A feedback control strategy is proposed for lane assignment at bottleneck locations. The strategy assumes that some vehicles equipped with vehicle automation and communication systems are capable of receiving and executing specific lane-changing orders or recommendations. From a previously proposed optimal control strategy based on a simplified multilane motorway traffic flow model and formulated as a linear quadratic regulator, a feedback control problem was designed. The aim was to maximize the throughput at bottleneck locations while distributing the total density at the bottleneck area over the lanes according to a given policy by optimal lane assignment of the vehicles upstream of the bottleneck. The feedback control decisions were based on real-time measurements of the traffic state and inflow. The proposed strategy was tested on a nonlinear first-order macroscopic multilane traffic flow model, which also accounted for the capacity drop phenomenon.

In the near future, vehicle automation and communication systems (VACS) are expected to revolutionize the features and capabilities of individual vehicles. Of the wide range of potentially introduced VACS, some may be exploited to interfere with driving behavior via recommending, supporting, or even executing appropriately designed traffic control tasks, providing unprecedented opportunities to improve traffic control performance (I). However, the uncertainty about the future development of VACS calls for the design of control strategies that are robust with respect to the different system types, as well as to their penetration rate. A promising new feature that can be exploited for traffic management is lane-changing control.

The problem of modeling 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) (2–8). However, this choice is also behavioral 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 (9, 10). Last, 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 are equipped with VACS having vehicle-to-infrastructure 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.

The problem of assigning traffic flow over lanes for motorways under fully automated or semiautomated driving has been studied in research during the past decades. To tackle the high complexity of the problem, several assumptions are typically made, including the known and constant prevailing speeds along the motorway and the absence of traffic congestion, thanks to the assumed (but not addressed) appropriate operation of other control actions (e.g., ramp metering) at the motorway entrances. In addition, structural assumptions are commonly considered to limit the (otherwise vast) space of potential path assignments. In his seminal work, Varaiya proposed a hierarchical framework for a fully automated motorway, in which the decisions on the lane-changing behavior of vehicles are addressed in the link layer, which consists of a set of parallel decentralized link controllers, each addressing a corresponding motorway link (of about 2 km in length) (11). Following this framework, several strategies have been proposed to solve the problem of lane assignment in the link layer; strategies include designing control methods suitable for real-time applications, including the definition of well-justified and structured heuristic rules (12); implementing lane-routing algorithms (13); and defining control laws to stabilize traffic conditions (14). Optimization methods for path planning through lanes have been developed, but the computation complexity of the proposed optimization problems makes them hardly applicable in a real-time context (15–18). Lane-changing control has also been considered, together with variable speed limits and ramp metering in integrated traffic management strategies (19–21).

Recently, a combined lane-changing and variable speed limits control strategy was developed by Zhang and Ioannou, with the purpose of avoiding lane changes in the immediate proximity of a bottleneck, which especially in the case of heavy vehicles, may lead to premature triggering of congestion (10). In particular, lane-changing commands delivered as recommendations to drivers are defined according to a set of case-specific rules. Schakel and van Arem proposed a system that aims at an optimal lane distribution in high flow conditions by sending advice on lane, speed, and headway to
vehicles equipped with an in-car advisory system (22). The advice is determined at a traffic management center, based on a newly proposed lane level traffic state prediction model. Furthermore, Guétau et al. proposed a multiagent decentralized framework with the aim of performing cooperative lane-changing tasks according to information exchange between vehicles and a roadside unit located in the proximity of a bottleneck (23).

An optimal feedback control strategy formulated as a linear quadratic regulator was recently proposed in Roncoli et al. (24). The solution is applied in the form of a linear state-feedback control law, which is highly efficient in real time even for large-scale networks. Different from other approaches, this strategy is based on a rigorous application of optimal control theory and does not involve the definition of heuristic rules. The control strategy aims at regulating the lane assignment of vehicles upstream of a bottleneck location so as to maximize the bottleneck throughput, targeting critical densities at bottleneck locations as set points. However, as a result, the traffic density distribution over different lanes may remain (roughly) constant under any demand scenario. Although this behavior would not produce any negative effect on the traffic performance, it may in some circumstances be undesirable. As an example, consider a two-lane motorway in which the two lanes have the same characteristics (i.e., same critical densities); targeting critical densities as set points would result in equal flows in both lanes for any traffic situation. This behavior is not permitted, for example, in European motorways, where vehicles are obliged to travel in the rightmost (for right-hand traffic) available lane, while overtaking is allowed only on the left side. For North American freeways this issue is less crucial since vehicle overtaking is allowed on any lane; however, also in this case, traffic authorities may for various reasons, prefer different specific lane distributions. To incorporate this feature, a method is proposed here that does not always aim at tracking the critical density, but through opportunely defined functions, it allows distribution of the total density at a bottleneck area over the lanes according to a given policy.

The rest of the paper is organized as follows. The control design framework for multilane motorways proposed in Roncoli et al. is presented first (24). Then, the control problem is reformulated, and a feedback control law is designed to achieve a different traffic density distribution for the various lanes at the bottleneck area. Simulation experiments using a first-order macroscopic traffic flow model featuring the capacity drop phenomenon are then presented to evaluate the effectiveness of the developed method and to highlight the different traffic behaviors in regard to flow distribution. The main results of the paper are highlighted in the concluding section, and further research challenges are proposed.

LANE-CHANGING-BASED OPTIMAL CONTROL OF MULTILINE MOTORWAYS AT BOTTLENECKS

Bottlenecks in Motorways

A motorway bottleneck is a location where the flow capacity upstream of the bottleneck location is higher than the flow capacity downstream. Bottleneck locations can be lane drops, merge areas, and zones with a particular infrastructure layout (e.g., strong grade or curvature and tunnels) or with external capacity-reducing events (e.g., work zones and incidents). The nominal bottleneck capacity is the maximum traffic flow that can be maintained at the bottleneck location if the traffic flow arriving from upstream is smaller than (or equal to) the bottleneck capacity. However, if the arriving flow is higher than the capacity or the lane-changing behavior leads to exceeding the capacity of at least one lane, the bottleneck is activated, generating congestion starting at the bottleneck location and spilling back for as long as the upstream arriving flow is sufficiently high. Empirical observations show that whenever a bottleneck is activated, the maximum outflow that materializes (also called discharge flow) may be some 5% to 20% lower than the nominal bottleneck capacity; the difference between these two values of flow is called the capacity drop (25, 26). To avoid or delay the activation of a bottleneck and the related capacity drop phenomenon, various traffic control measures have been proposed and applied (27). In this work, it is assumed that the proposed control strategy operates simultaneously with some other controller [e.g., ramp metering (28) or mainstream traffic flow control (29)] that guarantees that the flow approaching the bottleneck area does not exceed the overall capacity of the bottleneck and, therefore, assuming an appropriate operation of the proposed lane-changing controller, traffic congestion may be completely avoided.

Linear Multilane Traffic Flow Model

Consider a multilane motorway that is subdivided into \( i = 0, \ldots, N \) segments of length \( L \), while each segment is composed of \( j = m, \ldots, M \) lanes, where \( m \) and \( M \) are the minimum and maximum indexes of lanes for segment \( i \). Each element of the resulting grid is denoted as a cell (see Figure 1), which is indexed by \((i, j)\). The model is formulated in discrete time, considering the discrete time step \( T \), indexed by \( k = 0, 1, \ldots \), where the time is \( t = kT \). To account for any possible network topology, including lane drops and lane additions on the right and on the left sides of the motorway, it is assumed that \( j = 0 \) corresponds to the segment(s) including the rightmost lane. Consequently, \( m \) and \( M \) are defined as the minimum and maximum indexes \( j \), respectively, for which a lane exists in segment \( i \). For

![FIGURE 1 Hypothetical motorway stretch.](image)
example, looking at the hypothetical motorway stretch depicted in Figure 1, $m_0 = 0$ and $M_b = 4$, while $m_1 = 1$ and $M_1 = 3$. According to that definition, the total number of cells from the origin to segment $i$ is $H_i = \sum_{i=0}^{i}(M_i - m_i + 1)$, and the total number of cells for the whole stretch is $H = H_N$.

Each motorway cell $(i, j)$ is characterized by the traffic density $\rho_{ij}(k)$, defined as the number of vehicles present in the cell at time instant $k$ divided by $L_i$. Density dynamically evolves according to the following conservation law equation [see, e.g., Roncoli et al. (30)]:

$$\rho_{ij}(k + 1) = \rho_{ij}(k) + \frac{T}{L_i} \left[ q_{i,j+1}(k) - q_{i,j}(k) \right]$$

$$+ \frac{T}{L_i} \left[ f_{i,j+1}(k) - f_{i,j}(k) \right] + \frac{T}{L_i} d_{ij}(k)$$

(1)

where

$q_{i,j}(k)$ = longitudinal flow leaving cell $(i, j)$ and entering cell $(i + 1, j)$ during time interval $(k, k + 1]$,

$f_{i,j}(k)$ = net lateral flow moving from cell $(i, j)$ to cell $(i, j + 1)$ during time interval $(k, k + 1]$,

and

$d_{ij}(k)$ = external flow entering network cell $(i, j)$, either from mainstream or from on-ramp, during time interval $(k, k + 1]$.

Depending on the network topology, some terms of Equation 1 may not be present. In particular, the inflow $q_{i,j}(k)$ 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 term $f_{i,j}(k)$ exists only for $m_i \leq j < M_i$. Following previous considerations, the total number of lateral flow terms is $F = H - N$.

To guarantee numerical stability (since the discrete-time system described by Equation 1 may come from a discretization of a partial differential equation (31)), the time step $T$ must respect the so-called Courant–Friedrichs–Lewy CFL condition (32):

$$T \leq \min_{i,j} \frac{L_i}{v_{ij}^{max}}$$

(2)

where $v_{ij}^{max}$ is the maximum speed allowed in cell $(i, j)$.

Similar modeling approaches of multilane motorway traffic are also considered in Roncoli et al. (30), Mujal and Pipes (33), and Michalopoulos et al. (34). The net lateral flow $f_{ij}(k)$ is considered only in one direction, namely, from the right to left lanes; therefore, $f_{ij}(k)$ is actually the difference between the flow leaving and entering lane $j$ at its left side. This simplification is useful for the subsequent control problem formulation since lateral flows are treated as control inputs.

Consider the well-known relationship

$$q_{ij}(k) = \rho_{ij}(k) v_{ij}(k)$$

(3)

With Equation 3 replaced into Equation 1, the following is obtained:

$$\rho_{ij}(k + 1) = \left[ 1 - \frac{T}{L_i} v_{ij}(k) \right] \rho_{ij}(k)$$

$$+ \frac{T}{L_i} \left[ f_{i,j+1}(k) - f_{i,j}(k) \right] + \frac{T}{L_i} d_{ij}(k)$$

(4)

which treating speeds $v_{ij}(k)$ as known parameters, can be seen as a linear parameter varying system in the form

$$z(k + 1) = A(k)z(k) + Bu(k) + d(k)$$

(5)

where

$$x \equiv [\rho_{i,0} \cdots \rho_{i,M_b} \rho_{M_b,1} \cdots \rho_{N,M_b}]^T \in \mathbb{R}^n$$

$$u \equiv [f_{i,0} \cdots f_{i,M_b} f_{M_b,1} \cdots f_{N,M_b}]^T \in \mathbb{R}^d$$

$$d \equiv \left[ \frac{T}{L_i} d_{i,0} \cdots \frac{T}{L_i} d_{i,M_b} \frac{T}{L_i} d_{M_b,1} \cdots \frac{T}{L_i} d_{N,M_b} \right]^T \in \mathbb{R}^n$$

(6)

(7)

(8)

and time index $k$ is omitted to simplify the notation. $A \in \mathbb{R}^{n \times n}$, composed of elements $a_{sr}$, which represent the connection between pairs of subsequent cells connected by a longitudinal flow, and $B \in \mathbb{R}^{n \times d}$, composed of elements $b_{rs}$, which reflects the connection of adjacent cells connected by lateral flows, are defined as

$$a_{sr} = \begin{cases} 1 & \text{if } r = s \text{ and } (j < m_i \text{ or } j > M_i) \\ 1 - \frac{T}{L_i} v_{ij} & \text{if } r = s \text{ and } (i = N \text{ or } m_i \leq j \leq M_i) \\ \frac{T}{L_i} v_{i,j+1} & \text{if } r > H_i \text{ and } s = r - M_i - m_i + 1 \\ 0 & \text{otherwise} \end{cases}$$

(9)

$$b_{rs} = \begin{cases} T & \text{if } j > m_i \text{ and } s = r - i \\ \frac{T}{L_i} & \text{if } j < M_i \text{ and } s = r - i + 1 \\ 0 & \text{otherwise} \end{cases}$$

(10)

where $r = \sum_{i=1}^{i} H_i + j - m_i$.

### Optimal Control Problem Formulation with Constant Set Points

The linear system described in the previous section is used for formulating an optimal control problem with the purpose of manipulating the lateral flows to avoid creating congestion resulting from the activation of a bottleneck. Under the assumptions that the overall traffic flow entering the controlled area does not significantly exceed the bottleneck capacity and that the controller succeeds in avoiding the creation of congestion, one can assume that the speed in all cells remains at a constant value (e.g., the free-flow speed) $v_{ij}(k) \equiv \bar{v}$, $\forall i, j, k$. In addition, one can assume that the measurable inflows $d$ are constant; the actual slow time variation of $d$ will not affect the control performance significantly. With these assumptions, the system in Equation 5 can be viewed as a linear time invariant system

$$z(k + 1) = A(k)z(k) + Bu(k) + d(k)$$

Identifying the nominal capacity of a bottleneck is a nontrivial task. In fact, Elefteriadou et al. (35) and Lorenz and Elefteriadou (36) have demonstrated that the real flow capacity 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 prespecified capacity flow value may lead either to overload and congestion (on days in which the real capacity happens to be lower than its prespecified target value) or to underutilization of the
infrastructure (on days in which the real capacity happens to be higher than its prespecified target value). However, the critical density at which capacity flow occurs exhibits smaller variations, and it is therefore preferable to target a density set point (i.e., the critical density) at the bottleneck location (37). In Roncoli et al. (24), a control strategy was proposed that always targets the critical densities for each lane, and if they are unknown, an extremum-seeking algorithm was proposed to estimate them (38).

The following quadratic cost function (over an infinite time horizon) that accounts for the penalization of the difference between some (targeted) densities and the corresponding prespecified (assumed constant) set point values is defined, as well as a penalty term aiming at maintaining small control inputs, that is, small lateral flows (weighted by $\varphi$):

$$J = \sum_{k=0}^{\infty} \left\{ \sum_{i} \sum_{j} \alpha_{i,j} \left[ \bar{\rho}_{ij}(k) - \bar{\rho}_{ij} \right]^2 + \varphi \sum_{i,j} \sum_{r,j} \left[ f_{ir}(k) \right]^2 \right\}$$

(12)

where

$$(i,j) = \text{targeted cells}, \quad \bar{\rho}_{ij} = \text{desired set point}, \quad \alpha_{i,j} = \text{corresponding weighting parameter}.$

Equation 12 is rewritten in matrix form as

$$J = \sum_{k=0}^{\infty} \left\{ \left[ C_2(k) - \bar{\gamma} \right]^T Q \left[ C_2(k) - \bar{\gamma} \right] + \bar{\gamma}^T (k) R \bar{\gamma}(k) \right\}$$

(13)

where $Q = Q^T \geq 0$ and $\varphi = \varphi I > 0$ are weighting matrices associated with the magnitude of the state tracking error and control actions, respectively, while $C$, composed of elements $c_{rs}(k)$, where

$$c_{rs} = \begin{cases} 1 & \text{if density is tracked} \\ 0 & \text{otherwise} \end{cases}$$

(14)

reflects the cells that are tracked and $\bar{\gamma}$ is a vector that contains the desired set points. At first, only cells at the bottleneck locations (e.g., in Figure 1, $\bar{\rho}_{11}, \bar{\rho}_{21}$) are targeted.

The problem, defined as the minimization of the cost in Equation 13 subject to the linear dynamics in Equation 11, is solved through a linear quadratic regulator, under the assumption that the original system is, at least, stabilizable and detectable [see Lewis et al., chap. 2 (39)]. As shown in Roncoli et al., stabilizability is guaranteed for any network configuration, while to guarantee detectability, it is necessary to control the density of each cell that does not have any other cell downstream (24). To account for this issue, an additional dummy cell is placed immediately downstream of each lane drop, imposing it with an appropriate high penalty weight $\alpha_{ij}$ to have a density equal to zero. In the described case, the system is also observable. Further details are presented in Roncoli et al. (24).

The solution to the proposed linear quadratic regulator problem obtained via dynamic programming in Roncoli et al. results in the following feedback–feedforward control law (24):

$$u^*(k) = -K_{2}(k) + u_{\bar{\gamma}}$$

(15)

where

$$K = (R + B^T PB)^{-1} B^T PA$$

(16)

$$P = C^T QC + A^T PA - A^T PB (R + B^T PB)^{-1} B^T PA$$

(17)

$$u_{\bar{\gamma}} = K_{2} \bar{\gamma} + K_{d} d$$

(18)

$$K_{2} = (R + B^T PB)^{-1} B^T (I - (A - BK)^T)^{-1} C^T Q$$

(19)

$$K_{d} = -(R + B^T PB)^{-1} B^T (I - (A - BK)^T)^{-1} P$$

(20)

The optimal gain computed in Equation 16 and the algebraic Riccati equation computed in Equation 17 are the same as can be found in classic optimal control books [see, e.g., Anderson and Moore (40)]. Several methods have been proposed to efficiently compute the solution of the algebraic Riccati equation [see, e.g., Anderson and Moore (40) and Arnold and Laub (41)]. Also, for practical implementation, one may allow for the (measured) inflow $d$ to be time varying, in which case the feedforward term $u_{\bar{\gamma}}$ in Equation 15 also becomes time varying, obtaining (instead of Equations 15 and 18)

$$u^*(k) = -K_{2}(k) + u_{\bar{\gamma}}(k)$$

(21)

$$u_{\bar{\gamma}}(k) = K_{2} \bar{\gamma} + K_{d} d(k)$$

(22)

This corresponds to a model predictive control procedure, whereby the future inflow values are predicted to be equal to their current (measured) values.

The proposed feedback–feedforward control law is very effective for practical application since the computation of the feedback gain matrix $K$ and of $K_{2}$ and $K_{d}$ is effectuated only once, offline; whereas online calculations are limited to a few matrix–vector multiplications, as evidenced by Equations 21 and 22.

A similar optimal regulation problem, without guarantee of regulation to an a priori prescribed set point for state variables and non-zero mean disturbances, has also been considered in Haddad and Bernstein, in which a different formulation for the feedforward term is obtained (42). In fact, the solution to the optimal control problem in this paper is obtained with the dynamic programming principle, whereas Haddad and Bernstein use Lagrange multipliers (42). Although it is cumbersome to compare the two control laws analytically, they produce the same results in all tested examples presented in this paper.

The implementation of lane-changing actions may not be trivial in practice, even if all vehicles are connected with the control center. These actions can be implemented by sending lane-changing recommendations to an appropriate number of selected vehicles; the selection may be based on the known destinations of the vehicles and further criteria. Since for the foreseeable future, the lane change advice will not be mandatory, the assignment will have to account for the compliance rate, as well as for other, spontaneous lane changes decided by the drivers. The latter may be reduced by giving additional “keep-lane” advice to all equipped vehicles that do not receive lane-change advice. Clearly, any mismatch between the optimal lateral flows and the actually triggered lane changes may be partially compensated thanks to the feedback nature of the proposed controller.

Feedback Control Strategy for Density Distribution at Bottlenecks

An extended control strategy is proposed here that besides aiming at tracking the critical density (e.g., when demand is close to bottleneck...
capacity) aims at distributing the vehicles at the bottleneck area over the lanes according to a given policy.

To achieve that end, the control law is modified by choosing a time-varying set point \( \hat{y} \) as a function of the network inflow: \( \hat{y}(k) = \psi(d(k)) \); where the function \( \psi \) defines the pursued lane distribution policy. Thus, the feedback–feedforward control law in Equation 21 is maintained; however, the feedforward term of Equation 22 is replaced by

\[
I_u(k) = K_y \psi(d(k)) + K_d d(k) \tag{23}
\]

As an example, Figure 2 shows possible functions for defining the set points for the left (\( \hat{y}_L \)) and right (\( \hat{y}_R \)) lanes of a two-lane motorway. In this example, the authors impose that for low total inflow \( d_{in} \) entering the motorway network, a higher amount of traffic is assigned to the left lane by choosing \( \hat{y}_L = \rho_L^0 \) and \( \hat{y}_R = \rho_R^0 \) for \( d_{in} \leq \bar{d}_{sat} \). As a result, a higher outflow is expected from the left lane when the incoming demand is lower than \( \bar{d}_{sat} \), while the two lanes should simultaneously reach their capacity (i.e., operating at their critical densities) when the overall demand approaches the bottleneck capacity. Notably, the proposed controller is capable of achieving a desired distribution of traffic on the basis of any given functions, which would reflect different distribution policies. A constraint to be considered while defining such functions is that to obtain the best traffic performance, the (per-lane) density set points should be equal to the (per-lane) critical densities when the inflow approaches the bottleneck capacity.

As an alternative, the set point \( \hat{y}(k) \) may be varied via a total-density-dependent term \( \chi(p_{out}(k))p_{out}(k) \), where \( \chi \) is an opportunistically defined function and \( p_{out}(k) \) is the total (measured) density at the bottleneck area. In this case, \( \chi \) holds the portions of the total current density assigned to the corresponding lanes. The involvement of \( p_{out}(k) \), factually, to an additional (outer) feedback loop, which however has virtually no impact on the overall system stability, as numerical investigations have shown.

Finally, all proposed controllers are in the form of state-feedback regulators, which require availability of measurements for all state variables (densities for each cell) in real time. In the case of incomplete measurements, one may use a traffic state estimator to produce the missing measurements; in the context of connected vehicles, promising approaches are found in other research (43–46).

**SIMULATION EXPERIMENTS**

**Nonlinear Multilane Traffic Flow Model**

Presented next is the performance evaluation of the proposed control strategies based on simulation experiments using a first-order traffic flow model based on Ronconi et al. (30). The model is used for reproducing the traffic behavior for a lane motorway, and it features (a) nonlinear functions for the lateral flows of manually driven vehicles, (b) a cell transmission model–like formulation for the longitudinal flows, and (c) a nonlinear formulation to account for the capacity drop phenomenon (31). For completeness, a brief explanation is provided here of the model used.

Consider the conservation law described in Equation 1. Lateral flows owing to manual lane changing on adjacent lanes of the same segment are considered, and corresponding rules are defined to properly assign and bound their values. The net lateral flows are computed as follows:

\[
f_{ij}(k) = l_{ij+1}(k) - l_{ij-1}(k) \tag{24}
\]

where \( l_{ij}(k) \) is the lateral flow moving from cell \((i, j)\) to cell \((i, j')\) during time interval \([k, k + 1]\) and \( j' = j ± 1 \). Lateral flows \( l_{ij}(k) \) are computed according to the following:

\[
l_{ij}(k) = \min \left\{ \left. \frac{1}{D_{ij-1}(k) + D_{ij+1}(k)} ; D_{ij}(k) \right| \left. \psi_j(k) \right\}
\]

\[
S_j(k) = \frac{\bar{D}_j}{T} \left[ \rho_j^m - \psi_j(k) \right] \tag{26}
\]

\[
D_{ij}(k) = \frac{\bar{D}_j}{T} \psi_j(k) A_{ij}(k) \tag{27}
\]

\[
A_{ij}(k) = \mu \max \left\{ 0, \frac{P_{ij}^+(k) \psi_j(k) - \psi_j(k)}{P_{ij}^-(k) \psi_j(k) + \psi_j(k)} \right\} \tag{28}
\]

Equation 25 accounts for the potentially limited space that may not be sufficient for accepting the lateral flow entering from both sides of a cell, where \( S \) is the available space in regard to flow acceptance, and \( D \) is the lateral demand flow, which is computed via the definition of the attractiveness rate \( A \). The attractiveness rate is computed as a function of the densities for each pair of adjacent lanes; the factor \( P \) affects the distribution of vehicles over the lanes and should be calibrated to achieve the desired behavior, for example, with the use of real data as in Ronconi et al. (47). Choosing a value \( P = 1 \) implies that drivers always move toward a faster lane (leading also to equal densities over the lanes), but \( P \) may also be tuned to reflect particular location-dependent effects in which lateral flow may occur in the direction from a lower to a higher density (e.g., upstream of on- and off-ramps and lane-drop locations). Finally, parameter \( \mu \) is a constant coefficient in the range \([0, 1]\) reflecting the “aggressiveness” in lane changing.

Longitudinal flows are the flows generated in a segment that move to the next downstream segment while remaining in the same lane. A Godunov-discretized scheme similar to the scheme proposed in Ronconi et al. (30) is used, but it uses the nonlinear exponential function proposed in Messmer and Papageorgiou (48) to obtain a more realistic behavior at undercritical densities. The model also accounts for the capacity drop phenomenon via a linearly decreasing demand.
function for overcritical densities. Other modeling approaches can be used to improve the capability of reproducing capacity drop, obtaining comparable results [see, e.g., Han et al. (49) and Kontorinaki et al. (30)]. More details and calibration results related to this model are presented in Roncoli et al. (30, 47). Formally, the complete formulation for longitudinal flows reads

\[
q_{ij}(k) = \min \{ Q_{ij}^*(k), Q_{ij}^{m}(k) - d_{ij}(k) \} \tag{29}
\]

where

\[
Q_{ij}^{m}(k) = \begin{cases} 
v_{ij}^{\text{max}} \exp \left[ -\frac{\alpha}{\rho_{ij}^{*}} \left( \frac{\rho_{ij}(k)}{\rho_{ij}^{*}} \right) \right] \rho_{ij}(k) & \text{if } \rho_{ij}(k) < \rho_{ij}^{*} \\
\left( 1 - \gamma \right) \frac{Q_{ij}^{c}}{\rho_{ij}^{*}} + \gamma Q_{ij}^{c} & \text{otherwise}
\end{cases}
\]

\[
Q_{ij}^{c}(k) = \begin{cases} 
Q_{ij}^{c} & \text{if } \rho_{ij+1}(k) < \rho_{ij+1}^{*} \\
\rho_{ij}^{m} \left( N_{ij} - \rho_{ij}(k) \right) & \text{otherwise}
\end{cases}
\]

\[v_{ij}^{\text{max}} \text{ free speed,} \]
\[Q_{ij}^{c} \text{ capacity flow,} \]
\[\rho_{ij}^{*} \text{ critical density (i.e., density at which capacity flow occurs),} \]
\[\gamma \text{ capacity drop coefficient in [0,1], and} \]
\[\alpha = \left( \ln \frac{Q_{ij}^{c}}{v_{ij}^{\text{max}} \rho_{ij}^{*}} \right)^{-1} \tag{48}.\]

**Network Description and the No-Control Case**

A hypothetical motorway stretch is considered to test and evaluate the performance of the proposed strategy. In particular, consider the network depicted in Figure 3; the network is composed of seven segments. Segments 1, . . ., 5 feature three lanes, while Segments 6 and 7 feature only two, with a lane drop located downstream of cell (5, 1). All segments are characterized by a length of \( L_i = 0.5 \) km, and a simulation step is defined as \( T = 10 \) s. Different lanes feature different parameters, specifically a different fundamental diagram, which may reflect a different traffic composition (e.g., a higher rate of heavy vehicles reducing the capacity of a specific lane); the values used are shown in Table 1.

Traffic demand profiles are defined for a simulation horizon \( K = 480 \) (80 min), as shown in Figure 4. The overall demand entering the network is, at its peak, roughly equivalent to the total capacity of Segment 5, that is, the bottleneck capacity.

Running the macroscopic model described by Equations 1 and 24 through 31 without the use of any control actions eventually produces traffic congestion starting at the lane-drop area, as a result of the non-optimal spontaneous lane changes of the vehicles. Looking at the contour plots shown in Figure 5a, one can see that the density increases first in Lane 1 (the lane that is dropping) at about \( t = 20 \) because of the high demand arriving in the lane-drop area, while vehicles try to merge first into Lane 2 and, because density also increases in this lane, eventually also into Lane 3. In particular, most lane changes take place in Segments 4 and 5, while a small number of lane changes take place in Segment 6; there are virtually no lane changes in the upstream segments (see Figure 6a). Recall that according to Equation 28, with \( P_{i,j} = 1 \), the lane-changing model acts toward the homogenization of the densities between adjacent lanes. The detrimental effects of the congestion worsen as a consequence of the capacity drop that occurs; that drop is triggered here by overcritical densities at both lanes of Segment 5, causing a reduction in the outflow in both lanes during the high-demand period, as shown in Figure 7a.

The congestion created spills back, covering all lanes of Segments 4 and 5 (see Figure 5). The total travel time (TTT) over a finite time horizon \( K \) is used as the numerical evaluation criterion. It is defined as in Papageorgiou et al. as follows (28):

\[
\text{TTT} = T \sum_{k=0}^{K} \sum_{i=0}^{N} L_i \sum_{j=0}^{M-1} \rho_{ij}(k) \tag{32}
\]

obtaining for the presented no-control case a resulting overall \( \text{TTT} = 186.7 \text{ veh} \cdot \text{h} \).

**Application of Control Strategy with Constant Set Points**

Now, the optimal control strategy with constant set points is evaluated with the use of the previously described motorway scenario.

![FIGURE 3 Motorway stretch used for testing and evaluating proposed control strategy.](image-url)
FIGURE 4  Traffic demand $d$ entering Segment 1 (see Figure 3).

FIGURE 5  Contour plots of densities (a) in no-control case, (b) when control strategy with constant set points is applied, and (c) when proposed feedback control strategy for density distribution is applied (Lane 1 (left); Lane 2 (center), and Lane 3 (right)).
FIGURE 6  Contour plots of net lateral flows (a) in no-control case, (b) when control strategy with constant set points is applied, and (c) when proposed feedback control strategy for density distribution is applied (Lane 1 to 2 (left) and Lane 2 to 3 (right)).
The "application area," namely, the portion of the network where the designed strategy is applied, is defined as the area from Segment 3 to Segment 6 (see Figure 3). The outflow of the segments immediately upstream of the application area $q_{ij}$ is used as demand $d_i$. A dummy cell $(6, 1)$ is added immediately downstream of the lane drop to ensure system observability. The set point considered in the linear quadratic regulator thus includes the three cells in Segment 6.

According to the network topology and setting a constant speed $\bar{v} = 90$ km/h and cost weights $Q_{ii} = 1$ for $i = j = 2, 3, Q_{ij} = 100$ for $i = j = 1, Q_{ij} = 0, \forall i \neq j$, and $\varphi = 10^{-5}$ (obtained after some manual tuning of the controller aiming at achieving an efficient and smooth response), the gains according to Equations 16, 17, 19, and 20 are computed (offline).

Assuming that the critical densities at the controlled area are known, the set point vector $\hat{y}$ is built to consist of $\hat{\rho}_{6,2} = 32$ veh/km and $\hat{\rho}_{6,3} = 36$ veh/km, while for the additional dummy segment one defines $\hat{\rho}_{4,1} = 0$ veh/km.

Lateral flows $f_{ij}$ are computed as $y^*$, via the control law (Equation 21) and are then applied directly in the conservation law (Equation 1) of the simulation model, while longitudinal flows $q_{ij}$ are obtained from Equations 29 through 31 as in the no-control case.

From inspection of the resulting contour plots in Figure 5b, one can see that the controller is capable of avoiding the creation of congestion. This finding is the result of the fact that during the period characterized by high demand, the density at the bottleneck area is maintained at its critical value.

The optimal lateral flows are distributed quite homogeneously in the whole application area (see Figure 6b), thus avoiding high lane-changing flows close to the lane-drop location. Moreover, since all densities remain undercritical, the capacity drop phenomenon does not appear, and the system operates at the bottleneck capacity during the whole peak period (see Figure 7b). In this scenario, one obtains a TTT = 145.7 veh · h, which is a 22% improvement with respect to the no-control case.
However, as can be seen from Figure 8 (a), at the bottleneck area, the flow exiting Lane 3 is always higher than the flow exiting Lane 2, for any value of total flow. This finding is a result of the higher value of critical density used as a constant set point in the application of this control strategy.

Application of Proposed Feedback Control Strategy for Density Distribution at Bottlenecks

Now, the proposed control strategy aiming at distributing the total density at a bottleneck area over the lanes according to a given policy is tested. The set point vector \( \hat{y}(k) \) is computed via the functions depicted in Figure 2 with the use of a quadratic form for \( \hat{\rho}_{6,2}(k) \) and a linear term for \( \hat{\rho}_{6,3}(k) \) according to

\[
\hat{\rho}_{6,2}(k) = \begin{cases} 
\frac{1}{\nu d_{\text{tot}}} \left[ d_{\text{tot}}(k) \right]^2 + \frac{\nu \hat{\rho}_{6,2} + \hat{d}_{\text{tot}}}{\nu d_{\text{tot}}} & \text{if } d_{\text{tot}}(k) \leq \hat{d}_{\text{tot}} \\
\hat{\rho}_{6,2} & \text{otherwise}
\end{cases}
\]

(33)

\[
\hat{\rho}_{6,3}(k) = \begin{cases} 
\hat{\rho}_{6,3} & \text{if } d_{\text{tot}}(k) \leq \hat{d}_{\text{tot}} \\
\hat{\rho}_{6,3} & \text{otherwise}
\end{cases}
\]

(34)

where

\[
\hat{d}_{\text{tot}} = \frac{4}{5} d_{\text{cap}}
\]

(35)

The same configuration is maintained for the controlled system as in the previous case, computing the lateral flow as \( u^* \) via the feedback–feedforward control law in Equation 21, using Equation 23, however, to compute the feedforward term.

Similar to the previous case, the resulting contour plots in Figure 5c illustrate that the controller also avoids congestion and hence the capacity drop phenomenon during the whole peak period (see Figure 7c), while lateral flows are distributed quite homogeneously in the whole application (see Figure 6c). For this scenario, one obtains a TTT = 146.7 veh·h, which is a 21.4% improvement with respect to the no-control case.

In this case, however, one can see from Figure 8b that at the bottleneck area, the flow exiting Lane 2 is higher than the flow exiting Lane 3 for lower values of total flow (i.e., when the total flow is lower than about 3,500 veh/h); whereas, for higher values of total flow, the flow in Lane 3 exceeds the flow in Lane 2 until capacity flow is reached simultaneously. In Figure 8b, there are three equilibrium values (circled) for outflows at each lane, which can be identified as areas in which the marks appear thicker, which are representative of the respective periods of simulation characterized by low, intermediate, and high traffic demand (see Figure 4). The observed behavior is in full accordance with the goals of the policy used for lane distribution.

CONCLUSIONS

This paper presents an extended version of an optimal control strategy for lane-changing-based traffic control at bottleneck locations (previously proposed in Roncoli et al.) by including together with the capability to operate a motorway traffic system at its capacity, the possibility to distribute the traffic over the lanes at the bottleneck area according to a given policy (24). Simulation results demonstrate the effectiveness of the proposed control strategy in improving traffic performance, while also pursuing a prescribed lane flow distribution at the bottleneck area.

This method is currently being extended to account for unmeasured demand flows and incomplete measurements, as well as to incorporate a mainstream or ramp flow control strategy. Moreover, the case of mixed traffic, in which manual vehicles may not receive or may not follow the prescribed lane-changing commands, is being examined.

ACKNOWLEDGMENT

The research leading to these results received funding from the European Research Council under the European Union’s Seventh Framework Programme, project TRAMAN21.


The Standing Committee on Vehicle-Highway Automation peer-reviewed this paper.