT} - \bar{n}_R, t \in Tn$ | $\bar{n}_{AM} = \frac{q_A^T - Q_A^T}{q_A^T} t, t \in Tn$ |

| Summary of (Equivalent) Alternative Corrective Formulas (Method B) Derived from the Horizontal (Spatial) Queuing Model. |
Note also that several mixed cases can occur. But the HQM relations can always be developed from FOTM shockwaves by noting that the numerator of its speed represents a flow balancing and that conversion factors can be used to transform speeds in conventional space/time units into an implied speed in vehicle positions for an implicit density, at least for the neighboring densities (one for the length being reduced and other for that being increased) but also for any other convenient density (e.g. a referring density), that produces a correct accounting in vehicles. The clear example is that of the dissipation wave of Case II for which its wave speed \( w_{Ts}^{Tn} = -\frac{q_{S}^{Tn} - q_{n}^{T}}{K_{S}^{Tn} - K_{n}^{T}} \) that produces the extent of dissipation \( z_{St} = w_{Ts}^{Tn}.t \) can be translated into the number of vehicles taken from the queue as \( \tilde{n}_{At} = K_{n}^{T} \cdot z_{St} = \frac{\tilde{n}_{St}}{f_{Ts}^{Tn}} \) or the number of vehicles added to the free queue flow discharging length as \( \tilde{n}_{St} = K_{S}^{T} \cdot z_{St} = \frac{\tilde{n}_{St}}{f_{Ts}^{Tn}} \), with the alternative denominators as correction factors.

For the main mixed case of Case II, when both conditions occur, the queue length reduces simultaneously at both ends (forward for \( w_{r}^{Tn} \) and backward for \( w_{s}^{Tn} \)) and dissipate in an intermediate position \( Z_{AT} = \frac{Z_{AT}}{w_{r}^{Tn} + w_{s}^{Tn}} \) at an intermediate position \( Z_{i} = \frac{w_{s}^{Tn}}{w_{r}^{Tn} + w_{s}^{Tn}} \) from A (again using a switched part ratio of speeds). The normalization wave follows again, with speed \( w_{m}^{Tn} = \frac{Q_{n}^{Tn} - q_{n}^{T}}{K_{n}^{Tn} - K_{S}^{Tn}} \) (forward), as before, and normalization time \( t_{i} = \frac{Z_{i}}{w_{m}^{Tn}} = \frac{Z_{AT}}{w_{m}^{Tn}} \left( I + \frac{w_{r}^{Tn}}{w_{s}^{Tn}} \right) \) from the intermediate point (as implied by using...
\( w_{e}^{T_{n}} = -w_{r}^{T_{n}} \) for \( t_{m} \) in Case II). These results can be treated similarly for Case I because \( w_{s}^{T_{n}} = 0 \) of this pure case implies \( z_{i} = 0 \) for queue dissipation at A (consistent with this main mixed case). Other (secondary) mixed cases have major influence in intermediate states. First, a change in the front of the queue occurs in Case I if \( q_{A}^{T_{n}} = q_{A}^{T} \) can not be assumed, and then there will be a change in the queueing flow too (reducing the queue discharge flow if \( q_{A}^{T_{n}} < q_{A}^{T} \) otherwise increasing the queue discharge flow and resulting in the previous mix with Case II as \( q_{A}^{T_{n}} > q_{A}^{T} \), propagated by an internal wave in the queue with speed \( w_{q}^{T_{n}} = \frac{q_{A}^{T_{n}} - q_{A}^{T}}{K_{q}^{T_{n}} - K_{q}^{T}} \) (backward), transmitting this harder condition and finally changing \( w_{c}^{T_{n}} \) when the back of queue were reached.

The new wave speed \( w_{e}^{T_{n}} \) can propagate forward as long as \( Q_{A}^{T_{n}} < q_{A}^{T_{n}} \) applies (even if A keeps its role of a capacity bottleneck and \( w_{s}^{T_{n}} = 0 \) applies at A) and producing very similar patterns (otherwise a mixed peak period is observed as queues keep propagating backward when the new wave speed reaches the end of queue).

Second, another mixed case can occur if \( q_{A}^{T} < q_{A}^{T_{n}} \) but with \( Q_{A}^{T_{n}} > q_{A}^{T_{n}} \) as the flow discharging increases and transmits backward along the queue but the changes in queue discharging can not be surely translated in the normalization of the traffic operation. The new traffic condition with the change \( q_{n}^{T_{n}} = q_{A}^{T_{n}} \) can be very similar to normal operation (e.g. for \( q_{s}^{T_{n}} = S_{A}^{T_{n}} \)), instead of being described as congested operation, but the transmission wave can be slower than the congestion wave (if faster, a pattern similar to Case II results).

Otherwise the over-demand period will be extended a bit more, even if the degree of over-demand can decrease with \( Q_{A}^{T_{n}} < Q_{A}^{T} \) and \( q_{A}^{T} < q_{A}^{T_{n}} \) in the reduced congestion. Ignoring these intermediate and internal queueing states seems to be an acceptable simplification in most situations. Practically it can be assumed that \( z_{AM} = z_{AT} \) and \( t_{q} = 0 \) in all these particular mixed cases, distinguishing only Case II. Otherwise, at least when the end of queue keeps propagating backward, a multiple period analysis has to be carried-out by recognizing queue evolution and considering pre-existing queues.

3.1 Time Translation of Periods

Note that all the previous discussion remains on the supposition that time periods are similarly defined and observed along the corridor or segment that brings to A (the bottleneck). However, this supposition is inconsistent to the fact that finite speeds mean that events observed at some distance from A but inferred from traffic conditions referred to some period in A have to be adjusted backward in time if occurring upstream. As example, to be correct, the balancing on segments requires the time translation of periods by the unconstrained
travel times in $z_{AT}$ and/or $z_{AM}$, and, to be exact, it requires that unconstrained speeds are constant and uniform along periods and along the segment (what is the requisite for a “degenerate” shock wave with $w = \frac{\Delta q}{\Delta K} = V_u$ between subsequent unconstrained traffic periods, travelling at the unconstrained vehicle speeds, that preserves the duration and traffic condition of time periods $T$ and $T_n$, along the arrival link before queueing).

Under the assumption of constant and uniform unconstrained speeds, some further thoughts on the implications of the horizontal (spatial) queueing models can be developed. Then the time translation of periods by the unconstrained travel time is legitimate because, under these assumptions, time periods are similarly defined all along the corridor (after time translation) and no net flow of vehicles crosses the “degenerate” shock wave frontier between periods (corresponding to a fictitious limiting vehicle that travels at the instant that divides periods $T$ and $T_n$ with the common uniform unconstrained speed outside queueing). At least conceptually, the time translation of periods in this setting should deliver simple formulas, as periods delimit the same set of vehicles before and after queueing (as in vertical queueing models, warranting the similarity in some results as exhibited by input-output methods). This feature was previously noted (e.g. as in Hurdle, Son, 2001) but its interaction with the extent of physical queues was not properly represented in closed formulas.

In other conditions, the duration of time periods along the corridor also changes and the net flow through the frontier of periods has to be considered (as in the moving frames that track other waves). Frontiers of time periods can be defined by unconstrained speeds (or free flow speeds) or by some endogenous traffic speed (as the average traffic speed) or flow balancing (as a zero net flow) condition, but this feature can bring some analytical burden to any method (favoring numerical alternatives).

The time translation of periods means that data measured at section A can not be assumed for the same periods in other sections along the corridor. With uniform unconstrained arrival speeds, the frontier of time periods propagates with shock speed $w_T = w_{Tn} = V_u^T = Y_{Tn}$ and produces time translated periods with uniform duration $T$ and $T_n$ along the corridor, as shown in Figure 1c (otherwise, yet in a simplified view with uniform flow periods, unconstrained arrivals produce time periods with varying duration). But even for periods with general uniform pattern along the corridor, the time translation means that wave speeds generated at A (as congestion and dissipation waves) would meet the frontiers of periods in advance at the back of queue. As example (shown in Figure 1b), the congestion wave of period $T$ will experience the change of traffic conditions when the arrival wave is met at a distance $z_{AW} = w_c^T \delta_T = w_{Tn} \varepsilon_T$ with $\delta_T + \varepsilon_T = T$, that reduces to

$$\delta_T = T \cdot \frac{V_u}{w_c^T + V_u} \quad \text{and} \quad \varepsilon_T = T \cdot \frac{w_c^T}{w_c^T + V_u} \quad \text{(with the switched part ratio of speeds rule), or} \quad z_{AW} = \frac{w_c^T \cdot V_u}{w_c^T + V_u} \cdot T$$

for the queue at the instant when traffic conditions change. For Case I, queue recovery starts after $\delta_T$, instead
of $T$, from $z_{AW}$, instead of $z_{AT}$, predicting faster traffic recovery as $t_r = \frac{Z_{AW}}{W_c} - \varepsilon_T$ (by starting before). For
2 Case II, change of traffic condition delivers $z_{AT} = w_c^T \cdot \delta_T + w_c^{Tn} \cdot \varepsilon_T$, instead of $z_{AT} = w_c^T \cdot T$, a reduced
3 estimate if $Q_A^{Tn} < Q_A^T$ after the peak period, and $z_{AW} = \frac{w_c^T \cdot v_u}{w_c^T + v_u} \cdot T$ is again a better measure of queueing
4 produced during the peak period, but $t_q = \frac{Z_{AW} + w_c^{Tn} \cdot \varepsilon_T}{w_s^{Tn} - w_c^{Tn}} \cdot T$ (with $Z_{AT} \equiv Z_{AW} + w_c^{Tn} \cdot \varepsilon_T$ in general, warranting
5 similar values for $z_{AM}$ and $t_m$). Previous estimates can be assumed as an approximation for $w_c^T << v_u$
6 (meaning $\varepsilon_T << T$ for $\varepsilon_T \equiv 0$), the implied supposition accepted in the formulas developed before.
7 A similar effect (but opposite as the previous discussion applies to backward waves) occurs if a forward
8 wave (e.g. a recovery wave) faces a change of time period and can be evaluated by noting that the change of
9 traffic condition after $T$ as measured at a static section will be delayed to $\delta_T = T + \varepsilon_T$ at the back of the
10 forward wave with $w_r^T(T + \varepsilon_T) = V_u \cdot \varepsilon_T$, delivering $\varepsilon_T = \frac{w_r^T}{V_u - w_r^T} \cdot T$ for the additional time before the
11 change as seen at the back of the forward wave and then $\delta_T = \frac{V_u}{V_u - w_r^T} \cdot T$ is the total propagation time.
12 Both effects can be naturally interpreted as an adjustment of time periods as seen by wave shocks (the
13 practical approach adopted ahead). In general, formulas become more involved when embodying the effect of
14 the time translation of periods using the same principles of the previous development in any form (after the
15 lengthy discussion devoted to confirm the equivalence of the alternative approaches). However, taking again
16 the approach based on the balancing on segments (as discussed before), the previous analysis would be restated
17 as $(K_n^T - K_u^T)z_{AW} = (Q_A^T - q_A^T)\delta_T$ (that reproduces the previous expression for $z_{AW}$) and
18 $(K_n^T - K_u^T)z_{AW} = \left((q_A^T - Q_A^T)\varepsilon_T + (q_{A}^{Tn} - Q_A^{Tn})t_r\right)$ (where $\delta_T + \varepsilon_T = T$ with $\delta_T = T - \frac{V_u}{w_r^T + V_u}$ and
19 $\varepsilon_T = T - \frac{w_c^T}{w_c^T + V_u}$ as above). Using both and $\varepsilon_T = \frac{Z_{AW}}{V_u}$, implying $Q_A^T \cdot \varepsilon_T = K_u^T \cdot z_{AW}$ and $Q_A^{Tn} \cdot \varepsilon_T = K_u^{Tn} \cdot z_{AW}$,
20 delivers $(K_n^T - K_u^T)z_{AW} = (K_n^T - K_u^{Tn})z_{AW} = (Q_A^T - q_A^T)\delta_T - (q_{A}^{Tn} - Q_A^{Tn})\varepsilon_T - (q_A^T - Q_A^{Tn})t_r$ and
21 $Q_A^{Tn} \cdot \varepsilon_T - Q_A^T \cdot \varepsilon_T = (Q_A^T - q_A^T)\delta_T - (q_{A}^{Tn} - Q_A^{Tn})\varepsilon_T - (q_A^T - Q_A^{Tn})t_r$, as $K_u^{Tn} \cdot z_{AW} - K_u^T \cdot z_{AW} = Q_A^{Tn} \cdot \varepsilon_T - Q_A^T \cdot \varepsilon_T$,
that simplifies to $t_\epsilon = \tilde{i}_s$ (where $\tilde{i}_s = \frac{Q^T_A - q^T_A}{Q^T_A - q^T_A} \cdot T$), instead of $t_\epsilon = \frac{K_n^T - K^n_T}{K^T_\ell - K^T_\ell} \cdot \tilde{i}_s$, as previously obtained (meaning that the previous correction was the sole effect of ignoring the time translation of periods). Then, forcing the agreement of HQM dissipation time to the VQM estimate can improve its precision. Corresponding formulas can be obtained for $z_{AT}$, $z_{AM}$, $t_q$ and $t_m$ or their values can be recovered using $t_q = \frac{w^T_m}{w^T_m + w^T_n} \cdot t_\epsilon$

and $t_m = \frac{w^T_n}{w^T_m + w^T_m} \cdot t_\epsilon$ or $t_q = \frac{Z_{AM}}{W^T_n}$ and $t_m = \frac{Z_{AM}}{W^T_n}$ from $z_{AT} = z_{AW} + w^T_c \cdot s \cdot q$ and $z_{AM} = \frac{Z_{AT}}{1 - w^T_c / w^T_n}$ (as before). A generic procedure can be applied by setting $t_q = 0$ and $K^T_n = K^T_s$ in Case I (similarly, generic procedures as those described for intermediate queues and blocking time can be adopted too), even recognizing that this last condition is not queueing (as a basic difference between Cases I and II).

These computations remit to the general application of the Reynolds Transport Theorem to the (one dimensional) road segment (actually, Reynolds Transport Theorem is more general by admitting the use of a control volume whose frontiers move with known speeds).

However, the more usual method for considering the time translation of periods is implicit in another procedure. The alternative computation remits to the application of the Gradient Theorem to the Moskowitz cumulative counting function $N[z,t]$ for the (one dimensional) road segment, acting as a potential function on space-time frame with $k = \partial_z N[z,t]$ and $q = \partial_t N[z,t]$ (treated as a continuous functions) and warranting independence on “integration paths” of $k[z,t]$ and $q[z,t]$ in space-time (related to the application of the Green Theorem as highlighted in Daganzo, 1997, pg.100, and extensively applied in Yperman, 2007, and Gentile, 2010, among others). For example, under the simplifying assumptions that generate well defined time periods along the segments, the approach can be applied to a space-time window delimited by the frontiers of periods (i.e. travelling with the shock speeds $w = w^T_n = V^T_u = V^T_n$) between any pair of sections, usually based on the bottleneck section A.

The point made here is that the method has a physical basis. By construction, the Moskowitz function is able to recover the average traffic density from A to the extent $z$ at any instant $t$ as $K_n, z = N[z,t] - N[0,t]$ or the average traffic flow in a period $T$ at any section $z$ as $q^T, T = N[z,T] - N[z,0]$ (using the convention of Figure 1, it is an increasing function of both $z$ and $t$). As a potential function, the Moskowitz function can be used to balance densities and flows in the space-time diagram $[z,t]$ in orthogonal movements as $N[z,0] - N[0,T] = q^T, T + K_{zT}, z = K_{z0}, z + q^T, T$ as evaluated observing flow at $z$ along $T$ and then adding
vehicles from \( z \) to \( A \) at the end of period \( T \) or counting vehicles along \( z \) at time 0 and then observing flow at \( A \) along \( T \) (or any other “integration path” in the same space-time diagram window). The temporal “movement” is the conventional observation of flow in a static section. The spatial “movement” can be seen as an instantaneous change of position that respects the conservation of vehicles (i.e., counting the real number of vehicles between sections, even if seen as movements with “infinite” speeds and flows, as obtained by using spatial densities). The oblique integration paths are seen as relative movements and relative flows, counting vehicles as usual. Note also that there are no net overtaking for paths “traveling” at the average traffic speed (an average vehicle) and then the cumulative counting number is constant (e.g. meaning that \( N[z_{AW},-\varepsilon_T] = N[0,0] \), as both pair of values are on the travel path with travel time \( \varepsilon_T = \frac{Z_{AW}}{V_u} \) of the vehicle arriving in \( A \) at \( t=0 \)). For queueing vehicles (traveling with reduced speeds \( V_n < V_u \) during part of their path, at least), the average speed depends on the arrival time and on the exit (or travel and queueing) time (that can vary for each vehicle).

The translation of time periods can be treated as the use of general integration paths in the space time diagram but evaluated in a convenient alternative “integration path” delimited by period transition waves (a particular oblique “path” or its orthogonal components). Again, revising the approach based on the balancing on segments as example, observation on the parallelogram window \([z_{AW}, T]\) would be translated into

\[
N[0,T] - N[z_{AW}, -\varepsilon_T] = Q_A^T \varepsilon_T - K_u^T z_{AW} + q_A^T T
\]

(also \( N[0,T] - N[z_{AW}, -\varepsilon_T] = N[0,T] - N[0,0] = q_A^T T \) as \( Q_A^T \varepsilon_T = K_u^T z_{AW} \) for \( \varepsilon_T = \frac{Z_{AW}}{V_u} \), along the path of the vehicle arriving in \( A \) at \( t=0 \) and

\[
N[0,T] - N[z_{AW}, -\varepsilon_T] = -Q_A^T T + K_n^T z_{AW} + q_A^T T
\]

(then implying \((K_n^T - K_u^T)z_{AW} = (Q_A^T - q_A^T)\varepsilon_T \) and obtaining \( z_{AW} = \frac{(Q_A^T - q_A^T)}{(K_n^T - K_u^T)} \delta_T \) as before). Similarly the observation on the trapezoidal window \([z_{AM}, t_f]\) would be translated into

\[
N[0,T+t_f] - N[z_{AM}, T-\varepsilon_T] = -K_u^T (z_{AM} - z_{AW}) - K_n^T z_{AW} + q_A^T \varepsilon_T + q_A^T t_f
\]

and

\[
N[0,T+t_f] - N[z_{AM}, T-\varepsilon_T] = Q_A^T (\varepsilon_T + t_f) - K_u^T z_{AM}
\]

(then implying again

\[
(K_u^T - K_n^T)z_{AW} = (q_A^T - Q_A^T)\varepsilon_T + (q_A^T - Q_A^T)t_f \]

as before). The same revised results are then obtained.

Note that the convention of orienting \( z \) backwards (against the traffic flow direction) produces a negative sign for the contribution of densities when moving \( N \) with the traffic (a counterintuitive but correct sign, obtained as negative flows). Note also that moving with vehicles in a general setting (keeping \( N \) constant, as extensively done in Yperman, 2007, and Gentile, 2010) would require the calculation of intermediate values for vehicle trajectories. As example, for the vehicle leaving \( A \) at \( t=T \), \( N[0,T] = N[z_k, t_k] \) such that the arrival...
to the queue of vehicle \( k \) has \( z_k = w_c^T t_k \) and satisfies \( \frac{z_k}{w_c^T} + \frac{z_k}{V_n} = T \) (this condition, again after the use of a
switched part ratio of speeds rule, produces \( t_k = \frac{V_n}{w_c^T + V_n} T \) and \( z_k = \frac{w_c^T V_n}{w_c^T + V_n} T \) for \( N \) constant along a
trajectory without overtaking to \( A \) at \( t=T \), that can also be referred to the entry on a link but only by considering
the length of the link and exit conditions afterwards as determined by the overall traffic conditions after entry
time). And moving along wave frontiers would produce the relative flow as seen by the moving observer (as
practiced before), obtained along any oblique movement (considering its implied observer speed) or its
orthogonal components (i.e. for an observer in an oblique path with speed \( v_o \), travelling \( z_o = v_o \cdot t_o \) during an
observation time \( t_o \), \( q \cdot t_o \pm K \cdot z_o = K \cdot (V \pm v_o) t_o = \hat{q} \cdot t_o \), obtained with the relative flow \( \hat{q} = K \cdot (V \pm v_o) \);
along a wave frontier, the wave speed warrants that the relative flow is the same with traffic conditions on any
side of the wave). It is clear that, despite the reference to parallelogram or trapezoidal space-time windows, the
method implements a physical procedure for evaluating the Moskowitz function \( N[z, t] \) that reduces to
accumulating the relative flows between any two states in space-time (positive along or negative against axes;
even the use of densities amounts to evaluating consistent relative flows, when infinite relative speeds make
them indeterminate).

These theoretical remarks on the time translation of periods mix, however, with practical details usually
left unnoticed when measuring traffic flow at different sections and translating results between them (usually
done ignoring the time translation). Suppose that traffic demand before the bottleneck is measured at a section
just before the usual stretch of road with queueing conditions and the gathered traffic data is used to define
analysis periods. The simplest approach for the time translation of periods is exemplified in Hurdle, Son, 2001,
by offsetting time periods by the undisturbed travel time \( \tau_0 = L/V_{FL} \) as commanded by the implicit assumption
of an uniform arrival speed \( V_u = V_{FL} \) but being ineffective for evaluation of queues for intermediate queue
extents \( z < L \) (a variable to be obtained). The use of the data at the measured section contains an implicit time
translation that fits the current traffic situation (particularly the queue extent \( z < L \) ) but can change in new
traffic situations. However, for the evaluation of queues, the effective correction have to be consistent with the
queue extent as \( \tau_z = z/V_{FL} \) (even accepting \( V_u = V_{FL} \) ) A more general approach seems to be needed (at least
for further evaluating its practical relevance). Here some general comments on the implications and difficulties
found in this generalized setting were advanced above.

Note that the same kind of change on the traffic condition faced at the end of the queue will also be
commanded with other approaches (meaning that the acceptance of the approximation for \( w_c^T << V_u \) is very
convenient, if defendable). A correction based on the more detailed view can, perhaps, be specific. When analyzing discontinuous/interrupted flow (e.g. considering the effect of interruption cycles), a more radical simplification can be adopted by arguing that ignoring the effects of platoon dispersion along a corridor in discontinuous/interrupted traffic flow is not usually acceptable (only the assumption of unconstrained speeds independent of arrival demand for free platoons would be defendable but mixed with some effect of randomness). For continuous/uninterrupted flow, arguments would be switched (platoon dispersion can perhaps be ignored but unconstrained speeds usually depend on arrival demand), asking for alternative assumptions.

3.2 Queueing with Multiple Periods

Returning to the simplified view (ignoring changes of pattern and time translation along the road), periods for which pre-existing queues are present (the queue at the end of the previous period) have to be carefully considered. This is easy if avoiding a mix of inconsistent values: a field measure for the initial queue is a value consistent to a horizontal queueing model, not to a vertical queueing model (i.e. a $n_0$, not a $\bar{n}_0$). This feature is present in the previous discussion and must be applied when composing the queue evolution in the peak period and the next period. Avoiding (or correcting for) mixed values, no new problem arises from the presence of a pre-existing queue in the horizontal queueing model as can be seen in multi-period formulas as

$$n_{AT} = \sum_{j} \Delta \bar{n}_{At} = \sum_{j} \left( Q_{At}^{T} - q_{At}^{T} \right) T_{j} \geq \sum_{j} \frac{(Q_{At}^{T} - q_{At}^{T}) T_{j}}{1-K_{u}^{T}/K_{n}^{T}} \geq \sum_{j} \frac{(Q_{At}^{T} / V_{u}^{T}) / (N_{f} / f_{y})}{1-(Q_{At}^{T} / V_{u}^{T}) / (N_{f} / f_{y})} \quad (16a,b,c)$$

(where the time period $T$ was divided into sub-periods with $T = \sum_{j} T_{j}$, starting from a period for which $n_{0} = 0$). The previous formula simply reduces to $n_{T} = \sum_{j} \Delta n_{Tj}$ (that also can be related to $\bar{n}_{T} = \sum_{j} \Delta \bar{n}_{Tj}$ by jointly defining an average correction factor using both total values). When following the HCM tradition, the danger of mixing inconsistent values is the point to stress because there is the temptation of adding field values of queue (horizontal model) and flow imbalances (vertical model). Otherwise, some generalization is commanded by recognizing the specificities of Case II by using $\hat{n}_{At} = \Delta n_{At} - \Delta \bar{n}_{At} + \Delta \bar{n}_{At}$, for $n_{0} = 0$, or some other form that recognizes the different components of the queueing process at an approach (the growth of queue at the back of queue $\Delta n_{At}$, the dissipation of queue at the front of queue $\Delta \bar{n}_{At}$, and the final number of vehicles in free queue discharging flow yet heading to cross the stop line $\Delta \bar{n}_{At}$, if also taken as queueing).

A particular application of the temporal view is the measurement of traffic demand flow under queueing (volume or rate) from measures of traffic flow and vehicle queues. The previous discussion shows that traffic flow demand can be directly measured if queue estimates from a vertical queueing model are
available as $Q_A^T = q_A^T + \left( \tilde{n}_{A(t)} - \tilde{n}_{A(t)} \right) / (t_2 - t_1)$ (for $[t_1 ; t_2] \subseteq T$). If, however, queue measures were taken from field observation (horizontal queue values), the correct formula arises from

$$\Delta n_{AT} = \frac{\Delta \tilde{n}_{AT}}{f_{Qn}} = \frac{\Delta \tilde{n}_{AT}}{1 - K_u^T / K_n^T} = \frac{\left( Q_A^T - q_A^T \right) T}{1 - \left( Q_A^T / V_u^T \right) / K_n^T}$$

with $K_n^T \approx N_{\ell v}$, implying that

$$Q_A^T = q_A^T + \frac{f_{Qn} \Delta n_{AT}}{T} = \frac{q_A^T \cdot T + \Delta n_{AT}}{T + \Delta z_{AT}/V_u^T} \approx \frac{N_{AT} + \Delta n_{AT}}{T + \Delta n_{AT} \cdot \left( \ell v / N_u \right) / V_u^T}$$  (17a,b,c)

(as $\Delta n_{AT} = K_n^T \cdot \Delta z_{AT}$ relates the queue evolution to the evolution of the queue length in period $T$).

The pre-existing queue is usually obtained from field observation (as previously stressed, being a $n_0$, not a $\tilde{n}_0$). Then, the previously stated recursive formulas can be rewritten as

$$n_{AT} = n_{A0} + \sum_j \left( Q_A^T - q_A^T \right) T_j \approx n_{A0} + \sum_j \frac{\left( Q_A^T - q_A^T \right) T_j}{1 - Q_A^T / V_u^T \cdot \left( N_{\ell v} / \ell v \right)}$$  (18a,b)

or

$$\tilde{n}_{AT} = n_{A0} + \sum_j \left( Q_A^T - q_A^T \right) T_j \approx \tilde{n}_{A0} + \sum_j \left( Q_A^T - q_A^T \right) T_j$$  (19a,b)

(an available alternative to (17a,b) with $\tilde{n}_{A0} = n_{A0} \left( 1 - \left( Q_A^0 / V_u^0 \right) / \left( N_u^0 / \ell v \right) \right)$, if traffic conditions before the analysis period are missing and some initial condition should be assumed for subsequent corrections of $\tilde{n}_{A0}$).

Note, however, that the previous formulas implicitly convert vertical queues into horizontal queues (based on previous wave speeds or conversion factors). Any supposition for $n_{A0}$, if consistently averaged (e.g. by assuming similar conditions to the first period and maintaining the values for the previous and the first periods), do not change the results for $n_{AT}$ of the full period (e.g. taking $T = \sum_j T_j$ with an added previous period with duration $T_0$ for $\tilde{n}_{A0}$ under the assumed conditions). Also, previous formulas are better labelled as $n_{bAT}$ as only the queue change at the back of queue is considered. Unless a mixed (or extended) queue concept is adopted (treating all vehicles up to A as queueing and, forcefully having to consider changes in density or to accept a “fictitious” queue measure), the estimate should deal with changes at the front of queue as well and $n_{Ai} = n_{bAi} - \tilde{n}_{FAi}$ (of course, this observation is relevant only if this change occurs, as in Case II patterns). The option of using the concept of extended queue is similarly available (as $\tilde{n}_{Ai} = n_{bAi} - \tilde{n}_{FAi} + \tilde{n}_{FAi}$, implicitly treating the change of traffic density in the mixed “queue” and then avoiding “fictitious” queues).
Note that the expressions obtain changes in queue length and size, independently from the initial queue length or size (otherwise, the change in queueing extent $\Delta z_T = z_T - z_0$ would command time translations by $z_0/w_T$ and $z_T/w_{tn}$ at the start and end of $T$). Therefore, even implicitly assuming the previously discussed condition ($w^T_c << V_u$) for neglecting the time translation of the measurement periods by the travel speed along measurement sections (that are pattern preserving only if the vehicle speeds approaching the queuing sections are constant and uniform, by warranting the degenerated shock wave speeds that transmit changes of flow and speed with $w_T = w_{tn} = V_u$), the increased duration of queue is captured by $\Delta z_{AT}/V_u^T$. It also commands that measurements should avoid limiting intervals in the referred period (to avoid mixing data with adjacent periods, if not taking care of the time translation of measures made at some distance from the referring section). Again, the point to stress is that a more general approach has to translate time periods along sections (at least if the effect is felt as relevant), meaning that formulas (17a,b,c) have a partial (even if principal) correction.

In practical settings with over-demand, the multi-period formula is especially convenient not only because the traffic conditions change inside the analysis period $T$ but mainly because long queues on an initial stretch of road will usually reach points where demand converge or diverge to adjacent roads. This situation ask for balance of entering and exiting flow at the limit point before applying the procedure to the next segment and period. In practice, the only special point to note is that the time at which the intersection point is reached is not previously known and should be determined during field observation (or estimated using a simple iterative procedure). An example of this procedure is presented in the case study ahead and also applies to segments where the road conditions change (e.g. the number of lanes, as occurs in the case study approach too).

There is another major point to note here: the procedure works fine (and is sufficient) only if there is no bottleneck before the considered one (what can reveal itself by some intermediate queue relief followed by another queueing stretch). Otherwise, only the local manifest demand is then obtained. The adjustment for the upstream demand would also have to be made before evaluating the global manifest demand (in the sense of being that keeping their route, time and O/D decisions, even if not immediately served). Nevertheless, the intermediate adjustment follows the same approach and in general can be forwarded along the path (route and time) by splitting based on the observed maneuver proportions at the decision points of the path (ideally, local trips would also have to be considered, adding local departures and removing local arrivals but, at least for structural roads that serve a predominant proportion of through trips, local trips can usually be ignored).

Of course, the simple procedure breakdown if a more complex setting has to be considered (e.g. when queueing affects route choice). In general, the more complex settings require some kind of network approach, being the point in which this simple discussion touches the more general dynamic traffic assignment problem.
3.3 Delay Measures with Horizontal Queueing Models

An additional detail relates to the estimation of queueing delays and travel delays that are consistent with horizontal queueing models. Again, the conventional measures for queueing delay adopted by the HCM are implicitly adopting the vertical queueing model assumptions. More general measures of travel delays (e.g. those derived from path tracing or input/output methods) are not affected if the contribution of queueing is not a specific concern (perhaps due to its specific contribution to driver stress, fuel consumption or vehicle emission).

If the presence of queue is the only source of delay, the general definition of travel delay (as queueing delay) is \( d_n = t_{vn} - t_{v0} \) where \( t_{vn} \) is the travel time experienced with the presence of queue and \( t_{v0} \) is the travel time in the absence of queueing. Again, this expression is applied in the successive versions of the HCM by implicitly adopting assumptions of the vertical queueing model, delivering

\[
\tilde{d}_n = \frac{L}{V_u} + \frac{n}{q_n} - \frac{L}{V_u} = \frac{n}{q_n}
\]  

or really, even more simply, making

\[
\tilde{\tilde{d}}_n = \frac{L}{V_u} + \tilde{n} - \frac{L}{V_u} = \frac{\tilde{n}}{q_n}
\]  

(implicitly because no such discussion is made but the average queueing delay formulas can be deduced from the average queue obtained by conventional vertical queueing models and the simplified outflow model adopted throughout the HCM that sets \( q_n = C \), or \( q_n = C_s \) if distinguishing the capacity when queues are present, if \( n > 0 \), as average capacity \( C \) but also as instant capacity \( C[t] \) in queue sampling). The simplification is implicit in the analysis of discontinuous/interrupted flow conditions (then tailored to the two, three or four terms delay formulas adopted in the successive versions since the 1985 edition). There is no treatment of queues in the analysis of continuous/uninterrupted flow conditions, except when using macroscopic simulation models.

Using the horizontal queueing models, the queueing delay estimate is initially corrected to

\[
d_{st} = \frac{L - z_t}{V_u} + \frac{n_t}{q_n} - \frac{L}{V_u} = \frac{n_t}{q_n} - \frac{z_t}{V_u}
\]  

or also, by noting that \( z_t = \frac{n_t}{K_n} \equiv \frac{n_t}{N_{v}} \) and \( V_n = \frac{q_n}{K_n} \equiv \frac{q_n}{N_{v}} \), delivering

\[
d_{st} = \frac{z_t}{V_n} - \frac{z_t}{V_u} \cong \left( \frac{1}{q_n} - \frac{\ell_v}{N_{v}, V_u} \right) n_t
\]
(an expression that is conceptually consistent with most views on the effect of queueing on travel time). As before, the assumption that \( K_n \equiv \frac{N_t}{\ell_v} \) (i.e. that \( \bar{n} = \ell_v \) for vehicles in each lane of the queue) corresponds better to stopped queues (otherwise the average spacing along the queue must be evaluated, in general as a function of the queueing outflow \( q_n \) and the number of lanes \( N_t \)). As applied in the HCM formulas, these expressions could be applied to average traffic conditions by distinguishing each term of delay formulas for discontinuous/interrupted flow (meaning that queue would be \( \bar{n} \) or \( \bar{z} \) and the queueing delay would deliver the related average delay \( \bar{d}_n \)) or also to continuous/uninterrupted flow (as in 128, where it is shown that horizontal queueing models can reproduce first order macroscopic simulation models).

Average formulas can be developed after obtaining the area between the back and front of queue as

\[
A_{np} = \left( T + t_p \right) \frac{n_{AT}}{2} \tag{24}
\]

and by noting that it delivers the average time in queue for horizontal queueing models (instead of average queueing delay for vertical queueing models) as

\[
\bar{t}_n = \frac{A_{np}}{N_{T+t_p}} \text{ (per vehicle)} \tag{25a}
\]

with \( N_{T+t_p} = q_A T + q_n T_p \) (or \( N_{T+t_p} = Q_A \bar{T} + \hat{Q}_n T_p = \left( Q_A + \bar{K}_n w \right) T + \left( Q_n - \bar{K}_n w \right) T_p \) that translates into \( N_{T+t_p} = Q_A \bar{T} + K_u z_{AT} + Q_n T_p - K_u z_{AT} = Q_A \bar{T} + \hat{n}_{AT0} + Q_n T_p - \hat{n}_{ATn} \), where \( \hat{n}_{AT0}, \hat{n}_{ATn} \) are the number of free flowing vehicles along the length \( z_{AT} \) before and after queueing) and also as usual (alternatively, note that \( A_{np} = \left( T + t_p \right) \frac{z_{AT}}{2} \) would deliver \( \bar{t}_n = A_{np} z_{AT} / z_{AT} = \frac{T + t_p T_n}{2} \) as the average queueing time in sections of \( z_{AT} \) and \( z = A_{np} z_{AT} / z_{AT} = n_{AT} / K_n \) as \( n_{AT} = K_n z_{AT} \) and \( A_{np} = K_n T_{A} A_{np} \), also delivering \( \bar{t}_n = \frac{z}{V_{n}} = \frac{A_{np}}{V_{n}} \left( T + t_p T_n \right) \) as the average per vehicle). Then the average queueing delay is

\[
\bar{d}_n = \bar{t}_n - \frac{z}{V_{u}} = \frac{n}{\bar{q}_n} - \frac{z}{V_{u}} = \frac{n}{\bar{q}_n} - \frac{n}{V_{u} K_{n}^{T}} = \frac{n}{\bar{q}_n} \left( 1 - \frac{V_{u}}{V_{n}} \right) \tag{27a,b,c,d}
\]
with \( \bar{q}_n = \frac{q_A^T T + q_{tn}^T t_p n_{tn}}{T + t_p n_{tn}} \) as the average outflow from the queue and \( \bar{n}_v = \frac{\bar{q}_n^T}{K_n^T} \) as its average outflow speed. The average queueing delay can then be evaluated after obtaining queueing measures (using

\[ n_{At} = \left( Q_A^T - q_n^T \right) t / f_{Qn} T \] during \( T \), then \( n_{At} = n_{At} - n_{At}, n_{At} = \left( Q_{tn}^T - q_n^T \right) t / f_{Qn} \)

during \( T_p = \left[ T; T + t_{pn} \right] \subset T_n \), in a bottleneck imposing \( q_n^T \) and \( Q_A^T > q_n^T > Q_{tn}^T \), where some further assumption is required for evaluating \( K_n \) or \( V_n \) with \( q_n = K_n V_n \) for moving queues). As is well known, in this case, corrections for queue and delay cancel-out if queueing is the only source of delay.

However, the previous discussion makes easy to understand that this initial correction can only deal easily with Case I patterns. It is also applicable to Case II patterns but then variables should be clearly defined and evaluated. The basic point is that Case I permits to interpret the queue measure as either the queue faced at arrival or as the queue advance needed to leave the queue (as long as the front of the queue remains fixed in space and its outflow is constant at \( q_n \)). In Case I, this feature means that the same expressions can be applied to each vehicle, if ignoring over-takings along the queue, as function of \( n_1 \) or \( z_1 \) (the queue in vehicles or its length) faced at the time \( t \) when the vehicles arrives to the queue (both using vertical or horizontal queueing models). In Case II, these queue measures are eventually distinct. The first is the one directly observed at arrival but the other is the right variable to be used in delay estimation.

Using the horizontal queueing models, the relation between measures is easily obtained from the analysis of shock waves and a correct measure of queue at arrival. Based on the previous discussion, it is easy to conclude that the correct measure of queue at arrival is \( n_1 = n_{At} - n_{St} \), at least after the peak period \( T \) (before \( T \), it can be assumed that \( n_{St} = 0 \) ) and before queue dissipation at \( T + t_{qn} \) (after \( T + t_{qn} \) it can be assumed that \( n_1 = 0 \) ). Considering the movement of queue front, the time spent advancing along the queue is

\[ t_{nt} = \frac{n_1}{q_n}, \text{ where } \hat{q}_n q_n^T \text{ from the fixed front queue during } T \text{ but } \hat{q}_n = q_n^T + K_{tn} \cdot w_{tn} \text{ from the moving front queue during } t_{qn}, \text{ both taken while advancing along the queue after the arrival time } t. \text{ Then, the initial correction is enough for the vehicles that leave the queue during the peak period } T, \text{ i.e. those vehicles that arrive in } T_x = \left[ 0; t_x \right] \subset T \text{ such that } t_x + t_{nx} = T \text{ at } z_x = w_c^T t_x = V_n^T t_{nx}, \text{ where } t_{nx} \text{ is the time in queue for the vehicle that arrive at } t_x. \text{ For the remaining vehicles that arrive during } T_y = \left[ t_x, T \right] \subset T \text{ or } T_q = \left[ T; T + t_{qn} \right] \subset T_n \text{ the correction have to be}
supplemented by considering the period with $q_{n}^{Tn} = q_{A}^{Tn} + K_{s}^{Tn}w_{s}^{Tn}$, as a function of the arrival time $t$ of each vehicle (assuming $d_{nt} = 0$ for vehicles arriving during $T_{w} = [T + t_{q}^{Tn}, T + T_{n}] \subset T_{n}$). Under uniform conditions in $T$ and $T_{n}$, time in queue can be linearly interpolated from zero to $t_{x}$ and then to zero again, if needed, and the average time in queue is $\frac{Z_{x}}{2}$ both before and after $t_{x}$. Note, however, that vehicles that arrive to the queue at $t$ in $T_{x}$ have a queueing delay $\frac{n_{t}}{q_{n}^{T_{x}}}$ or $\frac{Z_{t}}{V_{n}^{T_{x}}}$, while vehicles that arrive to the queue at $t$ in $T_{y}$ have a queueing delay $(T - t_{x}) + \frac{n_{t} - q_{n}^{T_{x}}(T - t_{x})}{q_{n}^{T_{y}}}$ or $(T - t_{x}) + \frac{Z_{t} - V_{n}^{T_{y}}(T - t_{x})}{V_{n}^{T_{y}}}$, where $\hat{V}_{n}^{T_{y}} = V_{n}^{T} + w_{s}^{T_{y}}$ and $q_{n}^{T_{y}} = q_{n}^{T} + K_{s}^{T}w_{s}^{T}$, until $T$ (remembering that $n_{t} = n_{At}$ up to $T$, when $n_{t} = n_{AT}$). In period $T_{n}$, the queueing delay is $\frac{n_{t}}{q_{n}^{T_{n}}} or \frac{Z_{t}}{V_{n}^{T_{n}}}$ in $T_{q}$, until $T + t_{q}^{T_{n}}$ (remembering that $n_{t} = n_{At} - \tilde{n}_{St}$ in period $T_{n}$, until $T + t_{q}^{T_{n}}$, and then $n_{t} = 0$ in $T_{w}$ for the rest of the period $T_{n}$).

Avoiding this detailed calculation, the average queueing delay can be obtained by considering the overall queueing delay as related to the area between the back and front of the restrained queue as

\[ A_{nq} = T \cdot \frac{n_{AM}}{2} \quad (28) \]

and again noting that it delivers the average time in queue for horizontal queueing models (instead of average queueing delay for vertical queueing models) as

\[ \bar{t}_{n} = \frac{A_{nq}}{N_{T+T_{lq}}} \quad \text{(per vehicle)} \quad (25b) \]

with $N_{T+T_{lq}} = \hat{q}_{A}^{T_{n}}T + \hat{q}_{n}^{T_{n}}T_{q}^{T_{n}} = q_{A}^{T}T + (q_{A}^{T} + K_{s}^{T}w_{s}^{T})t_{q}^{T_{n}}$, that translates to

\[ N_{T+T_{lq}} = q_{A}^{T_{n}}T + q_{A}^{T_{n}}t_{q}^{T_{n}} + K_{s}^{T_{n}}z_{AM} = q_{A}^{T_{n}}T + q_{A}^{T_{n}}t_{q}^{T_{n}} + \hat{n}_{AMq} \]

where $\hat{n}_{AMq} = \tilde{n}_{AM}$ is the number of vehicles that fill the maximum queue extent $z_{AM}$ with the density $K_{s}^{T}$ after queue dissipation at $t_{q}^{T_{n}}$ (or

\[ N_{T+T_{lq}} = \hat{Q}_{A}^{T}T + \hat{Q}_{A}^{T_{n}}t_{q}^{T_{n}} = (Q_{A}^{T} + K_{u}^{T}w_{c}^{T})T + (Q_{A}^{T} + K_{u}^{T}w_{c}^{T})t_{q}^{T_{n}} \]

that translates to

\[ N_{T+T_{lq}} = Q_{A}^{T}T + K_{u}^{T}z_{AT} + \hat{Q}_{A}^{T}t_{q}^{T} + K_{u}^{T}(z_{AM} - z_{AT}) = Q_{A}^{T}T + \hat{n}_{AT0} + \hat{Q}_{A}^{T}t_{q}^{T} + \hat{n}_{AMq} - \tilde{n}_{AT0} \]

is the number of free flowing vehicles that fill the maximum queue length $z_{AT}$ with the before queueing density
1. \( K_n^T \) and \( \bar{n}_{AMn}, \bar{\bar{n}}_{ATn} \) are the number of free flowing vehicles that fill the maximum queue extent and maximum queue length with the after queueing density \( K_n^T \). Also, the average queue is

\[
\bar{n} = \frac{A_{nq}}{T + t_q}
\]

(26b)

4. as usual (alternatively, note that \( A_{zq} = T \cdot \frac{z_{AM}}{2} \) would deliver \( \bar{t}_{nz} = \frac{A_{zq}}{z_{AM}} = \frac{T}{2} \) as the average queueing time in sections of \( z_{AM} \) and \( \bar{z} = \frac{A_{zq}}{K_n^T} \) as \( n_{AM} = K_n^T z_{AM} \) and \( A_{nq} = K_n^T A_{zq} \), also delivering

\[
\bar{t}_n = \frac{\bar{z}}{V_n^T} = \frac{A_{zq}}{V_n^T (T + t_q)}
\]

as the average time in queue per vehicle, where \( V_n^T \) is the relative average queue outflow speed, considering \( V_n^T \) during \( T \) and \( \hat{V}_n^T = V_n^T + w_s^T \) during \( t_q^T \) as experienced by each vehicle while advancing along the queue). Then, similarly, the average queueing delay is

\[
\bar{d}_n = \bar{t}_n - \frac{\bar{z}}{V_u} = \frac{\bar{n}}{q_u} - \frac{\bar{z}}{V_u} - \frac{\bar{n}}{q_u} \cdot V_u \cdot K_n^T = \frac{\bar{n}}{q_u} \left( 1 - \frac{\bar{V}_n}{V_u} \right)
\]

(27c,f,g,h)

with \( \bar{q}_n = \frac{q_{AT}^T + q_{AT}^T t_q^T}{T + t_q^T} \) as the average relative outflow from the queue and \( \bar{V}_n = \frac{\bar{n}}{K_n^T} \) as its average relative outflow speed. Again, the average queueing delay can then be evaluated after obtaining queueing measures (using \( n_{AT} = (Q_A^T - q_n^T) t / f_q^T \) during \( T \), then \( n_{AT} = \bar{n}_{Stq} \) for \( t_q^T \) with

\[
n_{AT} = n_{AT}^T = \bar{n}_{AT} + (Q_A^T - q_n^T) t / f_q^T, \bar{n}_{St} = (q_n^T - q_n^T) t / f_q^T \) during \( T_q = [T; T + t_q^T] \subset T_n^A \), in a bottleneck imposing \( q_n^T, q_n^T > Q_A^T, q_n^T > q_n^T, q_n^T > Q_A^T \), where some further assumption is required for evaluating \( K_n^T \) or \( V_n \) with \( q_n^T = K_n^T V_n \), as before).

A more detailed form can be applied by distinguishing the stopped queueing delay \( d_{ns} \) (assuming \( V_n \neq 0 \)) from moving queue delay \( d_{nm} \) (assuming \( V_n = \bar{V}_m > 0 \)) as

\[
d_n = d_{ns} + d_{nm} = t_n + \frac{z}{V_m} - \frac{z}{V_u}
\]

where \( d_{ns} = t_n \) (the time spent stopped in queue) and \( d_{nm} = \frac{z}{V_m} - \frac{z}{V_u} \) (with the average speed while moving
in queue $\bar{V}_m$), also noting that $\frac{Z}{\bar{V}_n} = t_m + \frac{Z}{\bar{V}_m}$ implies that $\frac{1}{\bar{V}_n} = \frac{t_m}{z} + \frac{1}{\bar{V}_m}$ or $\bar{V}_n = \frac{1}{\left(\frac{t_m}{z} + \frac{1}{\bar{V}_m}\right)}$. A more general form can be applied by adding a term related to the free queue discharging flow regime (breaking down equivalence results based on input-output analysis). As before, improvements can also be made by adopting more general outflow models (this aspect is discussed ahead, when better considering the application of horizontal queueing models to each term of queueing and delay formulas for discontinuous/interrupted flow conditions that are usual at intersections). The development of delay measures for transient parts of the full period are also of interest. An initial effort in these directions is inserted in Appendix B.

4. SOME GENERAL CONCLUSIONS

Before leaving this preliminary treatment of horizontal queueing models, some further general comments are in case:

- For continuous/uninterrupted flow, the direct analysis using the first order macroscopic traffic model would be possible; however, data for the direct analysis would usually require some form of a Fundamental Diagram of Traffic for the normal section (perhaps also for the bottleneck section); if data is available, it should be viable, in addition to being probably a classic exercise proposed in any basic course on Traffic Flow Theory or any introduction to a Traffic Engineering course; actually, the solution is very similar to the numeric procedure proposed for the analysis of Freeway Systems as implemented in the FREEVAL computational tool by the HCM 2010;

- For discontinuous/interrupted flow, the case for proposing simplified formulas is strong; each interruption cycle generate at least a stopping and a starting wave; unless queues clear at every interruption cycle, several pairs of stopping and starting waves are simultaneously running along a stretch of road; in general, it is out of reason to try the tracking of these simultaneous waves generated under discontinuous/interrupted flow conditions for obtaining formulas to estimate the operational measures; even the numeric procedure embodied into the computational tool for the analysis of Urban Street Systems of HCM 2010 avoids such approach.

The interest here is to propose simplified formulas that avoid the need of knowing the Fundamental Traffic Flow Diagram and calculating speeds of propagation waves (in the spirit of the current HCM formulas), to be used for the analysis of intersections based on horizontal queueing models (instead of the current HCM formulas that are based on vertical queueing models), thus providing a more consistent and precise estimation method. Otherwise, the analysis was essentially that of a simplified queueing-theoretical traffic model similar to the procedures proposed for Freeway or Urban Street Systems in the current version of the HCM. A relevant detail is related to the effect of randomness, currently only tackled (not fully appropriately) in the analysis of
discontinuous/interrupted flow conditions, where it comes mixed with the effects of traffic control. In the
previous discussion, inflows and outflows were generically represented but, in the following, these values have
to be specified based on the discipline enforced by traffic control at intersections for achieving this purpose. A
more important practical point is, however, worth noting from the start: the distinction between Cases I and II is
much more subtle for discontinuous/interrupted flow conditions because then queues are always present, due to
the effects of interruption cycles. This point is detailed discussed in the next section.

It is now widely known (see e.g. El-Taha, Stidham, 1999) that most properties deduced by conventional
queueing theory (usually meaning vertical queueing theory as a rule) are independent of particular assumptions
on deterministic and/or stochastic features, by being valid to any realization of a queueing process (or even of a
more general input/output process). The previous discussion, even if simple in nature, shows that there is no
need for reconciling traffic analysis based on conventional queueing models (usually meaning vertical queueing
models) or first order traffic flow models (essentially horizontal queueing models) by other feature than the
horizontal (spatial) dimension of queues in road traffic. There is a large literature on this reconciliation that can
be dealt with and adjusted appropriately by understanding this point and applying the previous results (this point
was previously argued and exemplified in Hurdle, Son, 2001, without analytical justification, also including the
time translation of periods but using \( \tau_o = L/V_{FL} \) instead of \( \tau[t] = (L - z)/\sqrt{V_{ul}} + z/\sqrt{V_{nl}} = \tau_A[t] + \tau_o[t] \), as stated
here). Then, correcting for the effect of horizontal queueing is the only additional component procedure needed
to translate most results from queueing (or input/output) processes to traffic engineering contexts.

This understanding is even more important when noting that the current view that associates first order
traffic flow models (or other types of flow models) to deterministic models is misleading. For example, by
replacing the very simple outflow model \( q^T_A = \min\{Q^T_A, C^T_A\} \) usual in the HCM tradition with a convenient
alternative (see ahead), it is possible to translate the traffic analysis based on traffic flow models to traditional
HCM formulas for intersections, that are based on stochastic model assumptions (e.g. those implicit in
poissonian models or coordinate transformation formulas). As discussed in Brilon, 2008, for a particular
context but more widely relevant, current formulas for queues and delays hide some suppositions about the
implied behavior of traffic flow that can perhaps be better represented in alternative formulations. Modeling
options that make these assumptions more clear may favor the development of better approaches as well.

In general, the translation from vertical to horizontal queueing models is always a specific need of road
traffic applications. At this point, however, there is much more to be studied. The simple method presented in
this paper considers only the basic translation of estimated queues (and the implied adjustment in traffic demand
measures). This observation mostly applies to the dissipation times too (in the overall peak period or in each
interruption cycle under discontinuous/interrupted traffic flow conditions), where our previous discussion
showed that the dissipation time seen at M (the measure to be used in the evaluation of queue profiles) is
distinct from the dissipation time seen at A (the measure to be used in the evaluation of capacity) but here there
is a need to clarify and choose new concepts and assumptions before proceeding. This task is initiated in the
next section. Nevertheless, as can be readily checked from the previous discussion, several other formulas
could be improved for realism and precision by using the horizontal queueing model (as the estimates are
distinct from those obtained from the vertical queueing models) in general.

This paper shows that a simple revision of HCM formulas for queues at intersections based on
horizontal queueing models is readily available and can be applied without requiring additional data, improving
the understanding and the precision of current estimates. However, the paper also highlighted that several
missing points are brought by the introduction of horizontal queueing models. Among these points deserving
further studies, better consideration of alternative patterns of congestion dissipation (Cases I, II and the mixed
cases not differentiated by vertical queueing models) and revision of the time predicted for dissipation of queues
(as seen at the point of maximum queue extent or at the stop line of intersection approaches) seems to be the
more basic ones. Even if not discussed in this paper, corresponding revision of delay formulas are equally
relevant as well as formulation of improved outflow models that can embody stochastic assumptions.

Even if incipient, the paper shows that the recognition of the features derived from the horizontal
(spatial) dimension of road traffic queueing is able to reveal the specificity of these phenomena and provide a
better understanding of the adjustments needed to translated results from the conventional queueing models to
the field of traffic operation. It seems that a much larger attention and effort is justified by the importance of
these features than currently observed in the traffic engineering research area.
REFERENCES


Hurdle, V.F.; Son, B. (2001) – “Shock Waves and Cumulative Arrival and Departure Models – Partners without Conflict”. In Transportation Research Record 1776, pp.159-166.


Pietrantonio, H.


**FURTHER REFERENCES**


Kimber, R.M.; Hollis, E.M. (1979) – Traffic Queues and Delays at Road Junctions. TRRL-Transport and Road Research Laboratory, LR989, UK.


McTrans (19??) – TRANSYT 7F. McTrans . .


A. TABLES FOR CORRECTIONS TO DERIVE HORIZONTAL QUEUEING ESTIMATES

Given the previous discussion, further assumptions about the traffic states during queue formation and
dissipation are needed for reaching better estimates. Even if deserving further study, several assumptions are
readily available (see e.g. 107, App.D). As starting values, adjustments can be made by adopting saturation
flow around $S = 1800 \frac{v}{h} / \ln$ (assumed 1500to2000) with $V_s = 50km / h$ (assumed 40to60), then

$K_s = 36v / km / \ln$ (or a better value as $K_s = S/V_s$), jam density around $K_j = 146v / km / \ln$ (assumed
125to167) or with $K_j = N_v / \ell_v$ (evaluated using current HCM guidance for $\ell_v$). If needed, capacity can be
estimated around $C \geq 110S$ (assumed 1,05to1,15) with $V_c = 70km / h$ (assumed 60to80) or with

$K_c = 28v / km / \ln$ (assumed 25to30). For unconstrained flow conditions, $V_u = 0,90.V_p$ (assumed
0,80to1,00 of free flow speed) and then $K_u = Q/V_u$. For queueing flow, $K_n = 36 + 110 \left( \frac{q_n}{S} \right) v / km / \ln$
(or, more generally $K_n = K_j - (K_j - K_s) \frac{q_n}{S}$, with $K_s = S/V_s$ for free queue discharging flow $q_n = S$

with $V_n = V_s$ ; otherwise $V_n = \frac{q_n}{K_j - (K_j - K_s)q_n/S}$ for corresponding speeds from queueing flow $q_n \leq S$).

As example, these approximations permit to estimate wave speeds for unconstrained interrupted flow
conditions (i.e. flow conditions constrained only by periodic traffic interruptions), as $w_c = Q/\left( K_j - Q/V_u \right)$
for the stopping wave speed for unsaturated cycles, $w_s = S/\left( K_j - S/V_s \right)$ for the discharging wave speed, and

$w_m = (S-Q)/(S/V_s - Q/V_u)$ for the normalization wave speed (as discussed ahead, the stopping wave
speed is $w_{cs} = S/\left( K_j - S/V_s \right) = w_s$ for saturated cycles), in general meaning $w_c \sim 10km / h$ (in general
5to15 but largely variable with the demand to flow relation), $w_s \sim 20km / h$ (in general 15to25 or a constant
$S/(K_j - S/V_s)$ under the linear assumption for the forced flow regime), and $w_m > 30km / h$ (in general
25to40 or a constant $V_u$ under the uniform speed assumption for the normal flow regime).

The chosen method develops its correction factors based on the same assumptions, as summarized here
in Tables A1, A2, and A3 (that would be required for horizontal queueing model corrections).
Pietrantonio, H.

Table A1. Correction Factors for Queues based on Typical Parameters of a Horizontal Queueing Model

a. Unconstrained Arrivals \( q^T_n = Q^T_n \) (outside queues)

<table>
<thead>
<tr>
<th>( \ell_v ) (m)</th>
<th>( V^T_A ) (km/h)</th>
<th>( Q^T_A / \bar{N}_T ) ( (v/h/ln) ) with ( Q^T_A / \bar{N}_T &gt; q^T_n / \bar{N}_T ) = ( V^T_n / \ell_v )</th>
</tr>
</thead>
<tbody>
<tr>
<td>8(*)</td>
<td>60</td>
<td>1.0274</td>
</tr>
<tr>
<td></td>
<td>70</td>
<td>1.0234</td>
</tr>
<tr>
<td></td>
<td>90</td>
<td>1.0181</td>
</tr>
<tr>
<td>15</td>
<td>60</td>
<td>1.0526</td>
</tr>
<tr>
<td></td>
<td>70</td>
<td>1.0448</td>
</tr>
<tr>
<td></td>
<td>90</td>
<td>1.0345</td>
</tr>
<tr>
<td>25</td>
<td>60</td>
<td>1.0909</td>
</tr>
<tr>
<td></td>
<td>70</td>
<td>1.0769</td>
</tr>
<tr>
<td></td>
<td>90</td>
<td>1.0588</td>
</tr>
</tbody>
</table>

Obs.: Correction factor \( f_{cn}^T = \frac{\bar{n}_A}{\bar{n}_A} \) for \( n_A = f_{cn}^T \bar{n}_A \) corresponding to \( n_A = \frac{\bar{n}_A}{1 + \left( Q^T_A / \bar{N}_T \right) / V^T_\ell} \), where \( \ell_v = \bar{\ell}_n \) is the vehicle spacing in the queue (a function of the queue discharging flow \( q^T_n \leq S, V^T_n \leq V^T_\ell \)). (*) \( q^T_n \cong 0 \).

b. Free Queue Discharging \( q^T_A = S^T_A \) (from queues)

<table>
<thead>
<tr>
<th>( \ell_v ) (m)</th>
<th>( V^T_s ) (km/h)</th>
<th>( S^T_A / \bar{N}_T ) ( (v/h/ln) ) with ( S^T_A / \bar{N}_T &gt; q^T_n / \bar{N}_T ) = ( V^T_n / \ell_v )</th>
</tr>
</thead>
<tbody>
<tr>
<td>8(*)</td>
<td>40</td>
<td>1.4286</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>1.3158</td>
</tr>
<tr>
<td></td>
<td>60</td>
<td>1.2500</td>
</tr>
<tr>
<td>15</td>
<td>50</td>
<td>1.8182</td>
</tr>
<tr>
<td></td>
<td>60</td>
<td>1.6000</td>
</tr>
<tr>
<td></td>
<td>70</td>
<td>1.4737</td>
</tr>
<tr>
<td>25</td>
<td>60</td>
<td>2.6667</td>
</tr>
<tr>
<td></td>
<td>70</td>
<td>2.1538</td>
</tr>
<tr>
<td></td>
<td>80</td>
<td>1.8824</td>
</tr>
</tbody>
</table>

Obs.: Correction factor \( f_{cs}^T = \frac{\bar{n}_A}{\bar{n}_A} \) for \( n_A = f_{cs}^T \bar{n}_A \) corresponding to \( n_A = \frac{\bar{n}_A}{1 + \left( S^T_A / \bar{N}_T \right) / V^T_\ell} \), where \( \ell_v = \bar{\ell}_n \) is the vehicle spacing in the queue (a function of the queue discharging flow \( q^T_n \leq S, V^T_n \leq V^T_\ell \)). (*) \( q^T_n \cong 0 \).
Table A2. Correction Factors for Total Clearance Time of Queues based on Typical Parameters of a Horizontal Queueing Model After the Peak Demand Period for Case I or Case II Patterns

<table>
<thead>
<tr>
<th>$\ell_v$ (m)</th>
<th>$V_A^T$ (km/h)</th>
<th>$V_{A}^{Tn}$ (km/h)</th>
<th>$Q_A^T / N_f$ (v/h/ln) &lt; $q_A^{Tn} / N_f$ with $Q_A^T = f_Q Q_A^{Tn}, I : f_Q^T = 1.25$</th>
<th>$Q_A^{Tn} / N_f$ (v/h/ln) &lt; $q_A^{Tn} / N_f$ with $Q_A^T = f_Q Q_A^{Tn}, II : f_Q^T = 0.75$</th>
</tr>
</thead>
<tbody>
<tr>
<td>8(*)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>60</td>
<td>1.0069</td>
<td>1.0143</td>
<td>1.0222</td>
<td>1.0308</td>
</tr>
<tr>
<td>70</td>
<td>1.0059</td>
<td>1.0121</td>
<td>1.0188</td>
<td>1.0258</td>
</tr>
<tr>
<td>90</td>
<td>1.0045</td>
<td>1.0093</td>
<td>1.0143</td>
<td>1.0195</td>
</tr>
<tr>
<td>15</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>60</td>
<td>1.0133</td>
<td>1.0286</td>
<td>1.0462</td>
<td>1.0667</td>
</tr>
<tr>
<td>70</td>
<td>1.0113</td>
<td>1.0240</td>
<td>1.0383</td>
<td>1.0545</td>
</tr>
<tr>
<td>90</td>
<td>1.0087</td>
<td>1.0182</td>
<td>1.0286</td>
<td>1.0400</td>
</tr>
<tr>
<td>25</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>60</td>
<td>1.0233</td>
<td>1.0526</td>
<td>1.0909</td>
<td>1.1429</td>
</tr>
<tr>
<td>70</td>
<td>1.0196</td>
<td>1.0435</td>
<td>1.0732</td>
<td>1.1111</td>
</tr>
<tr>
<td>90</td>
<td>1.0149</td>
<td>1.0323</td>
<td>1.0526</td>
<td>1.1064</td>
</tr>
</tbody>
</table>

Obs.: Correction factor $f_{Q}^{Tn} = \frac{t_f}{t_s}$ for $t_f = f_{Q}^{Tn} \tilde{t}_s$ corresponding to $t_s = \tilde{t}_s \cdot \frac{1000 / \ell_v - (Q_A^{Tn} / N_f) / V_A^T}{1000 / \ell_v - (Q_A^T / N_f) / V_A^{Tn}}$, where

$\ell_v = \bar{s}$ is the vehicle spacing in the queue (a function of the queue discharging flow $q_n^T \leq S, V_n^T \leq V_s^T$); $f_{Q}^{Tn} = 1$ for $f_{Q}^T = 1$; $q_n^T / N_f = V_n^T / \ell_v$; $\tilde{t}_s = T(Q_A^T - q_A^{Tn}) / (Q_A^{Tn} - Q_A^T)$ with I: $Q_A^{Tn} < q_n^T$ or II: $q_A^{Tn} > q_n^T \cdot (*) q_n^T \equiv 0$.
Table A3. Correction Factors for Dissipation Time of Queues based on a Horizontal Queueing Model After the Peak Demand Period for Case II Patterns Only

a. Typical Parameters for $q_A^{T_{n}} / N_f = 2000 \, v / h / ln$ (as saturation flow/lane)

<table>
<thead>
<tr>
<th>$\ell_v$ (m)</th>
<th>$V_A$ (km/h)</th>
<th>$V_A^{T_{n}}$ (km/h)</th>
<th>$Q_A^{T_{n}}$ (v / h / ln) with $Q_A^{T} = f_q^{T} Q_A^{T_{n}}$, II : $f_q^{T} = 0,75$, $q_A^{T_{n}} / N_f = 2000 , v / h / ln$</th>
</tr>
</thead>
<tbody>
<tr>
<td>8(*)</td>
<td>60</td>
<td>0,7283</td>
<td>0,7231, 0,7177, 0,7121, 0,7062, 0,7000, 0,6935</td>
</tr>
<tr>
<td></td>
<td>75</td>
<td>0,7251</td>
<td>0,7157, 0,7049, 0,6920, 0,6760, 0,6550, 0,6248</td>
</tr>
<tr>
<td></td>
<td>90</td>
<td>0,7229</td>
<td>0,7108, 0,6966, 0,6793, 0,6574, 0,6283, 0,5865</td>
</tr>
<tr>
<td>15</td>
<td>60</td>
<td>0,7238</td>
<td>0,7135, 0,7023, 0,6902, 0,6769, 0,6624, 0,5944</td>
</tr>
<tr>
<td></td>
<td>75</td>
<td>0,7215</td>
<td>0,7083, 0,6933, 0,6760, 0,6552, 0,6293, 0,5944</td>
</tr>
<tr>
<td></td>
<td>90</td>
<td>0,7200</td>
<td>0,7048, 0,6873, 0,6667, 0,6414, 0,6090, 0,5644</td>
</tr>
<tr>
<td>25</td>
<td>60</td>
<td>0,7170</td>
<td>0,6984, 0,6769, 0,6519, 0,6222, 0,5867, 0,5432</td>
</tr>
<tr>
<td></td>
<td>75</td>
<td>0,7162</td>
<td>0,6969, 0,6748, 0,6491, 0,6188, 0,5820, 0,5351</td>
</tr>
<tr>
<td></td>
<td>90</td>
<td>0,7156</td>
<td>0,6957, 0,6728, 0,6462, 0,6144, 0,5752, 0,5240</td>
</tr>
</tbody>
</table>

b. Typical Parameters for $q_A^{T_{n}} / N_f = 1500 \, v / h / ln$ (as saturation flow/lane)

<table>
<thead>
<tr>
<th>$\ell_v$ (m)</th>
<th>$V_A$ (km/h)</th>
<th>$V_A^{T_{n}}$ (km/h)</th>
<th>$Q_A^{T_{n}}$ (v / h / ln) with $Q_A^{T} = f_q^{T} Q_A^{T_{n}}$, II : $f_q^{T} = 0,75$, $q_A^{T_{n}} / N_f = 1500 , v / h / ln$</th>
</tr>
</thead>
<tbody>
<tr>
<td>8(*)</td>
<td>50</td>
<td>0,6879</td>
<td>0,6735, 0,6557, 0,6319, 0,5966, 0,5329, 0,3548</td>
</tr>
<tr>
<td></td>
<td>60</td>
<td>0,6847</td>
<td>0,6661, 0,6423, 0,6100, 0,5617, 0,4773, 0,2758</td>
</tr>
<tr>
<td></td>
<td>70</td>
<td>0,6824</td>
<td>0,6608, 0,6331, 0,5953, 0,5394, 0,4444, 0,2386</td>
</tr>
<tr>
<td>15</td>
<td>50</td>
<td>0,6827</td>
<td>0,6625, 0,6381, 0,6071, 0,5645, 0,4949, 0,3221</td>
</tr>
<tr>
<td></td>
<td>60</td>
<td>0,6804</td>
<td>0,6572, 0,6285, 0,5913, 0,5385, 0,4516, 0,2571</td>
</tr>
<tr>
<td></td>
<td>70</td>
<td>0,6788</td>
<td>0,6534, 0,6218, 0,5803, 0,5213, 0,4254, 0,2258</td>
</tr>
<tr>
<td>25</td>
<td>50</td>
<td>0,6749</td>
<td>0,6448, 0,6079, 0,5615, 0,5000, 0,4106, 0,2403</td>
</tr>
<tr>
<td></td>
<td>60</td>
<td>0,6741</td>
<td>0,6433, 0,6058, 0,5584, 0,4949, 0,4000, 0,2160</td>
</tr>
<tr>
<td></td>
<td>70</td>
<td>0,6735</td>
<td>0,6420, 0,6036, 0,5548, 0,4889, 0,3893, 0,2000</td>
</tr>
</tbody>
</table>

Obs.: Correction factor $f_q^{T_{n}} = \frac{t_q}{t_A}$ for $t_q = f_q^{T_{n}} t_A$ corresponding to

$$f_q^{T_{n}} = \frac{(q_A^{T_{n}} - Q_A^{T_{n}})/(1 - K_u^{T}/K_u)}{(q_A^{T_{n}} - q_n)/(1 - K_u^{T}/K_u) - (Q_A^{T_{n}} - q_n)/(1 - K_u^{T}/K_u)}$$

by assuming $K_u^{T} = K_u^{T_{n}} = Q_A / V_A$, and

$$K_n = K_J - (K_J - K_S) q_A / S_A$$

or

$$q_n = S (K_J - K_n) / (K_J - K_S)$$

typical values

$Q_A^{T_{n}} = S_A = 2000; 1800; 1500 \, v / h / ln$, $V_S = 60; 50; 40 \, km / h \leq V_A$, $K_S = 33,3; 36,0; 37,5 \, v / km / ln$,

$K_J = 125 \, v / km / ln$ (variables without T/Tn are constant) assuming $N_f = 1$ (without loss of generality). (*) $q_A^{T_{n}} \geq 0$. 
Several strong results on delay estimates come from input-output analysis but these expressions have to be developed for more general settings. Most results from these analyses are stated for deterministic (and uniform) arrivals and departures but valid under more general settings (as justified by the sample path approach of El-Taha, Stidham, 1999) and can applied to queueing delay in the simplified case where queueing is the only source of delay, as corrections for queue and delay cancel-out in several known cases (e.g. as argued by Daganzo, 1983). However, the previous discussion has shown that this result does not fit the general patterns of Case II queueing, as there are other potential sources of delay during free queue discharging flow (then requiring a specific representation of queueing). Even other cases, not discussed here, can also be related to the impossibility of assuming that the input and output flow rates $Q_A, q_A$ are taken as pre-determined (meaning, for the discussion made here, independent of the queueing process, as negated when the traffic interaction in a network of road segments generates demand starvation and/or blocking of outflow in adjacent sections, or more generally under transient conditions where queue dynamics is dependent on the value of the initial queue as discussed in Bertsimas, Mourtzinou, 1997, including sample path arguments). Note, also, that queueing time (i.e. the time facing traffic operations in queueing) can be a better operational measure than queueing delay for evaluating some effects (as emissions and distress) other than the effect on travel time.

The deterministic version of the argument for the analysis of queueing in the peak and next periods, as previously adopted, can add vehicle detail by numbering their arrival and departure under the assumed traffic conditions (ignoring overtakings). Referring to section A, under VQM assumptions, the arrival time of vehicle $i$ can be obtained as $t_{Ci} = h_{A0}^T + (i - 1).h_A^T = h_{A0}^T + (i - 1)/Q_A^T$, with a synchronization start time $h_{A0}^T < h_A^T$, for $i \leq i_A^T = \left[\left(T - h_{A0}^T\right)/h_A^T\right]$, and $t_{Ci} = T + h_{A0}^T + (i - i_A^T - 1).h_A^T = T + h_{A0}^T + (i - i_A^T - 1)/Q_A^T$ for $i_A^T < i \leq i_A^{Tn} = i_A^T + \left[\left(T_n - h_{A0}^T\right)/h_A^{Tn}\right]$, where $h_{A0}^{Tn} = \left(1 - h_{At}^T / h_A^T\right)h_A^{Tn}$ with $h_{At}^T = T - i_A^T h_A^T$. Similarly, the departure time of vehicle $i$ can be obtained as $t_{Di} = h_{S0}^T + i.\bar{h}_S^T = h_0^T + i/q_A^T$, with a synchronization start time $h_{S0}^T \geq h_{A0}^T$ (related to the vehicle movement at the head of queue), for $i \leq i_S^T = \left[\left(T - h_{S0}^T\right)/\bar{h}_S^T\right]$, then $t_{Di} = T + h_{S0}^T + (i - i_S^T - 1).\bar{h}_S^{Tn} = T + h_{S0}^T + (i - i_S^T - 1)/q_A^{Tn}$ for $i_S^T < i \leq i_S^{Tn} = i_S^T + \left[\left(i_S - h_{S0}^T\right)/\bar{h}_S^{Tn}\right]$ and $t_{Di} = T + \bar{t}_s + (i - i_S^{Tn} - 1)h_A^{Tn} = T + \bar{t}_s + (i - i_S^{Tn} - 1)/Q_A^{Tn}$ for $i_S^{Tn} < i \leq i_S^T = i_S^{Tn} + \left[\left(T_n - \bar{t}_s\right)/\bar{h}_S^{Tn}\right]$, where
1 \[ h_{S0}^{Tn} = \left( 1 - \frac{h_{Sf}^T}{h_{Sf}^T} \right) h_{Sf}^{Tn} \] with \( h_{Sf}^T = T - i_{Sf}^T h_{Sf}^T \) and the dissipation time \( \tilde{t}_s \) is obtained by the condition

2 \[ \tilde{t}_s = \tilde{t}_{cis} = \tilde{t}_{Das} \] for the first time in \( Tn \) (being maintained after \( \tilde{t}_s \) too).

Figure B-1. Total Delay and Queueing Delay in Vehicle Trajectories
General statistics on queues and delay for the process at A are built from the instantaneous queue at t as

$$\tilde{n}[t] = \tilde{n}_c[t] - \tilde{n}_d[t],$$

with $$\tilde{n}_c[t] = \max\{i / \tilde{t}_{ci} < t\}; \tilde{n}_d[t] = \max\{i / \tilde{t}_{di} < t\},$$ and the individual delay

$$\tilde{d}_{ni} = \tilde{t}_{di} - \tilde{t}_{ci} = t_{ni},$$
as usual in Queueing Theory, being independent of any detail on the traffic operation but the pre-determined input and output flow rates and $$Q_A, q_A$$ the duration of time periods. However, in road traffic, actually these measures refer to vertical queue and total delay in a road segment (meaning that total delay is taken as queueing delay and time in queue is taken as queueing delay in the usual interpretation).

A detailed analysis of the space-time frame for HQMs can be related to that of VQMs by obtaining arrival and departure times of individual vehicles to and from (real) queues $$t_{ci}, t_{di}$$ as function of the corresponding arrival and departure times to and from vertical queues $$\tilde{t}_{ci}, \tilde{t}_{di}$$ (i.e. by tracking vehicle number in time by cumulating $$Q_A, q_A$$ as if occurring at section A), delivering general adjustments for headways as seen at the shock wave frontier as

$$\tau = \frac{1-K_c/K_d}{1-V_c/V_d}h_c = \frac{K_d/K_c-1}{V_c/V_d-1}h_d,$$

where locally C and D refer to traffic conditions before and after the shock wave (the expressions come from the effect on successive vehicles of shock waves travelling $$\delta = w.\tau$$ and satisfying

$$h_c + \frac{\delta}{V_c} = \frac{\delta}{V_D} + h_d$$

for backward waves or

$$h_c + \frac{\delta}{V_c} = \frac{\delta}{V_D} + h_d$$

for forward waves, also delivering

$$\delta = \frac{1-q_c/q_d}{1-V_c/V_d}e_c = \frac{q_d/q_c-1}{V_d/V_c-1}e_d$$

for backward waves and

$$\delta = \frac{1-q_c/q_d}{V_c/V_d-1}e_c = \frac{q_d/q_c-1}{1-V_d/V_c}e_d$$

for forward waves), as related to headways $$h = \frac{1}{q}$$ (and spacings $$e = \frac{1}{K}$$). Some well known formulas for shock wave speeds in the individual vehicular setting (e.g. Akçelik, Besley, Hoper, 1999) generalizes to

$$w = \frac{e_c - \tau V_c}{h_c - \delta V_c} = \frac{e_d - \tau V_D}{h_d - \delta V_D}$$

for backward waves (as $$w_{c}, w_{s}$$ )

and

$$w = \frac{\tau V_c - e_c}{h_c - \delta V_c} = \frac{\tau V_D - e_d}{h_D - \delta V_D}$$

for forward waves (as $$w_{e}, w_{m}$$ ).
Then for arrivals of the peak period $T$ to the back of queue, the conversion formulas are

$$t_{ci} = \frac{V_u}{V_u + w_c} t_{ci} - \frac{1 - K_u^T}{1 - V_n^T/V_u} t_{ci} \quad \text{or} \quad t_{ci} = t_{A0}^T + (1 - i)\cdot \tau_T^{ci} = t_{ci} - 1 + \tau_T^{ci},$$

after the first arrival at the back of queue, up to $T - \varepsilon_T$ at the back of queue (the last arrival with $t_{ci} = T - \varepsilon_T$ for $i = \varepsilon_T$ should deliver

$$t_{ci} = t_{ci} - \frac{T - \tau_T^i}{T} \cdot \varepsilon_T$$

with $\varepsilon_T = \frac{Z_{AW}}{V_u}$ by cumulating relative headways $\tau_T^c = \frac{1 - K_u^T}{1 - V_n^T/V_u} \cdot h_T^A$ for $h_T^A = \frac{1}{Q_A}$

and generating a transition synchronization with $t_{ci}^{A0} = T - \varepsilon_T + \frac{V_u}{V_u + w_c} h_{A0}^T = T - \varepsilon_T + \frac{1 - K_u^T}{1 - V_n^T/V_u} \cdot h_{A0}^T$

with $h_{A0}^T = \left(1 - \frac{h_{A0}^T}{h_A}\right) h_{A0}^T$ as before). The same process applies after adjusting traffic variables in $T_n$ to $T + t_p$ or $T + t_{pi}$ for Case I or up to $T + t_q$ for Case II) and then to

$$\tau_T^n = \frac{V_u}{V_u + w_c} h_A^n = \frac{1 - K_u^n}{1 - V_n^n/V_u} h_A^n$$

at the back of queue (after $T$ up to $T$ + $t_p$ or $T$ + $t_{pi}$ for Case I or $T$ + $t_q$ for Case II) or

$$\tau_T^n = \frac{V_u}{V_u - w_c} h_A^n = \frac{1 - K_u^n}{1 - V_T^n/V_u} h_A^n$$

at the back of the free discharging queue (from $T$ + $t_p$ up to $T$ + $t_{pi}$ for Case I) or

$$\tau_T^n = \frac{V_u}{V_u - w_c} h_A^n = \frac{1 - K_u^n}{1 - V_S^n/V_u} h_A^n$$

for Case II only). The procedure amounts to profiling the back of queue by tracking $W_c^T$ as $\Delta T_A^T = W_c^T \cdot \tau_T^c$ or

$$\Delta n_T^A = f_c^T \cdot \Delta n_A^T = \frac{(Q_A - q_A)^T \cdot \tau_T^c}{1 - K_u^T/K_n^T}$$

during the peak period $T$, then tracking $W_T^n$, $W_T^n$ as $\Delta z_A^n = -W_T^n \cdot \tau_T^n$ and

$$\Delta z_A^n = -W_T^n \cdot \tau_T^n \quad \text{or} \quad \Delta n_A = -f_c^n \cdot \Delta n_A^n = \frac{(Q_A - q_A^n)^T \cdot \tau_T^n}{1 - K_u^n/K_n^n} \quad \text{and} \quad \Delta n_A = -f_c^n \cdot \Delta n_A^n = \frac{(Q_A - q_A^n)^T \cdot \tau_T^n}{1 - K_u^n/K_n^n}$$

for

Case I or $W_T^n$, $W_T^n$ as $\Delta z_A = W_T^n \cdot \tau_T^n$ and $\Delta z_A = -W_T^n \cdot \tau_T^n$ or $\Delta n_A = f_c^n \cdot \Delta n_A^n = \frac{(Q_A - q_A^n)^T \cdot \tau_T^n}{1 - K_u^n/K_n^n}$ and

$$\Delta n_A = -f_c^n \cdot \Delta n_A^n = \frac{(Q_A - q_A^n)^T \cdot \tau_T^n}{1 - K_u^n/K_n^n}$$

for Case II (generically).
Correspondingly for departures of the peak period T from the front of the queue, the conversion formulas are \( t_{Di} = \tilde{t}_{Di} \) along T (as for any period with a stationary front of queue at A, with \( q_a = q_A \) measured at A) and \( t_{Di} = T + \frac{V_{Tn}^{a}}{V_{S}^{a} + w_{s}^{Tn}} \left( \tilde{t}_{Di} - T \right) \) with the first departure during queue dissipation on \( T_n \) at \( t_{D_{i}^{d} +1} = T + \frac{V_{Tn}^{a}}{V_{S}^{a} + w_{s}^{Tn}} h_{S0}^{Tn} = T + \frac{1 - K_{S}^{Tn}/K_{n}^{T}}{1 - V_{n}^{T}/V_{S}^{T}} (\tilde{t}_{Di} - T) \) and (incrementally) subsequent departures after \( t_{D_{i}^{d}} = \frac{V_{S}^{Tn}}{V_{S}^{Tn} + w_{s}^{Tn}} h_{S}^{Tn} = \frac{1 - K_{S}^{Tn}/K_{n}^{T}}{1 - V_{n}^{T}/V_{S}^{T}} h_{S}^{Tn} \) (from T up to \( T + \tilde{t}_{s} \) at the front of queue for Case II, while maintaining \( t_{Di} = \tilde{t}_{Di} \) for Case I). The procedure amounts to profiling the front of queue by tracking

\[ \Delta z_{Si} = w_{S}^{Tn} \tau_{D_{i}^{d}} \] or \( \Delta \tilde{n}_{Si} = f_{cn} \Delta \tilde{n}_{Si} = \left( q_{A}^{Tn} - q_{n}^{T} \right) \tau_{D_{i}^{d}} \] (taking generically \( w_{S}^{Tn} = 0 \) for Case I; a similar procedure can be used for profiling the internal change of queueing by tracking \( w_{T_{i}}^{Tn} \), if desired, even if not needed for the analysis of queueing). After \( T + \tilde{t}_{s} \) it holds that \( t_{c_{i}} = \tilde{t}_{c_{i}} = t_{Di} = \tilde{t}_{Di} \). Note that the dissipation wave at the front of queue \( w_{S}^{Tn} \) also describes the build up of the free discharging “queue”, that is cleared afterwards during the normalization of traffic (by the normalization wave \( w_{m}^{Tn} \)), but no further detail is needed for evaluating delays (as long as the free queue discharging speed \( V_{S}^{Tn} \leq V_{u} \) is considered along \( z_{Si} \)).

In general, queueing and delay measures can be derived from aggregate representations of these path trajectories. The conventional representation for VQMs is the queueing area, between the back and front of queue, evaluated as

\[ \widetilde{A}_{n} = \int_{T_{i}^{x}}^{T_{i}^{y}} \tilde{n}[t] dt = \int_{0}^{T_{i}^{x}} \tilde{n}[t] dt = \sum_{i=1}^{N_{y_{i}}} t_{n_{i}}^{y_{i}} = \sum_{i=1}^{N_{y_{i}}} t_{n_{i}}^{x_{i}} + \sum_{k=1}^{N_{k}} t_{n_{k}}^{x_{k}} \] (24a,b,c,d,e)

\[ \text{(using} \int_{0}^{T_{i}^{x}} \tilde{n}[t] dt \text{) =} \int_{0}^{T_{i}^{x}} \sum_{i=1}^{N_{y_{i}}+1} \delta_{n_{i}}[t] dt = \sum_{i=1}^{N_{y_{i}}+1} \int_{0}^{T_{i}^{x}} \delta_{n_{i}}[t] dt \text{) =} \sum_{i=1}^{N_{y_{i}}+1} t_{n_{i}}^{x_{i}} \text{ for the indicator function} \]

\[ \delta_{n_{i}}[t] \text{ that assumes the value 1 if vehicle i is queueing at time t or the value 0 otherwise) with} \]

\[ \tilde{n}[t] = \int_{0}^{T_{i}} (Q[t] - q[t]) dt = \int_{0}^{T_{i}} Q[t] dt - \int_{0}^{T_{i}} q[t] dt = N_{C}[t] - N_{D}[t] \] (25a,b,c)

and

\[ t_{n_{i}} = \tilde{t}_{D_{i}} - \tilde{t}_{c_{i}} = \tilde{d}_{n_{i}} \] (26a,b)
(e.g. as related to the definitions in Daganzo, 1997, pg.26-29; where $t_{ni}$ is the time in queue of vehicle $i$, assuming a queue discipline without overtakings, and $t_{nk}$ is the time with at least $k$ vehicles in queue), immediately delivering the average queueing delay in vertical queueing models as

$$\bar{d}_n = \frac{\bar{A}_n}{N_{T+Tn}} \quad \text{(per vehicle)} \quad (27a)$$

(by 24c,d) and the average vehicular queue in $T+Tn$ for vertical queueing models as

$$\bar{n}_t = \frac{\bar{A}_n}{T+Tn} \quad \text{(in time)} \quad (28a)$$

(by 24a), meaning that

$$\bar{d}_n = \frac{\bar{n}_t}{q_{in}} \quad \text{(per vehicle)} \quad (29a)$$

with

$$\bar{q}_n = \frac{N_{T+Tn}}{T+Tn} = Q_A^T \cdot \frac{T}{T+Tn} + Q_{An}^T \cdot \frac{Tn}{T+Tn} = q_A^T \cdot \frac{T}{T+Tn} + q_{An}^T \cdot \frac{\bar{t}_n}{T+Tn} + Q_{An}^T \cdot \frac{Tn-\bar{t}_n}{T+Tn} \quad \text{(for } Tn \geq \bar{t}_n),$$

dividing the periods in a saturated part $t_s = T + \bar{t}_s$ with $\bar{n} > 0$ and an unsaturated part $t_u = Tn - \bar{t}_s$ with $\bar{n} = 0$ where

$$\bar{d}_{ns} = \frac{\bar{A}_{ns}}{N_{T+Tn}} \quad \text{and} \quad \bar{d}_{ts} = \frac{\bar{A}_{ts}}{T+\bar{t}_s},$$

and

$$\bar{d}_u = \frac{N_{T+is}}{N_{T+Tn}} \bar{q}_{ns}, \quad \text{as} \quad \bar{d}_u = 0, \quad \bar{n}_{su} = 0 \quad \text{for delay from overflow demand).}$$

For VQMs, simplified results are intuitively obtained only for the pure Case I from

$$\bar{A}_n = \frac{\bar{n}_{AT}}{2} \left( T + \bar{t}_s \right) \quad (24e)$$

where

$$\bar{n}_{AT} = \left( Q_A^T - q_A^T \right) T \quad \text{and} \quad \bar{t}_s = \frac{\bar{n}_{AT}}{q_{Tn}^T - Q_{An}^T} \quad \text{with} \quad q_{Tn}^T = q_A^T, q_{An}^T \sim q_A^T, q_{Tn}^T = q_n^T \quad \text{(the extension to mixed Case I and to Case II patterns are discussed ahead, after detailing the picture drawn from HQMs).}$$

However, even in this simple setting, a basic point merits further discussion. The average queue face by vehicles arriving during the peak period $T$ is

$$\bar{n}_T = \bar{n}_{AT} / 2 = \bar{n}_z^T \quad \text{(as the advance in queue needed to leave queueing) and the average time spent in queue is} \quad \bar{t}_z^T = \bar{n}_z^T / q_n \quad \text{while the average queue face by vehicles arriving during the next period $Tn$ is} \quad \bar{n}_{Tn} = \bar{n}_{AT} / 2, \bar{t}_s / Tn = \bar{n}_{Tn}^z \quad \text{(similarly defined, as there is no queue after $\bar{t}_s$ during period $Tn$) and the average time spent in queue is} \quad \bar{t}_z^T = \bar{n}_z^T / q_n \quad \text{(as $q_{Tn}^T \sim q_A^T$ is taken as $q_{Tn}^T = q_n^T = q_n$ in pure Case I).}$$

Then, in the full period $T + Tn$ (for $Tn > \bar{t}_s$), the (time) average queue (faced at arrival) is

$$\bar{n}_t = \frac{\bar{A}_n}{T+Tn}$$
\[ \tilde{n}_t = \frac{T}{(T + T_n)} \tilde{n}_t^T + \frac{T_n}{(T + T_n)} \tilde{n}^T_{T_n} = \left( T + \tilde{t}_s \right) / (T + T_n) \tilde{n}_{AT} / 2, \]

while the average queue experienced by vehicles is

\[ \tilde{n}_z = \frac{N_{T} / N_{T+T_n}}{\tilde{n}_z^T} + \frac{N_{T} / N_{T+T_n}}{\tilde{n}_{T_n}^T} \frac{\tilde{n}_{z}^T}{\tilde{n}_{T_n}^T} = \frac{N_{T+T_n}}{N_{T+T_n}} \tilde{n}_{AT} / 2 \]

(then \( \tilde{n}_i = \tilde{q}_n / \tilde{q}_{Tn} \tilde{n} \) where

\[ \tilde{q}_{ns} = \frac{N_{T+T_n}}{(T + \tilde{t}_s)} = \tilde{q}_n \quad \text{and} \quad \tilde{a}_n = \frac{N_{T+T_n}}{(T + T_n)} = T / ((T + T_n) Q_A^T + T_n / (T + T_n) Q_A^T) \]

\[ \tilde{q}_n = \left( T + \tilde{t}_s \right) / (T + T_n) = \tilde{q}_{ns} + \left( T_n - \tilde{t}_s \right) / (T + T_n) Q_A^T \]

with \( \tilde{q}_{ns} = \tilde{q}^T_{n} / \tilde{q}_{n} = q_n^T / q_n \), meaning that the average time spent in queue as experienced by vehicles (the queueing delay in VQMs) is simply

\[ \tilde{t}_n = \frac{N_{T}/N_{T+T_n}}{\tilde{t}_n^T} + \frac{N_{T}/N_{T+T_n}}{\tilde{t}_{T_n}^T} = \frac{\tilde{n}_z / \tilde{q}_n = \tilde{n} / \tilde{q}_n}, \]

if using \( \tilde{n}_i = \frac{\tilde{A}_n}{T + T_n} \) with \( \tilde{a}_n = \frac{N_{T+T_n}}{T + T_n} \), by the same vein for the full period. Note that an alternative weighting based on the outflow of each period is better justified (otherwise outflows should be viewed as future values, perhaps subjected to revision), but delivers the same average when treating the full period (i.e. in \( T + T_n \) for \( T_n > \tilde{t}_s \), as

\[ N_{T+T_n} = Q_A^T T + Q_A^T T_n \tilde{t}_s = Q_A^T T + q_{T_n} T + q_{T_n} T_n \tilde{t}_s \]

as well). In general, note also that \( \tilde{q}_n^T, \tilde{q}_{n}^T \) can be endogenous (e.g. functions of queueing measures and other traffic conditions).

Conceptually, path tracing delay measures should be inflow-based values. Average delay from queue sampling is an option and can be related to inflow-based values. If a short period \( \Delta T \) sees \( Q \Delta T \) arrivals and \( q \Delta T \) departures and the queue evolution \( \tilde{n}_T = \tilde{n}_0 + (Q - q) \Delta T \), the queue sampling average delay is directly obtained as

\[ \bar{d}_T = \frac{\tilde{A}_n \Delta T}{q \Delta T}, \tilde{A}_{n \Delta T} = \frac{\tilde{n}_T + \tilde{n}_0}{2} \Delta T \]

and can be translated into the (forecasted) path tracing average delay as

\[ \bar{d}_T = \frac{\bar{d}_0 + \bar{d}_T}{2}, \bar{d}_0 = \frac{\tilde{n}_0}{q}, \bar{d}_T = \frac{\tilde{n}_T}{q} \]

for the inflow vehicles (under prevailing traffic conditions) of \( \Delta T \).

But the queue sampling formula have to be properly set as

\[ \bar{d}_T = \frac{\tilde{n}_T + \tilde{n}_0}{2 q_{\Delta T}}, \bar{q}_{\Delta T} = q_{\Delta T} \]

be translated to

\[ \bar{d}_T = \frac{\bar{d}_0 + \bar{d}_T}{2}, \bar{d}_0 = \frac{\tilde{n}_0}{q}, \bar{d}_T = \frac{\tilde{n}_T}{q}, \]

or

\[ \bar{d}_T = \frac{1}{2} \left( \frac{\tilde{n}_0}{q_{\Delta T}} + \frac{\tilde{n}_T}{q_{\Delta T}} \right) \]

where \( q_{\Delta T} \) is the average outflow during the period \( \Delta T \) and \( \bar{q}_{Tn} \) (or \( \bar{q}_{Tn}^{0}, \bar{q}_{Tn}^{n} \) for the first and last vehicles arriving in the period \( \Delta T \)) is the average outflow during the queueing time (then results in sub-periods of \( T \) or \( T_n \) can be different only if \( q_n \) changes along the period).

In this view, queue sampling estimates of delay can be used as estimates of the (path tracing) delay of inflow vehicles (i.e. of vehicles arriving during \( \Delta T \)) but should be subsequently revised by
\[
\Delta d_T = \frac{n_0 + n_T}{2} \left( \frac{1}{q_{Tn}} - \frac{1}{q_{\Delta T}} \right) \quad \text{or} \quad \Delta \bar{d}_T = \frac{1}{2} \left( \bar{n}_0 \left( \frac{1}{q_{n0}} - \frac{1}{q_{\Delta T}} \right) + \bar{n}_T \left( \frac{1}{q_{nT}} - \frac{1}{q_{\Delta T}} \right) \right)
\]

at the departure time from the queue if there was a change in outflow. It is clear that inflow-based values, even if conceptually preferred, must be a prediction while outflow-based values (as well as queue sampling measures) can be directly evaluated.

The corresponding representation for HQMs, as shown in Figure 1, for ordinates taken in vehicles, is

\[
A_n = \int_0^{T+T_n} n[t] \, dt = \int_0^{T+T_n} n[t] \, dt \cong \sum_{i=0}^{N_{T-T_n}} t_{ni} = \sum_{i=1}^{N_{T-T_n}} t_{ni} + \sum_{k=1}^{n_{\Delta T}} t_{nk} \quad (30a,b,c,d,e)
\]

and

\[
t_{ni} = t_{Di} - t_{Ci} \quad (31a)
\]

(where \(t_{ni}\) is the time spent in queue of vehicle \(i\) and \(t_{nk}\) is time of queueing with at least \(k\) vehicles), immediately delivering the average queueing time in \(T+T_n\) for horizontal queueing models (instead of the average queueing delay in vertical queueing models) obtained as

\[
\bar{t}_n = \frac{A_n}{N_{T-T_n}} \quad \text{(per vehicle)} \quad (32a)
\]

(by 30c,d) and the average vehicular queue in \(T+T_n\) for horizontal queueing models is

\[
\bar{n}_i = \frac{A_n}{T+T_n} \quad \text{(in time)} \quad (33a)
\]

(by 30a), also meaning that

\[
\bar{t}_n = \frac{\bar{n}_i}{q_{\Delta n}} \quad \text{(per vehicle)} \quad (34a)
\]

with \(q_{\Delta n} = \frac{N_{T-T_n}}{T+T_n} = Q_A^T \cdot \frac{T}{T+T_n} + Q_{A_n}^{Tn} \cdot \frac{Tn}{T+T_n} = q_{A}^T \cdot \frac{T}{T+T_n} + q_{A}^{Tn} \cdot \frac{Tn}{T+T_n} + \frac{T}{T+T_n} + \frac{Tn}{T+T_n} \) (for \(Tn \geq t\)), similarly dividing the periods in a saturated part \(T_s = T + t\) with \(n > 0\) and an unsaturated part \(T_u = Tn - t\)

with \(n = 0\) where \(\bar{t}_{ns} = \frac{A_n}{N_{T-T_n}}\), \(\bar{n}_{ns} = \frac{A_n}{T + t}\), then delivering \(\bar{t}_n = \frac{N_{T+T_n}}{N_{T+T_n}} \cdot \bar{t}_{ns}\), as \(\bar{t}_{nu} = 0\) and \(\bar{n}_{ns} = 0\) for queueing from overflow demand, from \(\bar{t}_n = \frac{N_{T+T_n}}{N_{T+T_n}} \cdot \bar{t}_{ns}\) equivalently). For HQMs, the general formulas also apply to for the full period (difference of time and vehicle averages and transient periods are considered ahead).

An alternative representation for HQMs is the corresponding space-time queueing area, between the back and front of queue in a space-time frame, taking Figure 1 for ordinates in queue length, being
(where $\overline{c}_{ni}$ is the average spacing in queue for vehicle i and $\overline{c}_{nk}$ is the average spacing for the k-th vehicle in queue), immediately delivering the average space-time in queue per vehicle in $T+T_n$ as

$$A_z = \int_0^{T+T_n} z[t] \, dt = \int_0^{T+T_f} z[t] \, dt = \sum_{i=1}^{N_{t+T}} \overline{c}_{ni} \cdot I_{t+i} = \sum_{i=1}^{N_{t+T_n}} \overline{c}_{ni} \cdot I_{t+i} = \sum_{k=1}^{n_{AT}} \overline{c}_{nk} \cdot I_{nk}$$  (35a,b,c,d,e)

and the average queue length in $T+T_n$ as

$$\overline{z}_t = \frac{A_z}{N_{T+T_n}} \text{ (per vehicle)}$$  (36a)

(by 35a and previous results), where $\overline{z}_t = \overline{a}_z \cdot \overline{q}_n$ with

$$\overline{q}_n = \frac{N_{T+T_n}}{T+T_n} = Q^T_A \cdot \frac{T}{T+T_n} + Q^{T_n}_A \cdot \frac{T_n}{T+T_n} = \frac{q^T_A}{T+t_f} + \frac{q^{T_n}}{T+t_f} + Q^T_A \cdot \frac{T_n-t_f}{T+T_n} \quad \text{(also } \overline{a}_{zs} = \frac{A_z}{N_{T+t_f}})\)$$

$$\overline{z}_{ts} = \frac{A_z}{T+t_f} = \overline{a}_{zs} \cdot \overline{q}_n, \quad \text{with } \overline{q}_n = \frac{N_{T+T_n}}{T+t_f} = Q^T_A \cdot \frac{T}{T+t_f} + Q^{T_n}_A \cdot \frac{T_n}{T+t_f} = \frac{q^T_A}{T+t_f} + \frac{q^{T_n}}{T+t_f} + \frac{t_f}{T+t_f} \quad \text{and the real average speed and density in queue } V_{ns}, K_{ns}, \text{ including the period without queues (in which queueing density and speeds are undefined). More importantly, the space-time average density in queue during saturation is }

$$K_{ns} = \frac{A_n}{A_z}$$  (38a)

(as $K_{ns} = \frac{1}{\overline{c}_{ns}}$, with $\overline{c}_{ns} = \frac{A_z}{A_n}$ by 35e and 30e), delivering $V_{ns} = \frac{q_{ns}}{K_{ns}}$ as the (real) space-time average speed in queue during saturation ($\overline{V}_{ns} = V_{ns}, \overline{K}_{ns} = K_{ns}$ if steady). Then, the average queueing time during $T+T_n$ can be used for obtaining the corresponding average queueing density and speed that warrants $\overline{t}_n = \frac{N_{T+t_p}}{N_{T+T_n}} \cdot \overline{t}_n$

per vehicle (as the queueing time is null outside $T+t_p$). Note that $A_n = \overline{K}_{ns}, A_z = \overline{K}_{n_z}, A_z$ requires

$$\overline{K}_n = \overline{K}_{ns}$$  (39a)

and the validity of the formula

$$\overline{t}_n = \frac{\overline{z}_t}{\overline{v}_n} \text{ (per vehicle)}$$  (34b)
requires that space-time apparent speed in queue during $T+T_n$ is $\nabla_n = \frac{q_n}{K_n}$ during $T+T_n$, as

$$\bar{\tau}_n = \frac{\bar{q}_n}{\bar{q}_n} = \frac{A_n/(T+T_n)}{\bar{q}_n/K_n} = \frac{Z_n}{V_n} \text{ for } \nabla_n = \frac{q_n}{K_n} \text{ and } \bar{\tau}_n = \frac{N_{T+T_n}}{q_n}(T+t_f) = \frac{A_z/(T+t_f)}{q_{ns}/K_{ns}}$$

for $\bar{\tau}_n = \frac{\bar{q}_n}{\bar{q}_n}$ (also $Z_n = \frac{A_z}{T+t_f}$, $q_{ns} = \frac{\bar{q}_n}{K_n}$ and $\tau_{ts} = \frac{\bar{q}_n}{\bar{q}_{ns}} = \frac{A_n/(T+t_f)}{q_{ns}/K_{ns}} = \frac{Z_n}{V_{ns}}$). The option of defining $K_n = K_{ns}$ seems to be artificial (as much as $V_n = V_{ns}$ alternatively) but more convenient. Note, however, that $\bar{\tau}_n \leq \tau_{ts}$, $\bar{\tau}_n \leq \tau_{ts}$, $\bar{\tau}_n \leq \tau_{ts}$ and $\bar{\tau}_n \leq \tau_{ts}$ could imply ambiguously $\nabla_n \leq \nabla_{ns}$ (instead of consistently $\nabla_n \geq \nabla_{ns}$, by increasing the proportion of non-queueing vehicles only).

For HQMs, several questions remain as the relation of time and vehicle averages, of queue faced at arrival and distance travelled in queue or the development of formulas for transient periods. As before, the detailed treatment of the speed of transitions (i.e. change of traffic conditions as represented by waves of FOTM or any other form) have the clear potential of distinguishing several patterns of queue formation and dissipation. But the peculiarities of Case I and Case II patterns are not explicit in general formulas (i.e. except for features implicitly embodied into the queueing profiles), meaning that a careful analysis can be justified.

In general, there is a clear aggregation method for vehicle averages and queueing area measures (e.g.

$$\bar{\tau}_n = \sum_k \tau_{nk} \frac{N_{Tk}}{N_T} = \sum_k \frac{A_{nk}}{N_{Tk}} \frac{N_{Tk}}{N_T} = \sum_k \frac{A_{nk}}{N_{Tk}} = \frac{A_n}{N_T}, \text{ perhaps distinguishing traffic conditions that gives}$$

$$\bar{\tau}_{nk} = \bar{\tau}_{nk}^0 + \bar{\tau}_{nk}^1 = \frac{A_{nk}}{N_{Tk}} + \frac{A_{nk}}{N_{Tk}} = \frac{A_{nk}}{N_{Tk}} \text{ with } A_{nk} = A_{nk}^0 + A_{nk}^1 \text{ for a common vehicle group } N_{Tk} = N_{Tk}^0 = N_{Tk}^1 \text{ ) as}$$

long as the queueing frame is split into a consistent partition of vehicle groups $k$ referred to arrival periods, departure periods or other partitions with $N_T = \sum_k N_{Tk}$ (or more generally

$$\bar{\tau}_n = \bar{\tau}_n^0 + \bar{\tau}_n^1 = \sum_k \frac{A_{nk}^0}{N_{Tk}^0} \frac{N_{Tk}^0}{N_T} + \sum_k \frac{A_{nk}^1}{N_{Tk}^0} \frac{N_{Tk}^1}{N_T} = \sum_k \frac{A_{nk}^0}{N_{Tk}^0} \frac{N_{Tk}^0}{N_T} + \sum_k \frac{A_{nk}^1}{N_{Tk}^0} \frac{N_{Tk}^1}{N_T} = \frac{A_n}{N_T}$$

as long as $N_T = \sum_k N_{Tk}^0 = \sum_k N_{Tk}^1$, if the distinction of traffic conditions commands different partitions) that do not translate generically to time averages (or space-time areas). Otherwise, any time splitting (either using arrival periods, departure periods or other partitions) can be made consistent by an aggregation procedure based on the space-time frame and on weighting by time proportion (e.g. $Z_t = \frac{A_z}{T} = \sum_k Z_{tk} \frac{T_k}{T}$ for $Z_{tk} = \frac{A_{nk}^1}{T_k}$ with
T = \sum_k T_k \text{ and } A_z = \sum_k A_{zk}, \text{ perhaps with } A_{zk} = A_{zk}^0 + A_{zk}^1 \text{ but assuming common time parts}

T_k^0 = T_k^1 = T_k, \text{ or } Z_i = \frac{A_{zk}^0 + A_{zk}^1}{T} = Z_i^0 + Z_i^1 = \sum_k Z_{ik}^0 \frac{T_k^0}{T} + \sum_k Z_{ik}^1 \frac{T_k^1}{T} \text{ for } Z_{ik}^0 = \frac{A_{zk}^0}{T_k^0}, Z_{ik}^1 = \frac{A_{zk}^1}{T_k^1} \text{ with}

T = \sum_k T_k^0 = \sum_k T_k^1 \text{ and } A_z^0 = \sum_k A_{zk}^0, A_z^1 = \sum_k A_{zk}^1, \text{ consistently but otherwise arbitrarily or at least flexibly) that do not translate generically to vehicle averages (or queueing areas). But the consistency does not mean that average delay measures corresponding to each part of the full time period are always correct (either for time average that is appropriate to each part of the time period or to its related vehicle average). The following discussion aims at identifying and understanding meaningful criteria from a detailed analysis.}

For Case I, the average formulas can be developed after obtaining the queueing area between the back and front of queue, again based in Figure 1b. In pure Case I, the queueing area is

A_n = \left(T + t_p\right) \cdot \frac{n_{AT}}{2} \hspace{1cm} (30f)

or A_n = \left(T + t_p\right) \cdot \frac{n_{AW}}{2} \text{ if considering the time translation of periods (for } T_n \geq t_p), \text{ with } \bar{t}_n = \frac{A_n}{N_{T+n}} \text{ per vehicle and } \bar{n}_t = \frac{A_n}{T + T_n} \text{ at arrival. The previous expression can be simply applied if } q_A^T \sim q_A^{Tn} \text{ can be assumed (as } q_A^T, q_A^{Tn} \text{ are usually similar, then } K_n^T, K_n^{Tn} \text{ and } V_n^T = \frac{q_A^T}{K_n^T}, V_n^{Tn} = \frac{q_A^{Tn}}{K_n^{Tn}} \text{ are usually similar too). As every vehicle travels the queue it faces at arrival, in each period } \bar{n}_t^T = \bar{n}_t = \frac{n_{AT}}{2}, \bar{Z}_t^T = Z_t^T = \frac{Z_{AT}}{2} \text{ and}

\bar{n}_z^T = \bar{n}_z = \frac{n_{AT}}{2}, \bar{Z}_z^T = \bar{Z}_z = \frac{Z_{AT}}{2} \text{ (during } t_p \text{ the saturated part of } T_n), \text{ as the front of queue is static at } A \text{ and arrivals are uniform in each period, meaning that } \bar{n}_{zs} = \bar{n}_{zs} = \frac{n_{AT}}{2}, \bar{Z}_{zs} = \bar{Z}_{zs} = \frac{Z_{AT}}{2} \text{ and } \bar{t}_{ns} = \frac{\bar{V}_n}{V_n} = \frac{\bar{n}_{zs}}{q_n} \text{ with}

\bar{Z}_{ns} = \frac{\bar{n}_{zs}}{K_n} \text{ applies with the assumed queueing conditions (with } K_{ns} = K_{ns}^T, \bar{V}_{ns} = V_{ns}^T \text{ and } \bar{q}_{ns} = q_{ns}^T \text{ as the real traffic conditions during queueing). In the overall period, for } T_n \geq t_p, \text{ then } \bar{t}_n = \frac{N_{T+t_p}}{N_{T+n}} \cdot \bar{t}_{ns} \text{ deliver the usual}
HQM estimates $\bar{t}_n = \overline{\bar{t}}_n = \frac{\overline{n}}{\overline{q}_n}$ for $\overline{q}_n = \frac{N_{Tn+T_n}}{T + T_n}$ or $\bar{t}_n = \overline{\bar{t}}_n = \frac{\overline{z}}{\overline{V}_n}$ for $\overline{V}_n = \frac{\overline{q}_n}{K^{n^T}}$. But generically (as in mixed Case I) similar expressions can be applied if a consistent average flow and speed $\overline{q}_n$, $\overline{V}_n$ (also $\overline{q}_{ns}$, $\overline{V}_{ns}$ in queue, during the saturated period $T + t_p$) in $T + T_n$ (or $T + t_p$) can be used, weighted for

$$N_{Tn+T_n} = N_{T+T_n} + N_{T-tp}$$

with $N_{T+T_n} = q_A^T + q_A^{tn} T_p$ and $N_{T-tp} = Q_A^T (T_n - t_p)$ or a more detailed breakdown of demand or flow (also $N_{Tn+Tn} = \hat{n}_{AT0} + Q_A^t T + Q_{A^n} T_n - n_{ATn}$, with $\hat{n}_{AT0}, \hat{n}_{ATp}$ as the number of free flowing vehicles along the length $z_{AT}$ before and after queueing, as

$$N_{T+T_n} = q_A^T T + q_A^{Tn} T_n = \hat{Q}_A^T T + \hat{Q}_A^{Tn} T_p = \left( Q_A^T + K_u^T z_{AT} \right) T + \left( Q_A^T - K_u^T z_{AT} \right) T_p$$

and

$$N_{T+T_n} = Q_A^T T + K_u^T z_{AT} + Q_A^{Tn} T_n - K_u T_n z_{AT} = Q_A^T T + \hat{n}_{AT0} + Q_A^{Tn} T_n - \hat{n}_{ATp}, \text{ and } \hat{n}_{ATn} = \hat{n}_{ATp}, t > t_p, \text{ if}$$

preferring measures based on arrivals), meaning that a space-time average is in general needed (as implied by aggregation of individual vehicle delay, that consider the distance travelled in queue $z_{AT}$ under distinct queueing conditions, perhaps adjusted by the queueing movement and its interaction with the internal transition wave).

Again in pure Case I, the area between the back and front of queue in a space-time framework is

$$A_x \approx \left( T + t_p \right) \frac{Z_{AT}}{2}$$

(35f)

or $A_x \approx \left( T + t_p \right) \frac{Z_{AT}}{2}$ if considering the time translation of periods (for $Tn \geq t_p$) with $\overline{a}_z = \frac{A_x}{N_{Tn+Tn}}$ per vehicle

and $\overline{z}_1 = \frac{A_x}{T + T_n}$ at arrival. In pure Case I, as $n_{AT} = K_u^T z_{AT}$ and $A_x = K_u^T A_z$, then $\overline{z}_1 = \frac{\overline{n}_n}{K^{n^T}_n}$ and

$$\overline{t}_n = \overline{\overline{t}}_n = \overline{\overline{z}}_n.$$ as before. But a general average is obtained in the space-time framework using

$$K_{ns} = \frac{A_n}{A_z} = K_n \text{ with (apparent) speed } \overline{V}_n = \frac{\overline{q}_n}{K^{n^T}_n} \text{ in } T + T_n \text{ or (real) speed } \overline{V}_{ns} = \frac{\overline{q}_{ns}}{K^{n^T}_{ns}} \text{ during } T + t_p, \text{ that}$$

can also handle mixed cases by implicitly taking the variations in densities when averaging $A_n$ by $A_z$ as

$$\overline{t}_n = \frac{\overline{n}_n}{\overline{q}_n} = \frac{\overline{z}_n}{\overline{V}_n} \text{ (implicitly averaging } K^{n^T}_n, K^{Tn}_{ns} \text{ as well). A spatial average queueing time in sections of a link}$$
length $L$ is $\overline{t}_{\text{at}} = \frac{A_z}{L} = \frac{T + t_p}{2} \cdot \frac{Z_{\text{AT}}}{L}$ for $L \geq Z_{\text{AT}}$ (or $\overline{t}_{\text{at}} = \frac{A_z}{Z_{\text{AT}}} = \frac{T + t_p}{2}$ in sections of $Z_{\text{AT}}$) and a temporal average queueing time in a section $z$ is $\overline{t}_{\text{at}} = \frac{Z_{\text{AT}}}{Z_{\text{AT}}} \cdot (T + t_p)$ for $z \leq Z_{\text{AT}}$ (as used for evaluating blocked time at $z$).

In the mixed Case I, $A_n$ and $A_z$ can be obtained by summing its parts (as split by time periods and the internal wave $w_i^{\text{Tn}}$). If ignoring the time translation of periods, $A_n = A_n^0 + A_n^1 = \frac{n_{\text{AT}}}{2} \cdot (T + t_p) + \frac{n_{\text{At}}}{2} \cdot t_p$ and

$A_z = A_z^0 + A_z^1 = \frac{Z_{\text{AT}}}{2} \cdot (T + t_p) + \frac{Z_{\text{At}}}{2} \cdot t_p$ (in a simple additive form), where $A_n^0 = K_n^T A_z^0$ and $A_n^1 = K_n^{\text{Tn}} A_z^1$ are used for weighting by densities in the space-time framework (with no concern about the relative flows to the internal wave $w_i^{\text{Tn}}$ as the effect cancels-out). Ideally aggregation should represent vehicle trajectories. Under uniform conditions in $T$ and $T_n$, time in queue under conditions 0 and 1 (as $t_{\text{mt}}$, $t_{\text{nt}}$ for a vehicle arriving at $t$) can be linearly interpolated between inflection points. The discussion ahead then focuses on sub-periods between the inflection points $t_x$, $t_y$ and $t_w$ of Case I. In all trajectories, vehicles remain queueing from arrival to the queue up to passing by section $A$, meaning that the queue extent faced at arrival must be fully travelled on the trajectory, but queueing conditions can change if meeting the intermediate wave $w_i$ that transmits the change of outflow inside the queue (from $q_n^T = q_n^T$ to $q_n^{\text{Tn}} = q_n^{\text{Tn}}$ if the supposition $q_n^T \sim q_n^{\text{Tn}}$ is violated).

For Case II, the relation between measures using the HQMs can be obtained after a correct measure of queue is adopted, and evaluated with the analysis of shock waves either from the tracing of vehicle paths or from the queue sampling from queueing diagrams. Based on the previous discussion, it is easy to conclude that the correct measure of queue at arrival is $n_t = n_{\text{At}} - \bar{n}_{\text{St}}$, at least after the peak period $T$ (before $T$, it can be assumed that $\bar{n}_{\text{St}} = 0$) and before queue dissipation at $T + t_q$ (after $T + t_q$ it can be assumed that $n_t = 0$), meaning that free queue discharging vehicles (\(\bar{n}_{\text{St}}\)) are not taking as queueing. Then there are major differences to consider: the movement of queue front on the time spent advancing along the queue (in $T_n$) and the operation regime during free queue discharging (another source of delay if $V_{\text{St}}^{\text{Tn}} < V_{\text{St}}^{T_n}$). Under uniform conditions in $T$ and $T_n$, time in queue and time in free queue discharging as $t_{\text{mt}}$, $t_{\text{nt}}$ (for a vehicle arriving at $t$) can be linearly interpolated between inflection points. Tracking shock waves has now the added task of
separating queueing delay and non-queueing delay, when present (i.e. at least if $V_s^{T_n} < V_u^{T_n}$). The time spent on queue is $t_{nt} = \frac{n_z}{q_n} = \frac{z}{\hat{V}_n}$, with convenient averaged flow and speed as seen at the moving front queue frontier where $\hat{q}_n = q_n$ and $\hat{V}_n = V_n$ for the fixed front queue during $T$ but

$\hat{q}_n = K_s^T \hat{V}_n = K_s^T \left( V_s^{T_n} + w_s^{T_n} \right) = q_A^T + K_s^T \cdot w_s^{T_n}$ and $\hat{V}_n = V_s^{T_n} + w_s^{T_n}$ for the moving front queue along $t_q$ (both taken during the time in queue for the vehicle that joined the queue at $t$ while advancing along the queue until leaving the queue at $t_i$ in $t_q$). As previously discussed, time in queue can also be evaluated as

$t_{nt} = \frac{n_z - n_{ii}}{\tilde{q}_n} = \frac{z - z_{ii}}{\tilde{V}_n}$ or for short $t_{nt} = \frac{n_{zt}}{\tilde{q}_n} = \frac{z_{nt}}{\tilde{V}_n}$ with the (real) flow and speed averaged along the distance travelled in queue for the vehicle that joined the queue at $t$ and left the queue at $t_i$ in the position marked by $n_{ii}, z_{ii}$ ahead of the arrival position and experiencing the changes of traffic conditions along the vehicle path when meeting shock waves (as averaged during its time in queue). The non-queueing delay is then incurred in the remaining path to $A$ as estimated by

$\bar{t}_{ni} = \frac{\bar{n}_{ii}}{\bar{q}_s} = \frac{\tilde{z}_{ii}}{\tilde{V}_s}$ (where $\bar{n}_{ii}$ is the number of vehicles until section $A$ operating under free flow discharging conditions, along $\tilde{z}_{ii} = z_{ii}$), if $V_s^{T_n} < V_u^{T_n}$.

For Case II, avoiding the detailed calculation, the average queueing delay can be obtained by considering the area between the back and front of the restrained queue. In pure Case II it is evaluated as

$A_n = T \cdot \frac{n_{AM}}{2}$ (30g)

(or more generally as $A_n = T \cdot \frac{n_{AM}}{2} + n_{AT} \cdot t_q^2$, ignoring the time translation of periods but admitting the change of demand on $T_n$) that delivers the average time in queue for HQMs (instead of average queueing delay for VQMs) as

$\bar{t}_n = \frac{A_n}{N_{T+T_n}}$ (per vehicle) and the average vehicular queue as $\bar{n}_n = \frac{A_n}{T + T_n}$ at arrival, meaning that

$\bar{t}_n = \frac{\bar{n}_n}{\tilde{q}_n}$, as usual, with $\tilde{q}_n = \frac{N_{T+T_n}}{T + T_n}$ (for $T_n > t_i$). The corresponding area under free queue discharging flow conditions is $\tilde{A}_n = T \cdot \frac{\bar{n}_{AM}}{2}$ (not taken as queueing in the strict concept), similarly delivering
\[ \bar{t}_n = \frac{\hat{A}_n}{N_{T+T_n}}, \bar{n}_i = \frac{\hat{A}_n}{T+T_n}, \text{ and } \bar{t}_n = \frac{\bar{n}_i}{q_n}, \] as well (even if not strictly queueing). For Case II, as seen at section A, \( N_{T+T_n} = q_{A\cdot T}^T + q_{A\cdot T}^{Tn} \left( t_q + t_m \right) + Q_{A\cdot T}^{Tn} \left( T_n - t_r \right) = q_{A\cdot T}^T + q_{A\cdot T}^{Tn} \cdot t_q + Q_{A\cdot T}^{Tn} \left( T_n - t_r \right), \) but now section A is not the front of queue. As seen at the front of queue, \( \hat{N}_{T+t_q} = \hat{q}_{ins} \cdot T + \hat{q}_{ins} \cdot t_q = q_{A\cdot T}^T + \hat{q}_{ins} \cdot t_q = q_{A\cdot T}^T + \hat{q}_{ins} \cdot t_q \) is the number of vehicles that experience queueing, translating to \( \hat{N}_{T+t_q} = q_{A\cdot T}^T + \left( q_{A\cdot T}^{Tn} + K_{S\cdot T}^{Tn} \cdot w_m^{Tn} \right) \cdot t_q = q_{A\cdot T}^T + q_{A\cdot T}^{Tn} \cdot t_q + K_{S\cdot T}^{Tn} \cdot z_{AM} = q_{A\cdot T}^T + q_{A\cdot T}^{Tn} \cdot t_q + \bar{n}_{AM} \), where \( \bar{n}_{AM} = K_{S\cdot T}^{Tn} \cdot z_{AM} \) is the number of vehicles that fill the maximum queue extent \( z_{AM} \) with the density \( K_{S\cdot T}^{Tn} \) after queue dissipation at \( t_q \) (or \( N_{T+t_q} = q_{A\cdot T}^T + \left( q_{A\cdot T}^{Tn} + K_{S\cdot T}^{Tn} \cdot w_m^{Tn} \right) \cdot t_q = q_{A\cdot T}^T + q_{A\cdot T}^{Tn} \cdot t_q + K_{S\cdot T}^{Tn} \cdot z_{AM} = q_{A\cdot T}^T + q_{A\cdot T}^{Tn} \cdot t_q + \bar{n}_{AM} \) as seen by the queue side; the same value can be obtained from demand as seen by the arrival or queue side of the congestion wave as \( \hat{N}_{T+t_q} = \hat{Q}_{Ac\cdot T} + \hat{Q}_{Ac\cdot T}^{Tn} \cdot t_q = \hat{q}_{nc\cdot T} + \hat{q}_{nc\cdot T}^{Tn} \cdot t_q, \) using the related variables). Extending the observation by the normalization time and returning to section A with the normalization wave, then \( N_{T+t\cdot t} = \hat{N}_{T+t\cdot t} \) is the number of vehicles that experience queueing or free queue discharging, with \( \hat{N}_{T+t\cdot t} = q_{A\cdot T}^T + q_{A\cdot T}^{Tn} \cdot t_q + \hat{q}_{Sm\cdot T} \cdot t_m \) translating to \( \hat{N}_{T+t\cdot t} = q_{A\cdot T}^T + \left( q_{A\cdot T}^{Tn} + K_{S\cdot T}^{Tn} \cdot w_m^{Tn} \right) \cdot t_q + \left( q_{A\cdot T}^{Tn} - K_{S\cdot T}^{Tn} \cdot w_m^{Tn} \right) \cdot t_q = q_{A\cdot T}^T + q_{A\cdot T}^{Tn} \cdot t_q + q_{A\cdot T}^{Tn} \cdot t_q = N_{T+t\cdot t} \), as \( z_{AM} = w_m^{Tn} \cdot t_q = w_m^{Tn} \cdot t_m \) (also \( \hat{N}_{T+t\cdot t} = q_{A\cdot T}^T + \hat{q}_{ins} \cdot t_q + \hat{q}_{Am\cdot T} \cdot t_m \) and \( \hat{N}_{T+t\cdot t} = \hat{q}_{nc\cdot T} + \hat{q}_{nc\cdot T}^{Tn} \cdot t_q + \hat{q}_{Am\cdot T} \cdot t_m \), both deliver \( N_{T+t\cdot t} = \hat{N}_{T+t\cdot t} \); from demand, \( \hat{N}_{T+t\cdot t} = \hat{Q}_{Ac\cdot T}^T + \hat{Q}_{Ac\cdot T}^{Tn} \cdot t_q + \hat{Q}_{Am\cdot T} \cdot t_m \) delivers \( \hat{N}_{T+t\cdot t} = q_{A\cdot T}^T + q_{A\cdot T}^{Tn} \cdot t_q \) and \( N_{T+t\cdot t} = \hat{N}_{T+t\cdot t} \) as well, after including vehicles experiencing free queue discharging).

Alternatively, the area between the back and front of queue in a space-time frame can also handle the peculiarities of Case II. In pure Case II, the space-time queueing area is evaluated as

\[ A_2 \leq T \cdot \frac{Z_{AM}}{2} \]  

(35g)

(or more generally as \( A_2 \leq T \cdot \frac{Z_{AM}}{2} + Z_{AT} \cdot \frac{t_q}{2} \)) with \( \bar{a}_n = \frac{A_z}{N_{T+T_n}} \) (per vehicle) and the average vehicular queue as \( \bar{Z}_t = \frac{A_z}{T+T_n} \) at arrival, meaning that \( \bar{Z}_t = q_{n\cdot T} \cdot \bar{a}_n \), as usual, with \( \bar{q}_{n\cdot T} = \frac{N_{T+T_n}}{T+T_n} \) (for \( T_n > t_r \)). As before, a spatial average queueing time in sections of a link length \( L \) is \( \bar{t}_{dL} = \frac{A_z}{L} \leq \frac{T}{2} \cdot \frac{Z_{AM}}{L} \) for \( L \geq \frac{Z_{AM}}{2} \) as the
average queueing time in sections of \( L \) (or \( \bar{t}_{z} = \frac{A_z}{z_{AM}} = \frac{T}{2} \) in sections of \( z_{AM} \)) and a temporal average

**queueing time in a section at** \( z \) is \( \bar{t}_{z} = \frac{Z}{Z_{AM}}.T \) for \( z \leq z_{AM} \) (as used for evaluating the blocked time at \( z \)).

The corresponding area under free queue discharging flow conditions is \( \bar{A}_z = T.\frac{Z_{AM}}{2} \), similarly delivering

\[
\bar{a}_n = \frac{\bar{A}_z}{N_{T+T_n}}, \quad \bar{z}_t = \frac{A_z}{T+T_n} \text{ and } \bar{\bar{z}}_t = \bar{q}_n, \bar{\bar{a}}_n \text{ as well (even if not taken as queueing in the strict concept).}
\]

For Case II, the previous discussion can also be used to justify apparent averages. For queueing, the saturated time is now \( T + t_f \) (even if only \( T \) at section A), delivering \( \bar{n}_{ts} = \frac{A_z}{T + t_f}, \quad \bar{t}_{ms} = \frac{A_n}{N_{T+t_f}}, \quad \bar{t}_{zs} = \frac{A_z}{T + t_f} \)

and justifying \( \bar{t}_{ns} = \frac{N_{T+t_f}}{N_{T+T_n}}.\bar{t}_{ms} \) (as \( N_{T+t_f} = \hat{N}_{T+t_f} \)), \( \bar{t}_{ns} = \frac{\bar{n}_{ns}}{\bar{q}_{ns}} \) with \( \bar{q}_{ns} = \frac{N_{T+t_f}}{T + t_f} \), \( \bar{t}_{ns} = \frac{\bar{z}_{ns}}{\bar{V}_{ns}} \) with \( \bar{V}_{ns} = \frac{\bar{q}_{ns}}{\bar{K}_{ns}} \)

and \( \bar{K}_{ns} = \bar{K}_n \) with the space-time averages \( \bar{K}_n = \frac{A_n}{A_z} \), then \( \bar{V}_n = \frac{\bar{q}_n}{\bar{K}_n} \), as before. For free queue discharging, the clearing time is just \( t_f \) (as seen at section A), delivering \( \bar{n}_{ts} = \frac{\hat{A}_n}{t_f}, \quad \bar{t}_{ms} = \frac{\hat{A}_n}{N_{T+t_f}}, \quad \bar{t}_{zs} = \frac{\hat{A}_z}{t_f} \) with

and justifying \( \bar{t}_{ns} = \frac{N_{T+t_f}}{N_{T+T_n}}.\bar{t}_{ms} \) (as \( N_{T+t_f} = \hat{N}_{T+t_f} \) during dissipation and normalization), \( \bar{t}_{ns} = \frac{\bar{n}_{ns}}{\bar{q}_{ns}} \) with \( \bar{q}_{ns} = \frac{N_{T+t_f}}{t_f} \),

\[
\bar{t}_{ns} = \frac{\bar{V}_{ns}}{\bar{V}_{ns}} \text{ with } \bar{V}_{ns} = \frac{\bar{q}_{ns}}{\bar{K}_{ns}} \text{ and } \bar{K}_{ns} = \bar{K}_n \text{ with the space-time average } \bar{K}_n = \frac{\hat{A}_n}{A_z} \text{ and } \bar{V}_n = \frac{\bar{q}_n}{\bar{K}_n}, \text{ similarly (but}
\]

using specific averages, except \( \bar{a}_n \), for free queue discharging).

Even for the pure Case II, with uniform traffic conditions inside the queueing and the free queue discharging regimes, the analysis is complicated by the need of separating the effects of each regime and the treatment of the moving front of queue during dissipation. If ignoring the time translation of periods,

\[
A_n = A_0 = \frac{n_{AT}}{2}.T + n_{AT}.\frac{t_q}{2} \quad \text{and} \quad \tilde{A}_n = A_1 = \frac{n_{AM}}{2}.(t_q + t_m) = \frac{n_{AM}}{2}.t_f \text{ (similarly, } A_2 = A_2 = \frac{Z_{AT}}{2}.T + Z_{AT}.\frac{t_q}{2} \)
\]

and \( \tilde{A}_z = A_2 = \frac{Z_{AM}}{2}.(t_q + t_m) = \frac{Z_{AM}}{2}.t_f \), where \( A_n = A_0 = K_n.A_z \) (queueing) and \( \tilde{A}_n = A_1 = K^*_n.A_z \) (in free queue discharging), each corresponding to \( t_{ms}, \tilde{t}_{ms} \) exclusively (and related travel delays). The moving front of
queue means that queue faced at arrival may be different from the travelled queue for vehicles that move in queue during periods in which the front of queue moved itself. The travel time to A after joining the queue at \( t \) with the moving front of queue is now
\[
\tau_{vt} = \tau_{nt} + \tau_{nt}
\]
with
\[
\tau_{nt} = \frac{Z_s - Z_v}{V}, \quad \tau_{nt} = \frac{Z_v}{V_s}
\]
are the queuing and clearing components (similarly \( \tau_{nt} = \frac{n_t}{q_r}, \tau_{nt} = \frac{n_t}{q_s} \)) where \( z_{ts} \) is the distance dissipated by the front moving dissipation wave (that reduced \( n_{ts} \) vehicles from the front of queue and added \( n_{ts} \) vehicles in free queue discharging flow in the stretch of road in the front of queue to A).

Of course, in a general setting, mixed cases can be generated by joining features of Cases I and II (with queue recovery and dissipation) or in a multiperiod framework (several periods of queue formation and of queue recovery and/or dissipation). However, the essential new features are the separation of traffic conditions and the effect of a front moving queue but more generally in the same transient framework considered for Case I. The discussion ahead then focuses on the very similar analysis of sub-periods between the inflection points \( t_x, t_y \) and \( t_w \) for Case II, in its peculiar setting, where the initial queueing conditions corresponds to restrained queues (queueing in the strict concept) and the final “queueing” conditions corresponds to free queue discharging.
C. PROTOTYPICAL APPLICATION TO A LITERATURE CASE STUDY.

The illustrative examples are taken from those presented in Hurdle, Son (2001), originally all examples of Case I patterns, here supplemented by similar examples with Case II patterns. The common setting deals with a link (B,A) of 10km divided in 3 segments \((x_0; x_1), (x_1; x_2), (x_2; x_3)\) of similar length (3.3km) in a freeway with normal capacity of \(C = 4000v/h; 2000v/h/ln\) and free flow speed \(V_f = 100km/h\) (both lanes). Section B at \(x_0 = 0.0km\) is the mainline entry of the time-varying demand of the freeway \(Q_0[t]\), while intermediate sections at \(x_1 = 3.3km\) and \(x_2 = 6.7km\) are the candidate locations of an on-ramp contributing a constant flow of \(q_R = 400v/h\). Section A at \(x_3 = 10.0km\) is controlled by a potential bottleneck that imposes a time-varying maximum flow \(\mu_A[t]\). All analyses start with a period with no pre-existing queues.

In the first example, the time-varying entry demand pattern is \(Q_0[t] = \begin{cases} 3200v/h, t \in [0;15 \text{ min}] \\ 2000v/h, t \in [15 \text{ min}; \infty] \end{cases}\).

The on-ramp is located at \(x_1\) (with its constant contributing flow) and the bottleneck effect is a fixed constrained flow \(\mu_A[t] = 3200v/h, t \in [6 \text{ min}; \infty]\) at a given density \(K_\mu = 72v/km\) (of a potential restrained queue flowing at \(V_\mu \approx 44.4km/h\)). The main analysis in Hurdle, Son (2001) uses the FOTM combining the track of shock waves with the graphical plot of waves and period frontiers (with initial time translation of periods by the unconstrained travel time). It is clearly a Case I pattern as the first period generates the queue build-up at A and the next period allows queue dissipation due to entry demand reduction. Assuming \(V = V_f = 100km/h\) for unconstrained flow, the free flow travel time of 6min is the exogenous translation of periods from B to A (similarly, the translation to each intermediate section is 2min from the previous one) and the traffic densities in intermediate sections for the peak period are \(K_{01}^T = \frac{3200}{100} = 32v/km\), \(K_{12}^T = \frac{3600}{100} = 36v/km\) (after the contribution of the on-ramp), and \(K_{23}^T = \frac{3600}{100} = 36v/km\) (without queueing); for the next period, the corresponding values are \(K_{01}^{Tn} = \frac{2000}{100} = 20v/km\), \(K_{12}^{Tn} = \frac{2400}{100} = 24v/km\), and \(K_{23}^{Tn} = \frac{2400}{100} = 24v/km\) (without queueing). The peak demand reaches A at \(t=6\text{min}\), starting queue build-up with the congestion wave speed \(w_c^T = \frac{3200 - 3600}{72 - 36} = (-)11.11km/h\) (taken
Pietrantonio, H.

backward). The maximum queue is originally evaluated with the graphical plot as \( z_c = 2.5\text{km} \) and \( n_c = 180\text{v} \) (i.e. queues are fully stored in the last segment). This result is consistent with the additional (endogenous) translation of periods as seen at the back of queue, meaning that \( \delta_r = \frac{100}{11.11 + 100} \times 1.15 = 13.5\text{min} \) (referred to the start of queueing), \( z_c = z_{AW} = 11.11 \times \frac{13.5}{60} = 2.5\text{km} \), and \( n_c = n_{AW} = 72 \times 2.5 = 180\text{v} \), as previously discussed (without the endogenous time translation of period, \( z_c = z_{AT} = 11.11 \times \frac{15}{60} = 2.7775\text{km} \) and \( n_c = n_{AT} = 72 \times 2.7775 = 200\text{v} \)). Then the reduction of demand starts the recovery time with the recovery wave speed \( w_{c}^{T} = \frac{3200 - 2400}{72 - 24} = 16.67\text{km/h} \) (forward), eliminating the queue after \( t_{r} = \frac{2.5}{16.67} \times 60 = 9\text{min} \). The analysis with VQM and HQM will both include the exogenous time translation of periods as seen at the bottleneck A (this is made explicit ahead). The VQM predicts queue build-up to \( \tilde{n}_{AT} = (3600 - 3200) \times \frac{15}{60} = 100\text{v} \) (~1.4km if assuming the given density of 72v/km in the queue) and the recovery time \( \tilde{t}_{s} = \frac{100}{3200 - 2400} \times 60 = 7.5\text{min} \). The HQM predicts queue build-up to \( n_{AT} = \frac{3600 - 3200}{1 - 36/72} \times \frac{15}{60} = 200\text{v} \) (losing the endogenous time translation of periods) and recovery time \( t_{r} = \frac{200}{(3200 - 2400)/(1 - 24/72)} \times 60 = 10\text{min} \). Except for losing the endogenous time translation of periods, the results from the HQM are clearly better.

The similar Case II for this detailed example can be set as \( Q_{S}[t] = 3200\text{v/h}, t \in [0;\infty] \) and \( \mu_{A}[t] = \begin{cases} 3200\text{v/h}, t \in [6;21\text{min}] \\ 4000\text{v/h}, t \in [21\text{min};\infty] \end{cases} \) keeping the given density \( K_{\mu}^{T} = 72\text{v/km} \) (of a potential restrained queue flowing at \( V_{\mu}^{T} = 44.4\text{km/h} \) in the peak period and assuming a free flowing queue of \( V_{\mu}^{T} = 64\text{km/h} \) (at a density of \( K_{\mu}^{T} = 62.5\text{v/km} \)). The shock wave analysis delivers the congestion wave speed \( w_{c}^{T} = \frac{3200 - 3600}{72 - 36} = (-)11.11\text{km/h} , \) taken backward, \( \delta_r = \frac{100}{11.11 + 100} \times 1.15 = 13.5\text{min} , \)
\[ z_{AW} = 11.11 \times \frac{13.5}{60} = 2.5 \text{km}, \text{ and } n_{AW} = 72 \times 2.5 = 180v \text{, as before, for the peak period. In the next period, } \\
\text{the dissipation wave speed is } w_{sn}^{Tn} = \frac{4000 - 3200}{62.5 - 72} = (-84.21) \text{ km/h (taken backward) and the dissipation } \\
time is \ t_q = \frac{2.5 + 11.11 \times 1.5/60}{84.21 - 11.11} \times 0.60 = 2.3 \text{ min (at the back of queue, with the effect of the endogenous} \\
\text{translation of period), meaning } z_{AM} = 84.21 \times \frac{2.3}{60} = 3.2 \text{ km with } \tilde{n}_{AM} = 3.2 \times 62.5 = 200v. \text{ After the} \\
dissipation, the normalization wave starts with speed } w_{sn}^{Tn} = \frac{4000 - 3600}{62.5 - 36} \text{ km/h (forward) and the} \\
\text{normalization time is } t_p = \frac{3.2}{15.09} \times 0.60 = 12.7 \text{ min (meaning 6+15+2.3+12.7=36min from the start of the} \\
analysis period). \\
The VQM predicts the same queue build-up to } \tilde{n}_{AT} = (3600 - 3200) \frac{15}{60} = 100v \text{ (1.4km if assuming} \\
\text{the given density of 72v/km in the queue) and the dissipation time } \tilde{t}_q = \frac{100}{4000 - 3600} \times 0.60 = 15 \text{ min (meaning} \\
6+15+15=36\text{min as with the FOTM for dissipation and clearing). Similarly, the HQM predicts the same queue} \\
\text{build-up to } n_{AT} = \frac{3600 - 3200}{1 - 36/72} \frac{15}{60} = 200v \text{ but the dissipation time and the normalization time are} \\
\text{t} \text{.} \text{60} = 2.3 \text{ min (as above), for} \\
\text{the additive form, } n_{AM} = 200 + \frac{3600 - 3200}{1 - 36/72} \frac{2.3}{60} = 204v, \text{ and } t_p = \frac{204}{(4000 - 3600)/(1 - 36/265)} \times 0.60 = 12.9 \text{ min. The same} \\
\text{results can be obtained using the (fictitious) reference queue (or correction factors; see Tables 1 and 2). Using} \\
\text{the additive form, } n_{AM} = 200 + \frac{3600 - 3200}{1 - 36/72} \frac{2.3}{60} = 231v \text{ corresponds to the maximum queue extent,} \\
t_q = \frac{231}{(4000 - 3200)/(1 - 62.5/72)} \times 0.60 = 2.3 \text{ min, } \text{ t} \text{.} \text{60} = 2.8 \text{ min to} \\
\text{the dissipation and normalization times, and } \tilde{n}_{AM} = \frac{(4000 - 3600)}{1 - 36/265} \frac{12.8}{60} = 202v \text{ to the effective number of} \\
\text{vehicles between M and A (as above). Results are similar to those of the FOTM. The precision is better}
because the effect comes mainly from the front of queue, not affected by the endogenous time translation of periods.

The use of auxiliary tables for converting VQMs into HQMs estimates can also be easily illustrated. For the peak period in Case I or II, taking $Q_A^T \sim 1400v/h/ln$ and $V^T_u \sim 90km/h$ for demand conditions and $\ell_v = \bar{e}_n \sim 25m$ as representative of queueing flow conditions generated from the bottleneck, the conversion factor for the estimated queue is seen as 1.6369 in Table A1a, meaning that $n_{AT} = 100v$ is translated into $n_{AT} = 164v$ (much akin to the value $n_{AT} = 180v$ obtained with the FOTM). Similarly, for the next period in Case II, taking $q_{A}^{Tn} \sim 1500v/h/ln$ and $V_{A}^{Tn} \sim 60km/h$ for saturation flow conditions and again $\ell_v = \bar{e}_n \sim 25m$ for queueing flow conditions, the conversion factor for the estimated dissipation time can be seen as 0.2160 in Table A3b, meaning that $t_q = 3.2min$ (much akin to the value $t_q = 2.3min$ obtained with the FOTM). Of course, auxiliary tables can be detailed for getting better estimates.

An illustration of the use of flow and queue data to measurement of demand can also be made using these examples. In the peak period, field observation at the bottleneck will measure the flow of 3200v/h and the growth of queue to 180v or 2.5km in 15min. Given the HQM relations, the field data can be translated into an estimate of demand of $Q_0 = \frac{3200x13.5/60+180}{13.5/60+2.5/100} = 3600v/h$, the assumed value (or if neglecting the endogenous time translation of period).

Of course, with the assumption of a free flowing queue of $V^T_s = 64km/h$ (at a density of $K^T_s = 62.5v/km$), the original values of density $K^T_\mu = 72v/km$ for the given bottleneck flow $\mu_A[t] = 3200v/h, t \in [6min; \infty]$ (of a potential restrained queue flowing at $V^T_{\mu} = 44,4km/h$) of Hurdle, Son (2001) breaks down the linear density relation usually suggested for the forced flow regime (a mere convenience). A sensitivity test can be made by enforcing the linearity assumption, evaluating $K^T_\mu = 200 - (200 - 62.5) \cdot \frac{3200}{4000} = 90v/km$ (with $V^T_{\mu} = \frac{3200}{90} = 35.6km/h$). Then $f_{qn}^T = 1 - \frac{36}{90} = 0.40$ and $n_{AT} = \frac{100}{0.40} = 250v$ (instead of 200v) for $z_{AT} = \frac{250}{90} = 2.78km$ (instead of 2.8km), or
\[ \delta_T = \frac{1 - 36/90}{1 - 35.6/100} \cdot 15 = 14 \text{ min} \] (instead of 13.5 min), and \[ n_{AW} = \frac{(3600 - 3200)}{0.40} \cdot \frac{14}{60} = 233.3v \] (instead of 180v) for \[ z_{AW} = \frac{233.3}{90} = 2.59 \text{ km} \] (instead of 2.5 km), in the first example, while

\[ n_{AT} = \frac{3600 - 3200}{1 - 36/90} \cdot \frac{15}{60} = 167v \] for 1.86 km (instead of 200v for 2.8 km),

\[ t_q = \frac{167}{(4000 - 3200)/(1 - 62.5/90) - (3600 - 3200)/(1 - 36/90)} \cdot 60 = 5.1 \text{ min} \] (instead of 2.3 min),

\[ n_{AM} = 167 + \frac{3600 - 3200}{1 - 36/90} \cdot \frac{5.1}{60} = 227v \] for 2.5 km (instead of 231v for 3.2 km),

\[ t_m = \frac{227}{(4000 - 3600)/(62.5/90 - 36/90)} \cdot 60 = 10.0 \text{ min} \] (instead of 12.8 min) and

\[ \hat{n}_{AM} = \frac{(4000 - 3600)}{1 - 36/62.5} \cdot \frac{10.1}{60} = 159v \] (instead of 202v), in the second example. Results are clearly sensitive to assumptions, even if keeping the magnitude and being an improvement over results with VQM assumptions.

Note that these results imply that there is no interaction between road elements even if the on-ramp is located in the nearer alternative position (but for a very small margin in the Case II pattern). Nevertheless, interactions will be present in the next example and requires the application of an event-based procedure for tracking waves (e.g. as generally proposed by Adamo et alii, 1999, for the general dynamic network loading problem). In this case, any new event is related to the generation of a new shockwave, either by the change of traffic conditions that identifies time periods (in the here proposed setting) or due to the meeting of previous shockwaves (i.e. of next meeting of adjacent shockwaves as obtained by the initial distance and relative speeds of their wavefronts). This event-based procedure requires information on the position of all the shockwaves present in each road segment at event times, with the traffic conditions adjacent to each wavefront (perhaps including some traffic dispersion or attenuation model can link traffic conditions of adjacent front conditions). A general procedure must also be complemented by node models that can represent the competition of traffic flows provenient from all the links reaching a conflicting area (i.e. an intersection, merging or diverging area), given the conditions commanded by road design and traffic control, as well as the consideration of upper decision levels that can define route/departure choices (among other demand features that may be affected by traffic conditions), both features ruled-out here by the assumption of an exogenous inflow at the on-ramp and by the simplified topology of a linear corridor with given demand conditions (yielding a simple procedure that is applied in the simulation of the following examples).
Pietrantonio, H.

a. Detailed Example

general data: 2 lanes, $C_H=4000v/h$, $V_f=100$km/h
on-ramp contributing flow: $400v/h$ (location $x_1$)

Case I: $3200v/h$, $2000v/h$
Case II: $3200v/h$, $3200v/h$

b. Further Examples

general data: 2 lanes, $C_H=4000v/h$, $V_f=100$km/h
on-ramp contributing flow: $400v/h$ (candidate locations $x_1$ or $x_2$)
Figure 2. Illustrative Examples from Hurdle, Son (2001) with Proposed Case II Versions

The last two examples presented in Hurdle, Son (2001) can be jointly analyzed, by differing only on the location of the on-ramp (at x₁ or x₂), for a new time-varying pattern of the entry demand as

\[ Q_0(t) = \begin{cases} 
3800v/h, & t \in [0;15\text{min}] \\
2600v/h, & t \in [15;30\text{min}] \\
2000v/h, & t \in [30\text{min};\infty] 
\end{cases} \]

(as before, in a freeway with 2 lanes, normal capacity of \( C_m = 4000v/h \), and free flow speed \( V_f = 100\text{km/h} \), the on ramp has a fixed contributing flow as \( q_R = 400v/h \), and the bottleneck effect is a fixed imposed flow given as \( \mu(t) = 3200v/h, t \in [6\text{min};\infty] \), for a given density \( K_\mu = 72v/km \) in a potential restrained queue flowing at \( V_\mu = 44.4\text{km/h} \). Now two queues are generated but, again, both clearly experience Case I patterns. Further assumptions on the pattern of queue discharging are also introduced (it is easy to check that the assumptions are generated by a simplified triangular fundamental diagram with capacity and saturation flow of 4000v/h at 40v/km, with normal speed of 100km/h, and the jam density of 200v/km, all considering the 2 lanes of the freeway, meaning a backward wave speed inside queue of

\[ w = \frac{S}{K_J - K_s} = \frac{4000}{200 - 40} = 25\text{km/h} \]

that will occur several times ahead).

Similarly, Case II patterns can be generated by changing the bottleneck condition to keep the given density \( K_\mu^T = 72v/km \) (of a potential restrained queue flowing at \( V_\mu^T = 44.4\text{km/h} \) in the peak period and assuming free discharging queue at \( V_\mu = 64\text{km/h} \) (and density of \( K_\mu = 62.5v/km \), distinctively in the next period. The entry demand is also changed to

\[ Q_0(t) = \begin{cases} 
3800v/h, & t \in [0;15\text{min}] \\
2600v/h, & t \in [15\text{min};\infty] 
\end{cases} \]

(generating a Case II pattern for the main queue).

In the following examples, Case I is examined only for on-ramp location at x₁ (the simplest scenario) but Case II is examined for on-ramp location at x₂ (turning it into a more complex scenario). Again, all analyses in Hurdle, Son (2001) start with a period with no pre-existing queues, using FOTM and combining the track of shock waves and the graphical plot of wave and period frontiers, with initial time translation of periods by the unconstrained travel time, as maintained ahead.

For the Case I, in the initial period the upstream bottleneck is formed at the on-ramp (at x₁ or x₂), because the on-ramp contribution reduces the available capacity for flowing the entry demand to
\[ \mu_R = C_m - q_R = 4000 - 400 = 3600 \text{v/h} \] (the normal capacity ahead less the assumed on-ramp flow), with an assumed density of \( K_{\mu R} = 56 \text{v/km} \), in a period with \( Q_0 = 3800 \text{v/h} \) and \( K_{01}^T = \frac{3800}{100} = 38 \text{v/km} \) (admitting that the outflow remains at the normal capacity of 4000v/h at the merge with the on-ramp flow; eventually a merging model can be used to evaluate the split of discharging capacity to the on-ramp and entry flows, with their operating conditions, and the effect of capacity drop at merging, with the operating condition of its free queue discharging flow). With the on-ramp at \( x_1 \) the first queue starts at \( t=2 \text{min} \) (for the on ramp at \( x_2 \) the first queue starts at \( t=4 \text{min} \), considering the exogenous component of the time translation of periods.

The first congestion speed is \( w_{c1}^T = \frac{3800 - 3600}{38 - 56} = (-)11.11 \text{km/h} \) (taken backward). Evaluating the additional (endogenous) translation of periods as seen at the back of queue, \( \delta_T = \frac{100}{11.11 + 100} \cdot 15 = 13.5 \text{min} \) referred to the start of queueing at \( t=2 \text{min} \) in \( x_1 \), then \( z_c = z_{AW} = 11.11 \times \frac{13.5}{60} = 2.5 \text{km} \), and

\[ n_c = n_{AW} = 56 \times 2.5 = 140 \text{v}. \] After the peak period, with \( Q_0 = 2600 \text{v/h} \) and \( K_{01}^T = \frac{2600}{100} = 26 \text{v/km} \), the recovery speed is \( w_{r1}^{Tn} = \frac{2600 - 3600}{26 - 56} = 33.33 \text{km/h} \) (forward) in the next period, that returns the road to the no queue condition after \( t_{r1} = \frac{2.5}{33.33} = 4.5 \text{min} \) at \( t=20 \text{min} \) in \( x_1 \), potentially (i.e. if no change occurs in the meantime; the same analysis would apply if the on-ramp is located at \( x_2 \) but the queueing starts at \( t=4 \text{min} \) and finishes at \( t=22 \text{min} \), potentially). Note the gating effect: after the on-ramp the outflow of 4000v/h at \( x_1 \) generates \( Q_3 = 4000 \text{v/h} \) for \( t \in [6;24 \text{min}] \) at the main bottleneck for an enlarged period of 18min, from the outflow in \( t \in [2;20 \text{min}] \) (would be \( t \in [4;22 \text{min}] \) for the on-ramp at \( x_2 \)), being enough to also trigger the second queue at the exit section where \( \mu = 3200 \text{v/h} \) (downstream of \( x_4 \)).

The analysis assumes that traffic quickly accelerates to the same unconstrained speed of 100km/h after merging, meaning that the downstream queue starts at \( t=6 \text{min} \) for the exogenous component of the time translation of periods, with \( K_{12}^T = K_{23}^T = \frac{4000}{100} = 40 \text{v/km} \), and the main congestion speed is

\[ w_{c2}^T = \frac{3200 - 4000}{72 - 40} = (-)25 \text{km/h} \] (taken backward). Evaluating the additional (endogenous) translation of
periods as seen at the back of queue, $\delta_T = \frac{100}{25+100} \cdot 18 = 14.4$ min referred to the start of queueing at $t=6$min in $x_3$, $z_d = z_{AW} = 25 \times \frac{14.4}{60} = 6$km, and $n_c = n_{AW} = 72 \times 6 = 432v$, potentially (if no change occurs in the meantime of $t \in [6;20.4$min$]$). As a conclusion, the maximum queue extent does not reach the on-ramp if located at $x_1$ even if filling the segment $(x_2; x_3)$ and reaching the segment $(x_1; x_2)$, after $\frac{3.3}{25} \cdot 100 = 8$ min (it would reach the on-ramp if located at $x_2$ in $t=14$min, 8min after the start of queueing on $t=6$min, but this event is analyzed ahead in the more complex situation for Case II). Then, with the on-ramp at $x_1$ the queues are independent, except for the gating effect of the upstream queue at the on-ramp merge that reduced the downstream demand to $4000v/h$. The maximum queue extent in the main queue occurs at $t=20.4$min and then demand falls to $Q_3 = 2600 + 400 = 3000v/h$ for $t \in [20;32$min$]$ at $x_1$ (a shortened period) and generates a shock wave with slow recovery speed $w_{r2}^T = \frac{3000 - 3200}{30 - 72} = 4.76$km/h (forward). Evaluating the additional (endogenous) translation of periods as seen at the back of queue, $\delta_T = \frac{100}{100 - 4.76} \cdot 1.2 = 12.6$ min, the queue reduction reaches $z_d = 6 - 4.76 \times \frac{12.6}{60} = 5$km. Then, a further decrease in demand to $Q_3 = 2000 + 400 = 2400v/h$ for $t \in [32; \infty$min$]$ at $x_1$, increases the recovery speed to $w_{r2}^T = \frac{2400 - 3200}{24 - 72} = -16.67$km/h, clearing the queue $\frac{5}{16.67} \cdot 60 = 18$ min after, at $t=51$min ($6+14.4+12.6+18=51$min) or 45min after the start of queueing at $x_3$. The interaction is more involved with the on-ramp at $x_2$ (see 4) but the discussion of Case II ahead deals with an even more complex situation.

For Case II with the on-ramp at $x_2$, the potential course of independent queue formation (from 0 to 2.5km for the upstream queue with $w_{c1}^T = \frac{3800 - 3600}{38 - 56} = (-)11.1$ km/h, taken backward, now on $t \in [4;22$min$]$ at $x_2$, and from 0 to 6km for the main queue with $w_{c2}^T = \frac{3200 - 4000}{72 - 40} = (-)25$ km/h, taken backward, from $t \in [6;20.4$min$]$) is disturbed because the downstream queue that starts at $t=6$min reaches the position of the on-ramp at $x_2$ (3.3km before $x_3$) after $\frac{3.3}{25} \cdot 100 = 8$ min of the main congestion (at
t=6+8=14min), joining the downstream and the upstream queues. Then, the flow reduction to 3200v/h of the main queue is imposed to the merging section and further reduces the flow to 2800v/h for the main road before the on-ramp (keeping the on-ramp contribution as 400v/h) from t=14min, flowing at an assumed density $K_{\mu R} = 88v/km$, from the head of queue before merging. The transmission of this internal change occurs inside the queue with an internal speed $w_{ic} = \frac{3600 - 2800}{56 - 88} = -25km/h$ starting at t=14min, when the initial congestion wave that starts at t=4min reaches km $160 \times 11.11 \times \frac{10}{60} = 1.85km$ from $x_2$. The same pattern of upstream congestion with $w_{ci} = -11.11km/h$ occurs for the remaining period with $Q_0 = 3800v/h$ (to be evaluated) while the change propagates till the back of queue, meaning that the internal change potentially lasts for the assumed triangular fundamental diagram). But the change of time period occurs before the effect of the internal wave meets the entry demand flow (otherwise the congestion speed at $x_2$ would increase after the end of internal change to $w_{ci} = \frac{3800 - 2800}{38 - 88} = (-)20km/h$, faster backward). As the maximum extent of the upstream queue occurs before, at t=17.5min (13.5min after the start of queueing in $x_2$ or 3.5min after the start of the internal wave), then $z_c = z_{AW} = 2.5km$ while $z_{ic} = 25x \frac{3.5}{60} = 1.46km$ for the internal wave at the time of the maximum (back of) queue extent. Then the upstream recovery wave starts with $w_{ri} = \frac{2600 - 3600}{26 - 56} = 33.33km/h$ (forward). The internal wave reaches the back of queue $t_{ic} = \frac{2.5 - 1.46}{25 + 33.33} \times 60 = 1.1min$ after the start of the recovery wave and changes its speed to $w_{ri} = \frac{2600 - 2800}{26 - 88} = 3.22km/h$ (forward but much slower) at t=18.6min and $z_{di} = 2.5 - 33.33x \frac{1.1}{60} = 1.9km$, thus requiring $t_{ri} = \frac{1.9}{3.22} \times 60 = 35.2min$ to potentially clear the upstream queue at $x_2$ (if no change occurs). In the given example in 4, the next change comes from a further decrease of demand at the back of queue that can recover the normal traffic condition for the overall road. In this corresponding Case II example, the next change is the increase in the bottleneck outflow to 4000v/h at the head.
of queue, at t=21min (15min after the initial period in $x_3$), while the slow recovery clear the joined queue and reaches $x_2$ (then potentially changing the recovery speed to $w_{s2}^{Tn} = \frac{3000 - 3200}{30 - 72} = 4.76\text{km/h}$ to $x_3$).

As before, in the next period, the dissipation wave speed is $w_{sl}^{Tn} = \frac{4000 - 3200}{62.5 - 72} = (-)84.21\text{km/h}$ (backward), starting at t=21min and can reach $x_2$ after $t_{sl}^{Tn} = \frac{3.3}{84.21} \cdot 60 = 2.4\text{min}$, well before the slow recovery wave. At the congested merging, the on-ramp is flowing 400v/h and the main road has the queueing flow of 2800v/h that can increase to 3600v/h after the arrival of the dissipation wave from $x_3$. Then, the dissipation wave speed changes to $w_{s2}^{Tn} = \frac{3600 - 2800}{56 - 88} = (-)25\text{km/h}$ (backward) at t=23.4min, when $z_{ei} = 1.9 - 3.22 \times \frac{4.8}{60} = 1.6\text{km}$, thus requiring further $t_{s2}^{Tn} = \frac{1.6}{25} \cdot 60 = 3.9\text{min}$ to clear at t=27.3min. The normalization starts then with wave speed and time are $w_{m1}^{Tn} = \frac{3600 - 2600}{56 - 26} = 33.33\text{km/h}$ (forward) with $t_{m1}^{Tn} = \frac{1.6}{33.33} \cdot 60 = 3\text{min}$ in the way to $x_2$, then changes the speed to $w_{m2}^{Tn} = \frac{4000 - 2600}{62.5 - 26} = 38.36\text{km/h}$ with $t_{m1}^{Tn} = \frac{3.33}{38.36} \cdot 60 = 5.2\text{min}$ in the way to $x_3$, potentially returning the road to normal traffic conditions flowing 2600v/h at t=35.5min. As the next change of time periods would occur at t=36min as seen at $x_3$, it occurs after normalization (otherwise, the normalization speed would experience another change).

It is well known and easy to understand that FOTM results are very sensitive to values assumed for some parameters, mainly the outflow at bottlenecks (104). However, the previous discussion shows that queue discharging densities also play a major role in predicting shock wave speeds. The parameters in 4 are based on a triangular fundamental flow diagram (given its theoretical interest in applying this modeling assumption). A previous departure from this model form was the assumed free queue discharging flow density (and speed) for the Case II patterns (the assumed saturation flow $S = 4000\text{v/h}$ at $V_{\mu}^T = 64\text{km/h}$ and $K_{\mu}^T = 62.5\text{v/km}$, meaning at least a trapezoidal form). A sensitivity test can be made by returning to the default model (then $V_{\mu}^T = 100\text{km/h}$ and $K_{\mu}^T = 40\text{v/km}$) or changing to an alternative assumption (e.g. $V_{\mu}^T = 80\text{km/h}$ and $K_{\mu}^T = 50\text{v/km}$). It is easy to check that, even keeping the same outflows and neglecting the capacity drop, the results with the FOTM vary widely. Thus the incipient knowledge about queue discharge flow models is a
major theoretical constraint and the linear hypothesis for queueing densities (i.e. \( K_n = K_J - (K_J - K_s) \frac{q_n}{S} \)) seems to be the simplest option. Then, for queue discharge flows \( q_n = 3200; 3600; 2800 \) v/h, keeping \( K_J = 200 \) v/km, the assumed densities of \( K_n = 72; 56; 88 \) v/km change to \( K_n = 90; 76.25; 103.75 \) v/km under the assumed condition or \( K_n = 80; 65; 95 \) v/km under the above option (the previous examples can be revised with these new values).

The analysis with the HQM is carried-out with the simplest method of Table 1 for correcting (the totally misleading results of) the VQM estimates (i.e. neglecting the endogenous component of the time translation of periods, as demonstrated in the detailed example). The solution proceeds in a similar way to the tracking of waves. First, the queues are examined as being independent and, if not, the interaction has to be considered also in a similar way. The elaboration of the Flow Diagram shown in Figure 2 (mimicking the illustrations in 4, but redrawn with the convention of Figure 1 and the new case data when distinct) is a friendly resource for the calculation process, as used in the following.

For Case I, the upstream queue generated by the imbalance at the on-ramp merging during the first 15 min (peak) period would generate a vertical queue estimate of \( \tilde{n}_{uT} = (3800 - 3600) \frac{15}{60} = 50 \) v and, using a free flow speed \( V_f = 90 \) km/h and typical moving queue spacing of \( \ell_v = 25 \) m (if not knowing that \( V_f = 100 \) km/h and \( \bar{c}_n = 35.7 \) m for \( K_n = 56 \) v/km), a horizontal queue estimate of

\[
\begin{align*}
n_{uT} &= \frac{50}{1 - 3800/2 \times 0.025/90} = 106 \text{v} \\
&= \frac{50}{1 - 3800/2 \times 0.036/100} = 158 \text{v},
\end{align*}
\]

against the value of 140v obtained with the FOTM) and then

\[
\begin{align*}
z_{uT} &= \frac{106}{2 \times 0.025} = 1.3 \text{km} \quad \text{(with the known parameters)} \\
&= \frac{158}{2 \times 0.036} = 2.8 \text{km} \quad \text{against the value of 2.5 km obtained with the FOTM).}
\end{align*}
\]

It is clear that results are very sensitive to the assumed parameters (i.e. to the queue discharge assumptions). The vertical queue dissipation time estimate is

\[
\begin{align*}
\tilde{t}_{us} &= \frac{50}{3600 - 2600} \times 60 = 3 \text{ min},
\end{align*}
\]

that can be accepted for a Case I pattern but the better estimate is

\[
\begin{align*}
t_{uT} &= \frac{1 - 2600/2 \times 0.036/100}{1 - 3800/2 \times 0.036/100} \times 3 = 5 \text{ min} \quad \text{(against the value of 4.5 min obtained with the FOTM).}
\end{align*}
\]

Ignoring the endogenous component of the time translation of periods, the enlarged period with gating...
Pietrantonio, H.

at 4000v/h is 20min (instead of 18min obtained with the FOTM, a difference which can be attributed to the endogenous component of the time translation of periods considered in the FOTM solution).

The downstream queue generated by the imbalance at the main bottleneck during the first 15min (peak) period would generate a vertical queue estimate of \( \tilde{n}_{dT} = \frac{4000 - 3200}{60} \times 20 = 267 \) by considering the enlarged period generated by the gating effect of the upstream queueing and, using free flow speed \( V_f = 90\text{km/h} \) and a typical moving queue spacing of \( \ell_v = 25\text{m} \) (if not knowing that \( V_f = 100\text{km/h} \) and \( \bar{e}_n = 27.8m \) for \( K_n = 72\text{v/km} \)), a horizontal queue estimate of \( n_{dT} = \frac{267}{1 - 4000/2 \times 0.025/90} = 606v \) (with the known parameters \( n_{dT} = \frac{267}{1 - 4000/2 \times 0.025/100} = 606v \), against the value of 432v obtained with the FOTM) and then \( z_{dT} = \frac{600}{2} \times 0.025 = 7.5\text{km} \) (with the known parameters \( z_{dT} = \frac{606}{2} \times 0.028 = 8.5\text{km} \) against the value of 6km obtained with the FOTM). Then, the results do not support the conclusion of independent queues by a medium margin, mainly because of the effects of ignoring the endogenous component of the time translation of periods (otherwise supporting the prediction of a small interaction effect, if present).

For Case II, of course the conclusion with a HQM is that the interaction effect is really strong (contrary to the prediction obtained with a conventional VQM and akin to the conclusion obtained with the FOTM). Then, the prediction of the interaction effects has to be checked.

The first task is predicting the time at which the downstream queue reaches \( x_2 \) and join the upstream queue. Again using free flow speed \( V_f = 90\text{km/h} \) and a typical moving queue spacing of \( \ell_v = 25\text{m} \) (if not knowing that \( V_f = 100\text{km/h} \) and \( \bar{e}_n = 27.8m \) for \( K_n = 72\text{v/km} \)), \( z_{dT} = 3.3\text{km} \) corresponds to \( n_{dT} = 2 \times 3.3 \times 0.025 = 264v \) that occurs when \( \frac{(4000 - 3200)t_d}{1 - 4000/2 \times 0.025/90} = 264v \) or \( t_d = 8.8\text{min} \) (with the known parameters, \( n_{dT} = 2 \times 3.3 \times 0.028 = 236v \) and \( \frac{(4000 - 3200)t_d}{1 - 4000/2 \times 0.028/100} = 236v \) or \( t_d = 7.8\text{min} \), a difference due to rounding against the value of 8min obtained with the FOTM).

The second task is predicting the next event till the end of the first period with the new traffic condition in the joined queue, at 6.2min (or 7.2min with better parameters) ahead in time (instead of 3.5min obtained with the FOTM, as the endogenous component of the time translation of periods is ignored). As seen from the downstream queue, this new period has queue formation with arriving flow from the upstream queue. Under the
simplified assumptions, vehicle spacings and densities were taken as equal inside moving queues (in all conditions, $\ell_v = 25 m$, meaning that only the discharging speed changes), what is inconsistent with a finite value for speed with the usual shock wave formula. As an option, with the linear hypothesis for queueing densities all shock waves inside queues have the same congestion wave speed ($w = 25 km/h$ in the example).

In a consistent theoretical framework, these hypotheses (constant queueing vehicle spacing or constant queueing wave speeds) must be viewed as alternatives (being inconsistent) and the linear hypothesis for queueing densities seems to be then preferable. In a practical context, however, they can be taken as approximations to be used when parameters are not available for applying more precise models. At least in this example, it seems clear that the error in any of these assumptions is less important than the error from neglecting the endogenous component of the time translation of periods.

Accepting the simple view, the next 6.2 min potentially reaches $z_{ui} = w_{ui} \cdot t = 25 \times \frac{6.2}{60} = 2.58 km$ with

\[ n_{ui} = N_r, \quad z_{ui} = 2 \times \frac{2.58}{0.025} = 207 v, \] meaning that the prediction is the full replacement of the initial queue by the new one before the end of the peak period (considering the upstream queue estimated as $106 v$ and $1.3 km$, in the peak period). At $t=8.8$ min, the initial queue is predicted as

\[ n_{ui} = \frac{3800 - 3600}{1 - 3800/2 \times 0.025/90} \times 8.8 \times 60 = 62 v \text{ and} \]

the replacement starts, then finishing when $62 + \frac{(3800 - 3600) t_e}{1 - 3800/2 \times 0.025/90} = 2 \times \frac{25 t_e}{0.025}$, delivering $t_e = 2.4$ min

with $n_e = 2 \times \frac{25 \times 2.4/60}{0.025} = 79 v$ at $z_e = \frac{79}{2} \times 0.025 = 0.99 km$ (3.8 min before the end of the peak period).

Thus, up to the change of the period, queue grows to

\[ n_r = 79 + \frac{3800 - 2800}{1 - 3800/2 \times 0.025/90} \times \frac{3.8}{60} = 134 v \text{ and} \]

\[ z_r = \frac{134}{2} \times 0.025 = 1.68 km \text{ at the end of the peak period.} \]

With the linear hypothesis for queueing densities, the parameters of the upstream and downstream queues are known ($K_n = 56 v/km$, meaning $\bar{e}_n = 36 m$, upstream for $q_n = 3600 v/h$ and $V_n = 64 km/h$; $K_n = 88 v/km$, meaning $\bar{e}_n = 23 m$, downstream for $q_n = 2800 v/h$ and $V_n = 32 km/h$; both consistent with the speed of $w_{ui} = 25 km/h$ for the internal wave, as assumed above). Then, the next 7.2 min potentially reaches $z_{ui} = w_{ui} \cdot t = 25 \times \frac{7.2}{60} = 3.0 km$ with $n_{ui} = N_r, z_{ui} = 2 \times \frac{3.0}{0.023} = 261 v$ vehicles added and
vehicles reduced from the queue, meaning again that the prediction is the full replacement of the initial queue by the new one before the end of the peak period (considering the upstream queue estimated as 158v and 2.8km, in the peak period), even if for a smaller margin. Now, at t=7.8min, the initial queue is predicted as
\[ n_{\text{f}} = 116 + \frac{3800 - 2800}{1 - 3800/2 \times 0.023/100} \times 4 = 234v \] and the replacement starts, then finishing when
\[ 82 + \frac{(3800 - 3600) t_e}{1 - 3800/2 \times 0.036/100} = 2.25 \times 0.023 \, t_e \], delivering \( t_e = 3.2 \, \text{min} \) with
\[ n_e = 2.25 \times 3.2/60 \times \frac{1}{0.023} = 116v \] at \( z_e = \frac{116}{2} \times 0.023 = 1.33 \, \text{km} \) (4min before the end of the peak period). Thus, up to the change of the period, queue grows to
\[ n_{\text{f}} = 116 + \frac{3800 - 2800}{1 - 3800/2 \times 0.023/100} \times 4 = 234v \] and
\[ z_{\text{f}} = \frac{234}{2} \times 0.023 = 2.69 \, \text{km} \] at the end of the peak period (a much larger estimate).

Estimates were, thus, qualitatively distinct (as explained by the large effect expected from neglecting the endogenous component of the time translation of periods) and sensitive to assumptions about queue discharging densities. However, most estimates (above referred to \( x_2 \)) keep the correct order of magnitude (much better than those obtained with VQMs) even in these awkward conditions.

The last task investigated in the examples deals with the specific effect of the Case II pattern (now referred to \( x_3 \)). Starting at t=21min, queue dissipation from the front of queue is generated by the increased outflow to 4000v/h (originally at the assumed speed of \( V_T^\mu = 64 \, \text{km/h} \), meaning an implied density of \( K_T^\mu = 62.5 \, \text{v/km} \) and \( \bar{c}_n = 32m \)). Again, using \( \ell_v = 25m \) for a moving queue, the estimate dissipation time can be obtained by evaluating the queue reduction initially as
\[ \bar{n}_q = \frac{(4000 - 3200) t_q}{1 - 4000/2 \times 0.025/64} \] and after \( x_2 \) as
\[ \bar{n}_q = \frac{(3600 - 2800) t_q}{1 - 3600/2 \times 0.025/64} \] (without distinction of the change in outflow condition at 3600v/h) and the extended queue estimated initially with
\[ \tilde{n}_q = \frac{4000/64}{2/0.025} \bigg\{ (4000 - 3200) t_q \bigg\} \] and after \( x_2 \) with
\[ \tilde{n}_q = \frac{3600/64}{2/0.025} \bigg\{ (3600 - 2800) t_q \bigg\} \] (without distinction of the change in outflow condition at 3600v/h).
The only further assumption was related to the speed at the queue discharging flow (assumed as 64km/h for S=4000v/h, then at 62.5v/km). Using more detailed information, it is known that the restrained queue is flowing at the density of 72v/km, then initially \( \bar{n}_q = \frac{(4000 - 3200)t_q}{1 - 62.5/72} \) and \( \bar{n}_q = \frac{62.5}{72} \times \frac{(4000 - 3200)t_q}{1 - 62.5/72} \),

and \( \bar{n}_q = \frac{(3600 - 2800)t_q}{1 - 56/88} \) and \( \bar{n}_q = \frac{56}{88} \times \frac{(3600 - 2800)t_q}{1 - 56/88} \) after \( x_2 \) (as the restrained flows of 3600v/h and 2800v/h are flowing at densities of 56v/km and 88v/h). The better information on queueing densities can be filled by any queue discharging model (with few parameters if accepting the linear hypothesis).

From the previous discussion, using the simplest assumptions, the queue from \( x_3 \) to \( x_2 \) has \( n_{ql} = 2 \times \frac{3.3}{0.025} = 264 \) and dissipates when \( \bar{n}_q = \frac{(4000 - 3200)t_q}{1 - 4000/2 \times 0.025/64} = 264 \) at \( t_{ql} = 4.3 \) min (with better information on queueing densities the condition is \( \bar{n}_q = \frac{(4000 - 3200)t_q}{1 - 62.5/72} = n_{ql} = 3.3 \times 72 = 240 \) and then \( t_{ql} = 2.4 \) min, the same value obtained with the FOTM). The number of vehicles in free flow queue discharging is evaluated as \( \bar{n}_q = \frac{4000/64}{2/0.025} = \frac{4000 - 3200}{1 - 4000/2 \times 0.025/64} \times \frac{4.3}{60} = \frac{62.5}{80} \times 264 = 206 \) (with better information on queueing densities the evaluation is \( n_{ql} = 3.3 \times 62.5 = 208 \), as with the FOTM). Similarly, the estimation process can follow the other scenarios of queue evolution.