Semiconductor Device Reliability Failure Models

International SEMATECH
Technology Transfer # 00053955A-XFR

Semiconductor Device Reliability Failure Models

Technology Transfer # **00053955A-XFR**
International SEMATECH
May 31, 2000

Abstract: No abstract.

Keywords: Reliability Modeling, Failure Mechanisms, Electromigration, Corrosion, Stress Migration

Authors: Richard Blish, Noel Durrant

Approvals: Noel Durrant, Program Manager
Dave Anderson, Director, Supplier Relations & Corporate Services
Dan McGowan, Technical Information Transfer Team Leader

SEMICONDUCTOR DEVICE RELIABILITY FAILURE MODELS

Abstract: This review paper brings together the most prevalent and important models for reliability failures in semiconductor devices. It gives a explanation of the failure, the failure model(s) and constraints of the models. Examples are presented showing the application of the models for deriving acceleration factors. Original references are included for each model.

Editors:
Richard Blish, Advanced Micro Devices
Noel Durrant, Intel Corporation

Contributors:

Ted Dellin, Sandia
Don Denton, Texas Instruments
Babak Sabi, Intel Corporation
Dennis Eaton, Agilent
Peng Fang, Advanced Micro Devices
Steve Huber, Intel Corporation
Jeff Kersey, Advanced Micro Devices
Nick Lycoudes, Motorola
Jack McCullen, Intel Corporation
Joe McPherson, Texas Instruments
Tony Oates, Lucent Technologies
Pramod Patel, Advanced Micro Devices
Carl Peridier, Lucent Technologies
George Prokop, IBM
TS Sriram, Compaq

Table of Contents

OBJECTIVES, SCOPE AND EXPLANATION OF SYMBOLS

Objectives

- Provide critical assessment of best known models to predict Acceleration Factors (AF), FITs & DPM for common IC failure mechanisms.
- Provide the SEMATECH Reliability Technology Advisory Board's consensus-accepted method to calculate AF, FITs & DPM corresponding to specific failure mechanisms between reliability stress conditions and different environmental conditions, e. g. office, mobile, or other market segment conditions.
- Provide literature citations to original, relevant source materials.
- Provide numerical examples to illustrate how to apply these models.

Scope

This document covers most commonly observed integrated circuit failure mechanisms; is not intended to be all-inclusive. In the future, additional models will be added to this document to cover additional failure mechanisms. Better models, when available, will be added for the existing failure mechanisms.

This document applies to the following:

1. aluminum alloy metallization (doped with small amounts of Cu and/or Si),
2. refractory metal barriers with thin anti-reflective coatings,
3. doped silica or silicon nitride interlayer dielectrics,
4. poly silicon or "salicide" gates (such as W, Ni & Co salicide to decrease resistivity),
5. thin SiO_2 gate dielectric
6. Si with p-n junction isolation
7. Packages, both wire bonded and C4, particularly fatigue failure and humidity/corrosion effects

Explanation of Symbols, Variables and Abbreviations

TF = time to failure
AF = Acceleration Factor = TF_{cond2}/ TF_{cond1}
A_0, B_0, C_0 = process/material-dependent constant or scaling factors
J = current density
J_{crit} = critical (threshold) current density which must be exceeded before failure will occur
N = current density, substrate current or Relative Humidity exponent
E_a = activation energy
k = Boltzmann's constant = 8.62×10^{-5} eV/ $^{\circ}$K
RH = Relative Humidity as % (100% = saturated)
T = absolute temperature in ^{o}Kelvin
FIT = Failures-in-time expressed as failure per billion device hours
M1 = first metal layer
M2 = second metal layer
<j> = average current density
Q_{crit} = critical charge for soft upset event
e= charge of an single electron

Additional symbols and variables are defined in the paper at the point of usage.

A. ELECTROMIGRATION

Due to momenta exchange between the current-carrying electrons and the host metal lattice, aluminum ions can drift in the direction of the electron current. Due to the presence of flux divergence centers, vacancies start to cluster, clusters grow into voids, and the voids can continue to grow until they block the current flow in the aluminum. Thus, the current is forced to flow through the supporting barrier layer and/or capping layer; the resultant increase in resistance leads to device failure. Since this is a mass conserving process, accumulations of the transported aluminum ions increase the mechanical stress in supporting dielectrics, and may eventually cause fractures and shorts to occur.

Constraints and Limitations for EM in Long Metal Lines

- For Al-alloy stripes, terminated by bonding pads and having no barrier metalization, the total time-to-failure is dominated by nucleation and N is observed to be equal to 2.
- For Al-alloy stripes on barrier metal and terminated by tungsten plugs, one may see both an incubation (nucleation) period, dominated by N=2, and a resistance rise (drift period) dominated by N=1 or a blend of both current exponents to produce an intermediate value. Larger N values can be observed if Joule heating is not properly considered.
- Under high current density test conditions, unaccounted self-heating can produce *apparent* current density exponents *much* greater than 2. Therefore, extreme care must be taken when extrapolating time-to-failure data from high to low current densities.
- The industry standard NIST bow-tie type EM test structure with simple bonding pad connections is grossly inadequate for present multilevel metal systems and generally gives overly optimistic EM results relative to via-fed test structures.
- EM kinetics for both lines & vias are different according to linewidth.
- Via-fed test structures must be carefully designed to avoid resistance saturation & reservoir effects, which can produce misleading t_{50} and σ values.
- Industry standard NIST bow-tie type EM test structures with simple bonding pad connections are inadequate for multilevel metal systems tending to give overly optimistic EM results relative to via-fed test structures.
- For aluminum alloy lines terminated by bond pads and having no barrier metalization, the total time-to-failure is dominated by nucleation and N is observed to be 2.
- Under high current density test conditions, unaccounted for self-heating can produce *apparent* current density exponents *much* greater than 2. Thus, extreme care must be taken when extrapolating time-to-failure data from high to low current densities.

Constraints and Limitations for EM in Vias and Contacts

Electromigration associated with vias and contacts must be investigated separately because they show characteristics unlike single leads fed by bond pads. For example,

- Vias can show different degradation rates depending on electron current flow direction (M2 to M1 versus M1 to M2)
- Via degradation rate is strongly dependent on via structure, number of vias and layout
- Via degradation may have a reservoir effect [1-2].
- Silicide formation and barrier type are extremely important for in contact electromigration
- Silicon (not Aluminum) may be the dominant diffusing species for contact failure [3-6].
- Contact failure occurs by increased barrier metal resistance for an Al-alloy stripe on barrier metal and terminated by tungsten plugs If no refractory barrier is present, failure time is controlled by the rate of drift of Al away from a contact, as junction spiking will be the dominant failure mechanism for transport of Si through the Al grains
- If self-heating is neglected, an apparent current density exponent much greater than 2 can be observed under high current density test conditions. Thus, extreme care must be taken when extrapolating time-to-failure data taken at high density to low current density.

Model

The generally accepted Black model to describe time-to-failure (TF) takes the form [7-13]:

$$TF = A_0 (J - J_{crit})^{-N} \exp (E_a / kT)$$

In this model

- J must be greater than J_{crit} to produce failure,
- failure criterion will be product-dependent, probably defined as the maximum tolerable increase in resistance for the worst-case path

- J_{crit} = critical (threshold) current densityJ_{crit} is inversely related to the Blech length for the line being evaluated, i. e., $J_{crit} * L_{crit} \cong 6000$ A/cm - for Al alloys this is typical (L_{crit} , JL product is approximately constant)
- when test stripe length is ~60 μm, then J_{crit} is comparable to normal EM stressing current densities near 1 MA/cm^2
- in layered metal systems or for those with relatively coarse pitch (linewidth > ~1 μm) N = 2, corresponding to an incubation period, but generally one should use N =1 for for deep sub-micron technologies. It is noteworthy that one may observe an apparent current density exponent of ~2 if J_{crit} is taken as zero.
- E_a = 0.5 – 0.6 eV for Al and Al with a small percent of silicon
- E_a = 0.7 – 0.9 eV for pure aluminum or aluminum/copper alloys

For unipolar current waveforms, J is the average current density < J > [14-15].

For bipolar waveforms a "sweep-back' recovery action takes place and the effective current density J is described by J = (< J+ > - r) < J- >, where < J+ > is the average of the positive polarity pulses and
< J- > is the average of the negative polarity pulses. The recovery factor r has a value of at least 0.7.

Electromigration Numerical Example

Objective: Calculate Acceleration Factor in a mobile environment versus an office environment.
Assumptions:

- Very long Al-Cu metalization stripes with large grain size vs line width (bamboo grain structure)
- Mobile = 80 °C chip temperature inside the laptop
- Office = 50 °C chip temperature inside the enclosure
- Mobile & Office current density of 2.5 & 2.0 X 10^5 A/cm^2 , respectively
- $J >> J_{crit}$
- E_a = 0.8 eV
- N = 2

The ratio of the TF values will be:

$$AF = \text{ratio of TF values, Office/Mobile} = (J_{Office} / J_{Mobile})^{-N} \exp [(E_a / k)(1/ T_{Office} - 1/ T_{Mobile})]$$

$$AF = (2.0/2.5)^{-2} \exp [(0.8 \text{ eV} / 8.62 \times 10^{-5} \text{ eV/°K}) \times (1/\{273 \text{ °K}+50 \text{ °C}\} - 1/\{273 \text{ °K}+80 \text{ °C}\})]$$

$$\text{AF, Mobile/Office} = 1.5 \times 10 = 15$$

So moving from the hot, high current density Mobile environment to the cool, low current density Office environment will increase the TF value to 15X that of the previous value. Current density accounts for a factor of 1.5, while temperature accounts for a factor of 10, giving a 15x overall increase.

References

1. C. Graas, H. Le, J. McPherson and R. Havemann, *IEEE-IRPS Proceedings*, p. 173 (1994).
2. J. McPherson, H. Le and C. Graas, *Microelectron. Reliab., Vol. 37*, p. 1469 (1997).
3. S. Vaidya, et. al., "Electromigration Induced Shallow Junction Leakage with Al/Poly-Si Metallization", *J. Electrochem. Soc., Vol. 130*, p. 496 (1983).
4. S. Vaidya, et. al., "Shallow Junction Cobalt Silicide Contacts with Enhanced Electromigration Resistance", *J. Appl. Phys. Vol. 55*, p. 3514 (1984).
5. S. Steenwyk and E. Kankowski, "Electromigration in Aluminum to Ta-Silicide Contacts", *IEEE-IRPS Proceedings*, p. 30 (1986).
6. J. Ondrusek, C. Dunn and J. McPherson, *IEEE-IRPS Proceedings*, p. 154 (1987).
7. H. Huntington and A. Grone, J*. Phys. Chem. Solids*, Vol. 20, p. 76(1961).
8. J. Black, *IEEE Trans., Vol. ED-16* (No. 4), p. 338 (1969).
9. F. d-Heurle and P. Ho, "*Thin Films: Interdiffusion and Reactions*," John Wiley & Sons, p. 243 (1978).
10. Blech, J*. Appl. Phys., Vol. 74*, p. 1203 (1976).
11. R. Filippi, G. Biery and M. Wood, *Mat. Res. Soc. Symp. Proc., Vol. 309*, p. 141 (1993).
12. R. Filippi, G. Biery and R. Wachnik, *J. Appl. Phys., Vol. 78*, p. 3756 (1995).
13. A. Oates, *IEEE-IRPS Proceedings*, p. 164 (1996).
14. J. Maiz, *IEEE-IRPS Proceedings*, p. 220 (1988).

15. L. Ting, J. May, W. Hunter and J. McPherson, *IEEE-IRPS Proceedings*, p. 311 (1993).

B. CORROSION

ULSI devices with aluminum or aluminum alloys with small percentages of copper and silicon metalization are subject to corrosion failures [1-2]. Corrosion occurs in these metalization in the presence of moisture and contaminants. Corrosion failures are usually classified as one of two broad groups: bonding-pad corrosion or internal corrosion. Bond pad corrosion is more common because the die passivation does not cover the metalization in the bond pad locations. Internal corrosion (internal to the chip, away from the bond pads) is attributed to weakness or damage in the die passivation, permitting moisture to reach the metalization.

Three industry-standard tests are used to accelerate corrosion failure mechanisms

- 85/85 (85 °C and 85% Relative Humidity)
- autoclave (or pressure cooker, usually ~ 2 atmospheres absolute pressure)
- HAST (Highly Accelerated Stress Testing) -typically 85% RH with steam pressure > ambient pressure

Constraints and Limitations

- Industry consensus is that the proper activation energy for aluminum corrosion is in the 0.7-0.8 eV range
- There is not a consensus for the humidity dependence due to lack of data below 85% RH.
- The relevant relative humidity (RH) is a local RH, the RH found at the interface between the silicon chip and the package, which is not necessarily the ambient RH. If power dissipation is small, then ambient RH is approximately the same as local RH. If the power dissipation is large, stress tests may need to employ a duty cycle <100% [3,4] to align local and ambient RH values well enough to get meaningful data.

Models

To extrapolate accelerated corrosion test results to field use-conditions, four models are commonly used, all of which are Eyring models (Eyring makes the practical assumption of mathematically separable, independent variables). The models use the product of an RH term, an applied voltage term & a temperature-based Arrhenius factor. Choice of model is largely personal preference as we generally lack data over a broad enough range of variation in temperature, voltage and RH to discriminate one model from another on the basis of statistical fit. However recent work [5] suggests the Exponential Model may be the most effective. There is industry speculation that a voltage term is needed in these models, but there are not yet supporting data to specify the term.

Reciprocal Exponential Model [6-9]

$$TF = C_0 \exp [b/RH] \; f(V) \exp [E_a /kT]$$

where:

C_0 = arbitrary scale factor,
b = ~300
E_a = 0.3 eV,
f(V) = an unknown function of applied voltage

Power Law (Peck) Model [10],

originally for Al corrosion, but applied to other failure mechanisms with different N & E_a values

$$TF = A_0 \, RH^{-N} \, f(V) \exp[E_a /kT]$$

where,

A_0 = arbitrary scale factor
N = ~2.7,
E_a = 0.7-0.8 eV (appropriate for aluminum corrosion with chlorides are present)
f(V) = an unknown function of applied voltage

Exponential Model [5, 11,12]

$$TF = B_0 \exp [(-a) RH] \, f(V) \exp [E_a /kT],$$

where,

B_0 = arbitrary scale factor,
a = 0.10 – 0.15 per %RH,
E_a = 0.7- 0.8 eV,
f(V) = an unknown function of applied voltage

A fairly recent comparison of four models [5] favored the exponential model with an "a" value of 0.12 – 0.15.

RH^2 (Lawson) Model [13]

$$TF = C_0 \, RH^2 \, f(V) \exp [E_a / kT]$$

where:

C_0 = arbitrary scale factor, (typical value 4.4×10^{-4})
RH = Relative Humidity as % (100% = saturated),
E_a = 0.64 eV,
f(V) = an unknown function of applied voltage

Corrosion Numerical Example

Objective: Calculate the Acceleration Factor for Al corrosion (on bond pads) due to chloride contamination comparing an office environment vs a HAST environment.

Assume

- Office = 50 °C chip temperature & 10% RH inside the enclosure (same absolute humidity as room at 20 °C/50%RH) & applied voltage of 5 V
- HAST environment is 130 °C & 85% RH inside the stress chamber & applied voltage is 6 V
- E_a = 0.75 eV
- Peck RH exponent = –2.7, which was experimentally determined for Al corrosion by Peck et al
- Corrosion rate is linear with applied voltage

Thus, the ratio of the TF values will be:

$$\text{AF, ratio of TF values, office/HAST} = (RH_{office}/RH_{HAST})^{-2.7} \times (V_{office}/V_{HAST}) \times \exp[(E_a/k) \times (1/T_{office} - 1/T_{HAST})]$$

$$AF = (10/85)^{-2.7} \times (6/5) \exp [(0.75 \text{ eV} / 8.62 \times 10^{-5} \text{ eV/°K}) \times (1/\{273 \text{ °K} + 50 \text{ °C}\} - 1/\{273 \text{ °K} + 130 \text{ °C}\})]$$

$$\text{AF, office/HAST} = 323 \times 1.2 \times 210 = 8.1 \times 10^{+4} = 81{,}400$$

This indicates that by moving from a HAST environment to the office environment will increase TF (time to fail) value to 81,000 times the accelerated stress (HAST) value, of which 300X is due to RH, 1.2X due to applied voltage and 250X is due to temperature.

References

1. G. Schnable and R. Keen, *IEEE-IRPS Proceedings*, p. 170 (1969).
2. D. Peck, *IEEE-IRPS Proceedings*, p. 81 (1970).
3. T. Ajiki, M.Sugumoto, H. Higuchi, and S.Kumada, "A New Cyclic Biased THB Test for Power Dissipating ICs, " IEEE International Reliability Physics Symposium Proceedings,, 1979, pp. 118-126.
4. C. G. Shirley and C.E. Hong, "Optimal Acceleration of Cyclic THB Tests for Plastic Packages, "*International Reliability Physics Symposium Proceedings*,, 1991, pp. 12-21.
5. J. McPherson, G. Bishel and J. Ondrusek, Proceedings of Third Intern. Symp. on Corrosion and Reliability of Electronic Materials and Devices, Electrochem. Soc., V. 94-29, p. 270 (1994).
6. J. Gunn, R. Camenga and S. Malik, IEEE-IRPS Proceedings, p. 66 (1983).
7. J. Flood, IEEE-IRPS Proceedings, pp. 95-99, 1972.
8. N. Lycoudes, Solid State Technology, p. 53 (Oct. 1978).
9. K. Stringy and A. Schelling, 31st ECC Proceedings, p. 238 (1981).
10. D. Peck, IEEE-IRPS Proceedings, p. 44 (1986).
11. C. Dunn and J. McPherson, J. Electrochem. Soc., p. 661 (1988).

12. H. Koelmans, IEEE-IRPS Proceedings, p. 168 (1974).
13. R. Lawson, British Telecom. Technol. J., Vol. 2, No. 2, p. 95 (1984).

C. TIME DEPENDENT DIELECTRIC BREAKDOWN

Time-Dependent Dielectric Breakdown (TDDB) is an important failure mechanism in ULSI devices. The dielectric fails when a conductive path forms in the dielectric, shorting the anode and cathode. The two models widely used in describing TDDB are field-driven (E-model) and current-driven (1/E - model).

Constraints and Limitations

- These models are intended for application to SiO_2 with thickness greater than approximately 40A (4 nm).
- Accuracy of these models for thinner (silicon) oxide is unknown.
- Models and parametric values to effectively model "hi K" oxides (based on materials or composites other than silica) are unknown.
- Recently, several low-field/long-term TDDB studies showed that TDDB data were described more effectively by the E-Model than the 1/E model [1-5].
- The E-model is now based on fundamental, physical parameters (not empirically fitted parameters) and fits TDDB data more effectively than the 1/E model [6-8].
- The good fit of the physics-based E-model to the low-field/long-term TDDB data strongly suggests electric field is the dominant degradation driver at low stresses characteristic of customer applications, and that constant current stress is not relevant to customer application.

E-Model [1-14]

In the E-Model, the cause of low-field (< 10MV/cm) TDDB is due to field-enhanced thermal bond-breakage at the silicon-silica interface. The E-field serves to reduce the activation energy required for thermal bond breakage and therefore exponentially increases the reaction rate for failure. Time-to-failure, inverse to reaction rate, decreases exponentially:

$$TF = A_0 \exp[- \gamma E_{ox}] \exp[E_a /kT]$$

where:

A_0 = arbitrary scale factor, dependent upon materials and process

γ = field acceleration parameter, is temperature dependent, $\gamma(T) = a/kT$ where a is the effective dipole moment for the molecule (see Note 1 below)

E_{ox} = electric field across the dielectric in MV/cm. It must be voltage compensated for band bending if an accumulation layer is formed, but no compensation is needed if an inversion layer is formed. E_{ox} is the quotient of the compensated voltage & the oxide thickness, $t_{ox.}$ Note that t_{ox} should be electrically or physically measured.

Previous work [14] provides values for $(\Delta H)_0$ and "a" and shows that γ has a 1/T dependence. This means that the activation energy reduces linearly with the electric field,

$$E_a = (\Delta H)_0 - a \ E_{ox}$$

where:

E_a = effective activation energy (eV). E_a may be nearly temperature-independent if several types of disturbed bonding states are present in the dielectric and the reaction rates are mixed during high-field and/or high-temperature TDDB testing.

$(\Delta H)_0$ = the enthalpy of activation for bond breakage in the absence of external electric field (~2.0 eV)

a = effective molecular dipole-moment for the breaking bonds which value is ~7.2 eÅ .

For intrinsic failures in SiO_2 dielectrics of thickness < 100Å, γ ~ 2.5-3.5 Naperians per MV/cm (~1.1-1.5 decades per MV/cm) and E_a = 0.6-0.9 eV. For extrinsic defects, effective oxide thickness can be quite thin and therefore the

effective field can be very high. This leads to the *apparently* lower activation energy of about 0.3 eV observed during burn-in.

1/E – Model [15-18]

The cause of TDDB, even at low fields, is postulated to be due to current through the dielectric by Fowler-Nordheim (F-N) conduction. F-N injected electrons (from the cathode) cause impact ionization damage of the dielectric as they accelerate through the dielectric. Additionally, when these accelerated electrons reach the anode, hot holes may be produced which can tunnel back into the dielectric causing damage (hot-hole anode injection mechanism). The time-to-failure is expected to show an exponential dependence on the inverse electric field, 1/E:

$$TF = \tau_o(T) \exp [G(T) / E_{ox}]$$

where:

$\tau_o(T)$ = a temperature dependent prefactor, ~ 1 x 10^{-11} sec

G = field acceleration parameter, ~ 350 MV/cm with a weak temperature dependence

E_{ox} = electric field across the dielectric in MV/cm. It must be voltage compensated for band bending if an accumulation layer is formed, but no compensation is needed if an inversion layer is formed. E_{ox} is the quotient of the compensated voltage & the oxide thickness, $t_{ox.}$ Note that t_{ox} should be electrically or physically measured.

TDDB Numerical Example Using the E-model

Objective: Calculate the Acceleration Factor for gate oxide failure by TDDB for an office environment (chip inside an enclosure) compared to a test structure in a highly accelerated environment.

Assume

- Office = 50 °C chip temperature & an electric field of 4 MV/cm inside the enclosure,
- Accelerated means 300 °C & 8 MV/cm
- γ value of 3 Naperians (logarithm) per MV/cm
- Activation energy of 0.75 eV (approximate center of the nominal range for E_a.)

The ratio of the TF values will be:

$$AF, \text{ ratio of TF values, office/accelerated} = \exp[- \gamma(E_{office} - E_{accel}) \times \exp [(E_a /k) \times (1/T_{office} - 1/T_{accel})]$$

$$AF = \exp[-3/(MV/cm) \times \{4 \text{ MV/cm} - 8 \text{ MV/cm}\}] \exp[(0.75eV / 8.62 \times 10^{-5} \text{ eV/°K}) \times \{1/(273 \text{ °K} + 50 \text{ °C}) - 1/(273 \text{ °K} + 300 \text{ °C})\}]$$

$$AF, \text{ office/accelerated stress} = [1.60 \times 10^{+5}] \times [1.27 \times 10^{+5}] = 2 \times 10^{10}$$

So moving from a highly accelerated test structure environment to office environment will increase TF (time to fail) value to 2 x 10^{10} times the accelerated stress value, of which 16,000 X is due to electric field and 12,700 X is due to temperature.

Note 1: Many papers in the literature may use and plot base 10 (rather than base e) when expressing the field acceleration factor. One should be careful to note whether base 10 or natural base e is being used. Some authors, for clarity reasons, will write the field acceleration as *decades* per MV/cm to emphasize that the base 10 is being used or *Naperians* per MV/cm to emphasize that the natural base e is being used. Many authors, however, may not emphasize this distinction to the reader and the reader must be cautious. The conversion factor between base 10 and base e is 2.3:1, i.e., $\gamma_{base\ e} = 2.3 * \gamma_{base\ 10}$. In this document, the natural base e is assumed.

References

1. K. Boyko and D. Gerlach, *IEEE-IRPS Proceedings*, p. 1 (1989).
2. J. Suehle, P. Chaparala, C. Messick, W. Miller and K. Boyko, *IEEE-IRPS Proceedings*, p. 120 (1994).
3. P. Charparala, J. Suehle, C. Messick and M. Roush, *IEEE-IRPS Proceedings*, p. 61 (1996).
4. J. Suehle and P. Chaparala, *IEEE Trans. Elect. Devices*, p. 801 (1997).
5. M. Kimura, *IEEE-IRPS Proceedings*, p. 190 (1997).
6. J. McPherson, V. Reddy and H. Mogul, *Appl. Phys. Lett., Vol. 71*, p. 1101 (1997).
7. J. McPherson and H. Mogul, *Appl. Phys. Lett., Vol. 71*, p. 3721 (1997).
8. J. McPherson and H. Mogul, *IEEE-IRPS Proceedings*, p. 47 (1998).
9. E. Anolick and G. Nelson, *IEEE-IRPS Proceedings*, p. 8 (1979).
10. D. Crook, *IEEE-IRPS Proceedings*, p. 1 (1979).
11. Berman, *IEEE-IRPS Proceedings*, p. 204 (1991).
12. J. McPherson and D. Baglee, *IEEE-IRPS Proceedings*, p. 1 (1985).
13. J. McPherson and D. Baglee, *J. Electrochem. Soc., Vol. 132*, p1903 (1985).
14. J. McPherson, *IEEE-IRPS Proceedings*, p. 12 (1986).
15. Chen, S. Holland and C. Hu, *IEEE-IRPS Proceedings*, p. 24 (1985).
16. J. Lee, I. Chen and C. Hu, *IEEE-IRPS Proceedings*, p. 131 (1988).
17. Moazzami, J. Lee and C. Hu, *IEEE Trans. Elect. Devices, Vol. 36*, p. 2462 (1989).
18. K. Schuegraph and C. Hu, *IEEE-IRPS Proceedings*, p. 7 (1993).

D. HOT CARRIER INJECTION

Hot carrier injection describes the phenomena by which carriers gain sufficient energy to be injected into the gate oxide [1-9]. This occurs as carriers move along the channel of a MOSFET and experience impact ionization near the drain end of the device. The damage can occur at the interface, within the oxide and/or within the sidewall spacer. Interface-state generation and charge trapping induced by this mechanism result in transistor parameter degradation, typically as switching frequency degradation rather than a "hard' functional failure.

Constraints and Limitations

- HCI-induced transistor degradation is well modeled by peak substrate-current for the n-channels and peak gate current for the p-channels, at least for transistors at > 0.25 μm.
- For sub-0.25 μm P-channel, the drive current tends to decrease like NMOS after hot carrier stress.
- For sub-0.25 μm P-channel, worst case lifetime occurs at maximum substrate current stress. The TF model is the same as N-channel.
- The drive-currents for the n-channel transistors tends to decrease after HCI stressing, the p-channel drive current tends to decrease.
- The off-state leakage can increase dramatically [8], especially for initially high current-drive p-channels.
- HCI-induced transistor degradation modeling seems to be accurate, but the extrapolation from transistor degradation to circuit-level degradation is uncertain and should be the focus of future research efforts.
- There is growing evidence that HCI physics may be starting to change at 0.25 μm and smaller, leading changing worst-case stress conditions[5].
- Precise voltage models (rather than substrate current or gate current) would be very useful.
- HCI evaluations are almost always performed on test structures rather than products and done under dc conditions, thus the calculated lifetime should be considered a Figure of Merit for process comparison.
- A short "lifetime" observed with dc test structures does NOT imply unacceptable product performance under ac conditions.
- Typically, HCI degradation causes reduced circuit speed rather than catastrophic failure, although clearly a large enough speed reduction can cause device failure.
- For products where the substrate or gate current is unknown, large voltage acceleration is possible because gate and substrate current are exponential to the reciprocal gate oxide electric field
- There have been reports that the temperature dependent of substrate current has a positive activation energy when V_{cc} is lower than 2.5V. The temperature dependent model for lower V_{cc} is still under investigation.

Models

Generally, degradation induced by HCI can be described by:

$$\Delta p = At^n$$

where:

- p is the parameter of interest (V_t, g_m, I_{dsat}, etc.),
- A = material dependent parameter
- t = time
- n = empirically determined exponent which is a function of stressing voltage, temperature and effective transistor channel length

N-channel Model

N-channel devices use an Eyring model (which makes the practical assumption of mathematically separable and independent variables):

$$TF = B(I_{sub})^{-N} \exp(E_a/kT)$$

where:

B = arbitrary scale factor (function of proprietary factors like doping profile, sidewall spacing, dimensions, etc.)
I_{sub} = peak substrate current during stressing
N = 2 to 4, typically 3
E_a = -0.1 eV to -0.2 eV (note E_a is negative)

P-channel Model

$$TF = B(I_{gate})^{-M} \exp(E_a/kT)$$

where:

B = arbitrary scale factor (function of proprietary factors, such as doping profiles, sidewall spacing dimensions, etc.)
I_{gate} = peak gate current during stressing.
M = 2 to 4
E_a = -0.1eV to -0.2eV (note this is a *negative* activation energy)

A "rough rule-of-thumb", for the substrate current versus voltage dependence of P-channel devices is peak substrate current doubles for each 0.5V increase in source-drain voltage (V_{ds}).

HCI Numerical Example

Objective: Calculate the Acceleration Factor for Hot Carrier Injection of an n-channel device for an office environment versus a test structure accelerated environment.
Assume

- Office = 50 °C chip temperature & substrate current of 1 μA
- Accelerated conditions are -40 °C & 10 μA substrate current
- N = 3
- E_a = -0.15 eV

AF, the ratio of the TF values will be:

$$AF, \text{ ratio of TF values, office/accelerated} = (I_{sub\ office} / I_{sub\ accel})^{-N} \exp[(E_a /k) \times [1/ T_{office} -1/ T_{accel}\}$$

$$AF = (1\mu A /10\ \mu A)^{-3} \exp [(-0.15\ eV /8.62 \times 10^{-5}\ eV/°K) \times\{1/(273\ °K+50\ °C)-1/(273\ °K-40\ °C)\}]$$

$$AF, \text{ office/accelerated stress} = 1 \times 10^{+3} * 8 = 8000$$

So moving from accelerated test structure environment to the office environment will increase TF (time to fail) value to by 8000 X of the accelerated stress value, of which 1000 X is due to substrate current and 8-fold is due to temperature.

References

1. S. Aur, A. Chatterjee and T. Polgreen, *IEEE-IRPS Proceedings*, p. 15 (1988).
2. E. Takeda, R. Izawa, K. Umeda and R. Nagai, *IEEE-IRPS Proceedings*, p. 118 (1991).
3. E. Snyder, D. Cambell, S. Swanson and D. Pierce, *IEEE-IRPS Proceedings*, p. 57 (1993).
4. J. Wang-Ratkovic, R. Lacoe, K. Williams, M. Song, S. Brown and G. Yabiku, *IEEE-IRPS Proceedings*, p. 312 (1997).
5. G. LaRosa, F. Guarin, S. Rauch, A. Acovic, J. Lukaitis and E. Crabbe, *IEEE-IRPS Proceedings*, p. 282 (1997).
6. J.T. Yue, ULSI Technology, McGraw-Hill, p. 657 (1996).
7. P. Fang, J. Yue, and D. Wollesen, "A Method to Project Hot-Carrier-Induced Punch Through voltage Reduction for Deep Submicron LDD PMOS FETs at Room and Elevated Temperatures", *IEEE-IRPS Proceedings*, p. 131 (1982).
8. C. Huang, T. Rost and J. McPherson, *IEEE-IRW Final Report*, p. 63 (1994).

E. SURFACE INVERSION (Mobile Ions)

Mobile ions can contaminate semiconductor processing materials. In SiO_2 they are very mobile even in the presence of modest electric fields (~ 0.5 MV/cm) and modest temperatures (100°C). An accumulation of drifted ions at the Si/SiO_2 interface causes surface inversion leading to device failure [1-6].

Sodium and potassium are the usual suspects, simply because of high mobility and their natural abundance. In a gate structure under bias, they can drift from the polysilicon (anode) to the silicon substrate (cathode). A buildup of positive ions at the Si/SiO_2 interface inverts the surface and severely degrades the oxide isolation. Ionic drift in SiO_2 gate dielectric can also cause premature TDDB. In the case of EPROMs, mobile-ion accumulation around the negatively-charged floating polysilicon can lead to data retention fails.

Devices showing isolation-leakage failures may recover during an unbiased high temperature bake. The bake causes a redistribution of the mobile-ions away from the accumulated Si/SiO2 interface or floating polysilicon in an EPROM device.

Constraints and Limitations

The activation energy for mobile-ion diffusion depends on several factors including

- the diffusing species,
- the medium through which the mobile ions diffuse
- the concentration of the contaminant
  - For low concentrations, and if deep interfacial traps exist, then one may see deep interfacial-trap dominated activation energy of ~ 1.8 eV for Na^+ diffusion through SiO_2.
  - If the concentration of Na^+ is high, then all interfacial traps may be filled and a lower activation energy will be observed (E_a ~ 0.75 eV for simple Na^+ drift through SiO_2).

Model

Mobile ions are influenced by both electric field and temperature. One can use Eyring model for temperature and one would compute the product of inverse to the ionic flux & an Arrhenius factor the time-to-failure (TF) described by:

$$TF = A\,(J_{ion})^{-1}\exp(E_a/kT)$$

A = material constant,

$J_{ion} = \langle(eD_0\rho E/kT) - (D_0\partial\rho(x,t)/\partial x)\rangle$ is the average flow of ions

- $(eD_0\rho E/kT)$ is the *drift* current
- $(D_0\partial\rho(x,t)/\partial x$ is the *back-diffusion* component
- E = electric field
- e = charge on the electron
- D_0 = diffusion coefficient
- ρ = density of mobile ions
- <> represents the time- averaged value of the time-dependent quantities enclosed,

E_a = activation energy

- depends upon the medium through which the ion must diffuse
- for Na^+ it ranges from 0.75 to 1.8 eV, with 1.0 eV being typical (0.75 eV for bulk silica)

Mobile Ion Numerical Example

Objective: Calculate the Acceleration Factor for surface inversion for an office environment compared to an accelerated stress environment.

Assume:

- Baseline or office conditions are
  - $[Na^+] = 10^{10}/cm^3$
  - electric field = 5 MV/cm
  - Office = 50 °C chip temperature
- Accelerated stress is
  - Temperature = 250 °C
  - $[Na^+] = 10^{17}/cm^3$
  - electric field = 8 MV/cm in the dielectric under stress
  - same distance & gradient in both cases
  - E_a = 0.75 eV
-

First consider whether the drift or diffusion term is dominant for flux. Since D_0 & concentration factors the same for both drift and diffusion, the ratio [drift/diffusion] becomes (eE/kT)/(1/typical distance). Plugging in some nominal values (electric field = 6.5 MV/cm, temperature = 150 °C , characteristic distance = 10^{-4} cm) then the unit-less ratio is ~10^{-7}. This means the diffusion flux is far more important than the drift flux for this particular case and therefore the electric field will not be a factor in calculating an acceleration factor

The ratio of the TF values will be:

$$AF \text{ (ratio of TF values, office/accelerated stress)} = (J_{ion\ office}/J_{ion\ accel})^{-1} \exp([E_a/k](1/T_{office} - 1/T_{accel}]$$

$$AF = (10^{10}\ /cm^{-3}/10^{17}\ /cm^{-3})^{-1} \exp([0.75\ eV / 8.62 \times 10^{-5}\ eV/°K](1/(273\ °K+50\ °C)-1/(273\ °K+250\ °C))$$

$$AF(accelerated/office) = 1 \times 10^{+7} * 3.2 \times 10^{+4} = 3.2 \times 10^{+11}$$

Moving from accelerated stress environment (deliberate contamination and high electric field) to the office environment will increase the TF value to $3.2 \times 10^{+11}$ times the accelerated stress value. A factor of 1×10^{7} is due to flux (concentration) while a factor of 32,000 is due to temperature.

References

1. J. McPherson, "Accelerated Testing," Electronic Materials Handbook, Volume 1, Packaging," *ASM International Publishing*, p. 887 (1989).
2. E. Schlegel, G. Schnable, R. Schwartz and J. Spratt, "Behavior of Surface Ions on Semiconductor Devices," IEEE Transactions on Electron Devices, *Vol. ED-15,* 1968, pp. 973-979.
3. E.H. Snow, A.S. Grove, B.E. Deal, and C.T. Sah, "Ion Transport Phenomenon in Insulating Films, "*J. Appl. Phys. Vol 36*, p. 1664 (1965).
4. E.H. Snow and B.E. Deal, "Polarization Phenomena and Other Properties of Phosphosilicate Glass Films on Silicon," *J. Electrochem. Soc., Vol 113*, p. 263 (1966)
5. P. L. Hefley and J. McPherson, "The Impact of an External Sodium diffusion Source on the Reliability of MOS Circuitry *IEEE-IRPS Proceedings*, p. 167 (1988).
6. D.A. Stuart, "Calculations of Activation Energy of Ionic Conductivity in Silica Glass by Classical Methods, "*J. Amer. Ceramic Soc*, " p. 573 (1954).

F. STRESS MIGRATION

The term stress migration describes the movement of metal atoms under the influence of mechanical-stress gradients. Generally, stress gradients can be assumed to be proportional to the applied mechanical stress. Little occurs until the stress exceeds the yield-point of the metalization. Then flux divergence associated with the metal movement causes voids in ULSI metal leads. The resistance rise associated with the void formation can cause electrical failures [1-6].

The role of stress and stress relaxation is very important in the nucleation and growth of voids in aluminum-alloy interconnects. Cu doping in the aluminum is somewhat effective in suppressing grain-boundary diffusion, but is much less effective if the grain size is large compared to linewidth, i. e. bamboo leads, as one observes slit-like void formation due to intra-grain diffusion.

Constraints and Limitations

- This model applies to Aluminum alloys (doped with Cu and/or Si) only, there is insufficient data on pure copper at this time.
- There is currently no standard industry test for SM. Typically, long (> 1000 μm) and narrow (< 2 μm width) stripes are stored (unbiased) at temperatures of 150-250^{o}C for 1-2 K hr and then electrically tested for resistance increases or reduction in breakdown currents. Electromigration stressing is optional after SM baking.
- The SM baking temperature should be carefully selected because the maximum creep rate is generally in the range between 150-250^{o}C. The maximum in the creep rate occurs due to the high stress but low mobility at low temperatures, and low stress but high mobility at high temperatures.
- Because the mechanical stress is temperature dependent, a straightforward determination of the diffusional activation energy is difficult to obtain. Generally E_a~0.5-0.6 eV is used for grain-boundary diffusion and ~ 1 eV for single-grain (bamboo-like) diffusion.
- The use of refractory metal barriers or layered metallization tends to nullify the severe damage caused by slit-like void formation in bamboo leads.
- Refractory metal layer acts as a redundant conductor, shunting the current and reducing the electrical resistance rise due to the void formation.

Mechanical Stress Model

The time-to-failure (TF) can use an Eyring model, for which one computes the product of a power law on mechanical stress & an Arrhenius factor [3]:

$$TF = A_o\ (\sigma)^{-n}\ \exp(E_a/kT)$$

where:

σ = constant stress load,
n = 2-3 for ductile metals,
n is usually ~5 if creep, thus implies $T < T_m/2$)
E_a= 0.5 - 0.6eV for grain boundary diffusion, ~ 1 eV for single grain (bamboo-like) diffusion,

Thermomechanical Stress Model

If the stress is generated by differing thermal expansion rates, then the stress is called as “thermomechanical stress” and is proportional to the change in temperature -

$$\sigma \propto (\Delta T)$$

Therefore the time-to-failure can be written [3],

$$TF = B_0\ (T_0 - T)^{-n}\ \exp(E_a/kT)$$

where:

T_0 = stress free temperature for metal (approximate metal deposition temperature for aluminum)
n = 2 – 3, (~5 if creep, thus implies $T < T_m/2$)
E_a = 0.5 - 0.6eV for grain-boundary diffusion, ~ 1 eV for intra-grain diffusion

Stress Migration Numerical Example

Calculate the Acceleration Factor for stress migration in an office environment compared to an accelerated environment.

Assume

- Office = 50 °C chip temperature inside the enclosure,
- Accelerated stress temperature = 150 °C.
- E_a = 0.55 eV,
- n = 2.5
- $T_{0 =}$ 250 °C
- Use the thermomechanical stress model $TF = B_0 (T_0 - T)^{-n} \exp(E_a/kT)$

The ratio of the TF values will be:

$$AF \text{ (ratio of TF values, office/accelerated)} = ((T_0 - T_{office})/(T_0 - T_{accel}))^{-n} \exp([E_a/k](1/T_{office} - 1/T_{accel}))$$

$$AF = ((300-50)/(300-150))^{-2.5} \exp([0.55 \text{ eV}/8.62 \times 10^{-5} \text{ eV/°K}](1/(273 \text{ °K}+50 \text{ °C}) - 1/(273 \text{ °K}+150 \text{ °C}))$$

$$AF(\text{office/mobile}) = 0.28 \times 105 = 30X$$

So moving from the accelerated environment to the Office environment will increase the TF value to 30-fold that of the accelerated stress value. Mechanical stress changes the TF value by 0.28 x (farther from stress-free temperature), while temperature TF by 105X the accelerated stress value.

References

1. J. Klema, R. Pyle and E. Domangue, *IEEE-IRPS Proceedings*, p. 1 (1984).
2. J. Yue, W. Fusten and R. Taylor, *IEEE-IRPS Proceedings*, p. 126 (1985).
3. J. McPherson and C. Dunn, J*. Vac. Soc. Technology B*, p. 1321 (1987).
4. J. Yue, Reliability Chapter, "ULSI Technology," McGraw-Hill, p. 674, (1996)
5. J. McPherson, "Accelerated Testing," Electronic Materials Handbook, Volume 1 Packaging," *ASM International Publishing*, p887 (1989).

G. TEMPERATURE CYCLING & THERMAL SHOCK

Fatigue failures can occur in ULSI devices due to temperature cycling and thermal shock. Permanent damage accumulates during thermal cycling, accumulating each time the device undergoes a normal power-up and power-down cycle. Such cycles can induce a cyclical stress that tends to weaken materials [1- 9], and may cause a number of different types of failures, including

- Dielectric/thin-film cracking
- Lifted bonds
- Fractured/broken bond wires
- Solder fatigue (joint/bump/ball)
- Cracked die
- Lifted die

Constraints and Limitations

- "Linearity" is assumed, modeling parameters are constant over range of interest (stress vs customer application)
- Alternative methods are needed for reliability estimates (AF or FITs) under certain conditions such as
  - if the temperature cycle range crosses a critical temperature, e.g. T_g (glass transition temperature of polymer)
  - if a material property changes dramatically over the temperature range of interest. For example the stress relaxation rate of lead-based solders changes substantially near room temperature (see note below).

Coffin-Manson Model [1-2]

For ductile materials, low-cycle fatigue data are described by the Coffin-Manson equation:

$$N_f = A_0 [1/\Delta\varepsilon_p]^B$$

where:

N_f = cycles to failure
A_0 = a material constant
$\Delta\varepsilon_P$ = plastic strain range (irreversible strain for one hysteresis loop, unitless)
B = an empirically determined constant

Low cycle fatigue is defined as a stress condition in which hundreds or thousands of cycles cause failure, while high cycle fatigue would require millions of cycles. The Coffin-Manson model was originally developed for ductile materials (iron and aluminum alloys for aircraft), but has been successfully applied to brittle materials.

Modified Coffin-Manson Model [3]

The standard Coffin-Manson equation works well, even for brittle material failures, where failure is dominated by crack initiation and growth rather than simple plastic deformation. However, during a temperature cycle, not all of the temperature range ΔT may be inducing plastic deformation. If a portion of the cycle, ΔT_0, is actually in the elastic range, then this should be subtracted from the total plastic strain range.

$$\Delta\varepsilon_P \propto (\Delta T - \Delta T_0)^{\beta}$$

Therefore, for temperature cycling with plastic deformation, the Coffin-Manson equation becomes:

$$N_f = C_0 [\Delta T - \Delta T_0]^{-q} \qquad \{1\}$$

where :

N_f = Number of cycles to failure
C_0 = a material dependent constant
ΔT = entire temperature cycle-range for the device
ΔT_0 = the portion of the temperature cycle range in the elastic region
q = the Coffin-Manson exponent, an empirically derived constant

If the elastic range (ΔT_0) is much smaller than the entire temperature cycle range (ΔT), then it may be dropped without significant error being introduced (usual practice). Most failure mechanisms will show the same failure rate for slow temperature cycling and for rapid thermal shock, but solder fatigue is a significant exception. Low melting point solders

are used at temperatures in excess of $T_{melting}/2$, so creep is significant & the mechanical properties are extremely time-sensitive. Conversely, strain rate for brittle materials (those with large q values) seems to be largely irrelevant.

Values for q for common ULSI materials are [4]:

	q value
ductile metal, e.g., solder	1-3
hard metal alloys / intermetallics (e.g. Al-Au)	3-5
brittle fracture (e.g. Si & dielectrics : SiO_2, Si_3N_4)	6-9

Temperature Cycling Numerical Example

Objective: Calculate the Acceleration Factor for wire bond intermetallics due to temperature cycling in an office environment compared to an automotive environment.

Assume:

- Office = once daily temperature changes of 20 °C
- Automotive = 4 cycles per day with a change of 80 °C
- q= 4
- $\Delta T >> \Delta T_0$
- Use $N_f = C_0 [\Delta T - \Delta T0]^{-q}$

The ratio of the TF values will be:

$$AF \text{ (ratio of } N_f \text{ values per stress cycle, office/automotive)} = (\Delta T_{office}/\Delta T_{auto})^{-4}$$

$$AF = (20/80)^{-4}$$

AF(office/automotive per cycle) = 256 per cycle, but more cycles/day for automotive also

Comparative evaluation of an automotive environment to office environment shows a time to fail value 1000-fold that of the automotive value, of which 256-fold is due to temperature excursion difference, while 4X is from cyclic frequency.

If a critical temperature is crossed during each temperature cycle, at least four different approaches can be used.

1 AF and FITs can be calculated for each zone where the modeling parameters are constant. Using available data [4,7,8], the Coffin Manson exponent can be allowed to vary between the high and low temperature zones.
2 The most commonly used method utilizes a Coffin-Manson exponent weighted over the temperature range.
3 An alternative is to use the Norris-Lanzberg approach, which adds additional multiplicative terms. Factors are a cyclic frequency factor as a power law (typ. with exponent ~1/3) and a mild Arrhenius-like temperature dependence (typically using an "activation energy" of ~0.01 eV with highest vs lowest cyclic temperature.
4 Perhaps the best method, but the most labor-intensive, is to use Finite Element Analysis (FEA) to derive the von Mises or maximum principal stresses to feed the simple Coffin-Manson equation {1}.

References

1. L. Coffin, Jr., *Met. Eng. Q., Vol 3*, p. 15 (1963).
2. S. Manson, "Thermal Stress and Low-Cycle Fatigue**,"** *McGraw-Hill*, New York, (1966).
3. C.F. Dunn and J.W. McPherson, "Temperature Cycling Acceleration Factors in VLSI Applications," *IEEE-IRPS Proceedings*., p. 252 (1990).
4. R. C. Blish, II "Temperature Cycling and Thermal Shock Failure Rate Modeling *IEEE-IRPS Proceedings*, p 110 (1997).
5. H. Caruso and A. Dasgupta, "A Fundamental Overview of Accelerated-Testing Analytic Models," *1998 Proc. of Annual Rel. and Maintainability Symp*.," pp 389-393.
6. J. McPherson, "Accelerated Testing," Electronic Materials Handbook, Volume 1 Packaging," *ASM International Publishing*, p. 887 (1989).
7. C. Y. Li, R. Subrahamanyan, J. R. Wilcox and D. Stone, "Damage Integral Methodology for Thermal and Mechanical Fatigue of Solder Joints," Solder Joint Reliability, Theory and Applications, ed. John H. Lau, 1991, pp. 261-288.
8. P. M. Hall, "Creep and Stress Relaxation in Solder Joints," Solder Joint Reliability, Theory and Applications, ed. John H. Lau, 1991, pp. 306-332.

9. K. C. Norris and A. H. Landzberg, “Reliability of Controlled Collapse Interconnections,” IBM J. R&D Vol. 13, 1969, pp. 266-271.

H. SOFT ERRORS (SINGLE EVENT UPSETS)

Soft Errors (also called Single Event Upsets)

Soft errors in memory devices were first mentioned by May and Woods in 1978 [1]. A 5 MeV alpha particle impinging on a memory cell will produce 1.4 million electron-hole pairs in silicon to a depth of 25 μm. Minority carriers are collected by the voltage across the depletion layer. If the number of collected carriers is greater than the Q_{crit} (critical charge -- charge necessary to determine cell state), the cell state will change resulting in a "soft" (reversible) error. A tutorial [2] in the IEEE Transactions on Reliability provides an excellent introduction to the subject.

To accurately determine the sensitivity of integrated circuits to radiation-induced soft errors or single event upsets (SEU), all sources of radiation must be accounted for. At terrestrial, Earth-based altitudes, the predominant sources of radiation include both cosmic radiation and alpha particle radiation from radioisotopic impurities in the package and chip materials. An overall assessment of a device's sensitivity to SEU/SER is complete ONLY when the alpha particle component AND the cosmic radiation component have been considered.

Constraints and Limitations

- The Q_{crit} value of memory devices of the 1970s, 1980 and early 1990s has changed from approximately 1 million electrons (~'70) to ~10K electrons (~'99))
- For alpha particle tests the test source should match the expected environment. For package impurities (ceramics, glass, epoxy filler) the isotope impurities are Th^{232} and U^{238}, so the alpha source should be Th^{232}. For flip-chip and C4, the impurities are Po^{210} and Pb^{210}. The preferred alpha source here is Am^{241}. The alpha particle source should always be MUCH larger than the sample to assure all alpha impinging angles are possible.
- For cosmic ray evaluation (terrestrial), neutron or proton sources of five to several hundred MeV should be used. These sources may be found at government, academic and medical facilities.

Cosmic rays (mainly high energy neutrons) at sea level are typically of little concern. CPUs in desktop computers run at a high enough V_{cc} that the Q_{crit} value is rarely exceeded by cosmic ray induced ionization. Cosmic ray intensity is 10 times worse at 2 Km above sea level and reaches a peak of approximately 300 times sea level at 10 Km [3]. Laptop computers are at risk in use durning flights on commercial airlines, which typically fly at ~10 Km. The hazard to devices in laptop computers is further exacerbated by the smaller Q_{crit} value resulting from a low V_{cc} voltage used to reduce power consumption.

The effects of alpha particles can be ameliorated by shielding and purity, but there is little defense against cosmic rays, short of architectural change (redundant bits & Error Correction Code, ECC). There is a fine review by Ziegler et al encompassing entire January 1996 IBM Journal of Research and Development [3]. Further discussion of the effects of cosmic ray neutrons and measurement techniques is contained in an article by McKee et al [6].

Evaluation for alpha particles has been done historically using Th^{232} or Am^{241} sources. A test method is described in a MIL-STD-883 procedure [5], but this will be supplanted by a new spec from JEDEC Committee 13.4 (letter ballot planned q1 2000, reference number TBD). Cosmic ray evaluation has been performed at high elevation locations. For example, Leadville, Colorado cosmic ray intensity is approximately 13 times that of New York City [4]. Cosmic ray effects can be simulated by an appropriate neutron source. Software programs are available for design simulations, but their accuracy is not yet proven.

Model

The basic relation for soft error failure rate (SER) is:

$$SER = (OER/AF) \times 10^9 \text{ FITS}$$

Where:

OER is the observed error rate (errors/hour) on the test sample using an alpha source,
AF is the acceleration factor for the setup.

A Curie is 3.7×10^{10} disintegrations per second, thus the source alpha particle emission is 0.1×10^{-6} Ci * 3.7×10^{10} or 3700 alpha/sec. The source area is 0.32^2 π/4 or 0.08 cm^2. Thus, the total source intensity is 3700/0.8 or 46,600 α/sec/cm^2. The ambient alpha intensity is 0.001/3600 or 2.78×10^{-7} α/sec/cm^2. The acceleration factor is the test alpha flux

divided by the ambient alpha flux or in this example 46,600/2.78 x 10^{-7} or 1.68 x 10^{11}. The SER of the device as normally packaged would then be (1,000/1.68 * 10^{11}) x 10^{9} or approximately 6 FITs.

Numerical Example

An alpha particle test might use a 0.1 μCi Am^{241} source to test a device typically packaged in a plastic package. A SER (Soft Error Rate) of 1,000 errors per hour is observed. Common molding compounds have a measured alpha emission rate of 0.001 α/hour/cm^2 and no die coating is used. The Am^{241} source is 1/8" (0.32 cm) in diameter. What would be the Soft Error Rate in FITs of the device as normally packaged? Warning … this numerical example illustrates Mil 883 Method 1032 rather than the new JEDEC spec.

References

1. May, Timothy C. and Woods, Murray H., "A New Physical Mechanism for Sift Errors in Dynamic memories," *IEEE-IRPS Proceedings*, pp 33-40 (1978).
2. Lantz, L., "Tutorial: Soft Errors Induced by Alpha Particles," *IEEE Transactions on Reliability*, Vol. 45, No. 2, June 1996.
3. Ziegler, J.F. et al, "IBM Experiments in Soft Fails in Computer Electronics (1978 -- 1994)," *IBM J. of Res. Develop*. No. 1, January 1996. Entire Issue is devoted to soft errors and associated topics.
4. Ziegler, J.F, "Terrestrial Cosmic Ray Intensities," *IBM J. of Res. Develop*., Vol. 42, No. 1, January 1998.
5. MIL-STD-883 -- Test Methods and Procedures for Microelectronics, Method 1032, Package Induced Soft Error Test Procedure (Due to Alpha Particles).
6. McKee, M.C., et al, "Cosmic Ray Neutron Induced Upsets as a Major Contributor to the Soft Error Rate of Current and Future generation DRAMs," *IEEE-IRPS Proceedings*, pp 1-6 (1996).

I.DESIGN OF EXPERIMENTS FOR DETERMINATION OF MODELING PARAMETERS

If the modeling parameters for a particular failure mechanism are not available in the literature or from internal sources or the parameters are uncertain, a designed experiment is needed to measure the appropriate constants. It is essential that a range of "over-stresses" be used to produce a statistically significant number of failures. Overstress conditions can be used to define technology margin and acceleration factors.

For example, assume that a failure mechanism is thermally activated, and follows a conventional Arrhenius model i.e.

$$\ln(\text{rate}) \propto (Q/k)*(1/T),$$

where Q = activation energy, k = Boltzmann's constant and T = Absolute Temperature.

Plotting the data as ln(rate) versus the reciprocal temperature will produce a straight only if a single mechanism is present. In order to get a reasonably accurate Q value, the 1/T values should be far enough apart to modulate the rate significantly (at least several-fold, preferably as modulated over as many orders of magnitude as you can manage).

How much difference in temperature stress is needed? If the expected Q ~0.7 eV and the relevant temperature is near 100 °C, every 20 °C change in temperature will change the rate by ~3-fold. So, ±20 °C will modulate the rate by ±3-fold. Similarly, every 40 °C change in temperature will change the rate by 9-fold with corresponds to ±40 °C modulating the rate by ±9-fold.

As a second example, assume the failure mechanism is some sort of mechanical fatigue, for which the conventional model is Coffin-Manson. The rate (or time to failure) is proportional to $\text{strain}^{\text{power}}$ and the usual surrogate for strain is the swing in temperature (ΔT) between hot and cold in temperature cycling or thermal shock.

Plotting the data as log(rate) vs log(ΔT) (or perform a power law regression fit) , the slope is the power (Coffin-Manson exponent). How much temperature swing will is needed? Let's say you expect a Coffin-Manson exponent of 4 and the relevant temperature swing (ΔT) is 100 °C. If your experiment were to explore ΔT values half as large and twice as large as the nominal 100 °C, the failure rate would be modulated by 2^4 = 16-fold, which would provide enough change to determine the Coffin-Manson exponent more accurately than your initial guess. As above in the Arrhenius example, you want to perturb the rate by as much as possible without changing to a different failure mechanism (verified by failure analysis).

As a third example, let's consider that your failure mechanism is electrically activated, perhaps something similar to TDDB, for which the conventional model is the "E-model." Thus, ln(rate) is proportional to the difference in electric field ($\Delta V/t_{ox}$). Plotting the data as ln(rate) vs $\Delta V/t_{ox}$ will produce a straight line (or one could perform an exponential regression fit), for which the slope is the γ value. How much modulation in $\Delta V/t_{ox}$ will you need to modulate the failure rate significantly? Let's say you expect a γ value of 1 decade per MV/cm of electric field. Thus, changes in electric field of ±1 MV/cm will modulate the failure rate by an order of magnitude. You might also want to modulate t_{ox} to see if the relevant stress is electric field or voltage. As above in the other examples, you want to perturb the rate by as much as possible without changing to a different failure mechanism.

As a fourth and final example, let's consider that your failure mechanism responds to two different stresses. For example, TDDB is sensitive to both temperature and voltage. The key design strategy is to assume that an Eyring model will work adequately (separable stresses). You now construct a "space" (temperature is one axis and voltage is the other) and make sure that your stress conditions explore a constrained area of that space, using factorial design. Your initial DOE (Design Of Experiments) might employ a 2x2 or 3x3 "box" or perhaps a box/star. A particularly good reference for all DOE, factorial design and statistics issues is "Statistics for Experimenters" by Box, Hunter & Hunter (Wiley).

J. TIME-TO-FAILURE DISTRIBUTIONS

Models for the several failure mechanisms are well suited for calculation of acceleration factors between one stress condition and another. The "hidden" assumptions that an acceleration factor is valid:

- each distribution is a single population AND
- the dispersion in failure times is the same for both stress conditions AND
- the same cumulative percentage failures are to be compared
- A "Stress to Fail" method is used to characterize and qualify the process or product, producing a statistically significant number of valid failures from which the failure mechanism and root cause can be determined

Several different distributions can be used to model failure rate under **appropriate** circumstances:

An exponential distribution implies a constant failure rate, which is generally not true if the product is insufficiently screened, improperly Designed for Reliability or if the product is past the bottom of the "bathtub," into wearout phase. The exponential distribution is simple, a good approximation and works well if there are very few or zero failures.

A Weibull distribution can be used to model "weakest link" and has two variants. The two parameter version derives a characteristic life for 63% fail and a shape parameter, usually called β. The three parameter version for Weibull retains t_{63} and β, but adds an "delay time" corresponding to the time required to initiate defects, which may then propagate until they cause failure. The solder fatigue & some other communities traditionally use Weibull, but most of the semiconductor community uses the lognormal distribution. There is a simple relationship to relate Weibull shape parameter (β) to lognormal σ ($\beta = 1.2/\sigma$) over a wide range of cumulative percent failing (~100 ppm to 99.99%) [1, but be advised that Wager's tutorial contains a ln(10) error, see below]. Since substantially all real data sets show identical correlation coefficient whether plotted as lognormal or 2 parameter Weibull, one can say that the choice is personal preference. However, lognormal is the preferred distribution for most failure mechanisms (except solder fatigue, use Weibull instead) on the basis of mathematical simplicity, physical basis, intuitive basis (Gaussian, etc.) and ease of use. Excel, or equivalent, is a powerful tool for plotting and curve-fitting either lognormal or Weibull. On the other hand, the Weibull distribution is mathematically complicated (nested logarithms), non-physical & not Gaussian.

The lognormal distribution is based on a normal distribution of failures vs logarithm of time. The characteristic fitting parameters are the time to 50% cumulative failure (t_{50}) and sigma, a measure of the time dispersion of the failures ($\sigma = \ln(t_{50}/t_{16})$). The preferred, graphical way to extract parametric data is to use Excel to plot NORMSINV(cume % fail) as abscissa vs. natural logarithm of stress time as ordinate. Then a least squares fit of a straight line (in Excel: Chart, Add Trendline, Options, Show Equation & Correlation Coefficient) will provide slope = σ and intercept = $\ln(t_{50})$, Providing a single distribution is present. The lognormal distribution is particularly well suited to changing failure rate, especially if a large number of failures are available. The existence of many failures is NOT bad news from a modeling perspective (actually necessary!). In fact, large number of failures is also preferable for determination of the product's Achille's Heel, such that Continuous Improvement, successful Failure Analysis and Corrective Action(s) can be implemented.

A lognormal plot is a very effective way to see if single or multiple populations are present.
Multiple populations are indicated if

- a substantial curvature to the plotted data (poor correlation coefficient), or
- a substantial change in slope, or
- an offset on the time axis (an arrest in cume % fail over some considerable time period), or
- subsequent failure analysis shows different failure modes/mechanisms.

ONLY if a single population is indicated, do the t_{50} and σ values have any physical significance. If multiple populations are present, one must parse the data such that each distribution is a single "pure" mechanism, which will show superior correlation coefficient (minimal divergence at the ends of the distribution) and no abrupt change in slope or offset.

A common example for multiple populations is that of the "defective subpopulation" [2]. A proportion of the total sample may contain defects and the remaining product is immortal for the failure mechanism in question. The usual indication for this situation is a large sigma value. Typical sigma values are: ~0.5 for wearout; ~1.5 for defects; > ~4 for improperly analyzed defective subpopulations. The key to successful analysis of a defective subpopulation is to make an estimate for the asymptotic value for complete failure of all defect-ridden product. An effective method to establish the asymptotic value is to plot cume percent failing vs time or vs log time. The asymptotic value for cume % is the size

of the defective subpopulation. To extract valid t_{50} & sigma values, one must recompute the cume percent failing with the sample size for the defective subpopulation ONLY. Then proceed with a lognormal plot as described above.

Tips for interpretation of lognormal plots. The abscissa (cume percent failing) is non-linear in cume %, but is linear in units of sigma. Thus, t_{50} corresponds to sigma = 0 (at the mean failure time), while t_{16} corresponds to sigma = -1 (one unit below the mean), $t_{2.3}$ corresponds to sigma = -2 (two units below the mean), $t_{0.1}$ corresponds to sigma = -3, (three units below the mean), etc. The ordinate is ln (time), but most people find it difficult to convert ln (time) values back to time values, whereas it is much easier to think in base 10. Accordingly, it is convenient to pick the scale factor along the ordinate (ln(time)) in increments of 2.3. Thus, each time increment is an order of magnitude, as ln(10) = 2.302.

There are two commonly used ways to quote failure rate. FITs (defined as number of Failures In Time per billion hours, American 10^9 hr, not British or European 10^{12} hr) and DPM (Defects Per Million over some stated time period). FITs are a rate, while DPM is an integral of rate over time. For example, if your data suggest constant failure rate (very few failures) & you had 100 FITs (100 /10^9 hr) for 10K hr (~1 yr), then the DPM is simply the product of FITs & time, producing 10^{-3} in this example, 1000 DPM. If the failure rate is known to be changing with time, having been modeled by the lognormal or Weibull distribution, then the FIT value is the instantaneous slope of cum % fail vs time.

It is important to quote FITs or DPM to a known Upper Confidence Level, invariant with number failing and with 60% as the commonly used value. However, some users specify 90% or 95% UCL. The chi-squared distribution, supported well by Excel, is an effective way to "level the playing field."

Numerical Example

Let's assume your objective is to calculate when 0.1% of the product will fail based on knowledge of the mean time to fail and some measure of the time dispersion (constant failure rate, lognormal sigma or Weibull beta). Let's further assume that your failure mechanism has been established to be well fitted by the lognormal distribution with a t_{50} = 100 years and a sigma of 0.7 (natural logarithm of 2 to illustrate the point). Since sigma = $\ln(t_{50}/t_{16})$, we know that every 2-fold shorter time will move the cum % fail down the normal curve by one unit (also called sigma, but a different quantity). Thus, 16% cum fail (-1 sigma) will occur at 50 years; 2.3% cum fail (-2 sigma) will occur at 25 years; 0.13% cum fail (-3 sigma) will occur at 12.5 years. This scenario (low sigma characteristic of wearout such as electromigration) supplies a reasonably acceptable time before an appreciable cum % fail is seen (taken as 0.1% here), but the situation is FAR more serious if the sigma value increases. For example, if the sigma value doubles to 1.4, each "milestone" along the cum % fail curve will occur 4-fold earlier in time … PER STEP. Thus, 50% cum will still occur at 100 years for this scenario, while 16% cum occurs at 25 years, 2.3% cum at 6.25 years and 0.13% cum at 1.56 years (4-fold sooner than if sigma were half the value!). Clearly, low sigma is indicative of more robust product, even if the t_{50} value is "good."

So moving from a MTTF (Mean Time To Fail or t_{50}) mentality to a "DPM" environment (when will the early fails occur & how many?) will produce results STRONGLY dependent on the lognormal sigma or Weibull beta (shape parameter) value, either of which is a measure of failure time dispersion.

Numerical Example

Let's assume your objective is to calculate FITs or DPM, given 0 fail from a sample of 2000 in HTOL (125 C for 24 hr), a use condition of 55C, a known acceleration factor of 50:1, and an elapsed time in the field application (use condition) of 1 year. Since we have no fails, we will assume the exponential distribution, for which the failure rate (FITs) is constant. The chi-square value for "0" fails at 60% Upper Confidence Level is 0.93. 24 hr HTOL is equivalent to 1200 hr in the customer use condition (AF=50:1), so FITs = # fail@60%UCL/device-hr = 0.93/(2000*1200hr) = 381 FITs (rate). The DPM (integral) is simply the product of the rate * time, which is 381 10^{-9} /hr* 8760 hr = 3.4 DPM.

Summary

Acceleration Factors are fine for comparing performance under different environmental stresses, but fail to answer questions about the time element. The most important question for a user is when will the first 1 ppm fail, when for 10 ppm, when for 100 ppm, etc? Customers will typically NOT be most interested in t_{50}, but rather more interested in the early failure rate (impacts warrantee costs, customer perception, reputation, etc.). Some customers will prefer the integral of failure rate to count the "dead bodies" for purposes of assessing the financial impact. Choice of the proper mathematical distribution will affect the results and proper choice will be a function of the number failing in the accelerated stressing or in the field application and whether the failure rate is constant, increasing (wearout), or decreasing (infant mortality) with time.

References

1. A. Wager, *IRPS Tutorial*, 1996
2. P.A. Tobias & D.C. Trindade, "Applied Reliability," 2nd Ed. 1996
3. J. R. King, "Probability Charts for Decision Making" TEAM (803) 323-8843 (1971)

International SEMATECH Technology Transfer
2706 Montopolis Drive
Austin, TX 78741

http://www.sematech.org
e-mail: info@sematech.org