CHAPTER 1
Introduction

1.1 Background
Reliability is an important requirement for almost all users of integrated circuits (ICs). Scaling for enhanced performance and cost reduction has pushed existing complementary metal-oxide-semiconductor (CMOS) materials much closer to their intrinsic physical and reliability limits. Reliability challenges have become more and more concerned with continuous scaling for 65-nm technology node and beyond also owing to the introduction of new materials and device architectures with unknown reliability behaviors. For example, the application of stress-memorization technique (SMT), high-$k$/metal gate in front-end-of-line (FEOL), and ultralow-$k$ porous dielectrics in back-end-of-line (BEOL) has a serious impact on the reliability performance.

A newly developed semiconductor technology node cannot be released to designers without going through a rigorous process and product-reliability qualification. The reliability qualification of a developing node usually starts in parallel with every phase of the process and device development and ends only when the full node of the process development is completed. In other words, any process and/or tool change during development should be confirmed to meet reliability requirements. After the process is released for mass production, the reliability still should be monitored periodically to ensure that there are no process/tool variations. The reliability also refines the design rule; for instance, the minimum channel length $L_{\text{min}}$ for a 2.5-V input/output (IO) device overdriven to 3.3 V is defined as the $L_{\text{min}}$ that meets the hot-carrier reliability requirement.

This book covers the areas of process reliability and its qualification of a deep submicron semiconductor process node from theory to the very practical industry practice for each area of reliability concern. It starts with initial spec definition and moves through test-structure design to analysis of the silicon data taken from the test structures and final qualification of the process. Many practical examples, including test-structure designs to qualify the front-end-of-line devices or the back-end-of-line of the interconnects, have been included to strengthen the reader’s understanding of reliability theory and how the theory is
used to resolve issues in the standard practice. In addition, this book includes some fundamental process flow and device physics when necessary to help in explaining process and device reliability concerns.

### 1.2 Process Reliability Items

In general, process integration engineers (PIEs) divide a semiconductor process into two parts, the FEOL and BEOL. The FEOL includes the process steps up to polysilicon, where as BEOL covers the process after the polysilicon. FEOL defines the metal-oxide-semiconductor (MOS) device and other commonly used active and passive devices, whereas BEOL defines interconnects.

The five major significant device-level failure items in IC technology FEOL are hot-carrier injection (HCI), gate-oxide time-dependent dielectric breakdown (TDDB), negative-bias temperature instability (NBTI), plasma-induced damage (PID) and electrostatic discharge (ESD). The three major interconnect failure modes considered most commonly in BEOL are electromigration (EM), stress migration (SM), and TDDB of intermetal dielectrics. Note that PID is drawing more and more attention because both FEOL and BEOL processes can lead to PID. There is another device-level failure mode, PID, that is drawing more and more attention because it affects the reliability of both FEOL and BEOL.

#### 1.2.1 FEOL

**Hot-Carrier Injection**

HCI is a phenomenon of short-channel devices in which high field in the channel near the drain region leads to impact ionization, followed by injection of energetic carrier into the gate dielectrics and degradation in the MOS $I-V$ characteristics and parameter shift. Since the 1980s, HCI degradation in N-channel metal-oxide-semiconductor (NMOS) nodes has become a critical issue because of the continuous scaling of transistor dimensions without accompanying scaling of the power-supply voltage $V_{dd}$ (maintained at 5 V).\(^3\) HCI immunity was achieved for scaling devices from 1- to ~0.5-μm nodes by new drain and gate architectures derived from the lightly doped drain (LDD) structures. The corresponding worst-case HCI degradation of NMOS nodes was clearly identified in correlation with the large amount of energetic carriers at the maximum of $I_{sub}$ condition at $V_{gs} = V_{ds}/2$ corresponding to the maximum of interface traps generated.\(^4\)

From the mid-1990s, together with the scaling of channel length and oxide thickness, $V_{dd}$ was forced to be reduced considering both reliability requirements and power-consumption reduction. Hence HCI damage became less of an issue for CMOS nodes down to 120 nm.
However, with CMOS node continuous scaling down to 100 nm and beyond, HCI degradation has become a concern again because $V_{dd}$ scaling is slowing or stopping, whereas gate length is continuing to scale down to 40 nm and beyond. In addition, the requirement for stressing an MOS device and passing HCI has been changed from $V_{gs} = V_{dd}/2$ to $V_{gs} = V_{dd}$, and this leads to a higher substrate current and worse HCI, whereas the temperature for stressing an MOS device and passing HCI also has shifted from room temperature to high temperature. Chapter 5 details the HCI phenomenon with emphasis on the 90-nm and beyond technology node, where temperature/oxide-thickness dependence and degradation mechanisms of thin oxide are somehow different from those of previous technologies. The HCI improvement methodology through design and process optimization will be explained.

**Gate-Oxide Integrity and Time-Dependent Dielectric Breakdown**

With the gate-oxide physical thickness becoming thinner and thinner (~1.2 nm at the 65-nm node with only a few monolayers of Si-O bonds), the oxide becomes more and more susceptible to gate current leakage (~100 A/cm$^2$ at 1.0 V), defect creation, and ultimately, dielectric breakdown. With thinner oxide, the electrical field across the gate oxides has been increased with each technology node and will increase further at future technologies, which leads to further challenges to reliability assessment and control. The International Technology Roadmap for Semiconductor (ITRS) predicts that the oxide electrical fields under normal operation can be up to 10 MV/cm. The TDDB behaviors of the gate oxide also have been changed and mixed with hard breakdown (HBD), soft breakdown (SBD), and stress-induced leakage current (SILC), which lead to harder interpretation of the testing data and lifetime extrapolation. The lifetime-extrapolation methodology and model also have been evolved from the $1/E$ model for thick oxide to $E$ model for intermediate-thickness oxide and to the current-power-law model for thin oxide. Chapter 6 will introduce the gate-oxide integrity (GOI) and TDDB phenomena, mechanisms, and models, with focus on the most advanced technology. The improvement methodology through process optimization also will be elaborated.

**Negative-Bias Temperature Instability**

NBTI on P-channel metal-oxide-semiconductor (PMOS) nodes causes a shift in threshold voltage $V_t$ and a decrease in the mobility of the inversion channel. The basic understanding is that the Si-H bonds at the Si-SiO$_2$ interface, which are needed for interface-state passivation, can become broken under NBTI stressing. NBTI has become more and more important today because electrical fields in the gate oxide are higher and devices are operating at higher temperatures owing to higher power consumption. In addition, the voltage headroom
Part 1: General

(difference between $V_g$ and $V_t$) becomes much smaller than in the past, and a smaller $V_t$ shift ($\Delta V_t$) can lead to a larger $I_{sat}$ shift ($\Delta I_{sat}/I_{sat}$) [details are given in Chap. 2, Eq. (2.50)]. The NBTI has a larger impact on the ring oscillators at lower operating voltages. Thus the NBTI impact on $V_{min}$ circuit operation can be quite significant and an important reliability concern. Chapter 7 will provide a comprehensive summary for the negative-bias temperature instability phenomenon, which is observed more clearly in thinner oxides. The corresponding mechanisms, models, and improvement methodology through process optimization will be detailed.

Plasma-Induced Damage
PID can be a serious reliability concern for state-of-the-art ICs of modern process generations. Although in recent years the publications on PID have been much reduced, PID is still detectable for the most advanced gate oxides, for example, with an oxide thickness of 1.6 nm. In general, specially designed test structures are used to monitor the severity of PID during manufacturing. However, it is hard to ensure that the test structures and the corresponding measurement methods are carefully designed so that the PID effects are properly detected. In addition, measurements for evaluating failure rates in the parts-per-million (ppm) regime are not possible because the size of test structures for PID investigation employed in one test vehicle is limited. Investigations in the literature demonstrate that dielectrics of a wide thickness range can degrade because of PID, leading to significant MOS transistor parameter shift such as threshold voltage $V_t$, transconductance $G_m$, and gate-oxide leakage currents $I_{g_{leak}}$. Different gate-oxide thickness indicates different PID sensitivity. To prevent IC failures because of PID, design rules (DRs) have been established as guidelines to ensure the layout robustness against a large amount of PID that typically occurs during processing. However, even with design rules in place, PID damage to gate dielectrics is still possible, attributed to some hidden layout weaknesses with respect to the PID DRs. Protection schemes such as protection diodes are widely employed to reduce the PID impact, but sometimes they are not properly placed. Chapter 8 details the PID mechanism, antenna ratio, test structures, characterization methods, and PID control/improvement methodology through design and process changes.

Electrostatic Discharge
The ESD protection for CMOS devices has become more and more challenging because there is a general downward trend in ESD robustness with scaling, including the use of silicide, channel-length, and gate-oxide thickness reduction; the use of fin field effect transistor (FinFET) and silicon on insulator (SOI); and so on. Chapter 9 provides give a summary for ESD phenomena and how to prevent ESD-induced damage.
1.2.2 BEOL

The transition from aluminum- to copper-based interconnect technology has been an important milestone for the continued scaling of micro-/nanoelectronics. The technological benefits of integrating Cu as the conductor material for interconnect metallization of advanced ultra-large-scale integration (ULSI) include, for example, the reduced resistance-capacitance (RC) delay and the improved EM resistance, which are quite accepted by the industry. For example, the EM lifetime for Cu is much greater than that for Al, by more than 100 times, owing to the lower diffusivity of Cu compared with Al. Copper interconnects will continue to be used for the 32- and 22-nm technology nodes.

To accompany the application of Cu interconnects, the dual-damascene (DD) process flow has been developed, which requires the use of electrolytic Cu plating and chemical mechanical polishing (CMP) techniques. Major concerns with Cu were its extremely high diffusivity in silica or polymer dielectrics and its vulnerability to corrosion. A sidewall barrier layer, usually a derivative of tantalum, titanium, or tantalum nitride, thus is proposed to effectively prevent Cu diffusion at temperatures below 873 K. All these approaches have created a complex electromigration, stress migration, and Cu/low-\(k\) dielectric breakdown behaviors.

Electromigration

EM is the migration of metal atoms in a conductor owing to an electric current and can lead to the formation of voids in the interconnects and finally failure of the circuits. The aggressive shrinking of Cu lines and vias, together with the usage of (ultra) low-\(k\) dielectric materials for newer technologies, has exacerbated EM challenges. With the smaller vias, the critical void size for circuit failure also was reduced. The interconnect failure also becomes more sensitive to void shape and location. Therefore, the process variation in the new technologies will lead to a broader EM failure time distribution, in particular, for interconnections without sufficient liner/via redundancy, imposing a larger EM challenge.

In DD interconnects, a high-aspect-ratio via has been recognized as the most complicated region for integration and also has been reported as the weakest link for reliability. Although it is hard, it is very critical to form a thin but conformal barrier in the via. Otherwise, a via with a nonuniform barrier with poor adhesion/integrity/mechanical strength may result in serious extrinsic EM reliability issues. Thus via-relevant processes such as via barrier coverage, via barrier etch, and via cleaning definitely will have an impact on EM performance.

Another critical aspect of DD interconnects is the Cu–cap-layer interface, which has been found as the fastest diffusion path during
EM stressing. In general, Cu interconnect reliability can be much improved by good liner and cap adhesion and barrier properties and integration practices that prevent contamination or corrosion of the metal films and their interfaces.

Accompanying the continuously decreasing $k$-value of the low-$k$ dielectric with shrinkage is a higher level of porosity, which further leads to a poorer thermal conductivity. The exponential dependence of thermal conductivity of low-$k$ dielectric on its $k$ value is given in Eq. (1.1)\(^\text{19}\):

$$\lambda \propto \exp(1.05^*k)$$  \hspace{1cm} (1.1)

where $\lambda$ is the thermal conductivity in mW/cm·K, and $k$ is the dielectric constant.

In contrast, the operating current density inevitably increases owing to the scaling trend. This increase in current density causes not only an acceleration of EM degradation but also a temperature rise owing to joule heating. The application of the low-$k$ dielectric makes the situation even worse. Chapter 10 details the EM phenomenon, including mechanisms, lifetime prediction methodology, process effect, and test-structure effects. This chapter also will show how the joule heating affects the interconnect EM performance and how to control the joule heating effect from a design point of view.

**Stress Migration**

SM can result in void formation under and/or within vias for advanced interconnect systems and can be accelerated by high-temperature baking (150–200°C)\(^\text{20}\). Generally, this has been an issue when a single minimum-size via is connected to a wide Cu lead. In this case, the high Cu diffusivity path at the Cu–cap-layer interface and stress gradients under the via enable voids to nucleate below the via and easily grow to a fatal size, which becomes even worse when there is no conductive shunt layer at the via-bottom–Cu–cap-layer intersection. On the other hand, SM also has been reported more recently for narrow metal leads\(^\text{21}\). With continued scaling of the via, via voiding with narrow metal leads indeed becomes an issue simply because fewer vacancies will be required to form a “killer” void for abrupt resistance rise and failure. Chapter 11 will present the SM mechanisms, finite element analysis as a tool for SM understanding, and then how to reduce SM damage through process modification and optimized interconnect design.

**Intermetal Dielectric Time-Dependent Dielectric Breakdown**

With a high electrical field continuously applied to the BEOL intermetal dielectrics (IMDs), similar to FEOL gate dielectrics, damage can be induced in IMDs and the interfaces between the interconnects and the IMDs, that will lead to the formation of a conduction path and
finally dielectric breakdown—commonly referred as TDDB. In older technologies, when the spacing between metal interconnects was relatively large, the IMD TDDB is a much less critical reliability concern because the electrical field across the BEOL dielectric is low. However, current-interconnect low-\(k\) dielectric minimum spacing can be 70 to 80 nm or less, which is similar to gate-oxide thickness approximates 20 years ago. In other words, the electrical field across the IMD increases continuously with continuous scaling, as shown in Fig. 1.1. Although the electrical field across the BEOL dielectric is still about one order lower than that across the gate dielectrics for 32-nm technology nodes and below, the breakdown strength of the BEOL dielectric is much less because the intrinsic quality of these low-\(k\) dielectrics is much less perfect than that of gate oxides in terms of both electrical\(^{22}\) and mechanical strength.\(^{23}\) There are a number of reasons for the low breakdown strength of BEOL dielectrics.\(^{13}\) First of all, there is a much higher defect density in the IMDs, especially for low-\(k\) dielectrics.\(^{22}\) Such preexisting defects serve as precursors for traps during TDDB stressing. Second, damage or contamination of the dielectric can be induced during processing such as CMP. Furthermore, Cu diffusion into the dielectric can take place if the integrity of the barrier layers is not well ensured during processing.\(^{24}\) Last but not least, patterning problems such as line-width roughness (LWR) or via misalignment can result in a locally high electrical field. As a result, IMD TDDB becomes a more and more critical reliability concern.

![Figure 1.1](image)

**Figure 1.1** A plot of the electrical field across FEOL gate dielectrics and BEOL dielectrics for minimum pitch of interconnects in terms of the technology node.
10 Part 1: General

when device dimensions are scaled and (ultra) low-\(k\) materials with high porosity are used as BEOL insulators. Chapter 12 summarizes the most updated knowledge and understanding of the intermetal dielectric breakdown mechanisms and how to control reliability through design and process optimization, with especial focus on low-\(k\) materials.

1.3 Process-Dependent Reliability

Reliability is definitely process dependent. Different process steps and conditions potentially can affect reliability and its related parameters. It is imperative for process engineers to understand the process-reliability correlation so as to make prudent process changes for reliability robustness. How process parameters affect reliability is one of the key focuses of this book. An example is the hydrogen-related processes in forming-gas annealing (FGA), which is a common practice to passivate the dangling bonds in the gate-oxide–channel interface—an approach to recover the plasma-induced damage. For FGA, the key parameter is the thermal budget, which can be characterized by the annealing temperature and time, deciding how much the hydrogen will be incorporated into the device. In general, the higher the thermal budget, the more hydrogen will be introduced. When considering the effects of reducing thermal budget (i.e., less hydrogen is incorporated), it is seen that

- The retention time of dynamic random-access memory (DRAM) is degraded, which is not desirable.
- On the other hand, the NBTI and HCI performance could be improved.

Therefore, it is a must (but very tricky) to optimize the processes for a highly reliable ULSI system in future generations in order to meet requirements in terms of both reliability and device characteristics. Note that the optimized region becomes smaller and smaller with technology advances, meaning that reliability is more and more process dependent.

1.4 Reliability Assessment Methodology

For reliability assessment, a common and realistic practice is to perform accelerating tests under conditions that are much more “severe” than those under operation conditions to obtain failure statistics within a reasonable testing time. The so-called severe conditions in accelerating tests in general mean a much higher stressing temperature or stressing current/voltage or both than the operation conditions. After obtaining the time-to-failure (TTF) and failure distributions on specially
designed test structures from the accelerating tests, the assessments of reliability under normal operation for a larger chip area then is realized through the relevant physical/mathematical models by extrapolation. There are three extrapolations from the accelerating test to the normal operation as follows:

1. *Toward lower percentiles (typically 0.1 percent or 1 ppm).* This extrapolation is done based on the assumed statistical distribution (Weibull or log-normal, etc.) because we always have a limited number of devices under test (DUTs).

2. *Toward larger chip areas (sometimes referred to as area scaling).* TTF of a smaller-area \( (A_1) \) DUT is tested, which is extrapolated to a much larger chip area \( (A_2) \) based on the following Poisson relationship

\[
\frac{TTF_1}{TTF_2} = \left( \frac{A_2}{A_1} \right)^{1/\beta}
\]

where \( \beta \) is an area scaling factor.

3. *Toward operating conditions.* This extrapolation is done based on acceleration models for the specific failure mechanism from the stress conditions (i.e., high voltage/current and/or high temperature) to operating conditions (i.e., relatively low voltage/current and/or low temperature). Table 1.1 shows

<table>
<thead>
<tr>
<th>Reliability Item</th>
<th>Acceleration Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gate-oxide TDDB</td>
<td>( \text{MTF} = AV^{-n} \exp \left( \frac{E_a}{k_BT} \right) )</td>
</tr>
<tr>
<td>Hot-carrier injection</td>
<td>( \text{MTF} = A(i_{sub})^{-m} \exp \left( \frac{E_a}{k_BT} \right) )</td>
</tr>
<tr>
<td>Negative-bias temperature instability</td>
<td>( \text{MTF} = A \exp(-B \cdot V_g) \exp \left( \frac{E_a}{k_BT} \right) )</td>
</tr>
<tr>
<td>Electromigration</td>
<td>( \text{MTF} = Aj^{-n} \exp \left( \frac{E_a}{k_BT} \right) )</td>
</tr>
<tr>
<td>Stress migration</td>
<td>( \text{MTF} = A(T_o - T)^{-n} \exp \left( \frac{E_a}{k_BT} \right) )</td>
</tr>
<tr>
<td>Intermetal-dielectric TDDB</td>
<td>( \text{MTF} = A \exp(-\gamma E) \exp \left( \frac{E_a}{k_BT} \right) )</td>
</tr>
</tbody>
</table>

**Table 1.1** Typical Acceleration Models for Various Reliability Items
some typical acceleration models for various reliability items. Therefore, the acceleration models are very critical for an accurate projection of lifetime from accelerated stress data to operating conditions. Indeed, these models, in conjunction with their reliability physics, have migrated along with the migration of the technology nodes because the models should properly reflect the physics of the reliability failure process. The migration of the models for each reliability item is, thus, one of the focuses of this book.

Figure 1.2 provides an example of IMD TDBD lifetime projection from stress conditions (voltage 32 V, 0.01 m) to operating conditions (voltage 3.63 V, 100 m). The voltage acceleration is based on the $E$ model, whereas the area/length projection is based on the area scaling given in Eq. (1.2). In this book, the test structures, the accelerating tests, and the corresponding physical/mathematical models will be elaborated for the reliability items in respective chapters.

1.5 Organization of This Book

This book consists of three parts as shown in Fig. 1.3. Part 1 includes some background knowledge for understanding semiconductor process reliability. Chapter 1 provides an introduction to semiconductor process reliability, its importance, the reliability items, and an overall view of the book. Chapter 2 discusses the basic device physics that serve as base for understanding process-reliability mechanisms. Chapter 3 presents the basic semiconductor manufacturing process flow, including FEOL and BEOL flows, knowledge of which will assist readers in understanding the process effects on reliability performance. Chapter 4 includes some measurement methodologies to
identify the interface trap and oxide trap in a gate-oxide dielectrics induced by reliability stressing and degradation, information about which is a prerequisite to understanding reliability mechanisms.

Part 2 discusses in detail FEOL reliability items, including hot-carrier injection (Chap. 5), gate-oxide integrity and TDDB (Chap. 6), negative-bias temperature instability (Chap. 7), plasma-induced damage (Chap. 8), and electrostatic discharge (Chap. 9).

Part 3 presents BEOL reliability items including electromigration, (Chap. 10), stress migration (Chap. 11), and intermetal dielectric breakdown and TDDB (Chap. 12).

1.6 References

Chapter 1: Introduction


