Behavior-rule dependence of capacity gains and hysteresis suppression in autonomous/human mixed cellular-automaton traffic

Mitsuyoshi Yagyu[a,*]

[a]*Department of Information and Data Science, Faculty of Information and Data Science, Notre Dame Seishin University, 2-16-9 Ifuku-cho, Kita-ku, Okayama City 700-8516 Japan*

Abstract

The dependence of capacity gains and hysteresis suppression on the choice of autonomous-vehicle (AV) behavior rule is investigated using a Nagel–Schreckenberg-type cellular-automaton model of mixed AV/human-vehicle (HV) traffic. Four AV following rules—naive (deterministic Nagel–Schreckenberg), IDM, ACC, and CACC—are implemented under a common collision-safety architecture, in which each vehicle computes a desired speed from its own rule and is subject to the same hard safety floor. Slow-to-start (VDR) dynamics are retained for HVs to induce hysteresis, and two independent detection methods are used to quantify the hysteresis loop width as a function of AV penetration rate r. The framework is validated against an exact two-species disordered-TASEP solution at $v_{\max} = 1$, showing agreement to within 0.53%.

The capacity gain from AV penetration is found to be highly sensitive to the calibration of the AV following rule rather than being an intrinsic property of the rule itself: under human-calibrated IDM parameters, AV penetration lowers capacity below the human-only baseline at every penetration rate tested, whereas doubling the acceleration/deceleration capability reverses this shortfall into a capacity gain. CACC, in contrast, saturates the common safety floor and becomes indistinguishable from the naive rule at high r, while ACC remains distinct owing to its longer effective headway. Furthermore, hysteresis suppression is shown to be threshold-like rather than

*Corresponding author.
Email address: `mitsu.yagyu@gmail.com` (Mitsuyoshi Yagyu)

proportional to r: the loop width remains largely unchanged up to $r \approx 0.4$ and vanishes only as $r \to 1$, indicating that AVs must dominate the population of stopped vehicles before slow-to-start asymmetry is eliminated. These results demonstrate that the widely assumed benefit of AV penetration for traffic flow is not a model-independent conclusion, but instead depends sensitively on the specific behavior rule and its calibration.

Keywords: cellular automaton, Nagel–Schreckenberg model, autonomous vehicles, mixed traffic, car-following models, hysteresis

---

1. Introduction

The Nagel–Schreckenberg (NaSch) cellular automaton [11] and its many variants [3] remain a standard, computationally efficient tool for studying macroscopic traffic flow phenomena—the triangular flow-density fundamental diagram, jam formation, and, in slow-to-start variants, metastability and hysteresis. With the emergence of autonomous and connected vehicles, a large body of work has asked how the AV penetration rate r reshapes the fundamental diagram, typically by giving AVs a deterministic (non-stochastic) update rule, sometimes combined with a reduced following gap. This question—AV penetration rate vs. fundamental diagram—is by now well covered by prior studies using a variety of specific AV control laws, including cyber-interaction-based communication rules [5], a fixed CACC control law combined with a calibrated human car-following model [8], and reinforcement-learning-trained AV policies [12, 7]. A simple re-derivation of the same qualitative conclusion—that AVs improve capacity—under yet another AV rule would add little.

We instead ask a comparative question that, to our knowledge, has not been addressed within a single, controlled CA framework: how much does the *answer itself* depend on which AV behavior model is assumed? We implement four car-following rules on the same NaSch lattice—a simple deterministic rule (no randomness, otherwise identical to a human-driven vehicle, HV), the Intelligent Driver Model [IDM; 6], and the Adaptive/Cooperative Adaptive Cruise Control laws calibrated directly from experimental vehicle data by Milanés and Shladover [10] (a calibration also connected to macroscopic flow characteristics by van Arem et al. [1])—under one common, collision-safe update architecture, and compare their effect on the fundamental diagram across AV penetration rate r.

A second, related but distinct question is whether AV penetration affects metastability and hysteresis. Plain (non-slow-to-start) NaSch traffic does not exhibit hysteresis; the effect requires a slow-to-start, velocity-dependent-randomization (VDR) rule in which standing vehicles restart with a different, typically lower, probability than moving vehicles brake [2, 9]. Whether and how AV penetration changes this metastability—a mechanism entirely orthogonal to the fundamental-diagram question above, since it concerns the *shape* of the flow-density relation and its dependence on initial conditions, not merely its height—has, to our knowledge, not been systematically reported.

Implementing multiple literature AV models under mixed traffic surfaced a methodological pitfall worth reporting on its own terms (Section 2.3): a common way of representing an AV's advantage—letting it tolerate a shorter following gap than an HV—is only collision-free under NaSch's fully parallel update rule if the vehicle ahead is guaranteed to move every timestep in which it might be adjacent to a follower. This holds for the standard single-cell safety margin but not for a reduced one, and the violation is invisible in a homogeneous AV population, where it happens to be safe, while occurring in a large fraction of timesteps once mixed with a stochastically braking HV population. We flag this because the same reduced-gap construction recurs across the mixed-autonomy CA literature and, to our knowledge, its collision safety in heterogeneous populations is not always explicitly checked.

The remainder of this paper is organized as follows. Section 2 describes the base model, the four AV car-following rules and their shared safety architecture, and the VDR slow-to-start rule together with two independent methods for detecting hysteresis. Section 3 presents the fundamental-diagram comparison across AV rules (Section 3.1) and the hysteresis-suppression results (Section 3.2). Section 4 independently validates the shared safety-floor architecture at $v_{\max} = 1$ against an exact disordered-exclusion-process solution. Section 5 discusses the implications and limitations, and concludes.

2. Model and methods

2.1. Base model

We simulate a single-lane, periodic (ring) NaSch lattice of L sites, each holding at most one vehicle. Vehicle i's state is its position x_i and integer velocity $v_i \in \{0, \ldots, v_{\max}\}$. Writing d_i for the gap (number of sites) to the vehicle ahead, the standard update rule advances the system by

$$v_i \leftarrow \min(v_i + 1,\ v_{\max}) \qquad \text{(acceleration)} \qquad (1)$$
$$v_i \leftarrow \min(v_i,\ d_i - g_{\min}) \qquad \text{(safety/gap constraint)} \qquad (2)$$
$$v_i \leftarrow \max(v_i - 1,\ 0) \text{ w.p. } p \qquad \text{(random slow-down)} \qquad (3)$$
$$x_i \leftarrow x_i + v_i \qquad \text{(movement)} \qquad (4)$$

applied in parallel to every vehicle, where $g_{\min}$ is the minimum allowed gap. Standard NaSch [11] is recovered with $g_{\min} = 1$ and a uniform p. A human-driven vehicle (HV) uses $v_{\max} = 5$, $p_{\mathrm{HV}} = 0.5$, $g_{\min} = 1$ throughout this paper.

2.2. Baseline validation

Before extending this engine to mixed AV/HV traffic, we validate our own implementation against the classic homogeneous-HV result it is built on. Figure 1 shows the simulated fundamental diagram at $v_{\max} = 5$, $p = 0.5$, $L = 1000$, giving $J_{\max} = 0.328$ at $\rho_{\mathrm{c}} = 0.075$, matching the qualitative shape—a steep free-flow branch, a sharp kink, and a slowly decaying congested branch—of the classic triangular fundamental diagram of Nagel and Schreckenberg [11].

Quantitative agreement was further verified directly against that paper for the identical parameters ($v_{\max} = 5$, $p = 0.5$): Sec. 5 of Ref. [11] states explicitly that "our maximum of the flow is only 0.32 vehicles per timestep" ($J_{\max} = 0.32$ vs. our 0.328, a ∼2.5% relative difference), and its Fig. 6—a zoomed, unambiguously labeled version of the same simulation data—shows the peak directly at $\rho_{\mathrm{c}} \approx 0.08$–$0.09$, $J_{\max} \approx 0.32$–0.33. The small remaining difference in ρ_{c} is consistent with our smaller system size and shorter relaxation/averaging than that paper uses for its Figs. 4/6 ($L = 1000$, $T_{\mathrm{relax}} = 1000$ here, versus $L = 10^4$, relaxation over $10 \times L$ steps, and averaging up to 10^5–10^6 timesteps there): a bulk quantity such as $J_{\max}$ self-averages over many vehicles and is far less sensitive to this difference than the precise location of a sharp kink on a finite density grid. This agreement confirms the simulation engine itself, prior to the introduction of any AV-specific extension.

2.3. A collision-safety pitfall in reduced-gap AV models

A common way to represent an AV's advantage over an HV is to let it tolerate a shorter headway, i.e., set $g_{\min,\mathrm{AV}} < g_{\min,\mathrm{HV}}$ in Eq. (2) (in the

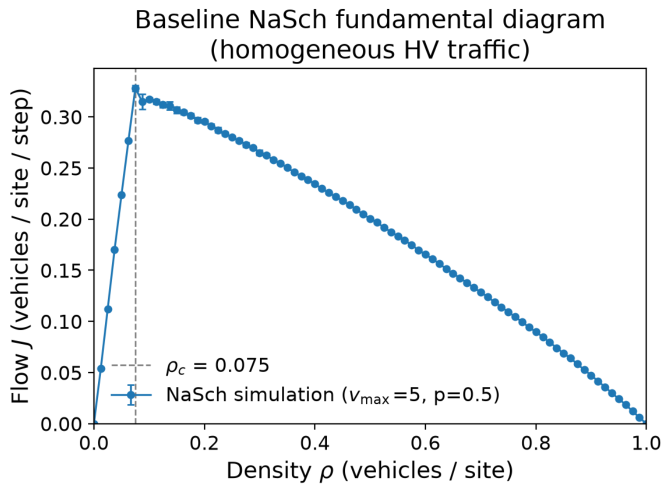

Figure 1: Baseline validation: simulated fundamental diagram for homogeneous HV traffic ($v_{\max} = 5$, $p = 0.5$), reproducing the classic NaSch triangular flow–density relation [11, 3].

extreme, $g_{\min,\mathrm{AV}} = 0$). Under the fully parallel update described above, this construction is collision-free only if the vehicle immediately ahead is guaranteed to move at least $g_{\min,\mathrm{HV}} - g_{\min,\mathrm{AV}}$ cells in any timestep in which it might be adjacent to a follower using the reduced margin, because Eq. (2) is evaluated against the leader's *pre-update* position, not its realized post-update position. With $g_{\min,\mathrm{AV}} = 0$, a follower may advance up to d_i cells, landing exactly on the leader's pre-update site if the leader itself realizes $v = 0$ in that same timestep—for instance, through an HV's random slowdown or the leader's own gap constraint.

We confirmed that this failure mode occurs in practice: instrumenting a homogeneous-AV population (all vehicles using $g_{\min} = 0$) shows zero collisions at any tested density, but a mixed population with as little as 20% HV shows overlapping (duplicate-position) vehicle states in the large majority of measured timesteps at moderate density (e.g., $\rho = 0.1$, $r = 0.2$: 2515 of 3000 measured steps corrupted). The pitfall is thus invisible precisely in the homogeneous configuration in which it would most naturally be first tested, and manifests only once a heterogeneous population is introduced—exactly the regime this paper's questions require. We therefore fix $g_{\min} = 1$ for *every* rule studied here, AV and HV alike; the AV/HV distinction is expressed entirely through the car-following dynamics described in Section 2.4, never through a shorter geometric margin. We are not aware of this specific mixed-population failure mode having been flagged in the prior mixed-autonomy cellular-automaton literature, and report it here as a general caution for related model designs.

2.4. Four AV car-following rules

All four rules share the same architecture: each vehicle computes a *desired* velocity from its own car-following law, which is then subject to one shared hard safety floor,

$$v_i \leftarrow \min(v_{i,\text{desired}},\ d_i - 1)\,, \tag{5}$$

identical in form to Eq. (2) with $g_{\min} = 1$, followed by random slow-down (Eq. 3, with $p = 0$ for every AV rule) and movement (Eq. 4). Only the first step—how $v_{i,\text{desired}}$ is computed—differs between rules.

2.4.1. Naive (baseline)

$v_{i,\text{desired}} = \min(v_i + 1, v_{\max})$, i.e. Eq. (1). This isolates the effect of removing random slow-down alone and serves as the point of comparison for the other three rules.

2.4.2. IDM

Following Helbing [6], Euler-discretized with $\Delta t = 1$ (one NaSch timestep):

$$v_{i,\text{desired}} = v_i + a\left[1 - \left(\frac{v_i}{v_0}\right)^{\delta} - \left(\frac{s^*(v_i, \Delta v_i)}{d_i}\right)^2\right], \tag{6}$$

$$s^*(v_i, \Delta v_i) = s_0 + \max\!\left(0,\ Tv_i + \frac{v_i \Delta v_i}{2\sqrt{ab}}\right), \tag{7}$$

where $\Delta v_i = v_i - v_{i+1}$ is the closing rate with the leader. We set $v_0 = v_{\max}$ (so all four rules share the same free-flow speed ceiling, isolating behavioral rather than desired-speed differences) and $s_0 = 0$ (the shared safety floor of Eq. 5 already reserves one cell, so an additional internal jam distance would double-count that margin). T, a, b are calibrated highway values verified directly from Treiber et al. [14] ($a = 0.73\,\text{m/s}^2$, $b = 1.67\,\text{m/s}^2$, $T = 1.6\,\text{s}$, converted to cell/step units using the standard NaSch convention of one cell $= 7.5\,\text{m}$ and one timestep $= 1\,\text{s}$), $\delta = 4$.

2.4.3. ACC and CACC

Following the experimentally calibrated control laws of Milanés and Shladover [10] (verified directly against the published equations and coefficients), with gap error $e_i = d_i - hv_i$:

$$\text{ACC:} \quad v_{i,\text{desired}} = v_i + k_1 e_i + k_2(v_{i+1} - v_i), \quad k_1 = 0.23,\ k_2 = 0.07,\ h = 1.1\,\text{s}, \tag{8}$$

$$\text{CACC:} \quad v_{i,\text{desired}} = v_i + k_p e_i + k_d(e_i - e_i^{\text{prev}}), \quad k_p = 0.45,\ k_d = 0.25,\ h = 0.6\,\text{s}. \tag{9}$$

The CACC control law requires the previous timestep's gap error e_i^{prev}, carried as per-vehicle persistent state. Because k_1, k_2, k_p, k_d are purely inverse-time quantities and h is a time (and one NaSch timestep is by convention one second), these coefficients transfer to cell/step units unchanged. We note for completeness that our initial implementation assumed a single shared PD-control functional form for both ACC and CACC with a leader-acceleration feed-forward term for CACC; on verifying directly against Milanés and Shladover [10], the two control laws are structurally different as given in Eqs. (8)– (9) (CACC is a velocity recursion without an explicit feed-forward term in the simplified traffic-flow model used here), while the numerical gain and time-headway values we had recalled were, once re-derived from the source, correct.

2.4.4. A further discretization detail matters

Because IDM/ACC/CACC accelerations are frequently well under 1 cell/step^2 (e.g. IDM's $a \approx 0.13$ cells/step^2), rounding each vehicle's velocity to an integer at every timestep silently discards this sub-unit progress every step; we found this produces permanent gridlock for a pure-IDM population under an integer-state implementation. We instead carry velocity as a persistent floating-point state across timesteps (exactly integer-valued for the naive/HV rule, genuinely fractional for IDM, ACC, and CACC) and round only the per-timestep *displacement* to an integer number of lattice cells; flux is measured from this integer displacement, not the continuous state.

2.5. Slow-to-start rule and hysteresis detection

To study metastability, we add the velocity-dependent-randomization (VDR) slow-to-start rule of Barlović et al. [2]: the random slow-down probability in Eq. (3) depends on whether a vehicle's velocity at the start of the timestep was zero,

$$p(v) = \begin{cases} p_0 & v = 0 \\ p & v > 0 \end{cases}, \qquad p_0 \geq p, \tag{10}$$

with HVs using Barlović et al.'s parameters ($p = 1/64$, $p_0 = 0.75$) and AVs deterministic regardless of velocity ($p = p_0 = 0$); both use the shared $g_{\min} = 1$ floor.

We use two independent methods to detect hysteresis, both used by Barlović et al. [2] to trace the same loop. *Method 1* (adiabatic ramp): starting from a low-density homogeneous configuration, density is increased by inserting vehicles into the continuing trajectory (no reset), tracing the upper (free-flow) branch; starting from a high-density jammed configuration, density is decreased by removing vehicles, tracing the lower (jammed) branch. *Method 2* (two initial conditions): flow is measured independently at each density from a fresh homogeneous initial condition (upper branch) and a fresh single-jam initial condition (lower branch). Method 1 is run only for the human-only ($r = 0$) baseline, since it is inherently sequential and expensive to repeat across many r; Method 2, cheaper and embarrassingly parallel, is used for the full AV-fraction scan.

3. Simulation results

3.1. AV behavior rule comparison

Figure 2 shows the fundamental diagram at $r = 1$ (a homogeneous population under each rule); Figure 3 and Table 1 show maximum flow $J_{\max}(r)$ across AV fraction $r \in \{0, 0.25, 0.5, 0.75, 1.0\}$, mixed with a common HV background ($p_{\mathrm{HV}} = 0.5$). Here and throughout, $J_{\max}(r)$ is the maximum, over a density grid refined to $\Delta\rho = 0.01$ near the low-density peak, of the flow averaged over 16 random initial conditions; the fine grid and seed averaging are needed because the IDM fundamental diagram peaks sharply at low density ($\rho \approx 0.06$), where a coarse grid or few seeds estimates the peak flow unreliably. As required by construction, the four rules agree at $r = 0$ (no AVs present, regardless of which AV rule is configured) within seed-to-seed noise, providing a built-in consistency check.

Two findings depart from the simple, monotonic "more AVs, more capacity" picture.

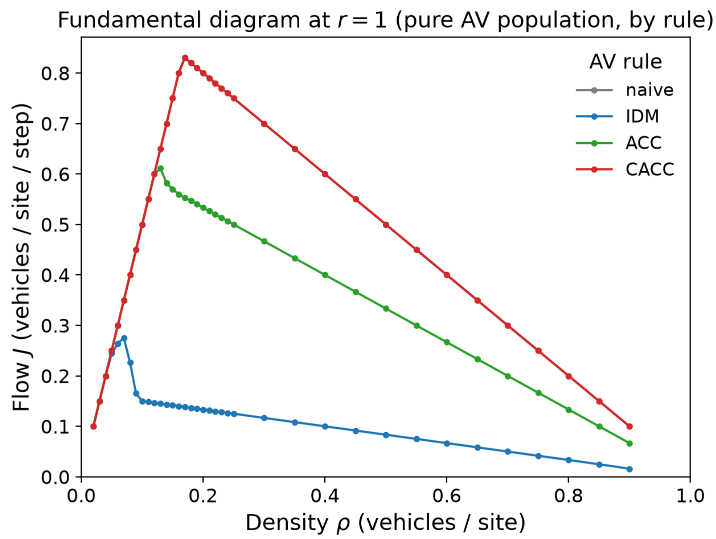

Figure 2: Fundamental diagram at $r = 1$ (homogeneous AV population) for each of the four car-following rules.

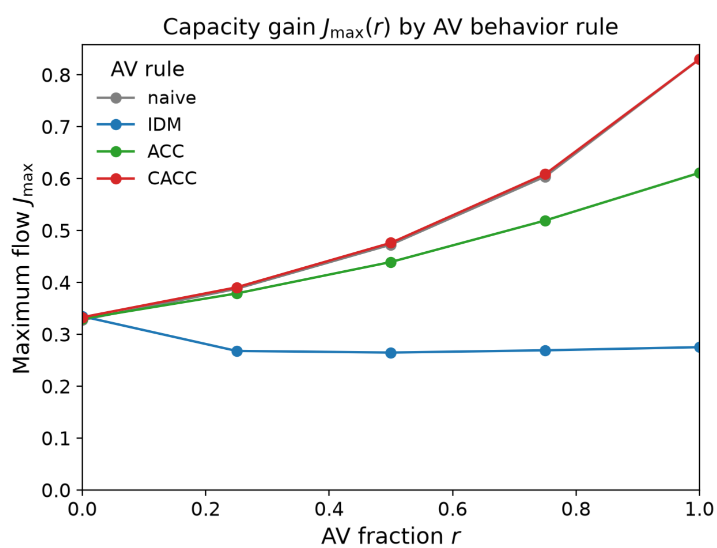

Figure 3: Maximum flow $J_{\max}$ as a function of AV fraction r, by AV car-following rule.

Table 1: Maximum flow $J_{\max}(r)$ by AV rule.

r	naive	IDM	ACC	CACC
0.00	0.328	0.334	0.330	0.333
0.25	0.388	0.268	0.379	0.390
0.50	0.473	0.265	0.439	0.476
0.75	0.604	0.269	0.519	0.609
1.00	0.830	0.275	0.611	0.830

3.1.1. Finding 1: IDM's capacity effect is highly sensitive to acceleration and deceleration calibration, and can be negative

With the IDM parameters of Section 2.4 ($v_0 = v_{\max}$, $s_0 = 0$, $a = 0.73\,\mathrm{m/s^2}$, $b = 1.67\,\mathrm{m/s^2}$, $T = 1.6\,\mathrm{s}$ – values verified directly from Treiber et al. [14], as is common practice when an IDM-based rule is assigned to an "AV" role), IDM gives lower $J_{\max}$ than the all-HV baseline at every $r > 0$ tested: rather than rising with r as the naive, ACC, and CACC rules do, $J_{\max}$ stays essentially flat near 0.27 across all penetration rates (Table 1), roughly 0.06 below the $r = 0$ baseline of ≈ 0.33 (e.g. $r = 0.5$: $J_{\max} = 0.265$). Because $v_0 = v_{\max}$ and $s_0 = 0$, IDM's own acceleration (Eq. 6) is never exactly zero except in the $d_i \to \infty$ limit—even a vehicle at $v_{\max}$ experiences residual deceleration from the interaction term at any finite gap—and this drag is amplified by the discrete lattice's coarse velocity quantization acting on small closing-speed fluctuations, since a, b are numerically small in cell units.

To test whether this is a fixed property of IDM or an artifact of reusing a human-calibrated a, b for an AV role, we re-ran the same r-scan under three additional parameter sets: doubled acceleration/deceleration capability ("responsive": $a = 1.46$, $b = 3.34\,\mathrm{m/s^2}$), a shorter desired time headway alone ("short-headway": $T = 0.8\,\mathrm{s}$, a, b unchanged), and both combined (Table 2, Figure 4). **Under the responsive calibration the effect reverses: IDM then exceeds the all-HV baseline across the penetration range, at both $r = 0.5$ ($J_{\max} = 0.343$ vs. the ≈ 0.33 baseline) and $r = 1$ (0.396).** Shortening T alone, with a, b held at the weaker human-calibrated values, instead makes things *worse*—especially at $r = 1$ (0.255, below even the unmodified baseline)—because demanding a shorter gap without the actuation capability to maintain it safely backfires; T and (a, b) interact rather than acting independently, as the combined set confirms: pairing the shorter headway with the stronger actuation recovers the gain and again exceeds the baseline at $r = 1$ (0.450). The correct statement of Finding 1 is therefore not that IDM reduces capacity as a fixed property, but that *reusing a human-driver-calibrated a, b for an AV role, without revisiting it for the AV's presumably faster actuation, makes an IDM-governed AV underperform human-only traffic, an effect that reverses once acceleration/deceleration capability is recalibrated, but only if the desired headway is not shortened faster than that capability can support.*

Table 2: IDM parameter sensitivity: $J_{\max}(r)$ under four calibrations.

Calibration	a (m/s^2)	b (m/s^2)	T (s)	$J_{\max}(0)$	$J_{\max}(0.5)$	$J_{\max}(1)$
baseline (verified, Treiber et al. 14)	0.73	1.67	1.6	0.328	0.268	0.273
responsive (2x a, b)	1.46	3.34	1.6	0.335	0.343	0.396
short-headway (lower T only)	0.73	1.67	0.8	0.330	0.264	0.255
combined (2x a, b + lower T)	1.46	3.34	0.8	0.332	0.337	0.450

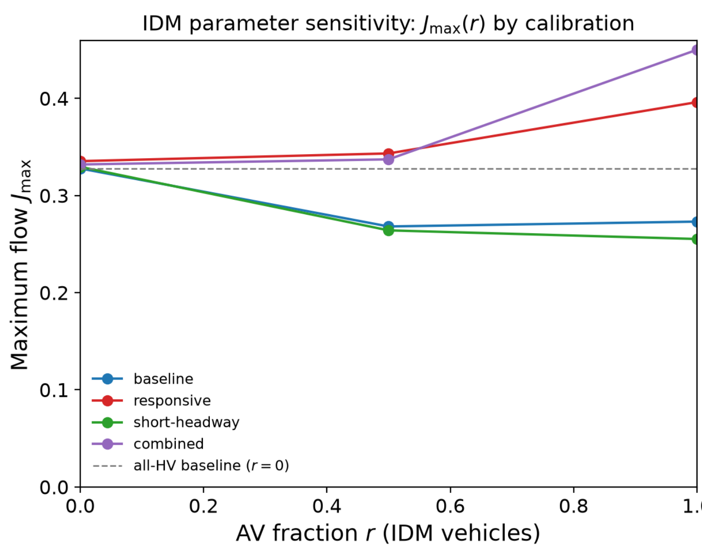

Figure 4: IDM parameter sensitivity: $J_{\max}(r)$ under four IDM acceleration/deceleration and time-headway calibrations.

3.1.2. Finding 2: CACC saturates against the shared safety floor

At $r = 1$, CACC's $J_{\max}$ (0.830) exactly matches the naive rule's, while ACC's (0.611) is meaningfully lower. At CACC's own control-law fixed point

($e_i = e_i^{\text{prev}}$, so the derivative term vanishes), the implied equilibrium velocity is $v_{\text{eq}} = d_i/h$. With $h = 0.6\,\text{s}$ and the moderate gaps typical of the densities studied, v_{eq} substantially exceeds $v_{\max}$, so CACC's own control law is floor-bound (Eq. 5) at essentially every density tested; since the naive rule's acceleration ramp converges to the same $\min(v_{\max}, d_i - 1)$ fixed point given enough relaxation time, the two rules become numerically indistinguishable in steady state. ACC's smaller gains ($k_1 = 0.23$ vs. CACC's $k_p = 0.45$) and longer time headway ($h = 1.1\,\text{s}$) leave it non-floor-bound more often within the measurement window, producing a visibly different, lower curve. This indicates that control-law gains and time headways calibrated for continuous, meter-scale following distances in flowing real-world traffic can lose their distinguishing behavior once mapped onto a coarser cellular-automaton lattice at low-to-moderate density—a discretization-interaction effect, not a defect in the underlying verified control law.

3.2. VDR hysteresis suppression by AV fraction

Figure 5 validates the VDR implementation at $r = 0$ against Barlović et al.'s (1998) own result (two-initial-condition method): a clear hysteresis loop is reproduced, with the homogeneous-IC branch far above the jammed-IC branch in an intermediate density window, converging outside it. The supplementary adiabatic-ramp check (Figure 6) shows a weaker, noisier version of the same signal: inserting vehicles at random empty sites into an already-relaxed homogeneous trajectory introduces local perturbations that accumulate and nucleate jams earlier than a fresh restart would, so the ramp's increasing branch does not sustain the same high-flow values as the two-IC method at comparable densities, though the decreasing branch does remain measurably above the increasing branch through most of the mid-density range. We rely on the two-IC method, Barlović et al.'s own primary approach, for the main AV-fraction result below.

Figure 7 shows the hysteresis loop at several AV fractions, and Figure 8/Table 3 summarize the loop width, defined as $\max_\rho (J_{\text{hom}}(\rho) - J_{\text{jam}}(\rho))$.

3.2.1. Finding 3: hysteresis suppression is threshold-like, not proportional

Loop width is roughly flat, within run-to-run noise, up to $r \approx 0.4$, then falls markedly above $r \approx 0.6$, vanishing exactly (not merely approximately) at $r = 1.0$: an all-AV population has $p = p_0 = 0$ uniformly, so the VDR rule reduces identically to the deterministic NaSch model, which cannot produce metastability by construction (there is no slow-to-start asymmetry left to

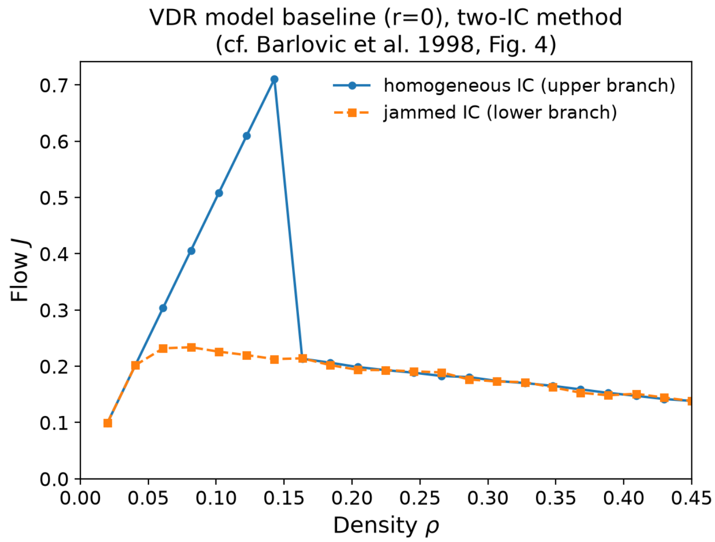

Figure 5: VDR model baseline ($r = 0$), two-initial-condition method, reproducing Barlović et al. [2]'s Fig. 4.

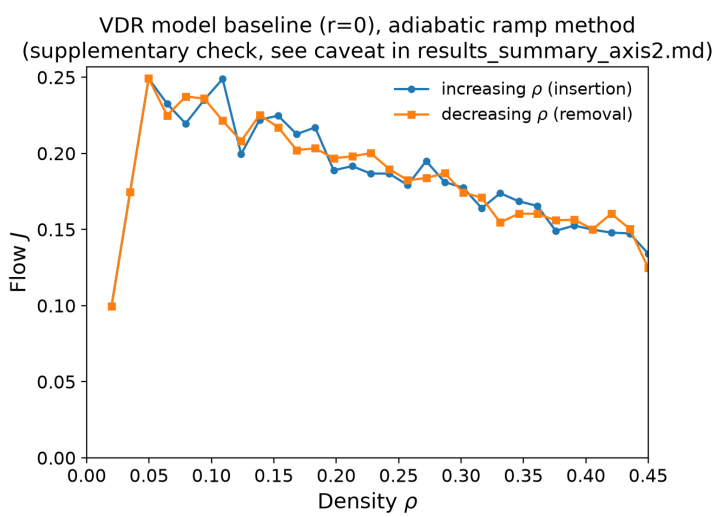

Figure 6: VDR model baseline ($r = 0$), adiabatic ramp method (supplementary check; see discussion of its weaker signal in the text).

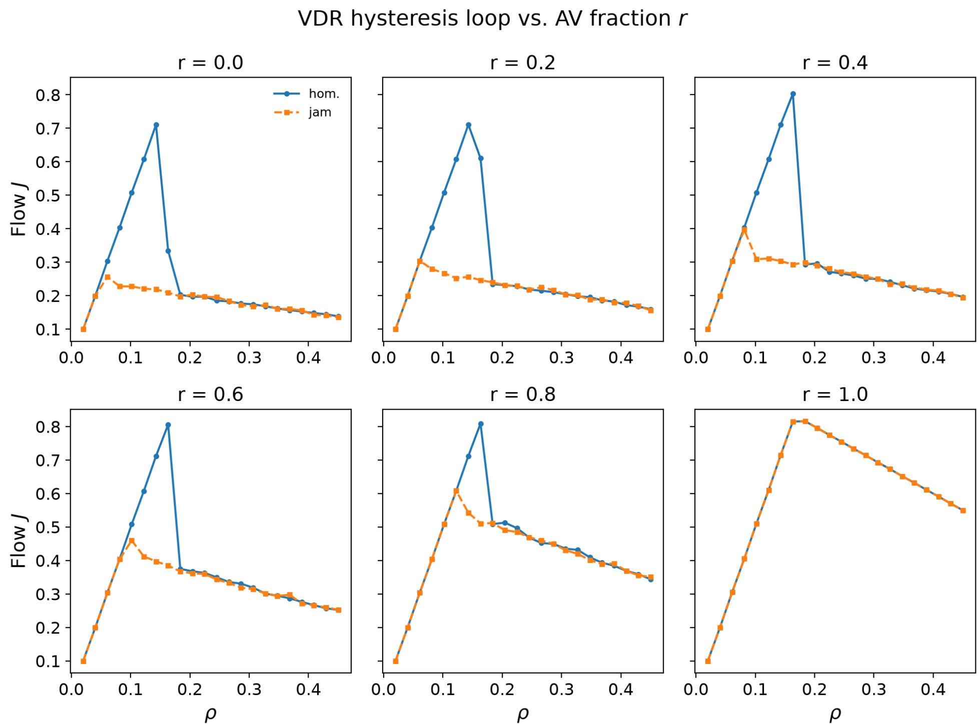

Figure 7: VDR hysteresis loop at several AV fractions r.

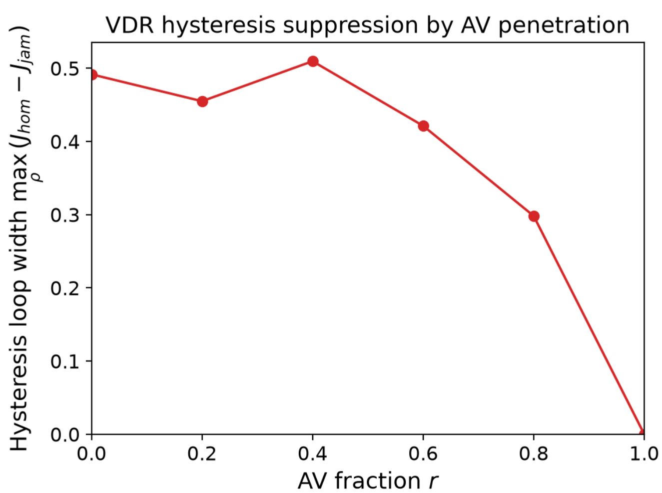

Figure 8: Hysteresis loop width vs. AV fraction r.

Table 3: Hysteresis loop width by AV fraction.

r	loop width
0.0	0.492
0.2	0.455
0.4	0.510
0.6	0.421
0.8	0.298
1.0	0.000

produce it) – a built-in consistency check on the scan rather than a separate finding. The qualitative shape differs from the fundamental-diagram results of Section 3.1, where capacity effects (for the naive/ACC/CACC rules) grow smoothly and continuously with r: suppressing hysteresis appears to require AVs to make up a majority of the standing-vehicle population before the slow-to-start asymmetry's aggregate effect on the flow-density curve is meaningfully diluted, since how often a vehicle is standing is itself density- and jam-structure-dependent rather than simply proportional to the AV fraction.

4. Analytical validation at $v_{\max} = 1$

The simulation results above (Section 3) all use $v_{\max} = 5$, for which no exact analytical solution of the mixed AV/HV model is available. As an independent check on the modeling framework itself—specifically, the naive rule and shared safety-floor architecture of Section 2.4—we additionally verify it against an exact result at $v_{\max} = 1$, where the model reduces to a special case of a well-studied class of exactly solvable exclusion processes.

4.1. Connection to disordered exclusion processes

At $v_{\max} = 1$, our mixed model (vehicle i moves one site forward if the site ahead is empty, with probability $q_\tau = 1 - p_\tau$ depending on type $\tau \in \{\mathrm{AV}, \mathrm{HV}\}$, $p_{\mathrm{AV}} = 0$, $p_{\mathrm{HV}} > 0$) is an instance of the totally asymmetric exclusion process (TASEP) on a ring with *particle-wise disordered hopping probabilities* under fully parallel (synchronous) updating. This disordered-TASEP class was solved exactly by Evans [4] for an arbitrary distribution of per-particle hopping probabilities q_μ (in Evans' notation, particle μ hops with probability p_μ; our q_τ plays the same role). Our two-type case is thus the special case of Evans's (1997) result restricted to a two-point distribution ($q = 1$

for AV, $q = q_{\text{HV}} < 1$ for HV, in fractions r and $1 - r$). The exact stationary gap-distribution weights [4, Eq. 4], the current/velocity self-consistency equation [4, Eq. 17], and the fact underlying both—that no two adjacent particles on a no-passing ring can sustain different long-run average velocities, since a persistent gap between differing average speeds would grow or shrink without bound on a bounded ring—are all already established there for the fully general disordered case. We are not aware, however, of this exact solution having been previously connected to the car-oriented mean-field theory (COMF) framework used in the traffic-CA literature [13], nor explicitly specialized to the two-type AV-fraction parametrization relevant to mixed autonomous/human traffic; we give both below, mainly for the reader's convenience and as a self-contained basis for the validation that follows, and we validate the resulting flux-density relation against our own simulation.

4.2. Two-type closed form and density constraint

Since $q_{\text{AV}} = 1$, adjacent vehicles' shared move probability (the current J/ρ, equivalently Evans's velocity v, here written g^* to match COMF notation) lets the AV gap distribution be read off directly from Evans's Eq. 4: $P_0^{\text{AV}} = 1 - g^*$, $P_1^{\text{AV}} = g^*$, $P_n^{\text{AV}} = 0$ for $n \geq 2$ (a deterministic follower's gap is supported only on $\{0, 1\}$), giving mean occupied width

$$\mathbb{E}[n + 1 \mid \text{AV}] = 1 + g^*. \tag{11}$$

For HV, the same equation gives the geometric tail $P_n^{\text{HV}} = (P_0^{\text{HV}}/p_{\text{HV}})\, \rho_{\text{HV}}^n$ ($n \geq 1$), with $\rho_{\text{HV}} = p_{\text{HV}} g^*/(q_{\text{HV}}(1 - g^*))$ and $P_0^{\text{HV}} = (1 - g^*) - (p_{\text{HV}}/q_{\text{HV}})\, g^*$, giving

$$\mathbb{E}[n + 1 \mid \text{HV}] = P_0^{\text{HV}} \left[1 + \frac{1}{p_{\text{HV}}} \left(\frac{1}{(1 - \rho_{\text{HV}})^2} - 1\right)\right]. \tag{12}$$

Since every vehicle occupies $n + 1$ sites, the density ρ satisfies

$$\frac{1}{\rho} = r\, \mathbb{E}[n + 1 \mid \text{AV}](g^*) + (1 - r)\, \mathbb{E}[n + 1 \mid \text{HV}](g^*), \tag{13}$$

a single scalar equation solved numerically for g^* (bracketed root-finding restricted to $g^* \in (0, q_{\text{HV}})$ whenever $r < 1$, since an HV's own move probability cannot exceed q_{HV}). At $r = 0$ this reduces exactly to the known single-species $v_{\max} = 1$ solution [13]; at $r = 1$ it reduces exactly to the standard deterministic triangular relation $J = \min(\rho, 1 - \rho)$.

4.3. Validation against simulation

Figure 9 and Figure 10 compare Eq. (13) (solved for $J = \rho g^*$) against direct simulation at $v_{\max} = 1$, $p_{\mathrm{HV}} = 0.5$, across $r \in \{0, 0.2, 0.4, 0.6, 0.8, 1.0\}$. Maximum relative error is below 0.53% at every r tested (mean below 0.2%), consistent with finite-size/finite-sampling noise rather than a systematic deviation. This confirms that the shared safety-floor architecture used throughout this paper (Section 2.4) behaves exactly as the underlying exclusion-process theory predicts in the one regime where an exact comparison is possible, supporting the correctness of the simulation framework used for the $v_{\max} = 5$ results in Section 3.

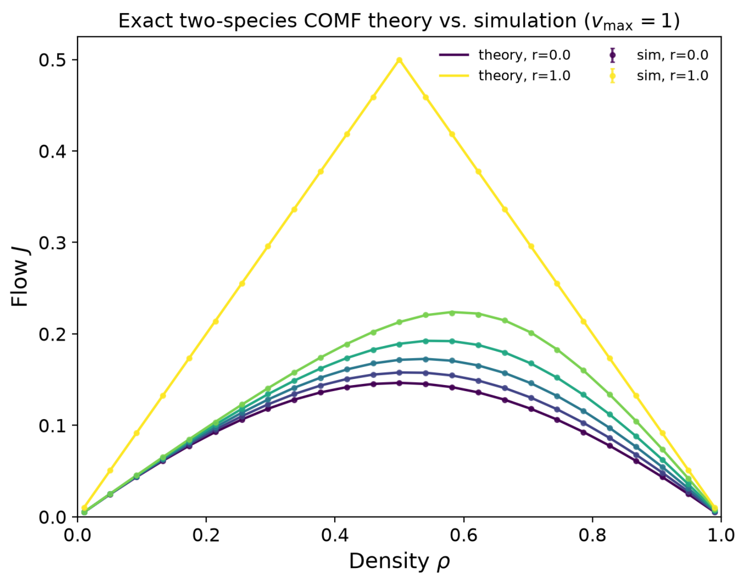

Figure 9: Exact two-type theory (solid curves) vs. simulation (points with error bars) at $v_{\max} = 1$, across AV fraction r.

5. Discussion and Conclusions

Taken together, these results argue that "AV penetration improves traffic flow" is not a single, model-independent statement. Both the magnitude (Section 3.1, Table 1) and even the direction of the capacity effect (the IDM finding) depend on which AV car-following model is assumed, and the hysteresis-suppression benefit (Section 3.2) is not proportional to penetration

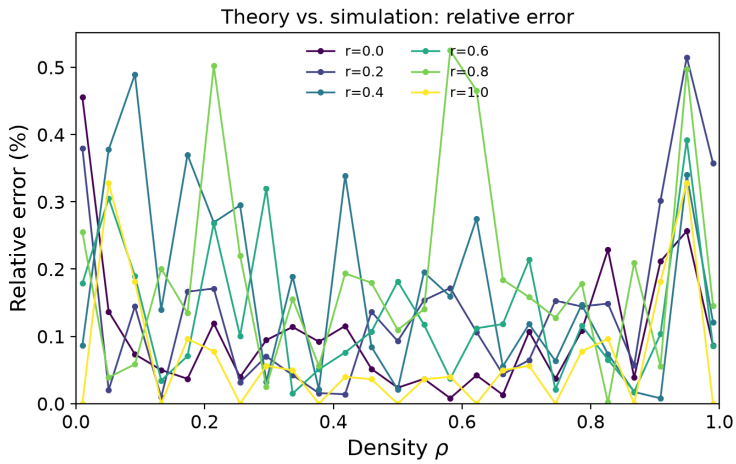

Figure 10: Relative error between theory and simulation, by density and AV fraction r.

rate. A policy statement of the form "X% AV penetration is needed for Y% flow improvement" therefore requires specifying both a car-following model and whether flow or metastability suppression is the target outcome measure; the two need not track each other, since one depends on the discretization-scale interaction of a control law's own gains (Section 3.1, Finding 2), while the other depends on how the slow-to-start asymmetry aggregates over the standing-vehicle population (Finding 3). Neither the sign nor the magnitude of the AV benefit is a property of AV penetration alone.

Note that both the ACC/CACC calibration-vs-lattice-scale effect and the exact vanishing of hysteresis at $r = 1$ trace back to the same underlying structural fact used in Section 4: on a single-lane ring with no passing, no two adjacent vehicles can sustain different long-run average velocities, since a persistent gap between differing average speeds would grow or shrink without bound on a bounded ring [4].

Limitations.. All results are for a single lane with no lane changing, a simplification we adopt for three reasons rather than for convenience alone. First, it preserves direct comparability with the mixed-autonomy NaSch literature cited in Section 1, which is itself single-lane in almost every case. Second, and more fundamentally, the exact result of Section 4 depends on the no-passing assumption in an essential way: the shared-velocity argument underlying both Evans's (1997) disordered-exclusion-process solution and our hysteresis-suppression result (Finding 3, Section 3.2) requires that a vehicle's identity, and hence its long-run average velocity, never changes; once passing is permitted this no longer holds, and no analogous exact solution is available. Third, no lane changing is the more defensible approximation specifically in the congested, jam-forming regime that both the hysteresis (Section 3.2) and analytical-validation (Section 4) analyses target, since gaps wide enough to support a lane change become increasingly rare as density approaches ρ_c. Note that the same assumption is correspondingly harder to defend in the free-flow regime, where a real vehicle could overtake a slower leader rather than being held behind it indefinitely; the model may therefore underestimate free-flow capacity, particularly for rule combinations in which a single slow leader suppresses an entire following queue (e.g., IDM at low r, Finding 1) that a multi-lane road would instead allow to pass. We speculate that the hysteresis-suppression threshold identified in Finding 3 would shift toward lower AV fractions once lane changing lets AVs dissolve HV-induced jams more efficiently, but this remains untested. Extending the present framework to multiple lanes, and re-examining whether the rule-dependence and hysteresis-suppression findings above survive once passing is allowed, remains an important direction for future work.

The ACC/CACC parameters are one literature-consistent choice (Section 2.4); we did not re-run a corresponding sensitivity check for them; the CACC/naive equivalence at $r = 1$ (Finding 2) is therefore not a demonstration that CACC's control law is inherently insensitive to calibration, only that the one calibration tested here is floor-bound, and it remains open whether Finding 2 is similarly calibration-dependent or a more robust consequence of the shared safety floor—this is a natural next step. For IDM, we already confirmed (Section 3.1, Table 2) that the qualitative sign of the capacity effect is calibration-dependent, not fixed, so Finding 1 should be read as a statement about the sensitivity of IDM-based AV rules to their acceleration/deceleration calibration rather than a fixed property of IDM. The slow-to-start (Section 3.2) and car-following-rule (Section 3.1) mechanisms were also not yet combined: we did not test whether IDM, ACC, or CACC's smoother, non-random deceleration dynamics interact differently with VDR-induced hysteresis than the naive deterministic rule tested in Section 3.2.

Implications for AV deployment.. We infer the practical implications of these results for near-term AV deployment. Current and near-term AV penetration rates on public roads remain well below the $r \approx 0.6$ threshold at which hysteresis suppression becomes appreciable in our results (Finding 3, Section 3.2); at these low penetration levels, the dominant lever available to planners is therefore the capacity effect of Section 3.1, not hysteresis suppression. Our results indicate that this capacity effect is not guaranteed to be positive: a fleet of AVs governed by an IDM-type rule calibrated on human driving data, without recalibrating acceleration and deceleration capability for the AV's own actuation, may reduce rather than improve flow relative to human-only traffic (Finding 1). The reduced-gap collision-safety pitfall identified in Section 2.3 is a similarly concrete caution: a control law verified safe in homogeneous AV testing is not thereby verified safe in the mixed populations that any real deployment will pass through en route to full penetration. We expect these calibration- and safety-related cautions to be at least as relevant at higher, future AV penetration rates as at present.

Summary.. We have compared four AV car-following rules—a simple deterministic rule, IDM, ACC, and CACC—on the same NaSch lattice under a shared, collision-safe architecture, and have shown that AV penetration's effect on capacity is strongly rule-dependent. This includes a case (IDM, under a human-driver-calibrated acceleration/deceleration setting) in which AV penetration reduces flow relative to human-only traffic, an effect that reverses under a more AV-appropriate recalibration and is therefore a calibration sensitivity rather than a fixed property of IDM, and a separate case (CACC) in which a literature-verified control law's distinguishing behavior is masked by discretization-scale saturation against the shared safety floor. We have also found, independently, that AV penetration suppresses slow-to-start-induced hysteresis in a threshold-like, not proportional, manner, with the effect concentrated above roughly 60% penetration and vanishing exactly at full penetration. Along the way, we identified and corrected a collision-safety pitfall in reduced-gap AV models that is invisible in homogeneous populations and manifests only once mixed with stochastically braking vehicles; we believe this methodological point is relevant beyond the present study. Finally, we validated the shared safety-floor architecture underlying all of the above against an exact result at $v_{\max} = 1$: our model reduces there to a special case of an already-known, exactly solvable disordered exclusion process [4], expressed here in car-oriented mean-field language [13] and specialized to the AV-fraction parametrization, with simulation matching this exact prediction to within 0.53% across all penetration rates tested.

As discussed above, the rule-dependence of both the capacity and hysteresis-suppression effects can be traced to how each behavior model interacts with the shared, collision-safe safety floor and, in the hysteresis case, with the aggregate slow-to-start dynamics of the standing-vehicle population; a single AV penetration rate cannot therefore be assigned a single capacity or stability benefit independent of the behavior model in question. Extending the present single-lane framework to multiple lanes with passing, and to the combined VDR-plus-behavior-rule regime noted above, remains an important direction for future work.

CRediT authorship contribution statement

Mitsuyoshi Yagyu: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Writing – original draft, Writing—review & editing.

Declaration of competing interest

The author declares that he has no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The simulation code, analysis scripts, and result data supporting the findings of this study are available from the corresponding author upon reasonable request.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

[1] van Arem, B., van Driel, C.J.G., Visser, R., 2006. The impact of cooperative adaptive cruise control on traffic-flow characteristics. IEEE Transactions on Intelligent Transportation Systems 7, 429–436. doi:10.1109/TITS.2006.884615.

[2] Barlović, R., Santen, L., Schadschneider, A., Schreckenberg, M., 1998. Metastable states in cellular automata for traffic flow. European Physical Journal B 5, 793–800. doi:10.1007/s100510050504. arXiv:cond-mat/9804170.

[3] Chowdhury, D., Santen, L., Schadschneider, A., 2000. Statistical physics of vehicular traffic and some related systems. Physics Reports 329, 199–329. doi:10.1016/S0370-1573(99)00117-9.

[4] Evans, M.R., 1997. Exact steady states of disordered hopping particle models with parallel and ordered sequential dynamics. Journal of Physics A: Mathematical and General 30, 5669–5685. doi:10.1088/0305-4470/30/16/011. arXiv:cond-mat/9705006.

[5] Guo, X.Y., Zhang, G., Jia, A.F., 2023. Study on mixed traffic of autonomous vehicles and human-driven vehicles with different cyber interaction approaches. Vehicular Communications 39, 100550. doi:10.1016/j.vehcom.2022.100550.

[6] Helbing, D., 2001. Traffic and related self-driven many-particle systems. Reviews of Modern Physics 73, 1067–1141. doi:10.1103/RevModPhys.73.1067.

[7] Kumar, P., Chakraborty, P., Peruru, S.S., 2026. Macroscopic characteristics of mixed traffic flow with deep reinforcement learning based automated and human-driven vehicles. arXiv preprint arXiv:2603.25328 .

[8] Ma, L., Qu, S., Ren, J., Zhang, X., 2023. Mixed traffic flow of human-driven vehicles and connected autonomous vehicles: string stability and fundamental diagram. Mathematical Biosciences and Engineering 20, 2280–2295. doi:10.3934/mbe.2023107.

[9] Maerivoet, S., De Moor, B., 2005. Non-concave fundamental diagrams and phase transitions in a stochastic traffic cellular automaton. European Physical Journal B 42. doi:10.1140/epjb/e2005-00001-9.

[10] Milanés, V., Shladover, S.E., 2014. Modeling cooperative and autonomous adaptive cruise control dynamic responses using experimental data. Transportation Research Part C: Emerging Technologies 48, 285–300. doi:10.1016/j.trc.2014.09.001.

[11] Nagel, K., Schreckenberg, M., 1992. A cellular automaton model for freeway traffic. Journal de Physique I 2, 2221–2229. doi:10.1051/jp1:1992277.

[12] Poudel, B., Li, W., Heaslip, K., 2023. EnduRL: enhancing safety, stability, and efficiency of mixed traffic under real-world perturbations via reinforcement learning. arXiv preprint arXiv:2311.12261 .

[13] Schadschneider, A., Schreckenberg, M., 1997. Car-oriented mean-field theory for traffic flow models. Journal of Physics A: Mathematical and General 30, L69–L75. doi:10.1088/0305-4470/30/4/003.

[14] Treiber, M., Hennecke, A., Helbing, D., 2000. Congested traffic states in empirical observations and microscopic simulations. Physical Review E 62, 1805–1824. doi:10.1103/PhysRevE.62.1805. arXiv:cond-mat/0002177.