Connected and automated road vehicles: state of the art and future challenges

Tulga Ersal a, Ilya Kolmanovsky b, Neda Masoud c, Necmiye Ozay d, Jeffrey Scruggs c, Ram Vasudevan b, a and Gábor Orosz a, c

a Department of Mechanical Engineering, University of Michigan, Ann Arbor, MI, USA; b Department of Aerospace Engineering, University of Michigan, Ann Arbor, MI, USA; c Department of Civil and Environmental Engineering, University of Michigan, Ann Arbor, MI, USA; d Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, MI, USA

ABSTRACT
The state of the art of modelling, control, and optimisation is discussed for automated road vehicles that may utilise wireless vehicle-to-everything (V2X) connectivity. The appropriate tools to address safety and energy efficiency are described and the effects on traffic dynamics are highlighted. Finally, the economical and societal impacts of the deployment of connected and automated vehicles are discussed.

ARTICLE HISTORY
Received 27 August 2019
Revised 26 February 2020
Accepted 3 March 2020

1. Introduction
Over the past several decades the auto industry has been moving toward greater degrees of automation. Automated vehicles are able to monitor their state and their environment using a plethora of sensors, and they may also communicate with each other and the traffic infrastructure using wireless vehicle-to-everything (V2X) communication. Based on the information collected they can make decisions, plan their motion, and follow the plan utilising a set of sophisticated controllers.

While the prime objective of such enhanced autonomy is to maintain the safety of vehicles, the implications of the decisions at the automation level are more far-reaching. On the one hand, these decisions also influence the vehicle operation at the lower level (including engine, transmission, brakes, air conditioning), which, if taken into account properly at the automation level, may improve the vehicles’ energy efficiency. On the other hand, since automation allows vehicles to respond to traffic stimuli with better control accuracy and reduced reaction time compared to their human-driven counterparts, automated vehicles may have a positive impact on the traffic dynamics. These benefits can be enhanced by utilising V2X connectivity, which enables vehicles to obtain traffic information from beyond the line of sight. For example, by mitigating stop-and-go traffic jams, automated vehicles may increase the traffic flux and decrease the energy consumption and emissions at the system level. Finally, connected and automated vehicles shall influence the economy
Table 1. Summary of SAE’s levels of driving autonomy for road vehicles [1].

<table>
<thead>
<tr>
<th>SAE level</th>
<th>Name</th>
<th>Execution of steering and acceleration/deceleration</th>
<th>Monitoring of driving environment</th>
<th>Fallback performance of dynamic driving task</th>
<th>System capability (driving modes)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>No automation</td>
<td>Human driver</td>
<td>Human driver</td>
<td>Human driver</td>
<td>N/A</td>
</tr>
<tr>
<td>1</td>
<td>Driver assistance</td>
<td>Human driver and system</td>
<td>Human driver</td>
<td>Human driver</td>
<td>Some driving modes</td>
</tr>
<tr>
<td>2</td>
<td>Partial automation</td>
<td>System</td>
<td>System</td>
<td>Human driver</td>
<td>Some driving modes</td>
</tr>
<tr>
<td>3</td>
<td>Conditional automation</td>
<td>System</td>
<td>System</td>
<td>System</td>
<td>Some driving modes</td>
</tr>
<tr>
<td>4</td>
<td>High automation</td>
<td>System</td>
<td>System</td>
<td>System</td>
<td>Some driving modes</td>
</tr>
<tr>
<td>5</td>
<td>Full automation</td>
<td>System</td>
<td>System</td>
<td>System</td>
<td>All driving modes</td>
</tr>
</tbody>
</table>

and the society in the long run, and by having the appropriate business models for their deployment, the benefits may be significantly increased.

This paper discusses some of the latest results on connected and automated vehicles. Rather than aiming for a comprehensive review of the field, which would take much more volume, we highlight some of the important concepts with references to related literature. To help the reader to navigate further in this field, we include some review papers in the reference list that focus on some specific aspects (like perception or energy efficiency).

In Section 2, we start with describing the modelling, analysis, and design tools for controlling the dynamics of automated vehicles. Section 3 discusses safety verification of controllers for automated vehicles, while Section 4 focuses on powertrain control and energy efficiency of connected and automated vehicles. Section 5 is dedicated to traffic dynamics while assuming different levels of penetrations of connectivity and automation, while Section 6 addresses the large-scale economical and societal impacts under different deployment strategies. Finally, in Section 7 we list some challenges that one shall overcome in order to make connected and automated vehicles successful on future roadways.

Before starting the technical description, here we define our nomenclature utilised in the rest of the paper. When distinguishing between human-driven vehicles (HVs) and automated vehicles (AVs), we categorise levels 1–5 of AVs according to the SAE categorisation given in Table 1. When referring to wireless V2X connectivity, we include vehicle-to-vehicle (V2V), vehicle-to-infrastructure (V2I), and even vehicle-to-cloud (V2C) communication. These communications may be executed using wifi-based technologies [2], or cellular technologies [3], as long as they satisfy the low latency and low packet drop requirements. If a human-driven vehicle is equipped with V2X connectivity, we call it a connected human-driven vehicle (CHV), while automated vehicles with V2X connectivity are referred to as connected automated vehicles (CAVs) – without ‘and’ between connected and automated; see Figure 1.

2. Vehicle dynamics and control for automated vehicles

An automated vehicle builds up its situation awareness based on the data collected via sensors; this process is often referred to as perception. Then it makes decisions regarding its manoeuvres and executes those with high precision; see Figure 2. Such planning and control are typically based on mathematical models. Indeed, automated vehicles are
expected to outperform their human-driven counterparts, which demands models that allow the design of high-performance controllers that can be verified for safety. Note that data-driven methods (e.g. machine and reinforcement learning approaches) can also be combined with a priori available or learned online models. In this section, we list a few models that can be used to describe the longitudinal and lateral motion of automated vehicles, and we discuss motion planning and control design.

### 2.1. Perception and localisation for automated vehicles

Vehicle perception and localisation has gone through a tremendous development during the last few decades and particularly accelerated since the last DARPA Grand Challenge [4]. Utilising sensory data an automated vehicle can recognise objects in its neighbourhood and localise itself and the other objects on digital maps while simultaneously updating the maps. Typical sensors include GPS, cameras, radars, and lidars. Perception and localisation is a broad topic on its own and it is beyond the scope of this paper. We rather refer the reader to a recent review paper on this topic [5]. We mention, however, that V2X communication may also be considered as a sensor that can reach beyond the line of sight of optical sensors and may provide more accurate information especially when it comes to velocity, angular velocity, acceleration, and angular acceleration [6].
2.2. Vehicle dynamics for automated vehicles

Traditionally vehicle dynamics focuses on describing the velocity state of the vehicle and the corresponding models are derived for the velocity of the centre of mass and the angular velocity [7–10]. These vectors are typically resolved in a reference frame fixed to the car body, that is, the velocity components are obtained by projecting the vectors to the axes of this frame. These so-called longitudinal, lateral and vertical components can be obtained from wheel rotation sensors and inertial measurement units fixed to the vehicle. Moreover, the models are typically linearised around the straight running, constant speed motion assuming small speed variations and small steering angles. This allows for the separation of the dynamics into three groups: (i) longitudinal dynamics; (ii) bounce and pitch dynamics; and (iii) lateral, yaw and roll dynamics (often referred to as handling). Despite such simplification, these models are adequate to evaluate the vehicle dynamics for driver comfort and safety and also allow the development of vehicle control functions like anti-lock braking (ABS) and electronic stability control (ESC). Such functions have been developed during the last few decades of the previous century and they are now standard options on every production vehicle. In this paper, we focus on groups (i) and (iii), as these are primary concerns for automated driving design.

As mentioned in Section 2.1 an automated vehicle shall localise itself (and other road participants) in physical space and establish some contextual knowledge about the world around it. Thus, the corresponding models shall include position and orientation coordinates of the vehicle in an Earth-fixed frame. Moreover, for many manoeuvres the small angle and constant speed approximations may not hold. Finally, in some situations, it is necessary to control the lateral and the longitudinal dynamics of the vehicle in a coordinated fashion, which may require integration of the corresponding dynamical models. These requirements demand vehicle models that include nonlinearities while still having low complexity so they can be used for control design. Here we review some modelling approaches that can be used to achieve this goal.

From the perspective of mechanics, an automated vehicle, as any other vehicle, can be modelled as multi-body system consisting of rigid bodies and flexible elements. Complexity changes with the number of bodies and with the number of constraints considered in the modelling process. When using the Newton–Euler approach, the constraints are manifested as reaction forces. After eliminating these, one may derive differential equations for the vehicle’s motion. When the constraints only depend on configuration coordinates, one may utilise the Lagrangian approach in order to eliminate these so-called geometrical constraints and obtain differential equations immediately. However, in case of vehicles, the constraints often include velocities resulting in so-called kinematic constraints. In order to eliminate these and obtain differential equations straight away, one may use the Appellian approach [11–13].

In the next three subsections, we present some details regarding the models describing the longitudinal, lateral and yaw dynamics of automobiles. These models will serve as the base for the later discussions on vehicle control and safety verification for automated vehicles. We emphasise that, while some version of these models have been widely used in the literature to evaluate the dynamics of human-driven vehicles, some extensions are necessary when one wishes to apply them for motion planning and control of automated vehicles. In case the reader is familiar with these models he/she may decide to skip these subsections and continue with Section 2.3.
2.2.1. Longitudinal dynamics

Consider the mechanical model shown in Figure 3(a). Omitting the flexibility of the suspension and the longitudinal deformation of the tires, and assuming that the wheels are rolling without slipping, one may eliminate the reaction forces acting between the vehicle body and the axles and obtain differential equations for the longitudinal motion:

\[
\begin{align*}
\dot{s} &= v, \\
\dot{v} &= -g \sin \phi - g \xi \cos \phi - \frac{k}{m} (v + v_w)^2 + \frac{T_w}{mR},
\end{align*}
\]

(1)

where the dot represents the derivative with respect time \( t \), while \( s \) and \( v \) denote the longitudinal position along the road and the longitudinal velocity. The mass of the vehicle is denoted by \( m \), the mass moment of inertia of the rotating elements is neglected, \( g \) denotes the gravitational constant, \( \xi \) is the rolling resistance coefficient, \( \phi \) is the grade, \( k \) is the air drag coefficient (incorporating the air density and vehicle frontal area), and \( v_w \) is the speed of the headwind. The last term contains the wheel torque \( T_w \) and the wheel radius \( R \). By calculating the reaction forces this model also allows one to determine the weight transfer between axles that can be included in the handling models.

To control the longitudinal vehicle motion, one can command the longitudinal acceleration \( u \) based on the states (position and velocity) as well as the environmental information collected by sensors and V2X connectivity (e.g. positions and velocities of other vehicles). Then, the wheel torque can be assigned as

\[
\frac{T_w(t)}{mR} = \text{sat}(u(t - \tau)),
\]

(2)

where \( \tau \) incorporates the sensory, communication and computation delays as well as the powertrain dynamics. We remark that the latter is often approximated by a first-order lag in the literature. The sat function saturates at \( a_{\text{min}} \) corresponding to the torque limitations of the brakes and at \( \tilde{a}_{\text{max}} = \min\{a_{\text{max}}, P_{\text{max}}/(mv)\} \) corresponding to the torque and power limitations of the engine; see Figure 3(b,c).

In order to simplify the control design, the grade and the headwind are often neglected and the model (1) is linearised about the constant longitudinal velocity \( v^* \), yielding

\[
\begin{bmatrix}
\dot{s} \\
\dot{v}
\end{bmatrix} =
\begin{bmatrix}
0 & 1 \\
0 & -2v^* \frac{k}{m}
\end{bmatrix}
\begin{bmatrix}
s \\
v
\end{bmatrix} +
\begin{bmatrix}
0 \\
1
\end{bmatrix} \tilde{u}(t - \tau).
\]

(3)

We remark that the term \(-2v^* \frac{k}{m}\) is typically very small, and thus, it may be omitted leading to a second-order integrator with input delay.

2.2.2. Lateral and yaw dynamics

When modelling the lateral and yaw dynamics of automobiles, one may utilise a so-called single track or bicycle model conceptualised in Figure 3(d). In this case, the front wheels are represented by a single wheel and the same applies to the rear wheels. The wheelbase is
denoted by $l$, the distance between the centre of mass $G$ and the centre of the rear wheel $R$ is given by $d$, while the steering angle is denoted by $\gamma$. In order to describe the configuration, one may use the yaw angle $\psi$ and the position of centre of mass $(x_G, y_G)$, or the yaw angle $\psi$ the position of the centre of the rear wheel $(x_R, y_R)$.

Assuming rigid massless wheels that are rolling without slipping, one may obtain the kinematic constraints that require the velocity of the wheel centre points $R$ and $F$ to be aligned with the vehicle planes; see Figure 3(e). When also assuming constant longitudinal
speed \( v^* \), one can derive the so-called kinematic bicycle model

\[
\begin{align*}
\dot{\psi} &= \frac{v^*}{l} \tan \gamma, \\
\dot{x}_G &= v^* \left( \cos \psi - \frac{d}{l} \sin \psi \tan \gamma \right), \\
\dot{y}_G &= v^* \left( \sin \psi + \frac{d}{l} \cos \psi \tan \gamma \right),
\end{align*}
\]

\[
\iff
\begin{align*}
\dot{x}_R &= v^* \cos \psi, \\
\dot{y}_R &= v^* \sin \psi.
\end{align*}
\]

While being very simplistic, this model is often used for motion planning purposes. Moreover, utilising the Lagrangian equations one may calculate the lateral forces acting at the wheel-ground contact points and provide some predictions about whether the friction is able to keep the vehicle on track.

The lateral deformations of the tire (omitted by the kinematic model) may have significant effects on the stability and performance of the vehicle. Due to these deformations the velocities of the wheel centres do not align with the wheel planes but form tire slip angles \( \alpha_R \) and \( \alpha_F \); see Figure 3(f). Then, using the lateral velocity \( \sigma \) and the yaw rate \( \omega \) as pseudo velocities, the Appelian approach yields the differential equations

\[
\begin{align*}
\frac{m(\dot{\sigma} + v^* \omega)}{J_G \dot{\omega}} &= F_R + F_F \cos \gamma, \\
\dot{\sigma} &= \omega, \\
\dot{x}_G &= v^* \cos \psi - \sigma \sin \psi, \\
\dot{y}_G &= v^* \sin \psi + \sigma \cos \psi,
\end{align*}
\]

where \( J_G \) represents the moment of inertia about the centre of mass and \( c = l - d \). The lateral forces \( F_R \) and \( F_F \) depend on the slip angles \( \alpha_R \) and \( \alpha_F \) according to the nonlinear characteristics shown in Figure 3(g), while the aligning moments are neglected here for simplicity. To complete the equations, the slip angles are derived from the velocity components using kinematics

\[
\alpha_R = - \arctan \frac{\sigma - d \omega}{v^*}, \quad \alpha_F = \gamma - \arctan \frac{\sigma + c \omega}{v^*}.
\]

We remark that when investigating the handling dynamics of human-driven vehicles, the last three equations in (5) are often dropped since the configuration coordinates \( \psi, x_G, y_G \) only appear here. That is, if the steering angle \( \gamma(t) \) is given one may solve the first two equations without considering the last three and these two are often used to characterise lateral stability of vehicles. This, however, is not the case for automated vehicles which utilise feedback laws of the form \( \gamma(\psi, x_G, y_G, \ldots) \) in order to avoid objects located in physical space (represented by \( \ldots \) in the feedback law). For motion planning, control, and safety verification of automated vehicles one must consider the whole five-dimensional state space \((\sigma, \omega, \psi, x_G, y_G)\).

For the sake of control design, the model (5)–(6) is often linearised about the straight running motion. In particular, linearising around the constant speed motion along the \( x \)
axis, we obtain

\[
\begin{bmatrix}
\dot{\sigma} \\
\dot{\omega} \\
\dot{\psi} \\
\dot{y}_G
\end{bmatrix} =
\begin{bmatrix}
-\left(\frac{C_R + C_F}{m v^*}\right) & \left(\frac{d C_R - c C_F}{m v^*}\right) - v^* & 0 & 0 \\
\left(\frac{d C_R - c C_F}{J_G v^*}\right) & -\left(\frac{d^2 C_R + c^2 C_F}{J_G v^*}\right) & 0 & 0 \\
0 & 1 & 0 & 0 \\
1 & 0 & 0 & v^*
\end{bmatrix}
\begin{bmatrix}
\sigma \\
\omega \\
\psi \\
\gamma_G
\end{bmatrix}
\times
\begin{bmatrix}
\sigma \\
\omega \\
\psi \\
y_G
\end{bmatrix} +
\begin{bmatrix}
\frac{C_F}{m} \\
\frac{c C_F}{J_G} \\
0 \\
0
\end{bmatrix} \gamma,
\]

(7)

where the so-called cornering stiffnesses are defined by \( C_i = \frac{d F_i}{d \alpha_i(0)} \) and the equation for \( x_G \) is omitted, since the linear terms are all zeros. We remark that one may try to linearise (5)–(6) about other motions, like circular motion with constant steering angle corresponding to negotiating a curve of constant radius. The resulting equation is of similar form as (7), but with more complicated coefficient matrices.

2.2.3. Vehicle models of higher complexity

When controlling both lateral and longitudinal motion of the vehicle, one may try to merge the models (1) and (4), yielding four differential equations, or merge models (1) and (5), leading to six differential equations. In the latter case, one may also take into account the lateral deformation of the tires using so-called combined slip models. The development of such nonlinear models is currently of interest to the vehicle dynamics community as these are necessary for planning and control of automated vehicles to be able to execute a large variety of manoeuvres with high accuracy.

2.3. Vehicle control for automated vehicles

As in the case of vehicle dynamics, automated driving techniques can be divided into three categories depending upon the dynamic coupling between the describing equations. The first and generally best-understood category is associated with the longitudinal dynamics of the vehicle and corresponds to classical, adaptive, or connected cruise control. The second category of automated driving is associated with the lateral dynamics of the vehicles and corresponds to lane keeping. The final category of automated driving is associated with a simultaneous description of the longitudinal and lateral dynamics and corresponds to fully autonomous vehicle control. This section describes the state of the art in designing controllers for each category. As explained below, the last category is more than a mere merge of the first two categories and it is currently receiving a lot of attention in the vehicle control research community.

2.3.1. Classical and adaptive cruise control

Cruise control was introduced to the market in the 1950s by fitting vehicles with flyball governors, which were eventually replaced by electronic control units that measure the wheel based velocity of the vehicle and control the throttle and brake torque. The primary
The primary objective of cruise control algorithms is to maintain vehicle speed despite variations in road grade, head wind, vehicle mass, or powertrain behaviour. By assuming that the desired wheel torque can be obtained through powertrain and brakes and the tire friction is not a limiting factor in the generation of traction force, one can use the nonlinear differential equation (1) to describe the longitudinal vehicle dynamics. Linearising this equation about a nominal speed yields the model (3) and one can apply a variety of classical linear control techniques to track a reference speed specified by a driver. To account for variations of parameters like mass, unmodelled powertrain dynamics and unknown disturbances due to the grade and headwind, estimation algorithms often run in concert with the cruise control system to identify relevant parameters. For instance, recursive least squares parameter estimation or adaptive control algorithms have been applied in cruise control systems [14–16].

Cruise control has also been extended to respond to the motion of the vehicle in front of the one being controlled. This extension of cruise control, which was developed in the 1990s, is entitled adaptive cruise control (ACC). Apart from measuring the longitudinal velocity, ACC requires the measurement of the distance \( h \) and the rate of change of this distance \( \dot{h} \) between the vehicle and the one in front of it; see Figure 4(a). This can be achieved with the help of radar or cameras. To accommodate these new quantities, the model (1)–(2) is rewritten as

\[
\begin{align*}
\dot{h} &= v_1 - v, \\
\dot{v} &= -g \sin \phi - g \xi \cos \phi - \frac{k}{m} (v + v_w)^2 + \text{sat}(u(t - \tau)),
\end{align*}
\]  

(8)

where \( v_1 \) denotes the velocity of the vehicle ahead. Often the resistance terms in the second equation are neglected to simplify the control design. Moreover, the time delay \( \tau \) is sometimes approximated by the first-order lag, resulting in \( \text{sat}(w) \) in the second equation and the additional equation \( \dot{w} = (u - w)/\tau \). Finally, in some cases the system (8) is augmented with \( \dot{v}_1 = a_1 \) and the acceleration of the car ahead is used as an input.

The primary objective of ACC is to match the longitudinal velocity to that of the vehicle ahead and to keep a distance specified by the range policy \( h = H(v) \). There are numerous possible range policies, the simplest being the constant distance range policy that is rarely used in practice due to its poor performance in disturbance mitigation. The most common range policy is the constant time headway policy. This is depicted in Figure 4(b) using the form of \( v = V(h) \), where \( V = H^{-1} \) on the domain where the inverse exists [17]. This representation highlights three separate regimes: for small headways the vehicle intends to stop for safety reasons; for large headways it intends to travel with the velocity \( v_{\text{set}} \) (i.e. it is running conventional cruise control); while for intermediate headways it tries to maintain the time headway \( \tau_{\text{set}} \). By shaping the function for intermediate headways, more complicated range policies may be created to improve driver comfort and/or safety; see Figure 4(c,d). The control objectives may be achieved by car-following feedback laws that typically rely on the range \( h \), range rate \( \dot{h} \), and the longitudinal velocity \( v \). In addition, other strategies (e.g. feedback linearisation, integral action or adaptive control) may be used to compensate for the resistance terms in (8).

While ACC is primarily aimed at improving driver comfort, keeping the vehicle safe is of utmost importance. Thus, the car-following feedback laws are often adjusted by a collision avoidance algorithm. In practice, one can synthesise and in fact verify that the feedback
laws that are applied while performing ACC are provably safe [18,47]. This process is discussed in more detail in Section 3. Another interesting question is what the impact of ACC driven vehicles is on the motion of other vehicles participating in the traffic flow. This leads to the notion of string stability and more details are provided in Section 5.

2.3.2. Lane keeping
Another comfort feature that has been introduced during the past two decades is lane keeping. To perform lane keeping, the vehicle must be able to detect the lane, which is typically accomplished by using a vision-based system that may also be augmented by high precision GPS and high definition maps. For control design, typically the linearised model (7) is used and lane keeping is accomplished using a variety of classical linear control techniques including PID control and linear quadratic optimal control [19]. The design becomes more challenging once the road curvature changes significantly and longitudinal velocity of the vehicle is high. In these cases, arclength parameterised representation of the centre line is needed in an Earth-fixed frame and the nonlinear model (5) is utilised to keep the vehicle on the road. Similar to the cruise control case, in practice one can synthesise and in fact verify that the feedback laws that are applied while performing lane keeping are provably safe [20–23]. Techniques have even emerged to verify the composition of controllers tasked with simultaneous ACC and lane keeping [24,25]. Some of these verification methods are described in greater detail in Section 3.

2.3.3. Fully automated driving
The primary purpose of fully automated driving is to generate a control strategy for the vehicle that avoids obstacles in the environment while ensuring comfort and efficiency. The generation of this controller is typically formulated as a motion-planning problem wherein a dynamically feasible trajectory from an initial state to some terminal state is computed.
while avoiding collisions with objects in the environment and satisfying traffic rules. The description of the initial and terminal states is usually left to some high-level planner.

Motion planning algorithms can be cast as either path or trajectory generation problems. In the path planning formulation, a path of given length is constructed in the configuration space \((x, y, \psi)\). The solution does not prescribe how fast the path should be followed, instead the velocity is prescribed by a lower-level feedback controller. In contrast, the trajectory generation problem constructs a trajectory in state space (parameterised by time) along a given time horizon. This enables one to incorporate vehicle dynamics and handle dynamic obstacles. We briefly describe the various algorithms that have been proposed to solve either the path or the trajectory generation problem; however, several recent papers provide longer introductions to these topics \([26,27]\).

Path planning formulations of the motion planning problem can be divided into incremental search and variational methods. Incremental search methods such as rapidly exploring random trees (RRTs) and lattice planners (LPs) generate a graph- or tree-based representation of the configuration space \((x, y, \psi)\) and trace out edges to generate a path between the initial and terminal configurations of the vehicle. RRTs, for instance, construct a tree data structure that is expanded stochastically, until the terminal configuration is reached \([28,29]\). Then the desired path is obtained by tracing the edges that lead to that terminal configuration from the initial one. To generate a path corresponding to the different edges in the graph, RRTs typically utilise a kinematic model of the vehicle (cf. (4)). As a result, these approaches often generate nonsmooth paths, which can be difficult to track well. In contrast, LPs discretise the configuration space of the vehicle and represent it as a graph \([30–33]\). Paths between the different edges in this graph are usually pre-computed while incorporating vehicle dynamics, which can enable one to pre-select dynamically feasible paths. The desired path is found by performing a search for a minimum-cost path on the graph. Since these methods require discretising the configuration space, they are usually restricted to working in specific scenarios, which can make transferring these approaches to a novel context at run-time challenging.

The trajectory generation algorithms formulate motion planning as a nonlinear optimisation problem where the decision variable is usually the control input that is applied to the vehicle \([34–39]\). Note that this optimisation problem can be formulated in a receding horizon fashion. The constraints in this formulation typically correspond to the vehicle dynamics, the initial and terminal states of the vehicle, and any environmental constraints that may exist. Since the optimisation variable and constraints are infinite dimensional, one typically discretises time to solve the problem, which results in a nonlinear finite dimensional optimisation problem. Solving these problems globally, at run-time, is usually infeasible. As a result, typically trajectories are pre-computed using environmental constraints that are fixed and known a priori (e.g. locations of sidewalks or the lane median). However such trajectories are unable to incorporate environmental information that may only be known at run-time (e.g. pedestrians on sidewalks).

Once a motion plan is generated, a feedback controller is constructed to follow the reference path or trajectory to ensure appropriate behaviour in the presence of modelling error. To generate this feedback controller, model predictive control is usually applied \([40–43]\). Since this feedback controller must be constructed at run-time, the dynamics of the vehicle are usually linearised about the reference motion and the model predictive control problem is formulated as a convex, quadratic program. Environmental constraints that can only be
measured at run-time (e.g. pedestrians and surrounding vehicles) are usually accommodated at run-time by simplifying them into linear polytopic constraints that are valid within some appropriate region of the reference trajectory.

Three classes of approaches have also been developed to verify the safe operation of controllers, which are synthesised for fully automated driving. The first class of verification techniques focus on simply checking whether a pre-computed controller can be safely applied during fully automated driving [44,45]. The second class of verification techniques tries to correct the control inputs at run-time using the so-called viability kernels [46]. The final class of techniques synthesises controllers that are verified to operate safely [18,47–53]. Some of these methods are described in greater detail in Section 3.

3. Safety verification for automated vehicles

As the autonomous driving features on passenger vehicles become prevalent, the risks associated with potential failures of such algorithms also become a more serious concern. While extensively testing these algorithms in diverse real-world scenarios has been the main validation approach to build trust in these algorithms, it is well-known that ‘correctness’ cannot be guaranteed just by testing. As an alternative, tools from formal methods, in particular, formal verification, falsification, and correct-by-construction synthesis can be used to provide formal guarantees on safety [54]. Formal verification is the task of algorithmically computing a certificate that the system satisfies its specification, while falsification is the task of algorithmically searching for initial conditions and external inputs within the allowable set of inputs (disturbances) that will make the system fail the specification. As an alternative to these approaches, correct-by-construction control synthesis aims to construct a controller together with a domain of validity under some assumptions on the system model and the implementation platform. When the system is initialised within this domain, the synthesised controller guarantees that the closed-loop system remains safe.

The main safety specification for automated vehicles is avoiding a crash. However, it is unrealistic to expect that an automated vehicle can prevent a crash under all circumstances. For instance, if a car suddenly cuts in front of an ACC equipped car on the highway and slams on the brakes, it might not be possible to slow down in time to avoid a crash. One advantage of safety verification and synthesis approaches is that within the specification, the assumptions under which safety is satisfied are explicitly specified. This explicit characterisation of assumptions provides a clear demarcation of the conditions from which safety is guaranteed; which in turn can be used for explaining potential failures/accidents and for identifying corner cases [55].

While modelling and control design have attracted considerable attention in the vehicle dynamics community, for safety verification, in addition to a model, a formal safety specification that is amenable to mathematical reasoning is also crucial. Temporal logics, including linear temporal logic (LTL) and signal temporal logic (STL), can be used in order to formally capture requirements [18]. While specifications written in plain English can have different interpretations (e.g. different people can interpret the meaning of ‘maintain a distance’ slightly differently), the specifications written using logics have unambiguous and mathematically precise meaning. This makes it easier within an engineering team to have a unified, common understanding of the requirements and assumptions. Moreover, logics enable mathematical analysis of the specification and algorithmic verification of a
system’s behaviour with respect to the satisfaction or violation of the specification. In particular, LTL is commonly used, as it is arguably not too far from plain English and is more favourable compared to real-time logics like STL in terms of computational perspectives. There are also readily available tools for verification, falsification, and controller synthesis from LTL specifications; see, for instance, [56–60].

LTL specifications are formulated with the help of temporal operators like □ (always) and ♦ (eventually) and logic operators like ∧ (and), ∨ (or), and → (implies). Some combination of the operators may be particularly useful, e.g. ♦□ (eventually always) can be used to characterise stability like properties, in particular, reaching a target operating condition and staying there indefinitely. To exemplify how temporal logics can be used for capturing the requirements relevant to automated driving functionality, we start with presenting a specification for the lane keeping controller [61,62]. Notice that in Equation (7) the quantity $y_G$ defines the distance of the centre of mass G from the lane centre. That is, keeping the vehicle within the lane at all times (always) can be expressed as

$$\square(|y_G| \leq \delta y),$$

where $\delta y$ denotes the allowable deviation from the centre, which is a function of lane width and the wheel track of the vehicle. This specification, in general, is not enough, as it allows the vehicle to yaw or to have excess lateral velocity, so further (point-wise in time) restrictions on the other states in (7) must also be enforced to ensure safety. We remark that such specifications can also be established when the automated vehicle follows a more general lane centre curve, in which case one needs to also specify the assumptions on maximum allowable curvature of the roads that the vehicle is expected to operate on.

For the ACC example (8), the ISO specification states ‘When the ACC is active, the vehicle speed shall be controlled automatically either to maintain a time gap $\tau_{set}$ to a forward vehicle, or to maintain the set speed $v_{set}$, whichever speed is lower’. We express this in LTL as

$$\psi_g = \square(a_{min} \leq u \leq a_{max}) \land$$

$$\square((\square(v_{set} \leq h/\tau_{set}) \rightarrow \diamond\square(|v - v_{set}| \leq \varepsilon)) \land$$

$$(\square(v_{set} > h/\tau_{set}) \rightarrow \diamond\square(|v - h/\tau_{set}| \leq \varepsilon)) \land$$

$$(v_{set} \leq h/\tau_{set}) \rightarrow ((0 \leq v \leq v_{max}) \land (0 \leq h))) \land$$

$$(v_{set} > h/\tau_{set}) \rightarrow ((v \leq h/\tau_{min}) \land (0 \leq v \leq v_{max}) \land (0 \leq h))).$$

Here (9a) captures the input bounds, that is the maximal braking and acceleration torque allocated to the ACC system that should be respected at all times (cf. Figure 3(c) with large $P_{max}$). In addition, (9b) states the preconditions and requirements for the case where the set speed should be maintained and (9c) states the preconditions and the requirements for the case where a desired time gap should be maintained (cf. Figure 4(b) with $h_{st} = 0$). The parameter $\varepsilon$ represents the allowable tracking error, while maintaining the appropriate speeds. Finally, (9d) and (9e) capture the domains where different constraints are active together with the speed limit $v_{max}$ and a minimum time gap $\tau_{min}$ that should be maintained in either situation. One reason the formula looks complicated is its precision in explicitly specifying the conditions (assumptions) under which the specification is expected to be satisfied. We can further include assumptions on the lead vehicle’s velocity and acceleration captured as

$$\psi_a = \square(v_{1,min} \leq v_1 \leq v_{1,max}) \land \square(a_{1,min} \leq a_1 \leq a_{1,max})$$

(10)
and the overall specification takes an assume-guarantee form $\psi_a \rightarrow \psi_g$. For this specification to be satisfied, the closed-loop system should satisfy its guarantees $\psi_g$ whenever the assumptions $\psi_a$ on the lead car behaviour are satisfied. On the other hand, when the lead car drives in a way that violates $\psi_a$, we do not expect the system to adhere to $\psi_g$.

Once the specifications are converted to an LTL formula, controller synthesis involves encoding the temporal logic formulae with an appropriate automaton that can be used to define a set of nested fixed-point operations on subsets of states that appear in the formula. These operations essentially propagate subsets in state space in a way consistent with the dynamics. This may be done either on the continuous state space or on a finite abstraction of it [59,60,63]. This can be seen as a generalisation of control invariant set computation [64]. Iteratively applying the operations until a fixed point is reached results in the validity domain of the controller. A control policy, which maps each state in the domain of validity to the set of all admissible control inputs, is also extracted during this process [18]. Figure 5(a) shows the validity domain (green region) for the ACC controller synthesised based on (9). For all states inside the green region, there is a controller, consistent with the speed and acceleration/braking constraints, that guarantees that a crash is avoided for all possible behaviours of the lead car compliant with $\psi_a$.

Now consider a scenario where a car that is initially driving with velocity $v_1$ cuts in front of the ego vehicle in a way that the initial headway becomes $h$ while the initial velocity is $v$. If $v_1$ and $h$ are small and $v$ is large enough, then the state $(h, v, v_1)$ may fall outside the green region. In this case there exists a future behaviour for the preceding vehicle (e.g. maximal braking within $\psi_a$), for which there exists no control policy for the ego vehicle that can prevent a crash. For this problem, it can be shown that the green validity domain is indeed the best possible validity domain, i.e. any other controller’s validity domain is a subset of this one. Contrarily, for any state in the blue region in Figure 5(b), the preceding vehicle, if driven adversarially yet within the assumptions $\psi_a$, can enforce a crash. Ability to compute validity domains is not only useful for comparing how safe different controllers are, but also useful in understanding the limits of safety and explaining the run-time behaviour.
Safety verification and control synthesis provide principled techniques for the design and analysis of automated vehicles, beyond simple driving functionality. Other relevant applications include powertrain control systems [65,66], semi-autonomous driving with human in the loop [67,68], and traffic flow control in a network of signalised intersections [69,70]. Recently similar techniques are also applied for verifying automated vehicle control that involves machine learning or perception modules within the control loop [71]. These techniques hold the promise of reducing the required testing time significantly before system deployment by bringing more formal analysis in earlier stages of the design process.

4. Powertrain control for connected and automated vehicles

Powertrain control has undergone a significant evolution since the introduction of automated fuel injection in the 1980s. Modern human-driven vehicles are equipped with sensors and on-board computers that can monitor and regulate the state of the powertrain in order to lower fuel/energy consumption and emissions and to improve vehicle driveability and comfort. Further improvements are possible by integrating V2X communication and exploiting automated driving [72]. Recent survey manuscripts [73–75] highlight the emerging opportunities in this domain and provide a comprehensive account of the existing literature. In what follows we highlight a few additional examples and approaches that exploits vehicle connectivity and automated driving technologies to improve engine and powertrain control. Note that the engine and powertrain control are downstream of automated driving decisions; see Figure 2.

The wheel torque $T_w = T_d + T_b$ in (1)–(2) is given by the driving torque $T_d$ put on the driven axle by the powertrain and the torque $T_b$ applied by the vehicle brakes. These depend on the type of powertrain, engine and transmission. For instance, for a conventional gasoline or a diesel powertrain without a torque converter, disregarding flexibility of the shafts, the engine dynamics can be modelled as

$$J_e \dot{\omega}_e = T_e - T_c, \quad (11a)$$

$$T_c = \frac{T_d}{i_t i_f \eta_t \eta_f}. \quad (11b)$$

Here $J_e$ is the rotational inertia of the engine, $\omega_e$ is the engine speed, $T_e$ is the torque provided by the engine, $T_c$ is the clutch torque (torque applied by clutch to the driveline), $i_t$ and $i_f$ are the transmission and final drive gear ratios, respectively, and $\eta_t$ and $\eta_f$ are the corresponding efficiencies. When the clutch is engaged and the wheels are rolling without slipping, we have $\omega_e = i_t i_f v/R$, where $v$ is the longitudinal velocity and $R$ is the wheel radius; cf. (1). Then combining (1) and (11) we obtain

$$\dot{v} = -\frac{mg}{m_{\text{eff}}} \sin \phi - \frac{mg}{m_{\text{eff}}} \xi \cos \phi - \frac{k}{m_{\text{eff}}} (v + v_w)^2 + \frac{i_t i_f \eta_t \eta_f T_e + T_b}{m_{\text{eff}} R}, \quad (12a)$$

$$m_{\text{eff}} = m + \frac{i_t^2 i_f^2 \eta_t \eta_f J_e}{R^2}. \quad (12b)$$

Note that one may additionally include the rotational inertia $J_w$ of the wheels and axles by adding $J_w/R^2$ to the effective mass $m_{\text{eff}}$.
Figure 6. A typical engine map showing the contours of the brake specific fuel consumption (BSFC) as a function of engine speed and engine torque. BSFC is a measure of engine fuel efficiency, defined as the ratio of fuel rate in g/sec to engine power in kW and then multiplied by 3600.

We also note that the engine torque \( T_e \) is a function of engine control inputs (fueling rate, air-to-fuel ratio, spark timing, valve timing, residual gas fraction, cylinder deactivation, etc.) and engine dynamics (e.g. manifold filling dynamics, turbocharger dynamics, etc.); see [76] for details. For the powertrains of electric vehicles or hybrid electric vehicles, the battery dynamics, the efficiency characteristics of motor-generators and planetary transmission, and regenerative braking need to be taken into account.

### 4.1. Powertrain control strategies at a single-vehicle scale

Eco-driving exploits the interplay between traction losses at the wheels and efficiency characteristics of the powertrain and vehicle components to maximise the tank-to-wheel efficiency as measured in miles-per-gallon (MPG) for conventional vehicles or in miles-per-gallon equivalent (MPGe) for electrified vehicles. Smoother driving which avoids unnecessary use of friction brakes, transmission downshifts and torque converter unlocks generally lowers fuel/energy consumption. At the same time, engine efficiency may change significantly depending on the engine speed and torque; see Figure 6. This has motivated research on control of the vehicle’s longitudinal motion together with its powertrain. For instance, pulse-and-glide driving strategies were proposed where the engine operates at a medium to high load for a shorter duration followed by operation at a low load for a longer duration [77,78]. Such unconventional strategies can lower fuel consumption of vehicles equipped with conventional and hybrid electric powertrains without compromising average travel speed but not of battery electric vehicles.

V2X communications can be used to inform a preview of road topography and a forecast of traffic conditions. Such a preview can be exploited by optimisation-based controllers to reduce fuel/energy consumption. Information on grade, speed limit and traffic for the anticipated route ahead is utilised for a vehicle equipped with an internal combustion engine in [79] and for a hybrid electric vehicle in [80]. In the latter case, optimal set-points for battery state of charge (SoC) are computed for each segment of the anticipated route. To simplify the computations, a virtual route is used, consisting of a few initial route segments and a virtual terminal segment. Virtual terminal segment represents average properties of the part of the route beyond the receding horizon. The first SoC set-point in the optimised
SoC set-point sequence is applied for the current segment. Once the vehicle reaches the next segment, information is updated and optimisation is repeated. Compared to the same SoC for all route segments, dynamic programming showed a benefit of 10.8% while receding horizon with virtual terminal segment demonstrated 8.1% benefit over an urban route in Cambridge, MA.

In [81] an approach to ambient condition sensing through wireless communications with a weather station is proposed and experimental vehicle implementation results are reported. The ambient humidity information can be exploited for improved engine spark timing control to achieve lower fuel consumption, to improve the accuracy of the air-to-fuel ratio (UEGO) sensor, and for heating, ventilation, and air conditioning (HVAC) control. The ambient pressure information can be used for diagnostics of barometric pressure sensor. A relatively unexplored aspect concerns the integration of the information about wind speed and direction into the route and vehicle speed planning. The cloud cover information can be used to improve the efficiency of HVAC control and enhance thermal comfort perception by the customers by adjusting HVAC operation in response to an estimate of solar radiation absorbed by the vehicle.

References [82,83] illustrate the benefits of coordinated control of engine actuators (spark timing, valve timing, waste gate) and dual clutch transmission while exploiting preview/communication with a traffic light. Assuming a preview of a launch event is available about 1 second before vehicle launch, engine speed is increased within preview time window to prepare for launch. With this approach, terminal vehicle speed at the end of the launch phase is improved 25% and 38% at standard and high altitude conditions, respectively, if trajectories of all actuators and pre-launch engine speed set-point are optimised versus if only the clutch torque is optimised. Including pre-launch speed set-point into the optimisation leads to 11% incremental improvement in terminal vehicle speed at the standard condition in case all other actuators are included in the optimisation.

A significant aspect of powertrain control and calibration in the automotive industry is ensuring acceptable vehicle driveability. According to [84], the ‘term driveability describes the driver’s complex subjective perception of the interactions between driver and vehicle’. For human-driven vehicles driveability metrics and requirements have been established through extensive past development efforts; see, e.g. [84]. For Level 4 and 5 automated vehicles, on the other hand, metrics and requirements need to address rideability, i.e. the perception of the interactions between a passenger riding an automated vehicle, rather than driveability, and their development represents an opportunity for continuing research [85].

4.2. Powertrain control strategies with multiple vehicles

Many opportunities exist for improved energy efficiency and enhanced powertrain control in vehicle platooning scenarios and from cloud-based services that do not require dense penetration but benefit from integrating information from a few vehicles. In vehicle platooning scenarios, the energy efficiency can be gained through speed control and gap control (aerodynamic drag reduction/slipstreaming). More details about the structure of such controllers are provided in Section 5. Vehicle-to-cloud connectivity offers new opportunities in diagnostics, prognostics and condition-based maintenance. Data from multiple vehicles can be fused on the cloud to estimate parameters, predict degradation trends and optimally schedule maintenance of powertrain and vehicle components; see, e.g. [86,87].
Currently, real-world benefits of speed trajectory optimisation for connected automated vehicles are limited by traffic unpredictability. For instance, a human-driven vehicle may cut-in in front of a CAV which follows an optimised speed profile, forcing the eco-driving CAV to brake aggressively, thereby degrading the fuel/energy efficiency. At dense penetration levels of CAVs, such occurrences could be reduced through distributed control and vehicle-to-vehicle coordination. Furthermore, much is to be gained by co-optimisation of vehicle speed profile and powertrain operation. For instance, in [88] simulation results for a real-world corridor of six intersections over Plymouth Road in Ann Arbor, Michigan, revealed that the eco-driving speed trajectory planning can reduce the energy consumption of CAVs by up to 29.8%, with an average saving of 13.1%, compared to a baseline hybrid electric vehicle with non-optimised rule-based powertrain controller. Additionally, for one selected vehicle, while up to 11.9% energy saving is observed from speed optimisation, further energy savings of up to 14.2% and 18.8% can be achieved after applying the HVAC thermal load and power split optimisations, respectively.

Powertrain control strategy, architecture and calibration can be greatly simplified, if all vehicles are automated and centrally coordinated. In particular, the control system of a slower accelerating and decelerating vehicle can rely to a larger extent on feedback and adaptation rather than calibration-intensive feedforward control. Smaller and more reliably predicted driveline side disturbances for vehicles with continuously variable transmission (CVT) can permit lowering belt clamping pressure and improved fuel economy. Vehicles with turbocharged gasoline engines may no longer need a compressor bypass valve (CBV) if aggressive decelerations that can cause compressor surge are eliminated (or may not need to open CBV as frequently, which degrades fuel economy). However, it is likely that at high penetration levels, where the vehicle speed can be reliably controlled, the design of engine and powertrain for CAVs can drastically change.

With ubiquitous penetration of connected and automated vehicles, it is conceivable that computations of vehicle manoeuvres will be greatly simplified and moved off-board to the cloud, in particular, if cyber-security concerns can be appropriately addressed. Furthermore, advanced sensors, which are currently expected to be integrated with the vehicles to provide information about the surrounding traffic, can be moved off-board to the infrastructure with information wirelessly communicated to the individual vehicles. This has a significant potential to lower the power draw from and cooling requirements (yielding 5kW estimated power loss in the current systems) of advanced electronics and computers, which perform automated driving functions and are currently located on-board of the individual vehicles.

4.3. Connected powertrain testbeds to facilitate the development of connected vehicles

As the above-mentioned algorithms are developed, evaluating them experimentally with high-fidelity becomes an important challenge, especially when fuel economy and emission improvements due to these algorithms need to be assessed with high accuracy under dynamic drive cycles. On the one hand, computer simulations are in general attractive as a controllable, repeatable, and scalable means for evaluating CAV algorithms. However, when it comes to predicting fuel consumption and emissions in highly transient
drive cycles, simulation models available for fast, system-level evaluations typically lack the required accuracy. As an example, estimates from purely simulation-based studies can be off by as much as 27% in terms of fuel economy and 350% in terms of emissions [89]. On the other hand, full-scale experiments with physical vehicles on actual roads offer the highest fidelity. However, they are expensive and difficult to orchestrate, control, repeat, and scale.

The connected testbed paradigm can help address this challenge. The connected testbeds concept refers to a cyber-physical experimentation paradigm that enables remote closed-loop access to engineering testbeds such as powertrain test cells. For example, a new CAV engine control algorithm developed in one location could be validated remotely on an engine testbed in another location. The concept is not limited to remotely accessing only one testbed; multiple geographically dispersed testbeds could be integrated to scale up the existing experimental capabilities or even create new ones. For instance, multiple engine testbeds could be integrated to emulate a platoon of CAVs, or an engine testbed could be remotely connected to a battery testbed to create a cyber-integrated hybrid powertrain experimentation capability.

The realisation of this paradigm brings about a new challenge, namely, the challenge of enabling such integration despite the delays introduced by the networks that are used to connect the testbeds. This challenge is important, because a connected testbed forms a closed-loop system, in which even small delays can significantly deteriorate the performance of the system, and even destabilise it. Thus, the high fidelity that is expected due to the inclusion of the physical powertrain elements in the loop can be lost because of the distortions in the closed-loop system dynamics due to the delays. Nevertheless, the literature has already demonstrated successful applications of connected testbeds [89–93]. With leveraging and further development of the connected testbed design strategies [94] and delay compensation techniques [95–97], connected testbeds can present a cost-effective, scalable, repeatable, and high-fidelity solution to evaluating engine scale CAV algorithms at various penetration levels.

5. Improving traffic dynamics with connectivity and automation

As described above, adaptive cruise control can improve the safety, fuel economy and driver comfort when responding to the motion of the vehicle ahead and these algorithms may be used for vehicles of different levels of automation. An important question is how these automated vehicles impact traffic dynamics. In order to qualitatively and quantitatively evaluate this impact, the notion of string stability is often used: a vehicle is said to be string stable if it attenuates the velocity perturbations arising from the vehicle ahead. There exist many different definitions of string stability based on what type of perturbations are applied and how their magnitudes are calculated [98,99]. The linearised version of the model (8) is often used to evaluate the string stability of controllers. In particular, by calculating the transfer function \( T(s) \) between the velocity perturbation \( \hat{v} \) of the ego vehicle and the velocity perturbation \( \hat{v}_1 \) of the preceding vehicle, one can evaluate the size of amplification/attenuation and ensure string stability by requiring

\[
|T(j\omega)| < 1, \quad \forall \omega > 0, \quad (13)
\]

where \( j^2 = -1 \); see [17] for more details.
Human drivers often exhibit string unstable behaviour due to their reaction time, and perturbations are amplified as they propagate backward along the chains of human-driven vehicles, often leading to the formation of stop-and-go traffic jams. In contrast, the ACC algorithm may be tuned to be string stable, and thus, mitigate traffic waves [100]. We remark that perturbations may arise at fixed spatial locations due to traffic incidents or changes in elevation. These may be compensated, however, by selecting appropriate range policies [101]. Indeed, if the longitudinal motion of all vehicles in the flow can be designed, string stability may be ensured for the overall vehicle string, while for partial penetration of vehicles the impact may be less prominent [102,103]. More precisely, when considering a mixed chain of human-driven and automated vehicles (see Figure 7(a)) one may define string stability between any two pairs of vehicles in the chain. Indeed, it is possible that the system shows string stability between two vehicles (say \(i\) and \(j\)) and in the meantime shows string instability between two other vehicles (say \(k\) and \(l\)). At the linear level, string stability between any two vehicles can still be evaluated using transfer functions. In particular, if the link transfer function between the velocity perturbations of vehicle \(i\) and vehicle \(i + 1\) ahead is denoted by \(T_{i,i+1}(s)\), then the \(j\)-to-\(i\) transfer function linking the velocity perturbations of vehicles \(i\) and \(j\) can be expressed as

\[
G_{ij}(s) = T_{i,i+1}(s) \ldots T_{j-1,j}(s). 
\]  

(14)

Then, the so-called \(j\)-to-\(i\) string stability can be ensured by

\[
|G_{ij}(j\omega)| < 1, \quad \forall \omega > 0,
\]  

(15)

which can be used to investigate how automated vehicles can compensate for the detrimental effects of human-driven vehicles and what penetrations of AVs is needed to make a mixed vehicle chains string stable [104].

In order to improve the performance of automated vehicles, one may utilise V2V connectivity. On the one hand, a connected automated vehicle may obtain motion signals (like acceleration, yaw rate) from vehicles within the line of sight, which may be challenging to obtain by sensors (lidar, radar, cameras). For example, having access to the longitudinal acceleration of the vehicle ahead allows one to improve the performance of the longitudinal controller of a CAV [105]. We emphasise that this does not require the vehicle ahead to be automated; it can be a human-driven vehicle equipped with V2V communication that we refer to as a connected human-driven vehicle (CHV). This is the simplest realisation of a concept called connected cruise control (CCC), where a CAV utilises motion information from human-driven vehicles ahead; see Figure 7(b). Indeed, when available, a CAV may utilise motion information from multiple CHVs within and beyond the line of sight (Figure 7(c,d)) in order to improve its safety and fuel economy [106,107]. This strategy can also have a positive impact on the overall traffic flow. In particular, one may design CCC algorithms so that CAV \(i\) attenuates the velocity fluctuations of the furthest CHV \(j\) it receives motion information from, that is, ensure \(j\)-to-\(i\) string stability (often referred to as head-to-tail string stability). This can be evaluated using the \(j\)-to-\(i\) transfer function \(G_{ij}(s)\) (also called head-to-tail transfer function), which can be calculated from link transfer functions \(T_{k,l}(s)\) for \(k, l \in i, \ldots, j\) representing the dynamical interaction when vehicle \(k\) utilises motion information from vehicle \(l\) (obtained by the human driver, sensors, or V2V connectivity). This calculation can be made very efficient for unidirectional information...
flows, i.e. when the CAV only uses information from vehicles ahead; see [108, 109]. Again, such design does not require the other vehicles to be automated and one may ensure that the performance of the network keeps improving with increasing penetration of CHVs and CAVs.

For higher levels of penetration of CAVs it may occur that multiple CAVs follow each other consecutively within the vehicle chain; see Figure 7(e). This provides these vehicles with the opportunity to cooperate and control their motion in a coordinated fashion, which is often referred to as cooperative adaptive cruise control (CACC) [110–118]. While calculating the head-to-tail transfer function is more demanding due to the bidirectional connections, these systems may show high level of performance for multiple reasons. Cooperative control may allow vehicles to travel at smaller time headway, which may help them to reduce air drag and increase their energy efficiency. Moreover, as each connected automated vehicle runs its own path planning algorithm, it can share it with other CAVs via V2V communication. This allows CAVs to utilise feedforward actions beside feedback and optimise their performance in a rolling horizon fashion. Challenges of CACC include the fact that multiple vehicles need to move in traffic while either keeping the formation (which may be difficult when performing manoeuvres like lane changes) or realigning the formation (that may be necessary when human-driven vehicles cut into the automated platoon).

If all vehicles become connected automated, many opportunities may arise for improving safety, energy efficiency and mobility [119, 120]. Since such scenarios may take many
decades to realise, there is significant interest on the field regarding the impact of connected and automated vehicles on the formation of large-scale traffic patterns such as congestion waves. Results in [121–124] indicate that these waves can be mitigated by CAVs even under partial penetration of automation and connectivity. In particular, V2X connectivity may be utilised by CAVs in order to obtain information regarding the large-scale traffic behaviour to which they may not have access using sensory information. This opens an avenue to Lagrangian traffic control, where CAVs are utilised for controlling the large-scale traffic dynamics leading to increased traffic flux and reduced travel time. This, in turn, may reduce the energy consumption and emissions at the system level in transportation systems and lead to tremendous societal benefits. How to incentivize CAVs to modify their motion from their local optimum toward system optimum remains a challenging open question at the moment.

6. Economical and societal impacts of connected and automated vehicles

The degree to which the economical and societal benefits of CAVs can be realised at the system level highly depends on their deployment strategy. CAVs will change the utility of travel, affecting multiple-choice facets including mode choice, destination choice, activity-chaining, and even choice of residence locations in the long run. As such, behavioural studies need to be conducted to capture the impact of the arising technologies on the supply and demand sides of the transportation networks, and the resulting equilibria. These can be formulated as large-scale optimisation problems, where individuals can schedule their daily activities, and their corresponding trips, by maximising their total daily utilities (i.e. utility of engaging in an activity plus the disutility associated with travel). The corresponding mathematical formulations allow for comparing activity and travel patterns under legacy vehicles and CAVs.

Despite the many anticipated benefits of CAVs at the vehicle and platoon level, in the absence of the right policies and solutions, CAVs could create a number of undesirable consequences at the system level, one of which is the potential increase in the system-wide vehicle miles travelled (VMT). A study in the San Diego County [125] demonstrated that, not accounting for induced activities and trips, having households replace their legacy vehicles with the optimal number of CAVs (the minimum number of vehicles required to satisfy household travel needs) will lead to a two-fold increase in the system-wide VMT, mainly due to the introduction of empty-haul trips.

Another way CAVs could increase the total VMT is by changing the trip-chaining behaviour of travellers. Trip chaining refers to forming tours of activities to be completed sequentially, in order to reduce the total time and VMT spent on travelling. For example, instead of making three separate trips from home to dry cleaning and back, from home to work and back, and from home to the grocery store and back, one might form a tour of home-dry cleaning-work-grocery store-home, substantially reducing the total VMT. With CAVs, over time it would be possible to send the vehicles to perform a portion of these daily tasks, e.g. picking up groceries. As such, trip chaining would not be used as often, leading to higher VMTs. Another potential source of increase in VMT is the expected change in residential patterns. It is envisioned that CAVs would allow for travellers to engage in productive activities while travelling, essentially reducing the time-cost of trips. This could
lead to a change in residential patterns in the long run, with people moving to more suburban areas looking for more affordable housing, and opting for longer work-based trips, thereby adding to system-level VMT. As such, when viewed from a system-level lens, it becomes clear that it is important for policy makers to anticipate the potential consequences of adopting a CAV system, and put in place the right policies, incentives, and solutions so as to curb some of the potential negative consequences to the extent possible.

One solution to reduce the increase in VMT in a CAV system is using ridesharing. It is expected that CAVs initially will be mostly deployed by ridesharing providers for both economic and practical reasons. The economic aspects mostly stem from CAVs’ higher price tags. The added cost of sensory and control systems in CAVs will likely prevent high levels of private CAV ownership at the onset. On the other hand, ridesharing systems are well-positioned to benefit from deployment of CAV fleets. Since driver compensation accounts for a large portion of fleets’ operational costs, ridesharing providers can offset the large capital cost of forming CAV fleets by the resulting reduction in their operational costs. Additionally, CAV fleets prevent scheduling complications posed by crew scheduling requirements (e.g. number of working hours, sequence of working hours, etc.), which can be especially limiting in long-haul trips.

System-level implications of a CAV fleet deployed in a ridesharing system highly depend on its implementation strategy. Specifically, a ridesharing provider can follow one of the two implementation strategies of private ridership or shared ridership of the fleet. In the private ridership implementation, a central ridesharing provider uses its fleet to serve one customer at a time. It is easy to see that the empty-haul trips between the drop-off location of a customer and the pick-up location of the next would cause an increase in total VMT, compared to the case where each customer uses her own privately owned vehicle. However, as the penetration rate of CAVs increases, the empty-haul trips will become shorter. Studies have shown that shared ridership of CAV fleet can substantially help to offset the increase in total VMT, rendering CAV deployment a more sustainable option [126,127].

6.1. Peer-to-peer transactions

Collaborative consumption of goods and services, also known as the sharing economy, underwent a substantial rise with the widespread availability of internet [128]. Facilitated by the sharing economy, V2X connectivity in transportation systems can enable real-time exchange of resources (e.g. trades) as well as real-time service offerings. Sharing economy has already shown to improve sustainability [129,130] and boost performance [131,132] in transportation systems.

Requests to use the transportation system are typically served on a first-come, first-served (FCFS) order. The FCFS order of serving requests originates from two considerations. First, since the transportation network is a public good, fairness considerations call for requests to be served in the order they arrive. Note that this general rule would preclude purpose-built, priority-access facilities, such as high-occupancy vehicle (HOV) and high-occupancy toll (HOT) lanes, expressways, etc. The second consideration is a practical one; since all requests arrive in real-time and no information on future requests exist at the time of resource allocation, the practical allocation strategy is to serve requests as they arrive. This, to some extent, can be remedied by using a forecast of arrivals. Forecasts obtained from historical data (often coupled with data collected by sensors in real-time) can be
used to enhance system-level performance. An example of incorporating demand forecasts into the decision-making process is adaptive traffic signal control, in which signal phasing and green times are allocated to opposing approaches based on a combination of historical travel patterns and real-time information. It should be noted that the degree to which demand forecasts can improve system performance depends on the quality of such forecasts – low-quality forecasts can generate solutions that are inferior to following an FCFS queuing discipline.

CAV technologies can help to increase system performance by addressing concerns about both fairness and future demand uncertainty. Let us consider the example of a ridesharing provider, in which requests for rides arrive in real-time and the demand is scheduled to be served based on an FCFS queuing discipline. Once a request is received, the system operator provides an itinerary for the ride and its corresponding price to the rider. Let us denote the perceived utility of rider $r$ from the proposed option as

$$u_r(v_r(\theta), p_r(\theta)) = v_r(x, \theta) - p_r(x, \theta),$$

(16)

where $v_r$ and $p_r$ are the valuation of rider $r$ from the proposed option and the price charged to this rider, respectively. A rider’s valuation of an option and the price charged are functions of his/her itinerary, denoted by $x$, as well as his/her private information (such as value of time/distance, penalty for detour from the shortest path, valuation of the activity to be completed at the end of the trip, etc.), denoted by $\theta$. Such private information can be directly solicited from riders and incorporated into incentive-compatible mechanisms to ensure participants cannot game the system. Once a rider is served, the resources allocated to her will be removed from the pool of available resources, that is, the itinerary of the vehicle dedicated to serving the rider will be partially fixed, and the available capacity of this vehicle will be reduced. This indicates less resources are available for the next request. That is, if the request of rider $r'$ arrives after $r$, it cannot be served due to lack of resources. However, if the resources were not previously allocated to $r$, they could have been utilised to serve $r'$ instead of $r$, or even both requests. Figure 8(a) displays an example of such a scenario.

Peer-to-peer (P2P) exchange, enabled by V2X connectivity among agents, provides an opportunity to increase the performance of the system, since it enables rider $r'$ to free vehicle $v'$ from its previous assignment to $r$ by engaging in a negotiation with rider $r$; see in Figure 8(b). This negotiation determines the monetary compensation that $r'$ offers $r$ for vehicle $v'$. The amount of this compensation can be determined based on the private types of agents, by maximising the social welfare function

$$W = \int \int (v' - c) \xi(c, v') dv' dc,$$

(17)

where $v'$ and $c$ denote the valuation of $r'$ from the new option (with the valuation of an outside option for $r'$ normalised to zero) and the opportunity cost of $r$ due to exchange, respectively. Function $\xi$ denotes the joint distribution of the private parameters of the two participants in the exchange, while the bounds of the integral identify the range of private parameters, for which the trade would generate positive social welfare (otherwise the trade will not take place). It can be shown that such bi-lateral trade mechanisms are individually rational (i.e. individuals would benefit from participating in the system, and thus,
Figure 8. An example of two requests, $r$ and $r'$, arriving at the system ($r'$ arriving after $r$) engaging in an exchange. Panel (a) displays the scenario where rider $r$ occupies the resource $v'$, leaving no resource left for serving $r'$. Panel (b) shows the result obtained after the exchange, where $r'$ purchases $v'$ from $r$, and $r$ is served by $v$ through an inferior itinerary compared to her original itinerary.

will voluntarily do so) and dominant-strategy incentive-compatible (i.e. individuals would maximise their utilities by being truthful in reporting their private parameters, such as value of time) [133].

Similar exchange mechanisms have been proposed for a wide range of systems and facilities, e.g. at signalised intersections to allow one traffic approach to collectively purchase ‘green time’ from the opposing approach [131], in freeways and facilities with parallel queues [134], among connected vehicles forming platoons [135], and in trading right-of-way [136]. These case studies demonstrate the increase in level-of-service and system performance obtained through P2P transactions.

7. Future challenges

Modelling vehicle dynamics and interactions in the age of automated vehicles contains many new challenges due to the large number of different scenarios that need to be covered and the high level of fidelity expected. While physics-based models will likely dominate the field during the next few years/decades, there are clear signs that these can be augmented by data-based models. This opens a door for bringing tools from machine learning (e.g. neural networks) to vehicles dynamics. However, one not only needs to ensure that these hybrid models outperform the physics-based counterparts, but also make sure that they generalise well for different environmental conditions.

Though techniques for controller design are well understood for automated vehicle systems when they can be modelled well as a linear system, much work needs to be completed before these methods can be suitably extended to arbitrary situations. In particular, numerical control synthesis techniques that are able to make guarantees about the discovery of approximately optimal or even feasible solutions in real-world environments, under real-time operation constraints are still being developed. Techniques for numerical control design also need to better accommodate uncertainties that may arise from the lack of accurate environmental models and the limitations of perception algorithms. Again, to be fully
applicable, any method for numerical control design that can accommodate uncertainties should make guarantees about the discovery of approximately optimal or even feasible solutions under real-time operation constraints.

Formal verification and correct-by-construction controller/supervisor synthesis hold great promise to build trust in the decision-making algorithms in connected and automated vehicles, from powertrain control-loops to autonomous driving functionality. There are several technical and usability challenges to take full advantage of these formal techniques. On the technical side, while verification of simple communication protocols for connectivity is somewhat studied, the verification of a large V2X system will require new modelling formalism to enable scalability. Another challenge is the verification of learning-based components (e.g. neural networks), which are already widely used in perception algorithms. In particular, how these components affect the overall system-level safety properties or how to pose verification questions with respect to datasets used in training these algorithms are open problems. On the usability side, not all automotive engineers are trained in formal verification, hence user-friendly software tools are needed to enable broad adoption in industry. Moreover, writing ‘correct’ specifications that fully capture the user’s intent can be challenging even for expert users. Therefore more research is needed also in usability and specification validation/testing. For example, domain-specific languages for intent capture can be useful for this purpose. Overall, we envision that initial adoption will require a combination of formal verification/synthesis and verification/synthesis guided validation and testing to reduce the potential test space while ensuring a high-coverage of safety critical scenarios.

In the area of improved powertrain control, many opportunities to exploit connectivity and take advantage of automated driving exist. The future challenges involve ensuring high reliability and low cost and complexity of the proposed solutions to make them realistic for implementation. In particular, resilience against sensor and actuator failures, communication dropouts, and cyber attacks need to be ensured. In order to further improve the safety and efficiency connected and automated vehicles cooperative driving algorithms may be utilised by exploiting V2X connectivity. Going beyond the simple ‘quasi-one-dimensional’ configurations of vehicle chains would require one to reformulate the trajectory planning problems for cooperative driving. Furthermore, cooperative control of CAVs may require one to redesign the V2X communication protocols and message sets which were originally established for human-driven vehicles.

Finally, there is a lot of uncertainty regarding the deployment of CAVs in terms of what levels of automation will be introduced to public roads and when. It is also an important question what type of new physical and cyber infrastructure these vehicles will demand in order to ensure safety and efficiency of themselves and other road participants. Evaluating the socioeconomic impacts under such uncertainties remains a huge challenge. Still the authors of this paper remain positive that connectivity and automation will be utilised to answer many of the transportation challenges of the twenty-first century.

**Disclosure statement**

No potential conflict of interest was reported by the authors.
Funding

T. Ersal was supported by the NSF [grant number 1646019]. I. Kolmanovsky was supported by the NSF [grant number 1931738]. N. Masoud was supported by the NSF [grant number 1837245] and by the USDOT [grant number 69A3551747105]. N. Ozay was supported in part by the NSF [grant numbers 1553873 and 1918123]. R. Vasudevan was supported by the ONR [award number 00014-18-1-2575] and by the NSF [grant number 1751093]. G. Orosz acknowledges the useful feedback from Sergei Avedisov, Tamás Molnár, and Dénes Takács.

ORCID

Tulga Ersal http://orcid.org/0000-0002-6811-8529
Neda Masoud http://orcid.org/0000-0002-6526-3317
Ram Vasudevan http://orcid.org/0000-0003-1978-0572
Gábor Orosz http://orcid.org/0000-0002-9000-3736

References


