

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,

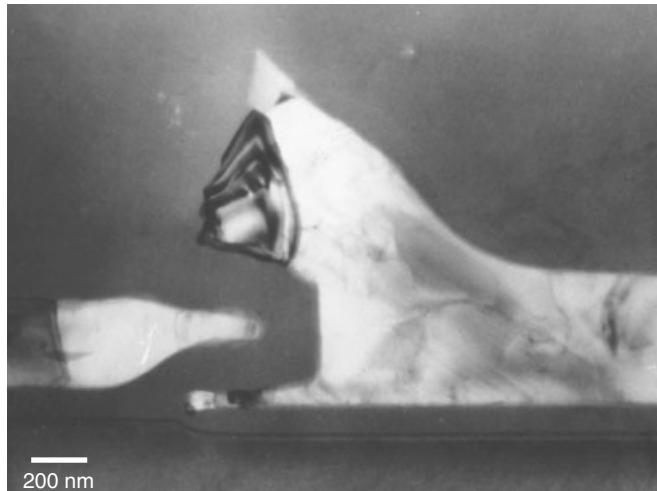
Singapore), and Dr. Ron Anderson (IBM, USA) for their technical advice. Our colleagues who helped in various experimental work include Wei-Kun Hung, Show-Ing Hsu, Ting Chou, Chin-Ting Chian, Ray-Chuan Chew, Su-Mei Chen, Chu-Mei Chang (ERSO/ITRI and Vanguard International Semiconductor Corp. Taiwan), 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,

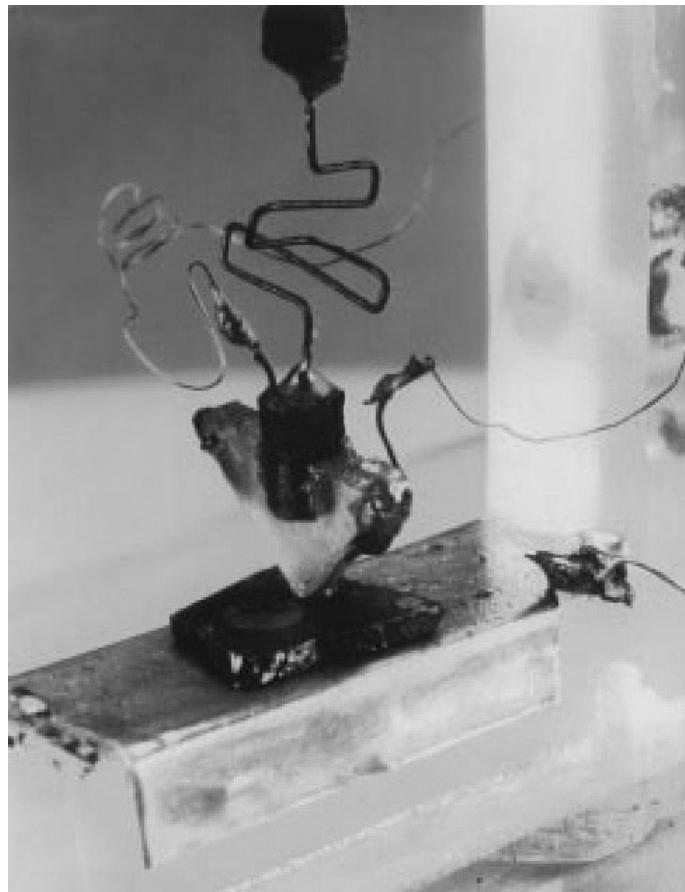


Figure 1.1 First transistor (point contact) in the world made by John Bardeen, Bill Shockley, and Walter Brattain in Bell Labs on December 16, 1947. (Reprint with permission from Lucent Technologies—Bell Labs)

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

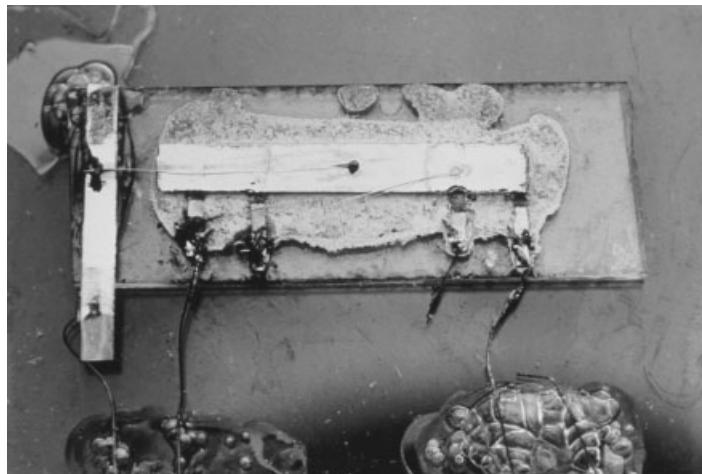


Figure 1.2 First integrated circuit (IC) in the world made by Jack Kilby in 1958. (Reprint with permission from Texas Instrument)

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

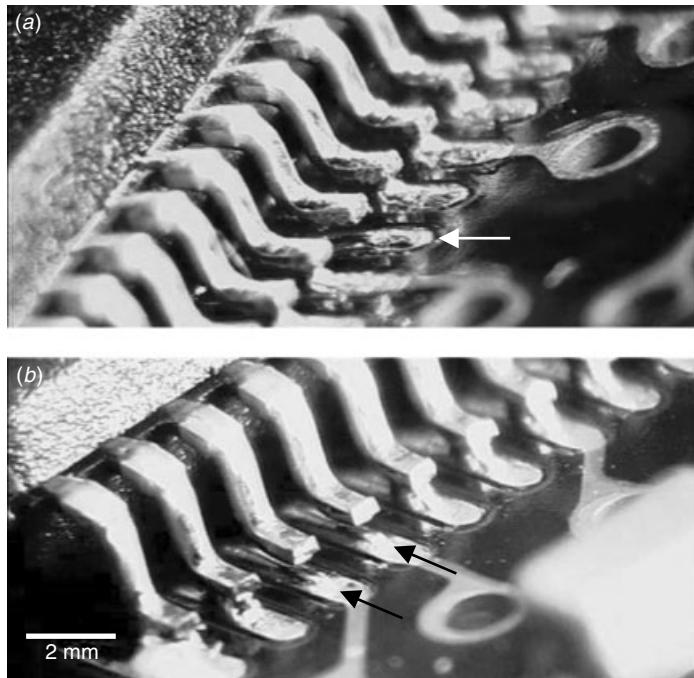


Figure 1.3 Optical stereo microscopic images show details of device lead soldering and soldering defects (as indicated) due to its large depth of focus and large depth of field. (Courtesy H. K. Hui, IME, Singapore)

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.

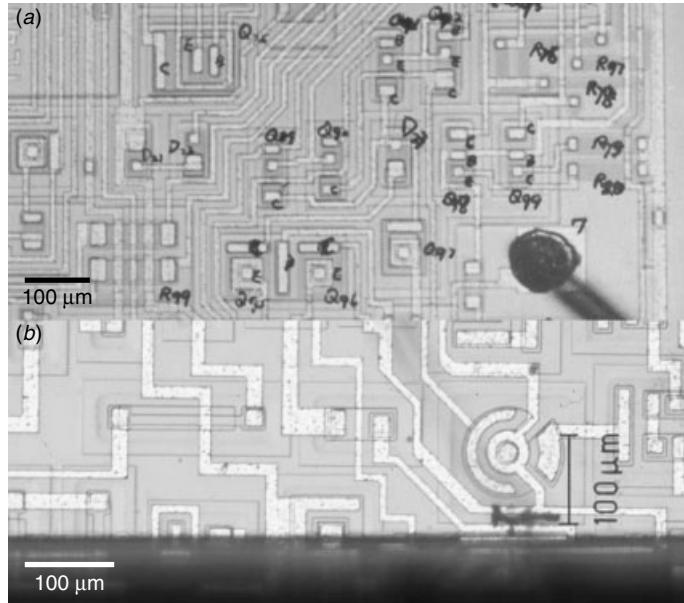


Figure 1.4 (a) Bipolar LSI devices with transistors and resistors indicated by handwritten Q (transistor) and R (resistor). Some of the transistor's electrodes, emitter (E), base (B), and collector (C) are also indicated. (b) Sample cross-sectioned (lower dark part) to reveal the device's structure.

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:

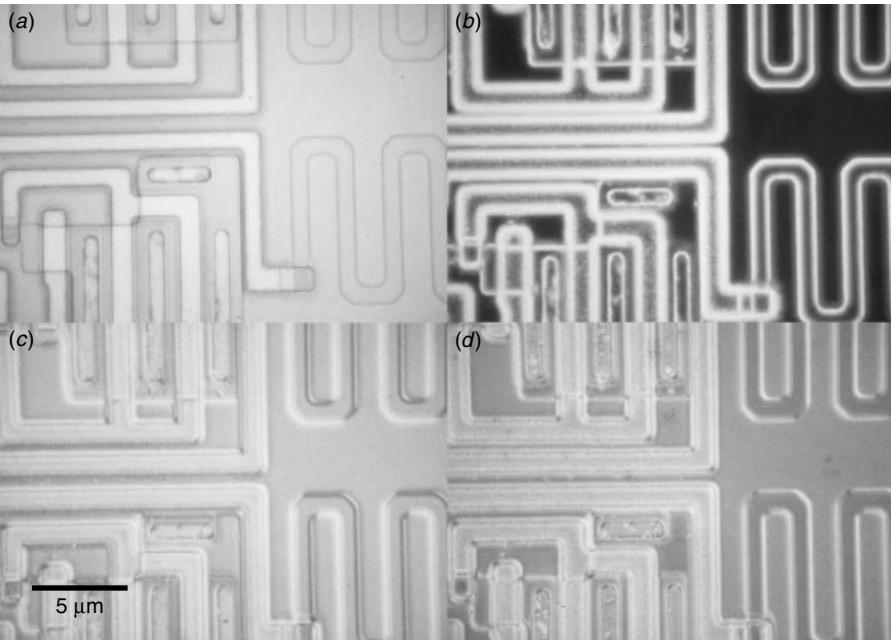


Figure 1.5 Optical images using various contrast: (a) Normal bright field, (b) dark field, (c) Nomarski I, (d) Nomarski II. The different contrast modes reveal different surface details of the sample. (Courtesy S. Y. Hsu, ERSO/ITRI, Taiwan)

- 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.

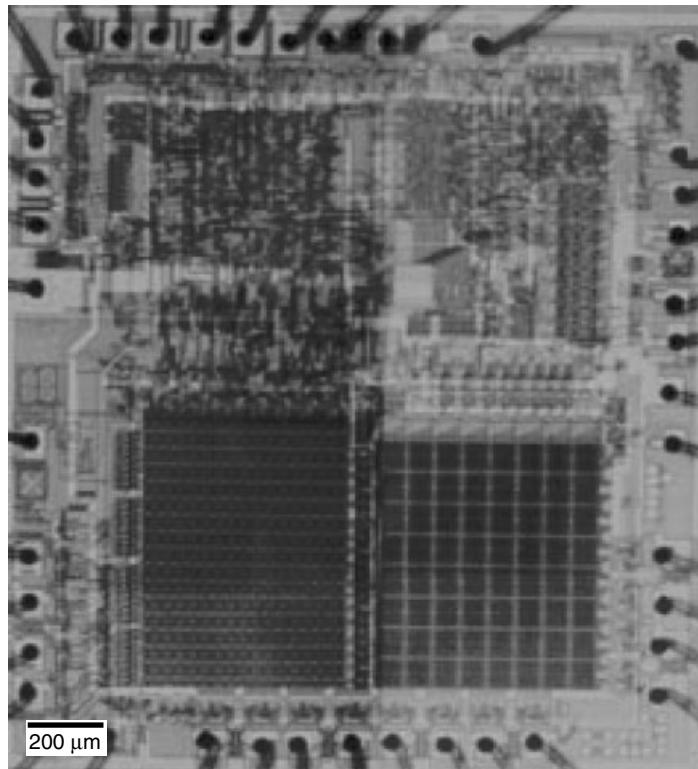


Figure 1.6 A VLSI device as seen with 0.5 μm process technology. Several distinctively different areas are apparent under the light microscope but it is nearly impossible to see details on the device using optical microscope due to the resolution limits of the visible light source.

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

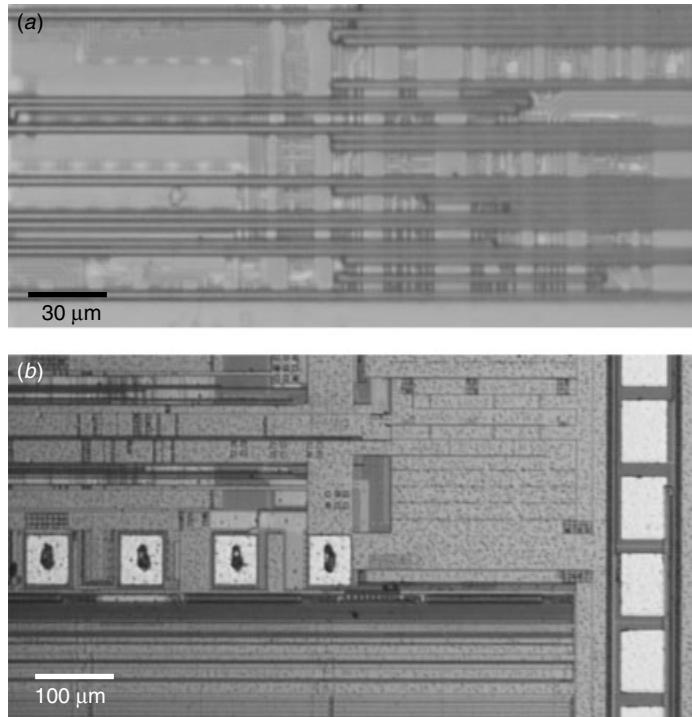


Figure 1.7 Besides the resolution issue another reason the light microscope is rarely used in analyzing the modern device is its multiple layers of interconnection. (a) high magnification optical image as observed under a light microscope showing multilayer interconnections. The circuit information can 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

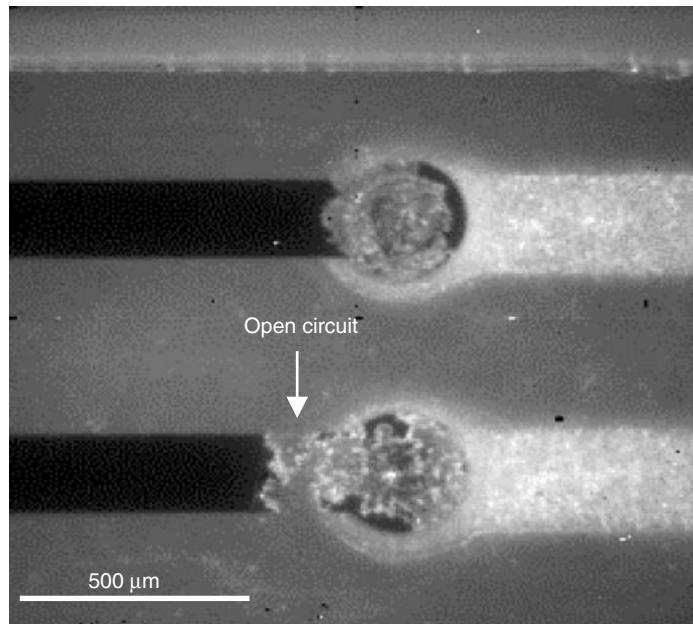


Figure 1.8 IR microscope imaging the flip-chip on board (FCOB) device's corrosion failure where the sample preparation is not needed and thus no damage or alteration of the failure mechanism. (Courtesy Dr. Alastair Trigg, IME, Singapore)

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

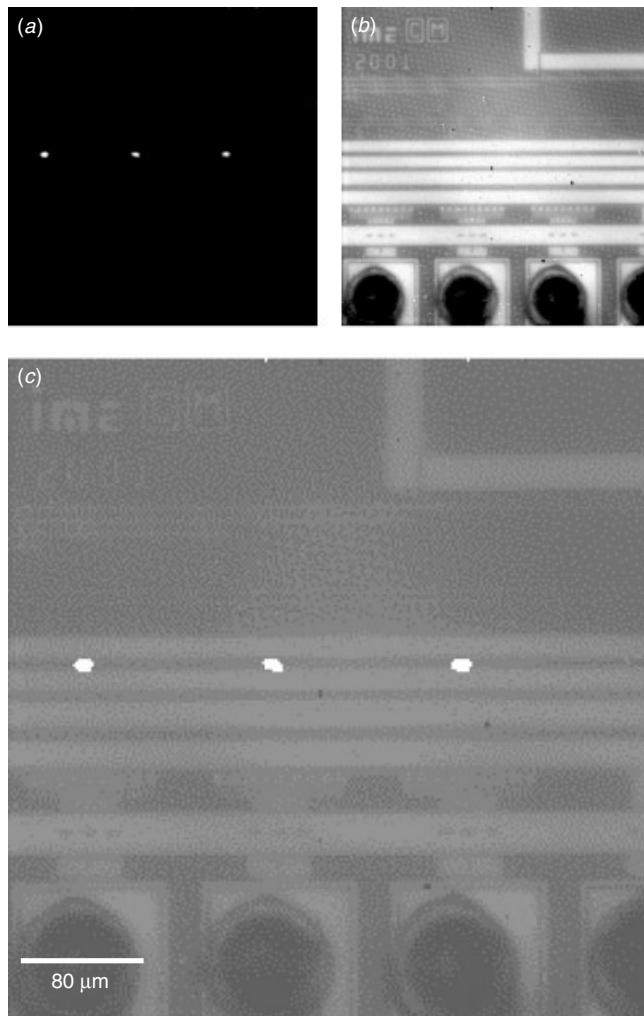


Figure 1.9 IR photo emission microscope reveals circuit hot spots where overlap occurs with the optical microscopic image and thus shorted diffusion (through salicide layer) between signal and 3V V_{DD} . (a) IR photon emission, (b) light microscope image, (c) combined image to show the leakage spots.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

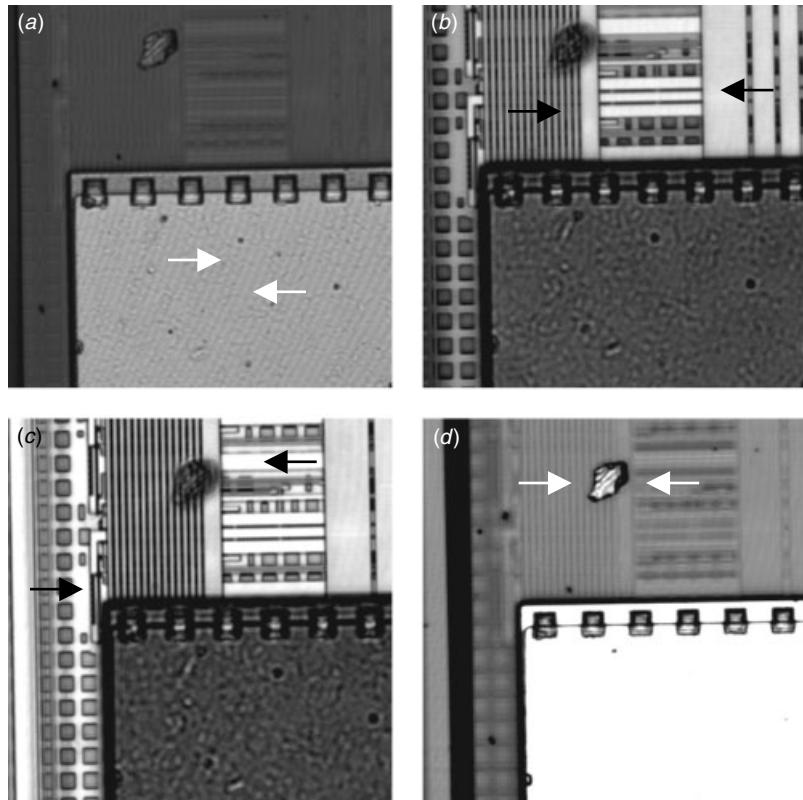


Figure 1.10 Laser confocal microscopic images. The arrows indicate the areas under focus. (a) bond pad, (b) intermediate metallization layers, (c) bottom metallization layers, (d) surface particle.

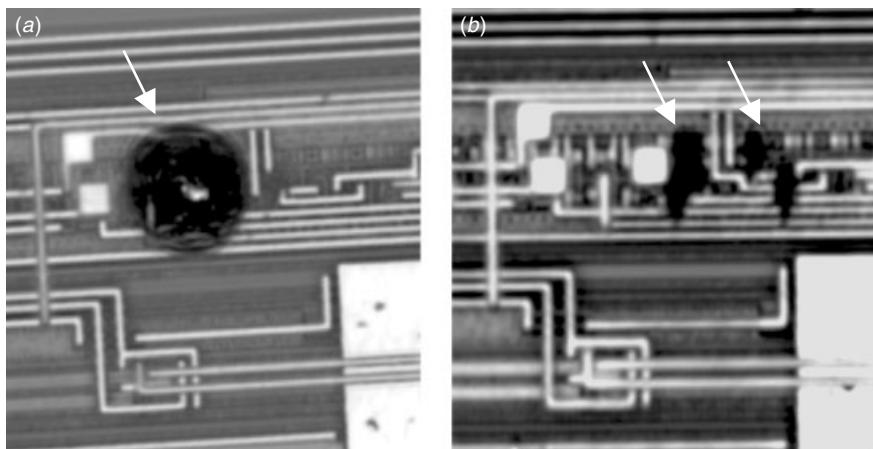


Figure 1.11 Laser confocal microscopic images from the chip's backside. The arrows indicate ESD/EOS damage spots. (Courtesy W. K. Ong, IME, Singapore)

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,

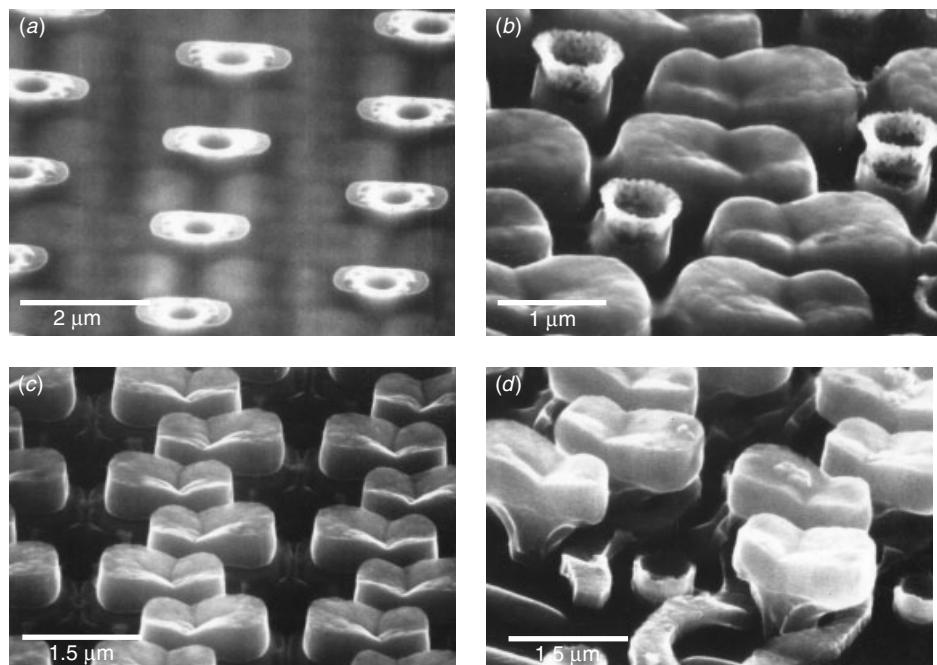


Figure 1.12 SEM images of a DRAM device with the layers removed. (a) Bit-line contacts, (b) bit-line contacts and stack capacitor, (c) stack capacitors, (d) stack capacitors, bit-line contact, and word-line polysilicon. (Courtesy Ying Chu Lee, VISC, Taiwan)

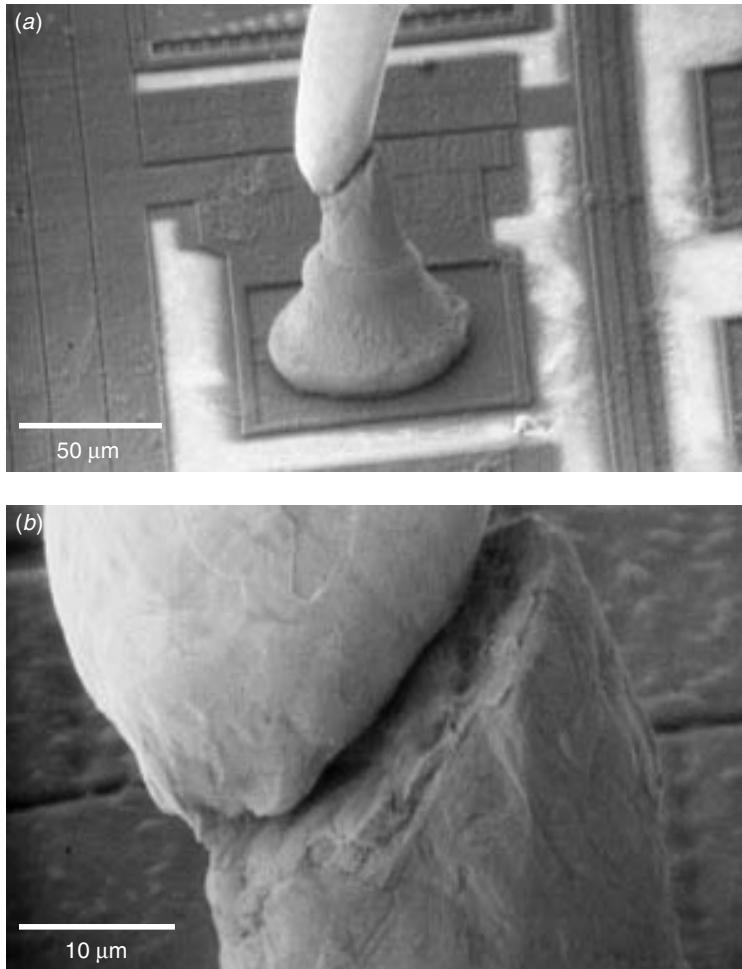


Figure 1.13 SEM images of a fractured gold wire. A typical failure analysis case uses SEM to observe topological features as no other techniques are as capable of revealing these details.

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.

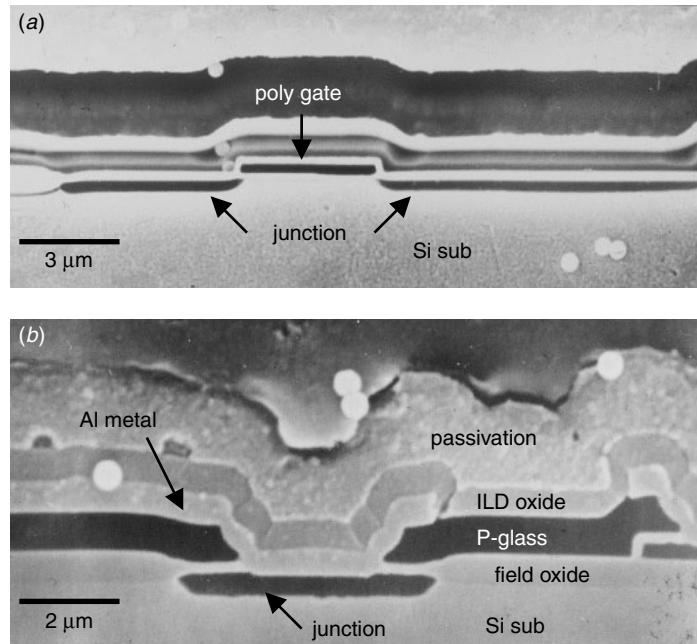


Figure 1.14 SEM cross section of an older device. The cross-sectional surface was polished and chemically stained to reveal the junctions and doped glass (dark areas). The small bright spheres are latex disks used as internal dimension standards. (Courtesy Showing Hsu, ERSO/ITRI, Taiwan)

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

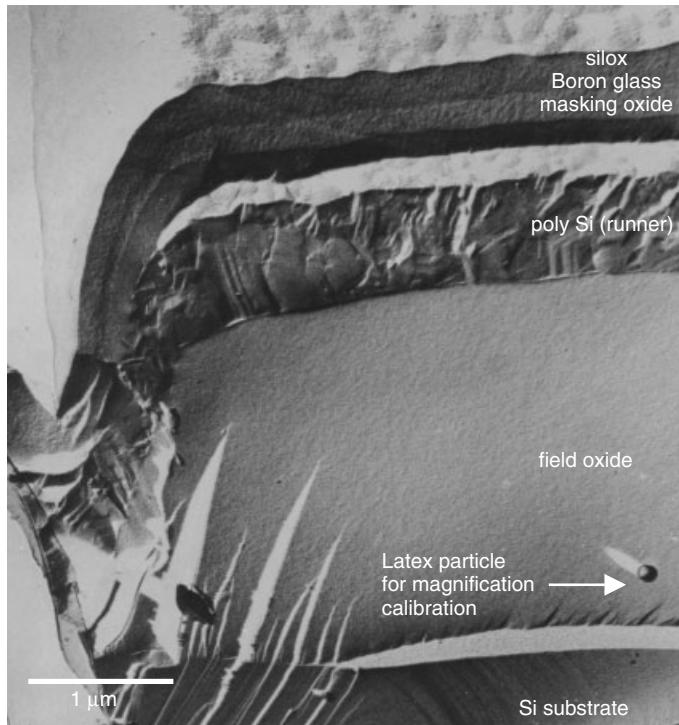


Figure 1.15 The very first cross-sectional TEM image of a silicon LSI device. The image was taken by George T. T. Sheng of Bell Labs in 1979 using the carbon replica technique.

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.

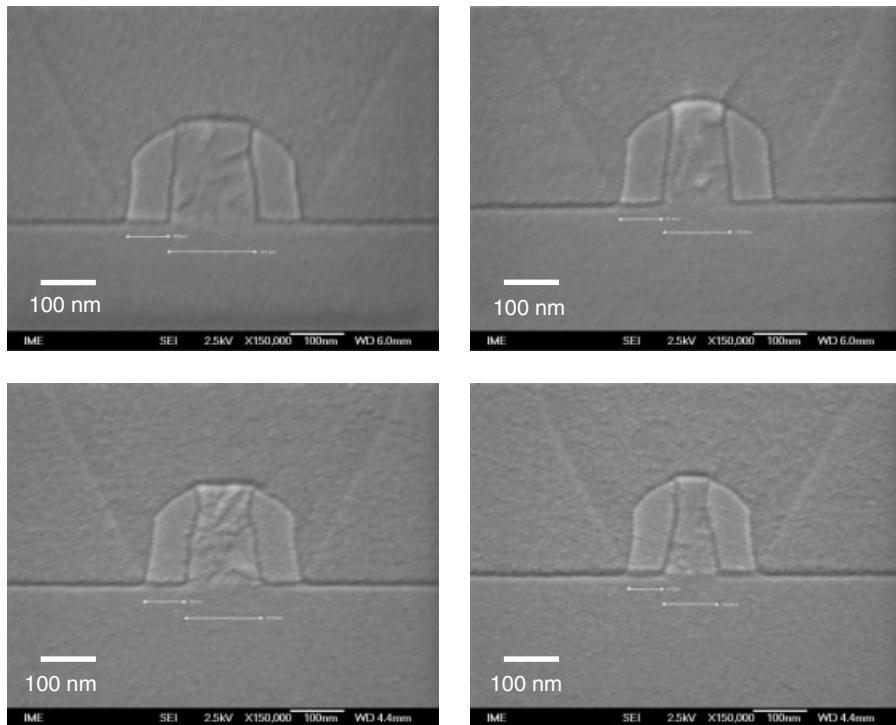


Figure 1.16 In-line SEM cross-sectional view of gate profiles. Details of the gate's morphology can be seen clearly and its dimensions measured, although some details, such as the gate's shape at the bottom corner and gate oxide's thickness, are not visible. (Courtesy Ranganathan Nagarajan, IME, Singapore)

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.

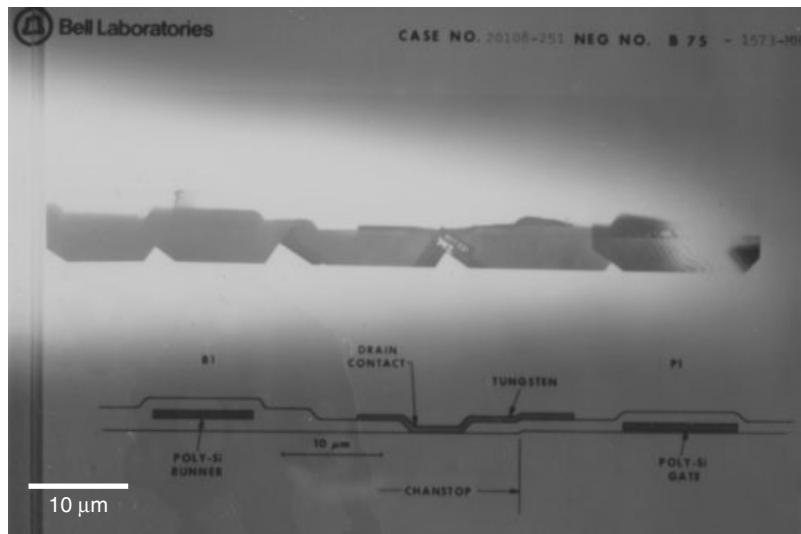


Figure 1.17 The first TEM cross-sectional image of a MOSFET device. The device's technology is 10 μm . The image is a montage of five micrographs. A schematic corresponding device is shown also. This work was done in Bell Labs in 1979 by George T. T. Sheng.

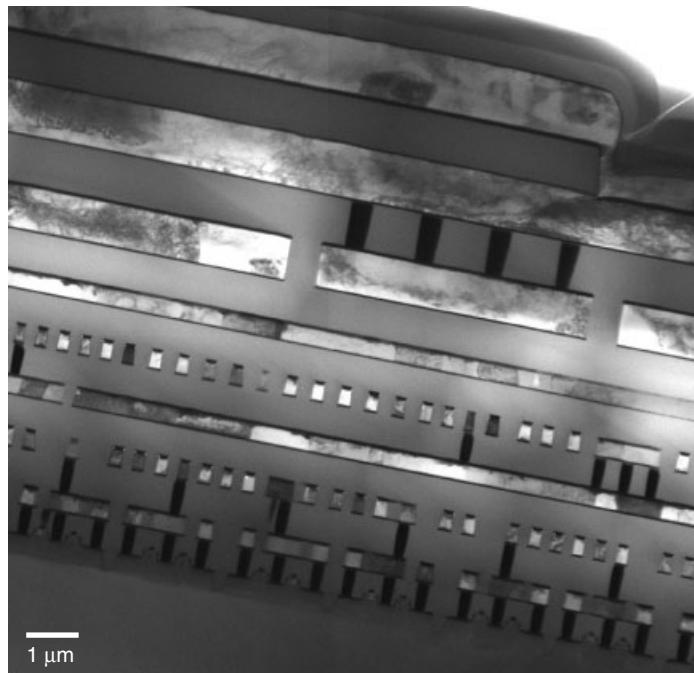


Figure 1.18 A more recent cross-sectional TEM image of a ULSI device. This is 8-metal layer technology and physical gate length is 0.17 μm . The first five metal layers are for signals (thinner) and top three metals (thicker) are for power and ground.

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

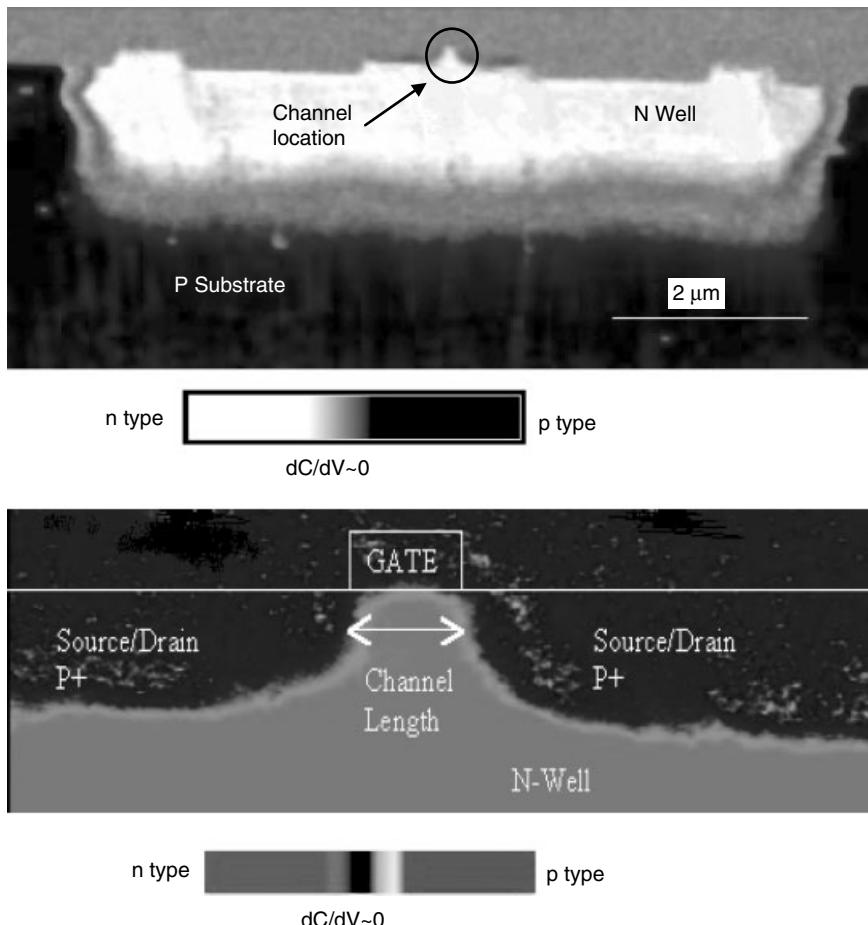


Figure 1.19 SCM image of a P-channel MOS (0.15 μm technology). The image shows a clear delineation of n- and p-type regions. (Courtesy Vennisa Lim and Dr. Alastair Trigg, IME, Singapore)

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

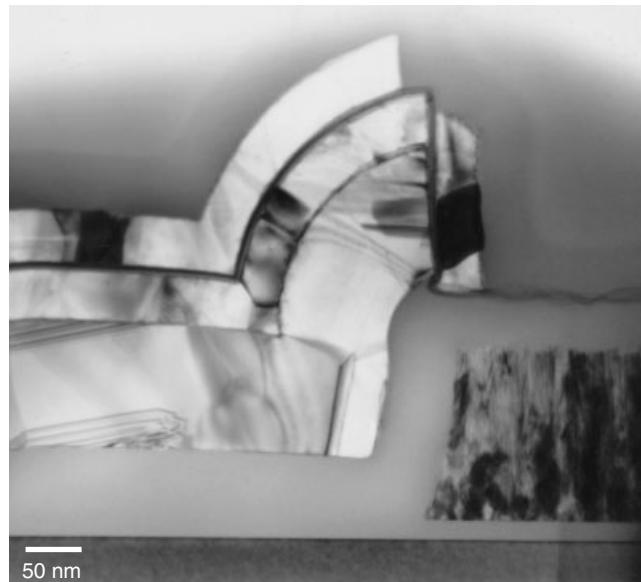
One of the most important assets of STM is its adaptability to various measurement needs for applications in semiconductors. More derivative techniques have recently been developed that allow physical properties to be visualized that had in the past either not been accessible at all or at least not for 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.

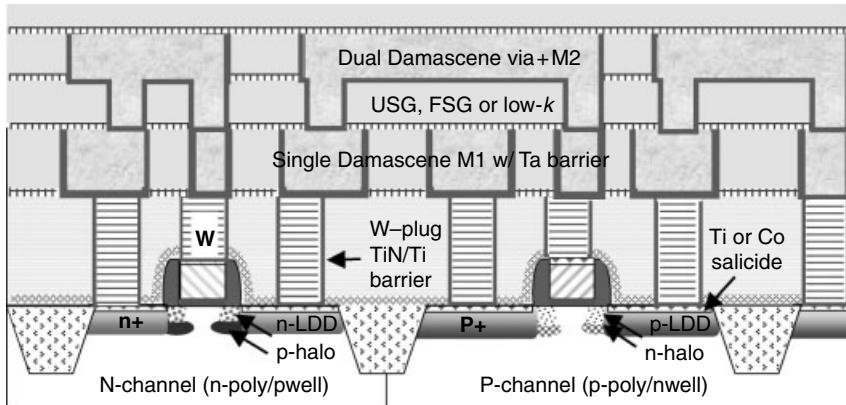


Figure 2.1 The basic plan of a contemporary VLSI CMOS process. The multiple interconnection layers often run up to seven or eight layers, though only two are shown here. Co salicide junctions, Shallow trench isolation, LDD structure, W-plug contact, damascene and dual-damascene Cu metallization, low-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

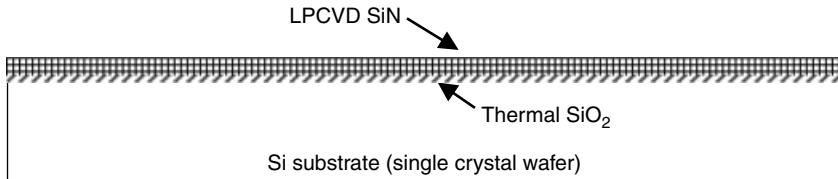
Figure 2.1 shows a schematic cross-sectional view of a modern CMOS process device. The typical process characteristics are as noted below:

- 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):

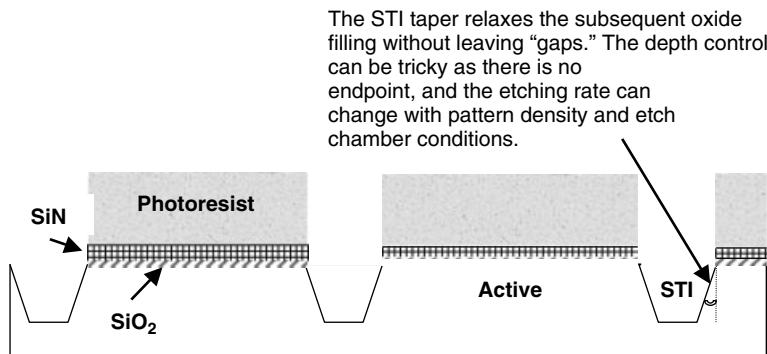
Step 1. Thermal $\text{SiO}_2/\text{LPCVD SiN}$ for STI Formation (Fig. 2.2). Formation of shallow trench isolation (STI), or in the old process local oxidation (LOCOS) is the first step in the whole process. Active or isolation areas are defined, beginning with the silicon nitride (SiN) layer deposited 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.

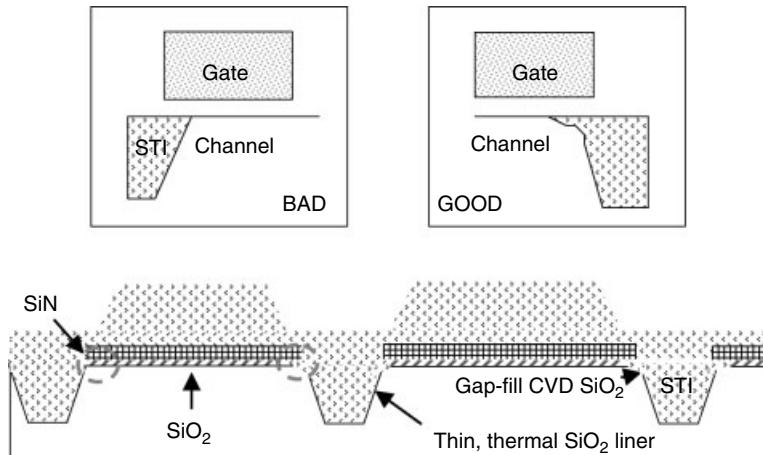


Figure 2.4 Diffusion clean surface with partial SiN undercut, on which is grown the thermal oxide trench liner and then the gap-fill CVD oxide deposited. The undercut of SiN results in a rounded top to the STI after the liner's oxidation, and this reduces the electric field when the subsequent gate overlaps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

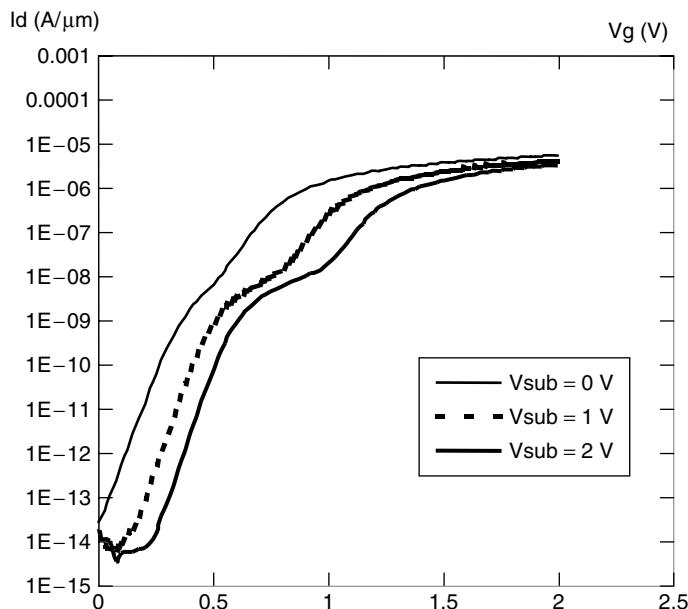


Figure 2.5 STI profile can affect I-V behavior. Gate characteristics of p-MOSFET under different substrate bias showing the subthreshold hump. The polygate overlaps the STI edge, concentrating the fields of the parasitic transistor. The STI top corner's rounding is one way to reduce the effect. (Courtesy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

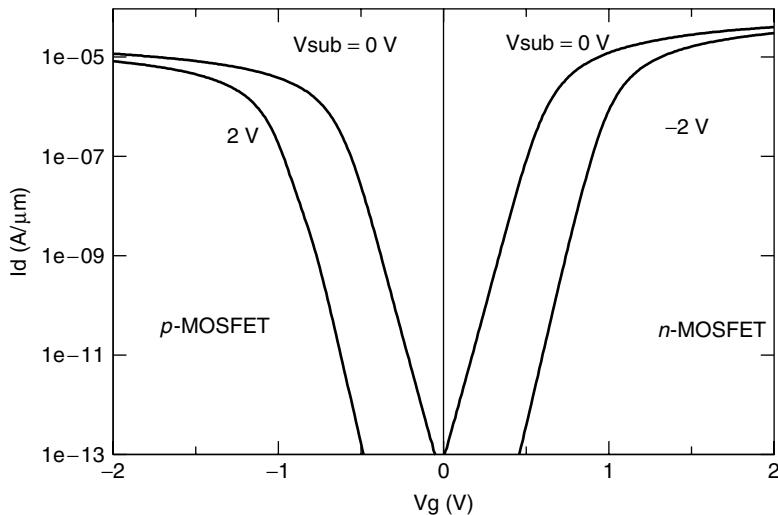


Figure 2.6 Gate characteristics of *n*- and *p*-MOSFETs in a wafer with a pad-oxide undercut etch. (Courtesy N. Balasubramanian IME Singapore)

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

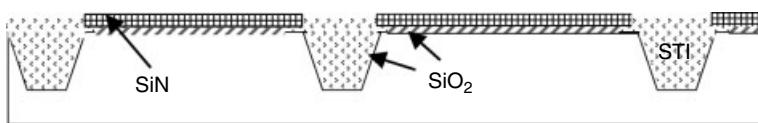


Figure 2.7 Chemical mechanical polish (CMP) of the excess oxide with a stop on SiN, producing a somewhat planar profile. The wide and narrow features have different starting topographies, so they complicate the CMP process. Some assistance comes from adding dummy (nonfunctional) active features to the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

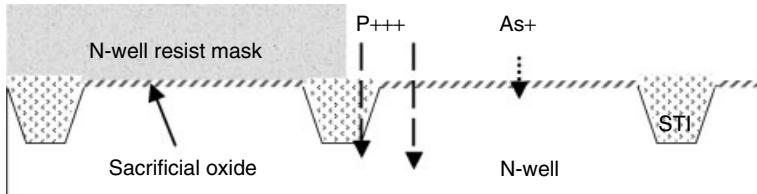


Figure 2.8 SiN and SiO_2 are wet chemically removed and a fresh sacrificial oxide is grown to serve as an implant screen. The high-energy well, shallow threshold (V_t), and punch-through ion (donor or n-type) implants are effected after the N-well resist masking to protect the P-well side. Very high 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

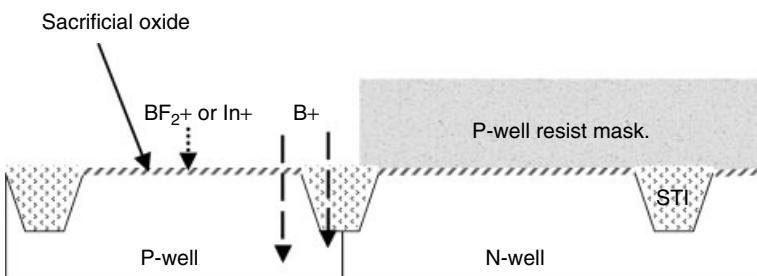


Figure 2.9 A fresh sacrificial oxide is grown to serve as an implant “screen.” N-well resist is stripped, and this is repeated for P-well using acceptor (p-type) ions.

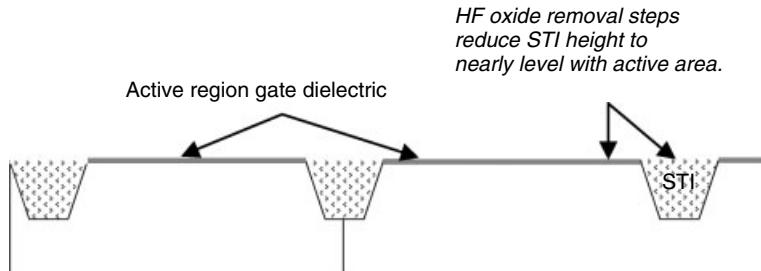


Figure 2.10 SiO(N) gate dielectric(s) are grown after the sacrificial oxide is stripped. In dual-voltage devices, two gate oxide thickness are targeted: first a thicker high-V oxide, which is masked and HF is removed in the low-V areas before the thinner SiO(N). Nitrogen can block boron from P-poly diffusing 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.

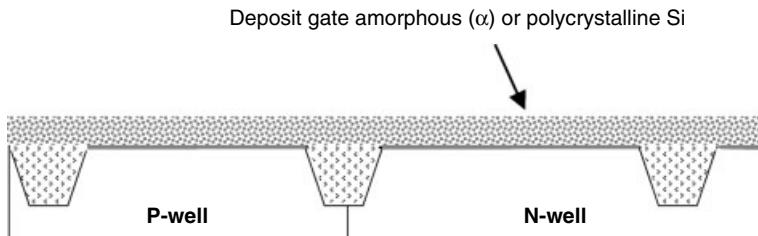


Figure 2.11 The amorphous Si condition is sometimes preferred in the CD and edge profile of the etched gate. There is some debate whether α-Si offers any benefit in gate oxide or gate doping.

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.

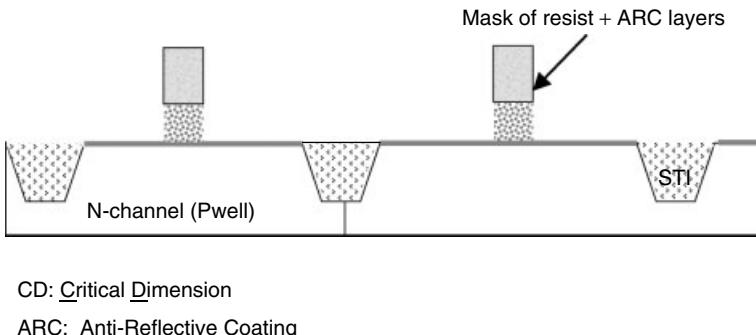


Figure 2.12 Resist mask gate, etch α /poly–Si stopping on a thin gate oxide. The ARC layer deposited over the α /poly improves CD control. The photomasks must be selectively sized and shaped for optical proximity corrections (OPC) or phase-shifting features for printing critical dimensions.

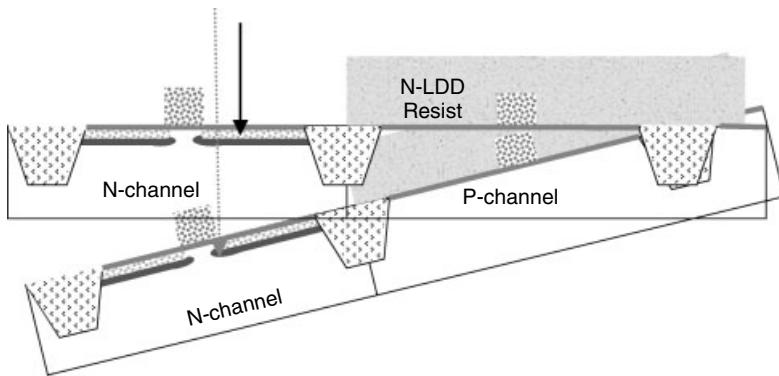


Figure 2.13 After the gate poly resist is stripped, and, a thin implant screen oxide grown, the masked nLDD (As)/“halo” (B) implants can begin. During the implants the wafer’s tilt can be adjusted to optimize the dopant’s profile under the gate edge without the need for thermal diffusion. The trade-offs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

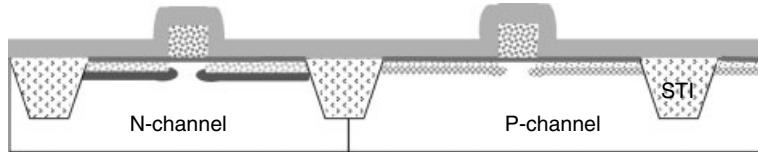


Figure 2.14 Masking and implant for P-LDD (BF_2)/halo (P) implants. Strip resist, deposit of thin oxide, and/or nitride films.

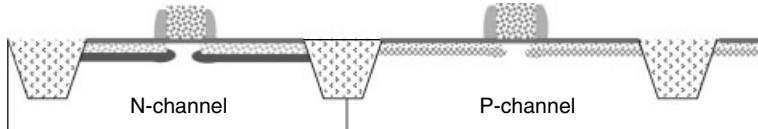
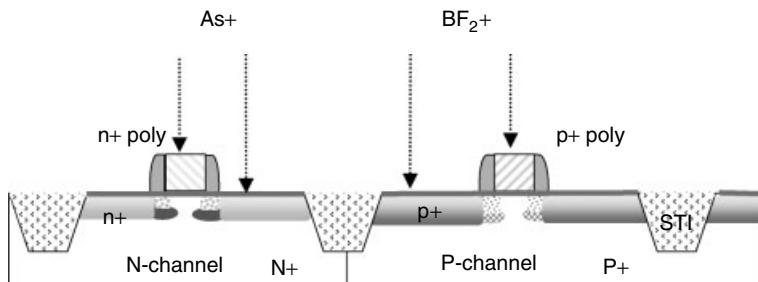


Figure 2.15 Anisotropic spacer etch deposit of gate sidewall spacers. The spacer is self-aligned to every polygate feature and separates the next implant (high dose for low resistance) away from the channel's edge.



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

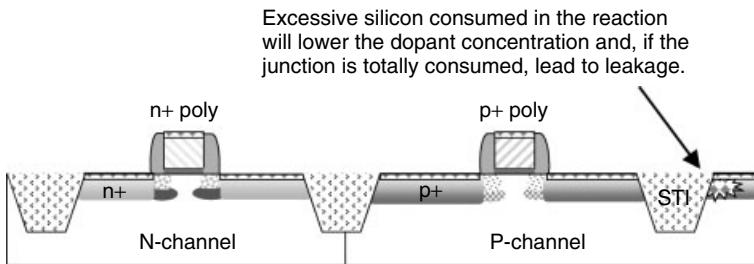


Figure 2.17 Source-drain and poly regions can be silicided (optional SiO/SiN dep and masking of select areas for silicide exclusion). An implant is often used to assist metal's reaction with silicon. The process involves cleaning the thin oxide, metal deposition (e.g., Ti and/or Co), implant (e.g., Si,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

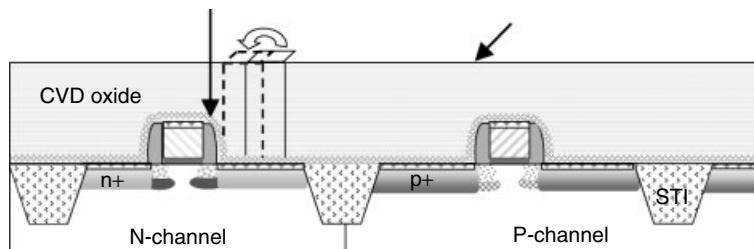


Figure 2.18 CVD oxide(s) and nitrides are deposited. and planarized by CMP. Nitride layers can serve as an oxide etch-stop in marginally aligned contacts. CMP planarization is critical in capturing small features but is limited by the lens' depth of focus.

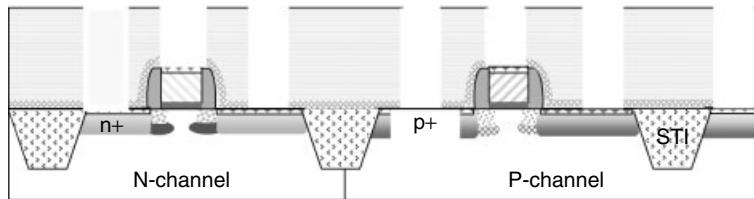


Figure 2.19 Contact masking, etching (using the SiN as an initial etch-stop followed by later removal), and cleaning.

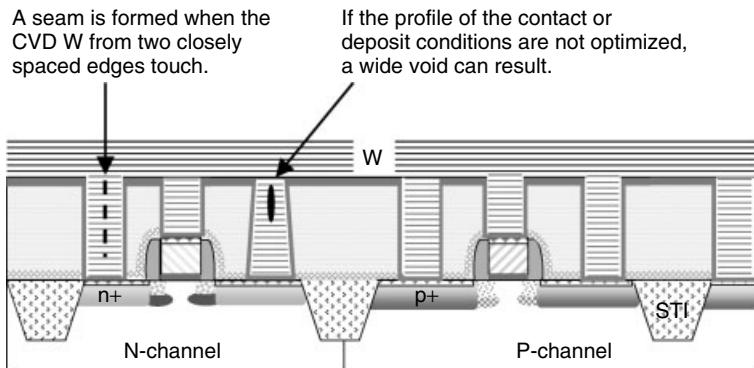


Figure 2.20 Contacts are metallized with combinations of sputtered and CVD barrier/seed metals (e.g., Ti/TiN) prior to the gap-filling CVD tungsten (W).

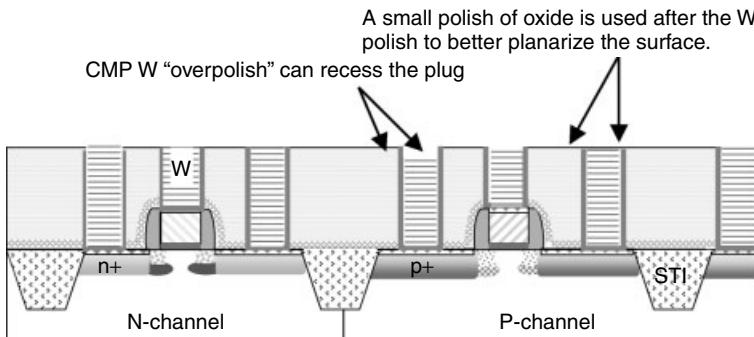


Figure 2.21 Contacts are left “plugged” with tungsten after the CMP removes the excess metal (W and barrier).

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.

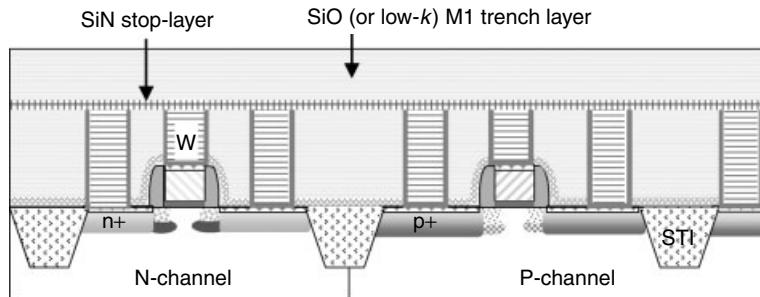


Figure 2.22 The first routing metal-line process begins. For aluminium, deposition of barrier and Al is done. For Cu damascene interconnects (illustrated), dielectric SiN stop-layer and oxide (or low- k) are deposited first.

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

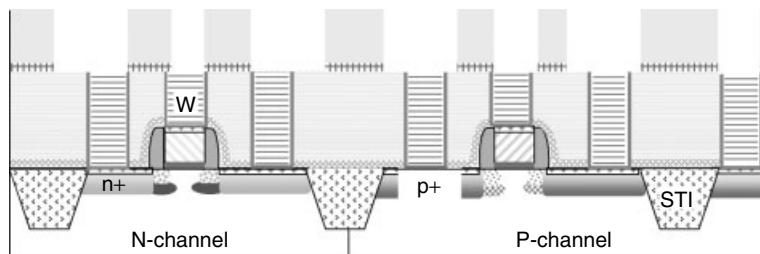


Figure 2.23 Cu will be filled into line trenches created by masking and the oxide/nitride etch. The trench mask has the reverse pattern of the desired final metal-line pattern. Nitride allows the oxide etch to “stop” in accord with a etch rate selectivity feature.

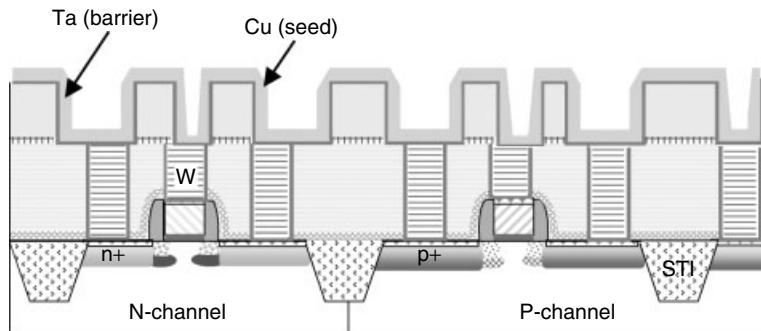


Figure 2.24 Cu process begins by removing surface oxide from underlying metal. Then a barrier film (e.g., Ta or TaN) is deposited to prevent diffusion into the oxide, and next a thin Cu seed layer, which will be contacted during subsequent electroplating. Both layers are typically sputter processes, 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).

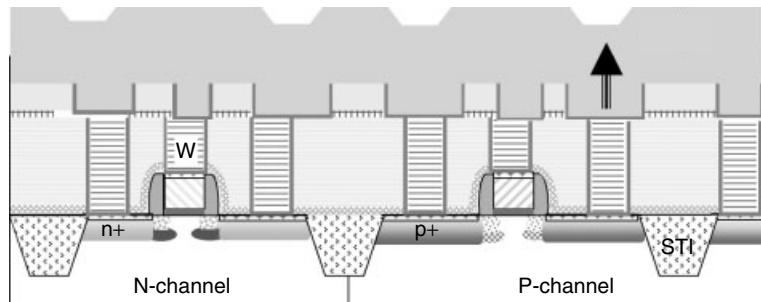


Figure 2.25 Electroplating Cu completes the fill of the line trenches. Various chemistry and flow/pulsing techniques are used to prevent gaps and excess topography. The idea is to plate bottom-up, with additives suppressing the plating rate on the elevated field areas.

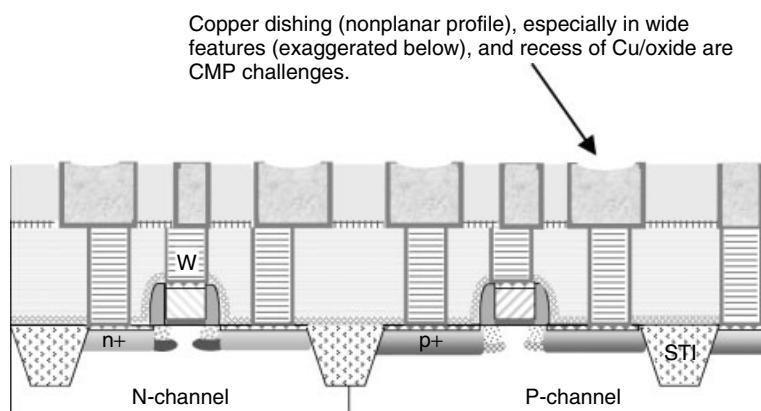


Figure 2.26 Copper grain size and polishing characteristics brought into a stable condition by use of an oxygen-free (N_2 or N_2/H_2) furnace anneal. CMP removes excess Cu and barrier.

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.

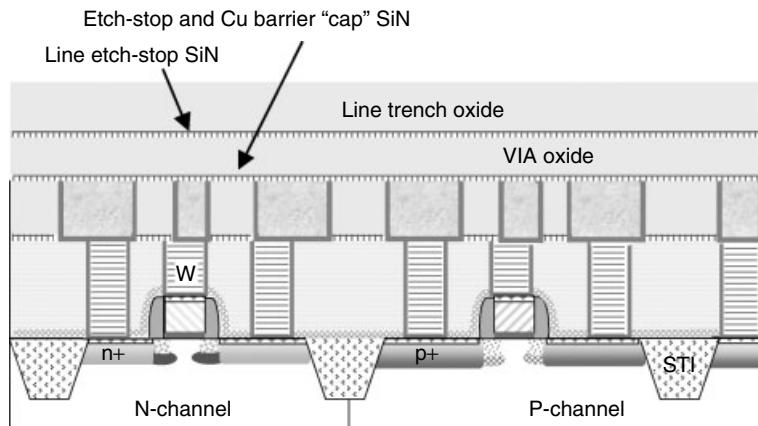


Figure 2.27 Subsequent layers of Cu proceed with both VIAs and line trenches etched before the simultaneous metal fill (dual damascene). Dielectric layers of SiN and SiO_2 are typically used for defining the VIA.

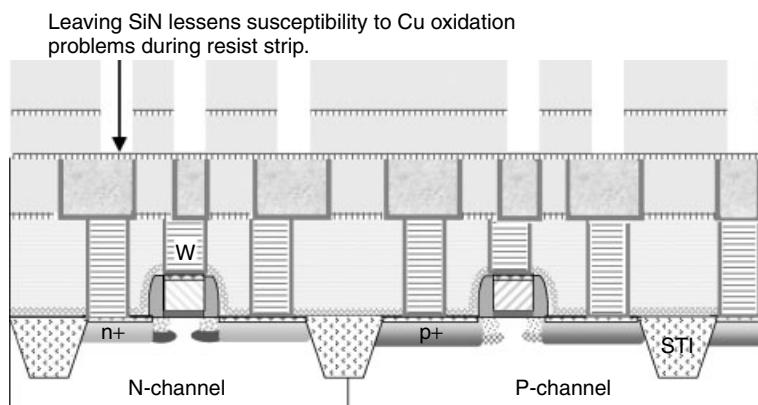


Figure 2.28 Patterning with a VIA mask first (before the line trench mask), as is common.

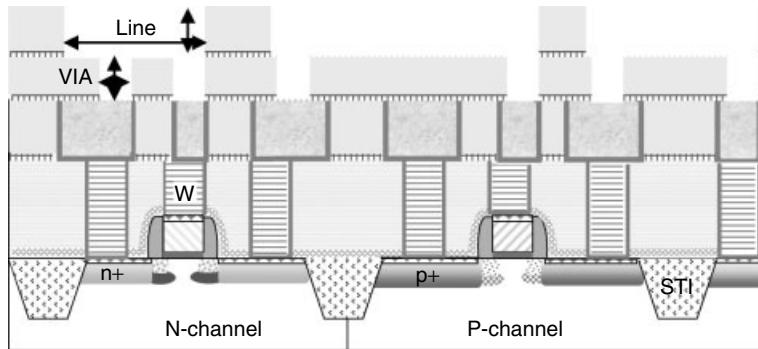


Figure 2.29 Resist masking and etching the line trench oxide followed by resist strip and nitride cap + etch-stop removal.

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

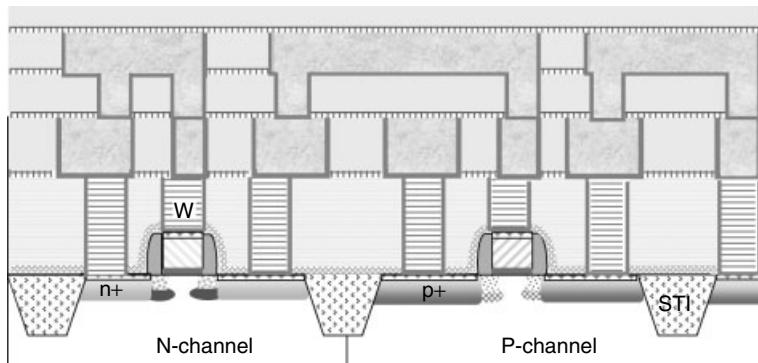


Figure 2.30 CuO removal followed by barrier and seed Cu deposit, electroplating, anneal, Cu CMP, and cap nitride layer. Passivation for bonding/assembly usually adds additional oxide/nitride and Al pad metal.

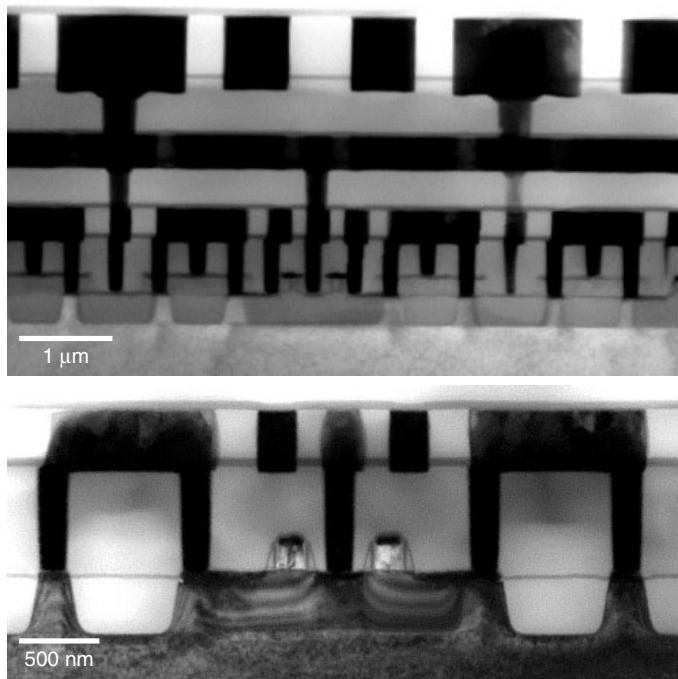


Figure 2.31 Cross-sectional TEM of a $0.13\text{ }\mu\text{m}$ technology device using Co salicide junctions, LDD structure, shallow trench Isolation, W–plug contacts, damascene and dual damascene full Cu metallization. As we illustrated previously by our process flow, these are the core technologies.

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.

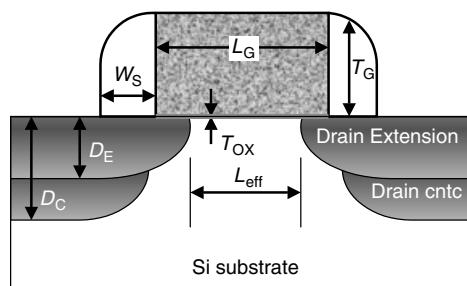
Front End-of-Line (FEOL) MOSFET

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

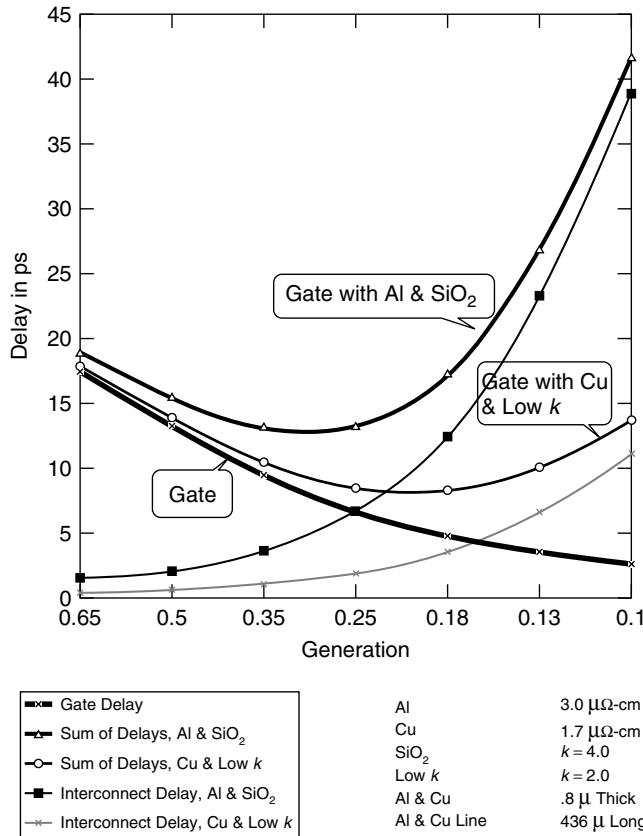


Figure 2.33 Resistance and Capacitance product ($R \times C$) can be the speed-limiting factor in circuit performance for deep submicron circuits. (M. Bohr, *IEEE IEDM Tech. Digest*, 241–242, 1999. Reprint with permission from IEEE)

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).

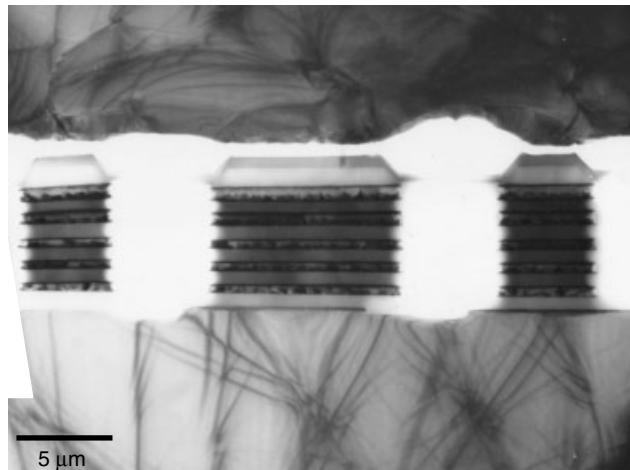
- Future lithographic technology: phase-shift masks and Ar-F 193 nm scanner.
- Implants at sub 1keV or heavy atoms (Sb, In) to reduce junction depth.
- Increasingly lower time-temperature diffusion using rapid thermal anneal (RTA), fast ramp furnaces and lower temperatures.
- CVD processes; CVD or modified 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

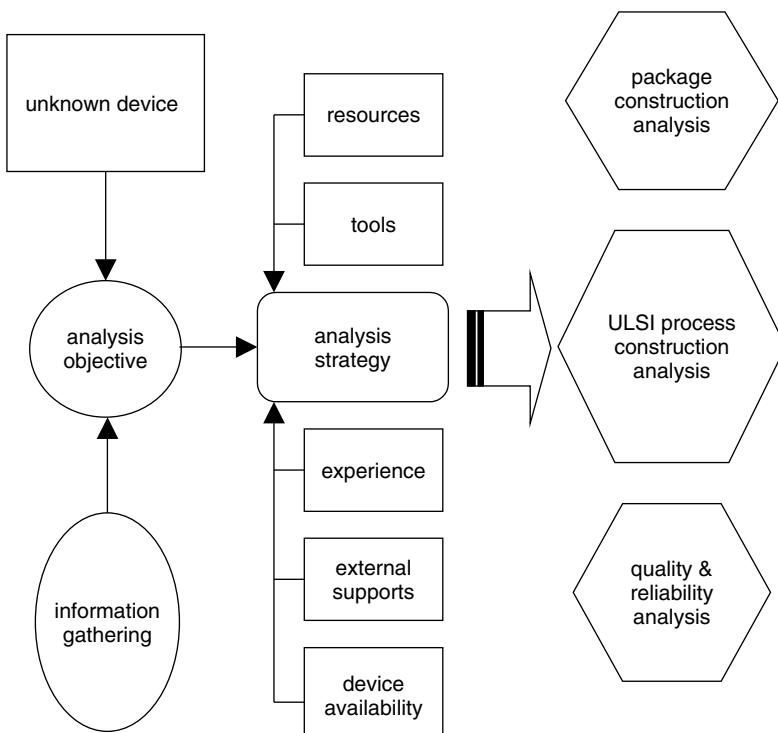


Figure 3.1 Flowchart used in construction analysis.

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.

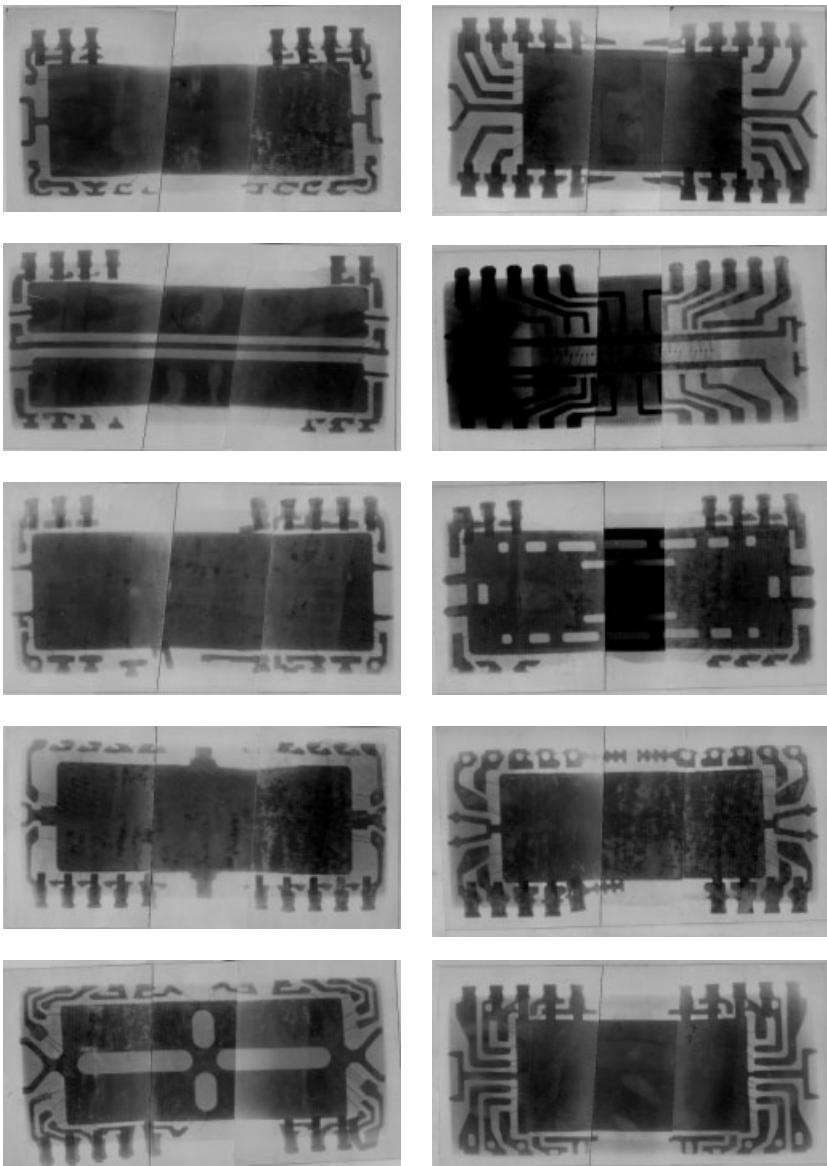


Figure 3.2 Soft X-ray images of the different 300 mil 28 leads of SOJ DRAM packages. The leadframe skeletons and wire bond design can thus be compared and examined in detail.

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

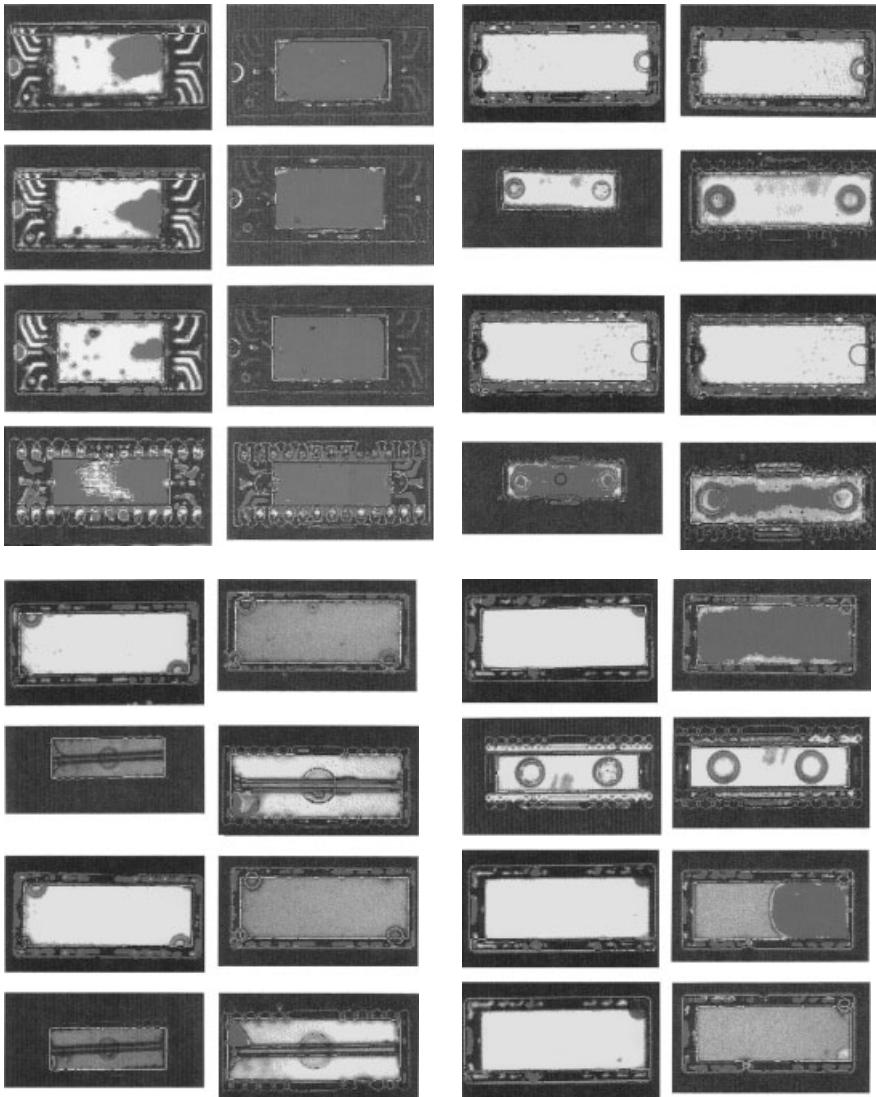


Figure 3.3 Scanning acoustic microscopy images of the front sides and back sides of various packages after different package reliability tests. By combining the results with X-ray images, the effectiveness of different leadframe design can be compared directly. Dark color indicates delamination within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

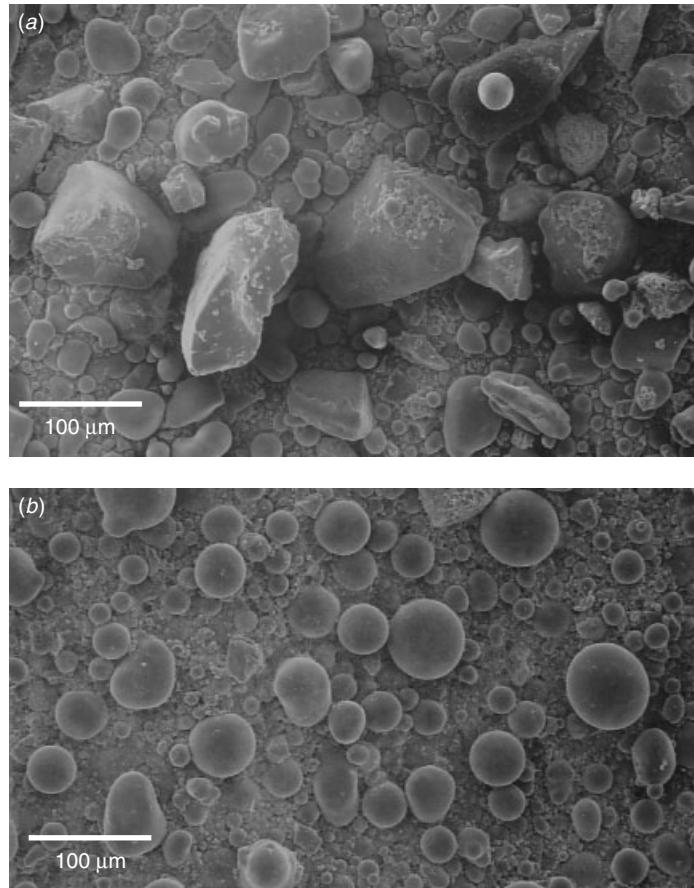


Figure 3.4 SEM images of the same molding compound but different lots show distinctively different filler shapes and sizes. The discrepancies are found to result in different reliability test results of final DRAM products.

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,

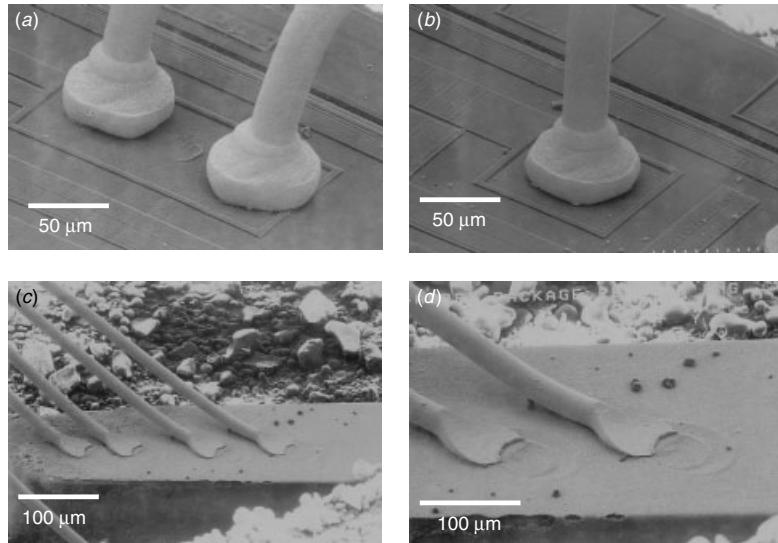


Figure 3.5 SEM image of wire bonds. Wire bonding analysis includes ball bond (first bond) as seen in (a) and (b), and stitch bond (second bond) as in (c) and (d). Abnormal shape, necking, stitch bond foot thinning, and so on, need to be noted and recorded during construction analysis.

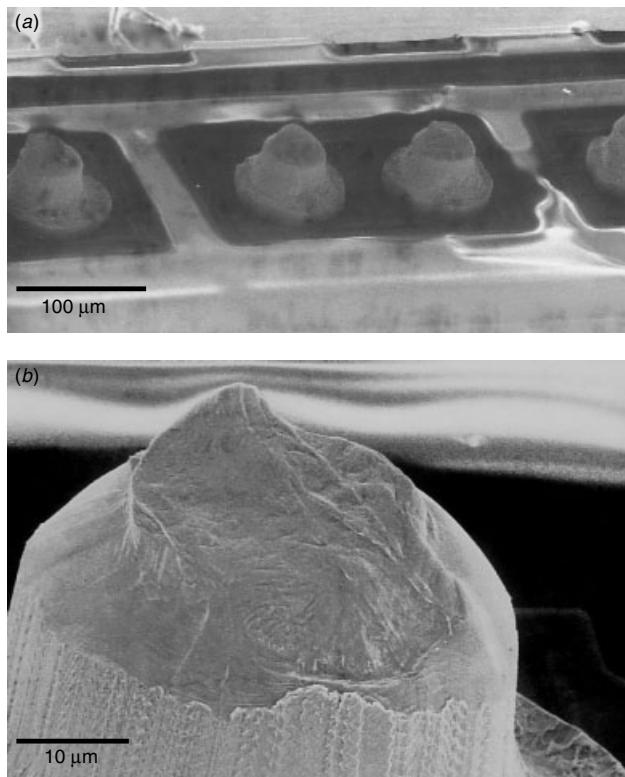


Figure 3.6 Wire bond quality examined by pull test and subsequent failure mode analysis using SEM. Healthy wire bonding should show failures as seen here.

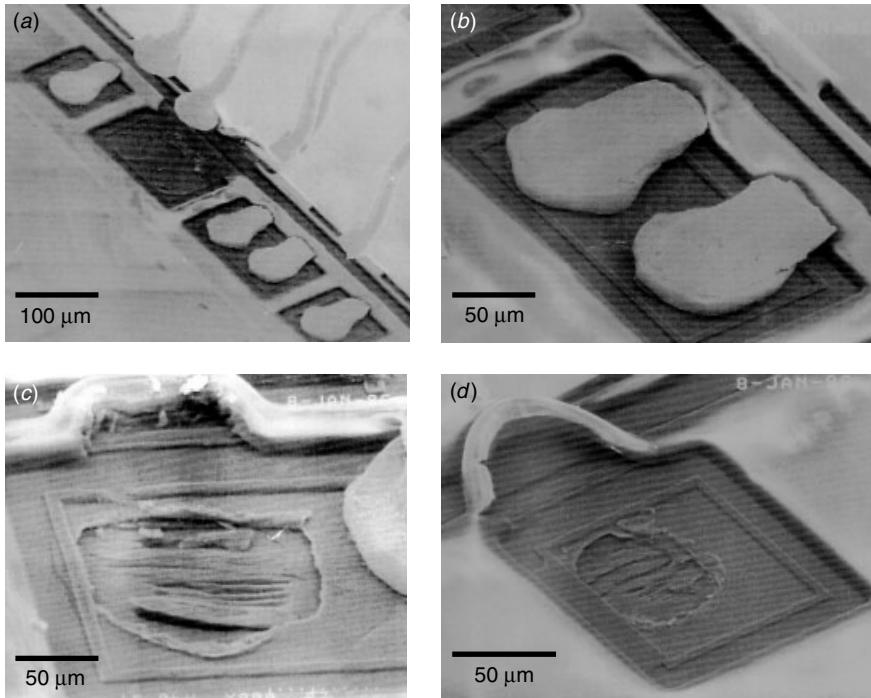


Figure 3.7 Wire bond quality was examined by ball shear tests and subsequent failure mode analysis. Normal failure mode (*a* and *b*) shows failures within gold balls, and abnormal or weak ball bonding (*c* and *d*) failure modes shows failures at the Au/Al interface or below.

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

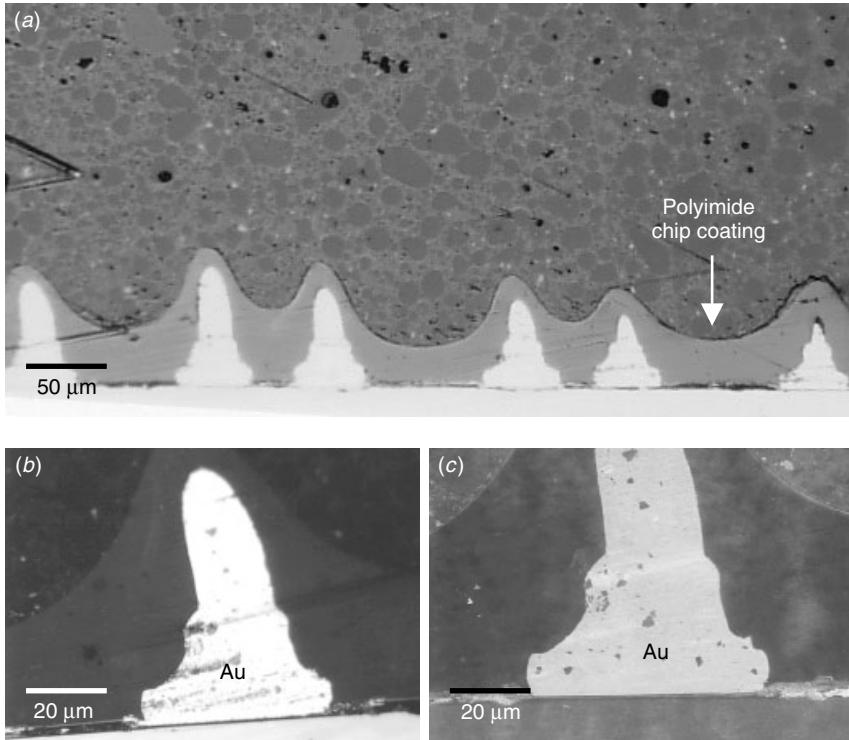


Figure 3.8 High-magnification SEM cross sections of the package device along an array of gold ball bondings. (a) Polyimide chip coating done after wire bonding is shown clearly. (b,c) Close-up at the Au ball bonding showing ball shape and Au/Al interface intermetallic and ball bonding quality. Kirkendall 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

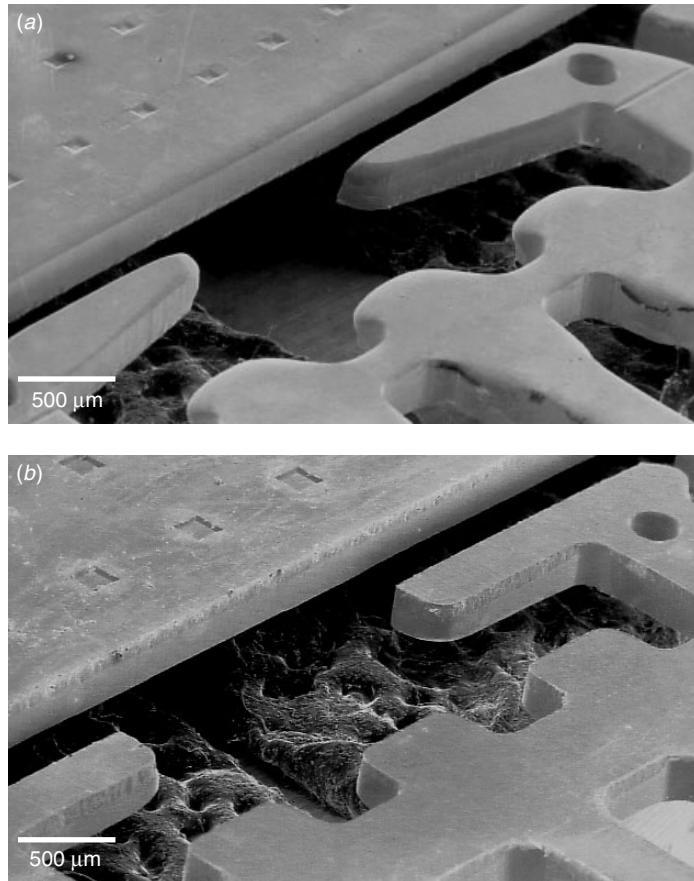


Figure 3.9 SEM images of the same leadframe design but two different leadframe manufacturing technologies: (a) Stamping leadframe and (b) etching leadframe. These details of leadframe design and manufacturing technologies are the keys to the inherent package quality.

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

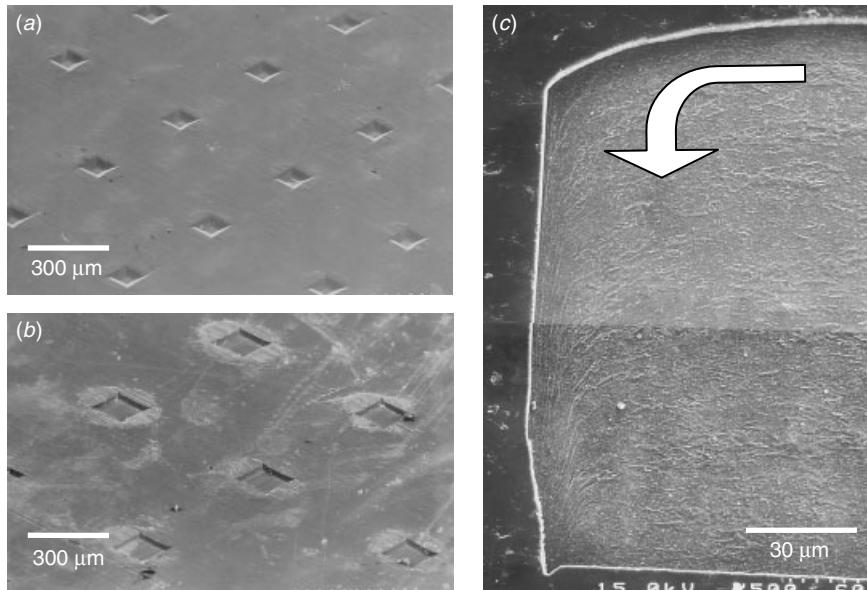


Figure 3.10 Die-pad back-side dimples on a (a) stamping leadframe and (b) etching leadframe, and (c) a cross section of a stamping leadframe. Its internal microstructure is revealed after chemical etching. Microstructure grains are squashed along the edge as indicated by the arrow.

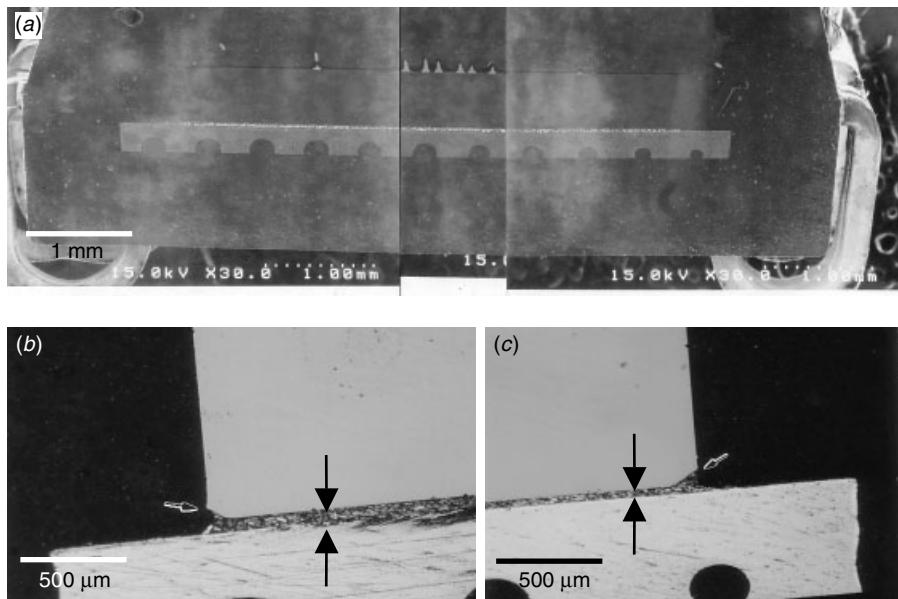


Figure 3.11 Cross section view of package device. (a) Low magnification SEM showing die-bond thickness, and leadframe die-pad thickness and back-side dimples. (b,c) Close-up of the die-bonding epoxy voids and variation thickness on both edges of the die.

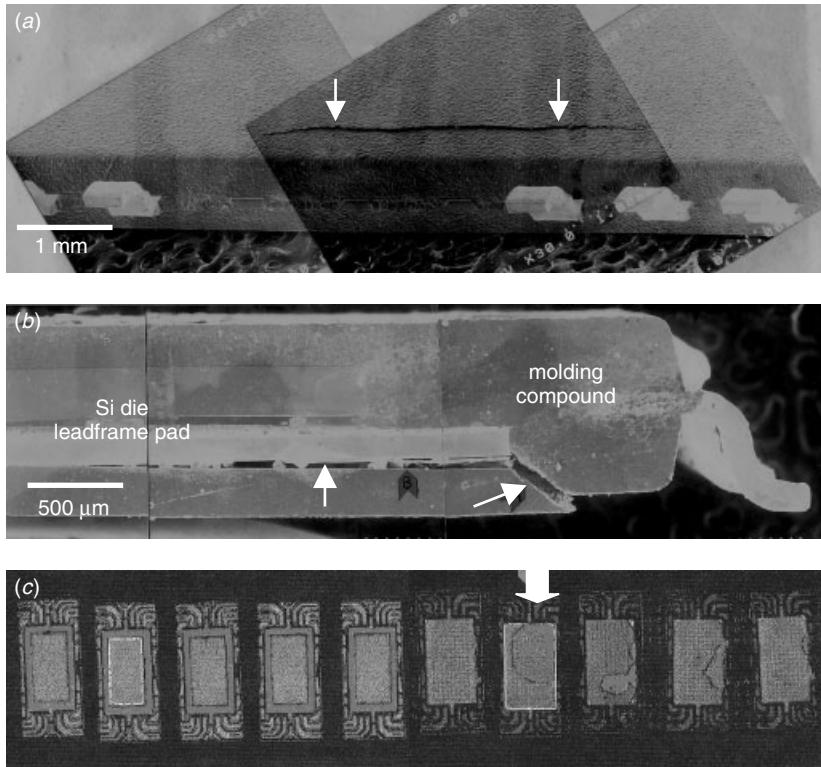


Figure 3.12 Popcorn failure is related by the SAM image (as indicated) of a TSOP package after preconditioning stress. (a) SEM tilted image, (b) cross section across the pop corn crack, and (c) SAM of the device indicating that the cracked die (arrow) has severe delamination at the edge of the package.

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.

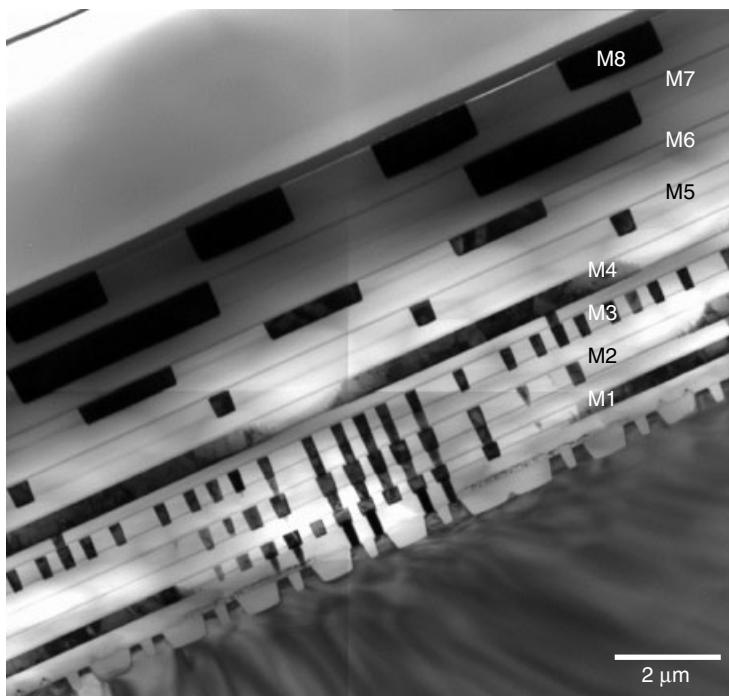


Figure 3.13 TEM cross section of microprocessor A. This is an eight-layer Cu metallization process.

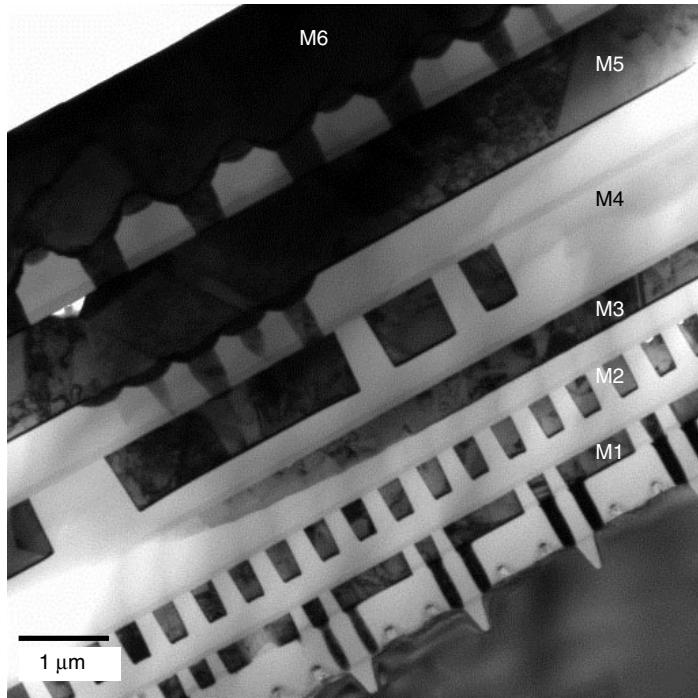


Figure 3.14 TEM cross section of microprocessor P. This is a six-layer Cu metallization process.

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

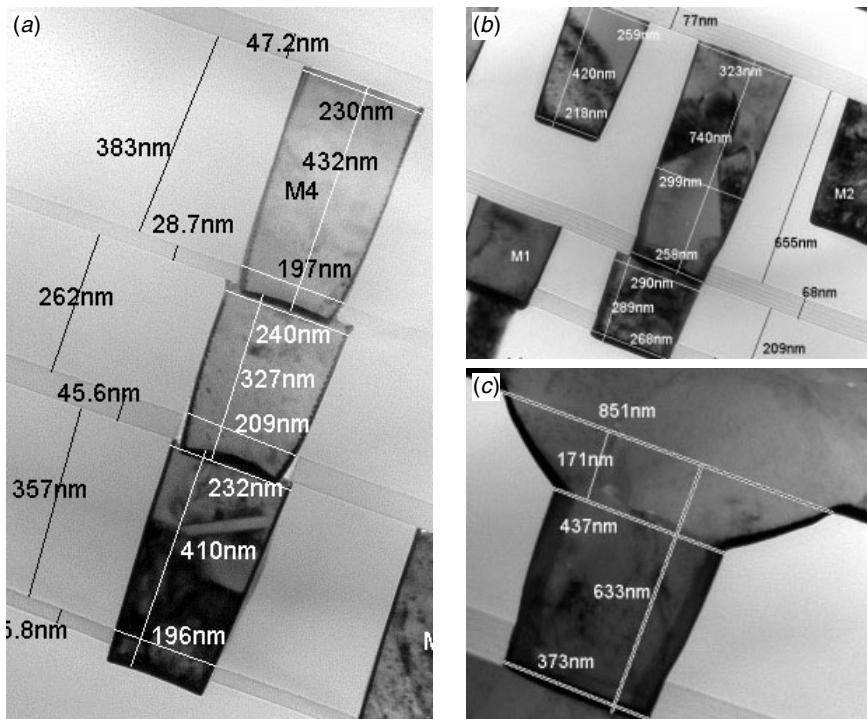


Figure 3.15 Metrology and measurement of each individual layer and element can be done in detail. This includes layer thickness as well as critical dimensions as shown in these examples.

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	12.5 ~ 18 Å				96 nm	75 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)			290 nm (top)	
					268 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)		323 nm (top at M2)	
					299 nm (bottom M2)	
					258 nm (bottom)	
					303 nm (Al M2)	
					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	

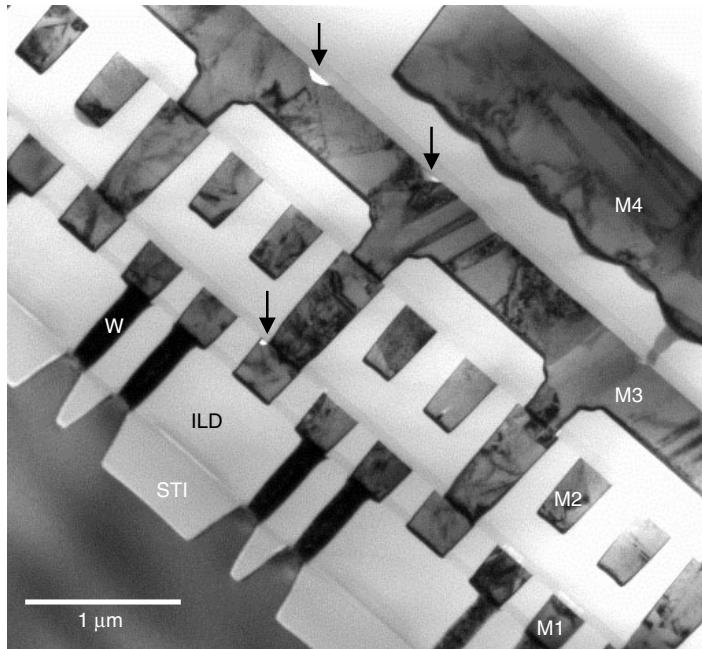


Figure 3.16 TEM cross section of the microprocessor A. Local interconnects using M1–M4 are shown. This is the SRAM Cache memory area. Note that M4 is thicker than M1–M3. Some random metal voids are also observed as indicated by arrows.

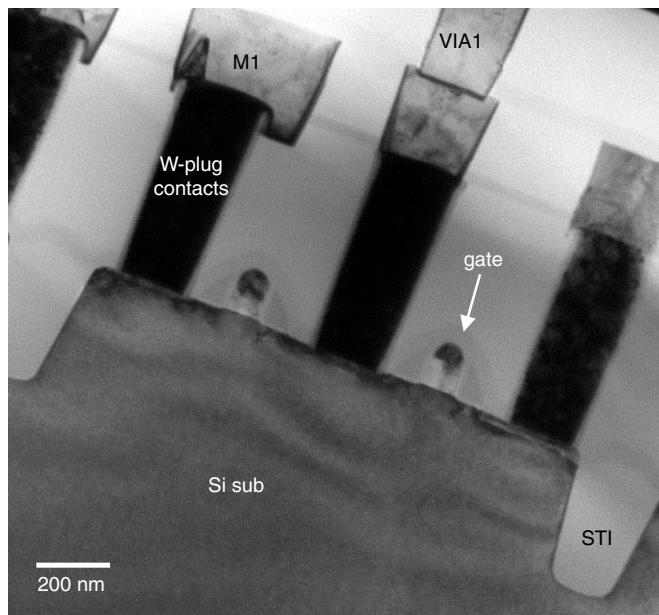


Figure 3.17 TEM cross section of the microprocessor A. Transistors, STIs, contacts, and metals are shown. Note that the gates are squeezed in between contacts in this SRAM area.

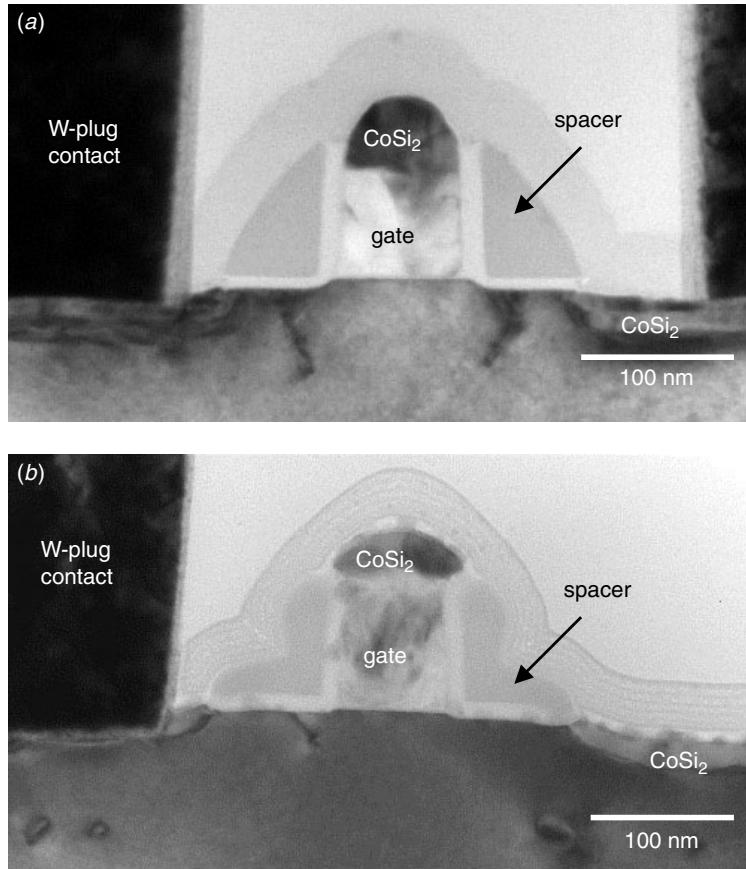


Figure 3.18 TEM cross section of the gate structures of microprocessors (a) A and (b) P. Note that the gate shape, spacer technology, and substrate defects are all different, even though both are of the 75 nm technology node.

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

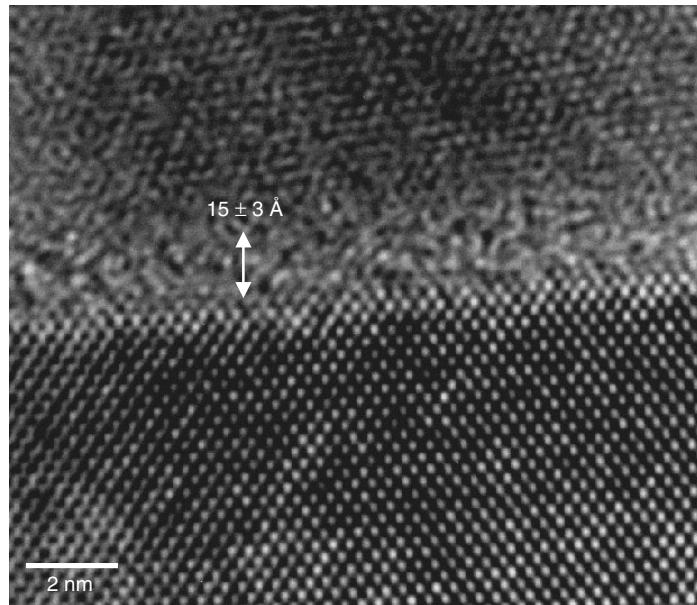


Figure 3.19 HRTEM of the gate oxide from microprocessor P. the Gate oxide's thickness measurement can be done by using Si(111) d spacing as an internal standard. $15 \pm 3 \text{ \AA}$ is the gate oxide's thickness.

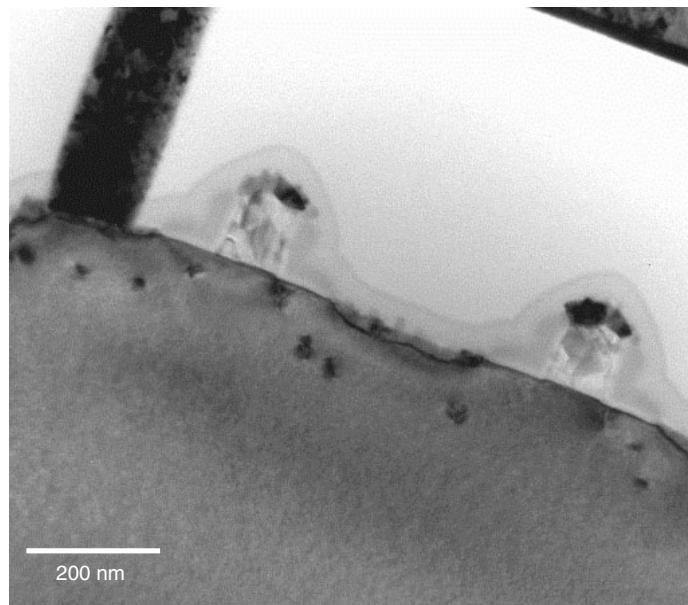


Figure 3.20 TEM cross section of the microprocessor P. Ion implantation induced substrate defects show the contour of ion implantation R_p range. The defects could be detrimental if the subsequent process stress is not controlled properly. In this case the device's function does not appear to be affected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

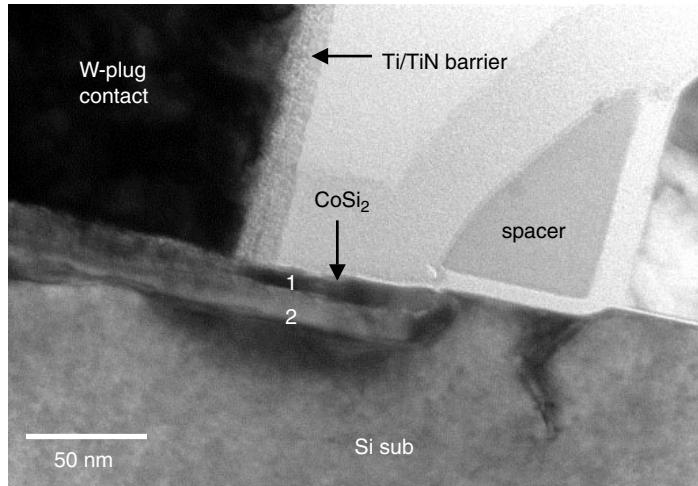


Figure 3.21 TEM cross section of the contact structure of microprocessor A. The CoSi₂ salicide contact has two distinctive crystalline layers.

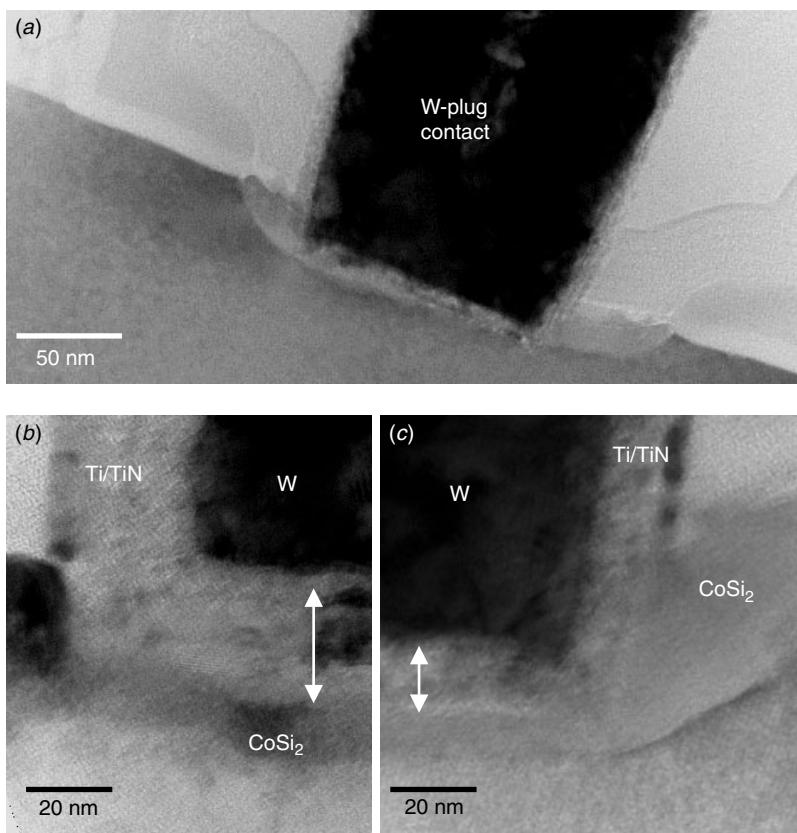


Figure 3.22 (a) TEM cross section of the contact structure of the microprocessors, (b) sample A, and (c) sample P. The Ti/TiN barrier thicknesses are very different in the two devices.

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

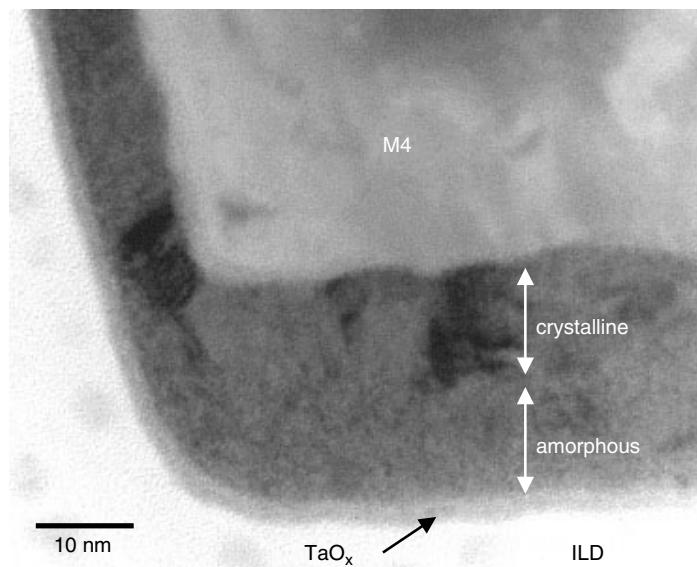


Figure 3.23 Ta barrier used for all the metals. The barrier is composed of two layers: a crystalline layer next to Cu and an amorphous layer next to the dielectric ILD. Each is about 10 nm in thickness. A thin dark layer between the amorphous Ta and dielectric is believed to be a Ta oxide that enhances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

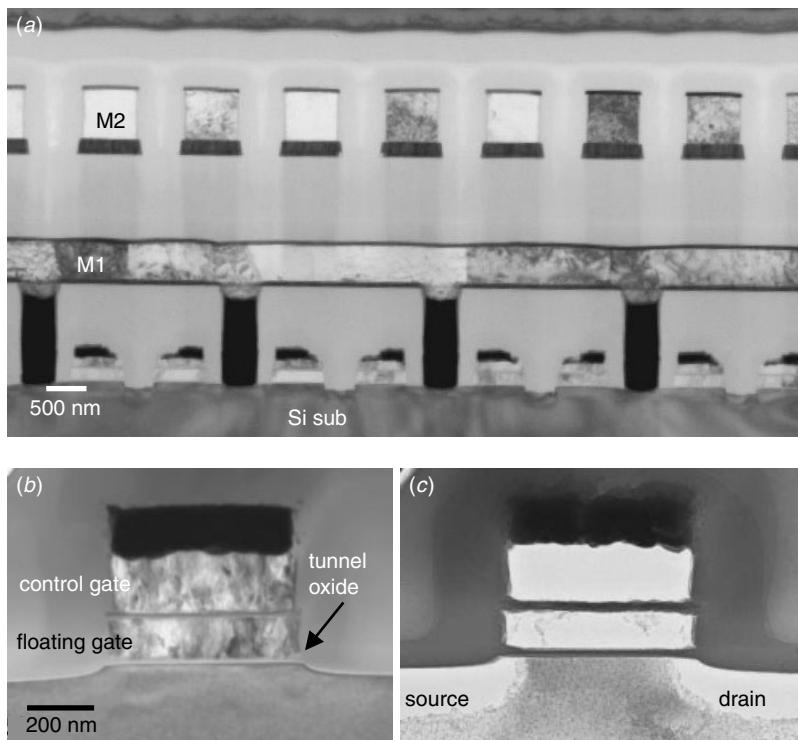


Figure 3.24 Flash memory devices using stack gate structure. (a) Overall device structure, (b) close-up at the stack gate, and (c) source and drain profiles after junction delineation.

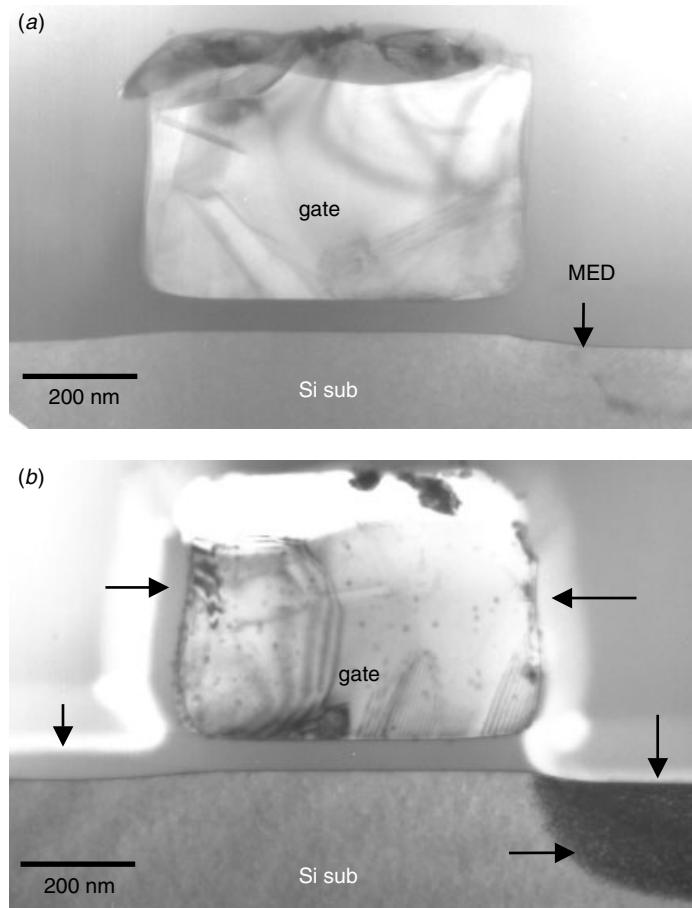


Figure 3.25 Construction analysis of LDMOS shows extra ion implantation on one side of the gate to produce asymmetrical doping. (a) Ion implantation-induced defects are observed in this region. (b) Oxide delineation reveals the asymmetrical gate sidewall oxide lining and its relationship to the substrates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

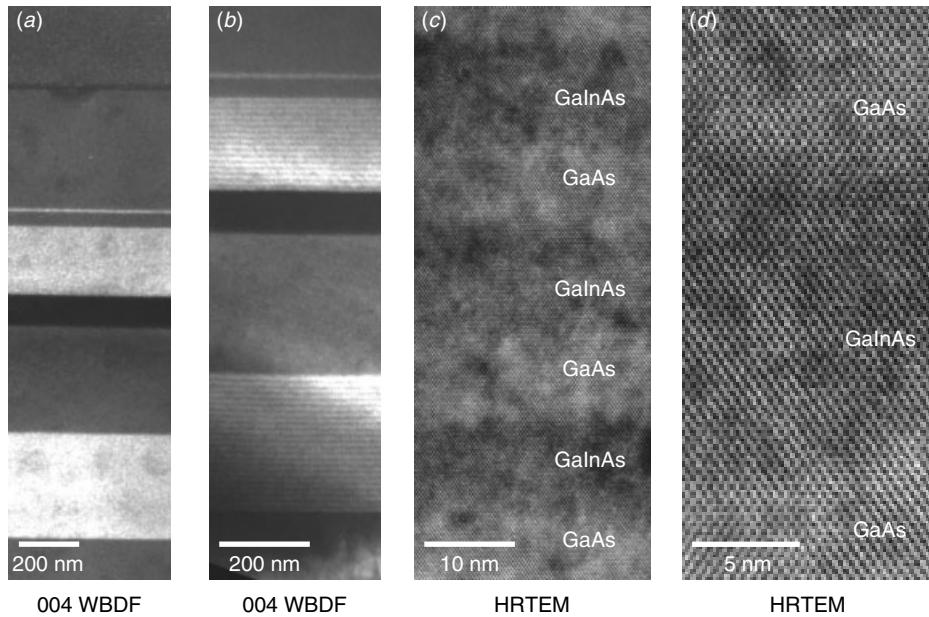


Figure 3.26 Example of non–Si based device construction analysis. GaAs and superlattice structure in a GsAs device. (a, b) Weak beam dark field (WBDF) on the overall layered structure; (c, d) on the HRTEM lattice images of the GaAs–InAs heterostructure.

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

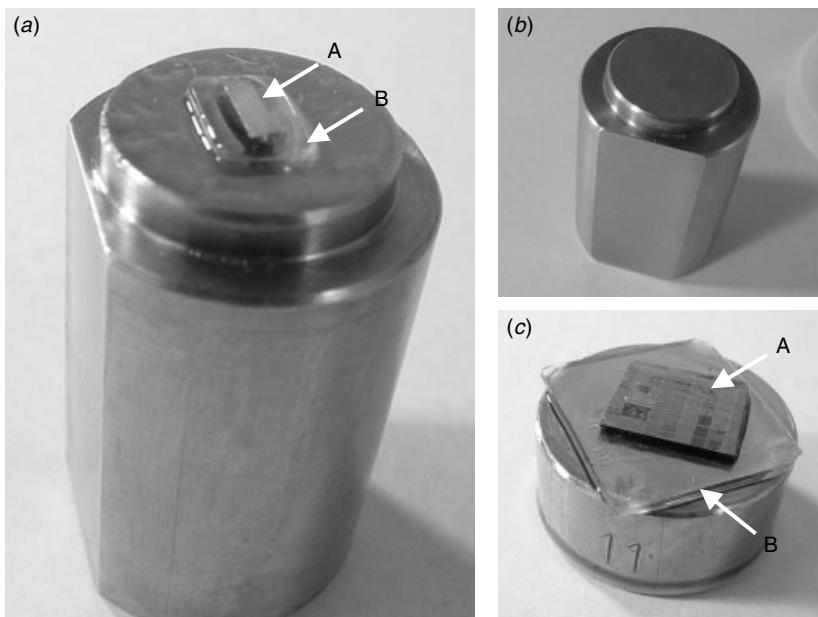


Figure 4.2 Sample preparation stages for nonprecision TEM sample preparation. Arrow A shows the sample and arrow B shows a glass buffer plate in between the sample stage and the sample.

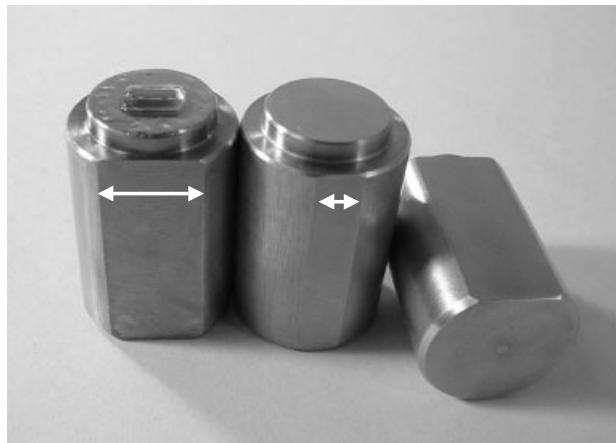


Figure 4.3 Sample preparation stages for nonprecision TEM sample preparation. The two vertical sidewall cuts along the cylinder provide for the holder to be positioned down and sample examined under an optical microscope.

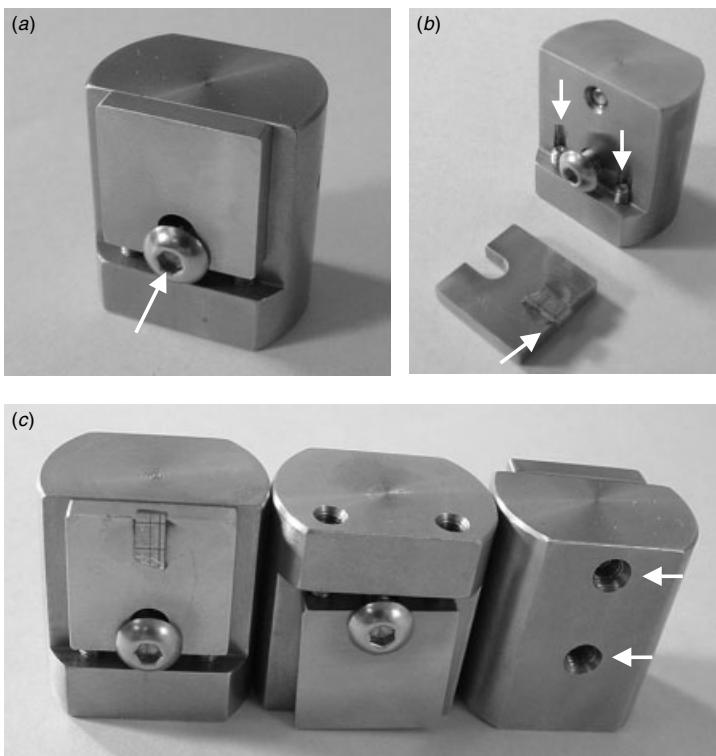


Figure 4.4 Sample preparation stages for precision TEM sample preparation. A sample stub was added on the truncated cylinder. Two screws on the back and another two screws from the bottom were added to provide for the tilting capability of the sample stub. (a) The affixing screw, (b) tilting screw and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

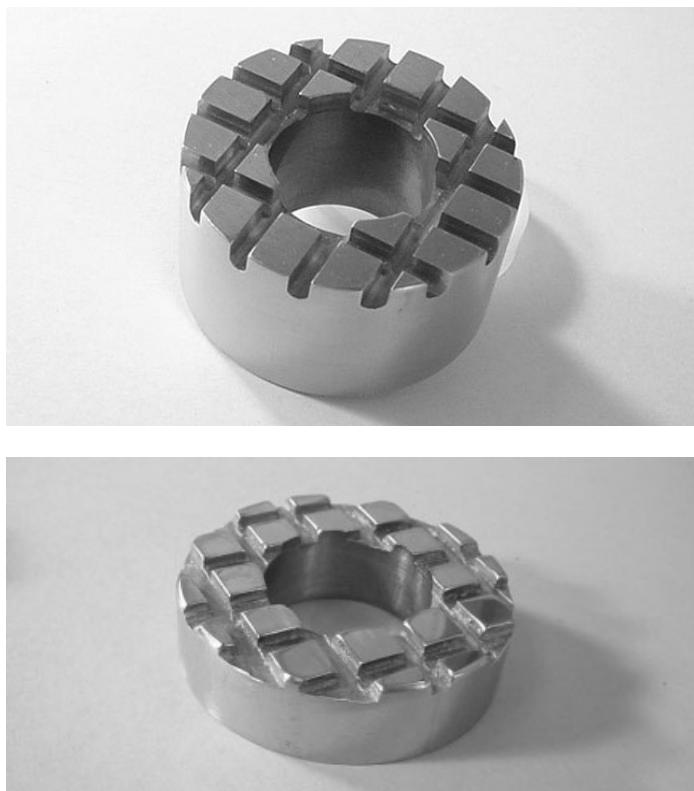


Figure 4.5 Sample preparation jigs. These donut-shaped cylinders allow the samples stages to fit through during mechanical polishing and grinding. The cross-cutting trenches are necessary for ducting the cooling and lubricating water.

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.

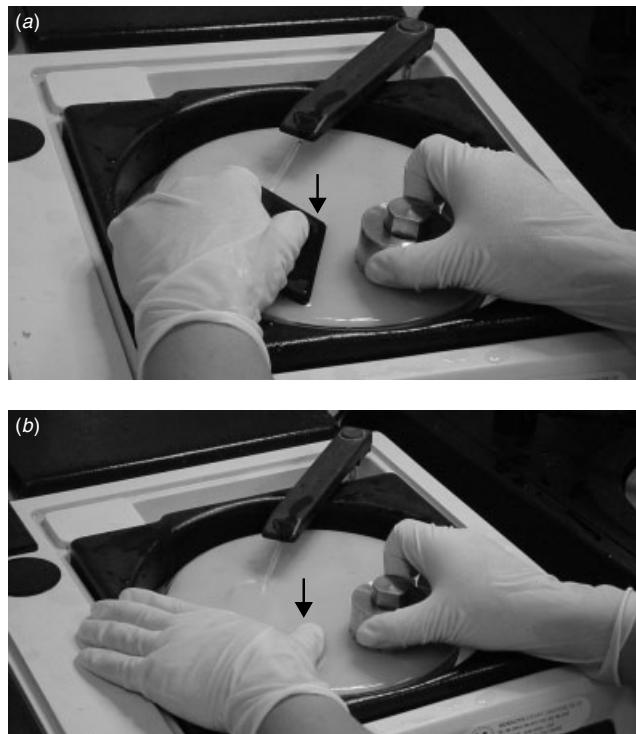


Figure 4.6 Debris can be effectively removed by (a) a rubber brush, or (b) hand with gloves to prevent scratches on the sample during the final polishing.

- 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.

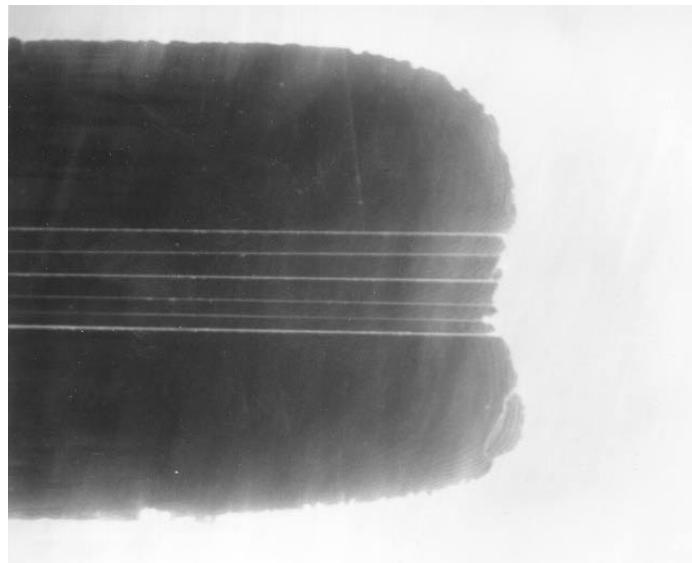


Figure 4.7 After final polishing the Si substrate should show orange to yellowish contrast under the stereo optical (binocular) microscope. If the sample is still totally dark and translucent, the sample is too thick and need further polishing.

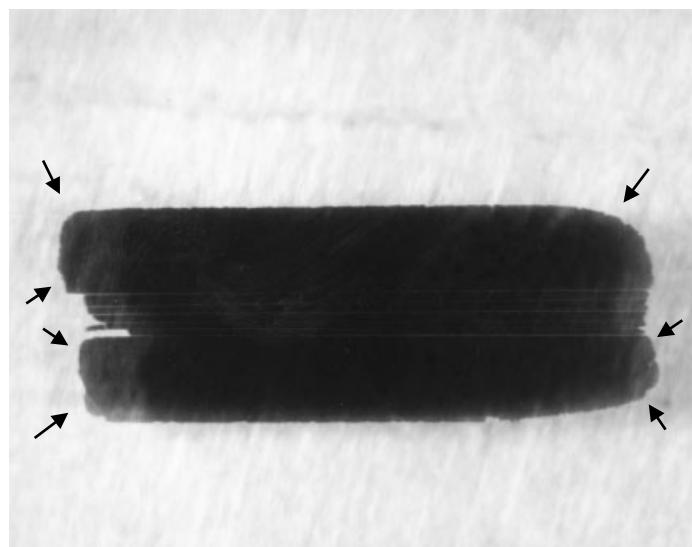


Figure 4.8 Thinned down sample that is reasonably uniformly until the rounding at all corners, which is a good indication that the sample is thin enough and ready for the next step: Cu grid reenforcement.

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

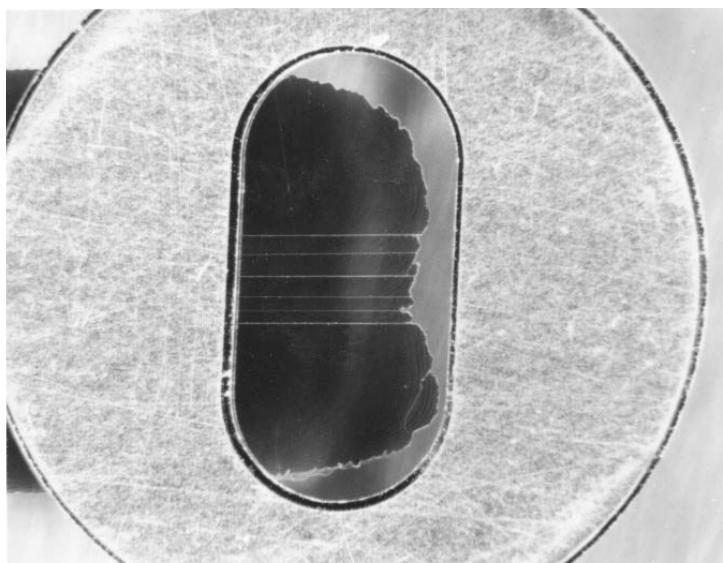


Figure 4.9 Large optically transparent areas indicate a thin sample. Often a few (3–5) minutes of ion milling (mainly for cleaning) are recommended before and the sample is ready for TEM examination. Almost the whole area within the 2×1 mm is usable for TEM analysis.

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

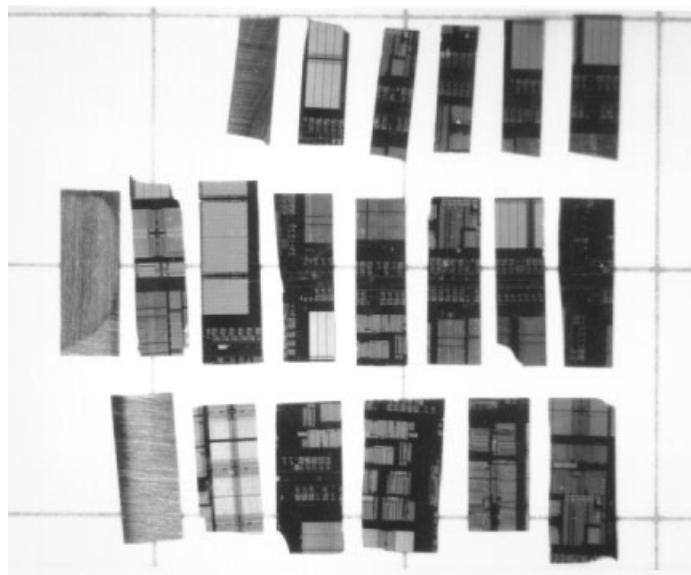


Figure 4.10 Wafer, or dice, sectioned into roughly 1×3 mm rectangles after back-side thinning, are ready for stacking. The exact shape and size of each dice is not critical. The background squares are 5×5 mm.

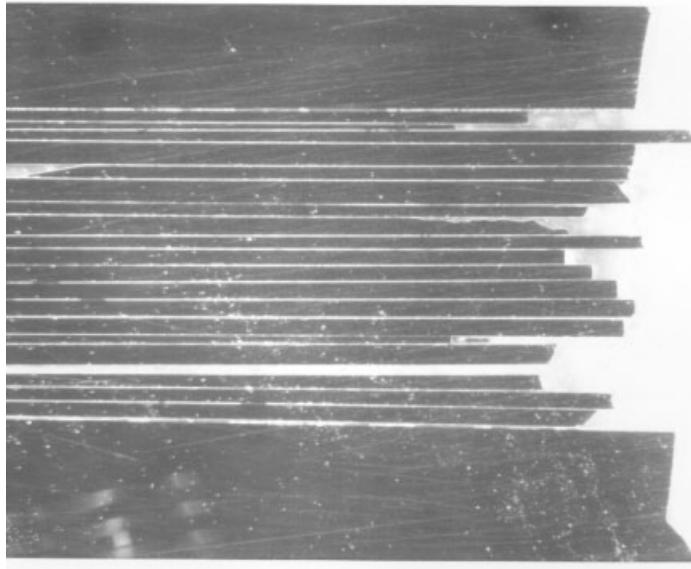


Figure 4.11 As many as 20 pieces can be stacked together, including the un-pre-thinned samples on both sides. This provides enormous advantages in efficiency. Notice that the stacking does not have to be perfect.

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:

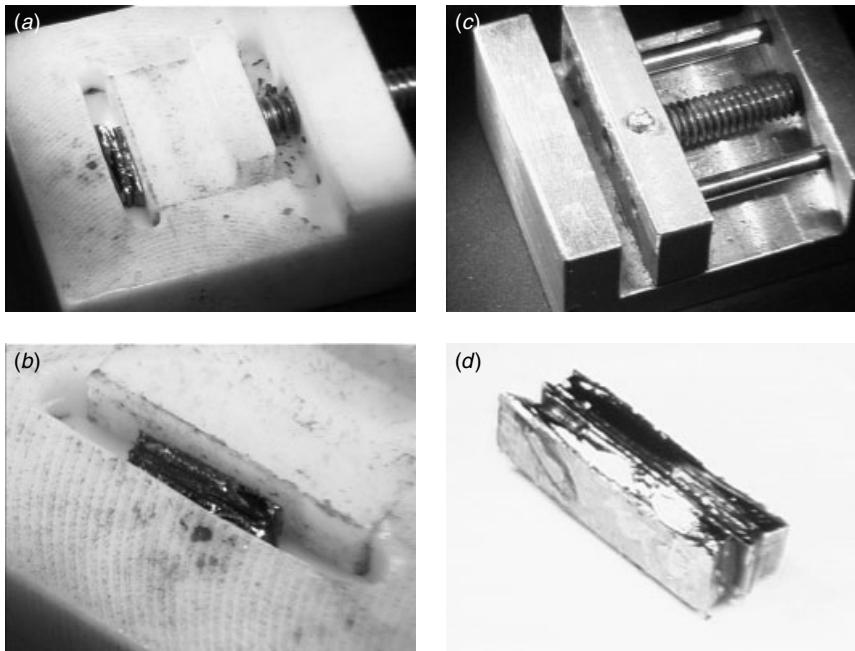


Figure 4.12 Mechanical clips for the stacked sample used during high temperatures. Teflon (a, b) and steel (c) are the usual materials. The cured sample is shown in (d).

- 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.

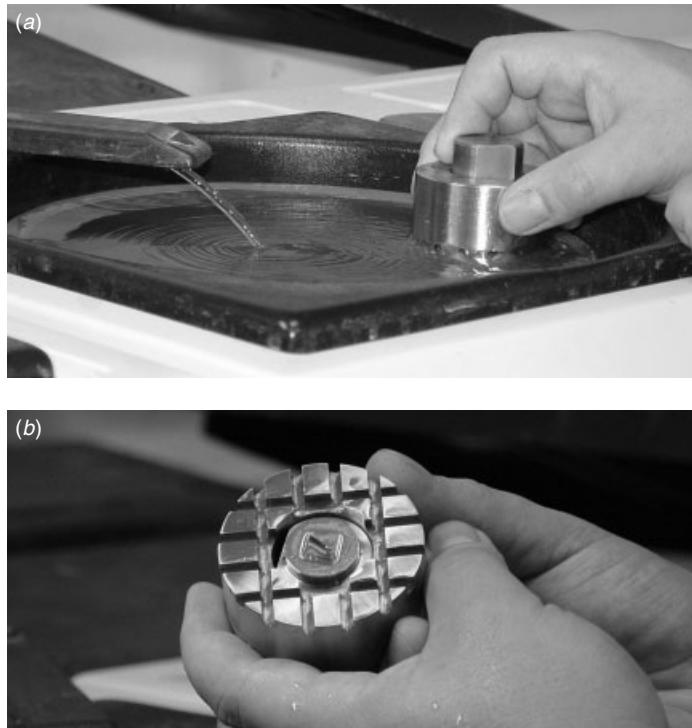


Figure 4.13 Minimum pressure is exerted on the polishing sample (*a*), and a constant check (*b*) is necessary for satisfactory results.

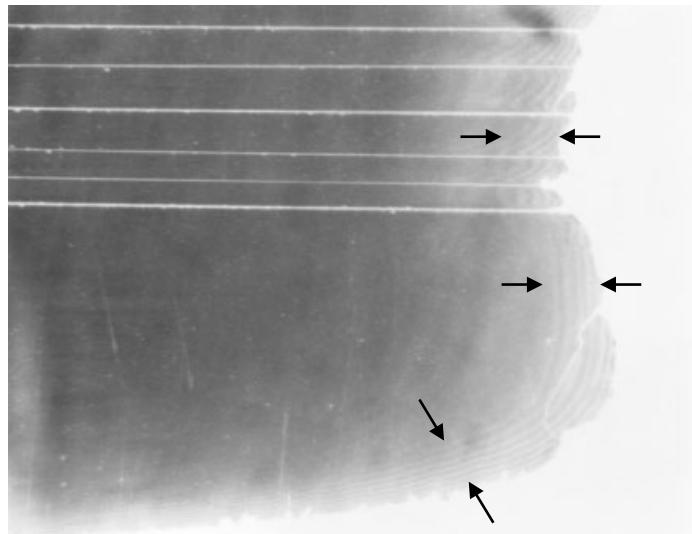


Figure 4.14 Sample with transparent yellowish fringes and contours indicates that the thickness is thin enough and polishing is nearing its end.

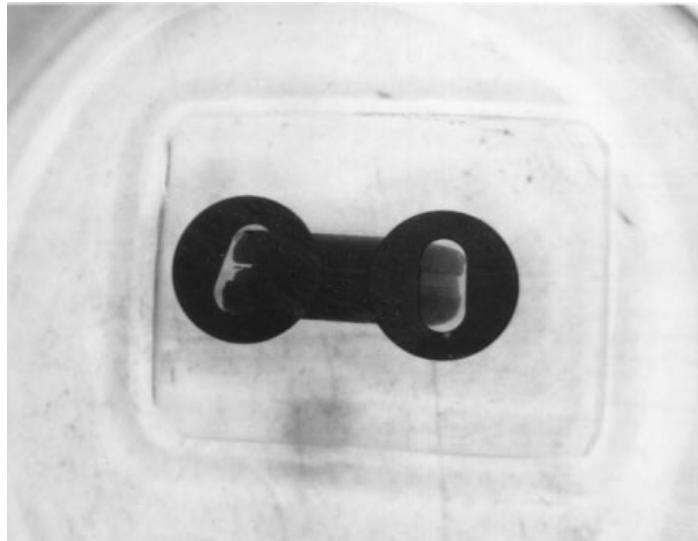


Figure 4.15 The glass piece along with the sample are lifted from the sample stage and rinsed into acetone to remove the thermal wax and clean the sample.

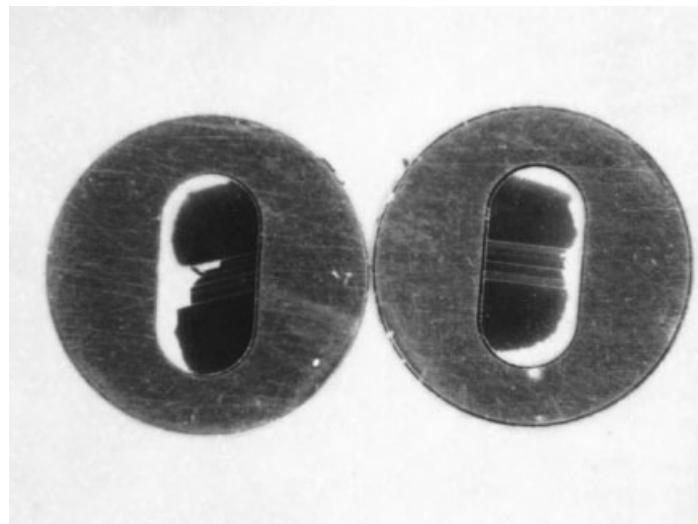


Figure 4.16 After trimming, two different samples are obtained. The thinner one is better than the other.

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

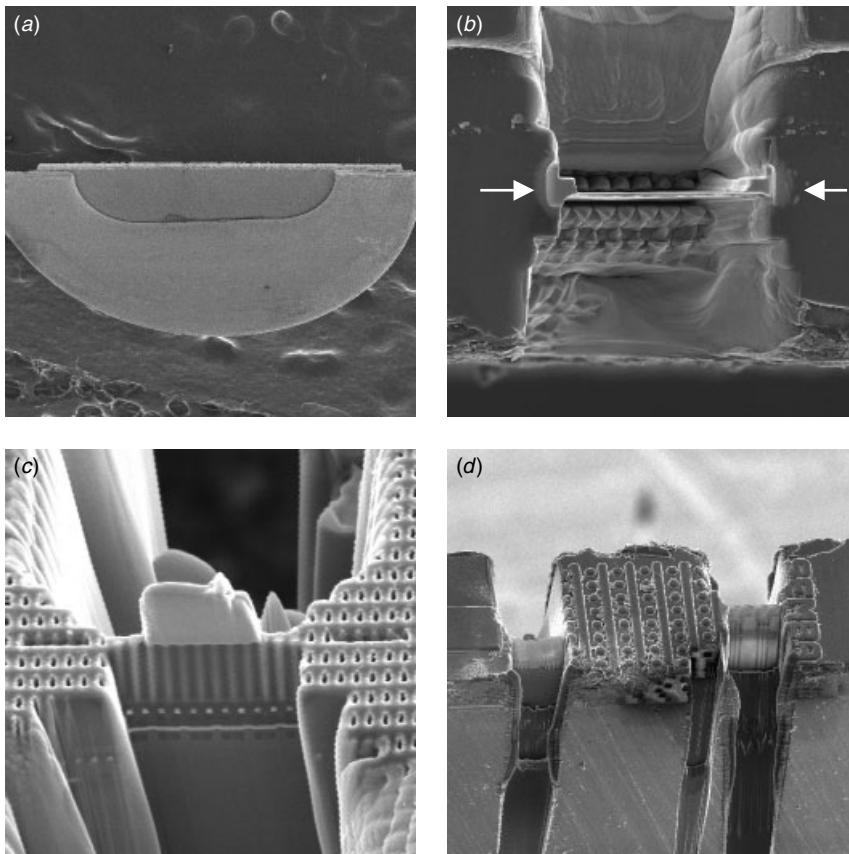


Figure 4.17 (a) Sample after mechanical pre-thinning down to around 20 μm and supported by half a Cu grid. (b) SEM top view of the sample after FIB thinning showing a thin membrane of the target area, as indicated. (c) A tilted view of the final TEM sample using SEM imaging capability in FIB. (d) Multiple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

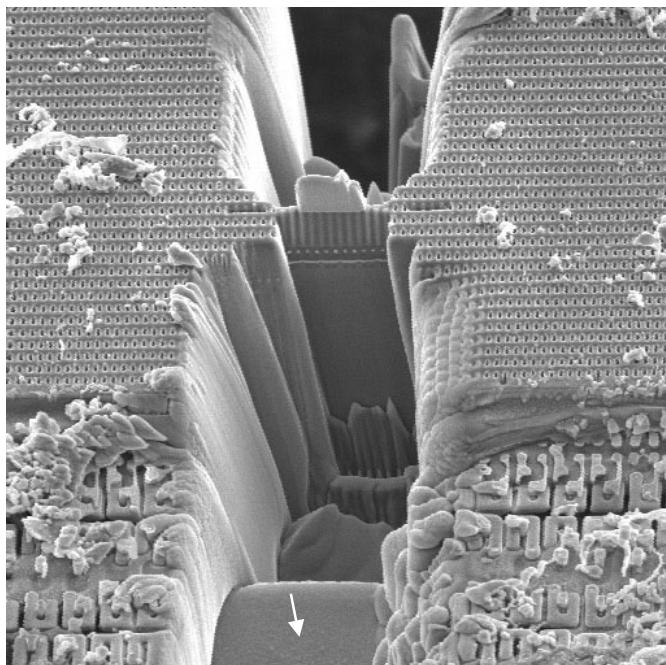


Figure 4.18 Progressively sharper ion beam milling at the target area creates progressively deeper cuttings. The re-deposition at the outer area, as indicated, may eventually block observation by TEM.

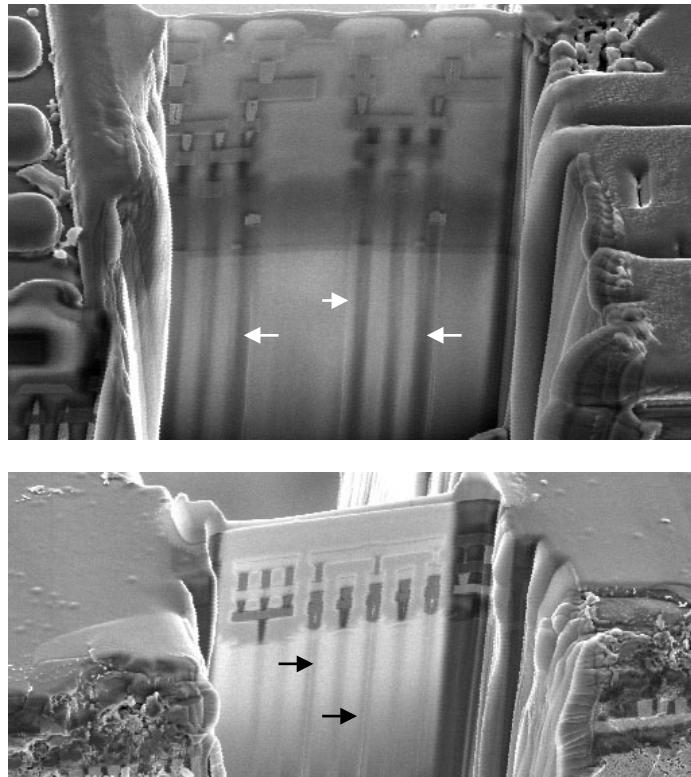


Figure 4.19 Shadowing effects in SEM images induced by the vertical shadows of W-plugs, a material that has much lower sputtering rate. A similar, but reverse, effect can result from void structures.

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

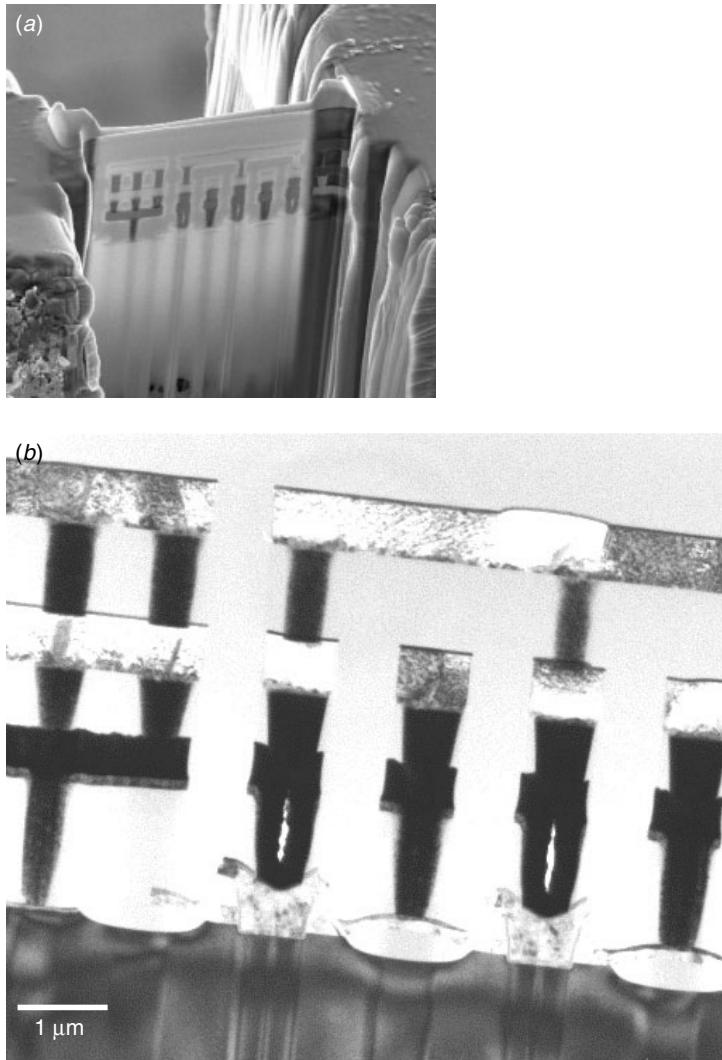


Figure 4.20 FIB cut TEM sample with an SEM image (a) and a TEM image (b) of the same area.

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

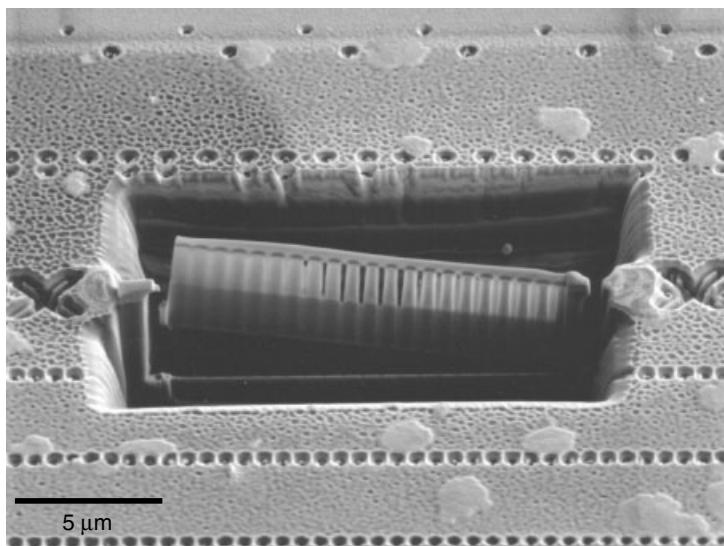


Figure 4.21 FIB cut target membrane for extraction. Notice that the sample membrane is nearly cut from the substrate and is tilting toward one side. Also note the two cross marks on both sides of the sample.

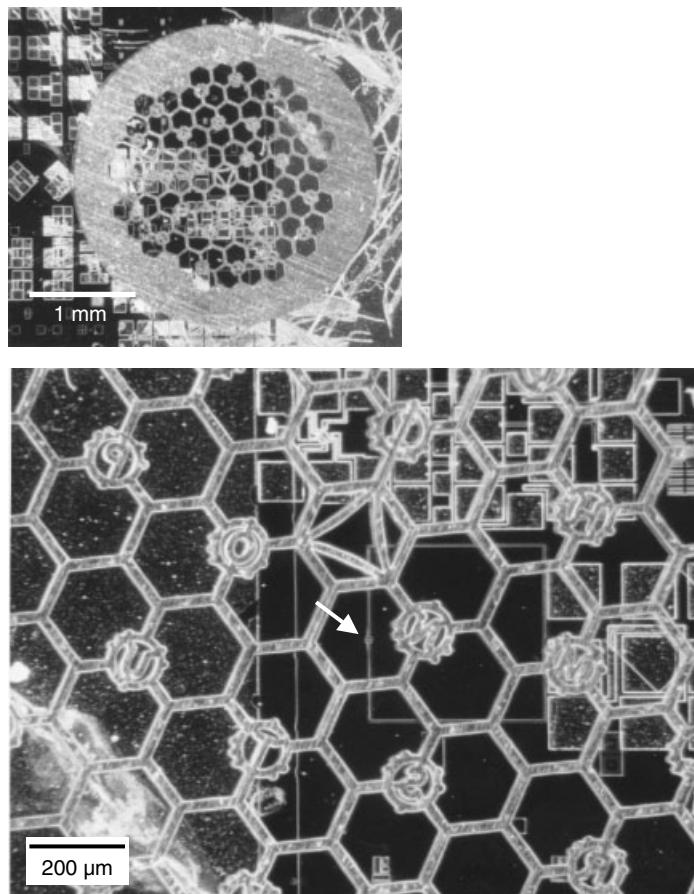


Figure 4.22 A finder's grid placed on top of the FIB cut area. In this case the sample is in the center of the N3 hexagon. The letters and other marks are used in identifying the sample's location and different samples within one grid. The sample's location is indicated by an arrow.

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

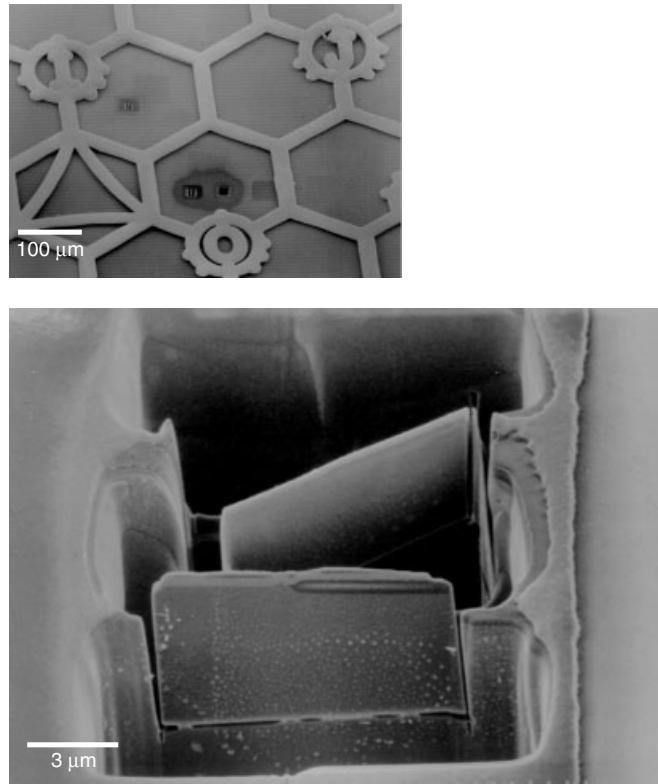


Figure 4.23 SEM image that includes nearby TEM sample membranes within a few μm apart can be prepared with this method. (Sheng et al. Reprint from IPFA, 92–96, 1997 with permission from IEEE)

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.

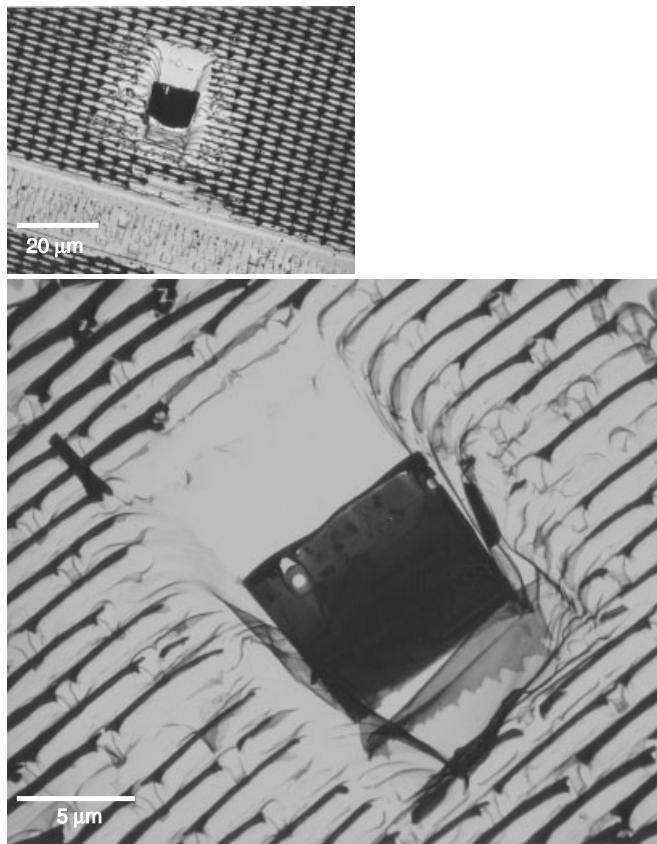


Figure 4.24 TEM membrane extracted by carbon replica method. The surface line feature of the device was replicated as is clearly visible. (Sheng et al. Reprint from IPFA, 92–96, 1997 with permission from IEEE)

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.

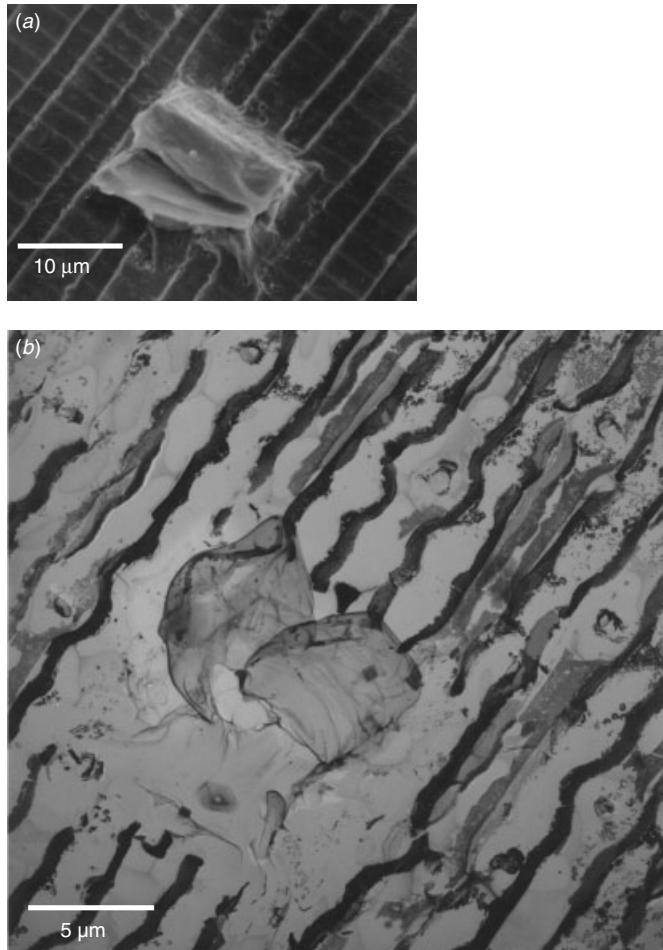


Figure 4.25 (a) SEM image of a sample in a failed extraction. (b) The failure carbon film packets in the sample does not contain the sample's membrane.

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.

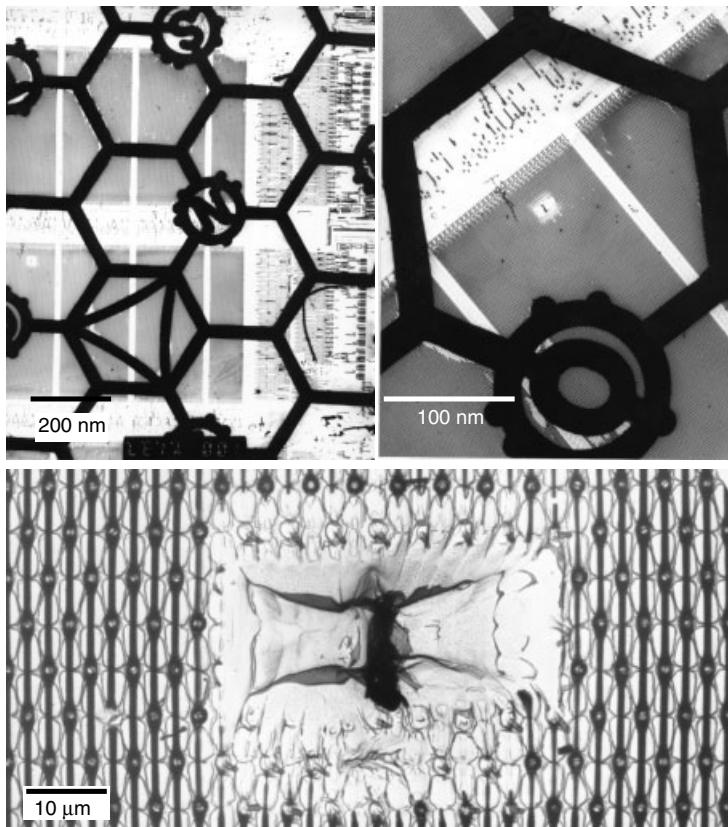


Figure 4.26 An extracted TEM sample membrane that includes the tungsten polycide bit lines. Carbon extraction not only can extract the FIB cut TEM membrane but also replicate any surface features. (Sheng et al. Reprint from IPFA, 92–96, 1997 with permission from IEEE)

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

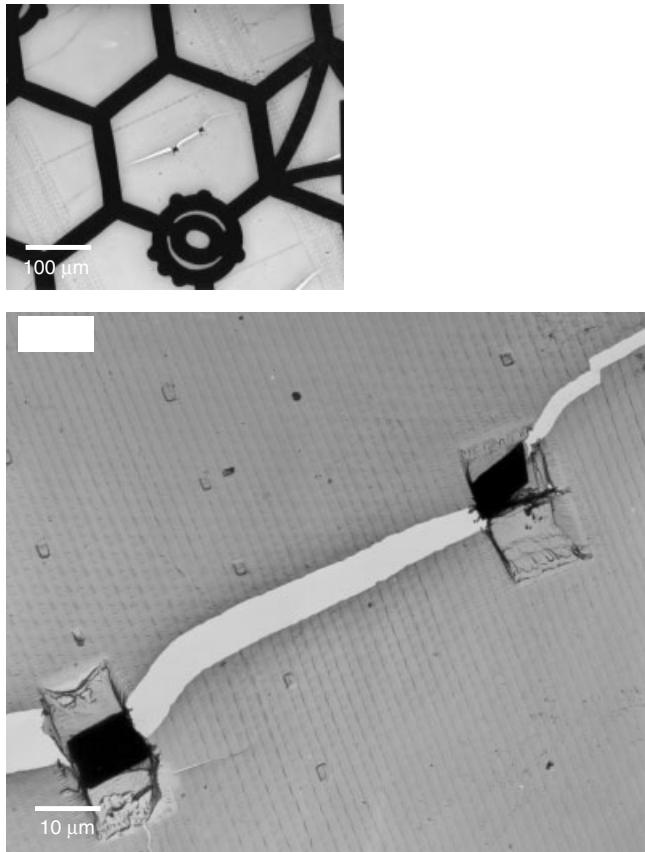


Figure 4.27 Extraction of two samples within a small area. Such an application is often handy in VLSI device defect analysis where multiple defects may exist within a confined area. (Sheng et al. Reprint from IPFA, 92–96, 1997 with permission from IEEE)

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

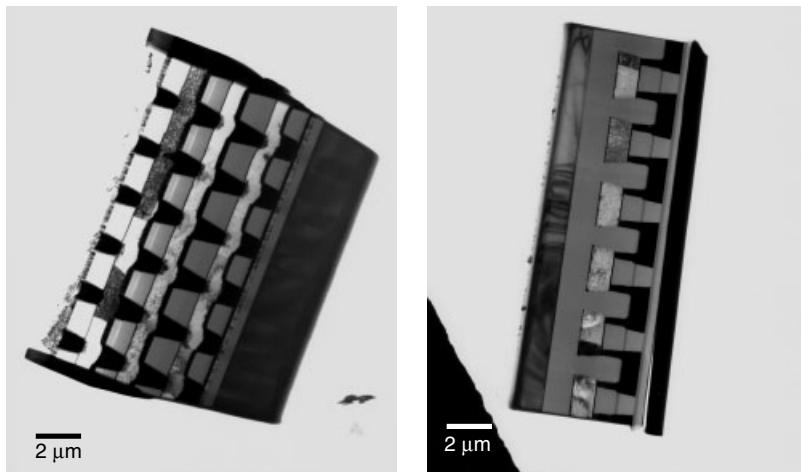


Figure 4.28 TEM sample membranes on carbon film extracted by the glass tip method.

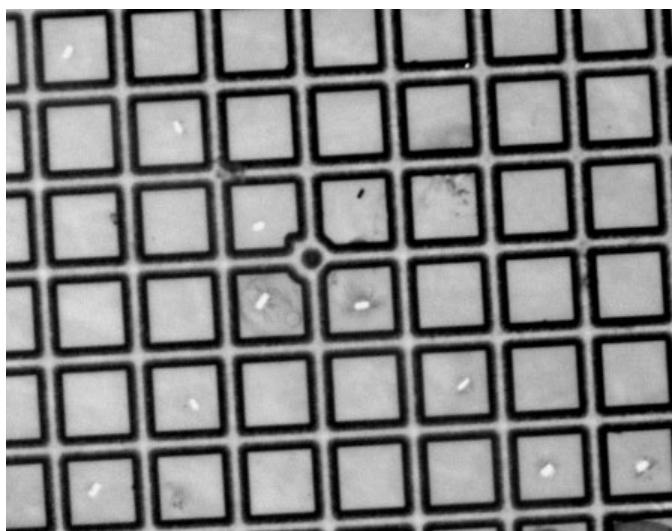


Figure 4.29 Multiple samples within a carbon-supporting Cu mesh.

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.

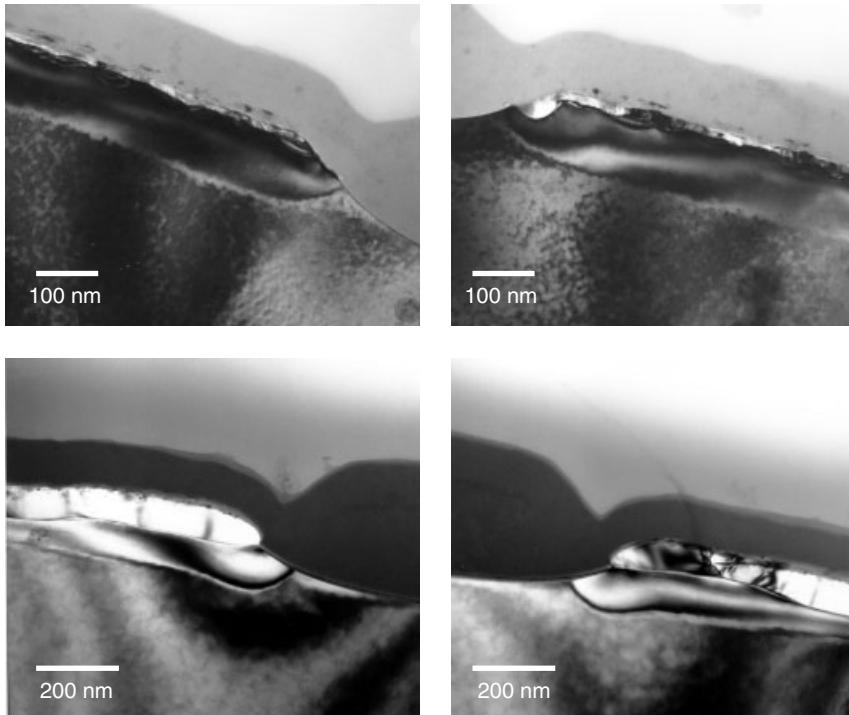


Figure 4.30 Delineation of the shallow junction of a salicide device's active area showing the TiSi_2 layer on top.

- 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

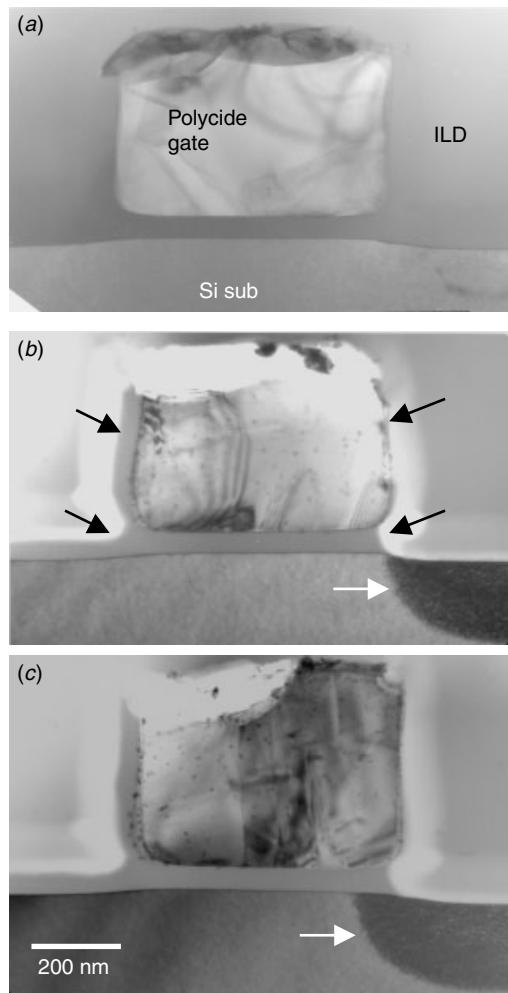


Figure 4.31 Power LDMOS device with asymmetrical device structure. (a) No difference can be seen in the as-milled sample, (b) BOE etching reveals the oxide spacer to be thicker on one side of the polycide gate. (c) BOE etching also reveals a heavily doped junction on one side but not on the other side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

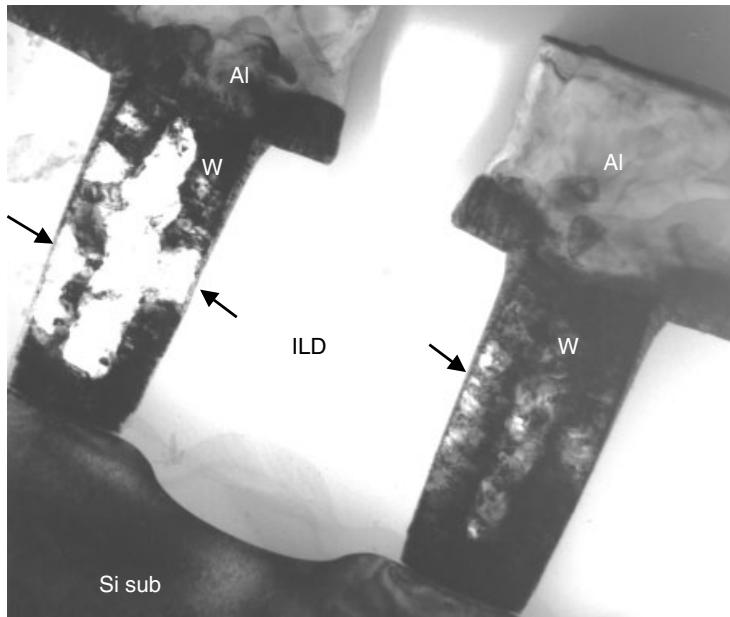


Figure 4.32 W–plug etched to reveal its detailed microstructure. Most important, the Ti/TiN barrier layer can be observed clearly after the W–plug is etched and thinned using chemical stain.

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.

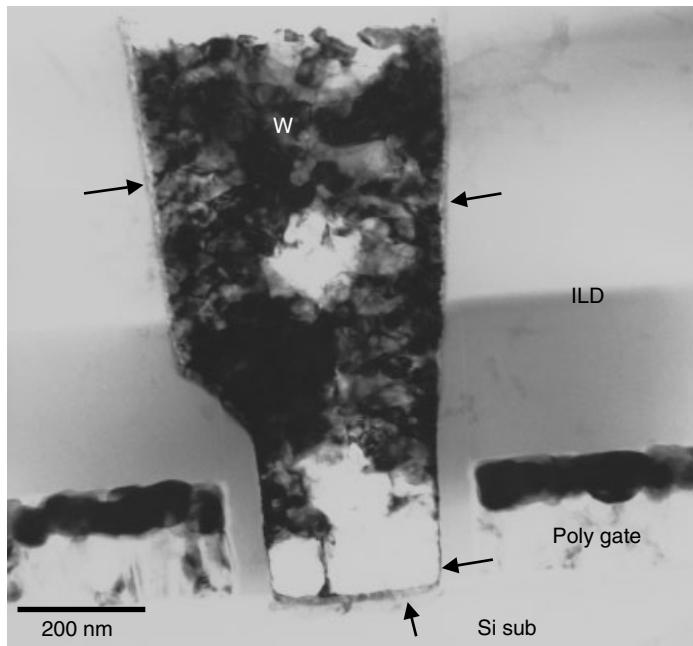


Figure 4.33 W–plug partially etched away. The Ti/TiN around the W–plug’s contact can be seen, as indicated, clearly even under such low magnification.

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

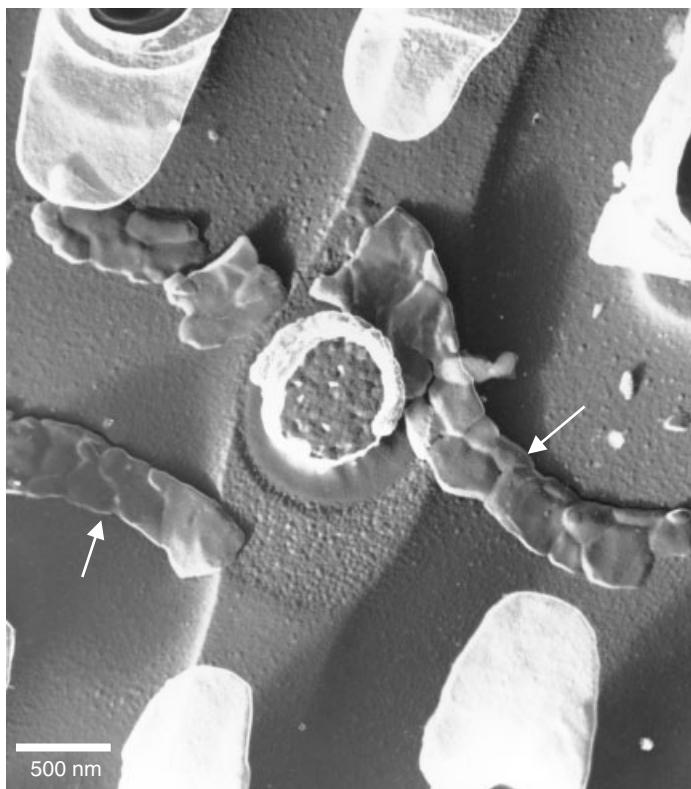


Figure 4.34 Carbon replica of a device's surface structure. Note that the polygate has also been extracted onto the carbon replication film, as indicated.

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

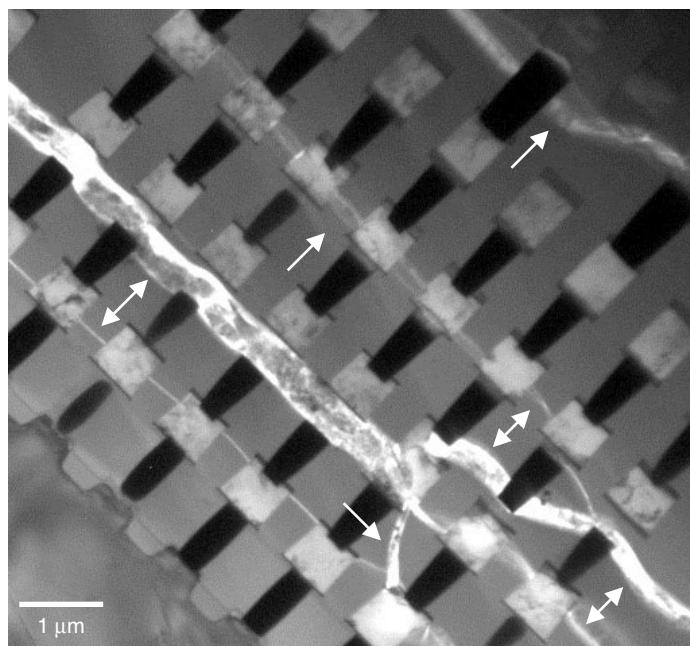


Figure 4.35 Mechanically polished TEM sample (with no in milling) showing interlayer and intralayer cracks at and near the thinnest areas.

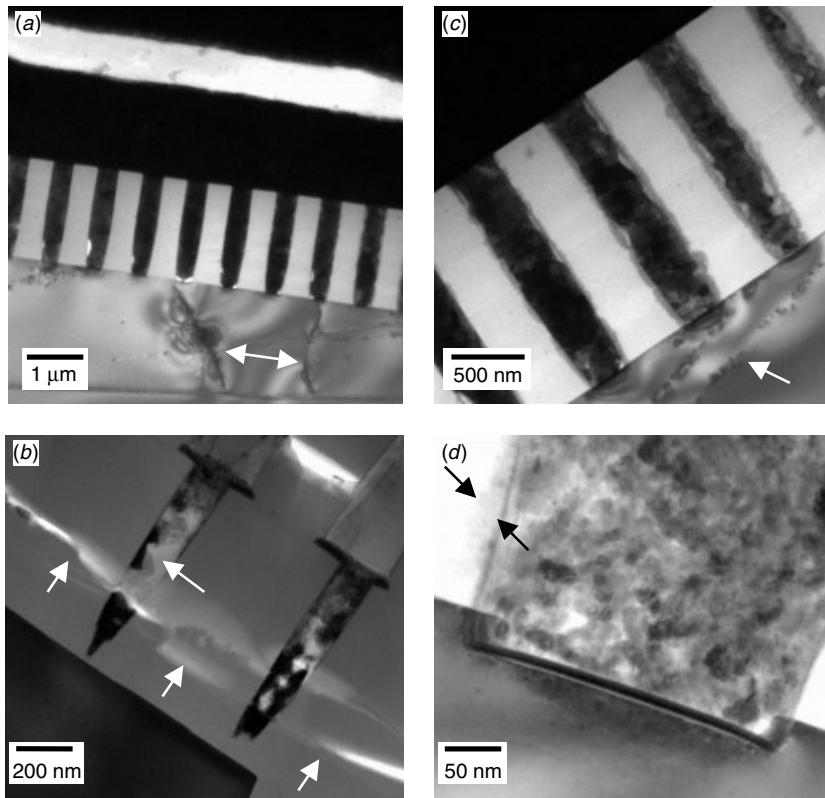


Figure 4.36 Some typical artifacts created by mechanical polishing. (a) Si-substrate cracks, (b) ILD cracks and W-plug scratches, (c) Si-substrate scratches, (d) smeared W grains.

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

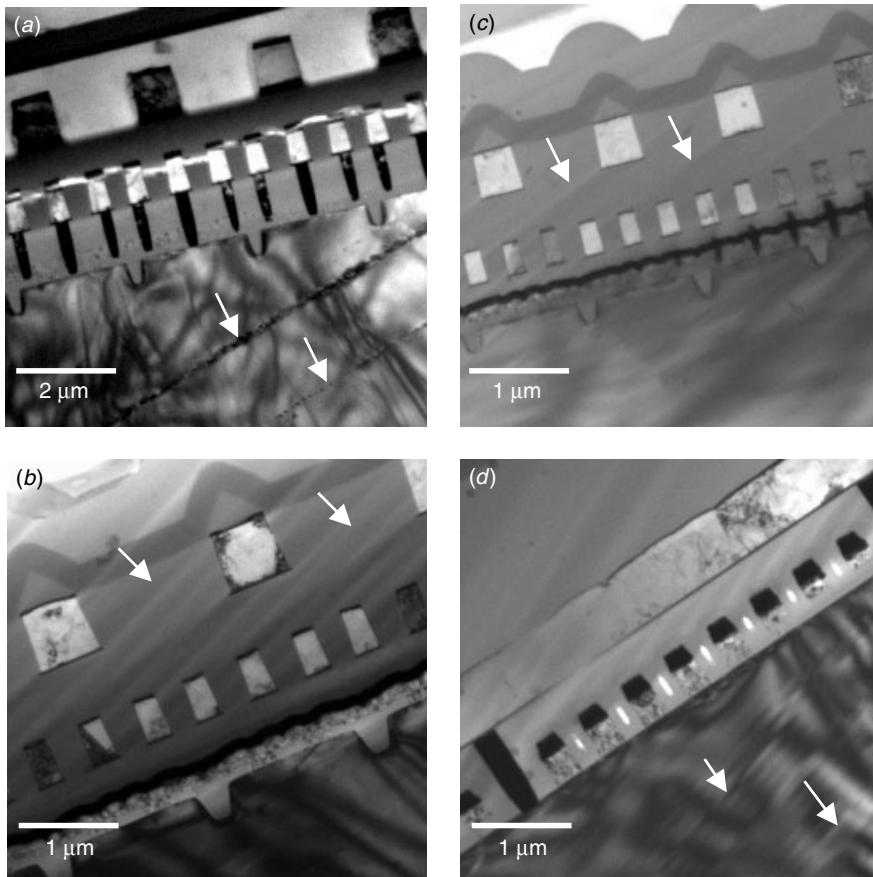


Figure 4.37 Artifacts created by mechanical polishing. (a) Si-substrate scratches, (b) fine polishing scratches due to diamond paste or Al slurry, (c) and (d) fine polishing scratches due to syton.

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

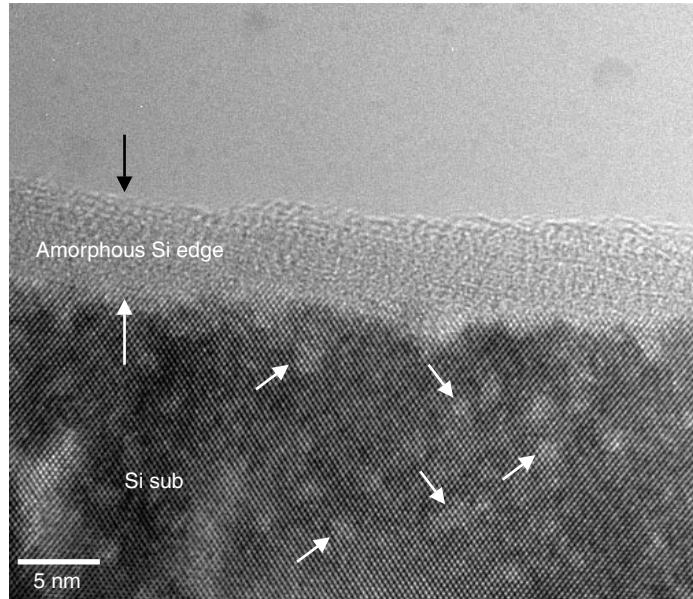


Figure 4.38 Ion milling induced amorphization layer on both surfaces of the sample. The amorphous layers extend to the edge of the sample and create an amorphous edge. Also observed are random spots with brighter contrast, as indicated. These are the ion milling sputtering craters.

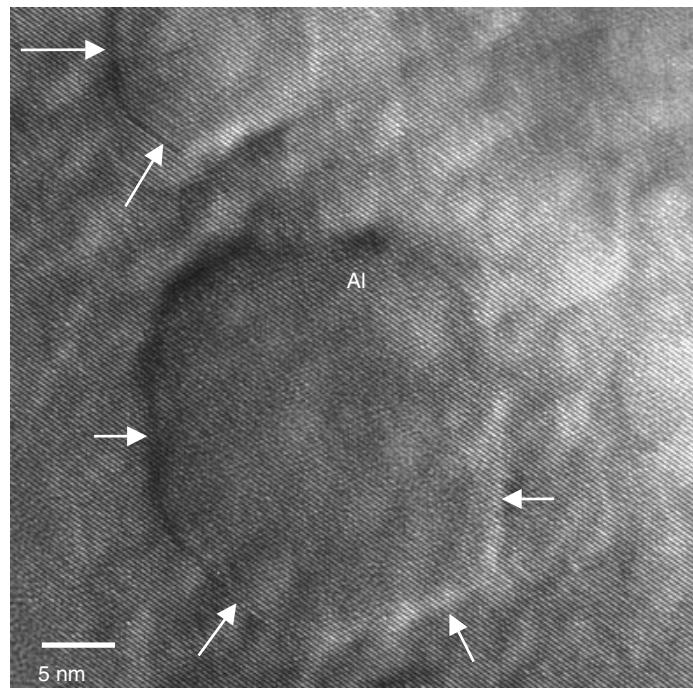


Figure 4.39 Al film with random dimples created by Ar ion milling. Notice that the craters have hexagonal shapes with edges aligned along the Al(111) lattice fringes, as indicated.

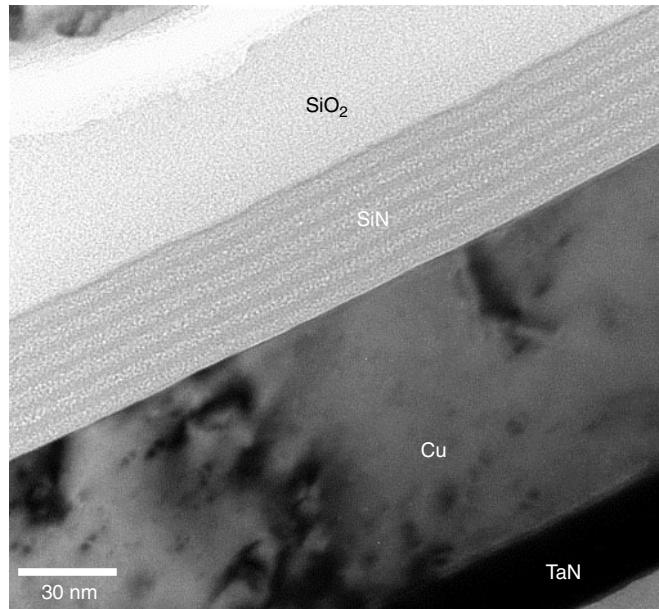


Figure 4.40 SiN laminated structure due to deposition process can be preserved after a short and well-controlled ion milling.

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

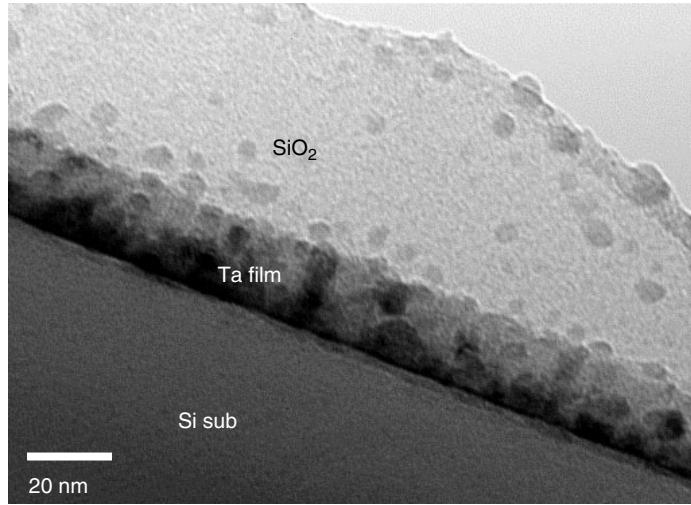


Figure 4.41 Local re-deposition due to ion milling can create spots on the sample. The re-deposited particles are mostly concentrated in insulating materials, such as SiO_2 , though some may be found within conductor layers and Si substrate.

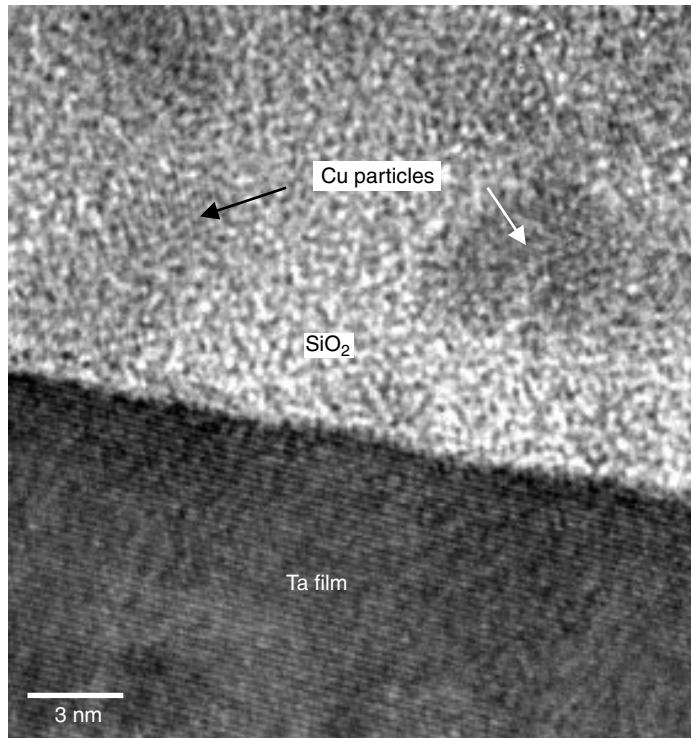


Figure 4.42 High-resolution images showed lattice fringes within the re-deposited particles, which indicates that they are crystalline. EDS shows these are mostly Cu particles.

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.

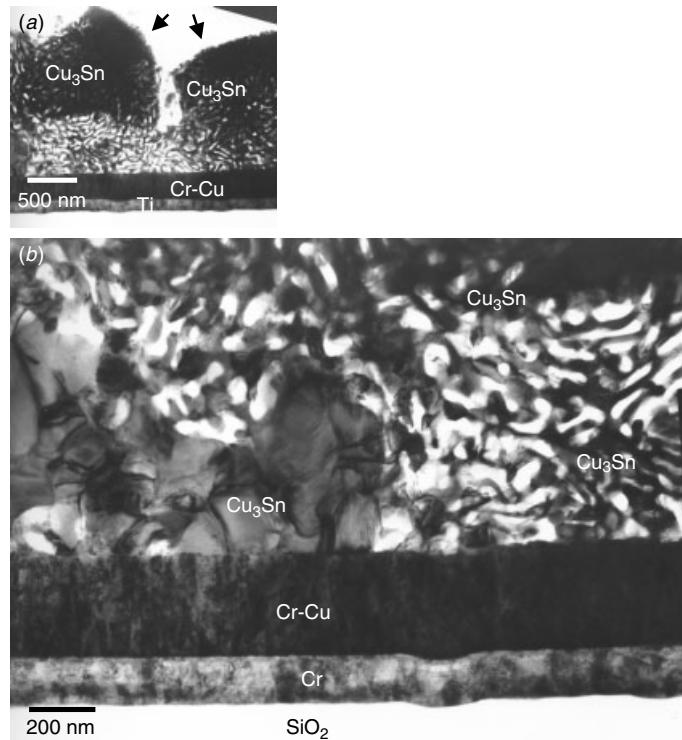


Figure 4.43 Microstructure of the Cu–Sn interaction phases can be altered with extensive ion milling. Cu_3Sn evolves into a porous structure resembling a coral reef.

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 $\text{Si}(111)$ 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

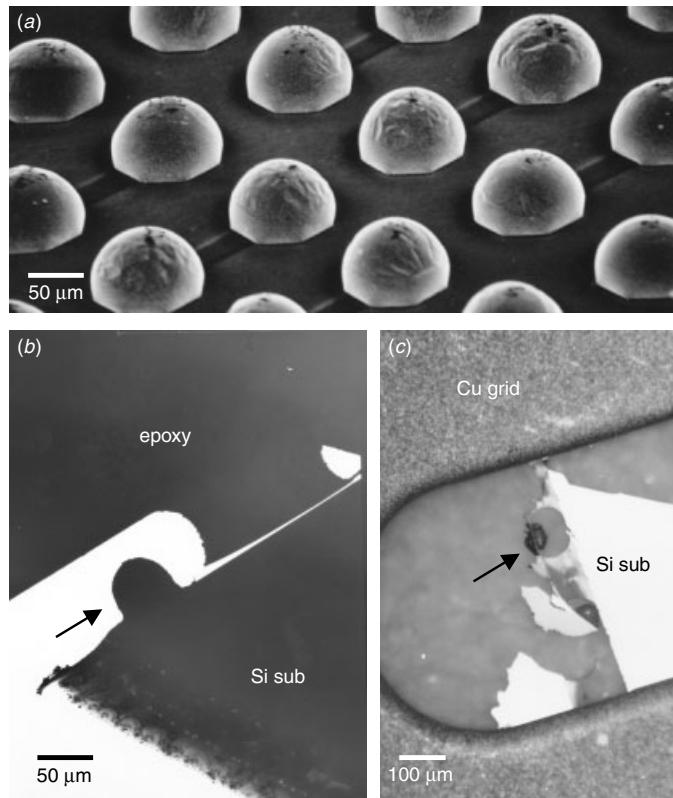


Figure 4.44 (a) Self-aligned solder balls arrayed on a Si substrate. (b) Catastrophic solder re-melting can result from electron beam irradiation. (c) Re-melting can also occur during ion milling.

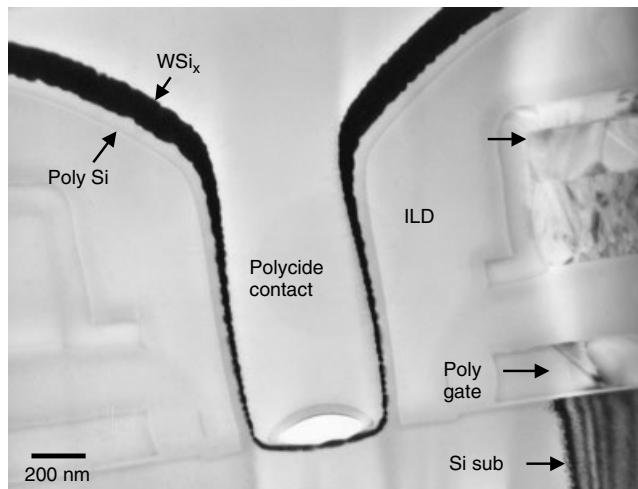


Figure 4.45 Total amorphization of all structural layers, including the Si substrate, polygates, and the W-polycide layer. The only remaining crystalline areas, as indicated, are away from the central targeted contact.

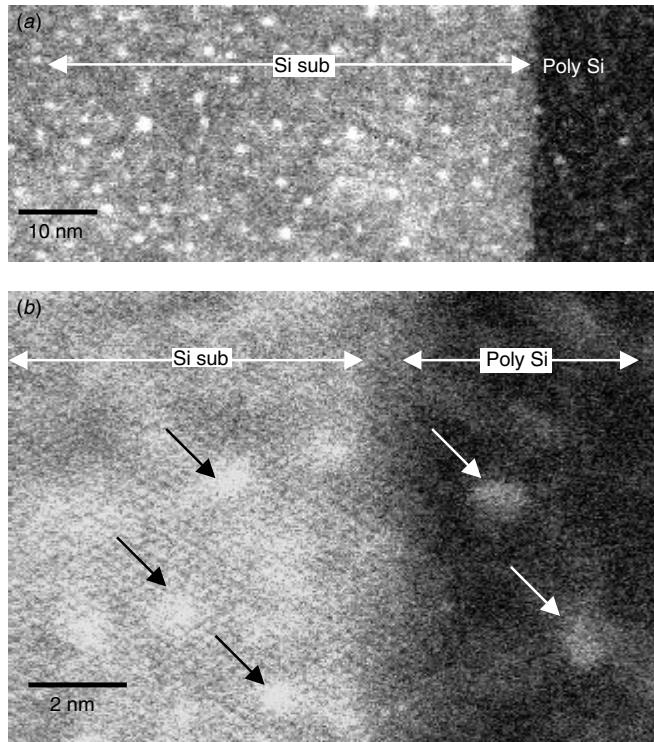


Figure 4.46 Bright patches in the STEM/HAAADF image are observed in FIB-prepared TEM sample. The bright patches, indicated by arrows, are Pt and Ga rich re-depositions as confirmed by EDS and EELS analysis.

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

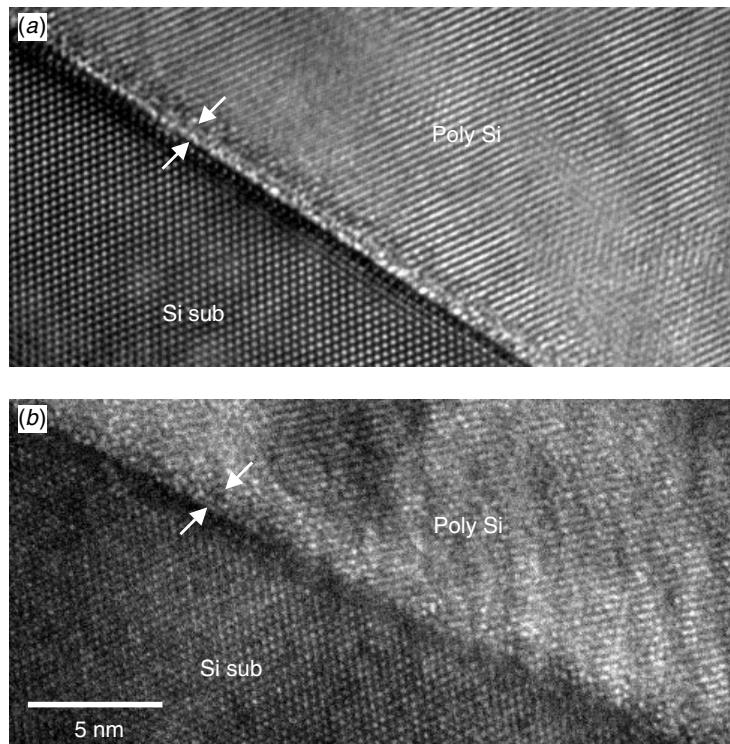


Figure 4.47 HRTEM cross sections of the native oxide layer analysis between polysilicon and Si substrate. (a) Sample prepared by mechanical polishing with less than 5 minutes ion milling; (b) FIB thinned sample. The thin oxide between poly and Si sub can be clearly observed in (a) but not as well in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

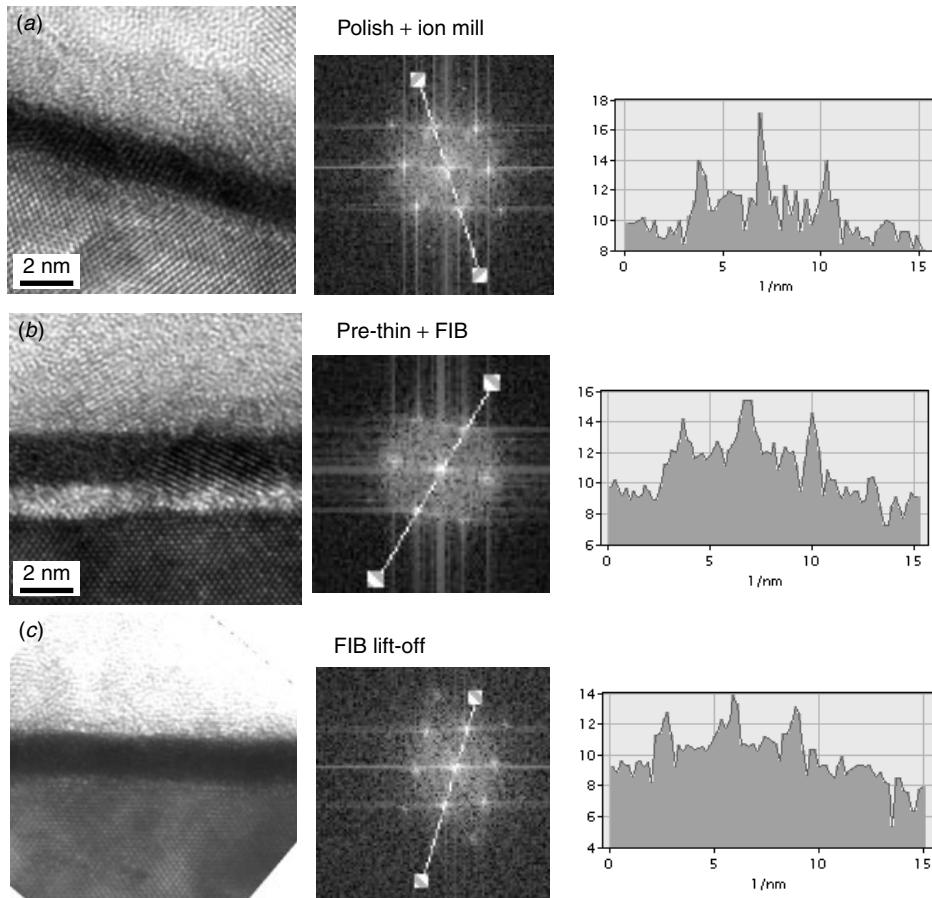


Figure 4.48 HRTEM cross sections of the high- k HfO_x oxide layer between the polysilicon and Si substrate. (a) Sample prepared by mechanical polishing with less than 5 minutes ion milling, (b) pre-thinned sample, and (c) lift-out FIB sample. Also shown are the corresponding FFT images (center) and their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

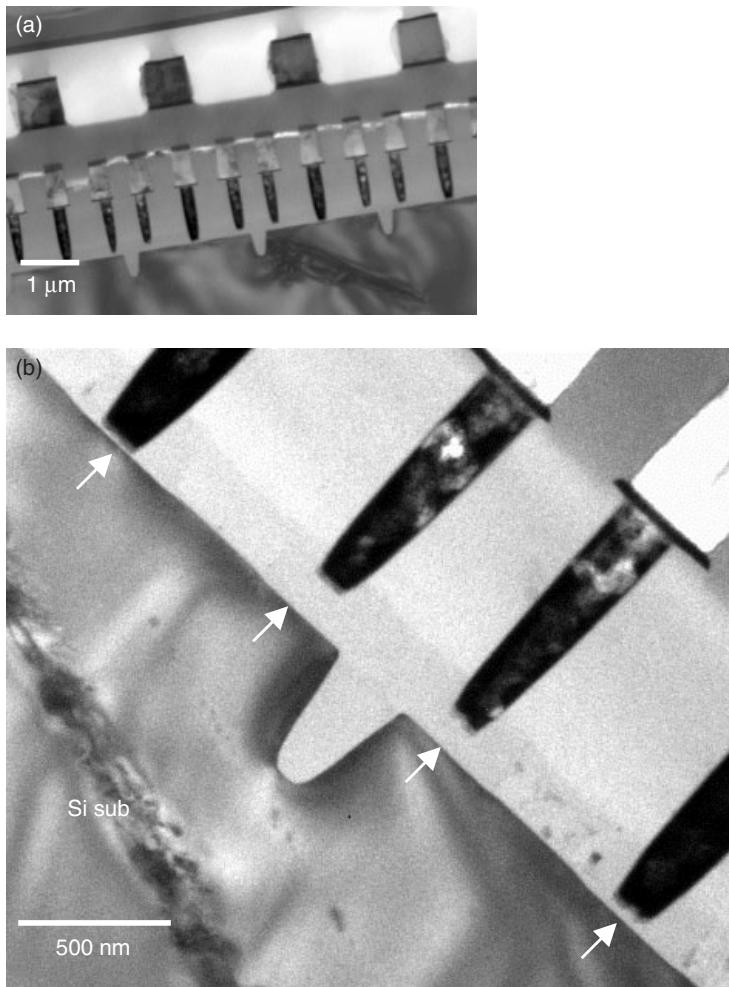


Figure 4.49 TEM analysis of contact under-etch issues. To check the under-etch contacts, as in (a), one needs to look through a large area of the cross section. (b) The contacts along the diameter show a severe contact etching under-etch problem, as indicated.

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.

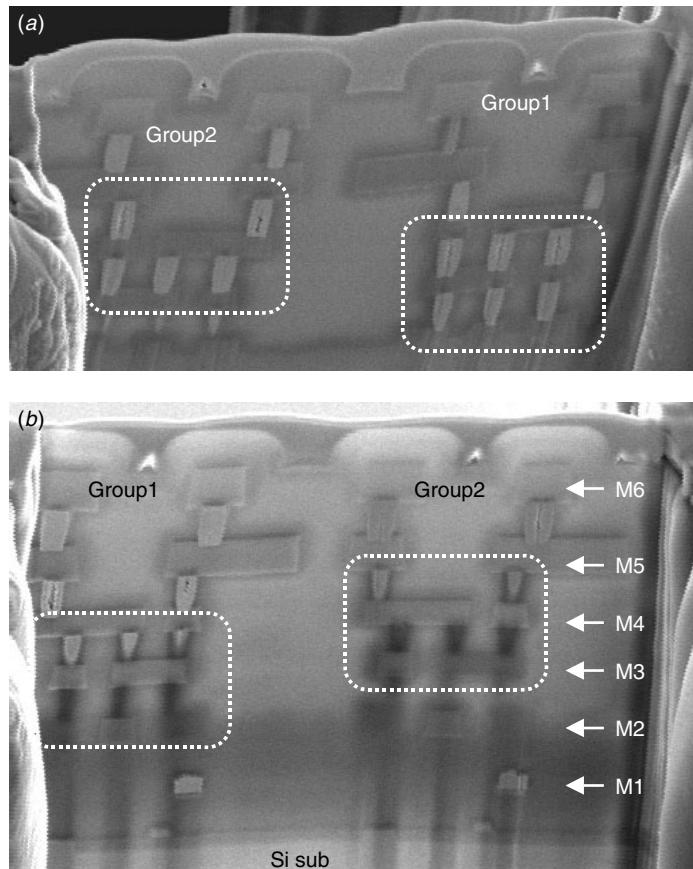


Figure 4.50 SEM images of both sides of an FIB cut TEM thin membrane. The VIAs and contacts appear exactly at the center on one side but do not appear at the center on the other side, as marked.

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

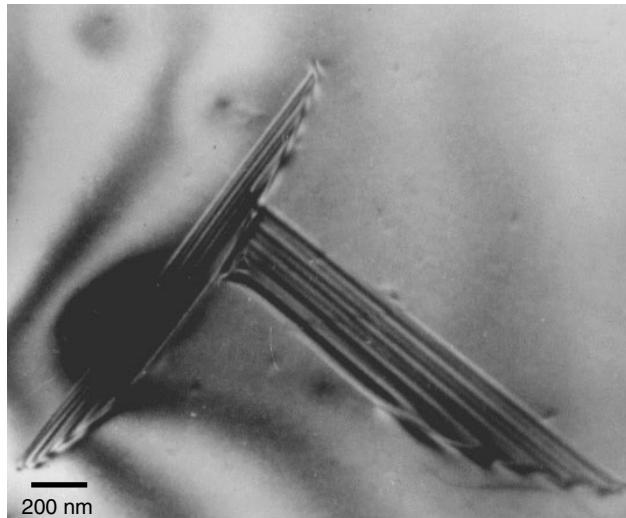
- R. Anderson, IBM, private communication, 1990.
- D. J. Barber, *J. Materials Sci.* **5**, 1–8, 1970.
- J. M. Brown and T. T. Sheng, *Materials Res. Soc. Symp. Proc.*, **115**, 29–38, 1988.
- C. J. Choi and T.-Y. Seong, *Proc. Microscopy of Semiconductor Materials Conf.*, Institute of Physics, 491–494,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

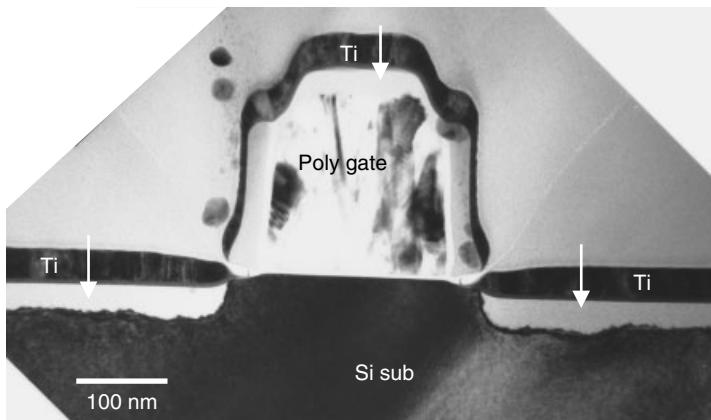


Figure 5.1 TEM cross section of an as-implanted salicide gate structure and its source/drain active areas. The ion implantation induced amorphization, as indicated, is clearly observed. The Ti metal was deposited immediately after the ion implantation. The spherical particles are TEM sample preparation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

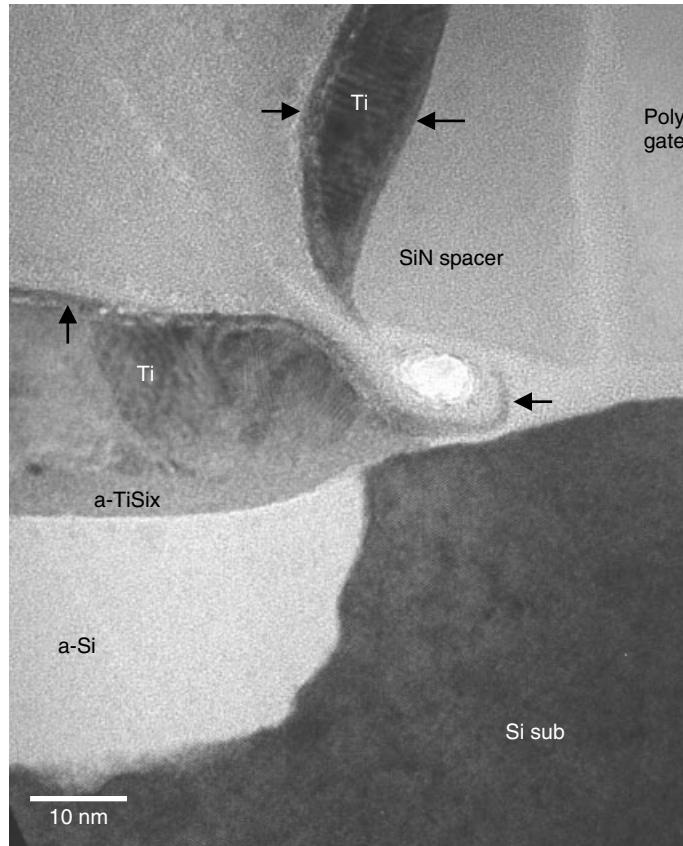


Figure 5.2 Close-up of TEM cross section at the as-implanted salicide gate structure's corner. The as-deposited Ti immediately reacts with amorphous Si and forms amorphous Ti silicide. Notice that the amorphous Si edge coincides with SiN spacer edge, which is due to the self-aligning implantation. The 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

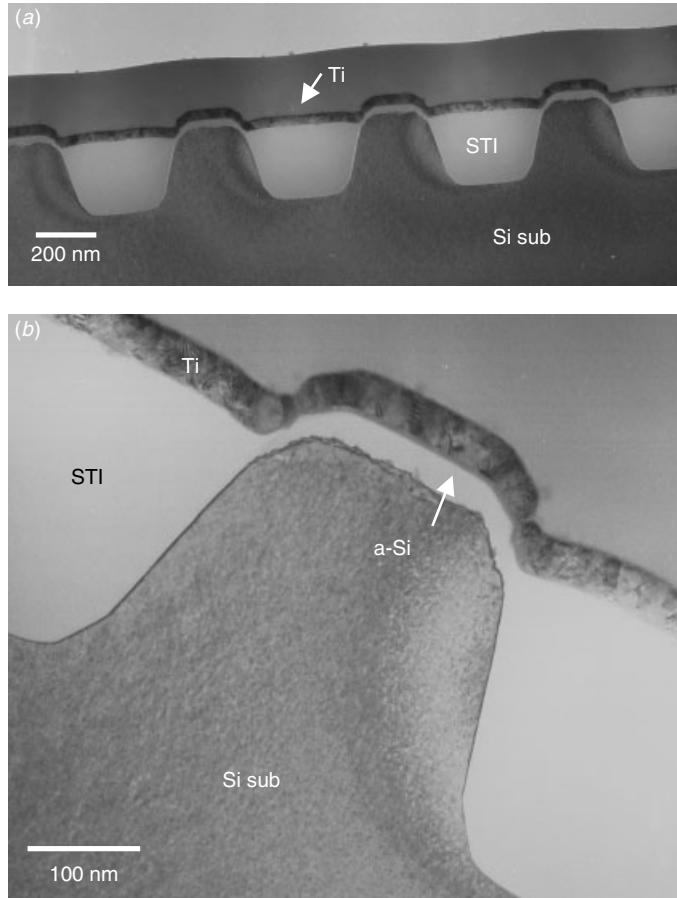


Figure 5.3 TEM cross section of as-implanted salicide active areas with STI isolation trenches. (Sample courtesy Prof. Kin Leong Pey, NTU Singapore)

(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

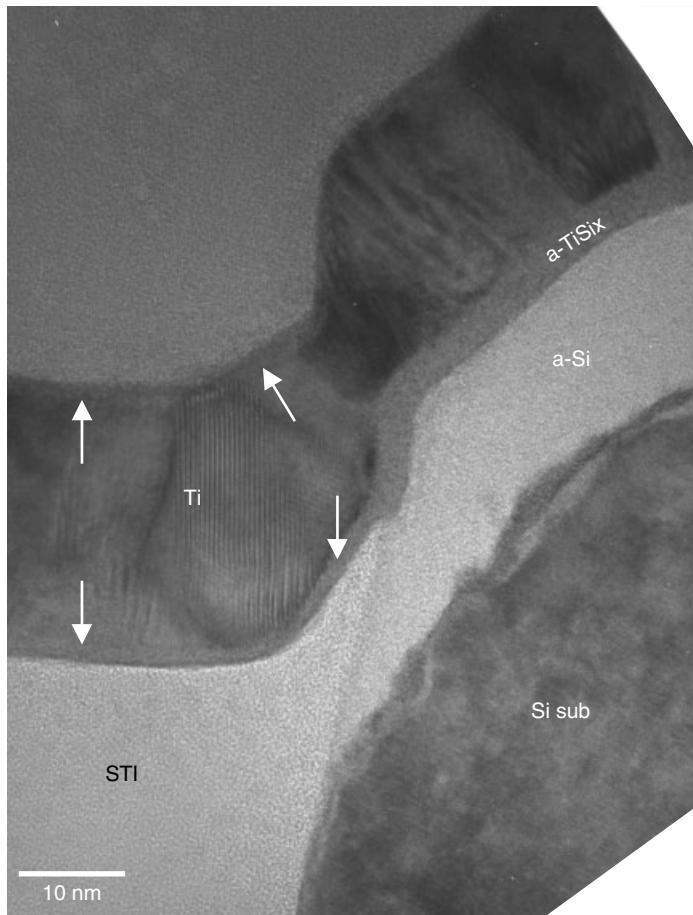


Figure 5.4 TEM cross section of the as-implanted salicide active area corner with the STI isolation trench. Notice that the a-Si follows the original Si-substrate surface's contour. Amorphous TiO_x has also formed on the STI oxide layer, as indicated.

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

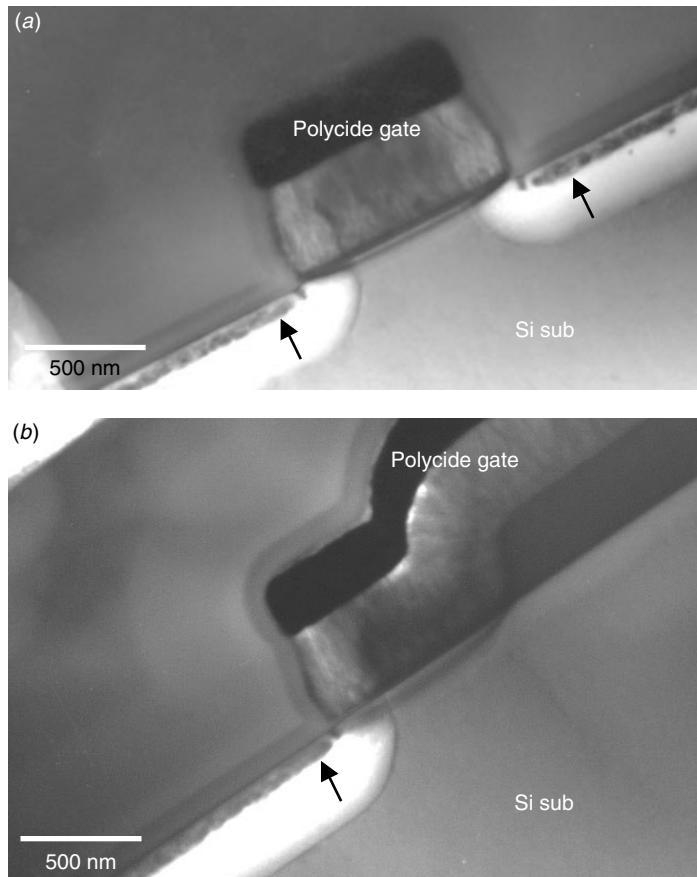


Figure 5.5 TEM cross section with junction delineation is used to reveal the junction profiling and location close to the polycide gate. Notice that the ion implantation induced substrate damages is close to the Si-substrate surface, as indicated.

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.

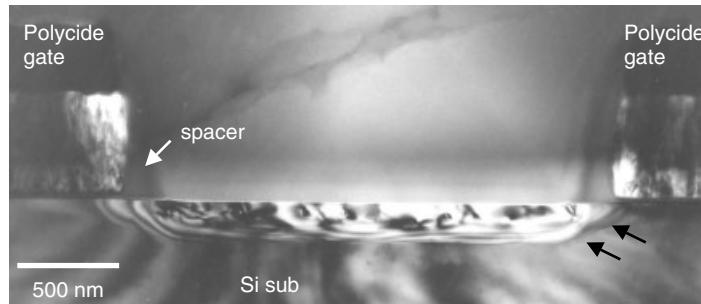


Figure 5.6 LDD structure as seen in a TEM cross section with delineation of the junction's shape and position near the polycide gate.

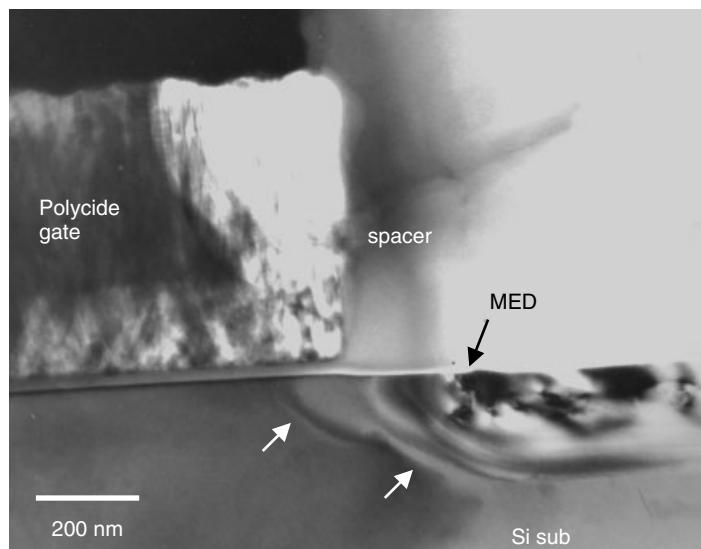


Figure 5.7 TEM cross section with junction delineation showing clearly the two-dimensional LDD junction, mask edge defects (MED), and spacer.

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

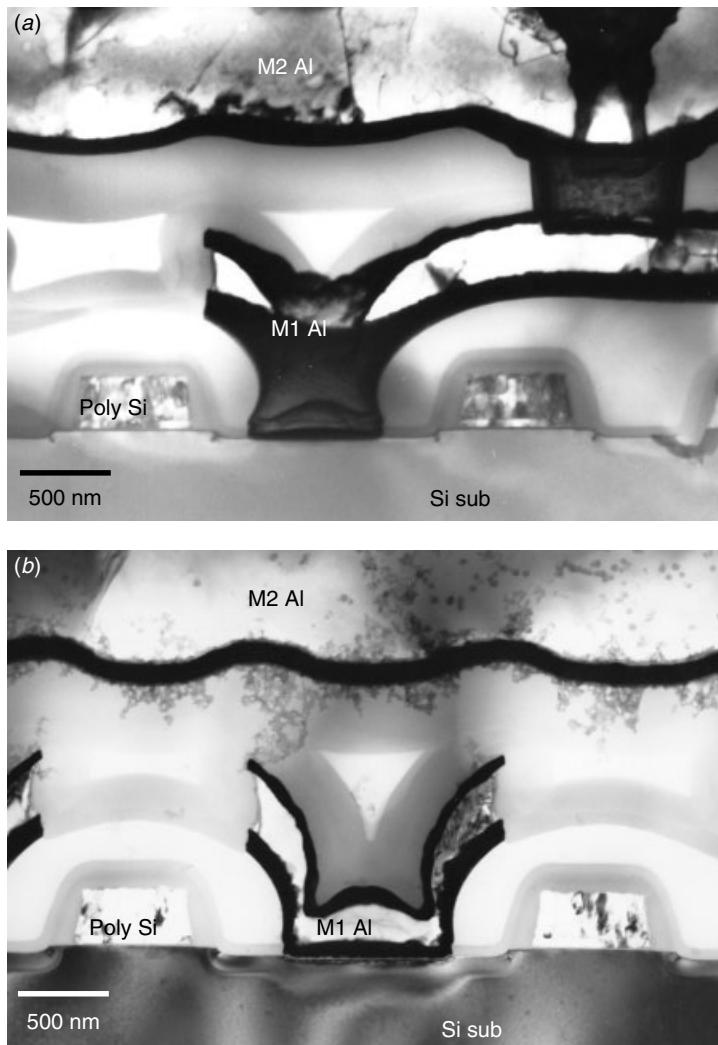


Figure 5.8 TEM cross sections (a) before and (b) after junction delineation. The junction delineation reveals the junction shape and position at and near a contact and two polygates.

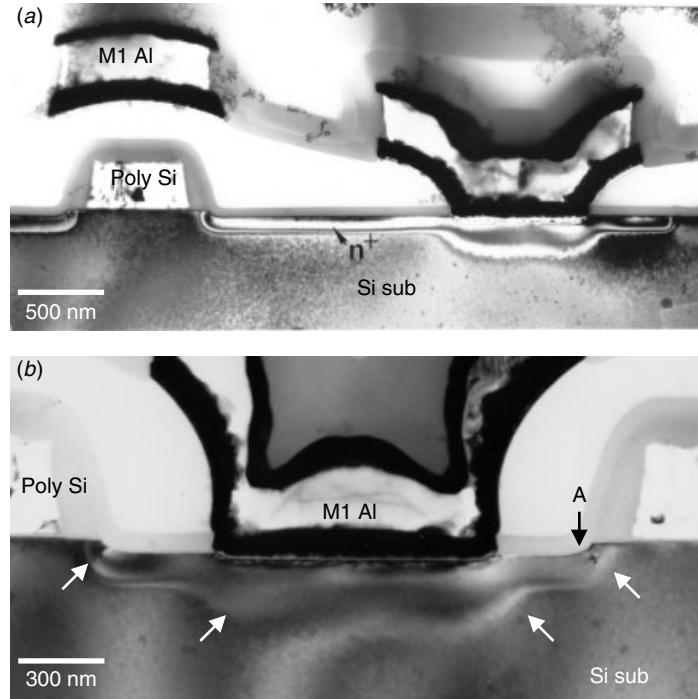


Figure 5.9 TEM cross-sectional views with junction delineation to reveal the junction under a metal contact. The junction right under the metal contact is deeper, as indicated, and is due to secondary ion implantation. Arrow A shows a mask edge defect right at the edge of the contact's overetch and spacer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

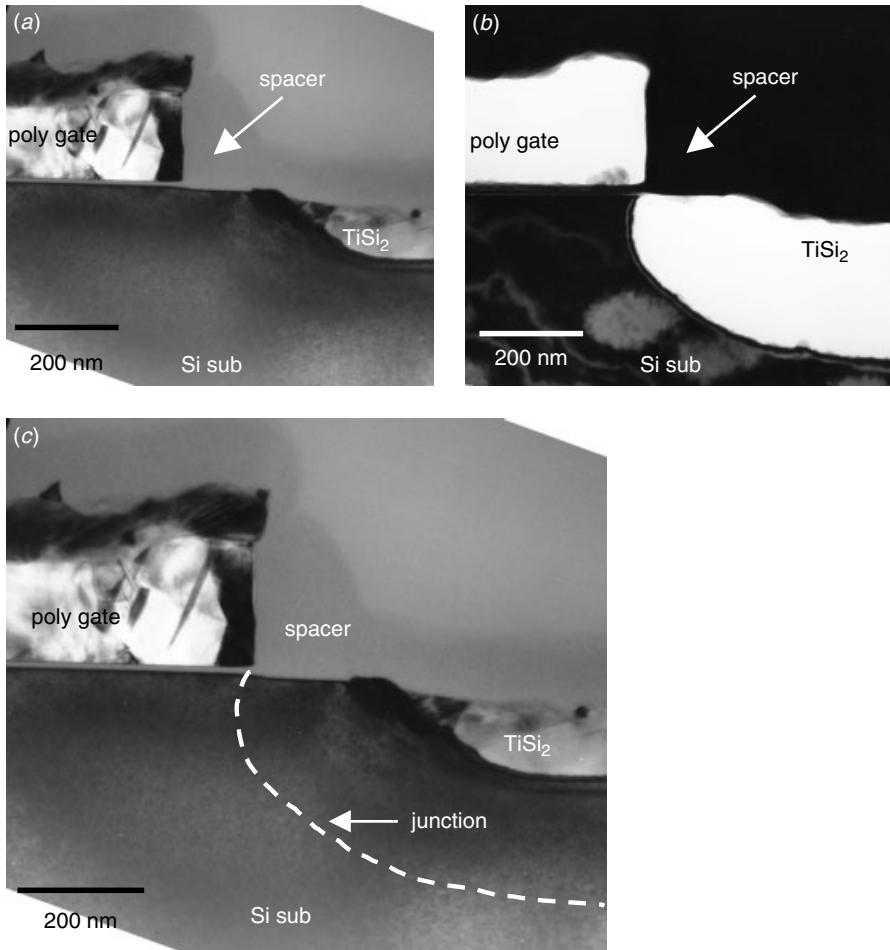


Figure 5.10 TEM cross-section views with junction delineation to reveal the junction near a salicide gate. (a) Before delineation, the spacer and salicide are clearly observed. (b) After delineation, the polygate, salicide, and doped active area in Si substrate are removed. Notice that the substrate 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

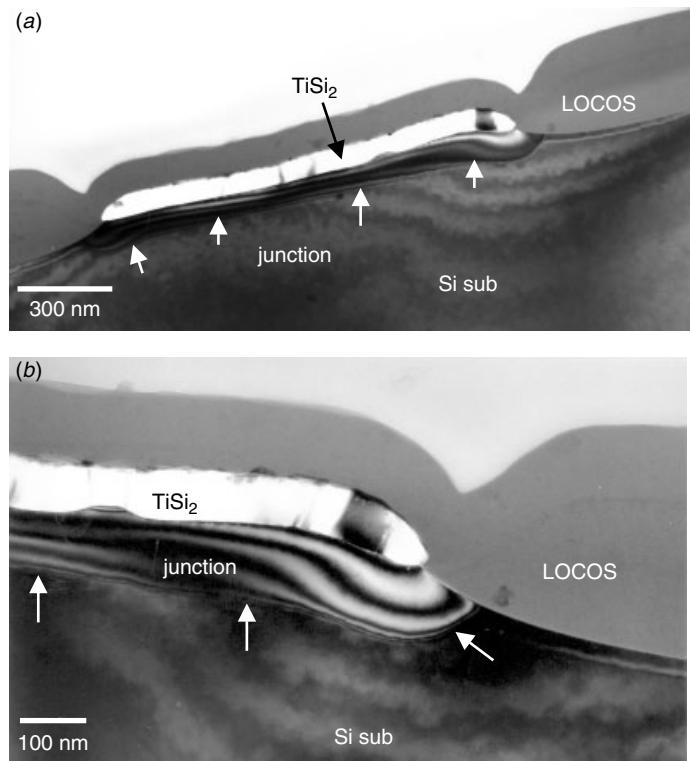


Figure 5.11 TEM cross-sectional views with junction delineation to reveal the junction near a LOCOS/salicide edge. A very shallow junction can be achieved (about 100 nm) by ion implantation into the salicide and diffusion out into the Si substrate to form the junction. (Sample courtesy Prof. Kin Leong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

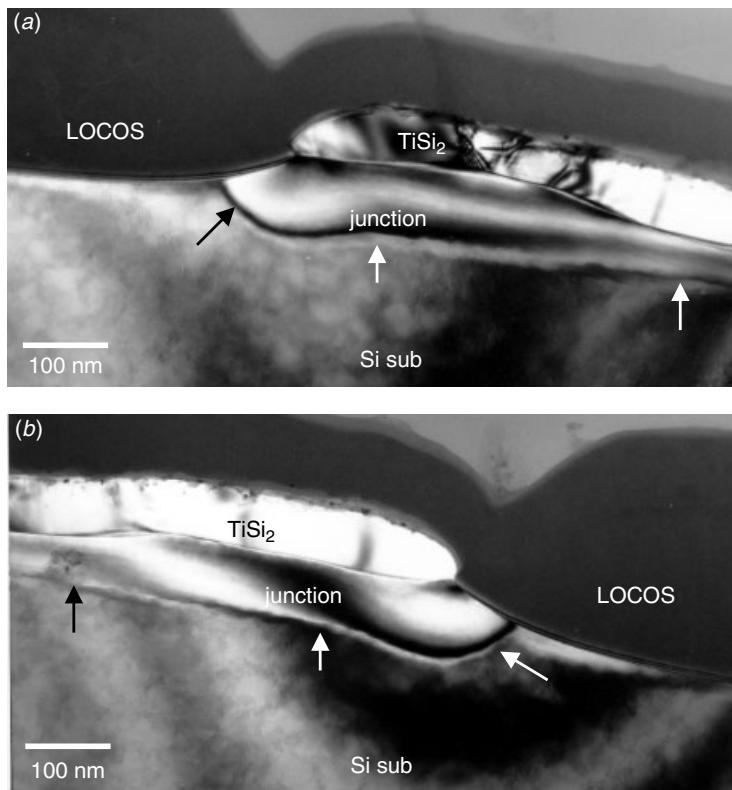


Figure 5.12 TEM cross-sectional views with junction delineation to reveal the junction profile near LOCOS/salicide edge. The junction basically follows the silicide's contour and is deeper at the salicide's edge and shallower at the center of the active areas, as indicated. (Sample courtesy Prof. Kin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

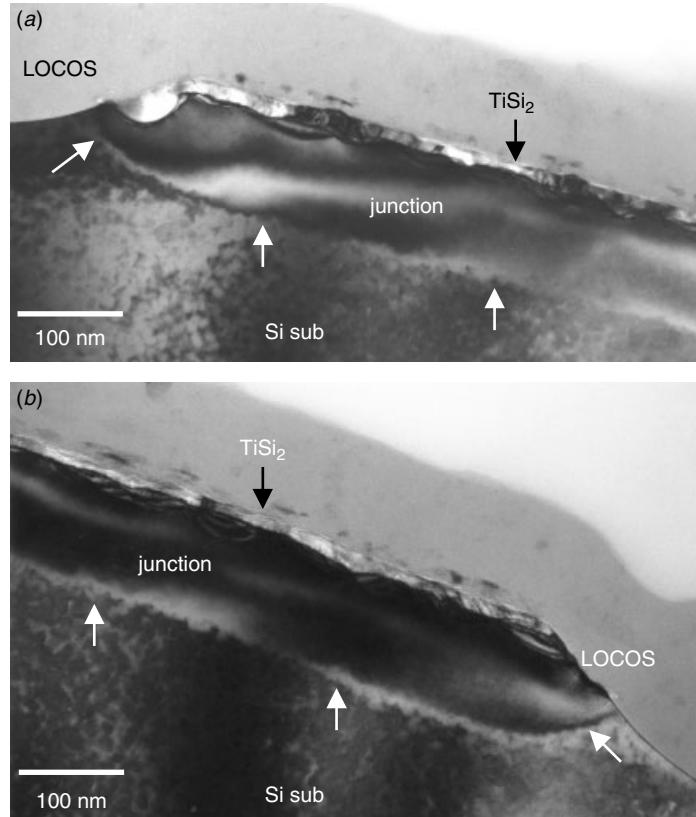


Figure 5.13 TEM cross-sectional views with junction delineation to reveal the junction profile near LOCOS/salicide edge. The ultra thin silicide is seen to have a shallow junction. Unlike the previous case, the junction profiles do not follow silicide's contour because the ion implantation was not done 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.

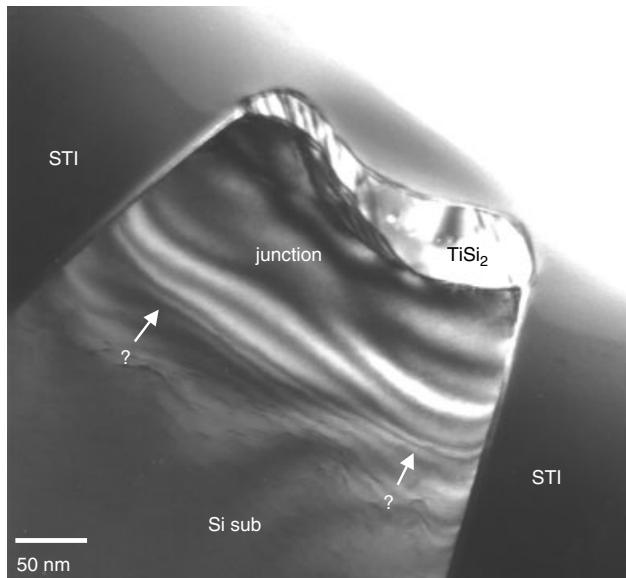


Figure 5.14 TEM cross-sectional view with junction delineation. Not all of the junction delineation can reveal a sharp junction interface, as in this case. The staining induced thickness contour changes gradually, so no single junction profile can be determined. (Sample courtesy Prof. Kin Leong Pey,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

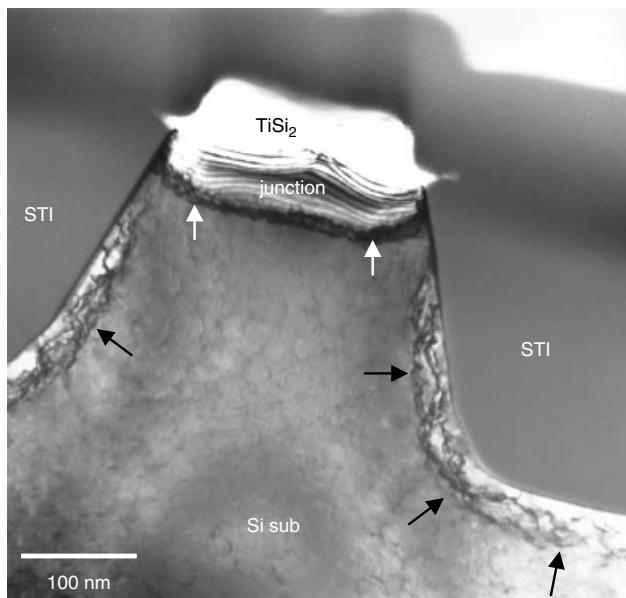


Figure 5.15 TEM cross-sectional view with junction delineation to reveal the junction profile near STI edge. Besides the active area junction, the STI implant that prevents cross STI leakage is observed, as shown by the black arrows. (Sample courtesy Prof. Kin Leong Pey, NTU Singapore)

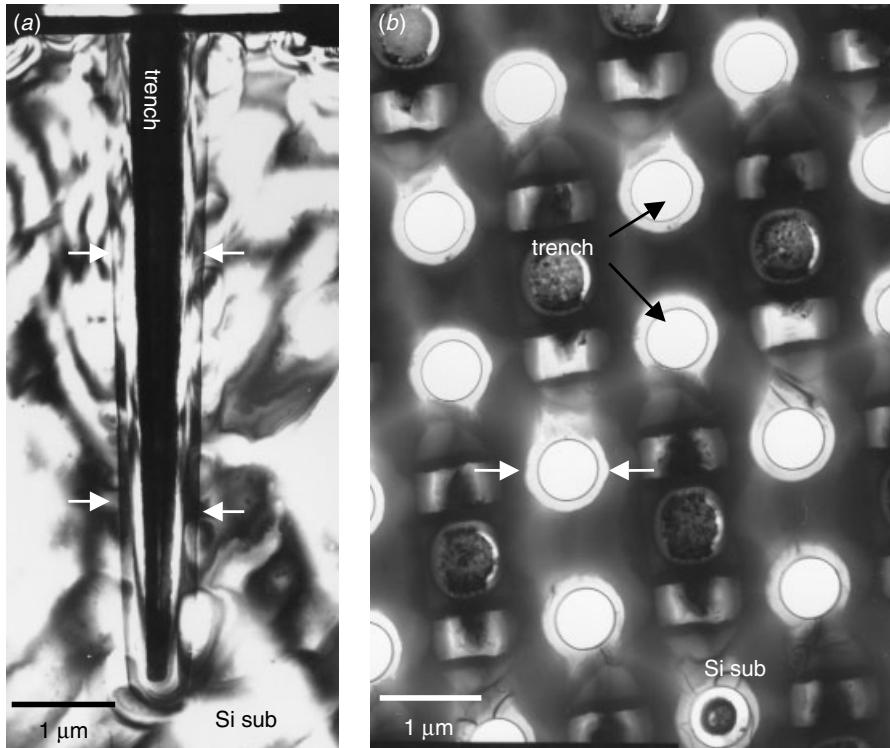


Figure 5.16 TEM cross section (*a*) and plan view (*b*) of a trench capacitor in the DRAM device's cell area. Notice that the centers of the trenches are removed in the plan view, and this is due to the strong chemical attack. The junction delineation reveals the doping profile to run the along deep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

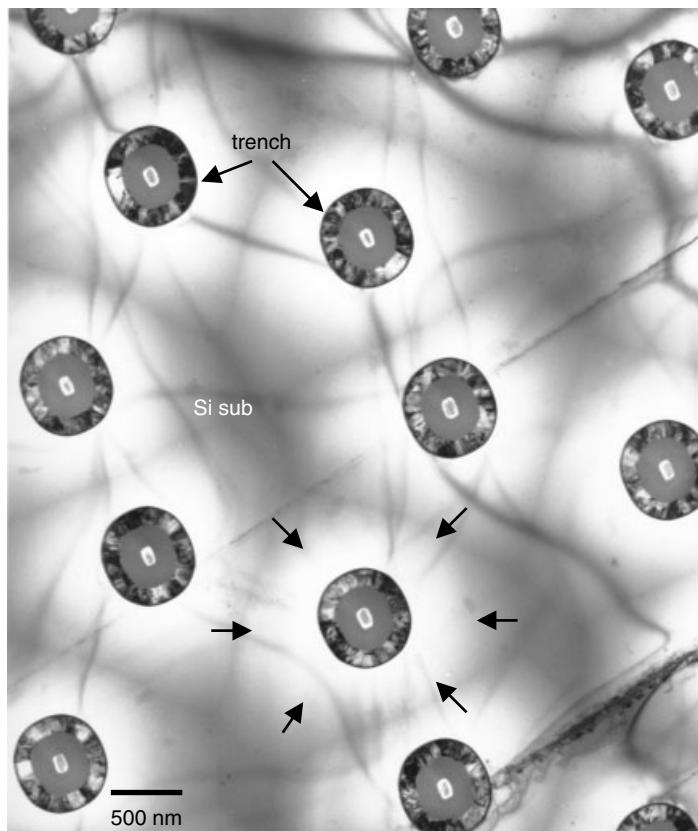


Figure 5.17 TEM plan view of the trench capacitor in a DRAM device's cell area. Junction delineation reveals the doping profile to be located at the circular trenches, as indicated by arrows, but the exact location of the junction is difficult to tell in this case.

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 SiO₂ 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 BF₂ or B11, the high-energy implantation forces BF₂ 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 F₂ gas bubbles. Both oxide precipitation in the n+ implantation

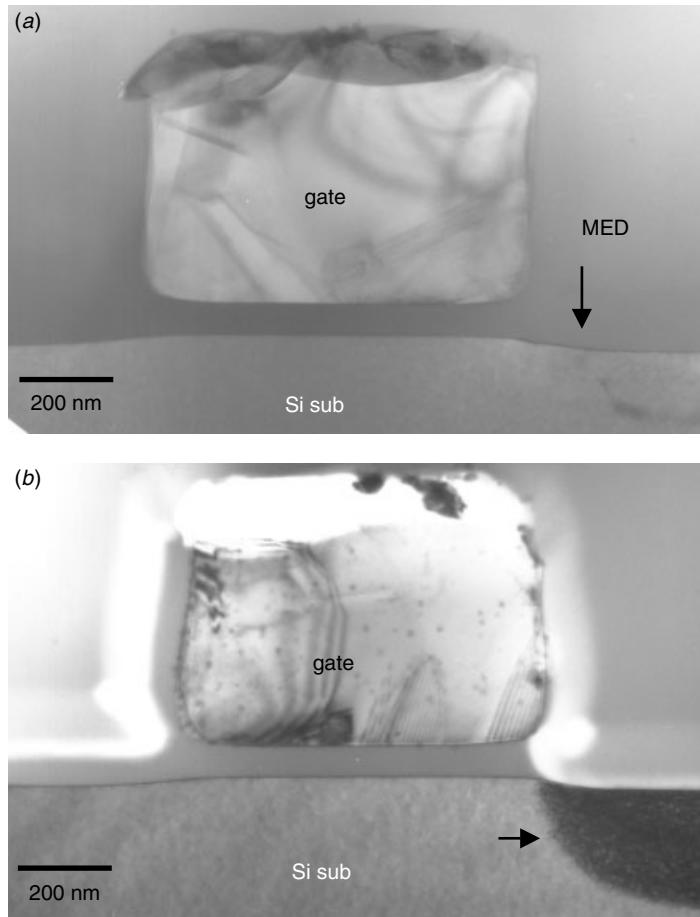


Figure 5.18 LD-MOS, TEM cross-sectional views with junction delineation taken to reveal the junction near a gate. Notice the rounded polygate with no sharp corners and heavily doped on one side of the device. To avoid stray capacitance, this doping cannot overlap with polygate.

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.

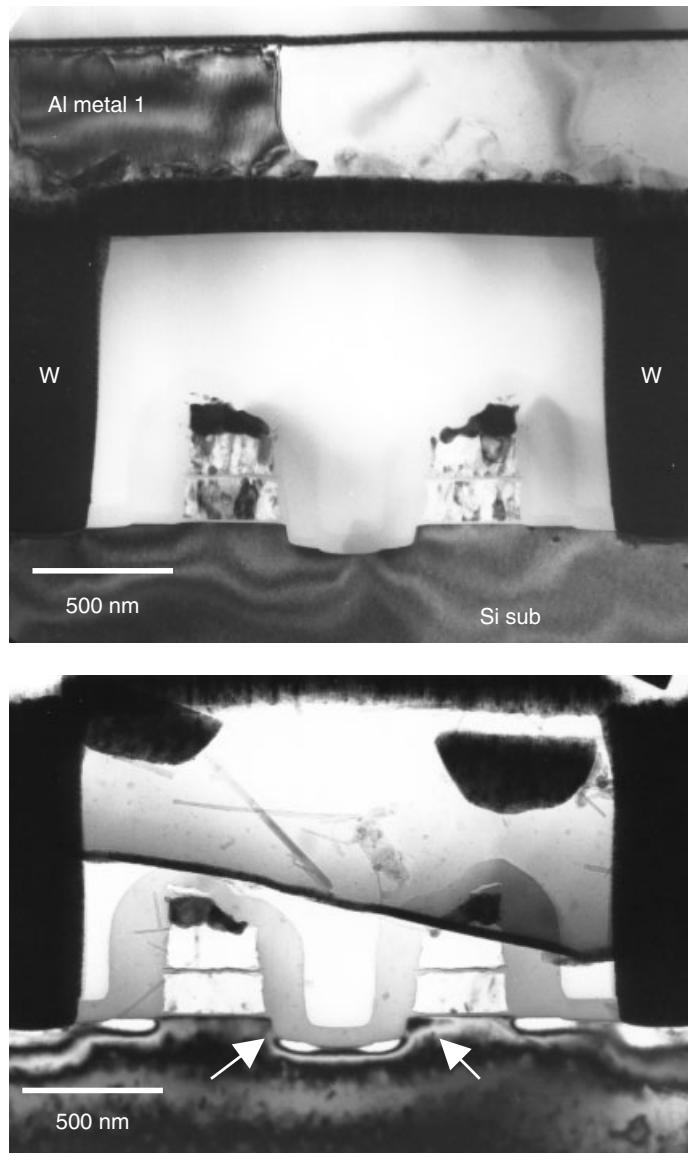


Figure 5.19 TEM cross-sectional views with junction delineation of the junction of a failed device. The junction reveals clearly that the device is unable to turn on because of the overetch dent in the middle of the image, as indicated. This common defect proves junction delineation to be a powerful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

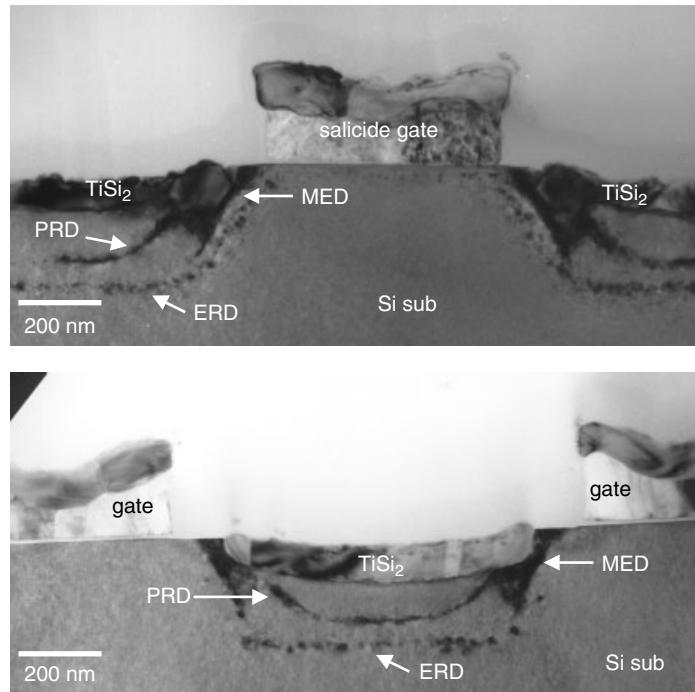


Figure 5.20 TEM cross section of the salicide gate and source/drain structures. The heavily implanted Si substrate shows clearly the mask edge defects (MED), projected range defects (PRD), and end of range defects (ERD), as indicated. (Sample courtesy Prof. Kin Leong Pey, NTU Singapore)

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.

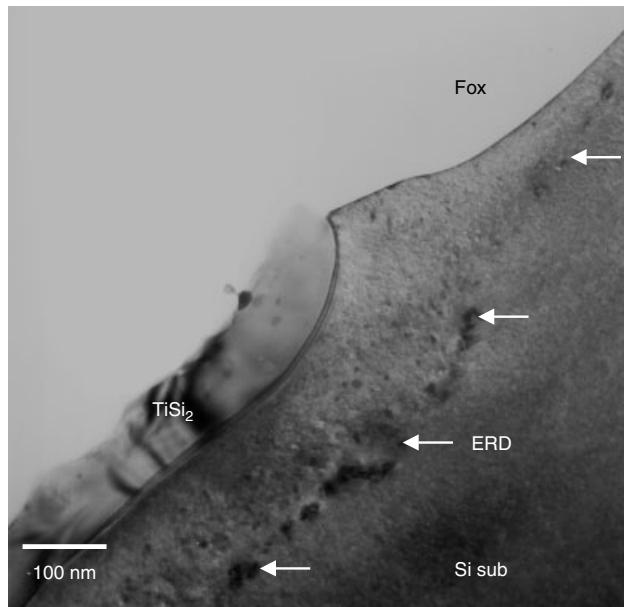


Figure 5.21 TEM cross section of the salicide's active area structure. The ion implantation with R_p in the Si substrate induces ERD well below $TiSi_2$, as shown. (Sample courtesy Prof. Kin Leong Pey, NTU Singapore)

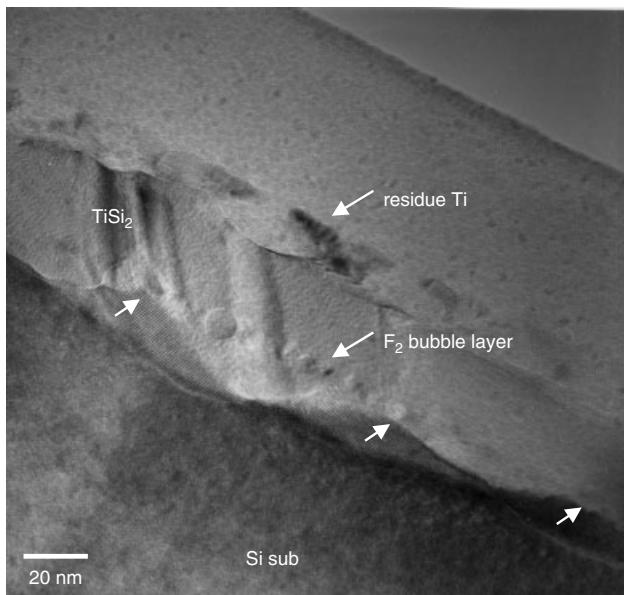


Figure 5.22 TEM cross section of the salicide's active area structure. The BF_2^+ implantation induced decomposed fluorine bubble is clearly visible. Usually the bubbles aggregate at or near the projected range, R_p , of the corresponding implantation. (Sample courtesy Prof. Kin Leong Pey, NTU Singapore)

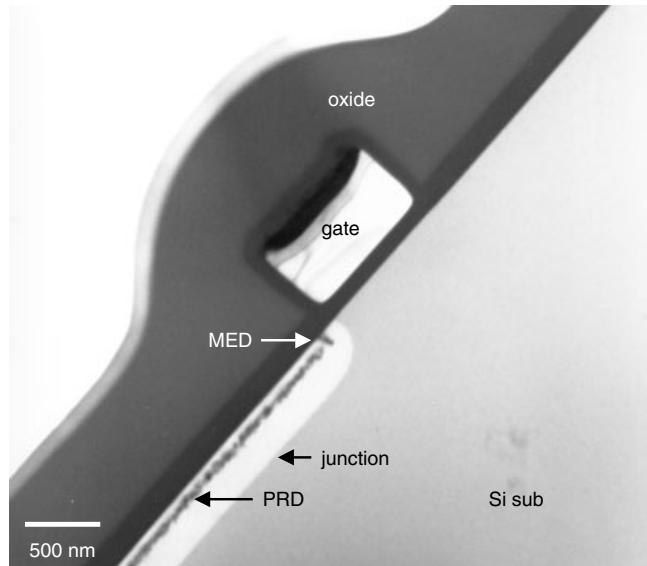


Figure 5.23 TEM cross section of an LDMOS at the polygate edge. Asymmetrical junction implantation is clearly seen here after junction delineation. Mask edge defects, projected range defects, and the junction profile are clearly visible.

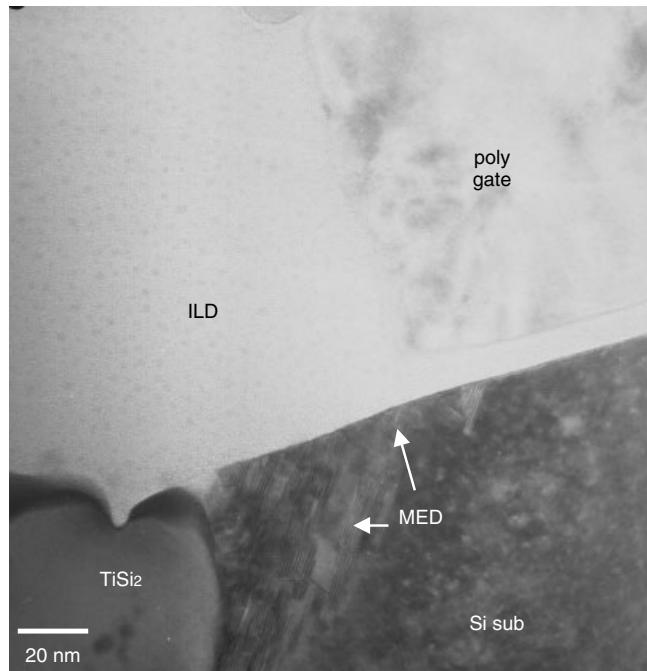


Figure 5.24 TEM cross section of a polygate edge. The mask edge defects formed at the edge of the polygate, and extending to TiSi₂ are clearly seen. No spacer was used in this technology.

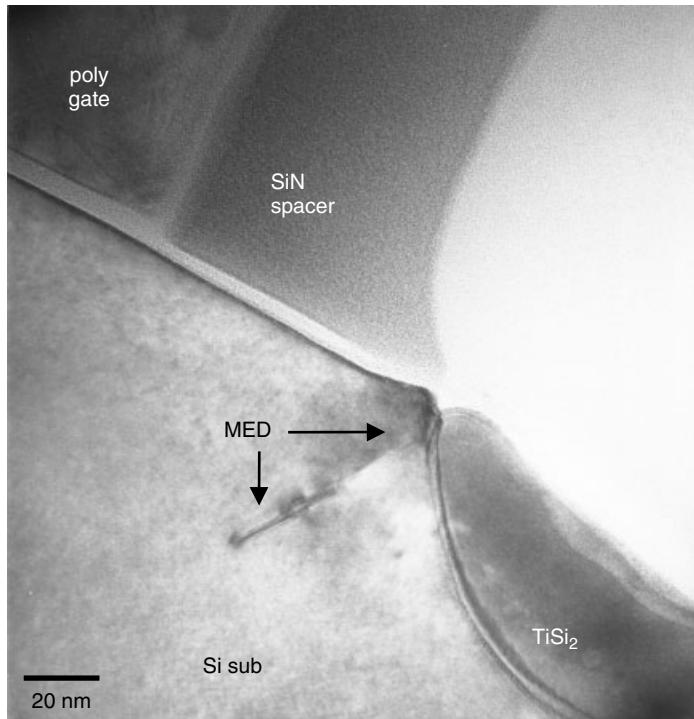


Figure 5.25 TEM cross section of a polygate edge. Mask edge defects are formed at the edge of SiN spacer/TiSi₂ interface.

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

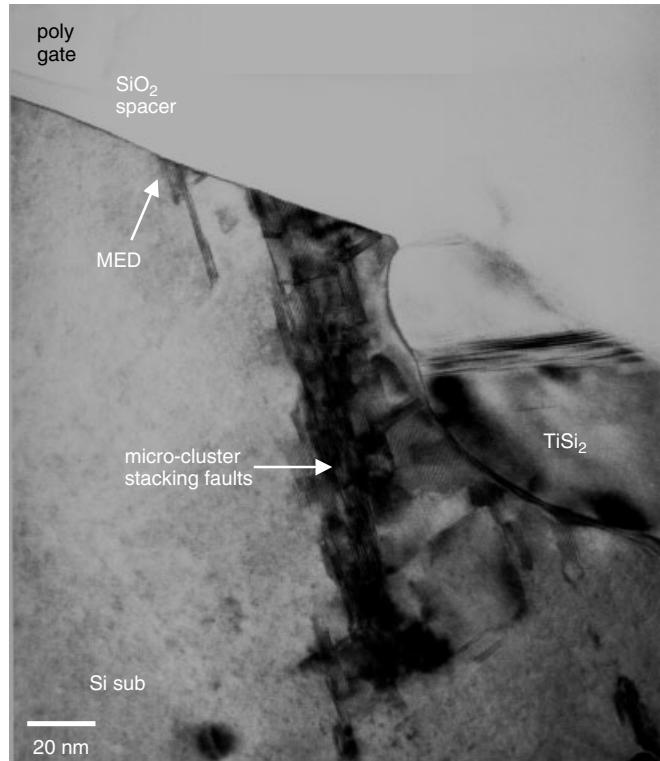


Figure 5.26 TEM cross section of a VLSI device's polygate edge. Mask edge defects are formed at the edge of SiO₂ spacer and extend into the TiSi₂ area.

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, $\mathbf{b} = \frac{1}{6}[011]$.

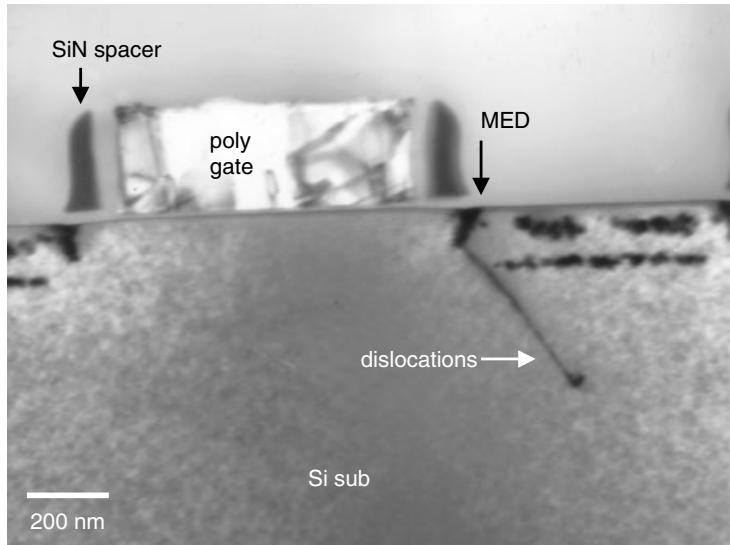


Figure 5.27 TEM cross section of the SiN spacer edge. The large stress between SiN spacer and ILD oxide layers cause the dislocations to nucleate from the MED.

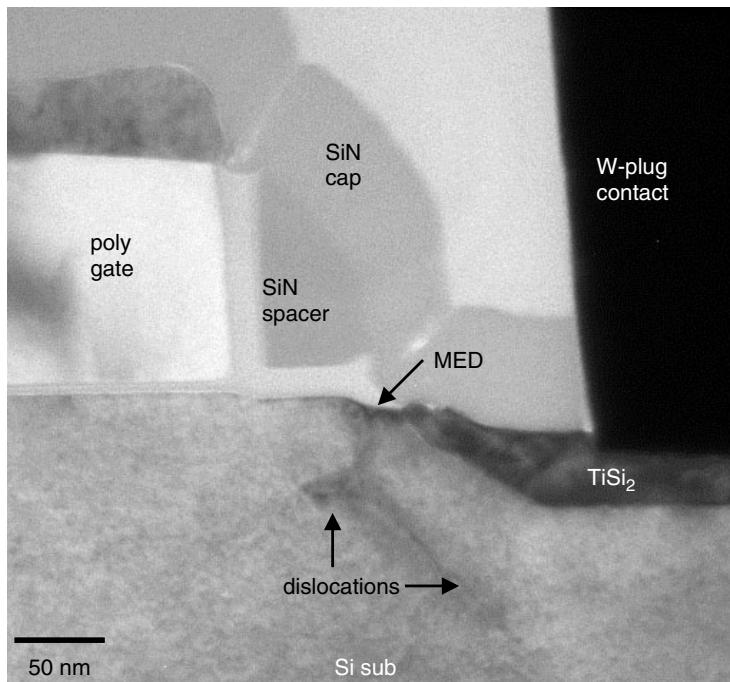


Figure 5.28 TEM cross section of the gate SiN spacer's edge. Mask edge defects appear to be the nucleation site for extended dislocations. A Si substrate surface step is associated with the MED.

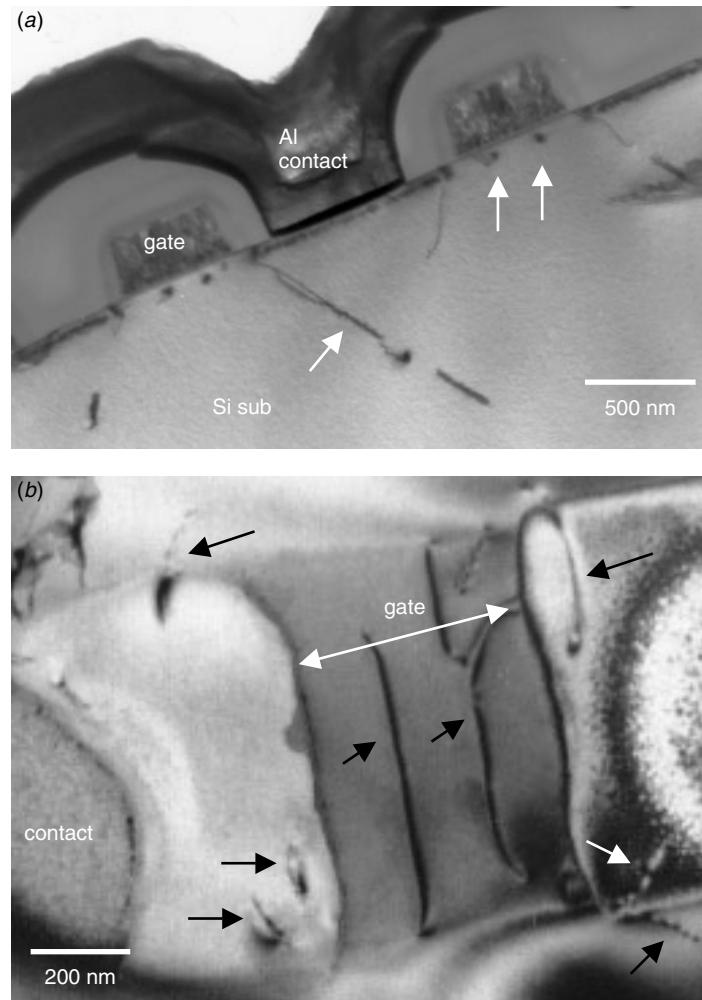


Figure 5.29 TEM cross section and plan view of MED-induced dislocations, as indicated. Dislocations going under the polygate can induce leakage and degrade the device's performance. These dislocations are edge dislocations with $\mathbf{b} = \frac{1}{2}[110]$.

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

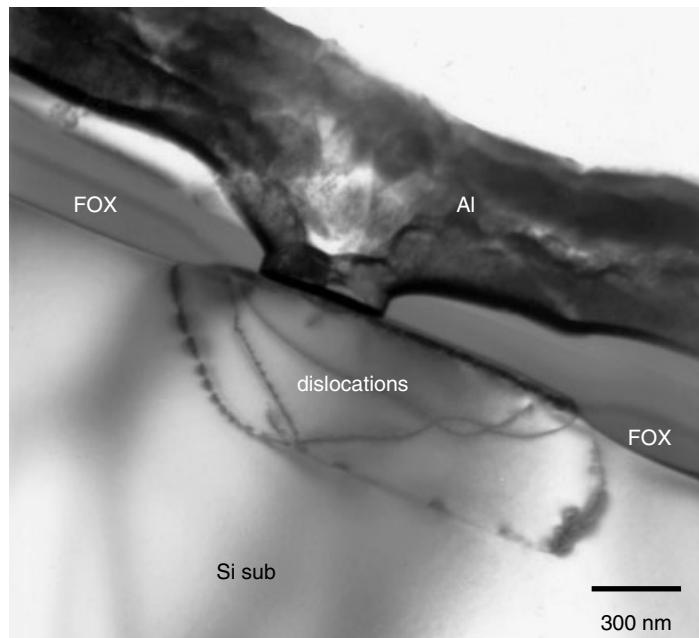


Figure 5.30 TEM cross section of tertiary dislocations, as indicated. Mixture of several types of dislocations are often observed, as dislocations start to interact with each other.

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).

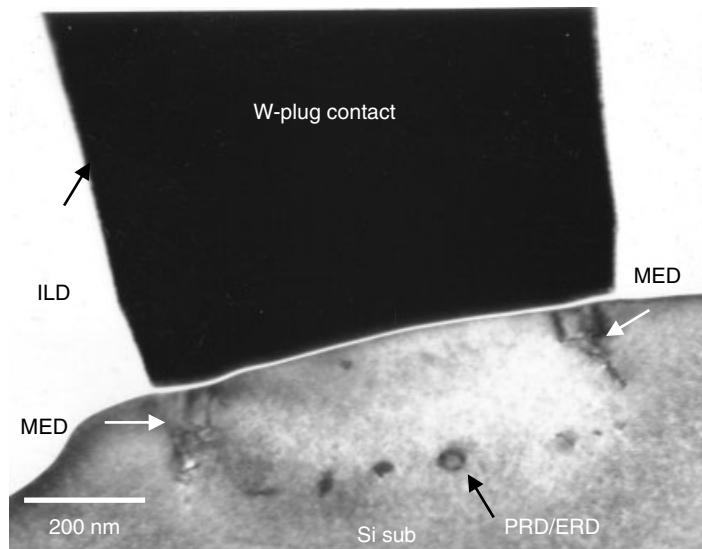


Figure 5.32 TEM cross section of the contact's bottom showing the PRD and ERD in contact to be deeper than those induced by s/d implantation. MED due to contact implantation is also visible.

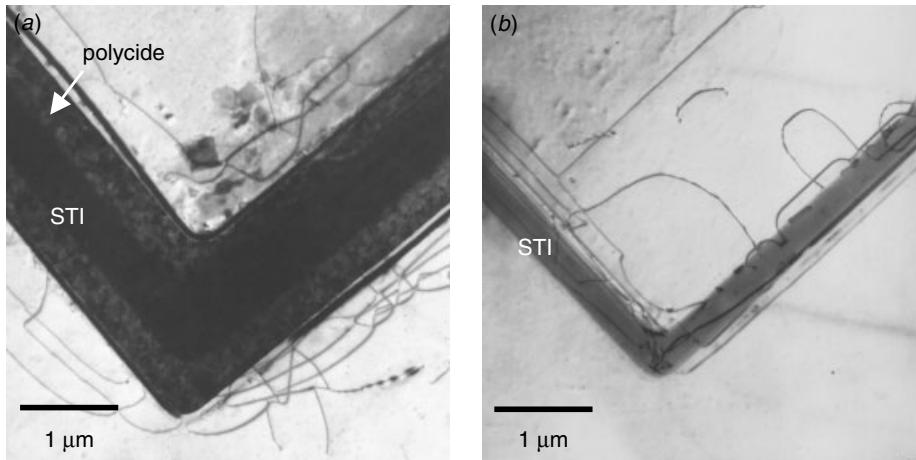


Figure 5.33 TEM plan view of the STI structure and its associated dislocation entanglements. (a) With the polycide runner covering STI and (b) with the polycide removed. Dislocations, apparently running in parallel with STI, appear to nucleate from STI sidewall as seen in (b).

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.)

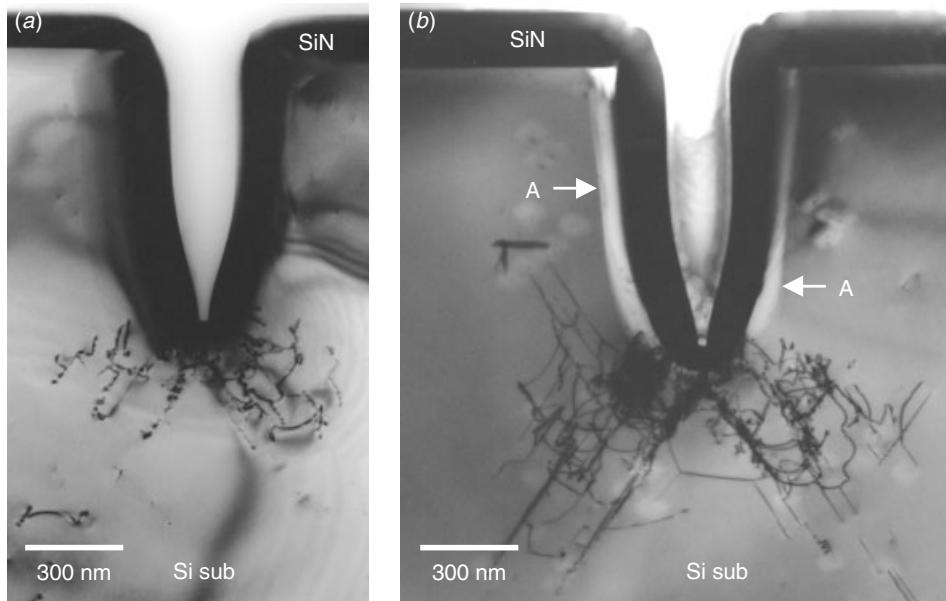


Figure 5.34 TEM cross section of a trench structure with a nitride lining. (a) Without junction delineation and (b) with junction delineation, where the junction is indicated by arrows. Dislocation tangles at the base of the trench are apparent.

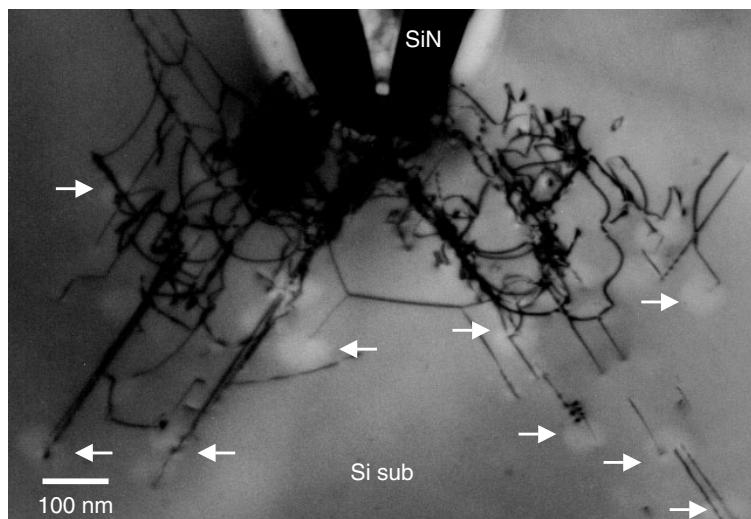


Figure 5.35 TEM cross section of the trench's base dislocations with junction delineation. Areas with local bright spots are delineated because of the high local concentration of dopants. The cross section directly proves that dislocations at or near junctions can act as diffusion pipelines and allow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

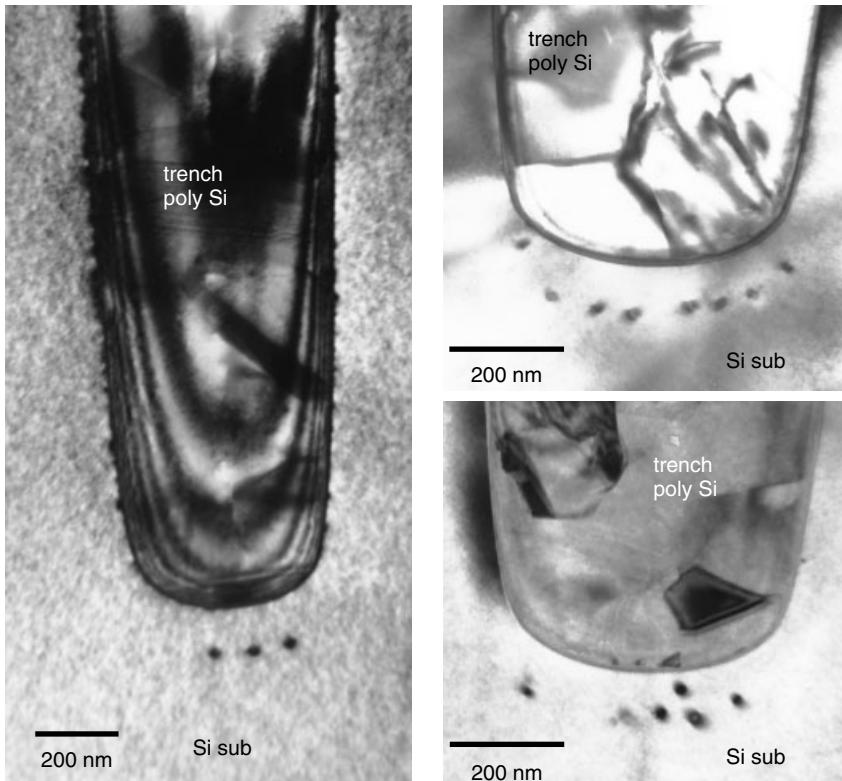


Figure 5.36 TEM cross section of the trench's capacitor structure with polysilicon fill. The defects distribution follows, more or less, the trench bottom's contours. These are the projected range defects (PRD) induced by the high dose ion implantation and post annealing.

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,

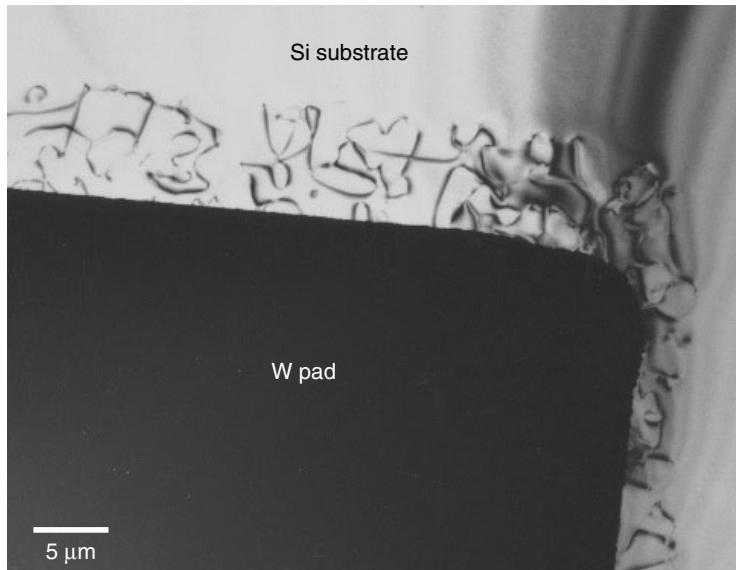


Figure 5.37 TEM plan view of a massive contact pad with a high dose of ion implant to Si substrate. PRD and ERD aggregation into the dislocation networks can be readily seen in the edge area of the contact implantation. The main contact array is covered by the W pad.

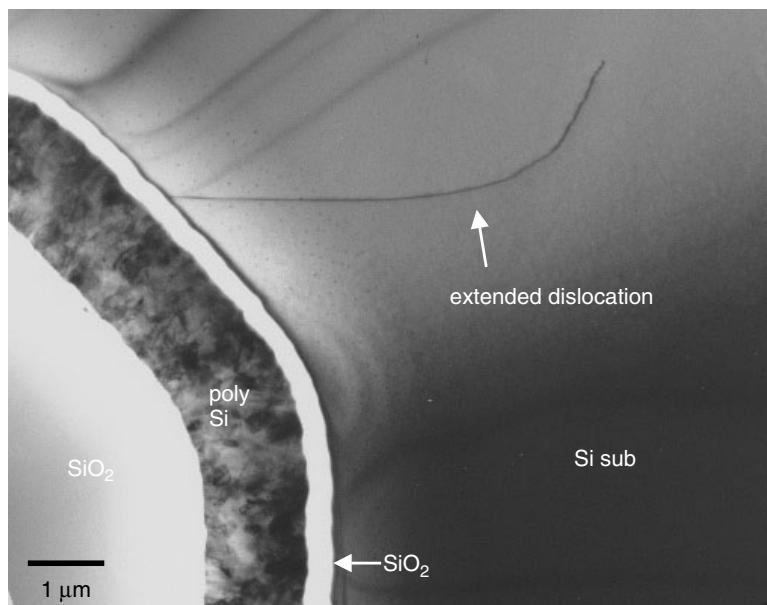


Figure 5.38 TEM plan view of a bipolar device isolation well. Extended dislocation from the isolation sidewall is observed. Stress induced by the thermal mismatch between the Si substrate and the SiO₂ trench fill is believed to be responsible for the dislocation.

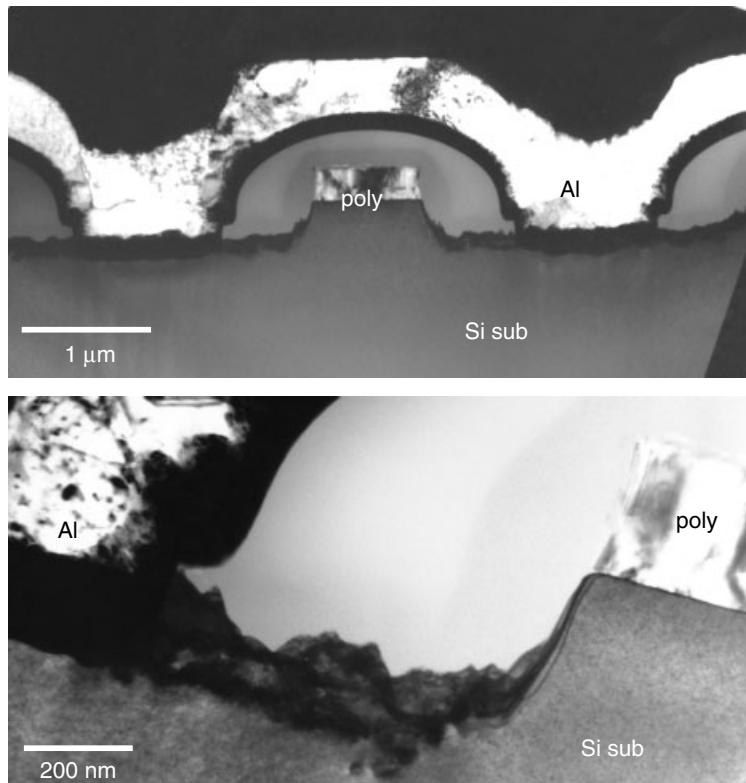


Figure 5.39 TEM cross section of a BiCOMS device's contact area with a large Si substrate overetch and a rough surface. Surprisingly, the uneven surface is not detrimental to the device's performance or reliability.

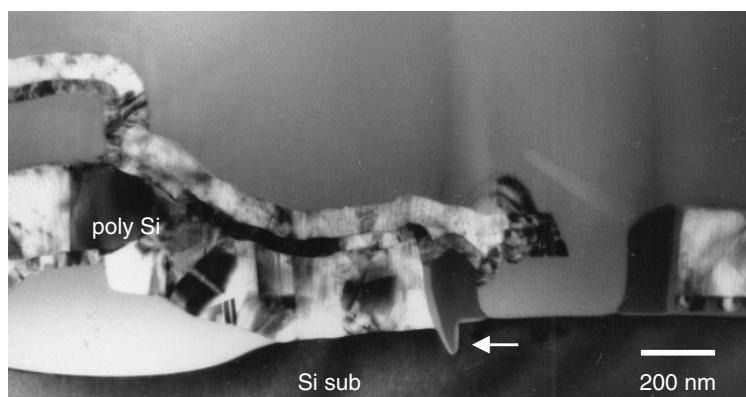


Figure 5.40 TEM cross section of an SRAM polysilicon buried contact. A groove next to the polysilicon contact, filled with spacer nitride, is observed.

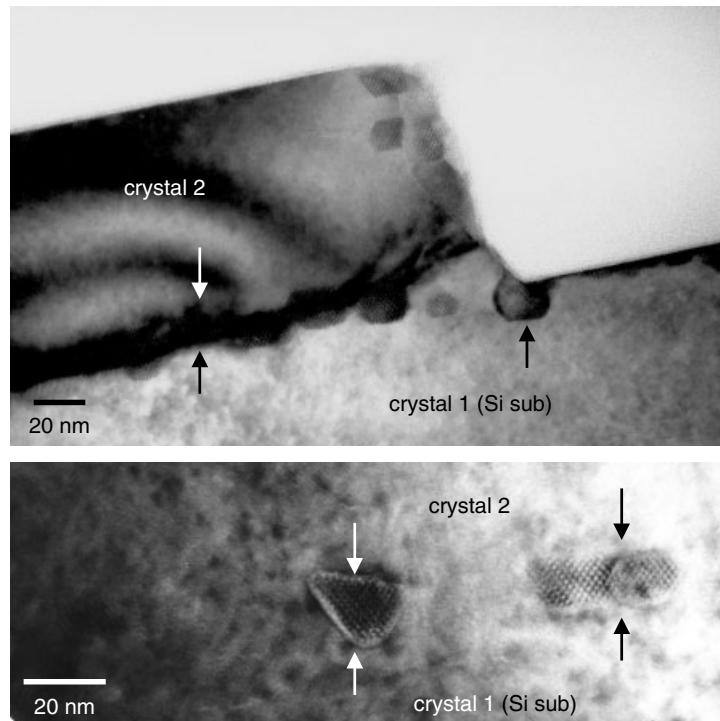


Figure 5.41 TEM cross section of a bicrystal joint. The interface defects along the joint can be identified and studied by conventional HR-TEM. The interface defects are identified to be oxide particles. (Sample courtesy and copyright Prof. King Ning Tu, UCLA)

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

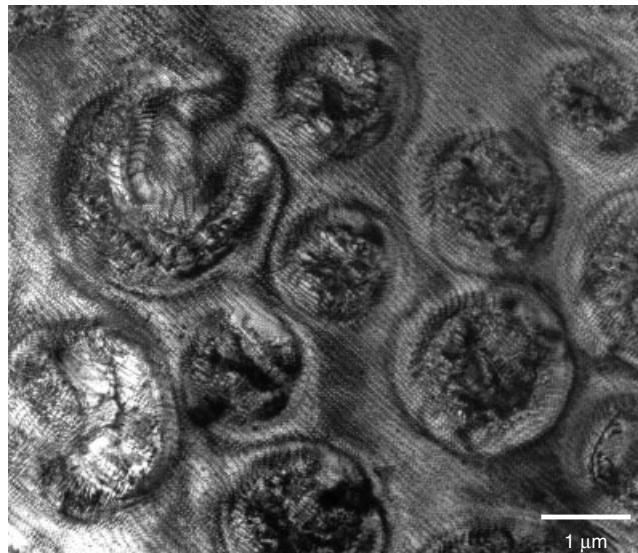


Figure 5.42 TEM plan view of a bicrystal joint. The epitaxial reaction starts randomly and spreads over to form circular islands with joining and non-joining areas in between. (Sample courtesy and copyright Prof. King Ning Tu, UCLA).

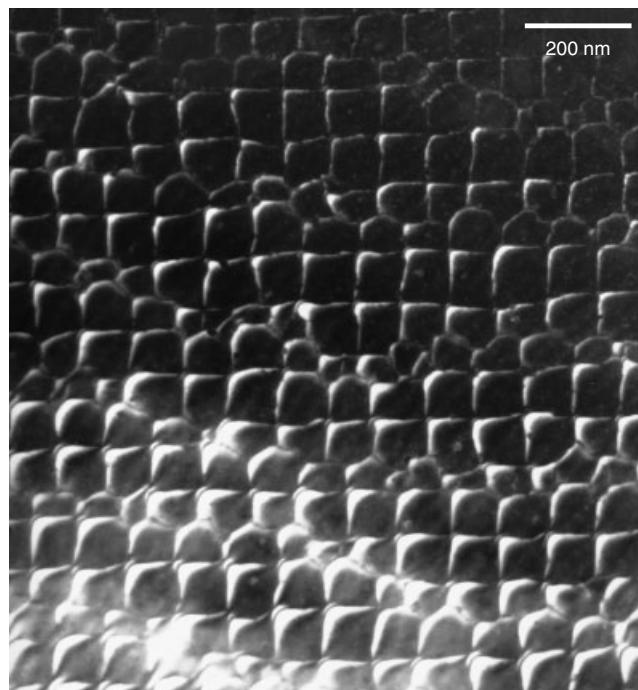


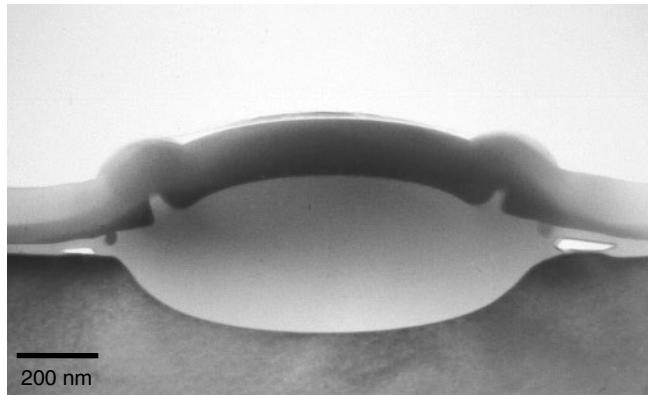
Figure 5.43 TEM plan view of a bicrystal crystal joining area. Dislocation networks (screw dislocations) are observed along the Si[110] direction. (Sample courtesy and copyright Prof. King Ning Tu, UCLA).

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

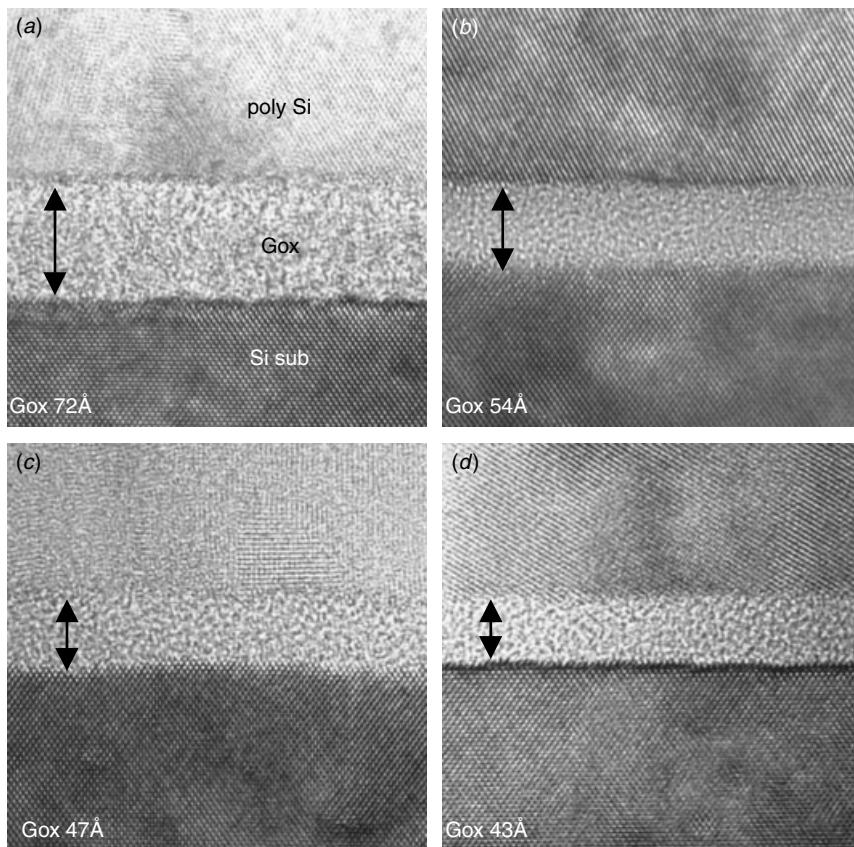


Figure 6.1 Some typical gate oxide layers in high-resolution TEM images. The Si substrate Si(111) cross lattice images provide an excellent internal standard for the exact gate oxide thickness measurement.

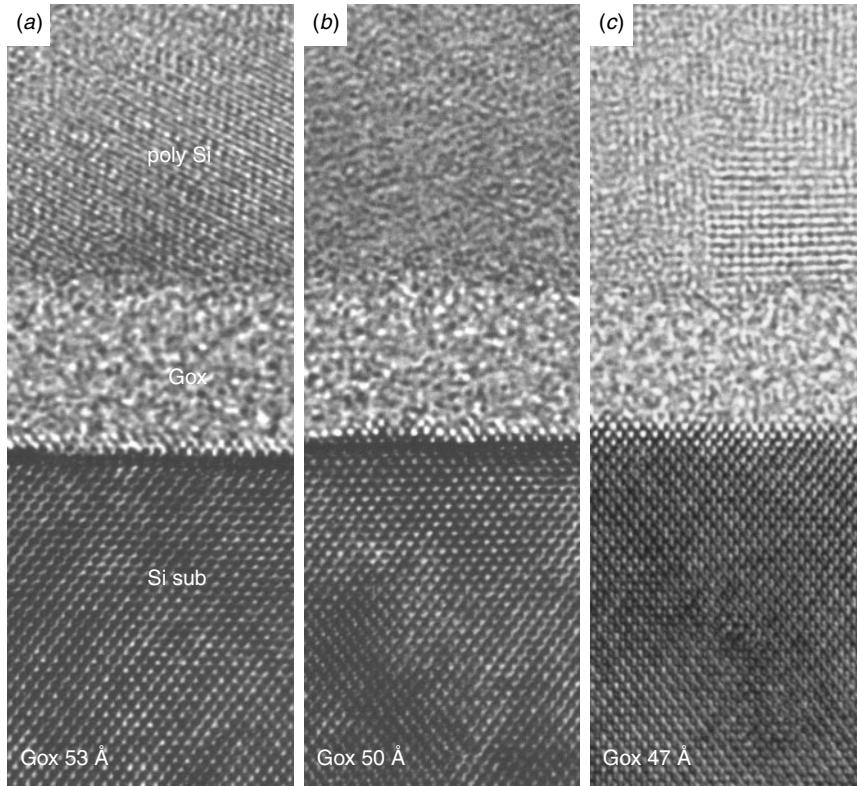


Figure 6.2 Three oxide layers with minor thickness differences. The polysilicon/oxide interface is laterally aligned so that the oxide thickness differences can be observed clearly in the Si sub/oxide interface among the three samples.

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

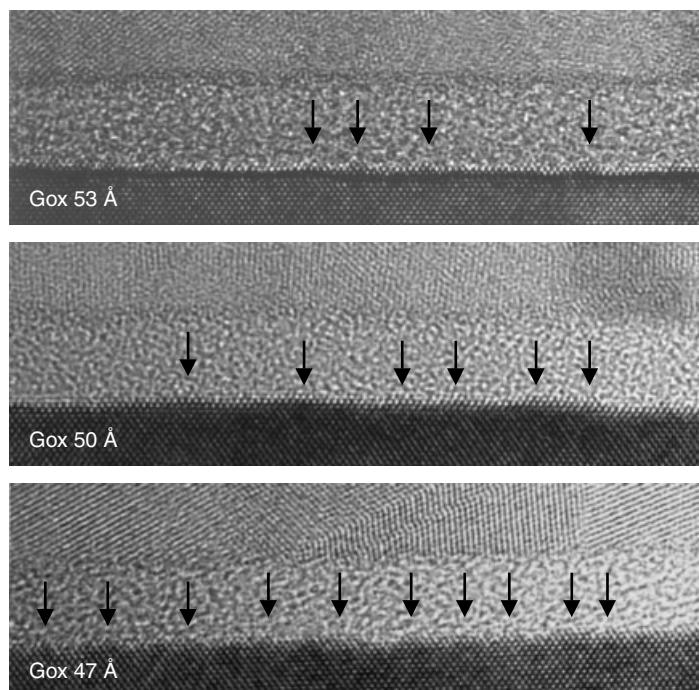


Figure 6.3 The Si sub/oxide interface's smoothness is an indicator of the oxide's quality. HRTEM captures directly any interface atomic ledges and terraces, as indicated by the arrows.

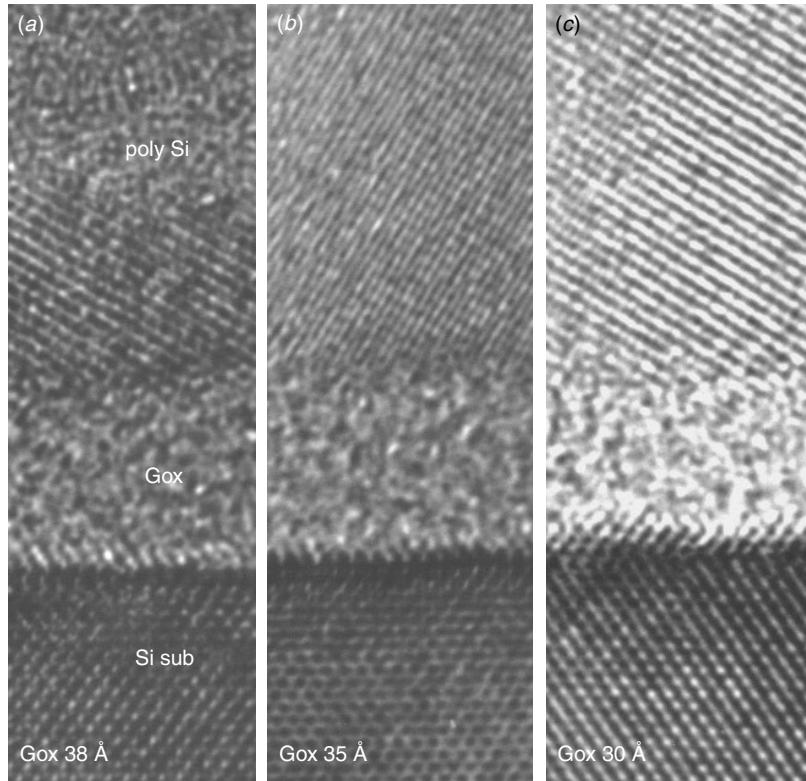


Figure 6.4 Three oxide layers with minor thickness differences. The polysilicon/oxide interface is laterally aligned so that the oxide thickness differences can be observed clearly in the Si sub/oxide interface among the three samples.

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

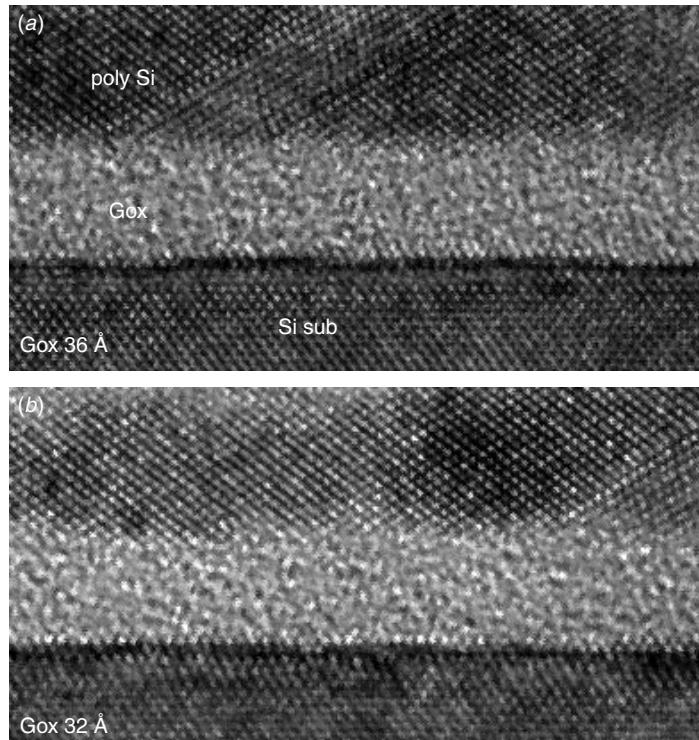


Figure 6.5 Two different thicknesses of gate oxides. Both micrographs show distinctive polysilicon grain cross-lattice fringes, so the gate oxide's thickness measurements can be done more accurately.

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

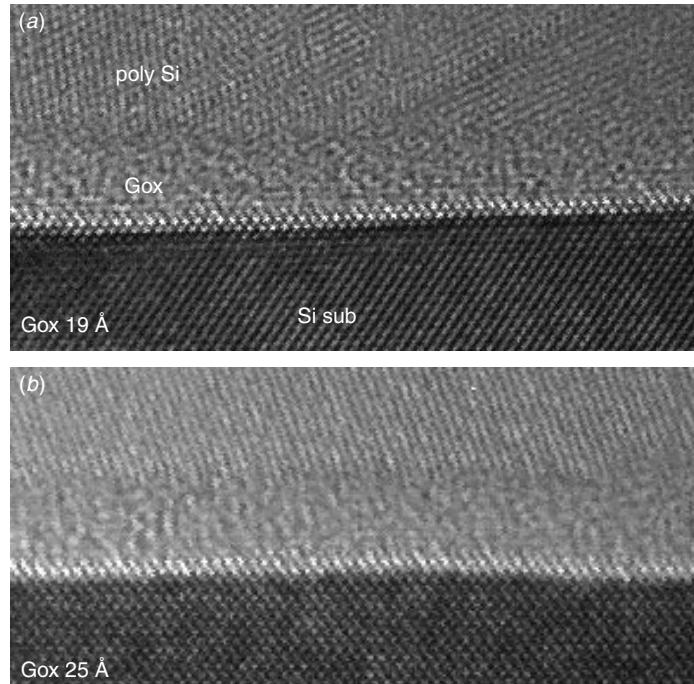


Figure 6.6 Two different samples of gate oxide thicknesses treated with nitridation, (a) 19 Å gate oxide and (b) 25 Å gate oxide. The Si substrate surface roughness appears to have a dominant effect on the local variation in the gate oxide's thickness.

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

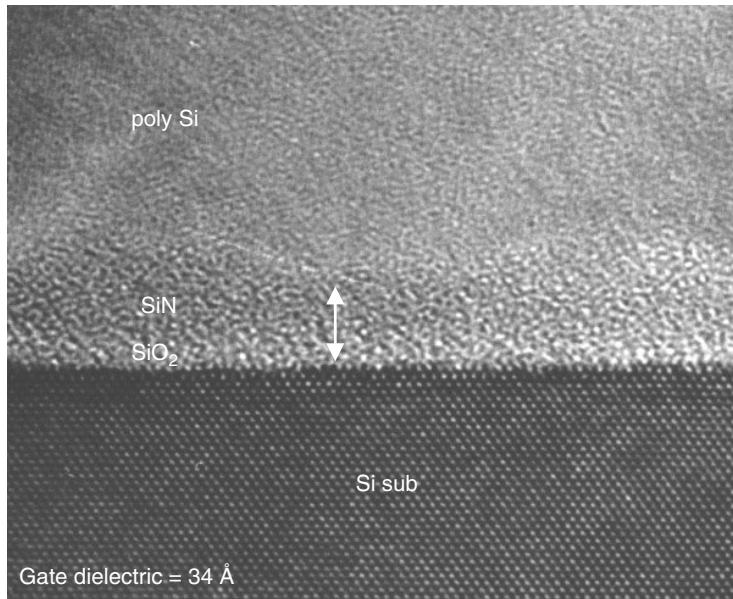


Figure 6.7 Silicon nitride/silicon dioxide double-layer gate material, a potential alternative to the single-layer oxide material, with or without nitridation annealing. The total thickness of SiN and SiO_2 is 34 Å.

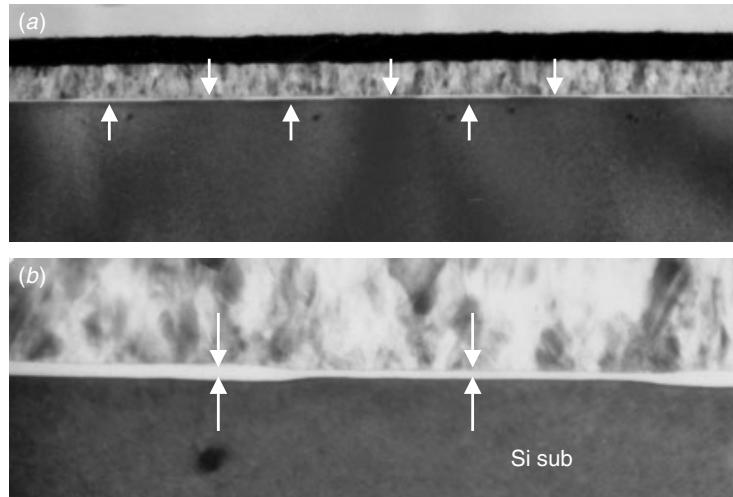


Figure 6.8 Dual gate oxide thickness as seen here is another strategy to improve reliability when the ultra thin gate dielectric is needed only in certain devices. Alternating gate oxide thicknesses are used next to each other to meet different device applications on the same chip and maintain overall 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.

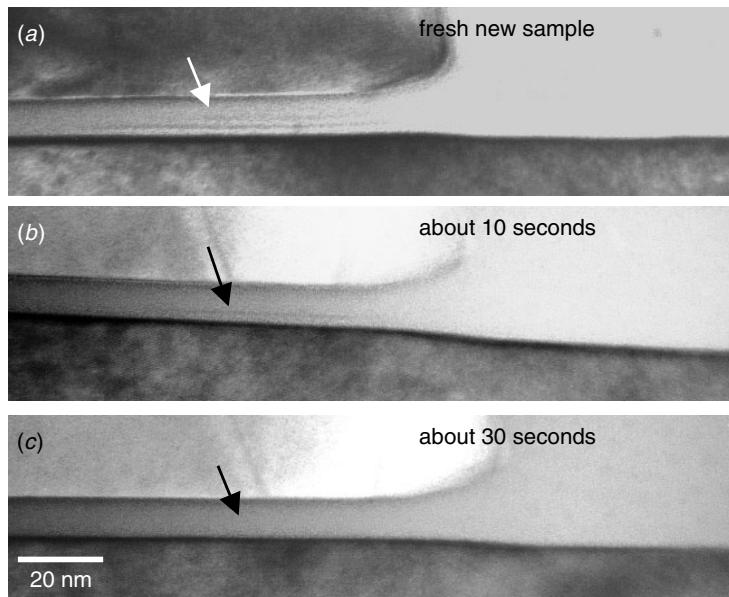


Figure 6.9 TEM cross section of a gate oxide breakdown point after the burn-in test. At least three parallel lines are observed near the high field's end corner of the device, as indicated. The defect is sensitive to the TEM electron beam observation and gradually disappears as the observation time increases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

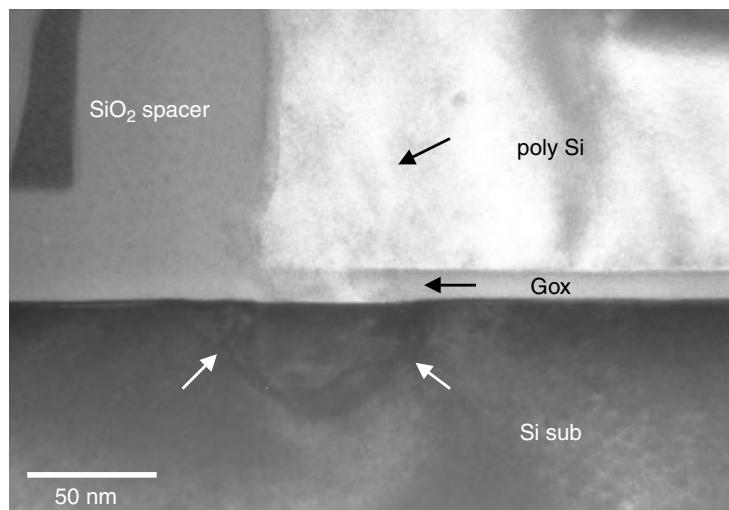


Figure 6.10 TEM cross section of a gate oxide's hard breakdown point after the burn-in test. The observable damages extend from the polysilicon gate's corner, through the gate oxide, and down into the Si substrate and the created groove in the Si substrate, as indicated. (Sample courtesy of Dr. Du An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

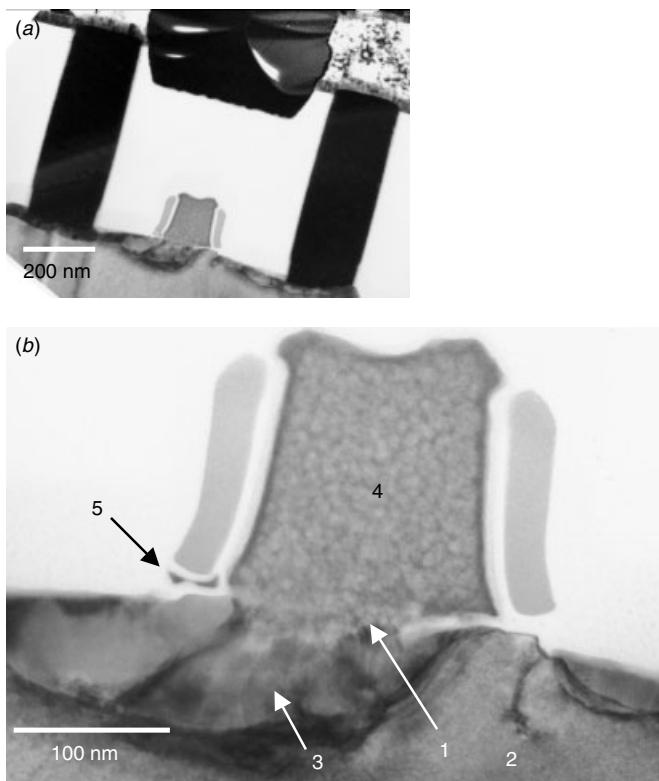


Figure 6.11 TEM cross section of a gate oxide breakdown point. At least five major failure mechanism are observed. (1) Ruptured Gox, (2) Si substrate damages and dislocations, (3) Migration and re-crystallization of silicide, (4) epitaxy on polycide gate, and (5) nitride spacer phase transformation. (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

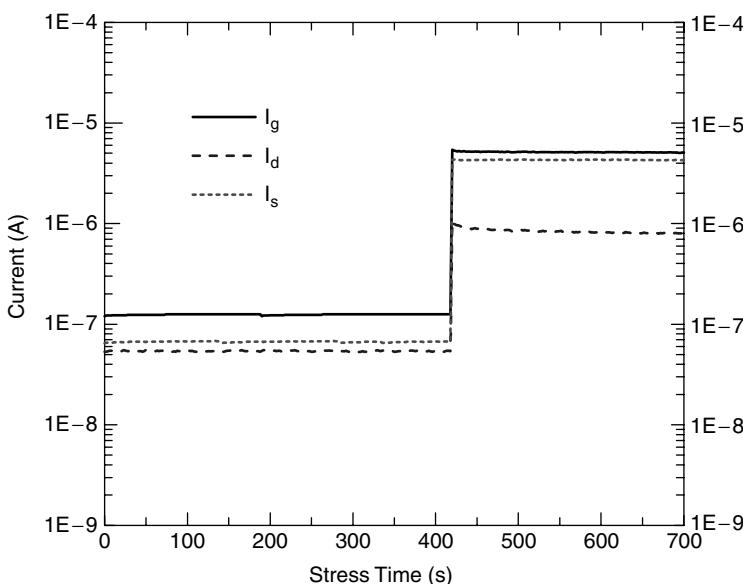


Figure 6.12 A current step in the SBD—for gate, source, and drain currents. The SBD has occurred near the source side, evidenced by the concurrent jump in the gate current and source current while that on the drain current is much less. (Reprint from *International Electron Devices Meeting Tech. Dig.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

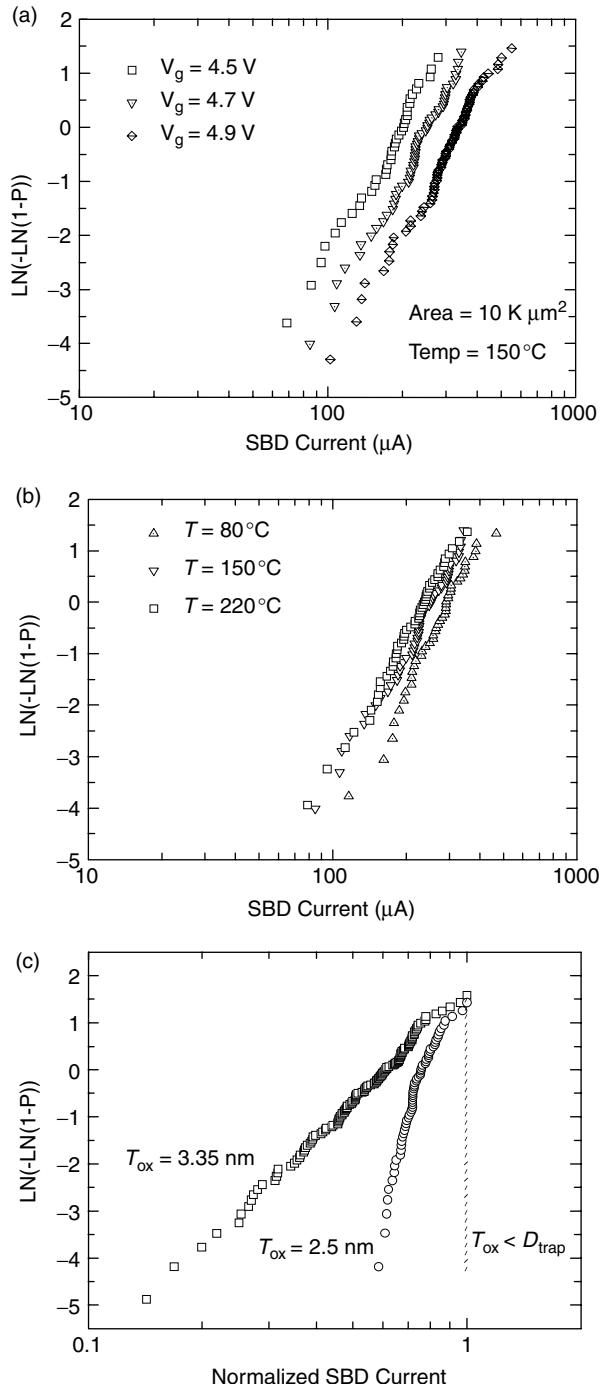


Figure 6.13 Failure distribution of the SBD current (a) for different stress voltages, (b) at different temperatures, and (c) for the oxide thickness of 33 \AA and 25 \AA . (The plot of $T_{\text{ox}} < D_{\text{trap}}$ included for reference.) (Reprint from *Microelectron. Reliabil.*, **42**, 565–571,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

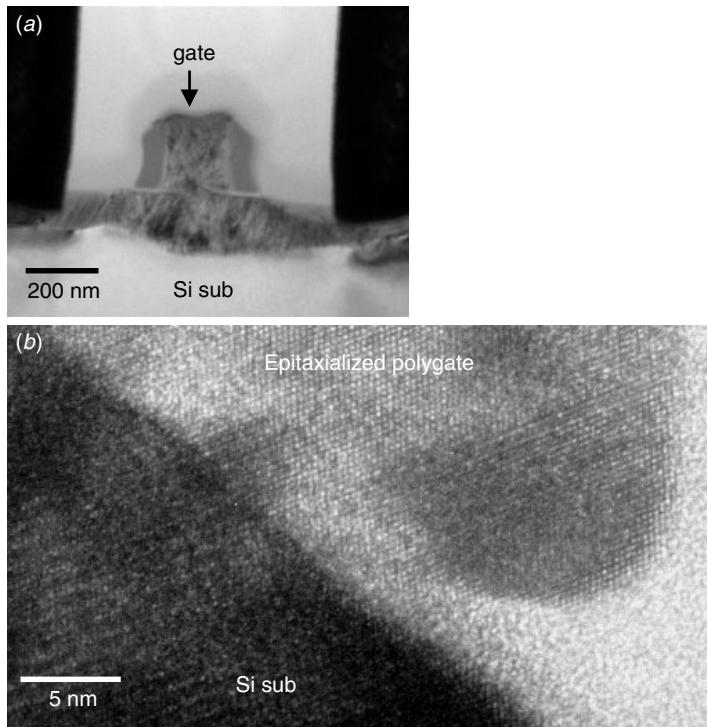


Figure 6.14 TEM micrograph of a breakdown of a Ti-silicide nMOSFET structure with 33 Å gate oxide. The stress voltage and current compliance was 5.1 V and 100 mA, respectively. (a) Total structure damage to the gate oxide, poly-Si gate, and substrate. (b) High resolution TEM showing the recrystallization 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

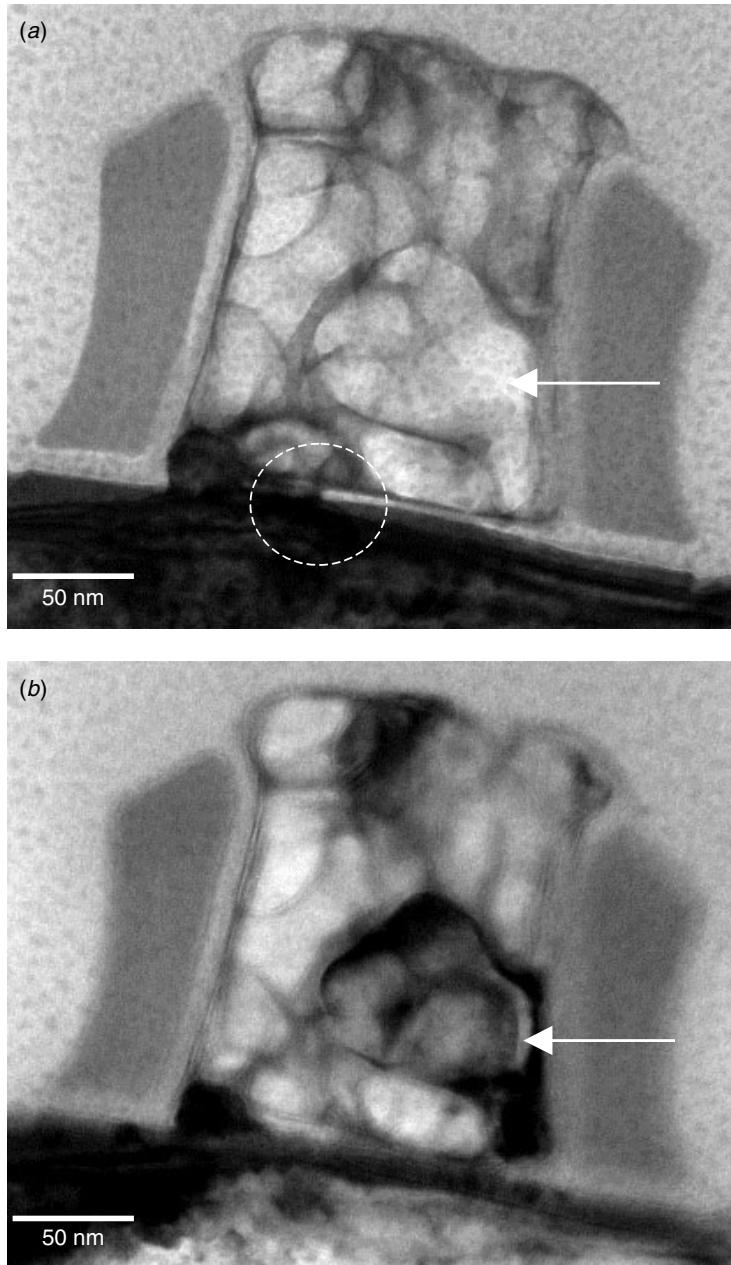


Figure 6.15 TEM micrograph of a Ti-silicide nMOSFET device stressed at a current compliance of 1 mA. Substantial damage to the polygate, including the polycide and the Si-substrate's channel region, and gate oxide can be observed easily. (a) At the Si sub (110) pole illumination; (b) at the tilted angle 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)

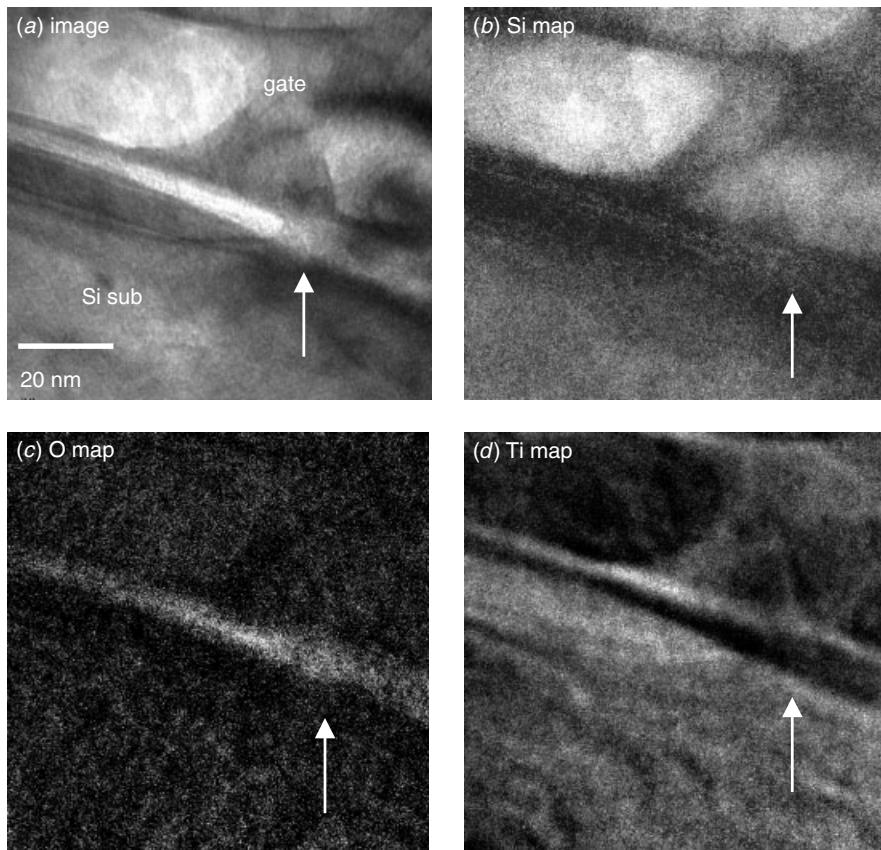


Figure 6.16 TEM micrograph of a Ti-silicide nMOSFET device stressed at a current compliance set at 1 mA. (a) Gate oxide's rupture, as indicated, (b) the Si EELS map, (c) the oxygen EELS map, and (d) the Ti EELS map. Ti is re-distributed within the polygate as well as migrates into the channel region. 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.

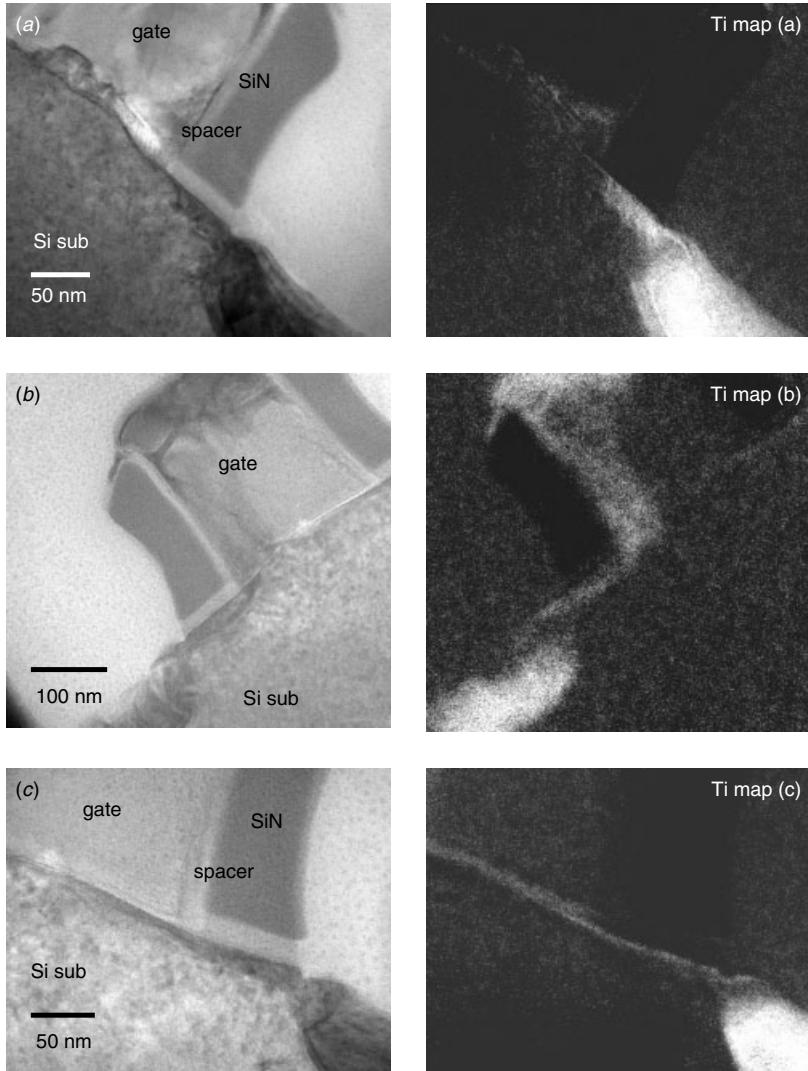


Figure 6.17 TEM micrograph of the Ti-silicide nMOSFET device stresses at current compliances set at (a) $50 \mu\text{A}$, (b) $10 \mu\text{A}$, and (c) $10 \mu\text{A}$. The right-hand side images are the corresponding EELS mappings of the Ti elements. Ti migration in these samples is obvious. (Reprint 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

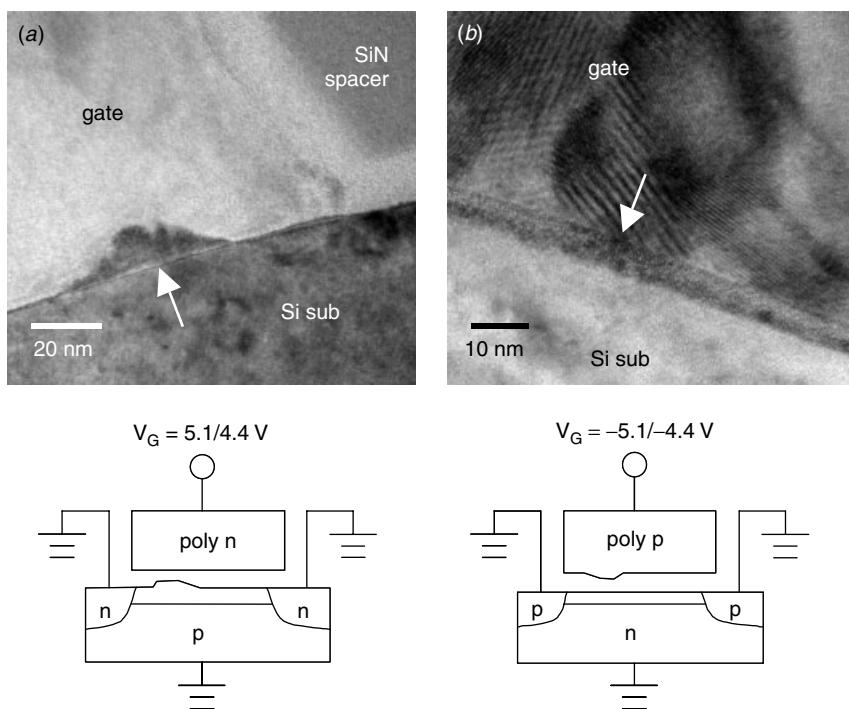


Figure 6.18 Low-magnification cross sectional TEM micrographs of a poly-Si gate structure showing damages at various compliance current levels. (a) 10 μA of a Co-silicided 25 \AA gate oxide nMOSFET and (b) 1 μA of a Ti-silicide 33 \AA gate oxide pMOSFET. The Si epitaxial extrusion is indicated by the 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

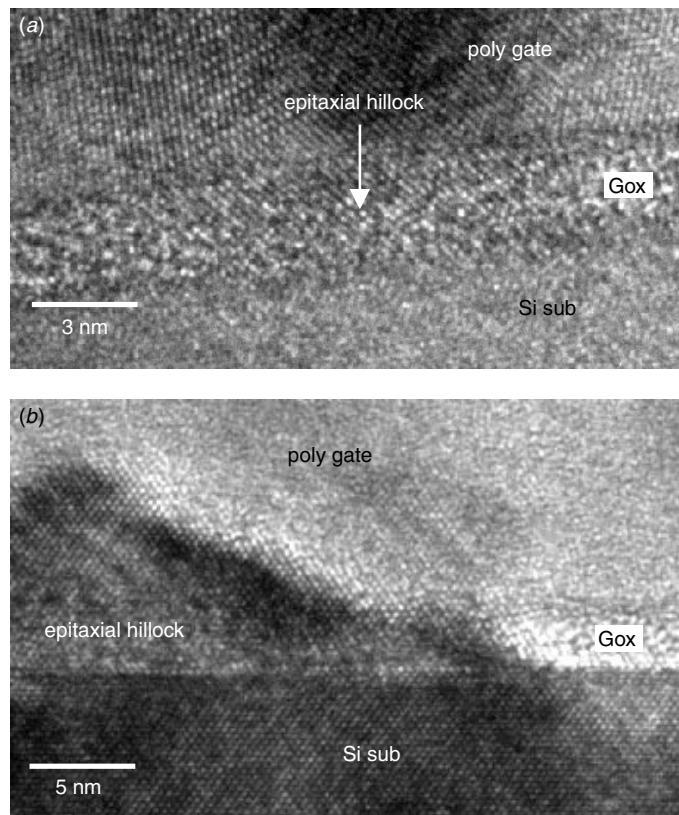


Figure 6.19 HRTEM lattice image showing a Si epitaxial growth at the breakdown point. The transistors were stressed with constant voltage and a current compliance was set at 1 μ A. An electrical test after the SBD confirmed that the device was still functioning but had a slightly higher gate oxide leakage. 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

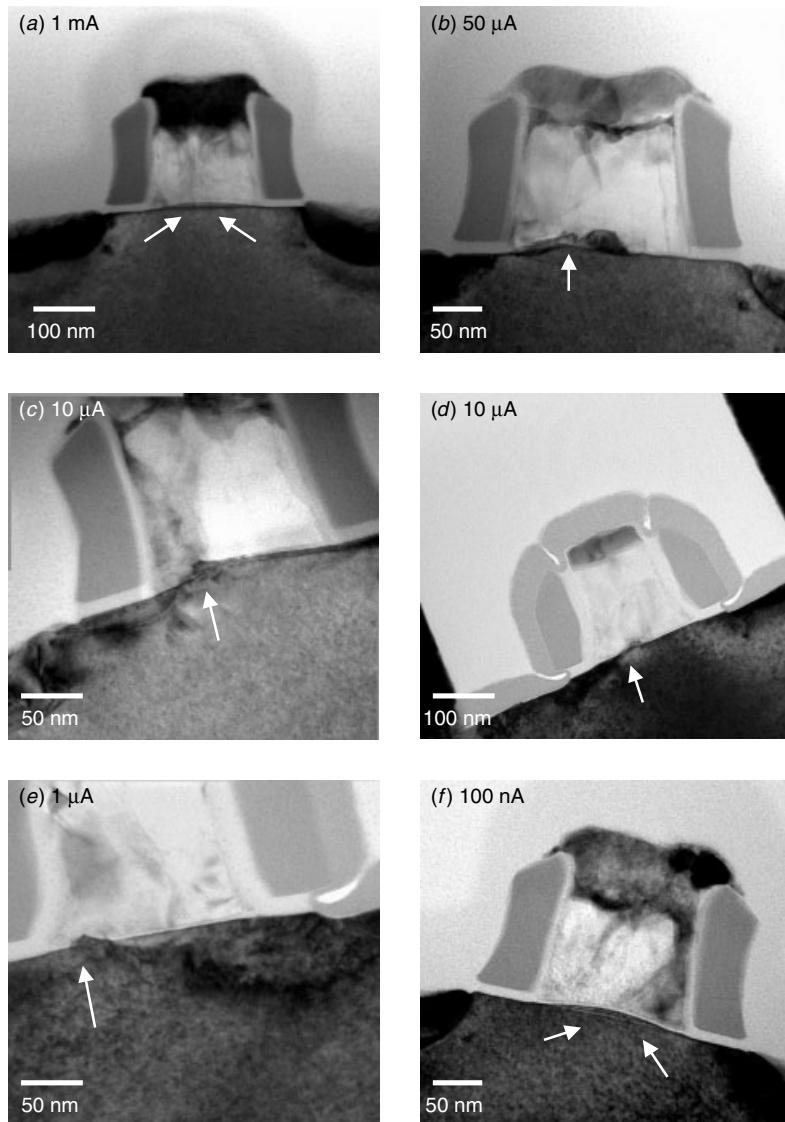


Figure 6.20 Low-magnification cross section TEM views of the gate structure damages at various compliant current levels. (a) 1 mA, Ti salicide, 33 Å Gox; (b) 50 μ A, Ti salicide, 33 Å Gox; (c) 10 μ A, Ti salicide, 33 Å Gox; (d) 10 μ A, Co salicide, 25 Å Gox; (e) 1 μ A, Co salicide, 25 Å Gox; and (f) 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

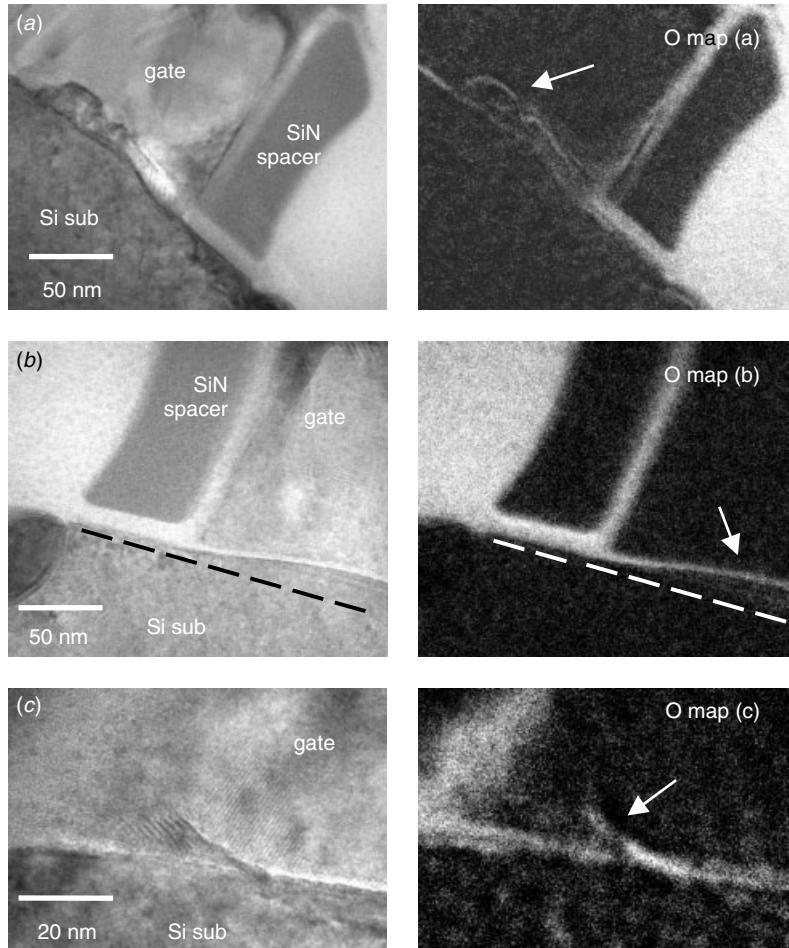


Figure 6.21 TEM images (*left*) and associated EELS oxygen mapping (*right*) showing the gate oxide pushed up along with the epitaxy islands, as indicated. (a) $50 \mu\text{A}$ HBD, TiSi process; (b) 100nA , SBD, TiSi process, the dashed line marks the original Si substrate position; and (c) $1 \mu\text{A}$,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

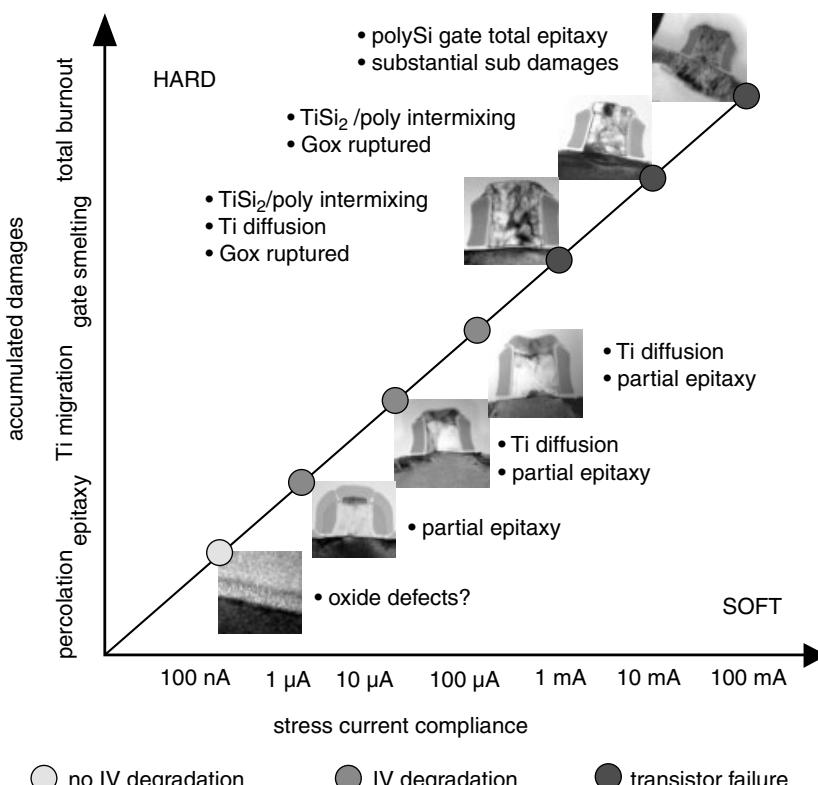


Figure 6.22 A summary of the physical damages generated by a gate oxide breakdown and the corresponding compliant current levels under a constant voltage stress (CVS) condition. (Pey et al., IEDM Tech. Dig. 163–166, 2002, reprint with permission from IEEE)

- 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.

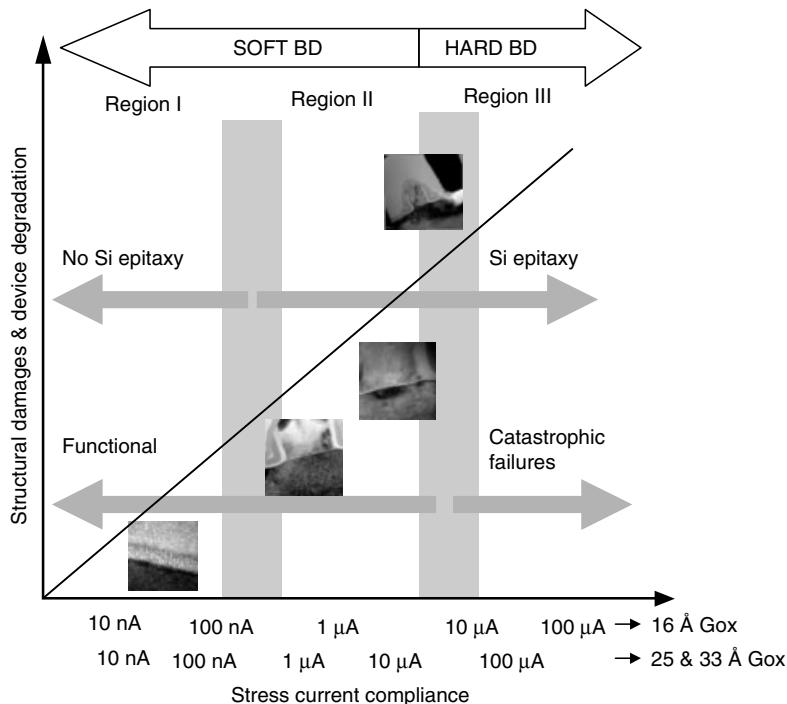


Figure 6.23 Effect of compliance current setting on the hardness of the oxide's breakdown and formation of DBIE in nMOSFETs. Regions I and II belong to SBD and Region III belongs to HBD. The micrographs shown are for 16 Å Gox transistors. Similar responses have been observed for the 25 and 33 Å Gox transistors 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).

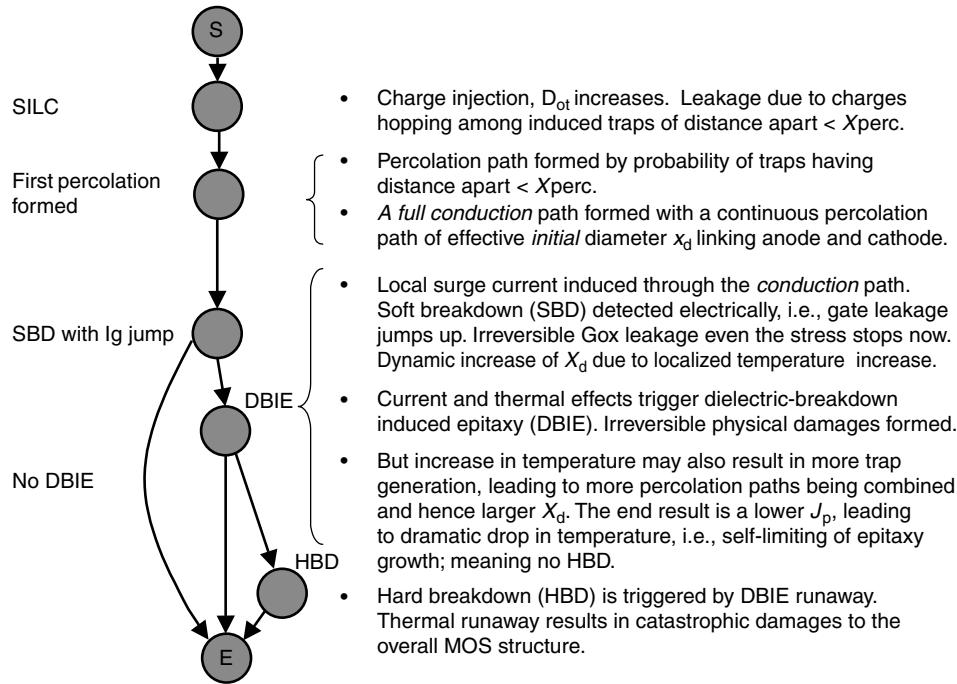


Figure 6.24 Sequential events of gate oxide breakdown under constant voltage stress. (Pey et al. *IEDM Tech. Dig.* 163–166, 2002, reprint with permission from IEEE)

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.

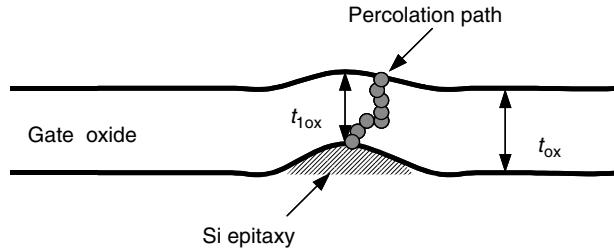


Figure 6.25 Schematics showing the possible effects of oxide deformation caused by the growth of DBIE on Gox failures. Localized oxide “thinning” occurs when the Gox thickness at DIBE point $t_{1\text{ox}} < t_{\text{ox}}$, which favors more percolation path formation in its vicinity. A proposed progression 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)

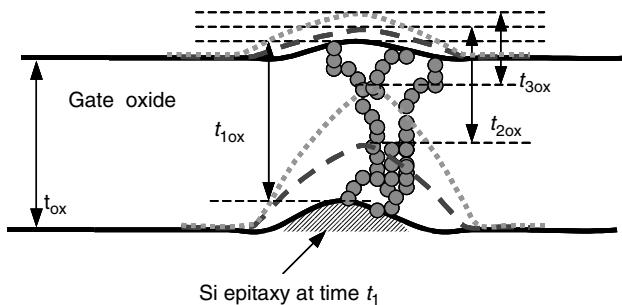


Figure 6.26 Schematics showing in more detail the effects of oxide deformation caused by the growth of DBIE on Gox failures. Localized oxide “thinning” occurs when the Gox thickness at DIBE point $t_{1\text{ox}} < t_{\text{ox}}$, which favors more percolation path formation in its vicinity. A proposed 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 Gox thinning rate. As a result a finite DBIE hillock will be formed, resulting in SBD DIBE without Gox 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 Gox are expected to be greatly altered compared to those of Gox elsewhere.

The possible sequential events of Gox 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 Gox 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 Gox 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 Gox . This means that the traps generated in Gox 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 Gox . However, because of the limited thickness of Gox 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 Gox , easily resulting in HBD. This explains narrower region II in Fig. 6.23 with decreasing Gox thickness. In summary, the threshold DBIE thickness causing HBD in an ultra thin Gox will be considerably smaller than in thicker Gox .

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 Gox thickness. This means that for the ultra thin Gox 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 Gox 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 Gox 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 a substrate active areas, polysilicon gate and runners, and polycide layers. Intermetal

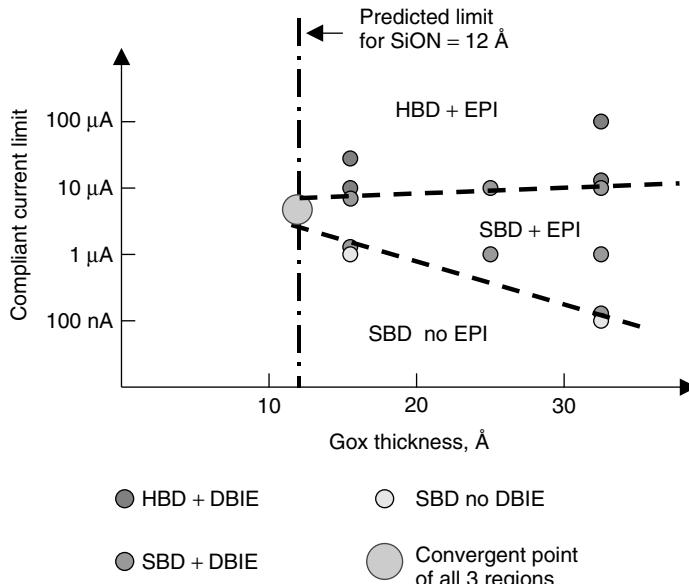


Figure 6.27 Convergence of HBD, SBD, and DIBE is expected to be below the predicted limit of the oxide's base gate dielectric at 12 Å. The circles represent the experimental data points from the current work: SBD without DBIE, SBD with DBIE, and HBD with DBIE. (Pey et al. *IEDM Tech. Dig.* 163–166, 2002,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

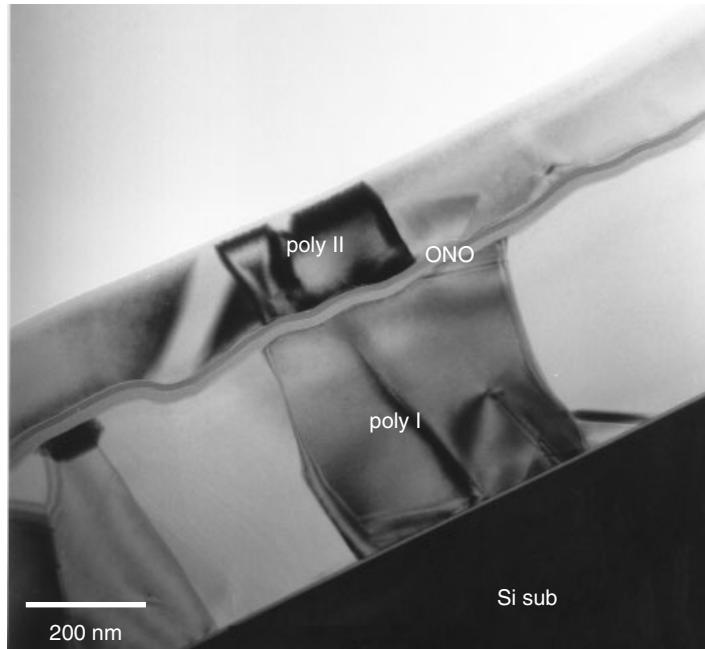


Figure 6.28 TEM cross section of a poly/ONO/poly stacked capacitor structure. The interface's roughness is a major challenge for TEM to measure the exact thickness of each oxide and nitride layer.

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

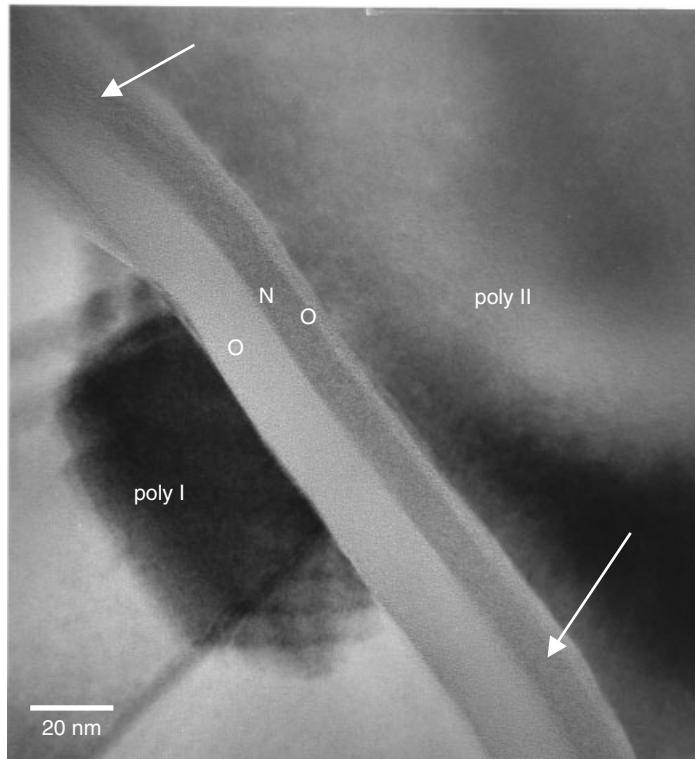


Figure 6.29 TEM cross section of a poly/ONO/poly stacked capacitor structure. The center area of the image shows clearly the ONO contrast, and this is convenient for thickness measurements. The arrows indicate the areas; however, the ONO contrasting smear is due to local polysurface roughness induced 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.

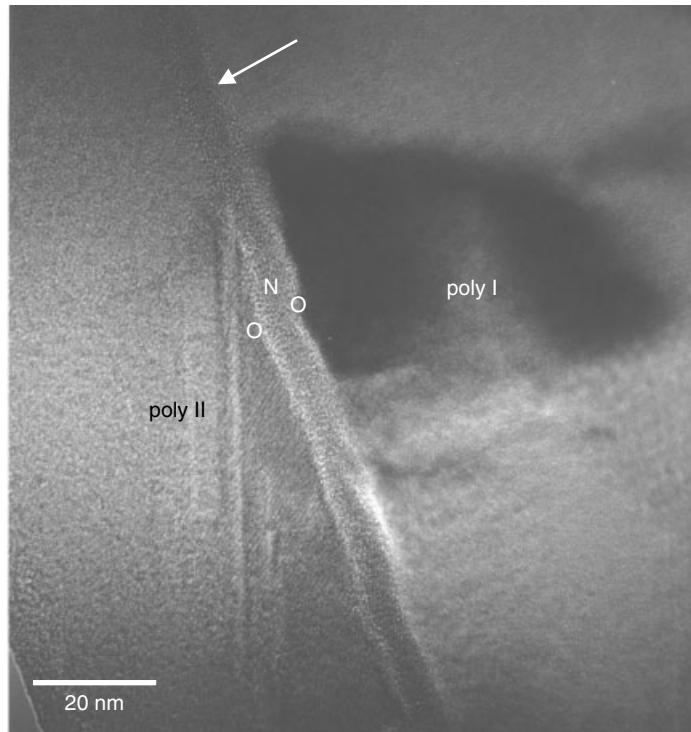


Figure 6.30 TEM cross section of a poly/ONO/poly stacked capacitor structure. The center area of the image shows clearly the ONO contrast. The arrows indicate areas where there is virtually no contrast.

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

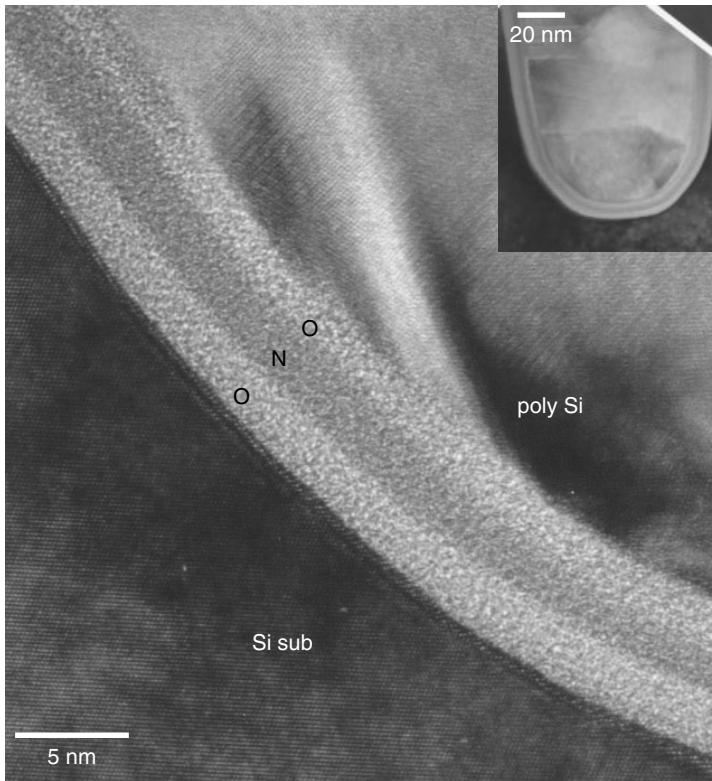


Figure 6.31 TEM cross section of a poly/ONO/Si–sub trench capacitor structure. The inserted image shows the faceted curvature of the trench’s bottom. ONO layers appear clearly. The layer’s thickness can be measured along the bottom curve.

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

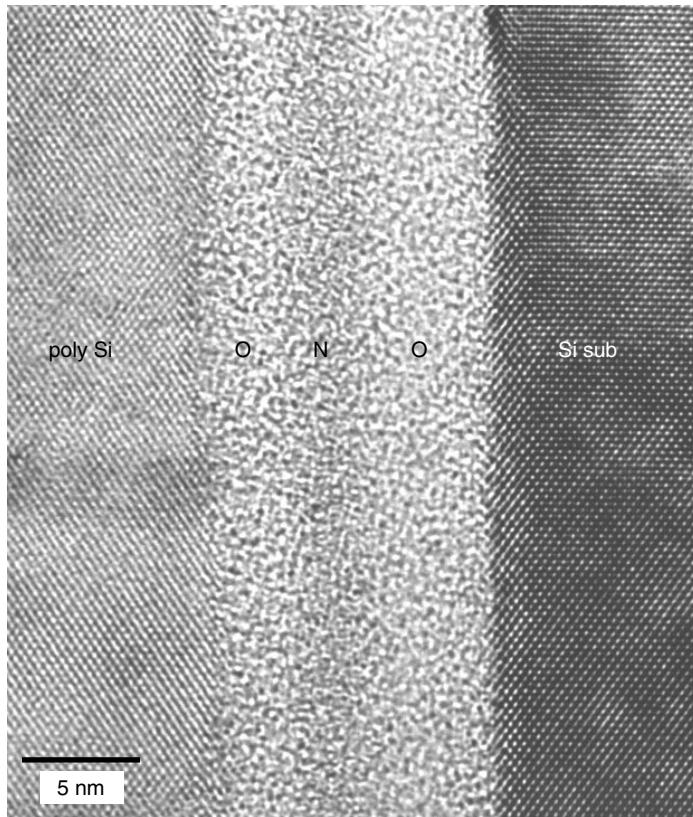


Figure 6.32 HRTEM cross section of a poly/ONO/Si–sub trench capacitor ONO structure. The oxide/nitride contrast is poor due to the multiple beam contrast mode used for high resolution, so the exact location of the oxide/nitride interface is difficult to determine.

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).

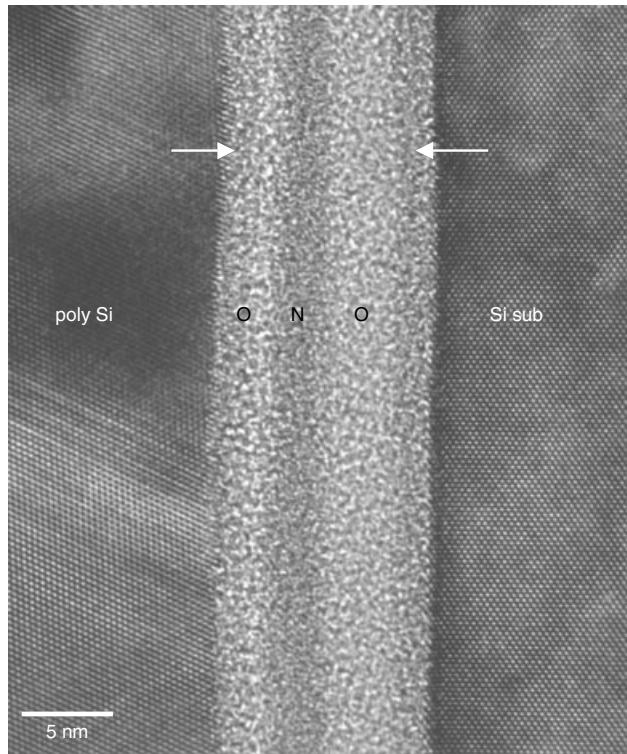


Figure 6.33 HRTEM cross section of a poly/ONO/Si–sub trench capacitor ONO structure. A special nitridation treatment of the oxide layer has created the distinctive dark contrast of oxide layers at both the Si sub and polysilicon, as indicated.

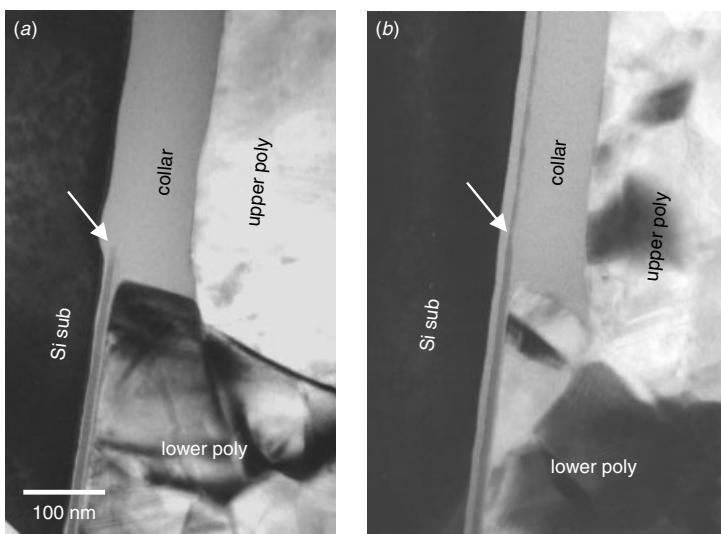


Figure 6.34 TEM cross section of poly/ONO/Si–sub trench capacitor ONO structures. The collar oxide (used to reduce parasitic transistor leakage) with the nitride removed (a) and the nitride retained (b).

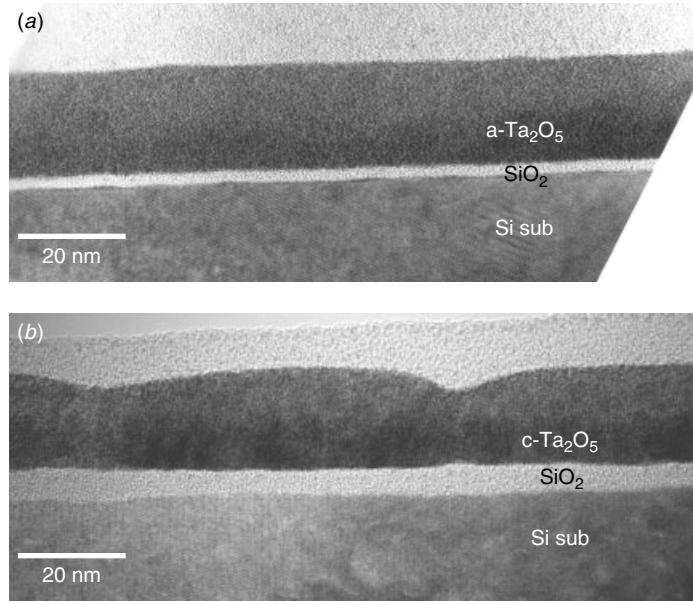


Figure 6.35 HRTEM of Ta₂O₅ process. Low-temperature annealing produces amorphous Ta₂O₅ with uniform thickness while high-temperature annealing shows crystalline Ta₂O₅ with local thinning at the grain boundaries.

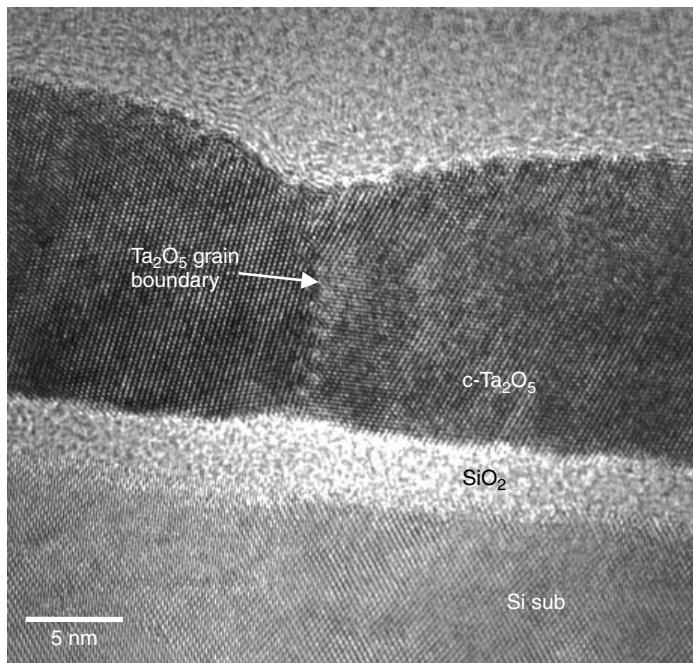


Figure 6.36 HRTEM of Ta₂O₅ process. High-temperature annealing sample clearly shows crystalline Ta₂O₅ with lattice fringes and the grain boundary. The weak spot oxidation is linked to the Ta₂O₅ grain boundary's diffusion, as shown in micrograph.

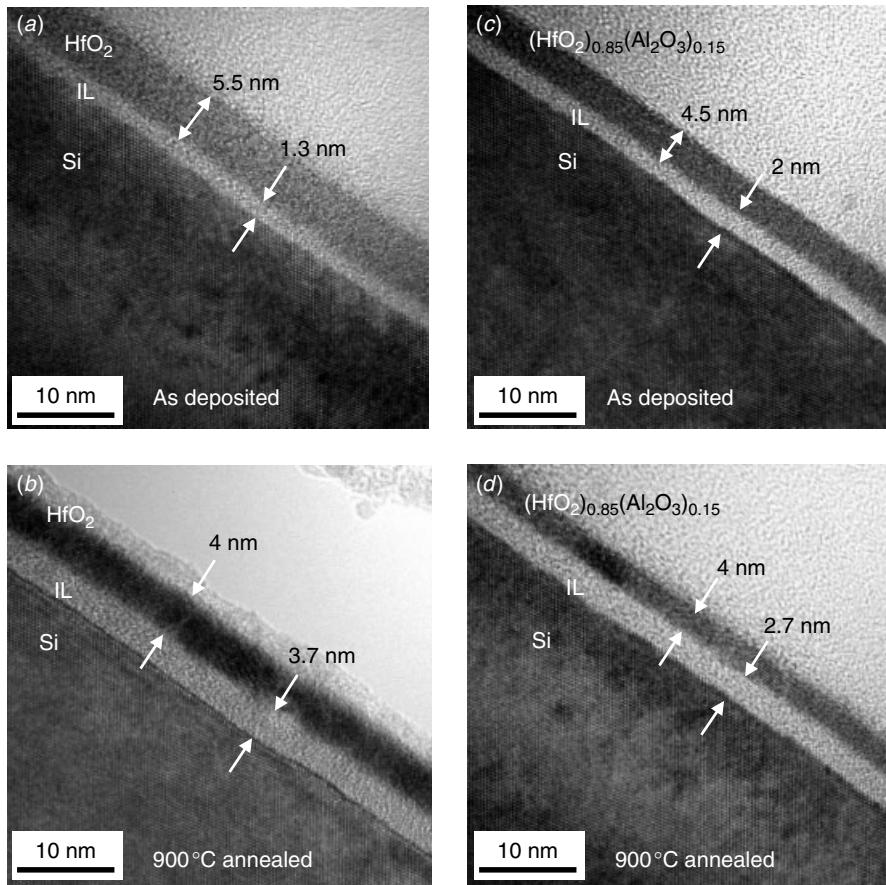


Figure 6.37 HRTEM images of (a) the as-deposited HfO_2 sample, (b) the $900^\circ\text{C}/\text{N}_2$ annealed HfO_2 sample, (c) the as deposited $(\text{HfO}_2)_{0.85}(\text{Al}_2\text{O}_3)_{0.15}$ sample, and (d) the $900^\circ\text{C}/\text{N}_2$ annealed $(\text{HfO}_2)_{0.85}(\text{Al}_2\text{O}_3)_{0.15}$ 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

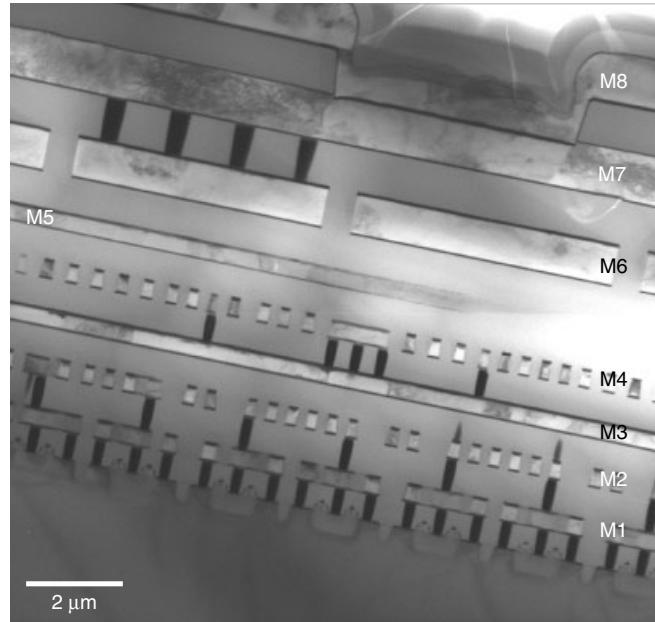


Figure 6.38 A modern ULSI device with eight metallization layers. Interlayer dielectric (ILD) and intermetal dielectrics (IMD) are a crucial part of the device's structure.

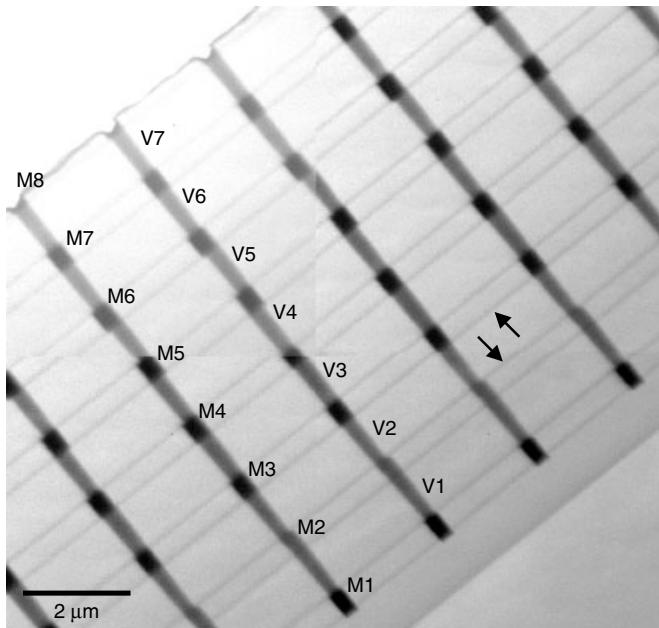


Figure 6.39 An eight-metal layer device using the Cu dual-damascene process. Metals (M) and VIAs (V) are marked accordingly. Barrier dielectrics (dark parallel lines as indicated by arrows) at the top and bottom of each low- k dielectric IMD are required. The process is more complicated than traditional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

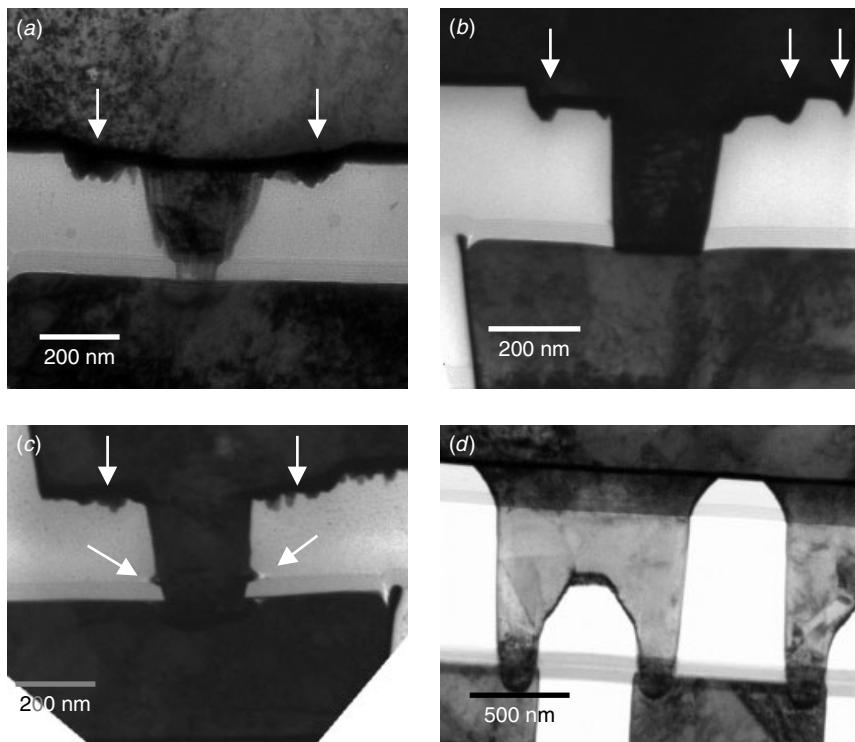


Figure 6.40 Three different low- k materials coupled with Cu dual-damascene process. (a,b) Carbon-doped SiO_2 material with severe etching pits (arrows) next to VIA. (c) A different carbon-doped SiO_2 low- k material showing pits not only on the film but also at the barrier dielectric interface (arrows).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

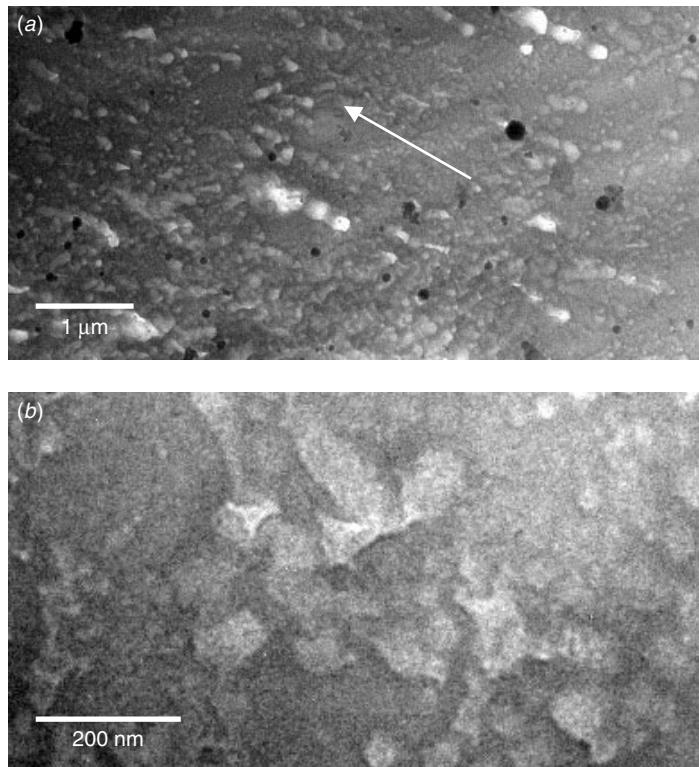


Figure 6.41 TEM of porous polyimide film. (a) The ubiquitous micro-pores are aligned toward the upper left, as indicated by the arrow. (b) Higher TEM magnification of the porous polyimide film shows them to be interconnected and about 50 to 100 nm in diameter.

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).

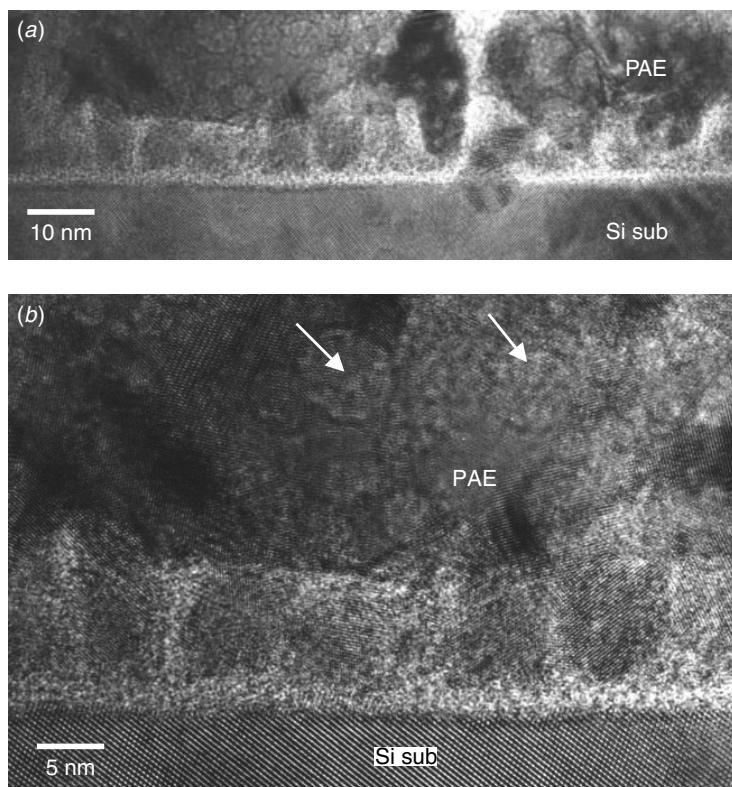


Figure 6.42 (a) Low magnification of polyacrylethers (PAE). Densely packed nanopores are observed within the film. (b) The same density of nanopores and nanocrystallines are observed within the film. The exact nature of the film remains to be studied. (*Appl. Phys. Lett.*, **75** (6), 1999 with permission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

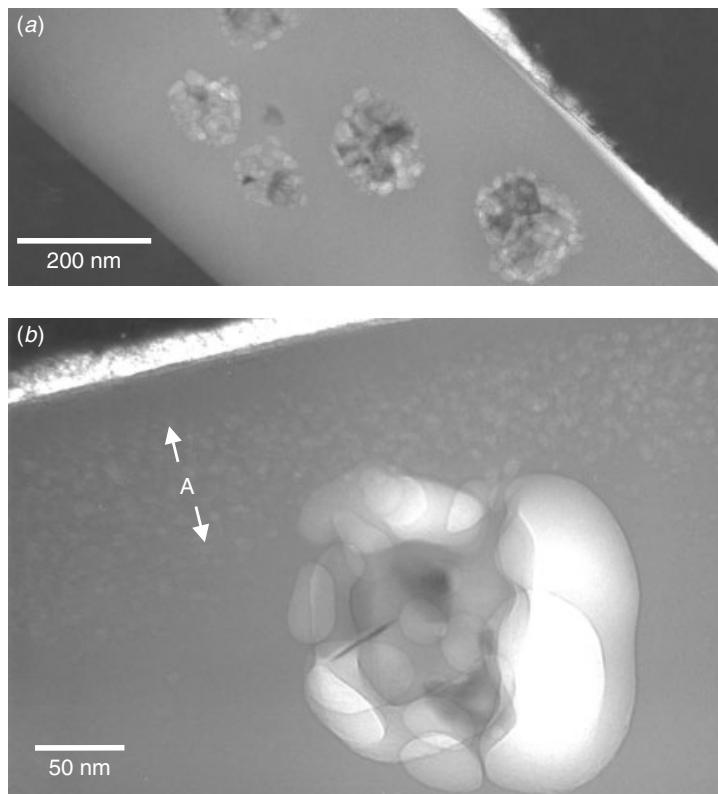


Figure 6.43 Cross section TEM of a microcrystalline particle residue showing micro-pore layers above the crystalline residues, as indicated by A, and large bubbles surrounding the crystalline particles. The crystalline particle, along with the large bubbles surrounding it, disappears after long TEM observation.

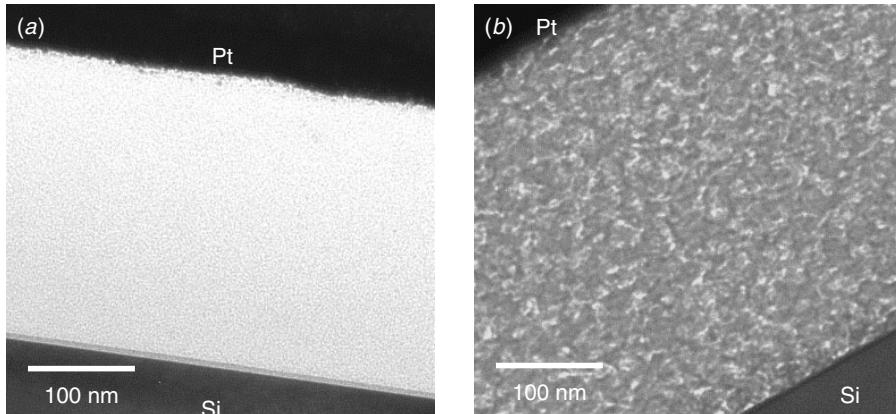


Figure 6.44 TEM cross sections of (a) low- k materials with $k = 2.8 \sim 3$ and (b) ultra low- k materials with $k = 2.4$. TEM/EELS, SIMS, and FTIR analyses showed the two materials to have identical basic chemical structures. TEM shows their microstructures to be very different. The nanoporous structure 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.

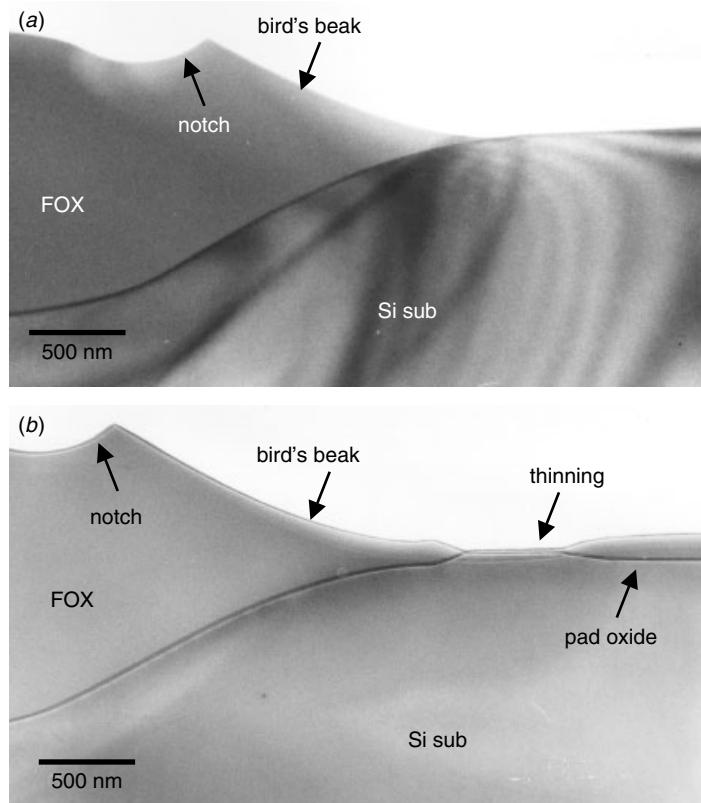


Figure 6.45 Cross section TEMs of the field oxide (FOX) shows a notch on field oxide surface, a thinning tail usually referred to as bird's beak. (a) Without local thinning and (b) with a local thinning on the pad oxide. (Courtesy Marcus and Sheng, 1983)

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.

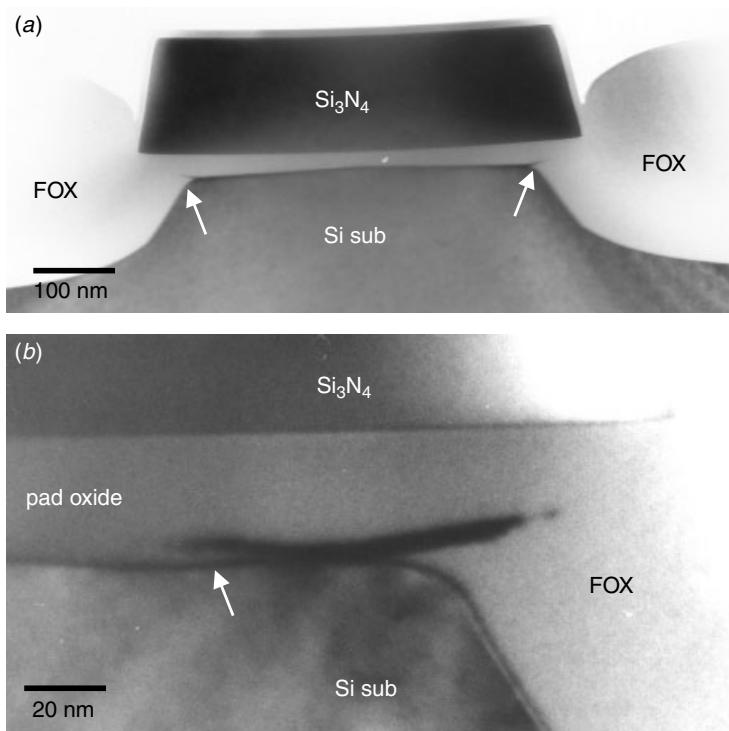


Figure 6.46 Cross section TEMs of the field oxide (FOX) bird's-beak area shows the micro masking "eyelash." (a) The two symmetrical micromasks on both sides of a Si_3N_4 mask give the eyelash form. (b) A detailed view shows the micro bird's beak on the opposite side of FOX, as indicated.

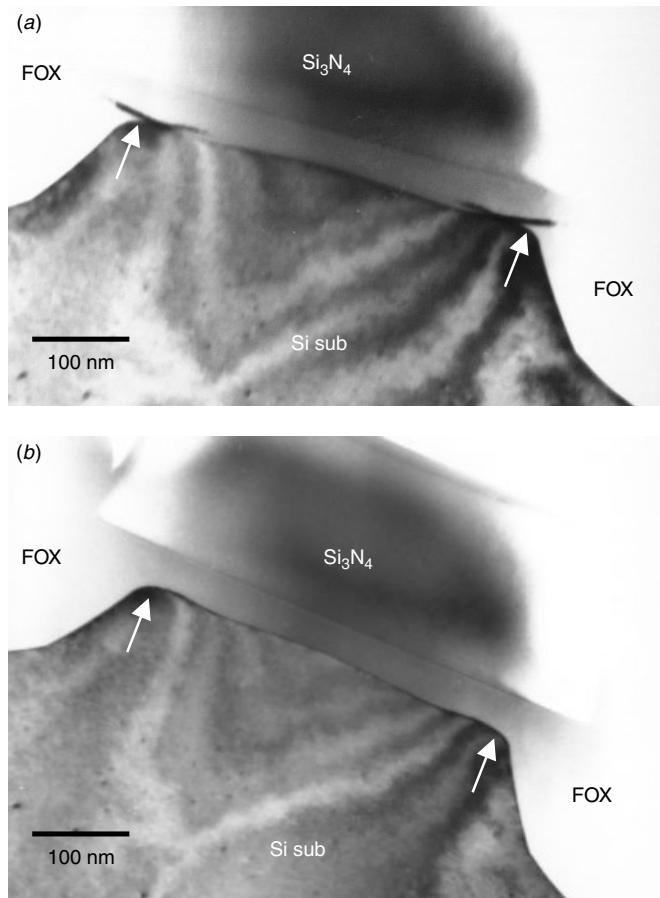


Figure 6.47 TEM cross sections of the field oxide (FOX) bird's beak area showing micro-masked eyelash. (a) The Si_3N_4 “eyelash” mask. (b) After 10 to 15 seconds of TEM observation, the eyelashes disappear.

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

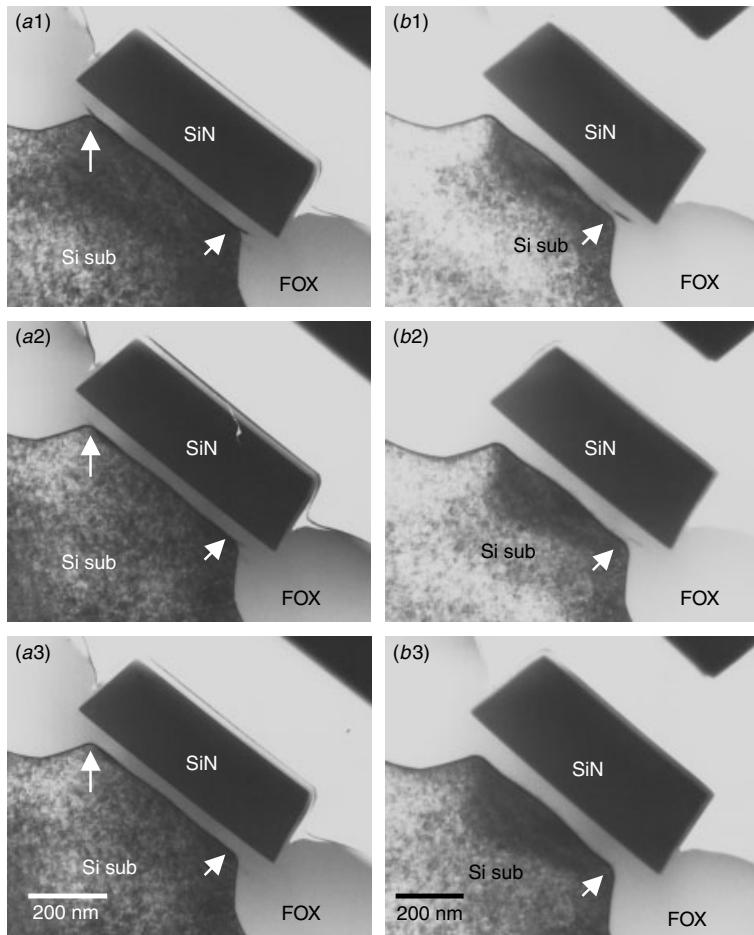


Figure 6.48 TEM cross sections of the field oxide (FOX) bird's beak area showing the micro-masked eyelash. (a) Two symmetrical micromasks on both sides of a Si_3N_4 mask with the characteristic “eyelash” form. (b) Asymmetrical micromasks on right-hand side. In both cases the micromasks diminished after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

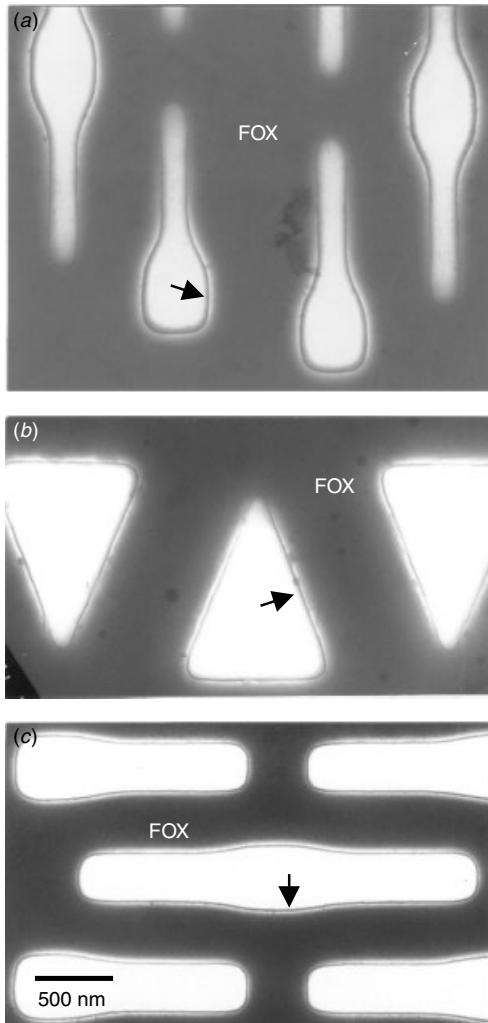


Figure 6.49 TEM plan view of various LOCOS islands. The dark areas are field oxide and the bright areas are gate oxide.

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

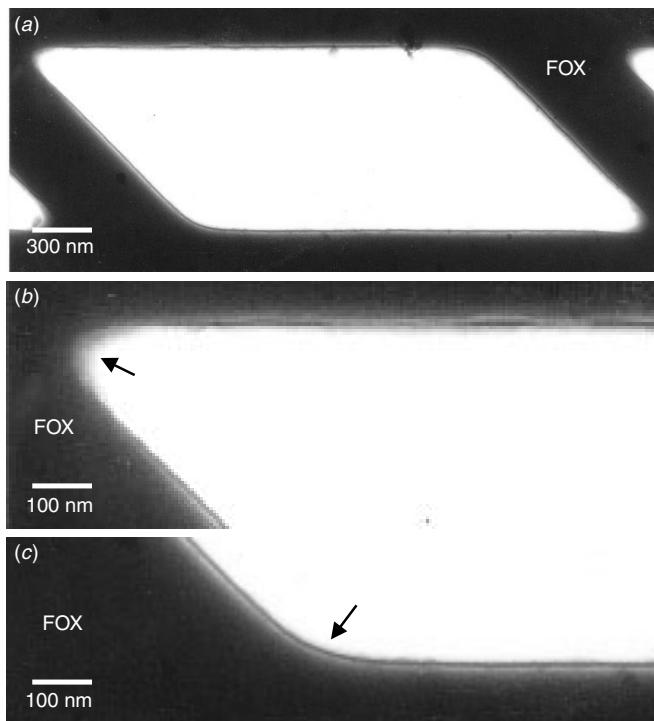


Figure 6.50 TEM plan view sample. (a) A black belt appears all around the edge of the LOCOS except at the acute inner angle; (b) Acute corner is enlarged to show the gap in the black belt, and (c) the black belt always appears in the obtuse inner angle corner. (Sheng et al. *J. App. Phys.* **75**(8),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

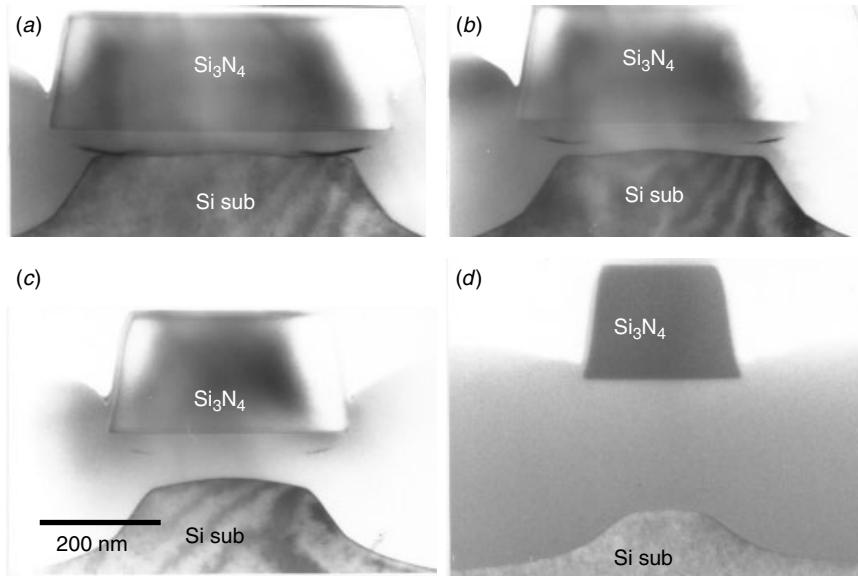


Figure 6.51 TEM cross section of inhomogeneous eyelashes and of the Si_3N_4 mask size dependence. (a) Si_3N_4 shows long and sticky (to the Si-substrate) eyelashes, and (b, c) floated and short eyelashes on both sides of the Si_3N_4 island. (d) The short Si_3N_4 mask shows no eyelash at all.

- 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.

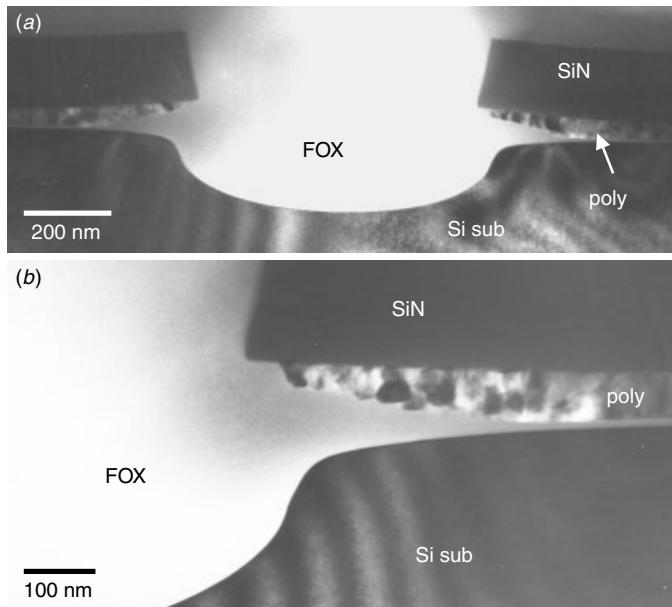


Figure 6.52 TEM cross section of a PBLOCOS isolation structure. Notice that the bird's beak is much shorter than that of the conventional LOCOS and that the substrate's sidewall is nearly vertical.

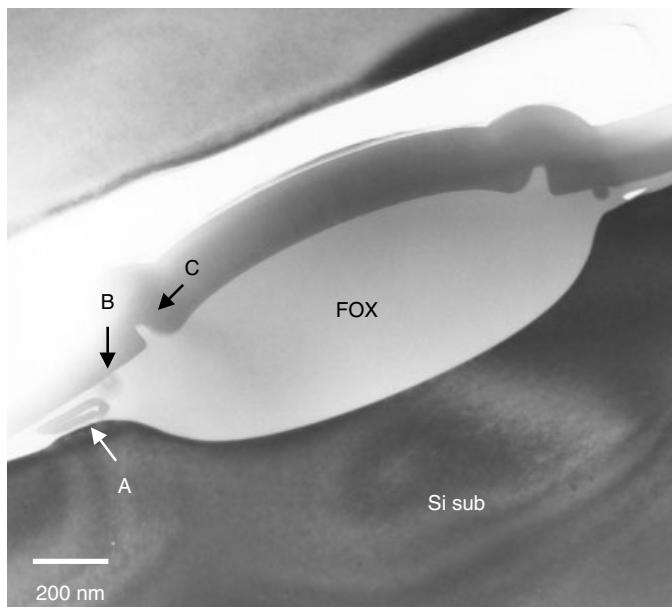


Figure 6.53 TEM cross section of the PBLOCOS structure with etching defects due to the poly layer's removal. The trapped voids surrounded by the nitride masking layers along with the pad-oxide local thinning (A), a pad-oxide ditch (B), and the field oxide protrusion on the bird's beak (C) altogether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

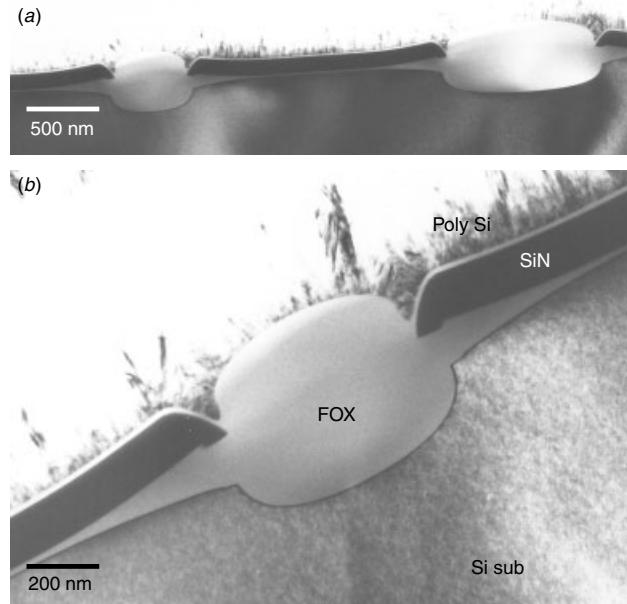


Figure 6.54 TEM cross section of the reverse L-shaped PBLOCOS structure. The nitride mask forms an upside-down-L-shape.

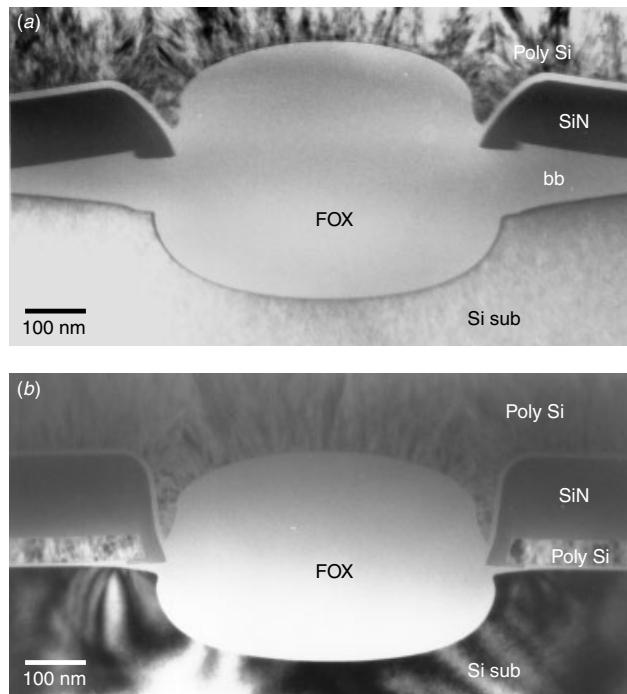


Figure 6.55 TEM cross section of the reverse L-shaped PBLOCOS structure. Further improvement is made by adding a polysilicon buffer layer, which can totally eliminate the bird's beak, as shown here.

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.

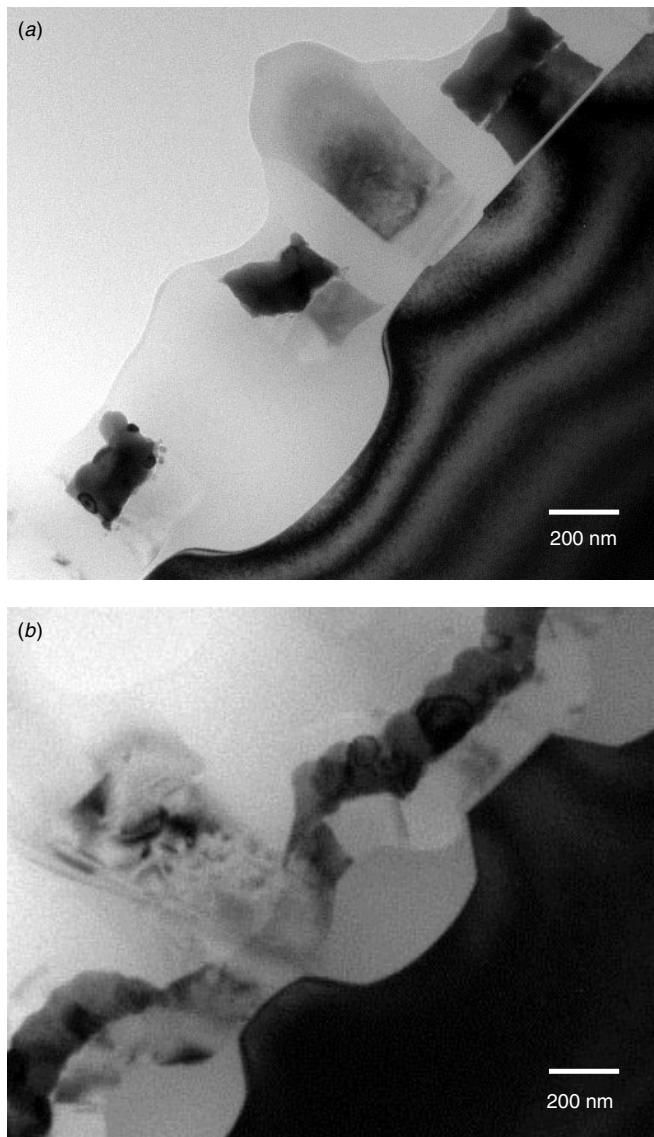


Figure 6.56 The $0.18 \mu\text{m}$ ULSI (DRAM process) technology with LOCOS isolation. This commercially available product shows that it is still justifiable to use LOCOS down to the $0.18 \mu\text{m}$ technology. The measured LOCOS length, including the bird's beak, is about 400 nm. The physical gate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

The scalability of a LOCOS isolation is questionable for $0.25\text{ }\mu\text{m}$ 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 $0.18\text{ }\mu\text{m}$ technology, as shown in Figs. 6.56 and 6.57. Figure 6.56 shows the physical gate length measured at $0.18\text{ }\mu\text{m}$. The measured LOCOS field oxide length, including the bird's beak, is $0.4\text{ }\mu\text{m}$ with $0.2\text{ }\mu\text{m}$ 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 $0.13\text{ }\mu\text{m}$. Notice also that the bird's beak shape is not the conventional curved slope but a straight slope on the Si substrate.

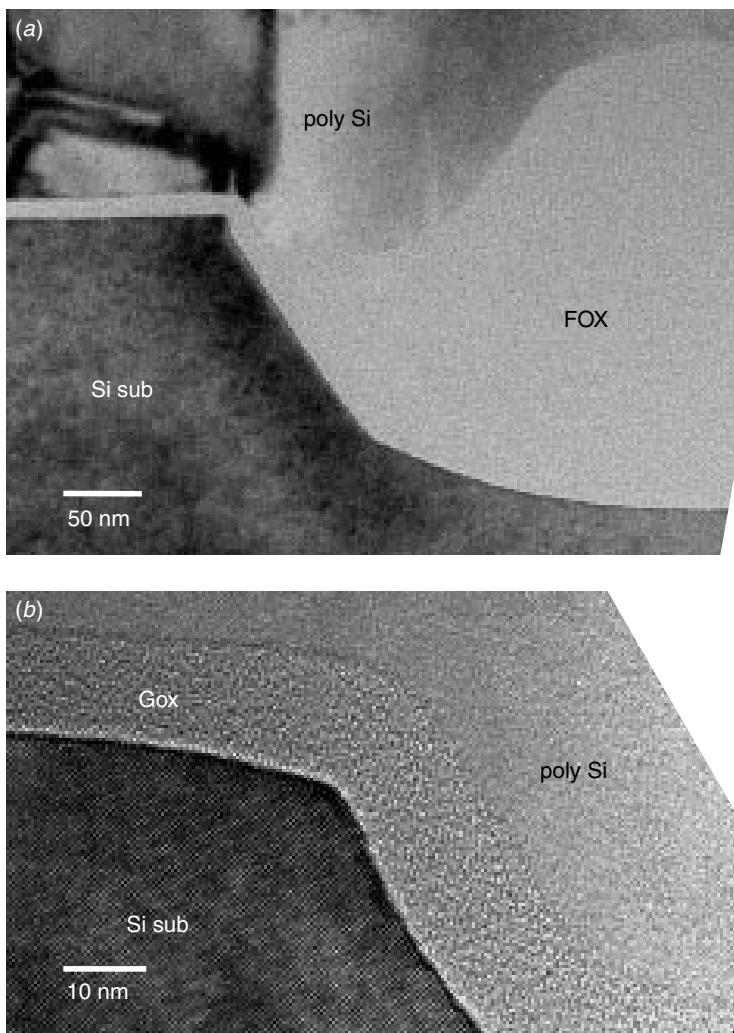


Figure 6.57 The $0.18\text{ }\mu\text{m}$ ULSI (DRAM process) technology with LOCOS isolation. (a) Close-up at the LOCOS corner shows detailed bird's beak shape and dimension; (b) no gate oxide thinning is observed at the bird's beak tip. The BB length measures about 125 nm .

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

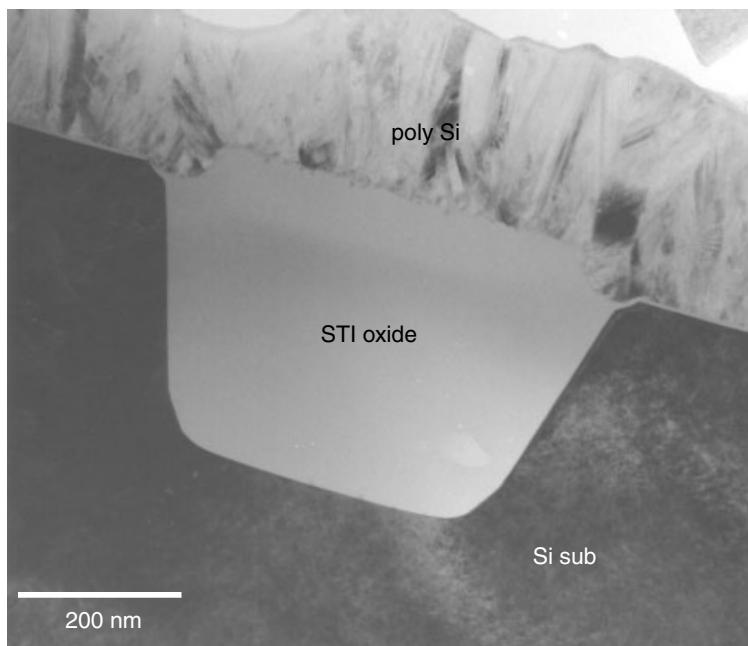


Figure 6.58 TEM cross section of a typical shallow trench isolation structure. Note the tapered sidewall configuration which facilitates the trench's filling with oxide and stress release.

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

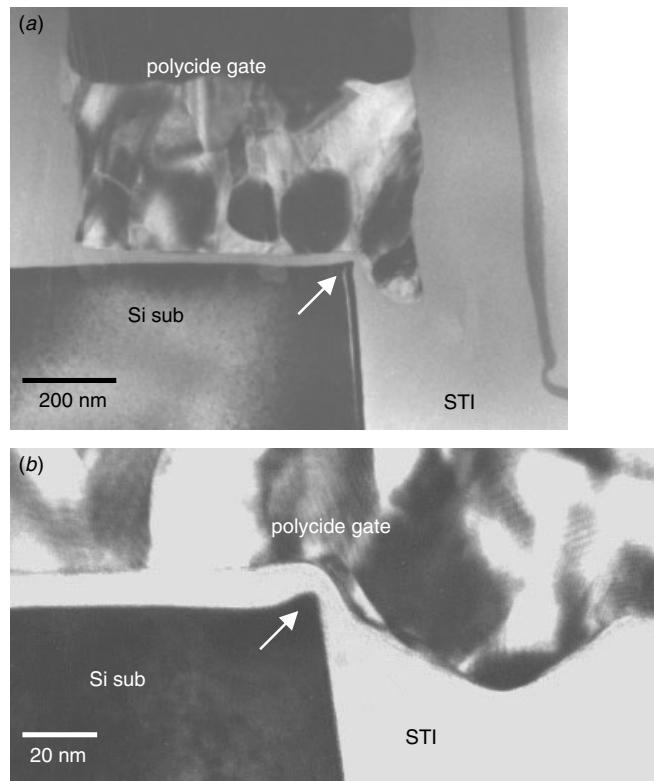


Figure 6.59 TEM cross section of STI corners from a commercially available device showing the STI corner thinning issue, as indicated.

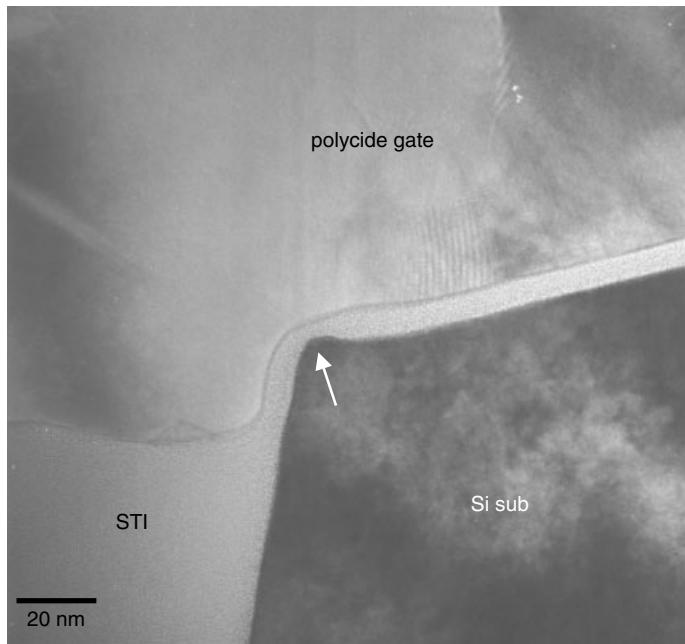


Figure 6.60 TEM cross section of STI corners showing the STI corner thinning issue, as indicated. The corner lift above the original Si substrate surface is due to the stress-induced rearrangement of the topology.

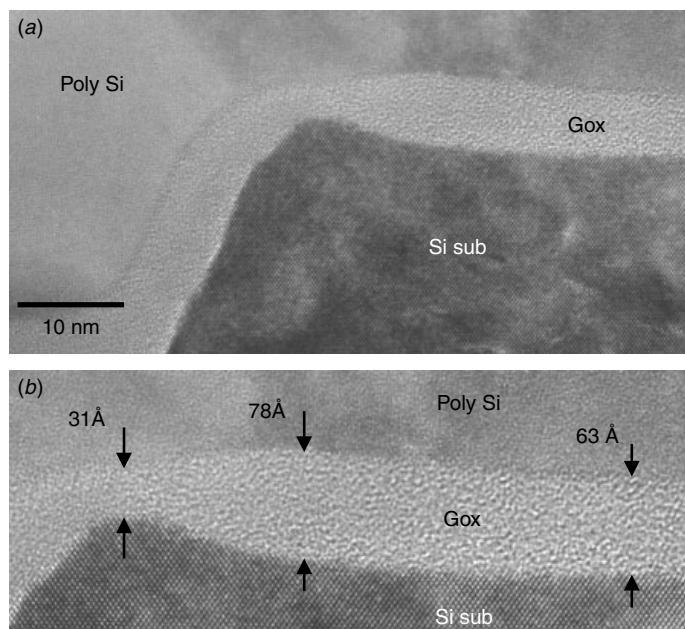


Figure 6.61 Close-up of the thinning area where the local thinning is accompanied by a local thickening area next to it. This indicates a surface rearrangement through an interface diffusion.

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

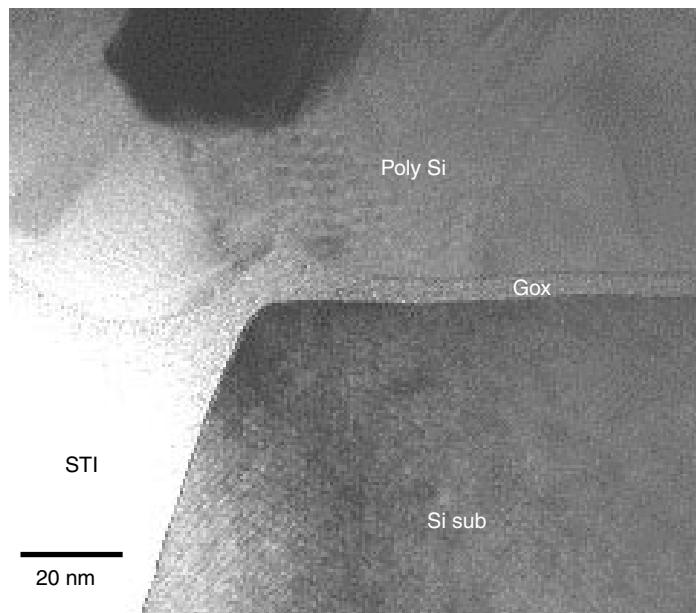


Figure 6.62 Corner rounding using high-temperature re-oxidation after CMP. The rounding radius depends mainly on the oxidation temperature contour.

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.

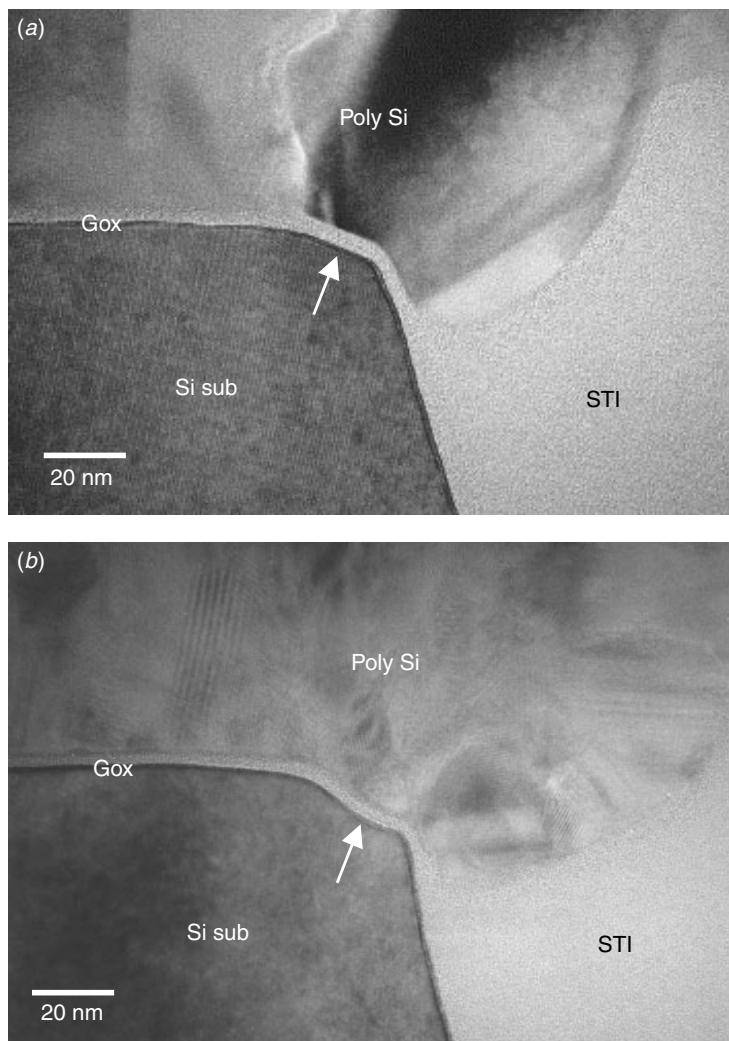


Figure 6.63 Short- and long-pad oxide undercut etching resulting in different STI corner shapes. (a) The short undercut shows facets, and (b) the long undercut shows a concave contour.

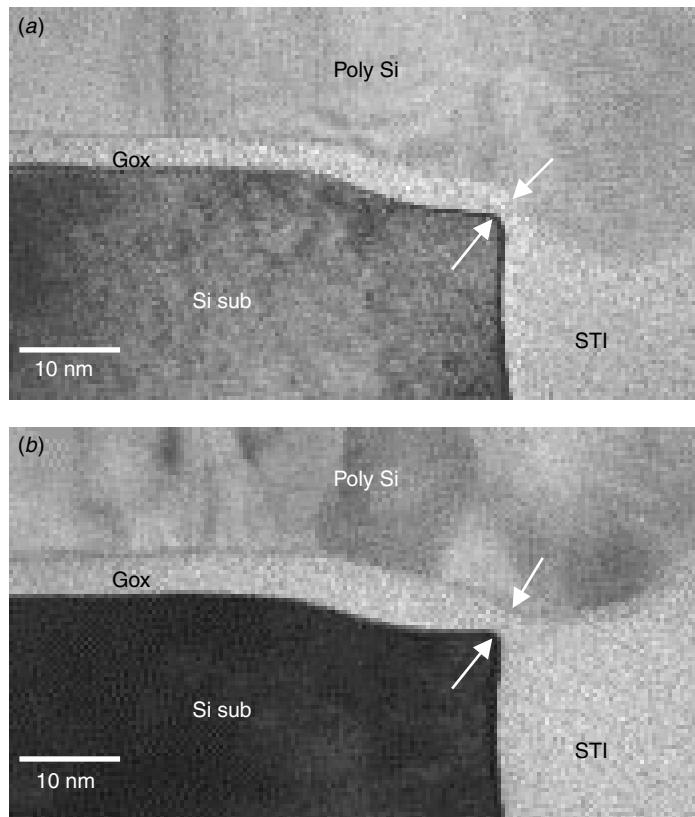


Figure 6.64 Low-densification temperature after trench refill showing a nonprotruding but sharp STI corner. Combined with the polysilicon wrap, the gate oxide thinning is still unavoidable. (a) 50 Å oxide, and (b) 70 Å oxide.

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

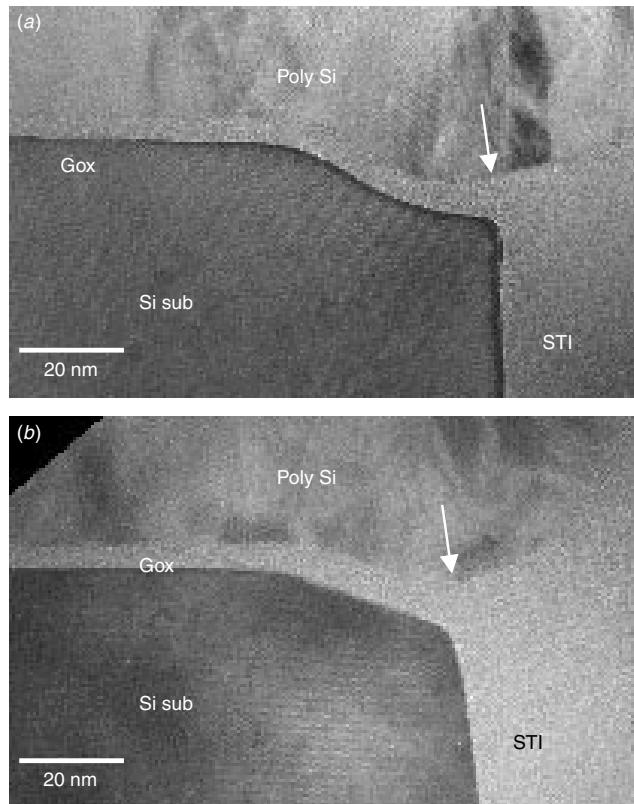


Figure 6.65 Without HF dip, the polysilicon does not wrap around the Si-substrate corner, and thus no gate oxide thinning occurs even when the Si-substrate corner is sharp, as in (a), due to low-temperature re-oxidation.

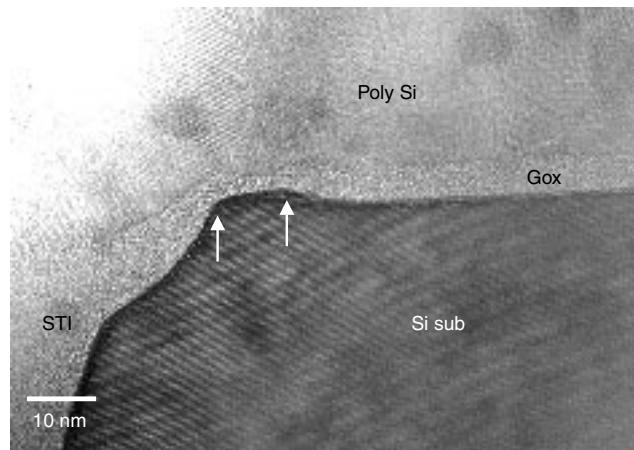


Figure 6.66 STI corner thinning due to the dry Kooi effect. Notice that the thinning area extends and forms a platitude instead of the usual peak. (Sample courtesy Dr. Bala, IME, Singapore)

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.

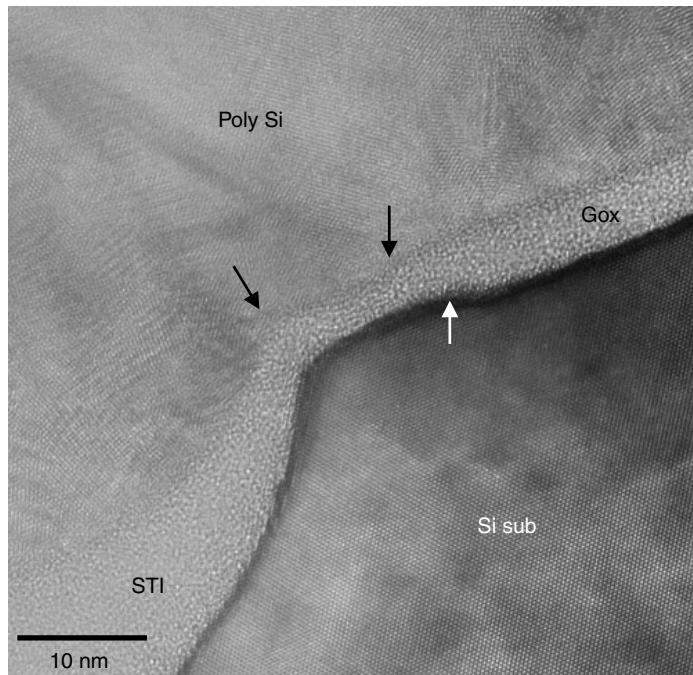


Figure 6.67 Detailed analysis of the corner thinning shows corresponding thinning from the polysilicon side (black arrows), and Sloped tailing on the plateau over the inner side (white arrow). All of these effects demonstrate a Kooi mechanism in action. (Sample courtesy Dr. Bala, IME, Singapore)

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

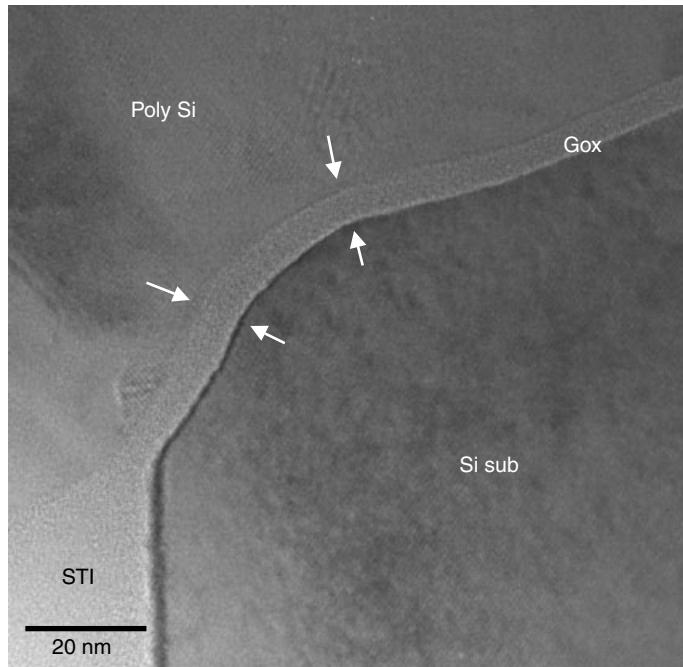


Figure 6.68 Thick sacrificial oxidation can eliminate the local thinning due to the dry Kooi effect. The local hump plateau still exists, as indicated, but not the gate oxide thinning. (Sample courtesy Dr. Bala, IME, Singapore)

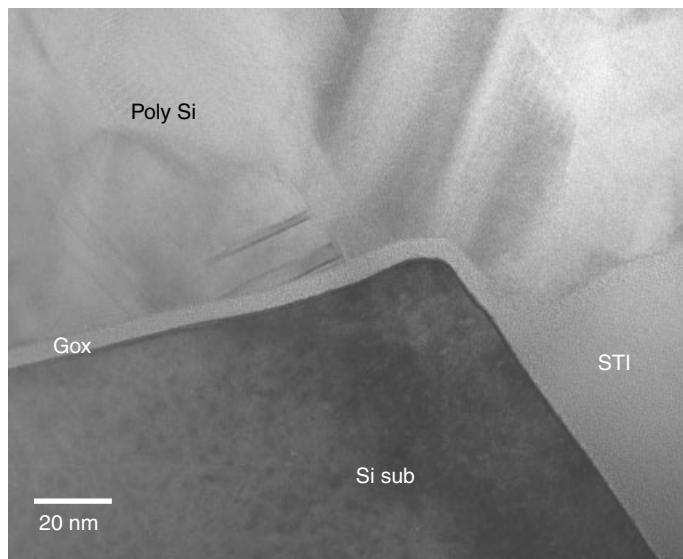


Figure 6.69 STI corner rounding by high-temperature oxidation without the pad oxide undercut. No gate oxide thinning is observed.

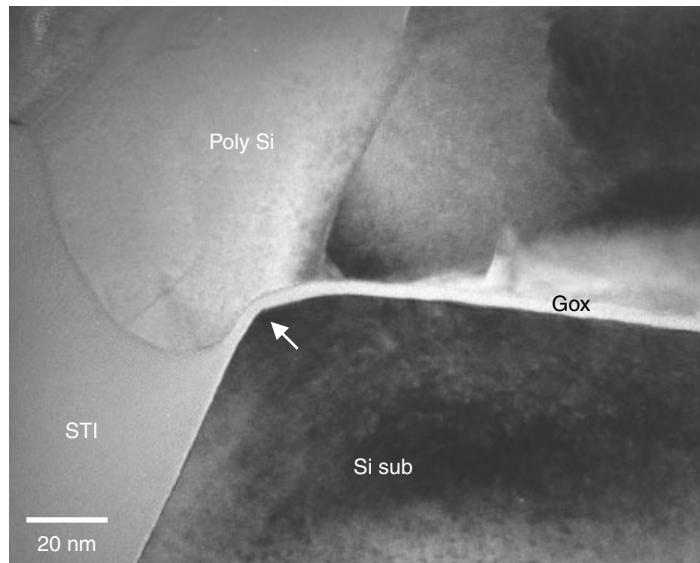


Figure 6.70 Short-pad oxide undercut to modify the corner's shape. Although the corner thinning is not obvious, a sharp tip at the corner can still result in a double hump and early breakdown in the gate oxide.

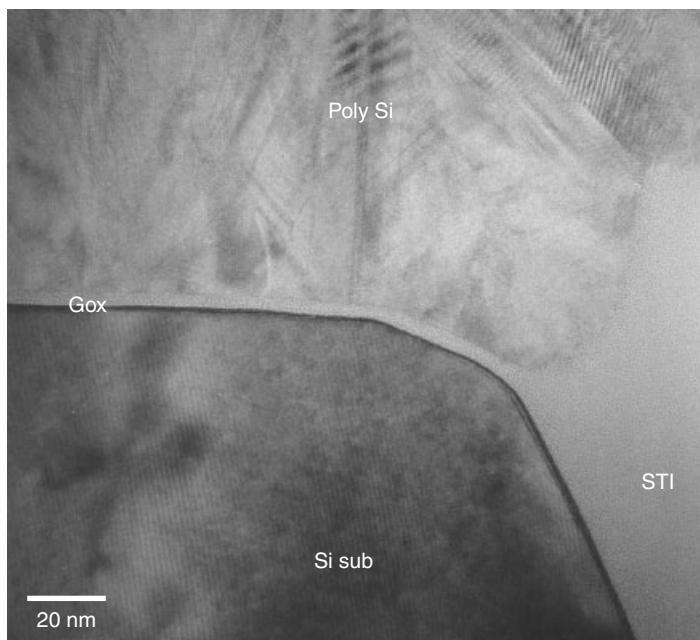


Figure 6.71 STI corner with short liner oxide pre-clean undercut and corner rounding. No thinning nor sharp corner is observed.

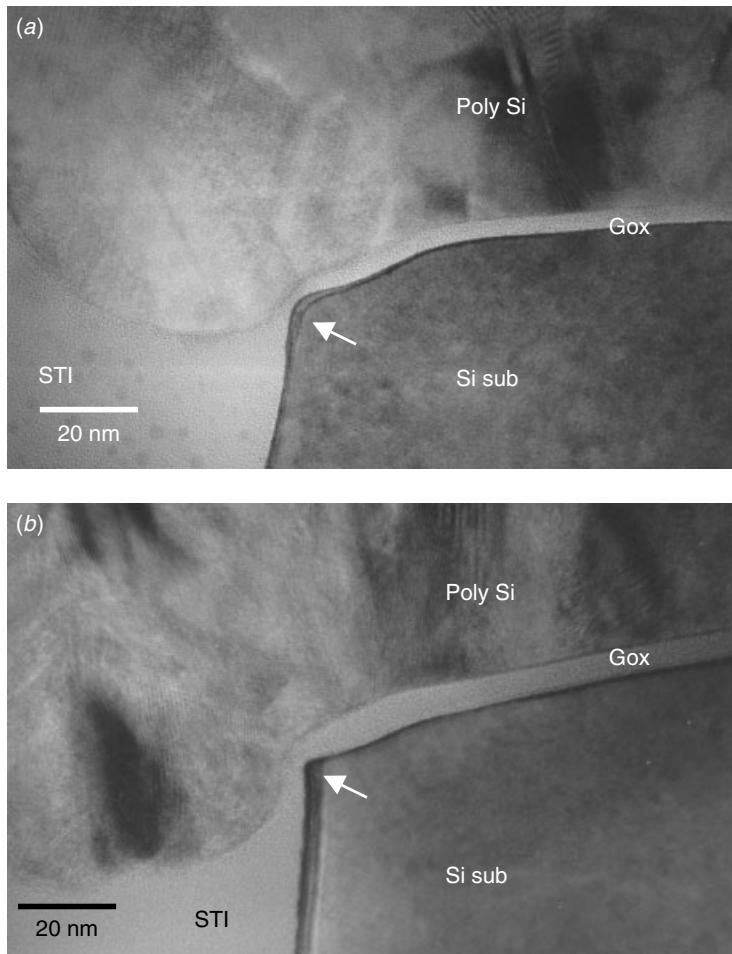


Figure 6.72 Long-pad oxide pre-clean undercut to introduce a concave corner contour. Gate oxide thinning appears due to the polysilicon wrap, and sharp corner.

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.

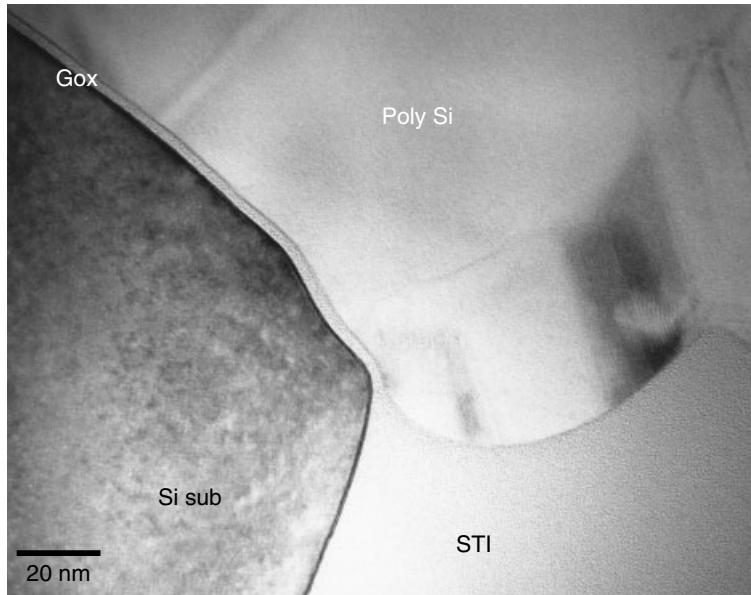


Figure 6.73 Long-pad oxide pre-clean undercut to introduce a concave corner contour. The corner rounding prevents the gate oxide from thinning even with the polysilicon wraparound.

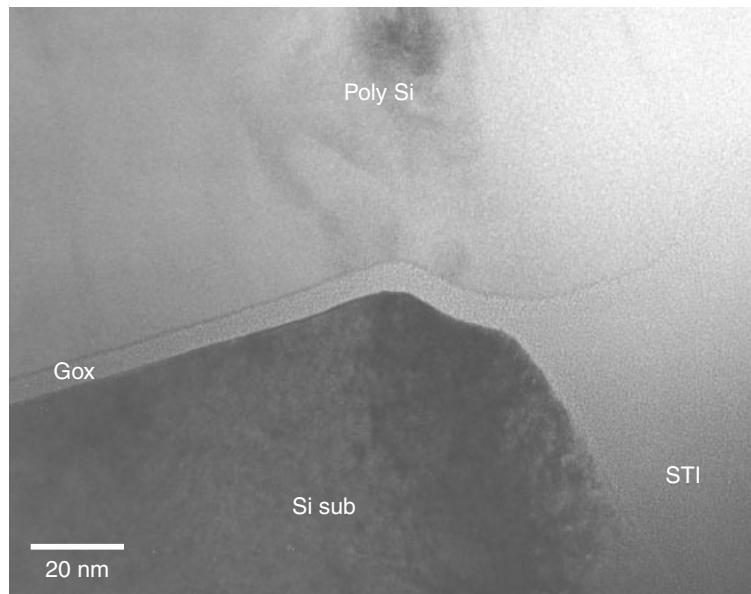


Figure 6.74 Long-pad oxide pre-clean undercut to introduce a concave corner contour. Corner rounding, without the polysilicon wrap, has prevented the gate oxide from thinning.

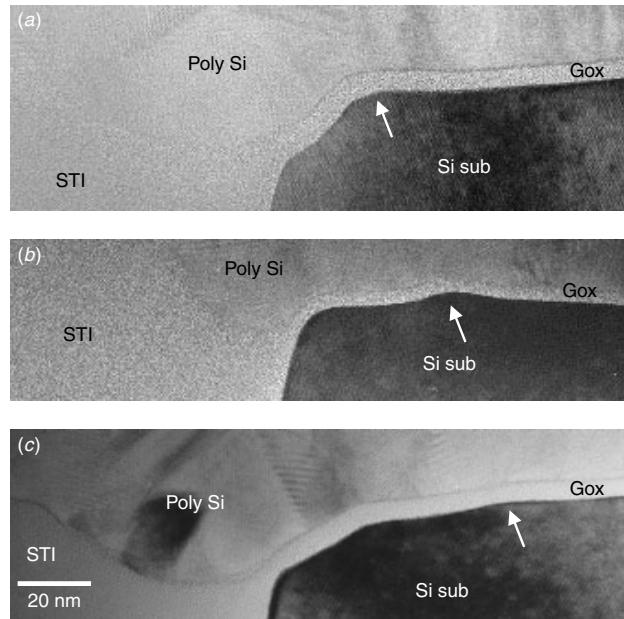


Figure 6.75 Dry Kooi effect observed along the STI near corner areas but at different positions, depending on different pad oxide under pre-clean timing and process details.

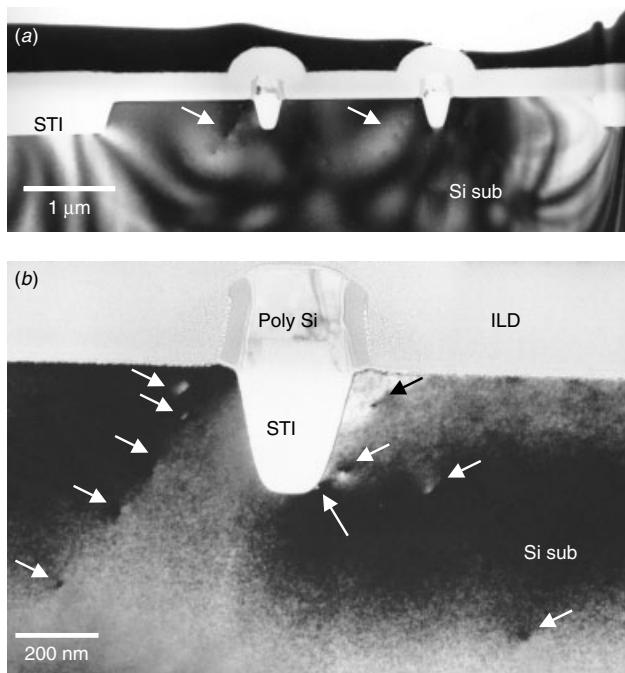


Figure 6.76 TEM cross section of the narrow STI lines showing Si-substrate defect arrays, as indicated. Noted that in (a) there are many dislocations associated with the isolated narrow STIs but no defect associated with the wider STIs. (Sample courtesy Dr. Bala, IME, Singapore)

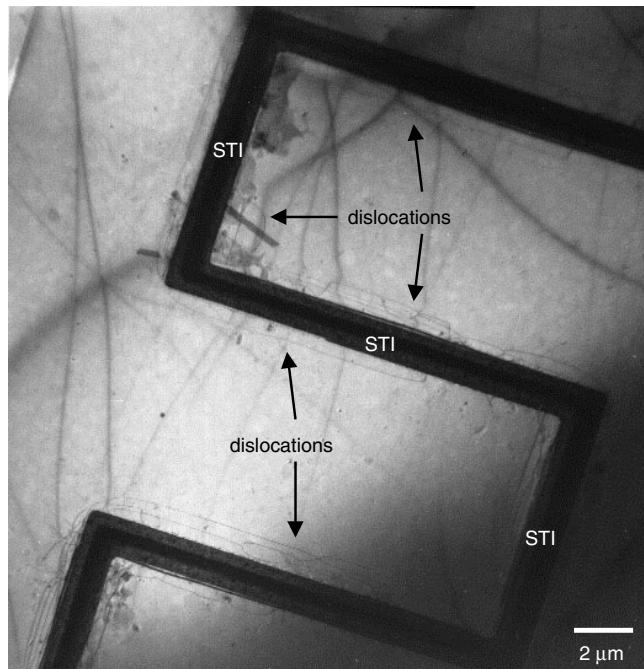


Figure 6.77 TEM plan view of narrow STI lines showing the Si-substrate dislocation line entanglement, as indicated. Note that the polyicide on top of the STI is not removed and overlaps with the STI. (Sample courtesy Dr. Bala, IME, Singapore)

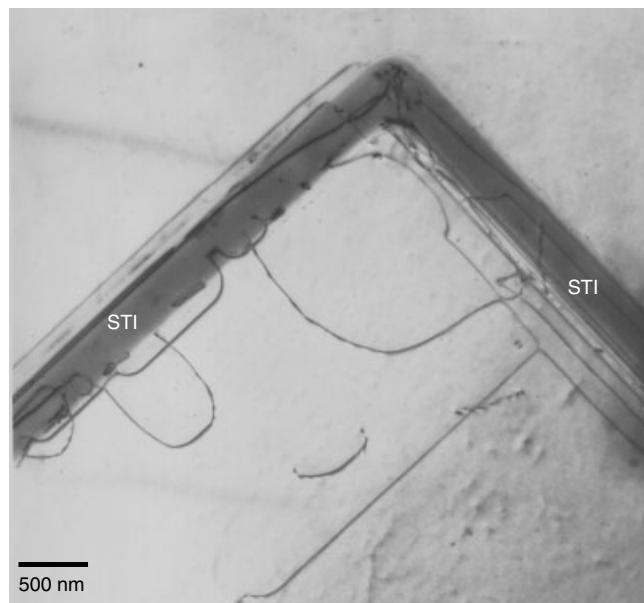


Figure 6.78 TEM plan view of narrow STI lines. Dislocation lines, loops, and arcs are interacting with the STI line. (Sample courtesy Dr. Bala, IME, Singapore)

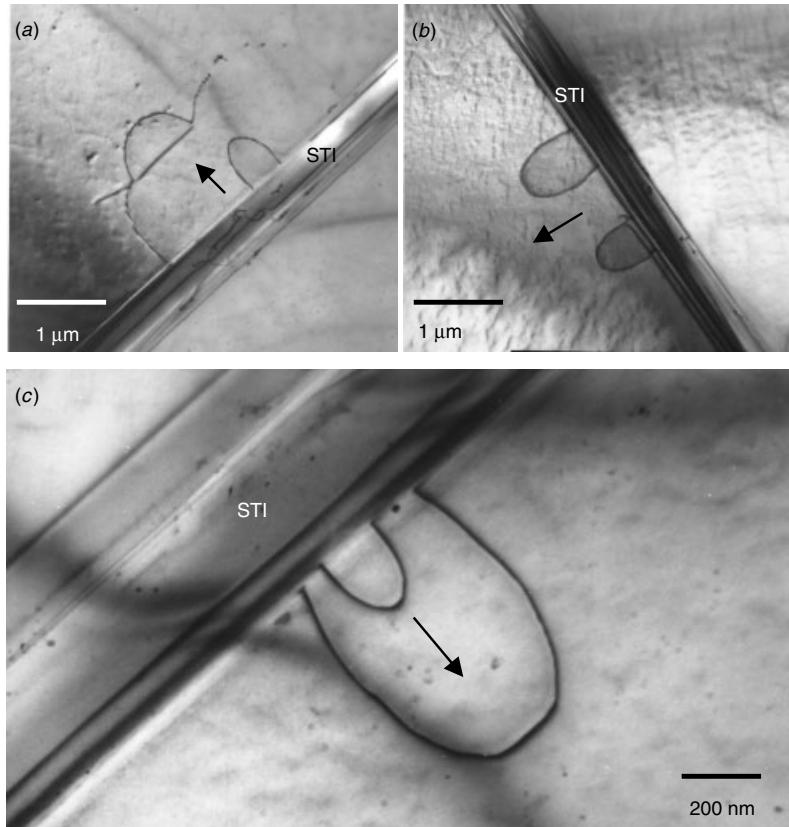


Figure 6.79 TEM plan view of narrow STI lines. The dislocations originate from and are pumped out of the STI sidewall. (Sample courtesy Dr. Bala, IME, Singapore)

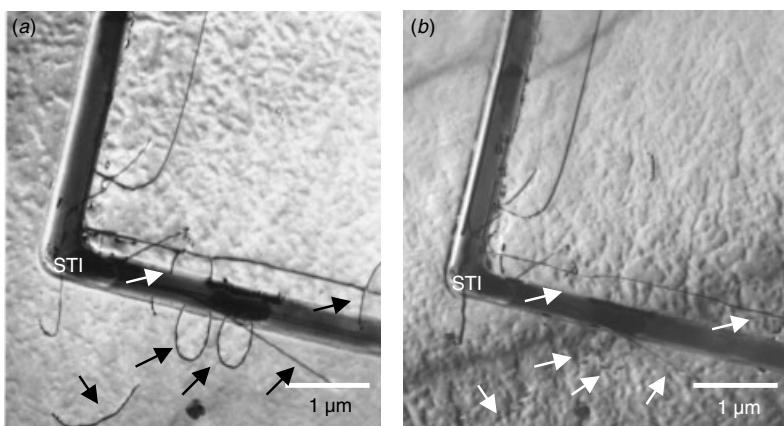


Figure 6.80 TEM plan view of narrow STI lines. The dislocations Burger vectors are determined under different illumination (\mathbf{g} vectors) (a) Various dislocations appearing at the (100) pole, (b) the same location but at a different \mathbf{g} shows some of the dislocations disappeared. (Sample 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

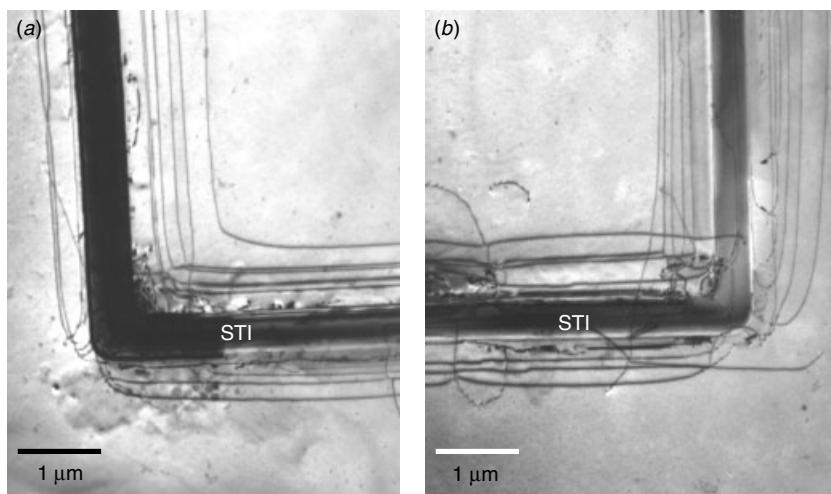


Figure 6.81 TEM planar view of narrow STI lines. The dislocation tangles run in parallel with the STI lines and are the result of a high-stress gradient due to oxidation on trench's sidewall and thermal annealing. (Sample courtesy Dr. Bala, IME, Singapore)

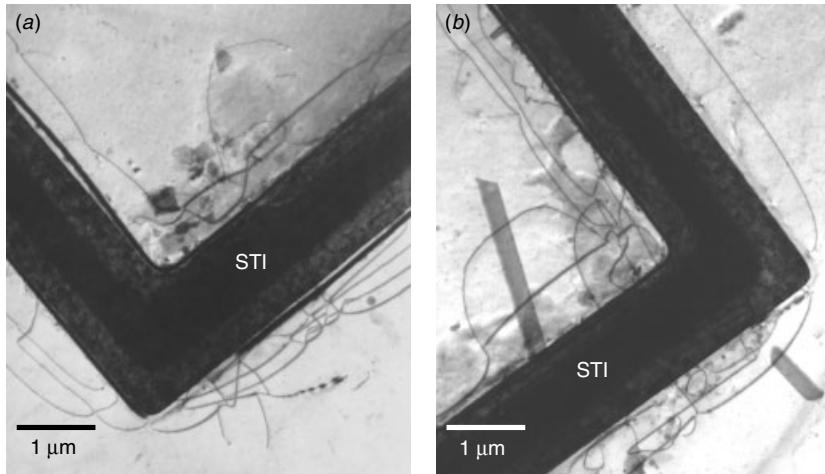


Figure 6.82 TEM plan view of narrow STI lines. The size and width of polycide layer on top of STI may have an effect on the dislocations. (Sample courtesy Dr. Bala, IME, Singapore)

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 $0.35\text{ }\mu\text{m}$ 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

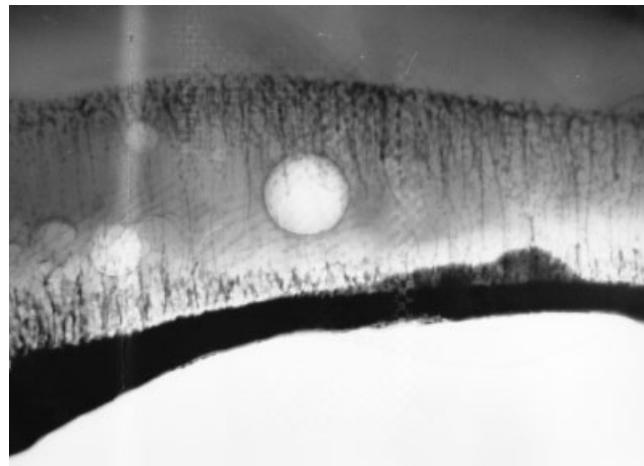
- M. A. Alam, B. E. Weir, P. J. Silverman, Y. Ma, and D. Hwang, *IEDM Tech. Dig.*, 529–532, 2000.
- N. Balasubramanian, E. Johnson, I. V. Peidous, Shiu Ming Jr., and R. Sundaresan, *J. Vac. Sci. Tech.*, **B18** (2), 700–705, 2000.
- N. Balasubramanian, E. Johnson, C. Perera, S. M. Chen, T. T. Sheng, 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

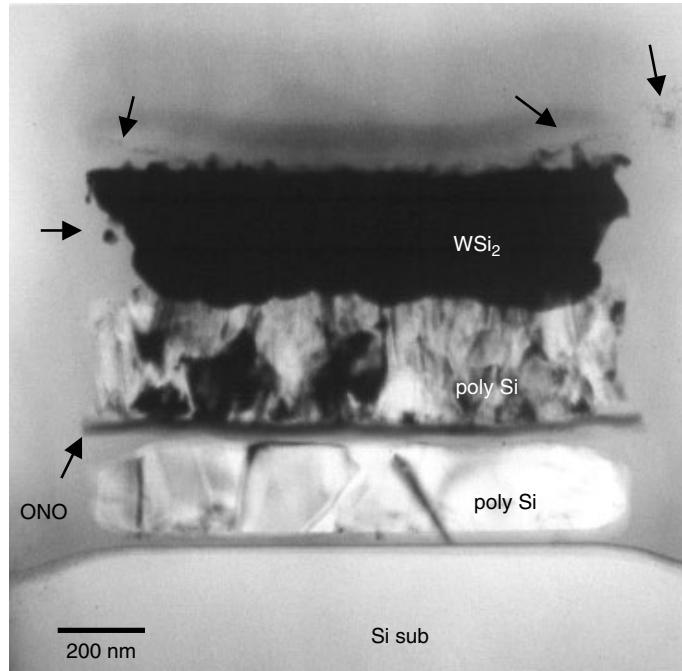


Figure 7.1 TEM cross section of W polycide gate structure in an EPROM stack gate device. Small WSi_2 whiskers and particles, as indicated, are typical of the WSi_2 process.

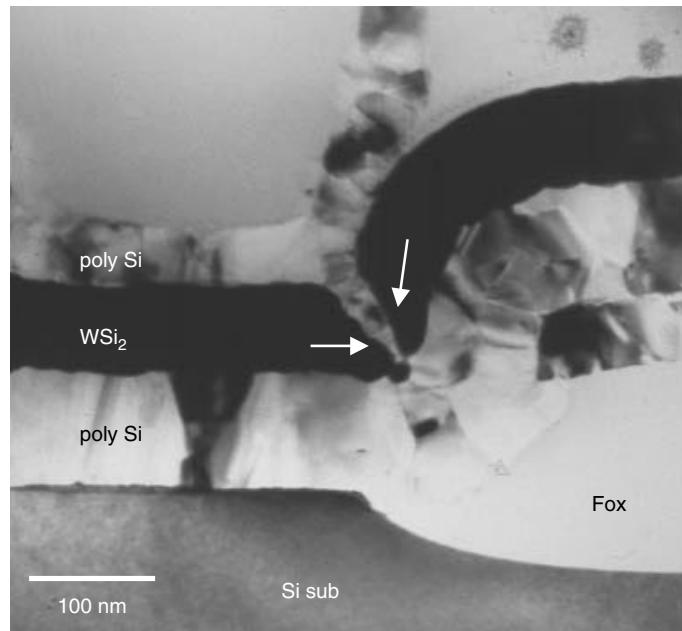


Figure 7.2 TEM cross section of the W-polycide film over a field oxide bird's beak. The WSi_2 film shows a discontinuity over the bird's beak, as indicated, that causes high-sheet resistivity.

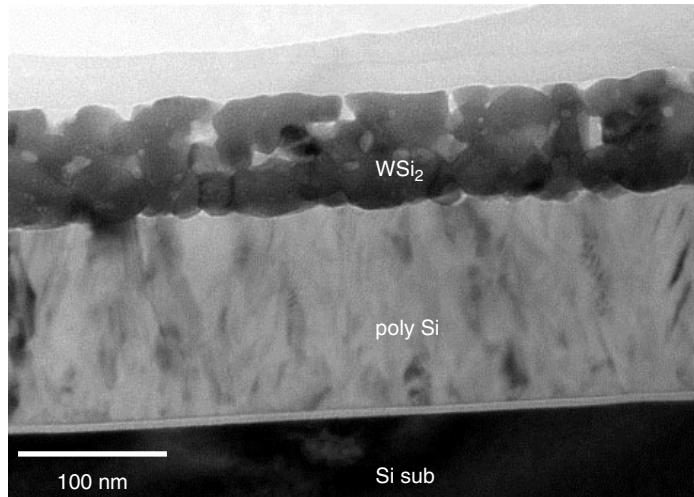


Figure 7.3 TEM cross section of the W-polycide gate structure. A typical W-polycide process produces the porous WSi_2 shown here, which is due to volume shrinkage within the film. Such shrinkage often results in an internal stress build up and thus void formation.

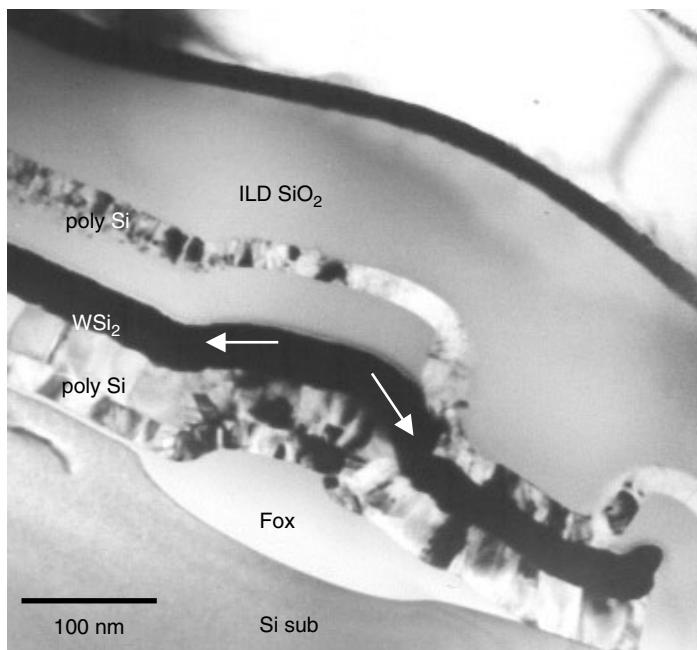


Figure 7.4 TEM cross section of the W-polycide film over a field oxide. The WSi_2 film is smooth, continuous, and shows no step coverage issue. The poly Si layer on top of the WSi_2 is another conductor layer in the SRAM structure.

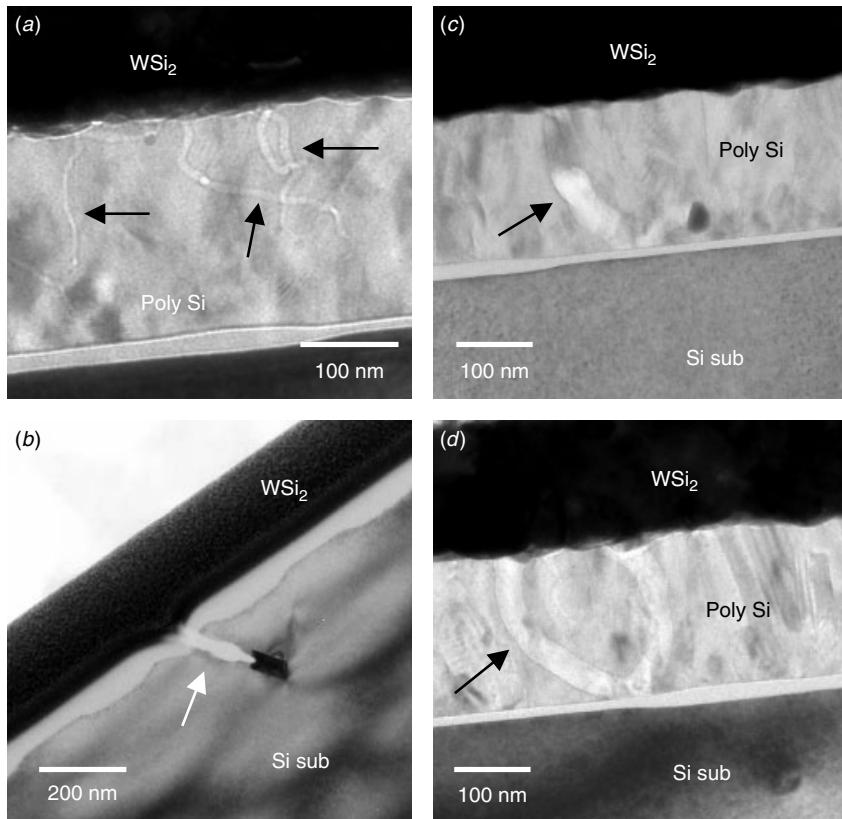


Figure 7.5 Wormholes (indicated by arrows) as observed in the production wafers. Most of the wormholes start from the WSi_2 and penetrate into the polysilicon (*a, c, d*). Some of the wormholes penetrate through the field oxide and into the Si substrate, as seen in (*b*). (Sample courtesy MXIC, Taiwan)

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.

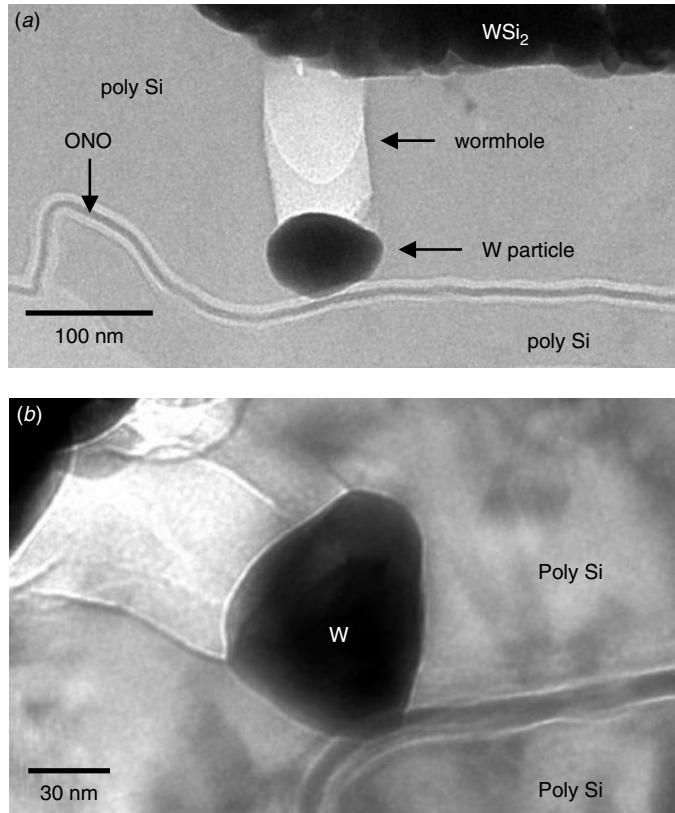


Figure 7.6 A W particle can be found at the bottom of each wormhole. Detailed analysis shows that the penetration was stopped by the ONO layers. (Sample courtesy MXIC, Taiwan)

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 Si₃N₄ 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 SiO₂ (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

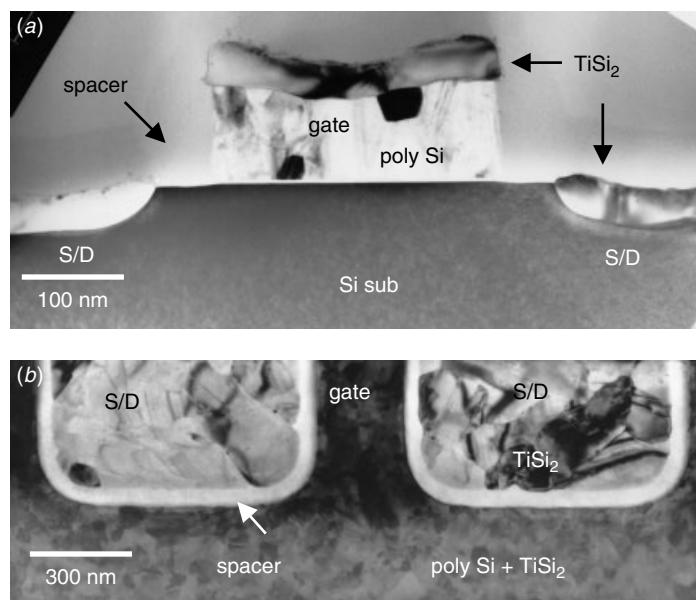


Figure 7.7 TEM cross section (a), and plan view (b) of TiSi_2 salicide basic device structure. The process utilizes an oxide spacer instead of the nitride spacer and thus the spacer cannot be readily seen. (Sample courtesy Prof. Kin Leong Pey, NTU, Singapore)

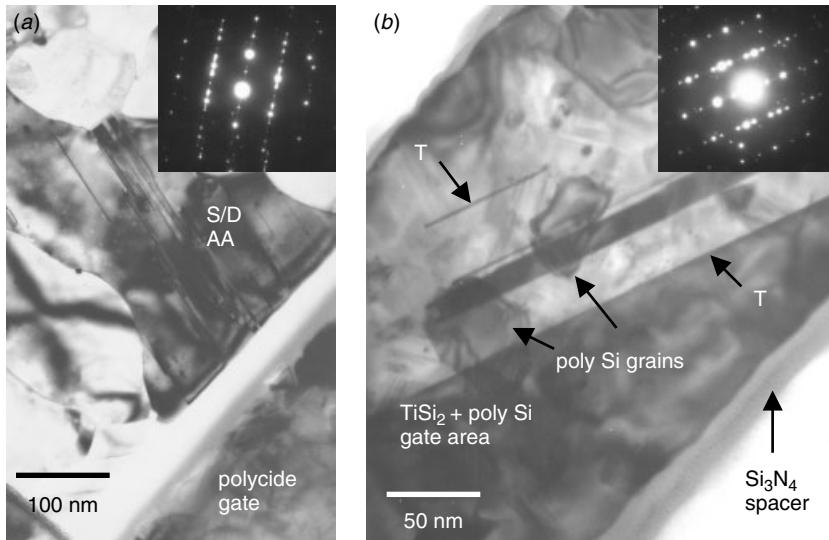


Figure 7.8 TEM plan view of the basic TiSi_2 salicide device structure and its associated diffraction patterns. Large C54 TiSi_2 silicide grains are observed on both active area (a) and the poly silicon gate (b). The grain size difference between polysilicon and TiSi_2 is best illustrated in (b). Arrows 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

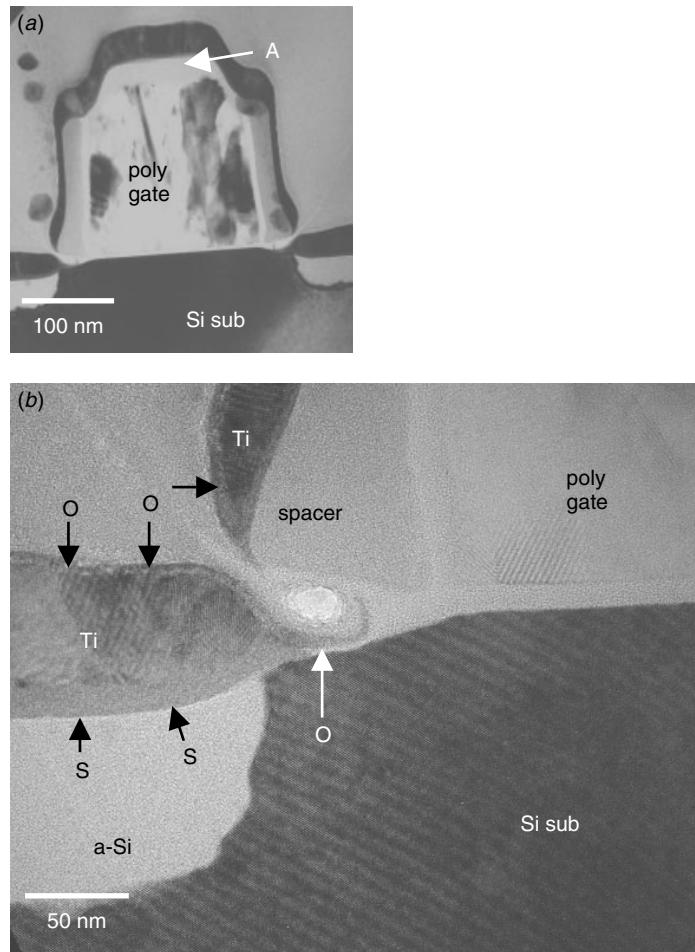


Figure 7.9 TEM cross section of the salicide gate structure after Ti deposition but before silicidation annealing. (a) Overall view of the gate where the ion implantation induced amorphization on top of polysilicon gate is observed, as indicated by arrow A. (b) Close-up at the gate corner, where a thin 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

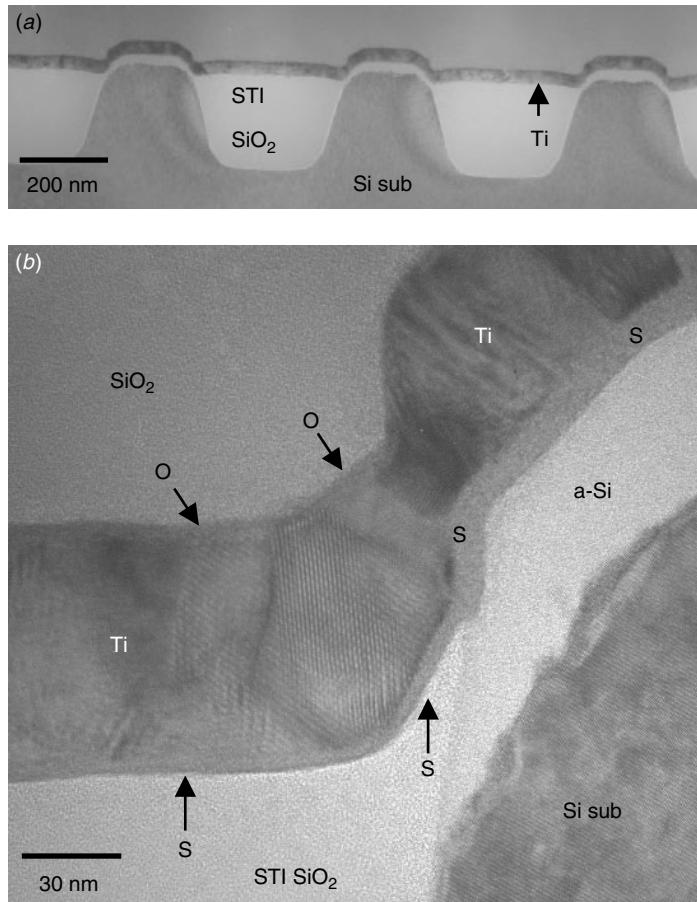


Figure 7.10 TEM cross section of the salicide structure near the STI after Ti deposition but before silicidation annealing. (a) Overall view where the ion implantation induced amorphization on top of active regions is observed. (b) Close-up at the STI corner where a thin amorphous silicide layer is observed, 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

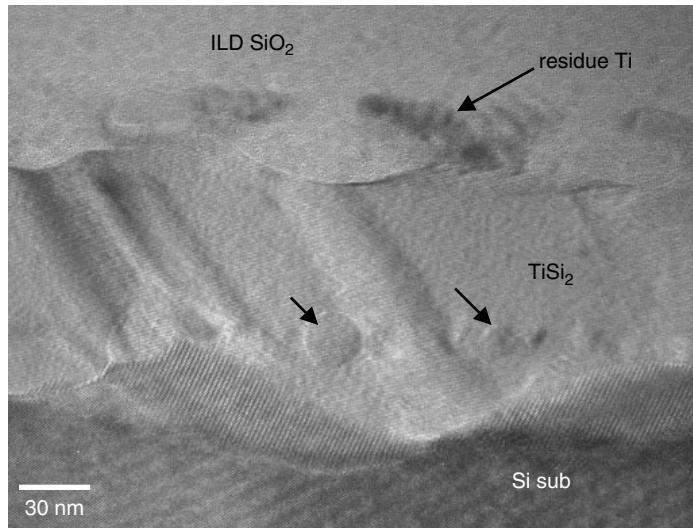


Figure 7.11 TEM cross section of the salicide structure. The sample was made after low-temperature annealing but before the removal of the residue Ti metal. The low-temperature annealing has formed TiSi_2 . The spherical objects (as indicated by the arrow) observed at the bottom of the TiSi_2 layer have a high fluorine content and are thought to be F_2 bubbles. (Sample courtesy Prof. Kin Leong Pey, NTU, Singapore)

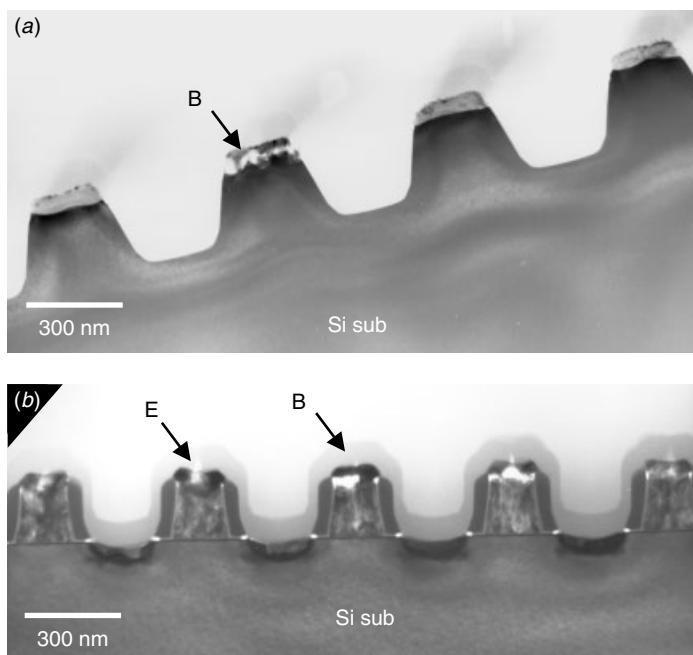


Figure 7.12 TEM cross section of the salicide structure. Voids formed in between the TiSi_2 and Si in (a) STI/actives, (b) gate and S/D. Arrow E shows voids that float to the surface and arrow B voids that are submerged in and under TiSi_2 . (*JVST*, **B19** (6), 2252–2257, 2001, with permission from AVS)

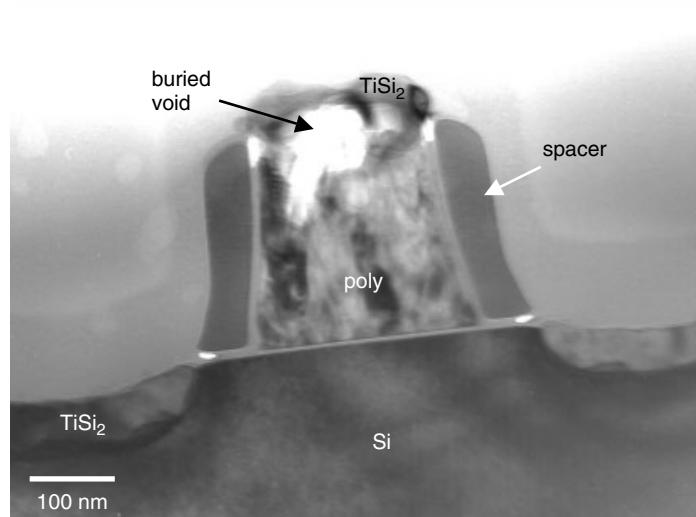


Figure 7.13 TEM cross section of the salicide structure showing voids formed in between TiSi₂ and Si. Such voids are invisible in a surface topology examination. (Electrochemical and Solid State Lett., **319**, 442–445, 2000 reprint with permission from ECS)

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

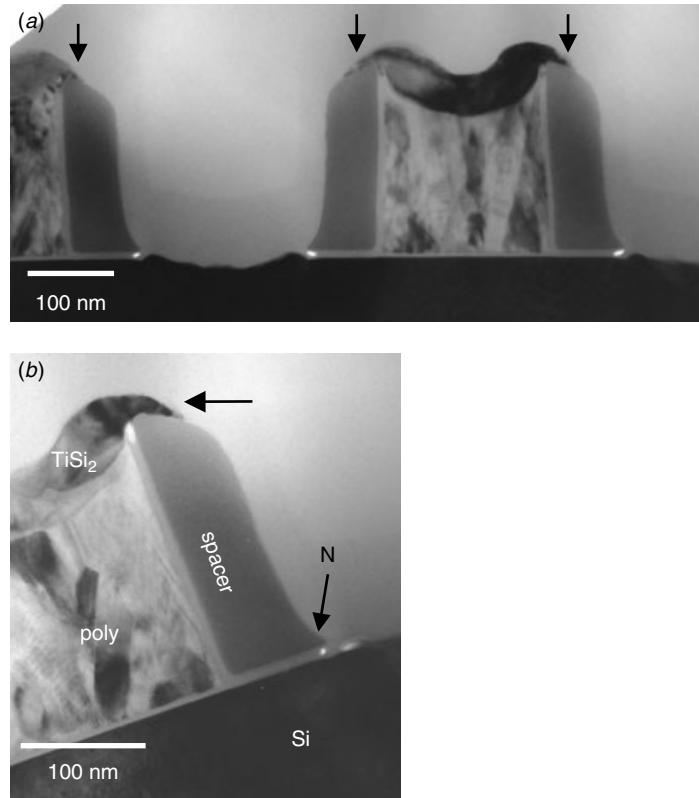


Figure 7.14 TEM cross section of the Ti silicide's encroachment on the poly-Si gate's nitride spacer (a). Close-up at the corner (b) shows clearly that the encroachment is the same crystal grain as TiSi₂ on poly Si, and not Ti oxide or metal, Ti residue. Note that there is no TiSi₂ growth at the spacer's 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.

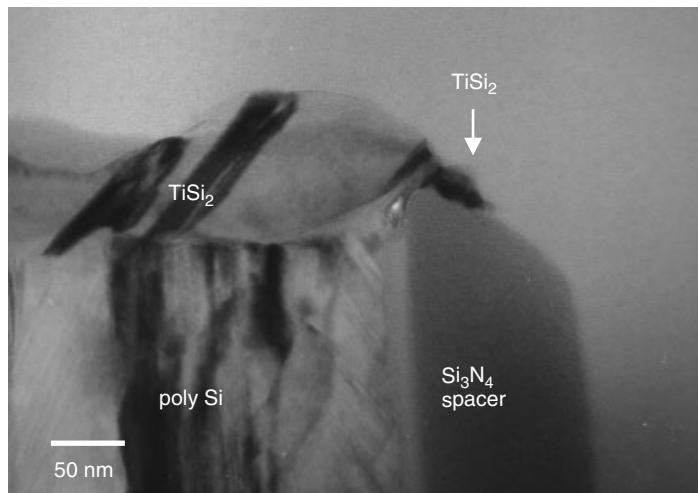


Figure 7.15 TEM cross section with the Ti silicide encroachment on the polygate nitride's spacer. The TEM image shows clearly that the encroachment is TiSi₂ crystalline. Most often it is of the same crystal grain as that TiSi₂ on the poly Si. (Sample courtesy Prof. Kin Leong Pey, NTU, Singapore)

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

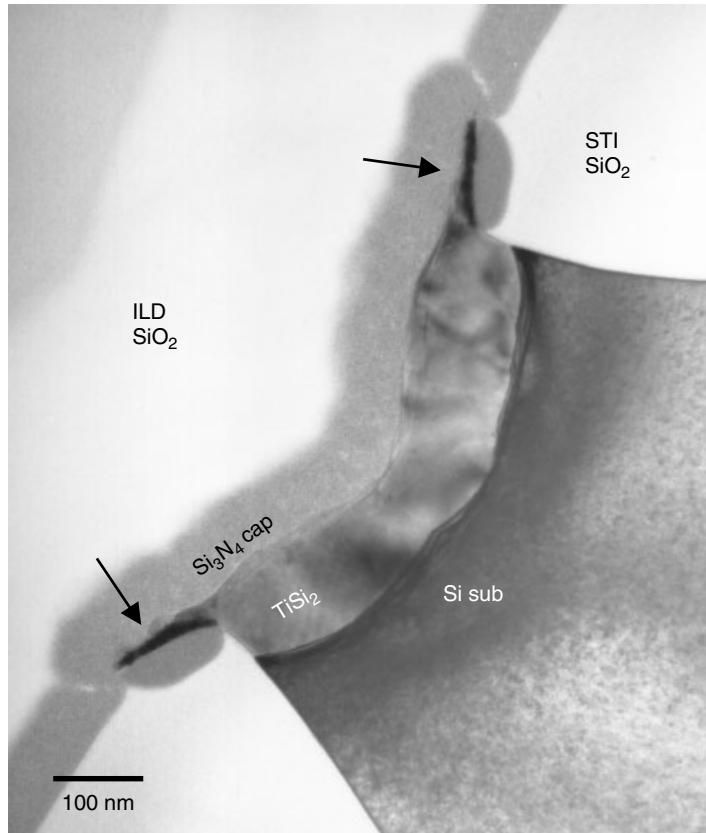


Figure 7.16 Close-up of the Ti silicide encroachment on the top STI corner. TEM image shows clearly that the encroachment is $TiSi_2$ crystalline. (Sample courtesy Prof. Kin Leong Pey, NTU, Singapore)

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

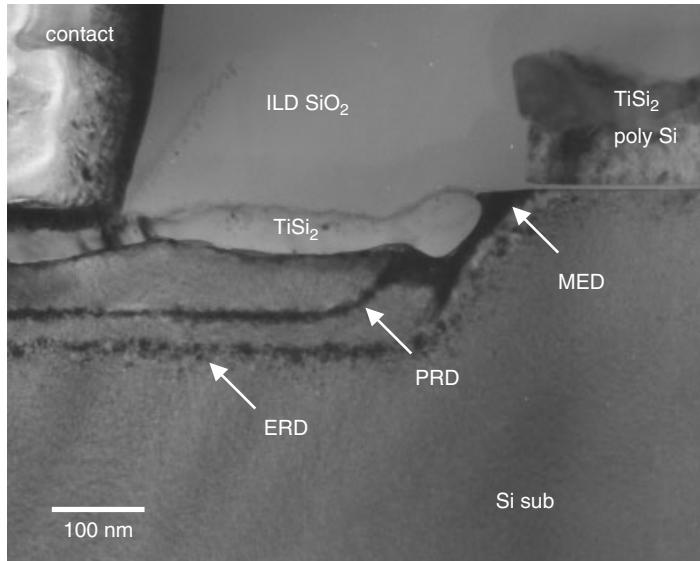


Figure 7.17 TEM cross section of the TiSi₂ salicide with pre-amorphization implant induced residual Si-substrate damages. Three types of damages are clearly observed. mask edge defects (MED), end of range defects (ERD), and projected range defects (PRD), as indicated.

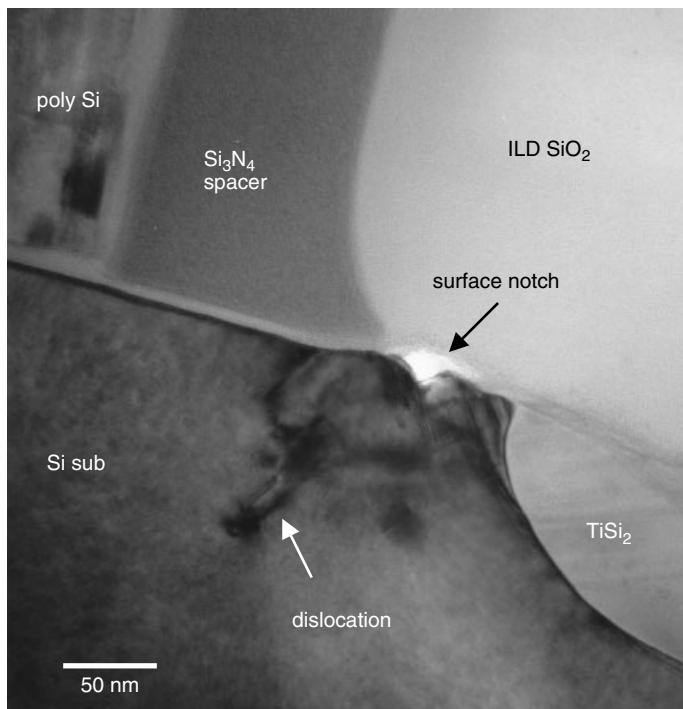


Figure 7.18 TEM cross section of the TiSi₂ salicide with pre-amorphization implant induced residual Si substrate damages. MED-induced extended dislocation and surface notch (both indicated by arrows) observed at the gate corner due to TiSi₂-induced stress.

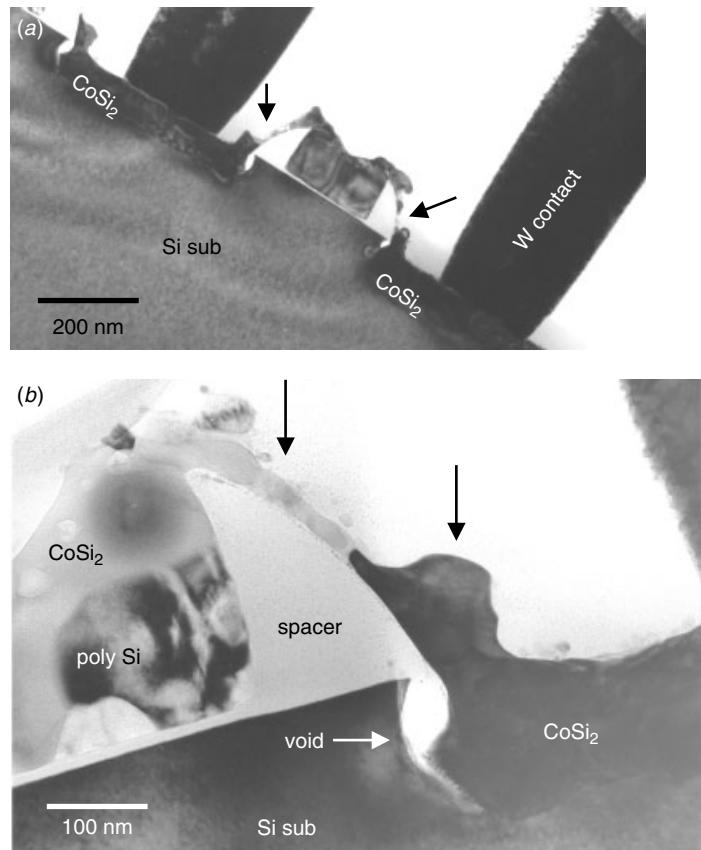


Figure 7.19 TEM cross section of the CoSi_2 salicide device. CoSi_2 extrusion at the gate corner and onto the spacer is observed, as indicated. The void formed at the active area's edge is due to Si depletion. (Sample courtesy Prof. Kin Leong Pey, NTU, Singapore)

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.

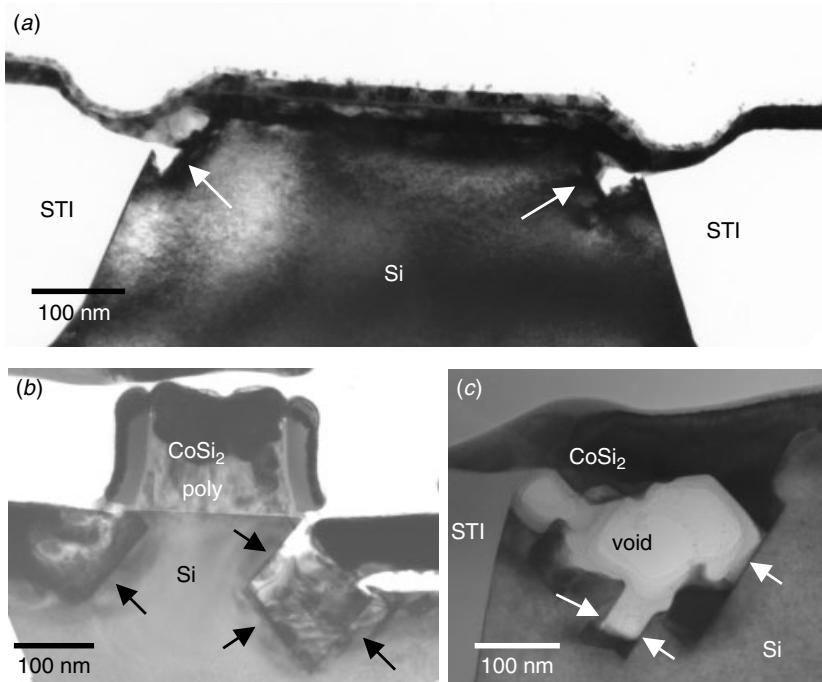


Figure 7.20 TEM cross section of the CoSi₂ salicide device. Close up the voids show faceted void/Si interface along the Si(111) planes, as indicated. The sample has a Ti cap layer. (*Mat. Res. Soc. Symp. Proc.*, **564**, 109–116, 1999 with permission from MRS.)

For samples with isothermal annealing at 520°C, an ultra thin CoSi₂ was observed. After annealing at 520°C for 4 minutes, seeds of voids could be clearly seen at the STI/CoSi₂/Si corner, as observed in Fig. 7.21. The voiding in samples with such a thin CoSi₂ 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 CoSi₂ formation is that the high ramp-up rate leads to a thick local CoSi₂ 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, CoSi₂ tends to form epitaxy with the Si substrate. CoSi₂ takes the CaF₂ structure with lattice parameter $a = 0.5365$ nm at room temperature, which is about 1.2% smaller than that of Si with $a = 0.5431$ nm. Moreover the lattice mismatch improves with increasing temperature because of the large difference in the thermal expansion coefficient of CoSi₂ and Si (the CoSi₂ coefficient is about four times that of Si).

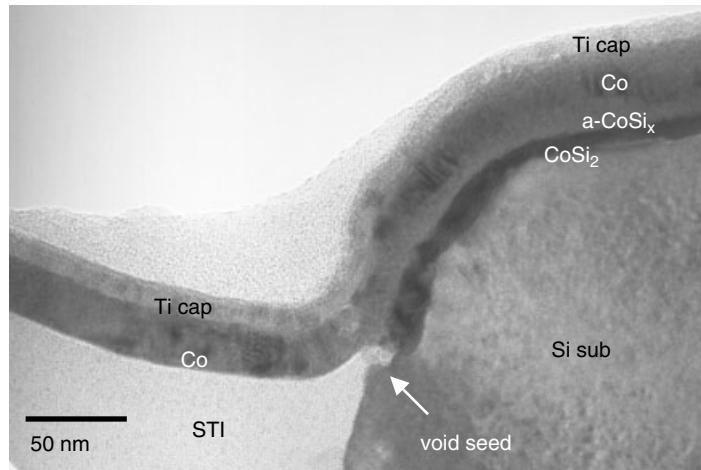


Figure 7.21 TEM cross section of the CoSi_2 salicide device after isothermal annealing at 520°C for 4 minutes. Close up at the STI corner the sample shows the ultra thin silicide phase and a void seed at the corner of the CoSi_2/Si edge, as indicated. (Sample courtesy Prof. Kin Leong Pey, NTU, Singapore)

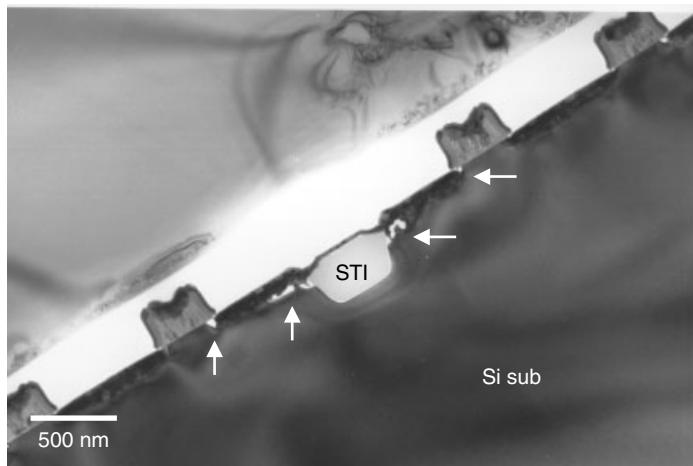


Figure 7.22 TEM cross section of the CoSi_2 salicide device. Voids occurred primarily at the STI corner and few or none at the poly Si gate spacer corners, as indicated. (Sample courtesy Prof. Kin Leong Pey, NTU, Singapore)

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

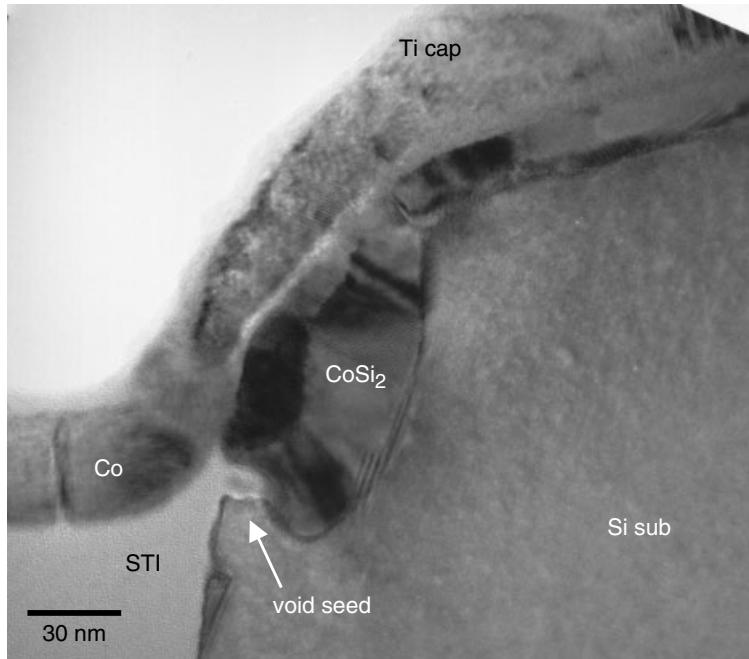


Figure 7.23 TEM cross section of the CoSi_2 salicide device with a high ramp-up rate. The close up of the STI corner shows the CoSi_2 silicide pocket and a void seed at the corner of the CoSi_2/Si edge, as indicated.

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

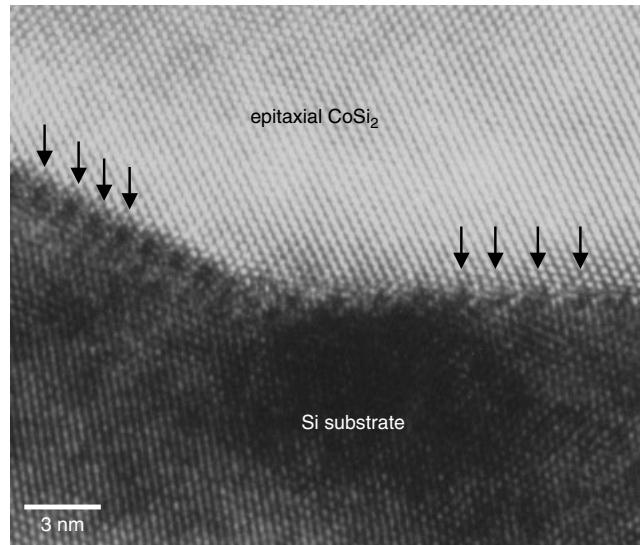


Figure 7.24 HRTEM cross section of the CoSi₂ epitaxial grain on Si substrate. Notice the faceted interface and the periodic mismatch dislocations (arrows) along the interface, as indicated.

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.

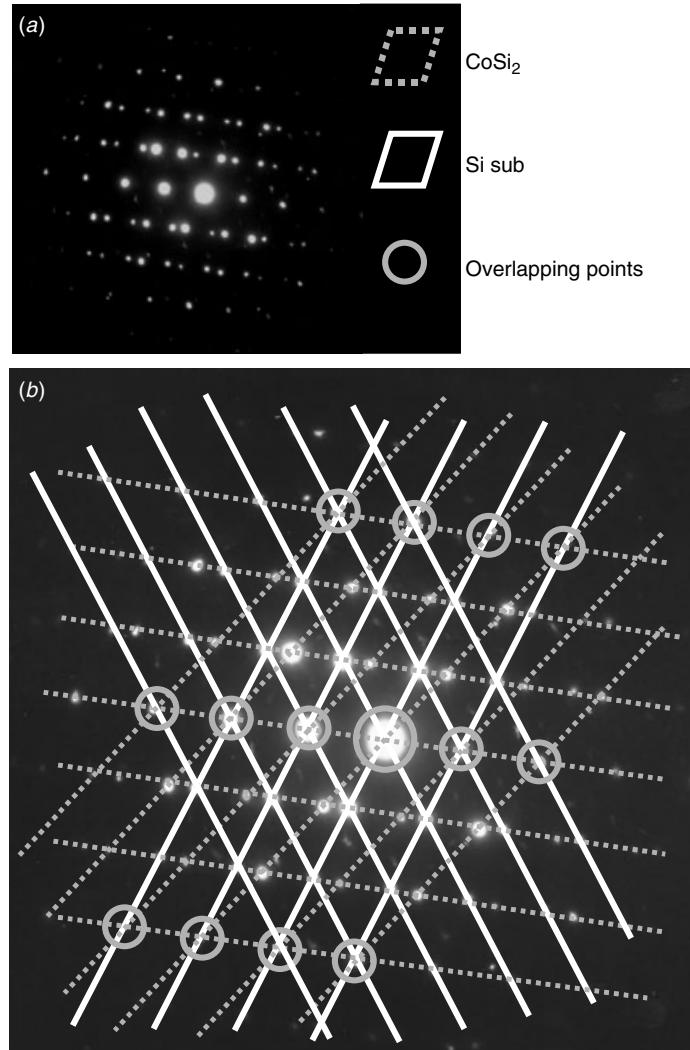


Figure 7.25 Electron diffraction patterns of the interface area of the CoSi₂ epitaxial grain on the Si substrate. (a) The original diffraction pattern and (b) individual lattices from CoSi₂ and Si sub, as marked.

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

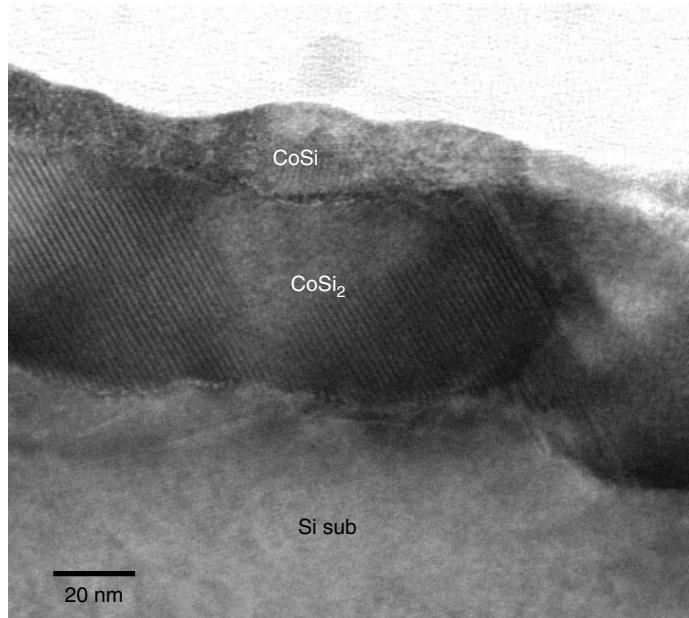


Figure 7.26 Sample *a* forming CoSi/CoSi₂ double silicide layers. The layer structure suggests that CoSi was followed by CoSi₂ nucleation in between CoSi and the Si substrate, which then consumed the CoSi. (Sample courtesy Prof. Kin Leong Pey, NTU, Singapore)

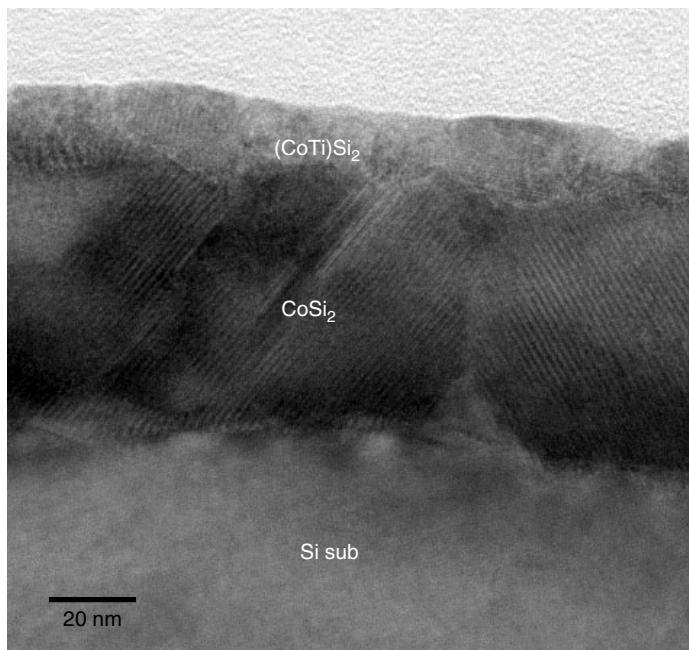


Figure 7.27 Sample *b* where the top layer was identified to be (CoTi)Si₂ instead of CoSi monosilicide. (Sample courtesy Prof. Kin Leong Pey, NTU, Singapore)

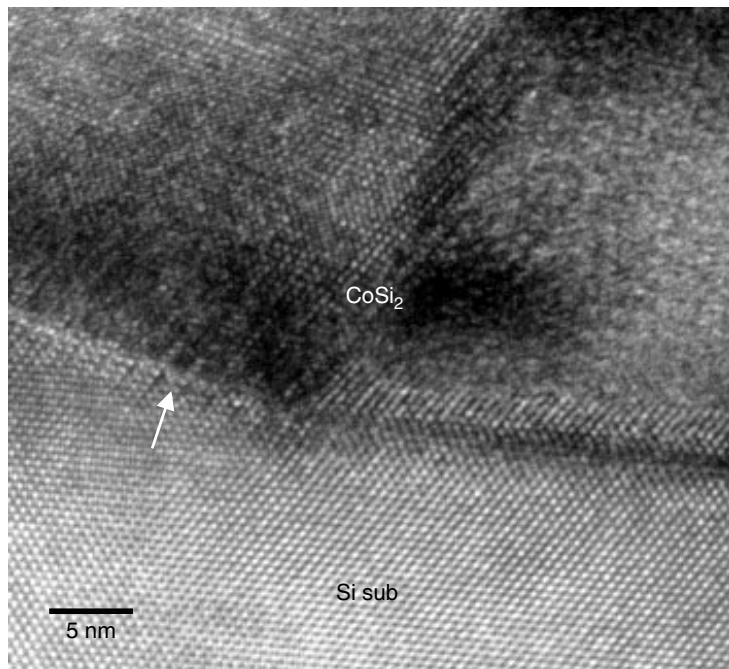


Figure 7.28 Sample *b* where the CoSi_2/Si interface shows partial epitaxy (left-hand side of the micrograph as indicated by arrow). (Sample courtesy Prof. Kin Leong Pey, NTU, Singapore)

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.

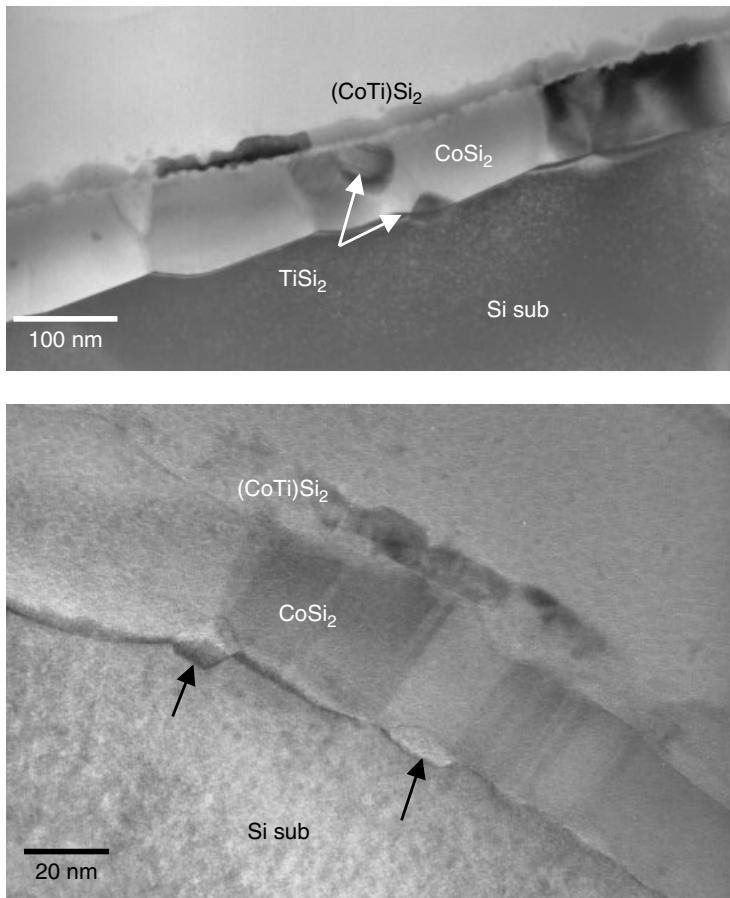


Figure 7.29 Sample *b* where the CoSi₂/Si interface shows void formation and poor epitaxy. Discrete TiSi₂ islands as well as voids are found scattered about the CoSi₂/Si, and CoSi₂/(CoTi)Si₂ interfaces, as indicated. (Sample courtesy Prof. Kin Leong Pey, NTU, Singapore)

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 Si_{0.7}Ge_{0.3} reacts with Ni, depending on the annealing conditions, monosilicide Ni(Si_xGe_{1-x}) can be formed, but Ge tends to segregate laterally and forms discrete islands of Si_{1-m}Ge_m with *m* much higher than 30% and often more than 50%, as shown in Fig. 7.36. On the other hand, when Si_{0.8}Ge_{0.2} reacts with Ni, monosilicide Ni(Si_xGe_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.

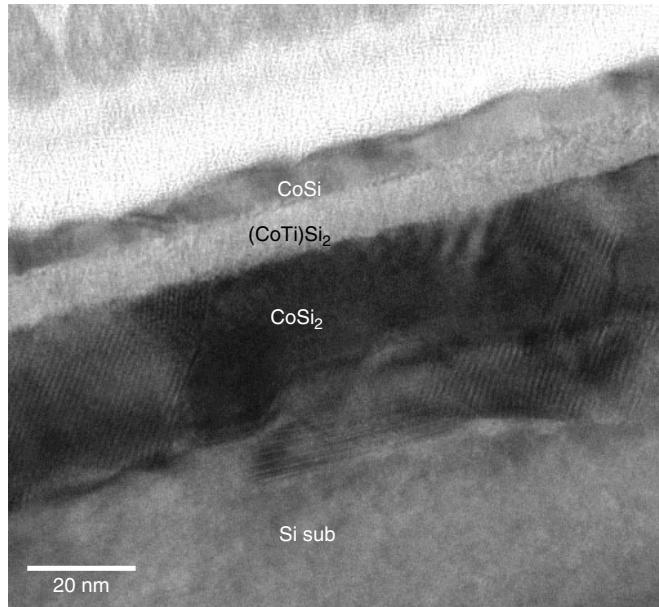


Figure 7.30 Sample *c* where CoSi formed and was followed by (CoTi)Si₂ and then CoSi₂. No epitaxy was observed between the CoSi₂ and Si substrate. The (CoTi)Si₂ appears to be amorphous. (Sample courtesy Prof. Kin Leong Pey, NTU, Singapore)

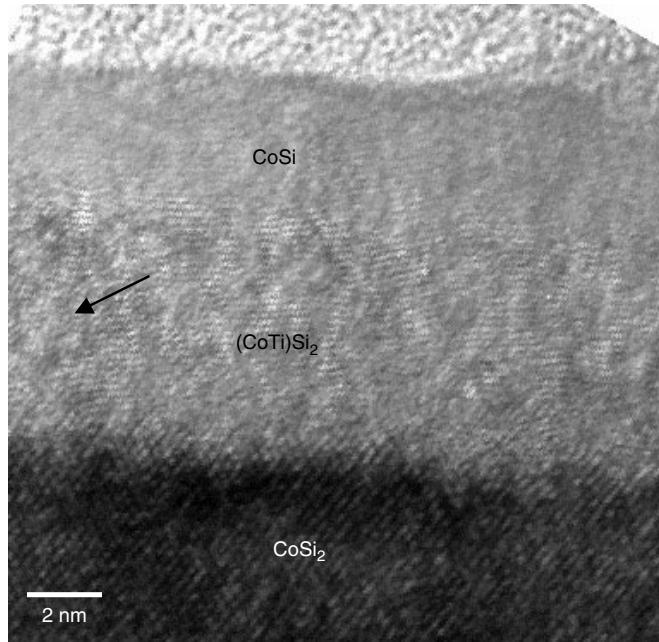


Figure 7.31 Sample *c* where the top layer (CoTi)Si₂ appeared to be amorphous but was confirmed to be crystalline with extremely fine grain, as indicated by the arrow. (Sample courtesy Prof. Kin Leong Pey, NTU, Singapore)

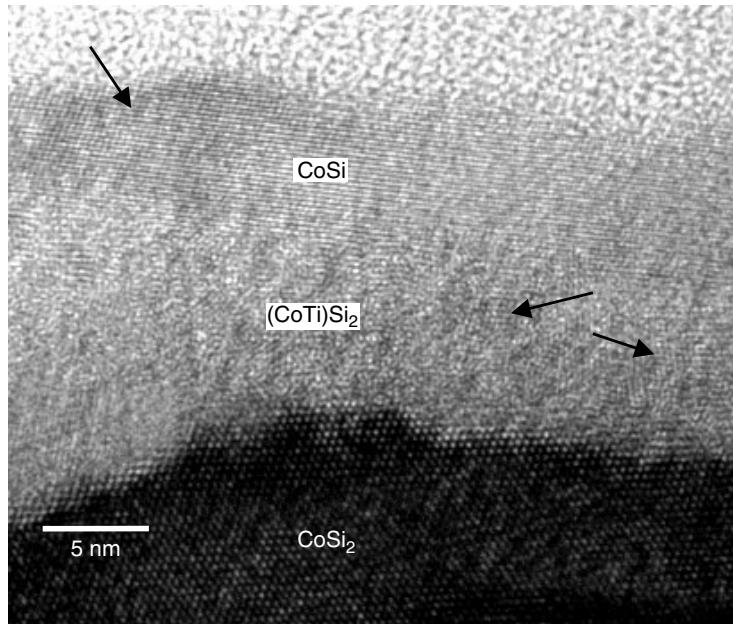


Figure 7.32 Sample *c* where both the top layer CoSi and the intermittent (CoTi)Si₂ are crystalline with nano-grains, as shown in this HRTEM image. (Sample courtesy Prof. Kin Leong Pey, NTU, Singapore)

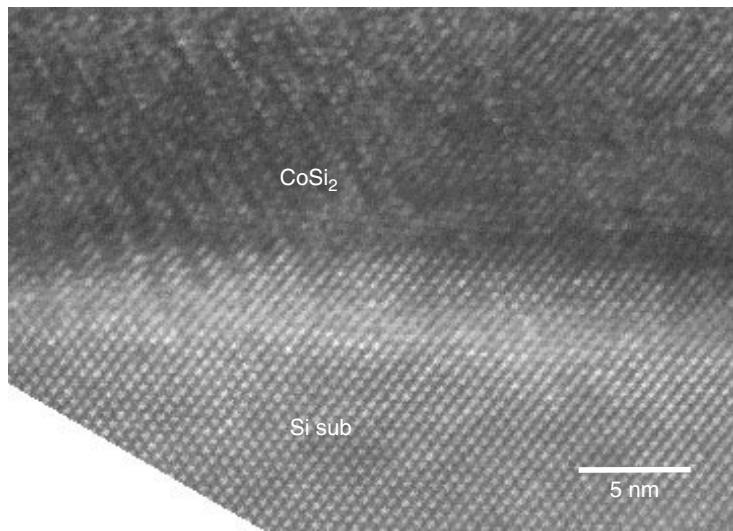


Figure 7.33 Sample *d* where the epitaxial CoSi₂ on Si substrate was formed at a temperature as low as 900°C. (Sample courtesy Prof. Kin Leong Pey, NTU, Singapore)

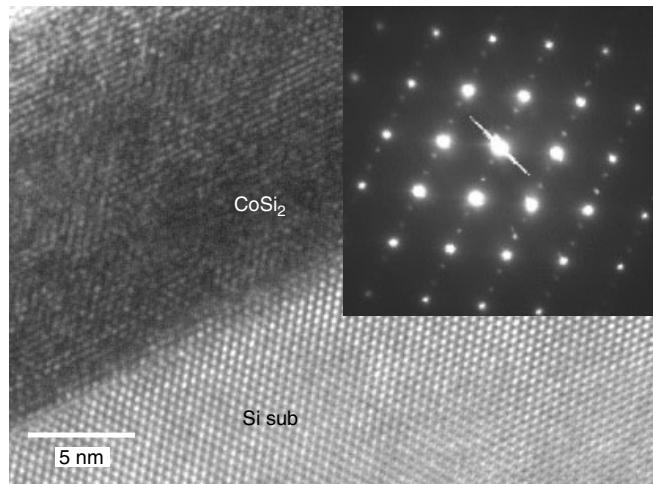


Figure 7.34 Sample *d* where with both the HRTEM image and electron diffraction the epitaxial relationship between CoSi₂ and Si substrate can be determined exactly. The inserted diffraction pattern shows twin spots ($\frac{1}{3}$ and $\frac{2}{3}$) in the CoSi₂. (Sample courtesy Prof. Kin Leong Pey, NTU, Singapore)

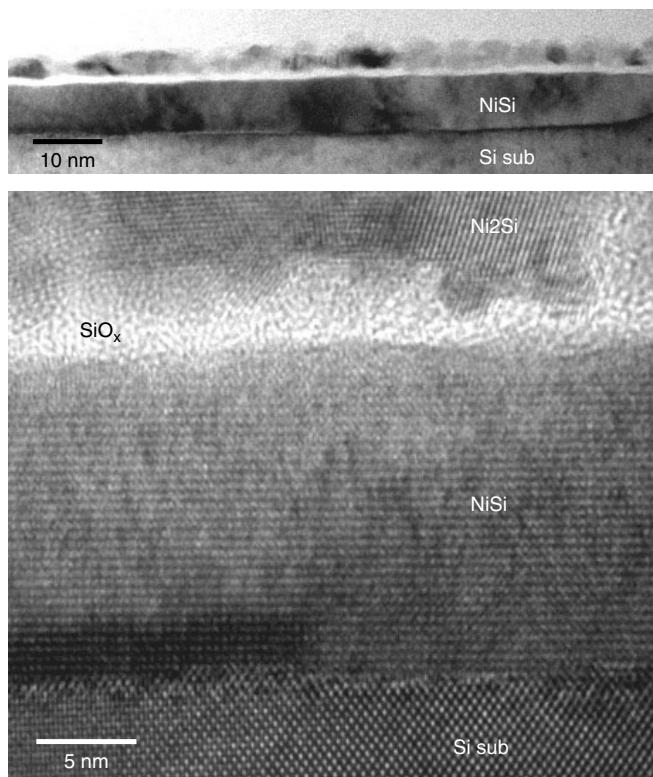


Figure 7.35 It is possible to form homogeneous and near-epitaxy NiSi monosilicide on the Si substrate. (Sample courtesy Prof. Kin Leong Pey, NTU, Singapore)

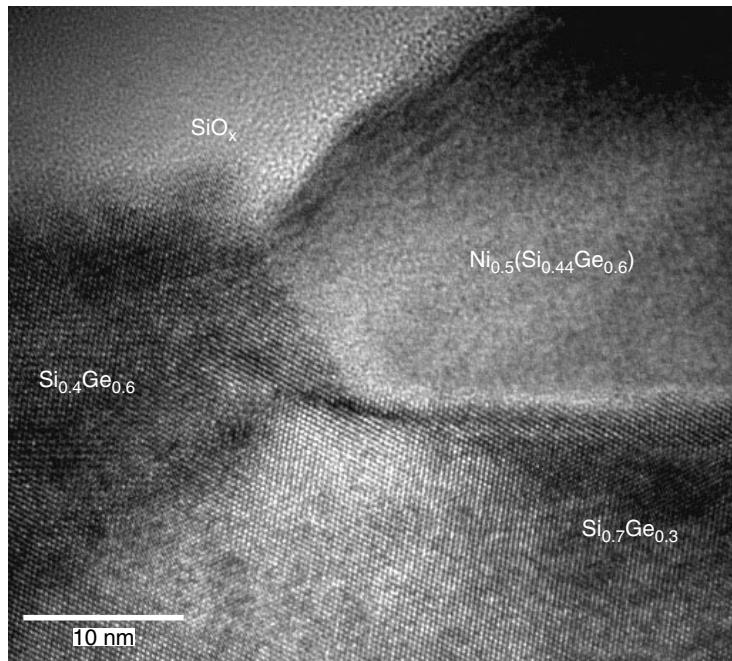


Figure 7.36 Ni forming silicide with $\text{Si}_{0.7}\text{Ge}_{0.3}$. The Ni forms silicide preferably with Si, and Ge is segregated to form high Ge concentrated SiGe grains in between the silicide islands. (Sample courtesy Prof. Kin Leong Pey, NTU, Singapore)

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.

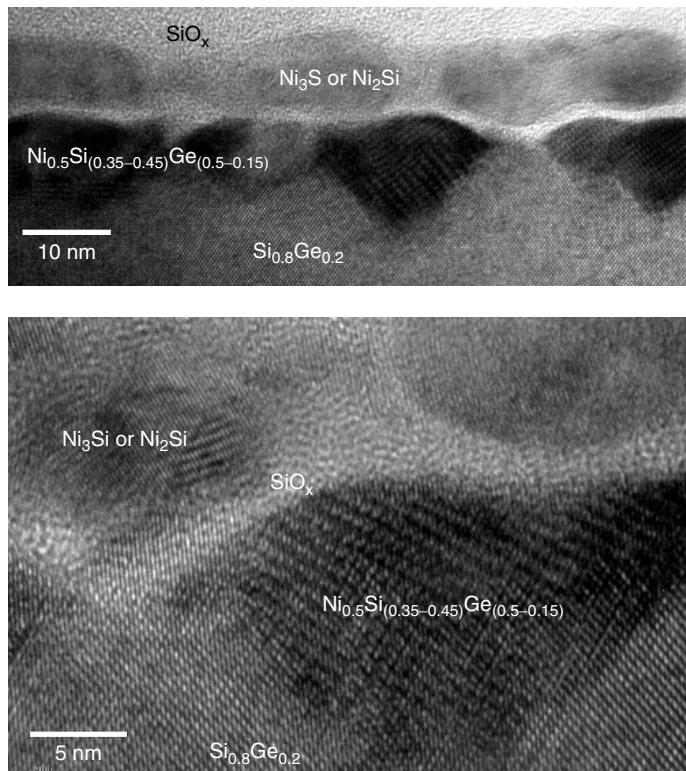


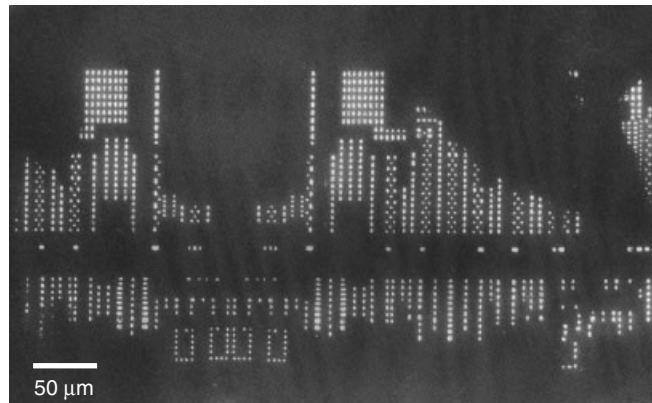
Figure 7.37 When the $\text{Si}_{0.8}\text{Ge}_{0.2}$ substrate is used, the silicidation reaction is different. The $\text{Ni}(\text{SiGe})$ monosilicide composition is found to be variable with the Ge content ranging from 5% to 15%. No major Ge segregation was found in this case. (Sample courtesy Prof. 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. Electro-chemi. 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

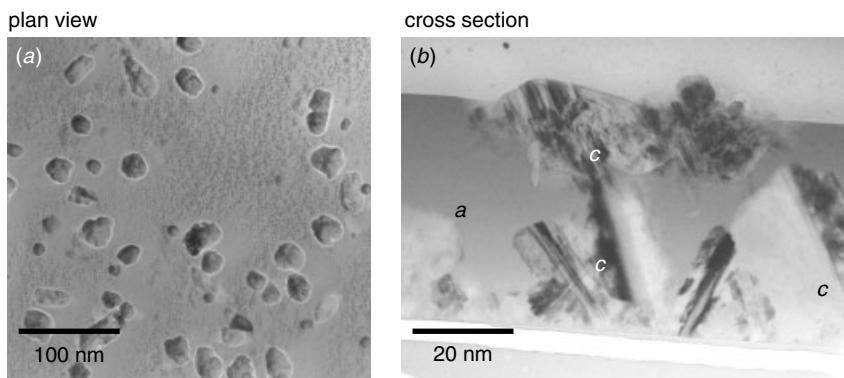


Figure 8.1 Amorphous silicon film recrystallizes randomly from film/substrate interface and from film surface. Location *c*, crystalline area; location *a*, amorphous areas. The crystalline area nucleated from film surface has introduced a surface bump.

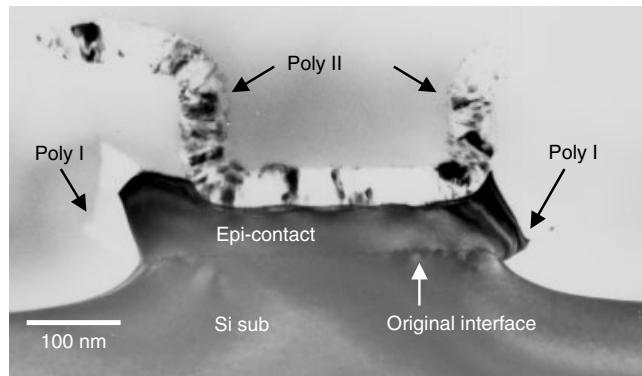


Figure 8.2 A good realigned polysilicon contact. The epitaxial realignment, as indicated, was by the poly I contact. Poly I was then contacted by another polycontact, poly II, on top.

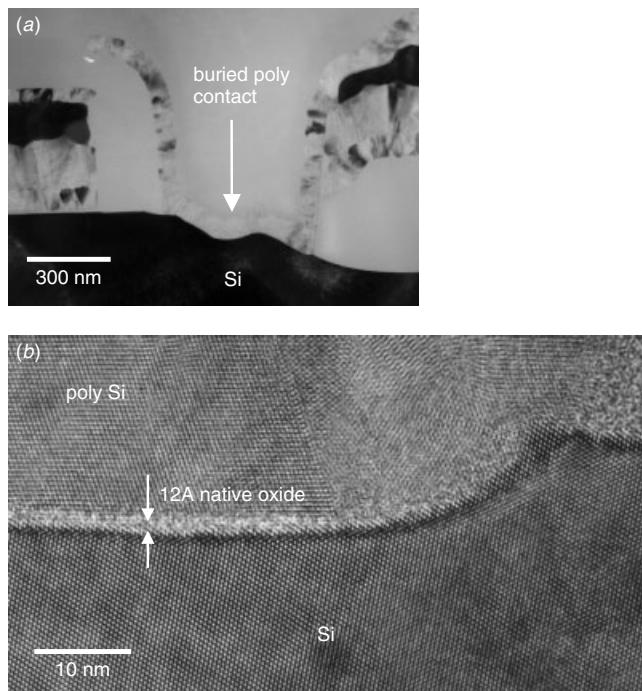


Figure 8.3 16M SRAM process technology with the problem of buried polysilicon contact high resistivity. The low-magnification TEM image shows there to be no epitaxial realignment. The HRTEM image shows the interface native oxide.

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

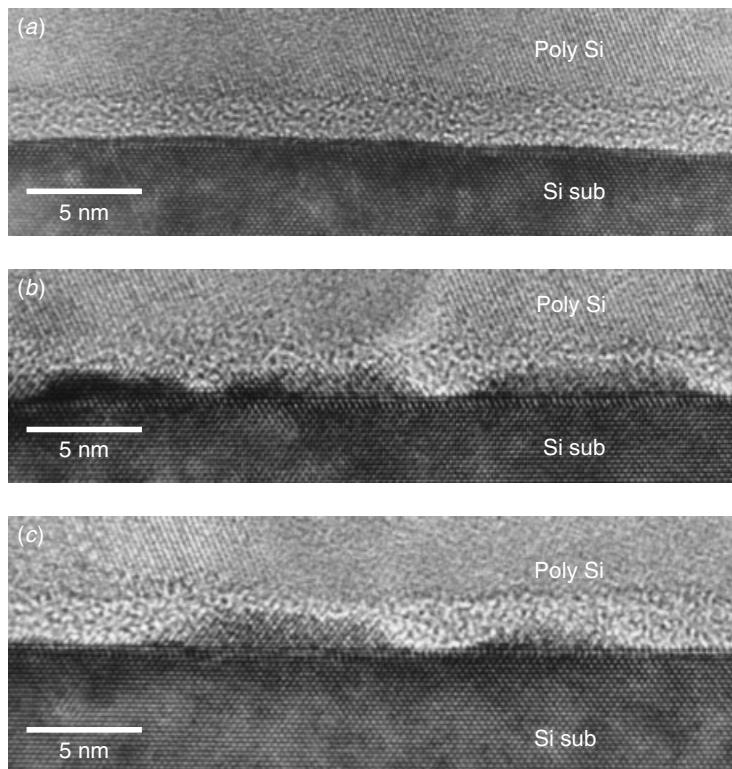


Figure 8.4 HRTEM image of the interface (a) without and (b) and (c) with realignment islands. (*J. Appl. Phys.*, **73** (5), 2543–2547, 1993 reprint with permission from AIP)

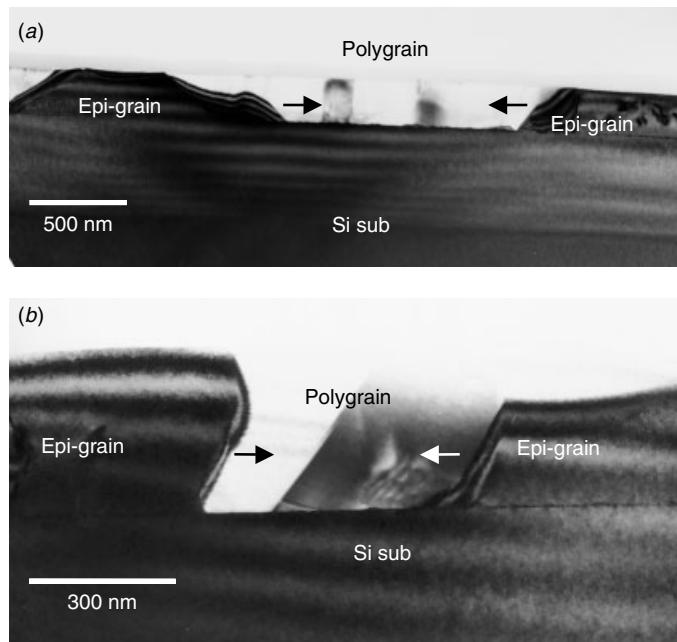


Figure 8.5 Epitaxy via lateral grain growth. The epitaxial grains spread, as indicated, and consume the polysilicon film.

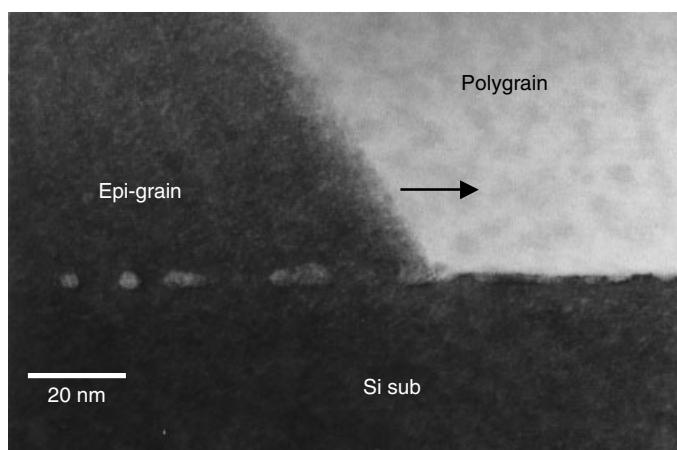


Figure 8.6 Epitaxy via lateral grain growth. Residue oxide particles can be seen in between Epi-grain and Si-sub.

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

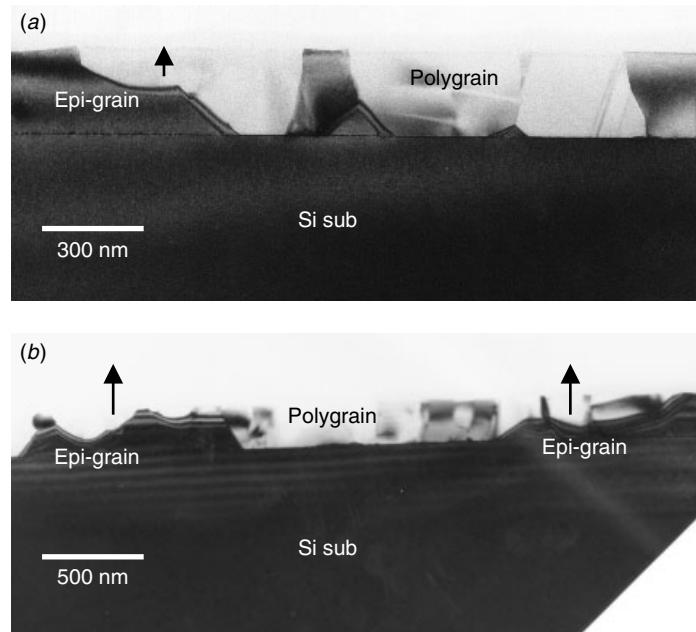


Figure 8.7 Epitaxy via vertical motion of the Si(100)-polysilicon interface toward the surface.

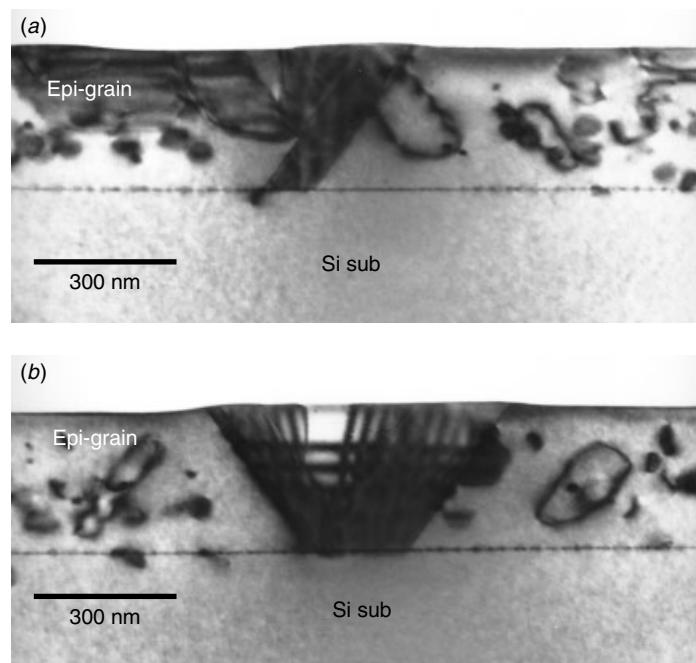


Figure 8.8 Epitaxial Si film with dislocation loops and stacking faults.

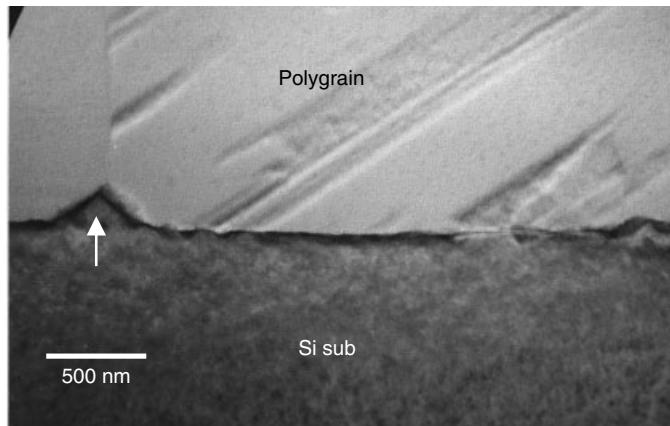


Figure 8.9 Epitaxial Si pyramid island as it extends from the interface. In this case the pyramid island started from a polysilicon grain boundary and gradually swept upward.

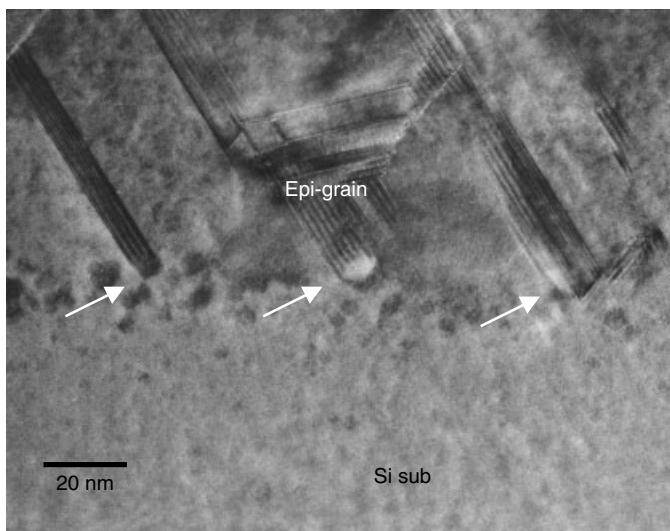


Figure 8.10 Epitaxial Si film with defects associated with the residue oxide particles.

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

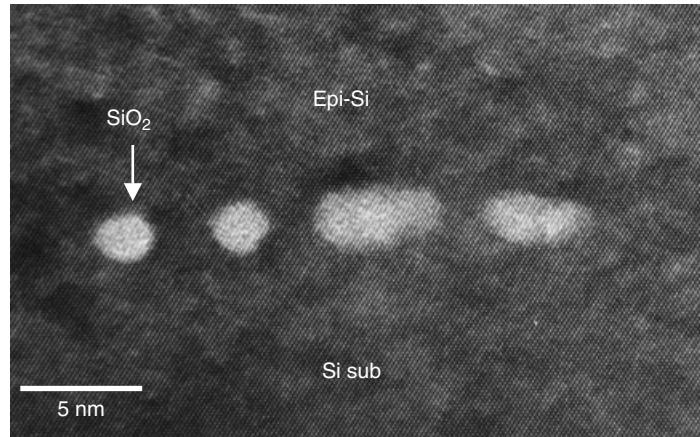


Figure 8.11 Epitaxial Si film with SiO_2 particles left at the original polysilicon-Si substrate interface.

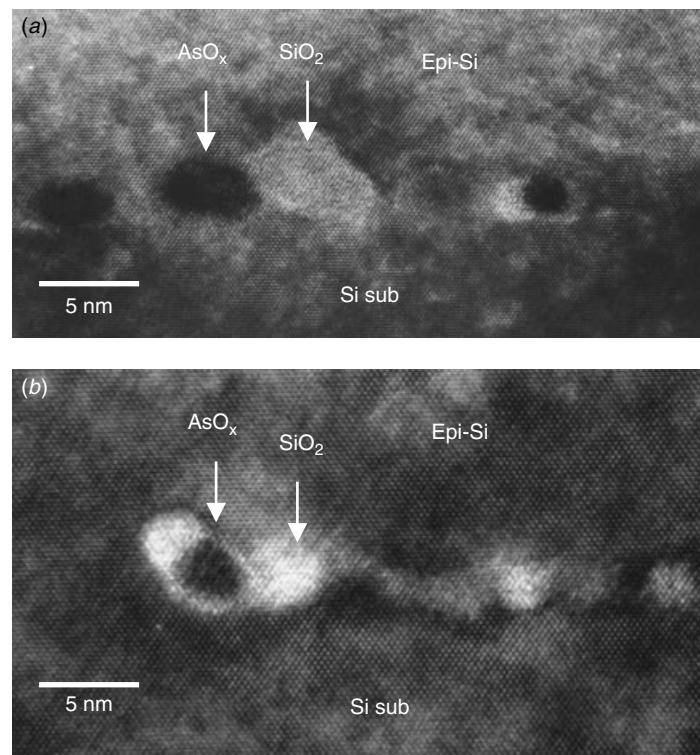


Figure 8.12 Epitaxial Si film with SiO_2 particles (white beads) co-precipitate with AsO_x (dark beads).

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.

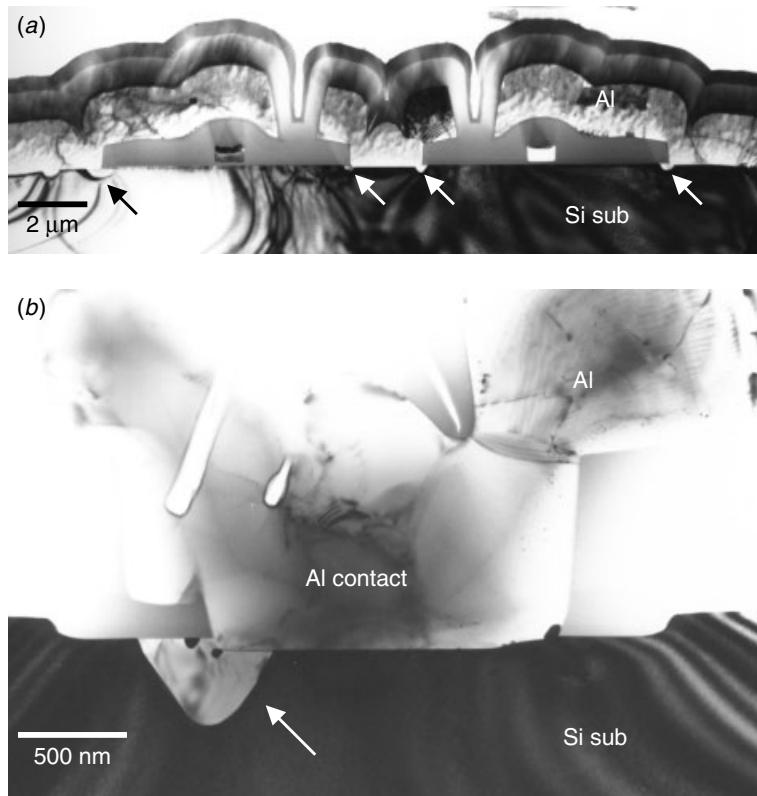


Figure 8.13 A power LDMOS device with an Al spiking problem. The Al penetrated into Si substrate with faceted spikes, as indicated. Such a problem normally occurs at or near the corner areas.

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

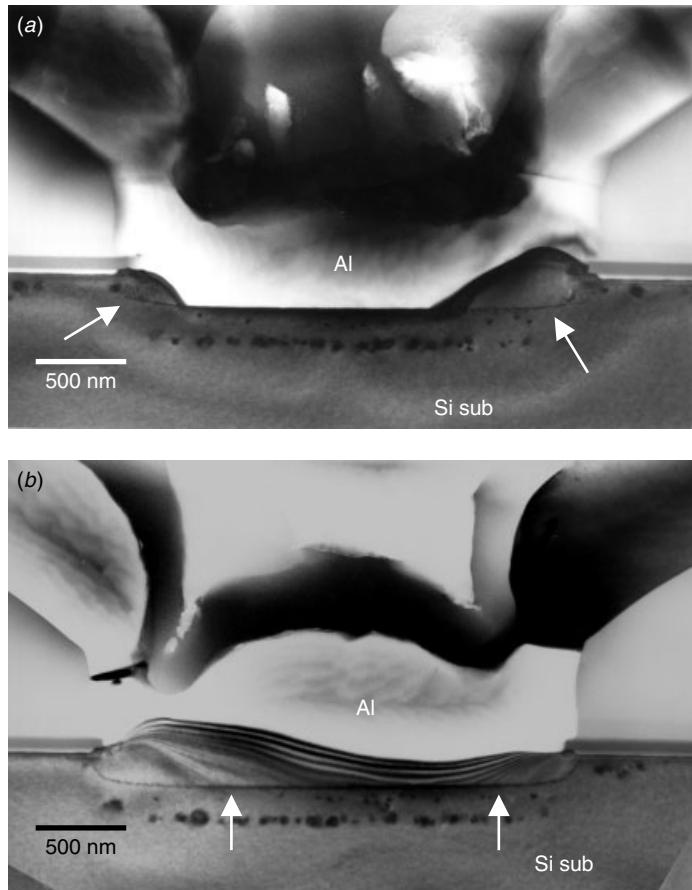


Figure 8.14 (a) Si nodules precipitate at the corner of contacts and form an epitaxy with the Si substrate, as indicated. (b) Si forms an epitaxy with the Si substrate and covers the whole contact area. This entirely changes the contact's electrical characteristics.

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

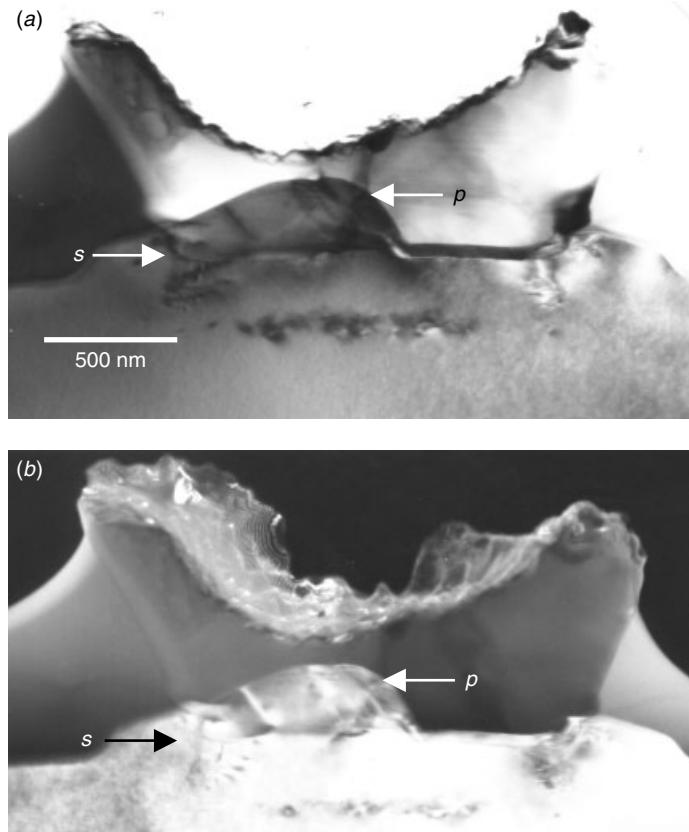


Figure 8.15 Si nodules precipitate (denoted *p*) at the contact corner where a spiking (denoted *s*) also occurs. Both (a) bright and (b) dark fields are shown to demonstrate the spiking and precipitation clearly.

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.

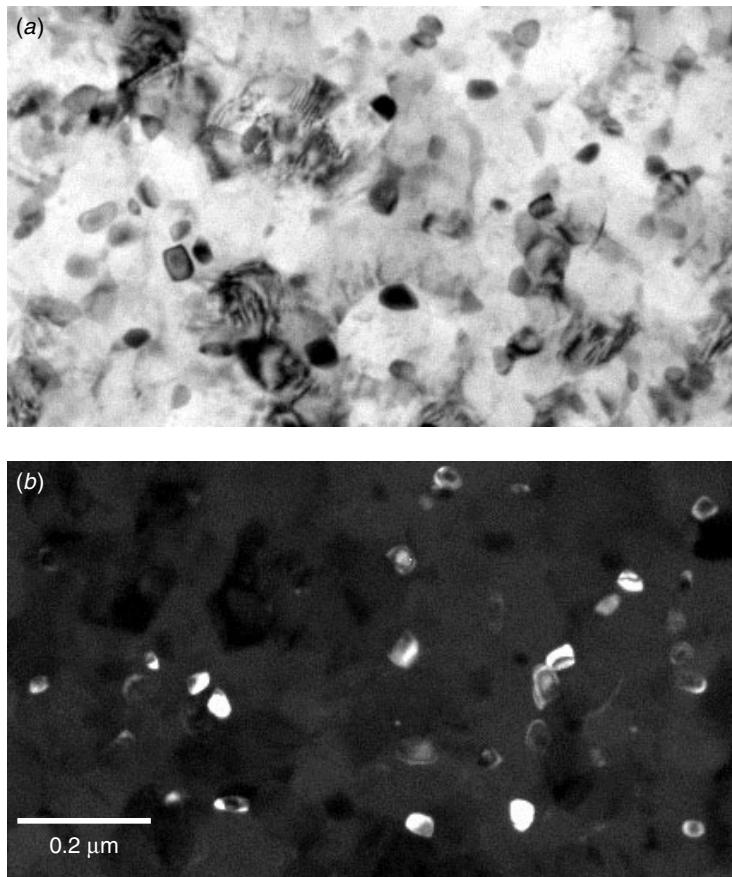


Figure 8.16 Al_2Cu (θ phase) precipitation seen in TEM plan view. (a) In the bright field image the particles appear to be dark, and (b) in the dark field, they appear to be bright.

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

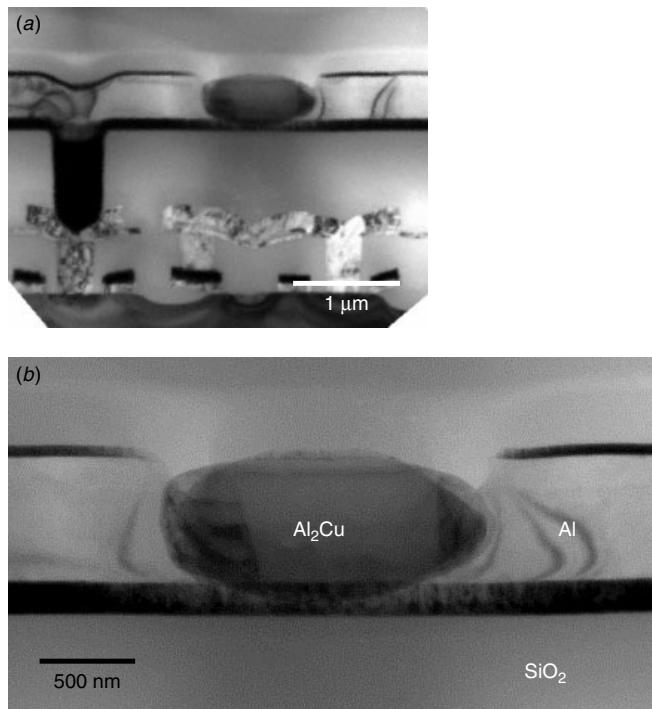


Figure 8.17 Al₂Cu (θ phase) precipitation can be larger than the Al film thickness, as shown here in a real device structure.

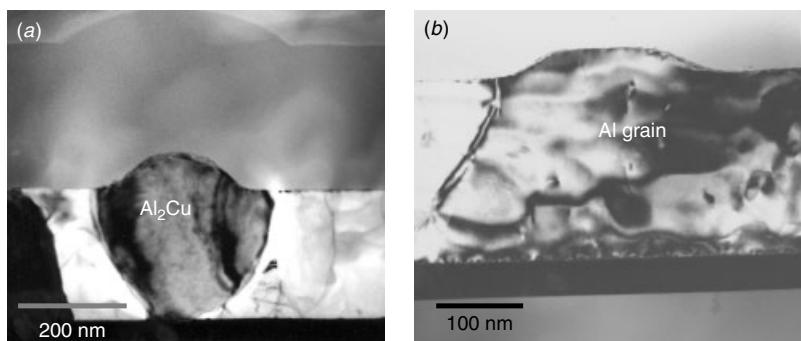


Figure 8.18 TEM analysis shows hillock formations, (a) some due to the Al₂Cu θ phase and (b) some due to the pure stress induced Al extrusion.

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 SiO₂ 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

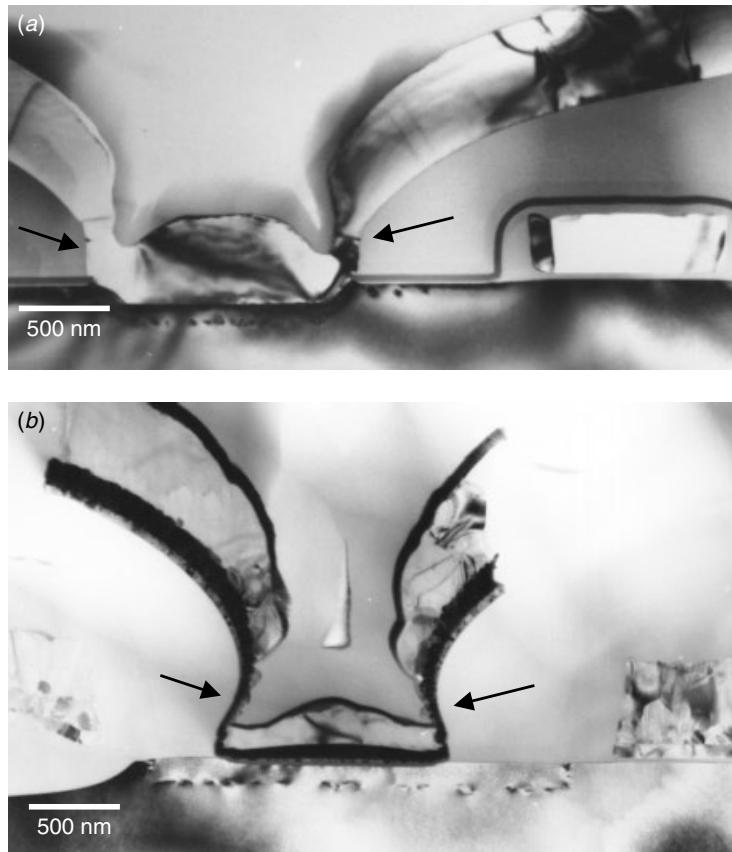


Figure 8.19 Al line step coverage issue in different generations of processes. (a) The Al line cross-sectional area is reduced to about 30% of the original thickness. (b) Al has totally disappeared in the area indicated, and only the barrier layer is left.

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.

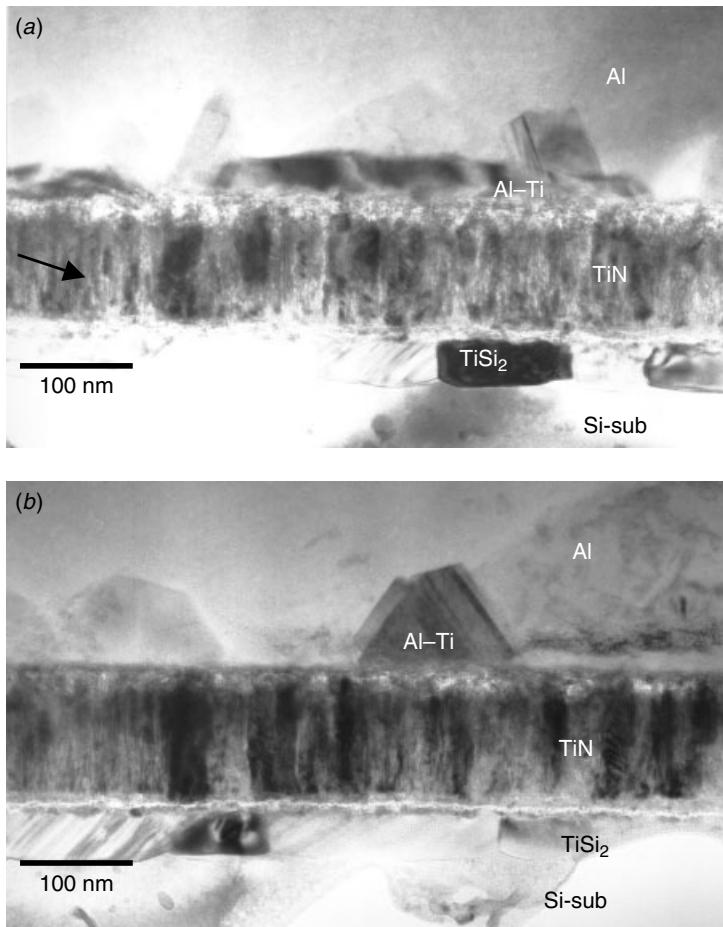


Figure 8.20 Al–Ti intermetallic formed at the Al/TiN interface during the hot Al processes. When nucleated, the Al–Ti intermetallic forms a coherent and faceted interface with the Al, as shown here.

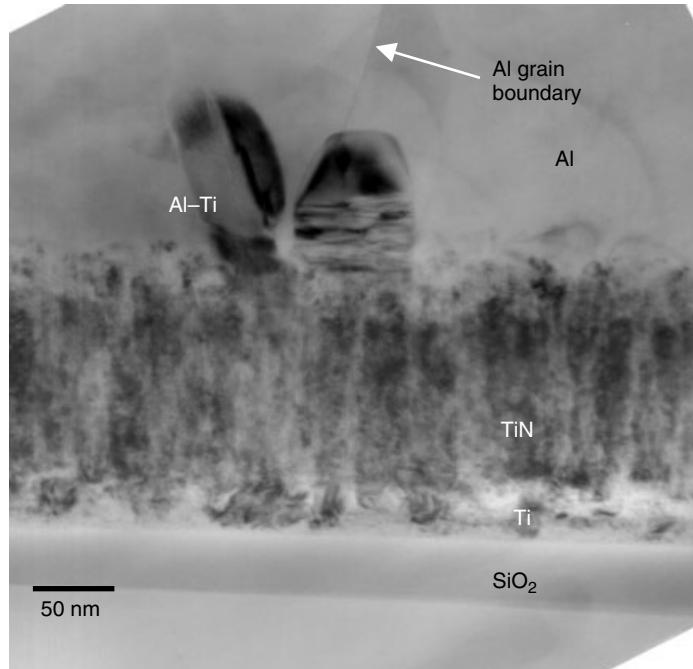


Figure 8.21 $\text{Al}-\text{Ti}$ intermetallic formed at the Al/TiN interface to the Al grain boundary intersection.

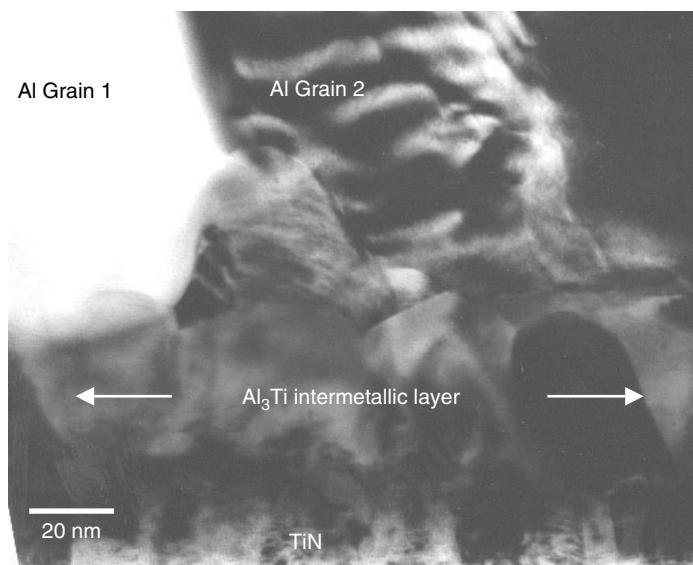


Figure 8.22 $\text{Al}-\text{Ti}$ intermetallic formed at the Al/TiN interface to Al grain boundary intersection.

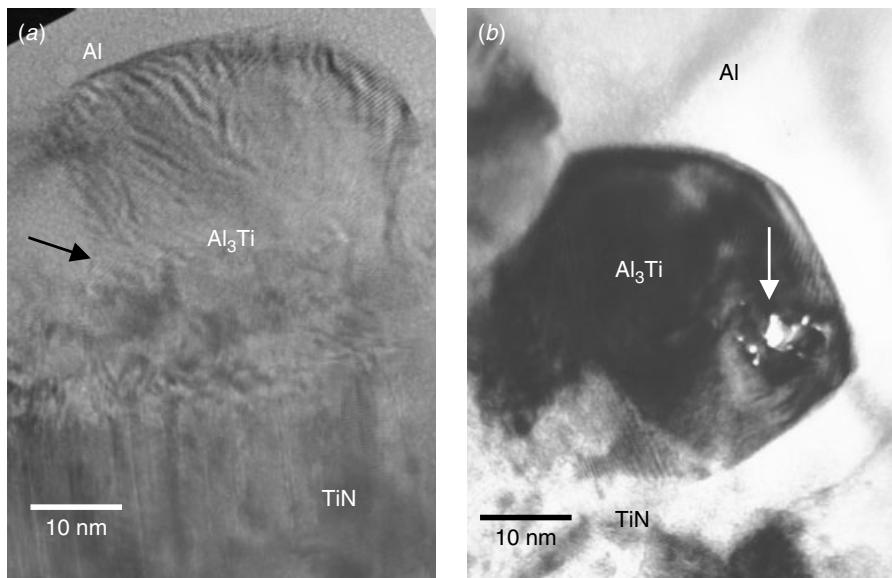


Figure 8.23 Al–Ti intermetallic formed at the Al/TiN interface to Al grain boundary intersection. Electron diffraction and nanoprobe EDS show the Al–Ti intermetallic formed at Al/TiN interface to be Al₃Ti. Note here that the high-energy electron beam induced local melting has drilled a hole in the Al₃Ti intermetallic, as indicated.

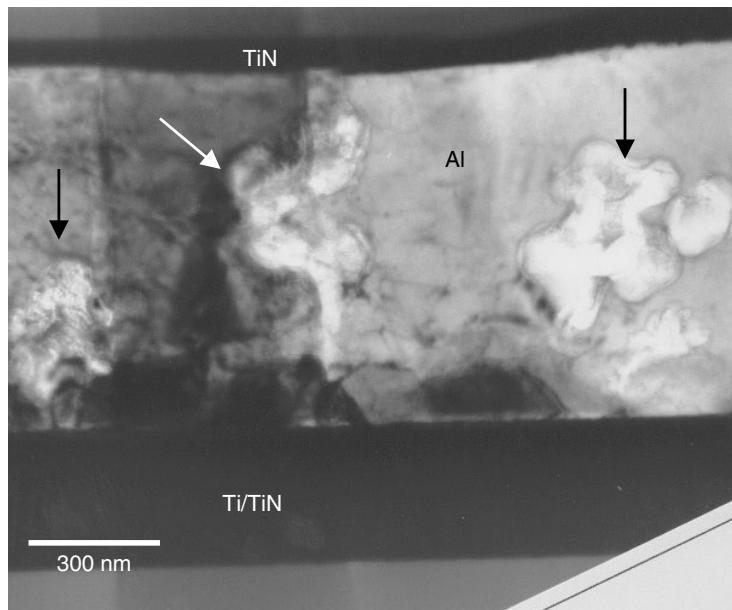


Figure 8.24 Local corrosion pits within the Al metallization layer, as indicated.

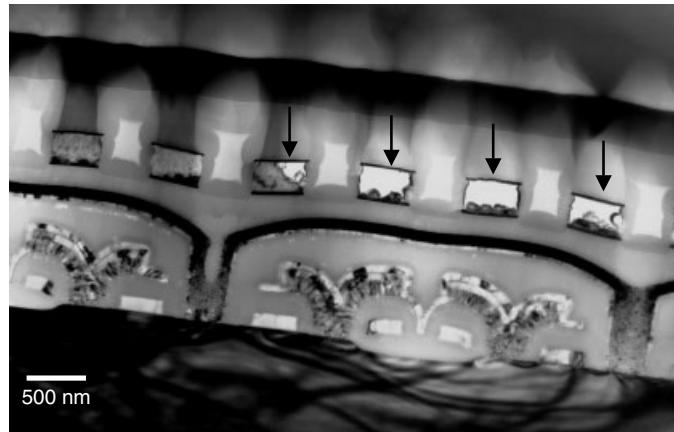


Figure 8.25 Local corrosion on the Al metallization layer, as indicated. In some instances the Al layer has been entirely consumed.

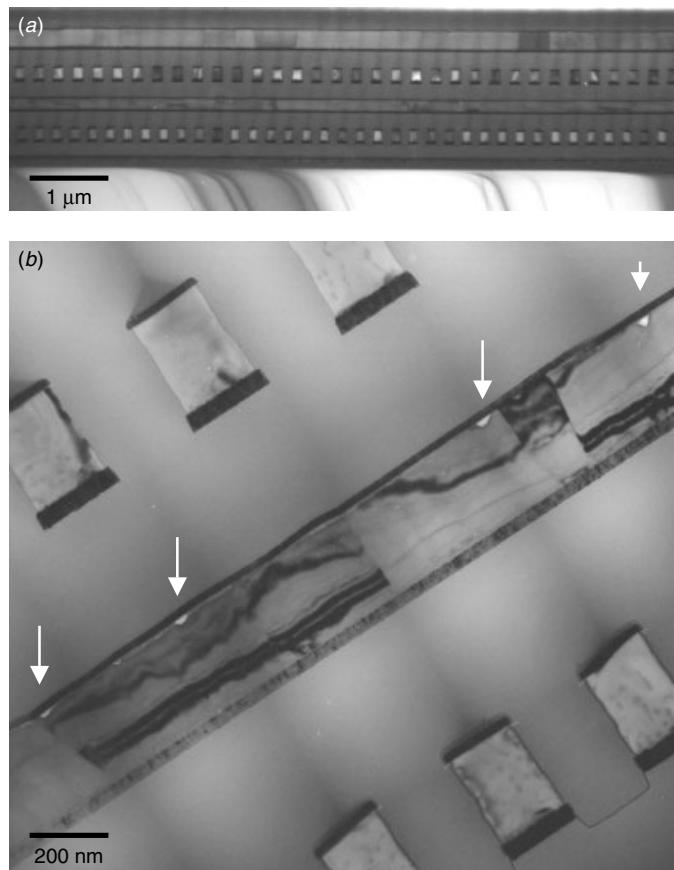


Figure 8.26 Detail of the corrosion starts from the upper interface between TiN and Al, as indicated by arrows.

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.

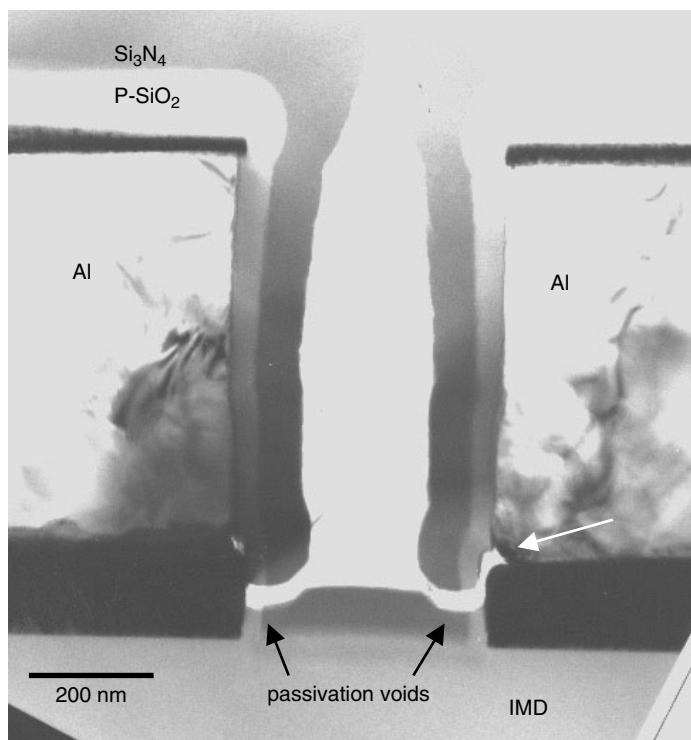


Figure 8.27 Passivation voids near the Al line bottom have exposed the Al to the environment and thus make the device vulnerable to corrosion.

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.

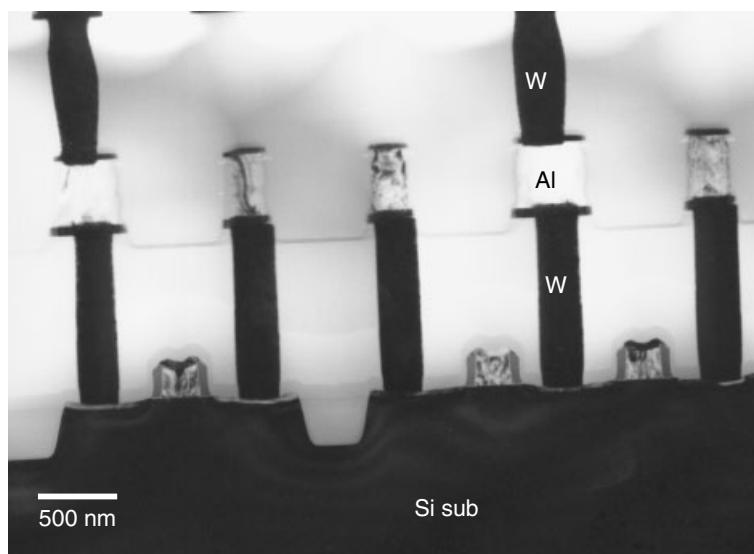


Figure 8.28 High aspect ratio contacts. The contact height is 1.2 μm with an aspect ratio of 4.

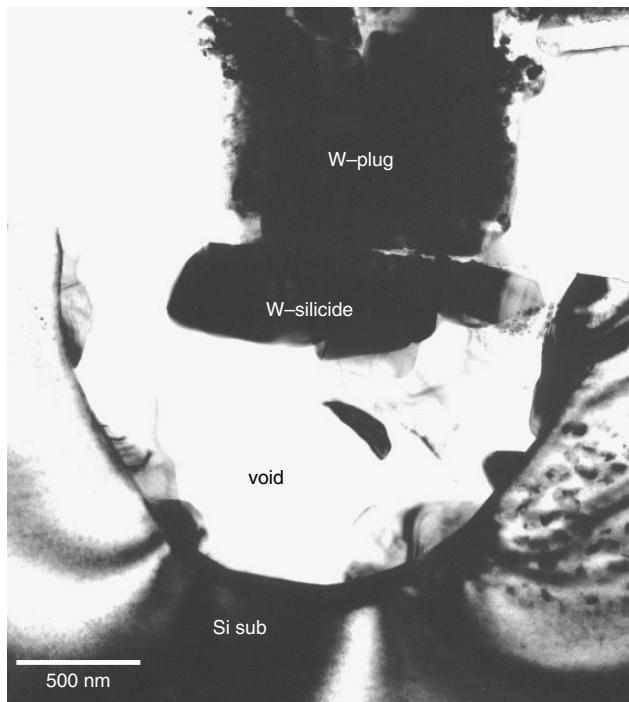


Figure 8.29 Tungsten–plug contact reacts with the Si substrate, forming W–silicide and consuming the whole substrate junction structure.

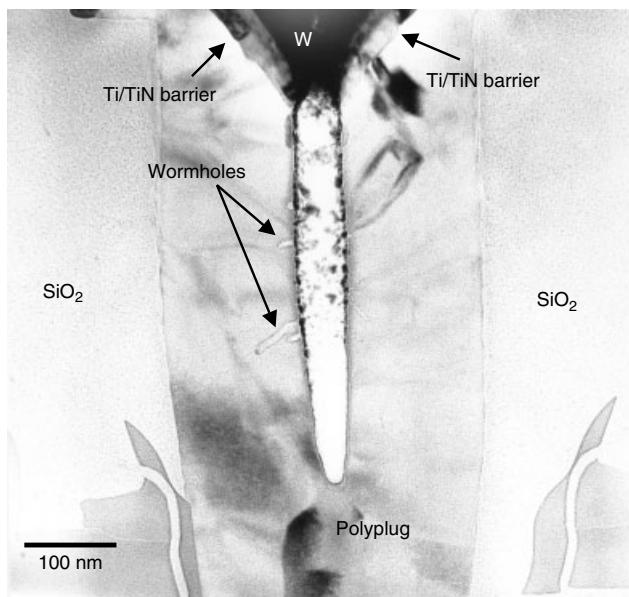


Figure 8.30 Tungsten–plug contact into polysilicon plug. As the barrier Ti/TiN did not get into the central gap within polyplug, W penetrates into the seam gap, reacts with the poly, and forms the wormholes (tunnels) at the sidewall of the polyplug, as indicated.

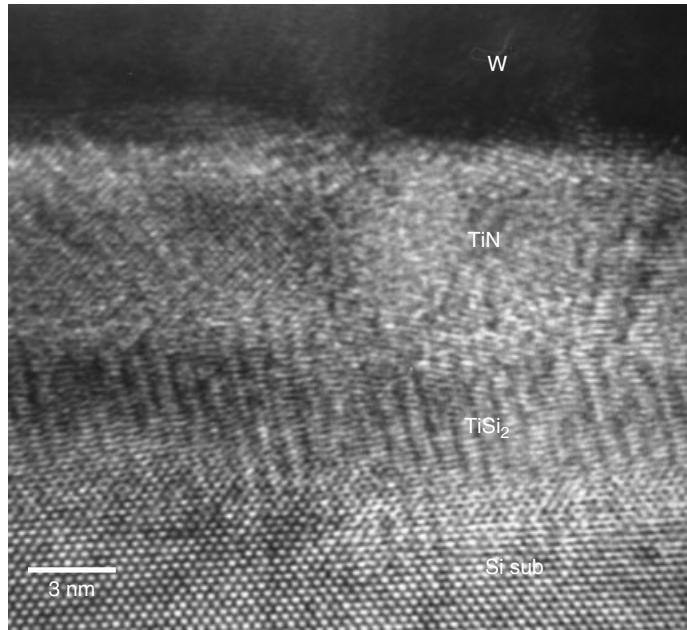


Figure 8.31 HRTEM image of the bottom of a W–plug contact. The layers observed are Si substrate, TiSi₂, TiN, and W.

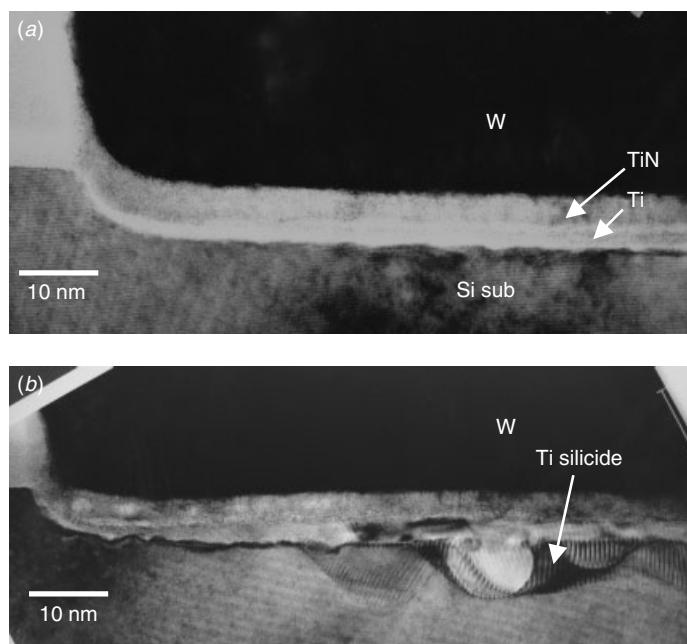


Figure 8.32 The bottom of W–plug contacts. (a) No Ti silicide formed at the interface, and the contact resistivity is high. (b) TiSi₂ crystalline phase formed in partial areas, and the contact resistivity is normal and low.

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

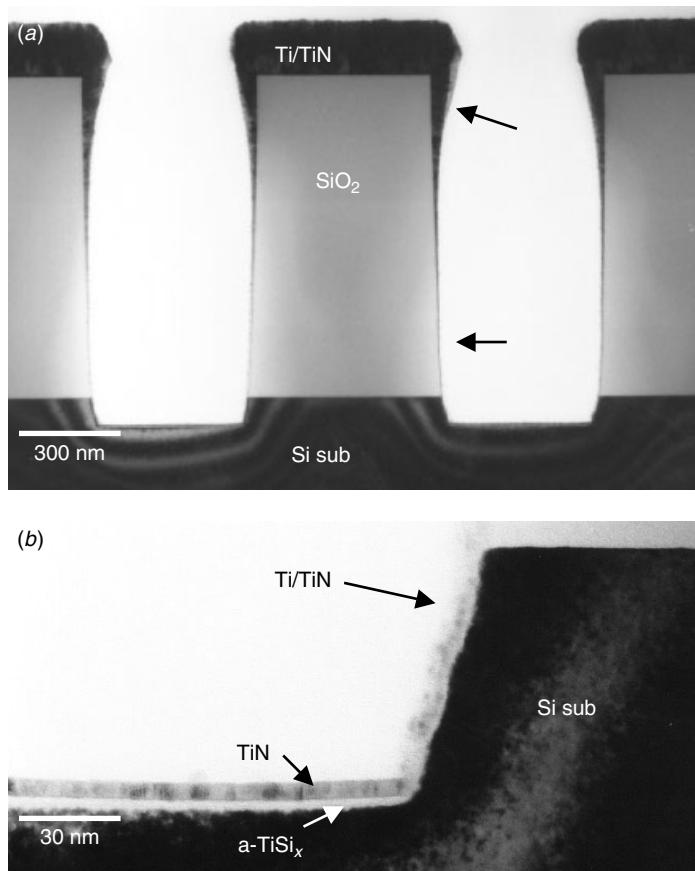


Figure 8.33 The W–plug contacts without W filling. (a) Step coverage of Ti/TiN layers is clearly observed. Ti/TiN thickness has diminished quickly to nearly zero at the contact’s sidewall, as indicated by arrows, in particular, at the contact’s bottom, as shown in (b)

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

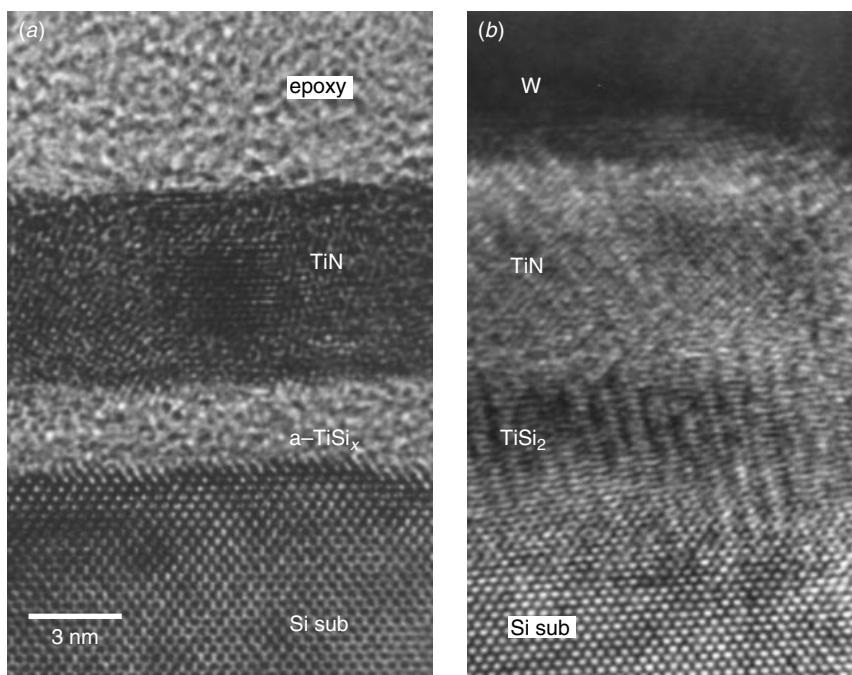


Figure 8.34 HRTEM image of the bottom of W-plug contacts, before and after W deposition and annealing. a- TiSi_x is observed between the TiN and Si sub, which has become crystalline TiSi_2 after the annealing.

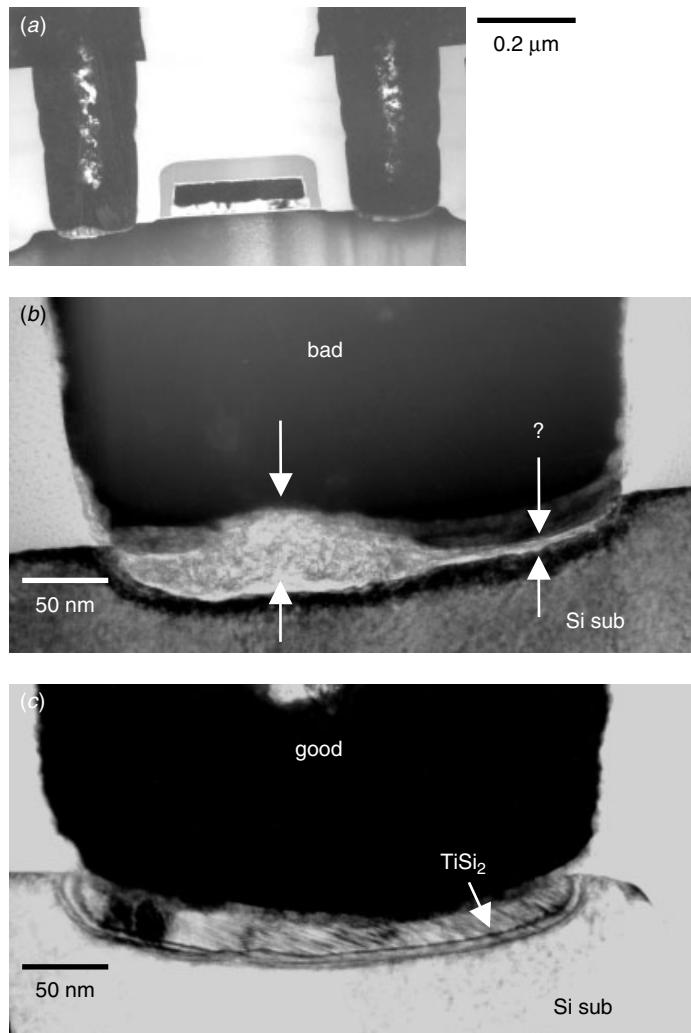


Figure 8.35 TEM cross section of a transistor with one electrically confirmed good contact (right-hand side) and one bad contact (left-hand side). Close-up images show that the bad contact has a thin foreign layer between Ti/TiN and Si sub, while the good contact has formed a thick TiSi₂ crystalline 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.

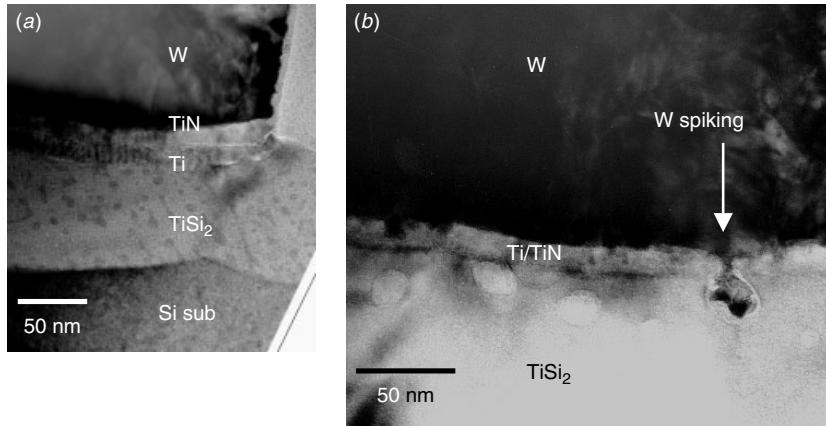


Figure 8.36 TEM cross section of one W–plug with Ti/TiN barrier layer and salicide contact. No further silicidation is necessary, since the substrate surface already has a thick layer of silicide. However, Ti/TiN is still necessary, since contact spiking can still occur if the Ti/TiN was not done properly, as indicated here)

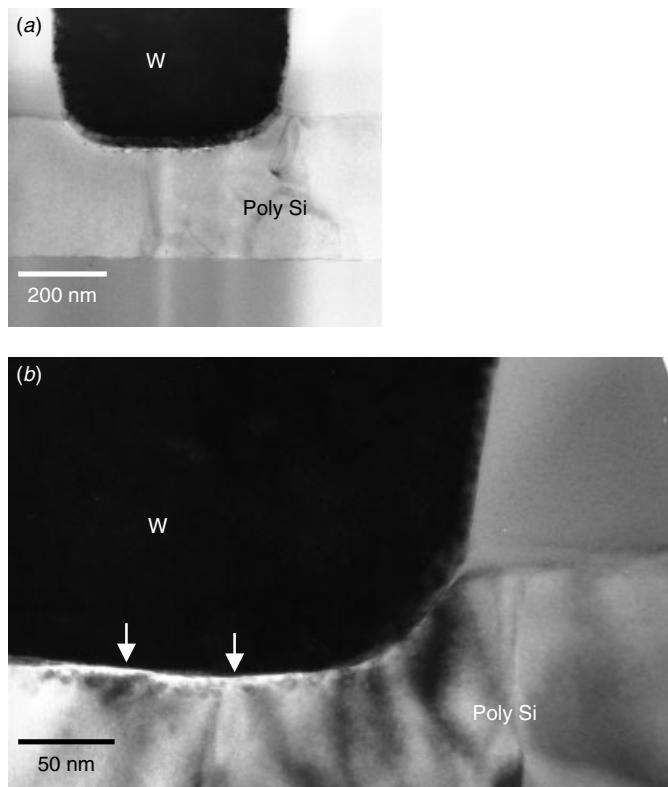


Figure 8.37 TEM cross section of one W–plug contact to polysilicon runner (conductor). Detailed interface micrograph shows no silicidation to the polysilicon, as indicated. The contact has a high-resistivity issue.

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,

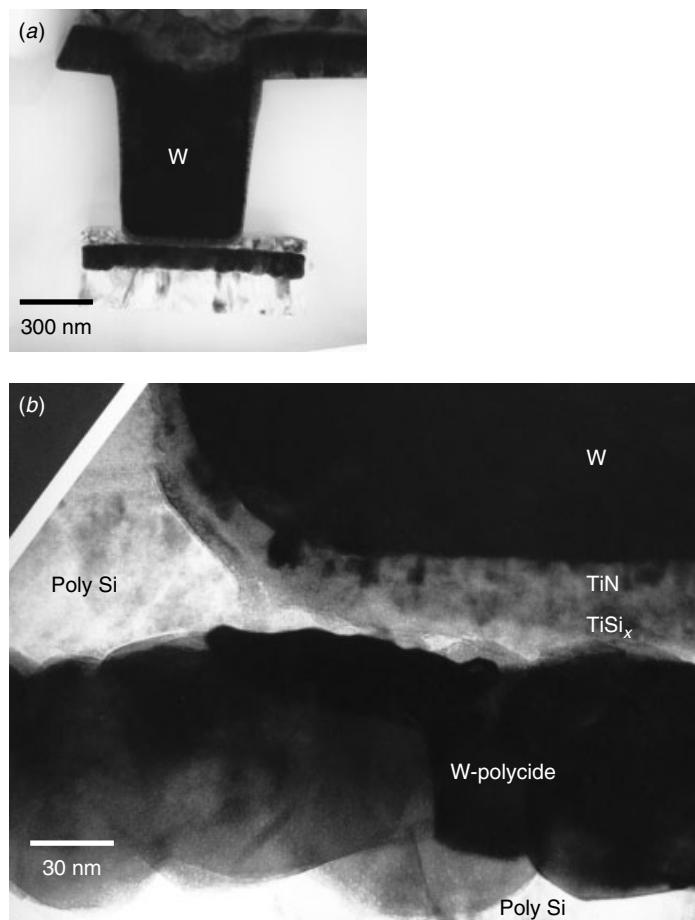


Figure 8.38 TEM cross section of one W–plug contact to the polycide runner with a poly on top. Detailed interface micrograph shows some Ti silicidation to the polysilicon at the interface.

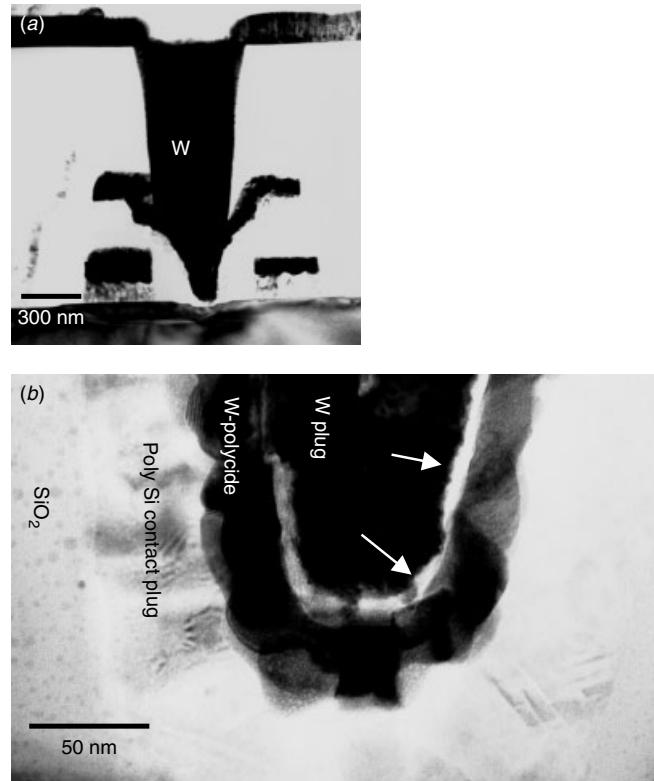


Figure 8.39 TEM cross section of one W–plug contact to a V-shaped polycide plug without surface poly layer. Detailed interface micrograph shows a thin gap between W and WSi_x , as indicated. No Ti silicidation occurs, and the contact resistivity is high.

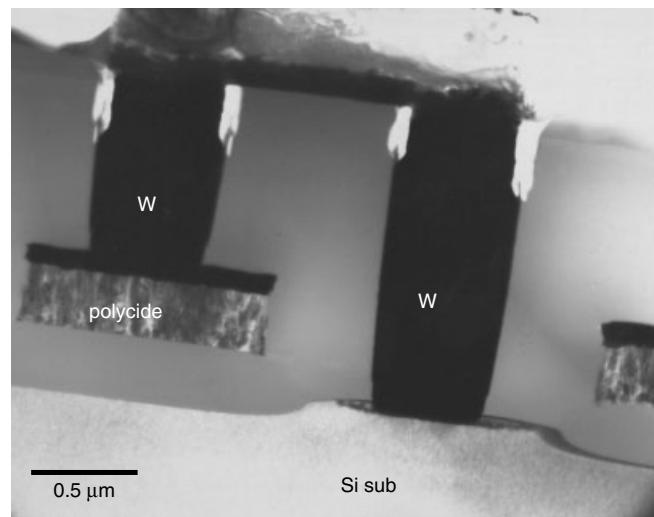


Figure 8.40 TEM cross section of W–plug contacts showing voids at the top corners due to the micro-loading effect.

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”

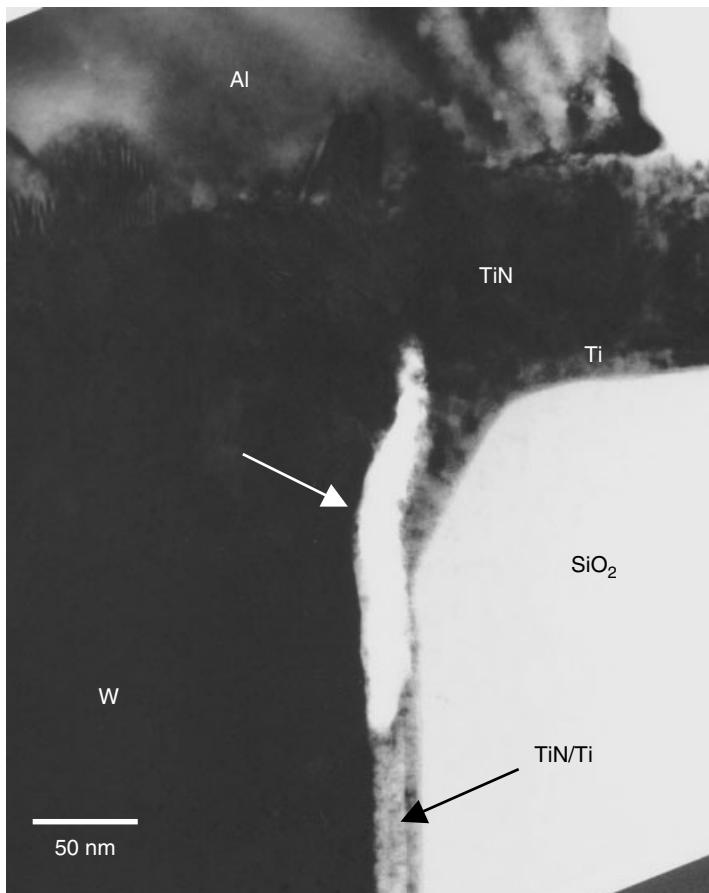


Figure 8.41 Close-up of the W–plug contacts at the top corner where the voids are embedded between the W and the TiN/Ti layer.

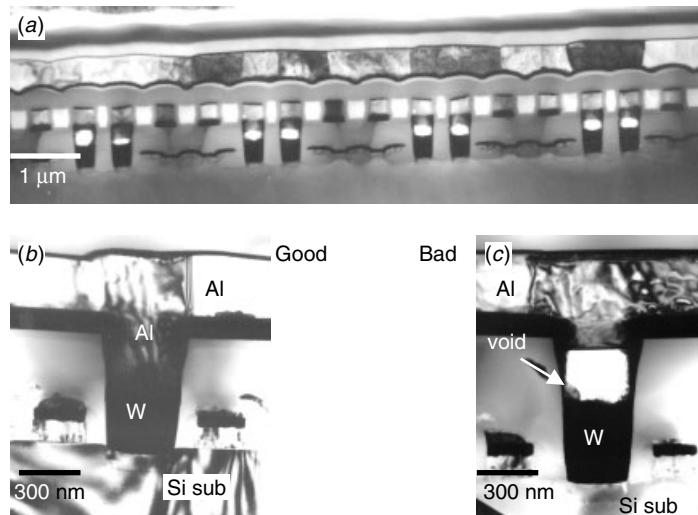


Figure 8.42 W–plug contacts with the Ti/TiN barrier layer. The voids formed between W and Al are due to W being over-etched back. Notice that the W only half-fills the contact holes for the electrically good contacts and the void form in the bad contacts.

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.

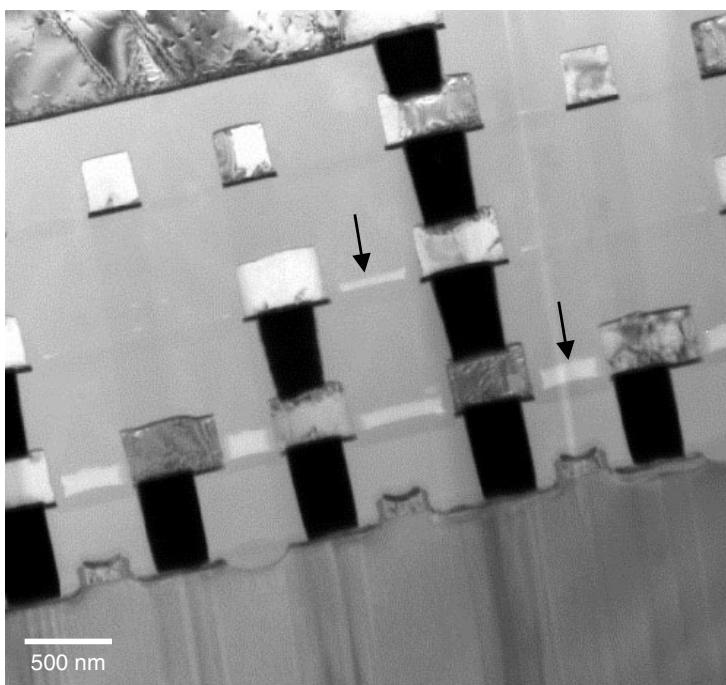


Figure 8.43 Multiple-level interconnection with SOG planarization seen in metal 1 and metal 2, as indicated.

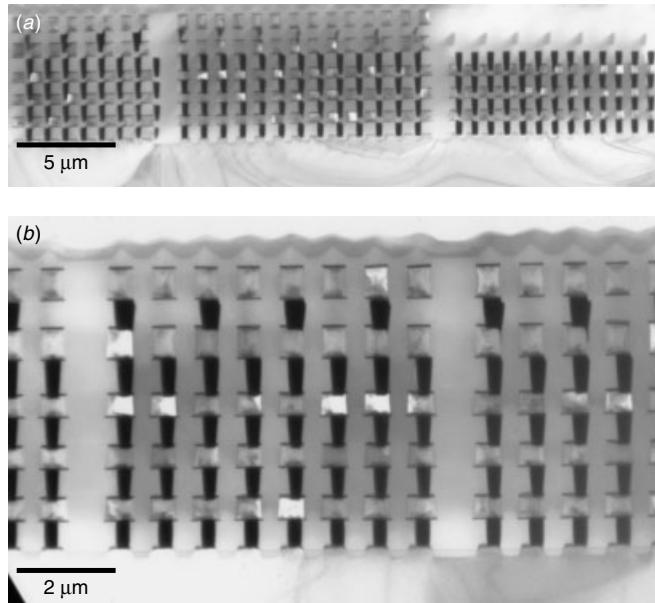


Figure 8.44 Multiple-level interconnect VIA chain test structure.

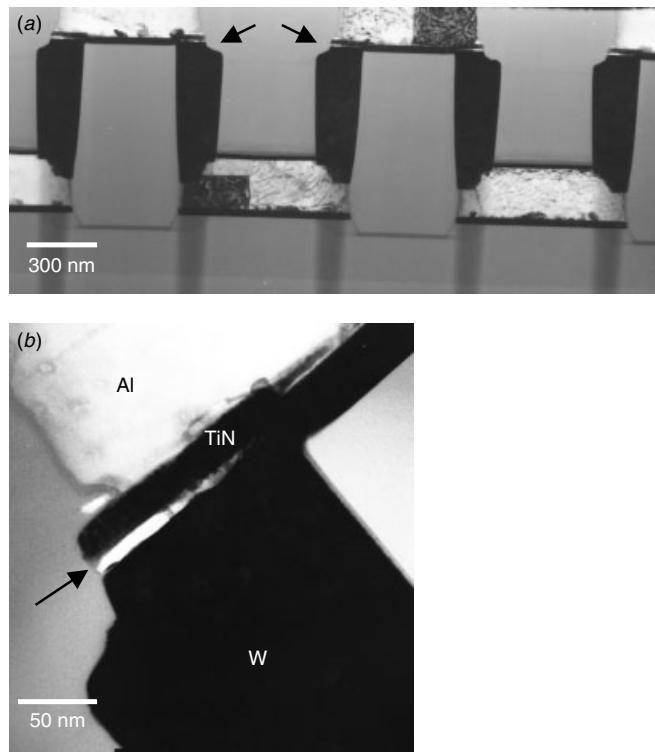


Figure 8.45 A VIA chain test structure shows the effect of the misalignment-induced W corner etch and the Ti-gap at the top corner of the VIAs. (*Materials Res. Soc. Symp. Proc.*, **564**, 451, 1999, reprint with permission from MRS)

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.

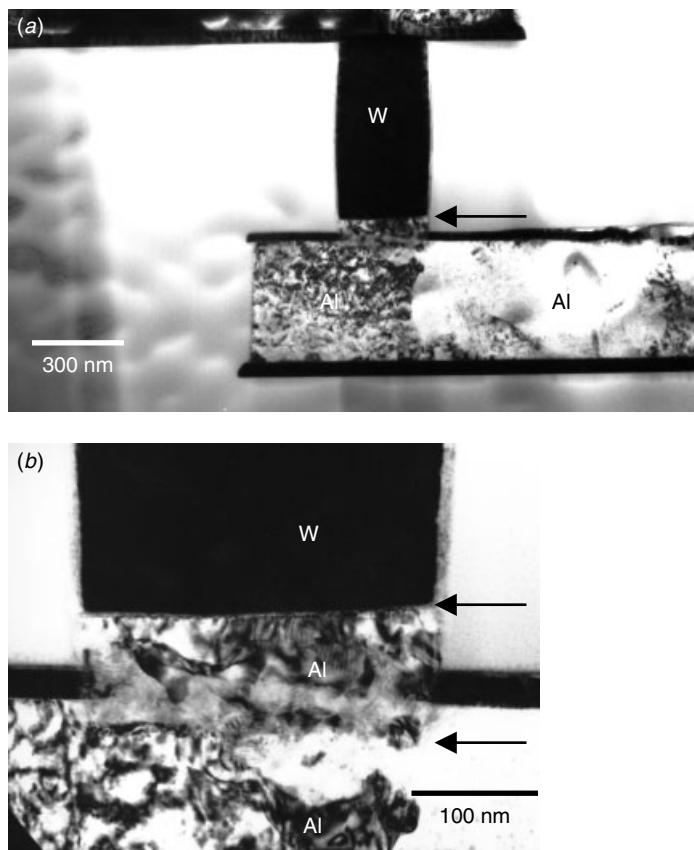


Figure 8.46 Al extruded into the VIA opening during W/TiN/Ti deposition, resulting in elevated VIA bottom above the lower metal line surface. The original VIA bottom and the elevated VIA bottom are indicated.

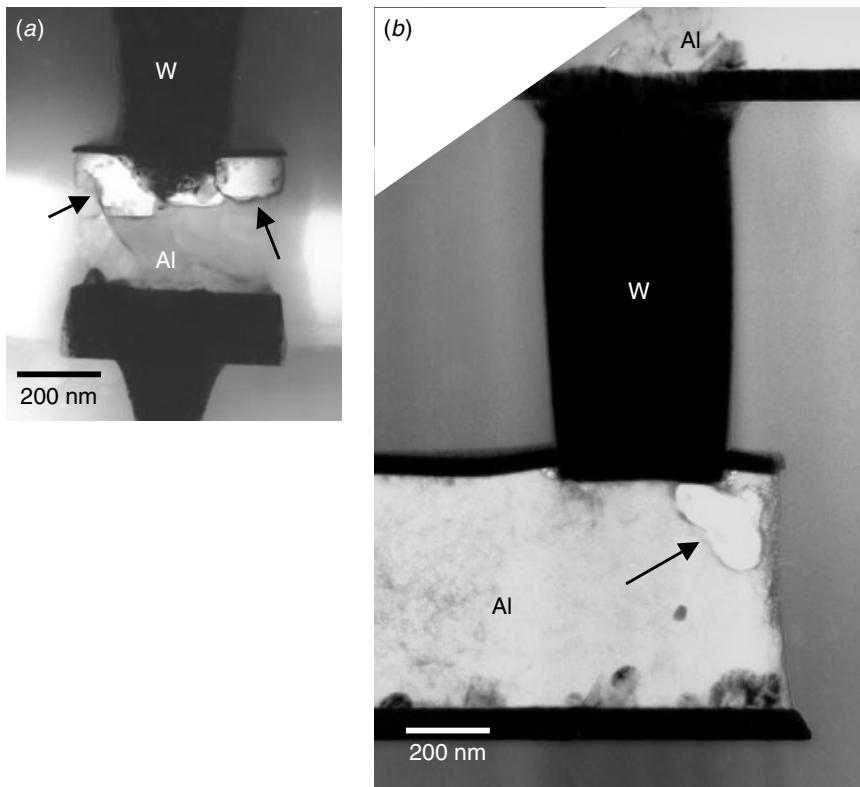


Figure 8.47 Al has shrunk and formed a void, as indicated, under the VIA, resulting in high VIA resistivity and even an open VIA (insert).

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:

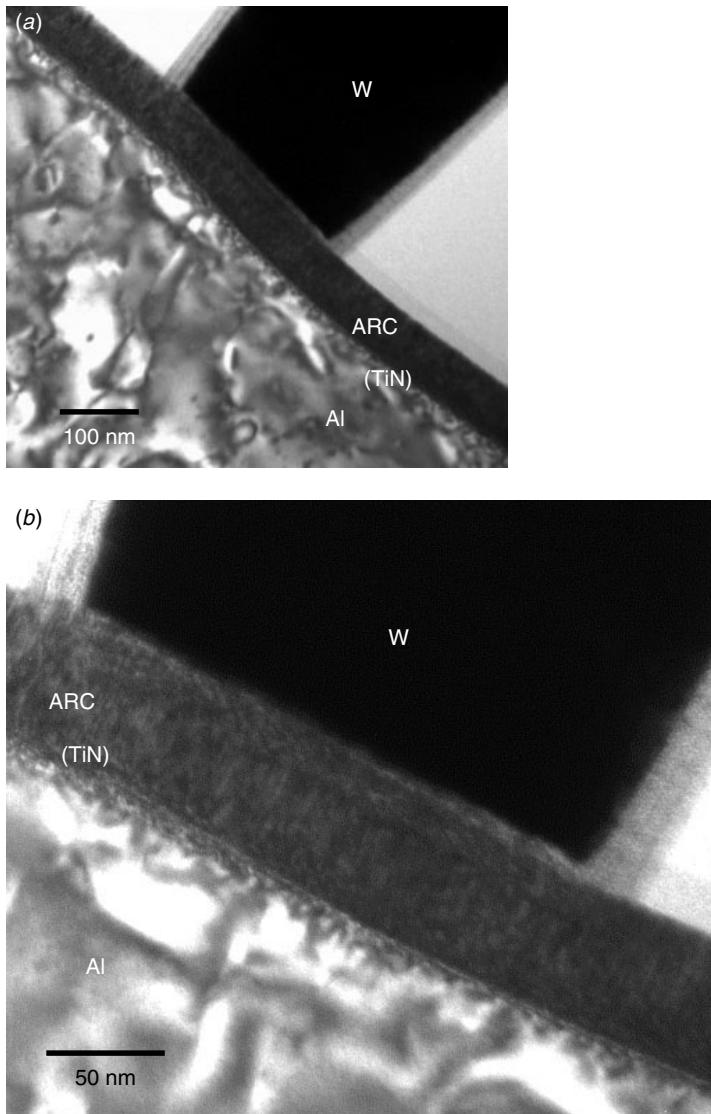


Figure 8.48 Incomplete ARC (TiN) removal during a VIA opening. The VIA opening has stopped within the TiN, as indicated.

- 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

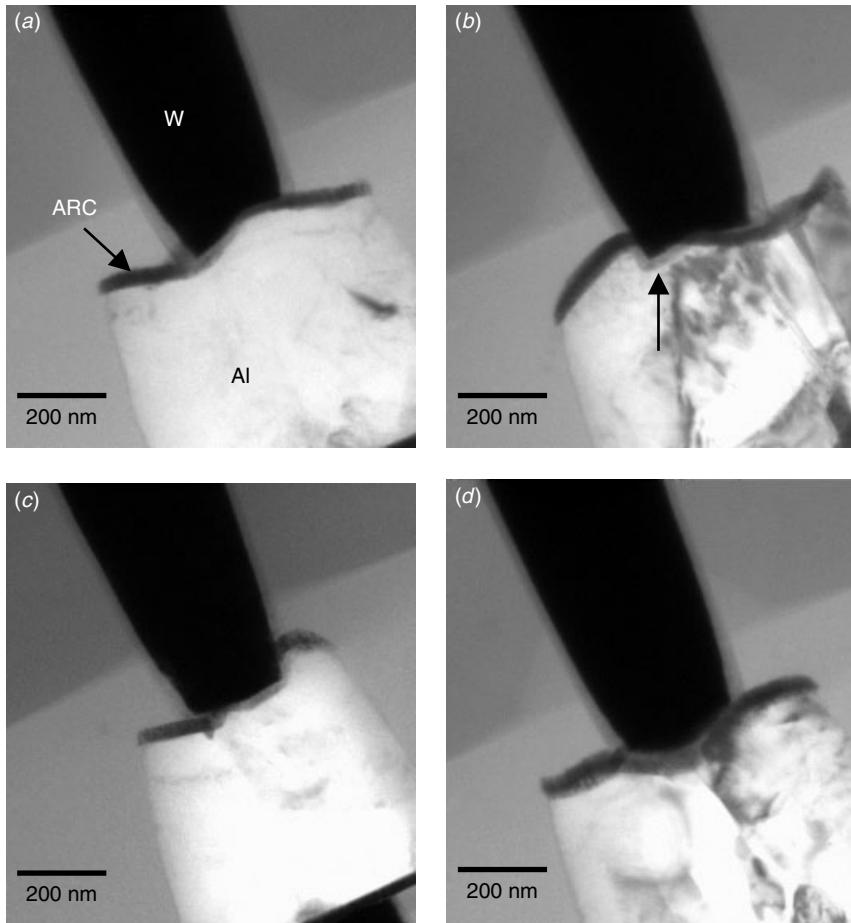


Figure 8.49 nonuniform VIA opening due to Al metallization surface roughness. Particularly in (b), half of the opening is within the Al, while the other half of the VIA is still outside the ARC layer.

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

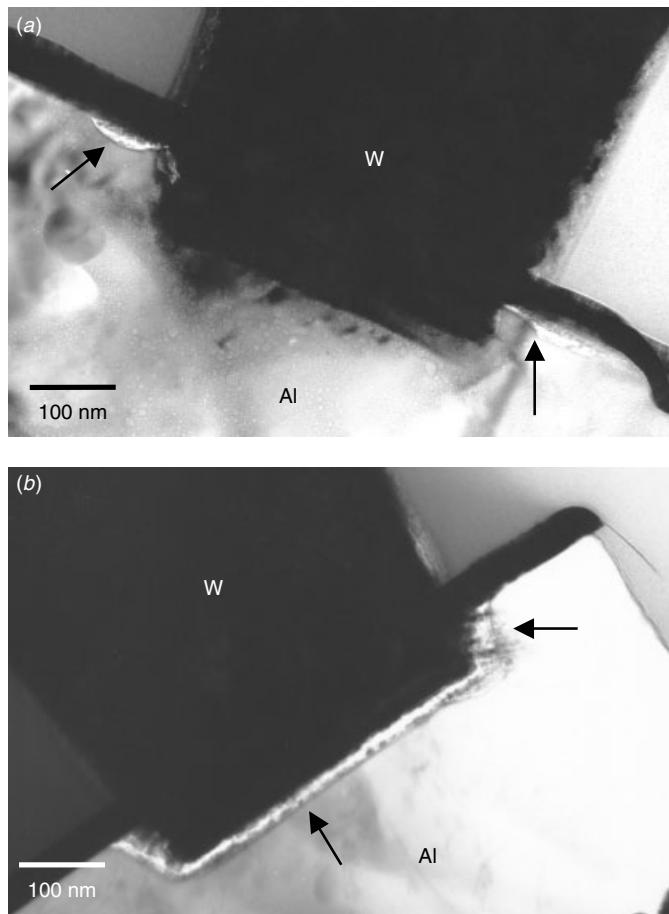


Figure 8.50 VIA bottom gap along (a) sidewall's bottom corner beneath the ARC TiN layer and (b) along the whole VIAs base.

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.

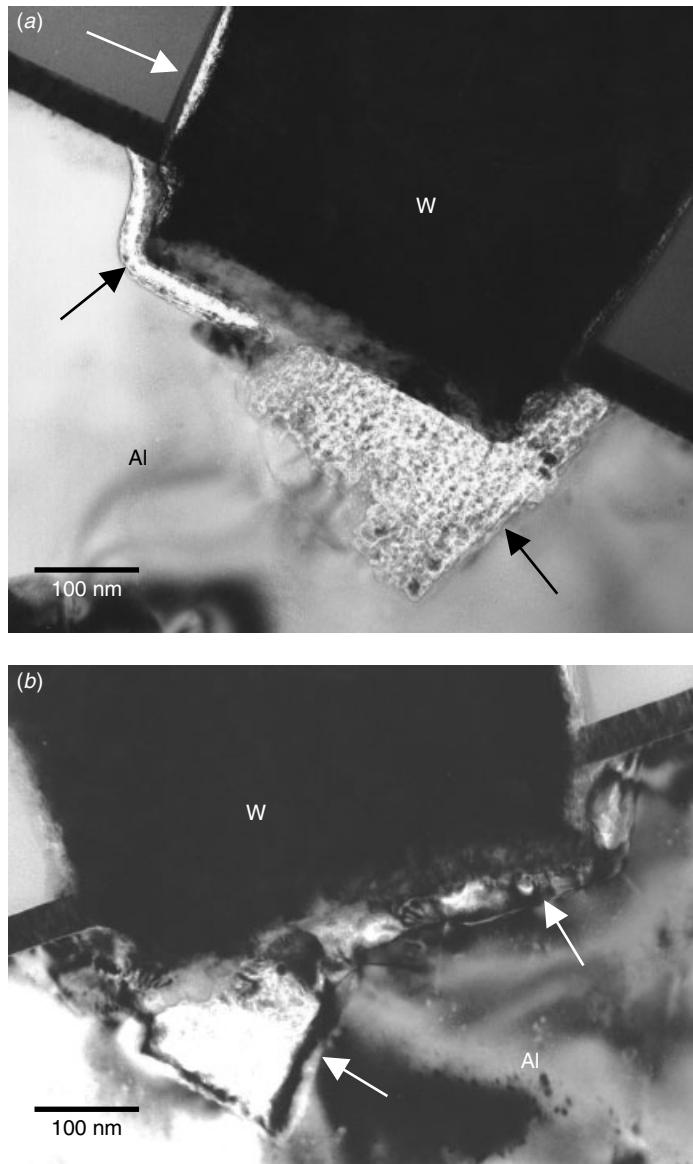


Figure 8.51 VIA bottom gap along (a) the whole bottom and (b) formed facted voids. Notice that in (a) the gap extends into the sidewall area along the SiO_2 , as indicated.

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

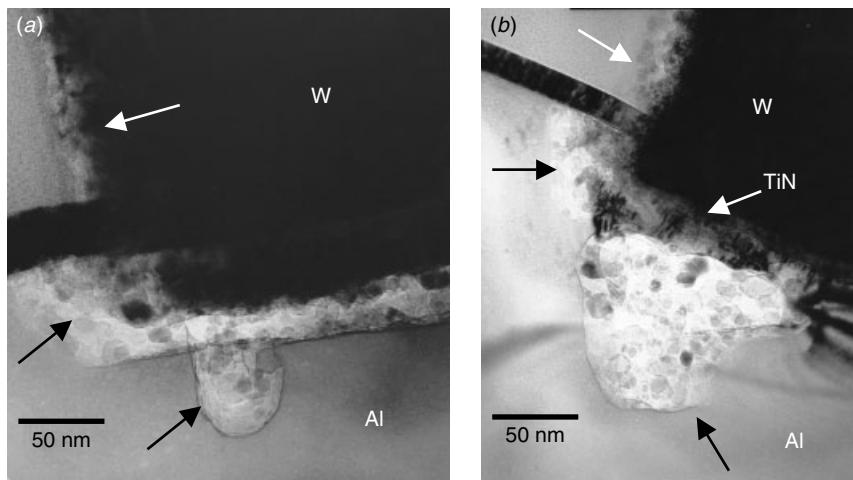


Figure 8.52 VIA bottom gap and voids (*a*) with a local irregular pit and formed under the TiN barrier layer (*b*), as indicated.

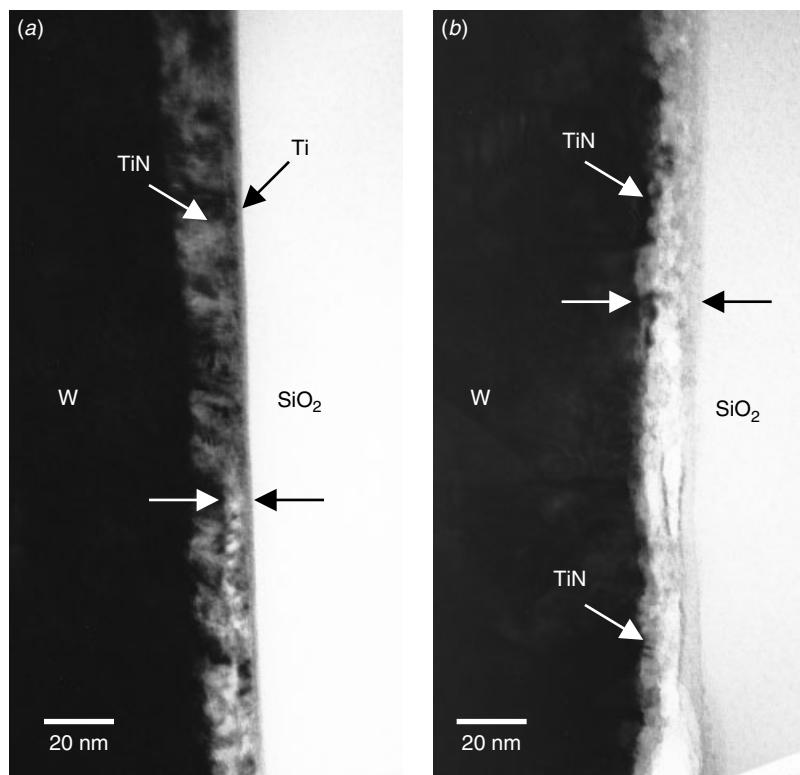


Figure 8.53 VIA sidewall interface viewed close up (*a*) at the upper half of the VIA and (*b*) at the lower half of the VIA. Notice that Ti is extremely thin even at the upper half of the VIA. As Ti disappears, a gap begins to form in between TiN and SiO_2 , as indicated in (*a*). The gap is obvious in (*b*).

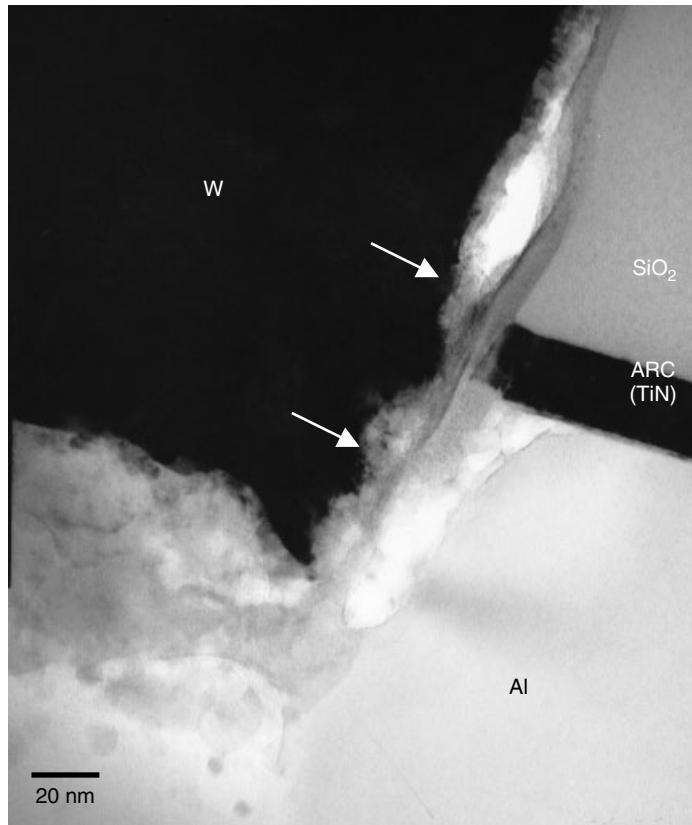


Figure 8.54 VIA sidewall to bottom corner. Both Ti and TiN have disappeared in this area. The gap between plug and SiO_2 extends into the interface of Al.

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

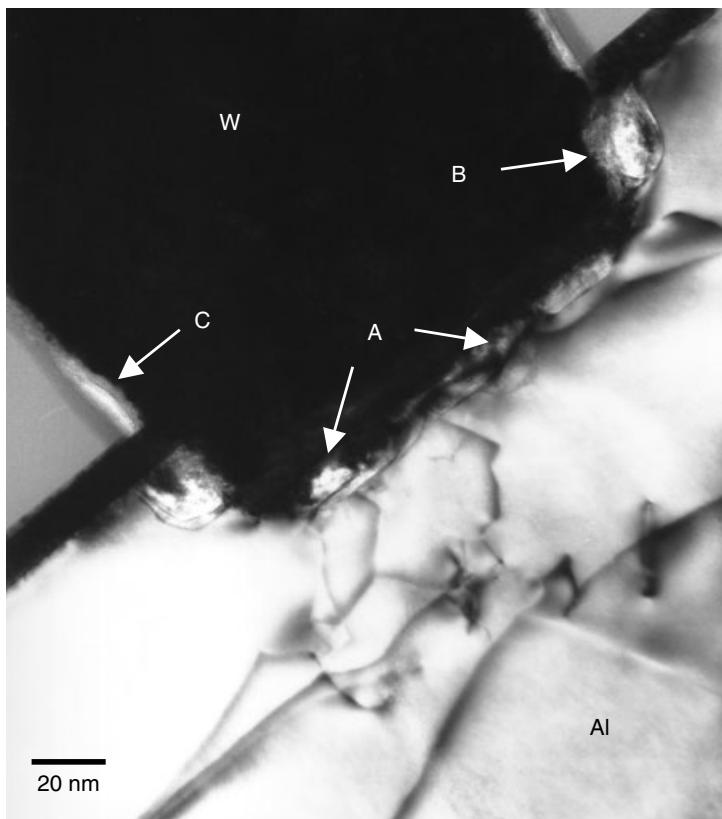


Figure 8.55 VIA bottom gap and voids. The attack includes the Ti at the VIAs bottom (arrow A), the Al metal (arrow B), and the sidewall SiO₂ (arrow C).

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,

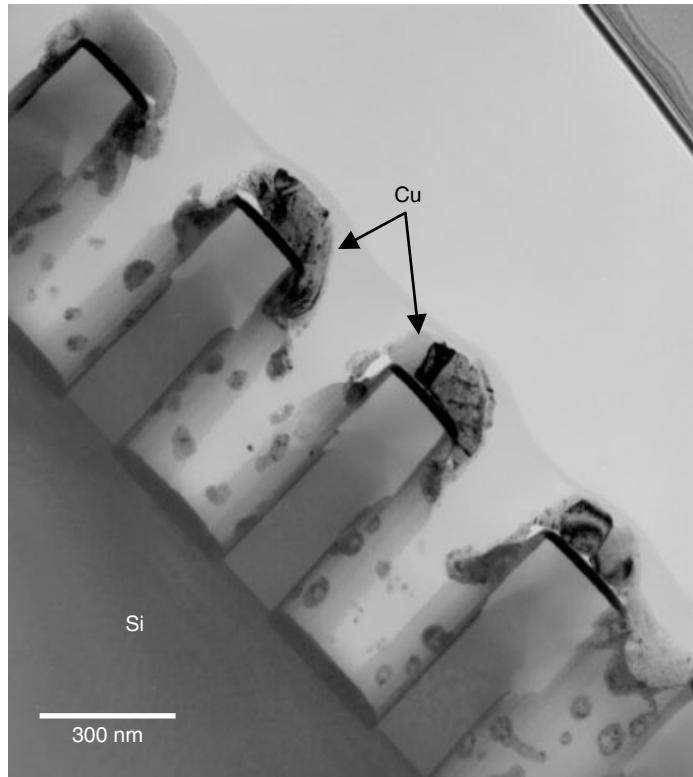


Figure 8.56 Cu metallization seed layer on the TaN barrier. The thin Cu seed layer, particularly within the contact sidewall where conformity is poor, agglomerates into small particles that attach to the contact's sidewall.

even after the whole Cu film is done. Figure 8.59 shows a total Cu film lift-off from a Ta barrier.

Bimodal Grain Size Distribution

Cu film, in as-electroplated condition, has been found to have an extremely fine grain structure. However, the grain size can grow to more than $1 \mu\text{m}$ over time, even at room temperature (Gignac et al. 1999). Such microstructure evolution is accompanied by sheet resistivity decrement, film internal stress relaxation, and volume shrinkage. Unfortunately, the grain growth within Cu film proceeds concurrently with recrystallization. This process leads to the bimodal grain size distribution, whereby large grains are surrounded by much smaller grains, and twin peaks results in the grain size distribution.

Re-crystallization (or secondary grain growth) is the process where some of the strain free grains grow at the expense of other grains. In the case of plated Cu thin films, the recrystallization process appears to occur simultaneously with normal grain growth, both of which happen at room temperature. Figure 8.60 shows a typical Cu film plan view microstructure. Some grains are clearly much larger than others. The

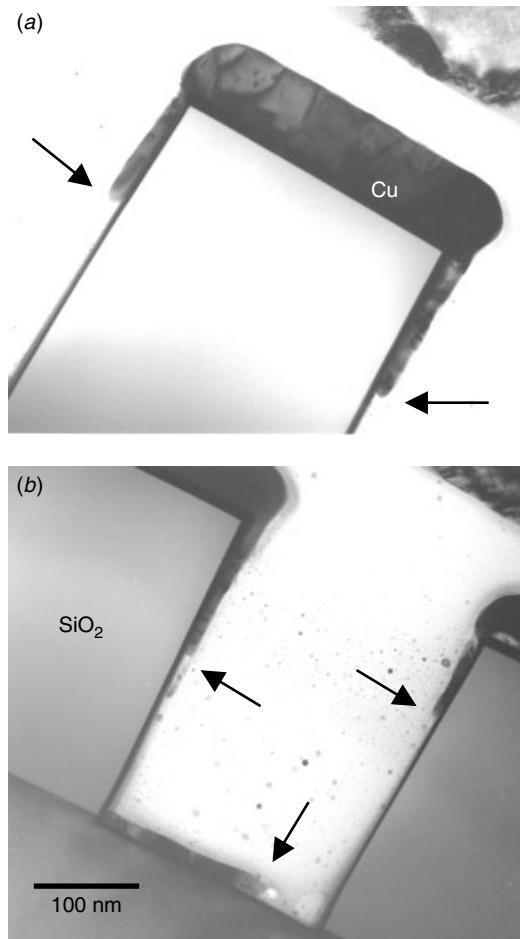


Figure 8.57 Cu metallization seed layer on the TaN barrier. Only the thin area Cu layer agglomerates into discontinuous islands. The thick Cu seed layer on the film surface does not have this problem.

large grains surrounded by much smaller grains are detrimental to electric and stress-induced migration since divergence of diffusion near the large grains will lead to early failure of the metallization. A solution to the problem is to add impurities to the Cu film. Impurities in the Cu matrix will act as grain boundary pinning centers during grain growth and thus slowing down or totally stopping the recrystallization (Harper et al. 1999).

Voids within Contacts

Voids within Cu contacts, particularly the small contacts with high aspect ratios, are known to form because of poor step coverage (Mikkola et al. 1999). Figure 8.61 shows a typical case of Cu contact voids. Notice that in the plan view, all of the voids appear to be at the center of the contacts, while in the cross-sectional view, the voids, mostly

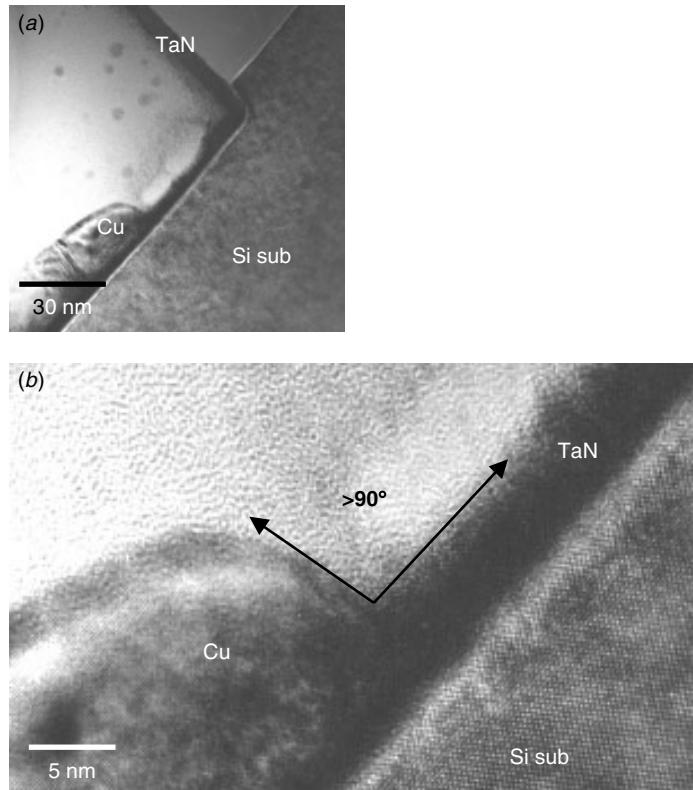


Figure 8.58 Cu metallization seed layer on the TaN barrier. At the contact's bottom, the Cu seed layer has de-wetted and formed islands. The wetting angle of Cu on TaN is measured to be more than 90° , implying that surface oxidation on the TaN at contact's bottom may have occurred.

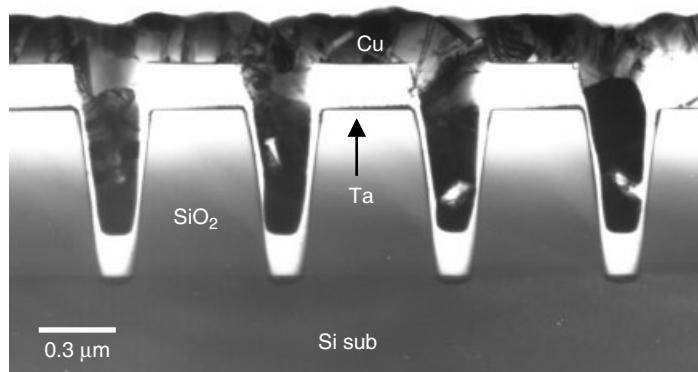


Figure 8.59 Cu metallization total lift-off from the Ta barrier metal due to improper Ta/Cu interface cleaning.

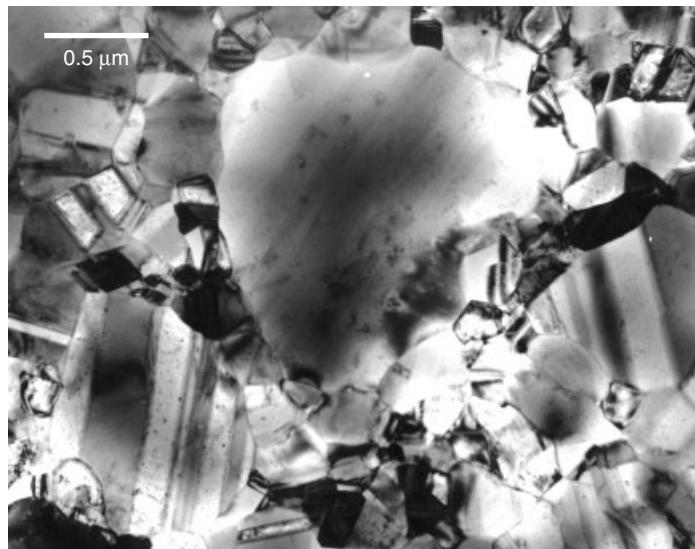


Figure 8.60 TEM plan view of Cu metallization. The bimodal grain size distribution shows large grains to be surrounded by much smaller grains. Divergent diffusion around large grains could be a potential reliability issue.

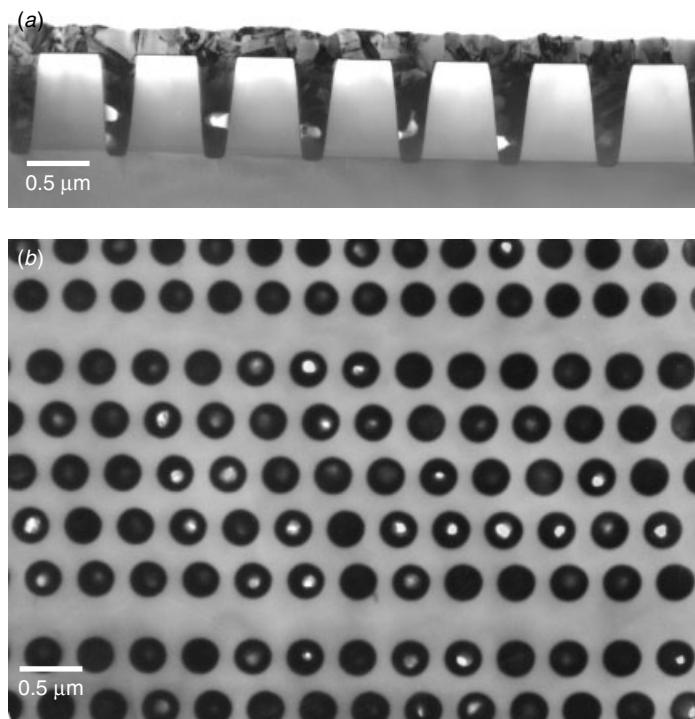


Figure 8.61 Both plan view and cross sectional TEM demonstrate that voiding occurs in almost every contact when the Cu deposition is not done properly.

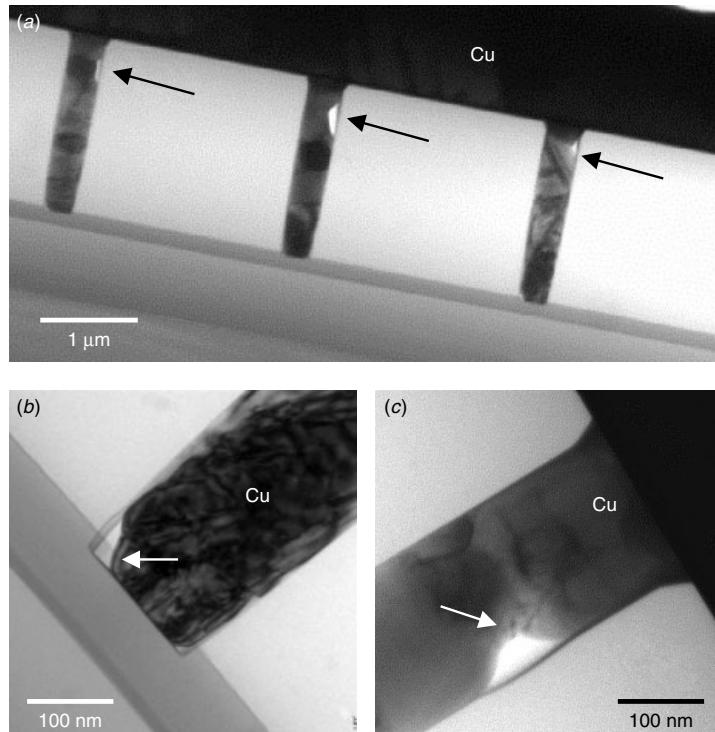


Figure 8.62 Cross section TEM shows voiding has occurred in the high aspect ratio (>6) Cu contacts, being mostly at the side wall and corners and not at the center of the contacts.

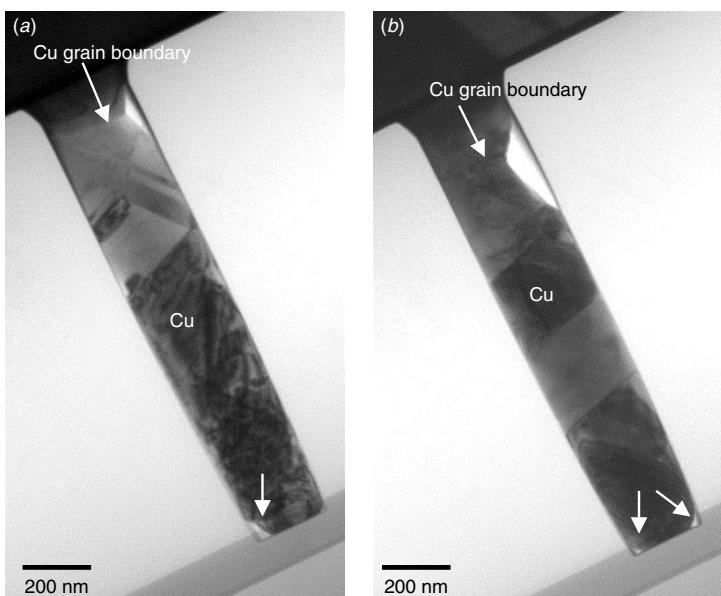


Figure 8.63 Close examination of the voids in high aspect ratio (>6) contacts shows that the voids are either at the contact's bottom corners or at the sidewall to the Cu grain boundary intersections, as indicated.

at the lower half of the contacts, are connected to the contact side wall. Other than poor step coverage, there is another effect that can cause voids in contacts.

As the Cu grains grow and recrystallize after electroplating, the grain size change is enormous. Typically the grain size increases from around less than 50 nm up to more than 1 μm . Such enormous grain size change is driven by grain boundary energy reduction. Evidence shows that the as-deposited Cu, either the CVD Cu seed layer or electroplated Cu, can be highly defective (Gignac et al. 1999) and contain twins, dislocations, and other defects. Surface topography suggests that some films may even contain microscopic pores and bubbles (Zhang et al. 1999; Boey et al. 1999; Liu et al. 1999). During crystal growth, microstructure densification can also occur. Defects and micro-voids condense and agglomerate into larger voids and become trapped within the film. Grain growth and recrystallization tends to start from the flat film and propagate into trenches and contacts (Jiang et al. 1999). This way the contact is sealed from the top. As the volume shrink due to microstructure densification, small voids inevitably form within the contacts. Figures 8.62 and 8.63 show this precisely. Closer analysis reveals the voids formed in this manner to be at Cu grain boundary intersections at the contact bottom corner or the side wall,

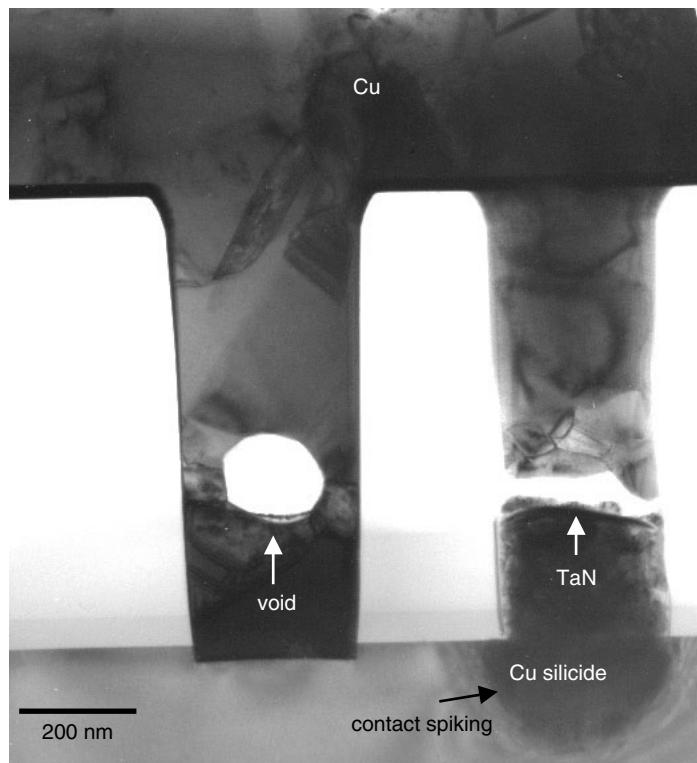


Figure 8.64 One cross-sectional TEM image includes Cu voids within the contact (*left*) and junction spiking into Si substrate (*right*). Notice that the contact spiking has formed Cu silicide and that the TaN barrier originally at the contact's bottom was lifted to a third of the height of the contact. A gap has formed above the lifted TaN.

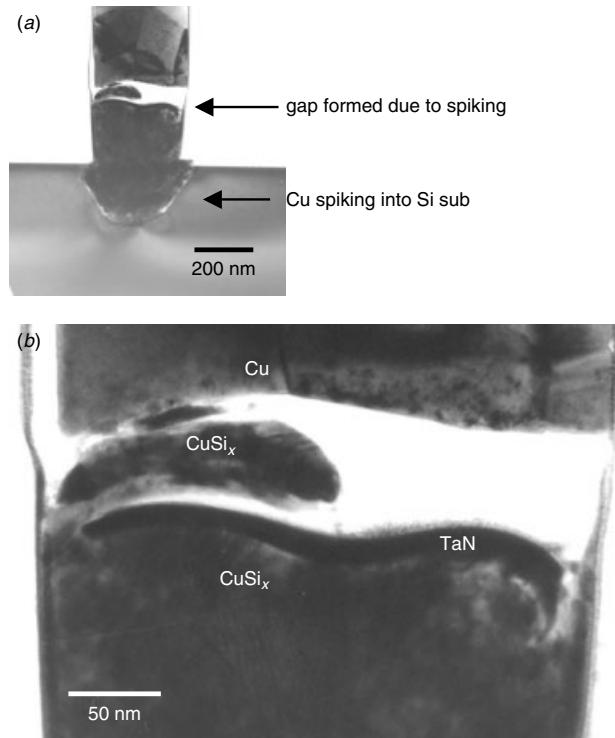


Figure 8.65 Spiking-induced contact opening. The Cu has formed a silicide with the Si substrate, and the total volume shrinks after the reaction. TaN has lifted in the middle of the contact, so the structure clearly shows the contact opening due to a gap in the middle because of the volume's shrinking.

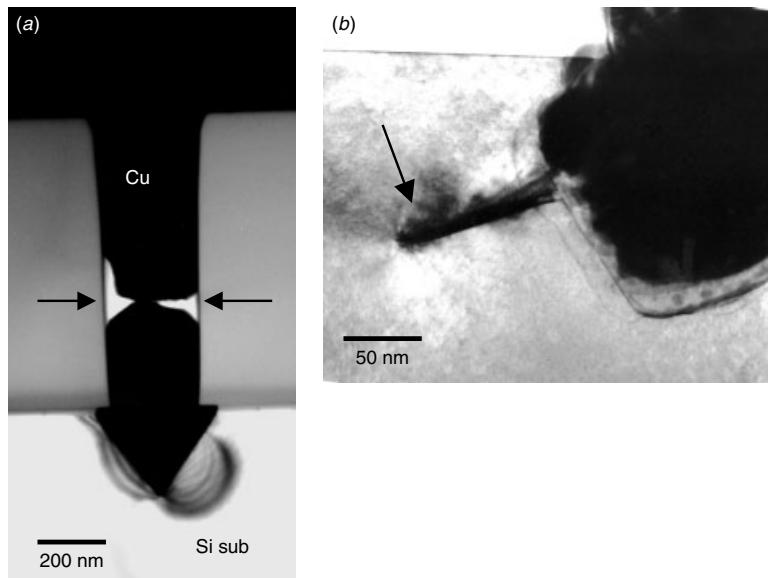


Figure 8.66 Spiking-induced the shape of a faceted arrow head (a). Local penetration through a crystallographic plane can also occur, as indicated in (b).

as shown in Fig. 8.63. These are locations where the nucleation energy for voids is minimal.

Spiking into Si Substrate

When Cu comes in direct contact with Si, it forms Cu silicide. This is why a diffusion barrier layer is required. If there are process defects on the barrier layer, Cu can

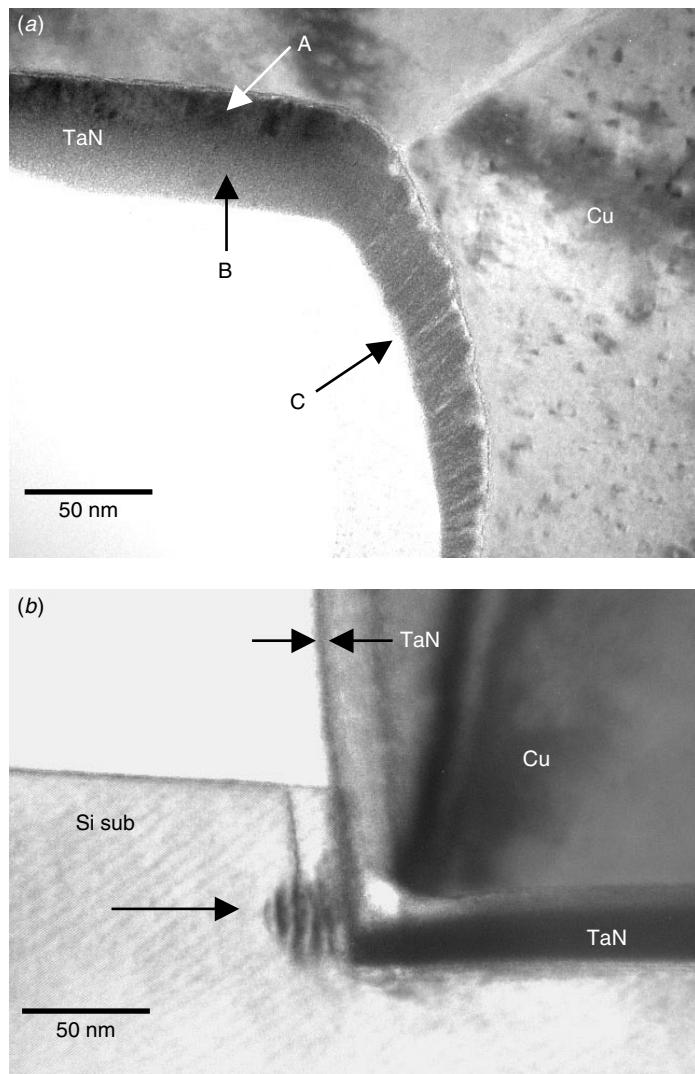


Figure 8.67 TaN barrier layer with different microstructures formed in different locations (a). The crystalline phase is observed in horizontal film near the Cu, location A. The amorphous phase is observed in horizontal film near the oxide, location B. An amorphous columnar feature is seen on the contact's sidewall, location C. The weak spot is observed at the contact's bottom corner where Cu seep through into the Si, as indicated (b).

penetrate and react with Si. As this happens in a contact hole, junction spiking occurs. Like the Al spiking issue discussed earlier, a proper barrier layer is the best solution to the problem. Figure 8.64 shows the problem in a real device failure. The micrograph in fact shows two critical issues that have occurred here. The void as discussed in the previous session, formed in the middle of the contact, with its nearly spherical shape, is clearly visible at the left-hand contact and the contact spiking is seen at the right.

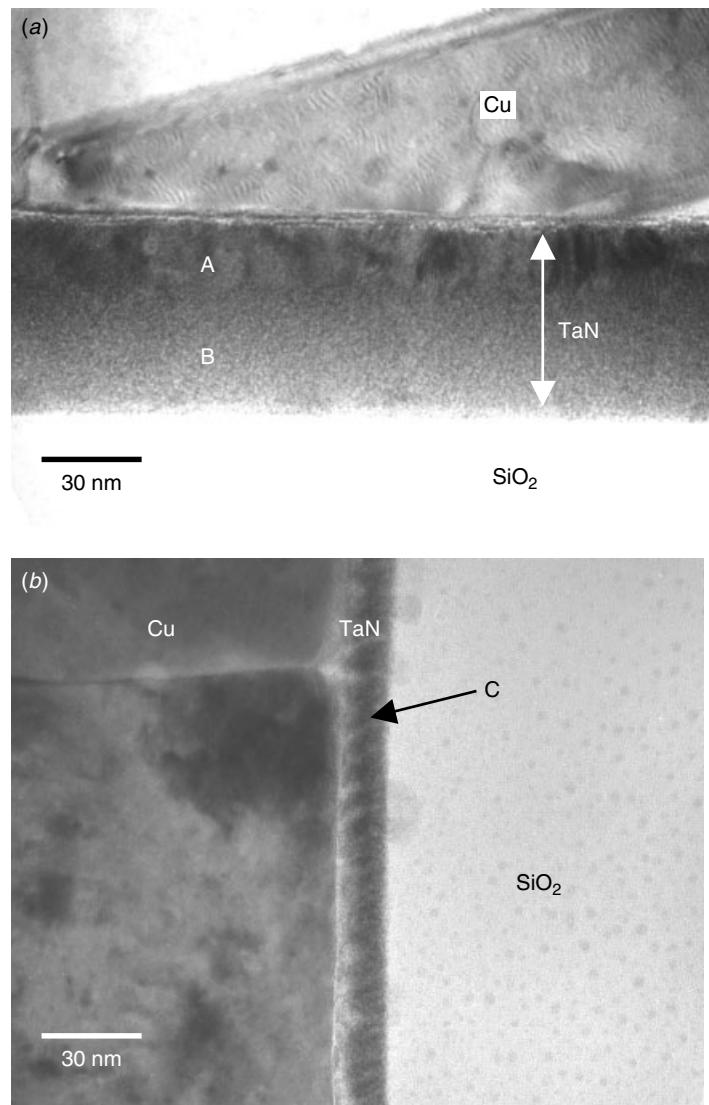


Figure 8.68 TaN barrier layer forms different microstructures in different locations. The crystalline phase is observed in horizontal film near the Cu, location A. The amorphous phase is observed in horizontal film near the oxide, location B. An amorphous columnar, and porous, area is observed on the contact's sidewall, location C.

As Cu forms silicide readily with Si, a volume shrink as a result of silicidation is expected, which is the case in most silicidation processes. The result of contact spiking, however, is contact opening, instead of junction short. Examination of the contact spiking in Figs. 8.64 and 8.65 reveals a thin layer of the barrier TaN visible in the middle of the contacts with spiking issue. The contact opening is due to the volumes shrinking and the separation of Cu silicide from pure Cu. Figure 8.66 shows a classic example of faceted spiking as an arrow head penetrating the Si substrate and also an example of local penetration through a specific lattice plane.

The location of the TaN layer within the spiking contact suggests that the spiking mechanism is due to weak points at the contact's bottom corner. Figure 8.67 shows where the problem has originated. As TaN is deposited, three different microstructures are formed. On the flat surface thin film, TaN forms a crystalline microstructure near the Cu interface and an amorphous microstructure near the SiO_2 interface. Such a combined crystalline and amorphous microstructure is believed to improve the robustness of TaN as a diffusion barrier. In contact holes, however, TaN has gradually transformed from a dense amorphous structure into a columnar, porous, and amorphous structure, as indicated in Figs. 8.67 and 8.68. This microstructural change is accompanied by a dramatic reduction of thickness. On reaching the contact's bottom, the TaN thickness diminishes to a nearly negligible film, Fig. 8.67(b). At the contact's very bottom the TaN thickness has regained a healthy level, but the contact's side wall at the bottom corners has become riddled with weak spots. As a result Cu can seep through and

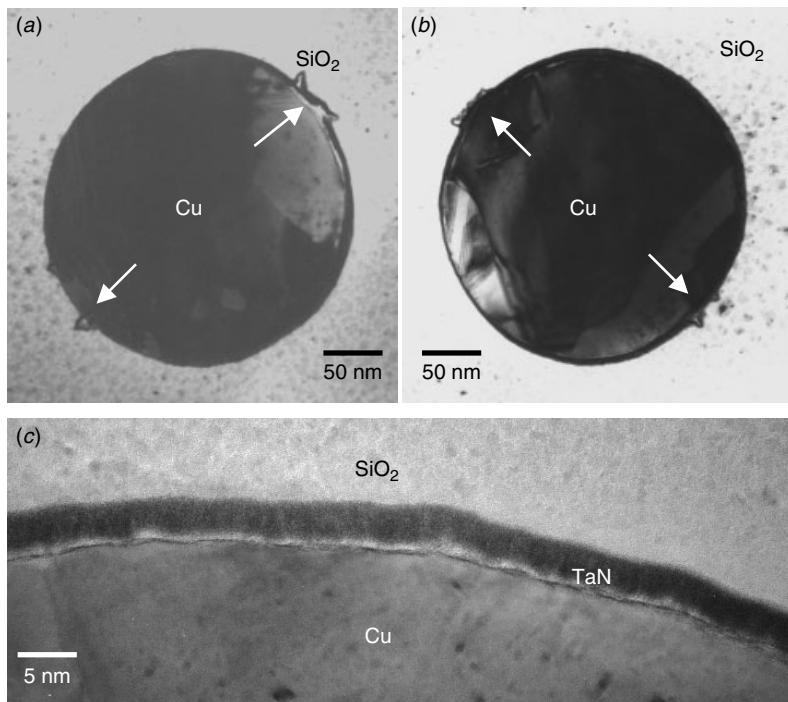


Figure 8.69 Porous TaN barrier layer observed in the horizontal TEM section. Thin and weakening TaN on the contact's sidewall may introduce a few weak spots, as indicated.

react with Si to form silicide, as observed in Fig. 8.67(b). The weakening of TaN at the contact side walls is better illustrated in Fig. 8.69. The weak spots may be introduced during or after TaN deposition. Once the Cu silicidation is nucleated at these weak locations, the newly formed silicide will gradually push the TaN up from the contact's bottom, resulting in the situation observed in Figs. 8.64 and 8.65.

The contact's side wall TaN, or Ta, barrier film quality holds the solution to the spiking problem. Improvements in the CVD process or even new barrier deposition techniques are needed to achieve higher film step coverage and lower film thickness.

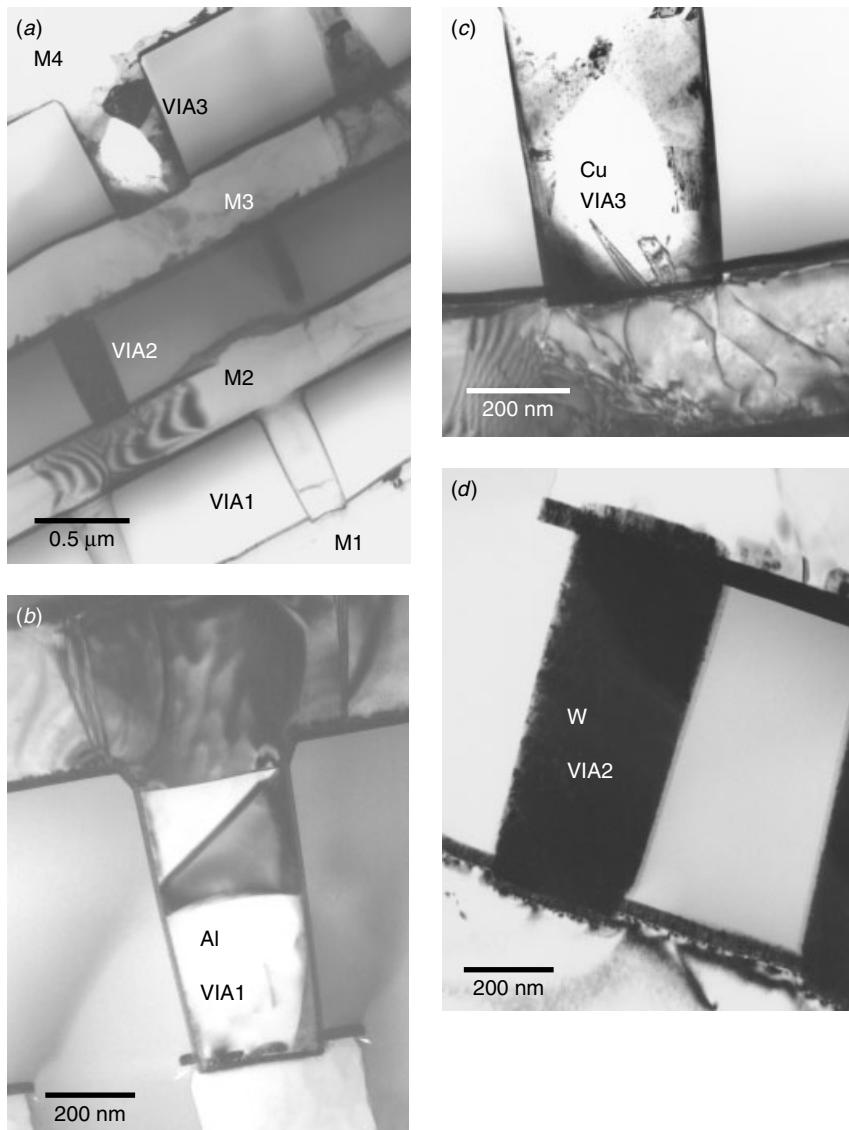


Figure 8.70 Integrating the Cu process into the conventional Al and W–plug process. M1, M2, and M3 are Al, and M4 is Cu. VIA1 is Al, VIA2 is the W–plug, and VIA3 is the Cu process.

Integration of Cu into Al Metallization Technology

In transferring from Al technology into Cu technology, most wafer fabrication manufacturers combined the processes as a transition system. While Al is still used in lower levels of metallization, in higher level metallization Cu is being used. This is because all aspects of the process must be tested and understood before a complete switch over to Cu technology. Figure 8.70 shows a typical mixed metallization scheme. Cu is used in the upper level interconnects mainly due to contamination concern and thus the advantage of using Cu metallization's low resistivity is not fully utilized. Contamination control using Cu technology is still a major concern in wafer fabs. Cross contamination from different process steps, equipment, and even wafer handling facilities needs to be understood at each point of the Cu process before it can be fully embraced by wafer fab.

REFERENCES

- J. S. Boey, G. L. Griffin, A. W. Maverick, and H. Fan, *Materials Res. Soc. Symp. Proc.*, **564**, 251, 1999.
- J. C. Bravemen, G. L. Patton, and J. D. Plummer, *J. Appl. Phys.*, **57**, 2779, 1985.
- L. Chen, B. Ekstrom, and J. Kelber, *Materials Res. Soc. Symp. Proc.*, **564**, 287, 1999.
- E. G. Colgan and K. P. Rodbell, *J. Appl. Phys.*, **75**, 3423, 1994.
- C. M. Dai and C. H. Tung, *J. Appl. Phys.*, **73** (5), 2543–2547, 1993.
- D. R. Frear, J. E. Sanchez, A. D. Romig Jr., and J. W. Morris Jr., *Metall. Trans.*, **A21**, 2449, 1990.
- R. Furlan, J. Spiegel, and J. W. Swart, *J. Electrochem. Soc.*, **138**, 2377, 1991.
- L. M. Gignac, K. P. Rodbell, C. Cabral Jr., P. C. Andricacos, P. M. Rice, R. B. Beyers, P. S. Locke, and S. J. Klepeis, *Materials Res. Soc. Symp. Proc.*, **564**, 373, 1999.
- R. G. Gordon, X. Liu, R. N. R. Broomhall-Dillard, and Y. Shi, *Materials Res. Soc. Symp. Proc.*, **564**, 335, 1999.
- J. M. E. Harper, C. Cabral Jr., P. C. Andricacos, L. Gignac, I. C. Noyan, K. P. Rodbell, and K. C. Hu, *Materials Res. Soc. Symp. Proc.*, **564**, 387, 1999.
- J. W. Hartman, H. Yen, H. A. Atwater, and I. Hashim, *Materials Res. Soc. Symp. Proc.*, **564**, 257, 1999.
- D. S. Hennan, M. A. Schuster, and R. M. Gerber, *J. Vac. Sci. Tech.*, **1**, 515, 1971.
- A. Jain, O. Adetutu, B. Ekstrom, G. Hamilton, M. Herrick, R. Venkatraman, and E. Weitzman, *Materials Res. Soc. Symp. Proc.*, **564**, 269, 1999.
- T. Kikkawa, H. Aoki, and J. M. Dryman, *IEEE Trans. Electron Devices*, **40**, 296, 1993.
- C. Kim and J. W. Morris Jr., *J. Appl. Phys.*, **72**, 1837, 1992.
- L. T. Koh, S. Y. M. Chooi, K. L. Chok, H. M. Li, and F. H. Ng, *Materials Res. Soc. Symp. Proc.*, **564**, 451, 1999.
- K. Y. Liu, M. S. Tse, and W. L. Goh, *Materials Res. Soc. Symp. Proc.*, **564**, 263, 1999.
- R. D. Mikkola, Q. T. Jiang, R. Carpio, and B. Carpenter, *Materials Res. Soc. Symp. Proc.*, **564**, 399, 1999.
- S. P. Murarka, *Metallization, Theory and Practice for VLSI and ULSI*, Butterworth-Heinemann, 1993.
- S. I. Ogawa, S. Okuda, and T. Yoshida, *J. Appl. Phys.*, **65**, 668, 1989.
- M. Park, S. J. Krause, and S. R. Wilson, *Appl. Phys. Lett.*, **64**, 721, 1994.

- P. E. Riley, T. E. Clark, E. F. Gleason, and M. A. Garver, *IEEE Trans. Solid Circuit Manuf.*, **3**, 150, 1990.
- C.-H. Tung, C. T. Chang, H. C. Chang, S. Chiu, B. Kok, and L. H. Chang, *J. Appl. Phys.*, **68** (4), 1592–1600, 1990.
- C.-H. Tung, R.-L. Chiu, and P.-H. Chang, *Scripta Materialia*, **34** (9), 1473–1477, 1996.
- C.-H. Tung, K. Y. Lee, R. C. Deng, and L. S. Lee, *Appl. Phys. Lett.*, **63** (9), 1197–1199, 1993.
- S. Wolf and R. N. Tauber, *Silicon Processing for the VLSI Era, Vol. 1: Processing Technology*, 403, Lattice Press, 1986.
- C. Y. Wong, A. E. Michael, R. D. Issac, R. H. Kastl, and S. R. Mader, *J. Appl. Phys.*, **55**, 1131, 1985.
- S. S. Wong, C. Ryu, and H. Lee, *Proc. IEEE 1998 IITC*, 107, 1998.
- N. Yamauchi and R. Reif, *J. Appl. Phys.*, **75** (7), 3235–3257, 1994.
- J. Zhang, D. Denning, G. Braeckelmann, G. Hamilton, J. J. Lee, R. Venkatraman, B. Fiordalice, and E. Weitzman, *Materials Res. Soc. Symp. Proc.*, **564**, 243, 1999.

PART III

9 ULSI Devices I: DRAM Cell with Planar Capacitor



This is TEM image of carbon replica with wrinkles. It resembles satellite image from Mars showing trenches on Mars surface, an evidence of mass water erosion during early days.

Dynamic random access memory (DRAM) is the flagship product of the semiconductor industry (Prince 1983). It is the memory product with the closest ties to the technology, and also the one with the highest volume of production up to date. The history of DRAM manufacturing has been about the struggle to increase the number of bits on a single wafer. The more DRAM bits that can be put on a single wafer, the lower the average cost per bit. The scaling of the production technology is the focus of the DRAM manufacturer, and as a result DRAM production is one of the driving forces of ULSI process technology innovation.

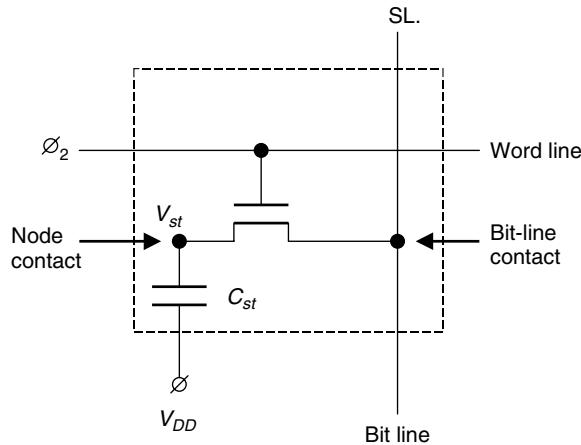


Figure 9.1 Schematic structure of one-transistor DRAM unit cell with the transistor as the control for read and write and the capacitor as the storage unit. The storage node contact and bit-line contact are marked.

Dynamic random access memories (DRAMs) are so named because their cells can retain information for a limited time before they must be read and refreshed at periodic intervals. For the modern device, each DRAM cell consists of one access transistor and one storage capacitor. Figure 9.1 shows a schematic of a typical DRAM unit cell. The interested reader is encouraged to refer to the enormous literature of books and original papers for details on DRAM cell designs.

The cell structure used in one transistor DRAM cell below 1 Megabit DRAM products ($1 \text{ Mb} = 10^6 \text{ bits}$) is called a planar DRAM cell. It is so named because the capacitor used in DRAM cells is placed horizontally beside the control transistor. Up to approximately 70% of the cell area can be used as the capacitor in such a layout design.

For higher densities, the planar diffusion plate does not have the capacity to hold enough charges for error-free storage. To increase capacitance, a so-called vertical cell capacitor is used. These capacitors are either stacked on top of the transfer gates as super-structures or they are buried deep into the substrate as trenches. We will discuss both varieties in later chapters. Here we discuss some of the fundamentals of the planar cell DRAMs.

9.1 PROCESS FLOW

Although the application of a DRAM device below 4 Mb is history now and its production has long been stopped, the principle of the DRAM cell operation has not changed, and it is instructive to see how the process and structure evolved from the earlier days to its present form. Figure 9.2 shows a schematic of its cross-sectional structure and corresponding circuit of a planar DRAM cell. Table 9.1 gives an outline of the essential processing steps. The memory array consists of n-channel devices, while the CMOS peripheral devices may be either n- or p-channel. The processing steps up to

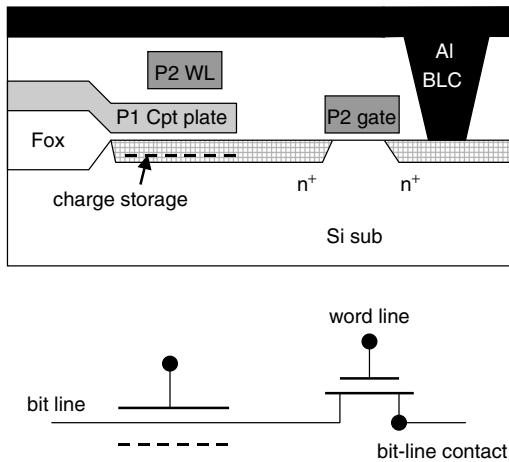


Figure 9.2 Schematic cross section of a DRAM planar cell structure and its corresponding schematic circuit. Poly 1 forms the capacitor plate. Poly 2 forms the word-line polygate and polyrunner. Al metal 1 is used for bit line and bit-line contact.

TABLE 9.1 DRAM Planar Cell Process Flow

Process Flow	Function
Implant ion	Forms N-well
Deposit field oxide	Forms LOCOS isolation
Mask, capacitor; implant boron	Implants high C to increase capacitance
Grow gate oxide	
Deposit• poly 1, mask, etch poly	Forms capacitor plate
Grow poly oxide	Forms insulation between poly layers
Deposit poly 2, mask, etch poly 2	Forms gates for cell and peripheral devices
Deposit BPSC	Forms ILD
Mask, contact	
Deposit metal, etch metal	Forms interconnect

the LOCOS oxidation are identical to those of a logic CMOS device. The storage area in the substrate is first defined by implanting arsenic in the p-type substrate to create a n+/p junction. Boron is implanted and driven in beyond this junction. This results in an increased doping density and hence increased depletion capacitance. The process is normally referred to as high capacitance (Hi-C) cell, and it was developed by Intel (El-Mansy and Burghard 1982). The capacitor's structure is completed by growing a thin layer of dielectric and by depositing and patterning a polysilicon counter electrode.

Next a layer of thermal oxide is grown a thicker layer of polyoxide is grown simultaneously on the polysilicon plate. On account of its higher impurity concentration, the oxidation rate is higher in the polyoxide. A second layer of polysilicon (or polycide)

is then deposited and etched. It serves as the gates for all the n- and p-channel devices and for interconnection as the word lines. The back-end process is again fairly conventional. A thick layer of boron-phosphorous glass is deposited and reflowed to smooth out the sharp steps. The contact holes are etched, and a metal layer is deposited and patterned. The metal lines reach into the holes and make contact with the diffusions (bit-line contacts) and the polysilicon gate. Metal lines are also used as bit lines in the memory array.

9.2 THE CELL STRUCTURE

Figure 9.3 shows a cross-sectional TEM micrograph of a planar DRAM unit cell and its corresponding schematic structure. The LOCOS field oxide acts as a diffusion isolation. The substrate together with the polysilicon plate and a thin layer of the thermal oxide constitute the capacitor structure. The transistor is connected serially with the capacitor diffusion VIA the source n+ diffusion, which is self-aligned to both the capacitor and the gate. A second polysilicon layer connects all the gates together along that line and constitutes the word line. In Fig. 9.3 we observe two word lines. One of them

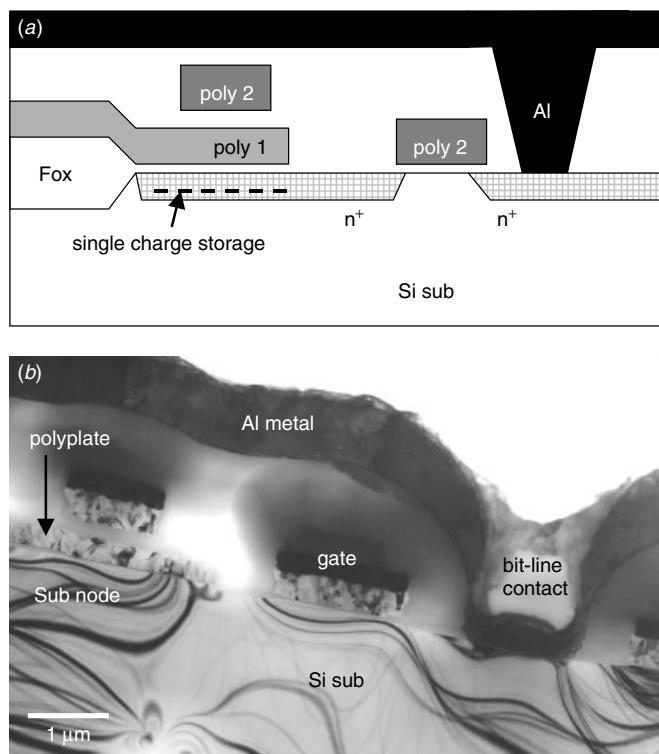


Figure 9.3 A typical TEM cross section of the DRAM planar cell structure. A direct comparison of (b) with the schematic of (a) shows how the device is built from the VLSI processes. Notice that the substrate's diffusion junction cannot be seen in the TEM cross section without delineation.

is on the thin gate oxide and serves as a transfer gate, while the second one is seen on the top of the capacitor and is the poly interconnect runner when the same poly line is used as control gate in the adjacent cells. The word lines are sometimes made up of a composite film of polysilicon and tantalum or tungsten silicide (usually called polycide). The reason for using polycide film is to reduce the word line resistance and increase access speed. A contact hole is seen opened to the drain side of the transfer device through the thick BPSG layer, and a metal bit-line contact is connected to it.

Cross-sectional TEM Analysis

As we mentioned above, a DRAM unit cell consists of one transistor and one capacitor. The transistor can be further broken down to a control gate, source and drain junctions, and their associated contacts. The capacitor consists of two electrodes, one dielectric layer in between and a contact to the transistor:

1. The control gate, usually made of a polysilicon film, also serves as a word-line interconnection. A low-resistivity word line is required to improve DRAM cell access time. As a result the polycide (a polysilicon-silicide composite) layer is often chosen to improve resistivity.
2. The two contacts into the drain and source of the control transistor are called bit-line contact and node contact, respectively. The bit-line contact connects to the cell to provide information read and write to the periphery. The node contact connects the control transistor to the capacitor node electrode, as seen in Fig. 9.1. In the planar capacitor cell the node contact is eliminated by connecting the transistor active diffusion directly to the capacitor node diffusion using self-align diffusion implantation. Figure 9.4 shows a planar unit cell before, Fig. 9.4(a), and after, Fig. 9.4(b), junction delineation to reveal how the source diffusion acts as the node contact and connects the transistor's source diffusion to the capacitor substrate plate. As a result there is only one physically distinctive node contact in each planar capacitor cell. This is clearly shown in the TEM micrograph of Fig. 9.3. The Al metallization film was chosen as the bit-line interconnect and the bit-line contact material.
3. The lower magnification cross-sectional TEM image (seen in Fig. 9.5) shows how the Al metal bit lines connect all the cells in one direction.
4. Figure 9.6 shows cross sections of the same device in two directions, one along word line, Fig. 9.6(a), and one along bit-line, Fig. 9.6(b), as a further illustration of the interconnecting relationship among neighboring cells. Figure 9.6(a) gives the cross-sectional view along the position seen in Fig. 9.6(b) at exactly the location of the bit-line contact and gate area. While Fig. 9.6(b) is a cross-sectional view along the position seen in Fig. 9.6(a) at exactly the location of the gate area and bit-line contact.
5. The capacitor was built by using a Si substrate as the capacitor node, and each capacitor node was isolated from the other nodes and from transistor active area by LOCOS isolation. Such isolation LOCOS is clearly seen in Fig. 9.6(a) at both ends of the capacitor and under the capacitor polyplate.

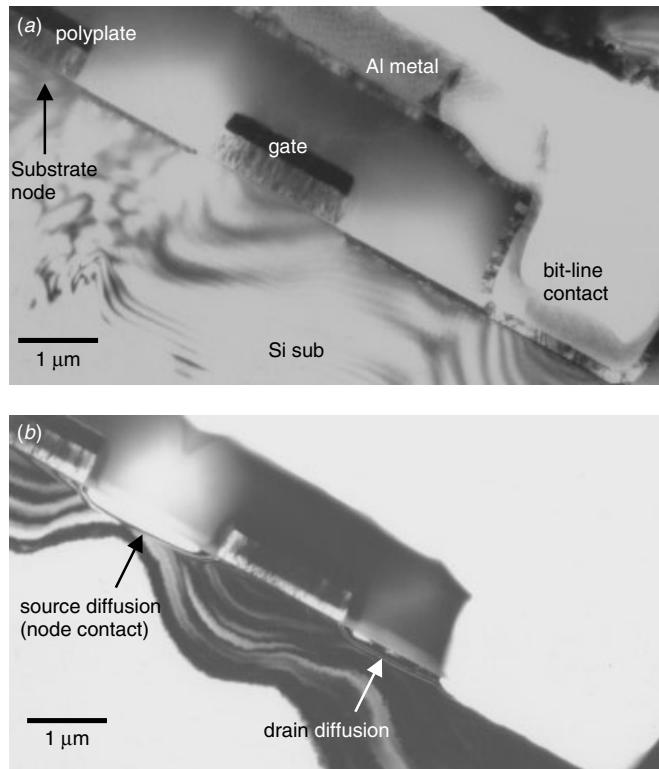


Figure 9.4 Cross section of a planar DRAM cell. (a) Basic structures of the transistor and the capacitor. (b) After junction delineation, the substrate's diffusion as the node contact (to connect transistor drain and capacitor plate) is revealed.

Plan View TEM Analysis

To further elucidate the cell arrangement and how the various constituents of the transistor and capacitor, such as LOCOS, contacts, gate, word line, bit line, and bit-line contacts, correlate, we next take a plan view and analyze each individual layer (polysilicon, interlayer dielectric (ILD), polycide etc.) step by step.

Step 1. The sample is prepared by removing all of the layers from the top of the Si substrate. This is done by dipping it into the HF or buffered oxide etchant (BOE). Figure 9.7 shows the remaining Si substrate visible by TEM. Si substrate with ion implantation reveals distinctive substrate residue dislocation loops. Mask edge defects appear as continuous lines at the boundaries of the active area, which is distinct from the normal substrate areas as they do not show any dislocation loops. Two different ion implantation pattern trajectories are observed:

- A strip pattern with a serpentine curve whose width remains more or less the same. This is the active diffusion area of the transistor source's implantation. This area also provides the node contact and connects the transistor to the capacitor substrate plate.

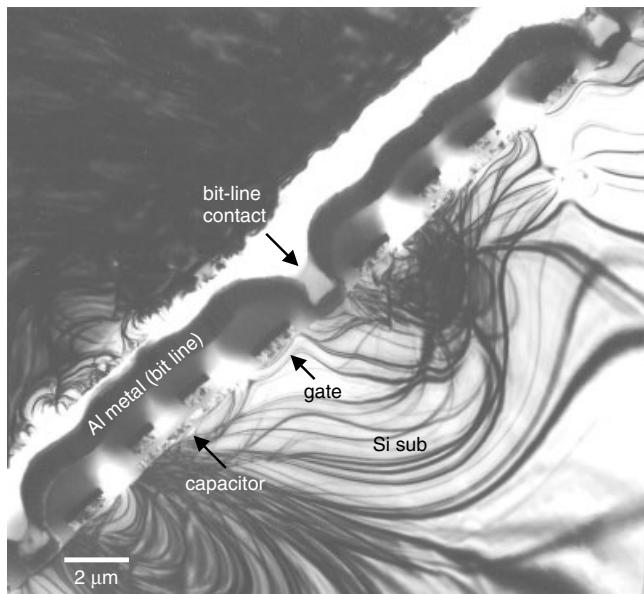


Figure 9.5 Low-magnification cross section of a DRAM planar cell revealing a long bit line, vertically sectioned word lines, and the cells' physical relationship to the neighboring cells.

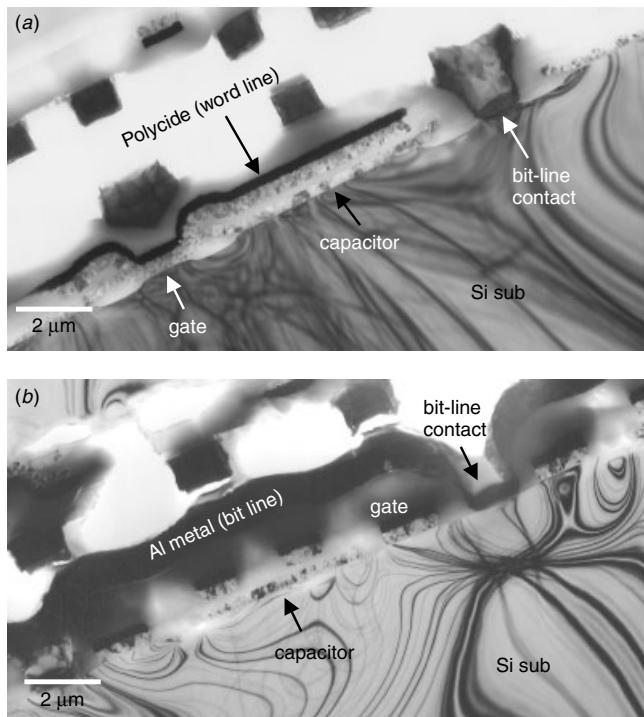


Figure 9.6 Low-magnification cross section of a DRAM planar cell, (a) along the word line and (b) along the bit line, to show the interconnections among neighboring cells. Notice in (a) only a section of word line is observed. This is due to the zigzag pattern of the word line.

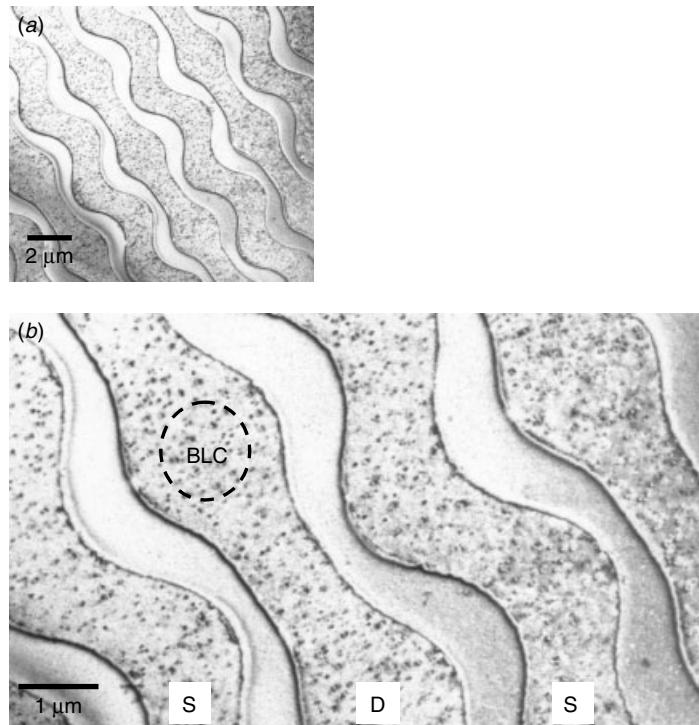


Figure 9.7 TEM plan view of the Si-substrate ion implantation imprint. The micro dislocation loops clustered within active area reveal clearly the active areas and their boundaries. The serpentine pattern (marked *S*) represents the source diffusion areas, which also act as node contacts. The straight (but changing in width) pattern (marked *D*) represents the drain diffusion where the bit-line contact (BLC) will connect into the center of the wider area, as indicated by a dashed circle.

- A strip pattern with a straight path whose width periodically changes. This is the active diffusion area of the transistor drain's implantation. The periodic widening occurs in areas where the Al bit line connects and makes contact with the substrate to form the bit-line contacts.

Step 2. The interlayer dielectric (ILD) layer is prepared. For the self-support plan view of the TEM sample, both polysilicon and Si substrate are chemically etched away using a poly-etch solution. The isolated ILD glass is then shadowed with Pt and carbon to enhance the surface features for TEM analysis. Figure 9.8 shows the result:

- The grainy texture of the structure observed through out the film is the characteristic of the surface grain topology of polysilicon. The ILD glass itself does not have any observable internal microstructure.
- The two parallel dark patches that run horizontally, Fig. 9.8(b), in the cell areas are the LOCOS isolation islands. They are much thicker than other areas and thus look much darker even without the poly topography imprint. Their thickness

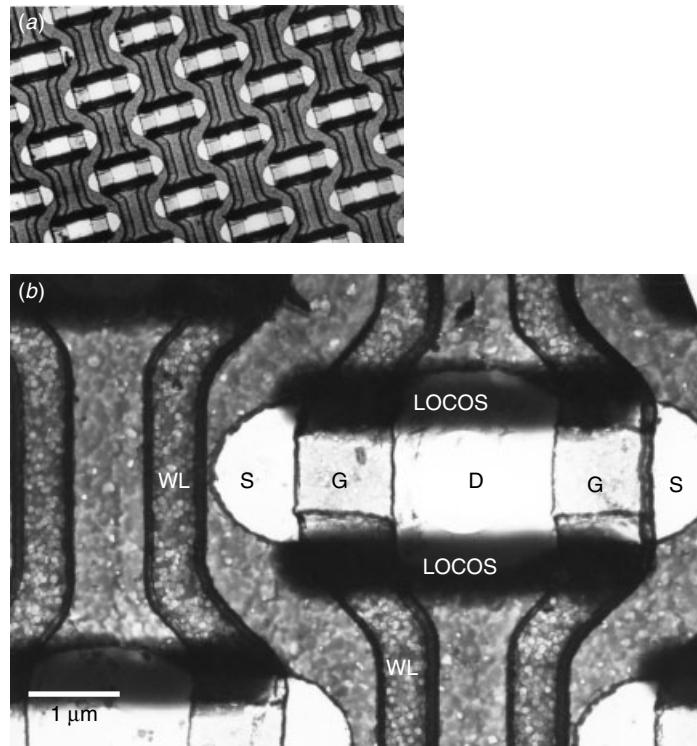


Figure 9.8 TEM plan view of the interlayer dielectric (ILD) silicate glass. The ILD was removed and isolated by wet chemical etching and subsequently coated with carbon. The texture observed in the gray areas replicates the polysilicon grain structure surface topology. The positions of various device structures are marked. source diffusion (*S*), gate (*G*), drain diffusion and bit-line contact (*D*), word line (*WL*), and *LOCOS* for the isolation field oxide (the *LOCOS* oxide are thick and thus showed much darker contrast).

decreases gradually along their edges and thus their boundaries do not show up in sharp contrast.

- Source, drain, and gate areas are all clearly imprinted on the ILD layer and are indicated accordingly in Fig. 9.8.
- The word line runs in the vertical direction as seen in Fig. 9.8(b). Notice that the word-line width differs in each area. When it acts as a gate, the width of the word line is much wider than when it is used as an interconnection line. This relationship is confirmed by cross-sectional images, such as that of Fig. 9.3(b).

Step 3. The polysilicon film forming the capacitor plate is a continuous sheet of poly-film. Figure 9.9 shows such a poly-film extracted by mechanical polishing and wet chemical etching.

- It is apparent that the poly-film covers the whole area except for the discrete oval shaped holes, where the control transistor's bit-line contact, two neighboring gates, and two source implantation (node contacts) are located.

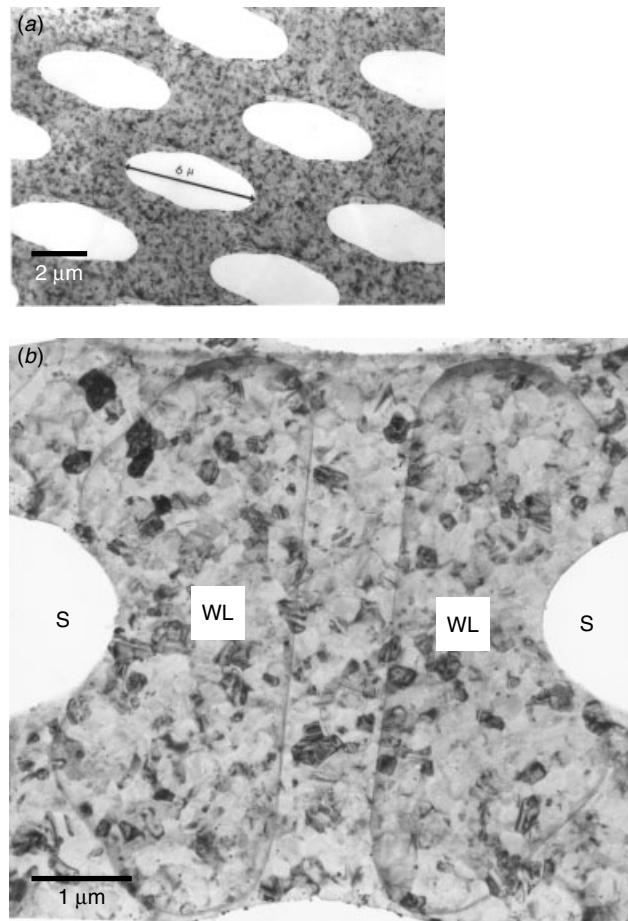


Figure 9.9 TEM plan view of the floated polysilicon film used for capacitor poly-Si plate. The poly-film was prepared using wet chemical etching. The continuous poly-film is shared among all the cells as the capacitor plate. The detailed view in (b) shows shadow lines cast by the parallel word lines running across the top of this area.

- The real capacitor area is smaller than the area covered by the continuous poly-film. As can be seen from the cross section (Fig. 9.6), the effective capacitor areas are smaller than the poly-film area because that part of the area was covered by the poly-film and used for the LOCOS isolation. The poly-film that extends onto the LOCOS does not contribute to the capacitor's effective area.

Step 4. Instead of peeling off a thin layer of laminated film by wet chemical etching, one could section the device in the horizontal direction and obtain a thin planar section of the device. Such a sample is useful in understanding the interlayer relationships and in identifying potential process issues. Figure 9.10, for example, shows such a planar section along the height near the gate and polyplate. In Fig. 9.10(a) the section includes WSi_x on top of the gate and thus shows distinctive dark squares as the gate

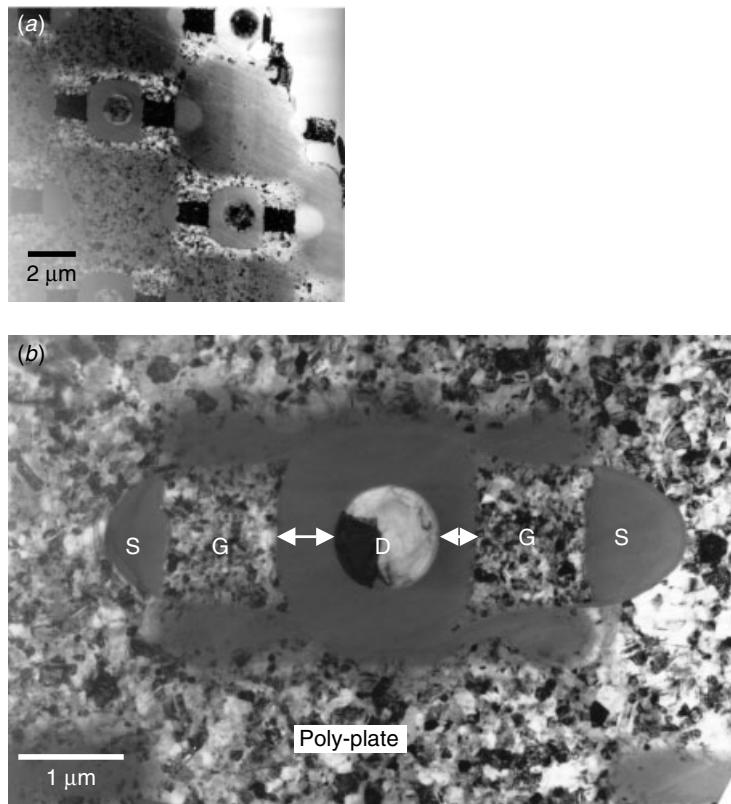


Figure 9.10 TEM plan view of a mechanically sectioned cell area showed polycide gates as dark squares in (a). The section at the lower part of the gate, as seen in (b), shows only polysilicon with no W-silicide, and thus the contrast is the same as that of the polysilicon plate. Source (S), gate (G), drain (D) and polyplate are marked. Notice that the G to D distances are different (as marked) between the two symmetrical devices. Such is also observed in the TEM cross-sectional view.

areas. Figure 9.10(b) shows the similar section but at a little lower area where the gate is polysilicon. Two process issues are revealed by this micrograph:

- As shown in the Fig. 9.10(b), for the two adjacent transistors the gate to bit-line contact distance is not equal. It is therefore quite apparent that the polygate is misaligned.
- Notice also the nearly connected gate-poly to the polyplate at the square polygate corners. The planar sectional view does not allow one to positively conclude the presence of this problem, but as will be apparent later in Fig. 9.15, the cross-sectional view of the same area can confirm it.

Step 5. Figure 9.11 shows all of the plan view TEM micrographs with corresponding cross sections of the same area. The reader is encouraged to cross-check each view and match the structures.

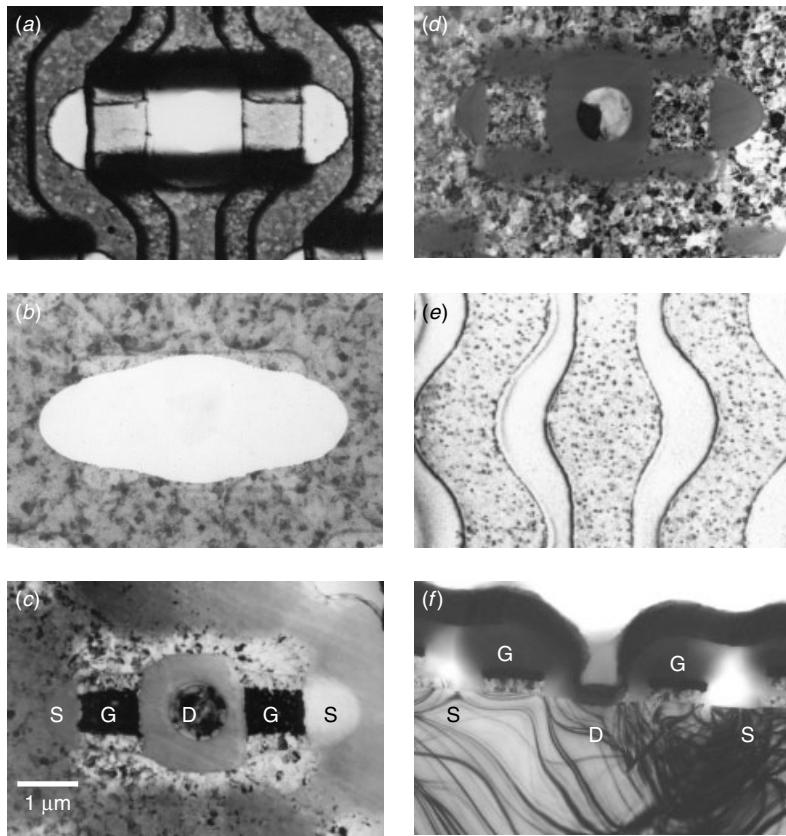


Figure 9.11 Combination of the TEM plan views (*a–e*) and cross section (*f*) clearly reveals all the details of the cell's structure. The *G* to *D* spacing asymmetry is observed in (*c*), (*d*), and (*f*).

Some Common Process Issues

From the early planar cell DRAM processes there are a few well-known process issues. These issues were new to the process engineers at the time for they only exist in DRAM processes.

Al Metal Step Coverage. Al metallization is known to have inadequate step coverage when contact size shrinks and contact aspect ratio increases. This problem became apparent in 1 M DRAM processes which used Al as the bit-line contact. Figure 9.12 gives an example of the problem. Step coverage, when decreased to 10% of the original design, as shown in the figure, can cause severe contact resistivity problems, and lead to faulty bit-line contacts. (More discussion of the Al step coverage issue and its solution is provided in the Chapter 8).

WSi_x Step Coverage. Although they share the same name, the WSi_x step coverage problem is due to a different reason than that of Al metal lines. As we saw in the Chapter 7, the WSi_x formed during silicidation can cause severe shrinkage and thus

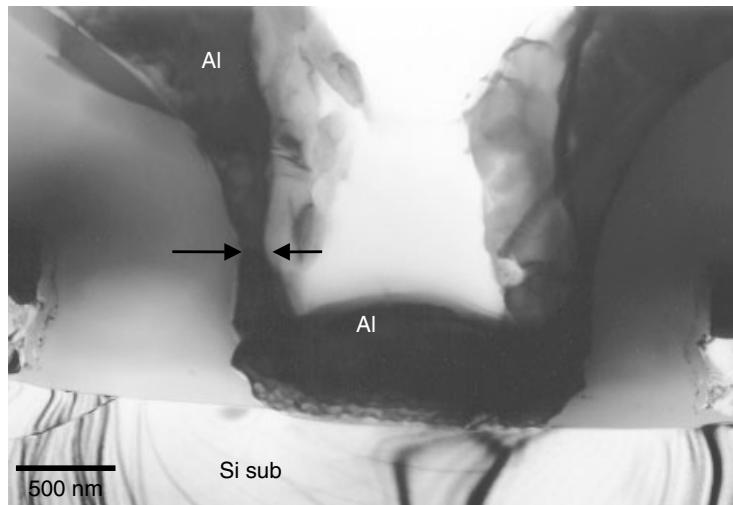


Figure 9.12 A bit-line contact using sputtered Al metallization shows a severe step coverage problem. Al thinning at the sidewall has reduced the Al thickness to nearly 10% to 15% of the original, as indicated. Such contacts suffer resistivity problems.

volume collapse. As the volume change occurs on flat surface, namely the film area, film thickness reduction results and micro-voids are formed. However, at a corner, the silicidation reaction on the step sidewall and step bottom compete for the silicon atoms, and as a result the corner is pulled apart during the volume's collapse. Figure 9.13 shows a good example of this effect. At the gate-to-field oxide corner, the step height is the field oxide plus the polyplate, so the issue is the worst in this area. WSi_x causes a gap at this corner, and the WSi_x film discontinuity leads to high sheet resistivity. The word-line polycide film, in this case, has resistivity more or less the same as that of normal poly film. The problem can be overcome by adding a thin layer of polysilicon film immediately after WSi_x deposition (either CVD W or WSi_x) but before annealing. Polysilicon film has excellent conformity. The extra poly can refill the gap and supply the much needed Si atoms in the corner area for WSi_x formation.

Polyplate to Word-Line Short. It was already shown in Fig. 9.10 by the horizontal section through the gate and polyplate that the distance between the gate (G) and the polyplate at the gate's corners is extremely short, and in some areas they overlap. The problem, however, is not as certain in the plan view because there is a possibility that the polyplate and gate are not on the same plane, so any overlap observed in plan view would be deceptive. To confirm the presence of this problem, a cross section through the corner area is needed. Figure 9.14 shows the result. The problem is confirmed that the corner has indeed nearly shorted. A sharp corner tip formed on the polyplate is also observed. Combined with the thin dielectric ILD, the tip-induced high field can cause leakage or even ILD breakdown. Indeed, further analysis with the processes has confirmed that this problem is indeed quite severe, as seen in Fig. 9.15.

As described in Table 9.1, the ILD layer is formed by the thermal oxidation of the polyplate (poly 1). The poly 1 grain deposited is not smooth, and thus when thermally

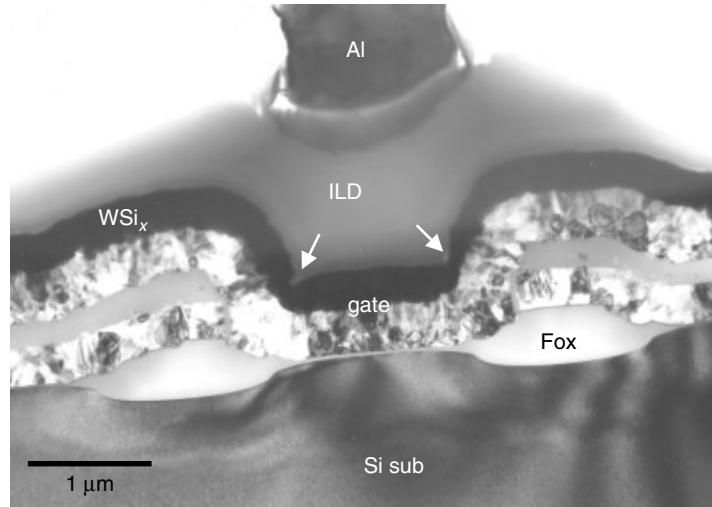


Figure 9.13 Tungsten polycide used as the word line to reduce word-line resistance. When WSi_x forms with the volume shrinkage, WSi_x gaps form at the step corners and give rise to resistivity issue, as indicated. The problem is particularly bad at the gate-to-LOCOS boundary, where the step height due to the field oxide plus the poly–plate is the highest.

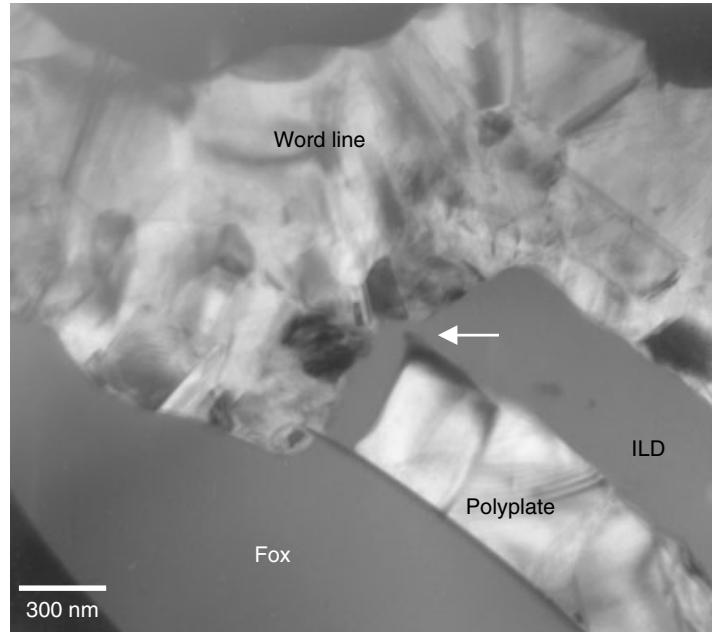


Figure 9.14 Close-up at the end corner of the polyplate shows that the polyplate has nearly shorted to the word line, as indicated.

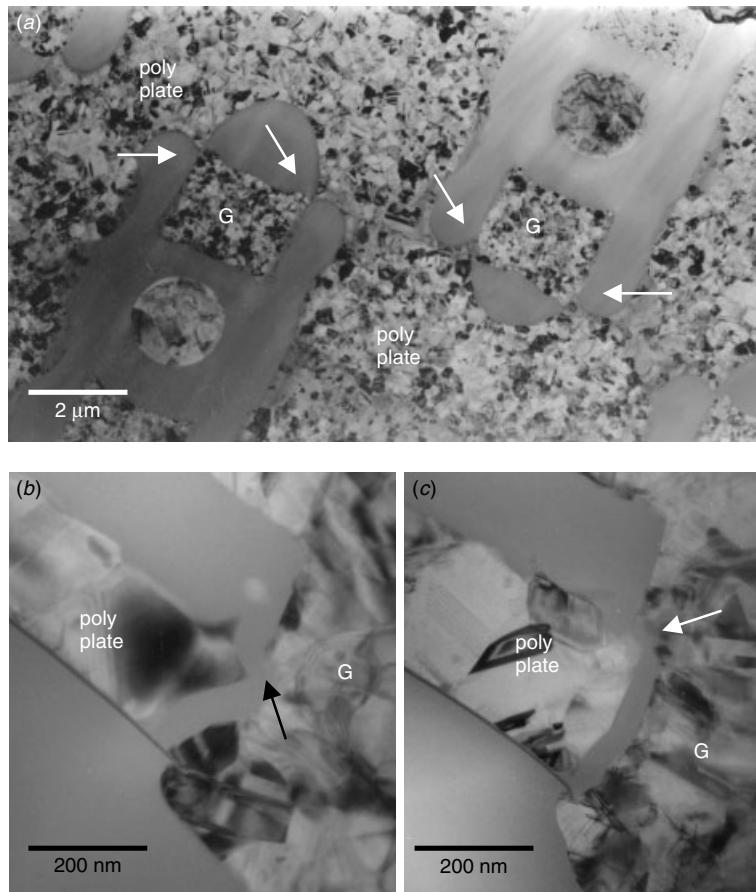


Figure 9.15 Close-up at the end corner of the polyplate shows the polyplate has nearly shorted to the word line, as indicated in plan view (a), and in cross section (b) & (c). The gap is less than 30 nm and a sharp corner tip is also observed. Intensive electric potential can build up, cause leakage, and induce ILD breakdown.

oxidized, the oxidation process will not be homogeneous. Poly stringers, poly residues, sharp tips at corners, and other similar defects can form within the thermal oxide. Close examination of Figs. 9.13, 9.14, 9.15(b), and 9.15(c) in fact reveals dark and light patches within thermal oxide. Different illumination angles (by tilting the sample) reveal these patches and sharp tips are poly stringers as they change contrast accordingly. The residue poly stringers are wrapped within the oxide. ILD breakdown, leakage, and even shorts can result from these poly stringers.

A way around this problem is to remove and regrow the thermal oxide. Additional oxide deposition (CVD) can be added. The final ILD layer will be clean and without sharp corners, as shown in Fig. 9.16.

Mask Edge Defects at Polyplate Edge. The planar capacitor plate formed by using poly I is later used as a self-align ion implantation mask for transistor drain-to-capacitor node junction implantation and diffusion. Mask edge defects inevitably form at the

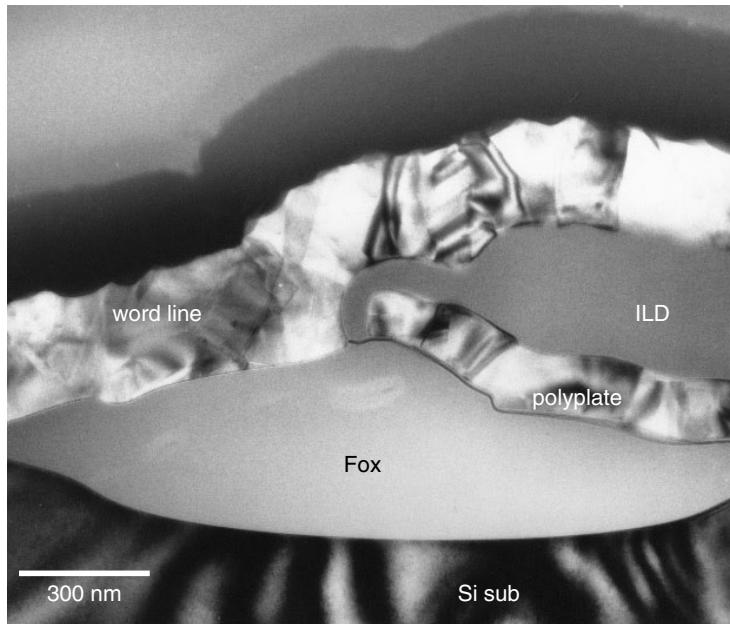


Figure 9.16 Clean ILD layer achieved by removing the first thermal oxide and re-growing a new oxide layer. Additional ILD oxide deposition is also observed on the right-hand side of the image, resulting in a different ILD thickness in this area.

edge of the drain-to-capacitor (node contact) corner. In most cases these mask edge defects do not cause any problems and are not considered as process defects. However, stress induced by subsequent processes, such as LOCOS, interlayer dielectrics, and metallization deposition and annealing can be large due to the high-density layout design. The mask edge defects will act as nucleation sites for generating extended dislocation. Impurities will aggregate along the extended dislocations and create an excellent leakage path. Junction leakage and even a direct short can result.

Figure 9.17 shows a segment of such extended dislocation. The area selected is from the polyplate to the transistor's drain edge, which is near the LOCOS. As is apparent in Fig. 9.17(b), the Si substrate surface slightly curves down, indicating that the area is close to the LOCOS structure. When device-shrinkage pushes the LOCOS structure right up to the mask edge defects, the dislocation can easily extend. A more sophisticated back-end process could release some of the stress concentration. Advancements in the device layout design may also help, as new design rules could take this mechanism into consideration and prevent the problem from happening.

9.3 PLANAR CELL STRUCTURE VARIATIONS

The basic 1M planar DRAM structure remained more or less the same among different manufacturers. A few minor modifications, however, can be found in the early DRAM market.

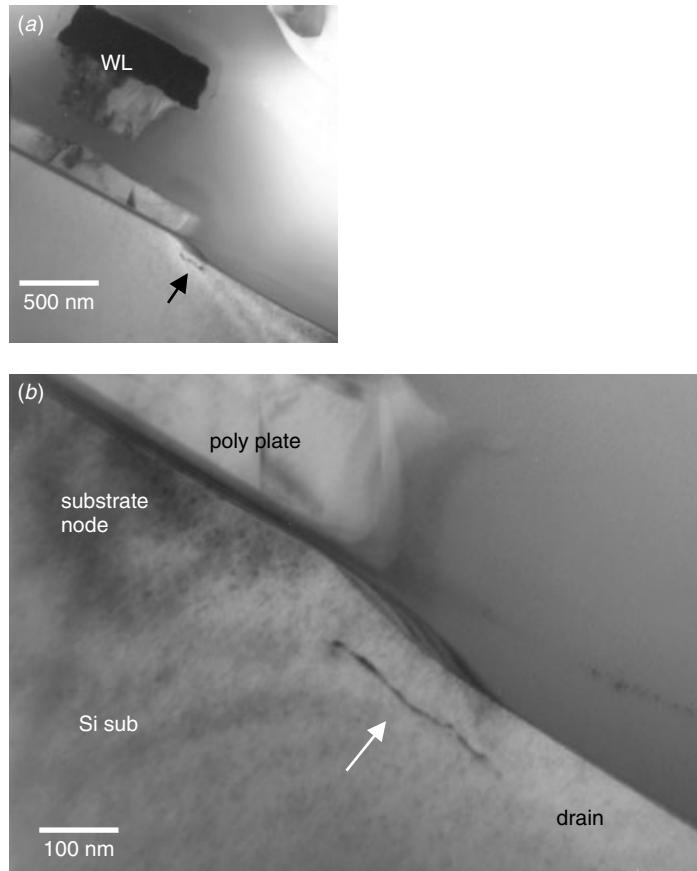


Figure 9.17 Extended dislocation nucleated from mask edge defect. LOCOS and back-end process induced local stress is the root cause of extended dislocation. The slight concave curve on Si substrate's surface indicates that the area is near a LOCOS structure.

Figure 9.18 shows an early planar DRAM device. Tungsten polycide (poly 1) is used as the capacitor plate and the same material is used again (poly 2) as the word line and the transistor gate. Al with a thin Ti barrier is used as the bit-line interconnection and bit-line contact.

Figure 9.19 shows a more advanced planar DRAM device. Both the capacitor plate and word line are made of polysilicon. Al is used for the bit-line interconnection. A rather large polygrain size, about 1 μm on average, is observed in both poly 1 and poly 2, indicating that both poly films have been heavily doped to reduce resistivity.

A later version of planar DRAM was shown in Figs. 9.3, 9.5, and 9.6, which illustrated basic planar DRAM features. This DRAM used polysilicon as the capacitor plate and tungsten polycide as the word line and gate materials. The polygrains in poly 1 are rather small, less than a quarter μm , indicating that a doping concentration less than that observed in Fig. 9.19 was used and thus that there is higher poly 1 resistivity.

A further development of the planar DRAM is shown in Fig. 9.20. Again, polysilicon (poly 1) is used for the capacitor plate and tungsten polycide (poly 2) as the word line

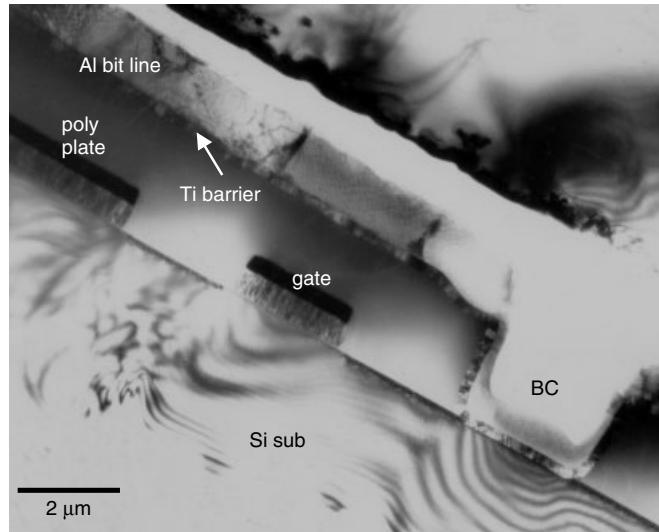


Figure 9.18 An early 256 K planar DRAM using tungsten polycide (poly 1) as the capacitor plate and a second tungsten polycide (poly 2) as the gate and word line. Al with Ti barrier is used as bit line and bit-line contact. The design rule is large and no word line is observed on top of the capacitor plate within this micrograph.

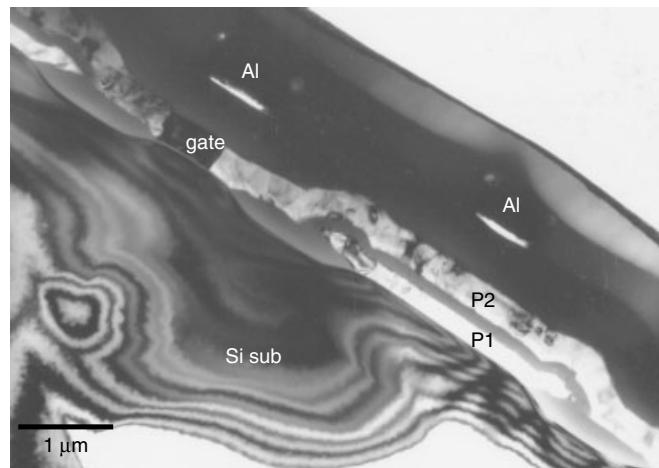


Figure 9.19 1M DRAM with polysilicon (P1) as the capacitor plate, polysilicon (P2) as the word line and transistor gate, and Al as the bit line and interconnect metallization.

and gate. A lightly doped drain (LDD) process with a nitride cap is used in this device with 1.2 μm technology. As a result a oxide spacer with a nitride cap can be seen on both sides of the polycide gate as well as at the word line on the capacitor plate. A void trapped at the edge of the polysilicon capacitor plate is also observed. Interestingly the LDD structure was designed to avoid the hot carrier effect. Such LDD and nitride cap

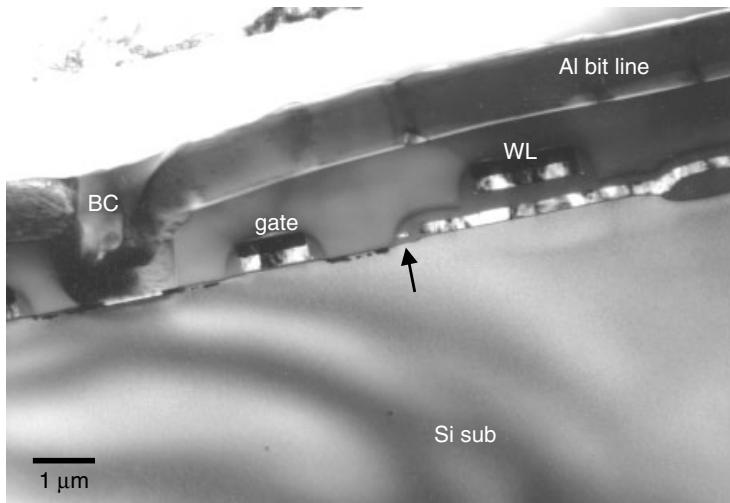


Figure 9.20 1M DRAM with polysilicon poly 1 as the capacitor plate, poly 2 (tungsten polyide) as the word line and transistor gate, Al as the bit line and interconnect, and the bit-line contact (BC). The LDD process with a nitride cap is used in this new technology. A void is observed at the corner of the capacitor plate (as indicated) due to the LDD process.

structures have become a standard in sub 0.25 μm devices. Notice in the figure that in this device the poly 1 polysilicon grain size is rather large (about 1 μm on average).

A final example of the planar cell DRAM is shown in Fig. 9.21. This version is particularly interesting because it represents a transition from planar cell DRAM to stacked cell DRAM. It uses a polysilicon (poly 1) capacitor plate and polysilicon (poly 2) as the word line and gate. Al metallization (metal 1) is used as the bit line, and another Al metallization (metal 2) running in parallel with and connecting to (using stitch contacts) poly 2 is used as the word line to reduce resistivity. An LDD structure and a nitride cap can be observed in this device also. Notice the changes that have occurred in this device:

- A layer of Al metallization was added. This new metallization layer runs in parallel with the poly 2 word line and connects to poly 2 periodically using stitch contacts. The main purpose of this metallization layer is to reduce word-line resistivity. The same design will be used almost ubiquitously in all later generations of stack and trench DRAMs.
- The LDD and nitride cap layers are present. Again, this is a technology that prevents hot carrier effect and device leakage between the word line and bit line. A similar design will be observed over and over again in stack and trench DRAMs.
- A thin oxide layer is observed in between the word line and polyplate. This thin oxide between the polysilicon layers may look insignificant in this device but will play a major role in the stacked capacitor cell DRAM. An ultra thin dielectric together with two polysilicon layers forms the stacked capacitor.
- In the early trench cell DRAM design, the trench sidewall was used as capacitor node, the same as in the planar cell DRAM. Thus, just like in the planar cell, the

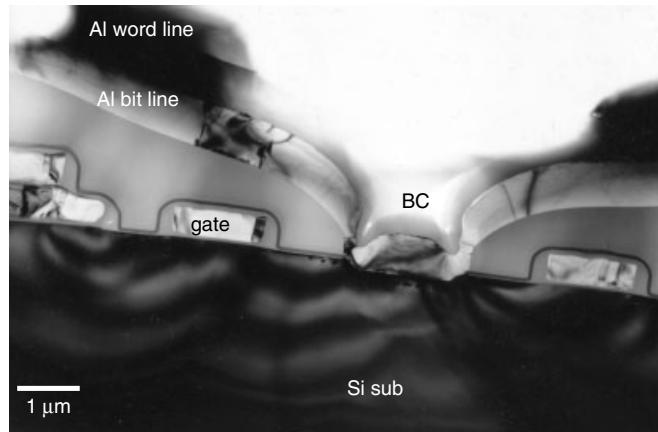


Figure 9.21 1M DRAM with a polysilicon poly 1 as the capacitor plate, a polysilicon poly 2 as the word line and gate, Al metal 1 as the bit line and interconnect metallization. Another Al metallization (metal 2) is added on top to help reduce the word-line resistivity. This is a transition design from the planar DRAM to the next-generation stack DRAM.

substrate node trench DRAM required no physically distinctive node contact and the control transistor diffusion connected directly to the capacitor node diffusion. This is another direct “link” between the design of trench DRAM and that of the planar cell DRAM.

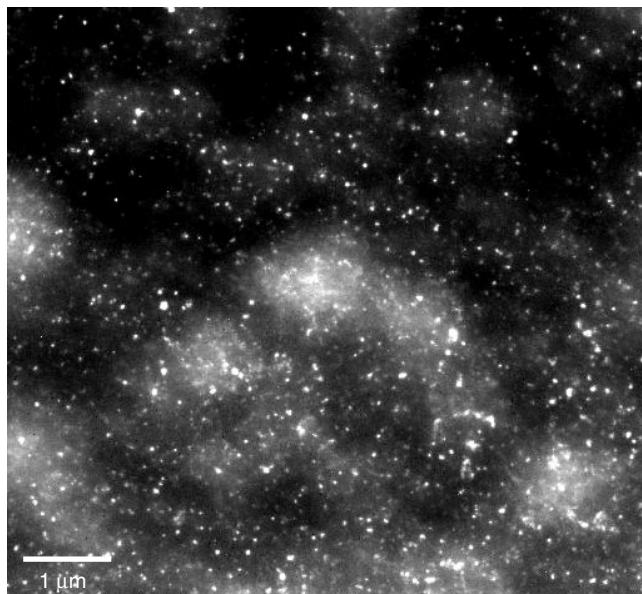
The submicron geometry MOS processes stretched the basic processing to accommodate increased layers, and this led to severe planarization requirements. By the time the 1 M DRAM became available on the market, it was clear that vertical device methods were necessary in order to allow the megabit chip densities to meet in size, and not significantly exceed, that of historic chip sizes. Ways had to be found to scale the area occupied by the capacitor on the surface of the wafer without reducing the effective area of the capacitor. Smaller horizontal geometries on the surface of the wafer could be achieved by making use of the vertical dimension, both above the surface of the silicon and below it (Prince 1995).

The aggressiveness with which industry undertook this challenge is best illustrated in the DRAM processes. By the stacked capacitor cell DRAM and trench capacitor cell DRAM processes development, the MOSFET processes were pushed to the limits. These were revolutionary ideas that stretched process innovation and the engineering imagination. To fully understand their characteristics it is helpful to learn how each evolved through succeeding generations of products and how the processes changed to accommodate the increasingly more stringent quality and cost requirements.

REFERENCES

- Y. A. El-Mansy and R. A. Burghard, *IEEE J. Solid State Circuits*, **17** (5), 1982.
- B. Prince, *Semiconductor Memories, A Handbook of Design, Manufacture, and Application*, 2nd ed., Wiley, 1983.

10 ULSI Devices II: DRAM Cell with Stacked Capacitor



Plan view Dark field ferromagnetic crystalline materials for FMRAM. It resembles milky way in a dark crispy night sky.

Developments in the ULSI process beyond submicrometer (submicron, <1 μm) geometry have basically extended the fundamental processing steps to accommodate the increased layers of lamination, severe planarization requirements, and new vertical device stacking methods. Megabit and gigabit chip densities are now being made in chip sizes not significantly larger than before. Innovative device structures incorporate, for example, deep trenches into Si substrate and stacked layers of devices rising high over the surface. Today, the main constraint is that the size of the chip cannot grow proportionally, even as the number of transistors in the chip increases exponentially.

10.1 PROCESS FLOW

For the DRAM with bit density less than one megabit, the planar type capacitor is used in the memory cell. Although the component density might increase, the amount of charges needed for a sufficient noise margin remains fixed. To increase the specific capacitance, two different development routes are undertaken: the first is to store electrical charge vertically in a trench into Si substrate; the second, allows the cell to shrink in size without a loss of storage capacity by stacking the capacitor on top of the access transistor. We will study devices with stacked capacitors in some detail in this chapter and treat the former in the next chapter.

A generic and simplified process flow for a stacked capacitor DRAM device is given in Table 10.1. It used a modified CMOS process. Only a skeleton of steps is provided. The purpose is to give some insight on how these devices are fabricated for a number of examples that will be treated later. In general, to make a stacked capacitor DRAM, one needs to have three to four polysilicon layers and one to two metallization layers. Poly 1 is normally used as the transistor gate structure. Poly 2 and 3 are normally used for capacitor electrode and plates. Poly 4 is used as the bit lines. Al lines (metal 1) are used in parallel with poly 1 to reduce the word-line resistivity. Metal 2 generally is not needed within cell area and is used for peripheral interconnects.

In order to increase the capacitance by reducing the capacitor dielectric layer thickness, laminated oxide-nitride-oxide (ONO) layers are used in between the polysilicon plate and polysilicon electrodes. The total thickness of the ONO layer is about 80 to 120 Å. Due to the lamination the pinhole defects generally found in thin thermal oxide are avoided effectively. Because it combines the geometric and topological change with polysilicon electrode, ONO provides an excellent dielectric layer and is used up to gigabit DRAM chips.

TABLE 10.1 DRAM Process of Stacked Capacitor Cell

Process Steps	Function
Form twin well	
Form LOCOS isolation	
Grow gate oxide, deposit poly 1	Form gates for cell and peripheral devices
n-implant, form spacer, n+/p+ implant	Form peripheral devices
Mask, node contact, etch oxide	Form node contact for capacitor
Poly 2, mask and etch	Form capacitor electrode
ONO deposition	Form capacitor dielectric
Poly 3 deposition, mask and etch	Form capacitor plate
BPSG deposition	ILD 1
Mask, bit-line contact, etch	Form bit-line contact
Poly 4, mask, and etch	Form bit line
BPSG deposition	ILD 2
Mask metal contact, etch	Form metal contact
Metal 1 deposition, Mask, etch	Form metal 1
IMD deposition	IMD 1
Metal 2 deposition, mask, etch	Form peripheral interconnects
Passivation, mask, etch	Form passivation and bond pad

10.2 THE UNIT CELL

Lu (1989), in an early review paper, provided excellent summaries on the basic DRAM cell structures, including stacked capacitor and trench capacitor DRAMs. A schematic cross section of the stacked capacitor DRAM cell and its associated circuit diagram is shown in Fig. 10.1. Notice that the circuit diagram remains the same compared to that of the planar cell. Figure 10.2 shows a typical example of a 4M DRAM unit cell structure with its corresponding circuit. All parts are marked for easy identification. There are a few key components within this unit cell.

The Transistor

Two active junctions and one polygate together form the control transistor.

1. A commonly observed problem is the extended dislocation generated by the active area mask edge defects, as seen in Fig. 10.3. The mask edge defects are inevitable and usually considered to be harmless. However, with the enormous stress induced by the back-end process (polycide, metallization, ILD, IMD, etc.), these mask edge defects can act as a dislocation source and, generate extended dislocation, cause device leakage and other issues. Minimizing the process stress can prevent such a defect.
2. Polysilicon runs in a zigzag pattern when observed in a plan view. This characteristic is similar to that observed in the planar cell DRAM, as seen in previous chapter. Part of the polyline is used as the polygate and part of it is used as the conducting runner. The polygate is a little wider than the poly-runner. The zigzag layout design maximizes the utilized cell area while reducing the overall cell

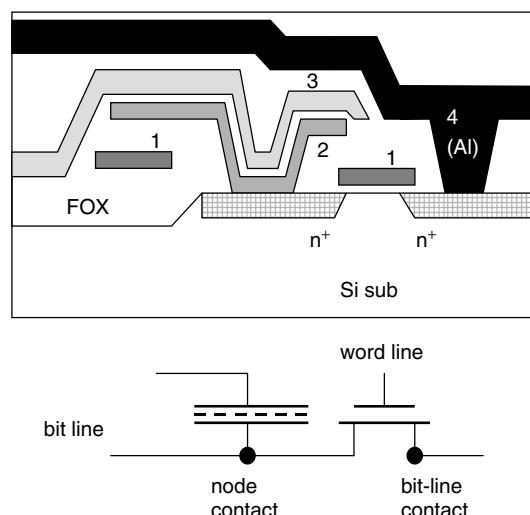


Figure 10.1 Schematic cross section of a conventional stacked capacitor cell DRAM structure and its corresponding schematic circuit. Polys 1 forms the gate structure. Polys 2 and 3 form the capacitor. Either poly 4 or metal 1 (W or Al) form the bit-line contact.

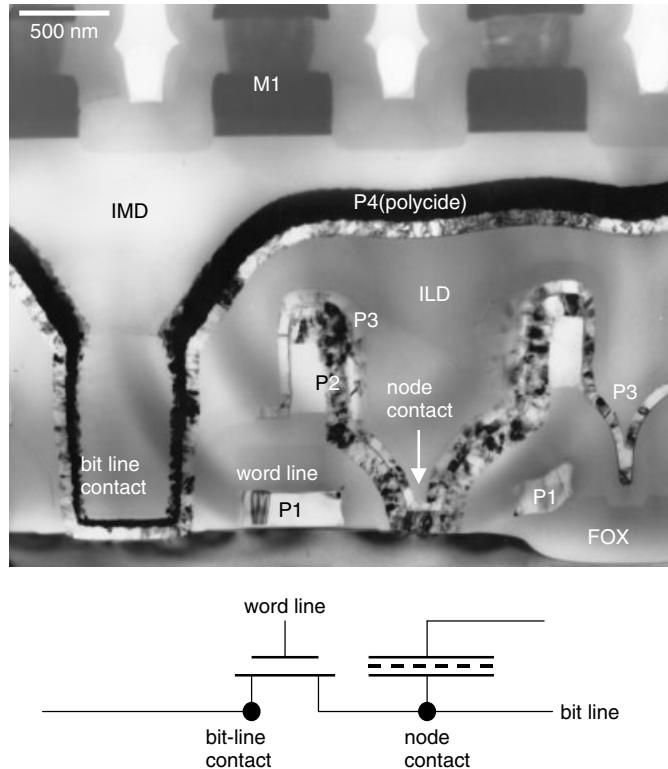


Figure 10.2 TEM image of a typical 4M and 16M DRAM cell's cross section. Each part of the layer is marked in accordance with the corresponding circuit.

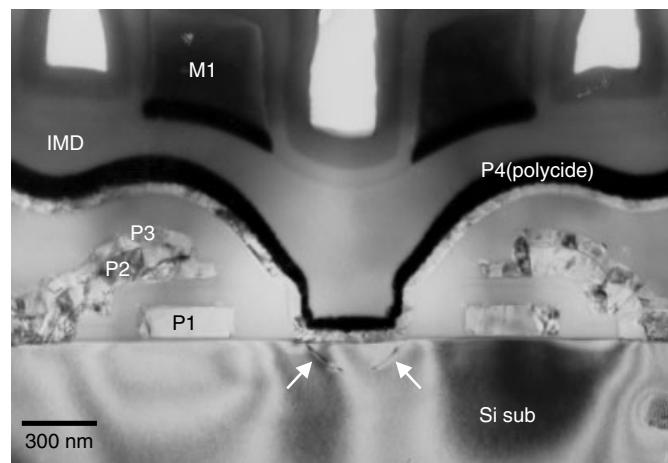


Figure 10.3 Extended dislocations observed under bit-line contact, as indicated. The dislocations start from the contact's mask edge.

area. It has been used since the first generation of the planar cell DRAM design and is still being used in the latest DRAM products.

3. When polysilicon instead of polycide is used for the polygate and runner, the polyline resistivity is high due to the nature of the material even with heavy doping. An easy solution of this word-line resistivity problem is to use Al metallization running in parallel with poly 1. When the Al metal line runs on top of the bit line, it does not occupy any cell area, as seen in Figs. 10.2 and 10.3. The only thing required of the Al metal is to make periodic contacts with poly 1, normally called “stitch contacts.”

The Capacitor

Unlike the planar capacitor, at least two layers of polysilicon (usually poly 2 and poly 3) are required for making the stacked capacitor. In between the two layers a thin dielectric layer is required to separate the two polysilicon layers. The structure of the capacitor is where most of the technology and design advancement is taking place. We will provide detailed discussion in a later section.

1. The lower poly-layer (usually poly 2) is also used as the “node contact,” and it is in direct contact to the active area of the control transistor. The poly-to-Si-substrate active area contact is vulnerable to the interface oxide issue and is one of the key process and yield control step. The poly/Si-substrate interface oxide was discussed in Chapter 8.
2. The upper poly-layer (usually poly 3) is not divided among cells. Instead, the whole memory section (contains several hundred or thousand of cells) can be made into one complete poly-plate. The only areas that poly-plate cannot cover is that of bit-line contacts. Apparently, bit-line contact areas are wasted as they do not contribute to storage capacitance. New generations of DRAM design using capacitors on, or over, the bit line (COB) utilize effectively the entire cell area.
3. The dielectric layer in between two polylayers usually consists of a laminated oxide-nitride-oxide (ONO) structure. Thin oxide layer grown from polysilicon tends to have a high density of defects and thus poor reliability performance. The laminated ONO structure allows the total thickness of the dielectric layer to be reduced significantly without increasing the density of defects. The thin ONO dielectric layer was discussed in Chapter 6.

The Contacts

There are two contacts in each of the stacked unit cells. One is called the “bit-line contact” where the bit line is in contact with Si-substrate active source. The other is called the “node contact,” where the poly 2 capacitor electrode is in contact with Si-substrate active drain.

1. The bit-line contact is usually made of the same material as the bit line. Polycide is generally chosen for the bit line, as seen in Fig. 10.4 (in certain early devices Al metallization was also used as bit-line material). The plug contact is usually not necessary for a process technology node above 0.35 μm . As the capacitor stacking height gets taller, the bit-line contact aspect ratio can grow very high,

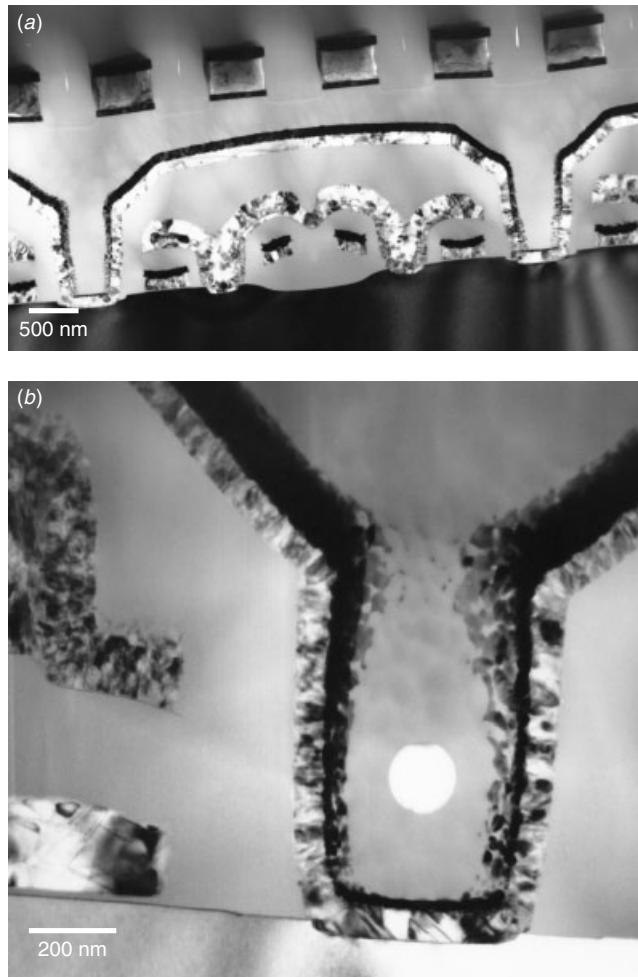


Figure 10.4 W–polycide bit-line contacts. The contacts are made directly from polycide’s thin film deposition, and hence the contact’s topology and step coverage are important. Voids may be formed, as indicated, as a result of the IMD oxide deposition.

making it difficult to fill the contact with the oxide layer, as seen in Fig. 10.5. An alternative to high-aspect ratio contact is to use plug contact technology. A polysilicon plug with W–plug provides a good solution, as shown in Fig. 10.6.

2. The node contact is usually made of the same polysilicon film layer as that of the lower poly–node of the capacitor. Contact resistivity and poly–sheet resistivity are the main concerns in this process. In-situ doped polysilicon is generally used. However, in order to be able to easily control the sheet resistivity and contact resistivity, more than one layer of polysilicon with different doping concentration may be needed. Additional poly–layers are used in shaping the lower poly–plate to create different topologies that increase the interface areas and thus the storage capacitance. More discussion on this design will be given later.

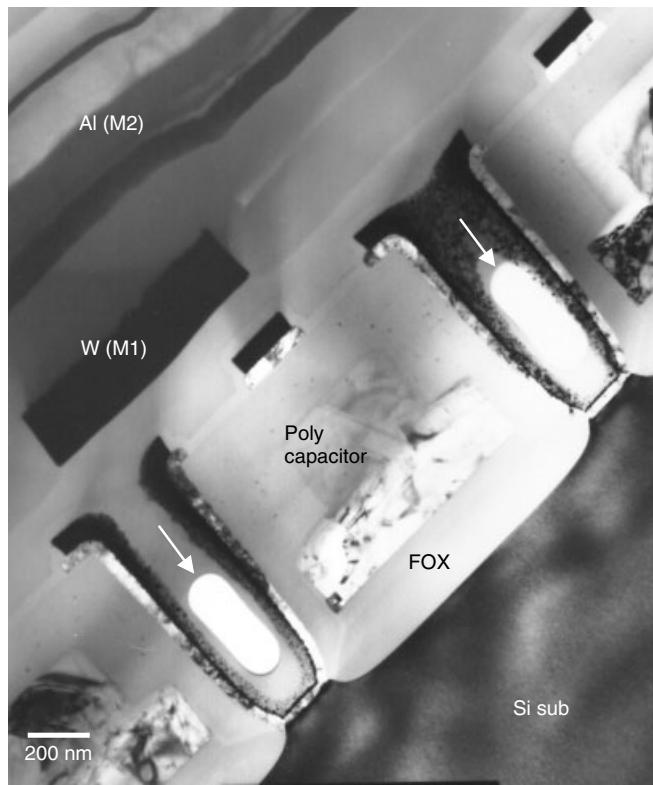


Figure 10.5 W–polycide bit-line contacts. The high aspect ratio makes it difficult to fill the contact with the IMD oxide, leaving holes in the center of all the contacts, as indicated.

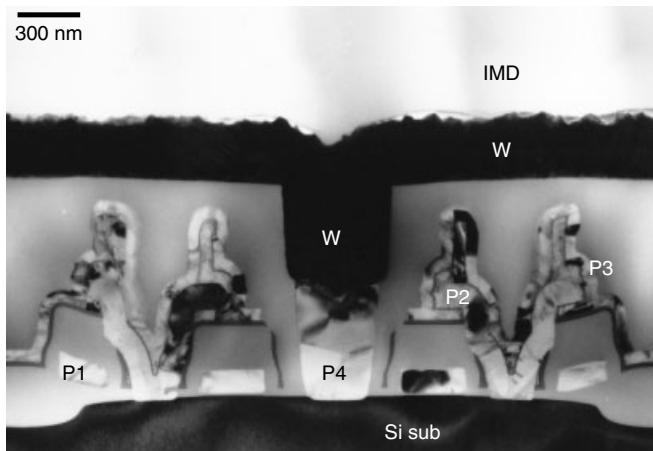


Figure 10.6 Poly–plug and W–plug for bit-line contact with W. The plug contact is the choice when the contact aspect ratio is high and the bit-line resistivity requirement becomes more stringent.

As we mentioned earlier, the main challenge is to increase the capacitance of the storage cell without significantly increasing the wafer surface area. There are basically three strategies:

1. Reduce the interface dielectric thickness, and thus increase the oxide capacitance. This is why ONO replaced straightforward oxide layer and also why different high- k materials like Ta_2O_5 are used.
2. Increase the doping concentration to increase the depletion capacitance. This is where the high-C implantation was introduced in the early planar capacitor design as mentioned in Chapter 9.
3. Increase the interface areas, and thus the storage capacitance. Tremendous effort has been devoted to increase the interface areas of the storage capacitor. The processing technologies used in the formation of the stacked capacitors, while complex, are still the variations on conventional Si MOSFET processing technologies.

10.3 CELL STRUCTURE EVOLUTION

The first-generation stacked capacitor remained quite flat and did not utilize effectively the space gained from stacking. Figures 10.7 and 10.8 show some early stacked capacitor designs. A few examples will be given here to illustrate the evolution of the cell structure.

Vertical Stack (Sidewall Effect)

The early stacked capacitor structure (Figs. 10.7 and 10.8) showed little utilization of the advantage gained in area. The first apparent area increment from the stacked capacitor was from the side wall and the lateral extension over the transistor gates. Figure 10.9 shows an example where the larger interface areas due to wrapping P3

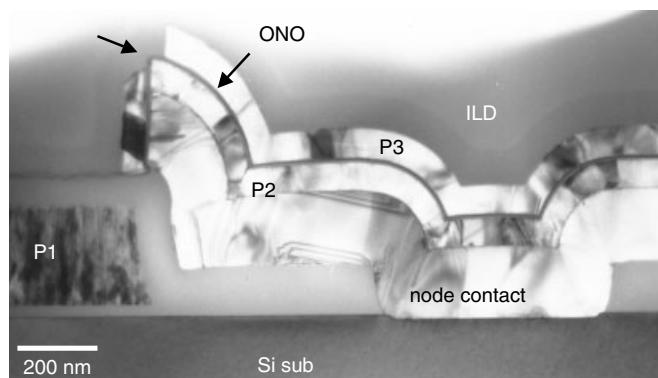


Figure 10.7 An early stacked capacitor design shows that the utilization of areas gained from the vertical sidewalls, curvatures, and overlap with the polygate is minimal. Note that this is failed process, with P3 etched open at the corner as indicated by the arrow.

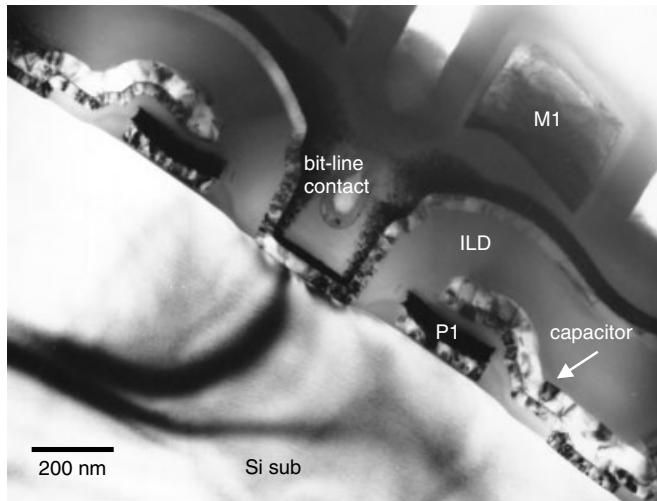


Figure 10.8 An early stacked capacitor design shows the minimal utilization of areas gained from vertical sidewalls, curvatures, and the overlap with the polygate.

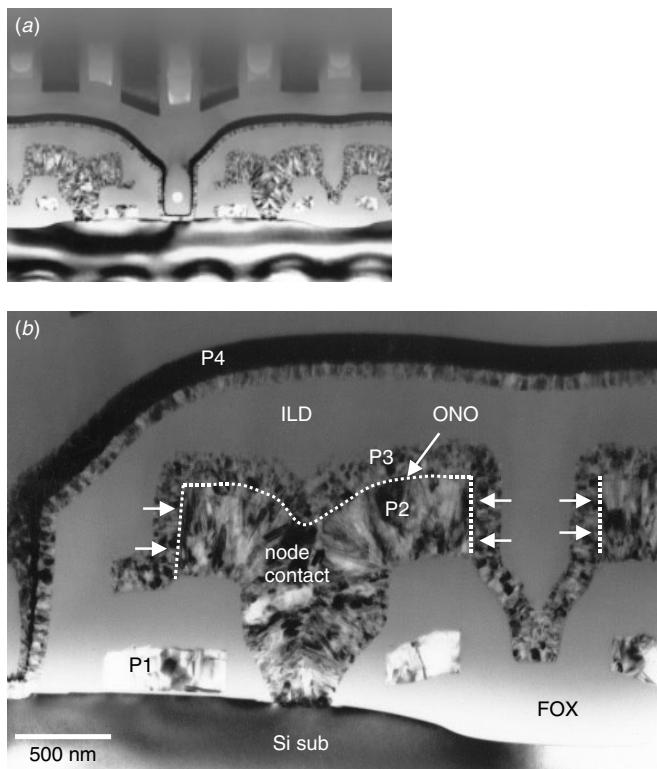


Figure 10.9 A stacked capacitor design shows the utilization of sidewall areas gained by wrapping the P3 plate around the P2 stud, as shown by the arrows in (b). The whole mushroom stud overlapping the polygate also helps reduce the cell area while increasing the capacitor area.

plate around P2 is raised by a stud shape. The entire mushroom stud overlaps with word-line polygate to reduce the cell real estate area while increasing the capacitor area (Kimura et al. 1988).

The next natural development was to use the volume of the mushroom stud. By creating a hollow stud, instead of solid stud, and filling P3 within the hollow center, the interface area was increased again. Figure 10.10 shows this design. Notice in Fig. 10.10 that the P2 was done with two poly-layers. The first layer creates two vertical blocks so as to raise the stacking profile and increase the side-wall areas. Any additional poly-layers may be used depending on the production cost. Figure 10.11 shows this alteration to Fig. 10.10. The development of vertical stud has reached its limit at this point. Further enlargement of the capacitor area had to rely on alternative developments in the cell design.

Horizontal Fin Structures

The horizontal fin structure (Ema et al. 1988) is a structure with multiple horizontal interlocking laminated polysilicon layers with a conformal dielectric ONO liner in between. In cross section it has the appearance of interlocking fingers.

The challenge and complexity of producing such a horizontally laminated structure is amazing. A cross-sectional TEM image is shown in Fig. 10.12 for the simple case where there is only one FIN. This is the same structure we mentioned before and showed in Fig. 10.11. Notice in Fig. 10.12(a) that both poly 2 and poly 3 have an arm on each side and that there is exactly one horizontal interface between poly 2 and poly 3 and thus one horizontal dielectric capacitor layer, as seen in Fig. 10.12(b). Notice further in Fig. 10.12 that the polysilicon grains (of both poly 2 and poly 3) are very

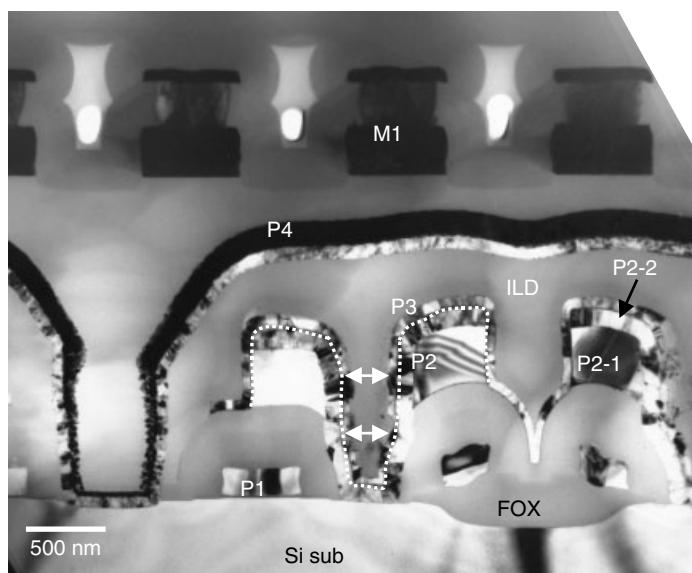


Figure 10.10 A stacked capacitor design shows the utilization of internal sidewall areas within the hollow stud, as shown by arrows. Notice that the P2 was done in two poly layers, P2-1 and P2-2 as shown. An additional mask layer is required for such a process.

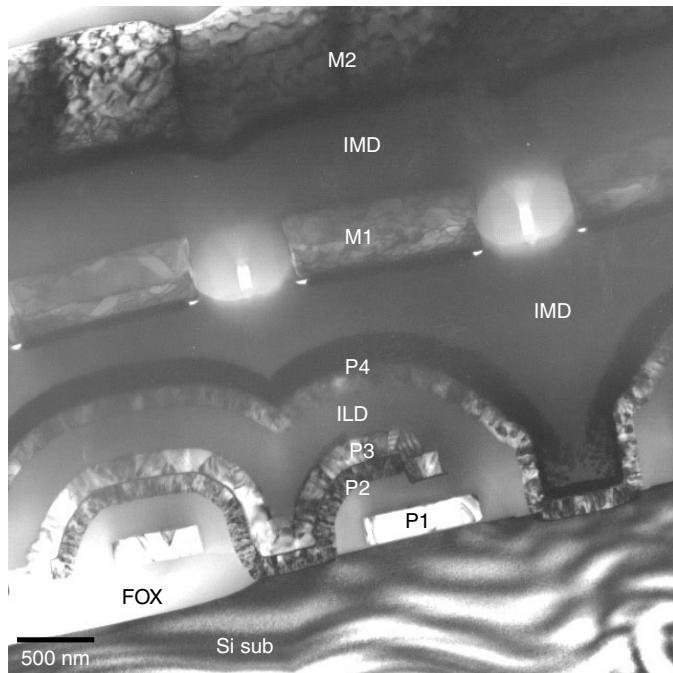


Figure 10.11 A stacked capacitor design showing simplified single P2 and P3 layers with sufficient interface areas for the capacitor.

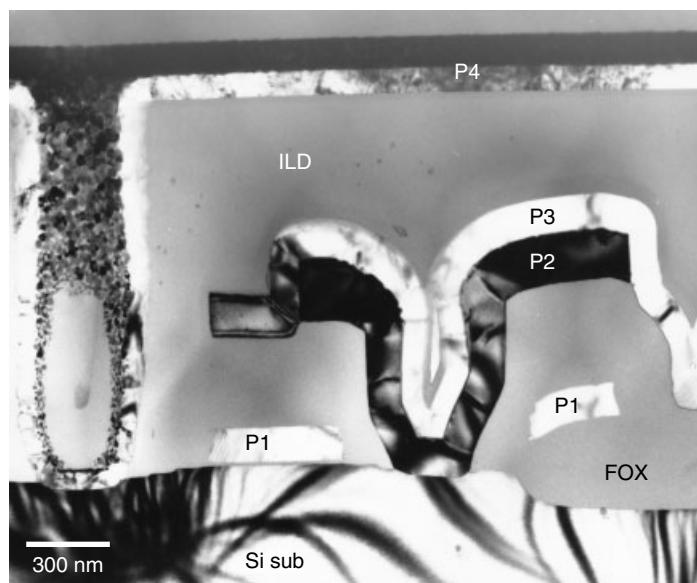


Figure 10.12 (a) A stacked capacitor design showing the single fin structure. Only one horizontal finger is found in both P2 and P3 layers. Notice that the grain size for both P2 and P3 is extremely large. Only one grain is found in each poly electrode. This is helpful in increasing the depletion capacitance and the storage charge.

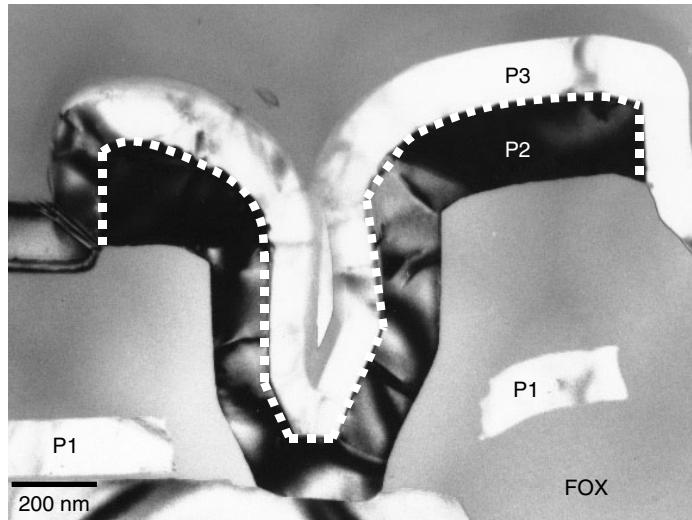


Figure 10.12 (b) Close-up of the capacitor interface. The dashed line indicates the location of the interface dielectric, which represents the effective capacitor areas.

large. There is virtually only one grain each for the entire capacitor electrode both in poly 2 and in poly 3. In polysilicon the large grain indicates a high concentration of doping, and this helps increase the depletion capacitance in a polysilicon capacitor and thus increase the storage charge.

Additional fins are painstakingly created layer by layer by deposition, patterning and etching, poly-re-deposition to form the central vertical cylinder connection, conformal dielectric layer deposition, and finally the polyplate deposition. Figures 10.13(a) and (b) show a one and one-half fin capacitor structure. There is one horizontal arm in poly 2 (poly-node), but there are two horizontal arms in poly 3 (poly-plate). There are two horizontal dielectric capacitor layers in this case, doubled when compared with the singular fin structure.

Similar process manipulation can be multiplied to add on more lateral layers creating more fins. An example is shown in Figs. 10.14(a) and (b) for structure with two fins. Now both poly 2 and poly 3 have two horizontal arms, and the total horizontal dielectric layers has become three, as shown from Fig. 10.14(b). Notice that there is still space between the poly 3 and poly 4 bit line as further layers can be added in a similar manner.

Crown-Shaped Stacked Capacitor

One disadvantage with horizontal fin structure is the complexity of the process. Multiple-deposition, photo-masking, and etching steps are required to achieve the laminated fin structure in the horizontal direction. Strict process control in the etching is critical to the success of the process. Due to these difficulties there is a need to create the fins/arms in the vertical direction. There are at least two advantages to making fins in the vertical instead of horizontal direction. First, more vertical fins can be created without adding layers, and thus the process tends to be simpler. Second, by using

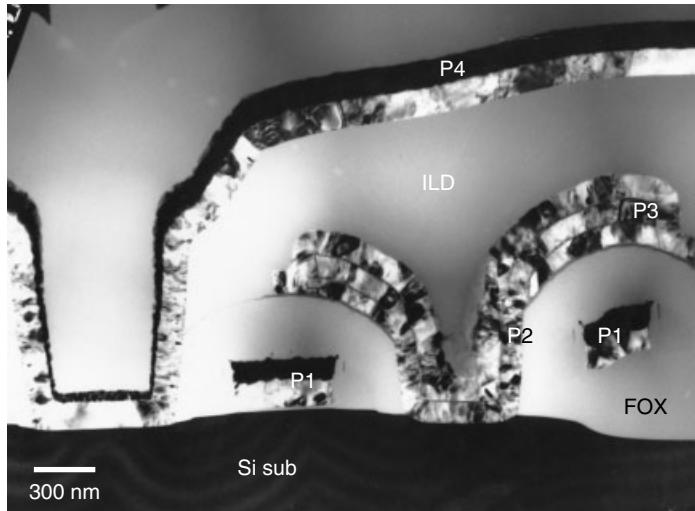


Figure 10.13 (a) A stacked capacitor design showing the one and a half fins capacitor structure. P2, which forms the node contact, shows only one layer while P3 (polyplate) shows two layers and wraps around P2.

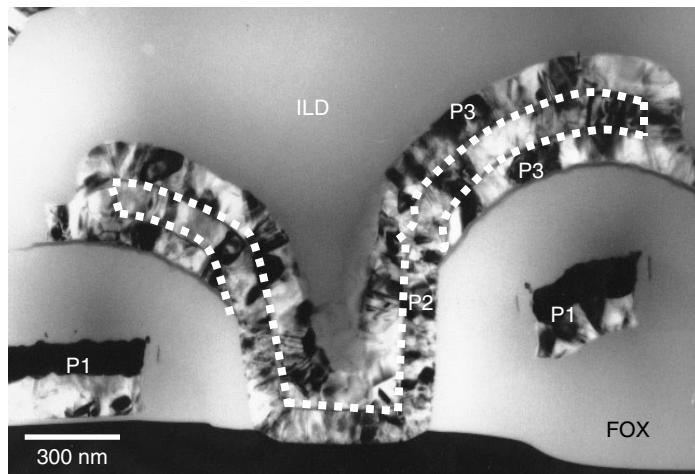


Figure 10.13 (b) Close-up of the capacitor interface. The dashed line indicates the location of the interface dielectric, which represents the effective capacitor areas.

self-aligned techniques, photo-masking steps can be eliminated. This simplifies the process steps, increases the yield, and reduces the process cost.

A first-generation crown structure is shown in Fig. 10.15. Plan view and cross-section are shown on the same structure. The term “crown” is obvious from the shape of these images. Noticed that the bit-line contact has changed to poly–plug plus tungsten–plug two-stage structure to accommodate the raised height due to the crown

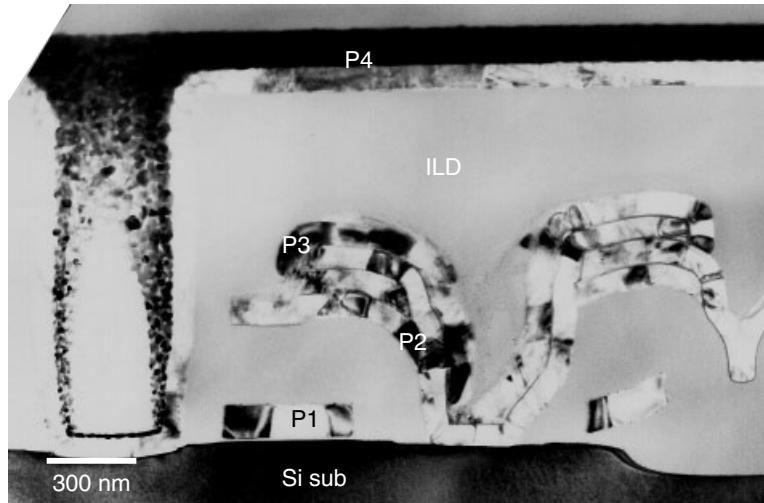


Figure 10.14 (a) A stacked capacitor design showing a two fins capacitor structure. P2, which forms the node contact, shows two layers and P3 (poly–plate) also shows two layers. The interlocking structure and the dielectric layers are clearly visible.

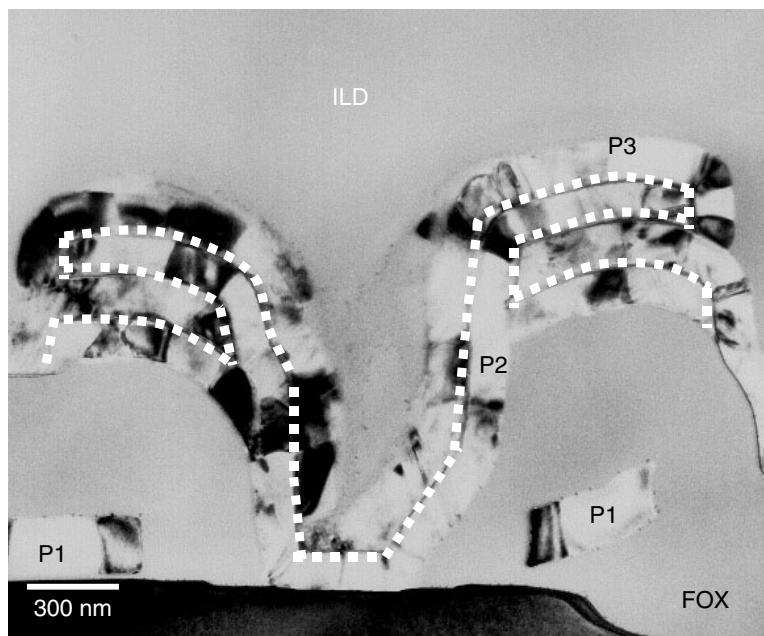


Figure 10.14 (b) Close-up of the capacitor interface. The dashed line indicates the location of the interface dielectric, which represents the effective capacitor areas.

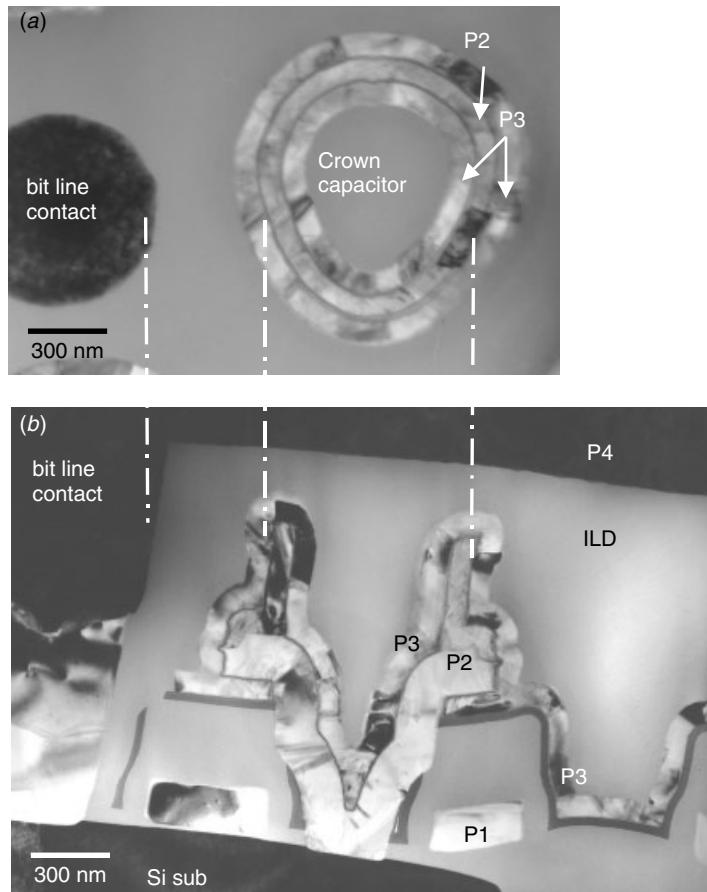


Figure 10.15 A crown capacitor (a) TEM plan view and (b) cross section. The circular crown-shaped capacitor was made by wrapping P3 around circular extrusion of the poly-node (P2).

structure. Figures 10.16 and 10.17 show the interface dielectric contours along the crown's vertical fin. It is obvious from the plan view micrograph of Fig. 10.17 that more circular layers can be added to the crown structure. In fact a self-aligned process (Kaga et al. 1991) has been developed for adding crown brim layers without adding any photo mask steps. And the approach has been demonstrated to be cost effective.

Textured Polysilicon Surface (or Hemispherical Grained, HSG, Si Structure)

Almost simultaneously, with the same driving force to increase the capacitor interface area, came a completely different approach. Instead of the structural variation over the capacitor in the device's scale, this alternative approach creates interface roughness between poly 2 and poly 3 on a nano-structural scale and thus increases the interface areas through the microscopically roughened structure. The terms "texturing" (Fazan and Ditali 1984, 1990; Fazan and Lee 1990), "rugged capacitor" (Fazan et al. 1992),

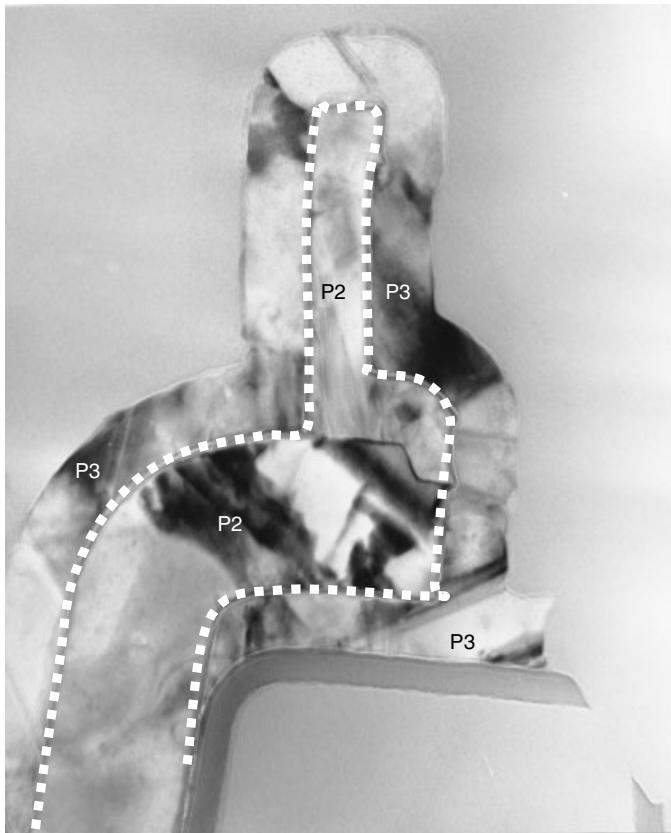


Figure 10.16 Close-up of a crown capacitor's vertical fin. The dashed line indicates the location of the ONO insulator layer wrapping the P2, and thus the effective capacitor area.

“Hemispherical grained Si,” or HSG (Watanabe et al. 1995), and “modulated stacked capacitor” (Jun et al. 1992) all refer to this approach, although some details of the process may differ.

The basic concept is to deposit a rugged surface polysilicon, instead of conventional smooth surface polysilicon, by low pressure chemical vapor deposition (LPCVD). By controlling the deposition temperature and pressure precisely so that the silicon film growth condition is between amorphous Si and polycrystalline Si, a rugged surface polysilicon film growth can be achieved (Ino et al. 1996).

Figure 10.18 shows an ordinary single horizontal fin stacked capacitor structure but with the HSG P2/P3 structure. Figure 10.19 shows a high-magnification close-up of the HSG structure at the P2 to P3 interface. Each hemisphere is 50 nm in diameter. As shown, most of them are not exactly the same shape and size due to the randomness of the growth/etching process. Notice that the shapes of the bumps are not exactly hemispheric. An ideal hemispheric surface area is half of $4\pi r^2$, or $2\pi r^2$, where r is the foot-print radius. In reality, the HSG structure can contribute an effective interface area of up to 3 times of the foot-print area, as seen in Fig. 10.19. Most of the bumps seen in this example have near-spherical shapes with a root undercut. In other words,

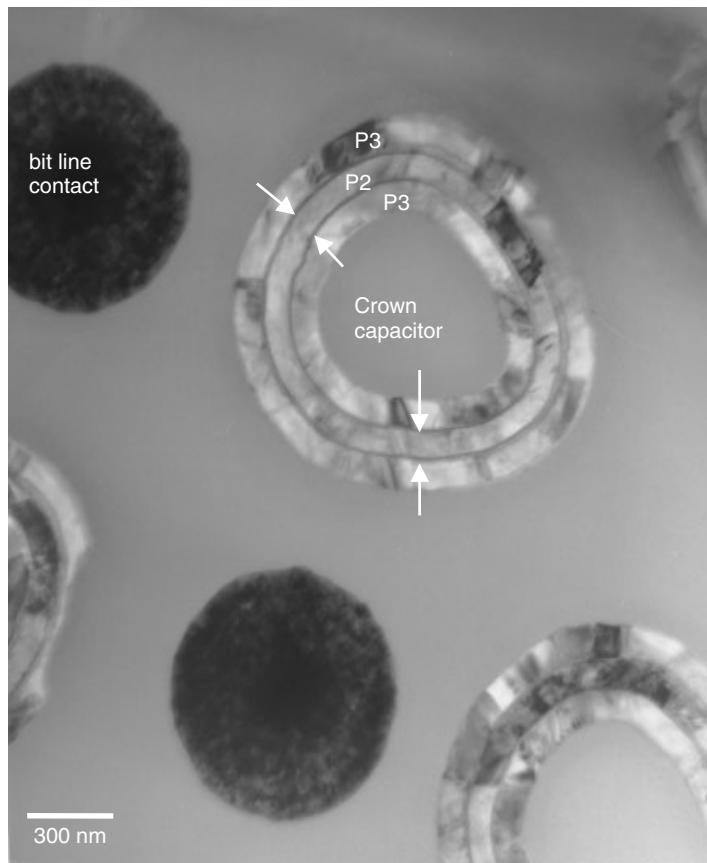


Figure 10.17 TEM planar section of a crown capacitor vertical fin. The name “crown” is obvious from planar section, since the capacitor’s vertical fins are circular in shape. The arrows indicate the ONO insulation layer between P2 and P3.

most of the bumps contribute to the effective area more than a half-sphere. Thus, whereas for a conventional flat interface, a circular area contributes a straightforward πr^2 interface area, with the hemispheric topography, the surface area increases up to between two to three times of the original surface area. The advantage of such an HSG approach is clear. No obvious geometric change in the device is necessary, and yet the capacitor’s effective storage area is doubled and even easily tripled without changing the basic design and most of the process steps. One step further to improve such an HSG structure is to extend the hemisphere into columns vertical to the node’s poly-surface. This is known as the modulated polysilicon capacitor approach (Jun et al. 1992), and the basic idea remains the same. The surface area’s increment can be up to 8 times that of the non-HSG plain surface node poly. However, it should be noted that the modulated poly cannot be easily combined with other technologies, like crown and fin structures. This is due to the fact that the modulated columnar poly structure was created using RIE vertical etch and thus can only be made in the direction normal to the substrate surface.

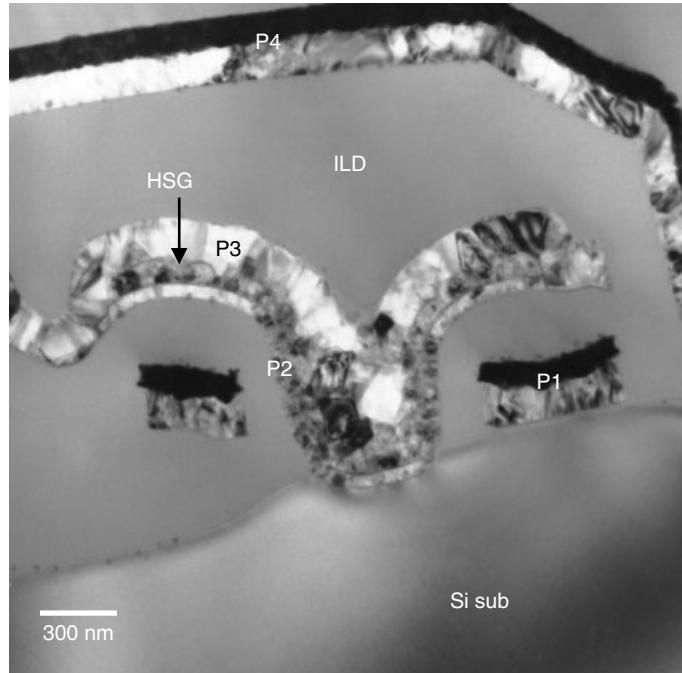


Figure 10.18 An ordinary stacked capacitor with the HSG structure. The low-magnification TEM cross section shows the capacitor's structure but does not reveal the poly 2 to poly 3 interface of the HSG microstructure.

The HSG has unique advantages. It is compatible with the other strategies that increase the capacitor effective storage area and thus can be implemented in conjunction with the other technologies, and the reliability of the thin dielectric layer on rugged poly has been proved to be quite satisfactory (Lo et al. 1992).

Capacitor on (or over) Bit Line (COB)

An important strategy for further increasing capacitor area that involves a major redesign of the process is to put the capacitor nodes on the bit line. This is an important design improvement and breakthrough as will be proved later. The new design scavenges the areas that were originally occupied by the bit-line contacts and increases the effective capacitor area. Exactly how big the areas that were occupied by the bit-line contacts? Figure 10.17 shows a good example in a plan view over the horizontal section through the effective capacitor area. The dark solid circles are the horizontal section of the bit-line contacts and the bright hollow circles are the crown structure capacitors. A rough estimation of the cell area occupied by the bit-line contacts, along with inter layer dielectric (ILD) with safe process and lithographic margins, is around 30% to 40%. This is why the capacitor on bit line (COB; sometimes called capacitor on data line) is so attractive.

Another big advantage of using the COB structure is that it frees the height limit that contains the capacitor structure, as the limit used to make bit-line contacts is

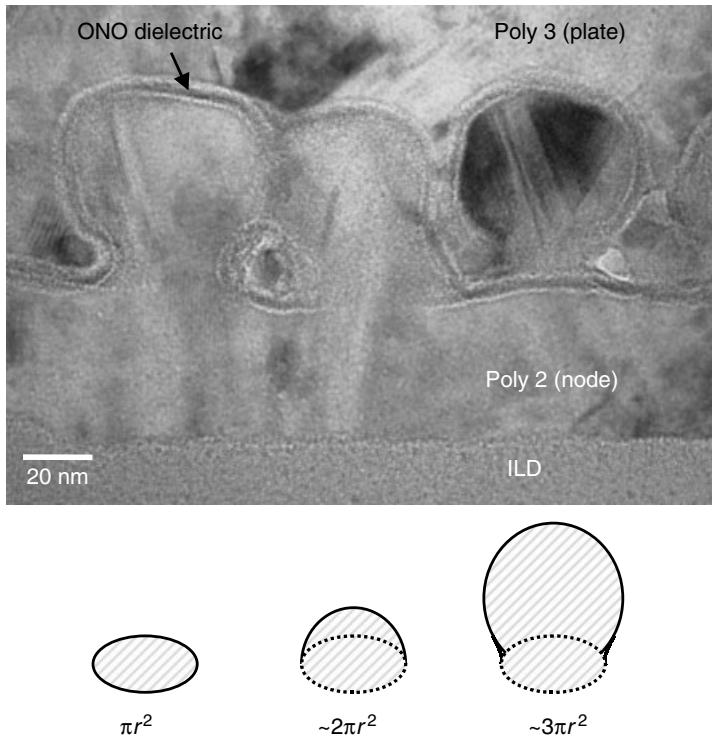


Figure 10.19 High-magnification close-up at the polysilicon interface showing the HSG microstructure. The hemispheric shapes are similar but not identical. Notice the undercut at each hemisphere's root. The image was slightly underfocused to enhance the ONO dielectric contrast. The schematics show the surface area increments due to poly-node surface topology changes.

the practical aspect ratio. In the conventional structure, the higher the capacitor is, the higher are the bit-line contacts, and accordingly a bigger bit-line contact diameter is needed; thus the bigger the areas will be consumed by the bit-line contacts. For COB structure, this limitation is removed and the capacitor structure can be made as high as is needed. This allows the capacitor areas to expand in another dimension effectively.

Figure 10.20 provides a direct comparison of the two different structures. The conventional capacitor under a bit-line structure, shown in Fig. 10.20(a), utilizes the cell area up to about 70%. The new capacitor on bit-line structure utilizes the cell area up to nearly 100%, as seen in Fig. 10.20(b). The capacitor's height has also increased without affecting the bit-line contacts, as seen in Fig. 10.20(b). As a result the utilization efficiency of the space on the z -axis has increased. As is clearly shown in Fig. 10.20(a) a large unused space is present, marked as IMD, between bit line and metal I (word line), whereas in Fig. 10.20(b) the device structure appears more compacted and densified.

Again, the beauty of the COB structure is that it can be combined with the other structures to increase the capacitor's area. Crown, fins, and HSG can be integrated with the COB structure.

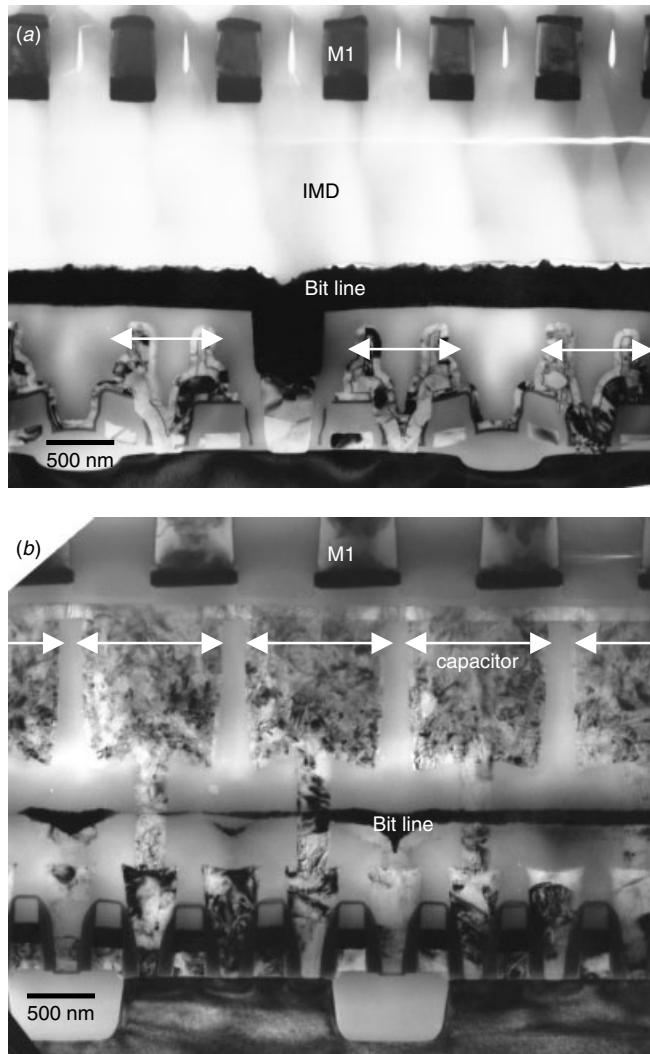


Figure 10.20 (a) Conventional stacked capacitor with a bit line on the capacitor's structure, and (b) the new capacitor on the bit-line (COB) structure that fully utilizes the chip cell's area. The two devices use a similar process technology ($0.35\text{ }\mu\text{m}$), yet the effective area for the capacitor (indicated by arrows) is much enlarged by the COB structure.

Combinations of Various Technologies

As we indicated earlier, several compatible technologies can be implemented simultaneously into the capacitor design to obtain a desired capacitance. Figure 10.21 shows an early COB structure. In this case the poly-node was manufactured with a cubical head to increase the lateral as well as vertical dimensions, as seen in the Fig. 10.21. This approach was sufficient for a capacitor cell in 16M DRAM with $0.5\text{ }\mu\text{m}$ process technology, and the chip size is small. The COB technology has simplified the process without using any other advanced technologies.

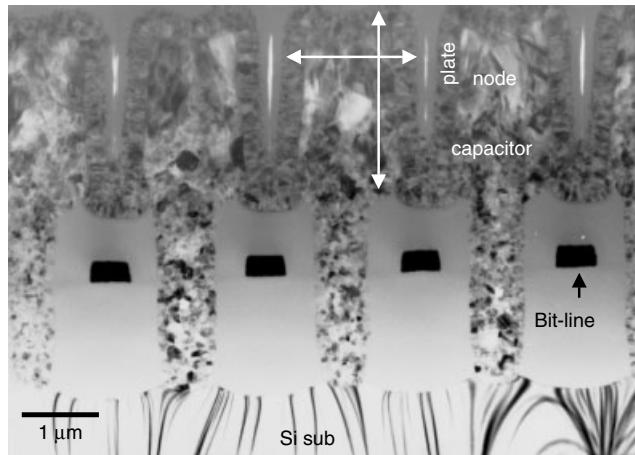


Figure 10.21 Early capacitor on bit-line structure combining the COB and the crown structure. The poly-node was manufactured with a cubical head to increase the lateral as well as vertical dimensions, as indicated by the arrows.

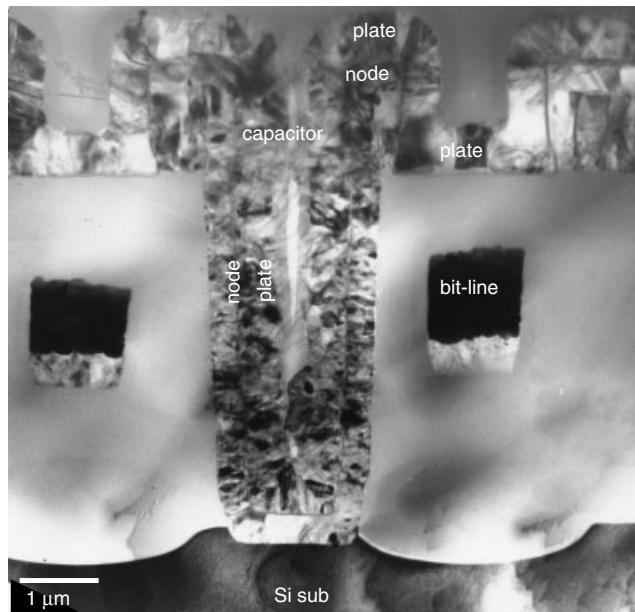


Figure 10.22 A capacitor on bit-line structure with a vertical stack (sidewall available for capacitor) and crown technologies implemented.

Figure 10.22 shows a similar COB structure, but a hollow poly-node contact is made using thin poly 2 as the sidewall, and the contact's internal cylinder is now used as the effective capacitor area. The structure further makes use of a crown structure, although not a very high crown, to increase the area. Since the node contact is high due to the COB structure, the capacitor area is large, and it includes the contact cylinder.

The disadvantage of this technology is that the node contact's diameter gets very big. As shown in Fig. 10.22, the contact diameter can grow from 0.6 μm , which is the norm, up to more than 1.5 μm (in using 0.5 μm process technology).

A further development along this line is a sophisticated structure called the planarized cylindrical stacked capacitor (PST) (Sim et al. 1997). This technology involves chemical mechanical polishing (CMP) for poly-plug contacts and the bit line. The advantage of using CMP is that it provides the fully planar stacked capacitor with cylindrical shapes rather than simple pillars. In the low magnification of such an example in Fig. 10.23, notice that poly-plug contacts were used for both node and bit-line contacts. Such simplification greatly reduces the process' complexity and contact problems. The PST capacitors are now between bit line and metal 1. Figure 10.24 gives a close-up of the capacitor structure's cross section in the bit-line and word-line directions. Note that the capacitor cylinders are not circular but oval (longer along the bit-line direction but shorter along the word-line direction). Note also from the TEM study that

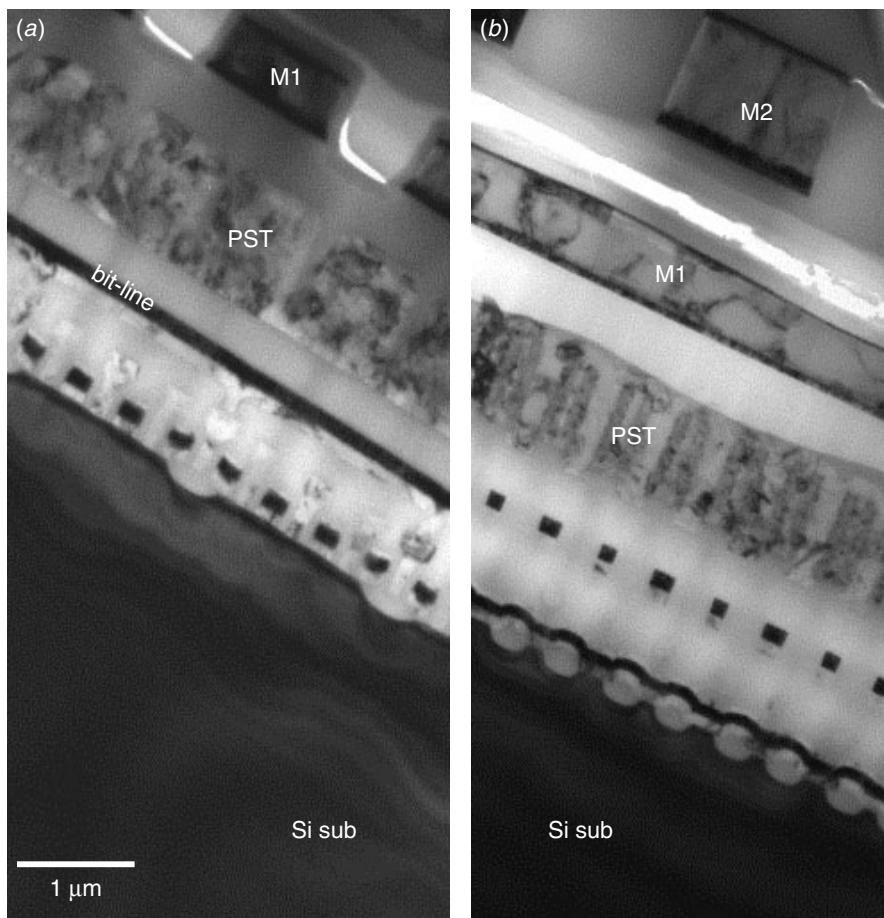


Figure 10.23 Low-magnification cross section of the PST cell structure, (a) along bit line and (b) along the word line. Notice that the planarity of the bit-line and PST structures is due to the CMP technology.

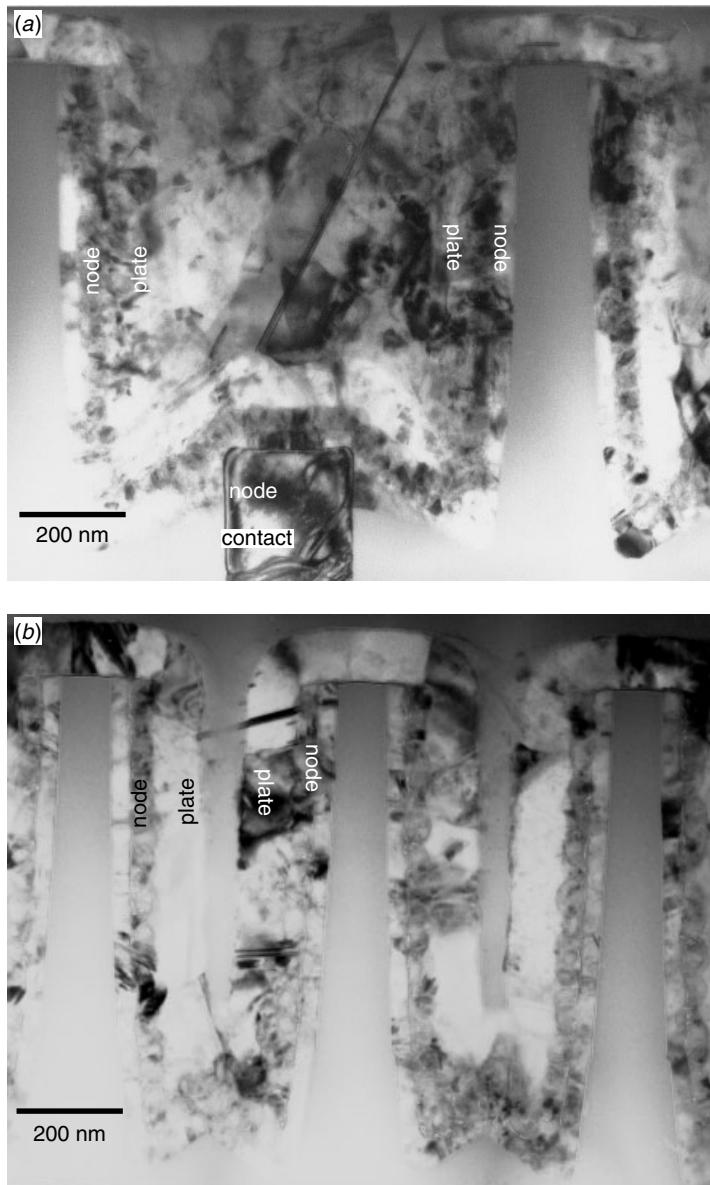


Figure 10.24 A capacitor on bit-line structure with the planarized cylindrical stack (PST) and HSG in combination. The cross sections were taken along the bit line (*a*) and along word line (*b*).

the poly-node was formed using at least two polysilicon layers. The first poly-node polyfilm (around 30 nm in thickness) forms the oval cup structure at the base. A second polyfilm is then deposited to form the HSG structure on one side of the poly-node surface. Figures 10.25 and 10.26 show the details of such multiple poly technology. It is clear from Figs. 10.25 and 10.26 that the poly-cylinders are not closely packed and



Figure 10.25 Close-up of the poly-node cup structure's edge showing the details of the HSG structure. Notice that the HSG is made on the poly-node with a separate polyfilm, as is necessary for this process.

oxide is used to fill the space in between. For the higher packing density PST capacitor, there is no oxide filling in between, as shown in Fig. 10.27. Further the number of crown brim can be doubled within each cylinder and HSG can be made on both sides of each crown structure. As a result the capacitor area has increased by a factor of 4, compared to the structure in Fig. 10.24. Figures 10.28 and 10.29 show details of the structure. A close-up at the double-sided HSG and ONO dielectric is provided in Fig. 10.30. Note that the HSG was made directly on the poly-node polysilicon film. No additional polyfilm was used to create the HSG. By combining HSG with ultra thin high- k dielectrics, high capacitance can be achieved within the limited cell areas (Fazan et al. 1992; Ino et al. 1994; Han et al. 1995; Kwon et al. 1996; Yamamichi et al. 1997).

The HSG + COB + PST structure is now used almost exclusively as the basic design for the stacked capacitor DRAM cell as with each generation the capacitor's height increases (COB technology has freed the height limit). The only variation among manufacturers are probably using different electrode materials (W, W silicide, etc.) and dielectric materials (Ta_2O_5 , with or without HSG, etc.).

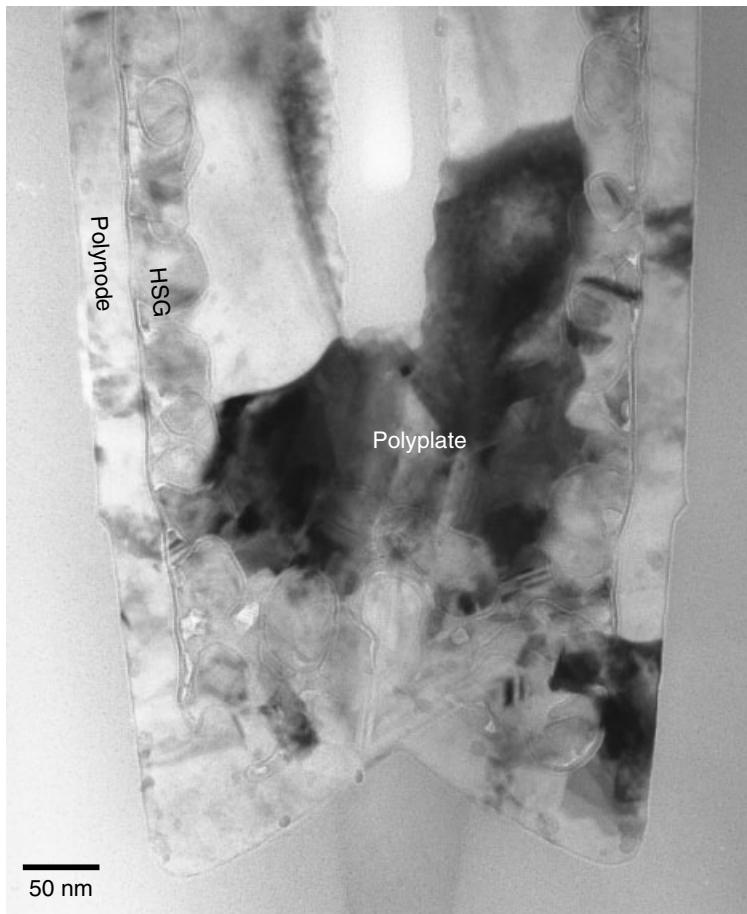


Figure 10.26 Close-up of the poly-node cup structure showing the details at the bottom of the HSG structure. Notice that the HSG was made of the poly-node by a separate polyfilm step. The layers are clearly differentiated.

Other Developments

A few brand new developments have been proposed that push the DRAM cell-packing density further into the Giga-byte range. SOI DRAM with a silicon-on-capacitor (SOC) structure has been shown to have scalable potential (Kim et al. 1997). A double self-aligned contact with Ge-added vertical epitaxial Si pads has also shown some advantages (Koga et al. 1997). X-ray lithography among other new developments, is providing new possibilities for conventional designs and processes.

Depending on the design methodology and device process technology used, there are always theoretical limits for each generation of technology. For example, for a supply voltage of 2.8 V and a planar cell structure, the limit for the physical gate's channel length is 0.42 μm (Lee et al. 1988) for the DRAM application. However, the decisions made for technologies used in mass production are not always due to

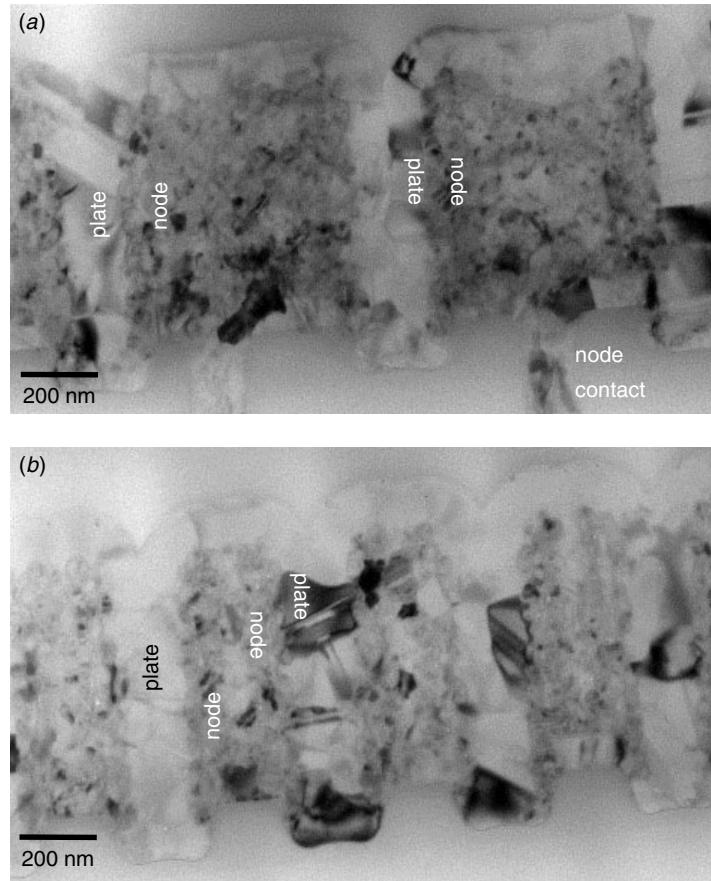


Figure 10.27 A capacitor on bit-line structure with a planarized cylindrical stack (PST) and HSG combined. A high-packing density PST structure is shown in this case with the cross sections along the bit line (a) and along word line (b).

capabilities of the technology involved. It has been suggested that the major technology challenges for near or below $0.1\text{ }\mu\text{m}$ concern threshold voltage control and lithography patterning (Asai and Wada 1997). However, problems can arise in a cost-effective production environment where the DRAM process undergoes simplification in manufacturing (Thakur et al. 1998). Technology that meets statistically controlled production standards is the base of the main stream technology used in the industry. The more elaborated and costly process does not survive in such a very competitive industry as DRAM manufacturing.

10.4 COMMON PROBLEMS IN DRAM PROCESSES

For DRAM manufacturing, low process cost and high yield are always the main objectives. The cell capacitor structures discussed above have been studied extensively, and most of the problems that can be attributed to failure in the capacitor area are due

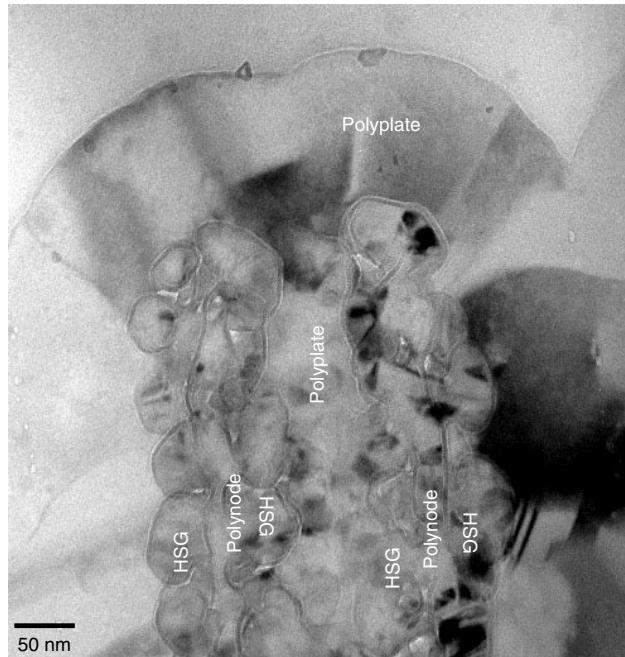


Figure 10.28 Close-up of the poly-node's cup structure. The cross section is along the word-line direction to show the details of the double crown and double-sided HSG structure within one cylindrical stack.

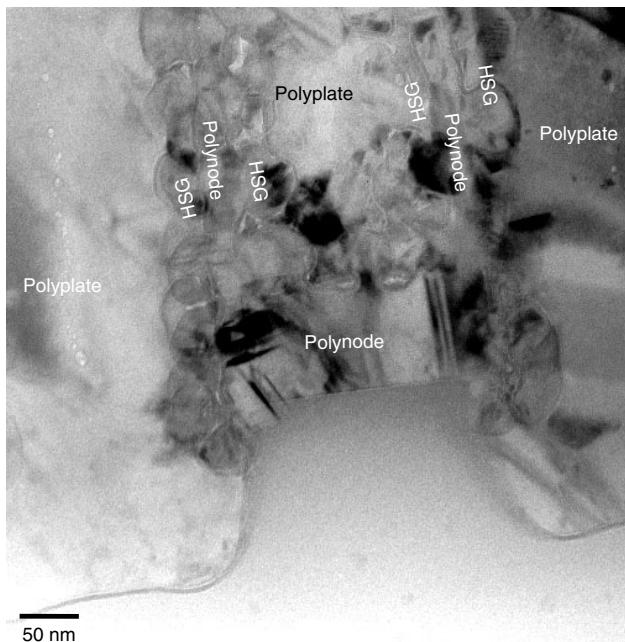


Figure 10.29 Close-up of the poly-node's cup structure at the bottom corner (along the word-line direction) showing the details of the root of the double crown and double-sided HSG structure within one cylindrical stack.

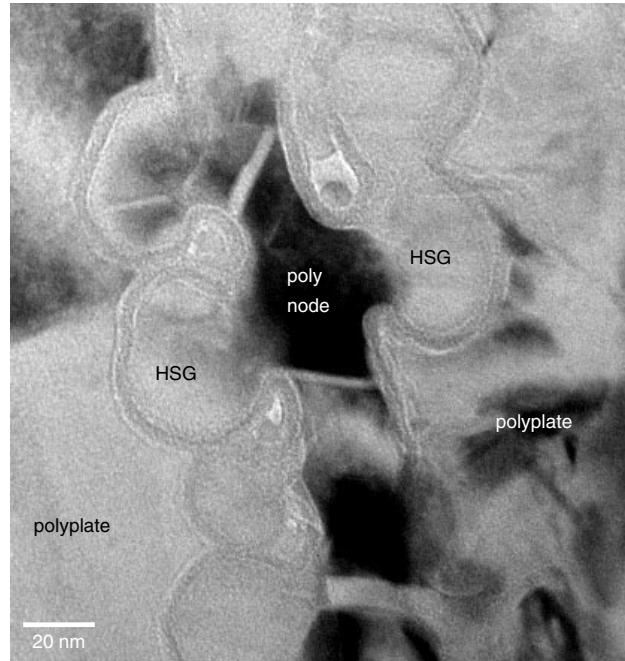


Figure 10.30 Close-up at the double-sided HSG structure and the ONO dielectric layer. The image was slightly underfocused to enhance the ONO contrast.

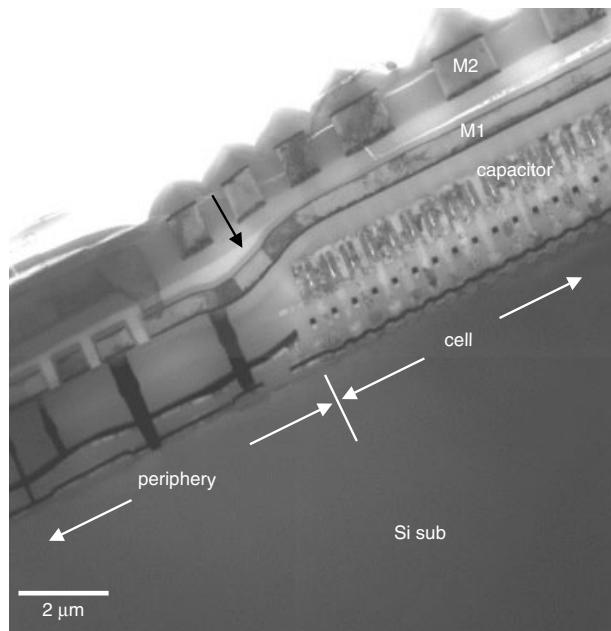


Figure 10.31 Low-magnification TEM cross section of the cell to its periphery boundary showing the step height difference between the two areas. The height difference is more distinctive in the metal 1 layer, as indicated. Apparently the step height difference increases with the higher capacitor.



Figure 10.32 Self-aligned poly-plugs used for both bit-line contacts and node contacts.

to dielectric leakage. Local dielectric rupture can be as much an issue as the particle contamination discussed in Chapter 13.

As we saw at the beginning of this chapter, a DRAM cell can be divided into a control transistor, a capacitor, and two contacts, namely the node contact and the bit-line contact. In this final section we consider a few other things worth mentioning in the DRAM structure and processes.

Planarization at the Cell-to-Peripheral Boundary

A serious problem can arise if the capacitor height keeps on increasing. Since the capacitors exist only within the cell area and not in the peripheral area in a DRAM device, a height difference as large as the capacitor height can result in the cell-to-peripheral transition area. This can severely compromise the lithography, which requires focusing for fine patterns. Other problems can arise if a back-end CMP process is used where such a large area height difference exists. Also a large step height difference can affect the device's reliability after packaging. As large stress concentrates

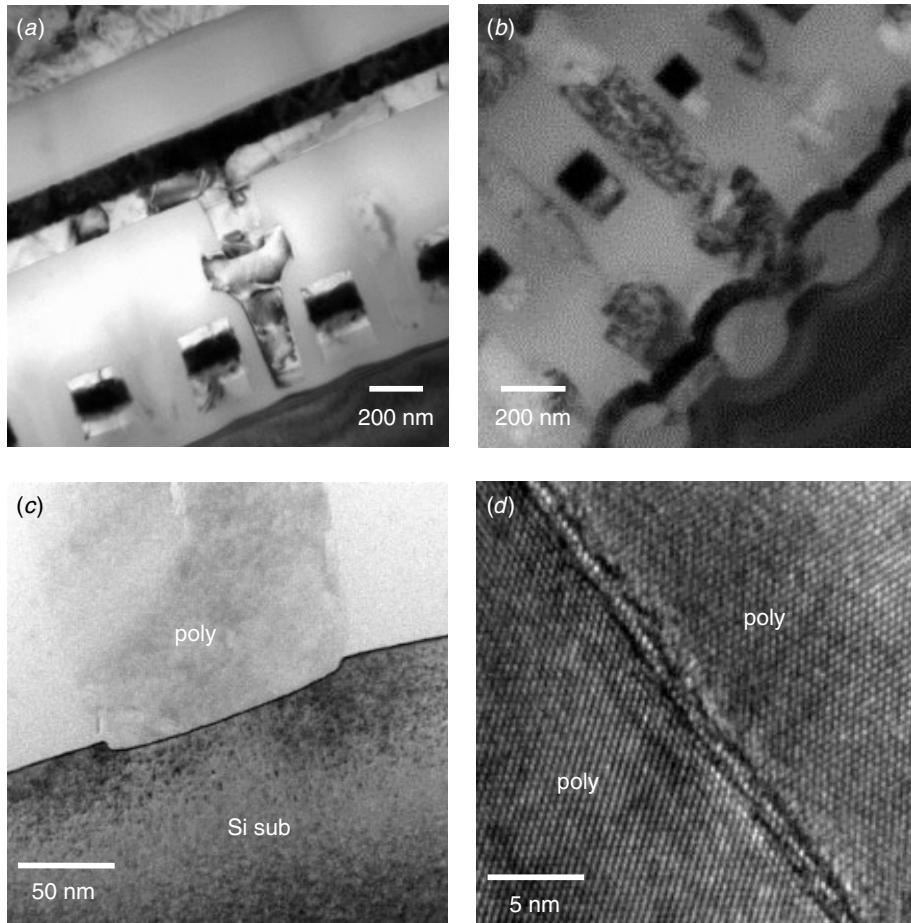


Figure 10.33 Bit-line contacts (a) and node contacts (b) are made of poly-plugs. Controlling the oxide layer in the poly-to-Si-substrate interface (c) and the poly-to-poly interface (d) is the key to a successful process.

at the boundary when the device is subjected to thermal stress, the increased boundary step can lead to system failure.

Figure 10.31 shows an example where effort was made to minimize the step height difference between the cell and the peripheral areas by putting up additional contact height in the peripheral. Nevertheless, as is quite apparent, a step height of about 0.5 μm exists across the cell-to-peripheral boundary. The problem is most obvious at the metal line across the boundary. A solution to this, as seen in Fig. 10.31, is to fill in the peripheral area with planarization oxide and CMP to planarize the boundary. A more elaborate approach is to include the damascene process that was originally designed for W bit-line. The sacrificial oxide layer in the peripheral area that was used for the storage node formation can also serve as a planarization oxide. For a large process margin with high process yield and increasing die size, use of CMP for global planarization and defect elimination is essential.

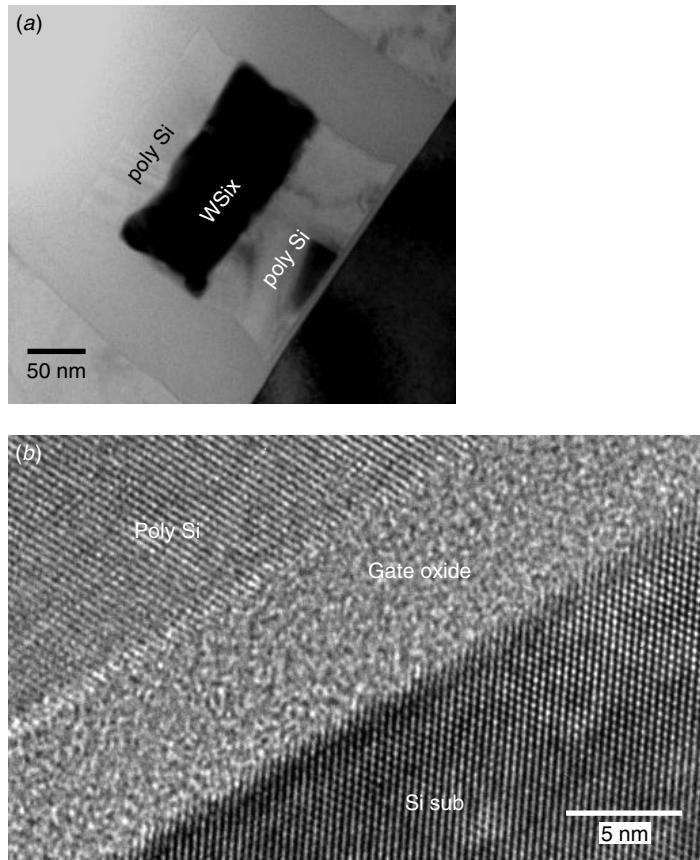


Figure 10.34 Close-up at the triple-layered polycide gate in (a) and a high-resolution gate oxide lattice image in (b).

Contacts and Contact Resistivity

As we saw in the preceding section, two contacts, bit-line contacts and node contacts are vital to the DRAM unit cell's performance. For the capacitor-over-bit-line (COB) structure, the scaling limit comes from the node contact's alignment with the bit-line, gate-poly, and active region. The problem is further complicated by a high-aspect ratio of the contact needed. One solution to such an issue is to use self-aligned poly-plug (SAC poly-plug), as seen in Fig. 10.32.

Selective oxide etching and the poly damascene process are used to maintain planarity in subsequent process steps. The SAC poly-plug itself involves no lithography and thus is a cheap and high-yield process (Sim et al. 1997). Both node contacts and bit-line contacts are done in one process. The only process concern is to control contact resistivity by maintaining clean interfaces between the poly-plug and Si substrate and between poly-plug and its subsequent plug contacts, usually W-plug or another poly-plug, as seen in Fig. 10.33. The interface oxide must be controlled because of its contact resistivity. A similar concern affects the use of the W-plug where high contact

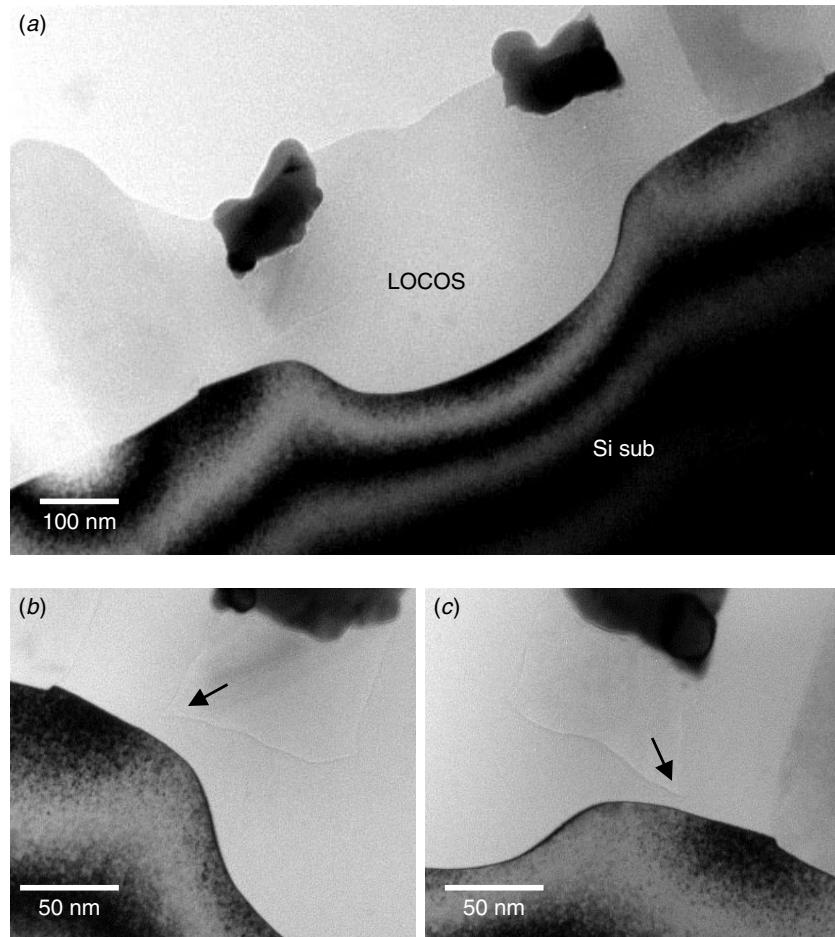


Figure 10.35 Close-up of the LOCOS structure and bird's-beak shape (*a*) along the bit-line direction. (*b,c*) The sharp tip at the polycide runner's bottom corner, as indicated, shows the bird's-beak shape.

resistivity and junction leakage can result. Ti/TiN as the barrier with Ti reacting with poly to form Ti-silicide needs to be optimized for a good contact W-plug process.

The problems concerning the contact interface of W-plug to poly-plug, poly-plug to Si substrate, and poly-plug to poly-plug were discussed in Chapter 8. The interested reader is encouraged to refer to the relevant sections for details.

Gate Structure and Gate Oxide

In recent DRAM structures where word-line resistivity is crucial and polycide is usually employed as the gate material, the typical gate structure is that shown in Fig. 10.34. The polycide gate with a polysilicon/WSi_x/polysilicon triple layer is usually accompanied by an oxide or nitride spacer. The reason for adding another polysilicon on top of WSi_x was discussed in Chapter 7. Since SAC poly-plugs are normally used, the contact

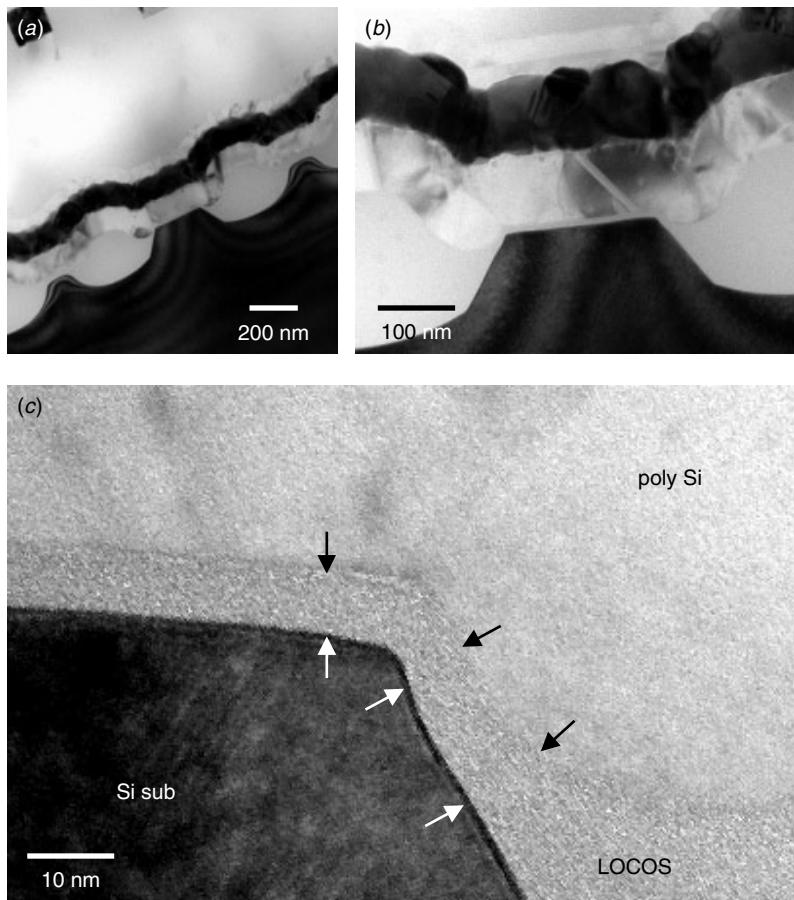


Figure 10.36 (a,b) Close-up of the LOCOS and bird's-beak structure along the word-line direction. The 0.18 μm technology using LOCOS isolation seems to be justifiable. The bird's-beak process without defect and gate oxide thinning is shown in (c), as indicated.

alignment and the possibility of a gate-to-contact short or leakage are less of a concern than in the old processes. Gate oxide, as usual, is still one of the major challenges in ULSI manufacturing yield and reliability, though this concern applies less to the DRAM than to a logic device. This problem is particularly urgent as gate oxide thicknesses reduce along with the device dimension shrinkage. (Chapter 6 provides a more detailed discussion.)

LOCOS and Bird's Beak

It was amazing to see a 0.18 μm (physically measured from gate physical length) technology DRAM device still employing LOCOS isolation technology instead of using shallow trench isolation (STI). As seen in Figs. 10.35 and 10.36, the LOCOS bird's beak was formed by a special LOCOS process, and there is no Kooi's thinning

effect observed at the bird's beak. Details on LOCOS, Kooi's effect, and STI processes are provided in Chapter 6.

REFERENCES

- S. Asai, and Y. Wada, *Proc. IEEE*, **85** (4), 505–520, 1997.
- T. Ema, S. Kawanago, T. Nishi, S. Yohida, H. Nishibe, T. Yabu, Y. Kodama, T. Nakano and M. Taguchi *IEDM Tech. Dig.*, 592–595, 1988.
- P. C. Fazan and A. Ditali, *IEEE IEDM Tech. Dig.*, 663–666, 1990.
- P. C. Fazan and R. R. Lee, *IEEE Electron Device Lett.*, **11** (7), 279–281, 1990.
- P. C. Fazan, V. K. Mathews, H. C. Chan, and A. Ditali, *IEEE Electron Device Lett.*, **13** (2), 86–88, 1992.
- L. K. Han, G. W. Yoon, J. Kim, J. Yan, and D. L. Kwong, *IEEE Electron Device Lett.*, **16** (8), 348–350, 1995.
- M. Ino, N. Inoue, and M. Yoshimaru, *IEEE Trans. Electron Devices*, **41** (5), 703–708, 1994.
- M. Ino, J. Miyano, H. Kurogi, H. Tamura, Y. Nagatomo, and M. Yoshimaru, *J. Vac. Sci. Tech.*, **B14** (2), 751–756, 1996.
- Y. K. Jun, S. K. Rha, S. C. Kim, J. S. Roh, W. S. Kim, and H. G. Lee, *IEEE Electron Device Lett.*, **13** (8), 430–432, 1992.
- T. Kaga, T. Kure, H. Shinriki, Y. Kawamoto, F. Murai, T. Nishida, Y. Nakagome, D. Hisamoto, T. Kisu, E. Takeda, and K. Itoh, *IEEE Trans. Electron Devices*, **38** (2), 255–261, 1991.
- I. K. Kim, W. T. Kang, J. H. Lee, S. Yu, S. C. Lee, K. Yeom, Y. G. Kim, D. H. Lee, G. Cha, B. H. Lee, S. I. Lee, K. C. Park, T. E. Shim, and C. G. Hwang, *IEDM Tech. Dig.*, 249–252, 1997.
- Y. K. Kim, S. K. Rha, S. C. Kim, J. S. Roh, W. S. Kim, and H. G. Lee, *IEEE Electron Device Lett.*, **13** (8), 430–432, 1992.
- S. Kimura, Y. Kawamoto, T. Kure, N. Hasegawa, J. Etoh, M. Aoki, E. Takeda, H. Sunami, and K. Itoh, *IEEE/IEDM 1988*, 596–599, 1988.
- H. Koga, N. Kasai, H. Hada, T. Tatsumi, H. Mori, S. Iwao, K. Saino, H. Yamaguchi, K. Nakajima, Y. Yamada, K. Tokunaga, S. Hirasawa, K. Yoshida, A. Nishizawa, T. Hashimoto, K. Ando, Y. Kato, K. Takemura, and K. Koyama, *IEDM Tech. Dig.*, 25–28, 1997.
- K. W. Kwon, C. S. Kang, S. O. Park, H. K. Kang, and S. T. Ahn, *IEEE Trans. Electron Devices*, **43** (6), 919–923, 1996.
- W. H. Lee, T. Osakama, K. Asada, and T. Sugano, *IEEE Trans. on Electron Devices*, **35** (11), 1876–1884, 1988.
- G. Q. Lo, V. K. Mathews, P. C. Fazan, and A. Ditali, *IEEE Electron Device Lett.*, **13** (4), 183–185, 1992.
- N. C. C. Lu, *IEEE Circuits Devices Mag.*, 27–36, 1989.
- R. P. S. Thakur, S. J. Dehoer, E. X. Ping, and C. A. Jesse, *IEEE Trans. Electron Devices*, **45** (3), 609–619, 1998.
- H. Watanabe, T. Tatsumi, S. Ohnishi, H. Kitajima, I. Honma, T. Ikarashi, and H. Ono, *IEEE Trans. Electron Devices*, **42** (7), 1247–1254, 1995.
- S. Yamamichi, P. Y. Lesaicherre, H. Yamaguchi, K. Takemura, S. Sone, H. Yabuta, K. Sato, T. Tamura, K. Nakajima, S. Ohnishi, K. Tokashiki, Y. Hayashi, Y. Kato, Y. Miyasaka, M. Yoshida, and H. Ono, *IEEE Trans. Electron Devices*, **44** (7), 1076–1083, 1997.

11 ULSI Devices III: DRAM Cell with Trench Capacitor



Plan view TEM on carbon replica of Fractured surface of magnesia oxide ceramic. It resembles satellite image from a rocky icy planet.

The 1M DRAM generation marked a transition period in MOSFET technology development. Planar capacitor technology had reached its limit and revolutionary process changes were needed to increase the cell density. The following 4M DRAM generation saw a flourish of new approaches. Two main developments evolved. One stacked the capacitor high above the Si surface, while the other dug deep into the Si substrate using trench technology. The stacked capacitor DRAM cells were discussed in the previous chapter. DRAM devices with trench-type capacitors are the subject of this chapter.

11.1 PROCESS FLOW

A simplified process flow is given in Table 11.1. Formation of isolation LOCOS (or STI) and N-well is followed by reactive ion etching (RIE) with Si_3N_4 as a mask to etch the trench a few micron deep into the Si surface. The etching is optimized for a tapered rectangular trench shape. The trench diameter can be rectangular, circular, oval, and even elongated oval. The sidewalls of the trench can be doped, and with oxide grown on them to serve as the dielectric material for the capacitor. (This was changed to oxide-nitride-oxide (ONO) in a later generation of trench capacitor). Polysilicon is then deposited and patterned for the gates and trench filling. Source and drain formation and metal interconnects follow.

Trench etching is a critical processing step. Deep, narrow trenches are required, about 3 to 4 μm in depth for 0.8 μm process technology DRAM, and about 5 to 7 μm deep for 0.5 μm technology 16M DRAM. In the early trenches the sidewalls had to be slightly tapered precisely at an angle of 7 degrees. Any derivations from this dimension

TABLE 11.1 Simplified Process for the DRAM Cell with a Trench Capacitor

Process Steps	Purpose and Comments
Form twin well	Defines n- and p-channel regions
Form LOCOS field oxidation	Defines active areas
Deposit nitride etch masking layer	—
Mask, trench, plasma etch	Form trench
Implant arsenic	Forms sidewall doping, storage electrode in trench
Deposit ONO	Form capacitor dielectric
Deposit poly 1 fill trench	—
Mask, capacitor plate, plasma etch	Form capacitor plate and array field plate
Deposit TEOS oxide	—
Mask, active area, oxide etch	Define active area
Grow gate oxide	—
Deposit poly 2, 3000A	—
Mask, poly 2, plasma poly etch	Form word line, gates for peripheral devices
Mask, storage node contact, oxide etch	Contact for storage node to substrate
Implant N-	Array S/D, and LDD implant for peripheral circuits
Form nitride spacer, mask, and implant N+ and P+	All n- and p-channel devices in the periphery
Deposit TEOS/BPSG	Forms ILD 1
Mask, bit-line contact, oxide etch	—
BPSG reflow	—
Deposit poly/silicide	—
Mask, bit line	Form bit line
Deposit TEOS/BPSG	Form ILD 2
Mask, metal contact	—
Metal 1 deposit and masking	—
Deposit IMD	—
Mask, VIA, oxide etch	Form VIA contacts
Metal 2 deposit and masking	—

or any bowing or kinks in the sidewalls due to etching damage would cause leakage (Wolf et al. 1990). The bottom of the trenches also had to be smooth and rounded at the corners, as fields of stress could easily be generated during oxide growth at sharp corners and cause defects. The trench depths had to be uniform. CVD deposited nitride was used as the mask for trench etching. The material deposited on top of the wafer was etched back and planarized by chemical mechanical polished (CMP).

In the latest generation of trench technology the trenches for 256M DRAM are at least 8 μm in depth and diameter less than 0.5 μm . The taper angle is nearly zero due to high aspect ratio. However, most of the other requirements has remained the same.

11.2 THE UNIT CELL

Figure 11.1 provides a schematic of the early generation trench cell DRAM with the three-dimensional trench capacitor (corrugated capacitor cell, or CCC cell) and the corresponding circuit. Silicon substrate is used as the capacitor node, which is directly connected to the drain of the control transistor through the substrate junction. The center of the trench was isolated by insulating the dielectric layer and refilled with polysilicon as the capacitor plate. Figure 11.2 shows a TEM cross section of a device

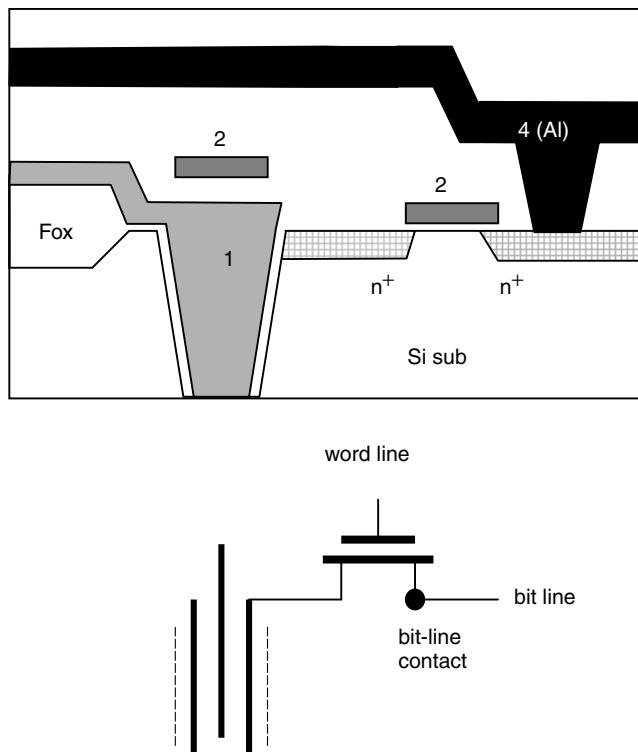


Figure 11.1 Schematic cross section of a conventional trench capacitor cell's DRAM structure and its corresponding schematic circuit. Poly 1 forms the poly–plate. Poly 2 is the gate poly and word line. The bit line is usually formed by the Al metal or another poly (or polycide) layer.

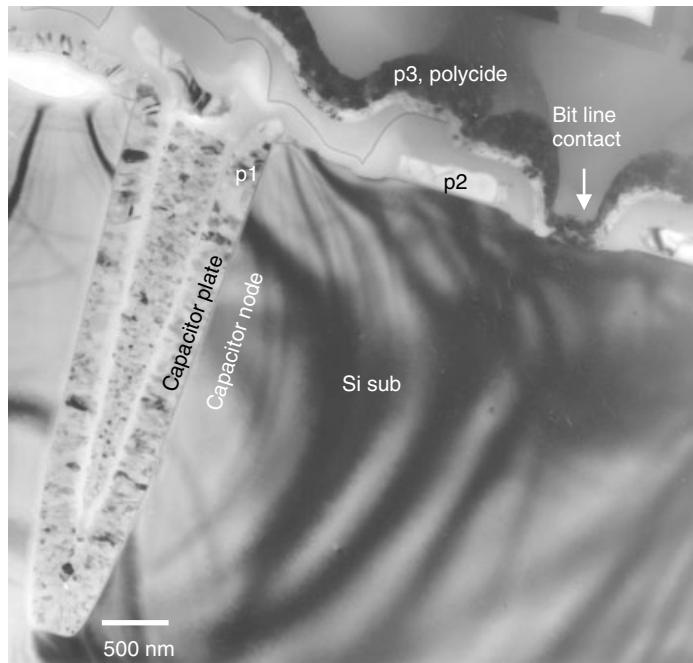


Figure 11.2 TEM cross section of an early generation trench technology. The Si substrate was used as the capacitor node and poly 1 as the capacitor plate. W polycide is the bit line and bit-line contact.

with such a structure. The key components within this unit cell are the transistor and the capacitor.

The Transistor

Two active junctions and one polygate together form the control transistor. This part is more or less identical to the traditional planar capacitor cell and even the stacked capacitor cell. The only difference is that the polygate used here is more often poly 2 instead of poly 1 as in the stacked capacitor process.

Since there are more process steps involved before the transistor gate oxide and polygate formation, the front-end process and thermal budget control becomes more critical in the trench capacitor DRAM processes. Polysilicon or polycide have been used as the word-line and gate materials, and Al metallization running in parallel with the word lines was used to reduce the word line resistivity. Figure 11.3 shows such a case clearly. The zigzag pattern of the word line again is used to facilitate the cell layout and maximize the utilization of the cell area.

The Capacitor

Unlike the stacked capacitor cell structure where two polyfilms were used as capacitor electrodes, most trench capacitor designs in the DRAM cell utilize a silicon substrate as one electrode and polysilicon as the other electrode to form the capacitor. As a

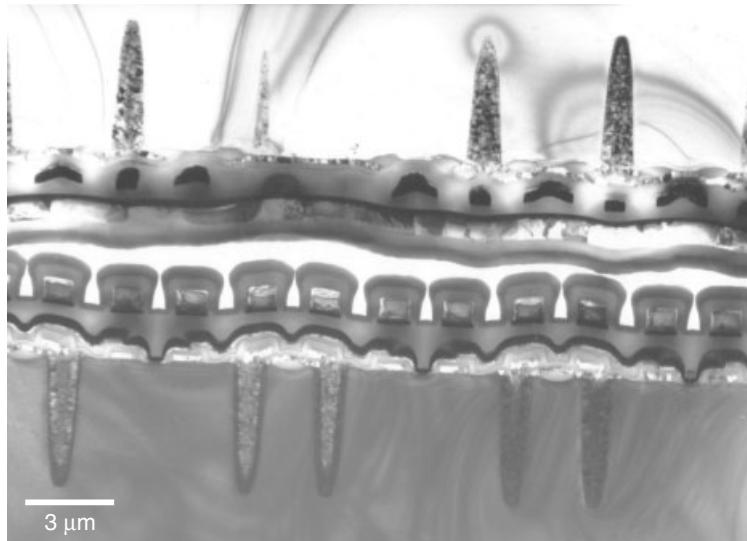


Figure 11.3 Low-magnification TEM cross section of two face-to-face 4M DRAM devices with trench capacitors. The lower device was cut along the bit line and the upper device along the word line. The arrangement of capacitor trenches, bit line and bit line contacts, Al metallization running in parallel with word line poly, are all clearly visible in this cross section.

result the number of polysilicon layers used in trench DRAM process tends to be less than that of the stacked capacitor process. Also due to the fewer stacking layers and lower vertical height, the device cell area tends to be much smoother than that of the stacked capacitor cell, so it requires much less planarization effort compared to that of the stacked capacitor cell process.

When silicon substrate is used as one electrode of the capacitor, doping to the substrate along the trench sidewall is necessary. Complex doping techniques are often needed (Sunami et al. 1982). In the early days this often resulted in substrate damages and defect generation along the trench's sidewall or bottom. As the technology matured, the doping technique was modified and the trench capacitors dislocation has ceased to be an issue.

The dielectric layer between the silicon substrate and the polyfilm electrode has a big effect on the capacitor's characteristics. To reduce the dielectric layer's thickness (and thus increase storage capacitance) while at the same time retaining the low-defect density, a composite dielectric such as oxide-nitride-oxide (ONO) or oxy-nitride (NO) is used in most commercial products. The excellent conformity and uniformity of ONO film deposition technology is one of the key factors in the success of trench capacitor technology. The later generation DRAMs used high- k dielectric materials as the capacitor dielectric. The production technology now includes Ta_2O_5/SiO_2 composite film.

The Polysilicon film serves as an electrode. It is deposited after the dielectric formation. The poly is doped *in situ* so that the deposited thickness, and thus the sheet resistivity, can be controlled. This procedure is followed by oxidation of the poly to fill the remaining gap inside the trench. Alternatively, another polyfilm can be deposited as an extra refill. As a result of such a simple process, the oxide that forms by oxidation of

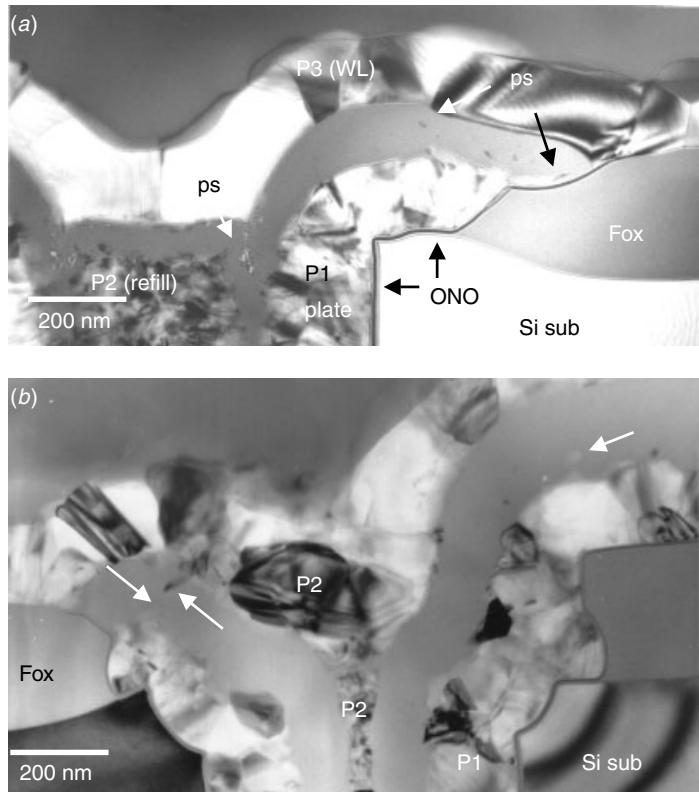


Figure 11.4 Poly stringers (residues) within SiO_2 between capacitor plate (poly 1) and word line (in this case, poly 2 or 3) are likely to cause leakage between them, as indicated. Both versions of the processes depend on the oxidation of the P1 polyfilm to form the dielectric layer and thus share the same poly stringer issue.

the poly 1 sometimes contains residue poly stringers, which can cause short or leakage between poly 1 (capacitor plate) and poly 2 (polygate and word line). Figure 11.4 illustrates this problem. Note that poly residue is not a problem in the substrate-plate trench-capacitor (SPT) cell. Poly 2 was used as the interconnect of poly 1, as is often the case where they are connected together.

11.3 CELL STRUCTURE EVOLUTION

From CCC Cell to SPT Cell

In the diagram of Fig. 11.1 the first-generation trench cell of the DRAM unit cell, called the CCC cell appears along with its corresponding circuit. A silicon substrate is used as the capacitor node and a polysilicon filling inside the trench's capacitor plate. This design quickly became adopted by the main stream trench technology for the 4M DRAM because of its simplicity. However, there was some concern about the trench's capacitor structure. As the signal charge was stored at the trench's surface in the bulk

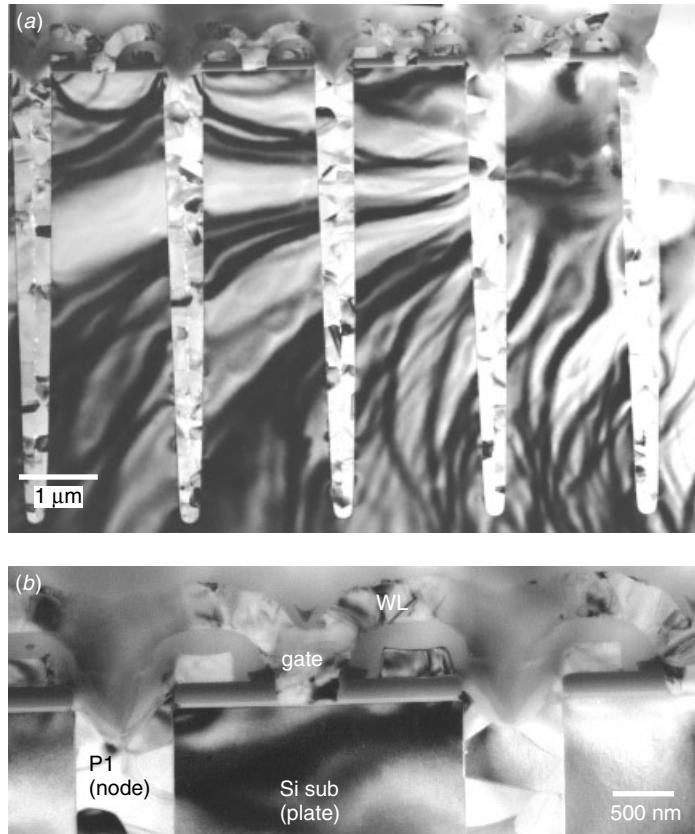


Figure 11.5 An early SPT cell's trench structure. Poly–fill acted as the capacitor node. It was isolated from the trench sidewall and connected to the transistor diffusion using the node contact.

silicon, the trench was highly susceptible to minority-carrier leakage. This disturbance would lead to surface leakage and a high soft error rate (SER) (Lu 1989). Several design and process improvements were implemented, such as Hi-C doping profiles at the trench's surface. However, other problems were introduced in the improvement, and the main stream of trench technology had to quickly shift to a new design that used a silicon substrate as the capacitor plate (substrate plate trench capacitor cell, or SPT cell). A typical unit cell of an early SPT cell is shown in Fig. 11.5. The main characteristic of an SPT cell is the contact that connects the capacitor node (poly inside trench) to the active area of transistor control, where as in a CCC cell the trench polyfills were connected to each other and grounded in the periphery area. The SPT cell quickly took over the CCC cell and became the dominant trench capacitor technology up to the present day.

Trench Morphology and Geometric Arrangement

Trench technology has evolved dramatically. Basically there is no rule as to how the trenches should be arranged, how the trench to transistor should be connected, and what

the trench's cross section should look like. Over several generations of trench design in commercially available products, one can find all kinds of designs and concepts being mass-produced. Here we give a few examples, which we show in Figs. 11.6 to 11.10. These examples were sampled by horizontal sectioning. TEM observation was done along the Si[001] pole direction. As will be discussed below, this observation angle provides excellent feel for how the trenches are arranged relative to transistors, contacts, and isolation structures.

Hexagonal Arrangement. Since the devices must be packed in the highest possible density, a typical and popular layout design is based on the simple closely packed geometry of the hexagon. Indeed, most of the early trenches were arranged in a hexagonal close-packing geometry. Figure 11.6 shows an example. The trenches were arranged symmetrically on both ends of a pair of transistors that share a bit-line contact in the center. The paired-up-transistor structure then repeats itself in the bit-line direction (left and right directions in the Fig. 11.6). As the two adjacent columns are shifted to each other by half the length of the paired-up transistors, a zig zag shift pattern results that repeats in the bit-line direction. Figure 11.6 was cut at near the Si substrate's top surface and thus the imprints of the gates and contacts can also be observed.

Rectangular Trenches. One way seen to increase the capacitor effective area without increasing the diameter of the trench was to change the shape of the trench's cross section from circular to rectangular, while the trenches are still arranged in hexagonal position as shown in Fig. 11.7. A more critical problem with this design was the sharp corners, which could cause leakage and compromise the reliability of the capacitor.

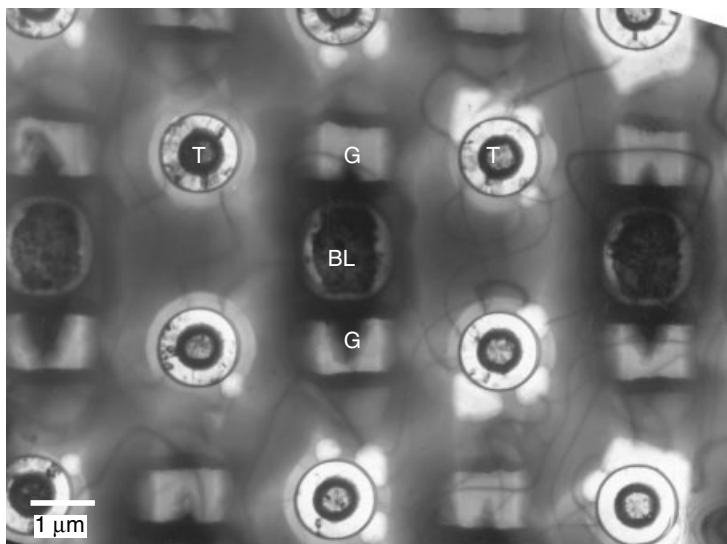


Figure 11.6 Planar section of a trench structure showing the hexagonal arrangement, poly and oxide fill, and the circular cross section: trenches (T), gate (G), bit-line contacts (BL). The diameter of the trench's cross section is about $0.8 \mu\text{m}$. Such trench designs were found in the 4M and 16M DRAMs. The word line runs from top to bottom in this image.

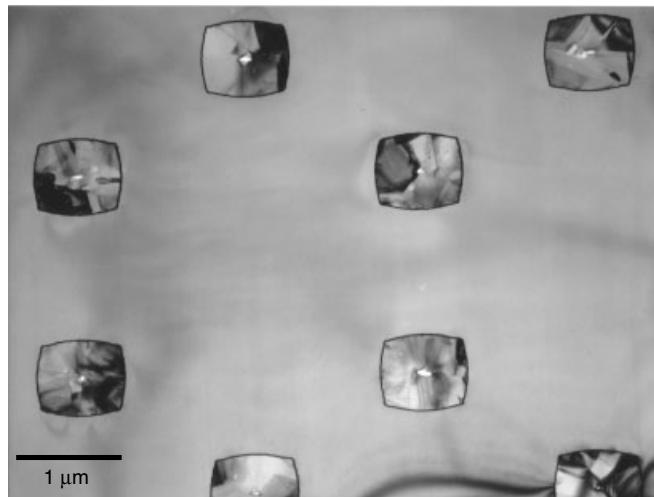


Figure 11.7 Planar section of a trench structure showing the hexagonal arrangement, solid poly-fill, and rectangular trench cross section. The cross section shows the dimensions of the trench to be roughly $0.8 \times 0.6 \mu\text{m}$. Such a trench design was found in some 16M DRAM products, though it was not a popular design. The word line runs from top to bottom in this image.

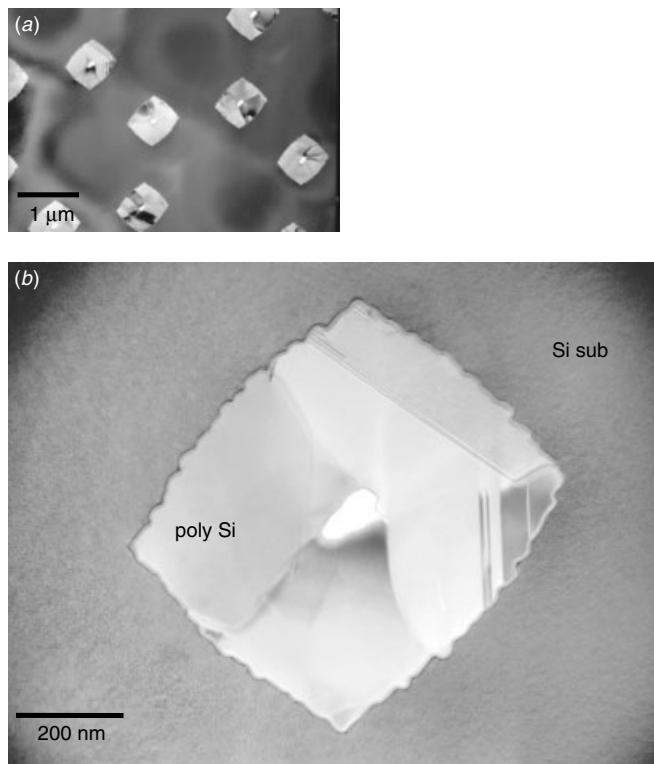


Figure 11.8 TEM planar section of another version of the rectangular trench showing a rugged sidewall. This creates additional interface areas of capacitor.

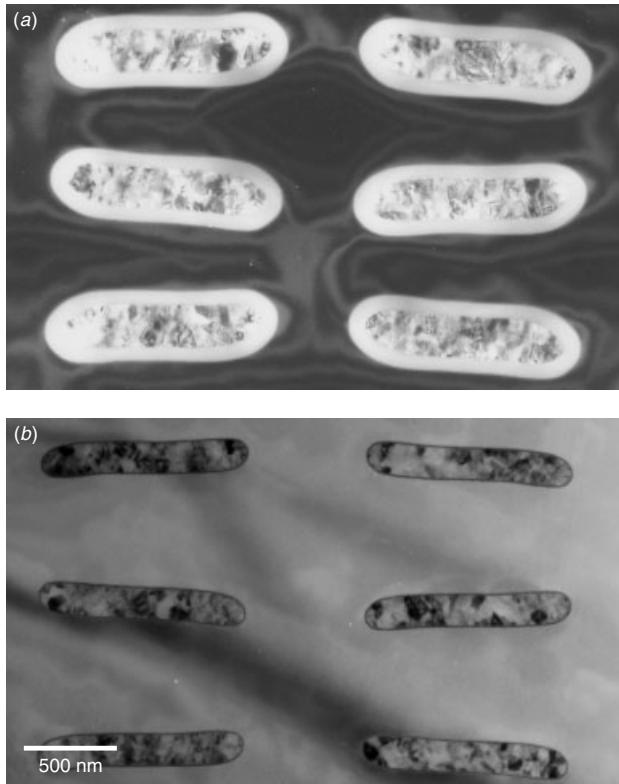


Figure 11.9 Planar section of a trench structure with a rectangular arrangement, Polyfill with a collar oxide on the upper part (*a*), and a chromosome-shaped cross section. The long and short axes of the trench's cross section are about $1 \times 0.3 \mu\text{m}$. Such a trench design was found in 16M to 256M DRAMs. The word line runs from top to bottom in this image.

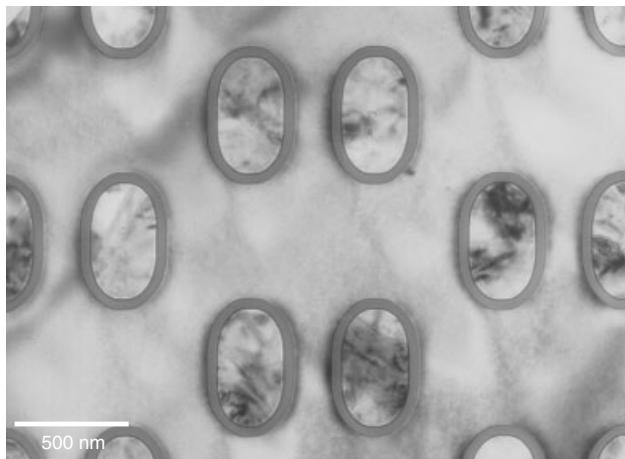


Figure 11.10 Planar section of a trench structure with a hexagonal arrangement, and an oval cross section. The long and short axes of the trench's cross section are about $0.6 \times 0.4 \mu\text{m}$. This trench design is found, in 64M DRAM products and beyond. The word line runs from top to bottom in this image.

ONO dielectric layer's quality is the successful aspect of this design. In order to round the corners and to further increase the effective area, the trench's sidewall can be turned into a rugged surface. This way up to 30% to 50% interface area was gained by a technology similar to that of HSG in the stacked capacitor. Figure 11.8 shows this treatment of the trench with its rugged sidewall interface.

Parallel Arrangement. A dramatically new design was implemented when the device technology allowed shrinkage below 0.5 μm . Apparently the diameter of the trenches could not be reduced beyond certain limits. New design ideas on the arrangement of transistors, contacts, and isolations were tested and implemented. In the end the trenches were moved from the two tips of the transistor pairs, as shown in Figs. 11.6 and 11.7, and placed side-by-side in line with the transistors. Each trench was sandwiched by two transistors and each transistor was sandwiched by two trenches. The performance details of this design will be discussed later. Here we are concerned with their physical layout. Figure 11.9 shows the planar section of this trench geometry. The elongated ovals (resembling chromosomes in shape) are arranged in parallel lines. Each trench is slightly twisted at both ends. The long axis of the trench is slightly tilted away from the Si(110) planes. This way the effective area was dramatically increased compared to the conventional equal-axis design. Figure 11.9(a) shows the planar section of this new trench at the upper part, and Fig. 11.9(b) at the lower part. This technology incorporated a buried plate trench (BPT) cell, which we will discuss later in this chapter when we review the 16M to 256M trench DRAMs.

Perpendicular Arrangement. The parallel, or side by side, arrangement in Fig. 11.9, increased the capacitor area by the elongated trenches; it did not make use of the close-packing geometry. As device's size shrank further, the hexagonal close-packing strategy was resumed. The trenches were moved back to the both ends of the transistor pairs. A slightly elongated oval shape of the trench was used, but the long axis was set perpendicular to the transistor axis, and not in parallel (as in Fig. 11.9). Figure 11.10 shows the new design. The beauty of this design is that the two slightly elongated trenches when coupled with transistor pair, appearing something like a dumbbell in shape, could perfectly dovetail and be arranged in a closely packed pattern. The trenches in Fig. 11.10 do not show much difference from those in Fig. 11.6 except that they have turned 90 degrees to each other. Both are in the same direction with the word line running from top to bottom, though the design concept behind them is entirely different. More discussion of this design will be provided later when we get to the 64M and higher giga-bit DRAMs.

Contact Technology

A distinctive difference between the stacked capacitor DRAM cell and the trench DRAM cell is the contacts that they use. For the stacked capacitor cell, the bit-line contact must conform with the capacitor's height and complicated capacitor stacks. Thus the aspect ratio tends to be high, and sometimes multiple plugs (e.g., poly–plug plus W–plug) must be used. The trench cell uses the much simpler bit-line contact. The borderless contact used for the 64M trench DRAM and beyond is also used by the stacked cell DRAM, so this is not unique to the trench cell. For the node contact, a

polyfilm (or polyplug) provides the contact in the stacked cell, while in the trench cell, different strategies have evolved to meet advances in the scaling down efforts. Thus bit-line contact technology appears to be crucial for the stacked cell and node contact for the trench cell.

The early trench cell used a substrate node design (Si substrate as the capacitor node), so it did not require any node contact. The transistors active area was in direct link to the trench sidewall diffusion, as seen in Fig. 11.1 and 11.2. Nearly the same design is found in the planar cell DRAM except that the capacitor node is vertical instead of horizontal. The poly–plate is a continuous polyfilm that is connected to the ground in the peripheral area. The trench cell, in this case, does not have any physically distinctive node contact (which is the same as in the planar cell DRAM). For the trench cell that takes the substrate plate design (SPT cell) and the more advanced buried plate design (BPT cell), the node contact and its design are crucial. As a review article by Alder et al. (1995) shows, the development of the trench DRAM cell by IBM was with the node contact (called strap contact).

The three generations of strap node contact as developed by IBM provide a useful illustration of the technology advancement here:

1. The straightforward approach was to use an in-situ doped polysilicon film or stud-plug as the contact, as shown in Fig. 11.11. As can be seen from this cross section, the stud requires precise control of the lithography. Only cells with n-type doped diffusion and n-type polysilicon trench fill can accommodate this conventional technology.

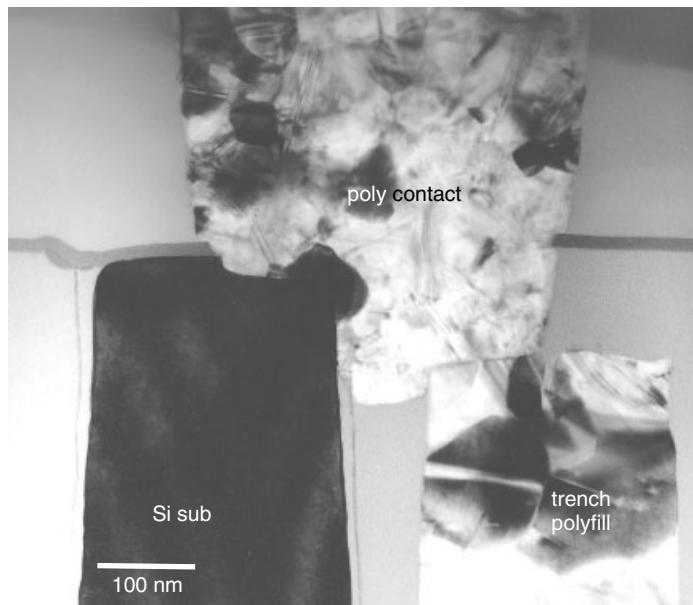


Figure 11.11 Conventional in-situ doped polysilicon stud contact that connects the poly capacitor node to the transistor's drain diffusion. Such contact is often used for n-type polysilicon trench fill.

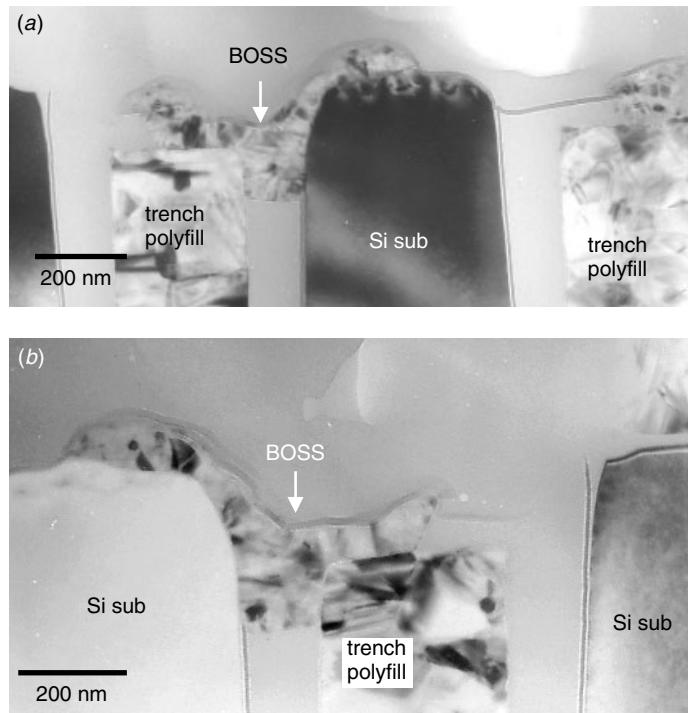


Figure 11.12 A boron outdiffused surface strap (BOSS) contact formed between the poly-node and drain diffusion. The irregular shape of this contact is due to its process nature.

2. In the technology generation for the 4M DRAM novel process technology called strap process was tried by IBM. Because of parasitic transistor leakage the technology was modified to accommodate the collar oxide in the 16M DRAM (Adler et al. 1995). The technology was called “boron out–diffused surface strap,” or BOSS. After the source-drain implantation, a thin layer of silicon nitride was deposited. Contact holes were etched through the silicon nitride and trench top oxide in each cell to expose the boron-doped trench to the polysilicon fill and p+ diffusions on the transistor drain. Thus, unlike the strap for 4M technology, the BOSS process did require a mask step. A blanket intrinsic polysilicon was deposited and the wafer annealed to diffuse and activate the boron. The boron in both the transistor drain and the trench polysilicon fill diffused up into the intrinsic polysilicon. The result was a boron-doped polysilicon layer bridging the trench and transistor drain diffusion. Selective wet etching then removed the remaining intrinsic polysilicon, isolating the cells from one another. Figure 11.12 shows a BOSS contact. Notice that the polysilicon strap contacts have a rough surface. This is due to the selective wet etch used to remove the intrinsic polysilicon, which has no well-defined and smooth boundary with that of the boron-doped polysilicon. The removal of the intrinsic polysilicon, and thus the isolation among different devices, relies solely on the selectivity of the wet etch process. The critical parameters are selectivity of the wet etch as well as the diffusion of boron in an undoped polysilicon film. An equivalent BOSS process for the cell with

n-type diffusion and trench polyfill was not available because a suitable doping-sensitive wet etchant was not available for n-type material (Adler et al. 1995). As a result a process similar to that used in Fig. 11.11 had to be adopted for the n-type poly.

3. A more aggressive contact scheme was developed to make a contact across the collar oxide without using any of the Si surface area. This was accomplished by a strategy called a “buried strap,” or BEST, contact. A BEST contact was fabricated after the recess etch of the second trench’s polysilicon fill. The top part of the oxide collar was removed so that a third polysilicon trench fill could be deposited and contact the silicon substrate just below the surface. The trench polysilicon was then recessed again. Isolation trench etch followed. At this process step the buried strap was formed, and this left only the trench’s sidewall next to the diffusion area of the storage node connected to the storage node’s trench with polysilicon fill. Figure 11.13 shows examples of the buried strap. Notice that the surface of the polysilicon fill and the buried strap form a continuous concave curve from the Si substrate to the drain’s diffusion area. The curve is cut straight

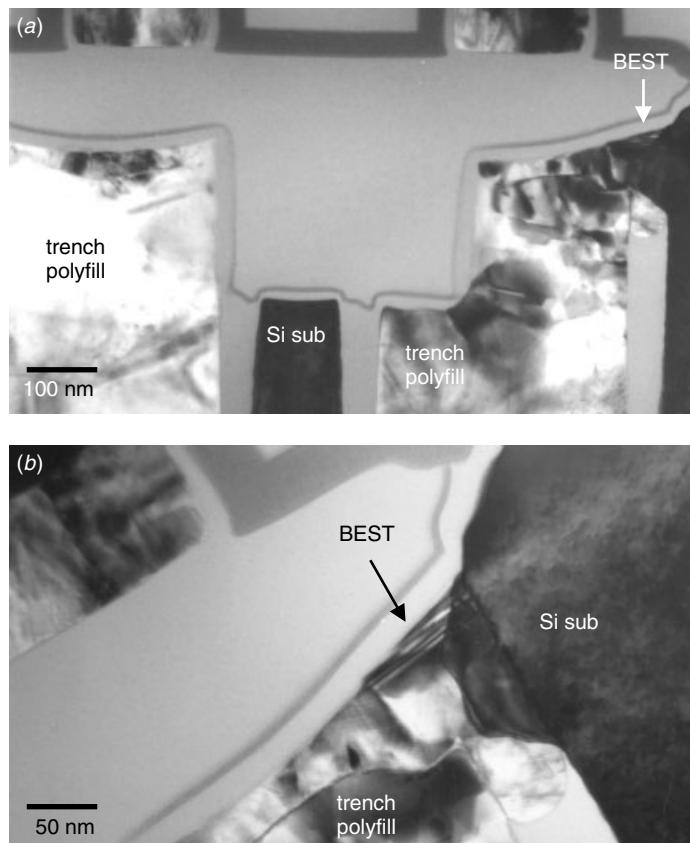


Figure 11.13 Buried strap (BEST) contact formed between the poly-node and drain diffusion. The buried strap is fabricated early in the process and has a diffused connection formed by creating a sidewall contact along one edge of the trench capacitor.

down near the center of the trench due to the formation of the isolation trench etch that defines the final shape of the buried strap and its associated trench with polysilicon fill. As will be shown later in a plan view of the same device, this forms a semicircular polysilicon trench top.

11.4 SOME EXAMPLES FROM COMMERCIAL PRODUCTS

A few popular DRAM device constructions are presented in detail in this section. The reader is encouraged to refer to the previous sections and obtain a full picture on how the DRAM product technology evolved over the various generations.

4M DRAMs

An early 4M DRAM is shown in Fig. 11.14. The cross sections are along word-line and bit-line directions. Compared with the 1M planar cell DRAM (e.g., see Fig. 9.3 in Chapter 9), this 4M trench cell DRAM does not have many process innovations except for those of the trench itself. The silicon substrate is still used as the capacitor node, poly 1 is still the capacitor plate, and poly 2 is still the word line and the gate. Both devices eliminate physically distinctive node contact by linking directly the transistor's drain diffusion area to the substrate node's diffusion area. The biggest difference is in the capacitor's trench structure. The horizontal capacitor is now wrapped into a cone and vertically inserted into the Si substrate. This way a huge amount of Si substrate

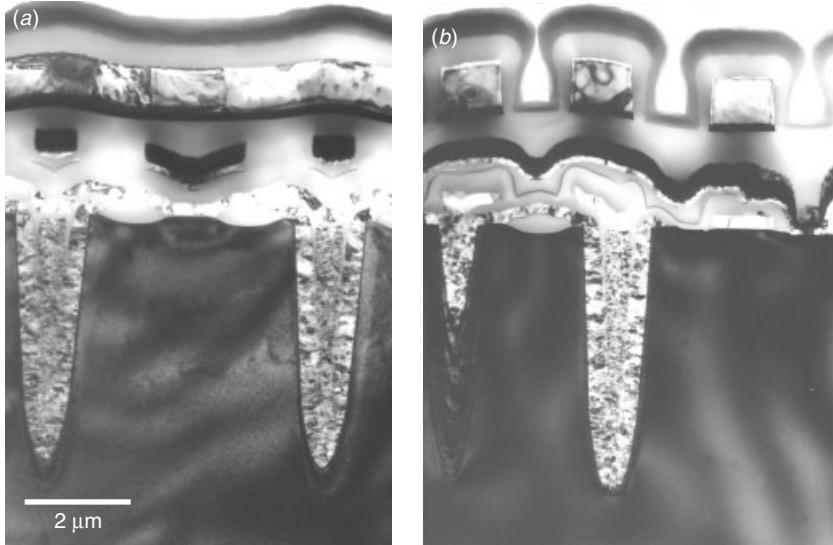


Figure 11.14 TEM cross section of the 4M DRAM trench cell (a) along the word line and (b) along the bit line directions. All layers are clearly seen; poly–plate and poly–fill, poly word line and gate, nitride cap, WSi_x bit line and contacts, Al metal as the parallel word line, and even oxide-nitride composite passivation. The trench's diameter is about $1.5 \mu\text{m}$, and the trench is about $5 \mu\text{m}$ deep. The trench's sidewall has two tapered slopes. The upper part is about 2.5° and lower part is about 7° to 9° .

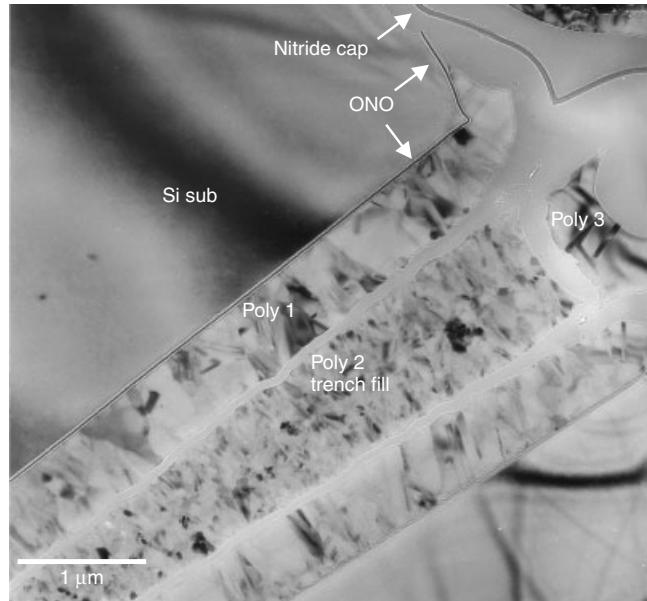


Figure 11.15 Close-up of a trench's cross section showing the node trench to be filled with two polysilicons. Poly 1 is used as the capacitor plate. A thin oxide is grown and then poly 2 is filled in. Poly 2 is used as a dummy trench fill material, and there is no electrical connection. A thin oxide is grown again, and then poly 3 as polygate materials is deposited. Also observed is that the capacitor dielectric uses an ONO (oxide-nitride-oxide) ultra thin dielectric to increase the capacitance.

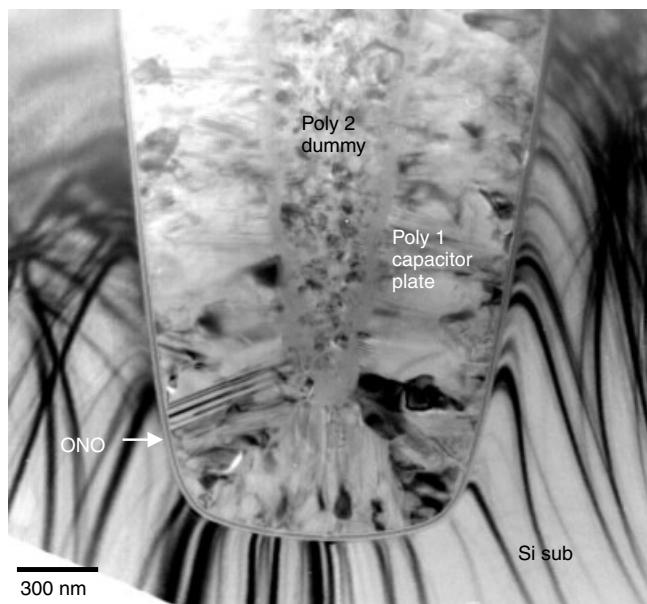


Figure 11.16 Close-up at the trench's bottom shows it to be flat with rounded corners to the tapered sidewall. The sidewall's tapered angle measures about 7°. The ONO dielectric layer can be seen clearly, and its uniformity can be examined in detail, particularly, at the trench's bottom.

surface space can be utilized as the cell's size is reduced. Additional changes involve the introduction of one more polyicide layer's as the bit-line (replacing Al in the 1M planar DRAM) interconnects and Al metallization as the word line running in parallel with poly 2.

A close look at the new trench's structure, Fig. 11.15, reveals some improvements as well. The composite oxide-nitride-oxide (ONO) is used as the capacitor dielectric layer. This reduces the dielectric thickness and thus greatly increases the capacitance. Notice that the poly 1 capacitor plate does not fill the trench completely but merely forms a thin film lining inside the trench. Thin oxide and additional poly-fill have filled the trench. Thus the process actually contains at least four polysilicon layers: poly 1 as the capacitor plate, poly 2 as the trench fill dummy layer, poly 3 as the word line and gate, and poly 4 (polyicide) as the bit line and bit-line contacts. In between poly 1 and poly 2, a thin oxide is formed by thermal oxidation of poly 1. The oxide layer conforms with poly 1's surface topology. But the problem is the unoxidized polysilicon stringers and residues within the thermal oxide, which can cause leakage (or a short) between the poly 1 and poly 2, which in turn shorts poly 3, the word line, as seen in Fig. 11.4(a). Figure 11.16 is a close-up view of the trench's bottom. Notice how close the poly 1 and poly 2 refills are to the trench's bottom. The ONO used for the

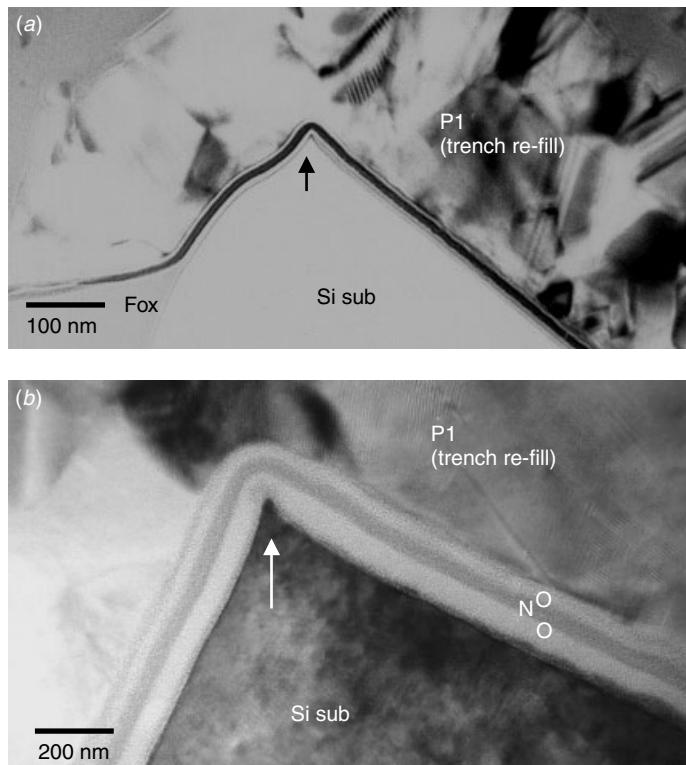


Figure 11.17 Close-up of the trench's upper corner. ONO was used as the capacitor's insulation material. The sharp corner, as indicated, at the substrate's edge creates a high corner field that has the potential to cause a reliability problem.

capacitor dielectric can also be seen clearly in this image. Another interesting aspect is the trench's bottom morphology. Instead of being semicircular, the trench's bottom is a rounded rectangle. This bottom trench morphology remained until the process technology advanced and trench's bottom diameter shrank to under 300 nm. Then the rounded corners of the bottom closed together and formed a semicircle.

A serious problem encountered in this early trench technology was not related to the bottom morphology but concerned control of the trench's top corner morphology. Recall that complicated trench's sidewall doping techniques are usually required to form a Hi-C doping profile at the trench's surface in order to prevent disturbance to the minority-carrier leakage and thus high leakage and the soft error rate. Inevitably a tilted substrate ion implant will damage the trench's top corner and form amorphous trench top corners as heavy ion implantation is applied into the trench's sidewall. The amorphous Si substrate at trench's top corners then re-crystallizes during the subsequent thermal treatment. Because of the large local stress imposed by the trench and the surrounding oxide structure, the re-crystallization will change the shape of the corners. In counteracting the mechanical stress the corners will taper to a sharp tip, as seen in Fig. 11.17. The re-grown Si substrate forms an elongated tip extruding into

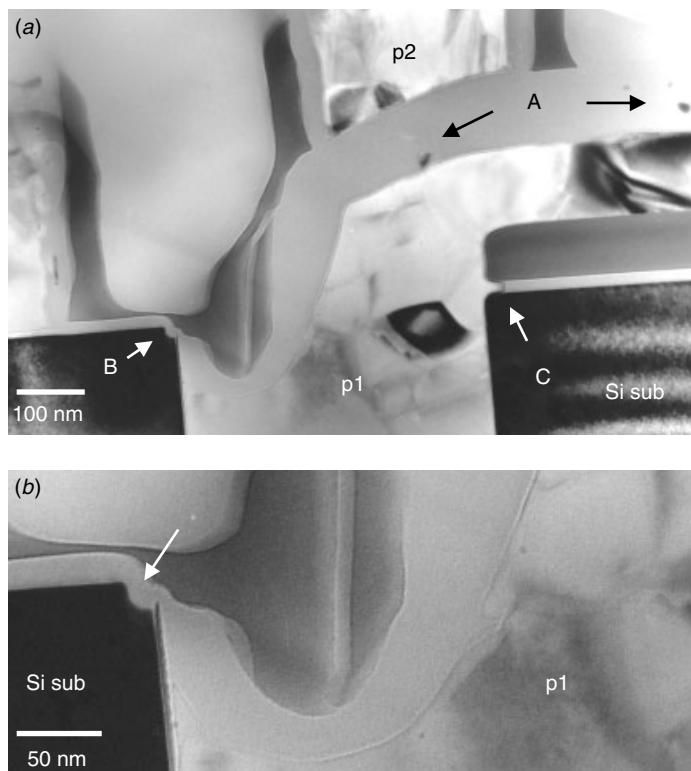


Figure 11.18 Close-up of the trench's top showing poly-stringers (A) in the oxide between the poly capacitor plate and the poly 2 word line, as indicated. Then trench's top corner mouse bite defects (B) and pad-oxide undercut (C) are also observed. A closer look at the mouse bite in (b) shows the defect forms after the ONO formation.

ONO dielectric. The local ONO thickness reduction due to this effect can be as high as 40%, which raises severe yield and reliability concerns. Other than the sharp protrusion, a defect called a “mouse bite” is often observed on the trench’s top corner, as seen in Fig. 11.18. Though not quite a reliability concern, the mouse-bite defect can usually be prevented by carefully controlling the etch process during poly 1’s capacitor plate formation.

Still another well-known issue at the trench’s top corner concerns lattice defects in the Si substrate. Figure 11.19 shows an example of such defects. The micro-twinning appearing along Si(111) lattice planes formed at the corner area to relieve the corner stress. Careful thermal process control to minimize the process-induced corner stress would have prevented such a defect from forming.

As the device’s trench capacitor shrinks in diameter and the refill oxide layer increases in thickness, only a very slim poly 2 refill is needed, as can be seen in Fig. 11.20. This development in fact simplifies the process by eliminating one polysilicon layer. A close-up of the top of this slim refill has been already shown in Fig. 11.4(b), where the slim refill was not due to an additional polysilicon layer but formed naturally through the poly 2 word line. The rest of the process remains the

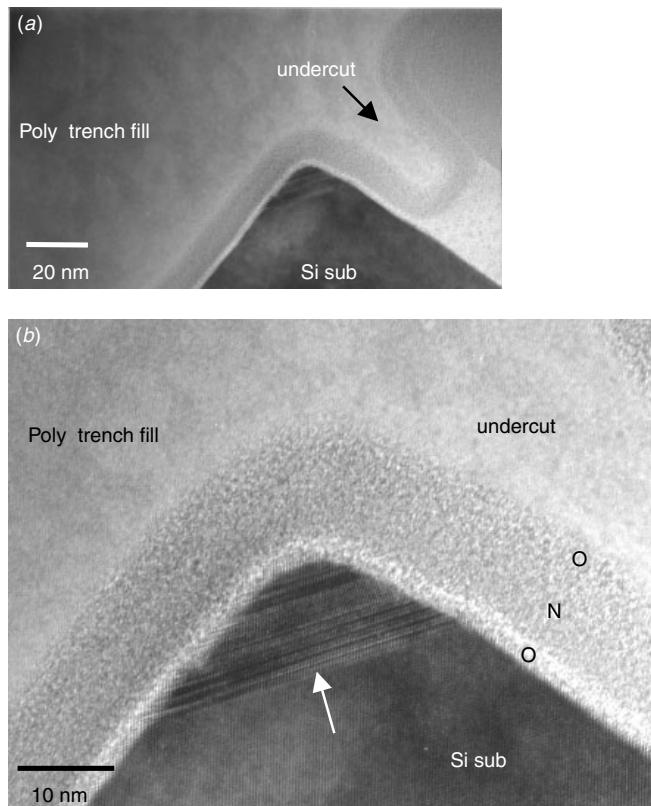


Figure 11.19 Close-up of the trench’s top corner with the pad-oxide undercut as in Fig. 11.18(b). The Si-substrate corner has micro-twinning defects, as indicated. These defects are due to the stress induced by the ONO formation and the poly refill of the undercut.

same, since the new process uses poly 1 as the capacitor plate and poly 2 as the word line without an additional dummy poly refill to fill the trench. Nevertheless, there is still concern about a possible leakage path formation between poly 1 and poly 2, as the slim poly 2 fills into the oxide gap can makes close encounter with poly 1. Close examination on Fig. 11.20 in fact reveals the bottom of poly 2 fill to be nearly touching poly 1.

A further development in the direction of a smaller trench diameter, was for it to be filled with poly 1 alone. Thermal oxidation on poly 1 polysilicon was also eliminated due to the stress it produced. Figure 11.21 shows more advanced 4M trench DRAM cells with all the trenches filled by only one polysilicon layer. When the polysilicon film was deposited into a narrow and deep trench, a seam gap formed at the center of the trench. This gap is inevitable, even though the polysilicon film deposition is known to form with excellent conformity. Polysilicon annealing induced grain growth tends to induce more grain boundary voids within the trench polysilicon, as can be seen in Fig. 11.22. These microscopic voids are considered harmless as long as they don't coalesce and form larger voids that block the trench's polysilicon.

Unique to the trench process is its sidewall doping uniformity and the residue defects that ion implantation generates. While computer simulation can help in understanding



Figure 11.20 Cross-sectional view of the trench capacitor. The slim poly 2 refill is in the center of the trench. There is no additional polysilicon dummy refill required. This slim poly refill is formed by the poly 2 polysilicon film, which forms the word-line and gate structures.

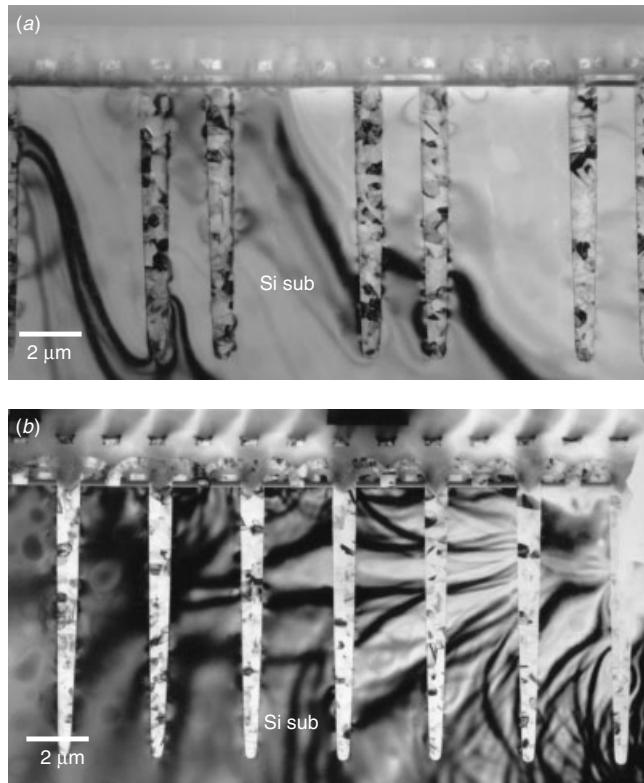


Figure 11.21 Cross-sectional view of the trench capacitor along (a) the bit line and (b) the word line. The way the trenches are arranged, slightly different tapered angles of the trench sidewall are observed along the different directions in this comparison. Notice that there is no oxide layer nor second polysilicon refill. The poly 1 itself fills the trench completely.

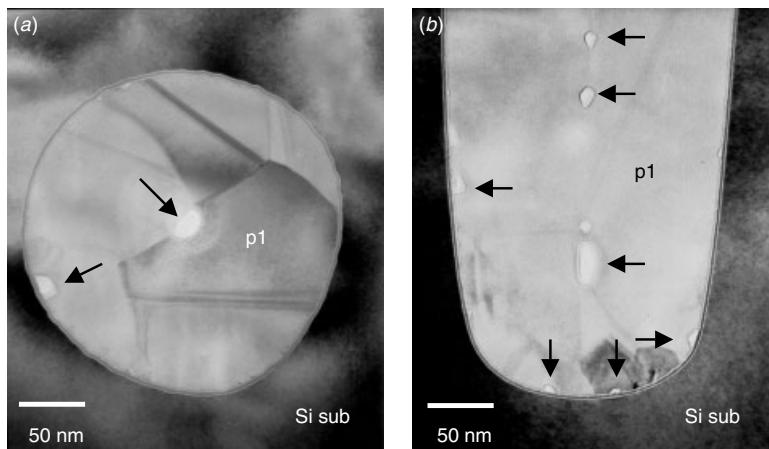


Figure 11.22 (a) Planar section and (b) cross section of the capacitor trench. Poly 1 entirely fills the trench without any oxide or additional polysilicon refill. Notice that there are micro-voids in the polysilicon formed at the trench center as well as at the sidewall and bottom.

and optimizing the ion implantation process (Kato 1988), the junction profile can be determined only experimentally. Visual observation of the doping profile along the trench's sidewall helps also confirm the computer simulation results. Secondary ion mass spectrometry (SIMS) depth profiling is the most accurate method used in determining junction profiles on fairly large planar surfaces. But SIMS analysis cannot be performed in the trench's sidewall. The most plausible technology to use is either scanning capacitance microscopy (SCM) or TEM with junction delineation. The correlation between SCM and TEM junction delineation was discussed in Chapter 5. Here TEM delineation will be used to determine the trench's sidewall junction profiles.

Figure 11.23 shows a planar section on a 4M DRAM trench cell area. With TEM delineation the junction around a trench can be seen clearly. A close-up at the trench in

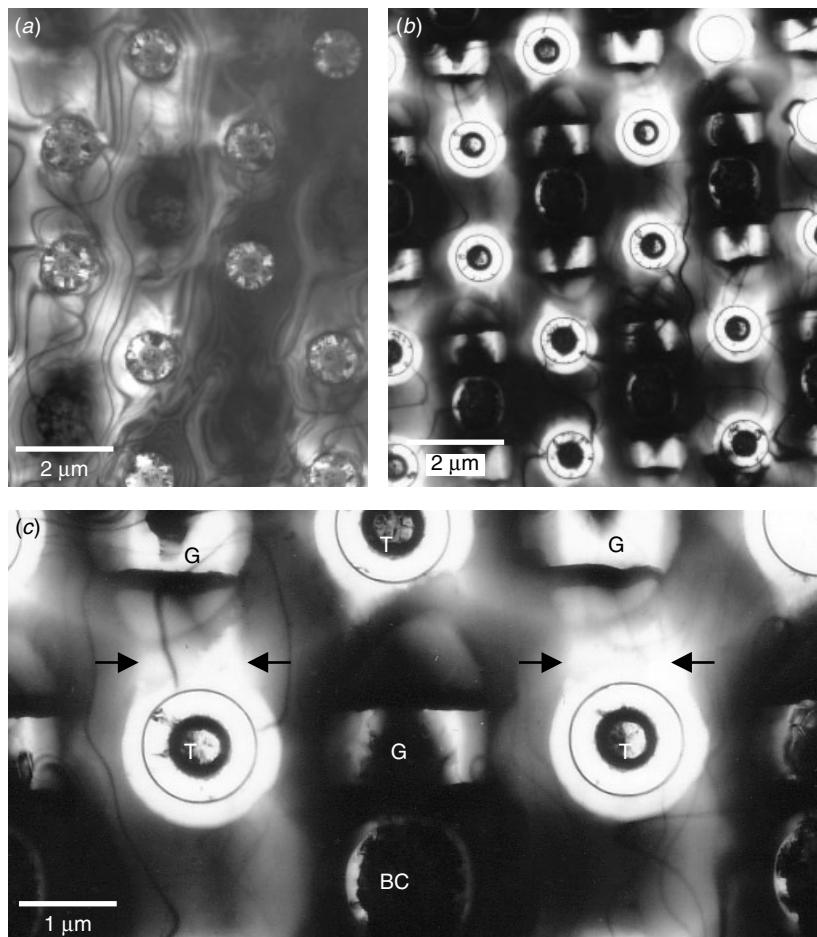


Figure 11.23 Planar section of the trench capacitor (a) as seen in TEM, and (b) with junction delineation to reveal the substrate node doping. (c) A close-up of the two trenches shows the doping profile and its link to device drain doping, as indicated by arrows: trench (*T*), gate (*G*), bit-line contact (*BC*). Notice that the polysilicon in the trench is also heavily doped, so they are also etched and delineated during the junction delineation.

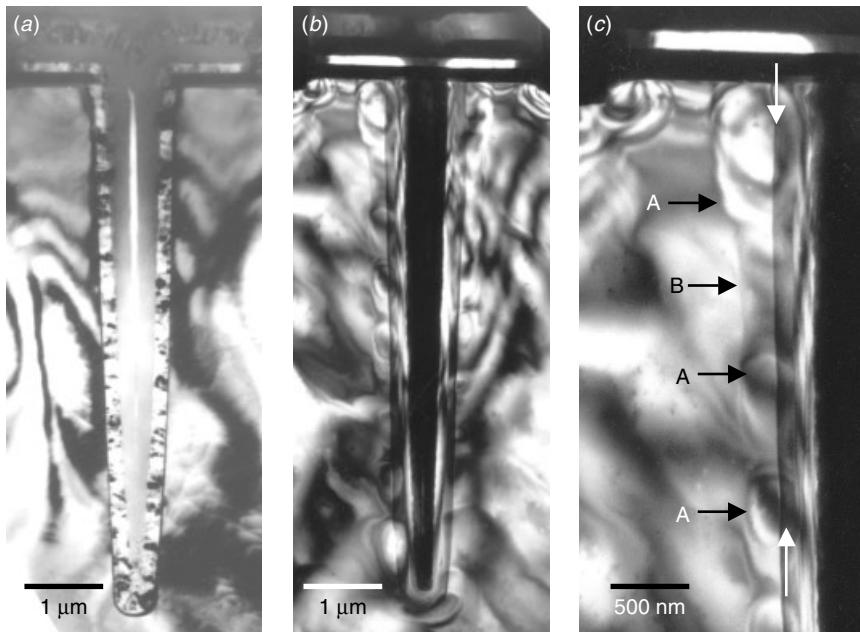


Figure 11.24 TEM cross section of the capacitor's trench structure (a) without and (b) with junction delineation. (c) A close-up at the sidewall of the delineated trench reveals the inhomogeneity of the doping as it proceeds depthwise. The clear straight line indicated by up and down vertical arrows in (c) is the overlapping shadow line from the Si substrate due to the thick sample effect. Arrows A shows lumps of higher concentration areas and arrow B shows the normal or lower concentration area.

Fig. 11.23(c) and its associated transistor shows the doping profile and its link to the device's drain doping. Notice that the polysilicon in the trench is also heavily doped and thus also etched and delineated during the junction delineation. When overetched, the poly-fill within the trench will disappear, and without any support, the central oxide fill will be lost too. This is shown in Fig. 11.23(b) at upper right-hand corner. Figure 11.24 shows the cross-sectional TEM image of the trench's structure with and without the junction delineation. The polysilicon fill in the trench was etched away, as can be seen in Fig. 11.24(b). A closer look at the trench's sidewall junction delineation profile, Fig. 11.24(c), reveals that the doping profile along the trench's sidewall is not as uniform as expected. Local high concentration lumps are revealed along the depth of the trench, probably due to multiple scattering during the tilted ion implantation. Figure 11.25 further compares the trench's structural view in cross section and plan views. The samples were taken after the junction's delineation. Noticed the junction profile is much clearer in cross-sectional than in plan view. This is because the image was taken from a much thicker sample area where the etching effect was delineated much better than in the thin area, as is the case for plan view sample. What is demonstrated here is that the doping profile along the trench's capacitor sidewall can be studied by cross-sectional and plan views in TEM analysis of the delineated samples. This is not yet possible for any other surface-sensitive analytical techniques, such as SIMS or spread resistance probe (SRP).

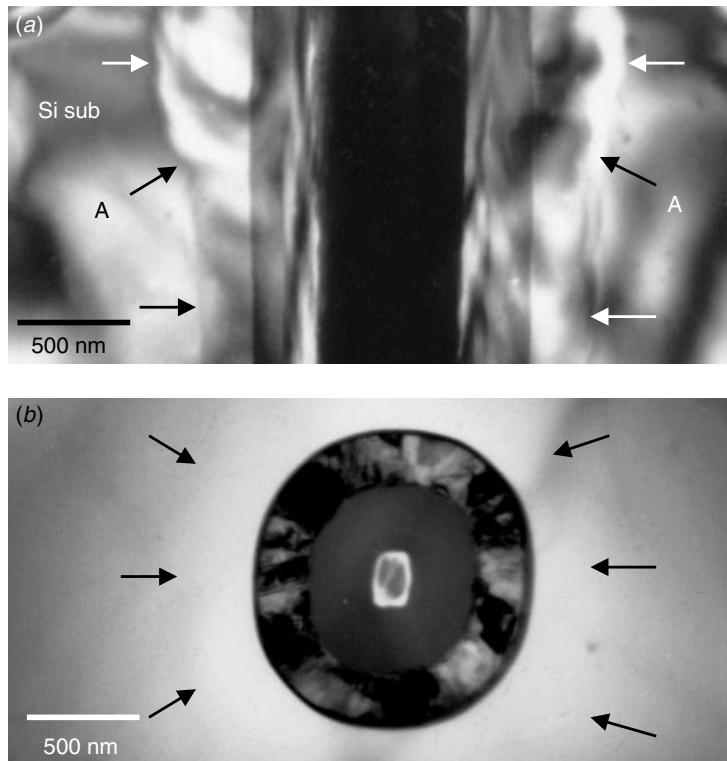


Figure 11.25 TEM (a) cross section and (b) plan view of the capacitor's trench structure. Both are after junction delineation and reveal the substrate's doping along the trench's sidewall. Arrows indicate the junction revealed by delineation, which is much clearer in cross section than in plan view. Arrow A shows the junction's position change along the trench's depth direction.

Another important issue in regard to trench doping is the substrate's defect formation. The tilting during ion implantation introduces amorphization along the trench's sidewall, which needs to be annealed and re-crystallized to restore its crystallinity. Residue dislocation loops can form, however, as is usual in ion implantation at the device's junction, and seen in Figs. 11.26 and 11.27. A detailed discussion of ion implant-induced substrate defects was provided earlier in Chapter 5.

16M to 64M DRAMs

Most early 16M trench DRAM cells were based on the 4M DRAM designs as these appear no distinctive differences in their basic structures. However, in a later generation of the 16M DRAM, a new design concept was adopted into the DRAM market. Eventually product incorporated several revolutionary design and process concepts that transformed the DRAM trench process. The new designs were adopted universally in all of the later generations of DRAM trench processes, and they are presently being used for DRAMs in size up to a Giga-bit.

The design revolution was started through a collaborative effort among three trench cell DRAM manufacturers, namely IBM, Toshiba, and Siemens (Nesbit et al. 1993).

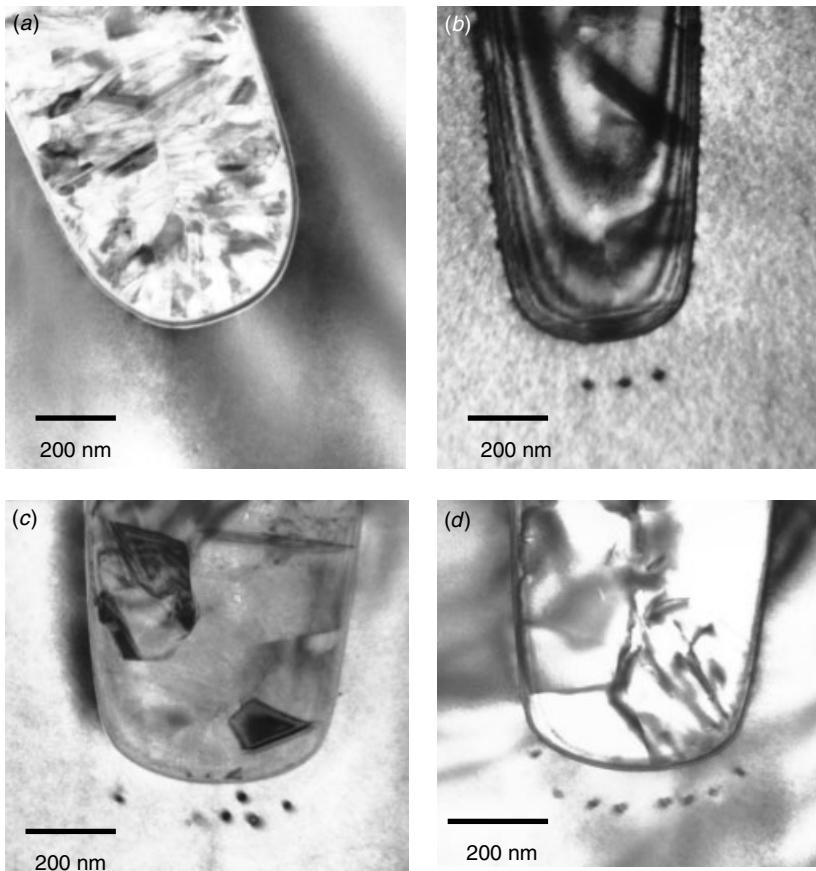


Figure 11.26 Bottom dislocation loops of the capacitor's trench due to ion implantation: (a) without defect and (b, c, d) with dislocation defects.

Although the published data show 0.25 μm technology in the 256M DRAM product, a much earlier version obtained from the market shows the 0.4 μm technology being implemented in a 16M DRAM product using the same design with minor modifications. Figure 11.28 shows the product's cross section in the word-line and bit-line directions. A few things can be noted from these low-magnification images:

- The trenches are 8 μm deep. They are elongated along bit-line direction. The short axis along word-line direction is only about 0.5 μm in diameter, making the aspect ratio (in depth direction) nearly 16.
- Shallow trench isolation is employed and merged with capacitor deep trench to save precious cell area.
- Polycide is used as poly 1 for the gate and word lines. While W metal and W plugs are for the bit line. Metal 1 is runs in parallel with poly 1 as the word line.

In general, the device already reveals some revolutionary design and process concepts. The trenches are no longer equiaxial (or near equiaxial). The long to short axis ratio

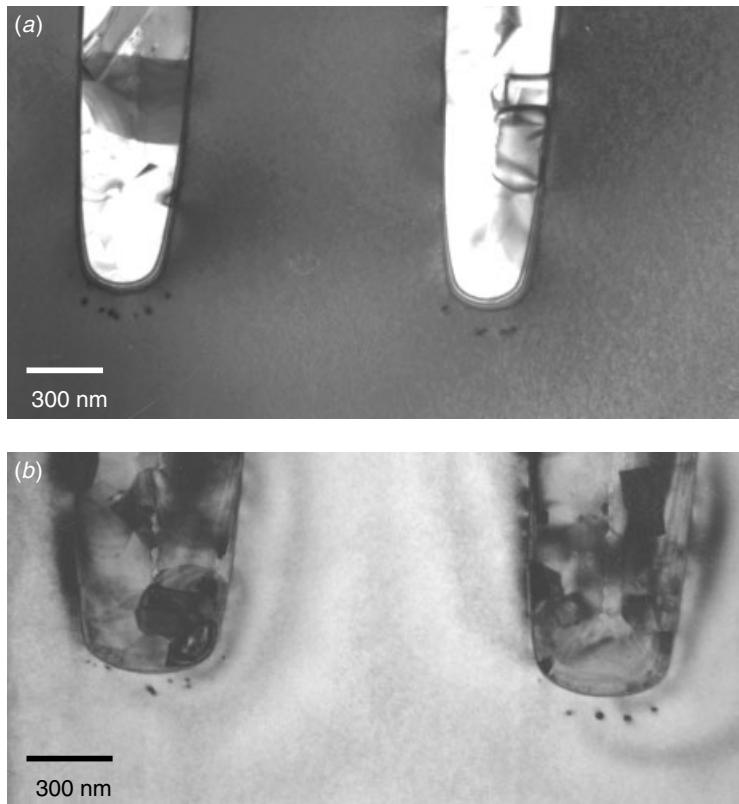


Figure 11.27 Bottom dislocation loops of capacitor's the trench due to ion implantation: (a) along the word-line direction and (b) along the bit-line direction. Images are taken in both direction to show that the defects are dislocation loops and not dislocation lines.

is more than 8. Further the trench's depth is about 8 μm and its shortest diameter is about 0.5 μm , rendering a depth aspect ratio of around 15. These are numbers beyond the imagination of just a generation before. The concept of merging shallow trenches and deep trench is also new. Combining this and the use of special sideway contacts (called BEST contacts, as we noted earlier), cell size can be reduced by 25% (Nesbit et al. 1993)

For a trench as deep as 8 μm , the Si substrate needs to be cleaned and denuded up to at least 15 μm in depth. Figure 11.29 shows this challenge achieved. A closer look at the trenches shows the trenches to be divided into two stages: upper trench and lower trench, Fig. 11.30. In the upper trench (with a higher tapping angle), a thick oxide, the so-called collar oxide, is seen wrapped around the polyfill, and in the lower trench the ONO thin dielectric liner is used around the polyfill. The depth of the upper trench is around 2.5 μm . The collar oxide is used at this upper stage to prevent parasitic transistor leakage (Nesbit et al. 1993). This leakage effect occurs where a vertical parasitic device has the trench n+ polysilicon acting as the gate, the surface strap as the drain, and the buried plate as the source. The collar oxide can prevent such a device from turning on by simply adding some thickness to the

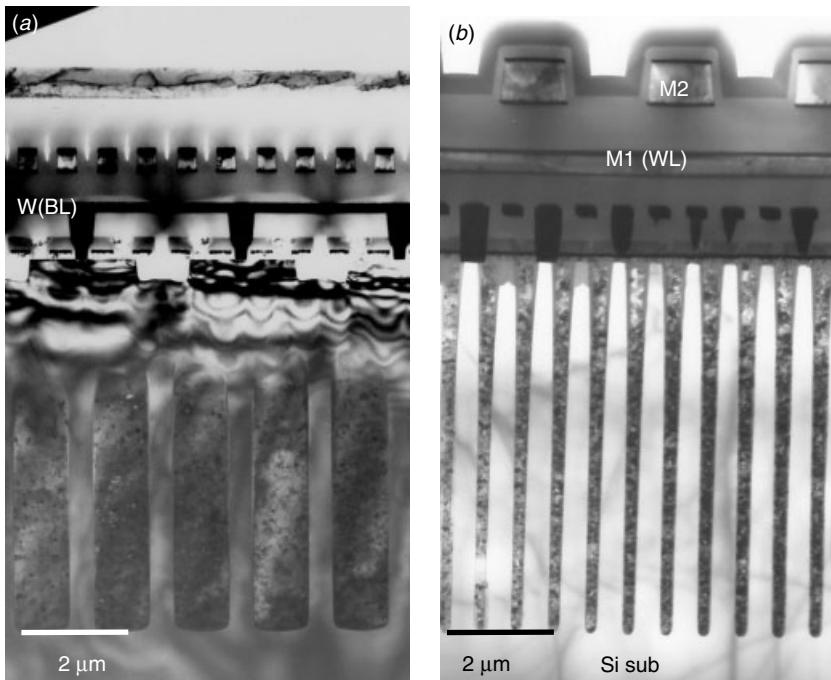


Figure 11.28 TEM cross section of new generation DRAM trench technology along (a) the bit-line direction and (b) the word-line direction. The trenches are elongated along the bit line. The trench depth is about 8 μm . The trench width (along the word line) is 0.5 μm and trench length (along the bit line) is about 1.8 μm . W-polycide is used as the word line and gate, W as the bit line and bit-line contacts.

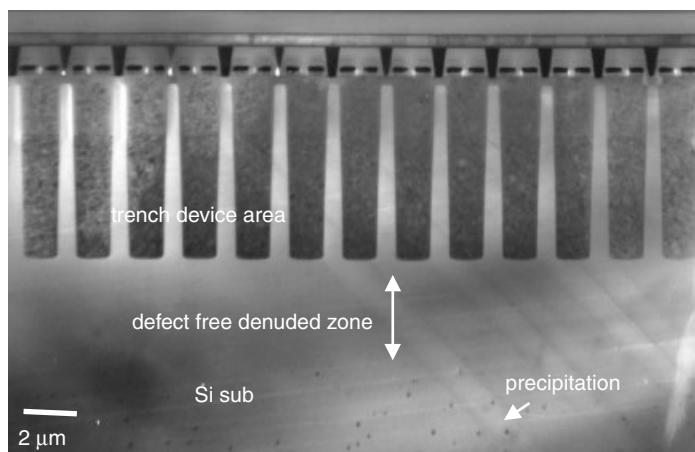


Figure 11.29 Low-magnification cross section shows the Si substrates defect-free denuded zone, below which dispersed precipitation can be seen clearly.

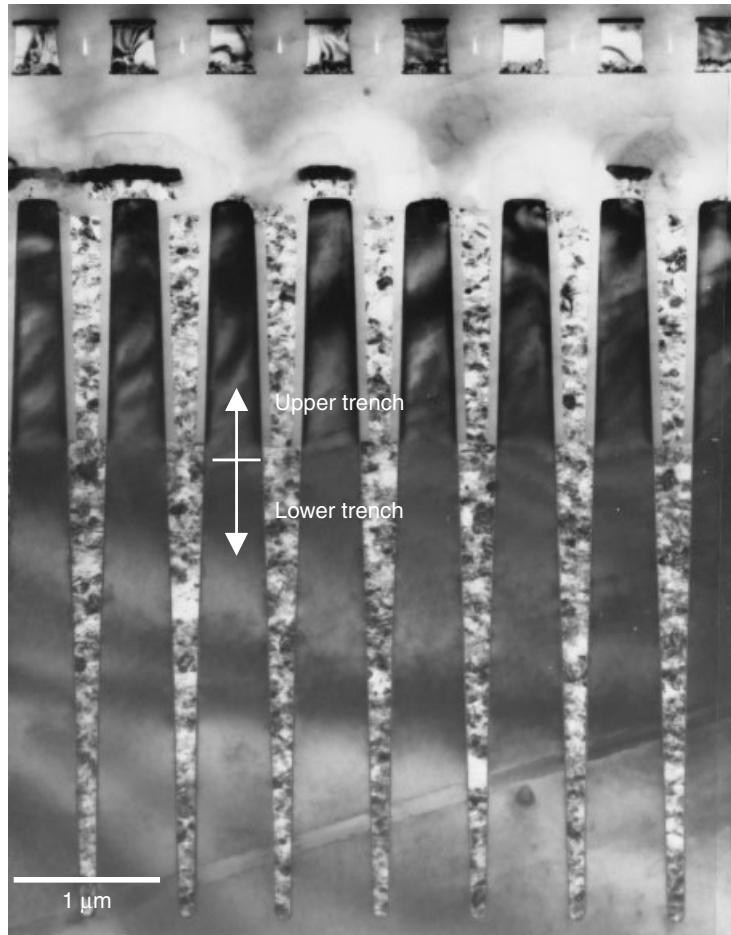


Figure 11.30 Cross section of the new generation DRAM trench technology along the word-line direction. Two stages are found within the trenches. The upper third has a thick oxide liner, which is called the collar oxide. The lower two-third, is the capacitor, which is filled with polysilicon.

gate oxide of the parasitic device. However, making the collar oxide also complicates the process by introducing a two-stage trench refill. The first challenge is to preserve the ONO's integrity. As can be seen in Fig. 11.31, as ONO is attacked, Fig. 11.31(a) it disappears, Fig. 11.31(b). A closer look at Fig. 11.31(b) shows that also the trench's sidewall Si substrate was attacked and etched and a shoulder step was created on the Si substrate. The usual ONO liner across this point should look like those shown in Fig. 11.32, where the ONO is continuous without any disturbance. An additional challenge concerns the buried poly–plate (lower poly–fill) to poly–connect (upper poly–fill) interface. This interface is a vital link in the series of connections between the surface strap and the poly–plate. Compared to the normal version in Fig. 11.32(b), the interface oxide in Fig. 11.32(a) has obviously a resistivity problem.

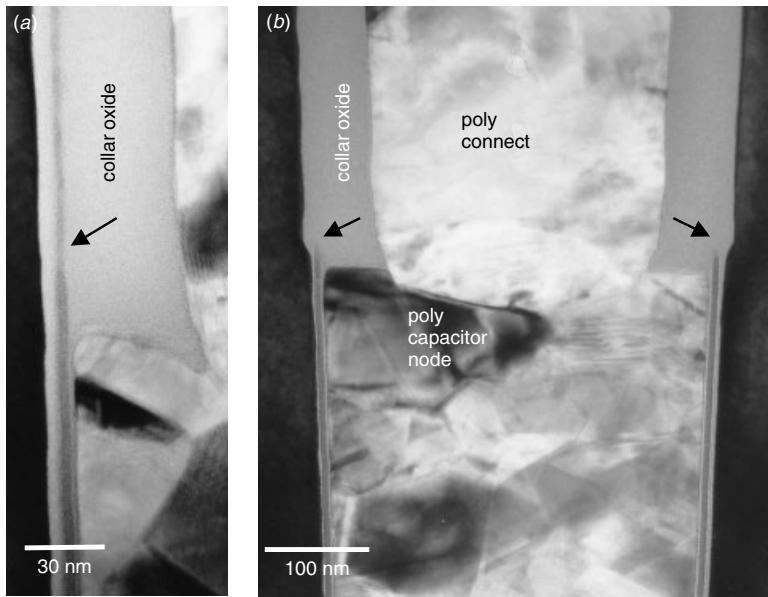


Figure 11.31 Cross section of DRAM trench technology along the word-line direction shows the poly–fill shoulder area with two different ONO structures. (a) the ONO continuous but nitride layer appears thinner within the collar oxide as indicated, and (b) the ONO nitride disappears within collar oxide, as indicated.

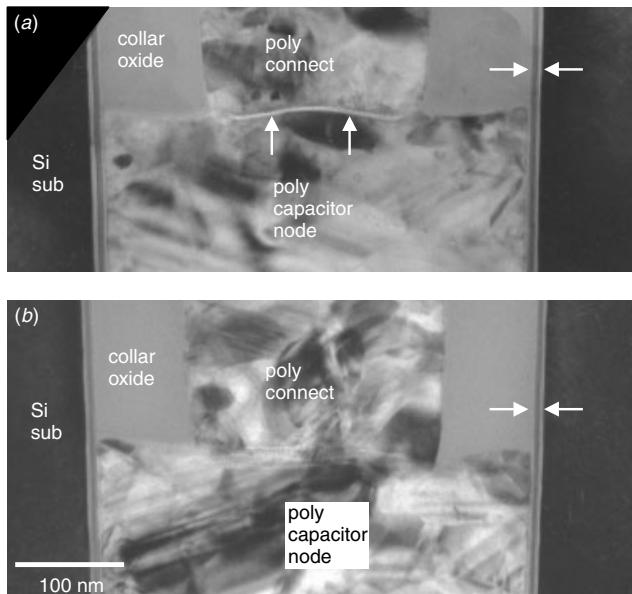


Figure 11.32 Cross section of DRAM trench technology along the word-line direction shows the poly–fill shoulder area. The ONO is homogeneous and continuous as compared to that of the previous figure. A thin oxide at the interface between the upper poly–fill (as the interconnection) and lower poly–fill (as the capacitor node) is shown in (a) as indicated. The problem was corrected in (b).

The two-stage trench, when observed through a planar section, shows an interesting dimensional transition in the depth direction. Figures 11.33 and 11.34 show the trench sectioned at different depths, namely the upper trench, the lower trench, and the very bottom. Noticed that the trench sectioned in horizontal direction looks like a string of chromosome, particularly in the lower part where there is no collar oxide wrapped around the poly. The horizontal section shows that the trench does not exactly take a straight line along Si(110) lattice plane. The two ends of the elongated trench are slightly tilted to the opposite sides, which is what causes the chromosome shape. The slight tilt was made to situate the transistor devices on both sides of the trench.

The veracity of the planar section is best illustrated by another detail. Figure 11.35 shows low-magnification plan view images on a periphery area of the DRAM device. The sample was sectioned through W-polycide gate structure, in Fig. 11.35(a), and the W-metallization in Fig. 11.35(b). The local interconnects, the transistors, the contacts, and VIAs are all indicated clearly. Circuit tracing and reverse engineering is usually done using optical microscopy or SEM. But, as the device shrinks and technology advances to under the 100 nm generation, this technique will become more

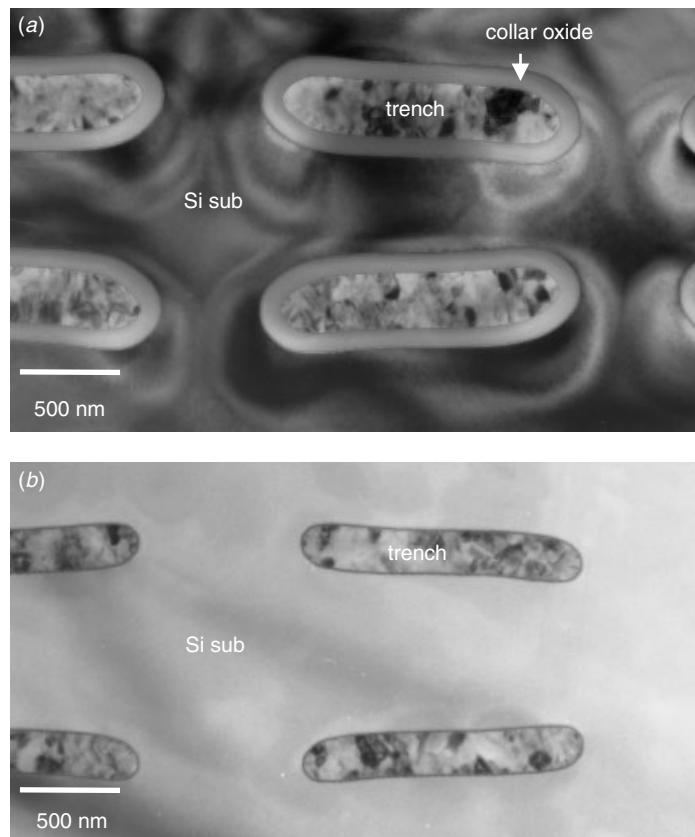


Figure 11.33 Planar sections of the upper (a) and lower (b) parts of the trench capacitor. The chromosome-shaped trench design increases the capacitor's surface areas. The trenches are slightly tilted away from Si(110) to incorporate the transistors and STIs in between.

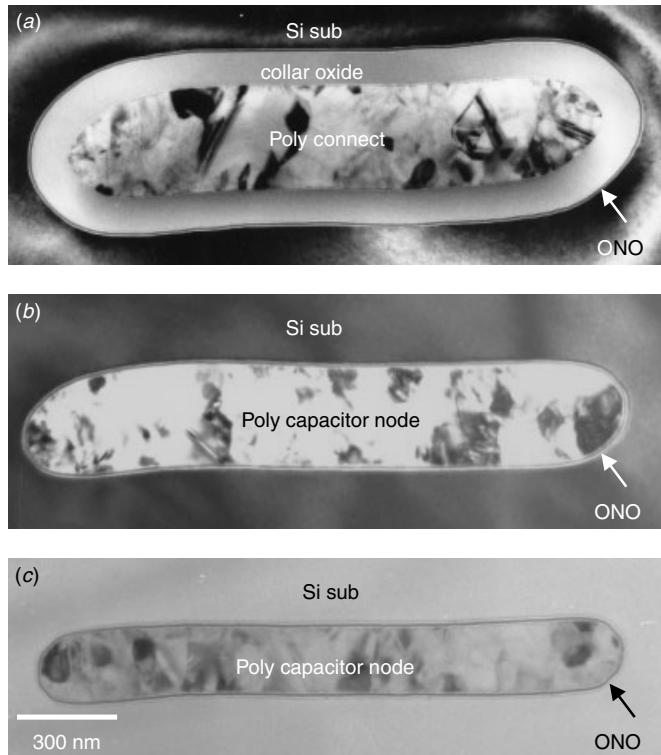


Figure 11.34 Planar section of the trench capacitor structure at different depths from the Si surface. (a) Upper trench where the collar oxide surrounding the poly–fill interconnects, (b) middle trench with ONO and capacitor polysilicon, and (c) lower part of the trench where the trench dimension shrinkage is observed as compared to the (b).

common in the near future. For DRAM reverse engineering, the TEM plan view has a distinct advantage over the conventional SEM. Figure 11.36 shows a DRAM planar section through the word-line and contact layers. The area is at the cell's edge in the transition to the periphery. The image shows a lot of information that cannot be obtained from SEM topological analysis. It clearly reveals the relationships among word line (W-polycide), bit-line contacts, transistor active areas, and surface strap contacts: how all of these end at the cell block's edge and how word lines make contact at the block's edge and make their transition to the circuits at the periphery.

The surface strap's contact is a key technology that is expected to enable a new cell structure. In this product the boron out-diffused surface strap (BOSS) is used as we mentioned earlier. Figure 11.37 shows a plan view and a TEM cross section of BOSS contacts. A few features should be noted:

- As seen in Figs. 11.36 and 11.37, the BOSS contacts run in a zigzag along the word-lines spaces at both sides. Recall that the BOSS contact is self-aligned with no photomasking steps involved. As a result their boundaries are determined by the oxide.

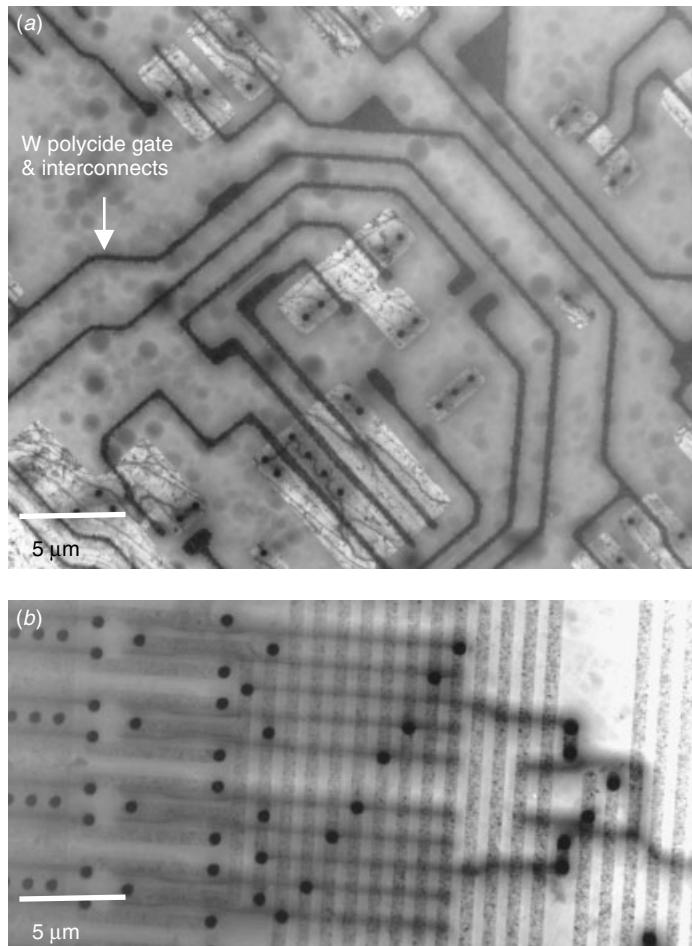


Figure 11.35 TEM plan view of the periphery areas shows (a) the W–polycide gate and interconnection and (b) the W interconnects (horizontal) and Al metal 1 (vertical) in the word-line driver area and how they connect to each other by W–VIAs. The dark circular dots are the W–plug contacts and VIAs, which are used as bit-line contacts in the cell area.

- The surface morphology and horizontal boundaries of the BOSS contacts are irregular. They are determined by the boron out-diffusion and its subsequent selective etching.
- A nitride cap is used on top of BOSS contact.

An important difference in this 16M DRAM is its transistor-to-capacitor arrangement. All previous designs (and even some subsequent designs) had put the trench capacitors at the transistor's tip. That is, the bit-line contact (source), gate, drain, and trench were all on the same straight line, and they usually aligned along the Si[110] direction. However, for this design the trenches were placed side by side with the transistors, which yielded two immediate advantages. First, the trench could be elongated to nearly the length of the transistor, and this greatly increased the capacitor's surface

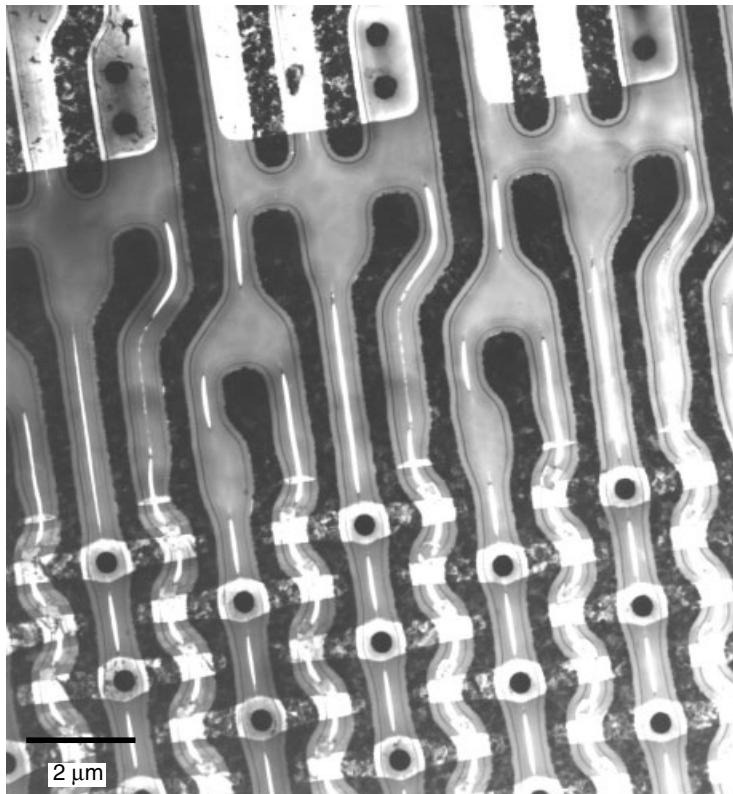


Figure 11.36 TEM plan view of the periphery to the cell transition area. The W polycide, which is the gate and word line in the cell area, is used as the local interconnects to the periphery area. Notice that the W polycide pattern is repeated every four word lines.

area and the capacitance. Second, the inter-transistor isolation (now using shallow trench isolation) can be merged with capacitor trench, and this practically eliminates the isolation area. Figure 11.38 shows a planar section along the near-Si-surface transistors, STIs, and upper trench structures. Notice that there is no STI needed in between the transistors, since the trench capacitors are filled in and take up nearly 100% of the space between the transistors. The only place for the STI is at the tip of the transistors, where the four transistors are separated by an STI. The upper trench's collar oxide thus serves as a local isolation. The result of this ingenious design is that a 64M DRAM unit can be easily put into a 400 mil plastic package by no less than the 0.4 μm design rule and process technology. This was a daunting challenge that could not be met otherwise by any concurrent technology, including the stacked capacitor cell design.

The 16M to 64M DRAM generations of trench devices are summarized in Fig. 11.39. In the sectioning through different layers, we can see one-to-one relationships among the building blocks of the bit cells. In understanding how each one of the building components can be fitted together, we can analytically re-construct the bit cell and even the whole DRAM device.

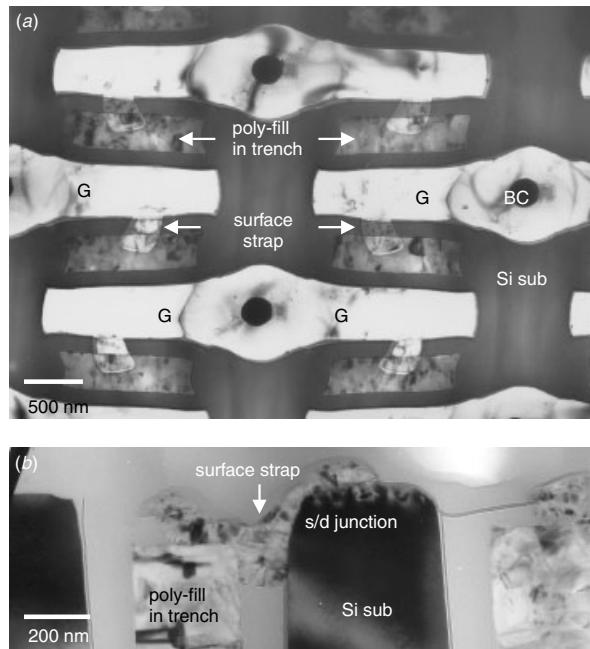


Figure 11.37 TEM (a) planar and (b) cross section of the surface strap contacts used for 16M to 64M generations of DRAM. The sample was delineated to enhance the contrast of the surface strap's contact. Transistors (G: gate, BC: bit-line contact) are aligned in parallel with the capacitor side by side, as can be seen from plan view image.

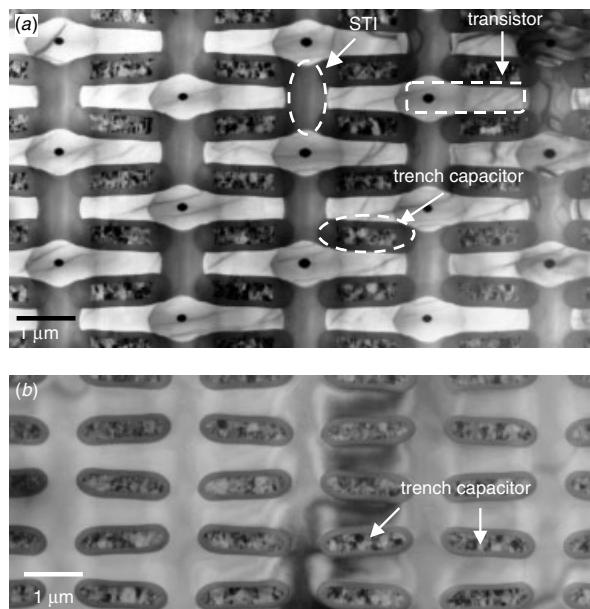


Figure 11.38 (a) Plan view of the geological relationships among the transistors, shallow trench isolations, and the trench capacitor. Each transistor area includes one bit-line contact (dark dot in the center) and two gates and drain areas at each side. The corresponding trench capacitor (planar section through a lower part of the device) is shown in (b).

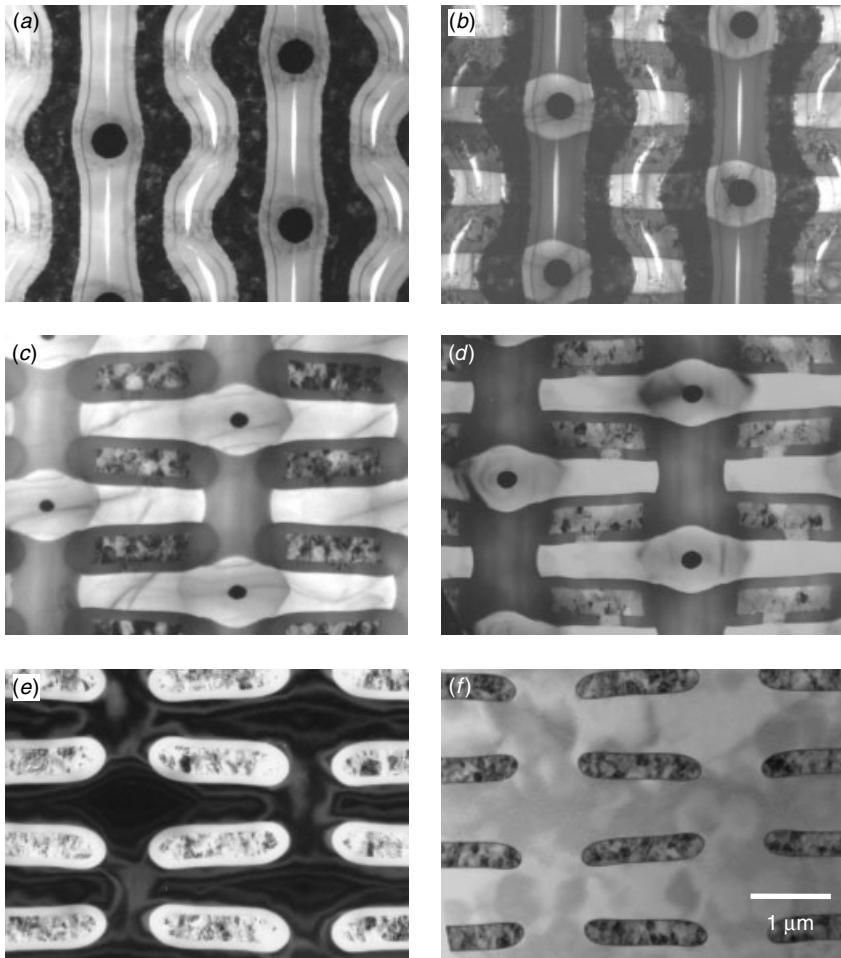


Figure 11.39 Planar section of the DRAM device's trench at different layers. The detailed structures of each device layers and their interrelationships are (a) word lines and bit-line contacts, (b) word lines, bit-line contacts, and transistor active areas, (c) transistors, STI, and trench capacitors, (d) transistors, trench poly–fill, and surface strap contacts, (e) trenches with collar oxides, (f) trenches with capacitor nodes.

64M to 256M DRAMs

Because of collaboration among IBM, Toshiba, and Siemens, trench cell DRAM technologies have converged into one main stream technology. In the late 16M DRAM and early 64M DRAM technology generations, no other technology had appeared in the commodity market except for that joint development by these three companies. To follow the development of trench technology is thus to follow their particular trench structure and see how it evolved through the different generations.

Figure 11.40 shows a cross-sectional view through word-line and bit-line directions. The first thing to notice is that the shape and arrangements of trenches have

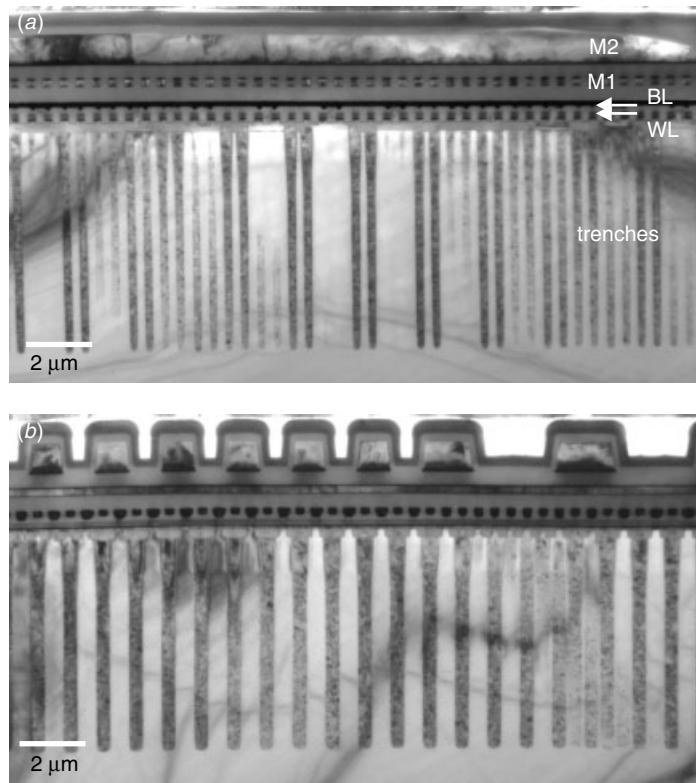


Figure 11.40 A 64M DRAM with a trench cell. The cross section was taken along (a) the bit-line and along (b) the word-line directions. Notice that the trench's shape and arrangement are different from what was observed in Fig 11.28.

been changed again. The long and short axes of the trenches are no longer so different and trenches are paired along the bit-line direction. W-polycide is used as the word line and gates. W is used as the bit line, while polysilicon plugs are used as bit-line contacts. Al metal-1 in the word line reduces the resistivity. Al metal-2 runs in parallel with the bit line as the group bit lines. A closer look at the trenches in Fig. 11.41 reveals that the upper and lower trench depth ratio is much smaller now, as the upper trench's only 0.8 μm deep and lower trench 5.2 μm . This gives a ratio of upper/lower trench equal to 6.5, whereas the previous generation's ratio was around 3 (see Fig. 11.30). Figure 11.42 gives a planar section of the trench capacitors. Chromosome-shaped trenches in a rectangular arrangement are no longer used (see Fig. 11.33). Instead, oval-shaped trenches in a honey comb arrangement are used. The long axis of the oval-shaped trenches is 0.63 μm and short axis is 0.5 μm . This gives a ratio of about 1.25, while for the previous generation, this ratio was around 3.3. If we use the short axis as the basis in calculating the trench's aspect ratio (diameter/depth), the previous generation gives us a ratio of 16 (0.5 $\mu\text{m}/8 \mu\text{m}$) while for this generation it is nearly 12 (0.5 $\mu\text{m}/6 \mu\text{m}$). All of these measurements show a very different way of making trenches and a completely different design and process concept. Such a leap in process development is rare in the ULSI process development history.

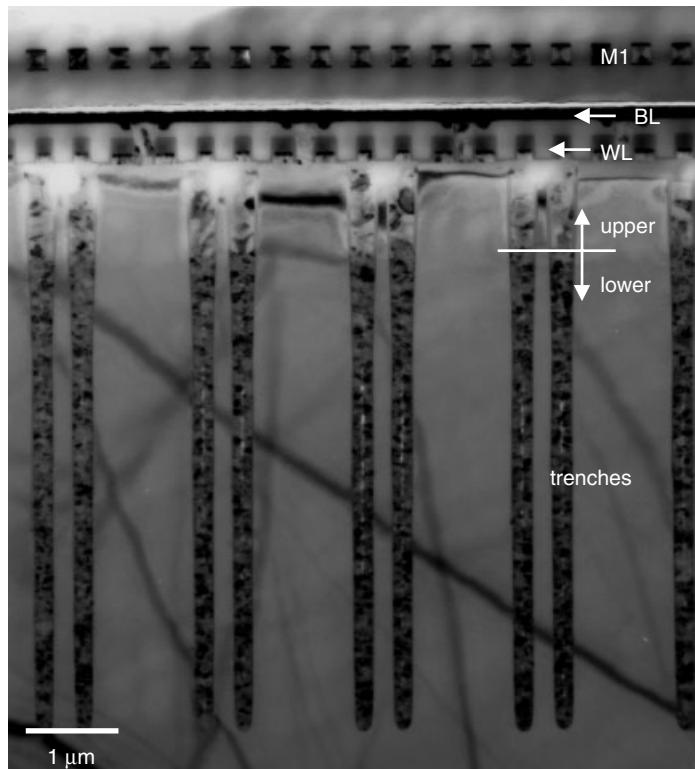


Figure 11.41 A 64M DRAM with a trench cell. The cross section was taken along the bit line, where the trenches appear to be paired.

In contrast to previous designs where trench was elongated and aligned in parallel with its control transistor, in this DRAM the trenches were returned to the ends of the transistor pair. Therefore it is possible to cross section the bit cell with the transistor and capacitor, all in the same view. The unit bit cell of this 64M DRAM is shown in Fig. 11.43. W is the bit line with a polysilicon plug as the bit-line contact. Notice that the poly-plug CMP and its subsequent etching have induced a circular ditch around the poly-plug, and this is seen in Fig. 11.43 as two tungsten dimples around the poly-plug. W-polycide forms the word line and gate. A large nitride head piece greatly increases the gate's height and a nitride cap covers the word lines. Self-aligned poly-plug bit-line contacts then follow. Despite its advanced design concept, the bit-cell structure has a deceptively simple appearance. The important components are the node contact and the bit-line contact. The BEST contact, as was described earlier, occupies virtually no surface space. Thus the design's self-aligned poly-plug bit-line contact also saves some space. The other area-saving aspect of the design comes from the STI. Figures 11.44 and 11.45 show the STI position viewed along the word line and the bit line, respectively. When combined, these two images easily tell us that the STI overlaps with the trench capacitor. Its position is just sufficient to isolate the two different active cell areas. Actually the STI is H-shaped (or rail-shaped) with two trench capacitors at each

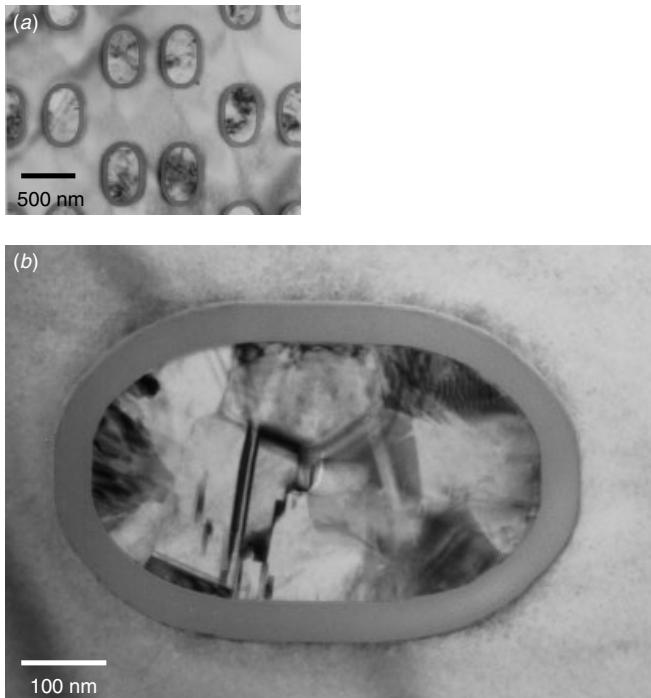


Figure 11.42 Planar section of a trench capacitor. (a) At low magnification the trenches are arranged on the Si substrate a honey comb pattern. (b) A close-up of one trench capacitor shows clearly the collar oxide and the oval shaped trench cross section.

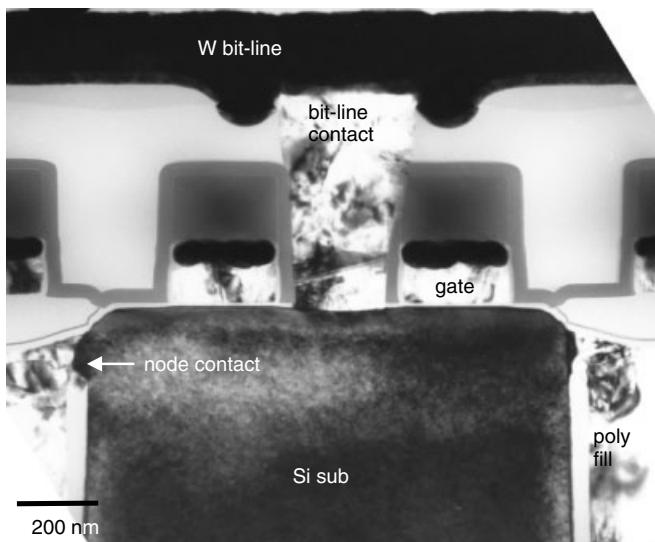


Figure 11.43 A 64M DRAM device's bit cells with W–polycide as the transistor gate. The self-aligned polysilicon plug as the bit-line contact and a W bit-line is used. The trench capacitor has a buried strap (BEST) node contact. The bit cell looks deceptively simple and straightforward.

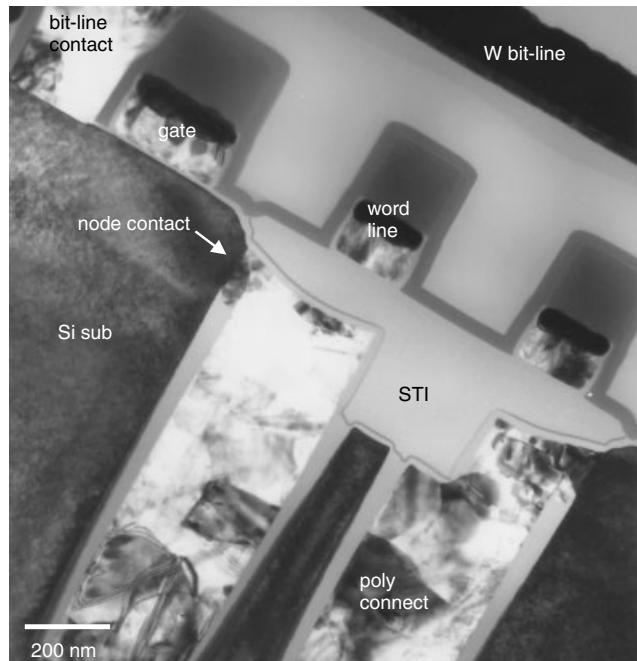


Figure 11.44 A 64M DRAM bit cell. In the trench structure, the BEST contacts do not occupy any surface area. The STI is made to overlap with the trench capacitor and thus does not occupy any surface area either.

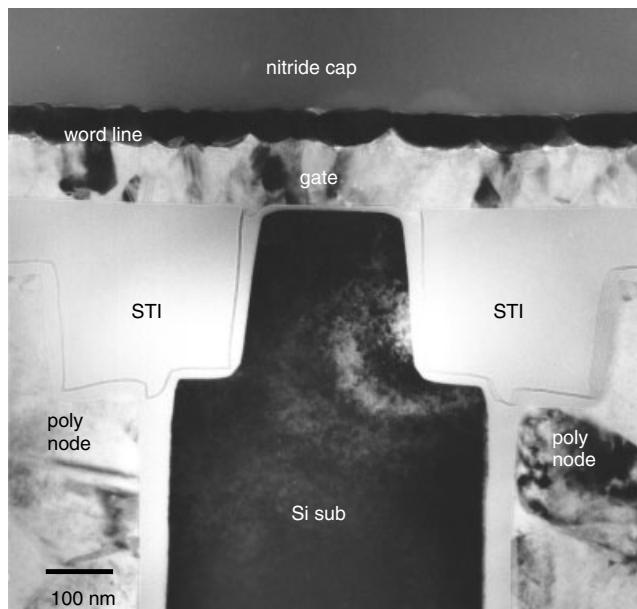


Figure 11.45 Cut along the word-line direction, the overlapping structure of the STI and trench capacitor is clearly visible. As combined with the previous figure, STI is basically situated half-way from the top of the trench's polysilicon and half-way from the top of the Si substrate.

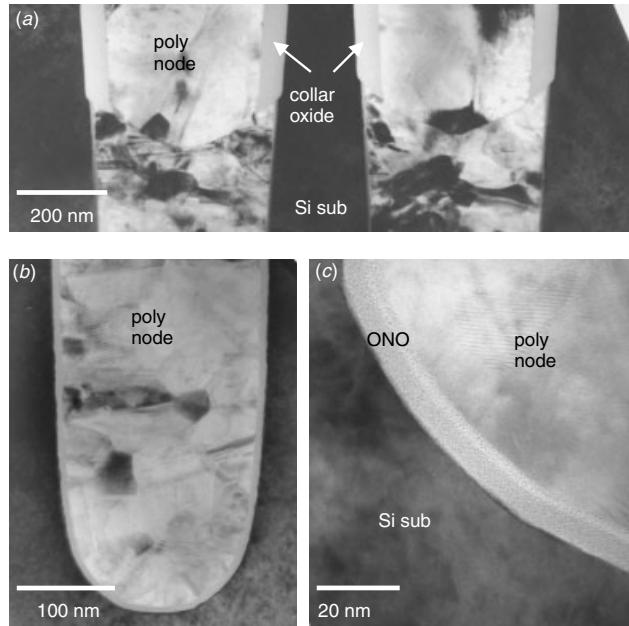


Figure 11.46 Close-up of the trench capacitor. (a) The shoulder area of the bottle shaped trench capacitor. (b) Trench's base which takes a semispherical shape. (c) HRTEM of the trench's base shows the ONO dielectric layer.

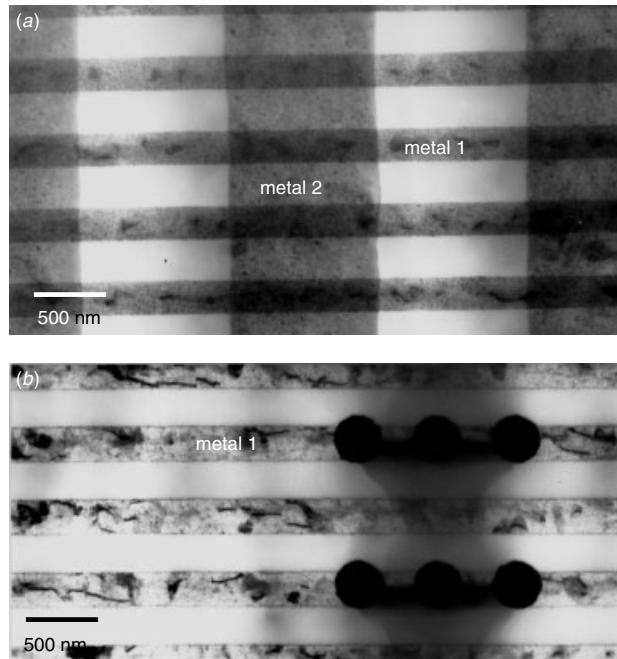


Figure 11.47 Planar section of the 64M DRAM cell area along metal 2 and metal 1. (a) Metal 2 and metal 1 are perpendicular to each other, but metal 2 is much wider than metal 1. (b) Metal 1 and W–plug contacts.

side while the two bit-line contacts sit on its top and bottom. Later we will come to view this characteristic of the structure when we perform planar section analysis. The STI oxide also connects with surface dent of the trench polyfill, and the whole oxide piece is wrapped, from side to side, and up and down, by the nitride cap, as seen in Fig. 11.44. This oxide piece altogether acts as an effective isolation structure.

This new trench configuration is called a “bottle trench.” The name is apparent for reason of the polysilicon fill resembling a beer bottle, with upper trench forming the long neck of the bottle when viewed in cross section along the word line. Figure 11.46(a) shows the shoulder of the bottle. There is no ONO on the upper trench now (again, unlike its precedents). Figure 11.46 also shows the trench bottom with its beautiful spherical shape and ONO lining.

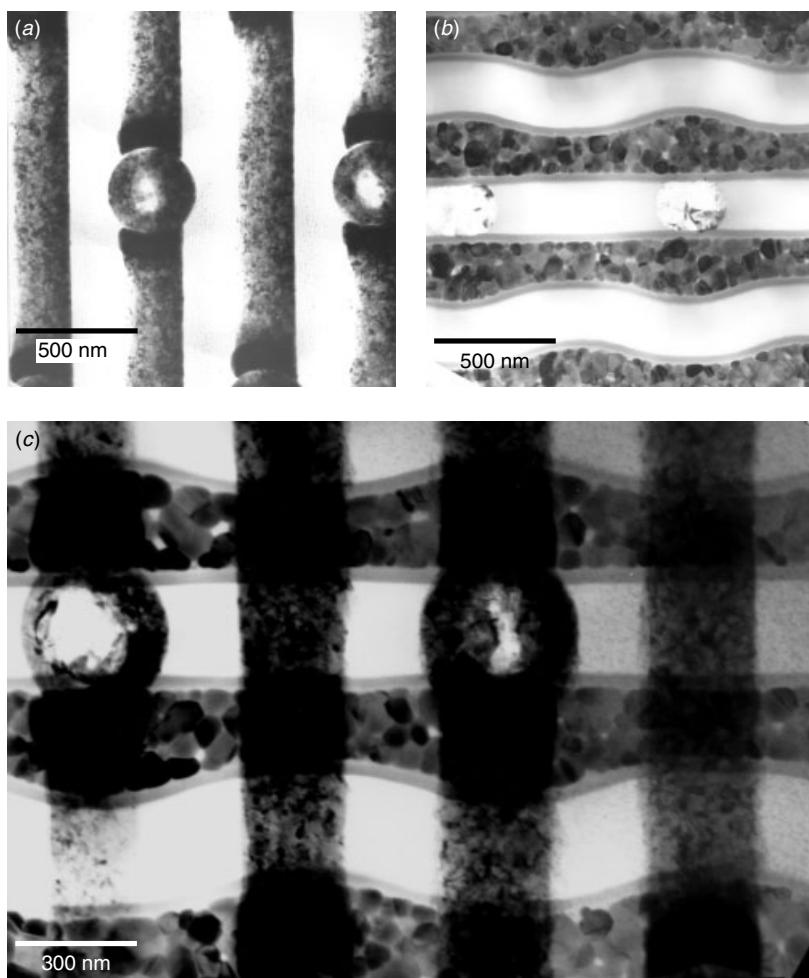


Figure 11.48 Planar section of the 64M DRAM cell area along (a) the W bit lines and (b) the W-polycide word lines. (c) Area enlarged where bit lines and word lines overlap. Notice that the bit-line contacts in (a) are seen in (b) but only smaller.

Planar sectional analysis, which is complimentary to cross-sectional analysis, was used to further illuminate the structure of this DRAM. In the planar section, Figs. 11.47–11.51, the following characteristics were noted:

- In the cell area, metal-2 is running in bit-line direction while metal-1 is running in word-line direction. The two are perpendicular to each other, as seen in Fig. 11.47. The same can also be observed in Fig. 11.40.
- Tungsten is used as the bit line and tungsten polycide as the word line. The two are perpendicular to each other, as seen in Fig. 11.48. As noted in Fig. 11.43, the bit-line contacts were formed by a poly-plug with a large tungsten ring ditch. The tungsten ring can be seen in Fig. 11.48(a) and (c). Poly-plug bit-line contacts can be seen in Fig. 11.48(b). The rings and the plugs can be matched exactly, as shown in Figs. 11.48(c).
- W-polycide and metal-1 both are used as word lines, and they run in parallel in exactly one-to-one correlation, as seen in Fig. 11.49(a) and (b). Periodic stitch contacts connect metal-1 to W-polycide, as seen in Fig. 11.49 (b).

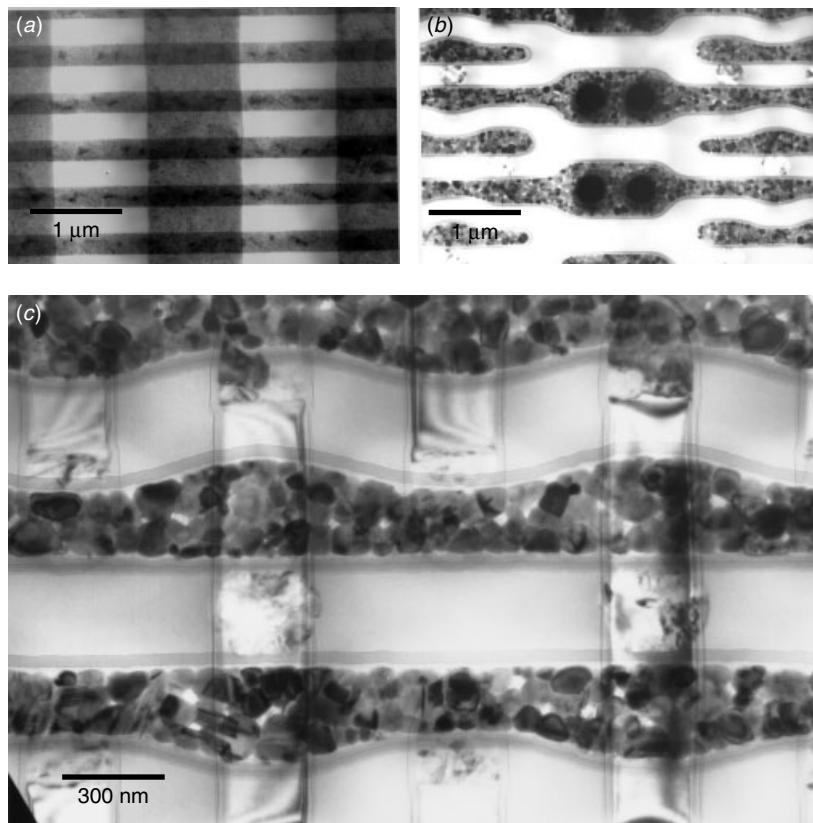


Figure 11.49 Metal 1 word line and W-polycide word line in a one-to-one correlation in (a) and (b). Also shown in (b) is the W-plug stitch contact that connects metal 1 to W-polycide word lines. (c) Word lines are related to the transistors and bit-line contacts.

- W-polycide, when viewed from top, does not run in a straight line. As shown in Fig. 11.48(c) and 49(c), W-polycide width varies from 0.26 μm in the gate areas to 0.16 μm in the areas on top of the trench capacitors.
- Noticed that the W-polycide gates are wrapped with nitride cap layers and no LDD spacer is used. This is the 0.25 μm process technology which does not use the LDD structure (Figs. 11.43 and 11.49).

Figure 11.50 shows the transistors, Fig. 11.50(a), the trench capacitors, Fig. 11.50(b), and how the two are coupled together, Fig. 11.50(c). Two trench capacitors sit at both ends of a transistor pair, making the structure, when viewed from top, look like a dumbbell. Each dumbbell consists of a bit-line contact at the midpoint, two gates next to the bit-line contact, two drain active areas follow, and finally two trenches at both ends.

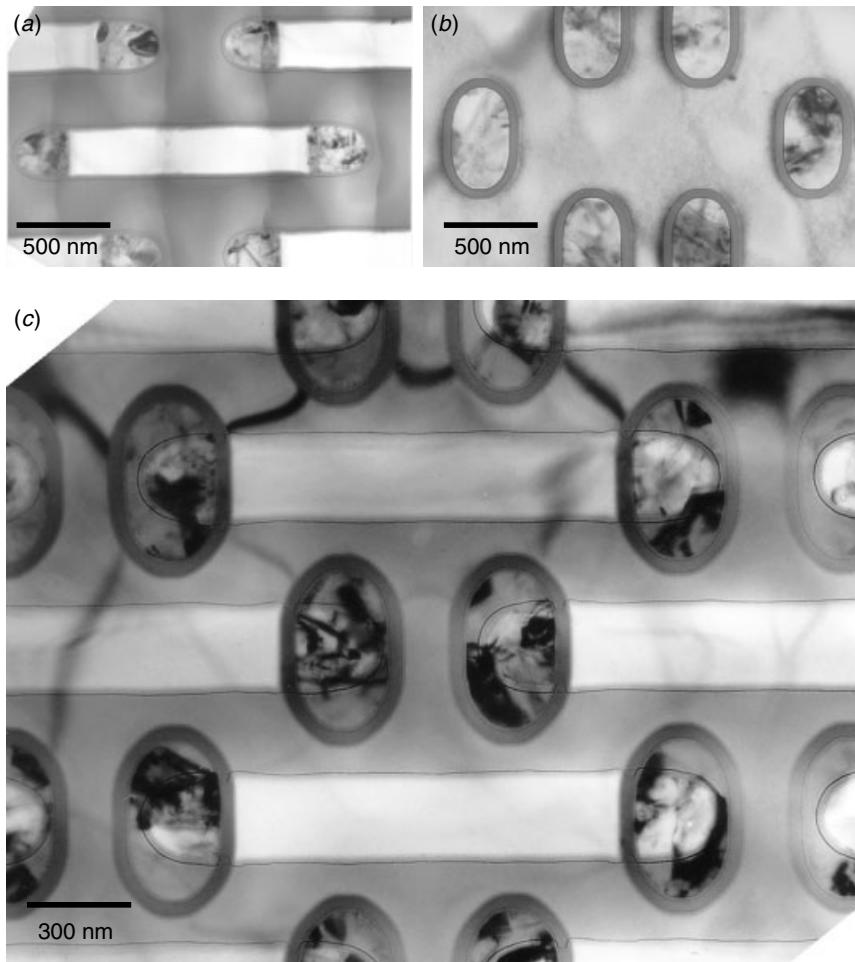


Figure 11.50 (a) transistor pairs and the BEST contacts at its both ends, (b) trench capacitor arrays. (c) putting these two together, one can clearly see how the two are connected to each other through BEST contacts and how they are arranged.

Figure 11.51(a) through (j) summarizes the structure of this DRAM, layer by layer, in planar section. All these layers were laminated in sequential order to reconstruct the entire DRAM cell structure. Bit-line contacts, and subsequently trench capacitors, can be used as the alignment marks, and the reader is encouraged to stack the images and see how each layer is related to the layers above and below it.

256M to 4Giga-bit DRAMs and Beyond

As we showed above, the 64M DRAM structure is deceptively simple in appearance. Indeed, close examination reveals that because the structure looks so simple, there may be plenty of room for the structure to be further reduced and improved by sophisticated design and process technology (Ozaki et al. 1995; Hamamoto et al. 1993). For example, a diagonal layout (or tilted array) and surface strap trench cell could increase its storage capacitance, and reduce the junction leakage current (Kajiyama et al. 2000; Hoenigschmid et al. 2000). The development of this trench technology seems be capable of bringing DRAMs well into the Giga-bit regime.

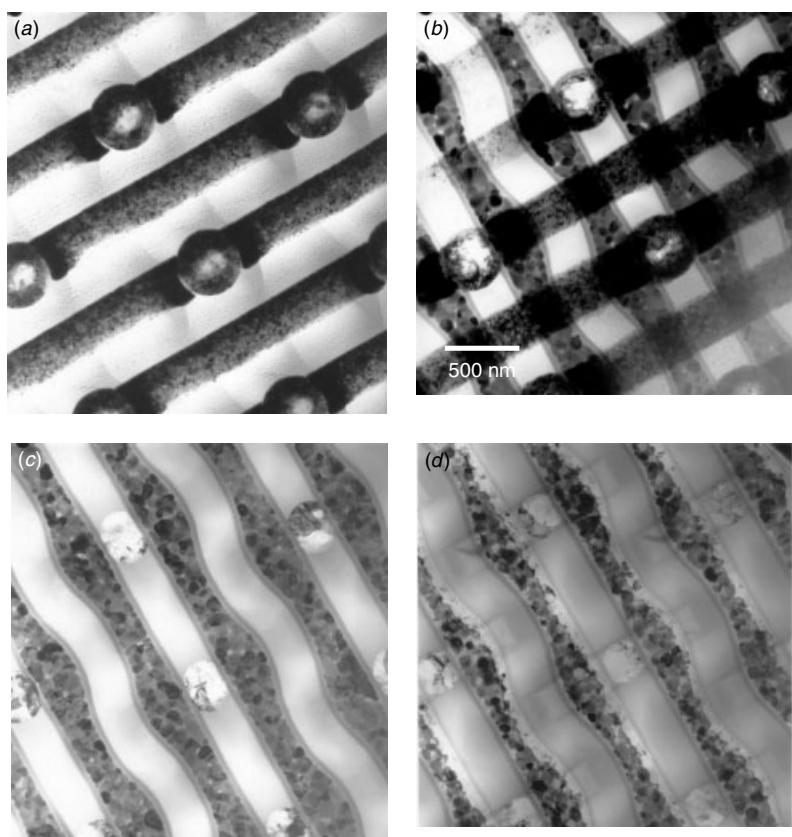


Figure 11.51 Planar section of 64M DRAM cell area (a) bit lines, (b) bit lines and word lines, (c) word lines and bit line contacts, (d) word lines and transistors, (e) word line and transistors, (f) transistors, (g) transistors and BEST contacts, (h) transistors and STIs, (i) transistors, BEST contacts, and trench capacitors, (j) trench capacitors.

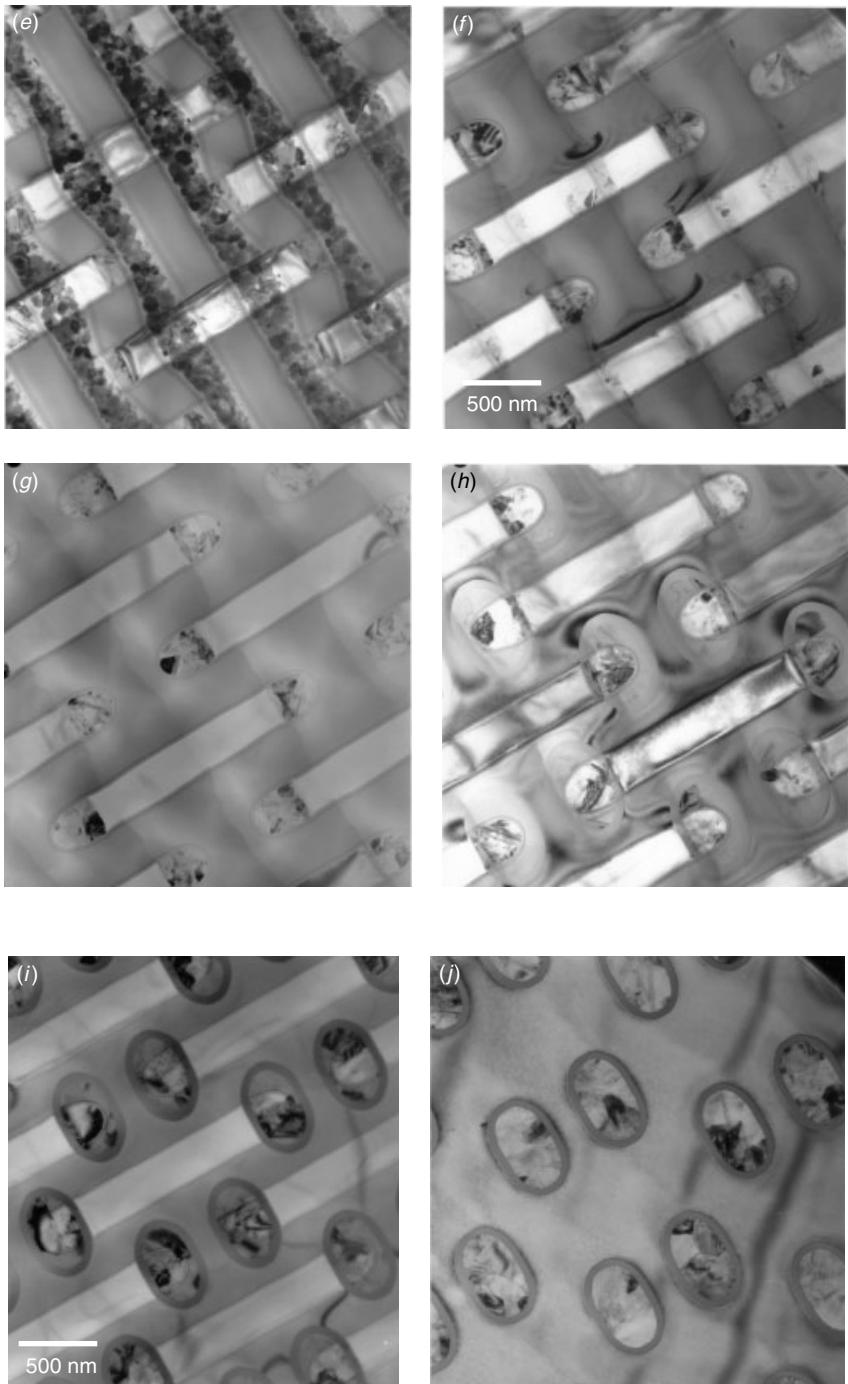
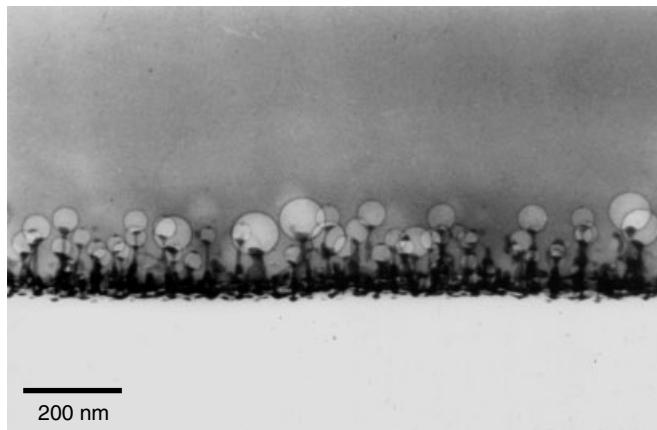


Figure 11.51 (*continued*)

REFERENCES

- E. Adler, J. K. Debrosse, S. F. Geissler, S. J. Holmes, M. D. Jaffe, J. B. Johnson, C. W. Koburger III, J. B. Lasky, B. Lloyd, G. L. Miles, J. S. Nakos, W. P. Noble Jr., S. H. Voldman, M. Armacost, and R. Ferguson, *IBM J. Res. Develop.*, **39**(1/2), 167–185, 1995.
- T. Hamamoto, T. Yamada, M. Aoki, S. Ishibashi, H. Kawaguchiya, Y. Ishibashi, H. Hashimoto, and H. Kanai, *IEDM*, 643–646, 1993.
- H. Hoenigschmid, A. Frey, J. K. DeBrosse, T. Kirihata, G. Mueller, D. W. Storaska, G. Daniel, G. Frankowasky, K. P. Guay, D. R. Hanson, L. L. C. Hsu, B. Ji, D. G. Netis, S. Panaroni, J. Alsmeier, W. Weber, M. R. Wordeman, *IEEE J. Solid State Cir.*, **35**(5), 713–718, 2000.
- T. Kajiyama, H. Aochi, Y. Asao, M. Morikado, H. Koyama, K. Sugimae, S. Ishibashi, K. Hosotani, Kito, A. Sato, M. Kido, M. Sakuma, M. Sato, S. Watanade, N. Miyawaki, and T. Hamamoto, *IEDM*, 357–360, 2000.
- K. Kato, *IEEE Trans. Electron Devices*, **35**(11), 1820–1828, 1988.
- N. C. C. Lu, *IEEE Circuits and Devices Magazine*, 27–36, 1989.
- L. Nesbit, J. Alsmeier, B. Chen, J. DeBrosse, P. Fahey, M. Gall, J. Gambino, S. Gernhardt, H. Ishiuchi, R. Kleinhernz, J. Mandelman, T. Mii, M. Morikado, A. Nitayama, P. Parke, H. Wong, and G. Bronner, *IEDM*, 627–630, 1993.
- T. Ozaki, M. Noguchi, M. Habu, M. Aoki, Y. Ishibashi, T. Shino, T. Hamamoto, D. Takashima, H. Niyyama, T. Nakasugi, T. Shibata, Y. Kato, E. Nishimura, K. Hattori, T. Magoshi, S. Sato, H. Yamaguchi, and K. Sugihara, *IEDM*, 661–664, 1995.
- H. Sunami, K. Shimohigashi, N. Hashimoto, *Solid-State Circuits, IEEE Journal*, **17**(2), 237–240, 1982.
- S. Wolf, *Silicon Processing for the VLSI Era, Vol. 2: Process Integration*, Lattice Press, 1990.

12 ULSI Devices IV: SRAM



An early silicide process showing TaSi_x needles, each with a bubble at the tip. Cross section TEM view resembles city lights in a foggy night.

Static random access memories (SRAM) are called static because, unlike DRAMs, they do not require periodic refresh signals to retain their stored data. The bit states are stored in a pair of cross-coupled inverters that forms a flip-flop circuit. The two stable states are designated as 1 and 0. Both states will be stable with neither branch of the flip-flop conducting as long as the dc voltage is applied. Thus no refresh is needed to retain data as in the case of DRAM. The interested reader should refer to the book by Prince (1995) for more details on the basic principles of SRAM operations and its associated circuit designs.

SRAMs have historically been a generation behind DRAMs in functional density due to the greater number of transistors in a static RAM cell. For example, a 256 k static RAM has about the same number of transistors as a 1 Mb DRAM, giving it approximately the same chip size in similar technology generations. This means that the yields in the production process are roughly equivalent. Early SRAM developed along three different technology paths, namely bipolar, NMOS, and CMOS. These products are very different in characteristics and are targeted at different markets. Bipolar SRAM is targeted for low density and very fast applications. It had a very small market share and will not be discussed here in this book. The reader interested in special (non-MOSFET) SRAMs should refer to the following references: Mele et al.

(1990), Bohr et al. (1994), Carns et al. (1995), and Kikuchi et al. (1995). Commodity SRAMs using MOSFET VLSI process technology is the focus of this chapter. It can be represented by three different generations of products. The earliest one used NMOS technology. This was followed by a so-called mix MOS or resistor load NMOS with CMOS periphery technology. The latest generation is known as full CMOS technology. By the mid-1980s the majority of all SRAMs were made using CMOS technology. The advantages of CMOS technology include low static power dissipation, superior noise immunity (compared to NMOS and bipolar), wide operating temperature, sharp transfer characteristics, wide voltage supply tolerance, improved reliability, ease of computer modeling, versatility in combining CMOS with bipolar and analog logic on the same chip, and low soft error rate. The disadvantages of CMOS historically have been high cost, slow speed, large chip size, added complexity in manufacturing, and a tendency to exhibit a so-called latch-up (Prince 1995).

The disadvantages of CMOS technology, however, have been minimized in submicron process technology. Dimensional shrinkage has lowered both the chip size and the cost of production but improved device speed dramatically. The introduction of submicron ($<1 \mu\text{m}$) process technology came approximately at the time when the full CMOS SRAMs replaced the mix MOS resistor load SRAM (i.e., around 1M to 4M SRAM product generations).

Before the device structures are introduced, it is important to understand how the basic constitution elements of an SRAM unit cell are connected together. Figure 12.1 shows a schematic of an SRAM unit storage cell. The data are stored as voltage levels with the two sides of the flip-flop (A and B) in opposite voltage configurations.

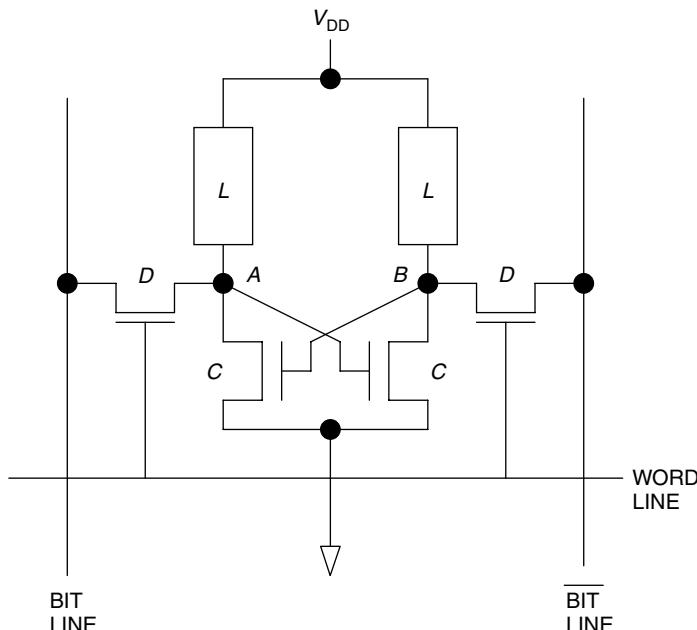


Figure 12.1 Schematic of SRAM unit storage cell. It consists of two load elements (L), two storage transistors (C), and two access transistors (D). Data are stored as voltage levels with the two sides of the flip-flop (A and B) in opposite voltage configurations.

with two sides of the flip-flop in opposite voltage configurations. The purpose of the load device is to offset the charge leakage at the drains of the storage and select transistors. When the load element is PMOS, the resulting CMOS cell has essentially no current drain through the cell except as it is switching. The select transistors and the storage transistors are always NMOS. The polysilicon load resistor, when used in place of the PMOS load transistor in the CMOS cell, can reduce the cell size up to a 30% reduction. This is true since the resistors can be stacked on the cell, therefore reducing the cell size by two transistors. This was the reason why resistor load mix MOS technology was preferred in the early days. However, the resulting cell had higher leakage in standby when not being accessed than a cell with PMOS load because a small amount of current always flows through the resistor.

12.1 PROCESS FLOW

The process flow for a 1 Mb CMOS SRAM memory with a polysilicon thin film transistor (TFT) is shown in Table 12.1. The Resistor load mix MOS process is much simpler than the full CMOS device.

The isolation LOCOS and twin well are first formed, followed by the mask and buried contact (BC) formation. The gate oxide is grown during the anneal step of the implant. The polysilicon or polycide (usually tungsten) is deposited and patterned to form the gate region of the transistor and interconnects. The high-dose n+ and p+ drain implants form the MOS transistors with the lightly doped drain area appearing under the edge of the gate structure as a result of the early implants. A self-aligned contact (SAC) is formed for the contact to n+. Poly 2 deposition (polysilicon or polycide) and mask etch combine to form the ground lines followed by ILD and VIA etching. Poly 3 deposition and mask combine to form the TFT gate layer followed by ILD and

TABLE 12.1 Simplified SRAM Process

Process Step	Purpose and Comments
Form twin well	
Form LOCOS isolation	Forms field oxide
Mask, buried contact (BC)	Form poly 1 local interconnect through BC,
Deposit first poly (WSi_x), mask etch poly 1	Form all device gates
Mask, implant N–, N+, P+	Complete all devices
Mask SAC	Etchs self-aligned contact to N+
Deposit poly 2 (WSi_x), mask, etch	Form ground lines
Mask, DC 2	
Deposit poly 3, mask, etch	Form TFT gates
Deposit TEOS	Forms gate oxide for TFT
Mask DC 3	
Deposit poly 4, mask, etch	Form TFT channel
Mask, metal contact	
Deposit tungsten plug, etch back	
Deposit metal 1, mask, etch	
Deposit IMD	
Deposit metal 2, mask, etch	

VIA etching. Poly 4 deposition and mask combine to form the TFT channel followed by ILD and contact etching. Tungsten–plug deposition and etch-back and Al metal 1 deposition and etching follow. Next come IMD deposition and Al metal 2 deposition and patterning. Finally passivation and bond pad patterning finish the process. The entire process is actually greatly simplified here. In reality an SRAM process can involve as many as 700 to 1000 process steps.

12.2 RESISTOR LOAD SRAM

Starting with 16K SRAM, the polysilicon load resistors have dominated the SRAM commodity markets. These SRAM have consisted of an NMOS transistor matrix with high-ohmic resistor loads and a CMOS periphery whose main advantage was low standby power consumption within the small die size of the NMOS SRAM. The R-load CMOS SRAMs were a logical spin-off from the CMOS technology, since the polysilicon resistors merely replaced the PMOS transistors in the CMOS cell. There is, however, a power dissipation penalty for R-load over CMOS, since the resistors always drain some current. This R-load CMOS was able to combine the scaling advantages of NMOS and the lower power advantage of CMOS over NMOS (Ohzone et al. 1985; Wada et al. 1987; Komatsu et al. 1987; Matsui et al. 1987; Minato et al. 1987; Kohno et al. 1988; Shimada 1988).

Unlike usual DRAM cells where there are only two major components, namely a transistor and a capacitor, an SRAM cell contains six major components: four control transistors and two load elements (resistors or transistors). The most common and economical way to link them together is through several interconnection layers and to squeeze within a small rectangular plane instead of along the lattice plane like DRAM. It is thus impossible to view the whole cell structure and its interconnecting relationship in a cross-sectional view, as we did in the DRAM analysis. Figure 12.2 shows a classical 64K SRAM device using 1.5 μm polycide technology. The sample was delineated using wet chemicals to reveal the junction. The tungsten polycide is poly 1 as the bulk MOSFET gate electrode. The same polycide layer is used for the PMOS gate electrode and is in direct contact with the s/d diffusion of the NMOS using a buried contact, as also shown in Fig. 12.2. An irregularly shaped substrate ditch next to the buried contact is revealed clearly. Such a ditch was a ubiquitous characteristic observed in almost all of the SRAMs. Its formation is illustrated in the following:

- As shown in Table 12.1, buried contact patterning was done after the LOCOS isolation formation and before the poly 1 deposition. Oxide etch was used to remove the pad/gate oxide in the buried contact areas and expose the junction's substrate.
- Poly 1 deposition immediately followed the definition of the buried contact.
- When poly 1 was masked and etched, gate oxide was used as the etch end point in the gate edges. The process ensures that the gate oxide is removed cleanly by overetching the substrate a bit.
- As the gate oxide was removed, the Si substrate was under attacked by the same process in the buried contact's mismatch windows as there was no thin oxide in these areas. The area attacked was the mismatch overlap between the buried contact windows and the poly 1 definition patterns.

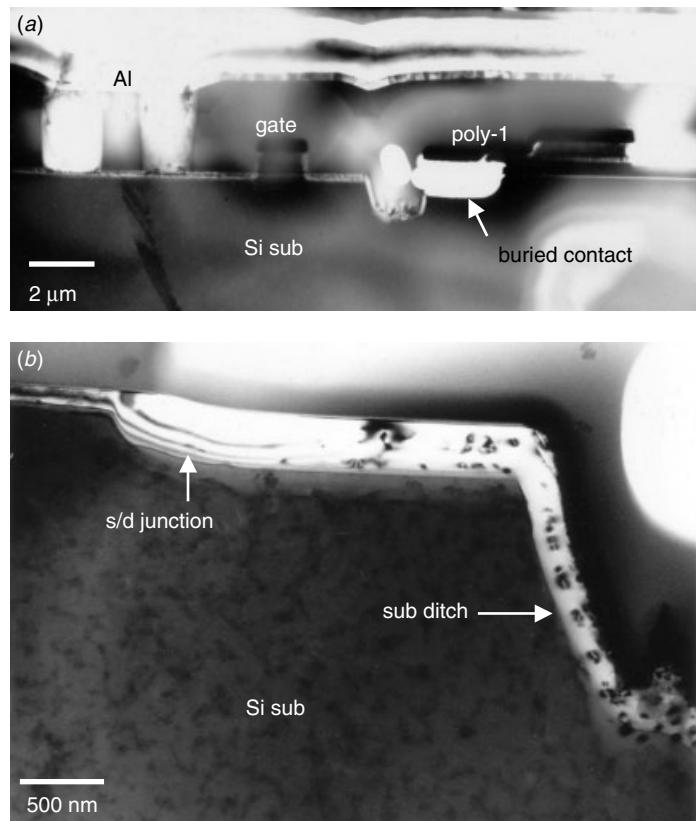


Figure 12.2 A classical 64K SRAM using 1.5 μm polycide technology. (a) Junction delineation reveals the *s/d* diffusion and polycide buried contact diffusion. (b) Close-up at the substrate ditch near the buried contact shows the substrate implant defects within the diffusion areas.

- Since the etching process was designed to remove polysilicon instead of gate oxide, the speed of attack on the Si substrate in this overlapped mismatch area was much faster than the oxide itself. As a result an irregular ditch was generated next to buried contacts in the substrate.
- The subsequent ion implantation and annealing will generate a junction layer at the ditch's base in a way similar to that of normal active junction areas.

Figure 12.2 shows that the delineated junction within the ditch is no different than that in the *s/d* active areas. Even the residual dislocation loops at the Rp range are identical to those in the normal *s/d* active areas. The important question is whether the overetched substrate ditch has any impact on the electrical connection between the *s/d* active area and the buried contact. Figure 12.3 shows a cross section of the area before and after the junction delineation. The substrate's diffusion/junction apparently runs from under the buried contact to the *s/d* active area of the transistor. This confirms that the overetch ditch, although unattractive in appearance and irregular in shape and depth, does not create any problem and thus can be ignored. In all of the subsequent

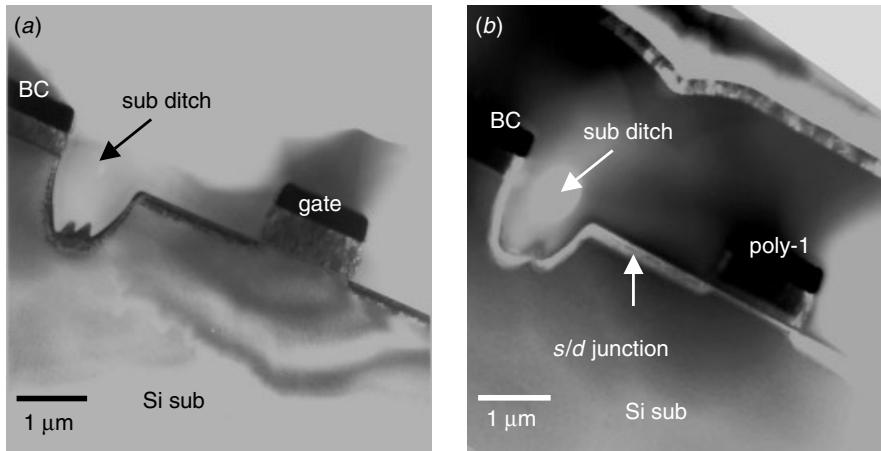


Figure 12.3 The transistor, its neighboring buried contact, and the associated ditch. (a) Before and (b) after junction delineation. The analysis confirm that the buried contact and transistor s/d active area are connected by substrate diffusion junction and not interrupted by the etching ditch.

examples, this buried contact ditch will reappear again, so it will not be discussed further unless necessary.

Another important aspect of the buried contact is its formation technology. As we mentioned above, the buried contacts are formed before ion implantation. There is no dopant in the substrate under the buried contact because it was blocked by the buried contact itself. To form a junction under buried contact in the Si substrate, the dopant must be driven in through the polysilicon during high-temperature annealing. In the annealing process, any interface native oxide between the polysilicon and Si substrate in the buried contact area will rupture and disintegrate. As the interface oxide ruptures, an epitaxy occurs between the polysilicon and the Si substrate. Figure 12.4 shows an

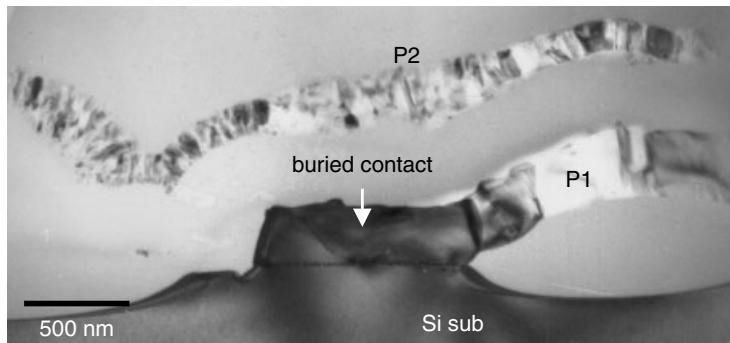


Figure 12.4 Close-up of the buried contact showing a large epitaxial grain in poly-layer. In-situ doped or implanted polysilicon was annealed to drive in and allow dopant to diffuse into the Si substrate to form the junction. The poly-grain tends to form an epitaxy with the Si substrate in the a high-temperature drive-in, as shown here.

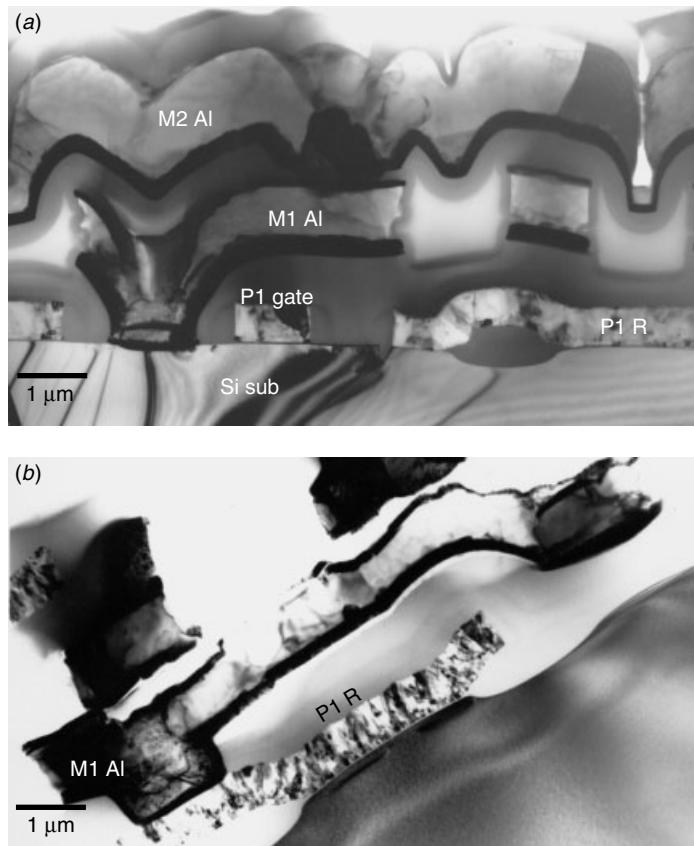


Figure 12.5 256K SRAM device with the polysilicon resistor load. As in the previous example, the process uses polysilicon poly 1 as the transistor gate and resistor load (P1 R). This simplifies the process but greatly increases the bit-cell area. Two metal layers were used for the interconnection, metal 1 as the local contact and metal 2 as the bit line.

epitaxy of a buried contact. The observed epitaxy is due to an interface oxide rupture and complete dopant diffusion, and this suggests that a healthy contact has formed. When no epitaxy is observed, the buried contact may still electrically function and work perfectly well. The mechanism behind the epitaxy in the buried contact of polysilicon to the Si substrate was discussed in Chapter 8.

Figure 12.5 shows another early SRAM using one polysilicon and two Al metal layers. The only polysilicon layer acts as a gate electrode as well as a resistor load. Al metal 1 is used for local contacts and interconnections and Al metal 2 is used for the bit line. The single-poly system was quickly abandoned because of a difficulty with dopant diffusion and resistivity control. It was replaced by double polysilicon layers in which poly-1 is the gate electrode and poly-2 is the load resistor. Figure 12.6 shows an example of a double-polysilicon and double-metal (2P2M) SRAM process. In the figure poly-1 is the gate electrode and buried contact and poly-2 is the polyload resistor layer. This early process utilized uniform and homogeneous doping within poly-1 and poly-2 without changing the doping concentration within the poly layer.

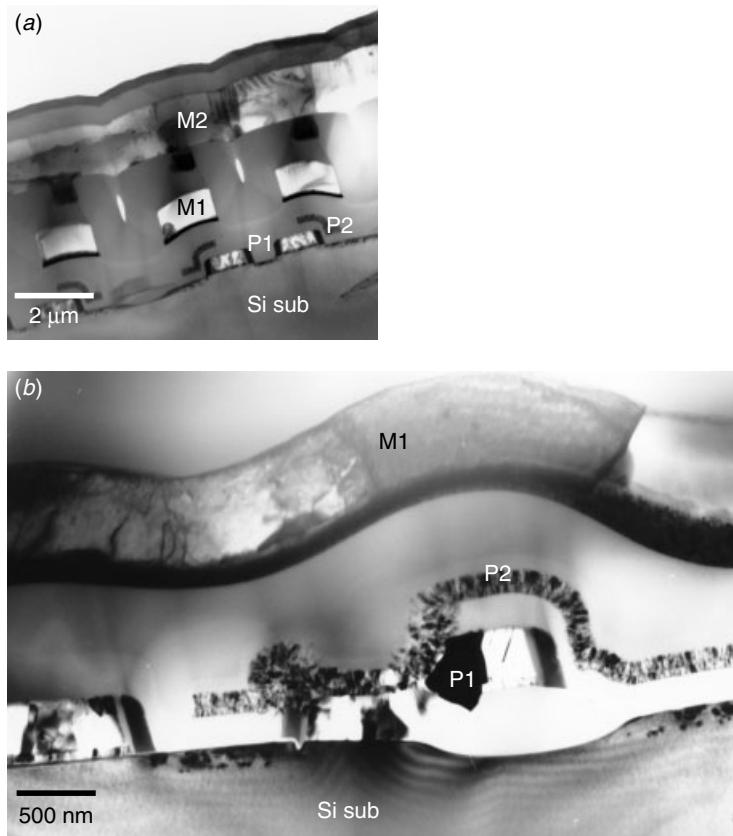


Figure 12.6 256K SRAM device with 2P2M. Poly 1 is used as the gate and buried contact, poly-2 as the load resistor, metal 1 as the bit line, and metal 2 as the row group select line.

One way to quickly tell the doping concentration of a polysilicon layer is by its polycrystalline grain size. The polysilicon grain size is roughly proportional to the doping concentration. The higher the doping concentration, the larger will be the grain size, and vice versa. This criterion is applicable disregard of the doping species, process deposition condition, and technology. It is an easy way also to tell (from the doping concentration) the resistivity of the polysilicon film, and it is particularly useful in relating and comparing the different polysilicon layers within the same device. The example shown in Fig. 12.6 is a good way to test this criterion. As seen in Fig. 12.6(b), poly-1 shows large grain sizes range from 200 nm to 500 nm. Resistivity is thus low, so they are suitable as the gate electrode. Poly-2, on the other hand, shows extremely fine grains, and thus is used for the high-resistivity load resistor. The same criteria will be applied over and over again in the SRAM examples that will be covered later in this chapter. The reader is encouraged to test these criteria using the examples presented.

As the device process technology entered the mega-bit era, the process technology and design concept continued to evolve. The first generation 1M SRAM thus looks very similar to the advanced 256K version. Figure 12.7 shows a 1M SRAM that looks just like the one in Fig. 12.6, the 256 K SRAM. Both use a nitride spacer LDD structure.

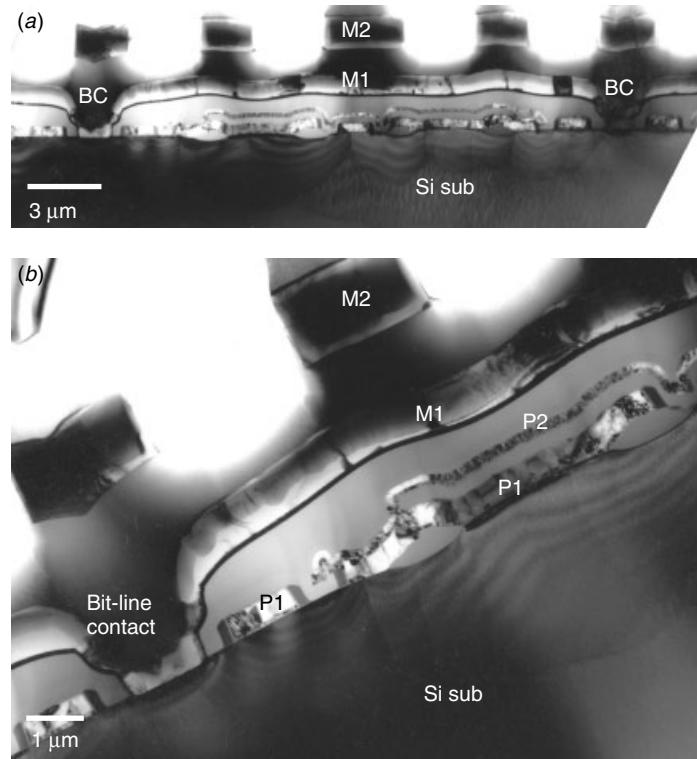


Figure 12.7 1M SRAM device with 2P2M. Poly 1 is used as the gate and buried contact, poly 2 as the load resistor, metal 1 as the bit line, and metal 2 as the row group select line.

Both use two polysilicon layers and two Al metallization layers. Both use poly-1 as the gate and buried contact and poly-2 as the resistor load. Finally, both use metal-1 as bit line and metal-2 as row group select line. The only apparent difference is the appearance of poly-2. Figure 12.8 shows a close-up at the poly-2 to poly-1 contact area. Poly-2 as the buried contact requires low resistivity (high doping concentration and thus large grain size) at the contact area. The large grain size Poly-2 is observed at the contact areas, as seen in the figure. On the other hand, poly-2 as a resistor requires high resistivity (low doping concentration and thus small and fine grain size) at the sheet resistor area, and thus a fine grain size is observed at the polyfilm sheet areas, also seen in Fig. 12.8. An extra masked ion implantation can accomplish such a task for two different doping concentrations on the same polyfilm. The process also eliminates the requirement for additional polyfilm. A comparison of Figs. 12.6(b) and 12.8 shows that the improvement is in the precise control over contact resistivity between poly-1 and poly-2. The former process as in Fig. 12.6(b) had an interface whose quality was difficult to control between the highly doped poly-1 and the lightly doped poly-2. The interface's native oxide tended to be inhomogeneous and sensitive to temperature, process ambient, and to vary locally in the same process lot even on the same wafer. In the improved version seen as in Fig. 12.8, the poly-1 and poly-2 in the buried contact area have similarly high doping concentrations, and thus the contact interface quality

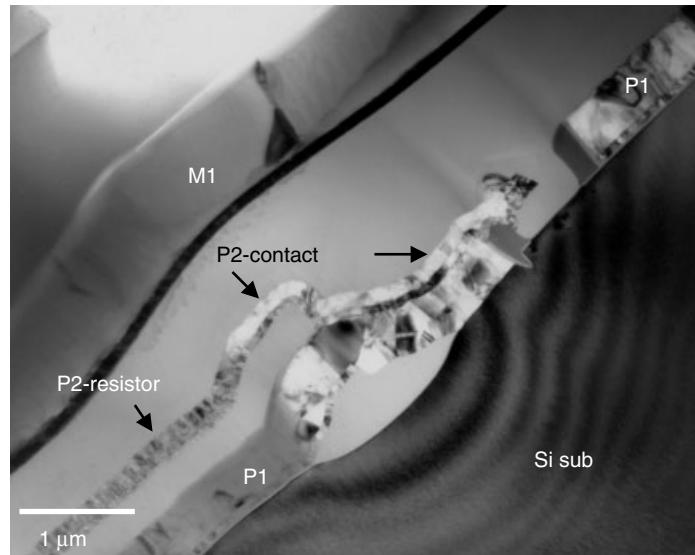


Figure 12.8 Close-up at the gate with an LDD structure, a buried contact, LOCOS, and a poly-Si load resistor. Poly 2 is used as the buried contact, and the resistor shows two distinctive areas of doping with apparently different grain sizes in the same poly layer, as indicated.

is consistent and process repeatability is much better. In fact poly-2 can be divided into three areas as illustrated in Fig. 12.9. The contact area is where poly-2 makes contact with poly-1, as seen in Fig. 12.9(a), arrow-A. The resistor area is where poly-2 acts as the resistor load, as seen in Fig. 12.9(a) and (b), arrow-B. The conductor area is where poly-2 used as local interconnection, as seen in Fig. 12.9(a) and (b), arrow-C.

A subtle, and interesting, detail is that both poly-1 and poly-2 are composite poly-films. As seen in Fig. 12.10, a very thin polyfilm laminated with a thick polyfilm is used in both the poly-1 and poly-2. A composite polyfilm is often used when precise control is needed over a dopant distribution along the film's thickness. As a gate material polysilicon poly-1 often encounters film sheet resistivity, and this is an issue because speed is one of the most critical parameters for SRAM products. The transistor gate's resistivity must be low, and this requirement is more critical here than in the DRAM device. W-polycide is often used to replace doped polysilicon for it can reduce the gate material's sheet resistivity. Figure 12.11 shows an SRAM product with W-polycide as the poly-1 gate and thin polysilicon poly-2 as the load resistor. The W-polycide thin film, as was discussed previously in the Chapter 8, requires a thin polysilicon layer over WSi_x to reduce the step coverage voiding and thus the sheet resistivity issue. A close look at the device shown in Fig. 12.11 reveals such a thin polysilicon layer on top of WSi_x ; this area is indicated in Fig. 12.12 by small arrows. The thin polysilicon on top of the gate and interconnection polycide should not be confused with the poly-2 polysilicon layer, since both of them are of similar thickness and grain size. In fact Fig. 12.12(a) and (b) reveals the confusing structure as on top of the poly-1 WSi_x layer there is a thin polysilicon layer. This polysilicon layer covers WSi_x and branches up vertically at a certain point to form the poly-2 resistor from that point. To understand this appearance, we need to understand the processes.

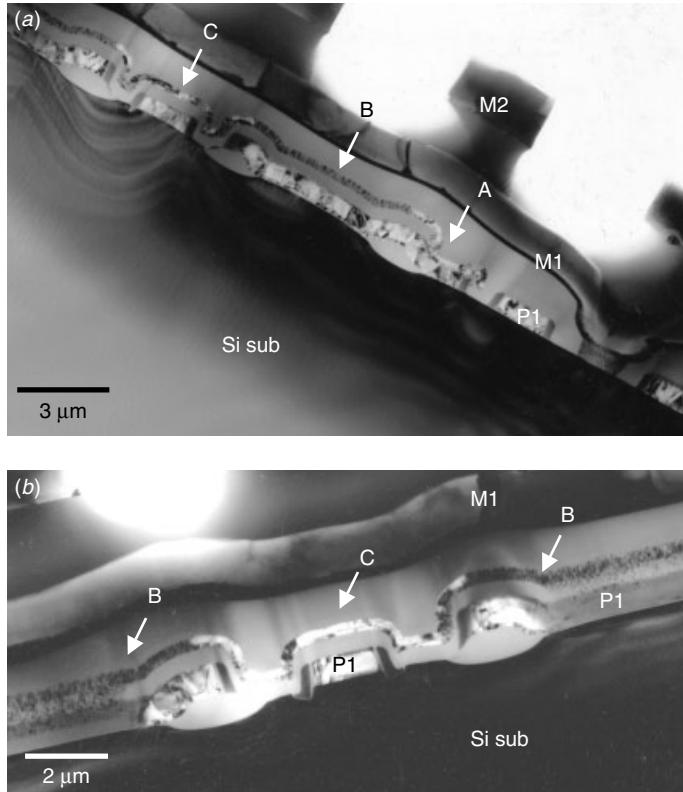


Figure 12.9 Cross section of poly 2 showing distinctively different areas. (a) Arrow A: contact to poly 1, arrow B: polyfilm resistor, arrow C: conductor line. (b) Close-up of poly 2 where there is a sharp conductor-to-resistor transition.

After poly-1 W-polycide process is completed, an interlayer-dielectric (ILD) covers the poly-1. Lithography and etching follow to open up the contact windows and re-expose the poly-1 in the contact window areas. As ILD is removed selectively in the contact windows, WSi_x acts as the etching stopper. The thin polysilicon film on WSi_x is also removed by the etching process, leaving a bared WSi_x in the contact windows (while in other areas WSi_x remains covered by the thin polysilicon film and ILD oxide). This procedure is followed by poly-2 deposition into the contact windows. The newly deposited poly-2 look similar to the original poly-1 polysilicon covering layer, and the structure appears as seen in Fig. 12.12. Interestingly, as Fig. 12.12 shows, the poly-2 covers not only the poly-1 in the buried contact but also the substrate ditch next to the buried contact, and it even extends into the other side of the ditch and up into the ILD oxide. Although there is no direct effect on the process yield and reliability, such poly-2 extrusion does pose an aesthetic issue and so is avoided by most SRAM manufacturers.

Figure 12.13 shows another SRAM using two poly and two metal process and poly-2 as the polysilicon load resistor. Dual poly-2 doping is, again, employed to make poly-2 partly interconnection and partly resistor. Two distinctively different grain sizes within poly-2 is observed clearly.

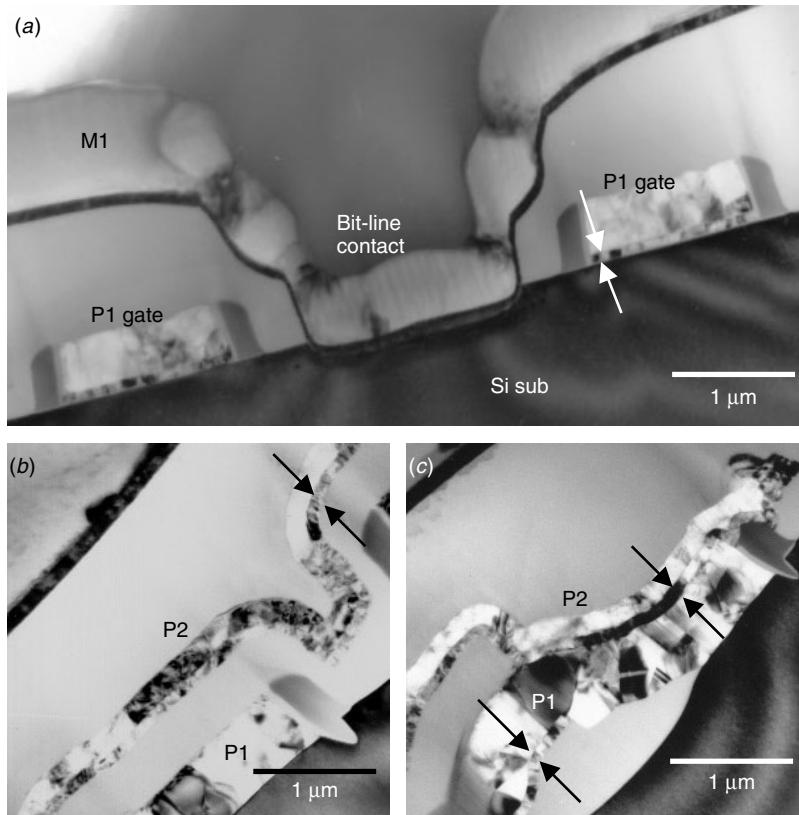


Figure 12.10 Closer up poly 1 and poly 2 appear to be composed of two poly layers. (a) Poly 1 used as a gate shows a thin polyfilm laminated with a thick polyfilm. (b, c) A similar laminated structure also observed in poly 2.

12.3 TFT LOAD SRAM

As device sizes have scaled down, the soft error rate (SER) has became the most serious reliability problem for SRAM. The problem is particularly a concern in below the 1 μm process technology generation. Experimental results from Kinugawa and Kakumu (1993) show that the SER failure rate has grown exponentially from 10 up to 100 as technology has evolved from 1 μm down to 0.1 μm . The high node voltage, V_H , of a cell has been accessed to drop from V_{DD} to that of the threshold voltage of the transfer gate transistor, V_{TT} . If the current supply from V_{DD} through the load devices is sufficient, V_H can be charged back to V_{DD} . However, where the resistivity of the polysilicon load resistor is very large, V_H cannot be charged sufficiently within the cycle time. This problem does not occur in the TFT-load bit-cell of SRAM, but in the polysilicon resistor load it has become insurmountable. Prince (1995) has summarized the problem in her book as due to the high node level in the memory cell needed to be traded off with the low stand-by current requirement:

If polysilicon resistors have high resistivity then the standby power will be reduced. The standby current requirement therefore sets a minimum value to the resistance. The

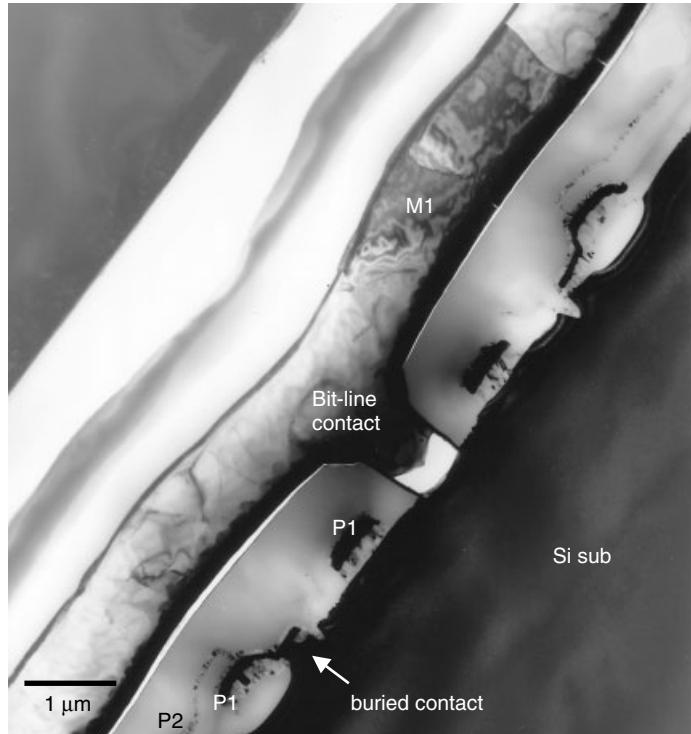


Figure 12.11 A different version of 1M SRAM using poly 2 as the poly resistor load, polycide poly 1 as the gate and buried contact, and Al metal 1 as the bit line and bit-line contact. Notice that the bit-line contact is not properly filled with Al, so the contact is electrically open.

minimum value increases with each density generation since the standby current requirement stays the same but the number of resistors increase four times. On the other hand, the resistance must be small enough to provide enough current to the storage node so that the stored charge does not leak away because of the sub-threshold in the storage transistors. This sets a maximum value on the resistance. Since the sub-threshold leakage of the storage transistors increases with decreasing threshold voltage and the threshold voltage drops as the geometry decreases, this maximum value of the resistance decreases with density. The gap between the minimum and maximum value of the resistance, therefore, decreases as the density of the chips increases making the process difficult to control.

The polysilicon load resistor that had served the SRAM commodity markets well for four generations (i.e., 16K, 64K, 256K, and 1M) is no longer adequate for both the low power and high speed of the 4 Mb generation. The low power and wide noise margins of the six-transistor cell are needed, and TFT-load cell is replacing the polyload at about this time. The thin film transistor stacked PMOS load transistor cell was a natural next phase for the manufacturers who used the polyload resistor cell and wanted the stability of full CMOS in the submicron geometries. Although the process complexity has increased, the two PMOS load transistors formed in the thin film polysilicon layers above the four NMOS transistors do not occupy more cell area and thus are justifiable (Prince 1995). More advanced TFT structures have been

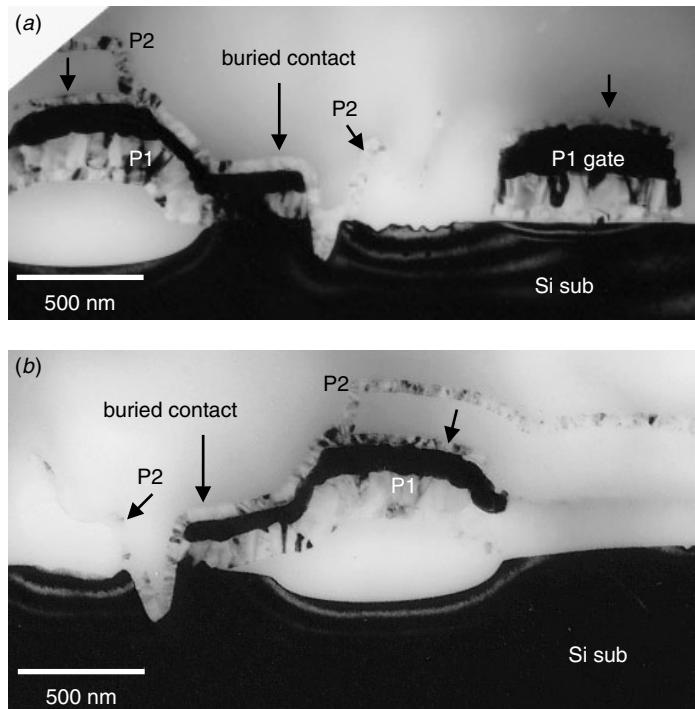


Figure 12.12 Closer up the poly 1, poly 2 appears to be formed by W-polycide with a thin poly on top (as indicated by small arrows). Poly 2 is an ultra thin polysilicon layer (about 50 nm thick). The buried contact is formed by both poly 1 and poly 2. A substrate ditch formed next to the buried contact is filled with poly 2.

proposed for TFT SRAM devices (Yamanaka et al. 1988; Ando et al. 1989; Yamanaka et al. 1990; Hayakawa et al. 1990; Ikeda et al. 1990; Batra et al. 1993; Naiki et al. 1993; Ikeda et al. 1993; Pfeister et al. 1994).

Figure 12.14 shows a typical TFT-load (4 NMOS + 2 TFT PMOS) cell. In this 0.5 μm technology device, 4P2M (4 polysilicon and 2 metallization) process was used. Poly-1 forms the gate and word line. Poly-2 forms the bit line and bit-line contacts. Poly-3 and poly-4 together comprise the thin film transistor (TFT). In contrast to the conventional bulk transistor, where the gate is always at the top of channel and active source/drain areas, the TFT transistor does not always follow this rule. Figure 12.15 shows an enlarged TFT part with each of the components indicated of the thin film transistor. Poly-3 is the gate electrode. Poly-4 forms the active source and drain and channel region in between. The thin oxide layer in between P3 and P4 is the gate dielectric. Notice that P3 and P4 (and the thin dielectric layer in between) are very thin compared to the P1 and P2. This is essentially why they are called thin film transistors. In this case P3 is 20 nm, P4 is 25 nm, and the oxide in between is 40 nm in thickness. P3 as the gate electrode controls the transistor channel on top of it (P4), and this is different from the conventional bulk transistor. Figures 12.16 and 12.17 show a similar yet slightly different TFT structure. In this case the TFT is lying directly on the poly-2 and not at its edge. The two slightly different TFTs are not

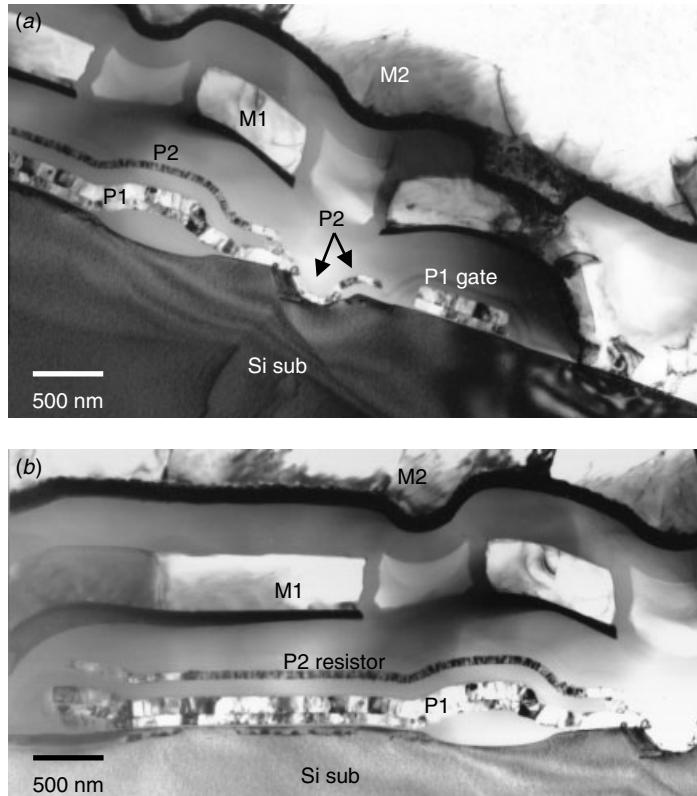


Figure 12.13 Another SRAM device using the laminated poly 1 as the gate, buried contact, and local interconnection, and the thin poly 2 layer as the resistor load. Notice that the buried contact is similar to but different from the previous case. Poly 2 also has filled into the substrate ditch, but the ditch is much wider compared to the previous case.

intentional. Figure 12.18 shows a lower magnification of the device's cross section along the word-line direction. Two TFTs are shown here compared with one seen in Fig. 12.15 and one in Fig. 12.17. The two TFTs coexist in an SRAM because there is no lithographic design purpose for P3 and P4 to be aligned accurately with P2. P3 and P4 together form the thin film transistor, and they must align exactly to each other in order to be functional. In addition they must relatively align with P1 so as to make contact with P1. P2, whose function is as bit-line contacts and local interconnection, on the other hand, does not have to line up with P3 and P4 in as critical a manner. As a result both P2 and TFT (i.e., P3 + P4) must be aligned with P1 but not with each other. That is why we have seen in the different cross sections TFT both at the P2 film edge and right on the P2 film. This bit of design freedom comes as a relief to circuit designers and layout engineers who can now treat TFT in a more versatile fashion. The size and shape of the TFT can be manipulated without worrying about location and the topography below it, as seen in Fig. 12.18. The relationship between P3 and P4 remains important because it determines the performance of the TFT device. P3 and P4 are essentially two long (and thin) polysilicon stripes crossing each other at

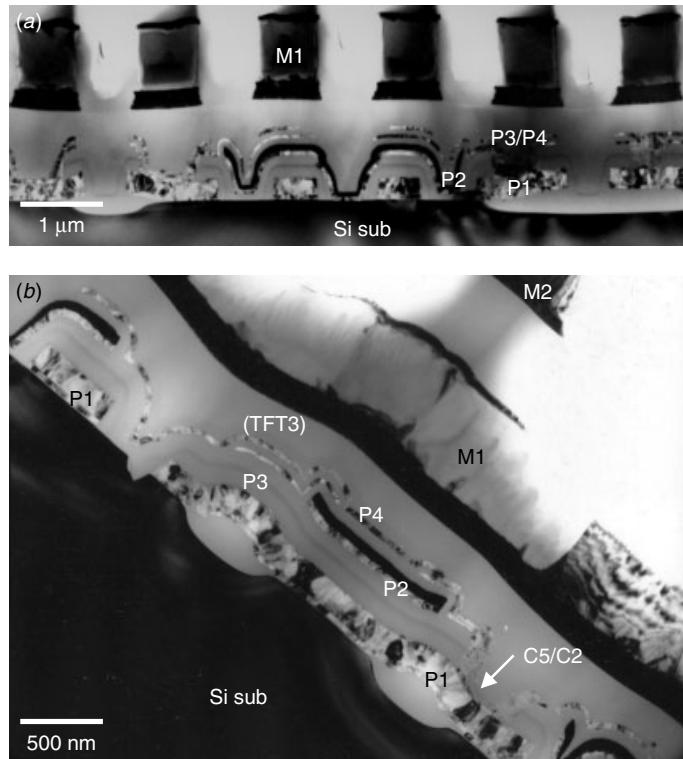


Figure 12.14 An early TFT-load SRAM. TEM cross sections along (a) the bit-line and (b) the word-line directions. The 4P2M process was used in this $0.5\text{ }\mu\text{m}$ technology device. The bracket TFT3 and arrow C5/C2 are explained in Fig. 12.20.

the point where the device gate structure forms. At this small rectangular region with two crossing thin polysilicon layers sandwich a thin oxide layer. A channel must be formed in this region on P4 side where there is no doping, just as in the bulk transistor channel where the conduction in the channel region is determined by the potential applied in the control gate conductor (in this case P3). Figure 12.19 shows a cross section along the bulk transistor bit-line direction and a close-up at the TFT shows long P3 running on top of P2. Unlike both Figs. 12.15(b) and 12.17(b), Fig. 12.19(b) shows only a short P4 on top of P3. In combining 12.17(b) with 12.19(b), it is possible to obtain a full picture of how the TFT is formed and how it is connected to the bulk transistors.

However, it is impossible to find this interconnection relationship within an SRAM cell using one or two cross-sectional views. The SRAM cell interconnection lies in the horizontal two-dimensional interconnection by multiple layers of polysilicons and metals that cannot be revealed directly in cross section. Plan view deconstruction analysis through several interconnection layers and circuit overlays and tracings are necessary to accomplish this. A cross-sectional view can be confusing, if not misleading, if one tries to determine how each transistor is connected to its mates in the same cell and functions as a bit cell. Nevertheless, cross section images can still provide

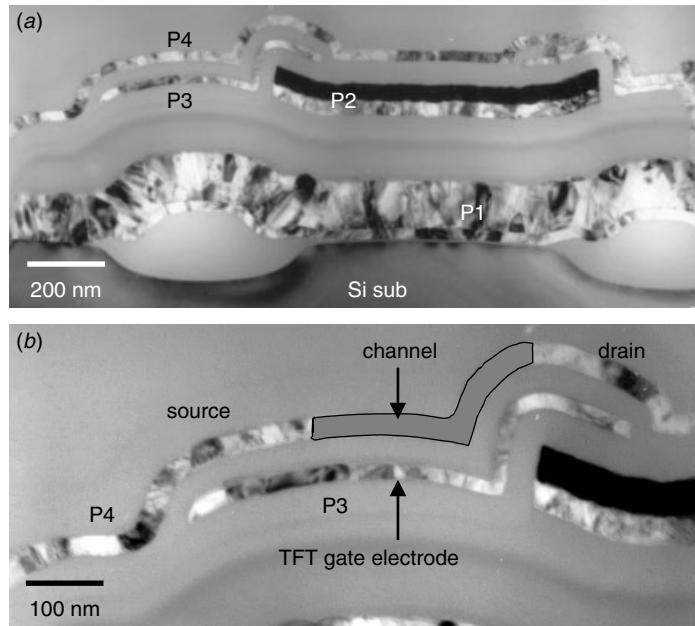


Figure 12.15 Thin film transistor (TFT) formed by using poly 3 and poly 4 as indicated. (a) TEM cross section image, (b) enlarged TFT part showing the TFT construction and composition. The channel, source, and drain are marked as seen in P4. P3 is used as the gate electrode.

vital information on how designers have ingeniously put together all the polysilicons and metals to accomplish an intricate network.

By replacing the two load components in Fig. 12.1, we can transform an SRAM bit cell into a six-transistor cell as shown in Fig. 12.20. Among the six transistors, T3 and T4 are the PMOS TFT load transistors and T1, T2, T5, and T6 are the access and driver NMOS transistors. While T1, T2, T3, and T4 form the flip-flop, T3 and T4 at each of their sources/drain are linked to Vdd (or Vcc) and T1 and T2 at each of their sources/drain are linked to Vss (or ground). T5 and T6 gates are linked to word line, and one of their source/drain is linked to the bit line. Although we are not able to see the whole interconnection, we can make out the bits and pieces by studying Figs. 12.14, 12.16, 12.18, and 12.20. Figure 12.20 provides the useful schematics for the names and each one of the contacts, gates, sources/drains, and their interconnections. The following summarizes the observations:

- Figure 12.14, TFT3 at one side of source/drain makes contact to P1. The C5 and C2 stacked contacts are marked in the schematic of Fig. 12.20. The stacked contact is poly-4 to poly-1 to active substrate area buried contacts.
- Figure 12.16 TFT4 at one side of the source/drain makes contact to P1. C6 and C1 stacked contacts are marked in the schematic of Fig. 12.20. The stacked contact is poly-4 to poly-1 to active substrate area buried contacts. Also shown in Fig. 12.16 is the other side of the TFT4 source/drain that makes contact to P1 at the other side. This is C7 to the Vcc (or Vdd) power supply.

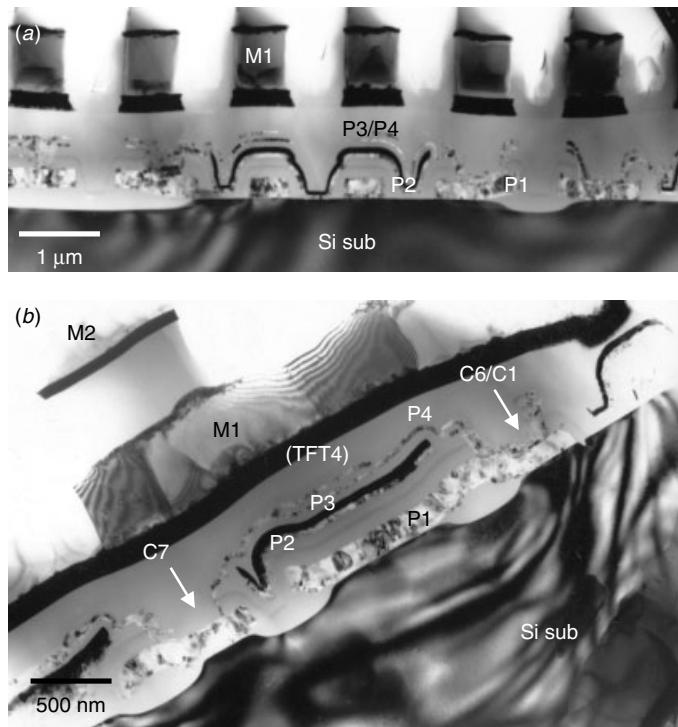


Figure 12.16 Another cross sections of the TFT-load SRAM. The TEM Cross sections are along (a) the bit-line and (b) the word-line directions. The 4P2M process was used in this 0.5 μm technology device. All components are the same as in Fig. 12.15 expect now the TFT is formed directly on top of P2 instead of at the corner of P2. The bracket (TFT4) and arrows C7 and C6/C1 are explained in Fig. 12.20.

- Figure 12.18 TFT3 and TFT4 share a center contact, which is again C7 to Vdd for the power supply.
- Figure 12.20 shows some intricate links: Figure 12.20(a) shows how T5 and T6 gates align together with a center contact (bit-line contact) marked as C4/C3. Figure 12.20(a) shows the TFT3 gate (P3) contact to poly-1 as the T1 polygate, which corresponds to C1 in the schematic. Figure 12.20(a) shows the TFT4 gate (P3) contact to poly-1 as T2 polygate, which corresponds to C2 in the schematic. Figure 12.20(b) shows the TFT4 poly-4 (source and drain) link at one side to poly-1 and then to the substrate. This is also observed in Fig. 12.16, and is the C1 to C6 to substrate buried contact. Figure 12.20(b) shows the TFT4 poly-4 (source and drain) link at the other side to poly-1 and this is marked as C7. The same C7 is shared by TFT3, and this C7 is to the Vdd power supply shared by all TFTs.

Not all of the TFT-load SRAM share the same contacts and interconnection structures. The roles of P2, P3, P4, and even M1 and M2 are intermixed and interchangeable and often without commonly accepted criteria. Rather, illustrated here are the basic techniques used to trace and find the intricate links among transistors.

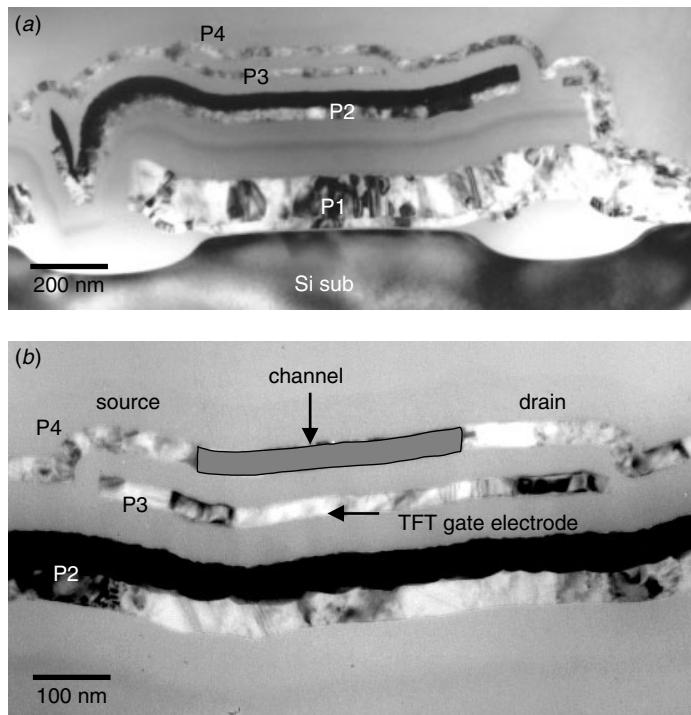


Figure 12.17 Thin film transistor (TFT) formed by poly 3 and poly 4, as indicated. (a) TEM cross section, (b) enlarged TFT part showing the TFT construction and composition. The channel, source, and drain are marked as seen in P4. P3 is used as the gate electrode.

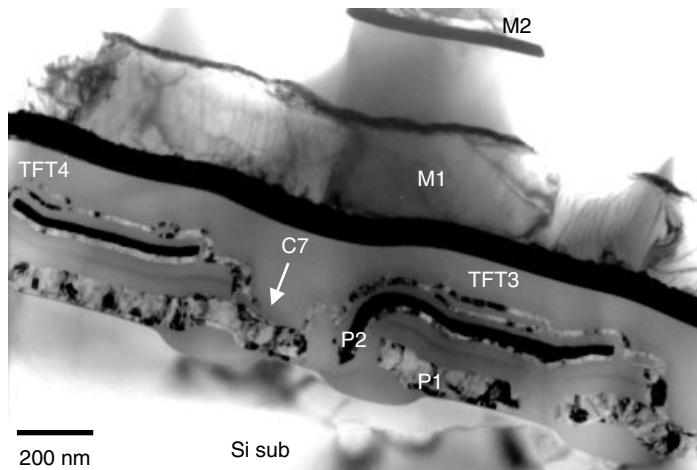


Figure 12.18 Closer up TFTs alone the word-line direction. Notice that there are two TFTs. One is directly on top of P2 (marked as TFT3) and the other is at the edge (marked as TFT4). P3 and P4 must align with each other to form a functional transistor. But they do not align rigidly with P2, as shown here. C7 indicates where the two TFTs share a contact to the Vdd power supply.

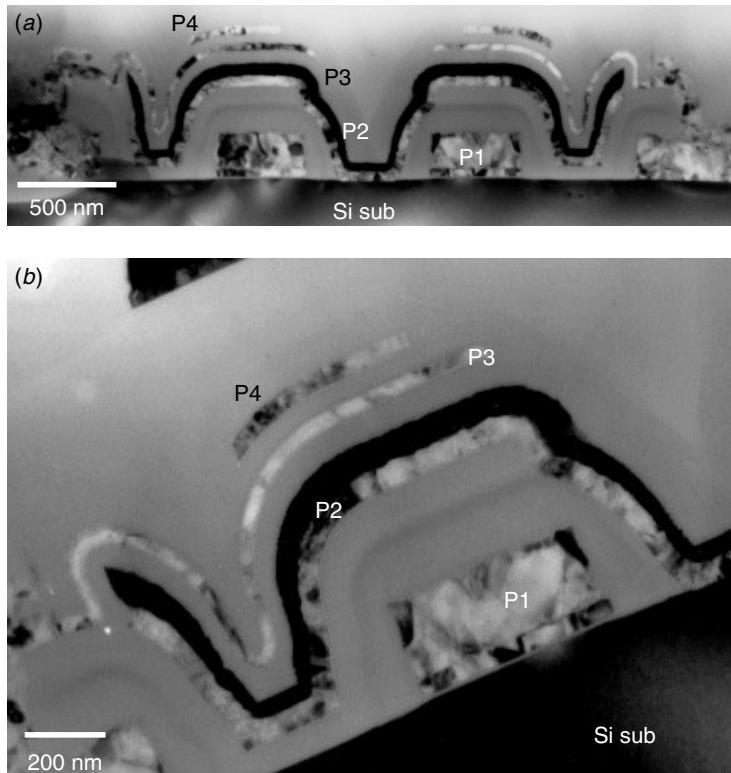


Figure 12.19 Closer up TFT cut along the bit line. P3 runs in parallel with P2 and connects to P1 as seen in (a). P4, on the other hands has only a short section in this figure.

An attempt to simplify the TFT structure and to reduce the number of layers of polysilicon has emerged during this time. Hayakawa et al. (1990) proposed a remarkable scheme that eliminates the TFT gate electrode polysilicon layer. The basic idea is quite simple: The bulk diffusion area acts as the drain of the driver transistor can be the gate electrode of the TFT. The TFT cell structure therefore does not need additional gate electrode layers, permitting a planarized structure as required for high-density SRAM cell. A commercialized version of such a technology was found in Fig. 12.21. The driver's bulk transistor drain is simultaneously used as the gate for TFT, which is formed using poly-2 as source-channel-drain. A large part of the TFT process development and design improvement efforts was concentrated on how to improve the TFT device performance. TFT by nature is not as effective as the bulk transistor, particularly when built on several polysilicon layers where the topography is irregular and uncontrolled. One way being explored to optimize the TFT performance is to build it on a controlled topography. An example is shown in Figs. 12.22 and 12.23. TFT is built by using poly-2 and poly-3 forms tight on top of LOCOS and the structure is highly symmetrical.

With decreasing device geometry, the TFT cell's construction will require more process complexity to meet the desired properties. In general, four polysilicon layers are needed. Poly-1 is used as word line and gate. Poly-2 usually uses a local interconnection

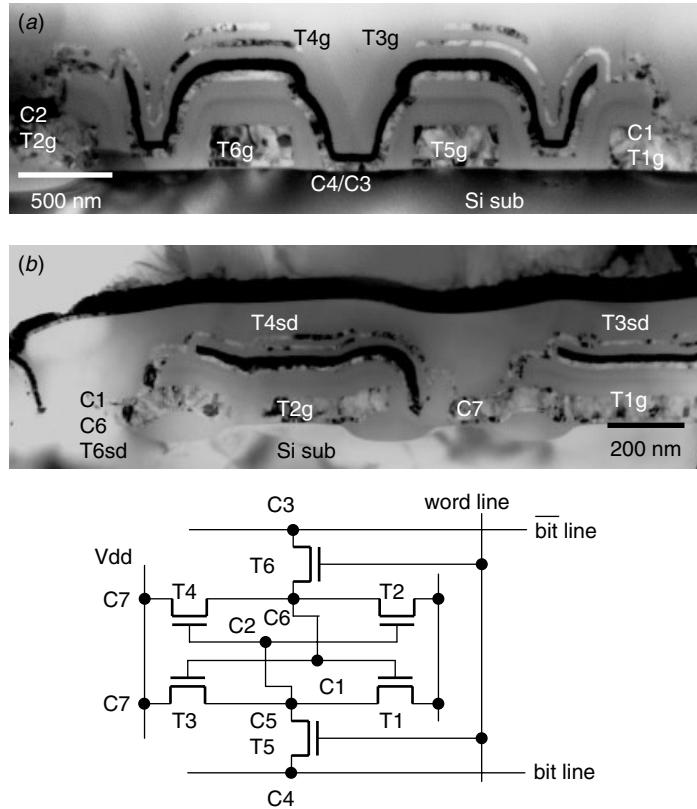


Figure 12.20 Cross section along (a) the bit line and (b) the word line revealing several key components of an SRAM bitcell. Compared to the schematic below, some of the components are observed and marked. Note are gate (*g*) and the source and drain (*sd*).

and/or bit-line contacts. Poly-3 and poly-4 are usually used together to form the TFT structures. Figures 12.24 through 12.26 are examples of such a development. Poly-1 and poly-2 as the word line, gate, and local interconnection require low sheet resistivity and thus the polycides are a reasonable choice. Poly-2 at the same time forms a self-aligned bit-line contact pad and effectively helps reduce the bit-line contact size, as seen in Fig. 12.27. Poly-3 and poly-4 are used to form TFT and thus normal polysilicon films are used. For TFT, the channel polysilicon needs to be thin, and thus usually one of the poly-3 and poly-4 will be very thin and used as the channel polysilicon. In fact most of the SRAM commodities available in the market seem to converge to such a design and process concept, as can be seen in Figs. 12.24 through 12.26.

As the TFT-load SRAM evolved further, the process development of SRAM appeared to verge with logic devices. Several of the new technologies incorporated in SRAM devices were developed independently for the CMOS logic device (Hayden et al. 1991; Subbanna et al. 1993). At this time there is not much new design and processes developed especially for SRAM production. The TFT-load SRAM device seems to have come to its development end.

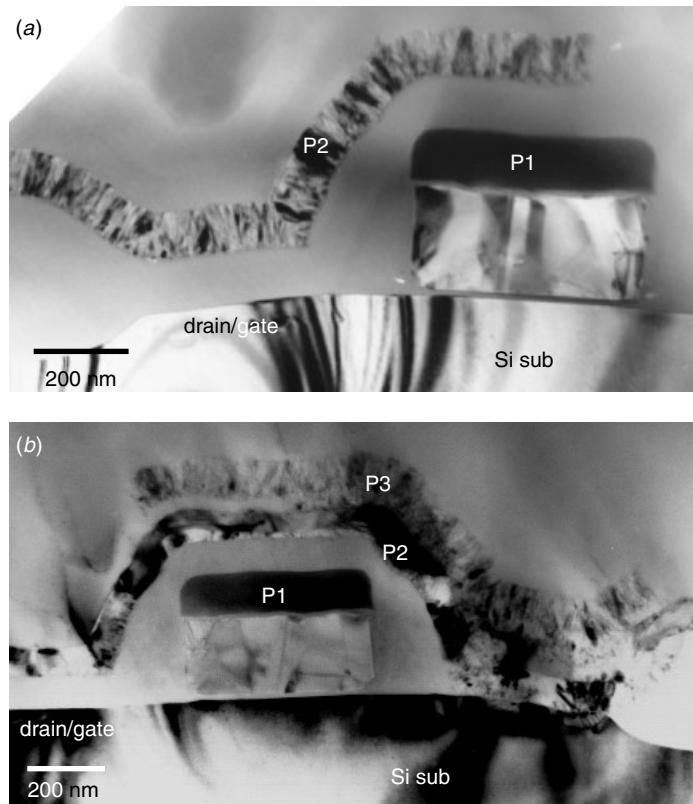


Figure 12.21 Simplified version of TFT load in the SRAM bit cell. The NMOS active substrate is the TFT's gate electrode and reduces the polysilicon layer, as shown.

12.4 BULK TRANSISTOR LOAD SRAM

Clearly, from the examples above the most important trade-off in SRAM cell design lies in the relationship between the cell size and process complexity. The PMOS thin film transistor (TFT) load in SRAM generally calls for highly sophisticated processes. Compared with the TFT load SRAM (4T cell) and bulk PMOS load SRAM (6T cell), the 4T cell seems to occupy much less cell area, although it does add to the process complexity. A more complex process can still produce a less expensive product provided that the increase in yield more than compensates for the increased wafer cost. This is why most major SRAM manufacturers are proceeding to reduce the SRAM cell size by adding the process complexity.

On the other hand, the 6T SRAM cell with six bulk transistors (using bulk PMOS loads rather than TFT PMOS load) has become the choice for on-chip storage in advanced microprocessors and other logic circuits because of its process compatibility with the logic device (Lage et al. 1996). Since it is often desirable to run SRAM and logic processes in the same wafer fab, compatibility between SRAM and logic processes is highly desirable (Watanabe et al. 1983; Verhaar et al. 1990; Hayden et al. 1991; Subbanna et al. 1993; Koike et al. 1994).

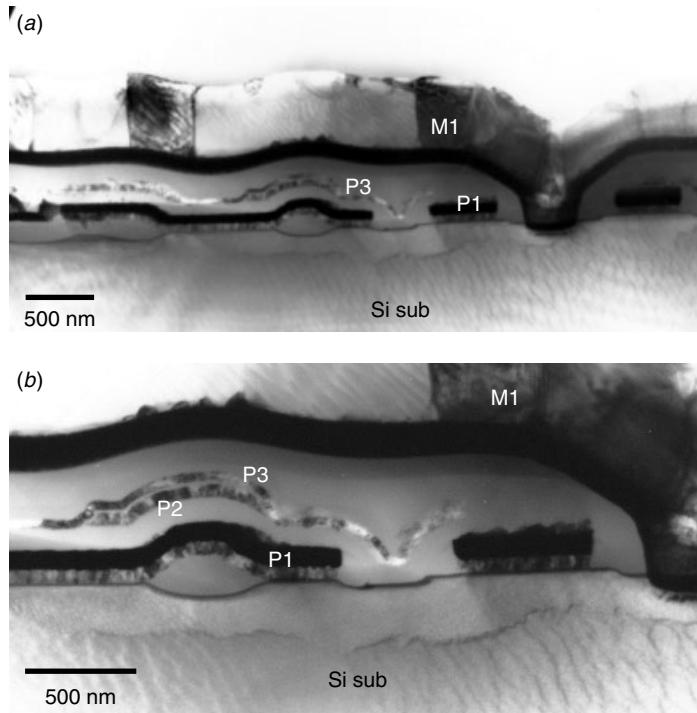


Figure 12.22 4M SRAM using the 3P2M process. Poly 2 and poly 3 together form the TFT load.

A big issue with the 4T cell is the TFT structure. Most TFTs used in manufacturing are rather weak devices, they reduce standby current draw and do very little to improve memory cell stability. The SRAM need to maintain a constant standby current per chip disregards the fact that capacities have increased with succeeding generations along with the threshold voltage of the bit-cell transistors decreased. The lower limit of operation voltage has increased with down scaling, while V_{DD} has decreased because of the reliability and power dissipation issues. For both R-load and TFT PMOS load cells, maintaining enough margin with $V_{DD} = 2.5$ V at the 0.25 μm technology is a challenge (Kinugawa and Kakumu 1993). It has proved to be very difficult to build a mass-production version of TFT with a current large enough to improve the cell's stability. With a basic 4T cell, it is very difficult to scale much below a supply voltage of 2.5 volts. On the other hand, 6T cells enjoy more stability, especially at low voltages. The 6T cell can provide enough margin because of the large current supply from the V_{DD} line to the high node by PMOS. The improved stability at low power supply voltages is the main reason why the 6T cell will become the dominant SRAM cell and replace the 4T cell (Lage et al. 1996).

Another problem with the TFT cell is with the on-state current. As the device was scaled down, the average cycle time has decreased with the down scaling. The shorter the cycle time, the higher must be the on-state current of the load device. It was observed for the 0.25 μm technology that in order to achieve V_H (averaged high node voltage) higher than or equal to V_{DD} , a minimum on-state current of 0.1

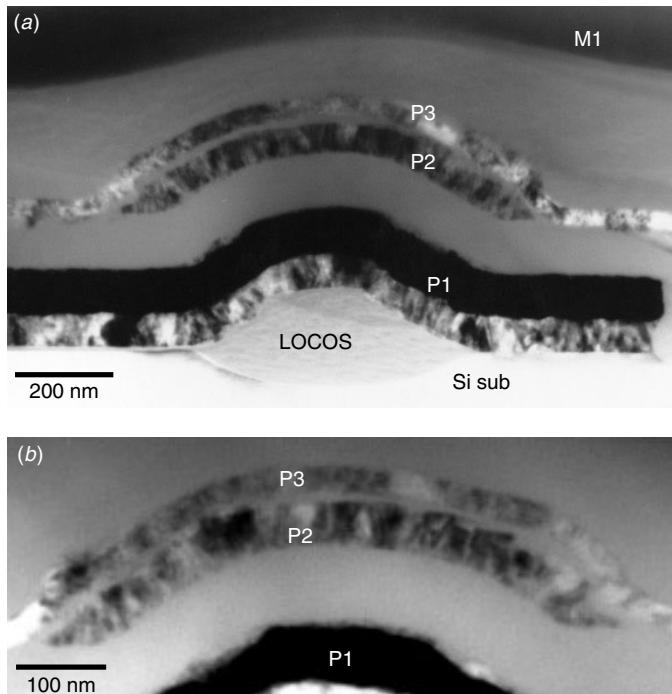


Figure 12.23 Close-up of the TFT cell showing poly 2 as the gate and poly 3 as the source-channel-drain. TFT is made on top of the LOCOS to ensure that the semicircular curve is used and a symmetrical TFT device is created.

μA is needed (Kinugawa and Kakumu 1993). It is very difficult to achieve by TFT technology such a high on-state current with well-controlled uniformity.

The bulk 6T cell SRAM does not show distinctive process differences from the normal logic devices. Figure 12.28 shows two examples of 6T bulk transistor SRAM cell. As can be seen, there is no difference from normal logic devices as far as the process is concerned. The only noticeable change is the highly periodic and repetitiveness of the SRAM areas when compared to the normal logic device areas where there are no periodic patterns seen in a cross section. The three metal layers together form the word lines, bit lines, and local interconnections. W-plug as the contacts and either conventional Al or full Cu metallization (4 metals) is used in the advanced 0.13 μm design rules and process technology, as seen in Fig. 12.28. Notice that each Cu metallization is buffered with a TaN barrier layer and wrapped by SiN layers. The SiN layers are there for damascene process and to prepare for the more advanced IMD low- k dielectrics where SiN would provide the necessary mechanical support and chemical protection for Cu metals. Figures 12.29 and 12.30 show snapshots of the 0.13 μm technology gate structure with conventional and L-shaped LDD spacer.

An analysis of the advanced 6T bulk transistor SRAM cell thus does not provide more insight on the advancement of ULSI process and technology evolution, at least not more than what can be learned from the logic devices themselves. The only advantage is that the high repetitiveness of the SRAM cell ensures a design rule and minimum feature measurement.

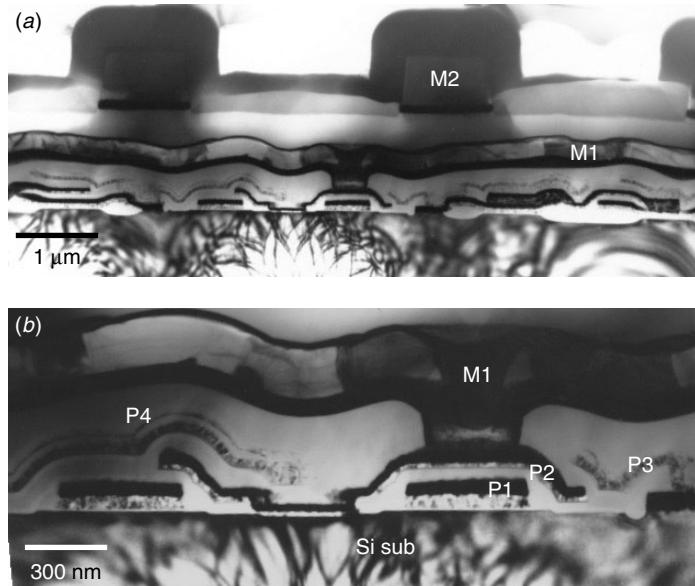


Figure 12.24 4M SRAM using the 4P2M process. Poly 1 is the word line and gate. Poly 2 is used for the local interconnects. Poly 3 and poly 4 together form the TFT load. Metal 1 is bit line and metal 2 for the global interconnects.

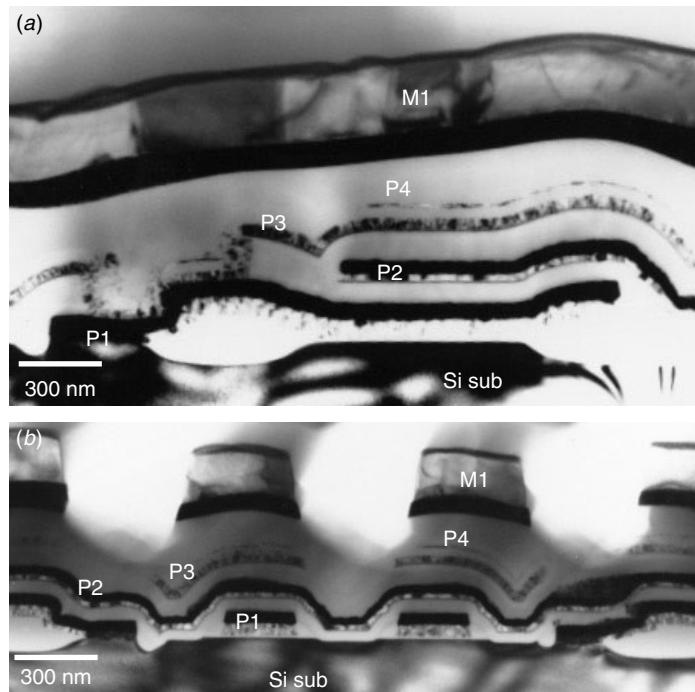


Figure 12.25 Close-up of the TFT cell showing poly 3 as the gate and poly 4 as the channel. Notice that poly 4 is very thin, only about 10 nm.

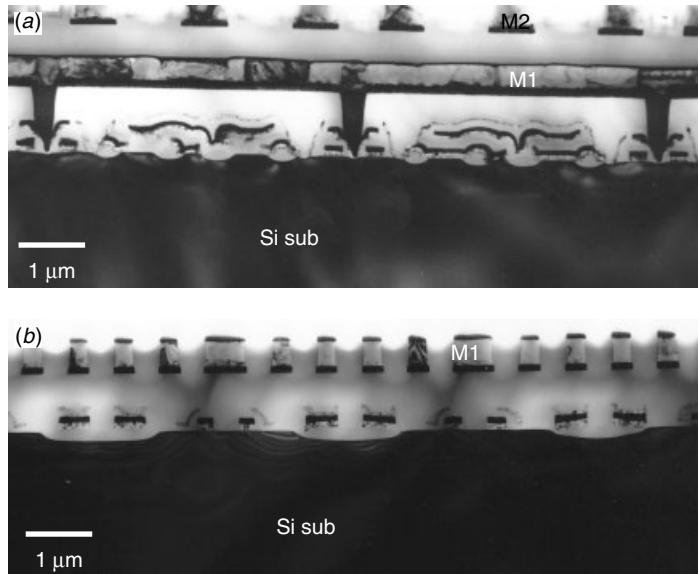


Figure 12.26 Late version of the 16M TFT-load SRAM using $0.25\text{ }\mu\text{m}$ technology, 4P2M process. Notice that poly 2 is used as a contact pad for the metal 1 bit-line contact.

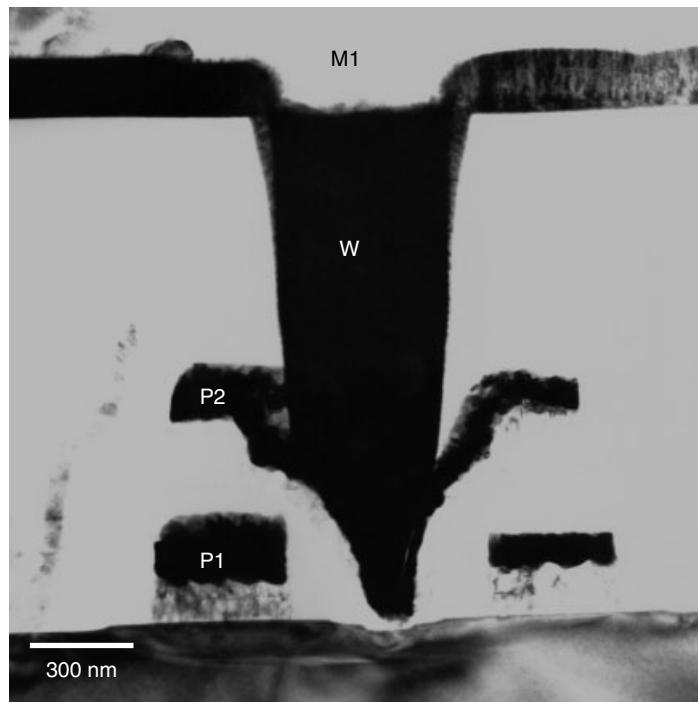


Figure 12.27 Poly 2 is used as the contact pad for the metal 1 bit-line contact. Notice that the contact's opening is formed by the self-aligned poly 2 contact to the substrate. This largely reduces the bit-line contact size, as shown here.

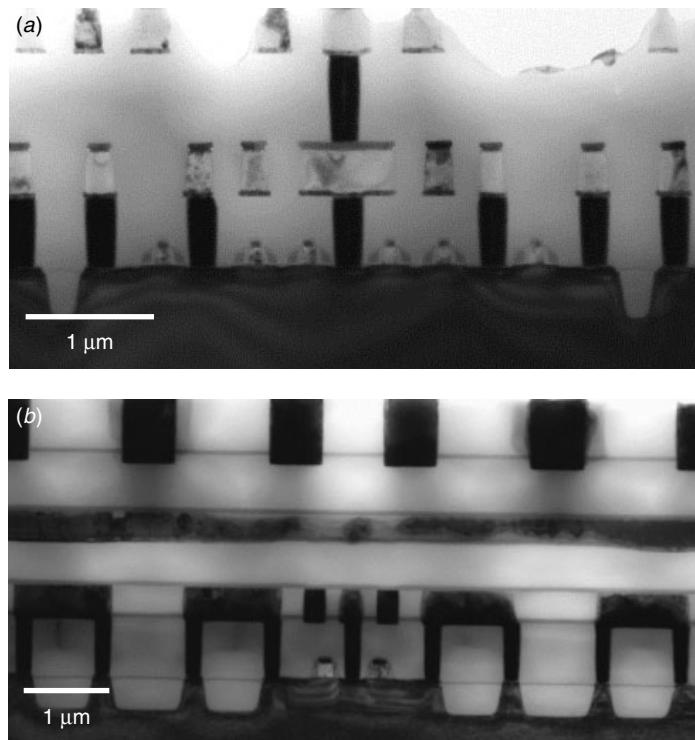


Figure 12.28 Two typical 6T bulk transistor SRAMs. The $0.13\text{ }\mu\text{m}$ design rule with (a) conventional Al metallization interconnects and (b) the full Cu metallization technology. Both use W-plug as the contacts.

12.5 SRAM AS THE NEW TECHNOLOGY DRIVER

As technology has evolved, the DRAM product is no longer always being used as the ULSI process technology driver. As device shrinks in dimension, the power supply decreases accordingly and the device performance and reliability improve. New device failure mechanisms are being observed with the interconnection line density increases. The smaller transistor dimensions have meant a rise in the interconnection lines that connect them, and as ever more transistors are fit on a single chip, eventually metallization resistivity will take over and limit the device's performance. Already this has occurred at about the 0.35 to $0.25\text{ }\mu\text{m}$ technology generation where approximately 100 million transistors are webbed by more than a few kilometers of interconnection lines in total. All are packed into a single silicon substrate no more than a few centimeters square. Revolutionarily new materials and processes are being introduced. As a result a vehicle device structure is needed for testing the new design layouts, process developments, evolving technology, and reliability. This device structure should have the following characteristics:

- Utilize new materials and process features in the same way as a normal logic device without modification.

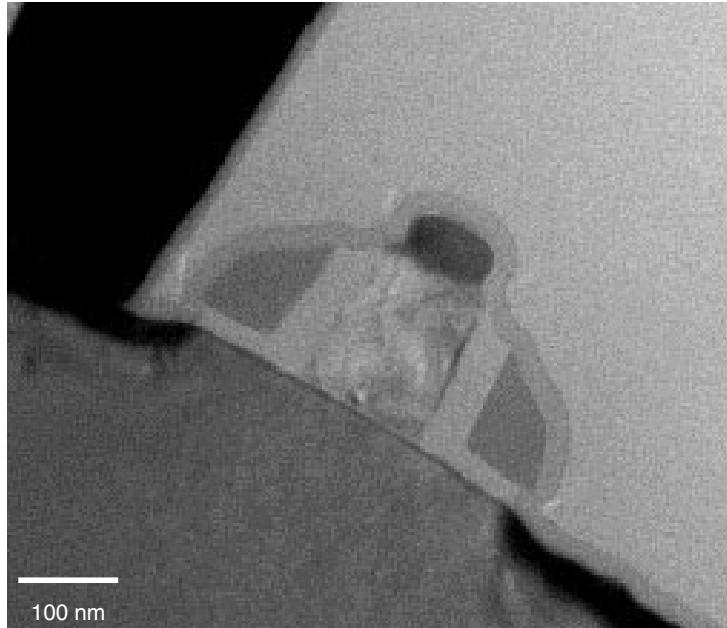


Figure 12.29 Transistor gate structure as corresponding to the device shown in Fig. 12.28(a). Nitride spacer is used in this 130 nm process technology.

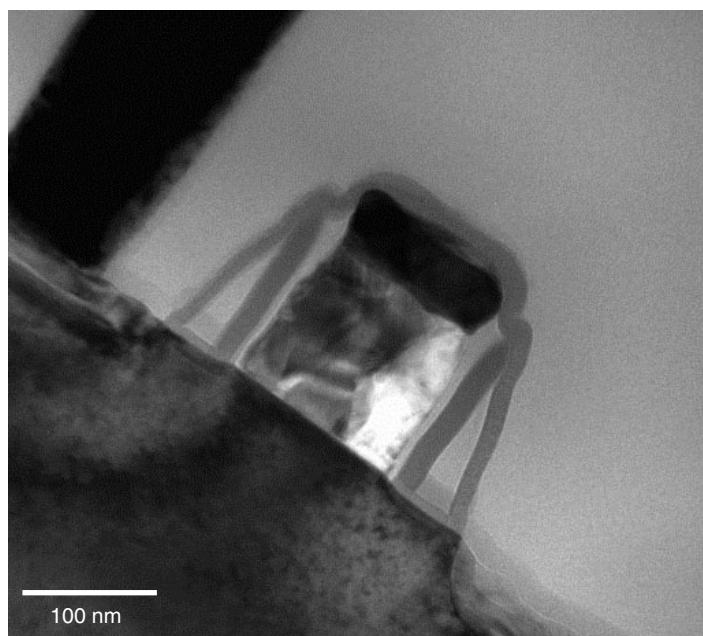


Figure 12.30 Transistor gate structure as corresponding to the device shown in Fig. 12.28(b). L-shape nitride spacer is used in this 130 nm process technology. The basic gate structure of the SRAM transistor is identical to logic device.

- Have repetitive features for easy process debugging and failure isolation.
- Have well known device parameters that can easily point to local and individual performance variations.
- Can be embedded on the wafer alongside the logic device and products.

The 6T SRAM seems to fulfill all of the requirements above. The current wafer FAB foundries even utilize SRAM as one of the standard units within the test wafer. These SRAM sits side by side along with many other test keys for process development and yield evaluation. Some manufacturers use these SRAM areas to evaluate the defect density for a certain technology and as a baseline for comparison among different wafer FABs in different locations. The test wafer SRAMs are also packaged for testing, burn-in, process qualification, and product reliability evaluation. Tremendous process knowledge and reliability information can be extracted and learned from the test wafer SRAMs. SRAM, though not yet a dominant commodity product in the memory market in terms of volume, is replacing DRAM as a technology driver in an unexpected way.

REFERENCES

- M. Ando, T. Okazawa, H. Furuta, M. Ohikawa, J. Monden, N. Kodama, K. Abe, H. Ishihara, and I. Sasaki, *IEEE J. Solid State Cir.*, **24** (6), 1708–1713, 1989.
- S. Batra, R. Maddox, L. Tran, M. Manning, C. Dennison, and P. Fazan, *IEEE Trans. Electron Devices*, **49** (11), 2125–2126, 1993.
- M. Bohr, S. U. Ahmed, L. Brigham, R. Chau, R. Gasser, R. Green, W. Hargrove, E. Lee, R. Natter, S. Thompson, K. Weldon, and S. Yang, *IEDM*, 273–276, 1994.
- T. K. Carns, X. Zhang, and K. L. Wang, *IEEE Electron Device Lett.*, **16** (6), 256–258, 1995.
- S. Hayakawa, M. Kakumu, H. Takeuchi, K. Sato, T. Ohtani, T. Yashida, T. Nakayama, S. Morita, M. Kinugawa, K. Maeguchi, and K. Ochil, *IEEE Int. Solid State Circuits Conf.*, 128–129, 1990.
- J. D. Hayden, F. K. Baker, S. A. Ernst, R. E. Jones, J. Klein, M. Lien, T. F. McNelly, T. C. Mele, H. Mendez, B. Y. Nguyen, L. C. Parrillo, W. Paulson, J. R. Pfeister, F. Pintchovski, B. M. Somero, and E. O. Travis, *IEEE Trans. Electron Devices*, **38** (4), 876–886, 1991.
- S. Ikeda, K. Asayama, N. Hashimoto, E. Fujita, Y. Yoshida, A. Koike, T. Yamanaka, K. Ishibashi, and S. Meguro, *IEDM*, 809–812, 1993.
- S. Ikeda, S. Hashiba, I. Kuramoto, H. Katoh, S. Ariga, T. Yamanaka, T. Hashimoto, N. Hashimoto, and S. Meguro, *IEDM*, 469–472, 1990.
- T. Kikuchi, Y. Onishi, T. Hashimoto, E. Yoshida, H. Yamaguchi, S. Wada, N. Tamba, K. Watanabe, Y. Tamaki, and T. Ikeda, *IEDM*, 923–926, 1995.
- M. Kinugawa and M. Kakumu, *IEDM*, 37–40, 1993.
- Y. Kohno, T. Wada, K. Anami, Y. Kawai, K. Yuzuriha, T. Matsukawa, and S. Kayano, *IEEE J. Solid State Cir.*, **23** (5), 1060–1066, 1988.
- H. Koike, Y. Unno, K. Ishimaru, F. Matsuoka, and M. Kakumu, *IEDM*, 855–858, 1994.
- T. Komatsu, H. Taniguchi, N. Okazaki, T. Nishihara, S. Kayama, N. Hoshi, J. I. Aoyama, and T. Shimada, *IEEE J. Solid State Cir.*, **22** (5), 721–726, 1987.
- C. Lage, J. Hayden, and C. Subramanian, *IEDM*, 271–274, 1996.
- M. Matsui, T. Ohtani, J. I. Tsujimoto, H. Iwai, A. Suzuki, K. Sato, M. Isobe, K. Hashimoto, M. Saitoh, H. Shibata, H. Sasaki, T. Matsuno, J. I. Matsunaga, and T. Iisuka, *IEEE J. Solid State Cir.*, **22** (5), 733–740, 1987.

- T. F. McNelly, J. D. Hayden, A. H. Perera, J. R. Pfiester, C. K. Subramanian, M. J. Blackwell, B. James, S. Ajuria, W. Feil, Y. C. Ku, T. Lii, J. H. Lin, F. Nkansah, C. Philbin, C. J. Sun, M. Thompson, and M. Woo, *IEDM*, 927–930, 1995.
- T. C. Mele, J. Hayden, F. Walczyk, M. Lien, Y. C. See, D. Denning, S. Cosentino, and A. H. Perera, *IEDM*, 481–484, 1990.
- O. Minato, T. Sasaki, S. Honjo, K. Ishibashi, Y. Sasaki, N. Moriwaki, K. Nishimura, Y. Sakai, S. Meguro, M. Tsunematsu, and T. Masuhara, *IEEE Int. Solid State Conf.*, 260–261, 1987.
- I. Naiki, M. Takizawa, M. Mano, T. Kimura, T. Ichikawa, M. Tsukamoto, S. Fujita, T. Nakayama, and M. Sasaki, *IEDM*, 817–820, 1993.
- T. Ohzone, M. Fukumoto, G. Fuse, A. Shinohara, S. Odanaka, and M. Sasago, *IEEE Trans. Electron Devices*, **32** (9), 1749–1756, 1985.
- J. R. Pfiester, J. D. Hayden, M. Thompson, F. Miller, J. H. Lin, M. J. Blackwell, K. Mocala, Y. C. Ku, C. K. Subramanian, W. Waldo, S. Ajuria, B. M. James, and B. Martino, *IEDM*, 623–626, 1994.
- B. Prince, *Semiconductor Memories: A Handbook of Design, Manufacture and Application*, 2nd ed., Wiley, 1995.
- H. Shimada, S. Kawashima, H. Itoh, N. Suzuki, and T. Yuba, *IEEE J. Solid State Cir.*, **23** (1), 53–58, 1988.
- S. Subbanna, D. Harame, B. Chappell, J. Comfort, B. Davari, R. Franch, D. Danner, A. Acovic, S. Brodsky, J. Gilbreth, D. Robertson, J. Malinowski, T. Lii, and G. Shahidi, *IEEE IEDM*, 441–444, 1993.
- R. D. J. Verhaar, R. A. Augur, C. N. A. Aussems, L. de Bruin, F. A. M. Op den Buijsch, L. W. M. Dingen, T. C.T. Geuns, W. J. M. Havermans, A. H. Montree, P. A. van der Plas, H. G. Pomp, M. Vertrgt, R. de Werdt, N. A. H. Wils, and P. H. Woerlee, *IEDM*, 473–476, 1990.
- T. Wada, T. Kirose, H. Shinohara, Y. Kawai, K. Yuzuriha, Y. Kohno, and S. Kayano, *IEEE J. Solid State Cir.*, **22** (5), 727–732, 1987.
- T. Watanabe, M. Hayasi, I. Sasaki, Y. Akatsuka, T. Tsujide, H. Yamamoto, O. Kudoh, S. Takahashi, and T. Hara, *IEEE J. Solid State Cir.*, **18** (5), 494–498, 1983.
- T. Yamanaka, T. Hashimoto, N. Hashimoto, T. Nishida, A. Shimizu, I. Ishibashi, Y. Sakai, K. Shimohigashi, and E. Takeda, *IEDM*, 48–51, 1988.
- T. Yamanaka, N. Hsegawa, T. Tanaka, K. Ishibashi, T. Hashimoto, A. Shimizu, N. Hashimoto, K. Sasaki, T. Nishida, and E. Takeda, *IEDM*, 477–480, 1990.

PART IV

13 TEM in Failure Analysis



Plan view on bipolar device IMD layer showing two slightly misaligned contact windows. It looks like Grand Pa's old clock with the pendulum lost.

Failure analysis in microelectronics often involves a few strategic steps. The first step is to identify and confirm the failure in the device and gather all the necessary information and background history. The second step is to measure and identify the failure modes. This step involves a few electrical tests and fault isolation techniques, such as liquid crystal, emission microscopy, voltage contrast, and e-beam tester. The third step is

failure mechanism analysis. Although the faulty location was identified and confirmed, detailed analysis on the failure site can reveal how and why the device failed. A lot of different analytical techniques can be used in this step, including optical microscopy (OM), scanning electron microscopy (SEM), focus ion beam (FIB), and transmission electron microscopy (TEM). Chemical information is vital in the failure mechanism's study and various techniques such as X-ray energy dispersive spectrometry (EDS), Auger electron microscopy (AES), X-ray photon spectrometry (XPS), and secondary ion mass spectrometry (SIMS) can also be involved depending on the analysis needs. The final step in a failure analysis is to find and correct the root cause based on the information gathered during the failure analysis. In practice, a complete failure analysis has been known to involve process and product engineers, testing engineers, failure analysis engineers, and even fab management individuals. Information sharing and communication are especially important during the process. In fact communication can become a bottleneck when turn-around time (TAT) is critical on a production line.

Among the techniques mentioned, TEM is becoming more and more important because the size of the device has been getting smaller. A $0.2\text{ }\mu\text{m}$ technology device failure often requires an observation on a faulty spot much smaller than $0.2\text{ }\mu\text{m}$. A combination of the resolution power of TEM and the capability of FIB to prepare the samples with pinpoint accuracy has made TEM applications in ULSI failure analyses practical and popular. The examples given here are in several important areas in failure analysis, including particle analysis and defect reduction, memory cell single bit failure analysis, EOS-induced damages, and ESD-induced damages.

13.1 PARTICLES ANALYSIS AND DEFECT REDUCTION

Particles or particle-like defects present one of the greatest challenges in modern ULSI process, as they are ubiquitous in the production environment. For the shrinking processes technology, the fatal particle size shrinks accordingly, and the number and density of them grow quickly (Menon 2000). In a typical process control procedure, a sophisticated optical in-line inspection system is the front-line for guarding particle and related defect issues. The system utilizes an optical imaging technology to view the device areas and compare them with a stored database. The computer will then determine if the differences between the reference images and the detected images are subtle or vital and remember the location whenever necessary. A large number of reference images are required in each and different layers of the ULSI processes to facilitate such comparison. Two important factors determine the effectiveness of the approach: the resolution limit of the optical system and the database along with the computer system in making the decision. Once the defect locations are identified, a statistical map (wafer map) of the defect locations and numbers is available in the computer for further scrutiny. With the predetermined database, the computer can roughly categorize the defects into several groups to facilitate the failure analysis if necessary. This greatly reduces the human effort—and error—in trying to determine the often mixed types of particles and related defects. Once the defect density and/or distribution exceeds the predetermined number or density, failure analysis of the defects becomes necessary. Modern ULSI wafer fabs allow the defect analysis to be done in-fab and at-line or on-line. The computer-generated defect maps are transferred automatically into a focus ion beam (FIB) for cross-sectioning. FIB with SEM imaging and EDS

capability now allow the analysis to be done all in one vacuum chamber without delay. If the particle defects are small, TEM is needed. Special tools are available to allow for TEM samples to be cut, lifted out, and transferred to proper sample stagings all in one FIB chamber with manual or automatic operation. STEM detectors are also fitted into the FIB chamber to allow for preliminary inspection of the target areas in the chamber. TEM not only possesses higher resolution when compared with SEM, but also the image contrast mechanism is different and provides insight into the microstructure, which is impossible by SEM.

After the particles are identified by the optical inspection system, normally preliminary analysis can be made. One important step that needs to be done immediately is determining whether the particle is a surface and loose defect or is an embedded defect. If it is a surface particle, a protective thin film is applied on it to keep the particle from moving or damaging during the sample preparation. Oxide, nitride, or even polysilicon thin films are among the most used materials for this purpose.

Polysilicon Particles

A large proportion of the particles found in the middle layers of a ULSI process are polysilicon particles. Polysilicon particles can be further categorized, by their nature, into three different categories: grown-in particles, etch residues, and foreign particles.

Particles and defects are detrimental not just because in the majority they can cause circuit short or leakage. The most detrimental effect they can have is to change the topography of the device's area surface. The topological change in one layer will be enhanced and enlarged by the addition layers deposited on it and eventually the device's features will be distorted in the area. Figure 13.1 shows an amorphous Si film with a "haze" problem. A high density of surface bumps is observed by optical microscopy. The foggy, or "hazy," surface is in fact due to local crystallization of the polycrystalline Si particles. Crystalline Si particles tend to nucleate on the Si/SiO₂ interface or at a amorphous film's free surface. When nucleated on the top surface, a surface bump is formed, as seen in Fig. 13.1(b). During the annealing, such surface bumps do not disappear but give rise to topological issues for the processes that follow. Figure 13.2 shows a surface bump after the amorphous Si film is completely transformed into a polycrystalline film.

The real challenge for particle analysis are the particles that are discrete and low in number in a device. These particles can be sufficiently lethal to kill a major part of the yield and yet difficult to catch and analyze. Figure 13.3 shows a typical particle analysis result. The particle was identified and allocated by an optical in-line inspection system after an interlayer dielectric (ILD) oxide layer deposition. The FIB/TEM sample preparation was followed by TEM analysis. As shown, the particle is a polysilicon particle, not WSi_x nor an oxide particle as was suspected at first. The reason that the particle was not found after polyfilm deposition or WSi_x deposition is that of the size of the particle was too small to be observed by the detecting system. Not until the particle size gets about three times bigger, after oxide deposition, will the inspection system be able to catch it. One advantage of using TEM is demonstrated here clearly: the internal texture of the particle reveals that the origin of the particle is at the polysilicon film's bottom, and not its top. This is a characteristic of a grown-in particle where the seed and nucleation site of the particle must to be identified in order to know "when" the particle was formed. Such information provides vital root cause identification and can thus lead

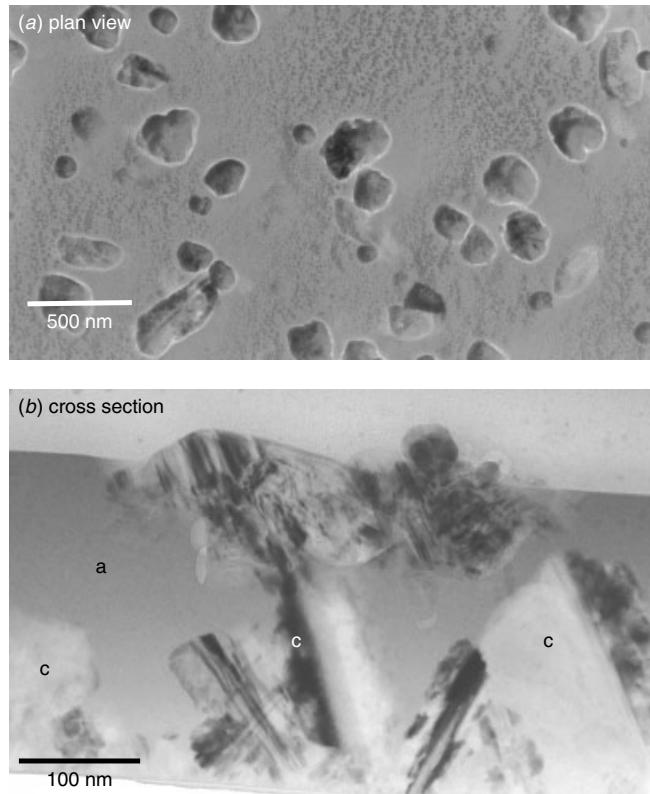


Figure 13.1 Plan view (a) and cross section (b) of an amorphous Si film showing a “haze”. TEM cross section, reveals that local crystallization nucleated on the films top surface has created surface bumps, which cause the haze.

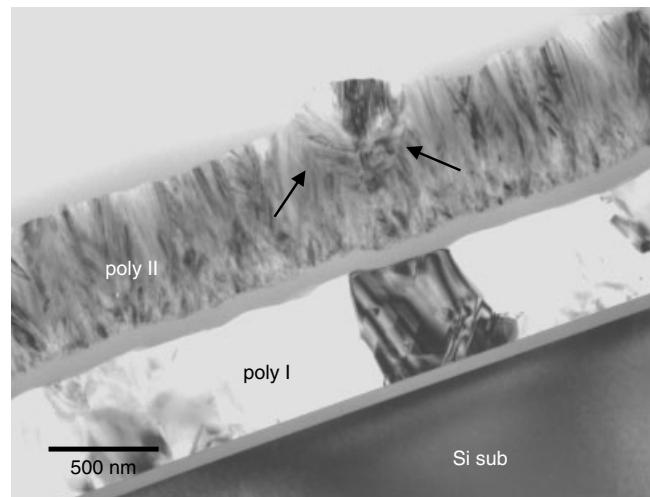


Figure 13.2 TEM cross section TEM of the polysilicon film showing surface bumps, as indicated, that remain after crystallization annealing.

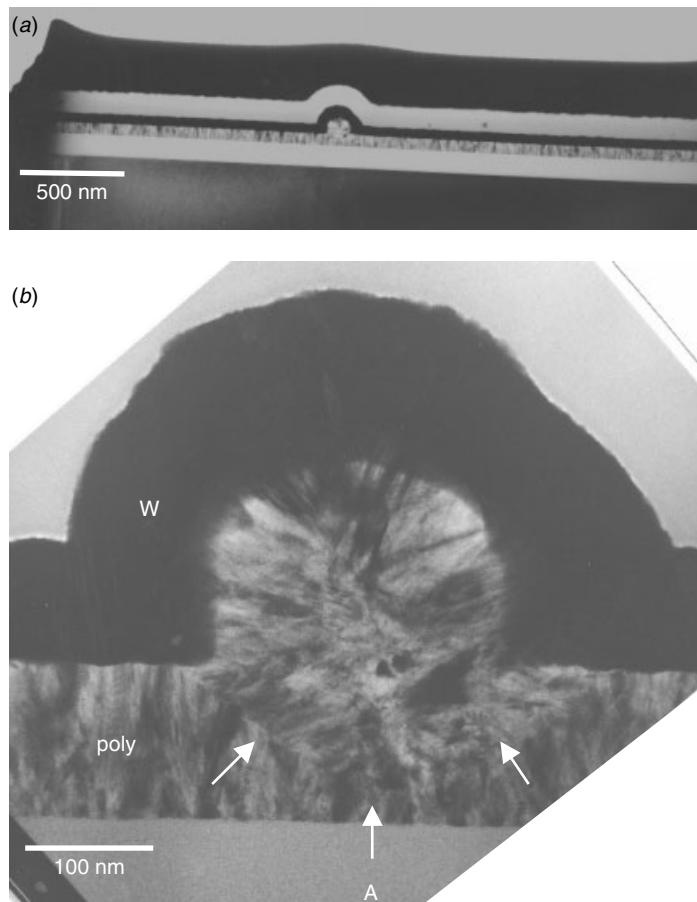


Figure 13.3 Particle defect growth from within the polyfilm. The microstructural grain texture shows that the particle may originate from the polyfilm's base and grow along with the polyfilm. Arrow A indicates the particle's origin.

to corrective action. Figure 13.4 shows another example of a grown-in particle. In this case the particle's origin is hard to determine based on the TEM analysis. Depending on the nature of the process, particles are mostly attracted to a wafer's surface by its static charge. These particles tend to sit in the corners or boundaries between the oxide and the conducting materials, for example, field oxide corners and metal line corners. As FIB/TEM cross sections are performed, the topology of the particle and the field oxide corner's shape can often be confused in the analysis result. This is better illustrated in Fig. 13.5. From the TEM micrograph one can still tell that the particle is a polysilicon particle and that the interface between particle and polysilicon film is roughly at a depth in parallel with the gate oxide's surface. But the exact location of the particle's origin is hard to tell because of the confused morphology. Figure 13.6 gives another example of an embedded then grown-in particle. Apparently the particle was dropped on the film near the beginning of the polysilicon deposition. Subsequent polysilicon deposition then grew and wrapped around the seed particle.

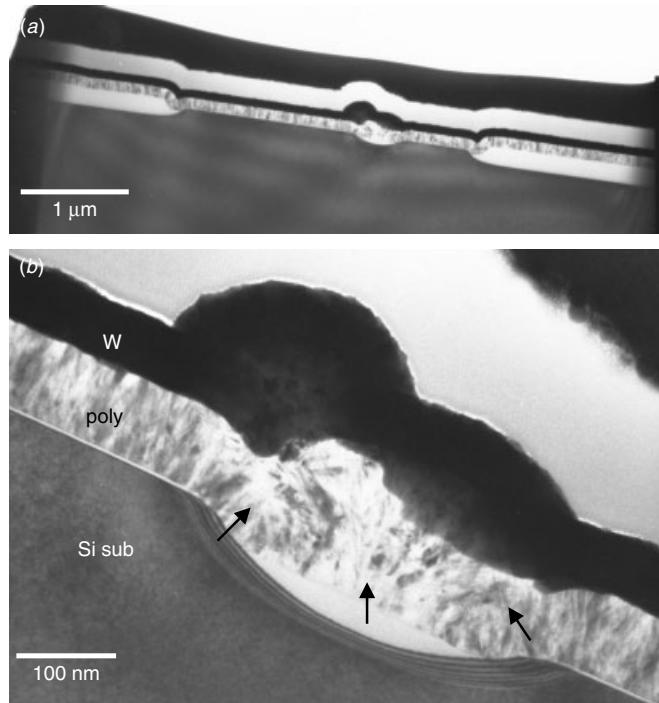


Figure 13.4 Particle defect growth from within the polyfilm. The image looks confusing because the particle was next to a field oxide corner. The film contour and the particle microstructure have intermixed, the origin of the particle cannot be easily deciphered.

Another category of polysilicon particles is the etch residue of polysilicon. Often such particles are extremely small and cannot be detected until several layers afterward. Figure 13.7 shows a few examples. Etch residue particles have one common characteristic. They all show flat bottom interfaces, with one exception as seen in Fig. 13.7(b) where the residue silicon has recrystallized and fused with the bottom polysilicon film, becoming a local extrusion from the bottom of the polyfilm. Etch residues are often symmetrical, as shown in Fig. 13.7(c), or shaped by a mask edge, as seen in Fig. 13.7(a).

Particles that originate from foreign sources make up a large proportion of the particle defects found in a process line. Figure 13.8 shows a typical example where the particle is spherical, indicating a foreign source rather than grown-in or etch residue. The particle sits on the SiO_2 surface and is embedded by polysilicon and WSi_x . Clearly, the particle was there at the beginning of the polysilicon deposition. The polysilicon and W silicide depositions, the patterning and cleaning that came afterward, all affected the particle and its surrounding materials and created the “bull’s eye” morphology seen in Fig. 13.8. Figure 13.9 shows a rare case where the foreign particle was deposited after the gate oxide’s formation but before the polysilicon gate’s deposition. Recall that this is one of the most strictly controlled process steps in terms of contamination and particle control. Again, TEM’s capability in revealing the grain texture directly helps us determine when the particle was deposited. Figure 13.10 shows an intriguing case.

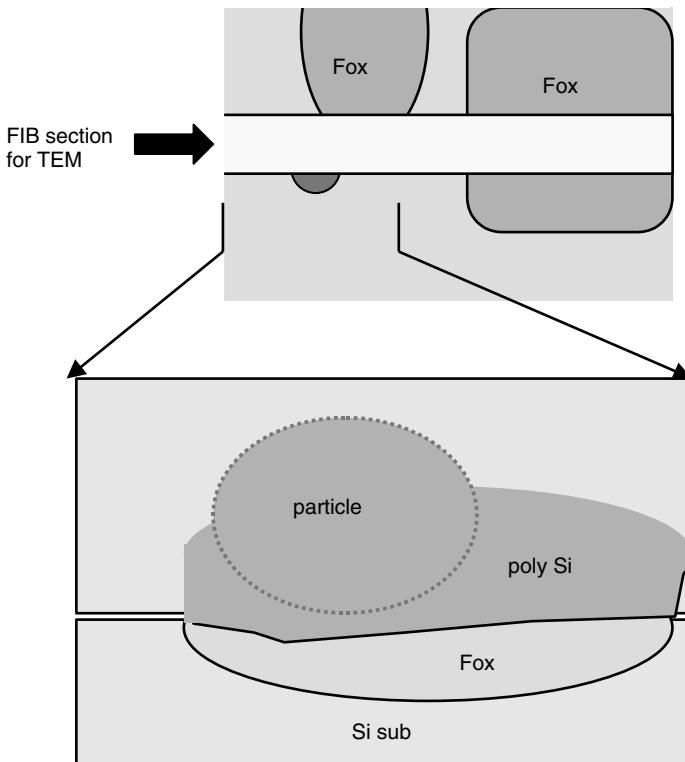


Figure 13.5 A schematic location of the particle in the previous figure. The FIB section for TEM has a thickness of about 100 nm. The polyparticle has overlapped with the polysilicon film on the bird's-beak slope, so the resulting image is confusing.

The particle's grain texture appears to be strikingly similar to the polysilicon grains immediately underneath. The particle must be a grown-in particle that later cracked and started to detach from the polysilicon film. Further analysis on similar samples have confirmed that detached particles can return back to the wafer and re-deposit as seen in Fig. 13.11. While the two cases, Figs. 13.10 and 13.11, show entirely different particle morphologies at different locations, the particles have originated from the same problem. The problem of why there is abnormal and local polysilicon film growth and later detachment from the film requires more study of polysilicon thin film deposition.

In most of the cases the exact layer where the particle sits can be determined by TEM cross-sectional analysis. The challenge is to trace back the source of the particle based on its location in between layers as there are generally numerous photo-stepping, etching, cleaning and drying processes involved. The process steps can be easily broken down into 10 to 20 main steps. Detailed analysis of the particle-to-substrate layer interface can sometimes provide help in finding the source of the particle. Figure 13.12

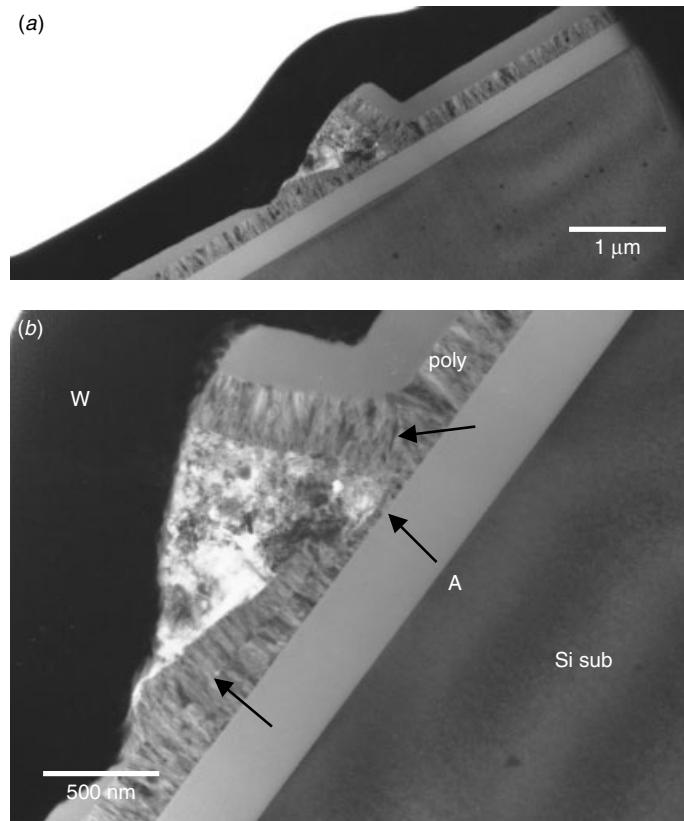


Figure 13.6 Particle defect grows from within the polyfilm. Arrow *A* indicates the particle's origin. Again, TEM reveals that the internal texture of microstructure helps to identify the root of the particle.

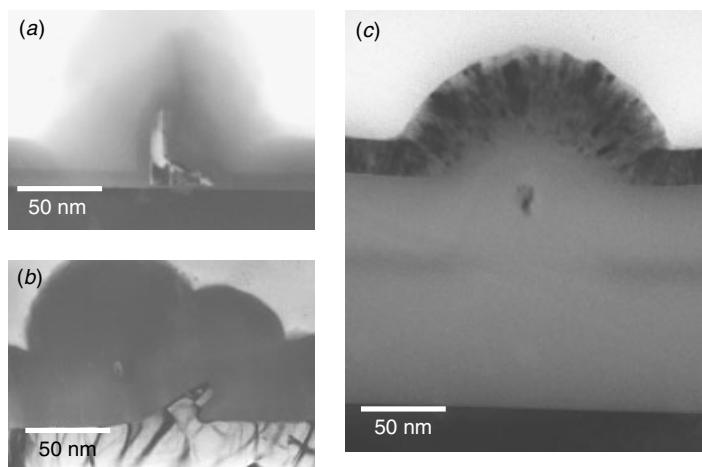


Figure 13.7 Typical etching residue particles on (a) the gate oxide, (b) the polysilicon thin film surface, and (c) the field oxide.

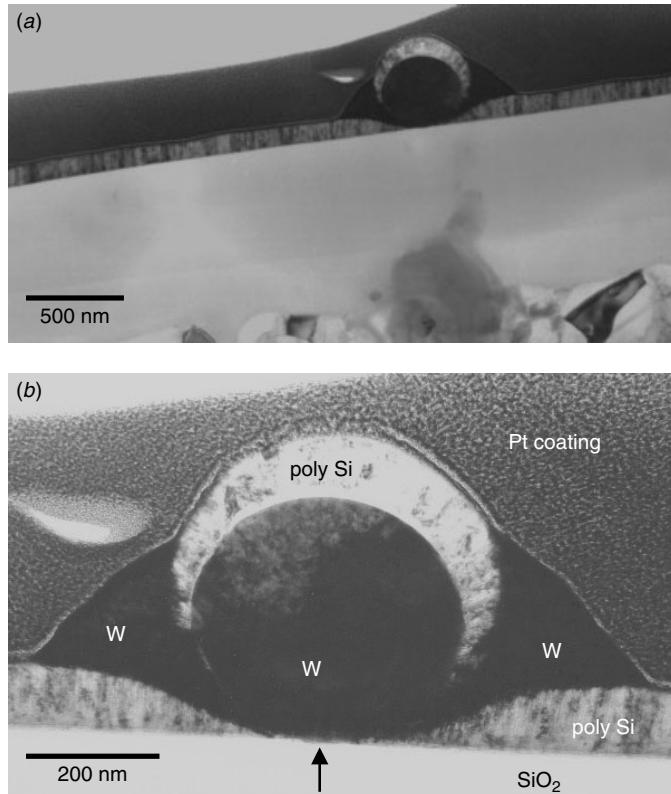


Figure 13.8 A polysilicon particle embedded within the W pile-up. The spherical polysilicon particle has most likely originated from a foreign source. The particle's base appears to sit on the SiO₂ surface, as indicated, suggesting that the particle was there from the beginning of the polyfilm deposition. (Sample courtesy Dr. An Yan Du, IME, Singapore)

shows such a case where the interface between the particle and the underlying polysilicon film is clear, indicating that the particle was deposited right after the polysilicon film deposition and likely before the wafer left the chamber.

Oxide Particles

Oxide particles, by nature, do not create leakage paths. However, there are a number of cases where leakage issues can be traced back to the oxide particle. The reason is that most particles are charged with electrostatic, and this makes them ideal attractants of contamination ions and wet chemical residues. So even oxide particles can give rise to leakage. Again, topological change is the most serious problem coming from any particle deposit.

Figure 13.13 shows oxide particles situated on a field oxide near a polysilicon runner. Close examination reveals some debris next to the particle, Fig. 13.13(b). More debris is found near-by, as seen in Fig. 13.13(c). Such a group of particles close to each other suggests a few suspect origins and is helpful in identifying the particle's source. Another kind of particle can arise from an extremely small seed area; usually

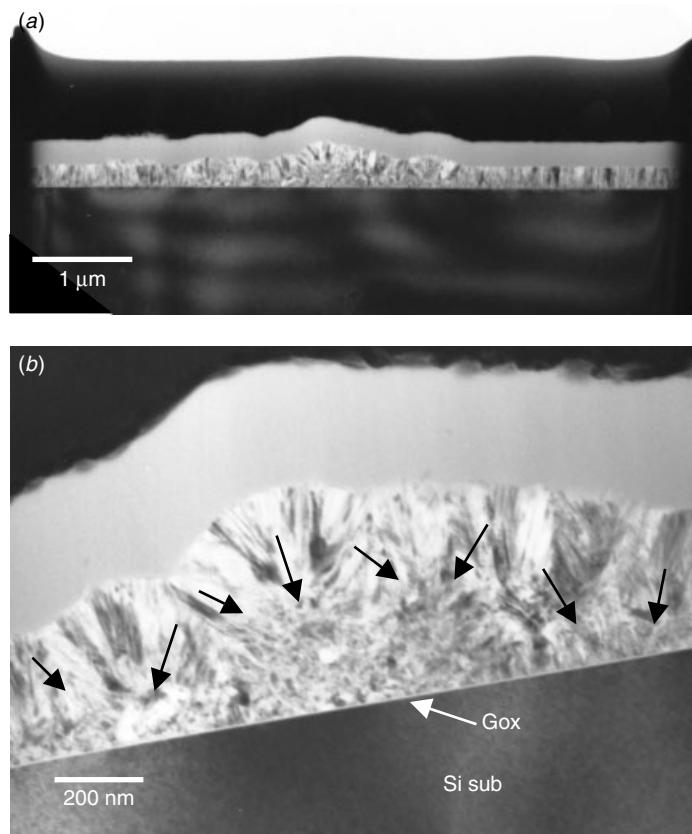


Figure 13.9 Rarely is the foreign particle deposited after the gate oxide's formation and before the gate polysilicon's deposition. The polygate has wrapped around the particle, as indicated.

these small particles, or even local chemical contamination, are enlarged by subsequent deposition and etching. Figure 13.14 shows a particle with symmetrical morphology and inner structure, suggesting that it is an etch residue from the spacer etching and formation process. The original seed is right in the middle. Besides the main particle some debris appears at the upper corner. This debris could have loosened from the main particle during the oxide layer deposition.

Other particles

Other than polysilicon and oxide particles, the particles can be metals, silicides, photoresist, polymers, and mixtures among them. Figure 13.15 is an example of metal residue on top of a silicide active region. With a sufficient database, the inspection system could be taught to recognize and categorize the particles from a previous database, and this would greatly reduce the analytical effort.

Some particles cannot be easily categorized. Figure 13.16 shows such an example. A particle was found a few layers after it formed within W polycide between WSi_x and the polysilicon interface. The nature of the particle was difficult to identify by

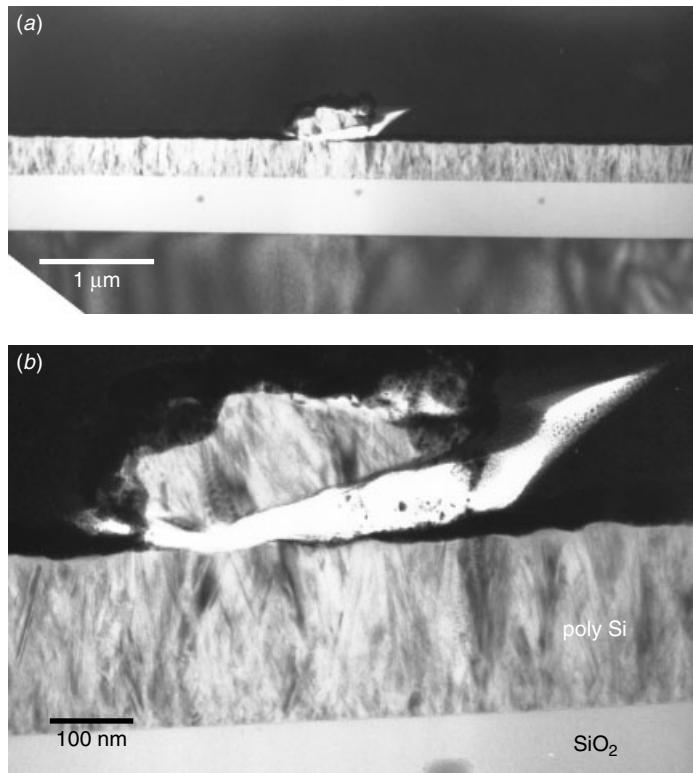


Figure 13.10 A particle sitting on top of the polysilicon runner (the polygate serves as a conductor on top of field oxide). The similar grain texture of the particle and the polygrains immediately underneath suggest the particle to be partially grown in and about to detach from the polysilicon film.

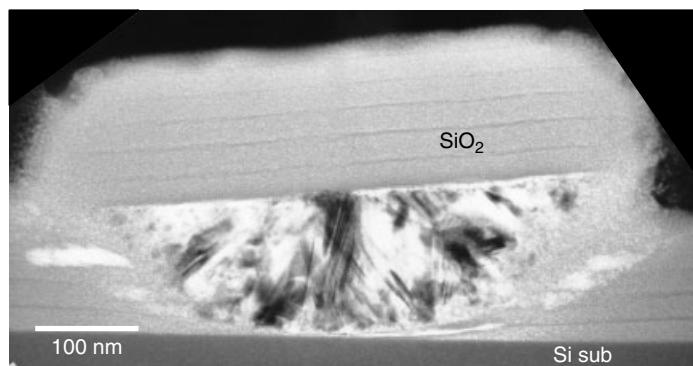


Figure 13.11 A particle that has detached from polysilicon film surface, and re-deposited onto the wafer's area where there is no polysilicon film. The particle was protected by the deposited oxide layer.

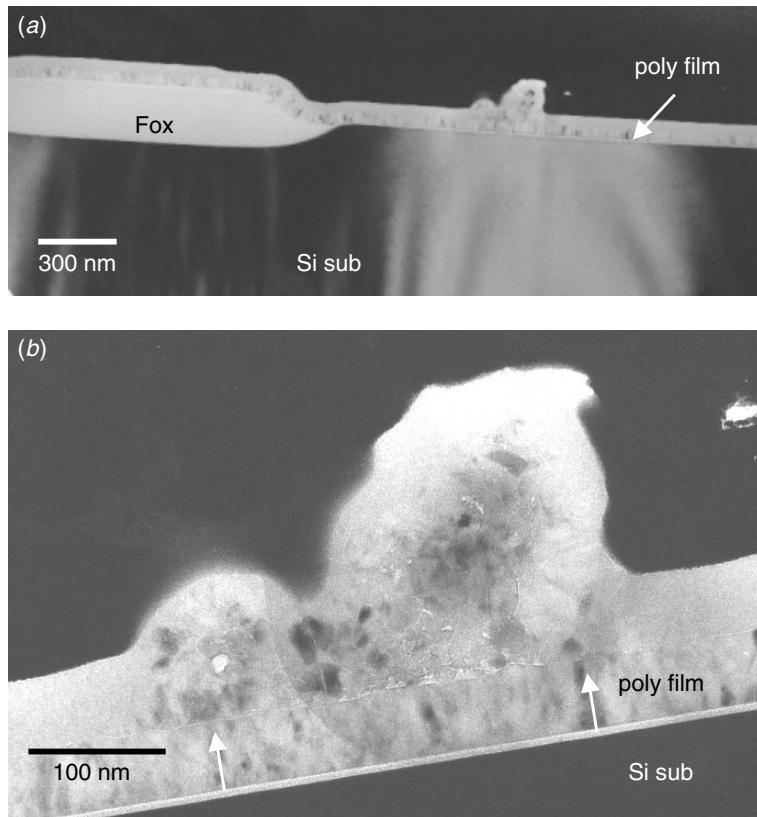


Figure 13.12 Detail of a particle on top of the polysilicon gate. The particle to polysilicon interface is very clean, as indicated, suggesting that the particle was deposited immediately after the film deposition and before the wafer left the chamber.

EDS analysis. Morphologically it seems to be a “soft” blister sitting on the polysilicon before WSi_x was deposited. However, its original nature is uncertain. The chemistry involved in WSi_x deposition process may have altered the nature of the particle.

Figure 13.17(a) shows examples where only a shadow of a particle is left. The particle had attached to the Si wafer’s surface right before or during the photo-stepping. Insufficient exposure on the photoresist or the later etching or oxidation in the particle area has resulted in the substrate curvature. Thus the sloped contour of the Si substrate is the shadow footprint of the original particle. Unfortunately, when an optical inspection system finds an abnormality, the original particle has long gone, and there is no way to find in which process step the particle problem was introduced and what kind of particle has caused the problem. Figure 13.17(b) shows another such example of a mysterious particle. The optical system has identified a small discoloration. TEM shows that the discoloring is due to a slight difference in oxide thickness. The change is subtle and gradual, as shown in Fig. 13.17(b). As most processes call for the wafer surface to be rinsed with water, this defect may be due to a water stain left by the cleaning and rinse processes.

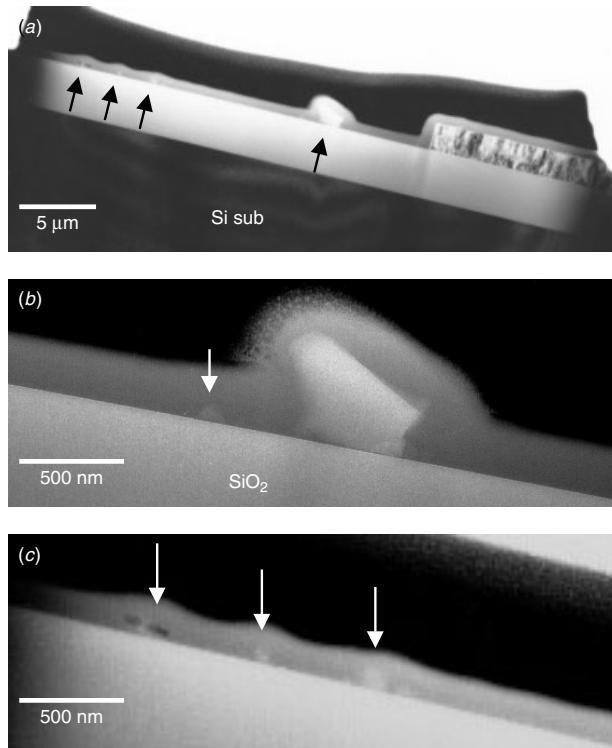


Figure 13.13 An oxide particle on the field oxide. Smaller debris is also observed, as indicated. Such accompanying particles are helpful in determining the particle's source and origin.

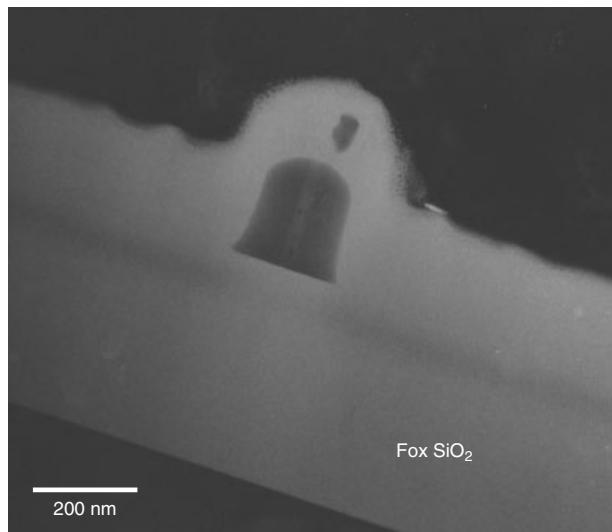


Figure 13.14 An oxide particle with symmetrical contour suggesting that this is etching residue. The seed particle seems to be right in the center (bright area) of the particle. The shape of the particle suggests it to be etch residue left during the spacer's formation. Smaller debris is also seen on the upper corner.

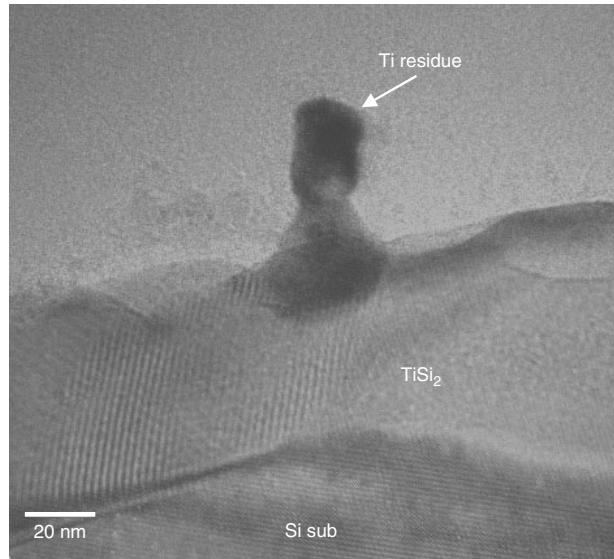


Figure 13.15 Metal residue defect on silicide area. EDS shows the particle to be pure Ti metal etching residue.

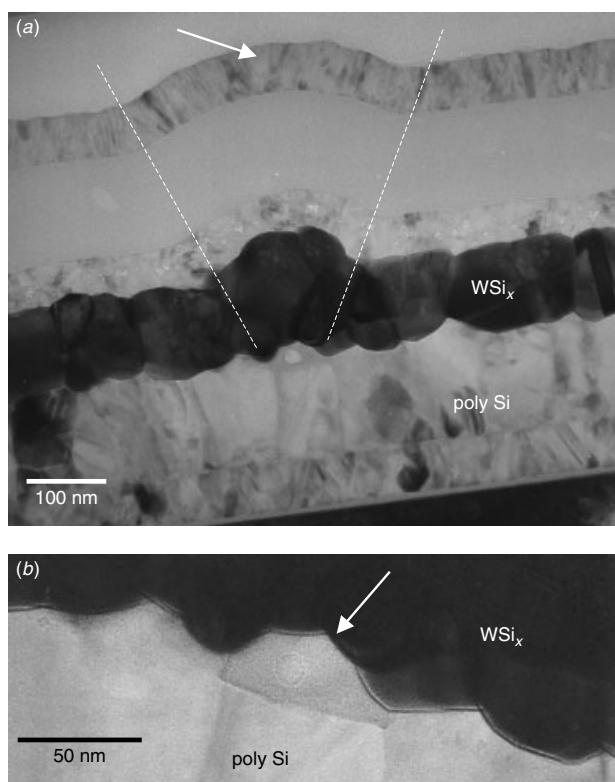


Figure 13.16 A particle-like defect identified by an optical inspection system. TEM result shows it is a blister between the polysilicon and WSi_x .

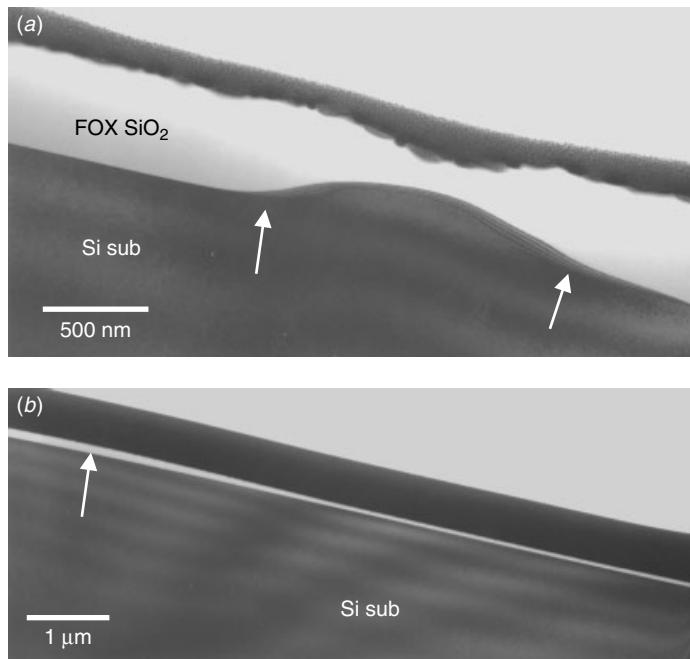


Figure 13.17 (a) Example of particle's imprint on the Si substrate. The particle was removed during photoresist stripping and is no longer present on the wafer. (b) Oxide thickness variation induced by a water stain, as indicated.

13.2 SINGLE-BIT FAILURE ANALYSIS

Most memory products, like DRAM, SRAM, EPROM, and Flash memory, suffer major yield loss in encountering single-bit or cluster-bits failures. This is a particularly difficult type of failure analysis. Among all the analytical tools available, FIB plays an important role in single-bit failure nowadays. Again, the combination of FIB and TEM provides one of the most powerful tools in single-bit debugging effort.

Nevertheless, FIB is not the only solution. Single-bit failure analysis depends largely on electrical testing and fault isolation in determining the subsequent physical analysis procedure and techniques that are required (e.g., see Hawkins and Soden 2001). Good judgment is important in single-bit failure analysis, since the critical area involved is often much less than a unit bit cell area. Determining how to finalize the failure mechanism to one, and only one, involves not only delicate electrical testing and fault isolation methodology but also engineering judgment and experience. After the possible failure mechanisms have been reduced to one. Careful planning of what physical analysis to conduct is required. The ultimate objective is to be able to “see” the defective area directly. But sometimes even indirect observation can be helpful. An example given in Fig. 13.18 is twin-bit failure. Electrically it was identified as a stacked capacitor ONO leakage. Physically, it is impossible to “see” a pinhole defect within the ONO layer even within one single-bit capacitor area. Any form of cross-sectioning is ruled out, as clearly the probability of success in sectioning through a pinhole, and then imaging it, is about zero. However, a wet chemical etching is very sensitive to such a

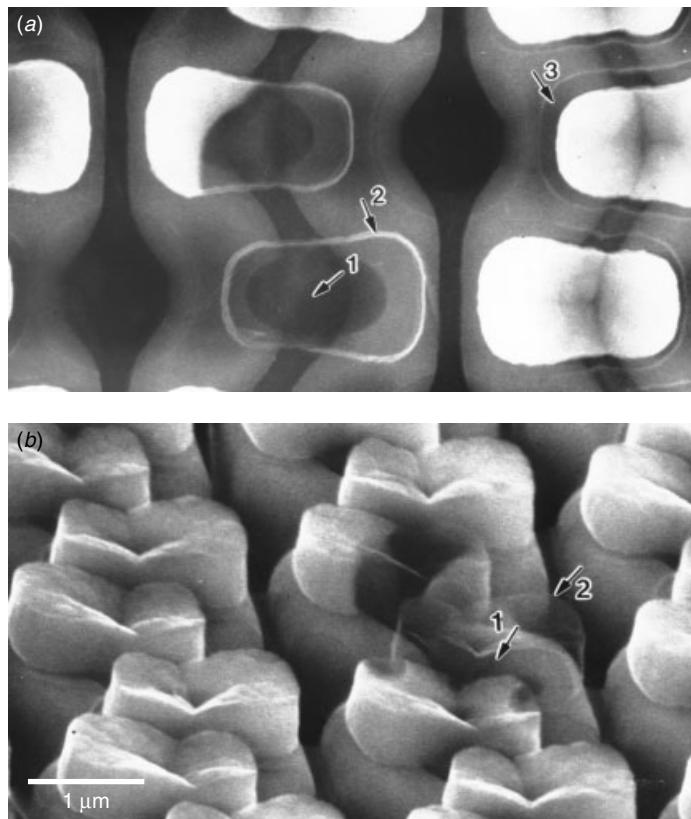


Figure 13.18 A twin-bit defect due to the stack capacitor's ONO leakage. (a) Top view of the etched defective capacitor. Arrow 1 shows the capacitor's contact to the substrate, and arrow 2 the stand-alone ONO shell. Arrow 3 shows a normal capacitor with the poly inside the ONO remaining intact. (b) Tilted view of the stand-alone ONO shell. (Courtesy of Ying Chu Lee of Vanguard International Semiconductor Corp., Taiwan)

microscopic defect. Indeed, the polysilicon etch solution can be used to eat away the outer polyplate (poly-3). In the normal capacitor the ONO protects the internal polysilicon (poly-2) from the wet chemical. In a defective ONO, the chemical can easily seep through the pinhole(s) and attack the inside poly. Figure 13.18 shows this clearly. The defective capacitors (the two adjacent ones) show poly-2 being attacked while the normal stacked capacitors show no poly-2 damage. Once the poly-2 is attacked, ONO stand alone as the distinctively transparent image seen in Fig. 13.18(b).

As we mentioned previously, TEM cross-sectional sample preparation techniques can be applied in the horizontal direction (or plan view sample). The sample is prepared with great precision. The x - y location is marked precisely either by laser cutter or by FIB, and the z -axis is sectioned precisely by referring to the layer structure and layout. This procedure can be combined with wet chemical etching to remove the layers to the designated layer. Figure 13.19 shows a typical single-bit failure that required a plan view TEM analysis. Laser marks are used in this case for positioning. The marks are easy to see during mechanical polishing and lapping. Normal FIB cut marks

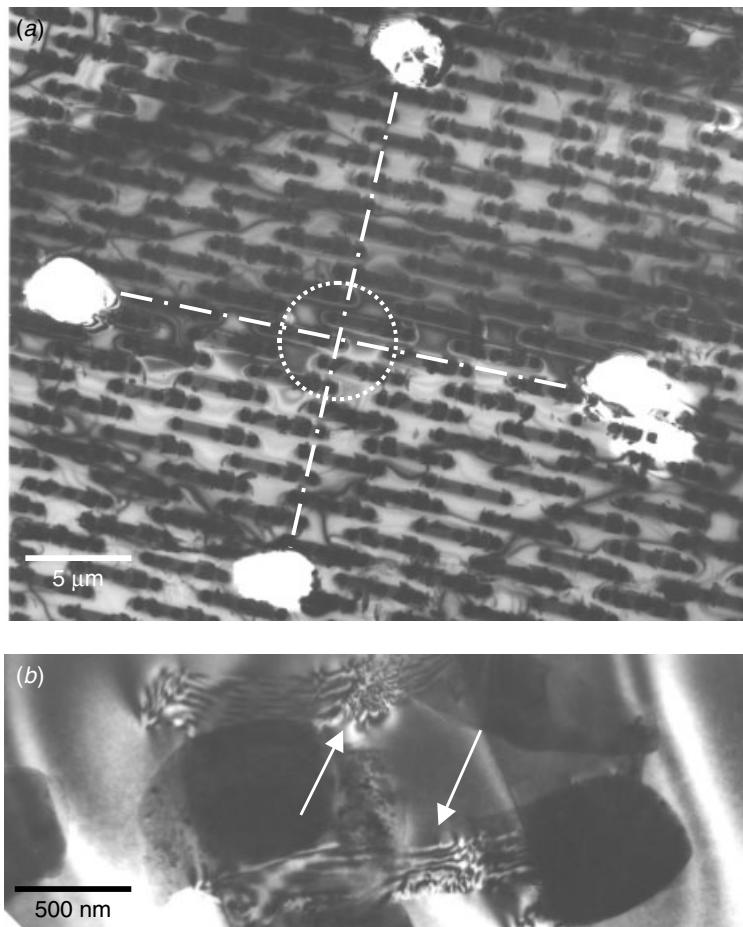


Figure 13.19 TEM plan view of laser marks deep into the Si substrate and easily identified under a low-magnification optical microscope. (a) The target structure is located in the central cross point. (b) Close-up at the target area shows the Si substrate defects that cause device failure.

are small and not easily seen under an optical stereo zoom microscope. After the laser marking, the sample is carefully horizontal lapped down to the desired layer, and a wet chemical etches away the oxide. Figure 13.19(b) shows a suspected defect area with elongated substrate defects extended across two adjacent active areas.

Cross-sectional TEM with FIB-assisted sample preparation provides the most powerful single-bit failure analysis capability in accuracy, spatial resolution, and turn-around time. This latest instrumentation technology promises complete analysis within an hour. Such quick and accurate analysis capability is indispensable in any modern wafer fabrication. Figure 13.20 shows a single-bit failure TEM/FIB cross-sectional micrograph. With the help of process and product engineers, the origin of defects can be identified and corrected before propagation. Another advantage to using FIB for the final sample preparation thinning is that it protects the defect area from the mechanical force that

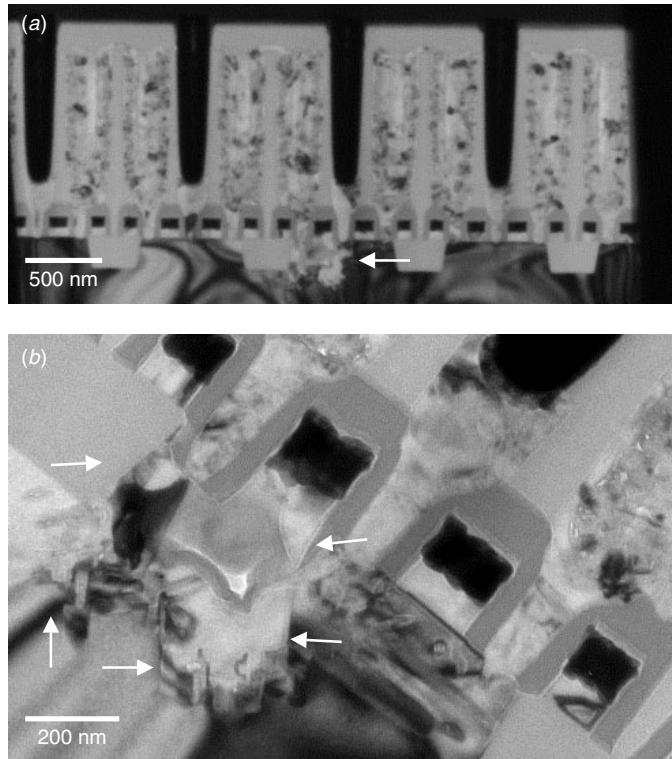


Figure 13.20 (a) Precision TEM cross section of a failure bit node contact, as indicated. (b) Abnormal Si substrate damage under the node contact and polycide gate clearly revealed. (Sample courtesy Dr. An Yan Du, IME, Singapore)

is inevitable during any mechanical polishing and lapping. A large number of failure mechanisms originate from thermal and mechanical stress, and lead to micro cracks and damages within the defected area. So much attention must be paid to preserving the physical integrity of the defective area. Mechanical polishing directly into the defect area should be avoided, as such an operation will change the defect area's morphology. Vital evidence that can relate the defect's history may be destroyed permanently.

13.3 EOS-INDUCED DAMAGE ANALYSIS

Devices are subjected to electrical stress during reliability tests and field applications. Their electrical functionality depends on their physical integrity. Once their physical integrity is altered due to electrical stress, a failure may occur. Electric overstress (EOS) is the term generally used to describe such failure. Depending on the circuit characteristics, EOS damage has been observed in several different locations. Gate oxide rupture and contact failure are the major failure mechanisms observed. Others like damages in substrate active areas and VIAs have also been observed.

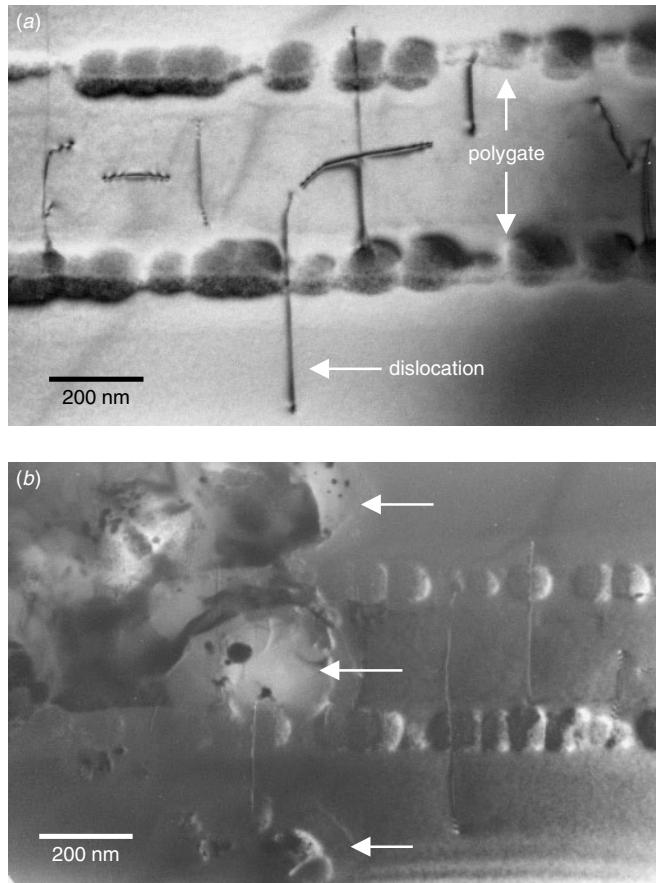


Figure 13.21 TEM plan view of a gate oxide's breakdown point. Areas without substantial damage show the dislocations to penetrate throughout the polygate (a). The dislocations are believed to be generated by local thermal stress due to EOS. Physical damages are observed in high field areas (b) as expected. (Tung et al., IPFA 2002, 65–69, reprint with permission from IEEE)

Gate Oxide Failure due to EOS

The physical failure mechanism of a gate oxide's breakdown has rarely been observed directly. This is mainly due to the difficulties involved in getting the breakdown point exactly located for a cross-sectional sample preparation. We gave in Chapter 6 a full physical analysis of ultra thin gate dielectric breakdowns using TEM. Figure 13.21 gives a plan view TEM that shows clearly an external damage spot outside the polygate toward the source's end. Interestingly a dislocation connects the external damage spot to the large damage spot within the polygate, as seen in Fig. 13.22. This is direct evidence that a dislocation within an active area is an excellent conductor (by gathering dopants), and can conduct high current swiftly between its two ends.

Figures 13.23 to 13.26 show an extreme case of gate oxide burnout. The location is a local gate oxide breakdown within a long serpentine gate structure with STI

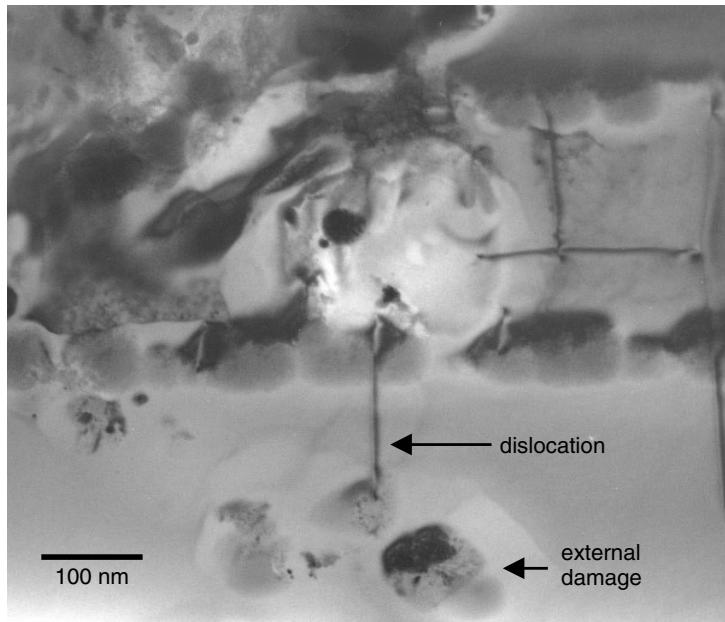


Figure 13.22 TEM plan view of a gate oxide's breakdown point. Same area shown as shown in Fig. 13.21(b). Detailed analysis shows the dislocation to have conducted the high field current and created a damage site outside the gate. This is a direct evidence that a dislocation can act as an excellent conductor and conduct a high current between its two ends.

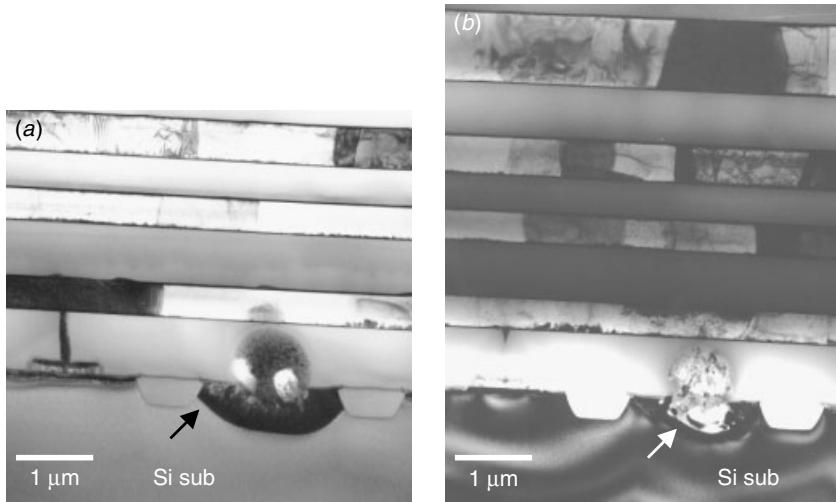


Figure 13.23 TEM cross section of a polygate burned-out structure. The burnout created a fireball within the ILD SiO_2 and thus a crater appears in the Si substrate underneath. (a) The fireball appearing when the TEM sample was still thick. (b) The internal structure of the fireball when the sample was further thinned down by ion milling.

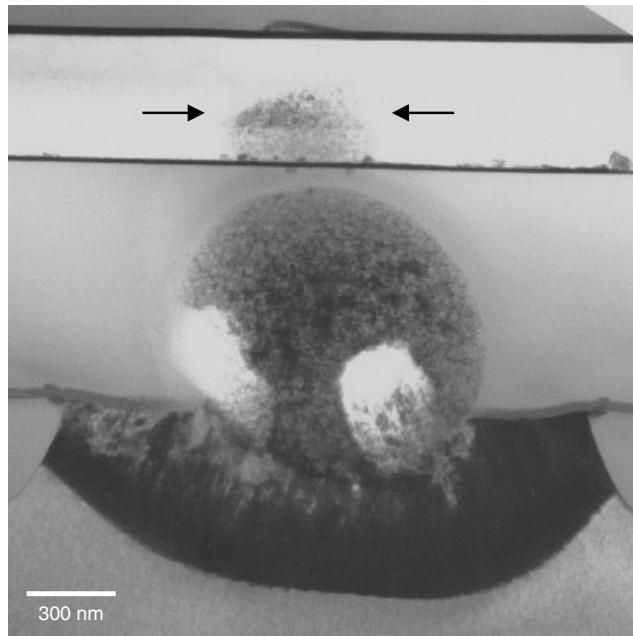


Figure 13.24 TEM cross section of a polygate burnout. The burnout has created a fireball within the ILD SiO₂ and a crater in the Si substrate underneath. The metal 1 Al lines directly on top of the fireball show signs of damage, as indicated. (Tung et al., IPFA 2002, 65–69, reprint with permission from IEEE)

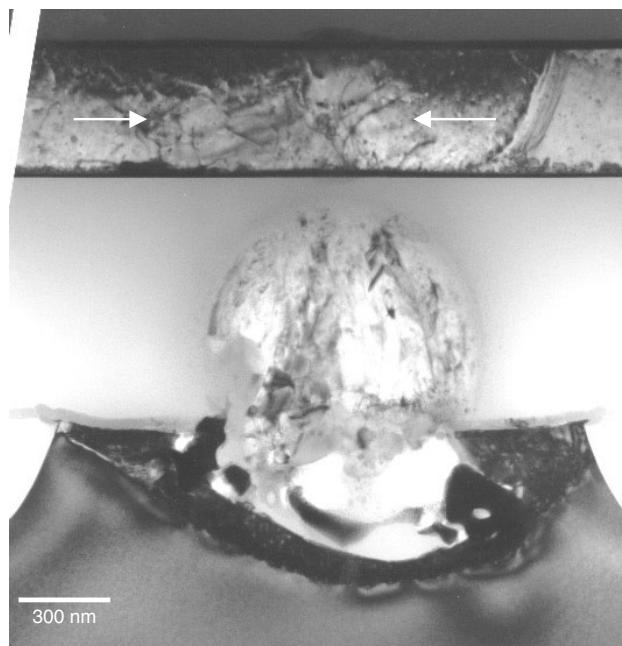


Figure 13.25 TEM cross section of a polygate burnout. The burnout has created a fireball within the ILD SiO₂ and a crater in the Si substrate underneath. Dislocations lines within the Al line were created by the stress due to the fireball, as indicated.

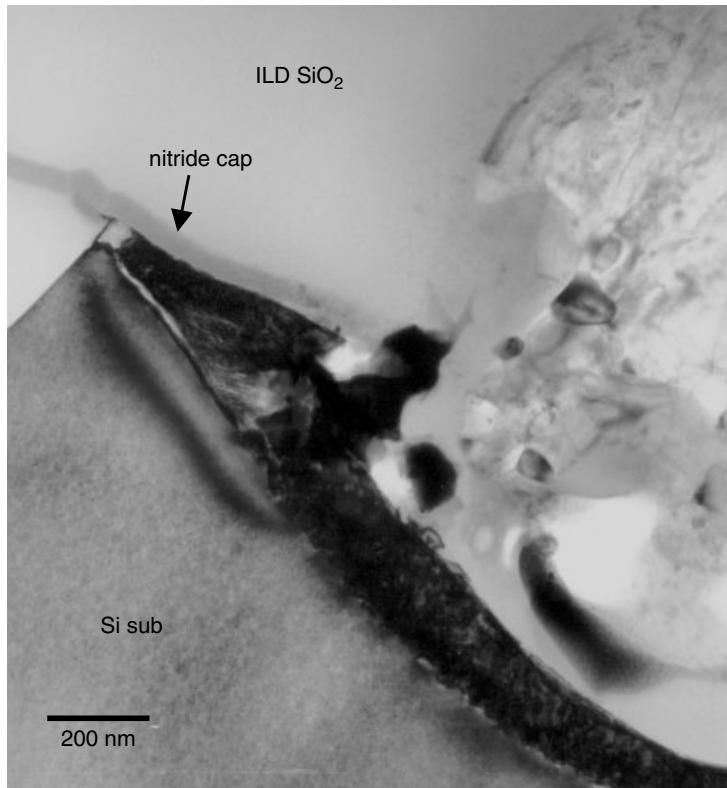


Figure 13.26 TEM cross section close-up at a corner of the fireball and the crater. Si-substrate smelting and damage create the slag, debris, and whiskers observed within the fireball. (Tung et al., IPFA 2002, 65–69 reprint with permission from IEEE)

isolation. There are four metal layers on top of the structure and thus the exact breakdown location cannot be identified using traditional methods like liquid crystal, or emission microscopy. The TEM sample was prepared by conventional polishing and grinding and subsequently ion milling. No FIB was involved in this analysis. Figure 13.23 shows the burned-out location's silicide gate, which was completely destroyed, and the destructive energy was so profound that a fireball was created within the solid state interlayer dielectric (ILD SiO_2). A crater was created under the fireball. Because of the massive stress from the fireball, the metal-1 Al metallization line directly above the fireball also showed internal damages and a high density of dislocations, as seen in Figs. 13.24 and 13.25. A close-up view of the fireball and the crater reveals Si and SiO_2 smelting, evaporation, and the presence of slag, debris, and whiskers, as seen in Fig. 26.

The energy involved in such a local gate oxide breakdown is very destructive. The analysis greatly improves our understanding of the gate oxide breakdown mechanism, identifies the initiation spot and thus the weak point along the current path, and helps our design and process optimization of the device structure.

Contact Failure due to EOS

When an external high current or high voltage stress is applied to a circuit, the interconnection node(s), where the resistivity is the highest throughout the circuit, will generate the most heat, transform the energy that is received into thermal energy, and dissipate it through out the surrounding materials. For aluminum metallization technology with W-plugs as contacts and VIAs, the high-resistivity points are usually at the lower interface of the contacts, that is, between the W-plug and the Si substrate, the polysilicon or polycide runners, or TiN ARC (anti-reflection coating) layer on top of the lower Al lines. In some cases, it could be within the W-plug itself.

Figure 13.27 shows typical cases where the contact was burned out and the smelting started within the W-plug. Interface heating has triggered a metallurgical reaction among W, TiSi_2 , and polysilicon, causing the W material's consumption. As W was consumed and reacted with polysilicon to form silicide, a continuous supply of W drove the W-plug to sink into the polysilicon line, and a notch, or necking, was created as seen in Fig. 13.27. The local high temperature caused the SiO_2 blistering. When the heating is fast, or the applied electrical stress is swift, the W contact necking may continue and eventually break open at the necking position. This is the case shown in Fig. 13.27(c) and (d). Figure 13.28 shows an interesting case where a VIA chain was damaged by EOS stress. All of the VIA/metal interfaces are damaged by the thermal energy. The final destruction point is at the M3/VIA3 interface. This is roughly at the midpoint of the entire VIA chain structure. A possible reason for the final burnout at the VIA chain midpoint is that this is the far point from the long metal lines (M1 and M5 in this case), which acts as a heat sink.

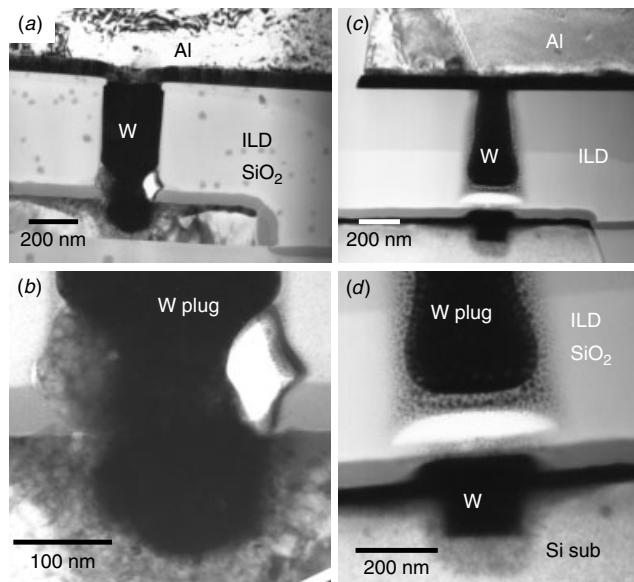


Figure 13.27 TEM cross section of burned-out W-plug contacts. (a, b) As the burnout began, the contact W-plug started to smelt and create a neck about $\frac{1}{4}$ the height from the polycide runner. (c, d) the W-plug has totally melted open. (Tung et al., IPFA 2002, reprint with permission from IEEE)

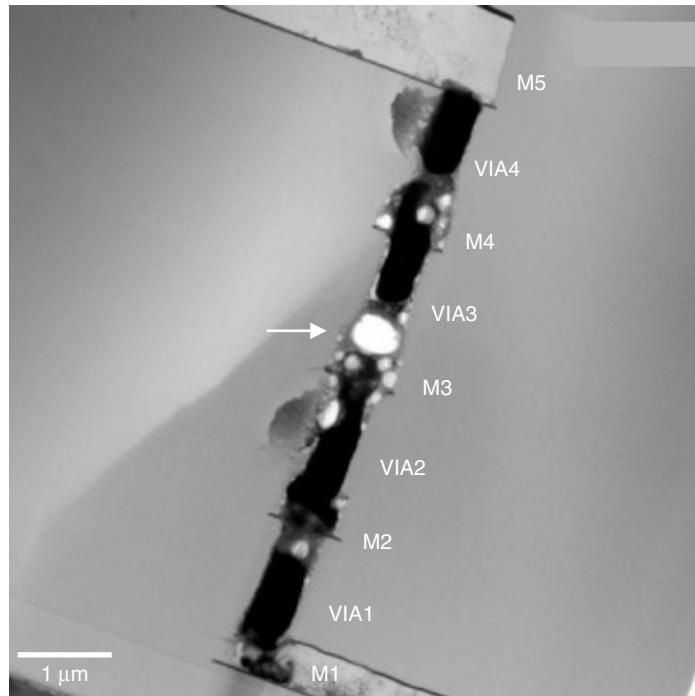


Figure 13.28 TEM cross section of the W–plug VIA chain EOS damage. Basically every VIA-to-metal interface has been damaged. The most severe damage occurs at the M3/VIA3 interface, as indicated. This is approximately the midpoint of the whole VIA chain structure.

Figure 13.29 shows the variety of VIA–pair EOS damage morphologies. As the overstress pulse hits the VIA–pair, both are vulnerable to the thermal damages, as seen in Fig. 13.29(a). Eventually one of them will be melted, and the result is an opening. For the component that is eventually burned open, the location is random (if the process is clean and free from high interface resistivity problems), as seen in Fig. 13.29(b) through (d). It could be Fig. 13.29(b) at the interface, Fig. 13.29(c), within the W–plug, or Fig. 13.27(d) where the whole VIA has evaporated.

When the applied electrical current is high and the pulse is short, the enormous energy supplied to the single contact needs to be transformed into thermal energy and dissipated. A local explosion, Fig. 13.30, may occur. The followings are observed:

- An explosion occurring at both the top of the polycide runner and below it, Figs. 13.30(a) and (b).
- A cleavage created at the end corner of polycide probably due to the explosion expansion force. The polysilicon material has melted and seeped into the cleavage gap; the temperature for such to happen is at least 1400°C, Figs. 13.31(a) and (b).
- A delamination gap created between polycide runner and nitride cap at a location about 3 μm away from the contact. This is probably due to local smelting and reaction between TiSi₂ and W, Fig. 13.30(a), arrow 4.

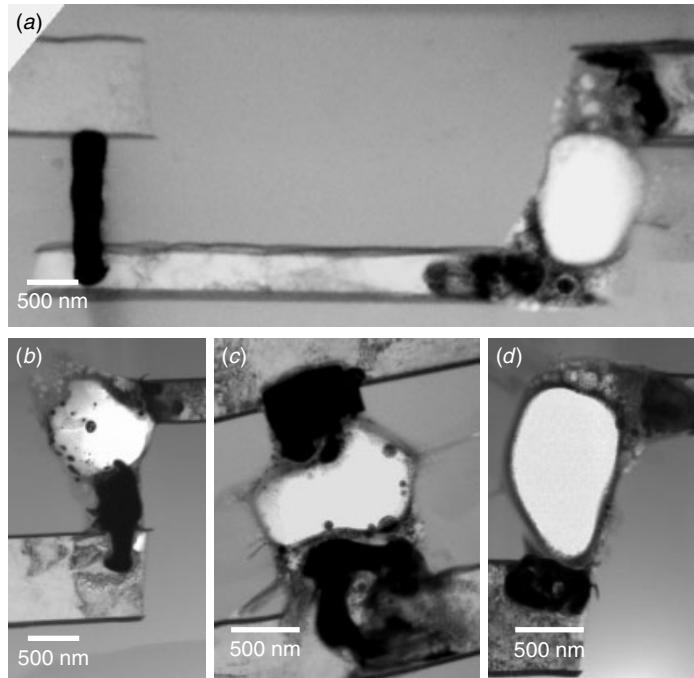


Figure 13.29 (a) TEM cross section of burned-out W–plug VIA–pairs. (b–d) Individual VIA burned out with various burnt morphologies: (b) The upper interface, (c) the mid point, and (d) the whole VIA.

- At the top of the fireball, SiO_2 is melted and the contact hole closed up entirely. The illusion created is that the contact hole does not come in contact with the metal on top. The temperature for SiO_2 to melt and flow is at least 1720°C , Figs. 13.30(a) and (b).
- At the top of the fireball, Al is damaged and buckling at the top, Fig. 13.30(a) arrow 5.

Studies of such typical EOS damage sites provide insight as to the energy consumption, phase transformation, and mechanical strength of the structural elements of a ULSI device. Understanding the destructive power and sequence of destruction in a microscopic area of a device has not only scientific importance but its engineering values.

13.4 ELECTROMIGRATION-INDUCED DAMAGE ANALYSIS

In an accelerated electromigration test, the failure usually occurs within the Al metal line if the line is long. But, if the line is short and connected with VIAs and contacts, a failure will normally occur at Al near the VIA structures. Figure 13.32 shows an electromigration-induced metallization failure near a W–plug VIA. The external current supply stopped when the resistivity became high. The TEM cross section shows that the lower metallization Al is nearly depleted. The materials left behind is mainly

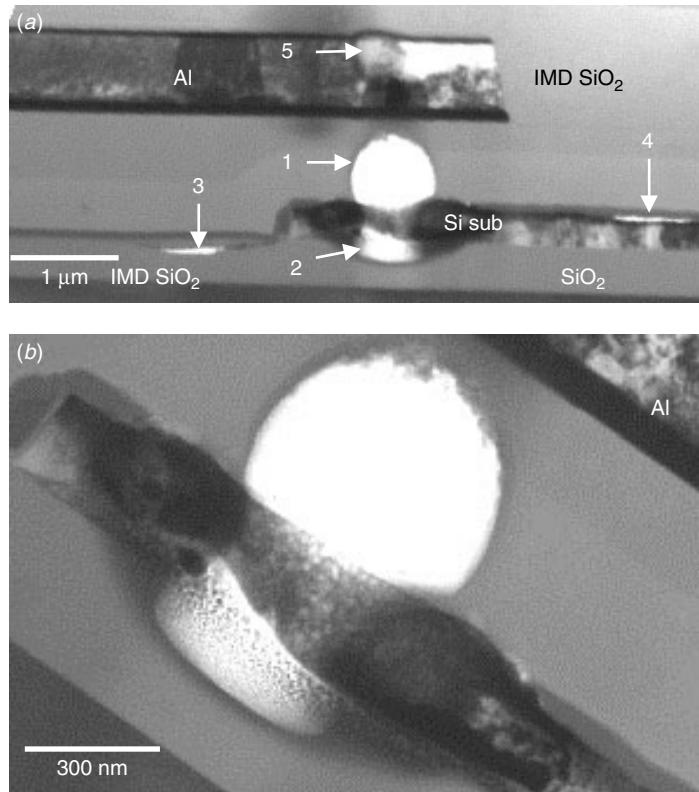


Figure 13.30 TEM cross section of a burned-out contact. (a) A micro-explosion has created a fireball above (1) and below (2) the polyicide runner, a cleavage (3) within ILD at the corner, a delamination between nitride cap and polyicide (4), and Al line buckling (5). (b) Close-up at the micro fireball shows the ball to be hollow.

the TiN/Ti barrier layer and TiN ARC layer. Only a small piece of Al under the W-plug is found. The barrier TiN/Ti buckles upward, and a small void on top of ARC TiN near the VIA is noticed. All of these damages are probably due to the thermal effect.

If the external current supply continues as resistivity becomes high, the migration effect will continue, and eventually an electrical open will spontaneously stop the test. For such a case the TEM results showed, as seen in Fig. 13.33, the followings:

- Both TiN/Ti barrier layer and TiN ARC layer depleted as well. The metal line area was swept clean without any trace of the conductive materials except at a corner behind the VIA structure, Figs. 13.33(a) and (c).
- W-plug material extruded like soft toothpaste, down into the Al line cavity and swept toward the electron-wind direction, Fig. 13.33(c).
- The upper corner of the VIA contact, originally protected and connected with TiN/Ti barrier layer, now ruptured and TiN/Ti barrier layers cracked and separated. Al has filled the crevice, Fig. 13.33(b).

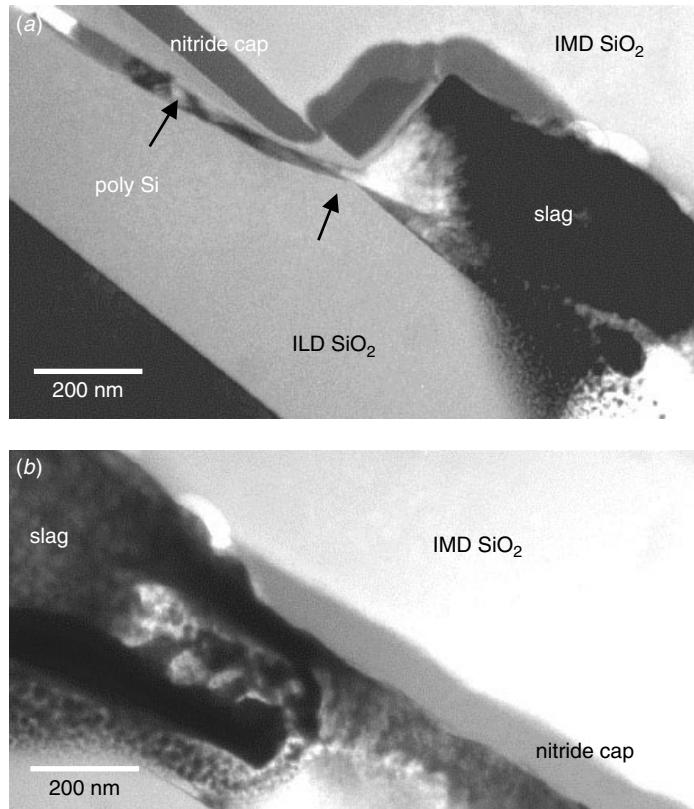


Figure 13.31 TEM close-up of Figs. 13.30. (a) A cleavage at the corner of polycide is observed, as indicated. (b) Local smelting and intermixing of W, TiSi₂, SiN, SiO₂, and polysilicon have created slag-like materials within the polysilicon layer.

Clearly, the electromigration damage is subtle and progressive compared to EOS damage. However, the powerful electron winds that have swept across the VIA and push the W-plug up like toothpaste is beyond imagination.

13.5 ESD-INDUCED DAMAGE ANALYSIS

In failure analysis electro-static-discharge (ESD) is usually categorized into the same group as electro-overstress (EOS). By their failure mode and failure mechanism they are usually undistinguishable, the difference being subtle. In a lot of failure analysis cases where the origin of failure is hard or impossible to identify, ESD and EOS are generally regarded as a single failure mode. However, physical analysis in the ESD/EOS failure parts shows that the two are distinguishable (Henry 2001). In comparing the examples given in previous session on EOS damages and examples that will be given below, it is easy to tell the difference between the two. Electrically they may show the same failure mode; physically, they can be distinguished, though sometimes the difference is subtle.

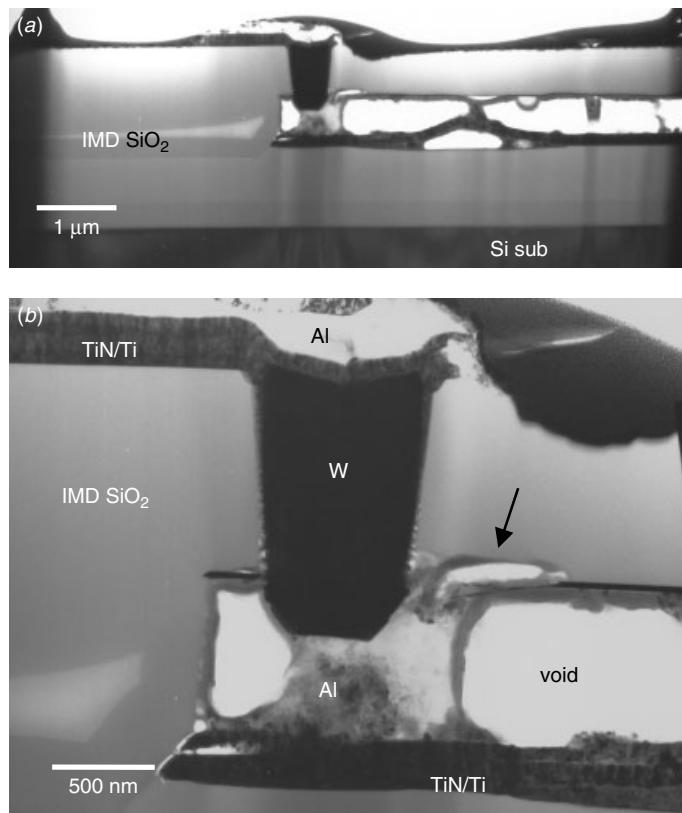


Figure 13.32 TEM cross section of an electromigration failure contact. The lower Al metallization (metal 1), shows excessive Al depletion. Almost all of the Al is gone except for a small piece under the W-plug. Notice that a small void extrudes outside of the ARC layer near the VIA, as indicated. This is probably due to thermal heating and rupturing of SiO₂.

As protection against ESD damages most modern ULSI devices are coupled with ESD protection circuits at their input/output bonding pads. Nevertheless, to test that the protection circuits are effective, practice allows a maximum ESD voltage to be applied to the protection circuit, industry standard for example, 2000 V for human body mode and 500 V for machine mode. If the device ESD protection circuit fails at certain simulated voltage, the protection circuit is modified and improved to meet these requirements. Unfortunately, often it was difficult for the circuit designers to modify and improve the existing design and circuit layout because the exact discharge path cannot be found, and the detail information on the discharge mechanism cannot be identified.

An ESD-induced part failure analysis normally starts with a fault isolation, where the exact leakage/damage location is identified by using liquid crystal, emission microscopy, or an e-beam tester. Because most damage sites are covered by metallization, the exact location of leakage is usually not certain. Figure 13.34 shows an optical and a plan view TEM micrographs at approximately the same magnification. ESD-induced damage spots at the four corner contacts are clearly visible. Figure 13.35 shows a plan view

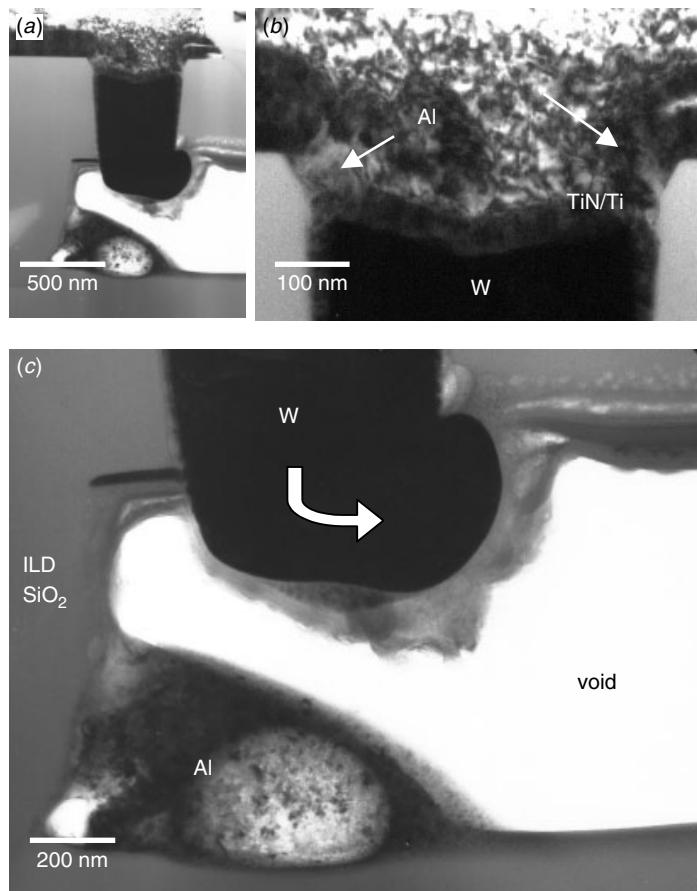


Figure 13.33 TEM cross section of an electromigration failure contact. The lower Al metallization (metal 1), shows excessive Al depletion. The contact plug tungsten materials, is extruded toward the electron direction just like toothpaste. At the contact's upper corner, the TiN/Ti layer is torn and separated, and Al has filled-in, as indicated in (b).

TEM micrograph. The Si substrate damages due to ESD power appear clearly. The damage shows where the heat, and thus the ESD current, has headed and is helpful in determining the discharge path. One may wonder why the contacts are damaged. The cross-sectional TEMs, seen in Figs. 13.36 through 13.38, reveal in detail the damaged mechanisms of the Si substrate and W-plugs throughout the discharge path. Notice that the Si substrate has become amorphous near the contact's bottom. Beyond this is a large area of damage spreading through the discharge path. This suggests that the damage started from the contact's bottom interface and gradually diminished as it reached the Si substrate. Again, what is demonstrated is that the contact's bottom interface is the highest resistivity point throughout the discharge path and thus is the initiation point of the discharge damage. When the contact's bottom interface resistivity is comparable with that of the W-plug, the W-plug material will share the discharge energy. Figure 13.38 shows a W-plug that has melted and is driven into the Si substrate, forming a boltlike

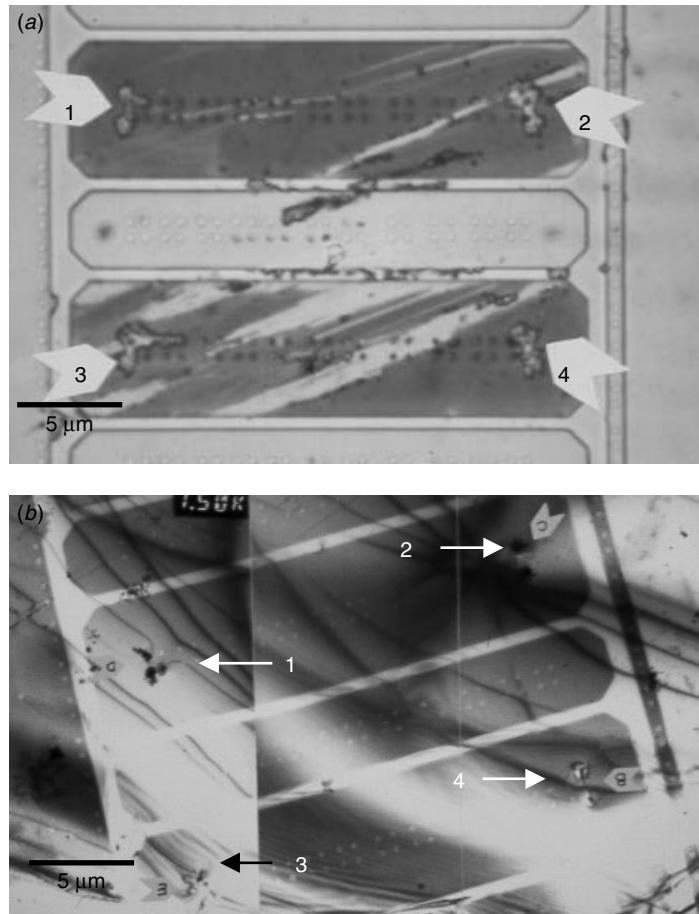


Figure 13.34 Optical planar view and TEM image at approximately the same magnification, showing four-corner ESD induced contact damages, as indicated.

structure within the Si substrate. W is a refractory material whose melting point is higher than 3400°C . This is the best demonstration of the destructive power of ESD discharge.

In studying ESD damage, our most important task is to determine the discharge path. We need this information to improve the ESD's protection circuit design. Figures 13.39 through 13.45 provide examples of a complete analysis. First, emission microscopy is used to determine the exact leakage circuit point, Fig. 13.39(a). After passivation removal using dry plasma, no physical damage is apparent on the upper metallization layer (metal 2), Fig. 13.39(b). However, further removal on IMD SiO_2 layer shows damages in four areas (Fig. 13.40):

- Metal 1 layer
- Field oxide
- ILD SiO_2 between the polysilicon line and a contact
- Polysilicon layers

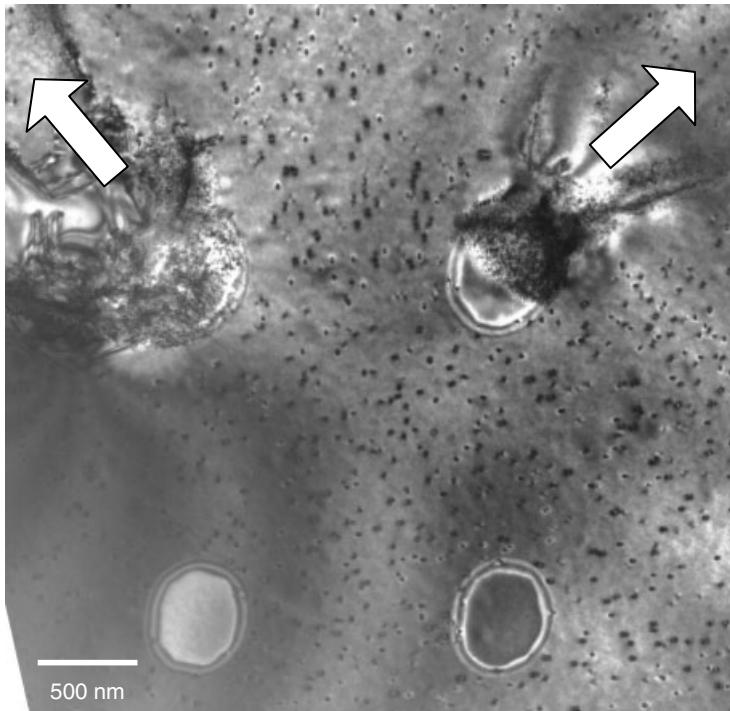


Figure 13.35 TEM plan view at the ESD damaged contact areas. The Si substrate damages, induced by the ESD, show exactly where the ESD current is heading, as indicated by white arrows. (Tung et al., IPFA, 2002, 65–69, reprint with permission from IEEE)

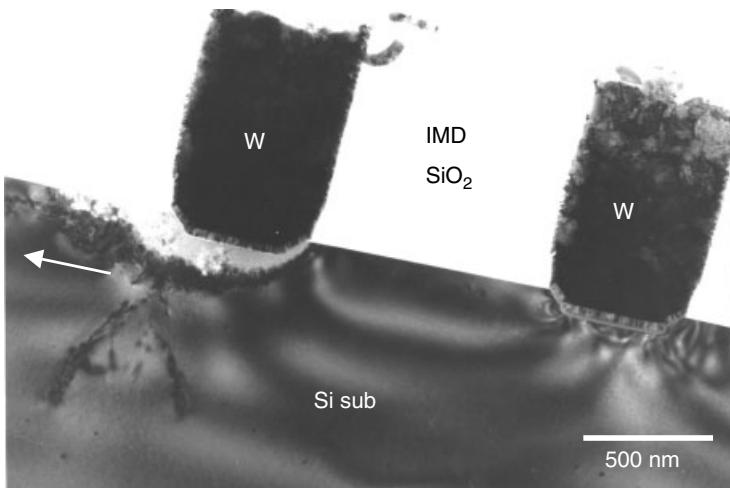


Figure 13.36 TEM cross section at the ESD damaged contact areas. Damages under the contact (left hand) spread toward left-hand corner of the image, as indicated. (Tung et al., IPFA 2002, 65–69, reprint with permission from IEEE)

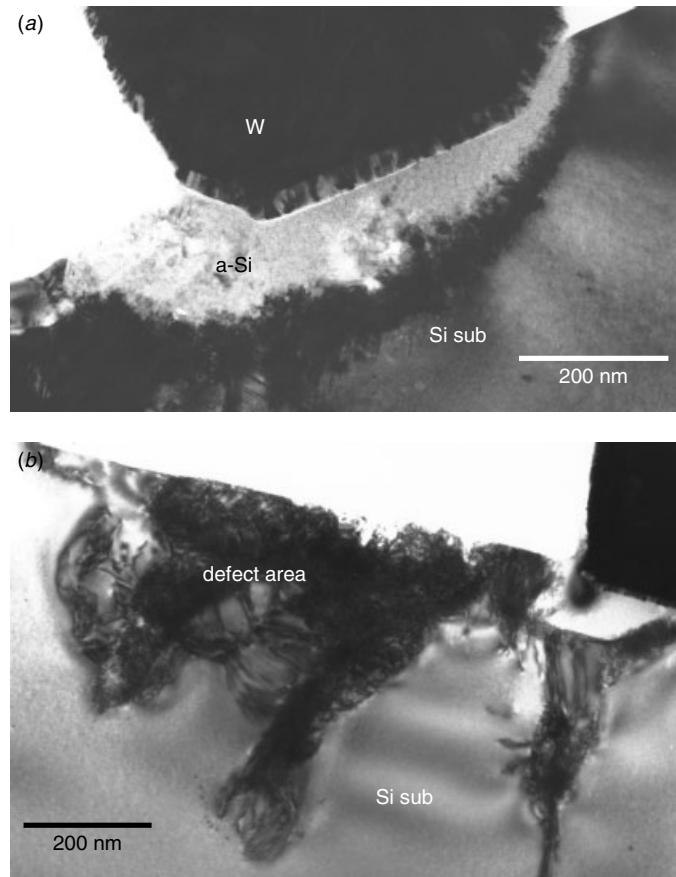


Figure 13.37 TEM cross section at the ESD damaged contact areas. (a) The Si substrate is totally amorphized at the contact bottom's interface, and has generated dense defects in the large area, as seen in (b). (Tung et al., IPFA 2002, 65–69, reprint with permission from IEEE)

Now removal of metal 1 and part of the polysilicon shows further damages (Fig. 13.41):

- Si substrate
- A gate oxide burn mark along the polysilicon lines
- ILD SiO_2 between the polysilicon line and a contact

The sample was next prepared for a plan view TEM. The plan view TEM reveals quite a lot of underlying damage:

- Si substrate damage under the field oxide resulting in $\text{Si}(111)$ rectangular facets, Fig. 13.41(a) and Fig. 13.43(a).
- Burn marks on Si substrate under the field oxide and a bird's beak, Figs. 13.41(b) and 13.43(a)
- Burn marks on the gate oxide along the polysilicon gate, Figs. 13.41(b) and 13.43(b)

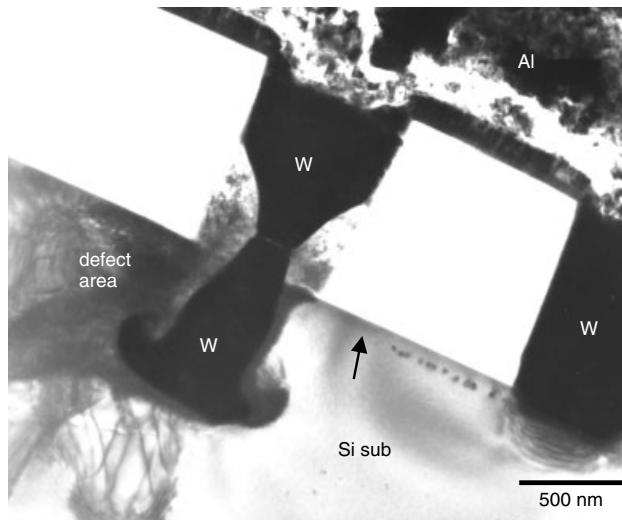


Figure 13.38 TEM cross section at the ESD damaged contact areas. The W–plug has smelted and plunged into the Si substrate. The damage is very similar to that of the EOS damage observed in the previous section. Notice that the implantation substrate defects have disappeared at the destroyed contact, as indicated. This may be due to local thermal annealing during the discharge. (Tung et al., IPFA 2002, 65–69, reprint with permission from IEEE)

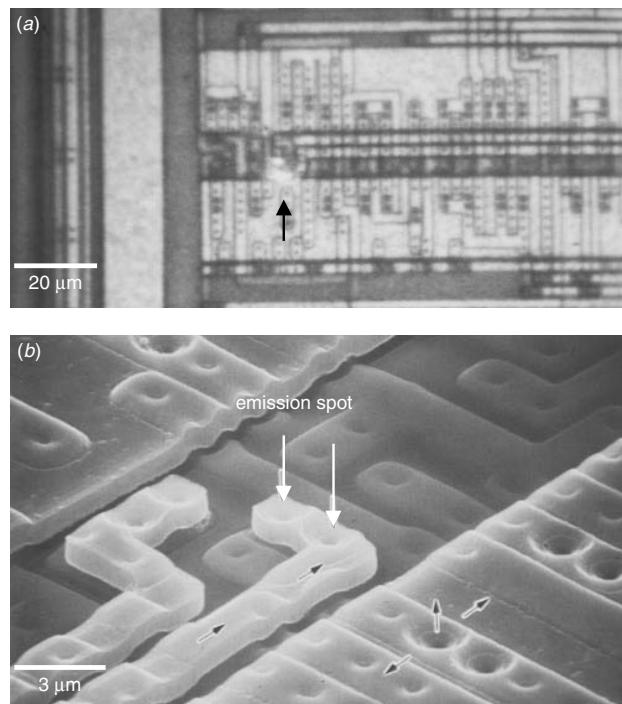


Figure 13.39 (a) Photon emission microscope image of an ESD induced leakage spot. (b) The corresponding SEM image at the emission spot after removal of passivation, which exposes the upper metallization layer. No apparent physical damage is observed on upper metal (metal 2) layer.

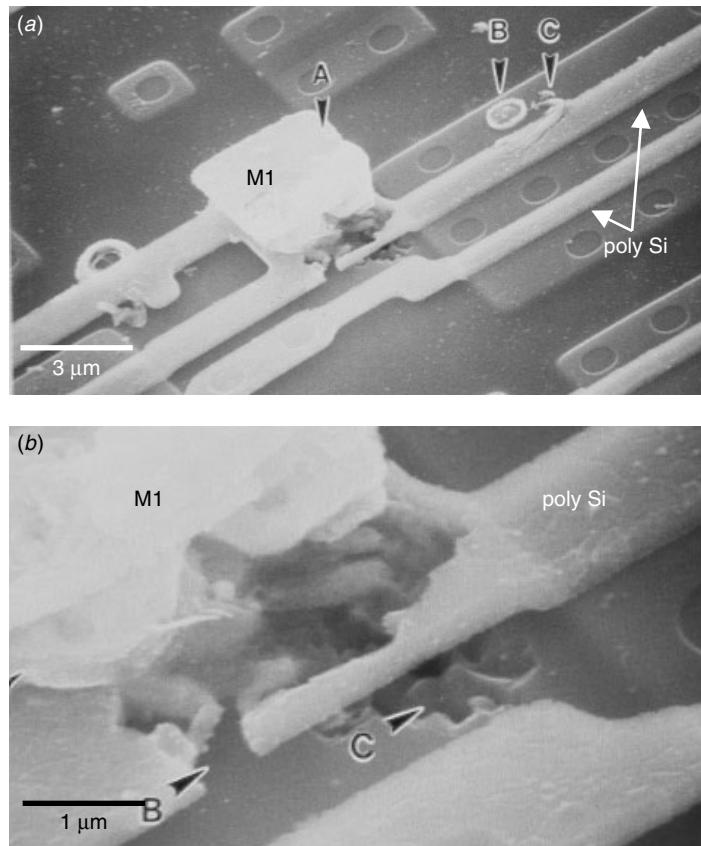


Figure 13.40 (a) Further removal of the IMD layer reveals metal 1 and poly Si damages at the emission's location. Arrows A, B, and C show the discrete damage locations. (b) The close-up SEM image at the metal 1 to poly Si damage shows that the discharge has gone all the way through the poly Si, arrow B, the field oxide, and into the Si substrate, arrow C.

- Polysilicon smelted and spilled out into ILD SiO_2 area, Figs. 13.43(b) and 13.44
- A burn contact indicated by a distinctive contact footprint, Figs. 13.43(b) and 13.44

The exact discharge path can now be traced explicitly, as shown in Fig. 13.45. Two distinctive discharge paths have contributed in parallel:

- The discharge current goes through metal 2, metal 1, polysilicon, and then penetrates the field oxide and discharges into the Si substrate, causing the Si substrate damages with $\text{Si}(111)$ facets.
- The discharge current also goes through metal 2, metal 1, polysilicon, and moves along the polysilicon, discharges through the gate oxide, causing the gate oxide break down, which leaves a zigzag burn mark on the gate oxide along the path.
- The same discharge path then continues along the polygate, punches through ILD SiO_2 and into the metal 1 contact. The contact is fried, leaving a Ti silicide

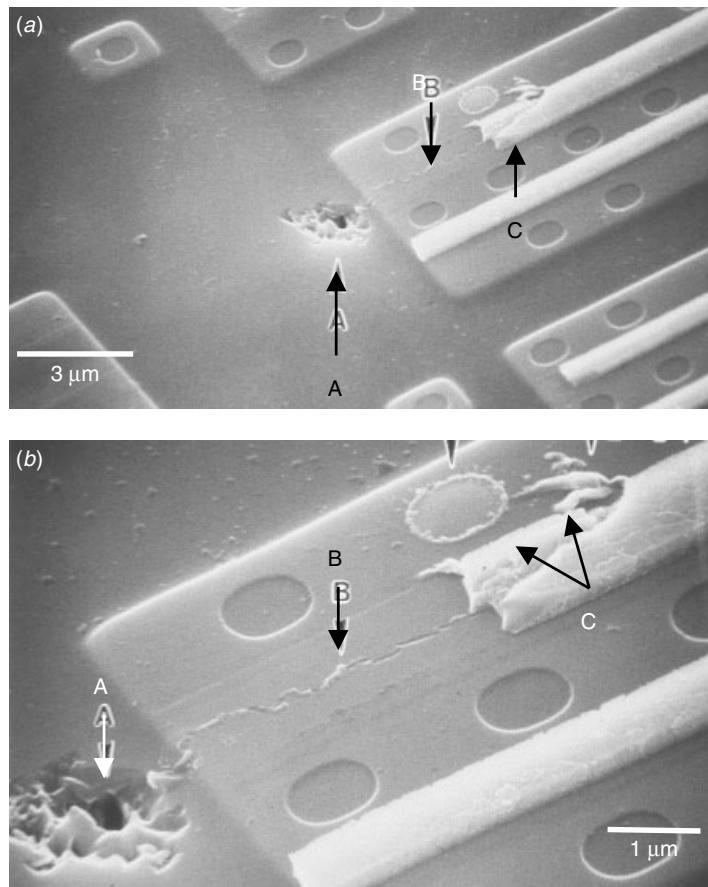


Figure 13.41 Further removal of the ILD, metal 1, and part of the poly Si. The three damage locations are clearly shown. A: substrate penetration; B: gate oxide punch through; C: polygate to contact dielectric breakdown. (Tung et al., IPFA 2002, 65–69, reprint with permission from IEEE)

imprint. The discharge current proceeds up to metal 2, which is connected to the device ground pad.

Not surprisingly, the two major discharge paths, to the Si substrate and to metal 2, that connect to the ground pad have similar resistivity impedance and therefore conduct the discharge power simultaneously.

Sometimes the discharge power may be subtle. Figures 13.46 and 13.47 show ESD discharge-induced polysilicon in the active area ILD breakdowns. Figure 13.46 shows that breakdown started at the field oxide's bird's beak. A polysilicon whisker has grown through the ILD SiO_2 dielectric and connects the active area to the polyrunner on the field oxide. Notice that there is no breakdown on top of the active area. The whisker nucleated from within the polysilicon line and grew out onto the bird's beak. The final punch through location is right at the bird's-beak tip area, where the oxide is weak and the field strongest. In a stronger current the whisker growth becomes

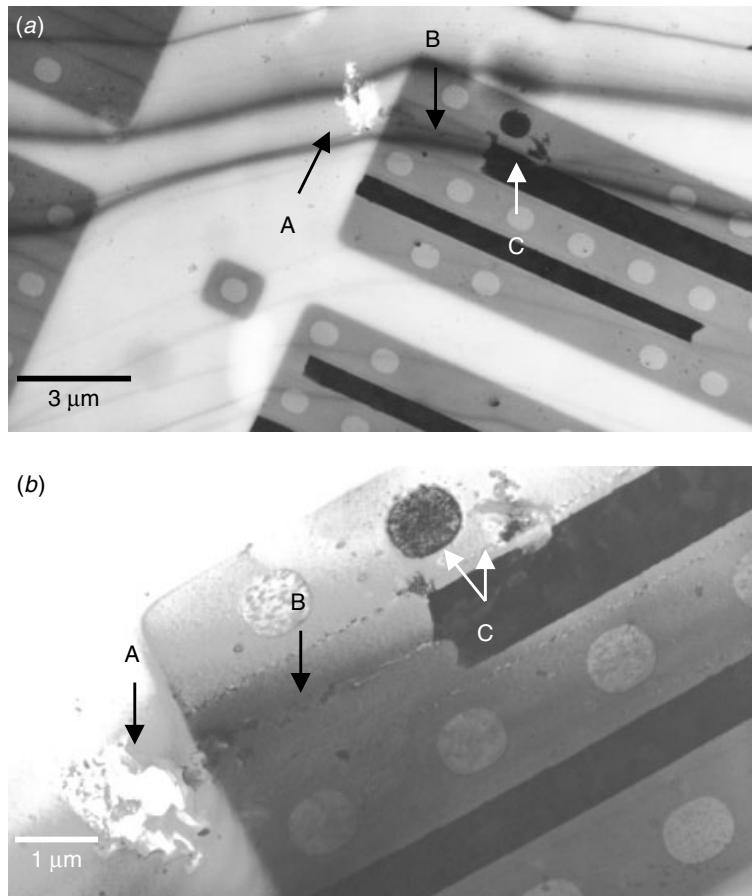


Figure 13.42 TEM plan view shows exactly what was observed in SEM as seen in Fig. 13.41. The three damage locations are clearly shown: *A*: substrate penetration; *B*: gate oxide breakdown; *C*: polygate to contact dielectric punch through.

massive and morphologically resembles lava flow, as seen in Fig. 13.47. It is apparent that the discharge is between the poly runner plate, area *A* in Fig. 13.47, and the active area, area *B* in Fig. 13.47. The poly lava growth gradually shifts from the polyrunner down to the polygate. Note that the poly lava growth could have been enhanced and exaggerated by the electrical tests performed to identify and locate the leakage spot. Further removal of the polysilicon and the field oxide reveals that the damage has crept deep into Si substrate. Figure 13.48 shows an SEM and the corresponding TEM image from the same sample as in Fig. 13.47. The damage is clearly concentrated along both sides of the polygate well into the field oxide's bird's-beak tip area. A close-up of the damage area, as seen in Fig. 13.49, shows that the ILD (or spacer oxide) fused to the Si substrate, forming fractal discharge pattern on the Si substrate near the polygate. Also observed in Fig. 13.49 is the exact location of bird's-beak tip damage. This is the location corresponding to the shortest conduction path in the high field due to the bird's-beak shape.

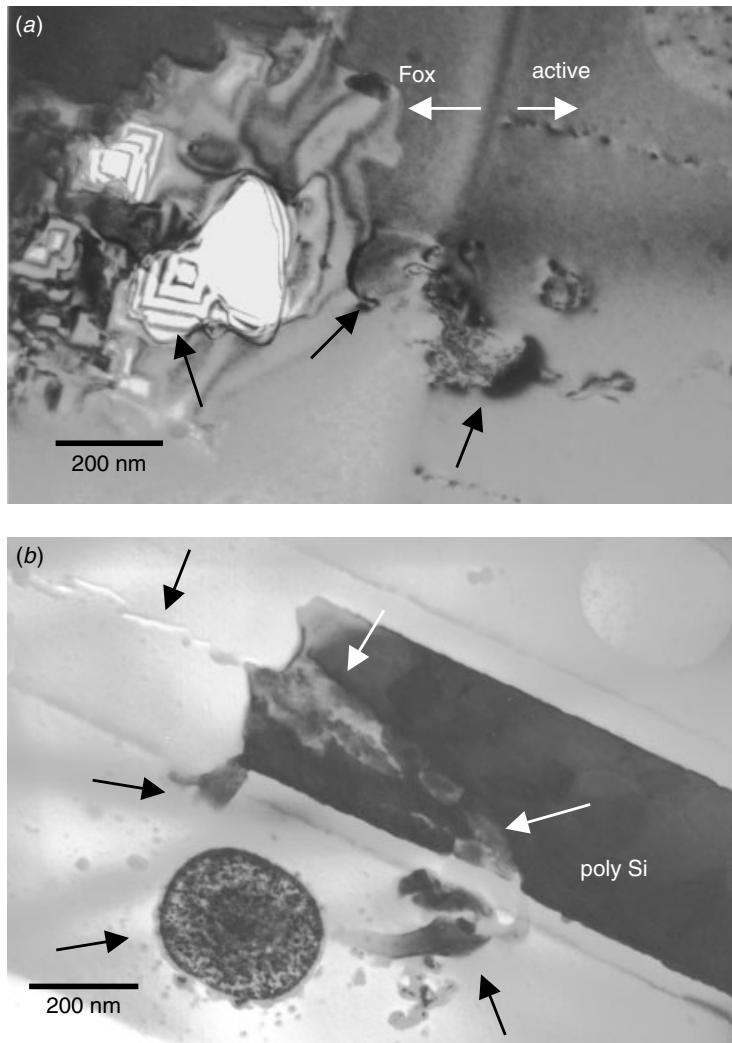


Figure 13.43 TEM plan view of the three damage locations. (a) The Si substrate penetration shows the Si(111) facet's rectangular contour. (b) The gate oxide breakdown occurs throughout the entire polysilicon line until the discharge punches through the gate to contact the ILD and discharge to the contact. (Tung et al., IPFA 2002, 65–69, reprint with permission from IEEE)

The gate oxide's edge usually is the location with the strong electrical field and low conduction resistivity (when overlapped with the s/d junction), so it is the place where most of the gate oxide's hard breakdown occurs. Figure 13.50 shows an ESD discharge-induced gate oxide edge breakdown. Both SEM and TEM plan view images show that the breakdown occurred right in between the polygate and the active area (on the drain side only). The damage spots are spread out as small cut marks on the Si substrate, seen in the Fig. 13.50. Figure 13.51 shows a schematic cross section of the gate structure and the ESD discharge path. Since such a breakdown mechanism is quite common, there has been a way to prevent the breakdown from happening. The

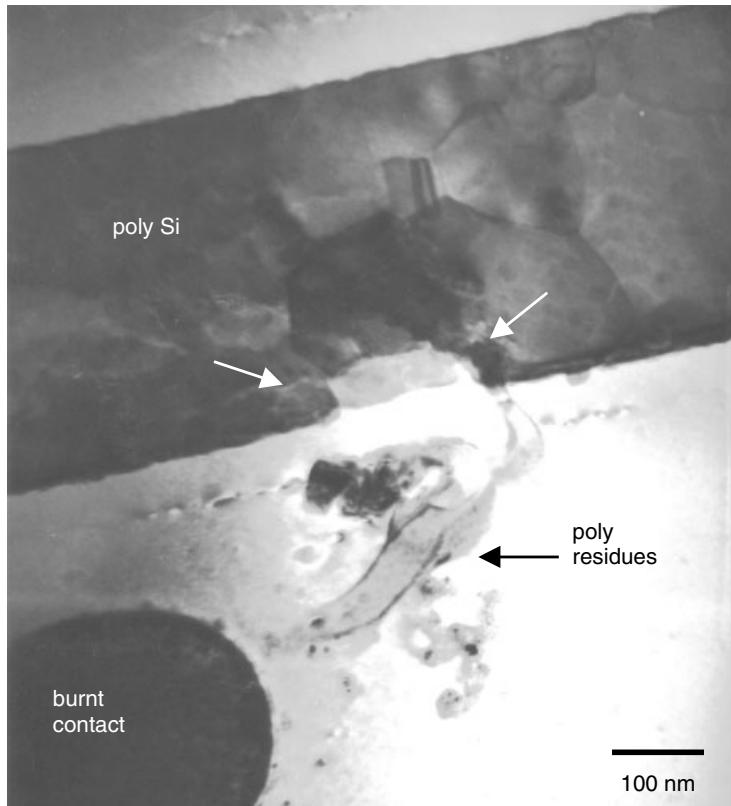


Figure 13.44 TEM plan view of the ILD punch-through location. The polysilicon was smelted and squeezed out into the ILD area, as indicated.

well-known approach is to improve the breakdown voltage by an extra ion implantation on the active area edge. This is usually called ESD implantation. However, the extra implantation can introduce more lattice defects. ESD implantation can cause severe mask edge defects, which then will evolve into extended dislocations in subsequent process steps. Figures 13.52 and 13.53 show ESD implantation induced mask edge defects that have evolved into extended dislocations. The dislocation lines can introduce junction leakage or at least greatly reduce the device capability to sustain the ESD surge stress.

The final example we give here is ESD damage upon the ESD protection circuit near the bonding pad. Most bonding pads are protected with an ESD protection circuit. If the protection circuit works properly, any ESD pulse coming in will be discharged properly by the protection circuit unless the incoming pulse exceeds a pre-designed level. That is to say, the ESD protection circuit will conduct all of the current/energy through a proper channel; if an incoming pulse exceeds the design, the ESD protection circuit will be damaged. A study of how such failure occurs is helpful in understanding the limits of the design, layout, and materials as each is used in the protection circuit and thus to help improve these factors. Figure 13.54 shows a low-magnification micrograph on an ESD discharge location near the pad area. The ESD destruction site is within the

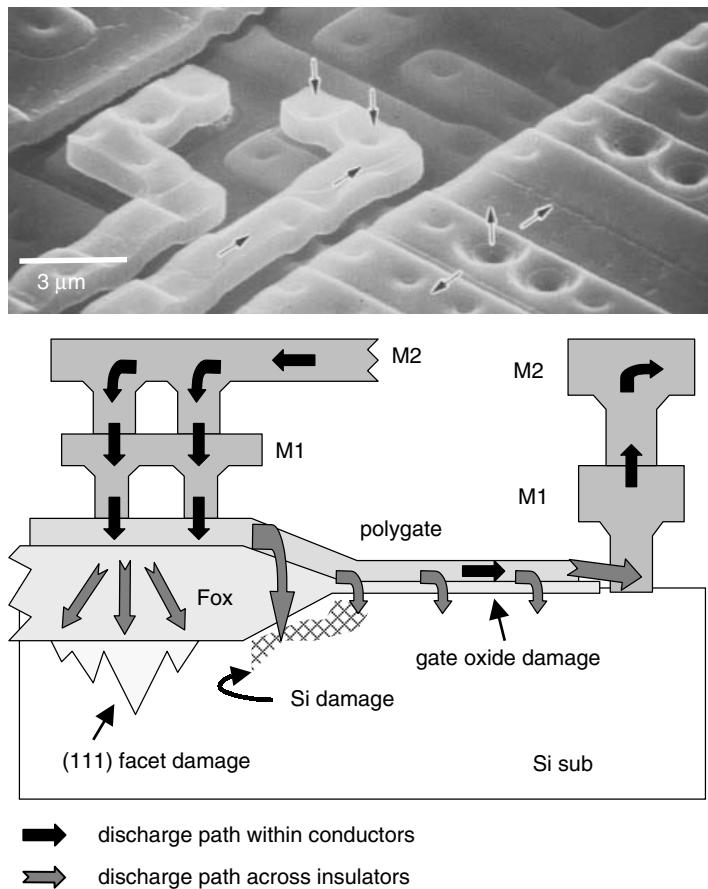


Figure 13.45 The discharge path can be exactly determined throughout the SEM and TEM analyses. The discharge partially goes into the Si substrate and partially comes back to metal 2 where the ground pad is connected. (Tung et al., IPFA 2002, 65–69, reprint with permission from IEEE)

ESD protection circuit transistor. The most burned location is a MOSFET transistor gate area, as shown in Figs. 13.55 and 13.56. To maximize the information obtained from such one of a kind sample, a carbon double-stage replica was used to replicate the burned location's surface topography and extract the surface materials. Figure 13.55 shows the original sample and the carbon replica samples side by side to indicate the one-to-one correspondence. A large piece of dark material, as seen in Fig. 13.55(b), was extracted from the burned surface's cavity. The EDS shows that this material is a mixture of fused SiO_2 , Si, Al, polyimide, and even the molding compound's epoxy materials.

The discharge current attempted to follow the gate and make a turn as the gate turned the corner. As a result the inner active area corner was burned away, and two or three spillover burning points were found at the outer active area, as seen in Fig. 13.56. Within the burning gate area, a few discharge spots were clearly observed, as in Fig. 13.57.

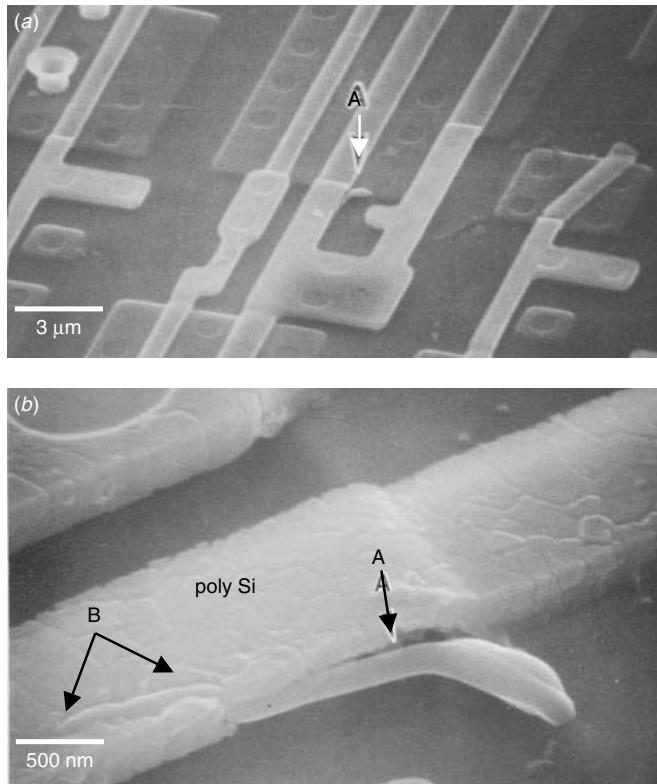


Figure 13.46 An ESD discharge-induced polysilicon to active area ILD breakdown near a field oxide bird's-beak area. Arrow *A* indicates the discharge-induced polysilicon whisker growth. Arrow *B* shows the whisker nucleated from within the polyline. (Tung et al., IPFA 2002, 65–69, reprint with permission from IEEE)

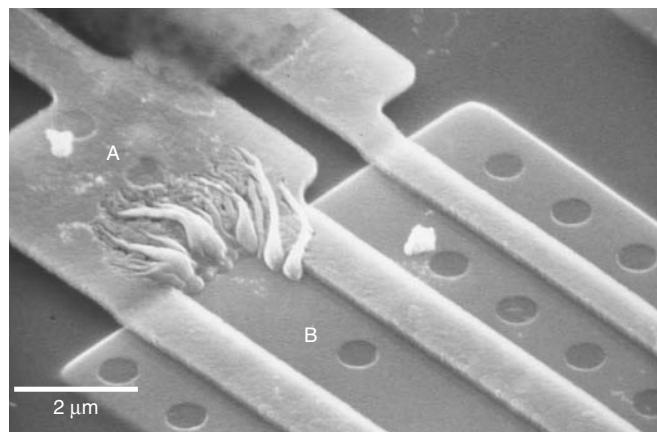


Figure 13.47 An ESD discharge-induced polysilicon to active area ILD breakdown near a field oxide bird's-beak area. Massive polysilicon migration and lava growth is observed compared to the effect in Fig. 13.46. (Tung et al., IPFA 2002, 65–69, reprint with permission from IEEE)

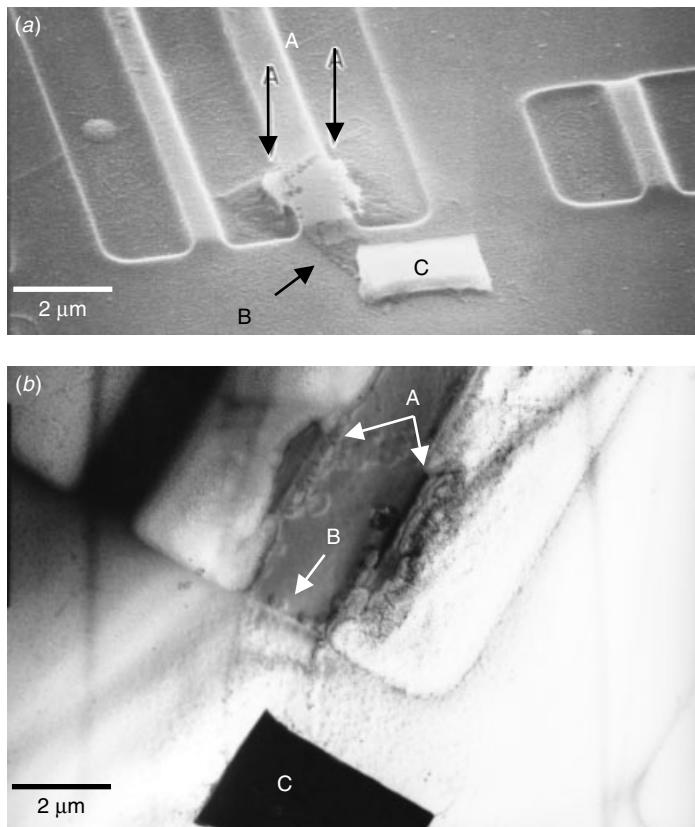


Figure 13.48 An ESD discharge-induced polysilicon to active area ILD breakdown near a field oxide bird's-beak area. Both SEM and TEM plan views show the breakdown at the polygate's two edges, arrow A, and field oxide bird's-beak tip, arrow B. Area C shows a burned and fused SiO₂ block that cannot be removed by the SiO₂ etch.

The evidence is that they gradually diminished with the largest cavity, which was created by the main discharge of energy. As it proceeds along the gate, the energy discharged diminishes quickly, leaving a trail of smaller cavities. In the heavy discharge areas, strange columnar structures are observed to have grown in perpendicular directions. Detail analysis shows these to be dendritic arms sprouting from the cavity's sidewall and bottom as they meet at the middle of the cavity, as seen in Fig. 13.58. Close examination of each discharge cavity reveals similar structures within them at smaller scales. The microstructure of these dendrite arms, as seen in Fig. 13.59, is seen to consist of inner hairline crosses at the middle, and resembling the eutectic solidification phase of a liquid phase within an area subjected to large thermal gradient. Amazingly the hairline dendrite arms all cross in the same direction. This suggests that they share the same growing substrate, and thus crystallographically they are the arms of one same crystal, the Si substrate. Both EDS and electron diffraction show the solidified dendrite arms to be a mixture of silicon and SiO₂ with a lot of metallic and organic impurities.

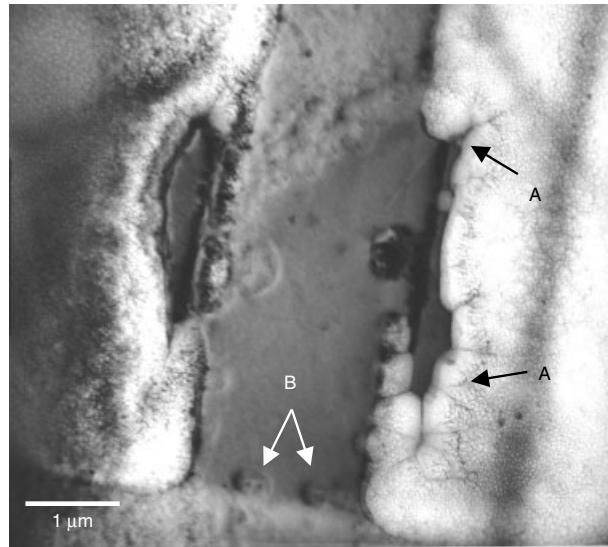


Figure 13.49 An ESD discharge-induced polysilicon to active area ILD breakdown. Total rupture of the ILD SiO_2 has squeezed the SiO_2 onto the Si substrate, forming fractal discharge patterns, as indicated by arrows A. The discharge on the bird's beak is shown by arrows B.

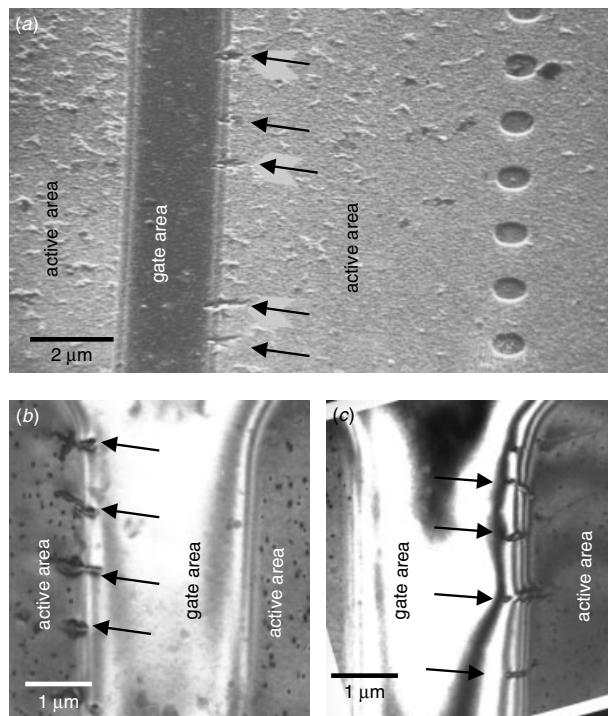


Figure 13.50 ESD discharge damages at the polygate's edge. The discharge occurs exactly between the polygate and the active area, leaving a damage mark on the Si substrate, as shown here.

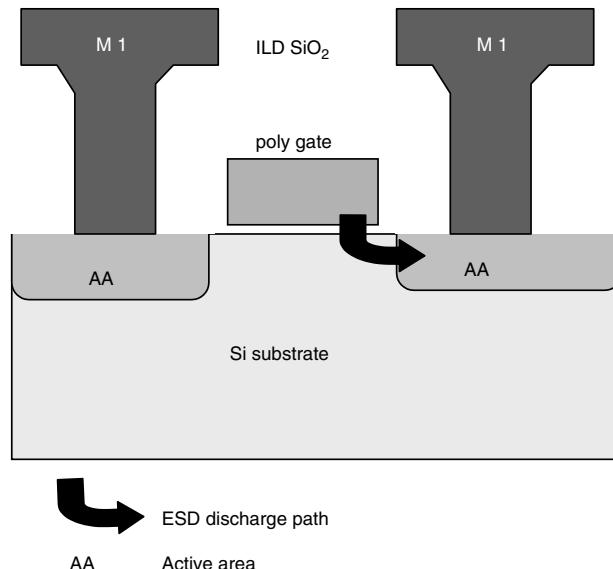


Figure 13.51 Schematic ESD discharge path over the gate oxide's corner breakdown.

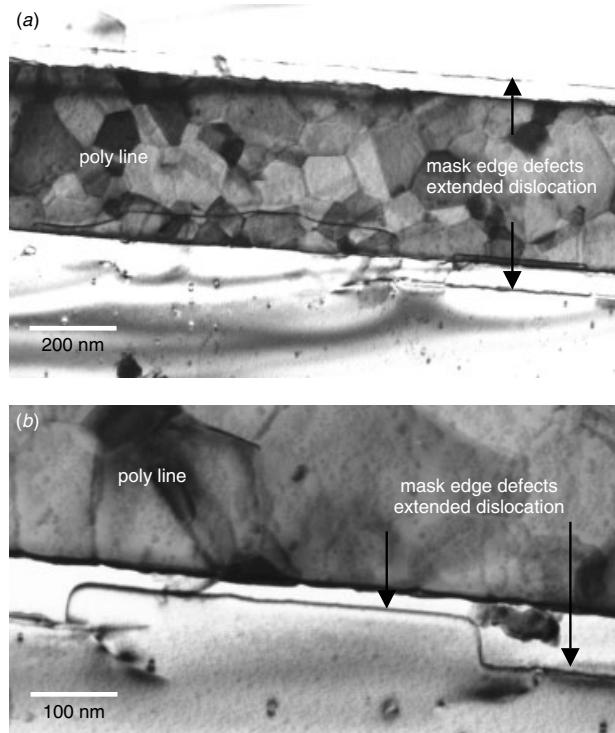


Figure 13.52 TEM plan view showing the ESD implantation-induced mask edge defect. The polygate is preserved for easy identification of the location of mask edge defects.

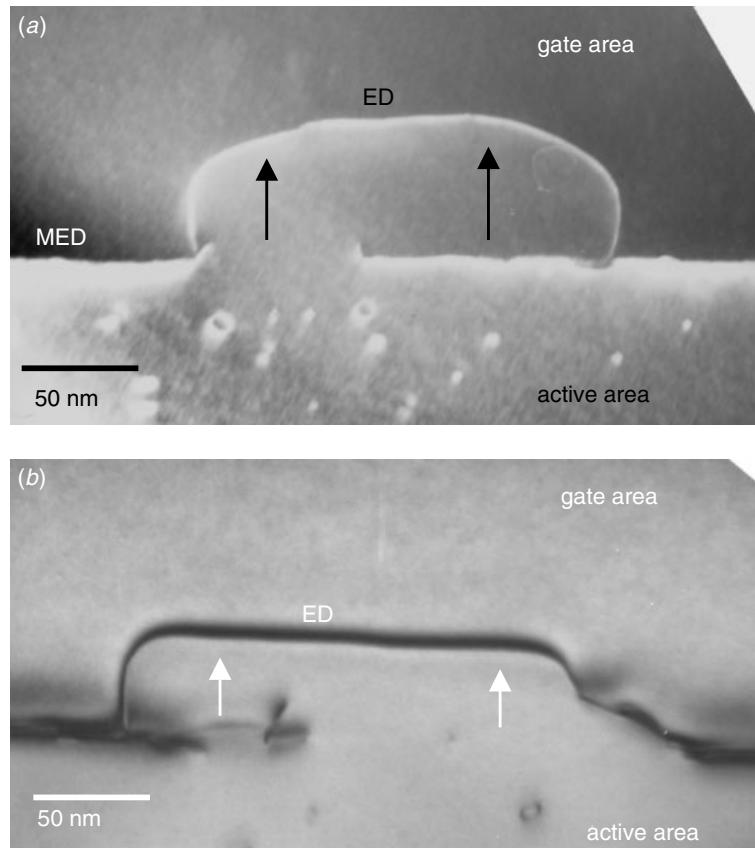


Figure 13.53 TEM plan view showing the ESD implantation-induced mask edge defect (MED). The mask edge defects act as a dislocation source and pump out the extended dislocations (ED).

13.6 THE DIFFERENCES BETWEEN EOS AND ESD-INDUCED DAMAGES

As we saw earlier, the failure modes and the differences observed in samples with EOS and ESD-induced damages are subtle, and thus the two are usually regarded as a single failure. However, TEM analysis, such as we showed in the examples above, clearly distinguishes the two failure modes (Tung et al., 2002). Instead of stopping the analysis at the circuit level, as most of the ESD/EOS failure analysis cases do, it is evident that the analysis of the failure mechanism must proceed at least to the device level. Polygate, gate dielectrics, contacts, isolations, bird's-beak corners, and so forth, all should be observed structurally down to their fullest details. Only this way can the differences between EOS and ESD damages be distinguished. The main failure mechanisms differences observed between EOS and ESD can be summarized as follows:

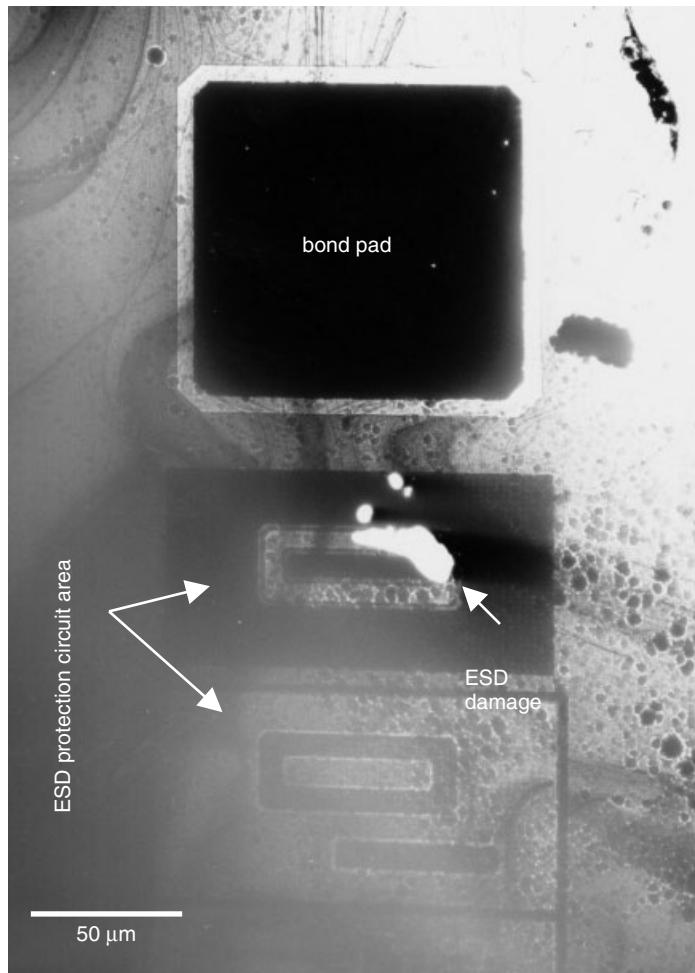


Figure 13.54 TEM plan view showing the ESD damage of a ESD protection circuit area near a bonding pad.

Discharge from EOS

- Usually a single hit
- Neat, clean, and concentrated
- Can be traced back to the point where smelting, fusing, and explosion all started

Discharge from ESD

- Usually with multiple hits
- Massy and intermixing

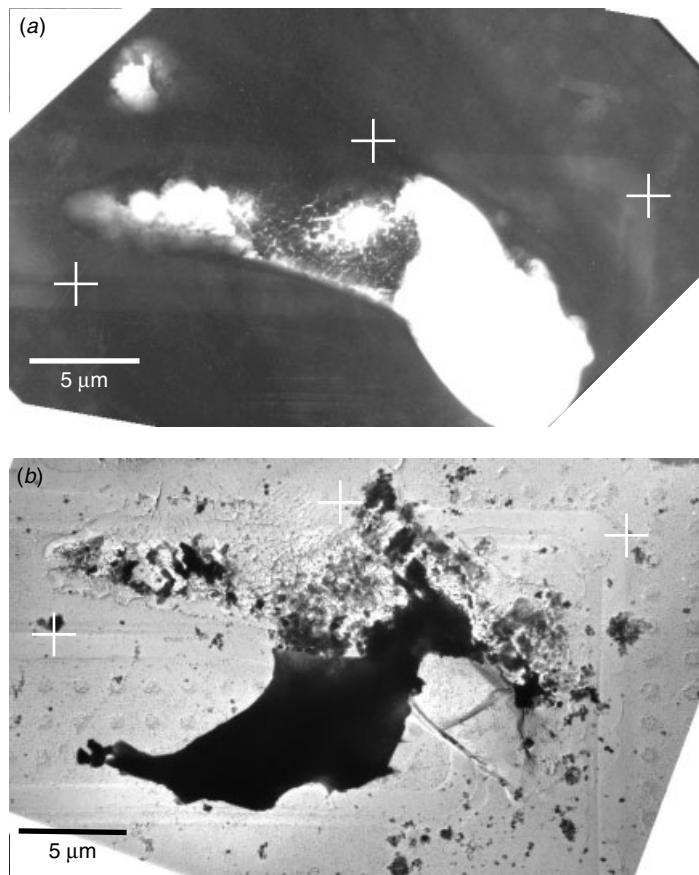


Figure 13.55 TEM plan view showing a damage spot. (a) TEM plan view where the ESD pulse has attempted to turn at the corner point, and (b) a carbon replica of the same area. The corresponding points are marked with crosses. The extracted materials on the carbon replica are helpful in determining the chemical and thermal reactions during the ESD discharge damage. (Sample courtesy by Dr. M. Natarajan & Dr. M. K. Radhakrishnan, IME, Singapore)

- Numerous Discrete hit points
- Difficult to trace back to a starting point, even though the path of discharge can be observed

The way to identify and distinguish an EOS damage from an ESD damage is by a complete and detailed failure mechanism study using SEM and TEM. Combined, these two analytical technologies not only reveal exactly the destruction morphology, the discharge path, and the circuit segment that absorbed most of the destructive energy, but also provide some important insights, for the design and process development teams, on the discharge characteristics of EOS and ESD. In particular, it is structural information such as this on device failure that leads to improvements in the next generation's devices.

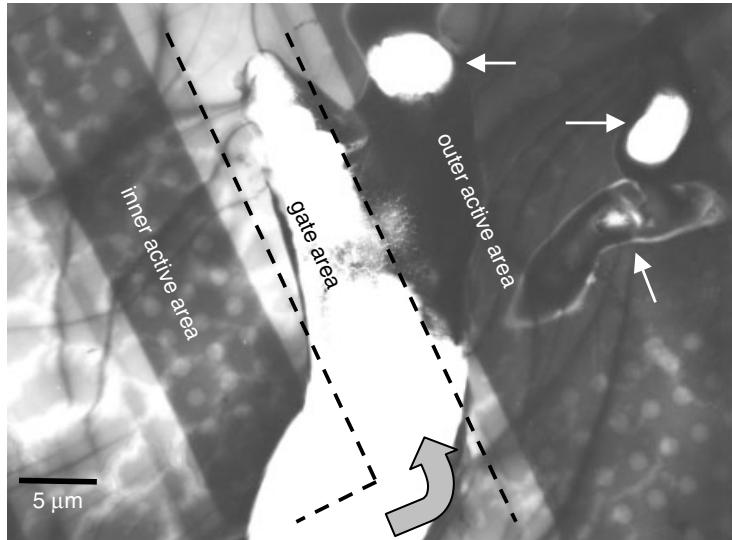


Figure 13.56 TEM plan view showing damage spots. The discharge pulse attempted to follow the gate to make a turn at the corner area but burned out the inner active area corner and spilled over into the outer active area, as indicated. (Sample courtesy Dr. M. Natarajan & Dr. M. K. Radhakrishnan, IME, Singapore)

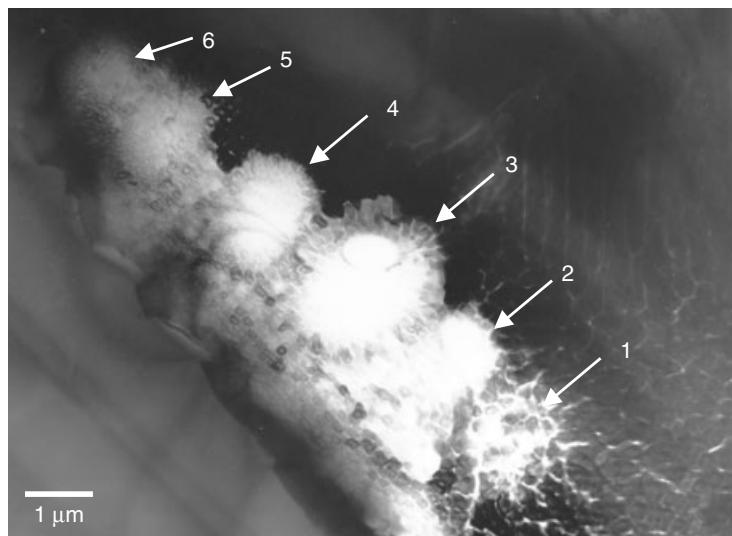


Figure 13.57 TEM plan view showing a detail of the microstructure where the main damage has occurred. At least six consecutive discharge centers can be identified. This implies, again, that the discharge current split into fractal patterns before it hit the Si substrate. (Sample courtesy Dr. M. Natarajan & Dr. M. K. Radhakrishnan, IME, Singapore)

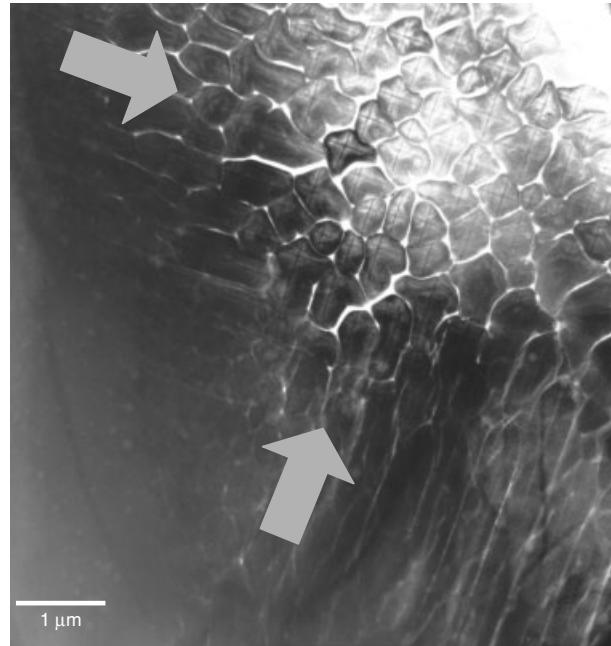


Figure 13.58 TEM plan view showing the detailed microstructure of the main damage area. Columnar dendritic arms are observed to grow from two equidistant perpendicular directions, as indicated. (Sample courtesy Dr. M. Natarajan & Dr. M. K. Radhakrishnan, IME, Singapore)

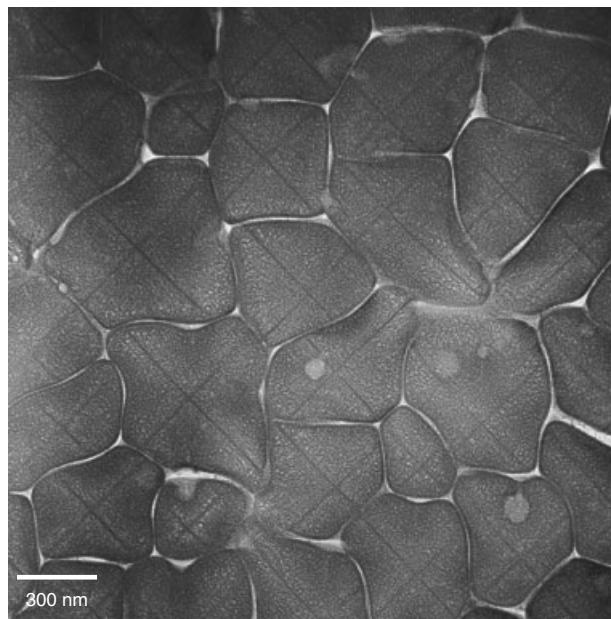
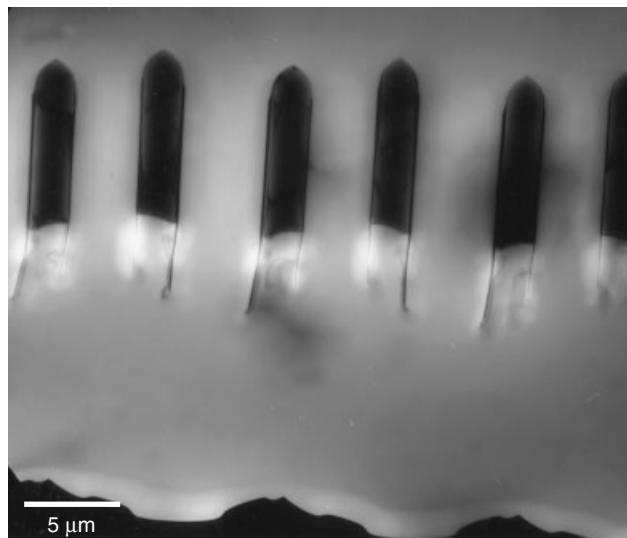


Figure 13.59 TEM plan view showing a detail of the dendritic arms' microstructure. Each column shows distinctive internal hairline crosses. In some cross fibers, dendrite arms has begun to develop. (Sample courtesy Dr. M. Natarajan & Dr. M. K. Radhakrishnan, IME, Singapore)

REFERENCES

- C. F. Hawkins and J. M. Soden, *Microelectronics Failure Analysis Desk Reference*, 4th ed., edited by R. J. Ross, C. Boit and D. Staab, Electronics Device Failure Analysis Society (EDFAS)/The Materials Information Society, 109–120, 2001.
- L. G. Henry, *Microelectronics Failure Analysis Desk Reference*, 4th ed., edited by R. J. Ross, C. Boit and D. Staab, Electronics Device Failure Analysis Society (EDFAS)/The Materials Information Society, 421–436, 2001.
- V. B. Menon, *Handbook of Semiconductor Manufacturing Technology*, edited by Y. Nishi and R. Doering, Marcel Dekker, 869–888, 2000.
- C. H. Tung, C. K. Cheng, M. K. Radhakrishnan, and N. M. Iyer, *Proc. 9th Int. Symp. Physical and Failure Analysis of Integrated Circuits (IPFA)*, 65–69, 2002.

14 Novel Devices and Materials



MEMS accelerometer sensor device (after releasing etching). Cross section TEM resembles Rockets take off.

Most of the semiconductor commodity market is dominated by MOSFET products, which include logic devices and memories. Nevertheless, semiconductor products, like any other commodities, are driven by market demand. The market is dynamic and changes daily. In the early 1980s DRAM was the technology driver, partly because it dominated the semiconductor memory market. As the device shrank to the critical size where its performance began to depend on the interconnect RC delay, the technology began to diverge toward logic and DRAM. Cu interconnects and low- k intermetal dielectrics (IMDs) became the focal point of research and development.

There are, however, other categories of semiconductor devices, where important special functions and market were fulfilled. These devices are not considered main stream because of their small market share. However, the technological challenges have been equally thrilling and with exciting results. This chapter presents some examples of these novel devices (and materials) in different special market niches. Most are undergoing further research and development. Some may eventually replace the current main stream technology.

14.1 FLASH MEMORIES

Semiconductor memory can be broadly classified into two categories: volatile and nonvolatile. Volatile memories such as DRAM and SRAM lose their stored information once the power supply is switched off. Nonvolatile memories, on the other hand, can retain their stored information. The nonvolatile memories include ROM (read-only memory), PROM (programmable ROM), floating gate memory, FRAM (ferroelectric RAM), and MRAM (magnetoresistive RAM). As explained by Sze (1999), for all the characteristics and performance that a semiconductor memory needs to fulfill, the memory that has the most desirable attributes is the flash memory in the floating-gate memory family.

The floating-gate concept was first proposed by Kahng and Sze in 1967. This was the very first time to recognize the possibility of rewritable nonvolatile semiconductor memory (NVSM). Figure 14.1 shows a cross-sectional view of a typical NVSM floating-gate structure. This simple stacked gate structure can perform programming (e.g., Fowler-Nordheim tunneling), storage, and erase (e.g., reverse FN tunneling). Several device structures were proposed in the early EPROM (erasable programmable ROM), for example, FAMOS (floating-gate avalanche injection MOS), and SAMOS (stacked-gate avalanche injection MOS) by Frohman-Bentchkowsky (1971) and Iizuka et al. (1976), respectively. SAMOS is an EEPROM (electrically erasable programmable ROM, or sometime called E²PROM) and for a typical E²PROM operation, we need two devices per cell, namely an E²PROM and a selection MOSFET. Therefore the cell size is relatively large. Masuoko et al. (1984) proposed a structure in the erase operation whereby the whole block is erased, thus the term “flash memory” was used. This particular device has only one device per cell and has the advantages of higher density, lower cost, and higher scalability compared to the conventional E²PROM.

A flash cell is an n-channel transistor whose threshold can change under certain conditions. Figure 14.1 shows a schematic cross section of a flash “stacked-gate” cell and the corresponding TEM image. The upper gate is called the “control gate,” and it behaves just like an ordinary NMOS gate. The lower gate is called the “floating gate,” as it is floated and not connected anywhere. When the floating gate is negatively charged, the overall transistor gets a higher threshold, and this is how a “bit” is stored in the cell. The flash bit-cell operation can be summarized as follows (Golla and Ghezzi 1998):

- In reading the cell, a constant voltage is supplied to the control gate and looks for the current drained by the cell.

If the current drain is low (or zero), the cell has a high threshold, and the bit is identified as a logic “0.” Thus we read a bit with “0.”

If a current exists or if there is a high current, the cell has a low threshold and is identified as a logic “1.” Thus we read a bit with “1.”

- In programming (or writing) the cell, the floating gate is charged by electrons, so the voltage threshold is increased (in writing a “0” into the bit). The bias is set up in such a way that the transistor source is grounded. The drain is at a voltage, which can attract an electron from the source side with enough speed. The control gate is biased to a voltage to get an appropriate electric field that attracts the fast electron from the drain side. The acceleration in the channel is obtained by the lateral electric field. The higher the field, the more electrons will have enough energy to climb the oxide’s potential barrier. This is known as “hot electron injection.”

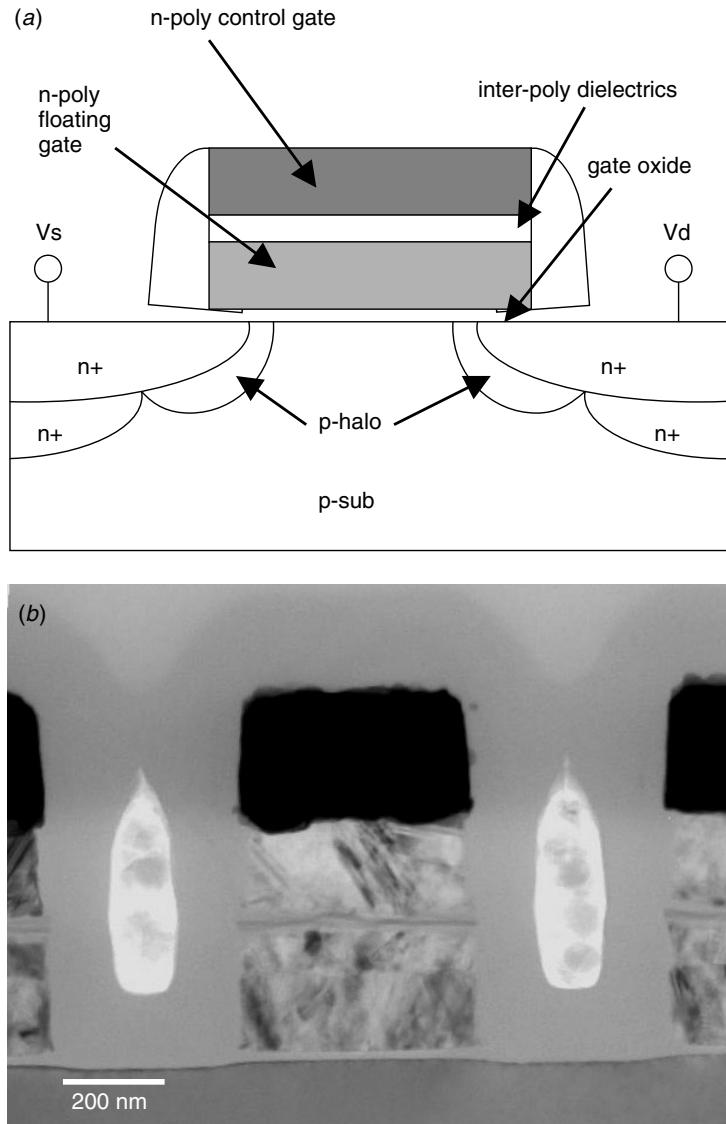


Figure 14.1 Typical flash memory device with floating gate, intergate dielectrics, and control gate stacking together. (a) P-halo in the substrate used to increase the generation of channel-induced substrate electrons. (b) TEM cross section of a typical device structure.

- In erasing the cell, the electrons in the floating gate are extracted back to the control gate using Fowler-Nordheim tunneling. The dielectric of a MOS capacitor becomes conductive when a high field is applied. Electrons in the floating gate are able to pass (tunnel) through the potential barrier. The problem with this process is that it is very slow compared to the writing process. To overcome the slowness, a great number of cells are erased together (this is why the device called flash memory).

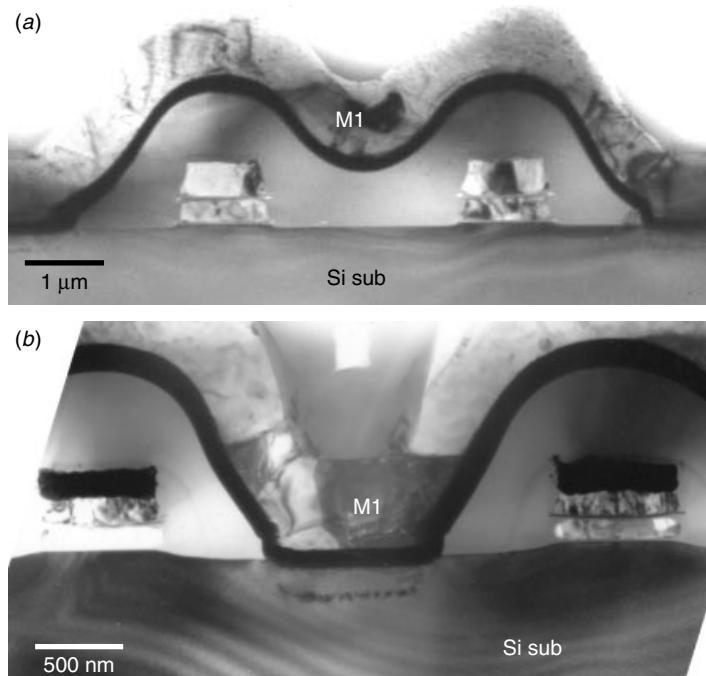


Figure 14.2 Two early flash memory structures. (a) Using normal polysilicon as floating and control gates and SiO_2 as interpoly dielectric. (b) Later version using W-polycide as the control gate and ONO as interpoly dielectric. Both use Al as the metal and contacts.

Figures 14.2 and 14.3 show some examples of early flash memory cell structures. Other than the stacked gate structures, only two metal layers were required and the device process was very simple. Metal 1 was usually used as the bit line. SiO_2 and ONO were usually used as the interpoly dielectric. Conventional thermal silicon dioxide served as the tunnel gate dielectric. Figure 14.3 shows the device after junction delineation and the junction profile at and near the gates. Notice that the shortest channel length between source and drain is not at the Si substrate surface but at a position slightly lower, at about one third of the junction's depth. Also notice that the contact junction is always deeper. A typical feature of early process was that the contacts were re-implanted after the contact opening etch to reduce the contact resistivity. Figure 14.4 shows that the Al contact was done by an additional contact etch step and a hot Al process to improve the contact yield as the contact size was further reduced and higher contact aspect ratio was needed. Further development on the contact aspect ratio required the W-plug process as shown in Fig. 14.5. Noticed that from Figs. 14.2 to 14.5, the flash memory cell stacked-gate structure has not changed much except for the shrinkage in dimension. The junction delineation on a later version of flash memory is shown in Fig. 14.6. W-plugs now replace the conventional Al contacts and CMP is used for planarization.

Compared with the DRAM and SRAM, flash memory certainly has a much simpler device structure and process. These features directly translate into high scalability and low process cost, and so make flash memories a very attractive alternative as a potential

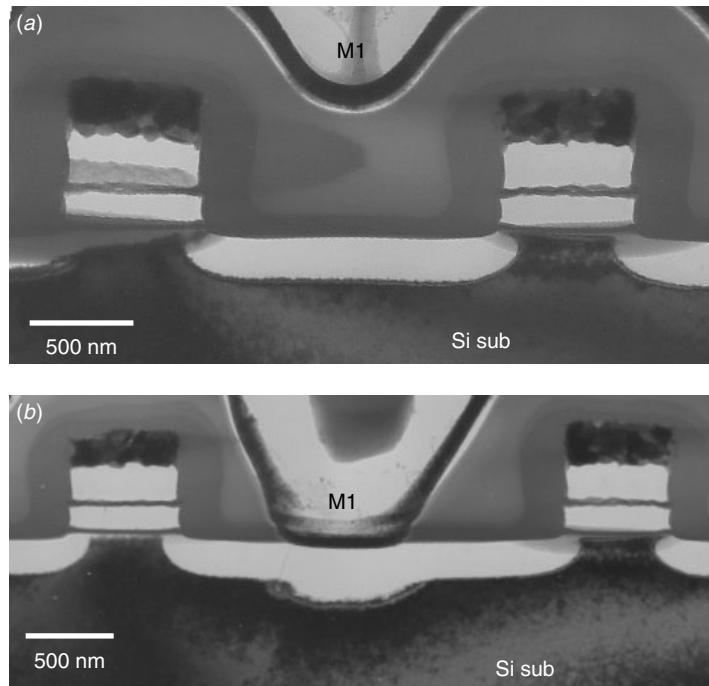


Figure 14.3 (a) Junction delineation reveals the junction profiles in substrate. (b) Notice that the contact junction is deeper due to the contact implantation in the early process to reduce the resistivity.

replacement for DRAMs and SRAMs. Further the possibility was demonstrated that more than one bit could be stored per cell. This allowed flash memories to be lower priced per bit than DRAMs, which made them very attractive for massive storage applications in portable personal computers (PCs) and mobile phones. However, as we noted above, the writing and programming operations involved both Fowler-Nordheim tunneling and hot carriers injection. Program/erase cycling endurance had to be achieved without degrading the data retention. The single-transistor structure exposed memory cells to array disturbances and to overerasure problems. A major yield and reliability concern with flash memories was the quality of the tunnel oxide, in terms of intrinsic properties and defect density. All these issues made flash process technology one of the most difficult to master, requiring very accurate process optimization and tight process control (Cappelletti 1998). This is why flash memories have not as yet taken over the DRAMs/SRAMs in the memory market.

14.2 SILICON-ON-INSULATOR (SOI) TECHNOLOGY

Silicon-on-insulator (SOI) technology was originally invented for a radiation-hard circuits niche. For designers and process engineers who are familiar with bulk-Si technology, SOI is not an entirely different technology. SOI is a special substrate on which the devices are constructed. CMOS circuits, bipolar transistors, and high-voltage

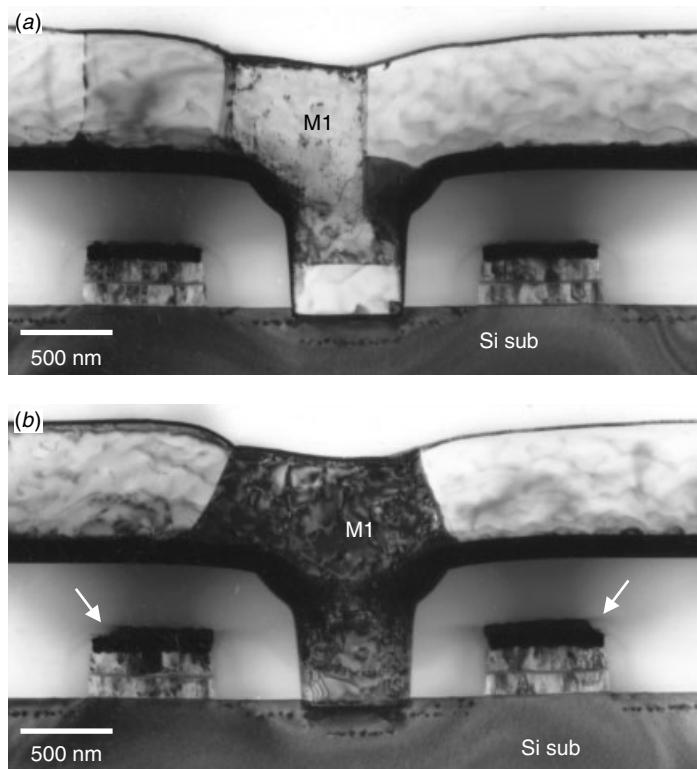


Figure 14.4 (a) An early hot Al M1 and contact used for the flash memory bit-line contact. (b) The same process but a slight overetched notch is observed on the polycide's top corners as indicated.

devices, for example, can all be built on an SOI substrate to achieve certain functions that otherwise cannot be achieved by bulk-Si technology. The current dominant SOI technology is called SIMOX (Separation by IMplantation of OXYgen). In which a Buried OXide (BOX) is synthesized by internal oxidation during the deep implantation of the oxygen ions into an Si wafer and followed by annealing at high temperature, 1320°C, for six hours as is necessary to recover a suitable crystalline quality of the Si film on top. The high-current implantation is conceivable for 8 in wafers with good thickness uniformity, low defect density, sharp Si–SiO₂ interface, robust BOX, and high carrier mobility (Cristoloveanu 1991). There are other technologies available for creating the SOI structure, the interested reader should refer to the references (e.g., listed in Cristoloveanu 2000).

The key difference between bulk-Si and SOI technologies is the oxide layer buried within the Si substrate. Thus our discussion of SOI will focus on the design, generation, and the properties of this oxide layer. Figure 14.7 shows an early SIMOX-generated oxide layer. Distinctive Si crystalline precipitation particles are observed embedded within the SiO₂ layer. Two different types of particles are present. The large spherical single crystal particles with internal twins are nearly the same in size as the buried oxide (BOX) layer; some are even attached to the Si on top or below, as seen in

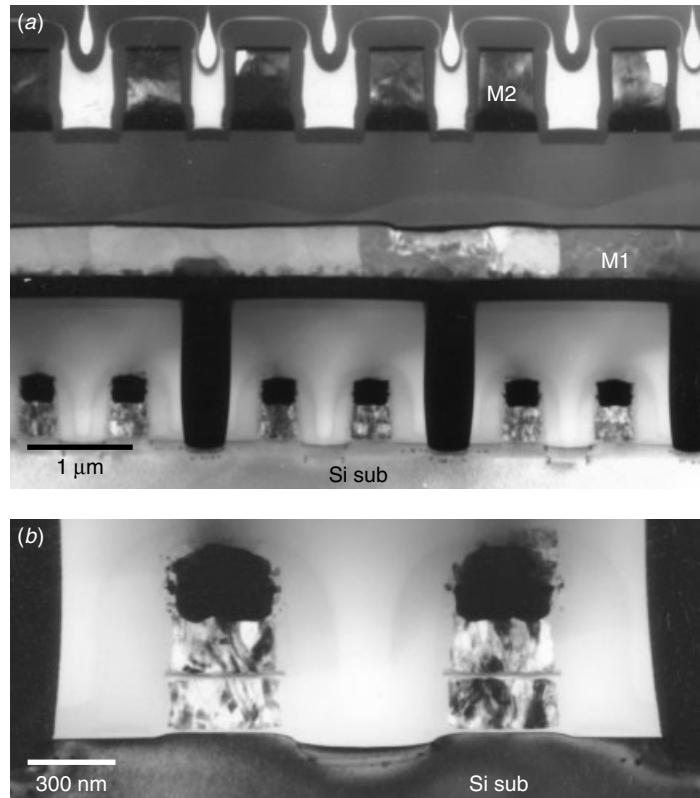


Figure 14.5 Close-up of the (a) cell structure and (b) gate structure of a flash memory. The floating gate is made of polysilicon, while the control gate is made of W-polycide. ONO is the interpoly dielectric and oxide the usual gate dielectric.

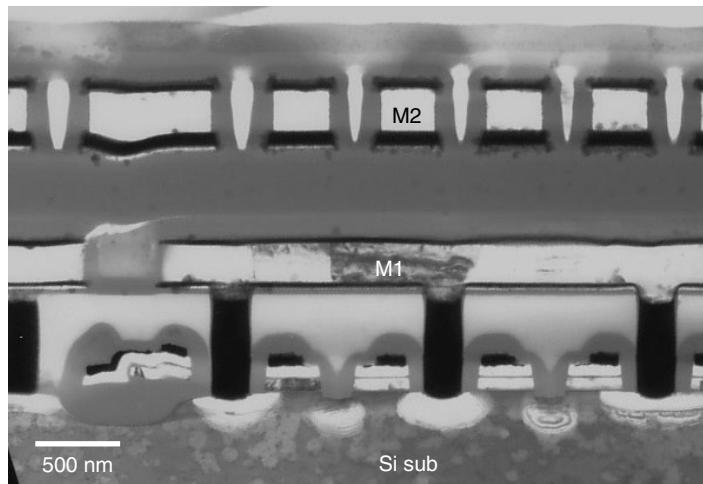


Figure 14.6 Later version of the flash memory device with the junction delineation revealing the junction structures. Again, the shortest channel length is slightly below the surface.

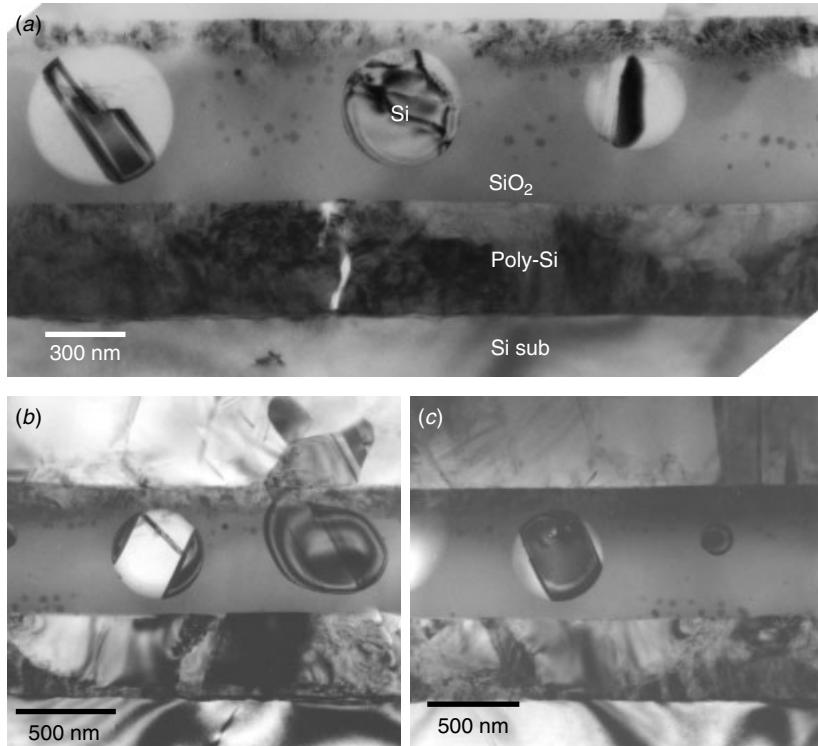


Figure 14.7 An early SIMOX (separation by implantation of oxygen) process-induced poly-Si layer and SiO_2 layer with the single crystal Si precipitation “balls.” The micro-twins within the Si balls show distinctive crystalline contrast.

Fig. 14.7(b). The small precipitation dispersed in between the large particles is the other category. Studies have found that with sufficient time and high enough temperature, all of the small precipitation will eventually be consumed and agglomerated into the large particles. The older SIMOX layers are of poor quality and cannot be used to build device applications. Improvements are needed in the SIMOX implantation and oxygen-rich annealing to enhance the BOX integrity. The typical SIMOX process involves a $1.8 \times 10^{18} \text{ O}^+/\text{cm}^2$ implant dose at 190 KeV and 650°C, which obtains a 0.2 and 0.4 μm thick BOX layer; a low dose $4 \times 10^{17} \text{ O}^+/\text{cm}^2$ and an additional oxygen-rich anneal for enhanced BOX integrity yield a 0.1 μm thick BOX layer (Cristoloveanu 2000).

Figure 14.8 shows a late SIMOX layer with much improved BOX layer quality. After partial annealing, dispersed small Si crystalline particles are observed and some are attached to the Si substrate’s surface. Note that the upper Si layer is polycrystalline, Fig. 14.8(a). When fully annealed, the remaining Si particles are concentrated into a layer near the Si substrate and the upper Si is now fully re-crystallized into a single crystal layer Fig. 14.8(b). Figure 14.9 shows at lower magnification the fully annealed BOX. The Si particles are visible mostly near the lower Si substrate’s interface, but scattered particles can be found near the upper Si interface, Fig. 14.9(a). Note in Fig. 14.9(b) that most of the particles are in dark contrast, indicating that they are in

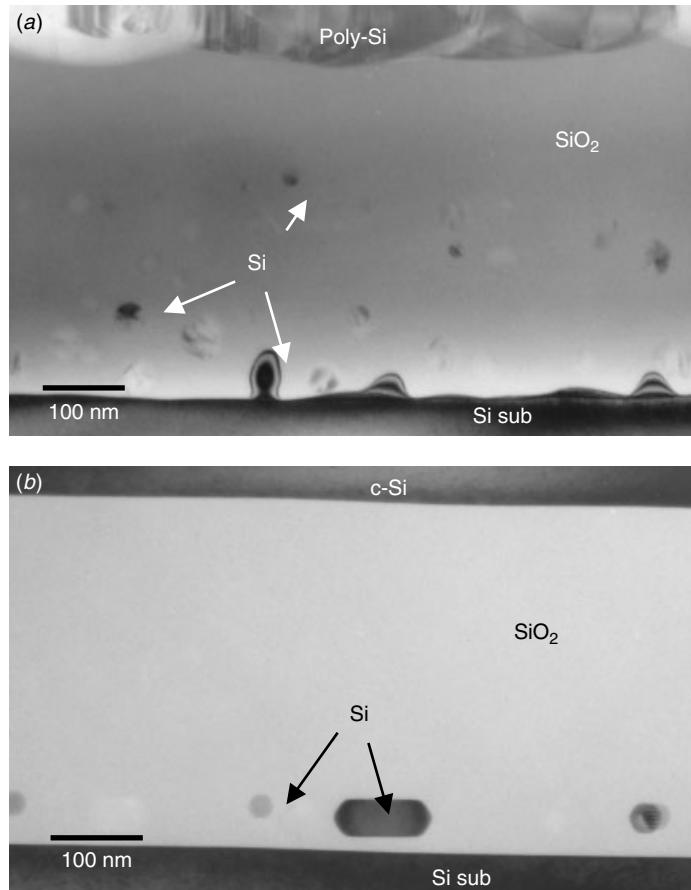


Figure 14.8 Improvement in the SiO₂ layer's quality and SiO₂–Si substrate interface. (a) Partial annealing shows the small Si crystalline particles scattered within the SiO₂ layer and Si bumps attached to the Si substrate. (b) Full annealing, a layer of the Si particle near the Si substrate is observed but no scattering Si particle or Si bumps. This is an apparent improvement compared to (a).

the same orientation as the Si substrate. Indeed, a detailed HRTEM and an electron diffraction analysis show these particles to be in epitaxial orientation to the Si substrate, Fig. 14.10. Removal of these interface particles has remained a challenge until recently, when a clean BOX layer free from particles and smooth Si/SiO₂ interface was achieved, as seen in Fig. 14.11. The HRTEM on the upper interface of this new BOX layer appears nearly atomically flat at the upper Si/SiO₂ interface, which is important as the device characteristics will be built on top of this layer.

Other than the SOI substrate, the devices built on the SOI substrate show no substantial differences from the conventional devices. This includes CMOS, bipolar, and some high-voltage devices. Submicron devices have been repeatedly demonstrated using SOI technology (Kikuchi et al. 1995; Adan et al. 1998; Voldman et al. 1999; Bagchi et al. 1999; Pae et al. 1999; Ito et al. 2000; Assaderaghi and Shahidi 2000). Figure 14.12

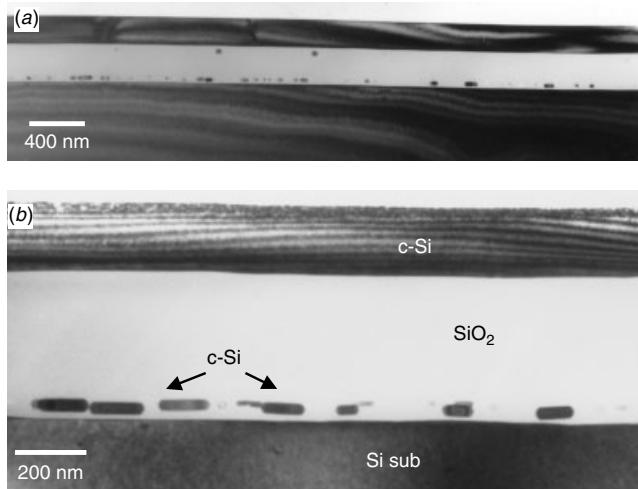


Figure 14.9 SIMOX technology creates an embedded SiO_2 (about 400 nm thick) with smooth upper and lower interfaces. A layer of crystalline Si disks is formed near the Si substrate. Another layer of crystalline Si particles, but with much lower density, also forms near the upper surface as shown in (a).

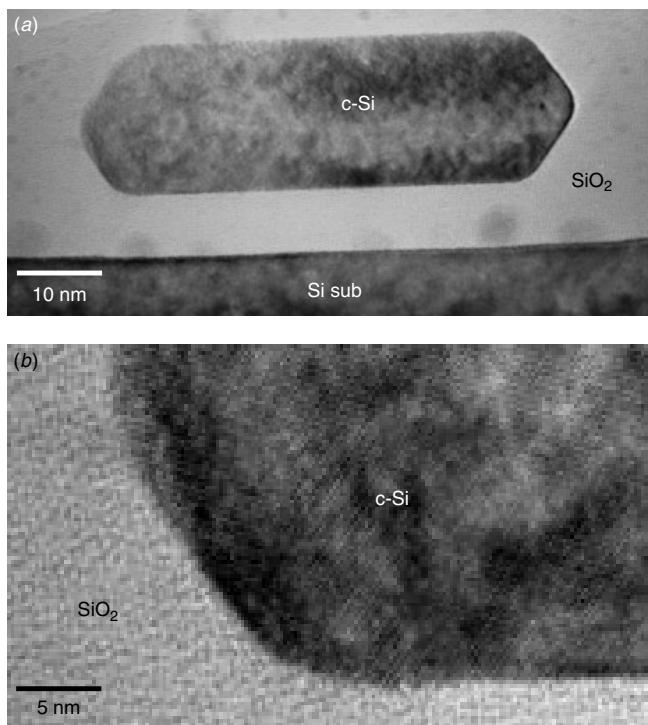


Figure 14.10 HRTEM of the crystalline Si particles near the Si substrate showing faceted surface morphology along the Si(111) lattice planes and the epitaxy nature of these particles. Their crystalline orientation matches exactly that of the Si substrate, yet they are not physically attached to Si substrate.

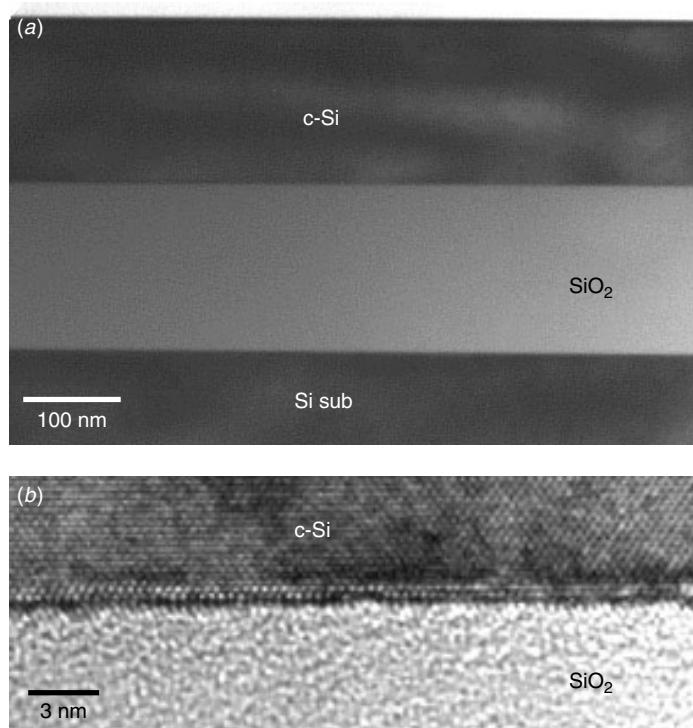


Figure 14.11 (a) The latest SIMOX technology with a defect-free buried SiO₂ (BOX) layer. No Si particles are observed. The upper and lower interfaces are nearly atomic in smoothness. (b) An HRTEM image of the upper interface.

shows some early SOI technology device TEM cross sections. Notice that the active devices are separated as oxide isolation areas. In these areas there supposedly are either LOCOS or STI structures in the bulk-Si technology. But now the isolation oxide areas are replaced by the BOX layer. Note also the rather thin active Si layer used in the SOI device, as shown in Figs. 14.12 and 14.13. These are so-called fully depleted SOI devices. Figure 14.13, in particular, shows that the active area, and thus the junction depth, is rather thin, only about 150 nm including silicide contact layer compared with the same generation's bulk-Si device.

The advantages of SOI should be apparent from the preceding structural analysis. The source and drain regions extend down to the BOX, and thus the junction surface is minimized. The consequent reduced leakage currents and junction capacitance translate into improvements in device speed, power consumption, and a wide range of operational temperatures. The limited extent of source and drain also allows the SOI devices to be less sensitive to short-channel effects, which originate from charge sharing between the gate and junctions. The SOI device also experiences a lower electric-field peak and is more immune to hot carrier damage. The SOI device is highly competitive in the low-power, low-voltage circuit operated with a single battery supply of 0.9 to 1.5 V. The small gate voltage gap is well suited to switching the transistor from the off to on state. The SOI presents the possibility of achieving a quasi-ideal subthreshold slope and a threshold voltage below 0.3 V.

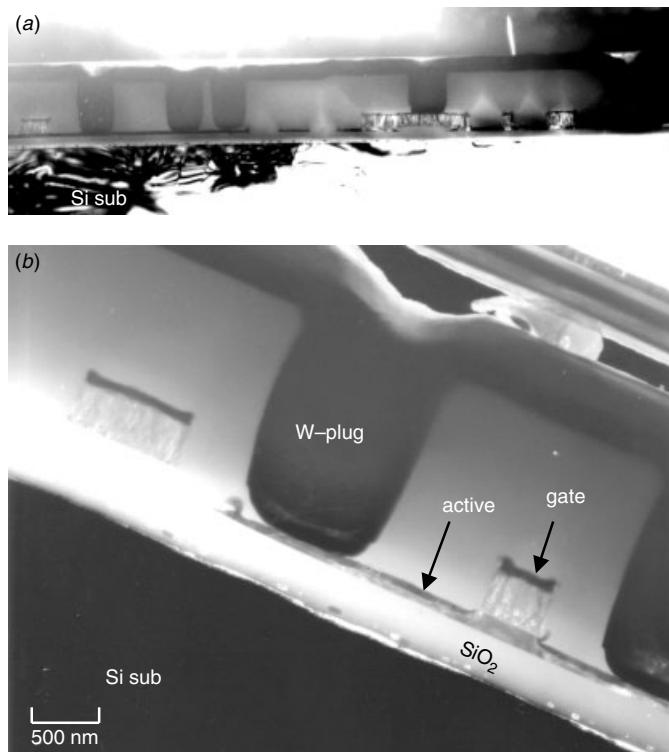


Figure 14.12 (a) An early SOI MOS device with W-plugs and $0.5 \mu\text{m}$ gate design rule. Notice that the active regions are basically floated on the SiO_2 BOX layer and that the individual BOXes are isolated from each other. (b) A closer look at the device.

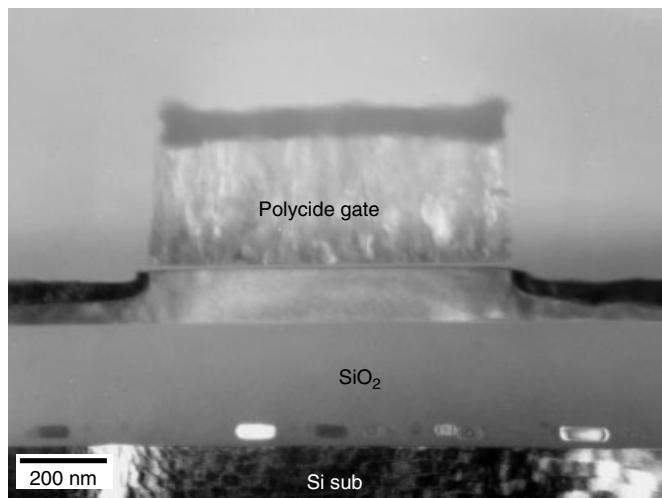


Figure 14.13 A closer look at the gate structure in a fully depleted SOI transistor. A self-aligned silicide is used in this device. Almost all of the traditional bulk-Si technologies can be implemented in SOI with minor or no modifications.

There are two unequivocal advantages of SOI over bulk-Si (Cristoloveanu 2000):

1. At similar voltage, operation of SOI shows a 30% increase in performance. Low-power dissipation can yield a 300% performance gain in SOI.
2. Some of the bulk-Si technologies can mimic a number of features that are native in SOI, for example, the double-gate configuration, low-high step doping, and dynamic threshold operation.

It remains to be seen how the SOI advantages in technology and cost will eventually gain a larger role in the semiconductor commodity market.

14.3 MEMS DEVICES

Micro-electro-mechanical systems (MEMS) were developed in the 1990s. Sensors, actuators, and electronics are integrated in this fast growing area of microelectronics. Basically MEMS are an extension of the technology of silicon-based photolithography to include mechanical devices of small dimension. Among the silicon micromechanisms that have been built are valves, springs, mirrors, nozzles, connectors, printer heads, circuit boards, heat sinks, and sensors for properties such as force, pressure, acceleration, and chemical concentration. Much of the interest in micromachining derives from the need for cheaper and more versatile sensors. The basis of micromechanics is that silicon, in conjunction with its conventional role as an electronic material, and taking advantages of an already advanced microfabrication technology, can be exploited as a high-precision, high-strength, high-reliability mechanical material. This is especially applicable wherever miniaturized mechanical devices and components must be integrated or interfaced with electronics (Petersen 1982).

There are several important differences between MEMS processing and IC processing that make it an exciting and rapidly evolving field (Hesketh 2000):

- Wider range of materials
- Wider range of fabrication processes utilized
- Use of three-dimensional structures
- Materials' properties not fully characterized
- Interdisciplinary expertise necessary for successful technology implementation
- CAD tools not yet fully developed for integrated thermal, mechanical, magnetic, optical, and electrical design

In general, silicon micromechanical structures are fabricated using two approaches: surface micromachining and bulk micromachining. In surface micromachining, the silicon substrate is used as a foundation and alternating layers of polysilicon and SiO₂ are deposited; the oxide is selectively removed and the remaining patterned polysilicon is the mechanical structure. In bulk micromachining, the single crystal silicon substrate is etched away to leave the structure. Bulk micromachining has typically been done using wet chemistry and reactive ion etching (RIE) processes (Shaw et al. 1994).

In the wet chemistry process, anisotropic wet chemical etches are accomplished by chemicals that attack certain crystal planes faster than other planes. The structure is shaped by an etching-mask design, wafer orientation, and by masked heavy ion

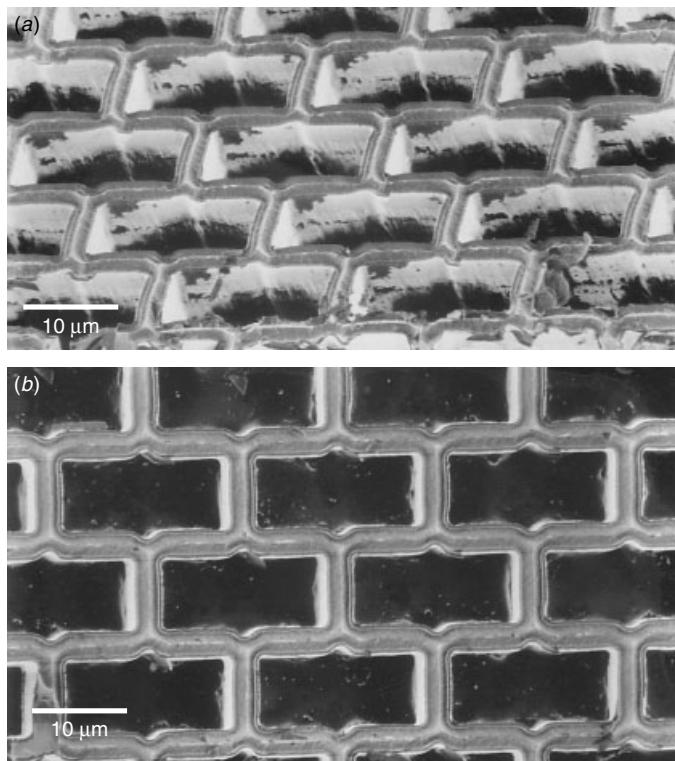


Figure 14.14 SEM tilted view (a) and plan view (b) of a MEMS device.

implantation. However, wet processes are difficult to control in submicron dimensions or in vertical high-aspect ratio devices. In contrast, in the dry bulk micromachining processes, a reactive ion etch is used to pattern and release structures, and only two masks are required to fabricate the suspended and movable structures. Figure 14.14 shows a MEMS device using such a dry process to create the suspending rectangular nets. Figure 14.15 shows the corresponding TEM plan view. Figure 14.16 shows a close-up view of the grid structure. Notice that the dark lines crossing the entire grid structure is the Si single-crystal bending contours. In other words, the whole grid structure, and in fact all the movable beams, fixed beams, and supporting spring grids, are composed from a single crystal Si. The technological beautifulness of such a device can only be appreciated by viewing TEM images such as Figs. 14.15 and 14.16.

The MEMS device involves movable parts within the Si substrate. When planar sections or cross sections are taken during TEM sample preparation, the movable parts simply do not have any support and therefore will disintegrate and fall apart. One way to resolve this is to embed and flood the sample with epoxy in a vacuum chamber, and cure it to form a rigid and self-supporting structure. Such an embedding process is typical in any TEM cross-sectional sample preparation. The epoxies developed by commercial TEM sample preparation product suppliers are mechanically tenacious and do not alter the device's structure in most cases throughout the sample preparation procedures and TEM analysis. Once embedded, the MEMS devices can be prepared by any known TEM sample preparation procedures as described in Chapter 4.

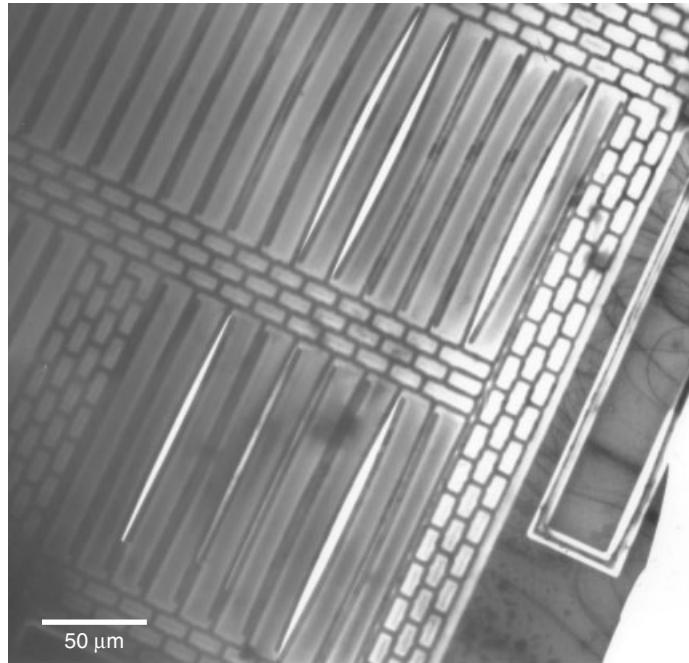


Figure 14.15 TEM plan view of a MEMS device showing the accelerometer sensor movable beams, fixed beams (forming the finger capacitor structure), supporting spring grids, and the silicon substrate.

A typical dry RIE approach for bulk micromachining process is summarized here, as shown in Fig. 14.17 (Shaw et al. 1994):

1. A mask oxide layer prepared by oxide deposition, photolithography, transfer patterning (oxide etching), and resist stripping
2. The first deep silicon etch using RIE and oxide as mask
3. PECVD oxide on sidewall
4. Conformal oxide removal (floor oxide removal)
5. The second deep silicon etch using RIE and oxide as the mask
6. Release etch (isotropic to undercut the thin-line patterns)
7. Sputter metallization (usually Al)

Notice that there is only one photo-mask layer used in the entire process (Shaw et al. 1994). Figure 14.18 shows a cross section of a sample after process step 5, the second deep silicon etching. Details of the sidewall's roughness and the etching profiles from the first and second etches are clearly visible. A closer look at the beam structure is shown in Fig. 14.19. One of the reasons we need to study the MEMS structure using TEM is the minuscule detail that TEM can reveal. This is important knowledge if the submicron MEMS process is to be developed. Take, for example, Fig. 14.19(b) where the beam mask oxide and the PECVD sidewall oxide are distinguishable. One can tell how the mask oxide and PECVD oxide determine the

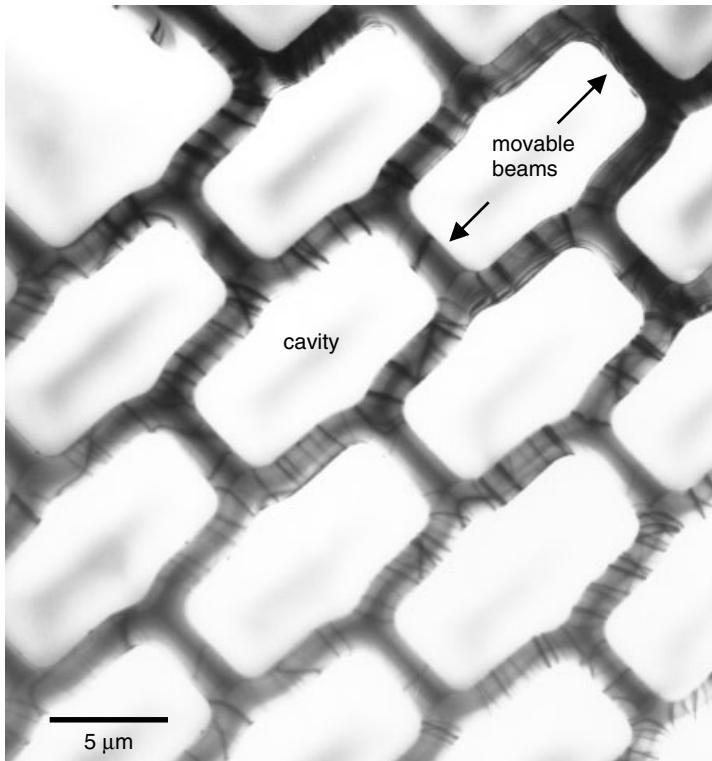


Figure 14.16 TEM plan view of a MEMS device showing the accelerometer sensor supporting spring grids. Single-crystal Si bending contours can be observed throughout the grid's structure.

shape and size of the movable beam. The beam sidewall's Si rough surface is due to the first deep silicon etch, and this helps in tuning the RIE process condition for optimization.

A successfully released beam structure is shown in Fig. 14.20. Note that the beam's bottom cavity height differs in Fig. 14.20(a) and (b). Such a difference can be manipulated by controlling the time durations of the first and second deep silicon etchings. Since the isotropic release etching is universal and ubiquitous, the normal base materials, such as thick capacitor plate beam are also etched. Figure 14.21 shows an example of such a process in poor control where the capacitor plate, which is not supposed to be released, is nearly cut off from the Si substrate. This will have detrimental effects in terms of MEMS device's yield and reliability. A successfully released movable beam structure is shown in Fig. 14.22. One prominent element noticed immediately is the overhanging oxide mask, which later will support the metallization layer and prevent it from electrically shorting to the exposed silicon under the beam. A MEMS device with a structure such as this has many applications. One example is as the accelerometer used in automobiles to trigger the airbag when there is sudden impact due to a car accident. The device's responses to mechanical impact and the movable beams shift to one side of the cavity due to inertia. The capacitance between the movable beam and the fixed capacitor plate changes immediately due to such movement. Thus

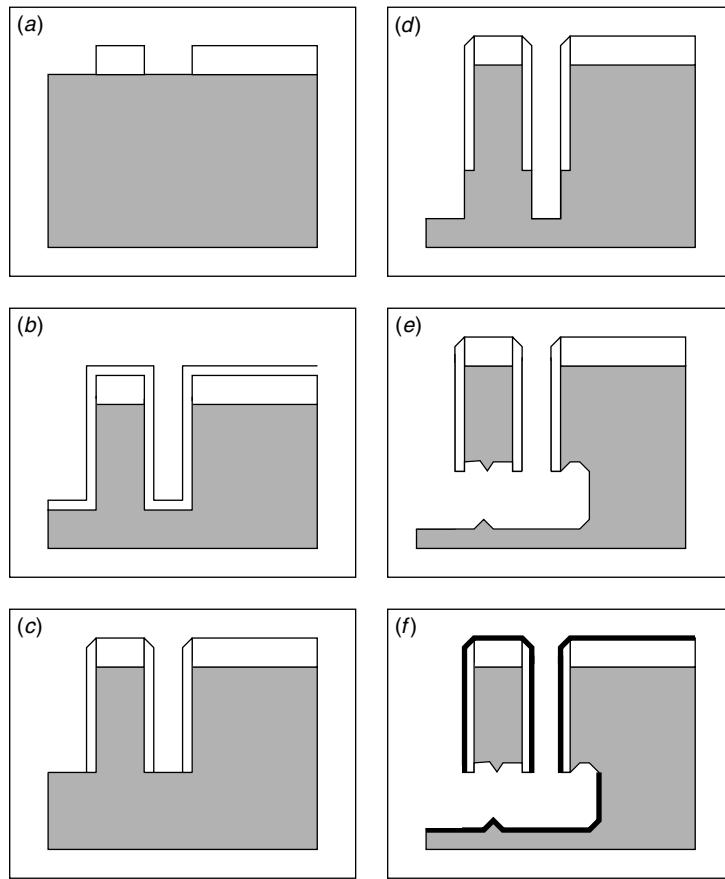


Figure 14.17 Schematic cross section of the dry etch bulk micromachining process flow. (a) Mask oxide formation, (b) first deep silicon etch and sidewall oxide deposition, (c) removal of floor oxide, (d) second deep silicon etch, (e) release etch, (f) sputtering metallization. (Shaw et al. 1994)

a mechanical signal is translated into an electrical signal, and the control circuit then triggers the airbag's inflation.

Figure 14.23 shows a similar but failed device where one set of the movable beams has lost its mechanical support and thus collapsed into the other set of beams due to static “sticking issue.” Detailed studies of the Si beam etch profiles, residue sidewall oxide, cavity shape, among others, as seen in Fig. 14.24, are of vital importance to the future developments in submicron MEMS device. TEM, as a complimentary analysis tool to SEM, provides additional insight to the MEMS device process optimization, yield enhancement, and failure analysis.

14.4 SiGe TECHNOLOGY

The concept of bandgap engineering has been applied in compound semiconductors for many years, especially in III–V and II–VI groups. Transistors built out

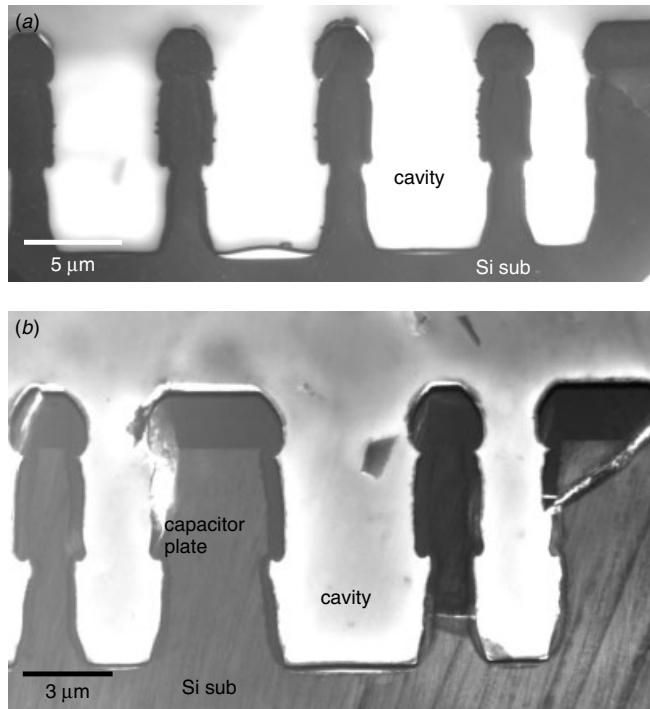


Figure 14.18 TEM cross section of the MEMS sample after the second deep silicon etching. The etching profiles, the mask oxide, and the irregular sidewall shapes are all clearly visible.

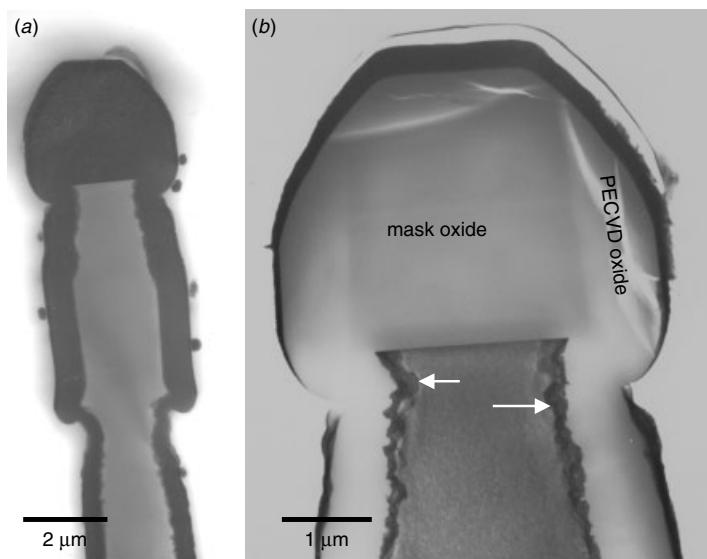


Figure 14.19 Close-up of the TEM cross section of the MEMS after the second deep silicon etch. (a) Overall profile of the beam's structure. (b) Close-up at the beam's headpiece revealing the mask oxide's shape, the PECVD sidewall's oxide, and the sidewall's rough contour (arrows) created by first deep silicon etching.

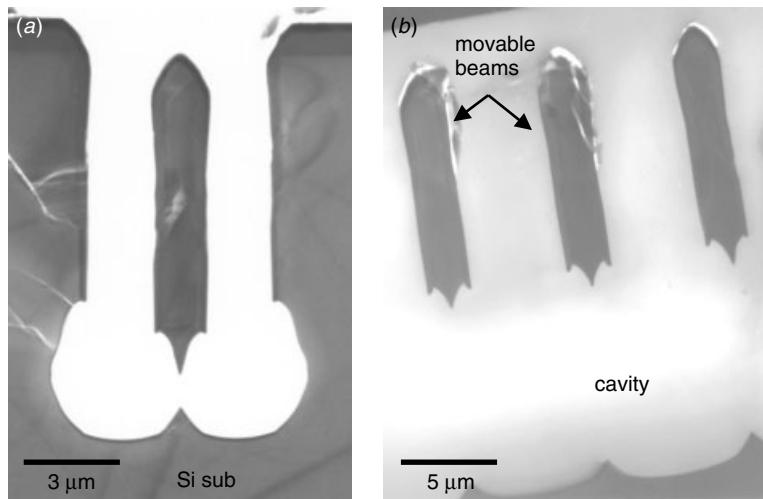


Figure 14.20 TEM cross section of the MEMS device after the beam release etch (process 6). (a) Notice that the isotropic etching at the bottom shows a nearly spherical etching cavity. (b) The distance between the released beams and the Si substrate can be controlled and adjusted by the depths of the first and second deep silicon etchings.

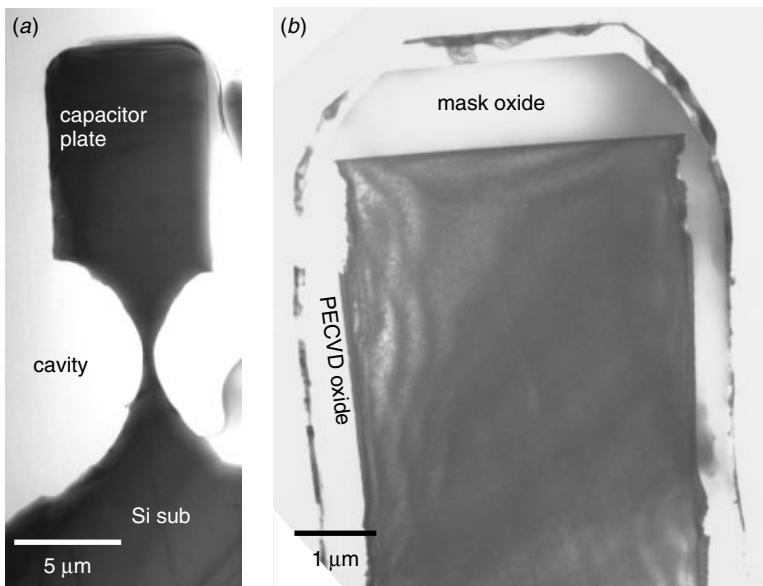


Figure 14.21 When the thin movable beams are release by the isotropic etch, the thick beams are also undercut. (a) A capacitor plate that is nearly cut lose and thus may cause yield and reliability concerns. (b) The mask oxide shapes of a capacitor plate and a movable beam, Fig. 14.19(b), are not exactly the same.

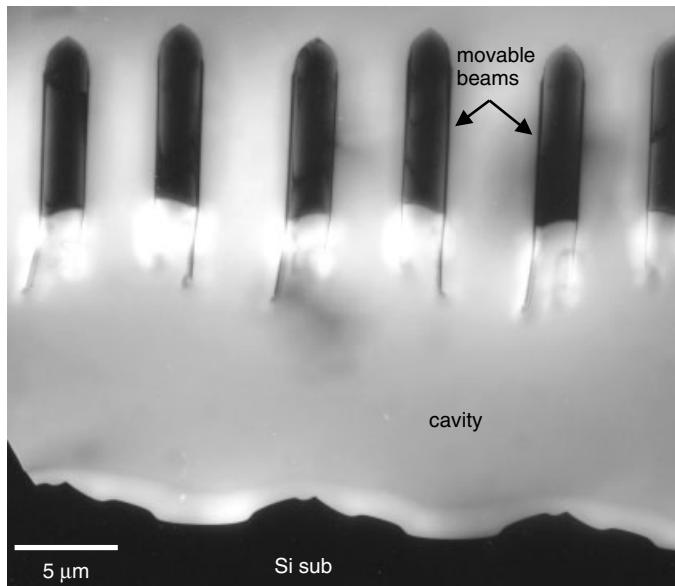


Figure 14.22 A collection of released beams. Notice that the oxide hangs over both sides of each beam. The movable beams are in a finger formation and are mechanically (and electrically) connected to different bases to form a finger capacitor.

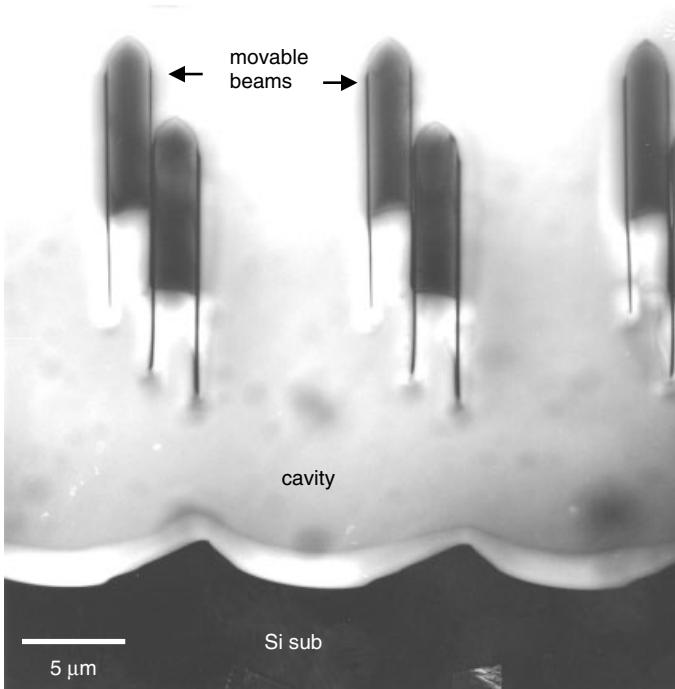


Figure 14.23 Under certain conditions the beams will irreversibly contact each other in pairs, indicating that one of the mechanical supports of the movable part has collapsed.

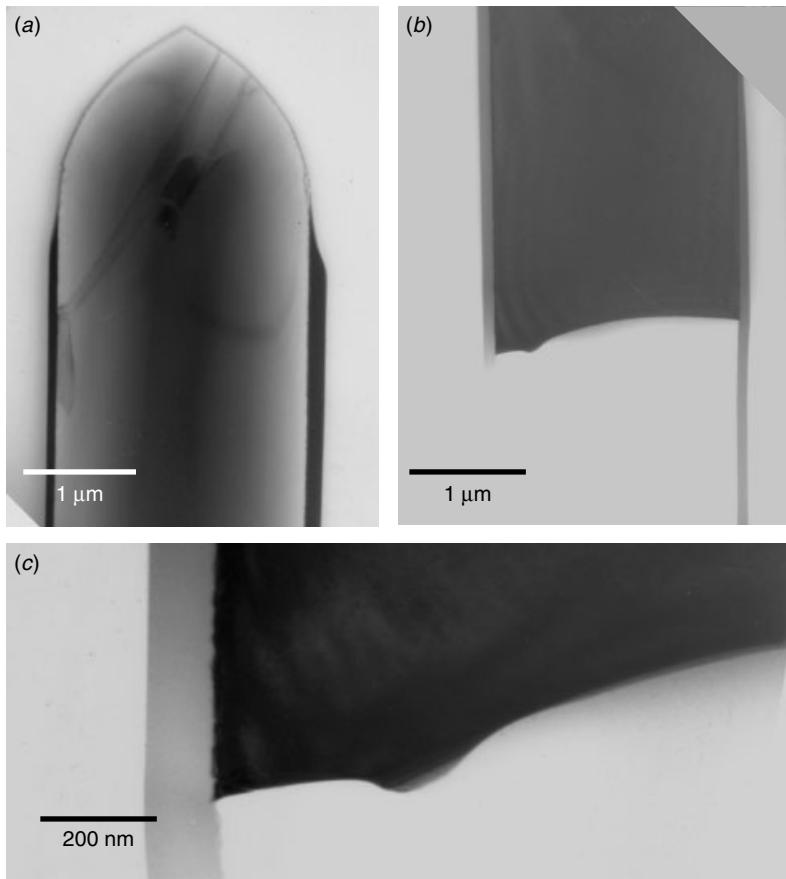


Figure 14.24 TEM close-up of the movable beam revealing the process and etching details of (a) the beam's top surface, (b) the beam's bottom with the oxide overhang, and (c) the oxide to Si undercut corner.

through bandgap engineering have improved the device's performance and performed tasks that are impossible for conventional Si-based devices. Examples are high-speed devices, high-frequency devices, and optical devices. However, compound semiconductor devices suffered one major disadvantage that can never match Si-based technology, cost. The difficulties in mass production of compound semiconductor devices, the low yield, and reliability concerns, all make for a very costly production of compound semiconductors. Device designers have long sought to combine the bandgap engineering techniques enjoyed in compound semiconductors technologies with the fabrication maturity, high yield, and low cost associated with conventional Si-based technologies. In this regard epitaxial silicon–germanium (SiGe) offers considerable potential for realizing viable bandgap-engineered transistors in the Si materials system, and allows Si electronics devices to achieve performances that were once thought impossible (Cressler 2000). The epitaxially grown Si layer on a strain relaxed SiGe buffer layer gives rise to a biaxially tensile Si film when its thickness is below a critical value. The strain is caused by the lattice mismatch of Si and SiGe. The strain in film

splits conduction band energy valley and degeneracy of (HL/LH) valence band. Both valley splitting and degeneracy reduces carrier effective mass along the transport direction. The separation of the conduction band minima reduces inter-valley scattering and deformation of valence band reduces interband scattering compared to bulk-Si. As a result the electron and hole mobilities are increased in strained Si (Bera et al. 2003).

The Si complementary metal-oxide-semiconductor (CMOS) field effect transistors (FET) are the most important building block of the digital integrated circuits. Their low power consumption and their mature technology are the advantages. The use of strained Si/SiGe heterostructure promises to improve the speed-power performance of CMOS by offering higher electron and hole mobilities. The challenge is thus to make the CMOS FET devices on top of SiGe heterostructure (Sadek et al. 1996). Figure 14.25 shows an example of MOSFET on top of the strained Si on strain relaxed SiGe heterostructure substrate. Notice all the dislocations are confined below the graded SiGe layer while the strain relaxed SiGe layer is defect free. As can be seen in Fig. 14.25, device built on top of strained Si layers on strain relaxed SiGe layer are very similar to the conventional Si MOSFET device. Such a device is sometimes referred to as a Si-based hetero-junction CMOS (HCMOS) technology device. In fact, the same Si-based MOSFET technologies, such as salicide contacts, LDD structure using spacer, and Cu dual damascene metallization, etc., all can be used without major modification. The process used is highly compatible with the conventional Si-based MOSFET process technology and can be easily adopted by any wafer fabrication facility that is capable of manufacturing Si-based MOS device.

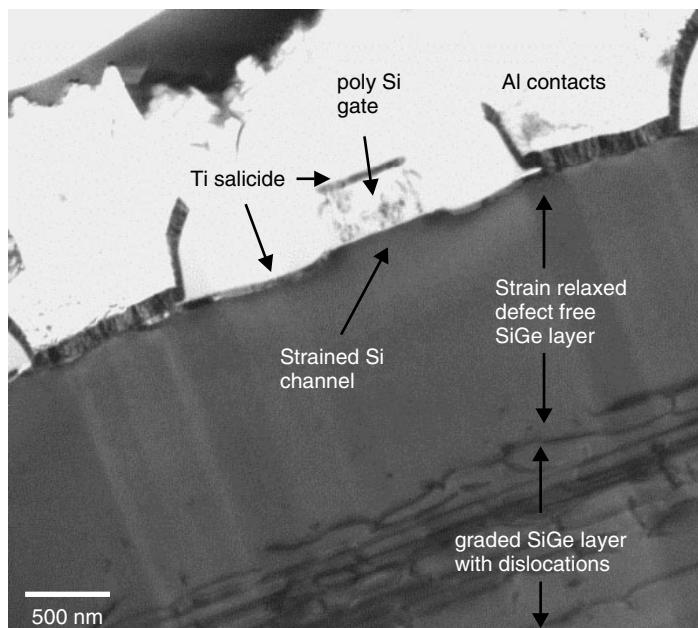


Figure 14.25 TEM cross section of a MOSFET device of the SiGe heterostructure substrate. The strain-relaxing misfit dislocations and interconnecting threading dislocations are restricted to the graded part of the buffer layer and below. The strain-relaxed SiGe layer is defect free. (Courtesy of AmberWave Systems Corp., USA)

There are, however, a few slight differences between the new SiGe device and conventional Si device. Ion implantation induced intermixing of Si and SiGe is one issue. Once disrupted by ion bombardment, the SiGe and Si laminated structure intermixing is thermodynamically irreversible. Annealing may recover the amorphization and defect formation but it cannot restore the hetero-junction layered structure. Other problems arise because the chemical and metallurgical reactivity between Si and Ge is different. An example is at the contact region where salicide formation taking place. A more detailed discussion on the germanosilicide has been given in Chapter 7.

Strained SiGe devices pose challenges to the TEM sample preparation. FIB assisted final thinning introduced irreversible damages. FIB induced surface amorphization can severely deteriorate the image quality. Examples on using high resolution TEM to the analysis of hetero-junction quantum well and quantum dot are given in Chapter 16. Careful mechanical polishing with limited ion milling time and sample stage cooling can reduce the ion milling induced artifacts and improve the image quality.

14.5 OPTICAL CD MATERIALS

The phase-change optical disk (or CD-rewritable, CD-RW) is one of the next generation's high-density memory media. The information stored is removable and rewritable with high density and capacity, and on average, it has a much lower unit cost than magnetic media. The CD-RW is lightweight, relatively reliable, and insensitive to the environment. In order to compete with its magnetic media counterparts, however, it is essential that the optical disk be capable of rewriting the information over more than a million times without deteriorating the basic materials and structures.

The basic operational principle for the phase-change optical disk is quite straightforward. A thin polycrystalline film is used as the information storage base material. The most popular materials are the Ge–Sb–Te ternary alloy or the Ag–In–Sb–Te alloy (Yamada et al. 1991). The material itself deflects laser photons differently when it is amorphous compared to when it becomes crystalline, and thus when the laser beam hits the materials, the detector can pick up a different signal in the amorphous area and in the crystalline area. To make it work properly, a very thin Ge–Sb–Te alloy, for example, 30 nm, is sandwiched between amorphous ZnS–SiO₂ layers (Kouzaki et al. 1997; Kouzaki 2000). One side of this lamination is usually coated with thick Al metallization to serve as a mirror and reflect light. The base Ge–Sb–Te material is crystalline, but when hit by a strong laser, it becomes amorphous locally, and this serves as a single digit information (0 and 1, as amorphous or crystalline, and vice versa). The laser beams used for read and write information are of different wavelengths and energy in such a way that only the writing laser can induce a phase transition but not the reading laser. As mentioned before, one million times of information rewrite simply means that the materials must go through one million times of crystalline–amorphous transition without apparent deterioration of the materials themselves. It is thus very important to design and control the quality of the lamination, the thermal dissipation, thin film deposition, and interface roughness, as well as the overall mechanical integrity. A detailed study of the microstructure and phase transformation of the metallurgical system is crucial to understanding reliability and related issues.

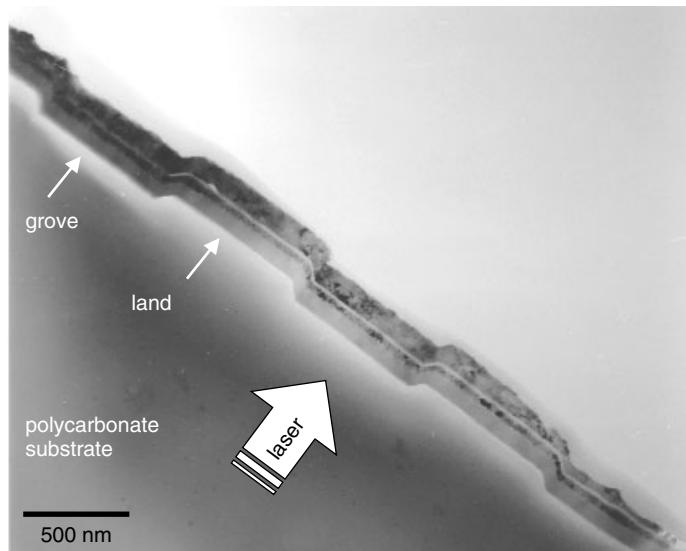


Figure 14.26 Cross section TEM image of phase-change optical disk. The zig-zag morphology are the tracks structures for data storage with pitch width 740 nm. Data are stored as marks on both grove and land tracks.

A typical cross section of such a CD-RW is shown in Fig. 14.26 and 14.27. The basic lamination is Al metal thickness of about 90 nm, a ZnS–SiO₂ layer of about 10 nm, the Ge–Sb–Te alloy film of about 30 nm, and another thicker ZnS–SiO₂ of about 65 nm, all laminated together. Polycarbonate is the external substrate material on both sides of the lamination, and UV resin is used for bonding the whole structure together. The zigzag pattern observed in the cross section, as seen in Fig. 14.26, forms the “tracks” that define the information read/write resolution. A typical track pitch is around 0.75 μm for current CD-RW but can be smaller if a blue laser (shorter wavelength) is used to increase the spatial resolution. Figures 14.28 and 14.29 show the plan view of the same track structure of a CD-RW disk. The polycrystalline tracks are clearly visible. The square windows are the track’s resetting marks, which serve as the starting points of each track. Notice from Fig. 14.28 that each is slightly shifted to its neighboring tracks for compensation. As we noted above, the Al film serves as a mirror to reflect the incoming laser beam back to the detector. It is thus critical to have a smooth mirror surface. The Al to ZnS–SiO₂ interface must be very smooth, as seen in Figs. 14.30 and 14.31. It is not important how thick the Al is as long as it reflects the laser photons properly. In fact nonhomogeneous Al film thicknesses, as shown in Fig. 14.31 as an example, are often found locally in CD-RW disks. The Ge–Sb–Te alloy film is the layer that possesses the information storage function. The film is kept very thin to facilitate the phase transformation reaction. Also important is the grain size, as it has two important functions:

- The finer grain size of Ge–Sb–Te alloy gives rise to higher reflectivity and thus is distinct from the amorphous phase.

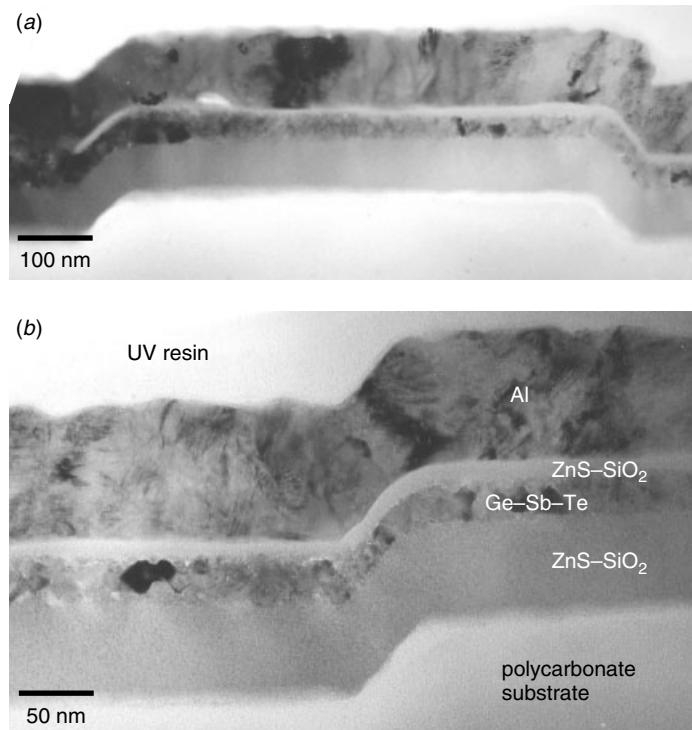


Figure 14.27 Close-up of the laminated structure showing details of each layer, as marked.

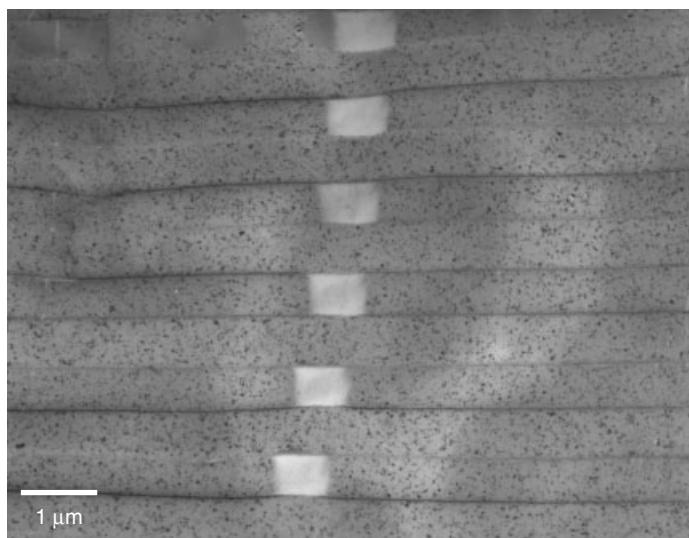


Figure 14.28 TEM plan view of the phase-change optical disk showing the polycrystalline contrast. The horizontal grid lines are the “tracks” on the optical disk for data read/write. The white squares in the center of the image are the reset alignment marks. They are slightly offset with each other for compensation.

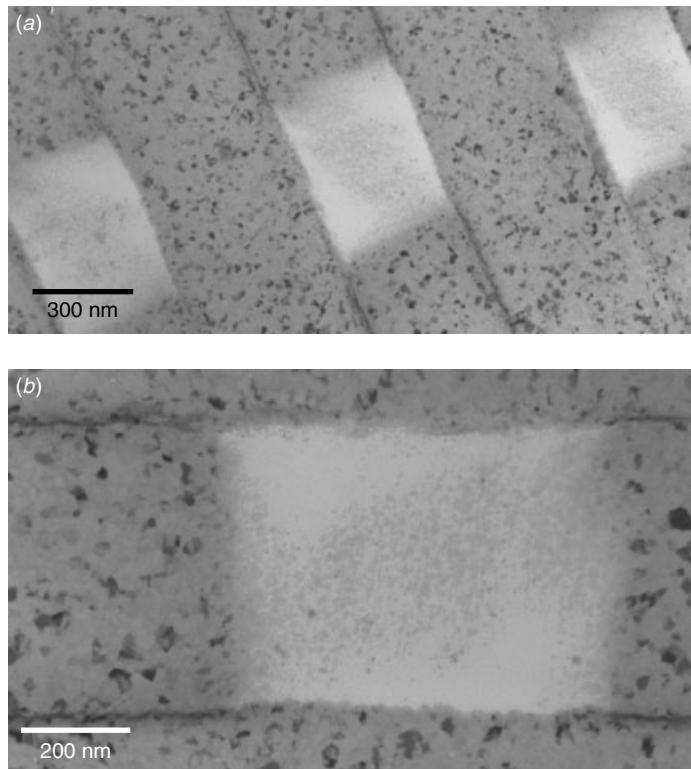


Figure 14.29 (a) TEM plan view of the phase-change optical disk showing the polycrystalline contrast. (b) A close-up of the reset alignment marks shows the rectangular windows of the amorphous and featureless contrast compared with the normal crystalline contrast.

- The finer grain size of Ge–Sb–Te also facilitates the back-and-forth phase switching between crystalline and amorphous phases.

It is thus desirable to have a nanocrystalline Ge–Sb–Te alloy film in the CD-RW disks. Figures 14.32 and 14.33 show the HRTEM images on the nanocrystalline Ge–Sb–Te alloy film. The alloy is in a multiphase equilibrium region with a low-temperature eutectic reaction. That is one reason why it is capable of swift back-and-forth transforms between crystalline and amorphous phases. Figures 14.32 and 14.33 show the fine-grain structures to have interlocking grain boundaries, as is typical of eutectic alloy microstructures.

An interesting question is what happens to the microstructure when the CD-RW disk is subjected to laser ablation during a rewriting procedure. As high-power laser hits the thin film laminate from one side, as shown in Fig. 14.31, it triggers a local phase transformation. The control of the laser's energy, timing, and duration are of critical importance. Figure 14.34 shows a plan view on the CD-RW disk with two distinctive laser ablation points. One thing that can be immediately noticed is the crack line through the middle of each ablation point. As also seen in Fig. 14.35, these cracks through the middle of the points are not the expected result and should be avoided. The cracks are

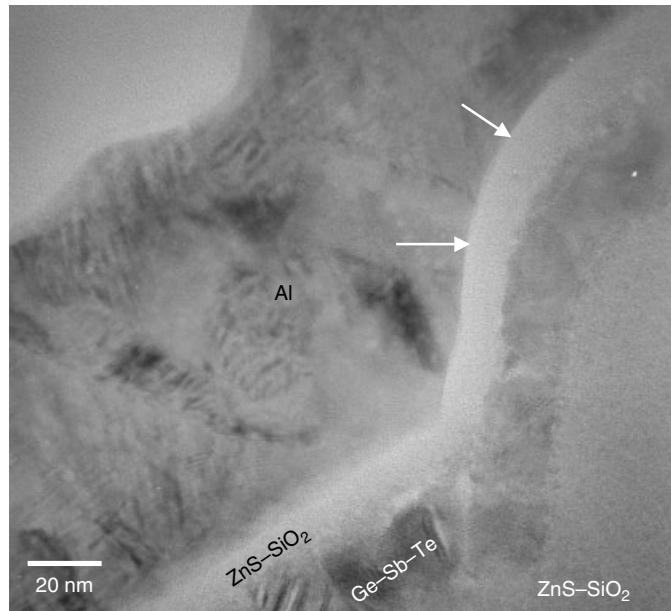


Figure 14.30 Close-up at the laminate structure showing some details of the Al/ZnS–SiO₂ interface and Ge–Sb–Te grain structures. An extremely smooth Al to ZnS–SiO₂ layer interface is important, as indicated by arrows.

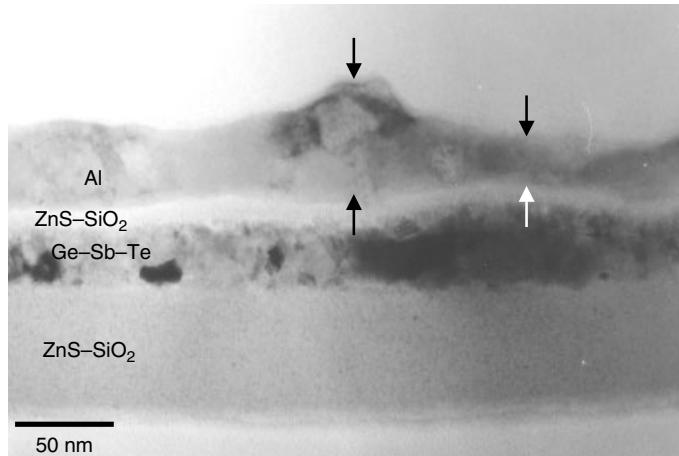


Figure 14.31 Cross section of the laminated layers. Al thickness variation, as indicated, does not affect the function of the optical disk, since it is used only as a reflection mirror.

due to prolonged laser time. The poor and insufficient thermal dissipation results in local overheating and thus the cracking of the polycarbonate. The polycarbonate cracks then propagate back to the laminate structure and introduce cracks on the Ge–Sb–Te alloy film, as seen in Figs. 14.34 and 14.35. The cracks further penetrate into the Al film. Since the cracks are irreversible and the damages are permanent, the cracks are

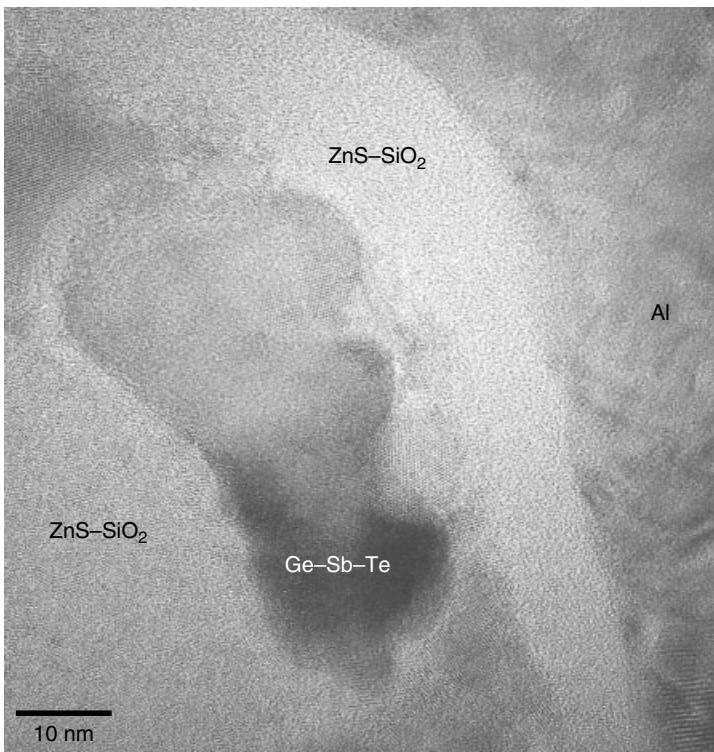


Figure 14.32 HRTEM image of the Ge–Sb–Te alloy film’s microstructure showing the lattice image of the crystalline grain of the Ge–Sb–Te alloy and its neighboring Al film.

considered a product defect. Note that on these plan views, the microstructures of the crystalline Ge–Sb–Te alloy and Al overlap. The two ZnS–SiO₂ layers and the UV resin are featureless due to their amorphous contrast.

Close examination of the plan view TEM images allows us to understand the microstructure change due to laser ablation in detail, and these changes can be summarized as follows: First, a crack appears through the center of the ablation point. In some cases, the crack can spread out to neighboring laser recording points and causing mechanical disintegration of the whole film structure. Next, the crack cuts through not only the Ge–Sb–Te film but also through the Al film. Any cracks are permanent and irreversible, and so are considered product defects or an overstressed (EOS) damage to the device.

Upon laser ablation, counting from the center of the point (ground zero), there are three phase transition zones that can be observed. These are named Z1, Z2, and Z3 accordingly.

- Z1 is the ground zero of the laser ablaze point. All Ge–Sb–Te grains have transformed into an amorphous phase and show clear and featureless contrast, as seen in Fig. 14.36 and 14.37. However, Al metal has much better thermal conductivity and a thicker film to absorb and dissipate the heat. Therefore it allows slow

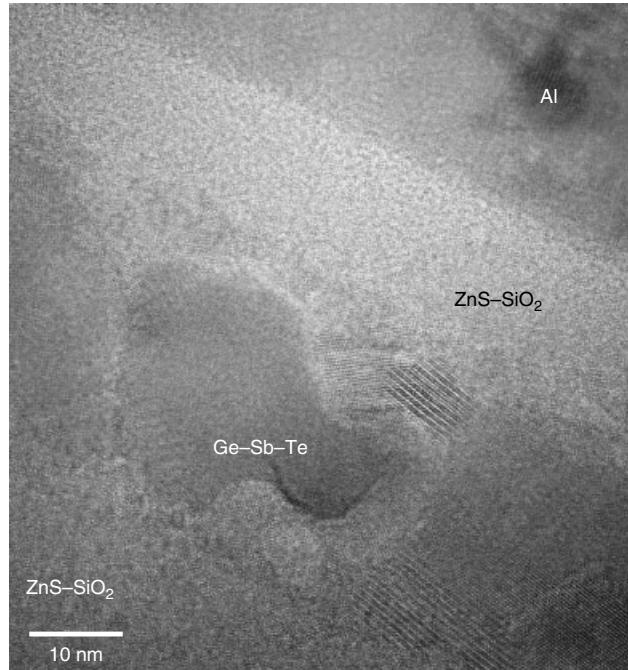


Figure 14.33 Another HRTEM image of the Ge–Sb–Te alloy film’s grain lattice. The grain size of the Ge–Sb–Te is shown. It is easier to trigger the amorphization reaction using such a fine-grain film structure.

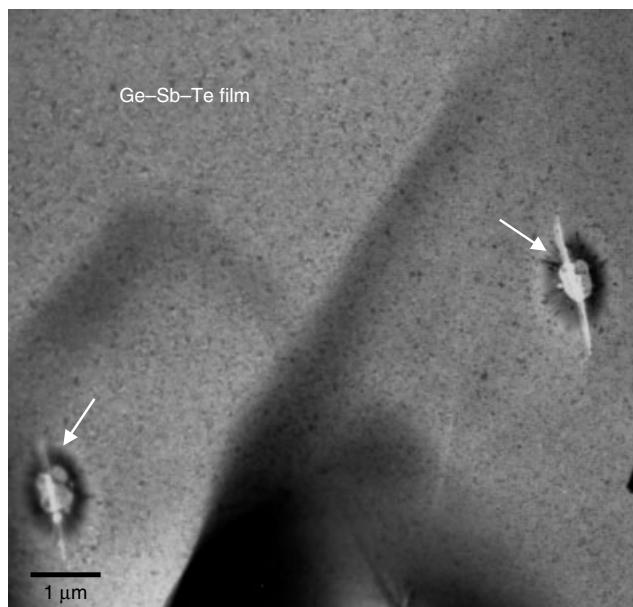


Figure 14.34 Plan view of the phase-change optical disk with two laser spots, as indicated. The spots are the record marks used to amorphize the Ge–Sb–Te film. The cracks are observed to extend from the center of the marks.

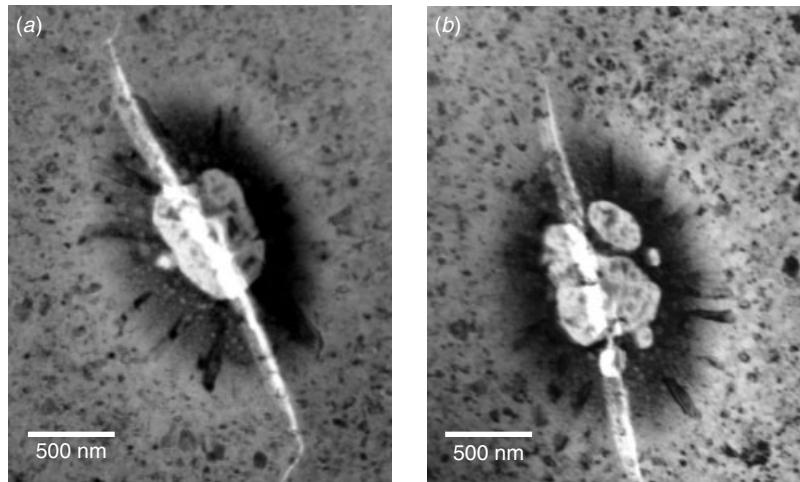


Figure 14.35 Detail of the laser record marks showing the center's re-melting and amorphization zone, with a large crack extending out from the center of the spot.

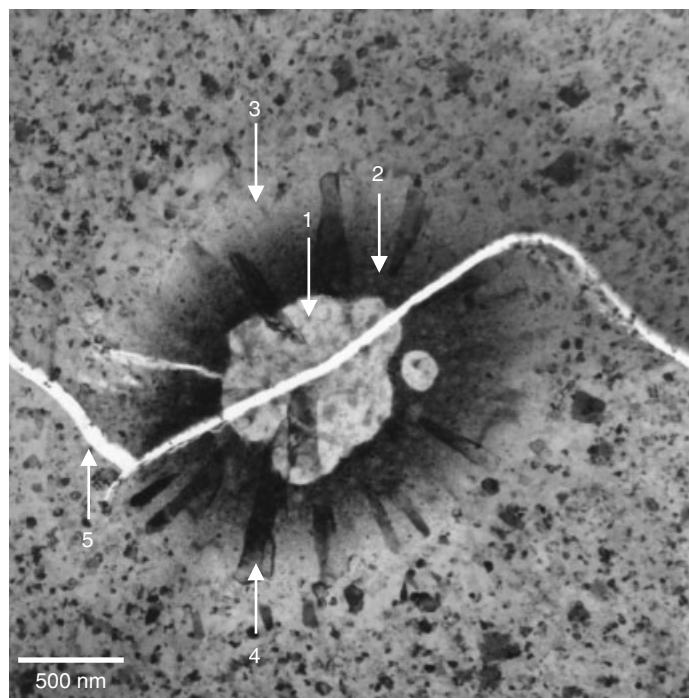


Figure 14.36 Close-up at the laser record spot showing (1) a bright central amorphous zone, (2) dark circular zone surrounding the central amorphous zone, (3) a bright featureless zone surrounding the dark zone, (4) large radiant grain structures overlapping the dark zone, and (5) a long crack through the center of the spot and zigzagging through the whole area.

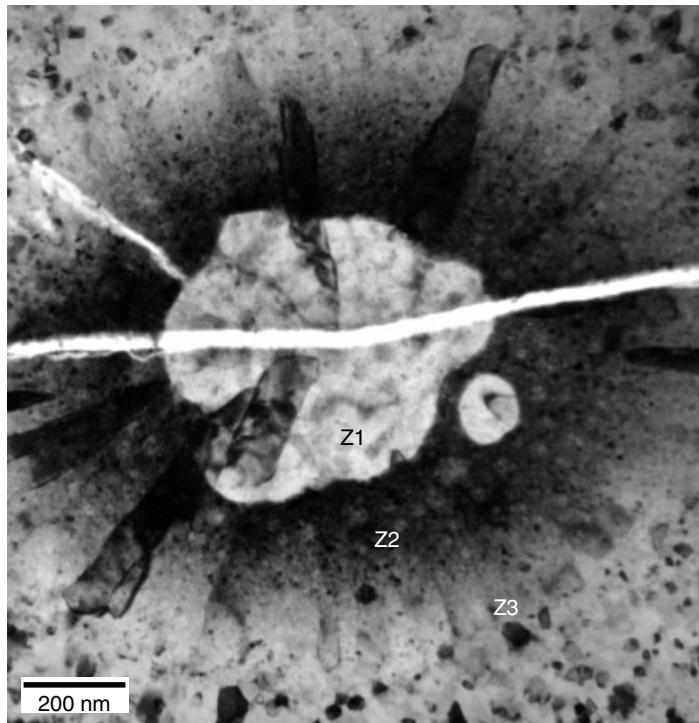


Figure 14.37 Close-up at the bright central amorphous zone and dark circular zone showing the dark zone to be composed of fine crystalline grains. The large radiant grains are Al layer grains with laser-induced grain growth. Three transition zones can be distinguished: Z1, Z2, and Z3, as shown.

cooling and thus forms a large and elongated grain structure upon laser ablation. This remains to be the only grain contrast in the Z1 area seen in Fig. 14.36.

- Z2 is the area immediately surrounding Z1. Overall the contrast is dark with some white patches within a dark and fine crystalline structure. This is the area where the amorphous (white patches) and nanocrystalline (dark areas) Ge–Sb–Te phases have intermixed. Notice that large elongated Al grains are also visible throughout this region. Note that local hot spot can introduce a local smaller Z1 within Z2, as seen in Fig. 14.35(b), 14.36, and 14.37.
- Z3 is an out skirt area of Z2. The contrast is clear and bright. This is the area where the Ge–Sb–Te retains its nanocrystalline phase. Al still appears as elongated large grains, and thus its grain structure is clearly visible in this region. The affected zone for Ge–Sb–Te includes Z1 and Z2. But the affected zone for Al is bigger and includes Z2, Z2, and Z3. This is reasonable since the Al film is thicker and has better thermal conductivity.

With the help of TEM analysis, the microstructure, the process, and the laser ablazing conditions can be studied and fine-tuned to achieve a comprehensive understanding on CD-RW materials, processes and reliability.

REFERENCES

- A. O. Adan, T. Naka, A. Kagisawa, and H. Shimizu, *Proc. 1998 IEEE Int. SOI Conf.*, 9–12, 1998.
- F. Assaderaghi, and G. Shahidi, *Proc. 2000 IEEE International SOI Conf.*, 6–9, 2000.
- S. Bagchi, Y. Yu, M. Mendicino, J. Conner, A. Anderson, L. Prabhu, M. Tiner, and M. Alles, *Proc. 1999 IEEE Int. SOI Conf.*, 121–122, 1999.
- L. K. Bera, S. Mathew, N. Balasubramanian, G. Braithwaite, M. T. Currie, F. Singaporewala, J. Yap, R. Hammond, A. Lochtefeld, M. T. Bulsara, and E. A. Fitzgerald, *Applied Surface Science*, (in press).
- P. Cappelletti, *Microelectron. Reliabil.*, **38** (2), 185–188, 1998.
- J. D. Cressler, in *The VLSI Handbook*, W. K. Chen, ed., CRC Press/IEEE Press, 1–14, 2000.
- S. Cristoloveanu, *J. Electrochem. Soc.*, **138**, 1313, 1991.
- S. Cristoloveanu, in *The VLSI Handbook*, W. K. Chen, ed., CRC Press/IEEE Press, 1–15, 2000.
- D. Frohman-Bentchkowsky, *Appl. Phys. Lett.*, **18**, 332, 1971.
- C. Golla and S. Ghezzi, *Microelectron. Reliabil.*, **38** (2), 179–184, 1998.
- P. J. Hesketh, in *The VLSI Handbook*, W. K. Chen, ed., CRC Press/IEEE Press, 1–71, 2000.
- H. Iizuka, F. Masuoko, T. Sato, and M. Ishikawa, *IEEE Trans. Electron Devices*, **23**, 379, 1976.
- M. Ito, K. Y. Yamagata, H. Miyabayashi, and T. Yonehara, *Proc. 2000 IEEE Int. SOI Conf.*, 10–11, 2000.
- D. Kahng and S. M. Sze, *Bell Syst. Tech. J.*, **46**, 1288, 1967.
- T. Kikuchi, Y. Onishi, T. Hashimoto, E. Yoshida, H. Yamaguchi, S. Wada, N. Tamba, K. Watanabe, Y. Tamaki, and T. Ikeda, *IEDM*, 923–926, 1995.
- T. Kouzaki, *Jap. J. Electron Microscopy*, **49** (1), 85–88, 2000.
- T. Kouzaki, K. Yoshioka, and E. Ohno, *Materials Res. Soc. Symp. Proc.*, **480**, 251–256, 1997.
- F. Masuoko, M. Asano, H. Iwashashi, T. Komuro, S. Tanaka, *Solid State Circuits Conf. Digest of Tech. Papers*, Vol. XXVIII, 168–169, 1985.
- S. Pae, J. P. Denton, and G. W. Neudeck, *Proc. 1999 IEEE Int. SOI Con.*, 108–109, 1999.
- K. E. Petersen, *Proc. IEEE*, **70** (5), 420–457, 1982.
- A. Sadek, K. Ismail, M. A. Armstrong, D. A. Antoniadis, and F. Stern, *IEEE Trans. Electron Devices*, **43** (8), 1224–1232, 1996.
- F. Schaffler, D. Tobben, H.-J. Herzog, G. Abstreiter, and B. Hoolander, *Semicond. Sci. Tech.*, **7**, 260–266, 1992.
- K. A. Shaw, Z. L. Zhang, and N. C. MacDonald, *Sensors and Actuators*, **A40**, 63–70, 1994.
- S. M. Sze, in *Future Trends in Microelectronics*, edited by S. Leuyi, Wiley International, 1999.
- S. Voldman, D. Hui, L. Warriner, D. Young, R. Williams, J. Howard, V. Gross, W. Rausch, E. Leobangdung, M. Sherony, N. Rohrer, C. Akrout, F. Assaderaghi, and G. Shahidi, *Proc. 1999 IEEE Int. SOI Conf.*, 68–71, 1999.
- N. Yamada, E. Ohno, K. Nishiuchi, and N. Akahira, *J. Appl. Phys.*, **69**, 2849–2856, 1991.

15 TEM in Under Bump Metallization (UBM) and Advanced Electronics Packaging Technologies



Plan view TEM image. VIA chain test key near bonding pad areas. The animal head shape was made of Cu metallization layer shrinkage and wrinkling. Part of the VIA chain can still be seen on the lower left hand side of the micrograph.

The application of TEM in packaging related materials and process analysis does not seem to be necessary until high pin count, high-density packaging, and, in particular, flip-chip technology are developed and fully used in mass production. A key component to the success of this technology is the use of under-bump metallization (UBM) materials. A barrier metal layer, normally refractory in nature, is applied in between the bonding pad (Al or Cu alloy) and the conducting materials used beyond the chip (normally PbSn solder, lead-free solder, or related materials) to prevent interdiffusion and metallurgical reaction. The UBM layer is crucial as the

solder materials generally have low melting points and are chemically and metallurgically reactive and mechanically soft. The UBM can provide mechanical bonding, prevent the chemical reaction from attacking the chip bond pad, and ensure the processibility and reliability for the device. For the performance mentioned, the UBM system has evolved from a single layer into several layers laminated together with each layer performing its particular function. A glue layer is first applied on top of the chip bond pads. This layer provides the adhesive strength for Al alloy (or Cu alloy) and chip surface passivation (SiO_2 , Si_3N_4 , polyimide, etc.). The materials generally used are Ti, Cr, and Al. A second layer of metal is then applied to provide the diffusion and reaction barrier. Refractory materials are used for this purpose, such as Ni, W, Ti, and Cu alloys. A top layer of metal is finally applied to prevent oxidation and ensure solderability. Materials like Au, Cu, and Ni are used here.

Different combinations of the metallizations provide different performances in specific application environments. Generally, there is no single UBM system that outperforms all others in every application and in reliability. In deciding on an UBM system, one considers factors such as process and materials cost, yield, and device re-workability. Different systems such as Cu/Cr–Cu/Cr, Cu/Ni(V)/Al, Cu/Ti, Cu/W(Ti), Au/Cu/W(Ti), electroless Ni/Cu/W(Ti), Au/Ni/Ti, and Au/electroless Ni have been developed. A few commercially available UBM systems are analyzed below as examples. These are by no means the only choices in the market. There are many tailored UBM systems specifically designed for certain application purposes.

The choices are enormous. Also discussed here is the Au–Al ball bonding system. Au wire bonding to Al has served the microelectronics industry for the last 40 years and has been proved to be the most durable and reliable interconnection system so far. In reality it still is and will remain to be the most important interconnection system on the market. However, it is the system that surprisingly we do not know very thoroughly. The Au–Al diffusion couple metallurgical interaction, of course, has been studied. But what happens at the interface of the Au wire bonding to the Al alloy in a real device, with different wire bonding conditions and later used in different reliability testing conditions, remains largely ambiguous. One simple reason is that the interfacial reaction is dynamic and changes with the thermal and chemical environment. Another reason is that the interface reaction is difficult, if not impossible, to study using SEM. As will be shown here, the interface reaction is complicated and involves several intermediate metastable phases. To identify these irregular interfacial phases is a difficult task for SEM.

15.1 Au–Al BALL BONDING SYSTEM

Package interconnection using Au wire bonding to the Al pad remained to be one of the most widely used and most studied metallization systems. Even in the present metallization era where the top metallization is Cu, Al is still employed to coat the top Cu layer so that the usual old-fashioned package Au or Al bonding can be employed. However, our understanding of the Au–Al interaction has remained limited to bulk diffusion and equilibrium thermal annealing where the conditions are quite ideal. There is very little understanding of the Au–Al interaction under the different ball bonding conditions, different reliability stresses, and different application environments. The

main challenge in such studies is the lack of a suitable analytical technique that can provide sufficient resolution and analytical power to investigate and understand in detail the microstructural changes in such confined areas.

Figure 15.1(a) shows a typical TEM cross section of a Au ball bonding to Al (Tung et al. 2002). At first glance, the Al appears to be almost consumed by the Au at the bottom of the bonding area. Figure 15.1(b) shows a different interface morphology formed due to a change in the bonding condition. In this case the reaction interface shows discrete islands as compared to the previously smooth interface. A closer look reveals that there is some Al retained in some bottom areas, Fig. 15.2(a). The bulk intermetallic formed was determined to be Au_4Al , as predicted by the conventional diffusion couple experiment where the final stable phase formed above 150°C is Au_4Al if sufficient Au supply is available. Kirkendall's voids are observed to be segregated in between Au and Au_4Al , as seen in Fig. 15.2. In the area where Al has remained, the interface has two more intermediate layers, as shown in Fig. 15.3. Through detail

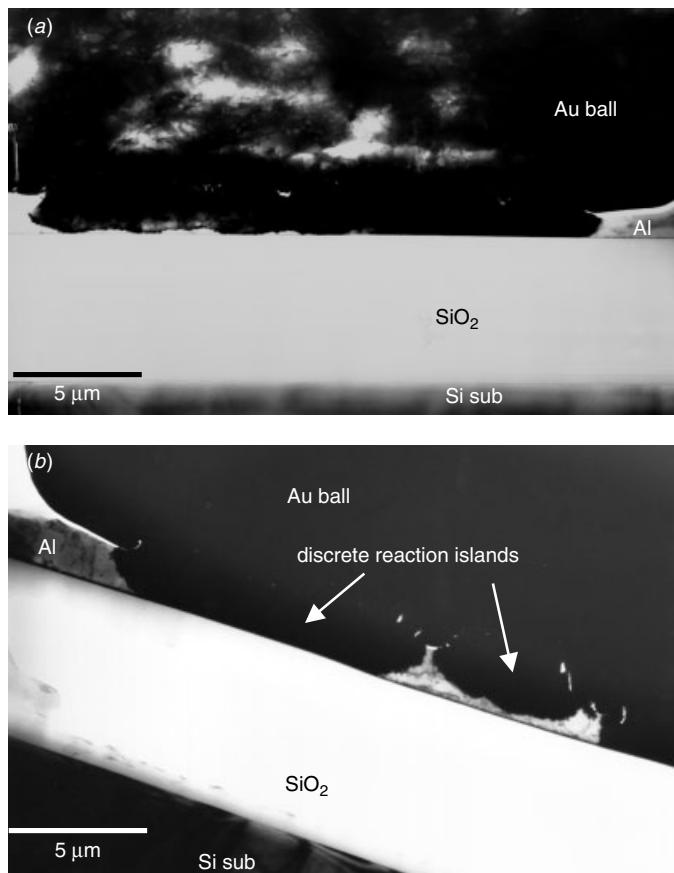


Figure 15.1 (a) Au ball bonding to the Al pad. The Au ball was thinned down to be partially transparent in some areas. (b) When the bonding condition changed, the interface morphology also changed. Discrete reaction islands are observed. (Tung et al., ISTFA 2002, 505–511, reprint with permission from ASM International)

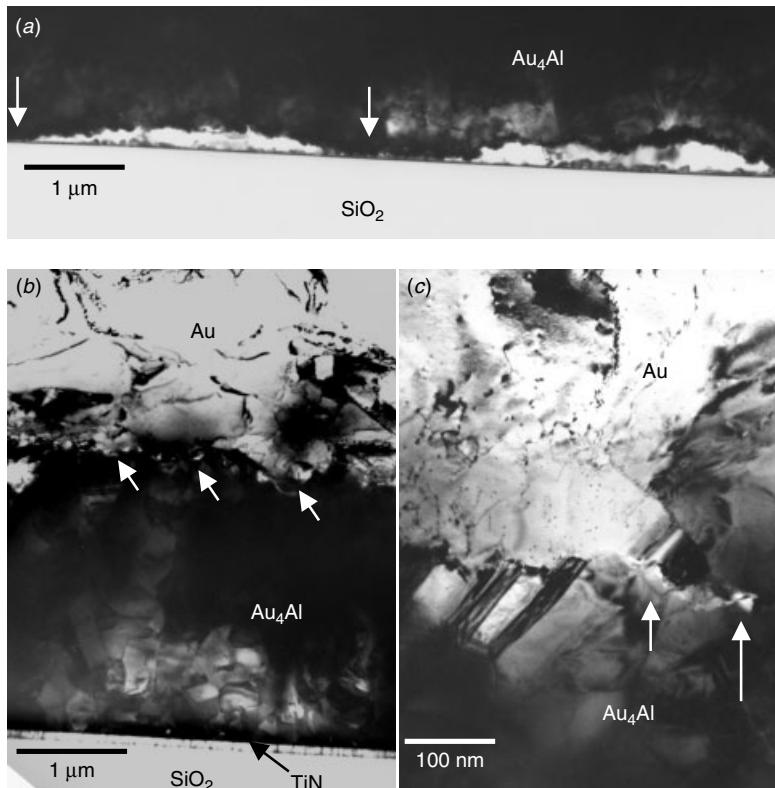


Figure 15.2 Au ball bonding to the Al pad. (a) In some areas, as indicated by arrows, Au has totally consumed the Al and reached the TiN at the bottom. (b, c) Bulk Au₄Al is formed where there is no Al left at the bottom. Kirkendall's voids are observed in between Au₄Al and Au, as indicated. (Tung et al., ISTFA 2002, 505–511, reprint with permission from ASM International)

EDS the two intermediate phases were determined to be Au₅Al₂ and Au₂Al. These two phases, again as observed in the conventional diffusion, must be two coexisting intermediate phases during diffusion.

When the Au ball bonding condition changes, for example, because of the substrate temperature or wire bonding parameters, the resulting interaction will also change. Closer examination of the interface of the Au₄Al and Au show a fairly smooth interface with Kirkendall's voids distributed along the interface. The overall thickness of this Au₄Al ranges from 1.7 to 2 times the original Al thickness. Interestingly, as Au₄Al hits TiN at the bottom of Al, the reaction is completed stopped by TiN, indicating that the TiN is an excellent stopper for Au from attacking anything below it, as seen in Fig. 15.4(a). A detailed analysis of the area where there is still Al left at the bottom has shown the intermediate phase to also alter (apart from interface morphology) as the ball bonding condition changes. AuAl is found to be the intermediate phase in this case, as shown in Fig. 15.4(b).

The conventional diffusion couple tells a different story, as illustrated in Fig. 15.5 (Campisano et al. 1970). The ideal diffusion couple shows Au₅Al₂ and Au₂Al to be

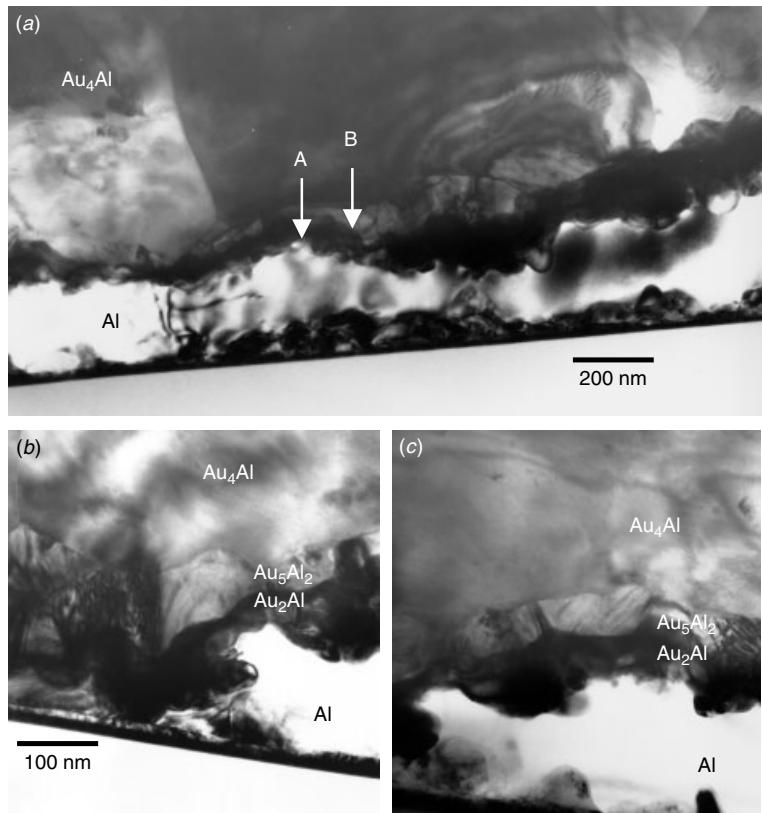


Figure 15.3 Au ball bonding to the Al pad. (a) Close examination reveals the existence of two intermediate layers, A and B, as indicated between the bulk intermetallic, Au_4Al , and pure Al phases. (b, c) EDS shows these layers to be Au_5Al_2 and Au_2Al , respectively. (Tung et al., ISTFA 2002, 505–511, reprint with permission from ASM International)

the first intermetallics formed and the reaction products gradually transformed from Au_2Al and Au_5Al_2 , into Au_4Al , depending on the temperature and reaction time. In comparing the idea case of a diffusion couple with that of wire bonding, a few points are to be noted:

- In areas where there is no Al left, Au_4Al is the final product phase, indicating that the local temperature within the wire bonding area cannot be less than 150°C .
- In areas where there is Al underneath, three intermetallics coexist, namely Au_4Al , Au_5Al_2 , and Au_2Al . Such coexistence is thermodynamically metastable, thus indicating that the system is not in equilibrium.
- Different wire bonding conditions create different intermediate phases, indicating that the reaction is highly sensitive to the wire bonding parameters, such as the substrate temperature, the transfer speed, and even the bonding force.

It is suspected that during the wire bonding, a temperature gradient builds up across the Au/Al interface. On the Au side, the temperature is much higher, since Au is melted

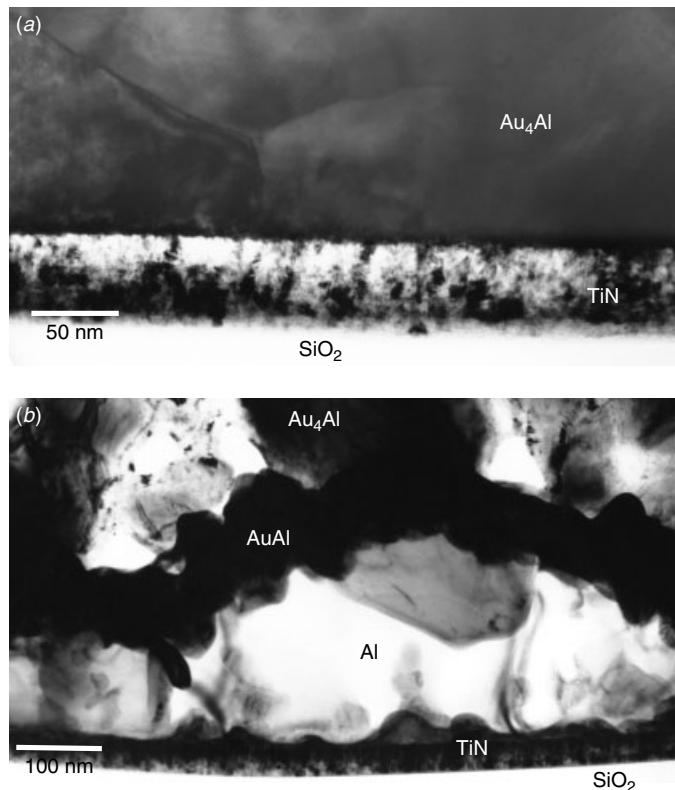


Figure 15.4 (a) Bulk Au_4Al is formed if no Al is left at the bottom. TiN at the bottom of Al has stopped the Au from further penetrating into the SiO_2 . (b) AuAl is found to be the intermediate phase, here between Al and Au_4Al . (Tung et al., ISTFA 2002, 505–511, reprint with permission from ASM International)

and aided by an ultrasonic power to form a ball at the wire bonder head (that is why this is called ball bonding). While transferring the Au ball in contact with the Al pad, Al is heated up from the Au side and the reaction is forced, starting from the Au/Al interface with highest temperature on the Au side. As the reaction continues, part of the heat is absorbed by the Al pad while the Si substrate acts as a heat sink where the interface temperature decreases rapidly. Subsequent intermetallics form under the swift temperature change, and the whole reaction stops when the temperature cools down below 100°C. The reaction is summarized in Fig. 15.6.

15.2 THE EUTECTIC PbSn SOLDER/SPUTTERED Cu/Ni(V)/Al SYSTEM

Ni-based under-bump metallization (UBM) is of interest in flip-chip technology because of a slow metallurgical reaction between the solder and Ni compared to Cu-based UBM systems (Liu et al. 2000; Kim et al. 1999). However, compared to the Cu system, the reaction between the solder and the Ni-based alloy is less understood (Jang et al. 1999; Mei et al. 1998).

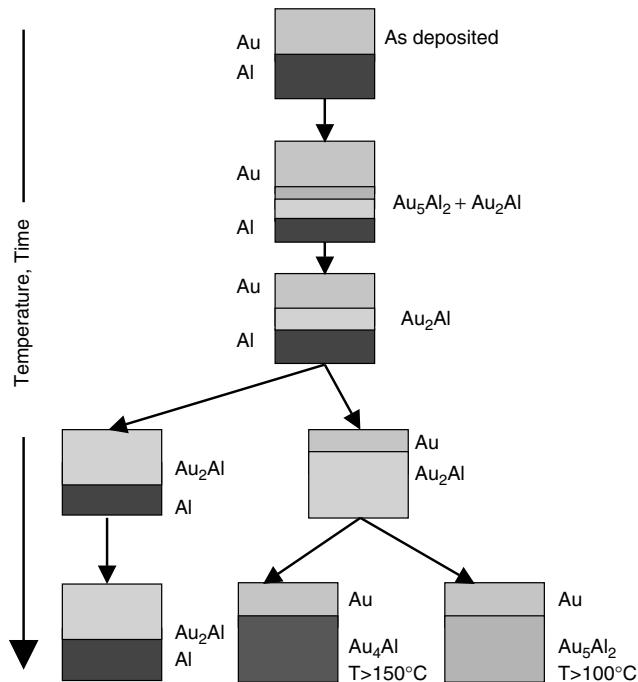


Figure 15.5 Schematic diagrams showing the phase formation in Au–Al thin film couples. Initially Au_5Al_2 and Au_2Al are formed. After the Al layer is consumed, as in wire bonding case, either Au_5Al_2 ($>100^\circ\text{C}$) or Au_4Al ($>150^\circ\text{C}$) is formed in the final stage. (Tung et al., ISTFA 2002, 505–511, Reprint with permission from ASM International)

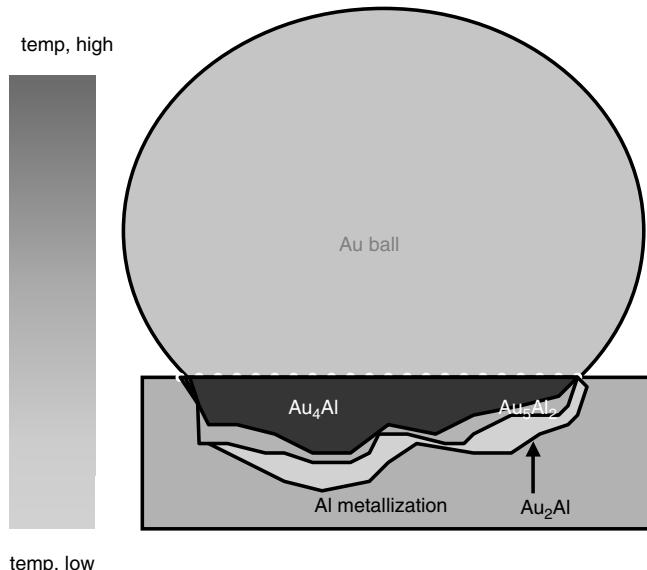


Figure 15.6 Schematic diagrams showing the phase formation in Au–Al wire bonding corresponding to the temperature gradient built up during the bonding process. (Tung et al., ISTFA 2002, 505–511, Reprint with permission from ASM International)

TABLE 15.1 Layered Structure of UBM Systems

UBM System	Sputtered Cu/Ni(V)/Al	Au/Electroless Ni(P)
Passivation	Silicon nitride	P-glass
Bumping method	Print and reflow	Print and reflow
UBM layer structure	Eutectic solder	Eutectic solder
	Cu 500 nm	Wetted Au
	Ni(V) 400 nm	Electroless
	Al 400 nm	Ni(P) 14 μm
	Al or SiN	Al or P-glass
	SiO ₂ or Si sub	SiO ₂ or Si sub

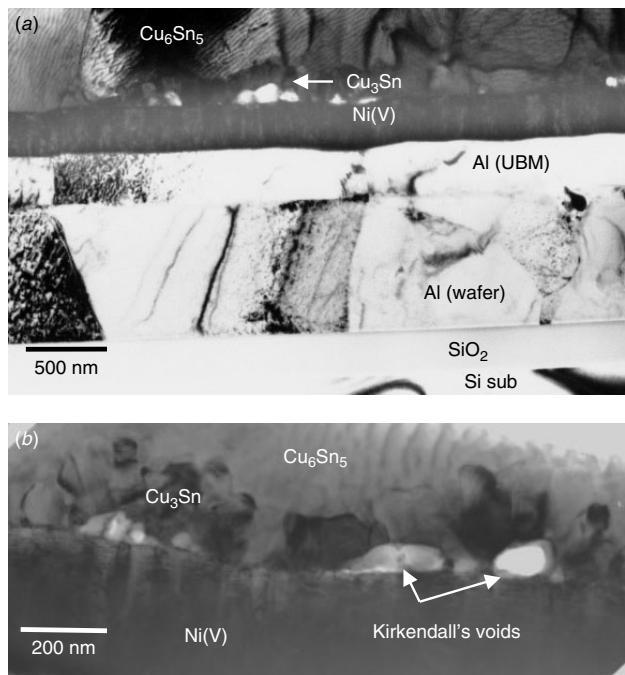


Figure 15.7 Eutectic PbSn solder/Cu/Ni(V)/Al UBM system in the as-reflowed condition. Thin Al was deposited immediately under the Ni(V) during the UBM deposition process. Cu₆Sn₅, Cu₃Sn and Kirkendall's voids are observed at the interface. Ni(V) has no observable damage by the solder. (Tung et al., ISTFA 2002, 505–511, reprint with permission from ASM International)

The basic thin film structures of the two different UBM systems to be discussed here are summarized in Table 15.1 (Tung et al. 2000, 2002). Cross-sectional TEM (XTEM) was used to study the samples that experienced different reflow cycles (as received, 1×, 3×, and 10×) as well as samples that were annealed at 150°C for 1000 hours. The TEM sample preparation was critical to the success of this analysis. One important concern was that the samples were highly temperature sensitive, so any prolonged exposure of

the sample in temperatures higher than 100°C would alter the microstructure. As our study showed that ion milling for more than 30 minutes will change the interface's microstructure, this was particularly true for the solder/Cu/Ni(V) system. For this reason special care was taken to ensure that the overall TEM sample preparation procedure was done at the temperature lower than 100°C and that the ion milling time was for less than 5 minutes. A liquid nitrogen cooled ion milling sample staging was used in some cases, although it was found not to be necessary.

The cross-sectional section TEM (XTEM) on the UBM pad without solder showed the original Cu coating thickness to be about 500 nm. In the as-reflowed sample, the Cu layer reacted to form Cu_6Sn_5 and Cu_3Sn compounds. The Cu_3Sn phase small grains aggregate at the interface between Ni(V) and large Cu_6Sn_5 grains. Kirkendall's voids accompanied and wrapped by the Cu_3Sn phase, as shown in Fig. 15.7. In some samples the residual Cu grains can still be observed below the Kirkendall's voids and right on top of the Ni(V) surface, as shown in Figs. 15.8 and 15.9. As reported previously, the concave nature of the $\text{Cu}_6\text{Sn}_5/\text{Cu}_3\text{Sn}$ interface strongly suggested the reaction is heading toward the Cu_6Sn_5 phase (Liu et al. 2000). Samples up to the 10× reflow did not show any major changes in their microstructures. Both Cu_6Sn_5 and Cu_3Sn remained after the 10 reflow cycles. The reason for the Cu_3Sn to survive after 10 reflow cycles may be that Cu does not have sufficient time to diffuse to the $\text{Cu}_6\text{Sn}_5/\text{Cu}_3\text{Sn}$ interface. However, two observable changes are noted. First, the Kirkendall's voids seem to grow and coalesce from 50 nm in the as-reflowed sample up to about 250 nm in the 10× sample. Second, there are two distinguishable new layers formed between Cu_3Sn and Ni(V), as shown in Fig. 15.10. EDS showed that the layers contain Cu, Sn, and Ni as the major components. The upper layer (layer 1) showed

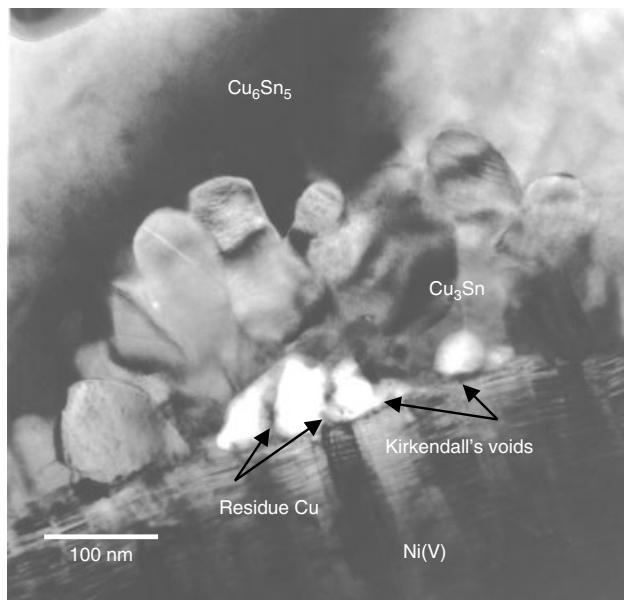


Figure 15.8 Eutectic PbSn solder/Cu/Ni(V)/Al UBM system in the as-reflowed condition. The close-up view shows Cu residue under the Kirkendall's voids, as indicated. (Tung et al., ISTFA 2002, 505–511, reprint with permission from ASM International)

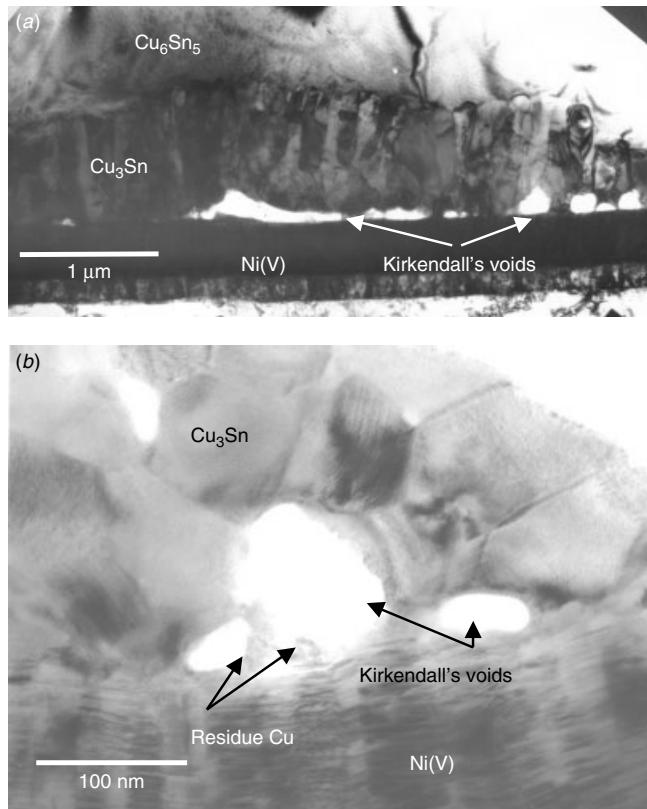


Figure 15.9 Eutectic PbSn solder/Cu/Ni(V)/Al UBM system in the as-reflowed condition. The close-up view shows Cu residue under the Kirkendall's voids, as indicated.

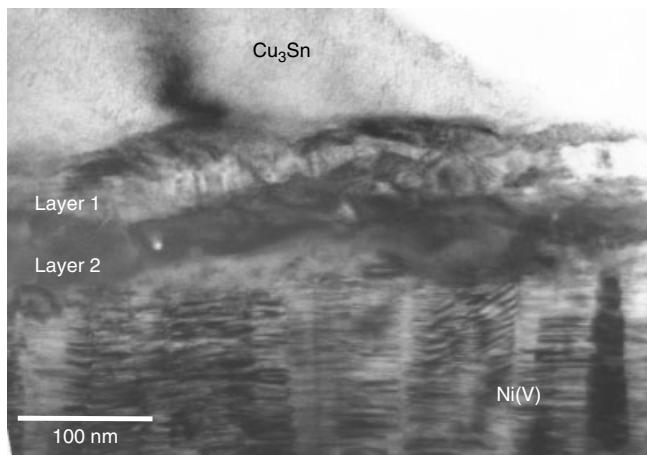


Figure 15.10 Eutectic PbSn solder/Cu/Ni(V)/Al UBM system after 10 reflow cycles. Two interfacial layers are observed: EDS on the top layer (1) shows that Sn : Cu : Ni is 32 : 66 : 2, and on the bottom layer (2) shows Sn : Cu : Ni is 53 : 33 : 14. (Tung et al., ISTFA 2002, 505–511, reprint with permission from ASM International)

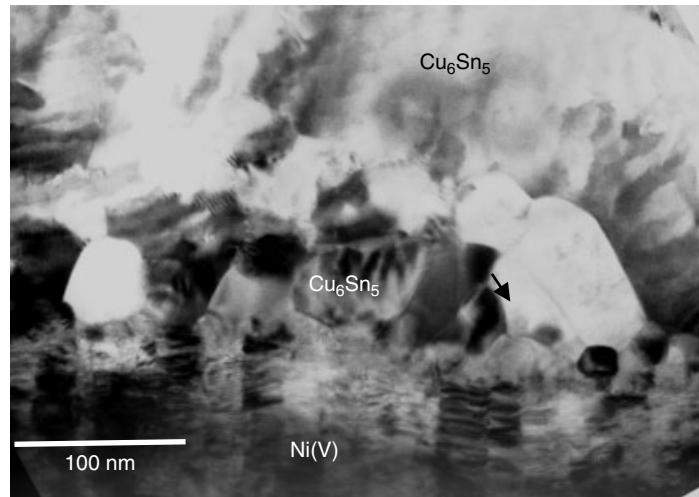


Figure 15.11 Eutectic PbSn solder/Cu/Ni(V)/Al UBM system after 1000 hours of 150°C annealing. No interfacial layer is observed. The small particles at the interface are identified to be Cu₆Sn₅.

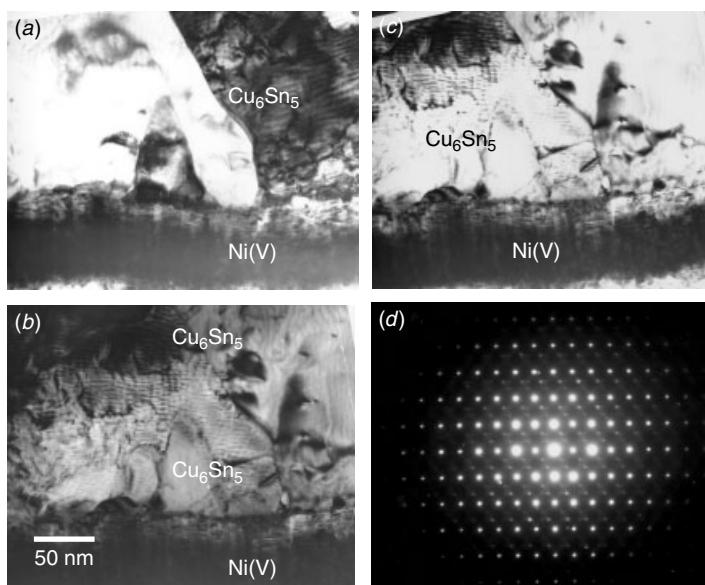


Figure 15.12 Eutectic PbSn solder/Cu/Ni(V)/Al UBM system after 1000 hours of 150°C annealing. Different illumination angles on the same area show different grain contrasts. The electron diffraction shows the small interfacial grains to have diffraction patterns identical to those of the large Cu₆Sn₅ grains.

$\text{Sn} : \text{Cu} : \text{Ni} = 32 : 66 : 2$ and the lower layer (layer 2) showed $\text{Sn} : \text{Cu} : \text{Ni} = 53 : 33 : 14$. But no positive phase identification was made from this study. For the sample with $150^\circ\text{C}/1000$ h high-temperature storage treatment, the interface microstructure is quite different. Figure 15.11 shows the TEM cross section of the sample. Only Cu_6Sn_5 is observed. There is no Cu_3Sn left anywhere at the interface. Detailed EDS and selective area electron diffraction showed the original Cu_3Sn small grains to have all transformed into Cu_6Sn_5 , as shown in Fig. 15.12. Furthermore, there are no Kirkendall's voids observed in the area. A similar result has been reported (Liu et al. 2000).

The disappearance of Cu_3Sn grains and Kirkendall's voids in the Cu/Ni(V)/Al system after high-temperature storage has been observed and explained by Liu et al. (2000). The observation is confirmed by the present study. The two additional layers observed in the sample after 10 reflow cycles imply that a reaction between Cu_3Sn and Ni(V) has started to take place. In the high-temperature storage sample, on the other hand, there is no additional layer, implying that there is no reaction between Cu_6Sn_5 and Ni(V). The difference in the Cu_3Sn and Cu_6Sn_5 reactions to Ni(V) is intriguing and requires further study. However, Cu_6Sn_5 is notorious for its scallop grain growth and will subsequently detach from the UBM interface and lead to the catastrophic solder bump's fall-off from the chip. The observation made here on the basic mechanism behind this problem should help in the possible solution.

15.3 THE EUTECTIC SOLDER/Au/ELECTROLESS Ni(P)/Al SYSTEM

The as-reflowed sample showed distinctive layers of Ni_3Sn_4 and Ni_3P at the solder/Ni(P) interface, as seen in Figs. 15.13 through 15.15. No Au or Au-related intermetallic was found anywhere near the interface. Electron diffraction on Ni(P) showed the film to be partially amorphous and partially crystalline (Ni face centered cubic phase), and P to be about 8 to 12 at% on average, as seen in Fig. 15.14. The two intermetallic layers identified as Ni_3Sn_4 and Ni_3P have different grain morphologies. Ni_3P is columnar with vertical gaps in between the grains; Ni_3Sn_4 is needle grain. It is believed that during the reflow operation, Ni reacts with Sn to form Ni_3Sn_4 . When surface Ni is depleted, Ni within Ni(P) outdiffuses to supply the Ni_3Sn_4 reaction. The result is that Ni(P) is depleted by Ni and forms Ni_3P when it reaches the 3 : 1 ratio. The vertical gaps in Ni_3P reflect a mechanism similar to Kirkendall's void formation. Since the reaction takes place in conjunction with the liquid phase solder, formation of vertical gaps in Ni_3P and a needle morphology of Ni_3Sn_4 can be expected (Tung et al. 2000, 2002). Both Ni_3P and Ni_3Sn_4 showed increases in thickness up to 10 reflow cycles, and Ni_3Sn_4 gradually transformed from a needle to an equiaxial grain structure, as shown in Fig. 15.16. The sample after $150^\circ\text{C}/1000$ h annealing showed a similar trend but at a slightly different growth rate in Ni_3P and Ni_3Sn_4 . The thickness of Ni_3Sn_4 measured after annealing is 2 μm corresponding to a reflow cycle of around 8 cycles, while the thickness of Ni_3P measured after annealing is 290 nm corresponding to a reflow cycle of only 2 cycles, as seen in Table 15.2 and Fig. 15.17. Also noticed is that the interface between Ni_3P and Ni_3Sn_4 begins to change after the additional thermal treatment. A thin crystalline layer is observed in both samples after 10 reflow cycles and 1000 hours of 150°C annealing, Fig. 15.18. EDS showed it to contain Ni, Sn, and P, but no positive phase identification could be made.

TABLE 15.2 Summary of the Ni(P) UBM System

	0×	1×	3×	10×	1000 h HTS
Ni ₃ Sn ₄	Needlelike 0.5 μm × 75 nm	Needlelike 1.3 μm × 160 nm	Scalloplike Thx > 0.9 μm	Grainlike Thx > 2.5 μm	Grainlike Thx ~ 2 μm Columnar + gap
Ni ₃ P	Columnar + gap Thx 180 nm	Columnar + gap Thx 260 nm	Columnar + gap Thx 350 nm	Columnar + gap Thx 450 nm	Thx 290 nm 160 nm Ni ₃ Sn ₄ + Ni ₃ P
Ni(P)	Thx 6 μm amorphous + c				
Ni(P)/Al	Penetration	Penetration	Penetration	Penetration	Penetration
Al	0.62 μm (68%) c-layer thx 37 nm 0.92 μm	0.48 μm (60%) c-layer thx 32 nm 0.82 μm	0.50 μm (60%) c-layer thx 33 nm 0.82 μm	0.58 μm (63%) c-layer thx 25 nm 0.92 μm	0.66 μm (66%) c-layer thx 36 nm 1 μm

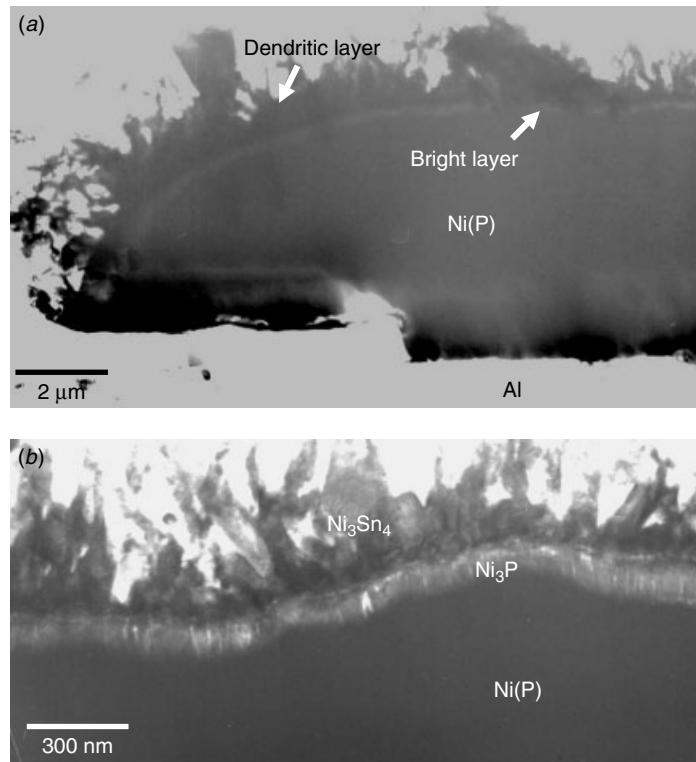


Figure 15.13 Eutectic PbSn solder/Au/Ni(P) UBM system in the as-reflowed condition. (a) The solder/Ni(P) interface shows a needle-grain layer and a bright layer, as indicated. (b) The solder/Ni(P) interface shows both Ni₃P and Ni₃Sn₄ intermetallic layers. (Tung et al., ISTFA 2002, 505–511, reprint with permission from ASM International)

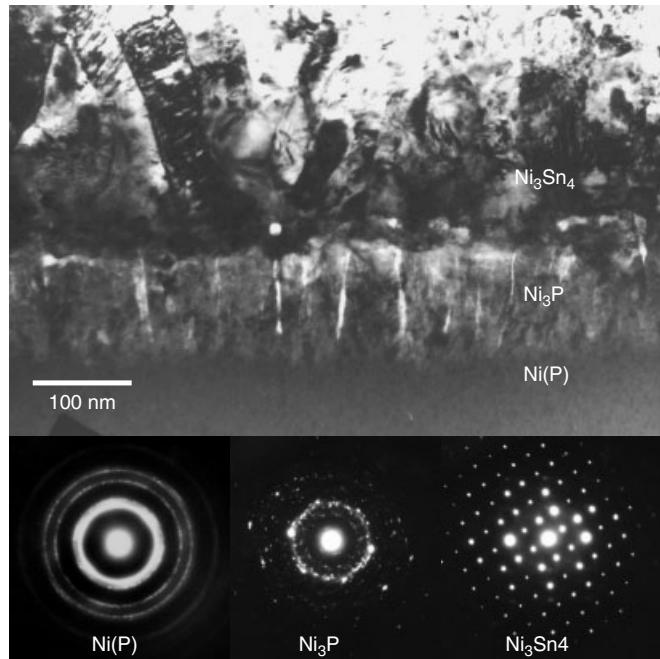


Figure 15.14 Eutectic PbSn solder/Au/Ni(P) UBM system in the as-reflowed condition. The solder/Ni(P) interface shows both Ni_3P and Ni_3Sn_4 intermetallic layers. Ni_3P showed columnar grain structure while Ni_3Sn_4 shows needle grains. The electron diffraction shows the positive phase identification of Ni_3Sn_4 , Ni_3P , and the pure Ni phase within Ni(P). (Tung et al., ISTFA 2002, 505–511, reprint with permission from ASM International)

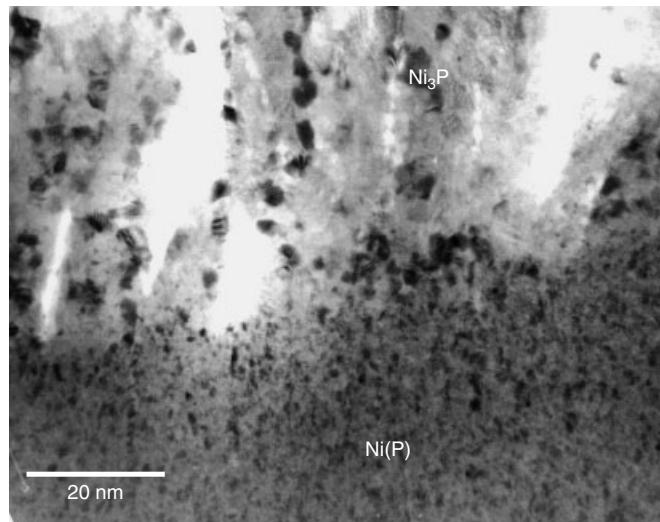


Figure 15.15 Eutectic PbSn solder/Au/Ni(P) UBM system in the as-reflowed condition. The close-up at the Ni_3P and $\text{Ni}(\text{P})$ interface shows that the columnar Ni_3P contains vertical gaps at the grain boundaries.

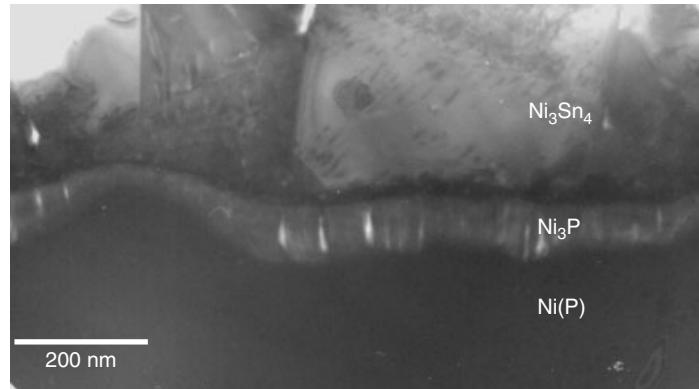


Figure 15.16 Eutectic PbSn solder/Au/Ni(P) UBM system after 10 reflow cycles. The Ni_3Sn_4 transformed into a more equiaxial and less needle-like morphology. The vertical gaps within Ni_3P remain.

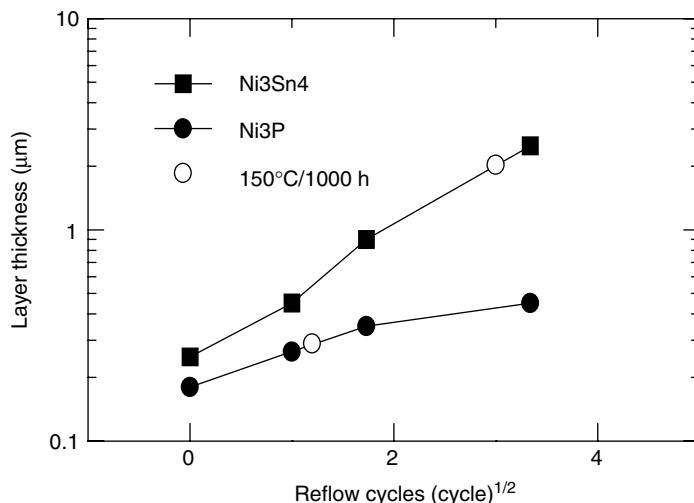


Figure 15.17 Linear thickness growth of Ni_3Sn_4 and Ni_3P versus the reflow cycles, but not in the same ratio for Ni_3Sn_4 and Ni_3P between the reflow cycles and high temperature storage annealing at 150°C . (Tung et al., ISTFA 2002, 505–511, reprint with permission from ASM International)

Another interface of importance is the interface between Ni(P) and Al pad. Figure 15.19 shows the TEM cross section images over this interface. Because of surface treatment before the electroless deposition of Ni(P), the Al surface was etched with irregular pitting, as seen in Figs. 15.19 through 15.21. A crystalline layer containing Zn, Cu, O, Al, and Ni was observed between the Ni(P) and the Al metal, and wrap around the etching pits, as seen in Fig. 15.21(a) and (b), while Ni(P) filled in the damage pits and formed seam gap at the center of the pits, Fig. 15.21(c) and (d). Clearly, the pits were there before the Ni(P) was deposited. It is believed that these

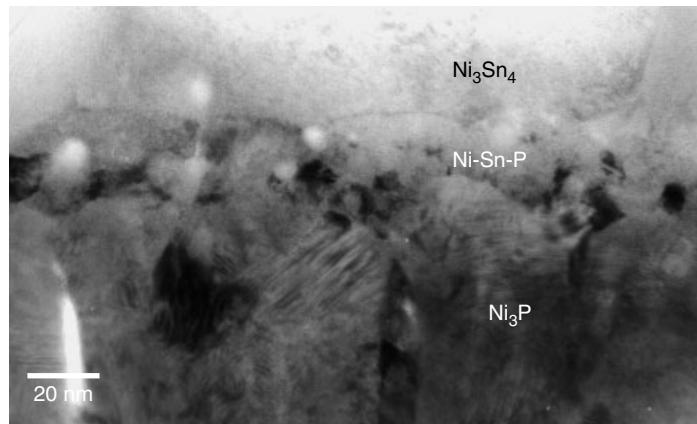


Figure 15.18 Eutectic PbSn solder/Au/Ni(P) UBM system after 10 reflow cycles. A very thin crystal layer, about 20 nm thick, was found at the Ni_3P to Ni_3Sn_4 interface.

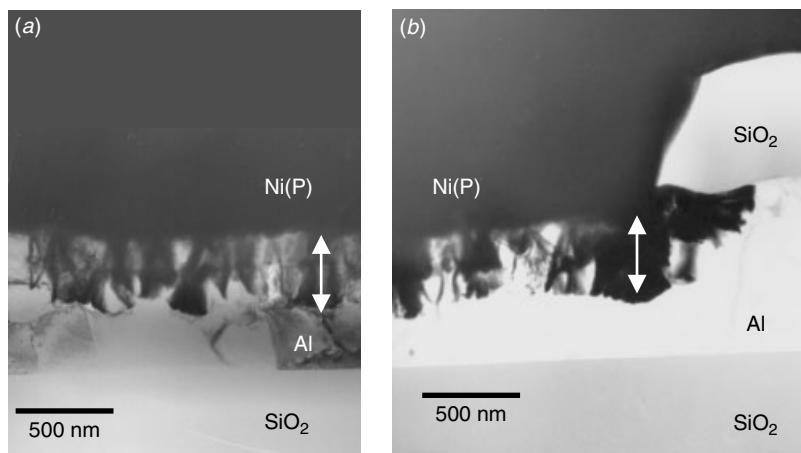


Figure 15.19 Eutectic PbSn solder/Au/Ni(P) UBM system in the as-reflowed condition. The Ni(P) to the Al metallization interface shows a severe attack of Al by Ni(P). The penetration depth can be as deep as 70% of the thickness of Al.

pits were formed during the Al surface treatment process (activation) before the Ni(P) deposition. The interface between Ni(P) and Al did not show any change during the reflow cycles and high-temperature storage annealing treatment. The only change was that the crystalline phase grain size seemed to grow a bit, but the change is subtle, Fig. 15.22. The damage and pitting do not show any degradation after the thermal treatments. Another interesting issue associated with the electroless Ni(P) UBM system is that the Ni(P) does not attach to phosphorous glass passivation very well and becomes, mostly delaminated at the passivation/Ni(P) interface. The present study confirms such an interface delamination. Figure 15.23 shows the interface between Ni(P) and passivation SiO_2 . At least three thin crystalline layers are observed in the as-reflowed

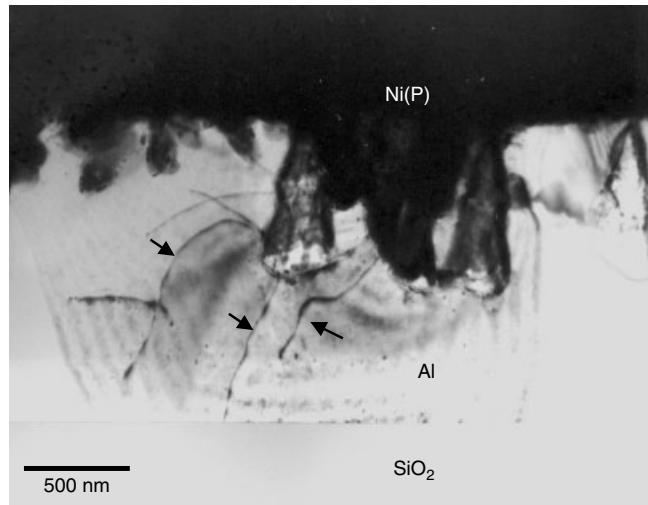


Figure 15.20 Eutectic PbSn solder/Au/Ni(P) UBM system in the as-reflowed condition. The close-up at the Ni(P) to Al metallization interface damage shows dislocations initiated from the pitting and extended deep into the Al, as indicated.

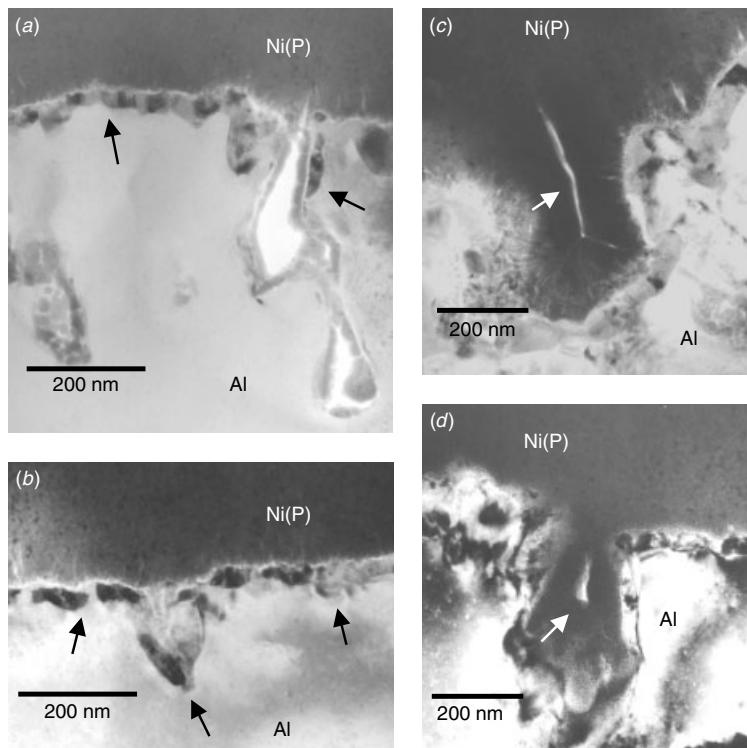


Figure 15.21 Eutectic PbSn solder/Au/Ni(P) UBM system in the as-reflowed condition. (a, b) The close-up at the Ni(P) to Al metallization interface pittings shows a thin crystalline layer containing Zn, Cu, O, Al, and Ni. The crystalline layer has wrapped around most of the pitting damages. (c, d) Ni(P) has filled the pits and formed a seam gap in the middle.

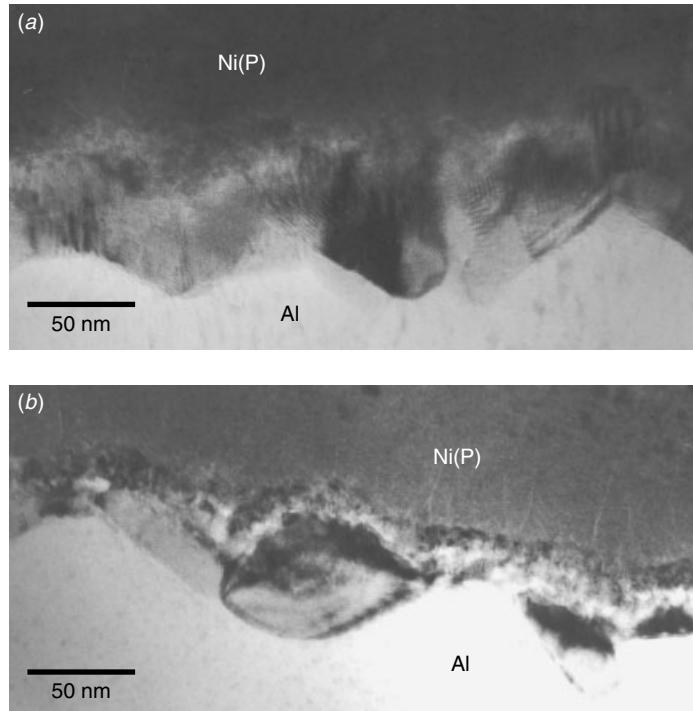


Figure 15.22 Eutectic PbSn solder/Au/Ni(P) UBM system after 1000 hours at 150°C annealing. The close-up at the Ni(P) to Al metallization interface shows the crystalline layer grain size to have remained unchanged after 1000 hours of high-temperature storage.

sample. The innermost one was found to contain Au and Ni, the middle one is Au rich and outer layer has Au and Sn, indicating that the interface detachment existed at least before or during the Au wetting layer deposition. During reflow the liquid solder further attacked this interface and formed Au–Sn intermetallics on the outside of the Au coating. After 10 reflow cycles over 1000 hours of 150°C annealing, the Au coating completely disappeared, so the only crystalline layer observed is Ni_3Sn_4 in Fig. 15.23(b). The question is, Where does all of the Au coating go? Within the solder and near the Ni(P) bump, dispersed Au-rich precipitation particles were found, as shown in Fig. 15.24. It is generally thought that the Au dissolved into the liquid solder during reflow and re-precipitated as Au–Sn intermetallic particles within the Sn-rich phase after solidification, as we have observed here.

The thickness growth of Ni_3Sn_4 and Ni_3P seems to be linear with the reflow cycles, as was shown in Fig. 15.17. However, compared with the high-temperature storage sample, the thickness ratio between Ni_3Sn_4 and Ni_3P does not seem to be in the same proportion among the reflow samples and the high-temperature storage samples. If we assume Ni and P to be conserved in the soldering and annealing reaction (Jang et al. 1999), such disproportion in Ni_3Sn_4 and Ni_3P thicknesses may be due to a difference in diffusivity between the solder reaction and the high-temperature storage annealing reaction. Furthermore the thickness of Ni_3P in the high-temperature storage sample, compared with that calculated from the published diffusivity data, should be far thicker

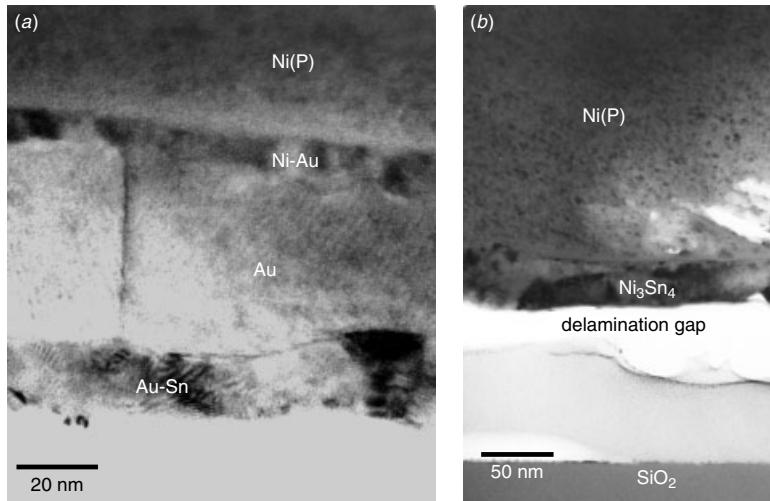


Figure 15.23 Eutectic PbSn solder/Au/Ni(P) UBM system. (a) The sample in the as-reflowed condition, where the Ni(P) to SiO₂ interface shows at least three crystalline layers. The innermost layer contains Au and Ni. The middle layer is mostly Au, and the outer layer shows Au and Sn. (b) The sample after 1000 hours of 150°C annealing, where the Ni(P) to SiO₂ interface shows that only a thin Ni₃Sn₄, has remained. The delamination gap was partially refilled by epoxy during the TEM sample preparation.

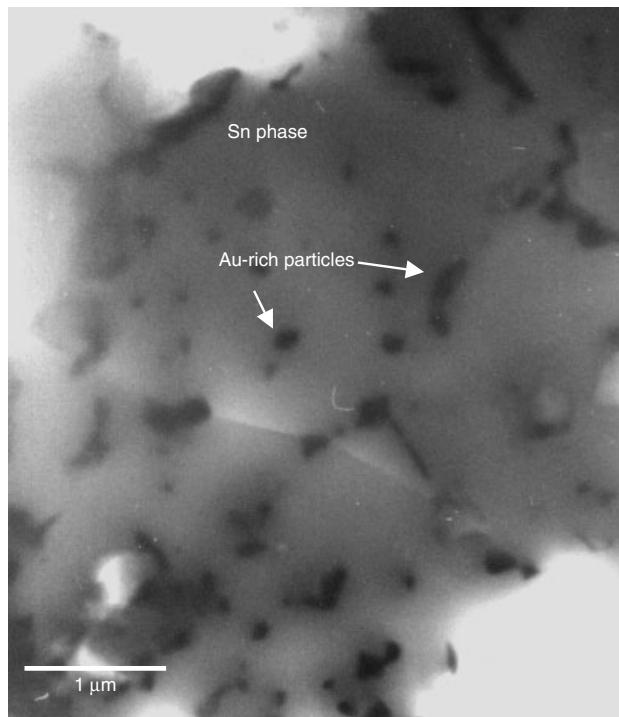


Figure 15.24 Eutectic PbSn solder/Au/Ni(P) UBM system in the as-reflowed condition. A close-up at the Sn-rich phase within the solder and near the Ni(P) bump shows dispersed Au-rich particles.

than what we measured in this study. Such a discrepancy implies that the growth of Ni_3P was controlled by the formation of Ni_3Sn_4 and thus by the supply of Sn from the solder phase. When in solid state such as in this case at 150°C, the Sn supply stops after a thick Pb-rich layer is formed near the interface. This stops the formation of Ni_3Sn_4 and the Ni_3P as well. The present work suggests that such an interruption may occur at the very early stage of the annealing. This makes the electroless Ni(P) system a particularly robust UBM system against thermal annealing (Tung et al. 2000, 2002).

15.4 THE EUTECTIC SOLDER/Cu/Cu-Cr/Cr UBM SYSTEM

This UBM system has remained one of the oldest and most studied systems (Fried et al. 1982). A special process is used to produce the Cu–Cr layer with phase-in structure to provide an additional interlocking mechanism. The system seems to be more suitable

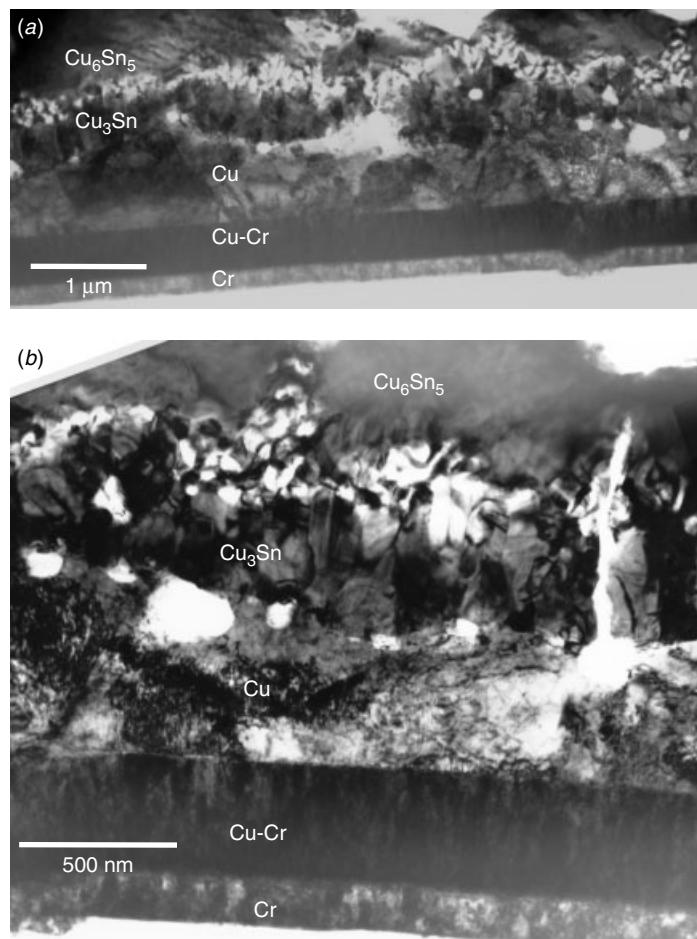


Figure 15.25 Eutectic solder/Cu/Cu—Cr/Cr UBM system in the as-reflowed condition. Cu_6Sn_5 , Cu_3Sn , and Kirkendall's voids are observed.

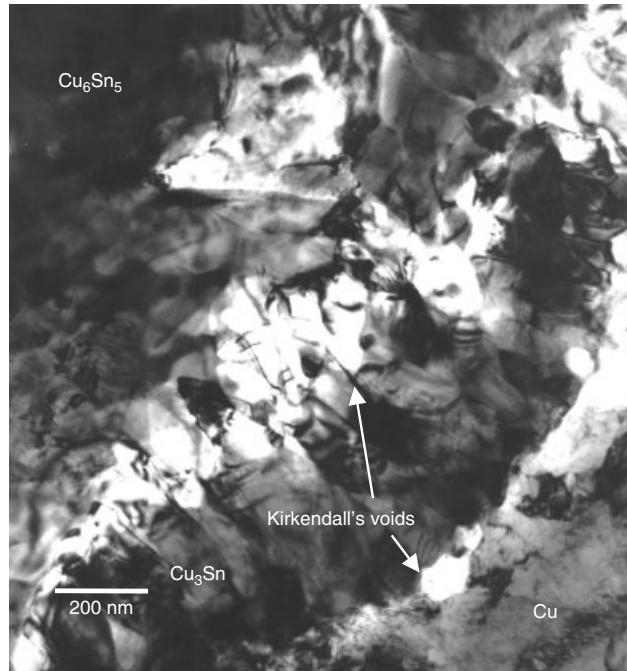


Figure 15.26 Eutectic solder/Cu/Cu—Cr/Cr UBM system in the as-reflowed condition. Kirkendall's voids are observed in between both Cu₆Sn₅/Cu₃Sn and Cu₃Sn/Cu interfaces.

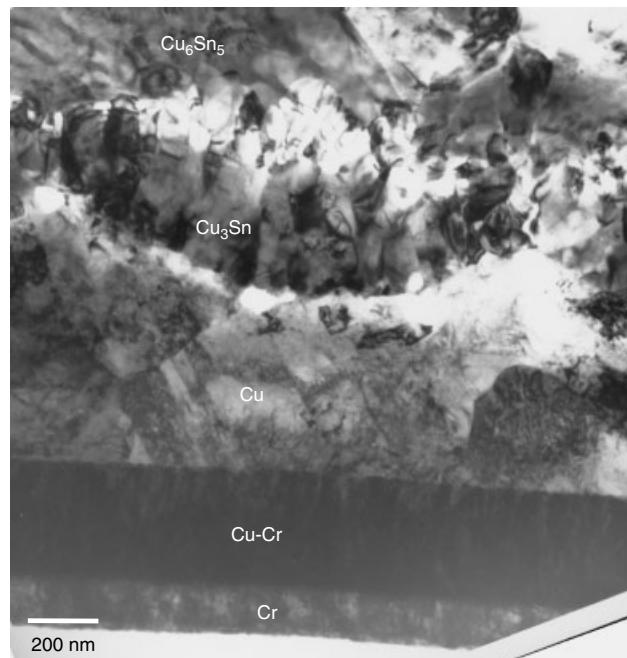


Figure 15.27 Eutectic solder/Cu/Cu—Cr/Cr UBM system in the as-reflowed condition. The morphology of the Cu₃Sn layer seems to consist of islands.

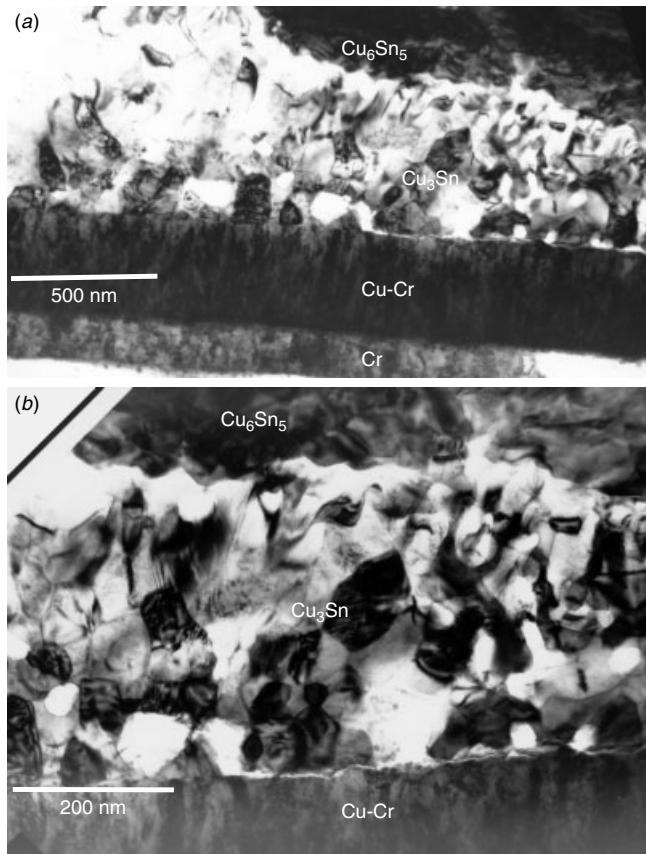


Figure 15.28 Eutectic solder/Cu/Cu—Cr/Cr UBM system in as-reflowed condition. Where there is no Cu residue, the Cu₃Sn is in direct contact with Cu—Cr.

for high Pb solder. Here the analysis will demonstrate why such a system works better with high Pb solder.

In the as-reflowed condition the thick Cu layer originally on top reacts with the Sn from the solder and forms Cu₆Sn₅ large grains on top and Cu₃Sn fine columnar grains in the middle, as seen in Fig. 15.25. While depending on the original Cu thickness and reflow temperature profile, a certain amount of Cu remains. Other than the intermetallics Kirkendall's voids are also observed. Interestingly there are two distinctive Kirkendall's void aggregation layers. The layer in between Cu₆Sn₅ and Cu₃Sn shows a much smaller size, about 50 to 100 nm in diameter, higher density, and clustering. The group between Cu₃Sn and residue Cu is bigger in size, and can be as large as 200 nm but much fewer in number, as seen in Fig. 15.26. Another interesting observation is the layer morphology of Cu₃Sn. The Cu₃Sn layer is not uniform but forms an islandlike structure with the large Kirkendall's voids sitting in between the Cu₃Sn islands. This is best seen in Figs. 15.25(a) and 15.27. The islands seem convex toward the residue Cu, implying that the Cu will be consumed in the end provided that sufficient thermal energy is supplied. Indeed, there are areas where no Cu is left and the Cu₃Sn is in direct

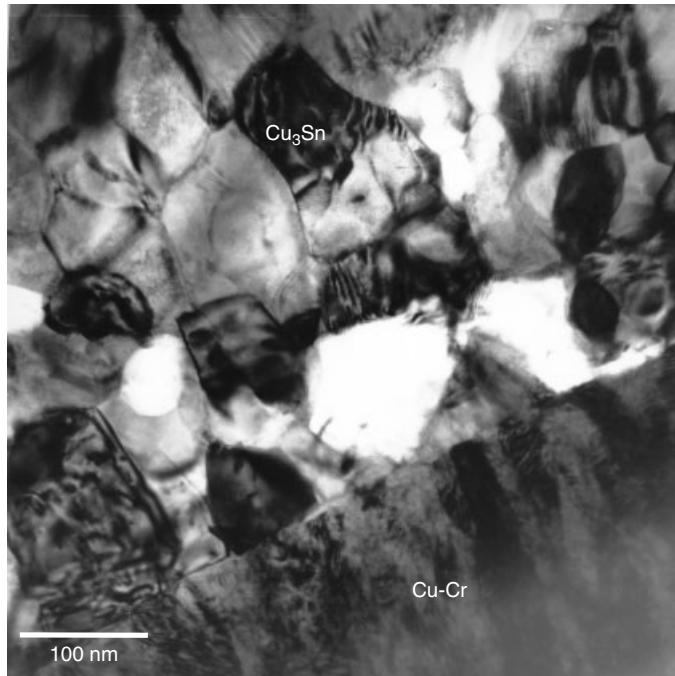


Figure 15.29 Eutectic solder/Cu/CuCr/Cr UBM system in the as-reflowed condition. A close-up at the Cu_3Sn to Cu-Cr interface shows no physical defects nor additional layers in between.

contact with Cu–Cr UBM, as in Fig. 15.28. A close-up of the interface between Cu_3Sn and Cu–Cr shows no apparent physical defects nor additional layers at the interface, Fig. 15.29. Notice in Figs. 15.28 and 15.29 that the Cu_3Sn layer seems to intermix with the Kirkendall void and forms a porous skeletal structure. The Cu–Cr layer remains intact after the reflow whether or not there is Cu left at the interface. The Cu within Cu–Cr is stable and does not react with Sn to form an intermetallic, at least in the as-reflowed condition.

Also observed is a mechanical crack, which initiates from a lower layer of Kirkendall's void, propagates through Cu_3Sn layer, and stops within Cu_6Sn_5 , Fig. 15.30. The source of this mechanical stress is unknown. TEM sample preparation could have introduced stress and triggered the vertical crack.

It is thought that as Cu reacts with Sn, Cu is the diffusing species and Kirkendall's voids must sit on the Cu side of the diffusion couple. However, during a reflow condition the situation is different from an ideal diffusion couple. There is a liquid phase involved, and depending on the reflow process, the sample heating may not be homogeneous. From this analysis it is quite obvious that in a certain area, a thick Cu layer has survived while in other area, no pure Cu remains. The morphology of the Cu_3Sn layer and the inhomogeneous distribution of Kirkendall's voids may be due to the reflow operation.

The Cu/Cu–Cr/Cr UBM system can react with eutectic solder to an extend where all of the top layer Cu is consumed during the initial reflow. The porous Cu_3Sn layer may cause stress and become reliability concern. On the other hand, if a high-Pb solder

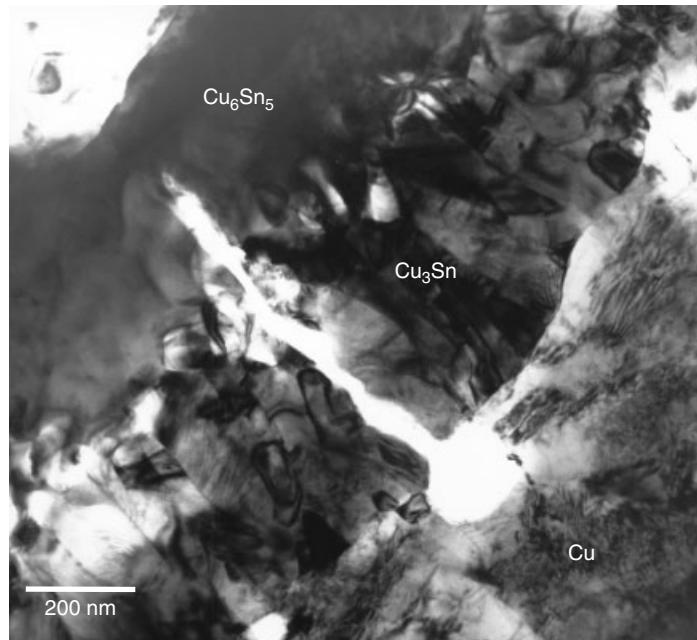


Figure 15.30 Eutectic solder/Cu/CuCr/Cr UBM system in the as-reflowed condition. A mechanical stress can initiate a crack. Here the crack has nucleated from a Kirkendall's void and penetrated through the Cu₃Sn. The crack has stopped within Cu₆Sn₅ in this instance.

is used with the Cu/Cu–Cr/Cr UBM system, the situation slightly changes. The supply of Sn in the high-Pb solder can be limited, and the remaining Cu will protect the Cu–Cr underneath. It is thought that the Cu/Cu–Cr/Cr UBM system performs better with the high-Pb solder than with the PbSn eutectic or Pb-free solder.

15.5 THE EUTECTIC SOLDER/Cu/W(Ti) UBM SYSTEM

Tungsten (W) alloys have a long history of being an ideal diffusion barrier. From the thin film ULSI process to the tape automatic bonding technology (TAB) Au bump process, W alloys have served as the preferred barrier substance (Fried et al. 1982). However, in the UBM process, a thick layer of Cu is applied on top of the W(Ti) alloy (with Ti around 9–10 at%), to provide the solderability. Thus the question is not whether W(Ti) is a good barrier but rather how the Cu layer will behave during the soldering processes, reliability tests, and application environment.

A first guess suggests that in the as-reflowed condition, the sample should look very similar to that of the eutectic solder/Cu/Cu–Cr/Cr UBM system since both have a very thick Cu on top. This is confirmed by TEM results. Figures 15.31 and 15.32 show a TEM cross-sectional view on the interface near the residue pure Cu. From the SEM images, it is clear that Cu₆Sn₅ and Cu₃Sn have formed. In TEM we observe that the interface between Cu₃Sn and the residue Cu contains a dispersed layer of Kirkendall's voids. The void size is 30 to 100 nm on average. Within the void-containing layer

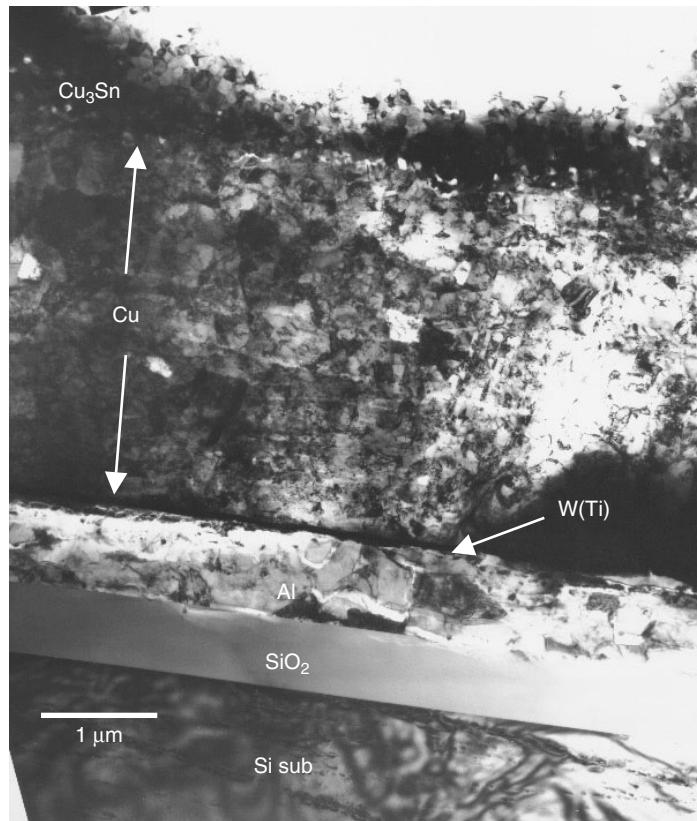


Figure 15.31 Eutectic solder/Cu/W(Ti) UBM system. The sample in the as-reflowed condition shows a layer of Cu₃Sn formed in between the Cu₆Sn₅ and pure Cu.

the materials are identified to be either Cu₃Sn or pure Cu phases. An interesting question is here the morphology of Kirkendall's voids between Cu₃Sn and pure Cu seems to be more like the one found in between Cu₆Sn₅ and Cu₃Sn as seen in the eutectic solder/Cu/Cu–Cr/Cr UBM system. The differences in the Kirkendall morphology seem to imply a difference in the interdiffusion behavior and thermal environment encountered by the two systems. Further study is required to get some understanding of such a difference.

The W(Ti) diffusion barrier layer was identified by EDS to be the W(Ti) alloy with W:Ti = 92:8. Such a composition is in the single phase region within the W–Ti binary phase diagram. The TEM bright field, dark field, and electron diffraction studies of this film show the film to be composed of large W(Ti) grains ranging in size up to 200 nm in diameter. Within the W(Ti) grains, subgrain microstructural defects are observed, as shown in Fig. 15.33. The interfaces between W to Al and W to Cu are surprisingly smooth and clean, indicating a film deposition process that is well controlled.

The sample after a few reflow cycles and high-temperature storage annealing shows astonishing microstructural changes. For samples with 10 reflows, there is no pure Cu left in the system. All of the Cu has converted into Cu₃Sn intermetallic phase,

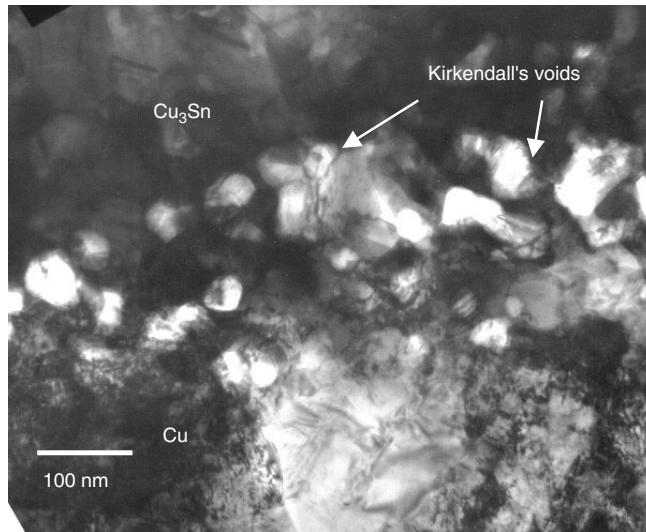


Figure 15.32 Eutectic solder/Cu/W(Ti) UBM system. The sample is in the as-reflowed condition. A layer of aggregated Kirkendall's voids has formed in between the Cu₃Sn and pure Cu.

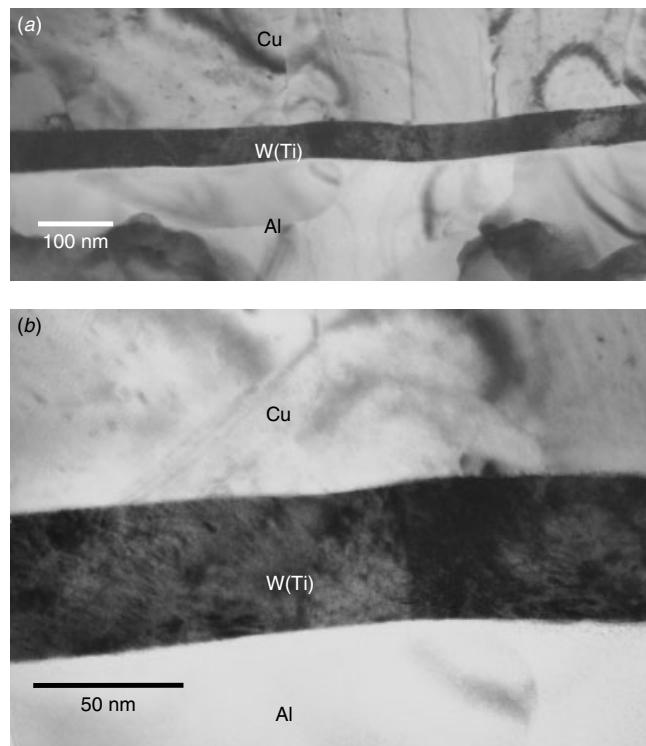


Figure 15.33 Eutectic solder/Cu/W(Ti) UBM system. The sample is in the as-reflowed condition. The W(Ti) layer was identified by EDS to be W:Ti = 92:8. The interface between Cu, W, and Al is extremely clean and smooth with no intermediate layers in between. A Close-up of the W(Ti) film shows large grain sizes.

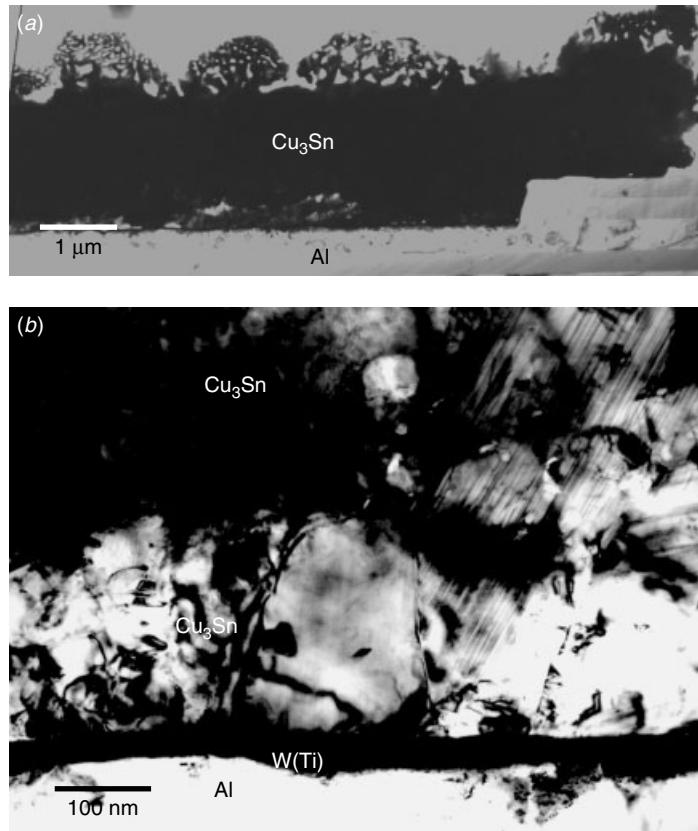


Figure 15.34 Eutectic solder/Cu/W(Ti) UBM system. The sample after 10 reflow cycles shows only the Cu₃Sn phase throughout the whole cross-sectional interface. No pure Cu is left in the sample.

Figs. 15.34 and 15.35. Such a dramatic change in microstructure has been known to be accompanied by mechanical property changes. Indeed, the mechanical shear and pull tests all show a fracture interface on top of the W(Ti) UBM layer (Teo et al. 2000). In addition the Cu₃Sn phase showed two different grain morphologies. One is the normal grain structure, as seen in Fig. 15.35. The other is a coral-like porous structure in the TEM micrographs, which is thus termed the “coral” phase. The coral phase in this sample is mostly found on top. Both EDS and electron diffraction have confirmed the coral phase to be Cu₃Sn, Fig. 15.36. Detailed microstructural studies of the coral phase and its interface with the bulk Cu₃Sn phase have shown that the coral Cu₃Sn are crystallographically continuous with the bulk Cu₃Sn, implying that the coral phase may be extending and growing from the bulk Cu₃Sn phase, as seen in Fig. 15.37. The origin of the coral phase is believed to be, at least partially, associated with TEM sample preparation artifacts as explained below.

As suggested by Anderson (2001), the Pb-rich phase in solder may corrode and dissolve in water during the TEM sample preparation using mechanical polishing where water is used as a lubricant and cooling agent. As a result the eutectic solder may

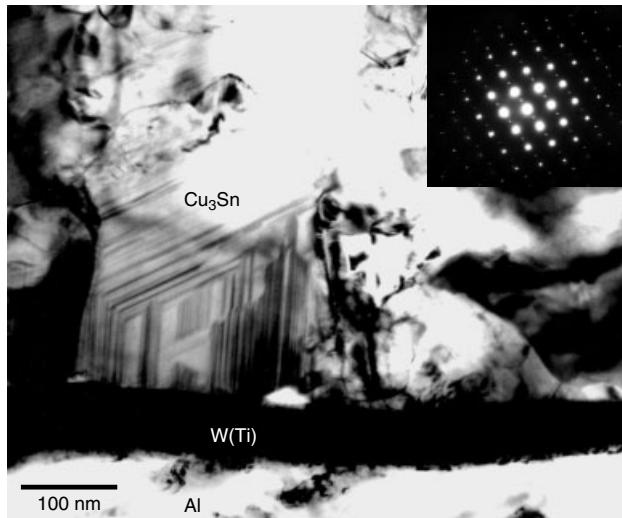


Figure 15.35 Eutectic solder/Cu/W(Ti) UBM system. The sample after 10 reflow cycles shows only the Cu_3Sn phase throughout the cross-sectional interface. The electron diffraction pattern confirms the intermetallic present near the W(Ti) interface to be Cu_3Sn .

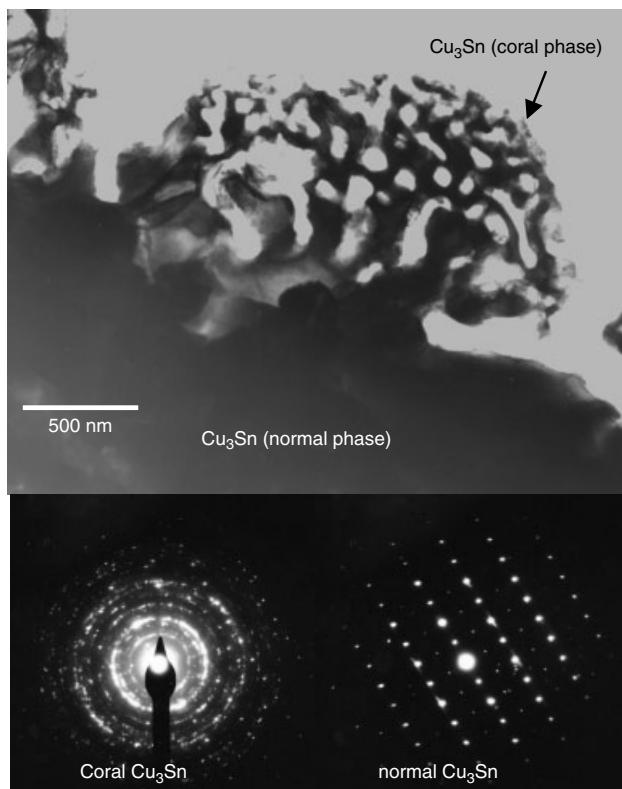


Figure 15.36 Eutectic solder/Cu/W(Ti) UBM system. The sample after 10 reflow cycles shows only the Cu_3Sn phase throughout the cross-sectional interface. Both EDS and the electron diffraction pattern have confirmed the coral phase to be Cu_3Sn .

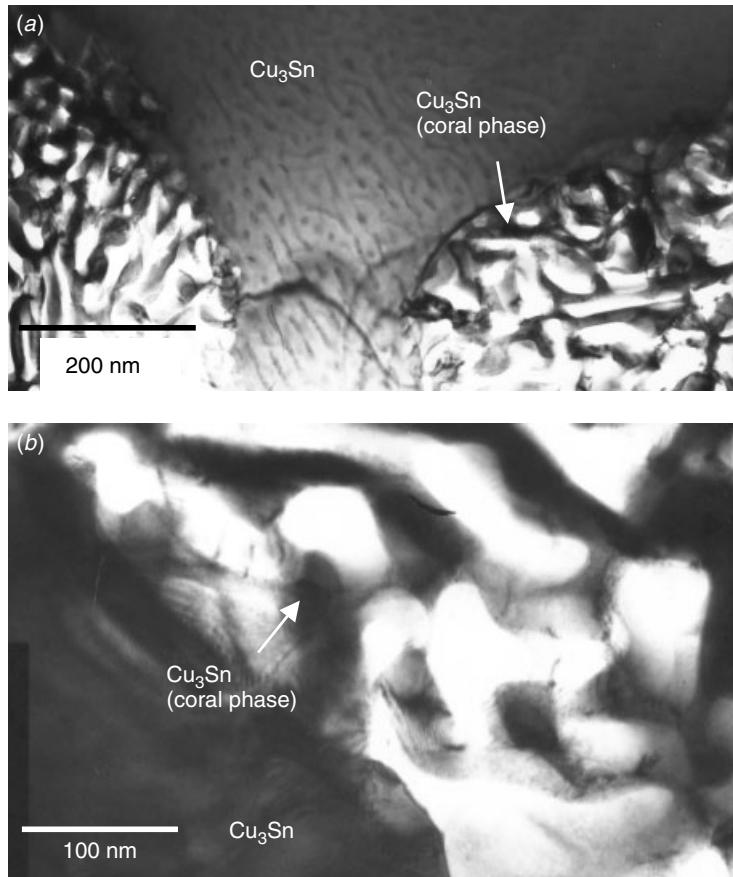


Figure 15.37 Eutectic solder/Cu/W(Ti) UBM system. The Cu₃Sn sample after 1000 hours of 150°C annealing. The interface between the coral Cu₃Sn and the normal Cu₃Sn shows no distinctive boundary. Some of the coral skeleton is even extended from the normal grain area. Absent the grain boundary, the coral grain likely nucleated and developed from the normal grain structure.

be depleted of the Pb-rich phase (and left with the Sn-rich phase) after mechanical polishing but before ion milling. As is known, the eutectic structure is a laminated structure with two alternative mutually undissolvable phases (miscibility gap) physically in contact with each other; the thin TEM sample slice at the end of mechanical polishing then results in Sn-rich skeletal layers without the Pb-rich phase. When the sample is put through ion milling, the local heating and re-deposition due to ion beam bombardment creates a favorable condition for the Sn phase to react with Cu₆Sn₅ and Cu to form Cu₃Sn. We believe this is the reason for the Cu₃Sn coral phase appearing in our samples and the lack of Cu₆Sn₅ and the pure Cu phase in some of these samples. Indeed, we found that a sample carefully prepared to not involve water during mechanical polishing will not elicit the coral phase formation but result in clean and well defined Cu₆Sn₅ top and Cu₃Sn intermediate layers, and pure Cu phases, as seen in Fig. 15.38.

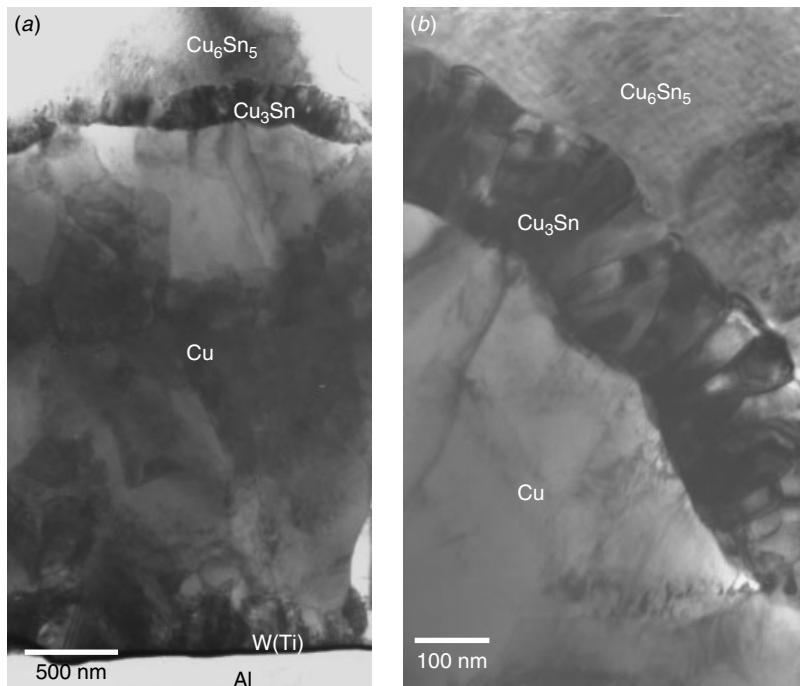


Figure 15.38 Eutectic solder/Cu/W(Ti) UBM system. The sample after three reflow cycles shows that with careful sample preparation, the formation of coral phase can be avoided and the pure Cu, the Cu_3Sn layer, and the Cu_6Sn_5 crystal on top preserved.

15.6 LEAD-FREE SOLDER/UBM SYSTEMS

Lead-free solder is becoming essential in the effort to create environmentally friendly microelectronic packaging (or green packaging). There are a few choices of lead-free solder in the market. Most of them are high Sn alloys with 0.5% to 4% additions of Ag, Cu, or both. A problem with high Sn solder is that it reacts vigorously with some of the widely used UBM systems, which creates new reliability issues. The Sn alloy, in general, has a higher melting point than the Sn-Pb solder. When solidified, Ag and Cu are partially precipitated and partially supersaturated in the Sn phase. These supersaturated Ag and Cu subsequently precipitate out as Sn–Ag and Sn–Cu intermetallic particles within the Sn grains or at the grain boundaries, as shown in Fig. 15.39. These precipitation particles greatly enhance the mechanical strength of the lead-free solder by pinning the Sn grain growth and by precipitation hardening. This is the reason why SnAgCu lead-free solder is mechanically stronger than traditional Sn–Pb solder, and its mechanical strength increases as thermal aging/stressing proceeds.

SnAgCu on Au/Sputtered Ni(V)/Ti UBM System

In its as reflowed condition this UBM system was readily attacked by the SnAgCu solder. Figure 15.40 shows a TEM cross section of the Ni(V)/Ti UBM layers. A thick layer of Ni_3Sn_4 is formed on top of the Ni(V) layer. Au has dissolved into the solder.

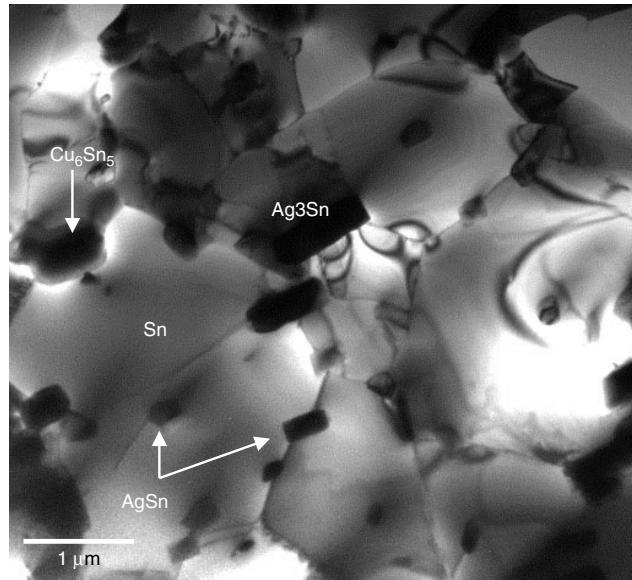


Figure 15.39 SnAgCu (Pb-free) solder after of 200 hours 150°C annealing. The matrix grain is in the Sn phase. Evident is inner and intergrain precipitation, which consists of either Ag–Sn or Cu₆Sn₅ intermetallics.

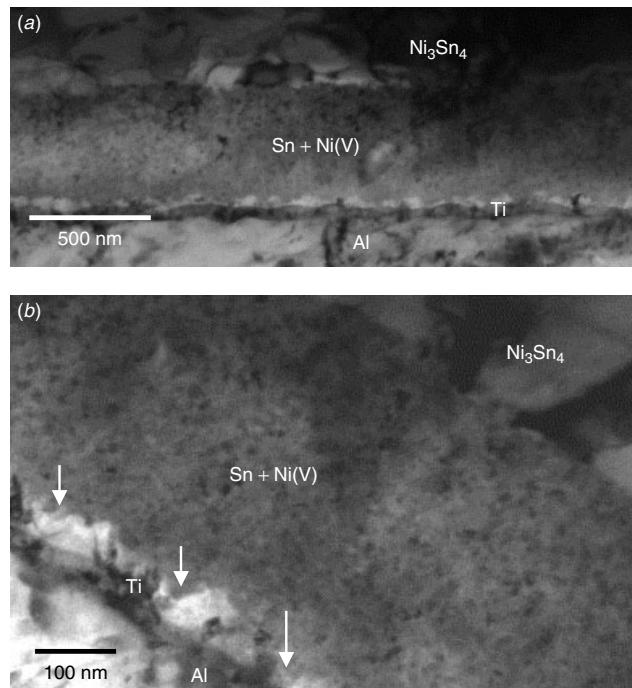


Figure 15.40 SnAgCu on Au/Ni(V)/Ti UBM, as-reflowed condition. A large amount of Sn has penetrated into Ni(V) and transformed the columnar Ni(V) into a fine-grain microstructure. Ni(V) to Ti interface has also been attacked, as indicated.

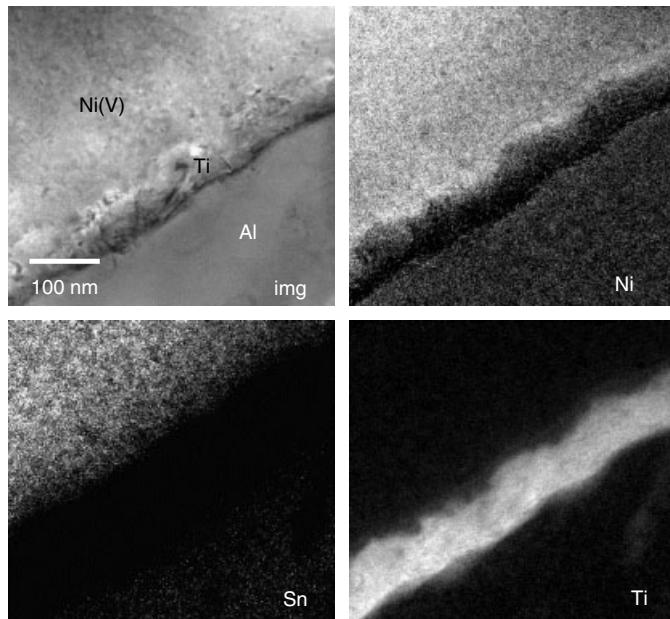


Figure 15.41 TEM/EELS mapping showing SnAgCu on Au/Ni(V)/Ti UBM, in as reflowed condition. A large amount of Sn has penetrated into Ni(V) and transformed the columnar Ni(V) into a fine-grain microstructure. The Ti map shows the Ni(V)/Ti interface to have been also attacked.

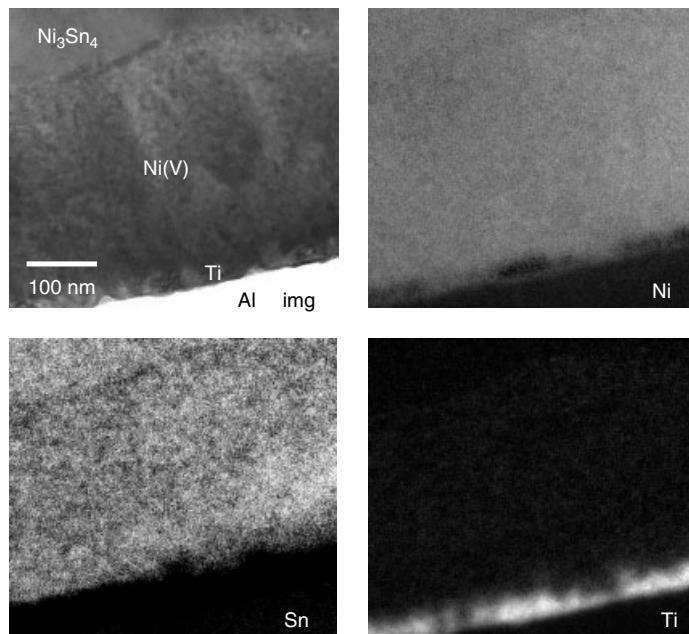


Figure 15.42 TEM/EELS mapping showing SnAgCu on Au/Ni(V)/Ti UBM, after 96 hours of annealing at 150°C. The Sn has further attacked the Ti layer, and the Ti layer has become thin and discontinuous.

Elemental mapping using electron energy loss spectrometry (EELS) shows that the Sn has penetrated all the way into the Ni(V) layer bottom and attacked the Ti layer, as in Fig. 15.41. Further annealing, for example, at 150°C high-temperature storage for 96 hours, has worsened the Sn penetration into Ni(V). Figure 15.42 shows that the Ti layer has gradually become discontinuous. Other experiments have shown the solder balls to fail in shear tests at the UBM interface (Tan et al. 2002). This UBM system apparently is not suitable for high Sn lead-free solder applications.

SnAgCu on Cu/Sputtered Ni(V)/Al UBM System

The system is very similar to the previous one except that the Au coating is replaced by a thicker Cu coating (500 nm) and the Ti is removed, since Ni(V) adhesion to Al is not a concern. Unlike the previous system, this UBM system survives after 1000 hours of 150°C high-temperature storage treatment. Sn still penetrates into Ni(V) layer but the extend damage is far less, as shown in Fig. 15.43. Note that a thick layer of Cu_6Sn_5 (with 5–10% of Ni in solid solution in the compound) is formed on top of

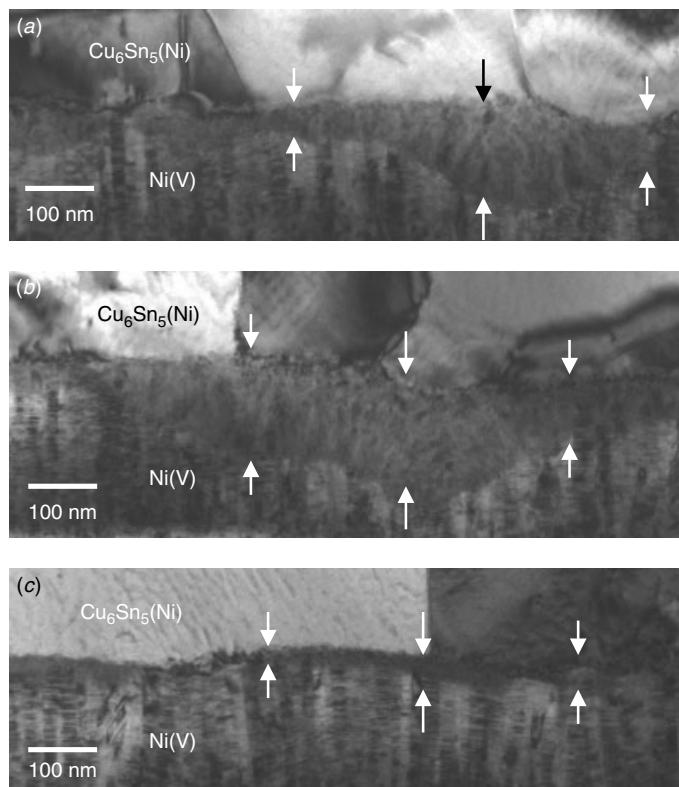


Figure 15.43 SnAgCu on Cu/Ni(V)/Al UBM samples annealed at 150°C for 1000 hours. The Sn has partially penetrated into the Ni(V) and transformed the columnar Ni(V) into a fine-grain microstructure, as indicated. The reaction was not homogeneous. Deep penetration is evident in certain areas, as seen in (a) and (b), and very shallow in others (c). (Courtesy Charles Lee, Infineon Technologies Asia Pacific Pte. Ltd., Singapore)

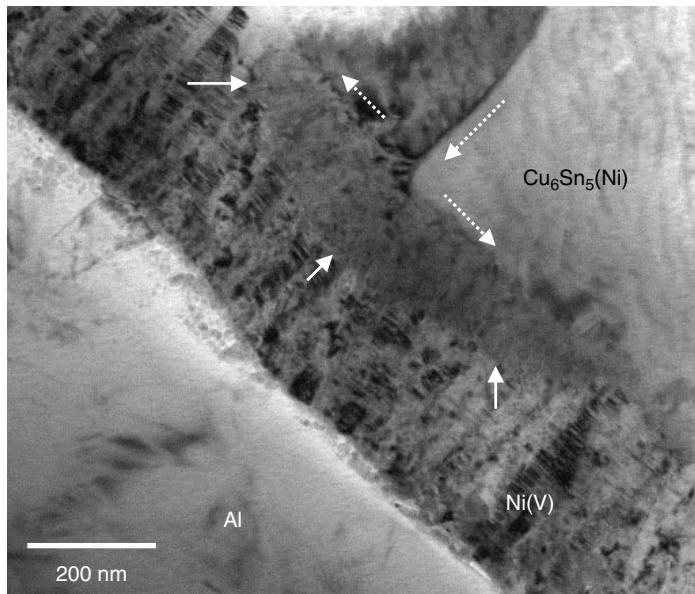


Figure 15.44 SnAgCu on Cu/Ni(V)/Al UBM samples annealed at 150°C for 1000 hours. The Sn penetration appears to have started from a Cu₆Sn₅ grain boundary and spread laterally (dashed arrows), which indicates that the reaction was controlled by the grain boundary's diffusion. (Courtesy Charles Lee, Infineon Technologies Asia Pacific Pte. Ltd., Singapore)

the Ni(V) layer. It is believed that this Cu₆Sn₅ greatly reduces the consumption of Ni and stops the Sn from directly reacting with and penetrating into the Ni(V). In other words, the supply of Sn is cut off by the Cu₆Sn₅ intermetallic compound and thus Ni(V) can survive even after 1000 hours of 150°C annealing. In most of the sample areas observed, Sn penetration into the Ni(V) always starts from the Cu₆Sn₅ grain boundaries, as shown in Figs. 15.43 and 15.44. This suggests that the supply of Sn reacting with Ni(V) is dominated by grain boundary diffusion. Cu₆Sn₅ as an effective diffusion barrier for Sn prevents the Ni(V) from being consumed. Figure 15.45 shows that the TEM image along with TEM/EELS mapping on the distribution of Sn is in fact controlled by the Cu₆Sn₅ grain boundary.

The analytic results in this and the previous sections clearly demonstrate that the effective diffusion barrier for Sn is the Cu–Sn intermetallic compound and not the Ni(V). The Ni(V) reacts with Sn readily and continuously, particularly in the presence of the high Sn solder. The case of SnAgCu in the Au/Ni(V)/Ti UBM system further confirms that the Ni(V) alone cannot prevent the Sn penetration as the UBM fails readily after a short annealing time. The effectiveness of the Cu–Sn intermetallic as the diffusion barrier is also confirmed by the eutectic solder/Cu/Ni(V)/Al system result described in previous section.

SnAgCu on Au/Electroless Ni(P)/Al UBM System

Electroless plated Ni is a commercially successful UBM system mainly due to its low-cost high throughput process, which yields excellent reliability. When Pb-free

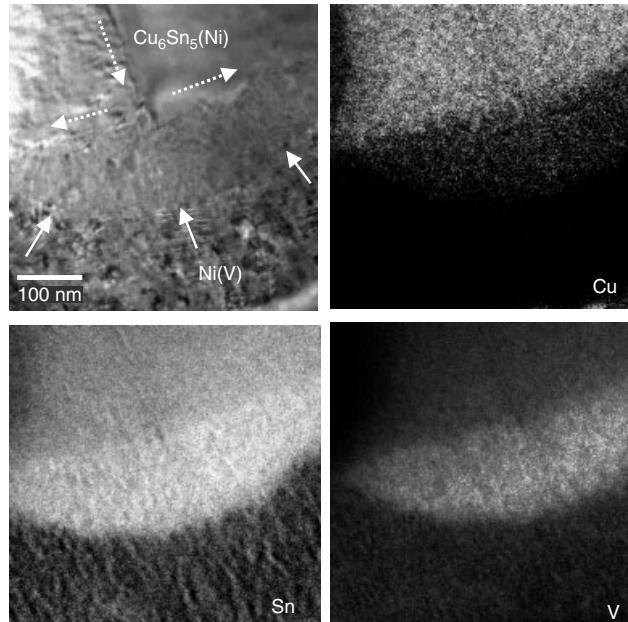


Figure 15.45 SnAgCu on Cu/Ni(V)/Al UBM samples annealed at 150°C for 1000 hours. The Sn penetration has started from a Cu_6Sn_5 grain boundary. Also shown are TEM/EELS element maps for Cu, Sn, and V. Note that the V content is higher in the area that has reacted with Sn. (Courtesy Charles Lee, Infineon Technologies Asia Pacific Pte. Ltd., Singapore)

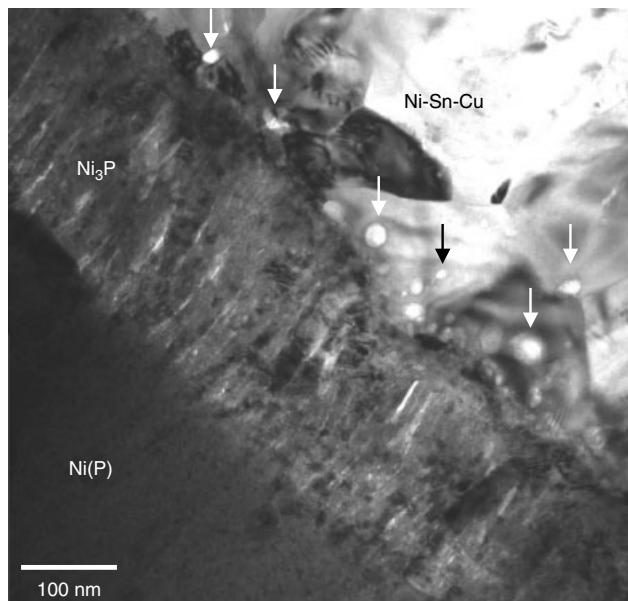


Figure 15.46 SnAgCu solder/Au/eNi(P)/Al UBM system in as-reflowed condition. The Ni_3P layer has formed between the Ni–Sn–Cu intermetallics and the Ni(P). Kirkendall's voids appear on the Ni_3P surface, as indicated. (Courtesy Charles Lee, Infineon Technologies Asia Pacific Pte. Ltd., Singapore)

solders are used, the interface reaction is not much different from that of the conventional eutectic solder as described in the previous session. Figure 15.46 shows an as-reflowed sample interface microstructure with the Ni_3P layer and Ni–Sn–Cu intermetallics formed on top. A large number of Kirkendall's voids appear in between the Ni_3P and Ni–Sn–Cu intermetallic. EDS shows that the intermetallic formed is by a Sn : Ni : Cu = 10 : 3 : 7 ratio and possibly a ternary IMD phase. No positive phase identification work was accomplished.

As the sample undergoes more reflow treatment, the Ni_3P layer grows thicker, from 250 nm in a 1 reflow sample up to 750 nm for the 10 reflow sample. Sn–Ni–Cu IMD also grows thicker with large columnar grains and smaller grains nucleated from the Ni_3P interface, as seen in Figs. 15.47(a) and (b), and the new Sn–Ni–Cu IMD phase is believed to start nucleating as well. The new intermetallic phase is higher in Ni and lower in Cu, $\text{SnNiCu} = 11 : 6 : 3$, as compared to the existing one with $\text{SnNiCu} = 10 : 3 : 7$. Both IMD are found to be within a $\pm 5\%$ range of variation in their Sn, Ni, and Cu compositions. The Kirkendall voids grow in density as the sample goes through more reflow cycles. In some cases, micro-cracks develop due to the external mechanical stress, Fig. 15.47(c). The high-density voids at the interface of Ni_3P are detrimental if the system is under substantial mechanical stress. However, reliability

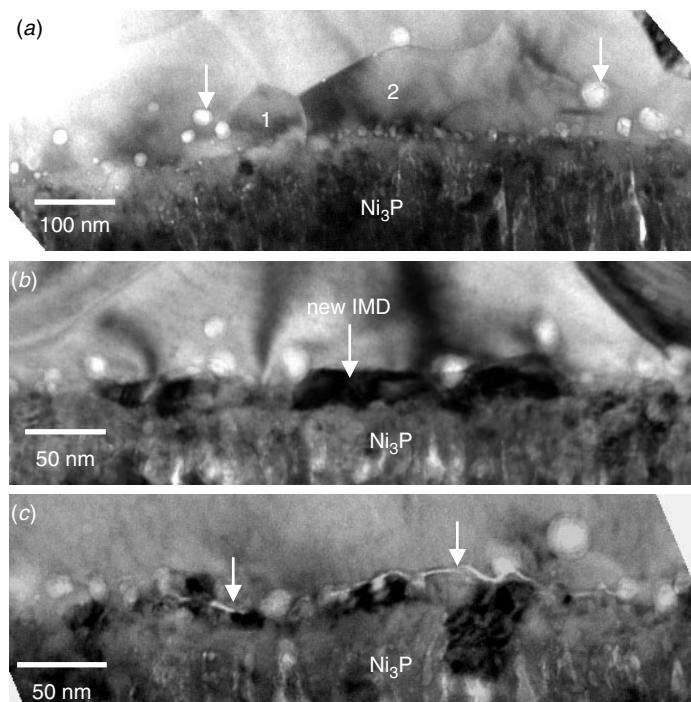


Figure 15.47 SnAgCu solder/Au/eNi(P)/Al UBM system after 10 reflows. (a) Kirkendall's voids, as indicated, are scattered near the interface between the Ni_3P and Ni–Sn–Cu intermetallics. Small Sn–Ni–Cu grains also nucleate from the interface, as indicated by numbers 1 and 2. (b) New Sn–Ni–Cu ternary intermetallics, as indicated, are forming. (c) Micro-cracks linking the Kirkendall's voids, as indicated, may eventually lead to the interface's delamination. (Courtesy Charles Lee, Infineon Technologies Asia Pacific Pte. Ltd., Singapore)

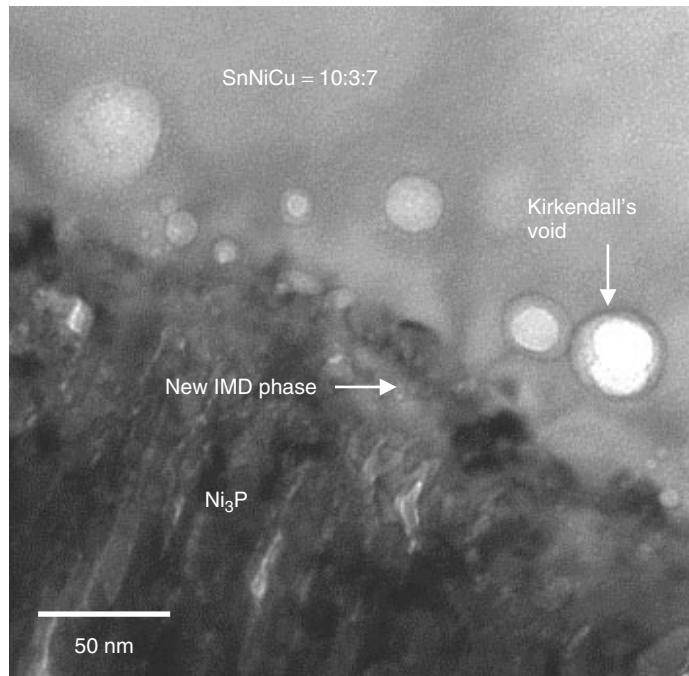


Figure 15.48 SnAgCu solder/Au/eNi(P)/Al UBM system after 10 reflows. Kirkendall's voids are formed within the SnNiCu intermetallic, suggesting that the Sn was extracted from this intermetallic to supply for the new phase forming at the NiP interface, as indicated. (Courtesy Charles Lee, Infineon Technologies Asia Pacific Pte. Ltd., Singapore)

tests did not reveal any mechanical failure arising from the UBM interface fractures. A more detailed look at the Kirkendall voids, is provided in Fig. 15.48, where the voids appear to mostly sit near but not exactly at the interface of the Ni₃P layer. The voids are inside the large Sn–Ni–Cu grains and usually at the interface to the smaller Sn–Ni–Cu grains, as shown in Figs. 15.47 and 15.48. This suggests that nucleation of the new smaller Sn–Ni–Cu grains comes at the expenses of the larger grains. In samples treated with high-temperature storage annealing at 150°C, the IMDs grown are thick and columnar, Fig. 15.49.

Nucleation at the Ni₃P interface of either the new phase or same phase new grains never stops. The supply of Ni from Ni(P) needs to be matched with the supply of Sn and Cu from IMD above and forming IMD at the Ni₃P interface. The resulting microstructure has the following characteristics: The IMD continues to nucleate from the Ni₃P interface, as seen in Figs. 15.47 to 15.50. The Kirkendall voids form with high concentration within old IMD (SnNiCu = 11 : 6 : 3) grains; most of the voids appear next to the newly nucleated grains, Figs. 15.48 and 15.50. On the Ni₃P side, the supply of Ni must travel from far below Ni(P) through the columnar gaps within Ni₃P. The Kirkendall voids formed in this way are far more gentle and slow growing, compared with the voids formed during reflow as Ni is extracted from Ni(P) in the reaction with the liquid Sn phase. There are thus two different Kirkendall's void morphologies in Ni₃P, as shown in Fig. 15.51;

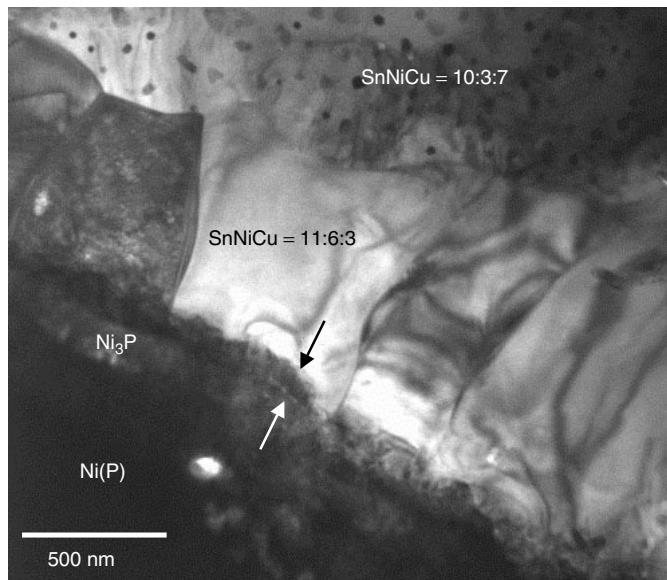


Figure 15.49 SnAgCu solder/Au/eNi(P)/Al UBM system after 200 hours at 150°C high-temperature storage. Different SnNiCu IMD phases have evolved and coexist. The nucleation of IMD at the Ni₃P interface has never stopped, as indicated by the arrows. (Courtesy Charles Lee, Infineon Technologies Asia Pacific Pte. Ltd., Singapore)

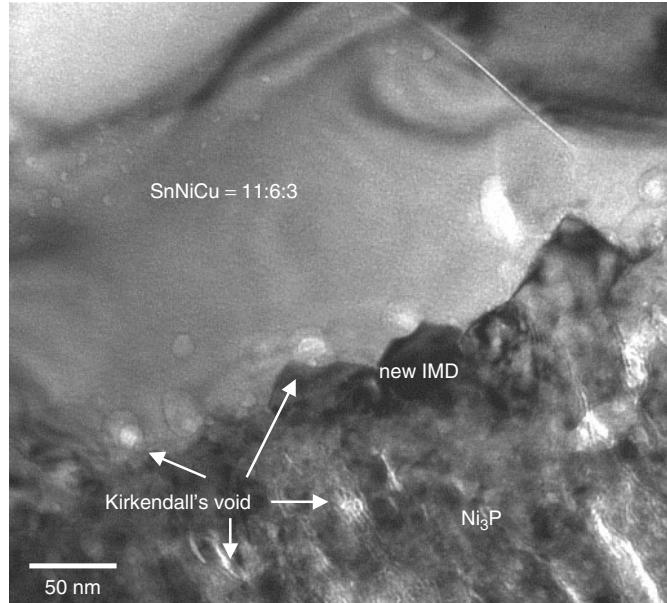


Figure 15.50 SnAgCu solder/Au/eNi(P)/Al UBM system after 1000 hours at 150°C high-temperature storage. The new IMD formation is supplied by the Ni from the Ni(P) below and the Sn and Cu from the IMD above, resulting in the Kirkendall's void formations seen on both sides of the new IMD. (Courtesy Charles Lee, Infineon Technologies Asia Pacific Pte. Ltd., Singapore)

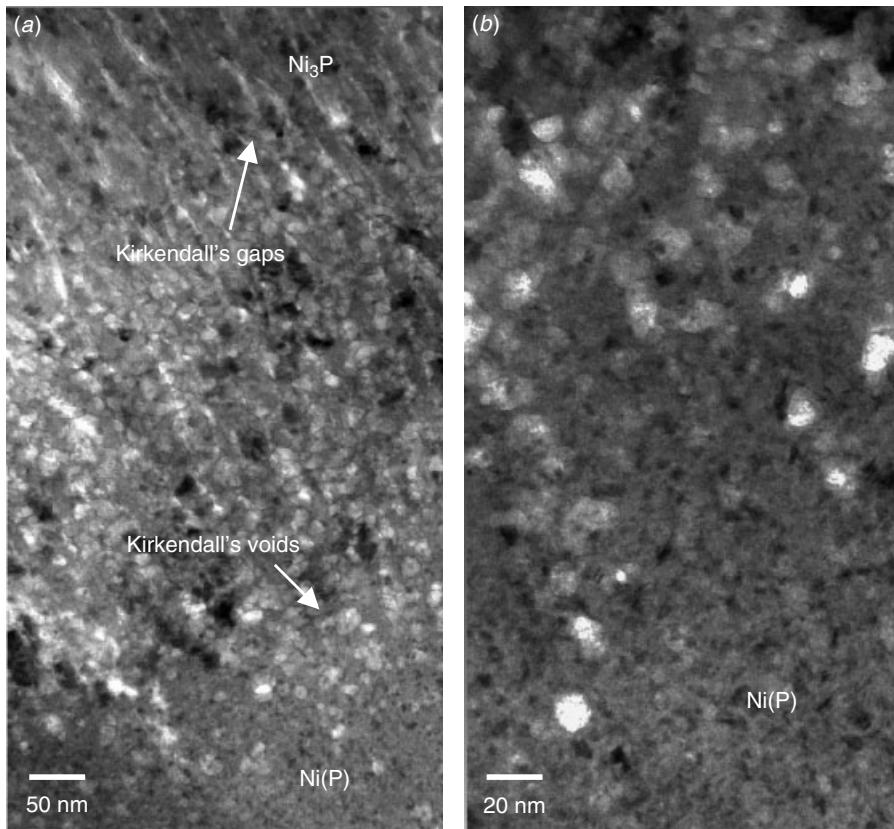


Figure 15.51 SnAgCu solder/Au/eNi(P)/Al UBM system after 1000 hours at 150°C high-temperature storage. Two different Kirkendall's void morphologies are found. One is the vertical gap seen at the top of (a). This is formed during the reflow cycles. The other is the spherical void seen in the lower part of (a) and in the upper part of (b). The spherical voids are formed during the 150°C annealing. (Courtesy Charles Lee, Infineon Technologies Asia Pacific Pte. Ltd., Singapore)

- Vertical gaps as reported in previous sections. These Kirkendall gaps form during the reflow process as Ni was extracted in the reaction with the liquid phase Sn. The temperature is high (usually higher than 220°C) and the liquid phase Sn reacted with Ni voraciously. Kirkendall's voids formed thus are aligned into vertical gaps.
- Normal spherical voids formed during high-temperature storage at 150°C or any other temperature treatment. Since the temperature is lower compared to that of reflow and there is no liquid phase involved, the reaction is far less vigorous. The Kirkendall's voids formed tend to be spherical, as seen in Fig. 15.51.

15.7 Sn WHISKERS AND Pb-FREE SOLDERS

In surface mount and flip-chip packaging technologies of electronic packaging, some Cu lead frames are finished with a layer of Pb-free solder. A large number of Sn whiskers

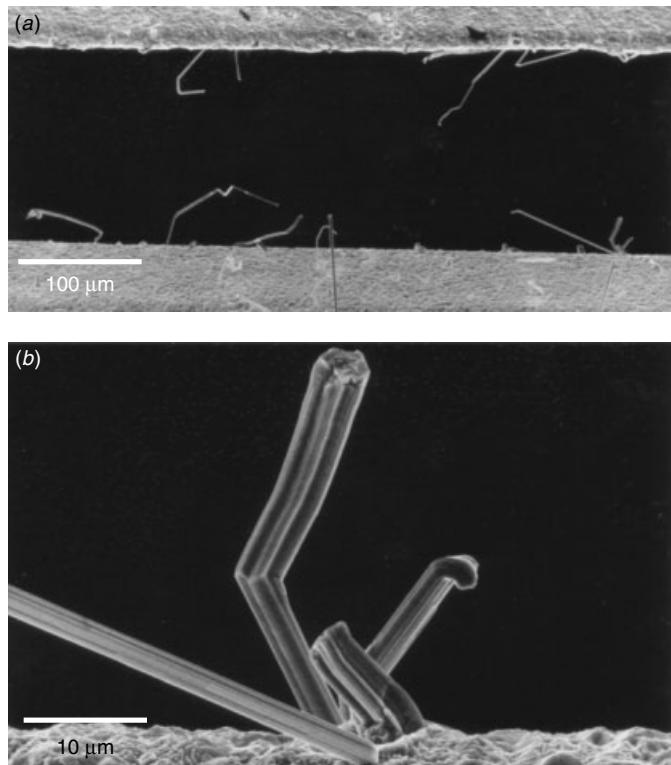


Figure 15.52 Sn whiskers grown on the Sn-plated surface Cu leadframes. (Courtesy Prof. K. N. Tu, UCLA, USA)

are found on the surface of the finish, especially for eutectic SnCu or pure Sn coatings. The whiskers grow spontaneously during room temperature storage and the consequences are detrimental. Figure 15.52 shows some typical whiskers grown on SnCu surface coating. Several features are noted:

- The growth of Sn whiskers is spontaneous.
- Sn whiskers grow much faster on the SnCu finish than on pure Sn finish coating.
- Parallel lines and irregular dots are often found on the body of each and every whisker surface, as seen in Fig. 15.53.
- The root of the whisker tends to be larger in diameter. Other than that, the whisker diameter remains more or less identical throughout the whole whisker, as in Fig. 15.53.
- Multiple angled bendings are often found, Fig. 15.53. The bending angles remained to be determined but they do not seem to follow any known crystallographic angles. For example Fig. 15.53(a) shows a bending angle of about 60° to 80° , while Fig. 15.53(b) shows bending angles to be less than 5° to 10° .

Cross sections of the Sn whisker in parallel and normal to the whisker growth are shown in Fig. 15.54. Both images show that the Sn whisker grows from a Sn grain

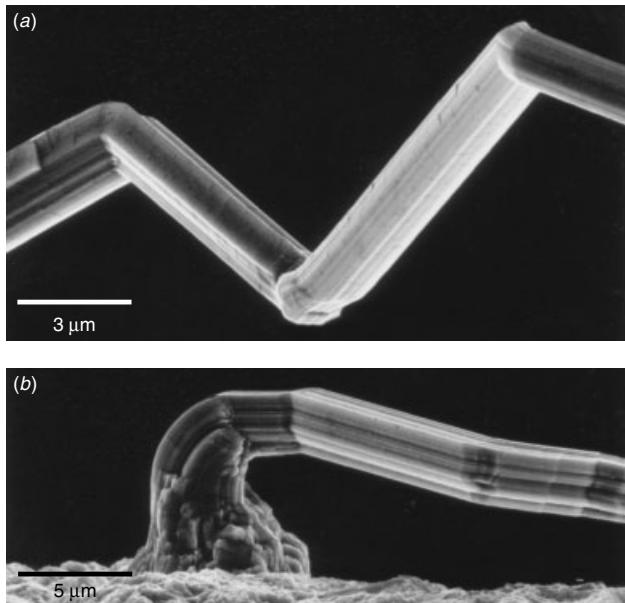


Figure 15.53 Sn whisker characteristics. Parallel lines with irregular dots are often found on each whisker. (a) Straight sections are joined by angled bends and (b) whisker roots tend to be larger in diameter, whereas the whisker itself has a diameter that remains relatively fixed. (Courtesy Prof. K. N. Tu, UCLA, USA)

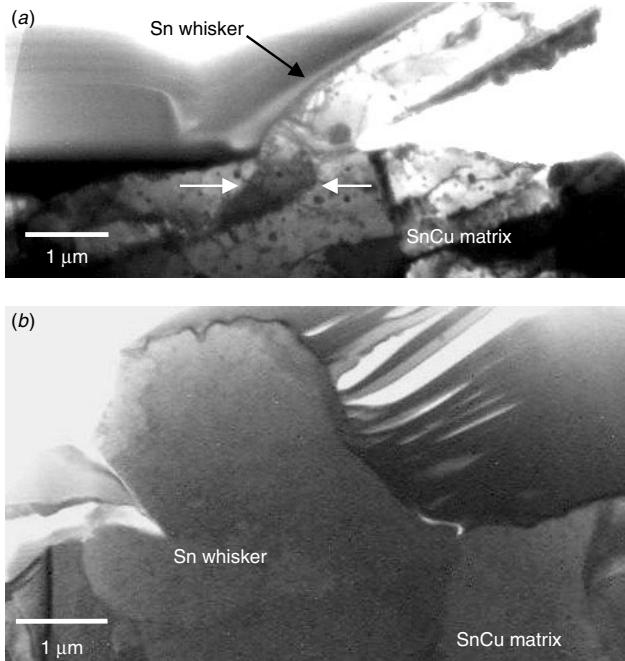


Figure 15.54 TEM cross section of the whisker root (a) parallel to the whisker's growth direction as the whisker initiates from a Sn grain boundary, as indicated, and (b) normal to the whisker's growth direction. (Courtesy Prof. K. N. Tu, UCLA, USA)

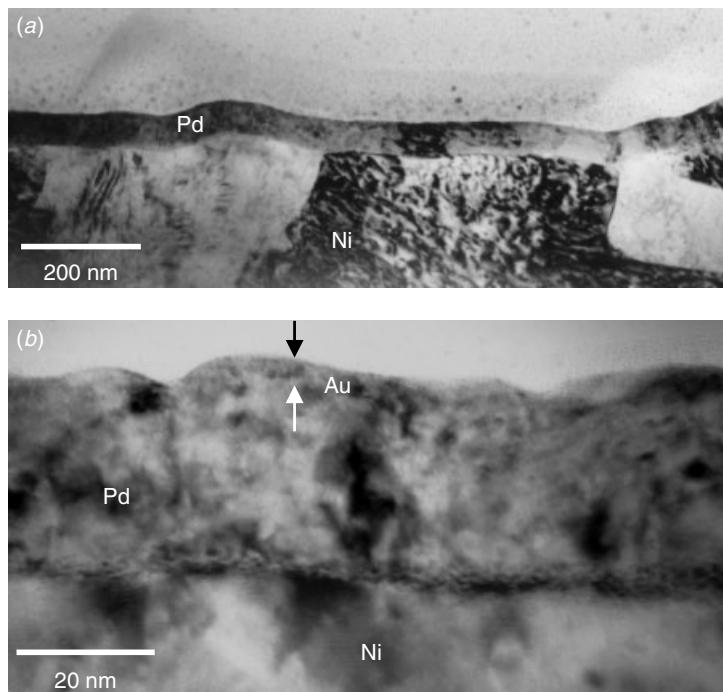


Figure 15.55 Cu leadframe coated with Ni, Pd, and a very thin Au. The microstructure of Pd and Au coatings can only be studied by TEM cross section, as shown here. (a) The Pd layer thickness is homogeneous, while the surface Au flash thickness depends on surface roughness, as seen in (b).

boundary, but the detailed microstructural relationship between the Sn whisker and the matrix Sn grains remained unclear. To understand the spontaneous Sn whisker growth, it is essential to understand the microstructures of the pure Sn or SnCu finish coating on Cu leadframes. A high density of Cu_6Sn_5 intermetallic compound precipitation was found in both the Sn and SnCu finish coatings, while the density in the SnCu finish is much higher. The intermetallic Cu_6Sn_5 phase was formed during electroplating in the as-plated condition. It is known that Cu and Sn react at room temperature to form the Cu_6Sn_5 intermetallic compound (Yu 1973; Mayer et al. 1975). The diffusion of Cu into Sn to grow Cu_6Sn_5 along the grain boundaries of Sn increases the volume and introduces a long-range compressive stress in the neighboring grains (Tu 1994; Lee and Lee 1998). To relieve the compressive stress, layers of atoms normal to the stress direction must be removed, and these atoms are driven by the stress gradient to diffuse and form a whisker. The whisker itself is stress free. The diffusion of the atoms can occur by long-range grain boundary diffusion, and the self-grain boundary diffusion of Sn at room temperature is fast enough for stress relief. On the other hand, Cu also diffuses rapidly along the Sn grain boundary at room temperature (and interstitially within the Sn grains) to supply Cu_6Sn_5 compound's growth. The continuous growth of the compound at room temperature will maintain as long as there is a compressive stress in the Sn matrix; hence the Sn whisker growth on the Cu leadframe is a spontaneous phenomenon (Sheng et al. 2002).

15.8 OTHER PACKAGE-RELATED APPLICATIONS

Various TEM applications in advanced packaging related analysis and studies were developed and employed recently. In this short section only a few examples are given to illustrate the potential of using TEM for package-related studies and the insightful information that can be obtained.

Au/Pd/Ni Coatings on Cu Leadframe

The Cu leadframe for microelectronics packaging has a long history of applications. Noble metals are used to coat the surface and prevent Cu from oxidation. Usually Ni is coated onto Cu first, and followed by either Au or Pd plus thin Au coatings. The advantages of Pd/Ni and Pd/Ni/Au leadframe coatings in basically removing the Sn, Pb, and Ag used in the process:

- Simplified the assembly process, reduced assembly cycle time, and increased yields with the removal of SnPb or Sn coatings. This has introduced environmental issues (Pb and its related coatings) and Sn whisker issues (Sn and its related coatings).
- No oxidation on the leadframe surface.
- Eliminating solder bridging in the fine pitch leadframe application.
- Eliminating the problems associated with Ag spot registration and Ag bleedout.

The Au layer on top of Pd/Ni ensures that no oxidation will occur up to a temperature of 450°C, and thus the leadframe can be used for the C4 joint (Yu 1998). Figure 15.55 shows a typical Au/Pd/Ni coating on the Cu leadframe in cross-sectional TEM images. The Pd coating's average thickness is around 20 to 30 nm and the Au coating is less than 3 to 5 nm. Only the cross-sectional TEM can and in determining the microstructures of these coating layers.

Cu Leadframe Nitridation

The Cu leadframe has been known to oxidize during storage, and the oxidized Cu surface has created notorious adhesion problems in its binding to molding compounds, wire bonding, or solder wetting. Various techniques have been developed that change the Cu leadframe surface properties in order to prevent the Cu leadframe shelf storage adhesion problems. One known technique is to introduce a surface treatment based on plasma technology (Wong et al. 2000; Gopalkrishnan et al. 1999; Wong et al. 2000). Figure 15.56 shows the cross-sectional microstructure of a Cu leadframe surface after a nitridation plasma treatment. The Cu surface has been roughened, so Cu nanodendritic structures are observed. Such surface morphological change has also provided a mechanical means of interlocking and thus has greatly improved adhesion in molding and wire bonding.

Cu/Polyimide Interface

Flip-chip packaging technology and many advanced chip-scale packaging technologies have relied on a substrate platform to support the ULSI die and to re-wire the chip's

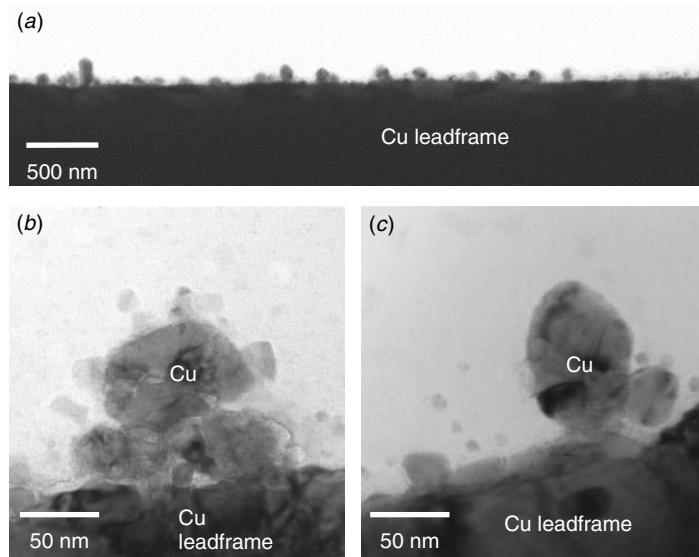


Figure 15.56 Cu leadframe surface after nitridation. (a) Lower magnification shows the surface roughness introduced by the treatment. (b, c) The surface bumps are revealed to be Cu grains with micro-dendritic structures.

interconnections. The flip-chip substrate is an important component in contemporary high pin count logic device packages. Basically the flip-chip substrate is a laminated structure with Cu forming the mechanical support as well as the interlayer interconnections and polyimide forming the dielectric layers. The Cu to polyimide interface and Cu's adhesion property are vital to the system's performance and reliability. To enhance adhesion, many different interface coatings have been proposed for improvement. Figure 15.57 shows an example of a Cu/polyimide interface treatment. A Cu layer of about 200 nm was first coated onto the Cu substrate, as seen in Fig. 15.57(a). The coated Cu was then treated by special chemistry to improve its adhesion to the polyimide. The chemical treatment produced a second thin layer of a complex Cu compound about 20 to 30 nm thick, as seen in Figs. 15.57(b) and (c). The nature of this Cu compound remains to be determined.

Another example using a more straightforward approach is shown in Fig. 15.58. First the Cu substrate's surface is very rough, Fig. 15.58(a). This surface roughness provides the first macroscopic level mechanical interlocking. Close up, the tip of one of the substrate's humps shows dendritic branches and arms of the Cu crystalline coating, as seen in Figs. 15.58(b) and (c). This secondary surface roughness provides a second microscopic level of mechanical interlocking. The bonding is thus purely mechanical between Cu and polyimide in this approach and not molecular level chemical bonding.

Solder Mask Fillers

Solder mask conformal coating is widely used in all the flip-chip substrates as well as for the conventional printed circuit board (PCB). As implied by its name, the solder mask coating is used to prevent any solder bleeding induced short/leakage among

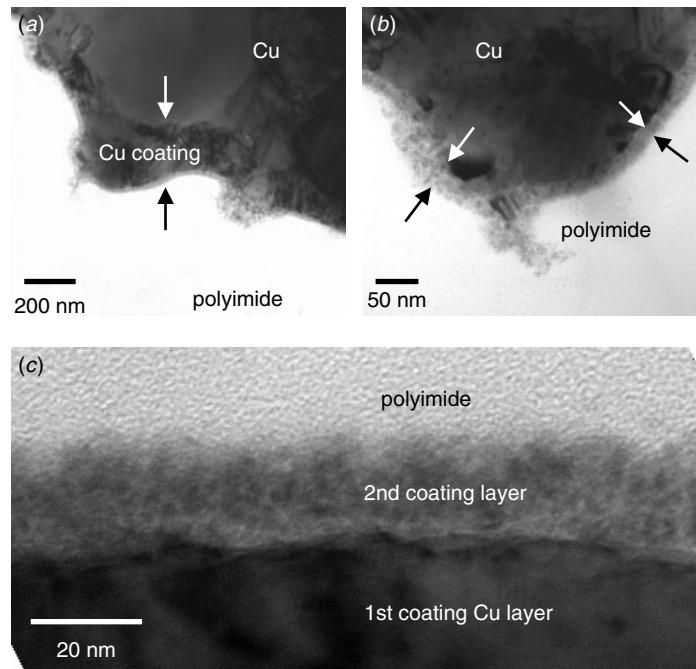


Figure 15.57 Cu/polyimide interface in a flip-chip substrate. (a) A thin Cu coating of about 200 nm is observed. (b) A second coating, about 20 to 30 nm thick, is observed. (c) Close-up of the second coating reveals a nanocrystalline microstructure.

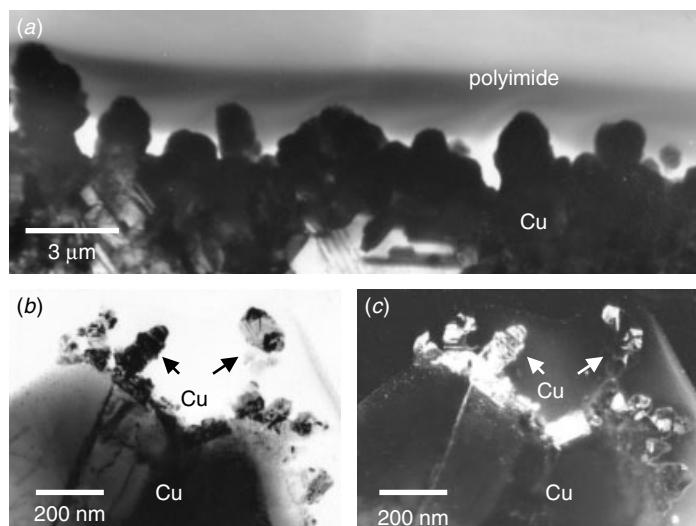


Figure 15.58 Cu/polyimide interface in a flip-chip substrate. (a) The mechanical adhesion between Cu and the polyimide is largely enhanced by the rough interface. (b, c) A close-up of the bright and dark fields at the tip of a Cu surface hump shows the dendritic microstructure to further enhance the Cu/polyimide adhesion.

the contact pads and conductor lines. It also prevents substantial contaminants, such as solder flux, from sticking to the substrate and inducing reliability issues at a later stage. A basic requirement for the solder mask is that its surface be nonwettable for solder and free of other contaminants. Usually the solder mask coating is a composite material with a polymeric based organic material for flowability and processibility. Oxide fillers are added to increase the coating's surface tension and to create a nonwettable surface. To enhance the effectiveness of the fillers in increasing surface tension, the filler's surface energy or surface area should be maximized. One way to achieve maximum surface area within a limited space is to add fillers with micro-pores. Figures 15.59 and 15.60 show two different solder mask fillers, both containing a large proportion of micro-pores within the filler. In Fig. 15.60, the filler is a crystalline grain with micro-pores following its hexagonal crystallographic lattice planes. A TEM analysis of these solder mask fillers clearly reveals the effects of a material's microstructure on a system's macroscopic performance.

Al Pad Surface Oxides

The Al bond pad surface is the only electrical component of the whole ULSI device exposed to the outside world and to which electrical interconnection can be made. Because of this importance to the device's ultimate electrical performance and reliability, the quality of the Al bond pad surface must be controlled, and the various

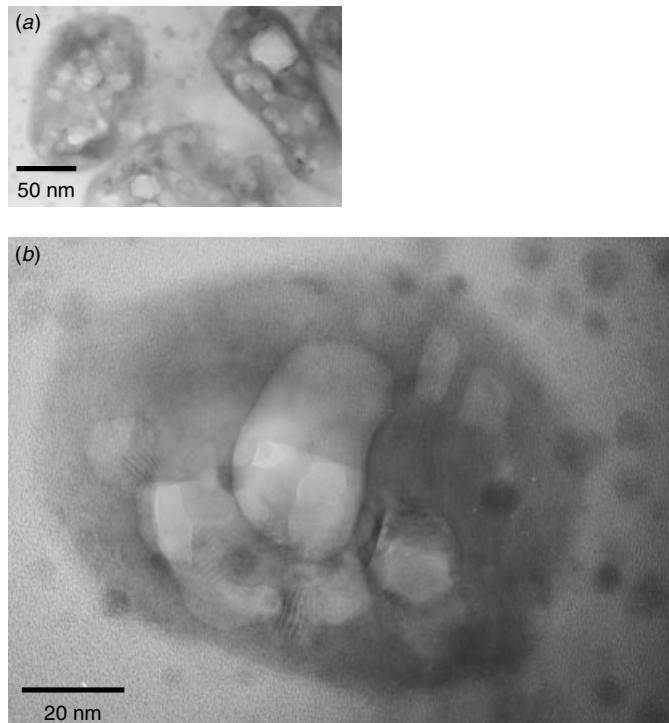


Figure 15.59 Solder mask filler morphology. (a) Each filler cluster has multiple pores within. (b) A close-up of one filler shows that the filler contains quite a number of micro-pores.

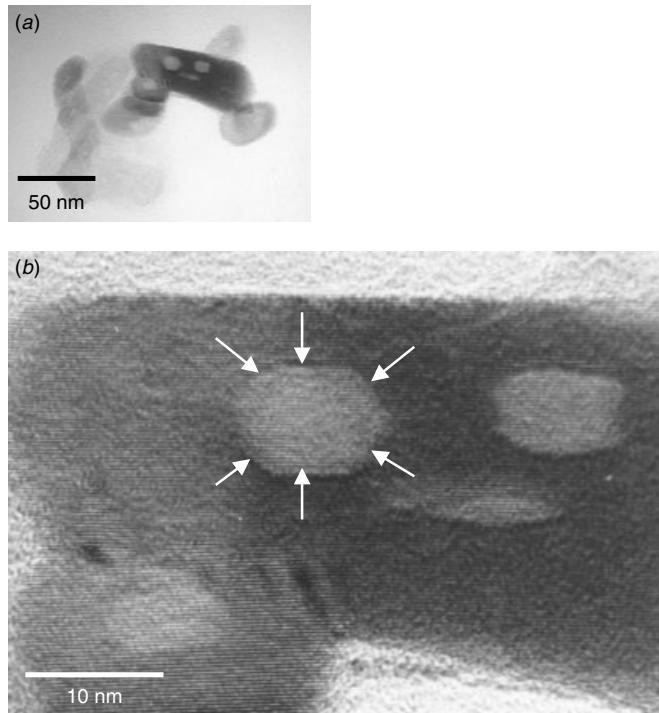


Figure 15.60 (a) A different filler with similar micro-pores, indicating these to be characteristic of all the fillers. (b) An HRTEM image of the filler showing the filler to be a single crystal grain with multiple pores within the crystal grain. The pores follow the hexagonal crystallographic orientation, as indicated.

factors that can affect its properties should be understood. It is now known that there is always some fluorine and carbon contamination as well as a thin aluminum oxide layer on the bond pad surface. These contaminants are regarded as normal and acceptable as long as they do not affect the subsequent processes, including testing and wire bonding. Figure 15.61 shows a “normal” Al bond pad surface with a thin native oxide observed in HRTEM. The surface oxide is about 1 to 3 nm. This surface oxide will protect and prevent the Al from further oxidizing. Upon wire bonding, this surface oxide will be broken by the wire bond’s ultrasonic power so that electrical and mechanical interconnections can be made. The presence of such a surface oxide on Al, however, does not always imply that the bond pad will have a wire bondability problem. The thickness of the oxide film and its chemical nature need to be characterized in order to understand their impact in the wire bond results.

15.9 A SHORT NOTE ON TEM SAMPLE PREPARATION FOR UBM AND PACKAGING RELATED DEVICES

TEM sample preparation for packaging related studies is more tricky than for ULSI devices for two good reasons:

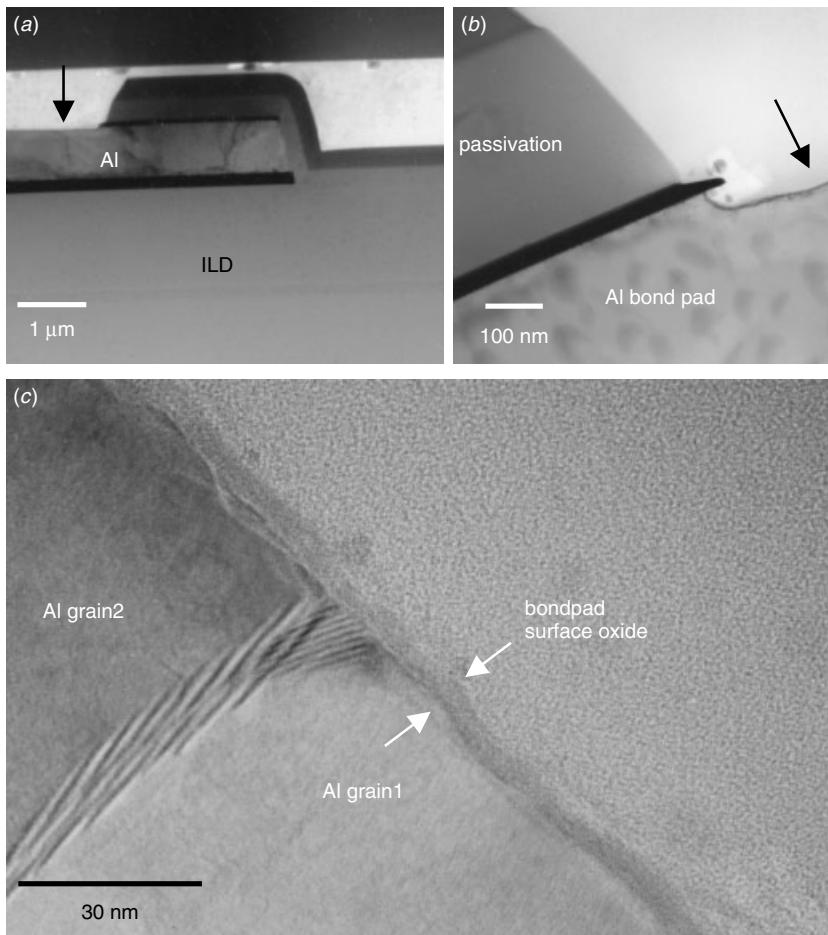


Figure 15.61 TEM cross section of the Al bond pad. (a) Low magnification of the overall bondpad structure. (b) Bond pad corner showing the Al surface undercut during the bond pad opening etching. (c) HRTEM image of the bond pad surface showing a thin oxide layer on the surface of Al.

- The materials involved are more sensitive to temperature. Solder materials melt at temperatures as low as 175°C. Phase transformation, intermetallic formation, and grain growth all occur at temperatures lower than the melting point, and often they occur at room temperature. Solder materials are also very soft compared to the ULSI device itself. Mechanical scratches are inevitable in most cases, Fig. 15.62. Special care must be exercised in in-situ cleaning of the polishing pad in order to avoid scratches the in solder ball, if the solder ball microstructure needs to be studied.
- Various organic materials are involved in most packaging structures. Many are reactive to organic solvents like acetone. Some are sensitive even to water, like Pb metal. Still others are sensitive to ion beam irradiation and thus cannot be processed using a focus ion beam or an ion miller.

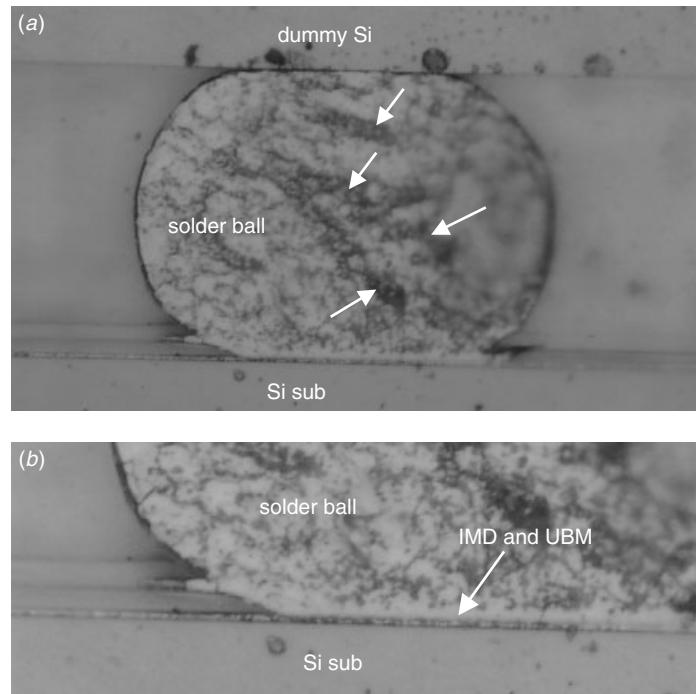


Figure 15.62 A mechanically polished solder bump cross section as seen under a light microscope. (a) Substantial mechanical scratches can be observed within solder ball, as indicated. (b) The solder/UBM interface reaction was not affected by the mechanical polishing.

Experience teaches us that the certain rules must be observed in processing package-related TEM samples:

- Use mechanical polishing with very limited ion milling. Even the most gentle ion milling can introduce local heating and solder re-melting, as shown in Fig. 15.63. It is advisable to use liquid N₂ cooling during ion milling, although it may not be necessary in some cases.
- Avoid water as a lubricant if the solder is Pb–Sn based and the solder's microstructure is to be studied. Note that in preparing the SEM samples, such precaution is not needed. In the SEM cross section, the Pb-rich phase attacked by water will enhance the contrast and help reveal the solder's microstructure.
- Use diamond lapping film instead of SiC, aluminum oxide, or any other polishing media. Avoid using polishing cloth and slurry, which will enhance the different erosion rates between hard and soft materials and create local thinning/rounding at the UBM interface. Soft slurry also embeds foreign particles into the solder and can confuse the analysis.
- Finish with polishing agents, such as cyton, that contain chemical enhance etching and polishing.
- Avoid FIB, unless the microstructure to be analyzed is not sensitive to both ion beam irradiation and local heating.

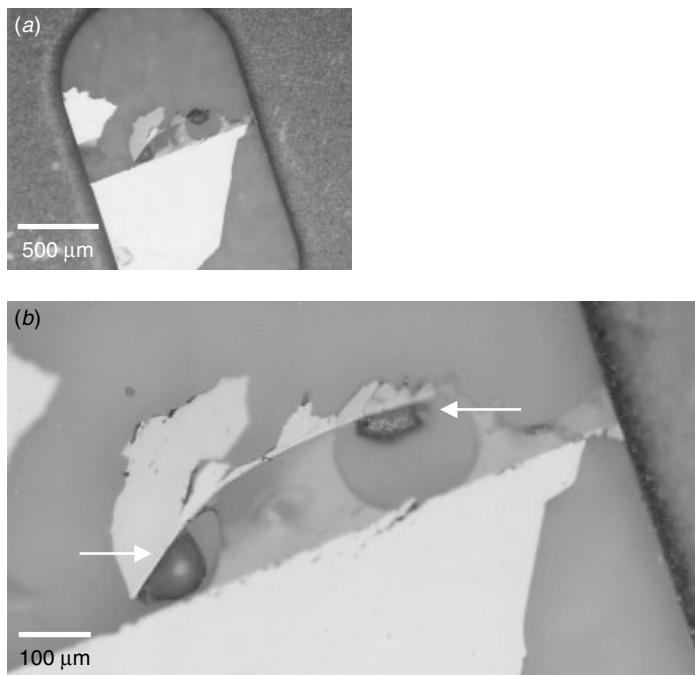


Figure 15.63 Low-magnification light micrographs show that the solder was re-melted during ion milling, as indicated.

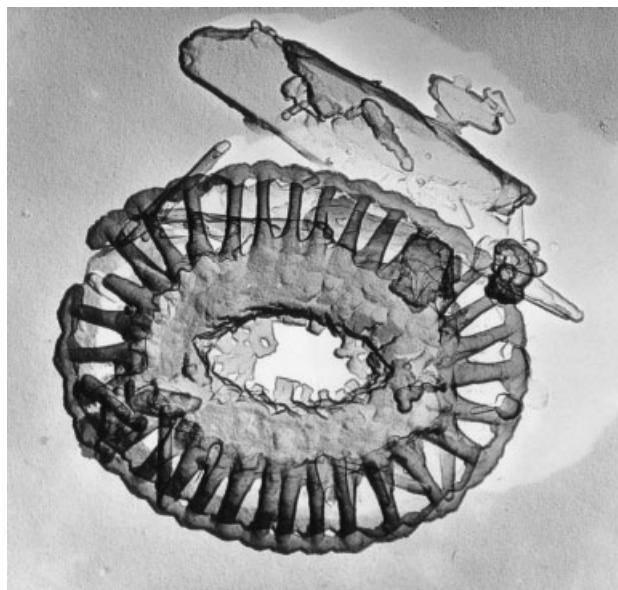
As packaging technologies continue to evolve with ever more novel materials introduced into the devices, TEM sample preparation for these new devices will need to be simultaneously studied and revised in order to meet the new challenges.

REFERENCES

- R. Anderson, IBM, private communication, 2001.
- S. U. Campisano, G. Foti, E. Riini, S. S. Lau, and J. W. Mayer, *Phil. Mag.*, **31**, 903, 1970.
- L. J. Fried, J. Havas, J. S. Lechaton, J. S. Logan, G. Paal, and P. Totta, *IBM J. Res. Develop.*, **26**, 362–371, 1982.
- R. Gopalkrishnan, K. Nyunt, T. C. Chai, A. Wong, and G. Sarkar, *Advances in Electronic Packaging, ASME EEP*, **26** (1), 1999.
- J. W. Jang, P. G. Kim, K. N. Tu, and D. R. Frear, 1999 *Int. Symp. Advanced Packaging Materials*, 252–255, 1999.
- P. G. Kim, J. W. Jang, T. Y. Lee, and K. N. Tu, *J. Appl. Phys.*, **86** (12), 6746–6751 1999.
- B. Z. Lee and D. N. Lee, *Acta Materialia*, **46**, 3701, 1998.
- C. Y. Liu, K. N. Tu, T. T. Sheng, C. H. Tung, D. R. Frear, and P. Elenius, *J. Appl. Phys.*, **87** (2), 750–754 2000.
- W. Mayer, M. Poate, and K. N. Tu, *Science*, **190**, 228, 1975.
- Z. Mei, M. Kaufmann, A. Eslambolchi, and P. Johnson, 1998 *Electronic Components and Technology Conf.*, 952–961, 1998.

- T. C. Tan, C. H. Tung, and T. C. Chai, *Proc. ISTFA, 28th Int. Symp. Testing and Failure Analysis*, 127–132, ASM International, 2002.
- G. T. T. Sheng, C. F. Hu, W. J. Choi, K. N. Tu, Y. Y. Bong, and L. Nguyen, *Journal of Applied Physics*, **92**(1), 64–69, 2002.
- P. S. Teo, Y. W. Huang, C. H. Tung, M. R. Marks, and T. B. Lim, *Proc. 50th Electronic Components and Technology Conf. (ECTC)*, 33–39, 2000.
- K. N. Tu, *Acta Metallurgica*, **21**, 347, 1973.
- K. N. Tu, Lecture Note of “Interconnection Challenges for ULSI Devices and Packages,” Singapore, 1998.
- K. N. Tu, *Physi. Rev.*, **B49**, 2030, 1994.
- C. H. Tung, Y. W. Huang, P. S. Teo, M. R. Marks, T. T. Sheng, T. B. Lim, and J. Wang, *IMAPS Advanced Technology Workshop on Flip Chip Technology*, Breselton, Georgia, USA, 2000.
- C.-H. Tung, George, T. T. Sheng, P. S. Teo, and M. C. Y. Lo, *Proc. ISTFA, 28th Int. Symp. Testing and Failure Analysis*, 505–511, ASM International, 2002.
- A. S.W. Wong, R. Gopalkrishnan, and G. Sarkar, *J. Vac. Sci. Tech.*, **A18** (4), 1–13, 2000.
- A. S.W. Wong, R. Gopalkrishnan, G. Sarkar, and L. F. Wang, *J. Electrochem. Soc.*, **147** (5), 1925–1929, 2000.

16 High-Resolution TEM in Microelectronics



Carbon replica of deep sea mud showing fossil skeleton of early microbes.

Besides the examples mentioned in the preceding chapters, there are many applications in the semiconductor process characterization that utilize high-resolution TEM (HRTEM). This is largely because process technology has evolved into the deep submicron and nano-technology era. All the dimensions diminish and the interface property dominates. HRTEM, for this reason, is becoming the most important means for characterizing and identifying possible failures in the process.

16.1 METROLOGY IN ULSI PROCESSES

As ULSI technology evolved below 100 nm, the device's physical dimensions and critical size, which affects the device's properties, reached proportions beyond the resolution capability of conventional SEM. Even though the improvements in SEM

technology show it to be catching up, TEM analysis and, in particular, HRTEM capability, are becoming indispensable tools. A good example is the use of TEM-related techniques to measure the thickness of ultra thin gate dielectrics. The gate oxide thickness required for future generation ULSI devices is challenging the physical and theoretical limits of SiO_2 . Conventional SiO_2 modified by introducing nitrogen is a temporary solution, and the high- k dielectric will eventually replace SiO_2 -based gate materials (see Chapter 6 for more details on this discussion). However, gate oxide physical thickness measurement is not a straightforward task even with HRTEM capability. Here we will discuss the difficulties and challenges that arise in using TEM-related techniques to measure the ultra thin gate dielectric thickness. The challenges and objectives are to determine the absolute physical thickness of the gate dielectrics (SiO_2 , SiON , and high- k), as well as the chemistry and distribution of nitrogen across the nitrided SiON gate dielectrics, and the oxygen content, distribution, and chemical bonding within the new high- k dielectrics, for example, Hf-Al-O compound oxides.

Conventional HRTEM

Figure 16.1 shows cross-sectional lattice images of an ultra thin SiON gate dielectric. Close examination of the lattice points reveals the interface roughness but the smooth and consistent interface line is difficult to determine. Interface atomic steps and atomic ledges, usually around 2 to 3 Å in thickness for $\text{Si}(100)$ substrate, have two effects on the thickness measurements. First, the interface roughness along the horizontal direction, as shown in Fig. 16.1, changes the oxide's thickness locally, so the exact thickness varies depending on how and where it is measured (Muller 2002). Second, the interface roughness along the TEM sample thickness direction can distort the lattice image, making the determination of the exact end-lattice-point very difficult, as illustrated in Figs. 16.1 and 16.2. Both problems are introduced by interface atomic steps, ledges, and terraces. The issue is not significant when gate oxide's thickness is more than 50 Å, where an atomic step is less than 5% of the overall thickness and, in general, can be ignored. The problem becomes serious as the gate oxide's thickness becomes thinner and thinner. For a 15 Å gate oxide sample, an atomic step can introduce more than 20% of local thickness uncertainty, a number apparently not acceptable for metrological purposes. A reliable HRTEM image should be accompanied by a focal series reconstruction (FSR) and an exit wave reconstruction (EWR) to obtain the best results (Principe et al. 2002).

Another often encountered problem when using the conventional HRTEM image to measure the gate oxide's thickness is shown in Fig. 16.3. A dark band within the Si substrate along the SiO_2/Si interface can be observed. The dark band often occurs in a certain sample thickness range but not in sample thicknesses that are either thinner or thicker than this thickness range. The phenomenon is due to extinction oscillations and can be avoided by choosing the very thin sample thickness areas, which have a thickness less than the extinction distance (Principe et al. 2002). Extinction oscillation should not be confused with contrast modulated dark bands induced by local lattice strain (Roy et al. 2001). Both extinction oscillations and strain-induced contrast modulation can generate a dark band at the Si/SiO_2 interface that is often difficult to distinguish in the image. Generally speaking, careful sample thickness measurement

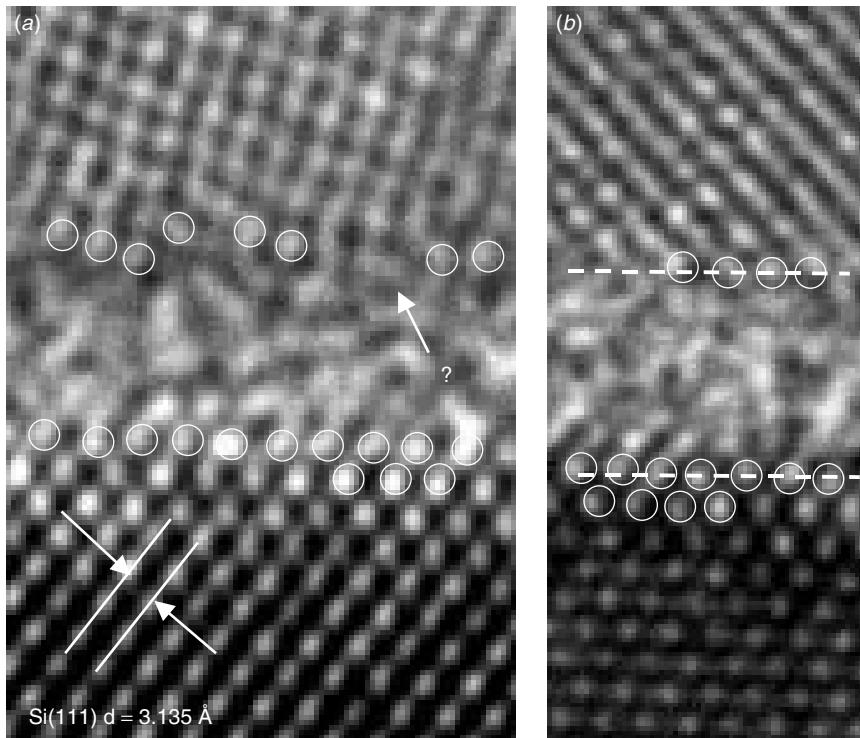


Figure 16.1 HRTEM lattice cross section showing ultra thin SiON gate dielectric thickness. (a) Notice that the interface roughness can interfere with lattice image and make it difficult to determine the interface's exact position. (b) When the interface is smooth, the exact thickness can be measured with reasonably consistent results. The “?” marks the area where the interface lattice point is hard to determine.

(using EELS) and lattice strain calculation (Pennycook 2002) are advisable when a dark band is observed in order to distinguish the ambiguities. The lattice strain calculation for the Si/SiO₂ interface will be discussed later.

High-Resolution Scanning TEM (HR-STEM), HAADF Detector, and EELS

An alternative way to measure ultra thin gate dielectric thickness is to use high-resolution scanning TEM (HR-STEM). Compared with the conventional HRTEM, HR-STEM makes use of the ultra fine scanning electron beam to probe the sample and pick up the transmitted electron signals. Channeling prevents probe spreading, and the resolution depends on the probe's size. The lattice observed in this case will be the true projection of the atomic columns. The latest commercially available field emission STEM provides an electron beam whose diameter is in the 1.4 to 1.9 Å range, and this is sufficient for a gate oxide metrological purpose. STEM imaging usually involves a special detector called a high-angle annular dark field (HAADF) detector. In principle, the detector collects incoherently scattered electrons. The technique is very sensitive

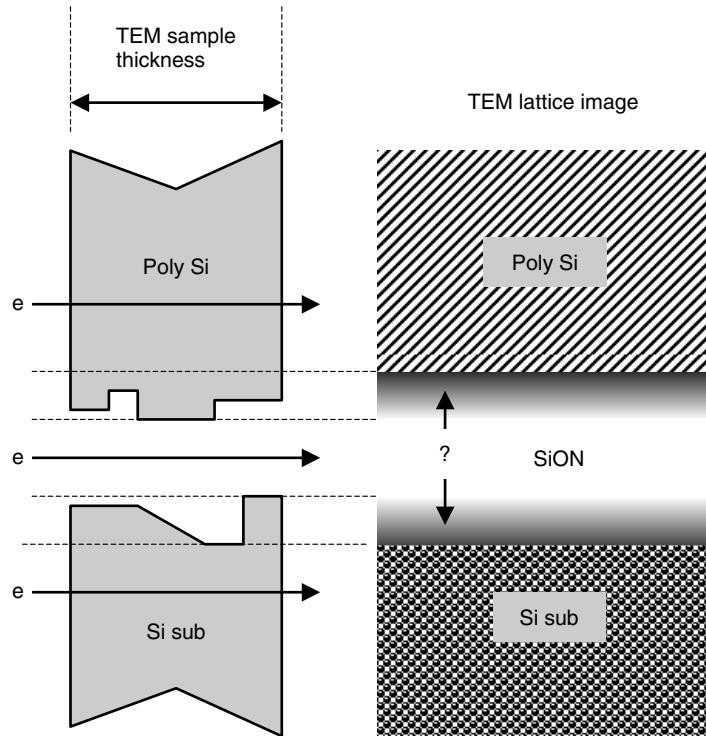


Figure 16.2 Schematics of the interface roughness. The atomic steps at the interface along the sample thickness direction directly affect the lattice image's quality, and it becomes a challenge to determine the interface's exact position, as was indicated by a "?" in Fig. 16.1.

to atomic numbers (Z-numbers) in the sample atomic columns, and thus it is usually called Z-contrast imaging. The Z-contrast is in compensated electron energy loss spectroscopy (EELS) analysis because both use exactly the same probe-forming optical conditions, and EELS is analytically sensitive to the lighter elements that usually do not show up well in Z-contrast images. The EELS spectrum also contains information on chemical structure such as plasmon and inter band transitions, dielectric functions, band gap states, bonding (coordination and local atomic environment), and oxidation states. This is critical information in ultra thin SiON and high- k gate dielectrics studies.

Figure 16.4 shows a STEM lattice image obtained by a HAADF detector. The image is not as impressive as a conventional HRTEM image and appears to be quite noisy. This is the nature of HR-STEM HAADF images. Contemporary STEM with fully computer controlled data acquisition capability enables us to take the HAADF image, EELS spectra, and EDS spectra all at the same time from exactly the same area. Such capability allows a full range of elemental information (from light to heavy elements), chemical bond information (EELS spectrum), and microstructural information (HR-STEM image) to be obtained simultaneously at atomic scales. Figure 16.5 shows an example with such capability. The sample is a Si substrate deposited with 30 Å of HfO₂ and 2000 Å of polysilicon. The challenge is to observe and analyze the HfO₂/Si-sub interface. It is known that SiO_x always forms in between the HfO₂ and Si substrate

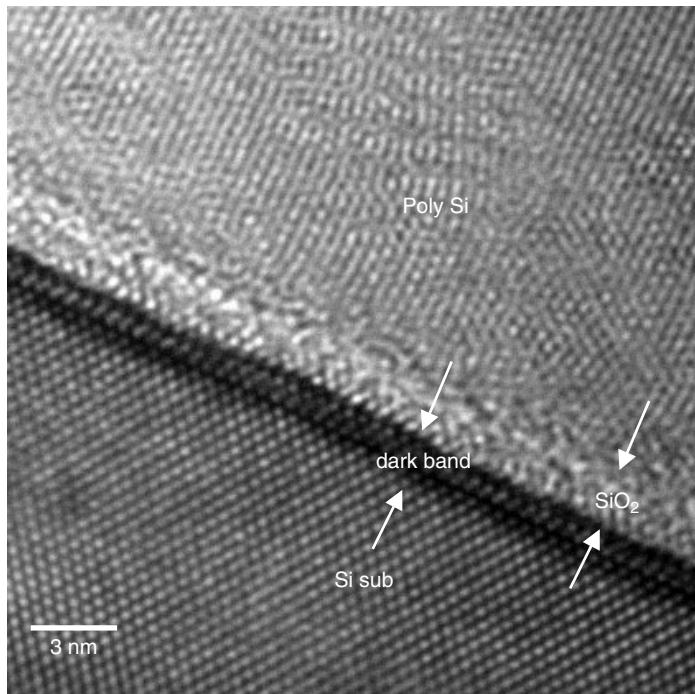


Figure 16.3 Conventional HRTEM image often shows a dark band along the interface in the Si substrate near SiO₂, as indicated. This dark band can interfere with the observation and make the image hard to interpret.

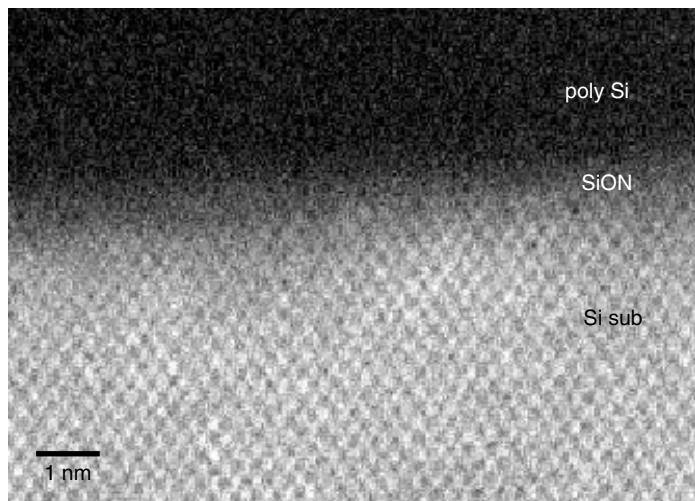


Figure 16.4 HR-STEM image as captured by the HAADF detector. The lattices observed are the Si(111) atomic planes. The image does not appear as impressive as conventional HRTEM images. However, a lot more chemical information can be extracted from it.

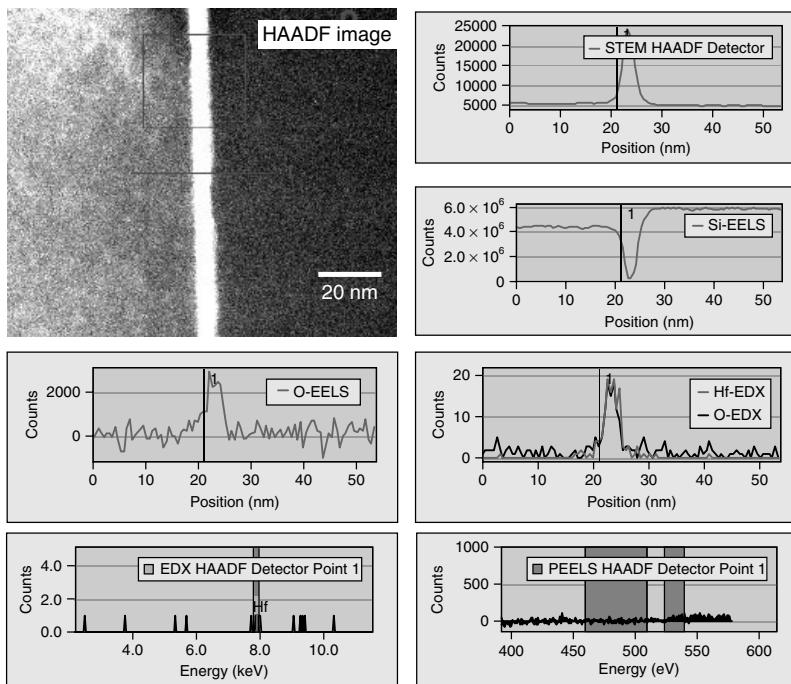


Figure 16.5 HR-STEM image taken by the HAADF detector and simultaneously both EDS and EELS spectrum for each and every pixels are also captured. Detailed chemical information can be retrieved by pointing to any position on the HAADF image. (Courtesy Dr. B. H. Freitag, FEI/Philips Eindhoven, The Netherland)

but its chemical state is not clear. This SiO_x layer is very thin, less than 7 Å, and most analytical techniques do not have the resolution power needed to study it.

An interesting test is that of the consistency of the gate oxide's thickness measured by different techniques in the same microscope. Since the different techniques use different principles and are essentially detecting different phenomena from the same sample region, one could expect to see different thickness results. Generally speaking, with careful aberration correction and measurement procedures, conventional HRTEM can achieve 0.4 to 0.7 Å accuracy in measuring SiO_2 or the SiON gate dielectric. STEM with Z-contrast, on the other hand, shows a slightly thicker gate oxide measurement result (about 2.5 Å thicker). Taking into account of interface roughness and other factors, the difference between HRTEM and STEM could simply reflect the difference in physical and chemical thickness of the gate (Principe et al. 2002).

16.2 SiGe AND GaAs HETERO-JUNCTIONS

The potential for using strain SiGe as strained channel substrate was discussed in Chapter 14. In addition silicon-Germanium (SiGe) technology has gained enormous attention for the potential application of the hetero-junction bipolar transistors (HBT) device to replace GaAs-based technology (Ning 1998). Other than the obvious device

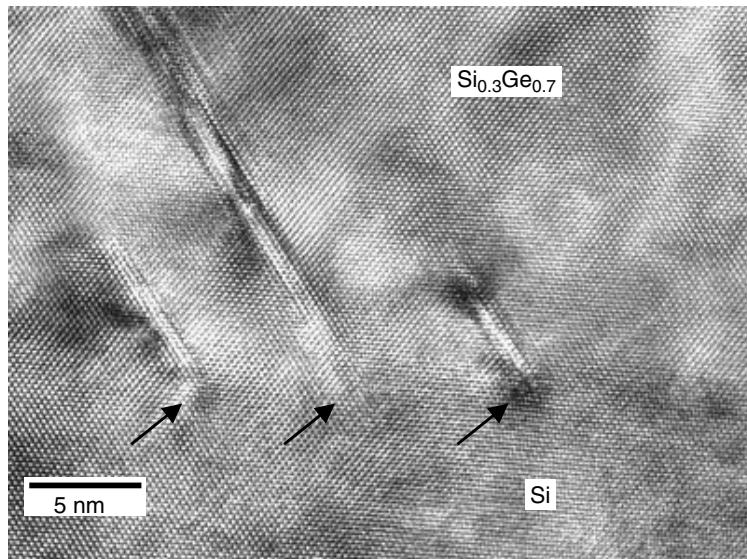


Figure 16.6 HRTEM of the low temperature Si–Ge process and defect study, showing the low-temperature $\text{Si}_{0.3}\text{Ge}_{0.7}/\text{Si}$ interface’s structural defect. HRTEM provides detailed structural information on the defect, but the contrast of the interface is difficult to determine directly from an HRTEM image.

characteristic advantages, the real advantage of the SiGe bipolar transistor lies in its compatibility with silicon VLSI processes. One example is that by growing SiGe (and GaAs) on top of a Si substrate, the bipolar device can be integrated into the conventional CMOS device. The SiGe lattice mismatch to the Si substrate is smaller than GaAs. Figure 16.6 shows a low-temperature $\text{Si}_{0.3}\text{Ge}_{0.7}$ on a Si substrate. The planar fault defects along $\text{Si}(111)$ originating from the interface can be studied with HRTEM, and these results can help improve the deposition process. Compared to the SiGe, GaAs will require more process effort to accommodate and eliminate the mismatch strain when applying the Si substrate. Figures 16.7 and 16.8 show a typical GaAs on a Si device. A buffering structure called strained layer superlattices (SLS) is employed to filter out the dislocations that are generated at the GaAs/Si interface. With the help of SLS, a clean and defect-free GaAs layer can be obtained (Augustus et al. 1988). However, the process is far more complicated than that of applying SiGe on Si, and it requires some intricate manipulation.

GaAs and other III–V compound semiconductor materials studies have always depended on TEM and HRTEM. There are many examples to be found in the biaural proceedings of the Microscopy on Semiconductor Materials conferences (e.g. Institute of Physics, UK, 1997, 1999). Different TEM techniques have been also combined, for example, HRTEM, convergent beam, weak beam dark field (WBDF), HR-STEM/HAADF, and EELS in order to study the various aspects of the heterostructure in detail. A typical example is shown in Figs. 16.9 and 16.10, where the complicated commercial GaAs device structure is revealed by using the weak beam dark field (WBDF) technique along with HRTEM. Such analysis not only has scientific value but also provides the GaAs device industry with a strong quality control tool. Note

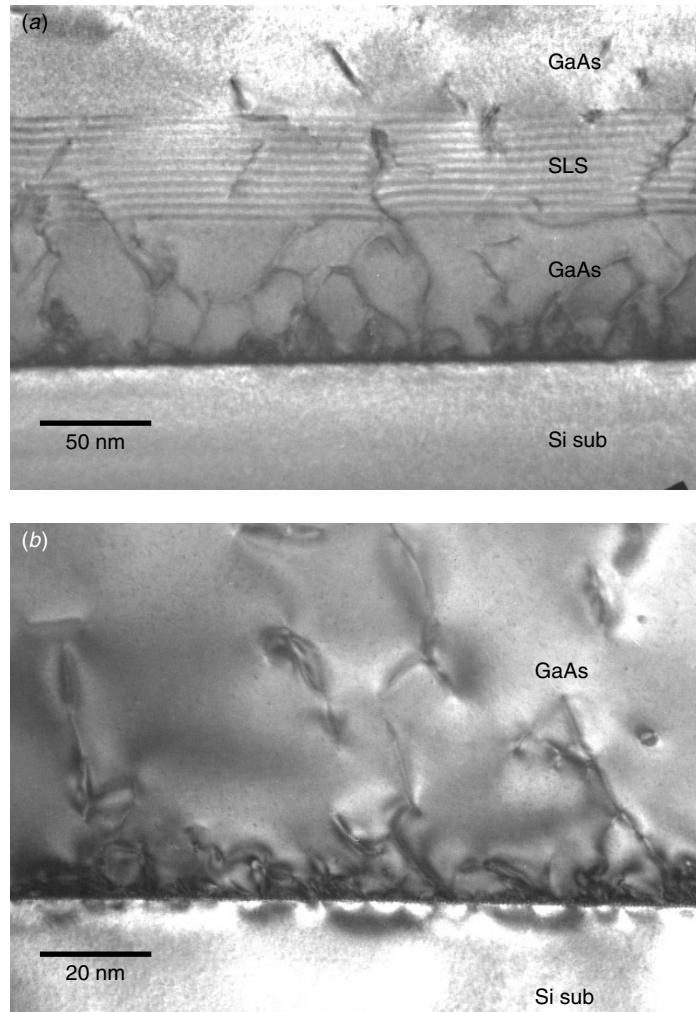


Figure 16.7 TEM cross section of (a) GaAs on the Si substrate with the SLS buffer structure. The SLS is designed to buffer out the dislocations generated at the original GaAs/Si interface, as shown in (b), to create a defect-free GaAs on the Si substrate.

that the interface between GaAs and GaInAs is not distinguishable under the normal HRTEM imaging condition along the (110) pole. This difficulty can be resolved by cutting the cross-sectional sample along the (100) plane. Figure 16.11 shows the difference immediately.

16.3 LATTICE STRAIN MEASUREMENT AND SELF-ASSEMBLED QUANTUM DOTS (SAQD's)

Local lattice strain in a confined area can be measured using high-resolution transmission electron microscopy. There are several computational techniques that can be

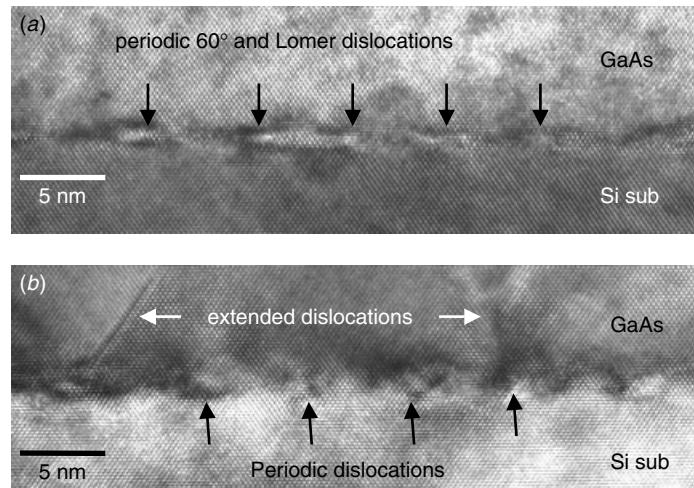


Figure 16.8 HRTEM of the GaAs on Si substrate with the SLS buffer structure. The GaAs/Si interface shows different periodical dislocations generated due to the lattice mismatch.

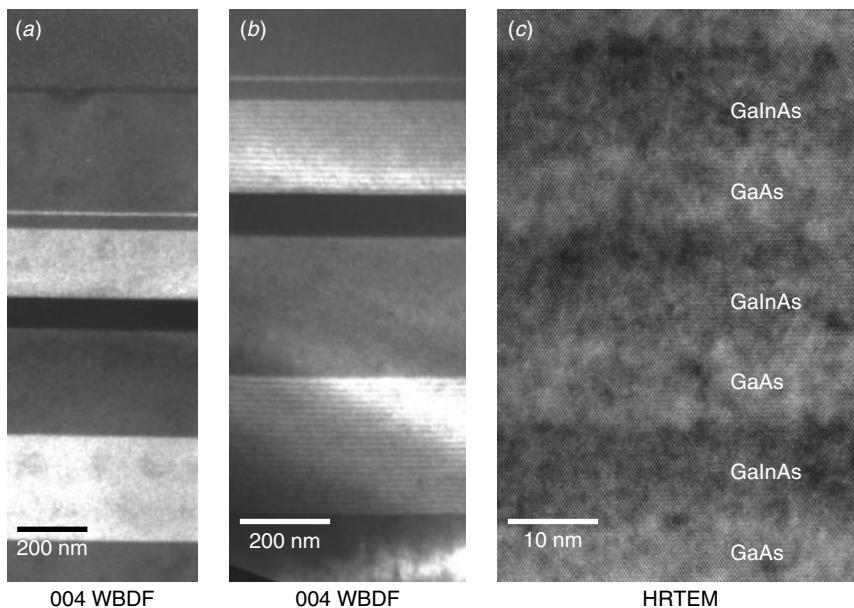


Figure 16.9 GaAs superlattice. (a,b) Weak beam dark field, and (c) HRTEM images showing both the contrast and resolution needed for process study and quality control.

used to extract the subtle lattice distortion from the lattice images and calculate the lattice strain. For example, high-resolution lattice images can be processed by taking the cumulative sum of deviations (CUSUM) of the lattice fringe's spacing from a target value and then calculating the local lattice strains. The strain distribution along different directions, say, the [001] direction in the (110) plane, can be measured with

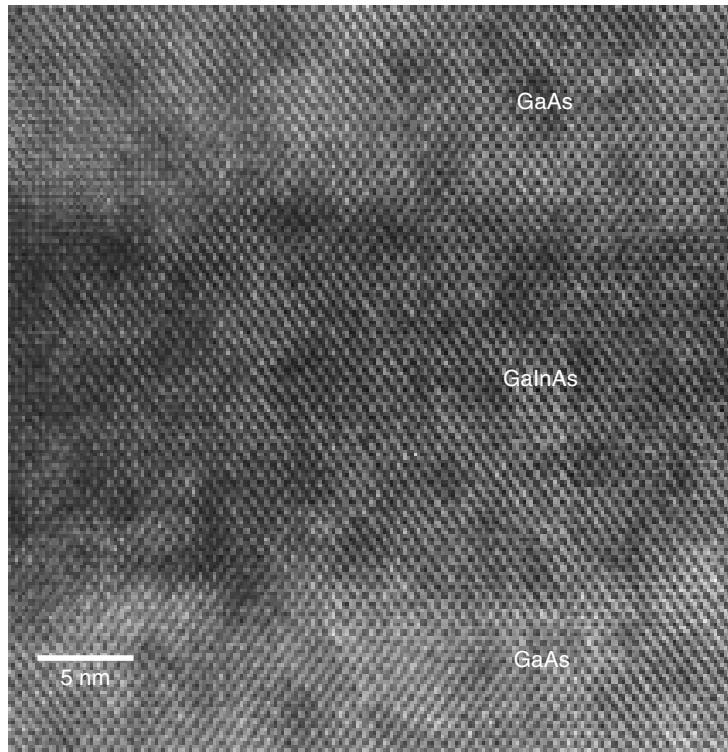


Figure 16.10 GaAs superlattice. The HRTEM image over the GaAs/GaInAs superlattice showed no dislocations. Notice that the GaAs/GaInAs interface is hardly distinguishable in the (110) pole of the HRTEM image.

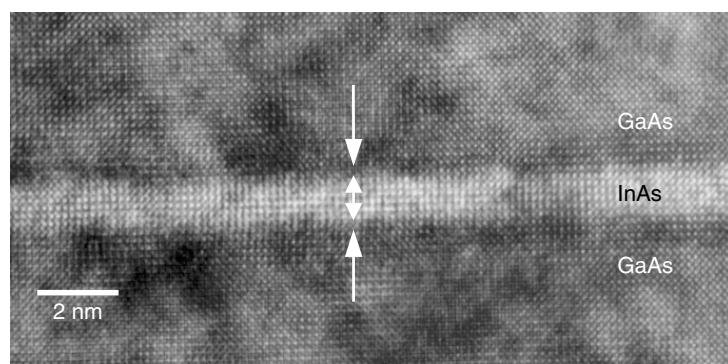


Figure 16.11 GaAs superlattice. The HRTEM image over the GaAs/InAs superlattice shows the distinctive interface position. The sample is cross-sectioned along the Si(100) lattice plane instead of the conventional Si(110), and the HRTEM image taken along the Si(100) pole. The lattice fringes of GaAs[200] can be observed. The exact thickness of the InAs layer can now be determined more clearly to be 14.2 Å.

results that are in agreement with theoretical predictions (Robertson et al. 1995a,b). An alternative method to calculate the local lattice strain is to obtain a moiré pattern of the two lattice images as they are overlapped. As the first lattice is given by the experimental image, the second one is calculated from an undisturbed area of the first. In the resulting moiré structure any local displacement of the crystal lattice with relation to the calculated lattice can be deduced (Bierwolf et al. 1993; Jouneau et al. 1994). This method has been used for identifying interfaces within the superlattice and mapping the ultra thin epitaxial layer(s) within the matrix lattice. In this section, examples on calculating local lattice strain within self-assembled (or aligned) quantum dots (SAQDs) will be given.

CUSUM Application in InAs/GaAs SAQDs

Self-assembled (or aligned) quantum dot (SAQDs) structures, either in the SiGe-based system (Peng et al. 1998) or the InAs–GaAs based system (Solomon et al. 1996), have demonstrated novel electronic and optical properties. Strain is known to be the driving force for the creation of Stranski-Krastanov grown self-assembled islands, or dots. But the size and size distribution of the dots with respect to the growth conditions need to be understood. Local crystal lattice stress/strain has been studied to understand the growth and formation mechanisms of the quantum dot islands (Solomon et al. 1996; Tersoff et al. 1996; Xie et al. 1995). For direct measurement of strains in epitaxial structures on the nanometer scale, high-resolution transmission electron microscopy images recorded by slow scan charge-coupled device (CCD) cameras and analyzed using Fourier filtering techniques have been shown to provide reliable, quantitative results that are independent of image contrast (Robertson et al. 1995a). The method is based on the analysis of the cumulative sum of deviations (CUSUM) of lattice fringe spacing from a target value (Robertson et al. 1995b). This method has been shown to be very sensitive to lattice strain, and independent of the surface relaxation within a reasonably large sample thickness range.

The InAs/GaAs sample was grown using a Riber 49 MBE system. A 4-inch semi-insulating GaAs (100) substrate was slightly etched before loading into the MBE system. The growth rates were 0.70 $\mu\text{m/h}$ for GaAs, 0.84 $\mu\text{m/h}$ for $\text{In}_{0.17}\text{GaAs}$, and 0.11 $\mu\text{m/h}$ for InAs. After the oxide's desorption, a 300 nm GaAs buffer was grown at the oxide's desorption temperature of 580°C. The temperatures were measured by a pyrometer. Next a 3 nm $\text{In}_{0.17}\text{GaAs}$ layer was grown at 500°C, followed by a 100 nm GaAs at 580°C. Then a 10 period 0.6 nm/10 nm InAs/GaAs quantum dot superlattice was deposited at 500°C. Finally, a 20 nm GaAs cap layer was grown at 580°C. Reflection high-energy electron diffraction (RHEED) patterns were used to monitor the growth. InAs was deposited at a rate of 0.1 monolayer (ML) per second with 8-second growth interruptions after every 0.4 ML InAs. The RHEED pattern changed from streaks into spots after 1.6 ML InAs was grown. A total of 2 ML InAs was grown. The RHEED pattern changed rapidly from spots to streaks and resumed the 2×4 pattern as a GaAs layer was grown over the InAs quantum dots. The substrate was rotated at a speed of about 20 rpm during the growth. Cross-sectional transmission electron microscopy samples were prepared using grinding and polishing techniques, and subsequent ion milling. To reduce possible artifacts, the ion milling time of the sample was kept below 10 to 15 minutes. High-resolution digital TEM images were first obtained, and fast Fourier transform (FFT) analyses were performed on the selected

areas. The (002) reflections were selected and filtered in for inverse FFT to obtain the (002) lattice fringes. The fringe intensities as a function of pixel position were digitized from the filtered (002) lattice fringe images, and their polynomial and peak positions were used for the CUSUM computation. The CUSUM technique calculates the sum of deviations of the individual lattice fringe spacing from a target value. The average fringe spacing is a useful choice for the target value. The regions of the CUSUM plot where the lattice fringe spacing is less than the average value will have negative slope, while regions where the lattice fringe spacing is greater than the average will have a positive slope. From the slopes in different areas in the same CUSUM plot, the misfit perpendicular to the interface, measured with respect to a reference region (assumed to be a strain-free lattice region in this study), can be calculated by a simple relationship:

$$e = \frac{(m_2 - m_1)}{(1 - m_2)}$$

where m_1 is the slope of the CUSUM plot in the strain-free matrix lattice area and m_2 is the slope in the area where the strain is to be calculated.

The cross-sectional TEM images of the sample showed distinctive vertically aligned InAs dot columns, as seen in Fig. 16.12. The island height, measured visually using image contrast, was found to be approximately 6 nm, and the in-plane dimension was

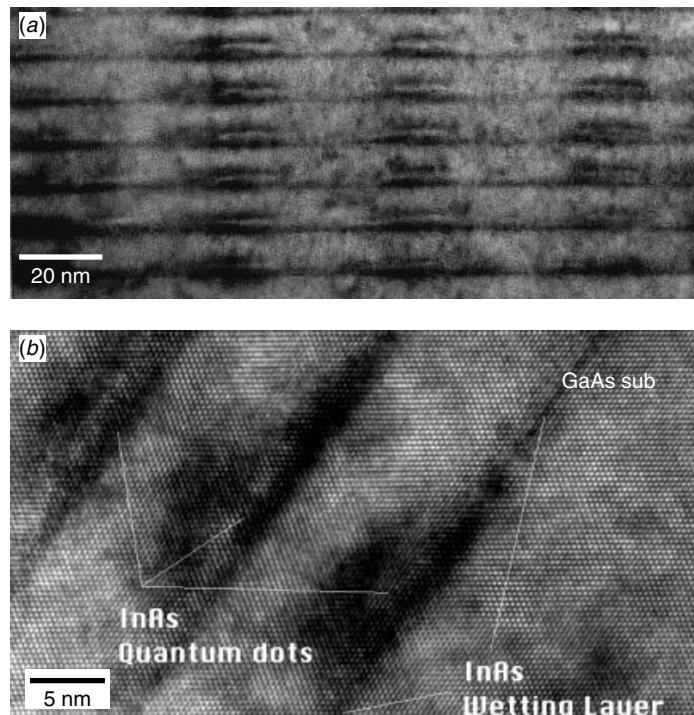


Figure 16.12 (a) Conventional bright field image showing InAs SAQDs to align with each other among the layers. (b) HRTEM of the InAs self-assembled quantum dots (SAQDs) formed in the GaAs matrix.

approximately 20 nm. The observed shape of the InAs dots is more like a liquid wetting a substrate with a wetting angle of approximately 30° than like a pyramidal used for theoretical calculations (Grundmann et al. 1995). It should be noted that it is difficult to accurately determine the island's height as the island peaks have only a small number of InAs atoms in cross section that contribute to the contrast, and the observed cross section may not necessarily be at the island's center. Exactly the same problem can be attributed to the strain calculations as will be discussed later.

Figure 16.13 shows a fast Fourier transformation (FFT) from the lattice image, a pair of filtered (002) spots, and an inverse FFT image on which are visible only the (002) lattice planes. As we saw earlier, the filtered (002) lattice plane can be used for the CUSUM plot. Figure 16.13(d) shows a typical single InAs dot (rotated 90°) along with the CUSUM plot extracted from the dot center area which is perpendicular to the interface. The strain as plotted in the CUSUM is calculated using the dot's center, the dot's edge, half a radius off the dot's center, and the wetting area as shown in Fig. 16.14. Table 16.1 shows how the strain's measurements based on the high-resolution TEM and CUSUM calculations compare with theoretical calculations (Grundmann et al. 1995).

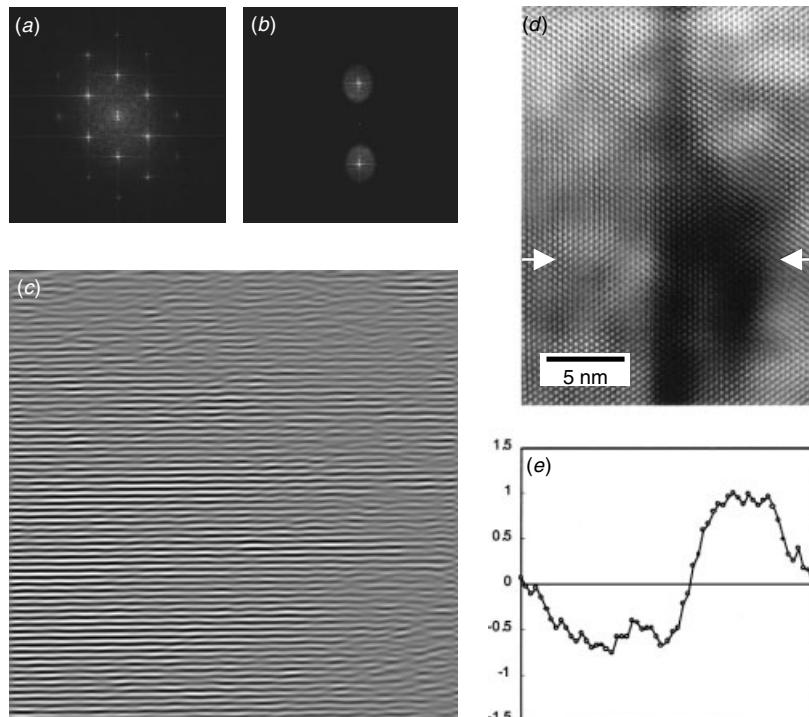


Figure 16.13 (a) Fast Fourier transform (FFT) of a selected HRTEM area, (b) a pair of filtered (002) lattice spots, and (c) a reverse FFT image that shows only the horizontal (002) lattice planes. (d) An example of the InAs SAQD lattice image (tilted 90° clockwise), and (e) the CUSUM plot corresponding to the InAs bump's center as indicated by arrows in (d). (Tung et al. *Microscopy Semicond. Mat. Conf.* 1999, 47–50, reprint with permission from Inst. of Physics, Publishing Ltd., London)

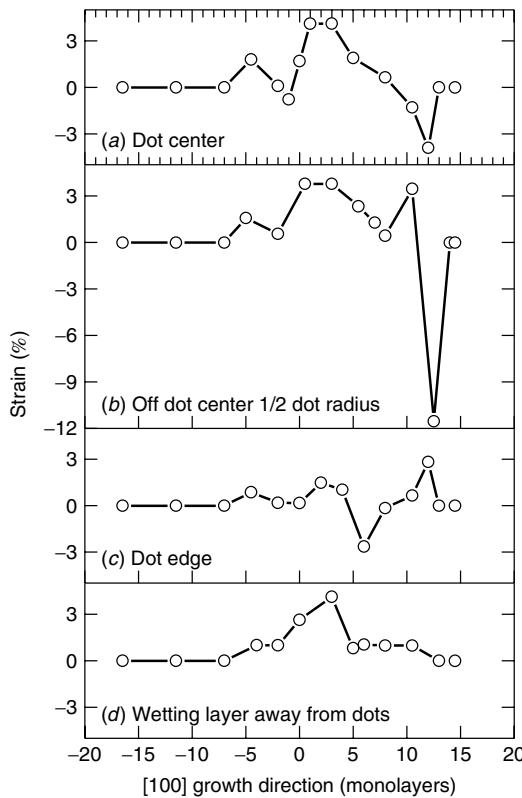


Figure 16.14 Lattice strain (%) calculated from the CUSUM plots. (a) Dot's center, (b) off the dot's center at about half a dot's radius, (c) dot's edge, and (d) wetting layer away from any dot. The horizontal axis is chosen so that the InAs wetting layer is from 0 to about 2.5 monolayers and the InAs dot is from 0 to about 11 monolayers of GaAs lattice. (Tung et al. *Microscopy Semicond. Mat. Conf.* 1999, 47–50, reprint with permission from Inst. of Physics, Publishing Ltd., London).

TABLE 16.1 Comparison Among Theoretical Predictions, STM Measurements, and Current HRTEM Measurements of Local Lattice Strain within and Around InAs Self-assembled Quantum Dots

	Strain (%)	Theory (Grundmann et al. 1995)	STM (Legrand et al. 1998)	HRTEM Current Study (1999)
Dot's center	maximum	3	7	4
	minimum	-9	-15	-4
Half off center	maximum	3.5	—	3.8
	minimum	-2.5	—	-11
Dot's edge	maximum	—	—	3
	minimum	—	—	-3
Wetting	maximum	7	—	4
	minimum	0	—	0

(Tung et al. *Microscopy Semicond. Mat. Conf.* 1999, 47–50, reprint with permission from Inst. of Physics, Publishing Ltd., London)

Also shown in the table for comparative purposes are the results from the scanning tunneling microscope (STM) measurements (Legrand et al. 1998). The highest measured tensile strain in the present study is approximately 4%, and it is located at the lower interface across the dot's center, in reasonable agreement with the theoretical calculations (Grundmann et al. 1995). The highest compressive strain of -11.5% is observed at the dot's top surface, about half a radius away from the dot's center. Although the theoretical calculations also predicted a compressive strain at the top surface (-9%), it was to occur at the apex of the pyramidal dot. It should be noted that in the theoretical model the dot was assumed to be pyramidal in shape whereas a crescent shape was observed in the experiments. An inaccurate determination of the dot's center in the images may contribute to this discrepancy. Note that the STM measurements show significantly larger magnitudes of both the compressive and the tensile strains. It is further possible to determine, from the CUSUM plots, the actual heights of the InAs dots and the wetting layer thickness. The present study showed that the InAs dot height is about 6.2 nm and the wetting layer thickness is about 1.5 nm. This is very close to some of the previously published data (Soloman et al. 1996; Grundmann et al. 1995). In effect, several important features have been experimentally verified in this study. These include the high tensile strain at the lower interface of the dot's center, the change in the sign of the strain (from tensile to compressive) at or near the center of the dot's cross section, and the high compressive strain at or near the top surface of the dot.

There are, however, several differences from the theoretical predictions among the current measurements. One noticeable difference not predicted by theory is the consistent presence of a small region with a tensile strain of about 1.5% near the InAs/GaAs interface, which is at the bottom area near the dot's center. Although the origin of this strain is not clear, its expansion into a hump may coincide with the large compressive strain on the dot's peak, as this induces the self-alignment mechanism for the dots in the subsequent layers. Another difference from the theoretical calculations and the measurements is the more gradual change in strain over the dot's center and the wetting areas. Because the high-resolution images were taken with the lattice spacings averaged over the sample thickness, a more-smoothed strain measure was expected. As the dot's diameter measures about 20 nm at the bottom area, the full TEM sample thickness, estimated to be about 50 nm in the present analysis, covers not only the InAs dot area but also the GaAs matrix lattice. This inevitably will average out the lattice strain and result in a less severe strain profile. The influence of such broader sampling needs to be studied quantitatively through HRTEM image simulation.

In summary, demonstrated this study has that the local lattice strain within and around InAs self-assembled quantum dots can be measured using high-resolution TEM images. The strains derived from the slopes of the CUSUM plots of lattice fringe spacings proved to be in reasonable agreement with theoretical predictions as well as with results from STM measurements (Tung et al. 1999).

SiGe SAQDs in Si Matrix

A similar effort has under taken to apply the CUSUM method to the SiGe SAQDs formed in the Si matrix (Tung et al. 1998). Ge self-assembled quantum dots (SAQDs) grown in Si/Ge short-period superlattices (SLs) by molecular beam epitaxy (MBE) had been reported to possess interesting quantum and optical properties (Peng et al. 1998a,b). Already the potential to fabricate efficient silicon light-emitting devices using

this material had drawn much research and development effort. SiGe island (also QDs) formation in Si grown by MBE was shown to be very sensitive to growth conditions. The islands' sizes and size distributions with respect to growth conditions still needed to be understood. Local crystal lattice stress/strain appeared to be key toward this understanding of the growth and formation mechanisms of quantum dot islands.

SiGe samples were grown in a V80S MBE system with structural SLs, $[(\text{Ge}_4\text{Si}_4) \times 4 + \text{Si } 5 \text{ nm}] \times 3$. The growth temperature was 550°C on a 100 nm Si buffer layer. The buffer layer was grown at 800°C on a Si(001) substrate. The 5 nm Si layer between the three $(\text{Ge}_4\text{Si}_4) \times 4$ SLs was used to reduce the strain energy. A total of 1 ML of Sb was deposited to increase the nuclei of the Ge islands. The growth process was monitored *in situ* by reflection high-energy electron diffraction (RHEED). The existence of quantum dots of a novel Ge self-organizing island structure was confirmed by photoluminescence (PL) and HRTEM images (Peng et al. 1998a,b). The cross-sectional transmission electron microscopy samples were prepared by conventional grinding and polishing, and subsequent ion milling. To reduce the possible artifacts, the ion milling time of the sample took no more than 10 to 15 minutes. The success of the samples thus depended mainly on the controlled ion milling time. The same CUSUM technique was then used to calculate the local lattice's distortion. Figure 16.15 shows the HRTEM image of the SiGe SAQD structure. A corresponding plan view HRTEM is shown in Fig. 16.16.

Qualitatively, in the wetting areas the Si to Ge transition slope appeared steeper than in the island peak areas. This feature corresponds to the stress relaxation mechanisms suggested by some island formation theories (Barabasi 1997). Our calculations indicate that the stress accumulates gradually and drops abruptly at the top of each SLs $(\text{Ge}_4\text{Si}_4) \times 4$. This stress transition is more dramatic in the wetting areas than in the island peak areas. We also observed that the lattice CUSUM curves reach their peak value at or near the third Ge_4 within each $(\text{Ge}_4\text{Si}_4) \times 4$. This was confirmed in

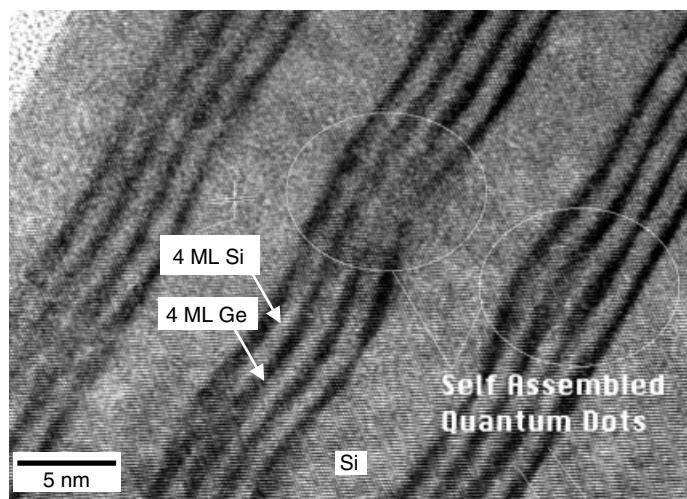


Figure 16.15 SiGe SAQDs form in the Si substrate. An SLs $[(\text{Ge}_4\text{Si}_4) \times 4 + \text{Si } 5 \text{ nm}] \times 3$ structure is formed. The bump strain can penetrate into the buffer Si (5 nm) and propagate to the next lamination.

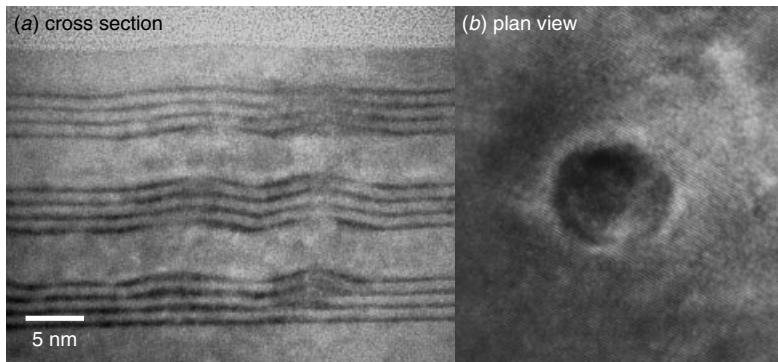


Figure 16.16 Si–Ge SAQDs form in the Si substrate. An SLs[$(\text{Ge}_4\text{Si}_4) \times 4 + \text{Si } 5 \text{ nm} \times 3$] structure is formed. The HRTEM plan view shows that a single dot and its associated lattice strain induced the dark contrast.

TABLE 16.2 CUSUM Calculated Lattice Distortion of each Ge_4 in a SL from HRTEM Images

Lattice Distortion (%)	1st	2nd	3rd	4th
Wetting area	4.55	2.35	2.37	-2.32
Island area	2.47	2.29	2.27	1.3

Note: “1st” denotes the Ge_4 at the bottom next to Si substrate and 4th represents the top Ge_4

both the wetting and the island areas. Coupled with lattice elasticity parameters, the lattice’s distortion, and thus local lattice’s strain in the wetting areas and in the island areas, could be calculated. The theoretical lattice misfit between Ge and Si was found to be around 4.2%. The amount of lattice misfit parallel to the interface normal was calculated as given in Table 16.2. The calculation was made from the bottom SL as the first SL, and it represents the first Ge_4 from the bottom of the SL. Overall the largest measured misfit appeared in the first Ge_4 of the wetting areas and the lowest misfit in the fourth Ge_4 of the wetting area. For both the wetting and island areas the highest strain occurred in the first Ge_4 and the lowest strain occurred in the last Ge_4 , implying a stress build up within the SLs that is subsequently relieved to the buffer Si.

Such calculations present, for the first time, a way to understand that the quantum optical property of the SAQD is directly linked to the local lattice’s microstructural distortion. Indeed, the strain as articulated by the quantum dots deduced from the phonon energies is consistent with the results of our HRTEM study (Peng et al. 1998c, 1999).

Derivatives of Displacement: An Alternative Lattice Strain Calculation

An alternative method to calculate lattice strain, using the same high-resolution lattice image, is called derivatives of displacement. The basic idea is to analyze the moiré

The author wishes to thank Dr. Roar Kilaas for his generous help and discussion during the course of acquiring and implementing the NCEM share ware in our TEM system.

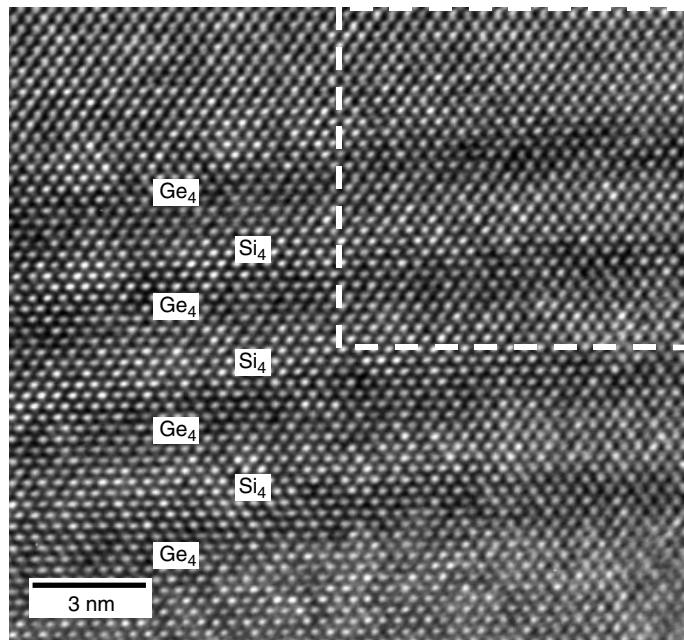


Figure 16.17 Si–Ge SAQDs form in the Si matrix. Within one SL, the strain relationship can be derived. The upper right-hand corner of this image (boxed) is used to calculate the lattice displacement as shown next in Fig. 16.18.

formed when two lattice images are overlapped. The first lattice is given by the experimental image, and the second one is calculated from an undisturbed area of the first. From the resulting moiré the local displacement of the crystal lattice with respect to the calculated lattice can be deduced. (Bierwolf et al. 1993; Jouneau et al. 1994) The distortions are due to the parameter differences between the two lattice materials and the elastic deformation of the strain layers. The basic procedure of this method is the same as that of the CUSUM plot, where internal reference points are picked from the strain-free matrix within the same image and used to eliminate any possible artifacts that come from the sample preparation, TEM illumination, and imaging errors. However, due to the need to calculate a large number of lattice points in order to obtain detailed position information and fewer artificial errors, the computation power required in this method is much more demanding. The computations involving a large area, for example, in SiGe to calculate all three SLs including the interim buffer Si in one single image frame, can be quite time-consuming. Figures 16.17 through 16.19 demonstrate this approach. A reasonably large area needs to be chosen from an HRTEM lattice image, Fig. 16.17. The lattice points are annotated individually by an image-processing algorithm. Within the area a reference section must be identified where the lattice does not have significant distortion. The reference lattice is then calculated and overlapped with the original annotated lattice points, Fig. 16.18. Finally the local displacement of each lattice point can be extracted (Kilaas et al. 1998), Fig. 16.19. Notice that depending on the reference matrix used, the displacement can be downward, upward, or even sideways. The excellent quantitative agreement obtained by the derivative of displacement compares well with the CUSUM method.

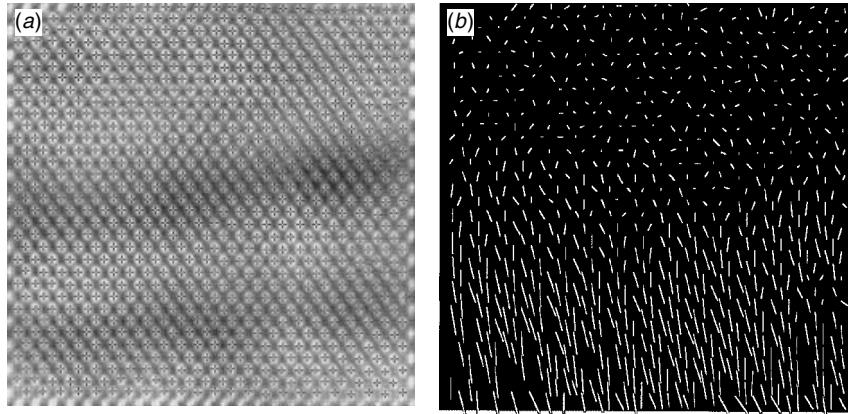


Figure 16.18 Si-Ge SAQDs form in the Si matrix. (a) The annotated lattice points are overlapped with the original lattice image. The upper half of this image is used as a reference matrix where the Si is regarded as a strain-free lattice. (b) The calculated displacement lines are shown on the right-hand side.

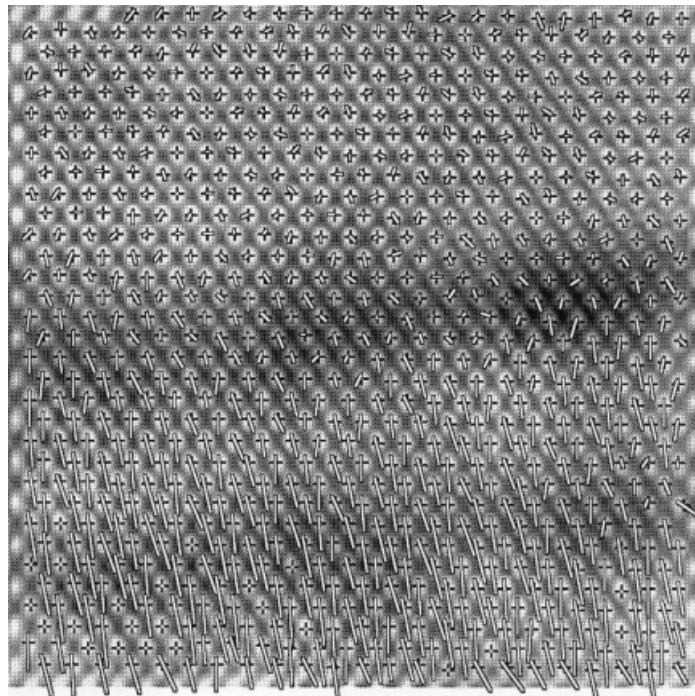


Figure 16.19 Si-Ge SAQDs form in the Si matrix. The annotated lattice points are overlapped with original lattice image as well as with the calculated displacement lines. The lower dark area corresponding to the Ge₄ is shown to be associated with the lattice displacement by the downward-pointing lines.

Defects and Defect Generation in SAQDs

When the strain in SAQDs is raised to a critical value, either by changing the process parameters or by introducing other impurities, the lattice distortion will no longer be effective in accommodating the stress, so lattice defects will start to form. One example that can be given in the SiGe system is a lattice defect that originates from one short-period superlattice (SLs) and then propagates through the Si buffer, multiplying into stacking faults within the next SLs, as seen in Fig. 16.20 (Peng et al. 1998a; Cullis 1996).

Another good example is the InAs–GaAs system. A detail of this system shows the defects generated within to have evolved into conical shapes; Fig. 16.21 (Carlino et al. 1996a,b). The defects are due to an increase of the height to width ratio of a particular island. The misfit dislocations and stacking faults eventually came to accommodate the highly strained area, as seen in Fig. 16.22. The TEM/EDS study has confirmed that the concentration of In within the defective area, and particularly at the defect's origin, is higher than in the defect-free area, which is evidence of the mechanism proposed earlier (Carlino et al. 1996a,b).

CUSUM and Si/SiO₂ Interface Lattice Strain in Ultra Thin Gate Dielectrics

It has been reported that the ultra thin gate dielectric using nitrided SiO₂ can introduce strain to the Si channel immediately below the dielectric (Roy et al. 2001). Different

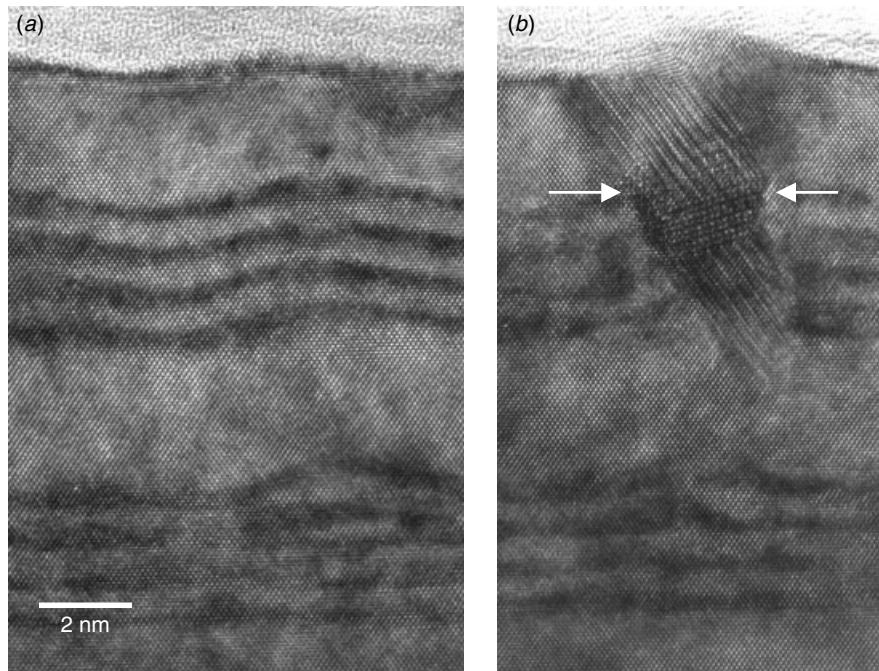


Figure 16.20 Si–Ge SAQDs form in the Si substrate: (a) Defect free area and (b) lattice defects. The defects originating from one of the SL can propagate through the Si buffer and multiply into stacking faults, as indicated, within the next SL.

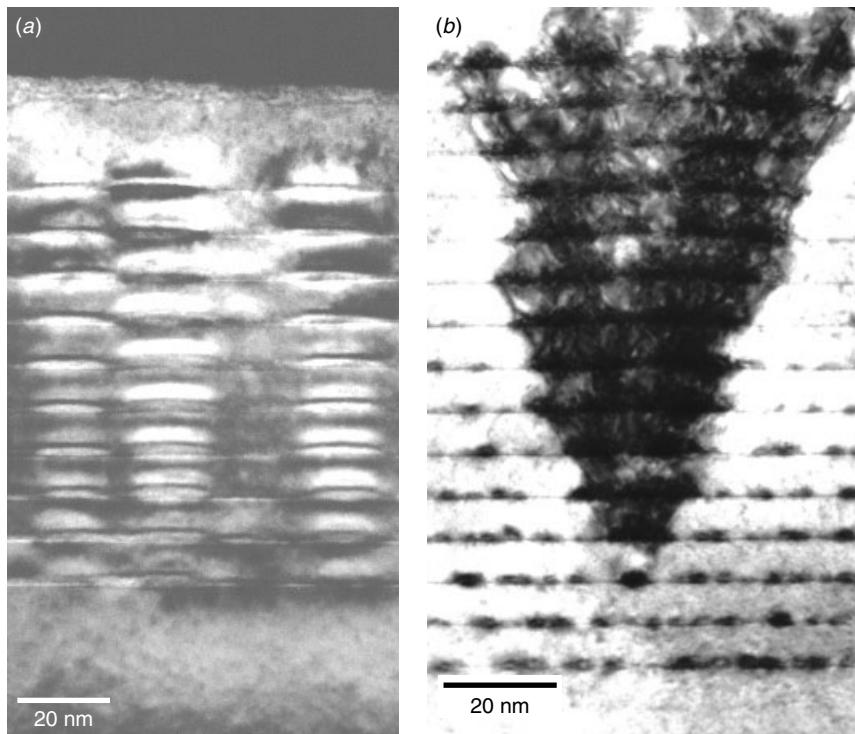


Figure 16.21 InAs–GaAs short-period superlattice (SL): (a) A defect free region and (b) with conical defects. TEM/EDS has confirmed that the In concentration is much higher within the defects.

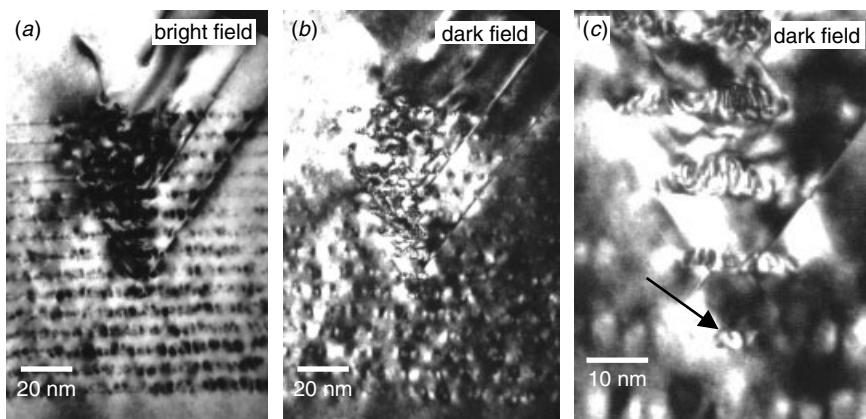


Figure 16.22 (a) InAs–GaAs short-period superlattice (SL) with conical defects (b) Conventional two-beam dark field and bright field showing the defect structure in SAQDs. (c) Defect origin is indicated by arrow. A high concentration of In was identified at the defect's origin.

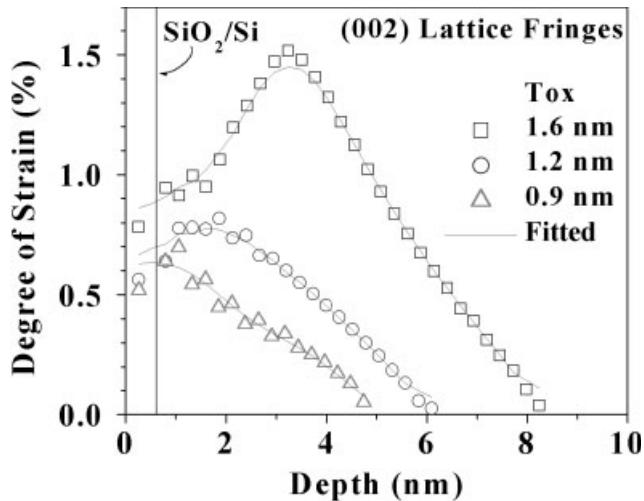


Figure 16.23 SiO_2/Si interface local lattice strain calculated using the HRTEM and CUSUM methods. The three different oxide thicknesses show different lattice strain profiles. Note that the thinner the oxides, the lower the strain value and the shallower the strain propagate into Si substrate.

oxidation and nitridation processes have different impacts on the distribution of strain and, thus on the device's performance (Kim et al. 2002; Ha et al. 1999). As a result an important step in understanding the critical impact on the final device is measuring the strain directly. Indirect measurements of the lattice strain has been reported (Kim et al. 1997), and TEM-related convergent beam techniques have also been used to measure the local strain near the channel region (Toda et al. 2001, 2000). However, none of the measurements were done at the atomic scale.

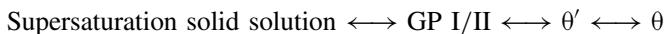
We applied the CUSUM method mentioned above to calculate the local lattice strain near the SiO_2/Si interface with atomic resolution. Figure 16.23 shows an example of this application. First, three different gate oxide thicknesses are compared. Where the oxide thickness is thin, the lattice distortion is lowest, and this is near the SiO_2/Si interface. As the oxide thickness increases, the maximum value increases, and its position progressively shifts to below the SiO_2/Si interface, as shown in the Fig. 16.23. The absolute strain value and strain profile may depend on the details of the gate oxide and nitridation processes. More study is needed to correlate the oxide process to the strain profile and to the final device characteristics.

16.4 OTHER HRTEM APPLICATIONS IN MICROELECTRONICS

The high-resolution TEM has been widely used in Si and non-Si semiconductor characterizations for many years. International conferences dedicated to related topics have been regularly held. The scope of this discussion is beyond what this book can cover. However, we give here a few special examples to show the variety and diversity of HRTEM applications in microelectronics.

GP Zone in Al Metallization

Al–Cu and Al–Si–Cu films are widely used as interconnects in contemporary ULSI process technology. Cu alloying in Al results in the formation of intermetallic Al_2Cu precipitates. The precipitation behavior of the Al_2Cu theta-phase from a supersaturated Al matrix phase in Al–Cu and Al–Si–Cu alloys in bulk materials has been extensively studied (Starink and van Mourik 1991). Precipitation proceeds VIA the following sequence:



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, and GP zones are small segregations formed by the atomic redistribution of a homogeneous solid solution over the crystal lattice sites. GP zones usually are not regarded as new phase precipitations since they do not have well-defined boundaries and their lattice is continuous into the parent phase structure. Experimentally GP zones in bulk

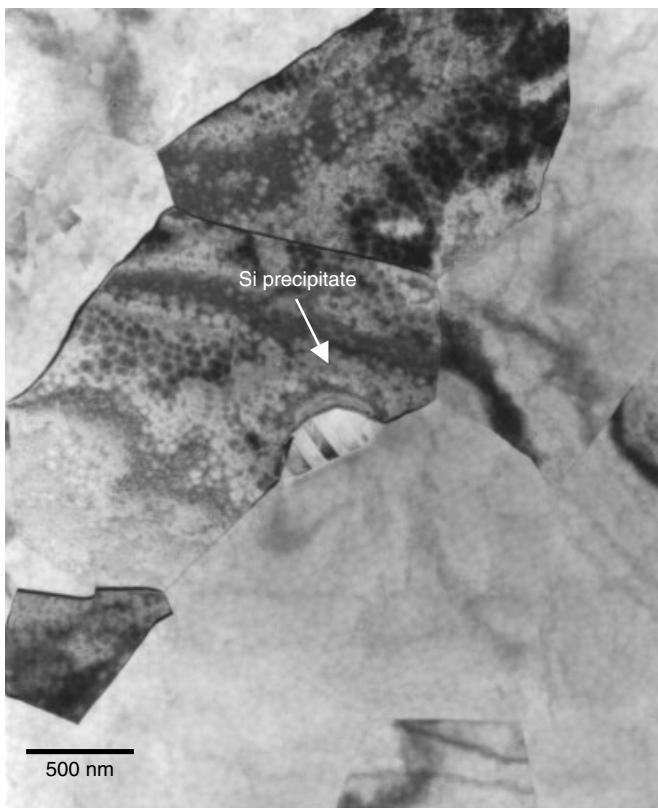


Figure 16.24 Al–1 wt% Si–0.5 wt% Cu thin film grain structure. The Al grain size is about 0.5–2.5 μm . The Si precipitation particles are observed at the Al grain boundary. (Tung et al., *Scripta Materialia*, **34** (9), 1473–1477, 1996, reprint with permission from Elsevier Science Ltd)

Al–Cu alloys are always observed to lie on Al{100} planes. (Lorimer 1978 Cohen 1986).

It is generally thought that GP zones do not exist in either Al–Cu or Al–Si–Cu thin films formed by the normal metallization process of VLSI (Kim and Morris 1992; Park et al. 1994; Colgan and Rodbell 1994). Depending on the deposition and thermal history, most studies show θ' and/or θ phases to have formed within the Al grains or along the grain boundaries, grain edges, and the substrate's interface. The absence of GP zones is attributed to low excess vacancy density in thin films (Frear et al. 1990; Colgan and Rodbell 1994) However, in an early report (Tung et al. 1996) we were able to demonstrate that with careful TEM sample preparation of the as-deposited Al–1 wt% Si–0.5 wt% Cu thin film, GP zones do appear in the as-deposited film.

First, we thermally oxidized silicon wafers of {100} orientation to form a 550 nm oxide, followed by chemical vapor deposition of a 600 nm borophosphosilicate glass (BPSG) layer at 720°C. Subsequently we deposited a 300 nm thick Al alloy film in a Varian DC magnetron sputtering deposition machine from a target with the nominal composition of Al–1 wt% Si–0.5 wt% Cu in an argon atmosphere with a deposition rate of 21 nm/s. Deposition was carried out at room temperature, but the actual wafer temperature was approximately 100°C due to the sputtering heat. No subsequent heat

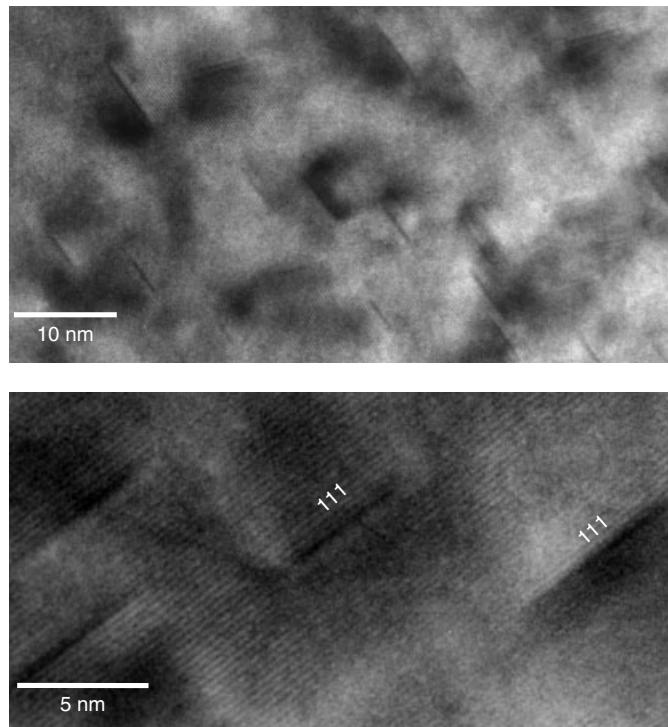


Figure 16.25 HRTEM of Al–1 wt% Si–0.5 wt% Cu thin film grain showing that the GP zones align along the Al{111} habit planes, instead of the normal {100}, as observed in the bulk materials. (Tung et al., *Scripta Materialia*, **34** (9), 1473–1477, 1996, reprint with permission from Elsevier Science Ltd).

treatment was carried out. The TEM samples were prepared by the sample preparation procedures discussed in Chapter 4. The highest temperature that the TEM samples ever experienced during the sample preparation was about 150°C for no more than 3 minutes when the sample were placed on a hot plate. Final thinning was performed by ion milling from the substrate's side. The milling time was controlled to avoid artifacts.

The as-deposited films were polycrystalline with grain sizes ranging approximately from 0.5 to 2.5 μm . Large Si precipitates were observed at the Al grain boundaries, as shown in Fig. 16.24. The HRTEM images of an Al grain's interior, Fig. 16.25, show discrete needle-like strain contrasts. The lattice image shows these to be Guinier-Preston zones along the Al{111} lattice planes, as seen in Fig. 16.25. The HRTEM image was viewed along the Al{110} direction and Al{111} lattice fringes with the lattice spacing $d_{111} = 0.234 \text{ nm}$ clearly resolved. High-density dark plates aligned along the Al{111} planes, which are observed to be about 2 to 4 Al{111} lattice plane thickness and about 5 to 10 nm long. The plates with dark contrast structurally cohere to the Al parent lattice without a discernible boundary within the present resolution limit. A selective area diffraction pattern, Fig. 16.26, shows streaks appearing along the $\langle 111 \rangle$ directions, and this is characteristic of disk-shaped precipitates lying on the Al{111} planes. The characteristics of the GP zones are clearly demonstrated alongside the associated diffraction pattern. Attempts to identify the chemical nature of the precipitates by high-resolution energy loss spectroscopy (EELS) failed because of the

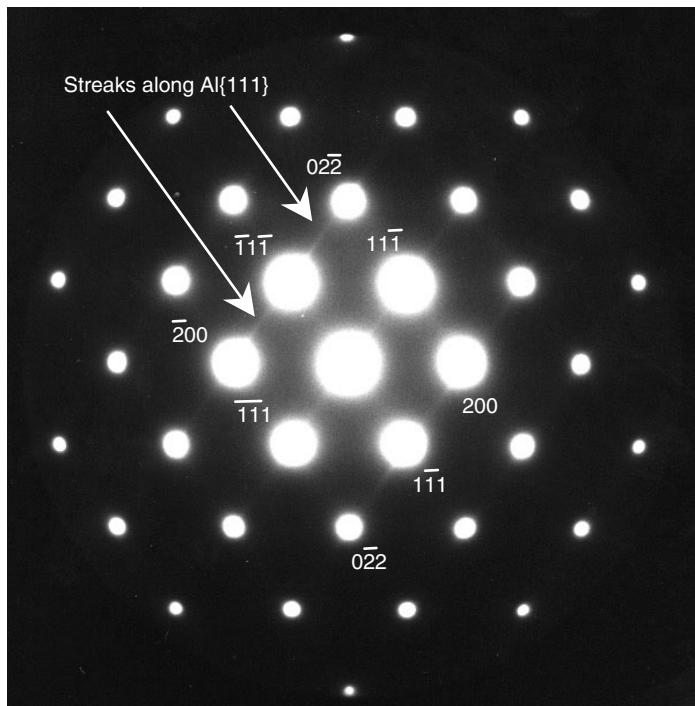


Figure 16.26 Al—1 wt% Si—0.5 wt% Cu thin film electron diffraction pattern shows streaking along $\langle 111 \rangle$, as indicated. (Tung et al., *Scripta Materialia*, **34** (9), 1473–1477, 1996, reprint with permission from Elsevier Science Ltd).

resolution limits in the instrumentation available at the time of analysis. It is uncertain whether the observed GP zones are due to Si or Cu precipitation. If they are Si plates, this is the first time that GP zones are observed to result from Si precipitation in thin film Al–Si alloy. The possibility of Si generating the GP zones cannot be ruled out as semi-coherent plates are frequently observed on Al{111} planes in quenched and aged Al–Si alloys (Westmacott and Dahmen 1985). However, it is unlikely that the GP zones formed in this sample are due to Si, because large Si nodules were already seen at the grain boundaries in Fig. 16.24.

The {111} orientation of GP zones may be due to Cu precipitation. For the Al–Cu bulk alloy, the GP zone plane is always on Al{100} (Cohen 1986). {111} is not the preferred orientation. The shape and orientation of a coherent inclusion in an isotropical crystal can be determined from the analysis of its strain energy (Mura 1982; Khachaturyan 1983). To take an example, consider the a single coherent cubic phase inclusion in an infinite cubic parent phase matrix, which is the case with GP zones in the Al matrix. Under the assumption that first, the dimension of the inclusion and the distance to the nearest inclusion is small compared with the dimension of the crystal (infinite isotropic matrix approximation). And second, external boundaries of the matrix and inclusions

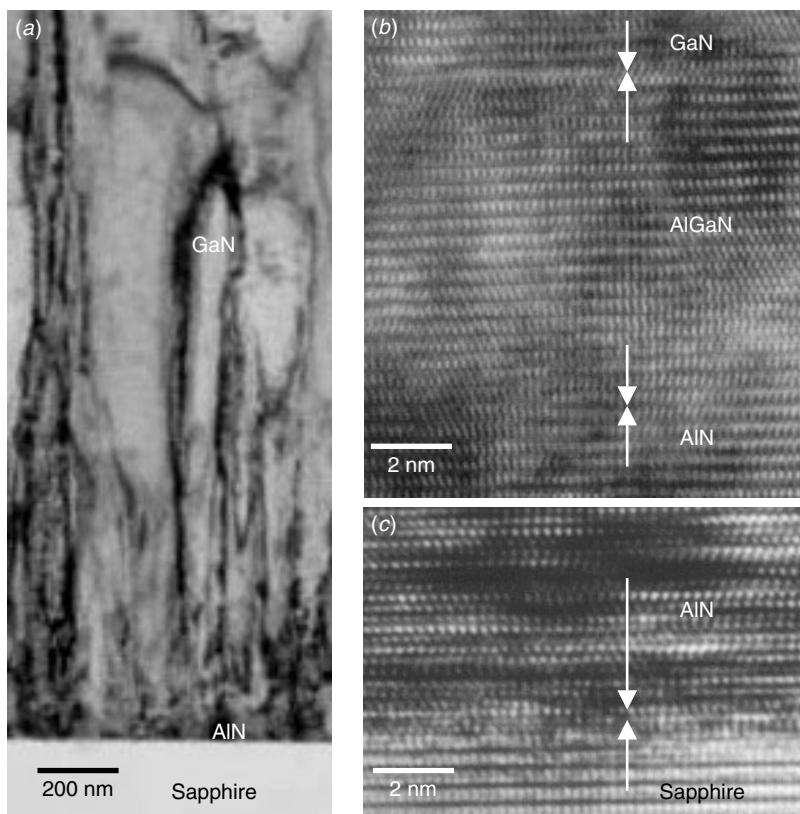


Figure 16.27 (a) TEM cross section of GaN thin film, about 2 μm thick, grown on a sapphire substrate. (b,c) HRTEM close-up at the substrate interface showing the AlN and AlGaN buffer layers.

mixtures are stress free. It has been shown that the habit plane of the inclusion is determined by the elastic anisotropic parameters defined as

$$\xi = \frac{C_{11} - C_{12} - 2C_{44}}{C_{44}}$$

where C_{ij} are the elastic constants of the matrix. For $\xi < 0$ the habit plane will be the {100} type and for $\xi > 0$ the habit plane will be the {111} type. For the GP zone in the Al–Cu alloy, the three nonzero elastic constants of Al are

$$C_{11} = 10.82 \times 10^{10} \text{ Pa},$$

$$C_{12} = 6.13 \times 10^{10} \text{ Pa},$$

$$C_{44} = 2.85 \times 10^{10} \text{ Pa}$$

Therefore $\xi = -0.365$, and the GP zones should have the {100} orientation, which is in agreement with all of the experimental observations made in the bulk materials. The situation is different, however, in the case of thin film because the theoretical assumptions may not be satisfied. Of the assumptions mentioned above, the last is probably the assumption most seriously violated because the thin metal film on the Si substrate is known to be under external stresses due to its thermal incompatibility.

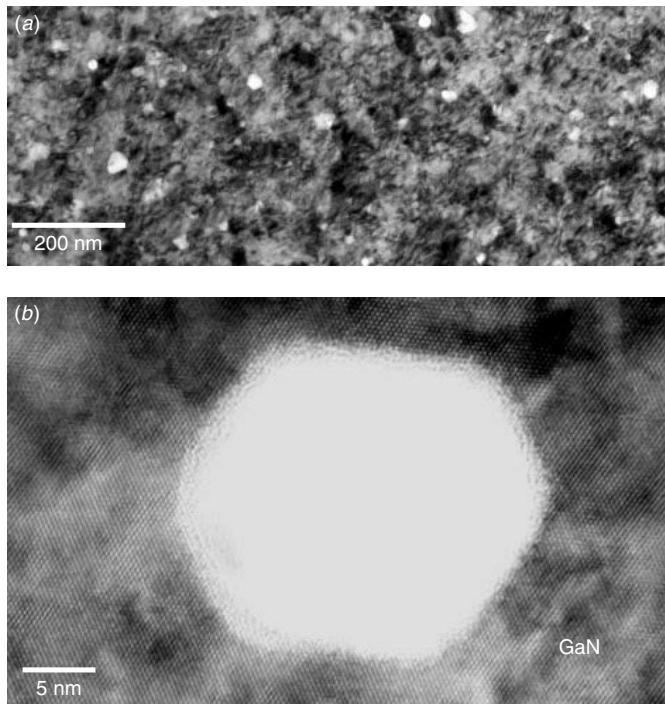


Figure 16.28 TEM planar section of nanopipe defects. (a) Low magnification shows the defect's (white dots) density and distribution. (b) The HRTEM image shows the hexagonal symmetry of the defect. The defect sidewalls follow approximately along the hexagonal [1100] lattice planes.

The validity of the second assumption is certainly less adequate for thin films than for bulk materials. For a thin film of 300 nm thickness and with GP zones about 5 to 10 nm in diameter, the infinite matrix assumption is questionable. Although further development of current theory to include stressed thin film is necessary to explain the observed {111} orientation, Cu GP zones seem to be capable of forming along Al{111} as has been shown in this study.

GaN, Nanopipes, and Nanopillars

Gallium nitride, GaN, and its related alloys (AlGaN and InGaN) are important wide band-gap semiconductors that have potential applications in both short wavelength optoelectronics and high-power high-frequency devices. These materials are excellent candidates for semiconductor UV/blue optoelectronics devices. The growth of GaN on Al_2O_3 (sapphire) or the Si(111) substrate has been realized by introducing AlN and AlGaN buffer layers, as shown in Fig. 16.27. As-grown GaN film is known to contain a high density of defects, in particular, threading dislocations (Qian et al. 1995; Romanov et al. 2003). These defects affect both electrical and optical properties of the material. It has been proved that the quality of GaN films improves significantly with optimized AlN and AlGaN buffer layers, and that the morphology of the GaN film is mostly governed by the specific growth parameters. Several techniques have been used

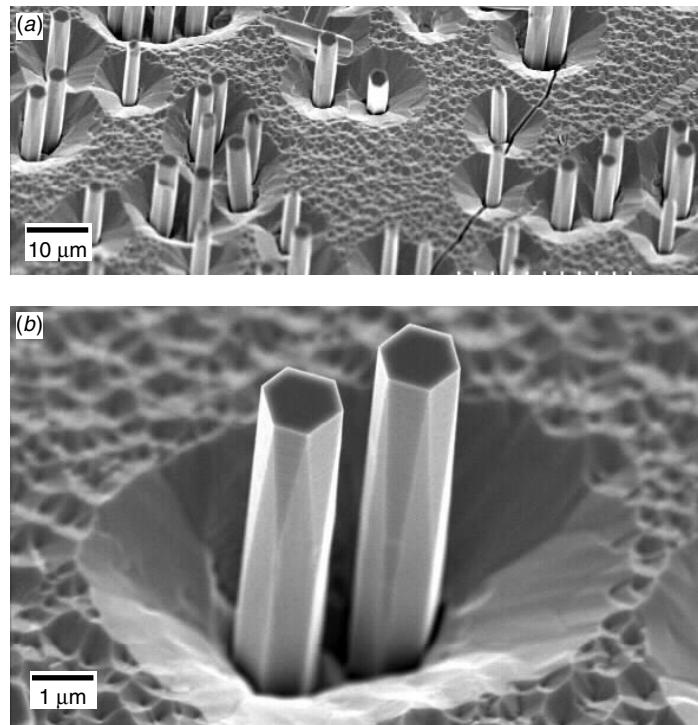


Figure 16.29 Nanopillar defects observed on thick GaN film. The SEM images show that the pillars grow randomly on the film's surface, as seen in (a). Each pillar has a surface crater associated with it, as seen in (b). Note that the nanopillars are hexagonal in most instances.

to grow high-quality GaN films, for example, metalorganic chemical vapor deposition (MOCVD), organometallic vapor phase epitaxy (OMVPE), and molecular beam epitaxy (MBE). Tremendous effort has been devoted to GaN crystal quality optimization (e.g., Gaskill et al. 1986; Sun et al. 1994; Feltin et al. 2003). Among many other parameters, the Ga to N ratio plays an important role in determining the film's microstructure. For example, columnar grains and whiskerlike microcrystals can be obtained in an N-rich growth environment (Sanchez-Garcia 1998).

Most of the dislocations observed in GaN are of the pure edge type, forming a low-angle twist boundary, which is the shadow observed in Fig. 16.27. Another interesting observed defect that has been reported is called a nanopipe (Qian et al. 1995). Figure 16.28 shows an example of a nanopipe defect observed in our sample analysis. The nanopipes are believed to be open-core screw dislocations. The nanopipes occur at the core of screw dislocations where dislocations intersect the surface, form spiral steps, and create hexagonal growth hillocks; the eventual effect is a nonplanar surface morphology. The defects are about 5 to 50 nm in diameter, and their diameters get progressively larger as the film grows thicker. No details have been reported as to how these defects evolve with the increases of film thickness.

As the GaN film increases in thickness, another type of film defect is observed. Figure 16.29 shows the defect in SEM. Pillars are observed to grow out of the GaN

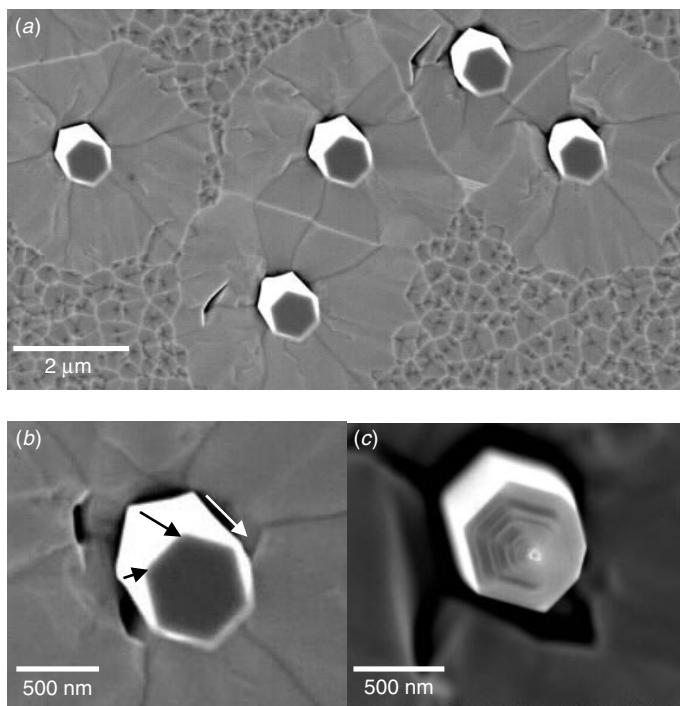


Figure 16.30 SEM plan views: (a) Each nanopillar has a hexagonal base associated with it. (b) A close-up shows the pillar's twisted growth morphology. The pillar's top surface is a perfect hexagon, but there is a twist between the pillar's tip and bottom. Notice also that all pillars twist in the same direction. (c) Some pillars show a stepped and sharp tip surface, suggesting that a screw dislocation could be the pillar's growth mechanism.

matrix. These submicron microscopic pillars (called nanopillars) grow randomly out of the GaN film surface, and each nanopillar has a surface dent (or crater) associated with it. A closer look at the pillars reveals the following details:

- Most of the nanopillars have hexagonal diameters and faceted sidewalls.
- The dent (crater) associated with each nanopillar also has hexagonal symmetry, as seen in Fig. 16.30(a).
- Some of the nanopillars show spiral growth, as seen in Figs. 16.30(a) and (b).
- Some of the nanopillars show a stepped and sharpened pillar tip morphology, as seen in Fig. 16.30(c).

Both spiral twist and sharp-tipped types of nanopillars suggest that the screw dislocation could be the nanopillar growth mechanism.

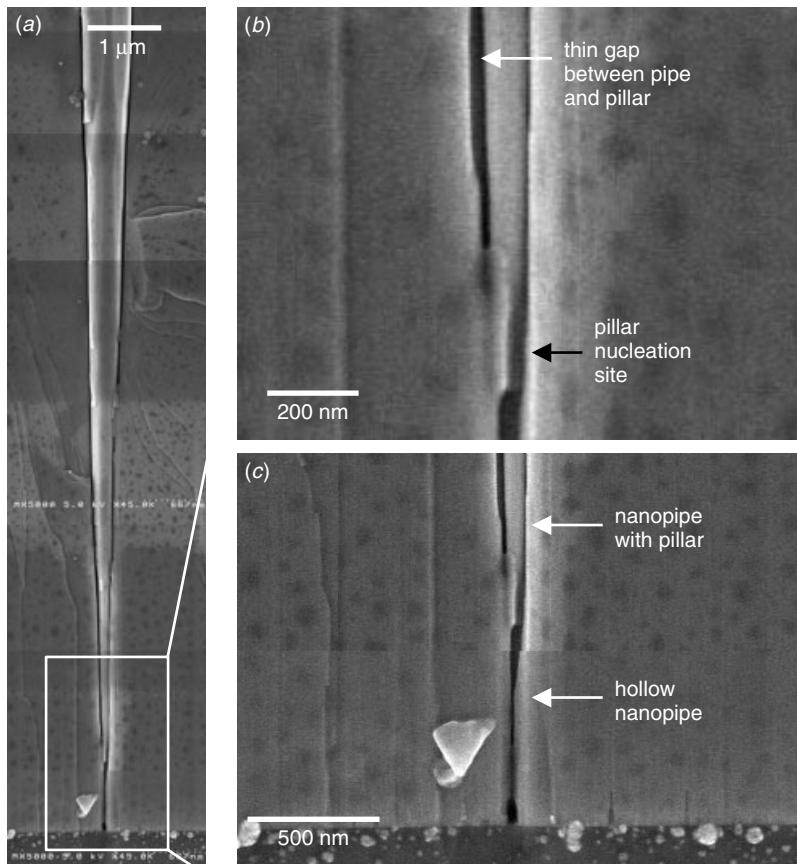


Figure 16.31 SEM cross section of the GaN film shows that the nanopillar nucleates and grows within the GaN film's matrix. (a) Low magnification shows the pillar diameters to increase along with the film. (b,c) A close-up at the pillar's base shows the pillar to have nucleated from the nanopipe's sidewalls and grow along inside the nanopipe.

A cross-sectional SEM image obtained by a precision cleavage that cut across the nanopillar is shown in Fig. 16.31. A few details are to be noted:

- The nanopillar diameter is small at the film's bottom and progressively has become larger with the increases of film thickness.
- Nanopillars nucleate from the nanopipe's sidewall at around 0.7 μm from the film's bottom, as seen in Fig. 16.31(c), and grow along the nanopipe.
- A thin gap is always found between the nanopipe and its nanopillar.

The TEM cross section, shown in Fig. 16.32, does not reveal much more about the thin gap because of the sample's thickness and the overlapping effects. A more useful image is obtained from the planar TEM section, seen in Figs. 16.33 and 16.34. If we disregard the pillar's density and shape, we see that each nanopillar has at least one bridge to connect itself and the matrix, and more often multiple bridges. A few other important observations can be made from Figs. 16.33 and 16.34.

- The distribution, sizes, and shapes of the nanopipes and nanopillars are random.
- The matrix surrounding each pillar contains substantial crystal defects. Defects segregate into low-angle boundaries that divide the substrate into approximately hexagonal symmetrical shapes, as seen in Fig. 16.34.

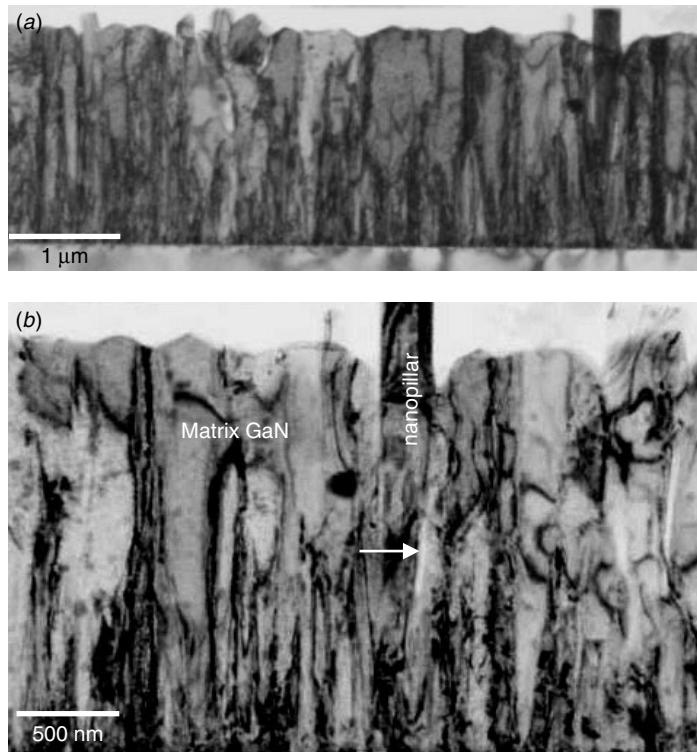


Figure 16.32 TEM cross section of the GaN film with nanopillars revealing thin gap in between matrix and nanopillar, as indicated.

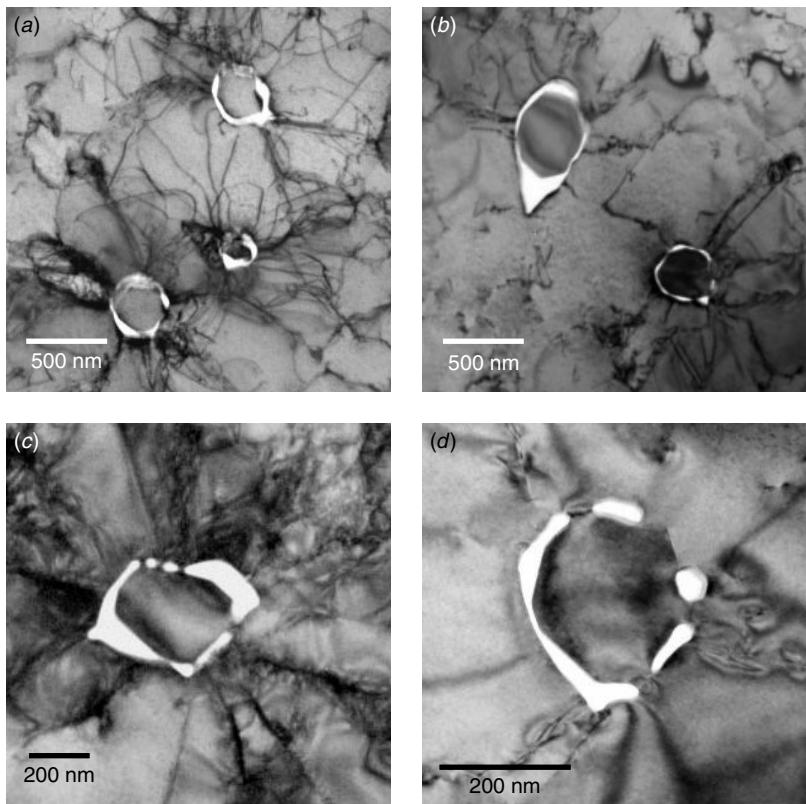


Figure 16.33 Planar section of nanopillars showing the nanopillars to be connected to the matrix GaN by single or multiple bridges. These bridging points are in parallel with the nanopillars throughout the film's thickness. Electron diffraction studies reveal the nanopillars to be in exactly the same crystal orientation with its surrounding matrix GaN. A high density of the crystal defects can be observed to aggregate around the nanopipe and nanopillar.

Figure 16.35 shows a close-up of a typical nanopillar. Two bridging points are revealed in this case. The electron diffraction shows the nanopillar and its matrix to be in the exactly the same crystal orientation. Through HRTEM analysis, Fig. 16.36, their lattice planes are seen to be continuous across the bridges, and the pillar and its matrix in an epitaxial relationship, although planar crystal defects are always found in the bridges between the nanopillars and their neighboring matrix.

It is possible that the nanopillar defects reported in this section are of a metamorphic defect type that has evolved from the nanopipes reported previously. This is on the following observations:

- Nanopillars only nucleate and grow within existing nanopipes, as seen in Fig. 16.31.
- In terms of defect crystallography, nanopipes were reported to be a metamorphosis of open-core screw dislocations. In our observation, nanopillars growing within nanopipes are also associated with screw dislocations, as seen in Fig. 16.30.

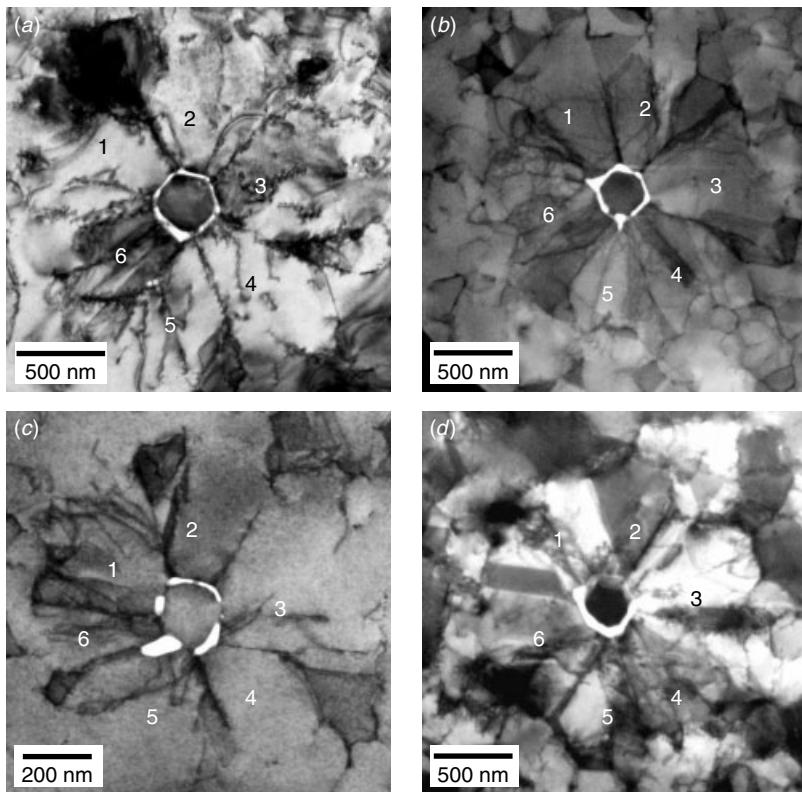


Figure 16.34 Lower magnification TEM plan view showing the defect distribution in the matrix crystal around the nanopipe and nanopillar to be in hexagonal symmetry, as marked in all figures, although some may not be perfect hexagons.

Similar nanowhiskers have been observed and reported (Wagner and Dorherty 1996; Yumoto et al. 1999). The growth of the whiskers was reported to be through a vapor–liquid–solid crystal growth mechanism. A seed impurity existing at a nucleation site can trigger a liquid–solid equilibrium and phase transition at the liquid–solid interface. In the present case the impurity existing in the nanopipes can trigger a similar growth mechanism, but the supply of reactants is limited by the nanopipes. Nanopillar growth in such a circumstance can only follow the nanopipes. However, there remain good questions to be purposed in future studies:

- What exactly is the role of a single or multiple bridges between the nanopillars and their associated matrix?
- How do the nanopillars nucleate within nanopipes?
- Why do the nanopillars not fill the gap created with the matrix? What stops the coalescence from happening?
- Why are there always hexagonal symmetric crystal defects in the matrix surrounding the nanopillar defects. What are the growth conditions and the stress/strain relationships that lead to this defect morphology?

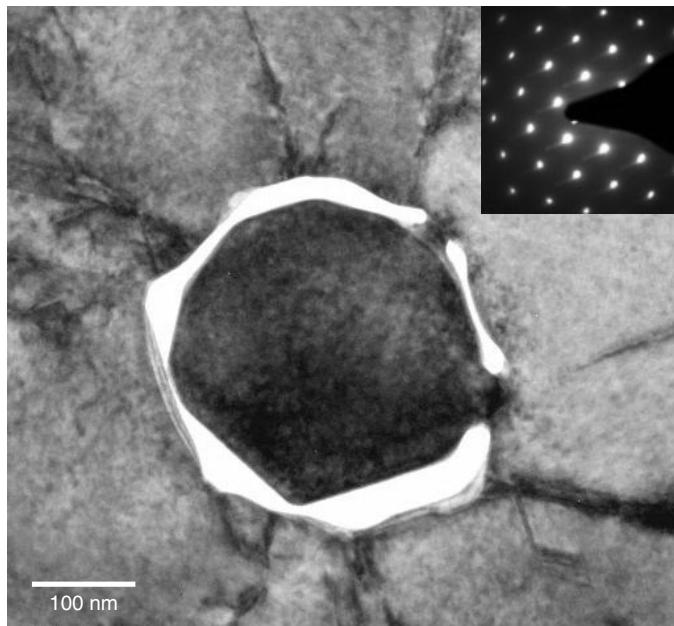


Figure 16.35 TEM plan view close-up of one single nanopillar showing the nanopipe, the nanopillar, and the two bridges across the thin gap in between. The electron diffraction (insert) shows that the matrix and pillar are in the same crystal orientation, (0001) basal plane.

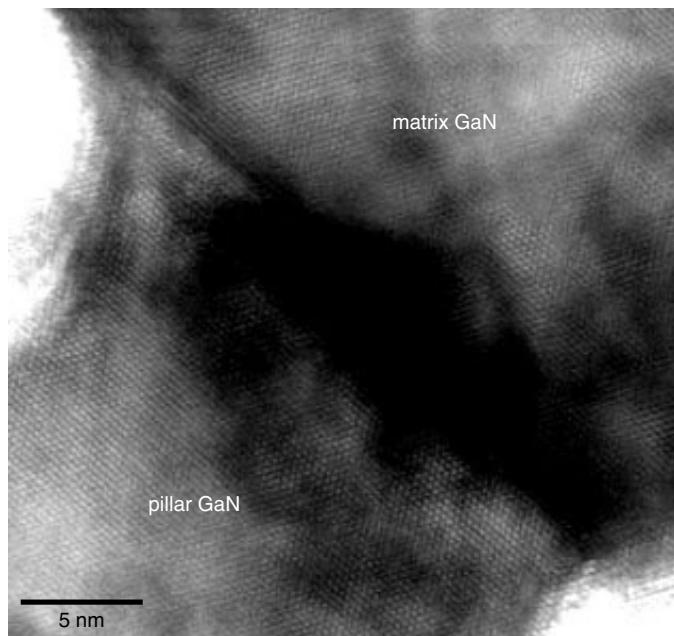


Figure 16.36 HRTEM image at the bridge between the matrix and pillar GaN showing the two to be in epitaxy. All lattice planes are continuous, although interface planar defects are found.

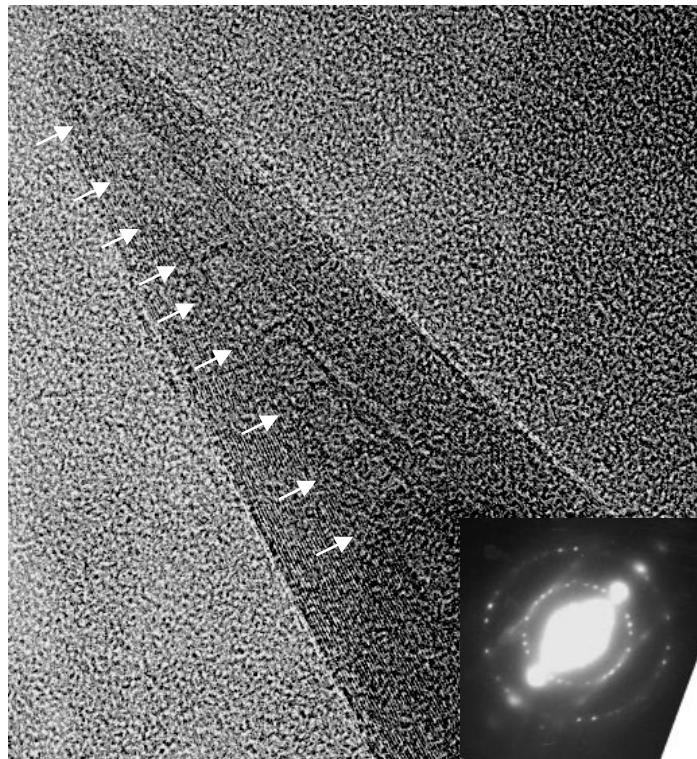


Figure 16.37 HRTEM image at the tip of a multiple-wall carbon nanotube. Layers of carbon walls can be observed clearly, as indicated. The corresponding electron diffraction is showing the symmetric structure of the carbon walls.

Many more such issues can be addressed in order to understand the defect formation mechanisms and thus eventually to avoid them during crystal growth process.

Carbon Nanotubes

For the first time in its history Si-based ULSI process technology is nearing a terminus in its development horizon. The process challenges are becoming exceedingly more difficult at both the front end of line (FEOL) and back end of line (BEOL). The development costs for each subsequent technology generation are becoming uneconomical for most of the wafer FABs. The International SEMATECH ITRS technology roadmap has forecasted the potential difficulties. Revolutionary changes appear to be needed for the conventional MOSFET structure in the next decade. By modifying the conventional device structures such as the multiple gate transistors, we may be able to push the development of the Si MOSFET to a few more generations (Geppert 2002). Yet many advanced research areas in biomaterials, semiconductor polymers, quantum computing using quantum mechanics, and so fourth that did not have much to do with conventional Si technology are now inverting ways to displace the Si-based device. One promising candidate is a carbon-based device that uses carbon nanotubes as the channel material. A single wall carbon nanotube field effect transistor (SWNT-FET)

has already been demonstrated in a conventional MOSFET structure and shown to be competitive in device characteristics (Wind et al. 2002).

Carbon nanotubes are macromolecules of carbon rolled into a cylinder. The ends of the nanotubes are fitted with hemispherical caps. Figure 16.37 shows the hemispheric tip of a multiple-wall carbon nanotube in an HRTEM image. Due to their chemical bonding, carbon nanotubes are light, strong, thermally stable, and inert to most chemicals. Basically there are single-wall nanotubes (SWNTs) and multi-wall nanotubes (MWNTs). Their optical and electrical properties are versatile, making them a promising material for almost every electronic and photonic application. One good way to use the multi-wall nanotubes is as a high-coherency and high-brightness electron beam source (Jonge et al. 2002). The study of carbon nanotubes and their applications in microelectronics technology are only in its inception. New applications and new process technologies are being reported constantly.

For microelectronics, it is hard, if not impossible, to predict what materials, technology, and innovative designs will eventually replace the conventional Si technology and predominate in microelectronics.

REFERENCES

- P. D. Augustus, P. Kightley, R. R. Bradley, and R. J. M. Griffiths, *Proc. NATO Advanced Research Workshop on the Evaluation of Advanced Semiconductor Materials by Electron Microscopy*, 335–347, 1988.
- A.-L. Barabasi, *Appl. Phys. Lett.*, **70** (19), 2565–2567, 1997.
- R Beanland, *Microscopy on Semiconductor Materials Conf.*, Institute of Physics Oxford, March 1999, 451–454.
- R. Bierwolf, M. Hohenstein, F. Phillip, O. Brandt, G. E. Crook, and K. Ploog, *Ultramicroscopy* **49**, 273–285, 1993.
- E. Carlino and L. Tapfer, *Appl. Phys. Lett.*, **69** (17), 2546–2548, 1996a.
- E. Carlino, C. Giannini, C. Gerardi, L. Tapfer, K. A. Mader, and H. von Kanel, *J. Appl. Phys.*, 1441–1447, 1996b.
- H. S. Chang, S. Choi, H. Yang, K. Y. Min, D. W. Moon, H. I. Lee, and H. Hwang, *Appl. Phys. Lett.*, **80** (3), 386–388, 2002.
- J. B. Cohen, *Solid State Physics: Advanced Research and Applications*, edited by H. Ehrenreich and D. Turnbull, Academic Press, 131, 1986.
- E. G. Colgan and K. P. Rodbell, *J. Appl. Phys.*, **75**, 3423, 1994.
- A. G. Cullis, *MRS Bulletin*, (April) 21–26, 1996.
- E. Feltin, B. Beaumout, P. Vennegues, M. Vaille, P. Gibart, T. Riemann, J. Christen, L. Dobos, and B. Pecz, *J. Appl. Phys.*, 182–185, 2003.
- D. R. Frear, J. E. Sanchez, A. D. Romig Jr., and J. W. Morris Jr., *Metall. Trans.*, **A21**, 2449, 1990.
- D. K. Gaskill, N. Bottka, and M. C. Lin, *Appl. Phys. Lett.*, **48** (21), 1449–1451, 1986.
- L. Geppert, *IEEE Spectrum*, (Oct.), 28–33, 2002.
- M. Grundmann, O. Stier, and D. Bimberg, *Phys. Rev.* **B52** (16), 11969–11981, 1995.
- Y. H. Ha, S. Kim, S. Y. Lee, J. H. Kim, and D. H. Baek, H. K. Kim, D. W. Moon, *Appl. Phys. Lett.*, **74** (23), 3510–3512, 1999.
- N. D. Jonge, Y. Lamy, K. Schoots, and T. H. Oosterkamp, *Nature*, **420**, 393–395, 2002.

- P. H. Jouneau, A. Tardot, G. Feuillet, H. Mariette, and J. Cibert, *J. Appl. Phys.*, **75** (11), 7310–7316, 1994.
- T. Kaga, M. Ohkura, F. Murai, N. Yokoyama, and E. Takeda, *J. Vac. Sci. Technol. B*, **13**, 2329, 1995.
- A. G. Khachaturyan, *Theory of Structural Transformations in Solids*, Wiley, 1983.
- R. Kilaas, S. Paciornik, and J. Bonevich, *User Manual of “The NCEM Image Processing Custom Function Plug-Ins for digital Micrograph,”* National Center for Electron Microscopy, USA, 1998.
- C. Kim and J. W. Morris Jr., *J. Appl. Phys.*, **72**, 1837, 1992.
- Y. P. Kim and S. K. Choi, *Appl. Phys. Lett.*, **71** (24), 3504–3506, 1997.
- W. S. Lau, M. T. Chandima Perera, P. Babu, A. K. Ow, T. Hang, N. P. Sandler, C. H. Tung, T. T. Sheng, and P. K. Chu, *Jap. J. Appl. Phys.*, **37**, L435–L437, 1998.
- B. Legrand, B. Grandidier, J. P. Nys, D. Stievenard, J. M. Gerard, and V. Thierry-Mieg, *Appl. Phys. Lett.*, **73** (1), 96–98, 1998.
- G. W. Lorimer, *Precipitation Processes in Solids*, edited by K. C. Russell and H. I. Aaronson, Metallurgical Society of AIME, 87, 1978.
- D. Muller, *Technical notes of Int. SEMATECH Analytical Lab Manager’s Council Meeting*, AMD Sunnyvale, 24–25 Jan., 2002.
- T. Mura, *Micromechanics of Defects in Solids*, Martinus Nijhoff, 1982.
- T. H. Ning, CMOS Performance and Density at Semi-log Trends Forever, and With SiGe, Who Needs GaAs, IMRE Workshops, Singapore, Oct. 1998.
- M. Park, S. J. Krause, and S. R. Wilson, *Appl. Phys. Lett.*, **64**, 721, 1994.
- C. S. Peng, Q. Huang, Y. H. Zhang, W. Q. Cheng, T. T. Sheng, C. H. Tung, and J. M. Zhou, *Physical Review B*, **57** (15), 8805–8808, 1998c.
- C. S. Peng, Q. Huang, Y. H. Zhang, W. Q. Cheng, T. T. Sheng, C. H. Tung, and J. M. Zhou, *Phys. Rev. Lett.*, 1999.
- C. S. Peng, Z. Y. Zhao, H. Chen, J. H. Li, Y. K. Li, L. W. Guo, D. Y. Dai, Q. Huang, J. Z. Zhou, Y. H. Zhang, T. T. Sheng, and C. H. Tung, *Appl. Phys. Lett.*, **72** (24), 3160–3162, 1998b.
- C. S. Peng, Q. Huang, W. Q. Cheng, J. Z. Zhou, Y. H. Zhang, T. T. Sheng, and C. H. Tung *App. Phys. Lett.*, **72** (20), 2541–2543, 1998a.
- S. J. Pennycook, *Technical notes of Int. SEMATECH Analytical Lab Manager’s Council Meeting*, AMD Sunnyvale, 24–25 Jan., 2002.
- E. L. Principe, A. Hegedus, C. Kisielowski, C. Song, B. Freitag, D. Hubert, T. Fliervoet, J. Gibson, J. Moulder, and D. Watson, *Technical notes of Int. SEMATECH Analytical Lab Manager’s Council Meeting*, AMD Sunnyvale, 24–25 Jan., 2002.
- W. Qian, M. Skowronski, M. D. Graef, K. Doverspike, L. B. Rowland, and D. K. Gaskill, *Appl. Phys. Lett.*, **66** (10), 1252–1254, 1995.
- M. D. Robertson, J. M. Corbett, J. E. Currie, and J. B. Webb, *Proc. Microscopy of Semiconductor Materials Conf. Institute of Physics*, 48–52, 1995a.
- M. D. Robertson, J. M. Corbett, J. E. Currie, and J. B. Webb, *Ultramicroscopy*, **58**, 175–184, 1995b.
- A. E. Romanov, P. Fini, J. S. Speck, *J. Appl. Phys.*, **93** (1), 106–114, 2003.
- R. K. Roy, Y. N. Chen, and S. Chetlur, *IEEE Trans. Electron Devices*, **48** (9), 2016–2021, 2001.
- M. A. Sanchez-Garcia, F. J. Sanchez, F. B. Naranjo, F. Calle, E. Calleja, E. Munoz, U. Jahn, and K. H. Ploog, *MRS Internet J. Nitride Semicond. Res.*, **3** (32), 1–8, 1998.
- G. S. Solomon, J. A. Trezza, A. F. Marshall, and Harris Jr., J. S. *Physi. Rev. Lett.*, **76** (6), 952–955, 1996.
- M. J. Starink and P. van Mourik, *Metall. Trans. A*, **22**, 665, 1991.

- C. S. Sun, P. Kung, A. Saxier, H. Ohsato, E. Bigan, M. Razeghi, and D. K. Gaskill, *J. Appl. Phys.*, **76** (1), 236–241, 1994.
- J. Tersoff, C. Teichert, and M. G. Lagally, *Physical Review Letters* 76 (10), 1675–1678, Q. Xie, A. Madhukar, P. Chen, and N. P. Kobayashi, 1995 *Phys. Rev. Lett.* **75** (13), 2542–2545, 1996.
- A. Toda, N. Ikarashi, and H. Ono, *J. Crys. Growth* **210**, 341–345, 2000.
- A. Toda, N. Ikarashi, H. Ono, S. Ito, T. Toda, and K. Imai, *Appl. Phys. Lett.*, **79** (25), 4243–4245, 2001.
- C.-H. Tung, R.-L. Chiu, and P.-H. Chang, *Scripta Materialia*, **34** (9), 1473–1477, 1996.
- C.-H. Tung, G. Muralidharan, T. T. Sheng, Y. H. Zhang, J. L.-F. Wang, and M. F. Li, *Proc. 4th Nat. Symp. Progress in Materials Research in Singapore*, 165–168, 1998.
- C.-H. Tung, G. Muralidharan, G. T. T. Sheng, Y. H. Zhang, J. L. F. Wang, W. J. Fan, C. H. Wang, and J. Jiang, *Microscopy Semiconductor Materials Conf.*, Institute of Physics, 47–50, 1999.
- R. S. Wagner and C. J. Dorherty, *J. Electrochem. Soc.*, **113** (12), 1300–1305, 1996.
- K. H. Westmacott and U. Dahmen, *Proc. Int. Symp. Behaviour of Lattice Imperfections in Materials—In Situ Experiments with High Voltage Electron Microscopy*, 137, 1985.
- S. J. Wind, J. Appenzeller, R. Martel, V. Derycke, and Ph. Avouris, *Appl. Phys. Lett.*, **80** (20), 3817–3819, 2002.
- H. Yumoto, T. Sako, Y. Gotoh, K. Nishiyama, and T. Kaneko, *J. Crys. Growth*, **203**, 136–140, 1999.

Index

- Accelerometer sensor, in MEMS device, 540, 541–542
- Acronyms, for analytical techniques, 8
- Active area junction, with salicide, 152–154, 155, 156, 157, 163
- Adhesion layer, 308, 311
- Airbag sensors, 541–542
- Al₂Cu particles, 295, 297–300, 301
- Alpha particles, in Rutherford backscattered spectrometry, 31
- Aluminum (Al), device scaling and, 56, 59. *See also* Au-Al ball bonding system; Cu/Ni(V)/Al system; Ni(P)/Al system
- Aluminum alloy plugs (Al-plugs), 308
- Aluminum alloys, 295–307
- Aluminum bond pad deposition, 51–52
- Aluminum extrusion, into multilevel interconnects, 321
- Aluminum metallization
- copper integration into, 340, 341
 - corrosion in, 305, 306, 307
 - GP zones in, 631–636
- Aluminum metal step coverage, 356
- Aluminum pad surface oxides, 603–604, 605
- Aluminum processes, hot, 302–306
- Aluminum spiking, 295, 296, 297, 299
- Amorphous Si, 42–43, 288
- CoSi₂ salicide and, 276, 277, 278
 - TiSi₂ salicide and, 263–264, 265
- Analytical techniques. *See also* Construction analysis (CA); Damage analysis; Failure analysis; Package construction analysis; TEM analysis; TEM cross-sectional analysis; TEM failure analysis
- acronyms for, 8
 - chemical, 24
 - for crystal defects, 25
 - electrical, 25
 - for gate oxide breakdown, 203–205
 - for junctions, 143–145
 - in microelectronics, 5–8, 9–10
 - in microscopy, 5–8
- for optical defects, 21–22
- for products, 62
- for topography, 18–20
- Anisotropic spacer etch, 44, 45
- Annealing, 189
- in resistor-load SRAM, 450
- Annular dark-field (ADF), 18, 24, 28
- Antireflective coating (ARC), 302–303
- electromigration-induced damage to, 502
 - multilevel interconnects and, 322–325
- Application-specific IC (ASIC), 63
- Argon (Ar) ion milling, artifacts from, 128, 129, 131–134. *See also* Ion milling
- Arsenic (As). *See also* GaAs entries
- polysilicon film with, 288–289, 293
 - TiSi₂ salicide and, 267
- Artifacts, sample-preparation, 125–138
- As-implanted condition, primary defects with, 145–147, 148
- Atomic resolution, of scanning tunneling microscopy, 33–34.
- Au–Al ball bonding system, 559–563, 564
- Auger electron spectroscopy (AES), 30–31
- in failure analysis, 478
 - ion implantation and, 144
- Au/Pd/Ni coatings, on copper leadframe, 600
- Back-end defects, 144
- Back end-of-line (BEOL) interconnection, 55–59, 75, 220
- Back-side chipping, 70
- Back-side lapping, 100
- Back-side thinning, 101
- Ball bonding, 65–67, 70
- Au–Al, 559–563, 564
- Bardeen, John, 3
- Barrier layers, tantalum, 329–330, 331, 332, 335, 336, 337, 338, 339–340
- Barrier materials, 328–329
- Beam structures, in MEMS devices, 540–542, 543, 544, 545, 546
- BiCOMS process, defects via, 173, 175, 176

- Bicrystal defects, 176–178
- Bimodal grain size distribution, in copper film, 330–331, 333
- Binning, G., 32–33
- Binocular stereomicroscopes, 8–9
- Bipolar devices, defects in, 173, 174, 176
- Bipolar SRAM, 445, 446
- Bird's beak dislocation. *See* LOCOS bird's-beak MED-induced dislocations
- Bit-line contacts (BLCs)
- in planar capacitors, 352
 - in stacked capacitors, 369–370, 371
- Bit lines
- capacitor on/over, 382–384, 384–385
 - in planar capacitors, 349, 351, 352
 - in planarized cylindrical stacked capacitors, 386, 387, 390, 394
 - in trench capacitors, 401, 433–436, 439
- Black belts, in localized field oxidation scheme (LOCOS) process step, 226–228, 229
- Blisters, from particles, 490
- Bohr, Mark, 56
- Boron outdiffused surface strap (BOSS), in trench capacitors, 411–412, 429–430
- Boron-phosphorus spin-on glass (BPSG), 301
- Bottle trench configuration, 439
- Brattain, Walter, 3
- Bright field (BF) imaging, 26–28
- Bright field mode reflection, 9, 11
- Buffer oxide etching (BOE), for junction delineation, 120–121
- Bulk micromachining, 538, 540
- of MEMS devices, 541–542
- Bulk transistor load SRAM, 466–471
- TFT load SRAM versus, 466
- Bull's-eye defect morphology, 482, 485
- Burger's vector, 251
- Buried contact (BC) formation
- in resistor-load SRAM, 448, 450–451, 453–454
 - in SRAM process flow, 447
- Buried oxide (BOX), 531, 533, 534, 536
- Buried plate (BPT) cell, 410
- Buried strap (BEST) contact, in trench capacitors, 412–413, 435, 436, 437, 441, 442–443
- C49 and C54 phases, of TiSi_2 salicide, 262–263, 267–269
- Capacitor dielectrics, ONO as, 210, 211
- Capacitor on data line, 382
- Capacitor on (over) bit line (COB), 382–384, 384–385
- DRAM cell design using, 369
 - with planarized cylindrical stacked capacitor, 388
- Capacitors, in DRAM cells, 346, 349, 350, 402–404. *See also* Planar capacitors; Stacked capacitors; Trench capacitor DRAM cells
- Carbon (C) evaporator, 123
- Carbon nanotubes, 643–644
- Carbon replica extraction, preparing samples via, 112–115, 116, 117, 118, 123–125
- Catastrophic breakdown, of gate dielectric, 190, 191, 197–203, 203–205, 205–208, 209
- Cavities, ESD-induced damage and, 517
- Cell-to-peripheral boundary, planarization at, 393–394
- Channel stop implant, 154
- Charge coupled devices (CCDs), 3
- Chemical analysis, SEM and TEM, 24
- Chemical delineation, in SEM and TEM junction characterization, 22–23. *See also* Junction delineation/stain
- Chemical mechanical polish (CMP), 38, 39, 41, 48–50, 51, 394. *See also* CMP planarization; Mechanical polishing
- device scaling and, 56, 58
 - multilevel interconnects and, 319
 - for planarized cylindrical stacked capacitor, 386
 - for trench capacitors, 401
 - in W-plugs, 317–318
- Chemical vapor deposition (CVD), 37, 308, 324, 325, 359. *See also* CVD entries; Low pressure chemical vapor deposition (LPCVD) SiN process step; Metal oxide chemical vapor deposition (MOCVD) devices; Selective chemical vapor deposition (SCVD)
- barrier deposition via, 328, 329
 - device scaling and, 56, 57, 60
 - on silicides, 256, 257
- Chromium (Cr). *See* Eutectic solder/Cu/Cu–Cr/Cr UBM system
- CMOS SRAM, 445, 446, 448, 457, 465
- CMOS wafer FABs, process flow technology for, 37–52
- CMP planarization, 46, 47
- Coatings, Au/Pd/Ni, 600
- Cobalt (Co) spiking, 272
- Cobalt silicide (CoSi_2), salicide process with, 269–279, 280, 281, 282, 283
- Cobalt–titanium (Co/Ti) interposing scheme, 270–272, 275–276, 277, 278, 280, 281, 282
- Cobalt-to-nickel transition, 280
- Collar oxide, in 16M to 64M DRAM devices, 424–428, 436
- Collodion replica, 113–114
- Compact disks (CDs), rewritable, 548, 551

- Complementary metal oxide semiconductor (CMOS) technology. *See also* CMOS entries; Hetero-junction CMOS (HCMOS) devices
 with planar DRAM cells, 346–348
 with stacked capacitor DRAM cells, 366
 STI and, 235
 TiSi_2 salicide in, 262
- Complementary metal oxide semiconductor field effect transistors (CMOS FETs). *See also* CMOS entries
 process flow technology for, 36, 38
 with SiGe technology, 547
- Composite polyfilms, in resistor-load SRAM, 454–455
- Conductive AFM (C-AFM), 34
- Conductors, in ULSI devices, 287
- Confocal microscopy, 13–16
- Constant voltage stress (CVS), 198
 in gate oxide breakdown model, 205–208
- Construction analysis (CA), 61–89. *See also*
 TEM applications
 described, 62–64
 of flash memory devices, 86
 flowchart for, 63
 of GaAs devices, 87–88
 of LDMOS devices, 86–87
 of leadframe design, 67–69, 72
 ULSI process, 73–86
 of wire bonding, 65–67, 68–69, 70
- Contact array areas, implantation on, 173, 174
- Contact bottom MED-induced dislocations, 168–169, 170
- Contact etching, 46, 47
- Contact failure, EOS-induced, 499–501, 502, 503
- Contact resistivity, in DRAM unit cells, 395–396
- Contact technology, trench DRAM cell, 409–413. *See also* Buried contact (BC) formation; Buried strap (BEST) contact; Copper contacts; High aspect ratio contacts; Interconnects; Metal contacts; Node contacts; Plug contacts; Stacked unit cell contacts; Stitch contacts; Tungsten (W)–plug entries; Via contact chains, sidewalls, and bottom gaps
- Contrast, in optical microscopy images, 11. *See also* Dark field (DF) imaging; Dark silicon imaging; Passive voltage contrast (PVC) technique
- Copper (Cu). *See also* Cu/Ni(V)/Al system; Eutectic solder/Cu/Cu–Cr/Cr UBM system; Eutectic solder/Cu/W(Ti) UBM system; SnAgCu entries
 adhesion to Ta barrier layer, 329–330, 331, 332
- aluminum alloys with, 297–300, 301, 307
 annealing of, 48–50, 51
 in Au–Al ball bonding system, 559
 deposition of, 48, 49, 51
 device scaling and, 56–57, 59
 electromigration of, 58
 electroplating with, 48, 49, 51, 57–58
 integration into Al metallization technology, 340, 341
 Ti/TiN/Al/TiN sandwich and, 302
 in ULSI process technology, 48–52
- Copper barrier cap, 50
- Copper contacts, voids within, 331–337
- Copper damascene, 47–48, 56, 57, 58, 217–219
- Copper grids. *See also* Finder's grid
 in glass needle extraction, 116, 117, 119
 gold and molybdenum grids versus, 120
 in making carbon replicas, 124
 TEM sample reinforcement with, 96, 98, 100, 103, 108, 109
- Copper leadframe, Au/Pd/Ni coatings on, 600
- Copper leadframe nitridation, 600
- Copper metallization, 74, 325–341
- Copper mesh. *See* Copper grids
- Copper oxide (CuO) removal, 51
- Copper/polyimide interface, 600–601
- Copper seed layer deposition, 48, 49, 51
- Corner rounding
 with STI, 238–240, 244, 245
 trench corner engineering and, 243, 245, 246, 247
- Corner thinning. *See also* Dry Kooi effect
 with STI, 236–238, 241, 242–243, 244, 245
 trench corner engineering and, 243–246, 247
- Corrosion, in Al metallization, 305, 306, 307
- Corrugated capacitor cell (CCC), 401, 404–405
- Crab-eyed morphology, 231
- Critical dimension (CD), metrology of, 74–82
- Critical-dimension SEM (CD-SEM), 21
- Cross-sectional samples
 FIB-assisted precision, 107–117, 119
 preparing, 101–105, 106–107
- Cross-sectional TEM (XTEM)
 of active area junctions, 152–154, 155, 156, 157, 158, 160, 161
 of contacts and plugs, 313, 314, 315–317
 of CoSi_2 salicide, 272, 273, 274, 275
 of crown-shaped stacked capacitors, 379
 of electromigration-induced damage, 504, 505
 of ESD-induced damage, 507, 508, 509
 of gate oxide failure, 495–498

- Cross-sectional TEM (XTEM) (*Continued*)
 of LDD junctions, 149–151
 of localized field oxidation, 223–228, 229, 230
 of MEMS devices, 543, 544
 of narrow STI lines, 248–250, 251
 in particle analysis, 480, 483–484
 of PBLOCOS, 230–234
 of phase-change optical disk, 549, 552
 of planar DRAM cells, 349–350, 351
 of SiGe devices, 547
 in single-bit failure analysis, 492–494
 of stacked DRAM cells, 374, 375, 392
 of TFT load SRAM, 462, 463
 of TiSi₂ salicide, 262–269, 270, 271
 of trench capacitors, 401–402, 403, 414, 425, 432
 of UBM systems, 565–566
 of voids within contacts, 333, 334, 335
 of W-plug contact damage, 499, 500, 501, 502, 503
 Crown-shaped stacked capacitor, 376–379, 380, 381, 385
 Crystal defect analysis, SEM and TEM, 25
 Crystal joint interface defects, 176–178
 C-SAM. *See* Scanning acoustic microscopy (SAM)
 Cumulative sum of deviations (CUSUM)
 for InAs/GaAs SAQDs, 619–623
 for interface lattice strain, 628–630
 Cu/Ni(V)/Al system, eutectic PbSn
 solder/sputtered, 563–569
 Curing damage, 70–71
 Curing samples, 101–102
 CVD oxide(s), 46, 180, 230
 CVD oxide STI fill process step, 39–41
 Damage, to gate structures, 190–209. *See also* Laser ablation
 Damage analysis. *See also* TEM failure analysis
 of electromigration-induced damage, 501–503, 504, 505
 of EOS-induced damage, 494–501, 502, 503–520
 Damascene, copper, 47–48, 56, 57, 58, 217–219. *See also* Polydamascene process
 Dark field (DF) imaging, 26, 28. *See also* Annular dark-field (ADF) technology
 Dark field illumination, 10, 11
 Dark silicon imaging, 28
 De Broglie wavelength, of electrons, 17
 Deep ultraviolet (UV) photoresist mask, 39
 Defect generation, in SAQDs, 628
 Defect reduction, in failure analysis, 478–491
 Defects. *See* Back-end defects; Back-side chipping; Crystal defect analysis; Curing damage; Damage entries; Dislocation entries; Ejector-pin induced damage; End-of-range defects (ERDs); In-line optical defect analysis systems; Mask edge defects (MEDs); Optical defect analysis systems; Particle analysis; Primary defects; Projected range defects (PRDs); Silicon-substrate damage; Stacking fault defects; Substrate defects
 Delamination, in contact failure, 500
 Dendritic arms, due to ESD-induced damage, 517, 524
 Design, leadframe, 67–69, 72
 “Design-in reliability,” 67
 Device characterization, via junction delineation, 156, 157, 158, 159, 160, 161, 172
 Device geometry limits, 134–138
 Device scaling, 52–60
 technology challenges associated with, 59–60
 Diagnosis, VLSI and ULSI process, 25–30.
 See also Failure analysis
 Diamond film, 95
 Dicing wafers, 101
 Die bonding curing, 70–71
 Die-bond thickness, construction analysis of, 69–71, 72
 Dielectric-breakdown-induced epitaxy (DBIE), 197–203
 Dielectric-breakdown-induced Ti migration (DBIM), 191–197, 204
 Dielectric cap layer, 58
 Dielectrics, 179–234. *See also* Gate dielectrics; Interlayer dielectrics (ILDs); Intermetal dielectrics (IMDs); Isolation; Oxide-nitride-oxide (ONO) dielectric; SiO₂ dielectric; Tantalum pentoxide (Ta₂O₅) dielectric
 device scaling and, 54–55, 56–59
 future materials for, 208–223, 224
 high-*k*, 179, 209, 212–214
 localized field oxidation, 223–234
 low-*k*, 47–48, 58–59, 179, 214–223, 224
 in stacked capacitors, 369
 ultra thin gate breakdown in, 184–189, 189–208, 209
 Die-pad backside
 delamination of, 71–73
 dimpling of, 69, 72
 Diffraction Kikuchi patterns, 28. *See also* Electron diffraction patterns
 Diffusion couple, in Au–Al ball bonding system, 561–562, 564
 Diode leakage, 246–252
 Discharge paths, in ESD-induced damage, 510–511, 513, 514–517, 518, 522, 523

- Dislocation loops
in polysilicon film, 292
in trench capacitors, 422, 423, 424
- Dislocations
MED-induced, 165–169, 170
shallow-trench-isolation, 169
stress-induced extended, 165–178
at trench capacitor bottom, 169–173
- Displacement, derivatives of, 625–627
- Doping
resistor-load SRAM, 451–452, 453, 455
stacked capacitor, 372
trench capacitor, 154–155, 158, 159, 160
- Drain-induced barrier lowering (DIBL), 54
- DRAM cell-packing density, developments related to, 389–390
- DRAM devices. *See also* DRAM products; Dynamic random access memory (DRAM) devices
4M, 413–422, 423, 424
16M to 64M, 422–433
64M to 256M, 433–442, 443
256M to 4Giga-bit, 442
flash memory and, 529–530
SRAM devices versus, 448
as technology drivers, 526
- DRAM processes, common problems in, 390–398
- DRAM products. *See also* DRAM devices
construction analysis of, 63–64, 75
external examination of, 64, 65, 66
history of, 345–346
with silicon-on-capacitor structure, 389
trench capacitor doping in, 154–155, 158, 159
- DRAM trench capacitor, 171, 213
- DRAM unit cells. *See also* Planar DRAM cells; Stacked capacitors; Stacked unit cell contacts
contacts and contact resistivity in, 395–396
with stacked capacitors, 365–398
structure of, 346–356
of trench capacitors, 401–404
- Dry etching. *See also* Etching
of MEMS devices, 541–542, 543, 544, 545, 546
in SEM and TEM junction characterization, 23
- Dry Kooi effect, 240–243, 244, 248. *See also* Corner thinning
- Dual-damascene process. *See* Copper damascene
- Dynamic random access memory (DRAM) devices, 345–364. *See also* DRAM entries; Planar capacitors; Stacked capacitors; Trench capacitor DRAM cells
process flow of, 346–348
- Ejector-pin induced damage, 70
- Electrical analysis, SEM and TEM, 25
- Electrically erasable programmable ROM (EEPROM, E²PROM), 527
- Electric overstress (EOS)-induced contact failure, 499–501, 502, 503
- Electric overstress-induced damage analysis, 494–501, 502, 503
ESD-induced damage and, 520–524
- Electric overstress-induced gate oxide failure, 495–498
- Electromigration, of copper, 58
- Electromigration-induced damage analysis, 501–503, 504, 505
- Electron beam-induced current (EBIC) technique, 25
- Electron diffraction patterns, 633. *See also* Reflection high-energy electron diffraction (RHEED) patterns
of CoSi₂ salicide, 277
of eutectic PbSn solder/sputtered Cu/Ni(V)/Al system, 568
of intermetallic compounds, 303
- Electron energy loss spectrometry (EELS), 24, 29, 132, 134, 612–614
of dielectric-breakdown-induced Ti migration, 195, 196, 202
of low-*k* dielectrics, 222–223
of SnAgCu on Au/sputtered Ni(V)/Ti, 589
- Electrons
analytical techniques using, 6
in Auger electron spectroscopy, 30–31
in FEG technology, 30
in flash cells, 527, 528
in SEM and TEM, 17–18
- Electroplating, with copper, 48, 49, 51, 57–58
- Electrostatic discharge (ESD)-induced damage analysis, 503–520
EOS-induced damage and, 520–524
- Emission microscope, 12–13
- End-of-range defects (ERDs), 144, 160–162, 163, 173, 174. *See also* Group I defects
MEDs versus, 165, 170
TiSi₂ salicide and, 268
- Energy dispersive X-ray spectrometry (EDS), 29, 132, 134. *See also* TEM with energy dispersive spectrometry (TEM/EDS)
of eutectic PbSn solder/sputtered Cu/Ni(V)/Al system, 566–569
in failure analysis, 478
of intermetallic compounds, 303, 305
in particle analysis, 478–479, 490
- Energy-filtered imaging, 29

- Energy-filtered mapping (EFTEM), 18, 189, 190
- Engineering, trench corner, 243–246
- Epitaxial film
- CoSi₂ salicide, 273, 274–275, 276–279
 - NiSi salicide, 284
- Epitaxial hillock extrusion, 198–202, 205.
- See also* Hillock formation
- Epitaxial islands, in polysilicon film, 291–295
- Epitaxial realignment, 290–293
- Epitaxial silicon–germanium (SiGe) technology, 542–548
- Epitaxy. *See also* Molecular beam epitaxy (MBE); Solid phase epitaxial recovery; Solid phase epitaxy (SPE)
- of aluminum alloys, 296, 298
 - polarity-dependent
 - dielectric-breakdown-induced, 197–203
- Erasable programmable ROM (EPROM), 527
- ESCA technology, 31
- ESD protection circuits, 504
- failure of, 514–515, 521
- Etching, 48, 51. *See also* Anisotropic spacer etch; Buffer oxide etching (BOE); Contact etching; Dry etching; Gate formation etching; Over-etch; Oxide etching; Reactive ion etch (RIE); STI etching; Trench etching; Undercut etching; Wet etching
- in crystal defect analysis, 25
 - device scaling and, 56
 - for junction delineation/stain, 118
 - of leadframe, 69
 - plasma, 150
 - of resistor-load SRAM, 448, 449
 - selective oxide, 395
 - silicon-substrate surface damage due to, 176
 - in SRAM process flow, 447–448
- Etching defects, 231
- Etching leadframe, 71
- Etch residues, 479, 482, 484, 489, 490
- Etch-stop, 50
- Eutectic PbSn solder/sputtered Cu/Ni(V)/Al system, 563–569
- Eutectic solder/Au/electroless Ni(P)/Al system, 569–577
- Eutectic solder/Cu/Cu–Cr/Cr UBM system, 577–581
- Eutectic solder/Cu/W(Ti) UBM system, 581–587
- Evaporators, carbon, 123
- Exit wave reconstruction (EWR), 610
- Explosions, in contact failure, 499–501, 502.
- See also* Fireball
- Extinction oscillations, with HRTEM, 610–611
- Eyelash structure, with bird's beak dislocation, 225–228, 229, 230
- Failure analysis, 477–478. *See also* Contact failure; Damage analysis; TEM failure analysis
- of gate oxide breakdown, 189–208, 209
 - junction delineation in, 156, 161, 172
 - single-bit, 491–494
- Fast Fourier transform (FFT), 134, 619–620, 621
- FEG technology, 30
- Ferroelectric RAM (FRAM), 527
- FIB-assisted precision cross-sectional samples, preparing, 107–117, 119
- FIB-induced artifacts, 131–134, 135, 136
- Field isolation, 154
- Field oxidation dielectrics 223–234
- Field oxide (Fox), dielectrics with, 223–234
- Fillers, solder mask, 601–603
- Finder's grid, 112–114
- Fin structures
- horizontal, 374–376, 377, 378, 380
 - lateral-layer, 376, 377, 378
 - vertical, 376–379, 380, 381
- Fireball, with gate oxide failure, 497, 498, 501, 502. *See also* Explosions
- Flash cells, 527–529, 530
- Flash memories, 527–530, 531, 532
- construction analysis of, 86
- Flip-chip of the board (FCOB) devices, 14
- Flip-chip packaging, 12, 14, 600–602
- Flip-chip UBM, 52
- Floating-gate avalanche injection MOS (FAMOS), 527
- Floating-gate memory, 527
- Fluorinated SiO₂ (FSG), as intermetallic dielectric, 217–219
- Fluorine (F₂) bubbles, 159–160, 163
- in TiSi₂ salicide, 264, 266
- Focal series reconstruction (FSR), 610
- Focused ion beam (FIB), 21–22, 89. *See also*
- FIB entries; Non-FIB samples
 - in carbon replica extraction, 112–115, 116, 117, 118
 - in failure analysis, 478
 - in glass needle extraction, 116, 117
 - milling via, 99–100
 - in particle analysis, 478–479, 483
 - in single-bit failure analysis, 491, 492–493
 - thinning via, 91
- Foreign particles, 479, 482–483, 485, 486
- 4Giga-bit DRAM devices, 442
- 4M DRAM devices, 413–422, 423, 424
- Fowler–Nordheim (FN) tunneling, 527, 528, 530

- Front end-of-line (FEOL) process
 metrology with, 75
 MOSFET, 52–55
- Front-end processes (FEP), 54
- Front-side parallel lapping, 100
- FTIR analysis, of low-*k* dielectrics, 222–223
- FUROX isolation technology, 233
- GaAs devices, construction analysis of, 87–88
- GaAs hetero-junctions, 614–616, 617, 618
- Gallium nitride (GaN), 636–643
- Gate dielectrics, 179–189
 catastrophic damage to, 190, 191, 197–203, 203–205, 205–208, 209
 device scaling and, 54–55, 56–59
 growth of, 42
 polysilicon and, 256–261
 ultra thin, 184–185, 189–208, 209–212, 213, 214, 215, 628–630
- Gate formation etching, 43, 44
- Gate oxide (Gox), 179–189, 396–397
 bird's beak and, 396, 397–398
 breakdown of, 189–208, 209
 EOS-induced failure in, 495–498
 ESD-induced damage in, 519
 in resistor-load SRAM, 448
- Gate oxide breakdown model, based on TEM evidence, 205–208, 209
- Gate oxide breakdown physical analysis, 203–205
- Gate smelting, 203–204
- Gate structure
 DRAM, 396–397
 in SIMOX technology, 537
- Geometry limits, 134–138
- Germanium (Ge), with NiSi salicide, 280, 284–285. *See also* Ge–Sb–Te alloy film; Silicon–germanium (SiGe) technology
- Ge–Sb–Te alloy film, in optical CDs, 548–556
- Glass, in planar capacitors, 353. *See also* Boron-phosphorus spin-on glass (BPSG). *See also* Highly doped phosphorus (HDP) glass; Interlayer dielectrics (ILDs); Phosphosilicate glasses (PSGs); Spin-on glass (SOG)
- Glass needle extraction, preparing samples via, 115–117, 119
- Gold (Au). *See* Au entries; SnAgCu on Au/sputtered Ni(V)/Ti UBM system
- Gold grid, for junction delineation, 120
- Grain growth, in copper film, 330–331, 333, 335. *See also* Lateral grain growth epitaxy; Normal grain growth epitaxy
- Grain size distribution, bimodal, 330–331
- Grain texture, in particle analysis, 482–483
- Grinding samples, 102–103
- Group I defects, 158. *See also* End-of-range defects (ERDs)
- Group II defects, 158–159. *See also* Projected range defects (PRDs)
- Group III defects, 158, 159. *See also* Mask edge defects (MEDs)
- Growth-in particles, 479–481, 482, 483, 484, 486, 487
 difficulty of finding, 480–481
- Guard ring, 154
- Guinier–Preston (GP) zones, 298–299
 in aluminum metallization, 631–636
- Hafnium aluminate (Hf-Al-O) dielectrics, 213–214, 217
- Hafnium dioxide (HfO₂) dielectric, 213–214, 217
- Hard breakdown (HBD). *See also* Catastrophic breakdown
 of gate dielectric, 190, 191, 198, 199, 204, 205
 in gate oxide breakdown model, 205–208, 209
- Hazy surface, due to particles, 479, 480
- Hemispherical grained (HSG) silicon structure, 379–382, 383
 with planarized cylindrical stacked capacitor, 387–389, 390, 391, 392
- Hetero-junction CMOS (HCMOS) devices, with SiGe technology, 547
- Hetero-junctions, SiGe and GaAs, 614–616, 617, 618
- Hexagonal trench arrangement, 406, 407, 408
- High-angle annular dark field (HAADF) detector, 132, 134, 611–614
- High aspect ratio contacts, 308, 331, 334
 in stacked capacitors, 371
- High capacitance (Hi-C) cells, 347, 416
- High-density SRAM cell, 464
- High-*k* dielectrics, 179, 209, 212–214
 in stacked capacitors, 372, 403
- Highly doped phosphorus (HDP) glass, as intermetallic dielectric, 217
- High-resolution scanning TEM (HR-STEM), 611–614
- High-resolution TEM (HRTEM), 29, 82, 83, 605, 609–646
 of 64M DRAM devices, 438
 of bicrystal defects, 176
 conventional, 610–611
 of CoSi₂ salicide, 275, 276–279, 282, 283
 of epitaxial growth, 199
 of gate oxide layers, 135, 136, 181, 182, 183, 193, 206
 of hafnium dioxide dielectric, 217
 history of, 609–610
 lattice strain measurement via, 616–630

- High-resolution TEM (HRTEM) (*Continued*)
 for metrology in conventional ULSI
 processes, 609–614
 microelectronics applications for, 630–644
 of ONO dielectrics, 210–211, 215
 of phase-change optical disk, 551, 553, 554
 of polyacrylethers, 221
 of polysilicon film, 289, 290
 of self-assembled quantum dots, 616–630
 of SiGe and GaAs hetero-junctions,
 614–616, 617, 618
 of SOI substrate, 534, 535
 of solder mask filler, 604
 of tantalum pentoxide dielectric, 216
 of W–plug contacts, 309, 312
 High-temperature re-oxidation process
 (HTR-STI), 238, 239
 Hillock formation, 299, 300, 301. *See also*
 Epitaxial hillock extrusion
 Holography, TEM, 24, 30
 Horizontal fin structures, in stacked
 capacitors, 374–376, 377, 378, 380
 Horizontal metallization, 287
 Horizontal section samples, 100
 Horseshoe dislocations, 251–252
 Hot aluminum processes, 302–306
 Howie detector, 30
 Hydrogen fluoride (HF) dip, 240, 241, 350
- Imaging
 bright field and dark field, 26–28
 dark silicon, 28
 energy filtered, 29
 Impedance noise, 325
 Implant through metal (ITM) technique,
 267–268
 Imprints on substrate, from particles, 491
 InAs–GaAs SAQDs, 628, 629
 CUSUM application for, 619–623
 Infrared (IR) microscopy, 12–13
 Infrared photon emission microscope
 (IRPEM), 13
 In-line focused ion beam, 21
 In-line optical defect analysis systems, 21–22
 Insulators, in ULSI devices, 287. *See also*
 Dielectrics
- Integrated circuits (ICs)
 device scaling of, 52
 history of, 3–5
 infrared microscopy of, 12–13
 invention of, 3, 5
 laser (scanning) confocal microscopy of,
 13–16
 light microscopy of, 10, 11, 12
 MEMS devices versus, 538
 SEM and TEM of, 19–21, 22, 23
- Interconnects
 back end-of-line, 55–59, 75, 220
 in bulk transistor load SRAM, 471
 copper, 325–329
 multilevel, 318–325
 in planar DRAM cells, 349, 350, 351
 in TFT load SRAM, 460–461
 types of ULSI, 287
- Interface area
 in eutectic solder/Au/electroless Ni(P)/Al
 system, 572–577
 in stacked capacitors, 372, 374, 375
 in trench capacitors, 407
- Interface heating, of W–plug contacts, 499,
 500
- Interface lattice strain, CUSUM and Si/SiO₂,
 628–630
- Interface reactions, W/TiN/Ti/Si, 311–315
- Interference patterns
 in infrared microscopy, 14–15, 16
 in TEM and HRTEM, 29
- Interlayer dielectrics (ILDs), 208, 218, 287
 device scaling and, 56
 particle defects in, 479
 of planar capacitors, 352–353, 357, 358,
 359, 360
 polygate burnout in, 497, 498, 511, 512,
 513, 514, 516, 517, 518, 519
 in resistor-load SRAM, 455
 in SRAM process flow, 447
 of stacked capacitors, 382
- Intermetal dielectrics (IMDs), 179, 208–209,
 214–223, 224, 287, 300, 477, 510,
 526
 device scaling and, 56
 multilevel interconnects and, 318, 319
 in SRAM process flow, 448
- Intermetallic compounds, 12
 aluminum–titanium, 302–306
- International Technology Roadmap for
 Semiconductor (ITRS)*, 219
- Ion implantation, 143–178. *See also*
 Substrate defects
 heavy, 173, 174
 junction analysis and, 143–145, 147–158
 n+ and p+ high dose, 44–45
 primary defects in as-implanted condition,
 145–147
 process steps in, 41–42, 43
 in resistor-load SRAM, 449
 substrate defects induced by, 82–85, 87
- Ion-implantation imprint, of a planar
 capacitor, 350, 352
- Ion milling, 98–99, 100, 105, 106–107, 606,
 607. *See also* Argon (Ar) ion milling
 artifacts induced via, 126–131, 132, 133
 for junction delineation/stain, 118–119
- Ions, analytical techniques using, 6
- Islands. *See also* Epitaxial islands; LOCOS
 islands; Pyramid islands

- in copper film, 331
- in wire bonding, 563
- Isolating materials, silicon dioxide as, 179.
 - See also* Dielectrics
- Isolation, shallow trench, 235–252. *See also* Dielectrics; Field isolation ring; Shallow trench isolation (STI); Trench isolation
- Isolation technologies, 233
- Jigs, for TEM samples, 91–97
- Johnson, Eric, 37
- Joule heating, 197, 205
- Junction analysis, ion implantation and, 143–145
- Junction characterization, SEM and TEM, 22–24
- Junction delineation/stain, 143, 144, 155, 156, 157, 158, 159, 160, 161, 172
 - in flash memory, 529, 532
 - of LDD junctions, 149–151
 - of n+p junctions, 148–149
 - preparing samples via, 117–123
 - in resistor-load SRAM, 449–450
- Junctions. *See also* Hetero-junction entries; Junction delineation/stain; N-type junctions; P-type junctions
 - active area, 152–154, 155, 156, 157, 163
 - ion implantation and, 147–158
 - LDD, 149–151
 - n+p, 148–149
- Junction spiking, 295, 296
- Kikuchi patterns, 28
- Kilaas, Roar, 610
- Kilby, Jack St. Clair, 3, 5
- Kilby principles, 4
- Kirkendall's voids, 67, 70, 560–561, 566, 567, 569, 577, 579–580, 581–582, 592, 593, 594, 595, 596
- Kooi's effect, 224–228, 229, 230, 397–398
 - dry, 240–243, 244, 248
- Laminated oxide-nitride-oxide (ONO) dielectric, 179
- Large scale integration (LSI), 4
- Laser ablation, of phase-change optical disk, 551–556
- Laser marks, in single-bit failure analysis, 492–493
- Lasers
 - FEG technology and, 30
 - phase-change optical disks and, 549
- Laser (scanning) confocal microscopy, 13–16
- Laser sources, in scanning confocal microscopy, 14, 16
- Laser spots, 554–556
- Laser thermal processing (LTP), NiSi salicide and, 284
- Lateral grain growth epitaxy, of polysilicon film, 290–291
- Lateral-layer fin structures, in stacked capacitors, 376, 377, 378
- Laterally doped MOS (LDMOS) devices
 - aluminum alloys in, 296
 - construction analysis of, 86–87
 - junction delineation and characterization of, 156, 160
- Lattice strain, 622. *See also* Interface lattice strain
 - calculating, 625–627
 - measuring, 616–630
- Layers, in UBM structure, 559. *See also* Polylayers
- Lead (Pb), in eutectic solder/Cu/W(Ti) UBM system, 584–586
- Leadframe design/manufacturing, 67–69, 72
- Lead-free solders, 596–599
- Lead-free solder/UBM systems, 587–596
- Leakage, diode, 246–252
- Licensing/patent technology analysis, construction analysis as, 62
- Lift-off method, 115. *See also* Glass needle extraction
- Lightly doped drain (LDD) junctions, 149–151. *See also* n-LDD/tilted wafer ion implantation; p-LDD/tilted wafer ion implantation
- Lightly doped drain structure, 38, 43, 44, 45
 - in bulk transistor load SRAM, 468
 - in planar DRAM cells, 362–363
 - in trench capacitors, 441
- Light microscopy, 8–12. *See also* Optical microscopy (OM)
- Line trench oxide mask
 - deposition of, 50
 - patterning etch on, 51
- Localized field oxidation dielectrics, 223–234
- Localized field oxidation scheme (LOCOS)
 - process step, 38, 135, 153, 154, 155, 156, 179, 223, 224–228, 229, 230. *See also* LOCOS entries; Poly buffered LOCOS (PBLOCOS)
 - with planar DRAM cells, 347, 348, 360, 361
 - with resistor-load SRAM, 448
 - in SRAM process flow, 447
- Local redeposition, ion milling artifacts from, 129–131, 132, 133
- LOCOS bird's-beak MED-induced dislocations, 165–168, 223, 224, 225–228, 229, 230, 231, 232, 233, 234, 242–243
 - in ESD-induced damage, 511–512, 516, 517, 518
 - gate oxide and, 396, 397–398

- LOCOS bird's-beak MED-induced dislocations (*Continued*)
 STI and, 235
 in W-polycide gates, 258
- LOCOS islands, 228, 229
 in planar capacitors, 352–353
- Log-normal distribution, of gate failures, 190, 192
- Long pad oxide etching time, 239, 246, 247
- Low-*k* dielectrics, 179, 214–223, 224, 325–326. *See also* Ultra-low-*k* dielectrics
 deposition of, 47–48, 58–59
- Low pressure chemical vapor deposition (LPCVD) SiN process step, 38–39, 223
- Lu, Chih-Yuan, xi
- Magnesia oxide ceramic, plan view TEM of, 399
- Magnetoresistive RAM (MRAM), 527
- Manufacturing, leadframe, 67–69, 72. *See also* Mass production
- Mask clean process step, 39–41
- Mask edge defects (MEDs), 144, 150, 151, 162–165, 166, 167, 168, 170. *See also* Group III defects; MED-induced dislocations
 due to ESD-induced damage, 519, 520
 in planar capacitors, 359–360, 361
 at polyplate edge, 359–360, 361
 in stacked capacitors, 367, 368
 TiSi₂ salicide and, 268
- Masking. *See* Line trench oxide mask;
 Photo-mask process step; Photoresist mask; Resist mask gate; Resist masking;
 Via mask
- Massive contact array areas, heavy implantation on, 173, 174
- Mass production, in ULSI and VLSI technology, 36–60
- Mass spectrometry, secondary ion, 32
- Materials. *See also* Dielectrics; Novel devices
 gate dielectric, 179–189
 isolating, 179
 optical CD, 548–556
- Mechanical polishing, 95–97, 109, 110–111, 606. *See also* Chemical mechanical polish (CMP)
 artifacts induced via, 125–126, 127
 locating particle defects after, 494
- Mechanical thinning, 91–97, 108
- MED-induced dislocations
 contact bottom, 168–169, 170
 polygate edge and LOCOS bird's-beak, 165–168
- Medium scale integration (MSI), 4
- Metal contacts, with n+ junctions, 151–152
- Metallization, 287–342. *See also* Under bump metallization (UBM)
 aluminum alloy, 295–307
 contact and plug, 308–318
 copper, 325–341
 polysilicon, 287–295
- Metallization layers, multiple, 51
- Metallography-type microscopes, 9
- Metal oxide chemical vapor deposition (MOCVD) devices, 88
- Metal-oxide semiconductor (MOS) devices, 364. *See also* CMOS entries;
 Complementary metal oxide semiconductor field effect transistors (CMOS FETs); Laterally doped MOS (LDMOS) devices
 ion implantation in, 145
 polarity-dependent dielectric-breakdown-induced epitaxy in, 197–203
- Metal-oxide semiconductor field effect transistor (MOSFET) devices, 364, 399
 bright field imaging of, 26, 27
 dielectric-breakdown-induced Ti migration in, 191–197
 front end-of-line, 52–55
 gate characteristics of, 41
 market dominance of, 526
 process flow technology for, 37
 with SiGe technology, 547
- Metal particles, 486
- Metal step coverage, 356
- Metrology, 74–82
 in conventional ULSI processes, 609–614
 SEM and TEM, 21–22, 23
- Micro-electro-mechanical systems (MEMS) devices, 538–542, 543, 544, 545, 546
 integrated circuits versus, 538
- Microelectronics, 3–8
 analytical techniques in, 5–8, 9–10
 HRTEM in, 609–646
- Micro-loading effect, W-plugs and, 315–317
- Micromachining, 538
 of MEMS devices, 541–542
- Micro-pores, in intermetallic dielectrics, 220–221, 222
- Microprocessor A
 construction analysis of, 74–86
 TEM structural analysis of, 74, 77–82, 84, 85
- Microprocessor P
 construction analysis of, 74–86
 TEM structural analysis of, 75, 77–80, 82, 83, 85
- Microprocessors, TEM structural analysis of, 77–82, 83, 84, 85
- Microscope, emission, 12–13

- Microscopy, 3–35. *See also* Scanning electron microscopy (SEM); Transmission electron microscopy (TEM)
 analytical techniques in, 5–8
 infrared, 12–13
 laser (scanning) confocal, 13–16
 light, 8–12
 optical, 6, 8–16
 photon emission, 509
 scanning tunneling and scanning probe, 32–34
- Milling
 artifacts induced via ion, 126–131, 132, 133
 focused ion beam, 99–100
 ion, 98–99, 100, 105, 106–107, 118–119
- Miniaturization, of integrated circuits, 4–5
- Misalignment, of via layers, 137, 138
- Mix MOS SRAM, 446
- Modulated stacked capacitors, 380
- Molding compound microstructure, 64, 67
- Molecular beam epitaxy (MBE), 88
- Molybdenum (Mo) grid, for junction delineation, 120
- Molybdenum silicide (MoSi_2), as gate dielectric, 257
- Moore’s law, 52
- Mouse bite defect, 416, 417
- Multilevel interconnects, 318–325
- Mushroom stud, in stacked capacitors, 373, 374
- Nanopillars, 636–643
- Nanopipes, 636–643
 defects in, 635
- Nanopores, in intermetallic dielectrics, 220–221, 223
- Nanotubes, carbon, 643–644
- Narrow STI lines, 246, 248–250, 251, 252
- NCPSL isolation technology, 233
- Near-field scanning optical microscope (NSOM), 14–15
- n+ high dose ion implantation, 44–45
- Nickel (Ni). *See* Cu/Ni(V)/Al system; Ni(P)/Al system; SnAgCu on Au/electroless Ni(P)/Al UBM system; SnAgCu on Cu/sputtered Ni(V)/Al UBM system
- Nickel (Ni)-based under-bump metallization, 563–569, 569–577
- Nickel silicide (NiSi , NiSi_2), salicide process with, 279–285
- Ni(P) UBM systems
 eutectic solder/Au/electroless, 569–577
 summary table of, 570
- Nitric acid (HNO_3), for junction delineation/stain, 118, 119
- Nitridation, copper leadframe, 600
- Nitride cap, 46, 51
 in planar DRAM cells, 362
 in trench capacitors, 441
- Nitride spacer. *See also* Silicon nitride (Si_3N_4) mask
 in SRAM device, 472
 TiSi_2 salicide and, 265–267, 268, 269
- n-LDD/halo tilted wafer ion implantation, 43–44
- NMOS SRAM, 445, 446, 448, 457, 458, 461
- Node contacts
 in stacked capacitors, 369, 370, 394
 in trench capacitors, 409–410, 411, 412
- Nomarski differential interference contrast, 10, 11
- Non-FIB samples, preparing, 106–107
- Nonprecision sample preparation, stages for, 92, 93
- Nonvolatile semiconductor memory (NVSM), 527
- Normal grain growth epitaxy, of polysilicon film, 290–291
- Novel devices, 526–557. *See also* Materials flash cells, 527–529, 530
 MEMS devices, 538–542, 543, 544, 545, 546
 SiGe technology, 542–548
 silicon-on-insulator (SOI) technology, 530–538
- N+/p junctions, 148–149
- N-type junctions
 delineation/stain of, 119, 149–151
 metal contacts with, 151–152
- Nucleation energy barriers, aluminum metallization and, 297
- N-well ion implant, 41–42
- On-state current, in bulk transistor load SRAM, 467–468
- Optical CD materials, 548–556
- Optical defect analysis systems, 21–22
- Optical interference microscope (OIM), 14–15
- Optical microscopy (OM), 6, 8–16. *See also* Light microscopy
 in failure analysis, 478, 493
- Over-etch
 in resistor-load SRAM, 449
 in W-plugs, 317–318
- Oxide etching, 37
- Oxide-nitride-oxide (ONO) dielectric, 209–212, 213, 214, 215, 261. *See also* Laminated oxide-nitride-oxide (ONO) dielectric
 in stacked capacitor DRAM cells, 366, 369, 372, 388, 392

- Oxide-nitride-oxide (ONO) dielectric
(Continued)
 in trench capacitor DRAM cells, 400, 403,
 414, 415–417, 424–427, 429, 438,
 439
 twin-bit failure in, 491–492
- Oxide particles, analysis of, 485–486, 487,
 489
- Oxides, thermal gate, 179–189
- Oxygen (O_2)
 in SIMOX technology, 531
 in thermal gate oxides, 179
- Oxygen precipitation, 159–160
- Oxynitride, as gate dielectric material,
 179–189. *See also* Oxide-nitride-oxide
 (ONO) dielectric
- Package construction analysis, 63–64, 64–73
- Package quality, evaluating, 71–73
- Package-related applications, of TEM,
 600–604
- Package-related devices, TEM sample
 preparation for, 604–607
- Packaging technology, TEM in, 558–608
- Parallel trench arrangement, 409
- Parasitic transistors, 39, 40, 211
- Particle analysis, in failure analysis, 478–491
- Passivation, 51–52, 307
- Passive voltage contrast (PVC) technique, 25
- Per-amorphization implant (PAI) technique,
 267–268
- Percolation process, in gate oxide breakdown
 model, 206, 207–208
- Perpendicular trench arrangement, 409
- Phase-change optical disk, 548, 549
- Phase identification, SEM and TEM, 24
- Phase-shift oscillator, first integrated circuit
 as, 3
- p+ high dose ion implantation, 44–45
- Phosphorus (P), $TiSi_2$ salicide and, 267. *See*
also Boron-phosphorus spun-on glass
 (BPSG)
- Phosphosilicate glasses (PSGs), 307
- Photo-mask process step, 39
 in MEMS device construction, 540–541
- Photon emission microscopy, of ESD-induced
 damage, 509
- Photons
 analytical techniques using, 6
 in emission microscopy, 13
- Photoresist mask, deep UV, 39
- Photoresist particles, 486
- Pinhole failure, 491–492
- Planar capacitors
 cross-sectional TEM analysis of, 349–350
 plan view TEM analysis of, 350–356
 stacked capacitors versus, 369
 structure of, 348–360
- Planar diffusion plate, 346
- Planar DRAM cells, 346. *See also* DRAM
 unit cells
 construction of, 346–348
 issues related to, 356–360
 structural variations in, 360–364
- Planarization, at the cell-to-peripheral
 boundary, 393–394
- Planarization oxide, 394
- Planarized cylindrical stacked (PST)
 capacitor, 386–388, 390
- Plan view TEM analysis, of planar DRAM
 cells, 350–356
- Plan view samples, 100
- Plasma-enhanced CVD (PECVD)
 as intermetallic dielectric, 217
 of MEMS devices, 540–541
- Plasma etching, 150
- p-LDD/tilted wafer ion implantation,
 43–44
- Plug contacts, 308–318. *See also* Tungsten
 (W)–plug entries
- PMOS devices, 447, 448, 457, 458, 460, 461,
 466
- Poisoned contacts, 312
- Polarity-dependent
 dielectric-breakdown-induced epitaxy
 (DBIE), 197–203, 204, 205
 in gate oxide breakdown model, 205–208,
 209
- Polishing. *See also* Chemical mechanical
 polish (CMP)
 artifacts from mechanical, 125–126, 127
 via ion milling, 98–99, 105, 106–107
 mechanical, 95–97, 109, 110–111
 of samples, 102–103, 104, 106–107
- Polyacrylethers (PAEs), as intermetallic
 dielectrics, 221
- Poly buffered LOCOS (PBLOCOS), 228–234
 process steps in, 229–230
 reverse L-shaped sealed, 230–234
- Polycide, 256–261, 361–362, 363. *See also*
 Polysilicon film; Tungsten
 (W)–polycide
 in planar DRAM cells, 347–348
 in resistor-load SRAM, 454–455
 salicides and, 261
 in stacked capacitors, 367, 368, 369, 370,
 383, 395
 in trench capacitors, 401, 402, 415, 423,
 428, 429, 431
 tungsten and, 313–315
- Polycrystalline Si, 42
- Polycrystalline Si damascene process, 395.
See also Damascene
- Polyfill
 with BEST contacts, 412–413
 boron-doped, 411–412

- in trench capacitors, 405, 408, 420–421, 426, 427, 429
- Polyfilm. *See* Composite polyfilms; Polysilicon film
- Polygate
gate oxide failure in, 495, 496, 497
in stacked capacitors, 367, 373, 374
- Polygate edge MED-induced dislocations, 165–168
- Poly 1 polysilicon, 361–364, 414, 415, 417–418, 419, 423
- Poly 2 polysilicon, 361–364, 414, 415, 416, 417–418
- Poly 3 polysilicon, 414, 415
- Poly 4 polysilicon, 415
- Polyimide, as intermetallic dielectric, 220–221
- Poly lava growth, in ESD-induced damage, 512, 516
- Poly layers, in stacked capacitors, 369–370. *See also* Layers
- Polymer particles, 486
- Polynodes, in stacked capacitors, 386–388, 389, 391
- Polyplate
mask edge defects in, 359–360, 361
in planar capacitors, 354–355, 357–359
of trench capacitors, 401, 426
- Polyplate-to-word-line short, with planar capacitors, 357–359
- Poly runners, in stacked capacitors, 367
- Polysilicon film, 75, 240, 241, 246, 247. *See also* Poly 1-4 entries; Polycide
boron-doped, 411
ESD-induced failure and, 516, 517, 518
in MOS devices, 287–295
in particle analysis, 479–485
in planar capacitors, 353–354, 361–364
in planar DRAM cells, 347–348, 348–349
in resistor-load SRAM, 450–456, 457, 458, 459
in stacked capacitors, 367–369
in TFT load SRAM, 456–466, 467, 468, 469, 470
in trench capacitors, 403–404, 410, 420–421
twin-bit failure in, 492
W/TiN/Ti and, 312–315
- Polysilicon load resistor, SRAM and, 457–458
- Polysilicon particles, analysis of, 479–485
- Polysilicon plugs (poly-plugs), 308, 309
in stacked capacitors, 370, 371, 393, 394
- Polysilicon silicide (Polycide). *See* Polycide; Polysilicon film; Silicides; Silicon (Si)
- Polysilicon surface, textured, 379–382, 383
- Poly stringers, in trench capacitors, 404, 416
- Poly wrap, 240. *See also* Polysilicon film
- Popcorn failure, 71–73
- Porous substrates, intermetallic dielectrics as, 220–221, 222–223
- Post-drive-in condition, secondary defects in, 158–165
- Power devices, with trench isolation, 173–176
- Pre-FA basic device structure/process analysis, construction analysis as, 62
- Precipitation, of Al₂Cu particles, 297–300, 301
- Precision cross sections, preparing, 106–107
- Precision plan view samples, preparing, 106
- Precision sample preparation, stages for, 92–94
- Pre-thinning, 108
- Primary defects, in as-implanted condition, 145–147
- Printed circuit boards (PCBs), solder mask fillers for, 601–602
- Process flow technology, for CMOS wafer FABs, 37–52
- Process technology, ULSI and VLSI, 36–60
- Product analysis, construction analysis as, 62
- Programmable ROM (PROM), 527
- Projected range defects (PRDs), 144, 160–162, 163, 173, 174. *See also* Group II defects
MEDs versus, 165, 170
TiSi₂ salicide and, 268
- P-type junctions, delineation/stain of, 119
- Punch-through ion implants, 41–42
- Punch-through location, in ESD-induced damage, 511–512, 513, 514
- P-well ion implant, 41, 42
- Pyramid islands, in polysilicon film, 293
- Quantum dots, self-assembled, 616–630
- Radiation-hard circuits, SOI technology for, 530
- Random access memory (RAM), 445
- Random cross-sectional analysis, 61
- Rapid thermal anneal (RTA), 45, 46, 60
- Rapid thermal oxidation (RTO), 189
- Rayleigh criterion, 17
- Reactive ion etch (RIE), 39
of MEMS devices, 538, 540
for trench capacitors, 400
- Read-only memory (ROM), 527
- Rectangular trench arrangement, 406–409
- Redeposition, ion milling artifacts from, 129–131, 132, 133
- Reflection high-energy electron diffraction (RHEED) patterns, 619
- Reliability, packaging, 71–73

- Resistance, of polysilicon film, 456–457
 Resistance and capacitance (RC) impedance noise, 325
 Resistance and capacitance (RC) product, device scaling and, 56, 57
 Resist mask gate, 44
 Resist masking, 51
 Resistor load (R-load) SRAM, 446, 448–456
 Resolution
 atomic Z-contrast, 29, 30
 optical microscopy, 11, 12
 Rutherford backscattered spectrometry, 31–32
 scanning electron microscope, 18, 33–34
 transmission electron microscope, 18
 Reverse engineering, construction analysis as, 62
 Reverse FN tunneling, 527
 Reverse L-shaped sealed PBLOCOS (RLS-PBL), 230–234
 Reverse whisker formation, 257
 Rewritable CDs (CD-RWs), 548, 551
 Rohrer, H., 32–33
 Rugged capacitor, 379–380
 Rutherford backscattered spectrometry (RBS), 31–32
 ion implantation and, 144
- Salicides, 261–285
 active area junction with, 152–154, 155, 156, 157, 163
 advantages of, 262
 CoSi_2 , 269–279, 280, 281, 282, 283
 formation of, 45–46, 261–262, 312
 NiSi , 279–285
 TiSi_2 , 262–269, 270, 271
 Salicide technology, 84, 85
 Sample preparation artifacts, 125–138
 device geometry limits and, 134–138
 FIB-induced, 131–134
 ion milling induced, 126–131, 132, 133
 mechanical polishing induced, 125–126, 127
 Samples. *See also* TEM sample preparation
 curing of, 101–102
 grinding of, 102–103
 non-FIB, 106–107
 polishing of, 102–103, 104, 106–107
 preparing cross-sectional, 101–105, 106–107
 preparing precision, 106–107
 stacking of, 101–102, 103, 106, 136, 143
 TEM stages and jigs for, 91–97
 Scanning acoustic microscopy (SAM)
 of die bonds, 69
 of DRAM products, 64, 66
 package quality and, 71–73
 Scanning capacitance microscopy (SCM), 30, 34, 122, 143, 144
 of trench capacitors, 420
 Scanning confocal microscopy, 13–16
 Scanning electron microscopy (SEM), 6, 17–25. *See also* SEM entries
 in carbon replica extraction, 112–115, 116, 117, 118
 chemical analysis via, 24
 crystal defect analysis via, 25
 of die bonds, 72
 electrical analysis via, 25
 of ESD-induced damage, 509, 510, 515, 517, 522
 in failure analysis, 478
 of FIB cuts, 111
 of FIB thinning artifacts, 136–138
 HRTEM and, 609–610
 ion implantation and, 144
 junction characterization via, 22–24
 of leadframe, 71, 72
 of MEMS devices, 539, 542
 metrology via, 21–22, 23
 modes of operation of (table), 18
 in particle analysis, 478–479
 phase identification via, 24
 principles of, 17–18
 of shadowing effects, 110
 surface morphology analysis via, 18–21, 22, 23
 as a tool for VLSI and ULSI process diagnosis, 25–26
 of wire bonding, 65–67, 68–69, 70
 Scanning optical microscope (CSOM), 14–15
 Scanning probe microscopy (SPM), 32–34
 Scanning spread resistance microscopy (SSRM), 30, 143, 144
 Scanning thermal microscopy (SThM), 34
 Scanning transmission microscope (STEM), 18, 29–30
 surface morphology analysis via, 19–20
 Scanning tunneling microscopy (STM), 32–34, 202
 applications for, 34
 invention of, 32–33
 Secondary electrons (SE), 17, 18
 Secondary ion mass spectrometry (SIMS), 32, 143, 144
 in failure analysis, 478
 of low- k dielectrics, 222–223
 of trench capacitors, 420, 421
 Secondary substrate defects, in post-drive-in condition, 158–165
 Selective chemical vapor deposition (SCVD), of TiSi_2 salicide, 269
 Selective oxide etching, 395

- Self-aligned (SAC) poly–plug, 393, 395–396
in SRAM process flow, 447
- Self-aligned salicide technology, 84, 85
- Self-aligned silicide (salicide), 261–285. *See also* Salicides
- Self-assembled quantum dots (SAQDs), 616–630
defects and defect generation in, 628
InAs–GaAs, 619–623
SiGe, 623–625
- SEM applications, ULSI process construction analysis as, 73
- Semiconductor devices. *See also* Complementary metal oxide semiconductor field effect transistors (CMOS FETs); Metal-oxide semiconductor (MOS) devices; Metal-oxide semiconductor field effect transistor (MOSFET) devices
- Semiconductor memory, types of, 527. *See also* DRAM entries; Dynamic random access memory (DRAM) devices; Flash memories; SRAM devices; Static random access memory (SRAM) products
- SEM with energy dispersive spectrometry (SEM/TDS), 24
- Separation by implantation of oxygen (SIMOX) technology, 531–533
- Shadowing effects, 110
from particles, 488
- Shallow threshold ion implants, 41–42
- Shallow trench isolation (STI), 134–135, 153, 179, 235–252. *See also* STI entries; Trench capacitor doping in 64M DRAM devices, 431, 432, 433, 435, 437, 439, 442–443
advantages of, 236
amorphization via, 145–147
corner rounding in, 238–240, 244, 245
corner thinning in, 236–238, 241, 242–243, 244, 245
in CoSi₂ salicide, 273
formation of, 38–39
narrow lines from, 246, 248–250, 251, 252
process steps in, 235
- Sheng, George T. T., xi, 19, 22, 26, 27, 91
- Shockley, William, 3
- Shockley partial dislocations, 166, 170
- Short pad oxide etching time, 239, 245
- Short-period superlattice (SL), 628, 629
- Si(110) poles, 28
- Sidewall, in MEMS device construction, 540–541
- Sidewall doping uniformity, in trench process, 418–420
- Sidewall effect, in stacked capacitors, 372–374. *See also* Via contact chains, sidewalls, and bottom gaps
- SiGe hetero-junctions, 614–616
- SiGe SAQDs, in silicon matrix, 623–625
- Silicidation, 312–315, 339–340
- Silicide particles, 486
- Silicides
in active area junctions, 152–154, 155, 156
defects in, 160–162
formation of, 45–46, 256
polysilicon, 256–261
self-aligned, 261–285
- Silicon (Si). *See also* Metal-oxide semiconductor field effect transistor (MOSFET) devices; Polysilicon film; Si(110) poles; W/TiN/Ti/Si interface reaction
in CMOS VLSI devices, 37–38
in gate deposition, 42–43, 44
in gate dielectrics, 179–189
in gate oxide breakdown model, 205–208, 209
hemispherical grained, 379–382, 383
infrared microscopy of, 12
in salicide formation, 45–46
in SIMOX technology, 531, 533, 534, 535, 536
- Silicon crystal and bicrystal defects, 176–178
- Silicon dioxide (SiO₂), 38–39, 40, 42. *See also* SiO₂ dielectric; Si/SiO₂ interface lattice strain
ESD-induced damage to, 506
fireball damage and, 501
as isolating material, 179–189
in SIMOX technology, 531, 533, 534, 535
- Silicon–germanium (SiGe) technology, 542–548. *See also* SiGe entries
- Silicon matrix, SiGe SAQDs in, 623–625. *See also* Silicon substrate
- Silicon micromechanical structures, 538
- Silicon nitride (SiN) deposition, 38–39, 40, 41, 42, 45
in bulk transistor load SRAM, 468
device scaling and, 57
- Silicon nitride (Si₃N₄) mask. *See also* Nitride spacer
Kooi’s effect and, 224–228, 229, 230, 242
salicides and, 261
for trench capacitors, 400
- Silicon nodules, 295–300
- Silicon-on-capacitor (SOC) structure, for DRAM cells, 389
- Silicon-on-insulator (SOI) technology, 530–538
advantages of, 536–538
- Silicon plugs (Si–plug), 308
- Silicon substrate. *See also* Silicon-substrate entries
in 4M DRAM devices, 413–415

- Silicon substrate. (*Continued*)
 in 16M to 64M DRAM devices, 424, 425, 436, 437
 aluminum alloys on, 295–307
 contacts and plugs and, 308–311, 311–315
 ESD-induced damage to, 505–514, 515, 517, 518
 gate oxide failure in, 495–498
 in MEMS devices, 539, 540, 541
 oxide-nitride-oxide dielectric and, 209–210
 in particle analysis, 483–484, 487, 492–494
 in planar DRAM cells, 349–360
 polysilicon film with, 288
 in resistor-load SRAM, 449–451
 in SIMOX technology, 531–536
 spiking into, 336, 337–340
 in stacked DRAM cells, 365, 366, 369–372
 in trench capacitors, 402–404, 405, 410, 412
 Silicon-substrate amorphization, 158
 via ion milling, 127–128, 133
 from shallow trench isolation, 145–147
 Silicon-substrate damage, 126, 127, 130, 132, 133, 242–243, 244, 245, 246–252
 from dielectric-breakdown-induced Ti migration, 191–197
 induced by wafer process steps, 173–178
 via ion implantation, 144, 145
 from polarity-dependent dielectric-breakdown-induced epitaxy, 197–203
 with STI, 236–238
 stress-induced, 165–178
 Silicon wafers, preparing TEM samples of, 100
 Silver (Ag). *See* SnAgCu entries
 Single-bit failure analysis, 491–494
 SiN stop layer, 47
 etching in, 48
 SiO₂ dielectric
 aluminum alloys and, 295–307
 deposition of, 47–48
 device scaling and, 54–55, 59
 etching in, 48
 as intermetallic dielectric, 217, 218, 219, 222, 242
 in polysilicon film, 293–295
 tantalum pentoxide with, 212–213, 216
 Si/SiO₂ interface lattice strain, 628–630
 16M DRAM devices, 422–433
 64M DRAM devices, 422–442, 443
 Small scale integration (SSI), 3–4
 SnAgCu on Au/electroless Ni(P)/Al UBM system, 591–596
 SnAgCu on Au/sputtered Ni(V)/Ti UBM system, 587–590
 SnAgCu on Cu/sputtered Ni(V)/Al UBM system, 590–591, 592
 Soft breakdown (SBD), of gate oxides, 189, 190, 192, 195, 197, 198, 199, 204, 205–208, 209
 Soft error rate (SER)
 resistance of, 456–457
 SRAM and, 456
 Soft X-ray imaging, of DRAM products, 64, 65
 Solder mask fillers, 601–603
 Solder, lead-free, 596–599
 Solid phase epitaxial recovery, 158–159
 Solid phase epitaxy (SPE), 144
 Spacer formation, nitride and/or oxide deposition for, 44, 45
 Spectrometry, Rutherford backscattered, 31–32, 144. *See also* Mass spectrometry
 Spectroscopy, Auger electron, 30–31, 144, 478
 Spiking, into silicon substrate, 336, 337–340.
See also Aluminum (Al) spiking; Cobalt (Co) spiking; Junction spiking
 Spin-on glass (SOG)
 as intermetallic dielectric, 217, 301–302
 multilevel interconnects and, 318, 319
 Spread-resistance probing (SRP), 143, 144 of trench capacitors, 421
 Sputtering
 of aluminum, 300–302
 ion milling artifacts from, 127, 128, 129–131
 SRAM devices
 DRAM devices versus, 448
 flash memory and, 529–530
 SRAM unit storage cells
 high-density, 464
 structure of, 446–447
 Stacked capacitor design, compatible technologies in, 384–389
 Stacked capacitors, 365–398
 cell structure evolution of, 372–390, 391, 392
 crown-shaped, 376–379, 380, 381, 385
 modulated, 380
 planar capacitors versus, 369
 planarized cylindrical, 386–388
 process flow of, 366
 structure of, 367–369
 trench capacitors versus, 409–410
 unit cells of, 367–372
 Stacked-gate avalanche injection MOS (SAMOS), 527
 Stacked unit cell contacts, 369–372
 Stack gate structure, 86
 Stacking fault defects, 143
 in polysilicon film, 292

- Stacking samples, 101–102, 103, 106, 136, 143
 CoSi₂ salicide, 276
 Stages, for TEM samples, 91–97
 Staining. *See* Junction delineation/stain
 Stair rod dislocations, 166, 170
 Stamping leadframe, 71
 Static random access memory (SRAM)
 products, 445–474. *See also* SRAM entries
 bulk transistor load, 466–471
 construction analysis of, 75–76
 defects in, 175
 history of, 445–446
 metal contacts in, 151–152
 polysilicon film in, 288
 process flow for, 447–448
 resistor load, 446, 448–456
 as technology drivers, 471–472
 TFT load, 456–466, 467, 468, 469, 470
 Step coverage, with aluminum, 300–302
 Stereomicroscopes, 8–9
 STI-associated dislocations, 169
 STI etching, 39
 STI stress-induced substrate defects, 246–252
 Stitch bonding, 65–67
 Stitch contacts, 369
 Strained SiGe devices, 548
 Strain-induced contrast modulation, with HRTEM, 610–611
 Strap node contacts, in trench capacitors, 410, 411
 Stress-induced extended dislocations, 165–178
 Stress-induced substrate defects, 246–252, 301, 416–417
 with CoSi₂ salicide, 270, 272–275
 with NiSi salicide, 279
 Substrate defects, 158–178. *See also* Ion implantation; Silicon-substrate entries
 bicrystal and crystal joint interface, 176–178
 ion implantation induced, 82–85
 secondary, 158–165
 STI stress-induced, 246–252
 tertiary, 165–178
 Substrate imprints, from particles, 491
 Substrate-plate-trench (SPT) capacitor cell, 404, 405, 410
 Supporting spring grids, in MEMS device, 540, 541
 Surface bumps, due to particles, 479, 480
 Surface damage, to silicon substrate, 176
 SWAMI isolation technology, 233
 Syton polishing, 126
 Sze, S. M., xi
- Ta(N) barrier deposition, 48, 49, 51
 Tantalum (Ta), device scaling and, 56–57
 Tantalum (Ta) barrier, 85–86, 329–330, 331, 332, 335, 336, 337, 338, 339–340
 Tantalum pentoxide (Ta₂O₅) dielectric, 54, 212–213, 216
 Tantalum silicide (TaSi₂), as gate dielectric, 257
 Tape automatic bonding (TAB) technology, 581
 Technology drivers
 DRAM devices as, 526
 SRAM devices as, 471–472
 TEM analysis. *See also* Transmission electron microscopy (TEM)
 of gate oxides, 179–189
 of ultra thin gate dielectric breakdown, 189–208, 209
 TEM applications, 61–89
 construction analysis and, 62–64
 package construction analysis, 64–73
 ULSI process construction analysis as, 73–86
 TEM cross-sectional analysis, 101–105, 144, 145–178
 TEM electron diffraction (TEM-ED), 24
 TEM failure analysis, 477–525
 of electromigration-induced damage, 501–503, 504, 505
 of EOS-induced damage, 494–501, 502, 503
 of EOS-induced versus ESD-induced damages, 520–524
 of ESD-induced damage, 503–520
 particle analysis and defect reduction in, 478–491
 process of, 477–478
 single-bit, 491–494
 TEM holography, 24, 30, 143, 144
 TEM-observed junctions, 147–158
 Temperature
 gate failure and, 191, 192
 infrared microscopy and, 12–13
 ion milling artifacts and, 126–127
 in wire bonding, 562–563, 564
 TEM sample preparation, 89, 90–139, 604–607. *See also* Sample preparation artifacts
 basic procedures for, 91–100
 by carbon replica extraction, 112–115, 116, 117, 118, 123–125
 conditions for, 90–91
 cross-sectional samples, 101–105
 for FIB-assisted precision cross-sectional samples, 107–117, 119
 via glass needle extraction, 115–117, 119
 by junction delineation/stain, 117–123
 plan view (horizontal section) samples, 100

- TEM sample preparation (*Continued*)
 precision cross sections, 106–107
 precision plan view samples, 106
- TEM samples, reinforcement with Cu grids, 98, 100, 103, 108, 109
- TEM sample stages/jigs, 91–97
- TEM test patterns, 88–89
- TEM with energy dispersive spectrometry (TEM/EDS), 24, 85. *See also* Energy dispersive X-ray spectrometry (EDS)
- TEOS oxide, multilevel interconnects and, 318–319
- Tertiary substrate defects, 165–178
- Textured polysilicon surface, 379–382, 383
- TFT load SRAM, 456–466, 467, 468, 469, 470
 bulk transistor load SRAM versus, 466
- Thermal emission, microscopy of, 12–13
- Thermal oxide, 179–180
 as gate dielectric material, 179–189
- Thermal oxide liner, 39–41
- Thermal SiO₂ process step, 38–39
- Thickness. *See also* Wafer thickness
 back-side thinning
 determining with HRTEM, 610–611
 die-bond, 69–71, 72
 in eutectic solder/Au/electroless Ni(P)/Al system, 572, 575–577
 of gate oxide layer, 180–189
- TEM sample, 90, 104
- thin film, 74–82
 Ti/TiN, 311–312
- Thin films, W-polycide, 257
- Thin film transistor (TFT). *See also* TFT load SRAM
 polysilicon film in, 288
 in SRAM process flow, 447–448
- Thinning. *See also* FIB thinning; Wafer thickness back-side thinning
 via FIB, 91
 via ion milling, 98–99, 105, 106–107, 134
 mechanical, 91–97, 108
- Three-dimensional imaging, for metrology, 21–22
- Tin (Sn), in eutectic solder/Cu/W(Ti) UBM system, 582–587. *See also* SnAgCu entries
- Tin solders, 587
- Tin whiskers, 596–599
- Titanium (Ti), in DRAM unit cells, 395–396.
See also Eutectic solder/Cu/W(Ti) UBM system
- Titanium caps, 272
- Titanium deposition
 in active area junctions, 152, 154, 156
 ion implantation amorphization and, 146–147, 148
- Titanium migration
 dielectric-breakdown-induced, 191–197
 titanium flux in 195–196
- Titanium nitride (TiN), in Au–Al ball bonding system, 560. *See also* Titanium/titanium nitride (Ti/TiN)
- Titanium silicide (TiSi₂), salicide process with, 262–269, 270, 271
- Titanium/titanium nitride (Ti/TiN). *See also* Ti/TiN/Al/TiN sandwich structure;
 Titanium nitride (TiN); W/TiN/Ti/Si interface reaction
 electromigration-induced damage to, 501–502, 505
 multilevel interconnects and, 319, 320, 321, 322–325
 thickness of, 311–312
 as a W–plug barrier layer, 46, 47, 85, 308–311
- Ti/TiN/Al/TiN sandwich structure, 302–306
- Topography analysis, SEM and TEM, 18–20
- Transistors. *See also* Metal-oxide semiconductor field effect transistor (MOSFET) devices; Thin film transistor (TFT)
 dielectric-breakdown-induced Ti migration in, 191–197
 in DRAM cells, 346, 349, 350
 as flash cells, 527
 gate oxide breakdown model for, 203–205, 205–208, 209
 invention of, 3, 4
 parasitic, 39, 40
 polarity-dependent dielectric-breakdown-induced epitaxy in, 197–203
 SiGe technology for, 542–548
 in SRAM devices, 445
 stacked capacitors and, 365
 stacked unit cell, 367–369
 testing gate dielectrics of, 189–190
 in trench capacitor DRAM cell, 402
 in ULSI devices, 76
- Transistor-to-capacitor arrangement, in 16M DRAM devices, 430–431
- Transmission electron microscopy (TEM), 6, 17–25. *See also* Cross-sectional TEM (XTEM); High-resolution TEM (HRTEM); Plan view TEM analysis; TEM entries
 chemical analysis via, 24
 of copper metallization, 332
 of CoSi₂ salicide, 272, 273, 274, 275
 of crown-shaped stacked capacitors, 379
 crystal defect analysis via, 25
 of DRAM products, 74
 electrical analysis via, 25
 in electronics packaging, 558–608

- of ESD-induced damage, 506, 508, 512, 513, 514, 515, 517–524
- of flash memory device, 528
- gate oxide breakdown model using, 205–208, 209
- of intermetallic compounds, 303
- junction characterization via, 22–24
- of magnesia oxide ceramic, 399
- of MEMS devices, 539, 540, 541, 542, 543, 544, 546
- metrology via, 21–22, 23
- modes of operation (table), 18
- of narrow STI lines, 248–250, 251
- of passivation voids, 307
- of phase-change optical disk, 550, 551, 556
- phase identification via, 24
- of planar DRAM cells, 348–349
- of polysilicon film, 289
- principles of, 17–18
- of stacked capacitors, 368, 392
- studying low- k dielectrics with, 220–223, 224
- surface morphology analysis via, 19–20
- of TiSi₂ salicide, 262–269, 270, 271
- as a tool for VLSI and ULSI process diagnosis, 25–30
- of trench capacitors, 401–402, 403, 413, 420–422, 425, 430, 431, 432, 433
- of voids within contacts, 333, 334, 335, 339
- of W-polycide gates, 258–259
- Trench capacitor bottom, dislocations at, 169–173
- Trench capacitor designs, 402–404
- Trench capacitor doping, 154–155, 158, 159, 160
- Trench capacitor DRAM cells, 399–444
 - examples of, 413–443
 - process flow for, 400–401
 - structural evolution of, 404–413
 - unit cell in, 401–404
- Trench capacitors
 - geometry of trenches in, 405–409
 - stacked capacitors versus, 409–410
- Trench corner engineering, 243–246, 247, 248
- Trench DRAM cell contact technology, 409–413
- Trench etching, 400–401
- Trench isolation, power device with, 173–176. *See also* Shallow trench isolation (STI)
- Trench morphology, 405–409
- TSOP package, popcorn failure in, 73
- Tung, Chih-Hang, xi
- Tungsten (W). *See* Eutectic solder/Cu/W(Ti) UBM system
- Tungsten (W)–plug contacts, 85, 308–318.
 - See also* W/TiN/Ti/Si interface reaction in DRAM unit cells, 395–396
 - failure of, 499–501, 503, 505–506
 - micro-loading effect and, 315–317
 - multilevel interconnects and, 319–320
 - over-etch and voids within, 317–318
 - in stacked capacitors, 370, 371
- Tungsten (W)–plug filling, 46, 47, 311
 - device scaling and, 56
 - for junction staining, 121–122, 123
- Tungsten (W)–polycide, 257, 258–259, 361–362, 363, 431, 434, 435, 436, 439, 440–441, 455. *See also* Polycide
- Tungsten silicide (WSi_x), 256
 - as gate dielectric, 257–261
 - with planar capacitors, 356–357, 358
 - polysilicon film with, 288
 - step coverage with, 356–357
- Turn-around time (TAT)
 - for construction analysis, 61
 - for failure analysis, 478
- Twin-bit failure, 491, 492
- 256M DRAM devices, 433, 442
- UBM systems. *See also* Under bump metallization (UBM)
 - eutectic solder/Cu/Cu–Cr/Cr, 577–581
 - eutectic solder/Cu/W(Ti), 581–587
 - lead-free solder, 587–596
 - SnAgCu on Au/electroless Ni(P)/Al, 591–596
 - SnAgCu on Au/sputtered Ni(V)/Ti, 587–590
 - SnAgCu on Cu/sputtered Ni(V)/Al, 590–591, 592
- ULSI devices
 - construction analysis of, 74–86
 - DRAM cell with planar capacitor, 345–364
 - DRAM cell with stacked capacitor, 365–398
 - DRAM cell with trench capacitor, 399–444
 - fundamental components of, 287
 - NiSi salicide and, 279
 - static random access memories, 445–474
- ULSI process construction analysis, TEM as a tool for, 73–86
- ULSI process diagnosis. *See also* Failure analysis
 - particle size and, 478
 - TEM as a tool for, 25–30
- ULSI process technology, 36–60. *See also* Process flow technology
 - device scaling, 52–60

- ULSI process technology (*Continued*)
 history of DRAM manufacturing in, 345–346
 metrology in, 609–614
 Ultra large scale integration (ULSI), ix, xi, 4.
See also ULSI entries; Very large scale integration (VLSI); VLSI entries
 process flow of, 36–60
 Ultra-low-*k* dielectrics, 222–223. *See also* Low-*k* dielectrics
 Ultra thin gate dielectrics
 breakdown of, 184–189, 189–208, 209
 ONO as, 209–212, 213, 214, 215
 Si/SiO₂ interface lattice strain in, 628–630
 Under bump metallization (UBM), 52, 131, 558–559. *See also* UBM entries
 Ni-based, 563–569, 569–577
 TEM sample preparation for, 604–607
 types of, 559
 Undercut etching, 238–239, 244, 245, 246, 247
 Unit cells. *See* DRAM unit cells; SRAM unit storage cells
 Vertical cell capacitor, 346
 Vertical fin structures, 376–379, 380, 381
 Vertical interlayer connection, 287
 Vertical stack, for stacked capacitors, 372–374, 375
 Very large scale integration (VLSI), ix, 4. *See also* ULSI entries; Ultra large scale integration (ULSI); VLSI entries
 process flow of, 36–60
 TiSi₂ salicide in, 262
 Via contact chains, sidewalls, and bottom gaps, 318, 319–325, 326, 327, 328, 329, 558
 failure in, 499–501
 Via layers, Misalignment of, 137, 138
 Via mask
 patterning with, 50
 in SRAM process flow, 447–448
 VLSI construction analysis, 64
 VLSI devices
 complexity of, 61, 74
 construction analysis of, 74–86
 VLSI process diagnosis, TEM as a tool for, 25–30
 VLSI process technology, 36–60. *See also* Process flow technology
 Voids
 in Al metallization, 307
 within copper contacts, 331–337
 in CoSi₂ salicide, 272–276
 Kirkendall's, 67, 70, 560–561, 566, 567, 569, 577, 579–580, 581–582, 592, 593, 594, 595, 596
 micro-loading effect and, 315–317
 in stacked capacitors, 370
 in TiSi₂ salicide, 264–267
 in W-plugs, 317–318, 321, 322–323, 324, 325, 326, 327, 328, 329
 Voltage stress, in gate oxide breakdown model, 205–208
 Wafer dicing, 101
 Wafer FABs. *See also* CMOS wafer FABs
 in particle analysis, 478
 process flow technology for, 36, 37, 52
 SRAM devices and, 472
 Wafer map, 478
 Wafer process steps, defects induced by, 173–178
 Wafer-scaled integration (WSI), 318
 Wafer thickness back-side thinning, 101
 Water, Kooi's effect and, 224–225, 242
 Weak beam dark field (WBDF) images, 88
 Weak breakdown, of gate oxides, 190. *See also* Soft breakdown (SBD)
 Weibull distribution, of gate failures, 190–191, 192
 Wet chemistry process, 538–539
 Wet etching, 238–239, 411
 of MEMS devices, 538–539
 of planar capacitors, 354–355
 single-bit failures and, 491–492
 Whisker formation, 257–260
 in ESD-induced damage, 511–512
 Whiskers, tin, 596–599
 Wire bonding, 65–67, 68–69, 70, 562–563, 564
 Wollaston prism, 10
 Word lines
 in planar capacitors, 351
 polyplate short to, 357–359
 in trench capacitors, 401, 406, 407, 408, 433–435, 437, 439, 440
 Wormhole process, 257–261
 W-plugs. *See* Tungsten (W)-plug entries
 W-polycide. *See* Tungsten (W)-polycide
 W/TiN/Ti/Si interface reaction, 311–315
 X-ray diffractometry, (XRD), 24
 X-ray imaging, of DRAM products, 64, 65
 X-ray photoelectron spectroscopy (XPS), 31
 in failure analysis, 478
 Z-contrast technique, 24, 612, 614
 atomic resolution, 29, 30
 Zigzag layout
 of stacked capacitors, 367–369
 of trench capacitors, 402
 Zigzag morphology, in phase-change optical disks, 549