



Harnessing connected and automated vehicle technologies to control lane changes at freeway merge bottlenecks in mixed traffic

Danjue Chen^a, Anupam Srivastava^b, Soyoung Ahn^{b,*}

^a Department of Civil and Environmental Engineering, University of Massachusetts Lowell, One University Avenue, Lowell, MA 01854, USA

^b Department of Civil and Environmental Engineering, University of Wisconsin-Madison, 1415 Engineering Drive, Madison, WI 53706, USA

ARTICLE INFO

Keywords:

Freeway merging
Connected and automated vehicles
Bottleneck throughput
Lane changing control

ABSTRACT

This paper proposes three control strategies for lane-changing (LC) at a merge bottleneck to improve bottleneck throughput by mitigating voids and speed disturbances in mixed traffic using connected and automated vehicle (CAV) technologies. Strategy 1 is ‘gap closure’ control, where an LC vehicle and its follower are controlled to close the void (extra time gap ahead) induced by the LC and prevent a backward-propagating speed disturbance. Strategy 2 is ‘batch LC’, where a group of LC vehicles are controlled to line up along a kinematic wave to minimize the total voids. Strategy 3 is ‘gap redistribution’ control, where extra gaps of vehicles are redistributed to periodically create large enough gaps for disturbance-free insertions. In a general control framework, the three strategies are integrated in different combinations exploiting their complementary nature. A numerical analysis shows that certain combinations, such as Strategy 2 and 3, can be very effective for improving bottleneck throughput. The analysis reveals insights on leveraging CAVs to develop traffic management strategies and/or policies and therefore improve system performance.

1. Introduction

In human-driven vehicular traffic, two traffic phenomena are critical for traffic flow efficiency and stability, and therefore system performance: reduction in bottleneck throughput (called ‘capacity-drop’) and stop-and-go oscillations. For the underlying mechanisms of these phenomena, numerous studies point to lane-changing (LC) and varying driver car-following behavior. This paper focuses on the LC behavior of vehicles. For LC, Laval and Daganzo (Laval and Daganzo, 2006) postulated that when the insertion/desertion speed of an LC vehicle is lower than the prevalent traffic, a permanent void can ensue due to the limited acceleration capability of vehicles and cause reduction in bottleneck throughput. Various studies have adopted this conjecture and produced capacity-drop values consistent with empirical measurements (e.g., Duret et al., 2011; Laval, 2009; Leclercq et al., 2016, 2011; Marczak et al., 2015; Marczak and Buisson, 2014; Chen and Ahn, 2018), suggesting that the conjectured mechanism is plausible. In contrast, Jin (Jin, 2010) proposed that LC can cause capacity-drop as the vehicles virtually occupy two lanes (current and target lanes) during LC. Jin and Laval (Jin and Laval, 2018) recently integrated the two mechanisms into a unified framework. Despite the different mechanisms conjectured, empirical evidence suggests that systematic LC maneuvers are closely linked to capacity-drop (Cassidy and Rudjanakanoknad, 2005; Elefteriadou et al., 2005; Sun et al., 2014). Numerous empirical studies have also shown that LC can trigger stop-and-go disturbances to form and grow over space (Ahn and Cassidy, 2007; Mauch and Cassidy, 2002; Zheng et al., 2013, 2011a, 2011b). These findings to date

* Corresponding author.

E-mail address: sue.ahn@wisc.edu (S. Ahn).

List of variables

a:	Constant acceleration rate
b:	Constant deceleration rate
C_0 :	Theoretical capacity ($=\frac{1}{h_0}$)
ΔC :	Unutilized capacity for control strategies
ΔC_{ML} :	Unutilized capacity in full-CAV scenario when control is extended to mainline vehicles
ΔC_{B1} :	Unutilized capacity in baseline Case B1
ΔC_{B2} :	Unutilized capacity in baseline Case B2
h_0 :	Minimum time gap in free-flow condition; i.e., equilibrium time gap
$\bar{h}$:	The mean of inter-vehicle gap $H (= \frac{1}{Q_m})$
H_0 :	Minimum time gap for disturbance-free insertion
$H^{s1,a}$:	The LC gap required in S1 control strategy to assure that the initial insertion is feasible
$H^{s1,b}$:	The LC gap required in S1 control strategy to assure that both the LC vehicle and the immediate follower maintain at least the equilibrium time gap
H:	Inter-vehicle gap size in mainline traffic
H_{LC} :	General desired LC gap under control
L:	Merge bottleneck segment length
n_A :	Number of CAVs in an LC batch
n_A^* :	Optimum number of CAVs in an LC batch to fully eliminate the LC void
Δn_A :	$n_A^* - n_A$
n_{bat} :	LC batch size
N_{max} :	Maximum allowable batch size determined by physical constraint of merge bottleneck
N_0 :	The number of vehicles that an LC has to wait to find an acceptable LC gap in Case B1
N_0^* :	Expected value of N_0 corresponding to the confidence bound Π
m:	Mainline platoon size
m^* :	Optimum platoon size
o:	Void created by an LC vehicle
o_{rs} :	Residual void after gap closure
o_{ex} :	The excessive void size after accounting for the tolerance level of the LC vehicle ($\Delta\eta_{LV}$)
Δo :	The closed void size achieved in void closure processes
o_{LC}^{s1} :	Final void size per LC vehicle under Strategy 1
$\hat{o}_{rs}$:	Residual void normalized by h_0
$\hat{o}_{rs,ML}$:	Void residual (in relative scale of h_0) in full-CAV scenario with mainline control
o_{B2} :	Void size of LC vehicles in Case B2
p:	Penetration rate of CAVs
p_{crit} :	Critical penetration rate of CAVs, above which, $n_A \geq n_A^*$ is possible
Q_m :	Mainline flow
Q_r :	LC flow
Q_{tot} :	Total discharge flow; $Q_{tot} = Q_m + Q_r$
Q_r^* :	Optimum LC flow that maximizes discharge flow when the mainline flow is given
$Q_{r,B1}^*$:	Maximum possible LC flow in baseline Case B1
$Q_{r,B2}^*$:	LC flow in baseline Case B2
u:	Free-flow speed
v_0 :	Initial LC speed
v_{max} :	Maximum allowable speed that exceeds u
v_{max}^* :	Optimal speed that minimize the duration of gap closure
T_{tl} :	The maximum tolerable duration during which the vehicle exceeds u
T_{tl}^* :	Minimum of T_{tl} achieved when the LC vehicle can reach v_{max}^* in gap closure process
T_{cru} :	Time period during which the LC vehicle is cruising at v_{max}
T_{cru}^* :	Optimum value of T_{cru} in case that v_{max}^* exceeds v_{max}
T_c :	Period of LC insertion in Case B2 ($= 1/Q_{r,B2}^*$)
T_{crit} :	Total time of an LC vehicle accelerating from v_0 to u, and its final wave, emanated just as it reaches u, traveling back the acceleration distance
η_{LV} :	Gap factor of the LC vehicle at insertion (as a fraction of the equilibrium gap)
$\Delta\eta_{LV}$:	The tolerance level of the LC vehicle ($= 1 - \eta_{LV}$)
η_F :	Gap scale of the immediate follower of an LC vehicle (as a fraction of the equilibrium gap)

$\Delta\eta_F$	Tolerance level of the follower
γ	Mainline flow ratio ($= \frac{Q_m}{C_0}$)
γ_{LB}	Lower bound of mainline flow ratio for controllable traffic
γ_{ML}^{crit}	Critical mainline flow ratio in full CAV scenario, above which, void-free can be achieved
$\gamma_{LB,ML}$	Lower bound of mainline flow ratio for controllable traffic when control is extended to mainline traffic
Π	Confidence bound of finding an LC gap
δ	Jam spacing in the Kinematic Wave theory
κ	Jam density in the Kinematic Wave theory
λ	Extra time gap between two vehicles ($=H - h_0$)

suggest that it is desirable to control LC maneuvers to reduce their impacts on traffic flow efficiency and stability and therefore improve the overall system performance.

The emerging connected and automated vehicle (CAV) technologies integrate sensing, communication, and automation technologies (Shladover, 2017), and offer great potential to improve traffic flow via precise control of CAV car-following and LC behaviors. It is expected that CAVs will co-exist with human-driven vehicles (HDVs) for the next two decades or even longer. Motivated by that, our paper aims to harness the automation and connectivity to develop strategies to control LC behaviors of CAVs in the mixed traffic (consisting of CAVs and HDVs) and therefore improve traffic flow.

Many strategies targeting car-following or LC behaviors have been proposed; see Rios-Torres and Malikopoulos (Rios-Torres and Malikopoulos, 2016) for a review on the use of coordinated control at intersections and highway merges, and Scarinci and Heydecker (Scarinci and Heydecker, 2014) who characterized control strategies by design components and focus. For LC behavior, the majority of research focused on developed merging assistance algorithms to improve safety and fuel efficiency; see Letter and Elefteriadou (Letter and Elefteriadou, 2017) for a detailed review. The algorithm of Letter and Elefteriadou (Letter and Elefteriadou, 2017) aimed to maximize the average travel speed, which controlled the merging vehicles at a constant speed prior to the insertion and also controlled the mainline vehicles to create gaps at the size of the minimum time gaps to accommodate the merging vehicles. It was found that the control has improved travel time while the throughput varied with the minimum time gap setting. Rios-Torres et al. (Rios-Torres et al., 2015) designed an algorithm to optimize the acceleration profile of all (controllable) vehicles on the mainline and the ramp within a control zone near the merge to maximize fuel economy. More specifically, the problem was formulated into an optimization problem regarding the merging order of the mainline and ramp vehicles. Scarinci et al. (2013) developed a cooperative ramp-metering strategy that controls some mainline vehicles to decelerate to achieve the optimum density and increase throughput (locally), while creating gaps for merging vehicles released from ramp-metering. Notably, the above works developed the control in a pure CAV setting assuming that all vehicles are controllable.

The control of LC behavior in mixed traffic is a much challenging task and there is only limited research. Along this line, Park et al. (2011) developed an algorithm that utilizes CV technologies to assist merging by controlling mainline vehicles near a merging area to create gaps needed for smooth insertions. Pueboobpaphan et al. (2010) developed a merging assistance algorithm to smoothen the speed of mainline traffic (by having mainline vehicles slow down early and smoothly) while creating gaps for on-ramp traffic. Duret et al. (2020) proposed bi-level merging strategies in the context of truck platoons, based on optimization models. The bi-level control addresses a tactical control for optimal vehicle ordering and gap optimization, and a lower-level control of CF to create optimal gaps for merging vehicles. Karimi et al. (2020a) developed two optimal merging controls for situations where 1) CAVs, and 2) HDVs are the majority class primarily engaged in the LC maneuvering. Karimi et al. (2020b) developed lower-level trajectory optimization for cooperative merging of CAVs in mixed traffic conditions.

While most studies cited above claim improvement of traffic flow, the underlying mechanisms and path to improvement are not clear. Additionally, none of the studies have directly addressed void creation by LC maneuvers coupled with disturbance growth. Thus, opportunities remain to develop control strategies that directly address the mechanisms of void creation and disturbance propagation. The present study explores such opportunities exploiting the uncharted potential of the CAV technologies. Specifically, this study aims to develop CAV-enabled control strategies for LC maneuvers in the context of mixed traffic to mitigate voids and the ensuing speed disturbances, thereby improving bottleneck throughput and traffic flow stability.

The proposed strategies are at the tactical-level; i.e., we illustrate what the vehicle trajectories should look like without going into the details of mechanical control (e.g., throttle and jerk). Specifically, this study proposes three control strategies: (1) S1: gap closure, (2) S2: batch LC, and (3) S3: gap redistribution (on mainline vehicles). S1 entails individual vehicle control: an LC vehicle and its follower are controlled to temporarily accept shorter, non-equilibrium time gaps and exceed the free-flow speed (LC vehicle only), thereby closing the void induced by the LC maneuver. S2 entails systematic grouping and positioning of LC vehicles: a batch of LC vehicles are positioned along a kinematic wave to minimize the total voids. S3 controls a group of vehicles in the mainline to redistribute extra time gaps and create large enough gaps for disturbance-free insertions. These strategies are complementary and can be integrated to further improve the control effectiveness and adapt to a wider range of traffic conditions. We develop a general control framework that can be reduced to individual strategies or different combinations of the strategies. Notably, while the strategies proposed are simple, the analysis reveals insights on leveraging CAVs to develop traffic management strategies and/or policies and therefore improve system performance.

Towards the end, this paper is organized as follows. Section 2 presents the research problem. Section 3 introduces the three control strategies. Section 4 develops a general control framework that can integrate the three strategies in different combinations. Section 5

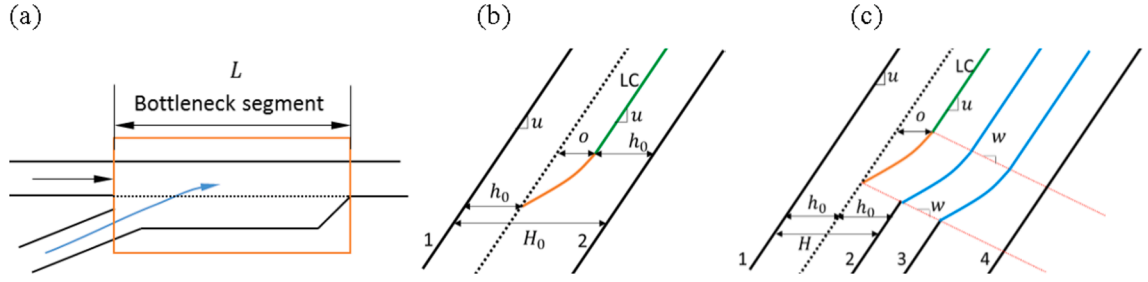


Fig. 2.1. (a) Merge bottleneck schematic; (b) Ideal insertion gap H_0 ; (c) Insertion gap with disturbance ($H < H_0$).

evaluates the benefits of the (integrated) control strategies. Section 6 discusses the limitations of this research and desirable future research. Conclusions are provided in Section 7.

2. Research problem

We study mixed traffic consisting of CAVs and HDVs. We consider a simple freeway merge bottleneck, with one mainline lane and a one-lane on-ramp (see Fig. 2.1 (a) for the sketch), where traffic is freely flowing. For the LCs of vehicles that are not controlled, we adopt the void creation mechanism proposed by Laval and Daganzo (Laval and Daganzo, 2006) that has been used by numerous studies and make the following assumptions. We assume HDVs will change lane with initial speed v_0 , smaller than free-flow speed u ¹. The LC vehicles accelerate at a constant rate of a to reach u and then travel at that speed thereafter. We assume for simplicity that all vehicles follow Newell's car-following model (Newell, 2002) to obtain key insights for control. We will discuss how to relax this assumption to consider more complex driver behavior in future research in Section 6.

When an LC vehicle inserts into the traffic stream at the equilibrium position, it will create a void (extra time gap), o , due to the bounded acceleration as illustrated in Fig. 2.1(b). The void can be derived based on the simple Newtonian equation of motion as follows:

$$o = (u - v_0)/a - \left(\frac{u^2 - v_0^2}{2ua} \right) = \frac{(u - v_0)^2}{2ua}. \quad (1)$$

Then the minimum LC time gap to avoid propagation of the speed disturbance by the LC vehicle, H_0 , is expressed as:

$$H_0 = h_0 + o + h_0 \quad (2)$$

where h_0 is the equilibrium time gap (also the minimum time gap at free-flow u corresponding to the capacity state). Specifically, in a triangular-shape fundamental diagram, $h_0 = \frac{1}{C_0}$, where C_0 is the *theoretical capacity*. Notably, here the time gap H_0 contains three components: the LC void o , the equilibrium gap needed by the LC vehicle h_0 , and the equilibrium time gap needed by the follower² (i.e., the lag vehicle in the mainline lane).

If an insertion gap, H , is smaller than H_0 , the insertion will induce a backward-propagating disturbance. Fig. 2.1(c) illustrates an example with $H = 2h_0$. The follower of the LC vehicle (vehicle 2) slows down immediately upon insertion, and the deceleration wave propagates upstream until it encounters extra gap(s) upstream, as depicted with vehicle 4 in the figure.

3. LC control strategies

For the baseline traffic (free-flow in the mainline and the on-ramp), we propose three control strategies, S1, S2, and S3, with increasing level of coordination to minimize the magnitude of voids and/or disturbances and thus maximize bottleneck throughput. S1 is 'gap closure' control at the individual vehicle level (microscopic) involving the LC vehicles and their immediate followers in the mainline to prevent persistent voids; S2 is 'LC grouping', involving coordinating multiple LCs to minimize creation of both voids and disturbances; and S3 is 'gap redistribution' control towards mainline vehicles to provide proper LC gaps and thus minimize disturbance propagation. These three strategies are complementary and can be combined for greater benefits. To capture the improvement on bottleneck throughput, we define unutilized capacity as the difference between the bottleneck throughput and the theoretical capacity. In our context, unutilized capacity can still exist if our control cannot fully eliminate the LC voids. Therefore, reducing LC voids is equivalent to reducing unutilized capacity and also equivalent to increasing bottleneck throughput.

This section will focus on introducing the underlying mechanism of each individual strategy assuming 100% CAV. Next section will

¹ This assumption is reasonable for merges where an acceleration lane is not sufficiently long and/or on-ramp vehicles are controlled by ramp-metering.

² In this paper, "follower" refers to the lag vehicle of a subject vehicle, and the "leader" refers to the immediate proceeding vehicle of a subject vehicle in the same lane.

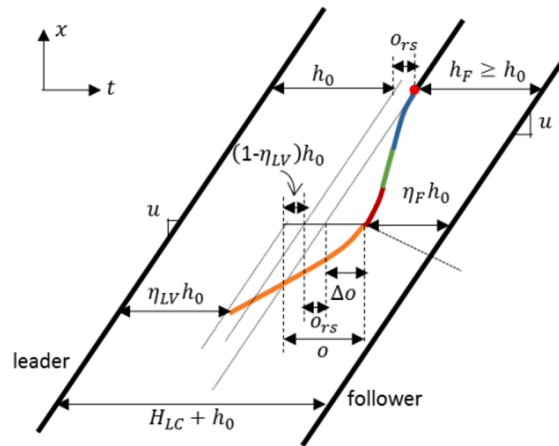


Fig. 3.1. Illustration of strategy 1 (S1): gap closure.

discuss how the strategies can be combined for more sophisticated control in mixed and full CAV traffic conditions.

Our control strategies assume the following features of CAVs: they can (1) precisely maintain certain time gaps, (2) insert into a gap smaller than equilibrium spacing, (3) temporarily exceed free-flow speed (to a limited extent) to close the gap ahead without a significant increase of safety risk, and (4) communicate with each other to share information such as the gap size. Note that (1) is already attainable with the adaptive cruise control (ACC) technologies (Serafin, 1996). (2) is already a common practice for some HDVs, which leads to the relaxation phenomenon (Duret et al., 2011; Laval and Leclercq, 2008; Leclercq et al., 2007) and would be suitable for CAVs given their smaller reaction time (as small as 40% of HDVs (Shladover et al., 2012)). (3) is feasible for CAVs to do with minimum risks as they can have precise control of their maneuvers. In fact, it is also commonly seen in HDVs. (4) is a common feature for connected vehicle technologies.

3.1. S1: Gap closure

In Strategy 1 (S1), we propose to control an LC vehicle and its immediate follower to mitigate the void induced by the LC and thus improve bottleneck throughput. S1 focuses on void reduction, and disturbance control is addressed in S2 and S3. The main principles of S1 are that (i) the LC vehicle and its follower³ can temporarily tolerate a smaller gap (than the equilibrium level); and (ii) the LC vehicle can slightly exceed u to close the void. This is illustrated in Fig. 3.1. Note that these types of behaviour are common in real traffic and can be controlled more precisely for better stability with automation technologies. Specifically, (i) is essentially the relaxation phenomenon commonly observed in HDVs during LC. To quantify this process, let η_{LV} capture the gap factor of the LC vehicle (as a fraction of the equilibrium gap) at insertion, and $\Delta\eta_{LV} (= 1 - \eta_{LV})$ represents the tolerance level of the LC vehicle upon insertion; i.e., the relaxation scale. The final desired position depends on the maximum allowable speed $v_{\max}$ ($v_{\max} > u$), and the maximum tolerable duration, denoted by, T_{ij} , during which the vehicle exceeds u . Note that the gap of the follower will be minimum when the LC vehicle reaches u at the end of the void creation, which is captured by the factor η_F (as a fraction of h_0). Also, at the end of the gap closure (i.e., red dot), the time gap of the LC vehicle and the follower should be equal to or larger than h_0 . Otherwise, the follower has to slow down, and the deceleration may propagate upstream; see Fig. 3.1.

Intuitively, the best strategy to close the void is for the LC vehicle to (1) accelerate at its maximum allowable rate a to recover u (see the orange trajectory segment), (2) continue to accelerate at the maximum rate to reach $v_{\max}$ ($v_{\max} > u$; see the red trajectory segment), (3) cruise at $v_{\max}$ for a time period of T_{cru} (to be derived shortly; see the green trajectory segment), and then (4) decelerate at the maximum allowable rate b to recover u (see the blue trajectory segment). For simplicity, we assume that the maximum allowable acceleration and deceleration rates are equal (i.e., $a = b$), which can be different in practice but will not change the main results. Note that (1) is the void creation process, and (2)-(4) are the gap closure processes. Throughout these processes, the follower maintains the constant speed of u .

Notably, depending on the initial condition (η_{LV} , v_0 , and a) and the process of gap closure (v_{max} , T_{tl} , a and b), a residual void, o_{rs} , may exist in front of the LC vehicle in the end as illustrated in Fig. 3.1, which satisfies the following relationship:

$$o_{rs} = o_{ex} - \Delta o, \quad (3)$$

where o_{ex} denotes the excess void size with respect to the equilibrium position, h_0 , and Δo denotes the amount of gap closure:

$$o_{\text{ex}} = o - \Delta\eta_{ly}h_0. \quad (4)$$

³ The control is also feasible if the follower does not experience the relaxation (i.e., $\eta_F = 1$). In this case, the control is less effective in void reduction and the LC vehicle does not have to exceed u .

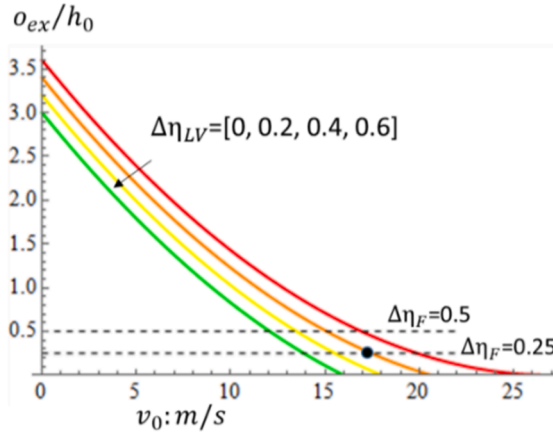


Fig. 3.2. comparison between o_{ex} and $\Delta\eta_F$ ($v_{max} = 115\%u$) (In all numerical calculation in this paper, we use $u = 30$ m/s, $a = 2$ m/s², $C_0 = 2000$ vph, $h_0 = 1.8$ s, wave speed $w = -5$ m/s).

$$\Delta\eta_{LV} = 1 - \eta_{LV} \quad (5)$$

$$\Delta o = T_{cru} \left(\frac{v_{max}}{u} - 1 \right) + \frac{(v_{max} - u)^2}{2ua} + \frac{(v_{max} - u)^2}{2ub}. \quad (6)$$

o_{rs} will be bounded by 0 as we can always have the LC vehicle decelerate less in the gap closure process. Δo depends on the LC vehicle's void closing capability, which is determined by v_{max} and T_{il} (T_{il} implicitly decides the available cruising time T_{cru}). We find that in most conditions, vehicles can fully close the void: i.e., $\Delta o = o_{ex}$, and $o_{rs} = 0$; see the detailed analysis provided in Appendix A. Thus, in the remaining of the paper, we assume that CAVs can close the void; i.e., $\Delta o = o_{ex}$ and thus $o_{rs} = 0$.

Referring to Fig. 3.1 for the control, the disturbance-free LC gap is given by the following:

$$H_0 = \max \{ H^{s1,a}, H^{s1,b} \} \quad (7)$$

$$H^{s1,a} = 2h_0 + o - \Delta\eta_{LV}h_0 - \Delta\eta_F h_0, \quad (8)$$

$$H^{s1,b} = o_{rs} + 2h_0, \quad (9)$$

where $H^{s1,a}$ is the required insertion gap to assure that the initial insertion is feasible (i.e., to accommodate the spacing needed by the LC vehicle and the follower), and $H^{s1,b}$ is to assure that the follower can eventually reach a gap equal to the equilibrium time gap. $H^{s1,a}$ and $H^{s1,b}$ can be computed as:

where $\Delta\eta_F (= 1 - \eta_F)$ denote the tolerance level of the follower. The condition that $H^{s1,a}$ dominates (i.e., $H^{s1,a} \geq H^{s1,b}$) is reduced to:

$$\frac{\Delta o}{h_0} \geq \Delta\eta_F. \quad (10)$$

This suggests that if the void reducing capability of the LC vehicle (captured by $\Delta o/h_0$) exceeds the tolerance level of the follower (captured by $\Delta\eta_F$), $H^{s1,a}$ will determine the minimum gap H_0 . Since vehicles are likely to fully close the void (i.e., $\Delta o = o_{ex}$), the inequality in Eqn. (10) is equivalent to $\frac{o_{ex}}{h_0} \geq \Delta\eta_F$. Fig. 3.2 shows the values of o_{ex} with respect to various factors such as the initial LC condition (v_0, a) and the tolerance level of the LC vehicle, $\Delta\eta_{LV}$. One can see that Eq. (10) holds in most cases unless v_0 or $\Delta\eta_{LV}$ is large. For example, with $\Delta\eta_{LV} = 0.2$, $\Delta\eta_F = 0.25$, $v_{max} = 1.15u$ (black dot in the figure), Eqn. (10) holds as long as $v_0 < \frac{17.45m}{s}$.

Note that when $H^{s1,a}$ dominates, the LC will still induce a void, though it is smaller than the non-control case. In contrast, when $H^{s1,b}$ dominates (i.e., $\frac{o_{ex}}{h_0} < \Delta\eta_F$), the LC will be void-free as o_{rs} equals to 0. The final void size under S1, o_{LC}^1 , can be concisely expressed as:

$$o_{LC}^1 = H_0 - 2h_0 = \max \{ o - (\Delta\eta_{LV} + \Delta\eta_F)h_0, 0 \}, \quad (11)$$

where H_0 is the disturbance-free LC gap given by Eqn. (7).

3.2. S2: Batch LC

In strategy S2, we propose to strategically position LCs to minimize voids and thus improve bottleneck throughput. This is motivated by our recent work (Chen and Ahn, 2018) that shows that the cumulative voids produced by LCs depend on the temporal-spatial distribution of LCs. Inspired by this work, the principle of S2 is to place LCs along the characteristics line of wave (i.e., $-w$) as a batch so that the cumulative voids of the LC batch are minimized. This is illustrated in Fig. 3.3. To simplify the illustration, we assume that LC

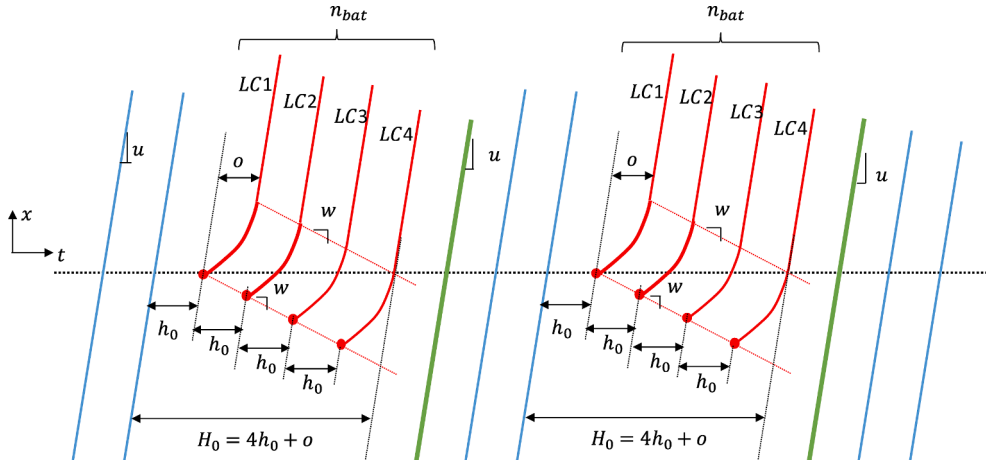


Fig. 3.3. Batch insertion for S2 (red trajectories denote LC vehicles and blue ones denote mainline vehicles). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

vehicles take the equilibrium position when they insert (i.e., $\eta_{LV} = 1$). By aligning LCs along the wave, each LC batch (referring to all insertions along the same wave) only produces one void o (given by Eq. (11)) in front of the first LC vehicle in the batch. Let n_{bat} denote the batch size (i.e., number of insertions in one batch). The total gap needed for all n_{bat} insertions to be disturbance-free is:

$$H_0 = n_{bat} * h_0 + o + h_0. \quad (12)$$

The effect of the control is equivalent to reducing the void per LC to o/n_{bat} ; i.e.,

$$o_{LC}^2 = o/n_{bat}. \quad (13)$$

Clearly, the control is more effective in reducing voids with larger n_{bat} . However, the allowable n_{bat} depends on the physical space available for merge at the bottleneck and operational conditions around the merge. Let N_{max} be the maximum allowable batch size, i.e., the upper bound of n_{bat} , constrained by the LC segment length:

$$N_{max} = \frac{L}{\delta}, \quad (14)$$

where L is the merge bottleneck segment length, and δ is jam spacing in the Kinematic Wave theory (Lighthill and Whitham, 1955; Richards, 1956) with a triangular shape fundamental diagram (i.e., $\delta = \frac{1}{k}$). Notably, N_{max} is usually large for typical ranges of parameter values. For example, existing studies (Punzo and Simonelli, 2005; Treiber and Kesting, 2013) suggest that δ is about 7m, and L is in the range of 50 m to 500 m (Leclercq et al., 2016, 2011), which results in N_{max} in the range of 7 to 71 vehicles.

Note that $n_{bat} = N_{max}$ corresponds to the case that LC vehicles in one batch occupy the entire bottleneck segment L , which may not be desirable in practice. Besides the physical constraint, operational feasibility may reduce allowable n_{bat} ; for example, drivers on the mainline may not accept a very large batch of LCs from the ramp. Thus, n_{bat} should be determined considering the mainline demand as well as the merging demand.

The implementation of this strategy would be similar to ramp metering. Specifically, merging vehicles can be held at an on-ramp to form a batch and then released at the pace to insert along the characteristics line. Also, the size of a LC batch should be provided to the mainline vehicles in advance to create a proper LC gap. Such information sharing may be achieved through vehicle-to-vehicle communication if the traffic has 100% CAVs. With mixed traffic, the information of LC batch size will be collected by roadside infrastructure (e.g., loop detectors) and shared with mainline vehicles through vehicle-to-infrastructure communication. The coordination between a batch of LC vehicles and the mainline gaps can follow a merging assistance design similar to Park et al. (Park et al., 2011) and Pueboobpaphan et al. (Pueboobpaphan et al., 2010).

3.3. S3: Gap redistribution

Strategy S3 aims to re-distribute gaps among vehicles in the mainline to create gaps at the desired size (e.g., H_0 for Strategy 1 per Eqn. (7)) so that they can be fully utilized. Specifically, with S3, mainline vehicles are either bunched to maintain a gap of h_0 or directed to create a desired LC gap, denoted by H_{LC} to be general. Vehicles in the former case form a “packed platoon”, implying that no insertions should occur, and vehicles in the latter case purposely leave appropriate extra gaps to allow insertions; see Fig. 3.4. This strategy is motivated by the facts that in baseline traffic (without control) some extra gaps would be too small for insertions and thus wasted; and some insertions could generate backward-moving disturbances if the inserting gaps are not sufficiently large. Thus, by bunching vehicles in advance, extra gaps can be accumulated to create a desirable gap at the right timing to accommodate an insertion without generating a disturbance. To achieve this, vehicles with gaps larger than h_0 but smaller than H_{LC} will be controlled to travel



Fig. 3.4. Control of Strategy 3: (a) baseline; (b) re-distributed gaps.

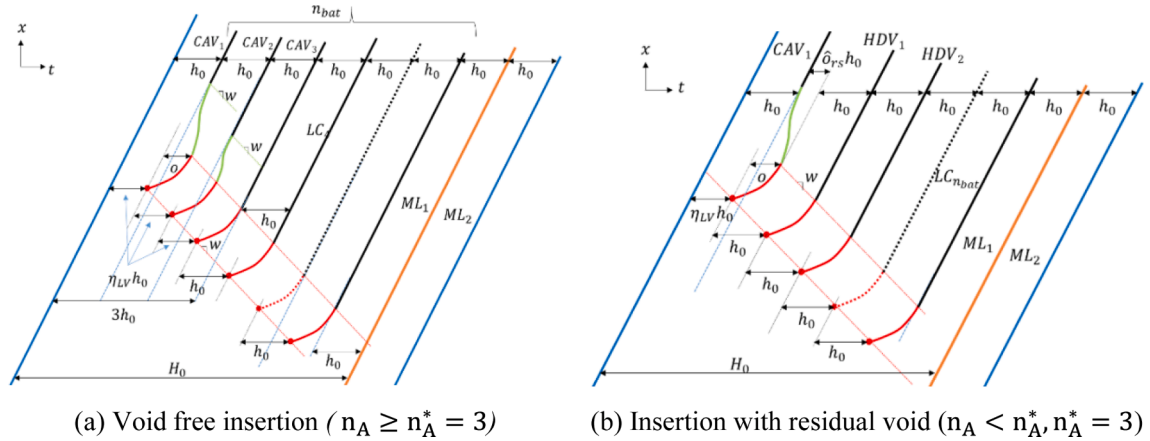


Fig. 4.1. Control for: (a) Case 1 – void free insertion; and (b) Case 2 – insertion with residual void.

slightly slower than u to shift the extra gaps downstream to cumulate a gap size of H_{LC} .

Specifically, the physical process is the opposite of the gap closure process in S1; i.e., a CAV in the mainline will be controlled to slow down to a lower speed, cruise at the lower speed for a certain interval, and then accelerate to resume u . Vehicles upstream of the controlled CAV will slow down accordingly once they become constrained by their leader. We omit the detailed trajectory path derivation for brevity. In the gap creation process, we assume that the HDVs, which are uncontrollable, will behave according to Newell's CF model when constrained. The impact of this assumption will be discussed in Section 6. We caution that the gap redistribution should preferably occur upstream of the merging segment and the mainline vehicles resume the free-flow condition before entering the merging segment. This assures that when the LC occurs, the mainline traffic is in a stable condition.

4. Integrated control

In this section, we integrate the three strategies, S1, S2, and S3. The main principle is to create usable gaps according to S3 (gap redistribution), considering batch insertions of LC vehicles according to S2 and the gap closure of CAVs according to S1. Note that the integration of the three strategies will use a general control framework, which can be reduced to individual strategies or different combinations of strategies by changing the different parameter settings. Analysis of the different combinations of the strategies will be considered in Section 5.

In our control design, we assume that the mainline flow, Q_m , is externally determined (and thus not controllable through the present strategy), but the inserting ramp flow, Q_r , is controllable, which can be achieved through ramp metering. The primary control objective is to minimize voids and thus maximizing bottleneck throughput while minimizing persistent propagation of disturbances. We assume that the CAV penetration rate in the traffic, for both the mainline and the ramp, is p .

We first introduce the design of the integrated control strategy and then discuss the two potential cases, insertion with void-free and with void residual. After that, we derive the valid domain of control and the optimum ramp flow that maximizes bottleneck throughput. Lastly, we examine the case with 100% penetration of CAVs, which allows for extending the control to mainline traffic to further increase bottleneck throughput. The benefits of the control compared to baseline cases will be quantified in more details in the following section.

4.1. Design of the integrated control for mixed traffic

Let n_A denote the number of CAVs in an LC batch, where $n_A = n_{bat}p$, ($n_A = n_{bat}$ if traffic consists only of CAVs). The control strategy consists of the following steps (illustrated in Fig. 4.1(a) below):

Step 1: Gap creation – control mainline CAVs to create LC gaps of desired size and at the proper frequency (to derive shortly).

Step 2: Batch insertion of CAVs to close the gap – we assume that vehicles on the ramp can be ordered as needed so that we can

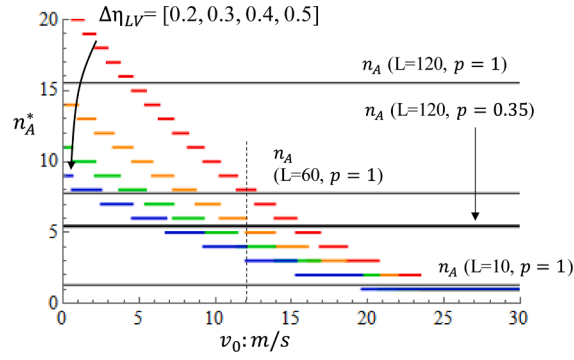


Fig. 4.2. Plots for n_A^* value and n_A (with $n_{bat} = N_{max}$).

position at least n_A^* (n_A) CAVs at the head of the merging batch, where n_A^* is the number of CAVs required to close the gap (to be computed shortly). The vehicle ordering can be achieved in practice if there are two lanes on an on-ramp or there is a turn-bay (e.g., similar to the control for transit signal priority in He et al. (2016)). Vehicles in the LC batch are positioned to insert into the mainline along the characteristic line as covered in S2. These first n_A^* CAVs will have to follow the gap closure process to close the void⁴ induced by the LC vehicle and take equilibrium positions as described in S1; see the green trajectories in Fig. 4.1(a). Specifically, these vehicles will be controlled to temporarily accept smaller time gaps, $\eta_{LV}h_0$, until they accelerate to u . Thereafter, they continue to accelerate beyond u to close the void. Notice that the gap closure process is the longest for the first CAV but becomes shorter for the upstream CAVs. The last vehicle in this set (i.e., the n_A^* -th vehicle) is controlled to accept a smaller time gap upon insertion but reach its equilibrium position by the time it accelerates to u . Thus, gap closure will be unnecessary for this vehicle and others upstream.

Step 3: Batch insertion of remaining LC vehicles – insert the remaining LCs in the batch ($n_{bat} - n_A^*$), regardless of the vehicle type, in the equilibrium positions along the characteristics line as in S2. For this set of vehicles, gap closure is not necessary.

The number of CAVs required to close the gap, n_A^* , in Step 2, is given by:

$$n_A^* = \left\lceil \frac{o/h_0}{\Delta\eta_{LV}} \right\rceil \quad (15)$$

n_{bat} will be an integer and n_A^* will take the floor integer.

Clearly, depending on the number of CAVs in an LC batch, n_A , there are two possible cases: $n_A \geq n_A^*$ and $n_A < n_A^*$, where the former can fully eliminate the LC void while the latter will have a residual void. We refer to them as Case 1 - void-free insertion and Case 2 - insertion with residual void, respectively. Fig. 4.1(b) provides an example of Case 2, in which $n_A^* = 3$ but $n_A = 1$. Note that, for Case 2, in Step 2 of the control, it is critical to have all CAVs in the beginning of the LC batch to maximize the void mitigation, because HDVs are not controllable to close a void systematically. Since HDVs do not accelerate beyond u to close the gap, when the void residual reaches the first HDV in the batch, it will persist in front of that vehicle, and upstream CAVs will not be able to reduce the void; see HDV₂ in Fig. 4.1(b).

For both cases, the desired LC gap for the LC batch, H_0 , is straightforward:

$$H_0 = n_{bat}^* h_0 + h_0 \hat{o}_{rs} + h_0, \quad (16)$$

where $\hat{o}_{rs}$ represents the void residual normalized by h_0 , given below.

$$\hat{o}_{rs} = \max \left\{ \frac{o}{h_0} - n_{bat} p \Delta\eta_{LV}, 0 \right\} \quad (17)$$

The effect of the control is equivalent to reducing the void (normalized) per LC to $\hat{o}_{rs}/n_{bat}$; i.e.,

$$\hat{o}_{LC}^{int} = \max \left\{ \frac{\frac{o}{h_0}}{n_{bat}} - p \Delta\eta_{LV}, 0 \right\}, \quad (18)$$

where $\hat{o}_{LC}^{int}$ is the control effectiveness of the integrated strategy.

Note that when $p < 1/n_{bat}$, not every LC batch will have a CAV, and the $\hat{o}_{rs}$ in Eqn. (17) gives the average void residual across multiple LC batches. Of note, this formulation (Eqn. (17)) needs a slight modification when $p < 1/n_{bat}$ and $\frac{o}{h_0} < \Delta\eta_{LV}$, suggesting that if

⁴ We assume all CAVs are able to fully close the gap as we mention in Section 3.1. This holds in most prevalent traffic conditions; see the analysis in Appendix A.

an LC batch has one or more CAVs, they are able to fully eliminate the void. Fortunately, this is not frequent in real practice. Thus, we only discuss the occurrence condition and provide the formulation in Appendix B.

4.2. Conditions for insertion with void-free and void residual

This sub section discusses the possibility of Case 1 ($n_A \geq n_A^*$) and Case 2 ($n_A < n_A^*$).

According to Eqn. (15), the required number of CAVs, n_A^* , depends on the spacing tolerance level, $\Delta\eta_{LV}$, and the LC void o , which boils down to the initial merging speed, v_0 . Clearly, n_A^* decreases quickly with $\Delta\eta_{LV}$ and with v_0 ; see the plots in Fig. 4.2. One can see that if $\Delta\eta_{LV} \geq 0.3$ (the orange plot) and $v_0 \geq 10\text{m/s}$, $n_A^* \leq 7$, which seems reasonably small for practice. On the other hand, n_A depends on the LC batch size, n_{bat} , which is externally determined and is bounded by N_{max} (i.e., $n_A \leq N_{max}$), and increases with the penetration rate p . Fig. 4.2 shows the upper bounds of n_A , which are achieved when n_{bat} is maximized to N_{max} and $p = 1$ (i.e., $n_A = N_{max}$). The figure shows n_A for three representative segment length: 10 m, 60 m, and 120 m. One can see that if $\Delta\eta_{LV}$ can reach 0.3 (the orange plot) and v_0 is at the moderate level (i.e., $v_0 \geq 12\text{m/s}$), $n_A > n_A^*$ holds for $L \geq 60\text{m}$; see the orange plot lying below the curve of n_A with $L = 60\text{m}$. Moreover, notice that $n_A^* \leq 5$, which seems very feasible for implementation. Note that $\Delta\eta_{LV} = 0.3$ is a feasible value as research on HDVs using the NGSIM data has shown that HDVs can deviate from the equilibrium spacing when experiencing stop-and-go waves and resume equilibrium later, with similar deviation magnitudes ($\Delta\eta_{LV} \sim 0.3$) (Chen et al., 2012).

Of course, if $p < 1$, n_A will be smaller and it is more difficult to satisfy $n_A \geq n_A^*$. More specifically, for a given n_{bat} , there exists a threshold penetration rate, p_{crit} , above which, $n_A \geq n_A^*$ is possible. p_{crit} can be easily derived.

$$p_{crit} = \frac{n_A^*}{n_{bat}} \quad (19)$$

Clearly, if $p \geq p_{crit}$, void-free insertion can be achieved. By contrast, voids are inevitable if $p < p_{crit}$. Fig. 4.2 shows the n_A plot for $L = 120\text{m}$ and $p = 0.35$; see the thick black line. One can see that if the merge segment length is moderate ($L = 120\text{m}$ or longer) and the penetration rate is moderate (0.35 or above), it is possible to achieve void-free; i.e., Case 1.

4.3. Valid control domains & improvement of bottleneck throughput

Below, we derive the control premise and formulate the improvement on bottleneck throughput. For simplicity, we assume that CAVs are distributed uniformly (according to their penetration rate), and batch insertions and mainline platoons are periodic and uniform in size; i.e., each LC batch (at size n_{bat}) inserts in front of a packed mainline platoon with platoon size m . This can be interpreted as the “average” dynamics of the traffic stream. This simplification allows us to easily scale the impact of microscopic control to the macroscopic level and quantify the control effect. The issue of stochasticity will be discussed in Section 6.

Regardless of whether the control is void-free or there is residual void, the control requires the following conditions to hold.

$$\frac{Q_m}{m} \geq \frac{Q_r}{n_{bat}}, \quad (20)$$

$$\frac{m}{Q_m} - mh_0 \geq n_{bat}h_0 + h_0\hat{o}_{rs}, \quad (21)$$

$$mp \geq 1 \quad (22)$$

These three conditions are the control premise. Specifically, Eqn. (20) assures that the gap frequency in the mainline (i.e., number of gaps per time unit) at least equals to the frequency of LC batches, and Eqn. (21) assures that the extra mainline gap is sufficient to accommodate the inserting vehicles and the void residual. Eqn. (22) requires that each mainline platoon has at least one CAV to create an LC gap. Otherwise, traffic is not fully controllable. Clearly, Eqn. (20) and Eqn. (21) impose constraints on the possible range of Q_r and thus affect the resulting bottleneck throughput; while Eqn. (22) imposes a constraint on the valid domain of control pertaining to the mainline traffic condition related to m and p .

We will derive the valid domain of control shortly. At this point, we discuss the optimum control on Q_r to maximize the total bottleneck throughput (and thus minimize the unutilized capacity) when Eqn. (20)–(22) are satisfied. Since Q_m is determined externally, intuitively, the total discharge flow will be maximized when equality holds for both Eqn. (20) and Eqn. (21); i.e., the LC batch frequency equals to the frequency of mainline platoon and the extra mainline gaps equal to the spacing required by the inserting vehicles and the void residual. This suggests that the platoon size of mainline traffic should be controlled at the optimum:

$$m^* = \frac{n_{bat} + \hat{o}_{rs}}{1/\gamma - 1}. \quad (23)$$

where $\gamma = Q_m/C_0$, is the mainline flow ratio (normalized by lane capacity), and $\hat{o}_{rs}$ is the residual void given by Eqn. (17).

Accordingly, from the equality in Eqn. (20) and Eqn. (21), we can derive the maximum possible value of Q_r , denoted by Q_r^* , and the corresponding unutilized capacity (normalized by C_0), ΔC :

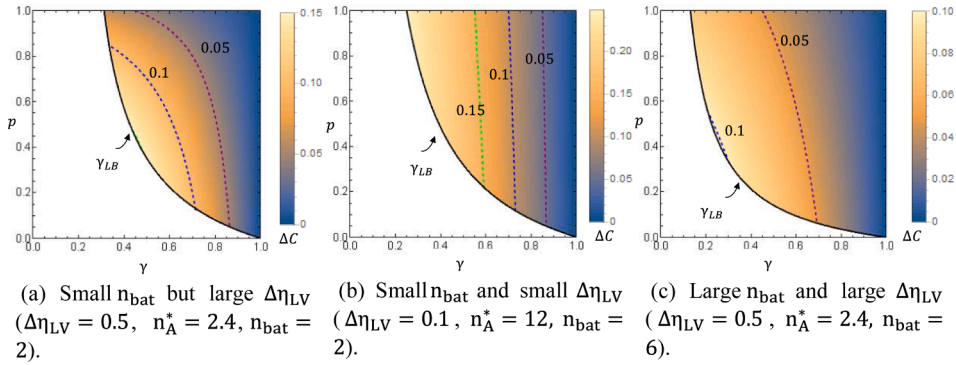


Fig. 4.3. Valid domains for control in mixed traffic with moderate void size ($o/h_0 = 1.2$).

$$Q_r^* = \frac{Q_m}{m} n_{bat} = C_0 (1 - \gamma) \frac{n_{bat}}{n_{bat} + \hat{o}_{rs}}, \quad (24)$$

$$\Delta C = \frac{C_0 - Q_r^* - Q_m}{C_0} = (1 - \gamma) \frac{\hat{o}_{rs}}{\hat{o}_{rs} + n_{bat}} = (1 - \gamma) \frac{\hat{o}_{LC}^{int}}{\hat{o}_{LC}^{int} + 1}, \quad (25)$$

where $\hat{o}_{LC}^{int}$ is the void (normalized) per LC under control. Note that Q_r is also bounded by the maximum possible flow corresponding to v_0 according to a triangular fundamental diagram; i.e., $Q_r \leq C_0 v_0 (u + w) / [u(v_0 + w)]$. However, this upper bound is usually large and does not impose a constraint. For example, for $v_0 = 0.15u$ (4.5m/s), the upper bound is $0.55C_0$, which exceeds the typical range of LC flow (Marczak & Buisson 2014; Leclercq et al. 2011; Marczak et al. 2015; Leclercq et al. 2016). Therefore, generally the valid range of Q_r is $[0, Q_r^*]$.

Note that if the mainline platoon size m exceeds m^* , the control is still valid though less effective, because the extra gap in the mainline is not fully utilized. In this case, Q_r^* is obtained by assuming the equality in Eqn. (20). Its value and the corresponding ΔC can be derived below:

$$Q_r^* = C_0 \gamma \frac{n_{bat}}{m}, \quad (26)$$

$$\Delta C = \frac{C_0 - Q_r^* - Q_m}{C_0} = 1 - \gamma \frac{n_{bat}}{m} - \gamma \quad (27)$$

By contrast, if $m < m^*$, Eqn. (21) is violated and the control can't sustain as the void residual will propagate upstream and alter the gaps created. In the remaining analysis, we assume that m is controllable and thus m^* can be achieved.

Regarding the valid domain of control imposed by Eqn. (22), since m can be controlled to be m^* , which is related to the mainline flow γ , this constraint imposes a lower bound of γ , denoted by γ_{LB} , which depends on the penetration rate p . Thus, the valid domain of control can be described in the 2-dimension plane of $\{\gamma, p\}$. γ_{LB} can be derived by integrating Eqn. (23) into Eqn. (22), which is given below:

$$\gamma_{LB} = \frac{1}{1 + p \left(n_{bat} + \hat{o}_{rs} \right)}. \quad (28)$$

$$\gamma_{LB} = \frac{1}{1 + p^* \max \left\{ \frac{o}{h_0} + n_{bat} (1 - p \Delta \eta_{LV}), n_{bat} \right\}} \quad (29)$$

Below γ_{LB} , there are not enough mainline vehicles to be controlled to create LC gaps at the desired rate. The physical meaning of the γ_{LB} is that our control needs mainline CAVs to create gaps at the desired size. If the mainline flow is too low, there won't be enough CAVs to execute the control. In this case, however, ramp flow can still merge as in the non-control case, but the inserting gaps in the mainline will not be precisely equal to the desired size and there will be unutilized capacity.

Note that in the formulation in Eqn. (28), γ_{LB} depends on two components, p and $n_{bat} + \hat{o}_{rs}$. The latter is the LC gap needed per batch, which consists of two elements, n_{bat} - the equilibrium spacing (normalized), and $\hat{o}_{rs}$ - the residual void. This component depends on n_{bat} , $\Delta \eta_{LV}$, and o . Fig. 4.3 illustrates the valid domains of control in the 2-dimension plane of $\{\gamma, p\}$ (shaded regions) and the unutilized capacity within the domain (scaled by colors). Without loss of generality, we consider a moderate size of void o (resulting from the $v_0 = 0.46u = 13.8\text{m/s}$). We obtain the following remarks:

R1: The valid domain grows as the LC gap per batch increases. More specifically, it shrinks with $\Delta \eta_{LV}$ (see Fig. 4.3 (a) vs. (b)) but

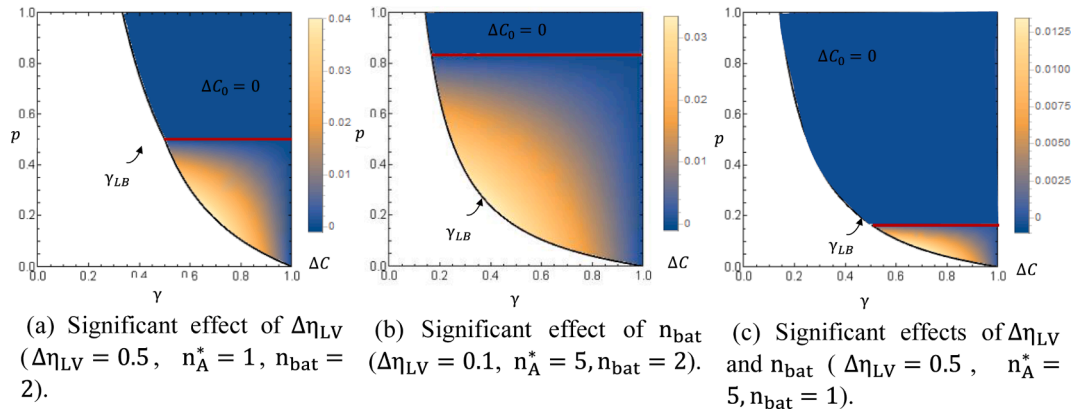


Fig. 4.4. Valid domains for control in mixed traffic with small void size ($o/h_0 = 0.5$).

grows with n_{bat} (see Fig. 4.3 (a) vs. (c)) and void size o (see Fig. 4.3 vs. Fig. 4.4). This is because when the LC gap per batch increases (due to decrease of $\Delta\eta_{LV}$ or increase of n_{bat} or/and o), the control requires a larger platoon size in the mainline, m^* , to create the gap needed. Accordingly, the penetration rate p needed to satisfy Eqn. (22) is smaller; i.e., the valid domain is larger.

R2: γ_{LB} is negatively correlated with p ; see the decreasing boundaries of the valid domain in Fig. 4.3 and Fig. 4.4. Note that the LC gap per batch decreases with p but the effect is much smaller than p itself; i.e., the increase of p dominates in the denominator of Eqn. (28). The physical meaning is straightforward: as the penetration rate increases, the mainline platoon size needed to satisfy the constraint in Eqn. (22) is smaller.

R3: Regarding the unutilized capacity ΔC , within the valid domain of control, ΔC decreases towards the upper right (with larger γ and larger p); see Fig. 4.3. This is expected: larger mainline flow leaves smaller share of roadway space for LC vehicles and thus a smaller chance of unutilized capacity. For the effect of p , as the penetration rate increases, the void residual $\hat{o}_{rs}$ is smaller and thus ΔC is smaller, which is also implied by Eqn. (25). It is worth noting that, when n_{bat} and $\Delta\eta_{LV}$ are small, the unutilized capacity is not very sensitive to p ; see Fig. 4.3 (b). One can see the three contour lines (with $\Delta C = 0.05, 0.1$, and 0.15) are almost vertical. This is because when both impact factors ($\Delta\eta_{LV}$ and n_{bat}) are small, there is a limited potential for CAVs to contribute even with $p = 1$. By contrast, when either of them is significant, the unutilized capacity clearly decreases as p increases; see Fig. 4.3 (a) and (c).

R4: A region of zero unutilized capacity (i.e., bottleneck throughput equals to C_0) can exist within the valid domain; see Fig. 4.4 for an example. It is straightforward that this corresponds to the case⁵ when $\hat{o}_{rs} = 0$. This is more likely to occur for small to moderate voids in which a large $\Delta\eta_{LV}$ (plot (a)) or n_{bat} (plot (b)) can fully eliminate the void. Of course, a combination of both factors leads to a larger region of zero unutilized capacity; see plot (c).

Notably, outside the valid domains of control, it is infeasible to do precise control to create desirable LC gaps and assure disturbance-free insertion because there are not enough CAVs in the mainline. However, it is still possible to implement “approximate control” with the same principle to reduce voids. Specifically, as in the case of $n_A < n_A^*$ above, CAVs from an on-ramp can be placed in the beginning of each LC batch to mitigate the void to the extent possible. Note that since we do not have full control of gap creation, it is likely that the actual LC gap provided by the mainline is smaller (disturbance would propagate upstream through mainline vehicles until it is resolved by an upstream gap), or larger (part of the gap remains unutilized though the disturbance does not propagate) than the desirable gap, H_0 .

4.4. Extended control to the mainline with full CAVs

In the case of full CAVs, the control can be extended to the mainline to further reduce the unutilized capacity when there is a void residual after the ramp control. The control principle is similar to the control for the merging vehicles on the ramp, except that now the mainline vehicles (CAVs) are controlled to tolerate a smaller spacing level (assumed η_{LV} for simplicity) and exceed u to close the gap. The number of mainline vehicles needed to fully eliminate the void, Δn_A , can be derived as:

$$\Delta n_A = n_A^* - n_A \quad (30)$$

Fig. 4.5 illustrates an example of control with $n_A^* = 3$ but $n_A = n_{bat} = 1$. For brevity, details of the mainline control are omitted here. After the control on the LC vehicle from the ramp, there is a void residual. Therefore, the control is extended to two mainline vehicles upstream, ML_1 and ML_2 and finally be resolved.

Note that the control premises in Eqns. (20)–(22) still apply with the formulation for void residual should be updated to capture the cooperation of the mainline vehicles. Let $\hat{o}_{rs,ML}$ denote the relative void residual, where the “ML” in the subscript indicates that it has

⁵ Another case is $\gamma = 1$, but this case is trivial as it suggests that the mainline flow equals to full capacity.

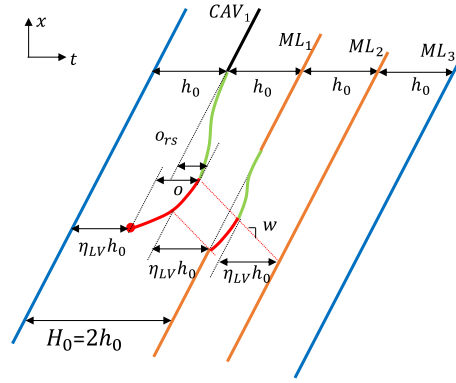


Fig. 4.5. Control extended to mainline in full CAVs scenario ($n_A^* > N_{\max}$, orange trajectories are mainline vehicles). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

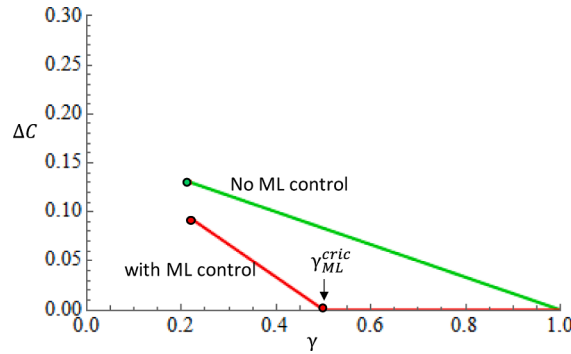


Fig. 4.6. Unutilized capacity in full CAV case with and without mainline control ($n_A^* = \frac{1.2}{0.2} = 6$, $n_{bat} = 3$).

extended the control to mainline vehicles. $\hat{o}_{rs,ML}$ is given as follows:

$$\hat{o}_{rs,ML} = \max\{o/h_0 - (n_A + m)\Delta\eta_{LV}, 0\}, \quad (31)$$

where m is the platoon size of mainline traffic, because all the mainline vehicles are CAVs and can be controlled to reduce the LC void. Accordingly, the optimal platoon size required to minimize unutilized capacity can be obtained below.

$$m_{ML}^* = \frac{\frac{o}{h_0} + n_{bat}(1 - \Delta\eta_{LV})}{1/\gamma - (1 - \Delta\eta_{LV})}. \quad (32)$$

Then, $\hat{o}_{rs,ML}$ and m_{ML}^* can be integrated into Eqns. (20)–(22) to obtain the optimal ramp flow (we omit this for brevity) and thus the unutilized capacity with mainline-control, ΔC_{ML} , given below:

$$\Delta C_{ML} = \frac{\max\{o/h_0(1 - \gamma) - n_{bat}\Delta\eta_{LV}, 0\}}{o/h_0 + n_{bat}(1 - \Delta\eta_{LV})}. \quad (33)$$

Of course, the lower bound imposed by the control premise in Eqn. (22) is still valid, but the value, denoted by $\gamma_{LB,ML}$, is larger as the residual void $\hat{o}_{rs,ML}$ is smaller than $\hat{o}_{rs}$ due to the contribution of the mainline vehicle.

$$\gamma_{LB,ML} = \frac{1}{(n_{bat} + 1)(1 - \Delta\eta_{LV}) + o/h_0} \quad (34)$$

It is straight forward that there exists a critical value of γ , denoted by γ_{ML}^{crit} , above which, void-free is always feasible. This critical value corresponds to the case that the platoon size m equals to Δn_A per Eqn. (30). This can be easily obtained by solving Eqn. (31):

$$\gamma_{ML}^{crit} = 1 - \frac{n_{bat}}{n_A^*} = 1 - \frac{n_{bat}}{\frac{o/h_0}{\Delta\eta_{LV}}} \quad (35)$$

Fig. 4.6 illustrates the unutilized capacity in full CAV with and without mainline control (red and green plot respectively). One can see that the extended mainline control has significantly reduced the unutilized capacity.

In summary, for the full CAV scenario, if the maximum allowable batch size is relatively large ($N_{\max} \geq n_A^*$), our control can achieve

void-free for LCs by controlling only the merging vehicles - as long as $\gamma \geq \gamma_{LB}$ (γ_{LB} is given by Eqn. (28)). If N_{max} is relatively small ($N_{max} < n_A^*$), LCs will produce voids and result in unutilized capacity, but complementary control on the mainline traffic can be used to further reduce or even eliminate unutilized capacity - as long as $\gamma \geq \gamma_{LB,ML}$ ($\gamma_{LB,ML}$ given by Eqn. (34)). Lastly, if the mainline flow is too low ($\gamma < \gamma_{LB,ML}$), traffic is not fully controllable.

5. Quantification

In this section, we compare the benefits of the proposed control strategies with the baseline cases (without the proposed control). We examine the benefits from two perspectives, bottleneck throughput and disturbance level, which are closely related. To maintain a reasonable scope, we have derived the analytical formulation for the unutilized capacity (and thus bottleneck throughput), but the analysis of disturbance level is left qualitative to emphasize on the main insight. We first analyze the baseline and then compare it with different combinations of the three proposed strategies.

5.1. Baseline cases

For the baseline, we consider the lower and upper bounds with different LC priority and minimum acceptable LC gaps:

- Lower bound - Case B1: LC vehicles choose gaps large enough to be disturbance-free. The minimum LC gap H_0 is given by Eqn. (2).
- Upper bound - Case B2: LC vehicles tolerate much smaller gaps to maximize the chances of merging into the mainline. The minimum LC gap H_0 is assumed to be h_0 .
- Upper bound with relaxation - Case B2': Like Case B2, LC vehicles tolerate much smaller gaps to maximize the chances of merging into the mainline. Moreover, upon insertion, the LC vehicles will experience the relaxation process, inserting initially with shorter spacing, $\eta_{LV}h_0$ ($0 < \eta_{LV} \leq 1$), and relaxing to the equilibrium h_0 . The minimum LC gap H_0 is assumed to be $(1 - \eta_{LV})h_0$.

Case B1 prioritizes minimizing roadway disturbance while Cases B2 and B2' prioritize the chances of insertion. These cases represent the upper and lower bounds of the baseline in terms of unutilized capacity and the lowest and highest level of disturbance, respectively. Below we specify the setting of each case and formulate their unutilized capacity level.

Case B1. In this case, we consider the number of vehicles that an LC vehicle must wait for, to find an acceptable LC gap, denoted by N_0 . We assume that the extra time gap, λ , follows an exponential distribution with mean $\bar{h} - h_0$; i.e.,

$$\lambda = H - h_0, \quad (36)$$

$$P(\lambda = X) = \frac{1}{\bar{h} - h_0} e^{-X/(\bar{h} - h_0)}, \quad (37)$$

where H is the stochastically distributed inter-vehicle gap (bumper to bumper), $\bar{h} = 1/Q_m$ is the mean of H . Thus, the probability of finding a gap greater than or equal to H_0 can be obtained as:

$$P(H \geq H_0) = P(\lambda \geq H_0 - h_0) = e^{-(H_0 - h_0)/(\bar{h} - h_0)} \quad (38)$$

Since each LC is disturbance-free and does not affect subsequent ones, the gap selection processes can be considered i.i.d. and follows a Geometric distribution with the success probability equal to $P(H \geq H_0)$. Accordingly, the mean of N_0 is derived as:

$$E[N_0] = \frac{1}{P(H \geq H_0)} = e^{(H_0 - h_0)/(\bar{h} - h_0)} \quad (39)$$

Since the gaps are stochastic, we consider the confidence of finding an acceptable gap, π . The corresponding number of waiting gaps is denoted by N_0^* . This is equivalent to the probability of failing to find an acceptable gap after N_0^* gaps not exceeding $(1 - \pi)$. Thus, the critical value of N_0^* can be derived as follows:

$$1 - (P(H \leq H_0))^{N_0^*} = \pi \quad (40)$$

$$N_0^* = \frac{\text{Log}[1 - \pi]}{\text{Log}\left[1 - e^{-(H_0 - h_0)/(\bar{h} - h_0)}\right]} \quad (41)$$

Accordingly, the maximum ramp flow, $Q_{r,B1}^*$, and the unutilized capacity, ΔC_{B1} , can be derived as:

$$Q_{r,B1}^* = \frac{1}{N_0^* \bar{h}} \quad (42)$$

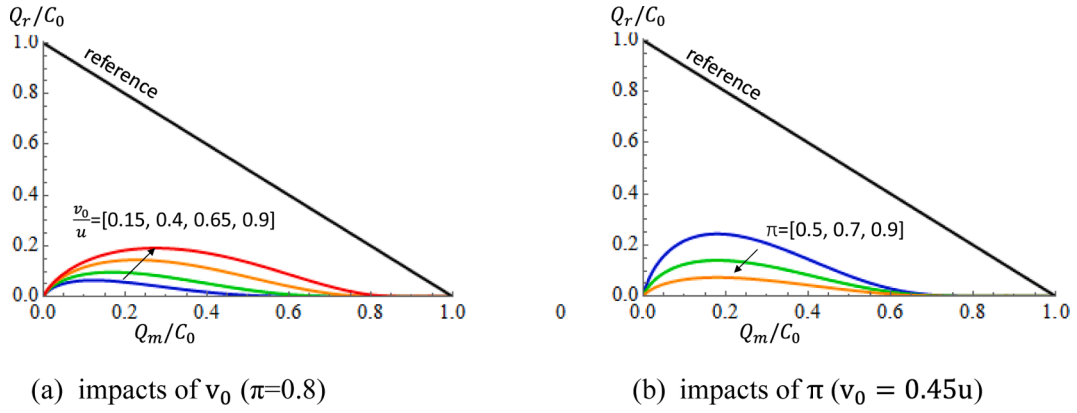


Fig. 5.1. Unutilized capacity in Case B1.

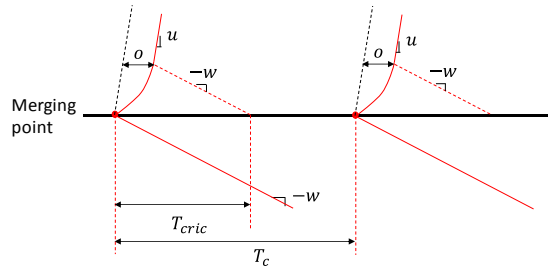


Fig. 5.2. Setting of baseline Case B2.

$$\Delta C_{B1} = \frac{C_0 - Q_{r,B1}^* - Q_m}{C_0} \quad (43)$$

We omit the elaborate forms to be succinct.

Fig. 5.1 shows the unutilized capacity in Case B1 and the impacts of the influential factors. The diagonal black line is the reference line with full capacity. The vertical or horizontal distance between a curve to the reference line captures the unutilized capacity. From plot (a) one can see that the unutilized capacity varies significantly with inserting speed v_0 that decides the size of initial void o and thus the LC gap H_0 . Particularly, when v_0 is small (e.g., the blue curve for $v_0 = 0.15u = 4.5\text{m/s}$), the maximum LC rate $Q_{r,B1}^*$ is very small ($<10\%$ of C_0). This is because the initial void o is very large ($\sim 3h_0$) and the probability of finding a large gap is very small. The unutilized capacity ΔC_{B1} also varies significantly with the confidence level π (see plot (b)) and becomes very large at high confidence levels (e.g., see the orange curve for $\pi = 0.9$). In general, $Q_{r,B1}^*$ is small even when v_0 is large and the mainline flow is small to medium, and that results in a significant unutilized capacity. Here Case B1 does not consider multiple LCs occurring into one gap, and thus it overestimates the unutilized capacity. However, the difference will be small considering that multiple LCs will require even larger gaps and the probability of finding such large gaps would be very small.

Case B2. For simplicity, this case assumes that (i) LC vehicles insert into the fully saturated traffic (i.e., $Q_m = C_0$) with mainline vehicles all in the equilibrium positions (headway is h_0); (ii) LC occurs at a fixed location⁶; (iii) upon insertion, an LC vehicle will take the equilibrium position in the mainline; and (iv) LCs occur periodically; see Fig. 5.2 for an illustration. Note that the setting of Case B2 implicitly implies that LC vehicles have higher priority than the mainline vehicles and thus the actual total LC flow equals to the “desired” LC rate to force into the mainline. Let $Q_{r,B2}^*$ be the LC rate. The problem setting is the same as in Chen and Ahn (2018) and we can apply the same principles to formulate the unutilized capacity, ΔC_{B2} . Specifically, ΔC_{B2} is equal to the void size per LC period:

$$\Delta C_{B2} = \frac{O_{B2}}{T_c} \quad (44)$$

⁶ Assumption (ii) is made to simplify the calculation without significantly compromising the accuracy. Specifically, in the control cases, n_{bat} will be mostly below 7 vehicles to be feasible in practice, which requires a very short segment ($L = 49\text{m}$ with $\delta = 7\text{m}$). In such a short segment, LCs in the baseline could be spatially distributed and result in smaller unutilized capacity than fixed point insertion but the difference is negligible as showed in Chen and Ahn (2018).

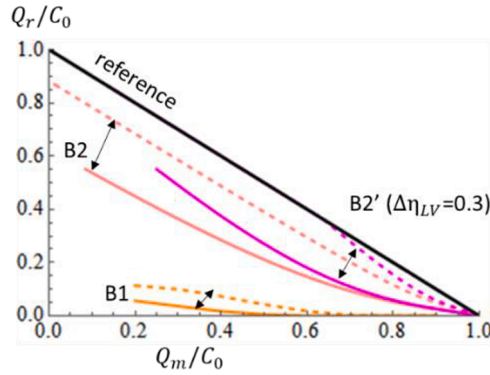


Fig. 5.3. Capacity of baseline cases (Orange curve: Case B1, $\pi = 0.8$; Pink curves: Case B2; Magenta curves: Case B2'; Solid lines: $v_0 = 0.15u$; Dashed lines: $v_0 = 0.5u$). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

where the void size, o_{B2} , can be derived according to the geometric relationship:

$$o_{B2} = \begin{cases} \frac{aT_c w + (u + w) \left(v_0 + w - \sqrt{2aT_c w + (v_0 + w)^2} \right)}{au}, & \text{if } T_c < T_{\text{crit}} \\ \frac{(u - v_0)^2}{2ua}, & \text{if } T_c \geq T_{\text{crit}} \end{cases}, \quad (45)$$

where T_c is the period of LC insertion ($= 1/Q_{r,B2}$) and $T_{\text{crit}} (= \frac{u-v_0}{a} + \frac{u^2-v_0^2}{2w.a})$ is the total time of an LC vehicle accelerating from v_0 to u , and its final wave, emanated just as it reaches u , traveling back the acceleration distance; see Fig. 5.2. The condition of $T_c < T_{\text{crit}}$ occurs when adjacent insertions interact and only cause a partial void; while the other condition occurs when adjacent insertions do not interact and a full-size void (given by Eqn. (1)) is created. The elaborate form of ΔC_{B2} has been derived in Chen and Ahn (2018) and we omit here to remain succinct.

Case B2'

This case is similar to Case B2 except that the final void is smaller due to the non-equilibrium insertion of the LC vehicle. The void per LC, $o_{B2'}$, can be derived in a similar fashion as Eqn. (11) except that mainline traffic is now congested:

$$o_{B2'} = \begin{cases} \max \left\{ -\frac{aT_c w + (u + w) \left(v_0 + w - \sqrt{2aT_c w + (v_0 + w)^2} \right)}{au} - \Delta\eta_{LV} h_0, 0 \right\}, & \text{if } T_c < T_{\text{crit}} \\ \max \left\{ \frac{(u - v_0)^2}{2ua} - \Delta\eta_{LV} h_0, 0 \right\}, & \text{if } T_c \geq T_{\text{crit}} \end{cases}, \quad (46)$$

Accordingly, the unutilized capacity in Case B2', $\Delta C_{B2'}$, is equal to the void size per LC period:

$$\Delta C_{B2'} = \frac{o_{B2'}}{T_c} \quad (47)$$

Fig. 5.3 shows the lower bound (orange curves), upper bound (pink curves), and upper bound with relaxation (magenta curves) of the unutilized capacity under low and high v_0 . Clearly, regardless of the v_0 , the unutilized capacity in upper bound - Case B2 is much smaller than lower bound- Case B1, and the unutilized capacity is even smaller if relaxation is considered per Case B2'. More importantly, when v_0 is small, there is still significant unutilized capacity in the upper bound; see the solid magenta curve.

Admittedly, the baseline investigation here is not the most comprehensive and accurate representation of the baseline traffic. Nonetheless, they can be treated as the upper and lower bounds of the baseline reference. Furthermore, the three cases considered yield important insights: with stochastic distribution of gaps, it is difficult to achieve a disturbance-free insertion even with moderate mainline flow (Q_m) if v_0 is small to moderate, and a significant part of the roadway capacity is not utilized. When LC vehicles are willing to tolerate smaller LC gaps, the roadway capacity is better utilized and bottleneck throughput is larger. Of course, that also comes with the cost that all vehicles, mainline and LC, have to bear more extensive disturbances.

5.2. Control vs. Baseline cases

With the baseline references established, we now turn our attention to the benefits of the proposed control strategies as compared to the baseline. Note that the control strategies can be combined differently to adapt to different available resources. We first examine the

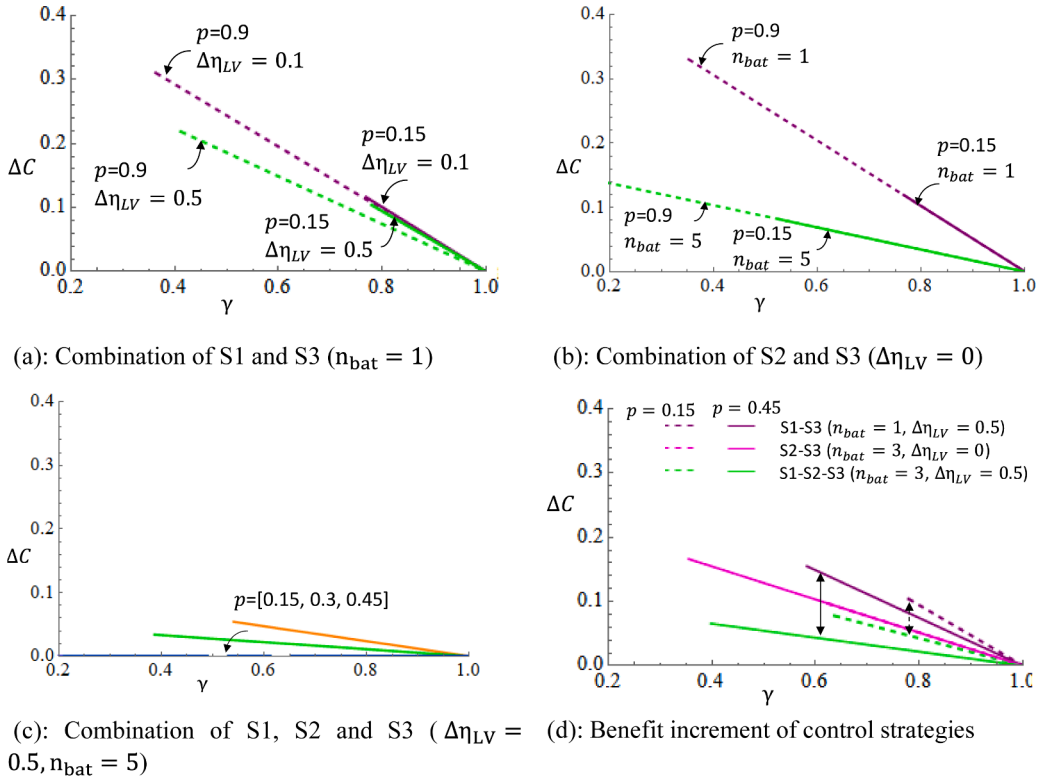


Fig. 5.4. Unutilized capacity in different control combinations ($v_0 = 0.5u$).

effectiveness of different combined strategies and then compare them with the baseline cases. Note that the different combinations of the strategies can be easily obtained by changing the different parameter settings of the general framework for the integrated control in Section 4, which can be reduced to different combinations of strategies.

5.2.1. Benefits of combined control strategies

For the individual strategies, while it's feasible to apply them individually, a single strategy has limited effectiveness in mixed traffic unless the penetration rate is high. This is expected because merging involves two continuous streams (mainline and on-ramp) while the individual strategies only pertain to one or a few vehicles from one stream. Therefore, integration is crucial for fully utilizing the control potential. Below we examine the effectiveness of three different combinations: (i) S3 combined with S1 (i.e., S1-S3), (ii) S3 combined with S2 (i.e., S2-S3), and (iii) all three strategies (i.e., S1-S2-S3). To obtain (i), we use the general framework for the integrated control in Section 4 but n_{bat} is set to 1 (i.e., no LC batching) and p and $\Delta\eta_{LV}$ vary. Similarly, to obtain (ii), $\Delta\eta_{LV}$ is set to be 0 (i.e., gap closure is not active) but p and n_{bat} vary. For (iii), $\Delta\eta_{LV}$, p and n_{bat} all vary. Additionally, for (i), since the LC vehicle will insert in front of a CAV in the mainline, we assume that the mainline CAV will tolerate the same level of spacing as the LC CAV (i.e., $\Delta\eta_F = \Delta\eta_{LV}$ per mechanism in Fig. 3.1).

Notably, since traffic is mixed with CAVs and HDVs, the control needs to handle the LC of HDVs. Specifically, for S1-S3, we assume that the gap re-distribution has created proper gaps to accommodate CAVs and HDVs from the on-ramp; i.e., each LC vehicle will merge into the mainline using a gap created by the mainline traffic per S3, and the gap size is given by Eqn. (7) for a CAV LC and by Eqn. (2) for a HDV LC according to Laval and Daganzo (Laval and Daganzo, 2006) described in Section 2. For S2-S3, the LC vehicles (CAVs or HDVs) will always merge in a batch into a gap created by the mainline traffic via S3, and the gap size is given by Eqn. (12). For S1-S2-S3, the merging process is fully described in Section 4.1.

We obtain the following remarks on the impacts of these different combinations based on the analysis results shown in Fig. 5.4:

- R1: For the combination of S1 and S3, the penetration rate p has the following impacts: as p increases, (a) the valid domain of control enlarges, and (b) the unutilized capacity decreases; see the solid line vs. dashed line in green ($\Delta\eta_{LV}$ is held at 0.5 but p increases from 0.15 to 0.9). As $\Delta\eta_{LV}$ increases, the unutilized capacity decreases; see the dashed lines in purple vs. green ($p = 0.9$, $\Delta\eta_{LV}$ increases from 0.1 to 0.5). These results are expected because the LC batches (each has only one LC) have a probability p to have a CAV and enjoy the control benefit (in void reduction), $p\Delta\eta_{LV}$, per $\hat{o}_{LC}^{int}$ given by Eqn. (18). As p (or $\Delta\eta_{LV}$) increases, the control benefit increases as well. Notably, given that the control benefit is given by $p\Delta\eta_{LV}$, when p is small, the variation of $\Delta\eta_{LV}$ does not make a significant difference (see the solid lines in purple vs. green which almost overlap) and vice versa. This suggests

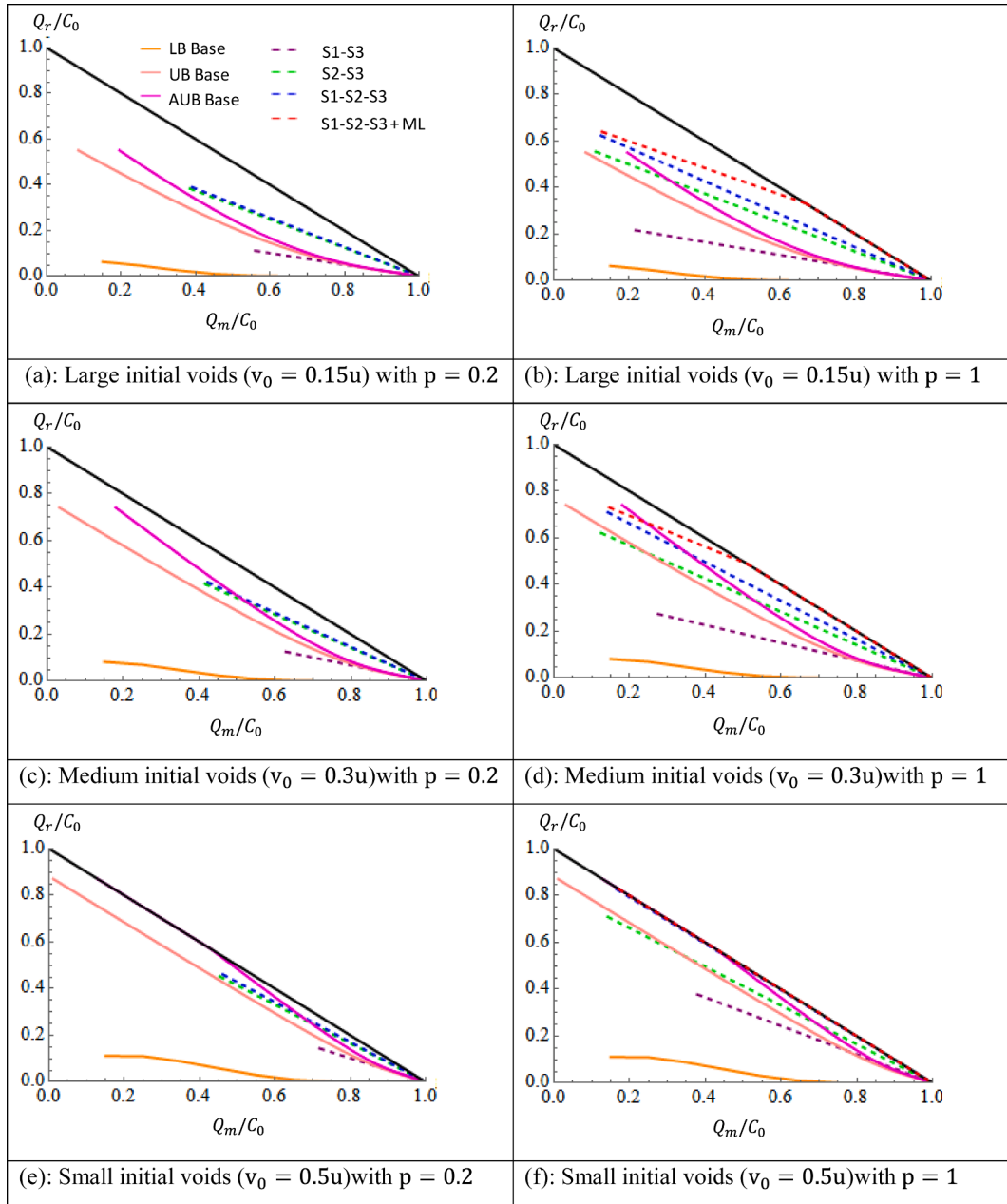


Fig. 5.5. Control vs. base cases: black solid line = reference line with $\Delta C = 0$, orange solid line = lower bound baseline (Case B1), pink solid line = upper bound baseline (Case B2), magenta solid line = upper bound with relaxation (Case B2') with $\Delta\eta = 0.2$. For S1-S3, $\Delta\eta_{LC} = 0.2, n_{bat} = 1, \pi = 0.8$; for S2-S3, $\Delta\eta_{LC} = 0, n_{bat} = 5$; for combination of S1, S2, and S3, $\Delta\eta_{LC} = 0.2, n_{bat} = 5$. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

that in low penetration rate, the control of S1 (implicitly supported by S3) is not very effective (i.e., small reduction of unutilized capacity and a small valid control domain). This is straightforward as only a small proportion of LC vehicles can be controlled.

- R2: For the combination of S2 and S3, p only affects γ_{LB} (and thus the valid domain of control) but not the unutilized capacity; see the solid vs. dashed lines in the same color (purple or green) in plot (b). This is because in such integration, there is no void closure by CAVs and p only affects the frequency of LC batches. However, n_{bat} has a significant impact on the unutilized capacity; see the solid lines in purple vs. green in plot (b). This is because an LC batch shares the initial void, and n_{bat} affects the void per LC in the batch, per Eqn. (13).
- R3: When all three strategies are combined, the control is very effective in reducing the void. From plot (c), one can see that the unutilized capacity decreases quickly and becomes zero after a certain penetration rate. Moreover, the control effectiveness varies

Table 5.1

Features of different combinations of the control strategies.

	S1-S3	S2-S3	S1-S2-S3
Void ($\hat{o}_{LC}^{int}$) per LC after control	$\max\{o/h_0 - p\Delta\eta_{LV}, 0\}$	$\max\{o/h_0/n_{bat}, 0\}$	$\max\{o/h_0/n_{bat} - p\Delta\eta_{LV}, 0\}$
Unutilized capacity after control	$\Delta C = (1 - \gamma) \frac{\hat{o}_{LC}^{int}}{\hat{o}_{LC}^{int} + 1}$		
Control setting	$n_{bat} = 1$ $0 < \Delta\eta_{LV} < 1$ for CAVs and $\Delta\eta_{LV} = 0$ for HDVs	$n_{bat} > 1$ $\Delta\eta_{LV} = 0$ for all LC vehicles	$n_{bat} > 1$ $0 < \Delta\eta_{LV} < 1$ for CAVs and $\Delta\eta_{LV} = 0$ for HDVs
Implementation requirement	Mainline-on-ramp coordination on insertion timing.	Requirements for S1-S3 + Roadside assistance to position vehicles along characteristic line and control the inserting location/timing.	Requirements for S2-S3 + The ramp should have more than one lane to reposition CAVs in the head of the LC batch.
Applicable conditions	Ramp geometry The on-ramp can be short. p level Moderate to high p LC demand LC demand that can be served is small when p is small to moderate.	The on-ramp needs to have an extended segment to store the LC batch. Large range of p Can serve a wider range of demand even p is small.	The on-ramp needs to have an extended segment to store the LC batch. Large range of p Can serve a wider range of demand even p is small.
Effectiveness in improving bottleneck throughput	Only effective for high p	Effective even for low p	Very effective.

with strategies. On plot (d), one can see the benefit of S1-S3 (solid magenta line) and S2-S3 (solid pink line). When S2 is added to S1-S3, the unutilized capacity decreases, and the benefit increment is the difference between the magenta and green solid lines captured by the vertical arrow. Note that the increment is larger with larger p . This is because, though the change of void per LC $\hat{o}_{LC}^{int}$ from S1-S3 to S1-S2-S3 does not depend on p , ΔC is not a linear function of $\hat{o}_{LC}^{int}$ per Eqn. (25). As p increases, $\hat{o}_{LC}^{int}$ decreases and ΔC increases. Such increase of ΔC in S1-S2-S3 is at a faster rate than the S1-S3 case. Fortunately, even with a small p , the increment is significant (see the dashed vertical arrow) as S2 is very effective in reducing the LC void and it does not depend on p . Additionally, one can notice that the increment between S2-S3 and S1-S2-S3 is not as significant when p is small (see the difference between the pink dashed (overlapping with the pink solid line) and green dashed lines. This is because S1 is not very effective in low penetration rate per R1. Therefore, one may opt for S2-S3 if adding S1 will significantly increase the implementation cost.

5.2.2. Benefits of combined control strategies vs. Baseline cases

Now we compare the impacts of the control cases in different combinations with the baseline cases. We present the results in Fig. 5.5, in which the three rows correspond to three void sizes from large to small, and the left and right columns represent small and large penetration rates. The plots show the allowed merging flow under each given mainline flow, and the distance between a curve and the diagonal reference line is the unutilized capacity. In the baseline cases, we consider the lower bound – Case B1 (in solid orange line), the upper bound – Case B2 (in solid orange line), and the upper bound with relaxation – Case B2' (in solid magenta line). Note that in the upper bound with relaxation, we assume that human-driven LC vehicles have a fixed relaxation scale with $\Delta\eta_{LC} = 0.2$. We observe the following remarks:

- R4: The combination of S1 and S3 (dashed purple line) results in discharge flow between the lower bound - Case B1 and upper bound - Case B2 as shown in Fig. 5.5 (a). This is because in Case B2 traffic upstream of the bottleneck is congested, and the voids ahead of LC vehicles are utilized to a large extent. Note that while this combination results in smaller discharge flow than Case B2, it assures that traffic is disturbance-free, and thus the overall traffic flow is more stable.
- R5: The combination of S2 and S3-gap redistribution, shown in dashed green line (S2-S3), is effective in reducing unutilized capacity, resulting in significantly higher discharge flow above the upper bound of baseline. This holds across a wide range of void size (green dashed lines in Fig. 5.5 (a), (c), and (e)) and a wide range of p (plots in the left vs. right columns in Fig. 5.5). Notably, the discharge flow of S2-S3 is also above the upper bound with relaxation (Case B2') in most cases (green dashed line above the magenta solid line, such as Fig. 5.5 (a), (b), (c)), except for the cases with very small initial void or large LC flow (e.g., $\frac{Q}{C_0} > 0.4$ in Fig. 5.5 (d)). Note that in large LC flow, while Case B2' will have smaller unutilized capacity, the mainline traffic is under frequent disturbances caused by LCs. By contrast, traffic under S2-S3 control will have a larger unutilized capacity but the mainline vehicles will be free from disturbance. Of course, the effectiveness of this combination varies with n_{bat} . In the comparison, $n_{bat} = 5$ is assumed for S2, which is feasible in most locations as it only requires that the LC segment length $L \geq 35m$.
- R6: The combination of all three strategies (blue dashed lines – S1-S2-S3) further helps increase bottleneck discharge flow, having a profound effect for high penetration rates (difference between the green and blue dashed lines in Fig. 5.5 (b)). It's worth noting that S1-S2-S3 achieves smaller unutilized capacity than Case B2' in most cases when penetration rate is medium to high (blue dashed lines above the magenta solid lines in Fig. 5.5 (b), (d), (f)). This shows the appealing potentials of the integrated strategies.
- R7: In full CAVs, extended control on the mainline helps to further increase bottleneck discharge flow; (difference between the red and blue dashed lines in Fig. 5.5 (b) representing the combination of three strategies with and without mainline control). Moreover, such extension achieves full capacity in a much wider range of LC flow. For example, even with large initial voids, it can achieve full capacity when LC flow is below $0.34C_0$ (see the red dashed line overlaps with the black reference line in Fig. 5.5 (b)). With small voids, full capacity can be achieved across all feasible range (Fig. 5.5 (f)).

5.2.3. Advantages and disadvantages of control strategies

Table 5.1 summarizes the advantages and disadvantages of the three types of combined strategies, highlighting trade-offs between the control effectiveness and the implementation efforts. For the combination of S1 and S3, the advantage is that the implementation seems relatively simple – the main task is to coordinate the inserting timing of the LCs. Also, this combination is applicable to ramps that are extremely short as it controls only one vehicle per batch. However, this combination is not very effective in reducing the unutilized capacity, especially when p is not high. Moreover, the upper bound of LC demand that it can serve is small when p is low as there are not enough CAVs to create gaps. In contrast, the effectiveness of the combination of S2 and S3 is superior, it can serve a wider range of LC demand, and it can work at much lower penetration rates (e.g., it is still quite effective when $p = 0.15$ as shown in Fig. 5.4). However, this combination requires more sophisticated implementation efforts; i.e., in addition to the mainline-ramp coordination on inserting timing, it also needs roadside assistance (e.g., ramp metering) to align LC vehicles along the wave characteristic line and control the inserting location. Besides, it requires an extended merging segment if a large LC batch is desired. The integration of the three strategies will further improve the control effectiveness over S1-S3 or S2-S3, which can be very profound if p is high. Its implementation, however, needs even more coordination efforts than S2-S3; i.e., it requires the on-ramp to have more than one lane to reposition CAVs in the head of the LC batch.

6. Discussions

Note that we made several simplifying assumptions for the purpose of developing a control framework, formulating potential

benefits, and understanding the impacts of influential factors. Future research is desired to relax these assumptions and fully capture their impacts. Specifically, we assumed a deterministic system where CAVs are distributed uniformly (according to its penetration rate), and batch insertions and mainline platoons are periodic and uniform in size. The proposed control framework should be extended in the future to address stochasticity in various system elements and design a more robust control framework. We conjecture that the control effectiveness will decrease as traffic becomes more stochastic (e.g., large variation in the distribution of CAVs and the mainline platoon size). Furthermore, we assumed that vehicles, including LC vehicles, follow Newell's simplified CF model. This would be feasible for CAVs but does not always hold for HDVs. In fact, [Chen et al. \(2014, 2012\)](#) found that HDVs can amplify disturbances when going through stop-and-go and the roadway throughput is below the theoretical capacity. This effect will affect the LC gaps needed for disturbance-free insertion (larger than the current setting). Fortunately, this effect can be accounted for in this current research framework, e.g., to scale the amplification factor and incorporate that into the LC gap formulation. Note that this effect actually emphasizes the importance of the proposed control strategies, as our control has minimized the propagation of disturbance induced by LC (disturbance is only limited to LC vehicles), which minimizes the probability of CF amplification at the origin. We believe that our integrated control strategies will still outperform the baseline cases, because Strategy 1 itself is similar to the relaxation effect and Strategy 2 essentially assures that the LC vehicles (except for the first one in the batch) do not induce additional voids.

In the control strategies proposed here, so far, we only consider one mainline lane. In the multi-lane setting, while the principles of the control still apply, thorough future investigations are needed. One potential issue is that the left lane traffic may take the gaps created in the shoulder lane for on-ramp vehicles. This is less likely a concern in free-flow traffic, where vehicles in the left lane have little incentive to change to the shoulder lane. However, if an off-ramp is located downstream close to the bottleneck, the exiting traffic could impose a problem. In such case, advisory may be beneficial to encourage exiting vehicles to change to the shoulder lane upstream of the merging segment and stay in the tail of the mainline platoons. In the case that the left lane traffic will inevitably merge into the shoulder lane gaps during the LC segment, precise control is infeasible and disturbance residuals are likely to propagate upstream. In this case, we can still control the LC vehicles from the ramp to line up along the characteristic line and insert following a certain mainline-on ramp ratio (such as alternating insertion with mainline vehicles). Nevertheless, we caution that the control for multi-lane traffic is a complex issue and requires extensive future research.

7. Conclusions

This paper has proposed control strategies for lane-changes (LCs) at a freeway merge bottleneck to improve bottleneck throughput by mitigating voids and backward-propagating speed disturbances. The control strategies are enabled by the CAV technologies, with which precise control of individual vehicles and coordinated control among vehicles would be possible. We first present three separate strategies. Strategy 1 is 'gap closure' control on each individual LC vehicle, which is controlled to accept a smaller time gap than an equilibrium gap and to exceed free-flow speed temporarily to (partially) close its void created during lane-changing. Its immediate follower in the mainline is also controlled to accept a shorter time gap temporarily to prevent the propagation of speed disturbance instigated by the LC. Strategy 2 is 'batch LC' that coordinates multiple LC vehicles and align them along kinematic waves to minimize the total voids among merging vehicles. Strategy 3 is 'gap redistribution' control in which some mainline vehicles are bunched to be at the minimum gap and others are directed to create the desired LC gaps (determined from Strategy 1 or 2) periodically to enable disturbance-free insertions.

Then we develop a general control framework integrating the three complementary strategies. Through different parameter settings, the general control framework can be reduced to individual strategies or different combinations of strategies. The integrated control can be applied to full CAV traffic or mixed traffic. The valid domains of control (within which traffic can be fully controlled) have been derived with respect to the mainline flow ratio and CAV penetration rate. It is found that the valid domain of control shrinks as the controlled vehicles have a larger tolerance level (i.e., tolerate a smaller time gap level) but increases with the LC batch size. It also increases with the size of initial LC voids. Furthermore, we have derived the maximum allowable LC flow from the ramp that maximizes bottleneck throughput and thus minimizes the unutilized capacity. We find that the unutilized capacity decreases with the tolerance level, LC batch size, and CAV penetration rate. Particularly, the compound effects of the first two can significantly increase the domain for full capacity.

We have examined the effectiveness of different control combinations and compared them with the baseline reference (without the proposed control). We use two baseline cases to obtain the upper and lower bounds of the unutilized capacity in the baseline. It is found that combination of Strategy 1 and 3 generally results in unutilized capacity between the lower and upper bound of the baseline reference. This combination is appealing as the implementation requires low level of efforts – we only need to control the LC vehicle and the immediate follower. However, this combination can serve only a small range of LC flow. The combination of Strategy 2 and 3 can be very effective in reducing unutilized capacity, generally smaller than the lower bound of the baseline. While the effectiveness depends on the LC batch size, significant benefits can be achieved for reasonable values. This combination can serve a wide range of LC flow. A main disadvantage, however, is that it requires coordination across LC vehicles through roadside assistance and coordination between LC from the ramp and mainline flow. Lastly, integration of all three strategies, and extension to mainline traffic, can help to further increase benefits in a wider range of traffic conditions (e.g., mainline flow rate and penetration rates).

Note that this paper aims to provide insights on the behavior mechanisms that can be leveraged to control CAVs and thus improve traffic flow. There are many issues left for future research. One important element to be studied is the impact of stochasticity in traffic on the control strategies proposed. Our formulations make simplifying assumptions on periodic distribution of platoons and on uniform distribution of CAVs within the stream. Stochasticity in traffic can undermine the performance of the proposed control strategies quantitatively, and lead to implementation difficulties. A thorough analysis of the impact of stochasticity would need to re-evaluate the

throughput in the base no-control, and in various control combination scenarios to estimate benefits more accurately. Since this is a complex analysis, we leave this outside the scope of the current work.

Another issue is that long platoons may exist in the mainline, which can occur due to (1) high flow in the mainline, (2) low CAV penetration rate, or (3) stochastic distribution of CAVs. In the former two cases, the occurrence of long platoons will be common, while long-platoons will be stochastic in case (3). If there are frequent long platoons in the mainline (due to case (1) or (2)), some LC vehicles from the ramp may experience excessive delay. From the perspective of system delay, S3 helps to reduce the overall delay. From the user perspective, admittedly, the delay may not distribute equally between the mainline and ramp vehicles under S3 and that may cause an equity problem. The severity of this problem varies with the platoon length and frequency of long platoons. Some remedies can be applied: (i) to reduce the LC batch size (e.g., to the minimum, 1) so that the gap needed in the mainline is smaller and thus the platoons created will be shorter; (ii) in multi-lane highway, to encourage mainline vehicles to stay on the left lanes unless they are close to their destination exits, which will help to reduce the flow of the shoulder lane; and (iii) to encourage CAVs to travel on the shoulder lane to increase the penetration rate on that lane. Note that such trade-off between system improvement and potential uneven benefit/cost distribution is a common problem pertaining to traffic control. A full-scale investigation to this problem is beyond the scope of this paper and is left for future research.

CRedit authorship contribution statement

Danjue Chen: Conceptualization, Methodology, Writing - original draft. **Anupam Srivastava:** Methodology. **Soyoung Ahn:** Methodology, Writing - review & editing, Supervision, Funding acquisition. .

Declaration of Competing Interest

The authors declared that there is no conflict of interest.

Acknowledgement

This research was sponsored by the National Science Foundation through Awards CMMI 1536599, 1932932, 1932921, and 1944369.

Appendix A

We show that that in most conditions, vehicles can fully eliminate the void; i.e.,

$$\Delta o = o_{ex}$$

Δo depends on the LC vehicle's void closing capability, which is determined by v_{max} and T_{tl} (T_{tl} implicitly decides the available cruising time T_{cru} ⁷). The balance of these two elements (o_{ex} and Δo) decides the residual void o_{rs} .

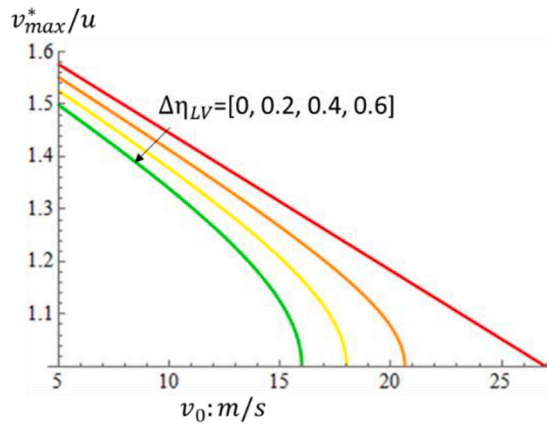


Fig. A1. v_{max}^* plot.

⁷ Here we assume that v_{max} is achievable within T_{tl} to a meaningful physical interpretation; i.e., $\frac{v_{max}-u}{a} + \frac{v_{max}-u}{b} \leq T_{tl}$. If not, we can simply replace v_{max} with the maximum possible speed within T_{tl} , which equals to $u + \sqrt{\frac{T_{tl}}{a+b}}$.

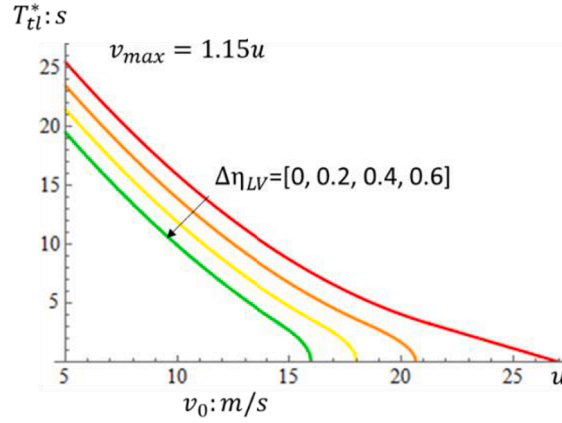


Fig. A2. T_{il}^* plot with $v_{\max} = 115\%u$.

If $v_{\max}$ is sufficiently large to exceed an optimal speed $v_{\max}^*$, the cruising process is not needed, and we don't need to use the full scale of T_{il} . Specifically, the LC vehicle can accelerate at its maximum rate to an optimal speed, $v_{\max}^*$ ($v_{\max}^* \leq v_{\max}$), and then decelerate at the maximum rate to recover u . Accordingly, $v_{\max}^*$ can be derived by applying void formulation (similar to Eqn. (1) in the manuscript) into Eqn. (6) in the manuscript:

$$\frac{(v_{\max}^* - u)^2}{2ua} + \frac{(v_{\max}^* - u)^2}{2ub} = o_{\text{ex}}.$$

$$v_{\max}^* = u + \sqrt{\frac{2uo_{\text{ex}}}{\frac{1}{a} + \frac{1}{b}}} = u + \frac{\sqrt{(a+b)b[(u-v_0)^2 - 2a\Delta\eta_{LV}h_0]}}{a+b}$$

The actual duration of exceeding u , T_{il}^a , can be obtained in a straightforward way ($T_{il}^a = \frac{v_{\max}^* - u}{a} + \frac{v_{\max}^* - u}{b}$). Fig. A1 below shows the values of $v_{\max}^*$ under typical parameter settings. One can see that $v_{\max}^*$ decreases with v_0 and $\Delta\eta_{LV}$. Specifically, if v_0 is above 15m/s, and $\Delta\eta_{LV} \geq 0.25$, $v_{\max}^*$ will be below 125% of u , and T_{il}^a will be no more than 5.5 s. Of course, it's obvious that $v_{\max}^*$ decreases with a and b .

In case that $v_{\max} \leq v_{\max}^*$, the full scale of T_{il} will be used to close the void (thus, $T_{\text{cru}} = T_{il} - \frac{v_{\max} - u}{a} - \frac{v_{\max} - u}{b}$). In this case, there exists a threshold value of T_{il} , T_{il}^* , below which it is impossible to fully eliminate the void and there will be a residual void, o_{rs} . The size of o_{rs} can be obtained by integrating T_{cru} into Eqn. (6) and then into Eqn. (9) in the manuscript; see the elaborate form in the manuscript. T_{il}^* can be derived by setting $o_{rs} = 0$ and solve the equations.

$$o_{rs} = \frac{(u-v_0)^2}{2ua} - \Delta\eta_{LV}h_0 - \left[T_{il} \left(\frac{v_{\max}}{u} - 1 \right) - \frac{(v_{\max} - u)^2}{2ua} - \frac{(v_{\max} - u)^2}{2ub} \right]$$

$$T_{il}^* = \left[\frac{(u-v_0)^2}{2ua} - \Delta\eta_{LV}h_0 + \frac{(v_{\max} - u)^2}{2ua} + \frac{(v_{\max} - u)^2}{2ub} \right] \frac{u}{v_{\max} - u}$$

Fig. A2 shows the T_{il}^* values with a conservative $v_{\max}$ setting ($v_{\max} = 115\%u$). One can see that T_{il}^* is below 25 s as long as v_0 is above 5m/s, which seems quite acceptable. Therefore, the results here suggest that it is likely that vehicles can fully eliminate the void in most typical conditions.

Appendix B

The formulation for $\hat{o}_{rs}$ (Eqn. (17)) needs a slight modification when $p < 1/n_{\text{bat}}$ and $o/h_0 < \Delta\eta_{LV}$; i.e., when the penetration rate is low, the full LC void o is small, and $\Delta\eta_{LV}$ is large. Specifically, solving these two conditions, it boils down to

$$\begin{cases} v_0 \geq v_0^{\text{crit}} = u \left(1 - \sqrt{\frac{2a}{u} \Delta\eta_{LV} h_0} \right), \\ p < 1/n_{\text{bat}} \end{cases},$$

where v_0^{crit} denotes the critical value. With typical numerical values such as $u = 30\text{m/s}$, $a = 2\text{m/s}^2$, $h_0 = 1.8\text{s}$, $\Delta\eta_{LV} = 0.5$, the above condition requires that $v_0 \geq 190.61\text{m/s}$ (43.86mph), which does not hold in most cases as the on-ramp speed limit is usually below 40mph. Nevertheless, in case that $v_0 \geq v_0^{\text{crit}}$ holds, the $\hat{o}_{rs}$ in Eqn. (17) should be updated to the following:

$$\hat{o}_{rs} = o(1 - p).$$

As the LC batches with CAV will not have any void and those without CAV will have a full void o . Of course, the valid control domain and the unutilized capacity should be updated accordingly using the above updated formulation for $\hat{o}_{rs}$. We omit the detailed forms here as such changes do not change the main insights derived in the main text.

References

- Ahn, S., Cassidy, M.J., 2007. Freeway traffic oscillations and vehicle lane-change maneuvers. Elsevier, New York, London, pp. 691–710.
- Cassidy, M.J., Rudjanakanoknad, J., 2005. Increasing the capacity of an isolated merge by metering its on-ramp. *Transp. Res. Part B Methodol.* 39, 896–913. <https://doi.org/10.1016/j.trb.2004.12.001>.
- Chen, D., Ahn, S., 2018. Capacity-drop at Extended Bottlenecks: Merge, Diverge, and Weave. *Transp. Res. Part B Methodol.* p. 108.
- Chen, D., Ahn, S., Laval, J., Zheng, Z., 2014. On the periodicity of traffic oscillations and capacity drop: The role of driver characteristics. *Transp. Res. Part B Methodol.* 59 <https://doi.org/10.1016/j.trb.2013.11.005>.
- Chen, D., Laval, J., Zheng, Z., Ahn, S., 2012. A behavioral car-following model that captures traffic oscillations. *Transp. Res. Part B Methodol.* 46 <https://doi.org/10.1016/j.trb.2012.01.009>.
- Duret, A., Bouffier, J., Buisson, C., 2011. Onset of Congestion from Low-Speed Merging Maneuvers Within Free-Flow Traffic Stream. *Transp. Res. Rec. J. Transp. Res. Board* 2188, 96–107. <https://doi.org/10.3141/2188-11>.
- Duret, A., Wang, M., Ladino, A., 2020. A hierarchical approach for splitting truck platoons near network discontinuities. *Transp. Res. Part B Methodol.* 132, 285–302. <https://doi.org/10.1016/j.trb.2019.04.006>.
- Elefteriadou, L., Roess, R.P., McShane, W.R., 2005. Probabilistic nature of breakdown at freeway merge junctions. *Transp. Res. Rec. Rec.* 80–89.
- He, H., Guler, S.I., Menendez, M., 2016. Adaptive control algorithm to provide bus priority with a pre-signal. *Transp. Res. Part C* 64, 28–44. <https://doi.org/10.1016/j.trc.2016.01.009>.
- Jin, W.-L., 2010. A kinematic wave theory of lane-changing traffic flow. *Transp. Res. Part B Methodol.* 44, 1001–1021. <https://doi.org/10.1016/j.trb.2009.12.014>.
- Jin, W.-L., Laval, J., 2018. Bounded acceleration traffic flow models: A unified approach. *Transp. Res. Part B Methodol.* 111, 1–18. <https://doi.org/10.1016/j.trb.2018.03.006>.
- Karimi, M., Roncoli, C., Alecsandru, C., Papageorgiou, M., 2020a. Merging in a mixed-traffic of connected and automated vehicles and conventional vehicles: A model predictive control framework. *Transportation Research Board Annual Meeting*.
- Karimi, M., Roncoli, C., Alecsandru, C., Papageorgiou, M., 2020b. Cooperative merging control via trajectory optimization in mixed vehicular traffic. *Transp. Res. Part C Emerg. Technol.* 116, 102663 <https://doi.org/10.1016/j.trc.2020.102663>.
- Laval, J.A., 2009. Effects of geometric design on freeway capacity: Impacts of truck lane restrictions. *Transp. Res. Part B Methodol.* 43, 720–728. <https://doi.org/10.1016/j.trb.2009.01.003>.
- Laval, J.A., Daganzo, C.F., 2006. Lane-changing in traffic streams. *Transp. Res. Part B Methodol.* 40, 251–264. <https://doi.org/10.1016/j.trb.2005.04.003>.
- Laval, J.A., Leclercq, L., 2008. Microscopic modeling of the relaxation phenomenon using a macroscopic lane-changing model. *Transp. Res. Part B Methodol.* 42, 511–522. <https://doi.org/10.1016/j.trb.2007.10.004>.
- Leclercq, L., Chiabaut, N., Laval, J., Buisson, C., 2007. Relaxation phenomenon after lane changing: experimental validation with NGSIM data set. *Transp. Res. Rec. J. Transp. Res. Board* 1999, 79–85. <https://doi.org/10.3141/1999-09>.
- Leclercq, L., Knoop, V.L., Marcjak, F., Hoogendoorn, S.P., 2016. Capacity drops at merges: New analytical investigations. *Transp. Res. Part C* 62, 171–181. <https://doi.org/10.1109/ITSC.2014.6957839>.
- Leclercq, L., Laval, J.A., Chiabaut, N., 2011. Capacity drops at merges: an endogenous model. 19th Int. Symp. Transp. Traffic Theory 17, 12–26. <https://doi.org/10.1016/j.sbspro.2011.04.505>.
- Letter, C., Elefteriadou, L., 2017. Efficient control of fully automated connected vehicles at freeway merge segments. *Transp. Res. Part C Emerg. Technol.* 80, 190–205. <https://doi.org/10.1016/j.trc.2017.04.015>.
- Lighthill, M.J., Whitham, G.B., 1955. On kinematic waves. I. flood movement in long rivers. *Proc. R. Soc. A Math. Phys. Eng. Sci.* 229, 281–316. <https://doi.org/10.1098/rspa.1955.0088>.
- Marcjak, F., Buisson, C., 2014. Analytical derivation of capacity at diverging junctions. *Transp. Res. Rec. J. Transp. Res. Board* 2422, 88–95. <https://doi.org/10.3141/2422-10>.
- Marcjak, F., Leclercq, L., Buisson, C., 2015. A Macroscopic Model for Freeway Weaving Sections. *Comput. Civ. Infrastruct. Eng.* 30, 464–477. <https://doi.org/10.1111/mice.12119>.
- Mauch, M., Cassidy, M.J., 2002. Freeway traffic oscillations: observations and predictions. Emerald Group Publishing Limited, Adelaide, pp. 653–674, 10.1108/9780585474601-032.
- Newell, G.F., 2002. A simplified car-following theory: a lower order model. *Transp. Res. Part B Methodol.* 36, 195–205. [https://doi.org/10.1016/S0191-2615\(00\)00044-8](https://doi.org/10.1016/S0191-2615(00)00044-8).
- Park, H., Bhamidipati, C., Smith, B., 2011. Development and Evaluation of Enhanced IntelliDrive-Enabled Lane Changing Advisory Algorithm to Address Freeway Merge Conflict. *Transp. Res. Rec. J. Transp. Res. Board* 2243, 146–157. <https://doi.org/10.3141/2243-17>.
- Pueboobpaphan, R., Liu, F., van Arem, B., 2010. The impacts of a communication based merging assistant on traffic flows of manual and equipped vehicles at an on-ramp using traffic flow simulation. 13th International IEEE Conference on Intelligent Transportation Systems 1468–1473. <https://doi.org/10.1109/ITSC.2010.5625245>.
- Punzo, V., Simonelli, F., 2005. Analysis and Comparison of Microscopic Traffic Flow Models with Real Traffic Microscopic Data. *Transp. Res. Rec. J. Transp. Res. Board* 1934, 53–63. <https://doi.org/10.3141/1934-06>.
- Richards, P.L., 1956. Shock waves on the highway. *Oper. Res.* 4, 42–51. <https://doi.org/10.2307/167515>.
- Rios-Torres, J., Malikopoulos, A., Pisu, P., 2015. Online Optimal Control of Connected Vehicles for Efficient Traffic Flow at Merging Roads. 2015 IEEE 18th International Conference on Intelligent Transportation Systems 2432–2437. <https://doi.org/10.1109/ITSC.2015.392>.
- Rios-Torres, J., Malikopoulos, A.A., 2016. A survey on the coordination of connected and automated vehicles at intersections and merging at highway on-ramps. *IEEE Trans. Intell. Transp. Syst.* 18 (5), 1066–1077.
- Scarinci, R., Heydecker, B., 2014. Control concepts for facilitating motorway on-ramp merging using intelligent vehicles. *Transport reviews* 34 (6), 775–797.
- Scarinci, R., Heydecker, B., Hegyi, A., 2013. Analysis of traffic performance of a ramp metering strategy using cooperative vehicles, in: 16th International IEEE Conference on Intelligent Transportation Systems (ITSC 2013). pp. 324–329. doi:10.1109/ITSC.2013.6728252.
- Serafin, C., 1996. Driver preferences and usability of adjustable distance controls for an Adaptive Cruise Control (ACC) system.
- Shladover, S.E., 2017. Connected and automated vehicle systems: Introduction and overview. *J. Intell. Transp. Syst.* 1–11 <https://doi.org/10.1080/15472450.2017.1336053>.
- Shladover, S.E., Su, D., Lu, X.-Y., 2012. Impacts of Cooperative Adaptive Cruise Control on Freeway Traffic Flow. *Transp. Res. Rec. J. Transp. Res. Board* 2324, 63–70. <https://doi.org/10.3141/2324-08>.
- Sun, J., Zhao, L., America, M., Zhang, H.M., 2014. The Mechanism of Early-onset Breakdown at Shanghai 's Expressway On-ramp Bottlenecks. *Transportation Research Board 93rd Annual Meeting Annual Meeting*.

Treiber, M., Kesting, A., 2013. *Traffic flow dynamics*. Springer.

Zheng, Z., Ahn, S., Chen, D., Laval, J., 2013. The effects of lane-changing on the immediate follower: Anticipation, relaxation, and change in driver characteristics.

Transp. Res. Part C Emerg. Technol. 26, 367–379. <https://doi.org/10.1016/j.trc.2012.10.007>.

Zheng, Z., Ahn, S., Chen, D., Laval, J., 2011a. Freeway traffic oscillations: Microscopic analysis of formations and propagations using Wavelet Transform. Transp. Res. Part B Methodol. 45, 702–716. <https://doi.org/10.1016/j.sbspro.2011.04.540>.

Zheng, Z., Ahn, S., Chen, D., Laval, J., 2011b. Applications of wavelet transform for analysis of freeway traffic: Bottlenecks, transient traffic, and traffic oscillations. Transp. Res. Part B Methodol. 45, 372–384. <https://doi.org/10.1016/j.trb.2010.08.002>.