CONSISTENCY OF SHOCK-WAVE AND QUEUING THEORY PROCEDURES FOR ANALYSIS OF ROADWAY BOTTLENECKS

Hesham Rakha[1] and Wang Zhang[2]

Word count: 3,913
Tables and figures: 2,500
Total words: 6,413
TRB Paper: 05-1763

ABSTRACT

A number of literatures have described queuing and shock-wave analyses as separate tools for solving bottleneck problems. Although some interrelationships between the two methods have been described in the literature, a number of these literature have claimed that deterministic queuing theory and shock-wave analysis are fundamentally different producing different delay estimates (McShane and Roess (*1*); Nam and Drew (*2*)). For example, Nam and Drew (*2*) mention that "*deterministic queuing analysis always underestimates the overall magnitude of delays compared to shock-wave analysis*." Furthermore, Nam and Drew claim that "*the area between the demand and capacity curves in a queuing diagram is analytically equivalent to the total vehicle-hours of travel in congestion as opposed to the widely accepted total vehicle-hours of delay*." Alternatively, this paper demonstrates the consistency in delay estimates that are derived from queuing theory and shock-wave analysis and highlights the common errors that are made in the literature with regards to shock-wave analysis delay estimation. Furthermore, the paper demonstrates that the area between the demand and capacity curves can represent the total delay or the total vehicle-hours of travel if the two curves are spatially offset (i.e. the count locations are at different spatial locations on the highway).

INTRODUCTION

A cumulative plot is the graph of a function *N(t)* that gives the cumulative number of vehicles (or other moving objects) that pass an observer at time *t* starting from an arbitrary initial count (Daganzo (*3*)) and are used to analyze the flow of items past a number of bottlenecks. Their usefulness in hydrologic synthesis was recognized for over a century and form the basis for a technique known as 'mass curve analysis' for determining capacity of reservoirs. Cumulative plots were introduced to the transportation arena by Moskowitz (*4*) and then again by Gazis and Potts (*5*), but Newell demonstrated its full potential as an analysis tool (Daganzo (*3*)). These cumulative plots are used to describe how items compete for service time through a node.

Traffic flow can be characterized using flow, density and speed through an analogy with fluid dynamics. Lighthill and Whitham (*6*), as well as Richards (*7*), made the first successful attempts at such a description. They both demonstrated the existence of traffic shock-waves and proposed a first theory of one-dimensional waves that could be applied to the prediction of highway traffic flow behavior. Equations 1 and 2 represent their model. The first equation defines the relation between flow, density, and speed that has

---

[1] Charles Via Jr. Department of Civil and Environmental Engineering, Virginia Tech, 3500 Transportation Research Plaza, Blacksburg VA, 24060. E-mail: hrakha@vt.edu.

[2] Virginia Tech Transportation Institute, 3500 Transportation Research Plaza, Blacksburg VA, 24060.

been developed from the application of fluid dynamics theory. Using Equation 1, Equation 2 was then developed to describe the speed at which a change in traffic characteristics, or a shock-wave, propagates along a roadway considering the conservation of flow at the shock-wave.

$$q_i = k_i \cdot u_i \quad [1]$$

$$w_{ij} = \frac{q_j - q_i}{k_j - k_i} \quad [2]$$

The main difference between shock-wave and deterministic queuing models is in the way vehicles are assumed to queue upstream of a bottleneck. While queuing analysis assumes vertical queuing, shock-wave analysis considers the spatial dimension of queues. The consideration of the horizontal extent of a queue enables the capturing of more realistic queuing behavior and the determination of the maximum queue reach, which is not possible with deterministic queuing models, as these models only track the number of queued vehicles, not their spatial location.

Delay is a measure of the additional travel time associated with traversing a specific distance relative to some base case travel time. Several literature have claimed that queuing theory and shock-wave analysis yield inconsistent delay estimates. For example, McShane and Roess (*1*) demonstrated that shock-wave analysis over-estimates average vehicle delays by as much as 60 percent in comparison with queuing theory; Chin (*8*) estimated differences in delay estimates in the range of 5 percent but concluded that because the differences were not too large that it could be argued that they were consistent. Finally, Nam and Drew (*2*) mention that "*deterministic queuing analysis always underestimates the overall magnitude of delays compared to shock-wave analysis*." Furthermore, Nam and Drew claim that "*the area between the demand and capacity curves in a queuing diagram is analytically equivalent to the total vehicle-hours of travel in congestion as opposed to the widely accepted total vehicle-hours of delay*." Unfortunately, these claimed differences result from inconsistencies in computing the base case travel time, which is used to compute the delay.

The main objective of this paper is to demonstrate, contrary to what is suggested in the some literature, the consistency in delay estimates between queuing theory and shock-wave analysis. Furthermore, the paper demonstrates that the claim by Nam and Drew that the area between the demand and capacity curves in a queuing diagram is equivalent to the total travel time as opposed to total delay is partially correct. The paper demonstrates when this claim is valid and when it is not valid. Finally, the paper corrects the equations that were derived in Nam and Drew's publication and demonstrates the consistency of the two formulations considering a fixed-capacity time-varying arrival rate bottleneck and a variable-capacity bottleneck.

In terms of the paper layout, the first section describes the traffic dynamics approach that was proposed by Nam and Drew (*2*) together with an example application of the model. Subsequently, corrections are made to the Nam and Drew proposed vehicle dynamics approach. The models are applied to two bottleneck problems, one for a constant capacity bottleneck with time-varying arrivals and another for a temporally varying capacity bottleneck. Finally, the conclusions of the paper are presented.

NAM AND DREW TRAFFIC DYNAMICS APPROACH

Traffic Dynamics Analysis

Nam and Drew (*2*) presented an example in which traffic flow on a freeway exceeds the capacity of a bottleneck. In this case an observer will observe two flows; the arrival rate and the queuing discharge rate. The flow and density of the approaching traffic stream are denoted as q_{n1} and k_{n1}, respectively, and those of the queuing flow are denoted as q_q and k_q, respectively, as illustrated in Figure 1. For a roadway segment between x_1 and x_2, the length of the queue discharge flow regime and arrival flow regime at time t are l_1 and

l_2, respectively. After some time Δt, the domains of the traffic regimes change to l_3 and l_4, respectively. Consequently, over the time interval Δt the congestion domain grows backward by l_3-l_1.

Using the principle of conservation of vehicles the queuing rate can be computed as

$$\left(q_{n1} - q_q\right)\Delta t \qquad [3]$$

The change in the number of vehicles traveling on the segment is given as

$$\left(k_q l_3 + k_{n1} l_4\right) - \left(k_q l_1 + k_{n1} l_2\right) \qquad [4]$$

Recognizing that the total length is constant (i.e. $l_1 + l_2 = l_3 + l_4$), [4] can be written as

$$\left(k_q - k_{n1}\right)\left(l_3 - l_1\right) \qquad [5]$$

Utilizing the principle of conservation of vehicles, [3] and [5] are equal

$$\left(q_{n1} - q_q\right)\Delta t = \left(k_q - k_{n1}\right)\left(l_3 - l_1\right). \qquad [6]$$

Re-arranging the terms of [6] the following relationship can be obtained:

$$\frac{l_1 - l_3}{\Delta t} = \frac{q_q - q_{n1}}{k_q - k_{n1}}. \qquad [7]$$

Here $\frac{l_1 - l_3}{\Delta t}$ is the speed of frontal boundary between the two flow regimes (i.e. it is the speed of the shock-wave between the arrival and queue departure regimes). In other words the speed of the shock-wave is equivalent to the difference in flow divided by the difference in density between the two flow regimes, as follows:

$$w_u = \frac{q_q - q_{n1}}{k_q - k_{n1}}, \qquad [8]$$

which is consistent with Equation 2 that was presented earlier.

If n is the number of vehicles in queue, the queuing rate in Δt is

$$\frac{\Delta n}{\Delta t} = \frac{k_q\left(l_3 - l_1\right)}{\Delta t}. \qquad [9]$$

Using [7] and [8], Nam and Drew manipulated [9] to derive

$$\frac{\Delta n}{\Delta t} = \left(q_{n1} - q_q\right) + \frac{l_3 - l_1}{\Delta t} k_{n1} = \left(q_{n1} - q_q\right) - w_u k_{n1}. \qquad [10]$$

Nam and Drew demonstrated that [10] is identical to the queuing rate obtained from shock-wave analysis. Using [8] they demonstrated that [10] is equivalent to

$$\frac{\Delta n}{\Delta t} = \left(q_{n1} - q_q\right)\left(1 + \frac{k_{n1}}{k_q - k_{n1}}\right), \qquad [11]$$

which implies that deterministic queuing analysis always underestimates the queue length by a factor of $\frac{k_{n1}}{k_q - k_{n1}}$.

After the conclusion of the peak demand, the traffic demand decreases below the capacity of the bottleneck and the queue length starts to decrease in size. Considering q_{n2} and k_{n2} to represent the flow and density of the reduced approaching flow, respectively, the discharging rate is opposite to the queuing rate and can be estimated as

$$-\frac{\Delta n}{\Delta t} = \left(q_q - q_{n2}\right) + w_d k_{n2} = \left(q_q - q_{n2}\right)\left(1 + \frac{k_{n2}}{k_q - k_{n2}}\right) \qquad [12]$$

where $w_d = \frac{q_{n2} - q_q}{k_{n2} - k_q}$ is the speed of the forward recovery shock-wave. Equation 12 demonstrates that deterministic queuing theory underestimates the queue recovery rate by a factor of $\frac{k_{n2}}{k_q - k_{n2}}$.

Comparative Analysis

Nam and Drew then compared queuing theory and shock-wave analysis procedures using a constant capacity bottleneck example illustration, as will be described in this section.

In a queuing diagram the queuing rate is equal to q_{n1}-q_q during the peak period, and the discharge rate is equal to q_q-q_{n2} during the discharge period, as illustrated in Figure 2. In the case of queuing theory, the maximum queue occurs when the traffic demand drops from q_{n1} to q_{n2} at time t_1 given that queuing theory considers that vehicles queue in a fictitious vertical queue. Congestion disappears at time t_2 when the queue is served. The spatial extent of the queue can be estimated by scaling the queuing diagram and dividing by the density of traffic in the queue k_q.

Considering a time-space diagram Nam and Drew constructed the spatial formation of queues considering that the queues build up at a rate of $(q_{n1}$-q_q-$w_u k_{n1})/k_q$ and that the queue dissipates at a rate of $(q_q$-$q_{n2}+w_d k_{n2})/k_q$., as demonstrated in Figure 2. Nam and Drew demonstrated that the queue reaches its maximum rate at time t_1-t_0, where t_0 is the travel time in the absence of congestion from the tail of the physical queue upstream of the bottleneck to the bottleneck. Similarly, the equivalent queuing diagram was constructed by multiplying by the traffic density within the congested regime k_q.

Using the geometry of Figure 2, the maximum queue length by shock-wave analysis is computed as

$$l_{\max} = -w_u\left(t_1 - t_0\right), \qquad [13]$$

where

$$t_0 = t_1 - \left[\frac{w_d}{w_d - w_u}\right] t_2 \text{ and} \qquad [14]$$

$$t_2 = \left[1 + \frac{q_{n1} - q_q}{q_q - q_{n2}}\right] t_1 . \qquad [15]$$

The total travel time (*TTT*) is computed as the area of the time-space domain of congestion multiplied by the density of traffic under congestion as

$$TTT = \frac{l_{\max} t_2}{2} k_q . \qquad [16]$$

Substituting [13] in [16] Nam and Drew derived

$$TTT = -\frac{w_u}{2} k_q \left(t_1 - t_0\right) t_2 . \qquad [17]$$

Using the modified arrival and departure curves in Figure 3, Nam and Drew computed the area as

$$\begin{aligned}\frac{1}{2}\left[Q_1 t_2 - Q_2\left(t_1 - t_0\right)\right] &= \frac{1}{2}\left[\left(q_{n1} - w_u k_{n1}\right)\left(t_1 - t_0\right)t_2 - q_q t_2\left(t_1 - t_0\right)\right] \\ &= \frac{1}{2}\left(q_{n1} - w_u k_{n1} - q_q\right)\left(t_1 - t_0\right)t_2\end{aligned} \quad . \qquad [18]$$

Subtracting [18] from [17] Nam and Drew demonstrated that the difference was zero as follows:

$$\begin{aligned}&\left[-\frac{w_u}{2}k_q\left(t_1 - t_0 t_2\right)\right] - \left[\frac{1}{2}\left(q_{n1} - w_u k_{n1} - q_q\right)\left(t_1 - t_0\right)t_2\right] \\ &= -\frac{1}{2}\left(w_u k_q + q_{n1} - w_u k_{n1} - q_q\right)\left(t_1 - t_0\right)t_2 \\ &= -\frac{1}{2}\left[w_u\left(k_q - k_{n1}\right) - \left(q_q - q_{n1}\right)\right]\left(t_1 - t_0\right)t_2 = 0\end{aligned} \quad . \qquad [19]$$

Consequently, Nam and Drew concluded that the area between the arrival and departure curves (or demand and capacity curves) in a deterministic queuing diagram is equivalent to the total vehicle-hours of travel in congestion as opposed to the widely accepted total vehicle-hours of delay.

Nam and Drew then derived formulations for the various descriptive variables based on queuing theory and shock-wave analysis, as summarized in Table 1and Table 2, respectively.

TRAFFIC DYNAMICS

Overview of Traffic Dynamics

Daganzo (*3*) mentions that "*since horizontal separations in the (t,N) diagram represent time and vertical separations represent accumulation, it should not come as a surprise that the area of the region enclosed by A(t), D(t) and any two vertical lines,* $t=t_0$ *and* $t=t_1$ $(t_0 < t_1)$*, should be the total wait time done in the system in the interval* (t_0, t_1)." In this analysis *A(t)* represents the cumulative arrivals at any instant *t* at some count station while *D(t)* represents the cumulative departures at any instant *t* at another downstream count station. Similarly, Daganzo demonstrates that the wait done by items N_0+1 through N_1, ($N_0 < N_1$) in a First-In-First-Out (FIFO) system is given by the area of the region enclosed by curves *A(t)* and *D(t)* and by the horizontal lines, $N = N_0$ and $N = N_1$. The second deduction does not hold for systems with passing except if the arrival and departure curves touch one another when $N = N_0$ and $N = N_1$, as is the case in Figure 3 because all items that entered must have departed within the identified time interval (t_0, t_1). Daganzo also mentions that the result is approximately true if the combined wait of items that only arrived or only departed in the observation period is a small fraction of the total wait.

It is important to note that delay is not necessarily equal to the time spent between the two count stations (time spent within the system). Specifically, the time spent within the system includes the time required to travel between the upstream and downstream count stations in the absence of congestion plus any additional time spent in the system as a result of congestion (known as delay). Here a virtual arrival curve can be introduced by shifting the arrival curve to the right by the free-flow travel time (distance d_q/v_f as shown in Figure 4). Therefore, for N_0, t_0 the distance between arrival curve *A(t)* and departure curve *D(t)* represents the total travel time within the system, while the distance between the virtual arrival curve *V(t)* and the departure curve *D(t)* represents the total delay within the system. It is important to note that if the arrival and departure curves touch one another at any instant then the area between the two curves represents the total delay and not total travel time given that there is no temporal offset between the two curves. Alternatively, if the two curves are temporally offset (e.g. located at different locations along a highway) the area between the two curves represents the total travel time.

The travel time between the arrival and departure count stations during the onset of congestion can be computed as

$$t_q = \frac{d_q}{v_q}. \qquad [20]$$

Here d_q is defined as the distance traveled in queue and v_q is the travel speed in queue. As discussed before the delay can be computed as

$$w = t_q - \frac{d_q}{v_f}, \qquad [21]$$

where v_f is the free-flow speed. Here the assumption is that the base travel is at free-speed. Consequently, the distance & time in queue is estimated as $d_q = \frac{w}{\frac{1}{v_q} - \frac{1}{v_f}}$ and $t_q = \frac{w}{1 - \frac{v_q}{v_f}}$, respectively.

The above computations assume that the base-case (no congestion) scenario involves travel at free-flow speed, which may not necessarily be true. Consequently, as part of this research effort we generalize the formulation to consider multiple traffic regimes for the base case, as described in the next section.

Proposed Modifications to the Nam and Drew Formulation

As was demonstrated in the previous section, the area between the arrival and departure curves can either represent the total travel time or the total delay depending on whether there is a spatial separation between the arrival and departure measurements. Consequently, Nam and Drew's conclusion that the area between the demand and the capacity curves in a queuing diagram is analytically equivalent to the total hours of travel in congestion as opposed to the widely accepted total vehicle-hours of delay is partially correct and only holds if the arrival and capacity curves are offset spatially, as was demonstrated by Daganzo (*3*). Alternatively, if the arrival and departure curves are not spatially offset, the area between these two curves represents the total delay. Daganzo demonstrated that the delay can be computed by offsetting the arrival curve a distance equal to the free-flow travel time to the right to create a virtual arrival curve. The area between the virtual arrival curve and the departure curve then represents the total delay. However, Daganzo's approach is valid only if the base case for which we are computing delay relative to involves travel at the free-speed.

In the case of the Nam and Drew derivation, because the demand and capacity curves (also known as arrival and departure curves) of Figure 3 coincided with each other at both ends of the figure, the area between these curves represents the total delay and not the total travel time as was suggested by Nam and Drew.

Furthermore, Nam and Drew claimed that the queue extent derived from queuing theory and shock-wave analysis in Figure 2 are different and thus they concluded that the delay estimates from both models would differ. However, it is important to note at this point that the queue extent that is derived from the queuing model represents the number of vehicles that would have been seen directly upstream of the restriction by time *t* if the physical extent of the queue had been eliminated and thus should not be compared with the shock-wave queue extent estimates. In other words the queuing theory queue extent is a fictitious queue extent. In order to compare the queue extents for the identical conditions, the queuing theory queue extent should be scaled by the ratio of the total travel time (*TTT*) to the total delay (*TD*). In doing so the extent of the queues are identical and thus the delays computed are also identical, as will be demonstrated in the following section.

Nam and Drew also made a third error in their analysis while computing the total delay using shock-wave analysis procedures. Specifically, they derived the total delay as

$$\frac{-w_u t_2}{2}(t_1 - t_0)(k_q - k_{n2}). \qquad [22]$$

Implicit in Equation 22 is the assumption that the base case traffic stream density in the absence of congestion is k_{n2}, as demonstrated in Figure 5. However, given that the initial arrival rate travels at a density of k_{n1} for the duration of t_1 and subsequently changes to k_{n2}, the base case condition involves two regimes with densities k_{n1} and k_{n2}, as illustrated in Figure 5. Consequently, with this in mind the total delay should be computed as

$$\frac{-w_u(t_1-t_0)}{2}\left[t_2\left(k_q-k_{n2}\right)-t_1\left(k_{n1}-k_{n2}\right)\right]. \qquad [23]$$

Equation 23 expands on Daganzo's procedures by considering a base case composed of two traffic regimes instead of the a single regime that involves travel at free-flow speed.

In addressing the errors of the Nam and Drew (*2*) formulations, we have developed new formulations for estimating delay using shockwave analysis and queuing theory, as summarized in Table 1 and Table 2. In the following sections, we shall demonstrate that the delay estimates from the shock-wave analysis and queuing theory are identical using the proposed formulations.

EXAMPLE APPLICATIONS

In this section we compare shock-wave and queuing theory delay estimates for two scenarios. The first scenario involves a time-varying arrival rate at a constant capacity bottleneck. This example is identical to the example presented by Nam and Drew. In this example we demonstrate the error in the Nam and Drew formulation and further demonstrate the consistency between shock-wave and queuing theory delay estimates. The second scenario compares shock-wave and queuing theory delay estimates for a constant arrival rate at a time-varying capacity bottleneck.

Variable Arrival Rate at a Constant Capacity Bottleneck

We use the example used by Nam and Drew in their paper. Specifically, consider a bottleneck on a six-lane (three in each direction) urban freeway. The capacity of the bottleneck is 1800 veh/h/lane and density of the congestion domain is 120 veh/mile/lane. The approaching flow is 6000 veh/h and its density is 120 veh/mile for the first 2 hours, and then the approaching flow changes to 4500 veh/h with a density of 75 veh/mile.

Nam and Drew Solution:

First, we present the solution that Nam and Drew presented in an earlier publication (Nam and Drew (*2*)). The solutions are based on the formulations that were presented in Table 1 and Table 2.

$$w_u = \frac{q_{n1}-q_q}{k_{n1}-k_q} = \frac{\left(\frac{6000}{3}-1800\right)}{40-120} = -2.5\,\text{mph}$$

$$w_d = \frac{q_q-q_{n2}}{k_q-k_{n2}} = \frac{\left(1800-\frac{4500}{3}\right)}{120-25} = 3{,}158\,\text{mph}$$

The time variables are computed as

$t_1 = 2.0\text{ h}$ (Given),

$$t_2 = \left[1+\frac{q_{n1}-q_q}{q_q-q_{n2}}\right]t_1 = \left[1+\frac{\frac{6000}{3}-1800}{1800-\frac{4500}{3}}\right]\times 2.0 = 3.333\text{ h}$$, and

$$t_0 = t_1 - \left[\frac{w_d}{w_d - w_u}\right] t_2 = 2.0 - \frac{3.158}{3.158 - (-2.500)} \times 3.333 = 0.140 \text{ h}.$$

The maximum queue length based on shock-wave analysis is computed as

$$l_{\max} = \left[\frac{q_{n1} - q_q - w_u k_{n1}}{k_q}\right](t_1 - t_0) = \left(\frac{6000 - 5400 - (-2.5) \times 120}{360}\right) \times (2 - 0.14) = 4.65 \text{ mi}.$$

Alternatively, the maximum queue length based on queuing theory is computed as

$$l'_{\max} = \left[\frac{q_{n1} - q_q}{k_q}\right] t_1 = \left[\frac{6000 - 5400}{360}\right] \times 2.0 = 3.333 \text{ mi}.$$

The total travel time (*TTT*) and total delay (*TD*) that are computed using shock-wave analysis are

$$TTT_s = \frac{(t_1 - t_0)^2}{2}\left(1 - \frac{w_u}{w_d}\right)(q_{n1} - q_q - w_u k_{n1}) = \frac{(2 - 0.14)^2}{2}\left(1 - \frac{-2.5}{3.158}\right)(6000 - 5400 - (-2.5) \times 120) = 2789 \text{ veh},$$

$$TD_s = \frac{-w_u t_2}{2}(t_1 - t_0)(k_q - k_{n2}) = \frac{2.5 \cdot 3.333}{2}(2 - 0.14) \cdot (360 - 75) = 2208.53 \text{ veh.h}.$$

Alternatively, the total travel time (*TTT*) and total delay (*TD*) based on queuing theory are computed as

$$TTT_q = \frac{t_1^2}{2}\left(1 + \frac{q_{n1} - q_q}{q_q - q_{n2}}\right)(q_{n1} - q_q) = \frac{2^2}{2}\left(1 + \frac{6000 - 5400}{5400 - 4500}\right) \cdot (6000 - 5400) = 2000 \text{ veh.h}, \text{ and}$$

$$TD_q = \frac{t_1 t_2}{2k_q}(q_{n1} - q_q)(k_q - k_{n2}) = \frac{2 \times 3.333}{2 \times 360}(6000 - 5400)(360 - 75) = 1583.2 \text{ veh.h}.$$

Proposed Model Solution:

Nam and Drew's formulation assumed the density of the base case to be constant and equal to the density of the second arrival rate, however, the base case density involves two density regimes for each of the arrival rates. Consequently, the proposed solution that is developed as part of this research effort and presented in Table 1 and Table 2 is summarized as follows.

The total travel time (*TTT*) and total delay (*TD*) by shock-wave analysis is computed as

$$TTT'_s = \frac{(t_1 - t_0)^2}{2}\left(1 - \frac{w_u}{w_d}\right)(q_{n1} - q_q - w_u k_{n1}) = \frac{(2 - 0.14)^2}{2}\left(1 - \frac{-2.5}{3.158}\right)(6000 - 5400 - (-2.5) \cdot 120)$$
$$= 2789 \text{ veh.h}$$

$$TD'_s = \frac{-w_u(t_1 - t_0)}{2}\left[t_2(k_q - k_{n2}) - t_1(k_{n1} - k_{n2})\right] = \frac{2.5 \cdot (2 - 0.14)}{2}(3.333 \cdot (360 - 75) - 2 \cdot (120 - 75))$$
$$= 1999.3 \cong 2000 \text{ veh.h}$$

Alternatively, the total travel time (*TTT*) and total delay (*TD*) estimated by queuing theory is as follows:

$$TTT'_q = \frac{t_1^2}{2}\left(1 + \frac{q_{n1} - q_q}{q_q - q_{n2}}\right)(q_{n1} - q_q) = \frac{2^2}{2}\left(1 + \frac{6000 - 5400}{5400 - 4500}\right) \cdot (6000 - 5400) = 2000 \text{ veh.h} \text{ and}$$

$$TD'_q = \frac{t_1^2}{2}(q_{n1} - q_q)\left(1 + \frac{q_{n1} - q_q}{q_q - q_{n2}}\right) = \frac{2^2}{2}(6000 - 5400) \cdot \left(1 + \frac{6000 - 5400}{5400 - 4500}\right) = 2000 \text{ veh.h}.$$

Note that the total travel time and total delay estimates are identical for the queuing theory formulation because the arrival and departure curves are not spatially offset. The example demonstrates that both approaches produce identical delay estimates and that the differences in delay estimates that were reported by Nam and Drew are attributed to the fact that their computation failed to recognize the fact that the base case scenario includes two density regimes. A summary of the various computations are provided in Table 3.

Constant Arrival and Time Varying Capacity Bottleneck

In addition to analyzing a constant capacity bottleneck we include an additional example which includes a time-varying capacity bottleneck. In this example we consider a signalized intersection with single lane approaches. The traffic signal timings include a cycle length of 60 seconds and an effective green time of 40 seconds, which corresponds to an effective red time of 20 seconds. The approach flow is 900 veh/h with a density of 15 veh/km. The saturation flow rate is 1800 veh/h with a density of 50 veh/km. The jam density is assumed to be 100 veh/km.

Shock-wave Analysis:

Based on Figure 6 the speeds of the various shock-waves can be computed as

$$w_{CB} = \frac{q_C - q_B}{k_C - k_B} = \frac{900-0}{15-100} = -10.6 \text{ km/h} ,$$

$$w_{BA} = \frac{q_B - q_A}{k_B - k_A} = \frac{0-1800}{100-50} = -36 \text{ km/h} , \text{ and}$$

$$w_{AC} = \frac{q_A - q_C}{k_A - k_C} = \frac{1800-900}{50-15} = 25.7 \text{ km/h} .$$

The time for the backward forming and backward recovery shock-waves to meet is computed as

$$w_{CB}\left(r + t_m\right) = w_{BA} t_m = d$$

$$t_m = 8.35 \text{ s} ; \; d = -\frac{w_{BA} \times t_m}{3.6} = 83.5 \text{ m} ; \; t_c = \frac{d}{w_{AC}} = 11.7 \text{ s} .$$

Finally, the total delay is computed as the area of the congested triangle multiplied by the difference between the congested and base densities, as

$$TD_s = \frac{d \times r}{2} * k_B + \frac{d \times \left(t_m + t_c\right)}{2} \times k_A - \frac{d \times \left(r + t_m + t_c\right) \times k_C}{2} = 100.2 \text{ veh} \cdot \text{s} .$$

Queuing Analysis:

Using Figure 7 the arrival rate is computed as

$$q = \frac{q_C}{3600} = \frac{900}{3600} = 0.25 \text{ veh/s} ,$$

while the saturation rate is computed as

$$s = \frac{q_A}{3600} = \frac{1800}{3600} = 0.5 \text{ veh/s} .$$

The maximum number of vehicles in queue is computed as

$$Q_m = qr = 0.25 \times 20 = 5 \text{ veh}$$

and the time to clear the queue is computed as

$$T = \frac{Q_m}{s - q} = \frac{5}{0.50 - 0.25} = 20 \text{ s} .$$

The total delay is computed as

$$TD_q = \frac{1}{2}(r + T)Q_m = \frac{(20 + 20) \times 5}{2} = 100 \text{ veh} \cdot \text{s} .$$

Again, the delay estimates as was the case with the variable arrival/constant capacity scenario, estimated by the queuing theory and shock-wave analyses are practically identical. Consequently, we demonstrate the consistency of both formulations in estimating delay at bottlenecks.

CONCLUSIONS

The delay computations using shock-wave analysis and queuing theory were compared for two example applications, namely (a) time varying arrival rate at a constant capacity bottleneck and (b) a constant arrival rate at a time varying capacity bottleneck. The results demonstrate the consistency between shock-wave analysis and queuing theory. Furthermore, the paper highlights the error in the Nam and Drew (*2*) computation and corrects the equations that were derived by Nam and Drew. In summary, the paper demonstrates that queuing theory provides a simple and accurate technique for estimating delay at highway bottlenecks.

ACKNOWLEDGEMENTS

The authors acknowledge the financial support of the U.S. DOT ITS Implementation Center in conducting this research effort.

REFERENCES

1. McShane W.R. and Roess R.P. *Traffic Engineering*, Prentice-Hall, Inc., ISBN 0-13-926148-6, 1990, pp. 623-632.

2. Nam D.H. and Drew D.R. Analyzing Freeway Traffic under Congestion: Traffic Dynamics Approach, *Journal of Transportation Engineering*, May/June, 1998, pp. 208-212.

3. Daganzo C.F. *Fundamentals of Transportation and Traffic Operations*, Elsevier Science Ltd., ISBN 0-08-0422785-5, 2000.

4. Moskowitz K. and Newman L. Notes on Freeway Capacity, *Highway Research Board*, No. 27, 1963.

5. Gazis D.C. and Potts R.B. The Over-saturated Intersection, *Proc. 2nd Int. Symp. On the Theory of Road Traffic Flow*, 1965, pp. 221-237.

6. LightHill M.J. and Whitham G.B. On Kinematic Waves: A Theory of Traffic Flow on Long Crowded Roads, *Proc., of Royal Society*, Series A, Vol., 229, No. 1178. London, 1955, pp 317-345.

7. Richards P.I. Shock Waves on Highway, *Operation Research*, 4 (1), 1956, pp 42-51.

8. Hoong-Chor Chin, A Reexamination of the Analysis of Freeway Bottlenecks, *Journal of Transportation Engineering*, Jan, 1996, pp. 30-35.

LIST OF TABLES

LIST OF FIGURES

Table 1: Queuing Theory Formulations

Variables	**Nam and Drew Formulation**	**Proposed Formulation**
Maximum number of vehicles in queue	$(q_{n1} - q_q)t_1$	$(q_{n1} - q_q)t_1$
Maximum queue length	$\left[\frac{q_{n1} - q_q}{k_q}\right]t_1$	$\left[\frac{q_{n1} - q_q}{k_q}\right]t_1$
Total delay	$\frac{t_1 t_2}{2k_q}(q_{n1} - q_q)(k_q - k_{n2})$	$\frac{t_1^2}{2}(q_{n1} - q_q)\left(1 + \frac{q_{n1} - q_q}{q_q - q_{n2}}\right)$ When arrival and departure counts are at different locations.
Total travel time (in congestion)	$\frac{t_1^2}{2}\left(1 + \frac{q_{n1} - q_q}{q_q - q_{n2}}\right)(q_{n1} - q_q)$	$\frac{t_1^2}{2}(q_{n1} - q_q)\left(1 + \frac{q_{n1} - q_q}{q_q - q_{n2}}\right)$ When arrival and departure counts are at different locations.
Average individual delay	$\frac{t_1}{2q_q k_q}(q_{n1} - q_q)(k_q - k_{n2})$	$\frac{t_1}{2q_q}(q_{n1} - q_q)$ When arrival and departure counts are at different locations.
Average individual travel time (in congestion)	$\frac{t_1}{2q_q}(q_{n1} - q_q)$	$\frac{t_1}{2q_q}(q_{n1} - q_q)$ When arrival and departure counts are at different locations.

Table 2: Shock-wave Analysis Formulations

Variables	Nam and Drew Formulation	Proposed Formulation
Maximum number of vehicles in queue	$\left(q_{n1} - q_q - w_u k_{n1}\right)\left(t_1 - t_0\right)$	$\left(q_{n1} - q_q - w_u k_{n1}\right)\left(t_1 - t_0\right)$
Maximum queue length	$\left[\frac{q_{n1} - q_q - w_u k_{n1}}{k_q}\right]\left(t_1 - t_0\right)$	$\left[\frac{q_{n1} - q_q - w_u k_{n1}}{k_q}\right]\left(t_1 - t_0\right)$
Total delay	$\frac{-w_u t_2}{2}\left(t_1 - t_0\right)\left(k_q - k_{n2}\right)$	$\frac{-w_u\left(t_1 - t_0\right)}{2}\left[t_2\left(k_q - k_{n2}\right) - t_1\left(k_{n1} - k_{n2}\right)\right]$
Total travel time (in congestion)	$\frac{\left(t_1 - t_0\right)^2}{2}\left(1 - \frac{w_u}{w_d}\right)\left(q_{n1} - q_q - w_u k_{n1}\right)$	$\frac{\left(t_1 - t_0\right)^2}{2}\left(1 - \frac{w_u}{w_d}\right)\left(q_{n1} - q_q - w_u k_{n1}\right)$
Average individual delay	$-\frac{w_u}{2q_q}\left(t_1 - t_0\right)\left(k_q - k_{n2}\right)$	$\frac{-w_u\left(t_1 - t_0\right)}{2q_q t_2}\left[t_2\left(k_q - k_{n2}\right) - t_1\left(k_{n1} - k_{n2}\right)\right]$
Average individual travel time (in congestion)	$\frac{1}{2q_q}\left(q_{n1} - q_q - w_u k_{n1}\right)\left(t_1 - t_0\right)$	$\frac{1}{2q_q}\left(q_{n1} - q_q - w_u k_{n1}\right)\left(t_1 - t_0\right)$

Table 3: Comparison of Queuing and Shock-wave Analysis for Example 4.1

Descriptive Items	Queuing Theory	Shock-wave Analysis	Difference	Units
Queuing Rate	600	900	300	veh/h
Discharging rate	900	1136.85	236.85	veh/h
Physical queuing rate	1.67	2.33	0.66	mph
Physical discharging rate	2.5	3.5	1	mph
Maximum number of vehicles in queue	1200	1674	474	veh
Maximum physical queue length	3.33	4.65	1.32	mile
Total veh-h of delay	2000	2000	0	veh-h
Total veh-h of travel in congestion	2000	2790	790	veh-h
Average individual delay	6.67	6.67	0	min
Average individual travel time	7.43	9.3	1.87	min

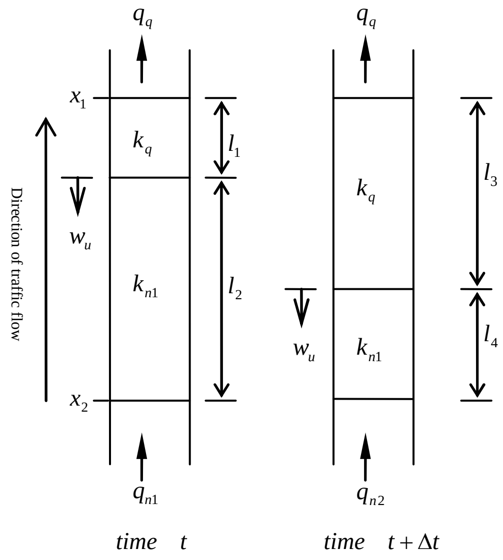


Figure 1: Schematic Representation of Congestion Buildup over *Δt* (Source: Nam and Drew, 1998)

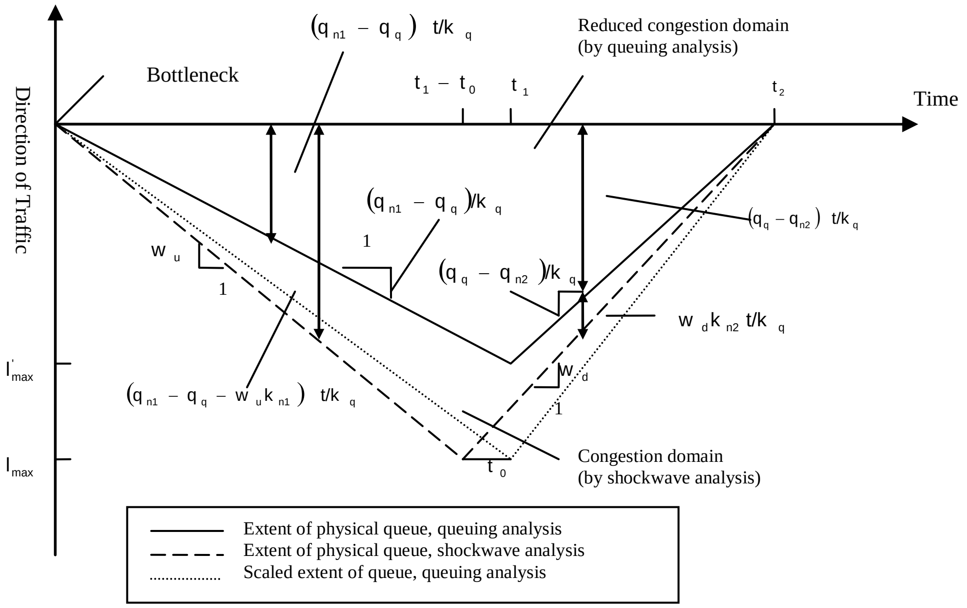


Figure 2: Superimposition of Time-Space Diagrams from Queuing and Shock-wave Analysis

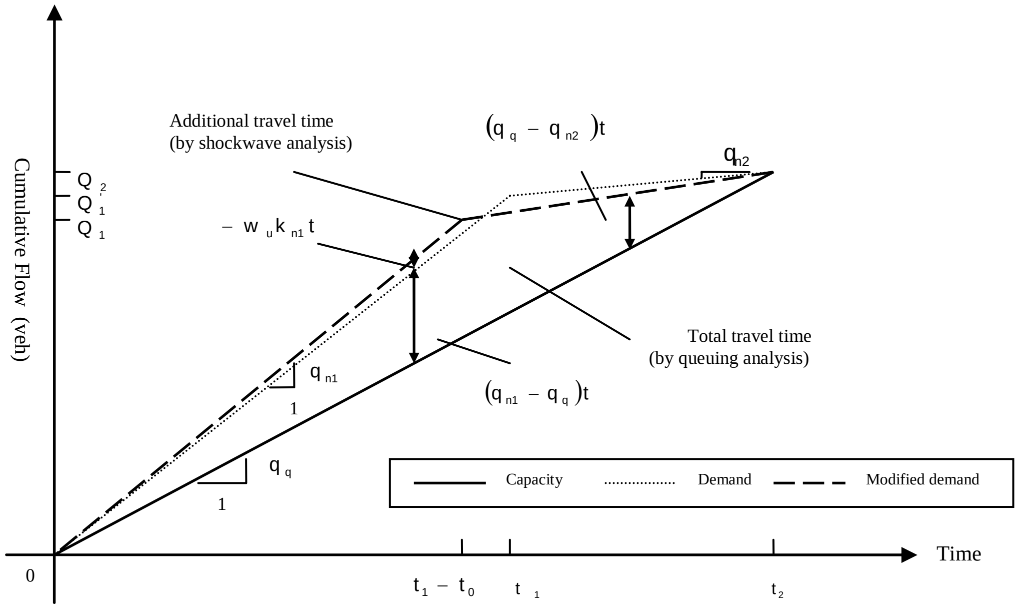


Figure 3: Superimposition of Queuing Theory and Shock-wave Analysis Diagrams

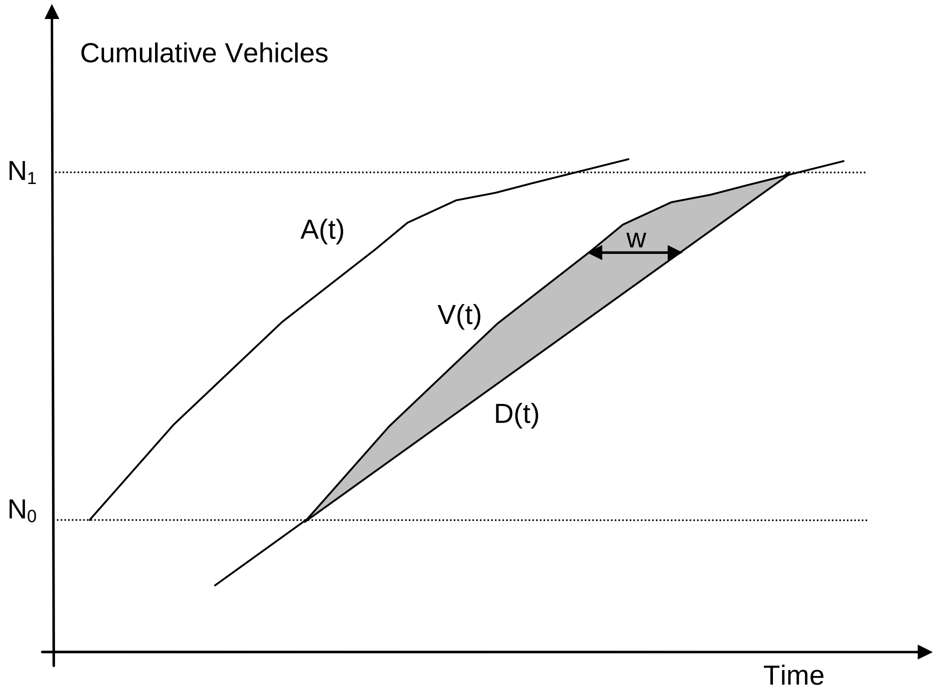
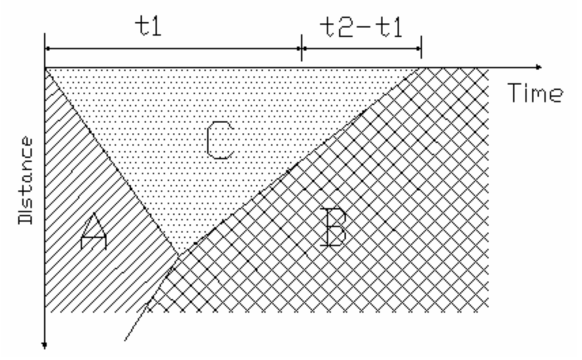
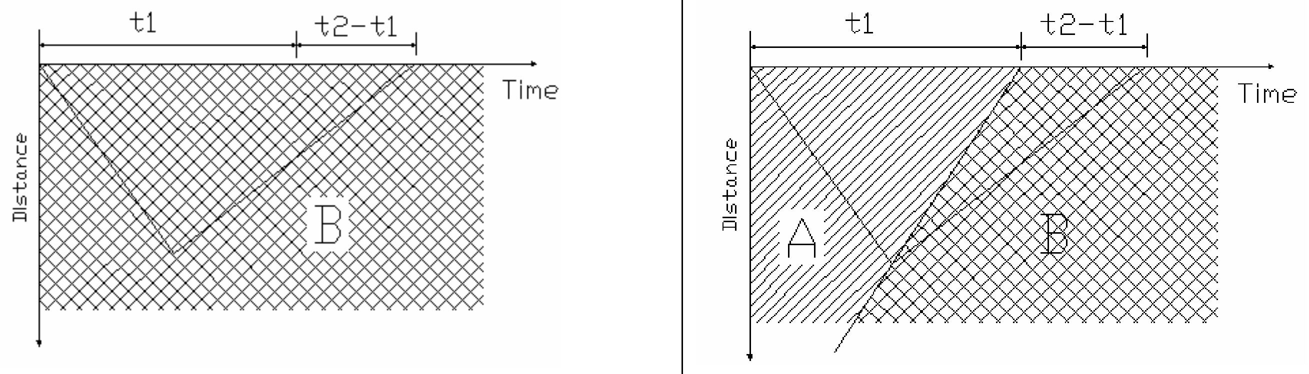


Figure 4: Arrival, Departure, and Virtual Arrival Curves for a Single Bottleneck with Variable Arrivals



(a) Traffic stream regimes with bottleneck



(b) Base no-bottleneck scenario as proposed by Nam and Drew (1998)

(c) Base no-bottleneck scenario

Figure 5: Traffic Flow Regimes with and without a Bottleneck

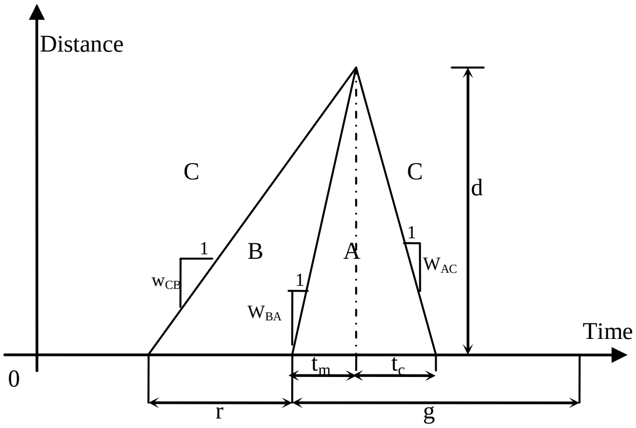


Figure 6: Time-Space Diagrams from Shock-wave Analysis (Constant Arrival Rate with Variable Capacity Bottleneck)

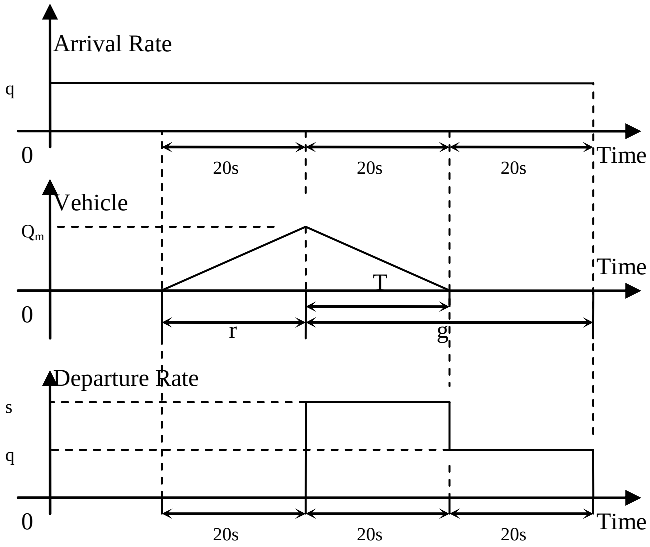


Figure 7: Queuing Diagram (Constant Arrival Rate with Variable Capacity Bottleneck)