IONIZING RADIATION EFFECTS IN MOS DEVICES AND CIRCUITS

Edited by

T. P. Ma
Department of Electrical Engineering
Yale University
New Haven, Connecticut

and

Paul V. Dressendorfer
Sandia National Laboratories
Albuquerque, New Mexico
MOS (metal-oxide-semiconductor) circuits are currently the cornerstone of the modern microelectronics industry. When an MOS device is exposed to a flux of energetic photons or particles, the resulting effects from this radiation can cause severe degradation of the device performance and of its operating life. These radiation effects are of both practical and scientific interest.

Since 1962, when the Telstar I communication satellite failed as a result of the detrimental effects from radiation in the Van Allen belts, there has been intensive study into the effects of ionizing radiation on semiconductor devices. This research, driven by the need for circuits that can function properly in a radiation-rich environment, has focused on the development of so-called radiation-hardened devices. Such circuits are essential for interplanetary space exploration, communication, meteorological, navigational, and surveillance satellites, advanced weaponry, instrumentation for nuclear power plants, and detectors for high-energy physics experiments.

Another driving force for research in this area has been the quest for ever higher density, higher performance integrated circuits. The advanced processing techniques required to manufacture such circuits often utilize energetic particles or photons (e.g., electron-beam lithography, X-ray lithography, reactive ion etching, and various plasma processes) and thus may cause significant radiation damage. In this regard, the practical objective of the research is to control and to remove the radiation damage in order to ensure proper functionality of the circuits after they are fabricated.

For both of the above applications, it is important to understand the fundamental mechanisms involved in the creation of radiation-induced charge carriers and bonding defects, and their effects (both transient and long term) on the electrical properties of devices. In addition, the studies of radiation effects have led to a better understanding of the nature of many commonly observed electronic defects in MOS device structures. Consequently, the use of radiation has become a valuable laboratory tool enabling researchers to study the generation, evolution, and annealing of these defects. For example, results from such studies have provided insight into hot-carrier-induced degradation, a significant reliability concern which can occur in VLSI (very-large-scale integration) circuits.

A wealth of literature has been created in this field over the last 25 years. However, knowledge about the significant advances made during this period can be gathered only from the original research contributions scattered in
scientific journals, technical reports, and conference proceedings. Review articles have been published in a few areas, but their scope is generally limited, and often they are written at an advanced level requiring a strong background in the area for complete understanding. A distinct need has arisen for a treatise that would provide a comprehensive account of ionizing radiation effects in MOS device structures and would cover introductory material as well as detailed specific developments. *Ionizing Radiation Effects in MOS Devices and Circuits* is intended to fill this gap in the literature.

Given the breadth and depth of coverage required, we considered it best to have the different topics written by experts in each area. Keeping in view the interests of the nonspecialist, special efforts have been made to ensure coherence among the chapters with a uniformity of presentation and notations, and to include sufficient introductory materials so as to be useful as a self-contained tutorial text. For the experts, on the other hand, the chapters are written so that specific topics of interest can be read and understood without reference to chapters which appear elsewhere in the monograph.

Each chapter is a critical review of and guide to the literature in a particular subject area, with an overall perspective and context provided by an expert in the field. The chapters are not a series of journal articles, but rather each presents an up-to-date overview of an area with references to the original articles for detailed derivations and extensive experimental results. The assessment and coverage of the literature is intended to be comprehensive, and more than 1300 citations of original works are made. This extensive bibliography should serve as a valuable source of reference and includes the titles of the works to assist in determining those articles which are of most interest. To make the book a useful guide for practitioners, figures and tables throughout the text are reproduced from actual experimental data whenever possible.

It is our hope that this book will be of value to solid state physicists, material scientists, semiconductor process engineers, reliability assurance and hardness assurance engineers, designers of and project managers for space and other hardened systems, VLSI project managers—indeed, anyone who is concerned with either the fundamental or the practical issues of radiation effects in MOS devices. It should be useful both to the neophyte to this area, providing a comprehensive, comprehensible overview, and to the expert, serving as a detailed reference to the vast literature base which has developed over the years. It is also aimed at graduate students who are preparing to enter any of these professions, and to those in other areas who wish to acquaint themselves with this important and rapidly growing field.

The first chapter provides primarily background material for those unfamiliar with either basic MOS device physics or the phenomenology of radiation effects. Chapter 2 next gives a historical perspective of the problem of radiation effects in MOS devices and follows the course of hardened technology development with some projections for future efforts as MOS integrated circuit technology evolves.
Chapters 3 and 4 are a comprehensive treatment of the basic mechanisms of radiation damage in MOS structures. These include charge generation, transport, and trapping, and the creation of interface trap levels at the Si/SiO₂ interface. The material in these chapters sets the foundation for a more in-depth appreciation of the material in Chapters 5 through 8.

Chapter 5 discusses in detail the practical effects of radiation on various MOS device and circuit parameters. An extensive set of examples of the changes which may be observed in various types of circuits is included.

Chapters 6 and 7 are concerned primarily with the fabrication of MOS devices and circuits. The impact of various processing parameters on the subsequent radiation response of MOS devices, strategies for integrating processes into a hardened technology, and issues for characterizing and maintaining that technology are covered in Chapter 6. In Chapter 7 the technologically important issues associated with the fabrication of VLSI circuits are detailed. Advanced processes which can contribute to radiation damage in such devices are described, as are means of annealing that damage. In addition, this chapter presents a section dealing with hot carrier effects and their implications for long-term reliability.

The radiation sources for both real system environments and those utilized in testing circuits are described in Chapter 8. Dosimetry principles and techniques for measuring the strength of radiation sources are explained.

Whereas Chapters 3 through 8 focus on total dose effects, Chapter 9 centers on the effects of transient radiation events. Such events can lead to both hard errors, where the device is permanently altered, and soft errors, where information in the circuit or device is temporarily disrupted. Chapter 9 details theoretical and experimental aspects of these phenomena, both for pulses of ionizing radiation and for single events caused by the passage of a highly energetic particle.

We would like to express our thanks and appreciation to a number of people who assisted us during the course of this endeavor. Dr. Jim Schwank, Dr. Dan Fleetwood, Dr. Peter Winokur, and Dr. Harry Weaver critically reviewed portions of the manuscript and contributed a number of valuable suggestions. Invaluable secretarial support was provided by Martha Tindell and Arlene Ciociola. Perhaps most of all, we would like to thank the authors of the various chapters, who worked many long hours in attempting to achieve a difficult goal.

In a book of this type, despite our best intentions, it is likely that omissions or errors have occurred. We welcome any comments, suggestions, or discussions on any of the topics contained herein.

Paul V. Dressendorfer
T. P. Ma
CONTENTS

Symbols xiii

1. Introduction 1
   T. P. Ma and Paul V. Dressendorfer

2. Historical Perspective 47
   H. Hughes

3. Electron-Hole Generation, Transport, and Trapping in SiO₂ 87
   F. B. McLean, H. E. Boesch, Jr., and T. R. Oldham

4. Radiation-Induced Interface Traps 193
   Peter S. Winokur

5. Radiation Effects on MOS Devices and Circuits 256
   Paul V. Dressendorfer

6. Radiation-Hardening Technology 333
   Paul V. Dressendorfer

7. Process-Induced Radiation Effects 401
   T. P. Ma

8. Source Considerations and Testing Techniques 443
   K. G. Kerris

   Sherra E. Kerns with contributions from B. D. Shafer

Index 577
1 INTRODUCTION

T. P. Ma
Department of Electrical Engineering
Yale University
New Haven, Connecticut

Paul V. Dressendorfer
Sandia National Laboratories
Albuquerque, New Mexico

1.1 Relevance of ionizing radiation effects
1.2 Introduction to MOS device physics
   1.2.1 The ideal MOS capacitor
   1.2.2 Practical MOS capacitors
   1.2.3 Physics of MOSFETs
1.3 Phenomenological description of ionizing radiation effects
   1.3.1 Total dose effects
   1.3.2 Transient radiation effects
1.4 Book overview
   References

Silicon MOS (metal-oxide-semiconductor) devices have become the mainstay of the semiconductor industry. When these devices are exposed to ionizing radiation, significant changes can occur in their characteristics. This book will detail those changes, ranging from the underlying basic mechanisms through their effects on circuits and their impact on technologies. In this introductory chapter, an overall groundwork will be laid for the succeeding chapters.

First the overall relevance of this field to scientific study, advanced technology, and practical applications will be discussed. The next sections of this chapter are oriented toward those readers not familiar with the field of ionizing radiation effects on MOS devices. A general discussion of MOS device physics is presented both to establish the nomenclature to be followed throughout this book and to provide enough elementary background material that the discussions in subsequent chapters can be easily followed by those with minimal knowledge in this area. Next a phenomenological description of ionizing radiation effects is given. This
section is intended to provide an awareness of the kinds of effects encountered, and allow those readers new to the field to obtain the background necessary to understand which effects may be of most interest to them. The chapter concludes with an overview of the contents of the following chapters, and a guide to most effective utilization of the book.

1.1 RELEVANCE OF IONIZING RADIATION EFFECTS

Ionizing radiation creates mobile electrons and holes in both the insulator and silicon substrate in MOS devices and leads to several kinds of damage in these devices. These properties have allowed its use as a tool for scientific study in a number of areas. The basic mechanisms of carrier transport in insulators have been very effectively explored by using various types of ionizing radiation to create mobile carriers and then monitoring their motion by electrical means. These studies have furthered our understanding of polarons, excitons, and trap-hopping processes (see Chapter 3). The generation of interface traps and oxide trapped charge in large numbers by ionizing radiation has allowed the identification of the atomic structures associated with these defects (see Chapters 3 and 4). By providing a means of altering the trapped charge at the SiO$_2$/Si interface in a given device, the interaction of mobile charge carriers in the channel of an MOS device with that trapped charge can be explored (see Chapter 5). By creating trapped charge distributions in the oxide layer which provide traps for carriers, tunneling and carrier capture phenomena can be effectively studied (see Chapter 3).

As the semiconductor industry progresses to the VLSI and ULSI era, the technological impact of ionizing radiation effects becomes more and more important. In order to produce the extremely fine geometries required at high levels of integration, the processes used in the manufacture of the integrated circuits themselves may produce ionizing radiation (see Chapter 7). Hot carrier effects caused by the high fields in small geometry devices cause effects very similar to those of ionizing radiation (see Chapter 7). At the small geometries of current and future integrated circuits, latchup [the triggering of the parasitic SCR (semiconductor-controlled rectifier) present in CMOS (complementary MOS) circuits on silicon substrates] initiated by normal operating conditions has become a major concern. This trend toward small devices has also made normal commercial ICs (integrated circuits) susceptible to single event upsets caused by ionizing particles created by the decay of residual radioactive material in IC packaging material (Chapter 9). Thus many of the concerns for radiation-hardened circuits are becoming a concern for standard commercial products. In addition, in order to make circuits for the specialized applications requiring operation in an ionizing radiation environment, significant modifications to the technology employed must be made (see Chapter 6).
There are a large number of specialized applications requiring ICs that have a known, predictable response to ionizing radiation. Satellite systems need electronic components that can operate in the harsh radiation environment around the earth and in space. Without such components, satellites for communication, surveillance, resource mapping, and the like would have extremely limited capabilities; it would be almost impossible to conduct interplanetary explorations such as those of the Pioneer and Voyager missions. Many weapons systems require hardened components to perform their tasks properly through an operational scenario. Nuclear power plants need instrumentation which can withstand the environment near the reactor and continue to provide reliable data. In the medical field, specially produced MOS dosimeters have been used to monitor the radiation dose delivered to patients undergoing radiation therapy [1]. A nonvolatile computer memory component based on the radiation damage created by an electron beam on MOS structures has even been proposed [2].

Thus a great deal of useful information can be gathered by using ionizing radiation as a tool to study the physical effects in insulators and semiconductors, by evaluating the effects of such radiation on devices, and by utilizing circuits that have been produced to be hardened to damage from ionizing radiation. The chapters that follow attempt to illustrate some of this information and provide a background for further study.

1.2 INTRODUCTION TO MOS DEVICE PHYSICS

While many excellent books [3–5] exist in which detailed treatments of MOS device physics can be found, we decided to provide a brief introduction in this section with sufficient background materials, so that subsequent chapters can be readily followed by those with minimal knowledge in this area.

As shown schematically in Fig. 1.1, the terminals of a typical metal-oxide-semiconductor field effect transistor (MOSFET) are designated as the gate \((G)\), the source \((S)\), the drain \((D)\), and the substrate \((SB)\). The portion of the semiconductor under the gate connecting the source and the drain is called the channel region. The current that flows in the channel depends on the voltage applied.

Depending on the type of conducting carriers flowing in the channel, a MOSFET can be classified as either an \(n\)-channel (electron conduction) or a \(p\)-channel (hole conduction) device. If the transistor is normally “off” (no current flowing at \(V_G = 0\)), it is called an enhancement mode device; if the transistor is normally “on,” it is a depletion mode device.

The operation of the transistor may be described qualitatively as follows, using the grounded source configuration (both the source and the substrate are tied to ground) of an \(n\)-channel enhancement mode MOSFET as an example. When the gate voltage \(V_G\) is set to zero, the source-drain current is
limited by the reverse saturation current of the two back-to-back $p-n$ junctions, which may be considered negligible for the purpose of this discussion. Therefore, the device is in its “off” state. When a small positive $V_G$ is applied, the majority carriers (holes in this case) in the channel underneath the gate oxide will be depleted, due to the field from the positive gate charge. This depletion region acts as an insulator, and again the source-drain current will be negligible. When a sufficiently large, positive $V_G$ is applied, however, a significant number of minority carriers (electrons in this case) will be attracted to the oxide–silicon interface. These electrons are supplied from both the bulk of the silicon and the source and drain regions. Once a sufficient number of electrons are accumulated there, a conduction channel is formed, allowing an appreciable amount of current to flow between the source and drain terminals. The device is now in its “on” state. Since the concentration of the electrons in the channel increases with increasing $V_G$, the magnitude of the current for a given drain voltage can thus be modulated by changing the gate voltage. The operation of a $p$-channel device can be similarly described, except for a change in the voltage polarity and in the sign of the charges.

The minimum gate voltage required to cause appreciable current to flow is called the *threshold voltage*, $V_t$, which will be defined more precisely in Section 1.2.4.

Thus, it is apparent that the key element in a MOSFET structure is the MOS capacitor (the structure in Fig. 1.1 without the source and drain), and the physics of a MOSFET becomes clear if one understands the operation of this simple two-terminal device.
1.2.1 The Ideal MOS Capacitor

To facilitate the discussion, the following assumptions are made for an ideal MOS capacitor: (1) there is no current flowing through the gate insulator; (2) there is no work function difference between the gate electrode and the semiconductor; (3) the semiconductor is doped uniformly; (4) there is no charge in the gate insulator; and (5) there are no interface traps.

Departures from these assumptions in a more realistic device will be considered later.

A useful starting point for visualizing the response of the MOS capacitor to an applied voltage is its energy band diagram.

1.2.1.1 Energy Band Diagram of an Ideal MOS Capacitor

As shown in Fig. 1.2(a), the ideal MOS capacitor is in thermal equilibrium at zero applied voltage, and the system is characterized by a constant Fermi level throughout all three materials: the metal, the oxide, and the silicon. As a result of the assumptions made earlier for the ideal MOS capacitor, the silicon is in its flatband condition for $V_G = 0$, and the electron (hole) density near the SiO$_2$/Si interface is the same as that everywhere else in the silicon.

When a voltage is applied to the gate, the flatband condition no longer exists, and the system can be characterized as in accumulation, depletion, or inversion, depending on the polarity and the magnitude of $V_G$. Consider the p-type semiconductor first. When a negative voltage ($V_G < 0$) is applied, holes will be attracted to the SiO$_2$/Si interface and pile up there. This is known as the accumulation condition. In accumulation, the majority carrier concentration is greater near the SiO$_2$/Si interface than in the bulk of silicon. The resulting band diagram is shown in Fig. 1.2(b).

When a small positive voltage ($V_G > 0$) is applied, holes are depleted in the vicinity of the SiO$_2$/Si interface, leaving behind a space charge region formed by the charged acceptors. This is known as the depletion condition. The resulting band diagram is shown in Fig. 1.2(c). The depth of the depletion region increases with increasing applied voltage; at the same time, electrons (minority carriers in this case) are attracted toward the interface, and their concentration also increases with $V_G$. But in the depletion region, the electron concentration may be considered negligibly small.

If one keeps increasing $V_G$, eventually the electron concentration at the SiO$_2$/Si interface will increase to a value equal to or even greater than that of the hole concentration in the bulk of the silicon. In other words, the silicon surface becomes as though it is a n-type material. This is referred to as the strong inversion condition. The applied gate voltage at which strong inversion just begins is known as the inversion voltage, $V_i$, which is equivalent to the threshold voltage for a MOSFET. The energy band diagram corresponding to strong inversion is depicted in Fig. 1.2(d).

After the inversion condition is reached, the depth of the depletion layer will no longer increase with increasing applied voltage (under static, equilib-
Fig. 1.2  Energy band diagrams of ideal MOS capacitors under the following conditions: (a) flatband; (b) accumulation; (c) depletion; (d) inversion.
rium conditions). This occurs because the higher concentration of inversion charge near the SiO₂/Si interface and its rapid rise with gate voltage effectively shield the interior of the semiconductor from any additional charge placed on the gate. As a result, a so-called maximum depletion depth is formed when \( V_G \) reaches strong inversion or beyond.

In summary, three distinctive biasing regions may be identified: accumulation, depletion, and inversion. For an ideal \( p \)-type MOS capacitor, accumulation occurs when \( V_G < 0 \), depletion when \( V_G > V_i > V_G > 0 \), and inversion when \( V_G \geq V_i \). The above voltage polarities are simply reversed for an ideal \( n \)-type MOS capacitor. The flatband condition, at \( V_G = 0 \), is the dividing line between accumulation and depletion. The bias voltage \( V_G = V_i \) marks the transition point between depletion and strong inversion.

### 1.2.1.2 Qualitative Description of MOS C-V Curves

The small signal AC capacitance-voltage characteristics provide a wealth of information about the properties of MOS devices. In Fig. 1.3, a high-frequency and a low-frequency C-V curve for a representative ideal \( p \)-type MOS capacitor are shown on the same graph. In the case of the high-frequency measurement, the modulation frequency is sufficiently high that the minority carriers are too slow to respond to the voltage modulation. In the low-frequency case, the frequency is sufficiently low that the minority carriers can readily respond to the voltage modulation. In practice, 1 MHz is almost universally accepted for the high-frequency measurement, and a quasi-static technique [6, 7] is widely used to obtain the low-frequency characteristics.

The behavior of the C-V curves shown in Fig. 1.3 can be qualitatively understood as follows. Consider the high-frequency case first. In the bias range \( V_i > V_G > 0 \), the MOS capacitor is in depletion. Since, to first order,
the depletion layer contains no free carriers, it can be treated as an
insulator, and the total capacitance can be visualized as consisting of two
capacitors in series: that of the oxide, $C_{ox}$, and of the space charge layer $C_{sc}$
(see Fig. 1.2(c)). Therefore, the total capacitance of the MOS capacitor is
$C = C_{ox}C_{sc}/(C_{ox} + C_{sc})$, where $C < C_{ox}$ and $C < C_{sc}$. As $V_G$ increases, the
depletion layer deepens, and the MOS device capacitance drops with
increasing $V_G$. When inversion is reached ($V_G \geq V_t$), the depth of the
depletion layer reaches a saturated maximum value, due to the screening
effect by the inversion charge as discussed previously, and the MOS device
capacitance assumes a flat minimum value. The depletion and inversion
regions of the high-frequency $C$-$V$ curve are depicted to the right of the
$V_G = V_{fb}$ axis in Fig. 1.3. Now go to the negative voltage regime. When a
large negative gate voltage is applied, strong accumulation is reached and a
very high density of majority carriers piles up at the silicon surface. These
majority carriers effectively shield the interior of the silicon from feeling the
change of the electric field arising from the modulation of the gate voltage.
Since there is no depletion layer in the silicon, and the AC field lines all
terminate at the silicon surface, the silicon behaves as a short circuit as far as
the AC signal is concerned. Therefore, in strong accumulation, the system
behaves as though it is a parallel-plate capacitor with the oxide as the
dielectric spacer, and the oxide capacitance, $C_{ox}$, is approached. This
 corresponds to the maximum capacitance of the system, as shown in Fig.
1.3. As one moves from strong accumulation to the right along the voltage
axis, the degree of accumulation decreases, the shielding becomes less
effective, and the AC field lines start to penetrate deeper into the interior of
the silicon. As a result, the system again must be treated as two capacitors in
series, of which the capacitance due to the silicon part causes the total
capacitance to drop. This decrease in capacitance as $V_G$ increases continues
through the depletion region. As mentioned before, the flatband condition,
at $V_G = 0$ for the ideal MOS capacitor, is the dividing line between accumu-
lation and depletion.

Now we are ready to discuss the low-frequency $C$-$V$ curve. Since the
majority carriers can follow both the high-frequency and the low-frequency
signals, the low-frequency curve coincides with the high-frequency one in
accumulation and depletion, because no minority carriers are involved. This
is evident from Fig. 1.3. When inversion is approached, however, an
appreciable number of minority carriers start to show up near the SiO$_2$/Si
interface. In the case of the high-frequency measurement just discussed, the
minority carriers do not respond to the AC signal and therefore do not
contribute to the capacitance. In the low-frequency case, however, some of
the AC field lines terminate at the minority carriers instead of the edge of
the depletion layer, causing the capacitance to increase. In the limiting case
of very strong inversion, the AC field lines essentially all terminate right at
the SiO$_2$/Si interface, and the oxide capacitance is again approached, as
shown in the far right region of Fig. 1.3.
1.2.1.3 Derivation of ideal MOS C-V Curves

The C-V curves for an ideal MOS capacitor can be calculated quantitatively and are derived below.

Let us first consider an ideal p-type MOS capacitor with a low-frequency AC signal applied to the gate of the device. The differential capacitance per unit area of the system is defined as

$$C = \frac{dQ_G}{dV_G}$$  \hspace{1cm} (1.1)

where $Q_G$ is the areal charge density on the gate.

Since the charge on the gate must balance the charge inside the semiconductor, $Q_G = -Q_{sc}$, where $Q_{sc}$ is the total semiconductor charge per unit area.

Equation (1.1) can therefore be written as

$$C = -\frac{dQ_{sc}}{dV_G} = -\frac{dQ_{sc}}{d\phi_s} \frac{d\phi_s}{dV_G}$$  \hspace{1cm} (1.2)

where $\phi_s$ is the surface potential of the semiconductor. Equation (1.2) suggests that the MOS device capacitance can be obtained once we are able to establish the functions $Q_{sc}(\phi_s)$ and $\phi_s(V_G)$. To find the expression for the function $Q_{sc}(\phi_s)$, let us start from the one-dimensional Poisson’s equation in the semiconductor

$$\frac{d^2\phi(x)}{dx^2} = -\frac{\rho(x)}{\epsilon_s} = -\frac{q}{\epsilon_s} \left[ p(x) - n(x) + N_D - N_A \right]$$  \hspace{1cm} (1.3)

where $\phi$ is the semiconductor potential, $\rho$ the volume charge density, $\epsilon_s$ the permittivity in the semiconductor, $p$ the hole concentration, $n$ the electron concentration, $N_D$ the charged donor concentration, and $N_A$ the charged acceptor concentration. In Eq. (1.3), we have included both $N_D$ and $N_A$ for generality.

Using the charge neutrality condition in the semiconductor bulk

$$0 = p_0 - n_0 + N_D - N_A$$  \hspace{1cm} (1.4)

or

$$N_D - N_A = -(p_0 - n_0)$$  \hspace{1cm} (1.5)

where $p_0$ and $n_0$ are carrier concentrations in the semiconductor bulk.

From the Boltzmann relations, the carrier density for a given $\phi$ can be expressed as (assigning $\phi = 0$ in the bulk)

$$p(x) = p_0 \exp[-\beta \phi(x)] \quad \text{and} \quad n(x) = n_0 \exp[\beta \phi(x)]$$  \hspace{1cm} (1.6)

where $\beta = q/kT$. Plugging Eqs. (1.5) and (1.6) in (1.3), we obtain
\[
\frac{d^2\phi(x)}{dx^2} = -\frac{q}{\varepsilon_s} \left\{ p_0[\exp(-\beta\phi) - 1] - n_0[\exp(\beta\phi) - 1] \right\} \tag{1.7}
\]

Integrating Eq. (1.7) from the bulk toward the surface

\[
\int_0^{d\phi/dx} \left( \frac{d\phi}{dx} \right) d\phi = -\frac{q}{\varepsilon_s} \int_0^\phi \left\{ p_0[\exp(-\beta\phi) - 1] - n_0[\exp(\beta\phi) - 1] \right\} d\phi
\]

yielding

\[
\left( \frac{d\phi}{dx} \right)^2 = \frac{2kTP_0}{\varepsilon_s} \left\{ \exp(-\beta\phi) + \beta\phi - 1 \right\} + \frac{n_0}{p_0} \left[ \exp(\beta\phi) - \beta\phi - 1 \right] \tag{1.9}
\]

Taking the square root of Eq. (1.9), and defining

\[
L_D = \left( \frac{kT\varepsilon_s}{p_0q^2} \right)^{1/2} \tag{1.10}
\]

and

\[
F(\phi) = \left\{ \left[ \exp(-\beta\phi) + \beta\phi - 1 \right] + \frac{n_0}{p_0} \left[ \exp(\beta\phi) - \beta\phi - 1 \right] \right\}^{1/2} \geq 0 \tag{1.11}
\]

we obtain the electric field

\[
\mathcal{E}(x) = -\frac{d\phi}{dx} = \pm \frac{\sqrt{2kT}}{qL_D} F(\phi) \tag{1.12}
\]

with positive sign for \( \phi > 0 \) and negative sign for \( \phi < 0 \).

The electric field at the semiconductor surface is obtained from Eq. (1.12) by letting \( \phi = \phi_s \), where \( \phi_s \) is the surface potential:

\[
\mathcal{E}_s = \pm \frac{\sqrt{2kT}}{qL_D} F(\phi_s) \tag{1.13}
\]

By Gauss' law, the areal charge density corresponding to this surface field is

\[
Q_{sc} = -\varepsilon_s \mathcal{E}_s = \pm \frac{\sqrt{2}\varepsilon_s kT}{qL_D} F(\phi_s) \tag{1.14}
\]

Now we have accomplished our goal of expressing \( Q_{sc} \) as a function of \( \phi_s \). The next step is to relate \( \phi_s \) to \( V_G \).

For an ideal MOS capacitor, the applied voltage will be partly across the oxide and partly across the semiconductor.

Thus,
\[ V_G = V_{ox} + \phi_s \]  
(1.15)

where \( V_{ox} \) is the voltage drop across the oxide and is given by

\[ V_{ox} = \varepsilon_{ox} d_{ox} = -\frac{Q_{sc}}{C_{ox}} = -\frac{Q_{sc} \phi_s}{C_{ox}} \]  
(1.16)

where \( \varepsilon_{ox} \) is the field in the oxide, \( d_{ox} \) the oxide thickness, \( \varepsilon_{ox} \) the permittivity of the oxide, and \( C_{ox} \) the oxide capacitance.

Therefore, Eq. (1.15) can be rewritten as

\[ V_G = -\frac{Q_{sc} \phi_s}{C_{ox}} + \phi_s \]  
(1.17)

At this stage, we have expressed both \( Q_{sc} \) and \( V_G \) in terms of \( \phi_s \), and the capacitance of the MOS capacitor can be obtained by using Eqs. (1.2), (1.14), and (1.17) for any applied gate voltage, \( V_G \).

The \( C-V \) curve derived above is the low-frequency curve, because we have allowed all the minority carriers to follow the AC modulation signal. To obtain the high-frequency \( C-V \) curves, one must remember that the minority carriers respond in a quasi-static manner to the DC voltage but \textit{not} to the AC modulation signal. Therefore, the same derivation as presented above may be followed, except that in taking the derivative, \( dQ_{sc}/d\phi_s \), all the terms in \( F(\phi_s) \) involving minority carriers should be set to zero.

An alternative but more intuitive method of deriving the high-frequency \( C-V \) curve is by the use of the depletion approximation. This method is only applicable in depletion and inversion. In the accumulation region and at flatband, because the contribution from the minority carriers is negligible, the high-frequency and low-frequency \( C-V \) curves coincide, and one may use the low-frequency curve already derived to represent the high-frequency one.

When the MOS capacitor is in depletion, the total capacitance of the system is a series combination of the oxide capacitance \( C_{ox} (= \varepsilon_{ox}/d_{ox}) \) and the semiconductor space charge capacitance, \( C_{sc} \):

\[ C = \frac{C_{ox} C_{sc}}{C_{ox} + C_{sc}} \]  
(1.18)

where \( C_{sc} \) is inversely proportional to the depth of the depletion layer, \( X_d \):

\[ C_{sc} = \frac{\varepsilon_f}{X_d} \]  
(1.19)

The depth of the depletion layer is a function of the surface potential and is given by the same formula as for the case of a one-sided abrupt \( p-n \) junction:
\[ X_d = \left( \frac{2\varepsilon_s \phi_s}{qN_A} \right)^{1/2} \]  

(1.20)

The gate voltage can be expressed as

\[ V_G = \phi_s + V_{ox} = \phi_s - Q_{sc}(\phi_s)/C_{ox} \]  

(1.21)

where \( Q_{sc} \) in the depletion approximation can be written as

\[ Q_{sc} = -qN_A X_d \]  

(1.22)

From Eqs. (1.18)–(1.22), one obtains the relationship

\[ C = \frac{C_{ox}}{\left( 1 + \frac{2C_{ox}^2}{qN_A \varepsilon_s V_G} \right)^{1/2}} \]  

(1.23)

Equation (1.23) indicates that, in the depletion region, the capacitance drops off approximately as the square root of \( V_G \) for large \( V_G \).

When \( V_G \gtrsim V_i \), inversion is reached, the semiconductor surface potential is pinned at \( \phi_s = 2\phi_b \), where \( \phi_b = (E_i - E_f)/q \) in the bulk, and the depletion depth reaches a maximum, \( X_{dm} \). Hence, the total capacitance of the system for \( V_G \gtrsim V_i \) is at a flat minimum, \( C_{\text{min}} \), which can be derived from Eq. (1.18) as

\[ C_{\text{min}} = \frac{C_{ox}(\varepsilon_s/X_{dm})}{C_{ox} + (\varepsilon_s/X_{dm})} \]  

(1.24)

where

\[ X_{dm} = \left( \frac{4\varepsilon_s \phi_b}{qN_A} \right)^{1/2} \]  

(1.25)

comes from Eq. (1.20) for the special case of \( \phi_s = 2\phi_b \).

Thus, Eqs. (1.23) and (1.24) approximately describe the portion of the high-frequency \( C-V \) curve corresponding to \( V_G > 0 \) in Fig. 1.3. The error resulting from the approximation is greatest near \( V_G = 0 \) and continues to reduce with increasing \( V_G \). Typically at a gate voltage where the depletion layer becomes thicker than the oxide thickness, the error may be considered negligible.

The inversion voltage, \( V_i \), for the ideal MOS capacitor can be obtained by setting \( \phi_s = 2\phi_b \) in Eq. (1.21).

\[ V_i = 2\phi_b + \frac{(4qN_A \varepsilon_s \phi_b)^{1/2}}{C_{ox}} \]  

(1.26)
1.2.2 Practical MOS Capacitors

A practical MOS capacitor usually deviates from the ideal one by one or more of the following parameters: (1) a nonzero, metal-semiconductor work function difference; (2) a noticeable immobile oxide charge; (3) mobile ions in the oxide; and (4) interface traps. Their effects on the MOS C-V characteristics will be discussed next.

1.2.2.1 Metal-Semiconductor Work Function Difference

Analogous to the built-in potential of a metal-semiconductor diode, the work function difference between the gate metal and the semiconductor in a MOS structure causes a band bending of the semiconductor surface and a built-in field in the oxide. In order to achieve the flatband condition, a gate voltage equal to the work function difference, $\phi_{ms}$, must be applied to counteract the built-in potential. The gate voltage necessary to bring about flatband is called the flatband voltage, $V_{fb}$. If there are no other nonideal factors, then

$$V_{fb} = \phi_{ms} \quad (1.27)$$

In the discussion presented above, we used flatband condition as a reference only for convenience. In fact, we could just as well have chosen any predesigned surface potential, $\phi_s$, and the gate voltage required to bring about a given $\phi_s$ is shifted by the amount of the work function difference, $\phi_{ms}$. Thus, in general,

![Diagram](image.png)

Fig. 1.4 Parallel shift of high-frequency MOS C-V curve due to metal-semiconductor work function difference and oxide charge.
\[ \Delta V_G|_{\phi_s} = \phi_{ms} \]  

(1.28)

In Eq. (1.28), \( \Delta V_{fb} \) is just a special case of \( \Delta V_G|_{\phi_s=0} \).

The effect of the work function difference on the MOS C-V characteristic is therefore to cause a parallel shift along the voltage axis of amount \( \phi_{ms} \), as illustrated in Fig. 1.4.

### 1.2.2.2 Immobile Oxide Charges

In addition to the work function difference, a practical MOS capacitor typically contains some charge in the oxide. Some of the charge species do not move under the influence of the gate voltage and are referred to as *immobile oxide charge*; the others can be forced to move and are referred to as *mobile ions*.

The immobile charge that exists in the as-grown oxide is called the *fixed oxide charge*, as opposed to the *oxide-trapped charge*, which is introduced by trapping holes or electrons. As an example, the oxide-trapped charge is often observed after exposing the MOS device to ionizing radiation, as described in Chapter 3.

The origins of these charge species, and how they are affected by the various processes, have been reviewed elsewhere [4] and will not be further discussed here. In particular, the radiation-induced oxide charge is discussed in detail in Chapter 3. Regardless of its origin, an immobile oxide charge will cause a parallel shift of the MOS C-V characteristic along the voltage axis, as is shown next.

Suppose the spatial distribution of the oxide charge density is \( \rho(x) \). This charge distribution will induce an image charge partly in the metal and partly in the semiconductor. Therefore, at \( V_G = 0 \), the semiconductor is no longer at flatband even without any metal-semiconductor work function difference. In order to bring about the flatband condition, one must apply an incremental gate voltage of appropriate polarity and magnitude. It can be shown [3–5] that this incremental voltage corresponds to

\[ \Delta V_{fb} = -\frac{1}{C_{ox}} \int_0^{d_{ox}} \frac{x}{d_{ox}} \rho(x) \, dx \]  

(1.29)

where the origin of the \( x \) axis is set at the metal/SiO\(_2\) interface.

Again, Eq. (1.29) may be generalized for any given \( \phi_s \)

\[ \Delta V_G|_{\phi_s} = -\frac{1}{C_{ox}} \int_0^{d_{ox}} \frac{x}{d_{ox}} \rho(x) \, dx \]  

(1.30)

In the presence of both the work function difference and the oxide charge, the total voltage shift in the C-V curve is then
\[ \Delta V_G|_{\phi_s} = \Delta V_{fb} = \phi_{ms} - \frac{1}{C_{ox}} \int_0^{d_{ox}} \frac{x}{d_{ox}} \rho(x) \, dx \] (1.31)

which is illustrated as curve (2) in Fig. 1.4.

In the special case that the oxide charge is a sheet charge, \( Q_{ox} \), located at a distance \( x \) from the metal/SiO$_2$ interface, Eq. (1.30) becomes

\[ \Delta V_{fb} = \Delta V_G|_{\phi_s} = -\frac{x}{d_{ox}} \frac{Q_{ox}}{C_{ox}} \] (1.32)

From Eq. (1.32), one can see that the voltage shift depends not only on the charge density but also on its location; the closer it is to the SiO$_2$/Si interface \( (x \rightarrow d_{ox}) \), the larger the shift. A maximum shift is obtained if the charge is located right at the SiO$_2$/Si interface \( (x = d_{ox}) \), and Eq. (1.32) becomes

\[ \Delta V_{fb} = \Delta V_G|_{\phi_s} = -\frac{Q_{ox}}{C_{ox}} \] (1.33)

In practice, it has been determined experimentally that, in most cases, the fixed oxide charge tends to reside very close to the SiO$_2$/Si interface. Therefore, it is a common practice to assume a sheet charge located right at the SiO$_2$/Si interface, and the effective density of this interfacial sheet charge, \( Q'_{ox} \), can be obtained from Eq. (1.33)

\[ Q'_{ox} = -\Delta V_{fb} C_{ox} = -\Delta V_G|_{\phi_s} \times C_{ox} \] (1.34)

where \( \Delta V_{fb} \) is the part of the flatband voltage shift that is caused solely by the oxide charge.

From Eqs. (1.29) and (1.34), the relationship between the effective oxide charge, \( Q'_{ox} \), and the actual charge distribution can be obtained

\[ Q'_{ox} = \int_0^{d_{ox}} \frac{x}{d_{ox}} \rho(x) \, dx \] (1.35)

Since it is very difficult to determine \( \rho(x) \) experimentally, the quantity \( Q'_{ox} \) is normally used to characterize a MOS device.

**1.2.2.3 Mobile Ions in the Oxide**

Unless extreme care is taken during cleaning, processing, and handling of the MOS device wafers, it is likely that the oxide may be contaminated with trace amounts of alkali ions, in particular sodium (Na$^+$), which are sufficiently mobile under normal operating conditions to cause device stability problems.
The effect of mobile ions on the MOS C-V characteristics can be best illustrated by the bias-temperature stress experiment described below.

Suppose that the mobile ion density is initially distributed in the oxide as $\rho_i(x)$. At any instance of time the effect of the mobile ions on the C-V curve is the same as though they are fixed charges, and therefore one expects, according to Eq. (30)

$$
\Delta V_G |_{\phi_s} = \Delta V_{fb} = -\frac{1}{C_{ox}} \int_0^{d_{ox}} \frac{x}{d_{ox}} \rho_i(x) \, dx 
$$

(1.36)

The corresponding C-V curve is shown as curve (1) in Fig. 1.5.

We now apply a negative voltage on the gate and suppose that the applied $V_G$ is sufficiently large to drive all the mobile ions to move to the metal/SiO$_2$ interface ($x = 0$). From Eq. (1.36) it is evident that the mobile ions in this case behave as though they are not present, and the resulting C-V curve would be to the right of curve (1) in Fig. 1.5, with a flatband voltage designated as $V_{fb}(-)$.

Similarly, if we apply a sufficiently large positive voltage to drive all the mobile ions to the SiO$_2$/Si interface ($x = d_{ox}$), the largest voltage shift of the C-V curve due to the mobile ions will be observed, and the flatband voltage shift becomes

$$
V_{fb}(+) - V_{fb}(-) = -\frac{Q_{mi}}{C_{ox}} 
$$

(1.37)

Fig. 1.5 High-frequency MOS C-V curves for a MOS capacitor containing mobile ions in oxide measured before (curve 1) and after the mobile ions are driven toward the gate metal (curve 2) or driven toward the SiO$_2$/Si interface (curve 3).
where $Q_{mi} = \int_0^{d_{ox}} \rho_i(x) \, dx$ is the areal density of the total mobile ionic charge, and $V_{fb}(-)$ and $V_{fb}(+)$ are the flatband voltages after negative bias stress and positive bias stress, respectively.

Since the transport of mobile ions in $\text{SiO}_2$ is a thermally activated, field-assisted process, the sample is usually heated during the bias stress experiment to ensure that the ions traverse through the oxide within a reasonable time period. This procedure is called the bias-temperature stress cycle, or $B$-$T$ stress in short. Typical conditions for the $B$-$T$ stress are $\varepsilon_{ox} \geq 2 \times 10^6 \text{ V/cm}$, $T \geq 100^\circ \text{C}$, and $t \geq 10 \text{ min}$.

Thus, by performing a $C$-$V$ measurement after a negative $B$-$T$ stress, followed by a positive $B$-$T$ stress and another $C$-$V$ measurement, the total mobile ionic charge per unit area can be determined from Eq. (1.37).

### 1.2.2.4 Interface Traps

Interface traps (also referred to as interface states or surface states, especially in earlier publications) are electronic energy levels located at the $\text{SiO}_2$/Si interface that can capture or emit electrons (or holes). These electronic levels arise because of the lattice mismatch at the interface, disconnected chemical bonds, or impurities. Throughout this book, only those interface traps with energy levels within the semiconductor band gap are considered. Those that lie above the conduction band edge, $E_c$, or below the valence band edge, $E_v$, have fixed charge states (independent of gate voltage) and can be treated as immobile oxide charges in as far as the $C$-$V$ characteristics are concerned.

Unlike the other nonidealities discussed so far, which all cause a parallel shift of the $C$-$V$ curve along the voltage axis, the presence of the interface traps will cause an uneven shift at different surface potentials. This is because the amount of charge trapped at the interface depends on the surface band bending. Just as is true for other electronic energy states in the semiconductor, the occupancy of interface traps is determined by Fermi statistics. At very low temperatures, essentially all the states below the Fermi level are filled with electrons, and above are empty of electrons. At room temperature, while the above statement is no longer strictly valid, it is still a good approximation for most practical purposes, and we shall assume its validity throughout this book unless otherwise specified.

Before we proceed, it is useful to define the two types of interface traps that will be encountered: a donor-like interface trap is one that is neutral when filled with an electron and positively charged when empty; an acceptor-like trap is negatively charged when filled with an electron, and neutral when empty. The same statement may be expressed in terms of Fermi level positions: a donor-like trap is neutral when below $E_F$ and positive when above; an acceptor-like trap is negative when below $E_F$ and neutral when above.

Figure 1.6(a)–(d) serves to illustrate the behavior of the interface traps and their effect on the MOS characteristics for a $p$-type and an $n$-type
Fig. 1.6 Charge state of interface traps (assuming all donor-like) in a MOS capacitor under (a) flatband; (b) accumulation; (c) depletion; (d) inversion.
sample. For the sake of illustration, let us assume these traps are all donor-like with a distribution in energy, $D_{it}(E)$, as depicted. In Fig. 1.6(a), the system is at flatband; the traps above $E_F$ are positively charged, and below are neutral. Figure 1.6(b)–(d) depicts the system in accumulation, depletion, and inversion, respectively. It is clear that the amount of the interface trapped charge (positive for donor-like traps) depends on the band bending. For an $n$-type sample, it is smallest in accumulation, and continues to increase as the device goes to flatband, depletion, and inversion; for a $p$-type sample, it is largest in accumulation, and continues to decrease as the device goes to flatband, depletion, and inversion.

Since a positive sheet charge at the SiO$_2$/Si interface causes a negative shift of the $C$-$V$ curve according to Eq. (1.33), we may write

$$\Delta V_G|_{\phi_s} = -\frac{1}{C_{ox}} Q_{it}(\phi_s)$$  \hspace{1cm} (1.38)

where $Q_{it}(\phi_s)$ is the interface-trapped charge for the given surface band bending $\phi_s$, and for the case that the interface traps are all donor-like,

$$Q_{it}(\phi_s) = q \int_{E_F}^{E_{CB} + q\phi_s} D_{it}(E) \, dE$$  \hspace{1cm} (1.39)

where $E_{CB}$ is the conduction band edge in the bulk, and $E_{CB} + q\phi_s$ is the conduction band edge at the silicon surface.

Figure 1.7 compares $C$-$V$ curves for three MOS capacitors having otherwise identical parameters, except that one is free of interface traps, one contains interface traps that are all donor-like, and one all acceptor-like. One can see that the influence of the donor-like interface traps is to stretch out the curve to the left [see curve (2)], in accordance with Eqs. (1.38) and (1.39).

If the interface traps are acceptor-like instead, then the trapped charge, $Q_{it}(\phi_s)$, will be negative and the voltage shift $\Delta V_G|_{\phi_s}$ in Eq. (1.38) would be in the positive direction. Equation (1.39) now becomes

$$Q_{it}(\phi_s) = -q \int_{E_{VB} + q\phi_s}^{E_F} D_{it}(E) \, dE$$  \hspace{1cm} (1.40)

where $E_{VB}$ is the valence band edge in the bulk, and $E_{VB} + q\phi_s$ is the valence band edge at the silicon surface.

The resulting $C$-$V$ curve for a device containing interface traps that are all acceptor-like is depicted as curve (3) in Fig. 1.7.

In actual devices, there has been strong evidence that the interface traps below the middle of the silicon band gap, $E_{mg}$, are donor-like, and those above are acceptor-like [8–10] (see Section 4.2.1 in Chapter 4 for a more detailed discussion). Therefore, at a gate voltage when the Fermi level
Fig. 1.7 Influence of interface traps on the high-frequency MOS C-V curve: curve 1, with no interface traps; curve 2, with only donor-like interface traps; curve 3, with only acceptor-like interface traps.

Intersects $E_{mg}$ at the silicon surface, called the *midgap voltage*, $V_{mg}$, the charge state of the interface traps is neutral, because all the donor-like traps are filled and all the acceptor-like ones are empty. Thus, the presence of the interface traps has *no effect* on $V_{mg}$, and $\Delta V_{mg}$ is only determined by the oxide charge and work function difference:

$$\Delta V_{mg} = -\frac{Q'_ox}{C_{ox}} + \phi_{ms} \quad (1.41)$$

In contrast, the flatband voltage, $V_{fb}$, and the inversion voltage, $V_i$, are affected by the interface trapped charge as follows:

$$\Delta V_{fb} = \phi_{ms} - \frac{Q'_ox}{C_{ox}} + \frac{q}{C_{ox}} \int_{E_i}^{E_F} D_{it}(E) \, dE \quad \text{for } n\text{-type substrates}$$

$$\Delta V_{fb} = \phi_{ms} - \frac{Q'_ox}{C_{ox}} - \frac{q}{C_{ox}} \int_{E_F}^{E_i} D_{it}(E) \, dE \quad \text{for } p\text{-type substrates} \quad (1.42)$$
\[
\Delta V_i = \phi_{ms} - \frac{Q_{ox}'}{C_{ox}} - \frac{q}{C_{ox}} \int_{E_{EF}}^{E_{iB}-2q\phi_b} D_{it}(E) \, dE \quad \text{for } n\text{-type substrates}
\]
\[
\Delta V_i = \phi_{ms} - \frac{Q_{ox}'}{C_{ox}} + \frac{q}{C_{ox}} \int_{E_{iB}+2q\phi_b}^{E_{EF}} D_{it}(E) \, dE \quad \text{for } p\text{-type substrates}
\]

where \( E_{iB} \) is the intrinsic Fermi level in the bulk.

Note that in Eqs. (1.41)–(1.43) the \( Q_{ox}' \) includes the mobile ionic charge at the instant of measurement.

The presence of the interface traps also affects the low-frequency \( C-V \) curves, which is discussed next.

As illustrated in Fig. 1.6, the charge state of the interface traps depends on the silicon surface potential. Under an AC modulation voltage signal, the surface potential changes back and forth at the modulation frequency. If the modulation frequency is sufficiently low that all interface traps respond to the change in the surface potential, then an additional capacitance component, due to the charging of the interface traps, is added to the semiconductor capacitance. The equivalent circuit corresponding to this situation is shown in Fig. 1.8, where \( C_{it} \) is the capacitance due to the interface traps.

By definition, the capacitance due to the charging of the interface traps is

\[
C_{it}(\phi_s) = \frac{dQ_{it}(\phi_s)}{d\phi_s} = qD_{it}(\phi_s)
\]

The total low frequency capacitance of the MOS system, including the effect of the interface traps, is now

---

**Fig. 1.8** Equivalent circuit representation of an MOS capacitor with effect of interface traps included.
\[ C(\phi_s) = \frac{C_{ox}[C_{sc}(\phi_s) + C_{it}(\phi_s)]}{C_{ox} + [C_{sc}(\phi_s) + C_{it}(\phi_s)]} \]  

(1.45)

Note that at a given \( \phi_s \) the presence of the interface traps causes an increase in the low frequency capacitance.

Depending on the device being measured, it is often necessary to use a frequency as low as a fraction of 1 Hz to satisfy the low-frequency requirement, which is difficult to do experimentally. Therefore, in practice, an alternative method, the quasi-static technique [6, 7] is more commonly used. Since this technique represents a slight departure from the usual AC capacitance measurements (the actual quantity measured is the displacement current, not the capacitance), and it is widely used, we thought it useful to describe briefly its basic principles here.

As shown in Fig. 1.9, the MOS capacitor to be measured is connected to a linear voltage ramp source and a sensitive current meter (an electrometer or a pico-ammeter is often used). The voltage output of the ramp source can be expressed as

\[ V = V_0 + rt \]  

(1.46)

where \( V_0 \) is the voltage at \( t = 0 \), and \( r \) is the ramp rate, typically in the range 5–50 mV/s. In response to the change in voltage, a displacement current flows in the MOS capacitor, which is

\[ i = \frac{dQ}{dt} \]  

(1.47)

Using the chain rule, we can write Eq. (1.47) as

\[ i = \frac{dQ}{dt} = \frac{dQ}{dv} \frac{dv}{dt} = rC \]  

(1.48)

where \( dQ/dv = C \) is the capacitance of the MOS capacitor, and \( dv/dt = r \) comes from Eq. (1.46).

![Fig. 1.9 Basic elements of the quasi-static C-V measurement apparatus.](image)
From Eq. (1.48) it is evident that the displacement current measured by the current meter is directly proportional to the capacitance of the MOS capacitor. Since there is no AC modulation signal involved, the capacitance corresponds to the limiting case of $f = 0$.

Combining the high- and low-frequency $C-V$ curves, we can determine the interface trap density over much of the semiconductor band gap by using Eqs. (1.44) and (1.45).

1.2.3 Physics of MOSFETs

Physically, the MOSFET is essentially a MOS capacitor with two $p-n$ junctions connected through a gate-controlled conduction channel. To establish the basic principles of MOSFET operations, let us use as an example an $n$-channel enhancement mode device whose cross-section was shown in Fig. 1.1.

1.2.3.1 Qualitative Description of MOSFET Operation

When the gate is biased such that the silicon surface under the gate is in accumulation or depletion ($V_G < V_t$, where $V_t = V_t$ is the threshold voltage), the channel region contains very few electrons, and no appreciable current can flow between the source and the drain. When $V_G$ is biased so that inversion occurs ($V_G \geq V_t$), a significant number of mobile electrons appear adjacent to the Si surface under the gate. Now a conduction channel connecting the source and drain is formed, as depicted in Fig. 1.10(a). The number of electrons piled up at the Si surface, and therefore the channel conductance, increases with $V_G$. Thus, the MOSFET structure allows one to modulate the source-drain conductance by the voltage applied to the third terminal – the gate.

Now let us examine the effect of the drain bias after the inversion channel is formed. For a small $V_D$ applied to the drain terminal, the conduction channel behaves just like a simple resistor, and the drain current will be directly proportional to $V_D$. This is the linear $I-V$ region, and is shown as the line from the origin to point A in Fig. 1.11. As one keeps increasing $V_D$, the potential drop along the channel due to the channel current starts to negate the inverting effect of the gate. As depicted in Fig. 1.10(b), the depletion region widens in going down the channel from the source to the drain, and the number of channel electrons decreases correspondingly. Thus, the channel conductance ($dI/dV$) decreases with increasing $V_D$, which is reflected as a decrease in the slope of the $I-V$ characteristic, as shown in the $A \rightarrow B$ portion of the curve in Fig. 1.11. The greatest increase in the depletion layer width, and therefore the greatest decrease in the number of channel electrons, occurs near the vicinity of the drain junction, and eventually for a sufficiently large drain voltage ($V_D \geq V_{D_{sat}}$), the inversion layer completely vanishes near the drain [Fig. 1.10(c)]. This condition is
Fig. 1.10 Schematic diagram of an n-channel MOSFET showing the conduction channel after the device is turned on: (a) \( V_D = 0 \); (b) \( V_{D_{sat}} > V_D > 0 \); (c) \( V_D = V_{D_{sat}} \); (d) \( V_D > V_{D_{sat}} \).

referred to as “pinch-off” because the normal conduction channel disappears adjacent to the drain. When the channel pinches off, current saturation is observed; that is, the slope of the I-V characteristic becomes approximately zero, as shown beyond point B in Fig. 1.11.

For drain voltages greater than the pinch-off voltage, \( V_{D_{sat}} \), the pinched-off portion of the channel widens from just a point into a depleted channel section, \( \Delta L \), as shown in Fig. 1.10(d), and the drain voltage in excess of \( V_{D_{sat}} \) is absorbed almost entirely across \( \Delta L \). If \( \Delta L \ll L \), the effective conduction channel from the source to the pinched-off region is essentially
unaffected by the excess drain voltage, and the channel current stays approximately constant. However, in the case where $\Delta L$ is comparable to $L$, the effective conduction channel decreases substantially with increasing $V_D$ and the current will increase monotonically with $V_D \geq V_{D_{sat}}$. This effect is called channel length modulation and is often observed in modern-day short channel devices.

The $I-V$ curve shown in Fig. 1.11 is for one gate voltage only. If the gate voltage is changed in a stepwise fashion, the familiar family of curves will be obtained (Fig. 1.12).

---

**Fig. 1.11** MOSFET channel current as a function of drain-source voltage at a given gate voltage after turn-on.

**Fig. 1.12** MOSFET channel current as a function of drain-source voltage for several different gate voltages.
1.2.3.2 Derivation of I-V Characteristics

To derive the quantitative $I_D$-$V_D$ characteristic, we assume an $n$-channel enhancement mode device with a long channel and that the diffusion current is negligible compared to the drift current. When $V_G \geq V_t$, the channel is turned "on" and the current density in the channel is (the $x$, $y$, and $z$ coordinates are as defined in Fig. 1.1)

$$J_n = -q \mu_n n \xi_y = q \mu_n n \frac{dV}{dy} \quad (1.49)$$

Integrating the current density over the cross-sectional area of the channel at an arbitrary point $y$ gives

$$I_D = \int_0^{x_c} \int_0^W J_n \, dx \, dz = W \int_0^{x_c(y)} J_n \, dx = q W \frac{dV}{dy} \int_0^{x_c} \mu_n(x, y)n(x, y) \, dx \quad (1.50)$$

where $x_c$ is the channel depth.

Note that both the electron mobility, $\mu_n$, and the electron density, $n$, are functions of $x$ and $y$. The electron mobility in the channel is determined by various scattering mechanisms due to phonons, impurities, and surface roughness; the surface roughness scattering depends strongly on the electric field and the distance of the electron to the surface, causing $\mu_n$ to be a function of both $x$ and $y$.

Using the standard averaging procedure, we may define an effective mobility of carriers at a distance $y$ from the source as

$$\bar{\mu}_n = \frac{\int_0^{x_c} \mu_n(x, y)n(x, y) \, dx}{\int_0^{x_c} n(x, y) \, dx} = -q \frac{1}{Q_n(y)} \int_0^{x_c} \mu_n(x, y)n(x, y) \, dx \quad (1.51)$$

where

$$Q_n(y) = -q \int_0^{x_c} n(x, y) \, dx \quad (1.52)$$

is the total carrier charge per unit area in the channel.

Thus, Eq. (1.50) can be simplified as

$$I_D = -W \bar{\mu}_n Q_n \frac{dV}{dy} \quad (1.53)$$

The current flowing in the channel, $I_D$, is independent of $y$ due to the current continuity requirement. If we assume that $\bar{\mu}_n$ is constant throughout the channel, we may rearrange Eq. (1.53) and perform the integration:

$$\int_0^l I_D \, dy = I_D L = -W \bar{\mu}_n \int_0^{V_D} Q_n \, dV \quad (1.54)$$

or
\[ I_D = - \frac{W}{L} \bar{\mu}_n \int_0^{V_D} Q_n \, dV \]  

(1.55)

To proceed further, we need to find the function \( Q_n(V) \). For the special case \( V_D = 0 \), the carrier charge density per unit area in strong inversion, \( Q_n \), can be obtained from analysis of a simple MOS capacitor:

\[ Q_n = -C_{ox}(V_G - V_t) \]  

(1.56)

For \( V_D \neq 0 \), the carrier density will be a function of \( y \) due to the potential distribution in the channel. At point \( y \) where the potential due to \( V_D \) is \( V(y) \), the carrier charge density per unit area is approximately

\[ Q_n(y) = -C_{ox}[V_G - V_t - V(y)] \]  

(1.57)

Substituting Eq. (1.57) into (1.55) and integrating, we obtain the \( I_D - V_D \) relationship

\[ I_D = \frac{W}{L} \bar{\mu}_n C_{ox} \left[ (V_G - V_t)V_D - \frac{V_D^2}{2} \right] \quad \text{for} \quad V_{D_{sat}} \geq V_D \geq 0 \]  

(1.58)

The post-pinch-off portion of the characteristic is approximately modeled by setting \( I_D|_{V_D=V_{D_{sat}}} = I_{D_{sat}} \), where \( V_{D_{sat}} \) corresponds to the drain voltage which causes \( I_D \) in Eq. (1.58) to be maximum.

Thus

\[ \frac{\delta I_D}{\delta V_D}|_{V_D=V_{D_{sat}}} = 0, \quad \text{which upon solution gives} \quad V_{D_{sat}} = V_G - V_t \]  

(1.59)

From Eqs. (1.58) and (1.59)

\[ I_{D_{sat}} = \frac{W}{L} \bar{\mu}_n C_{ox}(V_G - V_t)^2 \]  

(1.60)

While Eqs. (1.58) and (1.60) describe the first-order behavior of the \( I-V \) characteristics of a long-channel enhancement mode MOSFET, some approximations used in the derivation may cause an overestimate of the current magnitude, especially for high-drain voltages. The major factor that contributes to the error is the assumption that the depletion depth for all channel points from the source to the drain remains fixed for \( V_D \neq 0 \). In reality, as shown in Fig. 1.10(b)–(d), the depletion depth increases in progressing from the source to the drain when \( V_D \neq 0 \). Therefore, Eq. (1.58) must be modified to take this into account. The more accurate \( I-V \) characteristics including this effect have been worked out and can be found in Refs. 3 and 5.
1.2.3.3 Threshold Voltage Relationships

The threshold voltage, \( V_t \), is one of the most important parameters from an operational point of view. As discussed previously, the threshold voltage is the gate voltage when \( \phi_s = 2\phi_b \), or when strong inversion starts. For an ideal \( n \)-channel enhancement mode MOSFET, Eq. (1.26) applies, and we have

\[
V_t = V_i = 2\phi_b + \frac{2(qN_Ae\phi_b)^{1/2}}{C_{ox}}
\]  

(1.61)

Equation (1.61) suggests that \( V_t \) depends on the oxide thickness and the channel doping concentration. Since both terms in Eq. (1.61) are positive quantities, \( V_t \) for an ideal \( n \)-channel device is positive, as expected intuitively.

The threshold voltage for a practical device must include the effects of work function difference, oxide and interface charges, and may be expressed as

\[
V_t = 2\phi_b + \frac{2(qN_Ae\phi_b)^{1/2}}{C_{ox}} + \phi_{ms} - \frac{Q'_{ox}}{C_{ox}} - \frac{Q_{it}}{C_{ox}}
\]  

(1.62)

In Eq. (1.62), the terms \( \phi_{ms} \) and \(-Q'_{ox}/C_{ox}\) are typically negative quantities, while the term \(-Q_{it}/C_{ox}\) is typically positive for a \( n \)-channel device (due to charged acceptor-like interface traps, which make \( Q_{it} \) negative). Hence, depending on the balance of the positive and negative terms, it is possible to end up with a negative \( V_t \); that is, the device is “on” even at \( V_G = 0 \). As will be shown in subsequent chapters, one of the most significant consequences of the ionizing radiation effects is the shift of \( V_t \) due to the radiation-induced oxide and interface charges.

In addition to the terms discussed above, the threshold voltage can be changed by biasing the substrate contact relative to the source. In Fig. 1.1, if a negative bias, \( V_{sb} \), is applied between the substrate terminal and the source, then the surface potential must reach \( 2\phi_b + V_{sb} \), an additional band bending of \( V_{sb} \), before the surface inverts. Thus, the threshold voltage is modified from Eq. (1.62)

\[
V_t = 2\phi_b + \frac{[2qN_Ae(2\phi_b + V_{sb})]^{1/2}}{C_{ox}} + \phi_{ms} - \frac{Q'_{ox}}{C_{ox}} - \frac{Q_{it}}{C_{ox}}
\]  

(1.63)

The change of the threshold voltage resulting from the substrate bias is called the substrate bias effect, or the “body effect.” Some circuits are designed to make use of this additional control to place the threshold voltage in a more desirable range. Sometimes, however, the body effect may be introduced unintentionally due to the operation of adjacent devices, and it must be taken into account to ensure proper functioning of the circuit.
1.2.3.4 Transconductance

Another important parameter that characterizes a MOSFET is the transconductance, $g_m$, which is defined as

$$ g_m = \frac{\delta I_D}{\delta V_G} \quad (1.64) $$

From Eqs. (1.58) and (1.64)

$$ g_m = \bar{\mu}_n C_{ox} \frac{W}{L} V_D \quad \text{for} \quad V_D < V_{D_{sat}} \quad (1.65) $$

Assuming the effective mobility of the carriers is independent of $V_G$, it is apparent from Eq. (1.65) that $g_m$ increases linearly with $V_D$ but is independent of $V_G$. For a given device geometry at a given $V_D$, $g_m$ is proportional to $\bar{\mu}_n$. Since, among other things, $\bar{\mu}_n$ depends on the carrier scattering due to the oxide and interface charges, one of the important consequences of ionizing radiation effects is the reduction of the transconductance.

When $V_D \geq V_{D_{sat}}$, the transconductance is given by manipulating Eqs. (1.60) and (1.64) with $V_D = V_{D_{sat}}$

$$ g_{msat} = 2\bar{\mu}_n C_{ox} \frac{W}{L} (V_G - V_i) \quad (1.66) $$

Thus, $g_{msat}$ is independent of $V_D$ but linearly proportional to $V_G$.

In the above discussion, we assumed $\bar{\mu}_n$ to be independent of $V_G$. In reality, because the carrier scattering is a function of $V_G$, it may be necessary to modify Eqs. (1.65) and (1.66) to more accurately model the transconductance behavior. The physics involved in the various scattering processes for channel carriers has been extensively studied [11, 12] and will not be covered herein.

1.2.3.5 Subthreshold Characteristics

The theories for the drain current that we have considered so far assume that the channel does not contain any free carriers for $V_G < V_i$, and the current drops abruptly to zero once $V_G$ is below $V_i$. Experimentally, however, a small but finite current does flow through the drain at gate voltages below the threshold voltage. This current is called the subthreshold current, and it arises from the finite density of minority carriers in the channel at $V_G < V_i$.

Referring to the band diagram for the MOS capacitor shown in Fig. 1.2, one realizes that, for the $n$-channel device in the region $V_i > V_G > 0$, the majority carriers are depleted from the surface but the minority carriers are rising with $V_G$ so that the equilibrium condition $p \times n = n_i^2$ is maintained. When $V_G$ is small such that the Fermi level at the silicon surface lies in the lower half of the band gap, $p_s > n_s$, and the minority carrier density at the
surface is lower than the intrinsic carrier density, \( n_i < n_i \), which is negligibly small at room temperature. When a sufficiently large \( V_G \) is applied such that the Fermi level at the surface lies in the upper half of the band gap, however, the minority carrier density at the surface exceeds the majority carrier density there, \( n_s > p_s \), and we enter a regime called weak inversion before strong inversion is reached.

Because of the low carrier concentration, in weak inversion the drain current is dominated by diffusion, and it can be shown [3] that the \( I_D - V_D \) relationship can be approximated by

\[
I_D = \mu_n \left( \frac{W}{L} \right) a C_{ox} \left( \frac{n_i}{\beta N_A} \right)^2 \exp(\beta \phi_s)(\beta \phi_s)^{-1/2}[1 - \exp(-\beta V_D)]
\]  

(1.67)

where

\[
a = \frac{\sqrt{2}(\epsilon_i/L_D)}{C_{ox}}
\]  

(1.68)

and

\[
\phi_s = (V_G - V_{tb}) - \frac{a^2}{2} \left( \frac{1}{\beta} \right) \left[ 1 + \frac{4}{a^2} (\beta V_G - B V_{tb} - 1) \right]^{1/2} - 1
\]  

(1.69)

Equations (1.67)–(1.69) indicate that in the subthreshold regime the drain current varies exponentially with \( V_G \) and becomes substantially independent of \( V_D \) for \( V_D > 3kT/q \). A typical set of subthreshold characteristics is shown in Fig. 1.13 [13]. As it will be shown later in this subsection,

---

**Fig. 1.13** Experimental subthreshold characteristics at two drain voltages and three substrate biases. (After Troutman [13], © 1974 IEEE. Reprinted with permission.)
the subthreshold current could become a strong function of $V_D$ for short-channel devices.

A useful parameter in the subthreshold regime is the gate voltage swing, $S$, needed to reduce the current by one decade

$$S = \ln 10 \left[ \frac{dV_G}{d(\ln I_D)} \right] = \frac{kT}{q} \ln 10 \left[ 1 + \frac{C_{sc}(\phi_s)}{C_{ox}} \right] \left[ 1 - \left( \frac{2}{a^2} \right) \left( \frac{C_{sc}(\phi_s)}{C_{ox}} \right)^2 \right]^{-1}$$

(1.70)

In the case $a \gg C_{sc}/C_{ox}$, Eq. (1.70) reduces to

$$S = \frac{kT}{q} \ln 10 \left( 1 + \frac{C_{sc}}{C_{ox}} \right)$$

(1.71)

As might be expected, the parameters that cause the $I-V$ characteristics above the threshold to deviate from the ideal case also affect the sub-threshold behavior. The effect of the work function difference and the oxide charge is to shift the $I_D-V_G$ characteristic curve along the $V_G$ axis by an amount equal to that in the post-threshold regime. The interface traps, however, will cause a change in the slope of the $I_D-V_G$ curve, and thus the subthreshold swing.

Taking into account the interface traps, we modify the subthreshold swing as

$$S_{it} = S_0 \left[ 1 + \frac{C_{it}(\phi_s)}{C_{ox} + C_{sc}(\phi_s)} \right]$$

(1.72)

where $S_{it}$ and $S_0$ represent the quantity with or without interface traps, respectively, and $C_{it}(\phi_s) = qD_{it}(\phi_s)$ is the capacitance associated with the interface traps.

### 1.2.3.6 Short Channel Effects

The theories developed so far have been based on the long-channel model; that is, the channel length is assumed to be much longer than the depletion regions. If the channel length is comparable to the depletion layer widths of the source and drain junctions, a number of device parameters will deviate appreciably from those predicted by the long-channel theories. Of particular concern is the reduction of the threshold voltage. This is primarily due to the effect of the space charge from the source and drain regions extending into the channel region, causing a two-dimensional potential distribution, and charge sharing between the channel and source-drain regions. Hence, the amount of charge reflected to the gate electrode for a given silicon surface potential is reduced, resulting in a decrease of the threshold voltage.

In addition, the subthreshold characteristics are also different for short-channel devices. In contrast to the long-channel device, it has been demon-
strated both theoretically and experimentally that the subthreshold current for a short-channel device is a strong function of \( V_D \), due to the reduction of the effective channel length as \( V_D \) increases. For a given \( V_D \), the magnitude of the subthreshold current increases monotonically with decreasing channel length [3].

1.2.3.7 CMOS

While integrated circuits based on either \( n \)-channel or \( p \)-channel MOSFETs are suitable for various applications, of which the \( n \)-channel ones have the advantage of higher speed, the complementary MOS (CMOS) structure is rapidly becoming the dominant technology as the trend toward an even higher degree of integration continues. The most recognized advantage of a CMOS device compared to its \( n \)-channel or \( p \)-channel counterpart is its low standby power dissipation.

Basically, a CMOS structure consists of a \( p \)-channel MOSFET adjacent to a \( n \)-channel MOSFET, as shown in Fig. 1.14. To understand the reason for its low power dissipation, let us consider a simple inverter, based on which many complicated digital circuits may be built.

The basic CMOS inverter circuit and its voltage-transfer characteristics are sketched in Fig. 1.15, where the \( n \)-channel device has a positive threshold voltage \( (V_{th n} > 0) \) and the \( p \)-channel device has a negative threshold voltage. When the input voltage, \( V_i \), is “high,” the \( n \)-channel device is in its “on” state and the \( p \)-channel device is in its “off” state. Therefore, the channel resistance of the \( p \)-channel device is much higher than that of the \( n \)-channel device, which causes \( V_{DD} \) to drop mostly across the \( p \)-channel, and the output voltage, \( V_o \), is “low.” If the input voltage is now “low,” the \( n \)-channel device is “off,” and \( V_o \) assumes most of \( V_{DD} \), which represents the “high” state. Thus, in either of the two states, one or

---

Fig. 1.14 Schematic cross-sectional view of a CMOS inverter.
Fig. 1.15 (a) Basic CMOS inverter circuit. (b) Voltage-transfer characteristics of a CMOS inverter.
the other of the MOSFETs is always turned off and there is no DC path to carry current from the supply except for junction leakages. For this reason, the only time when significant power dissipation in CMOS circuits takes place is during switching transients.

A potential problem for the operation of a CMOS circuit is an undesirable condition known as latchup. When latchup happens the parasitic bipolar-transistor action causes a low-resistance path between the power supply and ground.

While a detailed treatment of the latchup mechanisms is beyond the scope of this chapter, it is nonetheless still useful to present a brief description here so that related materials in Chapter 9 may be more easily understood. As shown in Fig. 1.16(a, b) there are two parasitic bipolar transistors built in the CMOS structure, one p-n-p and one n-p-n. These two transistors are cross-connected so that the base-collector junctions are

---

**Fig. 1.16** (a) Cross-sectional view of a p-well CMOS inverter with parasitic p-n-p and n-p-n transistors indicated. The emitter contact resistances ($R_{pe}$ and $R_{ne}$), and the resistances in the substrate region ($R_s$) and in the p-well region ($R_w$) are also indicated. (b) Schematic representation of the cross-coupled parasitic transistors shown in (a).
common. Under active bias, the $p-n-p$ collector delivers current to the $n-p-n$ base, and the $n-p-n$ collector delivers current to the $p-n-p$ base. If the gains of these bipolar transistors are sufficiently large, both could be driven into saturation and the supply voltages become connected through a low-resistance path.

In general, latchup should not occur under normal operating conditions, because the base–emitter junctions for both parasitic bipolar transistors are zero-biased. However, if the CMOS structure is not designed to be latchup free, it is possible to initiate a latchup. Of particular interest to this book is ionizing-radiation-induced latchup, which will be discussed in detail in Chapter 9.

### 1.3 PHENOMENOLOGICAL DESCRIPTION OF IONIZING RADIATION EFFECTS

Ionizing radiation is that which possesses enough energy to break atomic bonds and create electron/hole pairs (i.e., cause ionization) in the materials of interest, which in the case of MOS devices are primarily silicon dioxide and silicon. This radiation may be in the form of photons with energies greater than the bandgap of the material of concern (1.1 eV for silicon; 9 eV for silicon dioxide), or in the form of particles such as electrons, protons, or atomic ions.

The effects due to these bond-breaking events are to be contrasted with those that can occur when atoms in a material structure are displaced from their original positions by the radiation (displacement damage). The typical cause of significant displacement damage is neutrons from a nuclear reactor or weapon burst. Displacement damage primarily causes a reduction in minority carrier lifetime in the silicon substrate; this in turn can have a severe adverse effect on the gain of bipolar devices. Since the properties of most MOS devices are not significantly affected by minority carrier lifetime, they are relatively insensitive to the damage caused by displacement events. Although the increased junction leakage currents that can occur in MOS devices are observable and at small geometries the effects of displacements caused by a single particle can be measured (see Chapter 9), in general displacement damage in MOS devices is of secondary concern. Therefore this book is devoted solely to the effects of ionizing radiation on MOS devices. For discussions of displacement damage effects in semiconductor devices, there are several books and reviews available (e.g., Refs. 14–16).

#### 1.3.1 Total Dose Effects

An MOS device exposed to an ionizing radiation environment typically suffers degradation in one or more of its performance parameters. MOS transistors experience a shift in threshold voltage and a decrease in gain, and
integrated circuits may slow down, show higher leakage currents, or even cease to function properly. These changes may not be constant with time after an irradiation and may depend on the dose rate at which the radiation is received. The damage responsible for these total dose effects occurs in the insulator (silicon dioxide) layers of the circuit structures. The following paragraphs will provide a phenomenological description and overview of these effects.

The radiation damage in the silicon dioxide layers consists of three components: the buildup of trapped charge in the oxide, an increase in the number of interface traps, and an increase in the number of bulk oxide traps. The possible processes by which the radiation creates these charge and trap sites are illustrated in Fig. 1.17. Electrons and holes are created within the silicon dioxide by the ionizing radiation or may be injected into the SiO$_2$ by internal photoemission from the contacts. These carriers can recombine within the oxide or transport through the oxide. Electrons are very mobile in SiO$_2$ and quickly move to the contacts; in contrast the holes have a very low effective mobility and transport via a complicated stochastic trap-hopping process. Some of the holes may be trapped within the oxide, leading to a net positive charge. Others may move to the SiO$_2$/Si interface, where they capture electrons and create an interface trap. Along with the electron-hole generation process, chemical bonds in the SiO$_2$ structure may be broken. Some of these bonds may reform when the electrons and holes recombine, whereas others may remain broken and give rise to electrically active defects. These defects can then serve as trap sites for carriers or as interface traps. Bonds associated with hydrogen or hydroxyl groups when broken can release these impurities, which are then mobile within the silicon dioxide. These impurities may then migrate to the SiO$_2$/Si interface, where they undergo a reaction which results in an interface trap. The defects created by the radiation may themselves migrate in the strained region near the SiO$_2$/Si interface and also result in the formation of an interface trap.

Since the number of electron/hole pairs generated is directly proportional to the amount of energy absorbed by the device material, the total damage (in the form of trapped charge or interface traps) is also roughly proportional to the total dose of radiation received by the device. The total dose is usually measured in units of rads(Si) or rads (SiO$_2$), a unit equal to 100 ergs absorbed per gram of the material specified.

Typically the net charge trapped in the oxide layer after irradiation is positive. The interface traps can exchange charge freely with the silicon substrate, and thus their charge state depends upon the bias applied to the device—more negative for a positive bias applied to the gate electrode than for a negative bias applied to the gate electrode. As the total dose to the device increases, the amount of oxide-trapped charge and number of interface traps monotonically increase. The radiation hardness of a device is determined by the rate at which these two damage measures build up as the total radiation dose increases. For radiation-hardened devices, the number
Fig. 1.17 Schematic diagram illustrating the possible processes by which ionizing radiation in an MOS device leads to the creation of oxide charge, neutral traps, and interface traps.
of radiation-generated holes trapped within the oxide layer may be less than a few percent of the number created within the oxide layer; the number of radiation-induced interface traps may also be less than a few percent of the number of electron/hole pairs generated within the oxide layer. For radiation-soft devices, the percentage of trapped holes can be greater than 50%; the number of radiation-induced interface traps may also increase, although typically not to as large a degree.

The effect of these radiation-induced charge components on the characteristics of an MOS capacitor is as described above in Section 1.2. More specifically, the oxide-trapped charge shifts the C-V curve in the negative direction. The interface traps tend to "stretch out" the C-V curve, so that a greater change in applied bias voltage is required to cause the same change in capacitance as before the irradiation (see Figs. 1.4 and 1.7).

These same effects occur in MOS transistors. In Fig. 1.18 is plotted the drain current of an n-channel MOS transistor vs. gate voltage before and after the device is irradiated. The curve shifts in the negative direction, just as in the case of the MOS capacitor. This means that the threshold voltage of the transistor is more negative, or that a less positive voltage is required to turn the transistor on. The curve is also less steep; that is, a greater change in applied bias is required to cause the same change in current as before the radiation. This effect is most pronounced in the subthreshold region of the curve and is again an effect similar to that in the MOS capacitor. The last point to note is that the total current passed through the device at a given voltage above the threshold is reduced; the gain of the MOS transistor is decreased by the radiation. The radiation-induced interface traps cause a degradation in mobility of the carriers in the channel of the MOS transistor; this leads to a reduction in channel conductance and transconductance for the transistor, and thus a decrease in gain (see Section

![Diagram](image)

**Fig. 1.18** Representative plot of the logarithm of the drain current of an n-channel MOS transistor as a function of gate voltage before and after irradiation.
1.2.3). These same types of change also occur for parasitic MOS elements present in integrated circuits, for example, the field oxide regions which may be used to isolate active MOS devices in a circuit.

The magnitude of the above changes depends on a number of factors: the total radiation dose and its energy; the rate at which it is delivered; the bias applied and the temperature during irradiation; the type of transistor; the length of time, the bias, and the temperature after irradiation; and the relative radiation hardness of the devices. Figure 1.19 illustrates the kinds of changes that can occur in threshold voltage for \( n \)-channel and \( p \)-channel transistors biased either "on" or "off" during irradiation; note that under certain bias and radiation conditions, the \( n \)-channel transistor threshold can shift to values more positive than its value before irradiation. In an integrated circuit, it is highly probable that individual devices will be under different biases during an irradiation and thus that similar devices which have the same threshold before irradiation may have very different threshold voltages after irradiation.

These changes in the properties of the transistors in an MOS integrated circuit can lead to profound changes in the circuit's characteristics, some of which may be very difficult to predict without extensive circuit simulations. The most drastic of these is that the circuit can completely cease to function. A more insidious change is that where the circuit performs some operations properly and fails to operate correctly for other operations. These catastrophic changes are usually presaged by changes in the values of various circuit parameters, such as standby power supply current, critical path delays or other timing parameters, TTL levels, output drives, etc. The details of how a given circuit will respond to a particular radiation environment depend on the specifics of that circuit's design and the exact response

Fig. 1.19 Representative plot of the threshold voltages of \( n \)- and \( p \)-channel MOS transistors as a function of radiation dose, for transistors biased "on" or "off" during the irradiation.
of the MOS devices within that circuit to the radiation. An example of the changes which might occur in standby power supply current and maximum functional frequency as a function of total dose received for a microprocessor circuit is shown in Fig. 1.20.

The above-described changes in MOS properties are not stable with time after irradiation. Either annealing of damage or increases in damage can occur, as can changes in the net charge trapped in the oxide. These
time-dependent phenomena are dependent on the temperature and bias applied to the device, and on whether or not the radiation source is still present during the “anneal” period. Figure 1.21 shows the effects on n-channel threshold voltage during and after irradiation for two devices processed differently and annealed under +10 V bias at 100°C following irradiation. The threshold voltage changes markedly after the radiation ceases, and can reach values appreciably above the preirradiation threshold (this phenomenon is known as rebound or superrecovery; see Chapter 5).

Figure 1.22 illustrates the effects a similar bias/anneal can have on an integrated circuit. The standby current decreases during the anneal period, but the maximum functional frequency also decreases, and the circuit even fails to function properly for a period of time during the anneal. These postirradiation effects and annealing phenomena can greatly complicate the characterization of integrated circuits for a radiation environment, since they cause the measured device response to depend sensitively on the dose rate at which the radiation is received and on the temperature and bias applied both during and after the irradiation.

Radiation-hardened circuits are generally considered to be those that pass a given set of specifications under all conditions of dose rate, bias, temperature, etc., for all doses greater than $10^5$ rads(Si). Radiation-soft circuits are typically considered to be those that fail a given set of specifications at $10^4$ rads(Si) or less.

### 1.3.2 Transient Radiation Effects

Transient radiation effects are those that depend on the rate at which an ionizing dose is delivered to a device or circuit rather than on the total
amount of dose delivered. The higher the dose rate, the more rapidly electron/hole pairs are created, both in the insulator layer and in the silicon substrate. These carriers can be separated by electric fields present in various regions of the device structure and cause electrical currents in the device and circuit. The magnitude of the current created within a given circuit structure depends on the dose rate, increasing for increasing dose rate. At high enough dose rates (greater than about $10^6$–$10^7$ rads/s), these currents may be large enough to affect the operation of the circuit.

One classical source of high-dose-rate radiation to be considered is that from a nuclear weapon. Dose rates in excess of $10^{12}$ rads/s are possible, which can lead to very large (greater than several amperes) radiation-induced currents in a circuit. This source irradiates the circuit uniformly, so that the “photocurrents” are generated globally throughout the circuit structures. Another source of high-dose-rate radiation is high-energy charged particles such as are encountered in a space environment (the cosmic ray spectrum) or are the decay products of residual radioactive elements present in the material used for IC packages. As an energetic ion passes through a device structure, it creates a high-density track of electron/hole pairs. These then can create a relatively large current in a localized region of the circuit.

The currents created by either of these environments may be larger than the currents seen by logic elements in the circuit under normal device operation. Thus they can overwhelm those normal currents or can cause large voltage drops along circuit interconnections, either of which can cause faulty operation. In particular, the contents of memory elements may be altered, or the logic operations of the circuit may be affected in such a way as to cause an error in the logic function being performed. This corruption of stored information or proper circuit operation, without permanent damage to the circuit, is called a soft error. When it is caused by a burst of ionizing radiation from an extended source, it is generally referred to as a transient upset; when caused by a single energetic ion, it is generally referred to as a single event upset (SEU).

The currents generated by high-dose-rate environments can also lead to hard errors, where the circuit is permanently damaged. This damage typically occurs when some element of the circuit is triggered into a very high current state—high enough that junction integrity may be destroyed or metalization lines may fail by electromigration.

The two elements most likely to be induced into this high current state are the following. In complementary MOS (CMOS) circuits (which contain both n-channel and p-channel devices) on bulk silicon substrates, there exist $p-n-p-n$ structures which form a parasitic SCR (semiconductor-controlled rectifier), as described in Section 1.2.3.7. Under normal conditions, this parasitic device is in its “off” condition, a high-resistance state. When a triggering current is applied, as by an ionizing radiation event, the SCR can be turned on, a low-resistance state where very large currents can flow at
low voltages (1–2 V). This situation is known as latchup. The effect is shown in Fig. 1.23, where the standby power supply current of a CMOS circuit is plotted as a function of voltage for the two possible states of the parasitic SCR. An MOS transistor itself can be placed into a similar high-current, relatively low-voltage state called snapback (see Chapter 9) by an appropriate trigger current; whether or not the holding voltage for this state (the minimum voltage that can sustain the state) is higher than the power supply voltage depends on the type of transistor and on the processing parameters used in the manufacture of the circuit. Typically the only way to remove these devices from the damaging high-current state is by removing power from them in the case of latchup or by turning off the transistor in the case of snapback.

In summary, total dose effects relate primarily to changes in device operational characteristics and performance caused by radiation-induced damage to the oxide layers in MOS circuits. Transient radiation effects are primarily the result of photocurrents generated in the substrate by high-dose-rate radiation. They can lead to momentary upset of circuit operation or corruption of stored data or to permanent catastrophic damage to the circuit.

1.4 BOOK OVERVIEW

Given the wide applications of studies of ionizing radiation effects on MOS devices, there have been a large number of investigations in this field and significant progress has been made over the last two decades. However, knowledge about these developments could be gathered only from original research articles scattered in scientific journals and conference proceedings. Although review articles have been published from time to time, their scope...
has usually been fairly limited, and they have often been written at an advanced level. This book is intended to provide a comprehensive account of ionizing radiation effects in MOS device structures and to cover introductory material as well as advanced developments.

Each chapter is a critical review of and guide to the literature in a particular subject area. Enough background material is included so that the non-expert can understand the principles and key features of the subject, yet enough detail is included so that the literature results may be put in perspective. The chapters do not represent a series of journal articles, but rather each provides an up-to-date overview of an area with references to the original articles for detailed derivations and extensive experimental results. Each chapter is relatively self-contained, so a reader desiring information in a particular area may focus on the relevant chapter without regard to the other chapters. The extensive bibliography in each chapter will allow the interested reader to pursue a given area in as much detail as desired, with the overall perspective and context provided by an expert in the field. Bibliographic references include the title of the work to aid in determining those articles which are of most interest.

Chapter 2 provides a historical perspective of the problem of radiation effects in MOS devices and details the course of hardened technology development. It begins with the discovery of total dose damage in the Telstar 1 satellite after a high-altitude nuclear weapon test. Early studies are recounted, along with the early attempts at producing hardened MOS devices and circuits. The chronological evolution of hardening techniques is described. The chapter concludes with some projections for future hardening efforts, taking into account the trends underway for MOS integrated circuit technology.

Chapters 3 through 8 are devoted primarily to the phenomena of total dose damage from ionizing radiation in MOS devices. Topics covered include the characterization of the radiation itself, the basic mechanisms of the interaction of the radiation with device structures, the effects on device response, and techniques for minimizing these effects. Chapter 9 departs from this theme and focuses on the effects of transient ionizing radiation on devices. Here the main effects are caused by the interaction of the radiation with the silicon substrate (i.e., photocurrents), rather than with the silicon dioxide insulator.

Chapters 3 and 4 contain a detailed discussion of the basic mechanisms for ionizing radiation damage in MOS devices, as outlined in Fig. 1.17 and briefly described in Section 1.3.1. Chapter 3 focuses on the charge generation and recombination mechanisms, hole and electron transport, and charge trapping in the insulator. Chapter 4 concentrates on the characteristics of and models for interface trap generation. It provides background on the general nature of interface traps and on the various techniques available for measuring their properties.

Chapter 5 proceeds to describe in detail the effects of ionizing radiation
on the characteristics of MOS devices and circuits. This is done by first considering the effects on simple MOS capacitor structures and then relating these to the effects on MOS transistors. Parasitic devices that exist in various technologies and layout geometries are also considered. Since the effects of irradiation are not necessarily stable with time after irradiation, a discussion of postirradiation effects and annealing is included. The response of integrated circuits to ionizing radiation is next detailed, and the relationship between the circuit and discrete transistor response discussed. The chapter concludes with a commentary on the effects of testing conditions on the measured radiation response of MOS devices, and on the differences between the test environment and the real environments typically of interest.

Chapter 6 contains an overview and compilation from the open literature of the effects of processing parameters on the subsequent radiation response of MOS devices. A discussion of the factors involved in integrating processes into a hardened technology follows, along with some considerations for maintaining and characterizing a hardened technology. Chapter 7 focuses on the processes in advanced technologies that themselves cause radiation damage to MOS devices during the course of their manufacture. The extent to which this process-induced damage may be annealed is described. Although not strictly a process-induced radiation effect, but rather an effect produced by device operation, hot carrier effects and their implications for long-term reliability are also discussed in some detail.

The radiation sources of interest both in real environments encountered by circuits and those of most use in testing the radiation response of devices are described in chapter 8. The concepts of dosimetry, or the amount of energy absorbed by the device structures, are explained, as are the techniques used in measuring the strength of these sources. Finally, the correct techniques for applying these principles to MOS device testing and characterization are covered.

Chapter 9 describes the effects of transient ionizing radiation on MOS devices. Here the radiation source is such that the ionization rate is very rapid, creating large numbers of electrons and holes in the device materials. As a result, the main effects of interest are not caused by damage to the silicon dioxide insulator layer, but rather by the photocurrents generated in the silicon substrate. The transient phenomena can lead to either “soft” errors, where information in the circuits or devices is disrupted but the device is not permanently damaged, or to “hard” errors, where the device is permanently altered. The sources of transient high-dose-rate radiation can be weapon environments or single highly energetic charged particles such as found in the cosmic ray spectrum. These environments can initiate latchup (the triggering of the parasitic SCR present in CMOS circuits) or snapback (a high current state of an MOS transistor), or upset of information in memory elements or logic chains. This chapter covers the mechanisms for these phenomena, describes the effects on MOS ICs and how these effects
are measured, and discusses techniques for eliminating or minimizing their impact.

REFERENCES


7 PROCESS-INDUCED RADIATION EFFECTS

T. P. Ma
Department of Electrical Engineering
Yale University
New Haven, Connecticut

7.1 Overview

7.2 Hot-carrier effects and long-term reliability problems
7.2.1 Nature of oxide traps, trapping kinetics, and measurement techniques

7.3 Process-induced ionizing radiation effects
7.3.1 Effects of E-beam and X-ray lithography
7.3.2 Effects of RIE and other plasma processes
7.3.3 Effects of sputtering and E-gun deposition

7.4 Annealing of process-induced radiation effects
7.4.1 Thermal annealing
7.4.2 Hydrogen-assisted thermal annealing
7.4.3 RF plasma annealing

7.5 Summary
References

7.1 OVERVIEW

In Chapter 6, the process flow for a typical CMOS technology was presented. Because of the stringent requirements in the dimensional control for the present and the next generation VLSI chips, advanced processing techniques such as electron-beam (E-beam) lithography, X-ray lithography, RIE (reactive ion etching) and other plasma processes have been introduced. Many of these techniques employ highly energetic particles or photons, which can cause radiation damage to the devices and circuits being fabricated.

As discussed in previous chapters, the two most commonly observed ionizing radiation effects in a MOS device are (a) buildup of positive charge in the oxide and (b) an increase in the interface traps. In this chapter, a third effect, the increase in the bulk oxide traps, will also be discussed. In terms of the device parameters, the radiation effects listed above could
cause a shift in the threshold voltage, a reduction in transconductance, an increase in the leakage current, and a degradation of the long-term reliability.

While the objective of the rad-hard (radiation-hardening) technology discussed in Chapter 6 is to develop processes that will minimize the degradation of the device and circuit performance while operating in a radiation environment, the focus of the present chapter is on the performance of the fabricated devices in a radiation-free environment. Because of the differences in the intended applications, the requirements for the two cases do not completely coincide. For example, in the latter, the process-induced radiation damage may be considered to be completely removed if the subsequent annealing treatment results in devices and circuits that will function electrically as well as those that have never received any radiation processes. In contrast, the former must also consider the change in the radiation sensitivity after the completion of the processing steps. It has been shown in Chapter 6 that residual radiation damage may remain in an MOS structure after an annealing treatment; such damage reveals itself as a degradation of the radiation hardness even though its electrical properties appear to have completely recovered. In this sense, the requirements for the annealing process in the rad-hard technology are much more stringent than those to be discussed in this chapter.

The effectiveness of the annealing depends on how it is implemented. Due to the restrictions imposed by the structural, material, and other considerations, it is often necessary to set a limit on the temperature and time duration of the thermal annealing process. As a result, some residual damage may remain after the annealing step, damage which may significantly degrade the device performance as well as its operating life. For example, in the process flow diagram presented in Chapter 6 (Fig. 6.1), the ion implantation steps are usually followed by a high-temperature thermal anneal to activate the dopant impurities, which normally serves to remove completely or nearly completely the radiation damage introduced prior to this step. On the other hand, if radiation processes are used after the metallization, such as metal delineation using E-beam lithography, X-ray lithography, and RIE, the temperatures used in the subsequent thermal annealing processes may be limited to avoid undesirable reactions between the metal and the underlying materials. In such cases, incomplete annealing of the radiation damage could become problematic. A more detailed discussion of the annealing processes is presented in Section 7.4.

In addition to the aforementioned radiation effects, a less commonly observed but potentially detrimental effect is the introduction of mobile ions in the oxide during plasma deposition and etching processes. This effect will be discussed in Section 7.3.2.

Of the radiation effects cited above, the bulk oxide traps, especially the neutral traps, are not detectable by the usual steady-state measurements used to obtain information about the oxide charge and interface traps. For
the same reason, these neutral traps also do not affect the device performance unless electrons or holes are introduced to the oxide and get trapped there. Because of the energy barrier between the gate oxide and the silicon (see Chapter 1; Fig. 1.2), only electrons or holes with sufficient energies (so-called hot electrons or hot holes) are able to surmount the barrier and enter the oxide. Unfortunately, however, the operating conditions of VLSI devices often produce such hot carriers, and some of these carriers will be injected into the oxide, of which a fraction will be trapped. Over a period of time, a significant drift of the device parameters, such as the threshold voltage and the transconductance, will occur, eventually leading to performance degradation or malfunctioning of the circuits.

Since it has been demonstrated experimentally that the radiation-induced neutral traps are much more difficult to anneal out than the oxide charge and interface traps [1–4], hot carrier injection and trapping have become a major concern. In the next section, the relevant hot-carrier effects will be discussed in some detail.

### 7.2 HOT-CARRIER EFFECTS AND LONG-TERM RELIABILITY PROBLEMS

When a MOSFET is biased on, the free carriers in the conduction channel flowing from the source to the drain will gain kinetic energy from the high field near the drain junction. At sufficiently large drain voltages, the carriers may gain sufficient energy to be emitted into the SiO₂ layer near the drain junction. Such an emission mechanism for hot electrons in an n-channel MOSFET is shown schematically in Fig. 7.1 [5]. It should be noted that, while the source-drain field is in the direction along the SiO₂/Si interface, carrier scattering could cause some hot carriers to assume a velocity component perpendicular to the interface. The electric field from the gate

![Fig. 7.1 Schematic cross-sectional view of a n-channel MOSFET showing channel hot-electron injection near the drain junction. (After Ning [5]; © 1978, Pergamon Press, Ltd. Reprinted with permission.)](image-url)
also tends to cause carrier motion in that direction. Both experimental results [5, 6] and theoretical modeling of two-dimensional transport [5, 7] showed that the injected charge tends to localize within a small region near the drain junction. This is the channel hot electron effect.

The emission probability for channel hot electrons is a function of a number of device parameters, including the source-drain voltage, the channel length, the channel doping concentration, the drain junction doping profile, and the device temperature. The dependence on these parameters

![Diagram](image)

Fig. 7.2 Schematic diagrams showing substrate hot-electron injection in an n-channel MOSFET: (a) cross-sectional view; (b) energy band diagram. (After Ning [5]; © 1978, Pergamon Press, Ltd. Reprinted with permission.)
has been studied, and interested readers are referred to Refs. 7–9 for
details.

Another way by which hot electrons may be created in silicon and
injected to the oxide is illustrated in Fig. 7.2(a) [5], where the source and the
drain are at the same potential, and the substrate is reverse biased with
respect to the source (drain). In this configuration, there is no channel
current flowing from the source to the drain, and therefore one does not
expect any channel hot-electron effect. However, due to the substrate bias,
a high field depletion region is formed beneath the channel, and electrons
entering this high field region will be accelerated toward the gate, some
becoming energetic enough to be emitted over the barrier. The energy band
diagram representing this mechanism is shown in Fig. 7.2(b). The electrons
are thermally generated within the depletion layer or by diffusion into the
depletion region from the bulk neutral region. This is the substrate hot-
electron effect [5]. Since the source of the hot electrons is thermally activated
in this case, the emission current rises rapidly with ambient temperature
[5, 8]. In contrast to the channel hot-electron effect discussed above, the
spatial distribution of the hot electrons in this case is relatively uniform
along the channel.

Once the hot electrons are emitted into the SiO₂ layer, most of them will
go through the insulator and be collected by the gate electrode. A small
fraction, however, will be captured by the traps in the SiO₂ layer, causing
the now well-recognized reliability problem of threshold voltage shift and
transconductance degradation. That fraction depends on the trapping prob-
ability, which in turn depends on the density, distribution, and the capture
cross-section of the traps.

In addition to the aforementioned trapping effect, the hot electrons may
also cause the generation of interface traps and positive oxide charge [9, 10],
and such effects are enhanced as the oxide field increases.

Hot hole injection has also been reported [11–13]. However, because of
the larger energy barrier for holes at the SiO₂/Si interface compared to that
for electrons, hole injection is much more difficult, although there has been
evidence suggesting that hole injection plays an important role in the
generation of interface traps and degradation of the transconductance
[12, 13].

### 7.2.1 Nature of Oxide Traps, Trapping Kinetics, and
Measurement Techniques

Electronic traps in SiO₂ are characterized by a capture cross-section, σ,
which is a measure of the ability of the potential of the trapping centre to
attract and capture a free carrier into a bound state. Depending on the
microscopic origins of these traps (e.g., broken bonds, H₂O or other
impurities, or strained bonds), they can be Coulombic attractive (10⁻¹⁴ ≤
σ ≤ 10⁻¹² cm²), neutral (10⁻¹⁸ ≤ σ ≤ 10⁻¹⁴ cm²), or Coulombic repulsive
\(10^{-21} \leq \sigma \leq 10^{-18}\ \text{cm}^2\). The nature of these various types of traps has been reviewed previously [14].

The radiation-induced neutral traps were first observed one decade ago [1, 15]. The capture cross-sections of these traps are in the range from \(10^{-18}\) to \(10^{-15}\ \text{cm}^2\), and they require a temperature of at least 500°C to anneal out thermally.

Various experimental techniques have been developed to measure these traps. Most of them involve the injection of hot electrons into the gate oxide while the change in the field distribution is monitored as trapping occurs. The various injection methods that have been used include avalanche injection [16], internal photoemission [17, 18], high field tunneling [19], channel hot-electron injection [5], and substrate hot-electron injection [5, 8]. In the first three methods mentioned above, a two-terminal MOS capacitor may be used, and the trap parameters are obtained by analyzing the \(\Delta V_G\) vs. \(Q_{\text{inj}}\) curve, where \(\Delta V_G\) is the change in the gate voltage to maintain a constant field at the injecting interface, and \(Q_{\text{inj}}\) is the total amount of injected charge over the period of the injection experiment. In the latter two methods, a multiterminal device is used, which provides the advantage that the gate voltage and the injection current may be independently controlled. Regardless of how the electrons are injected, the basic principles on which the trap parameters are extracted are essentially the same; these are briefly described below.

Let us first assume the oxide contains only one type of trap with a capture cross-section \(\sigma\). The trapping behavior may be described by a first-order kinetic rate equation given by [20]

\[
\frac{dn_t}{dt} = n v_t \sigma (n_0 - n_t)
\]  \hspace{1cm} (7.1)

where

- \(n_t\) = the density of filled traps
- \(n\) = the injected electron density
- \(v_t\) = the thermal velocity
- \(\sigma\) = the capture cross-section
- \(n_0\) = the total trap density

In order to relate Eq. (7.1) to the macroscopic current \(J\) measured in the external circuit, it is usually assumed that the thermal velocity of the electron equals its drift velocity so that \(J = q n v_t\), and Eq. (7.1) can be rewritten as

\[
\frac{dn_t}{dt} = \frac{J}{q} \sigma (n_0 - n_t)
\]  \hspace{1cm} (7.2)
Integrating Eq. (7.2) over the thickness of the SiO₂, one obtains

\[ \frac{dQ}{dt} = \frac{J}{q} \sigma (qN_t - Q) \]  (7.3)

where

\[ Q = q \int_0^{d_{ox}} n_i(x) \, dx \quad \text{and} \quad N_t = \int_0^{d_{ox}} n_0(x) \, dx \]  (7.4)

Equation (7.3) has the solution for a constant \( J \)

\[ Q(t) = qN_t \left[ 1 - \exp \left( -\frac{t}{\tau} \right) \right] \]  (7.5)

where \( \tau = (J\sigma/q)^{-1} \) is the capture time constant.

The trapped charge \( Q \) may be related to the flatband voltage shift (or, more generally, the shift in the voltage required to maintain a given surface potential)

\[ \Delta V_{fb}(t) = \frac{\bar{X}Q(i)}{\varepsilon_{ox}} \]  (7.6)

where \( \bar{X} \) is the centroid of the trapped charge, and \( \varepsilon_{ox} \) is the permittivity of the oxide.

Defining the effective trap density per unit area as

\[ N_{eff} = \frac{\bar{X}N_t}{d_{ox}} \]  (7.7)

one can combine Eqs. (7.5), (7.6), and (7.7) to yield

\[ \Delta V_{fb}(t) = \frac{qN_{eff}}{C_{ox}} \left[ 1 - \exp \left( -\frac{\sigma J t}{q} \right) \right] \]  (7.8)

where \( C_{ox} = \varepsilon_{ox}/d_{ox} \) is the oxide capacitance per unit area.

Equation (7.8) thus describes the time-dependent flatband or midgap voltage shift that one measures during the constant current injection experiment. If one takes the time derivative of Eq. (7.8), one finds

\[ \frac{d}{dt} \Delta V_{fb}(t) = \frac{\sigma J N_{eff}}{C_{ox}} \exp \left( -\frac{\sigma J t}{q} \right) \]  (7.9)

The trapping parameters \( N_{eff} \) and \( \sigma \) are then obtainable from a plot of the natural logarithm of \( \Delta V_{fb}(t) \) against elapsed injection time. The slope of this line yields \( \sigma J/q \), while the intercept at \( t = 0 \) yields \( \sigma J N_{eff}/C_{ox} \).
If more than one type of trap exists with different capture cross sections, Eqs. (7.8) and (7.9) are modified to become

\[ \Delta V_{fb}(t) = \frac{q}{C_{ox}} \sum_{m} N_{eff}^{m} \left[ 1 - \exp\left( -\frac{\sigma_{m} J t}{q} \right) \right] \]  

(7.10)

and

\[ \frac{d}{dt} \Delta V_{fb}(t) = \frac{J}{C_{ox}} \sum_{m} \sigma_{m} N_{eff}^{m} \exp\left( -\frac{\sigma_{m} J t}{q} \right) \]  

(7.11)

In this case, \( m \) straight line segments will appear in the \( \log([d/dt] \Delta V_{fb}) \) vs. \( t \) plot, and \( N_{eff}^{m} \) and \( \sigma_{m} \) will be obtained similarly.

Alternatively, the parameters \( N_{eff}^{m} \) and \( \sigma_{m} \) may be obtained by numerical techniques using a computer.

In the derivation presented above, it has been assumed that the hot-electron injection process does not cause the generation of interface traps, positive oxide charge, and new oxide traps. In a practical experiment, however, it is likely that new charge centers and traps may be created during the injection process unless extreme care is taken to prevent it. Therefore, it is very important to separate out the effects due to the traps that existed before the injection experiment from those newly generated ones. Attempts have recently been made to analyze the trapping dynamics in the presence of the trap generation processes [21].

In addition to the effective trap density and its capture cross-section, if one is interested in getting some information about the spatial distribution of the traps, the photo I-V techniques [18, 22] may be used, which provides the centroid of the trap distribution. However, this technique requires a semitransparent metal electrode to allow light penetration. Special precautions also must be taken to minimize photodetrapping and additional trapping during the measurement.

7.3 PROCESS-INDUCED IONIZING RADIATION EFFECTS

7.3.1 Effects of E-beam and X-ray Lithography

E-beam or X-ray lithography offers the advantage of higher resolution than the conventional optical lithography. This is because the wavelength of the radiation is much shorter, so that the optical diffraction effect, which limits the resolution of the ultraviolet (UV) lithography, does not play a significant role even for submicrometer feature sizes.

In the focused-beam electron exposure systems, the electron energy is typically in the range 20–50 keV. The sensitivity of the E-beam resist requires an exposure level of 10–100 \( \mu C/cm^2 \). For the X-ray exposure systems, the wavelength ranges from 0.4 to 5.0 nm, and an exposure level of
Fig. 7.3 Density of radiation-induced (a) oxide charge and (b) interface traps in a MOS capacitor as a function of E-beam dose. The sample has an Al-gate and a thermal oxide of 500 Å grown and annealed at 1050°C.
10–100 mJ/cm² is required for the X-ray resist. In both cases, the silicon wafer will have received an accumulated dose of over 1 Mrad(Si) after each exposure [e.g., for a 25-keV E-beam, a dose of 1 × 10⁻⁵ C/cm² corresponds approximately to 10 Mrads(Si)]. (See Fig. 7.3.)

Detailed descriptions of the E-beam and X-ray lithography systems and their technology developments can be found in Ref. 23.

The generation of oxide charge and interface traps by energetic electrons and X-rays was recognized two decades ago [24, 25]. Figure 7.3 [26] shows an example of the range of oxide charge and interface-trap densities for a commercial device resulting from exposure to a 25 keV E-beam. Experimental evidence has shown that X-ray radiation causes qualitatively the same effects. The possible mechanisms for the observed radiation effects have been discussed in detail in Chapters 3 and 4. The only point worth adding here is that in either E-beam or X-ray lithography, no external voltage bias is applied to the device being exposed.

Later the radiation-induced electron traps in SiO₂ were discovered [1, 15]. In addition to the Coulombic attractive traps, which are associated with the radiation-induced positive charge, many neutral traps with capture cross-sections ranging from 10⁻¹⁸ to 10⁻¹⁵ cm² were detected after either X-ray [4, 27] or E-beam radiation [2, 28]. While most of the radiation-induced positive charge can be annealed out at 400°C in forming gas, it has been found that the neutral traps are not completely removed even after annealing at 500°C [1–3]. Figure 7.4 shows an example of the effect of E-beam radiation on the trapping probability after thermal annealing. Table 7.1 lists the trap parameters after E-beam radiation and thermal annealing at 400°C for 30 min in forming gas for the samples reported in Ref. [1]. From the photo I–V measurements, it was suggested that the radiation-induced neutral traps are uniformly distributed spatially [4, 29, 30], as long as the incident radiation is absorbed uniformly through the oxide.

Following the initial discovery of the radiation-induced neutral traps, a

<table>
<thead>
<tr>
<th>Trap</th>
<th>Charge</th>
<th>( \sigma_{\text{tr}} ) (cm⁻²)</th>
<th>( N_{\epsilon_{\text{ox}}} ) (cm⁻²)</th>
<th>( \sigma_{\text{tr}} ) (cm⁻²)</th>
<th>( N_{\epsilon_{\text{ox}}} ) (cm⁻²)</th>
<th>( \Delta N_{\epsilon_{\text{ox}}} ) (cm⁻²)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>+</td>
<td>( 1.2 \times 10^{-14} )</td>
<td>( 2.0 \times 10^{10} )</td>
<td>( 1.0 \times 10^{-14} )</td>
<td>( 0.5 \times 10^{10} )</td>
<td>( 1.5 \times 10^{10} )</td>
</tr>
<tr>
<td>B</td>
<td>0</td>
<td>( 1.6 \times 10^{-15} )</td>
<td>( 4.4 \times 10^{10} )</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>C</td>
<td>0</td>
<td>( 2 \times 10^{-16} )</td>
<td>( 8.4 \times 10^{10} )</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>D</td>
<td>0</td>
<td>( 7 \times 10^{-18} )</td>
<td>( 2.58 \times 10^{11} )</td>
<td>( 3 \times 10^{-17} )</td>
<td>( 1.88 \times 10^{11} )</td>
<td>( 7.0 \times 10^{10} )</td>
</tr>
<tr>
<td>E</td>
<td>0</td>
<td>( 1 \times 10^{-18} )</td>
<td>( 4.7 \times 10^{11} )</td>
<td>( 1.4 \times 10^{-18} )</td>
<td>( 2.6 \times 10^{11} )</td>
<td>( 2.1 \times 10^{11} )</td>
</tr>
</tbody>
</table>

\(^{a}10^{-4} \text{ C cm}^{-2} \text{ at 25 keV.}

\(^{b}400^\circ \text{C 30 min in 90\% N}_2 \text{ and 10\% H}_2.

Source: Aitken et al. [1]. Reprinted with permission.
Fig. 7.4 Threshold voltage shift of E-beam irradiated and annealed polysilicon gate MOS-FETs as a function of injected electrons with annealing history as a parameter. The annealing period was 30 min at each temperature. (After Aitken [2]; © 1979 IEEE. Reprinted with permission.)

substantial amount of work from various groups has confirmed the earlier findings. The significance of the radiation-induced neutral traps and their impacts on the VLSI technology have been summarized [2], and the design constraints for the operating parameters in order to limit the hot-electron injection in VLSI devices have been discussed [31]. Effects of repeated E-beam radiation–thermal annealing cycles on the long-term stability of MOS devices [28], various annealing methods to remove the radiation-induced neutral traps, and ways to block the radiation during E-beam lithography have all been reported [32–34]. Some of these results are summarized below.

In a paper by Shimaya et al. [28] it was reported that, although the E-beam induced neutral traps seemed to disappear completely after a thermal annealing treatment at 700°C in dry N₂, the oxide becomes more susceptible to subsequent radiation damage, suggesting that residual structural defects still remained in the oxide after this high-temperature treatment. Some of their results are shown in Figs. 7.5 and 7.6.

Shimaya et al. also found that the long-term stability of the devices tended to degrade following E-beam exposure and subsequent thermal annealing. An example is shown in Fig. 7.7. A temperature as high as 1000°C was found necessary to recover the structure completely.

Since in a typical VLSI (very-large-scale integration) process it is likely
that the same oxide will receive multiple radiation exposures, Shimaya and colleagues' findings are of significant practical concern, although more work needs to be done to establish the generality of the effects they reported.

In principle, it is possible to place a blocking layer over the gate oxide during X-ray or E-beam lithography to prevent the penetration of incident radiation into the oxide. In practice, however, there are difficulties in its implementation. In an attempt to investigate the feasibility of this method, Shimaya et al. [34] used a scheme incorporating a heavy metal interlayer in their three-layer resist system. They found that, while the theoretical calculation showed that a 280-nm-thick platinum layer should be able to completely block the 20-keV E-beam, the underlying oxide still exhibited a damage level equivalent to approximately 20% of that without the blocking layer. Similar levels of radiation damage were observed for a thicker platinum layer. This was attributed to the X-rays that are generated by the bombardment of the E-beam and that penetrate deeper than the electrons. The radiation effects resulting from the secondary X-rays generated by the bombarding E-beam were also reported by others [35, 36].
Fig. 7.6 (a) Flatband voltage shift and (b) neutral trap density measured after each E-beam exposure and subsequent annealing process. (After Shimaya et al. [28]. Reprinted with the permission of the publisher, The Electrochemical Society, Inc.)
7.3.2 Effects of RIE and Other Plasma Processes

Plasma processes, including plasma etching, RIE (reactive ion etching), and plasma-enhanced chemical vapor deposition (PECVD), are gaining increasing importance in processing VLSI circuits. Common to all these processes is the use of low-pressure gaseous plasma from a RF (radio-frequency) or DC discharge. Since a typical processing plasma contains various energetic species, it is expected that some radiation effects will occur.

In a gaseous discharge, free electrons gain energy from an imposed electric field and lose this energy through collisions with gas molecules. The transfer of the energy to the molecules leads to the formation of a variety of new active species, including metastables, atoms, free radicals, and ions.

In the context of radiation effects, a typical processing plasma contains:

1. Charged particles, which include ions (both positive and negative) and electrons with various kinetic energies
2. Neutral particles, which include molecules, radicals, and atoms either in their ground states or in various excited states
3. Photons, which cover a wide range of wavelengths, from soft X-rays, through UV and visible, to IR (infrared)

Since the discharge initiates from the free electrons in the system, and because of their small mass, the electrons have the highest peak kinetic energy among all the energetic species. For a RF plasma, the highest electron energy (in electron volts) would correspond to the applied RF voltage, which in some systems could be as high as 1 or 2 keV. The high-energy end of the photons, i.e., the soft X-rays and VUV (vacuum ultraviolet) are generated by the electrons impinging upon a solid object, either the electrode plates, the chamber wall, or the sample. Ultraviolet photons are also emitted by deexcitation of gas molecules in the plasma. Therefore, the peak photon energy should track that of the electron energy. The ions, being charged particles, can gain kinetic energy in the presence of an electric field. Because of their much higher masses compared to electrons, the kinetic energies of the ions are generally much lower than the peak energy of the electrons. The molecules and other neutral species do not gain kinetic energy from the electric field and therefore have the lowest energies in comparison.

The ions in a processing plasma usually do not possess kinetic energies in excess of a few hundred electron volts, but because of their masses they are capable of causing displacement damage to the exposed wafer surface. Creations of an amorphous layer on single crystalline Si surfaces, structural disorder, dislocations, and implanted ions have all been observed [37–43], and some of these are serious problems to be dealt with. However, as stated in Chapter 1, this book is intended to focus on issues related to ionizing radiation effects, and therefore we will not address other forms of plasma damage here.

In discussing the ionizing radiation effects in SiO$_2$, we must examine the penetration depth of the various energetic species in the plasma. Because of their high energies and negligible momenta, the X-ray photons have the highest penetration depth among all the energetic species in the plasma. Figure 7.8 shows the photon absorption coefficient as a function of photon energy for Al, Si, and SiO$_2$ [81]. Based on such data, the penetration depth of photons with a given energy may be calculated. For example, the penetration depth in SiO$_2$ for 1 keV photons is approximately 3 $\mu$m, whereas for 350 eV photons it is approximately 0.4 $\mu$m. For photon energies $\geq$100 eV, the absorption coefficients for Al, Si, and SiO$_2$ are rather similar, and thus the penetration depths are also similar.

The electrons, although their energies are as high as those of X-rays, have a much shallower penetration [44], because of the momentum conservation requirement. The ions and neutrals, having even larger masses and
smaller kinetic energy, would have the least penetration depths. In a typical processing plasma environment, the penetration depths for various ions are less than 10 nm.

Based on the various ranges of penetration depths discussed above, we may assess the seriousness of the ionizing radiation effects after each plasma process step. If a thermal anneal process with sufficient temperature to remove the radiation damage can be used after a particular plasma step without creating undesirable side effects, then that step does not cause serious concern. On the other hand, if a high-temperature process following the plasma step is not compatible with other constraints, the residual radiation damage can cause serious reliability problems in the finished circuit. An example is any plasma process after the metallization step, which can introduce radiation damage that may not be completely annealed out due to the temperature limitations.

After the gate electrode is formed, the gate oxide is covered with a protective layer and, among the various energetic species in a processing plasma, only X-rays will be able to penetrate into the oxide and create radiation damage. It is therefore not surprising that the resulting radiation effects should be similar to those observed after X-ray lithography discussed in the previous subsection: generation of oxide charge, interface traps, and neutral oxide traps. All of these electronic defects have been observed experimentally [3, 30, 43, 45]. The only significant difference is in the spatial
distribution of the neutral traps along the thickness direction of the oxide, due to the more limited penetration depth of the soft X-rays generated in the plasma environment [40, 43]. Again, the radiation-induced neutral traps are more difficult to anneal out than the oxide charge and the interface traps [3, 30]. Figure 7.9 [3] shows an example of the electron trapping behavior in thermal SiO₂ after RIE and post-RIE thermal annealing treatment.

While the ions in the processing plasma may not cause significant ionizing radiation effects to the gate oxide due to their limited penetration depths, they may nonetheless cause a potentially more troublesome reliability problem, i.e., the release of mobile ionic charge in SiO₂. This phenomenon can occur when the surface of the SiO₂ is directly exposed to the processing plasma, as reported more than a decade ago [46]. In a related paper [47], Mc Caulhan et al. proposed a model involving ion–insulator interactions to explain the results. It was assumed that, initially, electrically neutral or coulombically bounded immobile impurities, such as Na, may be present on the SiO₂ surface. These neutral and immobile impurities do not affect the electrical properties of the MOS device and therefore are not detected by the usual C-V or I-V measurements. When the SiO₂ surface is exposed to the plasma, it may be bombarded by positive ions generated in the plasma. As these positive ions approach the SiO₂ surface to within a tunneling distance, some of them may be neutralized by electrons coming out of the

---

**Fig. 7.9** Electron-trapping-induced voltage shift as a function of injected charge for polysilicon MOS devices in which the 500-Å thermal SiO₂ was exposed to RIE plasma prior to thermal annealing at specified temperatures. (After Gdula [3]; © 1979 IEEE. Reprinted with permission.)
SiO₂ surface through Auger or resonance transitions [47]. As a result of this neutralization process, net positive charge is created in SiO₂. Either holes in the valence band of SiO₂, or positive ions (e.g., Na⁺) from the surface impurity sites are produced. In the case where Na⁺ are produced on the surface, they could diffuse to the bulk of SiO₂ and cause device instability. Experimental evidence in support of this model has also been reported [47].

The mobile ionic species, once they are introduced into the SiO₂, are extremely difficult to remove, and therefore precautions must be taken to prevent their occurrence. It should be noted that, according to the model discussed above, the necessary conditions for this effect to happen are (1) the oxide surface must be directly exposed to the plasma; (2) the oxide surface must initially contain immobile impurities, such as Na bonded to the SiO₂ network; and (3) the incoming ions must possess sufficient ionization potential so that electronic transitions from the SiO₂ surface can take place to neutralize the ions. Therefore, it should be possible to avoid this effect by designing a plasma process that does not satisfy one or more of the conditions listed above. Consequently, in practice this problem is not as serious as it appears to be.

Another reliability problem associated with plasma processes is the dielectric breakdown phenomenon that has been reported [46, 48]. This phenomenon has been found to be more serious for thin gate SiO₂ under plasma conditions in which the self-bias voltage and the ion energy are high, and is thought to be due to the excessive charging of the gate electrode [48]. More recently, however, evidence was presented which showed that this effect only occurred in process plasmas operated at low frequencies (≪1 MHz) [49]. In any case, since this is not due to ionizing radiation effects, it will not be covered beyond this paragraph.

### 7.3.3 Effects of Sputtering and E-gun Deposition

The radiation damage due to sputter etching and deposition has been studied in detail [46]. In many ways, the sputtering environment is similar to that of RIE and some other plasma processes, with the exception that an inert gas (typically Ar) is used for sputtering. Therefore, the ionizing radiation effects resulting from a sputtering process are expected to be qualitatively similar to those from other plasma processes discussed in the previous section.

In E-gun deposition, an electron beam of energy range 5–30 keV is used to bombard the source material, causing heating and evaporation of the material in a vacuum environment. The E-beam bombardment creates emission of X-rays and reflected and secondary electrons, which can cause radiation damage to the substrate on which the materials is being deposited. Again due to the larger penetration depth of the X-rays compared to the
electrons, it is usually the former that produce most of the radiation effects. Generations of oxide charge, interface traps, and oxide traps have all been observed experimentally [50–52]. Similar to the results obtained from other ionizing radiation sources, the neutral traps generated by the E-beam evaporation process cannot be completely removed after thermal annealing at 550°C [52]. Figures 7.10 and 7.11 show an example of the annealing behavior of the radiation effects after E-beam evaporation [51]. A detailed discussion of the effects of sputtering and E-gun deposition on the radiation hardness of MOS devices can be found in Chapter 6.

![Graph showing the density of radiation-induced positively charged traps caused by E-gun deposition of Al gate as a function of annealing time for three different annealing temperatures. These anneals were in addition to a postmetalization anneal of 20 min in forming gas at 400°C. (After Ning [51]. Reprinted with permission.)](image)

**Fig. 7.10** Density of radiation-induced positively charged traps caused by E-gun deposition of Al gate as a function of annealing time for three different annealing temperatures. These anneals were in addition to a postmetalization anneal of 20 min in forming gas at 400°C. (After Ning [51]. Reprinted with permission.)
Fig. 7.11  Density of radiation-induced neutral traps caused by E-gun deposition of Al gate as a function of annealing time for three different annealing temperatures. These anneals were in addition to a postmetallization anneal of 20 min in forming gas at 400°C. (After Ning [51]. Reprinted with permission.)

7.4 ANNEALING OF PROCESS-INDUCED RADIATION EFFECTS

From the discussions presented in Chapters 3 and 4, it is apparent that the radiation-induced oxide charge, interface traps, and oxide traps are associated with the microscopic structural and bonding defects in SiO₂ or at the SiO₂/Si interface. To remove the radiation damage requires some mechanisms by which bond reformation or some other chemical reactions can take place to make the defect sites electronically inactive. In general, some sort of energy is a necessary input to promote such annealing mechanisms.
7.4.1 Thermal Annealing

The most widely used technique is thermal annealing, in which thermal energy serves to initiate and sustain the various annealing mechanisms. The fact that the radiation-induced positive charge and interface traps in Al-gate MOS devices can be largely annealed out at 400°C or below was recognized long ago [24]. Since then, a good amount of more detailed work has been done to investigate the annealing process, and data on the time dependence and temperature dependence of the annealing behavior have appeared in numerous publications [53–60]. Figures 7.12 [54] and 7.13 [53] are examples of some of the earlier results.

Although a substantial amount of work was done in the 1960s and early 1970s, and a great deal of data were published that were useful from a practical applications point of view, they were less useful in terms of understanding the annealing mechanisms. This is primarily due to the following three factors: (1) the contribution of hydrogen in the thermal annealing process was not sufficiently recognized during the period and therefore was usually not taken into proper consideration; (2) the effects of oxide charge and interface traps were generally not separated; and sometimes (3) the difference between the flatband shift and the threshold voltage shift was not recognized. Because of these factors, the annealing data reported in the literature appear to vary widely among various experiments.

Fig. 7.12 Recovery of threshold voltage as a function of annealing time at several different annealing temperatures for p-channel MOSFETs irradiated with 1.5-MeV electrons. Arrows indicate the spread from sample to sample. The gate oxide thickness is 1200 Å. (After Danchenko et al. [54]. Reprinted with permission.)
Figure 7.13 shows a collection of isochronal annealing data taken from the open literature by Brown et al. [55], which serves to illustrate the enormous variations among the data when these factors are not considered separately.

The important role of hydrogen in the oxidation, annealing, and defect generation processes of SiO₂ is now widely recognized [61–70]. During the thermal annealing process, hydrogen species can undergo chemical reactions to form chemical bonds with certain defects, making them electrically inactive. Therefore, one would expect the annealing mechanism to be very different with or without the participation of hydrogen. Since the hydrogen species might come from within the SiO₂ itself, or from inward diffusion from external sources, it is very difficult to separate out the two annealing mechanisms. The problem is compounded if the annealing ambient and the hydrogen content in the oxide are not strictly controlled. This is part of the reason why the results reported vary widely from one laboratory to another.
Fig. 7.14 A collection of isochronal annealing data for MOS devices (solid curves) and bulk SiO₂ (dashed curves). (After Brown et al. [55]; © 1983 IEEE. Reprinted with permission.) These data were collected from open literature and laboratory results, and their origins can be found in Ref. 55.

Apart from the problem related to hydrogen, essentially all of the annealing data reported in the 1960s and 1970s used either the flatband shift, ΔV_{fb}, or the threshold voltage shift, ΔV_t, to represent the change in the oxide charge. As discussed in Chapter 1, both ΔV_{fb} and ΔV_t are influenced by the changes in both the oxide charge and interface traps. Since there is increasing evidence that the radiation-induced oxide charge and interface traps do not have the same annealing behavior [60, 71], the annealing data based on ΔV_{fb} or ΔV_t measurements cannot be used to obtain parameters (such as the activation energy or the annealing rate) related to the annealing mechanisms for either the oxide charge or the interface traps. In fact, it has been observed that, during thermal annealing of irradiated MOS devices, the interface trap density may increase for a period of time before it starts to decrease, while the radiation-induced oxide charge keeps decreasing with time [71]. An example of this annealing behavior is shown in Figs. 7.15 and 7.16 [71]. This phenomenon is related to the rebound effect discussed in detail in Chapter 5.

If the contributions from interface traps are not taken into account, such “discrepancies” in the annealing data between ΔV_{fb} and ΔV_t are not surpris-
Fig. 7.15  Thermal annealing of positive oxide charge in $n$-channel polysilicon gate MOSFETs irradiated by Co$^{60} \gamma$-rays. (After Sabnis [71]; © 1973 IEEE. Reprinted with permission.)

Fig. 7.16  Time-dependent change of the density of interface traps during thermal annealing for the same set of samples shown in Fig. 7.15. (After Sabnis [71]; © 1973 IEEE. Reprinted with permission.)
ing, because in the two measurements different amounts of interface trapped charge are included. Similarly, the same interface-trap distribution would exert a different influence on the $V_{fb}$ or $V_t$ for a $p$-channel device compared to an $n$-channel device. The importance of radiation-induced interface traps on the $\Delta V_{fb}$ and $\Delta V_t$ measurements has been demonstrated [72], and a more detailed discussion can be found in Chapter 4.

In addition to the oxide charge and interface traps, ionizing radiation causes an increase in neutral traps in the oxide. While the radiation-induced oxide charge and interface traps were identified soon after the discovery of radiation effects in MOS devices [24], the radiation-induced neutral traps were unnoticed until a decade later [1, 29], primarily because it is necessary to inject electrons to detect them. Once the radiation-induced neutral traps were discovered, it was soon found that they were much more difficult to anneal out than the oxide charge and the interface traps [1]. An example of the annealing behavior of radiation-induced neutral traps was shown earlier in Fig. 7.11 [51].

Because of the important role that hydrogen plays in the thermal annealing process, we will discuss in more detail some of the related work here.

### 7.4.2 Hydrogen-Assisted Thermal Annealing

The existence of hydrogen-related chemical bonds in thermal SiO$_2$, such as Si-H and Si-OH, was recorded and reported years ago [61, 62, 69]. Using infrared spectroscopy, it has been shown that both “wet” SiO$_2$ (thermally grown in wet O$_2$ or steam) and “dry” (thermally grown in “dry” O$_2$) contain significant amounts of hydrogen [69]. The role of hydrogen in oxidation, annealing, ionizing radiation, and hot electron effects in thermal SiO$_2$ has been extensively discussed [61–71].

After two decades of research, it is now well recognized that the active participation of hydrogen during the thermal annealing process greatly enhances the effectiveness of the anneal: the time and temperature needed to achieve a given reduction of the radiation-induced charges and traps can be significantly reduced.

The basic chemical reactions involving hydrogen in the annealing process have been proposed as [61]:

\[
\text{Si} = \text{Si} + \text{H} \rightarrow \text{SiH} \quad (7.12)
\]

\[
\text{Si} - \text{O} = \text{Si} + \text{H} \rightarrow \text{SiOH} \quad (7.13)
\]

Essentially, the “dangling” bonds, which are electronically active, are “passivated” by the hydrogen.

In the case that hydrogen must be introduced in the annealing ambient, forming gas is most widely used, although pure hydrogen is also occasionally used. The forming gas is typically made of 5–10% H$_2$ mixed in an inert gas,
Fig. 7.17 Isothermal annealing characteristics of interface traps at 280°C in various ambients for Al-gate MOS capacitors. (After Reed [73].)

Fig. 7.18 Proposed model for the annihilation of interface traps by a hydrogen species produced by an Al–H₂O reaction at the Al/SiO₂ interface. (After Deal [68]. Reprinted with the permission of the publisher, The Electrochemical Society, Inc.)
such as N₂, He, or Ar. To be effective, the active hydrogen species must be diffused into SiO₂ to react with the defects. Therefore, the thermal energy in the annealing process serves three purposes: (1) it promotes the generation of active hydrogen species, most likely atomic hydrogen; (2) it enhances the diffusion of same; and (3) it drives the defect reaction processes. Depending on the structure of the device being annealed and the annealing parameters, one or the other of the three may be the limiting factor in the overall annealing process.

If the oxide is covered with a layer of aluminum, as in the Al-gate technology, it has been found that the thermal annealing is equally effective whether or not the ambient contains hydrogen [63, 73, 74]. Figure 7.17 [73] shows an example of the insensitivity of the Al-gate MOS capacitor to the annealing ambient. This has been attributed to the fact that, during thermal annealing, the aluminum reacts with minute amounts of H₂O on the oxide surface to release the active hydrogen species, which then diffuse into SiO₂ and annihilate the bounding defects [63, 68]. This model is illustrated in Fig. 7.18 [68].

![Diagram](image)

**Fig. 7.19** Isothermal annealing characteristics of interface traps at 500°C in nitrogen and forming gas for polysilicon gate MOS capacitors. (After Fishbein et al. [75]. Reprinted with the permission of the publisher, The Electrochemical Society, Inc.)
In contrast, for polysilicon gate devices, the thermal annealing process goes faster in forming gas than in an inert atmosphere, as shown in Fig. 7.19 [75], but is generally less effective than that for Al-gate devices even with forming gas [2, 73, 76], due to the fact that it is difficult for hydrogen to diffuse through the polysilicon gate at the annealing temperatures (≤500°C). There has been evidence that, if the polysilicon gate is covered with a layer of aluminum, the defect annihilation rate is enhanced [2], presumably due to the production of active hydrogen species at the Al-polysilicon interface in a way similar to that at the Al/SiO₂ interface. These hydrogen species then diffuse into SiO₂ through grain boundaries of the polysilicon. Figure 7.20 [2] shows an example of this effect.

The possibility that the active species involved in the defect annealing process is atomic hydrogen was investigated in a more recent work [70]. In that experiment, the authors used atomic deuterium to annihilate the ESR

![Graph showing annealing characteristics](image-url)

**Fig. 7.20** Annealing characteristics of radiation-induced positive oxide charge for MOS capacitors with Al gate, polysilicon gate, or polysilicon gate coated with Al. The oxides are 350-Å thick and grown in dry O₂ at 1000°C. (After Altken [2]; © 1979 IEEE. Reprinted with permission.)
(electron spin resonance) spin center at the SiO₂/Si interface in the temperature range 25–230°C and concluded that atomic hydrogen reacted with trivalent silicon defects to form Si–H bonds. An example of their results is shown in Fig. 7.21. Detailed studies of the annealing kinetics in Al-gate MOS capacitors by Reed [73] also brought out the importance of atomic hydrogen in the annealing of interface traps.

The fact that the lack of active hydrogen species could hinder the defect annihilation process in SiO₂ is also evidenced in experiments involving MNOS (metal/nitride/oxide/Si) structures. It was found that, compared to the MOS structure, it is much more difficult to anneal out the oxide charge and interface traps in an MNOS structure [66]. Since silicon nitride is known to be an effective barrier for the diffusion of impurities, including hydrogen, the observed difficulty in thermal annealing for MNOS structures has been attributed to the blocking of needed hydrogen by the nitride layer. Figure 7.22 [66] shows an example of the incomplete removal of the radiation-induced oxide charge and interface traps in MIS structures containing a nitride layer after 550°C forming gas annealing.

Another piece of evidence which supports the model described above can be found in Ref. 77. In that work, Schols et al. introduced hydrogen into the oxide after nitride deposition. This was done by allowing the hydrogen to
Fig. 7.22  C-V plots at 25°C and −196°C for MIS capacitors with or without a Si₃N₄ layer, showing the effect of the nitride layer that blocks the hydrogen species (represented by the circles at the bottom of Al) from getting to the SiO₂/Si interface during post-Al anneal, resulting in higher densities of interface traps. The ledge on the −196°C curves for structures I, II, and III indicates high density of interface traps. (After Deal et al. [66]. Reprinted with the permission of the publisher, The Electrochemical Society, Inc.)
diffuse at a high temperature (950°C), through cutouts in the nitride layer, from the periphery of the gate into the gate oxide. They found that with this treatment the interface traps and oxide charge generated by ionizing radiation could be completely annihilated by a 30-min thermal anneal at 450°C. Samples without the hydrogen treatment, however, showed considerable amounts of residual defects after the annealing.

With recognition of the importance of hydrogen in the thermal annealing process, attempts have been made to more effectively introduce the active hydrogen species into the oxide. One way is the use of atomic hydrogen generated in a plasma environment, as mentioned previously [70]. Another

![Graph](image)

**Fig. 7.23** Effectiveness of thermal annealing (450°C, 30 min) as a function of pressure in an Ar/H₂ ambient for n-channel polysilicon gate MOSFETs (gate oxide 350-Å thick) exposed to radiation during E-gun aluminum evaporation. The vertical axis represents the voltage shift due to trapping of injected electrons by the positive charge and neutral traps in the oxide. (After Reisman and Merz [33]. Reprinted with the permission of the publisher, The Electrochemical Society, Inc.)
method is to perform the thermal annealing in a high-pressure environment containing hydrogen [33]. Using a specially designed high-pressure system, Reisman and Merz found that radiation-induced charged and neutral centers in Si-gate devices could be effectively removed by annealing in 50 atm of forming gas at 400°C for 30 min for radiation doses as high as $5 \times 10^7$ rads(SiO$_2$). In contrast, temperatures as high as 600°C are required to achieve similar results if the thermal annealing was done in 1-atm pressure. Figure 7.23 [33] shows an example of the dependence of the annealing behavior on H$_2$ pressure.

In addition to the thermal annealing and its variations, a very different annealing technique has been demonstrated to be effective in removing the radiation-induced charges and traps in SiO$_2$. Since this technique, called RF (radio-frequency) plasma annealing [78–84], takes an entirely different approach from that of conventional thermal annealing, it will be described in some detail here.

### 7.4.3 RF Plasma Annealing

The RF annealing apparatus and the experimental details can be found in a number of publications [81–84]. Basically, the annealing apparatus is similar to a parallel-plate RIE system but with some important differences. The most significant difference is that in the annealing system the sample experiences no DC self-biasing effect. This can be accomplished by having two equal-area parallel-plate electrodes with a wide spacing (≥8 in.) between them, and the sample is positioned in the center between the two electrodes. Other important considerations are that the pressure must be sufficiently low (<10 μm), the wafer surface must be perpendicular to the RF field, and both sides of the wafer must be exposed to the plasma. Departure from these conditions has been found to produce unsatisfactory results.

Figure 7.24 [83, 84] shows the effect of RF power as an annealing parameter, and the residual oxide charge and interface traps are plotted as a function of annealing time. Within the power range 0.2–0.6 W/cm$^2$, most of the anneal is completed in the first 10 min.

Similar annealing behavior was observed for MNOS samples [79]. The time required to achieve complete anneal, however, was longer for the MNOS capacitor.

The effect of RF plasma annealing on the radiation-induced neutral traps has also been studied [80]. It has been shown that there are several radiation-induced electron traps with capture cross-sections ranging from $10^{-18}$ to $10^{-14}$ cm$^2$ and they are completely annealed out in the RF plasma.

The temperature of the wafer during annealing is a function of the RF power and the annealing time. It has been shown [83, 84] that the wafer temperature saturates after a few minutes and it does not exceed 300°C even at 600 W, the highest RF power level used in these experiments. This
suggests that thermal annealing is not the dominating mechanism, because the radiation-induced neutral traps require a temperature over 550°C to anneal out thermally.

From the accumulated experimental data, and from the careful analysis of the various possible processes that may be taking place in the RF plasma, three essential components for the annealing have been identified [81, 83]: (1) plasma interactions with SiO₂; (2) RF field; and (3) induced wafer temperature. The primary role of the plasma is to serve as an excitation source for the generation of the electron-hole pairs in the oxide. The RF field controls the motion of these radiation-induced free carriers, modifying the defect-reaction coordinates such that a more favorable annealing reaction can be achieved. Furthermore, the RF field, along with the plasma-
wafer interactions, induces a moderate heating of the wafer, which also contributes to the annealing.

The plasma-induced excess carriers could participate in two possible annealing processes. The first involves the neutralization of the positive oxide charge centers through electron capture. The second involves the more complicated recombination-enhanced defect reactions (REDR), which have been observed by Kimerling ([85]; see also Refs. 86 and 87) in GaAs and other semiconductors, and could lead to the annihilation of the defects in the SiO₂.

According to the REDR theory [86, 87], the recombination of an excess electron with a hole through a defect center releases a substantial amount of energy, which could cause excitations of the vibrational state of the defect. If proper conditions are satisfied, this process could promote simple solid-state reactions involving the defects, such as diffusion, dissociation, and annihilation. Experimentally, the occurrence of the REDR mechanism helps to significantly reduce the temperature required for certain defect annealing processes.

In the RF plasma annealing treatment of MOS devices, the excess carriers in SiO₂ are generated by plasma radiation, and enhanced annealing can be expected through the nonradiative recombination at the defect sites. As was pointed out in Section 7.3.2, the energetic gas plasma, if it acts alone, is a source of radiation damage. It is only through the proper cooperative interactions of all three essential components that effective annealing is possible.

Although the RF plasma contains various energetic ions, neutrals, electrons, and photons with a wide range of energies, the only species that can penetrate through the gate electrode (Al or Si) and interact with the SiO₂ layer are photons with energies more than 15 eV (see Section 7.3.2).

Based on this observation, an experimental simulation of the RF plasma annealing process was performed in the absence of the RF plasma [81]. In that experiment, the three essential components for an effective anneal—(1) the X-ray radiation, (2) the RF field, and (3) the induced wafer temperature—could be independently controlled, and their individual and combined effects were investigated. It was found that the cooperative interactions of all three components were necessary to achieve good annealing results. The detailed descriptions of this experiment and its results can be found elsewhere [81–84].

A convenient test of the REDR theory is the analysis of the activation energies associated with the annealing process. A reduction of the effective defect reaction barriers is expected in the recombination-enhanced processes, which should be reflected in the measured activation energies [85, 86].

In the simulation experiment mentioned above, the activation energy involved in the annealing process could be conveniently studied. In Fig. 7.25 [81], the data are plotted for a pure thermal process together with the
recombination-enhanced at two different X-ray levels. These data were taken from an isochronal annealing experiment with a 2-min annealing time, using the setup described in Ref. 81. The annealing temperature ranged from 25° to 120°C. Although it is believed that more than one type of defect is involved in the annealing process, with possibly different activation energies, the data in Fig. 7.25 do indicate that an average activation energy may be assigned for each process. From the slope of the best linear fit, the average activation energy for the pure thermal process is approximately 0.4 eV, with a ±0.05-eV spread among several different samples. The other two curves (labeled “recombination enhanced”), however, both give rise to an average activation energy of approximately 0.13 eV, corresponding to a factor of 3 reduction when compared to the pure thermal process. This significant reduction in the activation energy is consistent with the REDR model.

Note in Fig. 7.25 that the difference in the X-ray flux for the two recombination-enhanced curves only affects the net annealing rates and does not affect the activation energy involved. This suggests that the recombination-enhanced annealing process is a fundamental characteristic of certain
defects whose altered activation energy does not depend on the number of excess electrons and holes that are present. The net rate of annealing, however, does depend on the X-ray flux. Within the range of X-ray flux used in that study [81], the radiation damage component seems to play a significant role, which increases with the X-ray flux and negates the annealing rate.

7.5 SUMMARY

Many of the advanced processing techniques involve the use of very energetic photons and charged particles, and as a result ionizing radiation effects are introduced to the devices and circuits being fabricated. Because of the ever-increasing demand for tighter dimensional control, it is expected that these techniques, including E-beam lithography, X-ray lithography, RIE, and other plasma processes, will assume more importance as the down-scaling trend continues. Accordingly, the process-induced radiation effects will become a serious issue of increasing practical concern.

Although the details of the radiation sources may be different, the fundamental mechanisms involved in the process-induced radiation effects are basically the same as those discussed in Chapters 3–6, where the emphasis was on issues related to the understanding and development of radiation–hard technology.

The most notable ionizing radiation effects in MOS devices, whether process induced or otherwise, include a buildup of positive charge in the oxide and an increase in the interface traps. In terms of device performance, these effects translate into a change in the threshold voltage, a degradation of the transconductance, and an increase in the leakage current. What is less notable, but by no means less important, is the radiation-induced neutral traps in the oxide. Being charge neutral, these traps are not detectable by the usual steady-state measurements used to probe interface traps and oxide charge; instead, injection of charge into the oxide is required to reveal their presence. These traps are of practical concern, because under certain operating conditions hot carriers may be produced in some MOS devices in a VLSI circuit, of which a portion may be injected into the oxide, causing reliability problems.

Whereas the objective of the radiation-hard technology is to minimize the degradation of the device and circuit performance while operating in a radiation environment, this chapter focuses on the performance of the fabricated devices (exposed to process-induced radiation effects) in a radiation-free environment. This difference in the intended applications makes a significant difference in the implementation of practical solutions. For example, in principle it does not matter how severe the process-induced radiation damage is, as long as the damage can be completely removed by a subsequent annealing process. Therefore, annealing is a much more im-
portant issue in the chapter than in others, and hence has received substan-
tial coverage.

Among the various possible annealing methods, thermal annealing is most widely used. In principal, the process-induced radiation damage can be completely removed if the annealing step is performed at sufficiently high temperatures. In practice, however, the applicable temperature for thermal annealing is limited by other considerations in a given technology. As a consequence, incomplete removal of the radiation damage may result.

There is a substantial amount of experimental data indicating that the radiation-induced oxide charge and interface traps can be completely annealed out at a temperature of 450°C or lower, but the neutral traps require a significantly higher temperature to be removed, sometimes higher than can be tolerated. This makes the radiation-induced neutral traps a major concern from a reliability point of view.

It has been established that the incorporation of hydrogen during thermal annealing can significantly improve its effectiveness. When sufficient active hydrogen species are introduced into the oxide, the temperature required to anneal out the radiation-induced charges and neutral centers can be significantly reduced. One way of introducing more hydrogen is to perform the thermal annealing in a high-pressure vessel containing hydrogen. This principle has been successfully demonstrated in some laboratory experiments.

In addition to thermal annealing and its variations, a technique based on the use of an RF plasma environment has been shown to be effective in removing the radiation-induced charges and traps in SiO$_2$ and at the SiO$_2$/Si interface. While it is a relatively low-temperature technique, it utilizes the photon energy in the plasma, coupled with the RF field, to drive the annealing process. It is particularly interesting to note that, in general, the RF plasma environment is a source of radiation damage, but under the proper setup and operated in the appropriate conditions it can be used to anneal out the radiation damage. The mechanisms involved in the RF plasma annealing process are in some ways similar to those involved in producing the ionizing radiation effects discussed throughout this book: the first step is the generation of electrons and holes in the oxide. The major difference is that in the annealing process the electrons and holes are driven by the RF field to recombine preferentially at defect sites, and the energy released by the recombination process subsequently drives the annealing process through the REDR (recombination-enhanced-defect-reactions) mechanism.

REFERENCES


