

# Integrated Model-Based and Data-Driven Diagnosis of Automotive Antilock Braking Systems

Jianhui Luo, *Member, IEEE*, Madhavi Namburu, Krishna R. Pattipati, *Fellow, IEEE*, Liu Qiao, and Shunsuke Chigusa

**Abstract**—Model-based fault diagnosis, using statistical hypothesis testing, residual generation (by analytical redundancy), and parameter estimation, has been an active area of research for the past four decades. However, these techniques are developed in isolation, and generally, a single technique cannot address the diagnostic problems in complex systems. In this paper, we investigate a hybrid approach, which combines model-based and data-driven techniques to obtain better diagnostic performance than the use of a single technique alone, and demonstrate it on an antilock braking system. In this approach, we first combine the parity equations and a nonlinear observer to generate the residuals. Statistical tests, particularly the generalized likelihood ratio tests, are used to detect and isolate a subset of faults that are easier to detect. Support vector machines are used for fault isolation of less-sensitive parametric faults. Finally, subset selection (via fault detection and isolation) is used to accurately estimate fault severity.

**Index Terms**—Antilock braking systems (ABSs), data-driven diagnosis, model-based diagnosis, nonlinear systems, parameter estimation, residuals, support vector machines (SVMs).

## I. INTRODUCTION

THE INCREASED complexity of automobiles has made system monitoring an inevitable component of future vehicles. There is a concomitant increased difficulty in the identification of the malfunction phenomena (subsystem failure modes and fault location ambiguity caused by cross-subsystem failure propagation). Consequently, the current rule-based monitoring systems used in the automotive industry are difficult to develop, validate, and maintain. The current practice in automotive diagnosis in the dealerships (or maintenance facilities) is based on reading diagnostic troubleshooting codes (DTCs), referring to the maintenance manuals based on observed DTCs, and then performing troubleshooting. In some cases, a single failure source can trigger more than four DTCs, and a single DTC can be caused by as many as 14 failure sources. In addition, current troubleshooting devices can only read static

“frame data” (e.g., engine revolutions per minute, temperature, etc.). The static data are unable to diagnose dynamic behavior of complex faults. This results in “trial-and-error” methods for troubleshooting, which, in turn, leads to long testing times, increased maintenance costs, and customer dissatisfaction. A continuous monitoring and early warning capability that detects and isolates faults and relates detected degradations in vehicles to accurate remaining lifetime predictions will minimize downtime, improve resource management via condition-based maintenance, and minimize operational costs.

The recent advances in sensor technology, remote communication and computational capabilities and standardized hardware/software interfaces are creating a dramatic shift in the way the health of vehicle systems is monitored and managed. The availability of data (sensor, command, activity, and error-code logs) collected during nominal and faulty conditions, coupled with intelligent health management techniques, ensure continuous vehicle operation by recognizing anomalies in vehicle behavior, isolating their root causes, and assisting vehicle operators and maintenance personnel in executing appropriate remedial actions to remove the effects of abnormal behavior. There is an increasing trend toward online real-time diagnostic algorithms embedded in the electronic control unit (ECU), with the DTCs that are more elaborate in reducing cross-subsystem ambiguities. The maintenance technician will be using an intelligent scanner with optimized and adaptive test procedures (e.g., test procedures generated by test-sequencing software, e.g., [1]) instead of precomputed static-paper-based decision trees, and maintenance logs (cases) will be recorded automatically. If the technician cannot isolate the root cause, the history of sensor data and symptoms will be transmitted to maintenance center for remote diagnosis. Future practice of automobile maintenance would be improved significantly due to these advances. In [2], Luo *et al.* provide a vehicle health-management architecture based on agent technologies and remote diagnosis.

The automotive industry has adopted quantitative simulation as a vital tool for a variety of functions, including algorithm design for ECUs, rapid prototyping, programming for hardware-in-the-loop simulation, production code generation, and process-management documentation. Applications of this model-based design include engine control and automatic transmission systems. With the availability of mathematical models and computer-aided engineering information, it is natural to integrate diagnostic processes into the initial design phase for vehicle health management. In parallel, research in model-based and data-driven fault diagnosis has gained increasing attention over the past four decades [3]. The major techniques

Manuscript received April 9, 2005; revised January 7, 2006, May 29, 2006, February 16, 2007, and June 11, 2008. First published November 17, 2009; current version published February 18, 2010. This work was supported by Toyota Technical Center. This paper was recommended by Associate Editor G. Biswas.

J. Luo was with the Department of Electrical and Computer Engineering, University of Connecticut, Storrs, CT 06269-2157 USA. He is now with Qualtech Systems, Inc., Wethersfield, CT 06109 USA (e-mail: jianhui.luo@gmail.com).

M. Namburu, L. Qiao, and S. Chigusa are with the Technical Research Department, Toyota Technical Center, Ann Arbor, MI 48105 USA (e-mail: setumadhavi.namburu@tema.toyota.com; liu.qiao@tema.toyota.com; shunsuke.chigusa@tema.toyota.com).

K. R. Pattipati is with the Department of Electrical and Computer Engineering, University of Connecticut, Storrs, CT 06269-2157 USA (e-mail: krishna@engr.uconn.edu).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TSMCA.2009.2034481

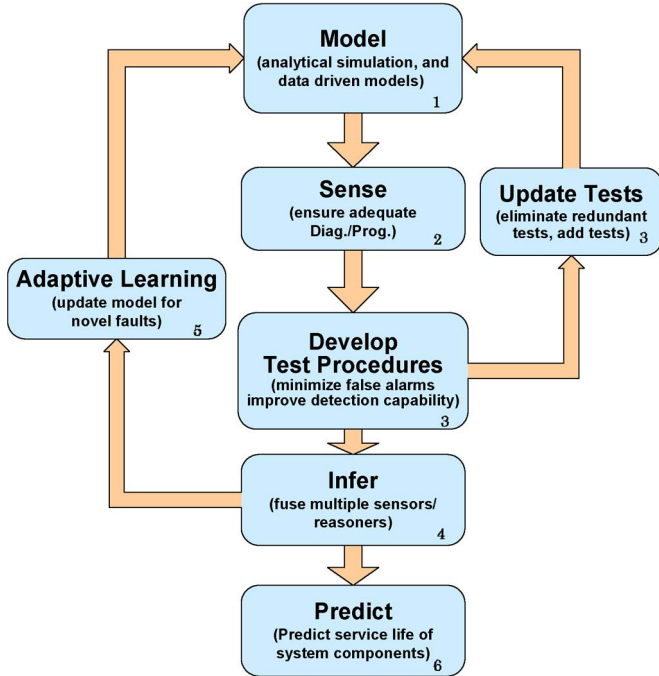


Fig. 1. Block diagram of intelligent diagnosis process.

for model-based diagnosis include parameter estimation, parity equations, output observers [4]–[6], and bond graph models [7]. Data-driven fault-detection-and-isolation (FDI) techniques have close relationship with pattern recognition, wherein one seeks to categorize the input–output data into normal or faulty classes. The most notable techniques for data-driven diagnosis include principal component analysis [8], Fisher discriminant analysis [9], partial least squares [10], and support vector machines (SVMs) [11]–[13]. While the use of different techniques to solve engineering problems is a classical approach, there is no “silver bullet” for fault diagnosis. Model-based and data-driven techniques provide the “sand box” that test designers can experiment with and systematically select relevant algorithms/techniques to solve their own FDI problems based on requirements on diagnostic accuracy, computational speed, memory, online versus offline diagnosis, and so on.

In [14], we have developed an intelligent diagnostic process for vehicle systems. Fig. 1 shows the block diagram of this iterative process. It contains six major steps: 1) model; 2) sense; 3) develop and update test procedures; 4) infer; 5) adaptive learning; and 6) predict. In [15], we have demonstrated step 6: predicting the service life of system components for an automotive suspension system. In this paper, we will focus on steps 1 through 4, and apply them to a realistic antilock braking system (ABS) model, which extends our work in [14]. There are very few research papers on advanced techniques for fault diagnosis in ABS. Pisu *et al.* [16] presented a model-based FDI scheme for sensor faults of a brake-by-wire system. They use linear observers to construct residuals. Borner *et al.* [17] illustrated model-based fault detection of solenoid ABS valves. A residual is constructed for armature stroke based on a simplified model (first-order differential equation) of solenoid valves. In this paper, we present an integrated model and data-driven diagnostic scheme that combines parity equations, nonlinear observer, and

SVM to diagnose faults in an ABS. This integrated approach is necessary, since neither model nor data-driven strategy could adequately solve the entire FDI problem for the ABS, i.e., isolate faults accurately.

This paper is organized as follows. Section II presents a model of the ABS. Section III provides an overview of the FDI scheme for the ABS. Section IV presents the generalized likelihood ratio tests (GLRTs) used for fault detection. Sections V-A and B discuss residual generation via parity equations and the nonlinear observer. Section VI presents the fault-isolation method based on SVM. Section VIII describes the parameter-estimation technique for fault-severity estimation. Finally, this paper concludes with a summary in Section IX.

## II. VEHICLE ABS MODEL

An important step in model-based fault-diagnosis design is the development and validation of a proper mathematical model of system dynamics. Although a detailed model is desired, it may be too complicated to use for fault diagnosis. For example, due to the limited computational power and memory of the automotive ECUs, it is difficult to achieve real-time diagnosis using detailed physics-based models of system dynamics. Thus, a design model [18], which captures the essential features of the system for FDI, should be considered.

Our development of the ABS model, shown in Fig. 2, follows that of [19]. We consider longitudinal braking with no steering and neglect the effects of pitch and roll. The model considers the wheel speed and vehicle speed as measured variables and the force applied to the brake pedal as the input. The wheel speed is directly measured, and vehicle speed can be calculated by integrating the measured acceleration signals, as in [20].

### A. Sliding-Mode Control

The control objective of the ABS is to regulate wheel slip ( $\lambda$ ) and to maximize the coefficient of friction  $\mu$  between the wheel and the road [21]. A sliding-mode controller design is adopted for the ABS because of its robustness to noise and disturbances. The sliding-mode variable  $S$  is defined as

$$S = e_\lambda + \alpha \cdot \dot{e}_\lambda \quad (1)$$

where  $e_\lambda = \lambda^* - \lambda$ ,  $\lambda^*$  is the target slip, and  $\alpha$  is a positive constant. The variable  $\dot{e}_\lambda$  is the derivative of  $e_\lambda$ , which is approximated via first-order forward difference as

$$\dot{e}_\lambda = \frac{de_\lambda}{dt} \approx \frac{e_\lambda(k+1) - e_\lambda(k)}{t_s} \quad (2)$$

where  $t_s$  is the sampling interval.

In order to avoid chattering associated with sliding-mode switching control laws, the switching control *sign* function is replaced by the pulsewidth-modulation (PWM) function as in [21]

$$\text{sat}(S) = \begin{cases} 1, & \text{if } S \geq \phi \\ \frac{S}{\phi}, & \text{if } S < |\phi| \\ -1, & \text{if } S \leq -\phi \end{cases} \quad (3)$$

where the boundary layer  $\phi$  is a small positive constant.

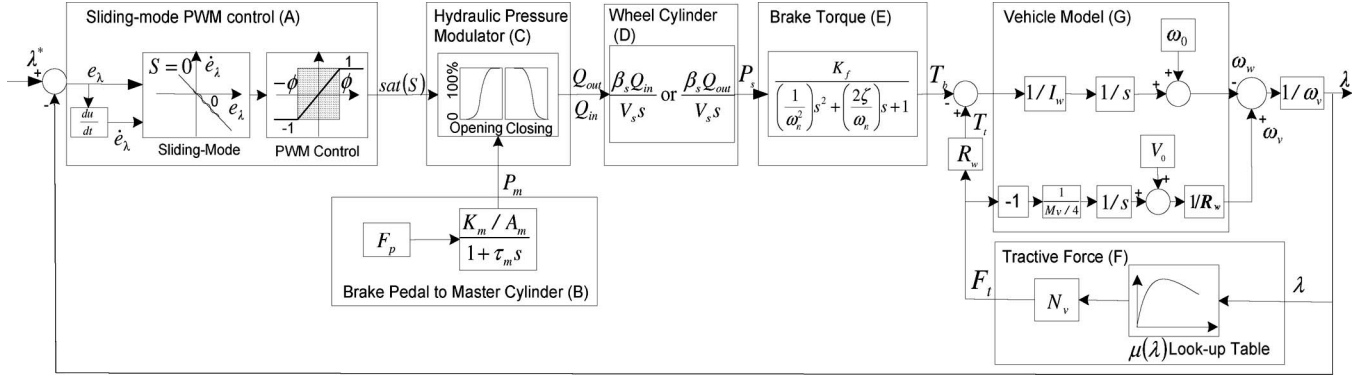


Fig. 2. Simulink block diagram of the ABS.

### B. Brake Pedal to Master Cylinder

The transfer function of brake force applied to brake pedal  $F_p$  is modeled as

$$F_p(s) = \frac{K_p}{1 + \tau_p s} \quad (4)$$

where  $\tau_p$  is the time constant and  $K_p$  is the steady-state force. The transfer function between  $F_p$  and the force applied to master cylinder ( $F_m$ ) is modeled as

$$\frac{F_m(s)}{F_p(s)} = \frac{K_m}{1 + \tau_m s} \quad (5)$$

where  $K_m$  is the steady-state gain and  $\tau_m$  is the time constant of the master cylinder. The oil pressure of the master cylinder  $P_m$  is modeled via an algebraic relation  $P_m = F_m/A_m$ , where  $A_m$  is the cross-sectional area of the master cylinder.

### C. Hydraulic Pressure Modulator

The pressure modulator consists of solenoid valves, brake oil pump, and low-pressure oil tanks. When the pressure is built up, the pump pushes the brake oil in the low-pressure tank back to the brake line. Solenoid valves are modeled by pressure build-up and release processes.

During the pressure build-up process ( $\text{sat}(S) > 0$ ), the brake oil flows from the master cylinder through the inlet throttle to the wheel cylinder. The flow of brake oil is modeled as flow through a restriction [22]. Denoting  $Q_{\text{in}}$  as the inflow rate to wheel cylinder, we obtain

$$Q_{\text{in}} = \text{sat}(S) C_{\text{in}} A_{\text{eff}} \sqrt{\rho(P_m - P_s)} \quad (6)$$

where  $C_{\text{in}}$  is the inlet discharge coefficient of the throttle,  $A_{\text{eff}}$  is the effective area of the throttle,  $\rho$  is the density of brake oil, and  $P_s$  is the brake oil pressure in wheel cylinder.

During the pressure-release process ( $\text{sat}(S) < 0$ ), the brake oil flows from the wheel cylinder through the outlet throttle to the low-pressure oil tank. Denoting  $Q_{\text{out}}$  as the outflow rate from the wheel cylinder, we obtain

$$Q_{\text{out}} = -\text{sat}(S) C_{\text{out}} A_{\text{eff}} \sqrt{\rho(P_s - P_{\text{lp}})} \quad (7)$$

where  $C_{\text{out}}$  is the outlet discharge coefficient of the throttle and  $P_{\text{lp}}$  is the pressure in the low-pressure tank. Here,  $P_{\text{lp}}$  is modeled as a constant.

The effective area of the valve is approximated by a hyperbolic function to reflect the opening and closing processes [19]. When the valve is opened, the effective area  $A_{\text{eff}}$  is

$$A_{\text{eff}} = A_0 (0.5 + 0.5 \cdot \tanh((t_{\text{op}} - t_{\text{opdelay}})/t_{\text{oprate}})) \quad (8)$$

When the valve is being closed, the area is

$$A_{\text{eff}} = A_0 (0.5 - 0.5 \cdot \tanh((t_{\text{cl}} - t_{\text{cldelay}})/t_{\text{clrate}})) \quad (9)$$

where  $A_0$  is the maximum open area of the valve;  $t_{\text{op}}$ ,  $t_{\text{cl}}$  are the operational times of the valve;  $t_{\text{opdelay}}$ ,  $t_{\text{cldelay}}$  are the time delays of the valve actuation; and  $t_{\text{oprate}}$ ,  $t_{\text{clrate}}$  are the parameters that determine how fast the valve is opened and closed, respectively.

### D. Wheel Cylinder

If the brake oil flows into the wheel cylinder through the inlet throttle, the dynamics of pressure  $P_s$  is modeled as

$$\dot{P}_s = \beta_s \frac{Q_{\text{in}}}{V_s} \quad (10)$$

where  $\dot{P}_s$  is the derivative of oil pressure inside the wheel cylinder,  $\beta_s$  is the effective bulk modulus, and  $V_s$  is the volume of the wheel cylinder.

If the brake oil flows out of the wheel cylinder, then

$$\dot{P}_s = -\beta_s \frac{Q_{\text{out}}}{V_s} \quad (11)$$

### E. Brake Torque

The transfer function between the brake torque and the brake pressure is of second order [23]

$$\frac{T_b(s)}{P_s(s)} = \frac{K_f}{(1/\omega_n^2)s^2 + (2\zeta/\omega_n)s + 1} \quad (12)$$

where  $T_b(s)$  and  $P_s(s)$  are the Laplace transforms of brake torque and wheel cylinder pressure, respectively,  $\zeta$  is the damping ratio,  $K_f$  is the torque gain, and  $\omega_n$  is the natural frequency.

### F. Tractive Force

The tractive force between the tire and the road surface is given by

$$F_t(t) = N_v \mu(\lambda) \quad (13)$$

where  $N_v$  is the normal tire reactive force. The friction coefficient  $\mu$  depends on the tire–road conditions and the value of the wheel slip  $\lambda$ . In this paper, we use a static 1-D lookup table for a dry pavement condition. This is not a restriction, since the current development of optoelectronic sensors can detect the presence of ice, water, or mud on the road surface condition using backscattered and reflected-light information [24]. The wheel slip is defined by

$$\lambda(t) = \frac{\omega_v(t) - \omega_w(t)}{\omega_v(t)} \quad (14)$$

where  $\omega_w(t)$  is the angular velocity of the wheel (in radians per second) and  $\omega_v(t)$  is the angular velocity of the vehicle defined as

$$\omega_v(t) = V_v(t)/R_w \quad (15)$$

where  $V_v(t)$  is the velocity of the vehicle and  $R_w$  is the radius of the wheel.

### G. Vehicle Model

The dynamic equations of quarter-car vehicle model are the result of Newton's law applied to the wheels and the vehicle [25]. The dynamics of vehicle speed is modeled as

$$\dot{V}_v(t) = \frac{-1}{M_v/4} F_t(t) = \frac{-1}{M_v/4} N_v \mu(\lambda) \quad (16)$$

where  $V_v(t)$  is the velocity of the vehicle and  $M_v$  is the mass of the vehicle. The dynamic equation of the wheel speed is determined by summing the rotational torques that are applied to the wheel via

$$\dot{\omega}_w(t) = \frac{1}{I_w} [T_t(t) - T_b(t)] \quad (17)$$

where  $I_w$  is the rotating inertia of the wheel,  $T_b(t)$  is the braking torque, and  $T_t(t)$  is the tractive torque between the wheel and the road surface. Here,  $T_t(t)$  is a function of  $F_t(t)$  and is expressed as

$$T_t(t) = R_w F_t(t) = R_w N_v \mu(\lambda). \quad (18)$$

In summary, the state-space model of the vehicle brake system is a six-state continuous-time nonlinear dynamic system with two sampled measurements as follows:

$$\begin{aligned} \dot{x}_1 &= \frac{1}{I_w} [R_w N_v \mu(\lambda) - x_3] + w_1 \\ \dot{x}_2 &= -\frac{4}{R_w M_v} N_v \mu(\lambda) + w_2 \\ \dot{x}_3 &= x_4 + w_3 \\ \dot{x}_4 &= \omega_n^2 [K_f x_5 - 2\zeta x_4 / \omega_n - x_3] + w_4 \\ \dot{x}_5 &= \begin{cases} \beta_s \frac{Q_{in}}{V_s} + w_5, & \text{if } sat(S) > 0 \\ -\beta_s \frac{Q_{out}}{V_s} + w_5, & \text{if } sat(S) \leq 0 \end{cases} \\ \dot{x}_6 &= \frac{K_m}{A_m \tau_m} F_p - \frac{1}{\tau_m} x_6 + w_6 \\ \mathbf{y}(t_k) &= \begin{bmatrix} x_1(t_k) \\ x_2(t_k) \end{bmatrix} \begin{bmatrix} v_1(t_k) \\ v_2(t_k) \end{bmatrix} \end{aligned} \quad (19)$$

TABLE I  
SUMMARY OF ABS MODEL PARAMETER VALUES

Parameter	Value	Units	Parameter	Value	Units
$\alpha$	0.01	—	$t_s$	1e-3	s
$\phi$	0.02	—	$K_p$	4.2	—
$\tau_p$	2e-2	s	$K_m$	5	—
$\tau_m$	0.02	s	$A_m$	2.5e-3	m <sup>2</sup>
$C_{in}$	2e-4	—	$\rho$	910	kg/m <sup>3</sup>
$C_{out}$	4e-4	—	$P_{lp}$	100	MPa
$A_0$	2.5e-3	m <sup>2</sup>	$t_{opdelay}$	4e-3	s
$t_{oprate}$	3e-4	s	$t_{cldelay}$	0.2	s
$t_{clrate}$	2e-4	s	$t_{cldelay}$	0.2	s
$\beta_s$	1.72e+9	MPa	$V_s$	2.5e-4	m <sup>3</sup>
$K_f$	1e+2	—	$\omega_n$	7.07	s <sup>-1</sup>
$\zeta$	3.54e+2	—	$I_w$	5	kgm <sup>2</sup>
$R_w$	0.318	m	$\lambda^*$	0.2	—

where

$x_1$  : wheel speed,  $\omega_w$ ;  $x_2$  : vehicle speed,  $\omega_v$   
 $x_3$  : brake torque,  $T_b$ ;  $x_4$  : derivative of  $T_b$ ,  $\dot{T}_b$   
 $x_5$  : oil pressure in wheel cylinder,  $P_s$   
 $x_6$  : oil pressure in master cylinder,  $P_m$   
 $v_1, v_2$  : measurement noises for states  $x_1$  and  $x_2$   
 $w_1 \sim w_6$  : process noise variables  
 $t_k = kt_s$ ,  $t_s$  = sampling interval

where

$x_1$  : wheel speed,  $\omega_w$ ;  
 $x_2$  : vehicle speed,  $\omega_v$ ;  
 $x_3$  : brake torque,  $T_b$ ;  
 $x_4$  : derivative of  $T_b$ ,  $\dot{T}_b$ ;  
 $x_5$  : oil pressure in wheel cylinder,  $P_s$ ;  
 $x_6$  : oil pressure in master cylinder,  $P_m$ ;  
 $v_1, v_2$  : measurement noises for states  $x_1$  and  $x_2$ ;  
 $w_1 \sim w_6$  : process noise variables  $t_k = kt_s$ ,  $t_s$  = sampling interval.

The equation for state  $x_1$  is obtained by substituting (18) into (17). The equation for state  $x_2$  is obtained by substituting (16) into (15). The dynamic equations for  $x_3$  and  $x_4$  are given by (12). The dynamic equation for  $x_5$  is obtained by combining (10) and (11). Finally, the equation for  $x_6$  is obtained using (5).

Table I lists the parameter values for the ABS model. To validate our model, we compared its performance with the SIMULINK 3-D handling model constructed using CarSimED, which is a product of the Mechanical Simulation Corporation. CarSimED's model includes detailed kinematical and compliance effects of the suspension and steering systems [26]. The same brake torque was applied to our model and to the CarSimED vehicle system dynamics model. In addition, the same controller is applied to both systems with a desired slip of 0.2. The responses of both models are shown in Fig. 3. Our model follows very closely the behavior of the detailed and complex CarSimEd model, except that in CarSimEd, the wheel



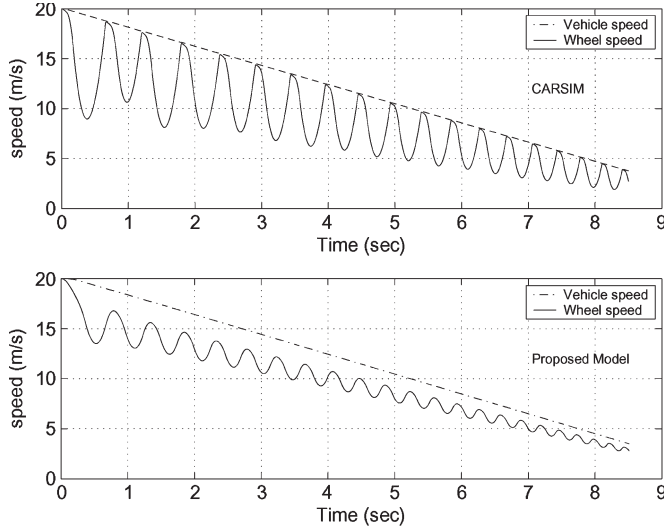


Fig. 3. Comparing the response of the model in (19) and the CarSimEd dynamic model.

speed is the same as vehicle speed when the brakes are cycled, which results in slip almost equal to zero. Our model output has better control performance, since slip is regulated closer to 0.2. Since failure modes cannot be specified in the CarSim model, fault-model-based comparison is not pursued. Encouraged by the good performance of our model, in the following sections, we will illustrate the FDI scheme and each individual FDI technique using the ABS model presented in this section.

### III. FDI SCHEME

#### A. Fault Universe

One commonly occurring sensor fault and four parametric faults are considered in this paper. There are some other components faults, such as those in the hydraulic valve (e.g., valve stuck closed or open) and the ABS ECU faults, which are not considered in this paper. These are easily detectable faults by the current diagnostic troubleshooting codes (DTCs). In the case of a wheel-speed-sensor fault, the sensor systematically misses the detection of teeth in the wheel due to incorrect wheel-speed-sensor gap caused by loose wheel bearings or worn parts. In order to model the wheel-speed-sensor fault (F1), we consider two fault-severity cases: greater than 0% but less than 5% reduction in the nominal wheel speed (F1.1), and greater than 5% reduction in the nominal wheel speed (F1.2). The four parametric faults (F2–F5) we considered are changes in  $R_w$ ,  $K_f$ ,  $I_w$ , and  $\tau_m$ . Notice that fault F2 models the tire pressure fault, F3 and F5 correspond to cylinder faults, while F4 is related to vehicle body. Faults corresponding to more than 2% decrease in  $R_w$  are considered. We distinguish between two  $R_w$  faults: greater than 2% but less than 20% (F2.1) decrease in  $R_w$ , and greater than 20% decrease in  $R_w$  (F2.2). The sizes for  $K_f$  and  $I_w$  faults considered are as follows:  $\pm 2\%$ ,  $\pm 3\%$ ,  $\dots$ ,  $\pm 10\%$ . The size for  $\tau_m$  fault corresponds to a more than 15% increase in the time constant. Table II shows the list of considered faults. The minimum fault magnitude is

TABLE II  
SIMULATED FAULTS LIST

F1.1	sensor fault ( $\pm 2\% \sim \pm 5\%$ decrease)
F1.2	sensor fault ( $\geq 5\%$ decrease)
F2.1	$R_w$ fault ( $\pm 2\% \sim \pm 20\%$ decrease)
F2.2	$R_w$ fault ( $\geq 20\%$ decrease)
F3	$K_f$ fault ( $\pm 2\% \sim \pm 10\%$ )
F4	$I_w$ fault ( $\pm 2\% \sim \pm 10\%$ )
F5	$\tau_m$ fault ( $\geq 15\%$ increase)

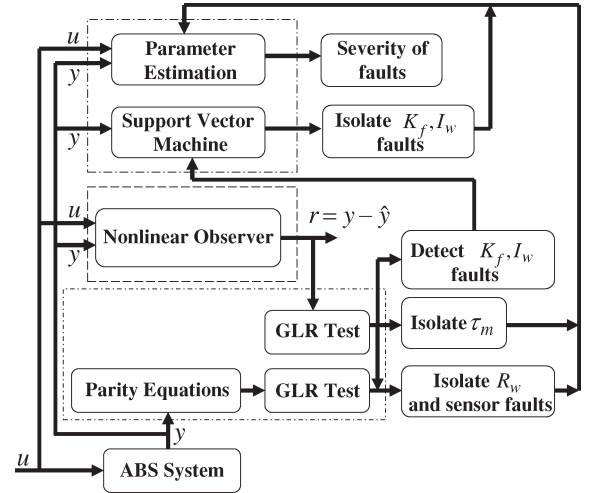


Fig. 4. FDI scheme for the ABS.

selected such that changes in the residual signals cannot be detected if we choose fault magnitude less than this minimum. The measurement noise variables for vehicle and wheel speed are white with variances of 0.004 each. The process noise variables are white with variance of 0.5% of the mean-square values of the states they are added to (which corresponds to a signal-to-noise ratio of +23 dB). This scaling follows the Fitt's law in psychophysics [27]. A small amount of process noise is added based on the fact that these states are driven by disturbances from chaotic combustion processes in the engine (unmodeled dynamics of wheel and vehicle speeds) and nonlinear effects in the ABS actuator (for brake torque and oil pressure).

#### B. FDI Scheme for the ABS

Figs. 4 and 5 shows the block diagram of our proposed FDI scheme for the ABS. The parity equations and GLRTs are used to detect severe  $R_w$  ( $\geq 20\%$ ) and wheel-speed-sensor ( $\geq 5\%$ ) faults. Then, a nonlinear observer is used to generate two additional residuals. The GLRTs based on these two residuals are used to isolate the  $\tau_m$  fault, less severe (small)  $R_w$ , and sensor faults. They are also used to detect  $K_f$  and  $I_w$  faults. Finally, we use the SVM to isolate the  $K_f$  and  $I_w$  faults. After parametric faults are isolated, an output-error method is used to estimate the severity of isolated faults. The rationale for this architecture and details of the FDI scheme are elaborated as follows.

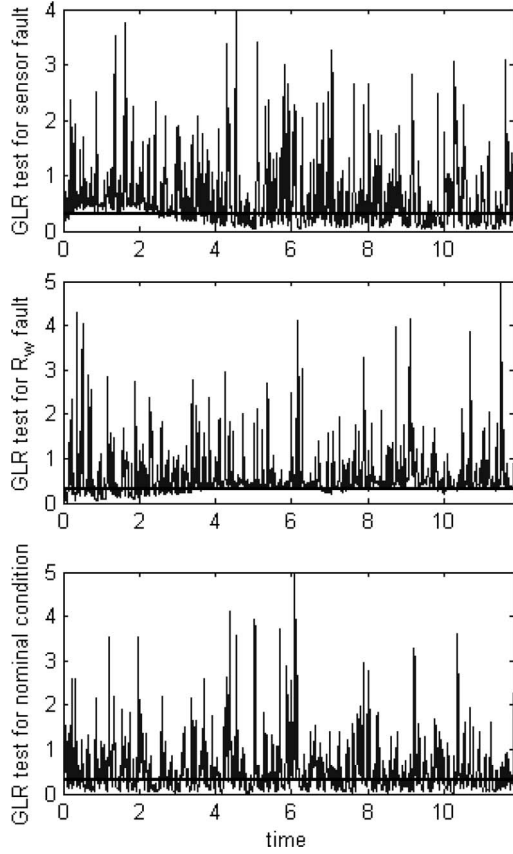


Fig. 5. GLRT with  $R_w$  and sensor faults for  $r_1$  generated by parity equation.

#### IV. GLRT-BASED FAULT DETECTION

A moving window GLRT [28] is used to detect changes in residuals. In this case, the mean  $\mu_0$  and variance  $\sigma^2$  before the change are known, and the mean  $\mu_1$  after the change is unknown. The log-likelihood ratio for residuals in a moving time window from time  $t_j = jt_s$  up to time  $t_k = kt_s$  (window size  $W = k - j + 1$ ) is

$$R_j^k(\mu_1) = \sum_{i=j}^k \ln \frac{p_{\mu_1}(r(i))}{p_{\mu_0}(r(i))}. \quad (20)$$

The ratio in (20) is a function of two unknown parameters: the change time and the value of mean after change. The standard statistical approach is to use the maximum likelihood estimates of these two parameters via

$$g_k = \max_{1 \leq j \leq k} \sup_{\mu_1} R_j^k(\mu_1). \quad (21)$$

We implemented the GLRT for a change in the mean, since the residual was found to be almost Gaussian under normal conditions. Then, the log-likelihood ratio in (20) simplifies to

$$R_j^k = \frac{\mu_1 - \mu_0}{\sigma^2} \sum_{i=j}^k \left( r(i) - \frac{\mu_1 + \mu_0}{2} \right) \quad (22)$$

where  $r(i)$  is the residual at time  $t_i = it_s$ . Denoting  $\eta = \mu_1 - \mu_0$ , the estimate of  $\eta$  is

$$\hat{\eta}_j = \frac{1}{k - j + 1} \sum_{i=j}^k (r(i) - \mu_0) \quad (23)$$

and the decision function in (21) is

$$g_k = \frac{1}{2\sigma^2} \max_{1 \leq j \leq k} \frac{1}{k - j + 1} \left[ \sum_{i=j}^k (r(i) - \mu_0) \right]^2. \quad (24)$$

An offline simulation is performed for nominal conditions to obtain  $\mu_0$  and  $\sigma^2$  for the residuals with a sample size  $N = 1250$ . When the GLRT is applied to the actual system, the mean  $\mu_0$  and variance  $\sigma^2$  can be computed from the recorded residuals under nominal conditions. Adjustable thresholds may be selected for the GLRT to declare a change, once the decision function exceeds the threshold for a specified time period. This minimizes false alarms at the cost of increased detection delay. Because of the nonlinearity and noise effects, the thresholds for GLRTs are tuned manually by inspecting simulation results. Due to model simplifications, the thresholds of tests may need to be retuned manually on an operational ABS. Unlike the embedded diagnostic logic in an ECU, a nice feature of the model-based approach is that the inference and model knowledge (fault and test parameters) can be separated; when thresholds change, all one needs to do is upload the model parameters (i.e., data) to the vehicle without changing the diagnostic software. The method for generating the residuals in the ABS is discussed next.

#### V. CONSTRUCTING RESIDUALS FOR FDI

In this section, two types of residuals are constructed for FDI. The first type is based on parity equations, while the second type is generated using nonlinear observers. This section provides the details of test design using these two types of residuals.

##### A. Parity-Equation-Based Residuals

Parity equations are rearranged forms of the input–output or state-space models of the system [3]. This approach checks for the consistency of inputs and outputs. Parity equations require less computational effort, since they only need algebraic computations based on measured inputs/outputs and their estimated derivatives. Other analytical redundancy techniques, such as an observer, need numerical integration (e.g., fourth-order Runge–Kutta formulas [29], also termed ode4), which is more expensive (and also more robust to noise). Algorithms to automatically generate the parity equations have been developed for differential algebraic equations [30], [31]. Moreover, automated approaches for nonlinear polynomial models are developed by Staroswiecki and Comtet-Varga [32], although they do not apply to the ABS problem.

For a continuous dynamic system with discrete-time measurements (such as in the ABS), the discrete-time residual equations require the estimates of time derivatives from sampled data. If the measurements are noisy, the noise amplification of the time derivatives of sensor data restricts this method to parity equations having third-order derivatives or less [33]. The derivative of noisy data also restricts this approach to problems with relatively large-sized faults. Conatser *et al.* [34] presented an FDI scheme for parametric faults in automotive electronic throttle control (ETC) system based on second-order parity equations. The range of parametric change in ETC is between 10% and 30% of nominal values. For the ABS in (19), parity-equation-based residuals can be used to detect gross faults: greater than 20% decrease in  $R_w$  and greater than 5% decrease in wheel-speed-sensor readings based on the fault simulation results.

Using the ABS model in (19), the first parity equation is obtained directly from the state equation for  $x_2$ . The second parity equation is obtained by eliminating the states  $x_3$  and  $x_4$  in (19) for  $\dot{x}_4$ . The two parity equations are

$$r_1(\mathbf{y}, \theta^1) = \dot{y}_2 + 4N_v\mu(\lambda)/(R_w M_v) \quad (25)$$

$$r_2(\mathbf{y}, \theta^2) = \ddot{x}_3 - \omega_n^2 K_f x_5 + 2\zeta\omega_n \dot{x}_3 + \omega_n^2 x_3 \quad (26)$$

where  $x_3 = -\dot{y}_1 I_w - R_w N_v \mu(\lambda)$  and  $x_5$  is constructed by integrating the right-hand side of state equation  $\dot{x}_5$  in (19) without considering process noise and with zero initial condition (because the brake pressure is zero before ABS is activated). The system parameters are as follows:  $\theta^1 = (R_w)$ ,  $\theta^2 = (K_f \ I_w \ \tau_m)$ . The other parameters are assumed to be fixed. Evidently, the two parity equations require the first- and third-order derivatives of measured variables, respectively.

In order to suppress noise, which may corrupt the estimation of derivatives, the measurement  $\mathbf{y}$  is filtered with a low-pass tenth-order Butterworth filter with a cutoff frequency of 10 Hz [35]. The cutoff frequency is determined from the largest eigenfrequency of wheel-speed measurement. Cubic spline regression is used to fit the measurements and is used to estimate the derivatives of measurements. However, in the presence of noise, the magnitude of second residual for the nominal case is greater than  $10^{11}$ . This causes this residual to contain almost no information about faulty conditions, since the faulty trends are immersed in the huge magnitude of the residual generated by noise.

Equation (25) can be written in the form of a linear regression, and the parameter  $R_w$  can be estimated by linear least square estimation by minimizing  $\sum_{i=1}^N \|r_1(\mathbf{y}, \theta^1)\|^2$ , where  $N$  is the number of samples. Equation (26) is a nonlinear function of  $\theta^2$ . Thus,  $\theta^2 = (K_f \ I_w \ \tau_m)$  can be estimated by nonlinear least squares by minimizing  $\sum_{i=1}^N \|r_2(\mathbf{y}, \theta^2)\|^2$ . However, the parameter estimates of nonlinear least squares are biased; the same phenomenon is reported in [14] and [36]. The biased estimates stem from the nonlinearity of the model, the measurement noise, higher order derivatives in the second parity equation, and closed-loop feed back control. Since we

have no *a priori* knowledge on how the parameter-estimation method based on parity equations will perform, this analysis based on parity equations was necessary to show that, although this method is computationally efficient, it does not work for the ABS system. Therefore, the residual  $r_2$  will not be investigated further. To obtain better parameter estimates, output-error-based parameter-estimation method should be used and will be illustrated in Section VIII.

### B. Nonlinear-Observer-Based Residuals

To overcome the drawback that measurements need to be differentiated in parity-equation-based residual-generation methods, a nonlinear-observer-based approach is used to generate two additional residuals [37], [38]. Because this approach is more robust to noise than the parity-equation-based approach, we can design tests that detect smaller variations in  $R_w$ , the wheel-speed-sensor values, and the  $\tau_m$  parameter than the parity-equation-based approach. Here, we employ a Luenberger observer for nonlinear time-invariant systems containing separable nonlinearities [39]. The dynamic equations for a nonlinear system with separable nonlinearities are

$$\begin{aligned} \dot{\mathbf{x}} &= \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{u} + \mathcal{NL}(\mathbf{x}, \mathbf{u}) \\ \mathbf{y}(k) &= \mathbf{C}\mathbf{x}(t_k) + \mathbf{D}\mathbf{u}(t_k). \end{aligned} \quad (27)$$

A system with separable nonlinearity of order  $n$  is observable in its linear part if the observability matrix

$$\mathbf{Q} = [\mathbf{C} \ \mathbf{F}^T \mathbf{C} \ \dots \ (\mathbf{F}^T)^{n-1} \mathbf{C}] \quad (28)$$

is full rank [40]. Here,  $\mathbf{F} = \exp(\mathbf{A}t_s)$ , where  $t_s$  is the sampling interval. For the ABS, the assumption of observability for a full-order observer design is not satisfied. Since our interest is to design an observer for  $[x_1, x_2]$  only, we can overcome this problem by designing a reduced-order observer. As can be seen from the system model in (19), it is a concatenation of two independent subsystems: the dynamics of  $x_1-x_5$  and the dynamics of  $x_6$ . Because the state  $x_6(P_m)$  is not coupled with other states, we can use  $P_m$  [generated by (4) and (5)] as an input ( $u$ ) and simplify the reduced subsystem to the form of separable nonlinearities as

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathcal{NL}(\mathbf{x}, u) \quad (29)$$

where  $\mathbf{x}$  includes the first five states in (19). The matrix  $\mathbf{B}$  is equal to  $\mathbf{0}$ . The matrix  $\mathbf{A}$  and the separable nonlinearities  $\mathcal{NL}(\mathbf{x}, u)$  are

$$\begin{aligned} \mathbf{A} &= \begin{bmatrix} 0 & 0 & \frac{-1}{I_w} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & -\omega_n^2 & -2\zeta\omega_n & \omega_n^2 K_f \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \\ \mathcal{NL}(\mathbf{x}, u) &= \begin{bmatrix} \frac{mgR_w}{4I_w} \mu(\lambda) \\ -\frac{g}{4R_w} \mu(\lambda) \\ 0 \\ 0 \\ \beta_s \frac{Q_{in}}{V_s} \text{ or } -\beta_s \frac{Q_{out}}{V_s} \end{bmatrix}. \end{aligned} \quad (30)$$

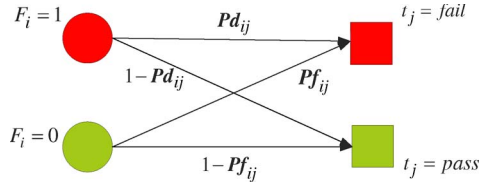


Fig. 6. Detection false-alarm probabilities of failure source  $F_i$  and test  $t_j$ .

Let

$$\mathbf{y}^h(t) = \sum_{k=1}^{\infty} \mathbf{y}(kt_s) [U(t - kt_s) - U(t - t_s - kt_s)] \quad (31)$$

where  $U(\cdot)$  is the unit step function. For the earlier system, we design a Luenberger observer

$$\dot{\hat{\mathbf{x}}} = \mathbf{A}\hat{\mathbf{x}} + \mathcal{NL}(\hat{\mathbf{x}}, u) + \mathbf{L}(\mathbf{y}^h - \hat{\mathbf{x}}^h) \quad (32)$$

where  $\mathbf{L}$  is the observer gain matrix, which can be adjusted to tradeoff the observers' settling time and its filtering demands [39], and  $\hat{\mathbf{x}}^h$  is the estimate of first two states. The two observer-based residuals are given by

$$\begin{pmatrix} r_3 \\ r_4 \end{pmatrix} = \mathbf{y}^h - \hat{\mathbf{x}}^h. \quad (33)$$

### C. GLRT Test Designs for Residuals

In this section, we will illustrate how to design GLRT tests for the earlier three residuals. Due to the presence of disturbance (process noise) and system nonlinearity, the tests based on GLRTs are unreliable [41], which means they have false alarms and miss detection.

Fig. 6 shows the detection and false-alarm probabilities of the designed tests, where  $Pd_{ij}$  and  $Pf_{ij}$  are the detection and false-alarm probabilities of test  $t_j$  and failure source  $F_i$ , respectively. For the automotive ABS system, it is desirable to minimize the number of false alarms for the designed test. This can be achieved by adjusting the decision thresholds and the GLRT window size to trade off the detection latency and probability of detection ( $Pd$ ). In addition, after the test designs are done, these  $Pd$ 's and  $Pf$ 's of the designed tests, which are related to the diagnosis accuracy, will be affected by the magnitudes of parametric faults and sensor faults, noise inflation and deflation, and GLRT chart window size. We will investigate the performance evaluation of these tests in the next section.

Note that for parity-based residual in (25), the parametric faults in  $\theta^2$  have no impact on the first residual. Relatively large-sized  $R_w$  ( $> 20\%$  decrease) and wheel speed sensor ( $> 5\%$  decrease) faults are detected by GLRTs on  $r_1$ . The moving window size  $W$  of GLRT is set to 50. Fig. 5 shows a plot of the GLRTs for a 25% decrease in  $R_w$  and a 5% decrease in wheel speed. The thick straight lines in the earlier two subfigures show the thresholds for  $R_w$  and sensor faults. We can see that, under nominal conditions, the GLRT sporadically exceeds the threshold. Such false alarms can be minimized by invoking the GLRT for a window size  $Z$ : The alarm will only be triggered when the GLRT is above the threshold during this

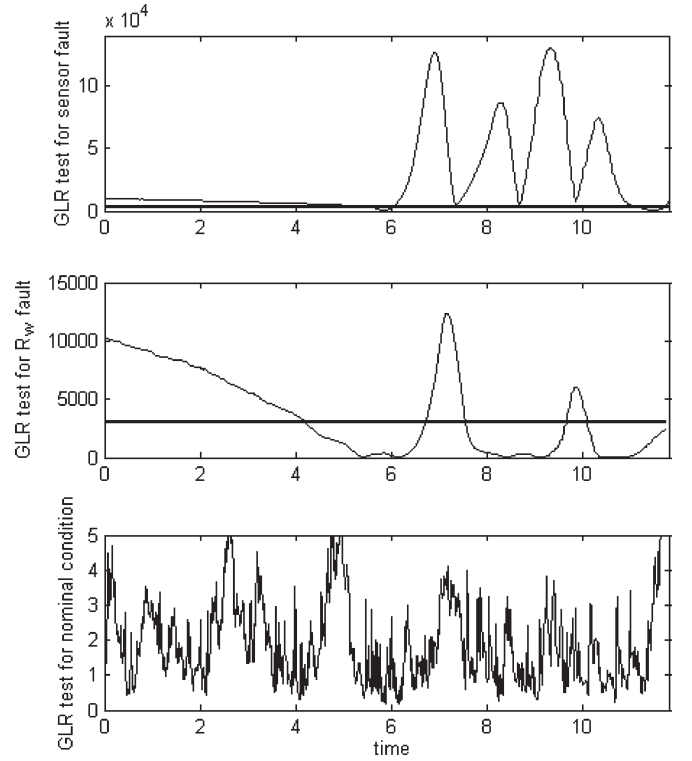


Fig. 7. GLRT with  $R_w$  and sensor fault for  $r_3$  generated by observer.

entire window. The decision rule for triggering an alarm by applying window size  $Z$  is

$$\left[ \frac{1}{Z} \sum_{i=1}^Z I_{[q, +\infty]}(\text{GLRT}_i) \right] = 1 \quad (34)$$

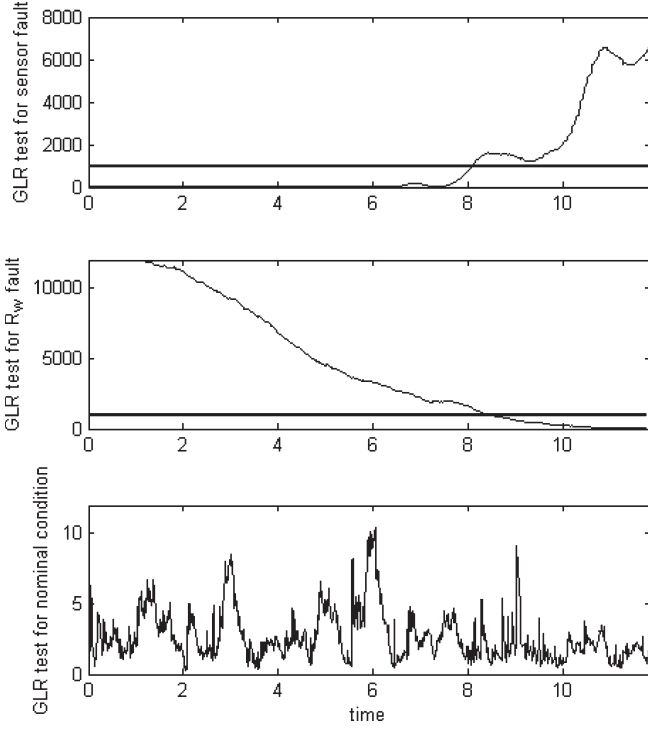
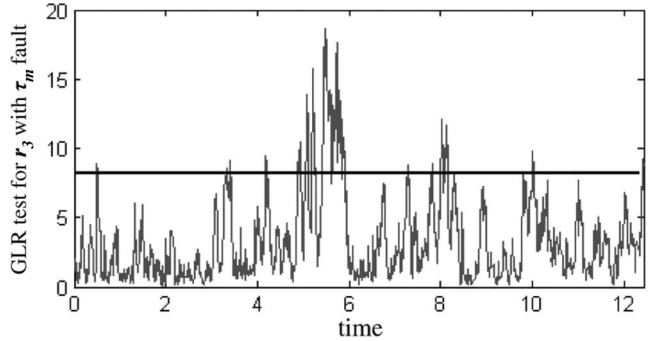
where  $I_{[q, +\infty]}(\text{GLRT}_i)$  is the indicator function (equal to one when  $\text{GLRT}_i \geq q$ , zero otherwise),  $\text{GLRT}_i$  is the  $i$ th sample of GLRT on residual, and  $q$  is a user-defined threshold. Indeed, applying a window  $Z$  is equivalent to applying a FIR filtering on the GLRT residual with tap length ( $Z$ ) and uniform weights ( $1/Z$ ).

In this paper, we name this test based on GLRT as  $G\_P_1$ . Thus, test  $G\_P_1$  will be set to "1" when the GLRT statistic exceeds the threshold for severe  $R_w$  and sensor faults; otherwise, it is set to "0." Additional tests are designed in the following section to discriminate between the  $R_w$  and the wheel-speed-sensor faults.

Figs. 7 and 8 show the GLRT plots for a 2% decrease in  $R_w$  and a 2% decrease in sensor readings, respectively. For these two GLRTs, the window size  $W$  is set to ten, since these residuals are more robust to noise. We can see that both GLRTs deviate from zero. Again, the thick straight lines in the earlier two subfigures show the thresholds. Here, we name these two GLRTs as  $G\_O_1$  and  $G\_O_2$ . Notice that the GLRT for residual four of sensor fault only exceeds the threshold after the ABS engages the valve at around 7 s. Therefore, we design a time-dependent test, which has the following logic:

$$\text{test outcome} = \begin{cases} 1, & G(t) > q, t > T \wedge G(t) < q, t \leq T \\ 0, & \text{otherwise} \end{cases} \quad (35)$$




 Fig. 8. GLRT with  $R_w$  and sensor fault for  $r_4$  generated by observer.

 Fig. 9. GLRT with  $\tau_m$  fault for  $r_3$  generated by observer.

where  $G(t)$  is the GLRT for residual,  $q$  is a user-defined threshold,  $T$  is a user-selected time constant, and  $\wedge$  is a conjunction (AND) operator. Let us denote this test for residual four as  $G_{O\_T_2}$ . Therefore, the test  $G_{O\_T_2}$  is “1” for sensor fault and “0” for  $R_w$  fault and other undetectable faults, thereby isolating these two faults. A fault in  $\tau_m$  can be detected by a GLRT on residual  $r_3$  at the beginning of the valve operation. Fig. 9 shows a plot for the GLRT on  $r_3$  with a  $\tau_m$  fault. The GLRT for  $r_3$  shows a change at 5 s. The  $\tau_m$  fault has no impact on residual  $r_4$  in the simulation. The change in  $\tau_m$  has very little impact (“sensitivity”) on the measurement  $y_2$ . This is because, in the observer design in (32), residual  $r_4$  is only dependent on  $\mu(\lambda)$ , where  $\mu(\lambda)$  is a function of wheel slip (which, in turn, is a function of wheel speed). Since the  $\tau_m$  fault has a very small impact on wheel speed (less than 2%), the change in the slip is negligible and cannot be detected by the GLRT.

For GLRT of  $r_1$  under nominal conditions, 5000 Monte Carlo simulations are performed. We used three window sizes ( $W = 50, 40, 30$ ). By setting proper thresholds based on different GLRT window size, the simulations result in a zero false-alarm

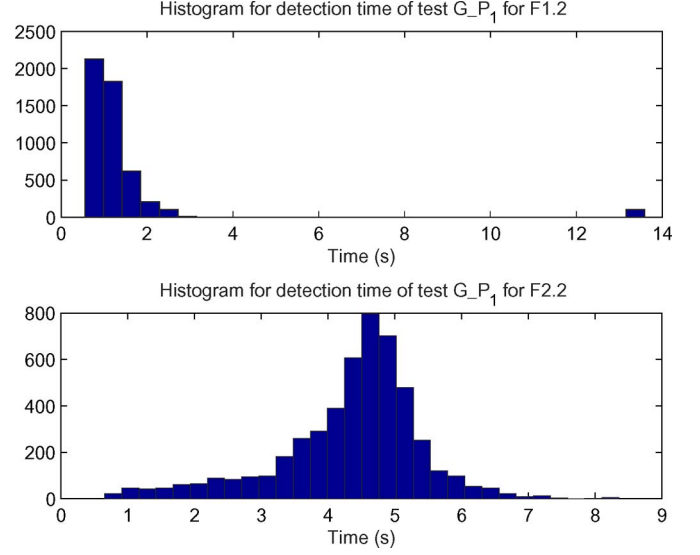

 Fig. 10. Histogram of detection time of  $G_{P_1}$  for fault F1.2 and F2.2.

 TABLE III  
GLRT PARAMETERS FOR RESIDUALS

Residuals	$\mu_0$	$\sigma^2$	W	Z	Threshold
$G_{P_1}$	0.097	1.26e+3	50	40	0.3
	0.097	1.26e+3	40	40	0.24
	0.097	1.26e+3	30	40	0.18
$G_{O_1}$	-1.07e-4	4.5e-3	10	5	8
$G_{O_2}$	-1.92e-4	4.0e-3	10	5	2000

$W$  is the moving window size for GLRT

$Z$  is the smooth window size to surplus the noise

rate ( $P_f$ ). For large-sized  $R_w$  (25%) fault (F1.2) and sensor fault (F2.2), 5000 Monte Carlo simulations are also performed.

Fig. 10 shows the histograms of detection time of  $G_{P_1}$  test on faults F1.2 and F2.2, respectively. Although this test has 100% detection, there are few cases where the detection time is slightly delayed. For a maximum detection delay of 7 s, the probabilities of detection  $P_d$  for both faults are 0.994 and 0.979, respectively.

The means  $\mu_0$  and variances  $\sigma^2$  of GLRT tests for residuals are determined by taking the mean value of the estimate of  $\mu_0$  for 5000 Monte Carlo simulation runs for nominal conditions. Table III summarizes the GLRT design parameters for these three residuals. From Table III, we can see that, under nominal conditions, the means of residuals ( $\mu_0$ ) are close to zero. The variance of the parity-based residual ( $G_{P_1}$ ) is significantly higher than the other two observer-based residuals due to its noisy derivatives. The sliding window size for parity-based residual  $G_{P_1}$  is set at 50, which means a 0.5-s detection latency for the GLRT. To surplus the additional noise in the GLRT output for residual  $G_{P_1}$ , a window size  $Z$  is set at 40 (0.4 s). This causes additional 0.4-s detection delay for this residual.

## VI. FAULT ISOLATION BY SVM

The remaining two faults,  $K_f$  and  $I_w$ , can be detected by the GLRT on residuals generated by the nonlinear observer. Once

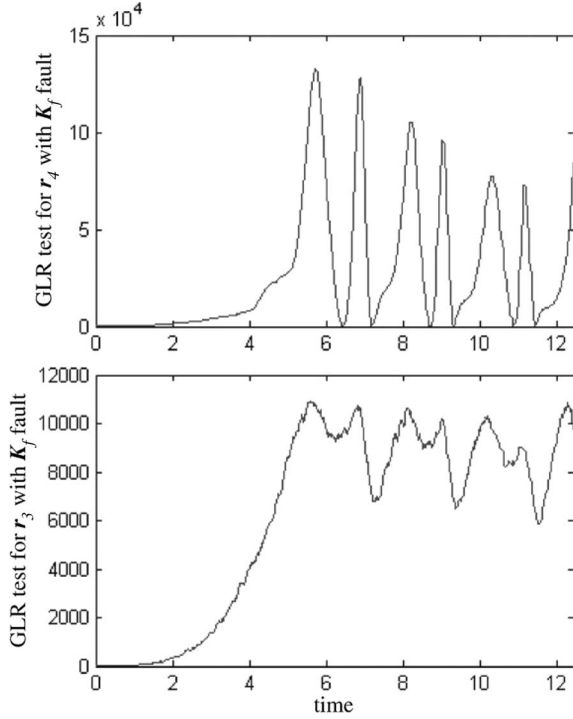


Fig. 11. GLRT with  $K_f$  fault for  $r_3, r_4$  generated by observer.

the fault occurs in one of the parameters, there is a substantial change in both residuals. Fig. 11 shows a plot of the GLRTs for  $r_3$  and  $r_4$  with a  $K_f$  fault. The GLRTs for  $I_w$  have a similar pattern as that for  $K_f$  fault. However, the  $R_w$  and sensor faults are also detected by these GLRTs. Note that, for  $r_3$ , there is a time lag for the GLRT to exceed the threshold. The time lag is the time between applying the brake and the response of the ABS system. This time lag for the GLRT is due to the inherent dynamics of the ABS system. Although the duration may vary under different operating conditions, this feature is inherent and significant enough to be used as a test statistic. Thus, we can design a test similar to (35) for  $r_3$ . Let us denote this test for  $r_3$  as  $G\_O\_T_1$ . Therefore, the test  $G\_O\_T_1$  is “1” for  $K_f$  and  $I_w$  faults and “0” for  $R_w$  fault, the sensor fault, and other undetectable faults.

Since the  $I_w$  and the  $K_f$  faults have similar effect on residuals, it is difficult to use the GLRT alone to isolate them, particularly with different fault severity levels. In order to isolate these two faults, we propose that once the fault is detected, the SVM method be used to isolate these two faults based on two sensor measurements. The SVM has been shown to be effective in a number of applications (e.g., handwriting recognition, face recognition, text categorization, fault diagnosis, etc.).

The theory of SVM used in this paper is provided in Appendix A. The training data consisted of a total of 720 simulated patterns. Out of these, 360 patterns are for the  $K_f$  fault, and the other 360 are for the  $I_w$  fault. The data set was generated as follows. For patterns with the  $K_f$  faults, 20 patterns are obtained for each size of fault ( $\pm 2\% - \pm 10\%$  in step of 1%), resulting in a total of  $20 \times 18 = 360$  patterns. The patterns for  $I_w$  are collected in a similar way. The test data consisted of 720 patterns collected by independent simulations. The data format is the same as that of training data. A plot of

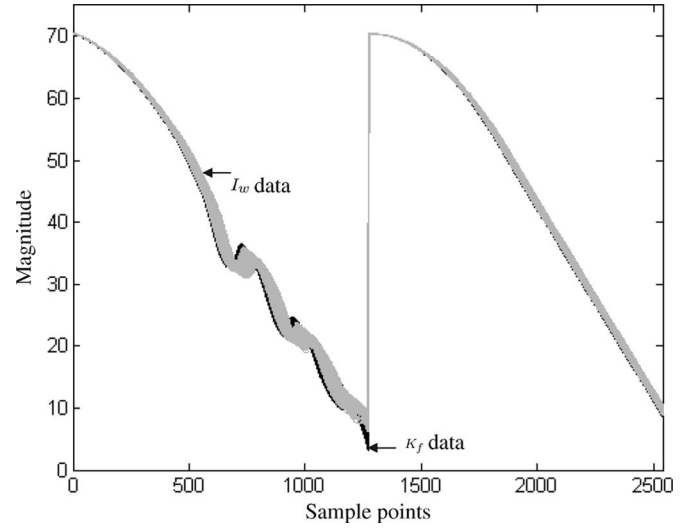


Fig. 12. Output for  $K_f$  and  $I_w$  faults.

TABLE IV  
DIAGNOSTIC MATRIX FOR THE ABS TEST DESIGN

Fault	$G\_P_1$	$G\_O_1$	$G\_O_2$	$G\_O\_T_1$	$G\_O\_T_2$	$S\_K_f$	$S\_I_w$
F0	0	0	0	0	0	0	0
F1.1	0	1	1	0	1	0	0
F1.2	1	1	1	0	1	0	0
F2.1	0	1	1	0	0	0	0
F2.2	1	1	1	0	0	0	0
F3	0	1	1	1	0	1	0
F4	0	1	1	1	0	0	1
F5	0	1	0	0	0	0	0

all the training patterns in Fig. 12 shows the similarity of the outputs for  $K_f$  and  $I_w$  faults. Fig. 12 shows two measurements (wheel speed and vehicle speed) augmented together. Wheel speed readings correspond to samples in 0–1250, while the vehicle speed readings correspond to the range of 1250–2500. The  $I_w$  data are marked with a gray line, and  $K_f$  data are marked with a black line. We can see that there is only a slight difference between the output signals.

The parameters for the SVM are  $\sigma = 20$ ,  $C = 1$ . After training, a total of 35 patterns are misclassified in the test data, which results in an error rate of 4.7%. Therefore, we can design two tests:  $S\_K_f$  and  $S\_I_w$  using the SVM, which assigns  $S\_K_f = 1$  when the data are classified as the  $K_f$  fault or assigns  $S\_I_w = 1$  when the data are classified as the  $I_w$  fault. Subsequently, a parameter-estimation technique is used after fault isolation to estimate the severity of the fault. When the SVM method is applied to an actual ABS, we can replace the training samples by recorded real samples and retrain the SVM classifier.

## VII. PERFORMANCE EVALUATION OF THE FDI SCHEME

Table IV shows the diagnostic matrix (D-matrix), which summarizes the test designs for all faults considered for the ABS. Each row represents a fault state, and columns represent tests. The D-matrix  $D = \{d_{ij}\}$  provides detection information, where  $d_{ij}$  is one if test  $j$  detects a fault state  $i$ . Here, F0

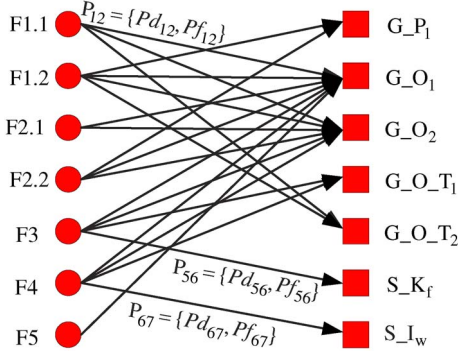


Fig. 13. Bipartite digraph model.

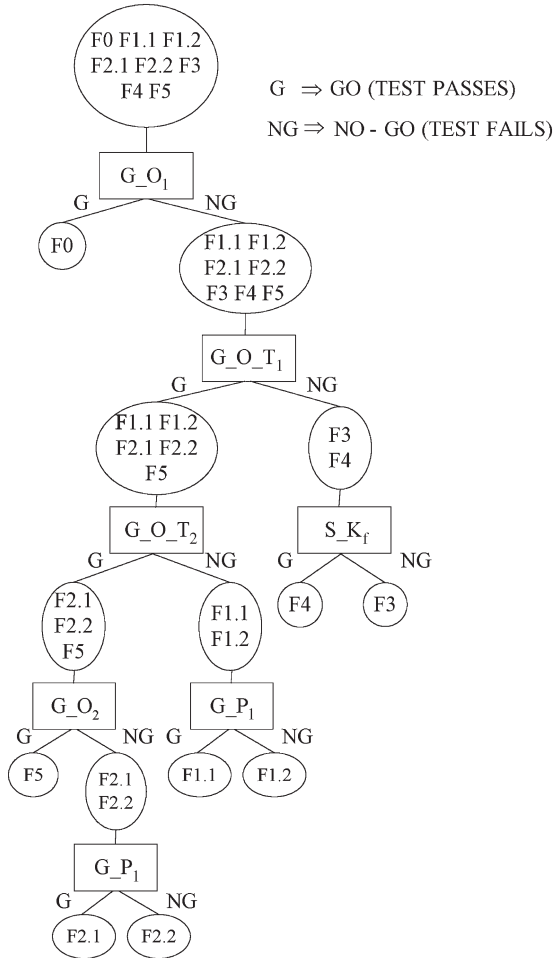


Fig. 14. Test sequence generation for the ABS.

represents the “System OK” status with all the tests having pass outcomes. Since there are no identical rows in this matrix, all the faults can be uniquely isolated. A directed graph (digraph) model of the system is constructed based on the D-matrix in TEAMS software [1], [42]. Fig. 13 shows the bipartite digraph model based on the D-matrix in Table IV. In Fig. 13,  $P_{ij}$  represents the set of probability pairs  $P_{ij} = (Pd_{ij}, Pf_{ij})$ . The AND/OR diagnostic tree, generated by the optimal test-sequencing algorithm of Pattipati and Alexandridis [43] and Raghavan *et al.* [41], [44], is shown in Fig. 14, wherein an AND node represents a test and an OR node denotes the ambiguity

group [43]. The goal is to use minimum expected cost tests to isolate the faults. In this tree, the branch which goes to the left/right below the test implies that the test passed (G)/failed (NG). It can be seen that if all six tests are performed online, all the faults can be detected and uniquely isolated. A passed test outcome can be used to discriminate faults in the diagnostic tree. For example, if test  $G_{O_1}$  fails, test  $G_{O_T1}$  passes and test  $S_{K_f}$  passes,  $I_w$  fault is isolated. Notice that  $S_{I_w}$  test is found to be redundant by the test sequencing algorithm for single-fault diagnosis, as evidenced by its absence from the tree. One feature of the diagnostic tree is that it shows the set of Go-path tests ( $G_{O_1}$ ) that can respond to any fault. The Go-path tests can be obtained by putting all the tests on the leftmost branches of the tree, leading to the “System OK” status. With this subset of tests residing in the ABS ECU, all the faults considered here can be detected. Therefore, to save computational load of ABS ECU, only the Go-path tests (which have full fault coverage) are executed in real time in the ABS ECU. The other tests are executed based on the results of Go-path tests. The diagnostic tree of the ABS is automatically translated into the “IF-THEN-ELSE” logic shown below, which can be embedded in the ABS ECU.

```

IF  $G_{O_1} == G$  DISPLAY(“SYSTEM OK”);
ELSE REQUEST  $G_{O_T1}$  TEST;
    IF  $G_{O_T1} == G$  REQUEST  $G_{O_T2}$  TEST;
        IF  $G_{O_T2} == G$  REQUEST  $G_{O_2}$  TEST;
            IF  $G_{O_2} == G$  DISPLAY[“ $\tau_m$  fault ( $\geq$ 
15%)”];
            ELSE REQUEST  $G_{P_1}$  TEST;
                IF  $G_{P_1} == G$  DISPLAY[“ $R_w$ 
fault ( $< 20\%$ )”];
                ELSE DISPLAY[“ $R_w$  fault
( $\geq 20\%$ )”];
            ELSE REQUEST  $G_{P_1}$  TEST;
                IF  $G_{P_1} == G$  DISPLAY[“sensor fault
( $< 5\%$ )”];
                ELSE DISPLAY[“sensor fault ( $\geq 5\%$ )”];
            ELSE REQUEST  $S_{K_f}$  TEST;
                IF  $S_{K_f} == G$  DISPLAY(“ $I_w$  fault”);
                ELSE DISPLAY(“ $K_f$  fault”);
    
```

When the D-matrix is not binary (e.g., when  $d_{ij}$  denotes the likelihood that test  $j$  detects a fault state  $i$ ), the resulting test-sequencing problem is a partially observed Markov decision problem. An efficient approach for implementing the dynamic-program recursion was presented to solve this problem in [41].

To evaluate the performance of the FDI scheme for ABS, all the tests and residual generation methods are modeled in Simulink, compiled through the Real-Time Workshop, and implemented in a hardware-in-the-loop environment [45]. Real-Time Workshop generates and executes stand-alone C code for developing and testing algorithms. In this paper, due to the lack of real ABS hardware, the resulting code is compiled for rapid prototyping of ECUs (Rtypes) [46] for real-time software-in-the-loop (SiL) testing of the target ABS ECU. The Simulink model for ECU includes the sliding mode PWM controller, the parity equations and the GLRTs, the nonlinear observers,

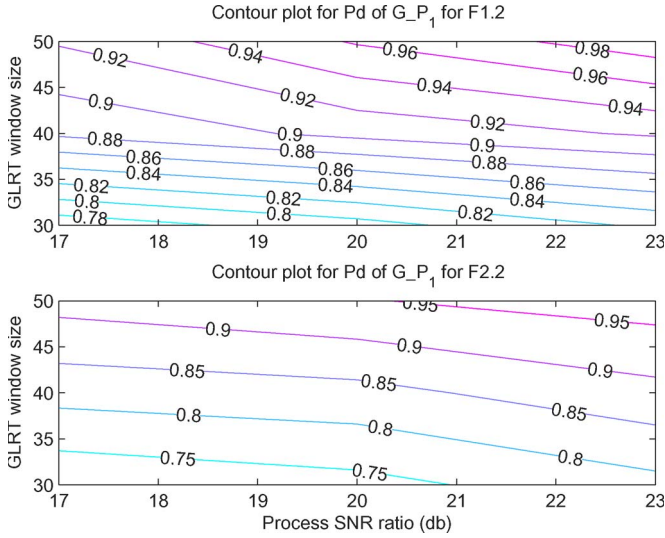


Fig. 15. Contour plot for detection probabilities of test  $G_{P1}$  for faults F1.2 and F2.2 under different noise levels and GLRT window sizes.

the SVM, and the parameter estimation in Fig. 4. The ABS system (except the sliding mode PWM controller) in Fig. 2 is implemented as an environment model. Since we only reconfigure the Simulink blocks (e.g., moving the sliding-mode PWM controller to the prototype ECU in our SiL testing, the FDI scheme is still the same as the one we used for simulation study. Therefore, the simulation results are still valid in the SiL testing environment. The objective of SiL testing is to evaluate the real-time performance and enabling the software to be tested in a completely simulated real-time environment. Notice that only test  $S_{Kf}$  (SVM) and parameter estimation are performed offline (since they require all the history of sensor data), and the other tests are performed online. The simulation is conducted using a fixed step with a step size of 1 ms. For the Rtype ECU, which has Pentium IV 2-GHz CPU, the execution time for the online algorithm is 0.1 ms, which is adequate for embedded diagnosis. For offline performance, the execution time for SVM test  $S_{Kf}$  and parameter estimation are 0.2 and 2 s, respectively.

To evaluate the sensitivity of FDI scheme under different operating conditions, such as different fault sizes, noise levels, and GLRT window size, we conducted extensive simulations to quantify the probability of detection and false-alarm rates of each individual tests. In this paper, measurement noise level is not considered as a changing variable, since it is typically known for ABS subsystem. The process noise levels are more important than measurement noise, since it represents the uncertainty in the ABS model.

For test  $G_{P1}$ , we investigated the effects of process noise levels and the GLRT window size on the detection probability. For process noise, we considered three levels, i.e., 23, 20, and 17 db. For window size effects, we consider three window sizes listed in Table III.

Fig. 15 shows a contour plot of detection probabilities of test  $G_{P1}$  for faults F1.2 and F2.2 under three different noise levels and the GLRT window sizes. Evidently, as the window sizes increase and noise level decrease, the detection probability increases.

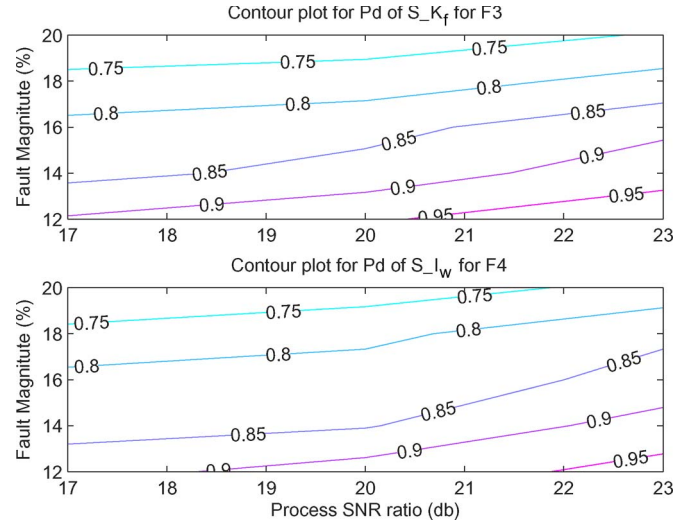


Fig. 16. Contour plot of detection probabilities of tests  $S_{Kf}$  and  $S_{Iw}$  under different noise levels and fault magnitudes.

For SVM-based tests  $S_{Kf}$  and  $S_{Iw}$ , we considered the impact of fault magnitude and process noise levels on probability of detection ( $Pd$ ) and false alarms ( $Pf$ ). We used the same levels of process noises as for the test  $G_{P1}$ . We considered fault magnitudes ranging from 10% to 20% with step change of 2%.

Fig. 16 shows a contour plot of detection probabilities of tests  $S_{Kf}$  and  $S_{Iw}$ . Due to the binary decision outcome of the SVM and the fact that the operating point is at the intersection of the major diagonal and the receiver-operating-characteristic curve (the so-called skill level [47]), the relationship between  $Pd$  and  $Pf$  for these two tests are as follows:  $Pd(S_{Kf}) = 1 - Pf(S_{Iw})$  and  $Pd(S_{Iw}) = 1 - Pf(S_{Kf})$ . As shown in Fig. 16, the probabilities of detection for both of these tests decrease as the noise level and fault magnitude increase.

To evaluate the overall performance of the diagnostic tree, we adopted the standard metrics of sensitivity and specificity that are used in the detection theory. The sensitivity of a diagnostic tree (the true positive rate or diagnostic accuracy under faulty conditions) is the ratio of the number of correct fault diagnoses to the total number of faults introduced. The specificity of a diagnostic tree (the true negative rate) is the ratio of the number of true negative diagnoses (i.e., System OK) to the number of negative samples. For each fault conditions, 5000 Monte Carlo simulations are performed. For no-fault (System OK) conditions, 10000 Monte Carlo simulations are performed. These simulations are also conducted for different noise levels to obtain these two metrics. The window size of all GLRT tests is set to 40. All the fault sizes are set to minimum according to Table II. The underlying assumption of these simulations is that the fault probabilities of simulated faults have uniform distribution (same number of simulations for each fault). The specificity and sensitivity may be affected if the tests have false alarms and missed detections because samples from fault-free and faulty cases will be different. Notice that the System OK prior probability does not affect sensitivity and specificity of the diagnostic tree. It only affects how many no-fault and fault samples will be drawn for Monte Carlo simulation. The 10000



TABLE V  
ACCURACY AND SENSITIVITY METRICS FOR DIAGNOSIS TREE

SNR Ratio (db)	17	20	23
Sensitivity (%)	99.92	99.96	99.98
Specificity (%)	99.76	99.84	99.99

Monte Carlo simulations for System OK state provides robust estimate of specificity.

Table V displays the sensitivity and specificity of the ABS diagnosis tree under three noise levels. We can see that these two metrics are all very high. The reason for high sensitivity and specificity is that the Go path (path from the top node of tree to System OK) test  $G_{O_1}$  has high specificity (this is an observer-based test that is robust to noise).

### VIII. SUBSET SELECTION FOR PARAMETER ESTIMATION

Fault isolation, based on process parameters, which are not directly measurable, requires parameter estimation. Two parameter-estimation techniques, viz., equation error method and output error method, can be used. Generally, equation error method is computationally more efficient than the output error method. However, it is less accurate than the output error method. In this paper, we employ the output error method discussed in Appendix B.

The ABS model has a number of physically interesting and interpretable parameters, and the measurement variables are not rich enough to reflect the individual effects of all the parameters in the system. When all the interesting parameters are estimated at the same time, even output error methods can lead to ill-conditioned estimation problems, particularly in the presence of additive noise. Burth *et al.* [48] examined the manifestations of such ill conditioning in the context of parameter estimation for a synchronous generator. Carefully chosen subset of parameters can lead to major improvements in estimation performance measured in terms of computational time as well as the standard deviations of the estimated parameters.

In the ABS, the nonlinear output error parameter-estimation method produces biased estimates when all the parameters are estimated as a block. Therefore, the subset parameter-estimation techniques are well suited for our application. The subset of parameters is selected by detection and isolation of the parametric fault using the GLRT and SVM. When a parametric fault is isolated, this parameter is estimated via the nonlinear output error method.

Table VI compares the accuracies of parameter estimation averaged over 20 runs via the two methods: estimating all the parameters versus reduced (one-at-a-time) parameter estimation after FDI. For a parametric fault, “true” values of the simulated faults are  $R_w = 0.286$ ,  $K_f = 95$ ,  $I_w = 4.75$ , and  $\tau_m = 0.023$ . Row *err* shows the mean relative errors of the estimated parameters with respect to their “true” values (in percent). Row *std* shows the normalized standard deviations of the estimated parameters (in percent), normalized by the “true” values.

From Table VI, it is evident that subset parameter estimation provides more precise estimates than the method which esti-

TABLE VI  
MEAN RELATIVE ERRORS AND NORMALIZED STANDARD DEVIATIONS IN PARAMETER ESTIMATION

		Block Estimation				Subset Para.
		$R_w$	$K_f$	$I_w$	$\tau_m$	Estimation
$K_f$	<i>err</i>	3.2	5.0	6.0	25.0	1.05
	<i>std</i>	1.2	3.5	6.8	22.2	0.12
$I_w$	<i>err</i>	2.0	4.5	4.0	19.0	0.52
	<i>std</i>	1.6	4.8	7.2	39.3	0.35
$\tau_m$	<i>err</i>	3.5	7.8	10.3	27.5	2.0
	<i>std</i>	2.4	5.2	5.6	46.5	0.80
$R_w$	<i>err</i>	0.39	0.33	2.98	279.33	0.004
	<i>std</i>	0.25	0.12	1.48	33.4	0.014

$$err = \frac{\text{mean relative error}}{\text{“true” value}} \times 100\%$$

$$std = \frac{\text{standard deviation of estimated parameters}}{\text{“true” value}} \times 100\%$$

mates all four parameters as a block. This is particularly significant with single-parameter faults; the full parameter estimation provides biased estimates for all the parameters.

### IX. SUMMARY

In this paper, a systematic design procedure was proposed for an integrated-model-based and data-driven FDI in nonlinear dynamic systems. The procedure starts with parity relations, which present major computational issues in the presence of noisy measurements. Then, a nonlinear observer was used to generate additional residuals to improve the diagnostic performance. Finally, a robust data-driven method, viz., SVM, was employed to isolate small-sized faults. The design procedure was successfully applied to an ABS to detect and isolate a sensor fault and four parametric faults of various sizes.

### APPENDIX A SVM

As a statistical learning theory, SVM has gained popularity in recent years because of its two distinct features. First, SVM is often associated with the physical meaning of the data, so that it is easy to interpret. Second, it requires only a small amount of training samples [9], [11]. SVM has been successfully used in many applications, such as pattern recognition, multiple regression, nonlinear model fitting, to name a few [49].

The essential idea of SVM is to transform the signal to a higher dimensional feature space and find an optimal hyperplane that maximizes the margin between the classes. We implemented a generalized form of SVM for overlapping and nonlinearly separable data.

Briefly, the training data for the two classes ( $K_f$  fault and  $I_w$  fault) are arranged as

$$\mathcal{F} = \{(\mathbf{y}_1, z_1), (\mathbf{y}_2, z_2), \dots, (\mathbf{y}_n, z_n)\}, \quad \mathbf{y}_i \in \mathbb{R}^N$$

$$z_i \in \{-1, 1\} \quad (36)$$

where  $\mathbf{y}$  is the augment of the measurement  $[\mathbf{y}_1 \ \mathbf{y}_2]$  and  $z_i$  is the class label.

For nonseparable case, a separating hyperplane must satisfy the following constraints:

$$z_i [\langle \mathbf{w}, \mathbf{y}_i \rangle + b] \geq 1 - \xi_i, \quad i = 1, 2, \dots, n \quad (37)$$

where  $\xi_i \geq 0$  is the slack variable. To determine vector  $\mathbf{w}$  and  $b$ , the following function is minimized:

$$\Psi(\mathbf{w}, \boldsymbol{\xi}) = \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^N \xi_i \quad (38)$$

subject to constraints in (37). The first term in the objective function represents the model complexity, and the second term represents the model accuracy [9]. Here,  $C$  is a regularization parameter to control the tradeoff between these two terms. The solution of (38) is given by the following dual optimization problem [49]:

$$\max W(\alpha) = \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j z_i z_j \langle \mathbf{y}_i, \mathbf{y}_j \rangle \quad (39)$$

subject to

$$\begin{aligned} \sum_{i=1}^N \alpha_i z_i &= 0 \\ \alpha_i &\in [0, C], \quad i = 1, \dots, n. \end{aligned}$$

If a nonlinear mapping  $K(\mathbf{y}_i, \mathbf{y}_j)$  is chosen *a priori*, the optimization problem in (39) becomes

$$\max W(\alpha) = \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j z_i z_j K(\mathbf{y}_i, \mathbf{y}_j) \quad (40)$$

subject to the same set of constraints. The nonlinear mapping (or kernel function)  $K$  is used to transform the original input  $\mathbf{y}$  to a higher dimensional feature space  $\Omega$  via

$$K(\mathbf{y}_i, \mathbf{y}_j) = \langle \psi(\mathbf{y}_i), \psi(\mathbf{y}_j) \rangle \quad (41)$$

where  $\langle \cdot, \cdot \rangle$  denotes the dot product.

The decision function becomes

$$f(\mathbf{y}) = \text{sgn} \left( \sum_{i \in SVs} \alpha_i z_i K(\mathbf{y}, \mathbf{y}_i) + b \right) \quad (42)$$

where  $SVs$  is an index set, which contains the indexes of the support vectors. In practice, various kernel functions are used, such as polynomial, radial basis functions (RBFs), and sigmoid functions. In this paper, we use the RBF kernel functions,  $K(\mathbf{y}_i, \mathbf{y}_j) = \exp(-\|\mathbf{y}_i - \mathbf{y}_j\|^2 / 2\sigma^2)$ .

#### APPENDIX B OUTPUT ERROR ESTIMATION METHOD

In this method, the parameters of ABS are determined by minimizing the sum of squares of the residual vector

$$\mathbf{r}(i, \boldsymbol{\theta}) = \mathbf{y}(i) - \hat{\mathbf{y}}(i|\boldsymbol{\theta}) \quad (43)$$

where  $\boldsymbol{\theta}$  denotes the  $n$ -vector of model parameters,  $\mathbf{y}$  is the  $M$ -vector of actual measurement (in ABS,  $M = 2$ ), and  $\hat{\mathbf{y}}(i|\boldsymbol{\theta})$  is the  $M$ -vector of model predictions for the measurements. These errors can be computed for  $i = 1, 2, \dots, N$  (in ABS,  $N = 1250$ ). The estimate  $\hat{\boldsymbol{\theta}}_N$  is then defined by minimizing

$$\hat{\boldsymbol{\theta}}_N = \arg \min_{\boldsymbol{\theta}} V_N(\boldsymbol{\theta}) \quad (44)$$

where the minimization criterion is defined by

$$V_N(\boldsymbol{\theta}) = \frac{1}{2} \sum_{i=1}^N \mathbf{r}^T(i, \boldsymbol{\theta}) \Lambda^{-1} \mathbf{r}(i, \boldsymbol{\theta}) \quad (45)$$

for some symmetric positive-semidefinite  $M \times M$  weighting matrix  $\Lambda$ . For the ABS, we choose  $\Lambda$  as a  $2 \times 2$  identity matrix. In general, the function in (45) cannot be minimized by analytical methods (e.g., for the ABS). The minimization problem in (45) is equivalent to a nonlinear least square problem. The solution is found by iterative numerical techniques [50]. In this paper, we use *lsqnonlin* function in the Optimization Toolbox of MATLAB. The optimization function uses the trust-region method and is based on the interior-reflective Newton method [51], [52]. The termination criterion for parameter estimation is  $V_N(\boldsymbol{\theta}) < c$ , where  $c$  is a small constant threshold (for the ABS,  $c = 1e - 5$ ). The use of output error method is advocated because, with the fault isolation from the test design, the complexity of parameter estimation is reduced considerably while improving the estimation accuracy.

#### ACKNOWLEDGMENT

The authors would like to thank the reviewers for constructive suggestions which significantly improved this paper.

#### REFERENCES

- [1] The QSI Website. [Online]. Available: <http://www.teamssqi.com>
- [2] J. Luo, K. R. Pattipati, L. Qiao, and S. Chigusa, "Agent-based real time diagnosis," in *Proc. IEEE Aerosp. Conf.*, Big Sky, MT, Mar. 2005, pp. 1–9.
- [3] R. Isermann, "Supervision, fault-detection and fault-diagnosis methods—An introduction," *Control Eng. Pract.*, vol. 5, no. 5, pp. 639–652, May 1997.
- [4] R. J. Patton, P. M. Frank, and R. N. Clark, *Issues of Fault Diagnosis for Dynamic Systems*. London, U.K.: Springer-Verlag, 2000.
- [5] S. Simani, C. Fantuzzi, and R. J. Patton, *Model-Based Fault Diagnosis in Dynamic Systems Using Identification Techniques*. London, U.K.: Springer-Verlag, 2003.
- [6] M. Fang, Y. Tian, and L. Guo, "Fault diagnosis of nonlinear system based on generalized observer," *Appl. Math. Comput.*, vol. 185, no. 2, pp. 1131–1137, Feb. 2007.
- [7] S. Narasimhan and G. Biswas, "Model-based diagnosis of hybrid systems," *IEEE Trans. Syst., Man, Cybern. A, Syst., Humans*, vol. 37, no. 3, pp. 348–361, May 2007.
- [8] M. Misra, H. H. Yue, S. J. Qin, and C. Ling, "Multivariate process monitoring and fault diagnosis by multi-scale PCA," *Comput. Chem. Eng.*, vol. 26, no. 9, pp. 1281–1293, Sep. 2002.
- [9] L. H. Chiang, M. E. Kotanchek, and A. K. Kordon, "Fault diagnosis based on fisher discriminant analysis and support vector machines," *Comput. Chem. Eng.*, vol. 28, no. 8, pp. 1389–1401, Jul. 2004.
- [10] T. Kourti, P. Nomikos, and J. F. MacGregor, "Analysis, monitoring and fault diagnosis of batch processes using multiblock and multiway PLS," *J. Process Control*, vol. 5, no. 1, pp. 3–21, 1995.

- [11] M. Ge, R. Du, G. Zhang, and Y. Xu, "Fault diagnosis using support vector machine with an application in sheet metal stamping operations," *Mech. Syst. Signal Process.*, vol. 18, no. 1, pp. 143–159, Jan. 2004.
- [12] B. Samanta, "Gear fault detection using artificial neural networks and support vector machines with genetic algorithms," *Mech. Syst. Signal Process.*, vol. 18, no. 3, pp. 625–644, May 2004.
- [13] J. Liang and R. Du, "Model-based fault detection and diagnosis of HVAC systems using support vector machine method," *Int. J. Refrig.*, vol. 30, no. 6, pp. 1104–1114, Sep. 2007.
- [14] J. Luo, F. Tu, M. Azam, K. R. Pattipati, L. Qiao, and M. Kawamoto, "Intelligent model-based diagnostics for vehicle health management," in *Proc. SPIE Conf.*, Orlando, FL, Apr. 2003, vol. 5107, pp. 13–26.
- [15] J. Luo, N. Madhavi, K. R. Pattipati, L. Qiao, M. Kawamoto, and S. Chigusa, "Model-based prognostic techniques," in *Proc. IEEE AUTOTESTCON*, Sep. 2003, pp. 330–340.
- [16] P. Pisu, A. Soliman, and G. Rizzoni, "Vehicle chassis monitoring system," *Control Eng. Pract.*, vol. 11, no. 3, pp. 345–354, Mar. 2003.
- [17] M. Borner, H. Straky, T. Weispfenning, and R. Isermann, "Model based fault detection of vehicle suspension and hydraulic brake systems," *Mechatronics*, vol. 12, no. 8, pp. 999–1010, Oct. 2002.
- [18] B. Friedland, *Advanced Control System Design*. Englewood Cliffs, NJ: Prentice-Hall, 1996.
- [19] M. Wu and M. Shih, "Simulated and experimental study of hydraulic anti-lock braking system using sliding-mode PWM control," *Mechatronics*, vol. 13, no. 4, pp. 331–351, May 2003.
- [20] C. Unsal and P. Kachroo, "Sliding mode measurement feedback control for antilock braking system," *IEEE Trans. Control Syst. Technol.*, vol. 7, no. 2, pp. 271–281, Mar. 1999.
- [21] C.-M. Lin and C.-F. Hsu, "Neural-network hybrid control for antilock braking systems," *IEEE Trans. Neural Netw.*, vol. 14, no. 2, pp. 351–359, Mar. 2003.
- [22] J. B. Heywood, *Internal Combustion Engine Fundamentals*. New York: McGraw-Hill, 1988.
- [23] T. C. Schafer, D. W. Howard, and R. W. Carp, "Design and performance considerations for passenger car anti-skid systems," presented at the Transaction Society Automotive Engineers, Detroit, MI, 1968, Paper 680458.
- [24] S. Park and T. B. Sheridan, "Enhanced human-machine interface in braking," *IEEE Trans. Syst., Man, Cybern. A, Syst., Humans*, vol. 34, no. 5, pp. 615–629, Sep. 2004.
- [25] C.-M. Lin and C.-F. Hsu, "Self-learning fuzzy sliding-mode control for antilock braking systems," *IEEE Trans. Control Syst. Technol.*, vol. 11, no. 2, pp. 273–278, Mar. 2003.
- [26] MSC, *CarSim Educational User Manual*, Mechanical Simulation Corp., Jan. 2000.
- [27] T. B. Sheridan and W. R. Ferrell, *Man-Machine Systems: Information, Control, and Decision Models of Human Performance*. Cambridge, MA: MIT Press, 1974.
- [28] M. Basseville and I. V. Nikiforov, *Detection of Abrupt Changes*. Englewood Cliffs, NJ: Prentice-Hall, 1993.
- [29] M. J. Maron, *Numerical Analysis: A Practical Approach*. New York: Macmillan, 1982.
- [30] P. Declerck and M. Staroswiecki, "Characterization of the canonical components of a structural graph—Application to fault detection in large scale industrial plants," in *Proc. Eur. Control Conf.*, Grenoble, France, 1991.
- [31] M. Krysanter and M. Nyberg, "Structural analysis utilizing MSS sets with application to a paper plant," in *Proc. 13th Int. Workshop Principles Diagnosis*, Semmering, Austria, May 2002.
- [32] M. Staroswiecki and G. Comtet-Varga, "Analytical redundancy relations for fault detection and isolation in algebraic dynamic systems," *Automatica*, vol. 37, no. 5, pp. 687–699, May 2001.
- [33] R. Isermann, "Process fault detection based on modeling and estimation methods—A survey," *Automatica*, vol. 20, no. 4, pp. 387–404, Jul. 1984.
- [34] R. Conatser, J. Wagner, S. Ganta, and I. Walker, "Diagnosis of automotive electronic throttle control systems," *Control Eng. Pract.*, vol. 12, no. 1, pp. 23–30, Jan. 2004.
- [35] A. V. Oppenheim and R. W. Schaffer, *Discrete-Time Signal Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1999.
- [36] Q. Zhang, M. Basseville, and A. Benveniste, "Early warning of slight changes in systems," *Automatica*, vol. 30, no. 1, pp. 95–114, Jan. 1994.
- [37] P. M. Frank, "On-line fault detection in uncertain nonlinear systems using diagnostic observers: A survey," *Int. J. Syst. Sci.*, vol. 25, no. 12, pp. 2129–2154, Dec. 1994.
- [38] E. A. Garcia and P. Frank, "Deterministic nonlinear observer-based approaches to fault diagnosis: A survey," *Control Eng. Pract.*, vol. 5, no. 5, pp. 663–670, May 1997.
- [39] D. Schröder, *Intelligent Observer and Control Design for Nonlinear Systems*. Berlin, Germany: Springer-Verlag, 2000.
- [40] G. Hostetter, "Sampled data updating of continuous-time observers," *IEEE Trans. Autom. Control*, vol. AC-21, no. 2, pp. 272–273, Apr. 1976.
- [41] V. Raghavan, M. Shakeri, and K. R. Pattipati, "Test sequencing algorithms with unreliable tests," *IEEE Trans. Syst., Man, Cybern. A, Syst., Humans*, vol. 29, no. 4, pp. 347–357, Jul. 1999.
- [42] S. Deb, K. R. Pattipati, V. Raghavan, M. Shakeri, and R. Shrestha, "Multi-signal flow graphs: A novel approach for system testability analysis and fault diagnosis," *IEEE Aerosp. Electron. Syst. Mag.*, vol. 10, no. 5, pp. 14–25, May 1995.
- [43] K. R. Pattipati and M. Alexandridis, "Application of heuristic search and information theory to sequential fault diagnosis," *IEEE Trans. Syst., Man, Cybern. A, Syst., Humans*, vol. 20, no. 4, pp. 872–887, Jul./Aug. 1990.
- [44] V. Raghavan, M. Shakeri, and K. R. Pattipati, "Optimal and near-optimal test sequencing algorithms with realistic test models," *IEEE Trans. Syst., Man, Cybern. A, Syst., Humans*, vol. 29, no. 1, pp. 11–26, Jan. 1999.
- [45] J. Luo, K. R. Pattipati, L. Qiao, and S. Chigusa, "Towards an integrated diagnostic development process for automotive systems," in *Proc. IEEE Int. Conf. SMC*, Big Island, HI, Oct. 2005, vol. 3, pp. 2985–2990.
- [46] M. Fukazawa, "Development of PC-based HIL simulator CRAMAS 2001," *Fujitsu Ten Tech. J.*, vol. 19, no. 1, pp. 12–21, 2001.
- [47] K. R. P. A. Pete and D. Kleinman, "Optimal team and individual decision rules in uncertain dichotomous situations," *Public Choice*, vol. 75, no. 3, pp. 205–230, Mar. 1993.
- [48] M. Burth, G. C. Verghese, and M. Vlez-Reyes, "Subset selection for improved parameter estimation in on-line identification of a synchronous generator," *IEEE Trans. Power Syst.*, vol. 14, no. 1, pp. 218–225, Feb. 1999.
- [49] A. J. Smola, P. L. Bartlett, B. Scholkopf, and D. Schuurmans, *Advances in Large Margin Classifiers*. Cambridge, MA: MIT Press, 2000.
- [50] L. Ljung, *System Identification: Theory for the User*. Englewood Cliffs, NJ: Prentice-Hall, 1999.
- [51] T. Coleman and Y. Li, "On the convergence of reflective Newton methods for large-scale nonlinear minimization and subject to bounds," *Math. Program.*, vol. 67, no. 2, pp. 189–224, 1994.
- [52] T. Coleman and Y. Li, "An interior, trust region approach for nonlinear minimization subject to bounds," *SIAM J. Optim.*, vol. 6, no. 2, pp. 418–445, May 1996.



**Jianhui Luo** (S'03–M'06) received the B.S. degree in automatic control from the Shanghai Institute of Railway Technology, Shanghai, China, in 1993, the M.S. degree in transportation system engineering, the M.S. degree in electrical engineering, and the Ph.D. degree in electrical engineering from the University of Connecticut, Storrs, in 2002, 2004, and 2006, respectively.

From 1993 to 2000, he was an Application Engineer and later Deputy Department Head with CASCO Signal Inc., Shanghai. Since 2006, he has been a Research Scientist with Qualtech Systems, Inc., Wethersfield, CT. His research interests include distributed fault diagnosis, fault diagnosis/prognosis, system modeling and simulation, real-time embedded systems, and machine learning.

Dr. Luo was the recipient of the Walter E. Peterson Award for Best Technical Paper at the IEEE International Autotest Conference in 2004 for his work on data-driven diagnosis. He was also the recipient of the Best Paper Award at the IEEE International Autotest Conference in 2005 for his work on inference for graphical diagnostic models.



**Madhavi Namburu** received the B.Tech. degree in electrical engineering from Jawaharlal Nehru Technological University, Hyderabad, India, in 2002 and the M.S. degree in electrical engineering from the University of Connecticut, Storrs, in 2006.

She is currently a Research Engineer with the Technical Research Department, Toyota Technical Center, Ann Arbor, MI. Her research interests include fault diagnostics and prognostics of engineering systems, data mining, text mining, knowledge discovery, and artificial-intelligence algorithms.

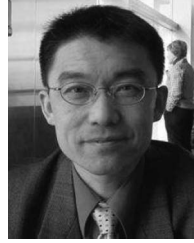


**Krishna R. Pattipati** (S'77–M'80–SM'91–F'95) received the B.Tech. degree in electrical engineering from the Indian Institute of Technology, Kharagpur, in 1975 and the M.S. and Ph.D. degrees in control and communication systems from the University of Connecticut, Storrs, in 1977 and 1980, respectively.

He is currently a Professor of electrical and computer engineering with the Department of Electrical and Computer Engineering, University of Connecticut. He has published over 300 articles, primarily in the application of systems theory and optimization

techniques to large-scale systems.

Dr. Pattipati served as the Editor-in-Chief of the IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS, PART B—CYBERNETICS during 1998–2001. He was the recipient of the Centennial Key to the Future Award in 1984 from the IEEE Systems, Man, and Cybernetics (SMC) Society. He was the recipient of the Andrew P. Sage Award for the Best SMC Transactions Paper in 1999, the Barry Carlton Award for the Best Aerospace and Electronic Systems Transactions Paper in 2000, the 2002 and the 2008 NASA Space Act Awards, the 2003 American Association of University Professors Research Excellence Award, and the 2005 School of Engineering Teaching Excellence Award at the University of Connecticut. He was also the recipient of the Best Technical Paper Awards at the 1985, 1990, 1994, 2002, 2004, and 2005 IEEE Autotest Conferences, and at the 1997 and 2004 Command and Control Conferences.



**Liu Qiao** received the B.E. degree from the Beijing University of Technology, Beijing, China, and the M.S. and Ph.D. degrees in electrical engineering from Tohoku University, Sendai, Japan.

Switching from a university faculty position, he started his automotive career as an Advanced Automotive Control System Expert, Advanced Technology Manager, e-Business Manager, and Research Manager. He successfully led a Canadian hybrid vehicle project and its market introduction. He is currently the General Manager and Chief Technologist

of the Technical Research Department, Toyota Technical Center, Ann Arbor, MI. He is an active member/supporter of many academic associations.



**Shunsuke Chigusa** received the M.S. degree in electrical engineering from Osaka University, Osaka, Japan.

He started his automotive career with Toyota Motor Corporation as an Autonomous Vehicle Control System Scientist and, later, as Intelligent Diagnosis System Scientist. He is currently a Senior Principal Research Scientist with the Technical Research Department, Toyota Technical Center, Ann Arbor, MI.