Transportation Research Part B 208 (2026) 103447

Contents lists available at ScienceDirect

Transportation Research Part B

journal homepage: www.elsevier.com/locate/trb

Unveiling traffic capacity in the mixed HV and CAV environment: A theoretical approach with CAV clustering intensity

Tongfei Li [a], Hongfei Zhu [a], Min Xu [b,*], Huijun Sun [c]

[a] Beijing Key Laboratory of Traffic Engineering, Beijing University of Technology, Beijing, 100124, China
[b] Department of Industrial and Systems Engineering, The Hong Kong Polytechnic University, Hung Hom, Hong Kong, China
[c] School of System Science, Beijing Jiaotong University, Beijing, 100044, China

ARTICLE INFO

Keywords:
Connected and autonomous vehicle
Mixed lane capacity
Clustering intensity
Mixed traffic flow

ABSTRACT

By integrating advanced communication and autonomous driving technologies, connected and autonomous vehicles (CAVs) are expected to have a shorter reaction time and drive in platoons to achieve a more compact safe time headway, and thus exhibit the great potential to increase road capacity. However, human-driven vehicles (HVs) and CAVs will coexist and share urban traffic infrastructure for a long transitional time, and the platoon size of CAVs is limited by technological constraints and safety concerns. These factors complicate the calculation and theoretical analysis of mixed traffic capacity. Unlike the fixed constant in a pure HV environment, traffic capacity in the mixed HV and CAV environment is a complex stochastic variable influenced by various factors, such as the CAV penetration rate, the maximum platoon size of CAVs, and the spatial distribution of heterogeneous vehicles in the mixed traffic flow. To provide a generalized methodology for calculating and analyzing mixed lane capacity, we propose an index named CAV clustering intensity. This index has a clear physical meaning and explicit expression, allowing for the quantitative characterization of the longitudinal distribution of heterogeneous vehicles without any simplified assumptions regarding car-following patterns and the spatial distribution of mixed traffic flow. Accordingly, a Markov chain model is established to theoretically derive an explicit expression for calculating mixed lane capacity, which is mathematically expressed as a multivariate function of CAV penetration rate, maximum platoon size, and CAV clustering intensity. The effect of all these factors on mixed lane capacity is explored by rigorously theoretical analysis. Then, the problem of calculating the upper and lower bounds of lane capacity in the yet-to-be-realized mixed traffic environment is equivalently transformed into a vehicle arrangement problem. Two linear programming models are established and theoretically derive the explicit mathematical expression of capacity bounds and the corresponding spatial distribution of mixed traffic flow in scenarios with unlimited and limited platoon sizes. Finally, the numerical characteristics (e.g., mean, variance, probability distribution) of stochastic lane capacity are also calculated to provide a more comprehensive understanding of its stochastic properties.

1. Introduction

Compared with human-driven vehicles (HVs), connected and automated vehicles (CAVs) exhibit great potential to improve road capacity as they are equipped with advanced autonomous driving and Internet of Vehicles technologies. Autonomous driving

* Corresponding author.
E-mail addresses: tfli@bjut.edu.cn (T. Li), hfzhu@emails.bjut.edu.cn (H. Zhu), min.m.xu@polyu.edu.hk (M. Xu), hjsun1@bjtu.edu.cn (H. Sun).

https://doi.org/10.1016/j.trb.2026.103447
Received 13 December 2024; Received in revised form 4 March 2026; Accepted 15 March 2026
Available online 23 March 2026

technology based on artificial intelligence endows CAVs with a shorter reaction time than HVs, and thus, they can safely drive closer to the preceding vehicle theoretically. Through real-time communication and cooperation, CAVs are expected to achieve a more compact safe time headway between each other when driving in the form of platoons. However, it is important to note that the platoon size of CAVs cannot be infinite due to the limitations in V2V communication range, string stability, and safety. Once the maximum platoon size is reached, any following CAV is prohibited from joining the existing platoon and instead initiates a new platoon as its leader (Du et al., 2025; Ren et al., 2024; Yao et al., 2023; Yu et al., 2023; Zhao et al., 2025). Moreover, HVs cannot be entirely substituted by CAVs in the short term. For a long transitional time, HVs and CAVs will coexist in the traffic system and share urban road resources.

The mixed HV and CAV environment greatly complicates the calculation and analysis of lane capacity. Different from the fixed value of the lane capacity in the pure HV environment, the lane capacity in the mixed HV and CAV environment is a stochastic variable affected by various factors. Specifically, due to the difference in technical characteristics of heterogeneous vehicles and the limited platoon size of CAVs, the safe time headway between two adjacent vehicles is related to the type of preceding and succeeding vehicles and the succeeding vehicle's driving pattern (i.e., driving alone or driving in the form of CAV platoons), namely car-following patterns. Since the average safe time headway directly determines traffic density and throughput, the lane capacity in the mixed traffic scenario is a variable related to the longitudinal distribution of two types of vehicles and the value of safe time headways corresponding to different car-following patterns. On the one hand, the sequence arrangement of heterogeneous vehicles on the longitudinal lane is inherently uncertain. As a qualitative concept, the spatial distribution of heterogeneous vehicles is difficult to quantify by a single index, and their sequence arrangement is affected by CAV penetration rate, maximum platoon size, etc. All the issues above make it difficult to theoretically derive the explicit expression for calculating mixed lane capacity and analyze its stochastic characteristics. On the other hand, the safe time headway corresponding to each car-following pattern is affected by many factors (e.g., the development level of CAV technology, and the level of public acceptance), both physical and psychological. In a yet-to-be-realized scenario with mixed HVs and CAVs, the exact value of safe time headways corresponding to different car-following patterns has not reached a consensus both in academia and in industry, and completely opposite conclusions may be derived based on different settings of safe time headways (Ghiasi et al., 2017; Zheng et al., 2023; Fan et al., 2025). To summarize, it remains a challenge to reveal analytical insights into the theoretical modeling and analysis of the mixed lane capacity, which significantly restricts the decision-making and implementation of traffic planning and management strategies in the mixed traffic scenario. Therefore, this study aims to answer the following questions:

- Is there an index with the explicit mathematical expression that can be used to quantify the spatial distribution of heterogeneous vehicles on the longitudinal lane? How do different car-following patterns and the CAV platoon with different sizes distribute on the longitudinal lane?
- Without any simplified assumption on car-following patterns and the value of safe time headways, how to theoretically derive the explicit expression for calculating lane capacity in the mixed HV and CAV environment? Moreover, how does the mixed lane capacity change with the variation of CAV penetration rate, maximum platoon size, and spatial distribution of heterogeneous vehicles?
- As a stochastic variable, what are the exact upper and lower bounds of the mixed lane capacity under any setting of related parameters (e.g., safe time headways, CAV penetration rate, maximum platoon size)? At that time, how are HVs and CAVs distributed on the longitudinal lane?
- To further reveal the stochastic properties of the mixed lane capacity, what are its mean, variance, and probability distribution? Moreover, how do these numerical characteristics change with the variation of the CAV penetration rate?

To address the above questions, after classifying different car-following patterns and recognizing corresponding safe time headways in the mixed HV and CAV environment, an index is proposed to characterize the longitudinal distribution of heterogeneous vehicles. Accordingly, the explicit expression for calculating mixed lane capacity is theoretically derived. The monotonicity of mixed lane capacity with various factors is rigorously examined based on the method of theoretical analysis and verified by simulation experiments. As a stochastic variable, two linear programming problems are established to theoretically derive the exact upper and lower bounds of mixed lane capacity and the corresponding spatial distribution of mixed traffic flow. Finally, under different technological scenarios, various numerical characteristics of the stochastic lane capacity are calculated for a comprehensive understanding of its stochastic properties.

1.1. Literature review

Based on different simplified assumptions of car-following patterns, safe time headways, and spatial distribution of the mixed traffic flow, a wide spectrum of research seeks to derive the calculation expression of the mixed traffic capacity through theoretical analysis and modeling (Levin and Boyles, 2016; Chen et al., 2017; Liu and Song, 2019; Yao et al., 2023; Ghiasi et al., 2017; Zhou and Zhu, 2021; Guan et al., 2023; Jiang et al., 2023). Under the assumption that the reaction time required for vehicle deceleration and braking is only related to the vehicle type itself, Levin and Boyles (2016) derived a calculation expression for mixed traffic capacity and found that the mixed traffic capacity is a function only related to the CAV penetration rate. Because the value of safe time headway is generally related to the type of succeeding and preceding vehicles rather than the type of succeeding vehicle alone, the traffic capacity is influenced by the spatial distribution of different types of vehicles in the mixed traffic flow. For simplicity, several different assumptions related to the longitudinal distribution of heterogeneous vehicles on the lane were proposed and adopted in the studies on mixed traffic capacity. By assuming that all CAVs appear periodically on the lane in the form of fixed-size platoons, Chen

et al. (2017) obtained the average headway per platoon cycle and thereby derived a calculation formula for the lane capacity in the mixed traffic environment. Under the assumption that the distribution of all vehicles in the mixed traffic flow is completely random, Liu and Song (2019) derived the formula for calculating the expected capacity of mixed lanes after regarding vehicle types along the longitudinal lane as a Bernoulli process, and Yao et al. (2023) extended it to the mixed scenario of four types of vehicles. The above assumptions and the corresponding expected value of the mixed traffic capacity function are widely adopted by existing literature on transportation network modeling in a mixed HV and CAV environment (Chen et al., 2026; Liu et al., 2021; Li et al., 2023, 2024; Liu and Song, 2019; Wang et al., 2019, 2021; Zhang et al., 2022, 2023; Zhou et al., 2024).

To derive the calculation expression of mixed traffic capacity in a more general scenario, Ghiasi et al. (2017) creatively proposed the concept of CAV platooning intensity to describe the aggregation degree of CAVs on longitudinal lanes in mixed traffic environments. This innovative index made it possible to formulate the spatial distribution of heterogeneous vehicles into the analysis of mixed traffic capacity. Without any simplified assumption of the spatial distribution of heterogeneous vehicles, a Markov chain model was established based on an exogenously given transition matrix and used to derive an approximation of the mixed traffic lane capacity as a function of CAV penetration rate and CAV platooning intensity. Moreover, they give a data-driven method for calibrating the proposed CAV platooning intensity. Later, Zhou and Zhu (2021) further extended the theoretical derivation of the calculation expression of mixed traffic capacity to a more general scenario by considering the maximum platoon size of CAVs. In a mixed semi-automatic and fully automatic traffic scenario, Guan et al. (2023) investigated the heterogeneous platooning effect of the two types of vehicles on the mixed traffic capacity. Recently, Jiang et al. (2023) provided a quantitative calculation formula for CAV platooning intensity, which was expressed as a multivariate function of the CAV penetration rate and the numbers of platoons with different sizes. Accordingly, they derived the calculation expression of mixed traffic capacity based on the assumption that all vehicles are completely randomly distributed in the mixed traffic flow.

Although the research community has recognized that mixed traffic capacity is a stochastic variable related to the spatial distribution of heterogeneous vehicles in the mixed traffic flow, studies on the stochastic character of mixed traffic capacity are relatively scarce in the literature (Mohajerpoor and Ramezani, 2019; Chen et al., 2022; Wang et al., 2022). By assuming safe time headways for all car-following patterns satisfy a given simplified relationship, Mohajerpoor and Ramezani (2019) and Wang et al. (2022) analyzed the longitudinal distribution of HVs and CAVs in the mixed traffic flow when the value of capacity reaches the maximum and minimum. Specifically, the mixed traffic capacity reaches the maximum value when all CAVs cluster together, while it reaches the minimum value when all CAVs are dispersed as much as possible across the lane. Accordingly, they obtained the calculation expressions for the maximum and minimum values of mixed lane capacity. Similarly, based on a given simplified relationship among different safe time headways corresponding to each car-following pattern, Chen et al. (2022) derived the lower and upper bound functions of the lane capacity in the mixed AVs and HVs environment. Then, based on the assumption that all vehicles are completely randomly distributed on the lane, they explored the probability distribution of the mixed lane capacity based on the simulation via probability distribution fitting. However, it is noted that, for convenience in analyzing the spatial distribution of heterogeneous vehicles when capacity reaches the extreme value, all the studies above simplified the classification of car-following patterns and the corresponding safe time headways by assuming that CAV platoons are not limited by the maximum platoon size. Finally, a detailed comparison between this study and the existing literature is listed in Table 1.

1.2. Research contributions and paper outline

This study aims to provide a generalized methodology for the calculation and analysis of traffic capacity in the mixed HV and CAV environment. The specific contributions are summarized as follows:

First, as shown in Table 1, few studies consider the limited platoon size of CAVs when calculating mixed traffic capacity, despite it being a critical factor. The only research that does so also relies on numerous simplified assumptions. Specifically, it is acknowledged that mixed lane capacity is directly related to the spatial distribution of heterogeneous vehicles, which complicates its calculation and analysis. To address this issue, two simplified methods are widely adopted: assuming that heterogeneous vehicles follow an ideal distribution (either fixed or completely random) and deriving the mixed lane capacity formula based on an exogenously given transition matrix and an index obtained by calibration (i.e., CAV platooning intensity). It may constitute a challenge to rigorously analyze the relationship between lane capacity and the spatial distribution of heterogeneous vehicles. To address this research gap, an index named CAV clustering intensity is proposed. This index, which has a clear physical meaning and explicit expression, quantitatively characterizes the longitudinal distribution of heterogeneous vehicles without relying on simplified assumptions about car-following patterns and spatial distribution in mixed traffic flow. It ensures that the spatial distribution of heterogeneous vehicles with the same CAV clustering intensity corresponds to the same lane capacity. Accordingly, a Markov chain model is established to theoretically derive the explicit expression for calculating mixed lane capacity, which is mathematically expressed as a multivariate function of CAV penetration rate, maximum platoon size, and CAV clustering intensity. During the model formulation, the transition probability is rigorously derived rather than exogenously given, as in previous studies. Furthermore, the effects of these factors on mixed lane capacity are explored by theoretically analyzing the monotonicity of the mixed lane capacity function with respect to different independent variables.

Second, since mixed lane capacity is a stochastic variable, calculating and analyzing its upper and lower bounds will contribute to the decision-making of transportation planning and management strategies in the mixed traffic environment. However, few studies have explored this issue, and those that have are based on strong assumptions regarding safe time headway values. As a yet-to-be-realized traffic scenario, there is no consensus in academia or industry on the exact values of various safe time headways, rendering existing methodologies inapplicable to future complex scenarios. To address the above issue, without drawing any assumption on

Table 1
A comparison between this study and the existing related literature.

References	Classification criteria of headways	Relationship among different headways	Limited platoon size	Vehicle distribution assumption	Vehicle distribution index	Characteristics of capacity
Levin and Boyles (2016)	Type of the succeeding vehicle	Aggressive	×	×	×	Capacity function
Chen et al. (2017)	Type of succeeding and preceding vehicles	Aggressive	✓	A fixed periodic distribution	×	Capacity function
Ghiasi et al. (2017)	Type of succeeding and preceding vehicles	Aggressive, moderate, conservative	×	×	CAV platooning intensity obtained by calibration	Capacity function
Bujanovic and Lochrane (2018)	Type of succeeding and preceding vehicles, CAV driving pattern	Aggressive	✓	Completely random	×	Capacity function
Liu and Song (2019)	Type of succeeding and preceding vehicles	Aggressive	×	Completely random	×	Capacity function
Mohajerpoor and Ramezani (2019)	Type of succeeding and preceding vehicles	Aggressive	×	Completely random	×	Capacity function, upper and lower bounds
Zhou and Zhu (2020)	Type of succeeding and preceding vehicles	Aggressive	×	×	CAV platooning intensity obtained by calibration	Capacity function
Sala and Soriguera (2021)	Type of the succeeding vehicle, CAV driving pattern	Aggressive	✓	Completely random	×	Capacity function
Zhou and Zhu (2021)	Type of succeeding and preceding vehicles, CAV driving pattern	Aggressive	✓	×	CAV platooning intensity obtained by calibration	Capacity function
Chen et al. (2022)	Type of succeeding and preceding vehicles	Aggressive	×	Completely random	×	Upper and lower bounds, mean, and probability distributions
Wang et al. (2022)	Type of succeeding and preceding vehicles	Aggressive	×	Completely random	×	Capacity function, upper and lower bounds

the value and relationship among safe time headways corresponding to different car-following patterns, the problem of calculating the upper and lower bounds of lane capacity in the mixed traffic environment is equivalently transformed into the arrangement problem of two types of vehicles on a single-lane ring road. In scenarios with unlimited and limited platoon sizes, two linear programming models are developed to formulate the proposed vehicle arrangement problems and theoretically derive the upper and lower bounds of mixed lane capacity, along with the corresponding spatial distribution of mixed traffic flow. The proposed method is validated in three technological scenarios distinguished by the values of different safe time headways (i.e., aggressive, moderate, and conservative), which represent different development stages of CAV technology from birth to maturity. Besides, under different technological scenarios, various numerical characteristics (e.g., mean, variance, probability distribution) of mixed lane capacity are calculated based on simulation experiments, which help us to comprehensively understand the stochastic properties of mixed lane capacity.

The rest of the paper is organized as follows. Section 2 introduces an index quantitatively characterizing the spatial distribution of heterogeneous vehicles and derives the explicit expression for calculating and analyzing the mixed lane capacity. Section 3 calculates and analyzes the mixed capacity bounds. Section 4 validates the theoretical analysis and analyzes the numerical characteristics of the stochastic capacity. Section 5 gives the conclusion and discusses future research directions.

Table 1
A comparison between this study and the existing related literature.(Continued)

Jiang et al. (2023)	Type of succeeding and preceding vehicles, CAV driving pattern	Aggressive	✓	Completely random	CAV platooning intensity calculated by an explicit expression	Capacity function
Li et al. (2023, 2024)	Type of the succeeding and preceding vehicles	Aggressive	×	Completely random	×	Capacity function
Yao et al. (2023, 2024)	Type of the succeeding vehicle, CAV driving pattern	Aggressive	✓	Completely random	×	Capacity function
This study	Type of succeeding and preceding vehicles, CAV driving pattern	Model: any relationship; Experiment: aggressive, moderate, conservative	✓	×	CAV clustering intensity calculated by an explicit expression	Capacity function, upper and lower bounds, mean, variance, and probability distribution

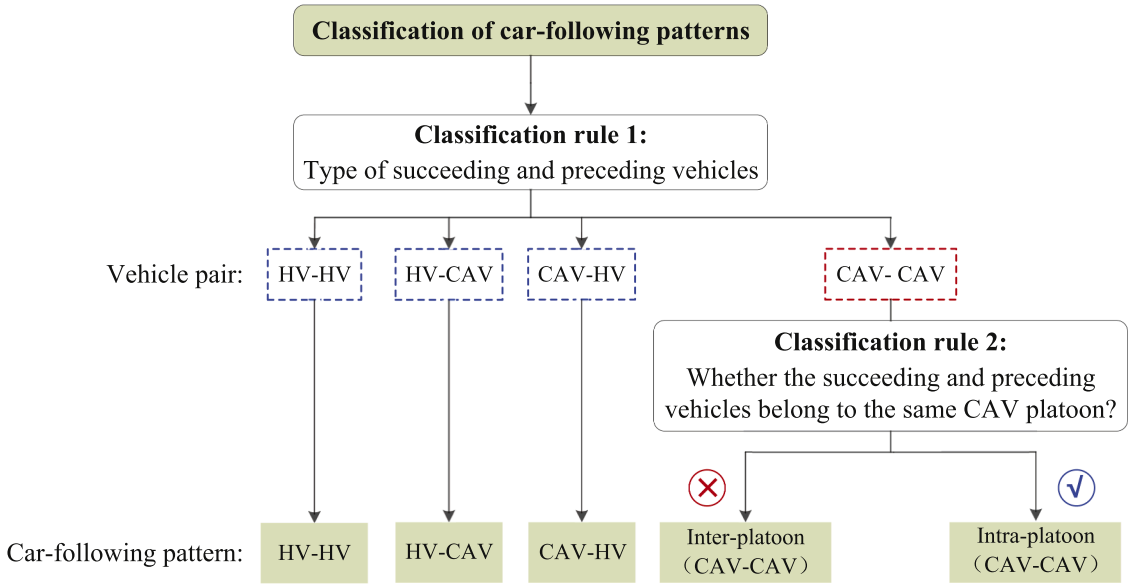


Fig. 1. Different types of car-following patterns and vehicle pairs in the mixed traffic flow.

2. Modeling of the mixed HV and CAV traffic

2.1. Car-following patterns and vehicle pairs in the mixed HV and CAV environment

Due to the different technical characteristics of HVs and CAVs, we can classify car-following patterns in the mixed traffic flow based on a combination of two rules, as shown in Fig. 1. Specifically, four types of vehicle pairs can be first recognized according to the type of succeeding and preceding vehicles, including the HV-HV vehicle pair (i.e., an HV following an HV), the HV-CAV vehicle pair (an HV following a CAV), the CAV-HV vehicle pair (i.e., a CAV following an HV), and the CAV-CAV vehicle pair (i.e., a CAV following a CAV). The first three types of vehicle pairs correspond to three types of car-following patterns, respectively. Because CAVs can form limited-sized platoons with each other, two adjacent CAVs may drive in different patterns, i.e., driving alone with each other or driving in the form of CAV platoons. Thus, the vehicle pair that a CAV following a CAV can be further classified into two categories according to whether the succeeding and preceding vehicles belong to the same CAV platoon: a CAV following a CAV platoon that reaches the maximum allowable platoon size (a maximum-size CAV platoon or a saturated CAV platoon for short), or a CAV following a CAV within the same CAV platoon. Generally, the succeeding vehicle in different car-following patterns maintains a different safe time headway with the preceding vehicle during driving in the mixed traffic environment. Thus, there are five types of safe time headways in the mixed HV and CAV environment, and each of them corresponds to a type of car-following pattern.

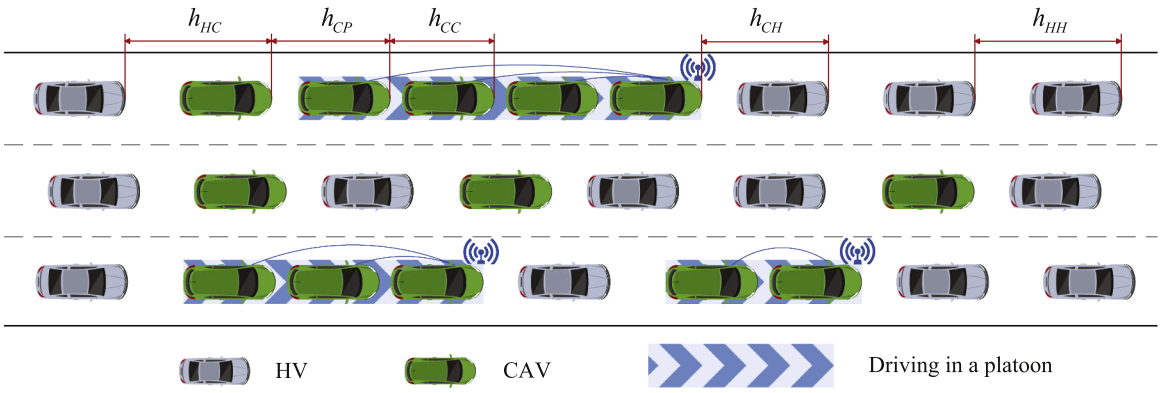


Fig. 2. Illustration of HVs and CAVs mixed on the road ($L = 4$).

Let L denote the maximum allowable platoon size of CAVs. When the size of a CAV platoon reaches L, the succeeding CAV cannot join the platoon anymore and has to drive alone without platooning or form a new CAV platoon with the following CAV, if any. Fig. 2 illustrates an example with $L = 4$. It shows that the fifth CAV has to drive alone without platooning. If we set the maximum platoon size to 1 (i.e., $L = 1$), it indicates that CAVs can not form platoons with each other on the road. All CAVs drive as individual units in the mixed traffic scenario. In contrast, if we set the maximum platoon size to positive infinity (i.e., $L = +\infty$), it means CAVs can form a platoon of any size and are not limited by the platoon scale anymore. Both cases are simplified cases that have received much attention in previous literature (Ghiasi et al., 2017; Liu and Song, 2019; Chen et al., 2022; Wang et al., 2022).

To clearly describe the spatial arrangement of vehicles, we introduce two complementary classification systems and symbolic representations:

(1) Symbolic representation of vehicle pairs

As illustrated in Fig. 1, the classification of vehicle pairs is only related to the types of succeeding and preceding vehicles. For ease of presentation of the vehicle pair and compact representation in the following mathematical derivation, "1" and "2" denote the HV type and the CAV type, respectively. The proportions of four types of vehicle pairs in the mixed traffic flow can be denoted by τ with different subscripts, where the combination and their sequence of "1" and "2" represent different types of vehicle pairs. Specifically:

- τ_{11}: Proportion of HV-HV pairs
- τ_{12}: Proportion of HV-CAV pairs
- τ_{21}: Proportion of CAV-HV pairs
- τ_{22}: Proportion of CAV-CAV pairs

(2) Symbolic representation of car-following patterns

However, the classification of car-following patterns is not only related to the types of succeeding and preceding vehicles, but also related to their platooning status as illustrated in Fig. 1. There is no one-to-one mapping relationship between vehicle pairs and car-following patterns. Therefore, different symbols are used in the classification of car-following patterns instead of "1" and "2". Here, "H" and "C" denote the HV type and the CAV type, respectively. Moreover, "P" is introduced to represent a saturated CAV platoon. Then, the combination of "H", "C", and "P" represents different car-following patterns, which can be used as superscripts in the model formulation. Specifically:

- HH: An HV following an HV
- HC: An HV following a CAV
- CH: A CAV following an HV
- CP: A CAV following a saturated CAV platoon (the inter-platoon car-following pattern)
- CC: A CAV following another CAV within the same platoon (the intra-platoon car-following pattern)

Similarly, variables and parameters related to five types of car-following patterns are distinguished by these superscripts. For example, let h with different subscripts represent the safe time headway in different car-following patterns as shown in Fig. 2, i.e., h_{HH}, h_{HC}, h_{CH}, h_{CP}, h_{CC}. It is noted that all headways in this study refer to safe time headways between two adjacent vehicles because our analysis focuses on lane capacity, which is similar to Liu and Song (2019). Let p with different subscripts represent the probability of different car-following patterns in the mixed traffic flow, i.e., p_{HH}, p_{HC}, p_{CH}, p_{CP}, p_{CC}.

(3) Relationship between the two classification systems

As illustrated in Fig. 1, the first three vehicle pair types correspond directly to car-following patterns. Therefore, we have the following relationship:

$$\tau_{11} = p_{HH}, \quad \tau_{12} = p_{HC}, \quad \tau_{21} = p_{CH}. \tag{1}$$

For CAV-CAV pairs, we must further distinguish whether the two CAVs belong to the same platoon. Thus, we can derive:

$$\tau_{22} = p_{CP} + p_{CC}. \tag{2}$$

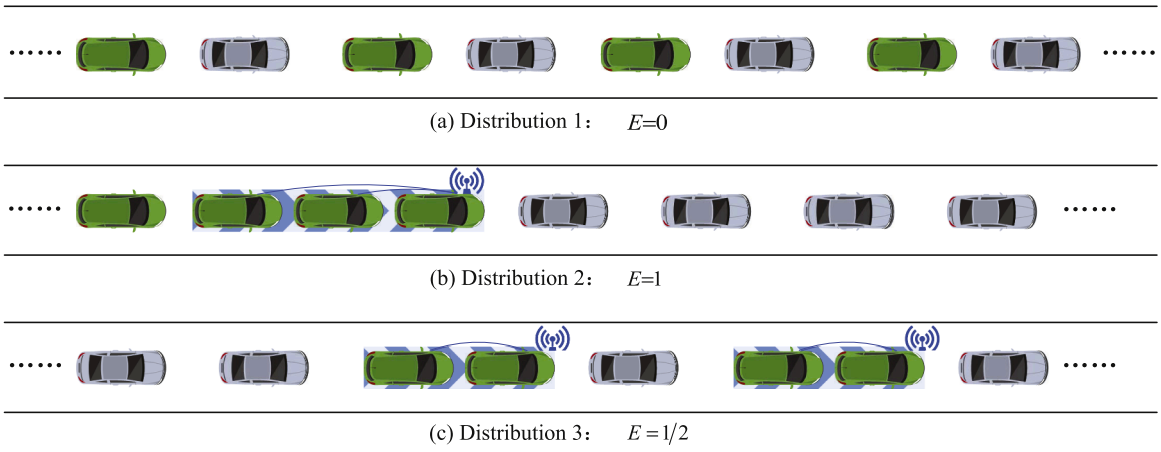


Fig. 3. Several spatial distributions of heterogeneous vehicles on the longitudinal lane with the same CAV penetration rate..

Table 2
Value of CAV clustering intensity corresponding to different spatial distributions in Fig. 3.

Distribution No.	Description	p_{CC}	τ_{22}	E
1	All CAVs driving alone on the longitudinal lane as shown in Fig. 3(a), which is one of the extreme cases of the spatial distribution of heterogeneous vehicles on the longitudinal lane.	0	0	0
2	All CAVs are clustered together on the longitudinal lane as shown in Fig. 3(b), which is another extreme case of the spatial distribution of heterogeneous vehicles on the longitudinal lane.	$\frac{1}{3}$	$\frac{1}{2}$	1
3	HVs and CAVs periodically distribute on the longitudinal lane as shown in Fig. 3(c)	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{2}$

2.2. Indexes on the spatial distribution of heterogeneous vehicles on the longitudinal lane

Under exogenously given the CAV penetration rate, an index is introduced to quantitatively characterize the spatial distribution of heterogeneous vehicles on the longitudinal lane without any simplified assumptions on car-following patterns and the spatial distribution of mixed traffic flow.

The index is defined as the clustering degree of CAV on the longitudinal lane in the mixed HV and CAV environment, which can be measured by the ratio of the number of vehicle pairs with both preceding and succeeding vehicles being CAVs to the total number of CAVs in the mixed traffic flow. Because we do not distinguish whether two adjacent CAVs are in the same CAV platoon or not in the calculation of the proposed index, the index is named CAV clustering intensity instead of CAV platooning intensity (Ghiasi et al., 2017; Zhou and Zhu, 2021; Guan et al., 2023; Qin et al., 2023; Ren et al., 2024). Let E denote the CAV clustering intensity. Mathematically, it is expressed as follows:

$$E := \frac{\tau_{22}}{P_C}, \tag{3}$$

where τ_{22} is the proportion of the CAV-CAV vehicle pair in the mixed traffic flow. We can see that if more CAVs are clustered together with each other on the longitudinal lane, the proportions of the CAV-CAV vehicle pair in the mixed traffic flow and the CAV clustering intensity will increase; otherwise, they will decrease.

The CAV clustering intensity E is a measurable quantity from field observations. Given a sequence of vehicle types obtained from trajectory data or advanced sensors, E can be estimated as the proportion of CAVs that are immediately followed by another CAV. This makes E a potential real-time diagnostic metric for monitoring the effective coordination level of CAVs in a traffic stream.

To acquire a better understanding of the proposed index above, we take an example in Fig. 3 to illustrate the specific meaning and the exact value of CAV clustering intensity in detail. To make sure the number of headways equals the number of vehicles, we consider a stream of heterogeneous vehicles driving on a single-lane ring road. The CAV penetration rate is 0.5 (i.e., $P_C = 0.5$), and the maximum platoon size is set as 3, (i.e., $L = 3$). Fig. 3 illustrates three different spatial distributions of heterogeneous vehicles on the longitudinal lane, where Fig. 3(a) and (b) are two extreme cases of the spatial distribution of mixed traffic flow, Fig. 3(c) is a normal case. For the convenience of calculations, we assume that the number of vehicles on the single-lane ring road tends to infinity. This is a standard theoretical assumption in traffic flow analysis to eliminate boundary effects and derive general, steady-state relationships. Consequently, the spatial distributions depicted in Fig. 3 are conceptual illustrations of repeating patterns. The eight vehicles explicitly drawn are for visual clarity, while the ellipsis (⋯) indicates that the identical pattern (i.e., Fig. 3(a) and (c)) or vehicle (i.e., Fig. 3(b)) extends indefinitely in both directions. Then, according to Eq. (3), the specific values of CAV clustering intensity corresponding to different spatial distributions are listed in Table 2. It is noted that since the number of vehicles is infinite, the value of CAV clustering intensity in Table 2 is the limiting value.

Here, we take the 2nd distribution in Table 2 as an example. We set the total number of vehicles to be N. Because vehicles are driving on a single-lane ring road, the number of vehicle pairs is also equal to N. As exhibited in Fig. 3(b), there are NP_C CAVs clustered together and $NP_C - 1$ CAV-CAV vehicle pairs on the longitudinal lane. The number of intra-platoon car-following patterns is $NP_C - 1 - \left\lceil \frac{NP_C}{L} - 1 \right\rceil$, which is equal to the number of CAV-CAV vehicle pairs (i.e., $NP_C - 1$) minus the number of inter-platoon car-following patterns (i.e., $\left\lceil \frac{NP_C}{L} - 1 \right\rceil$). Because the number of vehicles on the single-lane ring road tends to be infinite, the limiting value of p_{CC}, τ_{22}, and E can be calculated as follows:

$$p_{CC} = \lim_{N\to\infty} \frac{NP_C - 1 - \left\lceil \frac{NP_C}{L} - 1 \right\rceil}{N} = P_C\left(1 - \frac{1}{L}\right) = \frac{1}{3},$$
$$\tau_{22} = \lim_{N\to\infty} \frac{NP_C - 1}{N} = P_C - \lim_{N\to\infty} \frac{1}{N} = \frac{1}{2},$$
$$E = \frac{\tau_{22}}{P_C} = 1.$$

Then, we analyze the value range of the proposed CAV clustering intensity. According to the physical meaning of the proposed indexes and Eq. (3), it can be seen that when all CAVs on the longitudinal lane are clustered together, the CAV clustering intensity reaches its maximum value of 1.

If CAVs on the longitudinal lane strive to follow HVs as much as possible instead of clustering together, the spatial distribution of heterogeneous vehicles will achieve the lowest CAV clustering intensity. However, it is noted that the minimum value of the CAV clustering index is not always equal to zero but is directly related to the value of CAV penetration rate P_C. Specifically, if the number of HVs exceeds the number of CAVs (i.e., $P_C \leqslant 0.5$), one or more HVs can be inserted between any two adjacent CAVs. At that time, all CAVs drive alone on the longitudinal lane, with no adjacent CAVs clustering together. This corresponds to the minimum values of CAV clustering intensity, that is, $E_{\min} = 0$. If the number of CAVs on the lane exceeds that of HVs (i.e., $P_C > 0.5$), CAV and HV can not appear alternately one by one, and there must be some CAVs gathered together. When HVs are inserted as much as possible between adjacent CAVs, the proportion of CAVs clustering together is $(2P_C - 1)$. Among them, the proportion of the leading vehicle of CAV platoons is $\frac{1}{L}$. Thus, we can know $p_{CC} = (2P_C - 1)\left(1 - \frac{1}{L}\right)$, $\tau_{22} = 2P_C - 1$. At that time, CAV clustering intensity reaches the minimum value. By substituting p_{CC} and τ_{22} into Eq. (3), we obtain the minimum values of CAV clustering intensity, that is $E_{\min} = \frac{2P_C - 1}{P_C}$. Therefore, as for the value range of CAV clustering intensity, we have the following proposition:

Proposition 1. *Under any exogenously given CAV penetration rate, the value range of CAV clustering intensity is* $\left[\max\left(0, \frac{2P_C-1}{P_C}\right), 1\right]$. *Namely,* $E \in \left[\max\left(0, \frac{2P_C-1}{P_C}\right), 1\right]$, $\forall P_C \in [0, 1]$.

In contrast to the CAV platooning intensity discussed in previous studies, the range of CAV clustering intensity, particularly its lower bound, is influenced by the CAV penetration rate. Specifically, we define $E = 0$ to indicate that all CAVs are driving independently on the longitudinal lane, while $E = 1$ signifies that all CAVs are clustered together on the longitudinal lane. Consequently, when the CAV penetration rate exceeds 0.5, it is inevitable that some CAVs will cluster together. Therefore, the CAV clustering intensity must be greater than zero, i.e., $E > 0$. Especially, in a pure CAV environment (i.e., $P_C = 1$), the CAV clustering intensity is fixed at 1 according to the above proposition. This aligns with reality because all vehicles are CAVs and can only cluster together.

Finally, we can find that the index E provides distinct value over using raw proportions: (i) it normalizes the count of CAV-CAV adjacencies by the total number of CAVs, making it a rate comparable across different penetration levels; (ii) it has a clear physical interpretation as the probability that a randomly selected CAV is followed by another CAV; and (iii) its bounded range $[E_{\min}, 1]$ directly reflects the spectrum from completely dispersed to fully clustered CAV distributions, simplifying subsequent analysis.

2.3. Markov chain model

A Markov chain model is established to formulate the spatial distribution of heterogeneous vehicles on the longitudinal lane. Accordingly, the probability distribution of CAV platoon size and time headway is derived. It is noted that this study focuses on the longitudinal interactions of a mixed HV and CAV traffic stream on a single lane and the resulting mixed lane capacity. Accordingly, the proposed Markov-chain-based vehicle ordering model describes the stochastic longitudinal arrangement of heterogeneous vehicles along the vehicle sequence index, and the mixed lane capacity is derived from pattern-dependent safe time headways under this one-dimensional setting. Lateral maneuvers (e.g., lane changing), as well as network/intersection interactions (e.g., merging, weaving, and conflict-point constraints), are not explicitly modeled in the present framework. This single-lane longitudinal focus follows the convention adopted by a broad line of analytical mixed-capacity studies reviewed in Section 1.1 and summarized in Table 1, which commonly abstract away multi-lane and intersection-conflict mechanisms to obtain interpretable, closed-form capacity characterizations.

2.3.1. Model formulation

Since the spatial distribution of HVs and CAVs on the longitudinal lane is uncertain, the vehicle type at each position along the mixed traffic stream can be regarded as a discrete stochastic process. To formulate the relationship between the vehicle type at each position on the lane and the vehicle pair, a Markov chain model is established.

Fig. 4. Illustration of the mixed traffic flow on a single-lane ring road.

For the stream of mixed traffic flow, we number all vehicles in a front-to-back order and use the index $n \in \{1, 2, \dots, \}$ to represent the nth vehicle (i.e., the Markov "step" advances from one vehicle to the next and characterizes vehicle-type adjacency/ordering, rather than a time-recursive vehicle dynamics on a discretized space-time lattice as in DP-based trajectory optimization models such as Wei et al. (2017)). Let A_n denote the type of the nth vehicle, which is a stochastic variable and can be regarded as the state variable at step n. $A_n = 1$ indicates that the nth vehicle is an HV, and $A_n = 2$ indicates that the nth vehicle is a CAV. Thus, a sequence of discrete stochastic variables corresponding to the vehicle type at each location on the longitudinal lane $\{A_n, n \in 1, 2, \dots, \}$ can be regarded as a finite-state Markov chain.

All possible types of vehicles at each position constitute the state space of the Markov chain. Thus, the state space is:

$$S = [1, 2]. \tag{4}$$

The initial state of the Markov chain model is the type of the first vehicle on the longitudinal lane. In the mixed traffic flow, the probability of a vehicle being a CAV is P_C, and conversely, the probability of the vehicle being an HV is $(1 - P_C)$. Therefore, the initial state is:

$$\varpi = \left[1 - P_C, P_C\right]. \tag{5}$$

The one-step transition probability matrix of this Markov chain model is defined as the probability of the type of current vehicle transitioning to the type of the succeeding vehicle, which is formulated as follows:

$$T := \begin{bmatrix} t_{11} & t_{21} \\ t_{12} & t_{22} \end{bmatrix}, \tag{6}$$

where t_{sr} denotes the probability that a type-r vehicle is followed by a type-s vehicle. Thus, we have $t_{sr} = \Pr\left(A_{n+1} = s \middle| A_n = r\right) = \frac{\Pr\left(A_{n+1}=s, A_n=r\right)}{\Pr\left(A_n=r\right)}$, $\forall n \in \{1, 2, \dots, \}, s, r \in S$.

To calculate the one-step transition probability between different states, the assumption that all vehicles are driving on a single-lane ring road and forming a head-to-tail sequence in Section 2.1 is also adopted, as shown in Fig. 4. It ensures that the number of headways is equal to the number of vehicles, which can facilitate the establishment of the relationship between the vehicle penetration rate and the proportion of vehicle pairs.

Because only the succeeding vehicle in the "11" and "12" vehicle pairs is an HV, the sum of the proportion of these two vehicle pairs on the road should equal the HV penetration rate in the mixed traffic flow. Thus, we have the following equation:

$$\tau_{11} + \tau_{12} = 1 - P_C.$$

Similarly, only the succeeding vehicle in the "21", and "22" vehicle pairs is a CAV. Thus, the sum of the proportion of these two vehicle pairs should equal the CAV penetration rate in the mixed traffic flow. Then, we have:

$$\tau_{21} + \tau_{22} = P_C.$$

When we pay attention to the preceding vehicle of various types of vehicle pairs, it is found that only the preceding vehicle in the "11" and "21" vehicle pairs is an HV. Therefore, the sum of the proportion of these two vehicle pairs on the road should equal the HV penetration rate, which can be expressed as follows:

$$\tau_{11} + \tau_{21} = 1 - P_C.$$

Similarly, only the preceding vehicle in the "12" and "22" vehicle pairs is a CAV, and the sum of the proportion of these two vehicle pairs on the road should equal the CAV penetration rate in the mixed traffic flow. We can also have:

$$\tau_{12} + \tau_{22} = P_C.$$

Based on the established equations above, the proportions of different types of vehicle pairs can be represented as a function of τ_{22} and P_C as follows:

$$\tau_{11} = 1 - 2P_C + \tau_{22}, \tag{7}$$

$$\tau_{21} = P_C - \tau_{22}, \tag{8}$$

$$\tau_{12} = P_C - \tau_{22}. \tag{9}$$

It is worth noting that under a given CAV penetration rate P_C, the proportions of all vehicle-pair types are interlinked through conservation conditions (Eqs. (7)–(9)). Therefore, the single index E, which is a function of τ_{22}, together with P_C, provides a complete description of the vehicle-pair distribution $\{\tau_{11}, \tau_{12}, \tau_{21}, \tau_{22}\}$.

Besides, it also indicates that, for an exogenously given P_C, we can express the one-step transition probability t_{sr} as a function of CAV clustering intensity:

$$t_{11} = \Pr\left(A_{n+1} = 1 \middle| A_n = 1\right) = \frac{\Pr\left(A_{n+1} = 1, A_n = 1\right)}{\Pr\left(A_n = 1\right)} = \frac{\tau_{11}}{1 - P_C} = \frac{1 - 2P_C + \tau_{22}}{1 - P_C} = \frac{1 - 2P_C + EP_C}{1 - P_C}, \tag{10}$$

$$t_{21} = \Pr\left(A_{n+1} = 2 \middle| A_n = 1\right) = \frac{\Pr\left(A_{n+1} = 2, A_n = 1\right)}{\Pr\left(A_n = 1\right)} = \frac{\tau_{21}}{1 - P_C} = \frac{P_C - \tau_{22}}{1 - P_C} = (1 - E)\frac{P_C}{1 - P_C}, \tag{11}$$

$$t_{12} = \Pr\left(A_{n+1} = 1 \middle| A_n = 2\right) = \frac{\Pr\left(A_{n+1} = 1, A_n = 2\right)}{\Pr\left(A_n = 2\right)} = \frac{\tau_{12}}{P_C} = \frac{P_C - \tau_{22}}{P_C} = 1 - E, \tag{12}$$

$$t_{22} = \Pr\left(A_{n+1} = 2 \middle| A_n = 2\right) = \frac{\Pr\left(A_{n+1} = 2, A_n = 2\right)}{\Pr\left(A_n = 2\right)} = \frac{\tau_{22}}{P_C} = E. \tag{13}$$

We can find that the transition probability is rigorously derived as a function of the CAV penetration rate and CAV clustering intensity, rather than being subjectively designed as in previous studies (Ghiasi et al., 2017; Zhou and Zhu, 2020, 2021; Guan et al., 2023). To facilitate a more intuitive understanding, the specific meaning of the one-step transition probability is explained in three special cases of the spatial distribution of heterogeneous vehicles.

(1) Maximum clustering intensity of CAVs

In the first case, all CAVs are clustered together on the longitudinal lane, similar to the case depicted in Fig. 3 (b). Thus, the maximum value of CAV clustering intensity is achieved and equal to 1, i.e., $E = 1$. Substituting the value of CAV clustering intensity into Eqs. (10)–(13), we can have:

$$\begin{cases} t_{11} = t_{22} = 1, \\ t_{12} = t_{21} = 0. \end{cases}$$

This implies that when all CAVs are clustered together and the preceding vehicle is known to be a CAV, the probability of the succeeding vehicle also being a CAV is 1, while the probability of it being an HV is 0. Similarly, when it is known that the preceding vehicle is an HV, the succeeding vehicle must be an HV, and it is impossible that the succeeding vehicle is a CAV.

(2) Minimum clustering intensity of CAVs

According to the analysis in Section 2.2, we know that the minimum CAV clustering intensity is achieved when all CAVs are dispersed as much as possible, which is regarded as the second case. Because its exact value may vary with the value of the CAV penetration rate as stated in Proposition 1, we discuss the value of one-step transition probability and the CAV clustering intensity in the following two situations.

If $P_C \leqslant 0.5$, the minimum value of CAV clustering intensity is 0, which is similar to the case in Fig. 3 (a). Similarly, after substituting the value of CAV clustering intensity into Eqs. (10)–(12), we can obtain:

$$\begin{cases} t_{12} = 1, t_{22} = 0, \\ t_{11} = \dfrac{1-2P_C}{1-P_C}, t_{21} = \dfrac{P_C}{1-P_C}. \end{cases}$$

In this situation, it means that when all CAVs are dispersed as much as possible and the preceding vehicle is a CAV, the probability of the succeeding vehicle being an HV is 1, and it is impossible that the succeeding vehicle is a CAV. When it is known that the preceding vehicle is an HV, the probability of the succeeding vehicle being a CAV is $\frac{P_C}{1-P_C}$, and the probability of the succeeding vehicle being an HV is $\frac{1-2P_C}{1-P_C}$.

If $P_C > 0.5$, the minimum CAV clustering intensity is $E = \frac{2P_C-1}{P_C}$. After substituting the value of CAV clustering intensity into Eqs. (10)–(13), we have:

$$\begin{cases} t_{11} = 0, t_{21} = 1, \\ t_{12} = \dfrac{1-P_C}{P_C}, t_{22} = \dfrac{2P_C-1}{P_C}. \end{cases}$$

It means that when all CAVs are dispersed as much as possible, and it is known that the preceding vehicle is an HV, the succeeding vehicle must be a CAV instead of an HV. When it is known that the preceding vehicle is a CAV, the probability of the succeeding vehicle being an HV is $\frac{1-P_C}{P_C}$.

In summary, the one-step transition probability under any value of the CAV penetration rate is:

$$\begin{cases} t_{11} = \max\left\{0, \dfrac{1-2P_C}{1-P_C}\right\}, t_{21} = \min\left\{1, \dfrac{P_C}{1-P_C}\right\}, \\ t_{12} = \min\left\{1, \dfrac{1-P_C}{P_C}\right\}, t_{22} = \max\left\{0, \dfrac{2P_C-1}{P_C}\right\}. \end{cases}$$

(3) Completely random distribution

The assumption that all vehicles are completely random distributed on the longitudinal lane is widely adopted by almost all existing studies (Bujanovic and Lochrane, 2018; Liu and Song, 2019; Mohajerpoor and Ramezani, 2019; Sala and Soriguera, 2021; Chen et al., 2022; Wang et al., 2022; Yao et al., 2023, 2024), which is regarded as the third case. At that time, the types of two adjacent vehicles are independent of each other. Thus, we can directly obtain the following transition matrix:

$$\begin{cases} t_{11} = t_{12} = 1 - P_C, \\ t_{21} = t_{22} = P_C. \end{cases}$$

On the other hand, because vehicle types along the longitudinal lane in the third case can be regarded as a Bernoulli process, we can derive $\tau_{22} = (P_C)^2$ and $E = P_C$. Moreover, the Bernoulli process is only a special case of the Markov chain in this study. After substituting $E = P_C$ into Eqs. (10)–(13), we can derive the same result of the one-step transition probability as shown above. Therefore, different from existing studies based on the completely random assumption, the Markov chain model in this study provides a more general methodology to analyze the spatial distribution of heterogeneous vehicles.

2.3.2. Probability distribution of CAV platoon size

Although we know the probability of different types of vehicle pairs in the mixed traffic flow and the transition probability between two types of vehicles on the longitudinal lane, the probability of different time headways is still unknown because the CAV-CAV vehicle pair corresponds to different headways as illustrated in Fig. 2. For the CAV-CAV vehicle pair, we need to distinguish the driving pattern of the succeeding vehicle, i.e., driving alone or driving in the form of CAV platoons. In other words, the probability that a vehicle is at any position in CAV platoons should be calculated to identify whether these two adjacent vehicles belong to the same CAV platoon.

Let p_m denote the probability of a vehicle being the mth vehicle in a CAV platoon, $m \in \{1, 2, \dots, L\}$. Based on whether the vehicle is the leading vehicle of a CAV platoon, we classify the value of m into two categories: $m \in \{2, \dots, L\}$ and $m = 1$.

(1) $m \in \{2, 3, \dots, L\}$

It means that a vehicle is positioned within a CAV platoon but not as the leading vehicle. The probability of this vehicle being the mth vehicle in a CAV platoon can be expressed as follows:

$$p_m = \Pr\left\{A_m = 2, A_{m-1} = 2, \dots, A_1 = 2\right\}. \tag{14}$$

According to the conditional probability formula and the Markov property, it can be inferred that:

$$p_m = \Pr\left\{A_m = 2 \middle| A_{m-1} = 2, \dots, A_1 = 2\right\} \Pr\left\{A_{m-1} = 2, \dots, A_1 = 2\right\}$$

$$= \Pr\left\{A_m = 2 \middle| A_{m-1} = 2\right\} \Pr\left\{A_{m-1} = 2, \ldots, A_1 = 2\right\} = t_{22} p_{m-1}. \tag{15}$$

Because Eq. (15) holds for all $m \in \{2, 3, \ldots, L\}$, according to the mathematical induction, we can obtain the probability of a vehicle being the mth vehicle in a CAV platoon expressed as follows:

$$p_m = (t_{22})^2 p_{m-2} = \ldots = (t_{22})^{m-1} p_1. \tag{16}$$

(2) $m = 1$

It represents that a vehicle is the leading vehicle of a CAV platoon. At that time, its preceding vehicle should be either an HV or a maximum-size CAV platoon. Therefore, we have $p_1 = p_{CP} + p_{CH}$. As shown in Fig. 1, the probability of some types of vehicle pairs in the mixed traffic flow is directly equal to the probability of the corresponding car-following pattern, e.g., $\tau_{12} = p_{HC}$, $\tau_{21} = p_{CH}$. Therefore, based on Eqs. (8)–(9), we can derive $p_{CH} = p_{HC}$. Moreover, it also holds that $p_{HC} = \tau_{12} = \Pr\left(A_{n+1} = 1, A_n = 2\right) = P_C t_{12}$ according to Eq. (12). Therefore, we can obtain the following equation:

$$p_1 = p_{CP} + p_{CH} = p_{CP} + p_{HC} = p_{CP} + P_C t_{12}. \tag{17}$$

According to Eqs. (13) and (16), the probability of a CAV following a maximum-size CAV platoon is:

$$p_{CP} = \Pr\left\{A_{L+1} = 2, A_L = 2, \ldots, A_1 = 2\right\} = (t_{22})^L p_1. \tag{18}$$

By substituting Eq. (18) into Eq. (17), if $t_{22} \neq 1$, we can derive:

$$p_1 = \frac{P_C t_{12}}{1 - (t_{22})^L}. \tag{19}$$

It is noted that $t_{22} = 1$ is equivalent to $E = 1$ according to Eq. (13), which indicates the extreme case where all CAVs are clustered together on the longitudinal lane. Considering that the maximum platoon size L is relatively small compared to the number of vehicles on the lane or all CAVs can form integer maximum platoons, if $t_{22} = 1$ (i.e., $E = 1$), we can obtain the following equation:

$$p_1 = \frac{P_C}{L}. \tag{20}$$

In summary, we can get the probability that a vehicle is the leading vehicle in CAV platoons:

$$p_1 = \begin{cases} \dfrac{P_C}{L}, & E = 1, \\ \dfrac{P_C t_{12}}{1 - (t_{22})^L}, & E \neq 1. \end{cases} \tag{21}$$

By substituting Eqs. (19)–(20) into Eq. (18), we can obtain:

$$p_{CP} = \begin{cases} \dfrac{P_C}{L}, & E = 1, \\ \dfrac{P_C t_{12} (t_{22})^L}{1 - (t_{22})^L}, & E \neq 1. \end{cases} \tag{22}$$

Let p_l^{end} denote the probability that a vehicle at the end of a CAV platoon of size $l, l \in \{1, 2, \ldots, L\}$ in the mixed traffic flow. It also equals the ratio of the CAV platoon of size l to the total vehicle number. If $l \in \{1, 2, \ldots, L-1\}$, it denotes that the vehicle is at the end of a CAV platoon that does not reach the maximum platoon size, and its succeeding vehicle can only be an HV. Therefore, $p_l^{end} = \Pr\left\{A_{l+1} = 1, A_l = 2, \ldots, A_1 = 2\right\} = t_{12} p_l$. If $l = L$, it means that the vehicle is at the end of a maximum-size CAV platoon, and its succeeding vehicle can be either an HV or a CAV. Namely, the vehicle is the Lth vehicle in a CAV platoon, and thus it has $p_l^{end} = p_L$. Finally, we can obtain the following formula:

$$p_l^{end} = \begin{cases} t_{12} (t_{22})^{l-1} p_1, & l \in \{1, 2, \ldots, L-1\}, \\ p_L, & l = L. \end{cases} \tag{23}$$

Due to the uncertain distribution of heterogeneous vehicles on the longitudinal lane, CAV platoon size in the mixed traffic flow is a stochastic variable. According to Eq. (23), we can derive the proportion that CAV platoons with different sizes account for the total platoon is $\frac{p_l^{end}}{\sum_{l=1}^{L-1} t_{12}(t_{22})^{l-1} p_1 + p_L} = \frac{p_l^{end}}{t_{12} p_1 \frac{1-(t_{22})^{L-1}}{1-t_{22}} + p_L}$, $\forall l \in \{1, 2, \ldots, L\}$. Accordingly, the probability distribution of CAV platoon size can be expressed as follows:

$$\Pr\left(\text{Platoon size} = l\right) = \frac{p_l^{end}}{\sum_{l=1}^{L-1} t_{12}(t_{22})^{l-1}p_1 + p_L} = \begin{cases} \dfrac{t_{12}(t_{22})^{l-1}p_1}{t_{12}p_1\frac{1-(t_{22})^{L-1}}{1-t_{22}} + p_L}, & l \in \{1,2,\dots,L-1\}, \\ \dfrac{p_L}{t_{12}p_1\frac{1-(t_{22})^{L-1}}{1-t_{22}} + p_L}, & l = L. \end{cases} \tag{24}$$

After substituting Eqs. (12)–(13), (19)–(20), and (16) into Eq. (24), (24) can be further expressed as a function of E and L:

$$\Pr\left(\text{Platoon size} = l\right) = \begin{cases} E^{l-1}(1-E), & l \in \{1,2,\dots,L-1\}, \\ E^{L-1}, & l = L. \end{cases} \tag{25}$$

It is noted that Eq. (25) is applicable for any value of E and L. For example, $E = 1$ is an extreme case of the spatial distribution of heterogeneous vehicles on the longitudinal lane, which means all CAVs are clustered together on the longitudinal lanes. Because all CAVs form maximum-size CAV platoons, we have $l \in \{L\}$. Then, it can be inferred that $\Pr\left(\text{Platoon size} = L\right) = 1$.

Remark 1. Eq. (25) shows that the probability distribution of CAV platoon size depends on the CAV clustering intensity E and the maximum platoon size L. This implies that for a fixed E and L, the platoon size distribution remains invariant. Thus, E quantitatively characterizes the spatial distribution of the CAV platoon with different sizes on longitudinal lanes. We note that E itself is defined as a function of P_C and τ_{22}, and its feasible range varies with P_C as given in Proposition 1.

Fig. 5 intuitively illustrates the variation of the distribution probability of CAV platoon size against the CAV clustering intensity and the maximum platoon size. Specifically, for an exogenously given CAV penetration rate, i.e., $P_C = 0.5$, the CAV clustering intensity varies in the range from 0.2 to 0.8 with a step size of 0.2. Because when the maximum platoon size is 1, CAVs do not form platoons with other CAVs on the longitudinal lanes, and we always have $\Pr\left(\text{Platoon size} = 1\right) = 1$. Therefore, we set the maximum platoon size to vary in the range from 2 to 5.

We can find that the probability of CAV platoons with different sizes is independent of the value of the maximum platoon size except for the probability of CAV platoons with the maximum size. In other words, with the increase of the maximum platoon size, the probability of CAV platoons with non-maximum size remains unchanged. When the maximum platoon size is given, the probability of the maximum platoon size is monotonically increasing with the CAV clustering intensity.

2.3.3. *Probability distribution of time headway*

Due to the uncertain distribution of heterogeneous vehicles in the mixed traffic flow, the time headway between any two adjacent vehicles is also a stochastic variable, which is similar to CAV platoon size in Section 2.3.2. Let h denote the time headway between any two adjacent vehicles.

According to Eqs. (10)–(13), (16), and (22), the probability of different car-following patterns in the mixed traffic flow can be derived. Due to the one-to-one mapping relationship between car-following patterns and time headways, we can obtain the probability distribution of time headway as follows:

(1) The probability of the car-following pattern of an HV following an HV in the mixed traffic flow is:

$$\Pr\left\{h = h_{HH}\right\} = p_{HH} = \left(1 - P_C\right)t_{11} = 1 - 2P_C + \tau_{22}. \tag{26}$$

(2) The probability of the car-following pattern of an HV following a CAV in the mixed traffic flow is:

$$\Pr\left\{h = h_{HC}\right\} = p_{HC} = P_C t_{12} = P_C - \tau_{22}. \tag{27}$$

(3) The probability of the car-following pattern of a CAV following an HV in the mixed traffic flow is:

$$\Pr\left\{h = h_{CH}\right\} = p_{CH} = \left(1 - P_C\right)t_{21} = P_C - \tau_{22}. \tag{28}$$

(4) The probability of the car-following pattern of a CAV following a CAV in the same platoon is:

$$\Pr\left\{h = h_{CC}\right\} = p_{CC} = \sum_{m=2}^{L} p_m = \begin{cases} \dfrac{P_C(L-1)}{L}, & E = 1, \\ \displaystyle\sum_{m=2}^{L} t_{22}^{m-1}p_1 = \frac{P_C t_{22}(1 - t_{22}^{L-1})}{1 - t_{22}^{L}}, & E \neq 1. \end{cases} \tag{29}$$

(5) The probability of the car-following pattern of a CAV following a maximum-size CAV platoon in the mixed traffic flow is:

$$\Pr\left\{h = h_{CP}\right\} = p_{CP} = \begin{cases} \dfrac{P_C}{L}, & E = 1, \\ \dfrac{P_C t_{12} t_{22}^{L}}{1 - t_{22}^{L}}, & E \neq 1. \end{cases} \tag{30}$$

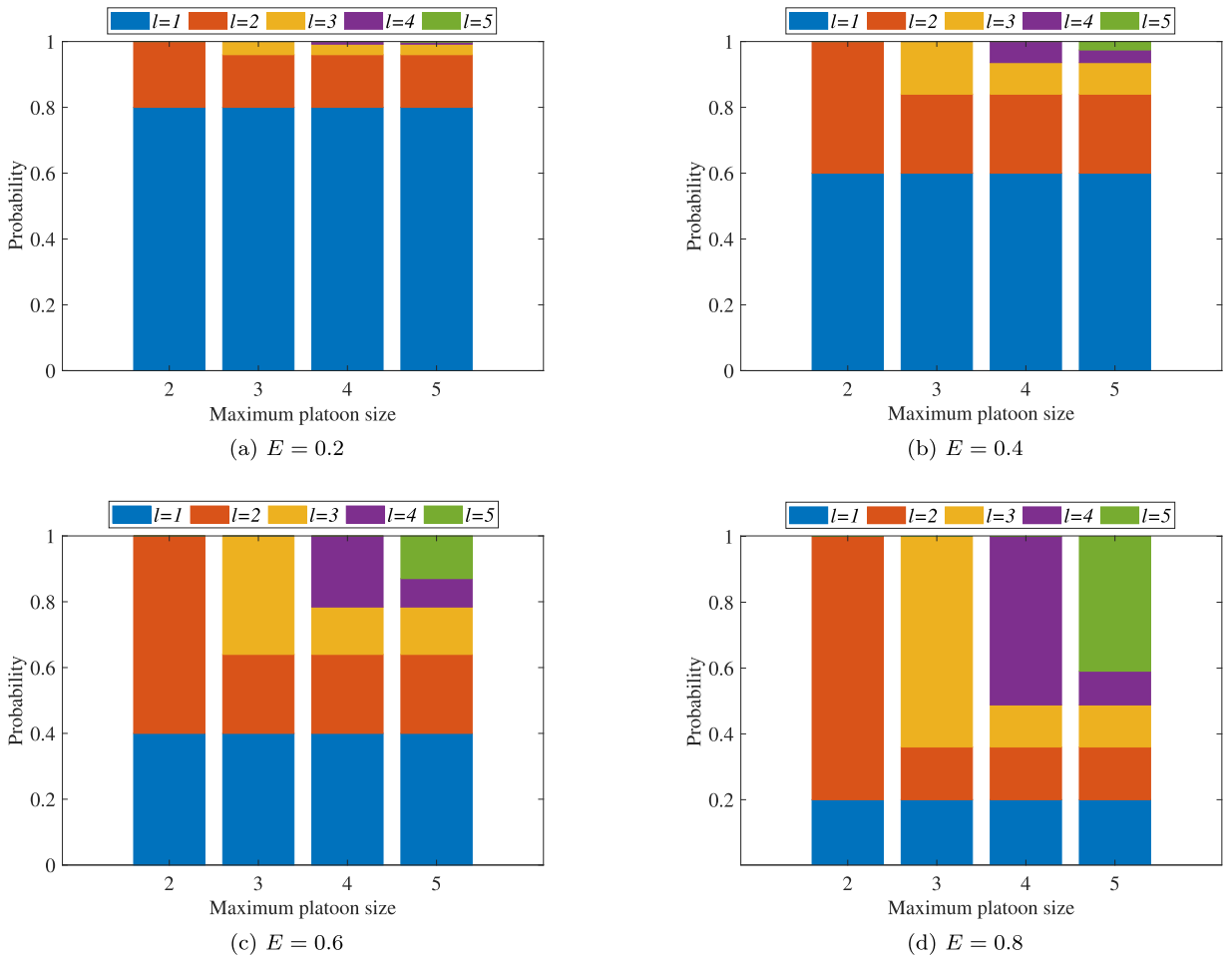


Fig. 5. Probability distribution of CAV platoon size under different values of clustering intensity and maximum platoon size.

It is noted that since the one-step transition probability and the probability of different vehicle pairs are functions of P_C and E as shown in Eqs. (3), (12), and (13), the probability of different time headways in Eqs. (26)–(30) also can be expressed as a function of P_C and E.

2.4. Lane capacity in the mixed HV and CAV environment

The explicit expression for calculating mixed lane capacity is derived in this subsection. Then, the effect of various factors on the mixed lane capacity is explored by theoretically analyzing the monotonicity of the mixed lane capacity function with different independent variables.

2.4.1. Explicit expression of mixed lane capacity

Let $\boldsymbol{p}$ denote the vector of the probability of different car-following patterns in the mixed traffic flow, i.e., $\boldsymbol{p} = \{p_{HH}, p_{HC}, p_{CH}, p_{CP}, p_{CC}\}^T$. $\overline{h}$ represents the average time headway on the longitudinal lane. Although the exact value of headway parameters remains controversial currently, once it is exogenously given, the average time headway on the longitudinal lane can be expressed as a function of $\boldsymbol{p}$ as follows:

$$\overline{h}(\boldsymbol{p}) = p_{CC}h_{CC} + p_{HC}h_{HC} + p_{CH}h_{CH} + p_{CP}h_{CP} + p_{CC}h_{CC}. \tag{31}$$

Let c denote the lane capacity in the mixed HV and CAV environment, referred to as mixed lane capacity for short. It represents the maximum mixed traffic throughput on a lane. Therefore, according to the relationship between the average time headway, traffic density, and traffic flow, we can derive:

$$c(\boldsymbol{p}) = \frac{1}{p_{HH}h_{HH} + p_{HC}h_{HC} + p_{CH}h_{CH} + p_{CP}h_{CP} + p_{CC}h_{CC}}. \tag{32}$$

Table 3
Assumptions adopted by previous literature in the derivation of calculation formulas of mixed lane capacity.

Assumptions	References
-	Eq. (33) in this study
$\tau_{22} = (P_C)^{2\ *}$	Bujanovic and Lochrane (2018), Sala and Soriguera (2021), Yao et al. (2023, 2024)
$L = +\infty$	Ghiasi et al. (2017), Zhou and Zhu (2020)
$L = +\infty$, $\tau_{22} = (P_C)^2$	Chen et al. (2022), Liu and Song (2019), Mohajerpoor and Ramezani (2019), Wang et al. (2022), Zhang et al. (2023)
$L = +\infty$, $h_{HH} = h_{HC}$, $h_{CH} = h_{CC}$	Levin and Boyles (2016), Li et al. (2023), Chen et al. (2022), Zhang et al. (2022)

$*$ $\tau_{22} = (P_C)^2$ indicates the distribution of heterogeneous vehicles on the longitudinal lane is regarded as completely random, and thus the sequence of vehicle types is a Bernoulli process.

Table 4
Comparison of vehicle distribution indexes adopted in the previous literature and this study.

References	Vehicle distribution index	Calculation	Value range	Transition matrix	Limited platoon size
Ghiasi et al. (2017), Zhou and Zhu (2020)	CAV platooning intensity	Calibration	$[-1, 1]$	Externally defined	×
Zhou and Zhu (2021), Guan et al. (2023), Qin et al. (2023), Ren et al. (2024)	CAV platooning intensity	Calibration	$[-1, 1]$	Externally defined	✓
This study	CAV clustering intensity	Explicit expression	$\left[\max\left(0, \frac{2P_C-1}{P_C}\right), 1\right]$	Rigorous mathematical derivation (Endogenous)	✓

After substituting the proportions of different time headways in the mixed traffic flow into Eq. (32), we can obtain the explicit expression of mixed lane capacity as follows:

$$c(P_C, L, E) = \begin{cases} \dfrac{1}{\frac{(1-E)P_C E^L}{1-E^L} h_{CP} + \frac{EP_C(1-E^{L-1})}{1-E^L} h_{CC} + (1-E)P_C(h_{HC}+h_{CH}) + (1-2P_C+EP_C)h_{HH}}, & E \neq 1, \\ \dfrac{1}{\frac{P_C}{L} h_{CP} + \frac{P_C(L-1)}{L} h_{CC} + (1-P_C)h_{HH}}, & E = 1. \end{cases} \tag{33}$$

Remark 2. As illustrated in Eq. (33), the mixed lane capacity $c(\cdot)$ is a multivariate function of CAV penetration rate P_C, maximum platoon size L, and CAV clustering intensity E. CAV penetration rate is an exogenous parameter representing the popularity of CAVs, which generally varies with time. Maximum platoon size is another exogenous parameter determined by the level of technological development, safety concerns, etc. For an exogenously given CAV penetration rate and maximum platoon size, the mixed lane capacity is a univariate function of CAV clustering intensity, which indicates it is only related to the clustering degree of CAVs on the longitudinal lane. Moreover, it demonstrates that, in terms of the calculation of mixed lane capacity, CAV clustering intensity is indeed an effective index representing the spatial distribution of heterogeneous vehicles in the mixed traffic flow.

Eq. (33) is a generalized explicit expression of mixed lane capacity, which is suited for any value of P_C, L, and E. However, for convenience in deriving the calculation formula of mixed lane capacity, various simplified assumptions are adopted in existing studies. For example, a large stream of literature does not consider the limited platoon size of CAVs ($L = +\infty$) or CAV platooning ($L = 1$), while some others assume heterogeneous vehicles are completely randomly distributed on the longitudinal lane.

Moreover, various simplified assumptions on the value of time headways have been adopted in previous studies. The existing derived calculation formulas of mixed lane capacity can be regarded as special cases of the explicit expression in Eq. (33). For an intuitive comparison, different assumptions adopted by previous literature in the derivation of calculation formulas of mixed lane capacity are summarized in Table 3.

Besides, it is noted that the index on the spatial distribution of heterogeneous vehicles plays an essential role in deriving the explicit expression of mixed lane capacity. To further illustrate the distinction between the CAV clustering intensity proposed in this study and the CAV platooning intensity defined in existing studies (Ghiasi et al., 2017; Zhou and Zhu, 2020, 2021; Guan et al., 2023; Qin et al., 2023; Ren et al., 2024), a detailed comparison between these two indexes from multiple perspectives is listed in Table 4.

We can find that the CAV platooning intensity has been widely adopted in previous studies. However, as an index obtained by the calibration method, its physical meaning and the corresponding spatial distribution of heterogeneous vehicles on the longitudinal lane may not be obtained directly. Moreover, the transition matrix in the model formulation of the Markov chain is exogenously designed as a function of the CAV generation rate and the CAV platooning intensity. It may be challenging to rigorously analyze the relationship between lane capacity and the spatial distribution of heterogeneous vehicles. In contrast, under exogenously given CAV

penetration rate P_C and CAV maximum platoon size L, the transition matrix of the Markov chain model is obtained by the rigorous mathematical derivation in Section 2.3.1. Accordingly, for the given CAV penetration rate and maximum platoon size, the mixed lane capacity can be expressed as a univariate function of the proposed CAV clustering intensity (i.e., Eq. (33)).

2.4.2. Effect of various factors on the mixed lane capacity

As expressed in Eq. (33), CAV penetration rate, maximum platoon size, and CAV clustering intensity are three crucial factors determining mixed lane capacity. Here, rigorously theoretical analysis is carried out to explore the effect of these factors on mixed lane capacity. Specifically, the monotonicity of mixed lane capacity with respect to CAV penetration rate, maximum platoon size, and CAV clustering intensity is discussed in Lemmas 1–6, and the corresponding proofs can be found in Sections A.1–A.6.

Lemma 1. *If the different time headways satisfy the following conditions, the mixed lane capacity $c(\cdot)$ is a monotonically increasing function of the CAV penetration rate P_C.*

I. $h_{HH} + h_{CP} - h_{HC} - h_{CH} < 0$,

II. $h_{HH} > h_{CP} \geqslant h_{CC}$.

Lemma 2. *If the different time headways and CAV clustering intensity satisfy the following conditions, the mixed lane capacity $c(\cdot)$ is a monotonically decreasing function of the CAV penetration rate P_C.*

I. $E\left[h_{CP} - 2h_{HC} - 2h_{CH} + 2h_{HH}\right] - 2h_{HH} > 0$,

II. $h_{HH} < h_{CP} \leqslant h_{CC}$.

Lemma 3. *The mixed lane capacity $c(\cdot)$ is a monotonically increasing function of the maximum platoon size L if and only if $h_{CP} > h_{CC}$. In contrast, the mixed lane capacity $c(\cdot)$ is a monotonically decreasing function of the maximum platoon size L if and only if $h_{CP} < h_{CC}$.*

Lemma 4. *If $h_{CP} > h_{CC}$, the mixed lane capacity $c(\cdot)$ is a concave function of the maximum platoon size L. In contrast, if $h_{CP} < Eh_{CC}$, the mixed lane capacity $c(\cdot)$ is a convex function of the maximum platoon size L.*

Lemma 5. *If the different time headways satisfy the following conditions, the mixed lane capacity $c(\cdot)$ is a monotonically increasing function of the CAV clustering intensity E.*

I. $h_{HH} + h_{CP} - h_{HC} - h_{CH} < 0$,

II. $h_{CP} \geqslant h_{CC}$.

Lemma 6. *If the different time headways satisfy the following conditions, the mixed lane capacity $c(\cdot)$ is a monotonically decreasing function of the CAV clustering intensity E.*

I. $h_{HH} + h_{CC} - h_{HC} - h_{CH} > 0$,

II. $h_{CP} \geqslant h_{CC}$.

2.5. Relationships with macroscopic traffic flow parameters

To strengthen the theoretical foundation of our model and facilitate its interpretation within established traffic flow theory, we provide explicit micro-to-macro mappings linking the microscopic parameters in our framework to the macroscopic parameters commonly used in the Cell Transmission Model (CTM) and cellular automata (CA) models, particularly Daganzo's CA(M) formulation (Daganzo, 2006). To avoid confusion with our capacity notation c, we denote Daganzo's CA(M) dimensionless update/memory parameter by γ in this subsection.

2.5.1. Pattern-level decomposition and mixture-level aggregates

For each car-following pattern ij, our safe time headway h_{ij} can be decomposed into a time-lag component and a minimum-spacing component:

$$h_{ij} = \tau_{ij} + \frac{d_{ij}}{v_f}, \tag{34}$$

where τ_{ij} is an effective time-lag (relaxation time scale), d_{ij} is the minimum spacing (jam spacing) associated with pattern ij, and v_f is the free-flow speed used for CTM/CA(M) parameter alignment.

Given the pattern probabilities p_{ij}, we define mixture-level aggregates:

$$\bar{h} = \sum_{ij} p_{ij} h_{ij}, \qquad \bar{\tau} = \sum_{ij} p_{ij} \tau_{ij}, \qquad \bar{d} = \sum_{ij} p_{ij} d_{ij}. \tag{35}$$

These quantities provide the natural link from heterogeneous microscopic interactions to macroscopic parameters. Note that when v_f is large relative to the operating speed in congested conditions, the headway is increasingly dominated by the time-lag term, i.e., $h_{ij} \approx \tau_{ij}$; however, we do not require this approximation for the mappings below because Eq. (34) is retained explicitly.

2.5.2. CA discretization scales: Cell size Δx and time step Δt

In CA/CTM-type discretizations, the road is partitioned into cells of a constant size Δx, each holding at most one vehicle at jam density, and the evolution proceeds in constant time steps Δt. To avoid making Δx pattern-dependent, we adopt a representative jam-spacing scale as the cell length:

$$\Delta x \equiv \bar{d}, \tag{36}$$

i.e., the cell size is anchored by the mixture-average minimum spacing. Each pattern-specific minimum spacing d_{ij} is then expressed relative to this global discretization scale via

$$d_{ij} = \alpha_{ij}\,\Delta x, \qquad \alpha_{ij} = \frac{d_{ij}}{\Delta x}. \tag{37}$$

In a strictly discrete CA implementation, α_{ij} can be represented by an integer cell occupancy (e.g., $\lceil \alpha_{ij} \rceil$ cells) or by a consistent rounding convention; for analytical mapping purposes, the continuous ratio in Eq. (37) is sufficient.

Following the standard CA(M) normalization (one-cell movement at free flow), the time step is chosen as

$$\Delta t = \frac{\Delta x}{v_f}. \tag{38}$$

2.5.3. CA(M) "memory"/time-lag parameter

In Daganzo's CA(M), the dimensionless parameter $\gamma = s v_f / d$ governs how many time steps a vehicle "remembers" the leader's past position in the lattice rule (with s being the relevant time scale in the CA(M) formulation). Under our heterogeneous setting, each interaction pattern ij has its own effective (τ_{ij}, d_{ij}), and in our mapping this CA(M) time scale is represented by the pattern-specific effective time-lag τ_{ij}, yielding the pattern-level CA(M)-consistent parameter

$$\gamma_{ij} = \frac{\tau_{ij} v_f}{d_{ij}}. \tag{39}$$

Based on Eqs. (36)–(38), the number of reaction steps in the discrete-time lattice can be written as

$$n_{ij}^{\text{steps}} = \frac{\tau_{ij}}{\Delta t} = \frac{\tau_{ij} v_f}{\Delta x} = \alpha_{ij}\,\gamma_{ij}. \tag{40}$$

Eq. (39) provides the direct CA(M) mapping at the pattern level, while Eq. (40) clarifies how the same time-lag appears in a CA discretization with fixed $(\Delta x, \Delta t)$ and pattern-dependent spacing ratio α_{ij}. If a strictly discrete CA(M) simulation is implemented, n_{ij}^{steps} is represented as an integer number of time steps (via a chosen discretization convention), whereas for parameter mapping the continuous value suffices.

2.5.4. Pattern-specific and macroscopic wave speeds

In this subsection, we adopt the standard signed convention for the upstream (backward) congestion wave speed, i.e., $w < 0$. In kinematic-wave theory with a triangular fundamental diagram, the signed upstream congestion wave speed can be written as $w = -d/s$, where s is the relevant time scale in Daganzo's CA(M) formulation. In our mapping, this time scale is represented by the pattern-specific effective time-lag τ_{ij}, and accordingly each pattern ij admits a pattern-specific wave speed:

$$w_{ij} = -\frac{d_{ij}}{\tau_{ij}}. \tag{41}$$

Combining Eqs. (39) and (41) further yields

$$w_{ij} = -\frac{v_f}{\gamma_{ij}}, \tag{42}$$

which makes explicit that the CA(M) memory/advance parameter directly controls the magnitude of the upstream wave speed. These wave speeds characterize how quickly disturbances propagate upstream under each interaction type.

At the mixture level, under a homogeneous congested-regime approximation (all vehicles operate in the congested branch and the mixture is statistically stationary), the overall macroscopic wave speed can be approximated by the ratio of mixture-average spacing and time-lag:

$$w_{\text{macro}} \approx -\frac{\bar{d}}{\bar{\tau}} = -\frac{\sum_{ij} p_{ij} d_{ij}}{\sum_{ij} p_{ij} \tau_{ij}}. \tag{43}$$

Eq. (43) reduces to the classical $w = -d/\tau$ in the homogeneous case. We emphasize that Eq. (43) is an approximation whose accuracy depends on the degree of homogeneity of the congested state; in mixed traffic, the spatial distribution quantified by E affects p_{ij} and can lead to local deviations around this aggregate value.

2.5.5. CTM calibration: Capacity, wave speed, and jam density proxy

For CTM with a triangular fundamental diagram, the essential calibration parameters are the free-flow speed v_f, capacity c, backward wave speed w, and jam density k_j. Our framework provides these quantities as follows.

First, the mixed lane capacity is obtained directly from the mixture-average headway $c = \frac{1}{\bar{h}}$ as exhibited in Eqs. (31)–(32). Second, the backward wave speed is given by Eq. (43), i.e., $w \approx w_{\text{macro}}$. Third, a natural jam-density proxy follows from the mixture-average

Table 5
Micro-to-macro relationships between our model parameters and CTM/CA(M) parameters.

Quantities in our model	CTM/CA(M) quantities	Relationship
Pattern min spacing d_{ij} and mixture average $\bar{d}$	Cell size Δx	$\Delta x = \bar{d}$, $d_{ij} = \alpha_{ij}\Delta x$
Pattern headway h_{ij}	(τ_{ij}, d_{ij}) decomposition	$h_{ij} = \tau_{ij} + d_{ij}/v_f$
Pattern time-lag τ_{ij} and spacing d_{ij}	CA(M) parameter γ_{ij}	$\gamma_{ij} = \tau_{ij} v_f / d_{ij}$
Fixed $(\Delta x, \Delta t)$ with $\Delta t = \Delta x / v_f$	Discrete reaction steps n_{ij}^{steps}	$n_{ij}^{\text{steps}} = \tau_{ij}/\Delta t = \tau_{ij} v_f / \Delta x$
Pattern parameters (τ_{ij}, d_{ij})	Pattern wave speed w_{ij}	$w_{ij} = -d_{ij}/\tau_{ij}$
	Wave-speed–CA bridge	$w_{ij} = -v_f/\gamma_{ij}$
Pattern probabilities p_{ij}	Macroscopic wave speed w_{macro}	$w_{\text{macro}} \approx -\bar{d}/\bar{\tau}$
	CTM capacity c	$c = 1/\bar{h}$
	CTM jam density proxy k_j	$k_j \approx 1/\bar{d}$

minimum spacing:

$$k_j \approx \frac{1}{\bar{d}}. \tag{44}$$

Therefore, given (v_f, c, w, k_j), a triangular fundamental diagram can be specified for CTM, where the mixed-fleet heterogeneity and longitudinal ordering enter through (p_{ij}) and thus through $(\bar{h}, \bar{\tau}, \bar{d})$.

2.5.6. Summary of relationships

Table 5 summarizes the key micro-to-macro mappings used in this subsection.

2.5.7. Implications for capacity interpretation

These mappings clarify the micro-to-macro bridge provided by our framework. The headway parameters $\{h_{ij}\}$ determine the mixture-average headway $\bar{h}$ and thus the mixed lane capacity via Eq. (32). Meanwhile, the same underlying interaction parameters $\{\tau_{ij}, d_{ij}\}$ determine wave propagation at both the pattern level (Eqs. (41)–(42)) and the mixture level (Eq. (43)). Because the pattern probabilities p_{ij} are governed by (P_C, L, E) through the vehicle ordering model, the macroscopic CTM/CA(M) parameters calibrated by $(\bar{h}, \bar{\tau}, \bar{d})$ inherit both the heterogeneity of interactions and the stochasticity of the longitudinal arrangement.

3. Upper and lower bounds of the mixed lane capacity

This section theoretically investigates the upper and lower bounds of mixed lane capacity under any setting of related parameters (e.g., time headways, CAV penetration rate, maximum platoon size) by formulating mathematical programming models that distinguish the scenarios with and without consideration of the limited platoon size of CAVs. Moreover, a series of theoretical analyses on the spatial distribution of heterogeneous vehicles on the longitudinal lanes is carried out when the upper and lower bounds of the lane capacity are reached.

As stated in the introduction section, the spatial distribution of heterogeneous vehicles on the longitudinal lane directly determines the probability of different time headways in the mixed traffic flow, which further determines the exact value of mixed lane capacity. Due to the uncertain distribution of heterogeneous vehicles in the mixed traffic flow, the mixed lane capacity is a stochastic variable. Here, the calculation problem of the upper and lower bounds of the mixed lane capacity is regarded as the following vehicle arrangement problem in a mixed HV and CAV environment:

Definition 1. [Vehicle Arrangement Problem] With exogenously given the CAV penetration rate and the maximum platoon size, how to arrange HVs and CAVs on the longitudinal lane to achieve the maximum and minimum value of the mixed lane capacity?

3.1. Capacity bounds in the mixed traffic environment with unlimited platoon size

Without considering the limited platoon size of CAVs (i.e., $L = \infty$), car-following patterns can be directly classified according to the type of preceding and succeeding vehicles. It means that there are four types of car-following patterns, and each of them corresponds to a type of vehicle pair. Naturally, there is also a one-to-one mapping relationship between time headways and vehicle pairs. Therefore, the proportions of each vehicle pair in the mixed traffic flow (i.e., $\tau_{11}, \tau_{12}, \tau_{21}, \tau_{22}$) can be regarded as the decision variables of the proposed vehicle arrangement problem, which is used to represent the spatial distribution of heterogeneous vehicles. To simplify the model formulation, due to the reciprocal relationship between the average time headway and the mixed lane capacity, the objective function of the vehicle arrangement problem is defined as maximizing or minimizing the average time headway. Moreover, constraints are established based on the relationship between the proportions of different vehicle pairs and the CAV penetration rate. Finally, we have the following mathematical programming model and proposition.

$$[\text{CBNL}] \min_{\tau_{11},\tau_{12},\tau_{21},\tau_{22}} / \max \quad h_{HH}\tau_{11} + h_{HC}\tau_{12} + h_{CH}\tau_{21} + h_{CC}\tau_{22}, \tag{45}$$

subject to:

$$\tau_{11} + \tau_{12} = 1 - P_C, \tag{46}$$

$$\tau_{21} + \tau_{22} = P_C, \tag{47}$$

$$\tau_{11} + \tau_{21} = 1 - P_C, \tag{48}$$

$$\tau_{12} + \tau_{22} = P_C, \tag{49}$$

$$\tau_{11}, \tau_{12}, \tau_{21}, \tau_{22} \geqslant 0, \tag{50}$$

where Eq. (45) represents the average time headway. It is expressed as a function of different time headways and the proportions of different vehicle pairs. Based on a single-lane ring road, Eqs. (46)–(49) are established in Section 2.3.1. Constraint (50) means the proportion of different vehicle pairs should be nonnegative.

Proposition 2. *The upper and lower bounds of the mixed lane capacity without consideration of the limited platoon size can be obtained by solving the linear programming model* [*CBNL*].

Because the objective function and constraint conditions are both linear expressions of decision variables, the above mathematical programming model [*CBNL*] is a linear programming model. The four-by-four coefficient matrix of constraint equations consisting of Eqs. (46)–(49) is denoted as $\mathbf{M_1}$. After performing an elementary row transformation, we can find the rank of $\mathbf{M_1}$ is three. τ_{11}, τ_{12}, τ_{21} can be selected as basic variables, while τ_{22} is the non-basic variable. Then, the basic variables can be expressed as the following function of non-basic variables:

$$\begin{cases} \tau_{11} = 1 - 2P_C + \tau_{22}, \\ \tau_{12} = P_C - \tau_{22}, \\ \tau_{21} = P_C - \tau_{22}. \end{cases} \tag{51}$$

By substituting Eq. (51) into Eq. (45), the objective function can be transformed into a univariate function of τ_{22}. Moreover, according to Eqs. (50)–(51), we can obtain the new feasible region related to τ_{22}. Thus, the original linear programming model [*CBNL*] can be equivalently transformed into a new linear programming model [*CBNL-T*] as follows:

$$[\text{CBNL-T}] \quad \min / \max_{\max(0,2P_C-1) \leqslant \tau_{22} \leqslant P_C} \quad (h_{HH} + h_{CC} - h_{HC} - h_{CH})\tau_{22} + P_C(h_{HC} + h_{CH}) + h_{HH}(1 - 2P_C), \tag{52}$$

subject to:

$$\tau_{22} \geqslant \max(0, 2P_C - 1), \tag{53}$$

$$\tau_{22} \leqslant P_C. \tag{54}$$

It is evident that [*CBNL-T*] is a problem of finding the maximum and minimum of a univariate function. By considering the relationship between the coefficient of the objective function in Eq. (52) and zero, we can directly derive the value of decision variables when the objective function reaches the upper and lower bounds. Finally, the upper and lower bounds of mixed lane capacity and the corresponding spatial distribution of heterogeneous vehicles on the longitudinal lane can be obtained. Let $\overline{c}$ and $\underline{c}$ denote the upper and lower bounds of the mixed lane capacity, respectively.

(1) $h_{HH} + h_{CC} > h_{HC} + h_{CH}$

When the different time headways satisfy the above relationship, the mixed lane capacity $c(\cdot)$ is a monotonically decreasing function of τ_{22} according to Eq. (52). Therefore, the upper bound of the mixed lane capacity $\overline{c}$ is achieved when $\tau_{22} = [\max(0, 2P_C - 1)]$, and it has:

$$\overline{c} = \begin{cases} [P_C(h_{HC} + h_{CH}) + h_{HH}(1 - 2P_C)]^{-1}, & P_C \leqslant 0.5, \\ [(1 - P_C)(h_{HC} + h_{CH}) + (2P_C - 1)h_{CC}]^{-1}, & P_C > 0.5. \end{cases} \tag{55}$$

The lower bound of the mixed lane capacity $\underline{c}$ is achieved when $\tau_{22} = P_C$, which is equal to:

$$\underline{c} = [(1 - P_C)h_{HH} + P_C h_{CC}]^{-1}. \tag{56}$$

(2) $h_{HH} + h_{CC} < h_{HC} + h_{CH}$

When the different time headways satisfy the above relationship, the mixed lane capacity $c(\cdot)$ is a monotonically increasing function of τ_{22}. Therefore, the upper bound of the mixed lane capacity $\overline{c}$ achieved when $\tau_{22} = P_C$, and we have:

$$\overline{c} = [(1 - P_C)h_{HH} + P_C h_{CC}]^{-1}. \tag{57}$$

The lower bound of the mixed lane capacity $\underline{c}$ is achieved when $\tau_{22} = [\max(0, 2P_C - 1)]$:

$$\underline{c} = \begin{cases} [P_C(h_{HC} + h_{CH}) + h_{HH}(1 - 2P_C)]^{-1}, & P_C \leqslant 0.5, \\ [(1 - P_C)(h_{HC} + h_{CH}) + (2P_C - 1)h_{CC}]^{-1}, & P_C > 0.5. \end{cases} \tag{58}$$

(3) $h_{HH} + h_{CC} = h_{HC} + h_{CH}$

In this case, the mixed lane capacity is not a variable anymore but a constant, which is independent of the spatial distribution of heterogeneous vehicles on the longitudinal lane and equal to:

$$\overline{c} = \underline{c} = [(h_{HC} + h_{CH})(1 - P_C) + h_{CC}(2P_C - 1)]^{-1}. \tag{59}$$

Table 6
Upper and lower bounds of the mixed lane capacity with unlimited platoon size.

λ	P_C	Upper bound Vehicle distribution	Mixed lane capacity	Lower bound Vehicle distribution	Mixed lane capacity
$\lambda > 0$	$P_C \leqslant 0.5$	$\tau_{11} = 1 - 2P_C$, $\tau_{12} = P_C$, $\tau_{21} = P_C$, $\tau_{22} = 0$	$\left[P_C(h_{HC} + h_{CH}) + h_{HH}(1 - 2P_C)\right]^{-1}$	$\tau_{11} = 1 - P_C$, $\tau_{12} = 0$, $\tau_{21} = 0$, $\tau_{22} = P_C$	$\left[(1 - P_C)h_{HH} + h_{CC}P_C\right]^{-1}$
	$P_C > 0.5$	$\tau_{11} = 0$, $\tau_{12} = 1 - P_C$, $\tau_{21} = 1 - P_C$, $\tau_{22} = 2P_C - 1$	$\left[(1 - P_C)(h_{HC} + h_{CH}) + (2P_C - 1)h_{CC}\right]^{-1}$		
$\lambda = 0$	$0 \leqslant P_C \leqslant 1$	Any distribution*	$\left[(h_{HC} + h_{CH})(1 - P_C) + h_{CC}(2P_C - 1)\right]^{-1}$	Any distribution	$\left[(h_{HC} + h_{CH})(1 - P_C) + h_{CC}(2P_C - 1)\right]^{-1}$
$\lambda < 0$	$P_C \leqslant 0.5$	$\tau_{11} = 1 - P_C$, $\tau_{12} = 0$, $\tau_{21} = 0$, $\tau_{22} = P_C$	$\left[(1 - P_C)h_{HH} + h_{CC}P_C\right]^{-1}$	$\tau_{11} = 1 - 2P_C$, $\tau_{12} = P_C$, $\tau_{21} = P_C$, $\tau_{22} = 0$	$\left[P_C(h_{HC} + h_{CH}) + h_{HH}(1 - 2P_C)\right]^{-1}$
	$P_C > 0.5$			$\tau_{11} = 0$, $\tau_{12} = 1 - P_C$, $\tau_{21} = 1 - P_C$, $\tau_{22} = 2P_C - 1$	$\left[(1 - P_C)(h_{HC} + h_{CH}) + (2P_C - 1)h_{CC}\right]^{-1}$

*Any distribution indicates that the value of mixed lane capacity is a fixed constant, and thus the value of its upper and lower bounds is independent of the spatial distribution of heterogeneous vehicles on the longitudinal lane.

To enhance the readability, under any value of time headways and the CAV penetration rate, the upper and lower bounds of mixed lane capacity and the corresponding spatial distribution of heterogeneous vehicles are summarized in Table 6. For the convenience of description, the following abbreviation is adopted: $\lambda = h_{HH} + h_{CC} - h_{HC} - h_{CH}$. Accordingly, although the mixed lane capacity is a stochastic variable affected by various factors, we can directly know the exact maximum and minimum value of the mixed lane capacity under any setting of related parameters and how are HVs and CAVs distributed on the longitudinal lane.

We can find that the results listed in Table 6 are in line with the proposed lemmas in Section 2.4.2. Specifically, when $\lambda > 0$, the time headways satisfy the conditions in Lemma 6, which indicates the mixed lane capacity $c(\cdot)$ is a monotonically decreasing function of the CAV clustering intensity E. As shown in the first row of Table 6, it is indeed true that the CAV clustering intensity is minimum (i.e., $E = \max\left(0, \frac{2P_C - 1}{P_C}\right)$) according to Eq. (3) when the mixed lane capacity reaches its upper bound, and the CAV clustering intensity is maximum (i.e., $E = 1$) when the mixed lane capacity reaches its lower bound. On the other hand, when $\lambda < 0$, the time headways satisfy the conditions in Lemma 5, which means the mixed lane capacity $c(\cdot)$ is a monotonically increasing function of the CAV clustering intensity E. As shown in the third row of Table 6, it is also indeed true that the CAV clustering intensity is maximum (i.e., $E = 1$) when the mixed lane capacity reaches its upper bound, and the CAV clustering intensity is minimum (i.e., $E = \max\left(0, \frac{2P_C - 1}{P_C}\right)$) when the mixed lane capacity reaches its lower bound. Besides, we can find that when the relationship between λ and 0 is reversed, the mathematical expressions of the upper and lower bounds of the mixed lane capacity and the corresponding spatial distribution of heterogeneous vehicles on the longitudinal lane are interchanged. This phenomenon demonstrates that a comprehensive theoretical analysis of the mixed lane capacity under any setting of parameters is indeed necessary before drawing any qualitative conclusions.

3.2. Capacity bounds in the mixed traffic environment with limited platoon size

In this subsection, we analyze the upper and lower bounds of the mixed lance capacity in a more general scenario, where the limited platoon size of CAVs is taken into consideration. The upper and lower bounds of the mixed lane capacity are obtained by the method of theoretical analysis.

As stated in Section 2.1, compared with the scenario without consideration of the limited platoon size of CAVs, more types of car-following patterns and time headways exist in the mixed traffic flow. Because whether two adjacent CAVs belong to the same platoon is identified, there are five types of car-following patterns and time headways. Besides, all CAV platoons in the mixed traffic flow can be classified into saturated and unsaturated platoons according to whether the maximum platoon size is reached, Fig. 6 illustrates all possible combinations of CAV platoons with any size and adjacent vehicles on the longitudinal lane. It can be seen that all unsaturated platoons should be followed by an HV, while all saturated platoons can be followed by an HV or a CAV in the leading position of another CAV platoon. On the other hand, the preceding vehicle of CAV platoons with different sizes can be an HV or a saturated CAV platoon.

To distinguish CAV platoons with different sizes and model the spatial distribution of heterogeneous vehicles in Fig. 6, several new variables are introduced. Let y_i denote the probability of a CAV platoon of size i in the mixed traffic flow, $i \in \{1, 2 \dots L\}$. Especially, y_1 denotes the probability of a single CAV without platooning in the mixed traffic flow. y_L denotes the probability of a CAV platoon reaching the maximum allowable platoon size in the mixed traffic flow. Because the saturated platoon can be further classified into two categories according to the type of its succeeding vehicle: a saturated platoon followed by an HV, and a saturated platoon followed by a CAV. In the mixed traffic flow, the probability of the former is denoted as y_L^H, and the probability of the latter is denoted as y_L^C.

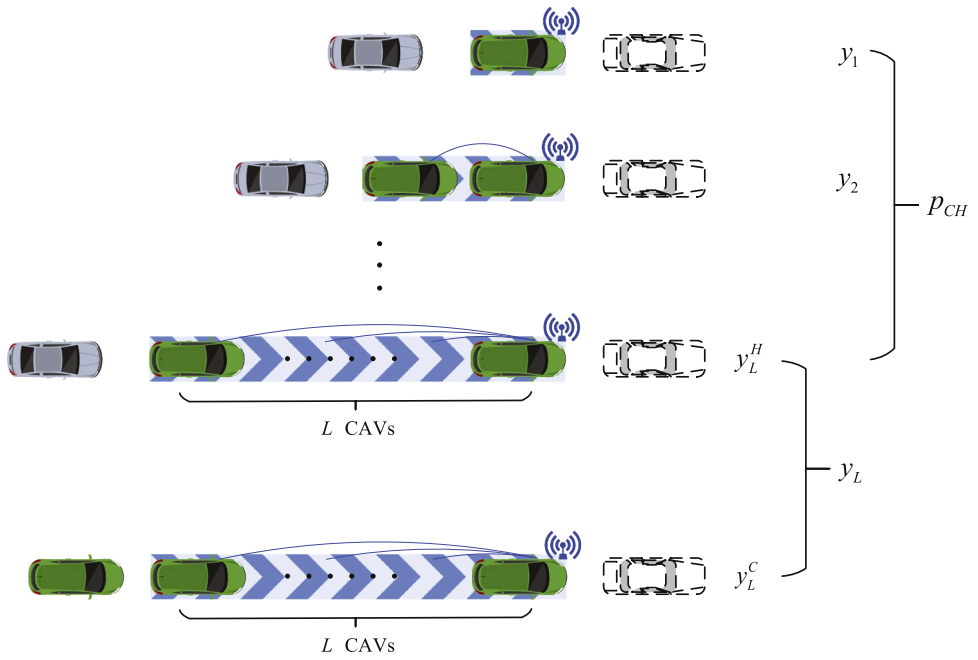


Fig. 6. Illustration of all possible combinations of CAV platoons with any size and adjacent vehicles on the longitudinal lane. The dotted vehicle contour represents either an HV or a saturated CAV platoon.

Thus, we have follow equation:

$$y_L^H + y_L^C = y_L.$$

Moreover, on a single-lane ring road, we have the following relationship between these new variables and the probability of different car-following patterns defined in Section 2.1. The probability of a saturated platoon followed by a CAV in the mixed traffic flow is equal to p_{CP}:

$$y_L^C = p_{CP}.$$

The sum of the product of the probability of the CAV platoon with different sizes and the corresponding number of "CC" car-following patterns in the CAV platoon is equal to p_{CC}, which can be expressed as follows:

$$y_2 + 2y_3 + \ldots + (L-2)y_{L-1} + (L-1)y_L = p_{CC}.$$

Since the unsaturated platoon must be followed by an HV, the sum of the probability of the unsaturated platoon and the probability of the saturated platoon followed by an HV is equal to p_{HC}:

$$y_1 + y_2 + \ldots + y_{L-1} + y_L - y_L^C = p_{HC}.$$

Finally, the vehicle arrangement problem in the mixed traffic environment with consideration of the limited platoon size can be formulated as the following mathematical programming problem.

$$[\text{CBL}] \quad \min_{p_{HH},p_{HC},p_{CH},p_{CP},p_{CC},y_1,\ldots,y_L,y_L^H,y_L^C} / \max \; h_{HH}p_{HH} + h_{HC}p_{HC} + h_{CH}p_{CH} + h_{CP}p_{CP} + h_{CC}p_{CC}, \tag{60}$$

subject to:

$$p_{HH} + p_{CH} = 1 - P_C, \tag{61}$$
$$p_{CP} + p_{CC} + p_{HC} = P_C, \tag{62}$$
$$p_{HH} + p_{HC} = 1 - P_C, \tag{63}$$
$$p_{CH} + p_{CP} + p_{CC} = P_C, \tag{64}$$
$$\sum_{i=2}^{L} (i-1)y_i = p_{CC}, \tag{65}$$
$$\sum_{i=1}^{L} y_i - y_L^C = p_{HC}, \tag{66}$$

$$y_L^C = p_{CP}, \tag{67}$$

$$y_L^H + y_L^C = y_L, \tag{68}$$

$$p_{HH}, p_{HC}, p_{CH}, p_{CP}, p_{CC}, y_1, y_2 \cdots y_{L-1}, y_L, y_L^H, y_L^C \geqslant 0, \tag{69}$$

where Eq. (60) denotes the average time headway. Constraints (61)–(64) are established in Section 2.3.1. Constraints (65)–(67) are explained in detail in the above content. Constraint (69) represents the probability of different car-following patterns should be nonnegative. We can find that [*CBL*] is a linear programming model and have the following proposition.

Proposition 3. *The upper and lower bounds of the mixed lane capacity with consideration of the limited platoon size can be obtained by solving the linear programming model* [*CBL*].

Let $\mathbf{X}$ denote the vector of decision variables, $\mathbf{X} = \left(p_{HH}, p_{HC}, p_{CH}, p_{CP}, p_{CC}, y_1, \ldots, y_L, y_L^H, y_L^C\right)^T$. $\mathbf{C}$ denotes the vector of the coefficients in the objective function, $\mathbf{C} = \left(c_1, c_2, \ldots, c_j, \ldots\right)$. c_j is the coefficient of jth decision variable, $j \in \{1, 2, \ldots, L+7\}$. $\mathbf{A}$ represents the coefficient matrix of constraints, $\mathbf{A} = \left(\mathbf{p_1}, \mathbf{p_2}, \ldots, \mathbf{p_j}, \ldots\right)$. $\mathbf{p_j}$ is the jth column of the coefficient matrix, $j \in \{1, 2, \ldots, L+7\}$. Then, [*CBL*] can be rewritten in the following vector form:

$$\min / \max \quad \mathbf{CX}, \tag{70}$$

subject to:

$$\mathbf{AX} = \mathbf{b}, \tag{71}$$

$$\mathbf{X} \geqslant \mathbf{0}. \tag{72}$$

3.3. Theoretical analysis

Before solving the linear programming model [*CBL*], theoretical analysis is carried out to discuss the upper and lower bounds of the mixed lane capacity and the corresponding spatial distribution of heterogeneous vehicles under different parameter settings as stated in Lemmas 7–9 and Corollary 1. The corresponding proofs can be found in Sections B.1–B.3.

Lemma 7. *If the different time headways satisfy the following conditions (I–II), all CAVs clustered together on the longitudinal lane can lead to the mixed lane capacity reaching its upper bound.*

I. $h_{HC} + h_{CH} - h_{HH} - h_{CP} \geqslant 0$,

II. $h_{CP} \geqslant h_{CC}$.

In contrast, if the different time headways satisfy the following conditions (III–IV), completely clustered CAVs can lead to the mixed lane capacity reaching its lower bound.

III. $h_{HC} + h_{CH} - h_{HH} - h_{CP} \leqslant 0$,

IV. $h_{CP} \leqslant h_{CC}$.

In any of the above cases, the mathematical expression of capacity bounds and the corresponding spatial distribution of heterogeneous vehicles are listed in the first row of Tables 7-8.

Lemma 8. *If the different time headways satisfy the following conditions (I–II), all CAVs forming maximum-size platoons and the platoons dispersed on the longitudinal lane as much as possible can lead to the mixed lane capacity reaching its upper bound.*

I. $h_{HH} + h_{CP} - h_{HC} - h_{CH} \geqslant 0$,

II. $h_{HH} + h_{CC} - h_{HC} - h_{CH} \leqslant 0$.

Otherwise, if the different time headways satisfy the following conditions (III–IV), the lower bound of mixed lane capacity will be achieved.

III. $h_{HH} + h_{CP} - h_{HC} - h_{CH} \leqslant 0$,

IV. $h_{HH} + h_{CC} - h_{HC} - h_{CH} \geqslant 0$.

In the above two cases, the mathematical expression of capacity bounds and the corresponding spatial distribution of heterogeneous vehicles are listed in the second row of Tables 7–8.

Lemma 9. *If the different time headways satisfy the following conditions (I–II), all CAVs avoiding forming large-scale platoons and the platoons dispersed on the longitudinal lane as much as possible can lead to the mixed lane capacity reaching its upper bound.*

I. $h_{HH} + h_{CC} - h_{HC} - h_{CH} \geqslant 0$,

II. $h_{CP} \geqslant h_{CC}$.

Otherwise, if the different time headways satisfy the following conditions (III–IV), the lower bound of mixed lane capacity will be achieved.

III. $h_{HH} + h_{CC} - h_{HC} - h_{CH} \leqslant 0$,

IV. $h_{CP} \leqslant h_{CC}$.

In the above two cases, the mathematical expression of capacity bounds and the corresponding spatial distribution of heterogeneous vehicles are listed in the third row of Tables 7–8.

Due to the similar proof procedure between Lemmas 7 and 8–9, the specific procedure for Lemmas 8–9 is not given. Besides, based on either Lemma 7 or Lemma 8, we can directly have the following corollary.

Table 7
Upper bound of the mixed lane capacity with limited platoon size.

Condition	Spatial distribution of CAVs	P_c	Decision variables related to the vehicle distribution	Upper bound
Lemma 7 (I–II)	All CAVs are clustered together	$0 \leqslant P_c \leqslant 1$	$p_{HH}=1-P_c,\ p_{HC}=0,\ p_{CH}=0,\ p_{CP}=\frac{P_c}{L},\ p_{CC}=\frac{L-1}{L}P_c,$ $y_1=0, y_2=0,\ldots,y_{L-1}=0, y_L=\frac{P_c}{L}, y_L^C=0, y_L^C=\frac{P_c}{L}$	$\left[(1-P_c)\,h_{HH}+\frac{P_c}{L}h_{CP}+\frac{L-1}{L}P_c h_{CC}\right]^{-1}$
Lemma 8 (I–II)	(1) All CAVs form maximum-size platoons as much as possible (2) All CAV platoons are dispersed on the longitudinal lane as much as possible	$P_c \leqslant \frac{L}{L+1}$	$p_{HH}=1-\frac{L+1}{L}P_c,\ p_{HC}=\frac{P_c}{L},\ p_{CH}=\frac{P_c}{L},\ p_{CP}=0,\ p_{CC}=$ $\frac{L-1}{L}P_c, y_1=0,\ldots,0, y_L=\frac{P_c}{L}, y_L^C=\frac{P_c}{L}, y_L^C=0$	$\left[\left(1-\frac{L+1}{L}P_c\right)h_{HH}+\frac{P_c}{L}\left(h_{HC}+h_{CH}\right)+\left(\frac{L-1}{L}P_c\right)h_{CC}\right]^{-1}$
		$P_c > \frac{L}{L+1}$	$p_{HH}=0,\ p_{HC}=1-P_c,\ p_{CH}=1-P_c,\ p_{CP}=\frac{L+1}{L}P_c-1,$ $p_{CC}=\frac{L-1}{L}P_c, y_1=0,\ldots,y_{L-1}=0,\ y_L=\frac{P_c}{L}, y_L^C=1-P_c,$ $y_L^C=\frac{L+1}{L}P_c-1$	$\left[(1-P_c)\,h_{HC}+(1-P_c)\,h_{CH}+\left(\frac{L+1}{L}P_c-1\right)h_{CP}+\frac{L-1}{L}P_c h_{CC}\right]^{-1}$
Lemma 9 (I–II)	(1) All CAVs avoid forming large-scale platoons as much as possible (2) All CAV platoons are dispersed on the longitudinal lane as much as possible	$P_c \leqslant \frac{1}{2}$	$p_{HH}=1-2P_c,\ p_{HC}=P_c,\ p_{CH}=P_c,\ p_{CP}=0,\ p_{CC}=0,$ $y_1=P_c, y_2=0,\ldots,y_{L-1}=0, y_L=\frac{P_c}{L}, y_L^C=0, y_L^C=0$	$\left[(1-2P_c)\,h_{HH}+P_c h_{HC}+P_c h_{CH}\right]^{-1}$
		$\frac{1}{2} < P_c \leqslant \frac{L}{L+1}$	$p_{HH}=0,\ p_{HC}=1-P_c, p_{CH}=1-P_c, p_{CP}=0,\ p_{CC}=2P_c-1,$ $y_1=0, y_2=0,\ldots,y_r=r+1-(r+2)\,P_c, y_{r+1}=0, y_{r+1}=$ $(r+1)\,P_c-r,\ldots, y_L^C=0,\ y_L^C=0$	$\left[(1-P_c)\,h_{HC}+(1-P_c)\,h_{HC}+(2P_c-1)\,h_{CC}\right]^{-1}$
		$P_c > \frac{L}{L+1}$	$p_{HH}=0,\ p_{HC}=1-P_c,\ p_{CH}=1-P_c,\ p_{CP}=\frac{L+1}{L}P_c-1,$ $p_{CC}=\frac{L-1}{L}P_c,\ y_1=0, y_2=0,\ldots,y_{L-1}=0, y_L=\frac{P_c}{L},$ $y_L^C=1-P_c, y_L^C=\frac{L+1}{L}P_c-1$	$\left[(1-P_c)\,h_{HC}+(1-P_c)\,h_{CH}+\left(\frac{L+1}{L}P_c-1\right)h_{CP}+\frac{L-1}{L}P_c h_{CC}\right]^{-1}$

Table 8
Lower bound of the mixed lane capacity with limited platoon size.

Condition	Spatial distribution of CAVs	P_c	Decision variables related to the vehicle distribution	Lower bound
Lemma 7 (III–IV)	All CAVs are clustered together	$0 \leqslant P_c \leqslant 1$	$p_{HH}=1-P_c,\ p_{HC}=0,\ p_{CH}=0,\ p_{CP}=\frac{P_c}{L},\ p_{CC}=\frac{L-1}{L}P_c,$ $y_1=0, y_2=0,\ldots,y_{L-1}=0, y_L=\frac{P_c}{L}, y_L^C=0, y_L^C=\frac{P_c}{L}$	$\left[(1-P_c)\,h_{HH}+\frac{P_c}{L}h_{CP}+\frac{L-1}{L}P_c h_{CC}\right]^{-1}$
Lemma 8 (III–IV)	(1) All CAVs form maximum-size platoons as much as possible (2) All CAV platoons are dispersed on the longitudinal lane as much as possible	$P_c \leqslant \frac{L}{L+1}$	$p_{HH}=1-\frac{L+1}{L}P_c,\ p_{HC}=\frac{P_c}{L},\ p_{CH}=\frac{P_c}{L},\ p_{CP}=0,\ p_{CC}=$ $\frac{L-1}{L}P_c, y_1=0,\ldots,0, y_L=\frac{P_c}{L}, y_L^C=\frac{P_c}{L}, y_L^C=0$	$\left[\left(1-\frac{L+1}{L}P_c\right)h_{HH}+\frac{P_c}{L}\left(h_{HC}+h_{CH}\right)+\left(\frac{L-1}{L}P_c\right)h_{CC}\right]^{-1}$
		$P_c > \frac{L}{L+1}$	$p_{HH}=0,\ p_{HC}=1-P_c,\ p_{CH}=1-P_c,\ p_{CP}=\frac{L+1}{L}P_c-1,$ $p_{CC}=\frac{L-1}{L}P_c,\ y_1=0,\ldots,y_{L-1}=0,\ y_L=\frac{P_c}{L}, y_L^C=1-P_c,$ $y_L^C=\frac{L+1}{L}P_c-1$	$\left[(1-P_c)\,h_{HC}+(1-P_c)\,h_{CH}+\left(\frac{L+1}{L}P_c-1\right)h_{CP}+\frac{L-1}{L}P_c h_{CC}\right]^{-1}$
Lemma 9 (III–IV)	(1) All CAVs avoid forming large-scale platoons as much as possible (2) All CAV platoons are dispersed on the longitudinal lane as much as possible	$P_c \leqslant \frac{1}{2}$	$p_{HH}=1-2P_c,\ p_{HC}=P_c,\ p_{CH}=P_c,\ p_{CP}=0,\ p_{CC}=0,$ $y_1=P_c, y_2=0,\ldots,y_{L-1}=0, y_L=\frac{P_c}{L}, y_L^C=0, y_L^C=0$	$\left[(1-2P_c)\,h_{HH}+P_c h_{HC}+P_c h_{CH}\right]^{-1}$
		$\frac{1}{2} < P_c \leqslant \frac{L}{L+1}$	$p_{HH}=0,\ p_{HC}=1-P_c, p_{CH}=1-P_c, p_{CP}=0,\ p_{CC}=2P_c-1,$ $y_1=0, y_2=0,\ldots,y_r=r+1-(r+2)\,P_c, y_{r+1}=0, y_{r+1}=$ $(r+1)\,P_c-r,\ldots, y_L^C=0,\ y_L^C=0$	$\left[(1-P_c)\,h_{HC}+(1-P_c)\,h_{HC}+(2P_c-1)\,h_{CC}\right]^{-1}$
		$P_c > \frac{L}{L+1}$	$p_{HH}=0,\ p_{HC}=1-P_c,\ p_{CH}=1-P_c,\ p_{CP}=\frac{L+1}{L}P_c-1,$ $p_{CC}=\frac{L-1}{L}P_c,\ y_1=0, y_2=0,\ldots,y_{L-1}=0, y_L=\frac{P_c}{L},$ $y_L^C=1-P_c, y_L^C=\frac{L+1}{L}P_c-1$	$\left[(1-P_c)\,h_{HC}+(1-P_c)\,h_{CH}+\left(\frac{L+1}{L}P_c-1\right)h_{CP}+\frac{L-1}{L}P_c h_{CC}\right]^{-1}$

Corollary 1. *If the different time headways satisfy the following conditions (I–II), the mixed lane capacity is no longer a variable but a constant, which is independent of the spatial distribution of heterogeneous vehicles on the longitudinal lane.*

I. $h_{HC}+h_{CH}-h_{HH}-h_{CP}=0$,

II. $h_{CP}=h_{CC}$.

It is worth noting that this study focuses on the longitudinal motion of vehicles on a single lane, under which the capacity bounds are analytically derived. While lateral maneuvers such as lane-changing are not explicitly modeled, they can provide a relevant physical interpretation for the capacity reductions observed in practice. As shown by Laval and Daganzo (2006), lane changes generate local voids and temporary moving bottlenecks that reduce discharge rates near bottlenecks. Within the framework of this study, such effects do not modify the analytically derived capacity bounds, which are obtained from the vehicle ordering problem under given car-following headways. Instead, lane-changing can be understood as a mechanism that prevents traffic systems from realizing the idealized vehicle arrangements that attain these bounds, leading to observed throughputs below the theoretical limits.

4. Validation and numerical analysis

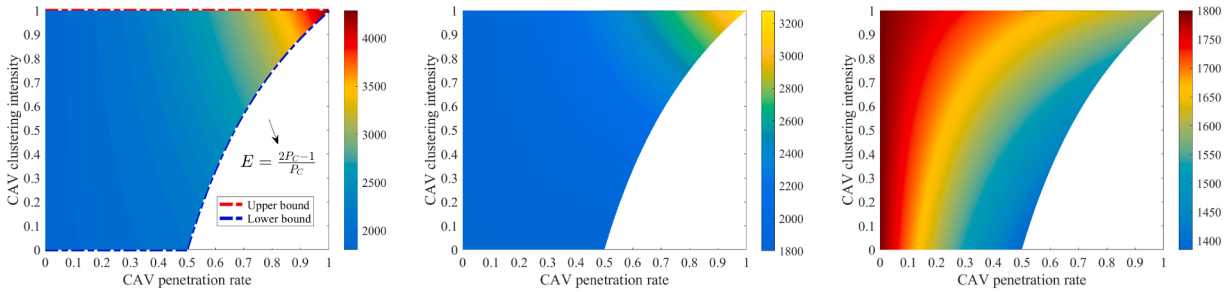
4.1. Safety time headways

Although the exact value of headways is not required in the above theoretical research, several sets of headway parameter values are given for the convenience of conducting numerical experiments to demonstrate the effectiveness of the proposed methodology. To make sure the value of different headways in the numerical experiments is representative, we classify three technological scenarios and the exact value of different headways in each technological scenario is given in Table 9 (Ghiasi et al., 2017). The unit for headway is seconds. Each scenario represents one development stage of CAV technology from birth to maturity as follows:

- *Aggressive scenario:* The CAV technology is already mature and has gained widespread trust from people. The real-time communication and cooperation between CAVs can contribute to reducing the headway, especially when they drive in the form of CAV platoons. Moreover, people all believe CAVs have superior performance in driving than humans. Because of the unpredictable and unstable driving behavior of human drivers, CAVs prefer to maintain a larger headway when following an HV during driving, and people who follow a CAV may prefer a more compact headway than following an HV.
- *Moderate scenario:* With the development of CAV technology, CAVs have the same performance in driving as human drivers. HVs and CAVs maintain the same time headway with the preceding vehicle. Moreover, real-time communication and cooperation between CAVs play a positive role in reducing headways, especially for vehicles in CAV platoons.

Table 9
Exact value of headways in different technological scenarios.

Limited platoon size	Scenarios	h_{HH}	h_{HC}	h_{CH}	h_{CC}	h_{CP}
×	Aggressive	2.0	1.2	1.0	0.8	–
	Moderate	2.0	2.0	2.0	1.0	–
	Conservative	2.0	2.4	2.8	2.2	–
✓	Aggressive	2.0	1.8	1.6	0.8	1.0
	Moderate	2.0	2.0	2.0	1.0	1.5
	Conservative	2.0	2.4	2.8	2.2	2.5



(a) Aggressive scenario (b) Moderate scenario (c) Conservative scenario

Fig. 7. Change in the mixed lane capacity with variations of the CAV penetration rate and the CAV clustering intensity under three technological scenarios ($L = 5$).

- *Conservative scenario:* In the early stage of the development of CAVs, the CAV technology is not mature yet. CAVs have a bad performance in driving compared with human drivers, and the immature CAV technology has not gained people's trust either. Human drivers usually behave more conservatively when following a CAV. Despite the bad performance of CAVs in autonomous driving, we believe that real-time communication and cooperation still play a positive role in reducing headways.

To facilitate comparison with macroscopic CTM/CA(M) formulations, the three technological scenarios in Table 9 (aggressive/moderate/conservative) can be mapped to CTM calibration parameters and CA(M) advancement (memory) constraints using the micro–macro relationships established in Section 2.5. For each scenario $s \in \{\mathrm{Agg}, \mathrm{Mod}, \mathrm{Con}\}$, the pattern-dependent headways $\{h_{ij}^{(s)}\}$ admit the decomposition $h_{ij}^{(s)} = \tau_{ij}^{(s)} + d_{ij}^{(s)}/v_f$. Given a chosen minimum spacing $d_{ij}^{(s)}$ (scenario-/pattern-specific or a representative value), the corresponding time-lag follows as $\tau_{ij}^{(s)} = h_{ij}^{(s)} - d_{ij}^{(s)}/v_f$, and thus determines the CA(M)-consistent parameter $\gamma_{ij}^{(s)} = \tau_{ij}^{(s)} v_f / d_{ij}^{(s)}$ that governs how restrictive the discrete advancement constraint is for interaction type ij. Importantly, the CA(M) update rule itself remains deterministic; the uncertainty arises from the stochastic vehicle ordering, so γ is a pattern-dependent random variable taking values $\gamma_{ij}^{(s)}$ with mixing weights p_{ij} governed by (P_C, L, E).

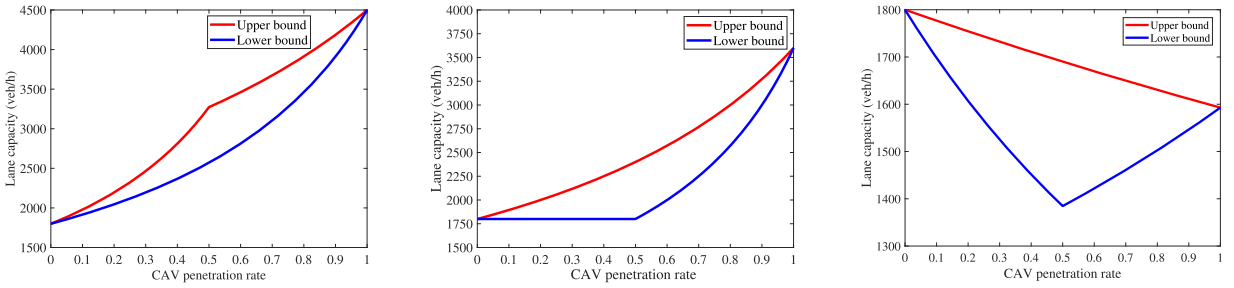
At the macroscopic level, using the same p_{ij}, define mixture averages $\bar{h}^{(s)} = \sum_{ij} p_{ij} h_{ij}^{(s)}$, $\bar{\tau}^{(s)} = \sum_{ij} p_{ij} \tau_{ij}^{(s)}$, and $\bar{d}^{(s)} = \sum_{ij} p_{ij} d_{ij}^{(s)}$. These provide direct triangular fundamental diagram CTM calibrations via $c^{(s)} = 1/\bar{h}^{(s)}$, $w^{(s)} \approx -\bar{d}^{(s)}/\bar{\tau}^{(s)}$, and (optionally) $k_j^{(s)} \approx 1/\bar{d}^{(s)}$. Therefore, the aggressive/moderate/conservative scenarios correspond to distinct, internally consistent CTM parameter sets while simultaneously determining the mixed distribution of CA(M) advancement constraints through $\{\gamma_{ij}^{(s)}\}$ and $\{p_{ij}\}$.

4.2. Explicit expression of mixed lane capacity

To intuitively exhibit the effect of various factors on the mixed lane capacity, Fig. 7 illustrates the variation of the mixed lane capacity against the CAV penetration rate and the CAV clustering intensity in different technological scenarios. The values of different time headways under three technological scenarios are referenced in Table 9, and the maximum platoon size is set to 5, i.e., L = 5 (Yu et al., 2023). The dashed lines in Fig. 7(a) represent the maximum and minimum value of CAV clustering intensity varying with the CAV penetration rate, which is in line with Proposition 1. The unit for capacity is veh/h.

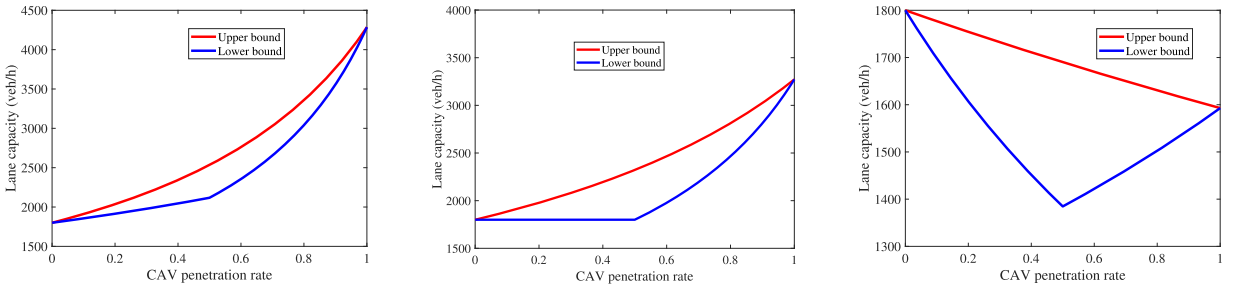
As depicted in Fig. 7, the mixed lane capacity keeps increasing with the increase in CAV penetration rates and CAV clustering intensity under aggressive and moderate scenarios, while it keeps decreasing as the CAV penetration rate increases under the conservative scenario. This phenomenon contrasts with the traditional view that mixed lane capacity always increases monotonically with the CAV penetration rate. It indicates that we'd better not draw any subjective conclusion before determining the exact value of time headways corresponding to different types of car-following patterns.

After giving the exact value of different time headways under three technological scenarios (i.e., Table 9), the change in the mixed lane capacity with the variations of various factors is preliminarily explored by conducting experiments in Section 2.4.1 and dedicated in Fig. 7. It can be checked that the results of numerical experiments in Section 2.4.1 are in line with the proposed lemmas in this section. Specifically, the time headways under the aggressive scenario satisfy the conditions in Lemma 1, which indeed leads to a monotonically increasing relationship between the mixed lane capacity and the CAV penetration rate as depicted in Fig. 7(a). The



(a) Aggressive scenario (b) Moderate scenario (c) Conservative scenario

Fig. 8. Variation of the mixed lane capacity bounds against the CAV penetration rate with unlimited platoon size..



(a) Aggressive scenario (b) Moderate scenario (c) Conservative scenario

Fig. 9. Variation of the mixed lane capacity bounds against the CAV penetration rate with limited platoon size ($L = 5$).

time headways under aggressive and moderate scenarios satisfy the conditions in Lemma 5, which indeed leads to a monotonically increasing relationship between the mixed lane capacity and the CAV clustering intensity as depicted in Fig. 7(a)–(b). Besides, from the lemmas above, it can be found that opposite conclusions can be derived based on different values of time headways corresponding to different car-following patterns. It is indeed necessary to carry out a comprehensive theoretical analysis of the mixed lane capacity before drawing any qualitative conclusions.

4.3. Upper and lower bounds of the mixed lane capacity

(1) Capacity bounds in the mixed traffic environment with unlimited platoon size

It can be checked that the time headways in the aggressive scenario satisfies $h_{HH} + h_{CC} > h_{HC} + h_{CH}$, while the time headways in the moderate and conservative scenarios $h_{HH} + h_{CC} < h_{HC} + h_{CH}$. By varying the CAV penetration rate from 0 to 1 with a step size of 0.02, a total of 51 sets of numerical experiments are conducted. After directly solving the linear programming model [*CBNL*], Fig. 8 intuitively illustrates the change of upper and lower bounds of the mixed lane capacity with the increase of the CAV penetration rate in the three technological scenarios with unlimited platoon sizes. The stochastic mixed lane capacity always varies within the range from the lower bound to the upper bound due to the uncertain distribution of heterogeneous vehicles on the longitudinal lane.

(2) Capacity bounds in the mixed traffic environment with limited platoon size

The maximum platoon size is set to 5 (i.e., $L = 5$). Similarly, by varying the CAV penetration rate from 0 to 1, the value of capacity bounds in the mixed traffic environment with limited platoon size can be calculated by directly solving the linear programming model [*CBL*]. Finally, the variation of the mixed lane capacity bounds against the CAV penetration rate with the limited platoon size is intuitively illustrated in Fig. 9.

The sometimes wide range between the upper and lower bounds of capacity (as exhibited in Figs. 8–9) is not merely a theoretical outcome but a critical practical insight. It quantifies the system sensitivity to the spatial organization of vehicles. Traffic planners can use the lower bound for robust, risk-averse design (ensuring minimum service levels), while the upper bound indicates the potential benefit of optimal coordination. The difference between them defines the manageable capacity gain achievable through traffic management strategies that influence vehicle distribution, such as platoon coordination or lane management.

Besides, to provide a more intuitive comparison of how imposing a maximum platoon size affects mixed-lane capacity, we quantify the relative differences between the capacity bounds obtained under an unlimited platoon size and those obtained under a finite platoon size in the aggressive scenario. Specifically, for each P_C, we report the percentage differences of $\overline{c}$ and $\underline{c}$ between the unlimited-platoon case and the finite-platoon baseline L = 5, using the L = 5 values as denominators. Fig. 10 plots these percentage differences for $\overline{c}$ and $\underline{c}$ across varying P_C.

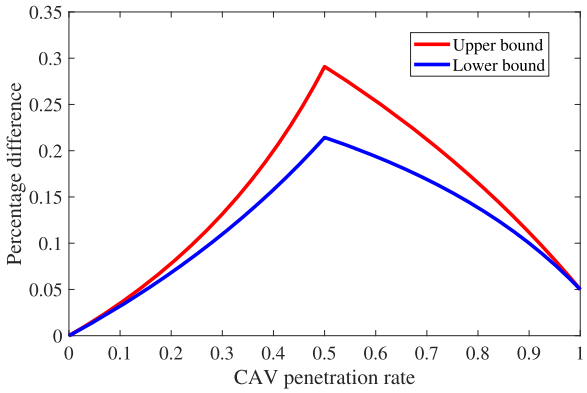


Fig. 10. Percentage differences of the upper bound $\bar{c}$ and lower bound $\underline{c}$ between the unlimited-platoon case and the finite-platoon baseline $L = 5$, computed as $(C_{L=\infty} - C_{L=5})/C_{L=5} \times 100\%$ for each bound.

Fig. 10 shows that the percentage differences between the unlimited-platoon case and the finite-platoon baseline ($L = 5$) are non-monotonic with respect to the CAV penetration rate P_C. For both the upper bound $\bar{c}$ and the lower bound $\underline{c}$, the gap first increases and then decreases as P_C increases, reaching its maximum at $P_C = 50\%$. At this penetration level, the upper-bound estimate differs by up to 29.1%, and the lower-bound estimate differs by up to 21.4% (both computed relative to the $L = 5$ denominators). These results indicate that adopting the unlimited-platoon assumption can noticeably overestimate the capacity bounds in mixed traffic, particularly at moderate CAV penetration rates.

4.4. Numerical characteristics of the mixed lane capacity

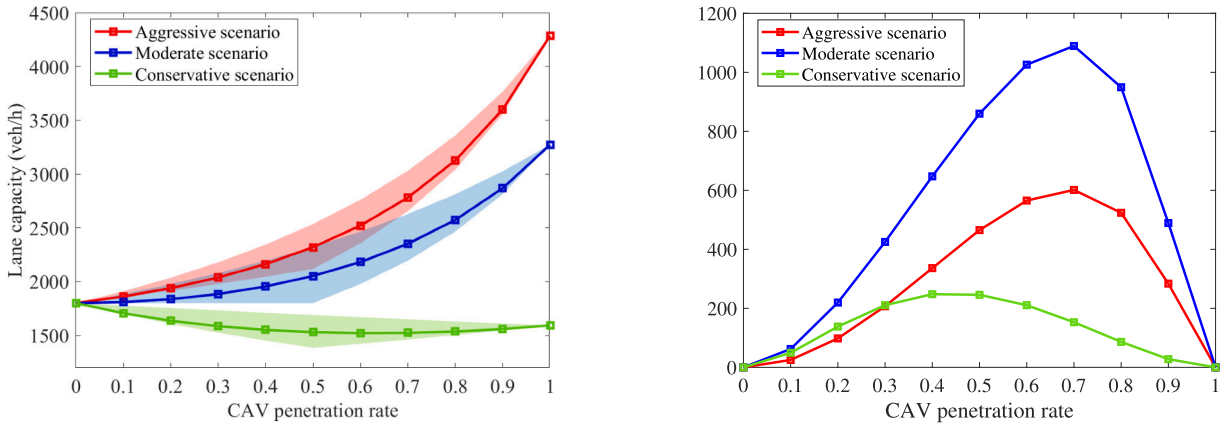
In this section, a series of numerical experiments are conducted to analyze the numerical characteristics of mixed lane capacity, i.e., mean and variance. By assuming the spatial distribution of heterogeneous vehicles on the longitudinal lane is completely random, the mathematical statistics method based on the numerical simulation is used to explore the stochastic properties of mixed lane capacity. The specific procedure of numerical simulation can refer to Algorithm 1. The parameters of simulation experiments are set as follows. The values of different time headways in three technological scenarios are referenced in Table 9. The maximum platoon size is set to 5 (i.e., L = 5), and the sample size of simulation experiments is 100,0000. Multiple potential vehicle spatial distributions are generated through simulation experiments, and statistical analysis of the simulation results is used to estimate the numerical characteristics of the mixed lane capacity. A total of 10,000 samplings are conducted.

Algorithm 1 Pseudo-code for the method of numerical simulation.

Input: Sample size W, total number of vehicles N, CAV penetration rate P_C, maximum platoon size L, value of different time headways $\boldsymbol{h}$, $\boldsymbol{h} = \{h_{HH}, h_{HC}, h_{CH}, h_{CP}, h_{CC}\}$
Output: Value of upper and lower bounds of the mixed lane capacity, i.e., $\bar{c}$, $\underline{c}$

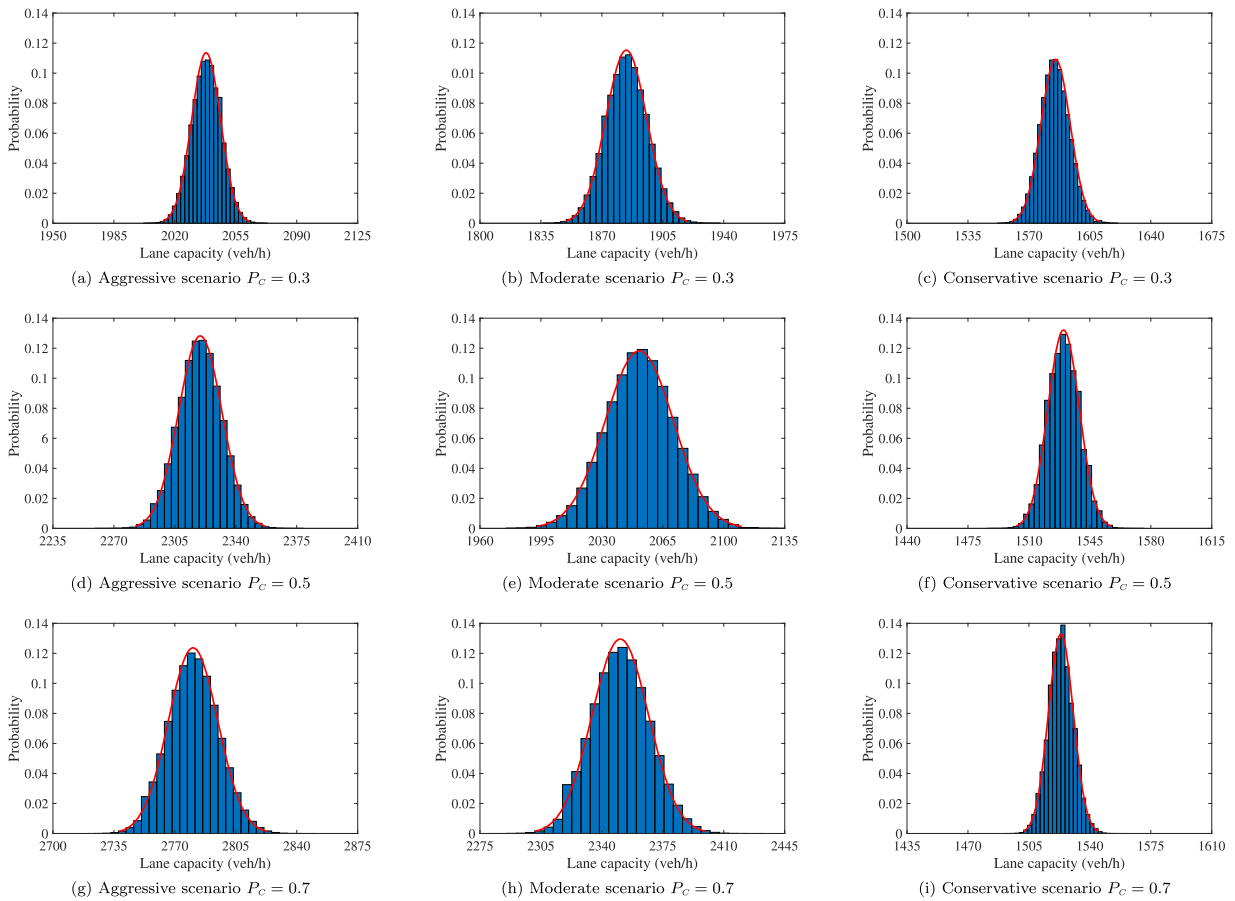
for $w = 1$ to W **do**
  $n \leftarrow 0$
  Step 1: Generate N vehicles with the CAV penetration rate P_C
  Step 1.1: Generate a pure HV platoon containing N vehicles, i.e., generate an initial array containing N zero elements
  Step 1.2: Randomly select NP_C HVs from the initial platoon and replace these HVs with CAVs, i.e., randomly choose NP_C positions from the initial array and change the element values at those positions from 0 to 1
  Step 2: Calculate the value of mixed lane capacity for the sample w
  while $n \leqslant N$ **do**
    Step 2.1: View the vehicle sequence as a head-to-tail ring and judge the car-following patterns between the nth and $n+1$th vehicle in turn based on the classification rules in Fig. 1
    Step 2.2: $n \leftarrow n + 1$
  end while
  Step 2.3: Obtain the proportions of different time headways $\boldsymbol{p}_w$
  Step 2.4: Calculate the value of mixed lane capacity $c_w(\boldsymbol{p}_w, \boldsymbol{h})$
end for

Based on the results of the samplings, we can obtain the sample mean and sample variance directly. Then, the population mean is represented by the sample mean, and the population variance is represented by the average of sample variances. For each technological scenario, eleven sets of experiments are conducted with the CAV penetration rate ranging from 0 to 1 in increments of 0.1. Fig. 11



(a) Mean of the mixed lane capacity (b) Variance of the mixed lane capacity

Fig. 11. Mean and variance of the mixed lane capacity with the increase of CAV penetration rate under the AH coordination.



(a) Aggressive scenario $P_C = 0.3$ (b) Moderate scenario $P_C = 0.3$ (c) Conservative scenario $P_C = 0.3$

(d) Aggressive scenario $P_C = 0.5$ (e) Moderate scenario $P_C = 0.5$ (f) Conservative scenario $P_C = 0.5$

(g) Aggressive scenario $P_C = 0.7$ (h) Moderate scenario $P_C = 0.7$ (i) Conservative scenario $P_C = 0.7$

Fig. 12. Probability distribution of the mixed lane capacity.

intuitively illustrates the variation of the mean and variance of the mixed lane capacity with the increase of CAV penetration rate in three technological scenarios. The shaded areas represent the varying range of mixed lane capacity, which is determined by the lower and upper bounds in Section 3.2.

As depicted in Fig. 11(a), the mean of mixed lane capacity keeps increasing with the increase of the CAV penetration rate in all technological scenarios except the conservative scenario. This is because of the over-cautious value of time headways in the conservative scenario. However, as the CAV penetration rate increases, the variances of mixed lane capacity in all technological

scenarios exhibit the same varying tendency. They all increase at first and then decrease as shown in Fig. 11(b). We can also find that the variance of mixed lane capacity in the moderate scenario is the largest, followed by the aggressive scenario, and then the conservative scenario. Because three technological scenarios represent different development stages of CAV technology from birth to maturity successively as stated in Section 2.1, the above phenomenon indicates although the development of CAV technology leads to a comprehensive enhancement of mixed lane capacity, a bigger fluctuation of mixed lane capacity is also followed by. Moreover, more attention should be paid to the moderate scenario, where the variance of mixed lane capacity is the largest. At that time, individuals will suffer huge uncertainty in travel time induced by the mixed HV and CAV environment.

Besides the mean and variance, the probability distribution of mixed lane capacity can reveal more information on its stochastic properties. Fig. 12 exhibits the variation of the probability distribution of mixed lane capacity against the CAV penetration rate in three technological scenarios. To save space, three representative levels of CAV penetration rate are chosen with values equal to 0.3, 0.5, and 0.7, respectively.

5. Conclusions and future research

This study provides a generalized theoretical framework for analyzing lane capacity in a mixed HV and CAV environment, where capacity is widely recognized as a complex stochastic variable influenced by multiple interacting factors. To explicitly capture the relationship between mixed lane capacity and the longitudinal distribution of heterogeneous vehicles, an index termed CAV clustering intensity is proposed. This index has a clear physical interpretation and an explicit analytical expression, enabling a quantitative characterization of vehicle arrangements without relying on simplified assumptions regarding car-following patterns or spatial distributions. Accordingly, a Markov chain model is established to theoretically derive the explicit expression of the mixed lane capacity, which is mathematically expressed as a multivariate function of CAV penetration rate, maximum platoon size, and CAV clustering intensity. As a generalized mathematical expression for calculating mixed lane capacity, the calculation formulas in many previous studies can be regarded as special cases of this study. It has also been demonstrated that the proposed CAV clustering intensity is an effective index representing the spatial distribution of heterogeneous vehicles in the calculation of mixed lane capacity. Moreover, the effect of various factors on the mixed lane capacity is explored by rigorously theoretical analysis. We find that opposite conclusions can be derived based on different values of time headways. It is indeed necessary to carry out a comprehensive theoretical analysis of the mixed lane capacity before drawing any qualitative conclusions. To explore the exact upper and lower bounds of the mixed lane capacity and the corresponding vehicle distribution under any setting of related parameters, a vehicle arrangement problem is proposed and formulated as two linear programming models in the scenario with unlimited and limited platoon sizes. Finally, various numerical characteristics of the stochastic lane capacity are calculated and employed to reveal its stochastic properties comprehensively.

Finally, it is important to clarify the role of the proposed analytical framework in relation to computational and discrete network-based modeling approaches. These two paradigms are not substitutes but are inherently complementary. The analytical model developed in this study aims to deliver explicit and interpretable relationships between mixed traffic capacity and key structural factors, offering closed-form results such as monotonicity properties, capacity bounds, and stochastic characteristics. Such mechanistic insights are difficult to isolate using purely simulation-based methods. In contrast, computational and network-level models are better suited for capturing large-scale interactions, complex topologies, control strategies, and detailed operational dynamics, typically through calibrated parameters and numerical experiments. From this perspective, the proposed analytical framework can serve as a theoretical benchmark and reference envelope for simulation-based studies, while computational models can extend the analysis to network-level phenomena beyond analytical tractability. Together, these approaches contribute to a more comprehensive understanding of mixed HV and CAV traffic systems.

Further research work can be undertaken in several aspects. These directions can be broadly grouped into (i) application-oriented extensions for planning and control, and (ii) model-oriented extensions that improve realism and generality. First, deploying dedicated CAV lanes or CAV/toll lanes on the road is an efficient way to improve traffic efficiency because it can entirely or partially separate CAVs from the mixed traffic flow. Therefore, we can examine the effect of these planning strategies on the stochastic properties of the lane capacity in the future. In particular, it would be valuable to quantify how dedicated-lane policies reshape the attainable CAV clustering intensity and the corresponding realizability of near-optimal vehicle arrangements under operational constraints. Second, the explicit bounds of stochastic mixed traffic capacity derived in this study can provide useful benchmark limits and quantitative inputs for robust infrastructure planning and traffic management under uncertainty. For example, these bounds can be incorporated into robust optimization models or microscopic simulation platforms (e.g., the cell-based virtual-track networks proposed by Lu and Zhou (2023)) to design resilient strategies such as CAV-dedicated lane deployment and adaptive intersection control (e.g., conflict-zone/cell-based control at intersections). In such cell-based and time-expanded formulations, our analytical upper/lower bounds of mixed lane capacity can serve as benchmark throughput limits and can be used to help define or calibrate the corresponding flow constraints, while operational rules such as lane-changing feasibility and conflict-zone exclusive-occupancy constraints determine whether the vehicle arrangements that attain these bounds are realizable in practice. Importantly, realizing this linkage requires translating our headway-based bounds into the discretized cell/time-step setting and, for intersections, specifying any additional conflict-zone clearance/separation terms under a chosen geometry and safety policy. Accordingly, our bounds can help parameterize (rather than uniquely determine) flow constraints in such discrete network models. Third, extending the proposed framework to model traffic streams with multiple sub-classes of CAVs (e.g., differing in capabilities or cooperation protocols) presents a valuable but complex future direction. This would require a generalized state space, a structured set of inter-vehicular interaction rules, and likely a family of indices to characterize multi-type clustering, posing significant analytical challenges beyond the two-class case studied here.

Fourth, the current model assumes perfect and instantaneous communication. Incorporating the effects of stochastic communication delays and packet loss (e.g., via advancement rules like CA(m)) on platoon stability and effective safe headways is a critical next step for assessing the realizability of the derived capacity under realistic V2V/V2I conditions. Such an extension would require explicit assumptions on delay/loss statistics and how they modify effective headways, and is therefore left for future work. Fifth, since several car-following models and lane-changing models have been proposed to formulate the performance of HVs and CAVs in the mixed traffic environment, we can further analyze the active platooning behavior of CAVs from the perspective of microscopic traffic flow and thus explore the spatial distribution of heterogeneous vehicles in detail. In particular, while lane-changing is not modeled in the present single-lane longitudinal framework, extending it to multi-lane settings requires explicitly modeling lane-changing as a coupling mechanism that reshapes the longitudinal vehicle ordering (and thus the observed CAV clustering intensity E) by continuously altering adjacent vehicle pairs. A promising direction is to integrate our macroscopic CAV clustering intensity index with high-fidelity microscopic simulators (such as the virtual-track network), where the clustering intensity can serve as an input parameter to initialize vehicle arrangements and evaluate the realizability of assumed spatial distributions under operational constraints. This provides a direct way to examine how lane-changing dynamics and intersection conflict constraints affect the attainable ordering patterns (e.g., the implied τ_{22} and E) and consequently the extent to which the analytically derived conditional capacity bounds can be achieved. Finally, beyond capacity analysis, the proposed CAV clustering intensity can serve as a foundational tool for investigating a wider range of traffic flow properties in future studies, including stability, safety, and energy efficiency in mixed traffic environments. Furthermore, it can inform the design of active platooning strategies-for example, by revealing how intentional clustering of CAVs (controlled via V2V communication) can be optimized to enhance road capacity, with the optimized strategies being tested and implemented in discrete network simulation and control frameworks that explicitly represent lane-changing and intersection conflict-zone constraints.

CRediT authorship contribution statement

Tongfei Li: Conceptualization, Methodology, Validation, Formal analysis, Writing – original draft, Writing – review & editing, Supervision, Project administration; **Hongfei Zhu:** Software, Formal analysis, Investigation, Writing – original draft, Visualization; **Min Xu:** Validation, Resources, Writing – original draft, Writing – review & editing, Funding acquisition; **Huijun Sun:** Resources, Writing – review & editing, Funding acquisition.

Data availability

Data will be made available on request.

Declaration of competing interest

None

Acknowledgments

This work was supported by the Humanities and Social Science Fund of Ministry of Education of China (24YJAZH068), the National Natural Science Foundation of China (72288101, 71901007), and the Research Grants Council of the Hong Kong Special Administrative Region, China (Project No. PolyU 15224824).

Appendix A. Effect of various factors on the mixed lane capacity

A.1. Proof of Lemma 1

Proof. Due to the reciprocal relationship between the average time headway and the mixed lane capacity, we first analyze the effect of the CAV penetration rate P_C on the average time headway $\bar{h}$ with the exogenously given maximum platoon size L and CAV clustering intensity E, which is discussed separately in two cases, i.e., $E = 1$ and $E \neq 1$. As shown in Eqs. (26)–(33), the average time headway in these two cases corresponds to different mathematical expressions.

(1) In the case of $E = 1$, the first derivative of $\bar{h}$ with respect to P_C can be expressed as follows:

$$\frac{\mathrm{d}\bar{h}}{\mathrm{d}P_C} = \frac{h_{CP} + (L-1)h_{CC} - Lh_{HH}}{L}. \tag{A.1}$$

If $h_{CP} \geqslant h_{CC}$ and $h_{HH} > h_{CP}$ as stated in Lemma 1 (II), we can know:

$$\frac{\mathrm{d}\bar{h}}{\mathrm{d}P_C} = \frac{h_{CP} + (L-1)h_{CC} - Lh_{HH}}{L} \leqslant h_{CP} - h_{HH} < 0.$$

Accordingly, since the mixed lane capacity is the reciprocal of the average time headway, we can derive that the mixed lane capacity is a monotonically increasing function of the CAV penetration rate.

(2) In the case of $E \neq 1$, the first derivative of $\bar{h}$ with respect to P_C is derived as follows:

$$\frac{\mathrm{d}\bar{h}}{\mathrm{d}P_C} = \frac{\partial \bar{h}}{\partial E}\frac{\partial E}{\partial P_C} + \frac{\partial \bar{h}}{\partial P_C} \tag{A.2}$$

$$= \left[\frac{\left(-(E)^L + LE^{L-1} - L(E)^L + E^{2L}\right)P_C h_{CP} + \left(1-(E)^L - LE^{L-1} + L(E)^L\right)P_C h_{CC}}{\left(1-(E)^L\right)^2} - P_C\left(h_{HC}+h_{CH}\right) + P_C h_{HH}\right] \tag{A.3}$$

$$\left(\frac{E}{P_C}\right) + \frac{(1-E)(E)^L h_{CP} + E\left[1-(E)^{L-1}\right]h_{CC}}{1-(E)^L} - E\left(h_{HC}+h_{CH}\right) + (E-2)h_{HH}. \tag{A.4}$$

Because $E \in \left[\max\left(0, \frac{2P_C-1}{P_C}\right), 1\right)$ as described in Proposition 1 and $h_{CP} \geqslant h_{CC}$, the right-hand side of Eq. (A.2) can be appropriately scaled up as follows:

$$\frac{\mathrm{d}\bar{h}}{\mathrm{d}P_C} < \left[P_C h_{CP} - P_C\left(h_{HC}+h_{CH}\right) + P_C h_{HH}\right]\left(\frac{1}{P_C}\right) + \frac{\left(1-(E)^L\right)h_{CP}}{1-(E)^L} < h_{HH} + h_{CP} - h_{HC} - h_{CH}.$$

According to Lemma 1 (I), we can know $\frac{\mathrm{d}\bar{h}}{\mathrm{d}P_C} < 0$. Therefore, the mixed lane capacity is a monotonically increasing function of the CAV penetration rate in the case of $E \neq 1$. The proof of Lemma 1 is complete. □

A.2. Proof of Lemma 2

Proof. Similarly, with the exogenously given maximum platoon size L and CAV clustering intensity E, the effect of CAV penetration rate P_C on the mixed lane capacity $c(\cdot)$ is discussed by analyzing the derivative of the average time headway $\bar{h}$ with respect to the CAV penetration rate P_C.

(1) In the case of $E = 1$, if $h_{CP} \leqslant h_{CC}$ and $h_{HH} < h_{CP}$ as stated in Lemma 2 (II), the first derivative of $\bar{h}$ with respect to P_C in Eq. (A.1) satisfies the following inequality relationship:

$$\frac{\mathrm{d}\bar{h}}{\mathrm{d}P_C} = \frac{h_{CP} + (L-1)h_{CC} - Lh_{HH}}{L} \geqslant h_{CP} - h_{HH} > 0.$$

Accordingly, the mixed lane capacity is a monotonically decreasing function of the CAV penetration rate.

(2) In the case of $E \neq 1$, the first derivative of $\bar{h}$ with respect to P_C is given in Eq. (A.2). Based on $E \in [\max\left(0, \frac{2P_C-1}{P_C}\right), 1)$ and $h_{CP} \leqslant h_{CC}$, the right-hand side of Eq. (A.2) can be appropriately scaled down as follows:

$$\frac{\mathrm{d}\bar{h}}{\mathrm{d}P_C} > \left[-P_C\left(h_{HC}+h_{CH}\right) + P_C h_{HH}\right]\left(\frac{E}{P_C}\right) + Eh_{CP} - E\left(h_{HC}+h_{CH}\right) + (E-2)h_{HH} > E\left[h_{CP} - 2h_{HC} - 2h_{CH} + 2h_{HH}\right] - 2h_{HH}.$$

Therefore, it has $\frac{\mathrm{d}\bar{h}}{\mathrm{d}P_C} > 0$ according to Lemma 2 (I), which indicates the mixed lane capacity is a monotonically decreasing function of the CAV penetration rate in the case of $E \neq 1$. The proof of Lemma 2 is complete. □

A.3. Proof of Lemma 3

Proof. When the CAV penetration rate P_C and the CAV clustering intensity E are exogenously given, the effect of the maximum platoon size L on the average time headway $\bar{h}$ can be analyzed as follows:

(1) In the case of $E = 1$, the first derivative of $\bar{h}$ with respect to L is:

$$\frac{\mathrm{d}\bar{h}}{\mathrm{d}L} = \frac{P_C\left(h_{CC} - h_{CP}\right)}{L^2}. \tag{A.5}$$

Therefore, if and only if $h_{CP} > h_{CC}$, the mixed lane capacity is a monotonically increasing function of the maximum platoon size in the case of $E = 1$.

(2) In the case of $E \neq 1$, the first derivative of $\bar{h}$ with respect to L is derived as follows:

$$\begin{aligned}\frac{\mathrm{d}\bar{h}}{\mathrm{d}L} &= h_{CP}P_C(1-E)\frac{\ln^E(E)^L}{\left[1-(E)^L\right]^2} - h_{CC}EP_C\frac{\ln^E(E)^{L-1}(1-E)}{\left[1-(E)^L\right]^2} \\ &= \ln^E(E)^{L-1}\left[\frac{EP_C\left(h_{CP}-h_{CC}\right)(1-E)}{\left[1-(E)^L\right]^2}\right].\end{aligned} \tag{A.6}$$

Since $E \in \left[\max\left(0, \frac{2P_C-1}{P_C}\right), 1\right)$, it has $\ln^E(E)^{L-1} < 0$, $\left[1-(E)^L\right]^2 > 0$, and $EP_C(1-E) > 0$. Thus, whether Eq. (A.6) is positive or negative depends on the relationship between $\left(h_{CP} - h_{CC}\right)$ and zero. Specifically, if and only if $h_{CP} - h_{CC} > 0$, $\frac{d\bar{h}}{dL} < 0$ can be derived. Thus, we can conclude that the mixed lane capacity is an increasing function of the maximum platoon size if and only if $h_{CP} > h_{CC}$ in the case of $E \neq 1$.

Similarly, we can also prove that the mixed lane capacity is a decreasing function of the maximum platoon size if and only if $h_{CP} < h_{CC}$. Finally, the proof of Lemma 3 is complete. □

A.4. Proof of Lemma 4

Proof. Due to the reciprocal relationship between the average time headway and the mixed lane capacity in Eq. (32), we can obtain the following equation on the second derivative of $c(\cdot)$ with respect to L.

$$\frac{d^2c}{dL^2} = 2\left(\bar{h}\right)^{-3}\frac{d\bar{h}}{dL} - \left(\bar{h}\right)^{-2}\frac{d^2\bar{h}}{dL^2}. \tag{A.7}$$

Then, with exogenously given the CAV penetration rate P_C and CAV clustering intensity E, the above lemma can be proved separately in the following two cases.

(1) In the case of $E = 1$, the second derivative of $\bar{h}$ with respect to L can be expressed as follows:

$$\frac{d^2\bar{h}}{dL^2} = \frac{2P_C\left(h_{CP} - h_{CC}\right)}{L^3}. \tag{A.8}$$

If $h_{CP} > h_{CC}$, it has $\frac{d^2\bar{h}}{dL^2} > 0$ and $\frac{d\bar{h}}{dL} < 0$ according to the proof of Lemma 3. Thus, $\frac{d^2c}{dL^2} > 0$ can be derived according to Eq. (A.7), which indicates the mixed lane capacity is a concave function of the maximum platoon size in the case of $E = 1$. In contrast, if $h_{CP} < Eh_{CC}$, we can have $h_{CP} < h_{CC}$, $\frac{d^2\bar{h}}{dL^2} < 0$, and $\frac{d\bar{h}}{dL} > 0$. Therefore, $\frac{d^2c}{dL^2} > 0$. It means that the mixed lane capacity is a convex function of the maximum platoon size in the case of $E = 1$.

(2) In the case of $E \neq 1$, we can derive the second derivative of $\bar{h}$ with respect to L as follows:

$$\begin{aligned}\frac{d^2\bar{h}}{dL^2} &= h_{CP}P_C(1-E)\frac{\left(\ln^E\right)^2\left[1-(E)^L\right](E)^L\left[1+(E)^L\right]}{\left[1-(E)^L\right]^4} + h_{CC}EP_C\frac{\left(\ln^E\right)^2(E-1)(E)^{L-1}\left[1-(E)^L\right]\left[1+(E)^L\right]}{\left[1-(E)^L\right]^4} \\ &= \frac{\left(\ln^E\right)^2\left[1-(E)^L\right](E)^L\left[1+(E)^L\right]\left[EP_C(1-E)\left(h_{CP}-Eh_{CC}\right)\right]}{\left[1-(E)^L\right]^4}.\end{aligned} \tag{A.9}$$

Since $E \in \left[\max\left(0, \frac{2P_C-1}{P_C}\right), 1\right)$, it can be inferred that $\left(\ln^E\right)^2\left[1-(E)^L\right](E)^L\left[1+(E)^L\right] > 0$, $\left[1-(E)^L\right]^4 > 0$. Thus, whether $\frac{d^2\bar{h}}{dL^2}$ in Eq. (A.9) is positive or negative depends on the relationship between $\left(h_{CP} - Eh_{CC}\right)$ and zero. If $h_{CP} > h_{CC}$, it is evident that $h_{CP} > Eh_{CC}$ and $\frac{d^2\bar{h}}{dL^2} < 0$, which indicates the mixed lane capacity is a concave function of the maximum platoon size if $h_{CP} > h_{CC}$. In contrast, if $h_{CP} < Eh_{CC}$, it has $\frac{d^2\bar{h}}{dL^2} < 0$. Finally, we can conclude that the mixed lane capacity is a convex function of the maximum platoon size if $h_{CP} < Eh_{CC}$. The proof of Lemma 4 is complete. □

A.5. Proof of Lemma 5

Proof. Because the mixed lane capacity is independent of the CAV clustering intensity in the case of $E = 1$, the above lemma is proved when E varies in the range of $\left[\max\left(0, \frac{2P_C-1}{P_C}\right), 1\right)$. When the CAV penetration rate (P_C) and the maximum platoon size (L) are exogenously given, the first derivative of $\bar{h}$ with respect to E is:

$$\frac{d\bar{h}}{dE} = \frac{\left[-(E)^L + L(E)^{L-1} - L(E)^L + (E)^{2L}\right]P_Ch_{CP} + \left[1-(E)^L - L(E)^{L-1} + L(E)^L\right]P_Ch_{CC}}{\left[1-(E)^L\right]^2} - P_C\left(h_{HC}+h_{CH}\right) + P_Ch_{HH}. \tag{A.10}$$

If $h_{CP} \geqslant h_{CC}$ as stated in Lemma 5 (II), the right-hand side of Eq. (A.10) can be appropriately scaled up as follows:

$$\begin{aligned}\frac{d\bar{h}}{dE} &= \frac{\left[-(E)^L + L(E)^{L-1} - L(E)^L + (E)^{2L}\right]P_Ch_{CP} + \left[1-(E)^L - L(E)^{L-1} + L(E)^L\right]P_Ch_{CC}}{\left[1-(E)^L\right]^2} - P_C\left(h_{HC}+h_{CH}\right) + P_Ch_{HH} \\ &= \frac{-(E)^L\left[1-(E)^L\right]P_Ch_{CP} + \left[1-(E)^L\right]P_Ch_{CC}}{\left[1-(E)^L\right]^2} + \frac{(E)^L\left(\frac{L}{E}-L\right)P_Ch_{CP} + (E)^L\left(L-\frac{L}{E}\right)P_Ch_{CC}}{\left[1-(E)^L\right]^2} - P_C\left(h_{HC}+h_{CH}\right) + P_Ch_{HH}\end{aligned}$$

$$
\begin{aligned}
&= \frac{-(E)^L P_C h_{CP} + P_C h_{CC}}{1-(E)^L} + \frac{(E)^L P_C\left[\left(\frac{L}{E}-L\right)h_{CP} + \left(L-\frac{L}{E}\right)h_{CC}\right]}{\left[1-(E)^L\right]^2} - P_C\left(h_{HC}+h_{CH}\right) + P_C h_{HH} \\
&= \frac{-(E)^L P_C h_{CP} + P_C h_{CC}}{1-(E)^L} + \frac{(E)^L P_C\left(h_{CP}-h_{CC}\right)\left(\frac{L}{E}-L\right)}{\left[1-(E)^L\right]^2} - P_C\left(h_{HC}+h_{CH}\right) + P_C h_{HH} \\
&< \frac{-(E)^L P_C h_{CP} + P_C h_{CC}}{1-(E)^L} + \frac{P_C\left(h_{CP}-h_{CC}\right)}{1-(E)^L} - P_C\left(h_{HC}+h_{CH}\right) + P_C h_{HH} \\
&< P_C\left(h_{HH}+h_{CP}-h_{HC}-h_{CH}\right).
\end{aligned}
$$

According to Lemma 5 (I), we can know $\frac{\mathrm{d}\bar{h}}{\mathrm{d}E} < 0$, which indicates the mixed lane capacity is a monotonically increasing function of the CAV clustering intensity. The proof of Lemma 5 is complete. □

A.6. Proof of Lemma 6

Proof. Similarly, the right-side of Eq. (A.10) can be appropriately scaled down based on $E \in [\max\left(0, \frac{2P_C-1}{P_C}\right), 1)$ and $h_{CP} \geqslant h_{CC}$ in Lemma 6 (II) as follows:

$$
\begin{aligned}
\frac{\mathrm{d}\bar{h}}{\mathrm{d}E} &= \frac{-(E)^L P_C h_{CP} + P_C h_{CC}}{1-(E)^L} + \frac{P_C(E)^L\left(h_{CP}-h_{CC}\right)\left(\frac{L}{E}-L\right)}{\left[1-(E)^L\right]^2} - P_C\left(h_{HC}+h_{CH}\right) + P_C h_{HH} \\
&> \frac{-(E)^L P_C h_{CP} + P_C h_{CC}}{1-(E)^L} + \frac{P_C(E)^L\left(h_{CP}-h_{CC}\right)}{1-(E)^L} - P_C\left(h_{HC}+h_{CH}\right) + P_C h_{HH} \\
&> P_C\left(h_{HH}+h_{CC}-h_{HC}-h_{CH}\right).
\end{aligned}
$$

According to Lemma 6 (I), we can know $\frac{\mathrm{d}\bar{h}}{\mathrm{d}E} > 0$. It means that the mixed lane capacity is a monotonically decreasing function of the CAV clustering intensity. The proof of Lemma 6 is complete. □

Appendix B. Theoretical analysis of capacity bounds

B.1. Proof of Lemma 7

Proof. We can find that the rank of the coefficient matrix of Eqs. (61)–(68) is seven. After finding rows in the coefficient matrix corresponding to Eqs. (61)–(62), (64)–(68) and columns corresponding to decision variables $p_{HH}, p_{HC}, p_{CC}, p_{CP}, y_L, y_L^C, y_L^C$, a basis matrix $\mathbf{B}_1$ can be obtained. Meanwhile, $p_{CH}, y_1, y_2, \ldots, y_{L-1}$ are regarded as non-basic variables. Let $\mathbf{C}_{\mathbf{B_1}}$ denote the vector of costs of the basic variables, the reduced costs of non-basic variables can be calculated as follows:

$$
R_j = c_j - \mathbf{C}_{\mathbf{B_1}}\left(\mathbf{B_1}\right)^{-1}\mathbf{p_j} = \begin{cases} h_{HC}+h_{CH}-h_{HH}-h_{CP}, & j=3, \\ \frac{L+5-j}{L}\left(h_{CP}-h_{CC}\right), & j \in \{6,7,\ldots,L+4\}. \end{cases}
$$

According to Lemma 7 (I–II), we can know the reduced costs of all non-basic variables are non-negative, which means the basic feasible solution of the linear programming model [*CBL*] with minimizing the average time headway is optimal. Conversely, according to Lemma 7 (III–IV), we can know the reduced costs of all non-basic variables are non-positive. Similarly, it means the basic feasible solution of the linear programming model [*CBL*] with maximizing the average time headway is optimal. After setting the non-basic variables to zero, one of the optimal solutions of the linear programming model [*CBL*] is $\mathbf{X} = \left(1-P_C, 0, 0, \frac{P_C}{L}, \frac{L-1}{L}P_C, 0, 0, \ldots, \frac{P_C}{L}, 0, \frac{P_C}{L}\right)^T$. Then, the mixed lane capacity is:

$$
c = \left[\left(1-P_C\right)h_{HH} + \frac{P_C}{L}h_{CP} + \frac{L-1}{L}P_C h_{CC}\right]^{-1}.
$$

Meanwhile, it can be found that $y_1 = 0, y_2 = 0, \ldots, y_{L-1} = 0, y_L = \frac{P_C}{L}$, $y_L^H = 0$, $y_L^C = \frac{P_C}{L}$, which indicates all CAVs are clustered together on the longitudinal lane. The proof is complete. □

B.2. Proof of Lemma 8

Proof. When all CAVs form maximum-size platoons, the spatial distribution of heterogeneous vehicles will only depend on the dispersion degree of platoons, which is directly affected by the CAV penetration rate. Therefore, we demonstrate the above lemma separately in the following two cases.

(1) $P_C \leqslant \frac{L}{L+1}$

In the first case, all maximum-size CAV platoons can be completely dispersed on the longitudinal lane. Thus, p_{HH}, p_{HC}, p_{CH}, p_{CC}, y_L, y_L^H, y_L^C are selected as basic variables, while the remaining decision variables are the non-basic variables. Let $\mathbf{B_2}$ represent the corresponding basis matrix, the reduced costs of non-basic variables are calculated as follows:

$$R_j = c_j - \mathbf{C_{B_2}}(\mathbf{B_2})^{-1}\mathbf{p_j} = \begin{cases} h_{HH} + h_{CP} - h_{HC} - h_{CH}, & j = 4, \\ \frac{j-L-5}{L}\left(h_{HH} + h_{CC} - h_{HC} - h_{CH}\right), & j \in \{6, 7, \dots, L+4\}. \end{cases}$$

According to Lemma 8 (I–II), we can know the reduced costs of all non-basic variables are non-negative. Conversely, Lemma 8 (III–IV) indicates the reduced costs of all non-basic variables are non-positive. These two conditions mean that the basic feasible solution of the linear programming model [*CBL*] with minimizing and maximizing the average time headway are optimal, respectively. Similarly, after setting the non-basic variables to zero, one of the optimal solutions of the linear programming model [*CBL*] is $\mathbf{X} = \left(1 - \frac{L+1}{L}P_C, \frac{P_C}{L}, \frac{P_C}{L}, 0, \frac{L-1}{L}P_C, 0, \dots, 0, \frac{P_C}{L}, \frac{P_C}{L}, 0\right)^T$. Meanwhile, the mixed lane capacity is equal to:

$$c = \left[\left(1 - \frac{L+1}{L}P_C\right)h_{HH} + \frac{P_C}{L}\left(h_{HC} + h_{CH}\right) + \left(\frac{L-1}{L}P_C\right)h_{CC}\right]^{-1}.$$

It can be found that $y_1 = 0, y_2 = 0, \dots, y_{L-1} = 0, y_L = \frac{P_C}{L}, y_L^H = \frac{P_C}{L}, y_L^C = 0$, which indicates all CAVs form the maximum-size platoons and the platoons are completely dispersed on the longitudinal lane.

(2) $P_C > \frac{L}{L+1}$

In this case, the maximum-size CAV platoons cannot be completely dispersed on the longitudinal lane anymore. Therefore, p_{HC}, p_{CH}, p_{CP}, p_{CC}, y_L, y_L^C, y_L^C can be selected as basic variables, and the associated basis matrix is denoted as $\mathbf{B_3}$. Then, the reduced costs of non-basic variables are calculated as follows:

$$R_j = c_j - \mathbf{C_{B_3}}(\mathbf{B_3})^{-1}\mathbf{p_j} = \begin{cases} h_{HH} + h_{CP} - h_{HC} - h_{CH}, & j = 1, \\ \frac{L+5-j}{L}\left(h_{CP} - h_{CC}\right), & j \in \{6, 7, \dots, L+4\}. \end{cases}$$

According to Lemma 8 (I–II) and (III–IV), we can know the reduced costs of all non-basic variables are non-negative and non-positive, respectively. One of the optimal solutions of the linear programming model [*CBL*] is $\mathbf{X} = \left(0, 1 - P_C, 1 - P_C, \frac{L+1}{L}P_C - 1, \frac{L-1}{L}P_C, 0, \dots, 0, \frac{P_C}{L}, 1 - P_C, \frac{L+1}{L}P_C - 1\right)^T$ and the corresponding mixed lane capacity is:

$$c = \left[\left(1 - P_C\right)h_{HC} + \left(1 - P_C\right)h_{CH} + \left(\frac{L+1}{L}P_C - 1\right)h_{CP} + \frac{L-1}{L}P_C h_{CC}\right]^{-1}.$$

Thus, $y_1 = 0, y_2 = 0, \dots, y_{L-1} = 0, y_L = \frac{P_C}{L}, y_L^H = 1 - P_C, y_L^C = \frac{L+1}{L}P_C - 1$, which indicates all CAVs form the maximum-size platoons and the platoons are dispersed on the longitudinal lane as much as possible. □

B.3. Proof of Lemma 9

Proof. If all CAVs avoid forming large-scale platoons on the longitudinal lane as much as possible, the number of CAV platoons with different sizes is directly determined by the CAV penetration rate. Therefore, the above lemma is discussed in detail according to the different value ranges of the CAV penetration rate.

(1) $P_C \leqslant \frac{1}{2}$

In the first case, all CAVs can be inserted one by one into the position between two adjacent HVs, and thus all CAVs drive alone on the longitudinal lane without forming a platoon with other CAVs. Therefore, $p_{HH}, p_{HC}, p_{CH}, p_{CP}, p_{CC}, y_1, y_L$ can be selected as basic variables. Let $\mathbf{B_4}$ denote the associated basis matrix, the reduced costs of non-basic variables are calculated as follows:

$$R_j = c_j - \mathbf{C_{B_4}}(\mathbf{B_4})^{-1}\mathbf{p_j} = \begin{cases} h_{HH} + h_{CC} - h_{HC} - h_{CH}, & j \in \{7, 8, \dots, L+4\}, \\ (L-1)\left(h_{HH} + h_{CC} - h_{HC} - h_{CH}\right), & j = L+6, \\ Lh_{HH} - L\left(h_{HC} + h_{CH}\right) + (L-1)h_{CC} + h_{CP}, & j = L+7. \end{cases}$$

According to $h_{CP} \geqslant h_{CC}$ as stated in Lemma 9 (II), it has:

$$Lh_{HH} - L\left(h_{HC} + h_{CH}\right) + (L-1)h_{CC} + h_{CP} \geqslant L\left(h_{HH} + h_{CC} - h_{HC} - h_{CH}\right).$$

Based on Lemma 9 (I), we can know the reduced costs of all non-basic variables are non-negative.

On the other hand, according to $h_{CP} \geqslant h_{CC}$ as stated in Lemma 9 (IV), it also has:

$$Lh_{HH} - L\left(h_{HC} + h_{CH}\right) + (L-1)h_{CC} + h_{CP} \leqslant L\left(h_{HH} + h_{CC} - h_{HC} - h_{CH}\right).$$

As stated in Lemma 9 (III), it means that the reduced costs of all non-basic variables are non-positive. Then, one of the optimal solutions of the linear programming model [*CBL*] is $\mathbf{X} = \left(1 - 2P_C, P_C, P_C, 0, 0, P_C, 0, \dots,\right)^T$, and the mixed lane capacity is $c = \left[\left(1 - 2P_C\right)h_{HH} + P_C h_{HC} + P_C h_{CH}\right]^{-1}$.

Moreover, it can be found that $p_{CC} = p_{CP} = 0$, $y_1 = P_C, y_2 = 0, \dots, y_L = 0$, which indicates all CAVs drive alone on the longitudinal lane without forming platoons with other CAVs.

(2) $\frac{1}{2} < P_C \leqslant \frac{L}{L+1}$

In this case, CAVs cannot be completely separated from each other anymore but may form CAV platoons with size r and $r+1$ on the longitudinal lane, $r \in \{1, \dots, L-2\}$. Therefore, $p_{HC}, p_{CH}, p_{CP}, p_{CC}, y_r, y_{r+1}, y_L$ can be selected as basic variables. Let $\mathbf{B_5}$ denote the associated basis matrix. Then, the reduced costs of non-basic variables are calculated as follows:

$$R_j = c_j - \mathbf{C_{B_5}}\left(\mathbf{B_5}\right)^{-1}\mathbf{p_j} = \begin{cases} h_{HH} + h_{CC} - h_{HC} - h_{CH}, & j = 1, \\ 0, & j \in \{6, \dots, r+4, r+7, \dots, L+4, L+6, L+7\}. \end{cases}$$

Lemma 9 (I) and (III) indicate that the reduced costs of all non-basic variables are non-negative and non-positive, respectively. Similarly, one of the optimal solutions of the linear programming model $[CBL]$ is $\mathbf{X} = \left(0, 1-P_C, 1-P_C, 0, 2P_C-1, 0, 0, \dots, r+1-(r+2)P_C, (r+1)P_C - r, 0, \dots,\right)^T$, and the corresponding mixed lane capacity is:

$$c = \left[\left(1-P_C\right)h_{HC} + \left(1-P_C\right)h_{HC} + \left(2P_C-1\right)h_{CC}\right]^{-1}.$$

Meanwhile, we can know that $y_1 = 0, \dots, y_{r-1} = 0$, $y_r = r+1-(r+2)P_C$, $y_{r+1} = (r+1)P_C - r$, $y_{r+2} = 0, \dots, y_L = 0$. It indicates that CAVs form platoons of size r and $r+1$ on the longitudinal lane, and the number of CAV platoons with size more than $r+1$ is zero.

(3) $P_C > \frac{L}{L+1}$

With the further increase of the CAV penetration rate, all CAVs form the maximum-size platoons, and the platoons cannot be completely dispersed on the longitudinal lane anymore. At that time, the selection of decision variables is consistent with the second case in Lemma 8, and the same mathematical expression of reduced costs of non-basic variables is obtained.

According to Lemma 9 (I–II), we can know the reduced costs of all non-basic variables are non-negative.

On the other hand, according to $h_{CP} \leqslant h_{CC}$ as stated in Lemma 9 (IV), it has $h_{HH} + h_{CP} - h_{HC} - h_{CH} \leqslant h_{HH} + h_{CC} - h_{HC} - h_{CH}$. According to Lemma 9 (III–IV), we can know that the reduced costs of all non-basic variables are non-positive. Then, one of the optimal solutions of the linear programming model $[CBL]$ is $\mathbf{X} = \left(0, 1-P_C, 1-P_C, \frac{L+1}{L}P_C - 1, \frac{L-1}{L}P_C, 0, \dots, 0, \frac{P_C}{L}, 1-P_C, \frac{L+1}{L}P_C - 1\right)^T$. The corresponding mathematical expression of mixed lane capacity is:

$$c = \left[\left(1-P_C\right)h_{HC} + \left(1-P_C\right)h_{CH} + \left(\frac{L+1}{L}P_C - 1\right)h_{CP} + \frac{L-1}{L}P_C h_{CC}\right]^{-1}.$$

Meanwhile, $y_1 = 0, \dots, y_{L-1} = 0, y_L = \frac{P_C}{L}, y_L^H = 1 - P_C, y_L^C = \frac{L+1}{L}P_C - 1$. It indicates that CAVs form the maximum-size platoons, and the platoons are dispersed on the longitudinal lane as much as possible. □

References

Bujanovic, P., Lochrane, T., 2018. Capacity predictions and capacity passenger car equivalents of platooning vehicles on basic segments. J. Transp. Eng. Part A: Syst. 144 (10), 04018063.

Chen, D., Ahn, S., Chitturi, M., Noyce, D.A., 2017. Towards vehicle automation: roadway capacity formulation for traffic mixed with regular and automated vehicles. Transp. Res. Part B: Methodol. 100, 196–221.

Chen, S., Wang, H., Xiao, L., Meng, Q., 2022. Random capacity for a single lane with mixed autonomous and human-driven vehicles: bounds, mean gaps and probability distributions. Transp. Res. Part E: Logist. Transp. Rev. 160, 102650.

Chen, Y., Qian, Z., Li, T., Dou, X., 2026. Deployment of dedicated CAV lanes and toll lanes in the mixed transportation network: a global optimization method. Appl. Math. Model. 153, 116629.

Daganzo, C.F., 2006. In traffic flow, cellular automata = kinematic waves. Transp. Res. Part B: Methodol. 40 (5), 396–403.

Du, W., Huang, Z., Li, S., Zhang, J., 2025. Fundamental diagram and stability analysis of mixed traffic flow considering platoon intensity and multiclass time delay. Physica A , 130847.

Fan, B., Xie, H., Li, T., 2025. Platoon communication power control under v2v data uncertainty: a robust DRL approach. IEEE Trans. Intell. Transp. Syst. 26(12), 21559-21573.

Ghiasi, A., Hussain, O., Qian, Z.S., Li, X., 2017. A mixed traffic capacity analysis and lane management model for connected automated vehicles: a markov chain method. Transp. Res. Part B: Methodol. 106, 266–292.

Guan, H., Wang, H., Meng, Q., Mak, C.L., 2023. Markov chain-based traffic analysis on platooning effect among mixed semi-and fully-autonomous vehicles in a freeway lane. Transp. Res. Part B: Methodol. 173, 176–202.

Jiang, Y., Zhu, F., Yao, Z., Gu, Q., Ran, B., 2023. Platoon intensity of connected automated vehicles: definition, formulas, examples, and applications. J. Adv. Transp. 2023 (1), 3325530.

Laval, J.A., Daganzo, C.F., 2006. Lane-changing in traffic streams. Transp. Res. Part B: Methodol. 40 (3), 251–264.

Levin, M.W., Boyles, S.D., 2016. A multiclass cell transmission model for shared human and autonomous vehicle roads. Transp. Res. Part C: Emerg. Technol. 62, 103–116.

Li, T., Cao, Y., Xu, M., Sun, H., 2023. Optimal intersection design and signal setting in the transportation network with mixed HVs and CAVs. Transp. Res. Part E: Logist. Transp. Rev. 175, 103173.

Li, T., Qian, Z., Fan, B., Xu, M., Sun, H., Chen, Y., 2024. Integrated optimal planning of multi-type lanes and intersections in a transportation network with mixed HVs and CAVs. Transp. Res. Part E: Logist. Transp. Rev. 192, 103814.

Liu, Z., Chen, Z., He, Y., Song, Z., 2021. Network user equilibrium problems with infrastructure-enabled autonomy. Transp. Res. Part B: Methodol. 154, 207–241.

Liu, Z., Song, Z., 2019. Strategic planning of dedicated autonomous vehicle lanes and autonomous vehicle/toll lanes in transportation networks. Transp. Res. Part C: Emerg. Technol. 106, 381–403.

Lu, J., Zhou, X.S., 2023. Virtual track networks: a hierarchical modeling framework and open-source tools for simplified and efficient connected and automated mobility (CAM) system design based on general modeling network specification (GMNS). Transp. Res. Part C: Emerg. Technol. 153, 104223.

Mohajerpoor, R., Ramezani, M., 2019. Mixed flow of autonomous and human-driven vehicles: analytical headway modeling and optimal lane management. Transp. Res. Part C: Emerg. Technol. 109, 194–210.

Qin, Y., Luo, Q., Wang, H., 2023. Stability analysis and connected vehicles management for mixed traffic flow with platoons of connected automated vehicles. Transp. Res. Part C: Emerg. Technol. 157, 104370.
Ren, X., Bai, L., Zheng, Y., Han, Y., Liu, P., 2024. Mixed traffic capacity estimation of autonomous vehicles impact based on empirical data. Appl. Math. Model. 135, 193–211.
Sala, M., Soriguera, F., 2021. Capacity of a freeway lane with platoons of autonomous vehicles mixed with regular traffic. Transp. Res. Part B: Methodol. 147, 116–131.
Wang, J., Lu, L., Peeta, S., He, Z., 2021. Optimal toll design problems under mixed traffic flow of human-driven vehicles and connected and autonomous vehicles. Transp. Res. Part C: Emerg. Technol. 125, 102952.
Wang, J., Peeta, S., He, X., 2019. Multiclass traffic assignment model for mixed traffic flow of human-driven vehicles and connected and autonomous vehicles. Transp. Res. Part B: Methodol. 126, 139–168.
Wang, J., Wang, W., Ren, G., Yang, M., 2022. Worst-case traffic assignment model for mixed traffic flow of human-driven vehicles and connected and autonomous vehicles by factoring in the uncertain link capacity. Transp. Res. Part C: Emerg. Technol. 140, 103703.
Wei, Y., Avcı, C., Liu, J., Belezamo, B., Aydın, N., Li, P.T., Zhou, X., 2017. Dynamic programming-based multi-vehicle longitudinal trajectory optimization with simplified car following models. Transp. Res. Part B: Methodol. 106, 102–129.
Yao, Z., Ma, Y., Ren, T., Jiang, Y., 2024. Impact of the heterogeneity and platoon size of connected vehicles on the capacity of mixed traffic flow. Appl. Math. Model. 125, 367–389.
Yao, Z., Wu, Y., Wang, Y., Zhao, B., Jiang, Y., 2023. Analysis of the impact of maximum platoon size of CAVs on mixed traffic flow: an analytical and simulation method. Transp. Res. Part C: Emerg. Technol. 147, 103989.
Yu, W., Hua, X., Ngoduy, D., Wang, W., 2023. On the assessment of the dynamic platoon and information flow topology on mixed traffic flow under connected environment. Transp. Res. Part C: Emerg. Technol. 154, 104265.
Zhang, F., Lu, J., Hu, X., Liu, X., Chen, J., 2022. Optimal design of differentiated credit charging links in mixed-autonomy transportation networks. Appl. Math. Model. 108, 646–669.
Zhang, F., Lu, J., Hu, X., Meng, Q., 2023. Integrated deployment of dedicated lane and roadside unit considering uncertain road capacity under the mixed-autonomy traffic environment. Transp. Res. Part B: Methodol. 174, 102784.
Zhao, P., Wong, Y.D., Zhu, F., 2025. Modeling and analysis of the platoon size of connected autonomous vehicles in a mixed traffic environment. Transp. Res. Part E: Logist. Transp. Rev. 199, 104130.
Zheng, Y., Xu, M., Wu, S., Wang, S., 2023. Lane management for asymmetric mixed traffic flow on bidirectional multi-lane roadways. Transp. A: Transp. Sci. , 21 (3), 2294492.
Zhou, J., Zhu, F., 2020. Modeling the fundamental diagram of mixed human-driven and connected automated vehicles. Transp. Res. Part C: Emerg. Technol. 115, 102614.
Zhou, J., Zhu, F., 2021. Analytical analysis of the effect of maximum platoon size of connected and automated vehicles. Transp. Res. Part C: Emerg. Technol. 122, 102882.
Zhou, W., Weng, J., Li, T., Fan, B., Bian, Y., 2024. Modeling the road network capacity in a mixed HV and CAV environment. Phys. A 636, 129526.