AUTOMOTIVE POWERTRAIN CONTROL — A SURVEY

Jeffrey A. Cook, Jing Sun, Julia H. Buckland, Ilya V. Kolmanovsky, Huei Peng, and Jessy W. Grizzle

Asian Journal of Control
Vol. 8, No. 3, pp. 237-260
September 2006
AUTOMOTIVE POWERTRAIN CONTROL — A SURVEY

Jeffrey A. Cook, Jing Sun, Julia H. Buckland, Ilya V. Kolmanovsky, Huei Peng, and Jessy W. Grizzle

ABSTRACT

This paper surveys recent and historical publications on automotive powertrain control. Control-oriented models of gasoline and diesel engines and their aftertreatment systems are reviewed, and challenging control problems for conventional engines, hybrid vehicles and fuel cell powertrains are discussed. Fundamentals are revisited and advancements are highlighted. A comprehensive list of references is provided.

KeyWords: Automotive, powertrain, modeling, control.

I. INTRODUCTION

Modern automobile engines must satisfy challenging and often conflicting requirements. Environmental concerns have motivated legislative action by governments around the world to reduce tailpipe emissions. Global commitments to CO₂ reduction require improved fuel economy. Customers demand performance and efficiency. All of these objectives must be delivered at low cost and high reliability.

These challenges are being met by modern controls, advanced aftertreatment devices and innovative powertrains. In this paper, we describe approaches to systems engineering, aftertreatment, and control of advanced technology gasoline and diesel engines, hybrid electric powertrains and automotive fuel cells. In each case, fundamental models are discussed and important control problems are illustrated by example. This survey, however, is far from exhaustive and interested readers are encouraged to refer to the proceedings of the recent IFAC workshops on “Advances in Automotive Control” [1–4], the NSF workshop on “Integration of Modeling and Control for Automotive Systems” [5], and the new monograph [6] on the subject.

1.1 A brief history of electronic powertrain control

In 1965, the US Congress passed an amendment to the Clean Air Act providing for the creation and enforcement of automotive emission standards. This was followed shortly by the establishment of the California Air Resources Board and, in 1970, the US Environmental Protection Agency. These regulatory developments spurred major efforts by automotive manufacturers to reduce fuel consumption and vehicle emissions, and brought about several technology breakthroughs in the 1970s. That decade saw the introduction of electronic engine control and the development of key engine control components such as the catalytic converter, exhaust gas recirculation and the common application of electronic fuel injection. Also in the 70s, emission regulations began to be introduced in Europe and Japan. In the 1980s, closed-loop air-fuel ratio control was made possible by the invention of the heated exhaust gas oxygen (HEGO) sensor, and the three-way catalytic converter became a standard feature on vehicles in Japan and Europe as well as North America. The 1980s also witnessed the increased application of control theory and modeling in the development of automotive powertrain systems. The 1990s defined the “systems” decade for powertrain development. Control intensive engine technologies such as variable valvetrains, direct injection and continuously variable transmissions required a multivariable approach to control. At the beginning of the twenty-first century, with even more stringent emission regulations, tightened fuel economy requirements and mandates on greenhouse gas emissions such as CO₂, hybrid electric and fuel cell powertrains appeared as potential solutions to the continued challenges of clean and efficient personal mobility.

1.2 Organization of the paper

This paper is organized as follows: in Section 2, models of the conventional port fuel injection (PFI) gasoline engine and its three-way catalyst (TWC) aftertreatment system are developed and the air-fuel ratio \((A/F)\) control
problem is motivated. Important issues in A/F control, together with representative control techniques, are described by reference. Two extensions of the basic engine model are presented for a variable cam timing engine and a turbocharged engine with electronically controlled wastegate.

Section 3 addresses modeling and control of direct injection stratified charge (DISC) gasoline engines. In this section, a DISC engine model and its lean NOx trap (LNT) aftertreatment system are described, and unique control problems due to the hybrid nature of the engine are presented. The problems of mode transition control, LNT adaptation, and fuel economy-emission tradeoffs are addressed. A computationally efficient dynamic programming solution is described to guide the DISC system design.

Section 4 covers modeling and control of diesel engines. Diesel engine controls, while they share some common features with gasoline engines, have many unique advantages and challenges. Several unique diesel control issues including sensor configuration, subsystem coordination and aftertreatment technology are reviewed in this section.

Sections 5 and 6 are devoted to hybrid and fuel cell powertrain systems. For fuel cell based automotive powertrain systems, the control of reactant supply, humidity and temperature are highlighted. For hybrid vehicles, different system architecture and associated control issues are reviewed, together with methodologies and tools for control strategy development. In both sections, references are given in lieu of the model description, due to the space limit.

II. PORT FUEL INJECTION ENGINE CONTROL

In the conventional PFI gasoline engine, fuel is metered to form a homogeneous and generally stoichiometric mixture based on measurements of inlet air flow or intake manifold pressure, and injected into the intake port of each cylinder upstream of the intake valve. Emission control relies primarily on a three way catalyst system to convert the HC, CO, and NOx emissions in the exhaust. This system may consist of several TWCs with different precious metal formulations (Pt and/or Pd, generally) and locations in the exhaust system to optimize emissions performance. It is characteristic of the three-way catalytic converter that high simultaneous conversion efficiencies for the three species occur only in a narrow band around stoichiometry, emphasizing the criticality of A/F control to minimizing tailpipe emissions. An overview of the challenges related to emissions control in the design and development of powertrain control systems for modern passenger vehicles may be found in [7].

Considerable effort as well is made to minimize engine out emissions to reduce the amount of costly precious metal required in the TWC. Typically, NOx reduction is accomplished by reducing combustion temperature through exhaust gas recirculation (EGR). EGR can be introduced externally via a valve that connects the intake and exhaust manifolds, or internally via variable camshaft timing (VCT) control. VCT can improve fuel economy in addition to reducing emissions, but presents control challenges that arise from dynamic interactions in the engine breathing process.

Turbocharged engines present similar challenges. The torque developed by a conventional gasoline engine is proportional to the air supplied to the cylinders, because the A/F is controlled to stoichiometry. In a turbocharged engine, the density of the cylinder air charge is increased. Consequently, engine displacement may be reduced at equivalent power, providing improvements to CO2 emissions and fuel economy. To achieve these benefits in a modern engine requires coordinated control of the throttle and wastegate actuators.

The following subsection will provide a brief review of models for the PFI engine and the TWC aftertreatment system. Control problems for A/F regulation, VCT torque management, and turbocharged gasoline engines will also be discussed.

2.1 PFI engine and aftertreatment models

A great deal of literature over many years describes the development of “control oriented” engine models: that is, linear and nonlinear low frequency phenomenological representations that capture the essential system dynamics required for control development, along with key static behavior such as emissions and volumetric efficiency that may be obtained experimentally from steady state mapping on an engine dynamometer. The four-stroke engine cycle naturally divides the physical process into four events comprising intake, compression, power generation and exhaust. This hybrid (that is, discrete event plus continuous dynamics) nature of the system is typically captured in the model by crank-angle based sampling. An introduction to engine modeling may be found in [6].

2.1.1 The fundamental PFI engine model

The mathematical representation of the conventional, naturally aspirated engine includes the following elements: (1) the throttle body, (2) the intake manifold, (3) torque generation and (4) engine rotational dynamics. The model may also include the EGR system, exhaust gas temperature and pressure dynamics, and feedgas emissions. The intake manifold dynamics are derived from the ideal gas law:

\[ p_t = K_i (W_s + W_{egr} - W_{cyl}) \]  

(1)

where \( K_i \) depends on the intake manifold volume and temperature, \( W_s \), \( W_{egr} \) are the mass flow rates through the throttle body and the EGR valve respectively; and \( W_{cyl} \) is the mean value of the flow rate at which the charge is inducted into the cylinders. The flows through the throttle body and
EGR valve are represented by a standard orifice equation:

$$W_a = \frac{A_{a,\text{th}} P_a}{\sqrt{T_a}} \phi \left( \frac{P_e}{P_a} \right), \quad W_{\text{gpr}} = \frac{A_{\text{gpr}} P_e}{\sqrt{T_e}} \phi \left( \frac{P_e}{P_a} \right)$$

(2)

where \(A_{a,\text{th}}, A_{\text{gpr}}\) are the effective flow areas for the throttle body and EGR valve respectively; \(P_a, P_e, P_e\) are intake manifold, exhaust manifold and ambient pressures; \(T_a, T_e\) are the ambient and exhaust temperatures. The function \(\phi\) represents the effects of the pressure ratio on the flow across the valve:

$$\phi(x) = \left\{ \begin{array}{ll} \gamma^2 \left( \frac{2}{\gamma + 1} \right)^{\frac{1}{2-\gamma}} & \text{if } x \leq \left( \frac{2}{\gamma + 1} \right) \\ \gamma^2 \left( \frac{2}{\gamma + 1} \left[ 1 - x^{1-\gamma} \right] \right)^{\frac{1}{2-\gamma}} & \text{if } x > \left( \frac{2}{\gamma + 1} \right) \end{array} \right.$$  

(3)

where \(\gamma\) is the ratio of specific heats, which takes different values for \(W_a\) and \(W_{\text{gpr}}\).

The amount of charge inducted into the cylinders, \(W_{\text{cy}}\), is a function of engine speed, intake manifold pressure and, possibly, temperature, where intake manifold temperature depends on mass air flow and EGR. \(W_{\text{cy}}\) is generally represented as a static regression equation based on steady-state mapping data for a particular engine.

Engine rotational dynamics follow the equation:

$$\frac{\pi}{30} J_n \cdot \dot{N} = T_b - T_f$$

(4)

where \(T_b, T_f\) are the engine brake and load torque in Nm, respectively, and the factor \(\pi/30\) is due to the unit conversion of engine speed, \(N\), (from rpm to rad/sec). The engine brake torque, \(T_b\), is the net torque available on the crankshaft to drive the rest of the powertrain, and can be decomposed into:

$$T_b = T_i - T_f$$

(5)

where \(T_i\) is the indicated torque, a measure of the total torque delivered to the piston by burning the fuel and \(T_f\) is the total friction which the engine has to overcome when delivering the torque to the crankshaft. The friction torque includes the pumping losses during the intake and exhaust strokes plus mechanical friction and may be regressed as a function of engine speed and intake manifold pressure.

Brake torque is generally represented as a regressed function of \(W_{\text{cy}}, A/F, N,\) and ignition timing.

2.1.2 Three-way catalyst model

Control oriented models of the TWC generally incorporate two parts: an oxygen storage mechanism to account for the modification of the feedgas \(A/F\) as it passes through the catalyst, and the standard steady-state efficiency curves driven by the tailpipe \(A/F\) computed from the oxygen storage model [8-11]. The following model is taken from [8].

First, consider the oxygen storage sub-model. Let \(0 \leq \Theta \leq 1\) be the fraction of oxygen storage sites occupied in the catalyst. \(\Theta\) is also referred to as the TWC oxygen loading. The oxygen storage mechanism is then modeled as a limited integrator:

$$\dot{\Theta} = \frac{1}{C(W_a)} \rho(\lambda_{FG}, \Theta) 0.23 W_a (t - \tau) \left( 1 - \frac{1}{\lambda_{FG}} \right) \quad 0 \leq \Theta \leq 1$$

$$\dot{\Theta} = 0 \quad \text{otherwise}$$

(6)

where \(W_a\) denotes the mass air flow rate, used to approximate the flow rate of the mixture entering the TWC and \(\tau\) is used to account for the transport delay. \(C\) represents the effective catalyst “capacity,” or the volume of active sites for oxygen storage, expressed in terms of the mass of oxygen that can be stored in the catalyst, as a function of \(W_a\). \(\rho\) describes the exchange of oxygen between the exhaust gas and the catalyst; and \(\lambda\) denotes the relative air-fuel ratio, with stoichiometry at \(\lambda = 1\) (the subscript \(FG\) refers to the feedgas).

The effective TWC volume parameter, \(C\), is expressed as a function of \(W_a\) in order to account for an observed increase in effective volume at high flow rates, specifically above 10g/s. For clarity, it should be emphasized that \(C\) does not represent the physical volume of the catalyst, often sized according to the engine displacement. For example, if there were no usable storage sites (i.e., if they were poisoned by substances such as sulfur or phosphorus), then \(C\) would be zero.

The oxygen storage function \(\rho\) is modeled as

$$\rho(\lambda_{FG}, \Theta) = \begin{cases} \alpha_{\lambda; f_L}(\Theta) & \lambda_{FG} > 1 \\ \alpha_{\lambda; f_R}(\Theta) & \lambda_{FG} \leq 1 \end{cases}$$

(7)

where \(0 \leq f_L \leq 1\) representing the fraction of oxygen from the feedgas attached to a site in the catalyst, and \(0 \leq f_R \leq 1\) representing the fraction of oxygen being released from the catalyst and recombining with the feedgas. In the oxygen storage function, \(f_L\) and \(f_R\) vary with the TWC oxygen loading and potentially with the space velocity (that is, the feedgas volumetric flow rate divided by the catalyst volume). In the model, \(f_L\) is assumed to be monotonically decreasing, with value one at \(\Theta = 0\) and zero at \(\Theta = 1\), and \(f_R\) is assumed to be monotonically increasing, with value zero at \(\Theta = 0\) and one at \(\Theta = 1\).

The quantity \(0.23 \times W_a \times \left( 1 - \frac{1}{\lambda_{FG}} \right)\) represents the differential total mass of oxygen in the feedgas with respect to stoichiometry. When multiplied by \(\rho\), it gives the mass of oxygen that is deposited in (or released from) the catalyst. By conservation of mass, the resulting equivalent tailpipe \(A/F\) can be directly computed:
\[ \lambda_{TP} = \lambda_{FG} - p(\lambda_{FG}, \Theta) \times (\lambda_{FG} - 1). \] (8)

### 2.2 A/F control for PFI engines

Three main problems arise in A/F control of the conventional PFI engine: accurate estimation of air charge, compensation for fuel puddling dynamics in the intake manifold runners and precise regulation of closed-loop A/F for good catalyst performance. A low frequency model of the induction process is described in [12], and compensation is developed for the relatively slow dynamics of the conventional hot-wire anemometer used to measure inlet air flow. Transient fuel characteristics for a PFI engine were first reported by Fozo and Aquino in [13]. In [14], a method of adaptive transient compensation for fuel wall-wetting dynamics is described that accounts for varying fuel properties. The technique requires only a heated exhaust gas oxygen (HEGO) sensor, which remains the prevalent feedback sensor for closed-loop A/F control. A HEGO sensor is essentially a switch, indicating that the A/F mixture is either rich or lean of stoichiometry, but not by how much. The basic idea of [14] is to use the feedback signal to evaluate changes in A/F during driver induced transients in closed loop, and store corrections to the compensation algorithm indexed by engine temperature for use in the next transient or during open-loop cold start operation.

In [15], it was shown that cylinder-to-cylinder A/F differences result in a closed-loop lean shift in controlled A/F due to preferential diffusion of H2 and CO across the HEGO sensor upstream of the catalyst. This control-point shift causes a dramatic reduction in NOx conversion efficiency due to the precipitous nature of the TWC characteristic away from stoichiometry. Typically, this effect is mitigated by biasing the A/F setpoint slightly rich, at a cost in fuel economy and conversion efficiency of the other exhaust constituents. In [16], an approach to achieving uniform cylinder-to-cylinder A/F control for a 4-cylinder engine in the presence of injector mismatch and unbalanced air flow due to engine geometry is presented. The method recognizes that the individual cylinder representation of the fueling process describes a periodically time varying system due to the unequal distribution of A/F from cylinder to cylinder. The key features of the controller are the construction of a time-invariant representation of the process and event-based sampling and feedback. In [17], the method was extended to an 8-cylinder engine in which exhaust manifold mixing dynamics were significant.

A significant advancement in A/F feedback control capability is the introduction in production vehicles of the Universal Exhaust Gas Oxygen (UEGO) sensor. Unlike the conventional HEGO sensor which simply switches about stoichiometry, the UEGO is a linear device that permits an actual measurement of A/F [18]. Control and diagnosis of catalysts using UEGO sensors is described by [19,20]. In [21], Fiengo and co-authors use the catalyst model described above along with pre- and post-catalyst UEGO sensors to develop a controller with two objectives: to simultaneously maximize the conversion efficiencies of HC, CO and NOx, and to obtain steady-state air-fuel control that is robust with respect to disturbances.

### 2.3 Control of engines with variable cam timing

Variable cam timing provides improved performance and reduced feedgas emissions using an electro-hydraulic mechanism to rotate the camshaft relative to the crankshaft and retard cam timing with respect to the intake and exhaust strokes of the engine. In this manner, the amount of residual gas trapped in the cylinder at the end of the exhaust stroke is controlled, suppressing NOx formation [22-24]. In addition, VCT allows the engine designer to optimize cam timing over a wide range of engine operating conditions, providing both good idle quality (minimal overlap between the intake and exhaust events) and improved wide-open throttle performance (maximum induced charge). Obviously, variable cam timing has a substantial effect on the breathing process of the engine. Properly controlled, the variable cam can be used to operate the engine at higher intake manifold pressures, reducing pumping losses at part throttle conditions to provide a fuel economy improvement. Uncompensated, however, VCT acts as a disturbance to the breathing process, compromising drivability and substantially reducing its effectiveness in reducing emissions.

Four versions of VCT are available: phasing only the intake cam (intake only), phasing only the exhaust cam (exhaust only), phasing the intake and exhaust cams equally (dual equal), and phasing the two camshafts independently (dual independent). A low order nonlinear model of a dual-equal VCT engine is derived in [25]. In [26], the model forms the basis for active compensation of VCT induced cylinder air charge variation employing electronic throttle control (ETC). The balance of this section will review the VCT model and describe the ETC compensation.

The basic equations of the VCT engine model are the same as those in Section 2.1, modified to incorporate the effects of the cam actuator on engine breathing. For the VCT engine, the mass air flow rate into the cylinders is represented as a function of cam phasing, \( \zeta_{cami} \) in addition to manifold pressure, \( P_m \) and engine speed, \( N \):

\[ W_{ci} = F(N, P_m, \zeta_{cami}) \] (9)

which, for the design model of [26], is approximated by a function affine in \( P_m \):

\[ W_{ci} = \alpha_1(N, \zeta_{cami})P_m + \alpha_2(N, \zeta_{cami}) \] (10)

where \( \alpha_1 \) and \( \alpha_2 \) are low-order polynomials in \( N \) and \( \zeta_{cami} \).

A block diagram of the VCT engine is illustrated in Fig.1, which shows the cam timing reference, \( \zeta_{cami} \), scheduled on engine speed and driver demanded throttle position, \( \theta_0 \). Typically, the cam schedule reaches maximal cam re-
tards at part throttle to provide maximal internal EGR; close to idle and at wide open throttle, the cam phasing is at zero or slightly advanced. Scheduling cam on throttle causes it to change when the pedal is depressed or released. It is this torque variation caused by the cam transient that results in undesirable engine response and drivability problems. Note that the throttle angle is comprised of the throttle position due to the driver’s request ($\theta_0$) and an additive term due to the compensation ($\theta^*$),

$$\theta = \theta_0 + \theta^*.$$

The throttle flow equation is represented as functions of pressure and flow geometry, $\phi(P)g(\theta)$, as in the conventional engine model.

A feedforward compensator is designed to recover the drivability of the conventional engine by eliminating the effect of the cam transients on cylinder mass air flow. The algorithm employs $\theta^*$ as a virtual actuator, according to [26]. That is, a control law is developed for $\theta^*$ such that the rate of change of $W_{\text{cyl}}$ coincides with that of the conventional engine. Specifically, compensation $\theta^*$ is evaluated:

$$\theta^* = g^{-1} \left( \frac{\partial \alpha}{\partial \alpha} \frac{P_i}{\phi(P)} + \frac{\partial \alpha}{\partial \alpha} \frac{\dot{\theta}}{\phi(P)} g(\theta_0) \right) - \theta_0,$$  \hspace{1cm} (11)

where $\hat{P}_i$ is a fictitious reference manifold pressure which should be equal to the manifold pressure of the conventional engine driven with the throttle angle, $\theta_0$, and engine speed, $N$. This reference manifold pressure is generated by

$$\hat{P}_i = K_i \left( \phi(\hat{P}_i)g(\theta_0) - \alpha_i(N, 0) \hat{P}_i - \alpha_i(N, 0) \right).$$  \hspace{1cm} (12)

Figure 2 shows the reduction of the torque fluctuation during cam transients achieved by the compensation.

2.4 Control of turbocharged gasoline engines

Turbocharging is an efficient method to boost intake pressure, as it extracts energy from the exhaust gases to drive a compressor to pressurize ambient air. In automotive applications, operating conditions vary over a wide range of speed and load. A design challenge is to develop a system that provides adequate boost at low speed and load without creating an over-boost situation at high speed and loads [27]. Typically, the amount of boost delivered by a turbocharger is controlled by a wastegate. In any event, the advantages of turbocharging are accompanied by an increase in complexity of the control design and calibration.

Complexity is also introduced by other phenomena associated with turbocharging. For example, increasing charge density increases propensity for engine knock, particularly at high loads. This phenomenon is alleviated in many applications by passive or active thermal management with a charge cooling device, such as an intercooler. In conventional gasoline engines, knock is further controlled by spark retard [29]. In direct injection engines, fuel injection control may also provide some benefit [28].

Transient response is another factor, as turbocharger inertia leads to a phenomenon known as “turbo lag.” Turbo lag describes the delay in torque response due to the time required for the turbocharger to change speed and thus affect boost pressure. Control objectives for fast response to minimize this effect are tempered by limits on boost pressure overshoot, which can lead to unacceptable torque disturbances [30,31].

Modern turbocharged gasoline engines have advanced technology actuators such as electronic throttle and variable valve timing, in addition to the wastegate. Coordinated control of these actuators is critical to achieve the full benefit of these combined technologies. Historically, literature that pertains to wastegate control in gasoline applications, such as [32,33,29], refer to systems with a mechanical throttle.
More recently, control with advanced actuators has received significant attention. Apart from [30,31], however, the focus has been on the turbocharged diesel engine (for example see [34,35,36]). Most of these results cannot be applied directly to the gasoline engine due to fundamental differences in actuators and system performance objectives. A notable exception is control oriented component modeling, for example the turbocharger model presented in [37]. Such component models are key to the system level models of turbocharged gasoline engines developed in [38-40].

Such control oriented models are all based, in principle, on the fundamental PFI engine model discussed in Section 2.1. The basic engine model is augmented with mathematical expressions representing a turbocharger, with wastegate and an intercooler.

A schematic diagram of a turbocharged gasoline engine is shown in Fig. 3. The representation of the turbocharger consists of models of the compressor, turbine and wastegate, and includes the dynamic coupling of the compressor and turbine. The mass flow rate through the compressor, $W_c$, is described by

$$W_c = f_c \left( \frac{P_a}{P_c} \right)^{\gamma} N_{tc} T_a,$$

where $P_a$ is the compressor exit pressure, typically referred to as boost pressure, $P_c$ and $T_a$ are the compressor inlet conditions, which in most cases are assumed to be ambient, and $N_{tc}$ is the turbocharger shaft speed.

The compressor exit temperature can be calculated as

$$T_c = T_a \left[ 1 + \frac{1}{\eta_{ic}} \left( \frac{P_c}{P_a} \right)^{\gamma} - 1 \right],$$

where $\eta_{ic}$ is the isentropic efficiency of the compressor.

The power consumed by the compressor, $Power_c$, is calculated via the first law of thermodynamics,

$$Power_c = c_{p,c} W_c (T_c - T_a),$$

where $c_{p,c}$ is the specific heat at constant pressure of the air in the compressor.

The turbine is described in a similar fashion. The mass flow through the turbine, $W_t$, is modeled as

$$W_t = \frac{P_t}{\sqrt{P_e}} f_t \left( \frac{P_c}{P_e} \right)^{\gamma} N_{te} T_e,$$

where $P_e$ and $T_e$ are the pressure and temperature at the inlet of the turbine, respectively, which are typically assumed equal to the exhaust manifold conditions, and $P_t$ is the turbine exit pressure.

The turbine exit temperature is given by

$$T_e = \left[ 1 - \left( \frac{P_e}{P_t} \right)^{\gamma} \right] \eta_{tc} T_e,$$

where $\eta_{tc}$ is the isentropic efficiency of the turbine.

The power generated by the turbine, $Power_t$, is calculated from the first law of thermodynamics,

$$Power_t = c_{p,t} W_t (T_e - T_t),$$

where $c_{p,t}$ is the specific heat at constant pressure of the gas in the turbine.

The dynamics of the turbocharger shaft are given by

$$\frac{\Delta \omega}{\Delta t} = \frac{Power_t - Power_c}{J_{tc} N_{tc}},$$

where $J_{tc}$ is the inertia of the turbocharger.

The wastegate can be modeled with the standard orifice flow equation, as described in (3). Measurements needed to derive the effective orifice area may be difficult to obtain; nonetheless, an effective model can be developed with selected use of estimated variables, such as exhaust flow rate.

Model integration requires an exhaust manifold model and a model to represent the volume between the compressor and the throttle. Both volumes are typically modeled in a fashion similar to (1), with variations to account for temperature dynamics and/or heat transfer [39,41], depending on the application.

A turbocharged system model of this type is used by the authors of [30] to analyze system characteristics and develop charge control algorithms for a wastegated turbocharged system equipped with electronic throttle. Boost pressure and intake manifold pressure are both measured.
and conventional decentralized PI control with feedforward on the wastegate is used to regulate these measured variables to desired setpoints, which are chosen to achieve fuel economy, emissions and driveability objectives. This approach produces acceptable performance, however the wastegate is prone to saturation. Multivariable control techniques can be used to analyze the system to guide formulation of a modified controller that maintains a simple structure desirable for implementation, and yet benefits from a centralized control methodology. Such an approach is described in [42].

III. LEAN BURN AND DIRECT INJECTION GASOLINE ENGINE CONTROL

Lean-burn engines may be a major enabling technology for improving fuel economy of gasoline engines. Engines operated with lean mixtures have lower throttling losses at low and part loads, resulting in reduced (up to 15%) fuel consumption and CO₂ generation. The major technical hurdles in extending the lean-burn limit of a PFI engine are combustion stability and NOₓ treatment. While the lean limit of a conventional PFI engine has been significantly extended by advanced combustion concepts (such as those that induce high turbulence), the maximum A/F that can be achieved in PFI engines without compromising other performance indices is around 22. This limit is substantially extended by direct injection and stratification made possible by technical advances in high-pressure fuel injection and combustion chamber design. The issues of NOₓ emissions associated with lean-burn (port or direct injected) engines arise because of the fact that conventional three-way catalysts are ineffective for air-fuel ratios even slightly lean of stoichiometry. Consequently, lean-burn engines use an actively controlled emission device called a lean NOₓ trap (LNT) to meet NOₓ emission standards. The incorporation of the LNT adds both cost and complexity, (2) dealing with the sulphur poisoning problem to maintain its efficiency, and (3) assuring that the LNT operates within its temperature window to maintain high efficiency and to avoid thermal degradation.

In this section, we will focus on three main control problems for direct injection stratified charge (DISC) engines equipped with LNT: mode transition, aftertreatment control and adaptation, and system optimization and integration. While the port fuel injected lean burn engine control problems will not be explicitly addressed here, it should be noted that the issues and solutions for direct injection engines are applicable to PFI lean-burn engines as well, with minor modification.

3.1 Unique features and control implications of DISC powertrain system

A DISC engine, like a diesel, injects fuel directly into the combustion chamber. It is different from a conventional PFI engine discussed in Section 2 in several respects. Most importantly, the DISC engine can, depending on speed and load, operate in one of three combustion modes: homogeneous stoichiometric (A/F ≈ 14.64), homogeneous lean (between stoichiometry and about 20) or stratified (≥ 20). A homogeneous A/F mixture is achieved by injecting fuel early in the intake stroke, while stratification is achieved by injecting late, during the compression stroke [43]. The torque and emission characteristics corresponding to homogeneous and stratified operation are so distinct that different control strategies are required to optimize performance in the two regimes [44,45]. Note also that, in addition to the usual control variables such as throttle position, ignition timing, exhaust gas recirculation (EGR) and fueling rate, the DISC engine requires new inputs including injection timing, fuel rail pressure and swirl control at a minimum [46]. Finally, the ultra-lean A/F operation of the direct injection engine mandates the use of a lean NOₓ trap (LNT) to manage oxides of nitrogen emissions. The LNT, as a NOₓ storage device, needs to be purged periodically to regenerate its storage capacity.

These special features of DISC engine operation have important control implications and lead to the following unique control problems:

- Mode transition: Depending on engine operating and LNT loading conditions, the DISC engine will either operate in stratified or homogeneous mode or switch between the two modes. The control must be capable of changing the combustion mode and the air-fuel ratio of the engine rapidly without causing noticeable disturbance to the driver.

- Aftertreatment control: The requirements for the aftertreatment control include (1) periodically running the engine rich of stoichiometry to regenerate its trap capacity, (2) dealing with the sulphur poisoning problem to maintain its efficiency, and (3) assuring that the LNT operates within its temperature window to maintain high efficiency and to avoid thermal degradation.

- Optimization and trade-off analysis: The inclusion of the storage device in the aftertreatment system changes the nature of the optimization problem. The interactive characteristics of the subsystems involved, together with the time and trajectory dependent nature of LNT operation, result in a high dimensional and dynamic optimization problem that demands new computational methodologies and tools.

The engine and aftertreatment models, to be discussed in the following subsection, facilitate the model-based treatment of these problems.

3.2 DISC engine and its aftertreatment system models

3.2.1 DISC engine model

References [44,45] describe modeling and control of a direct injection stratified charge (DISC) gasoline engine and discuss the fundamentally hybrid nature of the system. This model is illustrated in Fig. 4. On the surface, the model structure is not dissimilar to a conventional PFI
engine discussed in Section 2, consisting of the throttle, intake manifold dynamics, engine pumping, torque generation, rotational inertia and feedgas emissions. In fact, many of the equations used to describe the PFI engines in Section 2 can be applied here. Because of the different characteristics for homogeneous and stratified operation, the model is, in fact, hybrid in the sense that most components are represented by two continuous-variable sub-models with a discrete switching mechanism to select the appropriate characterization based on injection timing. Additionally, the injection-to-torque delay, fundamentally associated with the four-stroke engine cycle (intake-compression-power-exhaust), becomes a function not only of engine speed, but also of the operating mode that dictates the relationship between the injection and combustion events.

3.2.2 Lean aftertreatment model

The typical aftertreatment system for a lean-burn engine with a commonly used sensor configuration is shown in Fig. 5. It consists of a conventional three-way catalytic converter (usually closely coupled to the engine for optimal cold start performance) and an underbody LNT, with oxygen and temperature sensors in various locations.

The key chemical reactions involved in the LNT operation can be briefly discussed as follows. NO\textsubscript{x} storage phase: under lean conditions, NO is oxidized in the gas phase and the resulting NO\textsubscript{2} is then adsorbed on storage sites such as barium nitrate. As the NO\textsubscript{x} stored in the LNT increases, the storage efficiency drops and the trap must be purged to regenerate its capacity. LNT purge phase: under rich conditions, the barium nitrate becomes thermodynamically unstable and releases NO\textsubscript{2} and BaO. BaO then combines with CO\textsubscript{2} in the exhaust to form BaCO\textsubscript{3}, thereby regenerating the storage sites. The released NO\textsubscript{x} is converted to N\textsubscript{2} over the precious metal sites by reductants (CO or H\textsubscript{2}) in the engine exhaust stream.

A control oriented representation of the LNT exhaust aftertreatment system was first developed in [47]. In this model, the amount of NO\textsubscript{x} stored on the LNT is a state. Under lean conditions, the NO\textsubscript{x} storage capability is modeled by a limited integrator with the storage rate of NO\textsubscript{x} being a monotonically decreasing function of the state of the integrator.

In [48], the model is extended by modifying the purge model to capture the interactions between the oxygen storage and NO\textsubscript{x} storage mechanisms in the LNT. By separately modeling the releasing and conversion reactions during the purge phase, the integrated model is able to replicate experimentally observed NO\textsubscript{x} spikes during the purge phase [49]. In another modification to the original model, air-fuel ratio, \( \lambda \), is used instead of \( W_{CO, in} \) in the functions that represent the NO\textsubscript{x} release rate and conversion efficiency, making the model more amenable to control implementation.

3.3 Mode transitions for DISC engine control

Typically, stratified operation is limited to low- and part-load engine operating conditions where the maximum fuel economy benefits of a DISC engine can be achieved. At increasing loads, stratified combustion often results in increased smoke and hydrocarbon emissions, requiring a switch to homogeneous operation. Similarly, as the engine speed increases, a mode switch is also necessary as the time for mixing and breathing is reduced, making it infeasible to operate in stratified mode (stratified operation requires more air charge). Finally, the LNT aftertreatment system needs to be purged periodically to maintain high efficiency, and this is accomplished by transitioning to an air-fuel ratio slightly rich of stoichiometry. Consequently, mode switching between stratified and homogeneous combustion may be initiated not only when the engine torque demand increases, but also when the torque demand is small and constant, such as when the engine is idling. The mode transitions have to be accomplished in a manner that does not create a disturbance noticeable by the driver, while providing the desired value of the engine torque.
throughout the transition. In [50], a hybrid control scheme is presented to manage the transition. The controller consists of a high level Transition Governor that is used to determine the combustion mode and the setpoints, and a low level feedback controller that coordinates the spark timing, throttle, and fuel injection to ensure the desired value of the engine torque throughout the transition. In [45], the coordinating control is derived by minimizing the cost function that incorporates performance indices for torque delivery, charge control, spark control and EGR delivery, with the weighting for each individual performance index adjusted according to the desired mode of operation. Figure 6 shows typical $A/F$ and torque traces on a small DISC engine for constant torque combustion mode transitions. In the case of a transition from homogeneous to stratified, the transient $A/F$ requirement is relaxed, giving the fuel actuator substantial authority to maintain constant torque during the mode shift. On the other hand, the transition from stratified to homogeneous operation at stoichiometry requires tight control on $A/F$ to meet emission requirements.

The same control problem can also be solved using a Lyapunov based speed-gradient algorithm as in [51], and hybrid model predictive control [52] which optimally coordinates the actuators over a receding horizon. In [53], the continuously variable transmission (CVT) is exploited to provide an additional control actuation during mode transitions to manage wheel torque and mitigate the effect of torque disturbances. The study reveals, however, that an intuitively sound CVT gear ratio control strategy which attempts to completely cancel the engine torque disturbance, results in unstable zero dynamics. The same paper then proposes a control strategy that coordinates the engine control variables (spark and fuel) with the CVT gear ratio control to stabilize the zero dynamics while achieving seamless mode transition.

The multi-mode operation of a DISC engine also brings new challenges for the standard idle speed control problem, as well as opportunities for improved engine idle performance. In [54] an idle speed controller is designed for a DISC engine by exploring the use of electronic throttle, spark and fuel. A hierarchical control architecture is assumed, where a supervisory engine controller determines the combustion mode and the corresponding setpoints for all actuators, and all other control features strive to meet the demands set forth by the supervisory controller. Two different controller topologies, referred to as speed-dominant and air-fuel ratio dominant respectively, are developed to take advantages of the multi-mode nature of the DISC engine. Rapid completion of an LNT purge cycle was demonstrated while idling, even under considerable external load disturbances. In [55], idle speed is formalized as a constrained optimal control problem where fuel consumption is minimized. A sub-optimal, but easily implementable solution is obtained using a command governor.

### 3.4 Aftertreatment control and adaptation

To achieve the best tradeoff among competing requirements such as fuel economy, emissions and driveability, the LNT control strategy must manage the purge starting time, duration, and purge condition (such as $A/F$), and at the same time provide a bumpless transition between the lean and purge modes. The main challenges of LNT control stem from the lack of on-board measurements of key variables and the uncertainties in the characteristics of the key components. The NOx storage capacity of the LNT, one of the most critical parameters for control design and calibration, varies dynamically. In particular, the trap is susceptible to sulfur poisoning [56] and the capacity of the trap is reduced as sulfates accumulate. In addition, ambient conditions and component-to-component variations can affect the LNT operation and lead to deteriorated performance.

In the absence of real-time measurements, the control of the aftertreatment has to rely on feedforward and model-based control, making the system performance vulnerable to uncertainties and model inaccuracies. In [57], it is shown that the parameters of the LNT model [47] can be identified on-line using a conventional switching exhaust gas oxygen sensor. For the model structure and uncertainty representations used in [57], a nonlinear parametric model results. An on-line recursive algorithm is developed to improve the robustness of the model-based feedforward control and to ease the computational requirement of parameter
identification for the nonlinear parametric model. Persistent excitation, a condition normally required for parameter convergence, is established in [57] by changing purge thresholds.

In an effort to relax the computational intensity associated with the nonlinear parametric model used in [57], a new purge model [48] is exploited by the authors of [58] to develop an adaptive control strategy that is more feasible for real-time implementation in a computationally resource-constrained environment. By incorporating the physical properties of the system and properly choosing the structure for the LNT model and parameterization for the uncertainties, a linear parametric model is developed in [58] for on-line adaptation. Results show that, when integrated with model-based LNT control, the adaptation improves the aftertreatment control robustness by maintaining the desired tradeoffs between fuel economy and emissions.

### 3.5 System optimization and integration

For the DISC powertrain system incorporating NOx storage, a dynamic optimal control problem has to be formulated, because fuel consumption and emissions, evaluated over a specified driving cycle, are not simply functions of the instantaneous speed-load point, but of the operating history of the engine. The high degree of freedom introduced by the multiplicity of the control variables, coupled with time and trajectory dependency, leads to a very high dimension optimization problem. In [59] a method is introduced that dramatically reduces the computational burden of dynamic programming to make model-based design decisions for the lean-burn DISC powertrain. Results showing the sensitivity of the fuel economy performance objective at European Stage IV emission standards with respect to physical aftertreatment parameters, including the amount of oxygen storage in the TWC and the capacity of the lean NOx trap, are presented. In another trade-off study, control complexity is evaluated with respect to emissions benefit. Specifically, the optimal fuel economy, constrained by Stage III and Stage IV requirements, is evaluated to show the potential effects of eliminating the homogeneous lean combustion mode. It is determined, as illustrated in Fig. 7 of [46], that as NOx emission requirements become more stringent, the benefits of operating the engine in the homogeneous lean mode become less appreciable, up to a point where the incremental benefits may not be enough to justify the additional complexity.

The most important contributions of [59] are methodological. In particular, the computationally intense dynamic programming algorithm is rendered tractable by model simplification, state descretization, and analysis-based restriction on the search trajectories (called “calibrations”) along with careful treatment of computational details. The dynamic programming problem for a two-state system (TWC plus LNT) over an emissions drive-cycle was reduced to 40 minutes from 60 hours, while still achieving a near-optimal solution as shown in Fig. 7. These results are similar to the system optimization problems of hybrid vehicles, which will be discussed in more detail in Section 5. Stochastic dynamic programming and game-theoretic methods are explored for this purpose in [60,61].

Using dynamic programming, the authors of [62] also explore the benefits of air-fuel ratio profiling in achieving improved fuel economy, NOx and HC emissions tradeoffs. By allowing $A/F$ to vary during the purge phase, they show that substantial leverage can be achieved in reducing HC and NOx emissions, without a negative impact on fuel economy.

### IV. CONTROL OF AUTOMOTIVE DIESEL ENGINES

Diesel engines offer superior fuel economy compared to their conventional gasoline counterparts. Their drawbacks are associated with higher cost, and complexity of the aftertreatment system. Despite an earlier skepticism by even some of their developers, diesel engines have achieved a remarkable passenger car market penetration in Europe thanks to technology improvements. The consensus is that their penetration in North America will grow too, albeit at a slower pace due to differences in fuel cost and taxation.

![Fig. 7. Fuel economy versus NOx emissions of optimal policy with calibrations and full optimization over the Euro-cycle. The DISI engine and aftertreatment models are quasi-static. The LNT NOx filling and emptying is dynamically updated.](image)

---

1 Sir Harry Ricardo stated in 1925 that “...the exhaust from diesel engines ... has a characteristic pungent and disagreeable smell... the author cannot believe that the police will allow any large proportion of diesel-engined vehicles in the streets of, say, London.”
Diesel engines are typically turbocharged or supercharged to improve power density. A variable geometry turbocharger (VGT) enables optimal “sizing” of the turbine for each engine operating condition by opening or closing inlet guide vanes [63], resulting in both improved fuel economy and engine responsiveness. Electric boosting assist devices [64] have been developed for this purpose as well.

Diesel engines, operated on the compression ignition principle, have many different features compared to spark ignited gasoline engines. In particular, the following characteristics of diesel engines have strong control implications. First, they operate lean (\(A/F\) must usually stay above 22), and therefore require a different aftertreatment system. Second, NO\(_x\) control, to a much greater extent compared to conventional gasoline engines, relies on high EGR which, due to the lean operation, can contain significant amounts of combustible air. Third, the fueling rate is an independent and fast actuator for torque management, as long as the \(A/F\) is maintained within its limits. Modern common rail fuel injection systems permit fuel rate shaping and multiple injections per cycle for torque, noise and emission controls.

### 4.1 Diesel engine models

Mean value models and cylinder-by-cylinder diesel engine models have been utilized for control system design and validation. Mean value modeling of diesel engines has been covered in the review articles [65, 66] and in the book [6], while the cylinder-by-cylinder modeling is addressed in [66] and [67]. Different approaches to control oriented turbocharger modeling, including variable geometry turbochargers, are reviewed in the article [37]. References [68-70] explore the use of neural networks and related nonlinear identification techniques for diesel engine modeling.

A mean value model is developed in [71] for a diesel engine equipped with a VGT and an EGR valve. Compared to naturally aspirated gasoline engine models, diesel engine mean value models tend to be higher order. They capture the composition and temperature dynamics in the intake and exhaust manifolds and the turbocharger dynamics in addition to the manifold pressure dynamics. The engine torque is modeled as a static function of these states and inputs.

Cylinder-by-cylinder models predict cylinder pressure and engine torque with crank angle resolution. They use mass and energy balances to model the in-cylinder gas properties, in addition to manifold and turbocharger dynamics. In the simplest kinds of these models, the mass fraction of fuel burned is modeled as a function of the crank angle using Wiebe functions and the cylinder heat transfer is modeled using Hohenberg correlations. The intake and exhaust valve gas flows are modeled based on the orifice equations while the gas thermodynamic properties are captured using the Krieger-Borman relations. Reference [72] describes the use of a novel quadratic exponential fit for the mass of fuel burned and contains further references on the subject of cylinder-by-cylinder modeling. It also illustrates the use of a cylinder-by-cylinder model for a cylinder balancing application.

### 4.2 Control problems for diesel engines

Diesel engines provide many challenging control problems. The number of inputs (degrees of freedom) which needs to be dynamically controlled in a diesel engine ranges between 8 and 20, depending on the engine configuration. It can be even higher if individual cylinder behavior is taken into account. An increase in modeling, control and calibration complexity occurs with each added degree of freedom. Diesel engine dynamics are not only highly nonlinear but they are higher order than the ones for non-boosted gasoline engines. Static and dynamic interactions inherent to high order multi-input multi-output nonlinear systems complicate the control system development. Some of the control problems and pertinent solutions are briefly discussed here. The review articles [65, 73] and the book [74] also cover many of the aspects and literature on diesel engine control.

#### 4.2.1 Static and dynamic interactions

Figure 8 illustrates the effect of static interactions for the diesel engine with VGT and EGR valve. Note that at the operating point “b” when the EGR valve is fully open, opening the VGT results in an increase in the compressor flow. Exactly the opposite happens at the operating points “a” (when the EGR valve is closed) and “c” when the EGR valve is fully open and the VGT is open more than half way. This behavior is referred to as “dc gain reversal” and it complicates the control development [71, 75].

![Fig. 8. Steady-state dependence of compressor mass airflow, \(W_{c1}\), on VGT position, \(\chi_{VGT}\), for different positions of the EGR valve, \(\chi_{EGR}\).](image-url)
The dynamic phenomena important for control design have been illustrated in [71] where it is shown that the engine dynamics become slower when the EGR valve is more open, and that for the usual selection of outputs the system may exhibit non-minimum phase behavior. It is also shown through numerical optimal control-based analysis [75] that the optimal operating strategy of the VGT during a tip-in may not be its immediate closing (as the purely steady-state analysis would suggest). If the VGT is closed immediately during the tip-in, the exhaust pressure may increase rapidly in advance of the pressure increase in the intake manifold, thereby reducing the volumetric efficiency, increasing pumping losses, and increasing the turbo-lag. A more optimal operation of the VGT during this transient is to initially open it, then close it and reopen it again at higher rpm to prevent over-boost.

4.2.2 Selection of sensor configuration and control system architecture

In view of static and dynamic interactions in the diesel engine, the proper selection of sensor configuration and control system architecture is particularly important. Different internal variables may be used for feedback and they result in different levels of sensitivity to uncertainties and transient performance.

The simplest analysis procedure is to determine the steady-state sensitivities of key performance variables (such as fuel consumption and emissions) to the uncertainties for different sensor and controller configurations. The underlying assumption in this analysis is that a measured internal variable is maintained by the controller at the desired setpoint despite the effects of the uncertainties. In order for this analysis to lead to meaningful conclusions, the relative importance of performance variables and the expected size of uncertainties need to be established. Note also that the best sensor configuration or controller architecture may, in general, depend on the engine operating point, as was noted previously for DISC gasoline engines.

Other related procedures include the use of control-theoretic techniques such as Relative Gain Array (RGA) analysis [71] and $\mu$-analysis [76]. The value of $\mu$ is computed in [76] for different sensor configurations and at different operating points wherein low $\mu$ implied high robustness against uncertainties and small tracking errors. It is shown that although the numerical value of $\mu$ changes with the operating point, the relative ranking of the different configurations remains the same, thus permitting the identification of the best sensor configuration across the full engine operating range.

Besides formal procedures that consider the effect of uncertainties, the direct analysis of interactions and properties of the system may lead to an effective control architecture. In [34], the feedback architecture is designed based on consideration of available actuator authority at the optimal setpoints. It is shown that locally at these optimal setpoints, the EGR valve and the VGT become limited in their ability to independently affect the performance variables. This analysis led to a feedback controller architecture reliant on a single integrator instead of two. In reference [35], the exhaust pressure measurement is introduced to avoid the nonminimum-phase dynamics associated with the standard sensor configuration (compressor mass air flow and intake manifold pressure) and take advantage of the relative degree properties of the re-defined output set. This enabled application of effective robust nonlinear control design techniques. References [78, 79] propose combining switching logic and PID controllers to provide fast boost pressure response with small overshoot. Reference [80] utilizes an air-fuel ratio sensor positioned after the turbine and an LQG/LTR controller for the EGR valve in an engine with a conventional turbocharger. The use of the air-fuel ratio sensor can improve the system robustness and reduce calibration effort, although the transient performance may be limited due to the delay and sensor dynamics.

The guidelines resulting from numerical optimal control [75] can also be useful in comparing different controller architectures with each other in terms of their capability to generate an optimal behavior and for ease of subsequent controller calibration. For example, it is shown in [75] that the conventional decentralized architecture, wherein the VGT is controlled using a proportional plus integral feedback on intake manifold pressure and the EGR valve is controlled using a proportional plus integral feedback on the compressor mass air flow, is limited in its ability to generate the optimal behavior.

4.2.3 Coordinated EGR-VGT control

Coordinated control of the EGR valve and VGT has been a very active and recent research topic, with extensive literature on both linear and nonlinear control design approaches. Reference [81] compares several different linear and nonlinear control designs.

One of the controllers featured in [81] is a multivariable linear proportional-plus-integral (MIMO PI) controller for EGR valve and VGT position which uses the measurements of the intake manifold pressure and compressor mass air flow for feedback. This controller uses a decoupling transformation based on an inverse of the (static) gain of the plant for different operating conditions. Only 4 master gains need to be tuned on the engine while the decoupling transformation provides a mechanism for automatic gain scheduling.

Reference [35] develops a nonlinear controller for the diesel engine based on the method of Control Lyapunov Functions applied to a reduced order model of the diesel engine. The Control Lyapunov Function (CLF) is constructed as a Lyapunov function for the closed-loop system with a feedback linearizing controller; the CLF controller is then derived from the Lyapunov function for the desired mass flow rate of EGR and desired mass flow rate through the turbine. The EGR valve and turbine flow characteristics are inverted to backtrack the desired EGR valve and VGT.
positions from the desired flow rates. The CLF controller enjoys input uncertainty robustness properties such as infinite gain margin and 60 degree phase margin and highlights the advantages of using the exhaust manifold pressure measurement for feedback [35]. Reference [77] extends the CLF-based controller to a diesel engine model with delay using the method of Lyapunov-Krasovskiy functionals.

Authors of [82] propose to control the EGR valve using feedback on the error between estimated and requested cylinder fresh air flow while the controller for VGT is derived using feedback passivation ideas to enforce specified exhaust pressure dynamics. In addition, on-line parameter identification is employed to learn parameters in the cylinder flow and turbocharger models. Feedback passivation design using a master/slave approach is developed in [83]. A sliding mode controller is designed in [84] for the VGT and later extended to both EGR valve and VGT in [85]. A set of linear feedback controllers is designed in [86] and a switching logic is developed to control the engine response by selecting controllers in a sequence from this set. The design of each of the controllers in [86] relies on a polytopic representation of the model and the application of linear matrix inequality techniques. Reference [87] develops and implements a Model Predictive Control (MPC) algorithm for the coordinated control of EGR valve and VGT. It shows that the parameters in the cost function can be effectively used to shape the system transient response and demonstrates that the performance of the conventional controller has been either matched or exceeded. Bai and Yang [88] illustrate the benefits of a control algorithm which uses an estimate of cylinder air flow for feedback. Interactions between fueling and VGT is considered in [89]. It applies an inverse Nyquist array technique to analyze the interactions and design a controller for the system.

4.2.4 Composition estimation and fuel limiting

To avoid visible smoke emissions and reduce turbo-lag, a precise estimate of fresh air charge inducted into the engine cylinders is needed. The fueling rate can then be limited according to the fresh air charge estimate to maintain A/F above the smoke limit. The estimation of fresh air charge is complicated because the flow through the EGR valve and the gas mixture in both intake and exhaust manifolds contains both burned gas and fresh air.

Inasmuch as estimating the burned gas fraction is concerned, it is essentially unobservable from standard pressure and flow measurements in the diesel engine [71]. Therefore, an open-loop observer based on the burned gas fraction dynamic model [90], in combination with input observers [91], is used.

Charge estimation problems for diesel engines are studied in a number of other references. They include [92] which derives an adaptive observer for the cylinder flow in the diesel engine without EGR and demonstrates improvements over the conventional (open-loop) approach. Andersson and Eriksson [93] consider a related problem of the observer design for cylinder flow estimation in a diesel engine with a conventional wastegated turbocharger and without external EGR.

4.2.5 Aftertreatment control

Tailpipe NOx and particulate emissions (PM) represent particular challenges for diesel engines, because lean operation renders the conventional three-way catalyst ineffective. Much of present controls research is focused on the control of aftertreatment systems such as active lean NOx catalysts (ALNC), lean NOx traps (LNT), urea selective catalytic reduction (SCR), plasma catalysts and diesel particulate filters (DPF).

In an aftertreatment system with ALNC, engine fuel \( (i.e., \text{HC}) \) is injected upstream of the catalyst (typically by a special injector) to provide a reducing agent for the oxides of nitrogen in the ALNC. The control system must determine the quantity of the HC and control the temperature in order to maximize the ALNC conversion efficiency. The complicating factors are the hydrocarbon storage phenomenon in the catalyst and the interactions between hydrocarbon storage and temperature. A control oriented model for the ALNC is developed and extended in [94,95]. Dynamic programming is applied in [94] to generate a control law that minimizes the weighted sum of tailpipe NOx and spent fuel.

An LNT like that used in lean-burn gasoline applications can also be considered for diesel engine NOx control. This application, however, is particularly arduous as it has the same challenges faced by the lean-burn gasoline engine, in addition to the demands associated with the low operating temperatures of the diesel engine [95]. LNT temperature can be controlled with engine-based methods or by external methods, such as flow control devices in the exhaust and/or an oxidation catalyst placed upstream of the LNT. Each approach presents its own control challenges. Engine-based control has limited authority given competing objectives of fuel economy, performance and engine out emissions. Exhaust flow control devices involve additional hardware, including control valves, which increase cost and complexity, and introduce durability issues. An oxidation catalyst works well in a lean environment, but the duration of rich A/F conditions must be fairly short to avoid loss of authority.

A potential alternative to the LNT is SCR technology, where urea is injected upstream of a selective reduction catalyst [96]. Urea decomposes to ammonia, which serves as the reductant in the conversion of NOx. Accurate control of urea injection is critical for conversion efficiency and to avoid breakthrough of ammonia, which can lead to a foul odor at the tailpipe. The control problem is complicated by the transient nature of automotive applications. A control oriented model is developed in [97]. Observer based feedback control is implemented in [98], along with feedback from a NOx sensor. NOx measurement issues, including sensor sensitivity to ammonia, are discussed.
A DPF collects particulates emitted by the diesel engine. As particulates accumulate, backpressure increases, resulting in deteriorated fuel economy. To avoid the fuel economy loss, the DPF must be periodically regenerated by increasing its inlet temperature to a sufficiently high level to burn the stored particulates. Oxygen flow to the DPF must be carefully controlled during regeneration to avoid an over-temperature condition and damage to the DPF. The temperature increase can be achieved by fuel post-injection (i.e., injecting an extra amount of fuel late in the expansion stroke) and by coordinated control of the EGR valve, VGT and throttle to reduce the air flow through the engine. If an oxidation catalyst is available upstream of the DPF, injecting HC ahead of the catalyst creates an exothermic reaction which helps to increase DPF temperature. The key control problems for the DPF are estimating the soot level in the DPF (typically, from the measured pressure difference across the DPF), optimally deciding at which soot level to start regeneration, and controlling regeneration without affecting vehicle drivability and fuel economy or violating temperature limits for the DPF and oxidation catalyst. References [99] and [100] provide more background on the associated control problems.

V. ELECTRIC HYBRID POWERTRAIN SYSTEMS

Hybrid vehicles, especially hybrid electric vehicles (HEV), have demonstrated significant potential in reducing fuel consumption and exhaust emissions while maintaining driving performance. Hybrid powertrains may be viewed as a technology competing with variable valve timing, diesel, variable displacement and other fuel saving techniques. A natural question then arises: when would it make sense to choose a hybrid powertrain as opposed to other techniques (several of which are discussed in this paper)? Many “system-level” simulation studies have been conducted to compare the cost benefits of these techniques; see [101,102]. However, these system-level analyses are highly dependent on the underlying assumptions, such as fuel cost, and may not be that useful for predicting the future benefits and cost of ownership. Therefore, we will focus on a discussion of the fundamental performance benefits of hybrid electric powertrains.

By reviewing the design philosophy and functionality of existing HEVs, it is apparent that HEVs offer a few unique attributes in comparison to other engine-centric fuel saving techniques: (i) regenerating braking—energy that would otherwise be lost—which is only possible because a reversible secondary power source is present; (ii) component down-sizing or right-sizing—which is possible only when a competent secondary power source is present in parallel; and (iii) the fuel economy improvement (up to 100%) that has been demonstrated for hybrid vehicles. This improvement is available partly because of the first two attributes, and partly because of the control algorithm that properly coordinates the operation of the multiple power sources.

Due to the fact that a hybrid powertrain provides significantly increased flexibility, it is possible to size the components and integrate them together to achieve vastly different design targets. For example, for smaller passenger cars, which are more likely to be driven in an urban environment, fuel economy can be given the highest priority. For SUVs, on the other hand, improved launch performance (0~60 time) can be a decisive issue for a purchaser. For luxury sedans, the possibility of greatly improved NVH (noise-vibration-harshness) may be more important than the other potential benefits.

When fuel economy is the main design goal, as a general rule of thumb, a driving environment with lower average speed and frequent acceleration/deceleration is likely to see higher improvement. Larger vehicles (e.g., a large SUV) will probably see larger and faster market penetration, compared with smaller vehicles, because of their more favorable fuel saving returns.

5.1 Typical hybrid architectures and associated control issues

HEVs in general are classified into series, split and parallel hybrids; see Fig. 9. The performance potential of these different configurations and their associated control problems are quite different. For series hybrids, the mechanical power from the internal combustion engine is converted immediately to electrical form by a large generator. The electrical power is then distributed to the wheels with greater flexibility than with mechanical power.
employs heuristic control techniques, such as control rules, be roughly classified into three categories. The first type can be designed (e.g., [103,104]). A more sophisticated algorithm can be designed (e.g., [105]), but the likely improvement in fuel economy will be relatively small. Consequently, the control algorithms of most series HEVs have been designed on the basis of simple rule-based methods. In general, there is no or little component down-sizing problem that needs to be considered together with the control design due to the series configuration.

For parallel hybrids, a secondary power source exists in parallel with the ICE, thereby offering a greater level of flexibility in configuration, component sizing, and control. When the secondary power source is small ("mild" hybrids), the control problem becomes much simpler, as the two power sources do not operate simultaneously. The development effort has focused on hardware packaging and component efficiency (106,107). Key control decisions relate to the timing of engine start/stop and the execution of regenerative braking [108]. When the secondary power source is large ("strong" hybrids), the situation becomes much more interesting. The ultimate design procedure would involve the solution of the optimal design (component sizing) and optimal control problems simultaneously. In this paper, however, we will only discuss the solution of the optimal control problem, assuming that all the components have already been selected.

The third type of hybrid vehicle is the so-called split type. The most well-known examples include the Toyota Hybrid System [109] (used in the Prius, the Estima minivan, and the RX400H) and the Allison Transmission Electric Drives System [110]. Both of these hybrid systems use planetary gear(s) as the power summation device as well as the means to provide torque ratios, thereby eliminating power loss in transmission. Two electric motor/generators are used as the secondary power sources to sustain favorable operating conditions for the ICE as well as to augment the engine driving torque to satisfy the driver’s demand. The control of split-type hybrids, like their parallel counterparts, is frequently done on the basis of rules-based algorithms (e.g., [111]). The authors are not aware of any publications using optimal control techniques for split-type hybrids.

5.2 Control strategy development for parallel HEVs

Power management strategies for parallel HEVs can be roughly classified into three categories. The first type employs heuristic control techniques, such as control rules, fuzzy logic, and neural networks, for estimation and control algorithm development [112,113]. With these methods, the control designer must use his or her engineering judgement or experience to address the myriad tradeoffs presented by having multiple power sources and sinks. The second approach is based on static (point-wise-in-time) optimization methods. In this method, electric power is commonly translated into an equivalent (steady-state) fuel rate in order to calculate the overall fuel cost ([114,115]). The optimization scheme then determines the proper split between the two energy sources using steady-state efficiency maps. Because of the point-wise-in-time nature of the optimization problem, it is possible to extend such schemes to solve the simultaneous fuel economy and emission optimization problem [116]. The third approach to HEV control strategy development considers the dynamic nature of the system components—and the drive cycle—when performing the optimization ([117,118]). In particular, the optimization is with respect to a time horizon or time interval, rather than an instant in time.

Computational burden is a potential barrier to the widespread use of dynamic optimization in hybrid vehicles. While much work remains to be done in this area, progress is being made. Reference [119] reports on the results of a head-to-head comparison of a popular rule-based load-leveling approach to control law design for a parallel hybrid electric truck, versus a dynamic optimization method developed in [120]. On the same hardware, with testing conducted by an independent group, the rule-based strategy resulted in a fuel economy improvement of 31% and feedgas NOx reduction of 50%, whereas the strategy derived from dynamic optimization resulted in a 45% fuel economy improvement and feedgas NOx reduction of 54%. Vehicle drivability is similar in each case to the non-hybrid version of the vehicle.

In order to provide a better understanding of what is known and what needs to be discovered, the results of [120] and [121] will be overviewed in more detail.

5.2.1 Deterministic dynamic optimization over a drive cycle

This section describes an indirect method for dynamic optimization [120], with application to HEV control strategy development. It consists of setting up a deterministic dynamic programming problem over a specific drive cycle (vehicle speed versus time). The resulting optimal control policy requires advance knowledge of the drive cycle and is thus not implementable on an actual vehicle (the policy is non-causal). Nonetheless, analysis of the behavior of the optimal control policy for judiciously chosen initial conditions both on and off of the drive cycle yields near-optimal rules, which are implementable. The process is indirect because the user must carry out rule extraction on the basis of the non-causal optimal control policy.

The development of a dynamic vehicle model is the first step in the control design process. Typically, the model is developed in two stages. In the first stage, available component models are assembled with appropriate switching logic to represent the chemical, electrical, and
mechanical power paths in the vehicle, plus emissions production. The component models are typically a combination of ordinary differential equations, time delays, and maps (or tables) regressed against data. Low-level control laws in the electric motor, transmission, brakes, engine, etc., must also be included. The overall vehicle model is usually of fairly high order and, when “driven” over a test cycle, is assumed to accurately reflect the performance variables of interest to the designer. Hence, this model is called the detailed model.

A detailed model is not suitable for dynamic optimization because computation time grows exponentially with the number of states: “the curse of dimensionality”. Thus, the second stage of modeling is aimed at finding a simplified but sufficiently accurate vehicle model. Developing and validating the simplified model is a difficult process requiring extensive engineering judgement. It may be the most crucial step in the development of the control policy. The reference [120] develops a simplified model for a parallel hybrid electric truck, consisting of a V6 (5.5L) diesel engine, a 49 kW DC electric motor, and an 18 amp-hour valve-regulated lead-acid (VRLA) battery. Using the rule of thumb that when evaluating fuel economy and emissions over a long driving cycle (tens of minutes), dynamics that are faster than 1 Hz can be safely ignored, it was determined that a sufficiently accurate model could be constructed with only three state variables: the vehicle speed, transmission gear number, and battery state of charge (SOC). The simplified model is time-wise discretized at a sample period of 1 sec., and expressed as:

\[ x(k+1) = f(x(k), u(k)), \]

where \( u(k) \) is the vector of control variables such as desired output power from the engine, desired output power from the motor, and gear shift command to the transmission and \( x(k) \) is the state vector of the system.

The optimization goal is to find a charge-sustaining control policy that minimizes a weighted sum of fuel consumption and emissions over a given driving cycle

\begin{equation}
J(x_0) = \min_{u \in U} \sum_{k=0}^{N-1} [L(x(k), u(k))] + G(x(N))
\end{equation}

where \( N \) is the duration of the driving cycle, and \( L(x, u) \) is the instantaneous cost, including fuel use and engine-out (feedgas) \( NO_x \) and particulate matter \( (PM) \) emissions; \( G \) is a final-state penalty \(^3\) on terminal SOC, where \( SOC_f \) is the desired final SOC; and \( U \) is the set of control decisions that meet the vehicle speed equality constraint imposed by the drive cycle, plus a number of inequality constraints that ensure safe/smooth operation of the engine, battery, and motor; see [120]. For a fuel-only problem, the weighting factors are \( \mu = \nu = 0 \). The case of \( \mu > 0 \) and \( \nu > 0 \) represents a simultaneous fuel and emission problem. The optimal control policy is a time-varying state variable feedback, \( u^*(x, k) \) [122]. Numerical procedures for computing the optimal policy via dynamic programming are well known [122]. A numerical implementation of the dynamic programming algorithm described in [59], based on spatial discretization and interpolation, is used in [120] to compute the optimal policy for a parallel hybrid diesel truck, over a number of different drive cycles.

The optimal control policy itself cannot be implemented because it depends on the drive cycle (the control policy is non-causal or anticipative). However, the optimal feedback creates a family of optimal paths for all possible initial conditions of the model (20). By simulating the optimal policy for a range of initial conditions, it is possible to extract rules that are implementable. This indirect feedback design method of first formulating and solving a finite-horizon dynamic optimization problem over a fixed drive cycle and then extracting implementable rules is summarized in Fig. 10.

As an illustration of how rules are extracted from \( u^*(x, k) \), define the power split ratio \( (PSR) \) as \( PSR = \frac{P_{eng}}{P_{req}} \), which can be used to quantify the positive power flows in the powertrain, where \( P_{eng} \) is the engine power and \( P_{req} \) is the power request from the driver (that is, the power required for the vehicle to follow the drive cycle). Four positive-power operating modes are defined: motor-only \( (PSR = 0) \), engine-only \( (PSR = 1) \), power-assist \( (0 < PSR < 1) \), and recharging \( (PSR > 1) \). Figure 11 shows the result of plotting the power split ratio determined by the optimal policy versus the ratio of the requested power and transmission speed. Since the optimal points (dots) group nicely\(^4\) when plotted against the ratio of the requested power and transmission speed, regression (solid line) yields a rule for power split that is time invariant, near optimal,

\[ PSR = \begin{cases} \frac{P_{req}}{P_{eng}} & \text{if } PSR \leq 1 \\ \frac{1}{PSR} & \text{if } PSR > 1 \end{cases} \]

\[ \text{where } \alpha = 252 \]

\[ \text{and } \mu = \nu = 0. \]

\[ \text{The factors to use in regression were determined via subset selection, with forward selection and backward elimination.} \]

\[ \text{with } \text{forward selection and backward elimination.} \]

\[ \text{with } \text{forward selection and backward elimination.} \]

\[ \text{with } \text{forward selection and backward elimination.} \]

\[ \text{with } \text{forward selection and backward elimination.} \]

\[ \text{with } \text{forward selection and backward elimination.} \]

\[ \text{with } \text{forward selection and backward elimination.} \]

\[ \text{with } \text{forward selection and backward elimination.} \]
Fig. 11. An example of extracting the power split ratio (PSR) from the optimal control policy. This is for the UDDSHDV cycle. Similar functions are required for gear selection and regenerative braking. The ‘art’ in the extraction process is determining good regressors.

and easily implemented on the vehicle. A different choice of drive cycle would yield a different optimal policy, and thus different data (dots in Fig. 11) for extracting a rule for power split. However, reference [120] shows that, for the parallel hybrid truck under study, the power split ratio of Fig. 11 performs well over several common drive cycles when evaluated on the detailed model.

Even though the control laws obtained with the indirect method have performed well in a real hybrid electric vehicle [119], there are two drawbacks to this approach. First, this approach optimizes with respect to a specific driving cycle and might be neither optimal nor charge-sustaining under other cycles; secondly, the feedback solution to the deterministic dynamic optimization problem is not directly implementable and the rule extraction process can be time consuming. To overcome these drawbacks, a design procedure based on stochastic dynamic optimization is overviewed next.

5.2.2 Stochastic dynamic optimization

A direct method for dynamic optimization of hybrid powertrains has been presented in [121]. The key ideas are (1) to model the power requested by the driver, which is the equivalent of a drive cycle, as a stationary, finite-Markov chain, and (2), to formulate the optimization objective as an infinite-horizon, discounted-cost, stochastic dynamic programming problem. Specifically, the objective is to find a control policy \( \pi = (\pi_0, \pi_1, \cdots) \) that minimizes

\[
J_\pi(x_0) = \lim_{N \to \infty} E \sum_{k=0}^{N-1} \gamma^k g(x(k), \pi_k(x(k)), u(k)),
\]

for a model of the form

\[
x(k+1) = f(x(k), u(k), w(k)),
\]

where \( E \) is the expectation operator, \( w \) is a random variable from the stationary Markov chain model of the drive cycle, \( g(x, u, w) \) is the instantaneous cost, and \( 0 < \gamma < 1 \) is the discount factor. Under reasonable hypotheses, the optimal control law always exists and has the form of a time-invariant full-state feedback [122], and therefore can be directly implemented on the vehicle.

Just as in the deterministic approach, a simplified model is mandatory for computing the optimal policy (again, the curse of dimensionality), and a detailed model is desirable for evaluating the effectiveness of the strategy. Additional modeling effort is required to represent the planned vehicle use, that is, the drive cycle, as a stationary Markov chain. An illustration of the control design process on the parallel hybrid electric truck is presented in [121]. An illustration on a hybrid fuel cell vehicle (HFCV) is presented in [143]. As seen in [121] and [143], very ‘realistic’ random driving patterns can result from a Markov power-demand model. The method has not yet been evaluated on hardware.

5.2.3 Discussion on dynamic optimization

As opposed to deterministic optimization over a given driving cycle, the stochastic approach optimizes the control policy over a broader set of driving patterns: the best policy achieves a minimum of the expected cost, which is an average over all sample paths of the stochastic model. In other words, a benefit of this approach is that the control law is never a ‘cycle beater’. A second important benefit of the stochastic approach is the direct generation of an implementable feedback policy. This obviates the tedious process of extracting implementable rules.

The current formulation of the stochastic approach has several drawbacks as well. One is that future costs are discounted. This is done for mathematical expediency and is difficult to justify on engineering grounds. Since the control policy is optimal in an expected sense, even if the cost represents cumulative fuel and emissions, no guarantees on performance can be made for a given sample path of the model. In other words, even if the Markov model of power demand accurately represents the statistics of a particular government mandated drive cycle, a vehicle operated with an optimal control that meets the required emissions in an average sense, could fail the emissions test over the deterministic cycle. Hence, it would be desirable to solve the stochastic version of the problem with a deterministic performance constraint. Finally, computational techniques need to be developed to allow for use of a higher order model in optimization. The current limitation seems to be about three state variables, which makes it impossible to include a dynamic model of the aftertreatment system, for example.
VI. FUEL CELL BASED POWERTRAIN SYSTEMS

Fuel cells, as promising alternative power plants to internal combustion engines, have been pursued feverishly in recent years. In particular, Proton Exchange Membrane (PEM) fuel cells have been under intensive development for automotive applications. Considerable progress has been made on fuel cell system modeling, control design and system integration. In this section, we provide a brief overview of the main control challenges and relevant results, and draw readers’ attention to pertinent literature.

Unlike internal combustion engines, the PEM fuel cell is an electrochemical device that converts oxygen and hydrogen to electrical power, with water and heat as the only byproducts. As such, fuel cell based power systems are the ultimate clean power sources and hold great promise for automotive applications. There are, however, a number of technical hurdles, controls included, in making the fuel cell system a viable powertrain for automotive systems. To meet the robustness and reliability requirements for transient mobile applications, and to compete with internal combustion engines in both performance and cost, the fuel cell system has to be optimally integrated and effectively controlled to perform dependably under a wide range of operating conditions.

A schematic diagram of a fuel cell system and its main auxiliary components is shown in Fig. 12. The main sub-systems include the fuel cell stack, hydrogen and air supply systems, cooling system, humidification system, and the power conditioning system. Many fuel cell control problems have been discussed in [123,124]. In the subsequent discussion, we will highlight the key features of the control oriented models, the main control problems and the characteristics of the associated solutions.

6.1 Control oriented fuel cell models

Control oriented fuel cell models refer to those low order, phenomenological representations that capture both the nonlinear steady-state characteristics and the low frequency dynamic behavior from the control inputs to the performance variables. As illustrated in Fig. 12, the control inputs include fuel flow from the tank, air flow or compressor input, current drawn from the fuel cells, and control actuation for the temperature and humidity control systems. The performance variables are often concerned with the cell voltage, partial pressure of the air and fuel in the cathode and anode respectively, membrane humidity and temperature.

Several fuel cell models published in the literature have facilitated many successful control designs and applications [125-127]. In developing these models, electrochemical, thermodynamic and zero-dimensional fluid flow principles are used to characterize the dynamical and nonlinear fuel cell behavior. The electrical performance of PEM fuel cells is represented by a polarization curve, where the output voltage is a function of the current density, partial pressure of the reactants (oxygen and hydrogen), temperature and humidity. Electrochemical losses due to ohmic resistance, activation and concentration are accounted for in the polarization characteristics. In an attempt to minimize complexity and facilitate model-based control design, most of the control oriented models treat the fuel cell as a flow network consisting of lumped parameter volumes and pressure drops along the flow path. Mass and energy balances, together with other thermodynamic and fluid principles, are used to calculate the partial pressure of air, fuel and water in the reactant supply channel and in the cathode and anode. Water content, in both vapor and liquid states, is tracked by accounting for the water entering and leaving the stack, and that being produced with chemical reactions.

Different fuel cell concepts, such as high-pressure and low-pressure fuel cells, lead to different performance characteristics and operating constraints. While most of the system models are developed for high-pressure fuel cell models, a low-pressure system with an air blower has also been explored and its model and dynamic analysis are reported in [128]. Fuel cell systems integrated with fuel processing technology for mobile applications have also been investigated, and models have been developed and reported, for example, in [129,130].

Well regulated stack temperature and humidity are typically assumed in most of the system-level modeling efforts described in [125-127]. Other activities concentrating on specific phenomena such as water diffusion and transportation have led to other special purpose models [131-133]. Intensive studies are still underway to understand and characterize the complicated mechanisms and phenomena associated with water diffusion and transport across the membrane and along the reactant channel [124]. Much of the effort on CFD (computational fluid dynamic) modeling is expected to shed light on the humidity treatment of fuel cell systems at the microscopic level and aid in control oriented model development and control design [134].


6.2 Reactant supply and control

Fuel cells rely on the continuous supply of oxygen and hydrogen to maintain their optimal and safe operation. As electric current is drawn from the fuel cell stack, reactants (air and hydrogen) are consumed. Due to the dynamics present in the delivery system, consumed reactants cannot be replenished instantaneously, causing possible fuel cell oxygen or hydrogen starvation [125-136]. Reactant starvation not only leads to cell performance degradation, but also possible permanent membrane or bipolar plate damage, and therefore it has to be strictly avoided. On the other hand, excessive reactant supply adds parasitic losses to the system, thus reducing the overall efficiency. For fuel cell systems connected to a fuel processing system, excessive anode fuel not only causes a drop in efficiency, but can also lead to other environmental concerns if the anode exhaust is released to the atmosphere. Therefore, delivering the optimal reactants and achieving satisfactory load following performance are the key control objectives of the reactant supply system.

In [137], a detailed dynamic analysis of reactant supply systems is presented, together with an analysis of the control implications. The analysis reveals the transient performance limitations of the reactant supply systems, and provides guidelines for performance trade-off (between fast response and starvation protection) and sensor configuration selection.

Several control methodologies have been investigated to eliminate or mitigate the reactant starvation in the fuel cell, such as passive filtering of the load command [138], the use of a load governor for constraint enforcement [135], and model predictive control [136]. The need to protect the fuel cell from starvation while meeting the load following requirements strongly motivates the research on hybrid fuel cell based vehicles, where either a battery or a super capacitor is used to assist the primary power plant in its transient operation [139-142]. The power management problems and the optimization techniques are similar to those discussed in the previous section [143].

Another issue pertinent to the reactant supply control is sensor requirements. Given the cost-conscious nature of automotive applications, it is often desirable to minimize the number of sensors. Measuring the hydrogen for real-time control is not only prohibitive from the cost point of view, but also difficult from the technical perspective. Virtual sensing for fuel cell control using an observer has been explored in [125,144], where a model-based state estimation scheme is developed to support the sophisticated control implementation.

6.3 Temperature and humidity control

The PEM fuel cell membrane’s capability in conducting protons and thus producing electricity depends critically on the water content. As its water content decreases, the ionic conductivity of the membrane decreases, thereby leading to reduced cell electrical efficiency. Furthermore, this decreased electrical efficiency causes increased heat production and water evaporation, which in turn reduces the water content even further. Conversely, excessive water stored in the electrode obstructs fuel flow, resulting in flooding. Keeping an optimal temperature and humidity condition in the stack is thus critical to maintaining the efficient and safe operation of the cell.

Temperature control for the fuel cell system is challenging in several aspects. First of all, since the PEM stack is, compared to internal combustion engines, operating at a relatively low temperature of around 80°C, not much heat can be carried out through the fuel cell exhaust. Therefore, most of the heat rejection responsibility falls on the cooling system. Second, the heat transfer between the stack and water coolant is largely limited by the small temperature differential, given the low operating temperature of the stack. In addition to the coolant system, active cooling through the reactant flow and reactant inlet temperature control is often required to achieve effective temperature control. Finally, the temperature control system is expected to achieve fast stack warm-up without overshooting, while minimizing the power consumption of the cooling fan and coolant pump.

Modeling and control of the humidity of the fuel cell is a very complicated task, since the water vapor generation, transportation and condensation is a multi-phase process and involves many different mechanisms. It also has to be carried out in close coordination with temperature management, reactant flow control, and other fuel cell subsystem controls. Active and efficient humidity control will also depend on the availability of on-board humidity measurement. Given the difficulty in sensing the stack humidity, an accurate humidity model is very desirable. Several attempts have been reported in the literature. In [131], a lumped parameter model is developed to quantify the average vapor mass transport across the fuel cell and thus to predict the temperature and humidity. In [133] modeling and analysis are carried out for a fuel cell humidifier system suitable for automotive applications. These models, in combination with various flooding prevention solutions proposed in the literature (such as that in [145]), provide promising tools in tackling the humidity issues of the PEM fuel cell system. Nonetheless, due to the lack of reliable sensors and comprehensive models, membrane humidity modeling and control remain a weak link in fuel cell control, and intensive research is still underway.

It is important to note that humidity and temperature dynamics are inter-related phenomena and thus cannot be treated separately. It is generally believed that external controllable humidifiers are critical to ensure that the relative humidity of the inlet reactants is adequately controlled over a wide range of operating conditions of stack current, stack temperature and ambient conditions. Whether an external humidifier is used or not, the stack membrane humidity can be affected by several mechanisms. Excessive liquid water inside the cathode can be removed either by increasing the
excess ratio of the air, or by lowering the inlet air relative humidity. Similarly, increasing the water vapor in the stack can be accomplished either by humidifying the inlet air or by varying the flow rate of the humidity source (such as the stack exhaust air or cooling water). Excessive water inside the anode can also be removed by recirculation or purge. Any of these attempts to influence the humidity will also lead to changes in the temperature of the stack and of the inlet air, and possible changes in reactant flow. The authors are not aware of any publication reporting coordinated control of stack humidity, temperature and flow.

VII. CONCLUSIONS

Powertrain control has been, and remains a dynamic and exciting research subject. Advanced powertrain systems have served as benchmark problems for testing and evaluating many advanced control theories, methodologies, and development processes. The need to develop more reliable and efficient systems under stringent timing and cost constraints has motivated new algorithms, more efficient computational and design tools, and innovative control sensor/actuator designs. The rich literature surveyed in this paper is a testimony to the progress made by the powertrain control community, and we hope it serves to inspire new interests and research activities in this very important technological area.

ACKNOWLEDGEMENTS

The authors would like to thank their colleagues at Ford Research Laboratory and the University of Michigan, as well as those in the powertrain control community at large, for many years of collaboration and interaction. Professors Jessy Grizzle and Huei Peng also wish to acknowledge many stimulating discussions with their former and current PhD students, Jun Mo Kang, Chan-Chiao (Joe) Lin, and Ed Tate, on the topics of dynamic optimization and vehicle control.

REFERENCES

27. Watson, N. and M.S. Janota, “Turbocharging Concepts for Downsized DI Gasoline En-


Jeffrey Cook is a Technical Leader at the Ford Motor Company Research and Innovation Center. He is an Adjunct Professor of Electrical Engineering and Computer Science at the University of Michigan. His research addresses modeling and control of automotive powertrains for improved fuel economy and emissions, and improvements in systems engineering processes for the design of automotive controls. He holds 25 patents on automotive technology. He is a Fellow of the IEEE and one of three recipients of the 2003 IEEE Control Systems Technology Award. He is a Subject Editor for the International Journal of Robust and Nonlinear Control.

Jing Sun received her Ph.D. degree from University of Southern California in 1989, and her B.S. and M.S. degrees from University of Science and Technology of China in 1982 and 1984, respectively. From 1989-1993, she was an assistant professor in Electrical and Computer Engineering Department, Wayne State University. She joined Ford Research Laboratory in 1993 where she worked in the Powertrain Control Systems Department. After spending almost 10 years in industry, she came back to academia and joined the faculty of the College of Engineering at University of Michigan in 2003 as an associate professor. Her research interests include system and control theory and its applications to marine and automotive propulsion systems. She holds over 30 US patents and has co-authored a textbook on Robust Adaptive Control. She is an IEEE Fellow and one of the three recipients of the 2003 IEEE Control System Technology Award. She is also a subject editor for International Journal of Adaptive Control and Signal Processing.

Julia Buckland received B.S. and M.S. degrees in Aerospace Engineering from the University of Cincinnati in 1988 and 1993 respectively. She worked several years in the defense industry developing flight control systems for advanced military aircraft. In 1994, she joined Ford Motor Company, where she is currently a technical expert in the Powertrain Controls R&A Department. Her work focuses on modeling and control of advanced technology powertrain systems.

Ilya V. Kolmanovsky studied as an undergraduate at Moscow Aviation Institute, Russia. He received the M.S. degree and the Ph.D. degree in aerospace engineering, in 1993 and 1995, respectively, and the M.A. degree in mathematics, in 1995, all from the University of Michigan, Ann Arbor. He is currently a Technical Leader in Powertrain Control at Ford Research & Advanced Engineering of Ford Motor Company, Dearborn, MI. In addition to expertise in the automotive engine and powertrain control, his research interests include potential of advanced control techniques as an enabling technology for advanced automotive systems, and several areas of control theory, which include constrained control, optimization-based and model-predictive control, and control of nonlinear mechanical, nonholonomic, and underactuated systems.


Huei Peng received the Ph.D. degree in mechanical engineering from the University of California, Berkeley, in 1992. He is currently a Professor with the Department of Mechanical Engineering, University of Michigan, Ann Arbor. His research interests include adaptive control and optimal control, with emphasis on their applications to vehicular and transportation systems. He has been an active member of SAE and the ASME Dynamic System and Control Division. Dr. Peng has served as the chair of the ASME DSCD Transportation Panel from 1995 to 1997. He is currently an Associate Editor for the IEEE/ASME Transactions On Mechatronics. He received the National Science Foundation (NSF) Career award in 1998.

Jessy W. Grizzle received the Ph.D. in electrical engineering from The University of Texas at Austin in 1983. Since September 1987, he has been with The University of Michigan, Ann Arbor, where he is a Professor of Electrical Engineering and Computer Science. He jointly holds fourteen patents dealing with emissions reduction through improved control system design. He received the Paper of the Year Award from the IEEE Vehicular Technology Society in 1993, was elected a Fellow of the IEEE in 1997, received the 2002 George S. Axelby Award and the 2003 Control Systems Technology Award.