Research Article

Diagnosis of Intermittent Faults in IGBTs Using the Latent Nestling Method with Hybrid Coloured Petri Nets

Leonardo Rodriguez-Urrego, 1 Emilio García, 2 Eduardo Quiles, 2 Antonio Correcher, 2 Francisco Morant, 2 and Ricardo Pizá 2

1 Facultad de Ingeniería, Universidad EAN, Bogotá, Colombia
2 Departamento de Ingeniería de Sistemas y Automática, Instituto Universitario de Automática e Informática Industrial, Universidad Politécnica de Valencia, Camino de Vera, s/n, 46022 Valencia, Spain

Correspondence should be addressed to Eduardo Quiles; equiles@isa.upv.es

Received 14 October 2014; Revised 26 December 2014; Accepted 6 January 2015

Academic Editor: GuanJun Liu

Copyright © Leonardo Rodriguez-Urrego et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This paper presents a fault diagnosis application of the Latent Nestling Method to IGBTs. The paper extends the Latent Nestling Method based in Coloured Petri Nets (CPNs) to hybrid systems in such a manner that IGBTs performance can be modeled. CPNs allow for an enhanced capability for synthesis and modeling in contrast to the classical phenomena of combinational state explosion when Finite State Machine methods are applied. We present an IGBT model with different fault modes including those of intermittent nature that can be used advantageously as predictive symptoms within a predictive maintenance strategy. Ageing stress tests have been experimentally applied to the IGBTs modules and intermittent faults are diagnosed as precursors of permanent failures. In addition, ageing is validated with morphological analysis (Scanning Electron Microscopy) and semiqualitative analysis (Energy Dispersive Spectrometry).

1. Introduction

Nowadays digital electronic applications [1], power electronics [2], and even PCB [3] introduce IF diagnosis techniques for the analysis of faults by corrosion, contamination, overtemperature, overloads, electrochemical migration, and defects in manufacturing. IF diagnosis allows the utilization of preventive maintenance routines instead of corrective maintenance, so system reliability is increased.

PSD and particularly IGBT are fundamental in many industrial systems. Some of the most important IGBT applications include lighting controls, power supplies, computer systems, industrial control devices, voltage converters [4], motors, or electric generators [5–7]. Recent studies about IGBT diagnosis focus on optimizing their properties as power inverter [8], as a switch [9], aging [10], thermal fatigue [11], or manufacturing defects [12].

IF diagnosis in PSDs can be applied to predict the onset of permanent failures. Moreover it can be used to detect persistent IF episodes that degrade the operation of the system and can be considered as a failure. IF diagnosis applied to PSDs under stress tests predict the wearing-out of the component and can be contrasted with the aging related damage or morphological changes on the physical structure of the component allowing for the validation of the proposed diagnosis model. Then IF diagnosis allows for the estimation of the wear out phase in the hazard rate curve of electronic devices and can be applied in preventive maintenance procedures [13].

Low power equipment is subjected to lower levels of energy during operation and gets discontinued before reaching the wearing-out stage. On the other hand, high power electronics equipment (IGBTs) is subjected to higher energy levels and faces wearing-out due to aging. The main objective of the paper is to show the relevance of the LNM to detect IF in IGBTs.

Different methods have been proposed to diagnose semiconductor faults [14]. In [15] a study to characterize the IGBT
behavior under stress conditions using a SPICE model was introduced. The authors develop an IGBT test circuit and they tested it in two conditions: normal operation and under stress. It is important to note that this diagnosis does not allow predictive maintenance tasks.

In [16] it is discussed as a new method for IGBT fault detection based on gate voltage monitoring. This study takes into account only the degradation due to overcurrent or overtemperature. This analysis is very interesting and is taken into account for our prototype test development.

Another interesting work [17] shows different methods for the aging analysis, such as thermal cycling (TC), hot carrier injection on electrical stress, and dielectric breakdown of time-dependent stimulus. Two of these techniques are applied in our work as accelerated test methods.

The LNM was introduced by Garc´ıa et al. (2008) for the fault diagnosis in complex, large scale systems. LNM relies on CPNs as design platform and a method for nesting faulty marks in every place of the net. The formalization and methodology as well as some examples of the LNM can be seen in [18–21].

The LNM was developed to handle complex discrete event systems, but many systems can be better modeled with hybrid models. This paper will extend the LNM to hybrid systems so it could be applied to diagnose them.

Numerous studies have been carried out to explain hybrid process fault diagnosis using different methodologies [22–24]. New techniques need to be developed for diagnosis of IFs, like the residual analysis proposed in our method.

Furthermore, some researchers [25] analyzed fault models in hybrid PNs. Other authors propose an approximation of differential places to represent continuous places with negative markings (differential PNs [26]) in each place of latent nesting faults (PLN) in order to avoid unobservable transitions and allow faulty tokens of discrete type to be nested in places of continuous nature. The above provide advantages in solving hybrid systems of increasing complexity and finding failure times of each faulty token in the PV

IF diagnosis is carried out based on the work by [27] where the authors present a prognosis method to diagnose IF and predict the lifetime of electromechanical devices.

The paper is structured as follows. Section 2 introduces LNM for hybrid systems. It also includes a simple example to show its performance. Section 3 shows the IF diagnosis modeling based on LNM applied to IGBTs. Section 4 explains the test bench, the analysis, and experimental results. Finally, Section 5 draws some relevant conclusions.

### 2. Latent Nestling Method in Hybrid Systems

#### 2.1. LNM Definition in Hybrid Systems

LNM is a methodology for fault diagnosis of discrete event complex systems (see [18, 20, 21]). Because this paper introduces a hybrid model for the IGBTs (presented in Section 3.2) we present an update of LNM to handle hybrid systems.

The diagnoser will be a hybrid model of the system including normal and faulty behavior of each device in the system. In order to avoid the combinational explosion, [19] the model is built using hybrid colored PNs.

A hybrid CPN for fault diagnosis (HCPNFD) is defined as

\[
\text{HCPNFD} = \left( P, T, \text{Pre}, \text{Post}, M_0, C, C_0, \text{PLN}_f, \text{TF}, \text{PV}_f, S, \text{Tempo} \right),
\]

where \( P \) is a finite set of places, \( T \) is a finite set of transitions and \( \text{Pre} \) and \( \text{Post} \) are the input and output arc functions, with an additional argument \( C_k \) which is the color of the transition firing \( T_f \). Thus \( \text{Pre}(P_i, T_f/C_k) \) and \( \text{Post}(P_i, T_f/C_k) \) correspond in the general case to a linear combination of token colours related to place \( P_i \).

These functions can be divided into two subsets, depending on the transition-type behavior, namely, normal transition \( T \) or faulty transition \( TF \)

\[
\text{TF} = T_f \cup T_r,
\]

where \( T_f \) and \( T_r \) are the fault and recovery transitions, respectively. \( M_0 \) is the initial marking. \( \text{PLN}_f \) is the subset of fault latent nesting places, where \( \text{PLN}_f \subseteq P \). If \( M_0 \) includes a faulty token in \( P_i \) then \( P_i \) is now called \( \text{PLN}_{f_i} \). \( \text{PV}_f \) is the subset of fault verification places.

The places set and transitions set can be divided into two subsets

\[
P = P^D \cup P^C, \quad T = T^D \cup T^C.
\]

\( P^D \) is the set of discrete places and \( P^C \) is the set of continuous places. \( T^D \) is the discrete transitions set and \( T^C \) is the continuous transitions set.

\( P^D \) will represent discrete states of a device such that the device is on and off and is starting and stopping, and so forth. \( P^C \) will represent the continuous states of a device so it computes a differential equation model.

\( T^D \) will represent a discrete state change.

\( T^C \) will represent step execution of the model contained in a \( P^C \).

In addition, the normal behavior marks can have discrete or continuous nature:

\[
N = N^D \cup N^C.
\]

\( N \) will represent a normal behavior token of a device and its evolution through the diagnoser will show the device state.

\( C \) is the colour set assigned to different identifiers. \( C = N^D \cup N^C \cup f, \) where \( f = \{f_1, f_2, \ldots, f_n\} \) is the subset of coloured tokens representing the fault set.

Initial marking for a place \( \text{PLN}_{f_i} \) in \( P^D \) (called \( \text{PLN}^{D}_{f_i} \)) will be \( (N^D, f) \), and the initial marking for a place \( \text{PLN}_{f_i} \) in \( P^C \) (called \( \text{PLN}^{C}_{f_i} \)) will be \( (N^C, f) \). \( \text{Pre}^T : (P \times T) \rightarrow Q_+ \) or \( N, \text{Post}^T : (P \times T) \rightarrow Q_+ \) or \( N, Q_+ \) stands for the rational numbers (positives or zero). Then, for \( \text{PLN}^{C}_{f_i} \),

\[
\text{Pre}^T : \left( \sum_{i=1}^{k} \text{PLN}^{C}_{f_i} \times T_f \rightarrow f \cup \text{PV}_f \times T_r \rightarrow f \right).
\]
Let $\text{Pre}^{\text{TF}}$ be the input arc function corresponding to subset TF. Consider

$$\text{Post}^{\text{TF}}: \left( PV_f \times T_f \rightarrow f \cup \sum_{i=n}^{k} \text{PLN}^{\text{C}}_{f_i} \times T_r \rightarrow f \right).$$  \hspace{1cm} (6)

Let $\text{Post}^{\text{TF}}$ be the output arc function corresponding to subset TF. In $\text{Pre}^{\text{TF}}$ and $\text{Post}^{\text{TF}}$ case, the number of arc functions corresponding to TF subset of each PLN depends on the continuous places mutually influenced, such that $n$ is the initial continuous place influenced and $k$ is the last continuous place influenced.

This PLN$_{f_j}$ represents a continuously variable behavior and also allows the nesting of discrete type faults.

$\text{CO}_j: P \cup T \rightarrow \{ D, C \}$ is a composite function that is defined for every place of the net.

$S = (S_1, S_2, \ldots, S_h)$; it is the hybrid states set in the analyzed system. This set is composed of the operating states OS, fault signatures $S_f$, and recovery signatures $Sr$.

Tempo: it is a delay function that associates a rational number to each timed transition, where if for a function $f(T_j) = D$, tempo$(T_j) = \{1, d_j\}$ is a delay associated with the transition $T_j$, expressed in time units, if for a function $f(T_j) = C$, tempo$(T_j) = \{V(T_j), d_j\} = \{V_j, h\}$, such that $V_j$ represents the maximum firing speed associated with the transition $T_j$ and $h$ is the firing frequency that represents the sampling time. The method for delay fixing $(d_j)$ or fixing the frequency firing $(h)$ depends on the system behaviour.

**Definition 1.** A normal discrete transition in a HCPNFD is enabled at a marking $M$ if each place $P_i \in P^D$ in $0^{\text{TF}}_j$ meets the condition:

$$M(P_i) \geq \text{Pre}(P_i, T^D_f).$$  \hspace{1cm} (7)

**Definition 2.** A normal continuous transition in a HCPNFD is enabled at a marking $M$ if each place $P_i$ in $0^{\text{TF}}_j$ meets the condition:

$$M(P_i) \geq \text{Pre}(P_i, T^C_f), \quad \text{if } P_i \in P^D,$$

$$M(P_i) \in Q_+, \quad \text{if } P_i \in P^C,$$

where $0^{\text{TF}}_j$ is the set of the input places of discrete $T_j$ and $0^{\text{TF}}_j$ is the set of the input places of continuous $T_j$. Likewise, it is necessary to meet the condition $\forall t \in T^C$ and $\forall p \in P^D$, $\text{Pre}^T(p, t) = \text{Post}^T(p, t)$. $P_i, T_j/C_k$.

2.2. Initial Model and Fault Selection. The initial model is similar to that presented in the LNM. However, it includes continuous places where we could model the continuous behavior of the system variables. This step applies the techniques of modeling hybrid PNs [25].

According to [19], the sensors map is defined as sm: $M_0 \rightarrow SR$, where SR is the sensor readings, such that for $M_k$ marking the expression is given by $\text{SR}(M_k) = sr_1(M_k), sr_2(M_k), \ldots, sr_n(M_k)$. In the discrete case $\text{SROV}(M_k)$ is the set of sensor read output values for each discrete marking, such that

$$\text{SROV}(M_k) = \text{SROV}_{ev}(M_k) + \text{SROV}_{uev}(M_k),$$  \hspace{1cm} (9)

where $\text{SROV}_{ev}$, $\text{SROV}_{uev}$ are subsets of expected and unexpected values accordingly.

2.3. Latent Nestling Places and Trajectories of Fault Verification and Fault Recovery. Latent nestling places are defined by the LNM. However, in a hybrid system, there is a continuous place $P^C_f$ which represents an operating state during a certain time $t$ according to the states of the discrete places. Faults are assigned to this continuous place, such that $\text{PLN}^{\text{C}}_{p_f} \in P^C$. This implies that the faults have been generated by the anomalous behavior of the continuous variable, where the faults are nestling in the same continuous place now called $\text{PLN}^{\text{C}}_{f_j}$ owing to this hybrid character.

The trajectories of the faulty tokens are defined only by the fault and recovery transitions. The normal discrete and continuous transitions are defined by a classical method for modeling Hybrid PNs [25], as well as the firing rules for these transitions. Furthermore, fault and recovery transitions must be added to make restrictions that allow including both the place status as tokens of normal behavior.

**Definition 3.** A fault or recovery transition in a CPNFD or HCPNFD is enabled at a marking $M$ for discrete places if each place $\text{PLN}^{\text{D}}_{f_j}$ or $\text{PV}_j$ in $0^{\text{TF}}_j$ meets the condition: for fault transitions $T_f$

$$M \left( \text{PLN}^{\text{D}}_{f_j} \right) \geq \text{Pre} \left( \text{PLN}^{\text{D}}_{f_j}, T_f \right),$$  \hspace{1cm} (10)

for recovery transitions $T_r$

$$M \left( \text{PV}_j \right) \geq \text{Pre} \left( \text{PV}_j, T_r \right) \land M \left( \text{PLN}^{\text{D}}_{f_j} \right)$$

$$\geq \text{Pre} \left( \text{PLN}^{\text{D}}_{f_j}, T_r \right).$$  \hspace{1cm} (11)

Let $M'_f$ be the fault marking obtained after firing of transition $T_f$ with respect to the fault signature $Sr_f$. This fault marking is deducted from the marking $M_f$ by the following relation.

For fault trajectory,

$$M'_f \left( \text{PLN}^{\text{D}}_{f_j} \right) = M_f \left( \text{PLN}^{\text{D}}_{f_j} \right) + \text{Post} \left( \text{PV}_j, \frac{T_f}{\text{SROV}_{uev}} \right)$$

$$- \text{Pre} \left( \text{PLN}^{\text{D}}_{f_j}, \frac{T_f}{\text{SROV}_{uev}} \right), \forall \text{PLN}^{\text{D}}_{f_j}.$$

$$M'_f \left( \text{PLN}^{\text{C}}_{f_j} \right) = M_f \left( \text{PLN}^{\text{C}}_{f_j} \right) + \text{Post} \left( \text{PV}_j, \frac{T_f}{S_f} \right)$$

$$- \sum_{i=n}^{k} \text{Pre} \left( \text{PLN}^{\text{C}}_{f_j}, \frac{T_f}{S_f} \right), \forall \text{PLN}^{\text{C}}_{f_j}.$$


For recovery trajectory,
\[
M'_F(\text{PV}_f) = M_F(\text{PV}_f) + \sum_{i=1}^{m} \left( \text{Post}(\text{PLN}^{D}_{f_i}, \frac{\text{Tr}_j}{\text{SROV}_{ev}}) - \text{Pre}(\text{PV}_f, \frac{\text{Tr}_j}{\text{SROV}_{ev}}) \right) + \sum_{j=1}^{n} \left( \text{Post}(\text{PLN}^C_{f_j}, \frac{\text{Tr}_j}{\text{Sr}_k}) - \text{Pre}(\text{PV}_f, \frac{\text{Tr}_j}{\text{Sr}_k}) \right), \forall \text{PV}_f. \tag{13}
\]

\( m \) is the last PLN\(_{f_i} \), \( n \) is the initial continuous place influenced, and \( k \) is the last continuous place influenced.

In the example case of Figure 3 we have for fault verification
\[
M'_F(\text{PLN}_{f_i}) = M_F(\text{PLN}_{f_i}) + \text{Post}(\text{PV}_f, \frac{\text{Tr}_1}{\text{SROV}_{uev}}), \quad - \text{Pre}(\text{PLN}_{f_i}, \frac{\text{Tr}_1}{\text{SROV}_{uev}}), \quad M'_F(\text{PLN}^C_{f_j}) = M_F(\text{PLN}^C_{f_j}) + \text{Post}(\text{PV}_f, \frac{\text{Tr}_2}{\text{Sr}_k}) - \text{Pre}(\text{PLN}^C_{f_j}, \frac{\text{Tr}_2}{\text{Sr}_k}). \tag{14}
\]

And for fault recovery,
\[
M'_F(\text{PV}_f) = M_F(\text{PV}_f) + \text{Post}(\text{PLN}_{f_i}, \frac{\text{Tr}_1}{\text{SROV}_{ev}}), \quad - \text{Pre}(\text{PV}_f, \frac{\text{Tr}_1}{\text{SROV}_{ev}}) + \text{Post}(\text{PLN}^C_{f_j}, \frac{\text{Tr}_2}{\text{Sr}_k}) - \text{Pre}(\text{PV}_f, \frac{\text{Tr}_2}{\text{Sr}_k}). \tag{15}
\]

To find the residues, it is necessary to obtain the operation dynamic model of the continuous variables. Depending on the complexity, the models could be represented in state variables, as in the hybrid PN analysis [26]. In this case, the approach presented in our example introduces a series of residues of the form \( r(t) = y(t) - \hat{y}(t) \) in every continuous place. The residue is computed in the continuous place, while the residual evaluation is checked in each fault and recovery transition.

The definitions on states of hybrid operation, fault signatures, and diagnosability can be seen in [21].

3. IFs Diagnosis Using the LNM Based on HCPNFD

3.1. Temporal Modeling of IFs. The main purpose of diagnosing IFs is the generation of tools to perform preventive maintenance of devices in industrial systems. It becomes necessary to apply data obtained online to determine the best time to replace or repair a component. The basic idea is to employ prediction methods based on process fault information. This information is indicative of the deterioration that is suffering the component.

From this method we get two measures based on [27]: temporal failure density and pseudo period. Temporal failure density (DF or density in the rest of the paper) is defined as the average time a particular fault is active within a sliding time window of duration \( w \). DF computed at time \( C_T \) for failure \( F_j \) (DF\(_{C_T,F_j} \)) is defined as
\[
\text{DF}_{C_T,F_j} = \frac{\sum_{k=1}^{\text{CNT}} (T_{(k)F_j} + T_A)}{w}, \tag{16}
\]
where CNT is the number of faults inside the window, \( l \) stands for the index of the first fault detected inside the window \( \{l: FT_{(l)F_j} > (C_T - w) \} \) if it exists; otherwise \( l = \text{CNT} + 1 \) and \( T_A \) takes into account the duration of a failure occurred before \( C_T - w \) which continues active inside the window. Therefore,
\[
T_A = FT_{(l-1)F_j} + FT_{(l-1)F_j} - (C_T - w). \tag{17}
\]
Equation (17) is valid only if \( T_A \) is positive; otherwise \( T_A = 0 \), as this fact would indicate that the \( (l-1) \)th failure time is completely outside of the window. In a real system, DF tends to increase with time, thus confirming the hypothesis that IFs progressively damage the faulty device. In our case we only apply this measure DF with the LNM.

3.2. Initial Hybrid Model. For this case we will focus on a nonlinear model that represents the turn-on and turn-off switching waveforms and will get the \( V_{GE} \) and \( V_{GE} \) value that must have the IGBT. Some references that model different aspects of IGBTs and MOSFETs and the turn-on and turn-off waveforms can be seen in [28]. For each state (turn-on, turn-off) there are equations that define its operation.

For the turn-on these equations are as follows.

The increasing time constant from \( V_{GE} \) to \( V_{th} \) is limited by
\[
R_G: (C_{GE} + C_{CG}). \tag{18}
\]

The decreasing time constant from \( V_{CE} \) to \( V_{CE(sat)} \) is limited by
\[
\left( \frac{V_{GG} - V_{GE(sat)}}{R_G: C_{CG}} \right), \tag{19}
\]
where \( V_{GE(sat)} \) is the \( V_{GE} \) voltage when it reaches the maximum collector current \( I_{C_{sat}} \) and \( V_{GG}^+ \) is the voltage across the gate to the emitter of the transistor during conduction. The increasing time constant from \( V_{GE} \) to \( V_{GG}^+ \) is limited by
\[
R_G: (C_{CE} + C_{th}). \tag{20}
\]
The reverse transfer capacitance $C_{res}$ or $C_{CG,miller}$ is approximately equal to $C_{CG}$ because the emitter is connected to ground. Then we will use $C_{res} \equiv C_{CG}$ in our final model.

Based on the equivalent circuit of the IGBT gate, the gate current $I_G$ is deduced by

$$I_G(t) = C_{GE} \frac{dV_{GE}}{dt} - C_{CG} \frac{d[V_{CE} - V_{GE}]}{dt}. \tag{21}$$

Note that $I_G$ is directly affected by $C_{CG}$ which causes a large change in gate voltage.

For the turn-off the equations are as follows.

$V_{GE}$ falls from injected $V_{GG}$, to $V_{GEl}$, with a time constant given by (21). At this time, there is no change in the values of $V_{CE}$ or $I_C$.

Then $V_{CE}$ increases in this region and the rate can be controlled with $R_G$ as shown in the equation below:

$$\frac{dV_{CE}}{dt} = \frac{V_{GEl}}{C_{res} \cdot R_G}. \tag{22}$$

Then the value of $V_{CE}$ is maintained at $V_{th}$ while $I_C$ decreases at a rate defined by the following equation. The rate of increase can also be controlled with $R_G$

$$\frac{dI_C}{dt} = \frac{g_{fe} \cdot V_{GEl}}{C_{res} \cdot R_G}, \tag{23}$$

where $C_{res}$ is the input capacitance measured between the gate and emitter terminals with the collector shorted to the emitter for AC signals, $C_{res} = C_{GE} + C_{CG}$. The value of these fixed capacitances can be found in the data sheet of the manufacturer.

3.2.1. Hybrid Model Using Hybrid PNs. The hybrid model is implemented following the scheme of Figure 1. Continuous places $PLN_{f_1}^C$ and $PLN_{f_2}^C$ represent the ideal behavior of voltages $V_{CE}$ and $V_{GE}$, respectively. The continuous place $PLN_{f_3}^C$ represents the load voltage as a function of the collector current. Transition $T_4$ represents the activation of the IGBT (turn-on) and transition $T_5$ shows the switch off the IGBT (turn-off).

Anytime during IGBT switching this model represents voltages $V_{GE}$ and $V_{CE}$. This allows us to detect any small changes in these voltages during the stress tests. Depending on the experimental condition a complete cycle lasts from 20 to 100 ms as shown in the Results section.

As there are two continuous places, the model has two different residues that verify the same fault. It is important to nest in every place the same fault but with a different designation. Therefore we nested faults $f_1$ as $f_1^G$ if the fault is from the $PLN_{f_1}^C$ and $f_1^C$ if the fault comes from the $PLN_{f_1}^C$, likewise for fault $f_2$.

We consider two types of IGBT faults. The first fault is the device in open circuit. When there is a difference between $V_{CE}$ and $V_{GE}$, such that $V_{GE}$ remains in a positive value, it is considered that the system is in a fault mode because the IGBT does not respond to the control signal for some reason. This fault mode can be caused by two conditions: command level design or an internal failure of the component (intermittent fault). This fault is called $f_1$. When there is a difference between signals $V_{CE}$ and $V_{GE}$, such that $V_{GE}$ remains in a negative value, it is considered that the system is in a fault mode because the IGBT does not respond to the control signal for some reason. This fault mode can be caused by the same two previously defined conditions. This fault is called $f_2$.

The residues were analyzed using a nonlinear model based on HCPNFD.

Observing the model in Figure 2, the residues may be obtained using the sensor readings with the values of continuous places. In this model $g_{fe}$ is the forward transconductance.

Therefore, if the fault comes from the place $PLN_{f_1}^C$, the faulty mark nested is $f_1$ as $f_1^G$ and if the fault comes from the place $PLN_{f_1}^C$, the faulty mark nested is $f_1^C$, similarly to the faulty mark $f_2$. The proposed HCPNFD model has been verified by a reasonably good agreement with measurements. Figure 2 shows the resulting waveforms of the turn-on and turn-off. In this case the turn-on starts with $V_{CE}$ high, $V_{GE}$ zero or negative and constant gate charging current producing a linear increase of the gate voltage. With falling collector-emitter voltage $V_{CE}$ the gate bias current is utilized for changing the charge of $C_{CG}$ ($C_{CG} \times dV_{CE}/dt$) and the gate voltage remains constant. When the collector-emitter voltage has come down $C_{CG}$ becomes larger as much that also at reduced slope of $V_{CE}$ still all the bias supplied gate current is used up. Only when finally the current needed for charging becomes smaller than the bias supplied current the gate voltage rises again. The turn-off starts with $V_{CE}$ low, $V_{GE}$ positive or greater than the threshold voltage $V_{th}$. The gate voltage first decreases nearly linearly. With still low collector-emitter voltage $V_{CE}$ and with only moderate increase there is the strongest change (decrease) of $C_{CG}$. Decrease of a capacitance at constant charge increases the voltage. As there is a bias source which is drawing current out of the gate, the gate-emitter voltage remains constant. Subsequently $V_{CE}$ increases and most of the gate discharge current is used up for $C_{CG}dV_{CE}/dt$. The gate voltage further remains constant. The charge over process is finished when $V_{CE}$ roughly reaches the operating voltage. Now a further decrease of the gate voltage is possible again.

In this case, a residue signal is obtained that would be expressed as

$$r_{x,y} = |\nabla x - \nabla y|, \tag{24}$$

where $V$ is the real reading and $V'$ is the estimated reading. $x$ is the IGBT analyzed, and $y$ is the obtained residue number for this IGBT. In the case of the four IGBTs of our test, we obtain the following two residues for each IGBT:

$$r_{1,1} = |V_{GE1} - V_{GEi}'|, \quad r_{1,2} = |V_{CE1} - V_{CEi}'|. \tag{25}$$

Also, the implemented IF diagnosis needs the computing of DF with (17) and (18). So, some parameters must be computed:

(i) a counter CNT in the PV $f$ place for each type of fault;
(ii) a timer $t_f$ associated with each faulty mark. This timer is reset each time the fault is recovered;
(iii) a timer $t_d$ associated with the PV$_f$ place.

These parameters allow obtaining for each fault the temporal density and analyze the prediction of change for each IGBT. Therefore for each fault

$P$—timed place PV$_f$ is $\text{Tempo}(PV_f) = t_d = w$, where $w$ is the duration of the sliding window.

The counters for each fault are given by

$\text{CNT}_{f_1^C} = \text{number of times the fault type } f_1^C \text{ was isolated in PV}_f \text{ place in a window of duration } w$;

$\text{CNT}_{f_2^C} = \text{number of times the fault type } f_2^C \text{ was isolated in PV}_f \text{ place in a window of duration } w$;

An example of IFs can be seen in Figure 3 with a window of $w = 18$ units. There are two iterations of failure and recovery of type $f_1^n$. From left to right you can see the iteration number, the fired transition, the fault counter for that specific fault, time on the window, the timer of the fault, and finally the vector that stores timer information every time that a fault occurs.

Based on the analysis of continuous places of Section 2.3 and in Figure 2, we observe that $P_1^C$ and $P_5^C$ places are of isolated type; therefore
OS_f = (os_1, os_3), where os_1 = (S_f_1(k), S_f_3(k)), respectively, to P_1, P_3 which are the discrete places that influence the behavior of continuous place P_f^C. Consider the following.

os_3 = (S_f_7(k), S_f_9(k)), respectively, to P_7, P_9 which are the discrete places that influence the behavior of continuous place P_f^C. According to \( s_i \in S \mid S_f_i \in os \mid s_i = S_f_i \) for continuous isolated places, we obtain \( S_i = os_1, S_i = os_3 \) and each state \( S_i \) is equal to each fault signature \( S_f_i \); therefore,

\[
S_f_1(k) = \begin{cases} 
(f_1^C, S_1) & \text{if } r_{1,1}, \\
(f_2^C, S_1) & \text{if } r_{1,1}, 
\end{cases}
\]

\[
S_f_7(k) = \begin{cases} 
(f_1^C, S_7) & \text{if } r_{1,1}, \\
(f_2^C, S_7) & \text{if } r_{1,1}, 
\end{cases}
\]

\[
S_f_3(k) = \begin{cases} 
(f_1^C, S_3) & \text{if } r_{1,1}, \\
(f_2^C, S_3) & \text{if } r_{1,1}, 
\end{cases}
\]

\[
S_f_9(k) = \begin{cases} 
(f_1^C, S_9) & \text{if } r_{1,2}, \\
(f_2^C, S_9) & \text{if } r_{1,2}, 
\end{cases}
\]

Applying (13) to fault trajectory, os_1:

\[
M'_F(PLN_{f_1}^C) = M_F(PLN_{f_1}^C) + Post\left( PV_f, \frac{T_f_1}{S_f_1} \right) - Pre\left( PLN_{f_1}^C, \frac{T_f_1}{S_f_1} \right) + Post\left( PV_f, \frac{T_f_3}{S_f_3} \right) - Pre\left( PLN_{f_1}^C, \frac{T_f_3}{S_f_3} \right)
\]

(27)

Applying (16) to fault recovery,

\[
M'_F(PLN_{f_j}^C) = M_F(PLN_{f_j}^C) + Post\left( PLN_{f_j}^C, \frac{Tr_1}{Sr_1} \right) - Pre\left( PV_f, \frac{Tr_1}{Sr_1} \right) + Post\left( PLN_{f_j}^C, \frac{Tr_2}{Sr_2} \right) - Pre\left( PV_f, \frac{Tr_2}{Sr_2} \right) + Post\left( PLN_{f_j}^C, \frac{Tr_3}{Sr_3} \right) - Pre\left( PV_f, \frac{Tr_3}{Sr_3} \right)
\]

(28)

4. Analysis and Experimental Test Results

4.1. Hardware Implementation. The test bench is based on two test circuits. The first circuit is a direct operational model of activation with a resistive load. The second circuit has a driver that protects and regulates the current in the base of activation with a resistive load. The second circuit has two test circuits. The first circuit is a direct operational model with a range of 5°C to 540°C. Figure 4 shows a complete scheme of the assembly created for the test bench.

The basic driver circuit was based on the circuit presented in [15] that allows stress aging tests using thermal cycling (TC) and hot carrier injection (HCI).

The TC is strongly associated with failure by degradation and removal of welding. The HCI is another form of accelerated aging. This aging mechanism can be performed by...
applying high voltages at the gate of the IGBT or can also be produced by magnetic fields.

This circuit uses a $V_{CC} = 10$ V, $R_C = 10$ Ω, $R_G = 100$ Ω, $I_{CE} = 5$ A. The second circuit works as inverter in industrial installations for motor control or power generating systems. This driver circuit allows precision in the control signal at the IGBT gate. For aging tests using the technique of thermal cycling, it is necessary to limit the current out of the driver; therefore,

$$R_G (\text{min}) = \frac{V_{CC2} - V_{OH} \left( I_{OUT}/650 \ \mu A \right) - (V_{CL} + V_{EE})}{I_{OL/\text{peak}}}. \quad (30)$$

The maximum current driver is $I_{CL/\text{peak}} = 2.5$ A, the maximum switching voltage is $V_{CC2} = 15$ V, and $V_{OH} = 1.2$ V according to the manufacturer's data sheet. Using the low voltage output maximum $V_{CL,\text{MAX}} = 0.5$ V (manufacturer's data sheet), it has a $R_G (\text{min}) = 5.3$ Ω. $R_G$ modifies the voltage slope $V_G$ in the $t_{ON}$ and $t_{OFF}$. If $R_G$ is of greater value the transition in $t_{ON}$ and $t_{OFF}$ is slower. Therefore we have to employ small values for $R_G$. The maximum switching frequency is determined by

$$f_{\text{MAX}} = \frac{1}{t_{ON} + t_{OFF}}, \quad (31)$$

where

$$t_{ON} = t_d + t_r, \quad t_{OFF} = t_s + t_f. \quad (32)$$

Likewise, total dissipated power is given by

$$P_T = P_I + P_O, \quad (33)$$

where $P_I$ is the maximum input power dissipated, limited by $P_I < 150$ mW. $P_O$ is the maximum output power dissipated, limited by $P_O < 600$ mW.

Consider

$$P_I = V_{CC2} \cdot I_{CC1},$$

$$P_O = I_{CC2} \cdot (V_{CC2} - V_{EE}) + E_{SW} \cdot f_{SW}, \quad (34)$$

where $I_{CC1}$, $V_{CC2}$, and $V_{EE}$ are given by the manufacturer of the driver selected as our circuit. $f_{SW}$ is the maximum switching frequency of the driver and $E_{SW}$ is the power dissipated in a resistive load switching defined by

$$E_{SW} = \frac{V_{CC2} \cdot I_{C_{\text{MAX}}}}{6} \cdot f_J. \quad (35)$$

Knowing that $I_{C_{\text{MAX}}} = V_{CC2}/R_{CL,\text{MIN}}$, take a $R_{CL,\text{MIN}} = 5$ Ω for some aging tests as electrical overstress (EO) and TC methods; then $I_{C_{\text{MAX}}} = 3$ A. With these data we obtain $E_{SW} = 0.75 \mu J$. Finally solving (34) and comparing with the maximum input and output values of power dissipated, $P_I = 5 \text{ V} \cdot 16.5 \mu A = 82.5$ mW $< 150$ mW, $P_O = 5.5 \text{ mA} \cdot (15 - 0) + 0.75 \mu J \cdot 5 \text{ kHz} = 86.25$ mW $< 600$ mW.

Figure 5 shows the driver circuit for each IGBT.

In this case the maximum power dissipation is not exceeded according to the more demanding tests performed in our test bench.

Figure 6 shows the task execution blocks, interconnected to the data acquisition card and the test bench. Figure 7 shows the graphical user interface for fault diagnosis in the IGBTs test bench. Each red number designs information or task panel:

1. start/finish test;
2. test mode, switching frequency, and gate voltage;
3. input voltage signal;
4. measured gate voltage;
5. collector current (by shunt effect);
6. switching counter;
7. temperature display;
8. temperature zoom in;
9. collector current standard deviation;
10. type 1 and 2 faults switching counter.

Figure 8 shows in blue IGBT 22 failure in short circuit with a load short circuit fault current of 1.4 A.
Figure 5: IGBT driver circuit for aging and IFs analysis.

Figure 6: Software platform for diagnosis.

Figure 7: Diagnosis software environment for IGBTs in labview.
4.2. Results. In the IGBT fault-free, the first thing we get is the performance curve $I_{CE}$ versus $V_{CE}$ for several new IGBTs. Curve $I-V$ is commonly presented to show the performance of IGBTs (IGBT-IRG4BC30KDPBF).

Figure 9 shows in graph (a) the behavior of the current versus the collector voltage for a 7 V fixed value of the gate voltage. It can be seen in IGBTs 15 to 19 that the inclination angle of the curve $\beta > 45^\circ$ remains almost constant regardless of the initial resistive load (graphical detail (b)). Aging of IGBTs modifies $I_{CE}-V_{CE}$ curves as it can be seen in Figure 10(b).

In addition, it is presented as a morphological analysis and chemical analysis of some selected samples to determine the compounds of the IGBTs.

In this case the tests were performed at 4 IGBTs per sample. Each IGBT is carried out the algorithm for detection of IFs. In total 64 IGBTs are analyzed for different types of stress. Most tests were performed by TC and by load. We selected a test to give the best results in this case with the following characteristics: IGBT surface temperature of $250^\circ$C, switching frequency of $500$ Hz, gate voltage of 7 V, and load voltage of 10 V. Finally a condition monitoring by loading with $R = 5 \Omega$.

Figure 11 shows the open circuit fault $f_1$ better than the other test due to the intensity of the test. The graph (b) shows the initial faults and from hour 16 the intermittent faults due to wearing-out of the IGBT. At approximately 23.5 hours occurs the abrupt fault. The graph (a) shows the intensity
of the switching and that the last fault $f_1$ was detected in switching 198000.

Detail (a) of graph (a) in Figure 12 shows small short circuit IF. From $f_2$, fault number 20 onwards the IGBT only fails in short circuit leading to a short circuit permanent failure. Graph (b) in Figure 12 shows the initial short circuit faults and the $f_2$ wearing-out faults during the last 30 minutes.

Figure 9 shows the performance curve $I_{CE}$ versus $V_{CE}$ in order to see the aging curve. The curves represent the beginning of the normal curve, but the curve at hour 22 shows the wear on the IGBT. The detail (c) of the graph (b) shows that the state of the IGBT tends to open circuit as seen in the IFs shown in Figure 11. The end of the IGBTs life by short circuit at the 43rd hour of operation can be seen in the detail (d) of the graph (a).

Completing this IFs analysis we proceeded with the SEM/EDS analysis to the samples. This analysis corroborates the morphological and physical changes appreciated in the IGBTs structure. SEM analysis in Figure 13 shows an almost union separation and a grain size quite appreciable. Although this qualitative information is not very valuable, the semiquantitative information of EDS clearly shows that the compounds of silicon, copper, and tin are increased in the union. This increase is directly related to the test type, intensity, and hours of operation. While stress by TC and by load increases, the amount of these compounds increases too. These tests were conducted at 700 $\mu$m.

SEM/EDS analysis is also applied to the gate union in the IGBTs. In this case it was applied 600 $\mu$m testing. Figure 14 shows that deformations are very remarkable. Analyzing the results we observed in EDS that silicon and oxygen increased with the aggressiveness of the tests and copper and tin decreased in the same proportion.

5. Conclusions

Typical electronic devices such as IGBTs have a type B failure characteristic with an infant mortality followed by a constant or slowly increasing failure probability; therefore, they have no an identifiable wearing-out age. So an age limit is not applicable normally. The major contribution of our work is the inclusion of intermittent faults in the developed fault diagnosis model. Intermittent faults can be used as precursor symptoms of identifiable wearing-out age permanent failures.
in order to apply preventive or predictive maintenance in the electric and electronic devices. This paper shows the validity of this kind of intermittent fault diagnosis for IGBTs.

We have used models based in the Latent Nestling Method and HCPN. The dynamics of HCPN allow for the representation of transitions between transitory faults and fault-free states including quantitative measures.

Some conclusions can be drawn from the stress tests. The IGBTs condition and fault mode depend on the experimental procedure and stress level applied in the tests. Condition 1 of 10 Ohm/230°C/500 Hz had hardly any effect in the aging process of the components leading to no fault. Condition 2 of 5 Ohm/250°C/500 Hz produced open circuit intermittent faults leading to short circuit permanent failures. Our aging hypothesis has been confirmed by morphological and chemical analysis (SEM/EDS) carried out on the failed IGBTs.

**Nomenclature**

PCB: Printed circuit board
PSD: Power semiconductor device
TC: Thermal cycling
LNM: Latent Nestling Method
PLN\(_f\): Place of latent nesting fault
PV\(_f\): Place of fault verification
\(\mathcal{S}\): Set of hybrid states
\(f\): Set of faults
SR: Set of sensor readings
SROV\(_{ev}\): Subset of SR output of expected values
SROV\(_{uev}\): Subset of SR output of unexpected values
\(M_{f}\): Fault marking
\(S_{f}\): Fault signature
IF\(_{f}\): Intermittent faults
DF: Temporal failure density
HCPNFD: Hybrid coloured Petri net for fault diagnosis
$V_{GE}$: Gate-emitter voltage
$V_{CE}$: Collector emitter voltage
$C_{res}$: Reverse transfer capacitance
$C_{ies}$: Input capacitance between gate and emitter
CNT: Faults counter
t$_f$: Timer associated with the PV$_f$
t$_d$: Timer associated with the PV$_f$
HCI: Hot carrier injection
$I_{CL/peak}$: Maximum current driver
t$_{on}$: Turn-on
t$_{off}$: Turn-off
SEM: Scanning Electron Microscopy
EDS: Energy Dispersive Spectrometry
$f_{SW}$: Maximum switching frequency of driver
$E_{SW}$: Power dissipated in a resistive load switching
w: Sliding time window
CNT$_{t_f}$: Number of times the fault type $f_i$ was isolated in PV$_f$ in a window of duration w.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

This work was supported by the Spanish Ministerio de Ciencia y Tecnología Project DPI2009-14744-C03-03, by Generalitat Valenciana Project GV/2010/018, and by Universitat Politècnica de València Project PAID06-08.

References


