Introduction to

Metrology Applications in IC Manufacturing

Tutorial Texts Series

- *Introduction to Liquid Crystals for Optical Design and Engineering*, Sergio Restaino and Scott Teare, Vol. TT98
- *Digital Converters for Image Sensors*, Kenton T. Veeder, Vol. TT97
- *Laser Beam Quality Metrics*, T. Sean Ross, Vol. TT96
- *Military Displays: Technology and Applications*, Daniel D. Desjardins, Vol. TT95
- *Interferometry for Precision Measurement*, Peter Langenbeck, Vol. TT94
- *Aberration Theory Made Simple, Second Edition*, Virendra N. Mahajan, Vol. TT93
- *Modeling the Imaging Chain of Digital Cameras*, Robert D. Fiete, Vol. TT92
- *Bioluminescence and Fluorescence for* In Vivo *Imaging*, Lubov Brovko, Vol. TT91
- *Polarization of Light with Applications in Optical Fibers*, Arun Kumar, Ajoy Ghatak, Vol. TT90
- *Digital Fourier Optics: A MATLAB Tutorial*, David G. Voeltz, Vol. TT89
- *Optical Design of Microscopes*, George Seward, Vol. TT88
- *Analysis and Evaluation of Sampled Imaging Systems*, Richard H. Vollmerhausen, Donald A. Reago, Ronald Driggers, Vol. TT87
- *Nanotechnology: A Crash Course*, Raúl J. Martín-Palma and Akhlesh Lakhtakia, Vol. TT86
- *Direct Detection LADAR Systems*, Richard Richmond, Stephen Cain, Vol. TT85
- *Optical Design: Applying the Fundamentals*, Max J. Riedl, Vol. TT84
- *Infrared Optics and Zoom Lenses, Second Edition*, Allen Mann, Vol. TT83
- *Optical Engineering Fundamentals, Second Edition*, Bruce H. Walker, Vol. TT82
- *Fundamentals of Polarimetric Remote Sensing*, John Schott, Vol. TT81
- *The Design of Plastic Optical Systems*, Michael P. Schaub, Vol. TT80
- *Fundamentals of Photonics*, Chandra Roychoudhuri, Vol. TT79
- *Radiation Thermometry: Fundamentals and Applications in the Petrochemical Industry*, Peter Saunders, Vol. TT78
- *Matrix Methods for Optical Layout*, Gerhard Kloos, Vol. TT77
- *Fundamentals of Infrared Detector Materials*, Michael A. Kinch, Vol. TT76
- *Practical Applications of Infrared Thermal Sensing and Imaging Equipment, Third Edition*, Herbert Kaplan, Vol. TT75
- *Bioluminescence for Food and Environmental Microbiological Safety*, Lubov Brovko, Vol. TT74
- *Introduction to Image Stabilization*, Scott W. Teare, Sergio R. Restaino, Vol. TT73
- *Logic-based Nonlinear Image Processing*, Stephen Marshall, Vol. TT72
- *The Physics and Engineering of Solid State Lasers*, Yehoshua Kalisky, Vol. TT71
- *Thermal Infrared Characterization of Ground Targets and Backgrounds, Second Edition*, Pieter A. Jacobs, Vol. TT70
- *Introduction to Confocal Fluorescence Microscopy*, Michiel Müller, Vol. TT69
- *Artificial Neural Networks: An Introduction*, Kevin L. Priddy and Paul E. Keller, Vol. TT68
- *Basics of Code Division Multiple Access (CDMA)*, Raghuveer Rao and Sohail Dianat, Vol. TT67
- *Optical Imaging in Projection Microlithography*, Alfred Kwok-Kit Wong, Vol. TT66
- *Metrics for High-Quality Specular Surfaces*, Lionel R. Baker, Vol. TT65
- *Field Mathematics for Electromagnetics, Photonics, and Materials Science*, Bernard Maxum, Vol. TT64
- *High-Fidelity Medical Imaging Displays*, Aldo Badano, Michael J. Flynn, and Jerzy Kanicki, Vol. TT63
- *Diffractive Optics–Design, Fabrication, and Test*, Donald C. O'Shea, Thomas J. Suleski, Alan D. Kathman, and Dennis W. Prather, Vol. TT62
- *Fourier-Transform Spectroscopy Instrumentation Engineering*, Vidi Saptari, Vol. TT61
- *The Power- and Energy-Handling Capability of Optical Materials, Components, and Systems*, Roger M. Wood, Vol. TT60
- *Hands-on Morphological Image Processing*, Edward R. Dougherty, Roberto A. Lotufo, Vol. TT59
- *Integrated Optomechanical Analysis*, Keith B. Doyle, Victor L. Genberg, Gregory J. Michels, Vol. TT58
- *Thin-Film Design: Modulated Thickness and Other Stopband Design Methods*, Bruce Perilloux, Vol. TT57

(For a complete list of Tutorial Texts, see http://spie.org/tt.)

Introduction to
Metrology Applications in IC Manufacturing

Bo Su
Eric Solecky
Alok Vaid

Tutorial Texts in Optical Engineering
Volume TT101

SPIE PRESS
Bellingham, Washington USA

Library of Congress Cataloging-in-Publication Data

Su, Bo, author.
Introduction to metrology applications in IC manufacturing / Bo Su, Eric Solecky, and Alok Vaid.
pages cm – (Tutorial texts in optical engineering ; volume TT 101)
Includes bibliographical references and index.
ISBN 978-1-62841-811-8
1. Integrated circuits–Measurement. 2. Integrated circuits–Design and construction. I. Solecky, Eric, author. II. Vaid, Alok, author. III. Title. IV. Series: Tutorial texts in optical engineering ; v. TT 101.
TK7874.58.B63 2015
621.3815–dc23

2015016029

Published by

SPIE
P.O. Box 10
Bellingham, Washington 98227-0010 USA
Phone: +1 360.676.3290
Fax: +1 360.647.1445
Email: Books@spie.org
Web: http://spie.org

Printed in the United States of America.
First printing

SPIE.

Introduction to the Series

Since its inception in 1989, the Tutorial Texts (TT) series has grown to cover many diverse fields of science and engineering. The initial idea for the series was to make material presented in SPIE short courses available to those who could not attend and to provide a reference text for those who could. Thus, many of the texts in this series are generated by augmenting course notes with descriptive text that further illuminates the subject. In this way, the TT becomes an excellent stand-alone reference that finds a much wider audience than only short course attendees.

Tutorial Texts have grown in popularity and in the scope of material covered since 1989. They no longer necessarily stem from short courses; rather, they are often generated independently by experts in the field. They are popular because they provide a ready reference to those wishing to learn about emerging technologies or the latest information within their field. The topics within the series have grown from the initial areas of geometrical optics, optical detectors, and image processing to include the emerging fields of nanotechnology, biomedical optics, fiber optics, and laser technologies. Authors contributing to the TT series are instructed to provide introductory material so that those new to the field may use the book as a starting point to get a basic grasp of the material. It is hoped that some readers may develop sufficient interest to take a short course by the author or pursue further research in more advanced books to delve deeper into the subject.

The books in this series are distinguished from other technical monographs and textbooks in the way in which the material is presented. In keeping with the tutorial nature of the series, there is an emphasis on the use of graphical and illustrative material to better elucidate basic and advanced concepts. There is also heavy use of tabular reference data and numerous examples to further explain the concepts presented. The publishing time for the books is kept to a minimum so that the books will be as timely and up-to-date as possible. Furthermore, these introductory books are competitively priced compared to more traditional books on the same subject.

When a proposal for a text is received, each proposal is evaluated to determine the relevance of the proposed topic. This initial reviewing process has been very helpful to authors in identifying, early in the writing process, the need for additional material or other changes in approach that would serve to strengthen the text. Once a manuscript is completed, it is peer reviewed to ensure that chapters communicate accurately the essential ingredients of the science and technologies under discussion.

It is my goal to maintain the style and quality of books in the series and to further expand the topic areas to include new emerging fields as they become of interest to our reading audience.

James A. Harrington
Rutgers University

Table of Contents

Preface

We contemplated writing a book about IC manufacturing metrology for years. Each of us in parallel recognized a gap that existed in the field of metrology and inspection. Until now, metrology (not to be confused with meteorology) had no book from the perspective of an engineer in a manufacturing and development environment in the modern IC industry. The closest we could find was a chapter called "Metrology Methods in Photolithography," written by Laurie J. Lauchlan, Diana Nyyssonen from IBM Microelectronics, and Neal Sullivan from DEC, within a book titled *Handbook of Microlithography, Micromachining, and Microfabrication, Vol. 1* (SPIE Press, 1997). The other significant text comes from Alain Diebold, in a book titled *Handbook of Silicon Semiconductor Metrology* (CRC Press, 2001). Numerous other books, especially statistical books, briefly mention metrology in the context of gauge studies, precision-to-tolerance ratios (also called the gauge maker's rule), and repeatability and reproducibility.

The three of us combined have been working in semiconductor IC fabs for more than two decades, specializing in metrology in semiconductor manufacturing. This book allows us to share our learning, understanding, and experiences with our fellow engineers and managers. As in other disciplines, metrology is constantly improving, enhancing, and developing to meet ever-increasing needs in today's high-technology manufacturing.

In IC fabs, people refer to metrology equipment as process equipment tools or a fleet of tools. We will use that term in this book to mean metrology equipment or systems. Photomasks, masks, and reticles are all used to describe the photomasks used in IC manufacturing, interchangeably. Many acronyms in IC fabs are used in this book; please refer to the List of Acronyms for their exact meanings.

The contents of the book are organized not only for metrology engineers but also for other process engineers and fab managers to better understand metrology data and the uncertainties associated with those data. It also serves as a textbook for students and researchers who are interested in metrology in general or as a reference book on fundamentals and latest developments in the IC industry.

The first two chapters introduce metrology at its most basic level. The first half of Chapter 1 defines metrology, its origin, and its purpose regardless of the field of practice. The second half of the chapter discusses measurement methods and the measurement process and includes descriptions of operator, machine, and what is being measured along with the evolution of manual and automated measurements. It closes with a discussion on applications of industrial metrology. Chapter 2 explores metrology fundamentals as they pertain to traditional measurement system characterization and calibration. The word "traditional" is used because this is an area where much has changed since the 1997. Chapter 3 discusses the need to improve upon the traditional forms of measurement system characterization and calibration. Many newer concepts are introduced here in the areas of system matching, long-term stability monitoring, and accuracy. Chapter 4 focuses on a particular industrial application of metrology—the semiconductor industry—that has the tightest metrology requirements in the world. It explores areas such as the pervasiveness and value of metrology, target design, and process control. Chapter 5 features optical metrology measurement techniques such as ellipsometry and scatterometry. Chapter 6 presents charged-particle measurement techniques, with a primary focus on scanning electron beam metrology. Chapter 7 explores other measurement techniques not used inline in 1997, such as x-ray and *in-situ* metrology. Chapter 8 discusses the limits of metrology and the evolution of hybrid metrology. Hybrid metrology development and implementation is another major change. Chapter 9 discusses metrology in mask making, emphasizing the differences as compared to wafer metrology. The final chapter closes with trends, perspectives on future metrology challenges, and other considerations not covered in the other chapters.

Example Excel spreadsheets are provided in the accompanying CD in the area of measurement uncertainty analysis—specifically, precision, matching and relative accuracy. These files complement the textbook material and help readers understand the metrology concepts better (as well as leverage these spreadsheets in their own work, as needed).

Although there is a significant focus on the semiconductor industry, many of the concepts can be easily applied to other industrial fields. The metrology field has grown significantly over time, especially in the areas of semiconductor manufacturing. With this growth comes an increasing need for metrology expertise. This book is intended to introduce a new generation to metrology while also helping current practitioners. Almost every other major semiconductor discipline, such as lithography and etch, is taught in academia; there is very little metrology material taught currently, and we hope this book helps spark increased academic interest.

Acknowledgments

Bo Su would like to thank his many colleagues over the past 20 years. He would also like to thank his family (Xiaoyu Chen, Brooke Su, and Brian Su) for their love and support.

Eric Solecky would like to acknowledge three metrology pioneers: Diana Nyssonen, Charles (Chas) Archie, and Bill Banke.

> I first met Diana during my first few years at IBM as a new hire in my first professional assignment. I will never forget attending meetings with metrology suppliers, where Diana always knew the right questions to ask and often strongly influenced the supplier's future directions. I thought to myself, "wouldn't it be nice to have this ability?" As chance would have it, Diana became my mentor and taught me many valuable skills for the next few years before her passing. Also before then, Charles Archie joined IBM to lead the Advanced X-ray Lithography Facility, started developing an interest in metrology, and worked closely with Diana. Diana sensed her early passing and, as my mentor, brought Chas and me into a room and told him to take care of me, which is exactly what Chas did. I continued to learn from Chas until his retirement from IBM in 2012. He was unbelievably giving of his knowledge. If he did not know the answer to my questions (which was rare), he would think about it and always come back to me with an answer. Chas continued to mentor numerous other metrologists, and his industry influence was immense. Bill Banke worked in IBM Burlington in the standards lab. I got to know him through Chas, and we published a few papers together. Chas and Bill pioneered the TMU accuracy metric that relied on the Mandel regression. This breakthrough served as the catalyst for the other important measurement system analysis metrics: FMP and TMP. I was very lucky to be surrounded by these metrology pioneers.
>
> I considered publishing a book on metrology for over two years but never quite knew how to move forward. As fate would have it, being on an SPIE committee, I was selected to review a metrology book proposal from Bo Su. I reached out to Bo and explained my outline for a book comparable to his, and we had numerous

discussions. We agreed to combine his thoughts with mine and move forward together. I very much appreciate this opportunity he granted me. I reached out to Alok Vaid because I knew he could bring some great material to the table. I have a lot of respect for Alok from my interactions with him throughout the years, and he was the perfect third partner for the book. I would like to thank my family, especially my wife, Chrisanne, and son, Alex, for their support in writing this book.

Alok Vaid would like to acknowledge five industry pioneers: Chas Archie, Matthew Sendelbach, Cornel Bozdog, Olivier Vatel, and Carsten Hartig.

I was fortunate enough to have crossed paths with them at different stages of my career, and each of them in their own unique way helped shape my technical expertise and grow passion for metrology. Like Eric Solecky, I was fortunate to have a chance to work with Chas Archie at IBM Fishkill for several years before his retirement. He mentored me while we worked on projects aimed at expanding the sphere of CD-SEM metrology and solving the growing concerns of gaps in reference metrology. At same time, I also had the privilege of working with Matthew Sendelbach, who, in turn, helped grow my expertise in OCD as well as data analysis techniques that should be used to qualify various metrology techniques (e.g., TMU analysis). Together we conducted a couple of key next-generation metrology tool evaluations (that later became benchmarks) and also collaborated on and published several papers on advancements in OCD, one of which won the Best Metrology Paper award from SPIE.

I had the pleasure to know and work with Cornel Bozdog, who in my opinion is one of the best scientists in the field of optical metrology, but at the same time a very humble and charming personality. Cornel is an epitome of collaboration, and together we have been expanding and advancing the new field of hybrid metrology. Over the past years, I have learnt a lot from both him and Matthew Sendelbach and developed close friendships with them. I met Olivier Vatel a few years ago when he joined GLOBALFOUNDRIES as a senior executive in charge of the technology development group. I had several close interactions with him as he helped me develop the practice of never being satisfied with the status quo and understanding that true innovation happens when you have a motivated "team" (as opposed to silos) willing to collaborate. I also had the pleasure of working and collaborating closely with Carsten Hartig over the last decade. He has been and continues to be a technical mentor/coach who has helped me understand the fundamental capabilities and limits of various metrology techniques. I am always amazed by his deep knowledge of metrology techniques, especially CD-SEM and OCD, and even though he is based in Germany he remains a dear friend and my go-to person for technical advice.

I would like to acknowledge my co-author of this book, Eric, who gave me this opportunity to work on this book. I have known him for several years, including at IBM Fishkill, but I got the chance to work very closely with him only recently when he moved to GLOBALFOUNDRIES. Again, I am fortunate to have an expert like him sitting right next to my desk (literally), and together we have been challenging and expanding the boundaries of metrology/process control.

Finally, I would like to acknowledge my wife and fellow semiconductor engineer, Niti Garg, who has always been at my side through good and rough times over the years. She also happens to be my internal customer (as a Process Engineer at GLOBALFOUNDRIES), and my metrology techniques along with expertise, continue to be challenged to meet sub-angstrom specifications for her etch equipment!

We would like to thank SPIE Press for supporting technical publications like ours and for the opportunity to publish this Tutorial Text. We appreciate the great help and guidance from Tim Lamkins and Scott McNeill (our book editor).

Bo Su
Eric Solecky
Alok Vaid
August 2015

List of Acronyms

ADC	Automatic defect classification
ADI	After-development inspection
AEI	After-etch inspection
AFM	Atomic force microscope
ANOVA	Analysis of variance
APC	Advanced process control
BEOL	Back end of line
BMS	Benchmark measurement system
BSE	Backscattered electrons
CD	Critical dimension
CD-SEM	Critical-dimension scanning electron microscope
CDU	Critical dimension uniformity
CMP	Chemical and mechanical polishing
Cp and Cpk	Process capability indices
CTF	Contrast transfer function
DOE	Design of experiment
DOF	Degree of freedom
DR	Design rules
DRAM	Dynamic random access memory
DUV	Deep ultraviolet
FEOL	Front end of line
FinFET	Fin field-effect transistor
FMP	Fleet measurement precision
FOV	Field of view
e-beam	Electron beam
EBI	Electron-beam inspection
EBR	Electron-beam review
EUV	Extreme ultraviolet
GOF	Goodness of fit
HAR	High aspect ratio
HRXRD	High-resolution x-ray diffraction
HVP	High-volume production
IC	Integrated circuit
IC fab	IC fabrication plant
IR	Inspection report

ITRS	International technology roadmap for semiconductors
LCL	Lower control limit
LSL	Lower spec limit
MPU	Microprocessing unit
MS	Measurement system
MSA	Measurement system analysis
MTT	Mean to target
OAI	Off-axis illumination
OCD	Optical CD measurement technique
OPC	Optical proximity correction
POR	Process (or product) of record
PSA	Process-stressed artifact
PSM	Phase shift masks, including attenuated (or embedded) and alternating PSM
PTR	Precision-to-tolerance ratio
R&R (or gauge R&R)	Repeatability and reproducibility
RAE	Rotating analyzer ellipsometry
RET	Resolution enhancement technique (e.g., OPC, OAI, PSM)
rms	Root mean square
RMS	Reference measurement system
SE	Secondary electrons (in a SEM)
SE	Spectroscopic ellipsometry
SEM	Scanning electron microscope
SIS	Slope-induced shift
SMO	Source mask optimization
SNR	Signal-to-noise ratio
SoC	System on a chip
SPC	Statistical process control
Spec	Specifications
SRAF	Subresolution assistant feature
SRAM	Static random access memory
STEM	Scanning transmission electron microscope
SWA	Sidewall angle
TEM	Transmission electron microscope
TMP	Tool-matching precision
TMS	Test measurement system
TMU	Total measurement uncertainty
TuT	Tool under test
UCL	Upper control limit
USL	Upper spec limit
WIP	Work (or wafer) in process
XPS	X-ray photoelectron spectroscopy
XRF	X-ray florescence
XRR	X-ray reflectance

Chapter 1
Introduction

1.1 What is Metrology?

According to the Bureau International des Poids et Mesures (BIPM), or International Bureau of Weights and Measures, metrology is "the science of measurement, embracing both experimental and theoretical determinations at any level of uncertainty in any field of science and technology." It is particularly important in modern assembly-line manufacturing, such as the automobile industry, high-tech manufacturing, and the very-large-scale integrated circuit (IC) industry.

Metrology is a broad field, and it can be classified based on the method used (or the nature of the parameters being measured) or on its applications. The science of physical measurement is called *physical metrology*; likewise, the science of chemical measurement is called *chemical metrology*. The former measures physical parameters such as time, length, mass, velocity, electrical charge, electrical current, etc. The latter measures the qualitative and quantitative analysis of substances used in chemical, medical, biological, and environmental fields. Forensic analysis, environmental analysis, medical analysis, and analysis of food and agricultural products involve chemical metrology. Both physical and chemical metrology are sometimes needed in a specific set of analysis. For example, to review defects in IC manufacturing, both the physical size and composition of a defect are analyzed to pinpoint the root cause.

This book primarily addresses physical metrology in industrial manufacturing environments, particularly in an IC manufacturing fabrication plant (IC fab). A critical-dimension scanning electron microscope (CD-SEM) is used as an example to illustrate many of the concepts discussed herein.

Based on its applications, metrology has three distinctive branches:[1]

- **Scientific metrology** concerns the establishment of quantity systems, unit systems, units of measurement, the development of new measurement methods, realization of measurement standards, and the transfer of traceability from these standards to users in society. The BIPM maintains a database of the metrological calibration and measurement capabilities of various institutes around the world. These institutes, whose activities are peer-reviewed, provide the top-level reference points for metrological traceability. In the area of measurement, the BIPM has identified nine metrology areas, including length, mass, and time.

- **Applied or industrial metrology** concerns the application of measurement science to manufacturing and other processes, and their use in society, ensuring the suitability of measurement instruments, their calibration, and quality control of measurements. Although this branch emphasizes the measurements themselves, traceability of the calibration of the measurement devices is necessary to ensure confidence in the measurements.
- **Legal metrology** concerns activities that result from statutory requirements and measurement, units of measurement, measuring instruments, and methods of measurement that are performed by competent bodies. Such statutory requirements might arise from, amongst others, the need to protect human health, public safety, the environment, consumers, and fair trade, as well as enable taxation.

This book covers only topics related to industrial metrology—in particular, metrology usage in high-technology industries such as IC semiconductor manufacturing—although these concepts can be applied to other industries.

1.2 Measurements and Metrology

Measurement is as old as written human history. People long ago discovered how to measure time with the movement of the moon, sun, and stars. Humanity's inherent need to explore, find shelter, and make tools and weapons drove people to learn how to measure length and distances using human body parts, such as their arms, hands, or feet. The advent of trade required the measure of sizes, weight, and volume. The great monuments that were built thousands of years ago and survive today, such as the Egyptian pyramids and China's Great Wall, required precise measurements and demonstrate the engineering skill of those civilizations. Historically, the human body has been used to provide the basis for units of length (hence the imperial unit "foot"), although modern metrology has established such units using less variable means.

Metrology has a much shorter history. Its science and technology, associated with understanding the measurement process and data analysis of instrumental technologies primarily began after the Industrial Revolution in the 18th century. Weaponry development and repair of that era required more precise measurements to manufacture more precise parts. For example, a spherical cast-iron shot had to fit into a cannon bore; too large or too small a size could cause a catastrophic effect. For reasons related to accuracy, efficiency, and economy, all bores of any size of cannon had to have the same

Table 1.1 Major metrology instrument invention milestones of optical, SEM and AFM.

Instrument	Year invented
Telescope/microscope	1595
Ellipsometer	1887
Scanning electron microscope (SEM)	1932
Scanning tunneling microscope (AFM)	1986

diameter; likewise, the iron shot had to be the same size within limited variations.

Rapid development of metrology followed scientific and technological advancements in the early 20th century, especially the advancement of quantum physics and chemistry. Modern assembly-line manufacturing in the automobile industry and aerospace industry in the mid-20th century pushed metrology to new heights. The IC industry started in the early 1960s and saw an explosive expansion in manufacturing scale in the 1980s. It inherited all of the learning from the automobile and aerospace industries and pushed metrology into a new territory, especially with regard to small-dimension measurements and automation. New concepts are needed to deal with some of the unique issues in the IC industry.

Table 1.1 illustrates the rapid development of measurement systems that occurred due to patent filing. It took nearly 300 years for the ellipsometer to emerge after the microscope, whereas it took less than 50 years to invent the SEM after the ellipsometer.

In the 21st century, atomic-scale resolution instruments (such as the TEM and STM) have been used in the IC industry to measure transistor-gate oxide thickness on the order of nanometers (a few layers of atoms) and check the interface quality of the oxide and the gate by counting atoms. The measurement technology truly has been in the quantum regime.

1.3 Metrology in Daily Life

> "All forms of physical and chemical measurement affect the quality of the world in which we live." —BIPM Website

People constantly acquire, digest, and use various measurement-related information surrounding them: "what time is it?", "what is the outside temperature?", "what is my current location?", "how far away is my destination?", "how fast am I driving?", "is my car too close to other cars?", etc. That information is gathered either by instruments/sensors or by personal experiences and estimation. Decisions are then made based on those measured quantities: "what should I wear outside?", "do I have enough gasoline in my car to reach where I am going?", "when do I need to leave so that I can arrive before a meeting starts?" Modern society increases the need for such information, and modern technology is constantly evolving to fill that ever-increasing need. We can measure how many steps we walked today or where we have been in the last month thanks to GPS tracking devices, such as smartphones. The burgeoning Internet of Things (IOT) will push those frontiers even further.

Instruments are used everywhere to monitor our surroundings. At home, clocks tell us time, thermometers measure temperature, and timers monitor baking time. In cars, the speedometer tells us how fast we travel, how fast the engine spins, and how much gasoline is in the tank (which, when purchased, is measured by gallons). Groceries at the store are weighed in pounds. During a physical examination, a nurse measures our weight, height, blood pressure, body temperature, and heartbeat rate.

Some people assume that these measurements are exact and thus base their analysis and conclusions on that assumption. In most mundane cases, there is no harm with such an assumption because either the measurement is sufficiently accurate or the tolerance is sufficiently large. However, in areas of trade, healthcare, and environmental pollution, this assumption may cause adverse effects even though the probability is low. In modern manufacturing facilities, it can cause significant problems. There are always variations in any measurement system that affects individual measurement and thus the decisions based on the data. A significant part of metrology in modern manufacturing involves analyzing, characterizing, and minimizing those variations in measurement system development and system quality control.

More importantly, an increasing number of tests and measurements are involved in modern medicine to help doctors diagnose disease. Any misleading test results or measurements can produce erroneous diagnoses, and, in certain cases, a confirmation, reconfirmation, or second opinion is needed. Though it may be obvious, measurements and metrologies are a large part of everyday life. The appropriate application of metrology can save lives and significantly affect product quality (and thus profitability).

1.4 Applications in Modern Manufacturing

Modern manufacturing started in the automobile industry in the early 1930s. Ford Motor Company improved and perfected the assembly line for cars. The fundamental requirement for assembly-line manufacturing is interchangeable parts—any parts can fit into any assembled systems (e.g., cars) in any assembly lines, and the end-assembled products would meet predefined performance specifications. To ensure such interchangeable parts, each must be manufactured and qualified before it is used, which is where metrology systems come in—they control the part manufacturing process and qualify each part. Quality control is at the center of modern manufacturing, and metrology is the foundation of quality control.

Since the invention of integrated circuits by Jack Kilby (then at Texas Instruments) and Robert Noyce (then at Fairchild Semiconductor; he later co-founded Intel Corp.) in the early 1960s, the semiconductor IC industry has progressed at a historical pace following an empirical law first proposed by Gordon Moore of Intel in 1965. The eponymous Moore's law states that the transistor density in IC chips (MPU, DRAM, SoC, etc.) would double every two years (the period has grown longer as IC chips become more and more complicated over the years). Such rapid progress owes to the constant reduction of minimal feature sizes of IC chips (from ~10 μm in the 1970s to less than 20 nm today), gradual expansion of chip sizes, and, indirectly, a shift toward increasingly larger wafer sizes (from 10 mm in the 1960s to 300 mm in the late 1990s). Figure 1.1 illustrates Moore's law for the microprocessor transistor count from 1971 to 2011.

IC semiconductor devices are mostly CMOS-based digital devices and are built on silicon wafers with a round shape. From the bare silicon wafer to the last protection layer (called passivation layer), CMOS-based devices are

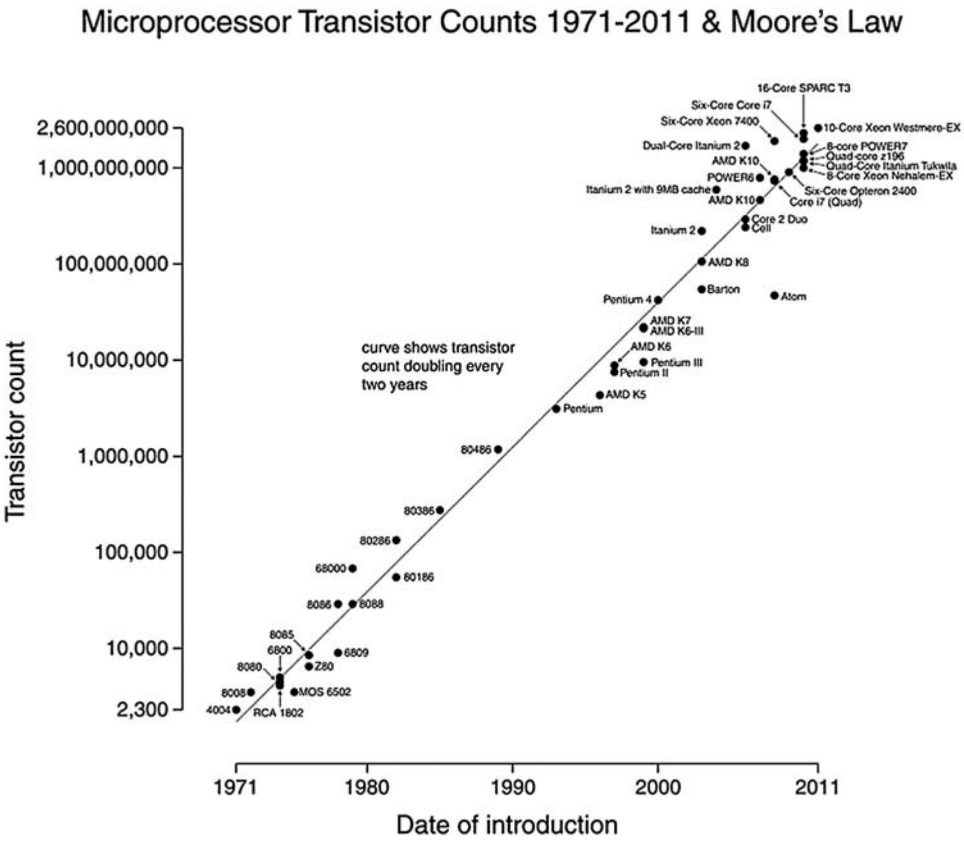


Figure 1.1 Moore's law for microprocessor transistor counts over four decades. Adapted from Wikipedia with permission under the terms of the GNU Free Documentation License.

manufactured layer by layer, through device isolation, gate definition, device formation, contact formation, and back-end interconnections. The processes involved are mainly film deposition, optical lithography, wet chemical and dry plasma etching, CMP (chemical and mechanical polishing), diffusion and ion implantation, cleaning, and rapid thermal processes. For most IC devices, there are typically 40–80 layers and 200–500 process steps. Today, the most-advanced IC devices contain over 800 process steps and take up to three months to completely fabricate a working chip. This is also one of the reasons that metrology becomes more important because more process steps mean more chances for process excursions to occur and thus higher cost if not caught in time.

When the minimal feature of IC devices (normally referring to a device gate width) was larger than 1 μm, a linewidth (or a spacewidth), called critical dimension (CD), was measured with optical-microscope-based measurement instruments (Optical CD tools). In late 1980s, when the IC device gate width shrank to less than 1 μm, optical CD tools were close to their resolution limits (directly related to the light wavelength for illumination) and had a hard time resolving features edges. CD measurement systems based on scanning electron microscopy (here denoted as CD-SEMs) (CD-SEM) were introduced into IC manufacturing. Over the next 5–6 years, as it became more and more mature, CD-SEMs had completely replaced optical CD tools in IC fabs. In 1990s, the CD-SEM became fully automated with robust pattern recognition technology for wafer alignment and measurement location navigation. From then on, the CD-SEM has experienced tremendous progress in image quality, advanced measurement algorithms, higher throughput, and ease of use. It has been a critical part of IC process development and control in IC manufacturing.

A CD-SEM is an image-based, direct CD measurement system. Due to the enhanced secondary electron (SE) emission from a feature edge (a line edge or a space edge), a strong signal from an edge region of any feature provides the basis for edge-detection algorithms to work. A CD-SEM is just one type of metrology tool used in IC manufacturing; many others measure other physical properties of the chip, such as overlay and film thickness, which are discussed in more detail in subsequent sections.

1.5 Standards and Traceability

The National Institute of Standards and Technology (NIST) is the principal standards organization in the United States, whereas the BIPM is the international organization that ensures world-wide uniformity of measurements and their traceability to the International System of Units (SI). As a part of the United States Commerce Department, NIST serves as the repository for most of the US physical and chemical measurement standards.

It also coordinates its measurement standards with the standards of other countries and distributes measurement and calibration procedures.

One of NIST's principal services is transferring measurements from its standards to other measurement systems. The procedure used to transfer the measurements is called a "calibration" procedure, and it is intended to bring other systems into agreement with the NIST measurements. The transfer process usually involves a hierarchical system of transfers, where each level relies on its own system of standards. At the highest level of the hierarchy is the *national standard*, held by NIST or other organizations (e.g., Los Alamos Scientific Laboratory) for NIST. Measurements are transferred from the national standard to the next level standard, which is called a *primary standard*. The primary standard has to be certified by NIST using its state-of-the-art calibration procedures. Once it is created, a primary standard may be obtained from NIST by any organization for any purpose. A primary standard is the direct link between the organization that owns it and the national standard owned by NIST.

The primary standards can be used to calibrate other measurement systems. However, primary standards are usually too expensive and too vulnerable to wear and tear for routine calibrations. Thus, a *secondary standard* is created by transferring measurements from the primary standard. Transfers to secondary standards can be performed by any organization with access to a primary standard. However, for the secondary standard to be traceable, the transfer must be performed with an appropriate calibration procedure.

From secondary standards, measurements may be transferred to another level of standard called *working standards*, which are usually used to calibrate the measurement systems found in production facilities. Working standards are also called *production standards* because of the direct link to production usage. Any standards in the lower hierarchy that can be connected back to the national standards through the proper use of calibration procedures are said to be traceable to NIST. Thus, according to Military Standard 45662 of the US Department of Defense, traceability of a standard is "the ability to relate individual measurement results to national standards or nationally accepted measurement systems through an unbroken chain of comparisons."

In general, standards become less accurate the further they are from the national standard. On the other hand, the standard becomes more robust to changes in its environment, thus making it cheaper and easier to maintain and use. Figure 1.2 is an example of a traceability chain that illustrates a kilogram standard's hierarchical levels and its relationship of precision trend.

Unfortunately for the semiconductor industry, because it deals with very-small-scale dimensions in micrometers and nanometers, there is no national standard that exists in NIST yet. The organization has been working to develop standards for the IC industry; however, it is an extremely challenging

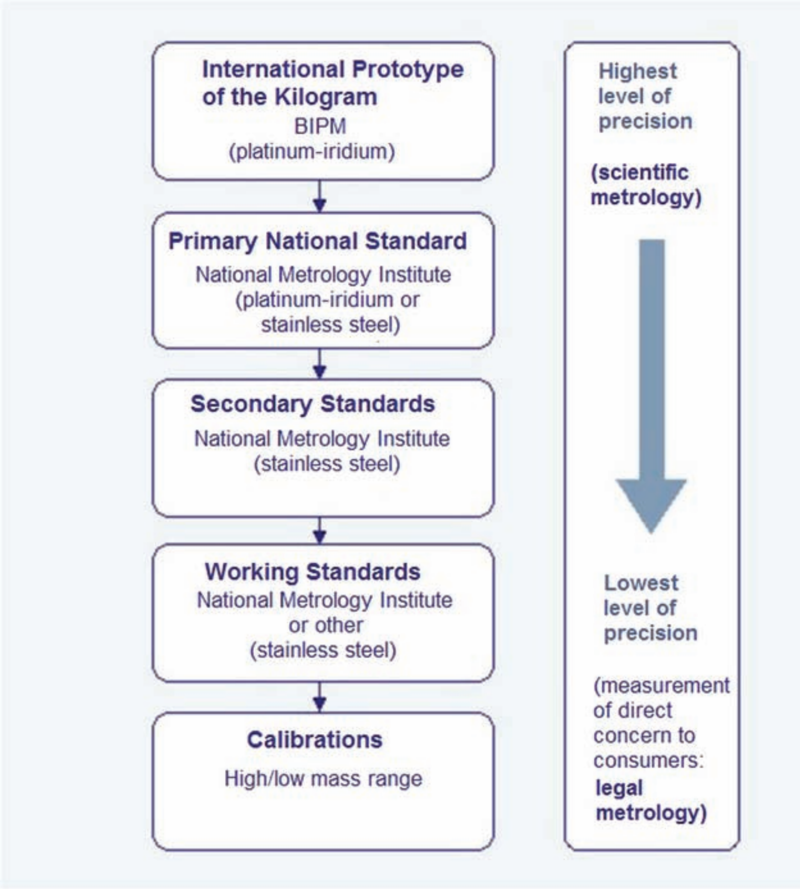


Figure 1.2 The kilogram standard hierarchical level and traceability flow (image used with permission from the BIPM).

task to manufacture and calibrate such small-scale (nm) objects from a limited material selection, and the IC industry has progressed so quickly (reducing the minimum features from 250 nm to 20 nm in less than 15 years) that it has left no time to develop standards. Fortunately, for semiconductor manufacturing process control, accuracy is usually not as important as precision, sensitivity, and matching. This point is elaborated further in Chapter 3.

1.6 Metrology Standards Related to IC Manufacturing

NIST has developed some standards related to IC manufacturing. For example, it has a SEM magnification calibration standard and other standards (only a few are listed here). SRM2059 was an old-generation standard for mask registration and has been published, but it is no longer offered by NIST today—it was replaced by SRM5001.

For example, a company called VLSI Standards, Inc. has specialized in metrology standards for the IC industry and other high-tech industries for the last 30 years. Their standards, which are traceable SI units of NIST, focus on the following areas:

- contamination standards for defect inspection,
- film thickness standards (such as silicon oxide and silicon nitride),
- dimensional standards for AFM/CD SEM [such as NanoCD, NanoCD for mask, NanoLattice (pitch standards), NanoLattice for mask, and step-height standards using quartz], and
- electrical standards, such as resistivity.

1.7 Measurement Methods and Process

Measurement involves a set of operations with the objective of determining the value of a quantity. A measurement result is the obtained value of that quantity acquired by a measurement system. Note that it is a *characteristic* of an object that is measured, not the object itself. A measurand is a specific quantity subject to measurement. Thus, a measurand consists of a measurement object, a quantity of the object (e.g., its length), and the measurement conditions. The first step of a measurement defines the measurand. The defined measurand often affects the accuracy of the measurement result; thus, all factors that influence the measurement process should be considered. This book uses the terms "measurement target," "specimen," and "measurand" interchangeably, with the assumption that the measurement conditions are pre-set.

1.7.1 Operator, measurement system, and measurement target

A measurement process normally consists of an operator that uses a measurement system (MS) to conduct a measurement on an object (a specimen or a measurement target). Therefore, an operator, a measurement system, and a measurement target form the basic elements of a measurement process.

An operator or an appraiser conducts the measurement. When different operators conduct the same measurement, any measurement error due to only the operator switch is called operator or human error. It has long been recognized that human error is usually one of the largest variations in measurement precision for many measurement systems. After the computer was introduced to modern complicated measurement systems (see Section 2.2), semi-automated and fully automated measurement systems have been developed and widely adopted in industrial manufacturing facilities. The measurement processes are now conducted by computer-controlled measurement systems to eliminate operator-related measurement errors. In a semi-automated system, an operator guides a machine to a measurement target; the machine collects the measurement signal by using a measurement algorithm to conduct a measurement, which eliminates the effect of the operator's judgment when determining a measurement value. An automated measurement system will find the measurement targets, conduct measurements, and output a measurement report automatically without operator assistance. Even though the operator error is largely eliminated in automated measurements, the operator's role in the measurement process still exists. Instead of directly taking a measurement, an operator's role (e.g., human decision making) involves constructing computer programs (usually called recipes) that dictate how to find a measurement target, how to measure it, and how to report the measured data set. Operators help perform any set up necessary for an automated measurement system to start. Automated measurement systems conduct the measurements with a preprogrammed computer to control the measurements; humans (engineers, technicians, and operators) are still needed to program the measurement systems' computers to make automation possible and react to tool or recipe issues when they arise.

The measurement process is part of a generic, measurement-data-driven decision-making process that can form a feedback loop. It normally goes through the following steps, starting with an instrument:

- measure,
- record data,
- analyze (statistical, plotting, comparing),
- conclude, and
- act,

which can be simplified as three phases—data collection, data to information, and information to decision. Data collection is accomplished by a measurement process; data to information is accomplished by data analysis, including statistical analysis, data trend plotting, and specification comparing; and information to decision is accomplished only by corresponding engineers and operators using the information from the previous step.

Measurement systems can be simply a ruler or a complicated, computer-controlled, automated measurement system, such as a CD-SEM. This book focuses on applications of the latter in high-volume industrial manufacturing, particularly semiconductor IC manufacturing.

1.7.2 Manual measurement versus automatic measurement

In a manual measurement process, an operator is in direct control of the entire process, from measurement-system selection, to calibration (if necessary), to operation and data generation. Human error can be introduced in any of these steps: improper MS selection (e.g., the measurement is at the end of the MS range), unnecessary calibration, improper data reading (e.g., taking too few eligible digits), accidentally recording the wrong data readings, etc.

Eliminating human error has been a consistent goal in measurement system development and progress in recent decades. Properly trained metrology engineers and technicians have a better understanding of particular measurement systems and can minimize some of the human errors; however, errors will always exist in manual measurements.

From manual measurements to semi-automatic and fully automatic measurements, the breakthrough of measurement systems has only been possible after advances in computer technology and system control technology. Consider a wafer CD-SEM as an example—the system must be able to

- automatically detect wafers in their cassette or other wafer container, such as a front open utility pod (FOUP);
- automatically transport a wafer from/to the cassette to/from its load lock chamber (a quick-pump-down chamber to maintain low pressure to the main chamber in high vacuum);
- align the wafer using its notch in a fixed orientation because of loading errors;
- load/unload a wafer to/from the loading lock chamber to the wafer stage inside the main vacuum chamber;
- perform coarse and fine alignment with less than a 0.1-μm error;
- find a measurement target as small as 20 nm using a pattern-recognition technique;
- auto-focus to achieve good focus in the area of the measurement;
- center the measurement target and scan the measurement area with predefined beam conditions; and
- take an automatic measurement with a preselected algorithm and move to the next target.

The computer program that controls the entire CD measurement process in a CD-SEM is called a recipe, which must be created and trained using a production wafer. New recent capabilities (called waferless recipe writing) allow the recipe to be built without using a production wafer before the production lot arrives, thus eliminating the tool time needed to build the recipe and the delays to production lots while they wait for a recipe to be built. A robust recipe is the critical part of automated CD-SEM systems.

Another advantage of automated measurement systems is the built-in data analysis capability. The computer in an automated measurement system is uniquely suited for statistical analysis of collected measurement data and can display plotted charts on its control screen with nearly no additional time. A decision can be made immediately, making the data-collection-to-decision step smooth and seamless.

Automated measurement systems not only eliminate human errors from measurement results but also increase measurement speed drastically and streamline the manufacturing processes. They free people on the manufacturing floor from boring and repetitive tasks. Automation in processing and in measurements is one of the greatest achievements in 20th-century high-volume manufacturing, and IC manufacturing exemplifies this achievement.

1.8 Applications of Industrial Metrology

Metrology in industry applications can be summarized by the following seven areas:

1. **Product/process equipment characterization (tool control):** Metrology and inspection equipment is routinely used to ensure that product/process equipment meets performance specifications at initial qualification, during routine use, and after maintenance and shutdowns.
2. **Product/process development:** Early product/process design and development phase and pilot production. Heavy measurements (oversampling) and quick data analysis are typically the characteristics involved in this stage. Product/process changes are allowed, and the impact of changes is quickly evaluated using metrology/inspection. For certain process development (e.g., etch process in IC manufacturing), destructive metrology must be used (breaking wafers for a cross-section SEM to access etch profile), and specific techniques such as DOE (design of experiments) are utilized to minimize material cost.
3. **Inspection of raw materials and parts:** Check that new materials and new parts meet predefined specifications (usually called "incoming inspection"). For example, for IC wafer fabs, incoming wafers are checked for surface flatness and cleanliness (particles).
4. **Product manufacturing:** At various stages of manufacturing processes, the selected (sampling) products from the overall production pool are measured and inspected at selected process steps before they go to the next step. This is called in-process inspection (product control). The product moves to the next step only if the measurement/inspection results conform to the specifications; otherwise, corrective actions are needed to bring the "out of spec" results to "in spec."
5. **Process monitoring:** One of the key applications of metrology, process monitoring reduces the sampling plan to detect key variations compared to the process development stage, partially because in the production stage, the process is relatively stable, and both quality and productivity are considered. For process monitoring, statistical process control (SPC) and other techniques based on SPC are extensively used to control the manufacturing process to ensure the product quality (or yield in IC manufacturing. Yield is the ratio of the good product produced over the entire product produced within a specified time period.). Statistical analysis of the measurements of certain characteristics of the product

drawn from the process is performed after each monitoring step and plotted in the SPC control chart. The average and standard deviation (mean and sigma) are typically plotted in the SPC control chart and compared to previous statistical data points. The specification limits for mean and sigma are typically also plotted in the SPC chart, so that any out-of-spec points can be clearly seen. Any drift or shift trends of the mean and/or sigma can be also easily spotted.

6. **Product maintenance:** Metrology tools are routinely used in production equipment maintenance and process corrective maintenance, when such actions are required.
7. **Model building and model verification:** Whenever process simulation is needed (e.g., OPC modeling), a process model must be calibrated, and a lot of model calibration and model verification data (can be orders of magnitude higher data volume, as compared to process development) must be collected by metrology tools. This is particularly true for the IC industry; fortunately, only one model is needed for a given process. For modeling purposes, the model accuracy directly depends on the data quality (i.e., the measurement uncertainties), which is why sometimes multiple measurements of a single gauge are desirable to average out "data noise."

References

1. "Metrology," Wikipedia, http://en.wikipedia.org/wiki/Metrology.
2. International Bureau of Weights and Measures, http://www.bipm.org/en/home, http://www.bipm.org/en/convention/wmd/2004.
3. National Institute of Standards and Technology, http://www.nist.gov/index.html.
4. VLSI Standards, Inc., http://www.vlsistandards.com.
5. U.S. Patent and Trademark Office, http://www.USPTO.gov.
6. Automotive Industry Action Group (AAIG), *Measurement System Analysis*, 2nd ed., Daimler Chrysler Corp., Ford Motor Co., and General Motors Corp., Distributed by AAIG, Southfield, MI (1995).
7. D. J. Wheeler and R. W. Lyday, *Evaluating the Measurement Process*, SPC Press, Knoxville, TN (1989).
8. H. Xiao, *Introduction to Semiconductor Manufacturing Technology*, 2nd ed., SPIE Press, Bellingham, WA (2012) [doi: 10.1117/3.924283].
9. K. M. Monahan, Ed., *Handbook of Critical Dimension Metrology and Process Control*, SPIE Press, Bellingham, WA (1993).
10. G. E. P. Box, W. G. Hunter, and J. S. Hunter, *Statistics for Experimenters—An Introduction to Design, Data Analysis, and Model Building*, John Wiley and Sons, New York (1978).
11. S. K. Kimothi, *The Uncertainty of Measurements: Physical and Chemical Metrology: Impact and Analysis*, American Society for Quality, Milwaukee, WI (2002).

Chapter 2
Metrology Fundamentals: Measurement System Characterization and Calibration Using Traditional Definitions

Many previously published documents exist describing how to perform measurement system analysis (also known as a gauge study). Some of the more prominent examples of this are the SEMI standard document SEMI E89-0707, a NIST technical note titled "Guidelines for Evaluating the Uncertainty of NIST Measurement Results,"[2] and the International Technology Roadmap for Semiconductors (ITRS) Roadmap for Metrology, which is published every year. This chapter introduces some basic concepts and summarizes some of the major elements covered in these documents. (Chapter 3 discusses limitations of the prior art and the need for addressing these limitations by creating new methodologies that are more appropriate for the IC industry today.)

2.1 Introduction

Two words are frequently used to discuss measurement data—precision and accuracy. What are the meanings of these two words and what is their relationship? Figure 2.1 illustrates a frequently used set of diagrams. A "precise" measurement means that the measurement data are close to each other. An "accurate" measurement means that the measurement data are close to the "true" value of the measurement target, which is a known reference. However, to fully calibrate a measurement system, other performance characteristics of a measurement system beyond precision and accuracy are needed to form the metrics.

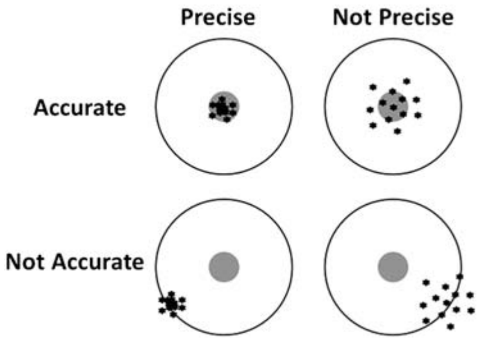


Figure 2.1 Illustration of "precise" and "accurate."

In modern manufacturing facilities, such as a semiconductor IC fab, automated measurement systems dominate manufacturing lines. It is critical to know the measurement systems' performance metrics—its precision (including repeatability and reproducibility), accuracy (or bias), stability, linearity, and throughput (related to measurement speed). Before any measurement system is put into a manufacturing line, it must be characterized and calibrated to qualify it to measure product. Such characterization and calibration are performed periodically to ensure measurement systems' validity.

All of the measurement-system performance metrics (precision, accuracy, stability, and linearity) are based on statistics and extracted following established characterization and calibration procedures. These procedures provide the general guidelines, but for certain types of measurement-system classes, the generally accepted procedures require modification due to either the interaction of the measurement tool and the sample or changes in the measurement sample over time due to the environment. For example, the relatively strong interaction between SEM-based CD measurement systems and their measurement targets—charging and carryover-induced CD variations—limit the number of measurements that can be performed, and periodic discharge and cleaning of the calibration artifacts are needed (see Chapter 4 for more details). This factor also complicates the mathematics: what part of the variability is measurement error, and what part is sample damage/changes?

There are normally two components in measurement variations—systematic and random variations. Systematic variations (e.g., variation with temperature changes) are generally well calibrated and well understood; thus, they are eliminated or minimized to a negligible level before any measurement instrument is put to use. In many cases, the primary goal is reducing random variation in measurement systems. Random variations of a given measurement data set generally follow a Gaussian distribution (a bell-curve normal distribution), as shown in Fig. 2.2. The normalized Gaussian distribution directly predicts the probabilities of any reading x, and the cumulated

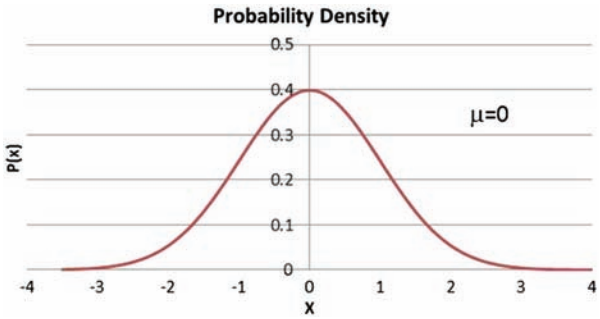


Figure 2.2 Typical Gaussian distribution with $\mu = 0$ and $\sigma = 1$.

probability distribution between two given values is just the integration of the probability function.

Gaussian distribution can be expressed in the following equation:

$$P(x,\mu,\sigma) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left[\frac{-(x-\mu)^2}{2\sigma^2}\right], \tag{2.1}$$

where μ is the mean (or mode), and σ is the standard deviation. The former determines where the peak of the distribution is, and the latter determines the width of the distribution. For symmetric distributions, such as the Gaussian, the mean (or average) is the same as the mode (the most probable value).

The cumulative distribution (probability) between any two given values x_1 and x_2 can be expressed as

$$C(x_1,x_2,\mu,\sigma) = \frac{1}{\sqrt{2\pi}\sigma}\int_{x_1}^{x_2}\exp\left[\frac{-(x-\mu)^2}{2\sigma^2}\right]. \tag{2.2}$$

For $x_2 = -x_1 = \sigma$, $C = 68.2\%$, which means that the cumulative probability of a value that falls between $-\sigma$ and σ is 68.2%. Pushing x_1 and x_2 farther away from the center (or the mean value) will cause the cumulated probability to rapidly increase. For example, for $x_2 = -x_1 = 2\sigma$, $C = 95.4\%$; for $x_2 = -x_1 = 3\sigma$, $C = 99.7\%$. The probability of a measurement value falling between -3σ to 3σ is 99.7%, and it has only a 0.3% chance of falling outside that range; thus, in reality, a process' control limits are normally set at $\pm 3\sigma$.

The symmetrical nature of the Gaussian distribution means that instead of specifying two boundaries, such as $-\sigma$ to σ or -3σ to 3σ, only one boundary is usually specified, such as one σ or three σ (which, for cumulative probability, means between $-\sigma$ and σ or -3σ and 3σ, respectively).

2.2 Precision

Precision is how statistically close the repeated measured values are to each other. Precision measures a measurement system's statistical variability when measuring the same specimen multiple times. It usually has two independent components: repeatability r and reproducibility R, and it is represented in the following equation:

$$\text{precision} = \sqrt{r^2 + R^2}. \tag{2.3}$$

Definitions for repeatability and reproducibility are classified in two categories—measurement systems with operator error (where an operator conducts measurements), and automatic measurement systems without operator error. In recent decades, automated measurement systems have dominated product manufacturing facilities; as such, the definitions should reflect such changes.

For a measurement system with operator error,

(a) **Repeatability** is the variation in measurements obtained with *one measurement instrument* when used several times by *one appraiser* while measuring the *identical characteristic* on the same part.
(b) **Reproducibility** is the variation in the measurements made by *different appraisers* using the *same measuring instrument* when measuring the *identical characteristic* on the same part.

For automatic measurement systems, the errors due to appraiser(s) are eliminated. Therefore, the definitions are modified to focus on measurement condition changes instead:

(a) **Repeatability (or static precision)** is the variation in measurements obtained with an *automated measurement instrument* repeating multiple times while measuring the *identical characteristic* on the same part without any measurement condition changes.
(b) **Reproducibility (or dynamic precision)** is the variation in the measurements made by an *automated measurement instrument* when measuring the identical characteristic on the same part with a change in the measurement condition.

For example, in an automated CD-SEM case, the test wafer is aligned after it is loaded into the measurement chamber, and then a measurement target is located so that repeated measurements can be performed for the predefined measurement sites. Its repeatability is determined through such repeated measurements. To test reproducibility, the reference wafer must be unloaded and reloaded to change the measurement conditions (due to variations in wafer alignment and measurement location finding) to reproduce the previous measurements. The time between measurements can also vary. Short periods between measurements are typically described as short-term precision (on the order of seconds to minutes between repeated measurements). Long periods between measurements are typically described as long-term precision (hours or days between repeated measurements). The timeframe between repeated measurements is chosen based on the behavior of

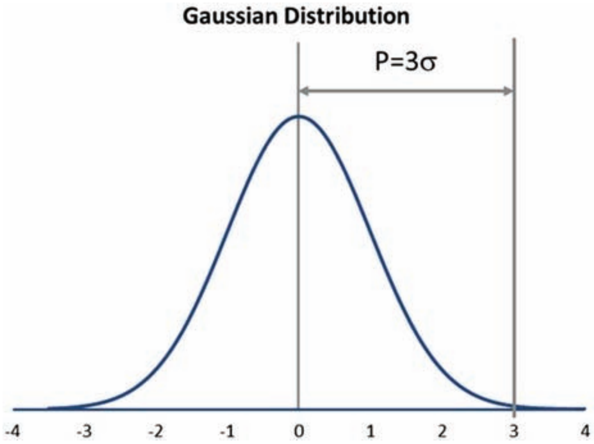


Figure 2.3 Illustration of precision (3σ) in a Gaussian distribution.

the tool the practitioner wants to characterize. For example, if the practitioner wants to know how the measurement system varies over the course of a week, then the precision would be described as long term. Using a Gaussian distribution, a measurement system's precision is directly related to its sigma (the width of Gaussian distribution), as illustrated by Fig. 2.3.

2.3 Long-Term Stability

Stability measures a measurement system's variation over an extended period of time. The time span for stability must be relatively long, measured in weeks, or even months and years. Any measurement systems that are subject to environment impacts (e.g., temperature variations, humidity variations, pressure variations, surrounding sounds and noises, ground vibrations, etc.) must be carefully calibrated for stability. In some cases, measurement systems have to be housed in tightly controlled environments to minimize environmental impact, such as temperature and humidity. Modern semiconductor manufacturing facilities (IC fabs) are well-controlled environments, particularly with respect to temperature, relative humidity, and cleanness. The IC production floor is usually called a cleanroom for that reason. The air in cleanrooms has been filtered using HPA filters to remove airborne particles until it reaches specified cleanliness classes; for example, cleanroom class 100 means there are fewer than 100 airborne particles of size 0.5 μm or larger in one cubic foot (U.S. customary measurement is used here). Most metrology and inspection systems, as well as process systems, have well-isolated foundations to minimize the impact of floor vibration in IC fabs. An individual metrology system may have other forms of protection to isolate environmental impact. For a CD-SEM system, electromagnetic shielding is necessary to minimize the influence that the

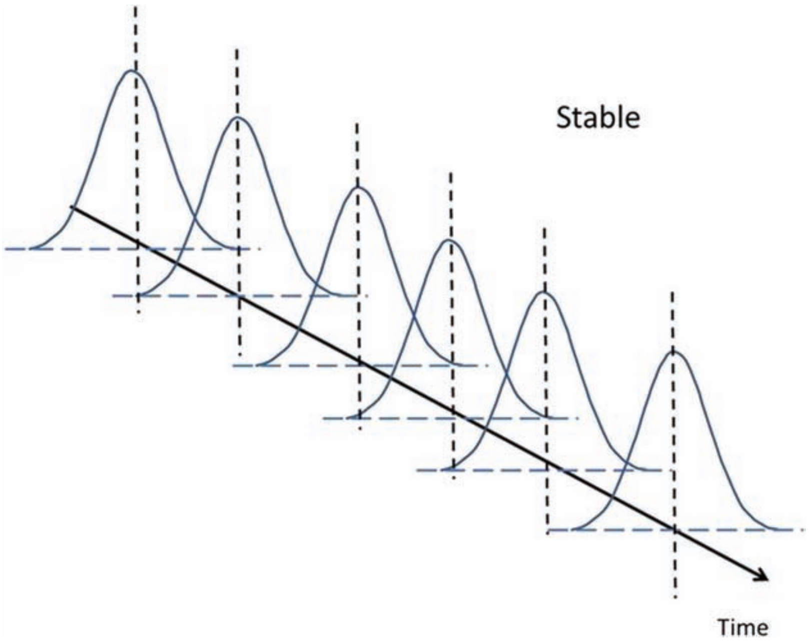


Figure 2.4 Illustration of a stable measurement system behavior.

environmental electromagnetic field has on a scanning electron beam and its generated signal.

Other factors that may impact metrology systems' stability include individual part lifetimes. For example, for optical-microscope-based metrology systems, light source stability can affect the measurement result. For a CD-SEM, a typical electron beam source (thermal emission field effect filament) has a lifetime of one year. At the beginning of its life and near the end of its life, the electron flux from the source is normally lower than its stable value, and the signal strength will be different, which changes the measurement results. A stable measurement system would show near-identical measurement results from a fixed calibrated specimen (both the average and the spread), as shown in Fig. 2.4. An unstable system, on the other hand, may show the average shifts with time and/or the measurement spreads change, as illustrated in Fig. 2.5.

2.4 Accuracy

Like precision, accuracy is a statistical concept. It measures how close a measurement system's average measurement value compares to the "true" value of a given measurement specimen. Figure 2.6 illustrates a measurement system's accuracy (or bias).

To calibrate a measurement system's accuracy, a known "true value" from a traceable reference is needed. Multiple measurements are also needed

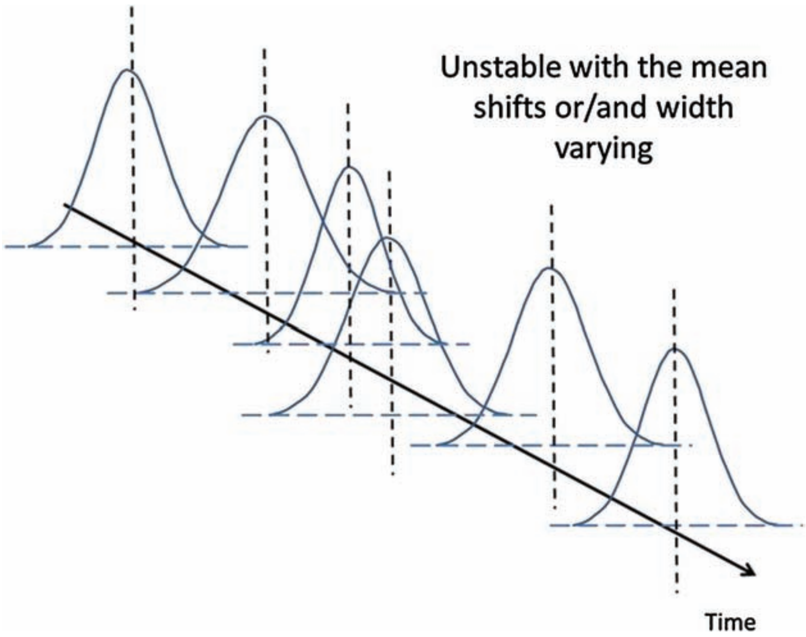


Figure 2.5 Illustration of an unstable measurement system behavior.

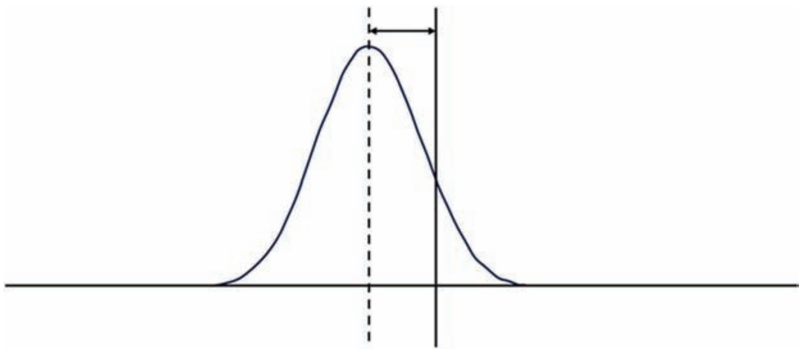


Figure 2.6 Illustration of accuracy or bias—the difference between the average measurement value (dashed line) and the true value (solid line).

to determine the average value of the measurements. Because the accuracy of a stable measurement system is expressed as a single number (which is partly why it is referred to as the bias), it is relatively easy to correct compared to the precision term because it is systematic.

Any given measurement system has a measurement range. For example, an optical-microscope-based dimension measurement system can measure from a few microns (limited by its optical illumination) to several hundreds of micrometers with different magnifications through different objective lenses. With such a wide measurement range, the accuracy must be assessed through multiple references to cover the entire measurement range. The overall measurement system accuracy over its whole measurement range is called linearity.

2.5 Linearity

Linearity refers to a measurement system's accuracy at multiple reference values. The linearity normally defines a measurement system's measurement range. Beyond that range, the measurement accuracy and linearity will deviate beyond the predefined measurement tolerance.

Linearity is important when measuring a characteristic of samples across a wide range. For example, in the IC industry, a CD-SEM routinely measures dimensions larger than 500 nm and smaller than 20 nm across more than an order-of-magnitude measurement range. An accurate measurement at 100 nm by a CD-SEM does not mean it can have an accurate measurement at 10 nm unless its linearity has been calibrated in its entire measurement range.

The typical representation of a measurement system's linearity is a simple linear fit between its measurement averages of multiple reference standard specimens versus their known "true" values. In case such reference standards do not exist, a calibrated reference measurement system can be used as the "reference standards." Figure 2.7 illustrates typical linearity plots of the test measurement system (TMS) and the reference measurement system (RMS). Three typical cases are illustrated in Fig. 2.7: ideal linear (with a slope of 45 deg and no offset; fixed offset is easy to correct), linear with slope (not at 45 deg), and nonlinear.

Keep in mind when a RMS is used as a "standard" that it has measurement uncertainties. When comparing the characterization data of a TMS to a RMS, the measurement uncertainties from both systems must be

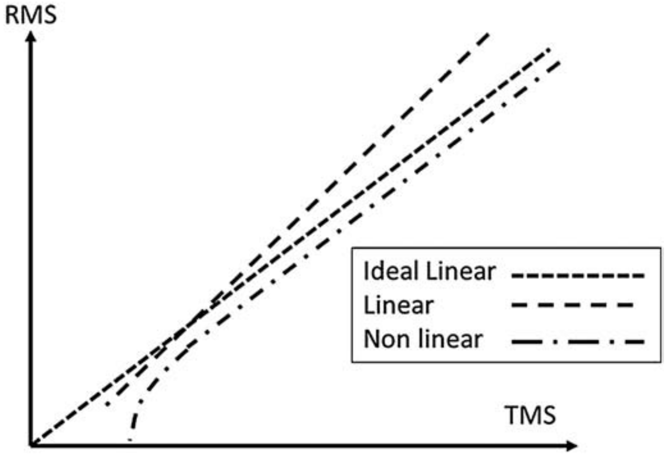


Figure 2.7 Illustration of measurement linearity versus a reference system (TMS).

considered, and an alternative analysis method must be used. Chapter 3 presents a regression technique that takes into account the uncertainty from both systems.

2.6 Tool-to-Tool Matching: Multiple Measurement Systems

For high-volume production facilities, such as IC fabs, multiple metrology systems are required to support a high level of production. To maximize productivity and working flow efficiency, it is a common practice to direct any product that requires a metrology check to any similar available metrology systems. Dedicated metrology systems for specific-process metrology checks are highly discouraged and usually prohibited. To ensure the validity of all measurement results from all measurement systems performed on the same parts, the entire fleet of same metrology systems must be matched using the same standards or reference artifacts, i.e., tool-to-tool matching. Any new metrology system that is added to the fleet must undergo the same matching procedure.

The composition of a metrology tool set determines the tool-matching procedures and reference tool standard. Ideally, the tool set consists of tools from a single manufacturer and model—all tools behave similarly. In reality, tools of different models (e.g., older- and newer-generation tools) and even from different manufacturers are often in the same tool set. This scenario complicates the tool-matching procedure; in certain cases, some tools must be excluded from the matching tool set due to hardware and feature differences/limitations.

Metrological tool matching essentially examines how well various tools agree with each other. A typical matching procedure involves selecting a "golden" tool (which serves as a reference), and all other tools match the golden tool or are compared to a fleet average. The golden tool should generally exhibit the best precision and accuracy on the matching specimen, but sometimes this is not the case due to a legacy issue. Oftentimes, the older-generation tool was the tool of choice and established as the in-house tool standard; a newer-generation tool with better precision and/or accuracy must match the in-house standard tool first. A new golden tool can be established with additional calibration work.

Before tool matching is performed, a matching specification must be established first. This process requires that all measurement systems produce the same measurement within a specified tolerance (matching specification). Figure 2.8 illustrates what a tool-to-tool matching specification should look like. It includes two tools with slightly different distribution widths and means. A match means that both Gaussian distributions are within the matching specification.

Tool-to-tool matching is a relatively new concept in industrial metrology, and it is not well covered in the existing metrology books. There are two ways

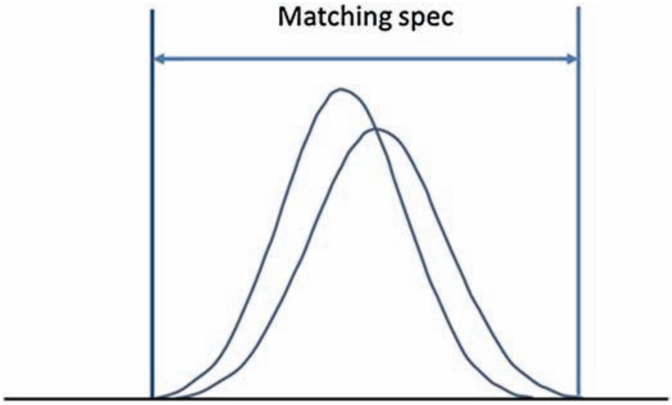


Figure 2.8 Illustration of matching specification for two tools.

to handle the tool-matching uncertainty. One way adds the tool-to-tool matching variation in the total measurement uncertainty as a new term. Another way introduces a new tool set precision—like single-tool precision, a tool set precision (called fleet measurement precision, or FMP; see next chapter) accounts for the overall tool set—a composition of all individual tools in the set. Figure 2.8 illustrates tool set precision. An extra term must be added in addition to the golden tool when calculating tool set precision.

The importance of tool-to-tool matching has been long recognized in the IC industry, and it has become a part of tool-selection criteria for both process equipment and metrology and inspection systems. The tool-matching requirements have been shifted to take place earlier in the equipment design and development phase and make it easier to install them on the manufacturing floor. New ideas and concepts regarding tool-to-tool matching are discussed in Chapter 4.

2.7 Classifying Components of Uncertainty

According to NIST, the uncertainty of the results of a measurement generally consists of several components that, via the Committee for Weights and Measures (CIPM) approach, may be grouped into two categories according to the method used to estimate their numerical values:

- **Type A:** Evaluated by statistical methods, such as making repeated measurements and statistically summarizing the results.
- **Type B:** Evaluated by other means. Without the ability to perform a Type A analysis, the measurement uncertainty is assessed in other ways, such as using the literature to estimate the uncertainty for similar situations or using prior experience.

Type A measurement-uncertainty determinations are preferable whenever possible.

2.8 Measurement Errors and Measurement Uncertainties

As stated earlier, the measurement process is a statistical process, and the measurement data must be analyzed using statistical methods (Type A). Measurement errors and uncertainties are also statistical in nature. All measurements have built in uncertainties (or variations), including the measurement equipment, the operators who perform the measurements (in manual measurement process), environmental factors, etc., with no exceptions. Associated with measurement uncertainty is its statistical probability—the likelihood that the measurement results resemble the true values. To make meaningful measurements and obtain meaningful measurement results, the measurement uncertainties must be carefully studied and controlled at every level of the process. Measurement variations cause measurement errors, of which there are two kinds: random and systematic errors.

2.8.1 Random errors

Random errors are the dominant kind in most measurement processes. If and only if random errors exist in a measurement system can a reliable measurement be achieved (provided they are below a certain tolerance). Random errors cannot be eliminated, but they can be reduced or minimized by careful measurement-system design, manufacturing, and automation, environmental control, and statistical sampling. A measurement system's precision is the metric used to capture the total random errors. A similar thing can be said about process control, except that when random errors exist in a process, the process is said to be in a state of statistical control (the resulting products are manufactured with an acceptable level of quality).

As mentioned earlier, random variations are described by Gaussian distribution. In most industrial metrology, Gaussian distribution is widely used in all statistics-related measurement setup and result analysis.

2.8.2 Systemic errors

Single-measurement-system accuracy (or bias) and tool-to-tool offset are typical examples of systematic errors that must be carefully calibrated and minimized to conform to predefined specifications. Systematic errors should normally be absent or at least minimized from any measurement systems in working condition. They are relatively easy to correct and eliminate in general, unlike random errors; thus, such an error component should not be present in normal measurement conditions. Whenever systematic errors start to appear in measurement results, one of two things has occurred, and corrective actions must be implemented immediately: either the measurement system and/or the measurement condition have been changed, or the manufacturing process has been shifted, e.g., the mean value of a characteristic of a SPC-chart product is constantly shifted up.

In other cases, strong interaction between measurement systems and measurement specimen can cause systematic measurement errors. We will describe such interaction in details in Chapter 4. When such an interaction occurs, extreme care must be taken to minimize the interaction.

2.8.3 Tool-to-tool variations

As we discussed in Section 2.6, tool-to-tool variations occur when multiple measurement systems in a metrology tool set perform same or similar measurements in the same manufacturing facility. These variations include both systematic biases and random errors. As a general rule, it is much more preferable to remove tool-to-tool differences through appropriate calibrations rather than managing tool-to-tool differences by correcting the measurement, e.g., with an offset lookup table. Good metrologists find the root cause of the variation(s) and calibrate it out, which requires a thorough understanding of the tool set and how it behaves.

2.9 Uncertainties and Risks

There are one-way specification limits and two-way specification limits, based on the measurement requirements and the decision to be made. In the former, a decision is made by a single value of a parameter and whether a measured value is smaller or larger than a specification. Many examples in environmental-safety-related areas are one-way specification limits, e.g., a certain harmful substance must be less than the safety limit. Highway speed limits and the checked-luggage weight limit at airports are two common examples of one-way specification limits. Two-way specification limits are more popular in industry product/process control. Consider, for example, CD control in IC industry—a specific CD value in a lithographical process must be controlled within 10% of its nominal value. Furthermore, the measurement error must be much less than the process' tolerance. A general rule of thumb called the gaugemaker's rule states that the measurement error cannot exceed more than 20% of the process tolerance. Use this guideline when no other measurement error specification is provided.

In two-way specification limits, a measured value is controlled between two values typically described as the lower specification limit (LSL) and the upper specification limit (USL). In some cases, either the LSL or USL is set by its natural limit and does not have to be explicitly expressed. For example, in SPC, the sigma value cannot be smaller than zero; thus, zero is its natural LSL, and only the USL needs to be specified. As another example, in the dry etch process of semiconductor manufacturing, the nature of plasma etch makes it nearly impossible to achieve an etch sidewall profile larger than 90 deg; thus, this angle (vertical profile) becomes the natural USL for the

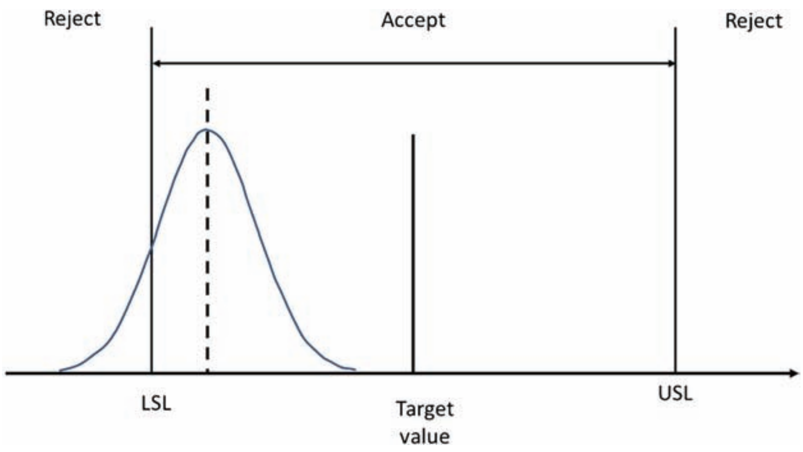


Figure 2.9 Illustration of probability of measurement uncertainty causing decision error.

Table 2.1 Type I and Type II errors.

	Actual Value (within Spec)	Actual Value (out of Spec)
Measured Value (within Spec)	No error	Type II error
Measured Value (out of Spec)	Type I error	No error

sidewall angle of an etch process, and only the LSL is specified in the etch-process profile control (e.g., no less than 85 deg).

Due to measurement uncertainties and the statistical nature of measurements, there is always the chance that the "actual" value of a monitored parameter of a product is within the specification but the measurement result says otherwise; even though such a chance is small, it is called a "false negative" or "alpha error" in statistics. Figure 2.9 illustrates such cases. When an "actual value"—represented by the dashed line—is close to the LSL, the measurement result may be outside the LSL due to measurement uncertainty (represented a Gaussian distribution). The area beyond the LSL of the left side of the Gaussian tail represents the probability of measurement uncertainty causing falsely rejected products. The opposite can also be true if it is assumed that when the dashed line in Fig. 2.9 shifts left beyond the LSL, the right tail of the Gaussian distribution still extends inside the LSL.

Table 2.1 illustrates the distinction between Type I and Type II errors. A Type I error will result in false rejecting a good part due to measurement uncertainty, whereas a Type II error will result in failure to identify a bad part due to measurement uncertainty. Both error types should be avoided by minimizing measurement uncertainties.

Similarly, when balancing detection sensitivity and inspection speed in the inspection domain, alpha and beta risks apply in a similar fashion as Type I and Type II errors, as illustrated in Table 2.2. Alpha risk is a "false alarm,"

Table 2.2 Alpha and beta risks in inspection systems.

	Actual Value (Good)	Actual Value (Not Good)
Measured Value (Good)	No risk	Beta risk
Measured Value (Not Good)	Alpha risk	No risk

which has a smaller cost-related impact compared to beta risk, which is "no alarm" or "miss detection." Alpha risk is less costly because the follow-up confirmation inspection would normally recover the false alarm and only delay the process. On the other hand, beta risk would continue the process with an assumed "good" product inline until future inspections catch it. The farther down the line that the defected products are processed, the more costly the process becomes. In a semiconductor IC process, the "missed" defects would be buried underneath the current layer, making it harder to recover from the inspection downstream all the way to the defective end products and resulting in lower yield. Thus, it is normal practice to tune inspection systems slightly more sensitive so that no misses happen, with tolerable false alarms.

2.10 Are These Traditional Metrics Enough?

Considering all of the traditional approaches and metrics used over the last 40 years, a question arises: are they sufficient for current and future metrology needs in the IC industry? Answering this question involves revisiting the measurement requirements. The measurement requirements can come from multiple sources. One source is the ITRS roadmap, which states the measurement uncertainty requirements for specific critical process steps, e.g., post-lithography CD and overlay or post-etch CD. Typically, a poly gate has the tightest process tolerances and therefore the tightest measurement uncertainty requirements. The supplier will also communicate the specifications that their tool set can handle. If there are no ITRS requirements or supplier specification for a given situation, then the gaugemaker's rule can be used. Because the measurement error consumes more than 20% of the process tolerance, the likelihood of detecting true process shifts decreases as it becomes more difficult to determine if the shift is a shift in the process or the result of measurement error. Bad decisions occur when this happens. Wafers that were measured could be falsely sent on (type II error) only to be found faulty later at test. Furthermore, if the measurement data is used for an APC, it becomes harder to converge to the desired result. Regardless of how the measurement uncertainty requirements are determined, what is not clear from the traditional metrics is how to determine which ones should be used to compare against these requirements. How does one assemble the data appropriately? This question is answered in the next chapter, as well as highlighting other gaps that appear when using traditional metrics and methodologies.

References

1. "Guide for Measurement System Analysis (MSA)," SEMI E89-0707, SEMI International Standards (1999).
2. B. N. Taylor and C. E. Kuyatt, "Guideline for Evaluating and Expressing the Uncertainty of NIST Measurement Results," NIST Technical Note 1297, The National Institute of Standards and Technology (1994).
3. ITRS 2013 Metrology section, http://www.itrs.net/ITRS%201999-2014%20Mtgs,%20Presentations%20&%20Links/2013ITRS/2013Chapters/2013Metrology_Summary.pdf.
4. Automotive Industry Action Group (AAIG), *Measurement System Analysis*, 2nd ed., Daimler Chrysler Corp., Ford Motor Co., and General Motors Corp. Distributed by AAIG, Southfield, MI (1995).
5. D. J. Wheeler and R. W. Lyday, *Evaluating the Measurement Process*, SPC Press, Knoxville, TN (1989).
6. G. E. P. Box, W. G. Hunter, and J. S. Hunter, *Statistics for Experimenters: An Introduction to Design, Data Analysis, and Model Building*, John Wiley and Sons, New York (1978).
7. J. L. Bucher, Ed., *Metrology Handbook*, 2nd ed., American Society for Quality, Milwaukee, WI (2012).
8. J. F. W. Galyer and C. R. Shotbolt, *Metrology for Engineers*, 5th ed., Cassell (1990).
9. S. K. Kimothi, *The Uncertainty of Measurements: Physical and Chemical Metrology: Impact and Analysis*, American Society for Quality, Milwaukee, WI (2002).

Chapter 3
Fundamental Metrology: Redefining Measurement System Analysis

3.1 Introduction

Why must measurement system analysis (MSA), i.e., gauge study, be redefined? The existing literature does not make clear which metric should be used to compare measurement system performance against the requirements or specifications. The requirements are straightforward to determine because they are either pulled from the Metrology ITRS roadmap, which typically specifies the requirements for only the most challenging situations, or they are derived from the target value or process tolerance of the given application. Expanding on the latter point, a good rule of thumb is to seek 1% of the target value or 10% of the process tolerance, ideally whichever is smaller. 2% of the target value or 20% of the process tolerance (whichever is smaller) is acceptable for automated process control, but continuous improvement should be pursued to achieve 1% of target or 10% of the process tolerance over time.

As for which metric to use to compare against the requirements, there are two scenarios to consider. For a single tool, the recommendation is to use the precision. For a fleet of n tools, where $n > 1$, the answer is not obvious—should each tool's precision be used? What about the offset between tools? This lack of clarity drove the creation of a new metric. It is desirable to represent a fleet of tools with one value that captures the measurement variation of the entire fleet; this value is called the fleet measurement precision (FMP), and it should be used to compare against requirements when $n > 1$. A fleet of tools is used to measure the various processes in most situations. Very rarely is one measurement tool used to control a given process except in certain circumstances, such as an integrated metrology module on a process tool, or when there is no other way to meet the requirements other than with a single tool. The FMP is the key metric used to describe the entire fleet, whereas the tool-matching precision (TMP) describes each tool in the fleet. TMP is a composite metric created from each tool's precision, offset, slope-induced-shift offset (SISoffset), and nonlinearity. This breakdown is different than the existing literature and previous chapter, which use terms like random and systematic variations but do not go far enough to better categorize the contributors. By categorizing the components in this manner, the assignable cause is easier to determine. The metrology toolset owner or supplier typically knows which "knobs" affect a given component(s) that is the dominant contributor. An offset issue is likely explained by a handful of "knobs," whereas a precision issue is likely explained by a different set of "knobs," and so forth. The TMP metric is a fundamental building block for improving the measurement performance of the fleet. It provides critical diagnostic information about where to focus efforts to improve the FMP, which is especially important when the requirements are not being met.

Another gap in the prior art applies to setting up control chart limits to monitor the long-term performance of the fleet. Conventional control-chart theory does not take into account the requirements that must be met. Calculations of the control limits using the conventional theory are based on a "stable" time period of, say, at least 30 data points for each tool. If the toolset is not meeting the requirement during this time period, then this error is built into the chart limits, and thus the likelihood of controlling the toolset to what is needed is low. A more-desirable determination of the control-chart limits considers the requirement that must be met and also takes into account the site averaging performed in the monitor. The averaging requires consideration because it reduces measurement error (as opposed to the measurement error of a single measurement point, as reported using the FMP); therefore, the control-chart limits must be rescaled accordingly (described by the central limit theorem[4]). In most fabs, the charted data is the average of many site measurements. This step is performed on a given monitor wafer running across all of the tools in the fleet at any point in time, measuring the same sites. This important breakthrough links the metrics used for measurement system analysis (precision, TMP, and FMP) with the control-chart limits used to maintain the toolset performance over time to ensure consistency between these independent events—more specifically, the measurement system analysis that assesses performance over a finite timeframe and the set-up of long-term control limits.

Accuracy (discussed further in subsequent sections) emphasizes relative accuracy through a metric called total measurement uncertainty (TMU). Ideally, absolute accuracy is most desirable, but it is sometimes difficult to achieve or unnecessary. The TMU metric was created to ensure that the variability of the measurement tool under test is properly assigned from a relative-accuracy perspective, given that the reference tool also likely has measurement variability. In the case of TMU, relative accuracy means that the offset between the tool under test and the reference tool is not factored into the equation. The offset is simply a fixed value, and if it were considered in the TMU equation (especially when it is large), it would dominate the metric and could lead to misinterpretation of the results. Not including the offset in TMU allows the metric to focus on how well the tool under test and the reference tool agree across the process variation.

In general, accuracy is not critically needed for process control. What is needed is good FMP and sensitivity to any process changes in order to keep the process centered. Accuracy is important for optical-proximity-correction model building. In this case, it is critical to ensure that the various CD-SEM measurements made in various patterning environments (e.g., isolated and dense areas) are free from bias. Therefore, the bias must be accounted for, such as by using an AFM, to minimize the impact on the final reticle build. The key is to be aware when accuracy is required and when it is not, based on the given scenario.

Two template spreadsheets with example data are provided on the accompanying CD: one for precision, TMP, and FMP analysis, and one for TMU analysis. These files allow either analysis to be performed with all of the calculations predefined; the user adds only the required raw data, and the rest is done automatically. They also ensure that everyone performs the analysis in the same way. Note that a special regression analysis called Mandel regression is used in the analysis. The regression considers the measurement error in both x and y axis variables that conventional least-squares regression (as in Excel) does not consider. The traditional metrics and methodologies are aligned with the new metrics and methodologies in Table 3.1.

Note that the ultimate reference for the measurements should be the electrical performance data. The inline measurements will ideally correlate with key electrical parameters to provide early detection of performance/yield issues. If this is not the case, the value of the measurement should be challenged unless it provides additional value, e.g., coming from automated process control. The things being measured should matter—a key consideration for metrology engineers.

Table 3.1 The relationship between traditional metrics and new metrics.

Category	Traditional Metrics	New Metrics
Single-tool variability	Repeatability and reproducibility (R & R)	Precision
Matching	Offset	FMP and TMP
Long-term stability	Control-chart limits calculated from conventional theory over a stable time period	Control-chart limits calculated from requirements and the averaging performed for the monitor
Accuracy	Offset, bias, linearity, and discrimination	TMU (relative accuracy)

3.2 A Metrologist's Core Activities

Before discussing the new metrics and methodologies, consider the core activities performed by a metrology engineer, as shown in Fig. 3.1. Each activity requires the use of a particular metric(s), as indicated by the possible criteria. All of the criteria mentioned under each activity are ideally used, but in reality, time constraints and resources may limit how many of the criteria are used in a given situation. The first box describes the process of bringing a new toolset to a fab. This process is typically executed at a given supplier's demo facility, or else a tool can be brought into the fab for evaluation to ensure that it meets key requirements before making a decision to purchase. Regardless, the typical criteria used would be precision, accuracy, and matching.

Once a decision is made to bring in a new toolset, the next step is to qualify the first tool. In this case, the primary criteria would be precision because there is only one toolset in the fleet at this point. Matching is sometimes used to compare against other similar types of tools already in the fab, e.g., a different model from the same supplier or a similar tool from a different supplier (this is defined as heterogeneous matching). After the first tool passes qualification, the long-term monitoring strategy is invoked to ensure that the tool is stable over time. More tools are then added to the fleet, and each must be qualified and monitored for long-term stability.

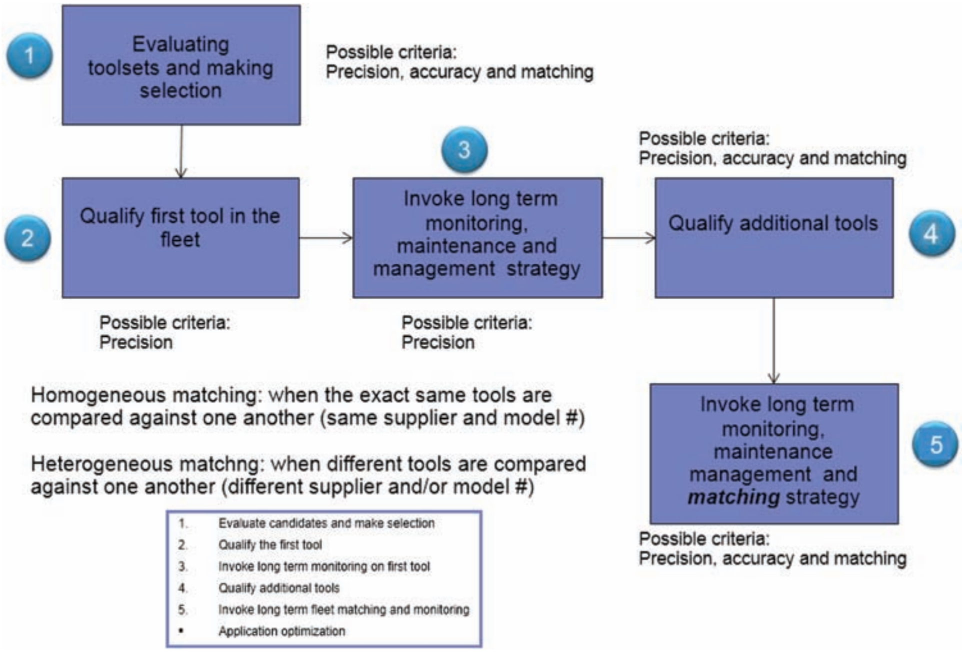


Figure 3.1 General list of activities and flow performed by a metrology engineer.

Each subsequent tool must match the other tools and is called homogeneous matching because all of them are from the same model and supplier. Each of the criteria listed for each step is critical to ensure that the tools meet requirements when released and over the life of the toolset.

3.3 Roadmap and Specifications

A roadmap is a document that describes the requirements that various metrology solutions need to meet. It is typically a function of technology node and the requirements listed are typically for the most demanding applications. This is published annually by the International Technology Roadmap for Semiconductors (ITRS). This document does not teach how to assemble the measurement data to compare it against the requirements; it simply communicates the needs. The following sections clarify new metric definitions and how to assemble the data so it can be used to compare against either the roadmap requirements, supplier specifications/requirements, or the needs of a given process step, as defined by the gaugemaker's rule.

The specification or requirement must be determined for the precision, FMP, or TMP metrics to be compared against. If there is no explicit requirement, then the requirement will generally be derived using the following methodology:

- The metrology tool(set) precision or FMP will not exceed 2% of the target value. For example, if the target CD for a 14-nm gate is 25 nm, then the precision or FMP will be no larger than 0.5 nm.
- Note that for overlay tool(set)s the target is zero, in which case 20% of the process tolerance is used. For example, if the overlay process tolerance for the 14-nm gate lithography is $\pm$10 nm, then the precision or FMP will be no larger than 2 nm.

3.4 Standards

Ideally, accuracy (or bias) compares measurements with known values on a NIST-traceable standard; however, this is often not possible. An International SEMATECH Manufacturing Initiative (ISMI) response (2) to the Automotive Industry Action Group (AIAG) MSA Guidelines states the following:

> In almost all meaningful cases, traceable semiconductor metrology standards are not available. The modern wafer fab's requirements have surpassed the traceability expectations described in the AIAG MSA. Today's fabs routinely employ many gauges costing millions of dollars, including state-of-the-art wavelength, interference comparators, and scanning electron microscopes. In most cases, reference standards simply do not exist. If reference standards were to be created, by the National Institute of Standards and Technology for

> example, the best gauges available to them at any price would be the very same gauges in hourly use in our member company wafer fabs. This eliminates the possibility of using MSA to ensure gauge accuracy. In any case, for most internal metrology needs, accuracy is least important. Only customer-measurable qualities need to be accurate, and for these cases we have and use traceable standards. Gauge matching is more important. The disparity between gauges used to measure the same characteristic must be minimal compared to the business decision limits used for product or process control.

Dimensional standards do not exist for the IC industry. Therefore, dimensional metrology systems such as a CD-SEM cannot be evaluated by the standard calibration procedure. A reference measurement system with nontraceable, manufacturer-specific standards is needed. Fortunately, the absolute accuracy (which needs a standard to calibrate) is not as important as relative accuracy (which does not need standards) because ultimately IC manufacturers care about device performance and device yield. So long as the measured CD values in IC process control have good correlation with the device performance and yield, then tightly reproducing those CD values in IC processes will produce good results. Additionally, because metrology tools are also used for process control, the measured values should be sensitive to the relevant process changes that need to be controlled.

The fact that IC device scaling follows Moore's law means that developing any CD standard is nearly impossible; however, it also has built-in tolerance for "absolute inaccuracy." Certified standards are desirable when they are available, but for the purpose of judging tool matching, golden parts, golden tools, or PSAs are sufficient.

3.5 Monitor Samples and Process Stressed Artifacts (PSAs)

A monitor is defined as a wafer or set of wafers used to ensure that a particular characteristic of the toolset maintains its desired performance. The characteristic of interest in this section is measurement stability. Monitor wafers are typically generated by removing them from a specific product lot at a particular process step and then placing them on a monitor route that continuously cycles at some frequency through the toolset being tested. The limitation of this method is that the removed wafers represent the process only at a single point in time. Meanwhile, other wafers from similarly processed lots moving through the fab at the same process steps will likely measure higher and lower. A better practice for generating monitor wafers incorporates this expected variability in one wafer or a wafer set, which are called process stressed artifacts (PSAs). When this wafer or wafer set is used to monitor the toolset, it better represents the process situations that the measurement tool is expected to see over time and serves as a more demanding and realistic test for the measurement toolset.

Most metrology tools are used for a wide variety of applications. It is not possible to monitor the performance of each one via monitors, so the following guidelines must be used:

1. Group similar applications together.
2. Rank the importance of application groups.
3. Create monitors for each of the most important groups to monitor all critical application groups with a minimum number of monitors.
4. Monitors should be PSAs whenever possible.

PSAs must be carefully chosen so that, ideally, all parameters of interest are monitored across the entire range of process variation. However, due to tool availability constraints, one must be careful to monitor the critical parameters in as few wafers/sites as possible. The order of preference when generating and using monitor wafers is as follows:

1. Single-wafer PSA, with intentional variation in one or more parameters of interest;
2. Multiple-wafer PSA, with intentional variation in one or more parameters of interest, changing wafer to wafer; and
3. Monitor wafer with normal process variation.

Some processes, such as lithography, can generate a wide range of process conditions on a single wafer, which allows for the preferred single-wafer PSA to be generated for toolsets such as a CD-SEM. Other processes, such as film deposition or etch, are not capable of generating variation across a single wafer, so toolsets such as ellipsometers must resort to one of the other options.

The following are the general guidelines for generating PSAs:

- Parameters of interest must be stable over time (feature dimensions or material composition). Unstable PSAs require additional work to monitor the true performance of the fleet, as well as periodic generation of new PSAs.
- Range of variation for parameters of interest should exceed the process tolerances by at least 10%, i.e., the artifact(s) should capture the entire expected process window expected over time.
- Data should be evenly distributed across the range of variation.

If it is not clear which parameters of interest a PSA must include and what the amount of induced variation should be, a set of wafers may be generated with broad variation across a multitude of parameters. Data can then be collected to determine which parameters correlate to true measurement problems and should therefore be monitored. This information can then be used to generate an improved version of a PSA that can become a sector monitor.

3.6 Test Vehicle Variability, Metrology-Induced Sample Damage, and Sample Stability

The analysis of a measurement system requires some critical components, such as (a) the measurement system, (b) test method, (c) test vehicle or PSA as defined earlier, and (d) specifications. Ideally, only the measurement system is an unknown variable, the test method and specification are not varied, and the test vehicle comprises a known variation. In the real world, the test vehicle becomes variable due to either natural processes or the act of measurement itself. The former is called sample instability, and the latter is called measurement-induced damage. Examples include (a) growth or loss of native films or components on the silicon wafer surface over time, i.e., oxidation, diffusion, etc., and (b) damage of films and structures due to e-beam charging or deposition, interaction with deep ultraviolet light, etc. that may, for example, shrink resists, deplete elements, or simply build charge. This behavior typically hinders the repeated measurement of the same site or structure, which is a fundamental requirement of the test method defined in a MSA. The analyst must be convinced that the error in measurement is not attributed to the measurement system but rather is associated with the test vehicle. Tests may have to be performed on multiple systems, wafers, and environments to conclude that the variability is not attributable to the tool under test.

The next step determines if the variability can be accounted for mathematically. Natural variations of the test vehicle will typically follow a simple function. Predefined analyses are available that account for simple trends in measured data induced by oxidation of the wafer film surface, for instance. Similar analyses have also been applied to cases where the e-beam from the measurement tool causes the elemental composition to vary systematically. Measurement-induced damage can also be avoided by dithering the measurement site so that no site is measured twice. The assumption here is that the spatial variation of the parameter being measured across the test vehicle is insignificant.

If the test period is relatively long, then natural variations of the test vehicle may appear unsystematic due to environmental changes. Such a scenario can be validated by conducting parallel tests on multiple systems. It is generally advised that such sensitive measurement analysis be done on more than one system where one of them is the tool under test (TuT). Note that the methods and examples described here are for non-measurement-induced damage or variability induced by natural processes. To cover these situations, special attention must be paid to the handling of the data.

3.7 Reference Measurement System

Despite the lack of relevant standards to calibrate inline metrology instruments, there is a strong and continuing need for calibration. Part of the solution to this problem involves constructing PSAs. The rest of the solution is an in-house reference measurement system (RMS), which is a combination of instrumentation and expertise that provides measurements with uncertainties consistent with the needs of semiconductor manufacturing and development. The need for accuracy, especially, comes at the expense of cost, speed, ease of use, and automation. The key attribute of a RMS is the intrinsic accuracy of the instrument. A workhorse instrument must be precise and fast, whereas a RMS must be accurate.

The TMU methodology (discussed further in the next section) calls for selecting an appropriate RMS. This is an application-specific decision. After the measurand (and thus the CD) is defined, then the correct RMS usually becomes obvious. One example is a system based on a cleaved cross-section. This traditional method sacrifices one or more wafers and often requires the customer (not a metrologist) to interpret the SEM image to determine dimensions. Furthermore, the sampling by cross-section is usually insufficient to assess the contribution from sample roughness. As a consequence, this RMS is often an inappropriate choice for calibration purposes.

Another example is a system based on the CD-AFM. Besides being a nondestructive measurement, this system can usually produce a sampling plan that is an "apples-to-apples" match for the TuT. For example, for scatterometry calibration and TMU assessment, this optical system makes a measurement that is an average linewidth across a grating that is many tens of micrometers on a side. The typical CD-AFM scans a region ~1 μm in extent. In order to provide a suitable measurement set, the sampling plan for the CD-AFM includes multiple measurements across the grating so that the grating average can be determined.

3.8 Precision

The most-basic analysis of a measurement system involves using an artifact to make a series of measurement repetitions n across a series of sites x on a single tool/gauge. Precision is a property of an individual tool. The traditional metrics that apply to this situation are repeatability and reproducibility, otherwise known as "gauge R and R." When gathering data, the practitioner knows the circumstances under which the data was acquired. For example, data is acquired without varying the gauge/tool focus or navigation. Alternatively, these factors can be varied along with time and operator. When this data is summarized, it is often analyzed differently among practitioners. In practice, it is desirable to standardize how data in this type of analysis is summarized regardless of the situation in which the data was acquired. The term precision is used to generally describe situations where the practitioner must state the circumstances when quoting actual precision values. For example, short-term dynamic precision applies when repeated measurements are made by loading the artifact wafer in and out of the tool multiple times and when time is typically less than a day or so. Likewise, longer-term dynamic precision applies when repeated measurements are made by loading the artifact wafer in and out of the tool multiple times over the course of many days. Regardless of the situation, the precision should be calculated using the following method. Once calculated, each raw data point has an error of ± this value; it is a 3σ value (i.e., 99.7% of the population falls within this range of data). There is only approximately a 3 out of 1000 chance that a variation that falls outside this range will be random and not systematic (assuming a normal distribution). Note that the units are the same as the raw data. When possible, the artifact used should be a PSA, as defined in a Section 3.5.

NOTE: this analysis assumes no metrology-tool-induced sample damage or sample instability effects. These factors are discussed in Section 3.9.

Using the variance calculation for measurement repetitions n on site x,

$$s^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}. \tag{3.1}$$

Calculate the variance for each site x:

$$(s_1^2, s_2^2, s_3^2, \ldots, s_x^2),$$

average the variances:

$$\frac{(s_1^2 + s_2^2 + s_3^2 + \ldots + s_x^2)}{n}, \tag{3.2}$$

take the square root, and multiply by 3:

$$3\sqrt{\frac{(s_1^2 + s_2^2 + s_3^2 + \ldots + s_x^2)}{n}}. \tag{3.3}$$

3.8.1 Precision example

Figures 3.2 and 3.3 illustrate an example from the spreadsheet included on the CD (called "Precision_TMP_FMP_SPC_chart_limits_spreadsheet"). The spreadsheet can handle up to 50 repeated runs per tool and up to 100 sites per run. In this example, 22 repeated runs with 32 sites were analyzed across a fleet of nine overlay tools. In order to calculate the precision, any of the tabs labeled T1 through T20 can be used. T1 was selected for this example. The raw data for each run must be entered into the areas labeled R1 through R20; R1 stands for run 1, R2 stands for run 2, and so on. Before entering any data in R1–R10, clear any data that may be there already.

In the following example, five repeated runs were performed with ten sites per run. After the data is entered, the spreadsheet automatically calculates each site's average and variance across the runs. It then averages all of the variances to calculate the 3σ precision value. This value (0.46 nm 3σ) is circled in Fig. 3.3. Remember that this value carries the units of the raw data. The calculation of

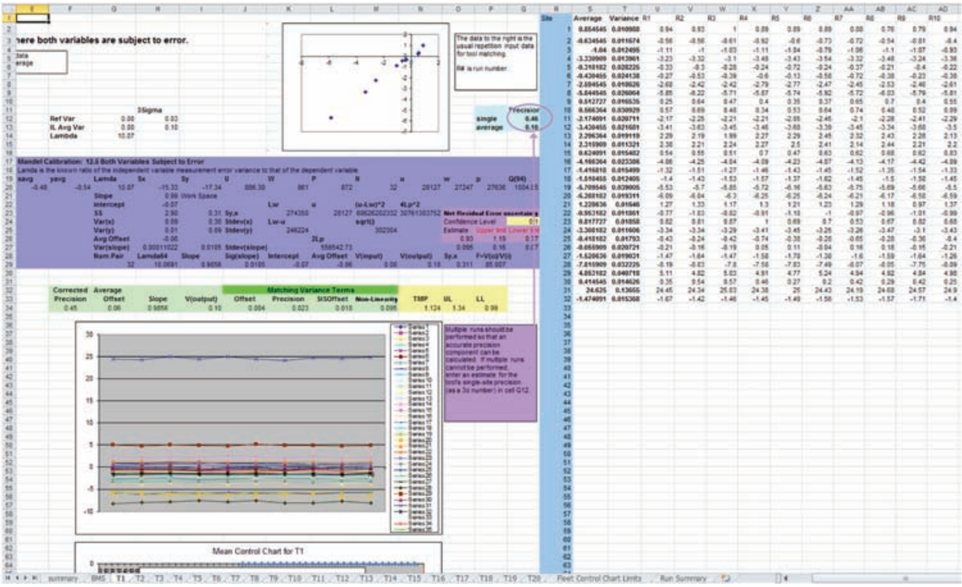
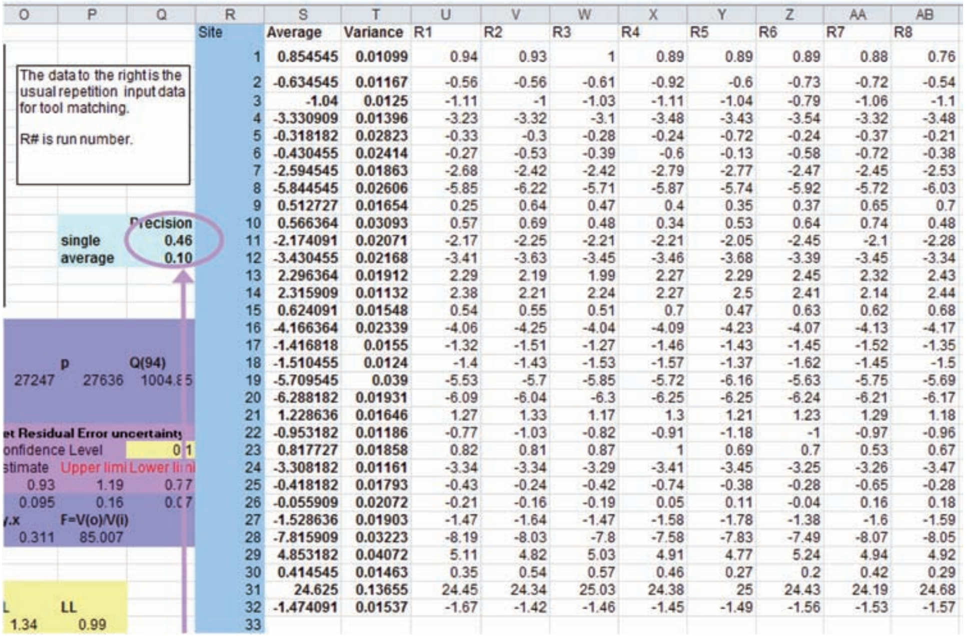


Figure 3.2 Example of the precision calculation in the included Excel spreadsheet.



Site	Average	Variance	R1	R2	R3	R4	R5	R6	R7	R8
1	0.854545	0.01099	0.94	0.93	1	0.89	0.89	0.89	0.88	0.76
2	-0.634545	0.01167	-0.56	-0.56	-0.61	-0.92	-0.6	-0.73	-0.72	-0.54
3	-1.04	0.0125	-1.11	-1	-1.03	-1.11	-1.04	-0.79	-1.06	-1.1
4	-3.330909	0.01396	-3.23	-3.32	-3.1	-3.48	-3.43	-3.54	-3.32	-3.48
5	-0.318182	0.02823	-0.33	-0.3	-0.28	-0.24	-0.72	-0.24	-0.37	-0.21
6	-0.430455	0.02414	-0.27	-0.53	-0.39	-0.6	-0.13	-0.58	-0.72	-0.38
7	-2.594545	0.01863	-2.68	-2.42	-2.42	-2.79	-2.77	-2.47	-2.45	-2.53
8	-5.844545	0.02606	-5.85	-6.22	-5.71	-5.87	-5.74	-5.92	-5.72	-6.03
9	0.512727	0.01654	0.25	0.64	0.47	0.4	0.35	0.37	0.65	0.7
10	0.566364	0.03093	0.57	0.69	0.48	0.34	0.53	0.64	0.74	0.48
11	-2.174091	0.02071	-2.17	-2.25	-2.21	-2.21	-2.05	-2.45	-2.1	-2.28
12	-3.430455	0.02168	-3.41	-3.63	-3.45	-3.46	-3.68	-3.39	-3.45	-3.34
13	2.296364	0.01912	2.29	2.19	1.99	2.27	2.29	2.45	2.32	2.43
14	2.315909	0.01132	2.38	2.21	2.24	2.27	2.5	2.41	2.14	2.44
15	0.624091	0.01548	0.54	0.55	0.51	0.7	0.47	0.63	0.62	0.68
16	-4.166364	0.02339	-4.06	-4.25	-4.04	-4.09	-4.23	-4.07	-4.13	-4.17
17	-1.416818	0.0155	-1.32	-1.51	-1.27	-1.46	-1.43	-1.45	-1.52	-1.35
18	-1.510455	0.0124	-1.4	-1.43	-1.53	-1.57	-1.37	-1.62	-1.45	-1.5
19	-5.709545	0.039	-5.53	-5.7	-5.85	-5.72	-6.16	-5.63	-5.75	-5.69
20	-6.288182	0.01931	-6.09	-6.04	-6.3	-6.25	-6.25	-6.24	-6.21	-6.17
21	1.228636	0.01646	1.27	1.33	1.17	1.3	1.21	1.23	1.29	1.18
22	-0.953182	0.01186	-0.77	-1.03	-0.82	-0.91	-1.18	-1	-0.97	-0.96
23	0.817727	0.01858	0.82	0.81	0.87	1	0.69	0.7	0.53	0.67
24	-3.308182	0.01161	-3.34	-3.34	-3.29	-3.41	-3.45	-3.25	-3.26	-3.47
25	-0.418182	0.01793	-0.43	-0.24	-0.42	-0.74	-0.38	-0.28	-0.65	-0.28
26	-0.055909	0.02072	-0.21	-0.16	-0.19	0.05	0.11	-0.04	0.16	0.18
27	-1.528636	0.01903	-1.47	-1.64	-1.47	-1.58	-1.78	-1.38	-1.6	-1.59
28	-7.815909	0.03223	-8.19	-8.03	-7.8	-7.58	-7.83	-7.49	-8.07	-8.05
29	4.853182	0.04072	5.11	4.82	5.03	4.91	4.77	5.24	4.94	4.92
30	0.414545	0.01463	0.35	0.54	0.57	0.46	0.27	0.2	0.42	0.29
31	24.625	0.13655	24.45	24.34	25.03	24.38	25	24.43	24.19	24.68
32	-1.474091	0.01537	-1.67	-1.42	-1.46	-1.45	-1.49	-1.56	-1.53	-1.57
33										

Figure 3.3 Magnified section of the image from Fig. 3.2.

the precision in the spreadsheet included in the CD uses Eqs. (3.1)–(3.3). The spreadsheet features a chart that plots the given tool's run averages, as shown in Fig. 3.4; it shows the upper and lower control limits for a tool based on a data set. The example here does not necessarily reflect the data used in Fig. 3.3.

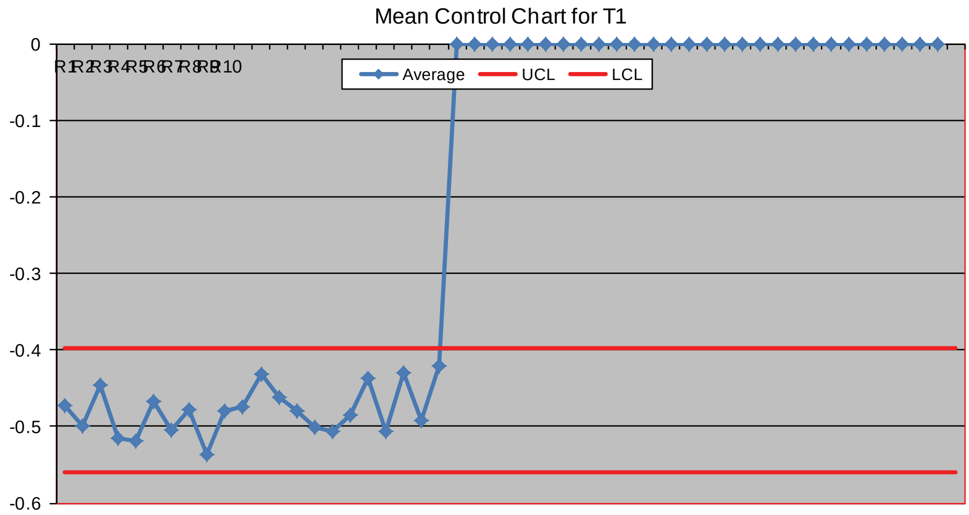


Figure 3.4 Example of a control chart for T1.

3.9 Calculating Precision in the Presence of Measurement-Induced Damage or Sample Changes over Time

Given the known measurement-induced damage imparted by a CD-SEM during repeated measurements, the measurement error must be separated from the variation induced by the measurement damage. One of the best ways to do this is the delta method. Consider the example shown in Fig. 3.5(a), ten sites are measured with five repeats each. The deltas between the first and second measurements are calculated for each site, followed by the variance of these deltas. The process is repeated for the second and third measurements, and so on. [Fig. 3.5(b)]. All of the variances are averaged together, the square root is taken, and then they are multiplied by three. In this case, the precision is 0.127 nm 3σ. The process results in a 3σ value that represents only the measurement error. Measurement-induced sample damage may be either linear or nonlinear, depending on the material being measured. Regardless of which applies, the delta method removes this systematic error and the remainder of the random error associated with the measurement tool. This methodology is not limited to a CD-SEM, although that is where this effect is most prevalent; it can be applied to any toolset where systematic measurement-induced damage or sample changes occur over time.

3.10 Mandel Regression

Subsequent sections on matching and accuracy use a different kind of linear regression, called the Mandel regression. Ordinary least squares assume that all of the error is present in the y axis. When comparing measurement tools, each one has its own error—something that ordinary least squares does not

Structure	1	2	3	4	5	6	7	8	9	10	Slope	Variance
	93.37	93.2	93.17	93.4	93.48						0.042	0.018
	93.51	93.33	93.33	93.61	93.81						0.088	0.041
	93.97	93.78	93.82	94.06	94.28						0.09	0.041
	92.42	92.34	92.46	92.62	92.8						0.104	0.034
	93.54	93.38	93.5	93.71	93.96						0.117	0.051
	93.76	93.68	93.7	93.9	94.09						0.088	0.029
	94.01	93.82	94.03	94.17	94.36						0.105	0.040
	93.51	93.35	93.34	93.47	93.71						0.052	0.023
	92.76	92.62	92.66	92.9	92.99						0.074	0.025
	92.54	92.37	92.4	92.63	92.68						0.054	0.019
Average	93.34	93.19	93.24	93.45	93.62					Average	0.08	0.03
Delta	-0.15	0.05	0.21	0.17							3 Sigma	0.535859

(a)

Deltas

	1	2	3	4	5	6	7	8	9	10	11	12	Average	Variance
Run 1-Run2	0.17	0.18	0.19	0.08	0.16	0.08	0.19	0.16	0.14	0.17			0.15	0.002
Run 2-Run3	0.03	0.00	-0.04	-0.12	-0.12	-0.02	-0.21	0.01	-0.04	-0.03			-0.05	0.005
Run 3-Run4	-0.23	-0.28	-0.24	-0.16	-0.21	-0.20	-0.14	-0.13	-0.24	-0.23			-0.21	0.002
Run 4-Run5	-0.08	-0.20	-0.22	-0.18	-0.25	-0.19	-0.19	-0.24	-0.09	-0.05			-0.17	0.005
Run 5-Run6														
Run 6-Run7														
Run 7-Run8														
Run 8-Run9														
Run9-Run10														
													Avg Var	0.00
													3 Sigma	0.127441

(b)

Figure 3.5 (a) Raw data from repeated runs, and (b) measurement tool error in the presence of measurement-induced damage, as calculated by the delta method.

take into account. Mandel regression is more appropriate for evaluating measurement systems because it accounts for errors from measurement systems on both axes. This regression is a key enabler in the matching and accuracy analysis. Refer to Mandel[7] for more about regression.

3.11 Accuracy

Accuracy is more concerned with the question "is the measurement the right one?" than the matching question "how similarly do tools measure?" or the precision question "how repeatable is the measurement?" It is challenging to make the measurement sensitive to actual changes in the primary characteristic of the measurand and insensitive to changes in any secondary characteristics. For example, the measurement of the bottom linewidth of a resist line (primary characteristic) should not change as the sidewall angle, resist height, or degree of nesting change (secondary characteristics). These properties of a measurement are called relative accuracy, and the TMU methodology is designed to produce the metric with the following properties:

1. Increases in value if the measurement fails to track the primary characteristic,

2. Increases in value if the measurement changes because of changes in secondary characteristics, and
3. Reduces to the precision of the instrument if the relative accuracy is perfect.

The TMU methodology compares measurements from a TuT with a reference measurement system by means of Mandel regression analysis using a PSA. Think of this methodology as a calibration exercise. The PSA has variation in the appropriate primary and secondary characteristics, and the RMS frequently has comparable measurement uncertainty to the TuT, in violation of an assumption in ordinary least-squares regression. A modified Mandel regression is generally used because it regresses based on the assumption that both variables have measurement uncertainty.

The output metrics of regression analysis include the regression slope, the intercept, and the net residual error. In TMU analysis, the average offset between the RMS data and the TuT data is considered a more-useful metric than the intercept. If $\bar{x}$ is the average of the TuT measurements and $\bar{y}$ is the average of the RMS measurements, then $\bar{y} - \bar{x}$ is the average offset. The metrics for any regression technique satisfy

$$\bar{y} = \beta \bar{x} + \alpha,$$

where β is the regression slope, and α is the intercept.

The net residual error is the root-mean-square of the residual errors. On a *xy* regression chart, the *y* distance between each data point and the best-fit line is the residual error for that data point. It is reasonable to assume that the contributors to the net residual error are the uncertainty of the RMS, the precision of the TuT, and the relative accuracy of the TuT. The TMU is calculated by subtracting the uncertainty of the RMS from the net residual error:

$$TMU = 3\sqrt{NRE^2 - u_{RMS}^2}.$$

where u_{RMS}^2 is the variance uncertainty of the RMS. A modified Mandel regression uses u_{RMS}^2 and the TMU as inputs for calculating the best-fit line. Note that because the TMU can only be determined after the regression analysis is performed, the full analysis seeks a self-consistent solution by iteration.

3.11.1 Accuracy example

Figure 3.6 shows an example from an Excel spreadsheet that has been constructed to perform the TMU analysis. The following screenshot shows a calibration exercise of a CD-SEM using reference data from a CD-AFM. The rectangular area from R21C4 to R24C16 contains the regression analysis as described by John Mandel. The three sets of inputs are the RMS and TuT data pairs, the measurement uncertainty estimates for each data

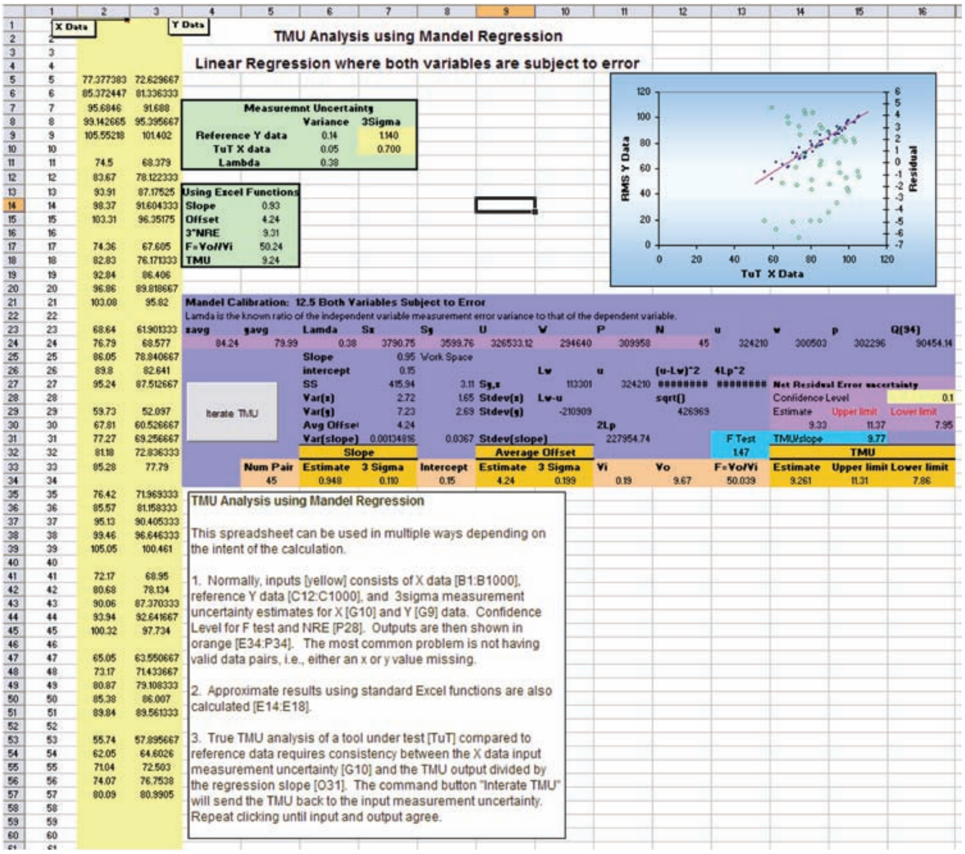


Figure 3.6 Example TMU exercise.

set, and the confidence level to be used for the estimations of the TMU upper and lower uncertainty bounds. The chart shows the data as closed blue circles and the residuals for each as open green circles to be read off the right axis. The values at the bottom of the Mandel analysis show the output. The regression slope, average offset, and TMU are particularly noteworthy. This Mandel regression engine is also used in the FMP analysis workbook.

The software button titled "Iterate TMU" sends the TMU estimate back to the TuT measurement uncertainty estimate after being rescaled by the regression slope. This process should be repeated until the TMU estimate no longer changes. If the regression slope is unity within its uncertainty estimate, then the offset (bias) is constant; otherwise, a fresh TuT measurement should be corrected by the slope and intercept. TMU describes how well this linear correction works.

In the case shown in Fig. 3.6, 45 data pairs (nanometer scale) were used to perform this analysis. The regression slope is 0.993 ± 0.12, which is consistent with unity. The average offset of 4.24 ± 1.4 nm can be used to correct all fresh measurements and eliminate bias. The measurement uncertainty of fresh measurements, including precision and accuracy, is 9.5 nm, based on TMU.

Many of these methods, such as TMU or FMP, can be used to optimize applications. For example, when using TMU, various scatterometry models can be compared to the RMS to determine which has the lowest TMU.

Although TMU analysis is fundamentally a calibration exercise, it is much better known as an extremely useful tool for determining how well a TuT correlates to an appropriate RMS, thereby acting as a success metric to quantify accuracy. TMU has been shown to be far more effective than the coefficient of determination (R2) at assessing the quality of the TuT measurement. TMU has dimensions, it takes into account (removes) the measurement error of the RMS, and it is a combined measure of both the precision and relative accuracy of the TuT. Figure 3.7 illustrates the flow of information from data collection through TMU analysis. The TMU metric, as calculated in this flow, could be replicated multiple times to quantify accuracy assessment of various TuTs or a single TuT but with different models/methodologies. This methodology of improving the metrology performance of a TuT by modifying or refining the measurement system is commonly called metrology optimization. Note the optional flow in Fig. 3.7, where the TMU metrics are fed back into the analysis of the raw signal in order to improve those same TMU metrics—this is accuracy optimization.

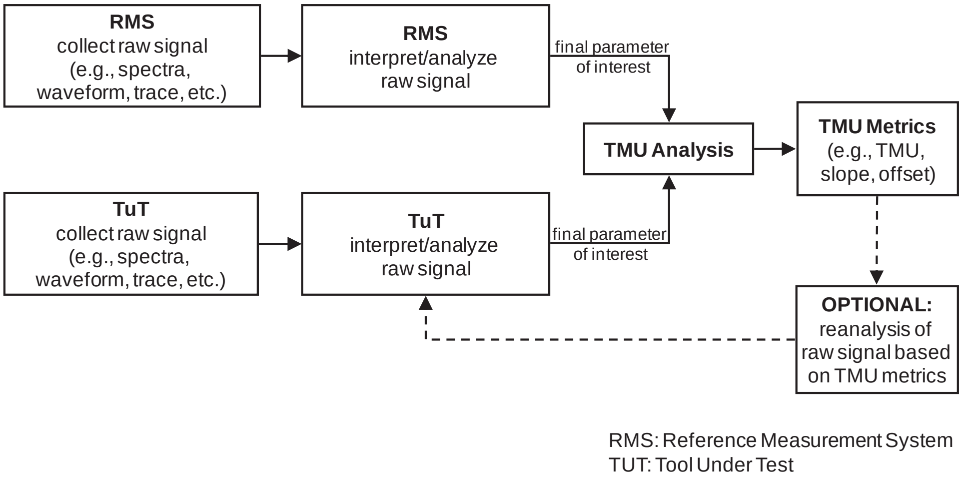


Figure 3.7 Example TMU exercise used to quantify the accuracy performance of a TuT versus a RMS.

3.12 TMU: An Alternative Definition for Clarification

Total measurement uncertainty has also been used in other contexts to describe measurement error. When it comes to describing overlay measurement error, the most notable example comes from KLA-Tencor:

$$TMU = \sqrt{\sigma^2_{precision} + \sigma^2_{TIS} + \sigma^2_{matching} + \sigma^2_{OMF}}. \tag{3.4}$$

In this case, TMU is constructed from four quantities: precision, matching, tool-induced shift (TIS) variability, and optical mark fidelity (OMF). Of these, the TIS component can be calibrated out by measuring the wafer at both the 0 and 180 rotations. Of the remaining three terms, two of them (precision and matching) do not address the accuracy, whereas the last term, the OMF, could be considered a contributor to accuracy. Therefore, for clarification, TMU used in this context is not a comprehensive measure of accuracy; rather, it is more of a holistic metric designed to primarily capture a toolset's precision and matching, where mark fidelity plays a role in increasing TMU when appreciable in magnitude. It therefore should be used in this context. The authors think that FMP is a better measure of a toolset's performance when accuracy is not under consideration. OMF is a diagnostic parameter that should be treated as such and not used to change the measurement error estimate. Mark fidelity might be the root cause when the toolset performance does not meet the requirement—one possible explanation for poor precision or matching results.

3.12.1 Mandel analysis: the building block for new matching terms

A core component of the matching methodology is the Mandel regression analysis, which, when combined with the PSAs, provides a more realistic estimate for each of the key matching metrics described later. Consider performing a calibration exercise between two CD tools across the comprehensive set of PSAs. For each such exercise the average offset and slope are important. Instead of using the standard Microsoft Excel® linear regression, the more generalized Mandel regression methodology is used. (Note that Mandel regression is not described in detail in this book, but key pieces of this analysis are used to accurately formulate the matching metrics.) Continuing on the calibration exercise using the Mandel regression, the first of two additional matching contributors emerges in addition to currently known ones, i.e.,, the precision and offset. From the Mandel regression one extracts the Mandel net residual error (MNRE), which is simply the variation of the paired data set across the two tools (or a fleet average) about the Mandel best-fit line. A key input to the Mandel regression is the variance (precision) estimate for both tools under test. The MRNE is then compared against the variance of both of these tools. The goal here is to determine if there is any scatter about the best-fit line in the plot of tool 1 against tool 2 that cannot be explained by the variances of each tool. The determination of this unexplained scatter is captured by the term nonlinearity, which is defined as the third matching contributor. It is this statistically significant scatter, insufficiently explained by both tool precisions, that allows quantifying the matching limitation between tools. Note that the current practices would have ignored this effect; rather they might have attempted to remove this effect, not realizing it was not due to precision or offset but in fact another effect captured by the nonlinearity. More specifically, the nonlinearity is defined as

$$\sigma^2_{nonlinearity} = \sigma^2_{MNRE} - \sigma^2_1 - \sigma^2_2 \tag{3.5}$$

where σ^2_{MNRE} is the MNRE and is treated as a variance term. The precisions of tool 1 and tool 2 are subtracted from the MNRE after a test for statistical significance is performed. The remainder of the subtraction is defined as the nonlinearity and the third term of matching assessment (see Fig. 3.8). The sample size used to determine the variance estimates for tool 1 and tool 2 must be made high enough to provide good confidence in these estimates. Unfortunately, each tool's variance estimate also has uncertainty: the fewer data points used for the sample size to determine these variances, the more the uncertainty in the estimates. An F-test could help gauge whether the two tools' variances are statistically similar or not, based on the sample size. The larger the sample size is, the better the variance estimate but the longer it takes to collect data across tools. The point here is to consider the effects of the sample size on the variance estimates for each tool. A balance between the quality of the variance estimate and the sample size must be determined. Lastly, it is important to understand the uncertainty for any terms being considered. For the remainder of this book, it will be assumed that the user takes this fundamental requirement into account.

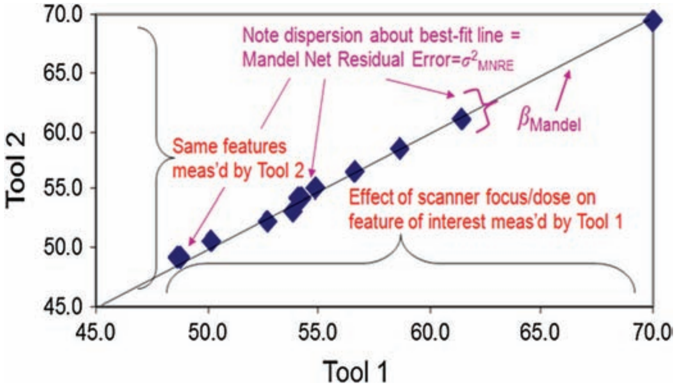


Figure 3.8 Description of nonlinearity where the variability in the data is intentional via a focus and dose modulation.

Next is the fourth matching term. This particular term is related strongly to the Mandel regression slope β, i.e., the β from the Mandel regression is extracted from this same set of data where tool 1 and tool 2 are plotted. In order to understand how β affects the matching process, recall the earlier comments regarding the PSAs. The intentional creation of the expected process variation for each artifact allows one to capture the effect not realized in the current practices. More specifically, when β deviates from unity, another matching contributor emerges. The fourth matching term, the slope-induced shift offset (SiSoffset), is thus introduced:

$$SiSoffset = \nu(Process\ Window)(1 - \beta), \tag{3.6}$$

where the *Process Window* is defined as the allowed variation over time for the given process. The fraction ν depends on the nature of the manufacturing process sampling of the process window over time. To clarify, as a particular process runs through the tools over time where no tool dedication exists, a population of lots is expected to run at or near process target. A population of lots is also expected to run significantly higher and lower than nominal but hopefully within the expected process limits. These lots should pass through any tools in the fleet at any given point in time. Figure 3.9 illustrates the SiSoffset concept. If β is not unity, the results can be described as an offset between the tools that is not constant. This means that if a lot sampled on tool 1 for a given process is near target and another lot sampled on tool 2 is also near target, then the effect is not witnessed; the offset is some value assumed to be constant. However, because a lot is sampled on tool 1 near the upper spec for the process and another lot is sampled on tool 2 also near the upper spec for the process, a different offset is obtained than what was determined between tool 1 and tool 2 near target. Similarly, when a lot is sampled on tool 1 near the lower spec and another lot is measured on tool 2 also near the lower spec, a different offset is obtained than the lots that measured near target.

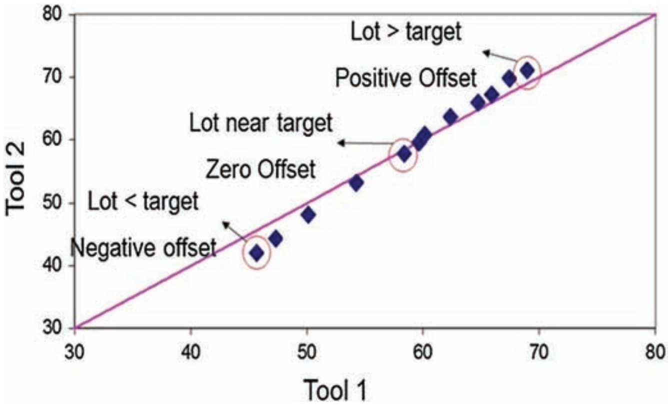


Figure 3.9 The SiSoffset concept, essentially a nonconstant offset between tools.

The offset in the latter case is similar in magnitude but opposite in sign with respect to the offset between lots than measured in the former case. This difference is a direct result of the nonunity β. A process engineer gauging the lot performance for this process when running through the tools over time may see the difference as process variation and attempt to improve the process further when it was really the metrology tools causing this variation. A worse situation involves lots that may be reworked or shipped based on biased data acquired on a tool that does not match the others with respect to β.

The PSAs allow one to capture this effect and quantify it. When β is subtracted from unity and multiplied by the *Process Window* scaled by ν for the specific artifact, the result is the SiSoffset, which is used to capture the expected matching contribution among tools due to β when this process runs over time in the manufacturing line across a fleet of tools. If β is unity between the tools, this term zeros out; if it is not, SiSoffset captures this contribution. Without the SiSoffset term, the metrology tool owner may have misunderstood the effect between tools based on current practices and not realized it is due to SiSoffset. This matching contributor is not captured by the average offset and therefore is difficult to diagnose without SiSoffset.

3.13 Matching

Matching is a term used to describe how a fleet of tools (more than one tool) measures. This process uses one or more artifacts to make a series of measurement repetitions n across a series of sites x on multiple tools. Matching studies can be performed on a homogeneous fleet (all gauges/tools in a fleet are the same model and from the same supplier) or a heterogeneous fleet (a mix of models and/or suppliers). The traditional metric that applies to this situation is the offset, described in Section 2.6. The same rules mentioned in Section 3.8 are also used here, specifically where the practitioner must state the circumstances under which the data was acquired when quoting matching values.

In the semiconductor industry, a fleet of tools is typically used to measure the various processes. Although the fleet can be described by using the offset for each individual tool relative to some standard like a golden tool or fleet average, it cannot be used to describe the entire fleet in a singular value. In practice, it is more desirable to describe the entire fleet with one value and then report all of the individual tool contributions and their breakdown, which would include the offset, among others, for diagnostic purposes. The new metrics were designed to capture these needs. Fleet measurement precision (FMP) is a metric that describes the entire fleet in one value and is therefore typically the best one to use to compare against the given specification/requirement. Each individual tool's contribution to the FMP is the tool-matching precision (TMP). Therefore, the TMP is reported for each tool along with each tool's TMP breakdown. The TMP breakdown is constructed from each tool's contributors: precision, offset, slope-induced shift offset (SIS offset), and nonlinearity. The SIS offset is a component that can be thought of most simply as a nonconstant offset; this term captures the amount of variation explained by a varying offset (nonunity slope). If the offset is constant, then this term zeros out, as shown in Eqs. (3.7) and (3.8). The nonlinearity team is best described as the remaining variability not explained by the precision, offset, or SIS offset. Core to the FMP methodology is the Mandel regression (Section 4.10) and the concept of the benchmark measurement system (BMS). Generally speaking, the best BMS to use to compare each tool in the fleet is the fleet average. The FMP would be the metric used to compare against the roadmap requirements, supplier specifications, or the process needs when the entire fleet (or any subset of tools) is used to control a given process or processes.

The regression analysis requires a large enough spread in data in the artifact(s) so that that the line-fit estimate is accurate. If this is not the case, the fitted line will have large error bars, which means that the SIS offset and nonlinearity terms will likely have poorly determined magnitudes. Therefore, these terms should not be relied on when making any conclusions; the practitioner should primarily rely on the precision and offset terms.

Matching studies should be performed just like a precision study. Specifically, multiple repeated runs should be performed on each tool participating in the matching study. This process contrasts the single-tool variability metric of precision with the matching metrics for significance. Note that this analysis assumes no metrology-tool-induced sample damage or sample instability effects. For reference, the following equations are already incorporated in the example spreadsheet included on the CD; the example data can be cleared so that new data can be loaded and analyzed.

The TMP for the i^{th} tool is determined by the following equation (in this example, the tool is compared to the BMS):

$$TMP_i = 3\sqrt{\beta_i^2\sigma_i^2 + offset_i^2 + SISoffset_i^2 + \sigma^2_{non\text{-}linearity,i}}. \tag{3.7}$$

where β_i is the regression slope; σ_i is the precision sigma;

$$offset_i = \bar{x}_{BMS} - \bar{x}_i;$$

$$\sigma^2_{non\text{-}linearity,i} = \sigma^2_{MNRE} - \sigma_i^2 - \sigma^2_{BMS};$$

and

$$SISoffset_i = \nu \times ProcessWindow \times (1 - \beta_\mathrm{i}),$$

where ν is the fraction of the linewidth variation expected in production for this artifact, and MNRE is the Mandel net residual error. All terms come from performing a Mandel regression analysis that compares the i^{th} tool to the BMS.

FMP estimates the matching quality for the fleet:

$$FMP = \sqrt{\sum_{i=1}^{N} \frac{TMP_i^2}{N}}. \tag{3.8}$$

3.13.1 Matching example 1

Assume in a two-tool fleet the following (e.g., qualifying tool 2 or assessing the performance of the two tools):

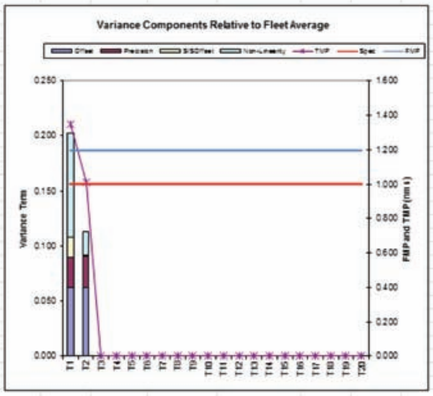
- Requirement is 1 nm;
- Tool 1 precision = 0.5 nm;
- Tool 2 precision = 0.5 nm;
- Tool 1 to tool 2 offset = 0.5 nm.

Is the 1-nm requirement met? No—as shown in Fig. 3.10, the requirement is exceeded by 20% when each tool's precision and the offset are combined. The FMP is 1.193 nm. Per the chart in the bottom right of Fig. 3.8, in order to get the FMP below 1 nm, either the offset or the nonlinearity (or both) should be reduced. If only the offset and/or precision were considered, it might appear that the requirement is being met when it is not.

3.13.2 Matching example 2

This example is included as a spreadsheet on the CD (called "Precision_TMP_FMP_SPC_chart_limits_spreadsheet"). It handles up to 20 tools for a matching study. The data for each tool is entered as described in Section 3.8.1. The tabs labeled T1 through T20 represent the data for tools 1 through 20. Any combination of tools can be used for matching from 2 to 20. Once the raw data for each tool in entered into the spreadsheet, the spreadsheet automatically calculates each tool's TMP and the FMP. The BMS is calculated and shown in the BMS tab; in this case, the BMS is the fleet average. Each tool's data is regressed against the BMS to determine its TMP, as the formulas describe.

Each tool's key data is summarized in a table format in the tab labeled "Summary," an example of which is shown in Fig. 3.11. Each tool's TMP is shown in the box highlighted under the TMP label, and the FMP is shown in the box under the FMP label. The needed specification is entered in the cell underneath the cell labeled "Spec." Nine overlay tools were used in this study. Also notice that each tool's TMP components are shown in the table: the raw and corrected precision, offset, SIS offset, and nonlinearity. The TMP



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1							Average				Use 20% Nomina	Spec	FMP	UL	LL
2							0.000				n	1.00	1.193	1.30	1.13
3		Include in			Raw	Corrected			Matching Variance Terms						
4	Active ?	BMS?			Precision 3	Precision 3	Offset	Slope	Offset	Precision	SISOffset	Non-Linearity	TMP	UL	LL
5	y	y	T1	tool 1	0.500	0.493	0.250	0.986	0.063	0.027	0.018	0.095	1.351	1.54	1.25
6	y	y	T2	tool 2	0.500	0.502	-0.250	1.003	0.063	0.028	0.001	0.022	1.010	1.07	0.98
7	n	y	T3	test											
8	n	y	T4	test											
9	n	y	T5	test											
10	n	y	T6	test											
11	n	y	T7	test											
12	n	y	T8	test											
13	n	y	T9	test											
14	n	n	T10	test											
15	n	n	T11	test											
16	n	n	T12	test											
17	n	n	T13	test											
18	n	n	T14	test											
19	n	n	T15	test											
20	n	n	T16	test											
21	n	n	T17	test											
22	n	n	T18	test											
23	n	n	T19	test											
24	n	n	T20	test											

Figure 3.10 Data summary for matching example 1.

						Average					Use 20% Nomina	Spec	FMP	UL	LL
						0.000					n	1.50	0.843	0.90	0.81
	Include in			Raw	Corrected			Matching Variance Terms							
Active ?	BMS?			Precision 3	Precision 3	Offset	Slope	Offset	Precision	SISOffset	Non-Linearity	TMP	UL	LL
y	y	T1	tool 1	0.458	0.452	0.063	0.986	0.004	0.023	0.018	0.095	1.124	1.34	0.99
y	y	T2	tool 2	0.368	0.369	-0.009	1.003	0.000	0.015	0.001	0.023	0.591	0.69	0.53
y	y	T3	test	0.384	0.387	-0.042	1.007	0.002	0.017	0.004	0.014	0.577	0.65	0.54
y	y	T4	test	0.481	0.486	-0.088	1.009	0.008	0.026	0.007	0.101	1.131	1.36	1.00
y	y	T5	test	0.295	0.299	0.040	1.013	0.002	0.010	0.015	0.049	0.821	0.97	0.73
y	y	T6	test	0.395	0.393	0.028	0.993	0.001	0.017	0.004	0.005	0.497	0.53	0.48
y	y	T7	test	0.487	0.490	-0.034	1.006	0.001	0.027	0.003	0.062	0.915	1.09	0.81
y	y	T8	test	0.546	0.539	0.014	0.987	0.000	0.032	0.015	0.081	1.073	1.27	0.96
y	y	T9	test	0.381	0.379	0.029	0.995	0.001	0.016	0.002	0.010	0.512	0.57	0.48
n	n	T10	test											
n	n	T11	test											
n	n	T12	test											
n	n	T13	test											
n	n	T14	test											
n	n	T15	test											
n	n	T16	test											
n	n	T17	test											
n	n	T18	test											
n	n	T19	test											
n	n	T20	test											

Figure 3.11 Example of a matching worksheet.

components shown under the "Matching Variance Terms" are the variance terms, which allow each term to be compared in order to determine which are dominant and thus the largest contributors to the FMP. These are charted visually in Fig. 3.12, shown as "Variance component relative to fleet average." This figure makes it easier to see each tool's contribution to the FMP and therefore decide how to proceed to drive improvements. In this situation, note

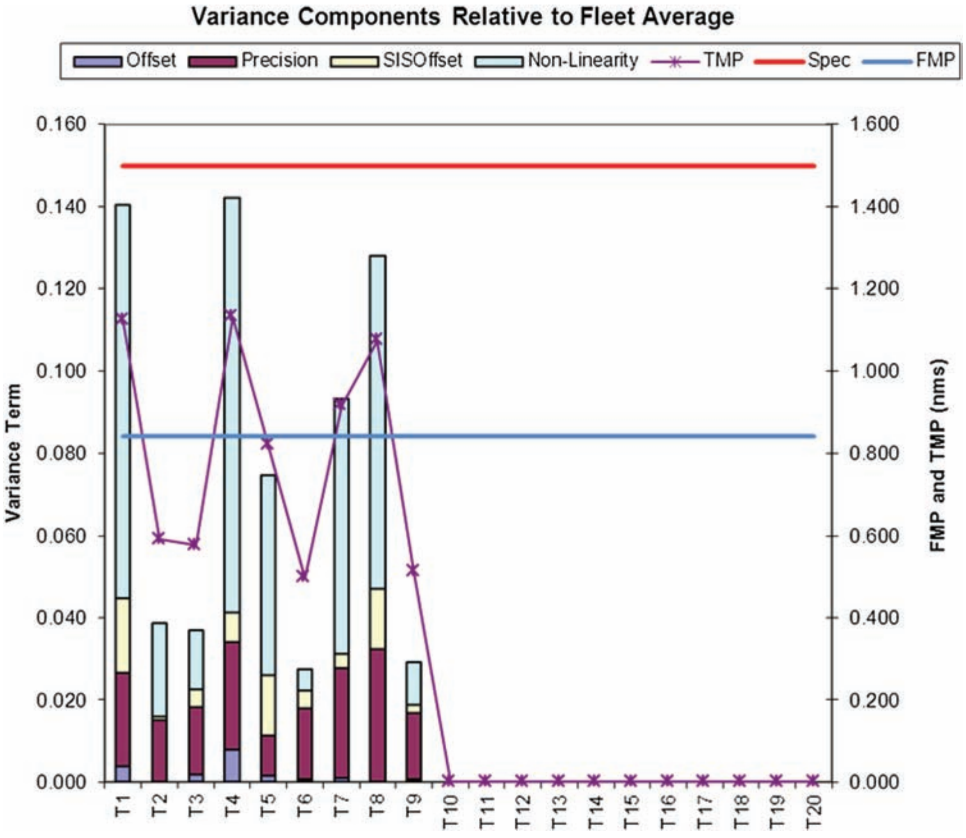


Figure 3.12 FMP and TMP breakdown of Fig. 3.11.

that the FMP is below (0.84 nm) the needed specification (1.5 nm). Also note that each tool's TMP is below the specification. If the practitioner wants to drive further improvements, the T1, T4, and T8 TMPs can likely be improved to the level of the other tools. These tools appear to have much more of a nonlinearity component than the other tools. Experiments should be performed to rectify these issues especially if further improvement is desired. Figure 3.13 shows how all of the run averages from all the tools are distributed. The two horizontal lines are the upper and lower control limits, calculated as described in Section 3.11. In the column labeled "Include in BMS?", by selecting "y" or "n," the user can selected what tools are included in the BMS. By selecting "y" for all of the tools that data was collected for, the BMS is the fleet average and includes all of the evaluated tools. By selecting "n" for all of the tools except one, then that tool is used as the BMS. Similarly, by selecting "y" or "n" in the column labeled "Active?," the practitioner can remove tools from the overall analysis. In this case, if T1, T4, and T8 were changed to "n," then the analysis would not consider data from these tools. This change allows the user to estimate the improvement in FMP while

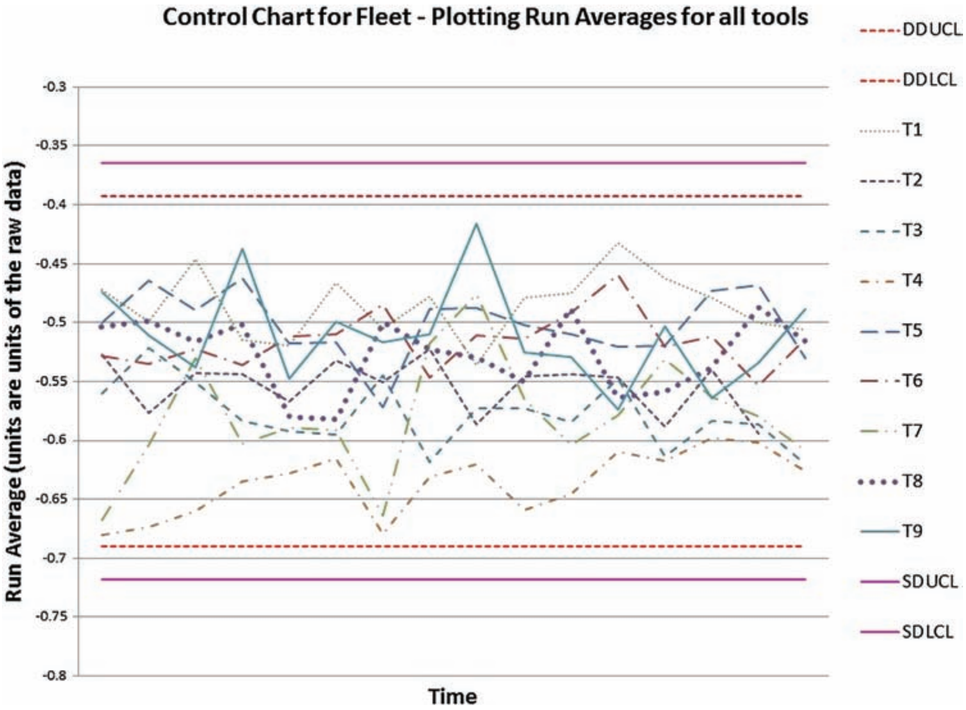


Figure 3.13 Plotting run averages.

ignoring these tools in production for the given application or fix them in line with the other tool's TMPs.

3.14 Sustaining/Stability/Statistical Process Control (SPC)

3.14.1 Using TMP and FMP methodology to calculate SPC control-chart limits

Statistical process control (SPC) is commonly used to ensure that the qualified measurement system continues to perform as expected, especially after tool qualification and after repair work is performed on the tool. Periodic measurements of a standard or stable PSA are performed with the same measurement recipe, wafer, and sites, and the mean and standard deviation are compared with a stable period of that tool's historical performance. Statistically derived 3σ control limits are added to the mean and sigma trend charts, and standard Western Electric run rules are applied to detect changes in the measurement tool, i.e., loss of stability, with the goal of inhibiting the measurement tool. When calculating chart limits this way on a metrological tool, it is difficult to determine whether the calculated SPC limits are consistent with the requirements. This difficulty is important because if the SPC control-chart limits were derived from a period of time when the desired requirements were not being met, as indicated by the FMP value, then the calculated limits will not maintain the toolset requirements. The chart limits will be inflated, and the tools won't be controlled to the extent needed.

Solving this problem involves the difference between how daily monitors are generally run in contrast to how MSA is performed. When running a given daily monitor, a number of sites are usually measured and averaged together. The FMP is a property of the individual measurement during MSA. More specifically, whenever a single measurement is made, the measurement error is expected to be ± the FMP value (it is a 3σ value). Ideally, the daily monitor calculated-chart limits are related to the single-measurement-point requirement. Thus, if a daily monitor remains within its limits, the toolset maintains the individual-measurement-point error requirement. Likewise, if the daily monitor drifts outside these limits, the toolset also fails the single-measurement-point measurement error requirements. The averaging performed by the daily monitor must be considered when calculating the chart limits. The formulas in the following subsections describe how to set limits for individual tools and the fleet. They ensure consistency between the individual-measurement error requirements used during MSA and the chart limits calculated for the long-term daily monitor(s). By scaling the limits by the square of n (where n is the number of sites being averaged in the given daily monitor), the chart limits and the individual-measurement-points requirements are self-consistent.

The chart limits for the fleet of tools should generally be derived from the FMP of the daily monitor data (or precision, in the case of a single-tool chart) or the fleet's FMP requirement, whichever is smaller.

3.14.2 Mean control-chart-limit calculations for each tool

The upper control limit (UCL) of a tool is defined as

$$UCL = \bar{x}_i + 3\sigma_i \text{ or } requirements/\sqrt{n},$$

and the lower control limit (LCL) as

$$LCL = \bar{x}_i - 3\sigma_i \text{ or } requirements/\sqrt{n},$$

where i is the i^{th} tool in the fleet, $\bar{x}_i$ is the mean of all of the data from i, $3\sigma_i$ is the precision of i (use either 3σ or *requirements*, whichever is smaller), and n is the number of sites in the daily monitor.

3.14.3 Mean control-chart-limit calculations for the fleet

The UCL of the fleet is defined as

$$UCL = \bar{x}_{fleet} + FMP \text{ or } requirements/\sqrt{n},$$

and the LCL as

$$LCL = \bar{x}_{fleet} - FMP \text{ or } requirements/\sqrt{n},$$

where $\bar{x}_{fleet}$ is the mean of all of the data from all tools, *FMP* is the fleet measurement precision (use either *FMP* or *requirements*, whichever is smaller), and *n* is the number of sites in the daily monitor.

3.14.4 Control-chart-limit setup example 1

This example involves setting up long-term control-chart limits for a monitor wafer to be used on an 11-tool fleet. The requirement that must be met is 1 nm. Figure 3.12(a) shows the results of the daily monitor; there are five data points for each tool, which represent the daily monitor being run each day for five days on the *x* axis. Each data point is the average of 32 sites measured in the daily monitor and is plotted against the *y* axis. The same arrangement is used across the fleet, measuring the same 32 sites. For reference, the 1-nm range is shown in the figure. Is the 1-nm requirement being met? Because there are some points outside the 1-nm range, the answer is no, but it might appear otherwise because the outlying points are not very far from the required range. Figure 3.14(b) illustrates the result if the same data is re-analyzed using the FMP methodology. The FMP is found to be 3.3 nm, which is much worse than the requirement and highlights the importance of considering all of the data appropriately.

As shown in Fig. 3.15, if the chart limits are calculated based on conventional control-chart theory, they would look like the dashed lines. Note

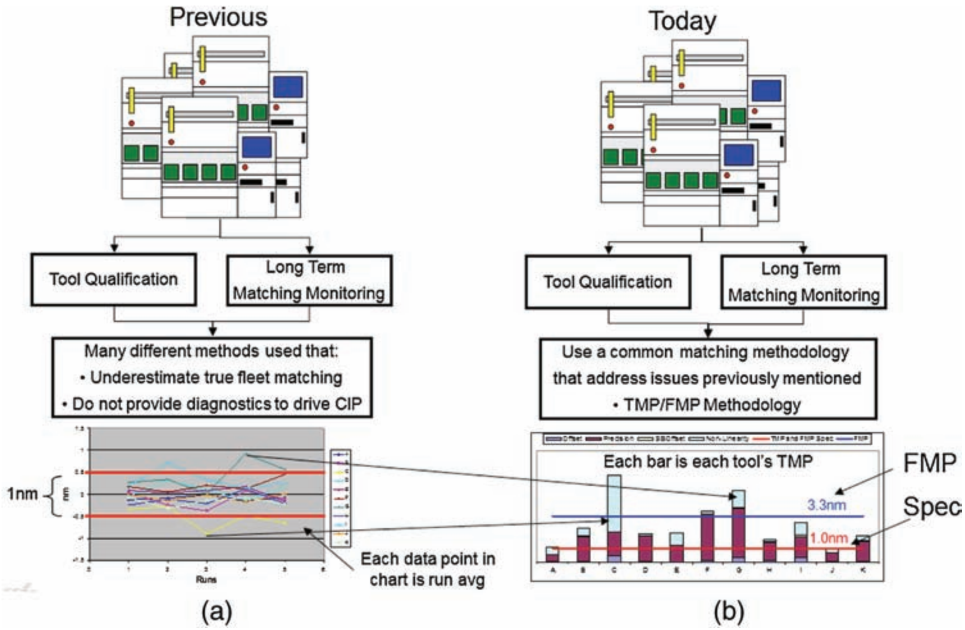


Figure 3.14 (a) Daily monitor data from 11 tools against the 1-nm range; (b) the same data re-analyzed using the FMP methodology.

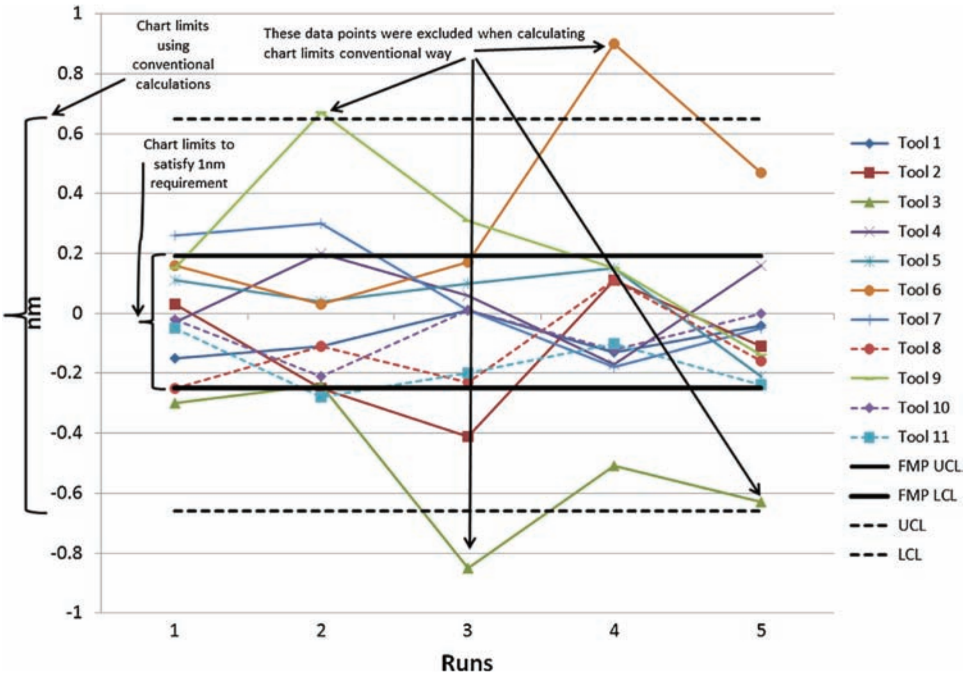


Figure 3.15 Control chart that illustrates the limit setup using conventional control-chart theory (dotted lines) versus the requirement-based limits (solid lines).

that three data points were excluded from this analysis because they were obtained when the given tool was in an unhealthy state. Because the requirement is not taken into account when setting these limits, the 1-nm requirement would never be achieved if these were used. In order for the toolsets to meet the 1-nm requirement, the limits must be calculated as follows:

$$UCL = \bar{x}_{fleet} + FMP \text{ or } requirements/\sqrt{n},$$

and

$$LCL = \bar{x}_{fleet} - FMP \text{ or } requirements/\sqrt{n},$$

where $\bar{x}_{fleet}$ is the mean of all of the data from all tools, FMP is the fleet measurement precision (use either FMP or *requirements*, whichever is smaller), and n is the number of sites in the daily monitor.

In this case, the mean of all the data from the fleet is -0.1 nm, the requirement is 1 nm, and n is 32 because there are 32 sites being averaged. According to the formula in this section, the chart limits should be $-0.1 \pm 1/\text{sqrt}(32)$, which is -0.1 ± 0.18. Therefore, the UCL should be 0.08, and the LCL should be -0.28 (both in nanometers), shown in Fig. 3.15

as the two solid lines. In order to ensure that the 1-nm requirement is met, all of the data points from each of the tools must be within these limits. The chart limits calculated using a conventional control chart would yield limits indicated by the two dotted lines. Note that numerous data points are outside the derived limits of these requirements, whereas most are within the conventional limits. The conventional limits will never drive measurements that meet the requirement; they are derived during a timeframe that is "stable," e.g., this "stable" time period does not meet requirements (recall that three data points were removed that did not represent stable data. The limits would have been even wider had these been included).

The requirement-based limits and the FMP/TMP methodology can be used to determine how to produce a toolset that meets the requirements. Figure 3.16 shows each tool's TMP. In order to drive improvements, the following should be performed:

- **Step 1:** A containment action plan inhibits tools B, C, D, F, G, and I, and runs product only on tools A, E, H, J, and K. This step makes the FMP = 1.1 nm, and then the investigation starts. (As noted earlier, the included spreadsheet allows the data to be easily re-analyzed by turning off unwanted tools).
- **Step 2:** Tools C, F, and G primarily focus on reducing fleet variation (FMP). Tools F and G have precision issues, and tool C has a nonlinearity issue—it is necessary to determine why and rectify these issues. The other tools are much better, proving that these tools can be improved.
- **Step 3:** To improve the FMP further, tools B and D should be investigated to have their precision reduced and determine why.

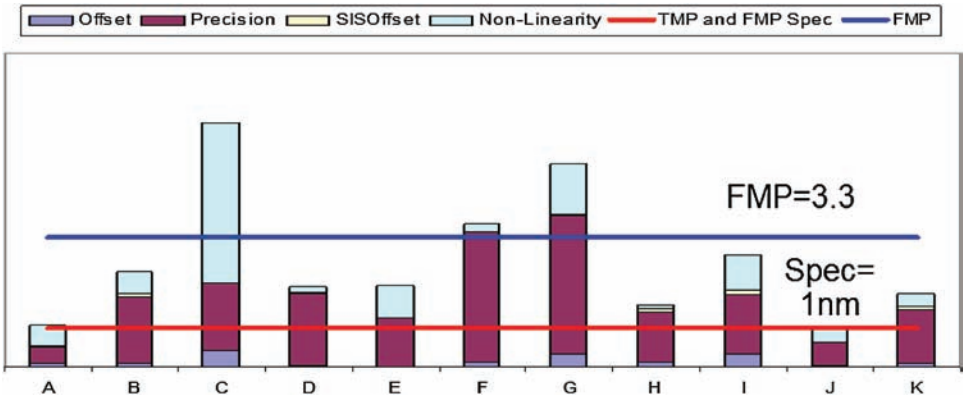


Figure 3.16 Daily monitor data re-analyzed using the FMP/TMP methodology, which provides a visual on which tools and TMP components are worst to help drive continuous improvements.

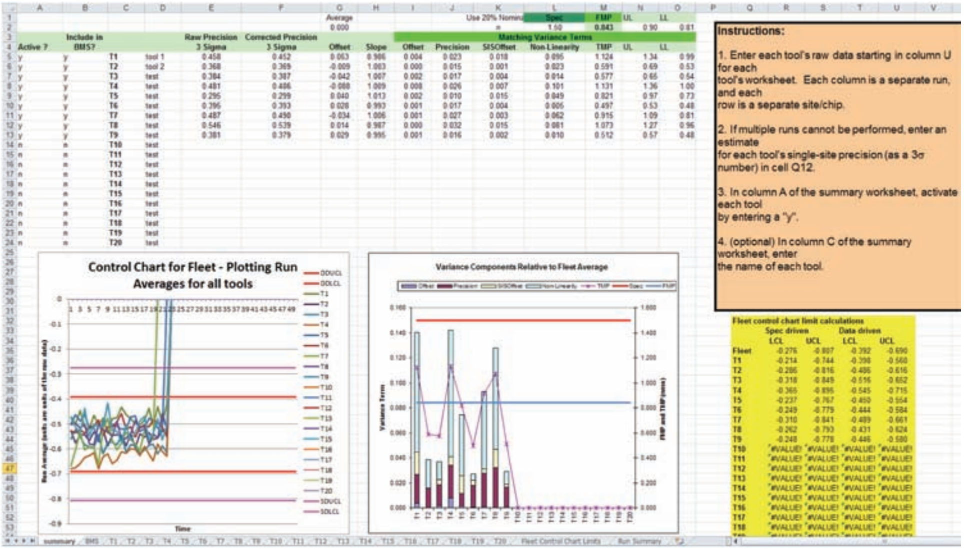


Figure 3.17 Calculated daily monitor SPC limits for a wafer used as a daily monitor (see cells R35 to U44).

3.14.5 Control-chart-limit setup example 2

If a wafer with the data shown in Fig. 3.17 were to be used as a daily monitor, then the SPC chart limits shown in cells R35 to U44 would be used. As mentioned earlier, because the FMP is lower than the requirement, then the "Data driven" chart limits would apply. If for some reason these limits are too aggressive to maintain (e.g., the supplier will not agree to limits tighter than the requirements), then at a minimum the "Spec driven" limits should be used. Either way, setting the limits using the "Data driven" values will ensure that the daily monitor (which averages 32 sites each run) maintains the fleet at the individual-measurement-point FMP value of 0.84 nm. If T1, T4, and T8 are improved in the future, then the chart limits can be reset even tighter as the FMP improves.

3.14.6 Leveraging FMP to determine the root cause to allow the fleet to meet the requirements

The following is an example that illustrates how to take the information learned from the TMP visual signal and drive the improvements needed to meet the requirements. This methodology can be applied to any tools or TMP components that are identified as being higher than they should be. Assume that the toolset is known to be failing the requirements because the FMP is higher than the requirements. Also assume that, based on the TMP data, it has been determined that one of the tools has a much larger TMP than the others due to the offset component. If this is a fleet of CD-SEM tools, then the

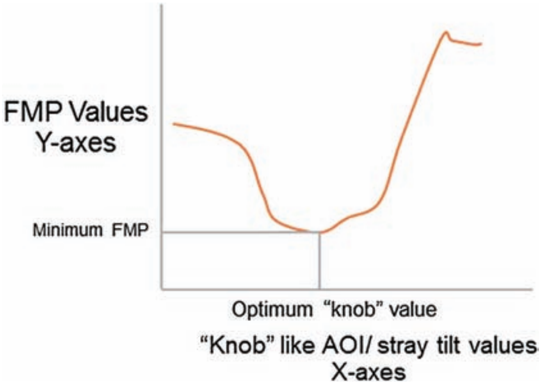


Figure 3.18 FMP response (*y* axis) when the parameter/knob is varied.

next issue to examine is what parameters ("knobs") on a CD-SEM can change the offset, e.g., magnification calibration, stray tilt, detector setup, etc., and then vary the one most able to explain the offset. For each setting changed for this one parameter, repeated data can be collected across the fleet and the FMP can be plotted (see Fig. 3.18). The x axis represents the modulation range for the parameter/knob in question, and the y axis represents the FMP values.

If the stray tilt in this scenario was varied in seven equally spaced increments on the problematic tool—from 0.5 deg to -0.5 deg—then repeated data can be obtained from each tool in the fleet and the FMP can be calculated for each of the seven increments (note that the stray tilt would not be changed for the good tools). The place where the FMP is minimal is the optimal "knob" value to use. Success is achieved if the FMP improves to a level at or below the requirements; the root cause has been found, along with the range within which this given parameter minimizes its impact on the FMP, which may need to be monitored. If the FMP improves but remains above the requirements, then another parameter should be varied and a similar experiment performed until the offset is further corrected.

Note in the situation where all of the tool TMPs and their components look nearly identical and the requirement is still not met, then a capability limit has likely been reached for this given application/situation. The only likely solution is to pursue a next-generation toolset and/or look for another supplier.

3.15 Sampling Plan: Catch-All Major Variations

In addition to using graphs, tables, and charts to display statistical measurement data (descriptive statistics), another application of statistics is sampling (using sample data to draw inferences about a population), also called inferential statistics. This method draws inferences about a universal set based on the study of sample data. The conclusions from the study of sample data are applied to the whole population, which makes it important to select representative data (sampling plan). Otherwise, the conclusions from the sample may be misleading. A carefully designed sampling plan is needed to determine what metrology will measure and how many measurements are needed to catch potential process variations and not add too much process time. An optimized sampling plan has minimal measurement sites to catch all potential process variations. In statistical terms, a sample (e.g., selected measurement sites) is selected to represent a population (all possible measurement sites—normally a very large number), and statistical tools must be used to analyze the data and draw conclusions from the sample to the population. A special sampling plan can be used for a special purpose, e.g., to catch a specific, high risk variation during a specific period of time.

In industrial manufacturing process control, measurement is considered a quality-control step; ideally, it does not alter or change products in their manufacturing process, although it does take time in the manufacturing process and contributes to the overall product cycle time. In that regard, metrology is normally considered a "non-value added step." Its value cannot easily be estimated from the product value itself. Its value is instead in "quality insurance," catching process variation (or quality deviation) early to prevent potential failures, including catastrophic failures. (See Chapter 6 for more details).

Three different sampling strategies can be illustrated by CD sampling while monitoring lithography. Note that due to the uniqueness of the lithographical process and its characteristic variations, three major CD variations exist:

- within-field CD variation [sometime called across-chip linewidth variation (ACLV)],
- wafer-to-wafer CD variation, and
- lot-to-lot CD variation.

A CD sampling plan must capture those variations. The batch process of the wafer flow in IC fabs means that the water-to-wafer CD variation is much smaller than the other two variations and that the sampling focus has always involved catching within-field and lot-to-lot variation.

3.16 Random Sampling

Random sampling is the simplest and most common sampling plan; all samples have the same chance of being chosen. When only one characteristic of a product is measured in a process control step, a random sampling is most suitable. Randomly selected products are measured and analyzed to catch potential variations. Statistically speaking, random sampling has the highest probability to catch process variations for a given number of sampling points. Some modifications can be applied to normal random sampling, such as dynamical random sampling, i.e., as sample rate increases, the standard deviation and range are dynamically calculated and compared until their changes are within predefined limits. This modification may help prevent under- or oversampling. Random sampling is not very useful in IC manufacturing because, in most process-monitoring schemes, it targets certain fixed variations that relate to the way Si wafers are processed in an IC fab.

3.17 Systematic Sampling

The designed image patterns in the IC manufacturing process (especially in image-transferring lithography and etch processes) have many features, e.g., line/space patterns, more-complicated line ends, line turns, and different 2D shapes, and 2D contacts/via patterns. In addition, there are dense lines/spaces/contacts (grating structures with nearby repeating lines/spaces/contacts), isolated lines/spaces/contacts (no nearby features within optical interaction range), and anything in between. Depending on the main feature in a specific layer, one or more features are selected in a sampling plan. For example, in a gate layer, both the dense and isolated lines are selected.

Silicon wafers are round, and due to the shape-related process nature (e.g., a spin coating photoresist and plasma etching), there are inherent center and edge differences in process conditions, thus creating process biases. Such biases have been intentionally minimized in process development, but they still exist and are sometimes alarmingly large in certain process steps (e.g., in lithographical and etch processes). To monitor such spatially related process biases, a sampling plan will have measurement sites that cover both the center region and edge region in different orientations. In an IC semiconductor manufacturing process, such shape-related variation is classified as within-wafer. The other two variations in IC manufacturing processes are within-chip (or across chip) and within-lot [also called wafer-to-wafer variation]. A lot refers to multiple wafers in a wafer cassette or a sealed wafer container called a front open utility pod (FOUP)—all wafers within a lot are processed together in a batch process and lot-to-lot. The total variation for a given lot is the combination of within-lot, within-wafer, and within-chip variations.

In order to catch these effects, the process engineer would ideally know where the dominant variation exists thanks to extensive process-characterization activities during set up. A sampling strategy that targets the key dominant terms would therefore be desirable. For example, if 90% of the variation originates within the wafer, then the sampling strategy should be primarily arranged to catch this variation. This approach minimizes metrology utilization to focus on the most important aspects of variation. In practice though, most sampling plans measure multiple wafers, multiple chips, and multiple structures in an effort to minimize risk at the expense of smarter metrology sampling plans to minimize unnecessary utilization.

3.18 Summary

Considerable care has been taken to redefine more-appropriate gauge metrics and methodologies. What metric should be used to compare measurement tool performance against requirements/specifications (minimally 2% of target, more ideally 1%) in gauge study/ MSA activities?

- For a single-tool fleet n, use precision.
- When $n > 1$, what should be used? Precision, offset, or both? The appropriate metric to use is FMP.

It is desirable to represent the fleet with one value that captures the measurement variation of the entire fleet. Because product is typically measured on the entire fleet, this value is the FMP, which should be compared against the requirements when $n > 1$. FMP is the key metric used to describe the entire fleet, and TMP is the key metric for each tool in the fleet. The latter is a composite metric created from each tool's precision, offset, SIS offset, and nonlinearity. These metrics are also the fundamental building blocks for improving the measurement performance of the fleet; TMP provides critical diagnostic information about how to improve FMP (especially important when requirements are not met).

How should daily monitor control limits be set up to ensure that requirements are being met? Conventional control-chart theory does not account for the requirements. The control-limit calculation, which uses conventional control-chart theory, is based on a "stable" period of time but does not consider whether or not the toolset meets the requirements. A more-desirable determination of control-chart limits would consider the requirement that must be maintained; this is achieved by deriving the control-chart limits from the requirements and taking into account the averaging performed in the daily monitor. This step is an important breakthrough that links the metrics used for MSA (precision, TMP, and FMP) with the control-chart limits used to maintain the toolset performance over time. Past methods treated MSA and setting up daily monitor control limits as independent events. In essence, maintaining requirements over time (long-term stability) is no different than a MSA/gauge study—it is just performed over a longer time period.

The provided template spreadsheet automatically performs the entire analysis mentioned earlier; the only required action is entering the raw data into the spreadsheet. This also ensures that everyone performs the analysis the same way.

Accuracy (TMU) should be strongly considered, especially in areas such as OPC model calibration and scatterometry model validation. Lastly, PSAs should always be used when possible.

Note that a special regression analysis called Mandel regression is used in this analysis; this regression analysis considers the measurement error in both x and y axis variables, something that conventional regression does not do.

References

1. International Technology Roadmap for Semiconductors (ITRS), www.itrs.net.
2. http://www.sematech.org/docubase/abstracts/35939.htm.
3. C. N. Archie et al., "Characteristics of accuracy for CD metrology," *Proc. SPIE* **3677**, 291 (1999) [doi: 10.1117/12.350818].
4. E. Solecky et al., "New comprehensive metrics and methodology for metrology tool fleet matching," *Proc. SPIE* **5752**, 248 (2005) [doi: 10.1117/12.602330].
5. J. Morningstar et al., "Metrology tool fleet management: applying FMP tool matching and monitoring concepts to an overlay fleet," *Proc. SPIE* **6152**, 615212 (2006) [doi: 10.1117/12.655578].
6. E. Solecky et al., "Monitoring measurement tools: new methods for driving continuous improvements in fleet measurement uncertainty," *Proc. SPIE* **7272**, 72721H (2009) [doi: 10.1117/12.814089].
7. J. Mandel, *The Statistical Analysis of Experimental Data*, Dover Books, New York (1964).
8. B. Su, G. Eytan, M. Padmanaban, and A. Romano, "Analyzing and characterizing 193nm resist shrinkage," *Solid State Technol.* **44**(5), 52 (2001).
9. M. Adel et al., "Performance study of new segmented overlay marks for advanced wafer processing," *Proc. SPIE* **5038**, 453 (2003) [doi: 10.1117/12.483477].
10. G. W. Banke, Jr. and C. N. Archie, "Characteristics of Accuracy for CD Metrology," *Proc. SPIE* **3677**, 291 (1999) [doi: 10.1117/12.350818].

Chapter 4
Metrology in the Semiconductor IC Industry

4.1 Pervasiveness of Metrology

Metrology measurement and defect inspection steps are ever-present in the semiconductor industry in every technology node (see Fig. 4.1). Given the increasing difficulty in manufacturing current- and future-generation chip technologies, process teams increasingly rely on metrology and defect-inspection engineers to develop manufacturable processes. A working chip takes months to manufacture and can include as many as 1800 steps in a route (a route is a complete set of steps properly ordered to manufacture a given chip). Chips are typically produced by a particular technology node, which defines the level of technology and complexity used to manufacture the chip. A 65-nm technology node is much less complex than a 22-nm or 14-nm technology node. As mentioned earlier, the number generally indicates the minimum size of the circuitry being printed. Figure 4.1 shows the percentage of the route used by metrology and defect-inspection steps from the 65-nm node to the 14-nm node at IBM and GlobalFoundries. These technology node designations are used internally at IBM but are similar to designations used at other companies. Note that the first three categories of steps shown are for e-beam metrology, inspection, and defect review. A CD-SEM is responsible for dimensional/structural measurements; e-beam review (EBR) is responsible for defect re-detection and classification; and e-beam inspection (EBI) is responsible for physical defect inspection.

The next three categories are optical thin-film measurement, overlay, and a group labeled as "the rest of metrology and inspection," which includes composition measurement, brightfield and darkfield inspection, etc. The last two categories are the lithography and the remaining unit process steps. As noted by the diamond symbols in Fig. 4.1, the metrology/defect steps at the 65-nm node accounted for $\sim$50% of all steps in the route. This percentage increases to $\sim$55% at the 22-nm node. There is a drastic change at 14 nm: the percentage increases to 70%. The square symbols in the figure

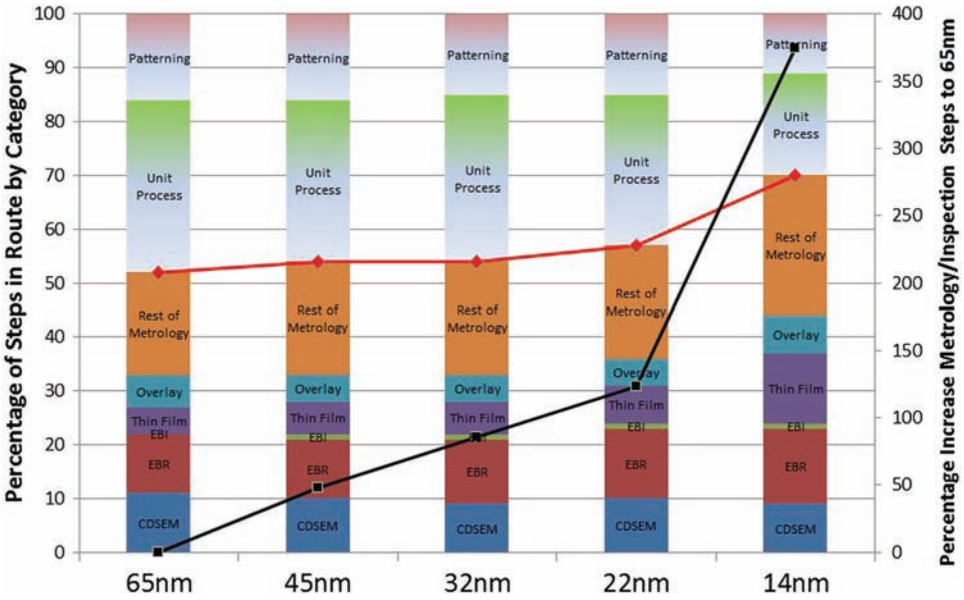


Figure 4.1 Percentage of steps in a route taken by metrology/defect process as a function of IBM technology nodes.

show the percentage increase in measurement/defect steps relative to the 65-nm node. Note that the number of metrology/defect steps almost doubles (~88%) from the 65-nm node to the 32-nm version, and quadrupled for the 14-nm node.

Note that not every lot/part that is manufactured goes through all of these metrology/defect steps each time. For the metrology measurement steps, if a given process step is stable (spread of data is well within the specifications, i.e., high Cpk), then there is no need to measure each lot. Likewise, if the process is not stable (spread of data is approaching or outside the specification, i.e., low Cpk), perhaps due to a very tight process tolerance, then it is likely that each lot is measured. Intelligent sampling is generally deployed to balance the risk in not measuring every metrology step in every lot versus measuring everything. A measurement recipe must be built for each measurement step in the route because they will all be used eventually to some degree. For the defect inspection steps, all of the steps are added to the route, but only a strategic fraction of those steps are turned on. As defect learning increases, these steps can be turned on or off as needed to maximize yield learning. Lots are sampled appropriately, and recipes are built as needed. Regardless, given the large percentage of steps taken up by metrology/defect steps, timely development of robust recipes for the metrology measurement and defect inspection steps are critical to the success of current and future chip technologies. This process is becoming much more challenging as significant

Table 4.1 Evolution of metrology techniques as a function of technology node to satisfy the increasing measurement demands coming from increasing complexity.

	65 nm	45 nm	32 nm	22 nm	14 nm	10 nm	7 nm
CD-SEM	X	X	X	X	X	X	X
Optical overlay	X	X	X	X	X	X	X
Film thickness (SE and R)	X	X	X	X	X	X	X
OCD (SE and R)		X	X	X	X	X	X
AFM		X	X	X	X	X	X
MBIR			X	X	X	X	X
XPS			X	X	X	X	X
LEXES			X	X	X	X	X
XRR				X	X	X	X
XRF				X	X	X	X
HRXRD				X	X	X	X
Hybrid					X	X	X
Speculation?							X

metrology measurement and defect inspection limits are being reached. Metrology is a very large part of the semiconductor industry and requires dedicated engineering to maximize its value in the fab.

Metrology's pervasiveness also increases over time with respect to the number of unique techniques needed to measure the increasingly complex technology nodes. Table 4.1 shows the evolution of measurement techniques needed as a function of technology node. There were only a few measurement techniques needed in the 65-nm technology node to satisfy all of the measurement demands. For the 22-nm technology node, the number of required measurement techniques has almost quadrupled. Metrology/defect tools in most state-of-the-art fabs account for 25–33% of all the tools in the fab. The amount of capital spent on metrology/defect tooling is usually 15–25% of the total tooling cost.

4.2 Metrology's Impact on Time to Market

Metrology's pervasiveness means that there are significant consequences on time to market. Insatiable metrology demand typically causes many metrology toolsets to be top cycle-time fab contributors. The primary effect of this insatiable metrology demand is over utilization, which increases the cycle time and therefore affects the "time to market." One of the keys to minimizing metrology's impact on the time to market is to ensure the utilization is sustainably maintained at an appropriate target. Figure 4.2 shows the relationship between toolset utilization and cycle time. As the utilization approaches 100% at the far right of the figure, the cycle time approaches infinity due to this exponential relationship. Therefore, the chosen utilization target must ensure that the fab cycle-time objectives are met. In this case, the target should be chosen as shown by the middle circular symbol so that the

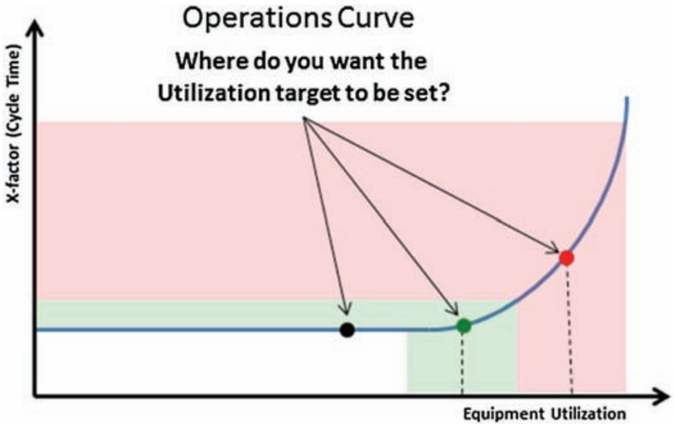


Figure 4.2 Relationship between toolset utilization and cycle time.

toolset may be used at the highest utilization target without negatively affecting the cycle time. This relationship between the toolset utilization and cycle time means that if the utilization target is poorly selected and allowed to exceed this value, each lot moves through the line slower, thus affecting the timely delivery of products to customers. When this happens to many metrology toolsets, each lot could be delayed days or even months, depending on the magnitude of the cycle time increase and the number of measurement steps affected. The development cycles of yield learning and the fab mean time to detect are also affected. When lots are moved slowly through the measurement tools, the yield learning data is obtained later, and the ability to detect fab issues is reduced.

4.3 Value of Metrology

Metrology primarily keeps processes centered in order to maximize yield and quickly detects process excursions that can detract the yield. A key measure used in many fabs is called process capability; the process capability for a given product is determined for each process step. In its simplest form, the process capability is the ratio of the process tolerance to the variability in that process. The higher the value is, the better the performance, thereby minimizing its impact on yield. The terms used to describe process capability are called the Cp and Cpk. The formulas for Cp and Cpk are denoted as follows:

$$Cp = \frac{(USL - LSL)}{6\sqrt{\sigma^2}}$$

and

$$Cpk = \frac{\min(USL - \mu, \mu - LSL)}{3\sqrt{\sigma^2}},$$

where *USL* is the upper specification limit, and *LSL* is the lower specification limit. Sigma is the measured variability of the given process. Cpk takes into account how well the process is centered about the process tolerances, whereas Cp does not. One of the key roles of the metrology engineer is to make sure their contribution to the sigma in the denominator of the Cp and Cpk formulas in minimized. If the precision to tolerance (P/T) ratio is not maintained, the impact on fab productivity can be significant. The Cp and Cpk formulas can be rewritten as follows for clarity:

$$Cp = \frac{(USL - LSL)}{6\sqrt{\sigma_{process}^{2} + \sigma_{metrology}^{2}}}$$

and

$$Cpk = \frac{\min(USL - \mu, \mu - LSL)}{3\sqrt{\sigma_{process}^{2} + \sigma_{metrology}^{2}}},$$

where sigma is separated into two components: the variability driven by the real process variation, and the variation driven by the metrology measurement error. The metrology sigma needs to be minimized by the metrology engineer to ensure the variability is true process variability so that reductions in the process variability can be achieved. Figure 4.3 illustrates the effect of measurement uncertainty on the Cp or Cpk for varying amounts of expected process variation. Improvements in metrology uncertainty are critical for tighter processes to achieve a good Cpk and, therefore, optimal process control. Note that the traditional rule-of-thumb value for a good Cp of 1.33 may be valid for

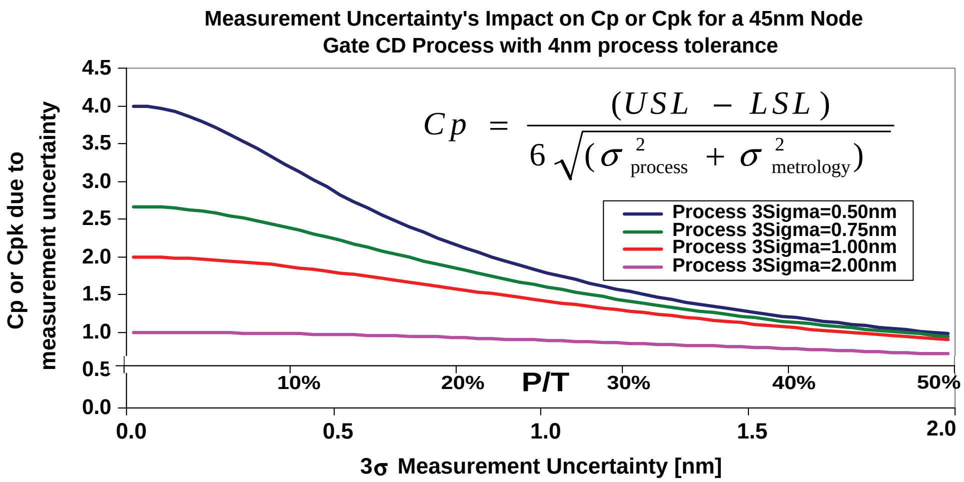


Figure 4.3 Measurement uncertainty's impact on the Cp and Cpk.

noncritical layers, but in the new paradigm of advanced process control (APC) and the need for as tight a process as possible, this value is insufficient; for tight processes, much can be gained by maximizing Cp.

Another way to look at the value of metrology is via APC feedback loops. For example, accurate, precise, and fast inline measurements of transistor-gate CDs and profiles are key drivers of world-class device performance. APC systems improve the electrical performance of devices by reducing the random CD and profile variation induced during predominant process steps. Real-time control loops feed inline CD and profile data to gate lithography, gate etch, and downstream processes (implant and anneal) to reduce L_{eff} variations and improve L_{eff} targeting. In fact, APC is a concept that gives metrology a key role in determining the quality of the final product. In the past, metrology monitored the outcome of processing and was primarily used to rework excursions and adjust processing "after the fact." APC improves this procedure by constantly guiding the processing with "mid-course corrections" to the target.

Figure 4.4 shows the typical gate-level APC loop. The metrology measurements are fed back to adjust lithographic settings, forward to adjust the gate etch based on lithography results, and can even be used to alter the RTA/implant steps to further correct for any excursions at the gate-etch step. This is a prime example of how metrology has become an integral part of the

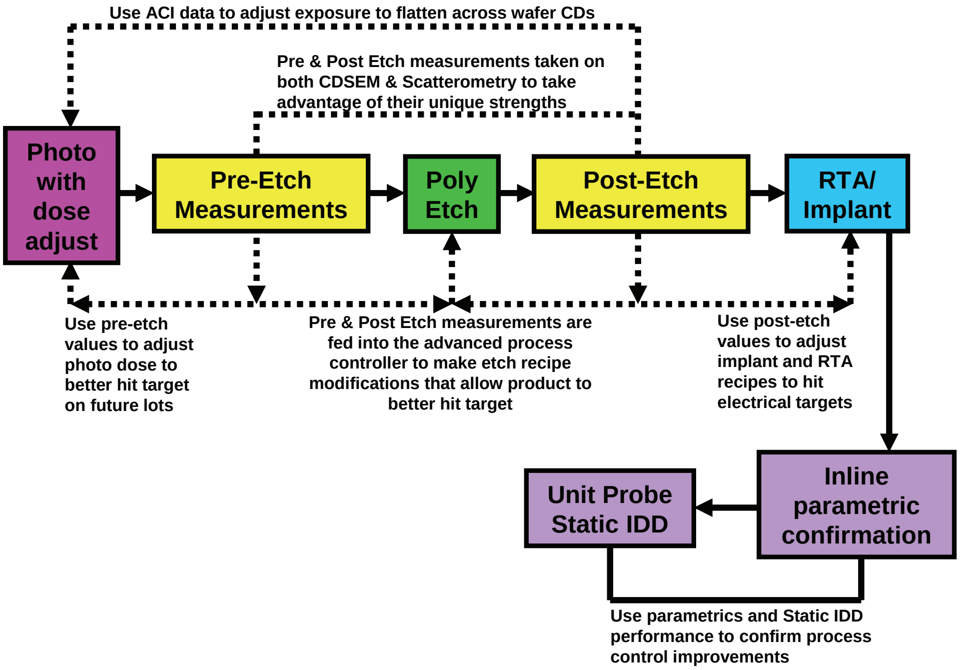


Figure 4.4 Flowchart for a typical gate-level APC loop. Note how metrology (pre- and post-etch measurements) is used everywhere.

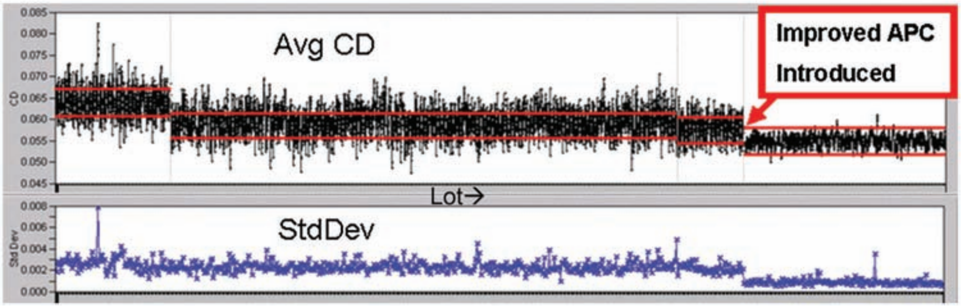


Figure 4.5 Demonstrated improvement in process control with the rollout of metrology-driven APC improvements in a gate module. Out-of-control sites (post-etch) were reduced from 32% to 2.5%, and the number of nonsaleable die was reduced by 85%.

modern lithography/etch process. If the metrology is inadequate, the scanners and etchers are out-of-control, expensive hardware; the metrology adjusts the entire flow toward success.

Figure 4.5 shows a 30% improvement in wafer-mean gate CD variation due to metrology-driven APC improvements. These improvements resulted in an 82% reduction of non-saleable die. Metrology data integrity is critical to realize the full benefit of APC and to prevent excursions caused by improperly fed APC loops. APC has demonstrated improvements across multiple semiconductor process areas.

One of the earliest adoptions of metrology APC involved the feedback of a correctable modeled overlay into the lithography stepper or scanner cells. It was realized early on that attempting to correct one of the overlay-modeled variables, such as "field magnification," using the actual modeled correction tended to overshoot the correction. As a result, partial corrections were employed in order to "walk" the correction toward zero with no overshoot. Currently, overlay tolerances as tight as 10 nm can be met in production with good overlay metrology and proper "tuning" of correction models.

Figure 4.6 is another example of how run-by-run control implemented on a thin-film toolset for post-polish ILD thickness can significantly reduce the process variability. In this case, the standard deviation was reduced by 52% and the Cpk improved by 56%. APC should be pursued wherever possible.

Figure 4.7 shows the APC trend for one sample fab. The x axis shows the technology node, and the primary y axis shows the number of metrology steps with and without APC. The secondary y axis shows the percentage of metrology steps with APC. Although the number of metrology steps using APC has increased as a function of the technology node, the percentage of metrology steps using APC has remained flat. This percentage must increase in the future for metrology to provide increased value. As mentioned earlier, linking the metrology data to the process in order to drive actions should be the goal of every metrology step.

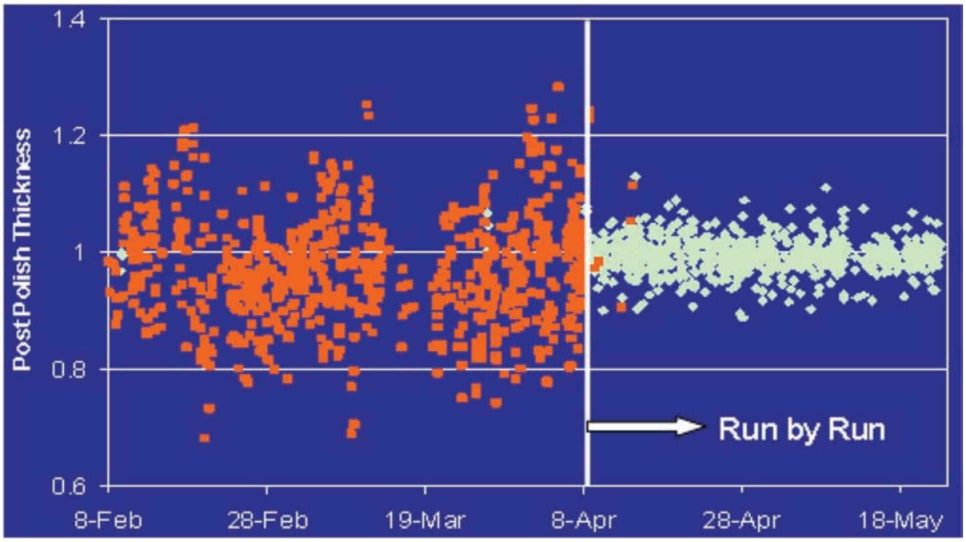


Figure 4.6 The post-polish ILD thickness-error standard deviation was reduced by 52% after run-by-run APC was implemented; the process capability (Cpk) improved by 56%. This example also shows that the same metrology principles are applicable to film metrology.

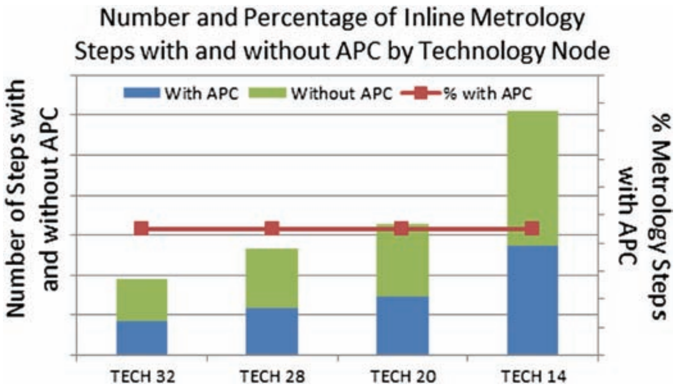


Figure 4.7 Displays the number and percentage of metrology steps using automated process control.

One more way to look at the value of metrology considers the impact of metrology toolset matching and long-term stability. In Fig. 4.8, the ultimate example of CD metrology value is seen, as the predictability of the speed sort maximizes the portion of parts that can be sold for higher profit margins. In this example, the metrology process can detect (and guide the rest of the process to create)—with high confidence—distinct product distributions that differ by 1-nm increments with little overlap. Because the CD in this product dictates speed and thus the value of the part, this scenario creates predictability of the different valued parts in the "speed sort." This is an

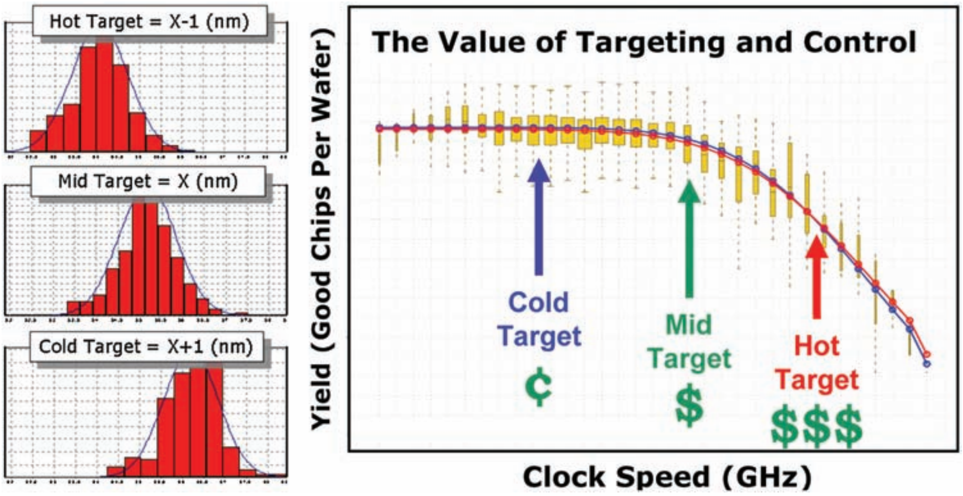


Figure 4.8 Quality CD metrology makes possible unprecedented levels of accuracy control, such as the ability to create distinct product distributions that differ by 1 nm with 6σ population distributions <2 nm, as shown on the left. The value of the targeting and control is shown on the right in the speed-sort yield distribution. As metrology improves to cross the steep part of the curve, the part value greatly increases.

example of process improvement, yield improvement, and improved knowledge of the parts produced.

The actual monetary value gained from the predictability of speed sorts, other final product characteristics, and the fine-tuning of the other process steps to attain these characteristics can be calculated through the improvement in yield. The model shown in Fig. 4.8 is one way to quantify the effect of measurement uncertainty on chip yield. Yield loss can be estimated using the number of wafer starts, the number of chips printed on the wafer, the defect-limited yield, the price per chip, the number of days in the year, and the difference in measurement performance [shown in Eq. (4.1)]:

$$\left(\frac{\text{No. of wafer starts}}{\text{day}}\right)\left(\frac{\text{No. of chips}}{\text{wafer}}\right)(\text{Defect-limited yield}) \times \left\{\begin{pmatrix}\text{Speed sort yield @ 0.5 nm,}\\ 3\sigma\,\text{measurement error}\end{pmatrix} - \begin{pmatrix}\text{Speed sort yield @ 1.5 nm,}\\ 3\sigma\,\text{measurement error}\end{pmatrix}\right\} \times \left(\frac{\text{Price}}{\text{chip}}\right)\left(\frac{\text{No. of days}}{\text{years}}\right). \tag{4.1}$$

Table 4.2 uses the cost model provided by Eq. (4.1) to illustrate one example of some reasonable estimates that are used in Fig. 4.9. The estimates were normalized to a 1000 wafer-start-per-day (WSD) fab. The provided inputs can be easily changed to predict the impact for any given

Table 4.2 Calculations based on Eq. (4.1).

Product	WSD	Chips/ wafer	Defect yield	Speed-sort yield @ 0.5 nm err	Speed-sort yield @ 1.5 nm err	Speed-sort yield delta	Cost/ chip	Days/ year	Cost/year
90 nm	1000	200	40%	65.70%	63.00%	2.70%	$50	350	$37,800,000
65 nm	1000	200	40%	65.40%	60.20%	5.20%	$50	350	$72,800,000
45 nm	1000	200	40%	64.70%	55.30%	9.40%	$50	350	$131,600,000

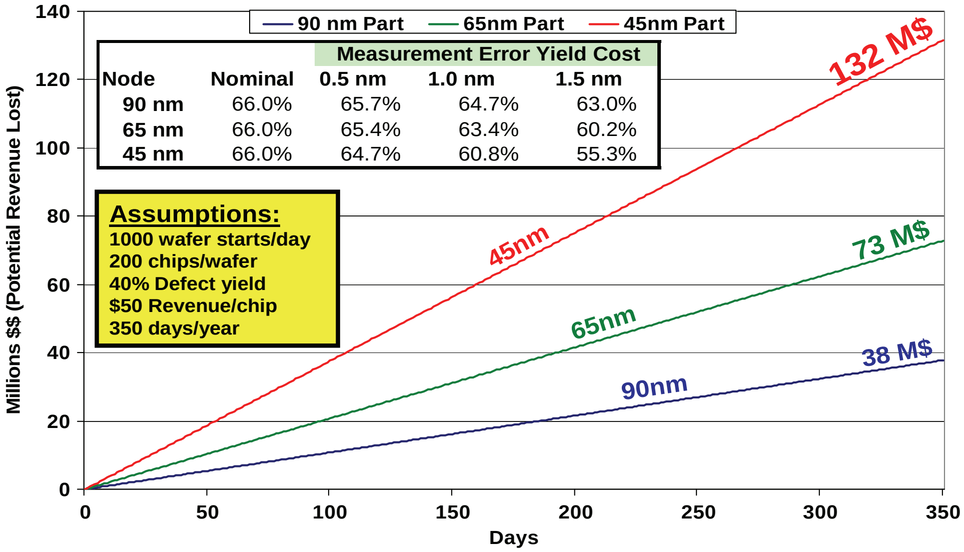


Figure 4.9 Revenue lost through CD metrology error effecting yield between two hypothetical fabs, one with a metrology error of 0.5 nm and one with 1.5 nm. Assumptions were selected based on reasonable to somewhat conservative values for 300-mm processing.

scenario. In reality, the accuracy of the inputs is not important; regardless of the fab's inputs, improvements to the measurement uncertainty quickly add up and improve the yield and therefore revenue. Figure 4.9 shows how two differing amounts of measurement uncertainty can affect the yield and thus revenue. These differing amounts, e.g., 0.5 nm and 1.5 nm, could represent differences in how well the CD measurement uncertainty in fab A is controlled relative to fab B for a gate process, respectively. Alternatively, they can represent the positive impact on revenue by improving the fleet matching through either improved practices that drive the tools closer together or by purchasing next generation measurement capability. The x axis is time in days, and the y axis is the estimated revenue lost by fab B versus fab A. The table inside the chart indicates the yield difference of the fastest parts between fab A and fab B as a function of measurement

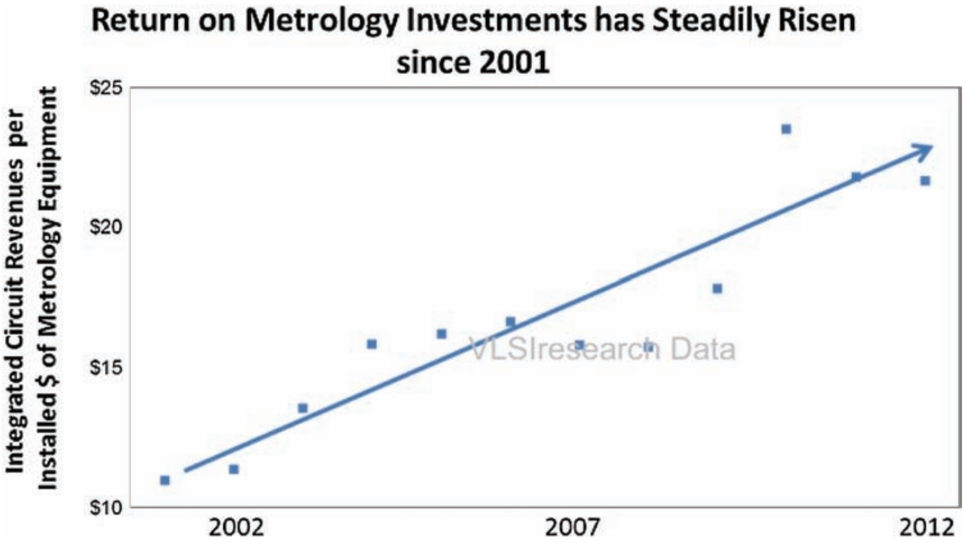


Figure 4.10 Metrology ROI increases since 2001 (data from VLSI Research).

uncertainty that lead to the revenue difference. Notice how the yield differences increase dramatically as the technology node shrinks, whereas both levels of measurement uncertainty remain the same. This situation causes the estimated revenue lost by fab B to dramatically increase because the measurement uncertainty becomes a larger fraction of the process tolerance, which means that more questionable product is sent through at the lithography step. In summary, fab B's higher measurement uncertainty prevents it from centering the line more aggressively to yield the more valuable, faster chips. The fab with the best metrology expertise, metrics, and methodologies will enjoy increased revenue through this relationship.

From a return-on-investment (ROI) perspective of metrology, VLSI Research has published a study showing a steady increase in metrology ROI since 2001, as shown in Fig. 4.10. In 2002, each dollar spent on metrology generated \$11 in IC sales. In 2012, each dollar spent on metrology generated \$22 in IC sales (double the return). That trend continues today. This data demonstrates the value of metrology, sometimes incorrectly regarded as a "nonvalue added process" in the semiconductor industry.

4.4 Metrology Target Design: an Element of Overhead

Metrology tools usually require dedicated targets that are typically spread out in the dicing/kerf area. There are a few exceptions to this, such as a CD-SEM, which can usually measure almost anything anywhere within a wafer. One of the primary reasons these dedicated targets are needed is because the beam/spot size and/or resolution of the illumination source is typically very large relative to the small dimensions that must be measured; therefore, these toolsets cannot and might never be able to measure the device structures directly, which is generally most desirable. For example, the state-of-the-art spot size of the illumination source for a thin-film or scatterometry tool is ~30–40 μm. In order to obtain a meaningful signal for analysis, the area illuminated needs to have certain properties. For a thin-film tool, the illuminated area must be free of any patterned features; otherwise, the signal will be influenced by the patterns, making it difficult to interpret the film stack. Special targets (protected areas) are thus designed on the order of 50 μm × 50 μm that are free of pattern, which allows the film stack to be measured in that area. Similarly, scatteromerty tools typically rely on an area with a periodic structure that is at least the size of the illumination area in order to get the information it needs to make proper analysis. Logic chips typically do not have large areas of undisturbed periodicity, so dedicated special targets are designed such that they mimic the sizes of the minimum features inside the chip. These targets are typically on the order of 50 μm × 50 μm.

Similar considerations are required for other metrology toolsets, such as overlay and x-ray tools, where the spot size and/or resolution limits may be larger or smaller than 50 μm × 50 μm. Given the large number of process steps that must be measured during chip fabrication, a significant amount of real estate is consumed by these target designs in the kerf/dicing area. These designs must be reviewed periodically to ensure that they are properly designed and that they generate useful data for process control, APC, and yield learning when they are measured.

There is an increasing desire to measure within the chip and not the kerf/dicing area for various reasons because the measurements from the target designs may not represent what is happening in the chip and therefore diminish the ability to respond appropriately to process shifts. This could be due to the target design having

- a larger dimension than the actual device dimensions,
- a different environment surrounding the targets in the kerf versus the device (they thus respond differently to process variation), or
- process influences that affect the target design in ways different than the device structures, such as CMP polishing.

Metrology tool suppliers have designed tools with a smaller spot size and/or better resolution in response to these issues. In response to this, metrology engineers are designing smaller and smarter target designs to better track the process changes within the chip. Additionally, where possible, shrunken target designs are being placed in the chip in strategic locations to better track the process across the chip field. Moving forward, innovative solutions are needed to generate smaller and smarter target designs that ensure high levels of process control.

4.5 Chip Scaling and an Introduction to Some Key Metrology Toolsets

The semiconductor industry, as described by Moore's law, has progressed tremendously in the last 50 years. Not only has the silicon-based-CMOS transistor density per chip increased from a few hundreds to more than billions but also the minimal feature (usually a transistor gate length) has been reduced from more than 2 μm to less than 20 nm today. As a result, the manufacturing process margins (or process windows—allowable process variations within the process specifications) have been shrunk continuously. The demand for metrology and inspection capable of process development and control has increased drastically. Only a handful of technologies are capable of measuring dimensions less than 100 nm; among them, only SEM-based metrology has been production proven for the last 25 years and is still the dominant choice for dimensional metrology.

The CD-SEM has been a key part of chip development and manufacturing for a long time. One of the primary reasons for this is because it is an image-based tool, meaning that the measurement is made directly from the acquired image. Unlike OCD, which is a scattering technique (also referred to as a computational or model-based tool), or possibly a future CD-SAXS (critical-dimension small-angle x-ray scattering) tool, what you see is what you get. Almost any single feature can be measured in any local environment. In contrast, OCD requires a large dedicated grating that is representative of what must be measured and ultimately averages the CD across a large area. This process has proven very valuable over the years because it is often necessary to image and/or measure discrete devices or structures to help understand local process variability and drive improvements. Because of this critical fundamental need, the CD-SEM has had a long and illustrious career, and its longevity is sometimes taken for granted.

Over the years, a large amount of activity has centered on the CD-SEM. Figure 4.11 summarizes the entire history of the inline wafer CD-SEM as best the authors could determine. The CD-SEM started becoming the primary CD measurement tool in the early 1990s. Prior to this time, optical tools were actually image-based instruments that could image the discrete device/structures and measure the CD directly from the image. In the late 1980s and early 1990s, the resolution required to image and measure the smallest devices/structures was beyond the optical tool's capabilities. Given the CD-SEM's resolution advantage, it began taking over these applications. CD-SEM tools completely replaced optical CD tools by the late 1990s. This means that the inline wafer CD-SEM has now been around for more than 20 years. During this time period, approximately 36 products were offered across at least 11 companies. (Refer to the details shown in Fig. 4.11 to better understand all of the changes and milestones that occurred.) As of 2015, only three suppliers remain: Applied

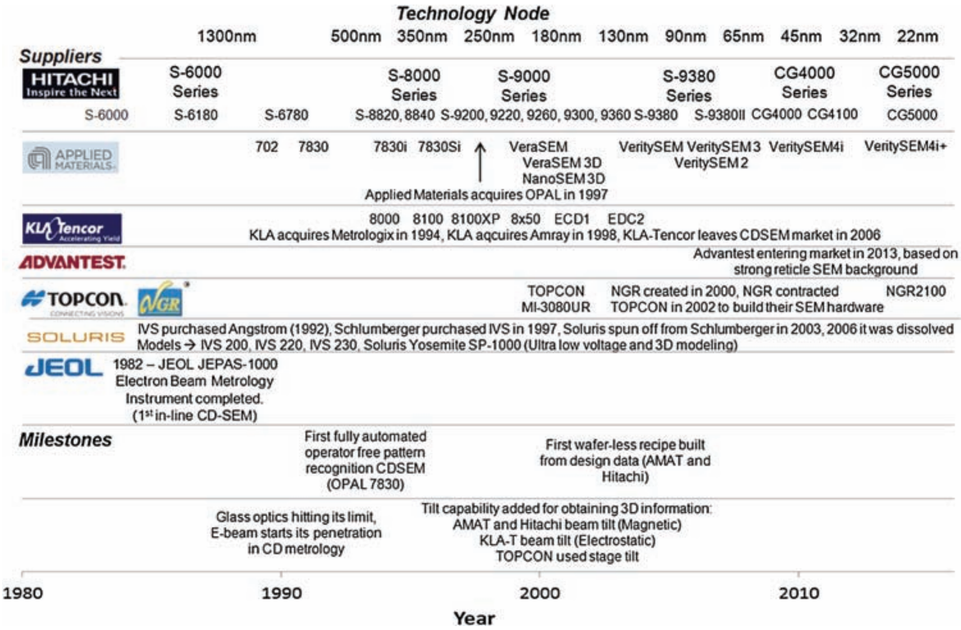


Figure 4.11 History of the CD-SEM.

Materials, Hitachi, and Nanogeometry Research (NGR). Advantest is entering the market based on their experience in SEM-based CD measurement on masks, and Holon Co. remains a CD-SEM supplier for masks.

CD-SEM metrology has progressed along with the IC industry over the last 30 years in resolution and precision, measurement algorithms and accuracy, automation and throughput, and charging reduction and tool matching. It will be the CD metrology tool of choice for foreseeable future.

Because the increased complexity of IC has scaled down in recent years, IC manufacturers (especially leading IC foundry companies) have accelerated process development to work on two nodes (or even three in anticipation of major design or lithography tool changes) simultaneously, instead of one. To do that, metrology tools must be ready ahead of the industry development curve, which is very hard to do. To gain advanced process data and samples, as well as share increased development costs, equipment venders and IC manufacturers have closely collaborated via various consortia.

In addition to CD, there are many other metrologies in other IC process controls, e.g., film-thickness measurements, overlay (the alignment error of two adjacent lithographic layers) measurements, etch-depth measurements, and many *in situ* measurements for process control, as well as electrical measurements (e.g., resistivity measurements).

in situ metrology (integrated metrology) is built inside process equipment or integrated in the process equipment or equipment clusters for dynamic process control. For example, advanced lithography scanners can dynamically detect

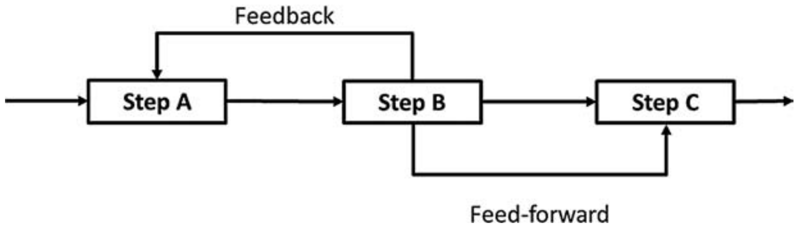


Figure 4.12 Typical metrology feedback and feed-forward loop in process monitoring and control.

the local distance of the scan area on wafer surfaces to the final lens and adjust its focal plane accordingly. Another example is ion flux monitoring in an ion implant machine to control the implant dose. Another important application of *in situ* metrology is feedback and feed-forward loops to dynamically adjust the previous process step (feedback) and the next process step (feed-forward), so that certain characteristics of IC devices can be adjusted for tighter control. Fig. 4.12 shows typical feedback and feed-forward process loops.

4.6 Vision System and Recipes

One key aspect present in a metrology tool not present in process tools that can add complexity is a vision system. No process tools except for exposure systems have a vision system. Because the metrology tool must precisely navigate to the measurement area, a vision system is required to calculate the errors in translation and rotation when the wafer is loaded into the tool. The basic elements of a recipe on any metrology tool are illustrated in Fig. 4.13.

The first step creates a wafer map, which is typically created once per product and shared across all recipes built for that product on the given toolset. The next step is global alignment, which corrects for the translational and rotational errors. The tool then drives very close to the measurement area and performs pattern recognition to further correct for navigational errors; it then focuses and measures. This process repeats across all programmed sites

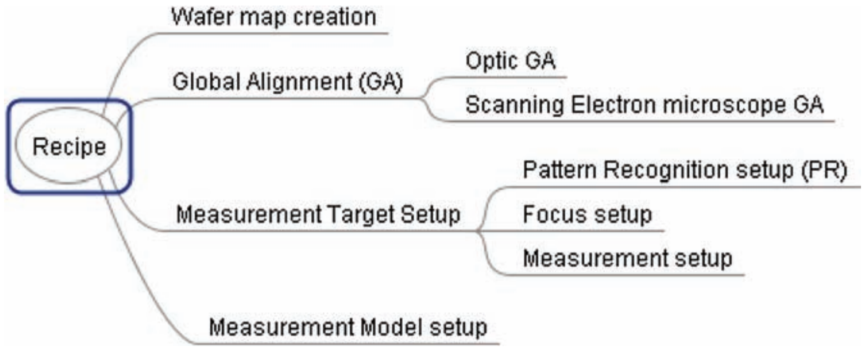


Figure 4.13 Brach diagram of the basic building blocks of a metrology recipe.

to be measured. Some metrology toolsets, such as thin film and scatterometry, require a model be built ahead of time, which adds complexity because there are opportunities to fail at any of these steps that can disrupt product flows throughout the production line. Metrology engineers must also ensure that all recipes are built as robustly as possible. There are always recipes that fail for various reasons—the metrology engineer must identify the reason, correct the issue, and determine if the solution should be applied elsewhere to prevent the issues from occurring again. In contrast, when a wafer is loaded into a process tool, there is no vision system to complicate things: the wafer goes in, processes, and comes out.

The vision system drives additional complexity over process tools. Because each product usually has a different chip size for a given technology node and therefore must navigate to different spots to find the measurement structure (even it is the same measurement step between products), recipes must be built for each product. For example, a CD-SEM has approximately 100 measurement steps in the 32-nm technology node, which means that 100 recipes must be built. If a fab has 20 different 32-nm products, then 2000 recipes (100×20) are required. If multiple technology nodes are running in the fab and producing many different products, the number of recipes can exceed 10,000 or more. This number does not include all of the recipes that must be built on the other metrology toolsets, such as thin film and overlay tools, which is also in the thousands. In contrast, the same recipe can be used on a process tool across products, reducing the number of recipes that must be built and supported by several orders of magnitude. It is a challenging task to maintain the performance of these recipes and ensure they always perform optimally.

Recipe errors can come from many different sources. It may not make sense to repair the recipe without some sort of additional investigation. Figure 4.14 lists all of the major factors that can cause recipe errors.

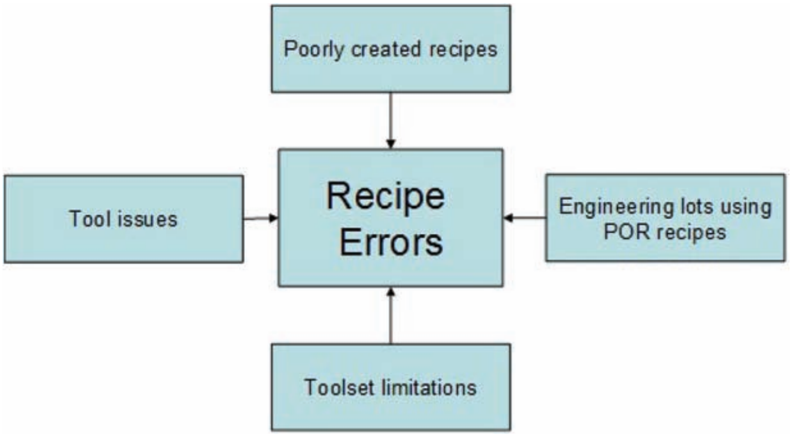


Figure 4.14 Recipe error contributors.

One source of recipe errors is poorly created recipes. Both manual and automatic production will result in a population of recipes that contain some poorly created recipes. In the former case, the experience of the recipe writer is generally to blame. A fab often cannot pick and choose the quality of recipe writers in the team, and a lot of experience is needed to understand what to do in each situation. Lack of experience leads to poor decisions that can negatively affect the recipe quality. In the case of a recipe created automatically through design data, algorithm intelligence limitations prevent the system from choosing an ideal combination of recipe components. For example, a poorly chosen pattern-recognition template or a poor focus location will cause recipe errors. Tool issues is another possible source of errors. Recipe errors can occur because one or more tools in the fleet is drifting or was poorly calibrated. For example, if the optical illumination of one tool drifts with respect to the others, then more global alignment failures would result from the tool that drifted. Another example occurs when one tool's stage is not navigating the same as another tool's stage, which may cause recipe errors in the form of pattern recognition errors because the problematic tool will not center the image properly in the field of view. Meanwhile, other tools running that same recipe on other lots have no issues. It is important to make sure that a tool issue is ruled out before repairing a recipe; otherwise, the recipe is assumed to be repaired when in fact there is another source for the recipe errors. In this case, it is advisable to shut down the offending tool and address the issue.

Toolset limitations can also cause recipe errors, although they are sometimes confused with poorly created recipes. For example, some navigation requirements are so challenging that it is not practical to build robust recipes to the point where they will not fail. Recipe errors resulting from this situation can only be overcome by fundamental improvements to the given toolset made by the supplier. Lastly, engineering lots could be run through the line using process-of-record (POR) recipes. In this case, the engineering lot may be missing and/or have extra process steps that make these jobs look abnormal. The engineering lot is measured on the metrology tool, and the recipe fails. These recipes fail because the wafers do not look like the POR wafer used for the initial recipe construction. The only solution here is to have a system that separates the recipes created for use on engineering lots and those used for POR lots. It is also important to note that POR recipes should not be "corrected" using these engineering lots because the POR lots will then fail.

In summary, investigation is needed to understand the root cause of the recipe errors so that the appropriate action can be taken (productivity suffers otherwise). The recipe should not be repaired until it is determined that this is the proper course of action. In recent years, waferless recipe writing has helped eliminate the need to hold the wafer to build a new recipe and help make them more robust. Design data is used to create a recipe without the wafer and is briefly discussed in the last chapter.

4.7 Toolset Recipe Portability Monitoring

Toolset recipe portability monitoring is the quest to understand all possible aspects of the hardware that can affect recipe performance. Ideally, when a recipe runs on one tool, it should run exactly the same on another tool in the fleet. This does not happen for many reasons. For example, each tool's stage may be navigating differently, making one tool more likely to fail pattern recognition because the pattern recognition feature is not centering in the field of view as well as the other tool. Additionally, if the optical microscope illumination conditions are not matched well between tools, there is a higher likelihood of global failure on one tool versus the other. These are just two examples of how hardware calibration differences or how hardware drifting over time can affect the overall success of the recipes. Figure 4.15 shows an example of hardware elements of a CD-SEM that can affect recipe performance. Many of these elements also affect other metrology tool types, and some may be unique to certain metrology tool types. It is up to the metrology engineer to determine what should be tracked for the given tool type.

Ideally, information about all of these hardware elements is automatically generated on the given metrology tool. It is better if this information is reported every time a measurement is made and then stored in the data file on the tool and reported to the fab database. Figure 4.16 shows an example of the information that can be extracted automatically on a CD-SEM from the pattern recognition, measurement, and global alignment image. For example,

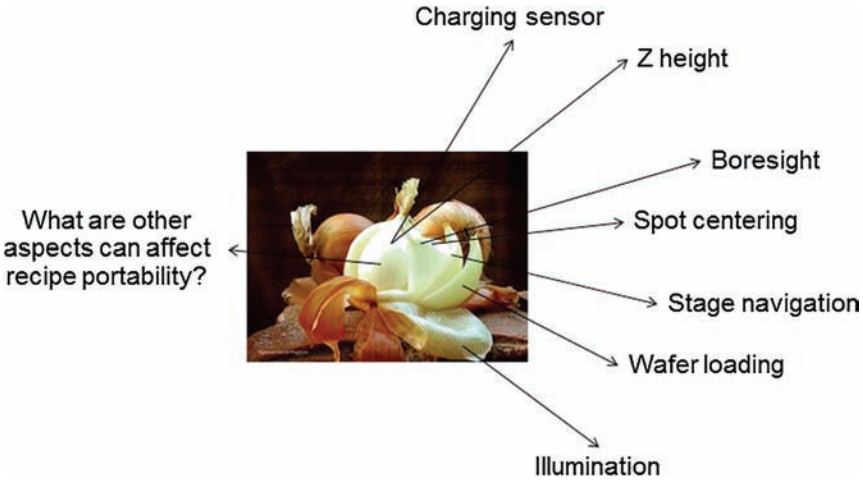


Figure 4.15 Hardware elements on a CD-SEM that can affect recipe performance.

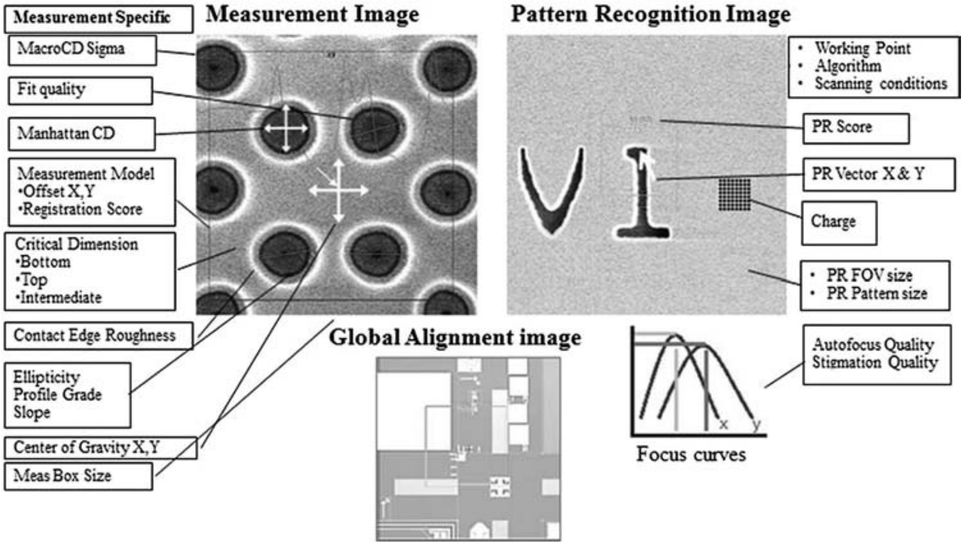


Figure 4.16 Information that can be extracted from images.

the PR vector X and Y, which represents the stage navigational error in x and y, is calculated from how well the pattern recognition center of gravity coincides with the center of the field of view. If they are perfectly coincident, then the navigational error is zero; if not, the X and Y displacements can be reported because the pixel size in the image is known. A monitor using the same wafer and recipe across all tools can then be used to measure many different sites; the navigational error measurement at each site can be reported and compared to a specification, as shown in Fig. 4.17. Note that tool 2 has many points outside of the specification compared to tool 1. If this data is reviewed regularly many times, one tool's cluster will appear to grow over time, and the engineer can proactively drive an action to address this situation before it significantly affects recipe performance and causes productivity issues.

If the toolset supplier reports this data automatically, it can be retrieved from the daily monitor already used to ensure measurement stability. Many other parameters can be monitored in similar ways, such as the optical illumination matching, wafer loading errors, and beam centering errors. If the supplier does not report these values automatically, then images can be arranged side by side across tools to subjectively evaluate these parameters, as shown in Fig. 4.18. These images are taken from the same wafer and recipe and represent the first global alignment image from each tool in the fleet. The first global alignment image carries information about illumination matching and wafer loading errors. Tool 1's image is much darker than tool 10, which will affect the global alignment performance of recipes. Tool 8's image is shifted further left than tool 9, illustrating wafer loading differences that could

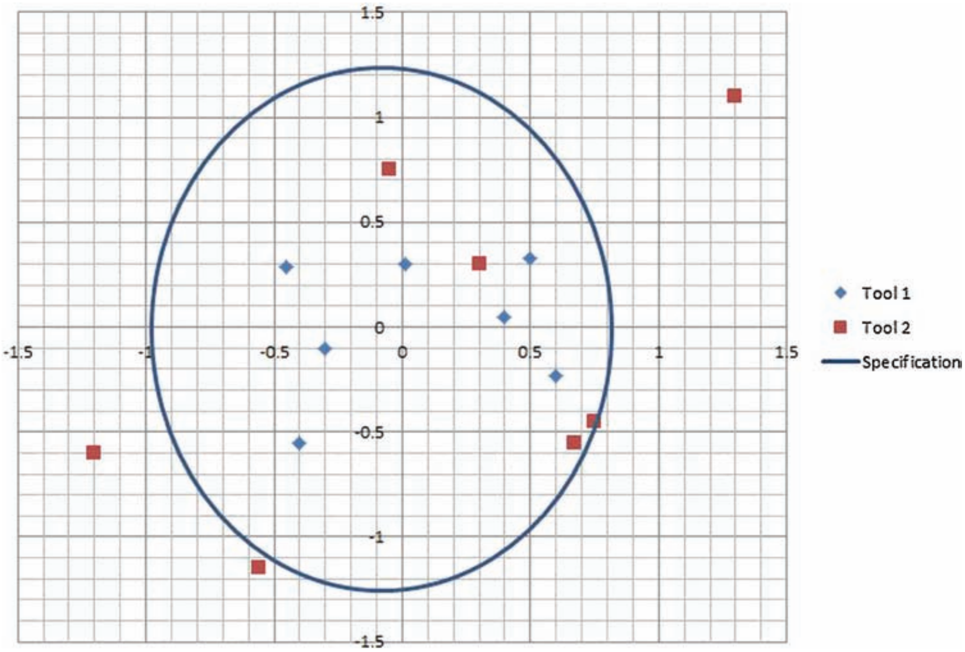


Figure 4.17 Navigational errors reported from two tools using a monitor recipe.

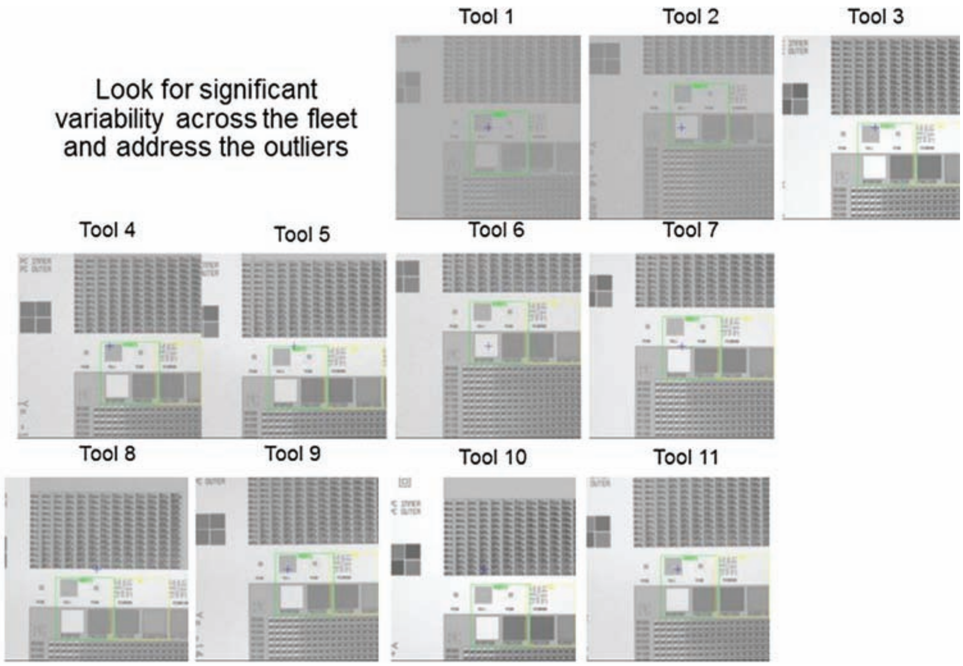


Figure 4.18 Global alignment images taken from the same wafer and site to compare illumination and loading errors across tools.

also affect the global alignment performance. When reviewed over time, these issues become obvious and can be proactively addressed.

The goal is to have the supplier automate this data output and analysis, review data regularly for each key hardware parameter, and drive continuous improvement based on the data. The metrologist can then determine how well the significant parameters can be controlled and then tighten specifications beyond supplier commitments over time to get the most from the toolset. Then the metrologist can continue to peel the onion for the next largest contributors and provide feedback to the supplier for future improvements.

4.8 Large Fleet Sizes: Fleet Management

For the most often used metrology toolsets in a fab with a large number of wafer starts, such as CD-SEM, overlay, and thin film, each fleet can contain fifty or more tools. Ideally, all of the tools in a fleet must act as one. Many activities need to be performed across the fleet by the metrology team:

- creating and optimizing recipes,
- optimizing measurements,
- qualifying tools,
- matching tools,
- monitoring tools,
- optimizing tools,
- determining problem areas, etc.

For any fleet of metrology tools, most of these activities can be placed into one of two fundamental categories: tool matching and recipe quality. Tool matching means ensuring measurement matching across the fleet of tools and proper tool calibration to allow recipes to run similarly across the fleet of tools (the latter is more appropriately defined as portability matching).

Recipe-quality management involves strategies in measurement and recipe optimization, and also determines which recipes chronically fail over time. Maintaining consistency among recipes where appropriate is desirable but hard to manage. Measurement optimization involves obtaining the best measurement accuracy, precision, and matching by optimizing the parameters that affect measurement quality for each application. Recipe optimization also means obtaining the best recipe performance by optimizing the recipe parameters to yield a robust running recipe. Many productivity detractors arise because of the difficulty in optimizing, creating, and managing a large number of recipes (as well as matching and monitoring a fleet of tools). In the end, both tool matching and recipe-quality management require significant attention over time. An overall strategy is needed in order to minimize productivity detractors caused by a lack of comprehensive solutions available today that address the root cause.

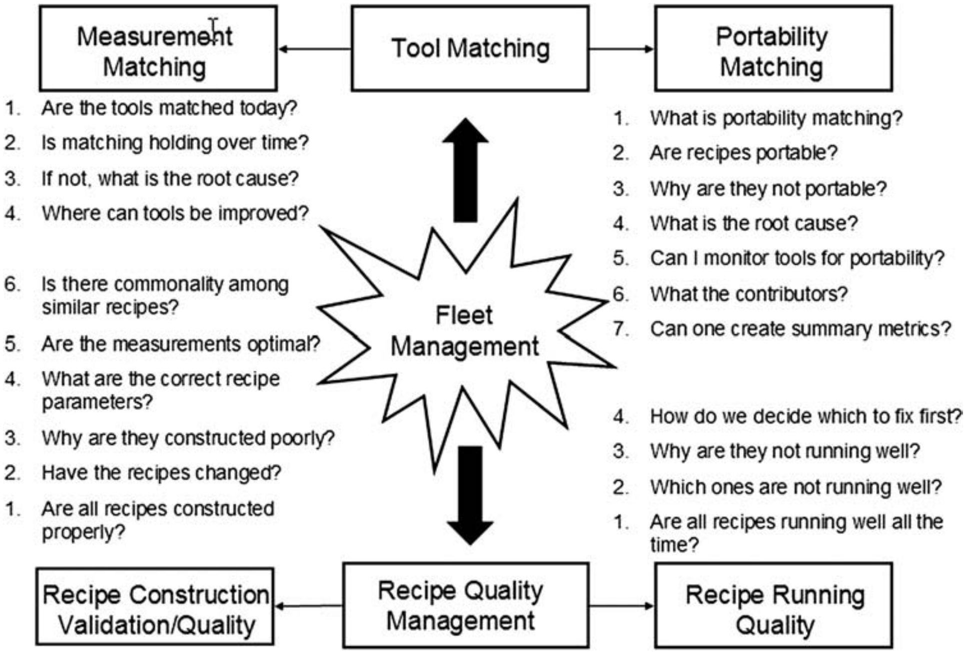


Figure 4.19 Fleet-management function blocks and questionnaire.

Metrology tool-fleet management is a strategy that captures the pursuit of minimizing productivity detractors arising from the fleet of metrology tools to optimize their performance. More specifically, fleet management collects the metrology tool activities mentioned earlier that are logically grouped so that solutions can be more easily realized. Figure 4.19 displays the categories of fleet management. Vertically above and below fleet management's core is tool matching and recipe quality management, respectively. Tool matching is further divided into two categories: measurement matching and portability matching. Recipe-quality management is also divided into two categories: recipe construction validation/quality and recipe running quality. Above and below each corner category is a series of questions that must be answered. If solutions exist to answer all of the questions under each category, then comprehensive fleet management has been achieved, and the productivity detractors have been minimized.

In reality, this is much easier said than done. Fleet management never ceases because there will never be a perfect fleet of tools. Spending the time necessary to create comprehensive solutions for each category is critical to minimize productivity detractors and realize the full potential of the fleet. Furthermore, all tools have their limits, which are more accurately determined by these comprehensive fleet management solutions so that future improvements can be more effectively implemented either on the existing fleet or the next generation of tools. The following sections of this chapter describe each of the aforementioned categories: measurement matching, recipe construction validation/quality, recipe portability matching, and recipe running quality.

4.9 IC Process Development and Control

The IC process goes through the following phases:

- development,
- early production (often called pilot line production), and
- process control in production.

The differences in each phase reflect in its process capability (measured by the process capability index) and its sampling plan. In the early process-development phase, metrology is heavily used to characterize and quantify all variations and modify the process to minimize some of the variations. Each time the process changes, another round of variation characterization and quantification occurs (with oversampling on metrology tools) until the process is stable and centered with a reasonable process window and process capability. Thus, the process development phase has the characterization of unstable and heavy metrology (or sampling). As the process becomes more stable and its variations (measured by the process distribution sigma) become smaller, the process moves to the early production phase (or pilot production phase). This phase has a process capability index larger than one, whereas the sampling plan can be reduced to process monitoring and control. However, the sampling plan in the early production phase intentionally oversamples in order to monitor all variations, rather than the main or dominant variations.

As the process becomes more mature and process variations are well controlled, the process enters the volume production phase. This phase has a process capability that is typically larger than 1.33, whereas the sampling plan has been minimized to only monitor a few major process variations, not all of them. All process-related parameters are usually fine-tuned and frozen in this phase to guarantee a good yield at the end.

IC manufacturing processes are generally batch processes—all wafers in a cassette or other wafer container (e.g., FOUP) are processed together at the same process step. In the era of sub-half-wavelength lithography, all RETs, especially OPC, become necessary in all designs to extend the lithography process window. OPC model calibration is an essential part of OPC. To get a good OPC model, an extensive set of modeling data must be acquired, which may include

- normal 1D line/space measurements with different pitches and duty cycles (L/S ratio), from isolated to fully nested (equal lines/spaces);
- slightly more complicated measurements (people sometimes refer to them as 1.5D features), such as a line/space end to a line/space and a line/space end to a line/space end of a different size; and
- some 2D features, such as line contact holes and posts with various sizes.

Sometimes the OPC model wafer may be shot with an expose dose (E) and focus (F) variations to perform process-window-aware OPC modeling, which will add additional CD measurements at different FE conditions. Such extensive CD measurements require a lengthy CD-SEM recipe setup and CD measurement time.

The desire to shorten that time has pushed people to explore other options, one of which attempts to use a SEM-image-based model-calibration method. The traditional model CD fitting becomes SEM image contour fitting (minimizing contour-to-contour EPE errors). Much less time is required to take a few SEM images than thousands of CD measurements. However, the burden is shifted to model-pattern selection—how to select one that covers enough pattern variations to ensure good model predictability?

4.10 Metrology in IC Manufacturing Process Control

The majority of modern IC semiconductor devices are CMOS-based digital devices, such as microprocessors and memory devices, which are manufactured layer by layer through modular process steps, as shown in Fig. 4.20. Those process steps consist of distinguished process modules, such as lithography, etch (dielectric layer etch and conductor layer etch), film deposition (dielectric layer and metal layer), diffusion and ion implantation, chemical and mechanical polishing (CMP), rapid thermal process (RTP), wafer cleaning and metrology, and inspection. Metrology and inspection are an integral part of IC manufacturing processes. A cross-section of the latest 16-nm CMOS device shows that layer-by-layer structure (Fig. 4.21).

IC manufacturing processes are normally divided into two major parts: the transistor formation, called the front end of line (FEOL), and the

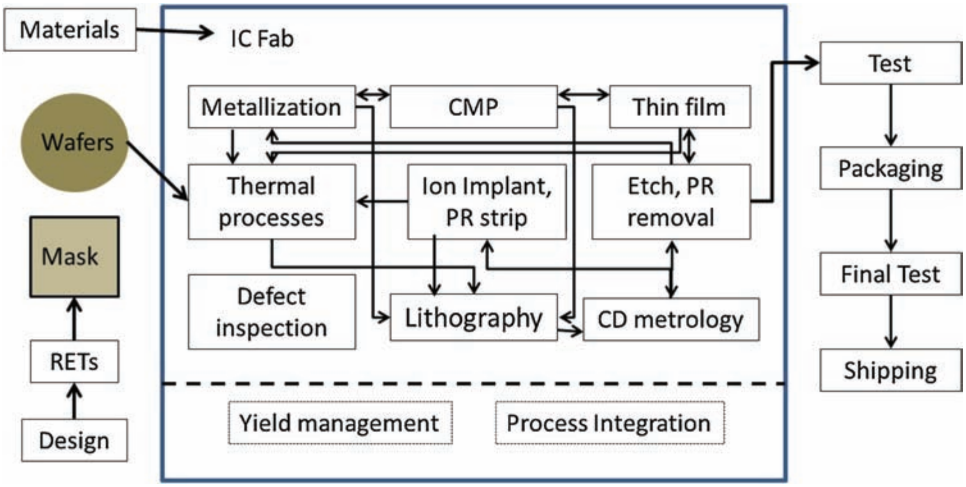


Figure 4.20 Block diagram of process modules in an IC fab.

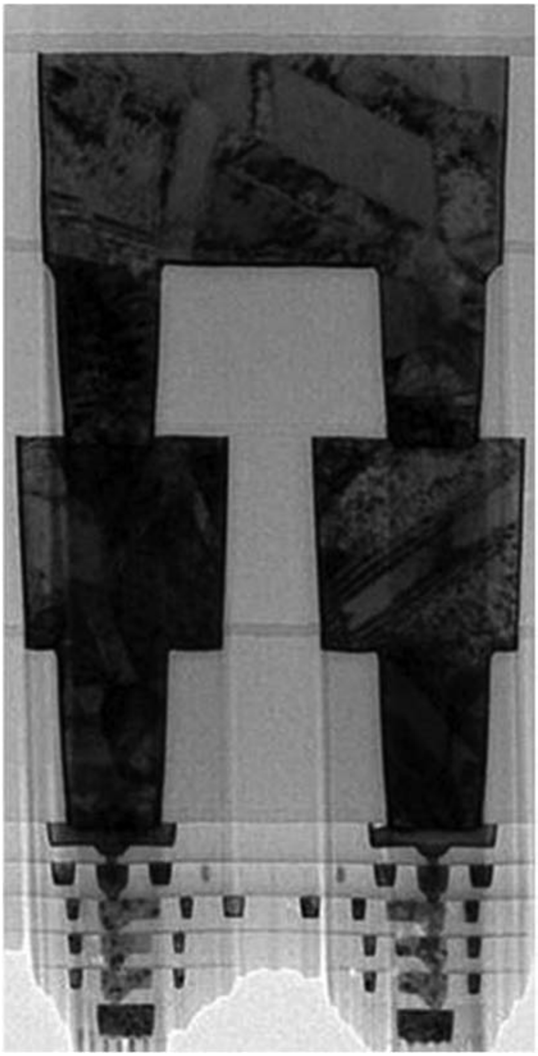


Figure 4.21 Cross-section SEM image of a CMOS device (16-nm FinFET SRAM, TSMC, IDEM 2013).

interconnected formation, called the back end of line (BEOL). The former refers to the bare silicon wafer to the first contact layer, whereas the latter refers to the first metal layer to the passivation layer (the top seal layer before the windows open to an outside wire connection). In order to get a good yield at the end of the IC process, every critical process step must be monitored and controlled through a careful metrology sampling plan and SPC control.

Lithography and etch processes are two critical processes in IC manufacturing to transfer the design patterns to wafer surfaces one layer at a time through photomasks. Thus, one of the measures of manufacturing flow speed is how many hours (or day) each mask layer requires. For lithography and etch processes, CD and overlay are two critical parameters used to monitor and control the processes. The other critical parameter is defect density, which is monitored and controlled using defect inspection systems. Other forms of IC manufacturing inline metrology involve film-thickness measurements for deposition processes, in addition to CD measurements and overlay measurements.

References

1. K. M. Monahan, Ed., *Handbook of Critical Dimension Metrology and Process Control*, SPIE Press, Bellingham, WA (1993).
2. B. Bunday et al., "Realizing 'value-added' metrology," *Proc. SPIE* **6518**, 65181K (2007) [doi: 10.1117/12.714216].
3. E. Solecky et al., "In-line E-beam wafer metrology and defect inspection: the end of an era for image-based critical dimensional metrology? New life for defect inspection," *Proc. SPIE* **8681**, 86810D (2013) [doi: 10.1117/12 .2010007].
4. E. Solecky et al., "Beyond measurement uncertainty: improving the productivity of metrology tools through recipe error analysis," *Proc. SPIE* **6518**, 65181J (2007) [doi: 10.1117/12.712850].
5. E. Solecky, "Metrology tool fleet management: a comprehensive discussion of requirements and solutions," *Proc. SPIE* **6152**, 61520U (2006) [doi: 10.1117/12.659589].
6. G. D. Hutcheson, "Metrology's value in the semiconductor industry," presented at ICFCMN 2013 by VLSI Research (2013).

Chapter 5
Metrology Toolsets in IC Manufacturing: Optical Metrology

5.1 Optical Film Characterization: Thickness and Composition

There are many different films in IC processes whose thicknesses must be tightly controlled following each deposition or growth step. In the FEOL and MOL, there are isolation films, such as silicon oxide (SiO_2) and silicon nitride (SiN); poly silicon; gate thin films, such as titanium nitride (TiN) and tantalum nitride (TaN); gate oxide; high-*k* oxide (usually involving hafnium oxide); etc. Low-*k* dielectric films and barrier films are deposited in the BEOL of the Cu damascene process. Furthermore, advanced transistors use compressive and tensile epitaxial films, such as SiGe (Narasimha[1]) and SiP, within source/drain regions to enhance the performance of advanced node devices (45-nm node and beyond technologies). All of these films require precise and accurate measurement of thicknesses and concentration to ensure tight process control of these processes. Spectroscopic ellipsometry (SE) and reflectometry are two major categories of optical systems used in the semiconductor industry to measure and monitor these films.

Most of the films mentioned here have varying levels of transparency (at least in the thickness ranges typically used in the semiconductor industry), and ellipsometric thin-film measurement systems are commonly used to measure them. Ellipsometry measures the change of polarization upon reflection or transmission; in most optical tools, it also compares results to a pregenerated model to predict the optical parameters of interest. This process is typically performed during the reflection set up of semiconductor metrology ellipsometers. The polarization change of light is determined by the sample's (structure on wafer) optical properties, such as thickness, complex refractive index, or dielectric function, commonly referred to in the industry as *n* and *k*.

When light passes through a medium, some of it will always be absorbed. This phenomenon can be conveniently accounted for by defining a complex index of refraction: $\tilde{n} = n + ik$. Here, the real part of the refractive index n indicates the phase velocity, whereas the imaginary part k indicates the amount of absorption loss when the electromagnetic wave propagates through the material (commonly called the extinction coefficient). Both n and k vary as a function of wavelength, and their variation pattern as a function of wavelength is called optical dispersion. The concentration/composition of films typically affects their dispersion.

Both a reflectometer and ellipsometer use the optical technique to measure film thickness and composition. Composition is estimated by assessing its impact on dielectric functions (n and k), and are therefore an indirect result of comparing the ellipsometry's experimental results and the model's predicted results. This is one of the reasons why optical metrology techniques are commonly referred to as "model-based" techniques, where the results, to some extent, depend on the model quality and characteristics (instead of a less-model-based, direct measurement technique, such as CD-SEM). The principle of an ellipsometer is illustrated in Fig. 5.1; this sample shows a rotating analyzer ellipsometry (RAE) setup.

Light emitted by a light source is shown to be linearly polarized by a polarizer (in reality, it could be polarized to any predetermined state). It then passes through a compensator (optionally, a retarder or quarter-wave plate) and reaches the sample (structure on wafer). After reflection, light passes through a compensator (optional), reaches an analyzer (second polarizer that determines the polarization state of light), and finally falls into the detector. The detector could be a prism followed by a CCD array, which would register the light's state in voltage as a function of time to interpret results. The incident and the reflected beam span the plane of incidence. Light that is polarized parallel to this plane is called p-polarized, whereas the perpendicular version is called s-polarized.

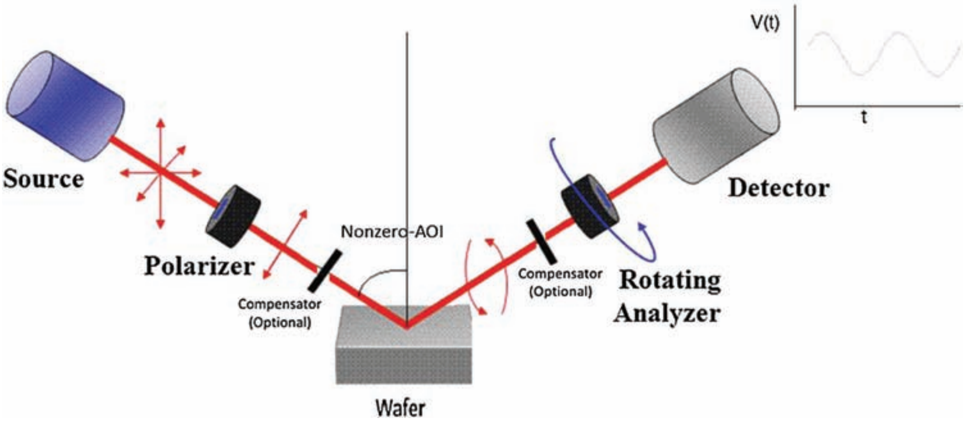


Figure 5.1 Illustration of the ellipsometric principle.

SE measures the ellipsometric parameters (Ψ, Δ) that are related to the ratio of the reflectivity of *p*- and *s*-polarization [see Eq. (5.1)]. Multilayer modeling analysis on the ellipsometric angles can estimate the complex dielectric functions that are the square of the complex refractive indexes. Model optimization typically includes regression-based chi-squared minimization to obtain parameters of interest, such as thickness and optical dispersions (used to compute composition). Like dispersions, experimental data ($\tan\Psi$ and Δ) are collected as a function of wavelength and are typically termed as spectra.

$$\rho = \frac{r_p}{r_s} = \left|\frac{r_p}{r_s}\right| \exp(i\Delta) = \tan\Psi \exp(i\Delta), \tag{5.1}$$

where ρ is the complex reflectance ratio that ellipsometry measures, r_s and r_p are the reflectance of the *s*- and *p*-components of light, respectively; $\tan\Psi$ is the amplitude ratio upon reflection; and Δ is the phase difference ratio upon reflection.

There are two kinds of ellipsometers commonly used in semiconductor optical tools: single-wavelength and spectroscopic. Single-wavelength ellipsometers use a monochromatic light source, such as a laser that runs at specific wavelengths, whereas spectroscopic ellipsometers use broadband light sources that can potentially span from the deep ultraviolet (DUV) to the visible and as far as the infrared (IR) wavelength regimes. The advantage of single-wavelength ellipsometers is that they often have very good precision (for Ψ, Δ), minimal uncertainties (high accuracy for single layers), and a smaller spot size (due to the laser); however, because they only operate on single wavelengths, they also have less information to potentially decouple complex stacks/thicknesses/dispersions. Spectroscopic ellipsometers, on the other hand, have much more information (Ψ, Δ as a function of various wavelengths), possibly a larger spot size, and are slightly less precise but much more widely prevalent in the semiconductor industry to resolve thicknesses and composition of complex stacks.

New optical equipment are being built that explore several ways to obtain more optical information and resolve for additional parameters that are needed to control advanced process such as thickness, composition, dielectric constant, porosity, roughness, dopant concentration, etc. The common trend to extend an optical metrology tool's capability is to look for additional optical channels to obtain more information that could potentially help improve sensitivity (to detect and track various intrinsic properties of film stack) and reduce correlation among multiple stack parameters (e.g., thickness and refractive index of a layer, or correlation of thickness among two optically similar layers, such as SiO_2 and SiN). Such parameter correlation is undesirable because it makes it more difficult to decouple those correlated parameters, which produces more measurement uncertainty. Additional optical channels bring in more information that could help potentially reduce

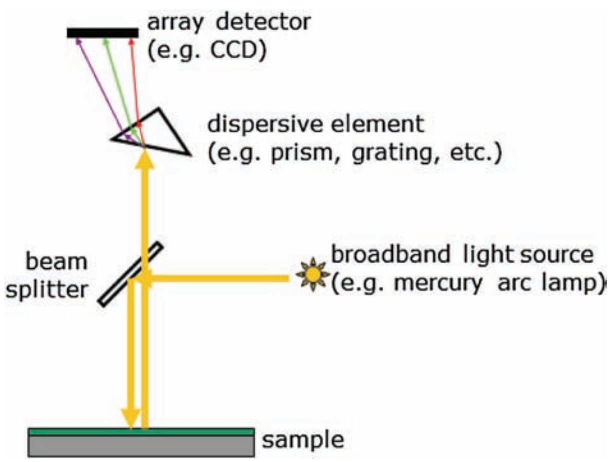


Figure 5.2 A spectroscopic reflectometer.

such correlations. However, such channels also make the tool more complex, slower (to collect data corresponding to all channels), and expensive. Example channels include:

- normal-incidence reflectometers,
- multi-angle-of-incidence ellipsometers,
- multi-single-wavelength ellipsometers used in unison with SE,
- infrared ellipsometers/reflectometers, and
- Mueller matrix SE, among others.

The specific characteristics of each of these channels change slightly depending on the configuration used, but the underlying principles stay the same: measure optical information as a function of predetermined setups on the sample of interest and compare the experimental results with the modeled results to deduce the sample's parameters of interest.

A simpler setup (as shown in Fig. 5.2) that is widely used in the industry—a spectroscopic reflectometer—acquires the reflectance of light as a function of wavelength. This method is usually performed in normal-incidence mode and is a relatively less expensive way to acquire optical data for thicker films (or it may serve as an additional channel with SE for complex layers, as described earlier). The information could be enhanced by using a polarizing reflectometer. Such optical setups are typically faster and less expensive, but they can only resolve simpler structures.

5.2 Scatterometry (OCD)

Scatterometry [also referred to as optical critical dimension (OCD)] technology relies on the diffraction pattern of an incident beam via a periodic grating on the wafer surface. The diffraction response is usually the plot for

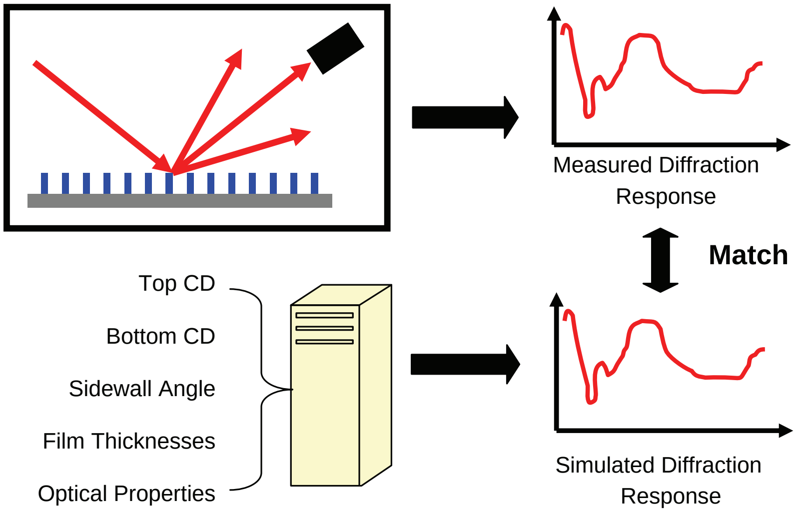


Figure 5.3 Block diagram of the scatterometry technique.

reflectance as a function of the wavelength. Measurements are made of the reflected light (hardware) and are used along with numerical simulation (software) to determine the profile of the grating structures (Fig. 5.3). Most scatterometry equipment/tool types share the same hardware and optics as optical thin-film metrology tools. Furthermore, the experimental data (commonly tan Ψ and Δ) collected in scatterometry are also similar to optical thickness metrology tools. However, scatterometers process these measured experimental data in a different way: numerical simulations are typically modeled using rigorous coupled-wave algorithms (RCWAs). These simulations are then compared with experimental results (spectra) via regression, and output parameters are then computed. These output parameters encompass 3D profile information, including the CD, sidewall angle, depth, thickness, and concentration of complex structures in a nondestructive high-throughput mode.

A beam of light incident on a periodic line-space structure (called "grating") is diffracted by the structure and measured. These beams of light will be referred to as "experimental/diffraction spectra." Again, the optical hardware setup used to acquire such a diffraction response is similar to those discussed in previous sections: SE, ellipsometers, reflectometers, and polarizing reflectometers. The experimental diffraction spectra are then compared to a "library" of simulated spectra. This library is prepared beforehand using RCWA diffraction simulations of the structure of interest. There is currently no known method for directly solving the grating profile parameters of a structure based on the measured/experimental diffraction response. Instead, forward simulations are run that take grating profile parameters as inputs and return the simulated diffraction response using a model. This model is built using stack information (such as n and k) for each layer, geometrical schematic, and parameter of interest. In order to produce the best model, it is often iterated on to attain the desired success metric performance per the application requirement (e.g., precision, TMU/accuracy, matching/FMP, process sensitivity); this process is referred to as model optimization and is the most time-consuming process in OCD flow.

After the model is ready, a set of forward simulations of diffraction responses are generated, i.e., a library of various permutations and combinations of various structure characteristic values. The experimental spectra are finally compared to simulated spectra, and the results of the parameters of interest (such as CD, SWA, thickness, etc.) are attained by ensuring the best match. Success metrics to quantify the level of matching between simulated and experimental spectra—such as goodness of fit (GOF) and chi-square (chisq)—are simultaneously generated to add further diagnostic information about the measurement. Figure 5.4 shows the offline scatterometry library setup and on-tool grating CD measurement flow, as described earlier.

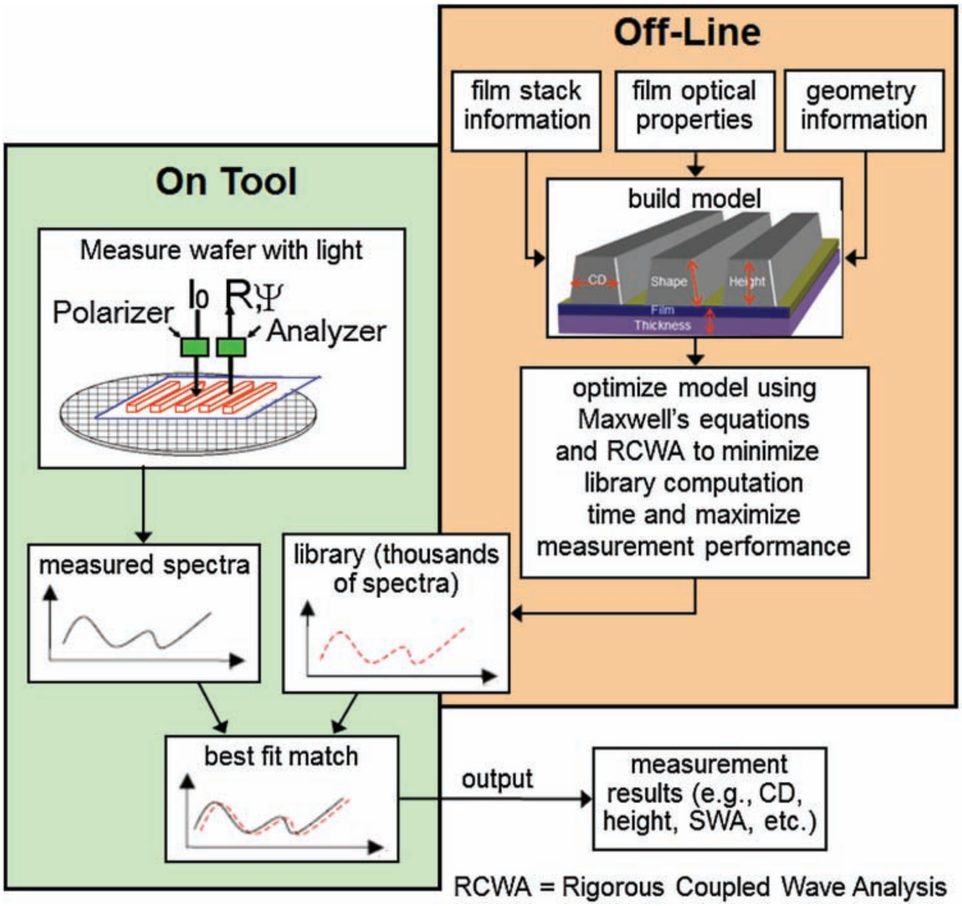


Figure 5.4 Flow of the scatterometry principle.

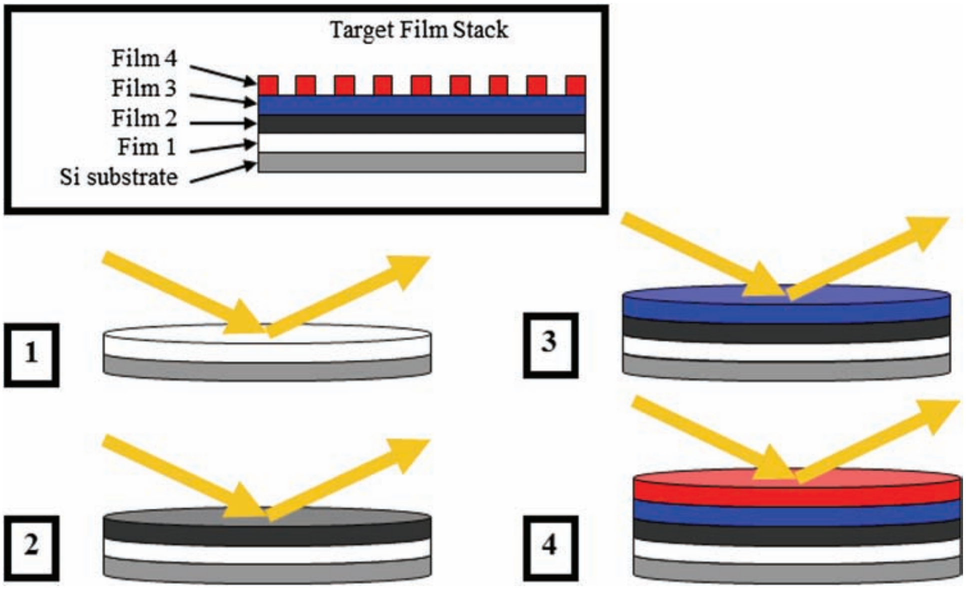


Figure 5.5 Additive stack method.

Scatterometry methodology requires a specific macrograting or an array of structures at least as wide as the spot size of the equipment (usually on order of 30–50 μm in *x* and *y*). These macrotargets should be generated and designed on the appropriate mask set beforehand. Targets are typically designed to closely mimic the device structure so that measurement results acquired by scatterometry are used to closely track the actual devices.

The first step of scatterometry determines optical constants—*n* and *k* for each of the layers within the structure—via various methods. An example of the simplest methods is the "additive stack" methodology shown in Fig. 5.5. Measurements are taken using advanced optical hardware throughout the additive stack process. As shown in Fig. 5.5, the stack contains four films on top of a silicon substrate; thus, four optical measurements are required. The optical properties (*n* and *k*) for each of these four layers are then extracted using all of the sets of data, typically using proprietary material-characterization software. Some of the more-advanced material characterization software allows the use of grating data along with additive spectra to be "regressed" simultaneously in order to potentially provide faster and more-accurate results.

The major benefit of scatterometry is that it provides a rich amount of information, i.e., CD, SWA, trench depth, film thickness, and other geometrical parameters, in a single measurement. Furthermore, it provides data at a high throughput and thus can be used to acquire inline measurement data in a very-high-sampling mode. In addition, it also enables scatterometry data to be used to actively control the processes in an APC model to ensure tighter

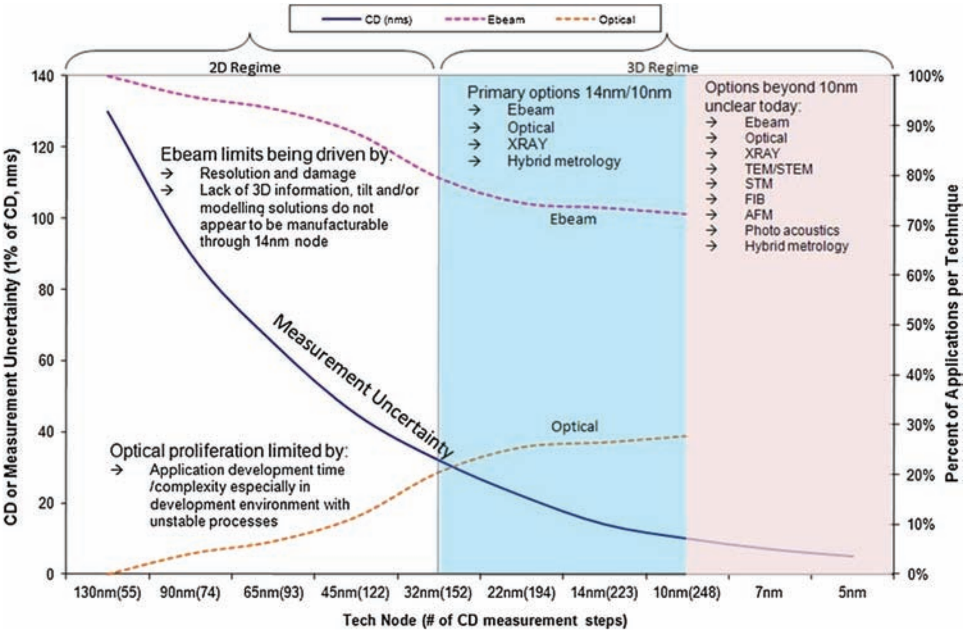


Figure 5.6 Trends of optical (OCD) steps as a function of nodes. Measurement uncertainty trends are also included for reference.

distribution of unstable processes. Based on these benefits, scatterometry is one of the fastest-growing metrology techniques in the industry in recent years, especially with the development of advanced nodes that require more measurements (more parameters of interest within a sample structure and more sampling across-wafer, wafer-to-wafer, and across lots). Figure 5.6 shows the trend of increasing optical (OCD) steps as sizes shrink toward advanced nodes. OCD tools are relatively small in size and thus can be integrated into process tools—such a setup is called integrated metrology (IM). Such IM tools could be integrated into lithography scanners, CMP polishers, and etch tools, to further reduce the mean time to detect, increase the APC use-case for wafer-level control, and reduce the overall cycle time.

The key challenge of scatterometry is a long model-optimization time and a requirement for reference data to help with model set up. Model optimization time is improving because higher levels of computational power and more-advanced automated and intelligent algorithms are now available. However, these gains are offset due to the increasing complexity of OCD models and the use of complex OCD hardware that use multiple optical channels, which are needed to satisfy the requirements of such complex models. Model complexity is increasing due to smaller structures that have more parameters of interest. As an example, a 90-nm OCD gate-etch structure would typically comprise five floating independent parameters, or five degrees of freedom (DOF). A typical 14-nm FinFET gate-etch structure would contain as many as 20–25 floating parameters with much smaller size and tighter pitch, thereby making such measurements a challenging process. Furthermore, reference metrology is much more challenging and time consuming because the only way to characterize such advanced small structures is TEM—destructive, time consuming, and requires several TEMs to produce a reference tool that resolves at an angstrom scale!

In order to resolve the challenges discussed here, similar to thin-film optical metrology equipment, additional channels are being added to scatterometry tools to keep pace with current and future requirements. Examples of such optical channels include multi-AOI, multi-azimuthal angle, DUV (deep ultraviolet) wavelength OCD, infrared OCD, Mueller matrix, hybrid metrology (see Chapter 8), etc. Furthermore, in order to improve scatterometry set up time and enhance the sensitivity to measure smaller dimensions, some scatterometry equipment manufacturers have recently started working on a modelless scatterometry solution. In this methodology, instead of generating conventional scatterometry models and comparing RCWA simulations with experimental spectra to compute the output parameters, the experimental data is directly correlated with output parameters (obtained through reference metrology during the set up process) to construct a mathematical model that could be used for scatterometry metrology.

5.3 Overlay

Overlay (also called registration error) is the measure of alignment errors in the *x* and *y* directions from one lithography layer to the next. Every lithographically patterned layer requires overlay measurement and control. Most overlay metrology is performed with an optical system that measures specially designed overlay measurement marks, which usually consist of two parts: one layer defines one part, and the next layer defines the other to complete the overlay marks (or overlay targets). Figure 5.7 shows three common types of marks; these designs are referred to as image-based designs

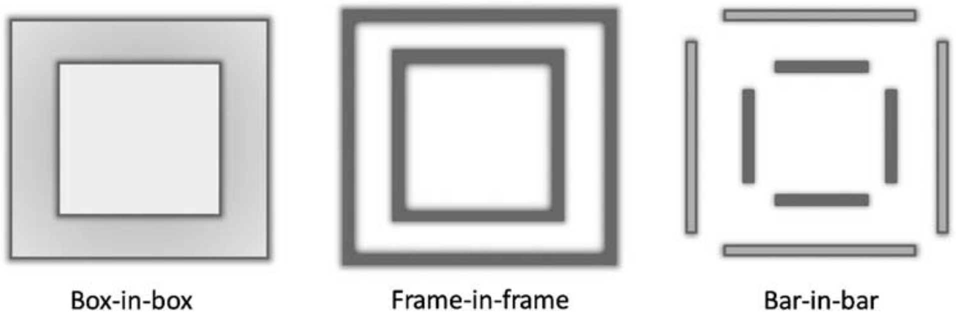


Figure 5.7 Typical overlay measurement marks.

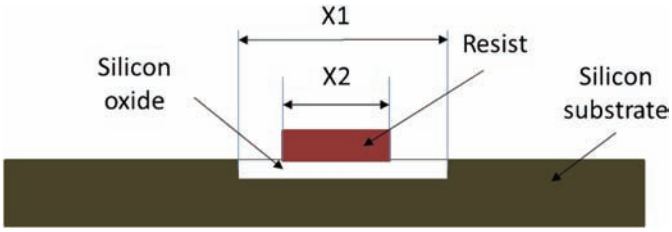


Figure 5.8 Cross-section of a box-in-box overlay mark.

because the x and y overlay measurements are derived directly from the image. Note that the overlay measurement marks have two different layers, and thus the two parts of the entire mark have different heights, as shown in Fig. 5.8. Note that the overlay marks consist of at least two edges in both the x and y direction, the center can be defined in each layer, and the distance between the two centers in the x and y direction are measured overlay errors in x and y—together they form an overlay vector at each overlay measurement location. Take the case in Fig. 5.8 as an example, wherein the overlay error measurement in x is $(x_1 - x_2)/2$.

Overlay errors are normally measured using confocal optical microscope systems, which can focus on two different mark layers and detect the edges of overlay marks, and then calculate and report overlay errors at each overlay mark site.

Overlay errors have the following sources:

1. True alignment error from a lithography machine (steppers or scanners). If two lithographic layers are printed using the same manufacturer and same model of scanners (most likely), with same alignment marks and alignment algorithm, the overlay error is normally limited by the machine performance, which includes, but is not limited to, the mask leveling relative to the wafer plane, dynamic focusing performance, alignment optics and algorithm, stage and wafer surface flatness, and lens aberrations.
2. The quality of the alignment mark used by the stepper for alignment. There are other process steps between two lithographic layers (e.g., film deposition, thermal process, etching and cleaning, etc.) and the alignment marks used by scanners could be changed slightly by those process steps (especially etching and cleaning process steps), and thus introduce additional overlay error from alignment marks' quality. In addition, with opaque layers (other than silicon oxide) and CMP processes, original alignment marks usually cannot be used without additional process steps (either open an window through a lithography and etch processes; or define a set of new alignment marks—cascade marks or cloned marks, using the original alignment marks—again through lithography and etch processes).
3. Overlay measurement system performance, like its precision and accuracy.

4. Distortions in the mark design measured by the overlay tool coming from previous processing steps (like CMP for example). These types of errors are difficult to correct for since they originate outside the lithography tool.
5. Relatively poor accuracy performance of overall recipes. Inaccuracies could be due to differences between the device minimum pitch and overall target pitches, target damage due to unoptimized target designs, or lower contrast on the overlay targets.

Overlay errors from the above mentioned sources have mixed random and systematic components. Those systematic errors from scanner systems can be modeled and corrected through systematic calibration, for example, scaling error, rotation error, trapezoid error, and translational error from focus error, mask leveling, alignment mark quality and lens aberrations. Figure 5.9 shows typical systematic overlay errors.

Because special overlay marks must be used in normal measurements, and those overlay marks are relative large (micron scale) compared to the real

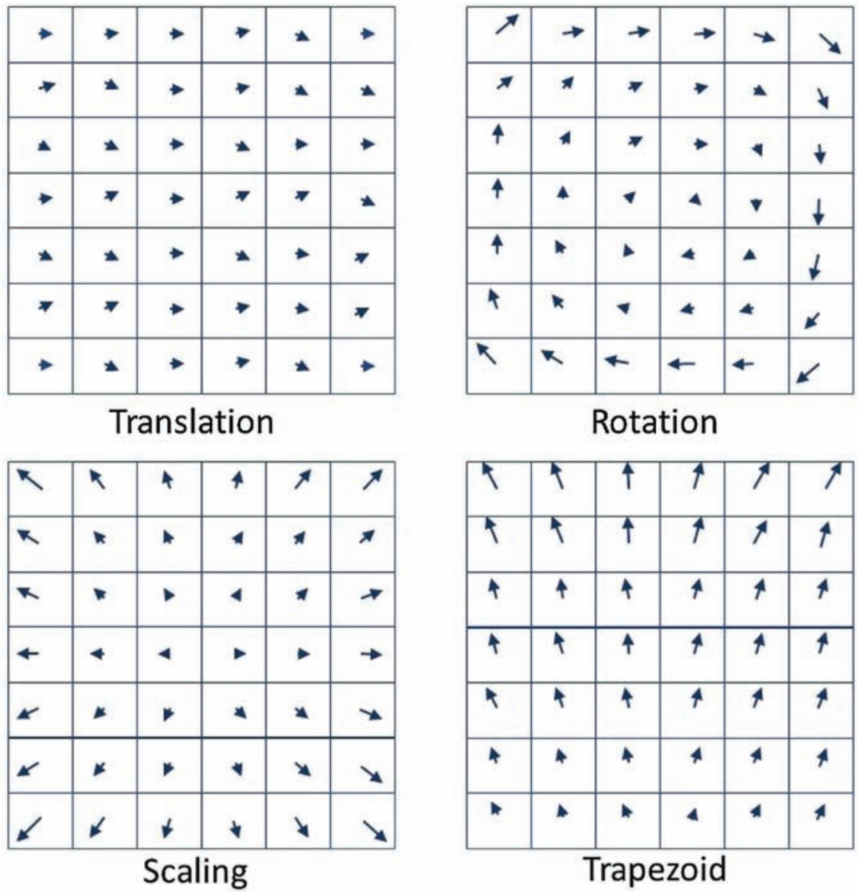


Figure 5.9 Typical systematic overlay errors.

circuitry pattern sizes (in the 10–100-nm scale), the overlay measurement results from those overlay marks may not correlate well with the much-smaller true overlay between the real circuitry patterns. Such suspicion has been confirmed in certain test cases by using CD-SEM as the overlay measurement tool on much-smaller overlay marks or by using actual circuitry-like real patterns. However, despite such correlation concerns, optical overlay measurements with relatively large overlay marks are still in use today, even with double-patterning or (future) multiple-patterning technologies. To meet the advanced node's tight overlay requirements, new target designs, called diffraction-based overlay (DBO) metrology, have been shown to have significantly reduced total measurement uncertainty (TMU) compared to image-based overlay (IBO), primarily due to the absence of measurable tool-induced shift (TIS). However, the advantages of this situation can be outweighed by increased susceptibility to wafer-induced shift (WIS) caused by target damage or process nonuniformities and variations. The path to optimal DBO performance requires well-characterized metrology targets, which are insensitive to process nonuniformities and variations, in combination with optimized recipes that take advantage of advanced DBO designs.

Because alignment is performed in lithography field by field, overlay must be checked inside each field (intrafield) and field-to-field (interfield). A typical sampling plan for overlay measurement in IC manufacturing involves five points per field (four corners and the field center), 5–12 fields per wafer to cover different areas of a wafer from center to edge, and 2–3 wafers per lot. To report the overlay measurement results, all intrafield overlay measurements are summarized as one field overlay vector and reported using (mean $+ 3\sigma$) as the maximal overlay error.

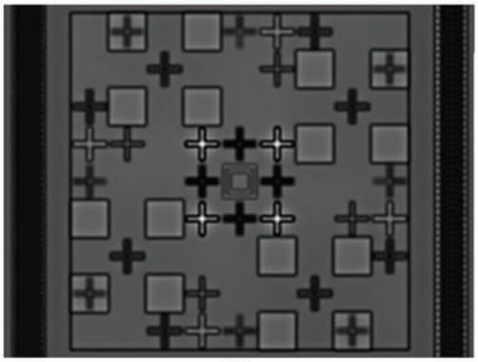
Overlay errors translate into CD changes in double-patterning technology; thus, the requirement for overlay error is very tight in IC manufacturing (at least half of the current overlay tolerance). To meet ever tighter overlay requirements in the upcoming multiple-patterning era, a new overlay monitor strategy may soon have to be developed and adopted.

Note that scanner machines have built-in overlay measurement capability, and that capability has developed along with IC manufacturing progress. In some sense, the overlay metrology inside scanners is better than standalone metrology measurement systems.

Figure 5.10 is a screenshot of different kinds of new multilayer targets: blossom and multi-AIM. These targets are a consequence of the industry's inability to make EUV lithography manufacturable. For a given 14-nm technology node's critical layer, e.g., the first metal layer, EUV could print this circuitry in one exposure step. Because EUV is not ready, the current state-of-the-art 193-nm immersion exposure tools need as many as three separate exposure steps to resolve the entire first-metal-layer circuitry, which complicates the measurement strategy at overlay (and CD). All three first

Overlay target type	Schematic	Feature width	Minimum Area required	Comments
Box-in-Box Bar-in-Bar (kerf)		1 μm to 2 μm	27 μm by 27 μm	Conventional (Image Based) scribeline target
AIM (kerf)		1 μm to 2 μm	31 μm by 31 μm	Conventional (Image Based) scribeline target
AIMid (in die)		0.5 μm to 2.0 μm	7 μm by 7 μm	In-die (Image Based) target
Multilayer (multi-layer, kerf & in-die		0.5 μm to 1.0 μm	5 μm by 20 μm	Multilayer overlay target
DBO		Pitch 400-600 nm Duty cycle 50%	10x10 um	Diffraction based

(a)



(b)

Figure 5.10 (a) Examples of prominent overlay targets. (b) Example of a blossom multilayer target.

metal layers must be printed in spec from an overlay and CD perspective with respect to each other and the underlying layers. These new targets enable the collection of this required information. The blossom target [Fig. 5.10(b)] is particularly innovative because the area is smaller than that used by conventional overlay targets—the same goes for other optimal multilayer overlay targets. The targets are made from gratings at or near the minimum ground rule in chip dimensions that the optical tool can still measure. They have been miniaturized and therefore consume less space; because they are a grating at or near the minimum ground rule, they generally correlate better to process shifts at the device level.

The advent of multipatterning has created numerous overlay and CD challenges that require credible solutions for pitch walking. Overlay budgets are shrinking to unprecedented levels. Multilayer overlay becomes challenging to manage with respect to design and data. Target design has numerous concerns for target robustness and correlation to device. Sources of overlay error from outside lithography have a higher impact on inaccuracy. In-field requires control, which means smaller target designs are needed that fit inside the field and minimize disruption. Furthermore, a CD-SEM is being used more often as a reference tool to provide post-etch overlay measurements on specialized target designs or directly on devices when the prior level is revealed. This reference information can then be used to select which target designs are best suited for use at the lithography step.

Expanding on advanced target designs, simulations must be used more often to evaluate the optimal mark design before printing. No longer can a given overlay design mark be assumed to work across all layers. Different mark designs might even need to be used during development versus manufacturing. Target design performance under different process variations and measurement conditions is critical to design correctly the first time. In many cases, this scenario requires high-density targets (consisting of many segments at or near the minimum ground rule) to withstand the process (layout, CMP assist, segmented space). Targets should undergo the same processing as the device (e.g., FinFet targets). One exception, process-sensitive targets, can sometimes be used to track nonlithographic processes. The target designs must better represent the device, which may involve measuring combinations of scribe lines and in-die, and include device structures. They must be made smaller (10×10 μm) to fit areas of restricted size with the die. Figure 5.11(a) summarizes three categories for overlay improvements in measurement algorithms and diagnostic metrics, target design, and tool capabilities. Figure 5.11(b) shows an example of modifications to a target design to make it more robust for processing and measurement. The provided

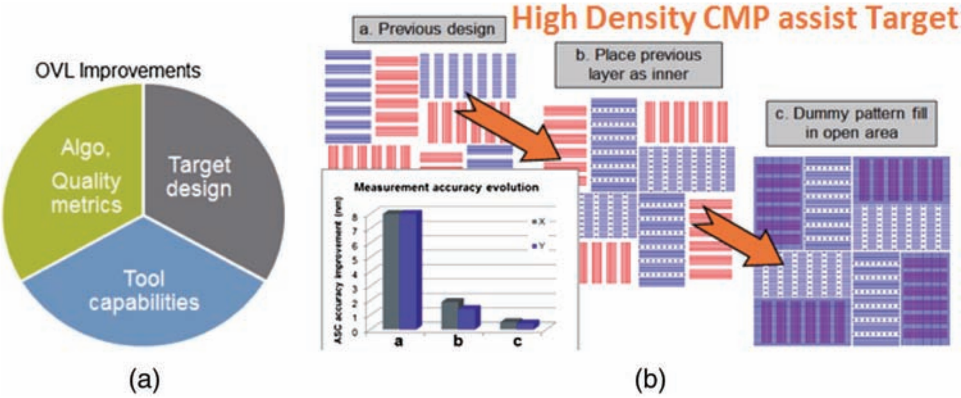


Figure 5.11 (a) Key overlay areas that require attention in the future. (b) An example of how a design might be modified to be more robust and accurate relative to the device.

Table 5.1 The ability (or lack thereof) of a given target-design approach to meet challenges.

1× nm Key Challenge	Imaging	Scatterometry/Diffraction
Process variation (20–30% within wafter in HVM)	Endures up to >30% variation. More desiable in R&D.	>15%, error-introduced. Stable process, better TMU.
In-field control	~8 × 8 μm in-die targets	10 × 10 μm in-die targets
Time to working target	First round (using simulation)	2–4
Multilayer	Proven	Gap (work in progress)
Nonlithographic process tracking (CMP/etch)	Process-sensitive targets (track OVL inaccuracies)	Uncertainty analysis
Error estimation/flags	Accuracy flags from raw signal. Target asymmetry, correction.	Accuracy flags from raw signals

data illustrates how much better the performance of the right-most design is compared to the others (these are guidelines for future consideration).

Note the interplay between image- and diffraction-based designs. Table 5.1 shows a selection of various 1× key challenges and contrasts the ability of image-based versus scatterometry/diffraction-based target designs to handle these challenges. Regarding process variation, image-based designs are better suited to handle the variations common in research and development than the scatterometry/diffraction-based designs; however, the latter are better suited in manufacturing because they represent the device better. Both designs can be miniaturized for placement in-die, with a slight advantage to image-based designs. Optimization of target design requires much less work for image-based versus scatterometry/diffraction-based designs. As many as four iterations may be needed for the latter, something that deserves future improvement. Image-based target designs are well suited for multilayer situations, whereas scatterometry/diffraction designs are still being refined. Lastly, regardless of design type, it is critical to utilize error flags while optimizing target designs and recipe setups.

Moving forward, overlay requires continued innovation to better track what happens in chips. Improvement in future optical toolsets will help, along with smaller innovative target designs intelligently placed in and around the chip; however, other solutions, such as CD-SEM etch overlay reference measurements, can help decide how to make the best target design selections at lithography. Overlay will require serious attention in the future and could be a limitation for future technology nodes.

References

1. S. Narasimha et al., "High-Performance 45-nm SOI Technology with Enhanced Strain, Porous Low-k BEOL, and Immersion Lithography," *Int. Electron Devices Meeting*, San Francisco, CA (2006).
2. A. Vaid et al., "Scatterometry as technology enabler for embedded SiGe process," *Proc. SPIE* **6922**, 69220U (2008) [doi: 10.1117/12.774564].

3. R. M. A. Azzam and N. M. Bashara, *Ellipsometry and Polarized Light*, 5th ed., Elsevier Science, Amsterdam (2003).
4. H. Fujiwara, *Spectroscopic Ellipsometry: Principles and Applications*, John Wiley and Sons, Hoboken, NJ (2007).
5. M. G. Moharam, E. B. Grann, D. A. Pommet, and T. K. Gaylord, "Formulation for stable and efficient implementation of the rigorous coupled-wave analysis of binary gratings," *J. Opt. Soc. Am. A* **12**(5), 1068–1076 (May 1995).
6. K. Lensing, "The Pattern Metrology Puzzle: Piecing Together a Control Strategy for 65-nm and Beyond," *IMA Meeting at AEC/APC Symposium XVII*, Indian Wells, CA (2005).
7. A. Vaid et al., "Product-driven material characterization for improved scatterometry time-to-solution," *Proc. SPIE* **7272**, 72720V (2009) [doi: 10.1117/12.814380].
8. P. Leray et al., "Achieving optimum diffraction based overlay performance," *Proc. SPIE* **7638**, 76382B (2010) [doi: 10.1117/12.848524].
9. T. Shapoval et al., "Influence of the process-induced asymmetry on the accuracy of overlay measurements," *Proc. SPIE* **9424**, 94240B (2015) [doi: 10.1117/12.2085788].
10. P. Leray et al., "Overlay metrology solutions in a triple patterning scheme," *Proc. SPIE* **9424**, 94240E (2015) [doi:10.1117/12.2087304].
11. S. Pandev et al., "Signal response metrology (SRM): a new approach for lithography metrology," *Proc. SPIE* **9424**, 94241P (2015) [doi: 10.1117/12.2086056].
12. C. Bozdog et al., "A holistic metrology approach: multi-channel scatterometry for complex applications," *Proc. SPIE* **7971**, 797113 (2011) [doi: 10.1117/12.881638].
13. M. Sendelbach et al., "Use of multiple azimuthal angles to enable advanced scatterometry applications," *Proc. SPIE* **7638**, 76381G (2010) [doi:10.1117/12.846692].

Chapter 6
Metrology Toolsets in IC Manufacturing: Charged-Particle Metrology Systems

The primary charged-particle-based metrology system used today is the CD-SEM. It accelerates electrons across the measurement structure and looks at the electron yield for imaging. For the last 15 or so years, ion-based imaging systems have been explored, but no production-worthy tools are available as of 2015, and no production-grade ion beam tools will be commercially available in the foreseeable future.

6.1 Electron-based Systems

The CD-SEM is the primary CD metrology tool for IC manufacturing. It has the resolution, precision, and accuracy to measure small-scale dimensions in nanometers. In addition, after 20+ years of development and improvements, it has the automation and throughput to be the sole production-proven CD metrology tool. CD metrologies are used in two critical image-transfer process steps: lithography and etch. CDs in different layers refer to different feature dimensions, such as linewidths, spacewidths, contact, via hole diameters, etc. Different measurement algorithms have been developed to measure different CDs.

Similar to all other SEM imaging machines, a CD-SEM column consists of the following basic elements:

- an electron emission (source),
- magnetic lenses to control beam scan and focus (beam control), and
- an electron detector (detection).

One type of CD-SEM column is shown in Fig. 6.1.

The primary electron beam, controlled by electromagnetic lenses, scans the sample surface of the predefined measurement area. Secondary electrons (SEs) and backscattered electrons (BSEs) are generated by the primary

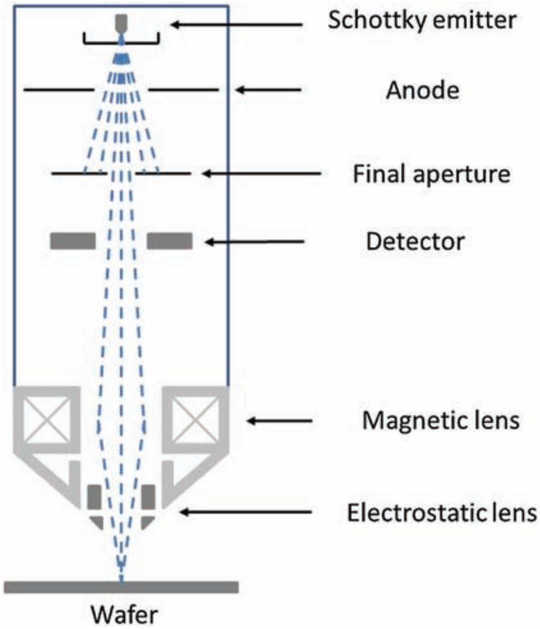


Figure 6.1 A CD-SEM column with an electron source, beam-control element, and in-column electron detector. Image used with permission of Applied Materials.

electrons from the sample surface, as shown in Fig. 6.2, from "the interaction sphere." The SEs have relatively low energy (typically less than 50 eV), and the BSEs have relatively high energy (larger than 50 eV). Because SEs have low energy, they escape from very shallow layers under the sample surface (only a few nanometers deep); thus, the low-energy SEs, which carry surface topology information, are used as measurement signals for CD measurements, whereas the higher-energy BSEs (which carry material-contrast information) are used for other purposes.

For CD measurement purposes, only the SE signal need be examined. The SE signal strength is directly related to SE emission from a sample surface. As shown in a typical electron-signal energy spectrum in Fig. 6.3, the SE peak is typically at less than 2 eV for materials in IC processing. A CD-SEM uses relatively low landing energy ($E_o < 1$ keV) and low beam current (in the 10–100 pA range) to avoid charging.

The SE emission is enhanced when a primary beam passes through a step (or the edge of a line or a space) and a SE signal peak is formed, as shown in Fig. 6.4. The SE emission enhancement can be viewed as the direct result of the interaction area enhancement on the sidewall of a step. As the primary beam scans through a line or a space, two edges from the surface topography change, producing two SE signal peaks, as in Figs. 6.4(a) and (b). An edge-detection algorithm can be applied to detect those edges [a rising edge (step up) or a falling edge (step down)]. A measurement algorithm can be applied to

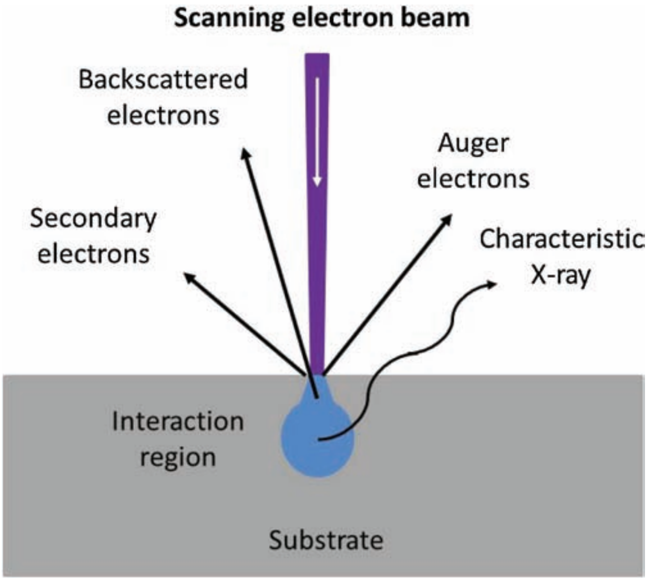


Figure 6.2 Primary electron beam interacting with a substrate through an interaction sphere and the signals emitted from the surface.

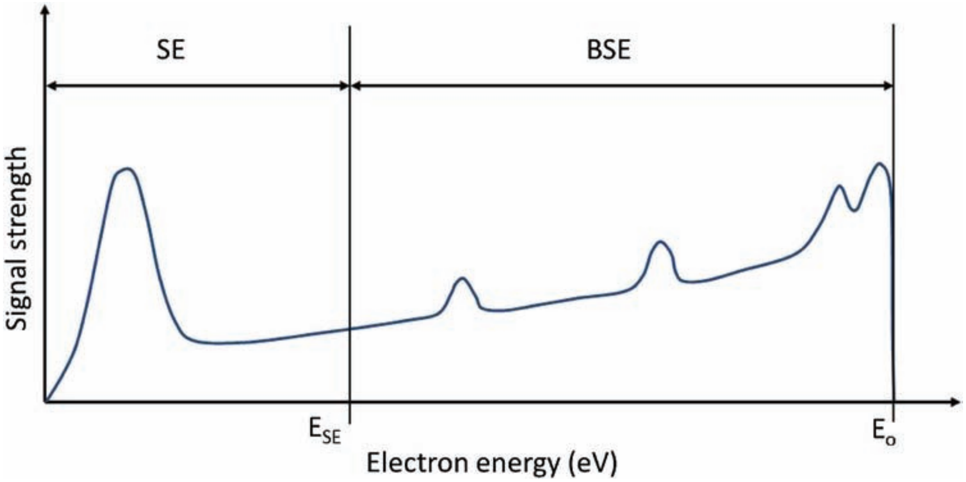


Figure 6.3 Electron signal spectrum of a typical SEM machine.

measure the distance between two edges, and a CD (a linewidth or spacewidth) value is obtained.

To increase the SNR, multiple scans are usually performed within a predefined length along a line or space, assuming a constant linewidth or spacewidth (which is true in all line/space patterns in IC manufacturing processes), as shown in Figs. 6.4(c) and (d). The average scan waveform comes

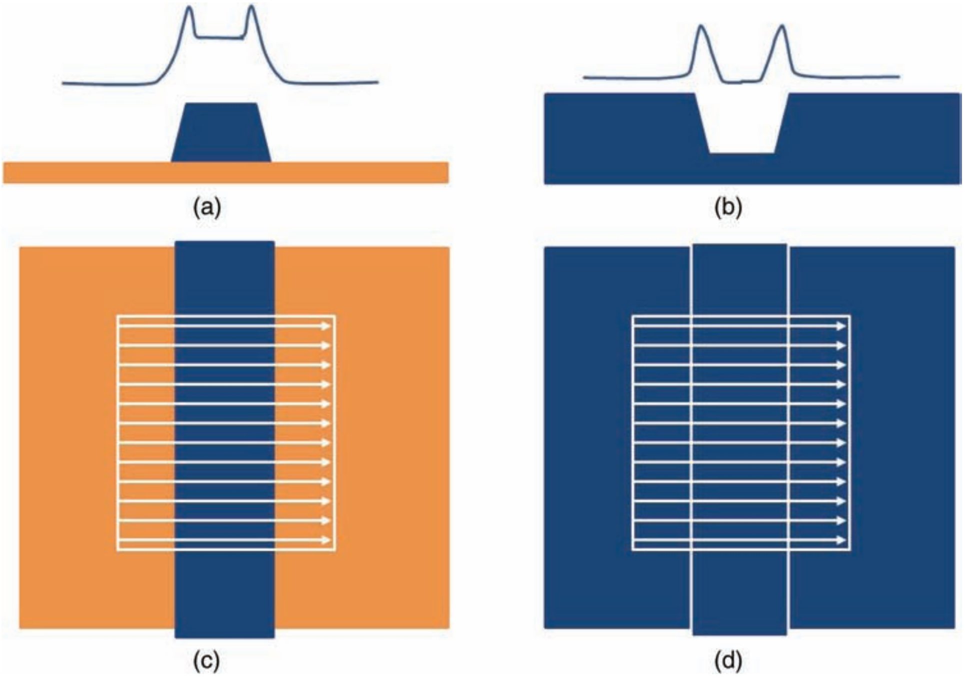


Figure 6.4 SE waveform formation through SE emission enhancement at edge regions for (a) a line and (b) a space. The multiple scan lines within the measurement gauge boxes for the (c) line and (d) space to increase the SNR.

from all of the scan lines, and the random noise is reduced by $\sqrt{m}$, where m is the total number of scans.

In addition to a 1D line/space pattern, more-complicated test patterns must be measured and more-complicated measurement algorithms developed for contact/via holes (measure a diameter or an area of a circle): line end to line end, line end to a line, space end to space end, etc.

6.2 CD-SEM and Sample Interactions

The CD-SEM is known to interact with the structure being measured and create an effect sometimes known as "carryover." More specifically, the structure being measured changes as a function of repeated measurements. Section 3.9 presented an analysis technique to separate out the carryover from the measurement error regardless of the trend of the carryover effect, which generally has two modes of operation. The first mode is nonlinear, as shown in Fig. 6.5. The CD changes in a nonlinear fashion due to repeated measurements. This effect is often described as "slimming" and is typically linked to 193-nm resists and EUV resist systems. If a line is being measured, it becomes smaller with each subsequent measurement (hence the term slimming) until it reaches a plateau, where the behavior changes yet again

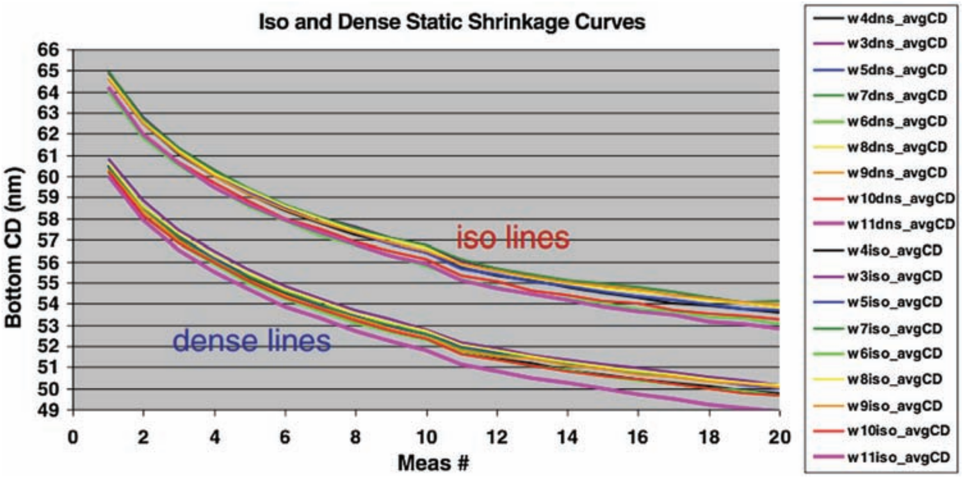


Figure 6.5 Resist shrinkage curves as a function of repeated measurement for various isolated and nested line resist structures.

(and vice versa for a space measurement), as illustrated in Fig. 6.5, where the behavior of nested and isolated lines is characterized as a function of repeated measurements. Note the sharp drop off in the first few measurements followed by a leveling effect. The second mode is linear; it is generally believed to originate from hydrocarbons present in the vacuum chamber that are attracted to the measurement location during the measurement process. It is essentially a deposition process and typically occurs at the post-etch measurement step and at 248 nm and higher photoresists.

The magnitude of this change is not much of a concern because it can be corrected for, but that is not true for very small, nanometer-scale measurements. Once the feature size is a few multiples of the slimming effect, there is concern that the structure will be distorted by the measurement process such that the measurement is not representative of the original feature. It may tilt or shrink in nonuniform ways, making it difficult to determine what the process is doing. Another concern is that the carryover effect is not constant across a fleet of CD-SEMs. If the carryover effect is 2× different between two tools, then as the process tolerance shrinks, this error factor becomes larger and negatively affects APC feedback loops and lot disposition. This error growth can be caused by many factors, such as different vacuum levels between tools or different amounts of hydrocarbons present in each tool. This is a significant concern that requires continued efforts to find manufactureable solutions.

6.3 Key Fundamental Challenges of the CD-SEM

As mentioned in the previous section, image-based OCD tools were replaced by the CD-SEM because they were not able to deliver the resolution needed at the time. Ironically, the CD-SEM currently faces a similar challenge and is

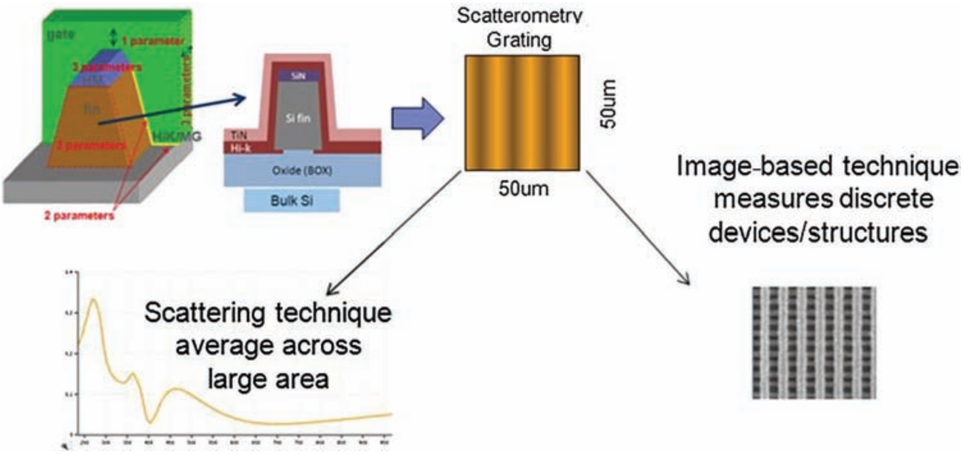


Figure 6.6 Two approaches for dimensional metrology: scattering (model)-based and image-based.

being replaced by a new breed of optical tool, i.e., scattering-based OCD tools. This very significant paradigm shift is driven by an industry need to capture more measurement parameters, such as undercut and feature height. These are parameters that the CD-SEM cannot provide. Figure 6.6 shows an example of the 3D measurement requirements needed at the fin and the gate modules for the FinFET device. Note that at least ten dimensional parameters are listed between these two modules:

- CD,
- undercut,
- line-edge roughness (LER),
- linewidth roughness (LWR),
- sidewall angle (SWA),
- height,
- pitch/walk,
- high-*k* thickness,
- TiN thickness, and
- height above the fin.

The CD-SEM has traditionally only measured CD, LER/LWR, and pitch. Most, if not all, of these parameters can be measured with the scattering-based OCD tools, but numerous challenges are involved. The scattering-based approaches require building a representative model for each given application, which is very time consuming and can only measure a global average for any of these parameters across large, predefined sacrificial targets. The modeling aspect significantly affects the ability to provide rapid process-development feedback when the process is in flux because the time it takes to build a very accurate model can be longer than the lifetime of a process experiment. Furthermore, it cannot provide measurements on discrete devices/structures that are required for optical proximity model building and many process debugging activities. For these reasons, and the fact that an image-based solution is ubiquitous, the industry requires both scattering- and image-based solutions to be able to measure all of the critical structural parameters to allow rapid process learning during process development and to maintain process control during high-volume manufacturing (HVM).

In the past, in order for the CD-SEM to be able to measure some of these additional parameters, some suppliers have pursued various forms of tilt (see Fig. 6.6). Unfortunately, this trend never caught on, possibly because there was insufficient detail in the image for the most-demanding applications to reliably extract additional parameters such as SWA. Suppliers have improved the CD-SEM resolution over time, but the question remains whether these improvements will meet requirements, especially if tilt is critical to measure these additional parameters. There is no standard that the authors are aware of for the resolution requirements needed over time, so Fig. 6.7 was created in this spirit. This chart examines the average resolution that the available supplier toolsets achieved from the 250-nm technology node to the 22-nm technology node. It is shown that the suppliers clearly improved their resolution over time, indicated by the square symbols, but the question is whether those improvements are sufficient to meet requirements. The measurement tolerance, defined as 1.5% of the minimum design rule (DR),

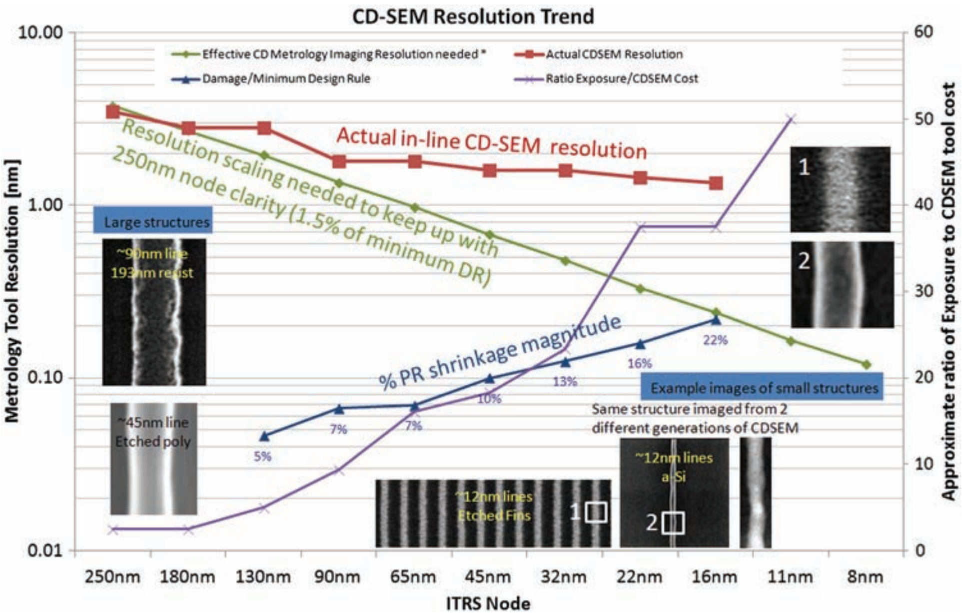


Figure 6.7 The quest for resolution with the imaging-based toolset and cost ratio between estimated tool prices of the lithography exposure to the CD-SEM tool.

for each technology node from the 250-nm node to the 22-nm node has also been plotted. When viewing this information on a log scale, a very disturbing trend emerges—note how the average resolution achieved does not keep pace relative to the resolution needed when anchored to the resolution–DR relationship in the 250-nm technology node. This means that in each successive node the critical features become less clear and more difficult to measure with certainty. At the current technology node (22 nm), the resolution consumes ∼6.5%, which is more than four times worse than what was achieved at the 250-nm node.

The images shown in the chart illustrate examples of lines from around 90 nm to 12 nm. On the left side of the chart there are two images of larger-line structures: a 90-nm 193-nm resist line and a 45-nm etched poly-silicon line. The right side of the chart shows some very small structures: a 12-nm etched fin and a 12-nm amorphous silicon (a-Si) line. Two images are shown for the latter, one from an older-generation CD-SEM and one from a newer generation. Note that the right image does not clearly delineate the two edges of the a-Si line, unlike the left image, which was made on the newer-generation CD-SEM and has the better resolution. This difference clearly illustrates how resolution improves the ability to reveal the needed details.

The 12-nm etched fin line image appears to be quite blurry. There is an obvious reduction in the detail that is revealed from the etched fin image relative to the a-Si line image. The reason for this is unclear because they are approximately the same size; perhaps there are material dependencies. This lack of consistent detail in the image makes it difficult for today's CD-SEMs to meet the needs of the process community. A resolution improved beyond that which is currently available for the smallest features is necessary to reveal the key process details required to rapidly develop and manufacture chips. Furthermore, resolution is a prerequisite before the CD-SEM can measure the other 3D parameters mentioned previously through the use of other techniques, such as tilt. Although many other issues exist for the CD-SEM, such as charging and the interaction volume, these are secondary issues relative to the resolution challenge.

It must be reiterated that the choice of 1.5% of a given ITRS node's minimum DR as the resolution goal is based on the historical performance of the 250-nm node, where the SEM performance was thought to be ideal. Whether 1.5% is the best choice as a resolution goal, or something larger (3% or 5%), is debatable, but that does not change the fact that the trends are divergent.

Furthermore, it should be noted that using the term "resolution" as the ultimate metric for comparison could be problematic, as there does not appear to be a single standard definition of the term accepted across the industry. Nor is there a standard method for its evaluation; various image analysis methods exist with different calculation algorithms, and results from these can be very sample-dependent. The use of a single resolution term is too simple, as the one number cannot begin to accurately describe all of the contributions to good imaging, including the size of the probe, interaction volume, material charging, and other particle/solid interactions. Although the term is helpful for comparison, it should be kept in mind that these values are somewhat relative and subjective, given the lack of standards. Estimates for ultimate resolution limits are in the 0.75–1.0-nm range for a conventional, low-voltage CD-SEM (LV-SEM). The current (2015) known state-of-the-art for a commercially available CD-SEM's resolution is 1.34 nm using the SMART algorithm, suggesting that there is potential for significant improvement to the existing technology. Further resolution improvements are expected with higher-voltage SEMs and ion microscopy. A future project of value to the industry would be to develop a consensus methodology and standard for resolution evaluation; work is underway with contrast transfer function (CTF) measurement, which could provide a standardized alternative metric. Significant progress has been made toward development of the methodology and sample design, but it is not yet complete.

It is interesting to plot another area where resolution is also paramount, specifically the lithography sector. The data shown by the X symbols is charted against the secondary y axis, which is the approximate ratio of lithography (exposure) tool cost to CD-SEM tool cost. Another very disturbing trend emerges, from the 250-nm node to the 11-nm node, we observe the ratio increases by a factor of about 25. Simply stated, the investment to improve resolution has not kept pace with the investment in the lithography tools that create the features; although parity is not logical, this seems to be a key indicator to help explain the resolution gap.

Lastly, it is useful to see how much the minimum resist feature measured per technology node is consumed by 193-nm resist slimming as a result of exposure to the e-beam. This was then plotted for each technology node by assuming a single measurement-exposure slimming magnitude (an assumption for lines/spaces was used) and dividing that by the technology node value. This is shown by the triangle symbols. We estimate 193-nm resists started being used around the 130-nm node. From the 130-nm to the 16-nm technology node, the percentage increased from $\sim$5% to $\sim$25%, which means that a minimum feature resist line in the 16-nm technology node changes its size $\sim$25% in the course of one to a few measurements, depending on various parameters. This is not a large concern if the slimming rate is stable on a given tool and across tools, but as this magnitude increases and differences arise among tools regarding the slimming rate, it will be much more difficult to control the process using e-beam CD data. This is only one mechanism of concern regarding e-beam exposure damage to devices/structures measured.

The Appendix categorizes all of the known e-beam damaging mechanisms, ranked from least damaging to most, and references. This is another key fundamental issue that needs to be addressed moving forward. Possible replacements for the current, conventional LV-SEM are the HV-SEM, environmental SEM, helium ion beam, proton microscope, and others, whereas transmission electron microscopy and scanning transmission electron microscopy are typical destructive reference metrologies; many of these different techniques have been subject to previous studies showing that their use might be, to varying degrees, more destructive than conventional LV-SEM. Some of these damage phenomena are local, meaning that only the region close to or within the area directly imaged exhibits such damage and thus influences the measurement (and maybe other device parameters) away from what it should nominally be, which may necessitate sacrificing planned features. Other damage phenomena might prevent further processing of the wafer that is truly destructive.

6.4 SEM-based Overlay Measurements

Interest in SEM-based overlay is increasing as device sizes continue to shrink in the 1× node. There is a growing concern in the disparity that may arise given that the device is much smaller than the size of the features measured in the overlay target designs. Due to the high-resolution imaging capability of a CD-SEM, SEM overlay targets can be designed to closely match the actual device dimensions and therefore can be potentially placed close to the device, given their smaller footprint. The size of the SEM-based target is also small enough to have fewer process-induced errors from CMP and etch, as well as reduce the chip size. Furthermore, if overlay can be performed post-litho on the CD-SEM tool, the lot cycle time can be significantly improved by removing the relevant overlay step(s) from the route and combining it with the existing CD-SEM measurement steps. The CD-SEM can also be used to provide reference metrology to evaluate the accuracy of existing optical-based overlay measurements, including image-based overlay (IBO) and diffraction-based overlay (DBO). An evaluation of the feasibility of SEM-based overlay post-litho is performed, and it demonstrates the capability of the CD-SEM for use with post-etch reference metrology.

A BEOL double-patterning layer is used for the evaluation. A high-energy SEM beam is used for post-litho measurement because a high-energy beam is needed to penetrate the photoresist and resolve the underlying pattern; a regular electron beam is used for post-etch measurement. At both the post-litho and post-etch step, the IBO (AIM test), DBO (SCOL and uDBO), and CD-SEM targets are measured across the full wafer, with 13 locations per die.

Figure 6.8 shows the results after lithography. SEM-based overlay results are highly correlated to other optical-based overlay measurements with an overall R^2 larger than 0.94 in both the x and y direction. Note that there is a difference in process sensitivity between the SEM-based overlay and other

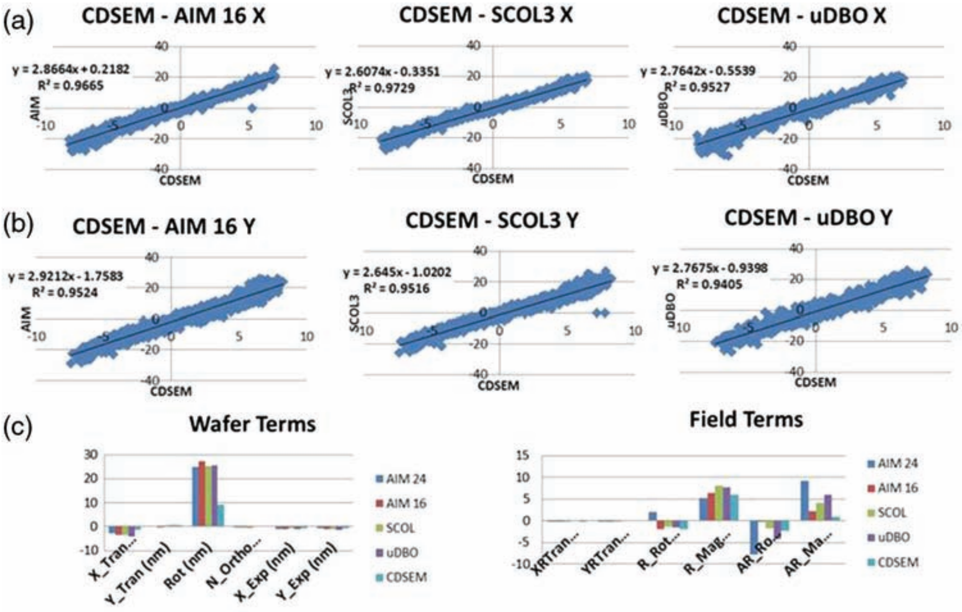


Figure 6.8 Post-litho result comparison: (a) correlation of overlay measurement on CD-SEM with AIM, SCOL, and uDBO in the *x* direction; (b) correlation of overlay measurement on CD-SEM with AIM, SCOL, and uDBO in the *y* direction; and (c) correctible terms after modeling.

optical overlay measurements. The slope is always between 2.6 to 2.9, with the SEM-based overlay measurements undergoing much less of an overlay change than the optical measurements on the optical targets. The reason for this is still under investigation; it is also reflected in the wafer terms after overlay modeling in a KT analyzer, as shown in Fig. 6.8(c).

This wafer is then etched and measured again in the same manner as post-litho. Figures 6.9(a)–(c) show the correlation of post-etch measurements from the SEM-based overlay to AIM, uDBO, and SCOL, and a significant difference in R^2 values has been observed. Note that only overlay_*x* data is shown because overlay_*y* data shows similar behavior. The IBO target (AIM) shows the best correlation with CD-SEM, with R^2 being 0.94 and slope close to unity, whereas the DBO targets show lower correlation, with the uDBO target showing the lowest correlation to SEM-based overlay measurements. This is not surprising because the DBO targets are not optimized for post-etch measurement. Pad edge damage is observed on the uDBO targets, which likely explains why this correlation is the lowest. Without the CD-SEM post-etch results, it would be difficult to determine which is best by comparing each optical measurement technique against one another, as shown in Fig. 6.9(d). The SEM-based overlay measurements can therefore help determine which is best by providing reference metrology.

Post-etch SEM-based overlay measurements are compared with the post-litho overlay measurements. In all cases the correlation is very high with a

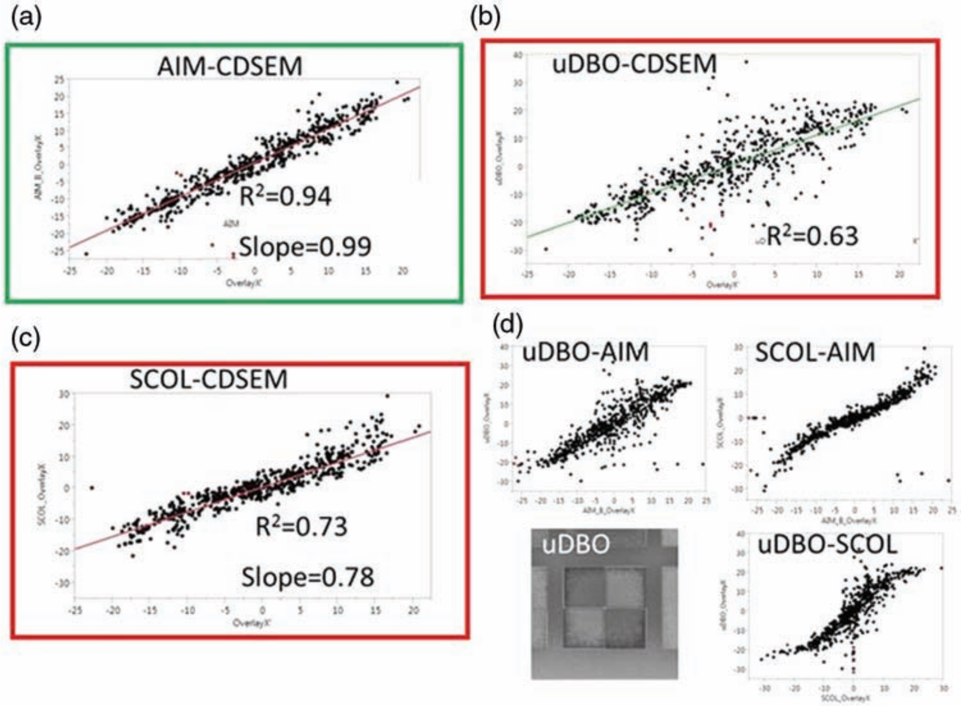


Figure 6.9 Overlay_*x* measurement results after etch, with correlation of (a) CD-SEM to AIM, (b) CD-SEM to uDBO, (c) CD-SEM to SCOL, and (d) intercorrelation of the optical overlay measurements and an image show pad damage of the uDBO target.

near-unity slope except when comparing the post-litho to post-etch SEM-based overlay measurements, as shown in Fig. 6.10. In this case, as expected, the slope is far from unity, given the process sensitivity observed at post-litho. When using the post-etch SEM-based overlay measurements as a reference to compare the different post-litho optical measurements, none stand out as being the best; they are essentially equivalent.

The precision of the SEM-based overlay has been evaluated, achieving 3σ of 0.93 nm at post-litho and 0.10 nm post-etch. The current overlay requirement is ∼0.3 nm at post-litho and post-etch, which means that the post-litho SEM-overlay measurement still requires improvement. Future work will focus on how to make the prior layer edges even clearer at post-litho because this factor likely contributes to the precision issue. Nevertheless, CD-SEM-based overlay may play a large part in the future regarding overlay measurements and should be investigated further.

In summary, the primary challenges facing CD-SEM involve

- increasing image fidelity;
- minimizing sample damage;
- providing device-relevant overlay measurements;

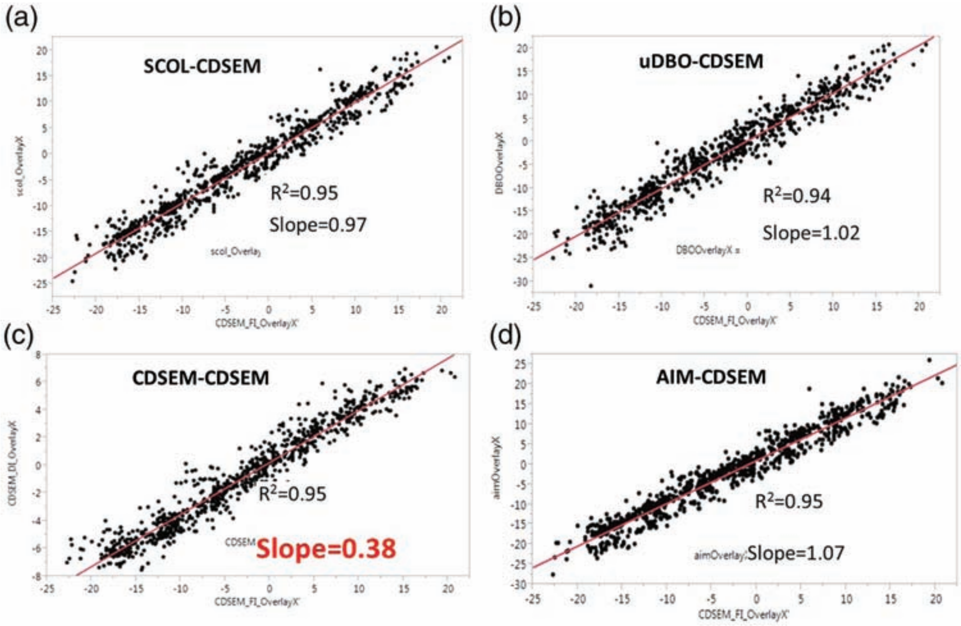


Figure 6.10 Correlation of overlay_*x* measurement post-litho on (a) SCOL, (b) uDBO, (c) CD-SEM, and (d) AIM in correlation to CD-SEM post-etch measurement as a reference.

- producing 3D measurements, e.g., height and sidewall angle; and
- using CD-SEM-based overlay to help validate the accuracy of the optical overlay measurements post-etch and possibly replace optical measurements at the lithography step.

Expanding on the latter point, tilting the beam can open the door to CD-SEM-based 3D measurements, but for the most-challenging layers (from an imaging-quality perspective) increased image fidelity and tilt are required to generate reliable and accurate measurements. CD-SEM-based overlay could help improve litho overlay control—this is the next frontier for the CD-SEM, and innovation is critically needed.

References

1. B. Su, R. Oshana, M. Menaker, Y. Barak, and X. Shi, "Shape control using sidewall imaging," *Proc. SPIE* **3998**, 232 (2000) [doi: 10.1117/12.386476].
2. B. Su, T. Pan, P. Li, J. Chinn, X. Shi, and M. Dusa, "Sidewall angle measurements using CD SEM," *Proc. ASMC*, 259 (1998).
3. E. Solecky and O. Adan, "Metrology toolset monitoring, matching, maintenance and management," SPIE short course material (2014).
4. B. Bunday et al., "Time-dependant electron-beam induced photoresist shrinkage effects," *J. Micro/Nanolithography, MEMS, and MOEMS* **11**(2), 023007-1 (2012) [doi: 10.1117/1.JMM.11.2.023007].

5. B. Su, G. Eytan, M. Padmanaban, and A. Romano, "Analyzing and characterizing 193-nm resist shrinkage," *Solid State Technology* (May 2001).
6. B. Su, G. Eytan, and A. Romano, "193-nm photoresist shrinkage after electron-beam exposure," *Proc. SPIE* **4344**, 695 (2001) [doi: 10.1117/12.436796].
7. X. Zhang et al., "Solving next-generation (1X node) metrology challenges using advanced CD-SEM capabilities: tilt, high energy and backscatter imaging," *Proc. SPIE* **9424**, 94240G (2015) [doi: 10.1117/12.2087267].

Chapter 8
Limitations of Metrology Techniques and Hybrid Metrology

8.1 Introduction

Dimensional metrology tools are increasingly challenged by the continuing decrease in the device dimensions, combined with complex disruptive materials and architectures. These demands are not being met by existing/forthcoming metrology techniques individually. Hybrid metrology (HM), which combines measurements from multiple toolset types to enable or improve the measurement of one or more critical parameters, has recently been incorporated by the industry to resolve these challenges. Before discussing HM, this chapter reviews the salient features of various inline metrology techniques based on the industry's current challenges and requirements use cases.

The complexity of device structures requires that the dimensions of intricate features be accurately and precisely determined for process control. Such features include the sidewall angle (SWA), undercut, thin sidewall films, pull-down, spacer width, composition, etc. As explained earlier, toolsets such as scatterometry/OCD, CD-SEM, and CD-AFM are typically used for inline CD metrology in the semiconductor industry. Each of these toolsets has its own set of limitations, merits, and assumptions based on the measurement technique and optimization algorithms. For example, the correlation between geometric parameters, variation in material properties (n and k), and long model-optimization time places limits on scatterometric metrology. Similarly, the shrinkage phenomenon, charging effect, and profile variations negatively affect the measurement performance of the CD-SEM toolset. A CD-AFM suffers from issues such as tip wear, tip characterization, low throughput, and inaccuracy when measuring dense and narrow structures and at the bottom of the profile. These toolsets also have their own application set in which they perform better relative to others, e.g., direct product measurement (CD-SEM),

Table 8.1 Summary comparison of various CD metrology techniques.

	Scatterometry	CD-SEM	CD-AFM	TEM/XSEM
What to measure	CD, profile, other	CD	CD, profile	CD, profile, other
Where to measure	Periodic grating	Any	Any	Any
Time to solution	Days to weeks	Minutes	Hours	Hours to days
Destructive	Negligible	Minor (resist)	No	Yes
Time to measure	Seconds	Seconds	Minutes	Days
Summary of strengths	• Fast measure • Most profile info	• Quick setup and fast measure • Measure anywhere	• Most profile information • High accuracy	• Full profile information • High accuracy
Assumptions and limitations	• Model assumptions • Trade-off between accuracy and precision • Requires grating	• Constant and uniform profile • No profile info • Difficult to measure the true bottom	• Tip wear and characterization • Large space • Low throughput	• Resolution is process dependent • Limited statistics
Typical fab usage	"Workhorse" for CD and profile	"Workhorse" for CD	• Good RMS • Partial inline	Absolute reference
Issues possibly alleviated by hybrid metrology	• Improved parameter correlation and sensitivity • Reduced setup time	Reduces measurement error due to profile variation	Eliminates restriction on structure (tight spaces)	

profile metrology (scatterometry and CD-AFM), tight precision (scatterometry and CD-SEM), etc. One toolset might be slower but better suited for full profile measurement, whereas another toolset might provide a faster answer that is limited to top-down CD only. Table 8.1 compares some of the high-level characteristics of these inline toolsets along with common destructive-reference metrology techniques [transmission electron microscopy (TEM) and cross-section scanning electron microscopy (XSEM)] for comparison.

Each technique acts like a "specialized filter," providing certain information about the measurand with high confidence. No technique provides all of the information with enough confidence. Each metrology technique employs some degree of signal modeling to extract the CD or profile values from the raw data. SEM techniques interpret a greyscale image to identify boundaries, a CD-AFM interprets physical traces, and scatterometry OCD interprets optical spectral traces. With all interpretations there is a varying degree of uncertainty from the assumptions made about the parameters (of measured structure) that can vary and are explicitly modeled, and from the parameters that are fixed and whose actual variation is neglected. Parameters that are assumed fixed by the model but actually vary are known to give rise to accuracy errors (type B); parameters assumed variable by the model give rise to precision errors (type A). These errors are briefly discussed in Chapters 2 and 3.

The following are characteristics of each metrology technique listed in Table 8.1:

1. Scatterometry can model and measure the minute details of a structure, such as footing/undercuts and re-entrant profiles. Scatterometry is sensitive to the full profile detail, as well as multiple layers underneath. The more parameters floating in the model, the more likely it is to capture all real variations of the structure (better accuracy) at the expense of interparameter correlation (worse repeatability). Furthermore, scatterometry can only measure specific periodic structures and requires a lengthy model set-up procedure. Within the scatterometry framework, more information can be added in order to reduce the correlation and increase the accuracy of the measurement—but a trade-off always exists.
2. A top-down CD-SEM can measure the CD with high precision, but it neither captures the profile (thus cannot detect the undercut for an accurate measurement of the bottom CD) nor provides a robust measurement of the top CD when there is significant top-edge erosion. Similar to scatterometry, there is some undesired correlation between the reported CD and the assumed values of top-edge erosion, bottom footing, or SWA. The assumption of fixed erosion or angle is implicitly considered in the measurement algorithm (image-processing setup). Actual variations of these fixed parameters ultimately affect the accuracy of the measurement. Top-down CD-SEM cannot measure the CD of re-entrant profiles, and the electron beam also shrinks the resist during measurement. On the positive side, CD-SEM does not require specific measurement targets (like scatterometry) and has a quick recipe set-up time.
3. A CD-AFM can provide statistically robust, nondestructive measurement of the CD and profile. However issues such as tip wear, tip characterization, low throughput, and inaccuracy when measuring at the bottom of structures limit a CD-AFM's measurement performance. The manufacturing challenge of flared CD probes with the desired geometry for the CD and profile measurement renders a CD-AFM unusable for certain structures (for example, when space is smaller than the 40 nm or high-aspect ratio structures).
4. Cross-section SEM and TEM images are often used for profile and CD reference measurements. Such techniques are destructive, require special sample preparation, and are more applicable in the case of post-etch patterns than patterns in resist. Cleaved samples of resist wafers can be used for XSEM imaging, but the image still suffers from possible sample-preparation-induced deformation and from the resist shrinkage due to the

energetic electron beam compromising the true CD measurement. Poor statistics are inherent in this technique, so sample roughness contributes to measurement uncertainty.

8.2 Hybrid Metrology: Synergies of Multiple Techniques

As discussed in the previous section, there is no "perfect toolset for all applications;" each toolset has specific strengths. A CD-AFM can "get" most of the profile right for low-to-medium aspect-ratio structures; a CD-SEM "gets" the CD right when the SWA and rounding are known; and scatterometry results improve when more parameters are known and fewer need to be floated. Combining the strengths of different techniques might produce an outcome that is greater than the sum of the parts.

Hybrid metrology is defined as the use of two or more metrology toolsets (Fig. 8.1) to measure aspects of the same dataset. Data obtained from one toolset is exchanged with another toolset and used in a complementary or synergistic way to enhance the resolving power of either or both tools and improve the overall measurement performance. Hybrid metrology could be implemented virtually, whereby data between tools is exchanged via host-automation systems or servers (see Fig. 8.2), or physically, whereby two techniques are mounted on same hardware.

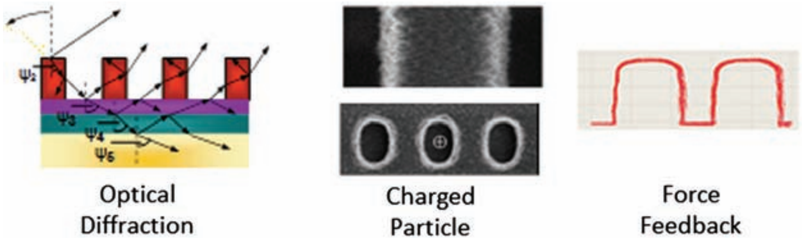


Figure 8.1 Diagrams of various techniques.

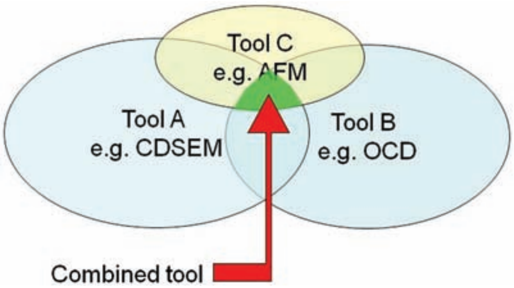


Figure 8.2 Virtual hybrid tool.

For example, both CD-SEM and scatterometry OCD can measure CD of a given structure. OCD also measures the SWA. By adding the CD information from a CD-SEM to the scatterometry model, the CD and SWA parameters can be better decoupled in OCD, thereby effectively improving the SWA measurement. Conversely, the SWA information measured by scatterometry can be used to improve the CD-SEM image-processing threshold settings for a better CD measurement. In a more-complex setup, the CD and SWA measurements from a CD-SEM and CD-AFM might both be added to scatterometry modeling to further reduce the number of floating parameters, thus enabling more-accurate measurement of weaker profile parameters, such as undercut/footing, bowing, arching, or re-entrant CD.

Another example from a film metrology use case: both optical techniques (such as ellipsometer/OCD) and x-ray techniques (such as XPS/XRF/XRD/XRR) can potentially measure the concentration and thicknesses of an ultra-thin layer. An x-ray measurement is usually a direct measurement of concentration; however, it's very slow and might not be sensitive to the thickness of layers below the top thin layer. Optical techniques, although very fast, are usually model-based techniques, i.e., indirect measurement. They may have limited sensitivity to the concentration of the ultra-thin layer through optical dispersion contrast and therefore would provide all thicknesses in a fast throughput mode (suffering from inaccuracies due to limited sensitivity). In an HM ecosystem, both x-ray and optical systems are used in unison, whereby results/data from an inline x-ray tool (with a limited sampling) are provided to optical tools to eliminate or reduce the parameter uncertainty by requiring the measurement of fewer parameters in their optical models. This combination provides a solution that is fast, manufacture-ready, and has the least measurement uncertainty.

Hybrid metrology can be employed two ways:

- recipe development (offline during metrology set up), and
- product measurement (online during actual product measurement).

The former uses data from multiple toolsets as a reference to optimize the recipe setup for the best individual performance. After a recipe is set up, the actual online measurement is still performed independently (the traditional way). The latter combines data from multiple toolsets in real time during actual product measurement.

When combining data from multiple metrology tools, the "secondary or source toolset" can be defined as the tool that measures first (generally) and provides the information to the "primary or receiver toolset" (refer to Fig. 8.3). In this scheme, depending on the attributes/traits of the structure to be measured, a particular toolset could be designated as either primary or secondary. The key is to provide previous information (from a secondary tool) to the primary tool during some stage of its measurement algorithm.

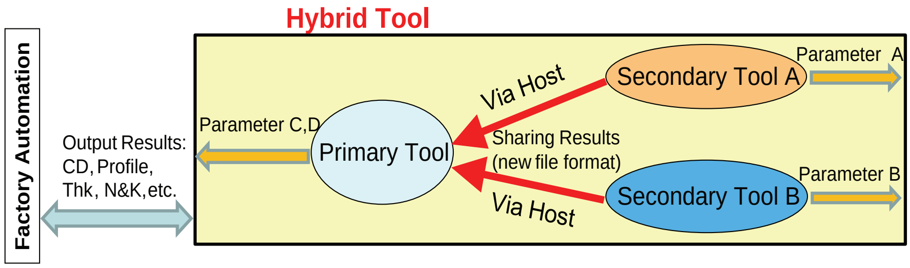


Figure 8.3 Simple HM schematic.

Data from a secondary tool serves two purposes: improving the measurement performance of the primary tool by providing additional data, and as reference system feedback, ensuring stable metrology performance.

The total measurement uncertainty (TMU) could be used to quantify the accuracy benefits realized by one metrology technique over another (including HM), as explained in Chapter 3. It is helpful, however, to clarify how TMU analysis is used in the context of HM, where there are two different tools under test (TuTs): the primary, or receiving, tool and the secondary, or source, tool. In HM, analyzed information [the shared parameter(s)] from the secondary tool is fed into either the primary tool, analysis software associated with the primary tool, or independent analysis software. This information, along with the raw signal information of the primary tool, is then analyzed by the software associated with the primary tool or the independent software, which is where the hybridization of the data occurs. The corresponding output (the final parameter of interest) is then fed into the TMU analysis, along with the final parameter of interest from the RMS. This data flow is shown in Fig. 8.4

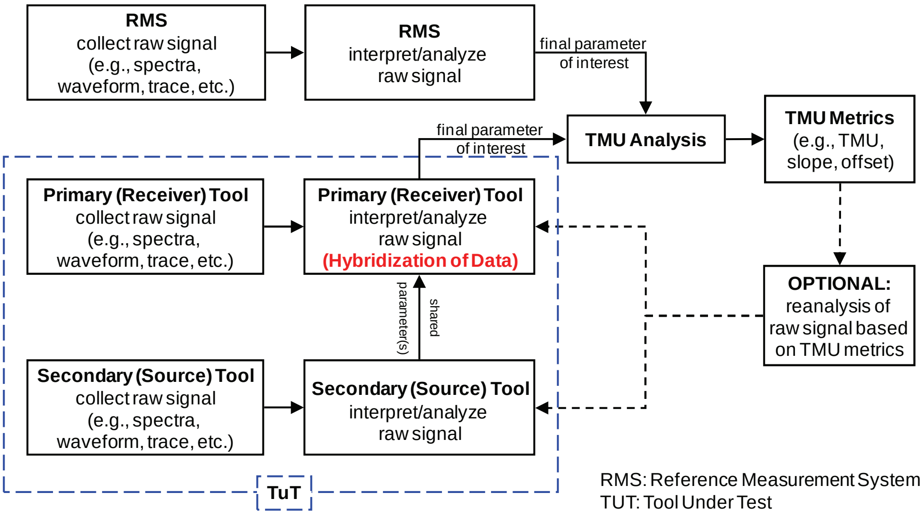


Figure 8.4 TMU analysis to evaluate hybrid metrology.

(compare to Chapter 3). Note that now the TuT is no longer defined as a single tool but as a combination of both the primary and secondary tools. There is still only one TuT. Also note that the flow of information between the primary and secondary tools in Fig. 8.4 is unidirectional. This type of HM, the simplest kind, is called “unidirectional hybrid metrology.” A scenario could exist where the information from the primary tool is sent to the secondary tool to enhance the secondary tool’s measurement. In this case, a feedback loop could be set up between the tools so that they could assist each other in optimizing their output to provide the best possible overall measurement. The labels “primary” and “secondary” begin to lose their meaning in “bidirectional hybrid metrology.” In principle, three or more tools could work in conjunction. If data feedback loops are set up between such tools, the most general case of “multidirectional hybrid metrology” would result.

8.3 Types of HM Implementation

After the HM solution/recipe has been verified and developed, it can be implemented in various ways for inline metrology. During inline implementation, a data exchange between secondary and primary toolsets can be accomplished either through fab automation systems (commonly called “host”) or independent, private channels between the toolsets, such as a hybrid computing server. Schematics of both implementation types are shown in Fig. 8.5. Both approaches have their own benefits and challenges, and the

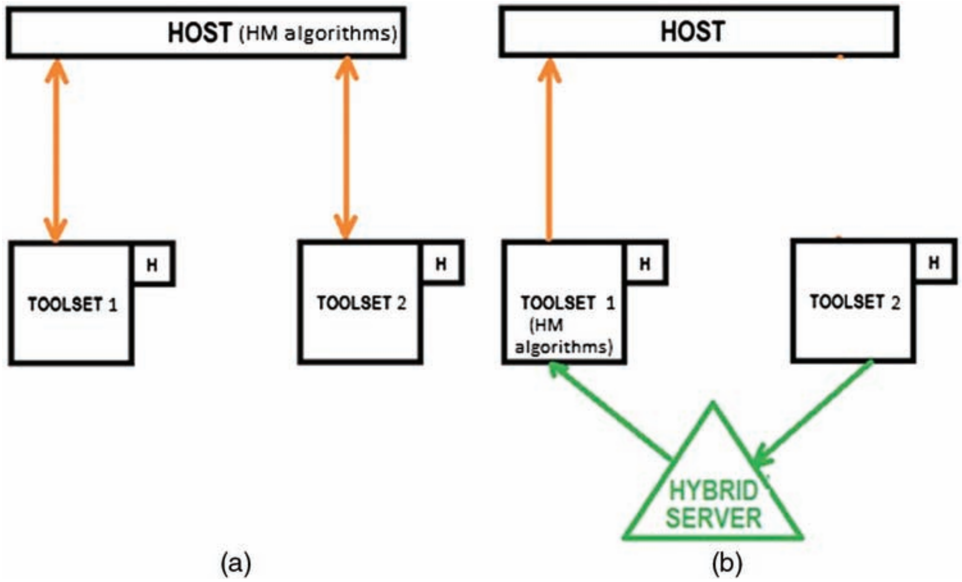


Figure 8.5 Schematics of HM implementation types: (a) host-based data exchange and (b) server-based data exchange.

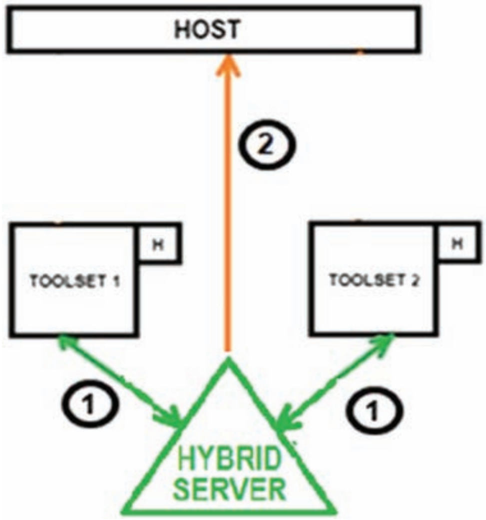


Figure 8.6 Schematics of co-optimization HM.

optimal implementation is based on a fab's specific requirements. The host-based implementation can utilize standard APC protocols that are already implemented in most fabs for controlling process tools. This approach may require some upfront investment from a fab to set up the automation infrastructure, but it provides a higher level of tool independence and flexibility because it is a relatively open frame architecture. A server-based implementation, however, implies that at least two metrology tool vendors must work together and collaborate with the fab. This arrangement may require relatively less upfront infrastructure investment (no change at the host software level), but it requires individual vendor cooperation and investment for each new vendor that will connect to this hybrid server system.

Recent advancements in the area of HM include co-optimization methodology, which is currently being explored for the most advanced and complex structures (such as 14-nm and smaller 3D structures). In cases where the difference in the physics of measurement techniques for the primary and secondary toolset might introduce more measurement noise than useful signal, the "standard" HM might fail to improve the results. When the raw data from both secondary and primary toolsets (shown as toolset 1 and 2 in Fig. 8.6) is simultaneously interpreted or optimized to improve a given success metric in a complementary methodology, such implementation is referred to as co-optimization hybrid metrology. Examples of raw data include measurement images (in the case of a CD-SEM), diffraction spectra (scatterometry), ellipsometric spectra (optical thickness metrology toolsets), and trace images (AFM toolset).

References

1. A. Vaid et al., "A holistic metrology approach: hybrid metrology utilizing scatterometry, CD-AFM, and CD-SEM," *Proc. SPIE* **7971**, 797103 (2011) [doi: 10.1117/12.881632].
2. A. Vaid et al., "Implementation of hybrid metrology at HVM fab for 20 nm and beyond," *Proc. of SPIE* **8681**, 868103 (2013) [doi:10.1117/12.2012339].
3. R. Silver et al., "Nested uncertainties and hybrid metrology to improve measurement accuracy," *Proc. SPIE* **7971**, 797116 (2011) [doi: 10.1117/12.882411].
4. B. N. Taylor and C. E. Kuyatt, "Guidelines for Evaluating and Expressing the Uncertainty of NIST Measurement Results," NIST Technical Note 1297 (1994).
5. ISO Guide to the Expression of Uncertainty in Measurement, (ISO, Geneva, 2006).
6. A. Vaid et al., "Hybrid metrology universal engine: co-optimization," *Proc. SPIE* **9050**, 905009 (2014) [doi:10.1117/12.2048939].
7. R. Silver et al., "Improving optical measurement accuracy using multi-technique nested uncertainties," *Proc. SPIE* **7272**, 727202 (2009) [doi: 10.1117/12.816569].
8. N. Rana and C. Archie, "Hybrid reference metrology exploiting patterning simulation," *Proc. SPIE* **7638**, 76380W (2010) [doi: 10.1117/12.848311].

Chapter 7
Metrology Toolsets in IC Manufacturing: Additional Metrology Systems

7.1 X-Ray Metrology

Most optical results, such as composition/concentration, etc., are based on indirect measurements extracted using a model-based approach. Although the metrological performance of such measurements is satisfactory for most cases, they become inadequate as the semiconductor industry moves toward ultra-thin layers and complex multi-stacks. In those circumstances, x-ray metrology solutions are becoming popular not only as lab solutions, but also as inline metrology solutions. High-resolution x-ray diffraction (HRXRD), x-ray fluorescence (XRF), x-ray reflectance (XRR), and x-ray photoelectron spectroscopy (XPS) are some of the techniques that are being developed for use (at various level of maturity) in inline metrology. Table 7.1 is a high-level summary of these four techniques, along with application use-cases and their strengths/challenges.

The XPS metrology technique was adopted as an inline toolset approximately when the semiconductor industry introduced the high-*k* metal gate (HKMG) process—a significant transistor-gate processing advancement adopted at the 45-nm/32-nm technology node that enhances the performance of semiconductor products. The XPS technique can accurately and precisely characterize ultra-thin high-*k* and gate metal layers within the HKMG stack, a key measurement that has a direct impact on product yield and performance. Similarly, the HRXRD metrology technique was introduced inline with the advent of strain process engineering that utilizes silicon germanium (SiGe) as a strain-enhancing material in the transistor's source drain to enhance the mobility of carriers (thereby improving transistor performance). HRXRD toolsets can accurately measure thickness, composition, and stress for such SiGe layers and allow semiconductor manufacturers to control the strain epitaxial processes. Table 7.1 presents some relevant

Table 7.1 High-level summary of major x-ray metrology techniques.

Attributes	HRXRD	XRR	XRF	XPS
Target size	~100 μm × 100 μm	~600 μm × 200 μm	~50 μm × 50 μm	~50 μm × 50 μm
Measured parameters	Thickness, composition, strain, relaxation	Density, thickness, roughness	Metal film composition	Thickness and composition of thin films
Sample applications	Epitaxial process (SiGe, Si:C)	HKMG layers, silicide, Cu seed and barrier	W:CMP, Cu:CMP, bump (far BEOL)	Gate oxide (Thk and N dose), high-*k*, metal gate films
Strengths	• Information on individual layers in a multistack. • High sensitivity to composition and strain.	• Information on individual layers in a multistack. • Various types of applicable films.	• Multi-elemental analysis. • Excellent precision. • Small spot size.	Can detect most elements and provide chemical bonding information
Challenges	• Films must be epitaxial and crystalline. • Accurate modeling required.	• Large spot size. • Relatively slower.	Requires standards for analysis calibration	Can only analyze the top portion of the surface (~10 nm)

examples of lab techniques that have been adopted as inline metrology techniques to meet the requirements of semiconductor processing advancements and manufacturing.

7.2 *In situ* and Integrated Metrology

All IC process equipment have some degree of *in situ* metrology built inside to precisely control process conditions, such as

- temperature monitoring in resist baking and rapid thermal processes,
- pressure and gas flow monitoring and control in etch and CVD chambers,
- ion energy and flux monitoring in an ion implant, and
- voltage and current monitoring in plasma chambers.

Not all of these examples are discussed here; rather, a select few are used to highlight typical *in situ* metrology cases in IC manufacturing processes, e.g., lithography clusters (scanner and resist track): light flux, dynamic focusing and leveling, alignment, and *in situ* CD measurement after development for etch bias adjustment (to put an after-etch inspection (AEI) CD on target). Etch-chamber endpoint detection is another example.

One of the trends emerging in semiconductor metrology involves the widespread use of integrated techniques (such as optical/scatterometry) to control complex, advanced processes. Integrated metrology typically means that a metrology toolset, compressed into a relatively smaller size, is integrated into the hardware of process tools, such as CMP polishers, etch tools, lithography clusters, etc. This feature helps minimize the overall wafer cycle time and the mean time to detect excursions. Furthermore, it allows more sophisticated control of advanced processes by reducing their variability with wafer-level APC. The measurement parameters could be thickness, CD, depth, overlay, scanner focus, scanner dose, etc. Given the increasing number of measurement steps (as described in previous chapters) for advanced semiconductor technology nodes, the integrated metrology toolsets help reduce the overall wafer cycle time and improve the mean time to detect (MTTD) excursions. These gains exist because the wafers are measured while they are being processed on the process equipment, thereby eliminating the need to measure on standalone toolsets.

7.3 Critical-Dimension Atomic Force Microscope (CD-AFM)

A CD-AFM is another high-resolution measurement system that can directly measure small features down to the nanometer scale. An AFM works by scanning a fine ceramic or semiconductor tip over a surface much the same way as a phonograph needle scans a record. The probe tip is brought into very close proximity of the sample and scanned across the surface. The tip is positioned at the end of a cantilever beam shaped like a diving board. As the tip is repelled by or attracted to the surface, the cantilever beam deflects; the magnitude of the deflection is captured by a laser that reflects at an oblique angle from the very end of the cantilever. A plot of the laser deflection versus the tip position on the sample surface provides the resolution of the hills and valleys that constitute the topography of the surface (Fig. 7.1). The AFM can work with the tip touching the sample (contact mode), or the tip can tap across the surface (tapping mode) like the cane of a blind person.

Forces between the atoms in the probe and the atoms in the sample cause the probe to deflect as the topography changes. Measuring the amount of probe tip deflection allows one to determine the surface profile. During normal CD-AFM operation, the device uses the tapping mode, and the AFM tip never touches the sample surface—it thus has minimal interaction with the samples it measures, which is one of the advantages that it has over a CD-SEM. To measure the CD of a defined feature, the special shaped tip (flared shape, as shown in Fig. 7.2) must be used, which can scan both sidewalls (with vertical or near-vertical profiles) and flat surfaces, as shown in Fig. 7.2.

In a CD-AFM scan, the spacing and time intervals between data points are not constant, as compared to traditional top-down AFM, because it needs to climb up and down sidewalls, as well as flat surfaces of a feature. A full 3D scan of a measurement area must be achieved with a raster scan (scan both directions). If the scan direction is defined as the xz plane by combining multiple scans in the y direction, one can get a 3D profile of the scan area after scan image rendering, as shown in Fig. 7.3.

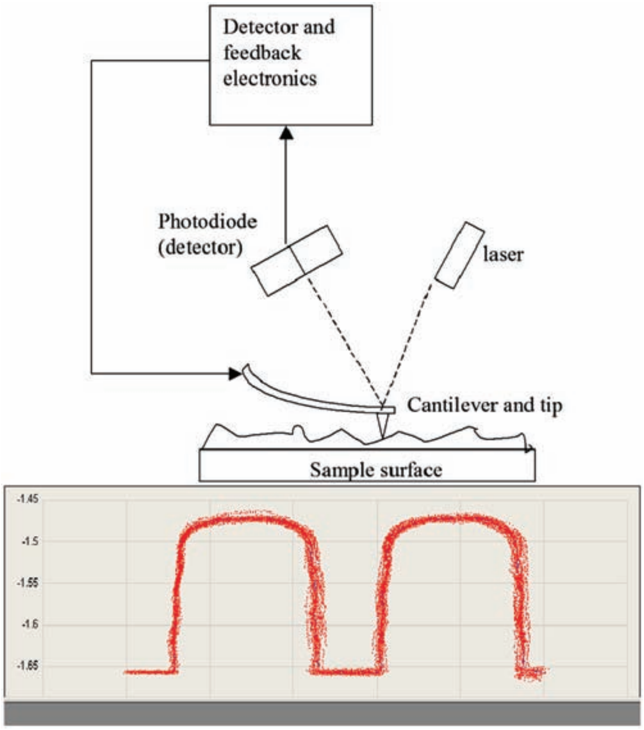


Figure 7.1 An illustration of the CD-AFM working principle and an example of an AFM trace image.

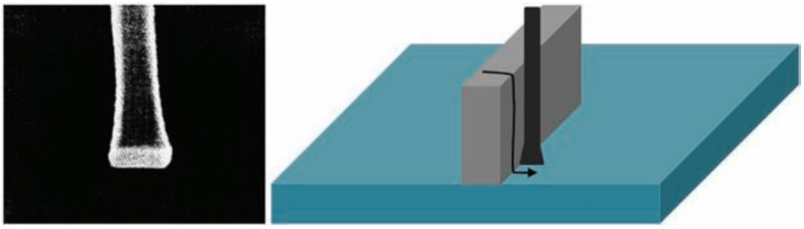


Figure 7.2 A flare-shaped AFM tip scans across a line to form a 2D shape in the *xz* plane.

In such scanning probe toolsets, impact of the tip on results is always a consideration that needs to be well comprehended. This tip influence could be due to tip wear, dilation or tip interaction with sample. Therefore, it is important to understand the tip geometry. The main parameters of interest for CD tips are the width, effective tip length, flare length (also known as overhang), and vertical edge height (VEH), as shown in Fig. 7.4 (top). The CD-AFM toolset can measure re-entrance profile structures (the bottom of a feature is narrower than its top) to a certain degree, whereas other top-down scanning measurement systems cannot. This ability is possible largely due to

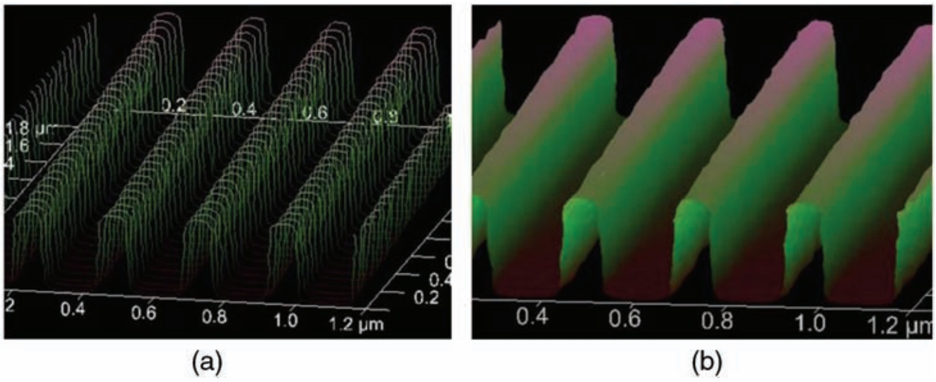


Figure 7.3 (a) AFM area scans and (b) rendered 3D scan image. Images adapted from Bunday[2] with permission.

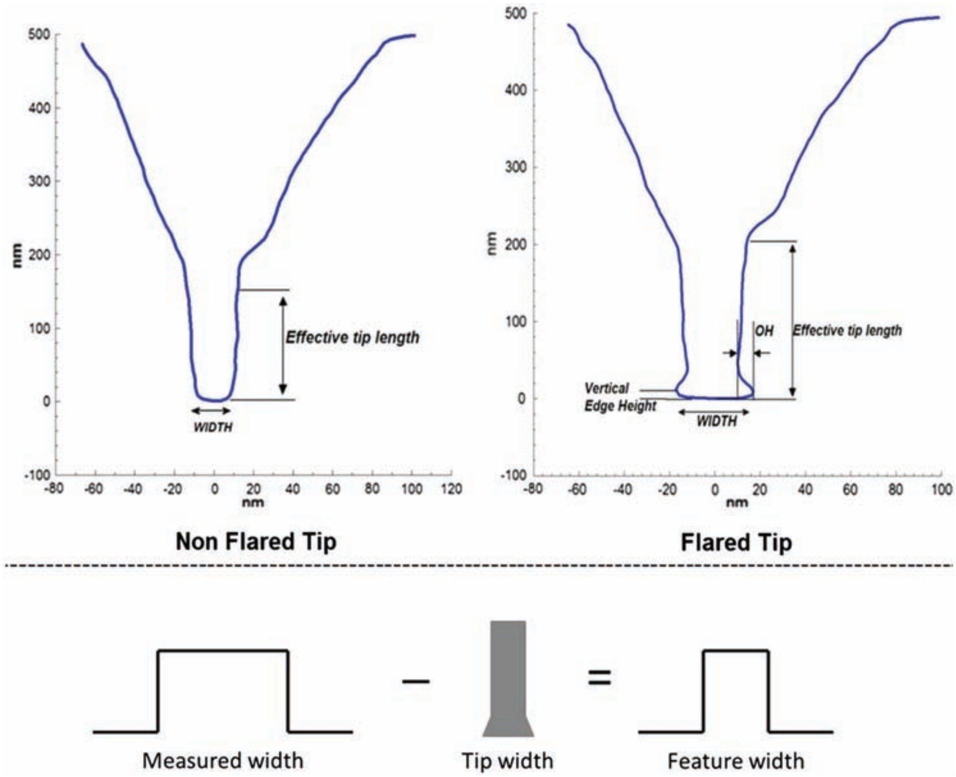


Figure 7.4 Top: Profiles of nonflared and flared tips. Bottom: AFM tip deconvolution to obtain the feature width from the scan width with a known tip width.

the shape of flared tip. Figure 7.4 (top) illustrates a diagram of a flared and a nonflared tip, highlighting some of the key geometrical parameters.

Before CD measurement scans, the tip shape must be carefully calibrated and extracted. The raw scan data is convoluted by the tip shape with the feature shape to be measured. Only after the tip deconvolution, the “true” feature shape can be revealed and relevant CD can be measured, as illustrated by Fig. 7.4 (bottom).

An AFM scan contains very rich information, such as CDs at different heights, depth or height, sidewall angle and profile, sidewall roughness, etc. However, such information is not fully utilized yet. The nondestructive, accurate cross-section profile measurement is considered one of the major advantages of the CD-AFM as a diagnostic tool and a reference measurement system to help qualify other, faster metrology systems. In particular, it is quite useful for an OCD (i.e., scatterometry-based CD measurement systems) to develop and qualify a library with different profiles.

A CD-AFM typically has three modes of operation, as shown in Fig. 7.5:

- Tapping mode uses *z* feedback with a constant speed and minimal sample interaction.
- Deep-trench (DT) mode is commonly used to help with the depth measurement of deeper structures, and a feedback algorithm is optimized to ensure the best measurement performance even though the scan speed is variable.
- CD mode is used to extract CD information using both *z* and *x* feedback and optimizing the scan speed and angle to obtain the most profile sensitivity, depending on geometrical characteristics of the sample structure.

A significant portion of CD-AFM measurement optimization involves identifying the optimum tip type, geometry, and size to ensure the most accurate and precise measurement as well as the least tip wear and longer tip life. This optimization of tip selection is not a trivial process, especially for advanced and smaller semiconductor structures because even though there are several types of commercially available CD tips, it is challenging to manufacture tips that have both a smaller width and smaller VEH with optimal overhang—geometrical tip attributes that are needed to measure

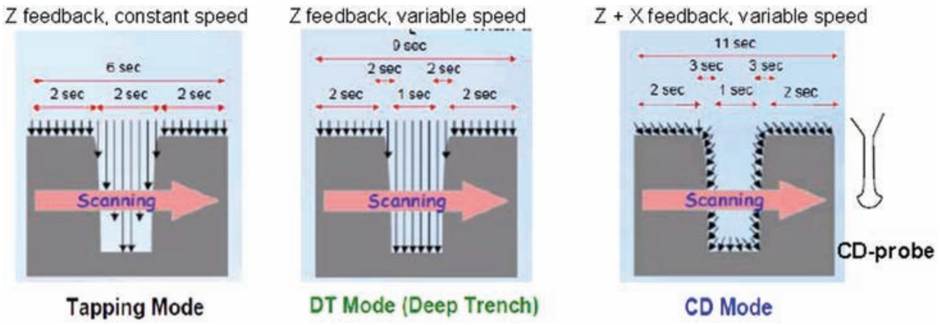


Figure 7.5 Illustration of typical AFM modes used in the semiconductor industry.

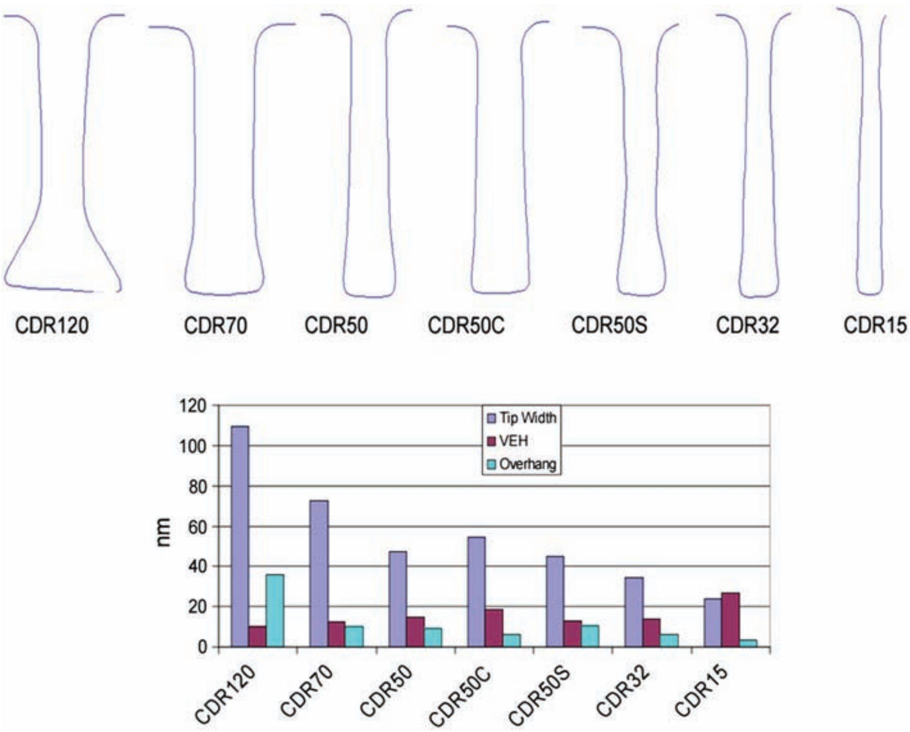


Figure 7.6 Top: Examples of tip types for CD-AFM metrology. Bottom: Relation of geometrical parameters of these sample tips.

small pitch structures with a higher aspect ratio (representing advanced 20-nm technology and beyond structures). Profiles of some sample tips with varying geometry are shown in Fig. 7.6 (top): CDR120, CDR50, CDR50C, CDR50S, CDR32, and CDR15. The number in these industry names corresponds to the nominal tip width (in nanometers). For example, CDR50C has a carbon coating, and CDR50S uses a special process that produces a shorter, stiffer neck. These tips differ in material and geometrical parameters. A smaller tip width is typically desired when measuring critical, smaller pitch structures, whereas wider tips are used to measure larger geometries. CDR15 is the smallest CD probe available in this sample. Figure 7.6 (bottom) shows that as the tip width reduces, the overhang reduces and the VEH increases, which typically leads to poor mechanical strength and an inability to measure advanced smaller pitch structure—therefore stressing the need for both optimal tip identification for measurement and improved, next-generation tip manufacturing.

A CD-AFM exhibits good accuracy compared to other scanning-scope-based metrology systems, such as a CD-SEM or scatterometry, thus making it a tool of choice for reference metrology (in cases where the tip can resolve the geometry of the structure in question). However, it also suffers from some disadvantages:

(a) the throughput disadvantage—accurately locating the measurement site, advancing the delicate tip toward the measurement surface, and a relatively slow scan speed all have a negative effect;
(b) the tip size limit when measuring small trenches and contacts;
(c) the tip suffers potential contamination, wear, and damage, which can affect CD measurements; and
(d) an AFM also suffers from relatively higher operational expenses in production implementation because tips/probes must be replaced as they wear. CD-AFM tip calibration uses known width or shape artifacts. A known-width silicon line calibrates a tip width, and then a silicon ridge calibrates the tip shape.

Although a CD-AFM has its limitations, an innovative scan mode can turn them into advantages. The DT depth can be measured rather accurately by changing the tip from a flared shape to a thin, long version and by switching the scan mode from "dithering" to HAR DT tapping mode. A CD-AFM is material independent in IC process applications, unlike CD-SEM or optical metrology systems. In additional to feature depth or height measurements, a CD-AFM can also be used for LER/LWR with incremental scans in the y direction.

References

1. B. Su, K. Rajikumar, and M. Agrawal, "Limitation of CD AFM on resist footing detection," *Proc. SPIE* **3332**, 151 (1998) [doi: 10.1117/12.308763].
2. B. Bunday et al., "Phenomenology of electron-beam induced photoresist shrinkage trends," *Proc. SPIE* **7272**, 72721B (2009) [doi: 10.1117/12.816249].
3. J. Foucher, N. Rana, and C. Dezauzier, "3D-AFM enhancement for CD metrology dedicated to lithography sub-28nm node requirements," *Proc. SPIE* **7638**, 763802 (2010) [doi: 10.1117/12.846550].
4. N. G. Orji et al., "Technique for AFM tip characterization," *Proc. SPIE* **9173**, 917305 (2014) [doi:10.1117/12.2062759].
5. N. Rana et al., "Reconciling measurements in AFM reference metrology when using different probing techniques," *Proc. SPIE* **7971**, 797117 (2011) [doi:10.1117/12.879516].

Chapter 9
Metrology in Mask Making

Photomask technology has become a critical part of IC manufacturing in recent years. It is the link between IC design and IC manufacturing through lithography and etch image-transfer processes. The heart of the enabling lithography technology (in keeping with Moore's law) involves transferring the pattern images from a mask to the wafer plane with acceptable distortion and maintaining reasonable pattern fidelity. Nearly all lithographic resolution enhancement techniques (RETs) involve masks in some degree. The exposure light path (OAI, immersion), light intensity (OPC), and phase (PSM) are all modified to increase the image information transfer through an optical system of scanners. Other areas, such as polarization and illumination shape adjustment to match the exposed mask patterns (SMOs), are also actively tapped to maximize the lithography process window. All of these techniques make the photomask production process and qualification process especially important. Furthermore, because a mask is used to expose every die on wafers, the quality of a mask directly affects pattern qualities on wafers and thus the product yield.

Mask qualification involves feature size measurements, pattern placement measurements (feature-to-feature position), etch depth and phase measurements, and mask defect inspection. Production-grade masks must be printable and defect free. This chapter focuses on mask CD measurements and mask inspection.

There are several mask types that have been proposed and studied over the years, including the Cr-on-glass (COG) binary mask, attenuated embedded PSM (EPSM), alternating aperture PSM (APSM), tri-tone PSM, chrome-less phase lithography (CPL), and pixelated phase mask. Among them, only two types are widely used in IC production: COG and EPSM (or some variations of EPSM). This chapter focuses on the latter with respect to mask metrology and inspection.

As the IC industry moves to advanced technology nodes below 20 nm, mask complexity has increased by orders of magnitude, due to design complexity increase and aggressive model-based OPC. As a result, the design data file size, mask writing time, and mask qualification time have all increased drastically. Significant effort has been spent to shorten the latter two.

The mask-making process is quite similar to one-layer-wafer processing with e-beam lithography (the majority of today's photomasks are exposed using variable-shaped e-beam mask writers). It involves

- substrate (called blanks, with a precoated Cr or MoSi) cleaning and inspection,
- resist coating and baking,
- variable-shaped e-beam mask writing,
- resist development, and
- RIE dry etch and cleaning.

Tri-layer masks require a second exposure and etch step, after which they go through the mask qualification process (optical mask inspection, defect review, mask repair, CD measurements) and then mask cleaning and inspection again. The pellicle is finally mounted on the mask to protect it from airborne dusts. A final mask inspection is performed to qualify the mask for shipment. Figure 9.1 shows the basic binary mask-making steps.

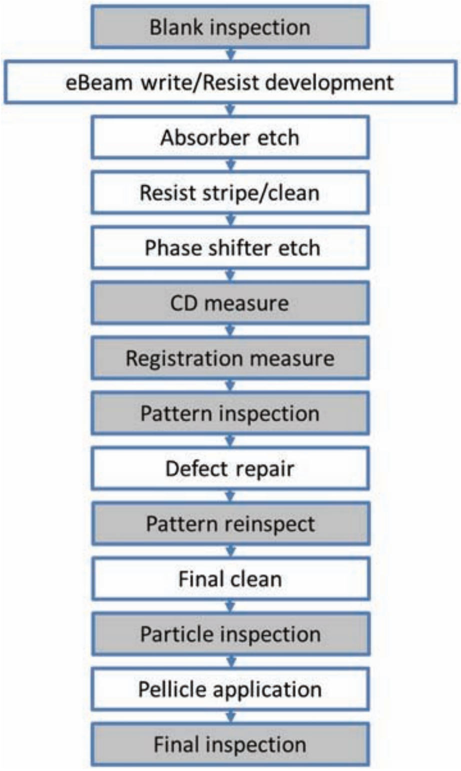


Figure 9.1 Basic binary mask-making process flow. Shaded steps are metrology and inspection.

The mask qualification process involves heavy CD measurements, feature placement measurements (registration measurements), and mask defect inspection. A mask inspection overview is provided here to complete the mask qualification.

9.1 Mask CD Measurements

At technology nodes 130 nm and above, the lithography process can tolerate optical proximity effects, such as corner rounding and line end shortening. Mask making involves rigorously reproducing design patterns—a period can be labeled as "what you see is what you get." Below 130-nm nodes, the process enters the sub-half-wavelength regime, and optical proximity effects are no longer tolerable, and rule-based or model-based OPC must be applied to masks, which largely alters the design data, and the patterns on mask no longer resemble the patterns in the data, i.e., "what you see is not what you get." At advanced nodes 20 nm and below, aggressive model-based OPC—even ILT (inverse lithography technique)—must be applied to the design data with double or triple patterning. The original design data must be used in the downstream steps to ensure the design intent.

Mask features are four times larger; however, sub-resolution assist features (SRAF) are much smaller and well below 100 nm in size. Masks with OPC decoration have forced optical CD metrology to transition to e-beam CD metrology, just like in wafer CD measurements.

Unlike a wafer CD-SEM, a mask CD-SEM has its own uniqueness in mechanical handling and charging control. The majority of masks in today's IC fabs are 0.25-in-thick, 6-in quartz masks, either COG binary masks or MoSi-on-glass attenuated PSMs. Sometimes both Cr and MoSi may be used for tri-layer masks. Therefore, a CD-SEM for masks involves many fewer substrate types as compared to wafers. The major mask CD-SEM makers are also different than wafer CD-SEM makers.

The strong insulating nature of quartz means that charging is a concern, especially for Cr/MoSi islands (difficult for charges to escape). Like a wafer CD-SEM, mask CD measurements use low beam energy (<1 keV) and low beam current (<10 nA). In addition, other charge-up suppression mechanisms, such as a local injection of gases to remove extra charges from the mask surface (e.g., Holon's wind SEM technology), are also used even though the effectiveness of the charge suppression and CD measurement signal may conflict. Another alternative is environmental or low-vacuum SEM, which uses gases in the chamber to relieve charging from the mask under measurement.

In terms of sampling, masks need more CD measurements than wafers because there is normally only one mask to qualify and all process variations are embedded in this one mask; its sampling reflects that. Modern high-end masks normally require thousands of CD measurements, from test patterns and real circuitry patterns to linearity and small feature biases. To achieve a fast throughput for mask CD measurements, a fast scan and accurate stage motion to locate the next measurement site is very important. Accurate stage navigation means skipping pattern recognition for fine beam alignment. Unlike wafers, a mask normally does not have repeating fields or dies, which means that high-accuracy stage navigation is very important to locate the predefined measurement sites. Laser-interferometer control-positioning stages are normal in a mask CD-SEM, and its navigation accuracy can reach the nanometer scale.

Higher e-beam intensity can cause mask film damage due to the very thin film thickness (<80 nm) of Cr or MoSi. The film thickness for advanced node masks can be even thinner. FOV and measurement-box selection must be careful to limit the e-beam intensity (the total electron dose divided by the measurement area). To avoid multiple exposures, a SEM image is often taken in an appropriate FOV to get multiple measurements.

In addition to a CD-SEM, scatterometry metrology tools are also utilized for mask CD measurements. Scatterometry is especially useful in cases where the profile and 3D characterization of mask structures is required. For advanced mask-manufacturing process control, characterization of optical properties and composition becomes important, thereby requiring scatterometry metrology.

9.2 Mask Inspection

Mask defects must be found and fixed before shipment. Real mask defects (also called killer defects) will cause problems on every die on wafers and create a product yield nightmare. Therefore, mask inspection is the critical step in mask qualification and normally done multiple times during mask making before shipment.

Although the main features on masks are four times larger due to strong model-based OPC, SRAFs on masks become very small. When designs processed with advanced OPC, such as ILT the boundary between main features and SRAFs is blurred. In short, OPC and other RETs have made already complex designs even more complex, which poses significant challenges for mask inspection.

Modern optical-inspection systems are still the primary inspection systems for mask inspection. Although the actinic wavelength is preferred in mask inspection, 193-nm-wavelength, stable sources are not always available for mask inspections for various reasons. Only in recent years have such systems become commercially available.

The main advantages of optical inspection systems are their integrated transmitted (T) and reflected (R) light inspection setup, as shown in Fig. 9.2.

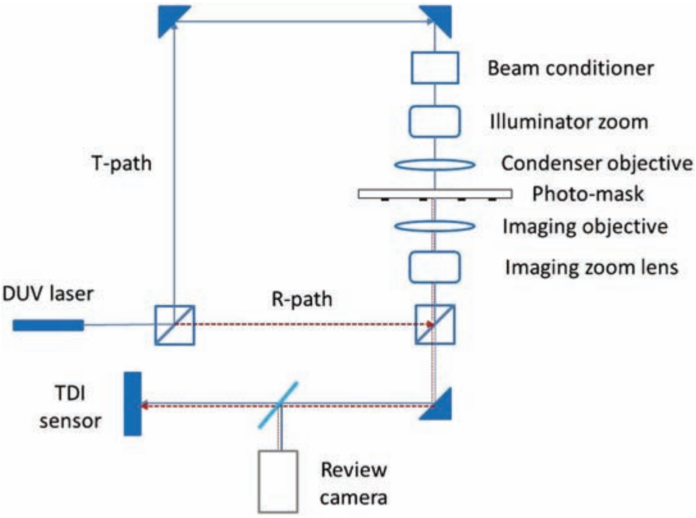


Figure 9.2 Typical mask inspection T/R image-acquisition system. An integrated T and R system allows quick switching to provide real-time images for defect and contamination detection.

Transmitted light inspection closely mimics the wafer lithography process with respect to the light intensity passing through a mask, while reflected light inspection picks up surface defects easily. Combined T and R inspection provides better defect detection and defect classification.

Mask inspection systems use laser raster scans, combined with the stage movement, to cover the entire inspection area. The useable area of a typical 150 mm × 150 mm square mask is 132 mm × 132 mm (outer frame). Lithography scanners further limited that space to 26 mm (104 mm on the mask scale). Therefore, the maximum lithography field size is limited by the scanner optics and mask to 26 mm × 33 mm.

A raster-scanned image is divided into swaths and patches, and the light intensity in each image is digitized into 256 (0 to 255) intensity levels. The light levels are calibrated each time before inspection—once at the dark background level, and once at the highest light intensity—so that the usable light intensity range utilizes the nearly entire range of levels. These patch images are compared with the corresponding references to detect intensity differences in both T and R images using defect-detection algorithms.

The defects are detected during mask inspection by using the image intensity differences with a direct image comparison between the test image and the reference image. The reference image can be the same in the adjacent die in the same mask (DD) or the rendered image from design (DDB). For single-die masks, DDB is the only choice for defect detection.

A mask contains the fiducial patterns, the logo, bar code and labels, alignment marks and test patterns, as well as device patterns, including main design patterns, and optical assist features and dummy fills. To reduce the inspection time, the areas with large frames, the logo, and labels can be blocked out as "do not inspect areas" (DNIAs) so that the inspection machine skips those areas and reduces the entire mask inspection time. In addition, certain areas can be desensitized to reduce the false error rate, whereas other areas can be made more sensitive to make sure there are no missing defects. Thus, previous knowledge or automatic detection of those areas during the data preparation step would help optimize the mask inspection strategy.

The sensitivity of an inspection machine depends on multiple parameters. The test images of the mask inspection are captured using a time-delay-and-integration (TDI) CCD scan sensor. Pixel size—the dominant factor in sensitivity—has been gradually reduced from 120 nm to 90 nm, 72 nm, and 55 nm as technology nodes progressed from 65/45 nm, to 32 nm, 22 nm, and even 16 nm and below. Smaller pixel sizes produce longer inspection times. A 55-nm pixel size will take 30% longer than a 72-nm version to inspect the same mask area.

A defect detected by either DD or DDB methods not only contains the location of the defects but also a rough estimation of other defect characteristics, such as defect size and shape, due to the light intensity difference and number of pixels involved. After the inspection is finished, an inspection report (IR) is generated that contains all defect information.

A mask inspection machine can be used as a live review system to take a closer look at the selected number of defects. This process helps determine detailed information about a defect (e.g., size or intensity difference). However, live review detracts from the inspection machine time and is usually a slow, manual process. Live reviews of expensive inspection machines in production environment are discouraged. Software-based automatic defect classification (ADC) solutions that use defect analysis and simulation have gained momentum recently.

The first separation of killer and non-killer defects is determined by their printability—will the detected defects appear in the wafer plane or not? If so, how big of an impact will they have on surrounding patterns (causing any CD changes, even a pattern short or break)? Either an aerial image measurement system (AIMS) or the simulation method can be used to explore these questions. AIMS mimics the optical setting (wavelength, NA, and sigma) of a scanner and checks the aerial image on the wafer plane (see Fig. 9.3). After a quick calibration (using a known feature to determine the threshold), a defect CD (or printability) can be easily evaluated. The simulation method uses the defect images from mask inspection (both T and R) to reconstruct the local mask-defect patterns using the inspection machine model and then forward simulate the aerial image of that local mask defect in the wafer plane.

Defects that are determined to have an impact on the wafer plane must be repaired. Typical mask repair includes two processes: extra-material removal

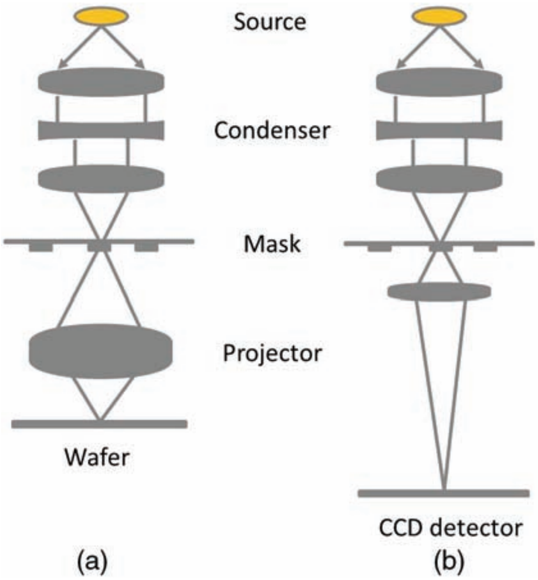


Figure 9.3 (a) Scanner optics and (b) AIMS optics. Note that the optical components are the same for both above the mask.

(local etch) and missing-material fill (deposition). Mask defects in wafer fabs can also be created by the loss of repaired material due to constant high-energy photon exposure. High energy photon exposure of masks also creates another problem: haze buildup on the mask surface. The haze reduces the amount of local exposure light intensity that reaches the wafer plane and can be a yield killer if it is not addressed immediately. For haze and contamination purposes, a mask inspection machine is added in contamination-detection mode, taking advantage of the contamination's strong signal in R and weak signal in T. Thus, periodic inspection (requalification) of masks in wafer fabs (to detect haze buildup and other progress defects) is needed. A mask that is detected with significant haze or contamination is usually returned to a mask shop for cleaning, inspection, and qualification before resuming use.

Electron-beam inspection systems are gaining momentum in mask inspection for their resolution. However, scanning speed, charging (sometimes even damaging mask features due to e-beam bombardments), and detecting only e-beam visible defects limit it to service as a supplement to optical inspection systems. For nano-imprint lithography plates, e-beam inspection is more suitable because it transfers patterns through nano-molding.

References

1. B. G. Eynon and B. Wu, *Photomask Fabrication Technology*, McGraw-Hill, New York (2005).

2. M. Malloy and L.C. Litt, "2012 Mask industry survey," *Proc. SPIE* **8522**, 852203 (2012) [doi: 10.1117/12.978704].
3. J. Y. Chen, "Transform design to chip, an end user point of view on mask making," *Proc. SPIE* **8522**, 852202 (2012) [doi: 10.1117/12.981267].
4. M. T. Postek, A. E. Vladar, T. Rice, and R. Knowles, "Potential for high pressure/environmental SEM microscope for photomask dimensional metrology," *Proc. SPIE* **5038**, 315–329 (2003) [doi: 10.1117/12.488115].
5. C. Y. Chen et al., "*In situ* repair qualification by applying computational metrology and inspection technologies," *Proc. SPIE* **8701**, 870108 (2013) [doi: 10.1117/12.2030688].
6. E. Guo et al., "Simulation-based mask defect repair verification and disposition," *Proc. SPIE* **7488**, 74880G (2009) [doi: 10.1117/12.829692].
7. E. Guo et al., "Simulation-based mask defect printability verification and disposition, Part II," *Proc. SPIE* **8166**, 81662D (2011) [doi: 10.1117/12.896879].

Chapter 10
Perspectives on Future Challenges and Considerations

This final chapter concludes with some suggestions for future metrology toolset improvements and other themes that require future consideration. Metrology has changed significantly in the past 15 years and will continue to change. It is hoped that these suggestions will become mainstream or at least motivate readers to pursue new ideas that drive positive changes.

10.1 Measurement Tooling Challenges

Figure 10.1 shows how the CD-SEM was the dominant choice in the past for most, if not all, CD applications. As the process complexity increased with 45-nm and 32-nm technology nodes, more information was required beyond the CD, such as the SWA, the intricate dimensional numbers (such as structure rounding, structure footing, depth, etc.) and a feature height. The CD-SEM was not able to handle the applications that required this additional information, and OCD started assuming this role. Furthermore (not shown in the chart), some applications require both techniques to measure a given application and create a better end result via HM. It is clear that beyond the 10-nm technology node the CD-SEM and OCD have sufficient limitations and warrant serious consideration of other techniques for many applications.

Given that the CD-SEM is an image-based technique, it is desirable for it to continue to evolve and overcome its limitations which are: image resolution, measurement-induced CD change, and limitations in providing 3D information. Regarding image resolution, as indicated previously in Chapter 6, there has been limited progress relative to how fast dimensions have been shrinking. Recent promising image-processing techniques are encouraging in their ability to enhance image fidelity, but they are ultimately “workarounds” for limited progress in improving the fundamental imaging capabilities. Cold field emission (CFE) has demonstrated some potential, although it is unclear when it will become a manufacturable solution and what its limitations are.

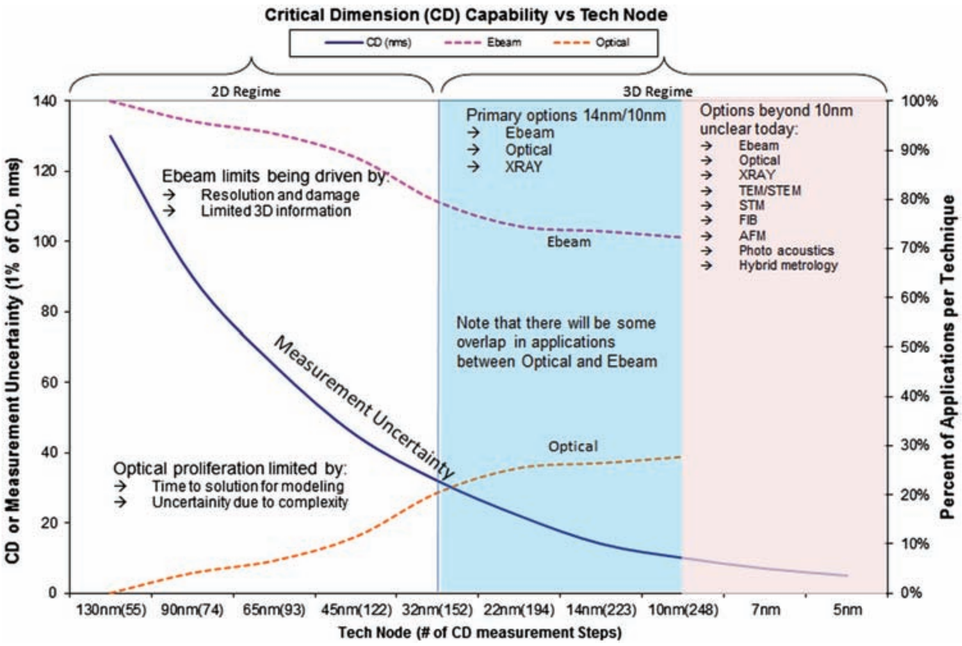


Figure 10.1 Evolution and future speculation of the CD metrology landscape.

Regarding measurement-induced CD change, so long as it is predictable and consistent across the fleet, it is generally manageable. There are concerns that the consistency across the fleet for a given application may have significant effects on process control. Given the tight process tolerances required for the most demanding application, measurement-induced CD differences across the fleet on the order of tenths of nanometers are possible. This factor requires a solution that is not yet available. Lastly, regarding the limitations in providing 3D information, beam tilt has shown promise in measuring the SWA and height. Tilt also suffers from image fidelity concerns, negatively affecting the transition from the slope to the top of the feature. Numerous innovative solutions are needed for the CD-SEM to maintain and increase its dominant role in the future of CD measurements.

OCD is a model-based technique and thus requires considerable overhead when developing applications—a trend increasing further with the advent of 3D devices and complex, multi-channel OCD hardware to resolve such 3D dimensional-measurement requirements. As discussed before, these overheads include building the model, verifying the model using reference metrology, and minimizing correlations between parameters. One of the emerging trends in OCD is the use of smart algorithms and advancements to process the OCD raw signals and reduce the modeling overheads. A couple of key recent advancements include

(a) model-less OCD, wherein a neural-algorithm-based quantitative model (instead of a conventional semi-physical RCWA model) is created that directly links the raw OCD spectra, and
(b) a reference-prediction capability that utilizes multi-perspective optical channels to gather relevant information based on smart algorithms to predict/generate reference metrology, thereby allowing a speedy OCD time-to-solution.

Some other potential future techniques may become mainstream. Numerous AFM advancements are aimed at overcoming its limitations (primarily speed and the ability to scan the tip in the smallest areas, e.g., current small-device spaces). At least one supplier is trying to overcome the speed gap by scanning multiple probes in parallel. Although this change may increase speed it does not address the concern of being able to scan the probe in the smallest areas. Another new advancement in AFM is the development of a mode that allows localized precise-force control and thus better mapping of cantilever action to possibly measure smaller spaces using sub-10-nm tip widths. Further development and innovation is needed for the AFM technique; otherwise, its role in mainstream metrology might be limited, although there is some potential in the defect review area where fitting the tip in the smallest dimensions is not a primary requirement but rather speed is.

Another ongoing trend is the need to measure/monitor the actual devices (or device-like targets) instead of measuring conventional surrogate structures on kerf. With the advent of 3D devices, such as FinFET and smaller structures, correlation of the kerf dimensional and composition measurements to those on actual devices is starting to break. Figure 10.2 shows one such example using data from kerf measurements and device measurements compared to the electrical results on the same wafers—device measurements

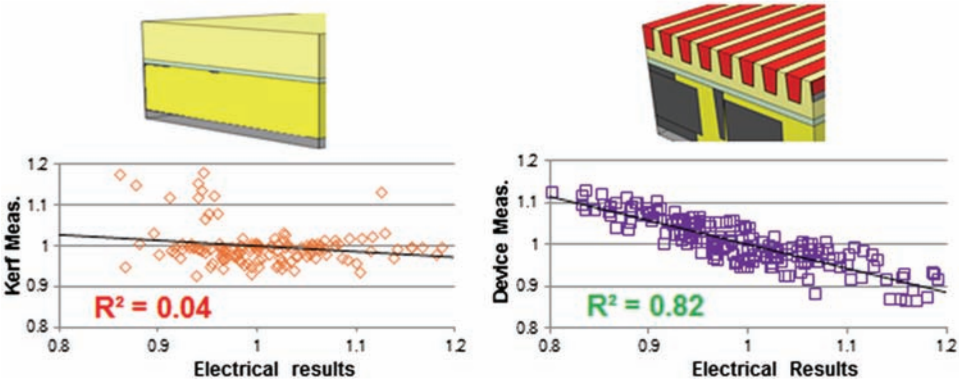


Figure 10.2 Improved electrical correlation using metrology on device targets versus kerf targets.

Table 10.1 Sources of variation for 14-nm FinFET node processes; the data is normalized. Note that the in-die (device) variation is the highest. This example reflects the early development of the 14-nm FinFET process, but it demonstrates the overall need for variation.

	Fin Height	Trench Depth	Gate Height
In-die	3.38%	3.07%	6.23%
Across water	2.34%	2.79%	2.66%
Wafer to wafer	1.04%	2.73%	2.81%

correlate much better with the latter. This is just one exceptional example where the results are extremely in favor of device measurements, but the overall trend is emerging regardless. This paradigm shift could primarily be due to complex CMP and etch processes on 3D devices, the effect of varying local pattern density on dimensions and topography differences across the lithographic exposure field. Furthermore, in-device (device-to-device, depending on its type) or in-die variation is becoming a significantly higher source of variation for leading-edge 3D FinFET nodes [Table 10.1]. The requirement to measure directly on devices or "targets very similar to devices" brings unique and significant challenges to most measurement techniques, such as OCD, ellipsometry thickness/composition, overlay, and x-ray techniques. Two common challenges that all of these techniques share when measuring actual devices are the need for an ultra-small spot size and the ability to resolve the device complexity. Along similar lines, OCD is starting to replace traditional ellipsometry measurements for certain steps due to its ability to measure a device like patterned targets (even if on kerf) as opposed to un-patterned/flat targets that might not correlate to the device.

To enable in-die metrology, ongoing work seeks to reduce the spot size of OCD and thickness tools while simultaneously using smart algorithms and multi-channel optical configurations to allow the measurement of device macros. Similarly, to allow overlay metrology, recent advances use small-spot techniques and augment or amend conventional overlay measurement by using CD-SEM overlay data on devices.

The future trend in hybrid metrology is to enable the simultaneous regression of raw data from multiple complementary techniques to enable the measurement of complex device structures. This methodology is being called "co-optimization hybrid metrology" and requires fab automation infrastructure development and localized computing resources to enable seamless connection of metrology tools. More collaboration between metrology suppliers of different techniques, as well as collaboration with semiconductor manufacturing companies, will allow the development and implementation of further HM use-cases to solve the growing need for 3D device metrology.

10.2 Fault Detection and Control

Fault detection and control (FDC) is a relatively new concept to increase the detection of toolset issues to near real time. The daily monitors of toolset health only run so often (once every day or two). Thus, there is significant risk of not catching an issue between monitor runs. FDC adds an additional layer of visibility to help complement the daily monitors by observing key hardware parameters in nearly real time. Catching issues with tools sooner can decrease the impact on production. The basic concept is as follows:

1. Hardware parameters and their values are sent from the given tool to the fab central database (e.g., the probe current on a CD-SEM is 10 pA on this date and time).
2. The critical parameters to monitor are identified (e.g., if the probe current changes, the tool will measure differently; this should be defined as a critical parameter).
3. SPC charts are set up to monitor these parameters via an FDC host/controller system in the fab (e.g., the probe current should be 10 pA $\pm$ 0.5 pA).
4. Actions are set up against the charts, such as shutting down the tool (e.g., if the probe current is outside the 10 pA $\pm$ 0.5 pA limits, shut the tool down).

The current trend involves implementing this type of capability on every metrology tool in the fab. The metrology toolset owner is responsible for leveraging this capability if the infrastructure exists in the given fab. If it is not available in the given fab, it should be pursued.

10.3 Virtual Metrology

Virtual metrology has become a critical component of semiconductor manufacturing control. The idea is to construct predictive models that can forecast the electrical and physical parameters of wafers, based on data collected from the relevant processing tools. In this way, direct measurements from the wafer can be minimized or eliminated altogether (thus the term "virtual" metrology). Challenges include the selection of the appropriate modeling method and the pretreatment of the raw data.

The deployment of virtual metrology is limited in the semiconductor industry. Although it has many benefits, particularly the reduction or elimination of metrology measurement steps, it has only been selectively deployed because there is significant overhead in its implementation. As discussed previously, there are hundreds of measurement steps in the CD-SEM for various technology nodes; if one were to implement virtual metrology on all of the process tools that these measurement steps support, it would require an army of individuals to implement. Therefore, until significant learning and efficiencies are achieved, the deployment of this method will likely remain in limited deployment in the near future. Regardless, it should be pursued wherever possible to reduce metrology costs.

10.4 Waferless Recipe Writing

Because hundreds of recipes are created for each new chip product, a large amount of time is spent building the recipes live while using the wafers. When each new recipe must be built, the lot is typically held to create the metrology recipe at the corresponding step in the process. The recipes only need to be built once on the first lot of each new product, but this can add up to days of delay. Additionally, metrology recipes are becoming so complicated that they do not perform well unless clear standards are in place about how to build these recipes so that they run robustly, measure well, and are consistent across products for a given layer in a technology node. Creating recipes offline without the wafer, before the lot needs it, is critical to eliminating waste. Furthermore, building them offline reduces the variability amongst recipe writers and should generally lead to better-performing recipes because the recipe builds are restricted to people with the right expertise.

Generally speaking, there are two forms of waferless recipe building:

1. Those that take design data and send them to a specially designed system provided by the metrology supplier that creates the recipe directly from the design data.
2. Those that take common elements from previously constructed recipes to synthesize future recipes for similar situations (this does not typically require a specialized system from the supplier).

It is up to the metrology toolset owner to determine the best way to implement waferless recipe writing for each toolset in a fab. This is a very important strategy to deploy across all of the metrology toolsets.

10.5 The Blurry Line between Metrology and Defect Inspection

Different e-beam tools are used for dimensional metrology and defect inspection (although in some cases they are from the same supplier). Does this really make sense from the perspective of the chip developer and manufacturer? A lot of resources are being used to manage each unique toolset. There are significant cost-of-ownership, route (multiple process steps), and cycle-time considerations in managing all these CD-SEM, EBR, and EBI tools independently at each fab. In recent years, the line between structural and defect metrology has become more blurred. Fundamentally, in all of these cases, an e-beam image is acquired and then either measured (in the case of dimensional metrology) or analyzed/interpreted for defects (in the case of defect inspection and review). Figure 10.3 illustrates an example from an EBI

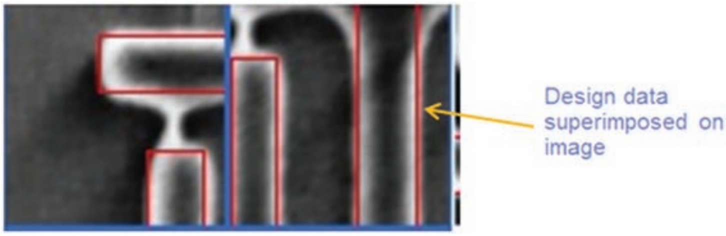


Figure 10.3 Design data superimposed on an e-beam image.

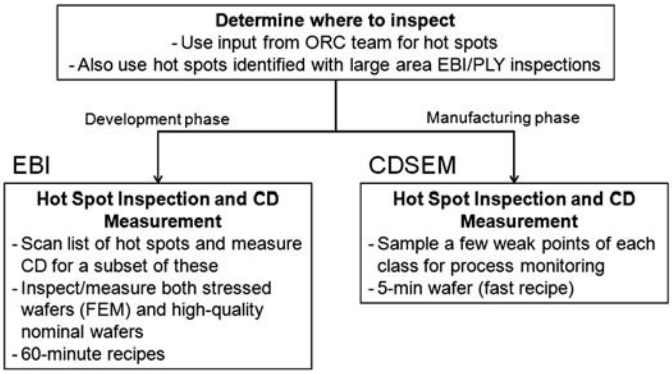


Figure 10.4 An EBI and CD-SEM in defect inspection during the development and manufacturing phases.

tool that shows how scumming was identified in a process-window mapping experiment. Notice that the overlaid design data in the image is used to help identify problematic defect regions by determining which pattern is outside this desired design region. While the EBI defect inspection tool looks for defects, the CD-SEM often measures these same or similar features to map the process window.

The authors feel that there is a better way to use these tools collaboratively. Figure 10.4 shows how the CD-SEM and the EBI could be used more synergistically. The flowchart describes how the EBI is used today to find 'hot spots,' which are locations where the process window falls off rapidly. The EBI tool will scan a large area, which takes a significant amount of time to do (on the order hours), limiting its ability to perform regularly. The information obtained is important to the process and design community because it lets them know which needs should be improved. Certain "hot spots" are particularly useful for process monitoring; those that are most sensitive to process changes are best suited to highlight when the process has drifted from its desired state. Measurements of specific devices/structures often do not catch issues elsewhere in the chip. The most-sensitive hot spots discovered by an EBI during the development phase could be fed to the CD-SEM for process monitoring during the manufacturing phase. Figure 10.5

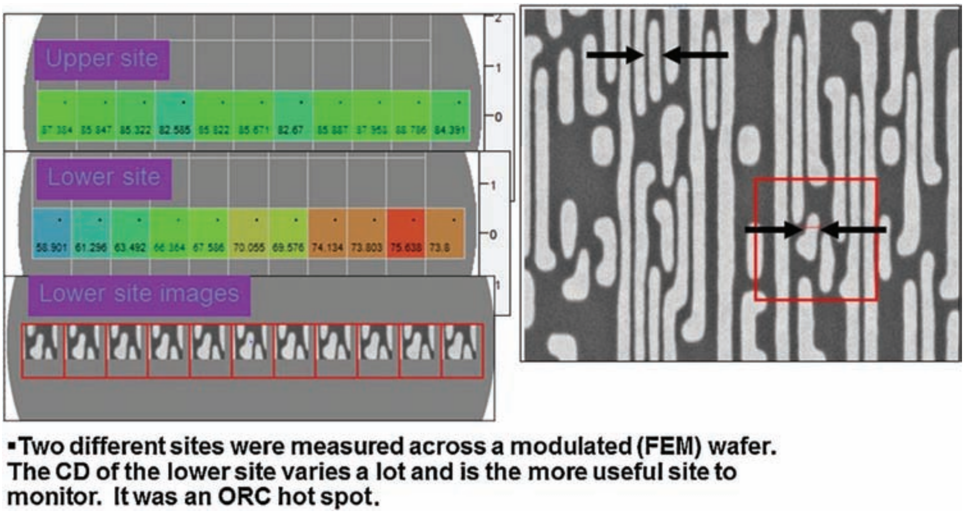


Figure 10.5 Two different structures measured, showing their different sensitivities.

shows an example that highlights two structures inspected on the EBI tool that have very different process sensitivities. The process was modulated across the wafer from left to right. Note that the upper structure does not change with these modulations, whereas the lower structure does. The lower structure is therefore a better structure to monitor this process over time on the CD-SEM. This synergistic relationship should be exploited because most fabs already have both CD-SEM and EBI tools. These fabs should also seek to minimize redundant process-window characterization work on the tools, which can be facilitated by clear communication between the process community and the metrology community.

Figure 10.6 shows the current e-beam offerings from each supplier as of 2013. The innermost circles that are shaded represent the current e-beam application landscape, which includes the e-beam critical dimension (ECD), e-beam inspect (EBI), and e-beam review (EBR). The circles that are not shaded represent each of the suppliers today; they intercept the shaded circles to show the application areas in which the given supplier has offerings. The rectangles represent the models available from each supplier. There is currently very little overlap between the ECD, EBR, and EBI toolsets, which is why there are so many different tool models. There are at least 12 platforms available today across these five suppliers [Applied Materials, Hitachi, Advantest, Nano Geometry Research (NGR), and Hermes Microvision, Inc. (HMI)]: three CD-SEMs, five EBI platforms, three EBR platforms, and one CD-SEM/EBI platform. NGR offers a platform being marketed as capable of performing an ECD and EBI simultaneously, but there appears to be limited knowledge about this tool in the United States. Note that the HMI eP3, although primarily an inspection tool, provides relative CD measurement capability.

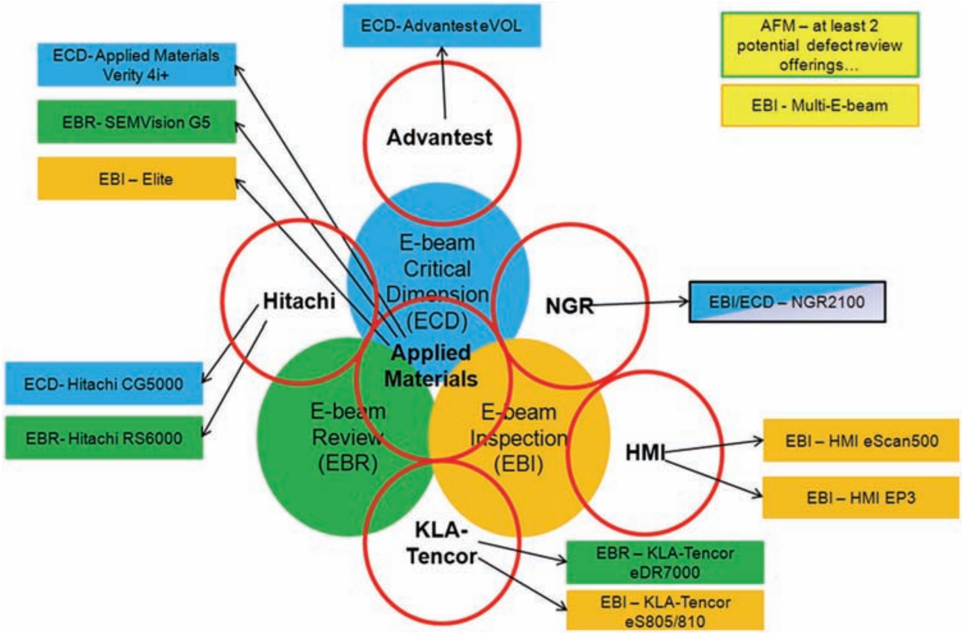


Figure 10.6 The e-beam metrology and defect landscape in 2013.

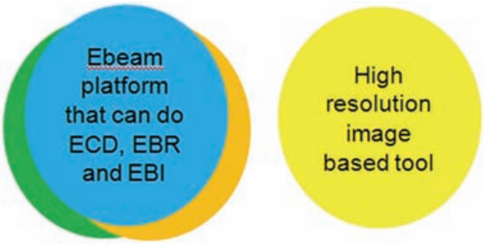


Figure 10.7 Future landscape recommendation for e-beam and dimensional metrology.

If this is not confusing enough, the multi-beam EBI tool is now part of supplier strategic roadmaps to address the speed issue that prevents the proliferation of more EBI applications; AFM manufacturers are also positioning new offerings in the area of defect review.

Focusing on the e-beam tools (because all of the information required to perform EBI and CD-SEM activities simultaneously is already present in the image), a better alternative to the previously described method is to create a single or cluster tool that can perform CD-SEM, EBR, and EBI activities well, as shown in Fig. 10.7. Some suppliers have begun exploring this area, such as NGR and HMI. It can be done successfully and significantly reduce the cycle time and cost of ownership. A single process step could be used for a CD-SEM, EBI, and EBR—in fact, the same image used for an EBI would be used for a CD-SEM. A larger fleet of the same tool would allow the toolset to be operated at a higher utilization. Maintenance requirements, such as spare part stocks, would also be simplified.

Suppliers should continue or start development on a multi-purpose e-beam platform that can perform all ECD, EBR, and EBI applications. Each fab can then decide how to best utilize this tool by either maintaining clear separation of the recipes between ECD, EBR, and EBI applications or strategically combining them to reduce the fab cycle time. Unfortunately, barring a dramatic improvement in CD-SEM capability, at least one other platform will be needed to address the CD-SEM resolution issues mentioned previously. This platform would only be used for the CD applications that require it. Approximately 75% of the CD-SEM measurement needs are less demanding and could be handled on this new multi-functional e-beam platform.

10.6 Line-Edge Roughness and 3D Metrology

The IC industry has long recognized that parameters other than traditional CDs (line/spacewidths, contact/via hole diameters) are needed. The SWA of resist, for example, is one parameter that directly affects the etch profile, and metrologists have tried multiple ways to measure it with a CD-SEM without much success. OCD solves the problem with database building and signal fitting to obtain multiple parameters, including the SWA. As the CD becomes smaller, the need to add more monitoring parameters grows.

Edge roughness has been in CD measurements throughout IC manufacturing history. All CD measurements, especially line and space measurements, are conducted with an average of multiple scan lines. However, the edge roughness portion is small enough and can be ignored when the CD is relatively large. Because edge roughness does not scale down alongside CD, the edge portion of roughness becomes large enough to ignore in advanced technology nodes. For example, the edge portion (assuming 3 nm per edge) is ~13% for a 45-nm line; it becomes 30% for a 20-nm line. Line-edge roughness (LER) and linewidth roughness (LWR) have been monitored together with line/space CD measurements.

10.7 Contour Metrology

As IC technology nodes progress below 20 nm, the designs become more and more complex due to aggressive model-based OPC before alternative design styles, such as gridded design rules (GDR), become mainstream. As lithographic process windows become smaller, it becomes much more desirable to monitor process shifts early with more sensitive process monitor 2D patterns than normal line/space CDs. Furthermore, with more adaptation

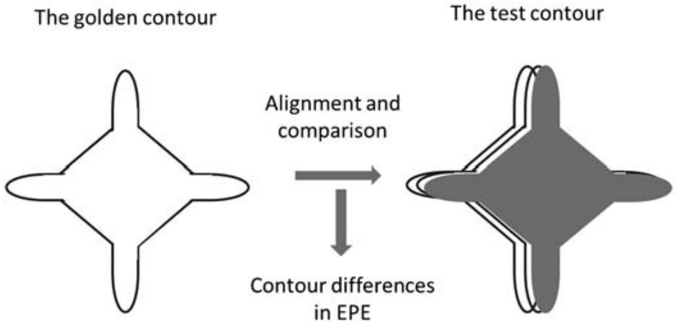


Figure 10.8 2D contour metrology.

of ILT style OPC, the traditional definition of CDs is less relevant. Thus, 2D contour overlay-based measurements will be used more often in lithographic and etch process controls, as well as in mask metrology.

2D-contour-based measurements require a reference 2D contour from a selected 2D pattern (sensitive to process variation) to be monitored. The reference can be obtained from the real printed 2D wafer pattern when the process is in good condition, or it can be simulated from a design/mask pattern with a lithographic model (i.e., OPC model). The reference serves as a golden image: the same 2D pattern will be imaged using a CD-SEM from production wafers as a test pattern. The contour from the test pattern will be extracted and aligned with the golden image contour for comparison and measurements. The contour-to-contour EPE can be computed and analyzed to report the largest and the range of all measured EPEs, as shown in Figure 10.8. Like CDs, the EPEs obtained from contour measurements can be monitored and controlled through SPC.

10.8 Lab-to-Fab and New Technology Trends

Further scaling down into below 14-nm technology nodes, new materials, new processes, new device structures, and continued dimensional shrinking will pose unprecedented challenges. In the recent past, conventional lab characterization techniques, such as XPS, XRD, interferometry, acoustic, LEXES, etc., have been incorporated as inline metrology tools to accommodate the upcoming insatiable demand to measure new parameters on new materials. As these lab tools become fab tools, certain challenges persist that hinder their widespread use (for some techniques more than others), e.g., lower throughput, lower reliability, and less automation relative to "workhorse" metrology tools such as a CD-SEM, ellipsometry, and OCD. Looking into the future, more emphasis on these aspects will allow a relatively smoother transition for these lab techniques to be incorporated as inline metrology tools and for process control.

References

1. A. Vaid and E. Solecky, "Strategic & Tactical Metrology Perspectives for Advanced IC Development & Manufacturing," *Metrology, Inspection, and Process Control for Microlithography XXIX*, San Jose, CA (2015).
2. C. Osorio et al., "Hybrid Metrology Universal Engine: Co-optimization," *Proc. SPIE* **9424**, 94241H (2014) [doi: 10.1117/12.2087233].
3. P. Timoney et al., "Comprehensive BEOL control using scatterometry and APC," *Proc. SPIE* **9424**, 94241G (2015) [doi: 10.1117/12.2087320].
4. Virtual metrology, http://bcam.berkeley.edu/research/dvm.htm.
5. R. Matsuoka, H. Mito, Y. Toyoda, and Z. Wang, "Development for 2D pattern quantification method on mask and wafer," *Proc. SPIE* **7638**, 76382L (2010) (doi: 10.1117/12.846720).

Appendix

All techniques using electron and ion beams change the sample to varying degrees. The effects manifest themselves as physical and/or electrical damage due to interactions with the sample. Damage can refer to several different undesired effects whereby the original target is changed in some way. Damage effects, ordered by increasing severity, are listed below, followed by corresponding illustrations:

1. Minor contamination: carbon or carbonaceous deposition.
2. Material shrinkage (also called slimming): energy transferred to the photoresist or high-*k* targets under SEM causes material changes to loss of mass or higher density, thus changing the CD.
3. Material swelling: the ion beam causes the material to expand locally.

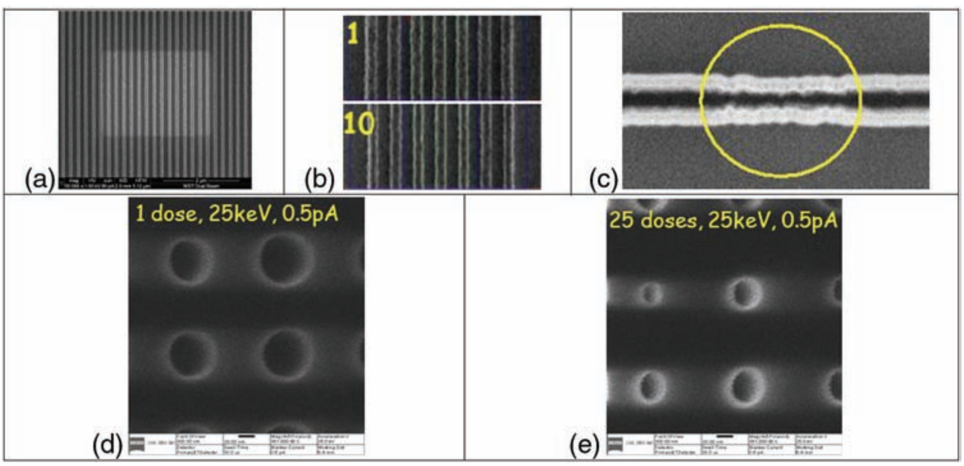


Figure A.1 Various SEM-induced damage phenomena: (a) contamination (the center box is a portion of a line/space grating of a-Si lines that was irradiated by an e-beam for 60 s; (b) shrinkage (the top image depicts immersion 193-nm photoresist lines after the first dose, and the bottom image depicts the same lines after the tenth dose); (c) etched Si line at the site of photoresist shrinkage due to SEM measurement of the precursor photoresist line (the bottom image illustrates material swelling of four 500-nm-deep HAR contact holes in oxide due to irradiation by HeIM); (d) HeIM image of four HAR holes after one imaging dose; and (e) HeIM image of the same four HAR holes in oxide after 25 imaging doses.

4. Electrical damage to devices: Vt (threshold voltage) shifts, which increases current leakage and traps dielectric energy.
5. Milling or knock-on damage: local target is physically destroyed.
6. Wafer structural damage or heavy metal contamination required, no further processing possible: dual-beam FIB, TEM sample preparation, etc.

SPIE.

Bo Su, Eric Solecky, and Alok Vaid,

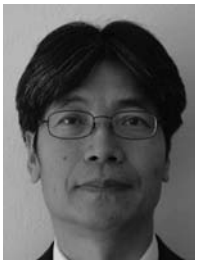
Introduction to Metrology Applications in IC Manufacturing
SPIE Press, Bellingham, Washington (2015).

The supplemental files for this eBook are available for download by copying and pasting the following link to a browser:

http:/spie.org/Samples/Pressbook_Supplemental/TT101_sup.zip

A zip file containing the supplementary materials will be available from the link.

Index



Bo Su has more than 20 years of semiconductor experience in multiple areas, different companies, and with different responsibilities. He has a B.S. and M.S. in Physics from Peking University and a Ph.D. in Solid State Physics from SUNY, Stony Brook. He started his career as a senior process development engineer in a R&D IC fab for National Semiconductor, responsible for metrology applications in lithography and etch processes. He later joined Applied Materials to oversee metrology equipment (CD-SEM) and its applications, new-generation CD-SEM definition and development, and product marketing. After that, Su joined KLA-Tencor to work on new equipment development and lithography simulation. He then worked at Anchor Semiconductor and Luminescent Technology, responsible for customer account management, new product development, and sales support. He is currently a senior director of product management at D2S, Inc., a private software company, developing unique, model-based mask-data-preparation (MB-MDP) solutions for leading-edge semiconductor companies and mask shops. Su has published more than 50 scientific and technical papers, and holds nine U.S. patents.

Eric Solecky is a senior engineer at IBM's microelectronics division in Fishkill, New York. He received his M.S. in Microelectronics from Rensselaer Polytechnic Institute, New York and his B.S. in Microelectronics from Rochester Institute of Technology, New York. He is currently on assignment at GLOBALFOUNDRIES to assist their metrology organization. Solecky started his career in lithography and then moved to metrology for the last 15 years. He currently serves as a committee member for SPIE and has also been a representative to research consortia such as SEMATECH. He was recently a guest editor of the metrology section of the *Journal of Micro/Nanolithography MEMS, and MOEMS*, focused on hybrid metrology. Solecky has published more than 30 papers in various conference proceedings and holds over 20 patents. He has given a keynote talk and presented an invited paper at SPIE conferences.

Alok Vaid is a Deputy Director/Senior Manager of the optical metrology group at GLOBALFOUNDRIES Fab8 in Malta, New York, where he oversees the company's strategic-metrology planning and roadmap. He received his M.S. in Mechanical Engineering from the University of Texas at Austin and his B.S. in Manufacturing Engineering from the University of Delhi. Prior to his current position, Vaid was a Senior Technical Staff Engineer at GLOBALFOUNDRIES (and previously AMD) for five years, responsible for developing metrology and control solutions for manufacturing advanced technical nodes. He has been working in the area of metrology research, development, and manufacturing for more than a decade and has published more than 50 scientific papers in various conference proceedings and journals. Vaid currently serves as a committee member at SPIE and IEEE/SEMI ASMC conferences, and has served as a representative to research consortia such as SEMATECH and SRC. He was recently a guest editor of the metrology section of the *Journal of Micro/Nanolithography MEMS, and MOEMS*, focused on hybrid metrology. He received the Diana Nyyssonen Memorial Award for the Best Paper in Metrology at the 2008 Advanced Lithography conference, and he has given a keynote talk at an SPIE conference.