Feedback-Based Ramp Metering and Lane-Changing Control With Connected and Automated Vehicles

Farzam Tajdari, Claudio Roncoli, and Markos Papageorgiou, *Life Fellow, IEEE*

***Abstract*— Aiming at operating effectively future traffic systems, we propose here a novel methodology for integrated lane-changing and ramp metering control that exploits the presence of connected vehicles. In particular, we assume that a percentage of vehicles can receive and implement specific control tasks (e.g., lane-changing commands), while ramp metering is available via an infrastructure-based system or enabled by connected vehicles. The proposed approach is designed to robustly maximise the throughput at motorway bottlenecks employing a feedback controller, formulated as a Linear Quadratic Integral regulator, which is based on a simplified linear time invariant traffic flow model. We also present an extremum seeking algorithm to compute the optimal set-points used in the feedback controller, employing only the measurement of a cost that is representative of the achieved traffic conditions. The method is evaluated via simulation experiments, performed on a first-order, multi-lane, macroscopic traffic flow model, also featuring the capacity drop phenomenon, which allows to demonstrate the effectiveness of the developed methodology and to highlight the improvement in terms of the generated congestion.**

***Index Terms*— Traffic control, connected and automated vehicles, lane-changing control, ramp metering.**

I. Introduction

IN THE last decades, a significant and increasing interdisciplinary effort by the automotive industry, as well as by numerous research institutions around the world, has been devoted to planning, developing, testing, and deploying new technologies that are expected to revolutionise the features and capabilities of individual vehicles in the future [1]. Among the wide range of available systems, few may actually have a direct impact on traffic flow, while the majority of them aims at primarily improving safety or driver's convenience [2]. We focus here on the integration of a promising new feature that can be exploited for traffic management in the presence of connected and automated vehicles, namely, lane-changing control, together with a conventional well-established traffic control measure, namely, ramp metering [3].

Manuscript received June 27, 2019; revised February 3, 2020 and May 19, 2020; accepted August 13, 2020. This work was supported by the Academy of Finland, via Profi4 and the FINEST Twins Center of Excellence (H2020) under Grant 856602. The Associate Editor for this article was V. Punzo. *(Corresponding author: Farzam Tajdari.)*

Farzam Tajdari and Claudio Roncoli are with the Department of Built Environment, School of Engineering, Aalto University, 02150 Espoo, Finland (e-mail: farzam.tajdari@aalto.fi; claudio.roncoli@aalto.fi).

Markos Papageorgiou is with the School of Production Engineering and Management, Technical University of Crete, 73100 Chania, Greece (e-mail: markos@dssl.tuc.gr).

Digital Object Identifier 10.1109/TITS.2020.3018873

The problem of modelling the distribution of vehicles over the lanes in the case of ordinary traffic has been addressed in research that shows that lane distribution is affected by some characteristics of the network layout (e.g., the total number of lanes) [4]–[6]. However, this choice is also behavioural since every single driver may autonomously decide to stay in a slower lane accepting the lower speed, stay in the slower lane and overtake when necessary (for lower densities), or travel constantly in a faster lane (in higher densities). In addition, particularly at bottleneck locations (e.g., lane drops and on-ramp merges), human drivers usually perform suboptimal lane changes on the basis of erroneous perceptions, which may trigger congestion and, thus, deteriorate the overall travel time [7]. Finally, some of the mentioned empirical investigations indicate that in conventional traffic, capacity flow is not reached simultaneously at all lanes, a feature that reduces the potentially achievable cross-lane capacity. It is, therefore, envisioned that if a sufficient percentage of vehicles have vehicle-to-infrastructure communication capabilities and appropriate lane-changing automatic controllers or advisory systems, the overall throughput at the bottleneck location may be improved by the execution of specific lane-changing commands decided by a central decision maker.

In the context of automated and connected vehicles or Automated Highway Systems (AHS), only a limited number of works have considered to optimize lane distribution ([8]–[14]). A number of other works addressed specifically the problem of deciding on efficient vehicle lane-paths for a motorway under fully automated (AHS) or semi-automated driving (e.g., [8], [9]). However, to tackle the problem complexity, a number of simplifying assumptions were typically made. Other works focused on computing lane-changing commands at vehicle level, which are obtained by sensing the surrounding environment [13], [16]–[21]. The intrinsic complexity of some of these approaches may be an impediment for real-time application while also considering additional options and features offered by emerging systems. An optimal feedback control strategy for lane-changing control is formulated in [12] as a linear quadratic regulator (LQR) and is extended in [14] to allow for different policies, i.e., achieving different traffic density distribution for the various lanes at the bottleneck area. However, in both works, it is assumed that another controller ensures that the overall traffic flow entering the area where lane-changing control is applied, does not significantly exceed the bottleneck capacity; in addition,

robustness of the controller to unknown disturbances is not explicitly considered.

On the other hand, the real-time control of motorway traffic via conventional means (e.g., traffic signal based ramp metering, road-side variable message signs for route guidance and variable speed limits) has been addressed in research studies (see, e.g., [3]) and practical implementations for decades [22], showing consistent benefits when they are properly implemented in traffic systems [23]–[25]. Ramp-metering and lane-changing control have been studied in [7] to investigate the effect of forced lane-changes at the bottleneck on capacity drop.

The method proposed in this article aims at bridging conventional traffic control, i.e., ramp-metering, with lane-changing control enabled by vehicle automation via a novel feedback-based integrated control strategy. From an operational perspective, we assume that equipped vehicles have the capability of bidirectional communication with the infrastructure, i.e., Vehicle-to-Infrastructure (V2I) communication, so that appropriate control actions are decided in a centralised manner by a Traffic Management Center (TMC) and are dispatched to specific vehicles for implementation (see, e.g., [10], [26]). We envision that, if lane-changing control capabilities are implemented in conjunction with more traditional traffic management strategies, such as ramp metering, in an integrated fashion, the resulting effectiveness, in terms of traffic performance, would be further increased, while allowing for rigorous investigations on the generalisability and robustness of the methodology for different networks topologies and under different disturbances. In particular, this article focuses on the development and testing of a novel methodology for integrated lane-changing and ramp metering control, considering and exploiting the presence of partly automated and connected vehicles. A preliminary version of this work is included in [27], which is extended here with a more rigorous formulation; a thorough discussion on controller stability properties for the proposed control law; a set of numerical investigations on robustness to parameter choices; and additional numerical experiments, including a scenario considering a state-of-the-art conventional control strategy and a scenario considering additional disturbances.

The paper is structured as follows. Section II describes the proposed traffic flow model. Section III includes the controller design, while Section IV presents a discrete-time extremum seeking scheme to achieve optimal set points value. Section V introduces the experiment setup, whereas, in Section VI, the obtained simulation results are presented and compared with a reference no-control case. Section VII concludes the paper, highlighting our main results and opening research directions for future work.

II. Linear Multi-Lane Traffic Flow Model

We consider a multi-lane motorway that is subdivided into N segments, indexed by $i = 0, \dots, N$, each of length L_i, while each segment i is composed of lanes, indexed by $j = m_i, \dots, M_i$, where m_i and M_i are the minimum and maximum indexes of lanes for segment i. We denote each element of the resulting grid (see Fig. 1) as "cell", which is indexed by (i, j). In order to account for any possible network topology, including lane-drops and lane-additions, both on the right and on the left sides of the motorway, we assume that index $j = 0$ corresponds to the segment(s) including the right-most lane. For example, looking at the hypothetical motorway stretch depicted in Fig. 1, $m_0 = 0$ and $M_0 = 4$, while $m_3 = 1$ and $M_3 = 3$. For controller design purposes, we assume that the last segment in our network, indexed by N, is the main bottleneck. The resulting total number of cells is $H = \sum_{r=0}^{N} (M_r - m_r + 1)$ and the number of cells at the bottleneck area is $S = M_N - m_N + 1$.

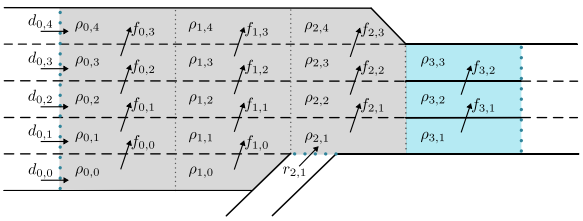

Fig. 1. A hypothetical motorway stretch.

The model is formulated in discrete time, considering the discrete time step T, indexed by $k = 0, 1, \dots, K$, where the time is $t = kT$.

Each motorway cell (i, j) is characterised by the traffic density $\rho_{i,j}(k)$, defined as the number of vehicles present within the cell at time instant k divided by L_i. Density dynamically evolves according to the following conservation law equation, (see, e.g. [10]),

$$\rho_{i,j}(k+1) = \rho_{i,j}(k) + \frac{T}{L_i}\big[q_{i-1,j}(k) - q_{i,j}(k)\big] + \frac{T}{L_i}\big[f_{i,j-1}(k) - f_{i,j}(k)\big] + \frac{T}{L_i} d_{i,j}(k) + \frac{T}{L_i} r_{i,j}(k), \tag{1}$$

where $q_{i,j}(k)$ is the longitudinal flow leaving cell (i, j) and entering cell $(i+1, j)$ during time interval $(k, k+1]$; $f_{i,j}(k)$ is the net lateral flow moving from cell (i, j) to cell $(i, j+1)$ during time interval $(k, k+1]$; and $d_{i,j}(k)$ is any external (uncontrolled) flow entering the network in cell (i, j), either from upstream of the considered stretch or from an on-ramp, during time interval $(k, k+1]$. Since a ramp, located within the considered stretch, is assumed to be controlled, we introduce $r_{i,j}(k)$ as the flow allowed to enter the network from the ramp located in (i, j) during time interval $(k, k+1]$ (e.g., $r_{2,1}$ in Fig. 1). Depending on the network topology, some terms of (1) may not be present. In particular, the inflow $q_{i-1,j}(k)$ does not exist for the first segment of the network, the outflow $q_{i,j}(k)$ does not exist for the last segment before a lane-drop, while lateral flow terms $f_{i,j}(k)$ exist only for $m_i \le j < M_i$. Following previous considerations, the total number of (controllable) lateral flow terms is $F = H - N$.

Let us now consider the well-known relation

$$q_{i,j}(k) = \rho_{i,j}(k)\, v_{i,j}(k). \tag{2}$$

Since we are designing a controller to operate in (and in fact maintain) congestion-free traffic conditions, we assume

that the speed in all cells is approximately constant (e.g., the critical speed) $v_{i,j}(k) \equiv \bar{v}_{i,j}, \forall i, j, k$. It is worth highlighting that this assumption is made only for controller design and analysis, while in the numerical experiments presented in Section VI, we use a nonlinear model that produces non-constant speeds, which are also different from the constant speed value set in the controller, without any significant negative impact on the results. Despite this may appear a strong assumption, we will see in simulation (Section VI) that the controller achieves good performance also when speed varies over time, thanks to the inherent robustness property of the designed feedback controller.

By replacing (2) into (1), we write the resulting system in the form of a Linear Time Invariant (LTI) system

$$\bar{x}(k+1) = \bar{A}\bar{x}(k) + \bar{B}u(k) + d(k), \tag{3}$$

where (time index k is omitted to simplify notation)

$$\bar{x} = \left[\rho_{0,m_0} \cdots \rho_{0,M_0} \;\; \rho_{1,m_1} \cdots \rho_{N,M_N}\right]^T \in \mathbb{R}^H, \tag{4}$$

$$d = \left[\frac{T}{L_0}d_{0,m_0} \cdots \frac{T}{L_0}d_{0,M_0} \;\; \frac{T}{L_0}d_{1,m_1} \cdots \frac{T}{L_0}d_{N,M_N}\right]^T \in \mathbb{R}^H, \tag{5}$$

$$u = \left[f_{0,m_0} \cdots f_{0,M_0} \;\; f_{1,m_1} \cdots f_{N,M_N}, r_{i,j}\right]^T \in \mathbb{R}^{F+1}. \tag{6}$$

Variables u and d are the controlled and uncontrolled inputs, respectively; u includes all the lateral flows $f_{i,j}$ and the ramp flow $r_{i,j}$ that are assumed controllable; while d includes the external flows that are not in u. Matrix $\bar{A} \in \mathbb{R}^{H\times H}$, composed of elements $a_{\bar{p},\bar{s}}$, reflects the connections between pairs of subsequent cells via a longitudinal flow; while matrix $\bar{B}$, composed of elements $b_{\bar{p},\bar{s}}$, reflects the interconnections among cells via their entering or leaving lateral flows. The matrices for the mentioned system can be described as

$$a_{\bar{p},\bar{s}} = \begin{cases} 1, & \text{if } \bar{p} = \bar{s} \text{ and } \left(j < m_{i+1} \text{ or } \; j > M_{i+1}\right) \\ 1 - \dfrac{T}{L_i}\bar{v}_{i,j}, & \text{if } \bar{p} = \bar{s} \text{ and } \left(i = N \text{ or } m_{i+1} \le j \le M_{i+1}\right) \\ \dfrac{T}{L_i}\bar{v}_{i,j}, & \text{if } \bar{p} > H_0 \text{ and } \bar{s} = \bar{p} - M_{i-1} + m_i - 1 \\ 0, & \text{otherwise} \end{cases} \tag{7}$$

$$b_{\bar{p},\bar{s}} = \begin{cases} \dfrac{T}{L_i}, & \text{if } j > m_i \text{ and } \bar{s} = \bar{p} - i \\ -\dfrac{T}{L_i}, & \text{if } j < M_i \text{ and } \bar{s} = \bar{p} - i + 1 \\ \dfrac{T}{L_i}, & \text{if } j = \bar{j} \text{ and } i = \bar{i} \\ 0, & \text{otherwise} \end{cases} \tag{8}$$

where $\left(\bar{i}, \bar{j}\right)$ defines the location of on-ramp flow.

Finally, the CFL condition [28]

$$\frac{T}{L_i}\bar{v}_{i,j} < 1, \forall i, j \tag{9}$$

should be respected for a stable discrete time and discrete space traffic flow model.

III. Lateral Flow and Ramp Metering Controller

A. Controller Formulation

We employ here the linear system described in Section II to formulate an optimal control problem, whose solution leads to a MIMO (multi-input multi-output) feedback controller. Specifically, the controller manipulates the lateral flows, as well as the flow entering from an on-ramp located upstream of the bottleneck, with the overall goal of avoiding the creation of congestion and maximising the bottleneck throughput. More specifically, lane-changing control aims at increasing the bottleneck capacity encountered with human-driver lane choices via more efficient lane distribution; while ramp metering aims at avoiding the creation of congestion and maintaining the bottleneck throughput at the increased capacity level.

Previous works (e.g., [29], [30]) have demonstrated that the real flow capacity (at which traffic flow speed is observed to break down due to overload) in a merge area may vary quite substantially from day to day even under similar environmental conditions, therefore, any control strategy attempting to achieve a pre-specified capacity flow value may either lead to overload and congestion (on days where the real capacity happens to be lower than its pre-specified target value) or to under-utilisation of the infrastructure (on days where the real capacity happens to be higher than its pre-specified target value). On the other hand, the critical density ($\rho^{\text{cr}}_{S\times 1}$), at which capacity flow occurs, appears to be more stable [31], and it is therefore preferable targeting a density set-point (i.e., the critical density) at the bottleneck location. Thus, to maximise the bottleneck throughput, densities at the bottleneck cells should be maintained around their critical values (ρ^{cr}), which are here supposed to be known.

In order to design our controller, we first assume the availability of nominal (desired) steady-state values $\bar{x}_{\text{d}}$ and u_{d}, for density states and input flows, respectively, which satisfy

$$\left(I - \bar{A}\right)\bar{x}_{\text{d}} - \bar{B}u_{\text{d}} - d_{\text{d}} = 0, \tag{10}$$

where d_{d} is the constant nominal uncontrolled demand. Variables $\bar{x}_{\text{d}}$ and u_{d} describe a desired behaviour of the controlled system; in our case, following the previous rationale, we define values in $\bar{x}_{\text{d}}$ corresponding to the cells at the bottleneck equal to ρ^{cr}. The need for specifying remaining desired steady-state state and control variables will be discussed in Section III-C.

Introducing the notation $\Delta\omega(k) = \omega(k) - \omega_{\text{d}}$, where ω replaces $\bar{x}$, u, and $\bar{d}$, we reformulate the system in terms of error dynamics, obtaining

$$\Delta\bar{x}(k+1) = \bar{A}\Delta\bar{x}(k) + \bar{B}\Delta u(k) + \Delta d(k). \tag{11}$$

In order to design a feedback control law without feed-forward terms, i.e., a control law that reacts to the impact of the disturbances on the controlled process rather than to disturbance forecasts, we assume $\Delta d(k) = 0$. In addition, since the system may be affected by disturbances, (e.g., upstream mainstream demand or uncontrolled lateral flows), in order to avoid offset at the stationary state, we employ an integral controller to reject constant disturbances [32], thus removing the need for measuring the external inflows.

Stressing on the usefulness of employing an integral controller, it is also worth pointing out that movements of conventional vehicles, such as, e.g., changing lanes or exiting at an off-ramp, can be essentially viewed as disturbances, e.g., as additive noise, affecting the traffic densities at each cells. In such case, due to the features of integral control, such disturbances are automatically rejected by our controller, without need of any additional measurement. Thus, we proceed by augmenting the system (11) with S (i.e., as many as bottleneck lanes) integral states, denoted as $z \in \mathbb{R}^S$, characterised by dynamics

$$z(k+1) = z(k) + \bar{C}\,\Delta\bar{x}(k), \tag{12}$$

where $\bar{C} = \begin{bmatrix} 0_{S\times(H-S)} & I_{S\times S} \end{bmatrix}$ extracts the elements of $\Delta\bar{x}$ corresponding to the bottleneck cells. This implies that, in order to compute the dynamics of z, only the nominal state values corresponding to the cells at the bottleneck, i.e., ρ^{cr}, are needed. These modifications lead to the following augmented system:

$$\Delta x(k+1) = A\,\Delta x(k) + B\,\Delta u(k), \tag{13}$$

where

$$\Delta x = \begin{bmatrix} \Delta\bar{x} \\ z \end{bmatrix}, A = \begin{bmatrix} \bar{A} & 0_{H\times S} \\ \bar{C} & I_{S\times S} \end{bmatrix}, \quad B = \begin{bmatrix} \bar{B} \\ 0_{S\times(F+1)} \end{bmatrix}.$$

Finally, we define the following quadratic cost function, over an infinite time horizon, which accounts for minimisation of integral state and control input errors:

$$\min J = \sum_{k=0}^{\infty} \Big[\Delta x(k)^T C^T \; Q \; C\Delta x(k) + \Delta u(k)^T \; R \; \Delta u(k) \Big], \tag{14}$$

where

$$Q = w_Q \; I_{S\times S}, \quad R = \begin{bmatrix} w_{R_1} \; I_{F\times F} & 0_{F\times 1} \\ 0_{1\times F} & w_{R_2} \end{bmatrix}, \tag{15}$$

$$C = \begin{bmatrix} 0_{S\times H} \; I_{S\times S} \end{bmatrix}. \tag{16}$$

Matrices Q and R are weighting matrices associated to the magnitude of the integral and control errors, respectively, defined by parameters $w_Q > 0$, $w_{R_1} > 0$, and $w_{R_2} > 0$, where w_{R_1} penalises lateral flow errors and w_{R_2} penalises ramp-metered flow error.

The resulting optimal control problem (14), (13) can be solved through a Linear Quadratic Regulator (LQR), which provides a stabilising feedback gain under the assumptions that the original system is, at least, stabilisable and detectable (see Chapter 2 of [33]).

B. Stabilisability and Detectability

We investigate stabilisability and detectability of system (13) by employing the Hautus-test [34].

We start by observing that the system matrix A is, by construction, lower triangular, implying that its eigenvalues λ are equal to the elements in the main diagonal. Since $\bar{v}$ is always positive and assuming (9) is respected, it results that all the modes related to cells for which another downstream cell exists are always stable ($|\lambda| < 1$), while the modes related to cells without any other cell downstream (i.e., at a lane-drop), as well as the integral state modes, are marginally stable ($\lambda = 1$).

In order to guarantee that the pair (A, B) is stabilisable, B must have more linearly independent columns than the number of non-stable modes ($|\lambda| \geq 1$) [34]. Depending on the network topology, we distinguish two cases: a) there are no lane-drops, thus non-stable modes are solely related to integral states; and b) there are lane-drops, thus non-stable modes are both related to integral states and to cells upstream of lane-drops.

In the first case (a), there are S marginally stable modes, corresponding to integral states, which, in order to satisfy the Hautus-test condition, require at least S controlled inputs. Since, in our system configuration, there are at least $S-1$ controlled lateral flows (e.g., the lateral flows at the bottleneck location), and a controlled on-ramp input, the stabilisability condition is satisfied. In the presence of lane-drops (b), in addition to the previous, stabilisability is guaranteed if there is at least one additional controlled variable that directly affects the dropping lane. Since we assume in our model that lateral flows are controlled for each lane of the network, also this condition is satisfied for any network topology following the structure shown in Section II.

We turn now our attention to the detectability of the pair $(A, C^T QC)$; since $Q > 0$, this is equivalent to investigating the detectability of the pair (A, C) [35]. In our case, the Hautus test condition is verified in case C has at least a non-zero element in each column corresponding to a marginally stable mode ($\lambda = 1$), which, in case there are no lane-drops, implies controlling all integral states, which is satisfied according to (15), (16). On the other hand, if the network features lane-drops, it is necessary to include in the cost function also terms to control cells that does not have any other cell downstream; this can be done by either defining an opportune set-point for the density in the lane-drop cell, or, as proposed in [12], by placing additional dummy cells immediately downstream of each lane-drop, imposing, with an appropriate high penalty weight, to have a density equal to zero. Note that, in all the described cases, the system is also observable.

C. Controller Design and Anti-Windup

The solution to the proposed LQR problem is the linear feedback control law

$$\Delta u(k) = -K\,\Delta x(k), \tag{17}$$

where

$$K = \left(R + B^T P B\right)^{-1} B^T P A, \tag{18}$$

$$P = C^T QC + A^T P A - A^T P B \left(R + B^T P B\right)^{-1}. \tag{19}$$

The optimal gain (18) and the Algebraic Riccati Equation (19) can be found in classic Optimal Control books (see, e.g., [36]). For practical implementation, gain K is appropriately split as

$$K = \begin{bmatrix} K_{\mathrm{P}} \;\vdots\; K_{\mathrm{I}} \end{bmatrix} \tag{20}$$

which allows to rewrite the control law as

$$\Delta u(k) = -K_{\mathrm{P}}\;\Delta\bar{x}(k) - K_{\mathrm{I}}\;z(k), \tag{21}$$

and, consequently,

$$u(k) = u(k-1) - K_{\mathrm{P}}\left[\bar{x}(k) - \bar{x}(k-1)\right] - K_{\mathrm{I}}\left[z(k) - z(k-1)\right]. \tag{22}$$

In practice, it may not be always possible to achieve the desired density set-point at the bottleneck (e.g., due to input saturation), therefore it is necessary to include an anti-windup scheme within our controller. We employ the scheme proposed in [37] (see also [38], [39]), which, in our case, modifies the integral part of the dynamic controller (12) as

$$z(k+1) = (I + \Lambda K_{\mathrm{I}})z(k) + \left(\bar{C} + \Lambda K_{\mathrm{P}}\right)\Delta\bar{x}(k) + \Lambda \mathrm{sat}(\Delta u(k)), \tag{23}$$

where $\Lambda \in \mathbb{R}^{S\times(F+1)}$ and $(I + \Lambda K_{\mathrm{I}}) \in \mathbb{R}^{S\times S}$. Since the saturation is implemented, in practice, on the actual system input $u(k)$, we may replace $\mathrm{sat}\left(\Delta u(k)\right)$ with $\mathrm{sat}\left(u(k)\right) + u_{\mathrm{d}}$, where the saturation operator is defined as follows

$$\mathrm{sat}(u_o) = \begin{cases} u_o^{\min}, & \text{if } u_o < u_o^{\min} \\ u_o^{\max}, & \text{if } u_o > u_o^{\max} \\ u_o, & \text{otherwise,} \end{cases} \tag{24}$$

where o is the index of controlled inputs inside vector u, and $u_o^{\min}$ and $u_o^{\max}$ are the lower and upper bound for the input u_o.

The final formulation of the dynamic regulator is therefore (22), (23), which is very effective for practical application since the computation of the feedback gains K_{P}, K_{I} may be effectuated offline, i.e., solving (18) and (19), while online calculations are limited to computing (22) and (23).

Note that, while the input $u(k)$ is not saturated, the dynamics (23) reduces to (12), implying that only steady-state desired values for bottleneck cells are required, i.e., critical densities. On the other hand, in case the input $u(k)$ is saturated, the computation of the integral state dynamics (23) requires desired values for all state and input variables. Such nominal values can be obtained from (10), by imposing values in $\bar{x}_{\mathrm{d}}$ corresponding to the cells at the bottleneck equal to ρ^{cr} and an arbitrary nominal demand d_{d}, deriving the remaining components via, e.g., a least-squares-based approach. Note that, our numerical experiments have shown that the choice of different nominal values has virtually no impact on the controller performance, which can be explained by the fact that the input is saturated mostly in non-critical situations, e.g., for the ramp input case, when the on-ramp demand is very low and ramp metering is, in fact, not needed.

D. Stability of the Closed-Loop System With Anti-Windup

A necessary condition for stability of the closed-loop system is that matrix Λ must be opportunely chosen so that $I + \Lambda K_I$ has stable eigenvalues $\bar{\lambda}$ [39]; this can be achieved, for example, via classical pole placement or via specific algorithm (see, e.g., [39]). Note that, when inputs are not saturated, system stability is guaranteed by the conditions stated in Section III-B, as the pair (A, B), and (A, C) are the same in both cases; on the other hand, when inputs are saturated, Λ must be properly designed as it may affect the stability of (23).

In order to investigate stability of the closed-loop system in case the controller (21), (23) is implemented and the input is saturated, we rewrite (13) as

$$\Delta x(k+1) = \bar{\bar{A}}\Delta x(k) + (B + R_{\mathrm{aw}}\Lambda)\mathrm{sat}(\Delta u(k)), \tag{25}$$

where

$$R_{\mathrm{aw}} = \begin{bmatrix} 0_{H\times H} \\ \hdashline I_{S\times S} \end{bmatrix}, \bar{\bar{A}} = \left[\begin{array}{c:c} \bar{A} & 0_{H\times S} \\ \hdashline \left(\bar{C} + \Lambda K_P\right) & (I + \Lambda K_I) \end{array}\right] \tag{26}$$

According to [40], if there exist a symmetric positive definite matrix $W_{\mathrm{aw}} \in \mathbb{R}^{(H+S)\times(H+S)}$, a diagonal positive definite matrix $S_{\mathrm{aw}} \in \mathbb{R}^{(F+1)\times(F+1)}$, and a matrix $Z_{\mathrm{aw}} \in \mathbb{R}^{S\times(F+1)}$, satisfying:

$$\Theta = \begin{bmatrix} W_{\mathrm{aw}} & -W_{\mathrm{aw}}K' & -W_{\mathrm{aw}}A' \\ -KW_{\mathrm{aw}} & 2S_{\mathrm{aw}} & SB' + Z'_{\mathrm{aw}}R'_{\mathrm{aw}} \\ -AW_{\mathrm{aw}} & BS_{\mathrm{aw}} + R_{\mathrm{aw}}Z_{\mathrm{aw}} & W_{\mathrm{aw}} \end{bmatrix} > 0, \tag{27}$$

then, for $Z_{\mathrm{aw}} = \Lambda S_{\mathrm{aw}}$, system (25) is globally asymptotically stable. In our case, we define matrix W_{aw} as follows

$$W_{\mathrm{aw}} = \zeta I_{(H+S)\times(H+S)}, \tag{28}$$

where ζ is a parameter to be opportunely calibrated. This result implies that, by evaluating condition (27) for a reasonable domain of $\bar{\lambda}$ and ζ, we can assess the global asymptotic stability of our system while designing controller parameters.

E. Activation Logic

For practical implementation, it is not necessary nor reasonable at low densities (far from ρ^{cr}) to order lane-changes, since it is not likely that a congestion would occur. We propose to implement an activation logic, by introducing variable $\Phi^{\mathrm{crt}}(k)$ that indicates if the controller is active at time k ($\Phi^{\mathrm{crt}}(k) = 1$) or not ($\Phi^{\mathrm{crt}}(k) = 0$). The proposed logic reads as follows

$$\Phi^{\mathrm{crt}}(k) = \begin{cases} 1, & \text{if } \sum_j \rho_{I,j}(k) > \rho^{\mathrm{act}} \\ 0, & \text{if } \sum_j \rho_{I,j}(k) < \rho^{\mathrm{deact}} \\ \Phi^{\mathrm{crt}}(k-1), & \text{otherwise,} \end{cases} \tag{29}$$

where ρ^{act} and ρ^{deact}, are the activation and deactivation thresholds, respectively. Essentially, the controller is activated when the total density at the bottleneck exceeds ρ^{act} and it remains active until the total density at the bottleneck becomes lower than ρ^{deact}. We consider selecting $\rho^{\mathrm{act}} > \rho^{\mathrm{deact}}$ in order to avoid unnecessary switches between activation and deactivation states, which may occur due to high frequency fluctuations in traffic demand. Note that, due to the fact that the congestion originates from the bottleneck area, it is sufficient considering only total density at the bottleneck as the trigger for the proposed activation logic.

IV. Optimal Set-Point Tuning via Extremum Seeking

The proposed controller requires opportune set-points for the per-lane densities at the bottleneck area. Note that the per-lane optimal set-points, i.e., the per-lane critical densities, may sum up to a higher value than the conventional cross-section

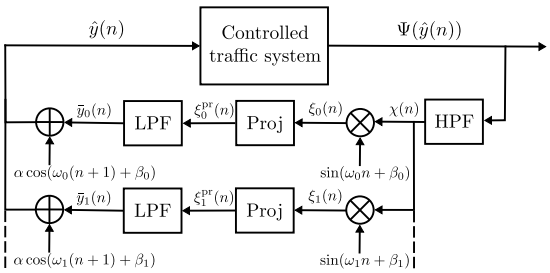

Fig. 2. The employed extremum seeking scheme.

critical density, which is typically used in ramp metering installations. This is because the lane-change control action may help avoiding pre-mature congestion at one lane, while density at other lanes is still under-critical. Finding appropriate set-points may be a non-trivial task that is commonly performed off-line, by collecting traffic data prior to the control application and analysing the obtained fundamental diagrams, in addition to some real-time fine tuning during the controller application. On the other hand, adaptive algorithms have been proposed and employed for on-line tuning the design parameters within urban control strategies (e.g., [41], [42]).

As in [12], we employ a methodology based on discrete-time extremum seeking, which is a model-free method for real-time optimisation that can be utilised for adaptively tuning set-points to achieve an optimal value of a cost utilising only real-time measurements of an appropriate cost function. Extremum seeking has been widely studied and used in several applications, e.g. [42]–[44]. In our case, in order to guarantee that the estimated set-points, denoted as $\hat{y}$, remain within a feasible interval, we incorporate also an orthogonal projection operator (34) that prohibits the estimated parameters from leaving the interval $\left[\hat{y}^{\min}, \hat{y}^{\max}\right]$ (see, for more details, [45] and [46]).

The employed control framework for multi-parameter extremum seeking, illustrated in Fig. 2, is formulated as

$$\chi(n) = -h\chi(n-1) + \Psi(\hat{y}(n)) - \Psi(\hat{y}(n-1)) \quad (30)$$

$$\xi_\ell(n) = \chi_\ell(n)\ \sin(\omega_\ell n + \beta_\ell) \quad (31)$$

$$\bar{y}_\ell(n+1) = \bar{y}_\ell(n) - \text{Proj}\left\{\gamma\ \xi_\ell(n);\ \hat{y}^{\min}, \hat{y}^{\max}\right\} \quad (32)$$

$$\hat{y}_\ell(n) = \bar{y}_\ell(n) + \alpha\cos(\omega(n+1)+\beta) \quad (33)$$

where

$$\text{Proj}\left\{\phi;\ \hat{y}^{\min}, \hat{y}^{\max}\right\} = \begin{cases} \frac{\bar{y}_\ell - \phi - \hat{y}^{\min}}{\delta}\phi, & \text{if } \bar{y}_\ell \le \hat{y}^{\min} + \phi + \delta \\ \frac{\hat{y}^{\max} - \bar{y}_\ell + \phi}{\delta}\phi, & \text{if } \bar{y}_\ell \ge \hat{y}^{\max} + \phi - \delta \\ \phi, & \text{otherwise.} \end{cases} \quad (34)$$

The cost function Ψ is evaluated while employing different set-points in the control strategy; $\ell = 0, \ldots$ is the parameter index; n is the iteration index of the extremum seeking algorithm; $\hat{y}^{\min}$ and $\hat{y}^{\max}$ are physically meaningful bounds for parameters $\hat{y}$; while h, ω_ℓ, β_ℓ, δ, γ, and α are parameters of the extremum seeking algorithm. The optimisation of the cost Ψ is achieved by sinusoidally perturbing each parameter estimate $\bar{y}_\ell$ and then employing a gradient-based optimisation by estimating, for each parameter, the gradient ξ_ℓ. The gradient is estimated by a high-pass filter (30) and a demodulator (31) with the same frequency of the perturbation (see, for more details, [44]).

Since the cost function must reflect the performance of the traffic system for given set-points $\hat{y}$, candidate functions are, for example, the Total Time Spent (TTS) [47] or the Total Throughput. Here, we employ TTS since minimising TTS is equivalent to maximization of a time-weighted sum of the total network output, which is known to be happening when the density is at its critical value (see, e.g., [48]). Finally, note that the computation time required to run the extremum seeking algorithm is negligible, as calculation is limited to solving (30)-(34) once every given interval during which the controller is applied; such interval could be, e.g., around morning or afternoon peak-hour or an entire day.

V. Experiment Set-Up

A. Nonlinear Multi-Lane Traffic Flow Model

In order to test and evaluate the performance of the proposed control strategy, we present simulation experiments using a first-order traffic flow model based on [10]. The model is used for reproducing the traffic behaviour for a multi-lane motorway and it features: (i) non-linear functions for the lateral flows of manually driven vehicles (which may also act as disturbances for the designed controller); (ii) a Cell Transmission Models (CTM)-like formulation for the longitudinal flows; and (iii) a non-linear formulation to account for the capacity drop phenomenon. Briefly, we consider the conservation law equation (1), where all variables are defined as in Section II. Lateral flows due to manual lane-changing, denoted as $f_{i,j}^{M}(k)$ are considered among adjacent lanes of the same segment, and corresponding rules are defined in order to properly assign and bound their values. They are computed as

$$f_{i,j}^{\rm M}(k) = l_{i,j,j+1}(k) - l_{i,j+1,j}(k), \quad (35)$$

where

$$l_{i,\bar{j},j}(k) = \min\left\{1, \frac{E_{i,j}(k)}{D_{i,j-1,j}(k) + D_{i,j+1,j}(k)}\right\} D_{i,\bar{j},j}(k) \quad (36)$$

$$E_{i,j}(k) = \frac{L_i}{T}\left[\rho_{i,j}^{\rm jam} - \rho_{i,j}(k)\right] \quad (37)$$

$$D_{i,j}(k) = \frac{L_i}{T}\rho_{i,j}(k) A_{i,j,\bar{j}}(k) \quad (38)$$

$$A_{i,j,\bar{j}}(k) = \mu\ \max\left\{0, \frac{G_{i,j,\bar{j}}(k)\rho_{i,j}(k) - \rho_{i,\bar{j}}(k)}{G_{i,j,\bar{j}}(k)\rho_{i,j}(k) + \rho_{i,\bar{j}}(k)}\right\}, \quad (39)$$

and $\bar{j} = j \pm 1$. Variable E denotes the available space, in terms of flow acceptance, while D denotes the lateral demand flow, which is computed via definition of the attractiveness rate A. Equation (36) accounts for the potentially limited space that may not be sufficient for accepting the lateral flow entering from both sides of a cell. In (39), the factor G is mostly equal to 1, which implies the intent of drivers to move towards a faster lane (which may lead also to balanced densities among

lanes), but may also be tuned to reflect particular location-dependent effects where lateral flow may occur in the direction from a lower density to a higher one (e.g. upstream of on- and off-ramps, lane drop locations, etc.); while μ is a constant coefficient in the range [0, 1] reflecting the "aggressiveness" in lane-changing.

Longitudinal flows are the flows going from a cell to the next downstream one, while remaining in the same lane. We employ the Godunov-discretised first-order model proposed in [10], employing however a non-linear exponential demand function for under-critical densities, to obtain a more realistic behaviour at low densities. The model accounts also for the capacity drop phenomenon, via a linearly decreasing demand function for over-critical densities [49] and a linear reduction of the maximum flow as a function of the entering lateral flows. The overall formulation for longitudinal flow is

$$q_{i,j}(k) = \min\left\{Q^D_{i,j}(k),\, Q^E_{i+1,j}(k) - d_{i+1,j}(k)\right\}, \quad (40)$$

where

$$Q^D_{i,j}(k) = \begin{cases} v^{\max}_{i,j} \exp\left[-\frac{1}{\alpha}\left(\frac{\rho_{i,j}(k)}{\rho^{\mathrm{cr}}_{i,j}}\right)^{\alpha}\right]\rho_{i,j}(k), & \text{if } \rho_{i,j}(k) < \rho^{\mathrm{cr}}_{i,j} \\ \frac{(1-\gamma)\,Q^{\mathrm{cap}}_{i,j}}{\rho^{\mathrm{cr}}_{i,j} - \rho^{\mathrm{jam}}_{i,j}}\left[\rho_{i,j}(k) - \rho^{\mathrm{jam}}_{i,j}\right] + Q^B_{i,j}(k), & \text{otherwise} \end{cases} \quad (41)$$

$$Q^E_{i+1,j}(k) = \begin{cases} Q^{\mathrm{cap}}_{i+1,j}, & \text{if } \rho_{i+1,j}(k) < \rho^{\mathrm{cr}}_{i+1,j} \\ w_{i+1}\left[\rho^{\mathrm{jam}}_{i+1,j} - \rho_{i+1,j}(k)\right], & \text{otherwise} \end{cases} \quad (42)$$

$$Q^B_{i,j}(k) = \gamma\; Q^{\mathrm{cap}}_{i,j} - \nu\left[l_{i,j+1,j}(k) + l_{i,j-1,j}(k)\right]. \quad (43)$$

Parameter $v^{\max}$ denotes the maximum speed, Q^{cap} is the capacity flow, ρ^{cr} is the critical density, while $\alpha = \left(\ln \frac{Q^{\mathrm{cap}}}{v^{\max}\rho^{\mathrm{cr}}}\right)^{-1}$ [50]. Parameter φ influences the extent of capacity drop due to overcritical densities, while ν affects the capacity drop due to entering lateral flows. Note that, setting $\varphi = 1$ and $\nu = 0$, we obtain a conventional first-order model, i.e., no capacity drop appears at the head of congestion.

The proposed model is characterised by constant parameters; specifically, it is assumed that the FD does not change during the simulation. However, in the presence of a percentage of connected and automated vehicles using driving assistance systems, such as adaptive cruise control (ACC) or cooperative ACC (CACC), these values may be altered, resulting in a different FD (see, e.g., [51]; this phenomenon is not captured by this model.

B. Network Description and Simulation Configuration

We consider a hypothetical two-lane motorway stretch, shown in Fig. 3, to test and evaluate the performance of the proposed strategy. In particular, we consider a network composed of 10 segments characterised by the same length $L_i = 0.5$ km, while we employ a time step $T = 10$ s. Different lanes feature different parameters, namely a different Fundamental Diagram (FD), which may reflect different traffic composition (e.g., a high number of heavy vehicles reducing

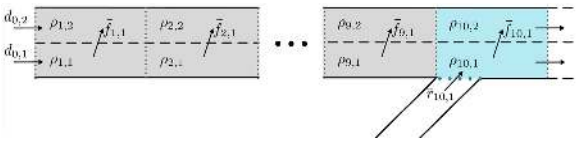

Fig. 3. Motorway stretch employed in the simulation experiments.

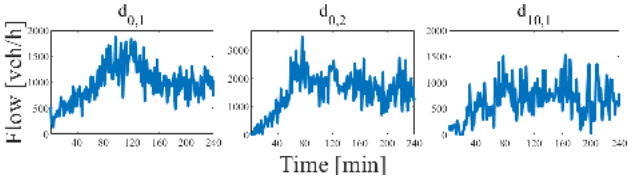

Fig. 4. Traffic demand used in the simulation experiments.

the capacity of a specific lane). In addition, the used traffic demand is depicted in Fig. 4.

We test our controller by considering only a percentage η of vehicles that are connected and automated, i.e., implementing the lane-changes dictated by the controller, while the other vehicles are assumed to behave according to the lane-changing model for manual vehicles described in (35)-(39). Moreover, while the controller is not activated ($\Phi^{\mathrm{crt}}(k) = 0$), all vehicles have the same lane-changing behaviour as manual vehicles. The lateral flow implemented in numerical experiments is therefore

$$\bar{f}_{i,j}(k) = \begin{cases} f^{\mathrm{M}}_{i,j}(k), & \text{if } \Phi^{\mathrm{crt}}(k) = 0 \\ \mathrm{sat}\left(f_{i,j}(k)\right) + (1-\eta) f^{\mathrm{M}}_{i,j}(k), & \text{if } \Phi^{\mathrm{crt}}(k) = 1. \end{cases} \quad (44)$$

Since ramp metering actions may create a queue outside the motorway network, we introduce the following dynamics for the queue length $w(k)$ (in veh)

$$w(k+1) = w(k) + T\left(d_{10,1}(k) - \bar{r}_{10,1}(k)\right), \quad (45)$$

where $d_{10,1}(k)$ is the on-ramp external demand during time interval $(k, k+1]$. The ramp flow implemented in our numerical experiments is

$$\bar{r}_{10,1}(k) = \begin{cases} d_{10,1}(k), & \text{if } \Phi^{\mathrm{crt}}(k) = 0, \\ \mathrm{sat}\left(r_{10,1}(k)\right), & \text{if } \Phi^{\mathrm{crt}}(k) = 1. \end{cases} \quad (46)$$

We consider the following bounds for the control inputs:

$$\mathrm{sat}(f_{i,j}) = \begin{cases} f^{\min}_{i,j} = -\frac{L_i}{T}\rho_{i,j+1}, & \text{if } f_{i,j} \le f^{\min}_{i,j} \\ f^{\max}_{i,j} = \frac{L_i}{T}\rho_{i,j}, & \text{if } f_{i,j} \ge f^{\max}_{i,j} \\ f_{i,j}, & \text{otherwise;} \end{cases} \quad (47)$$

$$\mathrm{sat}(r_{10,1}) = \begin{cases} r^{\min}_{10,1} = 0, & \text{if } r_{10,1} \le r^{\min}_{10,1} \\ r^{\max}_{10,1} = \min\left(\frac{w}{T} + d_{10,1},\, Q^{\mathrm{cap}}_{\mathrm{r}}\right), & \text{if } r_{10,1} \ge r^{\max}_{10,1} \\ r_{10,1}, & \text{otherwise.} \end{cases} \quad (48)$$

Bounds for lateral flows (47) account for the amount of vehicles in the demand cell that are available for lane-changing; whereas bounds for ramp flow (48) consider the total available flow, which depends on both vehicles queuing and on-ramp demand, as well as the ramp capacity $Q^{\mathrm{cap}}_{\mathrm{r}}$.

TABLE I
PARAMETERS USED IN THE NONLINEAR MULTI-LANE TRAFFIC FLOW MODEL

	$j=1$	$j=2$
$v^{\max}$ [km/h]	100	100
Q^{cap} [veh/h]	1800	2400
ρ^{cr} [veh/km]	22	26
ρ^{jam} [veh/km]	120	160
γ	0.6	0.6
ν	0.8	0.8
G	1	1
μ	0.6	0.6

Note that, since the longitudinal flow model implemented in this work assumes priority for the on-ramp flow (via (40)), the assigned on-ramp flow, for both the no-control and controlled cases, is not restricted by the amount of vehicles within the mainstream cell, while restrictions (the supply part of the FD) apply for the flow coming from upstream. This implies that, unless the ramp flow is limited by a controller, it is fully allowed to enter the mainstream; therefore, there is no need to specify bounds dependent on the mainstream traffic conditions. In addition, in the presented experiments (as well as in the majority or real situations), we assume that the ramp capacity is smaller than the mainstream one; in the opposite case, there may be need to consider for the presence of on-ramp queues also for the no-control case, thus $d_{10,1}$ should be saturated in (46).

As performance metric we employ the TTS over a finite time horizon K, defined as

$$\text{TTS} = T \sum_{k=0}^{K} \sum_{i=0}^{N} L_i \sum_{j=m_i}^{M_i-1} \rho_{i,j}(k) + T\, w(k). \tag{49}$$

Note that TTS is used both for numerical comparisons among scenarios and as cost function to be minimised within the extremum seeking algorithm.

C. ALINEA

To further investigate the benefits of considering lateral flow control, we include numerical experiments employing the ramp-metering feedback controller ALINEA [47] in an additional simulation scenario. The goal of ALINEA is to maintain the total (cross-lane) density at its critical value in the bottleneck segment, by manipulating only the ramp inflow. Accordingly, we implement the following control law

$$u(k) = u(k-1) - K_A \left(\rho_{\text{tot}}(k) - \hat{\rho}_{\text{tot}}(k)\right), \tag{50}$$

where, e.g., in our scenario, $\rho_{\text{tot}}(k) = \rho_{10,1}(k) + \rho_{10,2}(k)$; the set-point $\hat{\rho}_{\text{tot}} = \rho^{cr}_{10,1} + \rho^{cr}_{10,2}$; while the gain K_A may be selected via a trial-and-error tuning procedure.

D. Presence of Off-Ramps

In order to illustrate the robustness of the proposed controller to unmeasured disturbances, we develop an additional numerical experiment that accounts for the presence of an off-ramp. We consider the same network as in Section V-B, assuming the presence of an additional off-ramp located in segment 3 lane 1; this is formulated by adding a specific off-ramp term $-\gamma_{3,1} \sum q_{3,j}(k)$ in the conservation law (1) of the nonlinear model of Section V; see [10] for details. For this scenario, no modifications are made to the integrated controller. Note also, that in this case also the no-control case used for comparison needs to be modified to account for the presence of the off-ramp.

VI. EXPERIMENTAL RESULTS

A. No-Control Case

The no-control case is defined by implementing the nonlinear traffic model (35)-(43) for the described network. Looking at Fig. 5(a), one may observe that a strong congestion develops at the merge area (segment $i = 10$) and spills-back until segment $i = 1$. The congestion starts at the merge bottleneck area, where the density exceeds its critical value in both lanes (see Fig. 6).

The reasons for the congestion are a) the high inflow entering from the ramp, since the total demand during the peak period is about 4600 veh/h, while the overall capacity is 4200 veh/h; as well as b) the inefficient "natural" lane-changing flow. Capacity drop also occurs at the bottleneck cells of the stretch, which worsens the congestion.

B. Lateral Flow and Ramp Metering Controller

In order to test our proposed strategy, we apply the linear dynamic compensator (21), (23) to the non-linear traffic model (35)-(43). In the following experiments, we assume that the current system state (density) is exactly known, i.e., there are no measurement errors. To start with, we show here results for a percentage $\eta = 0.5$ of connected and automated vehicles. The controller is active during the entire simulation, thus $\Phi^{\text{crt}}(k) = 1, \forall k$.

As described in Section III-C, the anti-windup scheme requires nominal values for all state and input variables (i.e., x_{d} and u_{d}). Assuming steady-state conditions, we impose desired states at bottleneck cells equal to critical densities and we derive the remaining nominal states using a pseudoinverse (least-squares-based) approach.

We initially perform a set of experiments to select the controller parameters and to investigate the sensitivity of the controller to the choice of parameters w_Q, w_{R_1}, w_{R_2}, $\bar{\lambda}_1$, and $\bar{\lambda}_2$, using as evaluation metric the resulting TTS. As can be seen from Fig. 7, which shows how the obtained TTS varies by altering, in pairs, the weights used in the cost function (14), the controller is capable of obtaining a low TTS value (blue areas) for a wide range of such parameters. This indicates that the controller is little sensitive to the parameters choice, which is desirable for practical applications, since it implies there is no need for fine tuning in order to achieve satisfactory results.

Observing Fig. 8 (left), we see that the performance of the controller is not affected by values of $\bar{\lambda}_1$ and $\bar{\lambda}_2$ within the interval $(-1, 1)$. However, since these results are based on numerical experiments, there may be some conditions (e.g.,

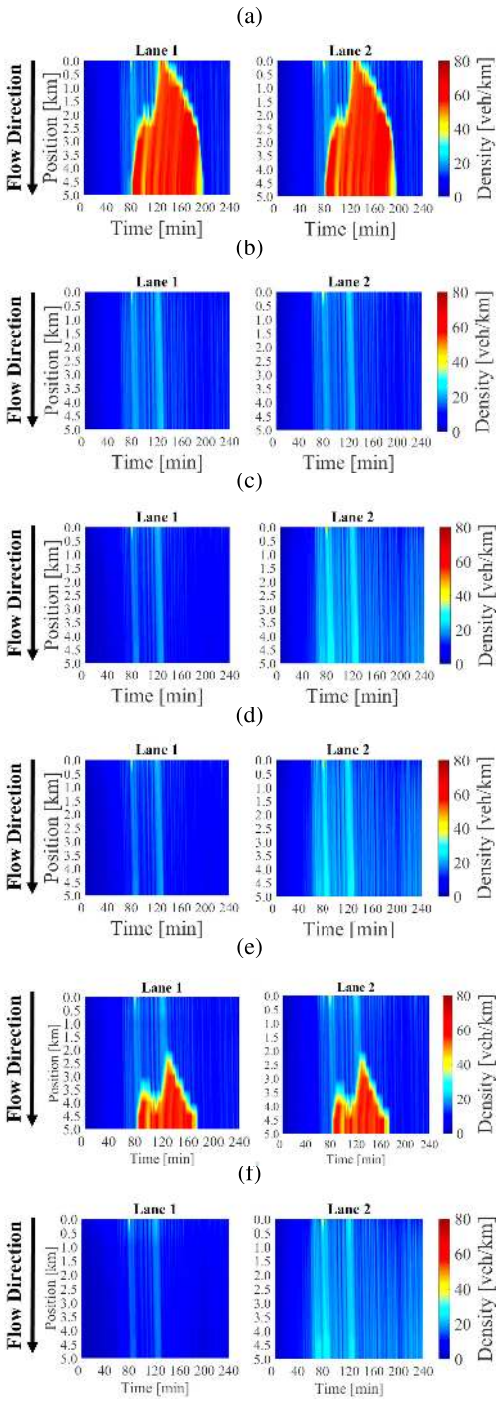

Fig. 5. Contour plots of per-lane densities for the different tested scenarios. (a) No-control case, (b) ALINEA controlled case ($\eta = 0.5$), (c) Controlled case ($\eta = 0.5$), (d) Controlled case, with activation logic ($\eta = 0.5$), (e) Off-ramp No-control case, (f) Off-ramp controlled case ($\eta = 0.5$).

saturation of some inputs), which are not appearing during our simulation runs, where the system becomes unstable, which would cause a deterioration of traffic conditions. Thus, we investigate the closed-loop stability of our system following the results presented in Section III-D; namely, we check whether matrix Θ in (27) is positive definite for a domain of $\bar{\lambda}_1$, $\bar{\lambda}_2$, and ζ. For the sake of brevity, we illustrate results for $\bar{\lambda} \equiv \bar{\lambda}_1 = \bar{\lambda}_2$ in Fig. 8 (right), where we can observe that there are no unstable eigenvalues for $-0.2 \leq \bar{\lambda} \leq 1$, which

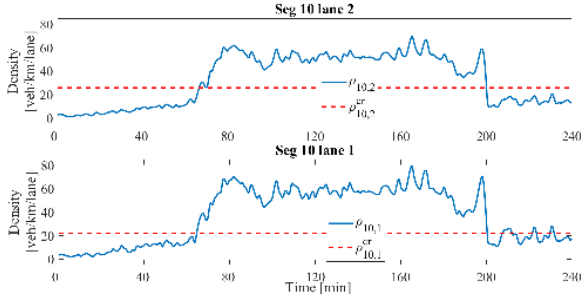

Fig. 6. Density at bottleneck cells in the no-control case.

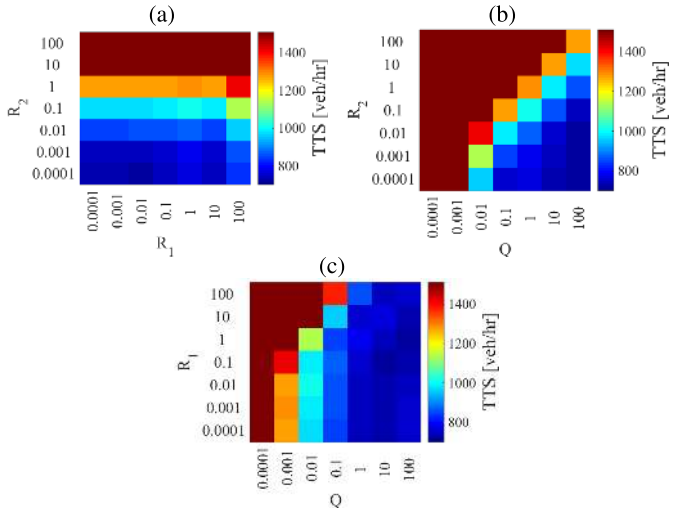

Fig. 7. Sensitivity analysis showing TTS for: (a) a domain of w_{R_1} and w_{R_2}, while $w_Q = 1$; (b) a domain of w_Q and w_{R_2}, while $w_{R_1} = 1$; (c) a domain of w_Q and w_{R_1}, while $w_{R_2} = 0.001$.

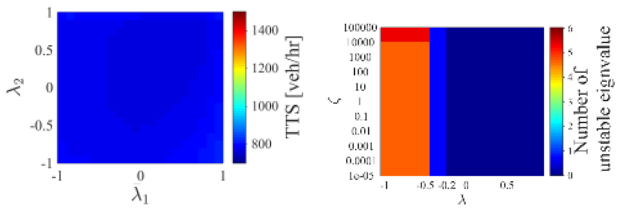

Fig. 8. TTS for a domain of λ_1 and λ_2 (left) and number of negative eigenvalues in Θ (right), while $w_Q = 1$, $w_{R_1} = 1$, and $w_{R_2} = 0.001$.

means that Θ is positive definite. As a result, this implies that assigning eigenvalues $\bar{\lambda}_1$ and $\bar{\lambda}_2$ within the range $(-0.2, 1)$ not only produces good results in term of TTS (Fig. 8 (left)), but also guarantees asymptotic stability of the closed loop system.

We now show detailed simulation results using $w_Q = 1$, $w_{R_1} = 1$, $w_{R_2} = 0.001$, $\lambda_1 = 0.75$, and $\lambda_2 = 0.75$. We can observe in Fig. 5(b) that the congestion fully disappears, while the entire network is basically in free-flow for the entire simulation. The controller is in fact capable of maintaining densities at the bottleneck areas at their critical values during the period with high demand (see Fig. 9). This is achieved by the integrated usage of lane-changes and ramp metering actions. The former are implemented during the entire simulation (see Fig. 10 (left)), assigning vehicles to move from

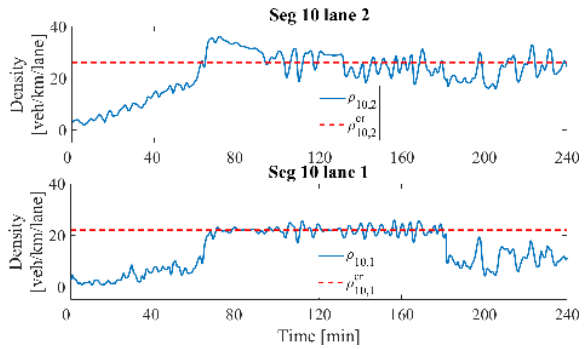

Fig. 9. Density at bottleneck cells in the controlled case.

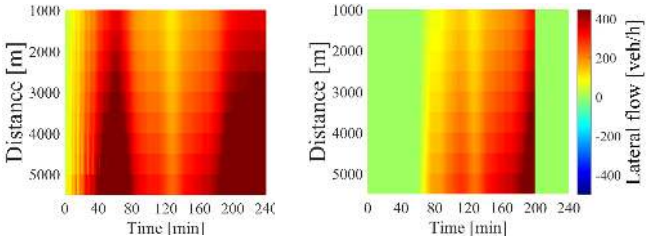

Fig. 10. Contour plots of lateral flows in the controlled case, for $\eta = 50\%$, without activation logic (left) and with activation logic (right).

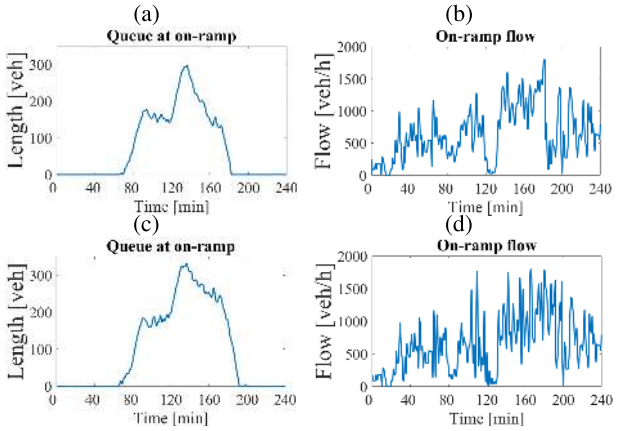

Fig. 11. On-ramp queue (a) and flow (b) in the controlled case, for $\eta = 50\%$; and on-ramp queue (c) and flow (d) in the ALINEA case.

lane $j = 1$ to lane $j = 2$, i.e., moving away from the merging lane; while the latter are implemented during the peak period only, creating a large queue (see Fig. 11), which is not upper-bounded in our experiments. The TTS improvement is about 26% (see also Table II).

C. Implementation of the Activation Logic

As can be observed from Fig. 10 (left), a large amount of lane-changes is assigned by the controller also during periods when the density at the bottleneck is low (see Fig. 9). In order to prevent from ordering these unnecessary actions, we test here the activation logic designed in Section III-E. In our experiments, we choose $\rho^{\text{act}} = 0.7 \sum_{j=1}^{2} \rho^{\text{cr}}_{10,j}$ and $\rho^{\text{deact}} = 0.5 \sum_{j=1}^{2} \rho^{\text{cr}}_{10,j}$, while we initialise $\Phi^{\text{crt}}(0) = 0$. Similarly as in the previous case, using the activation logic,

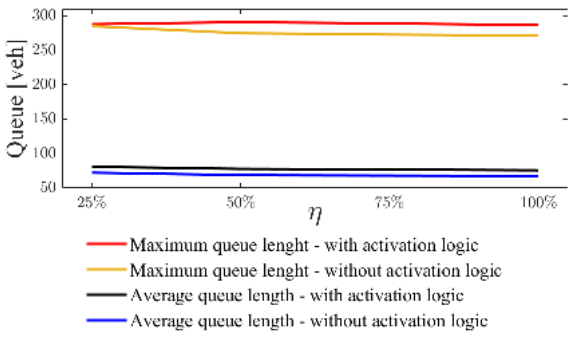

Fig. 12. On-ramp queue comparison for different η.

TABLE II
TTS FOR THE DIFFERENT SCENARIOS

	Penetration rate η (%)	TTS [veh·h]	TTS improvement (%)	Total Number of Lane Changes (N-LC)	N-LC improvement (%)
No control case		1060		15210	
Without activation logic	25	795	25	15163	0.3
	50	783	26.1	11926	21.6
	75	779.9	26.4	10969	27.8
	100	777.9	26.6	9938	34.6
With activation logic	25	821	22	5648	62.9
	50	810	23.6	4769	68.6
	75	806.6	23.9	4474	70.6
	100	802.7	24.2	4143	72.8
ALINEA	100	901.2	15	12982	14.6

the congestion fully disappears and the densities at the bottleneck area remain at their critical values, as can be seen in Figs. 13 and 5(d). Nevertheless, differently from the previous case, the controller is active only in the central part of the simulation, namely when demand is high (see Fig. 10 (right)). This has a minor impact on the TTS, which results slightly worse than without employing the activation logic, but still presents a considerable improvement of about 23% with respect to the no-control case (see Table II). On the other hand, employing the activation logic produces a considerable reduction of total number of lane-changes, in comparison with the controlled case without activation logic (see Table II), which is known having a positive effect on traffic smoothness and safety.

D. Different Penetration Rates of Controlled Vehicles

We test here a range of different penetration rates of connected and automated vehicles, which also shows the robustness of the proposed controlled to disturbances. We evaluate TTS for $\eta = \{0.25, 0.5, 0.75, 1\}$ and we report numerical results in Table II. Note that $\eta = 1$ corresponds to 100% connected and automated vehicles, which is therefore an upper bound for the effectiveness of the proposed strategy. Our results show that a 26.6% improvement is achieved for $\eta = 1$,

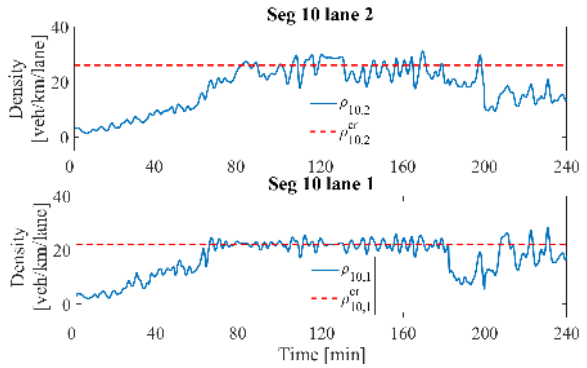

Fig. 13. Density at bottleneck cells in the controlled case with activation logic.

while TTS does not deteriorate significantly for the tested lower penetration rate. The resulting ramp-metering control actions for the different penetration rates η are very similar; we present in Fig. 12 the maximum and average values for the queue created at the on-ramp. From these results, we can deduce that, in the presence of a higher amount of vehicles able to perform controlled lane-changes, the controller needs to keep less vehicles queuing outside the mainstream, with a larger improvement in term of traffic conditions. Moreover, we can observe that applying the activation logic has minimal effect on the resulting TTS, therefore, in practical implementations, its usage should be preferred.

E. ALINEA Implementation

Considering (50), we set $K_A = 53$ tuned using some experiments. Applying the controller (50) in the model (35)-(43), we observe that, despite the ALINEA strategy is successful in maintaining total density at its critical value (see Fig. 14(a)) and avoiding the creation of congestion (see Fig. 5(b)), the per-lane densities are not maintained at their respective critical values (see Fig. 14(b)). Since per-lane capacity is achieved when the per-lane densities are at critical values, this implies that, in this case, the resulting total throughput is lower, causing also longer queues at the on-ramp (see Fig.11). As a consequence, the resulting TTS is higher than in the previous cases, namely TTS = 901.2 veh·h, proving the added benefit of applying lane-changing control in combination with ramp metering.

F. Off-Ramp Case

As it was envisioned, adding an off-ramp does not have a significant impact on the application of our controller. This is due to the fact that, since the outflow exiting the motorway at the off-ramp can be viewed as an unknown disturbance, this is automatically rejected by the controller in a similar manner as other disturbances affecting the system. The TTS for the related no-control case, i.e., where the off-ramp is considered, is TTS = 677 veh·h; corresponding density contour plots are presented in Fig. 5(e). Note the mildest congestion, due to the fact that some flow leaves the network at the off-ramp, which is located upstream of the bottleneck area. We present here only results applying the controller with $\eta = 0.5$, resulting

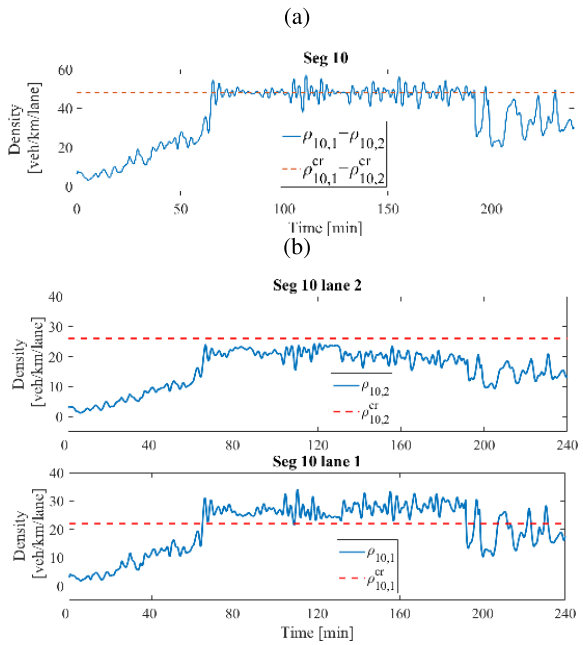

Fig. 14. Density at bottleneck cells in the ALINEA controlled case. (a) Summation of densities at bottlenecks, (b) Separated value of density at bottleneck cells.

TABLE III
PARAMETERS USED IN THE EXTREMUM SEEKING SCHEME

h	ω_n [rad]	β_1 [rad]	β_2 [rad]	σ [$\frac{\text{veh}}{\text{km}}$]	$\hat{y}^{min}$	$\hat{y}^{max}$	γ	α
0.9	$\frac{\pi}{1.5}$	0	$\frac{\pi}{2}$	2	0.8	1	0.05	0.5

in TTS = 545 veh·h, which is a 18% improvement with respect to the no-control case. The density results related to this experiment are shown in Fig. 5(f), where one can notice that the controller is capable of resolving the corresponding congestion in a similar way as in the previous cases.

G. Extremum Seeking Algorithm

The results presented so far are based on the knowledge of the exact value of critical densities at the bottleneck location. To overcome a situation where these values may be unknown, we test here the effectiveness of the extremum seeking algorithm for finding the optimal set-points proposed in Section IV. We employ the same network as described above, applying the controller with activation logic presented in Section VI-C, computing the TTS for the controlled area over a horizon $t = 240$ min, which corresponds to one iteration of the extremum seeking algorithm, and using the cost $\Psi = -\text{TTS}$. This procedure is iteratively performed for 600 times in our case. The set of parameters used in the algorithm is presented in Table III.

The set-points are initialised as $\hat{\rho}_{10,1} = 28$ veh/km, $\hat{\rho}_{10,2} = 24$ veh/km, which represents a non-optimal set-point configuration for our controller (i.e., set-points are different from the critical densities). As shown in Fig. 15 (left), the proposed algorithm achieves the optimal cost, while also reaching

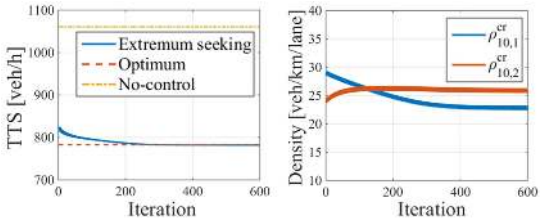

Fig. 15. The cost function (left) in the no-control case (yellow), optimum achieved setting critical densities as set-points (red), and applying the extremum seeking algorithm (blue); and the set-points achieved by the extremum seeking algorithm (right).

and maintaining the optimal density set-points (Fig. 15 (right)). Note that, once the extremum seeking algorithm converges to the true critical densities, the traffic pattern is the same as reported previously in this section, therefore no additional simulation results are reported.

VII. Conclusion

The methodology presented in this article demonstrates that integrated lane-changing and ramp metering control may be employed as an efficient management strategy, in a future where a percentage of connected and automated vehicles will be present in traffic. The proposed strategy is designed based on well-established feedback control methods, which guarantee its efficiency and applicability, while also requiring a very limited computation effort. The robustness and insensitivity to parameters choice, as well as the adaptive component, i.e., the extremum seeking algorithm, allow for an easy implementation, without necessity of lengthy and costly traffic observations and parameter tuning. The controller requires per-lane density data for the entire stretch under consideration, which may be obtained from spot detectors and/or connected vehicles. Various traffic state estimators have been proposed to produce such data, starting also from incomplete measurements, such as [52]–[55].

Despite the fact that the model used in the experiments does not consider explicitly effects of connected and automated vehicles on the FD shape and parameters, which may have an effect on the set-points that are then employed within our controller, the experiments with the extremum seeking algorithm demonstrate that, even without any knowledge of the nonlinear model structure or parameters, the set-points converge to the critical densities. Therefore, despite our experiments do not cover explicitly the case of varying capacities, we are confident that our method would behave appropriately in such cases as well. In order to further investigate this issue, we envision testing this strategy in microscopic simulation, which allows also to assess in more detail the impact of different penetration rates and different aspects related to traffic behaviour, such as, e.g., heterogeneity of drivers and compliance to control tasks.

Further developments include the incorporation in the control strategy of mainstream flow control, which may be implemented, for example, via variable speed limits, as well as accounting for the presence of multiple bottlenecks; the latter could, e.g., follow the works in [56], [57]. Another possible direction is to investigate the case of more complex networks, characterised by multiple destinations, where the behaviour of connected and automated vehicles is defined per-destination.

References

[1] R. Bishop, *Intelligent Vehicle Technology and Trends*. Norwood, MA, USA: Artech House, 2005.
[2] C. Diakaki, M. Papageorgiou, I. Papamichail, and I. Nikolos, "Overview and analysis of vehicle automation and communication systems from a motorway traffic management perspective," *Transp. Res. A, Policy Pract.*, vol. 75, pp. 147–165, May 2015.
[3] M. Papageorgiou, C. Diakaki, V. Dinopoulou, A. Kotsialos, and Y. Wang, "Review of road traffic control strategies," *Proc. IEEE*, vol. 91, no. 1, pp. 2043–2067, Dec. 2003.
[4] V. F. Hurdle, M. I. Merlo, and D. Robertson, "Study of speed-flow relationships on individual freeway lanes," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 1591, no. 1, pp. 7–13, Jan. 1997.
[5] V. L. Knoop, A. Duret, C. Buisson, and B. van Arem, "Lane distribution of traffic near merging zones influence of variable speed limits," in *Proc. 13th Int. IEEE Conf. Intell. Transp. Syst.*, Sep. 2010, pp. 485–490.
[6] A. Duret, S. Ahn, and C. Buisson, "Lane flow distribution on a three-lane freeway: General features and the effects of traffic controls," *Transp. Res. C, Emerg. Technol.*, vol. 24, pp. 157–167, Oct. 2012.
[7] Y. Zhang and P. A. Ioannou, "Integrated control of highway traffic flow," *J. Control Decis.*, vol. 5, no. 1, pp. 19–41, Jan. 2018.
[8] R. W. Hall and D. Lotspeich, "Optimized lane assignment on an automated highway," *Transp. Res. Part C: Emerg. Technol.*, vol. 4, no. 4, pp. 211–229, Aug. 1996.
[9] K. Kim, J. V. Medanić, and D.-I. Cho, "Lane assignment problem using a genetic algorithm in the Automated Highway Systems," *Int. J. Automot. Technol.*, vol. 9, no. 3, pp. 353–364, Jun. 2008.
[10] C. Roncoli, M. Papageorgiou, and I. Papamichail, "Traffic flow optimisation in presence of vehicle automation and communication systems–Part II: Optimal control for multi-lane motorways," *Transp. Res. C, Emerg. Technol.*, vol. 57, pp. 260–275, Aug. 2015.
[11] C. Roncoli, I. Papamichail, and M. Papageorgiou, "Hierarchical model predictive control for multi-lane motorways in presence of vehicle automation and communication systems," *Transp. Res. C, Emerg. Technol.*, vol. 62, pp. 117–132, Jan. 2016.
[12] C. Roncoli, N. Bekiaris-Liberis, and M. Papageorgiou, "Optimal lane-changing control at motorway bottlenecks," in *Proc. IEEE 19th Int. Conf. Intell. Transp. Syst. (ITSC)*, vol. 5, Nov. 2016, pp. 1785–1791.
[13] Y. Zhang and P. A. Ioannou, "Combined variable speed limit and lane change control for highway traffic," *IEEE Trans. Intell. Transport. Syst.*, vol. 7, pp. 1812–1823, 2017.
[14] C. Roncoli, N. Bekiaris-Liberis, and M. Papageorgiou, "Lane-changing feedback control for efficient lane assignment at motorway bottlenecks," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 2625, no. 1, pp. 20–31, Jan. 2017.
[15] W. J. Schakel and B. Van Arem, "Improving traffic flow efficiency by in-car advice on lane, speed, and headway," *IEEE Trans. Intell. Transport. Syst.*, vol. 4, pp. 1812–1823, 2014.
[16] W. Schakel and B. van Arem, "In-car tactical advice using delayed detector data," in *Proc. IEEE Intell. Vehicles Symp. (IV)*, Jun. 2015, pp. 982–987.
[17] A. Ghaffari, A. Khodayari, A. Kamali, F. Tajdari, and N. Hosseinkhani, "New fuzzy solution for determining anticipation and evaluation behavior during car-following maneuvers," *Proc. Inst. Mech. Eng., D, J. Automobile Eng.*, vol. 232, no. 7, pp. 936–945, Jun. 2018.
[18] A. Ghaffari, A. Khodayari, A. Kamali, and F. Tajdari, "A new model of car following behavior based on lane change effects using anticipation and evaluation idea," *Iranian J. Mech. Eng. Trans. ISME*, vol. 16, no. 2, pp. 26–38, 2015.
[19] F. Tajdari, A. Ghaffari, A. Khodayari, A. Kamali, N. Zhilakzadeh, and N. Ebrahimi, "Fuzzy control of anticipation and evaluation behaviour in real traffic flow," in *Proc. 7th Int. Conf. Robot. Mechatronics (ICRoM)*, Nov. 2019, pp. 248–253.
[20] D. Yang, S. Zheng, C. Wen, P. J. Jin, and B. Ran, "A dynamic lane-changing trajectory planning model for automated vehicles," *Transp. Res. C, Emerg. Technol.*, vol. 95, pp. 228–247, Oct. 2018.
[21] M. Ramezani and E. Ye, "Lane density optimisation of automated vehicles for highway congestion control," *Transportmetrica B, Transp. Dyn.*, pp. 1–21, 2019.
[22] C. Diakaki, M. Papageorgiou, and K. Aboudolas, "A multivariable regulator approach to traffic-responsive network-wide signal control," *Control Eng. Pract.*, vol. 10, no. 2, pp. 183–195, Feb. 2002.

[23] M. Papageorgiou, H. Hadj-Salem, and F. Middelham, "ALINEA local ramp metering: Summary of field results," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 1603, no. 1, pp. 90–98, Jan. 1997.

[24] I. Papamichail, M. Papageorgiou, V. Vong, and J. Gaffney, "Heuristic ramp-metering coordination strategy implemented at monash freeway, australia," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 2178, no. 1, pp. 10–20, Jan. 2010.

[25] A. Ferrara, S. Sacone, and S. Siri, *Freeway Traffic Modelling and Control*. Berlin, Germany: Springer, 2018.

[26] C. Roncoli, M. Papageorgiou, and I. Papamichail, "Traffic flow optimisation in presence of Vehicle Automation and Communication Systems - Part I: A first-order multi-lane model for motorway traffic," *Transp. Res. C, Emerg. Technol.*, vol. 57, pp. 241–259, 2015.

[27] F. Tajdari, C. Roncoli, N. Bekiaris-Liberis, and M. Papageorgiou, "Integrated ramp metering and lane-changing feedback control at motorway bottlenecks," in *Proc. 18th Eur. Control Conf. (ECC)*, Jun. 2019, pp. 3179–3184.

[28] R. Courant, K. Friedrichs, and H. Lewy, "Über die partiellen Differenzengleichungen der mathematischen Physik," *Math. Annalen*, vol. 100, no. 1, pp. 32–74, Dec. 1928.

[29] L. Elefteriadou, R. P. Roess, and W. R. McShane, "Probabilistic nature of breakdown at freeway merge junctions," *Transp. Res. Rec.*, no. 1484, pp. 80–89, 1995.

[30] M. R. Lorenz and L. Elefteriadou, "Defining freeway capacity as function of breakdown probability," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 1776, no. 1, pp. 43–51, Jan. 2001.

[31] M. J. Cassidy and J. Rudjanakanoknad, "Increasing the capacity of an isolated merge by metering its on-ramp," *Transp. Res. B, Methodol.*, vol. 39, no. 10, pp. 896–913, Dec. 2005.

[32] K. J. Åström and T. Hägglund, *PID Controllers: Theory, Design, Tuning*. International Society for Measurement and Control, 1995.

[33] F. L. Lewis, D. L. Vrabie, and V. L. Syrmos, *Optimal Control*. Hoboken, NJ, USA: Wiley, 2012.

[34] R. L. Williams and D. A. Lawrence, *Linear State-Space Control Systems*. Hoboken, NJ, USA: Wiley, 2007.

[35] J. P. Hespanha, *Linear Systems Theory*. Princeton, NJ, USA: Princeton Univ., 2009.

[36] B. D. O. Anderson and J. B. Moore, *Linear Optimal Control*. Upper Saddle River, NJ, USA: Prentice-Hall, 1971.

[37] K. J. Åström and L. Rundqwist, "Integrator windup and how to avoid it," in *Proc. Amer. Control Conf.*, Jun. 1989, pp. 1693–1698.

[38] M. V. Kothare, P. J. Campo, M. Morari, and C. N. Nett, "A unified framework for the study of anti-windup designs," *Automatica*, vol. 30, no. 12, pp. 1869–1883, Dec. 1994.

[39] N. Kapoor, A. R. Teel, and P. Daoutidis, "An anti-windup design for linear systems with input saturation," *Automatica*, vol. 34, no. 5, pp. 559–574, May 1998.

[40] J. M. G. da Silva and S. Tarbouriech, "Anti-windup design with guaranteed regions of stability for discrete-time linear systems," *Syst. Control Lett.*, vol. 55, no. 3, pp. 184–192, Mar. 2006.

[41] A. Kouvelas, K. Aboudolas, E. B. Kosmatopoulos, and M. Papageorgiou, "Adaptive performance optimization for large-scale traffic control systems," *IEEE Trans. Intell. Transp. Syst.*, vol. 12, no. 4, pp. 1434–1445, Dec. 2011.

[42] R. Kutadinata, W. Moase, C. Manzie, L. Zhang, and T. Garoni, "Enhancing the performance of existing urban traffic light control through extremum-seeking," *Transp. Res. C, Emerg. Technol.*, vol. 62, pp. 1–20, Jan. 2016.

[43] J.-Y. Choi, M. Krstic, K. B. Ariyur, and J. S. Lee, "Extremum seeking control for discrete-time systems," *IEEE Trans. Autom. Control*, vol. 47, no. 2, pp. 318–323, Feb. 2002.

[44] K. B. Ariyur and M. Krstić, *Real-Time Optimization by Extremum-Seeking Control*. Hoboken, NJ, USA: Wiley, 2003.

[45] P. Frihauf, M. Krstic, and T. Başar, "Finite-horizon LQ control for unknown discrete-time linear systems via extremum seeking," *Eur. J. Control*, vol. 19, no. 5, pp. 399–407, Sep. 2013.

[46] G. Mills and M. Krstic, "Constrained extremum seeking in 1 dimension," in *Proc. 53rd IEEE Conf. Decis. Control*, Los Angeles, CA, USA, Dec. 2014, pp. 2654–2659.

[47] M. Papageorgiou, H. Hadj-Salem, and J.-M. Blosseville, "ALINEA: A local feedback control law for on-ramp metering," *Transp. Res. Rec.*, vol. 1320, pp. 58–64, 1991.

[48] M. Papageorgiou, "A hierarchical control system for freeway traffic," *Transp. Res. B, Methodol.*, vol. 17, no. 3, pp. 251–261, Jun. 1983.

[49] M. Kontorinaki, A. Spiliopoulou, C. Roncoli, and M. Papageorgiou, "First-order traffic flow models incorporating capacity drop: Overview and real-data validation," *Transp. Res. B, Methodol.*, vol. 106, pp. 52–75, Dec. 2017.

[50] A. Messmer and M. Papageorgiou, "METANET: A macroscopic simulation program for motorway networks," *Traffic Eng. Control*, vol. 31, nos. 8–9, pp. 466–470, 1990.

[51] A. Bose and P. A. Ioannou, "Analysis of traffic flow with mixed manual and semiautomated vehicles," *IEEE Trans. Intell. Transp. Syst.*, vol. 4, no. 4, pp. 173–188, Dec. 2003.

[52] J. C. Herrera and A. M. Bayen, "Incorporation of lagrangian measurements in freeway traffic state estimation," *Transp. Res. B, Methodol.*, vol. 44, no. 4, pp. 460–481, May 2010.

[53] T. Seo, T. Kusakabe, and Y. Asakura, "Estimation of flow and density using probe vehicles with spacing measurement equipment," *Transp. Res. C, Emerg. Technol.*, vol. 53, pp. 134–150, Apr. 2015.

[54] N. Bekiaris-Liberis, C. Roncoli, and M. Papageorgiou, "Highway traffic state estimation per lane in the presence of connected vehicles," *Transp. Res. B, Methodol.*, vol. 106, pp. 1–28, Dec. 2017.

[55] S. Papadopoulou, C. Roncoli, N. Bekiaris-Liberis, I. Papamichail, and M. Papageorgiou, "Microscopic simulation-based validation of a per-lane traffic state estimation scheme for highways with connected vehicles," *Transp. Res. C, Emerg. Technol.*, vol. 86, pp. 441–452, Jan. 2018.

[56] Y. Wang, M. Papageorgiou, J. Gaffney, I. Papamichail, G. Rose, and W. Young, "Local ramp metering in random-location bottlenecks downstream of metered on-ramp," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 2178, no. 1, pp. 90–100, Jan. 2010.

[57] G.-R. Iordanidou, C. Roncoli, I. Papamichail, and M. Papageorgiou, "Feedback-based mainstream traffic flow control for multiple bottlenecks on motorways," *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 2, pp. 610–621, Apr. 2015.

Farzam Tajdari received the B.S. degree in mechanical engineering from the Khaje Nasir Toosi University of Technology (KNTU) in 2013 and the master's degree in mechanical engineering from the Amirkabir University of Technology, Tehran, Iran, in 2016. He is currently pursuing the Ph.D. degree with Aalto University, Finland. He is working on autonomous and connected vehicles. His research interests include control, robotics, optimization, and control of traffic systems.

Claudio Roncoli received the Ph.D. degree from the University of Genoa, Italy, in 2013. He is currently an Assistant Professor of transportation engineering with Aalto University, Finland. Before joining Aalto University, he was a Research Assistant with the University of Genoa, a Visiting Research Assistant with Imperial College London, U.K., and a Post-Doctoral Researcher with the Technical University of Crete, Greece. He has been involved in several national and international research projects as the principal investigator. His research interests include real-time traffic management, modeling, optimisation, and control of traffic systems with connected and automated vehicles, and smart mobility and intelligent transportation systems.

Markos Papageorgiou (Life Fellow, IEEE) was a Professor of automation with the Technical University of Munich from 1988 to 1994. Since 1994, he has been a Professor with the Technical University of Crete. He was a Visiting Professor with the Politecnico di Milano, the École Nationale des Ponts et Chaussées, MIT, the Sapienza University of Rome, and Tsinghua University, and a Visiting Scholar with UC Berkeley. His research interests include automatic control and optimisation theory and applications to traffic and transportation systems, water systems, and further areas. He is a fellow of IFAC. He received several distinctions and awards, including the 2020 IEEE Transportation Award and two ERC Advanced Investigator Grants.