Incipient Residual-Based Anomaly Detection in Power Electronic Devices

Qian Yang, *Student Member, IEEE*, Muhammed Ali Gultekin, *Student Member, IEEE*, Vahe Seferian, *Student Member, IEEE*, Krishna Pattipati, *Life Fellow, IEEE*, Ali M. Bazzi, *Senior Member, IEEE*, Francesco A. N. Palmieri, *Member, IEEE*, Ravi Rajamani, *Member, IEEE*, Shailesh N. Joshi, *Member, IEEE*, Muhamed Farooq, *Member, IEEE*, and Hiroshi Ukegawa, *Member, IEEE*

***Abstract*—Power electronics (PE) and high-frequency switching circuits are key to superior performance of electric vehicles. It is vital to monitor the condition of the PE components in real-time for safety and reliability. In this article, we propose two anomaly detection methods based on a combination of data preprocessing to suppress noise and outliers, multivariate regression models to predict signals of interest under nominal operation, and sequential analysis of residuals. In particular, the methods utilize median filtering to extract ON-state medians in each switching cycle in nonlinear autoregressive exogenous neural network models or filtered ON-state data in partial least squares-based models to represent the nominal circuit behavior. Optimal and approximate dynamic programming-based feature selection methods are developed to select the most informative signals or their transformations. Predictions from the learned models are used to generate the residuals for anomaly detection by Page's cumulative sum test. The proposed models and anomaly detection methods are validated on three accelerated aging experimental datasets, comprised of 60 power MOSFET devices with low-frequency and high-frequency switching under disparate operating conditions. Due to the simplicity and efficiency of the data-driven anomaly detection schemes, the proposed methods can potentially be embedded in real-time digital platforms.**

***Index Terms*—Cumulative sum (CUSUM) test, nonlinear autoregressive exogenous (NARX), online anomaly detection, partial least squares (PLS), power electronics (PE).**

Manuscript received September 3, 2021; accepted December 20, 2021. Date of publication January 7, 2022; date of current version February 18, 2022. This work was supported by the Toyota Research Institute of North America. The work of Krishna Pattipati was supported in part by the U.S. Office of Naval Research and in part by the U.S. Naval Research Laboratory under Grants N00014-18-1-1238, N00014-21-1-2187, and N00173-16-1-G905. Recommended for publication by Associate Editor F. Luo. *(Corresponding author: Qian Yang.)*

Qian Yang, Muhammed Ali Gultekin, Krishna Pattipati, and Ravi Rajamani are with the University of Connecticut, Storrs, CT 06269 USA (e-mail: qian.4.yang@uconn.edu; muhammed_ali.gultekin@uconn.edu; krishna.pattipati@uconn.edu; ravi.rajamani@uconn.edu).

Vahe Seferian is with the American University of Beirut, Beirut 1107 2020, Lebanon (e-mail: vss01@mail.aub.edu).

Ali M. Bazzi is with the University of Connecticut, Storrs, CT 06269 USA, and also with the American University of Beirut, Beirut 1107 2020, Lebanon (e-mail: alibazzi@ieee.org).

Francesco A. N. Palmieri is with the Dipartimento di Ingegneria Industriale e dell 'Informazione, Universita degli Studi della Campania "Luigi Vanvitelli", 81031 Aversa, Italy (e-mail: francesco.palmieri@unicampania.it).

Shailesh N. Joshi, Muhamed Farooq, and Hiroshi Ukegawa are with the Toyota Research Institute of North America, Ann Arbor, MI 48105 USA (e-mail: shailesh.joshi@toyota.com; muhamed.farooq@toyota.com; hiroshi.ukegawa@toyota.com).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TPEL.2022.3140721.

Digital Object Identifier 10.1109/TPEL.2022.3140721

I. Introduction

A. Motivation

ELECTRIC vehicles (EVs) have a number of advantages over traditional gas-powered automobiles in terms of efficiency, fuel economy, flexible charging, energy security, and beneficial impact on climate due to reduction in emissions. Power electronic (PE) converters, such as inverters and dc/dc converters, control the flow of electrical energy in EVs. For example, they control the speed and torque of the motor; they convert and distribute electrical power to other vehicle systems, such as heating and ventilation, lighting, and infotainment. Indeed, high-frequency switching circuits used in PE devices [1] are the key to superior performance of the motor drives. Consequently, the reliability of PE components, such as insulated-gate bipolar transistors (IGBT) and metal–oxide–semiconductor field-effect transistors (MOSFET), is becoming a major concern to motor manufacturers [1]–[4], because unscheduled maintenance ensues when faults occur in these devices. Therefore, it is essential to develop fault detection, diagnosis, and prognosis (FDDP) techniques that annunciate incipient device faults for proactive maintenance and increased vehicle availability. Anomaly detection is the first step in ensuring the operational integrity of PE devices [5]. Incipient detection of anomalies can prevent unplanned vehicle breakdowns, and lower the maintenance costs to vehicle owners [4].

B. Previous Work

Anomaly detection in MOSFETs is based on the premise that measurable signals of failed devices, such as the voltage, current, ON-state resistance, and temperature (body diode voltage), will change substantially when compared to normally operating devices [6]. Numerous anomaly detection algorithms for MOSFET devices have been developed in the literature [7]. These approaches can be broadly classified into three general categories: Condition-based, model-based, and data-driven [6], [8] methods.

MOSFET devices have three ports: Source (S), drain (D), and gate (G). The voltage between G and S, V_{GS}, controls the flow of current, I_{DS}, between the drain and the source. As V_{DS} is increased, the drain current I_{DS} increases, but due to the applied V_{GS}, the drain current is controlled up to a certain level

(saturation). MOSFET devices operate in three modes: ON when there is adequate V_{GS} to allow the current to pass, OFF when V_{GS} is below a threshold, and saturation when there is a constant current in I_{DS} independently of voltage, V_{DS}. Chen *et al.* [6] categorize failure modes in MOSFET devices into two distinct categories: Extrinsic failures, such as metallization reconstruction, wire-bonding, and solder joint fatigue, and intrinsic failures, including dielectric breakdown, hot carrier injection, and electromigration. Typically, one measures sampled time histories of V_{GS}, V_{DS}, I_{DS}, and body diode voltage V_{BD}; the latter serves as a surrogate for the case temperature, T_C of the device.

Condition-based methods detect or estimate or measure the electrical parameters, which indicate the degradation of the power device [8]. The parameters, such as damage-sensitive electrical parameters (DSEP) and thermosensitive electrical parameters (TSEP), are usually measured during operation. Fausto proposed a junction temperature estimation system using the V_{ON} measurement that can be integrated into the gate driver circuit [9]. The method showed high accuracy and high feasibility compared to the direct thermistor measurement of the temperature. Zhang [10] introduced a gate model of SiC MOSFET, an online I_G measuring method and a junction temperature estimation technique using I_G. It can be applied to various types of PE devices under different operating conditions.

Model-based approaches utilize the knowledge of MOSFET device's physics of normal operation and anticipated physics of failure mechanisms to develop an analytical or simulation model that relates physical parameters to the observed behavior of MOSFET. Kovacevic presented a physics-based model to simulate the stress–strain solder response under cyclical thermal loading [11]. An analytical PSpice model, applicable to commercially available SiC MOSFET power modules, is presented in [12]. The main advantage of a model-based approach is its ability to incorporate a physical understanding of the device's behavior into the anomaly detection scheme. Because of the nonlinear nature of switching circuit behavior and the complexity and variety of failure mechanisms, mathematically and precisely formulating the effects of faults and degradation in MOSFET devices is especially difficult [13].

A data-driven approach to anomaly detection is preferred when physics-based models are not available or are too complex for real-time use or are difficult to develop, but instead device health-monitoring data (e.g., currents, voltages, temperature) is available. In these cases, experimental data from an operating MOSFET device or from accelerated aging tests or simulated data from a detailed physics-based simulator will be the major source of device knowledge for anomaly detection. Neural network and statistical classification methods are illustrative of data-driven techniques. Significant amount of data from monitored variables under a range of nominal and faulty scenarios is needed for data-driven analysis to succeed. In addition, it is salient to know which signals or transformations of the signals ("features") must be monitored to detect malfunctions at an early stage of failure. There are several possible signatures and precursors, called condition indicators, for monitoring the degradation. These include the gate leakage current [14], I_{DS}, V_{DS}, changes in ON-state resistance ($R_{DS(ON)}$) [15]–[17], gate voltage, V_{GS} and so on [18]. A comprehensive reliability analysis of SiC MOSFETs under high-temperature power cycling was presented in [19], and the correlations between key precursor variations and degradation mechanisms have been assessed. The threshold voltage drift and increase in $R_{DS(ON)}$ is found to be related to the gate oxide degradation and sudden change in body diode voltage and $R_{DS(ON)}$ are due to packaging-related issues.

One of the most common ways of detecting failures involves comparing the signals in the fault-free and the faulty conditions. Particle filter [15], [17], [20] and extended Kalman filter [16] based techniques are employed to track the degradation state (e.g.,changes in $R_{DS(ON)}$) for anomaly detection and to compute the remaining useful life (RUL) of the device. Hidden Markov models (HMMs) are formulated for the healthy and anomalous behaviors of systems in [21] and the measured signals are classified as healthy or anomalous using the trained HMMs. Sutrisno *et al.* [22] extracted several electrical parameters and analyzed them to detect anomalies using the principal component analysis (PCA) and a K-nearest neighbor (KNN) classifier.

Extracting features from failure precursors is the commonly used data processing technique in anomaly monitoring and analysis. De Araujo Ribeiro *et al.* [23] developed a set of diagnostic matrices relating faults in voltage-source inverters used in motor drives (as rows) to errors in the phase voltages of the inverter pole voltage, line voltage, machine voltage, and the neutral voltage (as columns). Anomalies are detected by comparing the measured voltages to their predictions from an analytical model of the voltage-source inverter. Statistical feature estimation using Gaussian process regression (GPR) and genetic algorithm (GA) is proposed in [24] for online anomaly detection in dc/dc converters. In this article, the GPR model estimates the nominal output from training data and extracts seven features of the nominal output signals: The range, mean, standard deviation, skewness, kurtosis, entropy, and centroid of the nominal output signals. Since the nominal output is subject to parameter uncertainty due to component tolerances and instrument noise, GA is used to compute the minimum and maximum values of the seven output features under the nominal case. During online operation, the seven features of the actual output are compared with the extreme values of the nominal case found via GA. The anomaly will be detected online when any one of the seven features of the actual output deviates beyond the extreme values.

Patil *et al.* [25] proposed a Mahalanobis distance (MD)-based anomaly detector by monitoring the ON-state collector–emitter voltage $V_{CE(ON)}$ and current $I_{CE(ON)}$ in IGBTs. The sample mean and the sample covariance matrices used in MD calculations are obtained from the healthy data. The MD was transformed into a nearly Gaussian distribution using Box-Cox transformations and three standard deviation limits are used as thresholds for anomaly detection. Murphey [26] developed a neural network-based fault classification system for an electric drive inverter using data from a model-based simulator.

Residual-based techniques are able to capture the fault signatures and enable active adaptive learning. Chen *et al.* [6] used PCA to project the data onto orthogonal directions, thereby

removing multicollinearity in the data and reducing its dimensionality. Then, the Hotelling's T^2 and Q statistics-based tests are used for anomaly detection. In [27], PCA method was enhanced by combining it with an exponentially weighted moving average (EWMA) filter for tracking nonstationary and operating condition-dependent feature means and a probabilistic logic to improve the detection accuracy and to minimize false alarms.

C. Contributions and Organization of the Article

This article proposes two unsupervised data-driven approaches to learn the nominal model of a device on-line and use the deviations between actual measurements and predictions ("residuals" or "deltas") as an input to a sequential quickest detection scheme. Specifically, nonlinear autoregressive exogenous (NARX) neural network and partial least squares (PLS)-based models are used as two complementary methods for predicting the signals of interest during the ON-state of a switching device (e.g., current, voltage, ON-state resistance, body diode voltage, or case temperature), assuming nominal operation. The NARX model uses ON-state medians of signals in each switching cycle as inputs and outputs, while the PLS model works with filtered ON-state samples directly. Informed by physics-based models, optimal and suboptimal feature selection methods using dynamic programming (DP) are used to select the most informative signals and their transformations as inputs to the PLS models. The learned models of power MOSFET devices are applicable at both low-frequency and high-frequency switching and under various operating conditions. The predictions from the NARX and PLS models are used to compute the residuals; Page's cumulative sum (CUSUM) test monitors the changes in mean or variance of these residuals to detect anomalies. The data preprocessing steps used in the proposed anomaly detection methods suppress the impact of instrument noise via median filtering (for NARX) and a moving average (a form of finite impulse response) digital filter (for PLS), and increase the statistical power of anomaly detectors via extreme data removal. The NARX-CUSUM and PLS-CUSUM anomaly detection methods are validated on three distinct accelerated aging experimental datasets, comprised of 60 power MOSFET devices with low-frequency and high-frequency switching and disparate operating conditions. One of these accelerated test datasets with data from 27 MOSFET devices was generated at the University of Connecticut (UCONN) by setting up a hardware-in-the-loop (HIL) experimental facility. The proposed methods could be implemented both online and offline.

The rest of this article is organized as follows. Section II introduces the hardware setup for the three datasets and details of each dataset and how it was obtained is given. Section III describes the DP-based feature selection algorithms, and interprets the forward and backward feature selection algorithms as approximate DP (ADP) algorithms. Residual generation via NARX neural network and PLS regression models are described in Section IV. Section V briefly explains the CUSUM anomaly detection algorithm. In Section VI, NARX-CUSUM and PLS-CUSUM anomaly detectors are applied to three different datasets comprised of 60 devices and the detection performance results are presented. In Section VII, we discuss the implementation of PLS-CUSUM on a digital signal processor (DSP) and how we optimized the code for memory and computational speed. Finally, Section VIII concludes this article.

II. Hardware Setup and Dataset Description

Data-driven approaches to anomaly detection in MOSFET devices require datasets, spanning both healthy and faulty devices under disparate operating conditions that are expected to occur in real-world EV applications. To test our anomaly detection algorithms, we obtained accelerated aging test data from three experimental setups: SiC MOSFET accelerated aging test data, UCONN HIL data, and UT-Dallas data.

A. Failures in MOSFETs

Power devices degrade and fail in different ways. Failures can be categorized as extrinsic and intrinsic failures. Extrinsic failures are related to the packaging of single or multiple devices, whereas intrinsic failures are related to the device physics [28]. Large package areas and nonhomogeneously sourced materials make the devices vulnerable to thermal stress. Mismatches in thermal expansion coefficients in packaging manifest as wire bond and solder joint failures [29]. Intrinsic failures are caused by power cycling through silicon dies rather than the package itself or the fabrication process; they are listed in the JEDEC standards, as electromigration, corrosion, dielectric breakdown, hot carrier injection, etc. [30].

Both the intrinsic and extrinsic failures are nonobservable without special equipment or tests. In this article, failures are classified based on their electrical behavior, namely, open- and short-circuit failures. Classification of open- and short-circuit failures can be performed using simple electrical tests, but device decapsulation and/or imaging can provide root causes for each fault type and the location within a device.

B. SiC MOSFET Accelerated Aging Test Data

The low-frequency SiC MOSFET accelerated aging test data were generated using a combination of power cycling and thermal cycling tests. The test was set up using the 1500-A Mentor Graphics power tester, which provides automated power testing and cycling for semiconductor devices [6], [31]. Since electric machine drive circuits typically consist of six devices, six different devices, denoted as S1, S2, S3, S4, S5, and S6 (part number CCS050M12CM2 from Wolfspeed) with every pair connected in series, are used. The top switches, S1, S3, and S5, are connected to separate channels of the power tester, denoted as CH1, CH2, and CH3; the bottom switches, S2, S4, and S6, are connected to a common ground, denoted as COM. The reason for using separate channels, but the same voltage, current, and temperature conditions in each channel, is to avoid interaction between the three different half bridges and to ensure that a failure in one leg does not impact others [6]. The devices were mounted on

a temperature-controlled cold plate, the base plate temperature was set to 65 °C. The target junction temperature was set to 150° C, and the temperature was maintained as constant. To maintain a constant temperature, operating current is changed continuously. As for power cycling conditions, devices are switched at a very low frequency (0.02 Hz). The measured signals were case temperature (T_C), ON-state load current (I_{DS}), ON-state voltage (V_{DS}), body diode voltage at maximum junction temperature, T_j (right after switching off current), and minimum body diode voltage at minimum junction temperature, T_j (right before switching on current). The experiment was conducted in three phases. Data from 14 devices are analyzed in this article; three devices, numbered S3.3, S4.3, and S5.3 failed. Devices S3.3 and S5.3 failed in open-circuit mode, while S4.3 was shorted. Imaging analysis showed that the root cause of S3.3 and S5.3 failures was bond wire break and thermal interface material delamination, while the root causes of S4.3 were delamination of die attach and of the thermal interface. Details of the experimental setup and anomaly detection using PCA and EWMA-PCA may be found in our earlier work [6], [27].

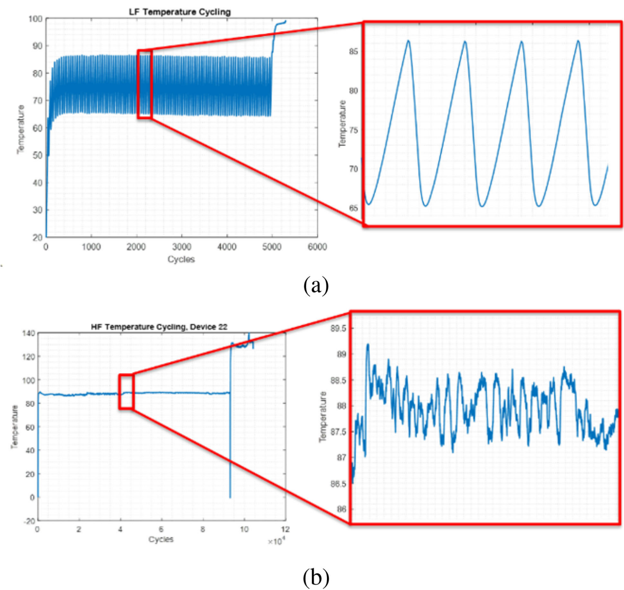


Fig. 1. Temperature variation during accelerated tests. (a) LF. (b) HF.

C. UCONN HIL Data

An accelerated power cycling testbed for Si MOSFETs was established at UCONN to 1) fail devices within hours or days instead of months or years, and 2) generate independent data in a repeatable and reproducible manner, while considering variations inherent in device fabrication, electrical conditions, and switching frequency [32]. Two test types are conducted, low-frequency (LF) and high-frequency (HF). A constant current is switched at 0.02 and 1 kHz with an 80% duty cycle for both the LF and the HF tests. The current is increased periodically to accelerate the device's failure. Temperature of the device's case is not actively controlled. However, in LF testing, wherein devices do not switch ON and OFF frequently, there will be some time for the devices to cool down, as opposed to HF switching, where the switching cycle time is four orders of magnitude smaller. Consequently, in HF testing, temperature is almost constant since the device's thermal time constant is large. Examples of temperature variations during LF and HF testing are shown in Fig. 1(a) and (b), respectively.

The UCONN HIL data are produced by the hardware setup shown in Fig. 2. Only one phase of the inverter is being used for testing purposes. The gate signal is produced by the FPGA, and it is transmitted to the isolation circuitry. After isolating the gate signal, it is fed to the inverter. The following parameters are collected from the inverter using FPGA's analog pins: V_{DS}, V_S, and I_{DS}. Gate signal is also measured, but it is measured directly from the FPGA for synchronization purposes. An EE type thermocouple is used for case temperature measurement, as well as plate temperature measurement. An NI data acquisition device is employed for this purpose. In short, Drain to source voltage V_{DS}, drain current I_{DS}, device case temperature T_C, heat sink temperature T_P, and gate to source voltage V_{GS} are the measured quantities. V_{DS} and I_{DS} are used to find $R_{DS(ON)}$ through Ohm's law.

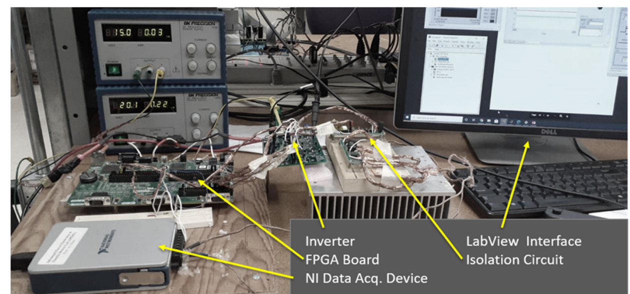


Fig. 2. Test setup.

D. UT-Dallas Data

UT-Dallas data were generated from a thermal cycling setup on C2M0280120D MOSFET devices. DUT's are heated by applying a load current until they reach a set temperature, and then they are cooled down in the absence of load current. Device measurements, taken after each thermal cycle, were provided to us by researchers in [19]. Data from six devices is analyzed (Devices numbered 2A-1 to 2A-7). The tests were stopped because gate charge measurement could not be done for 4 out of 6 devices (2A-1, 2A-2, 2A-5, and 2A-6), or gate control was lost (2A-3) or gate current exceeded the threshold (2A-7).

The datasets above represent 60 devices from multiple data sources (40 from UCONN data, 6 from UT-Dallas data, 14 from SiC MOSFET accelerated aging test data), and disparate operating conditions (different control strategies, low-frequency and high-frequency switching, varying ON–OFF times for low-frequency switching, measurement discontinuities at start–stop points, instrument noise and occasional extreme measurements).

III. Feature Selection Methods

A. Dynamic Programming

In data-driven methods, selecting parsimonious and efficacious features is often better, because it enables learning faster, reduces model complexity and overfitting, improves the accuracy of model predictions, and makes it easier to interpret and explain the model. In developing PLS models to predict I_{DS} from the other measured signals, we considered a number of features, such as $V_{GS}V_{DS}, V_{GS}, V_{GS}^2, V_{DS}$, and T_C as potential inputs to the estimator. The problem of selecting a ranked subset of k-best features (i.e., input signals and their transformations for predicting a desired output signal in our case) from a given set of N features ($k \leq N$) could be solved using branch and bound or DP or heuristics [33]–[35].

DP is a multistage optimization technique used to simplify a complex problem by decomposing it into simpler subproblems, and solving the original problem in a recursive manner. The concept of principle of optimality that states that the subsolutions of an optimal solution of the problem are themselves optimal solutions for their subproblems, in conjunction with a recursive functional equation that relates a solution to its subsolutions, is fundamental to DP. The functional equation of DP enables one to choose a ranked subset of features of specified cardinality with the maximum effectiveness (e.g., minimum mean square error (mse) between the nominal data and the nominal model predictions).

Let $F = (f_1, f_2, \ldots, f_N)$ be the set of available features, X_i, $(i = 1, 2, \ldots, k)$ be a subset of i features and there will be $\binom{N}{i}$ such subsets for each i. To select a subset of k best features among the entire set F, DP considers the subproblems of selecting i-best features, $(i = 1, 2, \ldots, k)$. The recursion also facilitates the ranking of the efficacious features by tracing best feature subsets from stages $i = 1, 2, \ldots, k$.

Let $J(X_i)$ be the cost of subset X_i. Note that, the cost (e.g., mse) depends only on the feature membership of X_i, and not on their order. To guarantee optimality of DP, the performance criterion, J, is assumed to be a strictly monotonic function of i. If this assumption is not valid, then order matters in cost computation and DP becomes a suboptimal feature selection method, because exhaustive search with factorial complexity is needed for optimal feature selection with nonmonotonic cost functions. The forward DP equation for selecting the i-best features based on the mse of a nominal model proceeds for $i = 1, 2, \ldots, k$ with $X_0 = \emptyset$ as follows:

$$J(X_i^*) = \min_{X_i \in F} J(X_i) = \min_{X_{i-1} \in F} \min_{\substack{f \in F \backslash X_{i-1}: \\ X_{i-1} \cup f = X_i}} J(X_{i-1} \cup f). \tag{1}$$

The optimal feature subset (X_i^*) is given as

$$X_i^* = \arg\min_{X_i \in F} J(X_i) \tag{2}$$

$$f_i^* = X_i^* \backslash X_{i-1}^*. \tag{3}$$

Once i best features for $(i = 1, 2, \ldots, k)$ are available, one can glean the ranking of features by traversing the subsets in (2) in a reverse order for $i = k, k-1, \ldots, 1$ as in (3). A feature of DP is that it provides the best feature subset for every $1 \leq i \leq k$. Unfortunately, implementation of such a search scheme requires an enormous amount of computation, especially for large N. Indeed, one needs to consider $\sum_{i=1}^{k} \binom{N}{i}$ nonempty subsets of features, and train and evaluate the concomitant models.

Note that, DP can be implemented as a backward recursion for $i = N-1, N-2, \ldots, k$ with $X_N = F$ as follows:

$$J(X_i^*) = \min_{X_i \in F} J(X_i) = \min_{X_{i+1} \in F} \min_{\substack{f \in X_{i+1}: \\ X_i = X_{i+1} \backslash f}} J(X_{i+1} \backslash f). \tag{4}$$

Note that, ranking of features with the backward method requires the recursion to be computed for $i = N-1, N-2, \ldots, 1$ with computational complexity $O(2^N - 1)$. Given the exponential complexity of DP, we also considered "backward" and "forward" feature selection methods, also known as the "knock-out" strategies. Compared with the DP method, "knock-out" strategies do not need to search all $\sum_{i=1}^{k} \binom{N}{i}$ combinations of features, but need to consider only $1 + \sum_{i=k}^{N-1} (i+1)$ subsets for the backward feature selection method, and $\sum_{i=1}^{k} (N - i + 1)$ subsets for the forward feature selection method.

B. Backward and Forward Feature Selection as Approximate Dynamic Programming

To avoid the exponential computational complexity of the optimal DP algorithm, one could reduce the search space by fixing, at each stage i, the optimal feature subset from stage $(i-1)$. This search process can be realized by starting with all the features and deleting one least effective feature at a time until k features are left, or starting with a null set and adding the highest effectiveness feature at a time to the current feature subset until k features are in the subset; the former is called backward feature selection method, while the latter is termed the forward feature selection method [36].

The backward "knock-out" strategy begins by evaluating the effectiveness of the subset containing all N features and "knocks out" the least effective one (i.e., the feature that increases the mse the least) at each stage, $i = N-1, N-2, \ldots, k$. Starting with the complete set $F = (f_1, f_2, \ldots, f_N)$, we delete the feature f_i^* from F that results in the least decrease in effectiveness (i.e., the least increase in mse). This process is continued until the desired number of features k is reached. The number of possible subsets at the ith stage is $N - i + 1$ and to select best k features, one needs to evaluate $1 + \sum_{i=k}^{N-1} i + 1 = 1 + \frac{(N+k+1)(N-k)}{2}$ subsets. The least effective feature at stage i, $i = N-1, N-2, \ldots, k$ is

$$X_i^* = X_{i+1}^* \backslash f_i^* \tag{5}$$

$$f_i^* = \arg\min_{f \in X_{i+1}^*} \left(J(X_{i+1}^* \backslash f_i) - J(X_{i+1}^*)\right)$$

$$= \arg\min_{f \in X_{i+1}^*} J(X_{i+1}^* \backslash f_i). \tag{6}$$

The initial feature subset is

$$X_N^* = F. \tag{7}$$

Indeed, the backward feature selection via (6) is an approximation to the DP recursion in (4), where minimization over

X_{i+1} is replaced by the subset X^*_{i+1}, thereby reducing the computational complexity to a quadratic function in the number of features. Feature ranking with backward feature selection method, however, requires search process to be continued until $k = 1$. The order of importance of k features using the backward feature selection would be $f^*_1, f^*_2, f^*_3, \ldots, f^*_k$. The computational complexity for ranking and selection in this case is $\frac{N(N+1)}{2}$. Since (6) can be viewed as an approximate DP recursion, the backward feature selection strategy can be used as a base heuristic in rollout and Monte Carlo tree search-based ADP algorithms to provide near-optimal feature subsets [37].

The forward feature selection is basically a bottom up approximate DP search procedure in (1), which adds a new best feature to the current feature set at each stage. At each iteration i $(i = 1, \ldots, k)$, forward feature selection strategy selects the optimal feature f^*_i that provides the largest efficacy, e.g., largest decrease in mse, and adds it to the set of already selected variables, denoted by X^*_{i-1}, initialized to be the empty set $X_0 = \emptyset$. The selection process ends when adding new features does not improve the performance significantly or the desired number of selected features k is reached. At stage i, forward feature selection method requires the evaluation of $(N - i + 1)$ feature subsets and the total number of subsets to be searched is $\sum_{i=1}^{k}(N - i + 1) = \frac{(2N-k+1)k}{2}$.

The subset of i optimal features is

$$X^*_i = X^*_{i-1} \cup f^*_i \tag{8}$$

and the most effective feature f^*_i at ith stage for $i = 1, 2, \ldots, k$ is

$$f^*_i = \arg \min_{f_i \in F \setminus X^*_{i-1}} J(X^*_{i-1} \cup f_i). \tag{9}$$

The forward feature selection method in (9) can be a heuristic for rollout and Monte Carlo tree search-based ADP algorithms to provide near-optimal feature subsets [37].

IV. Residual Generation Methods

In our approach to anomaly detection, a model is learned from data to describe the nominal behavior of the system. A quality index, e.g., residual, is used to compare the predicted output based on the nominal model and the actual output measurement to detect anomalies. The residuals are expected to be near zero and are only subjected to noise, component tolerances, and disturbances in fault-free devices, but deviate from zero substantially in the presence of faults in the system [38]. The residuals are calculated as

$$r(t) = y(t) - \hat{y}(t) \tag{10}$$

where $y(t)$ are the measurements and $\hat{y}(t)$ are the predictions at time t.

A. Data Preprocessing

Data exploration and descriptive statistics [6], data preprocessing, data reduction and transformation, and clustering are four major ways to improve the quality of data for model building and feature reduction. Central tendencies, variations, distributions, and ranges help in setting appropriate thresholds to triage incoming data, in identifying extreme measurements, and in selecting suitable predictive modeling techniques. Data preprocessing filters data to isolate signal components from noise, remove extreme measurements, and adjust scale differences among variables by autoscaling to obtain normalized data [39]. Data transformation, such as PCA and PLS, synthesize a smaller number of features from the redundant measurement data by preserving the correlation structures among the monitored variables and enable real-time implementation of data-driven prediction algorithms via compact memory footprint, with improved computational efficiency and generally enhanced prediction accuracy. The data transformation process enables physics-informed power transformations and basis functions of input signals to be incorporated into predictive models [40]. Data clustering may be viewed as a data reduction process that produces a concise model of data. For example, the context variables, such as the driving cycles, control strategies, system loads, and temperatures, can be grouped to create "cases" that can be used to adapt the model parameters in a new situation. A number of unsupervised clustering algorithms (e.g., K-means [41], Gaussian mixture models [17], and variational Bayes inference [42], SOM [43] can be used to summarize the data for later classification and regression.

In the present article, we used two data preprocessing methods: ON-state median extraction for NARX models to reduce the data size for training and deployment, and implementation of a digital filter to suppress noise in ON-state sampled data for PLS-based model. This is motivated by the observation that the ON-state voltage, current, and resistance have proved to be precursors indicating faults, such as wire-bond degradation [44], [45].

The ON-state raw signals $V_{DS}, V_{GS}, I_{DS}, T_C$, or V_{BD} have inherent noise, which obscures the anomalies underlying the samples. In order to filter out the noise present in ON-state signals, a moving-average filter (MAF) is applied to the ON-state data. The transfer function $H(z)$ of a linear digital filter can be expressed in the z-domain as (11)

$$Y(z) = \frac{\sum_{k=0}^{n_b} b_k z^{-k}}{1 + \sum_{k=1}^{n_a} a_k z^{-k}} X(z) \tag{11}$$

where n_a is the feedback filter order, n_b is the feed-forward filter order. If the denominator is unity (no feedback), (11) becomes a finite impulse response (FIR) filter. Assuming $x(n)$ is the raw signal at sample point n and $y(n)$ is the corresponding processed data sample after digital filtering, the general form of a FIR filtering operation is

$$y_n = \sum_{k=0}^{n_b} b_k x_{n-k}. \tag{12}$$

An MAF is a linear-phase low-pass FIR filter and is commonly used for smoothing noisy data [46]. MAFs are easy to implement and computationally effective. Assuming an MAF window length of n_b $(n_b \in Z^+)$, moving average filtering operation

is (14) [47]

$$y_n = \frac{1}{n_b} \sum_{k=0}^{n_b-1} x_{n-k}. \quad (13)$$

Due to the large computational cost of neural network training, a median filter is applied to reduce the data size for NARX. Median filters are simple and are a robust way to extract features from highly noisy data with occasional extreme values. A nonrecursive median filter is implemented by calculating median values of the samples in a given length window (ON-state cycle length in our case) [48]. Let $x(n)$ and $y(n)$ be the input and output of a median filter at point n and the window length is $2N+1, N \in Z^+$. The median filter operation can be expressed as

$$y_n = \text{median}(x(n-N), \ldots, x(n-1), x(n) \\ x(n+1), \ldots, x(n+N)). \quad (14)$$

B. Nonlinear Autoregressive Model With Exogenous Input Model

Most of the model-based anomaly detection approaches rely on linear discrete-time models, since it is difficult to obtain an accurate model for a complex nonlinear system in practice [49]. To overcome the limitations of the linearized models, neural networks have been introduced, which can work directly with data and have the ability to learn and represent nonlinear and complex relationships between the inputs and outputs, and deal with noisy or corrupted data with suitable data preprocessing and the use of regularization techniques in training [50]. The resulting models can be used to generate residuals, detect anomalies, diagnose faults, trend condition indicators, and estimate the residual useful life.

The NARX inputs neural network, a dynamic artificial neural network (DNN) with delays, can serve as a good predictive model of time series signals of MOSFET devices [51], [52]. NARX uses delays to model the dependence of the current (ON-state median) output on the current and previous values of the exogenous (ON-state median) inputs (features) and previous values of the (ON-state median) output. Consider a discrete-time multivariable nonlinear system with m exogenous inputs, u and r outputs, y

$$\hat{y}(t) = f[u(t-D_u), \ldots, u(t-1), u(t), y(t-D_y), \ldots, \\ y(t-1)] + v(k) \quad (15)$$

$$u(t) = [u_1(t), u_2(t), \ldots, u_m(t)]^T \quad (16)$$

$$y(t) = [y_1(t), y_2(t), \ldots, y_r(t)]^T \quad (17)$$

where $f(\cdot)$ is a nonlinear function, $u(t)$ is a vector of exogenous inputs, $y(t)$ is the vector output at time t and $y(t-D_y), \ldots, y(t-1)$ are the previous values of the output, D_u and D_y are the delay orders of input and output and $v(k)$ is the noise vector. In the context of MOSFET modeling, $y(t)$ could represent, for example, $R_{DS(ON)}$, and $u(t)$ could represent I_{DS}, V_{DS}, T, and V_{GS}.

The mapping function f is unknown, but is approximated by a multilayer perceptron (MLP). In layer j, input vector x_j, an output of the previous layer $(j-1)$, is multiplied by a weighted vector w_{ij} [51]. Then, the output z_i of neuron i of layer j is given by an activation function Ψ as (18)

$$z_i = \Psi\left(\sum_{j=1}^{n} x_j w_{ij}\right). \quad (18)$$

The weight parameters are optimized to minimize the mse between the observed output and model via a stochastic gradient descent method or their accelerated versions [53]–[57]. The gradient is computed via the backpropagation method [50].

C. Partial Least Squares Regression

Unlike the traditional least squares regression models, which assume that only the output is corrupted by noise, the partial least squares (PLS) regression models assume that both inputs and outputs can be corrupted by noise. PLS is particularly useful when the inputs are highly correlated and large in number. The PLS regression is popular in many fields, including chemometrics, biology, biomechanics, robotics, social sciences, and industrial process control [58], [59], where it has become an alternative to multiple linear regression and principal component regression (PCR) methods [60]. PLS regression has connections to conjugate gradient methods used in optimization [61].

Unlike PCR, which characterizes the latent space of only the inputs, PLS characterizes the latent spaces of both inputs and outputs and finds the regression coefficients from the resulting input–output latent space representations. Thus, the PLS regression projects the input feature (predictor) variables and the measured output variables to their respective input and output latent spaces by maximizing the variance of the output variables explained by the input variables or equivalently maximizing the covariance between input and output. Formally, let X be an $n \times m$ matrix of input predictors ($X \in R^{n \times m}$), and Y an $n \times p$ matrix of output responses ($Y \in R^{n \times p}$). The PLS algorithm extracts input and output factors successively from both X and Y, respectively, to maximize the covariance between the input and output. Specifically, PLS algorithm projects the X and Y into uncorrelated latent components of two sets of weights denoted by T and U to maximize the covariance between X and Y (or equivalently, covariance between T and U) as follows:

$$X = TP^T + E \quad (19)$$

$$Y = UQ^T + F \quad (20)$$

where $T \in R^{n \times r}$ and $U \in R^{n \times r}$ are the projections of X (X-scores) and Y (Y-scores), respectively; $P \in R^{m \times r}$ and $Q \in R^{p \times r}$ are orthogonal loading matrices of the input X and the output Y, respectively; and matrices $E \in R^{n \times m}$ and $F \in R^{n \times p}$ are PLS model residuals corresponding to X and Y, assumed to be independent and identically distributed normal random variables. The number of latent variables l can be estimated by using cross-validation or some other techniques [62]. The PLS

method finds the weight vectors w, c such that

$$[(t,u)]^2 = [(Xw, Yc)]^2 = \max_{\|a\|=\|b\|=1} [(Xa, Yb)]^2 \quad (21)$$

where $[(t,u)] = t^T u/n$ denotes the covariance between the score vectors t and u.

Multiple algorithms are available to compute the score and loading matrices, the most popular being the nonlinear iterative partial least squares (NIPALS) algorithm that extracts the X-scores and its loading factors, and the Y-scores and its loading factors, successively one at a time, and successive singular value decomposition of the cross-covariance matrix between input and output using deflation [58]–[61]. Once T, P, U, and Q matrices are computed, the linear regression model relating the input X and the output Y involves finding β by solving

$$U = T\beta \quad (22)$$

where β represents the regression matrix. By substituting the model into (20), we obtain

$$Y = UQ^T + F = T\beta Q^T + F = XP\beta Q^T \quad (23)$$

P, Q, and β can be used to predict y for a given x.

V. Cumulative Sum Anomaly Detection

Sequential change-point detection is concerned with the design and analysis of techniques for quickest online detection of a change in the output signals, subject to a tolerable limit on the risk of a false alarm. In this vein, predictions of the output signals from NARX and PLS models enable us to compute deviations from the actual observations, termed residuals (actual-predicted), which form the basis for change-point detection. Residuals are "small" as long as the behavior of the MOSFET device is normal. In sequential detection, with every new residual, one is faced with the question of whether to let the device to continue to operate or if the statistical behavior of the residual is believed to have altered; the goal is to detect the change as quickly as possible.

CUSUM test (also known as the quickest detection test), first proposed by Page [63], has been widely used in quality control to monitor shifts in the process mean and variance [64], [65]. Processes characterized exclusively by the presence of natural variability are said to operate in the "in-control" condition; when a special cause occurs, the process state is said to be "out-of-control" [65], [66]. The CUSUM test has emerged as a useful tool to detect two-sided shifts in both the process mean and variance and is able to identify the change points rapidly and accurately [66], [67].

For a time sequence data $X(t)$ $(t = 1, 2, 3, \ldots, n)$, which follows the normal distribution, define the null hypothesis that $X \sim N(\mu_0, \sigma_0)$. The target mean μ_0 and standard deviation σ_0 values are estimated from training residuals in our case, describing the normal behavior of MOSFET. The CUSUM test keeps track of the process mean $\bar{X}(t)$ $(t = w, w+1, \ldots, n)$ over w samples

$$\bar{X}(t) = \frac{1}{w} \sum_{j=0}^{w-1} X(t-j) \quad (24)$$

where $\bar{X}(t)$ is the tth sample mean and w is the window length. To monitor an increase in mean values from the in-control level μ_0 to a larger or lower mean μ_1, the following quantities are calculated [66]:

$$C^+(t) = \max(0, C^+(t-1) + \bar{X}(t) - \mu_0 - k) \quad (25)$$

$$C^-(t) = \max(0, C^-(t-1) - \bar{X}(t) + \mu_0 - k) \quad (26)$$

where $k = \frac{\delta\sigma_0}{2}$, δ is the amount of shift in the process mean that we wish to detect, expressed as a multiple of the standard deviation of the data points (which are the sample means). When either $C^+(t)$ and $C^-(t)$ exceeds threshold h, the process is out of control.

CUSUM test is also able to detect changes in variance over time. For a Gaussian time sequence data $X(t)$ $(t = 1, 2, 3, \ldots, n)$, define the null hypothesis that $X \sim N(\mu_0, \sigma_0)$. The CUSUM-$S^2$ test used to detect changes in the process variance, involves computing the sample variance $S^2(t)$ for $(t = w, w+1, \ldots, n)$ as follows:

$$S^2(t) = \frac{1}{w-1} \sum_{j=0}^{w-1} (X(t-j) - \bar{X}(t))^2. \quad (27)$$

To monitor an increase in variance from the in-control level σ_0^2 to a larger variance σ_1^2 $(\sigma_0^2 < \sigma_1^2)$, we use the CUSUM statistic [66]

$$C^+(t) = \max(0, C^+(t-1) + S^2(t) - k). \quad (28)$$

The CUSUM for a downward shift in variance $(\sigma_0^2 > \sigma_1^2)$ is

$$C^-(t) = \min(0, C^-(t-1) + S^2(t) + k) \quad (29)$$

where $k = \frac{2(\ln \frac{\sigma_1}{\sigma_0})\sigma_0^2\sigma_1^2}{\sigma_1^2 - \sigma_0^2}$.

A high-side CUSUM chart is used to detect an increase in σ^2 and to issue an out-of-control signal at the first t for which $C_t^+ > h^+$. Limit h^+ can be obtained using the parameter chart in [68]. A low-side CUSUM chart is used to detect a decrease in σ^2. It issues a signal at the first t for which $C_t^- < -h^-$. A flowchart of the CUSUM test to detect anomalies in residuals is shown in Fig. 3.

VI. Results and Discussion

In this section, the data preprocessing, residual generation, and anomaly detection mechanisms are applied to 60 MOSFET devices spanning three different MOSFET datasets. Case temperature T, drain-to-source voltage (V_{DS}), drain-to-source current (I_{DS}), and gate-to-source voltage (V_{GS}) are measured and used as attributes in the NARX and PLS regression models.

A. *Nonlinear Autoregressive Model-Based Anomaly Detection*

The proposed NARX-CUSUM anomaly detection method is summarized in the block diagram shown in the Fig. 4. The approaches, including data preprocessing (median filtering for NARX and moving average filtering for PLS), residual generation and anomaly detection, were performed on the collected data.

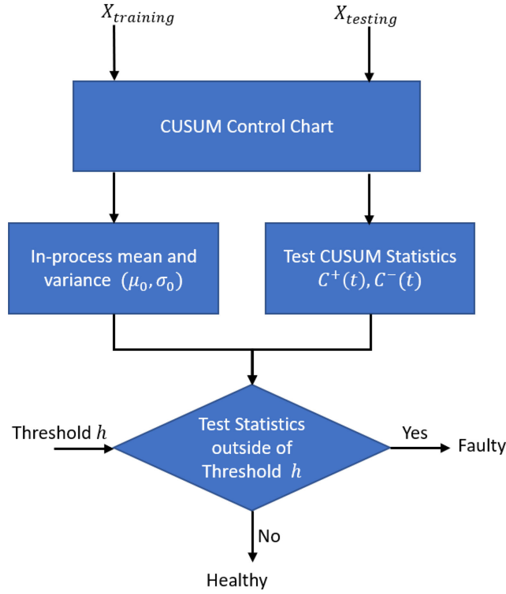


Fig. 3. Workflow of CUSUM chart.

Since every dataset has different numbers of measurements, but T, V_{GS}, I_{DS}, and V_{DS} are the most common and useful parameters, we reduced the feature set to these four. Drain-to-source resistance $R_{DS(ON)}$ is an important indicator of bond wire lift-off, heel cracking, and fractures. Thus, $R_{DS(ON)}$ is calculated from I_{DS} and V_{DS}. The UConn HIL datasets were preprocessed by down sampling the ON-state data to one median sample per cycle, as shown in Fig. 5.

The NARX model uses the past values of the same series, V_{DS}, I_{DS}, $R_{DS(ON)}$, T, and the past and current values of the exogenous input, V_{GS}, as the input. After preprocessing, a one-step delay nonlinear autoregressive model is applied to predict the current ON-state median sample of $T(t)$, $I_{DS}(t)$ and $V_{DS}(t)$, $R_{DS(ON)}(t)$ based on their previous samples and the present $V_{GS}(t)$ and the past sample of the $T(t-1)$, $I_{DS}(t-1)$ and $V_{DS}(t-1)$, $R_{DS(ON)}(t-1)$. The total training data ratio ranged from 10% to 30% of available data based on the data size. For very smaller datasets (as in UCONN data, where devices failed in substantially fewer number of cycles), higher ratio (30%) is used and for the The residuals are then calculated and used for anomaly detection in the CUSUM test. The CUSUM window size w is 100 for SiC MOSFET accelerated aging test data, UT Dallas data, and UCONN HIL high-frequency testing data and 20 for UCONN HIL low-frequency testing data, depending on the data size.

Fig. 6 shows the differences between $R_{DS(ON)}$ predicted values and the true values for four failed devices, S4.3 from SiC MOSFET accelerated aging test dataset, Devices 1, 6, and 18 from the UCONN dataset, which failed as short-circuit in low-frequency testing, open-circuit in low-frequency testing, and short-circuit in high-frequency testing.

CUSUM anomaly detection is applied to the residuals and changes in variance are monitored. Based on training residuals, the in-control level parameters μ_0 and σ_0 are obtained. Optimal amount of shift δ and threshold h are predetermined based on the desired detection delay and risk of false alarm. The amount of shift in the process variance δ was set to 5 and threshold is 10 for SiC MOSFET accelerated aging test datasets, 1 for UCONN-HIL due to the smaller amount of data points after the anomaly happens. Fig. 7 shows the CUSUM statistics for the same four failed devices as in Fig. 6. NARX-CUSUM can reduce false alarms due to the overshoot in transient state, but is less sensitive to small shifts, resulting in a missed detection in HIL 18. Since most of the testing devices stopped recording data when a device failed, the CUSUM statistics exceed the threshold h prior to the actual failure. Thus, NARX-CUSUM is able to detect incipient failures and ahead of the actual failure time as shown in Table I.

B. Partial Least Squares Regression-Based Anomaly Detection

The PLS-CUSUM anomaly detector workflow is depicted in the block diagram shown in the Fig. 8 and was tested on all three MOSFET datasets. In this process, the raw ON-state samples are filtered by an MAF filter with $n_b = 30$. Compared with NARX regression model, regular PLS only takes into account linear relationships among variables. However, the relationship between MOSFET measurements is usually described by a nonlinear function. For example, drain–source current I_{DS} is related to drain–source voltage V_{DS} and the gate–source current V_{DS}, as shown in [69]:

$$I_{DS} = \begin{cases} KF_x KP_x \\ \frac{(V_{GS}-VT_x)V_{DS} - \text{LINSAT}^{kx-1}(V_{GS}-VT_x)^{\frac{2-kx}{kx}}}{1+\theta_x(V_{GS}-VT_x)} \\ \text{if } V_{DS} < \frac{V_{GS}-VT_x}{\text{LINSAT}} \\ KP_x \frac{(V_{GS}-VT_x)^2}{2(1+\theta_x(V_{GS}-VT_x))}, \text{if } V_{DS} \geq \frac{V_{GS}-VT_x}{\text{LINSAT}} \end{cases} \tag{30}$$

$$kx = \frac{KF_x}{KF_x - \frac{\text{LINSAT}}{2}} \tag{31}$$

where KF_x, KP_x are, respectively, the high and low current gain factors, LINSAT is the linear to saturation transition parameter, θ_x is the transverse field mobility parameter, and VT_x is the threshold voltage, which is related to the temperature T. Based on this physics-based model, we assume a model of I_{DS} of the form

$$I_{DS} = a_0 + a_1 V_{GS} V_{DS} + a_2 V_{GS} + a_3 V_{GS}^2 + a_4 V_{DS} + a_5 T. \tag{32}$$

I_{DS} is not linearly related to other terms. Thus, feature transformation and selection are needed to deal with the nonlinearity and avoid the influence of irrelevant variables. As shown in (32), the transformed features are $[V_{GS}V_{DS}, V_{GS}, V_{GS}^2, V_{DS}, T]$.

Backward feature selection-based ADP, developed in Section III, is used to select the most salient transformed features. Table II presents the results. The “Remaining Features” lists the features included, “Deleted column” lists the features deleted according to the algorithm and “MSE” column shows the mse for the healthy data. The initial candidate set $[V_{GS}V_{DS}, V_{GS}, V_{GS}^2, V_{DS}, T]$ included all five features. The mse value increased significantly when the number of features is reduced to two from three, but did not change substantially

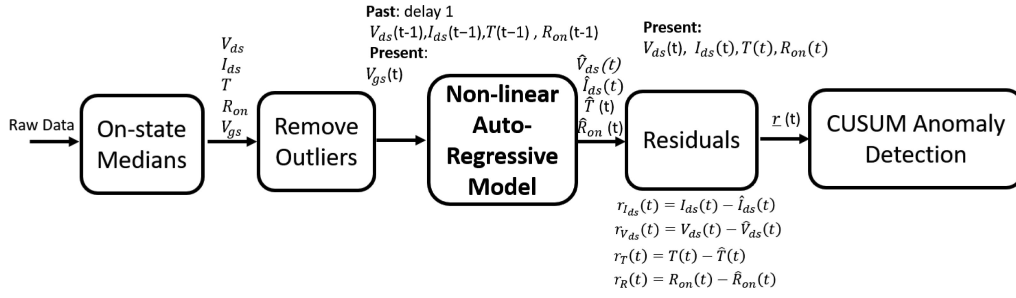


Fig. 4. Architecture of NARX-CUSUM anomaly detector.

TABLE I
SUMMARY OF TIME TO FAILURE FOR NARX-CUSUM

Dataset	Device	Cycle of prediction	Cycle of failure	Remaining useful cycles	Ratio
SiC MOSFET accelerated aging test	S3.3	11,107	12,016	909	92.44%
	S4.3	12,026	20,396	8,370	58.96%
	S5.3	24,051	24,995	943	96.22%
HIL Low Freq.	1	276	284	8	97.18%
	2	184	191	7	96.34%
	3	3,920	3,928	8	99.80%
	5	126	134	8	94.03%
	6	2,091	2,097	6	99.71%
	7	372	381	9	97.64%
	8	-*	4	-*	-*
	9	-*	241	-*	-*
	10	98	107	9	91.59%
	12	190	203	13	93.60%
	13	141	144	3	97.92%
	14	106	119	13	89.08%
	15	5,873	5,904	31	99.47%
	40	-*	58	-*	-*
	41	5,009	5,016	7	99.87%
	42	3,825	4,681	856	81.72%
	43	5,816	6,437	621	90.35%
	44	5,101	5,181	80	98.46%
	45	5,523	5,641	118	97.90%
	46	2,983	3,611	628	82.61%
HIL High Freq.	18	-*	80,400,000	-*	-*
	19	105,900,000	108,609,000	2,709,000	97.51%
	21	81,865,000	97,319,000	15,444,000	84.12%
	22	99,275,000	104,277,000	5,002,000	95.20%
	23	212,931,000	220,999,000	8,068,000	96.35%
	24	37,259,000	37,627,000	368,000	99.02%
	25	37,383,000	37,627,000	244,000	99.35%
	26	18,871,000	19,298,000	409,000	97.79%
	27	9,556,000	9,584,000	28,000	99.71%
	28	18,714,000	18,725,000	11,000	99.94%
	29	258,300,000	258,460,000	160,000	99.94%
	30	158,888,000	159,485,000	597,000	99.63%
	31	185,288,000	185,298,000	10,000	99.99%
	32	597,700,000	597,751,000	51,000	99.99%
	33	245,287,000	245,734,000	447,000	99.82%
	34	278,476,000	289,259,000	10,783,000	96.27%
	35	-*	296,526,000	-*	-*
	37	145,494,000	145,619,000	125,000	99.91%
	38	435,492,000	450,532,000	15,040,000	96.66%
	39	208,070,000	210,531,000	2,461,000	98.83%
UT Dallas	1	2,475	2,467	7 cycle detection delay **	-**
	2	4,771	4,750	21 cycle detection delay **	-**
	3	5,068	5,162	94	98.18%
	5	3,813	6,000	2,187	63.55%
	6	6,036	6,000	36 cycle detection delay**	-**
	7	4,563	7,250	2,687	62.94%

*Missed detection.
**Anomaly is detected after failure happens.

TABLE II
BACKWARD FEATURE SELECTION BASED ON MEAN SQUARE ERROR (MSE)

Num. of Features	Remaining Features	Deleted Features	MSE
5	$V_{GS}V_{DS}, V_{GS}, V_{GS}^2, V_{DS}, T$	-	**0.0312**
4	$V_{GS}V_{DS}, V_{GS}, V_{GS}, T$	V_{DS}	0.173
	$V_{GS}V_{DS}, V_{GS}, V_{GS}, V_{DS}$	T	0.035
	$V_{GS}V_{DS}, V_{GS}, V_{GS}, V_{DS}, T$	V_{GS}^2	0.0338
	$V_{GS}V_{DS}, V_{GS}^2, V_{DS}, T$	V_{GS}	**0.0337**
	$V_{GS}, V_{GS}^2, V_{DS}, T$	$V_{GS}V_{DS}$	0.0343
3	$V_{GS}V_{DS}, V_{GS}^2, T$	V_{DS}, V_{GS}	0.269
	$V_{GS}V_{DS}, V_{GS}^2, V_{DS}$	T, V_{GS}	0.0372
	$V_{GS}V_{DS}, V_{DS}, T$	V_{GS}^2, V_{GS}	0.037
	V_{GS}^2, V_{DS}, T	$V_{GS}V_{DS}, V_{GS}$	**0.0365**
2	V_{GS}^2, T	$V_{DS}, V_{GS}V_{DS}, V_{GS}$	0.294
	V_{GS}^2, V_{DS}	$T, V_{GS}V_{DS}, V_{GS}$	0.377
	V_{DS}, T	$V_{GS}^2, V_{GS}V_{DS}, V_{GS}$	**0.143**
1	V_{DS}	$V_{GS}^2, V_{GS}V_{DS}, V_{GS}, T$	**0.405**
	T	$V_{GS}^2, V_{GS}V_{DS}, V_{GS}, V_{DS}$	0.926

Bold values are MSE values for the best feature subsets.

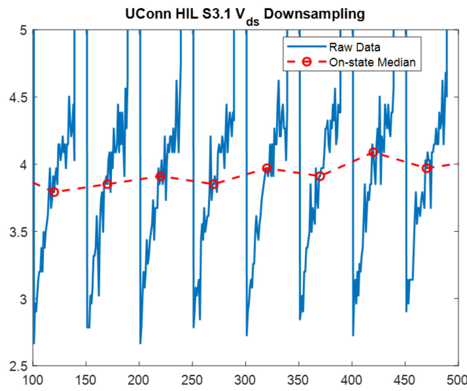


Fig. 5. Preprocessing of the raw data: ON-state medians.

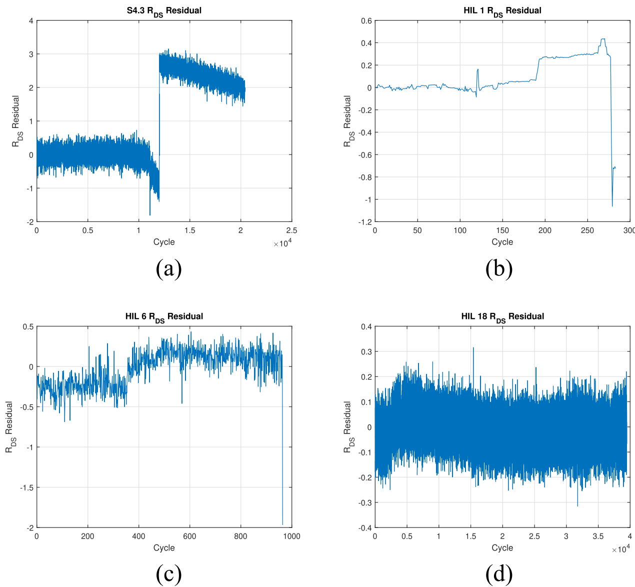


Fig. 6. NARX R_{ds} residuals. (a) Short circuit: SiC MOSFET accelerated aging test S4.3. (b) Short Circuit: Low-frequency HIL 1. (c) Open Circuit: Low-frequency HIL 6. (d) Short Circuit: High-frequency HIL 18.

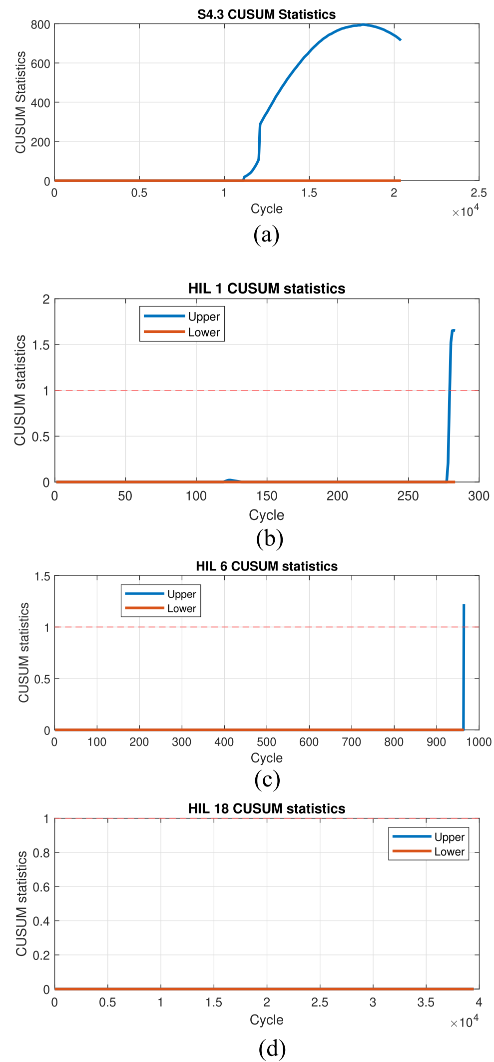


Fig. 7. NARX CUSUM statistics. (a) Short circuit: SiC MOSFET accelerated aging test S4.3. (b) Short Circuit: Low-frequency HIL 1. (c) Open Circuit: Low-frequency HIL 6. (d) Short Circuit: High-frequency HIL 18.

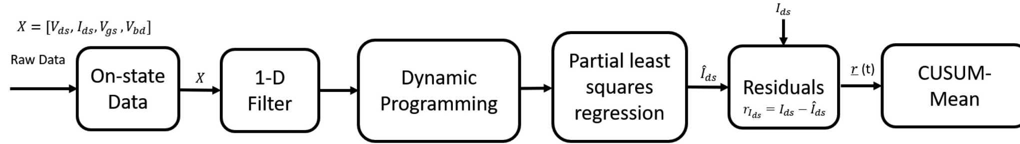


Fig. 8. Architecture of PLS-CUSUM anomaly detector.

TABLE III
PERFORMANCE EVALUATION FOR THE REDUCED MODELS

Feature subset	Accuracy	Missed detection rate
V_{GS}^2, V_{DS}, T	100%	0%
V_{DS}, T	96.7%	4.1%
T	86.7%	16.3%

TABLE IV
PERFORMANCE EVALUATION FOR THE MODEL WITH RAW FEATURES

Feature subset	Accuracy	Missed detection rate
V_{GS}^2, V_{DS}, T	100%	0%
V_{GS}, V_{DS}, T_c	100%	0%
$V_{GS}, V_{DS}, T_c, \Delta T$	85.2%	18.6%
V_{GS}, V_{DS}, T_c, T_p	83.3%	20.9%
$V_{GS}, V_{DS}, T_c, T_p, \Delta T$	83.3%	20.9%

when the number of features is three when compared to including all five features. Thus, we deleted the two features, $V_{GS}V_{DS}, V_{GS}$, with higher mse. The optimal *ranked* feature subset is $[V_{DS}, T, V_{GS}^2]$. The optimal DP and forward feature selection also gave the the same optimal ranked feature subset.

Our feature selection method provides the optimal subsets for different numbers of features. The anomaly detection performance of the reduced order models with 1, 2, and 3 features are shown in Table III. Evidently, a model with 3 features $\{V_{GS}^2, V_{DS}, T\}$ has the highest accuracy; consequently, we will use the model with three features in subsequent analysis.

To evaluate the effectiveness of the feature selection and feature transformation methods, the performance for the raw features without transformation are shown in Table IV. Since V_{GS} is almost constant in our experiments, the performance of $\{V_{GS}^2, V_{DS}, T\}$ and $\{V_{GS}, V_{DS}, T\}$ are the same. Including plate temperature T_p and the ΔT, the difference between plate T_p and case temperature T_c, decreases the performance. This is because the plate temperature is noisy.

As explained in Section V, the healthy data should be used for training, which involves the first few thousands of samples,depending on the data size, and the rest of the data can be used for testing to see if an anomaly has occurred. Fig. 9 shows the residuals for the same six failed devices as in Fig. 6. The residual for UCONN HIL 6 in Fig. 9(c) corresponds to an open-circuit failure mode. The remaining three devices, in Fig. 9(a), (b), and (d), fail in a short-circuit failure mode.

The results of CUSUM detection for mean changes are in Fig. 10. The high CUSUM value (upper statistic) detects a positive shift in mean and low CUSUM value (lower statistic) detects a negative shift in mean. If the process mean shifts upward, the upper CUSUM test statistic will eventually drift upward, and vice versa if the process mean decreases. Since the faults happen at the end of the data for all datasets,

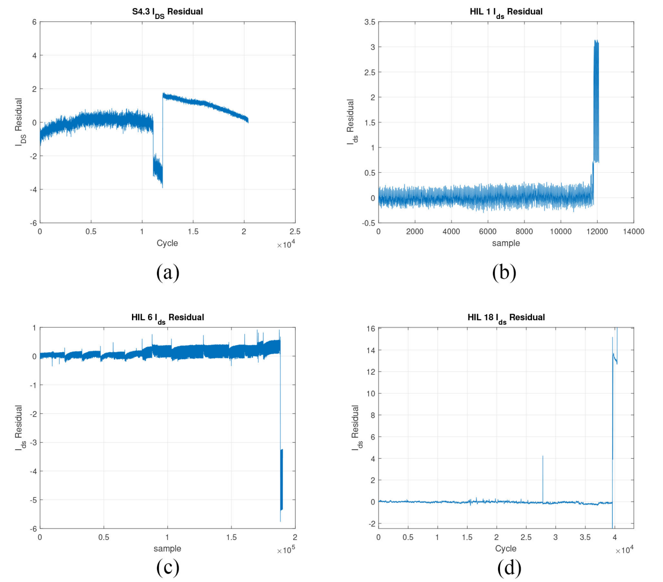


Fig. 9. PLS I_{ds} residual. (a) Short circuit: SiC MOSFET accelerated aging test S4.3. (b) Short Circuit: Low-frequency HIL 1. (c) Open Circuit: Low-frequency HIL 6. (d) Short Circuit: High-frequency HIL 18.

the CUSUM statistics exceed the threshold before the actual failure happens. This means the PLS-CUSUM method can forecast the failures before they happen. The number of remaining useful cycles is calculated from the cycle when either the upper or the lower CUSUM statistic exceeds the threshold to the cycle corresponding to the device failure, and are summarized in Table V. PLS-CUSUM has the ability to detect incipient failures of different types under different experimental conditions. The CUSUM window size w is 200 for SiC MOSFET accelerated aging tests data, UT Dallas data, and UCONN-HIL high-frequency testing data and 20 for UCONN-HIL low-frequency testing data, depending on the data size. Here, the CUSUM δ is defined as five standard deviations (99.99994% confidence) from the training I_{DS} residual mean and the threshold h is selected as 10.

We compared the performance of the proposed two methods with PCA and EWMA-PCA proposed by Chen *et al.* [6] and McMenemy *et al.* [27], respectively. The performance of the approaches are evaluated using accuracy (33), precision (34), recall (35), F-measure (36) [70], and the Cohen's Kappa statistic (37) [71] in Table VI. True and false positives (TP/FP) are the number of predicted positives that were correct/incorrect, and true and false negatives (TN/FN) are the number of predicted negatives that were correct/incorrect

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + FN + TN} \tag{33}$$

TABLE V
SUMMARY OF TIME TO FAILURE FOR PLS-CUSUM

Dataset	Device	Cycle of prediction	Cycle of failure	Remaining useful cycles	Ratio
SiC MOSFET accelerated aging test	S3.3	10,018	12,016	1,998	83.37%
	S4.3	12,570	20,396	7,826	61.63%
	S5.3	19,042	24,995	5,953	76.18%
HIL Low Freq.	1	274	284	10	96.48%
	2	185	191	6	96.86%
	3	3,911	3,928	17	99.57%
	5	100	134	34	74.63%
	6	2,090	2,097	7	98.78%
	7	377	381	4	98.95%
	8	3	4	1	75.00%
	9	235	241	6	97.51%
	10	59	107	48	55.14%
	12	191	203	12	94.09%
	13	97	144	47	67.36%
	14	61	119	58	51.26%
	15	5,862	5,904	42	99.29%
	40	39	58	19	66.78%
	41	4,906	5,016	110	97.81%
	42	4,635	4,681	46	99.02%
	43	5,797	6,437	640	90.05%
	44	5,113	5,181	68	98.69%
	45	5,563	5,641	79	98.61%
	46	2,696	3,611	915	74.66%
HIL High Freq.	18	77,579,000	80,400,000	2,821,000	98.69%
	19	104,800,000	108,609,000	3,809,000	97.13%
	21	80,660,000	97,319,000	16,659,000	98.24%
	22	97,870,000	104,277,000	6,407,000	93.86%
	23	212,880,000	220,999,000	8,119,000	96.32%
	24	37,417,000	37,627,000	210,000	97.45%
	25	36,439,000	37,627,000	1,188,000	99.08%
	26	18,645,000	19,298,000	653,000	96.62%
	27	9,504,000	9,584,000	80,000	99.17%
	28	17,719,000	18,725,000	1,006,000	94.63%
	29	241,142,000	258,460,000	17,318,000	93.30%
	30	152,719,000	159,485,000	6,766,000	77.21%
	31	173,611,000	185,298,000	11,687,000	96.16%
	32	582,531,000	597,751,000	15,220,000	97.45%
	33	244,932,000	245,734,000	802,000	99.67%
	34	279,015,000	289,259,000	10,244,000	96.46%
	35	292,054,000	296,526,000	4,472,000	98.49%
	37	112,588,000	145,619,000	33,031,000	77.32%
	38	435,824,000	450,532,000	14,708,000	96.66%
	39	207,585,000	210,531,000	2,946,000	98.60%
UT Dallas	1	2,387	2,467	80	96.76%
	2	4,789	4,750	39 cycle detection delay **	-
	3	5,089	5,162	73	98.59%
	5	5,255	6,000	745	87.58%
	6	4,984	6,000	1,016	83.07%
	7	4,581	7,250	2,669	63.19%

*Missed detection of B5025.
**Anomaly is detected after failure happens.

TABLE VI
OVERALL PERFORMANCE OF ANOMALY DETECTION ALGORITHMS

Method	Accuracy	Precision	Recall	F-meansure	Kappa
PCA	0.82	1	0.76	0.87	0.58
EWMA-PCA	0.84	1	0.76	0.88	0.65
NARX-CUSUM	0.92	1	0.90	0.95	0.90
PLS-CUSUM	1	1	1	1	1

$$\text{Precision} = \frac{TP}{TP + FP} \tag{34}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{35}$$

$$F_1 \text{ score} = \frac{2 \cdot TP}{2 \cdot TP + FP + FN} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \tag{36}$$

$$\kappa = \frac{\Pr(a) - \Pr(e)}{1 - \Pr(e)} \tag{37}$$

where $\Pr(a)$ is the actual observed agreement, and $\Pr(e)$ is chance agreement. The result of Cohen's Kappa can be interpreted as: $\kappa \leq 0$ as indicating no agreement and 0.01–0.20 as none to slight, 0.21–0.40 as fair, 0.41–0.60 as moderate, 0.61–0.80 as substantial, and 0.81–1.00 as almost perfect agreement.

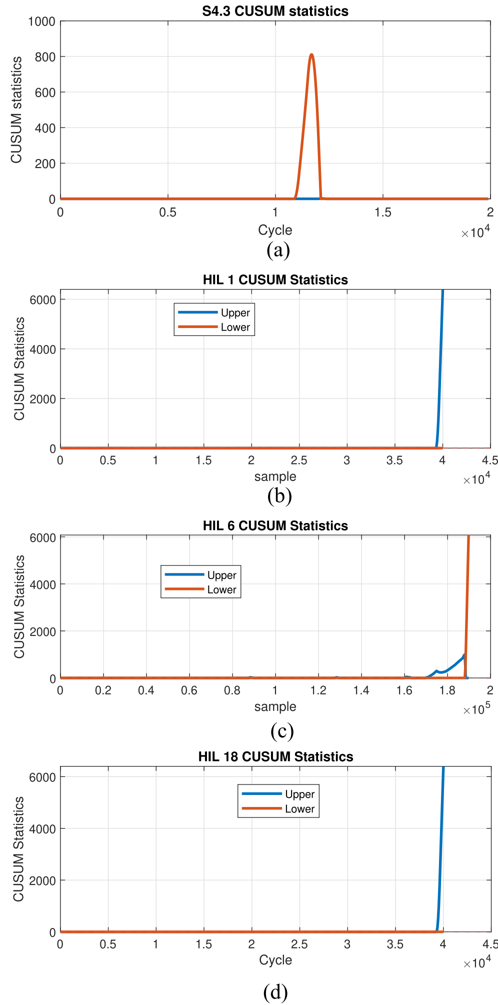


Fig. 10. CUSUM statistics of PLS I_{ds} residual. (a) Short circuit: SiC MOSFET accelerated aging test S4.3. (b) Short Circuit: Low-frequency HIL 1. (c) Open Circuit: Low-frequency HIL 6. (d) Short Circuit: High-frequency HIL 18.

TABLE VII
CROSS TABULATION OF NUMBER OF CORRECT AND WRONGLY CLASSIFIED FOR TWO APPROACHES

		PLS-CUSUM	
		Wrong	Correct
NARX-CUSUM	Wrong	$f_{11} = 0$	$f_{12} = 5$
	Correct	$f_{21} = 0$	$f_{22} = 55$

The results in Table VI clearly demonstrate that the NARX-CUSUM and PLS-CUSUM perform substantially better than PCA and EWMA-PCA methods.

We compared the performance of NARX-CUSUM and PLS-CUSUM using the McNemar test [72]. The McNemar test is based on a cross tabulation in Table VII and the null assumption is that the two populations for each outcome are the same for the NARX-CUSUM and the PLS-CUSUM. The McNemar test statistic follows a chi-squared distribution with one degree of freedom and is computed as follows:

$$\chi^2 = \frac{(f_{12} - f_{21})^2}{f_{12} + f_{21}} = \frac{(5-0)^2}{5} = 5. \tag{38}$$

Thus, the null assumption of equal performance is accepted under a 95% confidence level.

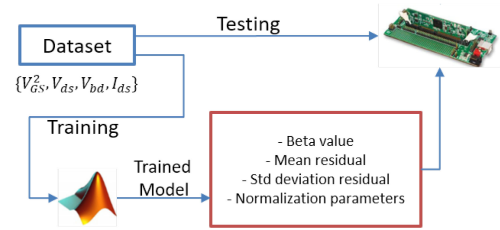


Fig. 11. HIL implementation.

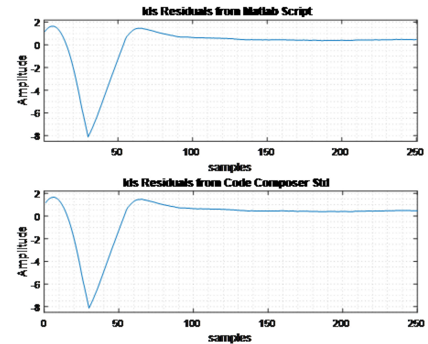


Fig. 12. HIL 18.

VII. HIL IMPLEMENTATION

We used processor-in-the-loop (PIL) testing on a DSP to evaluate the feasibility of the proposed anomaly detection methods. The implementation of the system is shown in Fig. 11.

MATLAB is used to communicate with the DSP and C-code is compiled and uploaded to the DSP. After we run the algorithms on the board, the results are sent back to MATLAB for visualization. The implementation of PLS-CUSUM on the PIL is realized by training the model in MATLAB, and the PLS coefficients, the mean and standard deviation values of features are stored as the nominal parameters. Then, the testing process is performed on the DSP and the CUSUM statistics C_n^+ and C_n^- are calculated. The implementation uses the HIL 18 as an example. The first 10% of the data points are used as the training data when the device is in a nominal condition, and the entire dataset is used as testing data. The data preprocessing, including median filtering, feature transformation, training the regression model, and calculating nominal mean and standard deviation values are performed on the computer hosting the MATLAB. The calculations in the DSP include preprocessing of the testing data, generating the residuals and computing the CUSUM statistics of the residuals, and evaluating the health condition of the device. The processing time of the DSP implementation takes around 2.01 ms on TMS320F28335 DSP board. It corresponds to less than 0.5% of the sampling time even with autocode generation. This small duty cycle assures us that there will not be any aliasing or overflow errors in between samples. The code performance can be further optimized (e.g., precomputing the past values needed in the NARX-CUSUM implementation during idle time) to guarantee the integrity of computations. The PIL generated PLS I_{DS} residual of first 250 data points of HIL 18 is compared with M-file generated results and shown in Fig. 12. The PIL test

demonstrates that the algorithm can be implemented online and that it gives identical outputs as those produced by MATLAB.

VIII. Conclusion

In this article, we proposed two residual-based machine learning models and algorithms, NARX-CUSUM and PLS-CUSUM, for incipient detection of anomalies in MOSFETs. The main goal is to manage the health of power electronic devices proactively rather than reactively after a system breakdown occurs. The data preprocessing and feature selection are performed to suppress noise, remove extreme data due to instrument errors and to obtain the best feature subset. The healthy data are used to train the NARX and PLS models and the training residuals yield the mean and standard deviation values of their distribution under normal conditions, termed the baseline. The residuals from the test sets are used for CUSUM detection by comparing the observed distribution with the baseline.

Datasets from both high- and low-frequency accelerated aging tests under a variety of experimental conditions (different control strategies, varying ON–OFF times for low-frequency switching, measurement discontinuities at start-stop points, instrument noise, and occasional extreme measurements) from three different datasets spanning 60 devices were used to demonstrate the utility of the anomaly detection approaches. The methods were able to achieve an overall 100% accuracy. This article illustrates how using residuals of multiple signals, and a sequential analysis technique to detect anomalies based on the healthy behavior provides improved early detection of failures in MOSFETs. Compared with the traditional PCA and EWMA-based PCA method of our earlier work, the proposed methods have higher accuracy and faster detection times. The results demonstrate that the proposed methods can effectively detect anomalies in MOSFETs regardless of circuit structures and failure types and are able to be implemented on a digital processor for real-time anomaly detection. In the future, we plan to pursue a number of research avenues, including 1) optimizing the methods on a digital processor for real-time anomaly detection; 2) prediction of remaining useful life; and 3) determining the fault types using a semisupervised active learning framework.

References

[1] H. Al-Sheikh, O. Bennouna, G. Hoblos, and N. Moubayed, "Study on power converters used in hybrid vehicles with monitoring and diagnostics techniques," in *Proc. 17th IEEE Mediterranean Electrotechnical Conf.*, 2014, pp. 103–107.

[2] R. Jayabalan and B. Fahimi, "Monitoring and fault diagnosis of multi-converter systems in hybrid electric vehicles," *IEEE Trans. Veh. Technol.*, vol. 55, no. 5, pp. 1475–1484, Sep. 2006.

[3] G. Breglio, A. Irace, P. Spirito, R. Letor, and S. Russo, "Fast transient infrared thermal analysis of smart power MOSFETS in permanent short circuit operation," in *Proc. IEEE Int. Symp. Power Semicond. Devices IC's*, 2006, pp. 1–4.

[4] J. He, N. A. Demerdash, N. Weise, and R. Katebi, "A fast on-line diagnostic method for open-circuit switch faults in SiC-MOSFET-based t-type multilevel inverters," *IEEE Trans. Ind. Appl.*, vol. 53, no. 3, pp. 2948–2958, May/Jun. 2017.

[5] F. A. Tobar, L. Yacher, R. Paredes, and M. E. Orchard, "Anomaly detection in power generation plants using similarity-based modeling and multivariate analysis," in *Proc. Amer. Control Conf.*, 2011, pp. 1940–1945.

[6] W. Chen, L. Zhang, K. Pattipati, A. M. Bazzi, S. Joshi, and E. M. Dede, "Data-driven approach for fault prognosis of SiC MOSFETS," *IEEE Trans. Power Electron.*, vol. 35, no. 4, pp. 4048–4062, Apr. 2020.

[7] J. M. Anderson and R. W. Cox, "On-line condition monitoring for MOSFET and IGBT switches in digitally controlled drives," in *Proc. IEEE Energy Convers. Congr. Expo.*, 2011, pp. 3920–3927.

[8] F. Foube, "Power devices health condition monitoring: A review of recent papers," *PHME_CONF*, vol. 6, no. 1, p. 15, Jun. 2021.

[9] F. Stella, G. Pellegrino, E. Armando, and D. Daprà, "Online junction temperature estimation of SiC power MOSFETS through on-state voltage mapping," *IEEE Trans. Ind. Appl.*, vol. 54, no. 4, pp. 3453–3462, Jul./Aug. 2018.

[10] Q. Zhang and P. Zhang, "An online junction temperature monitoring method for SiC MOSFETS based on a novel gate conduction model," *IEEE Trans. Power Electron.*, vol. 36, no. 10, pp. 11087–11096, Oct. 2021.

[11] I. Kovacevic, U. Drofenik, and J. W. Kolar, "New physical model for lifetime estimation of power modules," in *Proc. Int. Power Electron. Conf.-ECCE ASIA*, 2010, pp. 2106–2114.

[12] D. Johannesson and M. Nawaz, "Analytical PSpice model for SiC MOSFET based high power modules," *Microelectron. J.*, vol. 53, pp. 167–176, 2016.

[13] M. Baharani, M. Biglarbegian, B. Parkhideh, and H. Tabkhi, "Real-time deep learning at the edge for scalable reliability modeling of Si-MOSFET power electronics converters," *IEEE Internet Things J.*, vol. 6, no. 5, pp. 7375–7385, Oct. 2019.

[14] F. Erturk and B. Akin, "A method for online ageing detection in SiC MOSFETS," in *Proc. IEEE Appl. Power Electron. Conf. Expo.*, 2017, pp. 3576–3581.

[15] L.-F. Wu, Y. Guan, X.-J. Li, and J. Ma, "Anomaly detection and degradation prediction of MOSFET," *Math. Problems Eng.*, vol. 2015, 2015, Art. no. 573980.

[16] S. Saha, J. R. Celaya, V. Vashchenko, S. Mahiuddin, and K. F. Goebel, "Accelerated aging with electrical overstress and prognostics for power MOSFETS," in *Proc. IEEE EnergyTech*, 2011, pp. 1–6.

[17] J. Celaya, A. Saxena, S. Saha, and K. F. Goebel, "Prognostics of power MOSFETs under thermal stress accelerated aging using data-driven and model-based methodologies," *PHM_CONF*, vol. 3, no. 1, Sep. 2011.

[18] M. A. Rodríguez, A. Claudio, D. Theilliol, and L. Vela, "A New fault detection technique for IGBT based on gate voltage monitoring," in *Proc. IEEE Power Electron. Specialists Conf.*, 2007, pp. 1001–1005.

[19] E. Ugur, F. Yang, S. Pu, S. Zhao, and B. Akin, "Degradation assessment and precursor identification for SiC MOSFETS under high temp cycling," *IEEE Trans. Ind. Appl.*, vol. 55, no. 3, pp. 2858–2867, May/Jun. 2019.

[20] Y. Lu and A. Christou, "Prognostics of IGBT modules based on the approach of particle filtering," *Microelectron. Rel.*, vol. 92, pp. 96–105, 2019.

[21] E. Dorj, C. Chen, and M. Pecht, "A Bayesian hidden Markov model-based approach for anomaly detection in electronic systems," in *Proc. IEEE Aerosp. Conf.*, 2013, pp. 1–10.

[22] E. Sutrisno, Q. Fan, D. Das, and M. Pecht, "Anomaly detection for insulated gate bipolar transistor (IGBT) under power cycling using principal component analysis and K-nearest neighbor algorithm," *J. Washington Acad. Sci.*, vol. 98, no. 1, pp. 1–8, 2012.

[23] R. L. de Araujo Ribeiro, C. B. Jacobina, E. C. da Silva, and A. M. N. Lima, "Fault detection of open-switch damage in voltage-fed PWM motor drive systems," *IEEE Trans. Power Electron.*, vol. 18, no. 2, pp. 587–593, Mar. 2003.

[24] Y. Jiang, Y. Yu, and X. Peng, "Online anomaly detection in DC/DC converters by statistical feature estimation using GPR and GA," *IEEE Trans. Power Electron.*, vol. 35, no. 10, pp. 10945–10957, Oct. 2020.

[25] N. Patil, D. Das, and M. Pecht, "Anomaly detection for IGBTs using mahalanobis distance," *Microelectron. Rel.*, vol. 55, no. 7, pp. 1054–1059, 2015.

[26] Y. L. Murphey, M. A. Masrur, Z. Chen, and B. Zhang, "Model-based fault diagnosis in electric drives using machine learning," *IEEE/ASME Trans. Mechatronics*, vol. 11, no. 3, pp. 290–303, Jun. 2006.

[27] D. McMenemy, W. Chen, L. Zhang, K. Pattipati, A. M. Bazzi, and S. Joshi, "A machine learning approach for adaptive classification of power MOSFET failures," in *Proc. IEEE Transp. Electrific. Conf. Expo.*, 2019, pp. 1–8.

[28] W. Wu, M. Held, P. Jacob, P. Scacco, and A. Birolini, "Thermal stress related packaging failure in power IGBT modules," in *Proc. Int. Symp. Power Semicond. Devices and IC's*, 1995, pp. 330–334.

[29] A. Morozumi, K. Yamada, T. Miyasaka, S. Sumi, and Y. Seki, “Reliability of power cycling for IGBT power semiconductor modules,” *IEEE Trans. Ind. Appl.*, vol. 39, no. 3, pp. 665–671, May/Jun. 2003.
[30] *Failure Mechanisms and Models for Semiconductor Devices*, JEDEC Standard JEP122C, Mar. 2006.
[31] “Power tester 1500A,” 2014. [Online]. Available: http://s3.mentor.com/public_documents/datasheet/products/mechanical/power-tester-1500a-ds.pdf
[32] M. A. Gultekin *et al.*, “High-and low-frequency accelerated stress tests for aging assessment of MOSFET parameters,” in *Proc. IEEE Int. Electr. Mach. Drives Conf.*, 2021, pp. 1–4.
[33] A. Cohen and Y. Zigel, “On feature selection for speaker verification,” in *Proc. COST 275 Workshop Advent Biometrics Internet*, 2002, pp. 89–92.
[34] N. Acır, “Classification of ECG beats by using a fast least square support vector machines with a dynamic programming feature selection algorithm,” *Neural Comput. Appl.*, vol. 14, no. 4, pp. 299–309, 2005.
[35] M. T. Emmerich and A. H. Deutz, “A tutorial on multiobjective optimization: Fundamentals and evolutionary methods,” *Natural Comput.*, vol. 17, no. 3, pp. 585–609, 2018.
[36] S. Abe, “Modified backward feature selection by cross validation,” in *Proc. 13th Eur. Symp. Artif. Neural Netw.*, 2005, pp. 163–168.
[37] D. P. Bertsekas, *Reinforcement Learning and Optimal Control*. Belmont, MA, USA: Athena Scientific, 2019.
[38] J. Gertler, *Fault Detection and Diagnosis in Engineering Systems*. Boca Raton, FL, USA: CRC Press, 1998.
[39] S. García, S. Ramírez-Gallego, J. Luengo, J. M. Benítez, and F. Herrera, “Big data preprocessing: Methods and prospects,” *Big Data Analytics*, vol. 1, no. 1, 2016, Art. no. 9.
[40] C. J. Alonso-González, B. Pulido, M. Cartón, and A. Bregon, “A big data architecture for fault prognostics of electronic devices: Application to power MOSFETS,” *IEEE Access*, vol. 7, pp. 102160–102173, 2019.
[41] O. F. Eker, F. Camci, A. Guclu, H. Yilboga, M. Sevkli, and S. Baskan, “A simple state-based prognostic model for railway turnout systems,” *IEEE Trans. Ind. Electron.*, vol. 58, no. 5, pp. 1718–1726, May 2011.
[42] D. P. Kingma and M. Welling, “Auto-encoding variational Bayes,” in *Proc. 2nd Int. Conf. Learn. Representations*, 2014. [Online]. Available: https://arxiv.org/abs/1312.6114v10
[43] M. Wong, L. Jack, and A. Nandi, “Modified self-organising map for automated novelty detection applied to vibration signal monitoring,” *Mech. Syst. Signal Process.*, vol. 20, no. 3, pp. 593–610, 2006.
[44] A. Alghassi, S. Perinpanayagam, M. Samie, and T. Sreenuch, “Computationally efficient, real-time, and embeddable prognostic techniques for power electronics,” *IEEE Trans. Power Electron.*, vol. 30, no. 5, pp. 2623–2634, May 2015.
[45] X. Sun, M. Huang, Y. Liu, and X. Zha, “Investigation of artificial neural network algorithm based IGBT online condition monitoring,” *Microelectronics Rel.*, vol. 88, pp. 103–106, 2018.
[46] S. Golestan, M. Ramezani, J. M. Guerrero, F. D. Freijedo, and M. Monfared, “Moving average filter based phase-locked loops: Performance analysis and design guidelines,” *IEEE Trans. Power Electron.*, vol. 29, no. 6, pp. 2750–2763, Jun. 2014.
[47] C. Ya-Lun, *Statistical Analysis: With Business and Economic Applications*. Austin, TX, USA: Holt, Rinehart and Winston, 1963.
[48] G. Arce and M. McLoughlin, “Theoretical analysis of the max/median filter,” *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 35, no. 1, pp. 60–69, Jan. 1987.
[49] A. Fekih, H. Xu, and F. N. Chowdhury, “Neural networks based system identification techniques for model based fault detection of nonlinear systems,” *Int. J. Innov. Comput., Inf. Control*, vol. 3, no. 5, pp. 1073–1085, 2007.
[50] C. M. Bishop *et al.*, *Neural Networks for Pattern Recognition*. London, U.K.: Oxford Univ. Press, 1995.
[51] Z. Boussaada, O. Curea, A. Remaci, H. Camblong, and N. Mrabet Bellaaj, “A nonlinear autoregressive exogenous (NARX) neural network model for the prediction of the daily direct solar radiation,” *Energies*, vol. 11, no. 3, 2018, Art. no. 620.
[52] C. Robles Algarín, D. Sevilla Hernández, and D. Restrepo Leal, “A low-cost maximum power point tracking system based on neural network inverse model controller,” *Electronics*, vol. 7, no. 1, 2018, Art. no. 4.
[53] T. P. Vogl, J. Mangis, A. Rigler, W. Zink, and D. Alkon, “Accelerating the convergence of the back-propagation method,” *Biol. Cybern.*, vol. 59, no. 4/5, pp. 257–263, 1988.
[54] R. Battiti, “Accelerated backpropagation learning: Two optimization methods,” *Complex Syst.*, vol. 3, no. 4, pp. 331–342, 1989.
[55] M. D. Zeiler, “Adadelta: An adaptive learning rate method,” 2012, *arXiv:1212.5701*. [Online]. Avaialble: http://arxiv.org/abs/1212.5701
[56] D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” in *Proc. Int. Conf. Learn. Representations*, 2015, pp. 1–15.
[57] T. Tieleman and G. Hinton, “Lecture 6.5-RMSPROP: Divide the gradient by a running average of its recent magnitude,” *COURSERA: Neural Netw. Mach. Learn.*, vol. 4, no. 2, pp. 26–31, 2012.
[58] P. Geladi and B. R. Kowalski, “Partial least-squares regression: A tutorial,” *Analytica Chimica Acta*, vol. 185, pp. 1–17, 1986.
[59] S. De Jong, “Simpls: An alternative approach to partial least squares regression,” *Chemometrics Intell. Lab. Syst.*, vol. 18, no. 3, pp. 251–263, 1993.
[60] A. J. Burnham, R. Viveros, and J. F. MacGregor, “Frameworks for latent variable multivariate regression,” *J. Chemometrics*, vol. 10, no. 1, pp. 31–45, 1996.
[61] E. Elias and K. Pattipati, “A statistical foundation for the use of the conjugate gradient method in deconvolution,” in *Proc. 35th IEEE Conf. Decis. Control*, 1996, vol. 3, pp. 3131–3136.
[62] B. Li, J. Morris, and E. B. Martin, “Model selection for partial least squares regression,” *Chemometrics Intell. Lab. Syst.*, vol. 64, no. 1, pp. 79–89, 2002.
[63] E. S. Page, “Continuous inspection schemes,” *Biometrika*, vol. 41, no. 1/2, pp. 100–115, 1954.
[64] J. M. Lucas, “Combined Shewhart-CUSUM quality control schemes,” *J. Qual. Technol.*, vol. 14, no. 2, pp. 51–59, 1982.
[65] P. Castagliola, A. Artiba, P. Castagliola, G. Celano, and S. Fichera, “A new CUSUM-S2 control chart for monitoring the process variance,” *J. Qual. Maintenance Eng.*, vol. 15, no. 4, pp. 344–357, 2009.
[66] T. Chang and F. Gan, “A cumulative sum control chart for monitoring process variance,” *J. Qual. Technol.*, vol. 27, no. 2, pp. 109–119, 1995.
[67] G. A. Barnard, “Control charts and stochastic processes,” *J. Roy. Stat. Soc., Ser. B. (Methodological)*, vol. 21, no. 2, pp. 239–257, 1959.
[68] M. R. Reynolds Jr. and Z. G. Stoumbos, “Robust CUSUM charts for monitoring the process mean and variance,” *Qual. Rel. Eng. Int.*, vol. 26, no. 5, pp. 453–473, 2010.
[69] M. Mudholkar, S. Ahmed, M. N. Ericson, S. S. Frank, C. L. Britton, and H. A. Mantooth, “Datasheet driven silicon carbide power MOSFET model,” *IEEE Trans. Power Electron.*, vol. 29, no. 5, pp. 2220–2228, May 2014.
[70] D. Chicco and G. Jurman, “The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation,” *BMC Genomic.*, vol. 21, no. 1, 2020, Art. no. 6.
[71] M. L. McHugh, “Interrater reliability: The kappa statistic,” *Biochemia Medica: Biochemia Medica*, vol. 22, no. 3, pp. 276–282, 2012.
[72] Q. McNemar, “Note on the sampling error of the difference between correlated proportions or percentages,” *Psychometrika*, vol. 12, no. 2, pp. 153–157, 1947.

Qian Yang (Student Member, IEEE) received the B.S. degree in electrical and computer engineering from the Xi’an Jiaotong university, Xi’an, China, in 2016, and the M.S. degree from the Hong Kong University of Science and Technology, Hong Kong, in 2017. She is currently working toward the Ph.D. degree with Electrical and Computer Engineering Department, University of Connecticut, Storrs, CT, USA, under the advisement of Dr. K. Pattipati.

Her research interests include machine learning, anomaly detection, and diagnostics and prognostics.

Muhammed Ali Gultekin (Student Member, IEEE) received the B.E. degree in control and automation engineering, and the M.Sc. degree in electrical engineering from Istanbul Technical University, Istanbul, Turkey, in 2016 and 2020, respectively. He is currently working toward the Ph.D. degree in electrical engineering with the University of Connecticut, Storrs, CT, USA.

His research interests include modeling and control of electric machines and drives, condition monitoring of power electronic systems.

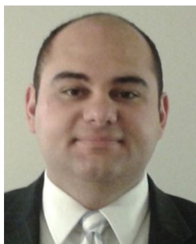
Vahe Seferian (Student Member, IEEE) received the B.E. and M.E. degrees in electrical and computer engineering from the American University of Beirut, Beirut, Lebanon, in 2013 and 2016, respectively.

During the graduate studies, he focused on cryptography and the security of smart grid networks. He is currently pursuing a career in embedded system design, where he is working on power and inverter applications.

Krishna Pattipati (Life Fellow, IEEE) received the B.Tech. degree in electrical engineering with highest honors from the Indian Institute of Technology, Kharagpur, India, in 1975, and the M.S. and Ph.D. degrees in systems engineering from the University of Connecticut (UCONN), Storrs, CT, USA, in 1977 and 1980, respectively.

He was with ALPHATECH, Inc., Burlington, MA, USA, from 1980 to 1986. Since 1986, he has been with the Department of Electrical and Computer Engineering, UCONN, where is currently the Board of Trustees Distinguished Professor and the UTC Chair Professor of systems engineering. He is a cofounder of Qualtech Systems, Inc., Rocky Hill, CT, USA, a firm specializing in advanced integrated diagnostics software tools (TEAMS, TEAMS-RT, TEAMS-RDS, TEAMATE, PackNGo), and serves on the board of Aptima, Inc, Woburn, MA, USA. He has authored or coauthored more than 500 scholarly journal and conference papers in these areas. His research interests include the application of systems theory, optimization and inference techniques to agile planning, anomaly detection, diagnostics and prognostics.

Dr. Pattipati was the co-recipient of the Andrew P. Sage Award for the Best SMC Transactions Paper for 1999, the Barry Carlton Award for the Best AES Transactions Paper for 2000, the 2002 and 2008 NASA Space Act Awards for "A Comprehensive Toolset for Model-Based Health Monitoring and Diagnosis," and "Real-time Update of Fault-Test Dependencies of Dynamic Systems: A Comprehensive Toolset for Model-Based Health Monitoring and Diagnostics", the 2005 School of Engineering Outstanding Teaching Award, and the 2003 AAUP Research Excellence Award at UCONN. He has served as the Editor-in-Chief for the IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS–PART B, from 1998 to 2001. He was selected by the IEEE Systems, Man, and Cybernetics Society as the Outstanding Young Engineer of 1984 and received the Centennial Key to the Future award.



Ali M. Bazzi (Senior Member, IEEE) received the B.E. and M.E. degrees in electrical engineering from the American University of Beirut, Beirut, Lebanon, in 2006 and 2007, respectively, and the Ph.D. degree from the University of Illinois at Urbana-Champaign (UIUC), Urbana, IL, USA, in 2010.

He is an Associate Professor of electrical and computer engineering with the University of Connecticut (UCONN), Storrs, CT, USA. In 2012, he joined as an Assistant Professor with UCONN, where he established and currently directs the Power Electronics and Drives Advanced Research Laboratory (PEARL). He was also an Associate Professor with the American University of Beirut, from 2018 to 2020, a Senior Power Electronics Electrical Engineer with Delphi Electronics and Safety, Kokomo, IN, USA, from 2011 to 2012, a Visiting Assistant Professor with UIUC, during spring 2011, an Engineer with Bitrode Corporation, St. Louis, MO, USA, in the summers of 2008 and 2009, and a Research and Teaching Assistant with UIUC, between 2007 and 2010. He has over 90 peer-reviewed and refereed technical publications and several patents. His research interests include power electronics design, control, optimization, fault diagnosis, and reliability modeling in motor drives, solar photovoltaics, and other applications. He is also interested in renewable energy integration in microgrids, and real-time control and optimization of energy systems in general.

Dr. Bazzi was the recipient of the NSF CAREER Award, in 2018, Mavis Memorial Scholarship at UIUC, in 2009, the Outstanding Teaching Award from the ECE Department at UCONN, in 2014, and the Research Excellence Award at UCONN, in 2015 and 2018. He was the recipient of the UTC Professorship for Engineering Innovation at UCONN, 2016. In 2021, he was the General Chair of the IEEE International Electric Machines and Drives Conference. He is a member of the IEEE Industry Applications Society, IEEE Power Electronics Society, IEEE Power and Energy Society, and IEEE Industrial Electronics Society. He has served on the organizing and technical committees of many IEEE conferences.

Francesco A. N. Palmieri (Member, IEEE) received the laurea degree in ingegneria elettronica cum laude from the Università degli Studi di Napoli Federico II, Naples, Italy, in October 1980, the M.S. degree in applied sciences and the Ph.D. degree in electrical engineering in from the University of Delaware, Newark, DE, USA, in 1985 and 1987, respectively.

From October 1980 to March 1982, he served as a second Lieutenant in the Italian Army in fulfillment of draft duties. In 1982 and 1983, he worked with the ITT firms: FACE SUD Selettronica, Salerno (currently Alcatel), Italy, and Bell Telephone Manufacturing Company, Antwerpen, Belgium, as a Designer of digital telephone systems. He was an Assistant Professor of electrical and systems engineering in 1987 with the University of Connecticut, Storrs, CT, USA, where he was awarded tenure and promotion to Associate Professor, in 1993. In the same year, after a national competition, he was appointed as a Professore Associato with the Dipartimento di Ingegneria Elettronica e delle Telecomunicazioni, Università degli Studi di Napoli Federico II, where he had been until October 2000. Subsequent to his participation in a national competition, he became Professore Ordinario di Telecomunicazioni with the Dipartimento di Ingegneria dell'Informazione, now Dipartimento di Ingegneria, Seconda Università di Napoli (SUN), Aversa, Italy, now Università della Campania Luigi Vanvitelli. His research interests include signal processing, communications, information theory, neural networks, and learning systems.

Dr. Palmieri was the recipient of a Fulbright scholarship, in 1983.

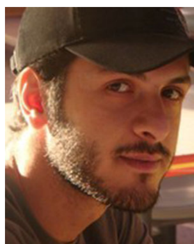
Ravi Rajamani (Member, IEEE) received the B.Tech. degree from the Indian Institute of Technology Delhi, New Delhi, India, the M.Sc. degree from the Indian Institute of Science, Bangalore, India, the MBA degree from the University of Connecticut, Storrs, CT, USA, and the Ph.D. degree from the University of Minnesota, Minneapolis, MN, USA.

He has held visiting research positions with the University of Connecticut and Cranfield University, Bedford, U.K.. He is an independent consultant working on applying model-based and data analytical techniques to complex engineering systems. He has worked with Meggitt, CA, USA, United Technologies Corporation (UTC), CT, USA, and GE, New York, NY, USA. He has authored or coauthored six books including *Electric Flight Technology: The Unfolding of a New Future* and *Unsettled Issues Regarding the Certification of Electric Aircraft*, and many book chapters, journal papers, conference proceedings, and patents.

Dr. Rajamani is a Fellow of SAE and of IMechE. He currently serves as the Editor in Chief for the *SAE International Journal of Aerospace* and is a part of the editorial board of two other journals. He is active within various SAE technical committees and serves as the Chair of the Integrated Vehicle Health Management Steering Group.

Shailesh N. Joshi (Member, IEEE) received the M.S. degree from the Rochester Institute of Technology, Rochester, NY, USA, in 2001, and the Ph.D. degree from Iowa State University, Ames, IA, USA, in 2005, both in mechanical engineering.

He is a Principal Scientist with Electronics Research Department, Toyota Research Institute of North America (TRINA), Ann Arbor, MI, USA. In 2005, He joined Hewlett-Packard in Houston as a Thermal Architect, where he developed cooling solutions for servers and datacenters. He has more than 90 issued patents and has authored or co-authored more than 50 articles in archival journals and conference proceedings on topics related to cooling of electronics and high-temperature bonding technologies. His research interests include developing novel power module architectures using WBG devices for vehicle power electronics along with investigating high-heat flux cooling solutions and advanced bonding technologies.



Muhamed Farooq (Member, IEEE) received the B.S. degree in information technology from Ishik University, Erbil, Iraq, in 2014, and the M.S. degree in computer and information science from the University of Michigan-Dearborn, Dearborn, MI, USA, in 2018.

He currently works as a Research Engineer with Hyundai Mobis, Plymouth, MI, USA. His research interests include developing driver monitoring as well as brain–machine interface technologies.

Hiroshi Ukegawa is currently a Senior Manager of the Power Electronics Research Group with the Electronics Research Department, Toyota Research Institute of North America (TRINA), Ann Arbor, MI, USA. His group investigates next generation power electronics technologies, including high performance cooling, compact packaging, high temperature tolerant bonding for semiconductor devices, fault diagnostics/prognostics, electric-machine integration, and wireless power transfer for future electrified vehicles. He is responsible for the direction of new power electronics research in the US as a part of Toyota's future mobility research and development effort. Prior to TRINA, he worked with Toyota Motor Corporation on the development of inverters and converters in numerous hybrid, plug-in hybrid, fuel cell, and electric vehicles. He developed the Intelligent Power Module for the 3rd and 4th generation. Prius, the 2nd generation MIRAI, and next generation electric vehicles. He also developed the motor/generator electronic control unit for next generation hybrid and plug-in hybrid vehicles.