Raising the Bar
Trends and Challenges in CMOS FEOL Technology for the 45nm Node and Beyond

18  A Balancing Act
23  (Feed)back to the Future
30  Fêting the Fin
48  The Winning Streak
64  Surface Watch
Maximize your yield to meet market demand.
RKET

market demand
Contents

COVER STORY

09 Raising the Bar
Trends and challenges in CMOS FEOL technology for the 45nm node and beyond

TECHNOLOGY TRENDS

18 A Balancing Act
Unveiling recent progress on strained silicon engineering

23 (Feed)back to the Future
Enabling transistor formation strategies with advanced process metrics

30 Fêting the Fin
Making MuGFET production a reality

36 A Sharper Image
Improving profile measurement accuracy via feed-forward spectroscopic ellipsometry

40 Keep it Flat
Reducing the effects of backside particle contamination

48 The Winning Streak
Advanced darkfield inspection for 65nm design rules and below

56 Stop that Leak
 Detecting crystal defects early in device fabrication using e-beam inspection

60 The Short Loop to Yield
Accelerating flash product inspections using electrical defect monitoring

64 Surface Watch
Generating high-speed, full-wafer maps of surface microroughness
PRODUCT NEWS

72  Viper 2435XP

73  SpectraCD-XT

DEPARTMENTS

06  Editorial
The Power of Innovation

16  CEO Interview
Priorities, Vision and Strategy

29  Awards
Puma 9000 Series

46  Spotlight on Lithography
Characterizing Overlay Errors

74  The Last Word
Industry Seeks Magic Lantern

WEB EXCLUSIVE

Towards Higher Performance
Controlling SiON gate thickness and composition using advanced metrics

Yield Management Solutions is published by KLA-Tencor Corporation. To receive Yield Management Solutions and web exclusive articles, subscribe online at: www.kla-tencor.com/ymsmagazine

For literature requests, visit:
www.kla-tencor.com/products

For information, visit:
www.kla-tencor.com

©2006 KLA-Tencor Corporation. All rights reserved. Material may not be reproduced without permission from KLA-Tencor Corporation. Products in this document are identified by trademarks of their respective companies or organizations.
The Power of Innovation

Judging from the technology news hitting the wires as I write this in mid-June, low power consumption was the overarching theme at the 2006 Symposium on VLSI Circuits. At the conference, leading companies such as Texas Instruments, IBM, Intel, IMEC, Freescale Semiconductor and others presented compelling, and often competing, visions on the future of the transistor. One headline in particular – Paths Diverge Toward Next-Gen Transistors – caught my attention.

Everything, of course, relates back to the stalling of CMOS performance largely because of the practical limits on power consumption in computers and mobile applications. As an IC Insights analyst puts it, “The power transistor market is set to drive the market to record levels in 2007.”

Given both the unprecedented interest in the future of the transistor and its role as the heart of consumer-driven applications we, at Yield Management Solutions, feel strongly about creating an issue with a special focus on transistor innovation. Our cover story, IMEC’s “Raising the Bar” leads the way towards creating metal gate electrodes with a fully silicided approach as a practical solution to continue transistor performance scaling.

New innovations are pushing CMOS transistors to their ultimate limits so they can effectively meet the incessant demand for higher density, performance, and power at lower costs. In “A Balancing Act,” Freescale Semiconductor discusses the recent progress made in terms of local and global strained silicon developments, notably enhancing carrier mobility, for boosting CMOS device performance without disrupting the delicate balance of power.

Freescale Semiconductor strongly believes that “extensive collaboration among IDMs, equipment and substrate suppliers, consortia, and universities is a critical factor in shortening cycle times, reducing development costs, and ensuring early entry into mainstream production.” This is an easy segue into important facets of semiconductor innovation: How do we know if it is economically feasible? Will the current process control technologies provide the required level of insight into the issues? Will new manufacturing challenges require innovative yield management technologies and strategies?

To date, laser-based darkfield inspection tools filled a key role in semiconductor inspection by providing high-throughput defect monitoring capability. As the industry moves beyond 65nm design rules and grapples with new challenges and continued cost pressures, conventional darkfield inspection technology struggles to meet manufacturers’ demands for cost effective inspection that provides the required sensitivity at production throughputs. “The Winning Streak” examines an innovative inspection technology that combines laser-based inspection with new darkfield imaging technology. Bolstered by years of technology innovation experience, the sky is the limit insofar as the range of applications that can be cost effectively addressed with this new inspector.

As clichéd as it sounds, knowledge is power. And we find knowledge in unexpected places. Previously, surface scattering was viewed mainly as a noise source for optical wafer inspection. More recently, wafer manufacturers and their customers used surface scattering measurements as a simple, single-value representation of surface quality, to accept or reject wafers. “Surface Watch” unveils a new product that leverages the system architecture of KLA-Tencor’s unpatterned wafer inspector to deliver surface-scattering data at unprecedented sensitivity. The measurement sensitivity of this system can be applied to detect changes in surface roughness for various surface types. This provides a wealth of information that is valuable for process development and may even be used for process monitoring.

Looking ahead, the prospect of massive transistor structure-related yield loss at the 45nm and 32nm node is driving the need for conjoint DFM and APC strategies. “(Feed)back to the Future” presents a compelling argument for linking design, layout, mask, and wafer processes with metrology. The article discusses in great detail how the increasing metrology needs of DFM and APC can be met by innovations in the measurement of pattern shape, profile, overlay, thickness, composition, and electrical properties.

Many articles in this issue present practical and innovative yield management approaches that have been successfully applied in production.

Clearly, if innovation is the key to our industry’s future, we must continue to work together (even if we don’t always agree) to enable radical new transistor technologies and structures that will ultimately revolutionize our world.

Vive l’innovation!

Uma Subramaniam
Editor-in-Chief
Join us at **Semicon West 2006** for the debut of
an exciting new forum from KLA-Tencor

The first annual Future Technology & Challenges forum explores emerging industry trends and challenges as viewed by industry luminaries from leading companies. A poster session runs concurrently with the oral presentations.

**Agenda - FTC Forum San Francisco**

- **Keynote: Nanoelectronics – From Vacuum Tubes to Nano Tubes**
  Dr. Pushkar Apte, Semiconductor Industry Association
- **Transistor Technology for 45nm and Beyond**
  Dr. Bich-Yen Nguyen, Freescale Semiconductor
- **Production Implications of Next Generation Lithography**
  Dr. Harry Levinson, Advanced Micro Devices
- **Break – Poster Session**
  Technical posters on display from 12:30 P.M. to 6:30 P.M. Authors are available for interactive discussions during the break and after the oral presentations.
- **DFM and Integration Challenges for the Fabless Manufacturer**
  Michael Campbell, QUALCOMM
- **Silicon Photonics – Opportunity, Applications & Recent Results**
  Dr. Mario Paniccia, Intel
- **“Be Amused” – Moore’s Law, Productivity, and PowerPoint**
  Don McMillan, Humorist, Technically Funny

**Located in Europe?**
For more information and registration for upcoming European Future Technology & Challenges forum events go to [www.kla-tencor.com/events](http://www.kla-tencor.com/events)
CMOS FEOL technology scaling, traditionally achieved by reducing gate length ($L_g$) and gate oxide thickness ($T_{ox}$), is down to a limit beyond which the gate leakage current becomes unacceptable. Strained silicon helps compensate for transistor performance loss from this reduced $L_g$ and $T_{ox}$ scaling. Beyond strain, high-k dielectrics, metal gates, millisecond anneals and new silicides are options for continued scaling. Each presents specific process challenges. This article reviews recent advances in fully silicided gates (FUSI) as an option for metal gate integration to continue transistor performance scaling down to 45nm and below.

Introduction

Traditionally, the industry has faced FEOL process challenges in three key areas:

- Narrowing of the gate dimensions by the use of advanced lithography and dry etch
- Thinning of the gate dielectric
- Control of junctions by ion implantation and short thermal anneal with contacts formed by self-aligned silicidation

The challenges alter dramatically as we move toward increasing transistor innovation at the 65nm node and below. The inversion layer thickness (proportional to $T_{ox}$) scaling trend from one technology generation to the next has changed when going from the 90nm node to the 65nm node, with a minor reduction of dielectric thickness compensated by a performance gain coming from strain engineering (figure 1). Whether it is achieved with recessed SiGe source/drain (S/D), strained liners or strain memorization techniques, each strain-boosting option presents integration and process control challenges. As we scale transistor dimensions beyond the 65nm node, we must not only maintain the gain obtained from strain boosters, but also adopt new ways to further increase device performance.
Enter Metal Gate Electrodes

There are many options to increase the MOS transistor drive current (figure 2). First, the mobility may be increased with the use of alternative substrates, recessed S/D or increased stresses of known strain-boosting techniques. The access resistance can also be reduced with new silicides that have reduced contact resistance, and by improved dopant activations obtained by co-implants or advanced anneals. Nevertheless, the prime candidate for achieving the required performance increase for the future generations is the transistor gate, where high-k dielectrics and metal gates offer the opportunity to resume the $T_{\text{inv}}$ reduction. During the last five years, many fundamental issues linked to high-k dielectrics have been solved. Using nitrided hafnium-silicates, the industry has found a viable candidate to reduce gate leakage, while maintaining the device performance with respect to SiON dielectrics. However, such a dielectric will likely require a metal gate electrode to avoid the compatibility issues of the high-k with doped poly-Si gates that leads, among other things, to unsuitable transistor threshold voltage ($V_t$). Hence the main challenge is to integrate the metal gate electrodes and fabricate devices such that the $V_t$ levels meet the specifications. The metal gate by itself, independent of the dielectric, is beneficial since it eliminates the so-called “poly-depletion” contribution to the $T_{\text{inv}}$.

The following sections present several metal gate integration approaches with their advantages and disadvantages. The FUSI device integration is then proposed as a practical solution for which process control challenges are briefly discussed.

Gate First or Gate Last?

In theory, the transistor $V_t$ is primarily determined by the work function (WF) of the gate electrode material present on top of the gate dielectric. Doped poly-Si has traditionally fulfilled this role very elegantly, by tuning the WF via the use of different species and concentrations to address all device $V_t$ requirements. Although very flexible, poly-Si has the drawback of an increased $T_{\text{inv}}$ by about 0.4nm due to the poly-depletion, which is becoming significant (~20% of the total $T_{\text{inv}}$) for state-of-the-art transistors (figure 3).
While the performance boost coming from the metal electrodes is well established, the challenge resides now in meeting the appropriate $V_t$ targets. The $W_f$ values for various metals are established and certain materials are identified with “band-edge” properties, i.e. with $W_f$ close to the valence or the conduction band of silicon. Those metals are susceptible to serve the needs for high, regular or even low $V_t$ transistors ($HVT$, $RVT$ and $LVT$, respectively).

The question now arises about how to integrate these materials as metal gates.

Two main options exist:

1. The “gate first approach” where the metal is deposited before the junction formation
2. The “gate last approach” where the metal insertion is done after

While the first option has the advantage of being the least disruptive from a process flow point of view, most of the metals will produce transistors with unacceptably high $V_t$, most likely due to material and interface modifications during the high thermal budget treatment (RTA) needed to activate the dopants. Hence, a gate first approach can only use thermally stable metals. On the other hand, a gate last approach requires, as detailed below, additional process steps that result in additional fabrication costs and yield issues.

**Deposited Metal or Full Silicidation?**

The gate last approach consists in first forming the device channel, junctions and contacts that require a high thermal budget, and then depositing the gate. Beyond Phase-engineered $V_t$ Although very elegant due to the relative simplicity of its implementation, the range of applications addressed by the phase engineering of FUSI is currently limited to the $HVT$ devices with HfSiON gate dielectric. Hence other options have been investigated to push further FUSI gates $W_f$ towards the Si valence and conduction bands. Dopants can be introduced by ion implantation but the $W_f$ variation obtained remains small for SiON dielectrics compared to the shift obtained on SiO$_2$ and non-existing on HfSiON. The Ni can also be alloyed with other metals or the gate poly-Si can be replaced by poly-SiGe. In this way, NiYbSi has been found to be very effective to tune the $W_f$ towards the conduction band on SiON and NiSiGe pushes the $W_f$ towards the valence band only in the case of HfSiON. Alternative metals have also been investigated such as Pt-rich silicide that exhibits pMOS compatible $W_f$ with $V_t$ values meeting $LVT$ requirements. Opportunities thus exist to modify FUSI gate $W_f$ beyond phase engineering to demonstrate FUSI even for $LVT$ applications both on HfSiON and SiON (figure A).
Here again, two options exist:

(1) The poly-Si is removed and the metal is deposited (referred to as the “replacement gate” approach).

(2) The metal is reacted with the poly-Si to form a silicide in contact with the gate dielectric (referred to as the FUSI approach).

The advantages of the latter include the relative simplicity to develop a self-aligned process through selective removal of un-reacted metal. For the replacement gate, however, additional fabrication steps are still needed to keep the metal only in the gate-region.

In recent years, FUSI has attracted considerable attention due to its relative practicality and, in the case of Ni-FUSI, its compatibility with existing processes. An example of such a gate last integration flow is given for the FUSI approach by the use of CMP and etch-back in figure 4.

**Ni-FUSI Phase Control is Important**

As Ni reacts with Si, it forms a Ni$_x$Si$_y$ silicide, $x$ and $y$ (i.e. the phase) being determined by the Ni-to-Si ratio and the reaction thermal budget. The possible phases are Ni$_2$Si, Ni$_3$Si, Ni$_3$Si$_2$, Ni$_2$Si, Ni$_3$Si$_1$, and Ni$_3$Si. It has been found that the phase of the silicide determines the Wf of the electrode formed with Ni-rich phases ($x>y$) closer to the Si valence band (figure 5) when on HfSiON dielectrics. The large Wf increase for Ni-rich phases on HfSiON has been attributed to the un-pinning of the Fermi level.

Hence, the Ni-to-Si ratio and the reaction temperature are the key parameters used to control transistor $V_t$. For the narrow gate, however, the Ni/Si ratio is not very well defined since the Ni present on the spacer can diffuse and the resulting phase will be Ni-rich while NiSi will form in wide-gate devices. This results in an unwanted $V_t$ variation from long to short channel transistors. The use of a two-step RTP process (RTP-1 + selective etch + RTP-2) helps reduce this gate length dependency by controlling the Ni supply by the first anneal step rather than by the deposited thickness (figure 6).

Once under control for all gate lengths, the phase is used as a tuning parameter to modulate the Wf and hence the $V_t$.

In figure 5, it is shown that on HfSiON...
gate dielectric, Ni-rich silicides have Wf values compatible with HV device requirements for pMOS while Ni monosilicide meets the same requirements for nMOS.

**Dual Wf FUSI Ring Oscillators**

A practical implementation of phase engineering to control the Wf selectively on pMOS and nMOS is illustrated in figure 7, presented by Lauwers et al. at IEDM 2005. From the Wf vs. phase observations, the appropriate phases can be obtained if higher Ni-to-Si ratio is achieved on pMOS compared to nMOS. Since the two-step RTP process is thermally limited, the Ni-to-Si ratio cannot be modulated by the Ni thickness. However, the poly-Si thickness available for the reaction can be selectively tuned by an etch-back in pMOS regions. The simultaneous silicidation of p- and n-MOS results in Ni-rich Si FUSI and NiSi FUSI, respectively. This implementation delivers a functional metal gate-based ring oscillator with $V_{t,sat}$ of 0.4V and 0.5V for p- and n-MOS, respectively.

**FUSI Process Control**

As discussed earlier, the control of the phase is critical to fabricate devices with the targeted $V_t$ values. First, the amount of Ni to react is determined by the RTP-1 temperature that impacts the phase formation (NiSi for nMOS). For a temperature that is too low, not enough Ni reacts, leading to a partial silicidation; for a temperature that is too high, extra Ni reacts and forms an unwanted Ni-rich phase. Figure 8 illustrates the determination of the temperature range for which NiSi forms in nMOS transistors (i.e. process window, PW) by electrical characterization. For temperatures below the lower bound of the PW, the capacitance equivalent thickness (CET) increases, indicating the presence of un-reacted poly-Si. For temperatures above the upper bound of the PW, narrow devices have a higher $V_t$.

Besides the RTP-1 temperature, the control of the poly-Si thickness and the spacer height at the time of silicidation are also important to achieve the desired phase in transistors. While it is obvious that the Ni-to-Si ratio is directly affected by the poly-Si thickness, the impact of spacer height is more subtle: In the case of a recessed spacer below the poly-Si top surface, the Ni deposited on the poly-Si sidewalls will also be available for reaction and an undesirable Ni-rich phase may be formed in nMOS devices.

In the CMP-based flow proposed in figure 4, the poly-Si available for...
silicidation is determined by the as-deposited thickness and the erosion occurring during the hard mask dry etch-back step. Similarly, the spacer recess depends on the spacer formation etch and the erosion during subsequent steps. Most of the erosion occurs during the gate hard mask dry etch back, due to the exposure of the spacers and the poly-Si once the oxide is removed, while an over-etch step must be applied to take into account process non-uniformities.

Figure 9 shows a wafer map of the oxide to be removed above the poly-Si gate after the CMP step. The area with the thinnest post-CMP oxide experiences more erosion, hence the available Si and recessed spacers shift the balance of the Ni-to-Si ratio towards Ni-rich phase formation. This is correlated with the nMOS device $V_t$, showing higher value (signature of the presence of Ni-rich phase) where the oxide thickness before etch-back was thinner.

The stringent requirements to control these parameters thus necessitate the development of new inline metrology techniques capable of measuring poly-Si thickness and spacer height in gate lengths down to 30nm, since those quantities are likely to be gate-length dependent.

The FUSI approach is arguably the most practical way to integrate metal gates in advanced CMOS.

**Conclusions**

Just as strain was introduced at the 90nm node and widely adopted at the 65nm technology node to continue transistor performance scaling, metal gates may fulfill this role for the 45nm node and beyond. Among the many options available to integrate them in transistors, the practicality of phase or dopant-engineered Ni-FUSI makes this the most mature approach, with functional ring oscillators demonstrated for HV$_t$ applications. The control of the NiSi phase (i.e. the Ni-to-Si ratio and thermal treatment) is found to be critical to achieve the targeted $V_t$ values, and requires particular attention in terms of process control and inline metrology. Beyond the phase engineering of the Wf, FUSI options exist to lower the transistors’ $V_t$ to levels suitable for LV$_t$ applications either with dopant, alloys of NiSi or alternative metals. Although many concerns remain, including yield and reliability issues, the FUSI approach is arguably the most practical way to integrate metal gates in advanced CMOS.
References

Author Biographies
Philippe Absil received his M.S. in physics engineering in 1997 from Université Libre de Bruxelles in Belgium, and his Ph.D. in electrical engineering in 2000 from University of Maryland in the field of micro photonics. In 2000, he joined Little Optics, involved in advanced micro photonics components fabrication. In 2003, he returned to Belgium and joined IMEC where he now manages the Bulk-Si CMOS FEOL integration research program.

E-mail: philippe.absil@imec.be

Thomas Hoffmann, IMEC vzw, Belgium; thomash@imec.be

Anne Lauwers, IMEC vzw, Belgium; anne.lauwers@imec.be

Anabela Veloso, IMEC vzw, Belgium; anabela.veloso@imec.be

Hongyu Yu, IMEC vzw, Belgium; hongyu.yu@imec.be

Malgorzata Jurczak, IMEC vzw, Belgium; malgorzata.jurczak@imec.be

Serge Biesemans, IMEC vzw, Belgium; serge.biesemans@imec.be

Jorge Kittl, Assignee from Texas Instruments at IMEC; jorge.kittl@imec.be

YMS Taiwan
August 23 – Ambassador Hotel, Hsinchu, Taiwan

YMS Singapore
August 25 – Raffles The Plaza Hotel, Raffles City Convention Center, Singapore

YMS Shanghai
August 29 – Ramada Plaza, Shanghai, China

YMS Beijing
September 1 – JinJiang Hotel, Beijing, China

YMS Japan
December 7 – New Otani, Makuhari, Japan

Attend KLA-Tencor’s Yield Management Seminar Series in Asia-Pacific & China 2006

For more details and registration for KLA-Tencor events, please visit www.kla-tencor.com/events
Rick Wallace, 46, was appointed CEO of KLA-Tencor in January 2006. Over the last 18 years, he has held a number of senior management positions at KLA-Tencor, including president and COO, and executive vice president, overseeing the company’s Reticle and Photomask Inspection Division, Films and Surface Technology Division, and Wafer Inspection Group. He has also served as CTO of the Software and Customer Groups, group vice president of the Wafer Inspection Group, as well as vice president/general manager and vice president of marketing for the Wafer Inspection Division.

Rick joined KLA-Tencor in 1988 as an applications engineer. Earlier, he built his expertise in lithography and yield management through engineering positions with Ultratech Stepper and Cypress Semiconductor. He has a BSEE from the University of Michigan and a master’s degree in engineering management from Santa Clara University.

We recently interviewed Rick on the future of KLA-Tencor and the industry. Here in Q&A format are his comments.

Q. What is the most important message you would like to communicate to KLA-Tencor’s customers?

A. We are now in the era of the consumer. Selling into the consumer market causes an increased focus on cost and time to market. I spend a lot of time in the field and with customers across the globe. What I hear about the most from customers is the increasing need to improve manufacturing efficiency, accelerate ramps, and speed the development effort to succeed in this consumer era.

Chipmakers want more agile platforms to enable better cost of ownership. To do this, we must engage with our customers earlier than ever before. The result is we understand our customer’s needs and know we’re solving the right problems. We want our products to be available ahead of customer requirements. Currently we have well over a dozen joint development projects with customers at the earliest stages of product development to ensure we will have the right products the market will need.

While our customers must continue to improve productivity to provide better, more cost effective products for their customers, we must be on top of optimizing and increasing productivity for chipmakers so they can maximize their profitability. We are doing this through our Global Support Services Group as well as our Process Control Solutions Group, which focus on providing applications expertise to our customers.

Q. As the new CEO of KLA-Tencor, what are your priorities?

A. I have spent a lot of time visiting customers, investors and our global employees.
My focus for the first six months has been on crystallizing a four-year global strategic plan to ensure that the company is prepared for the technical and market challenges ahead and at the same time redefining the corporate culture. The company’s strategic plan will enable KLA-Tencor to continue to grow and therefore allow us to continue developing leading edge products for our customers.

Q. What is your vision for the future of KLA-Tencor?
A. Our corporate vision is to “extend our leadership as the world’s best inspection and metrology company with differentiated technical solutions and customer expertise.” What this means to our customers is that they can count on us as an inspection and metrology company, and that we will continue to focus on creating best-of-breed products. KLA-Tencor excels at solving the really hard problems for our customers. We will continue to do just that. The problems are continuing to get harder and we are focused on developing the products needed to address these tough challenges.

We are committed to performance leadership and the customer experience. These two attributes are connected because you can’t be successful in just one area—you must do the whole job.

We have a renewed emphasis on quality, responsiveness, and supplying products and services that streamline operations and provide value in terms of increased yield and improved operational efficiency.

Q. Your R&D budget is one of the healthiest in the industry. What are your short term technology and product goals at KLA-Tencor?
A. From a process control perspective, the hurdles—at 65nm production, at 45nm development, and at 32nm R&D—all place additional emphasis on inspection and metrology because of the challenges developers face: new materials, new structures, new processes, and complex economic challenges.

Customers want more capable platforms and better cost of ownership. To do this, we must engage with them earlier than ever before. This is why we’re doing more advanced work at our customers’ facilities these days. This is one sure way to ensure we’re solving the right problems. You cannot ask customers what they need—it is not possible to survey innovation.

What we can do is determine what problems our customers anticipate they are going to have. Once we understand these problems, we can begin working on the solutions ahead of time. Even with a narrow focus, R&D is still expensive.

Another area of R&D focus for us is software development. Software is a major component of our product portfolio. For example, our current reticle inspection tool has ten million lines of code in it. Throughout the company, we have more than 100 million lines of active code, which is why 65% of our engineers are software engineers.

Q. What is your global strategy?
A. Today, innovation takes place around the world. A great deal of the work required is software and algorithm oriented, which can be done anywhere in the world. We are a global company, with product development in the U.S.A., Israel, China, and India. These positions support us in partnering with customers on a global scale. It puts us close to our customers (as well as our customers’ customers). The sun never sets on KLA-Tencor—product support and development are a 24/7 activity.

Q. What are you doing as a company to improve your manufacturing cycle time?
A. For a number of years we’ve been very focused on how to become more responsive to our customers’ needs. That has to do with cycle time. The average lead time, and customers’ requirements for when they want to turn on capability, has shortened over the years. We have to be in a position to be able to respond to that. We’ve done a number of things in manufacturing in terms of consolidating not just the manufacturing flow, but also moving toward common platforms and fewer suppliers, effectively leveraging those capabilities. Since 2004, we’ve been able to reduce our manufacturing cycle time on average by 25%, even though we’ve introduced new products with significantly greater degrees of complexity.

Q. KLA-Tencor announced plans to acquire ADE. Tell us why, and what do you see as the benefits of bringing ADE into KLA-Tencor?
A. ADE and KLA-Tencor present a good combination for customers. The acquisition will give both wafer manufacturers and IC manufacturers a broader product portfolio to choose from today and in the future. KLA-Tencor’s sales and service organization will bring ADE products to a wider global audience, thus giving our customers additional choices.
A Balancing Act
Recent Progress on Strained Silicon Channel Engineering for 65nm CMOS and Beyond


Much progress has been made by scientists in recent years to overcome the scaling limitation of classical planar CMOS transistors and maintain historical performance trends. New innovations are pushing CMOS transistors to their ultimate limits, beyond any previous predictions, to effectively meet the incessant demand for higher density, performance, and power at lower costs. This article discusses the recent progress made in terms of local and global strained silicon (Si) developments, and their successful insertion into today’s state-of-the-art CMOS technologies.

In the 2003 International Technology Roadmap for Semiconductors (ITRS) for CMOS 90nm and 65nm technology nodes, it was projected that neither new materials nor transistor structures were expected for 90nm and 65nm technologies. However, high-k gate dielectrics, dual metal gate electrodes, and elevated source/drain (S/D) structures were expected to extend planar CMOS beyond 65nm. Mobility enhancement – using either biaxial tensile-strained Si thin-film on relaxed SiGe virtual substrate for boosting CMOS performance or selective biaxial compressive-strained SiGe thin-film on Si substrate for boosting performance of p-channel transistor only – was also discussed. It was predicted that these biaxial strained Si or SiGe films would not be qualified for pre-production until mid-2006.

In reality, scaling planar silicon transistors beyond sub-50nm gate lengths (Lg) by increasing the channel and halo doping to suppress short channel effects and controlling off-stage leakage current (Ioff) has become extremely challenging, if not impossible, without some performance-power tradeoffs. Increasing substrate doping intensifies threshold voltage variation, junction leakage, capacitance, and degradation of carrier mobility. Thus, this approach is not an energy-efficient solution for portable electronic applications. In addition, portable electronics products demand microelectronic chips that are compatible with battery operation over longer and longer time intervals. One facet of the solution is carrier mobility (µ) enhancement. This can boost CMOS device performance (equation 1) without aggressively scaling Lg or gate oxide thickness (Tox) to meet the required performance at lower operating voltage (Vdd), while dramatically reducing the active and static power dissipations with low Vdd (equation 2).

\[ I_{d_{sat}} = \frac{W}{L} \cdot \mu \cdot C_{ox} \cdot (V_{dd} - V_t)^2 \]  

\[ P_{total} = \text{Active Power} \left( C \cdot V_{dd} \cdot f \right) + \text{Standby Power} \left( V_{dd} \cdot I_{off} \right) \]

Where \( I_{d_{sat}} \) is saturation drain/drive current, \( W \) is transistor width, \( L \) is transistor channel length, \( C_{ox} \) is inversion capacitance, \( V_t \) is threshold voltage, \( P_{total} \) is total power dissipation, \( C \) is total capacitance, and \( f \) is operating frequency.

Enhancing carrier mobility can be achieved by several techniques: Uniaxially strained Si using tensile or compressive stressors, biaxially strained Si on relaxed SiGe virtual substrates, or biaxially strained Si on insulator (SSOI). Innovation while maximizing re-use of existing materials, tools and device platforms has allowed the development of uniaxial stressors that boost both p-type (pMOS) and n-type (nMOS) channel devices in record time. Uniaxial stressors of embedded SiGe in the S/D regions of the bulk pMOS devices were inserted into the mainstream at 90nm in 2004 by Intel, and surely will be used as a p-type mobility booster for high-performance 65nm bulk or SOI circuits by many IDMs. Extensive collaboration among IDMs, equipment and substrate suppliers, consortia, and universities is a critical factor in shortening development cycle times, reducing development costs, and ensuring early entry into mainstream production.
The following sections present recent developments in local and global strained silicon developments, and discuss their effective insertion into cutting-edge CMOS technologies.

**Uniaxial Strain**

Due to the delay in identifying a pMOS performance enhancement solution, low threading dislocation density (TDD), pile-up defects (PU), and the lack of a cost-effective solution for the biaxial strained Si substrate, novel approaches have been identified and quickly inserted into mainstream products for boosting CMOS performance. These approaches are CMOS-compatible and utilize existing production tools with new processes or modified processes to provide the compressive or tensile stressor, which can boost carrier mobility. Significant stress (tensile or compressive) is imposed on the device in a preferred direction relative to the channel. The strain material distribution is typically localized to impact only one type (p- or n-) of transistor. This is achieved either by stressor incorporation in selected areas, or by locally altering the film characteristics (e.g., strain relaxation by patterned implantation) of an initially blanket stressor film.

Under strain conditions, semiconductor energy bands are shifted relative to each other, and band shapes are changed. When a state is reached with reduced inter-/intra-band scattering or with reduced effective masses, the carrier mobility is enhanced. The impact of strain on carrier mobility can be directly characterized with a piezo-resistance model by measuring mobility characteristics of conventionally built devices under an external mechanical strain. Intensive work has been accumulated in this area, with results characterizing the impact of multiple factors such as device type and channel orientation. Table 1 presents piezo-resistance results as a percentage of carrier mobility enhancement under 100MPa tensile/compressive strain for bulk p/nMOS with <110>/⟨100⟩ channel orientation. The data is directional for proper uniaxial stressor design. For example, for pMOS with ⟨110⟩ channel orientation, the best mobility enhancement is achieved by adding longitudinal (lateral channel direction) compression, while keeping the stress in the transverse or width direction under tension. However, for nMOS of the same orientation, tension in both channel and width direction is desirable.

<table>
<thead>
<tr>
<th>Device type</th>
<th>Channel orientation</th>
<th>Longitudinal stress (channel)</th>
<th>Transverse stress (width)</th>
<th>Vertical stress</th>
</tr>
</thead>
<tbody>
<tr>
<td>nMOS</td>
<td>⟨110⟩</td>
<td>3.1</td>
<td>1.8</td>
<td>(5.3)</td>
</tr>
<tr>
<td>pMOS</td>
<td>⟨110⟩</td>
<td>(7.2)</td>
<td>6.6</td>
<td>0.1</td>
</tr>
<tr>
<td>nMOS</td>
<td>⟨100⟩</td>
<td>10.2</td>
<td>(5.3)</td>
<td>(5.3)</td>
</tr>
<tr>
<td>pMOS</td>
<td>⟨100⟩</td>
<td>(0.7)</td>
<td>0.1</td>
<td>0.1</td>
</tr>
</tbody>
</table>

Table 1: Percentage mobility enhancement under 100MPa tensile/compressive stress.

Dual compressive and tensile Contact Etch Stop Layer (dESL) or Inter Layer Dielectric (ILD) as stressors have also proven to be viable solutions. Their rapid development times are due to the relative simplicity and reuse of existing manufacturing tools for further enhancing CMOS performance at sub-90nm technologies, especially when combined in a dual integration scheme. Combining the stress sensitivities of ⟨110⟩ channel orientation for pMOS devices with optimized transverse and lateral boundary placement can enhance the dESL performance gains in conjunction with the poly pitch effect. Figure 1 shows a high resolution TEM of the resulting dESL integration on a pMOS device with 70nm lateral boundary spacing. This particular TEM is from an integration in which the compressive film is formed first, and the tensile film second.

Figure 2 illustrates the 1.2V I<sub>off</sub>-I<sub>sat</sub> curves for pMOS and nMOS devices with a dESL integration combining high stress films with +400MPa tensile and -650MPa compressive lateral channel stresses.
stresses, respectively. More than 40% \( I_{\text{sat}} \) improvement was achieved for pMOS, but less than 10% \( I_{\text{sat}} \) gain for nMOS by implementing the dESL stress. As mentioned above, the pMOS device can be enhanced dramatically with a compressive stress in the lateral direction (parallel to current flow). The pMOS devices also prefer tensile stress in the transverse direction (perpendicular to current flow). In dESL integration, this can be achieved by placing the boundary between the compressive and tensile films close to the pFET in the transverse direction. Furthermore, all of these geometry effects need to be accurately modeled to maximize product performance gain.

Other uniaxial stressors can be formed by incorporating an epitaxial stressor layer in pre-recessed device S/D regions. The epitaxial material has a different lattice constant from the substrate. When the atoms of the grown film are well aligned with that of the substrate, and there are no (or negligible) misfit dislocations, the mismatch of the lattice constants of the substrate and refilling materials induces stress to the channel, resulting in mobility enhancement.

Epitaxial SiGe or silicon carbide (SiC) are the typical stressor materials in this case. SiGe has a larger lattice constant than that of the Si substrate. The S/D embedded SiGe (eSiGe) therefore induces the desirable lateral channel compression, while the S/D embedded SiC induces channel tension, which enhance hole or electron mobility and drive current of the transistor, respectively. Many reports on pMOS S/D eSiGe stressors have been disclosed in recent years for bulk and SOI technologies. Ghani et. al. reported that eSiGe for 90nm node bulk circuits were in production in late-2003 \(^3\). To couple eSiGe performance enhancement and SOI substrate benefits, Zhang et. al. reported an eSiGe stressor on a 65nm SOI platform (figure 3), with at least 20% gain in drive current and \( I_{\text{sat}} \), as shown in figure 4 \(^1\). Higher \( I_{\text{sat}} \) gain, up to 45%, can be accomplished by reducing the SiGe offset relative to the gate or by the increasing the Ge concentration.

Figure 4 shows that coupling of the eSiGe stressor with a compressive dESL stressor results in nearly linear enhancement combination, and more than 50% pMOS \( I_{\text{sat}} \) enhancement is demonstrated. The drawback of the uniaxial stressor is its strong dependence on geometry factors such as gate spacing, device width and density (figure 5). The study shows that incorporating eSiGe is important to maintain appreciable \( I_{\text{sat}} \) improvement at narrow device width, while the performance gain by dESL decays as device width dimension decreases.

Additional important benefits of the eSiGe stressor include its ability to retain hole mobility gains at high vertical fields, and to reduce device channel resistance (\( R_{\text{ch}} \)) and extension resistance (\( R_{\text{ex}} \)). This investigation shows that eSiGe not only...
enhances the mobility but also reduces both $R_{ch}$ and $R_{ex}$ while the compressive ILD stressor only reduces $R_{ch}$ (table 2). Processes and mechanisms that are similar to eSiGe, but that use tensile embedded SiC in the S/D region to boost n-type transistor performance, have received much recent attention due to the similarity to the existing eSiGe module, and have demonstrated potential for large performance gains up to 30% due to electron mobility enhancement.

**Biaxial Strain**

When a thin Si layer is grown pseudomorphically on a relaxed SiGe alloy buffer having larger lattice spacing than that of Si, the Si layer conforms to the SiGe template by expanding laterally and contracting vertically. This creates a biaxial stress, which enhances the transport properties of the Si layer due to altered band structure and electronic properties compared to unstrained Si. Stress reduces inter-valley and inter-band phonon scattering and effective hole mass due to band warping and preferential thermal population of electron states with light transport effective mass. Improvements to both electron and hole mobilities by applying biaxial tensile-strained Si as a transistor channel have been demonstrated. However, the fundamentally weak pMOS enhancement will pose scaling difficulties for global biaxial stressors in high performance CMOS. Although biaxial tension can produce modest hole mobility enhancement at low vertical effective fields, channel carrier sub-band splitting due to biaxial stress and its associated effective mass change leads to an undesired enhancement sensitivity to the vertical effective field (figure 6). The hole mobility enhancement under high effective gate fields is diminished and becomes negative when the fields are high. Piezo-resistance coefficients show that strong pMOS enhancement results when the undesired tension along the channel is reversed (table 1). Moreover, the transverse tension along $W$ should be preserved for pMOS performance. A novel in-plane stress engineering approach achieves the desired CMOS stress configuration as shown in figure 7, which would be more difficult to achieve by purely using uniaxial or biaxial stressors. The interactions and optimization between biaxial-uniaxial stresses-relaxation
and channel directions to obtain optimum performance gain for 65nm CMOS device has been proven (figures 8 and 9). Circuit simulation indicates that the logic circuits could achieve 1.2V circuit speed with 1V supply and 15% mobility for both n- and p-type transistors, and 40% dynamic power reduction at the same frequency (figure 10). Important progress has been made by SOI vendors in improving quality, availability and cost of SSOI substrates. Recent progress in TDD reduction and PU elimination is also promising, since these defects are potential yield killers, and also adversely impact device leakage and power dissipation.

**Conclusion**

Uniaxial stressors have been employed for boosting mostly pMOS performance. It will be more difficult to improve nMOS performance using the tensile stressor until cost-effective and manufacturable selective embedded SiC process and dual embedded S/D stressors integration are available. This implies changing the n:p ratio with technology scaling, which requires extensive library and circuit layout changes. Enhancing both nMOS and pMOS performance to retain the same n:p ratio is desirable. Interactions between biaxial lattice strain, uniaxial relaxation, process-induced stressor and channel orientation have been optimized to achieve the desired stress configurations for enhancing both short-channel SSOI nMOS and pMOS devices. Significant progress has already been made in meeting the performance, power and cost requirements for SSOI technology by joint collaboration between IDMs and substrate suppliers during the SSOI development and assessment phase.

**References**

As super-NA immersion lithography drives CDs smaller, process windows and yield entitlements are also expected to shrink. The prospect of massive transistor-structure-related yield losses at the 45nm node is driving the need for conjoint DFM and APC strategies. The success of these strategies will be critically dependent on feedback of accurate CD, overlay, and film metrology data. This article identifies innovative process metrics and trends, including simulation-based virtual metrology, that promise to be useful enablers for successful next-generation transistor formation strategies.

The primary goal of design for manufacturability (DFM) is to enlarge the process yield window, while the primary goal of multi-variate advanced process control (APC) is to keep the manufacturing process in that yield window (figure 1). This article discusses new technologies that will be needed for increasing yield as innovative transistor structures emerge at the 45nm node and below.

Enabling Transistor Innovation via DFM and APC
DFM requires feeding forward design intent, simulator output, layout clips, and design-rule-check (DRC) hot spots to expedite setup of measurement tools. Current DRC and aerial image modeling at best focus and exposure conditions are increasingly unreliable. In the future, process-window-aware approaches will require powerful full-chip simulators that can accurately predict and measure developed patterns in resist, along with accurate measurement feedback to calibrate the printability simulator. To control development costs, the conversion of data to information, knowledge, and decisions must be taken as far upstream as possible.

Implementing an APC strategy requires feeding forward both process context and measurement data. Looking ahead, we know that process context and measurement data must increase dramatically to support multi-variate control at the lot, wafer, field, die, and intra-die levels. Moreover, yield and performance losses are often caused by process integration issues or combinations of profile, shape, roughness, thickness, and pattern placement errors. Combined dispositioning and parametric yield analysis will require data from multiple metrology tools.

The case for linking design, layout, mask, and wafer processes with metrology is compelling. Greater complexity is offset by the advantage of greater access to adjustment. The increasing metrology needs of DFM and APC can be met by innovations in the measurement of pattern shape, profile, overlay, thickness, composition and electrical properties. As an example, examine the transistor drive current equation below:

\[ I_d = \frac{1}{2} \left( \frac{W}{L \cdot T} \right) \cdot (\varepsilon \cdot \mu) \cdot (V - V_t)^2 \]

Drive current at saturation depends on physical dimensions such as gate width W, gate length L, and gate oxide thickness T. It can limit the speed and, therefore, the average selling price of a device. Drive current also varies with electrical properties such as channel electron mobility \( \mu \), gate oxide dielectric constant \( \varepsilon \), and threshold voltage \( V_t \). These are in turn affected by such factors as strain, composition, and transistor architecture. Such performance-driven DFM and APC applications will require new measurement types, creating a need to decrease the cost and increase the yield-relevance of each measurement.

Conjoint APC and DFM Strategies Emerge
DFM and APC are likely to become a conjoint endeavor in the face of increasingly innovative transistor structures, and both will benefit from a wealth of new process metrics. More measurement types, more exotic technologies, and higher sampling will be required to support DFM and APC at the 32nm technology node.
Some examples are listed below:

- Overlay metrology using tiny, robust grating targets that can be inserted in the device in much the same way that CMP dummy structures are inserted now
- 3D multi-parameter profile scatterometry to measure critical dimensions in advanced planar and non-planar transistor architectures
- Virtual metrology utilizing calibrated litho simulators to assess the printability of complex RET structures, especially those designated for use with immersion lithography
- VUV spectroscopic ellipsometry to measure thickness in complex film stacks on patterned wafers

Changes in the process metrology landscape are accelerating, just as the wavelength reduction roadmap in lithography is decelerating. These changes are driven by the need to fill a widening design-to-process yield gap using DFM and APC strategies. These conjoint strategies, in turn, require new approaches to process measurement:

- Non-contact corona discharge technology to measure leakage and electrical properties of gate dielectrics and the low-k insulators used in advanced interconnects
- Common data analysis platforms with the ability to detect interactions between process errors measured on the same or different tools, especially CD and overlay at 32nm

Advanced Process Metrics for New Transistor Formation Strategies

By accelerating conjoint DFM and APC strategies, the following technologies promise to be useful enablers for transistor innovation at the 45nm node and below:

- Overlay metrology will augment scribe-only applications with in-die metrology, where small amounts of precision may be traded for more representative sampling, large reductions in model residuals, and improved overlay corrections that result in yield improvement.
- CD scatterometry will continue to exploit its advantage in profile metrology and increasingly powerful algorithms that will enable the characterization and control of advanced planar and 3D transistor architectures.
- Simulation will evolve into a calibrated platform for virtual metrology, enabling accurate prediction of downstream measurements due to upstream process variations such as focus or exposure excursions in lithography.
- Ellipsometry is evolving toward in-pattern VUV film thickness metrology and the use of sophisticated algorithms to extract more accurate film thickness measurements in the areas that affect yield most strongly.
- Electrical probe is filling a growing need for non-contact, in-line C-V measurements on new low-k and high-k dielectrics and is finding excursions in electrical characteristics that might otherwise go undetected.
- Data analysis is trending toward common platforms where multivariate techniques can find previously hidden interactions between electrical performance and physical measurements such as CD, overlay, and film thickness.
• **DFM applications** such as calibrated OPC/RET verification and design-based metrology (DBM) are developing rapidly. Accuracy is joining precision as a critical parameter for metrology capability, particularly if sub-0.5nm accuracy is required to calibrate OPC/RET models. However, local sample variations such as line-edge-roughness (LER), line-width-roughness (LWR), and film topography may be the ultimate limiters of measurement capability. The need to verify accuracy will continue to drive standards development, but we will also need to understand the “fundamental and practical limits of resolution, accuracy, and precision”.

• **In-die metrology** with CD SEM is already a required supplement to scribe metrology. Scatterometry (SCD), overlay, film thickness, and non-contact electrical metrology are likely to see more applications in this area. The dimensionality of CD measurement is increasing to the point where shape (SEM) and profile (SCD) are being discussed as targets for APC systems. Along with increased dimensionality, expect to see more use of multiple-measurement strategies and statistical metrology to improve yield-relevance, reduce cost per measurement, and increase overall measurement capability.

• **Optimal deployment** strategies are needed for SEM and SCD, as well as a clearer definition of applications for integrated metrology. As SCD takes a greater share of scribe-based APC applications, expect CD SEM to move gradually into the in-die 2D DBM, OPC, and DFM applications. Though SEM offers high-resolution imaging, it is occasionally limited by sample damage, charging effects, and measurement bias. The key advantage of SCD is that it can average over an array of features with high precision and accuracy, but SCD is occasionally limited by parametric model covariance, sensitivity gaps, and interference from underlying layers. Combining SEM, SCD, and other data offers the opportunity to improve overall metrology capability.

• **Fleet management** concepts may be extended to multiple metrology tool and data types. Both homogeneous (same tool type) and heterogeneous (e.g., SEM, SCD, AFM) calibration and matching are currently labor intensive. Recipe generation and data analysis are also labor intensive. In the future, data output will be automatically analyzed to select jobs that have the highest robustness to process variation, the most representative sample plans, or the best measurement capability. Tool recipes will reuse common elements and set themselves up without the need for a wafer. SCD could benefit from automated model analysis for parametric covariance, sensitivity gaps, and process robustness. Common analysis of data may also detect CD, overlay, and film interactions that might be yield limiting, especially in the case of dual-exposure, immersion lithography at the 32nm node.

The following sections describe six technologies (figure 2) with significant applications in conjoint APC and DFM strategies, along with key trends that make them applicable for enabling innovative transistor formation at the 32nm node.

**Overlay: From Scribe Line to In-Chip**

Traditional box-in-box (BiB) overlay metrology will evolve into more yield-relevant, grating-based overlay metrology (e.g., Archer AIM™). This will take measurement of pattern placement error to new levels of accuracy and enable combined CD and overlay dispositioning. At the 32nm node, BiB overlay metrology will suffer from extreme process sensitivity, particularly with respect to reticle fabrication error, asymmetric deposition and etching, and chemical mechanical planarization (CMP). Grating-based overlay technology (figure 3) can decrease process-induced measurement error by a factor of two. Remaining pattern placement error, including unmodeled intra-chip error, will be addressed.
with tiny in-chip grating targets. These enable more representative sampling and significant reduction of model residuals, arguably the largest remaining source of overlay metrology error. In some cases, such small overlay targets may be combined with line-end-shortening (LES) targets that are used to monitor focus and exposure excursions in lithography cells. The benefits are lower cost per yield-relevant measurement and higher temporal, spatial, and technology correlation for root-cause analysis. At the 32nm node, dual exposure-and-etch strategies may result in direct coupling of CD and overlay error, as in the following equation:

$$CDE = \frac{1}{2} CDE_1 + \frac{1}{2} CDE_2 \pm OLE$$

(2)

The combined CD error (CDE) results from the sum of edge placement errors in the first and second patterning steps, plus an additional contribution from intra-layer misregistration (OLE).

**SCD: From CD to Profile Metrology**

Scatterometry-based CD metrology (e.g., SpectraCD™) will evolve into more yield-relevant “profile metrology” and may become a reference tool for calibrating CD SEMs down to 13 nm or lower since SCD can accurately reproduce cross-section profiles imaged in a transmission electron microscope. The ability of SCD based on spectroscopic ellipsometry (SE) to accurately measure footing and notching at the base of gate structures has led to two-fold improvements in correlation to electrical L-poly and drive current. For this reason, SCD tools are currently displacing other metrology tools in feed-forward APC applications from lithography to etch. In control applications for shallow-trench isolation (STI), significant cost savings have been realized by metrology convergence. SCD tools are displacing CD SEM, AFM profile, and SE film thickness tools for the control and monitoring of isolation. The benefits are lower cost, shorter cycle-time and greatly reduced temporal, spatial, and technology de-correlation for the more yield-relevant, compound measurements such as aspect ratio. Currently, 3D SCD technology is being applied to measure the profiles and shapes of contact holes and other simple structures. At the 32nm node, the multiple simultaneous measurements provided by 3D SCD may be required for both DFM and control applications.
**Simulation: From Actual to Virtual Metrology**

Process modeling and simulation will evolve into yield-predictive “virtual metrology”. Even now, the measurement technologies discussed (SEM, SCD, and AIM) rely to some extent on simulation. Simulated SEM images assist with design-based pattern shape metrology. Rigorous coupled wave (RCW) algorithms generate libraries of ellipsometric spectra for comparison with actual SCD data. Overlay simulators predict the optical signatures of innovative overlay targets in order to maximize sensitivity and minimize response to process noise. Finally, robust printability of SEM, SCD, and AIM measurement targets is critical; so calibrated lithography models (e.g., PROLITH™), will be employed to assist in the initial target optimization. These models must use realistic mask data and comprehend the most aggressive resolution enhancement technologies, including immersion polarization and phase shift strategies (figure 4, two left images). Second, they must provide accurate, calibrated results for 193nm immersion lithography and enable rigorous virtual metrology through the focus-exposure window to supplement actual physical measurement (figure 4, two right images). The benefits are lower cost per measurement, in-line validation of physical metrology, and upstream pattern analysis to reduce design, mask, and wafer-level yield loss.

**Ellipsometry: From Film Stack to Pattern**

Spectroscopic ellipsometry (SE) has been the long-standing best-known-method for measuring the thickness of transparent films used in the semiconductor industry (e.g., SpectraFx™). Expect the wavelength range of these systems to be extended to the VUV, increasing sensitivity to ultra-thin films, high-k gates, high-k memory stacks, and 193nm anti-reflection coatings. Furthermore, in-pattern SE capability will be developed because of the large offsets observed between measurements made on a pad and actual film thicknesses in the die. The extraction of in-pattern thickness will require the use of unique and sophisticated algorithms. Such techniques may be applied to dielectric layers on patterned metal in 1D and 2D arrays (figure 5) and to embedded dielectric between...
time may make little sense until it is broken down by generalized ANOVA into systematic and random components at the cell, lot, wafer, field, and die levels. Another need is the correlation and calibration of physical measurements from multiple sources, such as SEM, SCD, AFM, and TEM. In the case of CD control in lithography, focus-exposure matrices from SEM, SCD, and overlay tools may be fitted to a CD response surface, enabling APC strategies that use feedback of focus and dose corrections. The 32nm node will get much more traction from analysis of correctable interactions between CD, overlay, and films. An example is shown in figure 6, where at least seven overlay, shape, profile, roughness, and topographic errors are interacting to increase the probability of gate-to-contact shorting in a transistor structure.

Conclusions
Virtually all of the technologies discussed above address the problem of hidden process error that could limit yield at the 32nm node. Many systematic variations of concern at 32nm are not observable with in-line metrology tools designed to control current-generation processes. Yield losses due to variation of in-die overlay, pattern profiles, pattern shape, film stoichiometry, in-pattern thickness and electrical properties can only be reduced if they are monitored. Clearly, the extent to which variations and their interactions can be analyzed, simulated, and corrected will determine yield and performance entitlements at 32nm and beyond.

Acknowledgements

The authors wish to acknowledge the contributions and comments of Amir Lev, Russ Huffman, John Robinson, Dan Wack, Mike Adel, David Tien, Amir Azordegan, WayneMcMillan, Andy Pindar, Adrian Wilson, Jianou Shi, and Chris Sallee.

References
Editors’ Choice Best Product Award

Presented to KLA-Tencor

KLA-Tencor’s Puma 9000 Series has received the 2006 Editors’ Choice Best Product Award for Semiconductor Manufacturing Excellence. The prestigious award is presented annually by Semiconductor International magazine to the industry's most innovative companies – specifically those that are truly making a difference in semiconductor manufacturing via excellence in design, engineering and production capabilities.

“Advances in semiconductor technology are only possible because of the kinds of products being honored in this year’s Editors’ Choice Best Product Awards program,” said Pete Singer, editor-in-chief of Semiconductor International. “Chipmakers rely on these products to create electronics that are smarter, smaller, faster, less expensive and more reliable. We congratulate the people and the companies that have had the insight and fortitude to bring these products to market.”

Advancing the state-of-the-art in wafer processing

The Puma 9000 Series is a family of UV laser-based, darkfield inspection systems for patterned wafers. The tools are used for cost- and performance-optimized defect monitoring on a broad range of layers at 65nm design rules and below.

The Puma 9000 Series is based on the revolutionary Streak™ technology, which combines multi-pixel sensor and line scanning technologies to produce high resolution darkfield imaging. The combination of the Puma 9000 platform’s unsurpassed noise suppression with powerful Streak technology produces a darkfield inspection tool with the highest sensitivity at throughput for a broad range of layers and defect types. The Puma 9000 platform’s configurable architecture optimizes cost of ownership for applications from tool monitoring to advanced etch. The Puma 9000 series shares a common user interface with other KLA-Tencor inspectors and review tools, to allow recipe exchange between platforms and speed yield learning.

Puma 9000 Series Benefits

• Streak technology with UV illumination provides the highest darkfield sensitivity at production throughputs on the broadest range of applications
• Low-angle illumination, selectable polarizations and programmable filters provide superior noise suppression capabilities for improved defect detection
• Common user interface with KLA-Tencor's 23xx and eS3x enables recipe transfer, lowered training costs and accelerated integration into production
• Flexible configurations provide application-specific solutions at the lowest possible cost of ownership
• Extendible technology protects customers’ capital investment
This article presents solutions to address the needs of Multi-Gate Field-Effect Transistor (MuGFET) metrology in a production-worthy fashion. A procedure to calibrate CD SEM to transmission electron microscopy (TEM) for accuracy is developed. CD SEM is used to automatically perform line width roughness (LWR) metrology of fins, while the three-dimensional (3D) information is obtained by means of scatterometry. Finally, the article discusses the application of design-based metrology (DBM) to MuGFET optical proximity correction (OPC) validation.

As we move toward the 45nm and 32nm node, MuGFETs are increasingly being considered as a necessary alternative to keep pace with Moore’s Law. If proven in production, MuGFETs could eventually replace conventional CMOS transistors. Given that the ability to perform proper and extensive metrology in a production environment is critical, this article investigates some key requirements of MuGFET metrology. Accuracy and LWR metrology will play an essential role, because of the small dimension of the features involved. 3D metrology is required when dealing with non-planar devices. Sophisticated OPC checks are needed in order to ensure that the design intent is respected.

The characterization of MuGFETs or other two-dimensional (2D) devices is a basic requirement in order to be able to adopt these innovative architectures. A robust metrology approach is essential to characterize these structures. Accuracy, line width and sidewall roughness, 3D characterization, and patterning optimization are some of the issues that need to be solved in order to transfer this technology from development to production.

Because the fins of a MuGFET device (figure 1) go down to 10nm geometries, the metrology tools have to guarantee accuracy in addition to the classical precision requirement. A 5nm accuracy error would correspond to a 30% change in critical dimension (CD) when dealing with a 15nm feature, which is unacceptable. In the current development phase, the accuracy requirement is often satisfied by expensive characterization techniques, such as TEM analysis. This approach is obviously not sustainable in a production environment.

LWR and sidewall roughness have a direct impact on device performance, calling for a robust metrology to characterize these elements in both development and production. The requirement of 3D characterization of these devices is not common to any planar device metrology, and it is complicated by the small dimensions both in terms of CD and height of MuGFETs. Finally, the accurate patterning of these small features requires a careful definition of the whole litho process.

This article proposes various solutions for some of the open issues related to MuGFET metrology. Accuracy standards ranging from 10-70nm are developed to calibrate CD SEM tools. Scatterometry is used to characterize the 3D structure of fins as small as 10nm. DBM and online LWR characterization demonstrably helps to optimize the litho process and to quantify roughness in various process steps, respectively.
The results indicate a clear need to deliver a comprehensive metrology solution for MuGFETs, which will enable reliable production of these advanced devices.

**Experimental Set-up**

All exposures are performed on an ASML PAS5500/1100™ step-and-scan system, interfaced with a TEL Clean Track Act8™. Maximum numerical aperture (NA) is 0.75. The total system is charcoal filtered to prevent airborne base contamination. Top-down CD SEM inspection is done using KLA-Tencor technology. For the baseline technology integration work at front-end of line (FEOL), a 193nm resist from JSR, AR237J at 230nm film thickness (FT), is used on Brewer Science ARC29a organic Bottom Anti-Reflective Coating (BARC), FT = 77nm. The stack for MuGFET patterning (active layer) is 65nm silicon on 150nm buried oxide (silicon-on-insulator, or SOI, stack). A 60nm TEOS oxide hard mask (HM) is used during the patterning process for two reasons: to provide etch resistance for the silicon etching, and to enable CD (HM) trimming. A binary mask (BIM) is used to print an active pitch of 350nm; the CD at mask level is 120nm. The litho target is set at 100nm. This target is chosen to have acceptable process latitudes (CD control) in lithography. Two exposure conditions are studied in more detail: a 0.63NA conventional 0.89 and a 0.75NA annular 0.89 outer σ and 0.65 inner σ. The scatterometry measurements are done on a KLA-Tencor SpectraFx 100™ using a polarized ellipsometer. The scatterometry target is 50x50mm².

**Accuracy is Key**

In principle, although the typical CD SEM resolution (2nm) does permit metrology on small features such as gate length or fin width in MuGFETs, it is essential to develop the methodology to guarantee proper accuracy. Historically, the main deliverable of CD SEM tools has always been precision. This is understandable when dealing with features that are 50nm or larger. In the case of a 15nm feature, however, even a 5nm error is unacceptable in the production environment.

Accuracy standards ranging from 10-70nm are developed to achieve this goal. The calibrated CD SEM is demonstrated to be sensitive to CD changes caused by process variations down to 10nm, and reference analysis performed on site previously measured by CD SEM confirmed the quality of the calibration.

Four different CD standards (70, 45, 25 and 13nm) are developed by depositing alternating layers of silicon and silicon oxide (figure 2, left). The wafer is then diced and rotated, and the oxide is etched (right). The uniformity of the CD is mainly dictated by the deposition uniformity, which can be carefully controlled. This procedure delivers standards with very low roughness, which helps create features with an extremely uniform and well-controlled CD over various millimeters.

The sample is then certified NIST traceable by using TEM analysis. The CD of the line is measured by comparing it to the lattice constant of the crystalline silicon of the wafer (figure 3).

These high quality standards help to optimize the measurement algorithm for accuracy, by mapping the total measurement uncertainty (TMU), as well as accuracy slope and intercept, as a function of the algorithm parameters. This procedure identifies a single set of parameters that guarantee the best CD SEM accuracy in the range of interest. The measured precision after accuracy calibration is observed to be less than 1nm, which can be further improved.

Figure 4 reports the CD maps of fins on TEOS wafers after the corner rounding step. The fins are observed to be smaller in the middle of the wafer (~10nm) as compared to the edge of the wafer (~30nm). The radial pattern is confirmed on other wafers. These results indicate the sensitivity of the accuracy setting to process variation of features as small as 10nm.

After performing a set of CD SEM measurements in well-defined locations, the devices are measured with various reference techniques. Figure 5 (left) compares the TEM on MuGFET devices before and after hydrogen annealing. The outer fins (first from the left) are clearly larger than the inner fins, in agreement with the CD SEM results. Similarly, the TEM analysis confirms that fins on the edge of the wafer are larger than those in the center, and that hydrogen annealing does reduce the fin’s CD in the case of TEOS wafers. Finally, figure 5 (right) compares the CD measured by the calibrated...
CD SEM to various reference techniques, including TEM, XSEM and scatterometry, as well as to the accuracy standards. The results clearly indicate the ability of the CD SEM to accurately measure in the range from 10-80nm.

**Patterning**

In MuGFETs, as for most structures, CD variations through pitch and as a function of length make the devices unreliable. The magnitude of corner rounding also directly affects performance, because increasing fin width and decreasing length impacts short channel effects. Different methods are used to reduce some of these effects, such as the addition of serifs or off-axis illumination settings.

A solid characterization of these approaches is needed, requiring a very large number of CD SEM measurements on different sites (extremely time consuming to set up). DBM can overcome this issue. This approach, originally introduced to improve design manufacturability, creates CD SEM recipes having hundreds of sites, starting from the design files. The recipes are created offline in a few hours, whereas it would take days of tool time to create them manually.

In addition, it is critical to guarantee data integrity by adopting a methodology for reliable measurements based on proper 2D algorithms for the various metrology needs. This study makes use of minimum/maximum gap and corner rounding algorithms. Gap algorithms are preferred for fin width and length measurements to standard line-width algorithms, which do not account for the rounding of the structures. The corner rounding algorithm determines the magnitude of the rounding (top-down) of the corners in a device. All algorithms can measure multiple structures within an image.

One of the main concerns with decreasing the size of the fins is the magnitude of corner rounding, which has an impact on both fin length and width. The rounding of the fin is characterized by the difference in area between the edge of the MuGFET and its bounding box. Corner rounding is reduced by off-axis illuminations as well as by serifs. For comparisons of the annular and standard illuminations with many different

---

**Figure 4:** CD maps of a MuGFET wafer after corner rounding. Fins in the center of the wafer are smaller. The map shows the sensitivity of the calibrated CD SEM to process variations down to 10nm.

**Figure 5:** TEM of MuGFET devices (left) before (a) and after (b) hydrogen annealing on TEOS wafers. CD SEM measurements (right) versus reference measurements, indicating the ability of the calibrated CD SEM to measure accurately.
OPC corrections, as well as comprehensive quantitative results of corner rounding analyses, please refer to the original paper “Comprehensive Approach to MuGFET Metrology” presented by the authors at SPIE 2006 Metrology, Inspection, and Process Control for Microlithography XX, Proc. SPIE 6152, 615219 (2006).

Results indicate that scatterometry can indeed be the tool of choice to extract 3D information for MuGFETs in production.

**Line Width Roughness**

Another specific issue of MuGFET metrology is LWR. In terms of top-down line edge roughness, various frequencies will impact device performance differently. Figure 6 reports an example of roughness on a fin. Low-frequency LWR will mainly impact CDU, while high-frequency components will impact device performance. It is then essential to fully characterize the spectral components and monitor LWR. This can be done online using recently developed roughness algorithms available on CD SEM, which can monitor LWR, correlation length and power spectrum.

The available CD SEM algorithms are tested to demonstrate sensitivity to MuGFET LWR. The results characterizing LWR for the hydrogen annealing process step on SiON and TEOS substrates are shown in figure 7. The LWR distributions clearly demonstrate that the annealing does not improve the roughness characteristics for SiON substrates (10%), while the improvement is significant on TEOS substrates (40%).

**3D Characterization**

For proper process control on MuGFET devices, it is crucial to be able to obtain 3D information on the device morphology. This requirement is quite complex when dealing with small features and large pitch, caused by resist trimming. The capability of scatterometry to measure these devices is investigated here.

The targets measured are not MuGFET devices, but gratings with the same design rules on a flat surface of oxide. The CD SEM measurements of the two types of targets used here are about 20nm and 40nm, respectively.
Patterning Fêting the Fin

Figure 8 reports the scatterometry signals as a function of wavelength, as measured on targets having line pitch ratio 100/100 and 150/150 after lithography. For comparison, the spectrum of bare field oxide is shown, as measured next to the scatterometry targets. The CD SEM analysis of the 150/150 target shows fin dimensions of 20-25nm with 300nm pitch (1:15 ratio). The difference in the signals indicates the presence of a design on the field oxide.

A scatterometry model is built to fit the spectra obtained. The model uses two simple trapezoids, simulating the fin and the recess in oxide, respectively. The thickness and the profile of the recess are fixed parameters. Figure 9 shows the comparison of the reconstructed profile obtained by scatterometry in good agreement with the SEM cross section. The model shown is not optimized, but demonstrates the feasibility of such measurements. In addition, preliminary tests indicate good repeatability, suggesting the possibility that the current limit for scatterometry (1:8) could be extended. These results are encouraging, and indicate that scatterometry can indeed be the tool of choice for industrial monitoring of 3D characterization for MuGFETs.

Conclusions

This article proposes several approaches aimed to build a robust production metrology for MuGFET devices. A methodology to accurately calibrate a CD SEM is demonstrated, along with the need to use DBM to fine tune the patterning of the devices, and the necessity of implementing online LWR analysis. Preliminary results indicate the potential of scatterometry to extract 3D information. The results indicate the feasibility of MuGFET metrology, although they do not satisfy all the metrology requirements for these devices.

Acknowledgements


References

Reticle Quality Assurance

Progressive reticle defects such as crystal growth and haze are an industry-wide reliability problem. Even if reticles are determined to be clean upon arrival from the photomask supplier, some of these reticles can show catastrophic defect growth over the course of production usage in the fab. The categories of defects that cause reticle-quality degradation over time are defined as progressive defects, commonly known as crystal growth, haze, fungus or precipitate. This progressive defect problem has been around for more than a decade and was observed at almost every lithographic wavelength.

![Image](image_url)  

Inspection with the STARlight™ system from KLA-Tencor clearly flags progressive defect growth on chrome and half-tone photomasks

This problem is especially severe at 193nm lithography. Triggering the increased severity are shorter wavelength lithography - where the photons are highly energized - and the concurrent transition to 300mm wafers, which require photomasks to endure more prolonged exposure as compared to 200mm wafers. Both embedded phase shift masks (EPSMs) and chrome-on-glass masks are affected by progressive defects.

On average, about 1 percent of binary masks (at 365nm lithography) and 6 to 15 percent of EPSMs (using DUV lithography) show a defect growth problem through the duration of their use in a typical fab.

These defects are generally found on the patterned surface underneath the pellicle (on clear, half-tone or chrome patterns), as well as on the backside surface of the photomasks. Past cases have indicated that this problem mainly starts on the clear areas of the reticle on the pattern side (this includes clear areas inside dies and scribes) with emerging semi-transmissive contamination that then propagates into the critical areas of the die while growing in both size and opaqueness. But current studies show an additional trend, where severe defect growth on half-tone and on chrome, as well as on edges of geometries (defect growth on chrome border) is also observed.

The rate at which these new generations of defects grow on half-tone and on chrome is also much higher than what was observed in the past for clear areas. At the same time, resolution requirements have driven the industry to implement very low-k1 lithography processes, which further worsen the impact of reticle errors. Hence this new trend of fast-forming “on half-tone” and “on chrome” progressive defects is of big concern to the industry, since an unnoticed defect formation on half-tone or chrome surface that may not impact the process window today may be critical in the very near future. An early warning on a defect growth problem will now need the complete knowledge of all the material surfaces of the reticle on the pattern side: on clear, on half-tone and on chrome.

To read an expanded version of this article, go to:  
www.kla-tencor.com/litho  
Your Patterning Process Control Resource
A Sharper Image
Feed-Forward Spectroscopic Ellipsometry Improves Profile Measurement Accuracy

Robert M. Peters, Suresh Lakkapragada – KLA-Tencor Corporation

This article presents a method for improving correlation of advanced gate lithography optical profile measurements to those of a critical dimension atomic force microscope (CD-AFM). By coupling a spectroscopic ellipsometry (SE) film measurement to a subsequent grating measurement, accurate values for underlying films are fed-forward to the grating measurement. This allows for reduction in the degrees-of-freedom during the grating measurement, which reduces parameter cross-correlation.

SE-based optical metrology methods have now gained a strong foothold for measuring the two-dimensional profiles of integrated device features. Optical metrology generally provides superior performance compared to other methods, but there still remain some challenges in terms of precision and accuracy requirements as device geometries continue to shrink at an aggressive rate.

Process engineers use several new techniques to meet device patterning requirements while maintaining manufacturing-worthy process windows. These include thinning the photoresist layers and adding under-layer films to act as hard masks for subsequent pattern transfer steps. Thinning the photoresist layer reduces the aspect ratio of patterned grating targets, which in turn reduces the signal-to-noise ratio (SNR) of the optical profile measurement. The additional films in the process stack below the gratings increase the number of optical interfaces that must be taken into account when building the optical model for the measurement. The increased complexity of the optical model increases the likelihood of cross-correlation between the underlying films and the grating profile parameters, such as CD, height and sidewall angle (SWA). Moreover, the lower SNR and higher cross-correlation negatively impact the precision and overall accuracy of the reported values for these grating profile parameters.

This paper discusses a methodology to overcome these issues. It involves performing a standard SE film thickness measurement on an open pad area in close proximity to the grating target of interest. The thickness values are then fed-forward to a subsequent SE measurement of the grating target. With the under-layer thickness values fixed based on the film thickness measurement, only the grating profile parameters are solved for during the grating measurement. Decoupling the under-layer film measurement from the grating measurement greatly reduces (or even eliminates) cross-correlation between parameters. Both SE measurements are completed within a total move-acquire-measurement (MAM) time of <10 seconds per pair, and the resulting values reported for CD, height and SWA are more accurate compared to reference metrology such as AFM.

Supporting data is presented from measurements taken on a 65nm technology node gate lithography process. Using the feed-forward process, the correlation and slope of profile parameters measured via SE compared to AFM measurements is greatly improved. Furthermore, systematic anti-correlation between resist height and SWA, observed during simultaneous measurement of the film stack and grating, is eliminated when the film measurement is decoupled and fed-forward into the grating measurement.

SE-based Measurement Technique
The SpectraCD™ measurement based on SE technology is described in detail in previous publications. For a SpectraCD measurement, a grating target is placed in the path of the SE beam. The grating comprises line/space features of uniform period, with the line width (CD) and period designed to represent the physical device feature under control.

SpectraCD measurements are completed using one of two separate methods: CDExpress™ (regression based) or library mode. Schematic representations of both methods are shown in figure 1. In both cases, the spectra measured by placing the grating target under the SE beam are compared against spectra based on a theoretical model of the grating. Both modes use process information (dispersion properties and nominal thickness of all films in the grating region) and estimation(s)
of the grating profile (pitch, nominal CD and height) to generate the theoretical spectra. Benefits and tradeoffs of the two measurement modes are described in previous publications.

Advanced Gate Lithography Process

Figure 2 details a typical gate after-develop-inspect (ADI) process stack at the 65nm node. An optimized combination of two separate, yet similar, dielectric films is deposited on top of the polysilicon that is to be etched. This achieves tight CD control while increasing etch selectivity. The combined thickness of these films is on the order of a few hundred angstroms. Using this dual hard mask stack allows for the photoresist to be thinned down to final patterned heights below 2,000 angstroms. Thinning the photoresist helps expand the exposure and focus process window that provides an acceptable resist profile for pattern transfer.

Correlation Issues with Single-Pass SE

In a typical SpectraCD measurement, all parameters of interest are simultaneously solved for in a “single-pass”. In the case of the aforementioned gate lithography measurement, the signal response and sensitivity to the underlying films requires that the thickness of both hard mask films be floated in the model, in addition to the CD, SWA and height of the resist grating. This leads to a five degrees-of-freedom (5 DOF) solution.

Here, SpectraCD measurements for the gate litho process are correlated against a CD-AFM that is routinely calibrated to a NIST traceable standard. The correlation results between SpectraCD and the AFM for a single-pass library match are shown (blue lines) in figures 3–5. The correlations are generally good for middle CD and resist height, with $R^2$ values $> 0.93$. However, the slope of the line for resist height does deviate ~20% from unity (figure 4). In contrast, the correlation for the resist SWA is very poor (figure 5).

The correlation between the optical profile and the AFM measurements is significantly improved.
Furthermore, when the SpectraCD site-by-site results for resist height and SWA are plotted on the same chart, a clear anti-correlation behavior is seen between these two parameters (figure 6, top). SWA decreases as resist height increases, and vice-versa. This is counter-intuitive to typical behavior of the resist profile through the focus/dose process window.

The poor correlation on SWA, slope deviation from unity on height, and the anti-correlation behavior between SWA and resist height suggest cross-correlation between parameters in the model. An effective method is needed to overcome this limitation.

Feed-forward SE Measurement (Profilm)

One potential method for reducing or eliminating cross-correlation is to minimize the number of parameters in the simultaneous solution. To get accurate solutions for the remaining parameters, however, you need to remove the influence of the excluded parameters from the measured signal or find a way to accurately account for some parameters and fix them during the measurement.

It is difficult to remove the influence of any individual parameters from the measured signal, but since the SpectraCD measurement is an extension of a standard SE measurement, it is possible to accurately measure the underlying film thickness values using SE and then feed those values forward to fix them during the SpectraCD grating measurement.

Figure 7 shows a schematic of this technique (known as Profilm). In step 1, a standard SE film thickness measurement is performed in an unpatterned region where the resist has been exposed. From this measurement, accurate values for the underlying polysilicon, hard mask 1, and hard mask 2 films are obtained. Typically, this measurement is performed at a location within a few hundred microns of the location of the grating target of interest.

On-tool software links the SE film measurement directly to the subsequent grating measurement and feeds the film thickness values into a CDExpress regression-based measurement of the grating. During the CDExpress measurement (step 2) the film thickness values are fixed, thus reducing the CDExpress...
solution from 5 DOF down to 3 DOF (CD, height, and SWA). The combined measurement of the films and gratings is completed in real time with a total MAM time of ~9 seconds per measurement pair. The underlying assumption is that the uniformity of the film thickness is such that the thickness values do not change significantly over the few hundred microns between the open region and the grating region.

**Profilm Improves Correlation Results**

By reducing the number of degrees of freedom in the SpectraCD measurement, the expectation is that small perturbations in the measured signal are handled more appropriately and assigned to the appropriate parameter of interest. Data in figures 3-5 (shown in orange) demonstrate that this concept works. Comparison of standard single-pass library results against Profilm results are shown for resist middle CD and resist height in figures 3 and 4 respectively. There is some improvement in $R^2$ for both parameters, but, more significantly, there is improvement in slope to values closer to unity. Using Profilm, the correlation between the SpectraCD SWA and the AFM SWA (figure 5) is significantly improved. The slope is much closer to unity, and the $R^2$ value is increased from 0.05 to 0.89. Finally, the site-by-site trend of SWA and resist height shows that the anti-correlation behavior between these two parameters is greatly reduced, if not completely eliminated (figure 6, bottom).

In sum, this article shows a method and supporting data to highlight an effective approach to reduce cross-correlation for a 65nm gate lithography process. Anti-correlation behavior between resist height and SWA is significantly reduced, while correlation of the optical profile measurements to CD-AFM is improved for all parameters of interest. These measurements are completed in real time in a full production environment with a MAM time of ~9 seconds per measurement pair.

**Acknowledgements**


**References**


Keep it Flat

The Influence of Backside Particle Contamination on Wafer Deformation during Chucking

Twan Bearda, Frank Holsteyns, Karine Kenis, Paul Mertens – IMEC
Aschwin van Meer – ASML Holding NV
Don Brayton, Lisa Cheung – KLA-Tencor Corporation

Backside particle contamination can cause deformations of the substrate during chucking. These defects, commonly called ‘hot spots,’ typically occur during lithographic exposures or CMP. Previous analysis has shown that the maximum wafer deflection is strongly reduced by plastic deformation of the particles. This article presents a promising study in which particles of different materials such as silica, silicon, tungsten and polystyrene latex (PSL), different sizes, and different densities are deposited on the backsides of 300mm wafers. The resulting changes in flatness during chucking are measured, and found to agree well with theoretical predictions.

Experimental Set-up

In the experimental procedure (figure 1) the flatness of two pairs of 300mm wafers is measured on the leveling system of an ASML TWINSCAN™. The scanner contains two identical chucks that are not flat, but covered by pins or ‘burls’ (figure 2). The diameter and position of these pins is designed to minimize the contact area with the wafer (approx. 3%) while maintaining good wafer flatness. The pins are made of a hard material and have a surface roughness that is negligible compared to the particle sizes relevant in this experiment. Prior to the flatness measurement, ten clean dummy wafers are cycled on each chuck to remove accidental residual contamination. The particle deposition – in different areas on the backside of one wafer of each wafer pair – is carried out by forming an aerosol from a particle-containing aqueous solution. For each deposition spot, the particle material, particle size, particle density, and deposition spot diameter are varied. It is difficult to control the size and density of the deposited particles because of the deposition method, the limited availability of the particles in the requested size range, and the stability of the aqueous solution. The actual values are easily determined, however, using a KLA-Tencor Surfscan™ SP1 unpatterned surface inspection tool equipped with a Backside Inspection Module (SP1-BSIM).

After particle deposition, the flatness of the two wafer pairs is measured again using the same procedure as before, with each wafer measured on the same chuck as in the first measurement. The uncontaminated wafer of each pair is measured before the contaminated one. Assuming that the intrinsic wafer flatness...
does not change significantly during the experiment, it is possible to use this sequence to correct for the drift of the leveling system and for the intrinsic flatness of the contaminated wafers. The resulting flatness change is considered to be solely due to the backside particle contamination.

Note that the two pairs of wafers are shielded by two “umbrella wafers” to minimize uncontrolled particle contamination during all transport. At each step, measurements using the SP1 or SP2 tools help to verify that no uncontrolled contamination occurs during the procedure. Also, before and after the flatness measurement, a KLA-Tencor 23xx bright-field inspection tool is used for extensive review of the deposited particles.

Figure 3 shows the light-scatter maps for the backsides of wafers 2 and 4 after particle deposition. The maps are mirrored to show the actual particle positions when the wafer backside faces down. The particle size histograms for each deposition spot are shown in figure 5 for wafer 2. Additional data are also available for wafer 4 (for comprehensive experimental data, tables and figures, please refer to the original poster presented by the authors at SPIE 2006). Moreover, for reasons that will become clear later, the figures also show the integrated particle volume as a function of particle size. The radii of the deposition spots are listed in table 1.

Figure 4 graphically represents the flatness changes of wafers 2 and 4. Different deflection areas corresponding to different deposition spots are clearly discerned. The detection limit for flatness changes is typically 25nm, however, the interpretation of the results only considers flatness changes larger than 50nm. Table 1 also shows the maximum deflection for each deposition spot, determining the area of the regions (DL) where the flatness change exceeds 50nm. The equivalent radii of these areas are also listed.

Figure 4: Graphically represents the flatness changes of wafers 2 and 4. Different deflection areas corresponding to different deposition spots are clearly discerned. The detection limit for flatness changes is typically 25nm, however, the interpretation of the results only considers flatness changes larger than 50nm. Table 1 also shows the maximum deflection for each deposition spot, determining the area of the regions (DL) where the flatness change exceeds 50nm. The equivalent radii of these areas are also listed.

Theoretical Considerations
A previous report considered the deflection of a wafer due to a point force. In this model, the impact of burls is considered negligible, so the chuck is modeled as a flat surface. The wafer deflection \( w(r) \) has a maximum at \( r=0 \) and extends over an area with radius \( a \). At \( r=a \), the wafer is no longer deformed (figure 6):

\[
w(r) = -\frac{qa^4}{64D} \left(1 - \frac{r^2}{a^2}\right)^2
\]

(1)

**Table 1:** Experimental sequence for particle contamination measurement and correction for flatness changes.

**Figure 1:** Experimental flow.

**Figure 2:** Schematic of the ASML TWINSCAN chuck, as used in this study.

**Figure 3:** Light-scatter maps of wafers 2 and 4 after particle deposition (using a recipe for particle detection in the range of 0.5 - 5µm LSE).
The constant $D$ is the flexural rigidity, which is determined by material properties and the thickness of the wafer. For standard 300mm wafers, $D=5.5\text{nm}$. In many cases, it is found that the force required for deflection exceeds the material strength of the particle. Therefore, plastic particle deformation occurs in order to reduce the deflection and to increase the contact area between wafer and particle. Assuming a cylindrical shape for the particle, it is possible to estimate the resulting deflection as a function of particle size.

Experience shows that backside particles often occur in clusters. In this experiment too, a large number of particles are present within each deflection area. Therefore, a discussion of particle clusters is desirable. The case of two particles that are located close to each other compared to the radius $a$ of the deflection area is discussed. This means that the deflection can still be considered as being caused by a point force. The particles have a cylindrical geometry as before and different sizes (see table 2 and figure 7).

The discussion assumes that the initial height $s_1 > s_2$. Initially, maximum deflection $w_{\text{max}} = s_1$ and no plastic deformation occurs. Then, particle 1 is deformed to reduce its height; the particle/substrate contact area is then given by $V_1/w_{\text{max}}$. This process continues until $w_{\text{max}} = s_2$. From this moment, the smaller particle also contacts the substrate, and the contact area increases from $V_1/w_{\text{max}}$ to $(V_1 + V_2)/w_{\text{max}}$. This argument helps deduce that in case of particle clusters, the total particle volume determines the wafer deflection.

**Discussion of Results**

The results clearly show a dependence of the wafer deflection on the particle material and particle density. In the case of tungsten or silicon particles, a deflection is always observed. PSL or silica spheres only cause a wafer deflection if the size of the deposited particles is large enough. This is expected because of the softness of the PSL and the brittleness of the silica (the chemically grown silica spheres are very porous). However, another explanation for the small wafer deflections is presented below.

<table>
<thead>
<tr>
<th>Deposition spot</th>
<th>Deposition radius (mm)</th>
<th>Spot deflection radius (mm)</th>
<th>Area maximum deflection (nm)</th>
<th>Deposition radius (mm)</th>
<th>Spot deflection radius (mm)</th>
<th>Area maximum deflection (nm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>12.5</td>
<td>DL</td>
<td>12.5</td>
<td>DL</td>
<td>12.5</td>
<td>DL</td>
</tr>
<tr>
<td>B</td>
<td>6.25</td>
<td>DL</td>
<td>6.5</td>
<td>DL</td>
<td>6.5</td>
<td>DL</td>
</tr>
<tr>
<td>C</td>
<td>12.5</td>
<td>12.3</td>
<td>302</td>
<td>12.5</td>
<td>DL</td>
<td>12.5</td>
</tr>
<tr>
<td>D</td>
<td>8</td>
<td>10</td>
<td>689</td>
<td>9</td>
<td>DL</td>
<td>9</td>
</tr>
<tr>
<td>E</td>
<td>12.5</td>
<td>DL</td>
<td>12.5</td>
<td>11.5</td>
<td>1360</td>
<td></td>
</tr>
<tr>
<td>F</td>
<td>7</td>
<td>DL</td>
<td>7.5</td>
<td>DL</td>
<td>14</td>
<td></td>
</tr>
<tr>
<td>G</td>
<td>12.5</td>
<td>7.3</td>
<td>154</td>
<td>14</td>
<td>DL</td>
<td></td>
</tr>
<tr>
<td>H</td>
<td>8.5</td>
<td>12.3</td>
<td>1386</td>
<td>10</td>
<td>14.9</td>
<td>1548</td>
</tr>
<tr>
<td>I</td>
<td>12.5</td>
<td>6</td>
<td>146</td>
<td>12.5</td>
<td>10.8</td>
<td>960</td>
</tr>
<tr>
<td>J</td>
<td>6.25</td>
<td>5.5</td>
<td>161</td>
<td>6.25</td>
<td>5.5</td>
<td>110</td>
</tr>
<tr>
<td>K</td>
<td>13.5</td>
<td>17</td>
<td>1729</td>
<td>12.5</td>
<td>4.8</td>
<td>69</td>
</tr>
<tr>
<td>L</td>
<td>8</td>
<td>10.7</td>
<td>1051</td>
<td>10</td>
<td>8.4</td>
<td>438</td>
</tr>
<tr>
<td>M</td>
<td>12.5</td>
<td>4.9</td>
<td>100</td>
<td>10.5</td>
<td>10</td>
<td>999</td>
</tr>
<tr>
<td>N</td>
<td>6.5</td>
<td>2.9</td>
<td>60</td>
<td>6.5</td>
<td>3.4</td>
<td>91</td>
</tr>
<tr>
<td>O</td>
<td>14</td>
<td>17.4</td>
<td>1766</td>
<td>15</td>
<td>3</td>
<td>66</td>
</tr>
<tr>
<td>P</td>
<td>9.5</td>
<td>13.8</td>
<td>2314</td>
<td>12.5</td>
<td>14</td>
<td>810</td>
</tr>
</tbody>
</table>

Table 1: Radii of deposition spots, and the characteristics of the corresponding deflection areas.

Figure 4: Flatness maps of wafers 2 and 4 as measured on an ASML TWINSCAN leveling stage.
Figure 5: Particle size distributions on wafer 2. Bars show the size histogram (bin size: 1µm), and lines represent the cumulative particle volume.
The model presented previously relates the maximum wafer deflection to the radius of the deflection area (Equation 1 with \( r = 0 \)). Figure 8 shows this relation along with the experimental data obtained in this experiment. Although the experimental data follow a trend that is very similar to the theoretical curve, most data points are located at the right of the curve. This is because the curve is derived under the assumption of a point force. In contrast, the particles are distributed over a large area in this experiment, thus artificially broadening the deflection area.

This discussion on clustered particles helps conclude that, for the analysis of the effect on wafer deflection, the total particle volume should be taken into account. This parameter is shown as drawn lines in figure 5 for wafer 2 (for complete wafer 4 data, refer to SPIE poster\(^*\)) . Clearly, although small particles are much more abundant than large particles, their effect on total particle volume is limited in our experiment. Also, the total particle volume is relatively stable at large particle sizes, so the experimental error can be considered small.

Since only 3\% of the wafer is in contact with the chuck of the ASML TWINSCAN, it is expected that only 3\% of the particles have an impact on the wafer flatness change. Ideally, these particles are identified by comparison of the coordinates of the chuck pins and the light-scattering results. However, both the placement of the wafer on the chuck, and the particle coordinates obtained from the defect reviews are not accurate enough for such an analysis. Therefore, the volume of the particles that have an impact on the wafer flatness is assumed to be 3\% of the total particle volume shown in figure 5.

The particle sizes in the figures are Latex Sphere Equivalent (LSE), i.e. they represent the sizes of PSL spheres that would have the same scattering efficiency as the actual particles. Because the scattering efficiency depends on the material that the particle is made of, the LSE size is different from the actual size. For particles larger than the SP1 wavelength (488nm), Mie scattering occurs. The intensity of the reflected light is

$$\frac{I_{s, Mie}}{I_0} = \pi R^2 \cdot Q_s\quad (2)$$

where \( I_0 \) is the intensity of the incoming light, \( R \) the radius of the particle, and \( Q_s \) the scattering efficiency. The product \( C_s = \pi R^2 \cdot Q \) is called the scattering cross-section of the particle. \( Q_s \) is obtained by numerical calculations\(^*\), and is shown in table 3 for the materials of interest in this study. Also shown is the correction factor applied to PSL-equivalent particle radii obtained from SP1 measurements.

Figure 9 shows the results of the deflection analysis for wafer 2 (results for wafer 4 in original SPIE poster\(^*\)). The drawn line is the theoretical prediction assuming yield strength of 1 GPa. The dashed line indicates the limit where the wafer deflection equals twice the particle radius (single-particle case). This shows wafer deflection as a function of particle radius (volume). For particle clusters, the particle radius is only an ‘effective radius’ representative of the cumulative volume of the cluster. The results from the complete data are very similar, indicating that the experiment is well controlled and reproducible. Comparing the different particle materials, it appears that the small deflections in the case of PSL and silica particles are mainly due to the small particle volume. Apparently the differences in particle yield strength are too small to be observed on a log scale in this experiment. In general, experimental data are of the same order of magnitude as the theoretical predictions, and they follow a similar trend.
Figure 10 shows the same results for wafer 2 as figure 9, but without applying the correction for differences in scattering efficiency (table 3). The line labeled “LSE spec” is a deflection limit based on the theoretical prediction, taking into account that LSE particle diameters may be 30% smaller than actual particle diameters. Clearly, without correction the wafer deformation would have been greater than anticipated, and the particle material dependence might have been wrongly attributed to different yield strengths. More importantly, during inspection the particle material is generally not known, and the particle diameter is expressed in µm as LSE. Critical particle diameters should be converted to this unit, which yields values that may typically be 30% smaller than the values for actual particle diameters.

A random review on defects before and after chucking indicates that most of the particles are not impacted because a chuck pin does not make contact with them. In at least one case, however, the particle is clearly crushed during the chucking. Figure 11 shows the same result in a larger area along with another particle (not crushed) in the same deposition spot.

### Table 3: Sizing of particles relative to PSL particles.

<table>
<thead>
<tr>
<th>Material</th>
<th>n</th>
<th>k</th>
<th>q_s</th>
<th>R_{actual}</th>
<th>R_{PSL}</th>
</tr>
</thead>
<tbody>
<tr>
<td>PSL</td>
<td>1.61</td>
<td>0</td>
<td>2.04</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>Silicon</td>
<td>4.42</td>
<td>0.02</td>
<td>1.44</td>
<td>1.19</td>
<td></td>
</tr>
<tr>
<td>Tungsten</td>
<td>2.47</td>
<td>1.92</td>
<td>1.45</td>
<td>1.19</td>
<td></td>
</tr>
<tr>
<td>Silica</td>
<td>1.46</td>
<td>0</td>
<td>2.20</td>
<td>0.96</td>
<td></td>
</tr>
</tbody>
</table>

### Conclusions

This study investigates wafer deflections due to backside particles during chucking. The experimental data show that the deflections depend on the material, size and density of the particles. The results are explained by plastic deformation of the particles. Low particle densities and small particle sizes are found to cause a smaller wafer deflection. In the case of particle clusters, the total particle volume determines the wafer deflection. As a rule of thumb, one may assume a critical particle volume of 1000µm^3 LSE for a wafer deflection of approximately 100nm.

### References

Overlay control is a vital part of lithography in semiconductor manufacturing. Errors in the overlay of different lithographic levels can cause many electrical problems, thus impacting yield. Ultimately, you can shrink die size with better overlay control. Thus, economics dictates that overlay specifications must shrink along with device geometries, requiring continuous improvement in measurement and control.

**HOW TO CHARACTERIZE**

Overlay is defined as the positional accuracy with which a new lithographic pattern prints on top of an existing pattern on the wafer. Overlay measurement involves the design of special patterns used on two different lithographic printing steps such that a metrology tool can measure overlay errors at that point on the wafer. The older pattern in common use is the “box-in-box” (BiB) target, where an outer box is printed during the first lithographic step and an inner box is printed during the second pass (figure 1). Recently, a new target called AIM has shown superior measurement results. The AIM bars have much higher measurement precision and are immune to processing errors that can damage a traditional BiB target.

The goal of overlay data analysis is two-fold: assess the magnitude of overlay errors and determine, if possible, their root causes. Root cause analysis (extracting knowledge about your lithography process from the measured data) involves explaining the data with a model that assigns a cause to the observed effect.

Figure 1: New AIM targets (right) demonstrate superior measurement results over typical BiB targets (left).
OVERLAY ERRORS

Consider three common sources of reticle overlay errors:

• Rotation of the reticle about an angle $\theta$
• Translation, where the entire reticle field is shifted in x and y by $\Delta x$ and $\Delta y$
• Relative magnification errors of $\Delta M_x$ and $\Delta M_y$ in x and y, respectively

Combining these sources gives this model for final overlay error:

$$dx = -\theta x y + \Delta x + \Delta M_x x$$

$$dy = \theta y x + \Delta y + \Delta M_y y$$

There are key differences between rotational errors applied to a reticle versus those applied to a wafer. Since the reticle field is repeated many times on one wafer, rotating the reticle is very different from rotating the wafer (figure 2). Similarly, a reticle magnification error (caused by the imaging tool) would yield a different signature than a wafer magnification error (caused by thermal expansion of the wafer). Translation errors, however, are exactly the same regardless of whether the offset is on the reticle or the wafer.

Measuring overlay errors is only one step in controlling overlay in a fab. By properly planning out the number and placement of the measurements to be made (sample planning), you can create a model to fit the resulting data such that the coefficients of the model represent physical error terms. These terms can then serve as correctables, which are fed back to the imaging tool to improve the overlay of the next set of printed wafers. The subsequent reduction in the total magnitude of overlay errors on the wafers results in major cost savings for a fab, while improving yield for a given design and allowing subsequent designs to shrink in size.

Chris Mack was Vice President of Lithography Technology for KLA-Tencor from 2000 - 2005. He currently writes and consults in Austin, Texas.

To read an expanded version of this article and for more lithography solutions, go to: www.kla-tencor.com/litho

Your Patterning Process Control Resource

---

Figure 2: Different types of rotation errors as exhibited on the wafer:
1. reticle rotation, and 2) wafer rotation.
The Winning Streak

Advanced Darkfield Inspection for 65nm Design Rules and Below

Catherine Perry-Sullivan, Ph.D., Christine Chua, Ph.D., Matthew McLaren, Ph.D. – KLA-Tencor Corporation

Increasingly complex technical and economic challenges continue to emerge at 65nm design rules and below, driving the need for inspection tools that provide cost- and performance-optimized defect monitoring on a broad range of layers. Conventional darkfield inspection tools based on acousto-optic deflector/photo-multiplier tubes (AOD/PMT) have reached their limit in sensitivity at throughput. This article describes an innovative new inspection technology that meets the sensitivity demands of next-generation semiconductor processing without sacrificing the high production throughput that distinguishes darkfield patterned wafer inspection.

As the semiconductor industry moves below 65nm design rules, it faces integration challenges associated with the introduction of new processes and materials as well as continued cost pressures. Shrinking design rules and process control windows require increased inspector resolution, while new substrates and device materials challenge inspectors’ noise suppression capabilities. The competitive environment, shorter product life cycles and single wafer processing techniques drive the need for cost effective manufacturing, including cost effective inspectors that provide the required sensitivity at production throughputs. To date, laser-based darkfield inspection tools have filled a key role in semiconductor inspection by providing high-throughput defect monitoring capability. However, as the industry moves forward, conventional darkfield inspection technology struggles to meet manufacturers’ inspection needs.

Conventional darkfield inspection tools illuminate the wafer surface with a focused laser spot. An acousto-optic deflector (AOD) sweeps the spot along one axis of the wafer surface while the stage moves perpendicular to the sweep direction in a serpentine pattern. Collectors use photodetectors, such as photo-multiplier tubes (PMTs), to detect the scattered light. Depending on optical configuration and feature implementation (i.e., polarizers, illumination angle), these systems can have excellent noise suppression, and can detect defects much smaller than the spot size. They also have high throughputs, making them ideal for patterned wafer tool-monitoring applications. As for any optical inspector, the resolution of these tools is determined by illumination wavelength (λ) and numerical aperture (NA). If the wavelength is not changed, the tool can be modified to resolve smaller features by increasing the NA. For traditional darkfield inspectors, increasing the NA corresponds to decreasing the spot size. Thus, as design rules shrink, it is necessary to shrink the spot size in order to maintain sensitivity to critical defects. Shrinking the spot size reduces the throughput, diminishing one of the key benefits of these inspectors. In addition, single scanning spot AOD/PMT systems have a maximum possible data rate of ~300Mpps, limiting the extendibility of these platforms to future semiconductor nodes. Introducing multiple spots on an AOD-scanning tool increases the throughput linearly with the number of spots, assuming each spot maintains sufficient photon density. However, multiple spots also increase the complexity of the system significantly, leading to potential problems with reliability and matching.

A laser-based inspector has been developed that incorporates a new darkfield imaging technology. This inspector meets the sensitivity requirements of 65nm and below processing technologies without sacrificing the high production throughputs distinguishing darkfield patterned wafer inspection. This article describes the fundamental features of this inspector, including the innovative darkfield imaging technology, illumination angle, polarizations, and Fourier filtering. Experimental data are presented that support the specific implementations of these features on this inspector. Additionally, several applications for this inspector are described, which highlight its capability for meeting the wafer inspection challenges beyond the 65nm processing node.

Darkfield Imaging Technology

With true brightfield imaging inspection technology, the wafer is flood illuminated through an objective with broadband light. The reflected specular beam is imaged onto a multi-pixel detector creating a high resolution image. The imaging resolution obtained by these tools provides a clear sensitivity advantage – making these tools the sensitivity
leaders in optical patterned wafer inspection. With traditional darkfield technology, the wafer is illuminated with a focused laser spot, and light scattered outside of the specular beam is detected with a PMT. These tools are considered the throughput leaders in patterned wafer inspection. A new darkfield imaging inspector has been designed which incorporates the imaging technology from broadband brightfield inspectors. A focused laser beam illuminates the wafer surface and scattered light is imaged onto a unique, patented multi-pixel sensor, instead of a ‘single-pixel’ PMT. This darkfield imaging tool provides the high resolution needed for today’s design rules without sacrificing the superior throughputs typically associated with darkfield tools. Figure 1 provides an illustration of brightfield imaging, the new darkfield imaging technology, and traditional darkfield scattering. These inspection technology illustrations depict only a subset (normal incidence brightfield and oblique incidence darkfield) of possible tool configurations.

The patented darkfield imaging technology used in the new inspector utilizes a UV laser as the illumination source. A collimated UV laser beam is focused onto a line on the wafer surface. This line is then imaged onto a linear multi-pixel sensor (figure 2). This high resolution, CCD-based sensor is capable of high data rates (>1Gpps) and enables large parallel collection. The optical elements of this darkfield imaging tool are unique to the industry and are critical enabling technologies for high resolution inspection at throughputs typically associated with traditional darkfield tools.

Traditional single scanning spot AOD/PMT systems collect only one pixel at a time, creating bandwidth requirements far in excess...
of those needed by the new darkfield imaging technology, which collects multiple pixels simultaneously. Single pixel data collection and AOD bandwidths limit traditional darkfield technology to maximum data rates of ~300Mpps. In contrast, the data rates of the linear multi-pixel sensor used in this darkfield imaging tool are ~1Gpps and are extendible, as there are no foreseeable limits on driving CCD-technology above 1Gpps. These data rates combine with potential upgrades in NA and λ to create a highly extendible architecture that allows for future resolution and throughput enhancements.

Figure 3 shows two examples of the resolution obtained with this darkfield imaging technology compared to conventional darkfield inspectors. The SEM images show the areas of a logic wafer and a DRAM wafer used in this comparison. Raw scattering images were gathered using a traditional AOD/PMT darkfield system and the new darkfield imaging system. The images taken with the new inspector qualitatively demonstrate the higher resolution of the logic structures and better definition of the array/periphery interface on the DRAM device. This higher resolution translates into increased defect sensitivity and improved inspection capability on smaller design rules.

Other features of this new inspector include low-angle oblique illumination with selectable incident polarizers to maximum surface selectivity and noise suppression. There are also up to three independent collectors – two low-angle collectors and one normal collector. The collectors have selectable polarizers and programmable Fourier filters to minimize pattern and nuisance noise. The following sections describe low-angle oblique illumination, polarization and Fourier filtering in further detail.

### Illumination Angle

Inspection tool sensitivity can be described as directly proportional to defect signal and inversely proportional to wafer noise:

\[
\text{Sensitivity} \propto \frac{\text{Defect Signal}}{\text{Noise}}
\]

It is critical that an inspection tool has a strong defect signal, and equally important that the inspection tool minimizes wafer noise sources. Potential noise sources on a wafer include color variation from film thickness variations, metal grain, and prior-level defects. If these noise sources are not sufficiently suppressed, false defects due to grain or color may be reported. The illumination angle utilized in darkfield inspection is an important design element for determining sensitivity, as it influences both the scattering signal from defects and the background noise characteristics.

Darkfield inspection tools utilize either normal or oblique illumination. Note that what distinguishes darkfield from brightfield inspection is whether the image is formed from the specular beam (brightfield) or the light scattered outside the specular beam (darkfield), not the illumination angle. With normal illumination, the incident laser beam is oriented perpendicular to the wafer surface. With oblique illumination, the incident angle can vary from high-angle, near-normal incidence to low-angle, grazing incidence. While normal illumination can provide strong darkfield defect signal, noise sources such as color, grain and prior-level defects often limit the ultimate sensitivity. The benefits of oblique illumination depend strongly on the exact incident angle. Low-angle (grazing-angle) oblique illumination has the advantage of providing both strong signal from current layer defects and superior noise suppression capability. Low-angle oblique illumination provides significantly higher signal from the wafer surface than from underlying layers, thereby minimizing noise from grain, pattern and color variation while providing surface selectivity to limit the detection of previous layer defects.

---

**Figure 3:** Images taken from a conventional darkfield inspection tool and the Puma 9xxx tool. The Puma 9xxx shows higher resolution of pattern structures on both logic and DRAM devices.
Applications for High Performance Darkfield Inspection

With its high sensitivity, noise suppression capability, and high throughput, the new darkfield imaging inspector is ideal for use in a broad range of semiconductor applications. Three applications are described in detail below. Additional applications include critical defect monitoring for CMP, etch and films.

Front-end DRAM Defect Monitoring: One semiconductor manufacturer used the new inspector for critical defect monitoring on front-end layers for 90nm DRAM production. The inspector was used for baseline inspection and excursion monitoring for residue defects in high aspect ratio (HAR) structures at a buried strap etch processing step. The sensitivity of the new tool for detecting voids at STI CMP was also evaluated. While the new darkfield imaging tool captured a subset of the void defects caught by the standard broadband brightfield inspection, it did so at much higher throughput with less susceptibility to previous layer noise. The results show that the new inspector can be used as a cost-effective void monitor.

Low Cost Photo-Cell Monitoring: Broadband brightfield tools have traditionally been used for photo-cell monitoring (PCM) applications. The critical microphotography defects on photo layers have low topography (stains, developer spots) or are very small (CD variations, bridging, single missing or deformed contacts). The high resolution broadband brightfield technology is ideal for detecting the widest range of these critical defects despite the slower inspection throughput.

The new inspector, with high resolution darkfield imaging capability, has demonstrated high sensitivity to critical defects on PCM wafers. Inspection results from a 70nm PCM DRAM wafer (figure A) show that the inspector surpasses the photo defect detection capability of conventional darkfield inspection and equals broadband brightfield performance at much higher throughputs. These data suggest that for design rules below 90nm, the optimum PCM strategy is a mix-and-match approach using both broadband brightfield inspection and the new darkfield imaging tool. With this new PCM strategy, applications that require the highest sensitivity and capture rate of critical defects (such as resist process development or incoming resist qualification) should utilize the highest resolution broadband brightfield inspectors. However, for daily tool monitoring, where the goal is to capture critical defects at high throughput, the new darkfield imaging inspector should be used. Thus, the typical photo cell monitoring timeline in a semiconductor fab would resemble that outlined in figure B. This strategy provides the maximum sensitivity at the most critical PCM steps, while minimizing overall cost of ownership by...

(continued on page 52)

Figure A: Defect Pareto for a 70nm DRAM PCM wafer showing the inspection results for broadband brightfield, Puma 9xxx (darkfield imaging), and traditional darkfield.

Figure B: Optimum photo-cell monitoring timeline in a fab. This strategy uses the Puma 9xxx as a daily tool monitor while high resolution broadband brightfield tools are used for monitoring the most critical PCM processing steps.
employing a high throughput tool that provides adequate sensitivity to critical photo-related defects.

**Back-end Logic Defect Monitoring:** Back-end logic defect monitoring presents a unique set of inspection challenges. Dense pattern structures with small design rules require high resolution inspection capability for detecting critical defects. Transparent dielectric films, rough metals and multi-layer film stacks test an inspection tool’s nuisance suppression capability by creating multiple noise sources, such as prior-level defects and metal grain.

The new darkfield imaging inspector has been widely adopted for back-end logic defect monitoring. Its superior back-end noise suppression capabilities, due to low-angle oblique illumination and selectable incident and collection polarizations, are illustrated in figures 4 and 5. Figure C further illustrates the inspector’s nuisance suppression by showing inspection results on a 90nm logic metal 4 copper CMP wafer relative to broadband brightfield inspection results. The inspector detects all defects of interest at higher throughput while providing 10x lower capture of hillock nuisance defects. This sensitivity at higher throughput enables higher sampling, while the reduction in nuisance defects improves overall time to results.

The detection capability of the new darkfield imaging inspector on back-end logic devices is further demonstrated in figure D. The bottom Pareto illustrates how the darkfield imaging technology of this new inspector greatly enhances detection capability when compared to a traditional darkfield inspector on a 65nm copper CMP wafer. In the top Pareto, the new inspector demonstrates sensitivity comparable to a broadband brightfield inspector at much higher throughputs on a 90nm trench etch logic wafer. This improved sensitivity combined with superior noise suppression capabilities makes the new darkfield imaging inspector ideal for back-end logic defect monitoring applications.
Oblique illumination also enables the use of selectable illumination polarizations, discussed in more detail in the following section.

Figure 4 presents experimental data demonstrating the noise suppression capabilities of low-angle oblique illumination. It shows raw scattering images from experimental test benches – one using low-angle oblique illumination and one using normal illumination. The top images are from a tungsten deposition DRAM wafer that exhibited significant film thickness variations. Images were acquired from a specific die location for two separate die – one near the center of the wafer, and one near the edge of the wafer. The images collected with low-angle oblique illumination show little sensitivity to the film-thickness variation of the wafer, while the normal illumination images show strong sensitivity to the film-thickness variations.

_Polarizations_

Every process level can have different defects of interest (DOI) and distinct sources of noise. Different illumination and collection polarizations on an inspection tool will provide varying levels of sensitivity to DOI and to noise sources. Without selectable polarizations, first introduced in the early 1990s on the Tencor wafer inspectors, the inspector may miss key DOI or capture false counts related to wafer-to-wafer process variation. Therefore, it is critical that a darkfield inspection tool has the flexibility of using different polarizations in order to maximize DOI capture.

One of the advantages of oblique illumination is that it enables the use of selectable incident polarizations. The incident laser beam can be filtered to generate S, P or C polarized light. With normal illumination there is no
Another example of the use of selectable polarizations is demonstrated with the following data collected on the new inspector on a metal 3 after develop inspect (ADI) wafer. The primary DOI on this wafer is broken resist lines. The primary nuisance source is prior-level, copper filled scratches. Table 1 shows SNRs calculated from images gathered at all nine polarization combinations for a DOI and a nuisance defect. The ideal polarization combination for inspection would be one that maximizes the SNR on the DOI while minimizing the SNR on the nuisance. From this table, it is easily seen that S/P provides the highest SNR (8.33) for the broken line defects and the lowest SNR (1.18) for the prior-level scratches. These results show that S/P is the best polarization combination for meeting the inspection goal of detecting broken resist lines while suppressing prior-level nuisance defects.

Table 1: Signal-to-noise ratios for a DOI (broken resist) and nuisance (prior-level scratch). Data were collected on a metal 3 ADI wafer with the Puma 9xxx at all nine polarizations and demonstrates the power of selectable polarizations in maximizing signal on DOI while suppressing nuisance and pattern noise.
Fourier Filtering

Certain semiconductor devices, such as the array regions of DRAM, SRAM and Flash devices, have areas of repetitive pattern structure. When illuminated with a coherent laser source having a wavelength on the order of the cell spacing of the pattern, this periodic structure gets imaged to discrete lines in the Fourier plane. A simple method of decreasing pattern noise and enhancing the defect signal in these array areas is to use an adaptable Fourier filter – a filter located at the Fourier plane that blocks the diffraction lines resulting from repetitive pattern areas. Using such a filter can significantly increase sensitivity in the array by reducing pattern noise.

Experimental data from the array region of a defect standard wafer—where programmed defects are arranged in a grid by type and size—illustrate the use of a Fourier filter using raw scattering images (figure 6). The image on the left shows an area of the defect standard wafer without Fourier filtering applied. The image is very bright with pattern noise, and it is difficult to distinguish the defects from the background pattern scatter. In the image on the right, the same wafer is shown with Fourier filtering applied. The diffraction pattern noise has been suppressed, and the defects are clearly distinguished from the background. These data clearly illustrate the pattern suppression and defect signal enhancement benefits of using a Fourier filter when inspecting array regions of a wafer.

The new inspector includes programmable, flexible Fourier filters. These are not fixed filters or pre-set masks. Rather, these filters are truly programmable based upon the unique scattering characteristics of each device. The Fourier filters automatically learn the exact location of the diffraction lines in the Fourier plane and then apply a filter to each individual diffraction line. This methodology effectively filters the repetitive pattern noise while minimizing the amount of detection area lost to filtering. Thus, defect signal is maximized while pattern noise is minimized, providing increased sensitivity in array areas.

Conclusions

As the semiconductor industry moves to 65nm design rules and below, semiconductor manufacturers face multiple challenges associated with new materials and tighter geometries. Cost pressures have escalated due to increased competition, shorter product life cycles and the move to single-wafer processing. These technical and economic challenges drive the need for inspection tools that provide cost- and performance-optimized defect monitoring on the broadest range of process layers.

A new laser-based darkfield inspection tool has been developed which utilizes unique, patented darkfield imaging technology to meet these inspection challenges. This technology employs line scanning and a multi-pixel sensor that provides both high resolution to detect critical DOI, and high data rates (>1Gpps) to meet required production throughputs. The tool also draws on a multitude of experimental data to maximize its sensitivity and noise suppression capabilities. As a result of these data, the tool incorporates low-angle oblique illumination for surface selectivity and noise suppression; selectable incident and collection polarizations for maximizing sensitivity to DOI while minimizing noise due to nuisance and film thickness variations; and true programmable Fourier filters to provide superior pattern suppression and sensitivity on memory devices.

Applications for the new darkfield imaging tool include front-end memory defect monitoring, back-end logic defect monitoring and low cost daily photo-cell monitoring, complementing higher sensitivity broadband brightfield inspections at critical photo steps. With its high sensitivity, high throughput and noise suppression capability, the new inspector meets the cost and performance requirements for defect monitoring on a broad range of semiconductor applications at 65nm and beyond.

Acknowledgements

The authors are grateful to Becky Pinto at KLA-Tencor for her critical feedback and valuable discussions on this paper.

References

Stop that Leak
E-Beam Inspection Detects Crystal Defects Early in Device Fabrication

O. Moreau, A. Kang – KLA-Tencor Corporation


An inline voltage contrast detection method utilizing an electron beam (e-beam) inspection tool and specially designed monitor structures is used to reveal crystal defects during the device fabrication process. The correspondence between bright voltage contrast defects and dislocations connecting the transistor source and drain is demonstrated using selective etching followed by SEM and TEM review. Finally, possible approaches to improve the capture rate of dislocations and their correlation to leakage current are discussed.

It has been widely reported that crystal defects can be very harmful in present-day silicon devices when they cause source-to-drain junction piping and subsequent transistor leakage. This failure is explained by anomalous dopant diffusion along the resistive path created by the defect. Defect formation is very often related to the isolation technology, which is responsible for the development of stress, and hence the generation and growth of dislocations. This effect becomes more and more important with shrinking device size and is dramatic when shallow trench isolation (STI) technology is used. It is necessary to identify a method to monitor the silicon crystal quality and to detect the formation of crystal defects at an early stage in the device fabrication process, in order to take the required corrective actions before the device is finished and tested.

Up to now, a combination of selective etching and scanning electron microscopy (SEM) inspection has been the only suitable methodology for identifying these crystal defects. This method has the disadvantage of being destructive and, in addition, it cannot be used for present-generation devices, both because of the reduced device size and because the increased dopant concentration confuses the etching results. This article describes how the KLA-Tencor eS31™ e-beam inspection tool can be used to detect and quantify crystal defects in the device fabrication process.

Experimental Details
Sample preparation: The monitor structures were fabricated using a non-volatile memory process flow, based on 0.13µm CMOS technology, using STI. The active area pattern was defined and etched to create the STI trenches, which were then filled by oxide deposition. After the active oxide growth, the gate electrode was defined. High dose arsenic and boron implantations were used to form the source and drain regions of n- and p-channel transistors, respectively. The implantation damage was annealed by a rapid thermal process (RTP). An oxide was deposited as the pre-metal dielectric; then the contacts were opened and filled with tungsten. The e-beam inspections were carried out at this step of the process flow. The process was completed with metal levels and passivation.

Structure description: Specific test structures for dislocation monitoring were designed to reproduce a critical pattern for dislocation formation and activation, consisting of transistor arrays with a high ratio (4:1) of active area corners to gates (figure 1). The rationale for this design is that the corner region of the active area pattern is the most critical in terms of mechanical stress: Two STI walls converge to an active...
area corner region, and during oxidation the related stress fields superimpose in this region. Dislocations are most frequently observed close to the corners of the active area patterns, and structures with a high density of corners are prone to generating dislocations. The transistor array was designed with a common source region and individual drain regions, connected by parallel metal stripes as shown in figure 1.

This work concentrates on n-channel transistors, because it has been systematically observed that defects are formed in n-channel regions only. In previous works, it was shown that dislocations nucleate in recrystallized amorphous regions produced by the high dose arsenic implantations that are used to form the source and drain regions of n-channel transistors. The formation of p-channel transistors does not involve amorphizing implantation; therefore, no dislocation defects are found in these regions.

Experimental techniques: E-beam inspection provides a real-time electrical test with the capability to scan a wafer for systematic signatures or random electrical failures. When the e-beam strikes a defective floating structure that has a path to ground as a result of a sub-surface failure, the local area can no longer hold the charge. As a result, the local yield of electrons sensed by the system detector is changed, signaling a fault at that spot. The eS31 features a unique extended range of electron optics settings that enable adaptable e-beam current and landing energy for the materials being inspected. Moreover, a bias can be applied to the wafer so that the secondary electrons produced by the beam interaction can be either maintained at the wafer surface or extracted. This is a crucial capability for this study, since it determines the electrical behavior of the source and drain contacts when shorted by the resistive path of a dislocation (figure 2). In the case of the test structure that was investigated, there is no real path to ground; however, since all source contacts are connected, the source line acts as a virtual ground compared to the floating drain contacts.

Some potentially defective structures, as identified by the e-beam inspection, were prepared for microscopy analysis and investigated with conventional microscopy techniques: selective etching followed by SEM inspection and by transmission electron microscopy (TEM). In some samples, the layers above the silicon surface were removed by HF immersion, then received a selective etch (Secco d’Aragona) to reveal defects, and the locations identified by the eS31 were reviewed by SEM. In other samples, a lamella for plan view TEM analysis was extracted at the locations identified by the inspection. A dual-beam focused ion beam (FIB) SEM was used to prepare TEM samples.

Dislocations are responsible for an increase in transistor leakage current under sub-threshold conditions (figure 3). To a good approximation, this comprises a source-to-drain current only, as the drain-to-substrate leakage current is negligible; hence, it is hereafter referred to as the “channel leakage current.” In non-defective structures, the sub-threshold current shows the usual strong dependence on gate voltage and substrate voltage. On the contrary, in defective structures the channel leakage current is weakly dependent on both substrate and gate voltages. Hence, it is possible to choose measurement conditions that reduce the sub-threshold current below the leakage current contribution due to a single dislocation, making any dislocations easily detectable. Table 1 reports the

| \( V_s (V) \) | 0 |
| \( V_d (V) \) | 3 |
| \( V_{sub} (V) \) | 0 |
| \( I_{nd} (A) \) | \((1-3)\times10^{-8}\) |
| \( I' (A) \) | \(10^{-7}\) |

Table 1: Bias voltages used for measuring the sub-threshold current of the structures in figure 1, with resulting sub-threshold and limit leakage currents.
electrical measurement conditions used to test the dislocation monitor structures, including the typical sub-threshold current for non-defective structures, $I_{nd}$, and the limit leakage current value $I^*$, above which the structure is assumed defective. Gate, substrate and drain voltages are also given.

**Experimental Results**

**E-beam inspections:** At the contact tungsten CMP layer, the e-beam inspector was optimized to enhance the voltage contrast signal. A hyper-extracting field of 1500V was applied to reverse-bias the junctions, revealing strong bright defects wherever dislocations facilitate an electron path from the source line to the drain contact (figure 4). Signal-to-noise was so high (3.4) that a four-minute inspection was sufficient to detect defects on all test structures over the entire 200mm wafer.

**Microscopy inspections:** Figure 5 compares the inspector image of a transistor array showing a bright drain contact, and the SEM image of the drain region after delayering and selective etching. Dislocation etch features are clearly identified. Figure 6 shows the TEM plan view of a transistor with a bright drain contact. The TEM image confirms the presence of a dislocation connecting the source and the drain of this transistor. In addition, the microscopy analysis showed additional defects not revealed by e-beam inspection.
Electrical measurement results: Two wafers whose e-beam inspection results gave significantly different voltage contrast defect densities, also showed significantly different electrical performance. Figure 7 shows the leakage current distributions of the dislocation monitor structures in these wafers.

The leakage current maps of the dislocation monitor structures and the voltage contrast maps obtained by e-beam inspection were superimposed for comparison (figure 8). It is observed that the e-beam inspection can discriminate between the low leakage and the high leakage wafer, and also identify the leaky region in the low leakage wafer. It was observed that 80% of the leaky structures were identified by voltage contrast inspection. The remaining 20% were undetected because they lay within an uninspected care area border prescribed by cell-to-cell inspection on the eS31. Future work could increase the inspection area (and its correlation to electrical results) by either using die-to-die inspection, or upgrading to an eS32™, which reduces the care area border.

In some cases, a bright voltage contrast defect unrelated to source-to-drain leakage was also observed. Under hyper-extracting conditions, shorts to ground also appear as bright voltage contrast defects. The shorting can result from a previous-layer particle, residue or pattern defect. Systematic FIB review would confirm the nature of these defects.

Conclusions

Voltage contrast detection using an e-beam inspection system was established as an inline method to detect piping dislocations in devices at an early stage of the devices have confirmed the ability of e-beam inspection to detect dislocations under hyper-extracting conditions. This is promising for the 45nm node, for which strain-induced defects are expected to be among the biggest challenges.

Acknowledgements

We would like to thank Jean Charles Mattlin and Pierre Parlour from ST Microelectronics in Roussset, and Carlo Severgnini from ST Microelectronics in Agrate for their critical support of this work.

Inspections on 65nm devices confirm the ability of e-beam inspection to detect dislocations under hyper-extracting conditions.

References

The Short Loop to Yield
Accelerating Flash Product Inspections using Electrical Defect Monitoring

David M. Price, Garrett Long, Doron Gal – KLA-Tencor Corporation
Laura Pressley, Mike Meyer – Spansion LLC

Floating gate word line structures on flash memory products have layouts that are uniquely suited to being tested using a special, high speed electron-beam inspection methodology previously confined to test structures. This article describes how Spansion’s Fab 25 effectively used this approach to detect small physical defects that were causing a yield excursion in the cobalt silicide (Co$_x$Si$_y$) layer.

In the last few years, flash memory has emerged as the fastest growing segment in the memory market. NOR random-access flash is used in high-performance applications such as networking, cell phones, and games. NAND sequential-access flash is used primarily in mass-storage applications, such as digital cameras, personal digital assistants, and other products requiring memory cards. As consumers continue to demand innovative, small, fast, cheap products, flash memory fills the need for low-cost, low-voltage memory with reasonably fast read-write times.

As the largest company devoted to flash memory products, Spansion is under strong economic pressure to ramp its new products quickly, attain high yields, and achieve a fast time to market—all at the lowest cost possible. Part of the company’s strategy is to detect and eliminate yield-critical front-end-of-line (FEOL) defects as quickly as possible during product development and ramp, and then to monitor defect levels in production to ensure that yield excursions are detected early.

In production, achieving low costs means monitoring defects using a high-throughput inspection system that can capture defects of interest (DOIs). While optical inspection systems normally perform cost-effective line monitoring, they do not easily detect some critical FEOL defect types. Very small physical defects, buried physical defects, and electrical defects can be detected only by using e-beam inspection.

When end-of-line testing at Spansion’s Fab 25 in Austin, TX, encountered a yield excursion that the optical inspectors had failed to detect, a team of engineers began to investigate the problem using e-beam inspection. While the engineers were able to find the defects, which proved to be Co$_x$Si$_y$ fibers less than 20nm in diameter that had shorted two Co$_x$Si$_y$ word lines, scan times took a few hours because a small pixel size was required. It was clear that ordinary e-beam inspection was ineffective because it would significantly limit wafer-level sampling. Hence, an alternative approach was required.

The team recognized that the structure of the Co$_x$Si$_y$ layer in which the fiber defects occurred mimics that of test structures from µLoop™, a non-contact electrical defect monitoring method developed by KLA-Tencor. µLoop uses specially designed test structures together with e-beam inspection to perform inline electrical tests. All defects detected by the method are necessarily yield-limiting, and because it detects the defects’ electrical properties rather than their physical properties, a large pixel size can be used, greatly reducing inspection times.

Because FEOL floating gate word line structures on flash memory products are similar to µLoop test structures, that methodology can be applied without having to manufacture special test structures. Using the method on product wafers enabled Fab 25 to detect killer defects in approximately an hour per wafer, increasing throughput dramatically. That accomplishment greatly accelerated the learning cycle, allowing Spansion to trace the sources of the defects quickly and alter the process to optimize sort yields.
of that signal so that only a portion of the structure must be inspected to locate defects.

The method performs two separate scans to fully characterize defect density. The first scan, called the assess scan, is used to determine the locations of electrical defects in the x dimension—in other words, the lines on which the defects lie. The assess scan can also be used to estimate electrical defect density. The second scan is used to locate the defects in the y dimension. Called the identification (ID) scan, it is made at right angles to the first scan along each line identified as having a defect. After the defect coordinates have been determined, e-beam inspection is used to generate detailed images of the defects for classification and prioritization. At Fab 25, an eS31™ e-beam tool from KLA-Tencor was used.

Since Spansion’s floating gate word line structures have layouts that mimic standard µLoop test structures, the use of the electrical defect monitoring method on product wafers followed the same inspection routine as that performed on test structures. As shown in figure 2, an assess scan (at left) was performed followed by an ID scan (right), which detected killer electrical defects only.

**Applying the Defect Monitoring Method to a CoₓSiᵧ Excursion**

During routine electrical testing of 110nm NOR flash product, Fab 25 discovered a yield excursion in the CoₓSiᵧ layer of the floating gate word line structures. A scanning electron microscope image of the defect is shown in figure 3. The culprit, illustrated in the schematic cross section in figure 4, proved to be a CoₓSiᵧ fiber defect that bridged the CoₓSiₙ word lines and caused a direct electrical short. The defect was difficult to detect using optical inspectors for two reasons: First, its composition and, therefore, its optical properties are very similar to those of the CoₓSiₙ lines on which it lies, and second, it had a diameter of <20nm. While fab personnel were able to

---

* Figure 1: Principle behind the µLoop method. The diagram shows the comb-type structures and an electrical short and open.*

Inspection times were reduced by more than 50%, while tool throughput increased by 71%.

Based on those results, Spansion became interested in deploying µLoop on the production floor. To that end, the team began testing the method on the CoSi layer of multiple production wafers. The application was set up in the same manner as a µLoop test chip, and initial inspection results were fast and accurate.

Figures 5a and 5b compare wafer maps of the CoSi process layer generated during standard e-beam inspection and µLoop inspection, respectively. The standard e-beam scan captured 304 total clustered and non-clustered defects, only seven of which were identified as CoSi fibers, while the µLoop scan captured only seven defects, all of which were found to be CoSi fiber defects. Since the electrical defect monitoring method had a similar capture rate for fiber defects as the standard e-beam method, also without nuisance defects, split lots were run using µLoop inspections to identify and resolve the root cause of the killer defect.

**Root-Cause Identification and Implementation of Process Change**

Employing the defect monitoring method on multiple process split lots, the team conducted several short learning cycles and quickly identified the root causes of the fiber defects. They determined that the interaction of three different process modules was responsible for the defects: shallow trench isolation, stacked gate mask, and poly 2 etch. The most manufacturable solution to the problem was to introduce a new middle-of-the-line (MOL) photoresist/masking process.

The electrical defect monitoring method can differentiate between DOIs—in this case, CoSi fiber defects—and artifacts caused by associated shorting. That ability resulted in the elimination of non-DOI defects. In contrast, a standard e-beam inspection tool captured 297 artifacts. By ignoring the non-DOI defects, the new method realized substantial inspection time and tool throughput improvements.
Once again, e-beam inspection and the defect monitoring method were employed to collect inline defect data from the Co$_x$Si$_y$ layer, with the goal of identifying the specific photoresist process change that eliminated the Co$_x$Si$_y$ fiber defects.

Figure 6 compares an overlay wafer map from the original process with a map from the process with the new MOL photoresist/masking step. The fiber defects from the original process are located in the vertical clustered streak on the lower left side of the map in figure 6a. In contrast, the map from the wafers processed using the new MOL photoresist process (figure 6b) is free of the streak, indicating that fiber defects are not present.

Further evidence that the new photoresist/masking process generates fewer defects than the standard process is presented in figure 7. Figure 7a compares the total number of defective die resulting from defects on the Co$_x$Si$_y$ layer for the standard versus the new process, while Figure 7b compares the end-of-line sort 0 yield bin correlated to the Co$_x$Si$_y$ defects for the old versus the new process. Both the defectivity data and the sort 0 yield loss data demonstrate that the new MOL photoresist/masking process eliminates the Co$_x$Si$_y$ fiber defects.

**Conclusion**

Floating gate word line structures on flash memory products have layouts that are uniquely suited to being tested using µLoop technology and e-beam inspection. Spansion’s Fab 25 used this approach to detect small physical defects that were causing a yield excursion in the cobalt silicide layer.

In addition to detecting the Co$_x$Si$_y$ fiber defects, the electrical defect monitoring method was used to conduct short-loop experiments to determine the root causes of the defect and to qualify a new MOL photoresist/masking process to correct the problem. The µLoop method was shown to eliminate artifacts, improving inspection times by more than 50% and throughput by more than 70%. Since this work was completed, the method has been employed in the fabrication of Spansion’s 90nm MirrorBit™ flash products.

**Acknowledgments**

This article originally appeared in the January 2006 issue of MICRO and is based in part on a paper that was presented at the Eighth Technical and Scientific Meeting of the Centre Régional d’Études en Microélectronique Silicium (CREMSI) 2005. Printed with permission from MICRO and CREMSI/ARCSIS.

The authors would like to thank several people for their contributions to this article, including Chris Foster, Dan Sutton, Mike Covert, Becky Pinto, and the Fab 25 contamination-free manufacturing (CFM) group, which helped define the yield enhancement experiments and collect the data.
Surface Watch
Generating High-Speed, Full-Wafer Maps of Surface Microroughness

Andy Steinbach, Alexander Belyaev, Becky Pinto, David Chen, Sanda Radovanovic, Gordana Neskovic, Hamlyn Yeh, Albert Wang, Jien Cao, Juergen Reich, Dan Kavaldjie, Prasanna Dighe, Rahul Bammi, Louis Vintro, David Bloom – KLA-Tencor Corporation

In an environment where nanometer and sub-nanometer surface topography variations have important ramifications for device performance, a new application of light scattering can provide angstrom-sensitivity, full-wafer maps of microroughness variations at speeds of tens of wafers per hour. These maps correlate strongly with atomic force microscope (AFM) measurements at discrete points. Process window characterization for grain size of interconnect films and unprecedented bare substrate microroughness characterization are now possible, as well as excursion monitoring and process control based on surface topography.

Previously, surface scattering was viewed mainly as a noise source for optical wafer inspection, a limiting factor in particle detection sensitivity. At best, wafer manufacturers and their customers used surface scattering measurements called ‘haze’ as a simple, single-value representation of surface quality, to accept or reject wafers. However, with the proper architecture and processing, an enhanced version of haze called ‘microhaze’ can divulge a wealth of information that is valuable for process development, and may even be used for process monitoring.

A new product called SURFmonitor™ has been designed, which leverages the system architecture of KLA-Tencor’s Surfscan SP2™ unpatterned wafer inspection system to deliver surface-scattering data at unprecedented sensitivity. The measurement sensitivity of this system can be applied to detect changes in surface roughness for various surface types. Sub-angstrom changes in root-mean-square (RMS) surface roughness can be detected for smooth front-end-of-line (FEOL) substrates and films. Excursions in grain size of rough interconnect films such as Cu, W, and polysilicon can be monitored using this technology. This article presents experimental microhaze data for a Cu film, confirming the theoretically predicted functional relationship of microroughness and scattering theory with a very high correlation coefficient.

The ability to effectively measure surface roughness variations over the entire wafer surface that were previously undetected, or detected only with long measurement times or over very small sampling areas, overcomes current sensitivity and throughput barriers to topographic process window characterization in development, and topographic process window control in production.

Measurement Principles

The Surfscan SP2 scanning surface inspection system works by scanning a laser spot over the surface of a wafer with either normal or oblique incidence, and then collecting the scattered light into a narrow or wide collection channel. From that scattering signal, both light point defects (LPDs) and haze can be extracted. Figure 1 shows a simplified means for distinguishing an LPD from haze, by applying a threshold.

The haze portion of the signal is often regarded as nuisance when localized defects are the aim of the measurement. As a result, grazing-angle systems with sophisticated algorithms were designed to suppress haze. Soon wafer manufacturers and their customers discovered that haze maps also contain
important information, because haze correlates with surface roughness. However, the value of haze information has been limited by lateral resolution and sensitivity, by the lack of haze standards, and by the visible wavelength employed by most unpatterned surface inspection systems.

To address these limitations and make best use of the information contained in the haze signal, SURFmonitor was developed to provide a haze map of unprecedented resolution, at UV wavelength. These microhaze maps show surface-roughness, contaminants, residues, film morphology, and even film-thickness variations.

The relationship between haze and surface roughness is summarized here, and can be examined in more detail by consulting the literature\(^2\)-\(^6\). To begin, RMS roughness, \( \sigma \) (or sometimes \( R_q \)), is most simply described as:

\[
\sigma = \lim_{L \to \infty} \frac{1}{L} \left( \int_{-L/2}^{L/2} [z(x) - < z >]^2 \, dx \right)^{1/2}
\]  

(1)

where \( L \) is the distance over which the measurement is taken, \( z(x) \) is the set of heights of data points collected at distances \( x \), and \( < z > \) is the average height. This definition works very well when comparative measurements are limited to one type of measurement system, such as a profiler or an AFM, and measurement parameters (such as scan size and tip radius) are held constant. However, when the intention is to compare widely different systems—in this case a scanning surface inspection system with an AFM—a more general treatment of surface roughness is warranted.

Every wafer surface comprises a continuum of features, with lengths (or spatial wavelengths) ranging from the distance between atoms to the size of the wafer. Equivalently, each surface can be described by a spectrum of spatial frequencies, PSD \((f)\).

In this nomenclature, the RMS roughness for a given measurement system can be obtained by integrating the PSD as follows:

\[
\sigma = \left( \int_{f_{\text{min}}}^{f_{\text{max}}} PSD(f) \, df \right)^{1/2}
\]  

(2)

where \( f_{\text{min}} \) and \( f_{\text{max}} \) describe the limits of the instrument used to measure the RMS roughness. (A one-dimensional description is presented for simplicity, but without loss of generality. In reality, one would integrate in two dimensions.)

For a scanning surface inspection system, \( f_{\text{min}} \) and \( f_{\text{max}} \) can be found by relating the scattering angles to the spatial frequency through the grating equation:

\[
f = \frac{\sin(\theta_s) - \sin(\theta_i)}{\lambda}
\]  

(3)

where \( \theta_i \) and \( \theta_s \) represent the angles of incident and scattered light, and \( \lambda \) is the wavelength of the light. In particular, the spatial frequencies encompassed by the SURFmonitor system are shown in figure 2. Note that four collection subsystems are given: normal-wide, normal-narrow, oblique-wide, and oblique-narrow. These correspond to different optical configurations of the system. Each has a unique spatial bandwidth.

On the other hand, a surface profiler or an AFM has a lower frequency limit equal to the inverse length of the scan,

\[ f_{\text{min}} = \frac{1}{L} \]

(4)

and an upper frequency limit given by the Nyquist criterion as

\[ f_{\text{max}} = \frac{1}{2d} \]

(5)

where \( d \) is the sampling distance or the minimum resolution of the system, whichever is larger. (This work involves taking relatively large AFM scans, so \( d \) is the sampling distance.)
Thus the spatial bandwidth of the AFM measurement will depend upon the scan size and the number of data points sampled per scan line.

Comparing the spatial frequencies encompassed by the SURFmonitor and the AFM (a Park Scientific Instruments M5™), figure 3 shows significant overlap, especially with the oblique-wide channel of the SURFmonitor. However, RMS roughness measurements by the two systems will correlate best when the data of both instruments are filtered to comprise the same spatial bandwidth—that is, when the integration limits of Equation 2 are equal. The green vertical lines in the figure indicate a possible band-pass filter, to compare the oblique-wide channel of the SURFmonitor with a 10mm or 20mm scan of the AFM.

The relationship between haze, $H$, and RMS roughness, $\sigma$, is given by

$$H = \left( \frac{4\pi \cos(\theta_i)}{\lambda} \right)^2 \cdot R_0 \sigma^2$$

where $R_0$ represents the specular reflectance of the surface, and is a function of $\lambda$, $\theta_i$, and polarization of the incident light. Haze, then, varies as the square of the RMS roughness.

For a given surface inspection system ($\lambda$, $\theta_i$, $\theta$, and polarization held constant), the haze measurement should correlate well with the RMS roughness given by an AFM, as long as the two data sets are corrected by the appropriate band-pass filter, to equalize their spatial frequency ranges.

**Surface Microroughness Measurements**

Several simplifications have been made in the preceding discussion. First, we have assumed that the collection of scattered light by the system optics is 100% efficient. In truth a roll-off near the band-pass edges will occur, but this will not have a large effect on the results. Second, we have assumed that surface height variations are small with respect to the wavelength of light—the definition of microroughness, and a condition that is likely well satisfied by any bare or blanket-film wafers we might consider, given that the SURFmonitor uses a wavelength of 355nm. Third, we have assumed that haze arises solely from microroughness effects. A counter example would apply if the system measures material transparent to UV radiation, such as most dielectric films. If the beam penetrates the top surface and reflects from underlying defects or interfaces, the haze measurement will include contributions from sub-surface scattering or thin-film interference effects, not just from microroughness. This kind of haze can also be exploited for process control, and our experiments with using this system to monitor film-thickness variations and detect residual material will be explored in future publications.
To ensure best success in correlating microroughness with microhaze, for our first experiment we chose a wafer that would be as immune to complicating effects as possible: a clean Cu film. At 355nm, the penetration depth of copper is less than 15nm, so haze ought to be a function of only the Cu film topography. In addition, this wafer shows an unusually large variation in grain size across its surface.

We then calculated ‘frequency-matched’ RMS roughness values from these PSDs, i.e. corrected to match SURFmonitor spatial bandwidths (using equation 2, with frequency values from figure 2). Figure 6 compares the resulting microroughness values derived from the AFM data, with the haze values measured by the normal-wide channel of the SURFmonitor. Note that no fitting parameters are used. The red ‘theory’ points are microhaze values predicted from equation 6, based on the frequency-limited AFM data. The green ‘experiment’ points are actual SURFmonitor microhaze measurements at locations where AFM measurements were taken. Note that both the theoretical and experimental data agree well with a common best-fit line—which has a slope of 1.98, very close to the predicted value of 2.

In figure 6 the experimental data show some scatter, which we hypothesize can be attributed primarily to AFM measurement error, such as tip convolution effects. We can conclude

**A full-wafer microroughness map allows engineers to monitor a process for grain-size excursions and thus ultimately ensure consistent device performance.**

Figure 4 shows the full-wafer SURFmonitor image of the Cu film wafer, a data set that can be collected at a throughput of tens of wafers per hour, including overhead for wafer handling. Circled points on the image indicate locations where eight AFM scans were taken. Figure 5a presents three representative AFM scans, showing significant variation in RMS roughness arising from grain size differences. We calculated PSDs from these scans and the others indicated on the wafer map in figure 4. The PSDs are given in figure 5b, with spatial bandwidths of the SURFmonitor optics overlaid.

We then calculated ‘frequency-matched’ RMS roughness values from these PSDs, i.e. corrected to match SURFmonitor spatial bandwidths (using equation 2, with frequency values from figure 2). Figure 6 compares the resulting microroughness values derived from the AFM data, with the haze values measured by the normal-wide channel of the SURFmonitor. Note that no fitting parameters are used. The red ‘theory’ points are microhaze values predicted from equation 6, based on the frequency-limited AFM data. The green ‘experiment’ points are actual SURFmonitor microhaze measurements at locations where AFM measurements were taken. Note that both the theoretical and experimental data agree well with a common best-fit line—which has a slope of 1.98, very close to the predicted value of 2.

In figure 6 the experimental data show some scatter, which we hypothesize can be attributed primarily to AFM measurement error, such as tip convolution effects. We can conclude

**A full-wafer microroughness map allows engineers to monitor a process for grain-size excursions and thus ultimately ensure consistent device performance.**

Figure 4 shows the full-wafer SURFmonitor image of the Cu film wafer, a data set that can be collected at a throughput of tens of wafers per hour, including overhead for wafer handling. Circled points on the image indicate locations where eight AFM scans were taken. Figure 5a presents three representative AFM scans, showing significant variation in RMS roughness arising from grain size differences. We calculated PSDs from these scans and the others indicated on the wafer map in figure 4. The PSDs are given in figure 5b, with spatial bandwidths of the SURFmonitor optics overlaid.
For metal films of interest, such as copper, aluminum and tungsten, microroughness is dominated by grain size. The correlation between the frequency-matched RMS roughness and the measured haze value can be used to construct a microroughness map of the wafer (figure 7). The translation from microhaze to microroughness is accomplished by applying the microhaze-to-roughness correlation as indicated in the figure. The red symbols in the correlation plot are ‘theory’ microhaze generated from equation 6, using the AFM data points indicated in figure 4. This data set is analogous to the red symbols in figure 6.

For metal films of interest, such as copper, aluminum and tungsten, microroughness is dominated by grain size. Figure 5b shows that the SURFmonitor collection frequencies overlay ideally with the broad peak in the PSD that indicates grain size distribution. Thus the SURFmonitor haze map may be uniquely suited to serving as a grain size measurement proxy. Grain size can affect the resistance of the interconnects, and ultimately, the speed of the IC device. Furthermore, grain size has been associated with integrated circuit reliability in metal films, and thus may also provide information useful for predicting the probability of circuit failures. Using the SURFmonitor as a means of generating a full-wafer microroughness map, process engineers may be able to map out a
process window for interconnect microroughness. Furthermore, the ability to generate these full-wafer maps at a speed of tens of wafers per hour enables process monitoring for grain-size excursions.

Future Work

We are currently investigating the relationship between SURFmonitor haze and microroughness for other films, FEOL materials, as well as substrates. For polysilicon, where microroughness affects electron-hole mobility in CMOS devices and capacitance in DRAM trenches, we have measured a similarly strong correlation between frequency-matched AFM microroughness and SURFmonitor haze. For nickel silicide, the changes in film morphology induced by the RTA step cause changes in surface roughness that show up in the value of the microhaze. Another application under study is the detection of pinholes in thin transparent films. In this case, clustered pinholes are in effect creating a surface roughness whose PSD is in the SURFmonitor detectable frequency range, if the density of pinholes is high enough.

The extension of this concept to monitoring the roughness of wafer substrate surfaces is more difficult, mainly because the surfaces are so much smoother, and the noise in the AFM measurements confounds the correlation to haze. We are currently collaborating with a major manufacturer.

SURFmonitor has immediate applications in process development, with the added potential to serve as a high-speed, low-cost process monitor.

Future Work

We are currently investigating the relationship between SURFmonitor haze and microroughness for other films, FEOL materials, as well as substrates. For polysilicon, where microroughness affects electron-hole mobility in CMOS devices and capacitance in DRAM trenches, we have measured a similarly strong correlation between frequency-matched AFM microroughness and SURFmonitor haze. For nickel silicide, the changes in film morphology induced by the RTA step cause changes in surface roughness that show up in the value of the microhaze. Another application under study is the detection of pinholes in thin transparent films. In this case, clustered pinholes are in effect creating a surface roughness whose PSD is in the SURFmonitor detectable frequency range, if the density of pinholes is high enough.

The extension of this concept to monitoring the roughness of wafer substrate surfaces is more difficult, mainly because the surfaces are so much smoother, and the noise in the AFM measurements confounds the correlation to haze. We are currently collaborating with a major manufacturer.
of silicon substrates, to investigate the feasibility of using SURFmonitor as a high-speed, full-wafer proxy for substrate microroughness measurements.

Summary
A new application of surface scattering uses broad collection optics, UV illumination, and new algorithms to produce full-wafer microhaze maps of angstrom-level sensitivity and unprecedented lateral resolution, at a throughput of tens of wafers per hour. We used the new system to measure a blanket Cu film, and demonstrated strong correlation between AFM microroughness measurements and SURFmonitor haze maps, when frequency matching was applied. Using SURFmonitor as a high-speed means of generating a full-wafer microroughness map, process engineers can quickly map out a process window for interconnect microroughness, and then monitor the process for grain-size excursions. In this way, SURFmonitor helps ensure more consistent device performance.

We are exploring further SURFmonitor microroughness applications, including extensions to blanket polysilicon films, silicides, substrate characterization, and detection of pinholes in thin films. In summary, SURFmonitor is finding immediate applications in process development, and also holds potential for serving as a high-speed, low-cost process monitor.

Acknowledgments
This article originally appeared in the July 2006 issue of Nikkei Microdevices.

References
With tolerances shrinking and processes constantly changing, maintaining reticle quality is a growing challenge. Take charge with KLA-Tencor litho control solutions. Get the precise data you need for making smart decisions quickly to ensure reticle quality. With the right reticle qualification strategy, you can pinpoint yield-relevant mask defects before they print. Before they affect yield or device performance. Result: a higher-yielding, reliable patterning process.

www.kla-tencor.com/litho
Your patterning process control resource
Fabs require accurate and rapid disposition decision making during manufacturing, as well as quick assessment of tool and process module output. Traditionally, operators at manual/semi-automated inspection stations performed this task. However, advanced 300mm manufacturing—involving smaller features, factory automation, and larger surfaces—now challenges the operators’ ability to reliably and repeatably assess the wafer and lot. This puts large quantities of wafers at risk.

The Viper 2435XP automated 300mm wafer and tool dispositioning system captures a broad range of defect types at high throughput. Delivering quick go/no-go decisions, the system enables fab engineers to take corrective action early, when wafers can be re-worked, or process tool problems can be repaired before additional lots are risked. Ideally suited for the lithography, CMP, etch, and films process modules, the Viper 2435XP can be rapidly and seamlessly integrated into a production environment.

Building on the strength of the Viper platform, the 2435XP boosts intra-lot throughput to 120 wph, enabling a denser lot sampling plan that delivers tighter control of process excursions. It also provides better inter-lot pipelining: The first wafer of the next lot is queued before the last wafer of the current lot has been unloaded. This increases tool utilization, thus lowering overall cost of inspection.

Questions about how the Viper 2435XP can address a specific use case or challenge? Please contact Bruce Johnson at bruce.johnson@kla-tencor.com

Historically (a), average throughput (dashed lines) was significantly below the maximum throughput spec because of the wafer loading/unloading time. Viper 2435XP (b) features both higher intra-lot throughput and more efficient inter-lot pipelining, enabling higher tool utilization and lower cost of ownership.

Poor coating and other low-contrast defects trigger dispositioning for rework.
Technology innovations and shrinking design rules reduce process tolerances, not only lot-to-lot and wafer-to-wafer, but also intra-wafer and intra-field. To accommodate tight tolerances and rising device complexity, sampling must dramatically increase, along with greater measurement precision and accuracy at higher speeds. Moreover, production line monitoring of profile is now especially yield critical. Chipmakers thus need a long-term solution that fulfills these tough measurement requirements while lowering production and advanced process control (APC) costs.

Offering industry leading reliability, cost of ownership (CoO), and time to yield-relevant data, SpectraCD-XT meets this critical demand for the 65nm node and below.

The SpectraCD-XT dedicated CD/profile metrology tool delivers optimal precision, stability, and matching at a throughput rate 2x that of traditional CDSEM metrology. This performance has consistently proven to improve CD APC in customer fabs over CDSEM-based systems. More importantly, the tool enables cost-effective 3D inline profile measurements for the complete range of critical layer structures.

Built using KLA-Tencor’s benchmark Archer overlay metrology platform with patented spectroscopic ellipsometry (SE) optics and a 50% reduction in move-acquire-measure (MAM) time, SpectraCD-XT effectively allows chipmakers to predict performance and yield on their most complex 65nm production devices, as well as anticipate issues for 45nm products in development. As these benefits ultimately optimize fab productivity and yield, SpectraCD-XT is gaining rapid adoption by semiconductor manufacturers worldwide, often replacing CDSEM metrology in mass production.

Questions about how the SpectraCD-XT can address a tough CD/profile measurement challenge? Please contact Matt Hankinson at matt.hankinson@kla-tencor.com
Sand, sand everywhere...all potentially high-grade silicon down the road. It is a cool spring night in 2016. A worried process engineer is walking in the desert near the new Silicon Oasis complex. Fab 1, the crown jewel of the complex, is ramping production of 22nm devices. It is not going well. The engineer flips through the micrographs for the thousandth time, shoves his PhotoPod back in his pocket, and kicks the sand in frustration. He yelps as his toe hits something solid, and bends over to yank the rock loose and throw it aside.

Hmm, it is not a rock! It is an old oil lamp, the kind that one typically hears about in mythological tales about genies. Hebuffs the metal with his sleeve, then jumps backward as a misty column pours out, rising to tower over his head. A deep voice booms, “I am the Genie of the Lamp! What is your command?”

The quick-thinking process engineer is smart enough to know that modest requests are safer than asking for money, women, or endless power. “You wouldn’t happen to have a tool for sidewall metrology in dense 22nm half-pitch arrays, would you?”

The column condenses and shrinks, collapsing into an elderly man with a rough beard, thick glasses and a pocket protector. The engineer gapes openmouthed, examining this apparition before him. “Would that be dimensional or compositional metrology?” the Genie drones. “Call me Omar, by the way.”

“Ummmm, both, please.” The engineer whips out hisPhotoPod to show Omar the offending structures, hastily adding, “With high throughput, if it isn’t too much trouble…”

Omar looks at the photos, then hands them back. “My fellow genies and I receive many tough requests, from ensuring universal peace to helping the Chicago Cubs win the World Series,” he explains. “Yours is especially difficult. You see, those fins are not necessarily vertical. You have to get a probe down inside the features, and sample all the way up. You must use light or maybe an electrical measurement too, since nothing else will be fast enough or nondestructive enough.”

The engineer nods, “I know.”

“Unfortunately, the wavelength of light is too long,” continues Omar. “You skim right over these features, instead of getting the beam inside. Test structures do not help, either, since what you are measuring depends on the environment around the feature.”

The engineer’s face falls. He looks really miserable. “I know.”

“I am sorry,” says Omar. “We have been working on this since 2006, but so far we have not got anything better than TEM or maybe focused ions. Which, I will bet you already know about. Is there something else I can do for you instead? A winning lottery ticket, perhaps?”

“Well, if I don’t use a FinFET,” insists the engineer. “There’s a planar structure I could use instead. No fancy sidewalls, just a stack of thin layers…”

“Thin, planar films? Piece of cake!” Omar exclaims, rubbing his hands together. “Just thickness and composition?”

“And a small spot size. And…,” the engineer hesitates.

“The spot size is tricky, but we can manage it. What else?”

“Well, everything depends on the interfaces. These layers are all interface – they’re only a few nanometers thick, and there are about three of them. We need to know how many hafnium-silicon bonds are at the interface, and maybe the interface trap density. Oh yes, and the interface we care about is at the bottom of the stack. Is that okay?”

Omar stares at the engineer. Takes his glasses off, wipes his face on his sleeve. Very slowly, he repeats, “You want to count the number of hafnium-silicon bonds. And interface traps. In a 22nm spot. At the bottom of a three-layer stack. To what precision? A few percent?”

“Yes, please,” the engineer says eagerly.

Omar stares up into the night sky for a long moment and mutters something under his breath. Finally, he says, “Let us have another look at those photos, okay?”

Katherine Derbyshire is writing an introduction to IC manufacturing for non-specialists, tentatively titled Semiconductor Manufacturing in Nontechnical Language. She has engineering degrees from the Massachusetts Institute of Technology and the University of California, Santa Barbara. She founded Thin Film Manufacturing, a firm that helps the industry manage the interaction between business forces and technology advances, in 2001. You can reach Katherine at P.O. Box 80229, Stoneham, MA 02180, USA. Tel: +1-781-4389779; E-mail: kderbyshire@thinfilmmfg.com
Come Visit KLA-Tencor at Important Industry Events 2006

For more details and registration for KLA-Tencor events, please visit www.kla-tencor.com/events
BE FIRST
Cut time to market and stay out front

www.kla-tencor.com