Accelerated Aging with Electrical Overstress and Prognostics for Power MOSFETs

Sankalita Saha, Member, IEEE, Jose. R. Celaya, Member, IEEE, Vladislav Vashchenko, Shompa Mahiuddin, and Kai. F. Goebel.

Abstract—Power electronics play an increasingly important role in energy applications as part of their power converter circuits. Understanding the behavior of these devices, especially their failure modes as they age with nominal usage or sudden fault development is critical in ensuring efficiency. In this paper, a prognostics based health management of power MOSFETs undergoing accelerated aging through electrical overstress at the gate area is presented. Details of the accelerated aging methodology, modeling of the degradation process of the device and prognostics algorithm for prediction of the future state of health of the device are presented. Experiments with multiple devices demonstrate the performance of the model and the prognostics algorithm as well as the scope of application.

Index Terms—Power MOSFET, accelerated aging, prognostics.

I. INTRODUCTION

POWER electronic devices have become a crucial component in most modern energy systems varying from wind turbines to hybrid electric cars as they are integral to power converter circuits. Leading research efforts related to these devices are focused on new device designs for performance optimizations and cost reduction. However, understanding the behavior of these devices as they undergo regular wear and tear during their lifetimes or unexpected operational conditions is equally important, if not more, for ensuring operational safety and preventing unexpected failures. In particular, for applications where remote deployment is used such as off-shore wind turbines, a clear understanding of the failure modes of these devices and corresponding health management techniques is required to ensure both performance and cost-effectiveness.

Degradation in the characteristic performance of the device can happen not only when they are subjected to operating conditions beyond their specifications but also due to regular usage and storage over time. Currently, most of the health management of the devices is based on reliability based maintenance [1]. However, in many cases, this is neither sufficient nor efficient. Each of these devices can go through very different life-cycles and hence different aging. Hence, if maintenance is done based purely on reliability analysis, in most cases the devices will either be discarded before they have reached their end of life (EoL) or worse, fail before their scheduled replacement. Given the criticality of the systems in which these devices are used, a more customized approach for individual devices based on the application profile and usage monitoring is required.

This paper presents such a customized approach for a power MOSFET (Metal Oxide Semiconductor Field Effect Transistor). The approach is based on prognostics based health management techniques used traditionally in mechanical systems. These techniques involve periodic monitoring in order to assess the remaining life of a degraded or aged system followed by making a prediction of the time of failure under anticipated future loading and operational conditions. The aging mechanism is emulated with an accelerated aging test procedure. Accelerated aging tests (ALTs) or highly accelerated aging tests (HALTs) are used very commonly in the reliability domain and the semiconductor industry to estimate the expected lifetimes of the device as well as limits on operating conditions [2]. However, the accelerated tests used for the experiments in prognostic health management are somewhat different as explained in later sections. The accelerated tests used in this paper target a particularly weak region in the power MOSFET — the gate interface — which is susceptible to damage from high voltage [3]. The results from the accelerated aging tests are analyzed to first identify a precursor of damage and then make estimates of the remaining useful life (RUL) for a given device based on the precursor trend. RUL estimates can then be used to make decisions for maintenance of the device as well as the enclosing sub-system.

Currently, most ALTs and HALTs for power electronics in the reliability domain focus on thermal stress either through temperature or power cycling [4] and mechanical stress [5]. Within the prognostics domain, the work done so far has mainly looked into electro-thermal cycling induced thermal fatigue ([6], [7] and [8]). However, purely electrical stress related failures are quite common to power MOSFETs as they are prone to damage from radiation, electro-static discharge (ESD) and lightning surges. A significant body of work has focused on radiation related faults [9] while recently, some work on analyzing the effects of ESD [10] and lightning surges [11] on power electronics have been reported. Exploring the effects of hot-carrier effects in MOSFETs have received considerable attention as well ([3], [12]). However, most of the work is not oriented toward prognostics and is limited to developing testbeds to trigger and study the phenomenon. To the best of our knowledge, this work represents the first attempt at developing a prognostic technique for electrical stress on power MOSFETs.
II. BACKGROUND

A power MOSFET is a special MOSFET that is designed to handle large amounts of power and most differ from normal MOSFETs in their structures. The main difference being in the fact that most power MOSFETs have a vertical structure as against the planar structure of normal MOSFETs as shown in Figure 1. Thus, the channel current flow is in the vertical direction in a power MOSFET. This leads to different physical limitations and failure modes compared to MOSFETs. Based on the fault modes, the main limits on operation are:

- **Gate oxide breakdown:** The gate oxide region of power MOSFETs is quite susceptible to damage mainly due to decrease in gate dielectric thickness to allow for faster switching speeds. Exceeding the limits on gate voltage reduces the lifetime of the device significantly and in many cases can lead to immediate failure of the device.
- **Maximum drain to source voltage:** Power MOSFETs have a maximum specified drain to source voltage, beyond which breakdown may occur. Exceeding the breakdown voltage causes the device to conduct in an uncontrolled mode, potentially damaging it and other circuit elements due to excessive power dissipation.
- **Maximum drain current:** The drain current must generally stay below a certain specified value (maximum continuous drain current) though it can reach higher values for very short durations of time (maximum pulsed drain current, sometimes specified for various pulse durations).
- **Maximum temperature:** The junction temperature must stay under a specified maximum value for the device to function reliably. This temperature is determined by MOSFET die layout and packaging materials.

In many cases, any combination of the above operation limits may be violated. For example, exceeding the limit on gate voltage could happen very easily through electrostatic discharge. As mentioned earlier, in order to provide reliable usage of these devices in a continuous manner in safety-critical applications, efficient management of their health conditions through prognostics is required. The methodology to do so involves multiple steps starting with identifying the main failure modes and mechanisms of the component as well as precursors to failure. The next and one of the most critical steps is to develop an accelerated aging methodologies to stress the devices under test. This could utilize single or a combination of stress conditions while continuously measuring key parameters. The accelerated aging methodology utilized in prognostics differ from the ones used in reliability tests or qualification tests used by device manufacturers. This is because the main aim in these aging experiments is to determine not only the device’s time to failure but also to estimate updates on the state of health of the device till failure. The results from the accelerated aging experiments are then analyzed to build a degradation model which is ultimately used in a prognostics algorithm to predict future behavior subject to anticipated usage and the device’s individual characteristics.

The accelerated aging could target extrinsic failure modes or intrinsic failure modes. Extrinsic failure modes include faults related to the packaging of the device, particularly due to mechanical stresses due to thermal structure. Thermal cycling is regularly used to accelerate the aging of the devices by cycling between temperatures considerably larger than those seen in normal operation. In this paper, we focus on an intrinsic failure mode caused by degradation due to overstress in the gate area. The need for faster switching speed and hence smaller input resistance has led to significant decrease in the gate width. This makes it susceptible to damage even with low levels of over voltage which can accumulate over time and ultimately lead to decrease in performance beyond accepted levels and finally complete switching failure. In our experiments, this intrinsic failure mode is triggered in isolation from other extrinsic faults in order to ensure that the degradation and prognostic models capture the former fully. In the following section, details of the aging methodology and the results obtained from that are presented.

III. ACCELERATED AGING TESTS

Electrical stress at the gate area of the power MOSFET can be implemented by applying high voltage. However, this also leads to rise in temperature within the package. Thus, the aging setup needs to ensure that the internal temperature is held within the safe operation levels which would then isolate the electrical stress from other factors. The aging setup used for our experiments allow for precise control of the device temperature thereby ensuring that no package related failures are induced. In addition, periodic measurements of electrical characteristics of the device were made to estimate its state of life. Note that in this setup, the device does not have any built-in gate protection. The aging circuit also does not provide any such protection, since the main aim is to stress the gate. For further details of the testbed, please refer to [13]. The accelerated aging testbed has 3 main parts:

- The electrical operation unit of the device which consist of gate drivers, power suppliers and function generator to control the operation of the device under test (DUT).
- The in-situ measurement unit to measure electrical and thermal parameters by means of commercially available measurement and data acquisition equipments.
- A thermal block section which allows for monitoring and control of the DUT temperature. The flow of heat is regulated in order to control the temperature of the device. Hence, the block can be used as a hot or cold plate depending on the configuration.
Details of the thermal block and the aging setup is shown in Figure 2 [13]. The accelerated stress test on a device was performed by periodically subjecting the device to high gate voltage using the above set up followed by electrical parameter characterization tests using a source measurement unit (SMU) from Keithley (Keithley 2410 series). These tests provide a measurement of the incurred damage:

- **Breakdown voltage $V_{BR(DS)}$:** This gives the voltage level at which the drain-source path of the device starts conducting drain current given that the gate is not biased.
- **Leakage Current $I_{DSS}$:** This is the current flowing from drain to source as the gate is shorted with the source (no gate bias).
- **Threshold Voltage $V_{th}$:** This voltage refers to the minimum gate bias required to strongly invert the surface under the poly and form a conducting channel for drain current to flow.

The experiments were carried out on commercially available power MOSFETs (IRF520Npbf). On analyzing the above characteristics of the aged devices it was observed that the threshold voltage started increasing with aging as shown in Figure 3. An increase in threshold voltage directly corresponds to increase in switching time and hence reduced frequency. Using such degraded device would affect any enclosing switching system. All these devices have minimum specifications for their threshold voltages specified by the manufacturers. When the device do not match the ratings they are considered faulty and not further functional. The change in threshold voltage as shown in Figure 3 can be considered as a precursor to failure. Note that in the characterization tests, leakage current and breakdown voltage were also measured, but significant change in their characteristics was not observed. Hence they were not considered further in the prognostics formulation.

Threshold voltage deviation is mostly considered a result of defects generated in the gate area which includes the gate oxide as well as the gate and semiconductor interface ([14],[3]). Though, most of the analytical work done in understanding the development of the defects have been performed for normal MOSFETs, important analogies – and hence insight into the faults – can be drawn from them for the power MOSFETs. In ([15], [16]), the authors provide a detailed analysis of how these defects are generated due to the presence of the high electric field at the gate. In general, application of very high field (due to high voltage at gate) or high negative voltage at drain, leads to the generation of hot carriers in the silicon substrate which could be electrons or holes. These carriers, due to their very high kinetic energy then can move around and get trapped near the silicon-silicon dioxide interface or the dioxide itself. This in turn, affects the the behavior of the device in particular the device transconductance.

In order to determine whether such gate oxide and gate interface traps can lead to threshold voltage deviation for our experiments, simulations were carried out using the numerical analysis tool DECIMM $^{TM}$ from Angstrom ([17], [18]). The power MOSFET was modeled as two-dimensional parameterized device structure with analytical implant profiles definition. Then a set of the equations for a drift-diffusion model was solved for the most typical parameters of the silicon material at room temperature conditions. In addition to typical mobility models (effect of the strong electric field, low field mobility), models for recombination (Shockley-Read-Hall generation recombination, Auger recombination and impact ionization) were also used for the pristine model of the device. The trench gate depth was defined to be $5 \mu m$. Next, traps were added at the silicon oxide interface – both positive and negative charges – to assess their effects on the device behavior. Figure 4 shows the effect on the threshold voltage for varying magnitudes of charge concentration. In this case, the trapped charges were uniformly distributed in the silicon oxide interface region. Introduction of positive charge traps leads to a decrease in $V_{th}$, while negative charge traps causes an increase. Since, the behavior observed in the aged devices reflected an increase in $V_{th}$, further simulations with varying magnitude of the negative charge concentration were explored. From the simulations results shown in the same Figure 4, the increase in the value of $V_{th}$ is correlated with increase in the negative trap charge density.

Further simulations with non-uniform distribution of the negative trap charges, i.e., localized trap charges, were also carried out. Figure 5 summarizes the result from these simu-
charge concentration is Fig. 5. Shift in threshold voltage with localized negative trap charges. The charge distribution is uniform.

Fig. 4. Shift in threshold voltage with varying density of positive and negative trap charges. The charge distribution is uniform.

Fig. 5. Shift in threshold voltage with localized negative trap charges. The charge concentration is $1 \times 10^{12} C/cm^2$.

lutions. The symbol $Y$ in Figure 5 refers to the depth in $\mu m$ as measured from the gate towards the drain. Thus, $5 < Y < 6$ implies the region between $5\mu m$ and $6\mu m$ below the silicon-silicon dioxide interface. $Y = 0$ refers to the start of the interface. Thus, the nearer to the interface the local trapped charges, the more the threshold voltage increases.

After determining the precursor to failure, the next step is to determine the EoL criteria based on the precursor trend. In this case, the EoL is determined by when these devices will be rendered non-functional. Though a complete failure may imply that the device stops switching completely, in most applications deviation beyond manufacturer’s ratings can be considered failures. Unfortunately, there are no guidelines on what constitutes their failure since these devices are used in a wide variety of applications; each application is expected to have different failure criteria based on the enclosing system’s tolerance for switching delays and errors. For the device used in our experiments, the threshold voltage for a functional device is expected to vary from a minimum of $3V$ to a maximum of $4V$. We set the failure criteria to be when the threshold voltage increases to more than an increase of $50\%$ at $6.1V$. Thus, EoL is reached by a device when its threshold voltage attains $6.1V$. In the next section, we develop a model that captures this shift in threshold voltage with aging cycles and combine it with a prognostic algorithm to predict the RUL for the aged devices.

IV. MODELING

Prognostics algorithms can be based on model-based techniques relying on a detailed degradation model of the unit, data-driven techniques where prediction models are created based on learning from past history or hybrid where a simpler model is used in conjunction with a data-driven algorithm. The hybrid methods have several advantages; the simplicity of the model makes it easier to implement especially for applications where computational resources may be scarce, while at the same time the support of data-driven methods allows to compensate for the simplicity of the model and capture complex or unmodeled behavior. In this work, a hybrid approach was used. The prognostics algorithm involves two steps; first the simple aging model is built offline based on training data sets which provide insights regarding the physical behavior of the degradation. Next, in online mode, the model is fit into a prediction algorithm – particle filters (PF) in our work – which tracks the behavior of the system for further tuning and then starts making predictions of the RUL. Typically, the prediction routine is triggered by a diagnostic flag and either continuous prediction of the state of health and RUL prediction is made or discrete time predictions are made which can then be used to make health management decisions.

Particle filtering is a sequential Monte Carlo (SMC) technique for implementing a recursive Bayesian filter using Monte Carlo simulations. It is primarily used for state estimation and tracking. The mathematical formulation for PF methods have been discussed in [19]. The basic idea is to develop a nonparametric representation of the system state probability density function (pdf) in the form of a set of particles with associated importance weights. The particles are sampled values from the unknown state space and the weights are the corresponding discrete probability masses. As the filter iterates, the particles are propagated according to the system state transition model, while their weights are updated based upon the likelihood of the measurement given the particle values. Resampling of the particle distribution is done when needed in order to prevent the degeneracy of the weights. Particle filters can be represented mathematically as follows:

$$x_k = f(x_{k-1}) + \omega_k, \quad (1)$$
$$y_k = h(x_k) + \nu_k, \quad (2)$$

where $k$ is the time index, $x$ denotes the state, $y$ is the output or measurements. Both $\omega$ and $\nu$ are samples from noise distributions which are picked from zero mean Gaussian distributions whose standard deviations are derived from the given training data, thus accommodating for the sources of uncertainty in feature extraction, regression modeling and measurement.

For state prediction purposes the same PF framework can be used by running only the state equation-based particle propagation step until the predicted state value crosses some predetermined EoL threshold. The predicted trajectory of each particle then generates an estimate of RUL, which can be combined with the associated weights to give the RUL pdf. The process is broken down into an offline learning part, and
Fig. 6. Flow of the particle filtering framework for prognostics

Fig. 7. Threshold voltage degradation curves i.e., aging curves for 3 devices obtained from accelerated aging test with $V_{GS} = 53V$.

an online tracking and prediction part. Figure 6 illustrates a simplified schematic of the process described.

A. Prognostics results

Accelerated aging was carried out for multiple devices, a few of which were used for training. The device used had a maximum gate to source voltage ($V_{GS}$) rating of 20V and the maximum power dissipation level of 50W approximately. The maximum drain current at room temperature ($25^\circ C$) is 9.7A with $V_{GS} = 10V$, while the maximum drain source voltage is rated at 100V. The devices were aged at $V_{GS} = 53V$ for periods of approximately one hour till $V_{th} > 6V$. The magnitude of the gate voltage for overstress was determined after repeated tests with different levels in order to determine the best setting which would allow considerable degradation in a reasonable time while ensuring that critical state information is not lost due to very high acceleration rate. Note that the periods of aging were not completely uniform and minor variations were present. The drain current was set at approximately 7.7A while the drain to source voltage was limited at 2.4V. Thus, the total power was limited to a maximum of 18.5W. This ensured that the stress was mainly focused on the gate due to the high gate voltage thereby isolating the degradation. From Figure 7, one may observe that the threshold voltage rises steeply at the beginning of the aging period and gradually reaches a plateau. Also, there are occasional dips in the value of $V_{th}$ instead of a monotonic increase. This is due to the rest periods between the aging cycles during which the device recovers slightly. However, these dips are not uniformly observed in all the devices since the rest periods were not uniform for all of them and also due to slight differences in device-to-device individual characteristics.

Figure 7 shows the sample aging curves of 3 devices that were used for training. Based on the aging curves from our training dataset, the following model for the threshold degradation with aging was used:

$$V_{th} = a_1 + b_1 \times (1 - \exp(-c_1 \times t)),$$

where $t$ is the aging time. Regression was used on the training aging curves to determine the parameters $a_1$, $b_1$ and $c_1$. The model fitting is shown in Figure 8. After training the threshold degradation model obtained is as follows:

$$V_{th} = 3.5 + 3.2 \times (1 - \exp(-0.005 \times t)).$$

This model was fed to the online process and the predicted values of $V_{th}$ are compared against EoL thresholds to derive RUL estimates. Note that there are variations in the curves as shown in Figure 7 due to noise arising from various uncertainties such as measurement noise, device-to-device slight variations and aging time variations. Though equation 4 does not capture these variations, it is expected that the particle filtering framework would be able to compensate for these variations through training.

Figure 9 shows the result of the RUL prediction from the particle filter based prognostics routine. The prediction is made after tracking for a few iterations which allows the prognostics algorithms to tune to the behavior of the current device compared to the model trained on different devices, as well as other environmental conditions. The dashed line represents the predicted EoL while the probability density function (pdf) around the predicted value represents the uncertainty bounds of the prediction which is expected to increase for more system noise. Note that the algorithm tracks the threshold voltage quite accurately, thereby adjusting for the rest periods as well.

Table I summarizes the prediction results for two devices at multiple prediction times. Note that multiple predictions were made and the mean values are shown in the table.

V. CONCLUSIONS

This paper presents a prognostics routine for power MOSFETs with intrinsic electrical degradation. This included a methodology for accelerated aging using purely electrical overstress and corresponding identification of a precursor of failure. The devices were aged using accelerated electrical overstress at the gate of the power MOSFET which triggered
Fig. 9. Particle filter based tracking and prediction of threshold voltage degradation and EoL of device

TABLE I

RUL PREDICTION RESULTS

<table>
<thead>
<tr>
<th>Device id</th>
<th>Time of prediction (mins)</th>
<th>Predicted RUL (mins)</th>
<th>Actual RUL (mins)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Device1</td>
<td>248</td>
<td>172</td>
<td>192</td>
</tr>
<tr>
<td>Device1</td>
<td>312</td>
<td>118.85</td>
<td>128</td>
</tr>
<tr>
<td>Device2</td>
<td>263</td>
<td>136.5</td>
<td>191</td>
</tr>
<tr>
<td>Device2</td>
<td>330</td>
<td>98</td>
<td>114</td>
</tr>
</tbody>
</table>

degradation in the threshold voltage of the device due to interface trap charge accumulation. Model for the degradation curves were derived and used in particle filtering framework to predict RUL estimates. The prognostics routine was robust and could handle system noise factors such as uneven rest periods between aging. Since the prognostics algorithm provides a RUL pdf, instead of a single value, the interpretation of the prognostic result is useful and intuitive. Future work includes refinement of the model and applying it to a larger training as well as test data set generation. Tighter integration of the the aging setup with the SMU and/or identification of other precursors that can be easily integrated to enable completely automated aging is also an important future direction of work.

REFERENCES


