Collaborative Multiagent Decision Making for Lane-Free Autonomous Driving

Dimitrios Troullinos
Dynamic Systems and Simulation Laboratory
Technical University of Crete
dtroullinos@dssl.tuc.gr

Georgios Chalkiadakis
School of Electrical and Computer Engineering
Technical University of Crete
gehalk@intelligence.tuc.gr

Ioannis Papamichail
Dynamic Systems and Simulation Laboratory
Technical University of Crete
ipapa@dssl.tuc.gr

Markos Papageorgiou
Dynamic Systems and Simulation Laboratory
Technical University of Crete
markos@dssl.tuc.gr

ABSTRACT

This paper addresses the problem of collaborative multi-agent autonomous driving of *connected and automated vehicles (CAVs)* in lane-free highway scenarios. We eliminate the lane-changing task, i.e., CAVs may be located in any arbitrary lateral position within the road boundaries, hence allowing for better utilization of the available road capacity. As a consequence, vehicles operate in a much more complex environment, and the need for the individual CAVs to select actions that are efficient for the group as a whole is highly desired. We formulate this environment as a multiagent collaboration problem represented via a *coordination graph*, thus decomposing the problem with local utility functions, based on the interactions between vehicles. We produce a tractable and scalable solution by estimating the joint action of all vehicles via the anytime *max-plus algorithm*, with local utility functions provided by *potential fields*, designed to promote collision avoidance. Specifically, the fields have an ellipsoid form that is most suitable for lane-free highway environments. This novel use of *max-plus* with *potential fields* gives rise to a coordinated control policy that exploits only local information specific to each CAV. Our experimental evaluation confirms the effectiveness of our approach: lane-free movement allows for increased traffic flow rates, and vehicles are able to achieve speeds that are both high and close to their desired ones, even in demanding environments with high traffic flow.

KEYWORDS

autonomous driving; lane-free traffic; coordination graphs; max-plus algorithm; artificial potential fields

ACM Reference Format:
Dimitrios Troullinos, Georgios Chalkiadakis, Ioannis Papamichail, and Markos Papageorgiou. 2021. Collaborative Multiagent Decision Making for Lane-Free Autonomous Driving. In *Proc. of the 20th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2021), Online, May 3–7, 2021*, IFAAMAS, 9 pages.

Proc. of the 20th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2021), U. Endriss, A. Nowé, F. Dignum, A. Lomuscio (eds.), May 3–7, 2021, Online.

1 INTRODUCTION

Vehicular traffic is an integral part of society and, with recent technological advances in the automobile industry, autonomous vehicles have already become a reality [10]. A major enhancement of this automation is the capability of communication across vehicles, introducing the *Connected and Automated Vehicles* (CAVs) [8].

Most traffic control research so far has focused on lane-based traffic [31]. However, the dawn of the CAVs era is opening up the possibility of autonomous driving in lane-free environments [26]. In this new paradigm, there is no need for lane-following and vehicles may be located in any arbitrary lateral position within the road boundaries, hence allowing for better utilization of the available road capacity.

Multiagent system (MAS) approaches have been proposed for autonomous driving [6, 34]. However, to the best of our knowledge, no MAS approach to date has tackled autonomous driving and the emergent need for vehicle coordination in lane-free environments. When one attempts to eliminate the restriction of lanes, the entire multiagent autonomous driving setting and its presentation is fundamentally altered. Instead of predefined lateral positions, now we need to consider the whole range of lateral placements, along with the existence of lateral speeds.

Our work in this paper tackles exactly this problem. We employ Coordination Graphs (CGs) to meet the need for intensive coordination in this environment. Autonomous vehicles in a lane-free setting will not be tied to lanes, and may be involved in critical scenarios where a "greedy" action of a single agent might be very inefficient for the adjacent agents. Vehicles in our examined domain operate with different desired speeds, resulting in conflicts in the traffic environment: an action of one single vehicle might lead to a congestion or even a collision between other vehicles. Coordination, or any notion of consideration for other vehicles' actions, should result in better policies. In our approach we use max-plus [19], an anytime message passing algorithm. Max-plus generates a joint policy for all involved agents by relying only on local information according to a provided coordination graph, and attempts to maximize the global utility. As such, the joint policy might limit the performance of a single agent in order to accommodate the overall performance, a property quite appropriate for this domain.

The local information of each vehicle needs to be quantified properly in local utility functions, as we illustrate in the following

sections. It is vital that the objectives of the vehicles are reflected on these local utilities, and therefore when the maximization process of max-plus results in a joint-action that "navigates" the whole multiagent environment to higher utilities, this should also correspond to higher efficiency of the system. We accomplish this by employing *artificial potential fields* [16]. These are designed to have a form that is suitable for this novel lane-free traffic domain and also fitting for our approach. The adopted potential field will be applied as a negative component of the local utilities, reflecting the danger of a collision between two respective vehicles.

Against this background, the main contributions of this paper are the following: we introduce a novel application of max-plus algorithm for a newly established traffic domain; we employ artificial potential fields to form local utility functions, and intertwine these with max-plus; our fields have an ellipsoid form that departs from the one typically used in the literature, and is designed to best fit the domain of interest; and build upon the existing Flow [33] framework and extend its functionalities to accommodate the need for connected vehicles in lane-free traffic simulation environments with SUMO [21].

Our simulation results confirm the effectiveness of our approach. Specifically, we show that by coordinated lane-free traffic we utilize more the available road capacity to allow for increased CAVs flow levels; and we achieve a level of vehicle speed optimization, as vehicles manage to a great extent to reach speeds that are close to their desired ones.

In what follows, in Section 2 we provide background and related work. In Section 3 we present our approach. Then, in Section 4 we showcase our experimental evaluation. Finally, in Section 5 we discuss future directions for this line of research, while Section 6 concludes this paper.

2 BACKGROUND & RELATED WORK

In this section we introduce the necessary background along with relevant related work.

2.1 Coordination Graphs & the Max-Plus Algorithm

Coordination graphs (CGs) for Multiagent Decision Making were first introduced in [9]. CGs model the local interactions in a multiagent system, allowing for scalability in the number of participating agents. The attractive property of CGs is that not all agents interact with one another, and thus the joint action $\mathbf{a}$ of a set of agents that maximizes a global utility (social welfare) $u(\mathbf{a})$ can be obtained much more easily. Agents' interactions are modeled through the graph's edges, and the density and form of the graph dictate the overall complexity of the system.

The max-plus algorithm [19] is an anytime message-passing algorithm that can be used for decentralized coordination of agents in a provided CG. The global utility $u(\mathbf{a})$ (sum of local utilities) is factorized as such: $u(\mathbf{a}) = \sum_{i \in N} f_i(a_i) + \sum_{(i,j) \in E} f_{ij}(a_i, a_j)$. This means that each agent $i \in N$, where N is the set of nodes (agents), has a local utility $f_i(a_i)$, while $f_{ij}(a_i, a_j)$ correspond to a shared utility related to the edge $(i, j) \in E$, where E is the set of edges. Max-plus is an iterative algorithm, where in every iteration each agent i sends locally maximized messages $\mu_{ij}(a_j)$ based on their current maximizing action a_i, to all agents j connected with i ($(i, j) \in E$), $\forall j \in N_i$. Each message is calculated as such:

$$\mu_{ij}(a_j) = max_{a_i} \left\{ f_i(a_i) + f_{ij}(a_i, a_j) + \sum_{k \in N_i \setminus \{j\}} \mu_{ki}(a_i) \right\} + c_{ij} \quad (1)$$

Convergence is only guaranteed when the CG does not contain cycles. However, as discussed in [18], a normalizing value of $c_{ij} = -\frac{1}{|N_k|} \sum_k \mu_{ik}(a_k)$ can be added to normalize the values of messages, so that they do not constantly accumulate when cycles exist in the graph.

Then, each agent i selects a new action a_i that maximizes the received local messages $\mu_{ji}(a_i)$ along with i's local payoff $f_i(a_i)$:

$$a_i = argmax_{a_i} \left\{ f_i(a_i) + \sum_{j \in N_i} \mu_{ji}(a_i) \right\} \quad (2)$$

When the actions are estimated, the algorithm proceeds to the next iteration, and is executed until convergence of the passing messages μ_{ij}, or until a stopping criterion is met. Note that these local utility functions are stateless, and the algorithm provides a coordination strategy only for a given state. Consequently, when used in a multiagent setting, we need to execute max-plus for every new state of the multiagent system, i.e., in every time-step. The anytime aspect of this algorithm allows us to limit the execution time or the number of iterations, with the trade-off being the quality of the solution.

2.2 Related Work

CGs are quite appealing for a highway scenario, as the assumption that CAV agents only interact with others nearby is quite sensible. One limitation in the use of CGs is that they are static, a property which cannot exist in our problem, since the varying position of agents determines the interactions with their surroundings. Additionally, with a static CG, we cannot easily examine road networks with varying sizes of vehicles. To the best of our knowledge, the only work that employs GCs, static or otherwise, for multiagent autonomous driving is [34], but for a lane-based traffic environment with assumptions that cannot be adapted to a lane-free environment. Specifically, [34] provides a dynamic CG approach, but their environment is lane-based, and the vehicles have a constant number of neighbors (with the use of fictitious agents when necessary). Moreover, they rely on the variable elimination (VE) [18] algorithm to find the joint action, instead of max-plus. Besides the time-related limitations that VE would impose to our approach, since the vehicles' constructed graph in our domain is much more dense, the fundamental reason why we do not adopt VE is the existence of cycles in our constructed CG. VE is an exact algorithm that requires an acyclic graph in order to converge to a solution. Thus, the solution in [34] is arguably tied to their specific environment and therefore cannot be transfered to this domain, where we consider a varying number of adjacent vehicles.

CGs, along with max-plus, have already been applied to address other traffic related issues, notably [20, 30], which provide a solution for urban traffic lights control. In another popular study from the field of Multiagent Systems for traffic establishes a novel multiagent approach for intersection management [7]. Besides the fact that the

aforementioned studies deal with different problems, they all base their formulation on a lane-based traffic environment.

CGs are commonly tied with Multiagent Reinforcement Learning (MARL) coordination approaches. As discussed in [6], only a small amount of studies exist that address multi-vehicle control using MARL techniques in terms of joint action learners. Moreover, to the best of our knowledge, no work to date used collaborative MAS techniques such as CGs combined with max-plus; and no multi-vehicle control work to date has intertwined max-plus with artificial potential fields for collision avoidance in multi-agent driving scenarios, as we do in this paper..

3 OUR APPROACH

In this section we first define the *Lane-Free Driving* environment setting, and then proceed to discuss our approach on this newly established domain.

3.1 Lane-Free Driving

We follow largely the approach outlined in [26] in order to formulate our lane-free driving environment. Specifically, we consider a highway environment populated with automated vehicles. These automated vehicles are equipped with certain capabilities, namely V2V (vehicle to vehicle) and V2I (vehicle to infrastructure) communication. Homogeneity across all vehicles is assumed, i.e., the vehicles have the same dynamics, meaning that they have the same characteristics with respect to their steering and acceleration capabilities. Now, each vehicle i wishes to attain or maintain a specific *desired speed* $v_{d,i}$. The environment is discrete-time, meaning that the vehicles take an action at discrete time steps, corresponding to a specified time-interval. However, the vehicle dynamics are continuous, and the control of each vehicle is performed by providing two corresponding accelerations for the longitudinal and lateral axis, respectively. Note that some specific aspects of vehicle control are automated by the environment itself. More specifically, the vehicles are automatically restricted within the road boundaries and negative speed is not permitted.

3.2 Coordination Graphs and Max-Plus for Coordinated Lane-Free Traffic

To adopt the *max-plus algorithm* [18, 19] for our setting, we need to construct a CG, i.e., define an undirected graph based on the local interactions between vehicles (agents). We first need to form an undirected graph, and then determine the local payoff values.

Naturally, the nodes depict each agent i from a set of agents N, and the edges E correspond to the need for coordination between all pairs of vehicles $\forall(i, j) \in E$. The main benefit of CGs is the ability to exploit these local interactions between agents, and avoid solving the problem by considering the total joint action space. Initially, each agent observes its surrounding environment, i.e., other nearby agents. We allow for a visibility (upstream and downstream traffic[1]) of o_m meters (longitudinal distance). Note that for a pair $(i, j) \in E$, where the agents i, j are within observation distance, it is not obvious that there is a need for this edge to exist, i.e., a need for i, j to coordinate their actions. This can be assessed according to the values of a_b, b_b (will be defined in the next Section 3.3) and the current longitudinal (dx_{ij}) and lateral (dy_{ij}) distance, meaning that if $(|dx_{ij}| > a_b \vee |dy_{ij}| > b_b)$, then this edge is omitted. Thus, for each agent i, the set of neighbors is:

$$N_i^p = \{j : j \in N, j \neq i, (|dx_{ij}| \leq min(a_b, o_m) \wedge |dy_{ij}| \leq b_b)\} \quad (3)$$

We further restrict the number of edges by allowing up to maximum number of edges e_f, e_b, for forwards and backwards, in order to have control over the graph's density, meaning that from a set of available agents N_i^p, a subset $N_i \subseteq N_i^p$ will be selected according to this limitation in terms of maximum number of edges. This selection process is performed based on the euclidean distance $dxy = \sqrt{dx^2 + dy^2}$ between i and each prospective agent $j \in N_i^p$. The choice of euclidean instead of the longitudinal distance is made since the number of edges to consider might be limiting, especially in wider roads or in dense traffic scenarios. For example, considering a pair (i, j), where j is in front of i and there is actual need for coordination, i may ignore j if there are more than e_f vehicles between i and j in terms of longitudinal position.

We define a compact representation for the current local state relevant to agents i, j as a tuple $s_{ij} = \{dx_{ij}, y_i, y_j, v_{x,i}, v_{x,j}, v_{y,i}, v_{y,j}\}$.[2] The required parameters $(dx_{ij}, y_i, y_j, v_{x,i}, v_{x,j}, v_{y,i}, v_{y,j})$ of $U_{ij}(s_{ij})$ depend on a given state s_{ij} of the respective agents. The problem is to obtain a joint action $\mathbf{a}$ that maximizes the global utility $u(\mathbf{a})$ (sum of local utilities), i.e.: $\mathbf{a}^* = \text{argmax}_{\mathbf{a}} u(\mathbf{a})$. The structure of the global utility, as defined in [18], is $u(\mathbf{a}) = \sum_{i \in V} f_i(a_i) + \sum_{(i,j) \in E} f_{ij}(a_i, a_j)$, where f_i corresponds to some local utility for agent i, and $f_i j$ is a function that models some local utility shared by agents i and j (as explained in Section 2.1 above). We consider a factorization of the utility only through the f_{ij}, as shown in Fig. 1, thus:

$$u(\mathbf{a}) = \sum_{(i,j) \in E} f_{ij}(a_i, a_j) \quad (4)$$

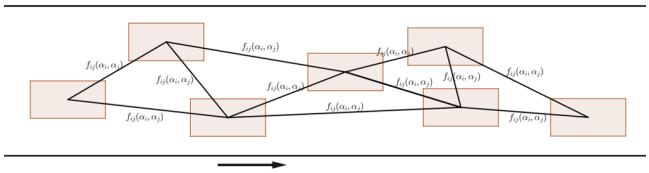

Figure 1: Example of a CG on a lane-free highway scenario.

To comply with max-plus, the actions of vehicles need to be discrete. Thus, we define a set of 5 possible actions for each agent:

- a_0: zero acceleration in both axes
- a_1: longitudinal acceleration of $c_a \frac{m}{s^2}$
- a_2: longitudinal deceleration of $c_d \frac{m}{s^2}$
- a_3: lateral acceleration $c_l \frac{m}{s^2}$ towards left
- a_4: lateral acceleration $c_l \frac{m}{s^2}$ towards right

Note that the local utility function f_{ij} depends on the joint action of these agents. We know the model of the environment, as we consider that vehicles move according to equations of motion. Therefore, we can predict the next state of the parameters

[1] Downstream (upstream) vehicles from the view of a focal vehicle i are located on the front (back) of it.

[2] Relative distances and speeds are calculated with respect to agent i, e.g., $dy_{ij} = y_j - y_i$.

(dx, dy, dv_x, dv_y) given a local joint action $\langle a_i, a_j \rangle$ according to equations of motion, meaning that for a local state s_{ij}, and a given set of joint actions $\langle a_i, a_j \rangle$, we can calculate the next state $s'_{ij} = \{dx'_{ij}, y'_i, y'_j, v'_{x,i}, v'_{x,j}, v'_{y,i}, v'_{y,j}\}$. The local state transition can be formulated with a known and deterministic transition function, which provides us with the next local state $T(s_{ij}, a_i, a_j) = s'_{ij}$.

Now, each agent's i goal is two-fold. Agents need to avoid collisions with nearby agents; and also operate according to a desired speed $v_{d,i}$. The local payoff f_{ij} should incorporate both goals as a local edge utility function. Essentially, we use the transition function for all combinations of joint action pairs, and then provide the value of the potential field for the resulting s'_{ij} to the local payoff $f_{ij}(a_i, a_j)$, intuitively to "inform" the agents on the results of their interaction.

Regarding the values of f_{ij}, we need to use negative values when vehicles (i, j) are "close". This effect is provided by the "repulsion" function U_{ij} that will be defined in the next Section 3.3. This function serves the purpose of quantifying the danger of collision between two vehicles (i, j). That is, when the vehicles are quite close, and therefore there is more danger, the value of U_{ij} is much higher. This is illustrated in Fig. 2. The difference between the two diagrams (a),(b) in Fig. 2 is based on the relative speeds (longitudinal and lateral) of the two vehicles. As shown in Fig. 2, in (a), where the relative speeds are both $0\frac{m}{s}$, i.e., both vehicles have exactly the same speed (longitudinal and lateral), the region where U_{ij} is higher is much smaller (compared to (b)), meaning that the vehicles are permitted to be closer to one another. In contrast, in (b), where there is a difference in relative speeds, we observe a much broader region, resulting in a more "conservative" utility function due to the additional safety needed when one vehicle approaches the other with higher speed.

The negative "collision avoidance"-related payoff is provided in the upper branch of Eq. (5) below, when the condition $(|dx'_{ij}| \leq a'_b) \wedge (|dy'_{ij}| \leq b'_b)$ is met, meaning that the vehicles are quite close. Values of a_b, b_b reflect a safety longitudinal and lateral distance respectively, which will be defined in the next Section 3.3. By contrast, if the joint action results in a state where there is no perceived danger of collision between these vehicles, the goal of achieving the desired speeds can be accommodated. This is achieved by providing a positive reward via the function $c_s \cdot r_{v,ij}$ as shown in the lower branch of Eq. (5). Thus, the local payoff function f_{ij} shared by i, j at local state s_{ij} is:

$$f_{ij}(a_i, a_j) = \begin{cases} -U_{ij}(s'_{ij}), & (|dx'_{ij}| \leq a'_b) \wedge (|dy'_{ij}| \leq b'_b) \\ c_s \cdot r_{v,ij}, & else \end{cases} \tag{5}$$

$$r_{v,ij} = r_{v,i} \cdot \frac{1}{|N_i|} + r_{v,j} \cdot \frac{1}{|N_j|} \tag{6}$$

where $|N_i|$ is the number of edges that contain agent i. Note that the form of $r_{v,ij}$ is influenced by the update rule of Sparse Cooperative Q-learning in edge-based decomposition schemes [18].

We define $r_{v,i}$ as a linear function based on current speed $v_{x,i}$, normalized according to the desired speed $v_{d,i}$. This speed utility component is defined as follows:

$$r_{v,i} = \begin{cases} \frac{v_{x,i}}{v_{d,i}}, & (v_{x,i} < v_{d,i}) \\ \frac{2 \cdot v_{d,i} - v_{x,i}}{v_{d,i}}, & else \end{cases} \tag{7}$$

To sum up, the local payoff f_{ij} is negative when there is a need for coordination due to an expected collision, and positive otherwise, reflecting the goal of reaching a desired speed. Since both the CG and the local payoffs are established, we can now directly apply max-plus with the anytime extension to our problem.

As discussed in [18], a CG with a cyclic graph is not guaranteed to converge to a solution, due to the accumulation of messages from the adjacent agents. However, in practice, graphs with cycles appear to converge in multiple applications, with the addition of a normalizing parameter c_{ij} for outgoing messages, as discussed in Section 2.1. Note that in our domain, we cannot avoid the existence of cycles in the constructed graph. We could try to dismiss an edge (i, j) that creates a cycle and thus not attempt to take (i, j) into our coordination problem. However, this may lead to a state where a collision is now inevitable. Consider for example a simple scenario with $N = 3$ agents that are nearby, and more specifically for agents i, j, k we have that $dx_{ij} = 0$, and both $dx_{ik} > 0, dx_{jk} > 0$, i.e., agents i, j are side by side, k is behind, and $v_{d,k} > v_{d,i} = v_{d,j}$. Now, this means that agent k wishes to accelerate and pass in between agents i, j. To coordinate in this scenario, we would need a fully connected graph, hence forming a cycle. Thought the existence of cycles is inevitable in our scenario, our results show that the approach is effective in practice.

3.3 Artificial Potential Fields

In this section we describe the artificial potential fields used in our local utility functions (Eq. (5)) to promote collision avoidance.

Consider two vehicles, i, j. We wish to quantify a potential collision between these two vehicles i, j. In a regular lane-based scenario, this could be established through metrics used in microscopic traffic models like headway, or time to collision (TTS), [14, 15] taking also into account a potential lane-change. For lane-free settings, we need to consider a 2-dimensional space, and how the relative longitudinal and lateral movement of these two vehicles can properly reflect a potential collision. *Artificial potential fields* [3, 16], used for robot path planning, and also for motorway traffic [13, 29] are appropriate for this type of formulation. Normally, repulsive fields are applied for obstacle avoidance. We wish to quantify a potential collision in a similar manner, in order to establish a local utility function. The formulation that we proceed to employ is In particular, for our approach, a repulsive field will correspond to a negative local utility.

Typically, repulsive fields are based on the distance of the obstacle from the respective agent [3]. For a highway scenario, the longitudinal and the lateral distance provide different information, and thus should be both taken into account. If we just account for the euclidean distance, we lose useful information. While two vehicles might have a certain distance dxy, we do not know if they are side by side, or front-back, or anything between these two extremes.

An ellipsoid function thus better captures a potential collision in this domain, since its shape can stretch over one dimension more

than the other. The particular ellipsoid form we adopt in this work, is the following [29]:

$$E(dx, dy) = \frac{m}{\left(\left(\frac{|dx|}{a}\right)^{p_x} + \left(\frac{|dy|}{b}\right)^{p_y} + 1\right)^{p_t}} \quad (8)$$

where dx, dy are the longitudinal and lateral distance of the respective center points of the two vehicles i, j. Parameters a, b adjust the range of the field for the two dimensions x, y respectively, and the exponents p_x, p_y, p_t affect the overall shape. Finally, m defines the magnitude when the distances are close to 0, i.e., $max(E(dx, dy)) = m$.

In our problem, we thus use two such ellipsoids to model the "repulsion" function U_{ij} that appears in Eq. (5). Specifically, U_{ij} is composed by a "critical region" ellipsoid and a "broader region" ellipsoid, as follows:

$$U_{ij}(s_{ij}) = E_c(dx_{ij}, dy_{ij}) + E_b(dx_{ij}, dy_{ij}, dv_{x,ij}, dv_{y,ij}) \quad (9)$$

Parameters of $E_c(dx, dy)$ are defined so that they account for the "critical region", i.e., when the two vehicles are at least within a certain range and are about to collide. Thus, a_c, b_c are based on the corresponding length and width of the two vehicles. Therefore, for $E_c(dx, dy)$ we define:

$$a_c = \frac{l_i + l_j}{2} + s_a \quad (10)$$

$$b_c = \frac{w_i + w_j}{2} + s_b \quad (11)$$

where $l_i(w_i)$ is the length (width) of vehicle i, and the quantities s_a, s_b provide additional safety.

The use of a field that only accounts for the critical region is inadequate for our approach, since it imposes a "myopic" view for the vehicles. We require an additional field $E_b(dx_{ij}, dy_{ij}, dv_{x,ij}, dv_{y,ij})$, that captures the "broader region", and can inform the vehicles when a collision is about to happen unless a proper reaction occurs. For this component, we determine the values of parameters a_b, b_b based on the relative speed (longitudinal and lateral) between the corresponding vehicles. Consequently, we include additional parameters for $(dv_{x,ij}, dv_{y,ij})$. This is crucial, and the parameter choices for $E_b(dx_{ij}, dy_{ij}, dv_{x,ij}, dv_{y,ij})$ have the potential to highly affect the performance.

Then, for $E_b(dx_{ij}, dy_{ij}, dv_{x,ij}, dv_{y,ij})$:

$$a_b = \frac{dv_{x,k}^2}{2 \cdot |u_x^{d,max}|} + \tau \cdot v_{x,b} + a_c \quad (12)$$

$$b_b = \frac{dv_{y,k}^2}{2 \cdot |u_y^{d,max}|} + b_c \quad (13)$$

and:

$$dv_{x,k} = \begin{cases} dv_x, & dv_x \cdot dx < 0 \\ 0, & \text{else} \end{cases} \quad (14)$$

$$dv_{y,k} = \begin{cases} dv_y, & (dv_y \cdot dy < 0) \wedge (|dy| > \frac{w_i + w_j}{2}) \\ 0, & \text{else} \end{cases} \quad (15)$$

where dv_x (dv_y) is the relative longitudinal (lateral) speed between the vehicles i, j, $|u_x^{d,max}|$ ($|u_y^{d,max}|$) is the maximum longitudinal (lateral) deceleration, and τ is a reaction time parameter. Essentially, a_b and b_b reflect the longitudinal and lateral safety distances respectively. In Eq. (12), the first term introduces a safety distance according to the longitudinal relative speed of the corresponding agents. For a pair of connected agents (i, j), a positive dv_x indicates that agent j has a higher longitudinal speed. In a scenario where agent j is in front of i, i.e., $dx > 0$, then there is no immediate danger of collision. Eq. (14) serves to take this condition into consideration. The condition $dv_x \cdot dx < 0$ is true only when the longitudinal speed of the upstream vehicle is higher, i.e., if we consider only the longitudinal axis, the vehicles would collide. The same condition is also applied in the lateral dimension, but with the additional condition $(|dy| > \frac{w_i + w_j}{2})$, since it should be applied when the agents are not directly one in front of the other. The second term of Eq. (12) is a safety distance relating the longitudinal speed $v_{x,b}$ of the vehicle on the back, and a parameter for reaction time τ. This term is introduced since the safety distance should be influenced by the current speed of the vehicle. Otherwise, the vehicles would react similarly irrespectively of their current speed.

In Fig. 2 we showcase a constructed potential field, for $v_{x,b} = 10\frac{m}{s}$ and two different configurations for $dv_{x,k}, dv_{y,k}$. In both configurations, the critical region near the center $(0, 0)$ is clearly distinguished from the outside region due to the $E_c(dx, dy)$ component in Eq. (9). In the first configuration (a), we have $dv_{x,k} = 0\frac{m}{s}, dv_{y,k} = 0\frac{m}{s}$, meaning that the additional safety outside of the "critical region" is due to the current speed $v_{x,b} = 10\frac{m}{s}$, as opposed to the second configuration (b), where the influence of $dv_{x,k}, dv_{y,k}$ is clear. In this case, $dv_{x,k} = 10\frac{m}{s}, dv_{y,k} = 2\frac{m}{s}$, and the additional stretch over both dimensions is quite distinct.

Essentially, the range of the field with respect to the distances (dx, dy) between two vehicles is affected by the relative speeds, and this range reflects the additional distance they should have when relative speeds are higher.

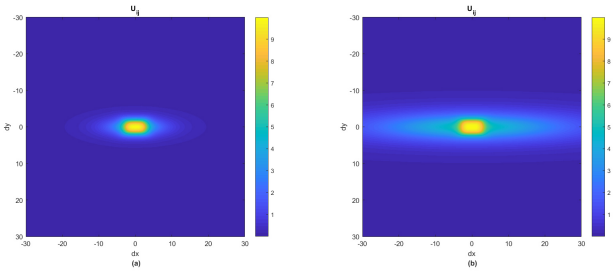

Figure 2: Heatmap of the constructed potential field (corresponds to the value of U_{ij}), for two different configurations of $dv_{x,k}, dv_{y,k}$. In (a): $dv_{x,k} = 0, dv_{y,k} = 0$, and in (b): $dv_{x,k} = 10, dv_{y,k} = 2$.

4 EXPERIMENTAL EVALUATION

Our experimental evaluation is conducted in SUMO [21], a popular traffic simulation tool. To properly examine lane-free traffic, we

Table 1: Potential field, coordination graph, max-plus & simulation parameters

Parameter	*Value*	*Parameter*	*Value*
s_a	1 m	max iterations	10
s_b	1 m	max execution time	2 s
τ	0.5 s	c_s	0.01
$p_x(E_c)$	4	c_a	2 $\frac{m}{s^2}$
$p_y(E_c)$	4	c_d	2 $\frac{m}{s^2}$
$p_t(E_c)$	2	c_l	1 $\frac{m}{s^2}$
$p_t(E_b)$	2	Highway length (width)	5 km (10.2 m)
$p_x(E_b)$	2	Vehicle length (width)	3.2 m (1.6 m)
$p_y(E_b)$	2	Simulation time	1 hr
o_m	50 m	time-interval	0.25 s
e_f	4	v_d range	$[25, 35]\frac{m}{s}$
e_b	4	$v_{x,init}$	$25\frac{m}{s}$
e_μ	0.001		

extended the code-base of the Flow Framework [33] to support lane-free traffic simulations, and incorporated the use of centralized and decentralized algorithms for custom vehicle control[3].

In more detail, we have designed and implemented the proper extensions that enable lane-free vehicle movement to run simulations with SUMO. Besides the lane-free movement, our implementation provides the necessary development for custom inflow rates, allowing for more flexibility in the examined domain. With an inflow parameter, we define the number of vehicles that will be spawn in one hour. Then, the vehicles spawn randomly in appropriate time-intervals (based on the provided inflow parameter) in any lateral position within the road boundaries.

Any type of lane-free controller for vehicles can be developed and adapted to our extension, provided that it yields the two accelerations (longitudinal and lateral) in every simulation time-step. We also integrate the desired speed of each vehicle, generated randomly within a provided range, and provide the simulation results based on these values. Finally, we allow for centralized controllers due to the flexibility they introduce, especially for coordinated vehicle moving strategies.

In our simulations, we include a simple heuristic rule for agents that have zero adjacent agents, $|N_i| = 0$. In such a case, these agents simply perform the most appropriate action in order to reach their desired speed $v_{d,i}$.

In Table 1, we provide the parameter settings related to the potential field, graph structure & max-plus and the simulation environment accordingly. We set up a highway scenario with the specified parameter choices, and then, for each scenario tested, we provide the *inflow* parameter that determines the number of vehicles that will be spawned within an hour. Every new vehicle i is spawned at a random initial lateral position within the road boundaries, with an initial longitudinal speed $v_{x,init}$ and a desired speed $v_{d,i}$ for i is defined by drawing a random sample from a uniform distribution based on the specified range of speeds (shown in Table 3)—i.e., $v_{d,i} \sim U(25, 35)$.

[3]The code for our work in this paper, including our extension of the Flow Framework, is available at: https://bitbucket.org/dtrou/flow_lanefree

The examined simulation inflow settings range from 1800 to $9000\frac{vehs}{hr}$. Just for reference, in a real-life highway, the peak flow would be $\sim 2000\frac{vehs}{hr \cdot lane}$ [24]. Note that the selected road width corresponds to a standard 3-lane road, hence to a peak flow of about $6000\frac{vehs}{hr}$ for lane-based traffic.

The effectiveness of our approach is evaluated with the following metrics. These are: *average speed*, *average speed deviation from the desired speed* along with the respective standard deviation, and *average delay*. Delay t_d is the time difference between the actual time t_a that the vehicle spent inside the motorway and the ideal time that the vehicle i would have spent with a constant speed equal to its desired speed $v_{d,i}$:

$$t_d = t_a - v_{d,i} \cdot l_r \tag{16}$$

where l_r is the length of the motorway.

In Fig. 3 we demonstrate the average speed of all agents. Naturally, the average speed is reduced when the inflow of vehicles is increased. However, due to coordination, the vehicles manage to mitigate this reduction[4], achieving a smooth, low-slope linear drop in average speed as inflow increases.

Then, in n Fig. 4 we include results from a regular lane-based traffic environment with SUMO, for the same road length, type of vehicles, and with the default SUMO parameters regarding the vehicles' control strategy. We need to clarify that for the lane-based environment, SUMO does not provide the ability to have desired speeds selected randomly as we do in our lane-free extension. Instead, we distribute the range of desired speeds (see Table 1) evenly to 3 and 10 different classes of vehicles, i.e., assign one desired speed value to each class.

The Fig. 4 lane-based simulation results show how the average speed reduces drastically as traffic inflow increases in regular lane-based scenarios; when viewed alongside the lane-free average speeds of Fig. 3, also underline the benefits of lane-free movement. In contrast to lane-free movement, lane-based traffic results to a sharp reduction in average speed for the inflow rates examined. Beyond a certain point, the average speed for lane-based traffic decreases even below the lowest desired speed ($25\frac{m}{s}$), exhibiting a behaviour worse even than the one we would obtain by using a single class of vehicles with a desired speed of $25\frac{m}{s}$. This is illustrated by the "slow-class" simulation results, where all vehicles belong in a class with desired speed of $25\frac{m}{s}$. We note that lane-based simulations cannot reach the flow of $9000\frac{vehs}{hr}$ that is achieved in lane-free conditions, and the last measurement in Fig. 4 for all cases corresponds to a 'peak' flow of about $7300\frac{vehs}{hr}$ that the lane-based simulation scenarios can handle given the capacity of each lane.

As discussed in Section 3.1, we are interested in vehicles that wish to attain or maintain a desired speed, and the coordination problem emerges exactly because of the different individual goals of the vehicles, i.e., we need to coordinate multiple vehicles with different speeds. So, the average speed by itself is not so informative of the overall performance, as we should also examine whether the vehicles tend to diverge substantially from their desired speed.

[4]In the unlikely (and also not very demanding with respect to coordination, and thus not very interesting) scenario that *all* vehicles have the same desired speed, the vehicles in our simulations maintain an average speed almost equal to their desired one.

Thus, we report in Fig. 5 the average deviation $v_{dev,avg}$ between the current speed $v_{x,i}$ and the desired speed $v_{d,i}$. It should be clarified that this speed deviation for a time-step t is calculated as: $v_{dev,t} = \frac{1}{|N_t|} \sum_{i=1}^{|N_t|} |v_{d,i} - v_{x,i}|$, where N_t the set of agents inside the road network at time-step t. This means that we do not distinguish between positive or negative deviations, and therefore the obtained results are more indicative of the overall effectiveness in terms of attaining the vehicles' desired speed.

Additionally, we examine the average speed deviation in more detail in Fig. 6, where we separate the vehicles' population based on their corresponding desired speed (That is, each line in Fig. 6 corresponds to the portion of the population having a specific desired speed range.). As expected, we observe that the slower vehicles tend to have a small speed deviation, even as the flow increases. But, this is not the case for faster vehicles, where the deviation is more affected by the flow. This is actually a desired effect, showing that coordination guides the vehicles to employing a more "cautions" behaviour when they find themselves within a *dense* traffic region.

Furthermore, in Fig. 7 we showcase the delay t_d as defined in Eq. (16). Based on the simulated results, the average delay exhibits a linear relationship with respect to the inflow rates we examined, even with the highest inflow rate of 9000 $\frac{vehs}{hr}$. This evidently corresponds to the increase in average speed deviation from the desired speed (Fig. 5). Yet, it clearly demonstrates the benefits of coordination, as delays are maintained very small even as the traffic flow increases substantially.

Clearly, the lane-free environment allows for higher utilization of the road capacity: as mentioned above, in our lane-based simulations the maximum inflow attained is about 7300 $\frac{vehs}{hr}$, whereas with coordination in lane-free traffic we can examine even higher inflows. Moreover, a frequent observation in simulation was the utilization of the full road width in cases where 4 vehicles were side-by-side, and were breaking formation to let faster vehicles pass through when necessary.

Regarding the max-plus algorithm, we limit both the execution time and maximum number of iterations,[5] and we require the value of the messages to differ by a maximum threshold e_μ at time t to determine convergence (as shown in Table 1). We observe that the algorithm almost always manages to converge for inflow rates up to 7200 $\frac{vehs}{hr}$ before it reaches either limit.

By disengaging agents whenever possible, as described in Section 3.2, the constructed CG is sparse, unless high densities of vehicles form in some regions (this is more frequent with higher inflows). So, with the highest inflow rate examined (9000 $\frac{vehs}{hr}$), max-plus does not always converge for the specified limits, since the vehicles would exhibit higher densities in larger regions, but nonetheless manages to provide joint actions of high quality, as demonstrated by our simulation results.

We remark that our lane-free simulations are collision-free up to 5400 vehs/hr. Notice that this inflow corresponds to the maximum capacity of three-lane real-world highways. However, for simulations with inflows of 7200 and 9000 vehs/hr, we observe 17 and 27 crashes respectively. These numbers correspond approximately

[5]Execution of the algorithm is terminated when either of these limits is surpassed, and the resulted joint policy is returned.

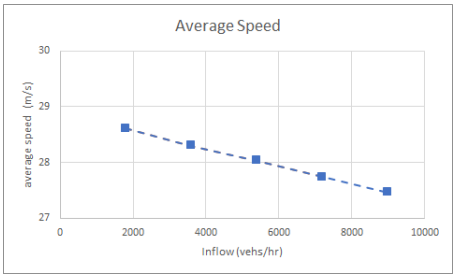

Figure 3: Average speed of vehicles for different inflow rates in lane-free traffic. Desired speeds v_d are uniformly distributed within the range $[25, 35]$.

Figure 4: Average speed of vehicles for different inflow rates on a regular lane-based scenario.

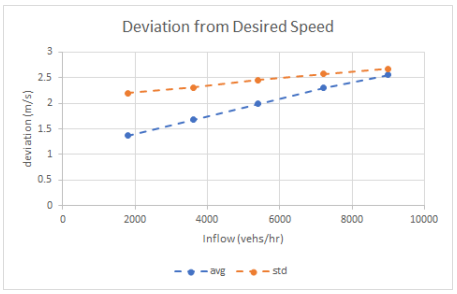

Figure 5: Average and standard deviation of speed deviation from desired speed for different inflow rates.

to only 0.47% and 0.6% of the total number of the inflow vehicles respectively; occur over a total of roughly hundreds of thousands of interactions (i.e., potential collisions) between vehicles; and are observed only in very dense regions of the simulations. One can easily make the fields more conservative to avoid collisions altogether, with the expected trade-off regarding desired speeds reached. In addition, the distribution of desired speeds of the simulated vehicles is an important parameter that can play a role in this trade-off. For instance, when *all* 7200 or 9000 vehicles are assumed to have the same desired speed (e.g., that of 35 m/s), we observe *no* collisions, even with our current potential field configuration.

5 FUTURE RESEARCH DIRECTIONS

Our work in this paper opens up a number of research directions for coordination and optimal control in lane-free environments,

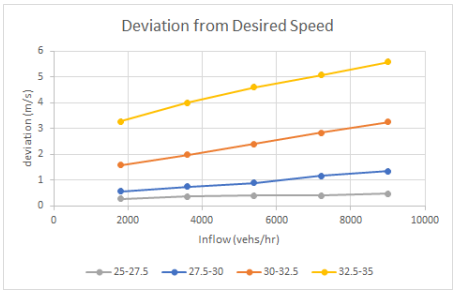

Figure 6: Average speed deviations from desired speed for different desired speed sub-populations for each traffic population inflow rate. Each line corresponds to approximately 25% of the inflow.

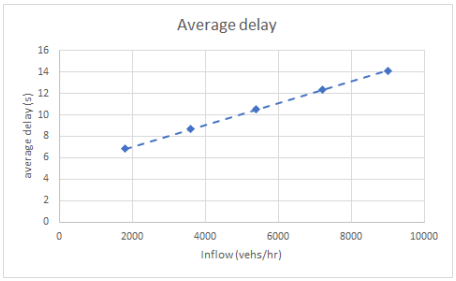

Figure 7: Average delay of vehicles for different inflow rates.

both in the microscopic and macroscopic traffic modeling levels (i.e., modeling both the behaviour of individual vehicles and their interactions, and the aggregate behaviour of the traffic flow). To begin, the most imminent extension is the inclusion of Reinforcement Learning to the current approach, and consequently the optimization of the joint policy for the expected (discounted) long-term rewards. Recent works on MARL that employ the max-plus algorithm could potentially be adapted [2, 5], but it should be noted that the above-mentioned studies are intrinsically different from our approach, as we have an environment with dynamic interactions and local rewards.

We believe that the homogeneity of the agents can be exploited in order to introduce dynamic interactions of agents as a Collaborative Multiagent MDP with edge-based decomposition [18] of the global $Q(\mathbf{s}, \mathbf{a})$ function, along with parameter sharing (as in [2]); and use Sparse Q-learning [17] with a neural network to approximate the local $Q_{ij}(s_{ij}, a_{ij})$ function. The envisioned approach could appropriately employ DQN [23], and then examine its various extensions—such as Double DQN [28], DQN with Dueling Architectures [32], Prioritized Experience Replay [27]—along with combinations of various recent techniques that aim to further improve the performance of Deep RL [1, 12]. Notice that one would also need to tackle the non-stationarity introduced in a multiagent learning scheme [11].

Another area of interest is the incorporation of continuous action spaces in max-plus. The current approach is limited to discrete actions, i.e., vehicles apply an acceleration instantly, therefore resulting in abrupt changes in *jerk* [25], which is deemed to be a metric for passenger convenience. To the best of our knowledge, no work to date has extended max-plus to continuous action spaces.

Moreover, our approach could benefit from *fallback mechanisms* to promote safety, and the careful incorporation of constraints relating to CAV decision making time. Such parameters could be set based on the current (or projected) capabilities of CAVs regarding computation, communication speed, and expected communication errors [22].

Furthermore, irrespective of the impending extensions or changes, one important aspect that needs to be examined is the consideration of a mixed-traffic environment, where the coordination strategy is applied to a subset $N \subseteq N_{total}$ of the existing vehicles. It would be important to formalize how the coordination strategy can be adapted in a mixed-traffic scenario with vehicles that do not participate in the coordination strategy. This is not a straightforward exercise, as non-participating vehicles' movement strategies could vary from having a constant speed, to act according to individual objectives, or even to adopting a different collaborative strategy. Likewise, our assumptions regarding these vehicles may also range from possessing specific, "hardwired", safety-related beliefs regarding the behaviour of other vehicles, to employing explicit opponent modeling techniques, i.e., attempting to learn the behaviour by observation, in order to produce joint actions based on these assumptions/predictions.

Finally, a natural model to include in mixed-traffic scenarios would be a human-driven one [4]. This is challenging, since the building of models based on human driving behaviour is available only for lane-based traffic. Indeed, there is currently no notion of how a human driver would operate and react in a lane-free traffic environment. This issue potentially calls for synergies with virtual reality researchers, to allow the study of human behaviour in lane-free traffic scenarios. It is also conceivable that, in lane-free environments with mixed traffic, human drivers are asked to observe certain rules—e.g., driving only along the outer road boundary, to reduce risks and render the human driving task easier.

6 CONCLUSIONS

In this paper we put forward a novel application of the max-plus algorithm for collaborative autonomous driving, with vehicles operating in a newly established lane-free traffic environment in which they are no longer tied to specified lane positions. This gives rise to a much more flexible environment, allowing for higher utilization of the available road capacity, but also introduces additional complexity. We tackle the coordination problem arising for the vehicles in such scenarios with the max-plus algorithm, and use artificial potential fields to construct local utility functions and utilize these for collision avoidance and vehicle speed optimization. Our experimental evaluation demonstrates the coordination capabilities for our setting and lays the foundations for interesting extensions in this important emerging real-world MAS problem domain.

ACKNOWLEDGMENTS

The research leading to these results has received funding from the European Research Council under the European Union's Horizon 2020 Research and Innovation programme/ ERC Grant Agreement n. [833915], project TrafficFluid.

REFERENCES

[1] Adrià Puigdomènech Badia, Bilal Piot, Steven Kapturowski, Pablo Sprechmann, Alex Vitvitskyi, Daniel Guo, and Charles Blundell. 2020. Agent57: Outperforming the Atari Human Benchmark. arXiv:2003.13350 [cs.LG]

[2] Wendelin Böhmer, Vitaly Kurin, and Shimon Whiteson. 2019. Deep Coordination Graphs. arXiv:1910.00091 [cs.LG]

[3] Farid Bounini, Denis Gingras, Herve Pollart, and Dominique Gruyer. 2017. Modified artificial potential field method for online path planning applications. In *2017 IEEE Intelligent Vehicles Symposium (IV)*. 180–185.

[4] Kyle Brown, Katherine Driggs-Campbell, and Mykel J. Kochenderfer. 2020. A Taxonomy and Review of Algorithms for Modeling and Predicting Human Driver Behavior. *arXiv e-prints*, Article arXiv:2006.08832 (June 2020), arXiv:2006.08832 pages. arXiv:2006.08832 [eess.SY]

[5] Jacopo Castellini, Frans A. Oliehoek, Rahul Savani, and Shimon Whiteson. 2019. The Representational Capacity of Action-Value Networks for Multi-Agent Reinforcement Learning. In *Proceedings of the 18th International Conference on Autonomous Agents and MultiAgent Systems* (Montreal QC, Canada) *(AAMAS '19)*. International Foundation for Autonomous Agents and Multiagent Systems, Richland, SC, 1862–1864.

[6] Xuan Di and Rongye Shi. 2020. A Survey on Autonomous Vehicle Control in the Era of Mixed-Autonomy: From Physics-Based to AI-Guided Driving Policy Learning. arXiv:2007.05156 [cs.AI]

[7] Kurt Dresner and Peter Stone. 2008. A Multiagent Approach to Autonomous Intersection Management. *J. Artif. Int. Res.* 31, 1 (March 2008), 591–656.

[8] David Elliott, Walter Keen, and Lei Miao. 2019. Recent advances in connected and automated vehicles. *Journal of Traffic and Transportation Engineering (English Edition)* 6, 2 (2019), 109 – 131. https://doi.org/10.1016/j.jtte.2018.09.005

[9] Carlos Guestrin, Daphne Koller, and Ronald Parr. 2002. Multiagent Planning with Factored MDPs. In *Advances in Neural Information Processing Systems 14*, T. G. Dietterich, S. Becker, and Z. Ghahramani (Eds.). MIT Press, 1523–1530. http://papers.nips.cc/paper/1941-multiagent-planning-with-factored-mdps.pdf

[10] P. A. Hancock, Illah Nourbakhsh, and Jack Stewart. 2019. On the future of transportation in an era of automated and autonomous vehicles. *Proceedings of the National Academy of Sciences* 116, 16 (2019), 7684–7691. https://doi.org/10.1073/pnas.1805770115 arXiv:https://www.pnas.org/content/116/16/7684.full.pdf

[11] Pablo Hernandez-Leal, Michael Kaisers, Tim Baarslag, and Enrique Munoz de Cote. 2019. A Survey of Learning in Multiagent Environments: Dealing with Non-Stationarity. arXiv:1707.09183 [cs.MA]

[12] Matteo Hessel, Joseph Modayil, Hado Van Hasselt, Tom Schaul, Georg Ostrovski, Will Dabney, Dan Horgan, Bilal Piot, Mohammad Azar, and David Silver. 2017. Rainbow: Combining improvements in deep reinforcement learning. *arXiv preprint arXiv:1710.02298* (2017).

[13] Hu Hongyu, Zhang Chi, Sheng Yuhuan, Zhou Bin, and Gao Fei. 2018. An Improved Artificial Potential Field Model Considering Vehicle Velocity for Autonomous Driving. *IFAC-PapersOnLine* 51, 31 (2018), 863 – 867. https://doi.org/10.1016/j.ifacol.2018.10.095 5th IFAC Conference on Engine and Powertrain Control, Simulation and Modeling E-COSM 2018.

[14] Serge Hoogendoorn and Victor Knoop. 2013. Traffic flow theory and modelling. *The transport system and transport policy: an introduction* (2013), 125–159.

[15] Sheng Jin, Zhi-yi Huang, Peng-fei Tao, and Dian-hai Wang. 2011. Car-following theory of steady-state traffic flow using time-to-collision. *Journal of Zhejiang University-SCIENCE A* 12, 8 (2011), 645–654.

[16] Oussama Khatib. 1990. *Real-Time Obstacle Avoidance for Manipulators and Mobile Robots.* Springer New York, New York, NY, 396–404. https://doi.org/10.1007/978-1-4613-8997-2_29

[17] Jelle R. Kok and Nikos Vlassis. 2004. Sparse Cooperative Q-Learning. In *Proceedings of the Twenty-First International Conference on Machine Learning* (Banff, Alberta, Canada) *(ICML '04)*. Association for Computing Machinery, New York, NY, USA, 61. https://doi.org/10.1145/1015330.1015410

[18] Jelle R Kok and Nikos Vlassis. 2006. Collaborative multiagent reinforcement learning by payoff propagation. *Journal of Machine Learning Research* 7, Sep (2006), 1789–1828.

[19] Jelle R. Kok and Nikos Vlassis. 2006. Using the Max-Plus Algorithm for Multiagent Decision Making in Coordination Graphs. In *RoboCup 2005: Robot Soccer World Cup IX*, Ansgar Bredenfeld, Adam Jacoff, Itsuki Noda, and Yasutake Takahashi (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 1–12.

[20] Lior Kuyer, Shimon Whiteson, Bram Bakker, and Nikos Vlassis. 2008. Multiagent Reinforcement Learning for Urban Traffic Control Using Coordination Graphs. In *Machine Learning and Knowledge Discovery in Databases*, Walter Daelemans, Bart Goethals, and Katharina Morik (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 656–671.

[21] Pablo Alvarez Lopez, Michael Behrisch, Laura Bieker-Walz, Jakob Erdmann, Yun-Pang Flötteröd, Robert Hilbrich, Leonhard Lücken, Johannes Rummel, Peter Wagner, and Evamarie Wießner. 2018. Microscopic Traffic Simulation using SUMO, In The 21st IEEE International Conference on Intelligent Transportation Systems. *IEEE Intelligent Transportation Systems Conference (ITSC)*. https://elib.dlr.de/124092/

[22] F. M. Malik, H. A. Khattak, A. Almogren, O. Bouachir, I. U. Din, and A. Altameem. 2020. Performance Evaluation of Data Dissemination Protocols for Connected Autonomous Vehicles. *IEEE Access* 8 (2020), 126896–126906. https://doi.org/10.1109/ACCESS.2020.3006040

[23] Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Alex Graves, Ioannis Antonoglou, Daan Wierstra, and Martin Riedmiller. 2013. Playing atari with deep reinforcement learning. *arXiv preprint arXiv:1312.5602* (2013).

[24] S. Notley, N. Bourne, and N. Taylor. 2008. Speed, flow and density of motorway traffic. In *TRL Insight Report INS003*. https://trl.co.uk/reports/INS003

[25] Ioannis A. Ntousakis, Ioannis K. Nikolos, and Markos Papageorgiou. 2016. Optimal vehicle trajectory planning in the context of cooperative merging on highways. *Transportation Research Part C: Emerging Technologies* 71 (2016), 464 – 488. https://doi.org/10.1016/j.trc.2016.08.007

[26] Markos Papageorgiou, Kyriakos-Simon Mountakis, Iasson Karafyllis, and Ioannis Papamichail. 2019. Lane-free Artificial-Fluid Concept for Vehicular Traffic. *arXiv preprint arXiv:1905.11642* (2019).

[27] Tom Schaul, John Quan, Ioannis Antonoglou, and David Silver. 2015. Prioritized experience replay. *arXiv preprint arXiv:1511.05952* (2015).

[28] Pedro A Tsividis, Thomas Pouncy, Jacqueline L Xu, Joshua B Tenenbaum, and Samuel J Gershman. 2017. Human learning in Atari. (2017).

[29] Panagiotis Typaldos, Kyriakos Mountakis, Ioannis Papamichail, and Markos Papageorgiou. 2019. Optimization-based path planning for automated vehicles. In *9th Intern. Congress on Transportation Research*. Athens, Greece.

[30] Elise Van der Pol and Frans A Oliehoek. 2016. Coordinated deep reinforcement learners for traffic light control. In *NIPS'16 Workshop on Learning, Inference and Control of Multi-Agent Systems*.

[31] Z. Wang, Y. Bian, S. E. Shladover, G. Wu, S. E. Li, and M. J. Barth. 2020. A Survey on Cooperative Longitudinal Motion Control of Multiple Connected and Automated Vehicles. *IEEE Intelligent Transportation Systems Magazine* 12, 1 (2020), 4–24. https://doi.org/10.1109/MITS.2019.2953562

[32] Ziyu Wang, Tom Schaul, Matteo Hessel, Hado Hasselt, Marc Lanctot, and Nando Freitas. 2016. Dueling network architectures for deep reinforcement learning. In *International conference on machine learning*. 1995–2003.

[33] Cathy Wu, Aboudy Kreidieh, Kanaad Parvate, Eugene Vinitsky, and Alexandre M Bayen. 2017. Flow: A Modular Learning Framework for Autonomy in Traffic. arXiv:1710.05465 [cs.AI]

[34] Chao Yu, Xin Wang, Xin Xu, Minjie Zhang, Hongwei Ge, Jiankang Ren, Liang Sun, Bingcai Chen, and Guozhen Tan. 2019. Distributed Multiagent Coordinated Learning for Autonomous Driving in Highways Based on Dynamic Coordination Graphs. *IEEE Transactions on Intelligent Transportation Systems* (2019), 1–14. https://doi.org/10.1109/TITS.2019.2893683