In-situ Sensors for Product Reliability Monitoring

Satchidananda Mishra*, Michael Pecht* and Douglas L. Goodman**
*CALCE Electronic Products and Systems Center; **Ridgetop Group, Inc.

ABSTRACT

Some advantages of predicting reliability include providing advance warning signs of failure, and the reduction of life cycle costs by reducing inspection and unscheduled maintenance. However, predictions can be inaccurate if they do not account for the actual environments that the product is subjected to in its life cycle. This paper describes an in-situ sensor (prognostic monitor) approach, which can be used to estimate the accumulated damage and the remaining life of semiconductor devices.

Keywords: In-situ sensor, prognostic cells, damage estimation, failure prediction, remaining life prediction, health monitoring, built-in-test.

1. INTRODUCTION

Reliability is defined as the ability of a product to perform as intended (i.e., without failure and within specified performance limits) for a specified time, in its life cycle application environment. The objective of reliability prediction is to support decisions related to the operation and maintenance of the product including to:

- Reduce output penalties including outage repair and labor costs
- Optimize maintenance cycles and spares holdings
- Maintain the effectiveness of equipment through optimized repair actions
- Help in the design of future products, by improved safety margins and reduced failures
- Increase profitability

During the past 25 years, there has been a tremendous improvement in the reliability of semiconductor devices. In fact Pecht’s Law suggests that semiconductor device reliability in terms of time-to-failure (TTF) is doubling every fourteen months based on activation energy trends of semiconductor devices (see Figure 1).

The evolution of technology as well as demands for increased warranties and the severe liability of product failures have supported Pecht’s law. Specific technology improvements, which have helped to increase device reliability, have included advancements in die passivation techniques, semiconductor manufacturing processes and encapsulation materials. However there is some serious concern is to what future will bring. The downsizing of semiconductor features has lead to decrease in the gate oxide thicknesses and the distance between metallization, resulting in higher electric fields across the gate and possibility of failure, such as gate oxide breakdown and hot carrier damage.

Furthermore, as manufacturers try to keep pace with performance requirements, reliability may be traded-off for increased functionality at an affordable cost.

2. THE RELIABILITY PREDICTION CHALLENGE

A reliability prediction for a product is dependent on its structural architecture, material properties, fabrication process, and the life cycle environment. Material properties and geometry of a product are not exactly the same for all the products coming out of a production line. There may be changes in these with variations in the fabrication process affecting the reliability and can be catastrophic for a given circuit grouping.

---

*a University of Maryland, College Park, Maryland-20742, USA; http://www.calce.umd.edu
**PO Box 35441;Tucson, Arizona-85740, USA; Email: doug@Ridgetop-Group.com; http://www.Ridgetop-Group.com

a Activation energy based models have been used to statistically correlate time-to-failure to environmental condition using activation energy. The higher the activation energy of a given component, the longer the time to failure as a function of steady state temperature.
Product life cycle environment describes the storage, handling and application scenario of the product, as well as the expected severity and duration of the load conditions for each scenario. Load conditions may include temperature, humidity, pressure levels, vibrational or shock loads, chemically aggressive or inert environments, acoustic levels, sand, dust, electromagnetic radiation levels, and loads caused by operational parameters such as current, voltage and power. These conditions, either acting individually or in various combinations, affect the reliability of the electronic products. If a product is connected to other products or sub-systems, the loads (i.e., external power consumption, current, voltage transients, voltage spikes, electronic noise, and heat dissipation) also play an important part in the life cycle environment of the product.

To circumvent the challenge to reliability prediction posed by variations in manufacturing and life cycle environment, there is a need to predict impending product failure of the product in its real-time fabrication and operating conditions.

### 3. HEALTH MONITORING AS A SOLUTION TO RELIABILITY PREDICTION

Health monitoring is a method of evaluating the extent of a product’s reliability in terms of product degradation in its life cycle environment. By knowing about impending failure, based on actual life cycle application conditions, procedures can be developed to mitigate, manage or maintain the product.

In health monitoring a product’s degradation is quantified by continuous or periodic measurement, sensing, recording, and interpretation of physical parameters related to the product’s life cycle environment and converting the measured data into some metric associated with either the fraction of product degradation or the remaining life of the product (in terms of days or distance in miles). A product’s degradation can be assessed in terms of physical degradation (e.g., cracks, deflection, delamination) and electrical degradation (e.g., increase in resistance, increase in threshold voltage), or...
performance degradation, such as deviation of the product’s operating parameters (e.g., electrical, mechanical, or acoustic) from expected values. Methods employed for health monitoring are non-destructive test (e.g., ultrasonic inspection, liquid penetrant inspection, and visual inspection) and operating parameter monitoring (e.g., vibration monitoring, oil consumption monitoring and thermography (infrared) monitoring).

An example of health monitoring is Boeing’s extended-range twin-engine operations (ETOPS) program. The ETOPS approach is based on the continuous monitoring of aircraft application conditions to identify problems before they affect aircraft operation or safety. Typical examples of ETOPS are engine condition monitoring (ECM) and oil consumption monitoring. ETOPS operators are required to use ECM programs to monitor adverse trends in engine performance and execute maintenance to avoid serious failures (e.g., events that could cause in-flight shutdowns, diversions, or turnbacks). The ECM programs allow for monitoring of engine parameters such as exhaust temperature, fuel and oil pressures, and vibration. In some cases, oil consumption data and ECM data are combined to identify problems in normal engine operation.

Built-in-test (BIT) concept is another technique employed for diverse applications. BIT is a hardware-software diagnostic mean to identify and locate faults. Two types of BIT concepts are employed in electronic systems, interruptive BIT (I-BIT) and continuous BIT (C-BIT). The concept of I-BIT is that normal equipment operation is suspended during BIT operation. Such BITs are typically initiated by the operator or during a power-up process. The concept of C-BIT is that equipment is monitored continuously and automatically without affecting normal operation.

Product degradation in semiconductor devices caused by semiconductor level mechanisms can be quantified by mounting prognostic cells that are located on the same chip but is subjected to accelerated conditions than the original circuitry. By proper inclusion of pre-designed, pre-calibrated prognostic cells, fail-safe operation can be achieved in real life applications.

4. PROGNOSTIC MONITOR APPROACH

A prognostic monitor is a pre-calibrated semiconductor cell (circuit) that is co-located with the actual circuit on a semiconductor device. The prognostic monitor thus experiences the same manufacturing process and the same environmental parameters as that of the actual circuit. Hence as long as the operational parameters are the same, the damage rate is expected to be the same for both the circuits. By incorporating these monitors as a part of the sub system we ensure that the cells see the same operational environment as the product from fabrication, to test to operation. It assures that any parameter that affects the product reliability will also affect the monitor causing its failure. This approach enables the cells to overcome the limitations of off-line tests, which are often performed to represent an average expected performance of circuits and have no means to account for the effects of the operational environment seen by the circuit in use.

Provided that both the circuits are designed and manufactured with the same technology, both circuits will have the same mean time to failure (MTTF). Although both the circuits are in the same environment, loading conditions can be altered in a controlled way by inducing variations in the operating parameters. This is achieved by increasing the current density inside the cells in a controlled manner. With the same amount of current passing through both circuits, if the cross sectional area of the current carrying paths in the cells is decreased, higher current density is achieved. Further control in current density can be achieved by increasing the voltage level applied to the cells. A combination of both of these techniques can also be used. Higher current density leads to higher internal (joule) heating causing higher stress on the cells. When a current of higher density pass through the cells as compared to the actual circuit, they are expected to fail faster.

---

b Damage to the prognostic sensors resulting from environmental parameters is approximately same as that of the actual circuitry due to their location on the chip. Hence if damage due to the operational parameters (current, voltage, power) is same, the combined damage will be the same for both the circuits.
The in-situ sensor approach can be described by referring to the idealized bathtub failure rate curves. Semiconductor reliability is often represented as an idealized bathtub curve, which can be divided into three regions, i.e., infant mortality region, useful life region and the wear out region (see Figure 2). The infant mortality region begins at time zero, which is characterized by a high but rapidly decreasing failure rate. Most failures in this region result from defects caused in the material structure during the manufacturing, handling, or assembly. These defects may take the form of missing metal from the side of a thin film conductor, internal cracks, and foreign inclusions. After the infant mortality region the failure rate decreases to a lower value and remains almost constant for a long period of time. This long period of an almost constant failure rate is known as the useful life period. Ultimately the failure rate begins to increase as materials start degrading and wear out failures occur at an increasing rate. This is a result of continuously increasing damage accumulation in the product. This region in the bathtub curve is commonly referred as wear-out or end-of-life period. Figure 2 shows two idealized bathtub curves one for the actual circuit and the other for the prognostic cell. The hazard rate, \( h(t) \) in the plot is defined as

\[
h(t) = -\frac{1}{n_s(t)} \frac{d[n_s(t)]}{dt}
\]

where \( n_s(t) \) is the number of surviving products at the end of time \( t \).

In the infant mortality region failures are often due to defects introduced during manufacturing, handling, and storage. Because the cells are designed to be a part of the actual chip, defects introduced in the product affects the prognostic cells in the same way as that of the actual circuitry. As a result infant mortality can also be expected for the cells.

Infant mortality effect can be increased in the prognostic cells by intentionally adding defects in them. This process known as error seeding introduces new defects, which can interact with the defects cause during manufacturing, handling and storage. Combined effect of both defects makes the chip fail during functional testing, which can be the indicative of infant mortality failure.
Because of accelerated failure mechanisms the prognostic cells have higher failure rates than the actual circuit, for the entire life period. Further since the failure mechanisms are accelerated in the prognostic cells, the wear out region occurs earlier than that of the actual circuit (see Figure 2). To predict the end-of-life period, the prognostic cells must fail prior to the actual circuit failure. As the failure of both the circuits are in the form of distributions, the failure distribution of prognostic cells on a particular chip must be before the failure distribution of the actual circuit. If the complete prognostic cell failure distribution is before the onset of the circuit failure distribution, then all prognostic cells will fail prior to the circuit failure there by predicting failure of the actual circuit. In other words the failure points of the prognostic cells must be calibrated in such a way that their failures occur before the actual circuit wear-out region.

5. MONITORING SEMICONDUCTOR FAILURE MECHANISMS

The major semiconductor failure mechanisms that can be monitored are time dependent dielectric breakdown (TDDB), electromigration, and hot-carrier aging\textsuperscript{14,15}. The degradation due to these mechanisms is attributed to damage accumulation in the product, which increases with time until it manifests itself as a failure.

5.1 Time Dependent Dielectric Breakdown

Time-dependent dielectric breakdown (TDDB) is the formation of low-resistance dielectric paths through localized defects in dielectrics, such as in MOS devices only thermally grown or other oxides\textsuperscript{14}. Failures typically occur at weaknesses in the oxide layer (such as presence of stacking faults), due to poor processing or uneven oxide growth. Oxide breakdown can also be due to charge accumulation in the oxide and to local thinning and discontinuities in the oxide caused by metal precipitates. TDDB is characterized by sudden, usually permanent, DC conduction in the dielectric of MOS capacitors. During use, applied voltages and currents generate defects and charge traps that accumulate both within the bulk oxide and at either interfaces. With time, charge accumulation raises the density of defects and traps to a critical level where sufficiently high local electric fields and currents are generated to cause a thermal runaway and melting of microscopic regions.

The effects of thin oxides on this mechanism is still under investigation and it appears that the high process control during the growth (only a few layers) of oxides have reduced the defects to a very low level and thereby reduced the potential points of failure. In thin oxides, it also appears that the dielectric failures can manifest itself as a soft error and recover without reaching TDDB.

5.2 Hot Carrier Aging

Hot carrier aging is degradation of device characteristics due to the trapping of charge in the gate dielectric. Carriers (electrons or holes) not in thermal equilibrium with the rest of the semiconductor crystal lattice, i.e., the carriers whose energy is above that of the conduction and valence band edges are called “hot carriers.” High energy hot carriers can cause a number of effects within a MOS device including impact ionization and avalanche multiplication in the drain region, resulting in increased drain current and a substrate current. Injection of the carriers into the gate dielectric region can also cause a gate current\textsuperscript{16}.

These currents are indicators of the presence of hot carrier effect. Over time, accumulation of charges in the dielectric can change the threshold voltage and transconductance of the device and when these two parameters change beyond a certain limit the device is considered “failed.” Hot carrier aging can be “reversed” during operation through release of the trapped charges. Changes in threshold voltages can cause delay faults in digital circuits that are difficult to locate.

5.3 Electromigration

Electromigration occurs in the interconnect metallization tracks. This diffusion phenomenon occurs when the current density in the tracks is sufficiently high to cause drift of the metal atoms in the direction of the electron flow\textsuperscript{17}. As the atoms migrate, there will be a net depletion “upstream” and an accumulation “downstream” of material. This process culminates in the formation of voids or extrusions that leads to an increase in electrical resistance or to a short between adjacent lines. Electromigration is an intrinsic wear out mechanism, but presence of defects changes the times to failure due to electromigration. The number of ions that migrate, i.e., the ion flux density, is dependent on:
• The magnitude of forces that tend to hold the atoms in place, and thus the elemental nature of the conductor or, crystal size, and grain boundary chemistry.
• The magnitude of forces that tend to dislodge the atoms, including the electron current density, temperature, and mechanical stresses.

Electromigration has been investigated extensively for aluminum interconnects and models for times to failure are available in the literature. The introduction of copper interconnects appears to have alleviated the problem of electromigration and research is underway in developing time to failure models for copper electromigration.

6. CALIBRATION OF PROGNOSTIC CELLS

The time difference between the prognostic event detection of the prognostic cell and the end-of-life of the actual circuit is known as the “prognostic distance.” As failure for the host circuit as well as the prognostic cells are not point values but distributions, the prognostic distance is also in the form of a distribution. This distribution is used to calibrate the cells using simulation techniques. Simulation techniques determine the load distribution for the circuits under a given life cycle environment.

Models for semiconductor failure mechanisms including time dependent dielectric breakdown, hot carrier aging, and electromigration are used to estimate cycles to failure for a circuit in its life cycle environment. These models make use of the load distribution to predict the cycles to failure (temperature, humidity, vibration, current, power) and hence estimate the mean time to failure (MTTF) for both the circuits under the expected life cycle environment. By knowing the acceleration mechanism and the time to fail for both the circuit, an acceleration factor can be defined for the prognostic cells as

\[
\text{Acceleration factor} = \frac{\text{Time to failure of the actual circuit}}{\text{Time to failure of the prognostic cell}}
\]

Higher current density passing through the cells act as the accelerating parameter. A higher value of acceleration factor corresponds to accelerated electrical conditions, which results in shorter time-to-failure of the prognostic cell.

Based on the earlier work conducted by US Air Force, Ridgetop Group has developed and refined the InstaCell™ library of prognostic cells that can provide indication of an impending failure for different semiconductor failure mechanisms (time dependent dielectric breakdown, hot carrier aging, radiation damage and electromigration). Dimension of these cells vary with the type of cell, but are in the range of 20 microns. For these cells failure point is calibrated to be of the statistical end-of-life of the actual circuit\(^\text{18}\). While the InstaCell™ library is currently oriented at embedded IC applications; there is some work being planned on applying the prognostic cells on power semiconductors such as power MOSFETs and IGBTs.

Software is available to simulate the load distribution of semiconductor chips numerically. Computer Aided Design of Microelectronic Packages (CADMP-II)\(^\text{19}\) is one such software that use dominant failure mechanisms determined using physics-of-failure models to calculate the time to failure for each of a series of potential failure mechanisms.

7. DAMAGE ESTIMATION

The installation of prognostic sensors mounted on the same chip as that of the actual circuitry enables multiple sensing with different acceleration factors. Prognostic cells with higher acceleration factor will fail faster than the others. Figure 3 shows the failure points of prognostic cells on a plot of acceleration factor vs. time. The failure points represents failure of different prognostic cells that are pre-designed with different acceleration factors to get different failure time.

\(^{\text{18}}\) CADMP-II is a component level virtual qualification tool developed by CALCE Electronic Products and Systems Center, University of Maryland, College Park
and are located on the same chip as that of the actual circuit. Acceleration factor “one” on the line corresponds to the failure of the actual circuit. Damage accumulated on the circuit at any point of time is given by

\[
\text{Accumulated damage ratio} = \frac{1}{\text{Acceleration factor}}
\]  

(3)

Given the plot, the remaining life of the actual circuit can be predicted, which is the time difference between the points with an acceleration factor equal to “one” and a time where the prediction is needed. In equation form

\[
RL_t = t_{AF=1} - t
\]

(4)

where \( RL_t \) is the predicted remaining life at time \( t \) and \( t_{AF=1} \) is the time at acceleration factor equals 1.

8. INTERFACING USING THE JTAG BUS STRUCTURE

The prognostic cells are accessed using a JTAG bus with a boundary scan method to provide means of accessing the information quickly and efficiently. Boundary scan test (BST) architecture can test pin connections and capture functional data while the device is operating normally\textsuperscript{20}. The device consists of a set of registers, five control pins, and core logic. Boundary scan cells in a device can force signals onto pins, or capture data from pins or core logic signals.

The BST circuitry has three types of registers; the instruction register, bypass register and boundary scan (or prognostic cell) register.
• The instruction register: This is used to determine the action to be performed and the data register to be accessed.

• The bypass register: This is a 1-bit register used to provide a minimum length serial path between TDI and TDO.

• The boundary-scan register or prognostic cell register: This is a shift register, composed of all the boundary-scan cells of the device. There may be a chain of these boundary scan registers, which interact with the core logic to perform the logic operation beyond printed circuit boards, to extend across the entire product.

Using the JTAG toolkit, it is possible to interface using the standard scan test bus that employs IEEE 1149.1. With this interface the five control lines can be used:

• TCK - Test Clock: This is the clock input to the boundary scan test (BST) circuitry, which controls the timing of the test interface independently from any system clocks. TCK is pulsed by the equipment controlling the test and not by the tested device. It can be pulsed at any frequency (up to a specified maximum), even at varying rates. TMS - Test Mode Select: The input pin that provides control signal to determine the transitions of the “Test Access Port” controller.

• TDI - Test Data In: This is the serial input pin for instruction, which supplies the data to the JTAG registers. Data is shifted at the rising edge of TCK.

• TDO - Test Data Out: This is the serial data output pin, which is used data output from the JTAG registers. It carries the sampled values from the boundary scan chain (or other JTAG registers) and propagates them. Data is shifted at the falling edge of TCK.

Figure 4: This schematic circuit shows the main data transfer lines in implementation of prognostic cells with JTAG data extraction.
• TRST – Test Reset: This is an optional pin, which initializes and disables the test interface.

The prognostic InstaCell™ is part of the boundary scan register (see Figure 4). Shifting of data in these registers occurs at rising and falling edge of the clock input. The input from TDI passes through a set of boundary scan registers and the core logic before reaching TDO. TMS is used to invoke a “self test” function to assure that the prognostic cell is functional. The TDO changes its logic state upon a detected failure event. This architecture allows the failure information of the cells to be extracted using maximum up to four pins for a particular acceleration factor.

9. APPLICATION OF PROGNOSTIC SENSORS IN AUTOMOTIVE ENGINE CONTROL

An automobile engine control unit (ECU) controls the vehicle’s injection and ignition. In order to burn completely the ratio between air and fuel injected to each cylinder should be kept constant. It is important to provide enough time for complete burning of the fuel, which is done by advancing the spark with respect to the instant when the piston is at its highest point. The time by which the spark is advanced is computed from the engine RPM. This parameter also affects fuel consumption and emission. The life of the ECU is normally the life of the vehicle. Unfortunately redundancy is seldom used due to cost considerations.

An ECU typically consists of a set of sensors to measure the engine status, an electronic unit that possesses the data coming from the sensors and drives the actuators, and actuators that execute the commands received from the control unit. A control strategy is implemented in the electronic unit to optimize the fuel injection and ignition. Failures of the ECU can have catastrophic effects, especially if the vehicle is moving when failure occurs.

The output from the cells can either be taken from the I/O port of the circuit cards or the same connector connecting the electronic unit and the sensors, actuators. The output is connected to JTAG boundary scan devices. The output of these devices are used as input to a software, which does two functions, 1) keeps track of the operation cycles of the car, 2) estimates the remaining life based on failure information of the sensors. Output of the software can be displayed in terms of life of the items as miles of the car or days of run.

The key benefits can be extended to other parts of the vehicle including the antilock braking system, transmission control or other critical modules. By reducing potential fault ambiguities, this methodology can provide reduction in mean time to repair (MTTR) and warranty expenses to the manufacturer.

10. SUMMARY

Prognostic monitors are pre-calibrated circuit cells that can be located on the same chip as that of the actual circuitry to predict semiconductor failures. These prognostic monitors are used along with a software algorithm to find out the damage accumulated in the actual circuit in terms of semiconductor failure mechanisms and the remaining life. By early failure prediction, the prognostic monitors can help in scheduling maintenance at the proper time. Determination of the onset of failure real time can help in avoiding the cost that may be encountered due to the fail of electronic products.

REFERENCES


