# Bottleneck

Kentaro Wada
University of Tsukuba
1-1-1 Tennodai, Tsukuba, Ibaraki 305-8573, Japan
wadaken@sk.tsukuba.ac.jp
+81-29-853-3614

Toru Seo
The University of Tokyo
7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan
seo@civil.t.u-tokyo.ac.jp
+81 3 5841 8090

Yasuhiro Shiomi
Ritsumeikan University
1-1-1 Nojihigashi, Kusatsu, Shiga 525-8577, Japan
shiomi@fc.ritsumei.ac.jp
+81-77-561-5094

**Abstract**

A bottleneck is a road segment whose traffic capacity is lower than that of its vicinity. Since traffic congestion occurs when an arrival traffic flow rate to the bottleneck exceeds its capacity, understanding and description of bottleneck behaviors are essential for establishing congestion mitigation strategies. This article contains discussions on various types of freeway bottlenecks, such as sag/tunnel, merge, diverge, and weaves. Specifically, the bottleneck activation mechanisms are discussed for each type, mainly from a macroscopic perspective. Furthermore, the current practice of traffic congestion management and possible advancing of it in the era of connected and automated vehicles are discussed.

**Keywords**

Freeway bottlenecks, Traffic congestion, Traffic control, Traffic management, Sag, Tunnel, Merge, Diverge, Weave

Chapter for *International Encyclopedia of Transportation*, Vickerman, R., Noland, R.B., Ettema, D. (Ed.), Elsevier Science.

August 17, 2020

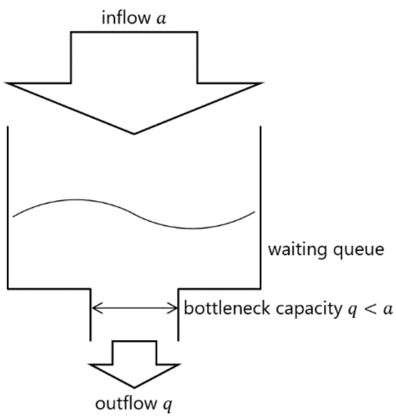


Fig.1 Bottleneck and waiting queue

## 1. Introduction

### 1.1 Definition of bottleneck

A bottleneck is defined as a road segment whose traffic capacity is lower than that of the road segments in its vicinity. The definition of traffic capacity of a road segment is the maximum traffic flow rate that can reasonably pass through the segment. Some evident examples of bottlenecks are lane drops, lanes blocked by car accidents, and signalized intersections with small green time ratios. Detailed discussions on types and mechanisms of bottlenecks are be described in the following sections.

The Highway Capacity Manual (HCM) (Transportation Research Board, 2016) provides a detailed discussion on traffic capacity and the concept of a bottleneck from practical perspectives. According to the HCM, traffic capacity is defined as “the maximum number of vehicles that can pass a given point during a specified period under prevailing roadway, traffic, and control conditions. This assumes that there is no influence from downstream traffic operation”, and a bottleneck is defined as “locations where the capacity provided is insufficient to meet the demand over a given period of time ” where “demand” refers to the flow of vehicles arriving at the bottleneck. The typical capacity of a basic highway segment is between 2250 to 2400 veh/h/lane, and that of a basic arterial segment with a signalized intersection is between 790 to 1110 veh/h/lane. These values vary among locations and conditions. Consequently, a location with a low capacity becomes a bottleneck.

Bottlenecks play an essential role in the description and understanding of traffic behavior, especially in uninterrupted flows (e.g., freeways, expressways). To be more precise, almost all of traffic congestions are caused by bottlenecks. Consider a bottleneck in an uninterrupted traffic. If an arrival flow rate (i.e., flow from the upstream section) to the bottleneck is larger than its capacity, then only a flow rate equal to the capacity passes through the bottleneck, and the remaining traffic forms a waiting queue at the bottleneck. A bottleneck is said to be active when it causes a waiting queue without queue extension from the downstream section.

See Fig. 1 for an intuitive illustration of a bottleneck. It shows an upturned bottle. The width of each cross-section signifies the capacity of the corresponding section. If an inflow is sufficiently large, the outflow will be limited by the neck. Futher, the medium (i.e., the vehicles) that cannot pass though the neck will be pooled (i.e., form an waiting queue) upstream from the neck.

### 1.2 Macroscopic mechanism of congestion due to a bottleneck

Typical congestion phenomena in uninterrupted traffic with a bottleneck can be explained as

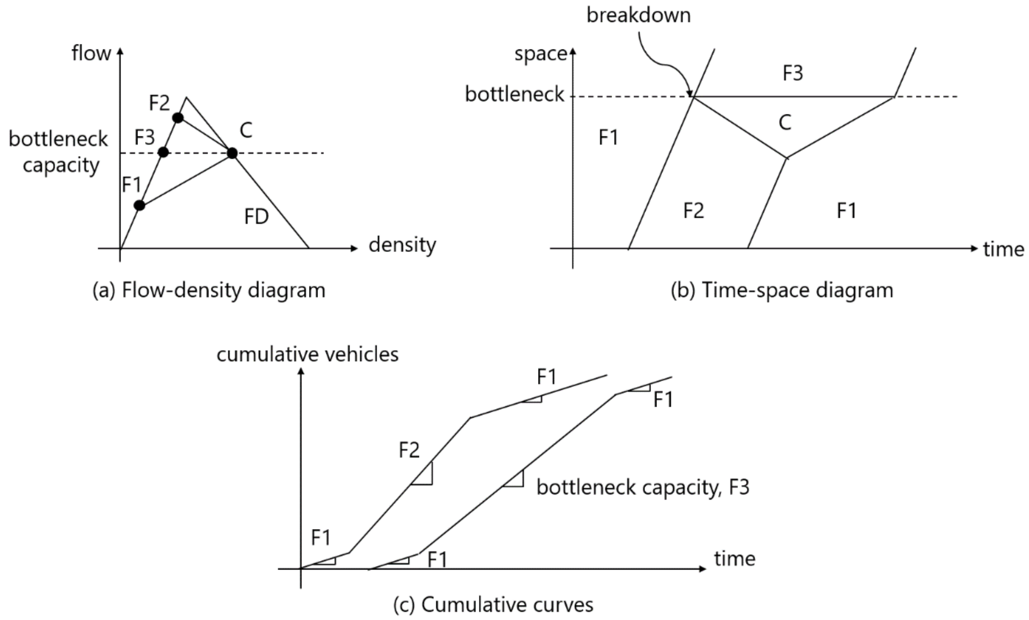


Fig.2 Typical congestion phenomena in uninterrupted traffic with a bottleneck

follows. Consider a freeway road section with a bottleneck. The inflow to the section can take two possible values: a flow less than the capacity (denoted as F1) and a flow greater than the capacity (denoted as F2). Congestion phenomena with time-varying inflow can be illustrated as in Fig. 2, based on the kinematic wave theory (Lighthill and Whitham, 1955; Richards, 1956; Newell, 1993), which is the most standard and simplest traffic flow model. Fig. 2(a) shows the relationship between flow and density, where the solid curve indicates the fundamental diagram (Greenshields, 1935) of the section except for the bottleneck, the dashed line indicates the bottleneck capacity, and the black dots indicate the observed traffic states. Fig. 2(b) shows a time-space diagram of traffic, where the horizontal axis indicates time (and towards the right indicates the future), the vertical axis indicates the space (and up indicates the downstream direction of traffic), and solid lines indicates boundaries between different traffic states. Regions F1, F2, F3 are free-flowing and region C is congested. Fig. 2(c) shows the cumulative curves of traffic, where the upper curve indicates the cumulative vehicle number counted at the upstream end of the section (referred to as the arrival curve) and the lower curve indicates the cumulative vehicle number counted at the bottleneck location (referred to as the departure curve).

As shown in Fig. 2(b), traffic at the bottleneck breakdowns (see the chapter "Flow Breakdown" for the details) when the arrival flow exceeds the capacity. As a result, a waiting queue ("C" area in Fig. 2(b)) is formed, and it keeps growing as long as the arrival flow exceeds the capacity. As soon as the arrival flow becomes less than the capacity, the queue starts to diminish. Thus, the severity of the congestion is determined by the bottleneck capacity and the demand (or arrival flow) profile.

The same phenomenon can be illustrated in a different way by cumulative curves (Fig. 2(c)). Note that the slope of a cumulative curve is identical to flow, and the vertical distance between arrival and departure curves denotes the number of vehicles between the two locations. Thus, if the slope of the arrival curve exceeds the bottleneck capacity, the vertical distance between two curves increases, signifying the development of a queue, and so on.

In the flow–density diagram (Fig. 2(a)), an arrival flow that is less than the capacity is represented as a dot labelled F1, and an arrival flow that is larger than the capacity is represented by

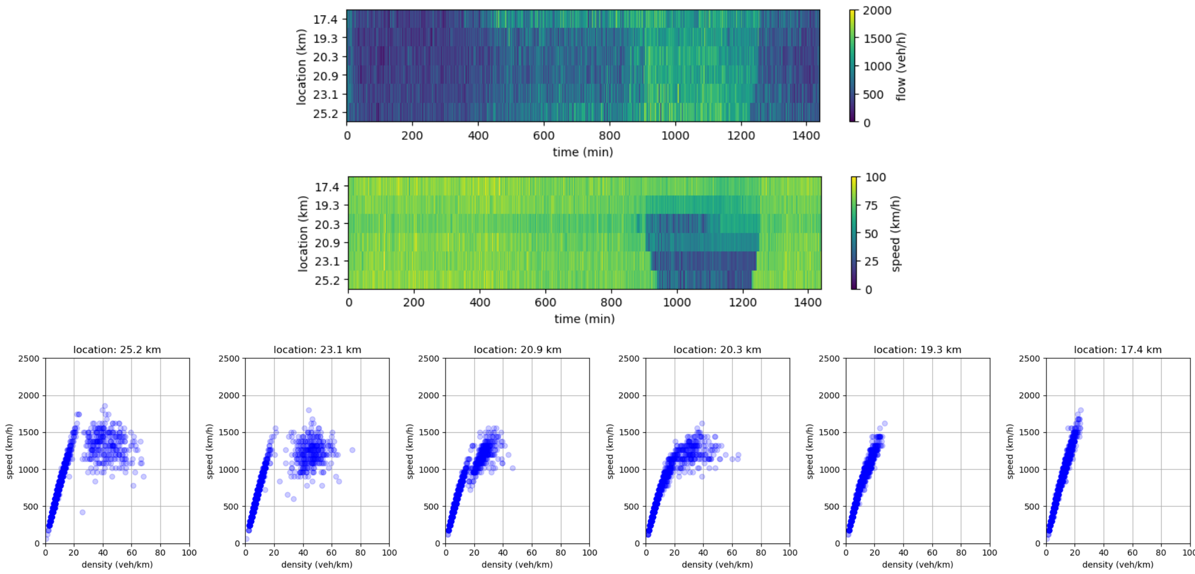


Fig.3 Time–space diagrams and flow–density diagrams at Chugoku Expressway (data from Shiomi et al., 2015).

dot F2, and the departure flow when the bottleneck is active is represented by dot F3, and the traffic state in the waiting queue is represented as C. When traffic data is collected by a traffic detector, the observed traffic state will be completely different depending on the location of the detector. If the detector is placed at upstream next to the bottleneck, states F1 and C will be observed. If the detector is placed far upstream of the bottleneck, states F1, F2, and occasionally C will be observed. If the detector is placed at downstream of the bottleneck, states F1 and F3 will be observed.

By leveraging this theory, we can detect a bottleneck from detector data. Suppose that the detector on a basic road segment observed (almost) free-flowing traffic only, and its upstream detector observed both free-flowing and congested traffic. Then, we can presume that a bottleneck exists between these two detectors and can roughly estimate its capacity as the upper bound of the downstream detector measurement.

Fig. 3 shows the actual traffic detector data collected along a road section near Takarazuka-West tunnel, Chugoku Expressway, an intercity highway in Japan. The area has no major merge or diverge sections. It is clear from the data that a waiting queue formed between 900 min and 1280 min due to a bottleneck. According to the flow–density diagrams, the location of the bottleneck appears to exist between 20.3 km and 19.3 km locations, and its capacity is approximately 1500 veh/km/lane.

## 2. Sag and tunnel bottlenecks

### 2.1 What is sag?

As noted in the first edition of the HCM (Highway Research Board, 1965), it has been known since the dawn of the traffic flow theory that the terrain type and grade can affect the traffic capacity. The main concerns in the HCM were the speed reduction of large vehicles and other adjustment factors that are caused by the grades. Since the early 1980s, it has been found that some of the sags and tunnels[1] that are the vertical alignment of a downgrade section followed by an upgrade section

[1] Although a sag and a tunnel can each be a bottleneck respectively and independently, sags and tunnels are located close together in many cases. It is therefore difficult to separate the effects of a sag and a tunnel, so we do not distinguish them explicitly in this manuscript.

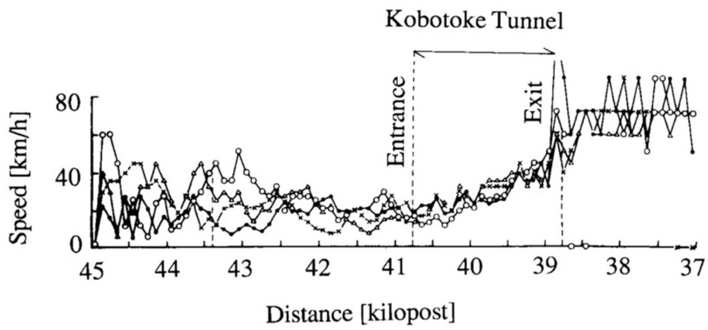


Fig. 4 Speed profile through the congestion at Kobotoke Tunnel (adapted from Koshi et al, 1992) Note: The dashed bertical lines on the distance-axis denote entrance of the tunnel at 40.7 kp and the exit at 38.7 kp. It takes a distance of 2 km to increase the speed from 20 km/h to 60 km/h.

can become active as bottlenecks on freeways (Koshi et al., 1992). The traffic congestion caused by sags and tunnels is common in hilly regions and subaqueous tunnels (e.g. Lincoln Tunnel and Holland Tunnel, see Edie and Foote, 1958). It is claimed that the capacity at a sag is up to 25 % lower than the expected value for the uninterrupted flow on a freeway (Koshi, et al., 1992). In Japan, almost 60 % of traffic congestion on freeways occurs at sags or tunnels (Xing et al., 2014), and considerable academic and practical efforts have been made that attempt to reveal the congestion mechanism and find the effective countermeasures. The next section describes the features of traffic dynamics at sags and tunnels, and its modeling approach. Practical control measures that can be taken to mitigate traffic congestion can be found in Section 4.

**2.2 Features of traffic dynamics at sags and tunnels**

A possible explanation of the bottleneck mechanism at sags and tunnels is as follow (Koshi, 1986): i) As traffic volume increases, the utilization ratio of the inner lane increases; ii) Some vehicles on the inner lane slightly decrease the speed on the upgrade section of sags or at the entrance of the tunnel; iii) Vehicles with a higher desired speeds catch up with slower vehicles and are forced to follow them; iv) The number of vehicles following behind the moving bottleneck increases, generating a platoon in which traffic density is locally high; v) Once a disturbance within a platoon occurs at a bottleneck section for some reason, the deceleration wave propagates upstream, and a breakdown in traffic flow occurs. As is shown in the Movie 1 (https://youtu.be/9tqF9O87GaI), which is the driver's view when passing through a congestion due to a sag, drivers cannot find any apparent causation for the congestion queue such as merging, diverging, weaving, lane drops, and work zones. In other words, it is quite unclear for drivers where the head of the congestion queue (i.e., bottleneck) begins. For this reason and due to an increase in gradient, drivers may be late in starting their acceleration and result in insufficient and bounded acceleration operation (see Fig. 4), which may cause a considerable capacity drop after the traffic breakdown (for more details, see the article "Flow Breakdown").

**2.3 Modeling approach**

Several models have been proposed to represent the congestion that occurs at sags from both microscopic and macroscopic perspectives. In the microscopic approach, a number of car-following models have been developed that consider the gradient of the road and the effect of gravity. For example, in order to discount the acceleration for General Motor (GM) based car-following model,

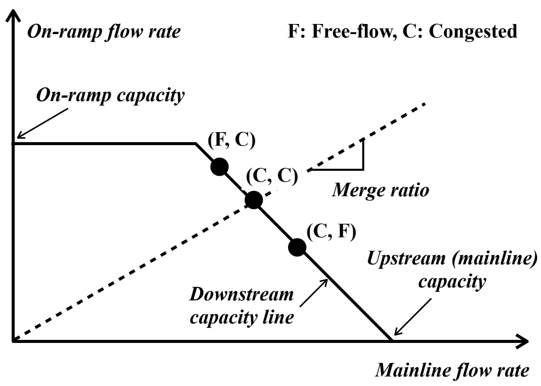


Fig.5 Merge diagram (reproduced from Daganzo, 1995) Note: Three black dots correspond to three possible states when the merge bottleneck is active. Each dot expresses a pair of mainline and on-ramp states, (mainline state, on-ramp state).

Koshi et al. (1992) considered the gradient term ($-\gamma \sin\theta$), in which $\gamma$ is a sensitivity parameter and $\theta$ is gradient difference at sag. Goñi Ros et al. (2016) considered the additive acceleration term in Intelligent Driver Model (IDM) that depicted a driver's compensation behavior and in particular regarding his operation of the accelerator pedal in response to a gradient change at the sag.

The macroscopic approach is rather limited when compared to the microscopic approach. Jin (2018) developed a behavioral kinematic wave (first-order) model taking the location-dependent time gaps and bounded acceleration into consideration to represent capacity reduction, capacity drop, and extremely low acceleration rates. Wada et al. (2020) extended Jin's model to a second-order one to represent the temporal transition process and the formation mechanism of the capacity drop.

## 3. Freeway network bottlenecks

This section contains discussions on various types of freeway network bottlenecks such as "merge," "diverge," and "weave". The activation mechanisms using a macroscopic level modeling approach are discussed for each type. Please refer to the chapters such as "signalised intersections" and "signalised roundabouts" for urban network bottlenecks.

### 3.1 Merge bottleneck

A merge bottleneck becomes active if the sum of traffic demands from upstream approaches (the total demand) exceeds the downstream traffic capacity. At an active merge bottleneck, a capacity drop likely occurs, and whether all or a subset of upstream approaches are congested depends on how the conflicting traffic streams share the downstream capacity. Such competing traffic flow behavior (under a constant downstream capacity assumption) can be explained by the following simple macroscopic theory (Daganzo, 1995), see Fig 5.

Suppose a merging section has two upstream approaches (i.e., a mainline and an on-ramp), and the total demand exceeds the downstream capacity. In Fig.5, a traffic state is expressed by a pair of mainline and on-ramp traffic flow rates. The solid lines indicate traffic states in which the traffic flow rate of both or either of the approaches equals to the capacity; if an initial state is located outside of these lines (i.e., demand exceeds capacity), the traffic state converges to a point on the lines as the initial state cannot be lasting. If both approaches have queues ((C, C) dot in Fig. 5), the traffic streams will merge according to some definite priority, i.e., the traffic volume of each approach shares a

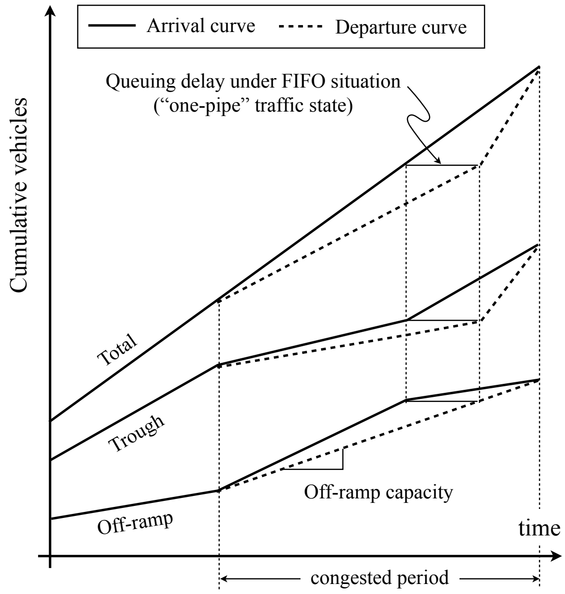

Fig. 6 Queue formation upstream of a diverge due to a fluctuation in traffic composition

constant percentage of the downstream traffic capacity. The ratio of the percentages of two approaches is called a merge ratio. Moreover, if the demand ratio is disproportionate to a fixed merge ratio, one of the approaches can be non-congested (i.e., the demand can be less than its capacity share). In this case, the other (congested) approach utilizes the remaining capacity ((F, C) and (C, F) dots in Fig. 5).

Is the merge ratio actually fixed, for example, irrespective of the severity of congestion? What is the main factor influencing the merge ratio? Cassidy and Ahn (2005) addressed the first question and they have observed that, at four merge sites in California, US and Toronto, Canada, upstream traffic streams enter a congested merge in some (nearly) fixed ratio. It suggests that the merge ratio mainly depends on merge geometry. Further, Bar-Gera and Ahn (2010) investigated 15 different merge sites in California and found that the merge ratio at each site can be reasonably estimated by its lane ratio (i.e., the ratio of the numbers of lanes of two approaches), which is similar to the capacity ratio proposed and (partially) validated by Ni and Leonard (2005).

**3.2 Diverge bottleneck**

A diverge bottleneck becomes active if the traffic demand to either one of the downstream branches exceeds its traffic capacity. If all lanes of the upstream approach are queued (called a "one-pipe" traffic state) then the traffic that advances into not only the demand-exceeding branch, but additionally the others, experiences the delay equally due to the first-in-first-out (FIFO) discipline. Under FIFO, the specific arrangement and order (or composition) of vehicles with different destinations (i.e., different downstream branches) arriving at the back of the queue must be kept until the point of leaving the queue; thus, the discharge flows of non-congested branches can change significantly without a change in that of the congested one (Daganzo, 1995). The composition change can also cause a sudden traffic breakdown, even if traffic upstream of a diverge (i.e., the total demand) is steady, see Fig. 6.

While the above FIFO situation more likely occurs at an active diverge bottleneck with a narrow upstream approach, Muñoz and Daganzo (2002) showed that it can happen even at a wider one with a 5-lane upstream approach. The reference further showed that non-FIFO situations

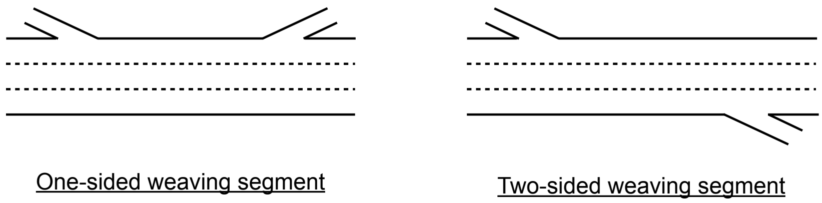


Fig.7 Examples of weaving sections

(calledthe “multi-pipe” traffic states) could persist for a long time. Congested multi-pipe traffic states, where queued lanes move at different speeds, possibly happen because drivers prefer different lanes depending on their destinations. Semi-congested traffic states, where some lanes are queued and others are not, also exist.

For modeling traffic behavior at a diverge, the simplified FIFO logic above (Daganzo, 1995) would be useful as a first approximation or for a large-scale network analysis. For a more detailed analysis, Daganzo (1997) and Daganzo et al. (1997) proposed a macroscopic multilane-multiclass theory that can approximate both FIFO and non-FIFO situations.

**3.3 Weaving bottleneck**

Weaving is defined as the crossing of two or more traffic streams traveling in the same general direction. Weaving sections are formed when a merge area is closely followed by a diverge area, or when an on-ramp is closely followed by an off-ramp and the two are joined by an auxiliary lane (Transportation Research Board, 2016). At weaving sections, a lot of lane changes can occur, and their intensity and spatial distribution depend on the weaving demands and their geometrical features of weaving sections such as configuration, length, and width (the number of lanes) (see, Fig. 7 for examples). Several empirical studies showed that the lane changes (of the weaving traffic) are more likely to be concentrated close to the merge (e.g., Cassidy and May, 1991). Such a concentration makes the lane changes more disruptive and may trigger activations of weave bottlenecks.

The impact of systematic lane changes on the capacity (or on the fundamental diagram) can be captured at a macroscopic level, as follows (Jin, 2010). For the duration of a lane change, a lane-changing vehicle occupies the current and target lanes. Therefore, the contribution of lane-changing vehicles to the total density should be doubled. In other words, lane-changing traffic effectively causes additional density, or equivalently, it causes a reduction in the effective number of lanes; and thus, the capacity reduces. In the same study (Jin, 2010), the capability of the theory was demonstrated by using vehicle trajectories by the NGSIM project (U.S. Department of Transportation Federal Highway Administration). Another explanation for the impact that lane-changing can have on overall traffic flow is that cautious lane changers at a weaving sections can act as moving bottlenecks that can cause a decrease in the overall traffic speed. With the bounded acceleration capability of lane changers, the moving bottlenecks can further contribute to the capacity drop (Laval and Daganzo, 2006).

Fig.8 Dynamic lane assignment: a variable indication of lane role by flood light at Tokyo Metropolitan Expressway.

**4. Management of bottlenecks**

**4.1 Current managements**

Numerous measures have been proposed or implemented to eliminate or alleviate bottleneck congestion. In this section, some of the typical measures are reviewed. The measures can be roughly categorized into two groups: supply-side measures and demand-side measures. Generally speaking, the former increases the capacity of a bottleneck, whereas the latter decreases the arrival flow to a bottleneck to prevent breakdowns and capacity drops.

The most direct and simple measure to alleviate bottleneck congestion is to physically increase the capacity of a bottleneck. This can be achieved by number of ways involving infrastructure construction, such as road expansion, lane addition, road geometry improvement, implementation of electric toll collection systems, and optimization of traffic signals. However, infrastructure construction is not always practically feasible because of cost or space limitations.

By introducing advanced traffic operation schemes, traffic capacity can be improved or congestion can be prevented. Such schemes are sometimes called as active traffic management. They mainly aim to increase the supply capability of the road operation without largely altering the road infrastructure. In the following text, notable examples of active traffic management schemes are introduced. Dynamic lane assignment, sometimes referred to as interchange merge control, is a control scheme in which the role of lanes is dynamically controlled, depending on the presence of nearby merging and diverging traffic, so as to improve efficiency of lane-changing. In Fig. 8, an actual implementation of dynamic lane assignment is shown. Lane change optimization also aims to improve efficiency of lane-changing at weaving sections in a more actively manner. It designates or informs drivers of optimal lane-changing locations (Mai et al., 2016). Hard shoulder running allows drivers to travel through a shoulder lane near a bottleneck section. It directly improves the physical capacity of a bottleneck.

As sags are the common cause of traffic congestion on freeways, several countermeasures have been taken in practice. Among the several countermeasures, a common idea has been to promote driver awareness and encourage drivers to accelerate more when they are on upgrade sections. Fig. 9 is an example adopted in Hanshin Expressway Route 3 in Japan, in which horizontal blue lines are painted on the walls to notify drivers that they are on an upgrade section. In addition, a

Fig. 9 Level lines and a signboard (Route 3, Hanshin expressway)


Fig.10 Moving-light-guide-system at Hanshin Expressway (image adapted from Masumoto et al., 2018)

signboard has been installed that indicates in words as well as with an illustration that the upgrade section continues for a further 300 m from this point. As another countermeasure, there has recently been an increase in the number of installations of the moving-light-guide-system (MLGS) shown in Fig. 10 and Movie 2 (https://youtu.be/EJ_hymy03vM) that creates a flow of LED light traveling with constant speed alongside the car and which stimulates the driver to match the speed of the light. The significant positive effects of MLGS on the traffic capacity have been confirmed in many bottlenecks. Currently, the possibility of further elaboration in operation and application of MLGS is being intensively investigated.

A variable speed limit intentionally changes the speed limit of particular section, especially in upstream sections of a bottleneck, at particular times. The purpose is to decrease the arrival flow to a bottleneck in order to prevent breakdown at the bottleneck. Therefore, it maintains free flowing of traffic and prevents potential capacity drops. Ramp-metering controls inflow to a highway at on-ramps to decrease the arrival flow to a bottleneck, as in the variable speed limit. High occupancy vehicles or toll (HOV/T) lane scheme is a demand-side measure to decrease traffic volume while maintaining passenger volume. It designates some lanes as HOV/T lanes, so that a vehicle with less passengers cannot enter these lanes freely. Although HOV/T lanes do not directly increase the capacity of a bottleneck, it does however encourage efficient usage of vehicles and thus, decreases traffic volume in the long term.

Traffic demand management aims to control traffic demand and limit the arrival flow to a bottleneck to a value that is below the capacity. This is a demand-side measure. Variable speed limit, ramp-metering, and HOV/T lane can also be considered as a traffic demand management. Furthermore, information provision and congestion pricing are also used to manage traffic demand on a larger scale. By providing travellers with information on current and predicted travel times for various routes, travelers' route choices can be improved. For instance, if travelers can avoid congested routes, the resulting network traffic assignment might be efficient. However, the efficiency of the information provision scheme is limited in the sense that it only aims to achieve the user equilibrium or user optimal assignment, not the system optimal. Congestion pricing aims to achieve the system optimal assignment. Under a congestion pricing scheme, drivers who travel particular road sections need to pay a particular toll, which is designed to achieve political goals such as the alleviation of congestion.

### 4.2 Future management

In the near future, more advanced management will become possible because of connected and automated vehicles (CAVs). CAVs will change the behaviors of traffic. Furthermore, we will be able to use CAVs to control traffic.

At an operational level, capacity will be increased by CAVs. For example, coorperative driving of CAVs will enable platooning and stable car-following behaviors. The platooning of CAVs will drastically increases the capacity by shortening the headway. Stable car-following behaviors will prevent breakdowns and the propagation of stop-and-go waves. Furthermore, merging will be optimized by controlling the CAVs' lane distribution and entrance timing (Roncoli et al., 2015).

At a strategic and tactical level, demand will be optimized by controlling CAVs. For example, the choice of route and departure time of a CAV could be optimized so as to achieve the system optimal dynamic traffic assignment. In fact, this is possible by introducing a demand management scheme called tradable bottleneck permits (Akamatsu and Wada, 2017). It has been shown that bottleneck congestion can be completely eliminated by implementing this scheme in which "the road manager issues a right that allows a permit holder to pass through the bottleneck at a pre-specified time period ('bottleneck permits')" and let the permits be traded in a free market. This can be considered as an ultimate management system for bottlenecks as it eliminates all congestions.

**Acknowledgement**

This research is partially supported by JSPS KAKENHI Grants (16K18164, 19K04637, 19H02268).

**References**

Akamatsu, T., Wada, K., 2017. Tradable network permits: A new scheme for the most efficient use of network capacity. *Transportation Research Part C* 79, 178–195.

Bar-Gera, H., Ahn, S., 2010. Empirical macroscopic evaluation of freeway merge-ratios. *Transportation Research Part C* 18, 457–470.

Cassidy, M.J., Ahn, S., 2005. Driver turn-taking behavior in congested freeway merges. *Transportation Research Record* 1934, 140–147.

Cassidy, M., May, A., 1991. Proposed analytical technique for estimating capacity and level of service of major freeway weaving sections. *Transportation Research Record* 1320, 99–109.

Daganzo, C.F., 1995. The cell transmission model, part II: Network traffic, *Transportation Research Part B* 29(2), 79–93.

Daganzo, C.F., 1997. A continuum theory of traffic dynamics for freeways with special lanes. *Transportation Research Part B* 31, 83–102.

Daganzo, C.F., Lin, W.H., Del Castillo, J.M., 1997. A simple physical principle for the simulation of freeways with special lanes and priority vehicles. *Transportation Research Part B* 31, 103–125.

Edie, L.C., Foote, R.S., 1958. Traffic flow in tunnels. In: *Highway Research Board Proceedings*, 334–344.

Greenshields, B. D., 1935. A study of traffic capacity. In: *Proceedings of the 14th Annual Meeting of the Highway Research Board*, 448–477.

Goñi Ros, B., Knoop, V.L., Shiomi, Y., Takahashi, T., van Arem, B., Hoogendoorn, S.P., 2016. Modeling traffic at sags. *International Journal Intelligent Transportation System Research* 14, 64–74.

Jin, W.-L., 2010. A kinematic wave theory of lane-changing traffic flow. *Transportation Research Part B* 44, 1001–1021.

Jin, W.-L., 2018. Kinematic wave models of sag and tunnel bottlenecks. *Transportation Research Part B* 107, 41–56.

Koshi, M., 1986. Capacity of motorway bottlenecks. *Journals of the Japan Society of Civil Engineers* IV-5(371), 1-7. (in Japanese)

Koshi, M., Kuwahara, M., Akahane, H., 1992. Capacity of sags and tunnels on Japanese motorways. *ITE Journal* 62, 17–22.

Laval, J.A., Daganzo, C.F., 2006. Lane-changing in traffic streams. Transportation Research Part B 40, 251–264.

Lighthill, M. J. and Whitham, G. B., 1955. On kinematic waves. II. A theory of traffic flow on long crowded roads, *Proceedings of the Royal Society A* 229(1178), 281–316.

Mai, T., Jiang, R., Chung, E., 2016. A cooperative intelligent transport systems (C-ITS)-based lane-changing advisory for weaving sections. *Journal of Advanced Transportation* 50 (5), 752–768.

Masumoto, H., Higatani, A., Kodama, T., Kitazawa, T., Suzuki, K., Tomoeda, Y., Lee, Y. H., 2018. Assessment and effective application of the moving light guidance system in hanshin expressway. *JSTE Journal of Traffic Engineering* 4 (3), B_1–B_9. (in Japanese)

Muñoz, J.C., Daganzo, C.F., 2002. The bottleneck mechanism of a freeway diverge. *Transportation Research Part A* 36, 483–505.

Newell, G. F., 1993. A simplified theory of kinematic waves in highway traffic, part I: General theory, part II: Queueing at freeway bottlenecks, part III: Multi-destination flows, *Transportation Research Part B* 27(4), 281–313, 1993.

Ni, D., Leonard, J.D., 2005. A simplified kinematic wave model at a merge bottleneck. *Applied Mathematical Modelling* 29, 1054–1072.

Richards, P. I., 1956. Shock waves on the highway, *Operations Research* 4(1), 42–51.

Roncoli, C., Papageorgiou, M., Papamichail, I., 2015. Traffic flow optimisation in presence of vehicle automation and communication systems – part II: Optimal control for multi-lane motorways. *Transportation Research Part C* 57, 260–275.

Transportation Research Board, 1965/2016. *Highway Capacity Manual*, Transportation Research Board, Washington, D.C.

U.S. Department of Transportation Federal Highway Administration, 2016. Next Generation Simulation (NGSIM) Vehicle Trajectories and Supporting Data. [Dataset]. Provided by ITS DataHub through Data.transportation.gov. Accessed 2019-11-18 from http://doi.org/10.21949/1504477

Wada, K., Martínez, I., Jin, W.-L., 2020. Continuum car-following model of capacity drop at sag and tunnel bottlenecks, *Transportation Research Part C* 113, 260-276.

Xing, J., Muramatsu, E., Harayama, T., 2014. Balance lane use with VMS to mitigate motorway traffic congestion. *International Journal Intelligent Transportation System Research*, 12, 26–35.