FINAL REPORT

US ARMY SBIR PHASE II

THE INFLUENCE OF TEMPERATURE ON MICROELECTRONIC DEVICE FAILURE MECHANISMS

Contract Number: DAAL01-91-C-0142

Ramsearch Company
2128 Espey Court - 8
Crofton, MD 21114

September 4, 1993

Prepared For:
US Army Research Laboratory
Electronics & Power Sources Directorate
M/S AMSRL-EP-RA (Edward B. Hakim)
Fort Monmouth, NJ 07703-5601

Prepared By:
Michael Pecht

Development of the Information contained herein was funded and/or technically sponsored by the U.S. Army, Research Laboratory, Electronics & Power Sources Directorate.
Table of Contents

1.0 Introduction
1.1 Microelectronic Device performance and Reliability: Role of Temperature 1
1.2 Single Activation Energy Arrhenius Models: What is the Problem? 2
1.3 Reliability Prediction Methodologies: What is the Problem? 3
1.4 Variability of Life versus Activation Energy 3
1.5 Screening and Accelerated Testing: What is the Problem? 3
1.6 Thermal Derating Guidelines: What is the Problem? 7
1.7 The Objective 7

2.0 Temperature Dependence of Microelectronic Device Failure Mechanisms 8
2.1 Introduction 8
2.2 Temperature Dependencies of Failure Mechanisms in the Die Metallization 10
2.3 Effect of Hydrogen (h2) and Helium (he) Ambients on Device Metallization versus Temperature 66
2.4 Temperature Dependencies of Failure Mechanisms in the Device Oxide 68
2.5 Temperature Dependencies of Failure Mechanisms in the Device 89
2.6 Temperature Dependencies of Failure Mechanisms in the Device Oxide Interface 96

3.0 Temperature Related Parameter Variations 115
3.1 Introduction 115
3.2 Temperature Dependence of Bipolar Junction Transistor Parameters 116
3.3 Temperature Dependence of Mosfet Parameters 126

4.0 A Physics-of-Failure Approach to IC Burn-in 147
4.1 Introduction 147
4.2 Burn-in philosophy 148
4.3 Problems with Present Approach to Burn-in 148
4.4 A Physics-of-Failure Approach to Burn-in 150
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.0</td>
<td>Derating Guidelines for Temperature-Tolerant Design of Microelectronic Devices</td>
<td>157</td>
</tr>
<tr>
<td>5.1</td>
<td>Introduction</td>
<td>157</td>
</tr>
<tr>
<td>5.2</td>
<td>Problems with the Present Approach to Device Derating</td>
<td>158</td>
</tr>
<tr>
<td>5.3</td>
<td>A Proposed Approach to Device Derating</td>
<td>159</td>
</tr>
<tr>
<td>5.4</td>
<td>Derating for Failure Mechanisms in Die Metallization</td>
<td>164</td>
</tr>
<tr>
<td>5.5</td>
<td>Derating for Failure Mechanisms in Device Oxide</td>
<td>185</td>
</tr>
<tr>
<td>5.6</td>
<td>Derating for Failure Mechanisms in the Device</td>
<td>195</td>
</tr>
<tr>
<td>5.7</td>
<td>Derating for Failure Mechanisms in the Device Oxide Interface</td>
<td>200</td>
</tr>
<tr>
<td>5.8</td>
<td>Derating for Failure Mechanisms in the Die and Die/Substrate Attach</td>
<td>201</td>
</tr>
<tr>
<td>5.9</td>
<td>Derating for Failure Mechanisms in First-level Interconnects</td>
<td>207</td>
</tr>
<tr>
<td>5.10</td>
<td>Derating for Failure Mechanisms in the Package Case</td>
<td>222</td>
</tr>
<tr>
<td>5.11</td>
<td>Derating for Failure Mechanisms in Lid Seals</td>
<td>225</td>
</tr>
<tr>
<td>6.0</td>
<td>Conclusion</td>
<td>228</td>
</tr>
<tr>
<td>7.0</td>
<td>References</td>
<td>234</td>
</tr>
</tbody>
</table>
There has been a common belief that reliable electronics can be achieved by lowering temperature. Elevated temperature has in the past been considered a dominant stress that lowers reliability, so every effort has been made to lower operating temperature until the desired reliability is achieved. The belief in the harmful effects of temperature has woven itself into today’s screening and thermal derating processes. High-reliability applications require that the microelectronic device be subjected to high-temperature stress screens, like burn-in, to improve the reliability of the product. Moreover, thermal derating measures for microelectronics often involve lowering temperature.

A problem arises when there is no way to evaluate the temperature sensitivity of device design. The design team has no choice but to follow the prescription of lowering temperature when there is no scientific tool to answer the questions:

- Is there a need for lowering temperature?
- If there is need for lowering temperature, what is the value of the lower temperature as a function of device architecture?
- How does the maximum operating temperature vary with microelectronic device
design?

• What device design modifications are necessary to vary the maximum allowable operating temperature for a desired mission life?

• How do manufacturing defect magnitudes affect the time to failure under temperature stress?

• Is it possible to change any other form of stress instead of temperature to achieve the desired mission life and still escape the penalty of the added cost and weight of a cooling system? This book attempts to address these questions.

The purpose of this book is to raise the level of understanding of thermal design criteria beyond simply lowering the operating temperature to achieve reliability. The goal is to provide the design team with sufficient knowledge to help them evaluate device architecture trade-offs and the effects of operating temperature, rather than just cooling the system to a lower temperature in expectation of higher reliability. The penalty of added cost and weight associated with lowering temperature can be minimized by exploring cost effective options in terms of derating other non-temperature stresses. This book will also assist with evaluating the effects of certain classes of manufacturing defects on operating life for high-temperature operation in order to tailor existing screens for maximum defect detection and overall device quality.

This book is directed to the reader interested in the damage mechanisms associated with various forms of temperature stress in microelectronic packages. The microelectronic package considered for the purpose of this investigation is assumed to consist of a bipolar or MOSFET (silicon) semiconductor device; first-level interconnects that may be wirebonds, flip-chip, or tape automated bonds; die attach; substrate; substrate attach; case; lid; lid seal; and lead seal. Failure mechanisms actuated under various forms of temperature stress, including steady-state temperature, temperature cycling, temperature gradients, and time-dependent temperature change, have been identified for each of the package elements. The temperature effects on electrical parameters of both bipolar and MOSFET devices have been investigated, and models quantifying the temperature effects on package elements have been identified. Temperature-related models have been used to derive derating criteria for determining the maximum and minimum allowable temperature stresses for a given microelectronic package architecture. The reader should have prior knowledge of bipolar and MOSFET device fundamentals, semiconductor device physics, fracture mechanics, and elasticity dynamics.

This book does not address damage mechanisms in device technologies other than bipolar and MOS, or assemblies such as circuit cards, printed wiring boards, sub-assemblies, and assemblies. This investigation covers damage mechanisms in
the temperature range of -55°C to 125°C. At temperatures much higher or lower than this temperature range, the damage mechanisms in the microelectronic package can change considerably with respect to their stress dependencies.

In Chapter 1, we discuss the motivation for the research work presented in this book. The existing strategies for modeling the effects of temperature on microelectronic device reliability have been discussed. Some of the problems with the existing modeling strategies have been outlined.

In Chapter 2, we discuss microelectronic device failure mechanisms in terms of their dependence on steady state temperature, temperature cycle, temperature gradient, and rate of change of temperature. The microelectronic package is considered to be an assembly of package elements, including the chip, chip metallization, operating devices on the chip, die attach, substrate, substrate attach, first-level interconnects (wirebond interconnects, tape automated bonds, flip-chip bonds), leads, lid, lead seal and lid seal. Common models used to characterize these failure mechanisms are identified and the variability in temperature dependence of each of the failure mechanisms is characterized. We also discuss the existence of temperature thresholds below which various failure mechanisms are not significantly activated.

In Chapter 3, the effect of temperature on the performance characteristics of MOS and bipolar devices is examined. The parameters investigated for bipolar devices include the current gain, I-V characteristics, collector-emitter saturation voltage, and voltage transfer characteristics. The parameters investigated for MOS devices include threshold voltage, mobility, drain current, time delay, leakage currents, chip availability, dc voltage transfer characteristics and noise margins.

In Chapter 4, the applicability of using high-temperature stress screens, including burn-in, for high-reliability applications is discussed. The burn-in conditions used by some manufacturers are examined, and a physics-of-failure approach is proposed. The proposed physics-of-failure approach addresses the dominant failure mechanisms in the device architecture and tailors the screening stresses to effectively remove defective devices.

Chapter 5 briefly overviews existing guidelines for thermal derating of microelectronic devices, which presently involve lowering the junction temperature. Then, the models presented in Chapter 2 for various failure processes, are used to evaluate the sensitivity of device life to variations in manufacturing defects, device architecture, temperature, and non-temperature stresses. Derating curves for constant device life are derived for mechanisms with complex dependencies on stresses and defects. The cumulative effect of competing failure processes on device life is used to determine the values of operating temperature and non-temperature related stresses. We give special thanks to
Charles T. Leonard of Boeing Commercial Airplanes for his constructive comments, criticisms and suggestions. We acknowledge the help of Joseph Kopanski from NIST; Fausto Fantini from Telettra (Italy); Sorin Witzmann from BNR; Canada; D.S.Campbell from Loughborough University of Technology (U.K); Pat O’Connor from British Telecom (U.K); J.R.Lloyd from Digital Equipment Corp.; Walter Winterbottom from Ford Motor Company; Michael Cushing from AMSAA; George L. Schnable from David Sarnoff Research Center; George A. Swartz from RCA Lab.; Steve Martell from SONOSCAN; Dennis Karr from Airpax Corporation; Vinod Maudgil from CETAR Ltd.; Anthony Gallo from Dexter Corp.; Dave Urech from Allied Signal Aerospace Co., Werner Engelmaier from Engelmaier Associates; Thomas Kwok from IBM Research Center; Robert Keyes from IBM Research Division; Lloyd Condra from Eldec. Corporation; Nick Lycoudes from Motorola Semiconductor Products; Imants Golts formerly from Tektronix.
The reliability of electronic devices has often been represented by an idealized plot called the bathtub curve.

Failure in microelectronic devices have been classified as results of failure mechanisms. The failure mechanisms have been classified according to failure sites at device level and first level package.

MTF is shorter for longer ON times, if device power dissipation is large [Ajiki, 1979]. MTF is not much affected by ON:OFF time ratio when power dissipation is small.

The amount of moisture absorbed increases with increase in OFF time, i.e., the more the time spent at lower temperature, the larger is amount of moisture absorbed. [Ajiki, 1979]

The average amount of moisture absorbed remains the same as long as the ON:OFF ratio (i.e., time @ higher temperature: time @ lower temperature) is the same.

Higher power dissipating device evaporate moisture due to junction temperature rise, irrespective of environmental conditions. [Ajiki, 1979]

Time to failure due to corrosion versus duty cycle ($k_o$). [Pecht, 1990]

The lifetime due to electromigration is a complex function of temperature and cannot be represented by a simple activation energy. The temperature acceleration can however, be represented by an apparent activation energy which changes with operating temperature.

Temperature dependence of vacancy supersaturation distribution [Rosenberg and Ohring, 1970]

The combinations of current density and temperature, which will produce electromigration damage in Ti-Pt-Au metallization [English, 1974].

Mean time to failure varies with current density in complex manner. Simple power laws can represent experimental observations over a small range of current densities. [Venables and Lye, 1972]

Temperature dependence has been represented by an apparent activation energy which changes with test conditions [Venables and Lye, 1972].
2.12. Influence of baseline temperature on mean time to failure. [Venables and Lye, 1972]


2.15. Effect of temperature on MTF vs grain size. [Attardo and Rosenberg, 1970]

2.16. MTF vs. metal line width @ 182°C for Al lines. [Kwok, 1989]

2.17. MTF vs. metal line width @ 182°C for Al-Cu lines. [Kwok, 1989]

2.18. Temperature dependence of TDDDB for electric field of 8 MV/cm. [McPherson, 1985]

2.19. The electric field acceleration parameter is inversely dependent on temperature. [McPherson and Baglee, 1985]

2.20. Change in resistivity of silicon Vs. temperature.

2.21. Influence of temperature on ionic current due to Na⁺ and K⁺ ions [Hillen, 1986].

2.22. Change in threshold voltage due to hot electrons is much greater at lower temperatures than at higher temperatures. [Matsumoto, 1981]

2.23. Substrate characteristics for a 5 μm device at 20°C and 100°C. $V_s = 3V$ and 7V, $V_t = 0.9 V$ [Hsu, 1984]; The variation in device characteristics between 20°C and 100°C indicates that degradation due to hot electrons is almost temperature independent in this range.

2.24. Drain characteristics for a 5 μm device at 20°C and 100°C. $V_s = 3V$ and 7V, $V_t = 0.9 V$ [Hsu, 1984]; The variation in device characteristics between 20°C and 100°C indicates that degradation due to hot electrons is almost temperature independent in this range.

2.25. Substrate current dependence of device lifetime at 20°C and 100°C [Hsu, 1984]. The small variation in device lifetime for temperature variation between 20°C and 100°C clearly indicates that the degradation due to hot electrons is temperature independent in this range.

vii
3.1. Temperature dependence of current gain for a Bipolar Transistor. The temperature sensitivity of current gain is a contributing factor to hot spot formation and effects the second breakdown energy limit. It is therefore advantageous to reduce the current gain temperature dependence by lightly doping the transistor base and limiting the phosphorus doped emitter surface concentration to about $7 \times 10^{19}$ cm$^{-3}$ [Kauffman and Bergh, 1968; Buhanan, 1969, RCA, 1978].

3.2. Threshold voltage variation of n- and p-channel MOSFETs (a). With bulk concentration $10^{15}$ cm$^{-3}$ and (b). with bulk concentration $3 \times 10^{16}$ cm$^{-3}$ [Wang, 1971]. The change in threshold voltage does not produce a very significant change in circuit performance, because even a 200 mV change in $V_T$ does not cause a larger percentage change in $V_{GS} - V_T$ [Hodges and Jackson, 1988].

3.3. Effective channel mobility characteristics for n and p-channel MOSFETs. Mobility is an inverse function of absolute temperature [Shoucair, 1986]

3.4. The variation the drain-to-source on resistance versus device temperature. [Blicher, 1981]

3.5. The variation in transconductance versus temperature in the temperature range of -55 to 125°C. [Blicher, 1981]

3.6. The variation in drain current in the equipment operating range of -55 to 125°C. The variation in drain current with temperature is not very significant in the temperature range of -55 to 125°C. [Blicher, 1981]

3.7. The variation the subthreshold parameter versus temperature. At temperature greater than 150-200°C, diffusion leakage currents completely dominate the weak inversion drain characteristics, causing the subthreshold parameter $n_0(T)$ to exponentially increase with temperature. Depending on the device design, the temperature at which the subthreshold currents start to increase exponentially, typically represents the practical upper operating junction temperature above which no reasonable device "turn-off" can be achieved. [Shoucair, 1989]

3.8. DC transfer characteristics of a typical CMOS inverter with temperature as a parameter. At temperature $> 270^\circ$C, curves degenerate into flat line due to onset of pnpn latchup phenomenon. [Shoucair, 1984]

3.9. The variation in the peak low-level output voltage versus steady state
temperature. The reduction in the output voltage $V_{OL}$, indicates an increase in the noise margins at higher temperatures. [Texas Instruments, 1987]

3.10. The variation in the peak high-level output voltage versus steady state temperature. The increase in the output voltage $V_{OL}$, indicates an increase in the noise margins at higher temperatures. [Texas Instruments, 1987]

3.11. Typical safe operating area of a power MOSFET, with the device protected from second breakdown. [Blicher, 1981]

5.1. Derating methodology for temperature tolerant design.

5.2. Derating curves of life under corrosion versus temperature and duty cycle. The curves identify the various combinations of temperature and duty cycle which will result in desired life. The paradigm of higher reliability associated with lower temperature is misleading, since the same mission life of 29 years can be obtained for any temperature from 40°C to 160°C depending on the duty cycle.

5.3. Derating curves for electromigration stress. The curves identify various combinations of current density and temperature which will result in desired life [Black, 1982].

5.4. The variation in time to failure versus aspect ratio and temperature shows that SDDV changes its temperature dependence from steady state to inverse temperature dependence, at a temperature threshold which is a function of the passivation temperature. [Kato, 1990; Niwa, 1990]

5.5. Due to the non-linear dependence of life under TDDB on temperature, worst case manufacturing defect magnitude, or electric field, derating the stress below a particular value may not result in noticeable benefit in terms of increased life because the time to failure is much beyond wear-out life of the device.

a. Derating curve for TDDB versus steady state temperature and effective oxide thickness. The curves identify the various combinations of temperature $T$, and effective oxide thickness $x_{eff}$, which will result in desired life. [Lee, 1988; Moazzami, 1989; Moazzami, 1990]

b. Derating curve for TDDB versus steady state temperature and electric field. The curves identify the various combinations of temperature $T$, and electric field $E_{ox}$, which will result in desired life. [Lee, 1988; Moazzami, 1989; Moazzami, 1990]
5.6. Time for heel to transform to intermetallic versus temperature. The temperature at which intermetallic formation has a dominant dependence on steady state temperature is a function of the bond geometry. [Philosky, 1970, 1971]

a. For gold wire bonded to aluminum metallization
b. For aluminum wire bonded to gold metallization
c. Time to failure versus temperature and bond pad thickness. The time to failure at any temperature may be much greater than the mission life depending on the bond pad thickness. [Philosky, 1970, 1971]

5.7. Various time-temperature products which will result in gold-aluminum intermetallic growth [Philosky, 1971]. A horizontal line on the graph at a thickness equal to the wire thickness at the bond pad gives the time-temperature product which will result in failure after time equal to the abcissa. The slope of each line gives the rate constants at various temperatures.

6.1. Derating curve for mission life versus operational parameters for dominant failure mechanisms, at a constant operating temperature of 125°C. This derating plot an be used to derate non-temperature operational stresses for cost-effective designs which are reliable at high temperature.

6.2. Derating curve for mission life versus steady state temperature for dominant failure mechanisms, at a specified values of non-temperature operational stresses. This derating plot an be used to derate non-temperature operational stresses for cost-effective designs which are reliable at high temperature.
List of Tables

1.1. Variability in activation energy for each failure mechanism is so much that the value of predicted reliability has little meaning.

2.1. Estimated MTF values for electromigration @ 85°C.

2.2. Summary of kinetic parameters for electromigration and extent of damage sustained by Al-2%Cu thin stripes. (current stressed @ 3 x 10⁶ A/cm²) (An almost constant activation energy for different heating rates suggests that the phenomenon is temperature independent in Hydrogen and Helium ambients)

2.3. Effect of H₂, He, Ar, N₂, ambients Vs. temperature.

2.4. Comparison of various time dependent dielectric breakdown models.

2.5. Steady state temperature dependence of device failure rates. (Normal operation assumed -55°C to 125°C)

2.6. The failure mechanisms dependent on steady-state temperature, magnitude of temperature change, temperature gradient, and time dependent temperature change in the range of -55°C to 125°C.

2.7. Dominant temperature dependency over steady-state temperatures from -55°C to 500°C. (Temperature dependency includes T (steady state temperature), ΔT (temperature cycle magnitude), V T (temperature gradient), δT/δt (time dependent temperature change).

2.8. Variation of nature of temperature dependency of the device from -55°C to 500°C.

3.1. Variation in intrinsic carrier concentration, epitaxial resistivity, and collector-emitter saturation voltage for temperatures between 20°C and 200°C. [Bromstead, 1991]

4.1. Burn-in time and temperatures

5.1. Derating Guidelines For Thermal Derating


5.3. Coating Integrity Index [Pecht and Ko, 1991]

5.4. Black’s Constants For Various Metallization Materials

5.5. Temperature acceleration and current exponents for various metallization materials

5.6. Comparison of Typical ESD Protection Thresholds

5.7. Empirical Models for TDDB

5.8. The dominant intermetallic compounds for common bimetal combinations

5.9. Rate constants and activation energies for common material combinations

INTRODUCTION

1.1 MICROELECTRONIC DEVICE PERFORMANCE AND RELIABILITY: ROLE OF TEMPERATURE

Temperature is a fundamental parameter associated with the performance and reliability of electronic equipment. Performance is defined by electrical parameters such as threshold voltage, propagation delay, leakage currents, and noise margins. Performance failures (out of specification limits), typically do not render the device non-operational, and must therefore be characterized most generally by impairment of electrical functionality resulting from parameter drifts. Temperature dependence of performance is thus, quantified by the variation of electrical parameters versus temperature stress including steady state temperature, temperature cycle, temperature gradient, time dependent temperature change.

Reliability is defined by the ability of a device to fulfill its intended function. Reliability is a function of failure mechanisms operative in the package architecture which results in circuit malfunction causing equipment failure. Reliability related failures render the device non-operational because of damage resulting from failure mechanisms. Reliability temperature dependence is thus
INTRODUCTION

quantified by the dominance of failure mechanisms versus temperature stress. While operation at high temperature may not affect the reliability of a device, it may reveal that at high temperature, the device does not meet performance requirements, due to threshold voltage drift, increased leakage currents, and increased propagation delay. This may indicate the need for a design change, or the unsuitability of the technology for high-temperature operation. [Semiconductor Reliability News 1990, BT 1984, CNET 1983, MIL-HDBK-217, NTT 1985, RPP 1988]

1.2 SINGLE ACTIVATION ENERGY ARRHENIUS MODELS: WHAT IS THE PROBLEM?

In the past, effect of temperature in its many forms has been neglected in terms of the effect on overall device reliability of steady-state temperature, temperature change, rate of temperature change, or spatial temperature gradient [Wong 1990, Blanks 1990, Witzmann and Giroux 1991]. Temperature-dependent models, such as the Eyring and Arrhenius models, originally proposed to model the effect of temperature on chemical reaction rates, have been widely applied to model overall microelectronic device reliability. The governing equation assigns an activation energy to the device which is used to predict reliability over all temperatures. These models are often used to show the effect of temperature on electronic component failure rates, under the assumption that the dominant component failure mechanisms depend on steady-state temperature. This assumption is carried further to derive activation energies or thermal acceleration factors for the device based on a weighted averaging methodology, and to characterize the device failure rate as an exponential function of temperature.

The use of a single activation energy to describe the thermal acceleration of device failure rate, in which failure mechanisms have temperature accelerations varying from 6.5 to 503, has been observed to be inappropriate and misleading [Setliff 1991, O'Connor 1990, Hakim 1990]. Moreover, new failure mechanisms can become dominant, depending on manufacturing processes, and the actual failure mechanism experienced by the device can vary. No failure mechanism can remain dominant, or even significant, for very long [Witzmann and Giroux 1991]. The continual use of a single activation energy of 0.7 eV for a device when hot electrons (with an activation energy of -0.06 eV) are the dominant failure mechanism will actually predict a decrease in the mean time between failures (MTBF) with rising temperature, when actually the MTBF will increase with increasing temperature; this causes the system designer to lower the temperature further, unintentionally decreasing reliability [Setliff 1991]. Due to the
multiplicity of temperature dependencies of failure mechanisms, geometric and structural variables, device materials, and processing variables, and because of the non-exponential dependence of lifetimes on temperature, it is not possible to derive meaningful models for thermal variation of lifetimes with a single activation energy for the device.

1.3 RELIABILITY PREDICTION METHODOLOGIES: WHAT IS THE PROBLEM?

The Arrhenius model derives its importance from the fact that it has been embedded in the part failure-rate prediction models of MIL-HDBK-217, "Reliability Prediction of Electronic Equipment." Most microelectronic failure rate models in MIL-HDBK-217 are of the form

\[ \lambda_p = \lambda_b \prod \pi_i \] (1.1)

where \( \lambda_p \) is the device failure rate, \( \lambda_b \) is the base failure rate, and \( \pi_i \) are the various factors for quality, temperature, and voltage. The temperature acceleration factor \( (\pi_T) \), which is an exponential function of temperature, is often the only stress term that determines the part failure rate. Lower temperature is thus associated with higher reliability, supporting the steady-state temperature dependence of device failure rates.

1.4 VARIABILITY OF LIFE VERSUS ACTIVATION ENERGY

The Arrhenius functional relationship between the MTBF and temperature is such that a variation of 0.1 eV may vary the MTBF by an order of magnitude. This is worsened by the wide variation in the activation energies described in the literature for various failure mechanisms and manufacturing defects. The variability of activation energy for the same failure mechanism is so great that the predicted reliability has little meaning (Table 1.1).

1.5 SCREENING AND ACCELERATED TESTING: WHAT IS THE PROBLEM?

The effects of temperature on microelectronic devices are often assessed by accelerated tests carried out at high temperatures. The results are then extrapolated to operating conditions (-55°C to 125°C) to obtain a value for the thermal acceleration of device failures. Implicit in the test strategy is the assumption that all the device failure mechanisms are exponential functions of
temperature, and that the failure mechanisms active at higher temperatures are also active in the equipment operating range of -55°C to 125°C. But the high temperature failure mechanisms (> 125°C) that do not exist in the equipment operating range can be operational at higher temperatures.

The reliability of electronic devices has often been represented by an idealized plot called the bathtub curve (Figure 1.1), which consists of three regions. In region A, the failure rate decreases with time, and is called the infant mortality or early-life failure region. In region B, the failure rate has reached a relatively constant level, and is called the constant failure rate or useful life region. In region C, the failure rate increases again, and is called the wearout region. Many modern semiconductors have been improved to the point where the infant mortality and useful life regions have failure rates so near zero that the bathtub curve "no longer holds water" [Wong 1990, Beasley 1990]. Indeed, the result of continuous reliability improvement is that most mature semiconductor products and integrated circuits do not require screens such as burn-in. Modern semiconductors should not reach the wearout portion of the curve when operated within specification limits and within reasonable equipment lifetimes. In fact, for many failure mechanisms the wearout portion of the curve has been delayed beyond the useful life of many devices [Hakim 1990, McLinn 1990, O'Connor 1990].

Over the years, the process of burn-in has deteriorated into an insurance policy used to check reliability or satisfy customer-imposed requirements. Burn-in procedures are often conducted without any prior identification the nature of the defects to be precipitated, the failure mechanisms active in the device, or their sensitivity to steady-state temperature stress, or without any quantitative evidence of the improvement achieved by the process. Current failure data indicates that burn-in prior to use does not remove many failures and, on the contrary, may cause failures due to additional handling. Chapter 4 discusses the applicability of high-temperature stress screens, such as burn-in, for improvement of overall product quality for high-reliability applications. It also proposes an approach to address the dominant failure mechanisms in a device architecture to tailor the burn-in stresses for effective removal of defective devices.
Figure 1.1 The reliability of electronic devices has often been represented by an idealized graph called the Bathtub Curve.
## INTRODUCTION

### Table 1.1 Activation Energies For Common Failure Mechanisms

<table>
<thead>
<tr>
<th>Failure mechanism/ manufacturing defect</th>
<th>Activation energy (eV)</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Die Metallization Failure Mechanisms</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Metal corrosion</td>
<td>0.3 - 0.6 eV</td>
<td>[Hakim 1989, Jensen 1982, Amerasekera 1987]</td>
</tr>
<tr>
<td></td>
<td>0.77 - 0.81 eV</td>
<td>[Peck 1986]</td>
</tr>
<tr>
<td>Electromigration</td>
<td>0.5 eV (Small grain Al)</td>
<td>[Black 1982]</td>
</tr>
<tr>
<td></td>
<td>0.43 eV (Al)</td>
<td>[Chate 1981, Towner, 1983]</td>
</tr>
<tr>
<td></td>
<td>0.35 - 0.85 eV (Al)</td>
<td>[Lloyd, 1987]</td>
</tr>
<tr>
<td></td>
<td>1.0 eV (Large grain glassivated Al)</td>
<td>[Nanda 1978, Jensen 1982]</td>
</tr>
<tr>
<td></td>
<td>0.24 - 0.57 eV (Al)</td>
<td>[Reimer, 1984]</td>
</tr>
<tr>
<td></td>
<td>0.7 eV (Al)</td>
<td>[Siato, 1974]</td>
</tr>
<tr>
<td></td>
<td>1.67 - 2.56 eV (Al-1% Si)</td>
<td>[Suehle, 1989]</td>
</tr>
<tr>
<td></td>
<td>0.58 eV (Al-1% Si)</td>
<td>[Schaaf, 1985]</td>
</tr>
<tr>
<td></td>
<td>0.96 eV (Al-1% Si)</td>
<td>[Fantini, 1989]</td>
</tr>
<tr>
<td>Metallization migration</td>
<td>1 eV</td>
<td>[Abbott 1976]</td>
</tr>
<tr>
<td></td>
<td>2.3 eV</td>
<td>[Jensen 1982]</td>
</tr>
<tr>
<td>Stress-driven diffusive voiding (constraint cavitation)</td>
<td>0.4 eV</td>
<td>[McPherson and Dunn, 1987]</td>
</tr>
<tr>
<td></td>
<td>1.0 - 1.4 eV</td>
<td>[Tezaki, et. al., 1990]</td>
</tr>
<tr>
<td><strong>Device and Device Oxide Failure Mechanisms</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ionic contamination (surface, bulk)</td>
<td>0.6 - 1.4 eV</td>
<td>[Amerasekera 1987]</td>
</tr>
<tr>
<td></td>
<td>1.4 eV</td>
<td>[Jensen 1982]</td>
</tr>
<tr>
<td>Hot carrier</td>
<td>-0.06 eV</td>
<td>[Hakim 1989]</td>
</tr>
<tr>
<td>Slow trapping</td>
<td>1.3 - 1.4 eV</td>
<td>[Jensen 1982]</td>
</tr>
<tr>
<td>Gate-oxide breakdown</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A. ESD</td>
<td>0.3 - 0.4 eV; 0.3 eV</td>
<td>[Baglee 1984, Crook 1979]</td>
</tr>
<tr>
<td></td>
<td>1 eV</td>
<td>[Hokari 1982]</td>
</tr>
<tr>
<td></td>
<td>0.3 eV</td>
<td>[Crook, 1979]</td>
</tr>
<tr>
<td></td>
<td>2.1 eV</td>
<td>[Anolick and Nelson, 1979]</td>
</tr>
<tr>
<td></td>
<td>0.3 - 1.0 eV</td>
<td>[McPherson, 1985]</td>
</tr>
<tr>
<td>C. EOS</td>
<td>2 eV</td>
<td>[Anolick 1979]</td>
</tr>
<tr>
<td>Surface-charge spreading</td>
<td>1.0 eV</td>
<td>[Hakim 1989]</td>
</tr>
<tr>
<td></td>
<td>0.5 - 1.0 eV</td>
<td>[Jensen 1982, Amerasekera 1987]</td>
</tr>
<tr>
<td><strong>First-level Interconnection Failure Mechanisms</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Au-Al intermetallic growth</td>
<td>0.5 eV</td>
<td>[Irvin 1978]</td>
</tr>
<tr>
<td></td>
<td>1.0 eV</td>
<td>[Hakim 1989, Jensen 1982]</td>
</tr>
<tr>
<td></td>
<td>1.1 eV</td>
<td>[Mizugashira 1985]</td>
</tr>
<tr>
<td></td>
<td>2.0 eV</td>
<td>[White 1978]</td>
</tr>
</tbody>
</table>
1.6 THERMAL DERATING GUIDELINES: WHAT IS THE PROBLEM?

Belief in the harmful effects of temperature and dependence on the Arrhenius equation have also woven itself into the thermal derating criteria. Existing guidelines for thermal derating of devices often suggest lowering the operating temperature [MIL-STD-883, Brummet 1982, Eskin 1984, Naval Air Systems Command AS-4613 1976, Westinghouse 1986]. Such derating criteria, in addition to misleading the designer to believe increased reliability has been obtained with lowered temperatures, also rule out the possibility of using cost-effective designs at elevated temperature. Furthermore, hard low-temperature values specified for a device’s derating criteria do not account for the effects of device architecture and design modifications on the required operating temperature for a given reliability. Chapter 5 uses the models investigated for various failure processes in Chapter 2 to evaluate the sensitivities of device life to variations in manufacturing defects, device architecture, temperature, and non-temperature stresses. Derating curves for constant device life are derived for mechanisms with complex stress and defect dependencies.

1.7 THE OBJECTIVE

A global methodology is presented here to address the effect of temperature on microelectronic reliability through a rational design approach to temperature and cost-effective reliability dictating the following elements in sequence:

- determining the dominant failure mechanisms in the microelectronic device;
- determining the dominant temperature dependencies of various failure mechanisms in terms of steady-state temperature, temperature cycle magnitude, temperature gradient, and time dependent-temperature change;
- providing derating guidelines for determining an upper temperature limit — based on actual negative effects of steady-state temperature, temperature cycling, temperature gradient, and time-dependent temperature change — that must be avoided for failure mechanisms having temperature dependencies.
2

TEMPERATURE DEPENDENCE OF MICRoeLECTRONIC DEVICE FAILURE MECHANISMS

2.1 INTRODUCTION

This chapter examines failure mechanisms for various package elements in terms of their temperature dependencies (Figure 2.1). Temperature effects have been broadly classified as steady-state temperature, temperature cycling, temperature gradient, and time-dependent temperature change. The microelectronic components examined here are the chip (die) and the device packaging. The chip has been further subdivided into die metallization, device oxide, device, and device-oxide interface. Device packaging has been subdivided into first-level interconnects, package case, leads, lead seals, die, and substrate attaches. First-level interconnects include wirebonded interconnects, TAB, flip-TAB, and flip-chip.
Figure 2.1 Failure mechanisms in microelectronic devices have been classified according to failure sites at die (or chip) level and first level package.
Failure mechanisms occurring predominantly at the die level include slow trapping, hot electrons, electrical overstress, electrostatic discharge, dielectric failure, oxide breakdown, and electromigration. First-level package failure mechanisms generally arise from corrosion, differential thermal expansion between bonded materials, large time-dependent temperature changes, and large spatial temperature gradients, all of which can cause tensile, compressive, bending, fatigue, and fracture failures. Corrosion-induced failure mechanisms include electrolyte formation and galvanic and ionic corrosion. Corrosion failures are complex functions of contamination, temperature, humidity, and bias. All failure mechanisms have been discussed here in terms of their dominant temperature dependencies. Failure mechanisms with a dominant temperature dependence in the range of -55 to 125 °C have also been identified.

2.2 TEMPERATURE DEPENDENCIES OF FAILURE MECHANISMS IN THE DIE METALLIZATION

2.2.1 Corrosion of Metallization and Bond Pads

Electrolytic Corrosion. Corrosion is typically defined as the chemical or electrochemical reaction of a metal with the surrounding environment. The mechanisms of corrosion can be divided into two types: dry -- such as oxidation of aluminum in air, and wet -- in which the reaction occurs in the presence of an electrolyte, a moist environment, and an electromotive force. Dry corrosion is of minor importance in semiconductor devices, since the corrosion process is self-passivating, forming a thin oxide film that prevents further oxidation. Wet corrosion, in the presence of an ionic contaminant, and moisture can provide a conductive path for electrical leakage between adjacent conductors, dendritic growth, or corrosion of the device metallization or bond pads. The ions most commonly found on the die surface that, in the presence of water, give rise to the electrolytic solution required to trigger the corrosion process include:

- halogens (especially Cl\(^{-}\)) deriving their origin both from inadequate removal of fabrication process residue and from plastic containers;
- alkalines (especially Na\(^{+}\)) derived from the diffusion ovens, glass containers, and the hands of operators; and
- phosphorus, which, unlike contaminating ions, is specially incorporated in the surface passivation glass to improve its mechanical characteristics and restrict the effects of Na\(^{+}\).
The importance of passivation in microcircuit metallization corrosion versus temperature has been evaluated by Commizolli, who found that the dependence of corrosion current on temperature appeared linear between 60°C and 100°C. The corrosion current at 90% relative humidity using passivated chips decreased with a decrease in temperature between 60°C and 100°C [1980].

Pecht modeled the time to failure due to corrosion as the sum of moisture ingress time and the time for corrosion attack and failure. The moisture ingress time for a hermetic package was determined based on the internal volume and the leak rate. The method for calculating the worst-case operation-independent moisture ingress time was based on standard testing procedures. After sufficient moisture ingress, a critical moisture content will be reached inside the package, and corrosion can initiate once the non-operating sealed package is exposed to temperatures below the dew point. At this time, the moisture inside the package condenses and combines with any ionic contaminant present to provide a conductive path between adjacent metallic conductors. The conductive path serves as a medium for the transfer of ions in the corrosion process. When the package is operating, the heat dissipated inside the package will typically elevate the temperature above the dew point. Consequently, the electrolyte will evaporate and no longer provide an electrolytic path between conductors. Elevated temperature due to device power thus acts as a mechanism to slow the corrosion process [Pecht and Ko 1990, RAC Report SOAR-3 1985].

When a packaged component is in the off mode, it can equilibrate with the ambient RH. However, when the packaged component reaches its operating temperature, the same moisture content results in a lower relative humidity. The low relative humidity at operating temperature may prevent the formation of an electrolyte and may negate exposure to corrosion entirely. The MTF is shorter for longer ON times, if device dissipation is large. For small—power devices, the MTF is not much affected by the ON:OFF ratio (Figure 2.2). The amount of moisture absorbed increases with an increase in OFF time (Figure 2.3), but the average amount of moisture, remains the same as long as the ON:OFF ratio is the same (Figure 2.4). Higher-power dissipating devices evaporate moisture due to junction temperature rise regardless of environmental conditions (Figure 2.5) [Ajiki 1979, Macheils 1991, Shirley 1991].
Figure 2.2  MTF is shorter for longer ON times, if device power dissipation is large [Ajiki, 1979]. MTF is not much affected by ON:OFF time ratio when device power dissipation is small.
Figure 2.3  The amount of moisture absorbed increases with increase in the OFF time, i.e., the more the time spent at lower temperature, the larger is the amount of moisture absorbed [Ajiki, 1979]

![Graph showing the relationship between weight change (ppm) and ON/OFF life time (hours). The graph indicates that weight change increases with increasing ON/OFF life time and decreasing ON/OFF time.]
Figure 2.4 The average amount of moisture remains the same as long as ON:OFF ratio is the same [Ajiki, 1979]
Figure 2.5  Higher power dissipating devices evaporate moisture due to junction temperature rise, irrespective of environmental conditions. [Ajiki, 1979]
The time to failure as a result of corrosion failure is given by

\[ \tau_c = \left( \frac{K_1 K_2 K_3}{K_4} \right)^{\frac{w^2 h n d F \rho}{4M V Z}} \]  

(2.1)

where \( M \) is the atomic weight of a metal conductor of density \( d \), width \( w \), height \( h \), and chemical valence \( n \); \( \rho/Z \) is the sheet resistance of the electrolyte; and \( V \) is the voltage applied. \( K_1 \) represents the physical and chemical properties of the metallization materials, \( K_2 \) is the coating integrity index, \( K_3 \) is the mission profile correction factor, and \( k_4 \) is the environmental stress correction factor. The effect of the environmental stress factor on the MTF is represented in Figure 2.6.

The only temperature-dependent k term is the environmental correction factor, \( K_4 \), utilized to determine the time to failure for various temperature and humidity conditions. The \( K_4 \) term is modelled as

\[ K_4 = \frac{(R H_R)^n \exp(E_d/K T_R)}{(R H)^n \exp(E_d/K T)} \]  

(2.2)

where \( R H_R \) is a reference relative humidity (%), \( E_d \) is an activation energy (eV), \( K \) is the Boltzmann constant (eV/°K), \( n \) is a material constant, and \( T_R \) is a reference temperature (°K).

### 2.2.2 Electromigration

A major VLSI failure mode is mass transport resulting from a momentum exchange between conducting electrons and the metal atoms in the conductor. The phenomenon of electrotransport, or "electromigration," is the result of high current density (typically of the order of \( 10^6 \) amperes/cm\(^2\) in aluminum) in metallization tracks, which produces a continuous impact on the grains in the metallization, causing the metal to pile up in the direction of the electron flow and produce voids upstream with respect to the electron flow [Schnable 1988]. The result is a net flux of metallization atoms that generally migrate in the same direction as the electron flux. Electromigration-induced damage in thin-film conductors usually appears in the form of voids and hillocks. Voids can grow and link together to cause electrical discontinuity in conductor lines, leading to open circuit failure. Hillocks can also grow and extrude materials, causing short-circuit failure between adjacent conductor lines on the same level, or in adjacent levels in multi-level interconnecting structures. Alternatively, electromigration can break through the passivation or the protective coating layers and lead to subsequent corrosion-
Figure 2.6 Time To Failure Due to Corrosion vs. Duty Cycle [Pecht, 1990]

The role of temperature in electromigration is extremely complex, especially within normal operating temperatures 125°C. The lifetime limitation due to electromigration is a complex function of temperature and cannot be represented by a simple activation energy. The temperature acceleration can, however, be represented by an apparent activation energy that changes with operating temperature (Figure 2.7). There have been insufficient tests on electromigration at temperatures less than 125°C, due to the difficulty in conducting such tests. Most tests have been performed at elevated temperatures in the neighborhood of 150°C or higher, with the results extrapolated to normal operating or room temperatures. Extrapolation of failure rates from stress results at higher temperatures to provide reliability estimates at lower temperatures will not be
Figure 2.7  Higher reliability cannot be attributed to lower temperature, since electromigration occurs with the same order of magnitude lifetimes at all temperatures.
accurate, since the physics-of-failure phenomenon is not the same.

Electromigration damage forms at sites of atomic flux divergence, of which the three main sources are structural defects, microstructural inhomogeneities, and local temperature gradients. Electromigration failures tend to be localized near sites of maximum temperature gradient [Lloyd 1988, Schwarzenberger 1988], even though typical field failures are characterized by structurally induced flux divergences, rather than temperature-gradient-induced flux divergences [Shatzkes 1986].

**Modelling temperature effects on electromigration.** The phenomenon of electromigration-induced mass transport is attributed to atomic flux resulting from electromigration in the lattice and at grain boundaries during the passage of current through a polycrystalline thin-film conductor. Diffusivity (D), exponentially dependent on temperature, is the only temperature-dependent term in the flux equation. The contribution due to electromigration in the lattice is [Huntington 1961]

\[ J_1 = \frac{1}{kT} N_l D_j \rho e Z^* \]  

(2.3)

The contribution of electromigration in the grain structure is given by

\[ J_b = \frac{1}{kT} \frac{N_b \delta b}{d} D_b j \rho e Z^* \]  

(2.4)

where \( N \) is the atomic density, \( D \) is the diffusivity, \( j \) is the current density, \( \rho \) is the resistivity, \( e Z^* \) is the effective charge, \( k \) is the Boltzmann constant, and \( T \) is the steady-state temperature [Ho 1974]. The subscripts \( l \) and \( b \) represent lattice terms and grain boundary, respectively. The quantity \( \delta \) is the effective boundary width (10 Å) for mass transport, and \( d \) is the average grain size. In aluminum, the transport is via the grain boundary, so grain boundary parameters, such as diffusivity, are important.

Electromigration damage can occur only where there is divergence in the electromigration flux, \( J \), caused by variations in any of the parameters on the right-hand side of Equations (2.3) and (2.4). For instance, if the grain size of the metallization changes, perhaps due to a change in the substrate, the flux-carrying capacity of the track is altered and flux divergence results. Cross-section changes in themselves do not lead to flux divergences, because a reduction in the cross-sectional area leads to a higher current density in the remaining section. However, a section change can cause a change in local self-heating and consequently, a
temperature-induced divergence.

Temperature changes in the stripe are important sources of flux divergence, because the flux depends exponentially on temperature [Schwarzenberger 1988]. Temperature changes may arise from changes in the thermal properties of the substrate as the metallization passes over other features on the substrate, or by changes in self-heating caused by section changes for instance, at steps over substrate features. Temperature changes can also lead to thermomigration (transport of material in a temperature gradient), though the magnitude of thermomigration is small compared to electromigration [Schwarzenberger 1988].

At moderate temperatures (much lower than $0.5T_m$, where $T_m$ is the melting temperature of the materials), the atomic flux from electromigration in the lattice is minute compared with that from the grain boundaries, so grain boundary electromigration becomes the dominant mode of mass transport for temperatures much lower than $0.5T_m$ [Ho 1989]. The mass transport in the grain boundaries in most metals occurs by vacancy diffusion [Kwok 1981]. The relative contributions of the atomic flux due to lattice diffusion and grain boundary diffusion can be estimated from the ratio of $J_1$ and $J_b$, given by [Ho 1989]

$$\frac{J_1}{J_b} = \frac{N_1 d D_1 Z_{1}^*}{N_b \delta D_b Z_{b}^*} \tag{2.5}$$

The subscripts 1 and b denote the electromigration parameters of the lattice and grain boundary, respectively. The measured values of $Z_{1}^*$ and $Z_{b}^*$ usually do not differ by more than one order of magnitude. For thin films with 1$\mu$m grain size at $0.5T_m$, where $T_m$ is the melting temperature of the materials [Ho 1989],

$$\frac{N_1}{N_b} = 1 \quad \frac{d}{\delta} = 10^3 \quad \frac{D_1}{D_b} = 10^{-7} \quad \frac{J_1}{J_b} = 10^{-4} \tag{2.6}$$

Attardo, while conducting experiments on 0.4-0.6 - mil wide and 10-12 mil long Aluminum films on silicon wafers with 800 Angstroms of thermally grown silicon dioxide and 2,000 Angstroms of sputtered quartz, found that mass transport during electromigration shifted from grain-boundary diffusion to lattice at threshold temperatures higher than $0.5T_m$ [1970]. The rate of electromigration was determined by the film's degree of preferred orientation.

Diffusion rates are highly anisotropic. The diffusion parallel to dislocations constituting tilt boundaries proceeds at several orders of magnitude greater than the diffusion perpendicular to the dislocation. The vacancy flux in the grain boundary, with contributions from electromigration and diffusion via grain...
boundary are given by

\[ J_v = -D_v \nabla C_v + J_b \]  

(2.7)

where \( D_v \) is the vacancy diffusivity in the grain boundary [Ho 1989]. The local variation of the vacancy concentration is given by the equation

\[ \frac{dC_v}{dt} = -\nabla J_v + \frac{C_v - C_v^*}{\tau} \]  

(2.8)

where the last term expresses the local deviation of the vacancy concentration from thermal equilibrium [Ho 1989]. The quantity \( \tau \) is the average lifetime of the vacancy, determined by the efficiency of the source and the sinks in creating and annihilating vacancies. Under a steady-state condition, \( dC_v/dt = 0 \),

\[ C_v - C_v^* = \tau \nabla J_v \]  

(2.9)

The vacancy flux is represented as

\[ \frac{dC_v}{dt} = D \left( \frac{d^2C_v}{dx^2} - \frac{dC_v Z^* eE}{kT} \right) - \left( \frac{D}{kT} \frac{dC_v}{dx} \right) \left( \frac{D C_v Z^* eE}{(kT)^2} \right) \]  

\[ \left( \frac{\Delta H_b (dT)}{T} \frac{d(\Delta H_b)}{dx} \right) \left( \frac{C_v - C_v^*}{\tau} \right) \]  

(2.10)

where \( \Delta H_b \) is the activation energy for the grain boundary diffusion [Attardo 1970]. The solution of this equation provides the rate of vacancy buildup at the point of divergence and the maximum vacancy concentration that can be achieved during steady-state electromigration. It is evident from Equation (2.10) that the damage from vacancy supersaturation can arise from any number of discontinuities other than temperature, such as structural variations between boundaries or between regions of the stripe with different structures [Attardo 1970]. The actual process of hole formation is a void growth process, and the time required for vacancy buildup to maximum supersaturation is orders of magnitude less than that needed for observation of holes. Rosenberg and Ohring calculated the vacancy supersaturations by considering the case of two boundaries of differing diffusion characteristics located in the isothermal region \( dT/\ dx = 0 \) and joining at \( x = 0 \), which is the accumulation site [1971]. The steady-state solution is represented in Figure 2.8, which demonstrates the effects of temperature on vacancy supersaturation. Assuming a product, for aluminum \( D_v Z^* \) to be \( 3 \pm 0.5 \times 10^{-2} \) cm²/sec, for current density of \( 1 \times 10^4 \) A/cm², Rosenberg and Ohring found the maximum
vacancy supersaturation to decreased from 27°C to 327°C [1971]. Once the maximum supersaturation is reached, the driving force for vacancy diffusion due to concentration gradient is lost. Rosenberg and Ohring observed
Figure 2.8 Temperature Dependence of Vacancy Supersaturation Distribution [Rosenberg and Ohring, 1970]
that, under current densities of $1 \times 10^6$ A/cm$^2$ and steady state temperature of 127°C, the maximum supersaturation was between 0.1 and 1.

Generally, electromigration has a linear dependence on temperature gradient, an exponential dependence on temperature for temperatures greater than 150°C, and a dependence on current density whose order varies from 2 to 14. The combinations of current density and temperature, which will produce electromigration damage in titanium-platinum-gold metallizations, have been characterized in Figure 2.9. The characteristics have been derived for not more than 0.14% cumulative failures (an average of eight failures/10$^6$ stripe hours) after twenty years, which is the three-sigma limit of the lifetime distribution. In figure 2.9, the lifetimes are assumed to have a fourth dependence on the current density [English 1974].

In an ideal case of a structurally uniform conductor with no temperature gradient, there is no flux divergence, so that electromigration damage will not occur [Ho 1974*, Ho 1989*]. The extent of deviation of vacancy concentration from equilibrium is proportional to the vacancy flux divergence. Whenever there is spatial variation in any of the parameters affecting the grain boundary electromigration, which may include structural defects in the form of non-uniform thickness of the conductor or temperature gradient, a divergence in the atomic flux occurs, giving rise to a local vacancy supersaturation or depletion (void formation).

**Temperature Dependence of Electromigration in Thin-film Conductors.**

**Steady-State Temperature Effects in the Presence of Cracks:** The effect of cracks in the conductor metallization perpendicular to current flow direction as correlated with temperature, has been predicted by Sigsbee [1973]. Grain-boundary electromigration, internal heat generation, and current crowding at growing voids dominate the rate processes that lead to failure. Assuming that the bottom interface below the substrate is held at a constant temperature, $T_0$, and that the conductor experiences a temperature rise $\Delta T_0$, above $T_0$ due to joule heating, for a crack perpendicular to the current direction, the current crowding is approximated by

$$J_s(X) = \frac{J_{\infty}}{1 - \frac{X}{W}}$$

(2.11)

Joule heating of the stripe causes an initial temperature rise, resulting in instability in the stripe, causing the vacancies in the stripe to migrate along grain boundaries.
Figure 2.9 Combinations of current density and temperature which will produce electromigration damage in Ti-Pt-Au metallization. [English, 1974]
and precipitate on a suitable boundary, forming elongated voids. The crack grows long by the accumulation of vacancies flowing along nearby grain boundaries. The atomic flux of vacancies in the crack is represented as

\[ J_v = J_{vo} \left( \frac{1}{1-L} \right) \left( 1 - \frac{\Delta H \Delta T_o \ln(1-L)+L}{kT^2 L} \right) \]  

(2.12)

where \( L = l/W \), \( l \) is the crack length, \( W \) is the conductor width, \( T_o \) is the bottom interface temperature, and \( \Delta T_o \) is the temperature rise due to joule heating in the metallization. \( J_{vo} \) is evaluated at initial film temperature \( (T_o + AT_o) \). The second term in brackets is due to current crowding; the third term in brackets is due to self-heating. Assuming that only the grain-boundary migration contributed to crack growth, the lifetime is calculated as

\[ \tau_f = \frac{cWT}{\delta N_a q^* p J_e D_o \exp \left( -\frac{\Delta h}{kT} \right) \Delta T_o} \left( 1 + \frac{0.2\Delta H}{kT^2} \right) \]  

(2.13)

where \( c \) is the crack width, \( W \) is the line width, \( k \) is the Boltzmann constant, \( T \) is the steady-state temperature, \( \delta \) is the effective grain-boundary width for transport, \( N_a \) is the atomic density, \( q^* \) is the effective charge, \( \rho \) is the resistivity, \( J_e \) is the current density at the crack tip region, \( D_o \) is the diffusion coefficient, \( \Delta H \) is the activation energy, and \( \Delta T_o \) is the initial temperature rise due to joule heating. This model neglects the effects due to the coefficient of resistance and the temperature sensitivity of \( q^* \); Sigsbee showed that these factors had negligible effect on the life prediction estimate. The lifetimes were found to have \( aJ_e^{-n} \) dependence, with \( n \) varying from unity at low \( \Delta T_o \) levels to 15 for high \( \Delta T_o \). The model has been shown to model grain-boundary electromigration for temperatures in the neighborhood of 260°C [Sigsbee 1973]. The above grain boundary grooving mechanism is typically noticed in silver metallizations [Venables and Lye 1972].

**Steady-State Temperature Effects in the Presence of Voids:** The atomic flux is dependent on temperature; therefore, local temperature gradients will cause a divergence in atomic flux. The depletion of mass occurs wherever the electron flow is in the direction of increasing temperature. Conversely, the accumulation of mass occurs wherever the electron flow is in the direction of decreasing temperature [Venables and Lye 1972]. Metallized stripes in good thermal contact
with their substrates have negligible temperature gradients [Venables and Lye 1972]. Electromigration in thin-metallized films is confined mainly to grain boundaries [Blech 1967, Agarwala 1970, Rosenberg 1968]. Thus, the steady-state temperature dependence of electromigration is of the same magnitude as that of grain-boundary diffusion [Blech 1967, Rosenberg 1968].

Voids form as a consequence of flux divergences at non-symmetrical nodes, and grow with time, eventually coalescing to form a gap across the conductor. The effect is particularly severe in films with a small grain size, because a large number of nodes are available to act as nuclei for void formation. On the other extreme, if the grain size is comparable to stripe width, the probability that a single grain will cover the entire stripe increases, which introduces an additional source of flux divergence acting as a barrier to atoms migrating from the negative side, and preventing the replacement of atoms transported away from connecting boundaries on the other side of the grain [Attardo 1970, Blair 1970].

The flow of current through stripes creates voids at grain-boundary nodes that are suitably oriented relative to the current flow direction and longitudinal temperature gradient. The resulting porosity increases as a function of the density of the grain-boundary nodes, current density, resistivity, and mobility of metal ions along grain boundaries, and is represented by

\[
\frac{dp}{dt} = C n j \rho \mu
\]

where \( C \) is a constant of proportionality, \( p \) is the porosity, \( n \) is the grain-boundary node density, \( j \) is the current density, \( \rho \) is the resistivity, and \( \mu \) is the mobility of metal ions along grain boundaries. Pore formation reduces the cross-sectional area of the metal stripe available for carrying the current, thereby increasing the local current density within the remaining section [Venables 1972].

\[
j = \frac{j_o}{(1-p)}
\]

where \( j_o \) is the initial current density in the pore-free stripe. The increased current density causes an increase in the current-enhanced motion of the metal atoms in the stripe, at the same time increasing the joule heating within the remaining conducting portions of the stripe. The local temperature in the stripe increases above the ambient temperature, \( T_o \), by an amount that is proportional to joule heating, given by [Venables 1972]
\[ \Delta T = T - T_o = \frac{j^2 \rho}{h} \]  

(2.16)

The temperature rise leads to a corresponding increase in mobility, \( \mu \), of the metal atoms along grain boundaries [Venables 1972].

\[ \mu = \frac{D}{kT} = \left( \frac{D_o}{kT} \right) e^{-\frac{Q}{kT}} \]  

(2.17)

where \( D \) is the diffusion coefficient for motion along grain boundaries, and \( Q \) is the activation energy for this process. The increase in temperature leads to a change in the resistivity, \( \rho \), of the metal [Venables 1972]:

\[ \rho = \rho_o (1 + \alpha (T - T_o)) \]  

(2.18)

where \( \alpha \) is the temperature coefficient of resistance. At a constant total current, the increase in resistivity causes an additional increase in the local rate of joule heating and in the effective electric field \( (j \rho) \) experienced by the atoms. Electromigration failure occurs where the grain-boundary migration and temperature gradient combine to create suitable conditions for porosity to develop until it exceeds the critical value, resulting in the melting of the stripe. Venables and Lye [1972] combined the above equations to give the time to failure as

\[ T_F = \frac{1}{2Cn} \left( \frac{\tau_o kT_o}{j_o \rho_o D_o e^{-\frac{Q}{kT}}} \right) \int_{x_o}^{x_T} \frac{e^{-Q/kT_o}}{x^2 (1 - x + xT_o)} dx \]  

(2.19)

where

\[ \tau_o = \frac{\Delta T_o}{T_o} = \frac{j_o \rho_o}{h T_o} \]  

(2.20)

and
$x = \frac{\tau_o}{(1-p)^2 + \tau_o(1-aT_o)}$\\
\hspace{2cm} = 1 - \left( \frac{T_o}{T} \right)$

(2.21)

$x = \frac{\tau_o}{1-\tau_o(aT_o-1)} \quad \Theta t=0$

(2.22)

$x_1 = 1 - \left( \frac{T_o}{T_m} \right) \quad \Theta t=T_F$

where $T_m$ is the melting temperature of the metal. They showed that time to failure versus current density varied complexly, and that a simple power-law dependence (as shown by Black) was inadequate to describe the experimental conditions over more than a small range of current densities. The results of Attardo [1970], Black [1968], and Blair [1970] were tangential to the results from the Venables and Lye model at a temperature of 210°C and current densities ranging from $1 \times 10^4$ A/cm$^2$ to $2 \times 10^6$ A/cm$^2$ (Figure 2.10). Venables and Lye [1972] showed that when the temperature dependence of times to failure due to electromigration was represented as an Arrhenius plot, although the curves appeared as accurate straight lines, the slopes yielded only apparent activation energy, which varied with test conditions, indicating that the time to failure was a complex function of baseline temperature and could not be represented by an Arrhenius plot to give an activation energy (Figures 2.11 and 2.12).

**Steady-State Temperature Effects Without Assumption Defect Magnitudes:** A general, but not universal, expression for the mean time to failure MTF or $t_{50}$, which is the time to reach failure of 50% of a group of identical conductor lines) is given in Equation (2.23). This is not a failure-rate expression, as the failure times are typically observed to follow a lognormal distribution:

$$MTF = A j^{-n} e^{\frac{E}{RT}}$$

(2.23)

where $A$ is a parameter depending on sample geometry, physical characteristics of the film and substrate, and protection coating; $j$ is the current density (A/cm$^2$);
Mean Time to failure varies with current density in a complex manner. Simple power laws can represent experimental observations over a small range of current densities [Venables and Lye, 1972].
Figure 2.11  Temperature dependence has been represented by an apparent activation energy which changes with test conditions [Venables and Lye, 1972]
Figure 2.12 Influence of Baseline Temperature on Mean Time to Failure [Venables and Lye, 1972]
and \(n\) is an experimentally determined exponent. This model is applicable only to conductor films that are wider than the average grain size of the aluminum film from which they are constructed. As the conductor width is reduced and approaches or becomes less than the average grain size, the structure begins to "bamboo," that is, most of the grain boundaries become normal to the electron flow. Black's relationship does not apply when bamooing starts to occur [Black 1982].

The electromigration lifetime test is carried out under a set of accelerated test conditions at elevated temperatures and with high-current-density stressing. The data are then extrapolated to device operating conditions, with current-density stressing below \(5 \times 10^5 \text{ A/cm}^2\), using the Arrhenius-like empirical equation cited above (originally formulated by Black [1969a,b]). Values of \(n\) that have been reported are

\[
\begin{align*}
\text{n} & = 1 \text{ to } 3 & \text{Chabra and Ainslie [1967]} \\
\text{n} & = 1.5 & \text{Attardo [1972]} \\
\text{n} & = 1.7 & \text{Danso and Tullos [1981]} \\
\text{n} & = 2 & \text{Black [1983]} \\
\text{n} & = 6 \text{ to } 7 & \text{Blair et al. [1970]}
\end{align*}
\]

Black characterized his data in the range \(0.5 \times 10^4 < j_o < 2.8 \times 10^4\), with an exponent of \(n=2\). Attardo reported \(n=1.5\) on the range \(10^4 < j_o < 10^6 \text{ A/cm}^2\). Blair reported a value of \(n=4-5\) in the range of \(10^4 < j_o < 2 \times 10^6 \text{ A/cm}^2\). Venables found that the simple power-law dependence of time to failure on current density could be used to describe experimental observations in a small range of current densities [Venables 1972].

Lloyd [Shatzkes 1986] treated electromigration failures by superimposing Fickian diffusion and mass transport due to electromigration force, and derived a modification of Black's equation:

\[
t_f = \left(\frac{2C_f}{D_o}\right)\left(\frac{k}{Z^* e \rho}ight)^2 T^2 j^{-2} e^{\frac{\Delta H}{kT}}
\]

(2.24)

where \(C_f\) is the critical value of vacancy concentration at which failure occurs, \(D_o\) is the pre-exponential factor for grain-boundary self-diffusivity, \(k\) is the Boltzmann constant, \(Z^*\) is the effective charge, \(e\) is the electronic charge, \(\rho\) is the resistivity, \(T\) is the steady-state temperature, \(j\) is the current density, and \(\Delta H\) is the activation energy. Equation (2.24) differs from Black's equation in that it has
a $T^2$ pre-exponential term, but it fits Black's data equally well.

Table 2.1 shows the estimated MTF of titanium: wolfram/aluminum (Ti:W/Al) and titanium: wolfram/aluminum + Copper (Ti: W/Al + Cu) film conductors at 85°C for $5 \times 10^3$ and $2 \times 10^4$ A/cm$^2$ current densities at 100% duty cycle. Even for the worst case $n = 1$, it has been anticipated that the actual time to failure will be greater than those predicted in Table 2.1 [Ghate 1981]. For current densities of $2 \times 10^5$ A/cm$^2$, temperatures of 125°C will not lead to failure in less than ten years with a typical exponent of $n = 1.7$.

Few studies at lower temperatures have been performed on unpassivated stripes, but these have shown that the phenomenon of electromigration shifts from grain-boundary migration to surface migration, with detachment of the stripe from the chip, at temperatures in the neighborhood of 223K to 347K [Rhoden 1991].

In order to improve the resistance of the conductor to surface electromigration, passivations consisting of glass overlays, metallic coatings, or natural oxides are used to cover the thin-film conductor. Typically, the electromigration lifetime of most conductors increases by one order of magnitude or more with complete surface coverage [Lloyd 1983, Felton 1985, Yeu 1985]. The use of transition layers such as titanium nitride (TiN) [Grabe 1983], chromium (Cr) [Levine 1984], and titanium-wolfram (Ti-W) [Fried 1982], has been reported to improve the lifetime of thin-film conductors by one order of magnitude. The transition layer provides a redundant structure, allowing void healing, and also acts as a metal diffusion-layer.

Temperature Gradient Dependence of Electromigration Lifetimes: Temperature gradients exist both globally and locally in thin-film conductors, due to heat generation from Joule heating and power dissipation from active devices on the chip. The global temperature gradient is small, except near electrodes or contact pads. Large local temperature gradients or hot spots can be caused by poor adhesion or contaminations at the interface between the metal film and the substrate, or by the variations in thickness of the metal film.

Temperature gradients are important sources of atomic flux divergence, since the flux depends exponentially on temperature. For example, at 200°C in aluminum, a 5°C change in temperature results in a change of more than 10% in the electromigration flux [Schwarzenberger 1988]. Studies on electromigration damage due to temperature gradients have revealed that void formation occurs in regions where the electron flow is in the direction of increasing temperature, and hillocks form in locations where the electron flow is in the direction of decreasing temperature [Blech 1967]. Temperature gradients can also lead to thermomigration, although migration due to thermomigration, compared to
emigration, is small for aluminum tracks in ICs.

Table 2.1 Estimated MTF Values for Electromigration at 85°C, as a Function of the Exponent Used in Black's Equation

<table>
<thead>
<tr>
<th>Current Density</th>
<th>Exponent</th>
<th>Ti: W/Al (years)</th>
<th>Ti: W/Al + Cu (years)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$5 \times 10^5$ A/cm²</td>
<td>$n = 1.0$</td>
<td>4</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>$n = 1.5$</td>
<td>5</td>
<td>17</td>
</tr>
<tr>
<td></td>
<td>$n = 2.0$</td>
<td>8</td>
<td>24</td>
</tr>
<tr>
<td>$2 \times 10^5$ A/cm²</td>
<td>$n = 1.0$</td>
<td>10</td>
<td>30</td>
</tr>
<tr>
<td></td>
<td>$n = 1.5$</td>
<td>23</td>
<td>68</td>
</tr>
<tr>
<td></td>
<td>$n = 2.0$</td>
<td>50</td>
<td>152</td>
</tr>
</tbody>
</table>

The temperature gradient dependence of electromigration failures has been confirmed by Lloyd [1988] and Schwarzenberger [1988]. Lloyd noticed that the electromigration failure location was typically near the location of the maximum temperature gradient, whereas the location of non-temperature-gradient-induced failure was randomly distributed for chromium/aluminum-copper (Cr/Al-Cu) conductors covered with polyimide passivation. He modeled the temperature of the stripe carrying current to produce significant joule heating as the balance between the heat generated in the stripe and the heat conducted away from the stripe to the surrounding thermal sinks, such as substrate, passivation, and the non-conducting portion of the metal stripe. The heat balance relation is given by

$$\rho j^2 = K \left( \frac{d^2 \Delta T}{dx^2} \right) - h \Delta T$$  \hspace{1cm} (2.25)$$

per unit volume, where $\rho$ is the metal resistivity, $K$ is the metal's thermal conductivity, $h$ is the parameter characterizing the thermal efficiency of the heat sinks, and $\Delta T$ is the temperature rise due to joule heating [Lloyd 1988]. The term on the left of Equation (2.25) is the heat generated due to current passing through the metal element; the first term on the right is the heat conduction along the metal stripe away from the heating element; and the last term is the heat conduction to the environment. If the heat conduction through the oxide (of thickness $t_{oxide}$) is considered alone, the value of $h$ is of the order
FAILURE MECHANISMS IN DIE METALLIZATION

\[ h = \left( \frac{K_{\text{total}}K}{t_{\text{total}}} \right) \]  \hspace{1cm} (2.26)

where \( K_{\text{total}} \) is the thermal conductivity and \( t \) is the stripe thickness [Lloyd 1988]. The heat sink is assumed to be at the ambient temperature. The solution to the heat conduction equation for a stripe with the origin in the center was given by Lloyd [1988] as

\[ T = A(1 - \frac{\text{Cosh}(Bz)}{\text{Cosh}(\frac{BL}{2})}) \]  \hspace{1cm} (2.27)

where

\[ B = \sqrt{(h - \frac{\rho_o \sigma j^2}{K})} \]  \hspace{1cm} (2.28)

\[ \rho = \rho_o(1 + \alpha \Delta T) \]  \hspace{1cm} (2.29)

\[ A = \frac{\rho_o j^2}{B^2 K} \]  \hspace{1cm} (2.30)

The location of failure was argued to be near the position of maximum flux divergence. The atomic flux is

\[ J = \frac{DF}{kT} \]  \hspace{1cm} (2.31)

where \( D \) is the diffusivity; \( F \) is the driving force for diffusion; the point of maximum flux divergence is

\[ \frac{dJ}{dx} = \frac{dC}{dt} = \left( \frac{dJ}{dT} \right) \left( \frac{dT}{dx} \right) \]  \hspace{1cm} (2.32)

and the condition of failure is when
The location of electromigration failure calculated from Equation (2.33) is near the location of maximum temperature gradient [Lloyd 1988]. The points of maximum flux divergence move from the edges, for higher current densities, to the middle for lower current densities. The change in the failure location was due to an exponential dependence of flux divergence on temperature, with only linear dependence on temperature gradient.

Figures 2.13 and 2.14 show the variation in the location of failure sites versus distance along the stripe, for low and high current densities. At high current densities (> 10^6 A/cm^2) accompanied by high joule heating, the highest flux divergence is very close to the location of highest temperature gradient, which is near the edge. The failure site is thus closer to the site of maximum temperature gradient. At lower current densities (< 10^6 A/cm^2), the failure location is more randomly distributed, since the effect of temperature gradient becomes small compared with other flux divergences induced by structural discontinuities.

The variation of flux between different regions of the conductor may lead to failure due to depletion in some regions. The time to failure is inversely proportional to the flux gradient within that region. The flux gradient, in terms of the temperature gradient, is given as

$$\frac{d^2 f}{dx^2} = \frac{d}{dx}\left(\frac{df}{dT}\right)\frac{dT}{dx} = 0$$

(2.33)

where N is the density of the ions, e is the electronic charge, \(\rho\) is the resistivity, Z is the effective ionic valency and represents the net effect of momentum exchange and electric field forces, \(D_\text{exp}(-E/kT)\) is the diffusion coefficient for ions in the conductor, k is the Boltzmann constant, T is the average steady-state conductor temperature, and \(E\) is the activation energy. Schwarzenberger et al. demonstrated the importance of temperature gradient as a source of electromigration flux divergence in metallization tracks and, therefore, the necessity of controlling the temperature profile in the integrated circuit to optimize its lifetime [Schwarzenberger, 1988; Oliver and Bower, 1970].

**Temperature Dependence of Electromigration in Multilayered Metallizations.** The definition of electromigration failure has changed with the introduction of multilayered interconnection metallizations with sensitive electrical circuits [Onduresk 1988]. Many studies on single-layer metallizations have used the
opening of the conductor as a criterion for failure, ignoring functional failures, including the resistance change of the metallization. The open criterion for layered metal systems may not be achievable in a test environment if one of the layers is not susceptible to electromigration.

Ondrusek derived a void-formation model for multilayered metallizations that allows calculation of void length from measured resistance and temperature coefficients [Onduresk 1988]. The temperature coefficient of resistance was
Figure 2.13  Temperature as a Function of Distance for High Current Condition [Lloyd, 1988]

probable failure sites are localized near regions of maximum temperature gradient for high current densities

Temperature (Degrees Kelvin)

Displacement From Center
Figure 2.14  Temperature as a Function of Distance for the Low Current Condition [Lloyd, 1988]

Temperature (Degrees Kelvin)

Displacement From Center

probable failure sites are randomly distributed for low current densities

Line Size = 250 microns
investigated to verify that the refractory layer remains undamaged throughout the voiding process. The temperature coefficient, $C$, for metal is defined by the equation

$$ R = R_i [1 + C(T - T_i)] $$

where $R$ is the resistance at temperature $T$ and $R_i$ is the resistance at some fixed initial temperature, $T_i$. In a simplified case of a void extending through the aluminum layer, all of the current must pass through the refractory metal layer. The total resistance of the metal stripe is the sum of the initial component due to the refractory material-metal/aluminum sandwich plus the additional series refractory metal resistance that results from the void. This is represented by the following expressions:

$$ R(\nu) = R_1 + R_2 $$

where

$$ R_1 = R_m \frac{L - \nu(t)}{L} [1 + C_m(T - T_i)] $$

and

$$ R_2 = R_r \frac{\nu(t)}{L} [1 + C_r(T - T_i)] $$

Taking the derivative of Equation (2.25) with respect to temperature gives

$$ \frac{\partial R(\nu)}{\partial T} = R_m C_m \frac{L - \nu(t)}{L} + R_r C_r \frac{\nu(t)}{L} $$

where $R_m$ is the initial resistance of the composite, $R_r$ is the initial resistance of the refractory metal layer, $C_m$ is the initial composite temperature coefficient, $C_r$ is the refractory metal temperature coefficient, $L$ is the total metal line length, and $\nu$ is the void length. Deviation from this model can occur due to intermetallic compound formation.

Reducing Electromigration Damage in Metallization Stripes. Solutions to reducing of electromigration do not lie in reducing steady-state temperature. The basic requirement for reducing electromigration damage is to reduce the local divergence of atomic flux. This can be accomplished, in principle, by reducing the magnitude of the atomic flux and/or the inhomogeneity of the parameters
controlling the mass transport. The magnitude of atomic flux is determined by the electromigration driving force and the grain-boundary diffusivity. Thus, to reduce the atomic flux, the option is to reduce either the driving force and/or the diffusivity. Reducing the driving force has some inherent difficulties, since it requires either a change in the scattering process responsible for the effective charge or a reduction in the current density. Because the current density is dictated by device functional requirements, the only choice is to reduce the grain-boundary diffusivity. The most common approach is by solute addition, which also results in improvements in conductor properties due to grain structure modification. Common examples are the addition of copper (Cu) or other solute elements such as manganese (Mn), magnesium (Mg), and titanium (Ti) to aluminum stripes [Ho 1989b].

There is a strong correlation between microstructure and electromigration lifetime in thin metal lines, which is particularly noticeable in VLSI technology when the line width and thickness of the metal lines are reduced to a submicron range comparable to the grain size. Larger lifetimes have been reported in large-grained aluminum (Al) films [Attardo and Rosenberg 1970] and in large-grained aluminum-copper films with bamboo structure [Vaidya 1980, Pierce 1981] (Figure 2.15). Annealing aluminum-copper (Al-Cu) lines at elevated temperature has been found to induce grain growth, increasing the electromigration lifetime. Structural modification of aluminum-copper (Al-Cu) films by adding titanium (Ti) and chromium (Cr) which is at about 400°C, results in the formation of intermetallic compounds such as (Al₃Ti) and (Al₇Cr), improves electromigration lifetimes, due to changes in microstructure that reduce damage formation and block void growth through the use of a redundant barrier that maintains current continuity [Kwok 1987].

Line width has a strong influence on electromigration lifetimes. Agarwala reported a linear relationship between line width and lifetime of aluminum wide lines with line widths down to 5μm [1970]. Subsequent studies have indicated that as line width reduces to below a critical value, lifetime is found to level off or reach a minimum and then increase, reversing the trend in wider lines (Figure 2.16). The critical width decreases with decreasing film thickness. Kwok found the critical line width to be about 0.75 μm for the 0.5-μm line thickness for aluminum-copper (Al-Cu) lines (Figure 2.17) [Kwok 1989]. The critical line width is sensitive to the thickness of the metal lines. Lifetime increases by a factor of 5 as line width increases from 1μm to 2 μm. The lifetime of aluminum-copper-silicon (Al-Cu-Si), chromium-silver-chromium (Cr-Ag-Cr), and aluminum/titanium (Al/Ti) lines levels off beneath critical line widths of around 2.0 μm, 1.5μm, and 1.2μm, respectively. The line width dependence is much
Figure 2.15 MTF Due to Electromigration vs. Grain Size @ 150 deg. C [Attardo and Rosenberg, 1970]

- **Temperature (1000/Kelvin)**
- **Mean Time To Failure (Hours)**
- **Grain Size = 2 microns**
  - $Q = 0.51 \text{ eV}$
- **Grain Size = 8 microns**
  - $Q = 0.73 \text{ eV}$
- **Current Density = 1E6 A/sq.cm**
- **T = 150 deg.C**
Figure 2.16  MTF Due to Electromigration vs. Metal Line Width @ 182 deg. C for Al lines. [Kwok, 1989]
Figure 2.17 MTF Due to Electromigration vs. Metal Line Width @ 182 deg. C for Al-Cu lines. [Kwok, 1987]

- **Metal Line Thickness**
  - □ 0.50 micrometer Al-Cu
  - ● 0.85 micrometer Al-Cu

**T** = 182 degrees centigrade
**J** = 1e+06 A/sq.cm

Median Time To Failure (Hours)

Metal Line Width (Micrometer)
stronger for aluminum-copper (Al-Cu) and aluminum-copper-silicon (Al-Cu-Si) lines than for chromium-silver-chromium (Cr-Ag-Cr) lines. The probability of alignment failure causing defects across a wide line is lower than for a narrow line; thus, it is more difficult for a crack to propagate across a wide line, and the expected lifetime increases. The dependence of lifetime on line width is also a function of pattern technique, metallurgy, and metal deposition conditions. Scoggan found that aluminum-copper (Al-Cu) lines patterned by chemical etch reached a minimum lifetime for line widths in the neighborhood of 5.5\(\mu\)m [1975]. The lifetime of aluminum-copper (Al-Cu) lines of the same dimension, patterned by metal lift-off, show a minimum lifetime for widths in the neighborhood of 3.5\(\mu\)m. Electromigration lifetime increases with decreasing thickness. The critical width decreases from 2.5 to 1.5\(\mu\)m when the film thickness decreases from 1.1 to 0.8\(\mu\)m in aluminum-copper-silicon (Al-Cu-Si) lines [Scoggan 1975].

Predicted interconnection failure rates at current density and temperature use conditions typically vary by several orders of magnitude, as they strongly depend on accelerated test data and model parameter selection. Generally, the failure time depends inversely on the temperature gradient and current density. Temperature acts as a strong accelerator of electromigration above temperatures of 150\(^\circ\)C. Electromigration failures in structurally uniform conductors cannot be accelerated in reasonable time frames at temperatures lower than 150\(^\circ\)C. The simple Black-type equation used for extrapolation of failure to use conditions can be used only over a small range of current densities. Moreover, the current exponent changes over various densities while the actual dependence of lifetime has been shown to be a complex function of current density and temperature that cannot be represented by an activation energy: the lifetime can be represented by an apparent activation energy that changes with operating conditions. Electromigration damage forms at sites of maximum atomic flux divergence, of which the three main sources are structural defects, microstructural inhomogeneities, and local temperature gradients. Because typical field failures are characterized by structurally induced flux divergences, geometric configurations, and metallization grain structures - but not reduced steady-state temperature - can reduce electromigration damage.

2.2.3 Hillock Formation

Hillocks in die metallization can form as a result of electromigration or extended periods under temperature cycling conditions (thermal aging) [LaCombe Christou 1982, Thomas 1983]. Hillock formation as a result of electromigration often occurs upstream in the electron flow from the area of voiding, but hillocks grow
fastest at the downstream edge closest to the source of the migrating metallization. Both voiding and hillock formation sometimes occur on top of each other at temperatures in the neighborhood of 140°C to 200°C [Thomas 1983].

Hillock formation due to extended periods under temperature cycling conditions (thermal aging) is believed to be due to a self-diffusion process that occurs in the presence of strains within the metallization [LaCombe Christou 1982]. These strains may be due to a mismatch in the thermal expansion coefficients of the metallizations gold (Au), titanium-tungsten (Ti-W), underlying refractory layer, silicon (Si), and (SiO₂). Hillock growth is more extensive in films deposited on room temperature substrates than on heated substrates, indicating that the effect may vary with grain size. Coating the metal with silicon nitride prevents failures and voids and hillocks from forming for at least 500 hours at 360°C. Hillocks form at random in aluminum films when heated to temperatures around 400°C during fabrication, and can cause electrical shorts between adjacent lines and fracture of the overlying dielectric film. In double-level metallized devices, hillocks can result in shorts between the underlying and overlying metal layers. Hillocks, which can be caused by electromigration, can result in thin dielectric sites that are susceptible to subsequent breakdown of the intermetal dielectric. Hillock formation is largely remedied by using alloyed metal, such as aluminum-copper (Al-Co) and improvements in the technologies of passivation and packaging. Hillock formation is more function of ΔT and temperature gradient, and is mildly dependent on temperature (in the neighborhood of 400°C).

2.2.4 Metallization Migration

Metal migration occurs between biased lands under conditions conducive to electrocrystallization. Dendritic growth is a common cause of failure. Conditions for metal migration include a level of around 10 A/cm² current density at the tip of the dendrite through spheroidal and parabolic diffusion, sufficient liquid medium such as condensed water, applied voltage that exceeds the sum of anodic and cathodic potentials in equilibrium with the electrolyte, and materials with defects that allow water condensation to satisfy the current density requirement.

For a given material and pore frequency distribution, the ionic current density or rate of mass transport per unit area is proportional to the fraction of the pore area containing condensed water. Diagiacomo proposed a model to predict failure as a function of environmental conditions and the physical properties of the package, and verified the predictions on the basis of migration failure data from accelerated tests [1982]. The fractional area in which condensation occurs can be represented by
FAILURE MECHANISMS IN DIE METALLIZATION

\[
A_{\text{fractional}} = \frac{1}{2} \text{erfc} \left( \frac{1}{2\sigma} \right) \ln \left( \frac{r_{\text{average}}}{r} \right)^2
\]  \hspace{1cm} (2.40)

where

\[
r = \frac{2\gamma v}{kT \ln \left( \frac{P}{P_o} \right)}
\]  \hspace{1cm} (2.41)

\[
\text{erfc} \left( \frac{1}{2\sigma} \right) = 1 - \frac{2}{\sqrt{\pi}} \int_0^{(\sigma^2)} e^{-u^2} \, du
\]  \hspace{1cm} (2.42)

\(\gamma\) = surface tension,
\(A_{\text{fractional}}\) = fractional area of condensation,
\(v\) = molar volume,
\(r\) = pore radius,
\(r_{\text{average}}\) = average radius,
\(p\) = saturated vapor pressure above meniscus,
\(p_o\) = saturated vapor pressure above flat surface,
\(p/p_o\) = \(H\) (relative humidity at which condensation occurs),
\(\sigma\) = \(\ln \left( \frac{r_{50}}{r_{16}} \right)\), sigma.

The average current density is expressed as a function of ionic concentration, electric field, diffusivity, temperature, fractional condensed area, and overpotential, using the Butler-Volmer equation of electrode kinetics and assuming diffusion control [DiGiacomo 1982, Barton and Bockris 1962, Price 1958, Adamson 1990].

\[
J = \frac{(Zf)^2 CDE}{2kT} \text{erfc} \left( \frac{1}{2\sigma} \right) \ln \left( \frac{kT (\ln H)r_{\text{average}}}{2\gamma v} \right)^2
\]  \hspace{1cm} (2.43)

where

\(C\) = ionic concentration in bulk,
TEMPERATURE DEPENDENCE OF DEVICE FAILURE MECHANISMS

To produce a metal dendrite, the current density at the whisker’s tip should be orders of magnitude higher than the average current density [DiGiacomo 1982]. The growth is possible only through parabolic and spheroidal diffusion focussing the ionic current on the dendrite tip, which can be expressed in terms of the radius of curvature:

\[ j_{up} = \frac{(ZF)^2CD}{2kT} \frac{V_{ext}}{r_d} \text{erfc} \left( \frac{1}{2\sigma} \right) \ln \left( \frac{kT (lnH) r_{\text{average}}}{2\gamma v} \right)^2 \]  

(2.44)

for dendritic growth \( j_{up} > j_{\text{critical}} \), the critical value of current density, or a mass transport rate of \( \partial Q_{up} / \partial t > \partial Q_{\text{critical}} / \partial t \). The mass transport rate is represented as \( Q_{up} = j_{up} / ZF \). Integrating the equation with respect to time and substituting for \( D = D_o \exp (-\Delta H/kT) \) gives the critical value of the mass transport rate - that is, i.e, the number of ions that must be transported to achieve dendritic growth across the gap as \( \partial Q / \partial t = \rho / Ml \), where \( l = \) distance between electrodes, \( \rho = \) density of dendrite, and \( M = \) atomic weight. The time to failure can be given as [DiGiacomo 1982]
FAILURE MECHANISMS IN DIE METALLIZATION

\[ t_f = \frac{\left( \frac{\rho}{M} \times \frac{ZF^2}{\overline{ZFD}} \right)^{1/2}}{\beta (V_{pot} - V_f) \frac{1}{kT} - \frac{\Delta H}{kT} \text{erfc} \left( \frac{1}{2\sigma} \ln \left( \frac{kT \ln Hr_{average}}{2\gamma} \right) \right)^2} \]  

(2.45)

where

- \( \rho \) = the density of the dendrite,
- \( M \) = the atomic weight,
- \( C \) = the ionic concentration,
- \( z \) = the ionic charge or the valence of the metal ions,
- \( D_0 \) = the diffusivity coefficient,
- \( l \) = the distance between electrodes,
- \( r_d \) = the dendrite radius,
- \( r_{average} \) = the average pore radius,
- \( \gamma \) = the surface tension,
- \( \nu \) = the molar volume,
- \( \rho \) = the saturated vapor pressure above the meniscus,
- \( \rho_o \) = the saturated vapor pressure above the flat surface \( H \) approaching \( p/p_o \) is the relative humidity at which condensation occurs and
- \( \sigma = \ln \left( \frac{r_{so}}{r_{sa}} \right) \) sigma).
- \( \beta \) = the fraction of metal surface at the anode which is susceptible to metal oxidation and therefore promoting metal migration
- \( V_{pot} \) = potential difference applied
- \( V_f \) = critical value of potential difference for dendritic growth

The average current, \( j \), is expressed as a function of ionic concentration, electric field diffusivity, temperature, and condensed area \( \beta \) is the fraction of metal surface at the anode which is susceptible to metal oxidation, and therefore promotes metal migration in regions satisfying current density requirements. \( \beta \) varies from metal to metal because of differences in device passivation and solubility, and also with testing time and temperature. Diagiacomo performed tests on silver (Ag) migration between tinned silver-palladium (Ag-Pd) leads in encapsulated packages with a polyimide surface coating, aluminum (Al) cap, and silicone rubber or epoxy backseal [1982]. The test was carried out under temperature/relative humidity conditions under bias. A failure was defined as a
resistance of less than $5 \times 10^6 \, \Omega$.

Metallization migration involves the formation of aluminum growths beneath the silicon dioxide layer. Metallization migration occurs during deposition of conductors on silicon dies due to the combined effects of elevated temperature and electrical stress. It has been reported that triangular aluminum growths causing local short-circuits form beneath the silicon dioxide layer when conductor metallization is deposited at high temperature on silicon substrate [Bart 1969, Lane 1970]. Lane [1970] found that triangles form within minutes after deposition of an 8,000-Angstrom-thick aluminum layer on silicon at temperatures from 500°C to 577°C. The time for growth formation decreases with increasing temperature. The high temperatures (500°C to 577°C) at which this failure phenomenon occurs makes it a recessive mechanism in the normal operation of microelectronic devices.

2.2.5 Contact Spiking

Contact spiking is the penetration of metal into the semiconductor in contact regions at temperatures typically above 400°C, causing increased leakage current or shorting. This failure mechanism is accompanied by solid-state dissolution of the semiconductor material into the metal or alloy of the metallization, with the metal semiconductor interface moving vertically or laterally into the semiconductor. Contact spiking in chips is observed at high chip temperatures or localized high contact temperatures. Localized high temperatures may cause failure of the chip-to-substrate bond, thus increasing the thermal resistance, producing a thermal runaway of the device, or resulting in a large magnitude of electrical overstress.

Contact spiking can occur when the device is exposed to high temperatures during fabrication. This failure mechanism can be minimized by using silicon containing aluminum alloys, such as Al-1%Si, or using barrier metals such as titanium-tungsten (Ti-W) [Chang 1988, Farahani 1987, T.I. 1987]. It is not possible to characterize such interdiffusion failure mechanisms by an activation energy because of their irregular behavior.

Migration of aluminum (Al) along silicon (Si) defects has been observed in NMOS LSI devices (logic gates), due to contact migration, also known as electrothermomigration. Failures are accelerated by elevated ambient temperatures in the neighborhood of 400°C. Contact migration is a major cause of failure in (GaAs) devices [DeChairo 1981, Christou 1982, Ballamy 1978, Christou 1980]. This failure mechanism may be dominant during VLSI manufacture and packaging when temperatures exceed 400°C.
2.2.6 Constraint Cavitation of Conductor Metallization

This failure mechanism is marked by the opening of the conductor metallization through the formation of slit-like voids or edge voids. Hinode and Owada while examining stress driven diffusive voiding (SDDV) failures in 0.9 μm wide Al-2% Si metallization stripes deposited on thermally oxidized silicon substrate, aged for various times at temperatures in the range of 200-295°C, found that open circuit failures occurred typically in passivated stripes. SDDV failure rate increases with increase in aging time and temperature till a critical storage temperature above which the failure rate decreases with further increase in temperature. The time to failure increases with increase in line width. Lifetime dependence on line width is characterized as: t \propto w^{2.7} \text{ (for line widths in the range of 1-2.5 } \mu \text{m). Stress driven diffusive voiding failures also manifest themselves in the form of resistance increase. Metallization stripes failing in high resistance mode are not broken, however, the resistance of the line increases stepwise during aging. Typical resistance increase steps are in the range of 1 kΩ to 50 kΩ. The resistance increase steps increase in magnitude with increase in aging temperature. [Hinode and Owada 1987; McPherson and Dunn 1987]}

There seems to be major disagreement in the microelectronic community as to the source of the stresses causing these voids. Various theories have been proposed to explain this failure mechanism, including:

- large silicon nodules in aluminum lines, causing metal voiding during room temperature storage;
- Coble and Nabaro-Herring creep of the conductor metallization;
- compressive stress of the plasma-enhanced silicon nitride passivation, combined with intrinsic stresses in nitrogen (N_2) - contaminated aluminum (Al) metal,
  causing voids at elevated temperatures [Klema 1984]
- compressive stresses of the overlying passivation, induced by cooling.

**Nitrogen (N_2) Contamination Induced Voiding.** Klema et.al., [Klema 1984] while evaluating MOS integrated circuits found that Al/Si metallization films deposited under conditions of nitrogen contamination coupled with subsequent silicon nitride (Si_3N_4) passivation resulted in open metal stripes. The opens occurred primarily at steps in narrow metal lines.

Failures were explained by the tendency of the metallization film to lower its energy. Energy stored in metallization includes: free surface energy equivalent
to the surface tension; interface energy; axial or compressive strain; and grain boundary or surface energy. Grain boundary were treated as cuts in a larger crystal. The atoms on the cut surface are chemically bonded to the solid on one side and not on the other, and thus sit in equilibrium potential wells that are at higher than the atoms in the bulk lattice. This extra potential energy constitutes the grain boundary or surface free energy. Smaller grain size thus indicates a greater potential for chemical reaction in order to minimize the overall film surface energy. Reactive species such as nitrogen (N$_2$) accumulate at grain boundaries by grain boundary diffusion both during and after deposition, and reduce mettallization film free energy principally by compound formation and strain relief. Compound formation involves the spurious reactive species chemically bonding to the available atomic orbitals of higher energy grain boundary edge atoms to produce compounds of host metal with typically large negative free energies. For aluminum mettallization the compounds include aluminum nitride (AlN) and aluminum oxide (Al$_2$O$_3$). Compound formation during sputter deposition thus reduces the driving force for grain growth. However, compounds such as aluminum oxide and aluminum nitride stop grain boundary motion and increase the mettallization resistivity and hardness.

Metal voiding was determined to be the result of the combination of intrinsic metal stress produced by increased brittleness of mettallization due to nitrogen contaminated sputtering process and thermal mismatch between the mettallization and overlying passivation. The failure rate due to stress driven diffusive voiding was found to be dependent on steady state temperature for temperatures below 180°C, and inversely dependent on steady state temperature for temperatures above 180°C. [Klema 1984]

**Silicon Nodule Formation Theory.** Curry et.al., [Curry 1984] while evaluating 64k dynamic rams, observed a mechanism of mettallization failures in both the flat and steep areas of the mettallization [1984]. The failures had two distinct morphologies: clean sharp breaks, and large, irregularly spaced voids. The observed failure mechanism was not affected by voltage and current density, and could be generated by high-temperature storage, though no failures were noticed in thermal cycling. The silicon content of the metal films was much greater than the solid solubility of silicon and aluminum, with the precipitates in the neighborhood of a quarter of a micrometer in diameter. The failures were attributed to the presence of impurities and silicon defects in aluminum films. The mechanisms responsible for the failures were speculated to be grain-boundary diffusion, temperature-assisted creep, and hydrogen-embrittlement fracture enhancement.
The failure of the metallization due to silicon nodule formation was also recognized by Donnell et al. who stated the problem in terms of the presence of large silicon nodules in aluminum which, in comparison to the cross-sectional area of the aluminum metallization, were large enough to restrict the current flow [1984]. Based on experimental evidence, they proposed the silicon nodule formation theory to explain the phenomenon of constraint cavitation. The theory is based on the fact that silicon alloys with aluminum at relatively low temperatures, and the solubility increases with temperature. The higher the temperature, the more silicon can be dissolved in the aluminum metallization before precipitates form. At normally high wafer-processing temperatures (aluminum is annealed at temperatures in the neighborhood of 475°C), more silicon can be dissolved in aluminum than can be retained at room temperature. Thus, as the wafer cools the excess silicon precipitates and forms nodules, epitaxial mounds, and epitaxial layers. The nodules form preferentially at dislocations in grain boundaries, and at stress points. The IC metallization was considered as a one-dimensional diffusion path, since in most applications the metallizations were long stripes. The diffusion of silicon in aluminum for thin films was found to be forty times that of bulk aluminum at elevated temperatures. This behavior was attributed to the polycrystalline nature of the aluminum metallization, which provided an enhanced diffusion path due to the larger number of grain boundaries. Silicon dissolved into the aluminum at the contact points, diffused throughout the strip, and was transported outward in all directions. The diffusion length of silicon into aluminum is expressed by the following equation:

$$L = [DT_a]^{1/2}$$ (2.46)

where $L$ is the length silicon can be transported through aluminum, $D$ is the diffusion coefficient expressed in cm$^2$ sec$^{-1}$, and $T_a$ is the annealing time. The diffusion coefficient of silicon is dependent on temperature. In these supersaturated aluminum films, the silicon nodules form on cooling, and further exposure to the high temperatures of glass deposition causes the precipitates to act as nucleation sites. Large nodules form by adsorbing smaller ones, and deplete the adjacent aluminum of its silicon. The size and shape of these precipitates depend on the crystalline structure of the aluminum and the rate of cooling. Slow cool-down results in large precipitates, while quick cooling results in the formation of numerous small precipitates. The temperature required to form appreciable precipitates is 200°C or higher, but the process is time-and temperature-dependent and can continue at room temperature, even though at a much reduced rate.

*Coble and Nabarro-herring Creep Theory*. The failures in long metallization lines
less than 4 micron wide, in the form of voids and cracks have also been attributed to Coble and Nabarro-Herring creep of the aluminum metallization [Turner 1985]. The source of this creep is a thermal expansion mismatch between the aluminum and the underlying silicon and silicon dioxide. The mechanism was complicated by the presence of silicon precipitates in the metallization. Turner et al. [1985] noted that the current density in the lines failing from this mode was much less than that noticed in electromigration failures; moreover, the failure rate characteristic decreased with time, unlike the electromigration failures. Turner tried to refute the earlier theory of Curry et al., which attributed the voids in the metallization to silicon nodule formation. The silicon nodule precipitate, being resistive, would give the metallization the electrical appearance of an open. Further, it was assumed that the precipitate would be removed in the top removal of the glass, leaving the appearance of a void or an open.

Turner et al. explained the failure of the metallization by reference to the thermal expansion mismatch between the aluminum and the silicon chip. The coefficient of thermal expansion of aluminum is about 26ppm/°C, while that of silicon is about 3ppm/°C, and that of silicon dioxide (SiO2) is about 0.5 ppm/°C. Aluminum alloys are typically deposited onto a substrate heated to about 300°C, then alloyed at 425°C, packaged at temperatures between room temperature and 350°C, burned in at 125°C, and operated at less than 70°C. Every time the temperature of the metal-substrate system is changed, the significant thermal expansion mismatch induces a stress in the system. If the temperature is raised, the metal is held in compression. If the system is cooled, the metal is pulled in tension. This stress is equal in both dimensions of the interface plane. The force generated by the thermal expansion mismatch is proportional to the length of the line, while the strength of the line is proportional to its width. At high temperatures, the stress in the aluminum is relieved by grain boundary migration, in which the atom in one lattice site jumps to an adjacent vacancy. In order for an atom to jump, it must have enough energy to overcome the energy separating the two sites. In equilibrium, this occurs frequently, but there is no migration because the electron jumps are at random; thus, there is no net mass flow. However, if the metal line is subjected to a stress field, some mass flow will occur to relieve the stress. Since grain boundaries have the highest vacancy density, migration along grain boundaries is most rapid. According to Gibbs, the stress relaxation rate by grain-matrix diffusion of thin polycrystalline films in tension is given by
where $\frac{d\sigma}{dt}$ is the stress rate, $M$ is the elastic modulus, $\Omega$ is the atomic volume, $B_0$ is the geometric factor (approximately equal to 10), $g$ is the grain size, $h$ is the film thickness, $D_{1}$ is the lattice diffusion coefficient, $f$ is the correlation factor for diffusion, $k$ is the Boltzmann constant, $T$ is the steady-state temperature, and $\sigma$ is the stress [Turner 1985]. The stress relaxation rate by grain boundary migration (Coble creep) is

$$\frac{d\sigma}{dt} = \frac{-B_1 \Omega MD_{g}\delta \sigma}{gh^2kT}$$

where $B_1$ is the geometric factor (approximately equal to 15), and $D_{g}\delta$ is the combined grain-boundary diffusion parameter [Turner 1985]. These two mechanisms work together in a thin film on substrate.

The stress may be relieved by two mechanisms - grain-boundary migration activated by grain matrix diffusion, and plastic deformation. Because there are few vacancies within the grain, grain matrix diffusion is relatively slow. Plastic deformation is fast, but operates only when the stress exceeds the yield stress of metal. Since grain-boundary migration can rapidly reduce the grain-boundary normal stress at higher temperatures, the metal rarely reaches the yield stress. Thus, at higher temperatures, grain-boundary migration rapidly relieves the grain-boundary normal stress, while grain-matrix diffusion slowly relieves the intra-grain stress. In unpassivated films, the edges of the lines act as vacancy sinks. In passivated films, the vacancies are trapped beneath the glass; if these vacancies collect in one location, they will form a void. Such voids are the first step in the formation of open metal lines, weakening the metal line and concentrating line stress in the metal around the void. At lower temperatures, neither grain-boundary diffusion nor grain-matrix diffusion can progress at appreciable rates [Turner 1985].

**McPherson and Dunn Model:** McPherson and Dunn model [McPherson and Dunn 1987] assumes vacancy to be an elemental unit of failure. The movement of vacancies results in clustering which results in void formation, void growth and conductor failure. Three primary sources of vacancies in VLSI Al-1%Si metallization are:

- Vacancy super saturation during metallization anneal in the neighborhood of
• Solid solubility of silicon in aluminum at room temperature is 0.5% therefore excess silicon during cool down from annealing temperature precipitates forming new nodules or increase size of existing nodules.

• Grain boundaries serve as potential sources and sinks for vacancies. Under passivation stress, gradients develop in the metallization transverse and parallel to the grain boundaries. Stress gradient serve as driving force for vacancy migration.

McPherson and Dunn model [McPherson and Dunn 1987] does not account for the effect of local yielding and work hardening during stress relaxation, and thus does not predict whether voiding will continue until the lead is totally severed. Stress concentration developed locally due to a void can result in local yielding. Further power cycling below recrystallization temperature can result in work hardening of metal and produce breakage. Further later stage of stress driven diffusive voiding can be dominated by rapid electromigration or fusing due to local rises in current. The time to failure is represented by: [McPherson and Dunn 1987]

\[ TF = B_o \sigma^{-1} e^{Q/k_B T} \tag{2.49} \]

where

\[ B_o = \frac{\ln \left( \frac{1}{f_c} \right)}{\left( \frac{\phi(b,t) - \phi(a,t)A_o}{N(t)} \right)} \tag{2.50} \]

where Q is the activation energy for vacancy diffusion (≈ 0.558 eV), \( K_B \) is the boltzmann constant (8.617 x 10\(^{-5} \) eV/K or 1.38 x 10\(^{-23} \) J/K), \( \sigma \) is the tensile stress in metallization (typical value 4 x 10\(^{9} \) dyne/cm\(^2\)), \( f_c \) is the \( N(t=TF)/N_o \), \( N(t) \) is the number of vacancies at time t, \( A_o \) is the area of surface bounding the volume of interest, a and b represent the void area in the metallization. The transport of vacancies in metallization is assumed to be Fickian and is represented by:

\[ J(x,t) = \mu n(x,t)F - D \left( \frac{\partial n(x,t)}{\partial x} \right) \tag{2.51} \]

where J is the vacancy flux, \( \mu \) is the mobility, and n is the density of vacancies,
FAILURE MECHANISMS IN DIE METALLIZATION

\[ J(x,t) = D \left( 1 - \frac{\partial n(x,t)}{\partial x} \right) \beta n(x,t) \sigma \]

\[ = \phi(x,t) \sigma e^{-\frac{Q}{K T}} \]  \hspace{1cm} (2.52)

F is the force acting on the vacancies due to stress gradients, and D is the diffusivity. The first term in equation (2.51) represents drift transport due to stress gradients, second term represents back diffusion due to concentration gradient. The model assumes that the force for vacancy transport, F, is derived from and proportional to the tensile stress \( \sigma \) in metallization, and that the drift component is much larger than the back diffusion component (\( \Phi \)). Further the temperature dependence of \( \Phi \) is assumed to be negligible compared to the exponential temperature dependence of the activation energy term.

**Passivation Constraint Theory.** Hinode [Hinode, Asano, Homma 1989] while examining stress driven diffusive voiding failures in Al-Cu-Si lines found that stress causing aluminum transport has two origins: thermal expansion mismatch between the aluminum and passivation films; compressive stress of the passivation. In fine lines, voids form due to thermal expansion mismatch, and in wide lines, voids form mainly due to compressive stress of the passivation. In wide lines (in the neighborhood of 30 \( \mu \)m) stress relaxation at higher temperatures is in form of bulge formation in the passivation containing the metallization, which causes a large deformation at the metallization edges. The temperature dependence of void growth rate is attributed to the temperature dependence of the deformation rate of passivation and mobility of aluminum. In fine lines (in the neighborhood of 2 \( \mu \)m), passivation has a lower compressive stress, and bulge formation in passivation becomes less dominant. Thermal expansion mismatch between the metallization and the passivation is the rate controlling process in fine lines. Voids in fine lines disappear when the lines are heated to a higher temperature and reappear during the cooling period [Hinode, Asano, Homma 1989]. The following models have been proposed based on the passivation constraint theory of stress driven diffusive voiding.

**Yue Model:** Yue et al. [Yue 1985] noticed that the voids in metallization were a result of certain device fabrication conditions, and that the density of these voids had a strong functional dependence on the compressive stress of the passivation film [1985]. They examined the effect of the overlying passivation and the rate
of cooling, and found that the rate of cooling from the passivation temperature had a significant effect on the void formation mechanism. The wafers, when cooled from 350°C to room temperature, developed a stress in the passivation, due to the difference in the coefficients of thermal expansion between the passivation and silicon. The maximum compressive stress was reached at room temperature and was independent of the cooling rate. The cooling rate of 22°C/min. allowed void formation, but a rapid temperature quench (approximately 65°C/min) did not allow sufficient time for the diffusion to form voids in aluminum. Yue hypothesized that since the missing aluminum from the voids was not found at any other location, as in hillock formation, the void space was a result of the coalescing of vacancies under a driving force. This driving force is comprised of large internal stresses and stress gradients developed in aluminum under cooling and enhanced by large applied passivation stress. In this process, aluminum (Al) covered by passivation is in a state of tension at room temperature. Annealing at 350°C causes the stress to become compressive, inducing multiaxial and localized stress gradients in the aluminum, especially near the line edges. A biaxial state of stress, consisting of both tension and compression, then coexists in the metal. Intergranular vacancy diffusion can occur from the boundaries in tension to the boundaries in compression diffusion creep. The coalescing of voids and vacancies can be expected to be greatest at the corners.

The metal void density can be analyzed by line fitting $D_l$ (metal void density) with the following expression:

$$D_l \propto \left( \frac{\sigma_{\text{pass}}}{G} \right)^n$$  (2.53)

where $G$ is the shear modulus of aluminum (Al), and $\sigma_{\text{pass}}$ is the compressive stress in the passivation film at room temperature [Yue 1985]. The steady-state creep - the strain rate, $\frac{de}{dt}$ - of bulk metal is related to the stress in the metal, $\sigma_{\text{metal}}$, by the following [Yue 1985]:

$$\frac{de}{dt} \propto \left( \frac{\sigma_{\text{metal}}}{G} \right)^n$$  (2.54)

For $n=1$, diffusion creep processes, such as Nabarro-Herring or Coble creep, take place; for the values of $n=4$ to 7, dislocation creep dominates. The volume of voids is proportional to the product of $\frac{de}{dt}$ (strain rate) and the fixed time and volume of the line. The stress in the metallization is not a result of silicon
FAILURE MECHANISMS IN DIE METALLIZATION

formation, since even pure aluminum (Al) films yield metal void formation [Yue 1985]. The reduction in density and size of the voids in hot substrate-sputtered aluminum/silicon (Al/Si) films is explained through the grain-size dependence of diffusion creep. In diffusion creep, the steady-state rate is related to the grain size by

\[
\frac{de}{dt} \propto \left| \frac{b}{d} \right|^n \left| \frac{\sigma_{\text{metal}}}{G} \right|
\]  

(2.55)

where \( n = 2, 3 \) for Nabarro-Herring and Coble creep, respectively; \( b \) is the Burger's vector; and \( d \) is the average grain size.

**Kato and Niwa Model**: Kato and Niwa theoretically estimated the stresses in an aluminum metallization under passivation arising from different coefficients of thermal expansion of the aluminum track and passivation film, using Eshelby's method-of-inclusion problem [Kato et al. 1990, Niwa et al. 1990]. The phenomenon of stress relaxation was found to be different at low and high temperatures.

Diffusion was practically inoperative at low temperatures, (defined as the temperature at which the time for diffusional relaxation is greater than \( 10^3 \) seconds; see Equations (2.81), (2.82), and (2.83)) and only instantaneous plastic deformation by dislocation glide was conceived as the relaxation mechanism [Kato et al. 1990, Niwa et al. 1990]. Kato et al. found that the stresses did not become hydrostatic after relaxation due to plastic deformation. The stresses and strains after plastic deformation as a function of aspect ratio are as follows.

For \( r > 0 \) and \( r < 1 \):

For aspect ratios, \( r \), such that \( r > 0 \) or \( r < 1 \), the strains after relaxation by plastic deformation are:

\[
\varepsilon_{11}^p = \frac{-(C_6 + C_2 C_7)\varepsilon_1^T + C_1(C_3 + C_2 C_9)}{(C_2(2C_3 + C_2 C_9) + C_1)}
\]  

(2.64)

and

\[
\varepsilon_{22}^p = C_1 + C_2 \varepsilon_{11}^p
\]  

(2.65)

where
Here, $\mu$ is the shear modulus of the metallization, $\nu$ is Poisson’s ratio of the metallization, $\varepsilon^T$ is the thermal strain given by Equation (2.85), $\sigma_y$ is the yield strength of the metallization, and $r$ is the aspect ratio of the metallization line ($r = a_2/a_1$; $2a_2$ is the metallization thickness and $2a_1$ is the metallization width) [Kato et al. 1990, Niwa et al. 1990]. Stresses resulting from plastic deformation are:
\[ \sigma_{11}^R = -\frac{\mu}{(1-v)} \left( \frac{2(1+v)\varepsilon_T}{1+r} + \frac{r(1-2v)+2(1-v)\varepsilon_{22}^R}{(1+r)^2} + \frac{r(1-2v)-2v\varepsilon_{11}^R}{(1+r)^2} \right) \]  

(2.73)

\[ \sigma_{22}^R = -\frac{\mu}{(1-v)} \left( \frac{2r(1+v)\varepsilon_T}{1+r} + \frac{r(1-2v)-2vr^2\varepsilon_{11}^R}{(1+r)^2} + \frac{r(1-2v)-2r^2(1-v)\varepsilon_{22}^R}{(1+r)^2} \right) \]  

(2.74)

\[ \sigma_{33}^R = -\frac{\mu}{(1-v)} \left( 2(1+v)\varepsilon_T + \frac{-2r-2(1-v)\varepsilon_{11}^R}{1+r} + \frac{-2-2r(1-v)\varepsilon_{22}^R}{1+r} \right) \]  

(2.75)

where \( \sigma_i^R \) for \( i = 1, 2, 3 \) are the stresses in the metallization line along its width, thickness, and length, respectively.

**For \( r = 1 \):**

For aspect ratios, \( r \), such that \( r = 1 \), the strains after relaxation by plastic deformation are:

\[ \varepsilon_{11}^P = \varepsilon_{22}^P = -\frac{1}{2} \varepsilon_{33}^P = -\frac{(1+v)\varepsilon_T + (1-v)\sigma_y}{\mu} \]  

(2.76)

where \( \mu \) is the shear modulus of the metallization, \( v \) is Poisson's ratio of the metallization, \( \varepsilon_T \) is the thermal strain given by Equation (2.85), \( \sigma_y \) is the yield strength of the metallization, and \( r \) is the aspect ratio of the metallization line (\( r = a_2/a_1; \) \( 2a_2 \) is the metallization thickness and \( 2a_1 \) is the metallization width) [Kato et al. 1990, Niwa et al. 1990]. Stresses resulting from plastic deformation are:

\[ \sigma_{11}^P = \sigma_{22}^P = -\frac{\mu}{(1-v)}((1+v)\varepsilon_T(1-2v)\varepsilon_{11}^P) \]  

(2.77)
\[ \sigma_{33}^p = \frac{-2\mu}{(1-\nu)}(1+\nu)e_T^p + (-2+\nu)e_{i1}^p \]  

(2.78)

where \( \sigma_i \) for \( i = 1, 2, 3 \) are the stresses in the metallization line along its width, thickness and length respectively.

For \( r = 0 \):

For aspect ratios, \( r \), such that \( r = 0 \), the strains after relaxation by plastic deformation are:

\[ \varepsilon_{i1}^p = \varepsilon_{33}^p = -\frac{1}{2}\varepsilon_{22}^p = -\varepsilon_T - \frac{(1-\nu)\sigma_y}{2\mu(1+\nu)} \]  

(2.79)

where \( \mu \) is the shear modulus of the metallization, \( \nu \) is the Poisson's ratio of the metallization, \( \varepsilon_T^p \) is the thermally strain given by Equation (2.85), \( \sigma_y \) is the yield strength of the metallization, and \( r \) is the aspect ratio of the metallization line \( (r = a_2/a_1; 2a_2 \text{ is the metallization thickness and } 2a_1 \text{ is the metallization width}) \) [Kato et al. 1990, Niwa et al. 1990]. Stresses resulting from plastic deformation are:

\[ \sigma_{11}^p = \sigma_{33}^p = \sigma_y \]
\[ \sigma_{22}^p = 0 \]  

(2.80)

where \( \sigma_i \) for \( i = 1, 2, 3 \) are the stresses in the metallization line along its width, thickness, and length, respectively.

At higher temperatures, (defined as the temperature at which the time for diffusional relaxation is less than \( 10^3 \) seconds; see Equations (2.81), (2.82), and (2.83)) further relaxation of stresses after plastic deformation was possible due to diffusional relaxation. [Kato et al. 1990, Niwa et al. 1990]. The stress state after diffusional relaxation (a function of metallization grain structure) was completely hydrostatic. Bamboo-structured metallization lines showed negligible diffusional relaxation at lower temperatures, while equiaxed grain structures demonstrated diffusional relaxation at all temperatures. The relaxation times for lattice, interfacial, and grain-boundary diffusion are given by

\[ \tau_L = \frac{\lambda^2kT}{8ED_L\Omega} \]  

(2.81)
FAILURE MECHANISMS IN DIE METALLIZATION

\[ \tau_I = \frac{a_2 \lambda^2 kT}{8EDw_i \Omega(1+r)} \]  
(2.82)

\[ \tau_B = \frac{d^2 kT}{32ED_p w_B \Omega} \]  
(2.83)

where \( \lambda \) is the length of grain along line axis (meter), \( k \) is Boltzmann’s constant \((8.617 \times 10^{-5} \text{ eV/k or } 1.38 \times 10^{-23} \text{ Joule/Kelvin)}\), \( T \) is temperature (Kelvin), \( E \) is the Young’s modulus of metallization line, \( r \) is the aspect ratio of the metallization line, \( \Omega \) is the atomic volume for conductor atoms (meter\(^3\)), \( D_L \) is the lattice diffusion coefficient, \( D_f \) is the interfacial diffusion coefficient, and \( D_B \) is the grain-boundary diffusion coefficient. Further relaxation of hydrostatic stresses was conceived to be due to several mechanisms, including decohesion or sliding at the aluminum (Al) line-passivation interface, and nucleation and growth of voids [Kato et al. 1990, Niwa et al. 1990].

The completely relaxed stress state was represented by

\[ \sigma_{11}^R = \sigma_{22}^R = \sigma_{33}^R = \sigma^R = \frac{-6\mu r(1+\nu)e^T}{(2r^2 + r + 2) - 4rv - 2(1-r^2)v^2} \]  
(2.84)

where

\[ \epsilon_{11} = \epsilon_{22} = \epsilon_{33} = \epsilon^T = \Delta \alpha(T - T_p) \]  
(2.85)

where \( \mu \) is the shear modulus of the metallization line (Pascal or N/m\(^2\)), \( r \) is the aspect ratio of the metallization line \( \{ r = (\text{metallization thickness})/(\text{metallization width}) \} \), \( \nu \) is the Poisson’s ratio of the metallization line, \( \Delta \alpha \) is the difference in coefficients of thermal expansion between the metallization line and the passivation (per°Kelvin), \( T \) is the temperature (Kelvin), \( T_p \) is the deposition temperature of the passivation, \( \epsilon^T \) is the thermal strain resulting from the mismatch between passivation and metallization, and \( \epsilon_{ps} \) is the strains in directions 1,2,3 (direction 1 = along line width; direction 2 = along line thickness; direction 3 = along line length) [Kato et al. 1990, Niwa et al. 1990].

The time to failure due to stress-driven diffusive voiding was considered to be the time during which the area fraction, \( A \), increases from \( A_{\text{min}} \) to \( A_{\text{max}} \) (Area fraction \( = \frac{r_o^2}{L^2}; r_o \) is the void radius, and \( 2l \) is the void separation).
TEMPERATURE DEPENDENCE OF DEVICE FAILURE MECHANISMS

\[
T_f = \frac{kT^3}{20Dgw\Omega}
\]

For \( \sigma \geq \frac{2\Gamma}{r_0} \)

\[
l = \frac{a_2}{r} \quad \text{for } r \geq 1
\]

\[
l = a_2 \quad \text{for } 0 < r < 1
\]

For \( \sigma < \frac{2\Gamma}{r_0} \)

\[
\text{no void growth}
\]

where \( T_f \) is the time to failure in seconds, \( k \) is Boltzmann’s constant (8.617 \times 10^{-5} \text{ eV/k or } 1.38 \times 10^{-23} \text{ Joule/Kelvin}), \( \Gamma \) is the surface energy per unit area \( (\text{J/m}^2) \), \( T \) is the temperature in Kelvins, \( D_g \) is the grain-boundary diffusivity \( (\text{meter}^2/\text{sec}) \), \( w \) is the grain-boundary thickness \( (\text{meter}) \), \( \Omega \) is the atomic volume for conductor atoms \( (\text{meter}^3) \), and \( \sigma \) is the stress in metallization given by the following:

for equiaxed grain structure:
\[
\sigma = \sigma_R
\]

for bamboo grain structure:
\[
\sigma = \sigma_{33} \quad t_{relax} \leq 1000 \text{ seconds}
\]
\[
\sigma = \sigma_R \quad t_{relax} \geq 1000 \text{ seconds}
\]

where \( \sigma_R \) is the hydrostatic stress after diffusional relaxation given by Equation (2.84), and \( \sigma_{33} \) is the stress along the metallization length given by Equations (2.75), (2.78), and (2.80).

2.3 EFFECT OF HYDROGEN (H₂) AND HELIUM (HE) AMBIENTS ON DEVICE METALLIZATION VERSUS TEMPERATURE

Reductions in the solid-state transport have been found possible through the interaction of active gases with thin-film metal conductors and bi-metallic diffusion couples (see Table 2.3). Studies indicate that a hydrogen ambient results in an improvement in electromigration rates, and stabilizes the intermetallic
formation due to interdiffusion.

Pasco and Schwarz considered several pathways for the gaseous ambients to affect mass transport mechanisms in metallization, including external surface diffusion, grain-boundary diffusion of metal atoms, and diffusion of vacancies into the bulk from defects present on external and internal surfaces, and from dislocations [Pasco 1983]. They studied the effects of heating rates in pure hydrogen and He-8.5%H\textsubscript{2}, relative to an inert helium environment, on the electromigration of Al-2%Cu conductors. For a heating rate of 1°C/min, the electromigration rate decreased by a factor of 10 for the hydrogen versus helium ambient. A decrease in the electromigration rate by a factor of five is achieved with only an 8 1/2% H\textsubscript{2} in the ambient. At higher heating rates (5°C/min), an increase in the degree of damage before the onset of failure increases with an increasing hydrogen constant. Table 2.2 shows the effect of heating rate on hydrogen and helium ambients [Pasco and Schwarz 1983]. The activation energy is constant, while the pre-exponential A varies, indicating that the ambient effects are independent of heating rate, since the activation energy is constant.

The process of electromigration is a combination of two processes: nucleation of voids, and growth of voids. Davis has shown that the nucleation of dislocation loops, with the quenching in the vacancy supersaturation, is a low-temperature process [1967]. At low temperatures in the neighborhood of 25°C, the process of nucleation of voids is transport-limited, and at moderately high temperatures (in the neighborhood of 100°C), the process is limited by a low driving force, due to the effectively lower supersaturation of vacancies. Dislocation loops can be precursors to the formation of microscopic voids in electromigration. Under current stressing, ions are transported along the grain-boundary generally, in the direction of the electron flow for aluminum (Al) alloys. In addition, vacancies flow in the opposite direction, producing local vacancy supersaturation. These vacancies can cluster to form dislocation loops and microscopic voids, given sufficient time at lower temperatures. The effect of the active ambient is to promote uniform nucleation and an increase in the nucleation rates. The ambient increases the rate of electromigration damage, up to the point where the growth of voids becomes the rate-limiting factor. This increase in the nucleation rate is actually beneficial, since many small voids are less likely to be harmful than a few large voids.

Void growth is decreased by the hydrogen ambient through a number of mechanisms [Pasco and Schwarz 1983]. Adsorbed hydrogen can become bound to vacancies, decreasing their migration rates, pinning them, or effectively reducing their concentration. In addition, the segregation of hydrogen to grain-boundaries decreases the grain-boundary energy and diffusion rates. These effect
a decrease in the overall transport and electromigration rates, due to the hydrogen ambient [Pasco 1983]. Higher heating rates decrease the time for nucleation and shift the electromigration damage to higher temperatures, which is reflected in the lowered pre-exponential A for higher heating rates (Table 2.2) [Pasco and Schwarz 1983]. In addition, hydrogen promotes uniform nucleation and, therefore, increases the amount of the damage that can be sustained before the rapid onset of failure in all cases.

Black observed that the cracking of the passivating glass overlayer as a result of hillock formation during electromigration damage could lead to a sudden increase in electromigration rates, due to the release of the compressive stresses in the conductor [1978]. These sudden increases in the electromigration rate can be avoided if the device is packaged in hydrogen ambient.

2.4 TEMPERATURE DEPENDENCIES OF FAILURE MECHANISMS IN THE DEVICE OXIDE

2.4.1 Slow Trapping (Oxide Charge Trapping and Detrapping)

Slow trapping is a failure mechanism, observed only in standard MOS transistors and certain types of memory devices, programmed by the transport of charge from the source or the drain through the gate oxide to the gate interface. High temperatures, combined with a high electric field, provide electrons with enough energy to cross the silicon-silicon dioxide (Si-SiO₂) interface [Nicollian 1974, Woods 1980]. Interstitial states at the silicon-silicon dioxide (Si-SiO₂) interface trap electrons and hold them in the oxide, permanently shifting the threshold voltage of the device. The presence of electrons permanently trapped at the oxide interface decreases the speed at which the device can be programmed by creating a field that opposes further electron flow through the oxide interface. The failure mechanism of slow trapping decreases the circuit speed and causes functional failures.

The trapped charge within the MOS oxide results in a C-V curve identical to that of the ideal structure, but shifted along the voltage axis by an amount equal to the flatband voltage shift. The flatband voltage shift is a function of the location of the oxide-trapped charge with respect to the silicon-silicon dioxide (Si-SiO₂) interface. The charges in the oxide induce equal and opposite charges divided between the silicon substrate and the metal gate. The closer the charge is to the silicon-oxide interface, the larger the charge induced in the silicon. The charge in the silicon alters the charge stored at thermal equilibrium, and thus alters
Table 2.2 Summary of Kinetic Parameters for Electromigration and Extent of Damage Sustained by Al-2%Cu Thin Stripes (current stressed at $3 \times 10^6$ A/cm$^2$). An Almost Constant Activation Energy at Different Heating Rates Suggests that the Phenomenon is Temperature Rate-Independent in Hydrogen and Helium Ambients

<table>
<thead>
<tr>
<th>Heating Rate</th>
<th>He(Helium)</th>
<th>He-H 1/2%H$_2$</th>
<th>H$_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 K/min</td>
<td>Q = 0.67 eV</td>
<td>Q = 0.67 eV</td>
<td>Q = 0.67 eV</td>
</tr>
<tr>
<td></td>
<td>A = 7.0 x $10^3$ s$^{-1}$</td>
<td>A = 1.4 x $10^3$ s$^{-1}$</td>
<td>A = 6.7 x $10^2$ s$^{-1}$</td>
</tr>
<tr>
<td></td>
<td>$\Delta R / R_{\text{heal}}$ = 0.67</td>
<td>$\Delta R / R_{\text{heal}}$ = 0.96</td>
<td>$\Delta R / R_{\text{heal}}$ = 1.2</td>
</tr>
<tr>
<td>5 K/min</td>
<td>Q = 0.67 eV</td>
<td>Q = 0.67 eV</td>
<td>Q = 0.67 eV</td>
</tr>
<tr>
<td></td>
<td>A = 1.4 x $10^3$ s$^{-1}$</td>
<td>A = 4.2 x $10^3$ s$^{-1}$</td>
<td>A = 1.1 x $10^4$ s$^{-1}$</td>
</tr>
<tr>
<td></td>
<td>$\Delta R / R_{\text{heal}}$ = 0.28</td>
<td>$\Delta R / R_{\text{heal}}$ = 0.56</td>
<td>$\Delta R / R_{\text{heal}}$ = 0.62</td>
</tr>
</tbody>
</table>

Table 2.3 Effect of H$_2$, He, Ar, and N$_2$, Ambients v. Temperature

<table>
<thead>
<tr>
<th>Atmosphere</th>
<th>Stripe Composition</th>
<th>Failure Mechanism/Failure Definition</th>
<th>Observation</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>argon(Ar), hydrogen(H$_2$)</td>
<td>aluminum</td>
<td>electromigration/resistivity</td>
<td>Electromigration rate ($dR/R_0 , dt$) decreased by a factor of 10 for hydrogen(H$_2$); atmosphere compared to argon(Ar) at $j = 0.7 \times 10^6$ A/cm$^2$ and $200^\circ$C.</td>
<td>Shih and Ficalora [1981]</td>
</tr>
<tr>
<td>Environment</td>
<td>Material</td>
<td>Mechanism</td>
<td>Description</td>
<td>Reference</td>
</tr>
<tr>
<td>-------------</td>
<td>-------------------</td>
<td>------------------------------------</td>
<td>-----------------------------------------------------------------------------</td>
<td>--------------</td>
</tr>
<tr>
<td>hydrogen(H₂), helium(He)</td>
<td>aluminum(Al), aluminum-copper(Al-Cu)</td>
<td>electromigration /resistivity</td>
<td>Al stripes: H₂ improves (decreases) electromigration rate by a factor of 13 to 16, as compared with helium(He). Al-Cu Stripes: H₂ improves electromigration rate by a factor of 4 to 7, compared with helium (He) environment.</td>
<td>Sardo [1981]</td>
</tr>
<tr>
<td>nitrogen(N₂), hydrogen(H₂) and H₂O</td>
<td>aluminum(Al)</td>
<td>hillock formation /reflectivity</td>
<td>Hillock formation at elevated temperatures during film deposition was greatly decreased for high partial pressures of H₂ and H₂O relative to that of N₂.</td>
<td>McLeod and Hartsough [1977]</td>
</tr>
<tr>
<td>hydrogen incorporated during E-beam deposition</td>
<td>aluminum(Al)</td>
<td>hillock formation</td>
<td>TTF of 40 hours at 7 x 10⁶ A/cm² for hydrogen incorporated samples; an anneal at 550°C produced fewer hillocks in the E-beam film.</td>
<td>Meyer [1983]</td>
</tr>
<tr>
<td>hydrogen(H₂), argon(Ar), oxygen(O₂), nitrogen(N₂), helium(He), and air</td>
<td>Au-Al thin-film diffusion couple</td>
<td>intermetallic compound formation</td>
<td>Au-Al intermetallic compound is formed at 350 to 400°C for all environments except H₂.</td>
<td>Shih and Ficalora [1978]</td>
</tr>
<tr>
<td>hydrogen(H₂), air</td>
<td>Cu-Sn, Ag-Sn, Ni-Sn thin-film diffusion couple</td>
<td>intermetallic compound formation</td>
<td>Rates of intermetallic formation reduced dramatically in H₂ environment compared to air. Reduction of rate by a factor of 3 to 10 observed.</td>
<td>Shih and Ficalora [1979]</td>
</tr>
<tr>
<td>oxygen(O₂), argon(Ar)</td>
<td>Au-Al thin-film diffusion couple</td>
<td>intermetallic compound formation</td>
<td>After 250°C anneal in 0.5 atm of either O₂ or Ar, SEM photography was used to evaluate the character of the intermetallic formation. Grain-boundary diffusion appears to be the dominant failure mechanism in O₂ environment; surface diffusion dominated in Ar environment.</td>
<td>Shih and Ficalora [1978]</td>
</tr>
</tbody>
</table>
the flatband voltage. The maximum value for the flatband voltage shift occurs when the charge is located at the silicon-oxide interface \((x_m = t_{ox})\), because the charge induced is contained entirely in silicon. In contrast, if the charge is located adjacent to the metal-oxide interface \((x_m = 0)\), there is no effect on the flatband voltage (Equation (2.89)). The flatband voltage shift of C-V curves is expressed by Balland and Barbottin [1989] and Muller and Kamins [1986] as:

$$\Delta V_G = \Delta V_{FB} = -\frac{1}{C_{ox}} \int_0^{t_{ox}} x' \rho(x') dx'$$

$$= -\frac{1}{\varepsilon_o \varepsilon_{ox}} \int_0^{t_{ox}} x' \rho(x') dx'$$

$$= -\frac{1}{\varepsilon_o \varepsilon_{ox}} x_m Q_{ox} \quad (2.89)$$

The gate bias voltage shift of the I-V curve is also a function of the distribution and location of the oxide-trapped charge. When a positive voltage is applied to the metal gate, the gate bias voltage shift is given by

$$\Delta V_G^+ = -\frac{x_m}{\varepsilon_o \varepsilon_{ox}} Q_{ox} \quad (2.90)$$

When a negative voltage is applied to the metal gate, the gate bias voltage shift is given by

$$\Delta V_G^- = -\frac{t_{ox}}{\varepsilon_o \varepsilon_{ox}} \left(1 - \frac{x_m}{t_{ox}}\right) Q_{ox} \quad (2.91)$$

where \(\varepsilon_o\) is the free space permittivity \((8.85 \times 10^{-14} \text{ F cm}^{-1})\), \(\varepsilon_{ox}\) is the relative permittivity of oxide [dielectric constant \(= 3.9\) (typical value)], \(t_{ox}\) is the oxide thickness (cm), \(x'\) is the distance away from the metal-silicon dioxide interface, \(x_m\) is the centroid of charge contained in the oxide (between 0 and \(t_{ox}\)), \(Q_{ox}\) is the density of the oxide-trapped charge over the thickness of the oxide per unit area of the silicon-silicon dioxide (Si-SiO) interface \((C \text{ cm}^{-2})\), and \(\Delta V_{FB}\) is the flatband voltage shift. The prime distribution moment \((X_m Q_{ox})\) increases or decreases with the increase or decrease in \(X_m\) or \(Q_{ox}\). If the trap distribution in the oxide is uniform parallel to the interfaces, the C-V curve shifts without distortion. For a non-uniform trap distribution parallel to the interfaces, the C-V curve distorts.
First-order Trapping Model. Balland and Barbottin [1989] related the drift and deformation in C-V and I-V characteristics to trap parameters in the oxide (\(N_T\), the spatial density of traps; \(E_T\), the fundamental energy level of traps; and \(\sigma_c\), the capture cross-section of traps). The first-order model assumed that the capture cross-coefficient and trap density remain constant during trapping.

Trap-Filling Kinetics: Trap-filling kinetics when re-emission is negligible can be predicted based on the following assumptions:

- the oxide has only one form of electron traps;
- the electron traps possess a single discrete level, \(E_T\);
- the electron traps do not act as generation-recombination centers; they can exchange electrons only with the conduction band;
- during the trap-filling phase, no re-emission takes place, i.e., \(e^{+}_{inj} = 0\), \(e^{-}_{sh} = 0\).

The oxide-trapped charge resulting from trap filling is given by

\[
Q_{ox} = qn_{TR}(t) = q \int_{0}^{t_{inj}} N_T(x)dx
\]

where \(N_T\) is the spatial density of traps possessing level \(E_T\) (cm\(^{-3}\)), \(q\) is the absolute value of the electronic charge (1.6 \(\times\) \(10^{19}\) Coulomb), \(t_{inj}\) is the length of the injection phase (second), \(\tau_c\) is the capture time constant (second) in the trap-filling experiment (Equation (2.93)), and \(t_{ox}\) is the oxide thickness (cm). The capture time constant is given by the equation [Balland and Barbottin 1989]

\[
\tau_c = \frac{qV_d}{\sigma_n J_{inj} v_{th}}
\]

where \(v_d\) is the drift velocity (cm s\(^{-1}\)), \(v_{th}\) is the thermal velocity (cm s\(^{-1}\)), \(\sigma_n\) is the capture cross-section for electrons (cm\(^2\)), and \(J_{inj}\) is the density of injected current (A cm\(^{-2}\)).

Trap Emptying: Trap emptying is represented as a separate process once the traps have been filled partially or totally [Balland and Barbottin 1989]. The emptying kinetics of traps when re-trapping is negligible can predicted based on the following assumptions:
FAILURE MECHANISMS IN THE DEVICE OXIDE

- only one type of carrier takes part in the exchange;
- only free electrons are released thermally or optically by traps;
- traps can exchange electrons only with the conduction band;
- the density of electrons is zero at the onset of depopulation;
- at all times, the density of free electrons is small compared to the density of carriers still trapped.

The charge trapped at time $t$ decreases exponentially with time

$$Q_{ox}(t) = (Q_{ox-0}) e^{-\kappa t}$$

(2.94)

where $e^d$ is the thermal emission coefficient ($sec^{-1}$), $t$ is the time in seconds (Equation (2.97)), and $Q_{ox-0}$ is the density of the oxide-trapped charge integrated over the thickness of the oxide per unit area of silicon-silicon dioxide (Si-SiO$_2$) interface at time $t=0$, i.e., at the onset of trap emptying given by Equations (2.95) and (2.96).

Two cases can exist at time $t=0$. First, the traps may be filled to saturation; and second, the traps may not be filled to saturation. If the traps are filled to saturation at time $t=0$ [Balland and Barbottin 1989],

$$Q_{ox-0} = q \int_{0}^{t_x} N_T(x) dx$$

(2.95)

If the traps are not filled to saturation at time $t=0$,

$$Q_{ox-0} = q \int_{0}^{t_x} n_T(x) dx$$

(2.96)

where $q$ is the absolute value of electronic charge($1.6 \times 10^{-19}$ Coulomb), $N_T$ is the spatial density of traps possessing level $E_T$ (cm$^3$), $n_T$ is the spatial density of filled traps of energy $E_T$, and $t_x$ is the oxide thickness [Balland and Barbottin 1989]. The thermal emission coefficient is a function of temperature given by
TEMPERATURE DEPENDENCE OF DEVICE FAILURE MECHANISMS

\[ e_n^* = \left( \frac{2\sqrt{3}(2\pi)^{3/2} m_n^* k^2}{\hbar^3} \right) \sigma_n T^2 e^{-E_n/kT} \]  \hspace{1cm} (2.97)

where \( m_n^* \) is the effective mass of the electrons in silicon dioxide (SiO₂), \( k \) is the Boltzmann constant \((8.617 \times 10^{-5} \text{ eV/K or } 1.38 \times 10^{-2} \text{ J/K})\), and \( \hbar \) is Planck’s constant \((6.62 \times 10^{-34} \text{ J.s})\) [Balland and Barbottin 1989].

**Hickmott Model:** Hickmott [1975] presented a model for thermally stimulated ionic conductivity due to trapped charges. The model is based on the assumption that at the start of the measurement, all the ions are present in the ion traps near one of the interfaces. When the temperature of the MOS capacitor is increased, the emission of ions from traps takes place at an increasing rate, which results in an increase in ionic current. After most of the ions have been emitted, the ionic current decreases. The flux of the ions emitted at time \( t \), based on the assumption that ion traps corresponding to one type of ions have a single energy, \( E_o \), is represented by

\[ -\frac{dN(t)}{dt} = s N(t) e^{-E_n/kT(t)} \]  \hspace{1cm} (2.98)

where \( N(t) \) is the density of ions trapped at time \( t \) \((\text{cm}^{-2})\), \( E_o \) is the single activation energy, \( s \) is the proportionality constant \((\text{sec}^{-1})\), and \( k \) is the Boltzmann constant \((8.617 \times 10^{-5} \text{ eV/K})\). The current density is represented by

\[ J_G(t) = q \frac{dN(t)}{dt} \]  \hspace{1cm} (2.99)

where \( q \) is the magnitude of electronic charge \((1.6 \times 10^{-19} \text{ coulomb})\), \( N(t) \) is the density of ions trapped at time \( t \) \((\text{cm}^{-2})\), and \( t \) is the time. The density of the ions still trapped follows from integration of Equation (2.98):

\[ N(t) = N_0 e^{-E_o/kT(t)} \]  \hspace{1cm} (2.100)

and the current density is represented by

\[ J_G(t) = q s N_0 e^{-E_o/kT(t)} \]  \hspace{1cm} (2.101)
The solution to equation (2.101) depends on how the temperature varies as a function of time, and exist for hyperbolic variation of temperature \(1/T = 1/T_0 - at\) [Hickmott 1975, Stagg 1977], and linear variation of temperature with time \(T = T_0 + bt\) [Hillen 1981]. Typically, temperature transients and temperatures in the neighborhood of 175°C or greater result in failures due to slow trapping [Gottesfeld 1984].

### 2.4.2 Gate Oxide Breakdown

Two forms of oxide breakdown can result from electrostatic discharge and electrical overstress or from time-dependent breakdown which occur during operation within rated conditions of voltage, temperature and power dissipation.

**Electrostatic discharge.** An integrated device can be considered equivalent to a circuit with multiple paths to the ground. When one pin is grounded, potentials sufficient to cause dielectric breakdown, junction shorts, or cracks between isolated regions may be discharged through the device. ESD discharges typically last less than 50 µs.

Electrostatic pulses can arise from contact with air, skin, glass or charge-carrying particles. Typically, ESD pulses damage gate oxides or melting of small amounts of the device, creating minute explosions, voids, cratering, and subsequent short circuit or open circuit on the device surface. Higher temperature result in a significant reduction in the electrostatic discharge resistance of the component [Kuo 1983, Hart 1980].

Typically, ESD failures result in fracture of the gate oxide in MOS devices, since voltage is in excess of the breakdown voltage (in bipolar devices, breakdown bulk occurs predominantly in the device). High currents through the breakdown site cause localized heating and usually produce a metal silicon alloy through the gate fracture site, forming a resistive short across the gate. The short can be a gate to drain short, source short, or substrate short, depending on the structure of and imperfections in the oxide. Most likely sites for ESD damage in defect-free oxides are the source or drain sites, depending on the polarity of the transient and biasing of the device. Gate-to-substrate shorts are more prevalent in devices with pre-existing oxide defects in the form of geometrical or dopant irregularities [McAteer 1989]. Typical PN-junction ESD damage occurs in reverse-biased conditions in the form of degraded PN-junction characteristics. The failed PN junction is characterized as cracked glass across the junction on the surface of the chip.
Metallization-open Voltage Threshold for ESD Failures: Typically, vaporized metal lines are indicators of EOS failures, but sufficiently large ESD pulses can also result in such failures. Open-circuit sites typically are at points of constriction, such as oxide steps. ESD failures causing metallization vaporization are uncommon, however, due to the presence in the discharge current path of other energy-absorbing elements which reduce the transient current through the metal below damaging current densities [McAteer 1989]. The condition for melting of non-stepped metallization stripes under adiabatic conditions is

\[ J^2 t_m = K_m \]  

(2.111)

where \( t_m \) is the time to melt in seconds, \( J_m \) is the current density in metallization (A/cm²), and \( K_m \) is a constant (Equation (2.112)).

\[ K_m = \frac{(\Delta T)H_c + H_f}{0.239\rho} \]  

(2.112)

where \( \Delta T \) is the temperature rise (= 660°C - 25°C = 635°C), \( H_c \) is the heat capacity per unit volume (= 0.637 cal/cm³ °C @ 240°C), \( H_f \) is the heat of fusion per unit volume (248.5 cal/cm³), and \( \rho \) is the volume resistivity of metallization (= 5.1 x 10⁴ Ω cm). The time constant of the RC circuit can be calculated based on resistance and capacitance (\( R = 1500 \Omega; C = 100 \text{ pF} \) values for a standard human-body model. The time to stripe melting can be approximated as five time constants:

\[ t_m(\text{max}) = 5 \tau \]  

(2.113)

The average current in the circuit can thus be calculated from

\[ J^2 = \frac{I^2}{A^2} = \frac{K_m}{t_m} = \frac{K_m}{5\tau} \]  

(2.114)

\[ I_{ave}^2 = \frac{K_m A^2}{5\tau} \]

The peak discharge current can be estimated from the modified Speakman model (Equation (2.108)) since the first term containing \( V_d \) does not apply):
FAILURE MECHANISMS IN THE DEVICE OXIDE

\[ P_{av} = \left( \frac{l_{\text{average}}}{R_{\text{average}}} \right)^2 \frac{R_f}{10} \]

where \( R_{\text{average(circuit)}} = (R_b + R_{\text{stripes}} + R_{\text{series}}) \), and \( R_{\text{series}} \) is any series load resistance in the circuit. The ESD voltage threshold then required for metallization melting is

\[ V = I_p R_{\text{average(circuit)}} \]

(2.116)

Devices are typically protected against ESD by protection structures shorting the ESD pulse voltage to ground or to the supply voltage to limit the current entering the critical junction. The damage voltages for protected devices are in the range of 3,000 to 9,000 V, depending on protection, as compared to as low as 100 V for unprotected devices. However, on-chip protection can reduce device performance.

**Time-dependent Dielectric Breakdown.** Time dependent dielectric breakdown (TDDB) is the formation of low-resistance dielectric paths through localized defects in dielectrics, such as thermally grown or other oxides in MOS devices only. Failures typically occur at weaknesses in the oxide layer due to poor processing or uneven oxide growth. Various studies have demonstrated a correlation between the low breakdown strength and the presence of stacking faults in the oxides [Lin 1983, Liehr 1988]. Other studies have attributed early oxide breakdown to charge accumulation in the oxide [Lee 1988, DiStefano 1975, Harai 1978, Ricco 1983, Holland 1984] and to local thinning and discontinuities in the oxide caused by metal precipitates [Honda 1984, 1985, Wendt 1989]. The mechanism is characterized by sudden, usually permanent, d.c. conduction in the dielectric of MOS capacitors. Typically, thin FET dielectric materials exhibit this breakdown failure mechanism, depending on latent defect density, temperature, electric field intensity, the ratio of the device operating potentials to the intrinsic dielectric strength, and the distances between the conductors in electronic packages defined by technology limits or electrical requirements.

**Fowler-Nordheim Tunneling-based Models for TDDB:** Fowler-Nordheim tunneling models are based on the failure of oxides due to TDDB as a consequence of
charge accumulation in the oxide. Breakdown of the oxide occurs when a critical charge density is reached in the oxide to trigger the breakdown process [Lee 1988, DiStefano 1975, Harari 1978, Ricco 1983, Holland 1984]. The critical charge density is proportional to the total number of electrons injected in the oxide and the probability of holes being generated by these electrons and trapped in the oxide [Chen 1985]:

\[ Q_{\text{critical}} = J \cdot t_{\text{BD}} \propto \eta \]  \hspace{1cm} (2.117)

where

\[ J = e^{-\frac{qE}{a}} \]
\[ a = e^{-\frac{qE}{\alpha}} \]  \hspace{1cm} (2.118)

where \( Q_{\text{critical}} \) is the critical charge density, \( J \) is the Fowler-Nordheim tunneling current, \( t_{\text{BD}} \) is the time to oxide breakdown, \( a \) is the hole-generation coefficient, and \( \eta \) is the hole-trapping efficiency (constant). Based on experimental data, the time to breakdown has been shown to be an exponential function of reciprocal of electrical field \( (E_{\text{ox}}) \) [Moazzami 1988, Lee 1988]:

\[ t_{\text{BD}} = \tau_0 e^{G/E_{\text{ox}}} \]
\[ = \tau_0 e^{G X_{\text{eff}}/V_{\text{as}}} \]  \hspace{1cm} (2.119)

where \( \tau_0 \) is the intercept of the \( \ln(t_{\text{BD}}) \) versus \( 1/E_{\text{ox}} \) plot, \( G \) is the slope of the \( \ln(t_{\text{BD}}) \) versus \( 1/E_{\text{ox}} \) plot \((G = B + H, \text{in Equation (2.118))}, X_{\text{ox}} \) is the oxide thickness, and \( V_{\text{as}} \) is the voltage across the oxide. In the case of defective oxides, the defects are modeled as localized oxide thinning. The electric-field dependence of time to breakdown for defective oxides is modeled as [Lee 1988]:

\[ t_{\text{BD}} = \tau_0 e^{G X_{\text{eff}}/V_{\text{as}}} \]  \hspace{1cm} (2.120)

where \( X_{\text{eff}} \) is the effective oxide thickness at the weakest spot in the oxide. This concept is also used to model asperities at the interface and localized areas having modified chemical composition, which may increase the charge-trapping rate or reduce the barrier height at the silicon/silicon dioxide (Si/SiO\(_2\)) interface. The natural logarithm of time to breakdown, \( \ln(t_{\text{BD}}) \), is a linear function of \( 1/E_{\text{ox}} \), and non-linear function of \( E_{\text{ox}} \) [Moazzami 1988, Lee 1988]. The electric-field acceleration factor is defined as the tangential slope of the \( \log_{10} t_{\text{BD}} \) versus \( E_{\text{ox}} \) plot [Lee 1988]:
\[ \gamma = \frac{d \log_{10}(t_{BD})}{d E_{ox}} = \frac{G}{(\ln 10) (E_{ox})^2 \left( \text{decades} \right)} \left( \frac{M\text{V} \text{cm}}{\text{MV} \text{cm}} \right) \]  

(2.121)

where \( \gamma \) is the electric field acceleration factor, \( t_{BD} \) is the time to oxide breakdown, \( E_{ox} \) is the electric field across the oxide, and \( G \) is the slope of the \( \ln(t_{BD}) \) versus \( 1/E_{ox} \) plot.

Recent studies have modeled the pre-exponential term, \( \tau_0 \), in Equations (2.119) and (2.120) as an exponential function of temperature with an activation energy \( E_b \) [Moazzami 1989, Moazzami 1990]:

\[ \tau_0(T) = \tau_0 e^{-\frac{E_b}{k \left( \frac{1}{300} \right)}} \]  

(2.122)

where \( \tau_0 \) is the room temperature value of the pre-exponential, \( \tau(T) \); \( E_b \) is the activation energy of the pre-exponential, \( \tau(T) \); and \( k \) is Boltzmann's constant \((8.617 \times 10^{-5} \text{ eV/K})\). The slope of the plot of \( \ln(t_{BD}) \) versus \( 1/E_{ox} \) \( G \), is also modeled as a Taylor expansion of a temperature exponential [Moazzami 1989, 1990]:

\[ G(T) = G \left( 1 + \frac{\delta}{k} \left( \frac{1}{T} - \frac{1}{300} \right) \right) \]  

(2.123)

where

\[ \delta = \frac{k d G(T)}{G d (1/T)} \]  

(2.124)

and \( G \) represents the room temperature value. Moazzammi et al. represented \( G \) by Equation (2.124) to allow \( t_{BD} \) to follow the Arrhenius relationship [1989]. The time to breakdown is thus represented by an apparent activation energy represented as follows:

\[ t_{BD}(T) = \tau_0 e^{\frac{E_{ad}}{k T}} \left( 1 + \frac{\delta}{k} \left( \frac{1}{T} - \frac{1}{300} \right) \right) \]  

(2.125)

where \( E_{ad} \) is the apparent activation energy represented by
The following values for $E_b$ and $\delta$ determined from experimental data:

<table>
<thead>
<tr>
<th>$\delta$</th>
<th>$E_b$</th>
<th>For</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0167 eV</td>
<td>0.28 eV</td>
<td>$25^\circ C &lt; T &lt; 125^\circ C$</td>
</tr>
<tr>
<td>0.024 eV</td>
<td>0.28 eV</td>
<td>$T &gt; 150^\circ C$</td>
</tr>
</tbody>
</table>

**Thermodynamic Models for TDDB:** McPherson proposed a thermodynamic model based on the assumption that when the dielectric breaks down, it undergoes an irreversible phase transition that transforms the material from an insulating phase to a conducting phase [1985]. The driving force for this transformation is the difference between the free energies of the conducting phase and of the insulating phase. The rate at which the reaction occurs is controlled by the free energy of activation associated with the growth of the conductive poly filament. McPherson represented a dielectric stored at a fixed field by a reaction rate constant, $k$:

$$k \propto \exp\left(\frac{-\Delta G^*}{K_B T}\right)$$  \hspace{1cm} (2.128)

where $\Delta G^*$ represents the free energy if activation associated with the breakdown processes. By treating the components as reactants and the broken-down components as reaction products, the time to failure of the dielectric is represented by McPherson as

$$T_F \propto \frac{1}{k} \propto \exp\left(\frac{\Delta G^*}{K_B T}\right)$$  \hspace{1cm} (2.129)

where $T_F$ is the time to failure, $k$ is the rate constant, $\Delta G^*$ represents the free energy if activation is associated with the breakdown process, $k$ is Boltzmann's constant ($8.617 \times 10^{-5}$ eV/K), and $T$ is the steady-state temperature.

The internal energy, $E$, of the dielectric under applied field stress is represented as
where $S$ represents the entropy of the system, $V$ is the dielectric volume, $N$ is the number of dipoles induced and/or oriented, and $P$ is the dielectric polarization. The Gibbs free energy is obtained from internal energy, using the Legendre transformation

$$G(T, E, \mu, p) = E - \left(\frac{\partial E}{\partial S}\right)_{V,N,P} - \left(\frac{\partial E}{\partial V}\right)_{S,N,P} - \left(\frac{\partial E}{\partial N}\right)_{S,V,P} - \left(\frac{\partial E}{\partial P}\right)_{S,V,N}$$

$$= E - TS + pV - \mu N - EP$$

$$= H_o - TS - \mu N - EP$$

where $H_o$ is the enthalpy of the dielectric. The Gibbs free energy is expressed in terms of intensive parameters including temperature, $T$; electric field, $E$; chemical potential, $\mu$; and pressure, $p$ (constant during dielectric stressing). On rearranging,

$$- TS - EP - \mu N = K_B T f(T) g(E) h(\mu)$$

where $f$, $g$, $h$, are functions of $T$, $E$, and $\mu$ respectively. Using thermodynamic relationships, the entropy, dielectric volume, and number of dipoles can be characterized by

$$S = - \left(\frac{\partial G}{\partial T}\right)_{E, \mu} = K_B g h f + T \frac{df}{dT}$$

$$P = - \left(\frac{\partial G}{\partial E}\right)_{T, \mu} = -K_B T f h d g \frac{dE}{dE}$$

$$N = - \left(\frac{\partial G}{\partial \mu}\right)_{T,E} = -K_B T f g h d \mu$$

Substituting the thermodynamic relationships into the above equation for Gibb’s free energy,
\[ \frac{T}{dT} \frac{df}{dE} + \frac{E}{dE} \frac{dg}{d \mu} + \frac{\mu}{d \mu} \frac{dh}{h} = 0 \] (2.136)

Solving the above equation,

\[ G(T,E,\mu) = H_0 + K_b T \sum \sum C_n E^* \mu^{n-1} \left( \frac{T}{T_0} \right)^m \] (2.137)

Considering the linear terms, McPherson represented the Gibbs free energy as

\[ \Delta G^* = \Delta H_0^* + K_b T \left( \frac{C_{01}}{\mu} + \frac{C_{11}}{T} \right) S \] (2.138)

\[ = \Delta H_0^* + K_b T \left( B + \frac{C}{T} \right) S \]

where \( \Delta H_0^* \) is the change in enthalpy required to activate the poly filament growth at breakdown, \( B \) and \( C \) are constants, \( S = E_b - E_s \), \( E_b \) is the breakdown strength of the dielectric, and \( E_s \) is the stressing in the dielectric. Thus,

\[ TF(f%) = Ae^{(\Delta H_0^* K_b T) \frac{1}{T_0}} \] (2.139)

where the field acceleration parameter, \( \gamma \), is the steady-state temperature-dependent parameter given by

\[ \gamma = B + \frac{C}{T} \] (2.140)

The effective activation energy of the field is given by

\[ (\Delta H)_{\text{eff}} = \Delta H_0^* + K_b C s \] (2.141)

**Empirical Models for TDDB:** Anolick [1981], Crook [1979], and Berman [1981] have proposed models to predict the time to failure based on these parameters. Secondary effects - especially dielectric thickness, electrode shape, materials, and other processing parameters - may also affect the dielectric breakdown mechanism. Anolick and Nelson [1979] modeled the time to failure as
where \( t_f(F) \) is the time to failure for \( F \) percent of the population, \( \gamma \) is the voltage form factor (determined by life testing), \( \Delta H \) is the activation energy, \( k \) is Boltzmann's constant, \( T \) is steady-state temperature, \( V_b(F) \) is the breakdown voltage for \( F \) percent of the population, \( V_a \) is the applied voltage, and \( A \) is a constant. The following table compares various models of time-dependent dielectric breakdown.

**Table 2.4 Comparison of Various Time Dependent Dielectric Breakdown Models**

<table>
<thead>
<tr>
<th>Reference</th>
<th>Oxide Thickness</th>
<th>Experimental Conditions</th>
<th>Observations</th>
<th>Model Predictions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anolick, Nelson</td>
<td>700 Å</td>
<td>( E_p = 1.3 \text{ MV/cm} ) ( E_s - E_p = 7 ) ( \text{MV/cm} )</td>
<td>( (\Delta H)_{\text{obs}} = 2.1 \text{ eV} ) ( \gamma = B + \frac{C}{T} )</td>
<td>( (\Delta H)_{\text{obs}} = 1.8 \text{ eV} ) ( \gamma = B + \frac{C}{T} )</td>
</tr>
<tr>
<td>1979</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Crook 1979</td>
<td>1100 Å</td>
<td>( E_p = 3.5 \text{ MV/cm} ) ( E_s - E_p = 3 ) ( \text{MV/cm} )</td>
<td>( (\Delta H)_{\text{obs}} = 0.3 \text{ eV} ) ( \gamma = 7 @ 25^\circ C )</td>
<td>( (\Delta H)_{\text{obs}} = 0.34 \text{ eV} ) ( \gamma = 6 @ 25^\circ C )</td>
</tr>
<tr>
<td>Berman 1981</td>
<td>( \geq 400 ) Å</td>
<td>linear ramp</td>
<td>( (\Delta H)_{\text{obs}} = 0.29 (E_p - E_s) ) ( \gamma = -5.4 + \frac{0.29}{T} )</td>
<td>( (\Delta H)_{\text{obs}} = 0.29 (E_p - E_s) ) ( \gamma = -5.4 + \frac{0.29}{T} )</td>
</tr>
<tr>
<td>Hokari 1982</td>
<td>100 Å</td>
<td>( E_p = 5 - 7 \text{ MV/cm} ) ( E_s - E_p = 5 ) ( \text{MV/cm} )</td>
<td>( (\Delta H)_{\text{obs}} = 1.0 \text{ eV @ } 6 ) ( \text{MV/cm} ) ( \gamma = 1.7 @ 250^\circ C )</td>
<td>( (\Delta H)_{\text{obs}} = 1.0 \text{ eV @ } 6 ) ( \text{MV/cm} ) ( \gamma = 1.5 @ 250^\circ C )</td>
</tr>
<tr>
<td>McPherson 1985</td>
<td>100 Å</td>
<td>( E_p = 6 - 8 \text{ MV/cm} ) ( E_s - E_p = 3 - 5 ) ( \text{MV/cm} )</td>
<td>( (\Delta H)_{\text{obs}} = 0.3 - 1.0 \text{ eV} )</td>
<td>( (\Delta H)_{\text{obs}} = 0.3 \text{ eV} ) ( \gamma = B + \frac{C}{T} ) ( \gamma = B + \frac{C}{T} )</td>
</tr>
</tbody>
</table>

Time-dependent dielectric breakdown has been a significant failure mechanism in...
metal-gate MOS integrated circuits, silicon-gate transistors and integrated circuits with two or more levels of polysilicon, and circuits with trench capacitors. The dielectric breakdown phenomenon has also been shown to be an important factor affecting the early life of dynamic memory devices [Barrett 1978]. TDDB is a function of voltage stress in thin-gate oxides, and also a function of temperature (Figure 2.18). TDDB has both a dominant voltage dependence and a weak temperature dependence. Typically, TDDB has a very low thermal activation energy (say 0.3 eV) but a very large voltage acceleration $10^7 \text{/megavolt/cm}$ [Crook 1978, 1979, Schnable 1988]. The activation of TDDB is a function of stressing
the electric field. The field acceleration is itself a function of temperature [McPherson 1985]. The apparent activation energy of TDDB decreases from 1 eV at low field stressing \((E_b - E_s > 5 \text{ MV/cm})\) to 0.3 eV at higher fields \((E_b - E_s < 3 \text{ MV/cm})\) (where \(E_b\) is the dielectric breakdown strength and \(E_s\) is the stressing electric field). The field acceleration also reduces from 6 decades/MV/cm at 25 °C to 2 decades/MV/cm at 150 °C [McPherson 1985] (see Figure 2.19). The projected failure rate due to time-dependent dielectric breakdown under temperature stress was found to be too low to be of any importance for 10 V and 55°C. At 10 V and 200°C operation, the projected lifetime was \(3 \times 10^6\) years, which represents a failure rate of \(10^2\) failures in \(10^9\) years [McPherson 1985, Boyko 1989, Swartz 1986]. For this reason, device screening procedures involve stressing the devices at higher voltages and room temperature, rather than at normal biases and higher temperatures [Crook 1978, 1979, Schnable 1988].

2.4.3 Electrical Overstress

Electrical overstress occurs when a higher-than-rated voltage or current induces a hot-spot temperature beyond specifications for short periods of time [Alexander 1978, Canali 1981, Smith 1978]. Hot-spot development typically occurs at a semiconductor junction as the current flow increases to accommodate the additional stress in the device. As the junction heats up, the increased temperature encourages even greater current flow, as the silicon resistance lowers at higher temperature; this in turn further heats the junction. The fundamental cause of failure is joule heating due to power dissipation along current-conducting paths. If the material reaches its melting temperature or its eutectic temperature, permanent damage may result. In a typical device, the greatest power dissipated per unit volume is either in the depletion region of the junction or in the lightest doped material. The temperature rises due to power dissipation, resulting in decreased silicon resistivity.

Runayan [1965] plotted the resistivity of silicon as a function of temperature for a variety of doping concentrations. Silicon initially exhibits a positive temperature coefficient of resistivity but reaches a peak value in the neighborhood of 160°C, and thereafter exhibits a large negative temperature coefficient of resistivity. As the temperature of a region rises, the initial increase in the resistance of the region tends to spread the current to the cooler region (Figure 2.20). If the situation continues, the hot-spot temperature may exceed the intrinsic temperature of silicon, beyond which the resistivity of silicon decreases greatly. This allows more current to pass through the hot spot, further increasing the temperature, and resulting in thermal runaway. That is, the temperature of the hot spot rises suddenly, while
Figure 2.19  The Electric Field Acceleration Parameter is inversely dependent on temperature [McPherson and Baglee, 1985]
Figure 2.20 Silicon resistivity versus temperature. [Runyan, 1965]
the resistance of the device drops. The process continues until the silicon in the hot spot melts, destroying the silicon crystal structure. If the electrical transient continues, the interconnects melt as well. If the transient is of high voltage, an electrical arc may occur when the metallization opens. The temperature in the vicinity of the arc can cause the metallization to vaporize. The energy required to raise the junction to an unstable temperature can be calculated using a power model [Pancholy 1978].

2.5 TEMPERATURE DEPENDENCIES OF FAILURE MECHANISMS IN THE DEVICE

2.5.1 Ionic Contamination

Ionic contamination causes reversible degradation phenomena, such as threshold voltage shift and gain reduction, due to the presence of mobile charge ions within the oxide or at the device-oxide interface, typically at temperatures above 100°C [Schnable 1988, Brambilla 1981, Johnson 1976].

P-channel devices are less sensitive to ionic contamination than n-channel devices. Figure 2.21 shows the variation of ionic current versus temperature in devices with Na\(^+\) and K\(^+\) ion implants (the curves have been generated using TSIC technique\(^1\)). The ionic current increases until 100°C for Na\(^+\) ions, and decreases to a negligibly low value for temperatures higher than 200°C. The K\(^+\) ion, on the other hand, shows two peaks in ionic current at temperatures in the neighborhood of 100°C and 300°C (Figure 2.21). The ionic current due to mobile ions is thus a complex function of steady-state temperature [Hillen 1986, Boudry 1979, Nauta 1978, Derbenwick 1977].

The mobility of the ions is temperature-dependent [Hemmert 1980, 1981]. Consequently, high-temperature storage bake and exposure to high temperature during burn-in screen out ionic contamination failures [Hemmert 1980, 1981, Bell 1980]. In particular, the position of ionic charges relative to the silicon-silicon dioxide (Si-SiO\(_2\)) interface greatly influences the effectiveness of altering the threshold voltage, \(V_T\). A uniformly distributed contaminant (Na\(^+\)) within the oxide is redistributed to the silicon (Si) channel surface due to the influence of an electric field, \(E\), and thermal activation energy. The change in threshold voltage as a function of steady-state temperature, \(T\), has been modeled by Wager [1984]:

---

\(^1\) Three techniques are generally used to measure the ionic current due to mobile charges versus temperature. These include BTS (Bais temperature stress), TSIC (Thermally stimulated Ionic Current), and TVS (Triangular Voltage Sweep).
\[ \Delta V_T = E \frac{q_1 q_2}{2} e^{-\Delta H/kT} \]  

(2.143)

where \( E \) is the electric field, \( t \) is the time under bias, \( \Delta H \) is the activation energy, and \( K \) is the Boltzmann constant. Higher chip temperature would result in more rapid \( \Delta V_T \) change and shorter times to saturation [Wager 1984].

N-channel MOSFETs are covered with phosphosilicate glass films to stabilize the threshold voltage against changes resulting from ionic contaminants (sodium). Hemmert investigated the temperature bias kinetics for sodium ion drift in phosphosilicate glass [Hemmert 1980, 1981]. Microelectronic devices may be subject to a variety of defects; one such defect is exposed gate oxide resulting from extraneous etched holes contiguous to the gate that subject the phosphosilicate glass film to either a high local ionic concentration or physical damage. The temperature at which ionic drift occurs in defective devices is not as large as in non-defective devices. Hemmert conducted experiments to measure ionic stability on a 1.5-mm MOS consisting of gate oxide, phosphosilicate glass film, and aluminum-copper (Al-Cu) metallurgy. He found that below 250°C, the plot of voltage versus temperature shows a polarization plateau that represents saturation PSG polarization voltage. The activation energy, \( \Delta H \), was 2.0 eV. This value is uncharacteristically high, compared to the value usually reported for sodium-related failure mechanisms (\( \Delta H = 1.0 \) eV); the cumulative percent defect predictions were found to be more accurate assuming a temperature-dependent defect level. In the temperature range of 250-350°C, flatband voltage shifts occurred. This temperature, called the break temperature, occurs when the contribution of flatband voltage shift to ionic drift and to polarization are equal. The maximum variation in the flat-band voltage is related to the mobile charge by the relation

\[ Q_m = C_x \Delta V_{FB} \]  

(2.144)

where \( C_x \) and \( Q_m \) represent the oxide capacitance and the mobile charge per unit area. The threshold voltage, \( V_T \), and the flatband voltage, \( V_{FB} \) are related by the equation

\[ V_T = V_{FB} + 2\phi_F + \gamma \sqrt{2\phi_F} \]  

(2.145)

where \( \phi_F \) is the difference in the Fermi level and the intrinsic Fermi level in the bulk of the semiconductor. It is evident that the threshold voltage depends on the
flatband voltage. Thus, the usual assumption that the defect level is temperature-independent was found to be invalid.

A common characteristic of contamination failures is reversibility under high temperatures (in the neighborhood of 150 to 200°C), which partially or fully restores device characteristics [Hemmert 1980, 1981, Bell 1980]. However, this simply reflects a disordering of charge accumulations resulting from ion mobility and applied bias. The problem of ionic contamination has been solved to a large extent by the use of high-purity materials and chemicals for processing, use of HCl during oxidation [Robinson and Heiman 1971], and use of phosphosilicate glass (PSG) or borophosphosilicate glass (BPSG) over the polysilicon to getter the ions [Schnable 1988]. Passivation layers can provide additional protection against the ingress of alkali ion contamination in completed devices [Schnable 1988].

2.5.2 Second Breakdown

Second breakdown is the transition to a state of higher conductance in a reverse-biased avalanching semiconductor junction. Second breakdown occurs predominantly in bipolar devices when the bias current density reaches a threshold value; in this case, the operating temperature in some region of the pn junction becomes high enough that thermally generated carriers can take over the conduction process. The avalanche is thus thermally quenched, and second breakdown occurs. A negative temperature coefficient of resistance, associated with thermally generated current, results in the current constricting and flowing through a narrow region in the junction, where the temperature rises significantly. If the temperature of the constriction is higher than $T_p$, (the temperature at which the resistivity of semiconductor materials is the highest), the high-power density region at the perimeter of the constriction will heat up, becoming intrinsic, and the constriction will elongate into the ohmic region. If the constriction temperature is less than $T_p$, the constriction will not extend itself into the ohmic region, and the current will flow through the constriction in the junction and fan out. Second breakdown is a limiting phenomenon in power transistors that are switched in the presence of inductive loads, and typically involves three stages [Budenstein 1972, Sunshine 1970, Schafft and French 1966, Reich and Hakim 1966, Chiang and Lauritzen 1970, Thornton and Simmons [1958]:

- nucleation of filament;
- growth of a relatively broad filament across the high resistivity region;
- growth of a second filament interior to the first, in which material is in molten state.
While, nucleation and growth of filament are non-destructive processes, formation of melt is damaging and irreversible. Typically, nucleation is accompanied by a negligible voltage drop. In reverse-biased diodes, the p-n junction is the hottest region and the place where the filaments nucleate. The increase in the temperature of the p-n junction increases the avalanche voltage, due to decrease in the collision ionization coefficients. However, as the temperature rises, the reverse saturation current decreases rapidly. Total saturation current is insensitive to the size of reverse saturation current until the latter is eight-tenths of the former; then the avalanche voltage drops to zero, and the reverse saturation current density equals the total current density, resulting in the initiation of filamentation [Budenstein 1972, Sunshine 1970, Chiang and Lauritzen 1970].

Filament growth involves the development of a moderately hot filament across the high-resistivity region. When the broad filament completely bridges the high resistivity region, melt initiates. Formation of melt involves the development of a melt channel through the hot central portion of the broad filament. The melt channel alters the device irreversibly, the degree of degradation depending on the size and location of melt. Broad filaments typically range from 25 to 50 μm across, while the melt channel, at currents near its formation threshold, is about 1 μm in diameter. Growth and formation of melt, however, is accompanied by appreciable voltage drop. Junction nucleation occurs when the reverse saturation current density at the local site on the junction becomes equal to the total current density at that site [Budenstein 1972, Sunshine 1970, Chiang and Lauritzen 1970]. Typically there is a time difference between the application of the high current pulse and second breakdown, called the delay time.

Second breakdown is characterized by two different failure mechanisms: forward second breakdown, and reverse second breakdown. Both forward-and reverse-bias second breakdown of diodes is accompanied by current filamentation and a high current transition. In reverse bias, the filament typically starts at the junction and reaches into the high-resistance side of the junction to the electrode. In forward bias, the filament typically starts well within the interior of the high-resistance region. The transition from avalanche to second breakdown of the base-collector junction occurs at a lower $V_{ce}$ when the emitter is forward-biased than when it is reverse-biased. Second breakdown damage manifests itself in the form of changes in I-V characteristics due to resistive shorts of one or both transistor junctions. The critical condition for filament formation occurs when the temperature in the high-resistance side of the junction reaches the value at which the resistivity of the semiconductor material is at maximum. Filaments are formed during forward and reverse bias at the same current levels. The onset of negative resistance occurs before the filament is clearly distinguishable from its
FAILURE MECHANISMS IN THE DEVICE


Forward Second Breakdown. Forward second breakdown is typically a result of thermal runaway at a point in the transistor. Current density and temperature increase jointly till thermal runaway occurs. A power transistor can be considered an aggregate of many elementary transistors that operate under different conditions. When a power pulse is applied to the transistor, the junction temperature rises. Since the resistivity of the silicon increases with temperature, the temperature rise causes the current to increase:

\[ dP_1 = V_{ce}dI_c \] (2.146)

The increase in temperature causes an increase in the power dissipation:

\[ dP_2 = \frac{dT}{R_{th}} \] (2.147)

If \( dP_1 < dP_2 \), a thermal equilibrium is reached. If, on the contrary, \( dP_1 > dP_2 \), the temperature and current increases enhance each other and thermal runaway occurs, causing transistor destruction.

Usually temperatures above 160°C are required to cause device failure due to second breakdown. The major cause of device destruction is localized heating. Second breakdown has also been found to occur in vertical power MOSFETs.

Reverse Second Breakdown. Reverse second breakdown is an internal phenomenon that occurs in power transistors during switching operations on an inductive load, and is characterized by a precipitous drop in voltage and a rapid increase in current. The second breakdown voltage increases mildly with an increase in temperature in the range of 25°C to 150°C. The breakdown voltage increases from 650v to 680v when the temperature increases from 25°C to 150°C [Beatty 1976].

Mathematical Model for Second Breakdown: Budenstein’s model [1972] assumes that the formation of a melt channel within the broad filament that spans the high-resistivity region marks the initiation of second breakdown. The model assumes that:
bullet the central portion of the filament is at the melting point of the silicon and the filament grows radially, while keeping a constant length. This allows the use of a cylindrical geometry for the heat transfer problem.

bullet the filament is about 1 \( \mu \text{m} \) in diameter at the end of transition for current amplitudes close to the threshold for filamentation.

bullet the filament is in parallel with a fixed resistance \( R_b \), of the remainder of the device.

The power dissipated in the filament is equal to the rate that the energy is absorbed in the latent heat of fusion plus the rate at which the energy is lost by conduction through the boundary surface of the filament:

\[
P_f = \frac{d(LM_f)}{dt} + CA_s \tag{2.148}
\]

where \( P_f \) is the electrical power dissipated in the filament, \( L \) is the latent heat of fusion for the semiconducting material, \( M_f \) is the mass of the molten filament (Equation (2.149)), \( C \) is the heat transfer coefficient (heat flow per unit time per unit area), \( A \), is the lateral surface area of the filament, and

\[
M_f = DH\pi r^2 \tag{2.149}
\]

\[
R_f = \frac{\rho_f H}{\pi r^2} \tag{2.150}
\]

\[
A_s = 2\pi rH \tag{2.151}
\]

\[
R_b = \frac{\rho H}{A} \tag{2.152}
\]
\[ P_f = I_f^2 R_f = \frac{I_f^2 R_b R_f}{(R_b + R_f)^2} \]  \hspace{1cm} (2.153)

Substituting Equations (2.155) to (2.159) into Equation (2.154),

\[ \frac{dr}{dt} = \frac{C}{LD} \left( \frac{I_f^2 R_b \rho_f}{2C} \frac{r}{(R_b \pi r^2 + \rho_f H)^2} - 1 \right) \]  \hspace{1cm} (2.154)

where \( r \) is the radius of the filament, \( C \) is the heat transfer coefficient, \( L \) is the latent heat for fusion of the semiconducting material, \( D \) is the mass density of the semiconducting material, \( I \) is the total current (sum of the currents through the filament and the semiconductor device), \( \rho_f \) is the resistivity of the filament, \( H \) is the length of the filament. The steady-state condition \( (dr/dt = 0) \) is assumed to be reached when power dissipated in the filament is at a maximum - that is, when \( R_f = R_b \) (obtained by differentiating Equation (2.153) with respect to \( R_f \)). The power dissipated in the filament in a steady-state condition is given by

\[ P_f' = CA_f' = \frac{V_f^2}{R_f} \]  \hspace{1cm} (2.155)

where the primes represent the steady-state values of the respective quantities. At steady-state \( (dr/dt = 0) \), the total threshold value of current for filamentation is

\[ I_t = 2\sqrt{2C} \left( \frac{\rho_f H^3}{R_b^3} \right)^{1/4} \]  \hspace{1cm} (2.156)

Substituting Equation (2.151) and (2.152), into Equation (2.155) and using the steady-state condition, \( R_b = R_f \),

\[ J_t = K \rho^{-3/4}, \text{where} \ K = 2\sqrt{2C} \rho_f^{1/4} A^{-1/4} \]  \hspace{1cm} (2.157)

where \( J_t \) is the threshold current density for filamentation, \( \rho \) is the resistivity of the semiconducting material outside the filament, and \( A \) is the area of the remainder of the device.

2.5.3 Surface-Charge Spreading
Surface-charge spreading, largely observed in MOS and memory devices, involves the lateral spreading of ionic charge from the biased metal conductors, along the oxide layer or through moisture on the device surface [Edwards 1982, Blanks 1980, Stojadinovic 1983]. An inversion layer outside the active region of the transistor is formed due to the charge, creating a conduction path between the two diffused regions or extending the p-n junction through a high-leakage region, resulting in leakage currents between neighboring conductors. The rate of charge spread increases with temperature. Surface-charge spreading failure, a wear-out mechanism, is usually observed at temperatures around 150°C to 250°C [Lycoudes 1980]. There are no existing physics-of-failure models to predict surface-charge spreading failures.

2.6 TEMPERATURE DEPENDENCIES OF FAILURE MECHANISMS IN THE DEVICE OXIDE INTERFACE

2.6.1 Hot Electrons

Hot electrons, or holes, are charge carriers that acquire energies from very high electric fields in excess of those that would be indicated by lattice or ambient temperature. Hot electrons are a phenomenon largely prevalent in high-density, small geometry, MOS memory devices because of small conduction channels and high voltages. It is accelerated at temperatures below 0°C [Stojadinovic 1983, Woods 1980]. Threshold voltage shifts or transconductance degradation can occur as renegade charge concentrations build up over time [Ning 1979]. Penetration of the oxide by hot electrons can lead to excess gate and substrate currents [Takeda 1983]. There are three possible mechanisms by which the electrons or holes are injected from silicon into the silicon dioxide (SiO₂) [Garrigues and Balland 1986]:

- **Direct tunnel emission**: Injection of electrons by direct tunneling is possible in thin oxides ( < 100 Å) and large electric fields (ε_x > 5 x 10⁵ V cm⁻¹).
- **Field assisted tunneling**: Electrons of energy close to but less than the energy barrier at the silicon-silicon dioxide (Si-SiO₂) interface may tunnel through the triangular energy barrier resulting from field in the oxide. For larger electric fields, the energy bands of silicon are so slanted that direct tunneling can occur, regardless of oxide thickness (field emission or Fowler-Nordheim tunneling).
- **Injection over the barrier**: Electrons may gain high potential energy due to high electric fields present in the conduction channel and cross the potential barrier at the substrate-oxide interface. An electron in conduction band of silicon (Si)
can be injected over the energy barrier into the conduction band of silicon dioxide \((\text{SiO}_2)\) if its energy \(\Delta E\) is greater than the difference in the conduction band energies of the conduction bands in the silicon (Si) and silicon dioxide \(\text{SiO}_2\).

Electron injection through field-assisted tunneling or injection over the barrier is possible by stimulating the electron thermally, optically, or with an electric field; field stimulation is the most common form of electron stimulation [Garrigues and Balland 1986].

Figure 2.22 shows the effect of ambient temperature on the hot carrier mechanism [Matsumoto 1981]. The temperature dependence of substrate current at fixed bias can have three origins: temperature-dependent channel current, temperature-dependent local electric field, and temperature-dependent ionization rate. The low field mobility follows \(-3/2\) power dependence on the temperature for typical n-channel MOSFETs [Sze 1981]. The reduced mobility increases the saturation voltage, which lowers the maximum-channel electric field at fixed-drain bias and results in a lower substrate current [Ko 1980, Hu 1983]. Tam et al. [1983] proposed that the device degradation rate can be related to the substrate current by

\[
\Delta V_T = C_1 \left( C_2 [I_s]^\alpha \right)^\beta
\]

(2.158)

where \(C_1\) is a structure-related factor, \(C_2\) is a material-related factor, \(\alpha\) is the energy factor (typically 2.9), and \(\beta\) is the factor from general kinetics (typically 0.6). The device degradation due to hot electrons very mildly decreases with the increase in temperature for temperatures between 20°C and 100°C, as shown in Figures 2.23 and 2.24 [Hsu 1984]. A plot of device lifetime as a function of temperature and substrate current is shown in Figure 2.25 [Hsu 1984]. A calculation of energy and proportionality factors in Tam's equation [1983] indicate that they are constant over the temperature range, suggesting that the physical mechanism for device degradation is the same over the temperature range 20°C to 100°C. The variation of lifetime versus substrate current for temperatures 20°C and 100°C, shown in Figure 2.25, indicates that hot-electron degradation is temperature-independent in this range.

The increase in the ambient temperature decreases the observed \(\Delta V_T\) shift from the channel hot-electron effect. For a given device design, Matsumoto has described this behavior as [1981]:
\[ \Delta V_T = AN_T(1 - e^{-BN_T}) \]  

(2.159)

where \( N_T \) is the effective oxide-trap density, \( N_{\text{inj}} \) is the density of injected electrons, and \( A \) and \( B \) are constants. From the equation, it is evident that the hot-electron threshold voltage shift increases with both the electrons injected into the oxide and the availability of traps to capture the electrons. Both these quantities decrease with increasing temperature, giving rise to reduced threshold shift due to hot electron trapping at elevated temperatures [Matsumoto 1981].

The Lucky Electron Model: This model ignores the overall description of the hot-carrier energy distribution and focuses on the distribution tail corresponding to those electrons with sufficient energy to overcome the potential barrier. The lucky electron model accounts for the influence of the electric field profile in silicon, the oxide field, and the lattice temperature.

A free electron is injected into silicon dioxide (SiO\(_2\)) if it reaches the interface with a component of its momentum, normal to the surface, greater than a critical value, \( p_c \), corresponding to the barrier energy, \( \phi_{\text{BS}} \). The probability of tunnel injection through the top of the barrier for an electron with momentum slightly lower than \( p_c \) is momentarily neglected. The electron can reach the interface with momentum \( p_c \) after travelling various possible trajectories. The trajectory consists of a series of free flights interrupted by various interactions that can be elastic or inelastic, during which the electron energy is modified by the electric field. The lucky electron model assumes that the most probable trajectories are those for which energy, \( \phi_{\text{BS}} \), is gained by the electron in the last free flight without any collisions before reaching the interface. The emission probability is thus close to
Figure 2.21
Influence of Temperature on Ionic Current due to Sodium and Potassium Ions [Hillen, 1986]
Change in threshold voltage due to hot electrons is much greater at lower temperature than at high temperatures. [Matsumoto, 1981]

Figure 2.22

- $T = -40 \text{ deg.C}, N_T = 2\times10^{12}$
- $T = 30 \text{ deg.C}, N_T = 1\times10^{12}$
- $T = 100 \text{ deg.C}, N_T = 5\times10^{11}$
Figure 2.23 Substrate characteristics for a 5 μm device at 20 deg.C and 100 deg.C
$V_g = 3V$ and $7V$, $V_t = 0.9V$ [Hsu, 1984]; The variation in device characteristics between 20 and 100 deg.C indicates that degradation due to hot electrons is almost temperature independent in this range.
Figure 2.24  Drain characteristics for a 5 μm device at 20 deg.C and 100 deg.C
V_g = 3V and 7V, V_t = 0.9V [Hsu, 1984]
The variation in device characteristics between 20 and 100 deg.C indicates
that degradation due to hot electrons is almost temperature independent
in this range.
Figure 2.25  Substrate current dependence of device lifetime at 20 deg.C and 100 deg.C [Hsu, 1984]; The small variation in device lifetime for temperature variation between 20 deg.C and 100 deg.C clearly indicates that the degradation due to hot electrons is temperature independent in this range.

Legend:
- $T = 15.5 \text{ nm}$
- $W/L = 100/1.5 \mu\text{m}$
- $V_b = 0\text{V}; V_g = 3\text{V}; \Delta V_{th} = 10 \text{ mV}$
the probability of the electron making such a trajectory. The minimum path length required to gain energy, before reaching the interface equals \( d \), the horizontal distance between the top of the barrier and the lower edge of the conduction band. It is assumed that the emission probability for an electron that has collision within a distance, \( d \), is negligible, irrespective of its previous energy. According to the lucky electron model, the emission probability of an electron is thus equal to its probability of travelling a distance greater than \( d \) without a collision [Ning 1977, Garrigues 1981]:

\[
P_e = P_0 e^{-\frac{d}{\lambda_{\text{min}}\left(\frac{E_a}{2kT}\right)}}
\]

where

\[
d = \sqrt{\frac{2e_{\text{sc}}\varepsilon_0}{qN_A} \left(\frac{1}{\Psi_S} - \frac{\Phi_{BZ}}{q}\right)}
\]

and \( P_e \) is the emission probability of an electron, \( P_0 \) is a constant (= 2.9 @ 300°K; 4.3 @ 77°K), \( d \) is the minimum path length of an electron required to attain critical energy \( \phi_{BZ} \), \( T \) is the steady-state temperature, \( E_a \) is the activation for mean free-path length (0.063 eV), \( \lambda_{\text{min}} \) is a constant (= 108 Å), \( \lambda \) is the electron mean free path between lattice interactions, \( K \) is Boltzmann's constant (8.617 x 10^{-5} eV/K or 1.38 x 10^{-23} J/K), \( \varepsilon_{\text{sc}} \) is the dielectric constant of the semiconductor (F cm^{-1}), \( \varepsilon_r \) is the free-space permittivity (8.85 x 10^{-14} F cm^{-1}), \( N_A \) is the concentration of acceptor doping atoms, \( q \) is the magnitude of electronic charge (1.6 x 10^{-19} Coulomb), \( \Psi_S \) is the surface potential (volts), and \( \Phi_{BZ} \) is the silicon-silicon dioxide (Si-SiO_2) barrier height for electrons, taking into account lowering due to the Schottky effect.
<table>
<thead>
<tr>
<th>Failure site</th>
<th>Failure mechanism</th>
<th>Dominant temperature dependence</th>
<th>Nature of steady state temperature dependence</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>wire</td>
<td>flexure fatigue</td>
<td>$\Delta T$</td>
<td>Independent of steady-state temperature function under normal operation</td>
<td>[Gaffney 1968], [Villela 1970], [Pecht 1989], [Harman 1974]</td>
</tr>
<tr>
<td>wirebond</td>
<td>Kirkendall voiding</td>
<td>$T$</td>
<td>independent of steady-state temperature below 150°C; independent of steady-state temperature above lower temperatures ($T &lt; 150^\circ$); in presence of halogenated compounds</td>
<td>[Newcombe 1976], [Philosky 1970, 1971], [Gerling 1984], [Khan 1986] [Pinnel 1972], [Feinstein 1979], [Pitt 1982], [Khan 1986], [Villela 1971]</td>
</tr>
<tr>
<td>die</td>
<td>fracture</td>
<td>$\Delta T, \nabla T$</td>
<td>primarily dependent on temperature cycle</td>
<td>[Tan 1987], [Hawkins 1987]</td>
</tr>
<tr>
<td>die adhesive</td>
<td>fatigue</td>
<td>$\Delta T$</td>
<td>independent of steady-state temperature under normal operation</td>
<td>[Chiang 1984], [Mahalingham 1984]</td>
</tr>
<tr>
<td>encapsulant</td>
<td>reversion</td>
<td>$T$</td>
<td>Independent of steady-state temperature below 300°C (glass transition for typical epoxy molding resin for plastic packages)</td>
<td>[Tummala 1989]</td>
</tr>
<tr>
<td></td>
<td>cracking</td>
<td>$\Delta T$</td>
<td>independent of steady state temperature ($T$) below the glass-transition temperature of the encapsulant; a $\Delta T, \nabla T$; driven mechanism.</td>
<td>[Nishimura 1987], [Fukuzawa 1985], [Kitano 1988]</td>
</tr>
<tr>
<td>package</td>
<td>stress corrosion</td>
<td>$dT/dt$</td>
<td>mildly steady-state temperature dependent under normal operation</td>
<td>[Tummala 1989]</td>
</tr>
<tr>
<td>Die metallization</td>
<td>dT/dt</td>
<td>Description</td>
<td>References</td>
<td></td>
</tr>
<tr>
<td>------------------</td>
<td>-------</td>
<td>-------------</td>
<td>------------</td>
<td></td>
</tr>
<tr>
<td>corrosion</td>
<td></td>
<td>only occurs above dew point temperature; mildly steady-state temperature dependent under normal operation</td>
<td>[Pecht 1990], [Commzizalli 1980], [Inayoaki 1979], [Sim 1979], [White 1969], [Schnable 1969]</td>
<td></td>
</tr>
<tr>
<td>hillock formation</td>
<td>T</td>
<td>Hillocks in die metallization can form as a result of electromigration or extended periods under temperature cycling conditions (thermal aging). Extended periods in the neighborhood of 400°C produce hillocks.</td>
<td>[LaCombe 1982], [Thomas 1983], [Amerasekera 1987]</td>
<td></td>
</tr>
<tr>
<td>metallization migration</td>
<td>T</td>
<td>independent of steady-state temperature below 500°C</td>
<td>[Diagiacomo 1982], [Bart 1969], [Lane 1970]</td>
<td></td>
</tr>
<tr>
<td>contact spiking</td>
<td>T</td>
<td>independent of steady-state temperature below 400°C</td>
<td>[Chang 1988], [Farahani 1987], [T.I. 1987], [DeChairo 1981], [Christou 1980, 1982], [Ballamy 1978],</td>
<td></td>
</tr>
<tr>
<td>constraint cavitation</td>
<td>T</td>
<td>steady-state temperature dependent above 25°C</td>
<td>[Yost 1988, 1989]</td>
<td></td>
</tr>
<tr>
<td>die electrical overstress</td>
<td>T</td>
<td>independent of steady-state temperature below 160°C (the temperature at which the coefficient of thermal resistance changes sign)</td>
<td>[Alexander 1978], [Canali 1981], [Smith 1978], [Runyan 1965], [Pancholy 1978]</td>
<td></td>
</tr>
<tr>
<td>Failure Mechanisms in the Device Oxide Interface</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----------------------------------------------</td>
<td>---</td>
<td>---</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Device Oxide</strong></td>
<td><strong>T</strong></td>
<td><strong>steady state temperature dependent above 175°C</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Slow Trapping</td>
<td><strong>T</strong></td>
<td>ESD voltage (resistance to ESD) reduces with temperature increase (from 25°C to 125°C); not a dominant mechanism in properly protected devices.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Electrostatic Discharge</td>
<td><strong>T</strong></td>
<td>steady-state temperature dependence is very weak; TDDB is a dominant function of voltage</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time Dependent Dielectric Breakdown</td>
<td><strong>T</strong></td>
<td>exogenous temperature-dependent above 200°C</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ionic Contamination</td>
<td><strong>T</strong></td>
<td>independent of steady-state temperature below 160°C</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Forward Second Breakdown</td>
<td><strong>T</strong></td>
<td>insignificant dependence of steady-state temperature; the breakdown voltage increases from 650V to 680V when the temperature increases from 25°C to 150°C</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reverse Second Breakdown</td>
<td><strong>T</strong></td>
<td>steady-state temperature dependent above 150°C</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Surface Charge Spreading</td>
<td><strong>T</strong></td>
<td>steady-state temperature dependent above 150°C</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Device Substrate-Oxide Interface</strong></td>
<td><strong>T</strong></td>
<td>steady-state temperature dependence decreases above -55°C. Temperature-independent in range of 20°C to 100°C.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

References:
<table>
<thead>
<tr>
<th>Failure mechanism</th>
<th>Nature of temperature dependence</th>
</tr>
</thead>
<tbody>
<tr>
<td>bondpad-substrate shear fatigue</td>
<td>( N_s = \frac{q}{\varepsilon' \tau_{\text{eq}}} )</td>
</tr>
<tr>
<td></td>
<td>( \tau_{\text{eq}} = Q \Delta T )</td>
</tr>
<tr>
<td></td>
<td>( Q = \left( \frac{G_p \Delta T}{b_p Z} \right) \left( \varepsilon - \varepsilon_0 \right) \left( \varepsilon - \varepsilon_0 \right) \left( 1 + \frac{E_s A_s}{E_p A_p \left( 1 - \nu_s \right)} \right) )</td>
</tr>
<tr>
<td>constraint cavitation</td>
<td>( \log D_s = 5.07 - \frac{1.02 \times 10^4}{(273 + T)} + \frac{2.88 \times 10^6}{(273 + T)^3} - \frac{3.24 \times 10^8}{(273 + T)^5} )</td>
</tr>
<tr>
<td></td>
<td>( T^{-1} ) for temperatures above 25°C</td>
</tr>
<tr>
<td>corrosion</td>
<td>( \tau_{\text{eq}} = \left( \frac{E_p \varepsilon_0 \xi}{E_\text{eq}} \right) \left( \frac{a_h d P}{4 M V Z} \right) )</td>
</tr>
<tr>
<td></td>
<td>( E_\text{eq} = \frac{(R H)^a \exp(E_a / K T_a)}{(R H)^b \exp(E_b / K T)} )</td>
</tr>
<tr>
<td>( \partial T / \partial t ) dependent</td>
<td></td>
</tr>
</tbody>
</table>
die and substrate adhesion fatigue

\[ N_f = 0.5 \left( \frac{\tau_x}{\tau_f} \right)^{1.7} \]

\[ \gamma_s = \frac{L|s_i - s_f| \Delta T}{X} \]

\[ N_f = 0.5 \left[ \frac{L(s_i - s_f) \Delta T}{k_m \gamma_f} \right]^{1.7} \]

\( \Delta T \)-dependent, T-independent

die fracture

\[ \Delta \sigma_{ap} = 10^{-6} |s_i - s_f| \Delta T \sqrt{\frac{EF_L}{X}} \]

\[ \frac{da}{dN} = A \left( \Delta \sigma_{ap} (\pi a)^{1/3} \right)^{1.9} \]

\[ N_f = \frac{2}{(n - 2)4(\Delta \sigma_{ap})^n a_i^{3/2} a_f^{3/2}} \]

\( \Delta T \)-dependent, T-independent
Temperature Dependence of Device Failure Mechanisms

**Electromigration**

Thin film metallizations:

\[ T_r = \frac{1}{2C_e} \left( \frac{T_r}{T_e} \right)^{\sigma_0} \int_0^{\sigma_0} e^{\frac{Q}{2T_r}} \frac{Q}{T_r} x^{2(1-x T_e)} \, dx \]

where

\[ \sigma_e = \frac{T_e}{T_r} = \frac{j \rho_e}{k T_r} \]

and

\[ x = \frac{T_e}{(1-p)^2 + T_e (1-a T_e)} \]

\[ = 1 - \left( \frac{T_e}{T_r} \right) \]

\[ x_e = \frac{T_e}{1-T_e (a T_e - 1)} \quad \text{if } \sigma = 0 \]

\[ x_i = 1 - \left( \frac{T_e}{T_r} \right) \quad \text{if } \sigma = T_r \]

**Multilayered Metallizations**

\[ \frac{\partial R(t)}{\partial t} = R_w C_w \frac{L^2}{L} + R_r C_r \frac{\omega(t)}{L} \]

Dependent on structural non-uniformity and temperature gradient at temperatures lower than 150°C. Steady-state temperature-dependent for temperatures above 150°C, even though dependencies on structural non-uniformity and temperature gradient still exist.

**Hillock Formation**

\( \Delta T \)-dependent, steady-state temperature dependent in the neighborhood of 400°C

**Hot Electrons**

\[ \Delta V_i = C_i (C_i H_i)^p \]

\( T^{-1} \)-dependent in the neighborhood of -55°C, and independent of steady-state temperature at higher temperatures (20°C to 100°C)
### Failure Mechanisms in the Device Oxide Interface

<table>
<thead>
<tr>
<th>Mechanism</th>
<th>Failure Expression</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wire-Bondpad Shear Fatigue</td>
<td>( N_f = C_\omega \tau_{\omega M}^{-n} )</td>
</tr>
<tr>
<td></td>
<td>( \tau_{\omega M} = \left( \frac{r^2}{2E_1A_\omega} \left( \frac{\cosh(z_\omega)}{\cosh(\omega L)} - 1 \right)^2 + \frac{\tanh^2(z_\omega)}{\cosh^2(\omega L)} \right) \frac{Q}{\Delta T} )</td>
</tr>
<tr>
<td></td>
<td>( \tau_{\omega M} = \left( \frac{WQ}{2\pi^2 A_\omega} \left( 1 - \frac{\cosh(z_{\omega L})}{\cosh(\omega L)} \right)^2 + \frac{(\varepsilon_f - \varepsilon)}{(E_1A)} + \frac{1}{(E_1A)} \right) \frac{Q}{\Delta T} )</td>
</tr>
<tr>
<td></td>
<td>( z^2 = \frac{G_\omega}{\beta_\omega} \left( \frac{r}{E_1A} + \frac{(1 - \nu)W}{E_1A} \right) )</td>
</tr>
</tbody>
</table>

\( \Delta T \)-dependent, \( T \)-independent

| Wire Fatigue            | \( N_f = A(\varepsilon_f^p) \)                                               |
|                        | \( \varepsilon_f = \frac{r}{P_{\lambda}} \left[ \frac{\cos^{-1}((1 - (\varepsilon_f - \varepsilon)\Delta T))}{\lambda_\omega} - 1 \right] \) |

\( \Delta T \)-dependent, \( T \)-independent
# TEMPERATURE DEPENDENCE OF DEVICE FAILURE MECHANISMS

Table 2.7 Dominant Temperature Dependency over Steady-state Temperatures from -55°C to 500°C

<table>
<thead>
<tr>
<th>Package element</th>
<th>Mechanism</th>
<th>Nature of temperature dependence</th>
<th>Nature of temperature dependence</th>
<th>Nature of temperature dependence</th>
</tr>
</thead>
<tbody>
<tr>
<td>wire</td>
<td>flexure fatigue</td>
<td>$\Delta T$</td>
<td>$\Delta T$</td>
<td>$\Delta T$</td>
</tr>
<tr>
<td>wirebond</td>
<td>shear fatigue</td>
<td>$\Delta T$</td>
<td>$\Delta T$</td>
<td>$\Delta T$</td>
</tr>
<tr>
<td>Kirkendall voiding</td>
<td>temperature-independent</td>
<td>T-dependent for $T &gt; 150^\circ C$</td>
<td>T-dependent for $T &gt; 150^\circ C$</td>
<td>T-dependent for $T &gt; 150^\circ C$</td>
</tr>
<tr>
<td>die</td>
<td>die fracture</td>
<td>$\Delta T$</td>
<td>$\Delta T$</td>
<td>$\Delta T$</td>
</tr>
<tr>
<td>die adhesive</td>
<td>die adhesive fatigue</td>
<td>$\Delta T$</td>
<td>$\Delta T$</td>
<td>$\Delta T$</td>
</tr>
<tr>
<td>encapsulant</td>
<td>encapsulant reversion</td>
<td>Temperature-independent</td>
<td>T-dependent for $T &gt; 300^\circ C$</td>
<td>T-dependent for $T &gt; 300^\circ C$</td>
</tr>
<tr>
<td></td>
<td>(plastic package only)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>encapsulant cracking</td>
<td>$\Delta T$, $\partial T/\partial t$ ($\geq 25^\circ C/\text{sec.}$)</td>
<td>T-dependent ($T &gt; 215^\circ C$), $\Delta T$, $\partial T/\partial t$, ($\geq 25^\circ C/\text{sec.}$)</td>
<td>T-dependent ($T &gt; 215^\circ C$), $\partial T/\partial t$, ($\geq 25^\circ C/\text{sec.}$)</td>
</tr>
<tr>
<td></td>
<td>(plastic package only)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>package</td>
<td>stress corrosion</td>
<td>$\partial T/\partial t$</td>
<td>$\partial T/\partial t$</td>
<td>$\partial T/\partial t$</td>
</tr>
<tr>
<td>die metallization</td>
<td>corrosion</td>
<td>mildly T-dependent, $\partial T/\partial t$</td>
<td>$\partial T/\partial t$</td>
<td>$\partial T/\partial t$</td>
</tr>
<tr>
<td></td>
<td>electromigration</td>
<td>structural non-uniformity dependent, $\nabla T$</td>
<td>T-dependent ($T &gt; 150^\circ C$), $\nabla T$-dependent</td>
<td>T-dependent ($T &gt; 150^\circ C$), $\nabla T$-dependent</td>
</tr>
<tr>
<td>Mechanism</td>
<td>hv</td>
<td>hv</td>
<td>hv</td>
<td></td>
</tr>
<tr>
<td>---------------------------------</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td></td>
</tr>
<tr>
<td>hillock formation</td>
<td>ΔT</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>metallization migration</td>
<td>T- dependent (for T &gt; 400°C)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>contact spiking</td>
<td>T- dependent (for T &gt; 500°C)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>constraint cavitation</td>
<td>T- dependent (for T &gt; 400°C)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>slow trapping</td>
<td>T- dependent</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>electrostatic discharge</td>
<td>T- dependent (in presence of protection circuits)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>time-dependent dielectric</td>
<td>voltage-dependent, weak T dependence</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>breakdown</td>
<td>voltage-dependent, weak T dependence</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Device ionic contamination</td>
<td>T- dependent (device not operational)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>forward second breakdown</td>
<td>T- dependent (for T &gt; 160°C)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>reverse second breakdown</td>
<td>mild T- dependent</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>surface-charge spreading</td>
<td>T- dependent (for T &gt; 150°C)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Device / Oxide interface</td>
<td>T- dependent</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>hot electrons</td>
<td>Temperature-independent</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Temperature-independent</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 2.8  Variation of the Nature of Temperature Dependency of the Device from -55°C to 500°C

<table>
<thead>
<tr>
<th>Nature of stress dependence</th>
<th>-55°C &lt; T &lt; 150°C</th>
<th>150°C &lt; T &lt; 400°C</th>
<th>T &gt; 400°C</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Number of Mechanisms</td>
<td>Number of Mechanisms</td>
<td>Number of Mechanisms</td>
</tr>
<tr>
<td>steady-state temperature (T)</td>
<td>1</td>
<td>8</td>
<td>11</td>
</tr>
<tr>
<td>temperature cycle (ΔT)</td>
<td>6</td>
<td>6</td>
<td>5</td>
</tr>
<tr>
<td>temperature gradient (V T)</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>time dependent temperature change (∂T/∂t)</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>voltage (V)</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>inverse temperature dependent (T^-1)</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>temperature independent (form of temperature dependence)</td>
<td>9</td>
<td>4</td>
<td>2</td>
</tr>
</tbody>
</table>
3

TEMPERATURE RELATED PARAMETER VARIATIONS

3.1 INTRODUCTION

This chapter discusses the temperature dependence of electrical parameter variations in bipolar and MOSFET devices. The parameters investigated for bipolar devices include intrinsic carrier concentration, thermal voltage, mobility, current gain, leakage current, collector-emitter saturation voltage, and VTC shift. The parameters investigated for MOSFETs include threshold voltage, mobility, drain current, time delay, strong inversion leakage, subthreshold leakage, and chip availability. The temperature thresholds above which temperature dependence renders the device inoperable have been identified for each of the parameters.
3.2 TEMPERATURE DEPENDENCE OF BIPOLAR JUNCTION TRANSISTOR PARAMETERS

3.2.1 Intrinsic Carrier Concentration, Thermal Voltage and Mobility

The ideal p-n junction (ideal diode equation) is affected by temperature due to variations in the saturation current, which is a function of the intrinsic carrier concentration and thermal voltage. Intrinsic carrier concentration is exponentially dependent on steady-state temperature (Equation (3.3), and thermal voltage is linearly dependent on steady-state temperature \( V_T = kT/q \). The ideal diode equation is represented by

\[
I = I_s(e^{V/qT} - 1) \tag{3.1}
\]

where \( I \) is the total forward bias diode current, \( V \) is the voltage across the diode, \( V_T \) is the thermal voltage \( (=kT/q) \), and \( I_s \) is the saturation current given by

\[
I_s = qA\left(\frac{D_pP_{no}}{L_p} + \frac{D_nP_{no}}{L_n}\right) \tag{3.2}
\]

where \( q \) is the electron charge \( (= 1.6 \times 10^{-19} \) Coulomb); \( A \) is the cross-sectional area of the p-n junction; \( D_p \) and \( D_n \) are the diffusion coefficients of the \( p \) and \( n \) regions, respectively; \( L_p \) and \( L_n \) are the diffusion lengths of the \( p \) and \( n \) regions, respectively (one diffusion length is the distance in the neutral region at which the minority carrier concentration is 0.37 of its value at the edge of the depletion, or space-charge, region); \( P_{no} \) is the minority hole concentration in the \( n \)-region; and \( N_{p0} \) is the minority electron concentration in the \( p \)-region. The effect of temperature on saturation current derives its origin from the variation in intrinsic carrier concentration, \( N_p \) which increases dramatically with temperature and is given as

\[
n_i = \sqrt{N_C N_V e^{-\frac{E_F}{kT}}} \tag{3.3}
\]

where \( N_V \) and \( N_C \) are the valence and conduction-band effective densities, respectively, given by
TEMPERATURE DEPENDENCE: BIPOLAR

\[ N_\nu = 2 \left( \frac{2 \pi m^*_k kT}{\hbar^2} \right)^{3/2} \]  
\[ N_c = 2 \left( \frac{2 \pi m^*_k kT}{\hbar^2} \right)^{3/2} \]  

(3.4)

(3.5)

Furthermore, the temperature dependence of the intrinsic carrier concentration may also be represented by

\[ n_i^2(T) = K T^{-3} e^{-V_p/V_T} \]  

where \( K \) is a constant, \( T \) is the absolute temperature (Kelvin), \( V_p \) is the band gap voltage (for silicon at 300 K, \( V_p \) is 1.11 V), and \( V_T \) is the thermal voltage (at 300K, \( V_T = 26 \text{ mV} \) [Hodges and Jackson 1988].

The temperature rate of change of the saturation current is represented by the temperature coefficient of saturation current, which describes the fractional change for \( I_s \) per unit change in temperature, represented by Hodges and Jackson [1988] as

\[ \frac{1}{I_s} \frac{dI_s}{dT} = \frac{1}{n_i^2} \frac{dn_i^2}{dT} \]  

\[ = \frac{3}{T} + \frac{1}{T} \frac{V_p}{V_T} \]

(3.7)

For silicon near-room temperature, the first term in Equation (3.7) is \( \approx 1\% \) per Kelvin, and the second term is 14\% per Kelvin. In other words, the saturation current approximately doubles for every 5°C rise. Moreover, the band gap voltage, \( E_G \), also varies with temperature:

\[ E_G(T) = E_G(0) - \frac{\alpha T^2}{T + \beta} \]

(3.8)

where \( E_G(0) \) is the band gap energy at \( T = 0 \) [Roulston 1988]. For silicon, \( \alpha \) is 7.02x10^{-4} and \( \beta \) is 1108 [Pierret 1989]. The band gap energy decreases with temperature.

Carrier mobility, a function of steady-state temperature, is a ratio of carrier
velocity to the electric field causing carrier motion. Mobility decreases with temperature and is represented by Roulston [1990] as

\[
\mu_p = 54 \left( \frac{T}{300} \right)^{-0.57} + \frac{1.36 \times 10^4 T^{-2.33}}{1 + \left( \frac{N_d}{2.35 \times 10^{17} \left( \frac{T}{300} \right)^{2.4}} \right) 0.88 \left( \frac{T}{300} \right)^{-0.146}}
\]

(3.9)

\[
\mu_n = 0.88 \left( \frac{T}{300} \right)^{-0.57} + \frac{7.4 \times 10^9 T^{-2.33}}{1 + \left( \frac{N_A}{1.26 \times 10^{17} \left( \frac{T}{300} \right)^{2.4}} \right) 0.88 \left( \frac{T}{300} \right)^{-0.146}}
\]

(3.10)

The temperature characteristics of these parameters is, however, of minor importance, because signal voltages in most cases are much larger than the variation of the diode voltage over the temperature range -55°C to 125°C [Hodges and Jackson 1988].

3.2.2 Current Gain

The current gain of the bipolar junction transistor is the collector current divided by the base current, otherwise known as \( \beta \). The d.c. current gain is expressed as

\[
B = \frac{I_C}{I_B}
\]

(3.11)

The common emitter d.c. current gain increases appreciably with increasing temperature because of improved emitter efficiency at higher temperatures. Emitter efficiency is exponentially dependent on steady-state temperature (Equation (3.12). The exponential term, represented by \( \exp \left( -\Delta E_g/kT \right) \), arises due to the energy-gap reduction of a highly doped emitter [Kauffman and Bergh 1968,
TEMPERATURE DEPENDENCE: BIPOLAR


\[ \gamma_E = \frac{D_n}{D_p} \left( \frac{w_s}{\int_0^{w_s} N_d dx} \right)^{-\Delta E_g/kT} \]  

(3.12)

For moderately or lightly doped emitters (concentrations below \(5 \times 10^{19} \text{ cm}^{-3}\)), \(\Delta E_g\) is very small, and the temperature effect on current gain is negligible. For doping concentrations above \(10^{19} \text{ cm}^{-3}\), the hole mobility in the emitter is relatively constant with respect to temperature. Einstein's relationship \((D_p/\mu_p = kT/q)\) predicts that the diffusion constant must increase with temperature. The minority carrier lifetime \(\tau_p\) increases with temperature, which causes the diffusion length, \(L_p\), to also increase with temperature. The increase in the diffusion length of minority carriers adds to the \(\Delta E_g\) effect by increasing the current gain. The variation in the d.c. current gain is depicted by Figure 3.1. The temperature sensitivity of current gain is a contributing factor to hot-spot formation and affects the second breakdown energy limit. It is therefore advantageous to reduce the current gain temperature dependence by lightly doping the transistor base and limiting the phosphorus-doped emitter surface concentration to about \(7 \times 10^{19} \text{ cm}^{-3}\) [Kauffman and Bergh 1968, Buhanan 1969, RCA 1978].

Ebers-Moll equations for emitter and collector current are represented as [Hodges and Jackson 1988]

\[ I_E = I_{ES} \left( e^{v_{BE}V_{BE} - 1} \right) - \alpha_R I_{CS} \left( e^{v_{BC}V_{BC} - 1} \right) \]  

(3.13)

\[ I_C = \alpha_F I_{ES} \left( e^{v_{BE}V_{BE} - 1} \right) - I_{CS} \left( e^{v_{BC}V_{BC} - 1} \right) \]  

(3.14)

where \(I_{ES}\) and \(I_{CS}\) represent the saturation currents for emitter and collector junctions, respectively; \(v_{BE}\) and \(v_{BC}\) represent the base-emitter and base-collector bias voltages; \(V_T\) is the thermal voltage \((= kT/q)\); and \(\alpha_R\) and \(\alpha_F\) are the common base current gains \((=I_C/I_E)\). Here, \(\alpha_R\) corresponds to the common base current gain with the base-emitter junction reverse-biased and the base-collector junction forward-biased; \(\alpha_F\) corresponds to the common base current gain with the base-emitter junction forward-biased and the base-collector junction reverse-biased.
The base current is obtained by substituting for $I_b$ and $I_c$ from Equations (3.13) and (3.14) [Marazas 1992]:

$$I_b = I_b - I_c \quad (3.15)$$

$$I_b = qA n_i^2 \left( \frac{D_E}{L_E N_E} \right) \left( e^{\frac{\Phi_E}{kT}} - 1 \right)$$

$$I_b = qA n_i^2 \left( \frac{D_C}{L_C N_C} \right) \left( e^{\frac{\Phi_C}{kT}} - 1 \right) \quad (3.16)$$

The expression for common emitter d.c. current gain, $\beta = I_c/I_b$, holds only for the active region of BJT operation. In the active region, the base-collector junction is reverse-biased, making the exponent $< 1$. On the assumption that the exponential base-collector term is negligible, the previous equations reduce to

$$I_b = qA n_i^2 \frac{D_E}{L_E N_E} e^{\Phi_E/V_T} \quad (3.17)$$

$$I_c = qA n_i^2 \frac{D_B}{W_B N_B} e^{\Phi_B/V_T} \quad (3.18)$$

where $n_i$ is the intrinsic carrier concentration; $A$ is the area of the base region; $q$ is the electron charge ($= 1.6 \times 10^{-19} \text{C}$); $D_B$ and $D_E$ are the diffusion coefficients of the base and emitter, respectively; $W_B$ is the base width; $L_E$ is the diffusion length of the minority carriers in the emitter; and $V_T$ is the thermal voltage.

Substituting the explicit temperature dependence into the above equations and dividing $I_c$ by $I_b$, $\beta$ is given as [Marazas 1992]

$$\beta = \frac{N_E k}{qN_B W_B} \sqrt{\frac{1}{T}} \mu(T) \tau(T) \quad (3.19)$$

### 3.2.3 BJT Inverter Voltage Transfer Characteristic (VTC)

Steady-state temperature alters the voltage transfer characteristic of the bipolar junction transistor (BJT). Typically, the transistor moves from the cutoff region to the active region, in the neighborhood of 27°C, when a voltage of
TEMPERATURE DEPENDENCE: BIPOLAR

approximately 0.7 volts ($V_{be}$) is applied across the base-emitter junction $V_{be}$. The
Figure 3.1 Temperature dependence of current gain for a bipolar transistor. The temperature sensitivity of current gain is a contributing factor to hot spot formation and effects the second breakdown energy limit. It is therefore advantageous to reduce the current gain temperature dependence by lightly doping the transistor base and limiting the phosphorous doped emitter surface concentration to about 7 x 10 cm [Kaufmann and Bergh, 1968; Buhanan, 1969; RCA, 1978]
minimum base-emitter voltage, \( V_{\text{beo}} \), required to obtain the same \( I_{\text{bo}} \) decreases with an increase in temperature. The temperature dependence of \( V_{\text{beo}} \) can be derived from Equation (3.17). This value, \( I_{\text{bo}} \), represents the minimum current necessary to turn on the transistor at 27°C. The equation for \( V_{\text{beo}} \) is [Hodges and Jackson 1988, Muller and Kamins 1986]

\[
V_{\text{beo}}(T) = \frac{kT}{q} \log\left[ \frac{I_{\text{bo}}L_d N_D q}{q n_i^2(T) A \mu(T) kT} \right]
\]  

(3.20)

There are three regions of operation for the BJT inverter, known as the cut-off, active, and saturation regions. In the cut-off region, the input voltage, \( V_{\text{in}} \), is less than the BJT turn-on voltage, \( V_{\text{beo}} \). The base current, \( I_b \), is zero, and thus the collector current and the emitter current are also zero. The output voltage is therefore equal to \( V_{\text{CC}} \):

\[
V_{\text{out}} = V_{\text{CE}} = V_{\text{CC}}
\]  

(3.21)

where \( V_{\text{out}} \) is the output voltage, \( V_{\text{CE}} \) is the collector-emitter voltage, and \( V_{\text{CC}} \) is the collector supply voltage. In the active region, \( V_{\text{in}} \) is greater than \( V_{\text{beo}} \). The base current can be found from the voltage drop across the base resistor

\[
I_B = \frac{V_{\text{in}} - V_{\text{be(on)}}}{R_B}
\]  

(3.22)

where \( R_B \) is the base-load resistor, \( V_{\text{in}} \) is the input voltage to the BJT inverter, \( V_{\text{be(on)}} \) is the turn-on voltage for the BJT, and \( I_B \) is the base current. The collector current can be calculated from the base current and the BJT current gain \( (\beta_f) \):

\[
I_C = \beta_f I_B
\]  

(3.23)

The output voltage, \( V_{\text{CE}} \), is given by

\[
V_{\text{out}} = V_{\text{CC}} - R_C I_C
\]  

(3.24)

The temperature dependence of the output voltage is calculated from Equations (3.22) through (3.24).
The third region of operation for the BJT is the saturation region. The collector current is given by

\[ I_C = \frac{V_{CC} - V_{CE}}{R_C} \]  

(3.26)

where \( V_{CC} \) is the collector supply voltage and \( V_{CE} \) is the collector voltage. However, when the transistor is not actually in the saturation region but on the edge of saturation, \( V_{CO} = 0 \), so

\[ I_C = \frac{V_{CC} - V_{BE(on)}}{R_C} \]  

(3.27)

Equation (3.27) represents the maximum amount of current allowed in the collector before the transistor actually goes into saturation. Additionally, the maximum allowed base current is

\[ I_B = \frac{I_C}{\beta} \]  

(3.28)

The input voltage can be derived from Equation (3.22):

\[ V_{in} = I_B R_B + V_{BE(on)} \]  

(3.29)

The temperature dependence of the critical input voltage that would push a BJT into saturation is represented as [Marazas 1992]

\[ V_{crit}(T) = \frac{R_B}{R_C} \left( \frac{V_{CC} - V_{BE(on)}(T)}{\beta(T)} \right) + V_{BE(on)}(T) \]  

(3.30)

where \( V_{crit} \) is the critical input voltage. In bipolar devices operating high-voltage current-switching applications, the collector current plays a major role in...
TEMPERATURE DEPENDENCE: BIPOLAR

determining transistor losses; Bromstead represents this by $I_{CE}$ in the following equation [1991]:

$$I_o = qA \left( \frac{D_n n_i^2}{L_n N_A} + \frac{D_p n_i^2}{L_p N_D} \right)$$  \hspace{1cm} (3.31)

The approximate fractional change in leakage current versus temperature is represented by

$$\frac{1}{I_o} \left( \frac{\partial I_o}{\partial T} \right) = \frac{1}{2} \left( \frac{3}{T} + \frac{E_o}{\sqrt{N_c N_v}} \right) @ V_R=constant$$  \hspace{1cm} (3.32)

where $E_o$ is the activation energy in eV, $N_c$ is the density of the conduction band, $N_v$ is the density of the valence band, and $T$ is the steady-state temperature in degrees Kelvin [Sevems and Armijos 1984]. The change in the collector leakage current is approximately 8 °C for every 12 °C rise in steady-state temperature at the junction for the temperature range 150 °C to 200 °C. Below steady-state temperatures of 150 °C, the leakage current exhibits a plateau-i.e., it does not change at temperatures of up to 150 °C ($I_{CE}(150°C) = 1.2 \times 10^{-4}$ amp, and $I_{CE}(200°C) = 1.5 \times 10^{-3}$ amp)[Bromstead 1991]. Emitter-base leakage current exhibits saturation in the lower temperature region (under 150°C). For temperatures higher than 150°C, the emitter base leakage current exhibits doubling for every 12°C rise in steady-state temperature($I_{CE}(150°C) = 1.3 \times 10^{-4}$ Amp, and $I_{CE}(200°C) = 2.3 \times 10^{-3}$ Amp)[Bromstead 1991].

3.2.4 Collector-Emitter Saturation Voltage

In high-voltage and current-switching applications, the saturation voltage, $V_{CE(sat)}$, is vital for determining transistor losses. Under normal switch or inverter conditions, the major part of the voltage drop across the transistor occurs in the non-conductivity-modulated region close to the $n^+$ epitaxial layer. Assuming that both the base and the collector have entered high-level injection, the total value of the C-E saturation voltage can be written as [Roulston 1990]

$$V_{CE(sat)} = \frac{I_{c0} W_R}{A} - 2V_i \left[ \frac{p(W_b)}{n_i} \right] + 2V_i \left[ \frac{n(0)}{n_i} \right]$$  \hspace{1cm} (3.33)

where $W_R = W_b - X_O$ ( $W_b$ - base width, $X_O$ - oxide thickness), $p(W_b)$ is the
hole concentration at the base-collector (B-C) junction, and \( n(0) \) is the electron concentration at the base-emitter (B-E) junction [Roulston 1990]. The resistivity of the epitaxial layer is given as

\[
\rho_{\text{epi}} = \frac{1}{q \mu_n N_{\text{epi}}} \tag{3.34}
\]

where \( \mu_n \) is the electron mobility and \( N_{\text{epi}} \) is the epitaxial layer density. The temperature dependence of electron mobility is represented as

\[
\mu_n = 88 \left( \frac{T}{300} \right)^{-0.57} + \frac{7.4 \times 10^8 T^{-2.33}}{1 + \left( \frac{N}{1.26 \times 10^{17} \left( \frac{T}{300} \right)^2} \right) 0.88 \left( \frac{T}{300} \right)^{-0.146}} \tag{3.35}
\]

where \( N \) is the electron density, and \( T \) is the steady-state temperature in degrees Kelvin [Roulston 1990]. The temperature dependence of \( V_{\text{CEsat}} \) is related to the intrinsic carrier concentration, \( n_i \), by

\[
n_i = \sqrt{N_c N_v} e^{-E_g k T} \tag{3.36}
\]

where \( N_c \) is the effective density of the conduction band, \( N_v \) is the effective density of the valence band, \( E_g \) is the activation energy in eV, \( k \) is Boltzmann’s constant (8.617×10\(^{-5}\) ev/K), and \( T \) is the steady-state temperature in degrees Kelvin [Pierret 1987]. Table 3.1 represents the variation of intrinsic carrier concentration, epitaxial resistivity, and collector-emitter saturation voltage for temperatures between 20 °C and 200 °C [Bromstead 1991].

### 3.3 TEMPERATURE DEPENDENCE OF MOSFET PARAMETERS

#### 3.3.1 Threshold Voltage

The threshold voltage is the minimum gate voltage at which the channel starts conducting. The closed-form value of the threshold voltage is derived from the current voltage characteristics. MOSFET behavior with respect to temperature can be derived by finding the incremental voltage drop along the channel, as a function of the channel current. At a distance \( y \) along the channel, the voltage with respect to the source is \( V(y) \) and the gate-to-channel voltage is \( V_{\text{GS}} - V(y) \). Assuming that the gate voltage exceeds the threshold voltage, the charge per unit area in the conducting channel at a point, \( y \), is represented by
TEMPERATURE DEPENDENCE: MOSFET

\[ Q(y) = C_{ox}(V_{GS} - V(y) - V_{th}) \]  

(3.37)

where \( C_{ox} \) is the oxide capacitance per unit area, \( V_{GS} \) is the gate-source voltage, \( V(y) \) is the voltage induced in the channel with respect to source, and \( V_{th} \) is the threshold voltage [Muller and Kamins 1986, Hodges and Jackson 1988]. The resistance, \( dR \), of this channel of length, \( dy \), is

\[ dR = \frac{dy}{W \mu_s Q(y)} \]  

(3.38)

where \( W \) is the width of the channel perpendicular to the length, and \( \mu_s \) is the average mobility of the electrons in the channel [Muller and Kamins 1986, Hodges and Jackson 1988]. The voltage drop along the channel of length \( dy \) is [Muller and Kamins 1986, Hodges and Jackson 1988]

\[ dV_c = I_D dR = -\frac{I_D dV}{W \mu_s Q(y)} \]  

(3.39)

Integrating along the path from the source to the drain for \( I_D \) as a function of applied voltage (for \( V_{GS} \geq V_{th}; \) and \( V_{DS} \leq (V_{GS} - V_{th}) \)),

\[ I_D = \frac{k}{2} \left( 2(V_{GS} - V_{th}) - V_{DS}^2 \right) \]  

(3.40)

where \( V_{th} \) is the threshold voltage, and \( k \) is the transconductance parameter \( (k = (\mu_s C_{ox} W)/L) \) [Muller and Kamins 1986, Hodges and Jackson 1988]. Equation (3.40) is the expression for drain current in the linear region of operation. In the saturation region, where \( V_{GS} \geq V_{th}; \ V_{DS} \geq V_{GS} - V_{th} \), the drain current is represented as [Muller and Kamins 1986, Hodges and Jackson 1988]

\[ I_D = \frac{k}{2} (V_{GS} - V_{th})^2 \]  

(3.41)

The total mobile electron charge is represented as [Muller and Kamins 1986, Hodges and Jackson 1988]
The threshold voltage is then represented by

\[ V_{th} = V_{FB} + V_C + 2|\Phi_p| + \frac{1}{C_{ox}} \sqrt{2\varepsilon_s q N_a (2|\Phi_p| + V_C - V_B)} \]  \hfill (3.43)

where \( V_{FB} \) is the flatband voltage, \( V_C \) is the voltage in the MOS channel, \( \Phi_p \) is the potential in the p-type material (i.e., the difference between the fermi level and the intrinsic concentration), \( C_{ox} \) is the oxide capacitance, \( N_a \) is the number of acceptors, \( e \) is the electronic charge (\( 1.6 \times 10^{-19} \) C), \( V_B \) is the base voltage, and \( \varepsilon_s \) is the silicon permittivity. The temperature dependence of threshold voltage is represented as

\[ V_{TN} = P_o T + Q_o \]  \hfill (3.44)

\[ V_{TP} = -P_o T - Q_o \]  \hfill (3.45)

where the constant \( P_o = -2.4 \times 10^{-3} \) and \( Q_o = 1.72 \) (for the temperature range of 25-250°C) [Shoucair 1986]. This equation applies to a typical MOSFET with \( V_t \) equal to 1 volt at 27°C. Typically, the threshold voltage of both NMOS and PMOS enhancement-mode transistors decreases in magnitude by 1.5 to 2 mV/°C with increasing temperature. Figure 3.2 shows the threshold voltage variation versus temperature for n- and p-channel MOSFETs with bulk concentration \( 10^{15} \) cm\(^{-3}\); 37(b) shows this for bulk concentration \( 3 \times 10^{16} \) cm\(^{-3}\) [Wang 1971]. However, the change in threshold voltage does not produce a very significant change in circuit performance, because even a 200 mV change in \( V_T \) does not cause a very large percentage change in \( V_{GS} - V_T \) [Hodges and Jackson 1988]. The threshold voltage of devices with heavily doped bulk is more sensitive to temperature variation than lightly doped substrates [Blicher 1981].

### 3.3.2 Mobility

The mobility of carriers in a channel of an MOS transistor is an inverse function of absolute temperature. The temperature dependence is represented by
where \( \mu \) is the mobility at 27°C in both cases and \( T_0 \) is 27°C [Shoucair 1986, Tsividis and Antognetti 1985, Muller and Kamins 1986, Hodges and Jackson 1988]. Thus, a 100°C rise in temperature may decrease the mobility by as much as 40%. The result is a proportional decrease in the drain current, \( I_D \), for fixed applied voltages. The current consumption of the whole circuit may decrease considerably at high temperature (Figure 3.3). The maximum speed of operation thus decreases in proportion. The change in threshold voltage and mobility affect the drain current, the transconductance, and the drain-source on resistance (Figures 3.4 and 3.5).

### 3.3.3 Drain Current

The drain current decreases with temperature. The temperature dependence of the drain current can be represented by

\[
I_{ds} = \frac{A}{1 + BT}
\]

where \( I_{ds} \) is the drain current, \( A \) and \( B \) are empirical constants, and \( T \) is the temperature (centigrade) [Chang 1987]. The temperature degradation coefficient, \( B \), in Equation (3.48), is a function of gate voltage \( (V_{gs}) \) and drain voltage \( (V_{ds}) \) (Figure 3.6). Typical values of the temperature degradation coefficient, \( B \), range from \( 1.03 \times 10^{-3}/°C \) to \( 6 \times 10^{-3}/°C \). This model fits experimental data for variations in drain current versus temperature for 10°C to 120°C. Theoretically, the temperature dependence of drain current versus temperature is represented by

\[
I_{ds} = C(T+273)^{-D}
\]

where \( C \) and \( D \) are fitting parameters and \( T \) is the temperature (centigrade). The relationship between the experimental data represented by Equation (3.48) and
Figure 3.2  Threshold voltage variation of n- and p- channel MOSFETs (a). With bulk concentration $10^{15}\text{cm}^{-3}$ and (b). With bulk concentration $3 \times 10^{16}\text{cm}^{-3}$ [Wang, 1971]. The change in threshold voltages does not produce a very significant change in circuit performance, because even a 200 mV change in $V_T$ does not cause a larger percentage change in $V_{GS}-V_T$. [Hodges and Jackson, 1988]
Figure 3.3  Effective channel mobility characteristics for n and p channel MOSFETs. Mobility is an inverse function of absolute temperature [Shoucair, 1986]
Figure 3.4 The variation in the drain-to-source ON resistance. The ON resistance increase by a factor of 3 over a temperature range of -40 to 120 deg.C [Blicher, 1981]
Figure 3.5 The variation in transconductance versus temperature in the temperature range of -55 to 125 deg. C.

[Blachier, 1981]
theoretical calculations represented by Equation (3.49) is

\[ D = 203B + 0.25 \]  

(3.50)

The correlation between the simplified first-order experimental equation and the theoretical equation is 0.998. The variation in drain current with temperature is not very significant in the temperature range of -55°C to 125°C (Figure 3.6) [Blicher 1981].

3.3.4 Time Delay

In a MOS circuit, the delay time due to charging and discharging of the load capacitor through a turn-on has been represented by Chang as [1987]:

\[ t_d = C_l \int \frac{dV_{ds}}{I_{ds}(V_{ds}, T)} \]  

(3.51)

where \( C_l \) is the load capacitance, \( V_{ds} \) is the drain voltage, \( I_{ds} \) is the drain current, and \( T \) is the temperature. The temperature dependence of time delay is represented by the linear temperature coefficient of delay time, \( K \), which is the time-change rate of delay time versus temperature

\[ K = \frac{dt}{t_o \, dT} \]  

(3.52)

where \( t_o \) is the time delay at reference temperature [Chang 1987]. Substituting into Equation (3.52) for \( I_{ds} \) from Equation (3.48) and for delay time, \( t_d \), from Equation (3.51),

\[ K = \frac{\int \frac{B}{V_i} dV_{ds} + \int \frac{B}{V_o} dV_{ds}}{\int \frac{dV_{ds}}{V_i} + \int \frac{dV_{ds}}{V_o}} \]  

(3.53)

where \( I_o \) is the \( I_{ds} \) at 0°C, \( V_i = F \times V_{cc} \), \( V_h = (1-F) \times V_{cc} \) for \( 0 < F < 1 \); \( V_i \) and \( V_h \) are the lower and upper bounds for the voltage swing used in delay-time calculation [Chang 1987]. The total swing during device switching has been
TEMPERATURE DEPENDENCE: MOSFET

divided into linear and saturation regions. Equation (3.53) is simplified as

\[ K = \frac{B_v \left( \ln \left( \frac{V_{dss}}{V_t} \right) V_{dss} + V_t \right) + B_{sat} \left( V_{dss} - V_t + \ln \left( \frac{V_{cc} - V_{dss}}{V_{cc} - V_h} \right) \right)}{\left( \ln \left( \frac{V_{dss}}{V_t} \right) - \ln \left( \frac{V_{cc} - V_{dss}}{V_{cc} - V_h} \right) \right) V_{dss} + \ln \left( \frac{V_{cc} - V_{dss}}{V_{cc} - V_h} \right) V_{cc}} \] (3.54)

where \( K \) is the linear temperature coefficient of delay time (the rate of delay time with respect to temperature), \( B_v \) is the temperature degradation coefficient at temperature 0°C, \( V_{dss} \) is the drain voltage at saturation, \( V_t \) and \( V_h \) is the lower and higher bounds of the voltage swing during device switching [Chang 1987]. The model can be extended to calculate the delay time for a CMOS ring oscillator, which consists of either seventeen or thirty-seven stages of directly coupled inverters. Each inverter consists of a pair of n-channel and p-channel MOSFETs. The total delay time, \( T_{ph} \), of a CMOS inverter can be represented by

\[ t_{ph} = t_r + t_f = t_f(1 + K_r T) \] (3.55)

where the subscripts represent the rise and fall times, respectively, and \( t_{ph} \) and \( t_{ro} \) are the corresponding delay times at 0°C [Chang 1987]. The total time-delay coefficient, \( K_{ph} \), is

\[ K_{ph} = \frac{K_f}{1 + \frac{t_{ro}}{t_{fo}}} + \frac{K_r}{1 + \frac{t_{fo}}{t_{ro}}} \] (3.56)

Another model for gate delay versus temperature dependence is given by Shoucair [1987]. The gate delay is defined as the time between the input and output waveforms at the \( V_{dd}/2 \) points of the gate in the ring oscillator (\( V_{dd} \) is the drain supply voltage) at temperature \( T \). The gate delay as a function of temperature is given by

\[ t_d(T) = \frac{1.8 C_L(T)V_{DD}}{\mu N(T) C_{ox} \left( \frac{W}{L} \right)_N} \left( \frac{1}{1 - \frac{V_{THN}(T)}{V_{DD}}} \right)^2 + \frac{1}{\left( 1 - \frac{|V_{THP}(T)|^2}{V_{DD}} \right)^2} \] (3.57)
where \( T_0 \) is the reference temperature, \( c(T) \) is the temperature rate of change of gate delay (also called the delineation factor, \( 1/K \) or \( 1/\gamma \)), \( V_{TH(n)} \) is the threshold voltage for n-channel, \( V_{TH(p)} \) is the threshold voltage for the p-channel, and \( C_L(T) \) is the total load capacitance. The delay increases linearly with temperature and can be calculated using the delineation factor, \( c(T) \), at any temperature ranging from 0.004/°C to 0.006/°C. At 250°C, the cell is half as fast as at 25°C. The delineation factor quantifying the temperature dependence is larger for cells driving loads dominated by temperature-dependent junction capacitances than for temperature-independent capacitive loads. The delineation factor is constant for a given cell over a temperature range of 25°C to 250°C [Shoucair 1987]:

\[
c(T) = \frac{1}{t_D(T_0)} \frac{dt_D(T)}{dT} \quad (3.59)
\]

\[
\frac{dt_D(T)}{dT} = t_D(T) \left( \frac{1.5}{T} + \frac{2}{V_{DD} - V_{TH}(T)} \frac{d_{TH}(T)}{dT} \right) \quad (3.60)
\]

Shoucair estimated the temperature dependence of the rise and fall times, \( t_r \) and \( t_o \), on a sample size of seven inverters [1984]. Assuming that \( |V_{TH(27°C)}| = 1.5 \text{ V}; |dV_{th}(T)/dT| = 3.5 \text{ mV/°C}, V_{DD} = +5\text{ V} \), the ratio of rise and fall times, \( t_r \) and \( t_o \), at 27°C and 250°C was estimated to be 1.67 (\( R_0 = t_0(250°C)/t_0(27°C) \); \( R_1 = t_1(250°C)/t_1(27°C) \); \( (R_1+R_2)/2 = 1.67 \)). This is the factor by which the speed of the stage will decrease in a high-clock-rate application.

## 3.3.5 Leakage Currents

Leakage currents represent the limiting factor for high temperature functionality of MOSFETs. \( I_D \) versus \( V_{GS} \) curves represent a significant upward trend at temperatures above 250°C because drain leakage currents have become comparable to drain channel currents. The drain leakage current is typically four to five orders of magnitude smaller at 25°C than at 250°C. The drain or source junction leakage current is generally dominated by generation-recombination
TEMPERATURE DEPENDENCE: MOSFET

leakage over a temperature range of 25°C to $T_{\text{max}}$ °C, and by diffusion leakage
Figure 3.6 The variation in drain current in the equipment operating range of -55 to 125 deg. C is not very significant. [Blicher, 1981]

- 80 μs PULSE TEST
- \(V_{DS} = 25\) V

Id, Drain Current (A)

Vgs, Gate-To-Source Voltage (V)
TEMPERATURE DEPENDENCE: MOSFET

above $T_{\text{transition}} \, ^\circ\text{C}$. Typically, in most CMOS processes, $T_{\text{transition}} \, ^\circ\text{C}$ lies in the range of 130°C to 150°C.

**Leakage Currents in Strong Inversion and Deep Depletion.** The leakage current components in an n-channel MOSFET in strong inversion comprise a drain and source well-bottom wall component, $I_{db}(T)$, which may be expressed as

$$I_{db}(T) = J_{db}(T)A_S = J_{db}(T)(W,d)$$  \hspace{1cm} (3.61)

where $I_{db}(T)$ is the component of leakage current through the bottom wall of the well, $W$ is the well width, and $d$ is well length; a drain and source well-sidewall component, $I_{sw}(T)$, given as

$$I_{sw}(T) = J_{sw}(T)A_{sw} = J_{sw}(T)(W,x)$$  \hspace{1cm} (3.62)

where $W$ is the well width, and $x$ is the well depth; a channel inversion-layer to body-junction leakage current component, $I_{ch}(T)$, given as

$$I_{ch}(T) = J_{ch}(T)A_{ch} = J_{ch}(T)(W,L_{\text{eff}})$$  \hspace{1cm} (3.63)

where $L_{\text{eff}}$ is the effective channel length; and a substrate-to-well component, $I_{npw}(T)$, which can be resolved into bottom and side wall components. For a device biased in deep depletion, no inversion layer is formed; thus the channel component of leakage current is eliminated. In a temperature range where the leakage currents are non-negligible, $T > 200 ^\circ\text{C}$, the temperature dependence of drain-to-body leakage currents, is represented as [Shoucair 1984]

$$I_{db}(T) = qA \frac{D_p}{L_p} \rho_i^2(T) \frac{1}{N_d} \alpha AT^3 e^{-E_g(T)/kT}$$  \hspace{1cm} (3.64)

where $A$ is the area of the drain-to-body p-n junction, $D_p$ is the diffusion constant (subscript p for p-channel or n for n-channel), $L_p$ is the diffusion length (subscript p for p-channel or n for n-channel), $N_i(T)$ is the intrinsic carrier concentration, $N_d$ is the substrate donor doping concentration (donor for p-channel or acceptor for n-channel), $\alpha$ is the constant of proportionality, and $E_g(T)$ is the band-gap energy, $\approx 1.2 - 2.73 \times 10^{-4} T$.

**Subthreshold Leakage Currents.** Subthreshold leakage current flows in a leakage path through the transfer device channel while the cell (word line) gate voltage is down and the device is off. The magnitude of the leakage current depends on how
well the transfer device channel can be turned off during standby, which in turn depends on the gate voltage and threshold voltage of the device. Even at the threshold voltage, the drain current is not truly zero and is of a $10^{-4}$ order of magnitude, multiplied by the width-to-length ratio of the device channel [Noble 1984]. Subthreshold leakage current is highly temperature-sensitive. In the subthreshold bias region, $|V_{GS}| < |V_{TH}(T)|$, and $|V_{DS}| > > kT/q$, the channel of the MOSFET is weakly inverted, and the drain current results from carriers diffusing from source to drain. The temperature dependence of drain current is represented by

$$I_D(T) = \mu(T) \left( \frac{W}{L} \right) C_{ox}^2 \left( \frac{n_s(T) kT}{e} \right)^{\frac{e(V_{GS}-V_{TH}(T))}{n_s(T) kT}}$$

(3.65)

where $C_d$ is the depletion capacitance of the space-charge region per unit area; and the temperature-dependent parameters are mobility $\mu(T)$, threshold voltage $V_{TH}(T)$, and subthreshold parameter $n_s(T)$ [Shoucair 1989]. The value of $n_s(T)$ increases exponentially with temperature beyond temperatures in the neighborhood of $150^\circ C$ to $200^\circ C$ (Figure 3.7). Beyond this range, the diffusion leakage currents completely dominate the weak inversion drain-current characteristics. Thus, the temperature at which the subthreshold parameter starts to increase exponentially denotes the reasonable upper operating limit of the device, above which no turn-off can be achieved. To make sure that the device does not make a significant contribution to node leakage at high temperatures, the device off-current must be maintained at $10^{-13}$ A, or five decades below the current at the voltage threshold level [Noble 1984].

### Table 3.1 Variation in Intrinsic Carrier Concentration, Epitaxial Resistivity, and Collector-Emitter Saturation Voltage for Temperatures Between $20^\circ C$ and $200^\circ C$. [Bromstead 1991]

<table>
<thead>
<tr>
<th>Temperature (°C)</th>
<th>Intrinsic carrier concentration, $n_i$ (cm$^{-3}$)</th>
<th>Epitaxial resistivity (ohm-cm)</th>
<th>$V_{CE(on)}$ (volts)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Measured value</td>
<td>Calculated value</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>$4.85 \times 10^8$</td>
<td>0.1103</td>
<td>0.55</td>
</tr>
<tr>
<td>80</td>
<td>$3.68 \times 10^{11}$</td>
<td>0.1486</td>
<td>0.60</td>
</tr>
<tr>
<td>140</td>
<td>$8.42 \times 10^{12}$</td>
<td>0.1952</td>
<td>0.72</td>
</tr>
<tr>
<td>200</td>
<td>$9.07 \times 10^{13}$</td>
<td>0.2496</td>
<td>0.85</td>
</tr>
</tbody>
</table>

$V_{TH}(T)$, and subthreshold parameter $n_s(T)$ [Shoucair 1989]. The value of $n_s(T)$ increases exponentially with temperature beyond temperatures in the neighborhood of $150^\circ C$ to $200^\circ C$ (Figure 3.7). Beyond this range, the diffusion leakage currents completely dominate the weak inversion drain-current characteristics. Thus, the temperature at which the subthreshold parameter starts to increase exponentially denotes the reasonable upper operating limit of the device, above which no turn-off can be achieved. To make sure that the device does not make a significant contribution to node leakage at high temperatures, the device off-current must be maintained at $10^{-13}$ A, or five decades below the current at the voltage threshold level [Noble 1984].
3.3.6 Chip Availability

The signal loss due to leakage current is proportional to time. To prevent the signal from falling below the magnitude needed for sensing, the cells in a dynamic random access memory (DRAM) need to be refreshed periodically. Chip availability is defined as the time that the chip is actually available for read and write operations, mathematically represented as

\[
\text{Availability} = \frac{t_1 - 2n \tau_y}{t_1} \times 100\% \tag{3.66}
\]

where \( t_1 \) is the refresh interval for a given cell, and \( \tau_y \) is the amount of time spent actually performing the refresh operation [Noble 1984]. If a device with a 30-nm cell-oxide thickness and a 200-mV loss of cell voltage due to leakage is maintained at temperatures below 85°C to 95°C, the chip availability is in the neighborhood of 98% to 99%. Once a chip designed for operation at 85°C to 95°C is operated at temperatures greater than 100°C, the chip leakage current becomes so high that most of the time is used just to maintain the data in the cells [Noble 1984].

3.3.7 DC Transfer Characteristics

D.C. transfer characteristics maintain their integrity up to temperatures in the neighborhood of 270°C [Shoucair 1984, Marazas 1992, Zhu 1992]. At 275°C, the transfer characteristic degrades into a flat line, due to the onset of pn-pn latchup phenomenon (Figure 3.8). The peak low-level output voltage, \( V_{\text{OLP}} \), decreases and the peak high-level output voltage, \( V_{\text{OHF}} \), increases with temperature in the range of -50°C to 150°C. The noise margins thus increase with an increase in temperature (Figures 3.9 and 3.10).

The safe operating area of a typical MOSFET is shown in Figure 3.11. The d.c. or pulse drain currents are limited only by the power dissipation slope (\( I_D = \text{max, allowable power dissipation}/V_{\text{DS}} \)) and maximum allowable temperature. The slope in Figure 3.11 is for a device protected from second breakdown by either an external or a built-in clamp like the zenner diode, which breaks in advance of the MOS transistor drain breakdown.
Figure 3.7 Subthreshold parameter versus temperature. At temperature greater than 150-200 deg.C, diffusion leakage currents completely dominate the weak inversion drain characteristics, causing the subthreshold parameter $n_0 (T)$ to exponentially increase with temperature. Depending on the device design, the temperature at which the subthreshold currents start to increase exponentially, typically represents the practical upper operation junction temperature above which no reasonable device turn-off can be achieved. [Shoucair, 1989]
Figure 3.8 DC transfer characteristic of a typical CMOS inverter with temperature as a parameter. At temperature > 270 deg.C, curves degenerate into a flat line due to onset of pnp latchup phenomenon. [Shoucair, 1984]
Figure 3.9 The variation in the peak low-level output voltage versus steady state temperature. The reduction in output indicates an increase in the noise margins at higher temperatures. [Texas Instruments, 1987]
Figure 3.10  The variation in the peak high level output voltage versus steady state temperature. The increase in the output indicates an increase in noise margins at high temperatures. [Texas Instruments, 1987]

- $V_{cc} = 5$ Volt
- $C_L = 50 \text{ pF}$
- $R_L = 0$
- Package = DIP

![Graph showing the variation of high level output voltage with temperature for different devices.](Image)
Figure 3.11
Typical Safe Operating Area (SOA) of a power MOSFET, with device protected from second breakdown [Blicher, 1981]
4

A PHYSICS-OF-FAILURE APPROACH TO IC BURN-IN

4.1 INTRODUCTION

Generally associated with microelectronic devices screening is a process that detects defects in a sub-population of devices that may exhibit early failure characteristics unlike the main population. Such defects occur due to variabilities in materials, manufacturing process, assembly process, workmanship, handling, or poor design. Defects are detected either through non-stress screens, like visual inspection and other non-destructive evaluation methods, or by stress screens, including burn-in, which involve the application of stress to precipitate defects on an accelerated basis.

Over the years, the process of burn-in has deteriorated into an insurance policy used to check reliability or to satisfy customer-imposed requirements. Burn-in procedures are often conducted without any prior verification of the nature of the defects to be precipitated, the failure mechanisms active in the device, their sensitivity to steady-state temperature stress, or quantitative evidence of the improvement achieved by the process. In fact, current failure data indicate that
burn-in prior to use does not remove many failures; on the contrary it may cause failures due to additional handling. This section examines the problems with existing burn-in methods and presents a physics-of-failure approach to burn-in.

4.2 BURN-IN PHILOSOPHY

Screening is a process to detect defects in devices in order to reduce or eliminate failures in the field, often under the assumption that the population of parts which fails early exhibit failure characteristics different from the main population, which lasts the designed-for life. Defects are either detected by non-stress screens, like visual inspection and other non-destructive evaluation methods, or by stress screens, which involve the application of stress (not necessarily service stresses) to precipitate defects often on an accelerated basis. Defects occur in devices due to variabilities in materials, manufacturing, process, assembly process, workmanship, handling, or poor design causing early occurrence of failures.

Good screens cause minimal damage to properly designed and manufactured components, while precipitating the defects that cause early failures in the life of the product. Additionally, the screen can aid in product development during prototype design, manufacture, assembly, handling, storage, and transportation, and arrive at the optimal combination of manufacturing and assembly process parameters. The development of a screen should be based on failure criteria and the results of a root-cause analysis. The failure criteria define device failure as an open, a short, or a limit on performance parameters, and the root-cause analysis determines the actual source of the failure in terms of improper manufacturing process parameters, design inconsistencies, failure mechanisms, and failure sites.

4.3 PROBLEMS WITH PRESENT APPROACH TO BURN-IN

Burn-in is a screen that subjects devices to extended operation at high temperatures in order to precipitate defects. A review of the burn-in practices used by some leading IC manufacturers reveals that even though burn-in has been regarded as a method for eliminating margin, devices with defects from manufacturing aberrations, the specifics of burn-in vary (Table 4.1). Most companies have their own burn-in specifications for commercial products; MIL-STD-883 is used to satisfy burn-in requirements for military products. Other companies use only MIL-STD-883, but the selection of Method 1015 burn-in procedures for quality assurance also seems to be arbitrary.

Burn-in at present is a generic procedure consisting of a combination of time,
steady-state temperature, and electrical stress, not targeted toward any specific defect types or failure mechanisms. The emphasis is on empirical analysis, without any pointers to cost-effective application or subsequent manufacturing or assembly process modifications. The low precipitation of defects from burn-in results because very few microelectronic device failure mechanisms are steady-state temperature dependent. Yet, increased steady state temperature for prolonged periods seems to be the common approach to burn-in.

The misapplication of burn-in has its roots in the association of higher reliability with lowered steady-state temperature, which is generally considered a key parameter in the design of electronic equipment. Cautions about steady-state temperature are widely disseminated. The widespread trust placed in the burn-in process is based on the assumption that failure mechanisms are steady-state temperature-dependent, even though it has generally not been determined whether steady-state temperature, temperature change, the rate of temperature change, or temperature gradients induce failure. Burn-in is typically a requirement imposed by the manufacturer to demonstrate higher product reliability; manufacturers have different burn-in procedures for the same class of components for military and commercial customers. For example, TriQuant Semiconductor [Allen 1991] found while testing their GaAs ICs that burn-in was ineffective in actuating any failure mechanisms. The reasons were traced to device architecture and failure mechanisms that had no dependence on steady-state temperature. Components inserted into and extracted from sockets, temperature chambers, and pre-and post-test procedures can suffer additional handling damage in the form of bent leads and electrostatic discharge. An example of burn-in-actuated failures was observed by Swanson and Licari [1986], who examined the effects of burn-in on hybrid package hermeticity. They found that low leak rates before screening in many hybrid packages increased by over an order of magnitude after burn-in. Packages that maintained a leak rate of $10^9$ atm-cc/sec after burn-in had a moisture content of 3,000 to 5,000 ppm after screening, compared with less than 1,000 ppm before. The increase in leak rates was attributed to the failure of glass-to-metal lead seals and lid seals, which were found to open during burn-in and close again. The increase in moisture was further attributed to outgassing of die and substrate attach epoxies inside the package.

Historically, ionic contamination has been the dominant mechanism precipitated by burn-in [Allen 1991]. Sodium, potassium, or ions in the oxide of silicon MOS devices under bias and temperature lead to junction leakage and threshold voltage shifts that cause failure. Cool-down under bias and retest within 96 hours of burn-in produces a relaxation of bias-induced charge separation. GaAs MESFET-based ICs have no oxide between the gate metallization and the surface of the channel -
- the interface is a Schottky diode. Similarly, the MESFET device is unipolar, majority-carrier conducting, and reliant on the semi-insulating bulk material to achieve device isolation. There are no junctions and no leakage to consider.

In an effort to improve reliability, microelectronic manufacturers have often subjected devices to increasingly longer periods of burn-in. However, Motorola noted that most of the failures precipitated by burn-in occur in the first 160 device hours, with few or no failures over the next 1,000 hours. This is controlled by the fact that the long-term projected failure rate, based on the number of failures over 1,000 hours, is of the same order of magnitude as the actual measured failure rate over 1,000 hours [Motorola 1990].

4.4 A PHYSICS-OF-FAILURE APPROACH TO BURN-IN

4.4.1 Understanding Steady-State Temperature Effects

Microelectronic reliability is typically modeled with steady-state temperature-dependent models, such as the Eyring and Arrhenius models, originally proposed to model the effect of steady-state temperature on chemical reaction rates. These models predict electronic component failure rates under the assumption that the dominant component failure mechanisms depend on the steady-state temperature. However, most microelectronic device failure mechanisms are not highly steady-state temperature-dependent in the equipment operating range [Pecht, et al. 1991]. The use of burn-in without attention to the dominant failure mechanisms and the nature of their temperature dependencies is a misapplication of reliability concepts. Such use of burn-in may cause failure avoidance efforts, without yielding anticipated overall results, or expensive system implementations whose costs and complexities exceed the anticipated benefits in reliability. While burn-in can be used for certain types of failures that can be accelerated by steady-state temperature effects, more insight into the failure mechanism can yield better solutions in terms of design and processes.

Microelectronic design and corrective action are often misdirected because of the confusion between reliability and performance. Reliability is a measure of the ability of a device to fulfill its intended function. Device performance, defined by electrical parameters such as threshold voltage, leakage currents, and propagation delay, is dependent on steady-state temperature. While burn-in may not affect the reliability of the device, it may uncover the fact that at high temperature, the device does not meet performance requirements. This may serve as an indicator for a design change, or for the unsuitability of the technology for high-temperature operation. Burn-in, where performance is checked after the temperature is
lowered, will not uncover this type of problem.

Temperature-cycle dependence is often confused with steady-state temperature dependence. Temperature cycle is a change in temperature that may impose a cyclic stress on the component, due to thermal mismatch at the interfaces of package elements; accelerate intermetallic formation at bimetallic interfaces; or cause stress-induced diffusive voiding in metallization, due to creep-stress relaxation mechanisms. Temperature-cycle stress imposes cumulative damage on the component in the form of thermo-mechanical fatigue. Burn-in is currently only a single temperature cycle.

4.4.2 Setting Up The Burn-In Profile

A physics-of-failure approach to burn-in considers the potential material defects, design inconsistencies, and manufacturing variabilities for each process that could cause defects in the product. The burn-in methodology is an iterative process consisting of the following major steps:

- **Identify failure mechanisms, failure sites and failure stresses:** Development of a burn-in program encompasses identifying potential failure mechanisms, failure sites, and failure stresses, active in a device technology. The burn-in process must be tailored to specific failure mechanism(s) at specific potential failure site(s) in order to be effective. The failure mechanism(s) and failure sites(s) depend on the materials and product processing technologies. Burn-in conditions are therefore specific to the manufacturing technology and hardware. The manufacturing sequence should be studied and possible defects, introduced due to the processing variabilities at each manufacturing stage, should be identified. The dominant failure stresses accelerating the failure mechanisms can be identified based on knowledge of the damage mechanics. The burn-in stress sequence will encompass those stresses that serve as dominant accelerators of the failure mechanisms.

- **Identify the combination of stresses to activate the identified failure mechanisms cost-effectively:** Typically, there may be a number of failure mechanisms dominant in a device technology; each may have a dominant dependence on a different stress. Thus, in order to activate all the failure mechanisms, the dominant stresses need to be applied simultaneously and cost-effectively. To quantitatively determine the magnitude of stresses necessary to activate the failure mechanisms and arrive at the desired cost-effective combination of stresses, models must be developed for each failure mechanism, as a function of stresses, device geometry, material, and magnitude of defects and design
inconsistencies. The quantitative models, however, can aid only in relating the magnitude of a particular type of stress to manufacturing flaws and design inconsistencies, based on physics-of-failure concepts. The real case is more complex, involving interactions of stresses causing failure earlier than predicted by superposition of different stresses acting separately. There may be more than one failure mechanisms in a device technology. Each of the mechanisms will have its own dependence on steady state temperature, temperature cycle, temperature gradient, and time dependent temperature change. An ideal case will be when an optimized combination of the relevant stresses is used to activate the failure mechanisms in a cost effective manner. The desired combination of stresses is a function of the physics of the failure mechanism, and response of package material and configuration. The approach to arriving at the desired stress level consists of subjecting the components to discrete stress levels of steady state temperature, temperature cycle, temperature gradient, time dependent temperature change, and voltage. The selection of the temperature stress level should be based on knowledge of designed-for temperature of the device, the temperature of the device during operational life and the thresholds for various failure mechanisms. Conducting step stress and HAST tests for various magnitudes of each type of stress applied separately will give failure results in terms of number of cycles to failure, failure mechanisms activated, and failure sites. From the test results the stress levels required for activation of failure mechanisms will be identified.

- **Conduct burn-in and evaluate effectiveness**: Burn-in should be assessed based on root cause failure analysis of the failed components, revealing the failure mechanisms, failure modes, and failure sites. Inappropriate burn-in stresses will either damage good components by activating mechanisms not otherwise noticed in operational life, or allow defective parts to go through. To make sure the stress level is right, the amount of damage (or the life consumed) in the case of the products without defects ("good" products) must be evaluated and the stress levels modified if necessary. Product reliability (due to the design improvements) must be used as the index for subsequent burn-in decisions. Physics of failure approach is used to determine the effective acceleration of device dominant failure mechanisms and is given by: \( A_{\text{eff}} = (A_T A_x A_y \ldots) \). Acceleration factors for dominant failure mechanisms are used to determine burn-in time \( t_b \) and temperature \( T_b \). The effective burn-in time is given by: \( t_{\text{eff}} = (A_T A_x A_y \ldots) t_b \). [Jensen 1982]

- **Decisions regarding burn-in modifications and in-process monitors**: The above steps should be repeated until all products have the required expected life, with an optimized return on investment. The burn-in process should be augmented
(supplemented or complemented) with in-line process controls to attain the desired quality and reliability. The physics of failure models along with the burn-in results will determine the optimal manufacturing stress levels and process parameters for minimal defect levels. The in-line process controls will ensure that the process parameters are maintained at their optimal values to minimize the occurrence of defects.

Economic analysis should indicate whether the burn-in should be continued or modified. A cost-effective burn-in program addresses all the relevant failure mechanisms by employing a minimum set of devices. Burn-in is recommended for all products which are in the development stage and do not have a mature manufacturing process. Burn-in at this stage not only improves the reliability of the products but also assists in determining product and process (manufacturing, assembly, testing) corrective actions. Products with a standardized design and a relatively mature process need burn-in only if the field failure returns indicate infant mortality. A cost analysis and return on investment is conducted to calculate the economics of the burn-in program. Analysis of cost and return of investments based on the customer satisfaction and the hidden factory costs (the costs associated with the factory-inputs which do not add value to the product, like product inspection, testing, rework, etc.) determine the profits to an organization. Burn-in economics are critical in convincing management about the benefits that accrue from burn-in and provide a benchmark for making improvements in the next burn-in "cycle".
Table 4.1 Burn-in Time and Temperatures

<table>
<thead>
<tr>
<th>Source</th>
<th>Min. Temp. $T_a$ (°C)</th>
<th>Time (hours)</th>
<th>Test Condition</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>125</td>
<td>Class S: 240 hours</td>
<td></td>
<td>Any of the specified combinations of time and temperature can be used for burn-in according to Method 1015 of MIL-STD-883. The various conditions of burn-in are defined by the electrical stress, steady state temperature (ranging from 100°C to 250°C), and time period (12 to 352 hours). Conditions include:</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Class B: 160 hours</td>
<td></td>
<td>- Test Condition A: Steady state temperature, reverse bias</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>- Test Condition B: Steady state temperature, forward bias</td>
</tr>
<tr>
<td></td>
<td>130</td>
<td>Class S: 208 hours</td>
<td></td>
<td>- Test Condition C: Steady state temperature, power and reverse bias</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Class B: 138 hours</td>
<td>A - E</td>
<td>- Test Condition D: Parallel excitation</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>- Test Condition E: Ring excitation</td>
</tr>
<tr>
<td></td>
<td>135</td>
<td>Class S: 180 hours</td>
<td></td>
<td>accelerated test</td>
</tr>
<tr>
<td></td>
<td>140</td>
<td>Class S: 160 hours</td>
<td>A - E</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Class B: 105 hours</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>145</td>
<td>Class S: 140 hours</td>
<td>A - E</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Class B: 92 hours</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>150</td>
<td>Class S: 120 hours</td>
<td>A - E</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Class B: 80 hours</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>175</td>
<td>Class B: 48 hours</td>
<td>F</td>
<td></td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>Class B: 28 hours</td>
<td>F</td>
<td></td>
</tr>
<tr>
<td></td>
<td>225</td>
<td>Class B: 16 hours</td>
<td>F</td>
<td></td>
</tr>
<tr>
<td></td>
<td>250</td>
<td>Class B: 12 hours</td>
<td>F</td>
<td></td>
</tr>
<tr>
<td>INTEL Corporation</td>
<td></td>
<td>Memory Products: 48 hours</td>
<td></td>
<td>Dynamic burn-in</td>
</tr>
<tr>
<td>Intel Spec.</td>
<td>125°C</td>
<td></td>
<td></td>
<td>[Intel, 1989; Intel, 1990]</td>
</tr>
<tr>
<td>MIL-STD-883 Method 1015</td>
<td>125°C</td>
<td>Military Products: 160 hours</td>
<td>Method 1015, Condition C, or D</td>
<td></td>
</tr>
</tbody>
</table>
## A PHYSICS-OF-FAILURE APPROACH

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>LSI Logic Corporation</td>
<td>MIL-STD-883 Method 1005</td>
<td>125°C 48 hours</td>
<td>Static or DC burn-in; Dynamic burn-in</td>
<td>The results of production burn-in are measured as a percentage fallout rate or PDA (Percent Defective Allowable). The PDA calculation is simply the reject rate, the number of failures divided by the total number of devices in the lot, and the result compared against target PDA. [LSI Logic, 1990]</td>
</tr>
<tr>
<td>Texas Instruments Inc.</td>
<td>MIL-STD-883 Method 1015</td>
<td>125°C Min. MOS Memory and LSI</td>
<td></td>
<td>PDA = 5% [Texas Instruments, 1988]</td>
</tr>
</tbody>
</table>
| Motorola Inc. Monitored Burn-in | 125°C 85°C | 6 hours 10 minutes | Electrical and parametric measurement | Monitored burn-in is a technique in which devices are operated at elevated temperature and voltage for an extended period, followed by short duration at a lower temperature and voltage during which parametric and functional tests are performed (Eschua, MOTOROLA, 1984). The test sequence consists of two stages which include:
* Temperature is increased from ambient to 125°C in 15 minutes and kept at that temperature for 6 hours.
* Temperature is lowered from 125°C to 85°C and allowed to stabilize for 10 minutes before electrical and parametric measurements are performed (MOTOROLA, 1991) |
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>IDT Spec.</td>
<td>125°C Min.</td>
<td>Commercial Products: 16 hours</td>
<td>Method 1015, Condition D</td>
<td></td>
</tr>
<tr>
<td>MIL-STD-883 Method 1015</td>
<td>125°C Min.</td>
<td>Military Hermetic Modules: 44±4 hours</td>
<td>Method 1015, Condition D</td>
<td></td>
</tr>
<tr>
<td>Cypress Semiconductor Inc.</td>
<td>150°C</td>
<td>Level 2 plastic and hermetic parts: 12 hours</td>
<td>Method 1015, Condition C or D</td>
<td>Either of the conditions is used (Cypress Semiconductor, 1990)</td>
</tr>
<tr>
<td>Cypress Semiconductor Spec. MIL-STD-883 Method 1015</td>
<td>125°C Min., or 150°C</td>
<td>JAN, SMD/Military Grade Products: 160 hours 80 hours</td>
<td>Method 1015, Condition C or D</td>
<td></td>
</tr>
<tr>
<td></td>
<td>125°C</td>
<td>Military Grade Modules: 48 hours</td>
<td>Method 1015, Condition C or D</td>
<td></td>
</tr>
</tbody>
</table>
5.1 INTRODUCTION

Derating is a technique through which either the stresses acting on a part are reduced, or the strength of the part is increased, in correspondence with allocated or rated strength stress factors. Manufacturers of microelectronic devices often specify supply voltage limits and threshold values for power dissipation, junction temperature, frequency, and output current. Given these rated values, the equipment designer often elects to arbitrarily derate, or lower the allowed values, to provide a safety factor or "factor of ignorance". When the equipment designer decides to select an alternative component or make a design change that maintains the operational condition for a specific parameter, such as temperature, consistently below the rated level, the component is said to have been derated for thermal stress. Thermal derating is one of the most common derating methodologies, but is not the only one. The reliability of electronic systems is,
obviously, very sensitive to component derating and exactly how the derating is applied.

5.2 PROBLEMS WITH THE PRESENT APPROACH TO DEVICE DERATING

Currently, the thermal derating criteria vary among government acquisition agencies. Westinghouse has a set of derating guidelines for electronic components, in which ECL, TTL, and CMOS digital components, linear devices, and hybrids are equally derated for thermal stress. Derating for thermal stress involves reducing the maximum operating temperature to around 0.6Tj (maximum junction temperature). Rome Laboratories has a set of derating guidelines in which microcircuits that have been grouped as: complex, digital, hybrid, linear, and memory are derated to operate at a lower operating temperature. The exact value of the operating temperature depends on the derating level. There are three derating levels, depending on the level of criticality of the component. In all cases, derating for thermal stress involves lowering the maximum junction temperature.

Implicit in such a derating methodology is the assumption that steady-state temperature is a dominant accelerator of microelectronic device failure mechanisms. Even in microelectronic technologies where temperature is a dominant failure accelerator, steady-state temperature thresholds, affecting device sensitivity to various forms of temperature stress, are a function of device architecture, materials, manufacturing defects, and other non-temperature-related operational stresses. Assigning a generic value of lower temperature to a technology, based on the assumption that all devices operating at that lower value of temperature will be reliable, is thus arbitrary.

Thermal-stress derating guidelines rule out the option of reliable system designs at higher temperatures. In addition, they give the designer a false sense of security about achieving increased reliability at lower temperatures. Lower temperatures may not necessarily increase reliability, since some of the failure mechanisms are inversely dependent on temperature; for example, device technologies with hot electrons as the dominant mechanism may have lower reliability at lower temperatures.

Current thermal derating guidelines do not account for the dominant failure mechanisms or their temperature dependencies. Temperature-cycle effects that have not been accounted for in device derating criteria are failure accelerators at mating interfaces. The maximum number of temperature cycles that the device can endure is a function of fatigue failure mechanisms, such as wire-interconnect
fatigue, die fracture, or die and substrate attach fatigue. These mechanisms may or may not be dominant in the device architecture, depending on the stresses generated at the mating interfaces, which are a function of interface geometry, and on material characteristics, including CTE mismatches.

Typically, localized temperature gradients exist in the chip metallization, chip, substrate, and package case, due to sudden variations in the conductivities of material produced by defects in the form of voids or cracks. The locations of maximum temperature gradients in chip metallization are sites for mass transfer mechanisms, including electromigration. Electromigration is also accelerated by temperature and current density. Therefore, device reliability, for mechanisms with a dominant dependence on more than one operating stress, (temperature and non-temperature) complicated by dependence on magnitudes of manufacturing defects, needs to be maximized by more than just lowering temperature, as existing derating criteria do.

The interaction of various temperature and non-temperature stresses modifies the dominant dependence of the failure mechanisms on one or more of the stresses. Temperature transients generated by the duty cycle (ON/(ON+OFF)) modify the dependence of the metallization corrosion on steady-state temperature. At low duty cycle values, metallization corrosion has a dominant dependence on steady-state-temperature. However, at higher values ( \approx \text{in neighborhood of 1.0}), metallization corrosion has a dominant dependence on duty-cycle and a mild dependence on steady-state temperature.

5.3 A PROPOSED APPROACH TO DEVICE DERATING

The derating problem for dominant stress accelerators for a particular device architecture requires derating the stress with maximum sensitivity with respect to life in order to achieve optimization. Some stresses can be easily controlled, while others may be hard to derate because they are a function of device architecture. Therefore, the practical situation may disclose that the most effective parameter is not the easiest to derate. The approach presented here allows the evaluation of the relative sensitivity of operating life to various stresses. Once the dominant stresses have been identified, the stresses which dramatically effect life, and which are also the easiest to control, are derated (Figure 5.1).

The derated values of steady-state temperature, temperature cycle magnitude, temperature gradient, and time-dependent temperature change, including non-temperature operating stresses, are determined for a desired device mission life.
Figure 5.1 Derating Methodology for Temperature Tolerant Design

Constraints
- Desired Mission Life (hours)
- Device Architecture (material, geometry)
- Performance Parameters (P : i = 1 to n)
- Worst Case Manufacturing Defect Magnitudes

Problem
- Derate \( T, \Delta T, \Delta T \frac{a}{2} \) and non-temperature stresses

Performance
- Identify critical performance parameters for the device technology
  - Bipolar
  - MOSFET

Critical dependencies of performance parameters
\[
S(j, i) = \begin{cases} \frac{T_{\text{e}0} \Delta T_{\text{e}0} \Delta T_{\text{e}0} \Delta T_{\text{e}0}}{2} - f(\Delta T_{\text{e}0}) - f(\Delta T_{\text{e}0}) & \\
\frac{T_{\text{e}0} \Delta T_{\text{e}0} \Delta T_{\text{e}0} \Delta T_{\text{e}0}}{2} - f(\Delta T_{\text{e}0}) - f(\Delta T_{\text{e}0}) & \\
\frac{T_{\text{e}0} \Delta T_{\text{e}0} \Delta T_{\text{e}0} \Delta T_{\text{e}0}}{2} - f(\Delta T_{\text{e}0}) - f(\Delta T_{\text{e}0}) & \\
\frac{T_{\text{e}0} \Delta T_{\text{e}0} \Delta T_{\text{e}0} \Delta T_{\text{e}0}}{2} - f(\Delta T_{\text{e}0}) - f(\Delta T_{\text{e}0}) & \\
\frac{T_{\text{e}0} \Delta T_{\text{e}0} \Delta T_{\text{e}0} \Delta T_{\text{e}0}}{2} - f(\Delta T_{\text{e}0}) - f(\Delta T_{\text{e}0}) \end{cases}
\]

Reliability
- Identify dominant failure mechanisms for the device architecture

Critical dependencies of dominant failure mechanisms
\[
S(j, i) = \begin{cases} \frac{T_{\text{e}0} \Delta T_{\text{e}0} \Delta T_{\text{e}0} \Delta T_{\text{e}0}}{2} - f(\Delta T_{\text{e}0}) - f(\Delta T_{\text{e}0}) & \\
\frac{T_{\text{e}0} \Delta T_{\text{e}0} \Delta T_{\text{e}0} \Delta T_{\text{e}0}}{2} - f(\Delta T_{\text{e}0}) - f(\Delta T_{\text{e}0}) & \\
\frac{T_{\text{e}0} \Delta T_{\text{e}0} \Delta T_{\text{e}0} \Delta T_{\text{e}0}}{2} - f(\Delta T_{\text{e}0}) - f(\Delta T_{\text{e}0}) & \\
\frac{T_{\text{e}0} \Delta T_{\text{e}0} \Delta T_{\text{e}0} \Delta T_{\text{e}0}}{2} - f(\Delta T_{\text{e}0}) - f(\Delta T_{\text{e}0}) \end{cases}
\]

Performance Related Stress thresholds for temperature and non-temperature operating stresses

For \( i = 1 \) to \( m, j = k \)
If \( \frac{\partial P}{\partial S_{j(i,j)}} < 0 \)
Calculate \( \min(S_{j(i,j)}) \) for \( P < P_{\text{max}} \)
OR
For \( i = 1 \) to \( m, j = k \)
If \( \frac{\partial P}{\partial S_{j(i,j)}} > 0 \)
Calculate \( \max(S_{j(i,j)}) \) for \( P > P_{\text{max}} \)

Reliability Related Stress thresholds for temperature and non-temperature operating stresses

For \( i = 1 \) to \( m, j = k \)
If \( \frac{\partial N_{j}}{\partial S_{j(i,j)}} > 0 \)
Calculate \( \min(S_{j(i,j)}) \) for \( N_{j} > N_{\text{ML}} \)
OR
For \( i = 1 \) to \( m, j = k \)
If \( \frac{\partial N_{j}}{\partial S_{j(i,j)}} < 0 \)
Calculate \( \max(S_{j(i,j)}) \) for \( N_{j} > N_{\text{ML}} \)

Maximum and Minimum allowable values of Stresses
\[
\begin{align*}
\max(S_{j(i,j)}) &= \max(\max(S_{j(i,j)}), \max(S_{j(i,j)})) \\
\min(S_{j(i,j)}) &= \min(\min(S_{j(i,j)}), \min(S_{j(i,j)}))
\end{align*}
\]
Calculated derated levels of temperature and non-temperature operating stresses for devices using approach are specific to a design architecture and cannot be generalized to a device technology. The derating problem can thus be stated as follows:

\[
\text{constraints:}
\]
\[
\begin{align*}
\text{desired mission life (hours)} \\
\text{device architecture (material, geometry)} \\
\text{performance parameters} (P_i; i=1 \text{ to } m) \\
\text{worst case manufacturing defect magnitudes}
\end{align*}
\]

\[
\text{problem:}
\]
\[
\text{derate } (T, \Delta T, VT, \frac{\partial T}{\partial t}) \text{ and non-temperature stresses}
\]

where \( P_i; i=1 \text{ to } m \) are the critical device performance parameters. Derating stress values involves addressing both performance and reliability requirements. While derating for performance evaluates stress influence on critical device parameters, derating for reliability involves evaluation of stress influence on device life under dominant failure mechanisms.

Device parameters considered in derating for performance are a function of device technology, which, for BJT devices, includes device thermal voltage, current gain, and the invertor voltage transfer characteristic and for MOSFET devices includes, threshold voltage, the invertor voltage transfer characteristic, and the invertor propagation delay:

\[
\text{Performance} = \begin{cases} 
\mathcal{F}(V_{\text{thermal}}, \beta, IVTC) & \text{for BJT devices} \\
\mathcal{F}(V_{\text{threshold}}, IVTC, IPD) & \text{for MOSFET devices}
\end{cases}
\]

(5.2)

where \( V_{\text{thermal}} \) is the thermal voltage, \( \beta \) is the device current gain, \( V_{\text{threshold}} \) is the threshold voltage, \( IVTC \) is the invertor voltage transfer characteristic, and \( IPD \) is the invertor propagation delay. Threshold values for temperature and non-temperature stresses that cause the critical parameters of the device technology to exceed acceptable ranges are calculated based on closed-form models relating device architecture to critical parameters.
where $S_p$ is a $(m \times n)$ matrix of threshold stress values for device performance. The $m$ rows correspond to $m$ (i.e., $p(i)$, for $i = 1$ to $m$) critical parameters for the device technology (such as threshold voltage, the inverter voltage transfer characteristic, and the inverter propagation delay, for MOSFET devices). The columns correspond to the $n$ temperature and non-temperature stresses that affect the critical parameters. The acceptable range of stress values for device performance is calculated as follows:

For $i = 1$ to $m$; $j = k$

If \[
\frac{\partial P_i}{\partial S_p(i,j)} \bigg|_{i=1 \text{ to } m, j=k} > 0 ; \text{ calculate } \min(S_p(i,j)) \bigg|_{i=1 \to m, j=k} \text{ for } P_i < P_{\max}
\]

(5.4)

For $i = 1$ to $m$; $j = k$

If \[
\frac{\partial P_i}{\partial S_p(i,j)} \bigg|_{i=1 \text{ to } m, j=k} < 0 ; \text{ calculate } \max(S_p(i,j)) \bigg|_{i=1 \to m, j=k} \text{ for } P_i > P_{\min}
\]

(5.5)

where $k$ represents a particular stress value.

The device life, determined by dominant failure mechanisms actuated by temperature and non-temperatures stresses, is used to derate for reliability. The influence of stresses on device life is quantified by closed-form models. The threshold stress values that result in a device life less than the desired mission life are calculated by
A PROPOSED APPROACH

\[
S_r(i,j) = \begin{pmatrix}
T_r(1) & \Delta T_r(1) & \nabla T_r(1) & \frac{\partial T_r(1)}{\partial t} & J_r(1) & E_r(1) & \ldots & S(1,n) \\
T_r(2) & \Delta T_r(2) & \nabla T_r(2) & \frac{\partial T_r(2)}{\partial t} & J_r(2) & E_r(2) & \ldots & \ldots \\
T_r(3) & \Delta T_r(3) & \nabla T_r(3) & \frac{\partial T_r(3)}{\partial t} & J_r(3) & E_r(3) & \ldots & \ldots \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ldots & \ldots \\
S(m,1) & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & S(m,n)
\end{pmatrix}
\]

where \( S_r \) is a matrix of threshold stress values for device reliability \((m \times n)\). The \( m \) rows correspond to \( m \) (i.e., \( p(i) \), for \( i = 1 \) to \( m \)) dominant failure mechanisms. The \( n \) columns correspond to the \( n \) temperature and non-temperature stresses that affect device life limitations resulting from the different failure mechanisms. The acceptable range of stress values for device reliability is calculated as follows:

For \( i = 1 \) to \( m \); \( j = k \)

\[
\text{If } \left. \frac{\partial N_f}{\partial S_r(i,j)} \right|_{j-k} \geq 0 \text{; calculate } \min(S_r(i,j)) \text{ for } i = 1 \text{ to } m \text{ for } N_f > ML
\]

(5.7)

For \( i = 1 \) to \( m \); \( j = k \)

\[
\text{If } \left. \frac{\partial N_f}{\partial S_r(i,j)} \right|_{j-k} \leq 0 \text{; calculate } \max(S_r(i,j)) \text{ for } i = 1 \text{ to } m \text{ for } N_f > ML
\]

(5.8)

where \( k \) represents a particular stress value, \( i \) represents a particular failure mechanism, \( N_f \) is the predicted time to failure due to a failure mechanism, and \( ML \) is the required device mission life. The maximum and minimum allowable values of stresses are then determined as:

\[
\max(S_d(i,j)) = \max[\max(S_p(i,j)), \max(S_r(i,j))]
\]

(5.9)

\[
\min(S_d(i,j)) = \min[\min(S_p(i,j)), \min(S_r(i,j))]
\]

where \( \max(S_r(i,j)) \) is the maximum allowable stress value derived from reliability
considerations, \( \text{max}(S_{ij}) \) is the maximum allowable stress value derived from performance considerations, and \( \text{max}(S_{ij}) \) is the maximum allowable stress value for the device. \( \text{Min} \) represents the minimum values of the individual stresses. Maximum values are specified for stresses directly proportional to detrimental effects on package reliability and performance. Minimum values are specified for stresses with an inverse influence on package performance or reliability.

Physics-of-failure concepts have been used to relate allowable operating stresses to design strengths through quantitative models for failure mechanisms. Failure models have been used to assess the impact of derating on the effective reliability of the component for a given load. The quantitative correlations outlined between derating and reliability will enable designers and users to tailor the margin of safety more effectively to the level of criticality of the component, leading to better and more cost-effective utilization of the functional capacity of the component.

5.4 DERATING FOR FAILURE MECHANISMS IN DIE METALIZATION

5.4.1 Corrosion of Die Metallization

Metallization corrosion has a dominant dependence on two temperature stresses - steady-state temperature and time-dependent temperature change. Temperature transients generated by the duty cycle \((\text{ON}/(\text{ON}+\text{OFF}))\) modify the dependence of the metallization corrosion on steady-state temperature. At low values of the duty cycle, metallization corrosion has a dominant dependence on steady-state temperature; however, at higher values \((\approx \text{in neighborhood of } 1.0)\), metallization corrosion has a dominant dependence on the duty cycle and a mild dependence on steady-state temperature. Time to failure due to metallization corrosion has been modelled as the sum of induction time and corrosion time [Pecht 1990]. Calculate \( T_{\text{max}} \) and \( (\partial T / \partial t)_{\text{max}} \) such that:

\[
T_{\text{operation}} < T_{\text{max}} \quad \frac{\partial T}{\partial t} < \left( \frac{\partial T}{\partial t} \right)_{\text{max}}
\]  

(5.10)
Table 5.1 Derating Guidelines For Thermal Derating

<table>
<thead>
<tr>
<th>Microcircuit Manufacturer/ Acquisition Agency</th>
<th>Component Type</th>
<th>Derating Criteria (max. allowable junction Temperature)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Westinghouse Electric [Westinghouse 1986]</td>
<td>digital: TTL/ECL</td>
<td>0.6 ( T_{j\text{(Rned)}} )</td>
</tr>
<tr>
<td></td>
<td>digital: CMOS</td>
<td>0.6 ( T_{j\text{(Rned)}} )</td>
</tr>
<tr>
<td></td>
<td>linear, amplifiers</td>
<td>0.6 ( T_{j\text{(Rned)}} )</td>
</tr>
<tr>
<td></td>
<td>linear, regulators</td>
<td>0.6 ( T_{j\text{(Rned)}} )</td>
</tr>
<tr>
<td>Rome Laboratories [RADC-TR-82-177 RADC-TR-84-254]</td>
<td>complex (LSI, VLSI, VHSIC)</td>
<td>derating level 1 85°C</td>
</tr>
<tr>
<td></td>
<td></td>
<td>derating level 2 100°C</td>
</tr>
<tr>
<td></td>
<td></td>
<td>derating level 3 125°C</td>
</tr>
<tr>
<td></td>
<td>digital</td>
<td>derating level 1 85°C</td>
</tr>
<tr>
<td></td>
<td></td>
<td>derating level 2 100°C</td>
</tr>
<tr>
<td></td>
<td></td>
<td>derating level 3 110°C</td>
</tr>
<tr>
<td></td>
<td>hybrid</td>
<td>derating level 1 85°C</td>
</tr>
<tr>
<td></td>
<td></td>
<td>derating level 2 100°C</td>
</tr>
<tr>
<td></td>
<td></td>
<td>derating level 3 125°C</td>
</tr>
<tr>
<td></td>
<td>linear</td>
<td>derating level 1 80°C</td>
</tr>
<tr>
<td></td>
<td></td>
<td>derating level 2 95°C</td>
</tr>
<tr>
<td></td>
<td></td>
<td>derating level 3 105°C</td>
</tr>
<tr>
<td></td>
<td>memory</td>
<td>derating level 1 85°C</td>
</tr>
<tr>
<td></td>
<td></td>
<td>derating level 2 100°C</td>
</tr>
<tr>
<td></td>
<td></td>
<td>derating level 3 125°C</td>
</tr>
<tr>
<td>Naval Air Systems Command [AS-4613, 1976]</td>
<td>digital: TTL/ECL</td>
<td>( T_{\text{max(ready)}} - 20°C )</td>
</tr>
<tr>
<td></td>
<td>class C</td>
<td>( T_{\text{max(ready)}} - 25°C )</td>
</tr>
<tr>
<td></td>
<td>class B</td>
<td>( T_{\text{max(ready)}} - 30°C )</td>
</tr>
<tr>
<td></td>
<td>class A</td>
<td>( T_{\text{max(ready)}} - 30°C )</td>
</tr>
<tr>
<td></td>
<td>digital: CMOS</td>
<td>( T_{\text{max(ready)}} - 20°C )</td>
</tr>
<tr>
<td></td>
<td>class C</td>
<td>( T_{\text{max(ready)}} - 20°C )</td>
</tr>
<tr>
<td></td>
<td>class B</td>
<td>( T_{\text{max(ready)}} - 30°C )</td>
</tr>
<tr>
<td></td>
<td>class A</td>
<td>( T_{\text{max(ready)}} - 30°C )</td>
</tr>
<tr>
<td></td>
<td>linear, amplifiers</td>
<td>( T_{\text{max(ready)}} - 20°C )</td>
</tr>
<tr>
<td></td>
<td>class C</td>
<td>( T_{\text{max(ready)}} - 25°C )</td>
</tr>
<tr>
<td></td>
<td>class A</td>
<td>( T_{\text{max(ready)}} - 30°C )</td>
</tr>
</tbody>
</table>
\[ T_{f(\text{corr})} > \text{mission life} \]  \hspace{1cm} (5.11)

where

\[ T_{f(\text{corr})} = \tau_{\text{INDUCTION}} + \tau_{\text{CORROSION}} \]

\[ \tau_{\text{INDUCTION}} = -\frac{4L^2}{\pi^2D} \ln \left(1 - \frac{P_{\text{in}}}{P_{\text{out}}} \right) \]  \hspace{1cm} \text{For nonhermetic package} \hspace{1cm} (5.12)

\[ = \text{From figure} \hspace{1cm} \text{For hermetic package} \]

\[ \tau_{\text{CORROSION}} = \frac{k_1 k_2 k_3 w^2 h n d F \rho}{k_4} \frac{p}{4MV Z} \]

and

\[ k_3 = \frac{1}{(1 - DC)^{\text{radc}} - 1} \]  \hspace{1cm} (5.13)

\[ K_4 = \frac{(RH_R)^a \exp(E / KT_R)}{(RH)^a \exp(E / KT)} \]  \hspace{1cm} (5.14)

where:

- \( RH_R \) is a reference relative humidity (%);
- \( E_a \) is an activation energy (eV);
- \( K \) is the Boltzmann constant (eV/°K);
- \( n \) is a material constant;
- \( T_R \) is a reference temperature (°K);
- \( DC \) is the duty cycle;
- \( M \) is the atomic weight of a metal conductor of density \( d \), width \( w \),
  height \( h \), and chemical valence \( n \);
- \( \rho / Z \) is the sheet resistance of the electrolyte;
Table 5.2 Physical and Chemical Properties of Metallization Materials
[Pecht and Ko 1991]

<table>
<thead>
<tr>
<th>Material</th>
<th>Physical Property Index ($k_1$)</th>
<th>Atomic Weight ($M$)</th>
<th>Density ($d$)</th>
<th>Chemical Valence ($n$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aluminum</td>
<td>0.1</td>
<td>27</td>
<td>2.7</td>
<td>3</td>
</tr>
<tr>
<td>Copper</td>
<td>0.7</td>
<td>64</td>
<td>8.9</td>
<td>2</td>
</tr>
<tr>
<td>Gold</td>
<td>1.0</td>
<td>197</td>
<td>19.32</td>
<td>3</td>
</tr>
</tbody>
</table>

$V$ is the voltage applied;
$K_1$ is the physical and chemical properties of the metallization material (from Table 5.2);
$K_2$ is the coating integrity index (from Table 5.3);
$K_3$ is the mission profile correction factor; and
$K_4$ is the environmental stress correction factor.

Derated stress values can be calculated from the curve of constant life versus temperature and duty cycle, obtained from Equation 5.12. (Figure 5.2 shows curves of constant life versus temperature and duty cycle.) It is evident that dependence of metallization corrosion on steady-state temperature and duty cycle is non-linear, and that reducing dominant stress magnitudes beyond a certain limit may not produce any observable benefit in terms of added reliability, because the time to failure is far beyond the wear-out life of the device. The paradigm of associating higher reliability with lower temperature is also misleading, since the same mission life of twenty-nine years can be obtained for any temperature from 40°C to 160°C, depending on the duty cycle (Figure 5.2).

Table 5.3 Coating Integrity Index [Pecht and Ko 1991]

<table>
<thead>
<tr>
<th>Coating Type</th>
<th>Coating Integrity Index ($k_2$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>No coating</td>
<td>1</td>
</tr>
<tr>
<td>Partially bonded</td>
<td>10 - 50</td>
</tr>
<tr>
<td>Completely bonded</td>
<td>100</td>
</tr>
</tbody>
</table>
Figure 5.2 Derating curves of life under corrosion versus temperature and duty cycle. The curves identify the various combinations of temperature and duty cycle which will result in desired life. The paradigm of higher reliability associated with lower temperature is misleading, since the same mission life of 29 years can be obtained for any temperature from 40°C to 160°C depending on the duty cycle. [Pecht, 1990]
5.4.2 Electromigration

Electromigration, a mass-transfer mechanism with a dominant dependence on current density, steady-state temperature, and temperature gradient, is a grain-boundary diffusion mechanism. The time to failure due to temperature stress for \( T < 150°C \) is much greater than the wear-out life. Typically, electromigration failures at \( T < 150°C \) occur at sites of maximum temperature gradient or structural non-uniformity. Steady-state temperature is a dominant stress accelerator for electromigration at \( T > 150°C \) for typical metallization geometries. The maximum temperature and current density combination that can be used for a given metallization can be calculated from the following models:

- Black's Model [Black 1982]
- Shatzkes and Lloyd Model [Shatzkes and Lloyd 1986]
- Venables and Lye Model [Venables and Lye 1972]

Calculate \( T_{\text{max}} \) (max. temperature) and \( j_{\text{max}} \) (current density), where

\[
T_{\text{operation}} < T_{\text{max}} \quad \text{and} \quad j_{\text{operation}} < j_{\text{max}}
\]

such that

\[
MTF > \text{mission life}
\]

where

\[
MTF = \frac{wt}{j^2 A e^{-\frac{E_A}{kT}}}
\]

and

- \( MTF \) is the mean time to failure (hours);
- \( w \) is the metallization width (cm);
- \( t \) is the metallization thickness (cm);
- \( A \) is a parameter depending on sample geometry, physical characteristics of the film and substrate, and protective coating;
- \( j \) is the current density (A/cm\(^2\));
- \( n \) is an experimentally determined exponent;
- \( E_A \) is the activation energy (eV); and
$T$ is the steady-state temperature (Kelvin).

### Table 5.4 Black’s Constants For Various Metallization Materials

<table>
<thead>
<tr>
<th>Metallization Material</th>
<th>$A$</th>
<th>$E_A$ (eV)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Al-2%Si</td>
<td>$0.3119 \times 10^{-14}$</td>
<td>0.558</td>
</tr>
<tr>
<td>Al-4%Cu-4%Si</td>
<td>$7.292 \times 10^{-13}$</td>
<td>0.703</td>
</tr>
</tbody>
</table>

The value of the current exponent derived in recent studies varies from 1.5 to 7. The commonly used current exponents are:

- $n = 1$ to 3 Chabra and Ainslie [1967]
- $n = 1.5$ Attardo [1972]
- $n = 1.7$ Danso and Tullos [1981]
- $n = 2$ Black [1983]
- $n = 6$ to 7 Blair et al. [1970]

The pre-exponential, current exponent, and activation energy for some common metallization materials reported in literature are presented in Table 5.5. Figure 5.3 shows the derating curves for electromigration stress; these identify various combinations of current density and temperature that will result in desired life. Shatzkes and Lloyd Model [Shatzkes and Lloyd 1986]

Calculate $T_{\text{max}}$ (maximum temperature) and $j_{\text{max}}$ (current density), where

$$
T_{\text{operation}} < T_{\text{max}} \\
j_{\text{operation}} < j_{\text{max}}
$$

such that

$$
T_f > \text{mission life}
$$

where

$$
t_f = \left( \frac{2C_f}{D_o} \left( \frac{k}{Z e \rho} \right) \right)^2 T^2 j^{-2} e^{\frac{\Delta H}{k T}}
$$

and where

$C_f$ is the critical value of vacancy concentration at which failure occurs;
occurs;

\( D_0 \) is the pre-exponential factor for grain-boundary self-diffusivity;

\( k \) is the Boltzmann constant;

\( Z^* \) is the effective charge;

\( e \) is the electronic charge;

\( \rho \) is the resistivity;

\( T \) is the steady-state temperature;

\( j \) is the current density; and

\( \Delta H \) is the activation energy.

Equation 5.20 differs from Black's equation in that it has a \( T^2 \) pre-exponential term, but it fits Black's data equally well. Venables and Lye Model [Venables 1972]

Given the allowable temperature rise of \( \Delta T_R \) in the metallization stripe, calculate \( T_{\text{max}} \) (maximum temperature) and \( j_{\text{max}} \) (current density), where

\[
\frac{T_{\text{operation}}}{T_{\text{max}}} < 1
\]

\[
\frac{j_{\text{operation}}}{j_{\text{max}}} < 1
\]

such that

\[
T_F > \text{mission life}
\]

where

\[
T_F = \frac{1}{2CN} \left( \frac{\tau_0 k T_0}{\rho D_0 e^{-\frac{\Delta H}{kT}}} \right) \int_0^{\Delta T_R} \frac{x^2}{x^2(1+x\alpha T_0)} dx
\]

and

\[
\tau_0 = \Delta T_R (1 + \alpha (T_0 - 300)) \left( \frac{1}{T_0} \right) \left( \frac{j_0}{1 \times 10^6} \right)^2
\]
\[
x = \frac{\tau_o}{(1-p)^2 + \tau_o (1-aT_o)}
\]

\(= 1 - \left( \frac{T_o}{T} \right) \)  

(5.25)

where \( \Delta T_m \) is the temperature increase in the metallization stripe; \( a \) is the thermal coefficient of resistivity for metallization (°C); \( T_o \) is the ambient temperature (°K); \( T_m \) is the melting temperature of the metallization stripe (°K); \( Q \) is the activation energy for grain-boundary diffusion (≈ 0.558 eV); \( K \) is Boltzmann’s constant \((8.617 \times 10^{-5} \text{ eV/K or } 1.38 \times 10^{-23} \text{ J/K})\); \( p \) is the porosity of the metallization stripe \((0.1 < p < 1.0)\); \( j \) is the current density \((\text{A/cm}^2)\); \( C \) is the constant for the metallization material; \( n \) is the density of grain-boundary nodes \((/\text{cm}^3)\); \( \rho_o \) is the resistivity of metallization material at ambient temperature \((\Omega \text{ cm})\); and \( D_o \) is the grain-boundary diffusivity of the metallization material \((\text{cm}^2/\text{s})\).
Figure 5.3 Derating curves for electromigration stress. The curves identify various combinations of current density and temperature which will result in desired life [Black, 1982].

Current exponent; $n = 2$
Width = $3e^{-4}$ cm
Thickness = $3e^{-4}$ cm

Temperature (Centigrade)

Current Density (A/cm²)

1 year
29 years
The current exponent for Black's equation (Equation 5.17) changes with the current density, $j$, in the metallization stripe. A constant value of the current exponent, $n$, is valid only over a small range of current densities [Venables 1972]. The Venables and Lye model accounts for the relationship of time to failure and baseline temperature, which is often simplistically represented by an Arrhenius relationship to give an activation energy. The model also addresses the dependence of current exponent on current density.

5.4.3 Hillock Formation

Hillocks in die metallization can form as a result of electromigration or extended periods under temperature cycling conditions (thermal aging). The phenomenon of simultaneous voiding and hillock formation occurs at temperatures in the neighborhood of 140 to 200°C [Thomas 1983]. Hillock formation due to extended periods under temperature cycling conditions is believed to be due to a self-diffusion process that occurs in the presence of strains within the metallization [LaCombe and Christou 1982]. Coating the metal with silicon nitride prevents failures from occurring and voids and hillocks from forming for at least 500 hours at 360°C. Hillocks form at random in aluminum films heated to temperatures around 400°C during fabrication.

5.4.4 Metallization Migration

Calculate the maximum allowable steady-state temperature to avoid metallization migration, where [DiGiacomo 1982].

$$T_{\text{operation}} < T_{\text{max}} \quad (5.27)$$

and

$$j_{\text{kap}} > j_{\text{critical}}$$

$$\frac{\partial Q_{\text{kap}}}{\partial t} > \frac{\partial Q_{\text{critical}}}{\partial t} \quad (5.28)$$

such that

$$t_f > \text{mission life} \quad (5.29)$$

where
Table 5.5 Temperature Acceleration and Current Exponents For Various Metallization Materials

<table>
<thead>
<tr>
<th>Material</th>
<th>Metallization Dimensions (μm)</th>
<th>Average Grain Size (μm)</th>
<th>Test Conditions</th>
<th>Activation Energy (Ev, eV)</th>
<th>Current Density Exponent (n)</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Single Layer Metallizations</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aluminum (Al)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>w = 10</td>
<td>l = 400</td>
<td>t = 1.2</td>
<td>1.3 - 3</td>
<td>75-350°C</td>
<td>0.41</td>
<td>[Satake 1973]</td>
</tr>
<tr>
<td>w = 15.4</td>
<td>l = 0.6</td>
<td>8</td>
<td>0.55 - 2.02</td>
<td>109-260°C</td>
<td>0.84</td>
<td>2</td>
</tr>
<tr>
<td>t = 1.0</td>
<td></td>
<td>-</td>
<td>-</td>
<td>0.56</td>
<td>2</td>
<td>[Black 1982]</td>
</tr>
<tr>
<td>w = 1.3</td>
<td>l = 1524</td>
<td>t = 0.8</td>
<td>0.1</td>
<td>0.75</td>
<td>175-275°C</td>
<td>0.34</td>
</tr>
<tr>
<td>w = 2.3</td>
<td></td>
<td>0.1</td>
<td>-</td>
<td>-</td>
<td>0.47</td>
<td>[Wu 1983]</td>
</tr>
<tr>
<td>w = 3.3</td>
<td></td>
<td>0.1</td>
<td>-</td>
<td>-</td>
<td>0.58</td>
<td>[Wu 1983]</td>
</tr>
<tr>
<td>t = 0.7</td>
<td></td>
<td>0.7-1.8</td>
<td>2</td>
<td>125-300°C</td>
<td>0.24-0.57</td>
<td>[Reimer 1984]</td>
</tr>
<tr>
<td>w = 4.5</td>
<td>l = 2000</td>
<td>t = 0.75</td>
<td>-</td>
<td>2</td>
<td>125-300</td>
<td>0.43</td>
</tr>
<tr>
<td>w = 10</td>
<td>l = 800</td>
<td>t = 1</td>
<td>-</td>
<td>1</td>
<td>160-250°C</td>
<td>0.57</td>
</tr>
<tr>
<td>w = 37</td>
<td>l = 1.2</td>
<td></td>
<td>8</td>
<td>0.46-0.99</td>
<td>109-260°C</td>
<td>1.2</td>
</tr>
<tr>
<td>w = 15-25</td>
<td>l = 800</td>
<td>t = 0.5</td>
<td>0.3</td>
<td>2</td>
<td>105-180°C</td>
<td>0.55</td>
</tr>
<tr>
<td>w = 7-20</td>
<td>l = 500-580</td>
<td>t = 1</td>
<td>0.8</td>
<td>0.5-2</td>
<td>130-200°C</td>
<td>0.7</td>
</tr>
</tbody>
</table>
### DERATING GUIDELINES

<table>
<thead>
<tr>
<th>Component</th>
<th>Width (w)</th>
<th>Time (t)</th>
<th>Temperature (°C)</th>
<th>Derating Factor</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Al-(0.79-3%)Si</td>
<td>w = 10</td>
<td>t = 0.7-0.73</td>
<td>110-245°C</td>
<td>0.55</td>
<td>Black 1978</td>
</tr>
<tr>
<td>Al-1%Si</td>
<td>w = 5</td>
<td>t = 1.1</td>
<td>162-215°C</td>
<td>0.58</td>
<td>Schaff 1985</td>
</tr>
<tr>
<td>Al-1.5%Si</td>
<td>w = 3.7-4.1</td>
<td>t = 1.2</td>
<td>277°C</td>
<td>1.67</td>
<td>Sudhe 1989</td>
</tr>
<tr>
<td>Al-2%Si</td>
<td>w = 10</td>
<td>t = 0.8</td>
<td>240-360°C</td>
<td>0.33</td>
<td>Liew 1989</td>
</tr>
</tbody>
</table>

**Notes:**
- Values are approximate and subject to variations based on specific conditions.
- Reference numbers indicate sources for detailed information.
<table>
<thead>
<tr>
<th>Material</th>
<th>Width (w)</th>
<th>Thickness (t)</th>
<th>Temperature Range (°C)</th>
<th>Derating Factor (α)</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Al-0.3%Cu</td>
<td>w = 7-8</td>
<td>t = 250</td>
<td>121-222</td>
<td>0.65</td>
<td>[d’Heurle 1972]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>t = 0.5-1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>w = 12.7</td>
<td>t = 1270</td>
<td>2-3</td>
<td>0.63-0.7</td>
<td>[Rodbell 1983]</td>
</tr>
<tr>
<td>Al-0.5%Cu</td>
<td>w = 1-4</td>
<td>t = 25000</td>
<td>1.2-4.5</td>
<td>0.5</td>
<td>[Sim 1979]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>t = 0.8-1.1</td>
<td>1.6-2</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>w = 2-5</td>
<td>t = 1000</td>
<td>1.8</td>
<td>0.73</td>
<td>[Levi 1985]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>t = 1</td>
<td>2.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>w = 2-5</td>
<td>t = 1000</td>
<td>2.9</td>
<td>0.84</td>
<td>[Levi 1985]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>t = 1</td>
<td>1.62</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>w = 7.5</td>
<td>t = 275</td>
<td>1.8</td>
<td>0.68</td>
<td>[Levi 1985]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>t = 0.5</td>
<td>1.82</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Al-2%Cu</td>
<td>w = 6</td>
<td>t = 380</td>
<td>1-2.3</td>
<td>0.7</td>
<td>[Ghate 1981]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>t = 0.8</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Al-4%Cu</td>
<td>w = 7-8</td>
<td>t = 250</td>
<td>1-4</td>
<td>0.83</td>
<td>[d’Heurle 1972]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>t = 0.5-1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>w = 15</td>
<td>t = 1250</td>
<td>-</td>
<td>0.76</td>
<td>[Learn 1975]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>t = 1</td>
<td>0.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Al-(5.1-5.4)%Cu</td>
<td>w = 10</td>
<td>t = 250-3380</td>
<td>1-3.1-5</td>
<td>0.48</td>
<td>[Kakar 1973]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>t = 1.5</td>
<td>1.3-1.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Al-1%Si-2%Cu</td>
<td>w = 6</td>
<td>t = 380</td>
<td>0.25</td>
<td>0.5</td>
<td>[Ghate 1981]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>t = 0.8</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Al-2%Si-4%Cu</td>
<td>t = 1.0</td>
<td></td>
<td>-</td>
<td>0.7</td>
<td>[Black 1982]</td>
</tr>
<tr>
<td>Al-Si-0.5%Cu</td>
<td>w = 3.3-7.7</td>
<td>t = 0.62-0.67</td>
<td>1.8-3.6</td>
<td>0.76</td>
<td>[Bukkett 1984]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.5-1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Material</td>
<td>w</td>
<td>l</td>
<td>t</td>
<td>D</td>
<td>Tmax</td>
</tr>
<tr>
<td>----------------</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>------</td>
</tr>
<tr>
<td>Al-Si-2%Cu</td>
<td>2.2-7.7</td>
<td>0.28-0.71</td>
<td>1.5-3.1</td>
<td>0.5-1</td>
<td>180-240°C</td>
</tr>
</tbody>
</table>
| Al-1%Si-0.35%
| (0.35-1.1)%Ti | w = 4 | l = 2000 | 0.5-0.7 | - | - | [Towne 1986] |

**Double-Layer Metallizations**

<table>
<thead>
<tr>
<th>Material</th>
<th>w</th>
<th>l</th>
<th>t</th>
<th>D</th>
<th>Tmax</th>
<th>C</th>
<th>( \Delta H )</th>
<th>Ref</th>
</tr>
</thead>
<tbody>
<tr>
<td>Al-1%Si/TiW</td>
<td>5.5</td>
<td>2550</td>
<td>0.6/0.22</td>
<td>0.5-2.2</td>
<td>100-200°C</td>
<td>0.49</td>
<td>2.5</td>
<td>[Onduresk 1988; Hoang 1988]</td>
</tr>
<tr>
<td>Al-0.3%Cu/TiW</td>
<td>12.7</td>
<td>1270</td>
<td>0.8/0.2</td>
<td>2.3</td>
<td>1.4</td>
<td>100-300°C</td>
<td>0.65</td>
<td>-</td>
</tr>
<tr>
<td>Al-4%Cu/TiW</td>
<td>12.7</td>
<td>1270</td>
<td>0.8/0.2</td>
<td>5</td>
<td>1.4</td>
<td>100-300°C</td>
<td>0.75</td>
<td>-</td>
</tr>
</tbody>
</table>

\[
\tau_f = \frac{\left(\frac{d}{MC}\right)\left(\frac{2r_d}{ZFD_o}\right)}{\beta(V_{pot} - V_\tau)\frac{1}{kT} e^{-\frac{\Delta H}{kT}} \text{erfc} \frac{1}{2\sigma} \ln\left(\frac{kTlnH \text{average}}{2\gamma v}\right)}
\]

\[r = \frac{2\gamma v}{kT \ln\left(\frac{p}{p_o}\right)}\]

\[\text{erfc}\left(\frac{1}{2\sigma}\right) = 1 - \frac{2}{\sqrt{\pi}} \int_0^{\frac{1}{2\sigma}} e^{-u^2} du\]

and where \( p \) is the density of the dendrite; \( M \) is the atomic weight; \( C \) is the ionic concentration; \( z \) is the ionic charge or the valence of the metal ions; \( D_o \) is the...
diffusivity coefficient; \( I \) is the distance between electrodes; \( r_d \) is the dendrite radius; \( r_{\text{average}} \) is the average pore radius; \( \gamma \) is the surface tension; \( \nu \) is the molar volume; \( p \) is the saturated vapor pressure above the meniscus; \( p_o \) is the saturated vapor pressure above the flat surface (\( H \) approaching \( p/p_o \) is the relative humidity at which condensation occurs; \( \sigma = \ln [r_{so}/r_{id}] \) sigma; \( \beta \) is the fraction of metal surface at the anode that is susceptible to metal oxidation and promotes metal migration; \( V_{\text{pot}} \) = potential difference applied; \( V_T \) = critical value of potential difference for dendritic growth; \( j_{\text{tip}} \) = current density at the whisker's tip; \( j_{\text{critical}} \) = critical current density at the whisker's tip; \( Q_{\text{tip}} \) = \( j_{\text{tip}} \)/ZF = mass transport rate; and \( Q_{\text{critical}} \) = critical mass transport rate.

5.4.5 Constraint Cavitation of Conductor Metallization

The lifetime due to SDDV becomes minimum at a certain temperature \( T_m \). The increase in lifetime when \( T \) increases or decreases from \( T_m \) (temperature where the minima occurs) results from a decrease in stress in the metallization for \( T \geq T_m \), or a decrease in diffusivity for \( T \leq T_m \), respectively. The allowable operation temperatures are calculated from

\[
T_{\text{operation}} < T_{\text{max}} \quad \text{for } T < T_{\text{passivation}}
\]

\[
T_{\text{operation}} > T_{\text{max}} \quad \text{for } T > T_{\text{passivation}}
\]

such that

\[
\tau > \text{mission life}
\]

where

\[
\tau = \frac{kGLWT e^{\frac{Q_s}{RT}}}{2AE_D(1-n)} \left[ \left( \frac{\sigma(0) - EW \tan \psi}{G} \right)^{1-n} - \left( \frac{\sigma(0)}{G} \right)^{1-n} \right]^{1-n}
\]

for \( n \neq 1 \)
\[
\tau = \frac{kGLWT}{2AED_0} e^{\frac{Q_d}{RT}} \left( \ln \frac{\sigma(0)}{G} - \ln \left( \frac{\sigma(0)}{G} - \frac{EW \tan \psi}{GL} \right) \right) 
\]

(5.36)

for \( n = 1 \)

where

\[
\sigma(t) = \sigma_o(T_d - T) \left( \frac{W_o}{W} \right)^m \left( \frac{H_o}{H} \right)^n
\]

(5.37)

and where

- \( k \) is Boltzmann's constant;
- \( G \) is the shear modulus of metallization;
- \( L/2 \) is the distance over which the stress due to void formation is relaxed;
- \( W \) is the width of the metallization;
- \( T \) is steady-state temperature;
- \( A \) is a factor independent of stress, temperature, and time;
- \( E \) is Young's modulus of metallization;
- \( D_o \) is the diffusivity constant;
- \( Q_d \) is the activation energy for diffusion;
- \( n \) is the stress dependency exponent;
- \( \sigma(0) \) = stress at \( t = 0 \);
- \( \sigma(t) \) = stress at time \( t \);
- \( m \) = \( p = 0.5 \);
- \( T_d \) = passivation deposition temperature;
- \( W_o \) = normalization width;
- \( H_o \) = normalization thickness;
- \( W \) = line width;
- \( H \) = line thickness; and
- \( \sigma_o \) = stress induced by temperature change of 1 degree centigrade, and
- \( \psi \) = half the angle of the void.

*Kato and Niwa Model [Kato et al. 1990, Niwa et al. 1990]*. This model accounts for plastic deformation and diffusional relaxation. While plastic deformation
occurs almost instantaneously after the passivation is laid on the metallization, diffusion is practically inoperative at low temperatures. (Low temperature has been defined as the temperature at which the time for diffusional relaxation is greater than $10^3$ seconds) [Kato et al. 1990, Niwa et al. 1990]. Kato et al. found that stresses did not become hydrostatic after relaxation due to plastic deformation. Time to failure due to stress-driven diffusive voiding is considered to be the time during which the area fraction, $A$, increases from $A_{\text{min}}$ to $A_{\text{max}}$ (Area fraction $= r_a^2/l^2$; $r_a$ is the void radius, and $2l$ is the void separation). The variation in time to failure versus aspect ratio and temperature shows that SDDV changes its temperature dependence from steady-state to inverse temperature dependence at a temperature threshold which is a function of passivation temperature (Figure 5.4). The steady-state temperature limits to avoid SDDV can be calculated from

\[ T_{\text{operation}} < T_{\text{max}} \quad \text{for} \quad T < T_{\text{passivation}} \]
\[ T_{\text{operation}} > T_{\text{max}} \quad \text{for} \quad T > T_{\text{passivation}} \]

such that

\[ T_f > \text{mission life} \]

where

\[ T_f = \frac{kT^3}{20D_p\omega\sigma\Omega} \]

\[ l = \frac{a_2}{r} \quad \text{for} \quad r \geq 1 \]
\[ l = a_2 \quad \text{for} \quad 0 \leq r \leq 1 \]

\[ \text{for} \quad \sigma > \frac{2\Gamma}{r_a} \]

\[ \text{no void growth} \]

where
\[ \sigma_{33}^R = -\frac{\mu}{(1-v)} \left( 2(1+v)\varepsilon^T + \frac{-2r-2(1-v)}{(1+r)} \varepsilon_{11} + \frac{-2-2r(1-v)}{(1+r)} \varepsilon_{22} \right) \] 

(5.56)

For \( r = 1 \):

\[ \varepsilon_{11}^P = \varepsilon_{22}^P = -\frac{1}{2} \varepsilon_{33}^P = -\frac{(1+v)\varepsilon^T + \frac{(1-v)\sigma_y}{\mu}}{(4v-5)} \] 

(5.57)

\[ \sigma_{11}^P = \sigma_{22}^P = -\frac{\mu}{(1-v)}((1+v)\varepsilon^T(1-2v)\varepsilon_{11}) \] 

(5.58)

\[ \sigma_{33}^P = \frac{-2\mu}{(1-v)}((1+v)\varepsilon^T+(-2+v)\varepsilon_{11}) \] 

(5.59)

For \( r = 0 \):

\[ \varepsilon_{11}^P = \varepsilon_{33}^P = -\frac{1}{2} \varepsilon_{22}^P = -\varepsilon^T - \frac{(1-v)\sigma_y}{2\mu(1+v)} \] 

(5.60)

\[ \sigma_{11}^P = \sigma_{33}^P = \sigma_y \]

\[ \sigma_{22}^P = 0 \] 

(5.61)

where

- \( T_f \) is the time to failure in seconds;
- \( k \) is Boltzmann's constant (8.617 x 10^-5 eV/k or 1.38 x 10^-23 Joule/Kelvin);
- \( \Gamma \) is the surface energy per unit area (J/m^2);
- \( T \) is the temperature in Kelvin;
- \( D_B \) is the grain-boundary diffusivity (meters^2/sec);
- \( D_L \) is the lattice diffusion coefficient (meters^2/sec);
is the interfacial diffusion coefficient (meter²/sec);

\( \lambda \)

is the length of grain along line axis (meter);

\( w \)

is the grain-boundary thickness (meter);

\( \Omega \)

is the atomic volume for conductor atoms (meter³);

\( \sigma_x \)

is the hydrostatic stress after diffusional relaxation;

\( \sigma_{33} \)

is the stress along metallization length;

\( \sigma_y \)

is the yield strength of metallization;

\( \sigma_i \)

for \( i = 1, 2, 3 \) are the stresses in the metallization line along its width, thickness, and length, respectively;

\( \varepsilon^T \)

is the thermal strain resulting from mismatch between passivation and metallization, and

\( \varepsilon_{123}^S \)

are the strains in 1, 2, 3 directions (direction 1 = along line width; direction 2 = along line thickness; direction 3 = along line length);

\( \mu \)

is the shear modulus of the metallization line (Pascal or N/m²);

\( r \)

is the aspect ratio of metallization line (\( r = (\text{metallization thickness})/(\text{metallization width}) \)); 2\( a_1 \) is the metallization width; 2\( a_2 \) is the metallization thickness;

\( v \)

is the Poisson's ratio of metallization line;

\( \Delta \alpha \)

is the difference in coefficients of thermal expansion between the metallization line and passivation (per deg Kelvin);

\( T \)

is the temperature (kelvin); and

\( T_p \)

is the deposition temperature of passivation.

5.5 DERATING FOR FAILURE MECHANISMS IN DEVICE OXIDE

5.5.1 Slow Trapping

Slow trapping, a phenomenon of MOS devices only, has a dominant dependence on the trap density of oxide. The trapping and detrapping kinetics of electrons cause threshold voltage shifts. The magnitude of threshold voltage shift is a function of oxide permittivity, oxide thickness, density, and location of oxide charge. Electrons, or holes trapped in oxide traps, are excited by the application of high steady-state temperature. The highest steady-state temperature for a maximum allowable flatband voltage shift can be calculated from

\[ T_{\text{operation}} < T_{\text{max}} \]  \hspace{1cm} (5.62)

such that
\[ \Delta V_{FB} < \Delta V_{FB(max)} \]  

(5.63)

where

\[ \Delta V_G^+ = \Delta V_{FB} = -\frac{x_m Q_{ot}}{\varepsilon_0 \varepsilon_{ar}} \text{ for positive gate voltage} \]

\[ \Delta V_G^- = \Delta V_{FB} = -\frac{t_{ar}}{\varepsilon_0 \varepsilon_{ar}} \left(1 - \frac{x_m}{t_{ar}}\right) Q_{ot} \text{ for negative gate voltage} \]

\[ Q_{ot}(t) = (Q_{ot-0}) e^{-\eta t^3} \]  

(5.65)

and

\[ \eta = \left( \frac{2\sqrt{3}(2\pi)^{3/2} m_n^* k^2}{h^3} \right) \alpha_n T^2 e^{-\frac{E_n}{kT}} \]  

(5.66)

where

- \( T \) is steady-state temperature (Kelvin);
- \( \Delta V_{FB} \) is the flatband voltage shift;
- \( \varepsilon_0 \) is the free-space permittivity \( (8.85 \times 10^{-14} \text{ F cm}^{-1}) \);
- \( \varepsilon_{ar} \) is the relative permittivity of oxide (dielectric constant \( \approx 3.9 \) (typical value));
- \( t_{ar} \) is the oxide thickness (cm);
- \( x' \) is the distance away from the metal-SiO\(_2\) interface;
- \( x_m \) is the centroid of charge contained in the oxide (between 0 and \( t_{ar} \));
- \( \eta \) is the thermal emission coefficient (sec\(^{-1}\));
- \( t \) is the time (seconds);
- \( E_n \) is the trap energy level;
- \( \alpha_n \) is the capture cross-section for electrons (cm\(^2\));
- \( Q_{ot(t=0)} \) is the density of oxide-trapped charge integrated over the thickness of the oxide, expressed per unit area of Si-SiO\(_2\) interface at time \( t=0 \);
- \( Q_{ot} \) is the density of the oxide-trapped charge over the thickness of the oxide, expressed per unit area of the Si-SiO\(_2\) interface (C cm\(^{-2}\)).
\[ m^* \] is the effective mass of the electrons in \( \text{SiO}_2 \);
\[ k \] is Boltzmann constant (8.617 x 10^{-3} \text{eV/K} \text{ or } 1.38 \times 10^{-23} \text{J/K}); and
\[ h \] is Planck's constant (6.62 \times 10^{-34} \text{J.s}).

The prime distribution moment, \( (X_m Q_m^\prime) \), increases or decreases with the increase or decrease in \( x_m \) or \( Q_m^\prime \). If the trap distribution in the oxide is uniform parallel to the interfaces, the C-V curve shifts without distortion.

### 5.5.2 Gate Oxide Breakdown

**Electrostatic Discharge.** Electrostatic discharge in MOS devices can produce gate-to-drain shorts, gate-to-source shorts, or gate-to-substrate shorts, depending on the nature of imperfection. The most likely sites for ESD damage in defect-free oxides are source or drain sites depending on the polarity of transient currents and the biasing of the device. Gate-to-substrate shorts are more prevalent in devices with pre-existing oxide defects, in the form of geometrical or dopant irregularities. For bipolar devices, typical ESD junction damage occurs in the device bulk under reverse-biased conditions resulting in degraded p-n junction characteristics. Failed p-n junctions appear as cracked glass across the junction on the surface of the chip. The device operating temperature affects the ESD damage threshold. The higher the operating temperature, the lower the damage threshold. The maximum allowable temperature for the device can be calculated from the Wunsch-Bell model [Wunsch and Bell 1968]

\[
T_{\text{operation}} < T_{\text{max}} \quad \text{for } V_{\text{ESD}} > V_{\text{protection}} \quad (5.67)
\]

where \( T_{\text{max}} \) is calculated from

\[
T_{\text{i}} = T_m - \frac{P/I}{A \sqrt{\pi k p C_p}} \quad (5.68)
\]

and where

- \( P/I \) is the power per unit junction area, calculated from the protection structure breakdown voltage;
- \( k \) is the thermal conductivity of the semiconductor;
- \( \rho \) is the density of the semiconductor;
- \( C_p \) is the specific heat of the semiconductor;
- \( T_m \) is the melting point of the semiconductor;
- \( T_{\text{i}} \) is the operating ambient temperature; and
is the ESD pulse time (seconds).
Figure 5.4 The variation in time to failure versus aspect ratio and temperature shows that SDDV changes its temperature dependence from steady state to inverse temperature dependence, at a temperature threshold which is a function of the passivation temperature. [Kato, 1990; Niwa, 1990]
The threshold voltage and power density, $P/A$, for the circuit can be calculated from the Speakman model [Speakman 1974]:

$$A = D_{base} \times L_{emitter}$$  \hspace{1cm} (5.69)

$$\tau = (R_b + R_d + R_o)C_b$$  \hspace{1cm} (5.70)

$$I_p = \frac{V_b - V_d}{R_b + R_d + R_c}$$  \hspace{1cm} (5.71)

$$P_{av} = \frac{1}{5\tau} \int V_d I_p e^{-\frac{t}{\tau}} dt + \frac{1}{5\tau} \int R_d I_p^2 e^{-\frac{2t}{5\tau}} dt$$

$$= \frac{V_d I_p}{5}(1 - e^{-2}) + \frac{R_d I_p^2}{10}(1 - e^{-10})$$  \hspace{1cm} (5.72)

$$= \frac{V_d I_p}{5} + \frac{R_d I_p^2}{10}$$

$$R_d = R_{sheet-base} \times \frac{t_{b-e}}{L_{emitter}}$$  \hspace{1cm} (5.73)

where

- $D_{base}$ is the depth of the base region;
- $L_{emitter}$ is the length of the emitter region;
- $T_{b-e}$ is the base-emitter separation;
- $I_p$ is the peak current of the discharge waveform;
- $V_d$ is the device voltage;
- $V_b$ is the discharge voltage on the human body;
- $C_b$ is the body capacitance;
- $R_{base-sheet}$ is the base-sheet resistance;
- $R_i$ are resistance components; where $i = b,c,d$ for body, contact, and device resistances, respectively;
- $t$ is the time (seconds);
\[ \tau \]
is the time constant of the RC circuit; and
\[ V_{\text{protection}} \]
is the voltage capacity of the device protection circuit (Table 5.6).

**Table 5.6 Comparison of Typical ESD Protection Thresholds**

<table>
<thead>
<tr>
<th>Protective Circuit Device</th>
<th>Device Technology</th>
<th>Protection Voltage Threshold (Volt)</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>double implant field isolation device in well</td>
<td>CMOS</td>
<td>8000 volts</td>
<td>[Nischizawa 1975]</td>
</tr>
<tr>
<td>thick oxide-diffused resistor-field plate</td>
<td>DRAM</td>
<td>5000 volts</td>
<td>[Duvvury 1983]</td>
</tr>
</tbody>
</table>
|                                                    | junction depth : 0.4 \( \mu m \)  
breakdown voltage : 20 V  
field oxide : 1 \( \mu m \) |                      |                 |
|                                                    | EPROM             | 6000 volts                          | [Duvvury 1983]  |
|                                                    | junction depth : 0.8 \( \mu m \)  
breakdown voltage : 26 V  
field oxide : 1.4 \( \mu m \) |                      |                 |
| diode-diffused resistor-field plate                | EPROM             | 4500 volts                          | [Duvvury 1983]  |
|                                                    | junction depth : 0.8 \( \mu m \)  
breakdown voltage : 26 V  
field oxide : 1.4 \( \mu m \) |                      |                 |
| thick oxide polysilicon resistor-field plate       | EPROM             | 3000 volts                          | [Duvvury 1983]  |
|                                                    | junction depth : 0.8 \( \mu m \)  
breakdown voltage : 26 V  
field oxide : 1.4 \( \mu m \) |                      |                 |
| diffused resistor-gated diodes and spark gap       | CMOS/SOS type A   | 800 volts                           | [Palumbo and Dugan 1986] |
|                                                    | CMOS/SOS type B   |                                    |                 |
| diode-resistor circuit                             | CMOS/SOS type C   | 1800 to 2000 volts                  | [Palumbo and Dugan 1986] |
| thick oxide                                         | CMOS, NMOS        | > 6000 volts                        | [Rountree 1988] |
| diode-resistor-diode                               | CMOS gate array   | 4000 volts                          | [Hull and Jackson 1988] |

**Time-Dependent Dielectric Breakdown.** Time dependent dielectric breakdown (TDDB) is the formation of low-resistance dielectric paths through localized defects in the MOS oxide. TDDB has a very weak dependence on temperature and a dominant dependence on the electric field across the oxide. The field acceleration itself is a function of steady-state temperature. The field acceleration is inversely dependent on steady-state temperature, and has been found to reduce
from 6 decades/MV/cm at 25°C to 2 decades/MV/cm at 150°C. The maximum allowable temperature and electric field across the oxide can be calculated from:

- Thermodynamic Model [McPherson 1985]

**Fowler-Nordheim Tunneling-based Models for TDDB.** The maximum allowable temperature and electric field can be evaluated by calculating predicted life versus steady-state temperature, worst-case manufacturing defect magnitude ($X_{def}$), and electric field (Figures 5.5 a and b). Due to the non-linear dependence of life under TDDB on temperature, defect magnitude, or electric field, derating the stress below a particular value may not result in a noticeable benefit in terms of increased life, because the time to failure is much greater than the wear-out life of the device. Furthermore, because of the multiplicity of dependencies of TDDB, reducing the temperature to a pre-specified value for a device technology, as is done now (Table 5.1), may not result in the desired reliability. The dependence of time to failure under TDDB is given by

$$T_{operation} < T_{max} \quad weak\ acceleration$$

$$V_G < V_{Eox(max)} \quad strong\ acceleration$$

such that

$$t_{BD} > \text{mission\ life} \quad (5.75)$$

where

$$t_{BD}(T) = \tau_0 e^{\frac{G X_{def}}{V_{Eox}} \left(1 - \frac{\delta}{1 - \frac{1}{T - 300}}\right) - \frac{E_b}{k \left(\frac{1}{T} - \frac{1}{300}\right)}}$$

(5.76)

$$\delta = 0.0167 \text{ eV for } 25^\circ C < T < 125^\circ C$$

(5.77)

$$\delta = 0.024 \text{ eV for } T > 150^\circ C$$

$$E_b = 0.28 \text{ eV}$$

where
\[ \delta = \frac{k d G(T)}{G d (1/T)} \]  \hspace{1cm} (5.78)

and where

- \( T \) is the steady-state temperature (Kelvin);
- \( G \) is the slope of the \( \ln(\tan) \) versus \( 1/E_{ex} \) plot (\( \approx 350 \times 10^6 \) volt/cm);
- \( X_{ox} \) is the oxide thickness (cm);
- \( X_{eff} \) is the effective oxide thickness at the weakest spot in the oxide (cm);
- \( V_G \) is the gate voltage across the oxide (Volt);
- \( V_{ax} \) is the voltage across the oxide (Volt);
- \( \tau_0 \) is the room-temperature value of the pre-exponential, \( \tau(T) (\approx 1 \times 10^{-11} \) sec);\)
- \( E_b \) is the activation energy of the pre-exponential (eV); and
- \( k \) is Boltzmann's constant (8.617 \times 10^{-5} \text{ eV/K}).

**Thermodynamic Models [McPherson 1985]**. The thermodynamic model is based on the assumption that when the dielectric breaks down, it undergoes an irreversible phase transition transforming the material from an insulating phase to a conducting phase. The driving force for this transformation is the difference between the free energies of the conducting phase and the insulating phase. The maximum temperature and electric field for desired mission life can thus be calculated as follows:

\[ T_{operation} < T_{max} \hspace{1cm} \text{weak acceleration} \]  \hspace{1cm} (5.79)

\[ V_G < V_{ax(max.)} \hspace{1cm} \text{strong acceleration} \]

such that

\[ t_{BD} > \text{mission life} \]  \hspace{1cm} (5.80)

where

\[ TF(\%) = Ae \left( \frac{\Delta R_0}{\Delta T_0} \right) \epsilon [\gamma(T)T] \]  \hspace{1cm} (5.81)

The field acceleration parameter, \( \gamma \), is the steady-state temperature-dependent parameter given by
\[ \gamma = B + \frac{C}{T} \]  \hspace{1cm} (5.82)

where

- \( T \) is the steady-state temperature;
- \( K_B \) is the Boltzmann constant (\( 8.617 \times 10^{-5} \text{eV/K or } 1.38 \times 10^{-23} \text{J/K} \));
- \( \Delta H^* \) is the change in enthalpy required to activate the poly filament growth at breakdown;
- \( B \) and \( C \) are constants, \( S = E_B - E_s \);
- \( E_B \) is the breakdown strength of the dielectric; and
- \( E_s \) is the stressing in the dielectric.

**Empirical Models [Anolick 1981, Crook 1979, Berman 1981].** The maximum temperature or the electric field across the oxide can be calculated from

\[
\begin{align*}
T_{\text{operation}} &< T_{\text{max}} & \text{weak acceleration} \\
V_G &< V_{\text{ax(max.)}} & \text{strong acceleration}
\end{align*}
\]  \hspace{1cm} (5.83)

such that

\[ t_{\text{BD}} > \text{mission life} \]  \hspace{1cm} (5.84)

where

\[ t_{f(F)} = \Lambda e^{\frac{\Delta H}{kT} e^{\gamma(V_{B(F)}-V_A)}} \]  \hspace{1cm} (5.85)

where

- \( t_{f(F)} \) is the time to failure for \( F \) percent of the population;
- \( \gamma \) is the voltage form factor (determined by life testing);
- \( \Delta H \) is the activation energy;
- \( k \) is Boltzmann constant;
- \( T \) is steady-state temperature;
- \( V_B(F) \) is the breakdown voltage for \( F \) percent of population; \( V_A \) is the applied voltage; and \( \Lambda \) is a constant. (Some typical values are given in Table 5.7)
Table 5.7 Empirical Models for TDDB

<table>
<thead>
<tr>
<th>Reference</th>
<th>Oxide Thickness</th>
<th>Experimental Conditions</th>
<th>Observations</th>
<th>Model Predictions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anolick and Nelson 1979</td>
<td>700 Å</td>
<td>$E_x = 1.3 \text{ MV/cm}$ $E_x - E_s = 7$ MV/cm</td>
<td>$(\Delta H)_\text{obs} = 2.1 \text{ eV}$ $\gamma = B + \frac{C}{T}$</td>
<td>$(\Delta H)_\text{obs} = 1.8 \text{ eV}$ $\gamma = B + \frac{C}{T}$</td>
</tr>
<tr>
<td>Crook 1979</td>
<td>1100 Å</td>
<td>$E_x = 3.5 \text{ MV/cm}$ $E_x - E_s = 3$ MV/cm</td>
<td>$(\Delta H)_\text{obs} = 0.3 \text{ eV}$ $\gamma = 7 @ 25^\circ \text{C}$</td>
<td>$(\Delta H)_\text{obs} = 0.34 \text{ eV}$ $\gamma = 6 @ 25^\circ \text{C}$</td>
</tr>
<tr>
<td>Berman 1981</td>
<td>≥400 Å</td>
<td>linear ramp</td>
<td>$(\Delta H)_\text{obs} = 0.29 (E_x - E_s)$ $\gamma = -5.4 + \frac{0.2}{T}$</td>
<td>$(\Delta H)_\text{obs} = 0.29 (E_x - E_s)$ $\gamma = -5.4 + \frac{0.2}{T}$</td>
</tr>
<tr>
<td>Hokari 1982</td>
<td>100 Å</td>
<td>$E_x = 5 - 7 \text{ MV/cm}$ $E_x - E_s = 5$ MV/cm</td>
<td>$(\Delta H)_\text{obs} = 1.0 \text{ eV @ 6 MV/cm}$ $\gamma = 1.7 @ 250^\circ \text{C}$</td>
<td>$(\Delta H)_\text{obs} = 1.0 \text{ eV @ 6 MV/cm}$ $\gamma = 1.5 @ 250^\circ \text{C}$</td>
</tr>
<tr>
<td>McPherson 1985</td>
<td>100 Å</td>
<td>$E_x = 6 - 8 \text{ MV/cm}$ $E_x - E_s = 3 - 5$ MV/cm</td>
<td>$(\Delta H)_\text{obs} = 0.3 - 1.0 \text{ eV}$ $\gamma = B + \frac{C}{T}$</td>
<td>$(\Delta H)_\text{obs} = 0.3 \text{ eV}$ $\gamma = B + \frac{C}{T}$</td>
</tr>
</tbody>
</table>

5.6 DERATING FOR FAILURE MECHANISMS IN THE DEVICE

5.6.1 Ionic Contamination

Ionic contamination occurs predominantly in MOS devices and results in reversible degradation, in the form of threshold voltage shift and gain reduction due to the presence of mobile ions within the oxide or at the device-oxide interface. P-
channel devices are less sensitive to ionic contamination than n-channel devices. The mobility of ions is steady-state temperature dependent. The threshold voltage shift increases with an increase in the steady-state temperature. However, a high-
Figure 5.5 Due to the non-linear dependence of life under TDDDB on temperature, worst case manufacturing defect magnitude, or electric field, derating the stress below a particular value may not result in noticeable benefit in terms of increased life because the time to failure is much beyond wear-out life of the device.

Figure 5.5a Derating curve for TDDDB versus steady state temperature and effective oxide thickness. The curves identify the various combinations of temperature $T$, and effective oxide thickness $x_{\text{eff}}$, which will result in desired life. [Lee, 1988; Moazzami, 1989; Moazzami, 1990]
Figure 5.5  Due to the non-linear dependence of life under TDBB on temperature, worst case manufacturing defect magnitude, or electric field, derating the stress below a particular value may not result in noticeable benefit in terms of increased life because the time to failure is much beyond wear-out life of the device.

Figure 5.5b  Derating curve for TDBB versus steady state temperature and electric field. The curves identify the various combinations of temperature $T$, and electric field $E_{ox}$, which will result in desired life. [Lee, 1988; Moazzami, 1989; Moazzami, 1990]
temperature storage bake in the neighborhood of 150°C to 250°C restores device characteristics. The maximum allowable steady-state temperature for the device is a function of the concentration of ionic contaminants in the oxide:

\[ T_{\text{operation}} < T_{\text{max}} \]  

such that

\[ \Delta V_T < \Delta V_{T(\text{max})} \]  

where

\[ \Delta V_T = \frac{1}{E^2 t^2} e^{-\frac{\Delta H}{kT}} \]  

and where

- \( \Delta V_T \) is the change in threshold voltage (volt);
- \( E \) is the electric field across the oxide (volt/cm);
- \( t \) is the time under bias (seconds);
- \( \Delta H \) is the activation energy (eV); and
- \( K \) is the Boltzmann constant (8.617 x 10\(^{-5}\) eV/K or 1.38 x 10\(^{-23}\) J/K);

N-channel devices are often covered with phosphosilicate glass films to stabilize the threshold voltage against changes resulting from ionic contamination.

### 5.6.2 Surface-Charge Spreading

This failure mechanism, occurring mostly in MOS devices, involves the lateral spreading of charge from the biased metal conductors along the oxide layer or through moisture on the device surface. The failure mechanism is manufacturing-defect-activated, due to the presence of ionic contaminants on the die surface. Derating for this failure mechanism involves controlling the amount of contaminant on the die.
5.7 DERATING GUIDELINES

5.7.1 Hot Electrons

The mechanism of hot electrons in MOS devices is inversely dependent on steady-state temperature. The minimum allowable steady-state temperature to avoid hot electrons can be calculated using the Lucky Electron Model [Ning 1977, Garrigues 1981]

\[ T_{\text{operation}} < T_{\text{min}} \]  \hspace{1cm} (5.95)

such that

\[ P_e = P_0 e^{-d / (\lambda \sqrt{E_a / kT})} \]  \hspace{1cm} (5.96)

where

\[ d = \sqrt{\frac{2 \varepsilon_{sc} \varepsilon_0}{q N_A} \left( \sqrt{\Psi_S} - \sqrt{\Psi_S - \frac{\Phi_{BS}}{q}} \right)} \]  \hspace{1cm} (5.97)

and where

\( P_e \) is the emission probability of an electron;
\( P_0 \) is a constant (\( = 2.9 \) @ 300°K; \( 4.3 \) @ 77°K);
\( d \) is the minimum path length of an electron required to attain the critical energy, \( \Phi_{BS} \);
\( T \) is the steady-state temperature (Kelvin);
\( E_a \) is the activation for mean free-path length (0.063 eV);
\( \lambda_0 \) is a constant (\( = 108 \) Å);
\( \lambda \) is the electron mean free path between lattice interactions;
\( K \) is Boltzmann’s constant (8.617 \( \times \) 10\(^{-5} \) eV/K or 1.38 \( \times \) 10\(^{-23} \) J/K);
\( \varepsilon_{sc} \) is the dielectric constant of the semiconductor (F cm\(^{-1} \));
\( \varepsilon_0 \) is the free-space permittivity (8.85 \( \times \) 10\(^{-14} \) F cm\(^{-1} \));
\( N_A \) is the concentration of acceptor doping atoms (cm\(^{-3} \));
\( q \) is the magnitude of electronic charge (1.6 \( \times \) 10\(^{-19} \) Coulomb);
\( \Psi_S \) is the surface potential (volt); and
\( \Phi_{BS} \) is the Si-SiO\(_2\) barrier height for electrons, taking into account the
DERATING FOR FAILURE MECHANISMS IN THE DEVICE OXIDE INTERFACE

lowering due to the Schottky effect.

5.8 DERATING FOR FAILURE MECHANISMS IN THE DIE AND DIE/SUBSTRATE ATTACH

5.8.1 Die Fracture

Westgaard-Bolger-Paris Equation-based Formulation. The maximum allowable temperature cycle should be less than the calculated temperature cycle magnitude:

\[ \Delta T_{\text{operation}} < \Delta T_{\text{max}} \]  \hspace{1cm} (5.98)

such that

\[ N_f > \text{mission life} \]  \hspace{1cm} (5.99)

where

\[ \Delta \sigma_{\text{app}} = 10^{-6}k|\alpha_s - \alpha_d|\Delta T \sqrt{\frac{E_s E_d L}{X}} \]  \hspace{1cm} (5.100)

\[ N_f = \frac{2}{(n-2)A(\Delta \sigma_{\text{app}})^n \pi^2} \left( \frac{1}{a_i^2} - \frac{1}{a_f^2} \right) \]  \hspace{1cm} (5.101)

and where

- \( A \) is the die-material coefficient;
- \( n \) is the die-material exponent;
- \( \alpha_s \) is the CTE of the substrate (/°C);
- \( \alpha_d \) is the CTE of the die (/°C);
- \( \Delta \sigma_{\text{app}} \) is the mode I applied-stress amplitude (psi);
- \( a_i \) is the initial crack size (inch); and
- \( a_f \) is the final crack length at failure, which may be taken to be equal to the critical crack size (inch).

Suhir-Paris Equation-based Formulation. The maximum allowable temperature cycle should be less than the calculated temperature cycle magnitude:
\[ \Delta T_{\text{operation}} < \Delta T_{\text{max}} \]  
\[ (5.102) \]

such that

\[ K_{v-\text{die}} < K_{C-\text{die}} \]  
\[ (5.103) \]

\[ K_{h-\text{die}} < K_{C-\text{die}} \]  
\[ (5.104) \]

and

\[ \sigma < \sigma_{\text{Rupture-die}} \]  \( \text{die middle} \)
\[ (5.105) \]

\[ \sigma_1 < \sigma_{\text{Rupture-die}} \]  \( \text{die edge} \)

and

\[ \min(N_{f(v-\text{die})}, N_{f(h-\text{die})}) > N_{\text{required}} \]  
\[ (5.106) \]

where

\[ N_{f(v-\text{die})} = \frac{1}{(1 - \frac{n_d}{2}) A_d^\frac{n_d}{2} \sigma^{n_d} \pi^2} \left[ a_{f-\text{die}}^{1 - \frac{n_d}{2}} - a_{l-\text{die}}^{1 - \frac{n_d}{2}} \right] \]  
\[ (5.107) \]

\[ N_{f(h-\text{die})} = \frac{1}{(1 - \frac{n_d}{2}) A_d^\frac{n_d}{2} P^{n_d} \pi^2} \left[ a_{f-\text{die}}^{1 - \frac{n_d}{2}} - a_{l-\text{die}}^{1 - \frac{n_d}{2}} \right] \]  
\[ (5.108) \]

and

\[ K_{v-\text{die}} = \sigma \sqrt{\pi} \frac{a_{v-\text{die}}}{a_{v-\text{die}}} F \]  
\[ (5.109) \]

\[ K_{h-\text{die}} = P \sqrt{\pi} \frac{a_{h-\text{die}}}{a_{h-\text{die}}} F \]  
\[ (5.110) \]
\[ F = \frac{2.85 \left[ 0.953 -2.369 \left( \frac{a}{t_d} \right) + 2.74 \tan \left( \frac{a}{t_d} \right) \right]}{3 + \frac{a^2}{c^2}} \]  

(5.111)

\[
\sigma = \frac{1}{t_d} + 3 (t_d + t_s) \frac{E_d t_d}{12 (1-v_d^2) D} \frac{E_d t_d}{12 (1-v_s^2) E_s t_s} \frac{1 - \frac{1}{\cosh(\Delta L)}}{4 D} \left[ \frac{t_d}{3 G_d} + \frac{2 t_a}{3 G_a} + \frac{t_s}{3 G_s} \right]^{-1} \frac{1}{2 D} (\alpha_s - \alpha_d) \Delta T
\]  

(5.112)

\[
p = -\left[ \frac{t_s t_d^3 E_d}{12 (1-v_d^2)} - \frac{t_d t_s^3 E_s}{12 (1-v_s^2)} \right] \left[ \frac{t_d}{3 G_d} + \frac{2 t_a}{3 G_a} + \frac{t_s}{3 G_s} \right]^{-1} \frac{1}{2 D} (\alpha_s - \alpha_d) \Delta T
\]  

(5.113)

\[
\tau = \left[ \frac{1-v_d}{E_s t_s} + \frac{1-v_d}{E_d t_d} \frac{(t_d + t_s)^2}{4 D} \right]^{-1} \left[ \frac{t_d}{3 G_d} + \frac{2 t_a}{3 G_a} + \frac{t_s}{3 G_s} \right]^{-1} \frac{1}{2} \tanh(\Delta L) (\alpha_s - \alpha_d) \Delta T
\]  

(5.114)

\[
A = \frac{\sqrt{\frac{1-v_d}{E_s t_s} + \frac{1-v_d}{E_d t_d} \frac{(t_d + t_s)^2}{4 D} \frac{t_d}{3 G_d} + \frac{2 t_a}{3 G_a} + \frac{t_s}{3 G_s}}}{4 D}
\]  

(5.115)
\[ D = \frac{E_d t_d^3}{12 (1 - v_d^2)} + \frac{E_a t_a^3}{12 (1 - v_a^2)} + \frac{E_s t_s^3}{12 (1 - v_s^2)} \]  
(5.116)

\[ \sigma_1 = \frac{P}{2} + \sqrt{\frac{P^2}{4} + \tau^2} \]  
(5.117)

and where

- \(E, G, \nu\) are the modulus of elasticity, the shear modulus, Poisson’s ratio, and the coefficients of thermal expansion (CTE), respectively;
- \(t\) is the thickness. The subscripts \(d, a,\) and \(s\) denote the die, attachment, and substrate, respectively. \(\Delta T\) is the temperature change;
- \(L\) is the half-diagonal length of the die;
- \(\sigma\) is the tensile stress in the middle of the die;
- \(P\) is the peeling stress at the die-attachment interface;
- \(\tau\) is the shear stress at the die-attachment interface;
- \(\sigma_1\) is the principal stress at the die edge;
- \(\sigma_{\text{rupture}}\) is the modulus of rupture of the die;
- \(K_v\) and \(K_h\) are the stress-intensity factors for vertical cracks on the top surface of the die and for horizontal cracks at the edge of the die;
- \(K_c\) is the fracture toughness of the die;
- \(a_{v-die}\) is the depth of the vertical crack on the top surface of the die; and
- \(a_{h-die}\) is the length of the horizontal surface crack at the edge of the die.

**5.8.2 Die Thermal Breakdown**

The calculated junction temperature should be less than the allowable junction temperature of the device:

\[ T_{\text{junction}} < T_{\text{junction, allowable}} \]  
(5.118)

where \(T_{\text{junction, allowable}}\) is the allowable junction temperature. \(T_{\text{junction}}\) is the effective junction temperature calculated from
FAILURE MECHANISMS IN DIE SUBSTRATE ATTACH

\[ T_{\text{junction}} = T_{\text{max}} + Q \sum_{i=1}^{3} \theta_i \]  
(5.119)

where \( T_{\text{max}} \) is the mean value of the maximum ambient temperature, and \( Q \) is the power dissipated by the devices. The effective thermal resistance of each layer, \( \theta_i \) (die, attachment, and substrate), is determined from

\[ \theta_i = \frac{1}{2k_i(L_i-W_i)} \left[ \ln \frac{L_i(W_i+2t_i)}{W_i(L_i+2t_i)} \right] \]  
(5.120)

where \( L_i, W_i, t_i \) are the lengths, widths, and thicknesses of each layer, and \( k_i \) is the thermal conductivity of each layer.

5.8.3 Die and Substrate Adhesion Fatigue

**Suhir-Paris-(Coffin-Manson) Equation-based Formulation.**

*For brittle-attach materials:* The maximum allowable temperature cycle should be less than the maximum calculated value, such that:

\[ \Delta T_{\text{operation}} < \Delta T_{\text{max}} \]  
(5.121)

such that,

\[ K_{h-\text{attach}} < K_{c-\text{attach}} \]  
(5.122)

where

\[ K_{h-\text{attach}} = p \sqrt{\pi} a_{h-\text{attach}} F \]  
(5.123)

\[ N_{f(h-\text{attach})} = \frac{1}{(1-n_{a})A_{a}^{n_{a}} p^{n_{a}} \pi^{2}} \left[ a_{f-\text{attach}} - a_{l-\text{attach}} \right]^{1-n_{a}} \]  
(5.124)

and where

- \( a_{l-\text{attach}} \) is the initial attach-crack size;
- \( a_{f-\text{attach}} \) is the final attach-crack depth;
- \( A_{a} \) and \( n_{a} \) are fatigue properties of the brittle attachment material;
\( p \) is the cyclic peeling stress determined in Equation 5.113;
\( F \) is the geometric correction factor given by Equation 5.111;
\( K_{\text{c-attach}} \) is the fracture toughness of the attachment; and
\( K_{\text{h-attach}} \) is the stress intensity factor for horizontal crack in the edge of the brittle attachment material.

For Ductile-attach Materials: The maximum allowable temperature cycle should be less than the maximum calculated value, such that:

\[
\Delta T_{\text{operation}} < \Delta T_{\text{max}} \quad (5.125)
\]

such that

\[
C_f' \sigma_1 < \sigma_{U-\text{attach}} \quad (5.126)
\]
\[
C_f'' \tau_M < \tau_{U-\text{attach}} \quad (5.127)
\]

where

\[
C_f' \sigma_1 = \frac{3p}{2} + \sqrt{\left(\frac{3p}{2}\right)^2 + (4\tau)^2} 
\]

\[
C_f'' \tau_M = \sqrt{(3p)^2 + 3(4\tau)^2} 
\]

\[
N_f-\text{attach}(t) = C_{\text{attach}} (C_f' \sigma_1)^{m_{\text{attach}}} 
\]

\[
N_f-\text{attach}(s) = C_{\text{attach}} (C_f'' \tau_M)^{m'_{\text{attach}}} 
\]

and where

\( p \) and \( \tau \) are the peeling and shear stresses given by Equations 5.113 and 5.114;
\( \sigma_{U-\text{attach}} \) and \( \tau_{U-\text{attach}} \) are the tensile strength and the shear strength of the attachment material;
\( C \) and \( m \) are the tensile fatigue constants;
\( C_{\text{attach}} \) and \( m'_{\text{attach}} \) are the shear fatigue constants of the attachment material;
\( C_f' \sigma_1 \) is the local cyclic principal stress; and
FAILURE MECHANISMS IN DIE SUBSTRATE ATTACH

\( C_f^w \tau M \) is the local cyclic von Mises’ stress.

5.9 DERATING FOR FAILURE MECHANISMS IN FIRST-LEVEL INTERCONNECTS

5.9.1 Wirebonded Interconnections

*Wire Fatigue.* The mechanism of wire fatigue involves the flexure of the wire about the reduced wire cross-section at the heel during temperature cycling. The mechanism has a dominant dependence on the magnitude of the temperature cycle and is independent of steady-state temperature. The maximum temperature cycle (\( \Delta T_{\text{operation}} \)) that the wire can be subjected to should be less than the calculated temperature cycle magnitude (\( \Delta T_{\text{max}} \)) [Pecht 1989],

\[
\Delta T_{\text{operation}} < \Delta T_{\text{max}}
\]

such that

\[
N_f > \text{mission life}
\]

(5.130)

where \( N_f \) can be calculated from either of the following models: *Pecht et al. Model* [Pecht et al. 1989]

\[
N_f = A(\varepsilon_f)^n
\]

(5.131)

and

\[
\varepsilon_f = \frac{r}{\rho_o} \left[ \cos^{-1}\left(\cos \lambda_0 (1 - (\alpha_w - \alpha_s)\Delta T)\right) - 1 \right]
\]

(5.132)

and where

\( r \) is the wire radius;
\( \lambda_0 \) is the angle of the wire with the substrate;
\( \alpha_w \) is the coefficient of thermal expansion of the wire;
\( \alpha_s \) is the coefficient of thermal expansion of the substrate;
\( \Delta T \) is the temperature cycle encountered by the structure;
\( \rho_o \) is the initial radius of curvature of the wire;
\( N_f \) is the mission life requirement; and
\( A \) and \( n \) are material constants.
Hu et al. Model [Hu, Pecht, Dasgupta 1991]

\[ N = C_w \sigma^{-m_w} \]  \hspace{1cm} (5.133)

where

\[ \sigma = 6 \frac{E_w r}{D(D^2 - 1)} \left( \frac{1}{2 \alpha_s} + \frac{\alpha_s - \alpha_w}{1 - \frac{D}{L}} \right) \Delta T \]  \hspace{1cm} (5.134)

and where

- \( 2L \) is the wire length;
- \( 2D \) is the wire span;
- \( E_w \) is the elastic modulus of the wire;
- \( \alpha_w \) and \( \alpha_s \) are the coefficients of thermal expansion of the wire and the substrate materials, respectively;
- \( \Delta T \) is the temperature change encountered during operation;
- \( \sigma \) is the bending stress (already calculated); and
- \( C_w \) and \( m_w \) are fatigue properties determined by tensile fatigue tests of the wire material.

**Wirebond Fatigue.**

**Bond Pad Shear Fatigue:** The maximum temperature cycle (\( \Delta T_{\text{operation}} \)) that the wire-wirebond assembly can be subjected to without causing shear fatigue failure of the bondpad should be less than the calculated temperature cycle magnitude (\( \Delta T_{\text{max}} \)) [Hu, Pecht, Dasgupta 1991]:

\[ \Delta T_{\text{operation}} < \Delta T_{\text{max}} \]  \hspace{1cm} (5.135)
DERATING FOR FAILURE MECHANISMS
IN FIRST-LEVEL INTERCONNECTS

\[ \Delta T_{\text{max}} = \left( \frac{N_z}{C'} \right)^{-\frac{1}{m'}} \frac{G_p \left( \frac{\alpha_w - \alpha_s - \left( \alpha_s - \alpha_p \right)}{1 + \frac{E_s A_s}{E_p A_p \left(1 - v_s \right)}} \right)}{b_p Z} \]

(5.136)

and

\[ Z^2 = \frac{G_p}{b_p} \left( \frac{r}{E_p A_p} + \frac{(1 - v_s) W_p}{E_s A_s} \right) \]

(5.137)

and where

- \( N_R \) is the mission profile requirement for the number of cycles to failure;
- \( C' \) and \( m' \) are the shear fatigue properties for the bond-pad materials;
- \( G_p \) is the shear modulus of the bond-pad material;
- \( b_p \) is the bond-pad thickness;
- \( Z \) is given by Equation 5.137;
- \( \Delta T \) is the temperature cycle magnitude;
- \( \alpha_s, \alpha_p, \alpha_w \) are the coefficients of thermal expansion for the substrate, pad, and wire, respectively;
- \( E_p \) and \( E_s \) are the modulus of elasticity of the pad and the substrate, respectively;
- \( v_s \) is Poisson’s ratio for the substrate materials;
- \( A_p \) is the cross-sectional area of the pad;
- \( A_s \) is the effective cross-sectional area of the substrate, equal to \( b_p (W_p + W_p)/2 \); and
- \( W_p \) is the width of the bond pad.

Wire Shear Fatigue: The maximum temperature cycle (\( \Delta T_{\text{operation}} \)) that the wire-wirebond assembly can be subjected to without causing shear fatigue of the wire should be less than the calculated temperature cycle magnitude (\( z\Delta T_{\text{max}} \)) [Hu, Pecht, Dasgupta 1991]:
$\Delta T_{\text{operation}} < \Delta T_{\text{max}}$  \hspace{1cm} (5.138)

where

$$\Delta T_{\text{max}} = \frac{\left( N_R \right)^{\frac{1}{m_w^2}}}{\left( \frac{r^2}{4Z^2A_w^2} \left( \frac{\cosh(z,x)}{\cosh(z,l_w)} - 1 \right)^2 + \frac{\sinh^2(Z,x)}{\cosh^2(Z,l_w)} \right)^{\frac{1}{2}} Q}$$  \hspace{1cm} (5.139)

$$Z^2 = \frac{G_p}{b_p} \left( \frac{r}{E_s A_w} + \frac{(1 - \nu_s)W_p}{E_s A_s} \right)$$ \hspace{1cm} (5.140)

$$Q = \left( \frac{G_p \Delta T}{b_p Z} \right) \left( \frac{(\alpha_w - \alpha_s) - \frac{(\alpha_s - \alpha_p)}{1 + \frac{E_s A_s}{E_p A_p}}} {\left( \frac{E_s A_s}{1 - \nu_s} \right)} \right)$$ \hspace{1cm} (5.141)

and where:

- $N_R$ is the mission profile requirement for the number of cycles to failure;
- $C_w$ and $m_w$ are the Coffin-Manson coefficients for wire material;
- $r$ is the wire radius;
- $Z$ is given by Equation 5.140;
- $A_w$ is the effective cross-sectional area of the wire at the bond = (0.6 wire diameter) x (1.5 wire diameter);
- $x$ is the position along the length of the bond ($x=0$ is the center of the bond);
- $l_w$ = half the length of the bond (total bonded length = 2 $l_w$);
- $Q$ is given by Equation 5.141;
- $G_p$ is the shear modulus of the bond pad;
- $b_p$ is the bond-pad thickness;
- $W_p$ is the width of the bond pad;
- $E_w$ is the elastic modulus of the wire;
DERATING FOR FAILURE MECHANISMS IN FIRST-LEVEL INTERCONNECTS

\( E_p \) is the elastic modulus of the bond pad;
\( E_s \) is the elastic modulus of chip;
\( A_s \) is the effective cross-sectional area of the substrate, equal to 
\( b_s(W_p + W_s)/2 \);
\( A_p \) is the cross-sectional area of the pad; and
\( \alpha_s, \alpha_p, \alpha_w \) are the coefficients of thermal expansion for the substrate, pad, and wire, respectively.

**Chip Shear Fatigue Leading to Cratering:** The maximum temperature cycle (\( \Delta T_{\text{operation}} \)) that the wire-wirebond assembly can be subjected to without causing shear fatigue of the chip should be less than the calculated temperature cycle magnitude (\( \Delta T_{\text{max}} \)) [Hu, Pecht, Dasgupta 1991]

\[
\Delta T_{\text{operation}} < \Delta T_{\text{max}}
\]  
(5.142)

where

\[
\Delta T_{\text{max}} = \frac{N_R}{C_s'} \left( \frac{1}{N_R} \right)^{1/N_R} \left( \frac{E_p Q}{2ZA_s} \left( 1 - \frac{Cosh(Z \alpha_w)}{Cosh(Z \alpha_w)} \right) + \frac{(\alpha_s - \alpha_p)}{(E_s A_s)} + \frac{1}{(E_p A_p)} \right)^2 + Q^2 \left( \frac{Sinh^2(Z \alpha_w)}{Cosh^2(Z \alpha_w)} \right)^{1/2}
\]

\[
Z^2 = \frac{G_p}{b_p} \left( \frac{r}{E_s A_w} + \frac{(1 - \nu_s) W_p}{E_s A_s} \right)
\]  
(5.144)

\[
Q = \left( \frac{G_p \Delta T}{b_p Z} \right) \left( \frac{1}{1 + \frac{E_s A_s}{E_p A_p}} \right) \left( \frac{\alpha_s - \alpha_p}{\alpha_s - \alpha_p} \right)
\]

(5.145)

and where
is the mission profile requirement for the number of cycles to failure;

and $m_w$ are the Coffin-Manson coefficients for the wire material;

is the wire radius;

is given by Equation 5.140;

is the effective cross-sectional area of the wire at the bond = (0.6 wire diameter) x (1.5 wire diameter);

is the position along the length of the bond ($x=0$ is the center of the bond);

is half the length of the bond (total bonded length = 2 $l_w$);

is given by Equation 5.141;

is the shear modulus of the bond pad;

is the bond pad thickness;

is the width of the bond pad;

is the elastic modulus of the wire;

is the elastic modulus of the bond pad;

is the elastic modulus of the chip;

is the effective cross-sectional area of the substrate, equal to $b_s(W_p+W_v)/2$;

is the cross-sectional area of the pad; and

are the coefficients of thermal expansion for the substrate, pad, and wire, respectively.

**Intermetallic Formation.** The maximum operating temperature for a bimetallic wirebond system can be obtained by using the parabolic relationship [Kidson 1961]:

$$t = \frac{x^2}{k}$$

(5.146)

The maximum allowable operating temperature is less than the calculated operating temperature:

$$T_{\text{operation}} < T_{\text{max}}$$

(5.147)

where


\[ T_{\text{max}} = - \frac{Q}{R \ln \left( \frac{x^2}{D_0} \right)} \]

(5.148)

\[ k = D_0 e^{-\frac{Q}{RT}} \]

and where

\( x \) is the critical intermetallic layer thickness (from Tables 5.8 and 5.9);

\( t \) is time to failure due to intermetallic formation;

\( k \) = rate constant, depending on the interdiffusion coefficients of the bonded materials (from Table 5.9).

**Table 5.8 The Dominant Intermetallic Compounds for Common Bimetal Combinations**

<table>
<thead>
<tr>
<th>Bimetal Combination</th>
<th>Intermetallic Compounds Formed</th>
<th>Dominant Intermetallic Compounds</th>
</tr>
</thead>
<tbody>
<tr>
<td>gold-aluminum (Au-Al)</td>
<td>( \text{Au}_2\text{Al}_3, \text{Au}_2\text{Al}, \text{AuAl}, \text{AuAl} ) [Philosky 1970, Philosky 1971]</td>
<td>( \text{Au}_2\text{Al}_3, \text{Au}_2\text{Al} )</td>
</tr>
<tr>
<td>gold-silver (Au-Ag)</td>
<td>significant silver diffusion into the wire bulk along the wire length. Rapid silver surface diffusion resulted in depletion at bond periphery [James 1977].</td>
<td></td>
</tr>
<tr>
<td>copper-gold (Cu-Au)</td>
<td>( \text{Cu}_2\text{Au}, \text{CuAu}, \text{CuAu}_3 ) [Hall 1975, Tu and Berry 1972]</td>
<td>( \text{Cu}_2\text{Au} )</td>
</tr>
</tbody>
</table>
The critical intermetallic layer thickness is calculated based on the intermetallic compounds that form the fastest for the bimetallic combination. The dominant compounds for some of the common material combinations are given in Table 5.8. The composition of the intermetallic compound when critical layer thickness is reached, and rate constants for various material combinations, are given in Table 5.9. The time to failure is calculated as the time for complete consumption of the bond pad in the intermetallic reaction. For example, in gold-aluminum bonds, the compound \( \text{Au}_3\text{Al}_2 \) forms the fastest and is the dominant product during intermetallic formation. Thus, for gold bond pads, the critical layer thickness is defined as the time to reach an intermetallic compound thickness of \( t_{\text{Au}} + (2/5) t_{\text{Au}} \), where \( t_{\text{Au}} \) is the thickness of the gold bond pad.

The time to failure versus temperature is then plotted, based on Equation 5.146. The maximum operating temperature of the device for a given mission life is calculated from the graph (Figures 5.6 a,b, and c). A similar procedure can be followed for each of the intermetallic combinations. Based on Philosky's observations of the layer thicknesses of intermetallic compounds at various temperatures (Table 5.10), a graph of intermetallic layer thickness versus square-root time to failure can be drawn (Figure 5.7) [1971]. A horizontal line on the graph at a thickness equal to the wire thickness at the bond pad gives the time-temperature product that will result in failure after time equal to the abscis. The allowable time-temperature for a desired mission life can thus be computed. The slope of each of the lines in Figure 5.7 gives rate constants at various
Figure 5.6 Time for heel of the wire to transform to intermetallic versus temperature. The temperature at which intermetallic formation has a dominant dependence on steady state temperature is a function of the bond geometry [Philosky, 1970, 1972]

Figure 5.6a For gold wire bonded to aluminum metallization
Figure 5.6 Time for heel to transform to intermetallic versus temperature. The temperature at which intermetallic formation has a dominant dependence on steady state temperature is a function of the bond geometry [Philosky, 1970, 1971]

Figure 5.6b For aluminum wire bonded to gold metallization

![Graph showing the relationship between time and temperature for aluminum wire bonded to gold metallization. The graph includes lines for different thicknesses (y) of the bond: 0.125 mil, 0.250 mil, 0.50 mil, and 1 mil. The temperature range is from 150 to 450 degrees Celsius, and the time range is from 0 to 600 minutes.]
Figure 5.6c Time to failure versus temperature and bond pad thickness. The time to failure at any temperature may be much greater than the mission life depending on the bond pad thickness. [Philosky, 1970, 1971]
Figure 5.7 Various time-temperature products which will result in gold-aluminum intermetallic growth [Philosky]. A horizontal line on the graph at a thickness equal to the wire thickness at the bond pad gives the time-temperature product which will result in failure after time equal to the abcissa. The slope of each line gives the rate constants at various temperatures.
temperatures. Intermetallic formation in aluminum-gold systems can be eliminated either by using a fritless gold composition that impedes intermetallic formation or by using diffusion barrier disks attached between the aluminum (Al) wire and thick film [Palmer and Gaynard 1978].

Table 5.10 Measured Thickness of Gold-Aluminum (Au-Al) Intermetallic Compounds. [Philosky, 1971]

<table>
<thead>
<tr>
<th>Temperature (°C)</th>
<th>Time (Seconds)</th>
<th>( \text{AuAl}_2 ) cm x 10³</th>
<th>( \text{AuAl} ) cm x 10³</th>
<th>( \text{Au}_2\text{Al}_3 ) cm x 10³</th>
<th>( \text{Au}_2\text{Al}_3 ) cm x 10³</th>
<th>( \text{Au}_4\text{Al}_3 ) cm x 10³</th>
<th>Total cm x 10³</th>
</tr>
</thead>
<tbody>
<tr>
<td>200</td>
<td>8700</td>
<td>0</td>
<td>0</td>
<td>0.68</td>
<td>0</td>
<td>0.69</td>
<td></td>
</tr>
<tr>
<td>200</td>
<td>27300</td>
<td>0</td>
<td>0</td>
<td>1.22</td>
<td>0</td>
<td>1.25</td>
<td></td>
</tr>
<tr>
<td>200</td>
<td>57600</td>
<td>0</td>
<td>0</td>
<td>1.57</td>
<td>0</td>
<td>1.63</td>
<td></td>
</tr>
<tr>
<td>200</td>
<td>349200</td>
<td>0.21</td>
<td>0</td>
<td>0.53</td>
<td>2.68</td>
<td>0.21</td>
<td>3.63</td>
</tr>
<tr>
<td>250</td>
<td>7200</td>
<td>0</td>
<td>0</td>
<td>0.72</td>
<td>2.38</td>
<td>0.17</td>
<td>3.38</td>
</tr>
<tr>
<td>250</td>
<td>75600</td>
<td>0.24</td>
<td>0</td>
<td>0.8</td>
<td>2.47</td>
<td>0.24</td>
<td>3.75</td>
</tr>
<tr>
<td>300</td>
<td>15900</td>
<td>0.11</td>
<td>0</td>
<td>0.71</td>
<td>2.38</td>
<td>0.17</td>
<td>3.38</td>
</tr>
<tr>
<td>300</td>
<td>58200</td>
<td></td>
<td></td>
<td>1.29</td>
<td>4.75</td>
<td>0.23</td>
<td>6.5</td>
</tr>
<tr>
<td>350</td>
<td>22200</td>
<td>0.27</td>
<td>0</td>
<td>0.1</td>
<td>1.65</td>
<td>0.11</td>
<td>2.13</td>
</tr>
<tr>
<td>350</td>
<td>24000</td>
<td>0.42</td>
<td>0</td>
<td>0.52</td>
<td>6.31</td>
<td>0.23</td>
<td>7.38</td>
</tr>
<tr>
<td>400</td>
<td>300</td>
<td>0.15</td>
<td>0</td>
<td>0.09</td>
<td>0.89</td>
<td>0</td>
<td>1.14</td>
</tr>
<tr>
<td>400</td>
<td>1500</td>
<td>0.32</td>
<td>0</td>
<td>0.29</td>
<td>2.32</td>
<td>0</td>
<td>2.94</td>
</tr>
<tr>
<td>400</td>
<td>6000</td>
<td>0.36</td>
<td>0</td>
<td>0.93</td>
<td>4.57</td>
<td>0</td>
<td>5.88</td>
</tr>
<tr>
<td>400</td>
<td>14400</td>
<td>0.33</td>
<td>0.12</td>
<td>1.48</td>
<td>6.46</td>
<td>0.11</td>
<td>8.5</td>
</tr>
<tr>
<td>400</td>
<td>36000</td>
<td>0.26</td>
<td>0.27</td>
<td>2.36</td>
<td>7.03</td>
<td>0.21</td>
<td>10.13</td>
</tr>
<tr>
<td>460</td>
<td>1200</td>
<td>0.26</td>
<td>0.27</td>
<td>1.24</td>
<td>2.36</td>
<td>0</td>
<td>4.13</td>
</tr>
<tr>
<td>460</td>
<td>6000</td>
<td>0.21</td>
<td>0.35</td>
<td>1.4</td>
<td>7.14</td>
<td>0.28</td>
<td>9.38</td>
</tr>
</tbody>
</table>

5.9.2 Tape Automated Bonds

*Thermally Induced Solder-joint Fatigue*. The temperature cycle magnitude to which the device is subjected should be less than the maximum allowable:
\[ \Delta T < \Delta T_{\text{max}} \]  

such that

\[ N_f > \text{mission life} \]  

where

\[ N_f = \frac{1}{2} \left( \frac{1}{2e_f'} \frac{K(L_D \Delta \alpha \Delta T_s)^2}{200A_{\text{eff}} h_{\text{solder}}} \right)^{\frac{1}{c}} \]  

and

\[ c = -0.442 - 6 \times 10^{-4} T_{sj} + 1.74 \times 10^{-2} \ln \left[ 1 + \left( \frac{360}{t_D} \right) \right] \]  

and

\[ T_{sj} = \frac{1}{4}(T_c + 2T_o + T_s) \]

and where

- \( T_s \) and \( T_c \) are the steady-state operating temperatures of the substrate and component, respectively;
- \( T_o \) is the temperature during off half-cycle;
- \( t_D \) is the half-cycle dwell time in minutes;
- \( K \) is the diagonal flexural stiffness of the unconstrained lead, determined by finite element analysis [Barker 1991] or strain-energy methods;
- \( \Delta \alpha \) is the difference in the coefficients of thermal expansion of the die and the substrate;
- \( \Delta T_{\text{max}} \) is the maximum allowable equivalent temperature range;
- \( e_f' \) is the fatigue ductility coefficient;
- \( A_{\text{eff}} \) is the effective solder-joint area, which is two-thirds of the vertical projection of the solder-wetted lead area;
- \( h_{\text{solder}} \) is the height of the solder joint; and
- \( t_D \) is the half-cycle dwell time in minutes.
5.9.3 Flip-Chip Bonds

The temperature cycle magnitude to which the device is subjected should be less than the maximum allowable:

\[ \Delta T < \Delta T_{\text{max}} \]  

\[ (5.156) \]

such that

\[ N_f > \text{Mission Life} \]  

\[ (5.157) \]

where

\[ N_f = \frac{1}{2} \left[ \frac{\Delta \gamma}{2e_f} \right]^\frac{1}{2} \]  

\[ (5.158) \]

\[ c = \alpha + \beta \ T_{SJ} + \gamma \ \ln(1+\frac{360}{t_D}) \]  

\[ (5.159) \]

\[ \Delta \gamma = k_g \delta \]

\[ = k_g \left[ \alpha_c(T_{jc} - T_{amb}) - \alpha_d(T_{jc} - \theta_{sof}(\frac{P_{chip}}{n}) - T_{amb}) \right] \]  

\[ (5.160) \]

\[ k_g = \frac{1}{h} \int_0^{\frac{1}{n}} \frac{dy}{\left(\pi r_{\text{critical}}^2\right)^{\frac{1}{n}}} \int_0^{\frac{1}{n}} \frac{1}{\pi \left(R^2 - y^2\right)} \]  

\[ (5.161) \]

\[ T_{SJ} = \frac{1}{4} \left[ 2T_{jc} - \theta_{sof}(\frac{P_{chip}}{n}) + 2T_{amb} \right] \]  

\[ (5.162) \]
and where:

- $\varepsilon_f$ is the fatigue ductility coefficient of the solder material;
- $\alpha$, $\beta$, and $\gamma$ are factors to be determined empirically;
- $T_{sv}$ is the mean cyclic solder-joint temperature (°C);
- $t_D$ is half-cycle dwell time in minutes;
- $k_s$ is the geometric parameter;
- $\delta$ is the total displacement;
- $d$ is the distance of farthest solder joint from neutral axis;
- $\alpha_c$ and $\alpha_s$ are the coefficients of thermal expansion of the chip and the substrate, respectively;
- $T_{jc}$ is the junction temperature;
- $T_{amb}$ is the ambient temperature;
- $\theta_{sod}$ is the thermal resistance of the solder;
- $P_{chip}$ is the chip power rating;
- $n$ is the number of solder bumps on the chip;
- $h$ is the height of the solder;
- $r_{critical}$ is the critical interface radius;
- $K$ is the thermal conductivity of the solder; and
- $A_{eff}$ is the effective area of cross section of the solder;

### 5.10 DERATING FOR FAILURE MECHANISMS IN THE PACKAGE CASE

#### 5.10.1 Cracking in Plastic Packages

The maximum temperature during thermal shock to which the device can be subjected can be calculated from:

\[
T < T_{\text{max}} \tag{5.164}
\]

such that

\[
\sigma_{\text{max}} < \sigma_{\text{crit}(T_{\text{max}})} \tag{5.165}
\]

where
\[ \sigma_{\text{max}} = 6K\left(\frac{a}{t}\right)^2 P \]  
\[ P = H_1 P_{\text{sat}}(T_{\text{max}}) \]

and where

- \( H_1 \) is the relative humidity of the saturation ambient prior to temperature shock;
- \( T_{\text{max}} \) is the peak temperature of the thermal shock;
- \( K \) is a dimensionless stress concentration factor that depends on the aspect ratio of the die pad (\( K = 0.05 \) for a square pad);
- \( a \) is the length of the short side of the die pad;
- \( t \) is the thickness of the molding compound under the pad;
- \( P \) is the vapor pressure in the cavity; and
- \( \sigma_{\text{crit}} \) is the value of critical stress.

### 5.10.2 Reversion or Depolymerization of Polymeric Bonds

Reversion is a steady-state temperature-dependent phenomenon that actuates above the reversion temperature for the encapsulant or molding compound. Derating for reversion involves limiting the maximum temperature for device operation to lower than the depolymerization temperature:

\[ T < T_{\text{depolymerization}} \]

### 5.10.3 Whisker and Dendritic Growth

Cases coated with tin are prone to whiskers and dendrites. Whiskering can, however, be reduced by increasing the coating thickness during manufacture. Typically, hot-dipping is used to obtain a stress-free tin coating. If electroplating is used, tin is reflowed to remove residual stress, and 2 to 3% lead is added to the tin to retard whisker growth. This is a contamination-actuated mechanism and has no temperature dependence.
5.10.4 Modular Case Fatigue Failure

The temperature cycle magnitude to which the device is subjected should be less than the maximum allowable:

\[ \Delta T < \Delta T_{\text{max}} \]  

such that

\[ N_f > \text{mission life} \]  

where

\[ \tau_{\text{max}} = 2\sigma (2N)^b \]  

and

\[ \tau_{\text{max}} = \left[ \frac{1-v_h}{E_h t_h} + \frac{1-v_w}{E_w t_w} + \frac{(t_w+t_h)^2}{4D} \right]^{1/2} \left[ \frac{t_w}{3G_a} + \frac{2t_a}{3G_a} + \frac{t_h}{3G_h} \right]^{1/2} \tanh(\alpha L) \Delta \alpha \Delta T \]  

\[ A = \sqrt{\frac{1-v_h}{E_h t_h} + \frac{1-v_w}{E_w t_w} + \frac{(t_w+t_h)^2}{4D}} \]  

\[ D = \frac{E_h t_h^3}{12(1-v_h^2)} + \frac{E_a t_a^3}{12(1-v_a^2)} + \frac{E_w t_w^3}{12(1-v_w^2)} \]  

and

- \( v \) is Poisson's ratio;
- \( E \) is the modulus of elasticity;
- \( t \) is the thickness;
- \( G \) is the shear modulus;
- \( L \) is the length of the joint;
\( \Delta \sigma \) is the difference in the CTEs of the case wall and header materials; and Subscripts \( h, w \) and \( a \) denote the header, wall and attach material, respectively

5.11 DERATING FOR FAILURE MECHANISMS IN LID SEALS

5.11.1 Thermal Fatigue of Lid Seal

*Lid and Case Material are the Same.* Most lid-seal failures are overstress failures, fatigue is not a concern in most applications. The temperature cycle magnitude to which the device is subjected should be less than the maximum allowable:

\[
\Delta T < \Delta T_{\text{max}}
\]  

(5.175)

such that

\[
\sigma_1 < \sigma_{\text{ult(seal)}}
\]  

(5.176)

The maximum principal stress in the seal is calculated using Mohr's circle, where the shear stress at any seal cross-section at a distance, \( x \), from the center cross-section is

\[
\tau = \frac{\Delta \alpha \Delta T}{\sqrt{\kappa \lambda}} \frac{\sinh \left( \frac{\lambda}{\kappa} x \right)}{\sinh \left( \frac{\lambda}{\kappa} \right)}
\]  

(5.177)

and the normal stress at any cross-section at a distance \( x \) from the mid-section of the seal is

\[
\sigma(x) = \frac{\Delta \alpha \Delta T}{\lambda'_{\text{seal}}} \left[ 1 - \frac{\cosh \left( \frac{\lambda}{\kappa} x \right)}{\cosh \left( \frac{\lambda}{\kappa} \right)} \right]
\]  

(5.178)

where

\( \Delta \alpha \) is the difference in the coefficients of thermal expansion of the lid and case material and the seal material;

\( \Delta T \) is the difference between the sealant melting point (stress-free...
is the in-plane compliance of the joint \( \lambda = [(1-\nu_{\text{case}})/E_{\text{case}}(h_{\text{case}}+t_{\text{lid}})] + [(1-\nu_{\text{case}})/E_{\text{case}}] \); 
\( \kappa \) is the interfacial compliance, given as the sum of the individual compliances for the seal and the case or lid material; that is, \( \kappa = \kappa_{\text{seal}} + \kappa_{\text{case}} \), while \( \kappa_{\text{seal}} = 2(1+\nu_{\text{seal}})t_{\text{seal}}/3E_{\text{seal}} \) and \( \kappa_{\text{case}} = 2(1+\nu_{\text{case}})(h_{\text{case}}+t_{\text{lid}})/3E_{\text{case}} \). \( l_{\text{seal}} \) is half the seal length (or width).

**Lid and Case Material are Different.** The temperature cycle magnitude to which the device is subjected should be less than the maximum allowable:

\[
\Delta T < \Delta T_{\text{max}}
\]  
(5.179)

such that

\[
\sigma_1 < \sigma_{\text{ult(seal)}}
\]  
(5.180)

where the principal stress in the seal is

\[
\sigma_1 = \frac{P}{2} \pm \sqrt{\frac{P^2}{4} + \tau^2}
\]  
(5.181)

\[
\tau = \left[\frac{1-\nu_{c}}{E_{c}t_{c}} + \frac{1-\nu_{l}}{E_{l}t_{l}} + \frac{(t_{c}+t_{l})^2}{4D}\right]^{-1/2} (G')^{-1/2} \tanh(Al_{\text{seal}}) \Delta \alpha \Delta T
\]  
(5.182)

\[
p = -\left[\frac{t_{l}E_{l}}{12(1-\nu_{l}^2)} - \frac{t_{c}E_{c}}{12(1-\nu_{c}^2)}\right] \frac{1}{2G'D} \Delta \alpha \Delta T
\]  
(5.183)

and where

\( p \) is the peeling stress;

\( \tau \) is the shear stress in the seal;

\( \nu, E, l, \text{ and } t \) denote Poisson's ratio, Young's modulus of elasticity, the length, and the thickness, respectively; subscripts \( l, c \) and \( s \) denote that the symbols refer respectively to the lid, case, and seal;
\[ \Delta \alpha \] is computed as \((\alpha_c - \alpha_i)\), where \(\alpha\) is the coefficient of thermal expansion;

\[ \Delta T \] is the temperature excursion with respect to the stress-free state of the seal, or the sealing temperature; \(A\), \(D\), and \(G'\) are defined as

\[
A = \sqrt{\frac{1-v_c^2+1-v_i^2+(t_c+t_i)^2}{E_c^f E_i^f 4D G'}} \tag{5.184}
\]

\[
D = \frac{E_c^f t_c^3}{12(1-v_c^2)} + \frac{E_i^f t_i^3}{12(1-v_i^2)} + \frac{E_s^f t_s^3}{12(1-v_s^2)} \tag{5.185}
\]

\[
G' = \frac{t_c}{3G_c} + \frac{2t_s}{3G_s} + \frac{t_i}{3G_i} \tag{5.186}
\]

where \(G\) is the shear modulus of elasticity.
CONCLUSION

The temperature dependencies of the failure mechanisms existing at various package elements have been investigated in terms of steady state temperature, temperature cycle, temperature gradient, and time dependent temperature change (Tables 2.8, 2.9). It has been found that temperature dependencies for the same mechanism are not the same at all operating temperatures. The mechanisms have been broadly grouped into the steady state temperature ranges of -55°C to 150°C, 150°C to 400°C, and \( T > 400°C \) (Tables 2.10, 2.11). The investigation demonstrates that there is no steady state temperature dependence for any of the failure mechanisms in the equipment operating range of -55°C to 125°C, but the steady state temperature dependence increases for temperatures above 150°C as more mechanisms assume a dominant steady state temperature dependence. An overview of the effects of temperature on microcircuits can be categorized as follows:

**Steady State Temperature Effects**
- Changes in semiconductor characteristics due to
  - steady state temperature dependence of semiconductor characteristics such as resistivity which can lead to failures in form of thermal runaway (e.g. electrical overstress).
- steady state temperature dependence of contaminants in the semiconductor that affect electronic functions or compatibility with the circuit parameters (e.g., ionic contamination).
- manufacturing defects which will cause device failures after prolonged operation under voltage and steady state temperature (e.g., TDDB).

- Increase in susceptibility to failure due to an environmental stress, at higher temperatures (T dependence)
  (e.g. electrostatic discharge).
- Increase in susceptibility to failure due to a failure mechanism at lowered temperature, i.e the failure mechanisms are inversely dependent on steady state temperature (e.g., stress driven diffusive voiding, and hot electrons).
- Increase in susceptibility to failure under steady state temperature stress above a threshold value of temperature (e.g., hillock formation, metallization migration, contact spiking, encapsulant reversion, electromigration).

**Temperature Cycle Magnitude Effects**

- Temperature cycles result in cyclic fatigue failures due to mechanical stresses and dimensional changes caused by thermal mismatches between mating surfaces (e.g., wirebond shear and flexure fatigue, die fracture, die adhesive fatigue).

**Temperature Gradient Effects**

- Sites of maximum temperature gradient provide most probable sites for failure due to mass transport mechanisms. (e.g., electromigration damage at high current occurs at sites of maximum temperature gradient)

**Time Dependent Temperature Change Effects**

- Time dependent temperature changes produce very large stress transients resulting in material failure. (e.g., encapsulant cracking)
- Temperature change in time may activate or de-activate a failure mechanism. (e.g., duty cycle serves as an ON/OFF switch for the corrosion process by evaporation of the electrolyte at higher temperatures)

The use of a simple Arrhenius expression to model microelectronic device reliability at all steady state temperatures is not correct because the temperature dependencies of the device are different at different steady state temperatures. It also improper to assess the thermal acceleration of the device by stress tests at elevated temperatures and extrapolate the results to lower temperatures, because the failure mechanisms (Table 2.10) are not uniformly active for all steady state temperatures.

The generalized association of lowered temperature with higher reliability may not be true for all device technologies. Two possible exceptions may exist. First,
the mechanism may not be dependent on steady state temperature, and second, the mechanism may have an inverse dependence on steady state temperature. In either case the steady state temperature will not be a driver for microelectronic reliability.

In cases where temperature can be directly related to parameter drift, premature aging, and even catastrophic failure, the temperature influence on reliability can be inarguable. Although examinations of case histories indicate that temperature's relationship to reliability may be more a case of exposing incompatibilities between operational requirements and design or manufacturing processes. In other words, the product as designed is not suitable for operation in the desired environment without changes. Failures mechanism such as reversion or depolymerization, contact spiking, metallization migration occur at high temperatures encountered during fabrication or assembly are well above normal operating temperatures. In such cases certain questions must be addressed before taking action:

- Will lowering the maximum operating temperature by itself avoid or decelerate the experienced failures? If so, how much should it be lowered and how is the conclusion reached?
- Will lowering the magnitude of the temperature cycle or change avoid the experienced failures? If so, how much should it be lowered and how is the conclusion reached?

Thus, for technologies exhibiting an obvious dependence of reliability on temperature, lowering the maximum operating temperature may be one of the measures of for enhanced reliability, however a generalization of the concept for universal application is incorrect.

Burn-in has been routinely used as a screen, with higher reliability attributed to burned-in parts without questioning the objectives, necessity or benefits. It has been used as a customer imposed requirement to supposedly ensure higher reliability. The emphasis has been on empirical analysis, without any analysis as to the real or "root" cause of failure in terms of improper manufacturing parameters or design inconsistencies. The stresses applied in the burn-in process have not been tailored to the dominant stress dependencies of the failure mechanisms in the device technology.

The goal of screening, and burn-in in particular, is to remove defective devices which would fail abnormally early in the field, and to implement corrective actions to avoid the occurrence of defects. Devices with mature design and manufacturing processes should experience few if any such failures. In many cases current material manufacturing and processing have reached such a state of maturity that few failures occur within the specified device lifetime.
The device derating criteria have been critically evaluated. A physics of failure based device derating criteria has been proposed to aid the designer enhance device life by means other than lowering steady state. The Derating approach allows the user to examine the critical stresses in the multichip module package and reduce their value to obtain a desired mission life. The acceptable values of stress are represented as graphs of life versus magnitudes of dominant operating stresses. Such graphs are called derating curves. Derating to achieve desired life involves examining the dominant failure mechanisms. By varying the operating point on the derating curve, different stress combinations which would result in desired life can be identified. The operating point can be the functionally most acceptable combination.

To evaluate the sensitivity of device life towards the temperature and non-temperature stresses, the user can plot the device life versus percentage change from a nominal stress value (Figures 6.1, 6.2). This menu allows the user to identify stress derating thresholds below which lowering stress magnitudes will produce no additional benefit in terms of added life. These thresholds can be identified as values of stresses for which the projected time to failure is well beyond the specified mission life of the module.
Figure 6.1 Derating curve for mission life versus operational parameters for dominant failure mechanisms, at a constant operating temperature of 125°C. This derating plot can be used to derate non-temperature operational stresses for cost-effective designs which are reliable at high temperature.
Figure 6.2 Derating curve for mission life versus steady state temperature for dominant failure mechanisms, at a specified values of non-temperature operational stresses. This derating plot can be used to derate non-temperature operational stresses for cost-effective designs which are reliable at high temperature.
REFERENCES


Alexander, D. R. An Electrical Overstress Failure Model for Bipolar Semiconductor Components.


REFERENCES


Campbell, Michael, Monitored Burn-In Improves VLSI IC Reliability. Computer Design, 24 (1985), 143-144.


REFERENCES


DELCO Data, (as of 11-2-88) ETDL LABCOM-from 636 nos IC's analyzed 513 nos Bipolar &


Ghate, P. B. and Blair, J. C. Electromigration Testing of Ti:W/Al and Ti:W/Al-Cu Film


Hall, P. M., Panousis, N. T., and Menzel, P. R. Strength of Gold Plated Copper Leads on Thin Film Circuits under Accelerated Aging. *IEEE Transactions on Parts, Hybrids, and Packaging*,


Hillen, M. W. thesis, University of Groningen, the Netherlands, (DAI-C 81-70, 066), 1981; cited


Howes, M. J. and Morgan, D.V. *Reliability and Degradation: Semiconductor Device and Circuits*,...


Integrated Device Technology, High Performance CMOS Databook Supplement, Santa Clara, CA, S3-8, (1989);


REFERENCES


Moazzami, R., Lee, J., Chen, I.C., and Hu, C. Projecting the Minimum Acceptable Oxide


Motorola Semiconductor Products Sector, Reliability and Quality Handbook, Phoenix, AZ, 1990


NASA, Requirements to Preclude the Growth of Tin Whiskers, information received from Jack
REFERENCES


Naval Air Systems Command, Department of the Navy, Application and Derating Requirements for Electronic Components, AS-4613, 1976.


Newsome, J. L. and Oswald, R. G. Metallurgical Aspects of Aluminum Wire Bonds to Gold Metallization, 14th IRPS, (1976), 63-74.


Nishimura, A., Tatemichi, A., Muira, H., and Sakamoto, T. Life Estimation for IC Plastic


Peeters, J. Burn-In for Increased Automotive Reliability, Test, 13(3) (1991), 24-25.


Schwarzenberger, A. P., Ross, C. A., Evetts, J. E., and Greer, A. L. Electromigration in


Thornton, C. G. and Simmons, C. D. A New High Current Mode of Transistor Operation. *IRE


Tu, K. N. and Berry, B. S. Journal of Applied Physics, 43 (1972), 3283.


Wu, C. J. and McNutt, M. J. Effects of Substrate Thermal Characteristics on the Electromigration


Zhu, Lily H., "High Temperature Modeling and Thermal Characteristics of GaAs MESFETs on Diamond Heat Sinks", *M.S. Thesis directed by Electrical Engineering Department and Mechanical Engineering Department, University of Maryland at College Park*, 1992