Some particular aspects of reliability evaluation of monolithic ICs

T. I. Băjenescu, * M. I. Bâzu **

* Company for Consulting  C.F.C., La Conversion / Switzerland
** National Research Institute for Microtechnology I. M. T., Bucharest, Romania

tmbajenesco@bluewin.ch
MBazu@imt.ro

Abstract: The paper gives an overview of the present situation in the domain, the front line in the battle for the best products. After a short introduction, the main problems arisen in reliability evaluation of monolithic integrated circuits, the evaluation itself and some new points of view concerning the dynamic life testing, screening and burn-in, accelerated tests, physics of failure of plastic encapsulated microcircuits (PEM) and process reliability are presented.

Key words: Accelerated tests, best reliability of products, dynamic life testing and burn-in, evaluation of monolithic integrated circuits, physics of failure of PEM.

1 Introduction

The huge progress obtained in the integrated circuit field led to smaller dimension electronic equipment and reduced costs, but also to improvements in power capability, reliability and maintainability. The predictions are that the strong world-wide increase of computer and communication market will lead to an even higher growth rate of semiconductor industry in the next decade, 20% per year, with a level of $300 billion in the year 2001 [1].

The complexity of ICs increased every year. In fact, Gordon Moore, in the 70’s, talked about a doubling of IC complexity every 18 months, with a corresponding decrease in cost per function. This became the so-called Moore’s Law, proved to be true for more than the subsequent twenty years. Many factors contributed to keeping this model on course: the improvement of design tools and manufacturing technologies, but also the permanent growth of the reliability level. The intrinsic reliability of a transistor from an IC improved with two orders of magnitude (the failure rate decreases from $10^{-6}$ h$^{-1}$ in 1970, to $10^{-8}$ h$^{-1}$ in 1997). But also, in the same period of time, the number of transistors per device increased with 9 orders of magnitude! Therefore, the IC reliability increased even faster than the prediction given by Moore’s Law.

The model for reliability growth was called “Less’s Law”, taking into account the known philosophy from the architectural design: “Less is More” [1]. Actually, Less’s Law means a tremendous increase of the requirements for the IC’s failure rate: from 1000 failures in 10$^9$ devices x hours (or 1000 Fits), now only some Fits in a single digit are required. It is worthwhile to note the change in the predictions made by the Semiconductor Industry Association (SIA) in the editions 1994, 1995 and 1997 of the National Technology Roadmap for Semiconductors [1][3]. From Table 1, one may see that the forecast was overpassed by the reality: the performances previewed in 1994 and 1995 for 1998 were attained earlier, in 1997.

Basically, two types of integrated circuits were developed since now: bipolar and MOS, taking into account the basic cell: bipolar transistors or MOS ones, respectively. In the beginning, the MOS ICs were $n$-channel MOS ICs or $p$-channel MOS ICs, but
1.1 Modeling IC reliability

First, only simulators for one or two subsystems or failure mechanisms were arise, such as: RELIANT [14], only for predicting electromigration of the interconnects and BERT EM [15]. Both use SPICE for the prediction of electromigration by derivating the current. Other electromigration simulators were CREST [15], using switch-level combined with Monte-Carlo simulation, adecuated for the simulation of VLSI circuits and SPIDER [16].

Other models were built for hot-carrier effects: CAS [17] and RELY [18], based also on SPICE. An important improvement was RELIC, built for three failure mechanisms: electromigration, hot-carrier effects and time-dependent dielectric breakdown [19]. A high-level reliability simulator for electromigration failures, named GRACE [20], assured a higher speed simulation for very large ICs. Compared with the previously developed simulators, GRACE has some advantages [20]:

A high-level reliability simulator for electromigration failures, named GRACE [20], assured a higher speed simulation for very large ICs. Compared with the previously developed simulators, GRACE has some advantages [20]:

- an orders-of-magnitude speedup allows the simulation of VLSI many input vectors;
- the generalised Eyring model [21] allows to simulate the ageing and eventually the failure of physical elements due to electrical stress;
- the simulator learns how to simulate more accurately as the design progresses.

If the typical failure mechanisms are known, by taking into account the degradation and failure phenomena, models for the operational life of the devices can be elaborated. Such models, in contrast with the regular CAD tools determining only wearout phenomena, predicts also the failures linked to the early-failure zone.

A Monte-Carlo reliability [26] simulation for IC (Fig. 1), incorporating the effect of process-flaw, test structure data, mask layout and operating environmental conditions was proposed by Moosa and Poole [13]. The device was divided into subsystems (metallisation, gate oxide, wire bonds and packaging), affected by various failure mechanisms. Further on, these systems were divided into elementary objects (e.g. for metallisation: metal runs, vias, contacts), which may have various failure modes/mechanisms.

The reliability-measures of the objects are obtained by accelerated life testing on specially designed test structure. Then the data are extrapolated at the subsystem and device level.

The simulation procedure is detailed in Fig. 1.

This simulator was used for a two-layers metal interconnect subsystem and the typical failure mechanism was electromigration. The effect of various grain size distributions and defect (voids) size distributions was checked and the results (given as cumulative failures vs. system failure times) agree well with previously reported results.
Table 1. Predictions for Si CMOS technology development: 1994, 1995 and 1997 editions of the National Technology Roadmap for Semiconductors.

<table>
<thead>
<tr>
<th>Year</th>
<th>Minimum feature (µm)</th>
<th>Lithography</th>
</tr>
</thead>
<tbody>
<tr>
<td>1995</td>
<td>0.35</td>
<td>0.35</td>
</tr>
<tr>
<td>1997</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>1998</td>
<td>0.25</td>
<td>0.25</td>
</tr>
<tr>
<td>1999</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>2001</td>
<td>0.18</td>
<td>0.18</td>
</tr>
<tr>
<td>2003</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>2004</td>
<td>0.13</td>
<td>0.13</td>
</tr>
<tr>
<td>2006</td>
<td>-</td>
<td>0.10</td>
</tr>
<tr>
<td>2007</td>
<td>0.10</td>
<td>0.10</td>
</tr>
<tr>
<td>2009</td>
<td>-</td>
<td>0.07</td>
</tr>
<tr>
<td>2010</td>
<td>0.07</td>
<td>0.07</td>
</tr>
</tbody>
</table>

Table 2. Incidence of main failure mechanisms (in %) arising in infant mortality period.

<table>
<thead>
<tr>
<th>Failure mechanism</th>
<th>Commercial circuits</th>
<th>Memories</th>
<th>Western Electric ICs</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>TTL</td>
<td>CMOS</td>
<td>TTL (Beam lead)</td>
</tr>
<tr>
<td>Electrical overcharge</td>
<td>4</td>
<td>60</td>
<td>17</td>
</tr>
<tr>
<td>Oxide defects</td>
<td>2</td>
<td>1</td>
<td>51</td>
</tr>
<tr>
<td>Surface defects</td>
<td>18</td>
<td>-</td>
<td>24</td>
</tr>
<tr>
<td>Connections</td>
<td>37</td>
<td>5</td>
<td>7</td>
</tr>
<tr>
<td>Metallisation</td>
<td>30</td>
<td>34</td>
<td>-</td>
</tr>
<tr>
<td>Various</td>
<td>9</td>
<td>-</td>
<td>1</td>
</tr>
</tbody>
</table>

Fig. 1. Monte-Carlo reliability simulation procedure for Ics.
2. Reliability evaluation

2.1 Some reliability problems

From a theoretical viewpoint, a higher reliability is expected for integrated circuits than for discrete components. In practice, these expectations were surpassed, because some basing conditions were fulfilled:

\[
\begin{array}{|c|c|}
\hline
\text{Material} & \text{Current density (mA/cm}^2\text{)} \\
\hline
\text{Cu} & 0.7 \\
\text{Al} & 0.5 \\
\text{Others} & 0.3 \\
\hline
\end{array}
\]

Fig. 2. Evolution of the metallisation technology and corresponding allowed current densities.

- high level and automated industrial fabrication,
- constant quality materials, screening tests for finished products.

Before stating which type of integrated circuit is adequate for a specific use, one must carefully analyse the design, the most important parameters, the dimensions, the costs and the limits imposed by the reliability. Such studies must take into account:

- a comparison of the total costs (development, manufacturing and testing costs),
- a comparison of the important parameters (resistor tolerances, temperature coefficient, speed, voltage level) with limitations specific to each type,
- an evaluation of the dissipated power of the circuit and of the thermal resistance of the encapsulated circuit, because an acceptable temperature of the substrate must be assured.

2.2 Evaluation of integrated circuit reliability

Generally, three main problems arise at the evaluation of integrated circuit reliability.

1. For modern devices, the failure rates decrease under a certain limit and the conventional methods become less usable. To overcome these difficulties, two solutions may be discussed:

- To perform on a very high number of integrated circuits reliability tests in normal operational conditions, with the duration of a couple of years. Obviously, this solution is unacceptable. As an example, if one has to test a failure rate of $10^{-9}\text{h}^{-1}$ (called also 1 FIT), 1000 devices must be tested for 114 years and only one device to be found defect.

- To perform on some integrated circuits reliability tests in higher than normal conditions, the so-called accelerated tests. This method may be applied only if at the accelerated tests the failure mechanisms are the same as for normal operational conditions. And this fact must be indubitably demonstrated.

The accelerated tests are used in the purpose to obtain quickly and with a minimum of expense information about the reliability of the product. The used stresses are higher than for normal operational conditions, the results are extrapolated and the failure rate for normal conditions is obtained. Usually, the accelerated tests contain combinations of stresses such as: temperature, bias, pressure, vibrations, etc. [11]. If the temperature is the only variable of the accelerated tests, the Arrhenius model may be used. To obtain reliable results, relatively short testing times must be used. So, using various levels of the same stress factor one may follow the real behaviour. The analysis of the physico–chemical process leading to failure allows obtaining the correlation between the speed of these phenomena and the stress and, as a result, the real dependence of time to failure on stress levels.

2. The rapid development of the manufacturing technology for integrated circuits, needed by the aim to improve the control and to reduce the costs, makes difficult the reliability evaluation. Usually, any modification in the technology or used materials is followed by the appearance of a new failure mechanism. Consequently, any manufacturing modification must include a new reliability evaluation.

3. The last problem is linked to the increasing complexity and costs of integrated circuits vs. discrete devices. Although the cost of a certain electronic function decreases substantially by integration, the basic costs are always higher for an integrated circuit than for a discrete component fulfilling the same function.

The definition of the failure criteria is, unavoidably, very difficult because the complexity of ICs is increasingly higher. Even for a simple device, like a transistor, it is hard to define the failure limits. For an integrated circuit, the basic parameters are more complex and hard to be measured and the degradation of these parameters differs from an utilisation to another.

\footnote{Even if the purpose is to minimise the testing time, a too stronger stress level must not be used, because new failure mechanisms may be induced.}

\footnote{If the time is the accelerated variable, this means that an hour of tests at high stress level produces the same effect on the component reliability as $n$ hours at normal operation time.}
For evaluating the various stresses able to be used in reliability accelerated tests, the following aspects must be taken into account [7][8]:

The stress must be encountered in the operational environment. In principle, one must note that the failure rate of integrated circuits is influenced by the thermal, electrical and mechanical conditions of the operational environment. But for common industrial use, mechanical shock and vibrations have a little influence on the integrated circuits encapsulated in epoxy packages, able to assure the necessary mechanical stability and a good protection. For instance, the acceleration measured at a sudden stop of a running car reaches 40g, for airplanes take-off and landing – up to 5g and for missiles – up to 50g. Compare these values with the acceleration level used for periodic tests: 30,000g. Consequently, mechanical factors will be used only rarely for accelerated tests. On the contrary, the temperature is the most used stress for this kind of tests. The experimentally observed correlation between failure rate and temperature is based on the fact that the speed of chemical reactions arising in the device is thermally increased.

The failure mechanisms must be always those arising in the operational environment.

All samples of integrated circuits used in accelerated tests must behave in the same way at a stress modification: the same circuits should be the first to fail at any stress level.

3. Dynamic life testing

The failure of an IC in operational life is an unpleasant event not only because the owner of the equipment must replace it, but also because this failure may induce serious damages to the equipment, loss of important information or even of human life. Therefore, it is desirable to replace a IC before failure. From economical reasons, this replacement must take place shortly before the anticipated failure. This implies that the lifetime of the IC be accurately estimated. This operation may be done only if laboratory tests simulating as closed as possible the real operational life are performed. In this respect, in the laboratory, not only static, but also dynamic testing must be done. The purpose is to quantify the performance degradation during IC operation. An example of such testing is given by Son and Soma [9].

First, the IC parameters which will be monitored during dynamic life testing are chosen by two criteria: (i) to be measurable at existing pin-outs, and (ii) to predict progressively IC degradation. Than, the typical failure and degradation mechanisms must be studied. In fact, there are two major types of degradation mechanisms: electrical ones (such as: latchup, ESD, hot-carrier effect, dielectric break-down, electromigration, etc.) and environmental ones (produced by thermal and mechanical stress, humidity, etc.). By means of appropriately chosen electrical parameters (such as: static / transient current level change, noise level in current, cut-off frequency, input offset voltage of CMOS differential amplifier, etc.), these mechanisms are monitored during dynamic life testing.

Eventually, aging models for various failure mechanisms must be elaborated. In [9] a model for hot-carrier effect is given. Starting from a widely accepted empirical relationship between parameter deviation and the elapsed stress time for the hot-carrier degradation mechanism, given in [10], an aging curve due to hot-carrier effect under the static or periodically repeated AC stress was obtained [9], defined by the equations:

\[ \Delta V/V_o = k \cdot t^a \]  (1)

\[ k = C \cdot \left[ I_d/w \cdot \exp(-\Phi/(q \cdot p \cdot E_{eh})) \right]^a \]  (2)

where \( C \) is a constant, \( a \) and \( k \) are coefficients of the aging curve (depending on the technology and on IC structure), \( I_d \) is the drain current, \( w \) is the channel width of a MOS transistor, \( \phi \) - the minimum energy required to cause impact ionisation, \( q \) - the electron charge \((1.6 \times 10^{-19} \text{C})\), \( p \) – the hot electron mean free path, \( E_{eh} \) – the channel electric field, \( \Delta V/V_o \) – circuit aging and \( t \) – elapsed stress-time. In a \( \log(\Delta V/V_o) \) vs. \( \log t \) plot, a straight line with slope \( a \) and y-intersection \( \log k \) is obtained (see Fig. 3).

![Fig. 3. A log (\( \Delta V/V_o \)) vs. log t plot for hot-carrier degradation mechanism.](image_url)

From the case study given in [9], one may understand the procedure for IC replacing before to occur a failure by hot-carrier effect. For a 31-stage inverter chain, designed according to MOSIS 0.8 µm HP technology rules, the device operation was simulated, the ageing being modeled by randomly changing device parameters. Based on this model, the probability of survival until the next inspection may be quantified at each inspection of dynamic-life testing. Then, the optimal moment for replacement may be calculated with respect to maintenance cost (the recovery cost of an unanticipated failure and the wasted cost of replacing an IC too early).

4. SCREENING AND BURN-IN

To better understand the role of screening tests for the reliability estimation, it will be given an example
concerning the failure causes. Assume that a printed circuit board (PCB) has 60 integrated circuits (ICs), and the probability of failure for an IC is 2%; it is considered that all the ICs are statistical independent. It results that the probability to find at least one defect IC is $1 - 0.9860^60 = 0.7$. Some reasons can lead to components failures; for example, if the components are very old, or if they are overloaded. In these cases, the screening tests have no sense. Other defects result from the intrinsic weaknesses of the components. These weaknesses are surely unavoidable and - for very good defined limits - are accepted even by the manufacturing. With the aid of electrical tests and/or operating tests (during the fabrication or before the delivering) these components with defects can be identified and eliminated. Nevertheless it remains a small percentage of components with hidden defects which - although still operational - have a small reliability and influence negatively the reliability of the components batch. The role of the screening tests is to identify the components partially unreliable, with defects that do not lead immediately to non-operation.

For at least two reasons it is difficult to define a cost-effective screening sequence, while: (i) it may activate failure mechanisms that would not appear in field operation; (ii) it could introduce damage (transients, electrostatic discharges ESD) which may be the cause of further early failures.

Generally, the selection is a 100% test (or a combination of 100% tests), the stress factors being the temperature, the voltage, etc. followed by a parametric electrical or functional control (performed 100%), with the aim to eliminate the defect items, the marginal items or the items that will probably have early failures (potentially unreliable items). To overcome these problems, recently, a method was proposed [13]. The method was called MOVES, an acronym for Monitoring and Verifying a Screening sequence. MOVES contains five procedures, namely VERDECT, LODRIFT, DISCRIM, POSE and INDRIFT.

By definition, an accelerate test is a trial during which the stress levels applied to the components are superior to these foreseen for operational level; this stress is applied with the aim to shorten the necessary time for the observation of the component behavior at stress.

5. Accelerated tests

The first accelerated tests were made in the early 60’s and tried to shorten the time period necessary to obtain significant results from life tests. The failure mechanisms must be investigated with great care, because it is essential that the failure mechanism acting at the higher stress level to be the same with that acting at normal stress level. Lately, two tendencies were observed: the increase of the stress level and accelerated tests performed early on the manufacturing process (even at the wafer level). The main problems which must be solved in the next years are: (i) the identifying of the accelerated laws with different stress factors; (ii) the taking into account at the design of the accelerated tests of the synergies between the stress factors encountered in the operational environment [14]; and (iii) the dependency of the activation energy on the stress level.

6. Physics of failure

The failure mechanism identification is essential for the reliability accelerated testing, because the obtained degradation laws must be extrapolated beyond the time period of the test and the extrapolation must be made separately for each population affected by a failure mechanism. The subject is still modern, taking into account that new failure mechanisms are discovered and even the old ones are not completely examined. Consequently, a series of tutorial papers on failure mechanisms were published from 1973 by IEEE Transactions on Reliability, almost in each issue. Most of these papers were written by the research team led by Michael Pecht, from Maryland University (USA). Trends in microsystems (integrating electronic, micro-electro-mechanical, electro-optical and micro-fluidic devices) are taking miniaturization close to its physical limits and creating a need for extensive reliability physics effort to identify and counter failure mechanisms in new devices [14].

7. Plastic encapsulated microcircuits (PEM)

The great effort made for reliability evaluation is best by PEM. Also, the current use of these devices is an example of reliability engineering responding to both technology trends and customer policies [5]. The Acquisition Reform policy encouraged U. S. Military to use PEM over other packages. On the technical side, user of PEM are employing Highly Accelerated Stress Testing (HAST) and acoustic microscopy to screen out flawed devices. While the reliability of PEM is constantly improving, the variability between suppliers remains a problem [16].

8. Reliability prediction

In the concurrent engineering era, Reliability Prediction Procedures (RPP) are valuable tools for designers and users of any product. The designer needs RPP to avoid the lag in feedback occurring when the predictions are made by a reliability team. The user, which in turn is a designer of a complex system, wants to have correct information about the part reliability. All the usual RPP (the most known being MIL-HDBK-
217) have some common features which diminish the prediction accuracy:

(i) a constant failure rate model is used;

(ii) the failure mechanisms (FM) are not analysed. Lately, improved RPP - with failure rate models other than the exponential and starting from the physics of failure - were considered desirable and as "a change in direction for reliability engineering" [15]. In 1995, an improved methodology, called SYRP, for predicting the failure rate of a lot was proposed, with the following characteristics: (a) a log-normal distribution for each FM used involved, and (b) the interaction (synergy) between the technological factors, depending on the manufacturing and control techniques, is considered.

9. Conclusion

This overview showed the dynamic in the field of the evaluation of monolithic integrated circuits reliability. Starting from some reliability problems, the main aspects arisen in dynamic life testing, screening and burn-in, accelerated tests, physics of failure of plastic encapsulated microcircuits (PEM) and process reliability were presented.

References


