PAPER No. 981292

A SIMPLE, GENERALIZED METHOD FOR ANALYSIS OF A TRAFFIC QUEUE UPSTREAM OF A BOTTLENECK

By

Alan L. Erera
Institute of Transportation Studies
University of California, Berkeley
109 McLaughlin Hall #1720
Berkeley, CA 94720-1720
Phone: (510) 642-9907 Fax: (510) 642-1246
E-mail: alerera@ieor.berkeley.edu

Tim W. Lawson
Institute of Transportation Studies
University of California, Berkeley
109 McLaughlin Hall #1720
Berkeley, CA 94720-1720
Phone: (510) 642-4609 Fax: (510) 642-1246
E-mail: tlawson@uclink4.berkeley.edu

Carlos F. Daganzo
Institute of Transportation Studies
University of California, Berkeley
109 McLaughlin Hall #1720
Berkeley, CA 94720-1720
Phone: (510) 642-3853 Fax: (510) 642-1246
E-mail: daganzo@euler.berkeley.edu

Submitted for publication,
Transportation Research Record

Words: 5,917 + 6 figures = 7,417
(including Abstract, 183 words)

April 1, 1998

ABSTRACT

This paper generalizes an approach for enhancing a standard input-output diagram to represent graphically the time and distance that vehicles spend in a queue upstream of a bottleneck. The approach requires the construction of a curve depicting the cumulative number of vehicles to have reached the back of the queue as a function of time. The original technique, described in a previous paper, is reviewed for bottlenecks with constant capacity, and for those where capacity changes once. The paper then generalizes the approach to allow multiple changes in bottleneck capacity, and relaxes the original assumption of a triangular flow-density (*q*-*k*) relation to one which is piecewise-linear concave. Although it is consistent with the kinematic wave theory of traffic flow, the proposed approach is simpler to apply to complex problems because it avoids the laborious construction of a time-space diagram. It allows the estimation of several measures required in the evaluation of the impacts of bottlenecks, including the (accurate) number of vehicles in queue and the physical extents of queues at any time, and the total time spent by vehicles in different traffic states.

Key Words:

Traffic Flow Theory, Input-Output Diagram, Evaluation, Queue(s), Bottleneck(s)

INTRODUCTION

This paper generalizes a method of enhancing a standard input-output (I/O) diagram to represent graphically the time and distance that vehicles spend in a queue upstream of a bottleneck. The time that a vehicle spends traversing a queue ("time in queue" or "waiting time") is greater than its *delay*, which represents the difference between the time a vehicle takes to traverse a distance and the time it would have taken if unobstructed. Unlike the standard I/O diagram, which quantifies only delay and an imprecise estimate of the number of vehicles in queue, the modified diagram proposed provides valuable additional information needed to evaluate the impacts of bottlenecks, including the total time (and distance) spent by vehicles in different queued states, the number of vehicles changing between states, and the physical extent of the queue at *any* time.

Reference (*1*) introduces the concept of modifying an I/O diagram by constructing a "back-of-queue" curve, $B(t)$. This curve represents the time that each vehicle reaches the upstream extent of the physical queue formed behind a bottleneck, or equivalently, the cumulative number of vehicles to have arrived at the back of the queue. Methods for constructing $B(t)$ in the case of a bottleneck with constant capacity and in the case of a restriction whose capacity changes once are presented in (*1*), and shown to be considerably simpler than conventional techniques for locating a back-of-queue trajectory on a time-space diagram.

Building on the ideas in (*1*), this paper describes a method for constructing back-of-queue curves for bottlenecks whose capacity may change multiple times. Furthermore, it generalizes the assumption that a linear relationship exists between flow and density in queued traffic. The proposed approach for constructing back-of-queue curves is shown to be consistent with the kinematic wave theory of traffic flow (*2*,*3*), but simpler to apply to complex problems.

The kinematic theory allows one to predict the evolution of traffic given initial and boundary conditions. The theory has three main components. The first is the assumption that some functional relation exists between traffic flow q and density k, which holds everywhere in the time-space (t,x) plane, such that $q = Q(k, x, t)$. For a homogeneous highway, this function is independent of time and space and the relation is simplified to $q = Q(k)$. If it is assumed that all vehicles travel at the space-mean speed, the relationship $q = kv$ further implies that wherever density is k, vehicles travel at speed $v = \frac{Q(k)}{k}$.

The second component of the theory predicts the characteristics of the discontinuous boundaries, or *interfaces*, between regions of the (t,x) plane where the traffic state, (q,k), varies smoothly. These interfaces, sometimes referred to as *shock waves*, usually do not remain stationary in space. The speed at which an interface travels is defined to be $\frac{\Delta q}{\Delta k}$, where Δq and Δk are the changes in flow and density between the separated states.

The third component of the kinematic theory centers on the idea of *stability*. For any physically meaningful problem of traffic dynamics, there may exist numerous solutions that satisfy the above two mathematical conditions, namely that the traffic state in each region in the (t,x) plane lies on the q-k curve and the interfaces between regions in different states have the correct slope. The complete wave theory predicts, however, that only one of these solutions—the *stable* solution—can exist in reality. A solution is stable if infinitesimal perturbations in the input data at time $t = 0$ (or to the solution itself at some intermediate time) induce only infinitesimal changes in the ensuing solution.

This paper begins by reviewing the results presented in (*1*). Methods for constructing back-of-queue curves are described, and an interpretation of the modified input-output diagram is presented. The paper then presents a generalization of the original approach to allow construction of back-of-queue curves for problems which have multiple bottleneck capacity changes and a linear q-k relation in queued traffic. Finally, the technique is extended to handle the more general case where the q-k relation is piecewise-linear concave.

SPATIAL AND TEMPORAL CHARACTERISTICS OF A QUEUE

Summarizing the results of (*1*), we begin by reviewing methods for constructing a back-of-queue curve for a bottleneck with a constant capacity, and for one whose capacity changes once at a known time. For both cases, it is assumed that a constant free-flow speed v_f holds for all uncongested traffic, independent of flow. Consistent with the kinematic theory, it is further assumed that a q-k relation holds for queued traffic, so that vehicles in the queue behind a bottleneck serving at a maximum rate m will travel at a constant speed v_m ($< v_f$), dependent only on the capacity. Further, it is assumed that speed changes occur instantaneously, and that vehicles neither enter nor leave the traffic stream while queued upstream of the bottleneck.

Constant Bottleneck Capacity

Consider first the case of a bottleneck with a constant capacity m. Reference (*1*) demonstrates that a relationship exists between a vehicle's time spent in queue upstream of a bottleneck, t_Q, and its delay, w:

$$t_Q = \frac{w}{1 - \frac{v_m}{v_f}} \qquad (1)$$

This relationship can be used to easily modify a standard input-output diagram showing vehicles encountering a bottleneck in order to represent graphically spatial and temporal queue characteristics. Figure 1 shows the desired (or *virtual*) arrival times of individual vehicles at some bottleneck as the curve $V(t)$. (These virtual arrival times can be derived from measurements at an upstream observation point by translating each vehicle's observed arrival time downstream by the free-flow travel time to the bottleneck.) The departure times of vehicles, shown by the curve $D(t)$, indicate that they clear the bottleneck at a maximum rate m when a queue exists. The horizontal separation between $V(t)$ and $D(t)$ represents the delay for a given vehicle number n, and is denoted w_n. (See, e.g., (*4*) for a more detailed explanation of I/O diagram construction).

The time at which each vehicle reaches the back of the queue is then plotted by "extending" the delay of each vehicle, w_n, to the left from the departure curve $D(t)$ by the factor in Equation 1 (See Figure 1). The locus of these points for all vehicles produces the back-of-queue curve, $B(t)$.

The $B(t)$ curve allows one to determine several practical quantities directly from the input-output diagram. The total time spent by all vehicles in queue, T_Q, is represented by the "area" between the $B(t)$ and $D(t)$ curves (just as the total delay to all vehicles, W, is the "area" between the $V(t)$ and $D(t)$ curves). The total distance traveled in queue by all vehicles, D_Q, is therefore the product of this area and v_m. The modified diagram also conveniently displays the (accurate) number of vehicles in queue at any time t as the vertical separation between the $B(t)$ and $D(t)$ curves. The distance spanned by the queue at time t is proportional to the horizontal separation between $B(t)$ and $D(t)$; this length equals the distance that the vehicle m, which joins the queue at time t, travels in queue (= $t_{Q,m} \cdot v_m$). Improved estimates of the maximum queue length, and the time at which this maximum occurs, can also be derived. Furthermore, the simplicity of the approach allows it to be incorporated easily into a spreadsheet (see (*1*)).

Bottleneck Capacity Which Changes Once: Linear $Q(k)$ Relation

Construction of $B(t)$ is also possible in cases where the bottleneck capacity changes once at a known time, given that the q-k relation for queued traffic in the homogeneous highway section upstream of the bottleneck is linear (*1*). Figure 2(a) depicts an example of a such a linear $Q(k)$ curve.

Consider a particular vehicle encountering a queue upstream of some bottleneck. In the constant capacity case, the departure rate at the bottleneck when this vehicle first enters the queue will be the same rate that is in place when it passes the bottleneck. Therefore, the queue is *homogeneous*; each vehicle travels at a constant speed while queued. If the capacity changes, this is no longer the case. Every vehicle in the queue at the moment when the capacity changes joined the queue and traveled in it at speed v_1 corresponding to the initial capacity, yet they will all depart the bottleneck at speed v_2, corresponding to the new capacity. Since these vehicles travel at two different queued speeds, Equation (1) can no longer be applied directly.

Interestingly, reference (*1*) shows that it is still relatively simple to construct $D(t)$ and $B(t)$ in this case, given the linear q-k relation. As explained in that reference, an interface (or shock wave) begins propagating back through the traffic queue at the time when the bottleneck capacity changes. Until this wave reaches the upstream extent of the queue, entering vehicles have no "knowledge" that the capacity has changed, and therefore they join the queue at the same time they would have *had the capacity never changed*. Likewise, vehicles arriving after the wave have no knowledge that the bottleneck ever discharged at the earlier rate, and they join the queue accordingly.

These observations suggest a piecewise approach toward constructing $B(t)$ once a departure curve for the bottleneck has been obtained in the usual manner. Suppose that the capacity changes from m_A to m_C at time t_J, as in Figure 2(b). If the original capacity never changed, the initial segment, $D_A(t)$, of the departure curve would extend through point J at slope m_A, yielding the extrapolated departure curve $D_A'(t)$. This extrapolated curve is then used to estimate each vehicle's virtual delay, w_n', as the horizontal distance between $V(t)$ and $D_A'(t)$. The quantity w_n' represents the delay a vehicle would have incurred had the capacity never changed, which is clearly equal to actual delay only for vehicles departing before the capacity change. By applying the original procedure using the virtual delays and queued speed v_A, $B_A(t)$ is constructed to the right from point P. Next, one constructs the $B(t)$ curve for vehicles that will only encounter state C. Reference (*1*) showed that $B_C(t)$ is constructed to the left from point Q by again applying Equation (1), but instead using the second segment of the departure curve, $D_C(t)$, together with m_C and v_C. Curves $B_A(t)$ and $B_C(t)$ will intersect, usually at a single point such as K on the figure. It should then be intuitive that $B(t)$ is composed of the portion of $B_A(t)$ to the left of K and the portion of $B_C(t)$ to the right of K.

In some instances, $B_A(t)$ and $B_C(t)$ may intersect at multiple points. Since the relevant intersection should occur at the time when the shock wave created by the change in capacity reaches the back of the queue, reference (*1*) suggests identifying the correct intersection by extending a line segment from J with a slope equal to the rate at which queued vehicles pass the wave. Given a triangular q-k relation, any wave created by a change in bottleneck capacity will pass vehicles at a constant rate m_I (*5*):

$$m_I \equiv \frac{q_{\max}}{1 - \frac{q_{\max}}{k_j v_f}} \tag{2}$$

Although locating an interface on an I/O diagram may be less familiar than constructing a shock wave line on a (t,x) diagram, the two construction methods are perfectly consistent. Since the interface segment on the I/O diagram identifies the locus of times when vehicles encounter the shock wave, it should be clear that the intersection of this segment with $B_A(t)$ (and $B_C(t)$) occurs at the time when a vehicle simultaneously encounters the shock and the back of the queue. After this time, arriving vehicles have no knowledge that the bottleneck ever discharged at the earlier rate, and therefore $B_C(t)$ represents the new curve of back-of-queue arrivals.

It should be noted for the degenerate cases, where $m = 0$ or $m = q_{\max}$, that Equation (1) breaks down. As explained in reference (*1*), these special cases can be handled by repeating the derivation of the expression using limits, where $m \to 0$ or $m \to q_{\max}$.

MULTIPLE CHANGES IN CAPACITY: LINEAR *Q*(*k*) RELATION

The approach in (*1*) is now extended to the analysis of bottlenecks whose capacity changes more than once, again assuming a linear *q*-*k* relation in queued traffic. The new method developed here is an iterative procedure in which at each change in capacity, a new back-of-queue curve is combined with a current "working" curve.

In the preceding section, it was shown that a piecewise procedure creates a valid $B(t)$ curve for a bottleneck with a single capacity change. Suppose now that the bottleneck depicted in Figure 2(b) changed capacity yet again, at some time $t_L > t_J$. Such a change would create at time t_L another shock wave propagating back through the queue. Since the *q*-*k* relation in queued traffic is linear, this wave will travel at a speed identical to the first wave, and will pass queued vehicles at the same rate m_I, given by Equation (2). By an earlier argument, it should be clear that the original curve $B(t)$ must continue to be a valid back-of-queue curve until the arrival of this new wave at the back of the queue. Vehicles arriving after the new shock have no knowledge of the earlier bottleneck capacities; therefore, the times when these vehicles join the queue can again be predicted using the departure curve and queued speed corresponding to this final capacity.

These arguments justify the following iterative procedure for construction of $B(t)$. To begin, it is imagined that the first bottleneck capacity *never changes*, and a corresponding extrapolated departure curve and a back-of-queue curve are constructed. This initial back-of-queue curve is labeled $B_1(t) = B_w(t)$, the "working" curve. Suppose that the first change in bottleneck capacity occurs at time t_J. $B_w(t)$ will be valid until the time, t_K ($> t_J$), when the shock wave created by the capacity change arrives at the back of the queue. The intersection of $B_w(t)$ with a line segment extended from point $(t_J, D(t_J))$ with slope m_I correctly identifies this arrival time. Next, as in the original procedure, an extrapolated linear departure curve corresponding to the new capacity is constructed from t_J onward. Using this new departure curve, a new back-of-queue curve segment $B_2(t)$ is created, starting at time t_K; of course, $B_w(t)$ will intersect $B_2(t)$ at time t_K. The union of $B_2(t)$ with the portion of $B_w(t)$ for $t < t_K$ is the new working back-of-queue curve. Iteration of this procedure until each bottleneck capacity has been processed yields a final working curve, which is the desired back-of-queue curve, $B(t)$.

Equivalently, one can construct $B(t)$ by first extending a family of parallel interface lines with slope m_I from the points on the departure curve where the capacity changes, and then using these interfaces to determine the portions of each curve $B_j(t)$ that should be connected to form $B(t)$. [The relevant portion of curve $B_j(t)$ lies between interface lines (j-1, j) and (j, j+1), where line (i, j) corresponds to a change in bottleneck capacity from m_i to m_j].

Example of an improving bottleneck

The proposed method is now applied to an example. Figure 3(a) depicts the $V(t)$ and $D(t)$ curves for a queue upstream of a bottleneck, where the capacity improves from m_A to m_B

at time t_{AB} and from m_B to m_C at t_{BC}. Such a diagram could result from a traffic accident on a busy highway. For example, suppose that a major incident occurs, initially restricting flow severely to rate m_A. After the incident has been partially cleared, rubber-necking may occur, partially restricting flow to rate m_B. Once it has been completely cleared, the roadway returns to its inherent capacity, m_C.

To begin construction of $B(t)$, an extrapolated departure curve $D_A'(t)$ is created using the initial departure rate m_A. Assuming the same q-k relation in Figure 2(a), the appropriate queued speed, v_A, is determined. Figure 3(b) shows the initial back-of-queue curve, $B_A(t)$, created by extending the (virtual) delays of vehicles to the left from $D_A'(t)$ by using Equation (1) in the normal manner. This is the initial working curve, $B_w(t)$.

At t_{AB}, the bottleneck capacity changes from m_A to m_B. From t_{AB} onward, an extrapolated departure curve $D_B'(t)$ is constructed assuming the capacity never again changes, as shown on Figure 3(c). Next, the time when the capacity change shock reaches the back of the queue is determined by extending a line segment with slope m_I from point J until its intersection with $B_w(t)$, point L. Using $D_B'(t)$ and the queued speed, v_B, corresponding to state B, $B_B(t)$ is then plotted to the right from L. The new working estimate of the back-of-queue arrival curve is created by combining the portion of $B_w(t)$ to the left of point L with $B_B(t)$.

The final capacity change occurs at time t_{BC}. The arrival of the shock wave for this capacity change occurs at point M on Figure 3(d). Using $D_C(t)$ and queued speed v_C, we construct $B_C(t)$ from M onward. The new working estimate is the union of $B_C(t)$ and the segment of $B_w(t)$ to the left of point M. Furthermore, since m_C is the final bottleneck capacity, the new working curve $B_w(t)$ is the desired back-of-queue curve, $B(t)$, shown on Figure 3(e).

Interpretation of the modified diagram

In the multiple capacity case, an I/O diagram enhanced by adding $B(t)$ and a set of interfaces allows many important measures to be read directly from the graph, including:

- (maximum) number of vehicles in queue at any time,
- (maximum) length of time any vehicle spends in queue,
- total time spent in queue by all vehicles,
- number of vehicle-miles or vehicle-hours spent by all vehicles in a particular state,
- number of vehicles that undergo a particular change in state, and
- physical extent of the queue at any given time.

The first two measures can be determined as explained for the constant capacity case. As before, the total time spent in queue by all vehicles is the “area” between $B(t)$ and $D(t)$, but this value can be stratified to give the time spent by all vehicles in each of the queued states. The points where a horizontal line corresponding to a generic vehicle *n* intersects the dashed lines *JL* and *KM* in Figure 3(e) demarcate the times that the vehicle spends in each particular queued state. Thus, the total time all vehicles spend in queue in a given traffic state is the "area" contained by $B(t)$, $D(t)$ and any interface line(s) bordering that state. In Figure 3(e), for example, the total time spent in state A is represented by the "area" *PJL*, the total time in state B is "area" *JKML*, and the total time in state C is "area" *KQM*. The total vehicle-miles traveled in any particular queued state is the product of the time in that state and the appropriate queued speed (v_A, v_B, or v_C).

The vertical extent of an interface line provides an estimate of the number of vehicles that undergo a particular change in state. For example, the *vertical* separation between points *J* and *L* represents the number of vehicles which change from state A to state B. This measure, and the above estimates of the total distance traveled in each queued state, are two parameters usually required in the evaluation of the emissions and energy consumption impacts of bottlenecks (see, e.g., (*6*)).

Figure 3(e) also permits a very simple interpretation of the distance that any vehicle travels in queue; i.e. the physical extent of the queue. The horizontal separation between $B(t)$ and $V(t)$ for some vehicle *n* represents the time that it would need to traverse the length of the queue at free-flow speed, v_f. This should be clear, since the separation represents the difference between the time when the vehicle arrives at the back of the queue and the time it would have arrived at the bottleneck traveling at v_f. Thus, the distance that vehicle *n* travels in queue, $d_{Q,n}$, can be read directly from the diagram by rescaling the time axis using v_f, as shown in the figure. Of course, this distance also represents the physical extent of the queue at the time that vehicle *n* reached the back of the queue, t_n. Hence it is possible to determine the physical length of the queue at *any* time, by extending an imaginary line vertically from the desired time on the time axis to $B(t)$, and then reading the horizontal separation between $B(t)$ and $V(t)$ using the (rescaled) distance axis. The maximum rescaled separation between $B(t)$ and $V(t)$ clearly will correspond to the maximum physical extent of the queue.

MULTIPLE CHANGES IN CAPACITY: CONCAVE *Q*(*k*) RELATION

Looking again at Figure 3(e) raises an interesting question: what if segments *JL* and *KM* were not parallel? This situation could arise if the *q*-*k* relation in queued traffic were not linear. In this case, the shock waves resulting from capacity changes could traverse the queue at varying speeds. This complicates the picture, since waves might merge within the queue to create new shocks, and/or diverge from the bottleneck creating new traffic states. Although this wave behavior introduces some added complexity to the procedure, the proposed recipe for constructing back of queue curves remains largely unchanged. This section presents a description and justification of the modified method for handling problems with a piecewise-linear, concave flow-density relation in queued traffic.

When the *q*-*k* relationship for queued traffic is nonlinear, interface speeds and the rates at which queued vehicles pass interfaces depend on the separated traffic states.

Suppose a bottleneck changes capacity from m_A to m_B, creating a wave moving at speed $v_{AB} = \frac{m_B - m_A}{k_B - k_A}$. The rate at which vehicles pass this wave, m_{AB}, can be calculated using the simple formula for flow past a moving observer (see, e.g., Equation 4.32 and Figure 4.10 of reference (*7*)):

$$m_{AB} = m_A - k_A v_{AB} \equiv m_B - k_B v_{AB} \quad (3)$$

Equation (3) should be used to determine the slope of interface line segments drawn on the I/O diagram. Furthermore, since the segments for different capacity changes may not be parallel, a line segment drawn at the appropriate slope may intersect a segment for an earlier shock wave before reaching the back of the queue. In this case, the waves *merge* to create a new wave. For example, if a capacity change from m_A to m_B creates a slow-moving interface, and then another change from m_B to m_C creates a faster wave, the second interface may overtake the first. At the point when the shocks intersect, traffic state B dissipates and a new shock wave separating states A and C is formed. Vehicles arriving at the back of the queue never encounter queued traffic state B upon joining, and therefore $B(t)$ will have no portion corresponding to m_B.

[We note as an aside that the kinematic wave theory predicts that any shock wave created by the dissipation of a state similar to B is *stable* if the *q*-*k* relation is concave; i.e. any such shock will not change in character and will continue to separate states A and C as it moves through the queue. In other cases, the dissipation of a state can trigger the appearance of other states and the method described below does not hold. This is why we require $Q(k)$ to be concave.]

Handling merging shock waves when the *q*-*k* relation is concave requires a modification to the original construction recipe. The key issue is to determine the set of queued traffic states that *actually influence* $B(t)$; these states are those that do not dissipate before reaching the back of the queue. To do so, first plot the true departure curve, $D(t)$, for the bottleneck. Next, construct the full set of interface segments emanating from the capacity change points on the departure curve, using Equation (3) to determine the slope of each interface from the two traffic states it separates. When two interface segments intersect, indicating the dissipation of the intervening traffic state, it is necessary to create a new wave at the intersection with a slope determined by the newly adjoining traffic states. The end result will be either a tree of line segments, or a disjoint set of trees (a *forest*), with the leaves intersecting the departure curve. The forest of interface segments partitions the portion of the (t,n) plane above the departure curve into regions of constant (queued) traffic state that describe the constitution of the queue. In other words, these regions demarcate the time periods during which each departing vehicle (horizontal line with ordinate n) would experience a particular traffic state, assuming that the vehicle is in the queue.

Figure 4 depicts the solution of a hypothetical problem, including the (truncated) trees and the $B(t)$ curve. Note from the geometry of our forest that immediately after the $B(t)$ curve meets an interface (as occurs for vehicle number n in Figure 4): (i) the state on the future side of the interface survives, and (ii) all of the prior traffic states prior have dissipated. Thus, the next few vehicles following n will meet this surviving state and join the queue as if the bottleneck served at this rate in perpetuity; e.g. in the figure, state "3" survives and curve $D_3'(t)$ would be used to construct the back-of-queue curve after vehicle n. Generalizing, if the j^{th} arc of the $B(t)$ curve, $B_j(t)$, corresponds to the s_j^{th} state of the $D(t)$ curve, this arc can be drawn by applying Equation (1) to the linear extension of the s_j^{th} segment of $D(t)$; see Figure 4. This suggests the following iterative procedure for constructing $B(t)$.

Assume that the bottleneck capacities have been numbered consecutively, "1, 2, 3, ..." and denote by s_j the j^{th} traffic state encountered at the back of the queue. If all states propagate to the back of the queue as occurs in the linear case, then s_j will always equal j, otherwise $s_j \geq j$. The proposed procedure is initialized as follows: j = 1, $s_1 = 1$, the working curve $B_w(t) = B_1(t)$ corresponding to the initial capacity, and P_1 as the initial intersection between $B_1(t)$ and $D(t)$; see Figure 4. The j^{th} step of the procedure is as follows:

(i) Find the point to the right of P_j where $B_w(t)$ first intersects either: (a) a branch in the forest of interfaces, or (b) the departure curve $D(t)$. In case (a), call the intersection P_{j+1} and let s_{j+1} be the state on the future side of this point. Otherwise, the intersection between $B_w(t)$ and $D(t)$ indicates that the queue has vanished, so the procedure terminates with $B(t) = B_w(t)$.

(ii) Eliminate the part of $B_w(t)$ that extends to the right of P_{j+1}. Construct $B_{j+1}(t)$ to the right of P_{j+1} (using the linear extension of $D_{s_{j+1}}(t)$) and let the union of this curve and the remaining part of $B_w(t)$ be the new $B_w(t)$. Increment j by 1 and repeat from (i).

If desired, this procedure can be streamlined to avoid building those parts of the interface forest that are not meaningful in the final solution, i.e. the portion above $B(t)$. For example, one could take advantage of the fact that the past cannot be influenced by the future and build time slices of the interface forest and the $B(t)$ curve jointly, step after step. However, because a description of these details is not particularly insightful, it is not provided here.

Stability in concave *q*-*k* problems

Although the proposed method for constructing back-of-queue curves on I/O diagrams described in the previous sections is clearly consistent with the first two postulates of the kinematic wave theory outlined in the introduction, the stability of the solution has not been verified yet.

For problems with a triangular *q*-*k* relationship, reference (*1*) showed that the solution generated for a bottleneck with a single change in capacity was the physically meaningful kinematic wave solution. Therefore, the construction of Figure 2(b) is "stable." Similar logic reveals that the procedure used to construct Figure 3 yields a stable solution. This is easy to check because the triangular *q*-*k* relationship ensures a constant wave speed between queued states. Reconsider the problem in Figure 3. If the final bottleneck capacity, m_C, were perturbed infinitesimally, the speed of the wave created by the capacity change would be unaffected and only curve $B_C(t)$ would change. This would occur from point *M* onward, and the displacements of curve $B_C(t)$ would be infinitesimal. Consideration shows that *any* infinitesimal change in the bottleneck capacity (or any other infinitesimal perturbation in $D(t)$ or $V(t)$) only results in infinitesimal changes to $B(t)$; i.e., the solution is stable.

Let us now examine the concave case. Figure 5 illustrates that in this case, a small perturbation in bottleneck capacity could lead to drastic changes in the solution as time progresses. Therefore, a blind application of the previous procedure to problems with nonlinear *q*-*k* relations could lead to unstable solutions. Suppose, for example, that this has been done for a highway with the $Q(k)$ curve of Figure 5(a), and the data ($V(t)$ and $D(t)$ curves) of Figure 5(b); i.e. we assume that the original procedure is used to obtain $B(t)$ when the bottleneck capacity changes from m_A to m_C. The resulting $B(t)$ curve would be composed of two segments, one constructed using m_A and the other using m_C. Not shown in the figure, this curve would be the result of extrapolating arcs $B_A(t)$ and $B_C(t)$ of Figure 5(b) toward the center until the shock wave arrival.

If, however, the input data were perturbed ever so slightly, such that the bottleneck capacity changed to an intermediate level (e.g., m_B) for an infinitesimal period of time (instead of moving directly from m_A to m_C), it turns out that a significantly different solution results. This is shown in Figure 5(b), which is the result of applying the proposed construction procedure to the perturbed data. Note that an arc $B_{AC}(t)$ corresponding to state B, and substantially different from the extrapolated $B_A(t)$ and $B_C(t)$ curves, would arise as part of $B(t)$. This is logical since the interface between A and B would travel faster than the one between B and C (see Figure 5(a)). Clearly then, the number of vehicles queued in state B would increase with time, and many vehicles would encounter state B when they joined the back of the queue. Since this solution is significantly different from that of the original (i.e., unperturbed) problem, the original solution is not stable.

When the $Q(k)$ curve is concave, this problem can be avoided by assuming that any capacity changes take place instantaneously, but *smoothly;* e.g., that the change from m_A to m_C was gradual, but infinitely fast. This is easy to do for piecewise-linear, concave relations like the one in the figure because the changes along the straight segments of the *q*-*k* curve do not induce changes in wave speed. Thus, each capacity change can be treated as a series of consecutive but infinitely fast jumps through all the intermediate breakpoints of the *q*-*k* relation. In the example of Figure 5(a), a capacity change from m_A

to m_C would require an instantaneous step through breakpoint state B, i.e., m_B, before the procedure presented in the previous subsection can be applied. The reader can verify that the resulting solution is indeed stable by introducing different infinitesimal perturbations in this example.

Although the procedure proposed in this paper could, in principle, be extended to non-concave $Q(k)$ relations, this is not done here. In this case the "stable" interfaces passing through the queue do not always form a forest, and this complicates the way in which the $B_j(t)$ curves have to be constructed.

CONCLUSIONS

This paper has presented a generalized method of enhancing a standard input-output (I/O) diagram to represent graphically the time and distance that vehicles spend in a queue upstream of a bottleneck, by adding a "back-of-queue" curve to the diagram. The ideas of a previous paper (*1*) are extended by allowing multiple changes to the bottleneck capacity, and then by relaxing the original assumption of a linear flow-density relationship in queued traffic to one that is piecewise-linear concave.

The proposed approach for constructing back-of-queue curves on an I/O diagram is consistent with the kinematic wave theory of traffic flow, but is simpler to apply to complex problems because it does not require the laborious construction of the time-space diagram. The approach allows the estimation of several practical measures, including the (accurate) number of vehicles in queue at any time, the time that individual and aggregate vehicles spend in queue, and the distance that individual and aggregate vehicles travel in queue. The total time that all vehicles spend in queue can be further stratified to give the total time in each queued state, which permits the evaluation of the energy and emissions implications of bottlenecks. A much simplified way of determining distances traveled in queue, and thus physical queue lengths, allows one to assess the impacts of queue spillback on the surrounding road network (e.g., when freeway traffic queues block an upstream off-ramp).

REFERENCES

1. Lawson, T.W., D.J. Lovell and C.F. Daganzo. (1997) Using the Input-Output Diagram to Determine the Spatial and Temporal Extents of a Queue Upstream of a Bottleneck. In *Transportation Research Record 1572*, TRB, National Research Council, Washington, D.C.

2. Lighthill, M.J. and G.B. Whitham. (1955) On Kinematic Waves II, A Theory of Traffic Flow on Long Crowded Roads. *Proceedings of the Royal Society*, Vol. A 229, pp. 317-345.

3. Richards, P.I. (1956) Shock Waves on the Highway. *Operations Research*. Vol. 4, pp. 42-51.

4. Newell, G.F. (1982) *Applications of Queueing Theory*. Second Edition. Chapman and Hall, New York.

5. Newell, G.F. (1993) A Simplified Theory of Kinematic Waves in Highway Traffic. *Transportation Research*, Vol. 27B, No. 4, pp. 281-313.

6. Dale, C.W. (1980) *Procedure for Estimating Highway User Costs, Fuel Consumption and Air Pollution*. Office of Traffic Operations, Federal Highway Administration, Washington, D.C.

7. Daganzo, C.F. (1997) Fundamentals of Transportation and Traffic Operations. Pergamon, Oxford, U.K.

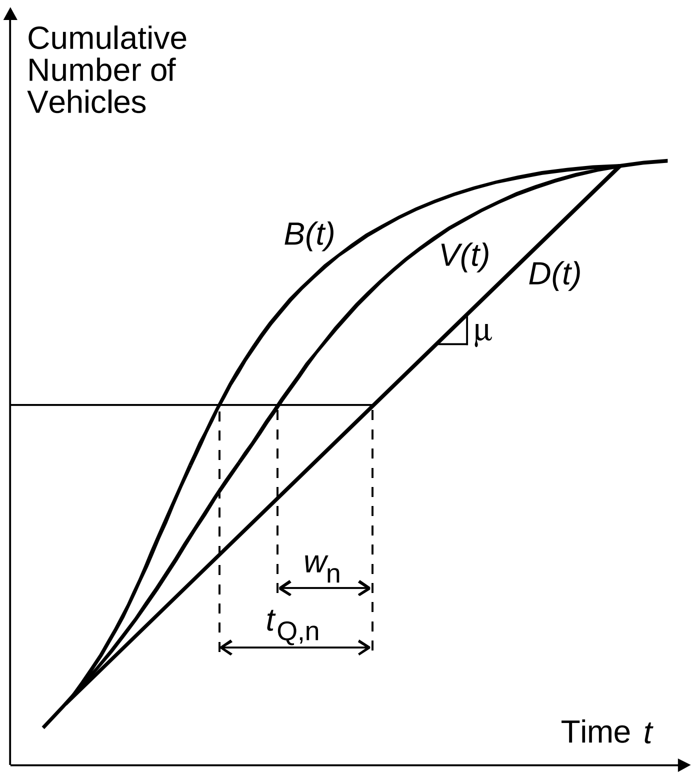


FIGURE 1 Locating Back-of-Queue Curve, B(t), for Constant Bottleneck Capacity

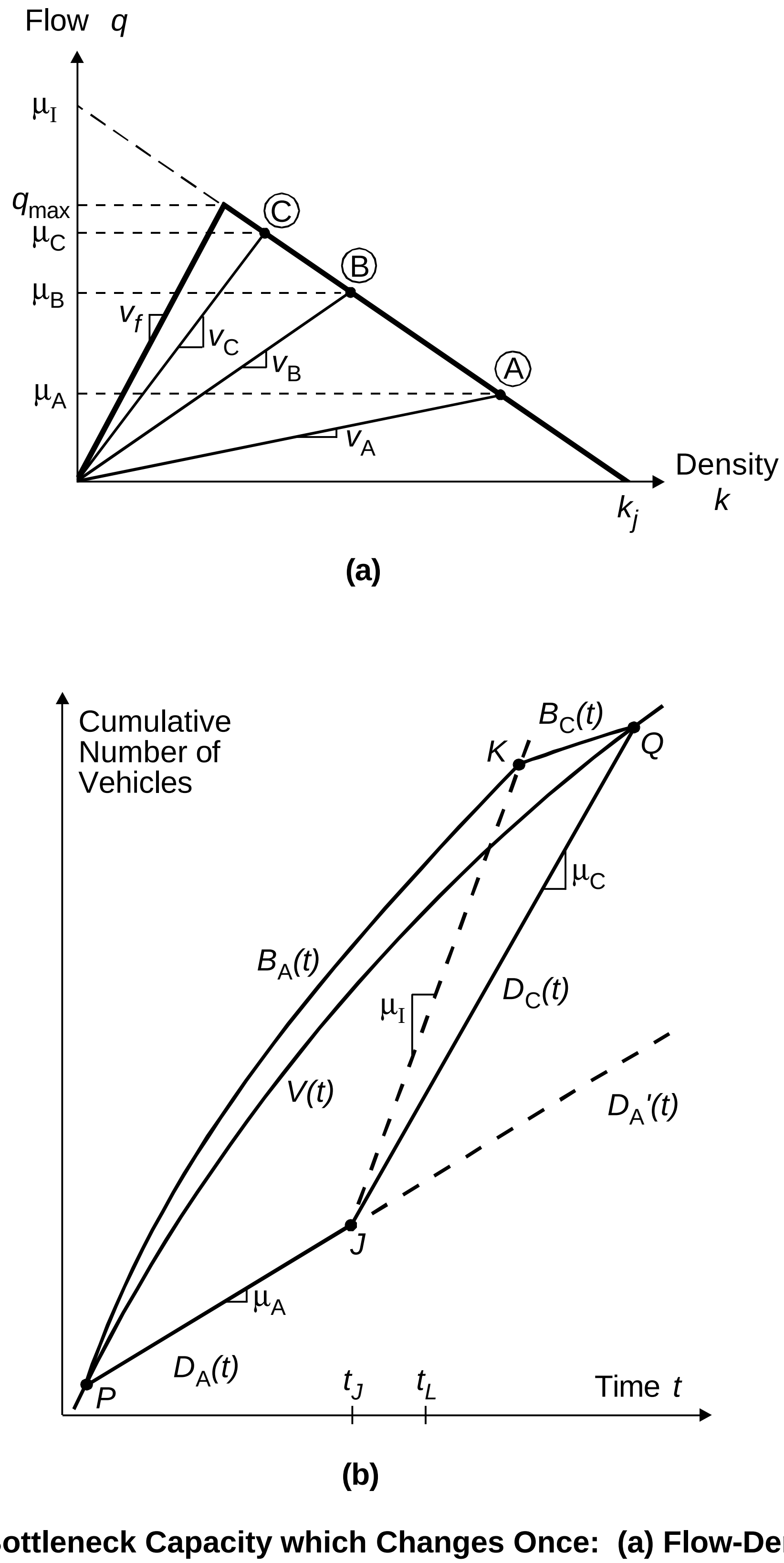


FIGURE 2 Bottleneck Capacity which Changes Once: (a) Flow-Density Relation, (b) Construction of B(t)

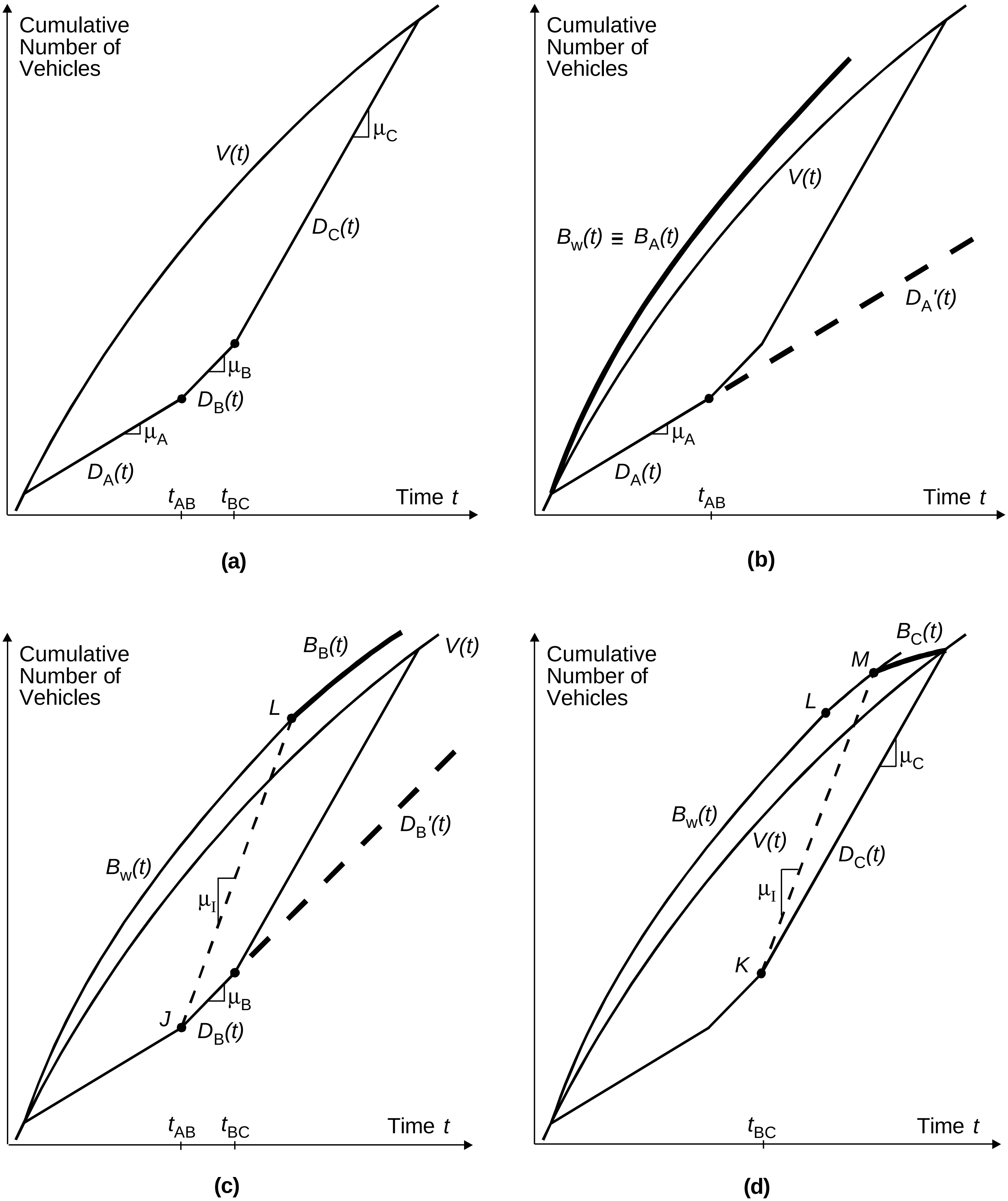
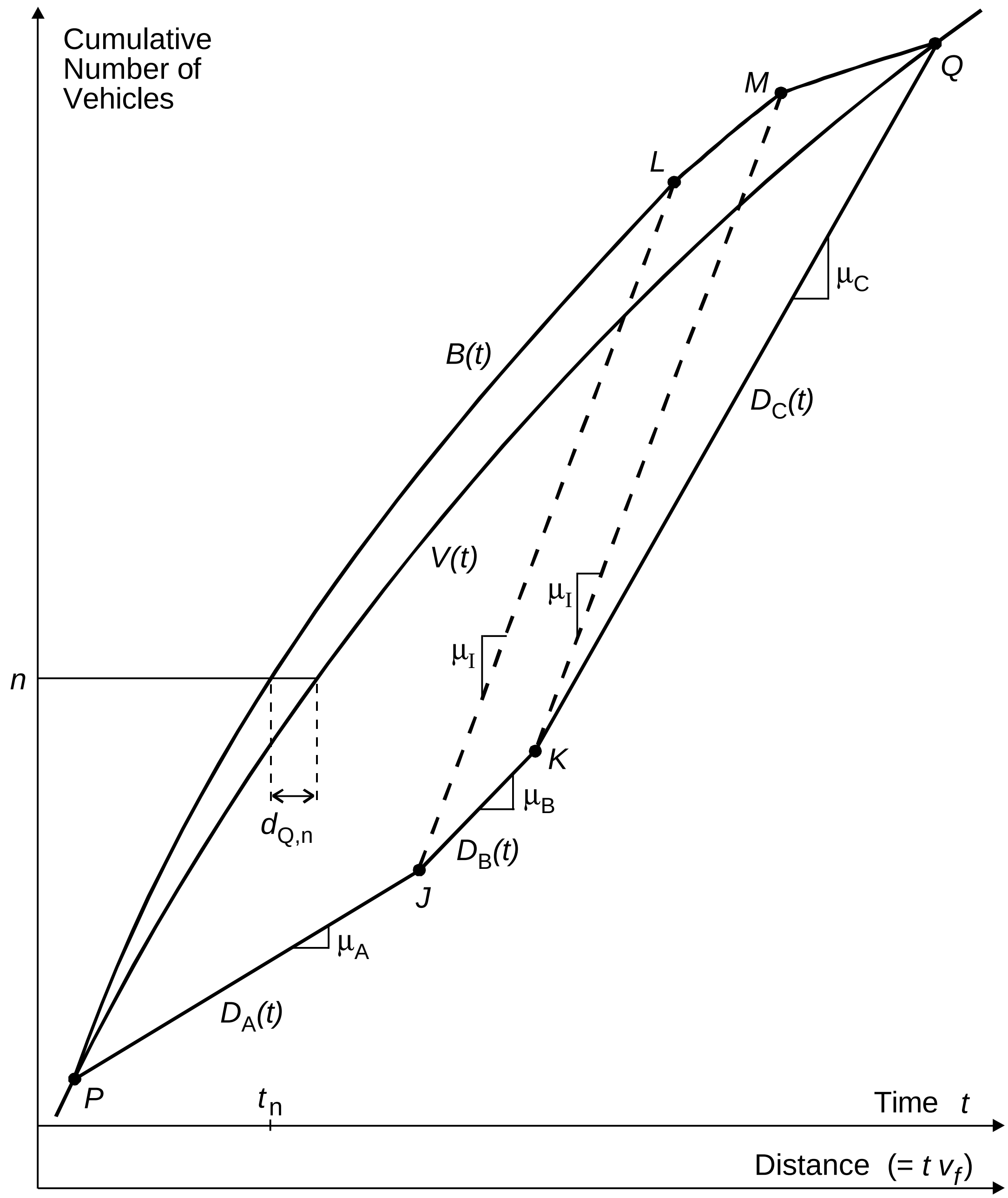


FIGURE 3 Multiple Changes in Bottleneck Capacity: (a) V(t) and D(t), (b) Capacity A, (c) Capacity B, (d) Capacity C



(e)

FIGURE 3 Multiple Changes in Bottleneck Capacity: (e) Completed Diagram

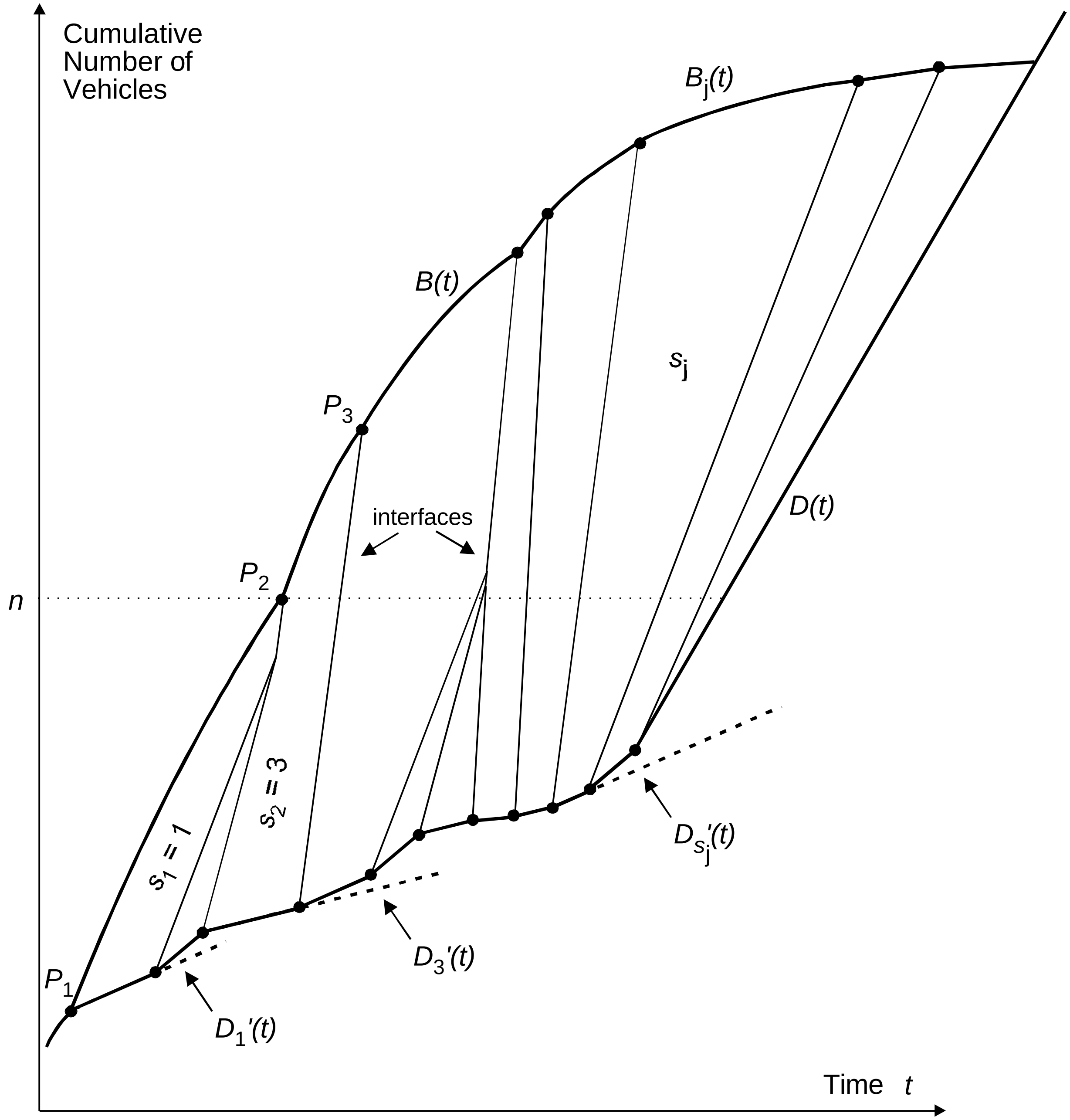


FIGURE 4 Hypothetical Solution to a Problem with Concave Q(k) Relation

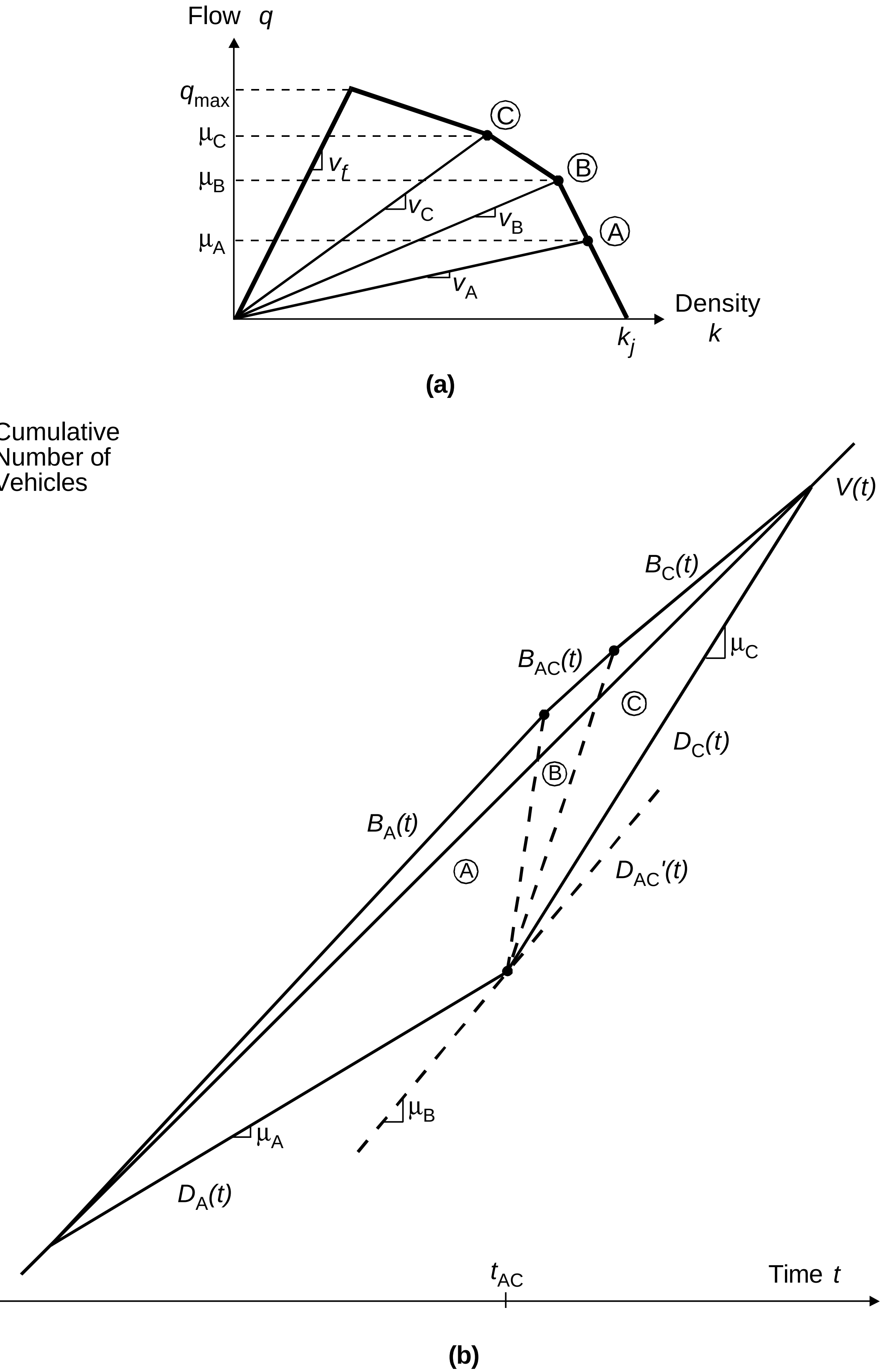


FIGURE 5 Example of Capacity Perturbation: (a) Concave Q(k) Relation, (b) Construction of B(t)