PROCEEDINGS OF SPIE

38th European Mask and Lithography Conference (EMLC 2023)

Uwe F. Behringer
Jo Finders
Editors

19–21 June 2023
Dresden, Germany

Organized by
VDE/VDI GMM—The Society for Microelectronics,
Microsystems and Precision Engineering (Germany)
UBC Microelectronics (Germany)

Published by
SPIE

Volume 12802

Proceedings of SPIE 0277-786X, V. 12802

SPIE is an international society advancing an interdisciplinary approach to the science and application of light.

The papers in this volume were part of the technical conference cited on the cover and title page. Papers were selected and subject to review by the editors and conference program committee. Some conference presentations may not be available for publication. Additional papers and presentation recordings may be available online in the SPIE Digital Library at SPIEDigitalLibrary.org.

The papers reflect the work and thoughts of the authors and are published herein as submitted. The publisher is not responsible for the validity of the information or for any outcomes resulting from reliance thereon.

Please use the following format to cite material from these proceedings:
Author(s), "Title of Paper," in *38th European Mask and Lithography Conference (EMLC 2023)*, edited by Uwe F. Behringer, Jo Finders, Proc. of SPIE 12802, Seven-digit Article CID Number (DD/MM/YYYY); (DOI URL).

ISSN: 0277-786X
ISSN: 1996-756X (electronic)

ISBN: 9781510668607
ISBN: 9781510668614 (electronic)

Published by
SPIE
P.O. Box 10, Bellingham, Washington 98227-0010 USA
Telephone +1 360 676 3290 (Pacific Time)
SPIE.org
Copyright © 2023 Society of Photo-Optical Instrumentation Engineers (SPIE).

Copying of material in this book for internal or personal use, or for the internal or personal use of specific clients, beyond the fair use provisions granted by the U.S. Copyright Law is authorized by SPIE subject to payment of fees. To obtain permission to use and share articles in this volume, visit Copyright Clearance Center at copyright.com. Other copying for republication, resale, advertising or promotion, or any form of systematic or multiple reproduction of any material in this book is prohibited except with permission in writing from the publisher.

Printed in the United States of America by Curran Associates, Inc., under license from SPIE.

Publication of record for individual papers is online in the SPIE Digital Library.

SPIEDigitalLibrary.org

Paper Numbering: A unique citation identifier (CID) number is assigned to each article in the Proceedings of SPIE at the time of publication. Utilization of CIDs allows articles to be fully citable as soon as they are published online, and connects the same identifier to all online and print versions of the publication. SPIE uses a seven-digit CID article numbering system structured as follows:

- The first five digits correspond to the SPIE volume number.
- The last two digits indicate publication order within the volume using a Base 36 numbering system employing both numerals and letters. These two-number sets start with 00, 01, 02, 03, 04, 05, 06, 07, 08, 09, 0A, 0B … 0Z, followed by 10-1Z, 20-2Z, etc. The CID Number appears on each page of the manuscript.

Contents

vii *Conference Committee*

EUV LITHOGRAPHY

12802 01 **Emergence of next generation EUV optics: status, outlook and future (Invited Paper)** [12802-3]

12802 02 **Investigation of the resolution limit of Talbot lithography with compact EUV exposure tools** [12802-22]

MASK PATTERNING AND PROCESSING

12802 03 **Multi-beam mask writing opens up new fields of application (EMLC 2023 Best Paper Award)** [12802-45]

12802 04 **Automatic laser writing parameters setting for mature mask production** [12802-7]

MASK AND RESIST OPTIMIZATION

12802 05 **Predicting resist pattern collapse in EUVL using machine learning (ZEISS Award for Talents in Photomask Industry - EMLC 2023)** [12802-29]

12802 06 **Test pattern generation by optimization of the feature space signature** [12802-20]

12802 07 **A deep learning facilitated approach for SEM image denoising towards improved contour extraction for 1D and 2D structures** [12802-48]

12802 08 **Understanding the impact of rinse on SEM image distortion and hole patterning variability** [12802-18]

12802 09 **Training dataset optimization for deep learning applied to optical proximity correction on non-regular hole masks** [12802-24]

OPTICAL AND E-BEAM MASKLESS LITHOGRAPHY AND METROLOGY

12802 0A **LITHOSCALE features accomplish dual exposure and high-resolution patterning (Invited Paper)** [12802-32]

12802 0B **New wave front phase sensor used for 3D shape measurements of patterned silicon wafers** [12802-27]

12802 0C **High-performance computing architecture (HW/SW) for mask CDSEM design based metrology** [12802-10]

12802 0D **Development of an e-beam/i-line stepper intra-level mix and match approach with the photoresist mr-EBL 6000.5 for PIC related structures such as waveguides, ring resonators and coupling structures** [12802-12]

12802 0E **Fabrication of gate electrodes for scalable quantum computing using CMOS industry compatible e-beam lithography and numerical simulation of the resulting quantum device** [12802-36]

MASK METROLOGY, TUNING AND INSPECTION

12802 0F **Opportunities of polarization-resolved EUV scatterometry on photomasks (EMLC 2023 Best Poster Award)** [12802-34]

12802 0G **EUV mask defect inspection for the 3nm technology node (Invited Paper)** [12802-44]

12802 0H **Divide et impera: a short talk about the importance of feature and model engineering** [12802-1]

12802 0I **Efficient mask characterization through automated contour and corner rounding extraction** [12802-9]

12802 0J **Image sharpness score CD-SEM CDSEM monitoring** [12802-35]

12802 0K **Actinic EUV mask qualification for next generation lithography** [12802-50]

12802 0L **SEMI-CenterNet: a machine learning facilitated approach for semiconductor defect inspection** [12802-15]

12802 0M **3D SEM metrology of microstructures for high volume manufacturing** [12802-26]

NANO-IMPRINT LITHOGRAPHY (NIL)

12802 0N **Direct patterning of functional materials using nanoimprint lithography** [12802-4]

DATA ANALYTICS

12802 0O **The image lab sandbox, pulling image computing in wafer fab metrology environment (Invited Paper)** [12802-33]

12802 0P **Improving the on-product overlay performance after optimization of the etch-induced contributions** [12802-8]

12802 0Q **The use of cross-validation for overlay model selection** [12802-13]

12802 0R **YOLOv8 for defect inspection of hexagonal directed self-assembly patterns: a data-centric approach** [12802-16]

Conference Committee

Conference Chairs

Uwe Behringer, UBC Microelectronics (Germany)
Jo Finders, ASML Netherlands B.V. (Netherlands)

Conference Co-chairs

Brid Connolly, Toppan Photomasks GmbH (Germany)
Chris Gale, Applied Materials GmbH (Germany)
Naoya Hayashi, Dai Nippon Printing Co., Ltd. (Japan)

Program Chairs

Ines Stolberg, Vistec Electron Beam GmbH (Germany)
Andreas Erdmann, Fraunhofer-Institut für Integrierte Systeme und Bauelementetechnologie IISB (Germany)
Hans Loeschner, IMS Nanofabrication (Austria)

Program Co-chairs

Jan Hendrik Peters, bmbg Consult (Germany)
Daniel Sarlette, Infineon Technologies Dresden GmbH (Germany)

Additional Program Committee Members

Frank E. Abboud, Intel Corporation (United States)
René Born, Photronics, Inc. (United States)
Albrecht Ehrmann, Carl Zeiss SMT GmbH (Germany)
Dave Farar, HOYA (United Kingdom)
Reinhard Galler, EQUIcon Software GmbH Jena (Germany)
Bertrand Le Gratiet, STMicroelectronics S.A. (France)
Harry J. Levinson, HJL Lithography (United States)
Michael Muehlberger, PROFACTOR GmbH (Austria)
Nico Noack, Advanced Mask Technology Center GmbH Company KG (Germany)
Laurent Pain, CEA-LETI (France)
Kurt Ronse, imec (Belgium)
Serap Savari, Texas A&M University (United States)
Thomas Franz Karl Scheruebl, Carl Zeiss SMS Ltd. (Germany)
Ronald Schnabel, VDE/VDI-GMM (Germany)
Steffen Schulze, SIEMENS (United States)
Rolf Seltmann, Carl Zeiss SMT GmbH (Germany)
Yoshitake Shushuke, NuFlare Technology (Japan)

Frank Sundermann, STMicroelectronics (France)
Raluca Tiron, CEA-LETI (France)
Martin Tschinkl, Toppan Photomasks GmbH (Germany)
Jacques Waelpoel, ASML Netherlands B.V. (Netherlands)
Stefan Wurm, ATICE LLC (United States)
Larry Zurbrick, Keysight Technologies (United States)

Session Chairs

1 Tutorial on EUV Lithography
Albrecht Ehrmann, Carl Zeiss SMT GmbH (Germany)
Hans Loeschner, IMS Nanofabrication (Austria)

2 Student Presentations
Andreas Erdmann, Fraunhofer IISB (Germany)
Laurent Pain, CEA-LETI (France)

3 1st Plenary
Ines Stolberg, Visted Electron Beam (Germany)
Stefan Wurm, ATICE LLC (United States)

4 DUV and EUV Lithography
Jo Finders, ASML (Netherlands)
Albrecht Ehrmann, Carl Zeiss SMT GmbH (Germany)

5 Mask Patterning and Processing
Frank E. Abboud, INTEL (United States)
Martin Tschinkl, Toppan Photomasks (Germany)

6 Mask and Resist Optimization
Kurt Ronse, IMEC (Belgium)
Andreas Erdmann, Fraunhofer IISB (Germany)

7 Poster Session
René Born, Photronics MZB (Germany)
Uwe Behringer, UBC Microelectronics (Germany)

8 2nd Plenary
Naoy Hayashi, DNP (Japan)
Daniel Sarlette, INFINEON (Germany)

9 Maskless Lithography and Metrology
Jan Hendrik Peters, bmbg Consult (Germany)
Rolf Seltmann, RS litho consult (Germany)

10 Mask Metrology, Tuning and Inspection
Thomas Franz Karl Scheruebl, Carl Zeiss SMT (Germany)
Nico Noack, AMTC (Germany)

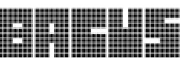
11 Nano-Imprint Lithography (NIL)
Rolf Seltmann, RS litho consult (Germany)
Michael Muehlberger, Profactor (Austria)

12 Data Analytics
Bertrand Le Gratiet, STMicroelectronics (France)
Reinhard Galler, EQUIcon (Germany)

Sponsors and Cooperating Partners

The VDE GMM and the members of the EMLC 2023 Program Committee of the 38th European Mask and Lithography Conference, EMLC 2023, would like to express their sincere appreciation to all the sponsors and cooperating partners mentioned below for their support.

EMLC 2023 Cooperating Partners



EMLC 2023 Sponsors

vistec
Electron Beam

Invited Paper

Emergence of next generation EUV optics: status, outlook and future

Daniel Golde*[a], Björn Butscher [a], Paul Gräupner[a], Peter Kürz[a], Dirk Jürgens[a], Olaf Conradi[a], Jan van Schoot[b], Judon Stoeldraijer[b]
[a] Carl Zeiss SMT GmbH, Rudolf-Eber-Str. 2, 73447 Oberkochen, Germany;
[b]ASML Netherlands B.V. (The Netherlands) De Run 6501,
5504 DR Veldhoven, The Netherlands

ABSTRACT

With already more than 160 EUV scanners operational worldwide, the promise of EUV lithography became a high-volume-manufacturing reality in the past few years. Moreover, EUV lithography has now become the main enabler for the latest generations of chips we all know and use.

ZEISS and ASML keep on developing the capability of EUV tools to further enable upcoming generations of chips. The next step is an increase of the numerical aperture (NA) of our optics from currently 0.33 to 0.55. These high-NA tools will support the shrink prescribed by Moore's Law to continue well into the next decade, by allowing lithographers to print 8nm half-pitch in a single exposure.

We will give an update on the current production status at ZEISS: not only on mirror surface polishing, coating, metrology, but also on mirror handling and integration. Moreover, we will also present the current status and prospects of 0.33-NA optics.

Keywords: EUV, optics, imaging, lithography, high-NA

1. INTRODUCTION

EUV lithography has become an essential part of state-of-the-art chip manufacturing during past years. ASML’s EUV scanners with ZEISS optics are widely used by all major high-end chip manufacturers for producing leading edge electronic devices [1]. Using EUV lithography, the number of costly multi patterning steps can be reduced significantly due to much higher image resolution compared to conventional DUV techniques.

The resolution limit of the patterning process is, in large part, determined by the properties of the optics modules within the scanner. The smallest printable half-pitch CD of an optical system of sufficiently small aberration and flare level is given by the well-known formula [2]

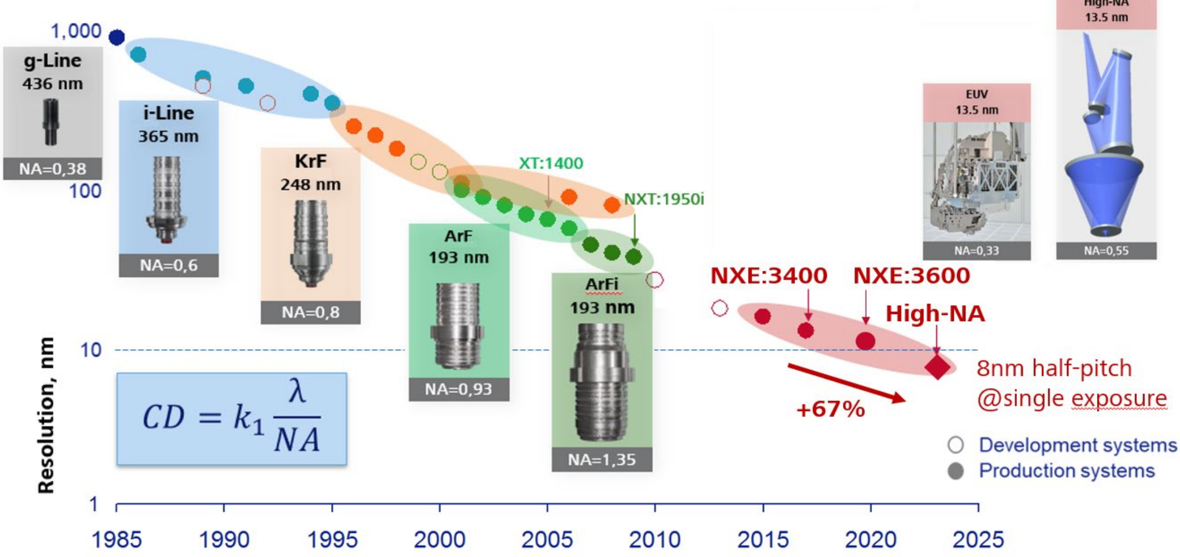
$$CD = k_1 \cdot \frac{\lambda}{NA} \tag{1}$$

where λ is the wavelength of the light used for patterning, NA is the numerical aperture of the projection optics, and k_1 is the so-called process factor. The latter one includes, among other things, properties of the illumination of the reticle like angular distribution and pupil fill ratio.

If we have a look at the lithography systems over the past decades, we see that the resolution has been improved within each product generation by decreasing k_1 while the large steps from one generation to the next one has been realized by addressing the wavelength and the numerical aperture (Figure 1) [3]. Our present and future product strategy follows the same logic. The current state-of-the-art EUV systems support further chip shrink by improving k_1. At the same time, we are preparing the next major technology step by increasing the NA from 0.33 to 0.55. Our high-NA EUV systems are about to be to be delivered to support Moore’s law for the next decade at least.

*daniel.golde@zeiss.com

38th European Mask and Lithography Conference (EMLC 2023), edited by
Uwe F.W. Behringer, Jo Finders, Proc. of SPIE Vol. 12802, 1280201
© 2023 SPIE · 0277-786X · doi: 10.1117/12.2673952



$$CD = k_1 \frac{\lambda}{NA}$$

Figure 1. Lithography roadmap of the past decades following Moore's law by improving λ, NA, k_1.

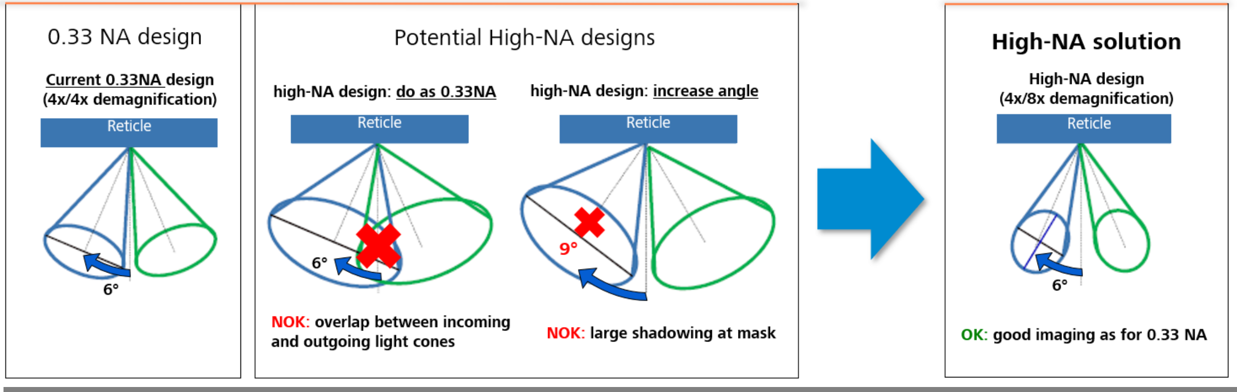


Figure 2. Due to increased angles at reticle for high-NA systems, the demagnification in scan direction needs to be increased from 4x to 8x resulting in an anamorphic lens design.

2. HIGH-NA SYSTEM DESIGN CONCEPTS

One of the major challenges during the design phase of the high-NA optics system was the increased angles of incidence of the EUV light at reticle level. When increasing the NA from 0.33 to 0.55, the light cones at reticle level of incoming and reflected light will overlap if you stick to a 0.33NA-like design, see Figure 2. The straightforward solution of this issue is to just enlarge the chief ray angle (i.e., the angle of incidence in the center of the light cone). In this situation, both light cones would be separated from each other well enough in order to be able to position the respective optical elements. Such a design, however, would suffer from large shadowing effects at the mask absorber due to the large occurring angles of incidences. Imaging assessments showed [4] that the contrast of the image at wafer level would be so low that such a design could not support 8nm half-pitch exposure and, hence, it could not make use of the increased NA for the patterning process. The problem can be solved by increasing the demagnification factor of the lens from the common 4x to 8x which decreases the apertures of the reticle light cones sufficiently to achieve spatial separation of the light cones while not suffering from mask shadowing effects. Since the conflict consists only in one direction (i.e., the scan direction), the 8x demagnification has been introduced in that direction only while the 4x is kept in the direction perpendicular to that. Because of this anamorphic design concept, the mask structures need to be stretched in one

direction by a factor of two in order to obtain the desired image at the wafer. It is important to note that the aperture is still circular at wafer level yielding an isotropic imaging resolution both in scan and non-scan direction [4][5][6][7].

Another consequence of the increased NA compared to the previous EUV systems is a large angle spread of the EUV rays at the second to last mirror, see Figure 3. Although, there are solutions to design a Bragg reflector for such situations, a large angle spread intrinsically results in a huge transmission loss and consequently in an unacceptably poor productivity of the scanner [8][9]. If one would "allow" the light to go through the mirrors from the backside, this complete problem would not exist. This is, of course, not possible but you can do something quite similar: By drilling small holes in the last mirrors, and carefully focusing and threading the light through these holes, the angles of incidence can be decisively reduced. This is exactly the solution that has been chosen for the high-NA system. Consequently, the system has a central obscuration meaning that there will be no light falling on the wafer in perpendicular direction and surrounding angles. In other words, the exit pupil has a central "hole" where no light is transmitted through the optics.

Due to the hole in the exit pupil, the wavefront aberrations of the projection optics cannot be described using the well-known Zernike polynomials any more in a robust way. That is why high-NA aberrations are described using Tatian-Zernike functions which are an orthogonal set of base functions on a ring-like exit pupil [10][11]. At the same time, they exhibit the same basic properties as the standard Zernike polynomials, see Figure 4.

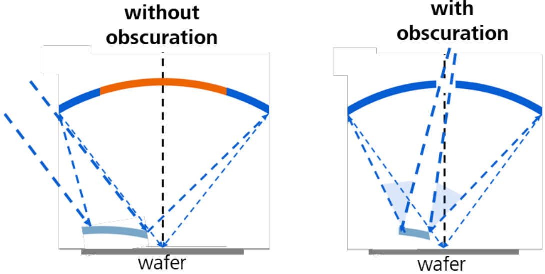


Figure 3. A large aperture at the wafer results in large angles of incidences at the second to last mirror and, hence, to a reduced EUV transmission of the lens. By drilling a hole in the last mirror, angles can be reduced and EUV transmission improved.

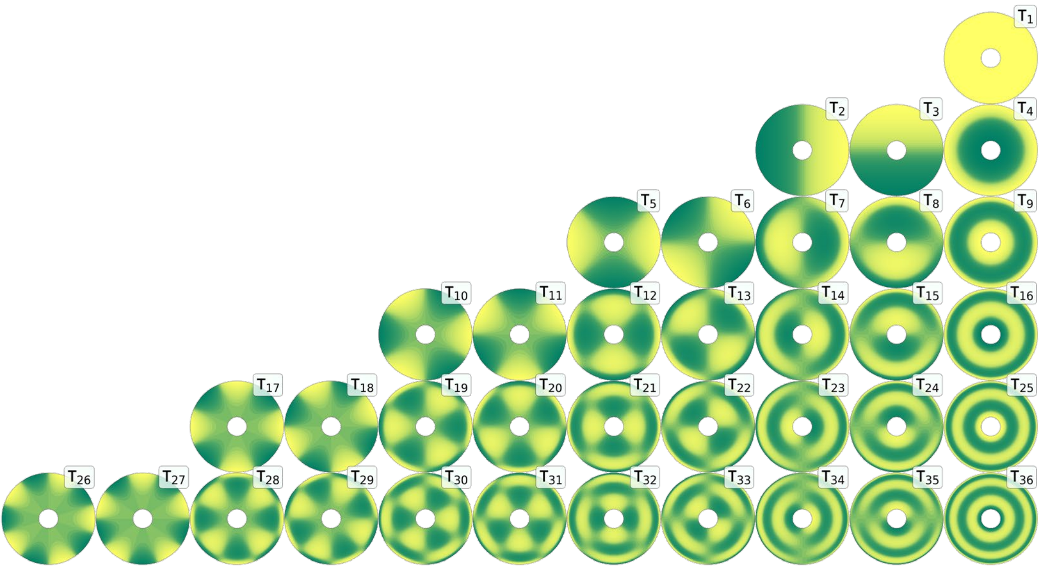


Figure 4. The first 36 Tatian-Zernike polynomials used to specify and describe the aberrations of the high-NA optics.

3. MANUFACTURING THE HIGH-NA OPTICS SYSTEM

There are basically two key drivers that represent the major challenges for the manufacturing processes of an EUV system. First of all, the mirrors need to be mounted and position-controlled very accurately in order to meet the tight overlay specification of the mask structures on the wafer. Just to give an example: Our mirror position control needs to be precise enough to guide a light beam to the moon and hit a baseball located there. Secondly, the mirror figure must be produced and measured with a sub-nm accuracy of over eight decades of spatial wavelengths to support the needed imaging quality of the chip features. This is a precondition to achieve the feature resolution as given by Equation (1). These two challenges come together with the overall increased dimensions of the high-NA design compared to our previous EUV products. Consequently, we could not just re-use existing technology and processes but needed to build a completely new manufacturing and integration infrastructure.

Figure 5 gives an overview of our high-NA integration infrastructure. On the left-hand side the two parallel branches of the integration process are shown: The coated mirrors are mounted in their individual mirror modules and qualified on single-module level. At the same time, the frames are assembled and qualified before the mirror modules are integrated in the frame and, after final cleaning and testing, shipped to ASML.

In order to reduce risks of the integration process, the complete sequence has been tested by building forerunners for projection optics and illumination system. These proto systems have been delivered to ASML and have already been integrated in the first high-NA scanner, see Figure 6.

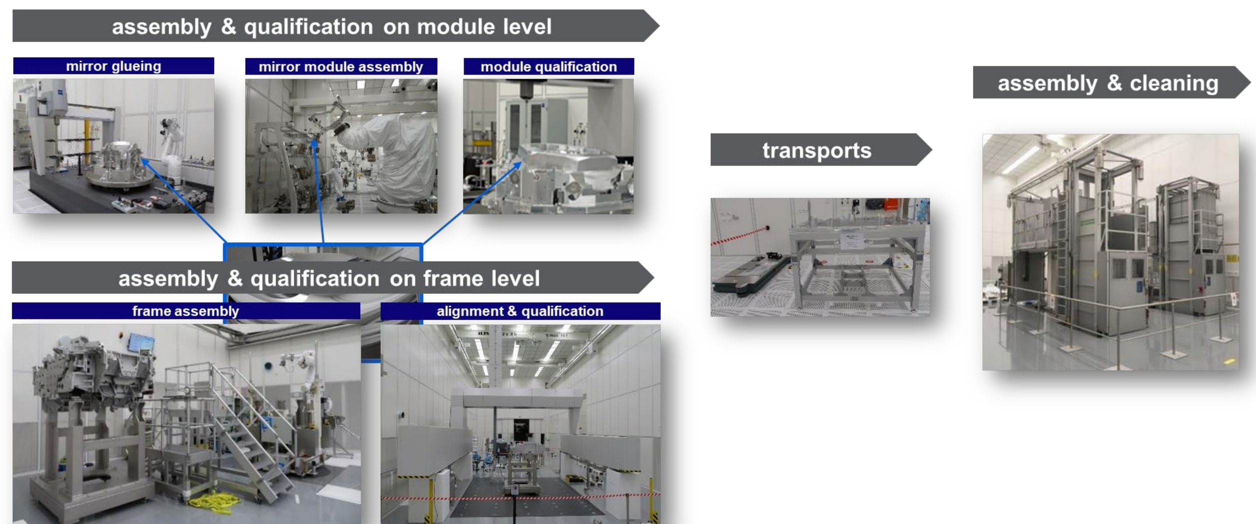


Figure 5. Integration infrastructure of the high-NA systems.

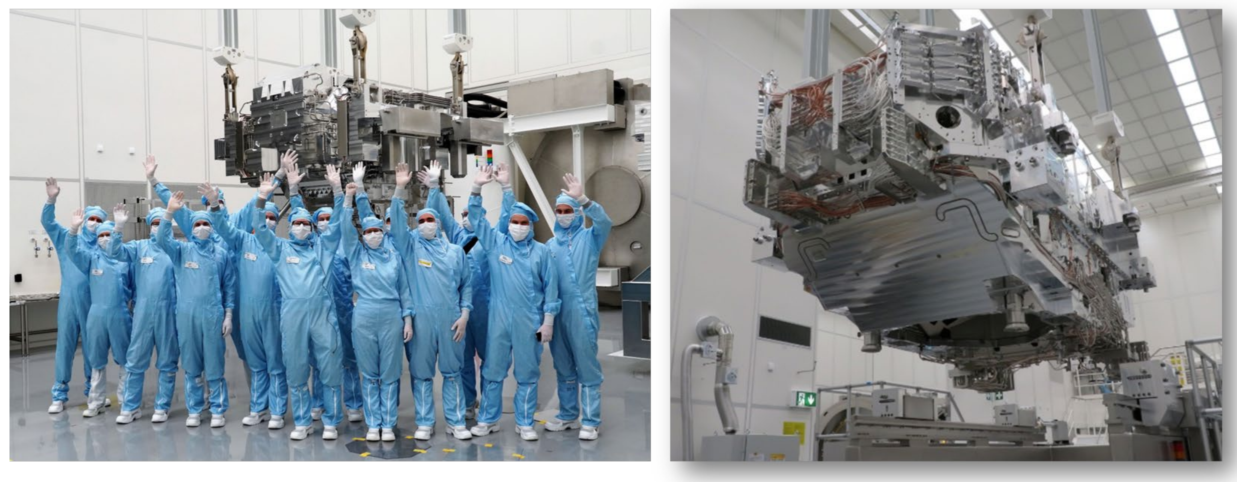


Figure 6. Shipment of first illuminator (left) and projection optics forerunner (right).

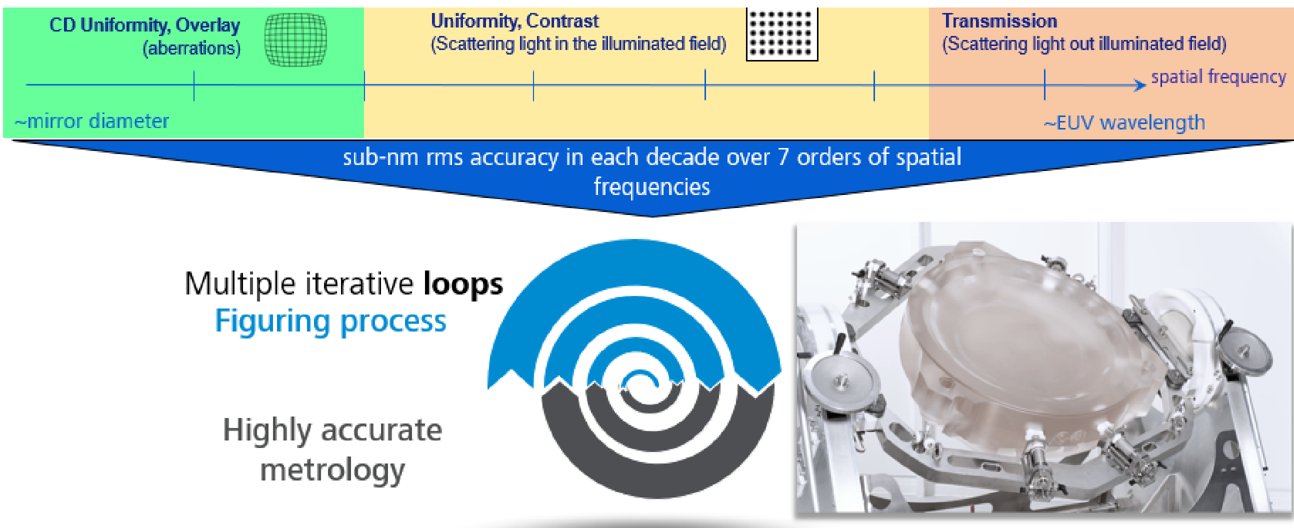


Figure 7. Using iterative figuring and metrology loops, we achieve sub-nm surface accuracy over a wide range of spatial frequencies on the mirror.

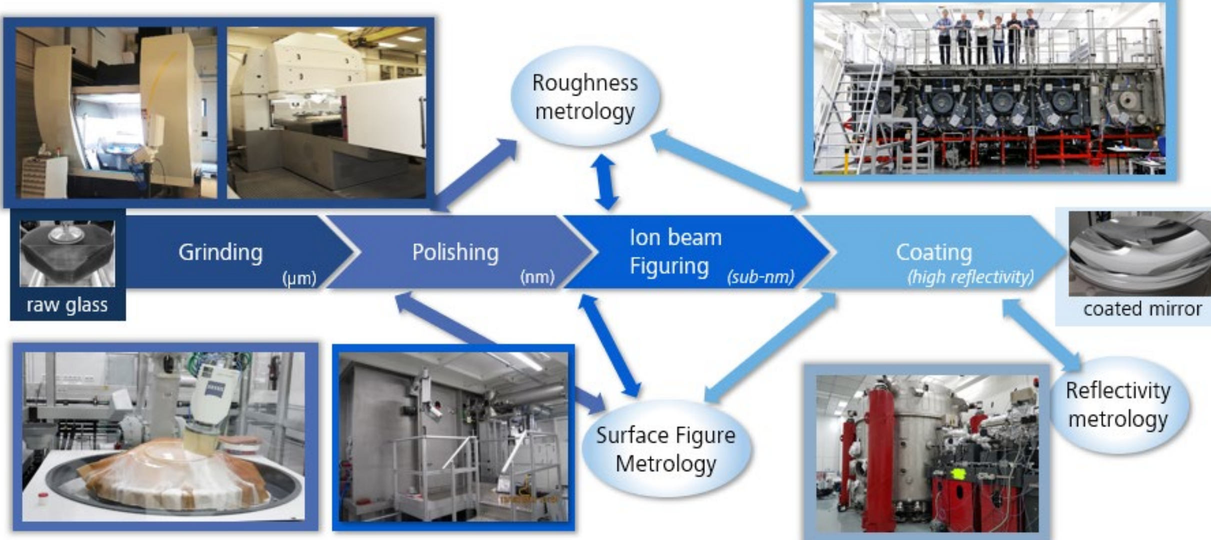


Figure 8. Schematic of the iterative figuring process of high-NA mirrors.

For figuring our mirrors to the desired sub-nm surface accuracy level, we use an iterative approach consisting of alternating figuring and metrology steps. Every figuring step is followed by a measurement that gives the recipe for the next figuring loops and so on. This approach allows us to come from the mm world to the pm world within a few process steps. Figure 8 shows an overview of our various mirror manufacturing technologies. The order of these techniques is chosen such that the fast but comparably inaccurate come first and we end with the high-precision processes.

The manufacturing process of the high-NA mirrors is confronted with tougher challenges compared to the previous EUV systems: larger mirrors with up to 1m diameter and stronger aspheres. Similar as for the integration processes also the mirror production demanded new infrastructure for handling, figuring, coating, and measuring the mirrors.

Meanwhile, the mirrors for the first systems are finished and within desired specifications. They are currently being integrated in their frames and the first systems are about to be delivered to ASML. Our focus has now transferred to increase the production capabilities in order to satisfy the challenging shipment demands of our customers. At the same time, we are preparing the next steps of the high-NA roadmap to further improve imaging performance and productivity of our systems.

4. SUMMARY

In recent years, the EUV systems by ASML and ZEISS have become an essential part of producing leading edge electronical devices. To further support the continuation of Moore's law, we are already working on and are about to deliver the next generation of EUV-systems called high-NA. By increasing the numerical aperture of the optical system from 0.33 to 0.55, a single-exposure half-pitch resolution of 8nm can be achieved.

To cope with fundamental challenges caused by the increased NA, two major design changes are introduced in our high-NA systems for the first time. Other than in previous systems, we changed the demagnification from 4x to 8x in scan direction while 4x is kept in the direction perpendicular to that (anamorphic system design). Another change is the introduction of a central obscuration causing a central hole in the exit pupil where no light is transmitted. Both changes were done to optimize both imaging performance and EUV transmission.

Besides these design features, high-NA has new challenges also for our internal production processes. A completely new infrastructure has been set up and production is ongoing at full speed. Important milestones so far were: (i) shipment of the first mechatronic forerunner systems to ASML and (ii) in-spec production of the complete mirrors sets for the first systems. High-NA roadmap extensions as well as production ramp-up are the fields that we are concentrating on now.

ACKNOWLEDGEMENT

We would like to thank all people that have helped to bring the high-NA program to that point where it currently is and will help to make it reality also for our customers. Special thanks to our colleagues in Veldhoven and Oberkochen for many fruitful discussions.

This work has received funding form the ECSEL Joint Undertaking (JU) / the German Federal Ministry of Education and Research under grant agreements No. 662338 / 16ESE0036K , 692522 / 16ESE0072K, 737479 / -, 783247 / 16ESE0287K, 826422 / 16ESE0377K, 101007254 / 16MEE0192K as well as from the German Federal Ministry for Economic Affairs and Climate Action under grant 16IPCEI205. The JU receives support from the European Union's Horizon 2020 research and innovation program and the member and associated states. Responsibility for the content lies solely with the authors and neither the ECSEL JU nor the German Federal Ministries may be made responsible for information it contains.

REFERENCES

[1] Christophe Smeets, Guido Salmaso, Joe Carbone, Marcel Mastenbroek, Nico Benders, Roderik van Es, and Roelof de Graaf "0.33 NA EUV systems for high-volume manufacturing", Proc. SPIE PC12051, Optical and EUV Nanolithography XXXV, PC1205103 (13 June 2022); https://doi.org/10.1117/12.2614220

[2] Burn J. Lin "Where Is The Lost Resolution?", Proc. SPIE 0633, Optical Microlithography V, (20 August 1986); https://doi.org/10.1117/12.963701

[3] Bartosz Bilski, Jörg Zimmermann, Matthias Roesch, Jack Liddle, Eelco van Setten, Gerardo Bottiglieri, and Jan van Schoot "High-NA EUV imaging: challenges and outlook", Proc. SPIE 11177, 35th European Mask and Lithography Conference (EMLC 2019), 111770I (29 August 2019); https://doi.org/10.1117/12.2536329

[4] Neumann, J.T., Roesch, M., Graeupner, P., Migura, S., Kneer, B., Kaiser, W., van Ingen Schenau, K., "Imaging performance of EUV lithography optics configuration for sub-9nm resolution," Proc. SPIE 9422, (2015).

[5] Migura, S., Kneer, B., Neumann, J.T., Kaiser, W., van Schoot, J.B.P., "EUV lithography optics for sub 9nm resolution," International Symposium on Extreme Ultraviolet Lithography, Washington D.C. (2014).

[6] Jan van Schoot, J.B.P., Valentin, C., van Ingen-Schenau, K., Migura, S., "EUVL lithography scanner for sub 9nm resolution," International Symposium on Extreme Ultraviolet Lithography, Washington D.C. (2014).

[7] J. van Schoot, J., van Ingen Schenau, K., Valentin, C., Migura, S., "EUV lithography scanner for sub 8 nm resolution," Proc. SPIE 9422, (2015).

[8] Paul Graeupner, Peter Kuerz, Thomas Stammler, Jan van Schoot, and Judon Stoeldraijer "EUV optics: status, outlook and future", Proc. SPIE 12051, Optical and EUV Nanolithography XXXV, 1205102 (26 May 2022); https://doi.org/10.1117/12.2614778

[9] Jan van Schoot, Eelco van Setten, Gijsbert Rispens, Kars Z. Troost, Bernhard Kneer, Sascha Migura, Jens Timo Neumann, and Winfried Kaiser "High-numerical aperture extreme ultraviolet scanner for 8-nm lithography and beyond," Journal of Micro/Nanolithography, MEMS, and MOEMS 16(4), 041010 (30 October 2017). https://doi.org/10.1117/1.JMM.16.4.041010

[10] Laurens de Winter, Timur Tudorovskiy, Jan van Schoot, Kars Troost, Erwin Stinstra, Stephen Hsu, Toralf Gruner, Juergen Mueller, Ruediger Mack, Bartosz Bilski, Joerg Zimmermann, and Paul Graeupner "Extreme ultraviolet scanner with high numerical aperture: obscuration and wavefront description," Journal of Micro/Nanopatterning, Materials, and Metrology 21(2), 023801 (21 April 2022). https://doi.org/10.1117/1.JMM.21.2.023801

[11] Berge Tatian, "Aberration balancing in rotationally symmetric lenses", Journal of the Optical Society of America, Volume 64, Number 8, (1974).

Investigations on the resolution limit of Talbot lithography with compact EUV exposure tools

Bernhard Lüttgenau*[a,b], Sascha Brose[a,b,c], Serhiy Danylyuk[c], Jochen Stollenwerk[a,b,c], Carlo Holly[a,b,c]
[a]RWTH Aachen University TOS - Chair for Technology of Optical Systems, Steinbachstr. 15, 52074 Aachen, Germany; [b]JARA - Fundamentals of Future Information Technology, Research Centre Jülich, 52425 Jülich, Germany; [c]Fraunhofer ILT - Institute for Laser Technology, Steinbachstr. 15, 52074 Aachen, Germany

ABSTRACT

This paper focuses on the design and fabrication of phase-shifting transmission masks tailored for high-resolution nanopatterning using a compact EUV exposure tool. The authors analyze various factors that influence the achievable resolution, aiming to push the boundaries towards the sub-10 nm range, approaching the theoretical resolution limit.
The demand for high-resolution nanoscale patterns spans across diverse applications, driving the need for compact exposure tools and lithographic concepts. The developed EUV exposure tool can be operated at either 10.9 nm or 13.5 nm exposure wavelengths depending on the specific use case. This capability allows for large area nanopatterning with enhanced throughput as well as industrial resist qualification with focus on highest resolution.
The utilized discharge-produced plasma (DPP) EUV source offers partially coherent radiation. For this radiation type, the (achromatic) Talbot lithography has proven to be the most effective with resolution in the sub-30 nm range and a theoretical resolution limit of less than 10 nm. To optimize the intensity distribution in the wafer plane, the authors use rigorous coupled-wave analysis (RCWA) simulations to fine-tune the material composition and geometry of the masks. Various factors influencing the achievable resolution are identified and presented.
In addition to simulative optimization, the fabrication of dense periodic nanopatterns poses increasing challenges for smaller periods. In this work, the mask fabrication process is optimized to produce stable and high-resolution periodic mask patterns, leading to record resolutions for both line and contact hole periodic nanopatterns with the presented setup.

Keywords: laboratory exposure tool, achromatic Talbot lithography, interference lithography, holographic lithography, resist characterization, phase-shifting masks, transmission gratings, aerial image contrast

1. INTRODUCTION

Dense periodic nanostructures with feature sizes below 100 nm are essential for numerous industrial and scientific applications, such as photoresist qualification for achieving high resolution[1,2] and the production of metamaterials, biosensors, quantum dot arrays, and artificial crystals[3-5]. Fabricating these nanostructures is a complex task, requiring advanced patterning technologies like electron beam lithography, nanoimprint lithography, or industrial projection lithography. Each method has its strengths and limitations. Electron beam lithography offers high resolution but suffers from limited throughput. Nanoimprint lithography provides high throughput but lacks pattern fidelity for large area nanopatterning. Industrial projection lithography ensures high pattern fidelity and throughput, but its application is mostly limited to major companies in the semiconductor industry due to high costs. Qualifying next-generation photoresists for extreme ultraviolet (EUV) lithography presents even greater challenges, as it often demands resolutions beyond the state-of-the-art. Research on photoresists and related processing relies mostly on high-volume manufacturing (HVM) tools[6] and synchrotron facilities[7-10]. To optimize the photoresist development process and ensure quality control, sensitivity, contrast and resolution tests are necessary. However, using high-volume manufacturing (HVM) tools for these tasks is often impractical due to their high operational cost and complexity. Additionally, operating such tools at the resist supplier's site may not be feasible.

*bernhard.luettgenau@tos.rwth-aachen.de; phone 0049 241 8906-301; fax 0049 241 8906-121; tos.rwth-aachen.de

38th European Mask and Lithography Conference (EMLC 2023), edited by
Uwe F.W. Behringer, Jo Finders, Proc. of SPIE Vol. 12802, 1280202
© 2023 SPIE · 0277-786X · doi: 10.1117/12.2675605

Despite their advantages, the limited number of synchrotron facilities and the limited availability of beamtime slow down the necessary iteration steps for this research and development, resulting in a long time to market.
To address these issues, compact nanopatterning tools using EUV radiation are a promising solution. The developed compact EUV exposure tool at RWTH Aachen University employs a discharge-produced plasma (DPP) EUV source for large-area nanopatterning and industrial EUV photoresist qualification. Talbot lithography has proven to be an effective method for this setup, utilizing all diffraction orders from a transmission grating for efficient patterning[11,12]. The use of interference effects results in homogeneous and high-contrast intensity modulations along with pattern demagnification of the mask pattern and an effective local defect compensation[13,14]. Talbot lithography offers the usage of coherent and partially coherent radiation with finite relative spectral bandwidth, making it ideal for plasma-based EUV sources[15,16] and synchrotrons[17]. The achievable resolution of the nanopatterns depends on several factors, including the optical properties of the EUV radiation, stability of the mask-wafer-positioning system, the photoresist performance, and the transmission mask.
This study focuses on the transmission mask and its impact on the resolution limit of achromatic Talbot lithography for the presented compact EUV exposure tool. Different transmission mask materials and designs are explored to achieve high-contrast intensity distributions. Phase-shifting transmission masks are particularly effective in suppressing the 0th diffraction order, resulting in a classical two-beam interference pattern with maximum intensity distribution contrast of 100%. However, the fabrication of ever smaller structures becomes challenging due to increasing aspect ratios. So far the fabrication of mask structures down to 40 nm half-pitch for contact holes and 65 nm for lines and spaces was demonstrated, resulting in 28 nm half-pitch[18] and 32.5 nm half-pitch[19] patterns in the exposure results, respectively. By optimizing mask geometries or using alternative mask materials, further improvement in resolution can be achieved.
The study involves a rigorous simulative evaluation of phase-shifting masks for achromatic Talbot lithography at EUV wavelengths. The goal is to increase the resolution by optimizing the transmission mask geometry and material system, ultimately enabling sub-30 nm patterning. These advancements pave the way for a compact EUV exposure setup suitable for both industrial EUV photoresist qualification and large-area nanopatterning for various scientific applications.

2. ACHROMATIC TALBOT LITHOGRAPHY

Achromatic Talbot lithography is based on the self-imaging of a grating pattern via the interference of propagating grating diffraction orders[20,21]. This method is particularly suitable for plasma-based EUV sources and synchrotron facilities, allowing high-resolution nanopatterning of dense periodic structures over large areas. The underlying principle of (achromatic) Talbot lithography is depicted in Fig. 1a.
The incoming EUV radiation is diffracted by the grating structures of the transmission mask, leading to a complex intensity distribution behind the mask. The bright spots in this distribution correspond to intensity maxima in the interference pattern. At specific distances behind the mask, the intensity distribution at the contact point ($z = 0$) is replicated in the aerial image. These distances, known as Talbot distances, can be calculated using $z_{T,n} = n \cdot \lambda / (1 - (1 - (\lambda^2 / d^2))^{1/2})$, where λ represents the radiation wavelength, d is the grating period, and n is a positive integer. When irradiated with a certain spectral bandwidth $\Delta\lambda$, the intensity maxima elongate along the propagation axis (z-axis) with increasing distance from the mask. At the achromatic Talbot distance $z_a = 2 \cdot p^2 / \Delta\lambda$, the intensity maxima are fully convergent, forming a stable intensity distribution with an up to two times reduced periodicity compared to the mask's periodicity (as indicated by the blue square). This effect enables a two times pattern demagnification of the mask pattern in the exposure result[16,17]. In addition, a certain amount of displacement or tilt can be tolerated due to the stable intensity distribution in the achromatic Talbot regime, which reduces the precision requirements for mask-wafer positioning. The Talbot lithographic approach is ideally suited for partially coherent radiation with a relative spectral bandwidth of 2% to 4%. Consequently, it finds direct application in nanopatterning using radiation from plasma-based EUV sources, especially those operating with an Argon/Xenon gas mixture, resulting in a primary wavelength of 10.9 nm[15,22]. In cases where EUV radiation with a main wavelength of 13.5 nm is needed (e.g., for photoresist characterization), the source can be operated with Xenon mainly, and the emitted radiation can be spectrally filtered for the target wavelength and spectral bandwidth using a tailored multilayer mirror[23].
The maximum mask-to-wafer distance, z_{max}, in achromatic Talbot lithography depends on the spatial coherence length of the emitted radiation. It is given by $z_{max} = l_s \cdot p / (2 \cdot \lambda)$, where l_s is the spatial coherence length, p is the grating period, and λ is the wavelength. This parameter, along with the spectral bandwidth, determines the depth of the distance window available for exposure[16]. In plasma-based EUV sources, spatial coherence is limited and can only be increased through spatial filtering or by extending the mask-source distance. However, this results in reduced intensity in the wafer plane,

so the spatial coherence is typically maintained between 20 and 40 µm. Consequently, the exposure distance window remains below 100 µm, requiring precise positioning of the mask and wafer.
To achieve high-resolution patterning within this distance window, it is essential to maximize the intensity contrast of the aerial image, defined by $((I_{max} - I_{min}) / (I_{max} + I_{min}))$. Practical resolution limits are influenced by the availability of high-contrast photoresists and the stability of the mask-wafer positioning during exposure. Enhanced throughput and imaging quality, leading to higher resolution, can be achieved using an efficient illuminator[24]. The aerial image quality primarily depends on the transmission mask design and directly influences the attainable resolution. To optimize the achievable resolution, a comprehensive understanding of the optical processes and simulation of the resulting intensity distribution behind the mask is required.

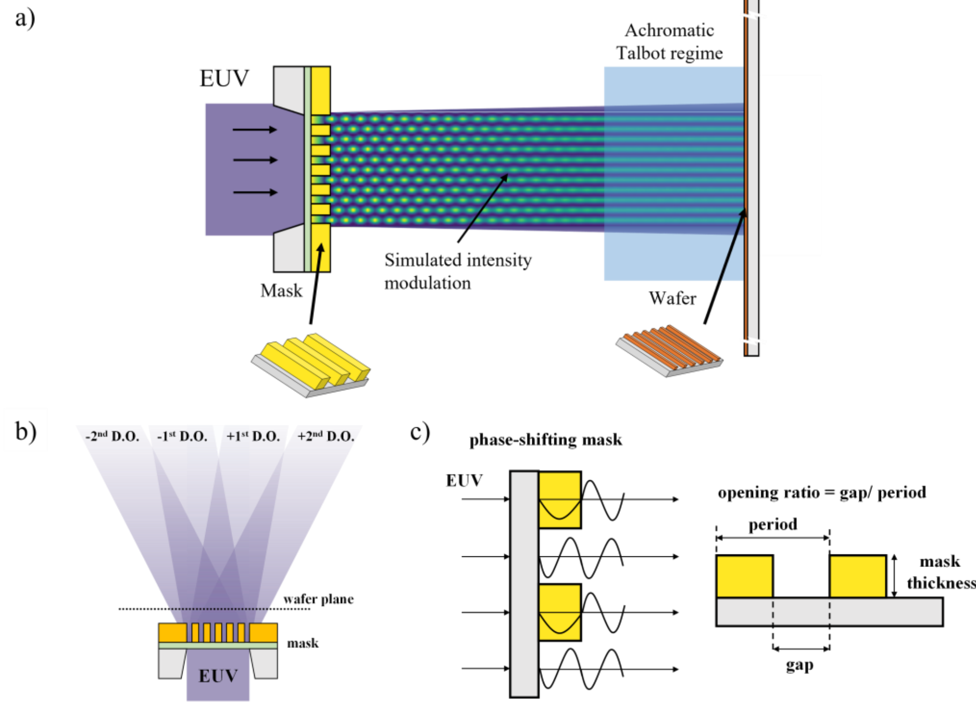


Figure 1. Principle of Talbot lithography with the achromatic Talbot regime indicated by the blue shaded box near the wafer plane. The bright spots represent intensity maxima in the interference pattern. In the achromatic Talbot regime, a pattern demagnification by a factor of two is achievable for line structures (a). Schematic presentation of the different diffraction orders (DO) behind the mask and the position of the wafer plane (b). Schematic representation of the working principle of a phase-shifting transmission mask and the definition of the geometrical parameters (c). Figure adapted from Lüttgenau et al.[18].

For advanced mask materials with small half-pitch sizes close to the exposure wavelength, analytical calculations become unreliable. In such cases, rigorous simulation tools like RCWA are needed to display the complex intensity distribution behind the mask. For optimal aerial image contrast in achromatic Talbot lithography, phase-shifting transmission masks are preferred. These masks are designed to induce a π-phase-shift in the radiation passing through the grating material. When accurately designed, this results in destructive interference between the radiation passing through the grating material and the gaps, causing an efficient suppression of the 0th diffraction order (see Fig. 1c). Consequently, the background intensity of the intensity distribution is reduced, and a higher aerial image contrast is achieved. In this configuration, the interference pattern can be understood as a two-beam interference, where the ±1st diffraction orders dominate. To achieve a complete suppression of the 0th diffraction order, the beam passing through the grating material must experience a precise π-phase-shift, and the intensities of the beams passing through the material and gaps need to be equal, accounting for the unavoidable absorption of EUV radiation in the grating material[25]. Meeting these requirements requires high precision and quality in both mask design and fabrication processes. Additionally, analytical calculations cannot be applied anymore if the structure dimensions are below 100 nm and rigorous simulations need to be applied, as further explored in the subsequent chapter.

3. SIMULATIVE OPTIMIZATION

In the currently used mask design, an electron beam resist is used as the phase-shifting material. This allows a simple and robust fabrication process[25] without the need for an additional etching transfer process. With this approach, the fabrication of 28 nm half-pitch contact hole patterns in a standard EUV photoresist could be demonstrated[16]. The basis of the transmission mask consists of a silicon wafer piece coated with 25 nm silicon nitride and an etched 1 x 1 mm² membrane window in the sample center.

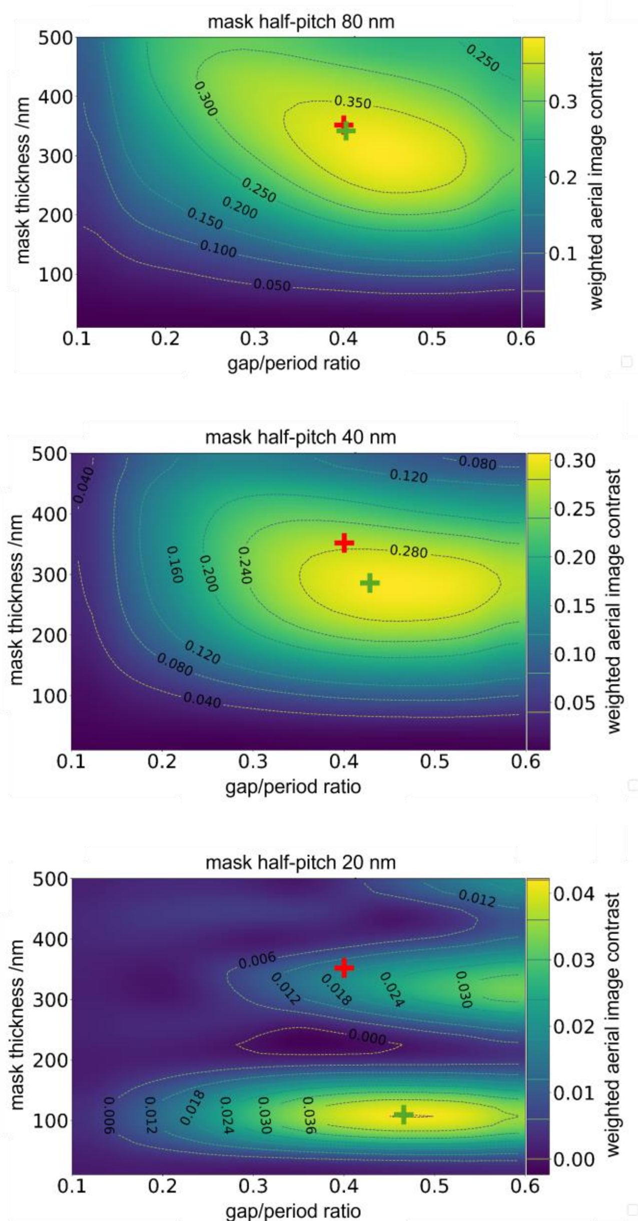


Figure 2. Rigorous simulation analysis of the aerial image contrast for different mask thickness gap/period ratios and mask half-pitches of 80 nm, 40 nm, and 20 nm. The analytical mask thickness is indicated by a red cross. The rigorously optimized mask geometry is indicated by a green cross. It is in good accordance with analytical result for a mask half-pitch of 80 nm but differs strongly for 40 nm and 20 nm half-pitch. Additionally, the achievable transmission weighted aerial image contrast decreases from ~0.38 for 80 nm half-pitch to ~0.04 for 20 nm half-pitch.

In a first step the substrate is coated with the electron beam resist, then patterned by an electron beam at 100 kV acceleration voltage and developed at a temperature of 0 °C. The developer is then replaced by isopropanol as a development stopper and the isopropanol is evaporated on a hot plate. In a last step, the mask is hard baked at 100 °C to increase the radiation stability of the resist that serves as the grating material during the EUV exposure. As a result, the electron beam resist gets more stable and can be used directly as a phase-shifting mask material.

The phase-shift and the resulting aerial image contrast in the wafer plane strongly depend on the mask geometry. Following analytical calculations based on Schnopper et al.[26], a π-phase-shift is achieved for a mask thickness of 350 nm and a gap/period ratio of ~0.4 for a wavelength of 13.5 nm. However, these calculations neglect the thickness of the mask and the effects that arise when the structure approaches the wavelength. To optimize the aerial image contrast for different mask half-pitches, a rigorous simulation is therefore conducted for lines and spaces phase-shifting masks with half-pitches of 20 nm, 40 nm, and 80 nm (see Fig. 2). The transmission weighted aerial image contrast at achromatic Talbot distance is calculated for irradiation with 13.5 nm main wavelength and a relative spectral bandwidth of 4%. Weighting the aerial image contrast by transmission ensures a mask geometry that is usable for an exposure with the limited EUV intensity provided by the DPP EUV source. For each mask half-pitch, the mask thickness is varied from 10 nm to 500 nm and the gap/period ratio is varied from 0.1 to 0.6. The analytically optimized geometry based on Schnopper at al.[26] is marked by a red cross. For a mask half-pitch of 80 nm the rigorously optimized geometry is in good accordance with the analytical result. The maximum weighted aerial image contrast is 0.38 close to the analytical calculation. Additionally, the simulation shows a large area of geometrical parameters that still allow high aerial image contrast. For 40 nm half-pitch, the weighted aerial image contrast decreases to a maximum of 0.31 and the optimal mask geometry deviates stronger from the analytical result since the relation $d >> \lambda$ does not hold true anymore[27]. The highest weighted aerial image contrast is achieved for a mask thickness of ~270 nm and a gap/period ratio of 0.43. This deviation gets even stronger for 20 nm half-pitch, where the highest weighted aerial image contrast is achieved for a mask thickness of ~115 nm and a gap/period ratio of 0.48. However, the weighted aerial image contrast does not exceed 0.05 for this mask half-pitch.

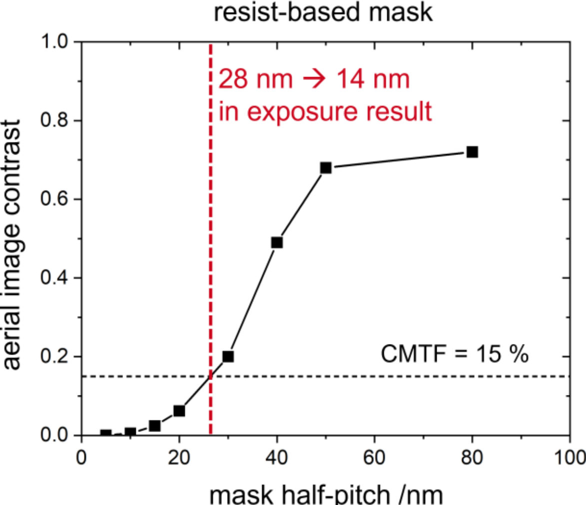


Figure 3. Rigorous simulation results for the aerial image for different mask half-pitches of a resist-based transmission mask. For each mask half-pitch, the mask geometry was optimized using rigorous simulations analogous to Figure 2. For an EUV photoresist with CMTF 15%, the resolution limit is at 28 nm mask half-pitch, which will lead to 14 nm in the exposure result due to the achromatic Talbot effect.

To observe the decrease in aerial image contrast with decreasing mask half-pitch more precisely, the rigorous optimization of the aerial image contrast is performed for mask half-pitches from 5 nm to 80 nm (see Fig. 3). For each mask half-pitch, the highest achievable aerial image contrast is plotted. The aerial image contrast drops from ~75% for 80 nm mask half-pitch to ~50% for 40 nm mask half-pitch to below 10% for 20 nm half-pitch. To evaluate the resulting resolution limit for resist patterning, the critical modulation transfer function (CMTF) of the utilized photoresist needs to be considered. This parameter describes the minimal aerial image contrast required to transfer the aerial image into a

resist topography for a resist with given sensitivity and contrast. If an industrial EUV photoresist with CMTF of 15% is used, a mask pattern with a half-pitch of 28 nm would still lead to a sufficient aerial image contrast for pattern transfer. Due to the achromatic Talbot effect, the realized pattern on the wafer would be demagnified to ~14 nm half-pitch. However, it is evident that the resist-based mask design has limitations towards the target resolution of sub-10 nm half-pitch, which would be necessary for resist qualification for future technology nodes. Therefore, alternative mask materials are investigated.

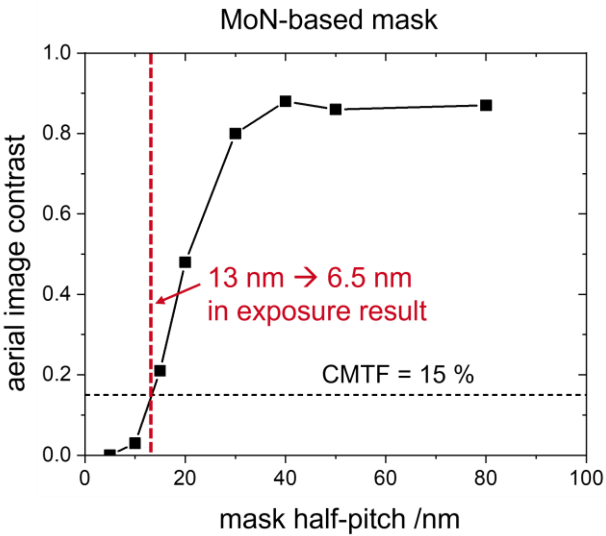


Figure 4. Rigorous simulation results for the aerial image for different mask half-pitches of a MoN-based transmission mask. For each mask half-pitch, the mask geometry was optimized using rigorous simulations analogous to Figure 2. For an EUV photoresist with CMTF 15%, the resolution limit is at 13 nm mask half-pitch, which will lead to 6.5 nm in the exposure result due to the achromatic Talbot effect.

In earlier investigations, a MoN-based mask was shown to exhibit highest aerial image contrast[29] and good patterning performance in etching processes[30]. Therefore, the rigorous analysis shown for the resist-based mask is repeated for a MoN-based lines and spaces transmissions mask with mask half-pitches from 5 nm to 80 nm (see Fig. 4). The mask geometry was optimized for each individual mask half-pitch to exhibit the highest transmission weighted aerial image contrast. The aerial image contrast is above 80% for mask half-pitches down to 30 nm and then decreases strongly. If an average EUV photoresist with CMTF of 15% is assumed, a mask pattern with a half-pitch of 13 nm would still have a sufficient aerial image contrast for pattern transfer. Due to the achromatic Talbot effect, the realized pattern on the wafer would be demagnified to ~6.5 nm half-pitch. For a mask half-pitch of 20 nm, that leads to a resolution of 10 nm half-pitch in the exposure result, the aerial image contrast is still ~50%. The MoN-based mask would therefore suffice to reach the targeted sub-10 nm resolution for resist qualification for future technology nodes.

4. CONCLUSION AND OUTLOOK

In this contribution, the authors present the simulative investigation of factors that influence the resolution limit of Talbot lithography with compact exposure tools. The design of optimized phase-shifting transmission masks for high-resolution nanopatterning with a compact EUV exposure tool is demonstrated. Several influencing factors on the achievable resolution are identified and characterized. The applied (achromatic) Talbot lithography has proven to be a valid interference scheme for partially coherent radiation with a theoretical resolution limit below 10 nm. To move further towards this limit, metallic phase-shifting masks are considered and characterized by rigorous simulations. The results demonstrate that with MoN-based phase shifting masks resolvable structure sizes down to 6.5 nm can be achieved.

ACKNOWLEDGEMENTS

This work was realized by cooperation activities in the frame of the Jülich Aachen Research Alliance for Fundamentals of Future Information Technology (JARA-FIT). The presented activities received funding by the German Research Foundation (DFG, support code: DA 990/4-1).

REFERENCES

[1] Simone, D. de, Vesters, Y. and Vandenberghe, G., "Photoresists in extreme ultraviolet lithography (EUVL)," Advanced Optical Technologies 6(3-4), (2017).

[2] Naulleau, P., Anderson, C., Chao, W., Bhattarai, S., Neureuther, A., Cummings, K., Jen, S.-H., Neisser, M. and Thomas, B., "EUV Resists: Pushing to the Extreme," J. Photopol. Sci. Technol. 27(6), 725–730 (2014).

[3] Banerjee, A., Maity, S. and Mastrangelo, C. H., "Nanostructures for Biosensing, with a Brief Overview on Cancer Detection, IoT, and the Role of Machine Learning in Smart Biosensors," Sensors (Basel, Switzerland) 21(4), (2021).

[4] Tempeler, J., Danylyuk, S., Brose, S., Loosen, P. and Juschkin, L., "Structural properties of templated Ge quantum dot arrays: impact of growth and pre-pattern parameters," Nanotechnology 29(27), 275601 (2018).

[5] Bi, K., Wang, Q., Xu, J., Chen, L., Lan, C. and Lei, M., "All-Dielectric Metamaterial Fabrication Techniques," Adv. Optical Mater. 9(1), 2001474 (2021).

[6] Goethals, A.-M., Foubert, P., Hosokawa, K., van Roey, F., Niroomand, A., van den Heuvel, D. and Pollentier, I., "EUV Resist Process Performance Investigations on the NXE3100 Full Field Scanner," J. Photopol. Sci. Technol. 25(5), 559–567 (2012).

[7] Kazazis, D., Tseng, L., and Ekinci, Y., "Improving the resolution and throughput of achromatic Talbot lithography," J. Vac. Sci. Technol. B 36, 06J501 (2018).

[8] Watanabe, T., and Harada, T., "Research activities of extreme ultraviolet lithography at the university of Hyogo," Sync. Rad. News 32(4), 28-35 (2019).

[9] Fan, D. and Ekinci, Y., "Photolithography reaches 6 nm half-pitch using EUV light," Proc. SPIE 9776, 97761V (2016).

[10] Tasdemir, Z., Vockenhuber, M., Mochi, I., Olvera, K. G., and Meeuwissen, M., "Chemically amplified EUV resists approaching 11nm half-pitch," Proc. of SPIE 10583, 10583W (2018).

[11] Yang, S., Zhao, J., Wang, L., Zhu, F., Xue, C., and Liu, H., "Influence of symmetry and duty cycles on the pattern generation in achromatic Talbot lithography," J. Vac. Sci. Technol. B 35, 021601 (2017).

[12] Isoyan, A., Jiang, F., Cheng, Y., C., and Cerrina, F., "Talbot lithography: Self-imaging of complex structures," J. Vac. Sci. Technol. B 27(6), (2009).

[13] Sato, T., Yamada, A., Suto, T., Inanami, R., Matsuki, K., Ito, S., "Printability of defects in Talbot lithography," Microelectron. Eng. 143(21-24), (2015).

[14] Lohmann, A., Thomas, J., "Making an array illuminator based on the Talbot effect", Appl. Opt. 29, 4337-4340 (1990).

[15] Bergmann, K., Danylyuk, S. V., and Juschkin, L., "Optimization of a gas discharge plasma source for extreme ultraviolet interference lithography at a wavelength of 11 nm," J. Appl. Phys. 106, 073309 (2009).

[16] Brose, S., Tempeler, J., Danylyuk, S., Loosen, P. and Juschkin, L., "Achromatic Talbot lithography with partially coherent extreme ultraviolet radiation: process window analysis," J. Micro/Nanolith. MEMS MOEMS 15(4), 043502 (2016).

[17] Solak, H. H. and Ekinci, Y., "Achromatic spatial frequency multiplication: A method for production of nanometer-scale periodic structures," J. Vac. Sci. Technol. B 23(6), 2705-2710 (2005).

[18] Lüttgenau, B., Brose, S., Danylyuk, S., Stollenwerk, J., Loosen, P., Holly, C., „Design and realization of an in-lab EUV dual beamline for industrial and scientific applications," Proc. SPIE 11854, 1185419 (2021).

[19] Lüttgenau, B., Brose, S., Danylyuk, S., Stollenwerk, J. and Holly, C., „Toward the resolution limit for Talbot lithography with compact EUV exposure tools," Proc. SPIE 12292, 1229208 (2022).

[20] Talbot, H., "Facts relating to optical science," Phil. Mag. 9(56), (1836).

[21] Lord Rayleigh, F., "On copying diffraction gratings and on some phenomenon connected therewith," Phil. Mag. 11(67), (1881).

[22] Vieker, J., and Bergmann, K., "Influence of the electrode wear on the EUV generation of a discharge based extreme ultraviolet light source," J. Phys. D: Appl. Phys. 50 (345601), 8pp (2017).
[23] Brose, S., Danylyuk, S., Bahrenberg, L., Lebert, R., Stollenwerk, J., Loosen, P., and Juschkin, L., "EUV-LET 2.0: a compact exposure tool for industrial research at a wavelength of 13.5nm", Proc. of SPIE 10957, 109571K (2019).
[24] Lüttgenau, B., Panitzek, D., Danylyuk, S., Brose, S., Stollenwerk, J., Loosen, P. and Holly, C., „Design of an efficient illuminator for partially coherent sources in the extreme ultraviolet," Appl. Opt. 61(11), 3026-3033 (2022).
[25] Brose, S., Danylyuk, S., Bahrenberg, L., Lebert, R., Loosen, P., Juschkin, L., "Optimized phase-shifting masks for high-resolution resist patterning by interference lithography," Proc. SPIE 10450, 104502A (2017).
[26] Schnopper, H.W., Van Speybroeck, L. P., Delvaille, J. P., Epstein, A., Källne, E., Bachrach, R. Z., Dijkstra, J. and Lantward, L., "Diffraction grating transmission efficiencies for XUV and soft x-rays," Appl. Opt. 16, 1088 (1977).
[27] Gremaux, D., Gallagher, N., "Limits of scalar diffraction theory for conducting gratings," Appl. Opt. 32(11), 1948pp (1993).
[28] Plummer, J. D., Deal, M. D., Griffin, P. B., „Silicon VLSI Technology: Fundamentals, Practice, and Modeling", 1st Edition, Prentice Hall Electronics and VLSI Series (2000).
[29] Lüttgenau, B., Brose, S., Danylyuk, S., Stollenwerk, J. and Loosen, P., „Novel high-contrast phase-shifting masks for EUV interference lithography," Proc. SPIE 11323, 113231Q (2020).
[30] Wang, L., Kirk, E., Wäckerlin, C., Schneider, C. W., Hojeij, M., Gobrecht, J. and Ekinci, Y., „Nearly amorphous Mo-N gratings for ultimate resolution extreme ultraviolet interference lithography", Nanotech. 25 (23), (2014).

Multi-Beam Mask Writing opens up New Fields of Application

Mathias Tomandl, Christoph Spengler, Christof Klein, Hans Loeschner, Elmar Platzgummer

IMS Nanofabrication GmbH, Wolfholzgasse 20-24, 2345 Brunn am Gebirge, Austria

imsoffice@ims.co.at; phone: +43 2236 373 10; www.ims.co.at

ABSTRACT

Multi-Beam Mask Writers (MBMW) from IMS Nanofabrication disrupted the mask writing technology in the past decade by offering this technology to the industry with a range of benefits over the preceding variable shaped beam (VSB) technology. The MBMW-101 enabled write times independent of pattern complexity, usage of low sensitivity resists at high throughput, and providing superior resolution and critical dimension uniformity (CDU) capabilities. With these benefits, the technology enabled high volume extreme ultra violet (EUV) mask manufacturing for logic and memory applications to the industry.

The MBMW-201 is today's standard technology for leading edge photo masks patterning and used in the most advanced mask shops around the globe. Its superior robustness and powerful write modes allow an unprecedented writing efficiency and resolution capability.

Now IMS broadens the spectrum of application for this technology and releases two new products:

The MBMW-100 Flex is a versatile mask writer to open multi-beam benefits to mature and intermediate nodes application at high throughput and beneficial total cost of ownership, targeting nodes from 32nm down to 10nm.

The MBMW-301: the next generation leading edge mask writer for ultra-low sensitivity resists with resolution and CDU capabilities meeting EUV high numerical aperture (NA) requirements targeting nodes down to 2nm and beyond.

This article will delve into the transformational journey of multi-beam mask writing, from its early beginnings to its current status as the cornerstone of EUV mask production and provide an overview on the two new models with performance data and lithography results.

Keywords: Mask writer, e-beam, multi-beam, curvilinear mask, EUV mask

1. INTRODUCTION

As we mark the tenth anniversary of installing the world's first Multi-Beam Mask Writer at IMS Nanofabrication **[1]**, it is fitting to reflect on the revolutionary impact it has had on the semiconductor industry. In the past decade, the landscape of mask writing has been irrevocably transformed. Through the adoption of this technology, setting new standards and enabling of high-volume production of EUV masks.

In 2013, IMS Nanofabrication installed the MBMW Alpha tool - the first e-beam multi-beam mask writer - that aimed to overcome the limitations of the then prevalent VSB writers **[2]**.

The VSB technology was struggling to keep pace as write times were projected to grow exponentially mainly due to two reasons. First, higher dose requirements of new resists and second, the increasing complexity of mask pattern due to extensive use of optical proximity correction (OPC). The utilization of curvilinear features for inverse lithography technology (ILT) seemed unreachable with VSB technology.

These issues were all addressed in 2016, when IMS Nanofabrication launched the High-Volume Manufacturing (HVM) version of the multi-beam mask writer – the MBMW-101 **[3]**, heralding a new era in mask writing technology. This significant milestone in the journey of mask making was followed in 2019, when IMS enhanced this technology with the second generation of multi-beam mask writer that further increased throughput and improved resolution further. This second-generation writer, the MBMW-201, is the most successful multi-beam mask writer and being used for EUV mask production and is used from 7nm nodes down to 2nm node research applications today.

38th European Mask and Lithography Conference (EMLC 2023), edited by
Uwe F.W. Behringer, Jo Finders, Proc. of SPIE Vol. 12802, 1280203
© 2023 SPIE · 0277-786X · doi: 10.1117/12.2678532

2. APPROACHES TO MULTI-BEAM TECHNOLOGY TODAY

2.1 Multi-beam mask writing, the fundamental concept

While many different designs to multi-beam writing have been developed in the early 2000s **[4]**, the design approach of IMS Nanofabrication was the only one succeeding being widely adopted by the industry since 2016 **[5]**. In 2022, NuFlare Technology (NFT) also released a multi-beam mask writer to the industry. This system is based on the same fundamental concept, illustrated in Figure 1.

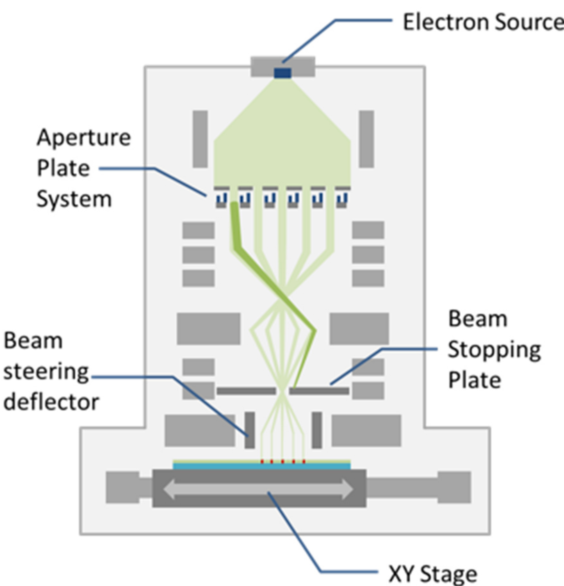


Figure 1: Fundamental concept of multi-beam mask writing.

A single electron source generates a single beam of electrons. This beam is then collimated to a wide parallel beam in the condenser lens system. This wide parallel beam is passing through the aperture plate system (APS), that consists of an aperture plate (APP) and a blanking chip (BLC).

The APP, divides the beam into more than 260 thousand individual beamlets that are then passing through the BLC. Using CMOS electronics combined with MEMS formed electrodes in the BLC, each individual beamlet can be deflected to hit the beam stopping plate in the lower section of the column. This allows to effectively switch individual beams on or off, since if deflected, they do not reach down to the substrate anymore. By controlling the time of each individual beam's deflection, different gray-levels, e.g., dose-control, can be realized.

Following the beam path through the projection optics, the system performs a 200x demagnification of the beam array onto the substrate. In tandem, the XY-Stage moves the reticle during exposure, while the beam steering deflector positions the beam array to precisely raster the designated exposure area. Effectively this approach realizes a massively parallelized rastering spot-beam writer.

2.2 Different approaches to Acceleration optics and implications

Both systems available today are using the same design approach, described in Section 2. Furthermore, even the detailed configuration of most relevant aspects is fully identical. This includes, acceleration voltage, demagnification factor, aperture sizes, beam pitch, number of beams, size of beam array and consequently also the size of image field down at the substrate. However, despite the many similarities, there are some significant differences that are crucial for the performance of the writer and even more for the scalability of the technology to future requirements. These differences are outlined in Figure 2.

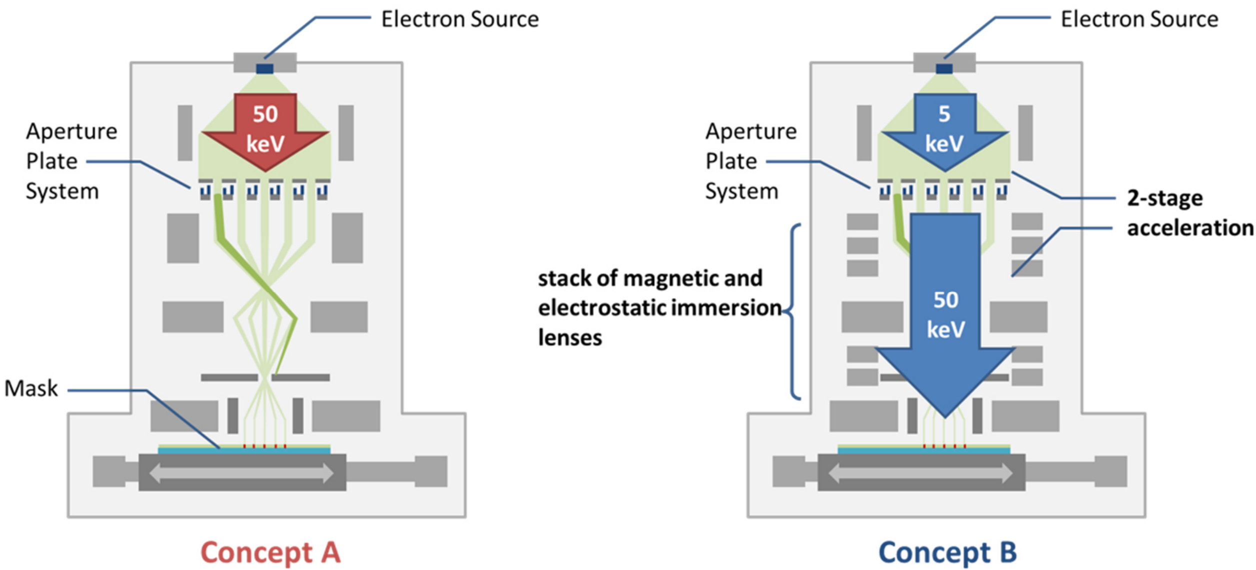


Figure 2: Two concepts to optical system and beam acceleration method used in today's multi-beam mask writers. While Concept A is the simpler design, Concept B on the other hand, while being more complex has clear advantages realized by 2-stage acceleration system and a stack of magnetic and electrostatic immersion lenses.

2.3 Superior lifetime of critical components due to 2-stage acceleration system

While Concept A as followed by NFT allows a simple design in which the beam is fully accelerated to 50keV above the Aperture Plate System (APS), the design invented by IMS Nanofabrication, Concept B, utilizes a 2-stage acceleration where, the beam is initially accelerated to 5keV, and the acceleration to full energy is carried out within the projection optics.

While at first glance, the simpler approach of Concept A seems appealing and beneficial due to simplicity as claimed recently **[6]**, Concept B has significant advantages - with regards to longevity of APS lifetime and general optics performance.

The first benefit is the absence of x-ray radiation at the aperture plate system in Concept B. Here, the energy level remains adequately low, consequently there is no ionizing radiation created at this stage that could harm the sensitive CMOS electronics of the blanking chip.

This allows Concept B to perform without any risk for electronics damage for many years. The fleet wide average lifetime is above 2.5 years, following a predictive maintenance approach individual units are already proven to have exceeded 5.5 years without any issues.

While earlier tools following Concept A could not deal with the radiation damage of the BLC electronics, recently NFT reported to achieve lifetimes of one year using mitigation strategies that limit the impact by radiation **[6]**. Reducing the amount of ionizing radiation seems to be managed, but any remaining radiation can bare a risk to electronics damage leading to slow but continuous APS degradation over time, and thus impact write-performance quality such as CDU and pattern fidelity.

Finally, the more robust design following Concept B allows longer APS lifetimes and result in lower maintenance needs and potential cost savings, by extending the components productive lifespan.

Besides the APS lifetime, there is a second advantage in Concept B, that lies in the control of the beam acceleration within the projection system using a strategically placed combination of electrostatic immersion lenses and magnetic lenses enabling reduced aberrations and beam distortion related blur, and thus superior imaging quality.

2.4 Superior aberration and distortion related blur due to stack of magnetic and electrostatic immersion lenses

Different to light optics, in charged particle systems the resolution is not dominantly dependent on the illuminating wavelength. In charged particle systems the Coulomb blur becomes the dominant factor to limit resolution when increasing the beam intensity required for high-throughput productivity. Accordingly, all other contributions to blur need to be reduced to the very minimum in order to achieve minimum half pitch and assist features.

The following three contributors to blur are defining the resolution performance of multi-beam technology: First, spherical aberration blur, second, distortion-related positional blur, and third Coulomb interaction or space charge blur as shown in Figure 3. Each of these contributors needs to be minimized in best performance multi-beam mask writing.

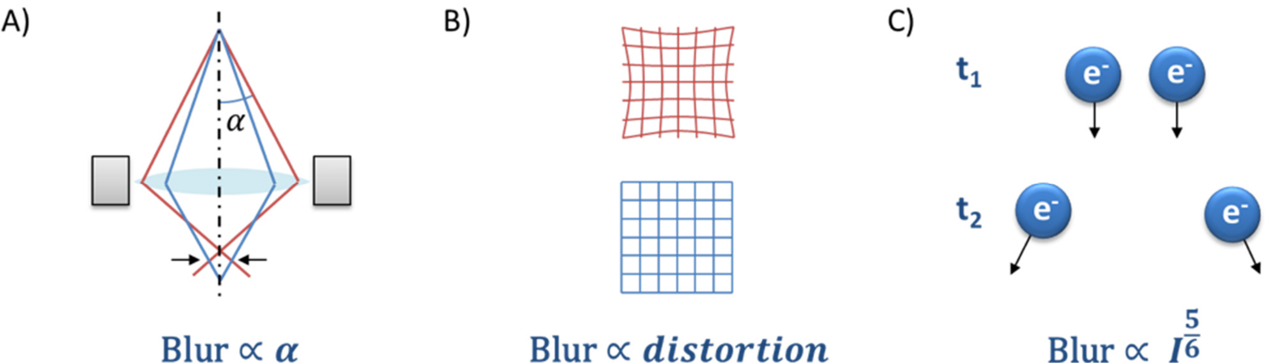


Figure 3 Three important components of blur in multi-beam mask writing: A) spherical aberration blur, B) distortion-related positional blur, and C) space charge induced blur by beam current I.

The two-stage acceleration system allows utilizing combinations of electrostatic immersion and magnetic lenses, presents distinct advantages in managing these blur types. It enables effective aberration correction and allows to minimize both aberration blur and beam field distortion at the same time.

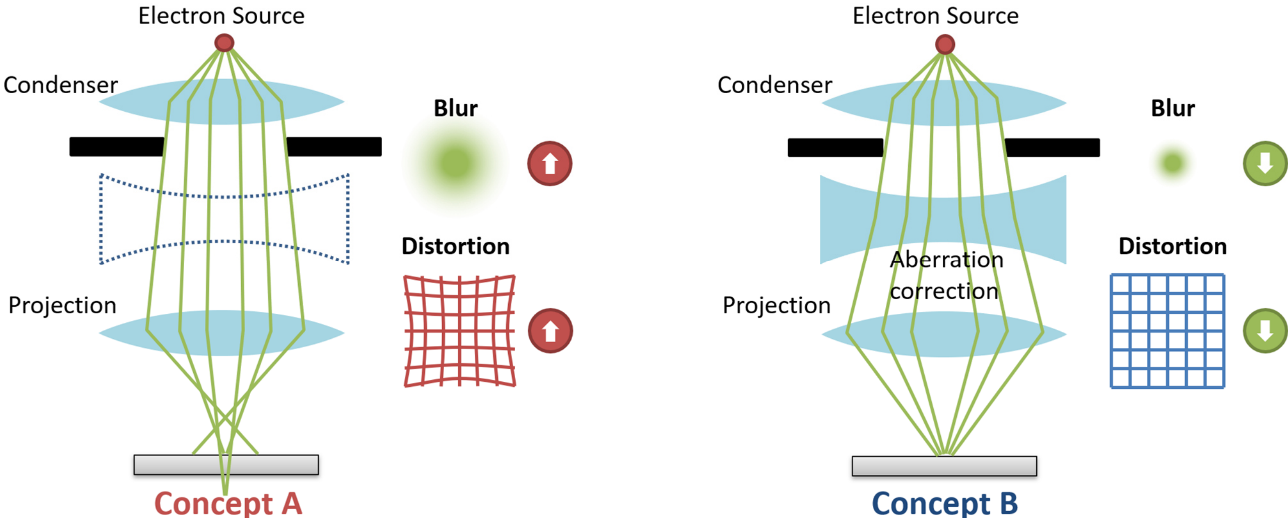


Figure 4 The utilization of divergent electrostatic lenses allows to realize a concept for aberration correction, commonly known in light optics, in the charged particle domain. By combining concave and convex lenses this approach allows to compensate for spherical aberrations and minimize beam field distortion, two contributors for increasing blur and consequently limiting resolution.

Spherical aberrations, meaning a different focal length for paths close to the optical axis compared to paths further away from the axis, are a widely known phenomenon in optical systems. Magnetic lenses are an easy way to realize strong optical lenses and are typically used in many e-beam projection systems. Since they do not require any high voltage, but only a current generating the magnetic field, they are a simple and easy to control lens solution. However, magnetic lenses due to the nature of how a magnetic field acts on a charged particle, always act as a focusing lens, never as a divergent lens. In consequence, they alone cannot realize good aberration control with the method described in Figure 4.

While the beam distortion in most optics applications only causes a displacement, in multi-beam applications beam field distortion is a significant contribution to the total blur as illustrated in Figure 5. As multi-beam technology makes use of massive averaging to limit imperfections of individual beamlets. These averaging overlaps different regions of the image field to write a single structure on the mask. By that different regions of the image field are locally combined on the mask. A larger residual distortion, the individual local displacements result in an additional blur component.

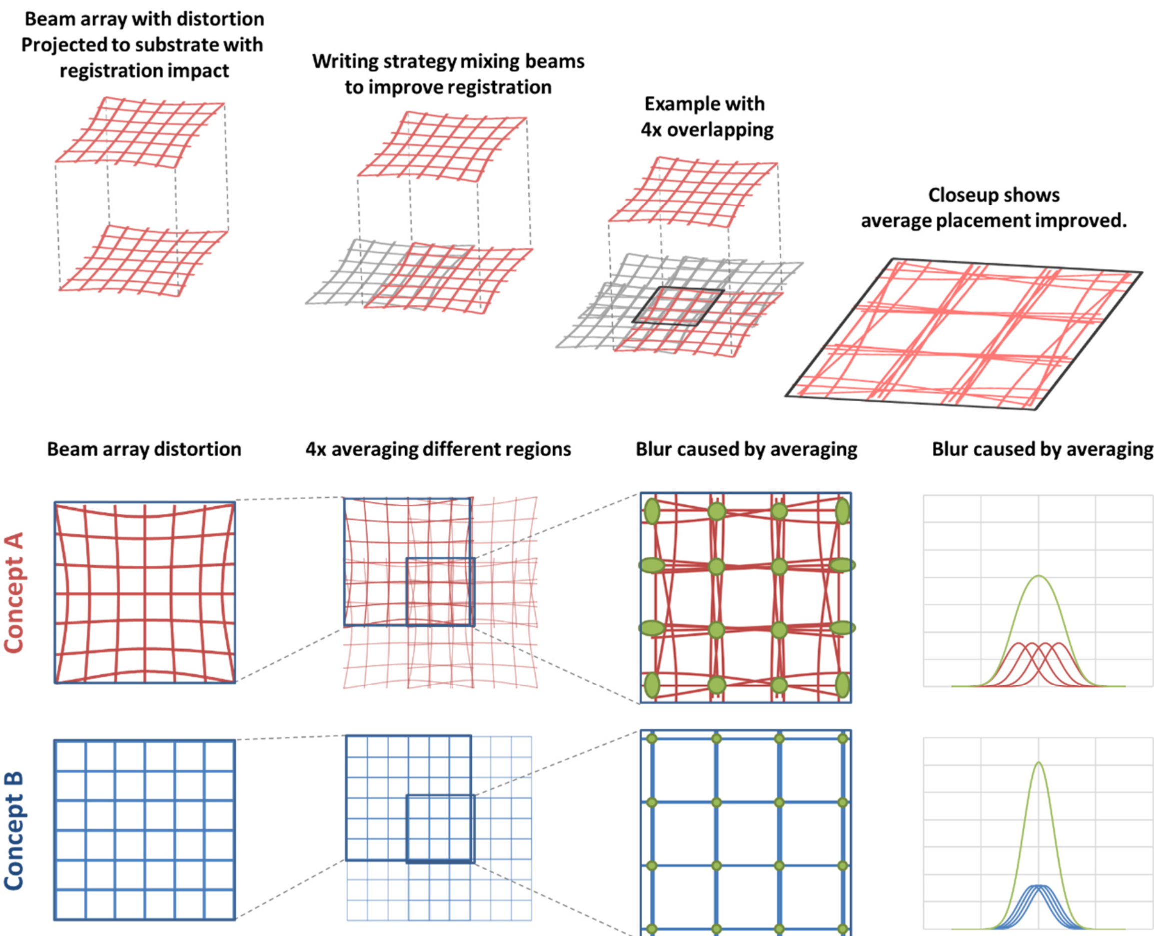


Figure 5 Examining the effects of distortion of the beam array on the writing quality. On the top, representing Concept A, in red we see a higher distortion. When we are overlapping multiple regions of the beam array to write a structure, as it is done in multi-beam writing processes by e.g. multi-passing or overlapping shots, this displacement between individual contributions, results in a broader blur distribution as illustrated with the envelope in green. Conversely, when the distortion is minimized, this effect is substantially reduced (Concept B).

Upon inspecting overlapping passes closely, we find that registration remains surprisingly unaffected due to the center of the averaging. However, other parameters such as blur, dose slope, CDU, pattern-fidelity, resolution, and process window suffer from this higher distortion.

Controlling aberration and distortion blur together is the general challenge in multi-beam writing, together they define the optics-baseline. On top we need to add the Coulomb or space-charge related blur that is dependent on the current through the projection optics. It becomes evident that having a lower optics-baseline allows to accept a higher current in the projection system without compromising resolution as illustrated in Figure 6. Resolution, on the other hand, is key to resolve small feature sizes and to achieve small half-pitch dimensions as needed for the next generation EUV and high-NA EUV masks.

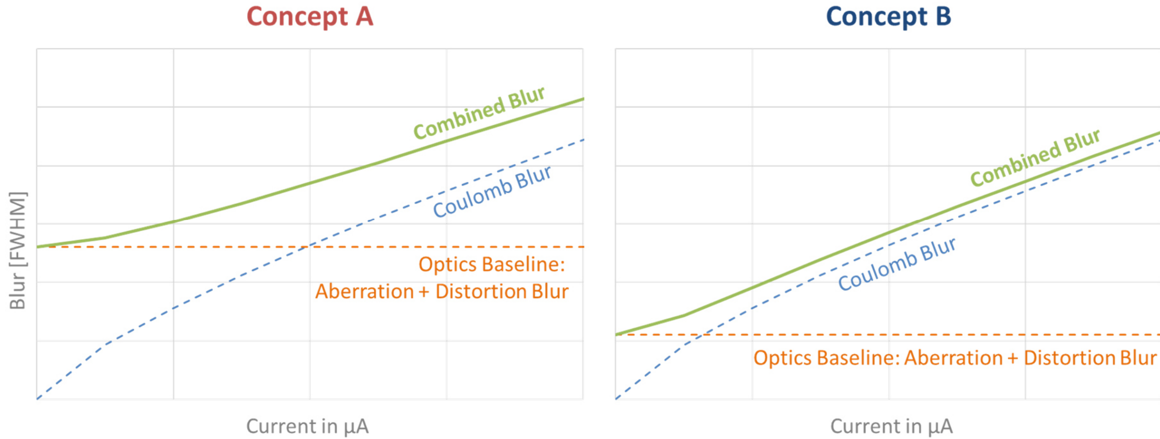


Figure 6 Comparison on blur performance with different concepts. The aberration correction and low distortion blur allow Concept B to operate at higher beam current to meet the same blur value.

In summary, Concept B's two-stage acceleration system ensures maximum APS health due to the absence of radiation, and also guarantees the lowest possible aberrations and distortion, thereby proving superior for resolution and throughput.

3. EXTENDING MULTI-BEAM MASK WRITING TO NEW FIELDS

Its low beam field distortion and aberration, allowing high resolution patterning at high throughput made MBMW-201 the most successful multi-beam mask writer so far. This year, IMS released two new product lines to extend the multi-beam technology to new fields of application, Figure 7.

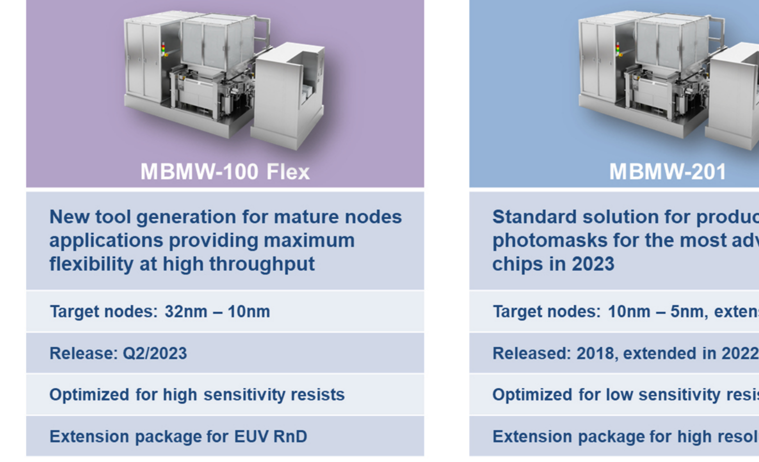


Figure 7 Three products provided by IMS Nanofabrication in 2023: The MBMW-100 Flex for mature nodes, the MBMW-201 for most advanced nodes in 2023, the MBMW-301 for leading edge nodes.

3.1 MBMW-100 Flex enabling multi-beam for mature and advanced nodes applications

There is a significant growth on chips in the mature nodes, mainly driven by the automotive industry, where highly efficient mask writers are needed. Here, MBMW-100 Flex can deliver many advantages from multi-beam technology to applications to mature and intermediate nodes, allowing to avoid running into the limitations of the VSB technology. Based on the well-established MBMW-201 platform, MBMW-100 Flex was optimized for mature and intermediates node application requirements providing three different write modes to be applied for 32nm node masks up to 10nm node masks at write times down to 3h to 5h (104 x 132mm²) dependent on the write mode, independent of pattern complexity.

This tool is designed for use with a wide range of different resists, from high sensitivity resists starting at 10µC/cm² as well medium and low sensitivity resists up to around 50µC/cm². The newly developed eSource system and write modes were specifically designed to meet these requirements. Besides the new eSource, the Flex tool mostly builds on hardware of the MBMW-201, with its demonstrated performance and durability over many years.

The MBMW-100 Flex was released in the second quarter 2023, and the first unit was successfully installed at customer.

3.2 MBMW-301 redefining leading edge mask writer

The MBMW-301 is a revolutionary step in mask writer capabilities and comes the next generation APS, a substantially smaller spot size and a significantly faster data path. This technological advancement is required to meet the requirements for 2nm node and below and High-NA EUV lithography **[7]**, with regards to placement, CDU and resolution for half pitch and assist features below 20nm, while keeping the write time comparable to the predecessor.

With that, the MBMW-301 represents a revolutionary step for mask writing technology, embodying significant advancements over any other mask writer. These enhancements promise to redefine industry benchmarks and ensure unmatched precision and efficiency in mask writing processes.

The MBMW-301 is utilizing an evolved e-source, which increases the current on the substrate by a factor of two compared to the MBMW-201. This higher current in combination with the efficient writing strategy is essential to achieve the write time requirements on ultra-low sensitivity resists. The 3rd generation APS controls close to 600,000 individual beamlets at a three times higher data rate compared to previous models. The tool's air-bearing stage from JEOL allows writing at double the stage speed at the required accuracy.

The MBMW-301 Alpha tool currently in operation and Beta tools are being shipped to clients. High Volume Manufacturing (HVM) units will be ready for deployment in the first quarter of 2024.

REFERENCES

[1] Platzgummer, E. (2010, April). Maskless lithography and nanopatterning with electron and ion multibeam projection. In *Alternative Lithographic Technologies II* (Vol. 7637, pp. 17-28). SPIE.

[2] Lee, S. H., Choi, J., Min, S. J., Kim, H. B., Kim, B. G., Woo, S. G., & Cho, H. K. (2010, May). Mask writing time explosion and its effect on CD control in e-beam lithography. In *Photomask and Next-Generation Lithography Mask Technology XVII* (Vol. 7748, pp. 165-174). SPIE.

[3] Klein, C., & Platzgummer, E. (2016, October). MBMW-101: World's 1st high-throughput multi-beam mask writer. In *Photomask Technology 2016* (Vol. 9985, p. 998505). SPIE.

[4] Platzgummer, E., Klein, C., & Loeschner, H. (2013). Electron multibeam technology for mask and wafer writing at 0.1 nm address grid. *Journal of Micro/Nanolithography, MEMS, and MOEMS*, *12*(3), 031108-031108.

[5] Chandramouli, M., Liu, B., Juday, R. K., Shamoun, B., Korobko, Y., Sowers, A. T., ... & Abboud, F. E. (2020, March). Development and deployment of advanced multi-beam mask writer. In *Novel Patterning Technologies for Semiconductors, MEMS/NEMS and MOEMS 2020* (Vol. 11324, pp. 60-68). SPIE.

[6] Yasuda, J., Nomura, H., Matsumoto, H., Nakayamada, N., & Yamashita, H. (2023). Recent progress and future of electron multi-beam mask writer. *Japanese Journal of Applied Physics*.

[7] Shamoun, B., Chandramouli, M., Liu, B., Juday, R. K., Bucay, I., Sowers, A. T., & Abboud, F. E. (2021). Multi-beam mask writer in EUV era: challenges and opportunities. *Novel Patterning Technologies 2021*, *11610*, 64-82.

Automatic laser writing parameters setting for mature masks production

Giorgio Borzini[a], Andrea Galbiati[a], Luca Sartelli[a], Hidemichi Imai[a],
[a] DNP Photomask Europe S.p.A. (Italy)
Via C. Olivetti 2/A, 20864 Agrate Brianza, MB, Italy
Phone: +39-347-5989321; FAX: +39-039-65493-100; e-mail: borzini-g@mail.dnp.co.jp

Abstract

An automated system has been developed to calculate dose and bias XY, based on the relevant characteristics of the single mask layout, material parameters, general process inputs from the production SPC and equipment-related input, such as the QC test masks..

Mask input parameters are Exposure Bias, feature size, local exposure ratio (1500µm x 1500µm square), micro exposure ratio (25µm x 25µm square) and coordinates of its CD measurement points. Feature size is used to centre the writing process based on CD linearity. The exposure ratio is used to correct over/under exposure related to process loading effects. Micro exposure ratio (MER) is also used to find the best exposure condition to minimize the effects of Iso-Dense bias. To calculate the exposure Dose, the global expected CD Error (CDE) is fed to the CDE vs. dose curve, obtained from QC test masks for different feature types (Hole, Iso Line, Iso Space, Dense Line&Space). Exposure Biases are set to work around iso-focal dose or for other needs. The calculated dose is used to determine the correction for X/Y bias, which depends on the dose itself.

Based on the model we developed an automated system (called ADOC – ALTA Dose Calculation) which automatically calculates writing parameters and sets the job request directly on the writing tool. Thanks to this system we increased writing parameters' reliability and obtained a CD yield close to 100%. This system greatly simplifies the activity of writing operators, with fewer manual inputs: the writing step requires only blank loading and running ADOC by simply providing the unique mask code identification.

In this paper, we discuss the writing parameters definition model, the application software and its interface with the production line's database and with the i-line writing tool. We conclude with the achievements we got in terms of yield and process robustness.

38th European Mask and Lithography Conference (EMLC 2023), edited by
Uwe F.W. Behringer, Jo Finders, Proc. of SPIE Vol. 12802, 1280204
© 2023 SPIE · 0277-786X · doi: 10.1117/12.2675407

Introduction

The recent semiconductor situation requires a high volume of mature products with very short delivery times. In particular, the lower technological nodes could be affected by a significant increase in demand against an almost zero increase in production capacity[1], placing further pressure on this type of products.

NODE SIZE DEMAND/CAPACITY GROWTH

Semiconductor demand and capacity forecast

2022–2026

NODE SIZE	2022	2026	CAGR* Demand	CAGR* Capacity	TREND**
> 250 nm			+2.4%	+0.0%	
> 180 nm			+3.3%	-0.7%	
> 130 nm			+3.4%	-0.7%	
> 90 nm			+3.1%	-1.4%	
> 45 nm			-1.3%	+3.7%	
> 28 nm			+7.3%	+3.1%	
> 16 nm			-10.7%	-7.0%	
> 10 nm			+8.4%	+5.7%	
≤ 10 nm			+18.3%	+16.6%	

No shortage · Minor shortage · Medium shortage · Major shortage

* Compound annual growth rate 2022–2026 |
** Based on development 2022–2026, subject to creation of further new capacities

Source: omdia, Q3/2022 Silicon Market Trackers, Porsche Consulting Analysis

Source: omdia Silicon Market Trackers, Porsche Consulting Analysis

The increase in cost, related to mask making (blanks, pellicles, tools maintenance, electrical power), significantly reduced the profit margin. It is then extremely important to push the productivity related to mature products to the limit, reducing any kind of loss, either technical or human-related, and decreasing cycle time.

In this article the design of an automated system for the ALTA laser writing tools is described, ADOC – ALTA Dose Calculation, which automatically calculates writing parameters and sets the job request directly on the machine.

The system, based on the characterization of the process provided by test mask analysis and fine-tuned through SPC results, takes as input information on some relevant features of the Critical Dimension (hereinafter called CD) measurement points of the mask to be written and provides dose and XY bias to be used for exposure.

Materials and Method

The system is constituted by an Oracle Script that calculates the dose and XY bias of the masks, and by a series of Excel Workbooks, one for every tool, that provides data analysis and optimization of the parameters used in the calculation.

The exposure dose of a mask is calculated by averaging the doses calculated for each measurement point of the mask, according to the following features of the CD site: dose setting bias (characteristic of the whole mask), target, proximity of the structures, local density of the pattern and position on the mask. The XY bias depends on the dose and the tool conditions.

The parameters used for the calculation come from the evaluation of a test mask, designed to provide the process characterization of the main pattern features of a mask, and from feedback coming from production results.

The test used for process characterization is the EXPOFOCUS[2], a native exposure test embedded in the ALTA laser writing systems object of the article, in which various pattern types (iso – Hole, Iso-Space, Iso-Line and Line & Space) are written at different targets, doses and focus level.

The fundament of the calculation system is the CD Error vs Dose curve, calculated for every pattern type, and it is obtained by averaging the results at the best focus for each target (600, 800, 1000, 1200, 1400, 1600, 3200nm) at every dose (65.1, 71.6, 78.7, 86.5, 95, 104.4, 114.7, 126.1, 138.5mj).

The curve is parametrized as dose function of the CDE with a 3rd degree polynomial function.

eq 1. $DOSE = a * CDE^{3} + b * CDE^{2} + c * CDE + d$

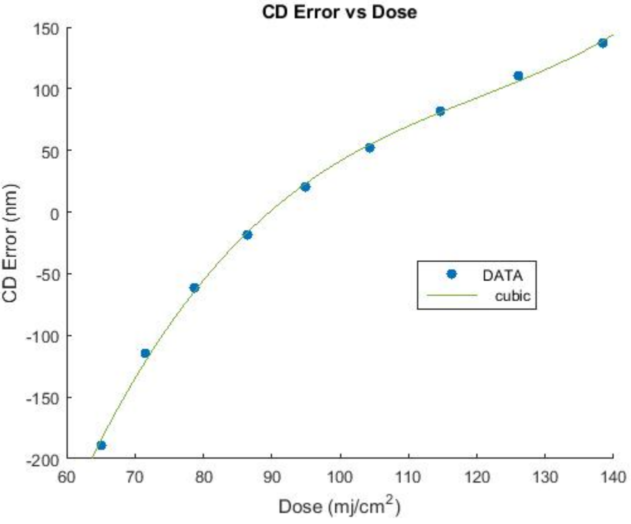


Fig 1. *CDE vs dose and its cubic fitting*

Since the process is dependent on proximity, for every process line, the dose is calculated for isolated and dense structures and the resulting dose for a given point is given by a linear combination of the two doses. To evaluate the proximity is used the glass coverage of the pattern (PGC) in a square area with a side of 25µm (same as the EXPOCUS pattern area), calculated with areachk, an Hotscope[3] command

eq 2. $DOSE = k_{dense} * DOSE_{dense} + k_{isolated} * DOSE_{isolated}$

Where k_{dense} and $k_{isolated}$ are functions of the PGC according to pattern tonality

eq 3.

If the tonality is clear, then:
$k_{isolated} = 1 - (PGC / 50)$
$k_{dense} = PGC / 50$

If the tonality is dark, then:
$k_{isolated} = (PGC - 50) / 50$
$k_{dense} = 1 - (PGC - 50) / 50$

if PGC passes the 50% threshold on the dense side (according to tonality) only the $DOSE_{dense}$ is considered for the calculation.

The CDE, and the related dose, is calculated for every measurement point of the mask, as the sum of the following contributes: dose setting bias of the mask, target, local glass coverage (1500µm) and positional coordinates (both X and Y) on the mask.

Dose setting bias
Data bias and the following dose setting bias are tuned by data preparation to write around the iso-focal dose of the tool, about 100mj.

Target
As for the CDE vs DOSE curve, from the EXPOFOCUS test and for every pattern type, a linearity curve (CDE vs Target) is obtained averaging the results at the best focus for each dose at every target. Since CDE differences among pattern types are already managed through different doses, the curve is normalized to the 1600nm target. The linearity curve is then parametrized as the CDE function of the target through an exponential function.

eq 4. $CDE_{target} = b + a * e^{(k * TARGET)}$

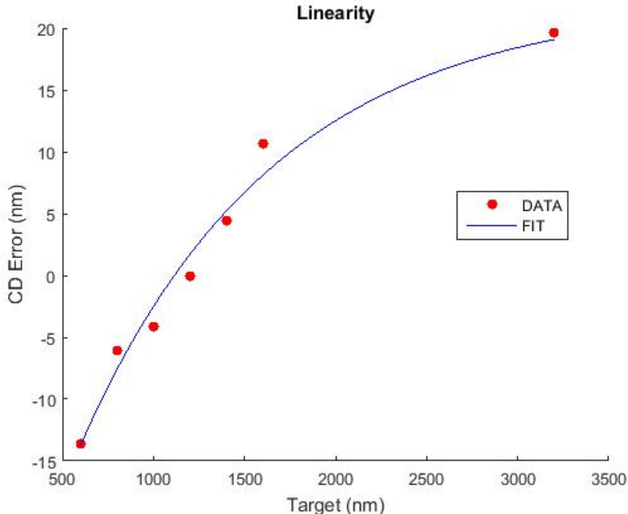


Fig 2. Linearity and its fitting provided by eq 4

As for dose calculation, because of the dependence of linearity by pattern type, the target CDE contribution is a linear combination of the contribution of isolated and dense values.

eq 5. $CDE_{target} = k_{dense} * CDE_{target_dense} + k_{isolated} * CDE_{target_isolated}$

Local glass coverage

To consider the influx of CDE provided by the local pattern conformation, has been empirically chosen to evaluate the glass coverage (LGC) in a square area with a side of 1500µm around the measurement point. A linear dependence is considered in the CDE correction relative to LGC.

eq 6. $CDE_{lgc} = a * LGC$

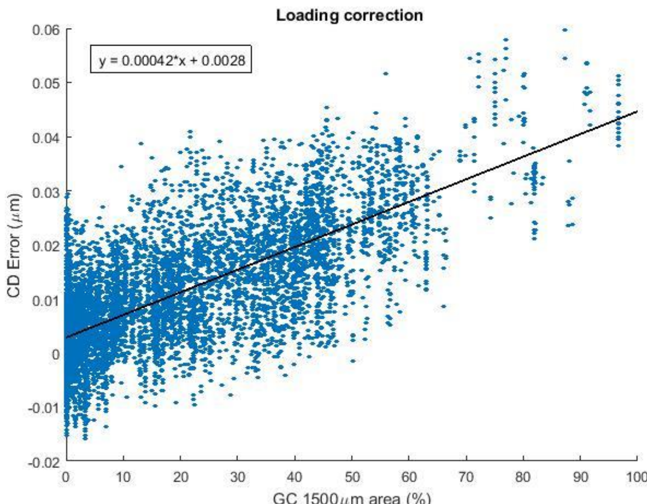


Fig 3. Loading correction impact on production data (simulated)

Coordinates

CDE relative to the positional error is calculated through a 2nd degree polynomial function of both X and Y.

eq 6. $CDE_{pos} = a * X^{\wedge 2} + b * X + c * X * Y + d * Y + e * Y^{\wedge 2} + f * r$

Total CDE is thus calculated as follows and used in the calculation of the exposure dose of the point.

eq 7. $CDE_{tot} = - (Bias + CDE_{target} + CDE_{lgc} + CDE_{pos})$

The mask exposure dose is calculated as the average of the exposure dose of the points.

Dose calculation is based on the CDE vs Dose curve obtained on X-axis measurement.

XY bias is calculated by dividing the CDE delta, obtain comparing X and Y CDE vs Dose curves at the final exposure dose, by 8, a value provided by AMAT to correlate the movement of the stage with the CD obtained on Y- axis.

eq 8. $XY_{bias} = CDE_{delta}/8$

Parameters Optimization

The parameters calculated through the test plate analysis are optimized thanks to production results feedback, to compensate the test limitation (fixed pattern features) and process drifts. CD results of the measurement

points on a fixed time range (6 months, to get hundreds to thousands of data points) are taken in account to minimize their global 3 sigma value, and to have their average mean to target equal to 0.
The optimization work is performed by the "solver" add-ins present on Excel, on variables that modify the above functions as follow:

CDE vs dose - optimization
A SHIFT parameter is added to CDE vs dose points to keep the mean to target of the production results equal to 0. Since the data set is usually much larger than the fluctuation that affects the production lines, the average mean to target is weighted on time (newer masks are more relevant than the older ones). The weight is calculated with a normal distribution (the 3sigma and average value of the normal distribution are in turn estimated according to the number of points measured in the last 30 days). The CDE vs dose curve is then parametrized on these new values.

The CDE vs dose curve is further tuned through a linear term, whose slope coefficient (M) is optimized to minimize the 3-sigma of the data set, and its zero-degree term (Q) is set to keep the average CDE of the points of the curve same as before.

eq 9. $CDE_{new} = a * DOSE^{\wedge 3} + b * DOSE^{\wedge 2} + c * DOSE + d + M*DOSE + Q$

New points for the CDE vs dose curve are simulated, and this is calculated again with the eq.1

Target - optimization
To ensure a good balance between isolated and dense measurements a 3sigma minimizing parameter ($PROXIMITY_{corr}$) is added to $CDE_{target_isolated.}$
Overall target contribution is tuned to minimize (always through minimization of 3sigma) the error related to linearity. Final CDE_{target} results as follow:

$CDE_{target} = LINEARITY_{corr}(k_{dense} * CDE_{target_dense} + k_{isolated} * CDE_{target_isolated} * PROXIMITY_{corr})$
Local glass coverage & coordinates – optimization
Both LGC in CDE_{lgc} and the coefficients of CDE_{pos} are directly calculated through the minimization of the global 3 sigma of the data set.

XY Bias – optimization
XY bias is optimized, considering the mean XY error of the last period and the weighted corrections applied so far.

The optimization is performed on simulated CDE, as the optimized parameters vary. The Simulation is obtained through the sensitivity curve, the first derivative of the CDE vs Dose curve, which provides the course of the CD correction (nm/mJ) as a function of the dose.

CDE vs Dose curve is fitted both as a 2nd degree and a 3rd degree polynomial function, the first derivative calculated for both, and the result of the two derivatives averaged. The points thus obtained are calculated in turn as a 2nd degree polynomial function.

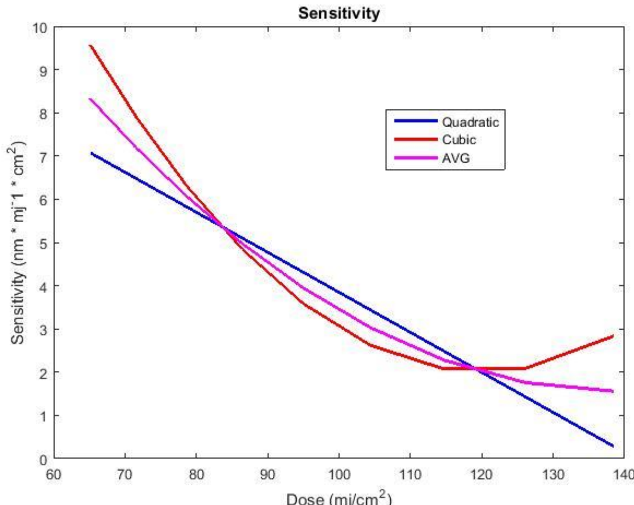


Fig 4. sensitivity curves, calculated starting from quadratic and cubic fitting, and an average of the two

SPC

The system provides an additional correction based on SPC evaluation through Western-Electric rules[4].

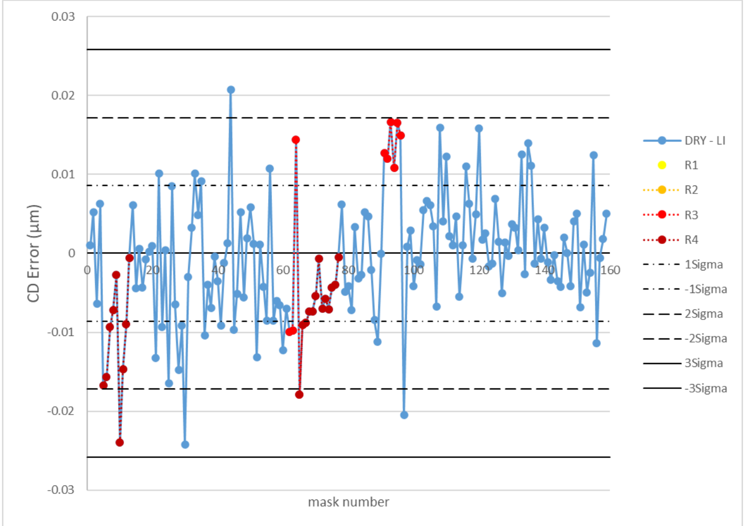


Fig 5. SPC with western electric rules.

According to the severity of the rule activated (rule 4 > rule 1), a correction between 25% to 75% of the MTT of the masks involved is calculated with the average dose and the sensitivity function and applied to the relative production line.

Results & Discussion

A production line (one of the most populated, representative, and challenging of the whole production) is taken as an example, and its production results of 3 months (from 9/3/2023 to 9/6/2023) used as the data set:

DRY_LI_LINE CLEAR on ALTA 59, with 159 masks for a total of 8200 CD points

Since the fundament of the calculation system is the CD Error vs Dose curve, the associated CDE error trend is shown. Doses have been centred around the iso-focal dose of the machine, and, as can be seen from the chart below, the error associated to the dose is negligible.

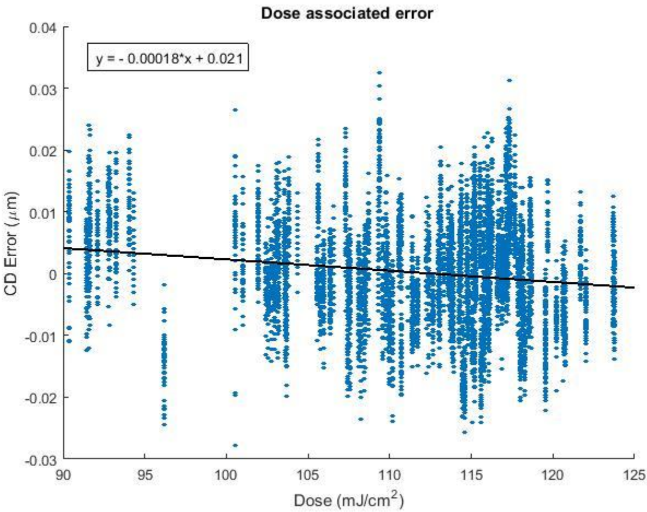


Fig 6. CDE Error on production results displayed by writing dose

To investigate the effect of the other corrections, results simulated through sensitivity are shown, with and without the relative correction applied (as for the system works, the correction and simulation are applied to single points independently)

To validate the method, the error vs dose is shown. Since points are free to move, to minimize the error, the trend results flatter, but the general behaviour in relation to dose remains the same.

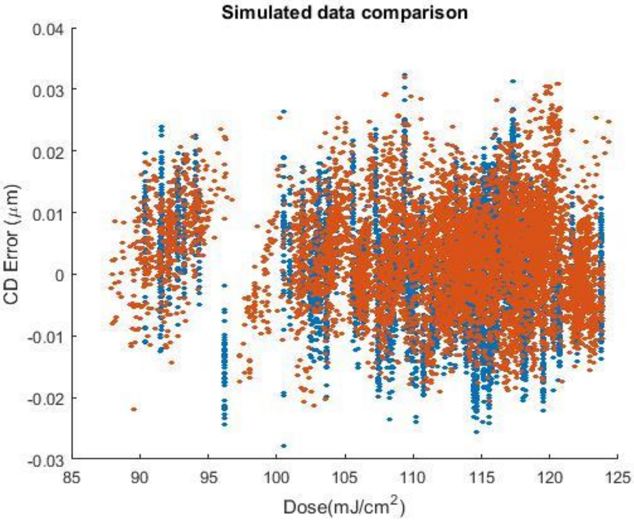


Fig 7. Data simulated and corrected, displayed by dose, and compared with original data

Along with it the relation trend of the simulated dose versus the dose used in production, shows little deviation between the two.

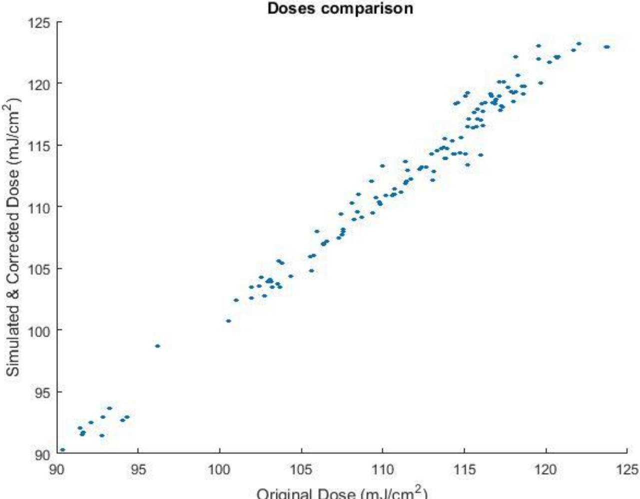


Fig 8. Original and corrected writing dose correlation.

Following, the error associated to target, glass coverage and position are illustrated.

Target

Production results show a relatively flat behaviour with the target, with a just noticeable residual error at the extreme of the range considered. The correction, as shown below reduce the CDE by more than 25nm.

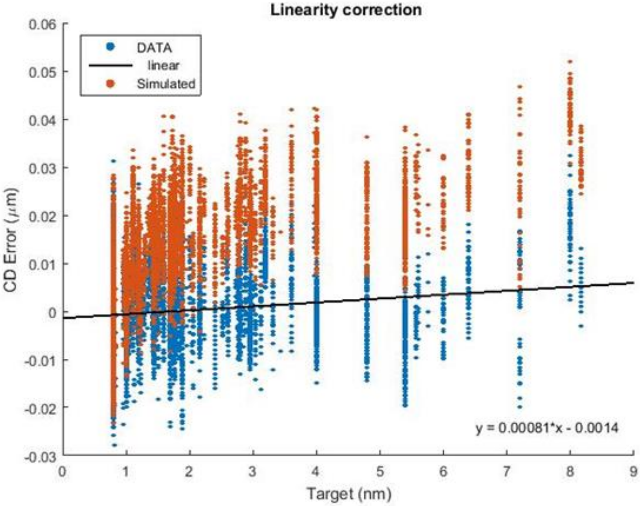


Fig 9. Production results, displayed by target, compared with the simulated data without target correction

Glass Coverage

Like the target, glass coverage data show a negligible residual error on production results, and a significant impact of the correction, 0,4nm per percentage point, that lead to about 40nm range error compensated through the whole glass coverage range.

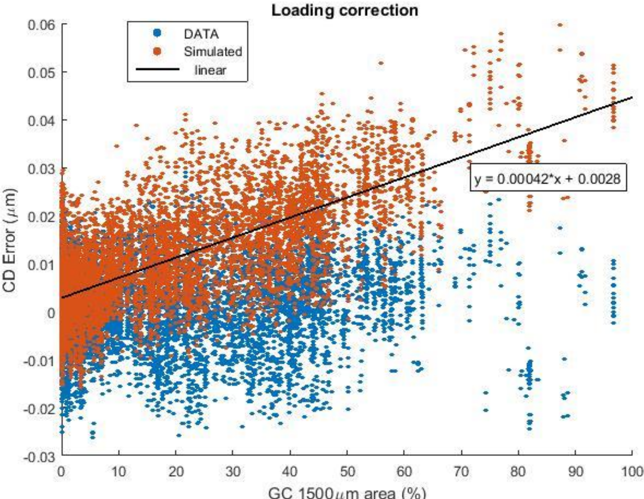


Fig 10. Production results, displayed by local glass coverage, compared with the simulated data without loading correction

Position

Position, the less impactful of the corrections performed, provide usually less than 10nm from the centre to the side of the mask on the Y axis, as shown. X axis is not reported but its impact is even lower.

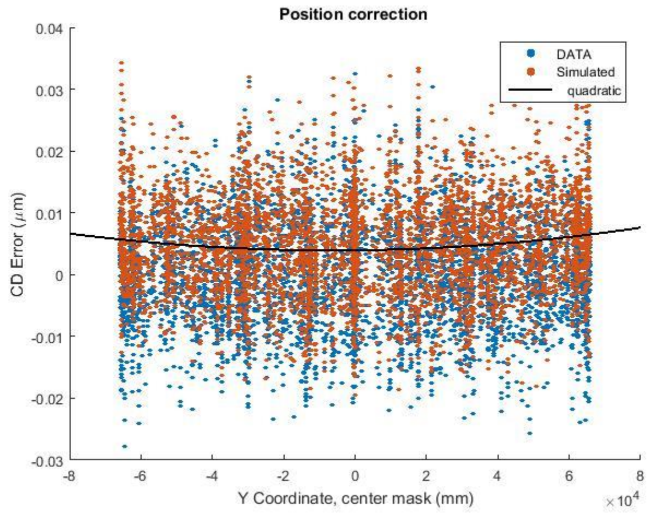


Fig 11. Production results, displayed by Y coordinate, compared with the simulated data without radial correction

The total result of the corrections, applied to the example line, delivers an average MTT of 0,6nm, a 3-sigma value of 20,42nm with no points of the 8200 measured, outside the tightest specification of CD deviation in use (+/- 35nm).

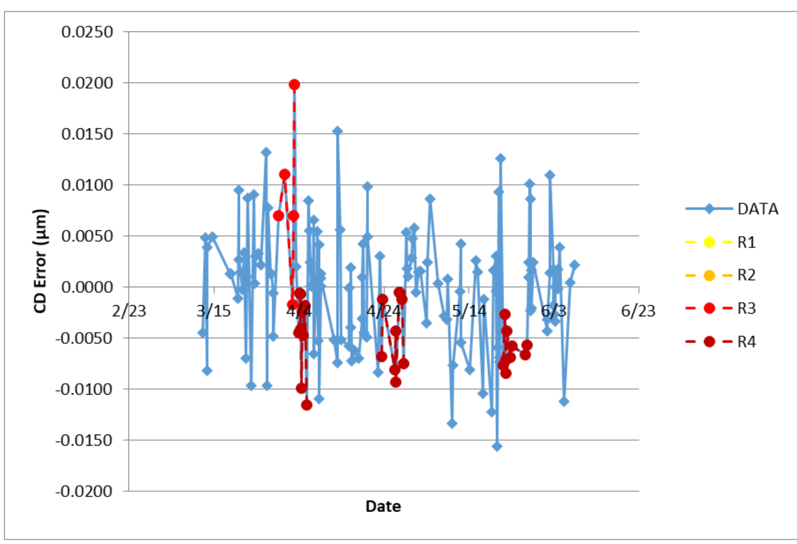


Fig 12. SPC results of three months of results of the DRY_LI_LINE CLEAR on ALTA 59 production line, displaying the MTT of the masks

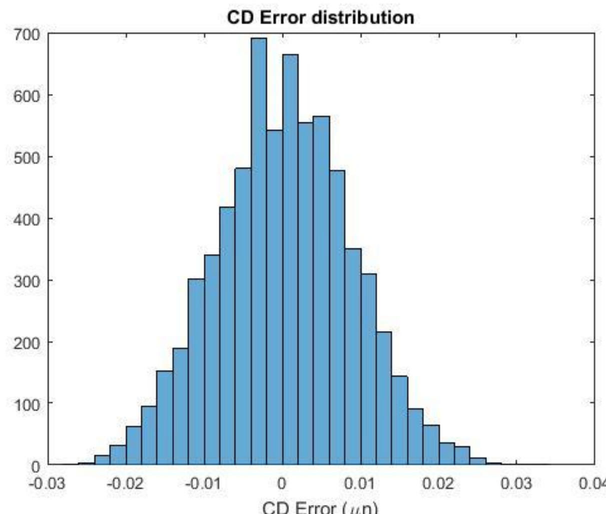


Fig 13. Error distribution of all the CD points measured in the period considered for data collection.

Conclusions

The system is completely automated and without any form of exception management, saving both production and engineering time. The final yield, up to now, is 100%.

References

[1] omdia, 03/2022 Silicon Market Trackers, Porsche Consulting Analysis
[2] a native exposure test embedded in the ALTA laser writing systems (Applied Materials)
[3] Shogo Narukawa, Kiyoshi Yamasaki, Yuji Machiya, and Naoya Hayashi "Photomask pattern viewer and analyzer: HOTSCOPE", Proc. SPIE 5853, Photomask and Next-Generation Lithography Mask Technology XII, (28 June 2005); https://doi.org/10.1117/12.617155
[4] Western Electric Company (1956), Statistical Quality Control Handbook. (1 ed.), Indianapolis, Indiana: Western Electric Co., p. v, OCLC 33858387

Predicting resist pattern collapse in EUVL using machine learning

Sean D'Silva[a], Raghunandan Arava[a], Andreas Erdmann[a], Thomas Mülders[b], and Hans-Jürgen Stock[b]

[a]Fraunhofer Institute for Integrated Systems and Device Technology (IISB), Erlangen, Germany
[b]Synopsys GmbH, Aschheim, Germany

ABSTRACT

Pattern collapse has become a problem area in lithographic manufacturing due to the added complexity as we move towards extreme ultraviolet (EUV) focused mass production of semiconductor devices. Collapse is influenced mainly by the geometry and mechanical properties of the pattern. Photoresist patterns with higher aspect ratios (ARs) or lower feature spacing (dense features) are prone to collapse. This, combined with a lower stiffness value (Young's modulus), adds to the undesirable deformation of patterns at the wafer level as we move towards advanced technology nodes. Such a deformation in the resist is seen after the rinsing stage and is caused mainly due to the non-uniform drying of the rinse liquid after chemical development. The main causes of this are the unbalanced Laplace pressure ΔP difference across the liquid-air interface and the surface tension force (STF) along the three-phase line. In addition to that, the sidewall surface roughness leads to localized regions of higher aspect ratios which makes collapse modeling in EUV lithography more challenging. The irregular variation in the aspect ratios increases the risk of collapse and also requires additional model considerations. A machine learning (ML) based approach is introduced to predict deformation characteristics for rough cross-sections (XZ plane) which is then utilized for the computation of the overall deformation of an entire lines and spaces (L/S) pattern. The 3D profiles are converted into a 2D representation using modified Fisher vectors (FVs) and labeled based on the estimated deformation of the given rough profile. A convolutional neural network (CNN) is then trained with the data generated and used to predict collapse probabilities for a given data set. For the prediction of collapse in pillar patterns, a slightly different ML-based approach is used based on the cross sections in the XY plane. A finite element method (FEM) model is implemented to calculate the deformations δ for a given pillar arrangement which then serve as labels for the training data. The cross-sections are stacked together along the height/thickness and a 3D convolutional neural network is trained and used for collapse prediction.

Keywords: photoresist, pattern collapse, mechanical effects, deformation

1. INTRODUCTION

Feature geometries are based on particular application requirements and the mask layout. Typical resist features include lines and spaces (L/S) patterns, pillars and contact holes. Thickness of the resist has an impact on etch stand-off and can cause defects if the resulting resist pattern is not etch resistant. Thicker resist features tend to have higher aspect ratios ($AR = H/w$) which contribute to pattern instability and collapse.

Figure 1 shows the most extreme case where the rinse liquid is accumulated in between two lines and the neighboring regions have no liquid around them. This is caused due to the uneven drying of the rinse liquid after development and will result in pattern deformation or collapse. Also, when there is an uneven level of liquid around a line, there is a Laplace pressure ΔP imbalance and this would lead to some amount of deformation, subject to the aspect ratio and polymer material properties.

$$\Delta P = -\gamma . \kappa_C \tag{1}$$

Further author information: (Send correspondence to Sean D'Silva)
Sean D'Silva: E-mail: sean.dsilva@iisb.fraunhofer.de, Telephone: +49 9131 761-246

38th European Mask and Lithography Conference (EMLC 2023), edited by
Uwe F.W. Behringer, Jo Finders, Proc. of SPIE Vol. 12802, 1280205
© 2023 SPIE · 0277-786X · doi: 10.1117/12.2675700

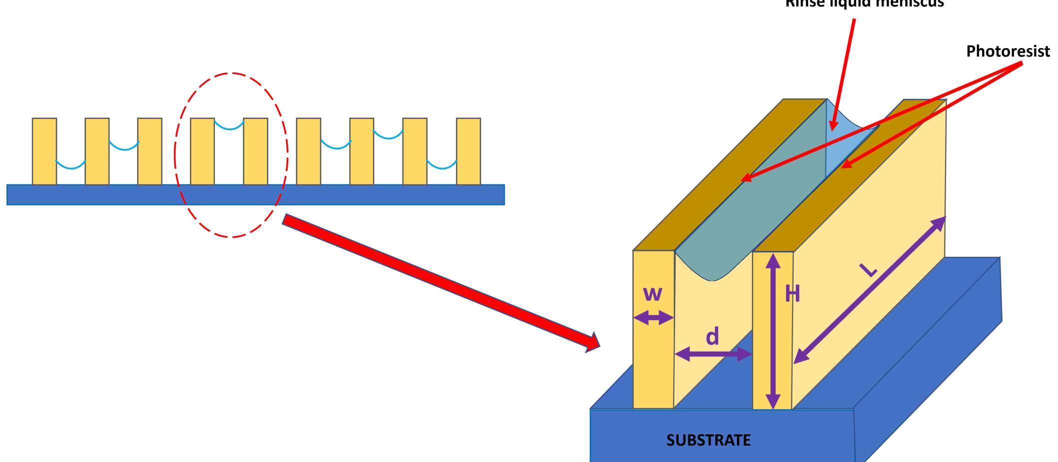


Figure 1: Rinse liquid in the space d between two lines with width w, height/thickness H and pattern length L. Adapted from Reference[1]

The pressure differential ΔP across the liquid-air interface produces a convex or concave surface depending on whether the rinse liquid is hydrophobic or hydrophilic. In Equation 1, κ_C is the interface curvature and γ is the surface tension value for the respective liquid. κ_C is negative for a concave surface which means the pressure outside the liquid is higher than it is inside and similarly κ_C is positive for a convex surface.

1.1 Capillary forces during resist rinsing

In addition to the Laplace pressure along the outer sidewalls of the pattern, there are capillary forces acting along the three-phase line (where solid, liquid and vapor phases coexist)[2] as shown in Figure 2. Since the rinse liquid is hydrophilic, the contact angle θ is less than 90°. Analogous to lines and spaces (L/S), pillar patterns

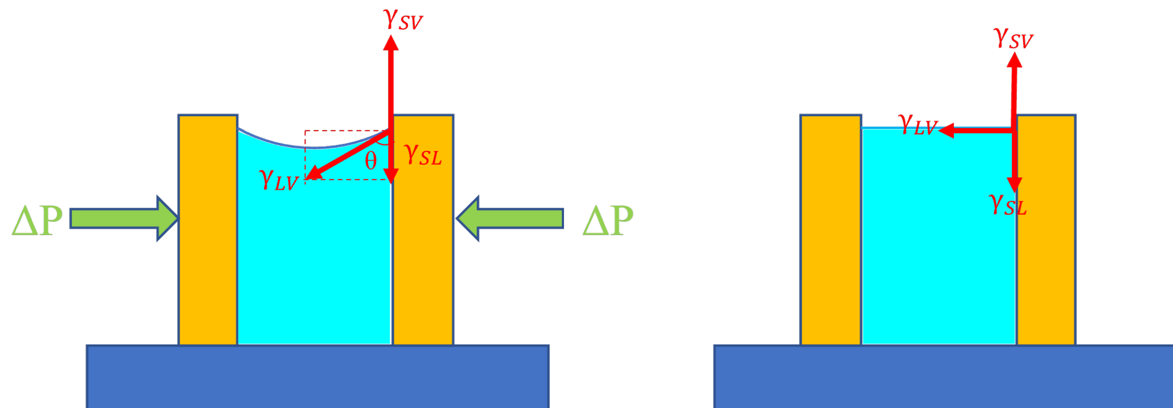


Figure 2: Left: Laplace pressure ΔP and capillary forces (γ_{SL}, γ_{SV} and γ_{LV}) for a lines and spaces pattern with rinse liquid contact angle θ. Right: Capillary forces for a rinse liquid with a contact angle θ of 90°. Adapted from Reference[1]

are equally prone to collapse due to their high surface area to volume ratios. A 2D array of isolated pillars can collapse and assemble into clusters leading to defects that can greatly impact the overall pattern at the wafer level. The self assembly of pillar features could have some interesting applications as well, but has a negative impact on lithographic performance.[3] Specific to the lines and spaces case, the Laplace pressure is given as:

$$\Delta P = \frac{2\gamma cos\theta}{d} \tag{2}$$

The Young equation for the three-phase line is given by:

$$\gamma_{SL} - \gamma_{SV} = \gamma_{LV} cos\theta \tag{3}$$

here γ_{SL}, γ_{SV} and γ_{LV} are surface tensions of the solid-liquid, solid-vapor and liquid-vapor interfaces respectively. The $\gamma_{LV} sin\theta$ component is the effective horizontal force acting along the three-phase line. Collapse is also observed when the pattern is completely surrounded by the rinse liquid as shown in Figure 3. This is mainly due to the capillary meniscus interaction force[4] present when particles disturb a liquid surface. The force between the partially immersed pillars can be approximated as:[3]

$$F_C = -\frac{\pi\gamma R^2 cos^2\theta}{\sqrt{(\frac{X}{2})^2 - R^2}} \tag{4}$$

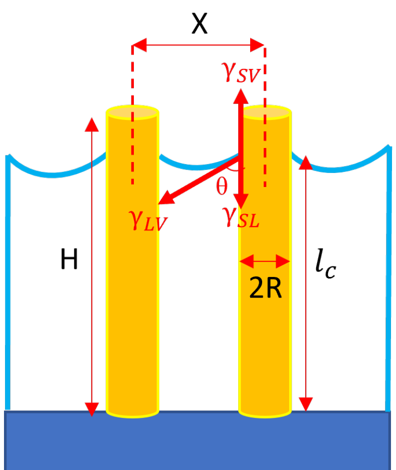


Figure 3: Capillary forces (γ_{SL}, γ_{SV} and γ_{LV}) for a uniformally immersed pillar feature of height H, capillary length l_c, radius R, pitch X with rinse liquid contact angle θ.

2. PREDICTING PATTERN COLLAPSE USING MACHINE LEARNING

Pattern collapse in itself can have numerous causes and predicting its probability in EUVL lithography entails numerous challenges. Due to the random or stochastic nature of features in EUVL, predicting the pattern shapes and contours require newer modeling approaches[5] compared to the ones used previously in DUVL. The modeling approaches used in DUVL were based on flat-sided profiles without roughness. This approach would, however, not work in the case of EUVL due to the roughness and varying sidewall angles of resist features. The task of predicting pattern deformation or collapse in cases where there is substantial roughness due to line width roughness (LWR) and line edge roughness (LER) is compounded significantly due to stochastic effects. The SEM image shown in Figure 4 depicts how a defect like pattern collapse can critically hamper pattern transfer to the wafer. The collapsed lines make it impossible to etch the feature on the substrate. Rigorous simulations can be quite accurate in their predictions but are generally very tedious and slow especially for the use cases expected in EUVL. One way of circumventing this hurdle and bridging the gap between accuracy and computation speed is to use machine learning. Deep learning (DL) is a widely used computational approach in the field of ML and is capable of achieving very good results in various application areas.[6] Depending on the type of training data and learning model used, various deep learning approaches can be implemented to predict, classify and estimate outcomes for a wide-spectrum of scenarios.[7,8]

There are numerous contributors to LWR and LER in EUV patterning ranging from roughness in the lithographic mask, photon shot noise, fluctuations in acid generation and reaction with the quencher molecules due to the random locations of the PAG and quencher.[10–13] Using a FEM-based model for simulating deformation

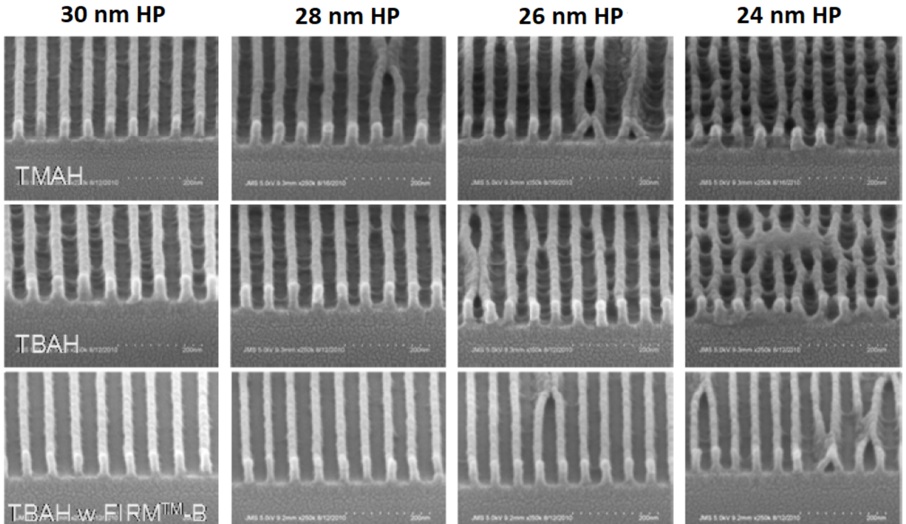


Figure 4: SEM images of collapsed lines and spaces features using different chemical developers. Taken from Reference[9]

in rough profiles would require complicated meshing and discretization techniques in order to capture all the irregularities in the geometry. The FEM solver would need to solve a large system of equations due to the large number of higher order mesh elements making the overall simulations extremely slow. A total simulation time of several hours was observed while simulating a single rough profile for predicting collapse. Predicting resist defects with the help of simulations therefore requires newer strategies due to the overall randomness caused due to the numerous optical and chemical effects. This is where machine learning could help speed up the process.

2.1 Lines and spaces

2.1.1 Generation of training data

Sidewall roughness in EUVL is basically a random deviation from the expected feature edge position. The standard deviation, however, is not the only parameter that can characterize resist roughness. SEM images of resist patterns in EUV lithography have shown that there can be varying levels of roughness depending on the resist blur. The first step, therefore, is to replicate the randomness in the profile contours mathematically to help simulate such behavior. Figure 5 shows the impact of LER and LWR on the final shape of a line resist feature. The LER is deviation of the profile edge from the expected shape and LWR is seen as an inconsistency in the feature width or critical dimension (CD). An exponentially decaying auto-correlation function shown in Equation 5 can be used to depict the standard deviation of an edge based on a line-length dependence.[14]

$$R(r) = \sigma_r^2 \exp(-(r/\xi)^{2H_r}) \tag{5}$$

here, σ_r is the standard deviation or the root mean square (RMS) value for the LWR, H_r is the roughness exponent, ξ is the correlation length and r is the Euclidean distance ($|x|$ for one dimension and $\sqrt{x^2+y^2}$ for two dimensions). A power spectral density (PSD) function can be used to describe the frequency f dependent characteristics of roughness seen in photoresists. A single or multi-dimensional Fourier transform of Equation 5 yields the respective PSD function. Analytical solutions to the Fourier transform were, however, only derived for certain parameter values of $H_r = 0.5$ and $H_r = 1$.[15] Moreover, experimental data shows that roughness exponent values between 0.5 and 1 represent most of roughness aspects in EUVL.[16] The Fourier transform in one dimension applied to the autocorrelation function of Equation 5 returns a 1D PSD which can be analytically solved for $H_r = 0.5$ as:[17]

$$PSD(f) = \frac{2\sigma_r^2 \xi}{1 + (2\pi f \xi)^2} \tag{6}$$

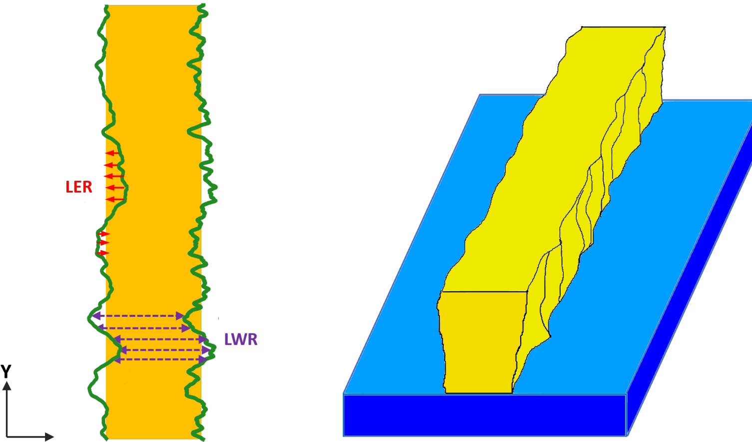


Figure 5: Left: Impact of line edge roughness (LER) and line width roughness (LWR) on resist contours. Red arrows and purple dashed lines represent LER and LWR respectively for the top view (XY plane) of a line feature. Right: A 3D representation of a rough line feature

In a similar way, the 2D Fourier transform of the autocorrelation function for a roughness exponent H_r of 0.5 yields:[16]

$$PSD(f) = \frac{2\pi\sigma_r^2\xi^2}{[1+(2\pi f\xi)^2]^{3/2}} \tag{7}$$

Figure 6 demonstrates how the characteristics of rough lines change based on varying ξ and shows the corre-

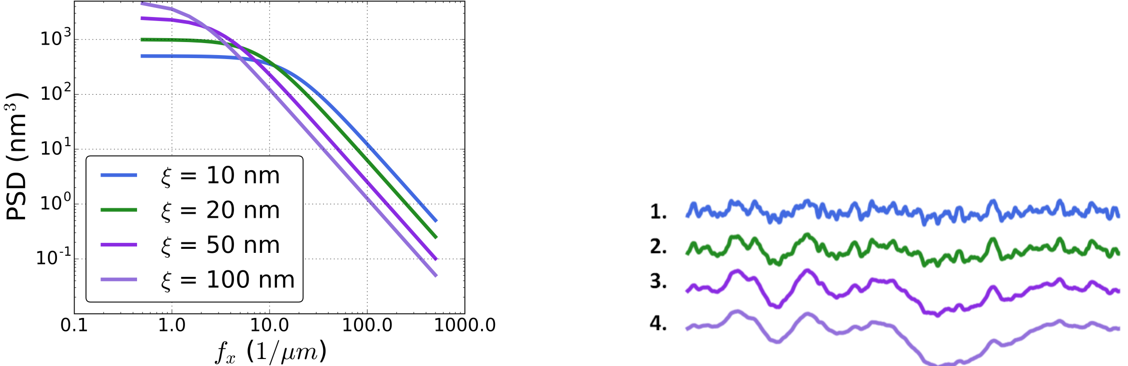


Figure 6: Left: The power spectral density (PSD) on a log-log scale for varying correlation lengths ξ and Right: corresponding rough lines generated for fixed parameter values: roughness exponent $H = 0.5$, standard deviation $\sigma_r = 5$ nm

sponding PSDs used to generate them. The PSDs from Equations 6 and 7 are used to generate rough lines and surfaces respectively. The parameters used in the PSD can be varied so as to calibrate a resist model based on experimental data. A three-dimensional line resist feature can be generated using a combination of 1D and 2D PSDs by superimposing the surfaces generated on the sidewalls of rough edges as shown in Figure 7. Generating the displacement fields using finite element modeling for such rough features is very cumbersome and time consuming due to the irregular and uneven sidewall profiles. The simulations would take very long (up to several hours for a single feature) since the mesh quality in the sidewall regions would need to be very high in order to capture the unevenness of the surfaces. In order to circumvent this, a machine learning approach is developed to predict collapse probabilities in line features.

In order to generate training data for the prediction model, an interpolation method is used to calculate the displacement field of the entire line feature. The first step would be to generate enough 2D cross-sectional profiles and calculate their individual displacement fields that can then serve as the building blocks for the

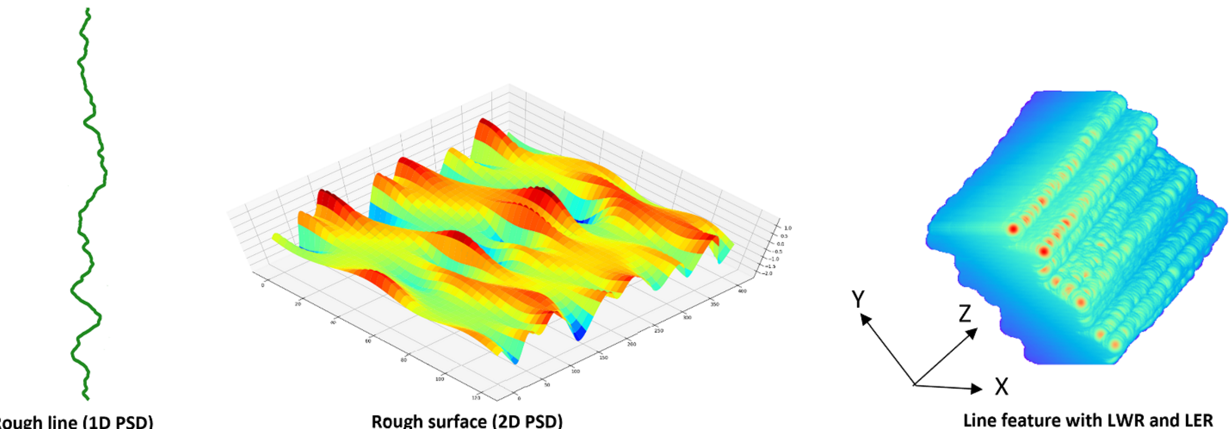


Figure 7: Left: A rough line is generated using a 1D PSD (left) and a 2D PSD (centre) is used to produce a rough surface for the resist sidewalls. Right: A combination of both the rough lines and rough surfaces generates a line feature with LER and LWR.

deformation prediction. The lower part of the resist receives significantly less light leading to fewer photons being absorbed within a smaller region. Since negative-tone development (NTD) is used, the exposed regions remain after chemical development and the lower photon absorption at the bottom leads to a narrower width at the bottom compared to the top. This undercutting of the resist features is quite prominent in EUVL due to the low optical absorbance of light reaching the bottom of the resist and the differences in photochemistry.[18] Figure 8 shows such undercut lines which are used for the generation of training data. These cross-sectional discretized profiles represent the numerous cross-sections of rough lines along the Y-axis. The discretized profiles have varying widths $w1$ - $w5$ along five equidistant positions (with height h) along the height resulting in sidewall angle variations. All the profiles were generated on a scale and grid having a minimum discretization of 1 nm throughout, in order to maintain a good level of accuracy and balance with the computation time. The Laplace pressure ΔP is defined perpendicular to the individual line width subsections and surface tension force (STF) is acting along the three-phase line at the top. The cross-sections cumulatively make up a rough line and their displacement fields combined, constitute the overall displacement field of a line feature with roughness.

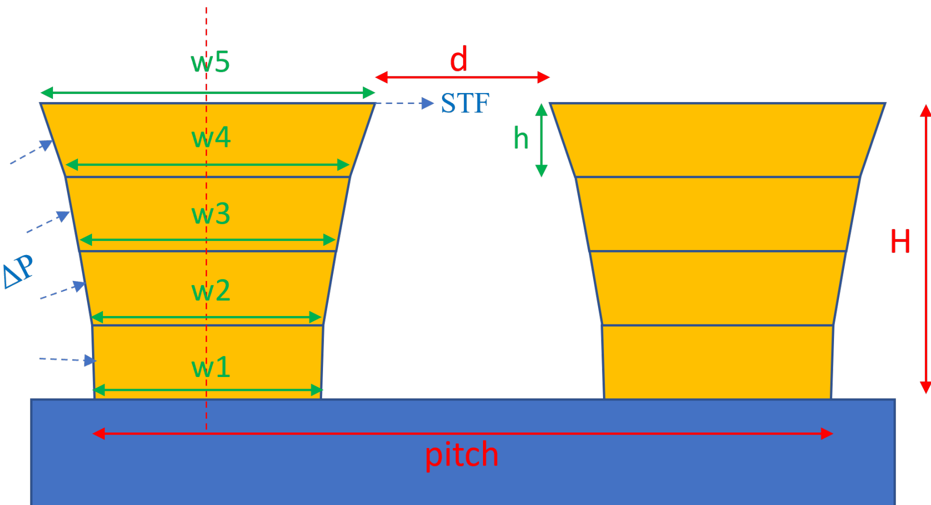


Figure 8: A dense lines and spaces use case discretized along the height H into four equal slices of height h. d is the line spacing at the top and $w1$ - $w5$ represent the varying line widths of each discrete feature section. The Laplace pressure ΔP is acting perpendicular to the sidewall and the surface tension force (STF) is acting at the three-phase line on top.

2.1.2 Predicting resist displacement fields using graph neural networks (GNNs)

An interpolation method making use of graph neural networks (GNN) was implemented[19] to predict the individual cross-sectional displacement fields of the L/S pattern along the length. The interpolation method was validated for accuracy using rigorous FEM simulations and showed a relative displacement error of $\approx 4\%$. The discretized profiles that were used to train the GNN had no roughness along their length in the Y direction and had a longitudinal aspect ratio (LAR) i.e., L/d = 20. For LAR <20, the displacement field would not be representative of a real line feature with L$\gg$d.[1] Moreover, the cross-sectional displacement fields from the FEM simulations could be used directly to estimate the displacement field in some cases based on the profile shape. This helped save some time in the computation, since relatively less rough profiles were estimated by the GNN.[19] The interpolation technique is capable of estimating displacement fields for low to medium level of roughness observed in line features in EUVL. The correlation length ξ is used to determine the height h discretization for the profile. The correlation length ξ and standard deviation σ_r values for the rough profiles were therefore not very large in order to maintain acceptable error limits of less than 5%.[1,19,20] The reason for using GNNs in this particular case is due to their high relational inductive bias when compared to other traditional neural network constructs such as Fully Connected Neural Networks, Convolutional Neural Networks, Recurrent Neural Networks, etc.[21] Some of the advantages of GNNs are listed below,

- It can be used for defining arbitary relationships
- It is permutation invariant
- It is basically a form of combinatorial generalisation
- It is scalable

The specific type of framework that has been employed for this kind of problem is a Message Passing Neural Network (MPNN) with certain modifications according to how the data is available. The basic framework of MPNN consists of three functions, that are applied in series, i.e., function that computes messages, function that aggregates the messages, and function that updates the node embeddings based on the aggregated message received.

The message passing layer was modeled in such a way that the features of the nodes were updated to their new positions in x and y coordinates. To calculate the displacement, the difference between the output and the

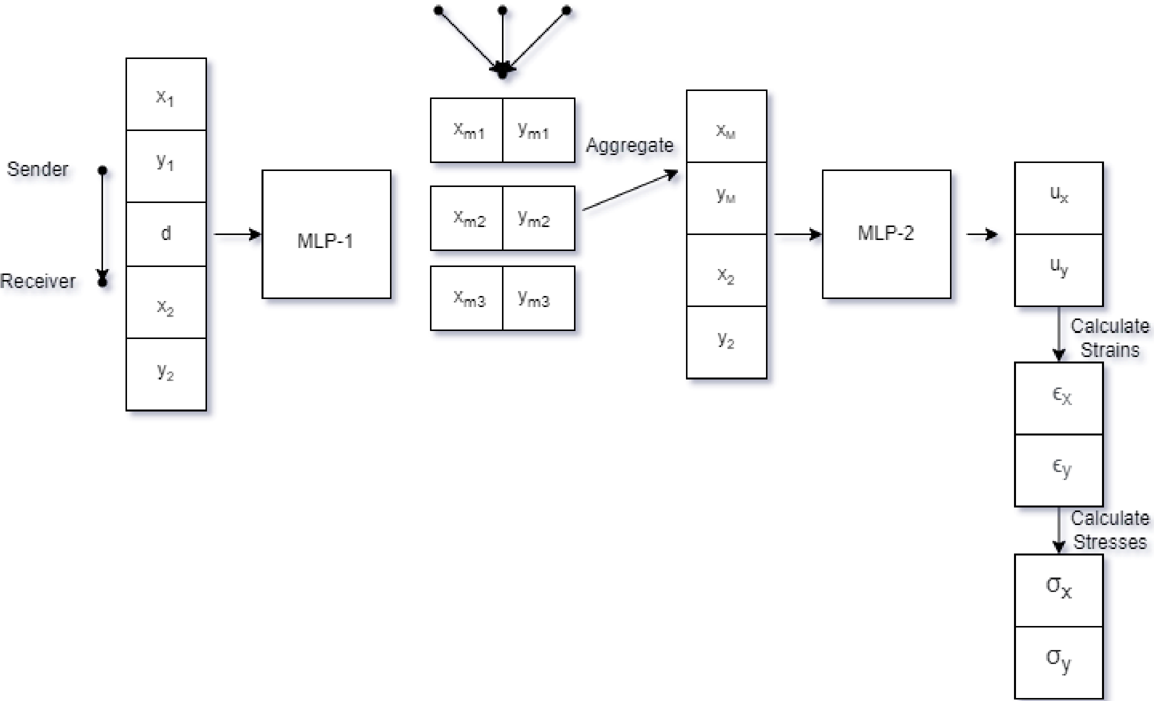


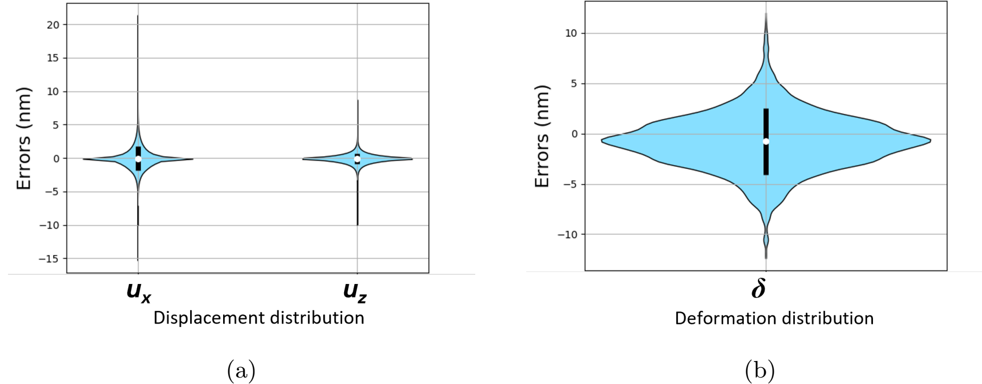
Figure 9: Message Passing Layer: Elasticity Model

input features was done, which were compared with the actual values for formulating the loss function during the training of the model.

To further enhance the predictions, elasticity was modeled in the loss function as shown in Fig 9. Instead of predicting the final positions of the nodes, the model directly predicts the displacements associated with the nodes in x and y, from which their corresponding strains are calculated. In turn, from the strains, the stresses are calculated. These predictions from the model are compared with the actual values to formulate the loss function and to use it in the back propagation step of the training of the model. Both the predicted strains and stresses and the actual strains and stresses are calculated based on the linear elasticity theory. Figures 10a and 10b show the error distribution of the displacement fields in the X and Z directions and the overall deformation error distribution respectively. These results were obtained from the GNN while comparing cross-sectional data of a given L/S pattern with rigorously simulated FEM data.

The hyper-parameters such as the learning rate, loss function, aggregation mechanism, and intermediate activation function used for training the final model are 0.001, Huber loss, sum aggregation, and ReLU respectively. Although the performance of the model with 0.001 learning rate seems to work well, Huber loss does not perform well with it. Thus, the learning rate is chosen to be 0.0001. With these parameters, the model is trained for 300 epochs. In almost all cases, a GNN has non-convexity of training, its convergence is usually guaranteed at a linear rate. With the introduction of depth in the model, the training is accelerated.[22] The reason for training it beyond 100 epochs is because the model has the capacity to converge to a near zero error and still continues with reducing loss. A test was performed with a rough resist profile generated using ANSYS[23] software. The feature pitch is 28 nm and height is 35 nm. The random cross-sections' displacement fields and their predicted displacement fields are presented in Figure 12. The predicted displacement fields are closer to the ground truth displacement fields in this case. Specifically, the values at the three-phase line at the top, which are important for the pattern collapse classification problem, are close to the ground truth values.

Figure 11 shows the complete predicted displacement field in 3D of a rough resist profile of height 35 nm and pitch 42 nm. The cross-sections of the point cloud are pre-processed by Delaunay triangulation and then converted into graphs. These graphs are used as an input in the trained elasticity model for predicting their



(a) (b)

Figure 10: Left: Violin plot showing the error distribution of the displacement in the X and Z directions and Right: corresponding deformation error distribution. The thick black lines represent 80% of the distribution and the white circle in the middle represents the median error.

respective displacement fields in X and Y. Once all the predicted displacement fields of the cross-sections are obtained, they are assembled and plotted as shown in Figure 11. As it can be observed, there are certain cross-sections, where there are outliers predicting a much higher displacement. The reason for such predictions could be because the corresponding cross-sections have a smaller average CD due to the variation in roughness along the length of the profile. In reality, the cross-sections where the outliers lie should be restricted from displacing further by the adjacent cross-sections as per the expected behavior of a solid material. Because of this behavior, the collapse labeling of the resist is performed in such a manner that only if more than 3% of the cross-sections have collapsed, then the entire resist profile is labeled as a collapsed pattern.

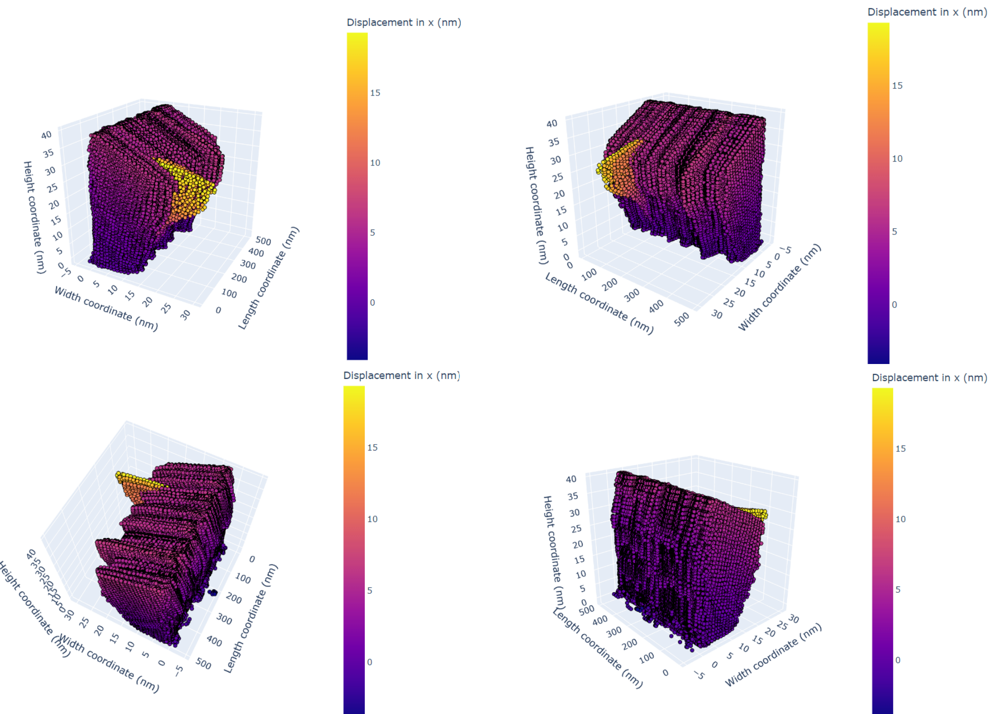


Figure 11: Predicted displacement field of a rough resist profile of height 35 nm and pitch 42 nm from different viewing perspectives in the X, Y and Z directions.

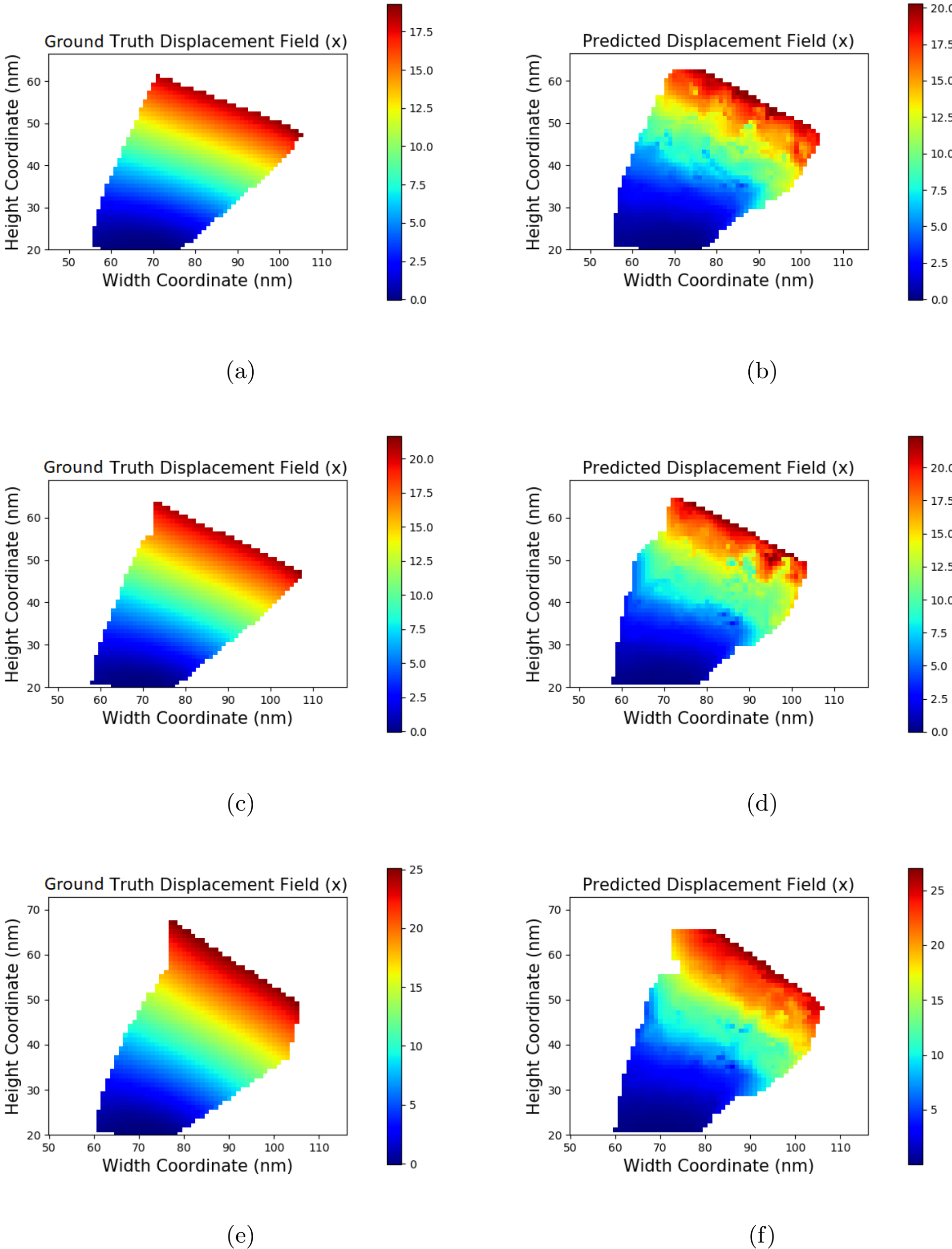


Figure 12: Left images: Ground truth displacement fields. Right images: Predicted displacement fields. The feature pitch is 28 nm and height is 35 nm. The three different rows represent random cross-sections from a rough line.

2.1.3 Collapse prediction

After generating the displacement fields and their corresponding labels based on the calculated deformation, the next step is to produce the input data for a CNN that can predict the collapse probabilities of a given resist line feature with roughness. Due to the randomness created by the roughness in the resist features, there are inconsistencies in the size of the 3D profiles required for the training data. The rough profiles are converted into a consistent set of 2D data with fixed dimensions and used as an input to the 2D convolutional neural network. The resist profiles generated using the method explained in Section 2.1.1 are converted first into 3D point clouds.

This unstructured and point-wise unordered representation of the rough resist poses challenges in deep learning applications. The point clouds need to be represented in a more structured and ordered manner to be effectively used as an input to a CNN.

Point clouds were previously represented using Gaussians[24,25] using a technique that maps the distribution of points by utilizing a local normal arrangement of several overlapping grid cells. To generalize a resist point cloud in an input format suitable for a CNN, a new strategy employing a 3D modified Fisher vector (3DmFV)[26] was used. A Gaussian mixture model (GMM) in three dimensions is described as a uniform set of Gaussians centered on a 3D grid. The 3DmFV describes the points based on their deviation from the GMM[27] and the various components are averages of the given collection of points which are generalized over symmetric functions.

A number of symmetric functions can be used in a Fisher vector, with the summation be the main symmetric function used. Using only the summation function, however, might not be very favorable for larger point cloud data, since it does not give complete information about the input. Additional symmetric functions including the minimum and maximum are included in the Fisher vector representation and have proven to improve the accuracy in the description of a given point cloud.[26,28] The final 3D modified Fisher vector (3DmFV)[26] resulting from the three symmetric functions is given below:

$$\mathscr{G}^{X}_{3DmFV_{\omega}} = \begin{bmatrix} \sum_{n=1}^{N} L_{\omega} \nabla_{\omega} log u_{\omega}\left(\boldsymbol{p}_{t}\right)|_{\omega=\alpha,\mu,\sigma} \\ \max_{n}\left(L_{\omega} \nabla_{\omega}\left(\boldsymbol{p}_{t}\right)\right)|_{\omega=\alpha,\mu,\sigma} \\ \min_{n}\left(L_{\omega} \nabla_{\omega}\left(\boldsymbol{p}_{t}\right)\right)|_{\omega=\mu,\sigma} \end{bmatrix} \tag{8}$$

For $K = m^3$ number of Gaussians in a 3DmFV, there are $20 \times K$ components which can then be expressed as a $20 \times K$ matrix as shown in Figure 14. The 20 components that make up the 3DmFV are a set of scalar values representing the three dimensions of the 3D point cloud. The summation and max function yield three components each along the three axes X,Y and Z i.e. $3 * (3 + 3)$, whereas the min function yields just a constant weight and is therefore having 2 components. The total here is a set of 20 3DmFV components along with the $4 \times 4 \times 4$ number of Gaussians shown in Figure 14. The regions on the left and right of the 3DmFV have zero values since they represent spaces between the lines and have no points present. The slight inaccuracies due to the averaging of the Fisher vector components are compensated for by using it in combination with a CNN.[26] The CNN used here is a classification network that is capable of classifying a given rough profile into

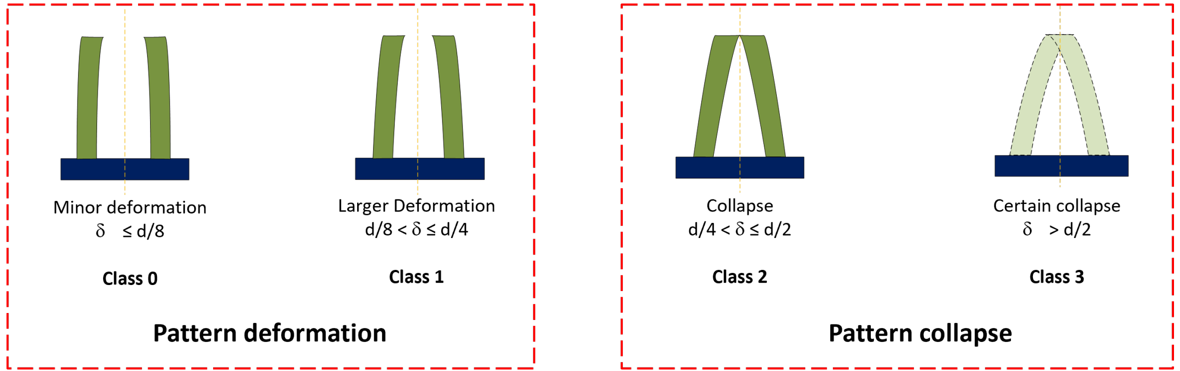


Figure 13: Pattern collapse classification into four different classes based on the deformation/deflection δ ranging from minor deformation (class 0) to a guaranteed collapse scenario (class 3).

the different categories explained in Figure 13. The CNN is the last component of the entire collapse prediction and classification procedure. A complete schematic showing how the training data is generated and classified is represented in Figure 15. The first step would be to generate the training data for the classification network. For this, a certain number of practical use cases are considered which closely resemble real profiles from experimental SEM images. The 3D line profiles generated are then converted into a 3DmFV which is a 2D representation of

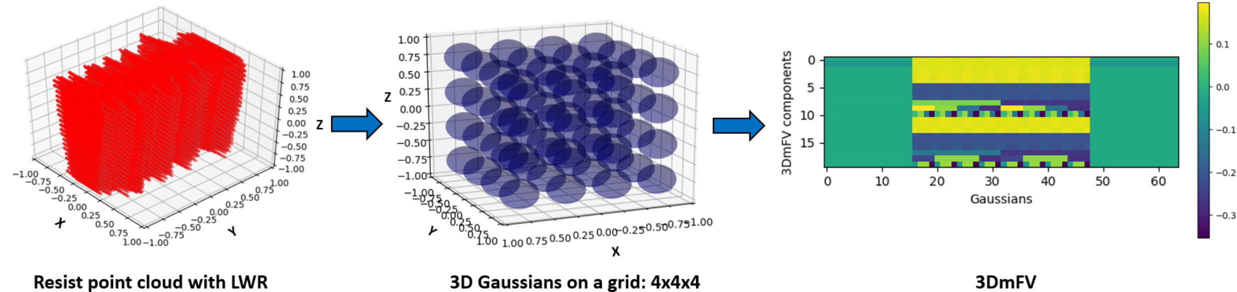


Figure 14: Schematic showing the generation of a 3D modified Fisher Vector (3DmFV). Point cloud of a line resist feature (left) inside a 3D Gaussian 4x4x4 grid (center) is used to create a 3DmFV (right).

the same profile. Since we are using a supervised learning scheme, the 3DmFVs are given their corresponding labels based on the interpolation of the cross sections in combination with a graph neural network (GNN).[19]

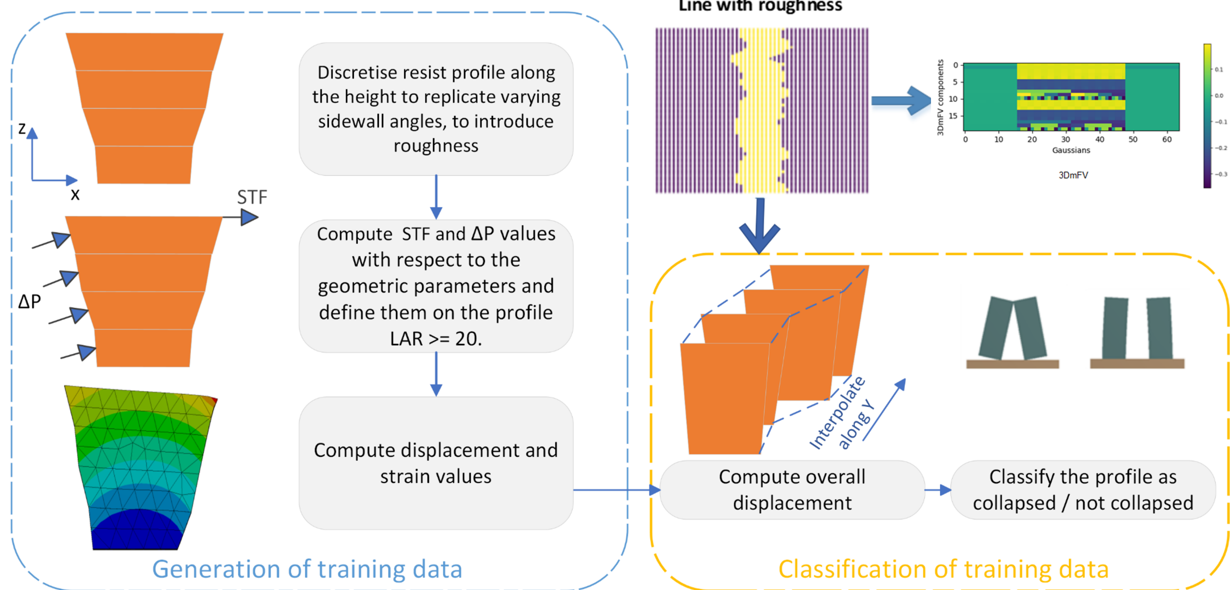


Figure 15: Pattern collapse data classification using the training data generated based on discretized line features having a longitudinal aspect ratio (LAR $\geq$ 20). The training data is classified using an interpolation method together with the generation of the 3DmFVs from the point clouds.

The training data for a rough profile is generated using the criteria given below. The parameter values are varied within the ranges given to produce a sufficiently large enough data set for training a CNN for image classification.

- Half width w = [6, 15] nm
- Height = [20, 35] nm
- Pitch = [28, 42] nm
- Correlation length ξ = [8, 11] nm
- Standard deviation σ_r = [2, 5] nm

- Roughness exponent $H_r = 0.5$, 1
- Young's modulus $E = 0.3$ GPa[29]
- Poisson's ratio $\nu = 0.4$[30]

The CNN architecture used for the 3DmFV image classification is shown in Figure 16. The network input is given as a 3DmFV (2D array) with a size of 20x64x1. The 3DmFV has a single channel with 20 components and 64 Gaussians (4x4x4 grid). A 2D convolution[31,32] is performed using a 3x3 kernel size with a progressively increasing number of filters. Instead of a max or average pooling layer, a 2D convolution with a stride of 2 is used. Such a strategy was proven to be on par or even out compete typical CNNs on several object recognition datasets by employing a simpler network architecture.[33]

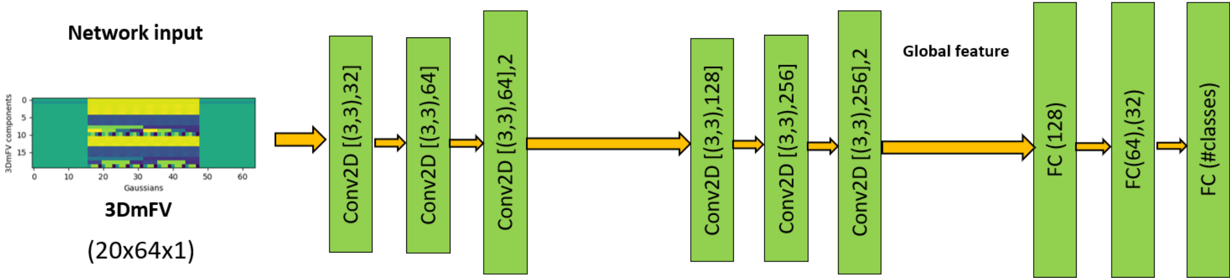


Figure 16: Convolutional neural network architecture for the prediction of collapse probabilities in a lines and spaces (L/S) use case in EUVL.

Dropout layers[34] were used in between the fully connected layers as a regularization method to help prevent network over-fitting.[35] Some of the other hyper-parameters selected for the training were determined based on various tuning runs and optimizations are given below:

- Batch size = 10, 50
- Learning schedule = 'Exponential decay'
- Learning rate = 0.001
- Decay rate = 0.75
- Decay steps = 1000
- Optimiser = 'Adam'[36]
- Loss function = 'Categorical Cross-Entropy'

The training was performed on NVIDIA Tesla V100 and NVIDIA Quadro RTX 5000 graphics processing units (GPUs) with the batch size being adapted accordingly to suit the memory available in each GPU. A total of 6000 input 3DmFVs ($6000\times 20\times 64$) were used during the training phase with a network training time of ~ 5 hours. Each class has a total of 1500 data points. A validation set of 800 data points was used during the training while monitoring losses and accuracy as shown in Figure 17. The figure also shows the overall accuracy of the trained network on a random test set that classifies collapse probabilities based on the 4 different classes described in Figure 13. The first 2 classes (classes 0 and 1) represent profiles that do not collapse, where $\delta_T < d/2$ and classes 2 and 3 combined represent scenarios where collapse is imminent ($\delta_T \geq d/2$). Certain data augmentations[37] such as jitter, rotations and scaling were used to make the network more robust and prevent over-fitting. As an extension to the 3DmFV, an additional symmetric function, namely the mean was added to test the network performance. Adding more symmetric functions to the 3DmFV can add more information to the input data, but

can also increase computation time. Training with a larger input array in our case, however, showed almost no improvement in the prediction accuracy and was therefore not used. The 'max' and 'min' symmetric functions proved to be the best implementation in our use case scenario, confirming the observations from Reference.[26] The left and center plots in Figure 17 show a gradually decreasing validation loss and gradually increasing validation accuracy respectively. Both reach a level of saturation and stabilize after a certain number of epochs which shows that the network is not over-fitting or under-fitting.

The best overall test accuracy achieved here is 86%, with better accuracy seen in cases where collapse is

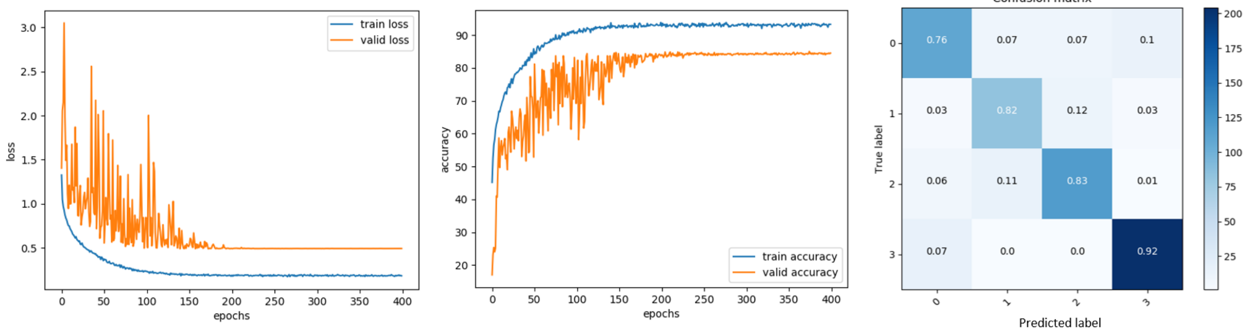


Figure 17: Collapse prediction for a lines and spaces use case. Left: Loss vs epochs during training and validation. Center: Accuracy of the training and validation data based on the number of epochs. Right: The confusion matrix for a four-class L/S pattern collapse classification

expected. The accuracy seen is in line with other 3D data classification networks from literature.[38–40] The main reason for this is that collapse is more probable when the features are denser i.e. have a smaller spacing d between the features. The network can easily identify such cases, since the 3DmFVs would have smaller non-zero regions in the 2D array used as the input. On the other hand, less dense features that actually do collapse are due to localized regions of larger aspect ratios ($AR = H/w$). These local 'hot-spots' along the line are not always picked up accurately by the network to make a decisive collapse prediction. The network therefore falls behind a little for these use cases and predicts them as features not likely to collapse. The line features chosen for generating the training data are assumed to be symmetric i.e., the lines are exactly the same on either side of the spacing. Also line features with a relatively low to medium level of roughness are considered, since the interpolation of the displacement fields along the length using the method proposed in Reference[19] holds true only for cases where the correlation length ξ is not very large.

2.2 Pillar patterns

2.2.1 Generation of training data

The capillary forces F_C acting on perfectly cylindrical/ideal pillar patterns were explained in Section 1.1 and Equation 4 was used to calculate those forces. The same equation can be used to calculate the capillary forces in EUV resists as well. Figure 18 shows the two main pillar configurations most commonly used in lithography applications. The capillary forces F_C described using Equation 4 act on the top sidewalls along the three-phase line as shown by the colored arrows. The capillary interaction forces at the top of the pillar are responsible for the bending, deformation or collapse of pillars during the rinsing stage. These forces are dependent on the pillar radius R, surface tension force γ, rinse liquid contact angle θ and pattern pitch p. Variations in the cross-sectional dimensions at the top of the feature can lead to unbalanced capillary forces causing collapse. Since the forces are based on the proximity to other pillars, they only act in a direction that has another pillar in close proximity. In an orthogonal configuration, the capillary forces act in four perpendicular directions and not along the diagonals since the diagonals are further apart ($\sqrt{2}p$ vs. p). On the other hand, for a hexagonal configuration, there is an equal distance between the pillars along one diagonal resulting in another opposing pair of capillary forces. This makes such a pillar configuration more prone to collapse compared to an orthogonal configuration.

For predicting collapse in pillar profiles using machine learning (ML), training data that captures the stochastic effects influencing resist profiles needs to be generated. Figure 18 shows the 3D profiles of pillars in an

orthogonal and hexagonal configuration. The pillar profiles are approximated as circular cross-sections and are discretized along the height. The discretization is used to vary the cross-sectional diameters creating undercut profiles in order to mimic the feature roughness seen in experimental data.[18,41] A four-pillar arrangement is used for the collapse study and prediction where the cross-sectional diameters are randomly generated for each of the four pillars. The cross-sections are plotted for a single pillar at six different height locations along the thickness.

A training data set of 6000 input arrays containing the cross-sectional profiles of each pillar similar to the

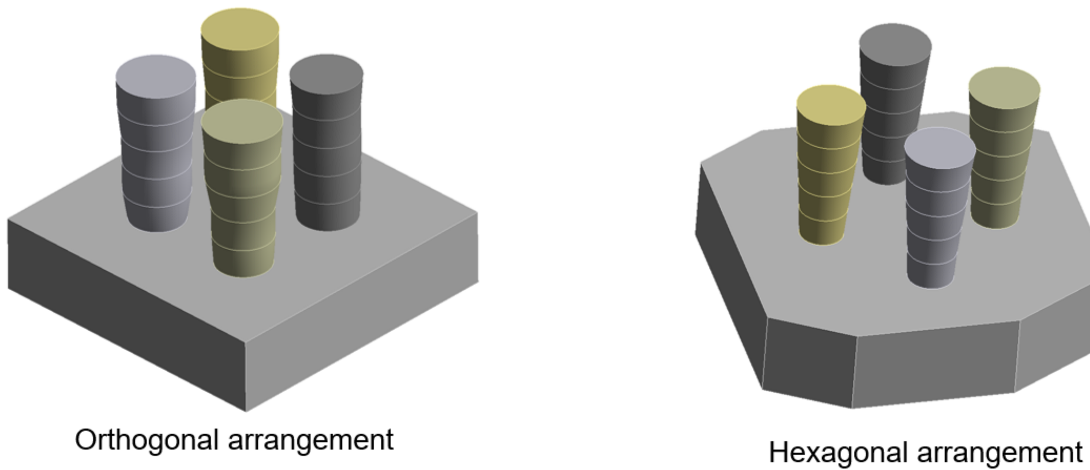


Figure 18: Three-dimensional representations of undercut pillars in an orthogonal arrangement (left) and a hexagonal arrangement (right)

one shown in Figure 18 is used for training a two-class classification CNN. The validation and test sets each contain 450 data points. The two labels are split evenly within the data so as to generate a balanced data set for training and testing. The dimensional and material properties of the pillars used during training are given below:

- Pillar diameter d = [15, 30] nm
- Height = 35 nm
- Pitch = [25, 70] nm
- Feature density ratio: $d/pitch \leq 0.6$
- Correlation length ξ = [7, 11] nm
- Young's modulus E = 0.2 GPa[29]
- Poisson's ratio ν = 0.4[30]

The material properties including E and ν given above are taken from literature[29,30] and experimental data and used in simulations for generating the required training data. The thickness/height of the resist is fixed at 35 nm with the pillar diameters and pitch varying within the ranges mentioned above. The data is generated to simulate mostly dense features i.e. where $d/pitch \leq 0.6$, since denser features are more prone to collapse. The thickness of each discretized unit (3D cylindrical disc) is in the range of the correlation length ξ to account for the thickness variations more accurately.

2.2.2 Collapse prediction

In the current study, a CNN is used for training and predicting collapse in pillar arrays for EUVL. The hyper-parameters used for training the CNN are identical to the ones used for predicting collapse in L/S and were described in Section 2.1.3. The training is initially performed using a CNN with 3D convolutions and the input is constructed as a 3D input array ($200 \times 200 \times 6$) which contains the six cross-sectional data. The trained network is tested on a test set which yields the two-class confusion matrix. Class 0 represents a pattern that does not collapse, while class 1 represents a collapsed pattern. The maximum accuracy achieved here is 83% with a total training time of $\sim$ 2 hours. 3D convolutions help with feature extractions in three dimensions but are generally slower.

Another approach using a CNN with 2D convolutions and multiple channels was used to investigate for possible network performance improvements. The multiple channels in this case represent the six different cross sections of the pillars. Data augmentation strategies were used here to make the network more robust and prevent over-fitting. The network architecture is shown in Figure 19 with the final prediction results shown in Figure 20 in the form of a confusion matrix. Two different input array sizes of $200\times 200\times 6$ and $400\times 400\times 6$ were used to see if there was any improvement in prediction accuracy by increasing the input size. The first two dimensions of the array specify the number of pixels for the pillar in the XY plane and the third dimension represents the number of cross-sections along the thickness (Z direction). Their prediction accuracy were almost the same at 97% and 98% respectively. The CNN with 2D convolutions and data augmentation showed better results in terms of prediction accuracy for the same input. The training time was also reduced significantly to $\sim$ 20 minutes. The 2D CNN performed better and would be the preferred choice overall for such a problem. The pillars considered in this study are centered around a fixed vertical axis and therefore the 3D CNN classification network would not benefit much from the use of 3D convolutions. 3D convolutions are used to learn the differences in the features along three dimensions, whereas 2D convolutions do the same in two dimensions. The 3D CNN also has more network parameters and is therefore slower in learning the features compared to a 2D CNN.

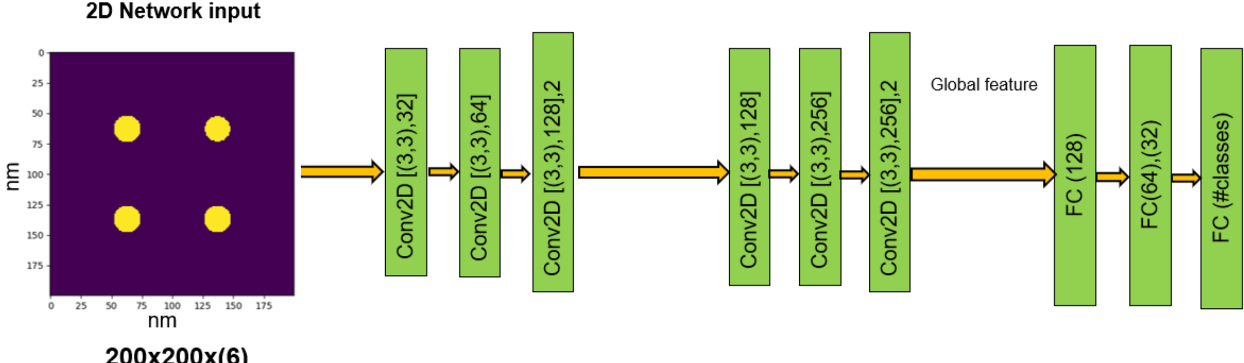


Figure 19: CNN network architecture using 2D convolutions for predicting collapse in orthogonally arranged pillars using a 2D input array (200×200) with 6 channels.

After establishing that a 2D CNN is more efficient in predicting collapse in pillar features, the same network was used for a hexagonal configuration to test for its prediction accuracy. Here Figure 21 demonstrates the network collapse prediction accuracy. The hexagonal configuration of pillars, as stated previously, is less stable due to the additional capillary forces acting along the diagonal pillars. The prediction accuracy is in line with the one seen for the orthogonal pillar configuration.

The results show how photoresist collapse can be predicted using machine learning for the two main use cases/patterns in photolithography. The trained model can be used to predict collapse behavior in commercial lithographic simulators by simply providing the 3D resist profiles generated after chemical development as an input. Simulating an outcome directly using a rigorous method like FEM would be otherwise very time consuming and computationally intensive. Integrating such a trained network into a modern lithographic simulator could

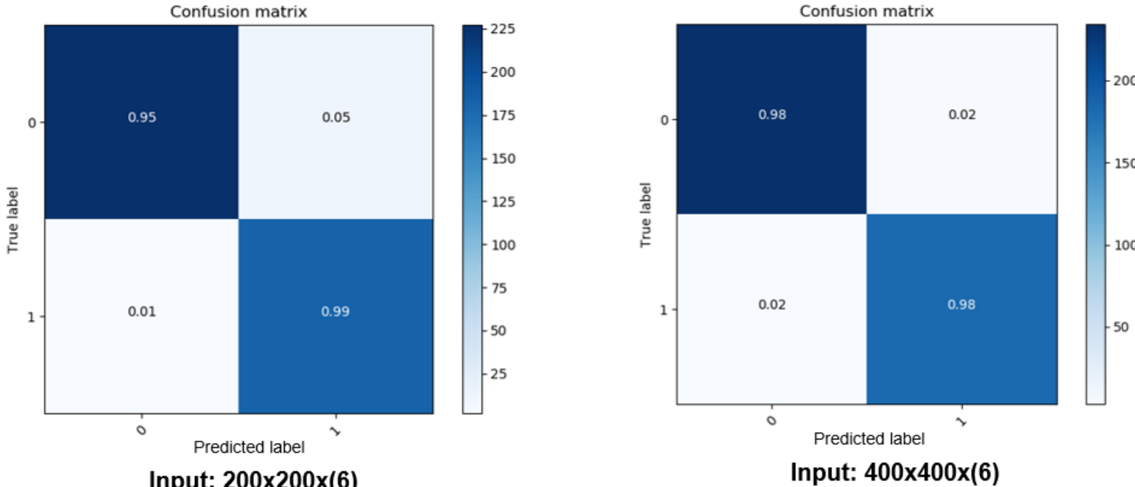


Figure 20: Confusion matrices from a 2D CNN with different input sizes for a two-class classification of pattern collapse in orthogonally arranged pillars with inputs of size $200\times200\times6$ (left) and $400\times400\times6$ (right).

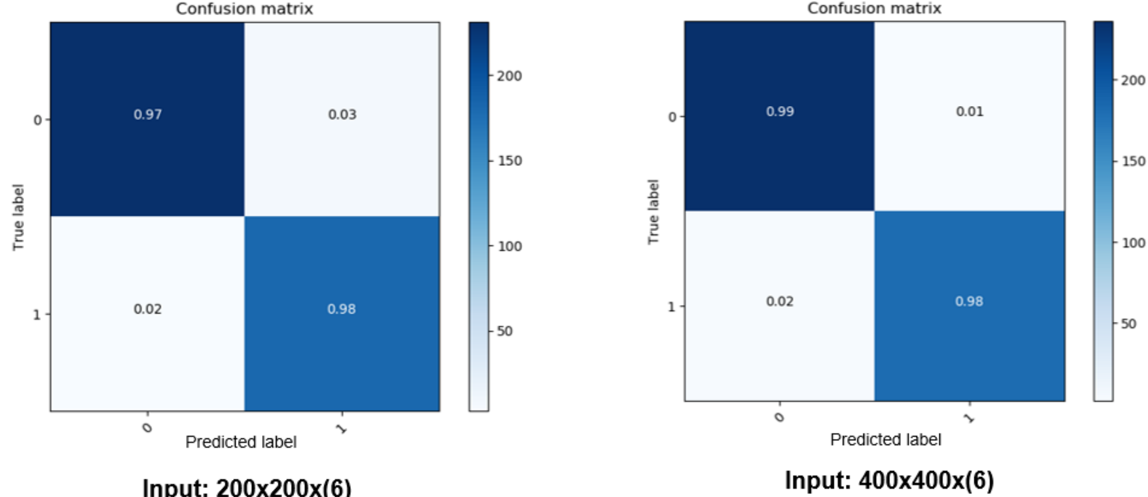


Figure 21: Confusion matrices from a 2D CNN with different input sizes for a two-class classification of pattern collapse in hexagonally arranged pillars with inputs of size $200\times200\times6$ (left) and $400\times400\times6$ (right).

open the doors for a more accurate understanding of pattern stability and aid in manufacturing approaches to make photoresist patterns more stable and prevent defects at the wafer level well in advance.

3. SUMMARY AND OUTLOOK

Models and simulation methods used to determine deformation and collapse in deep ultraviolet lithography (DUVL) cannot always help in predicting collapse in EUVL due to the numerous stochastic effects impacting photoresist processing. In EUVL there is an increased risk of collapse and pattern instability due to local regions of a rough profile with increased aspect ratios. Moreover, resist features in EUVL tend to have an undercut due to lower light absorption at the bottom of the resist making the pattern more unstable. Directly using a rigorous FEM model to predict collapse probabilities would be a slow and computationally intensive process. To circumvent this issue, a new machine learning (ML) based approach was developed to predict collapse probabilities in EUVL. Collapse for a L/S use case was predicted by using a novel approach combining 3D modified Fisher vectors (3DmFV) with a convolutional neural network (CNN). The 3D resist profiles in the form of point clouds were converted into a 2D array using 3DmFVs and passed as an input to a CNN for collapse prediction and classification. Another approach was used for pillar patterns, where the 2D cross sections were stacked into multi-channel 2D input arrays for a CNN. The network helped predict collapse in two different pillar configurations, namely the orthogonal and hexagonal configurations. The results from this study can help anticipate pattern stability directly after the simulated resist profiles are generated using modern lithographic simulators by integrating the trained network within the simulation environment.

Roughness does not scale linearly with reducing feature sizes and generally has a much higher order of magnitude compared to the feature dimensions for EUV resists. The impact of LWR can be studied using the results of this investigation as a foundation to help understand a more complete picture of collapse, particularly in EUV lithography. These stochastic effects are detrimental and critical to lithographic performance and hence need to be simulated more intensively.

REFERENCES

[1] Chini, S. F. and Amirfazli, A., "Understanding pattern collapse in photolithography process due to capillary forces," *Langmuir* **26**, 13707–13714 (Aug 2010).

[2] Nealey, P., Stoykovich, M., Yoshimoto, K., and Cao, H., "Nanolithographic polymer structures: Mechanical properties," in [*Encyclopedia of Materials: Science and Technology*], Buschow, K. J., Cahn, R. W., Flemings, M. C., Ilschner, B., Kramer, E. J., Mahajan, S., and Veyssière, P., eds., 1–9, Elsevier, Oxford (2003).

[3] Chandra, D. and Yang, S., "Stability of high-aspect-ratio micropillar arrays against adhesive and capillary forces," *Accounts of Chemical Research* **43**(8), 1080–1091 (2010). PMID: 20552977.

[4] Kralchevsky, P. A., Paunov, V. N., and Ivanov, I. B. N. K., "Capillary meniscus interaction between colloidal particles attached to a liquid-fluid interface," *Journal of Colloid and Interface Science* **151**, 79–94 (1992).

[5] Naulleau, P., Bhattarai, S., and Neureuther, A., "Understanding extreme stochastic events in euv resists," *Journal of Photopolymer Science and Technology* **30**(6), 695–701 (2017).

[6] Alzubaidi, L., Zhang, J., Humaidi, A. J., Al-dujaili, A., Duan, Y., Al-Shamma, O., Santamaría, J., Fadhel, M. A., Al-Amidie, M., and Farhan, L., "Review of deep learning: concepts, cnn architectures, challenges, applications, future directions," *Journal of Big Data* **8** (2021).

[7] Sarker, I. H., "Deep learning: A comprehensive overview on techniques, taxonomy, applications and research directions," *SN Computer Science* **2**(6), 1–20 (2021).

[8] Sathya, R. and Abraham, A., "Comparison of supervised and unsupervised learning algorithms for pattern classification," *International Journal of Advanced Research in Artificial Intelligence* **2** (Feb 2013).

[9] Petrillo, K., Huang, G., Ashworth, D., Georger, J., Ren, L., Cho, K.-Y., Montgomery, W., Wurm, S., Kawakami, S., Dunn, S., and Ko, A., "Line width roughness control and pattern collapse solutions for EUV patterning," *Proceedings of SPIE - The International Society for Optical Engineering* **7969** (Mar 2011).

[10] Garidis, K., Pret, A. V., and Gronheid, R., "Mask roughness impact on extreme UV and 193nm immersion lithography," *Microelectronic Engineering* **98**, 138–141 (2012). Special issue MNE 2011 - Part II.

[11] Patsis, G., Drygiannakis, D., Constantoudis, V., Raptis, I., and Gogolides, E., "Stochastic modeling and simulation of photoresist surface and line-edge roughness evolution," *European Polymer Journal* **46**(10), 1988–1999 (2010).

[12] Bhattarai, S., Neureuther, A. R., and Naulleau, P. P., "Study of shot noise in photoresists for extreme ultraviolet lithography through comparative analysis of line edge roughness in electron beam and extreme ultraviolet lithography," *Journal of Vacuum Science & Technology B* **35**(6), 061602 (2017).

[13] Kozawa, T., Yamamoto, H., and Tagawa, S., "Relationship between line edge roughness and fluctuation of acid concentration in chemically amplified resist," *Japanese Journal of Applied Physics* **49**, 096506 (Sep 2010).

[14] Mack, C. A., "Reducing roughness in extreme ultraviolet lithography," in [*International Conference on Extreme Ultraviolet Lithography 2017*], Gargini, P. A., Naulleau, P. P., Ronse, K. G., and Itani, T., eds., **10450**, 104500P, International Society for Optics and Photonics, SPIE (2017).

[15] Williams, P., Lubnau, D., and Elliott, D., "Generation of fractal images and comparison of their psds with several models," in [*Twenty-Second Asilomar Conference on Signals, Systems and Computers*], **1**, 495–499 (1988).

[16] Mack, C. A., "Analytic form for the power spectral density in one, two, and three dimensions," *Journal of Micro/Nanolithography, MEMS, and MOEMS* **10**(4), 040501 (2011).

[17] Constantoudis, V., Patsis, G. P., Tserepi, A., and Gogolides, E., "Quantification of line-edge roughness of photoresists. ii. scaling and fractal analysis and the best roughness descriptors," *Journal of Vacuum Science & Technology B: Microelectronics and Nanometer Structures Processing, Measurement, and Phenomena* **21**(3), 1019–1026 (2003).

[18] Fallica, R., Stowers, J. K., Grenville, A., Frommhold, A., Robinson, A. P. G., and Ekinci, Y., "Dynamic absorption coefficients of CAR and non-CAR resists at EUV," in [*Extreme Ultraviolet (EUV) Lithography VII*], Panning, E. M., ed., **9776**, 977612, International Society for Optics and Photonics, SPIE (2016).

[19] Arava, R., *Prediction of Photoresist Displacement Fields Using Graph Neural Networks*, Master's thesis, Friedrich-Alexander-Universität Erlangen-Nürnberg, Germany (2022).

[20] Chini, S. F. and Amirfazli, A., "Collapse of patterns with various geometries during drying in photolithography: numerical study," *Journal of Micro/Nanolithography, MEMS, and MOEMS* **11**(3), 1 – 11 (2012).

[21] Battaglia, P. W., Hamrick, J. B., Bapst, V., Sanchez-Gonzalez, A., Zambaldi, V., Malinowski, M., Tacchetti, A., Raposo, D., Santoro, A., Faulkner, R., Gulcehre, C., Song, F., Ballard, A., Gilmer, J., Dahl, G., Vaswani, A., Allen, K., Nash, C., Langston, V., Dyer, C., Heess, N., Wierstra, D., Kohli, P., Botvinick, M., Vinyals, O., Li, Y., and Pascanu, R., "Relational inductive biases, deep learning, and graph networks," (2018).

[22] Xu, K., Zhang, M., Jegelka, S., and Kawaguchi, K., "Optimization of graph neural networks: Implicit acceleration by skip connections and more depth," (2021).

[23] ANSYS Inc., *Mechanical User's Guide*, 2021 r1 ed. (January 2021).

[24] Biber, P. and Strasser, W., "The normal distributions transform: a new approach to laser scan matching," in [*Proceedings 2003 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2003) (Cat. No.03CH37453)*], **3**, 2743–2748 vol.3 (2003).

[25] Lalonde, J.-F., Vandapel, N., Huber, D. F., and Hebert, M., "Natural terrain classification using three-dimensional ladar data for ground robot mobility," *Journal of Field Robotics* **23**(10), 839–861 (2006).

[26] Ben-Shabat, Y., Lindenbaum, M., and Fischer, A., "3DmFV: Three-dimensional point cloud classification in real-time using convolutional neural networks," *IEEE Robotics and Automation Letters* **3**(4), 3145–3152 (2018).

[27] Sánchez, J., Mensink, T., and Verbeek, J., "Image classification with the fisher vector: Theory and practice," *International Journal of Computer Vision* **105** (Dec 2013).

[28] Charles, R., Su, H., Kaichun, M., and Guibas, L. J., "Pointnet: Deep learning on point sets for 3D classification and segmentation," in [*2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*], 77–85, IEEE Computer Society, Los Alamitos, CA, USA (Jul 2017).

[29] Yoshimoto, K., Higgins, C., Raghunathan, A., Hartley, J. G., Goldfarb, D. L., Kato, H., Petrillo, K., Colburn, M. E., Schefske, J., Wood, O., and Wallow, T. I., "Revisit pattern collapse for 14nm node and beyond," *Advances in Resist Materials and Processing Technology XXVIII* (Apr 2011).

[30] Mülders, T., Stock, H.-J., Küchler, B., Klostermann, U., Gao, W., and Demmerle, W., "Modeling of NTD resist shrinkage," in [*Advances in Patterning Materials and Processes XXXIV*], Hohle, C. K., ed., **10146**, 147 – 157, International Society for Optics and Photonics, SPIE (2017).

[31] Simonyan, K. and Zisserman, A., "Very Deep Convolutional Networks for Large-Scale Image Recognition," *arXiv e-prints* , arXiv:1409.1556 (Sept. 2014).

[32] Liu, S. and Deng, W., "Very deep convolutional neural network based image classification using small training sample size," in [*2015 3rd IAPR Asian Conference on Pattern Recognition (ACPR)*], 730–734 (2015).

[33] Springenberg, J., Dosovitskiy, A., Brox, T., and Riedmiller, M., "Striving for simplicity: The all convolutional net," in [*ICLR (workshop track)*], (2015).

[34] Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., and Salakhutdinov, R., "Dropout: A simple way to prevent neural networks from overfitting," *Journal of Machine Learning Research* **15**(56), 1929–1958 (2014).

[35] Thanapol, P., Lavangnananda, K., Bouvry, P., Pinel, F., and Leprévost, F., "Reducing overfitting and improving generalization in training convolutional neural network (cnn) under limited sample sizes in image recognition," in [*2020 - 5th International Conference on Information Technology (InCIT)*], 300–305 (2020).

[36] Kingma, D. P. and Ba, J., "Adam: A method for stochastic optimization," in [*3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*], Bengio, Y. and LeCun, Y., eds. (2015).

[37] Perez, L. and Wang, J., "The effectiveness of data augmentation in image classification using deep learning," *ArXiv* **abs/1712.04621** (2017).

[38] Wang, P.-S., Liu, Y., Guo, Y.-X., Sun, C.-Y., and Tong, X., "O-CNN: Octree-based convolutional neural networks for 3D shape analysis," *ACM Trans. Graph.* **36**(4) (2017).

[39] Riegler, G., Ulusoy, O., and Geiger, A., "Octnet: Learning deep 3D representations at high resolutions," in [*Proceedings IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 2017*], 6620–6629, IEEE, Piscataway, NJ, USA (July 2017).

[40] Maturana, D. and Scherer, S., "Voxnet: A 3D convolutional neural network for real-time object recognition," in [*2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*], 922–928 (2015).

[41] De Simone, D., Blanc, R., Van de Kerkhove, J., Tamaddon, A.-H., Fallica, R., Van Look, L., Rassoul, N., Lazzarino, F., Vandenbroeck, N., Vanelderen, P., Lorusso, G., Van Roey, F., Charley, A.-L., Vandenberghe, G., Ronse, K., Lee, K., Lee, J., Park, S., Lim, C.-M., and Park, C.-H., "Staggered pillar patterning using 0.33 NA EUV lithography," in [*Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*], *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series* **10957**, 109570T (Mar. 2019).

Test Pattern Generation by Optimization of the Feature Space Signature

Francois Weisbuch[a], Jirka Schatz[a]

[a]GlobalFoundries, Wilschdorfer Landstraße 101, 01109, Dresden, GERMANY

ABSTRACT

Background: Well-designed test patterns are required to ensure a robust patterning involving mask, lithography and etch. They are expected to anticipate potential process challenges while representing well the layout diversity.

Problem: We expect good test patterns to have a high design space coverage with minimal redundancy. Is it possible to get optimal design contents while keeping a small footprint and following the design rules?

Approach: In this work, after defining specific optical and geometrical features and discretizing the pattern as a binary matrix, we propose to use the signature of the pattern in the feature space to assign a score measuring the usefulness of the pattern. The score is used as a cost function to drive an iterative optimization of the pattern shape based on a differential evolution algorithm.

Conclusion: We demonstrated how to generate compact test patterns with high design diversity customized to specific applications that should help to anticipate, represent or monitor well process challenges.

Keywords: Pattern generation, test pattern, image parameter space, pattern coverage, OPC, feature, differential evolution.

1. INTRODUCTION

Test patterns are a very important component of the design-to-mask process to ensure robust patterning. They have to fulfill three main functions: anticipate, predict, and monitor. First, they should help to validate the design rules and explore any potentially challenging design configurations. Then, when used to train models, especially machine learning-based ones, test patterns must be sensitive enough to the process to simulate and diverse enough to guarantee good stability in the predictions. Finally, test patterns can serve for process control to detect excursions (like pinching/bridging) or even as unit tests to improve design to mask procedures like Optical Proximity Correction (OPC) or Optical Rule Checks (ORC).

However, at the same time, test patterns are constrained by the area available on the mask, the metrology cycle time, and the measurable field of view. Therefore, it would be of high interest to be able to design test patterns that combine high conciseness and are tailored for a particular application. In other words, the problem boils down to the question: how to get the best design content for a given footprint? This means achieving high design coverage with low redundancy while respecting the design rules constraints.

The first approach consists in generating parametric test patterns manually or collecting clips of known hotspots. Unfortunately, this method is quite slow, tedious, and not exhaustive. Another possibility is to extract the optimal test patterns either from a broader set of artificially generated layouts [1] or directly by analyzing production layouts using complex clustering [2, 3], similarity grouping [4], or data reduction techniques [5, 6, 7]. But these methods may not be exhaustive if the database to profile is not representative enough of the design space. In addition, the patterns are limited to known configurations not suitable to anticipate new patterning challenges. However, with the help of machine learning, new solutions emerge to augment the design space covered by known layouts [8, 9]. Finally, the test patterns can be synthesized directly without any layout support, only constrained by design rules [10].

In this work, we do not pose the problem as an optimal selection but instead, as an optimal generation of test patterns to address both the challenge of design space coverage of the test pattern and low redundancy in the design space. The proposed method does not require any a priori knowledge about the design style but imposes to follow the design rule constraints like minimum pattern width, space, and area. The idea is to start from a discrete representation of a test pattern and then extract relevant optical [11, 12] and geometrical features [13]. We define a cost function that analyzes the extent

38th European Mask and Lithography Conference (EMLC 2023), edited by
Uwe F.W. Behringer, Jo Finders, Proc. of SPIE Vol. 12802, 1280206
© 2023 SPIE · 0277-786X · doi: 10.1117/12.2675601

of the pattern signature in the feature space. By minimizing the cost function, we propose to generate highly diversified test patterns valid in terms of design rules but not necessarily realistic in terms of functional design.

2. TEST PATTERN DEFINITION

As mentioned in the introduction, we intend to generate test patterns whose applications can target different steps of the patterning process. We propose a general method, but for the sake of clarity and without restricting its generality, this work will be focused on a specific technology and layer application. In addition, we want each test pattern to have a small footprint -typically a few square micrometers- to be easily embedded in a reticle and fit in a typical SEM field of view.

2.1 Process context

In this study, we consider a metal layer for a positive tone development (PTD) resist associated with an advanced binary dark field mask. The lithography is based on a 193nm immersion scanner operating at 1.35 numerical aperture using a polarized free-form illumination. .

2.2 Test pattern parametrization

The test patterns are created to follow typical standard CMOS Manhattan design rules, which means the pattern edges must be either horizontal or vertical. Moreover, to ensure the robust printability of the patterns, we enforced a few important geometrical design rules like the pattern minimum width, minimum space, minimum area, and minimum enclosure. To get a simple parametrization of the test patterns taking into account these constraints, we propose to define each test pattern as a binary matrix where each elementary element, or "pixel" has a size equal to the minimum width/space (see Fig. 1a). The zeros of the matrix indicate the mask background, whereas the ones indicate the mask main feature transmission. The detailed geometry of the test pattern is described in Table 1. Note that the mask is assumed to have a value of 0 (background) around the pattern. Besides, for simulation purposes when extracting the cutline (see section §3), the matrix needs to be extended with a margin of 2 pixels per side.

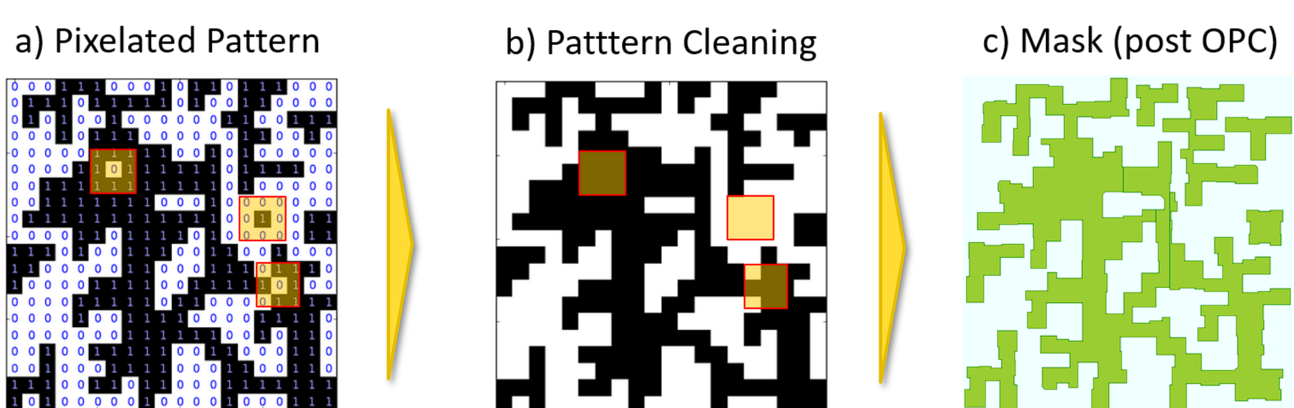


Fig. 1: (a) *Test pattern parametrization as a binary matrix. (b)Pattern cleaning operation to remove 'kissing corners' and to enforce minimum area/enclosure design rules. (c) Applying Optical Proximity Correction to improve the pattern printability.*

Table 1: Test pattern constraints and lithography process conditions for the simulation.

Constraints	Value
Min width (px_size)	50 nm
Min space (px_size)	50 nm
Field of view (n_px.px_size²)	1 um²
Mask type	Adv binary
Starting density (n_px_black/n_px)	0.5

Constraints	Value
Min area/enclosure	3 px
No "kissing corners"	-
Mask background	0
Mask main feature	1
Mask polarity	Dark field

Constraints	Value
Source	Free form
Scanner	193i
Numerical Aperture	1.35
Focus/Dose	nominal
Optical Proximity Correction	yes
Resist	PTD

The test pattern creation starts with a random initialization of the binary matrix where the average density of the pattern, i.e., the relative ratio of ones in the matrix, can be controlled. Then the test pattern is cleaned to enforce the geometrical constraints of minimum area/enclosure and avoid any "kissing corners" configuration. The design is intended to be printed on a photoresist after applying an Optical Proximity Correction (OPC) for the final mask shapes to achieve a good pattern fidelity. Due to the complexity of the pattern design, and the absence of regularity -like long straight lines-, we decided to use no positive or negative Sub Resolution Assist Features (SRAF) too challenging to place properly.

2.3 Test pattern simulation

In the next section, we plan to extract specific optical and geometrical properties from the patterns in order to characterize the response of the lithography and etch processes. These properties are derived from the aerial images and the 3D resist profiles of the pattern simulated with the help of S-Litho™ rigorous simulation software from Synopsys from a model calibrated beforehand for the metal layer (see Table 1 and Fig. 2b,c). This work requires many simulations of test patterns. In order to get a reasonable run time with acceptable accuracy, the mask was considered thin (no mask 3D effects are taken into account) and the mask, the aerial image, and the resist were represented with a grid of 3nm in all directions. Note that the resist 2D profile at a given height in the photoresist can be approximated by thresholding the aerial image (see Fig. 2) with the advantage of being much faster to compute than the rigorous full 3D resist profile.

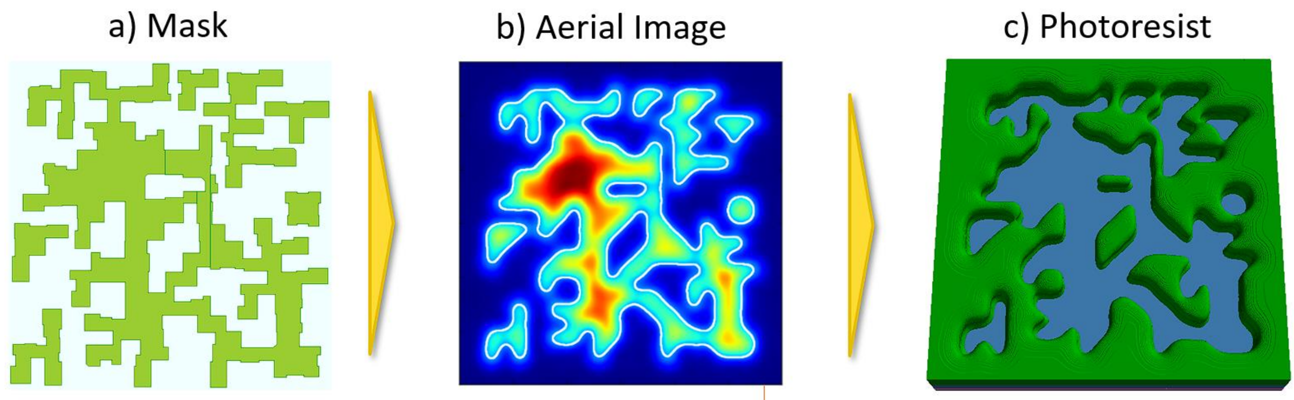


Fig. 2: *Using rigorous simulation to generate from the mask(a) the Aerial Image of the pattern(b) and the 3D resist profiles(c). The resist 2D profiles can be approximated by thresholding the Aerial Image (white contour).*

3. PATTERN FEATURE SCORE

3.1 Pattern feature definition

The complexity and diversity of the test patterns obtained with the method described in the previous section are considerable even for patterns of limited size and following specific design rules. To simplify the description of a pattern, it is convenient to represent it in a feature space. Although the features can be extracted directly from the design geometry [2], it is more efficient to rely on features that are related to the physics involved in the patterning. Features based on aerial image intensity measurement (IPS) [11, 12] or optical model kernels [4] are well adapted to represent the lithography process. More advanced methods use machine learning to derive feature vectors directly from the aerial image [6]. Additional features were introduced [8, 13] to include the resist 3D profiles – like resist top loss, or diffusion effects occurring during the resist bake - or local density important to describe the etch process. More features like mask,OPC or multi-layer can also be combined to increase the feature space [3].

Ideally, the features should be independent, i.e., uncorrelated, and sufficient to make it possible to discriminate between any pair of patterns. The features used in this work associate optical and geometrical properties extracted from the pattern. The definition of this set of features is mainly empirical, arbitrary, depending on the aspect of the patterning process to represent, and cannot be considered rigorously as a basis to represent exactly any pattern.

The first set of features illustrated in Fig. 3 are good descriptors of the lithography process. They are extracted only at specific points of the aerial image that are located along a contour of reference (see Fig. 3a), usually corresponding to an approximation of the resist bottom 2D profiles derived by thresholding the aerial image at a constant value. These points of interest correspond to the edges of the pattern. A cutline of length 200nm is moved along the reference contours to analyze locally at each point the variation of the aerial images normal to the contour. We define the following features shown in Fig. 3b,c,d: the slope, the local range of the aerial image, as well as the variation of the aerial image slope when changing the scanner focus. The values of these features are evaluated at each point of the reference contours and are displayed on the contour plots of Fig. 3e,f,g with a colored scale.

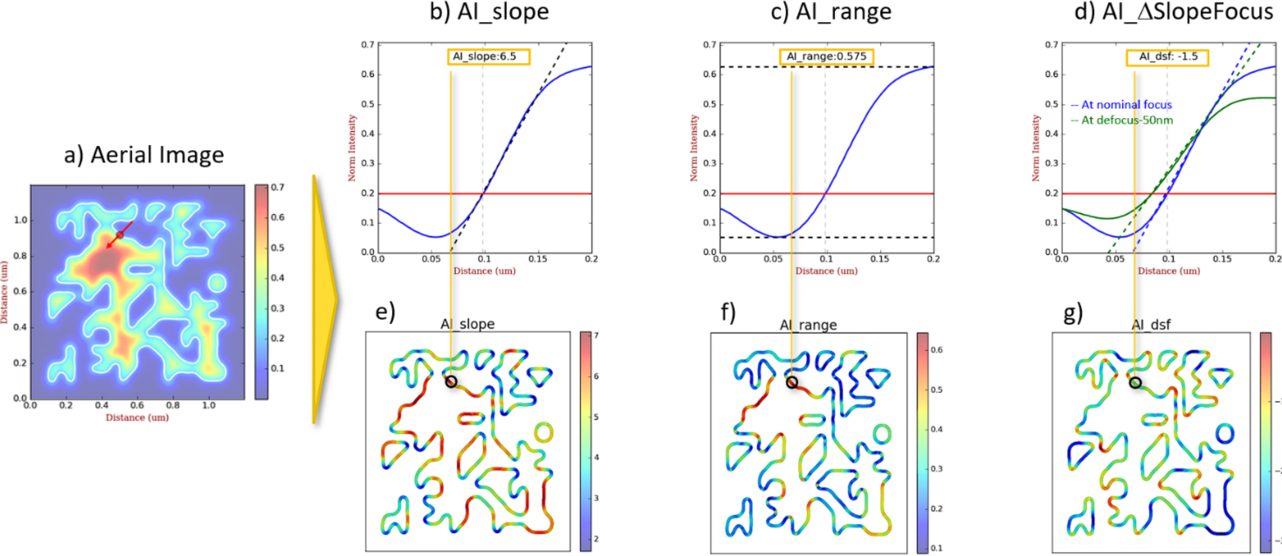


Fig. 3: *Lithography features extracted from the aerial image of the pattern(a). A cutline (red) normal to the reference contour (white) derives locally the optical properties of the aerial images (b,c,d). The contour plots (e,f,g) are generated by moving the cutline along the reference contours while evaluating the different features.*

The second set of features, based on the 2D geometry of the resist bottom contours (see Fig. 4a), are important resist properties influencing both the lithography and the etch processes. Again, these features are evaluated at each point of the reference contours. The first two features, R_int and R_ext (Fig. 4b,c), are sensing locally the width and the space of the resist perpendicularly to the contour integrating over a small-angle cone (typically 15°) to avoid a too sharp transition of

the feature values and reflect the spatial resolution of the etch. The last feature measures the curvature of the resist contour (see Fig. 4d).

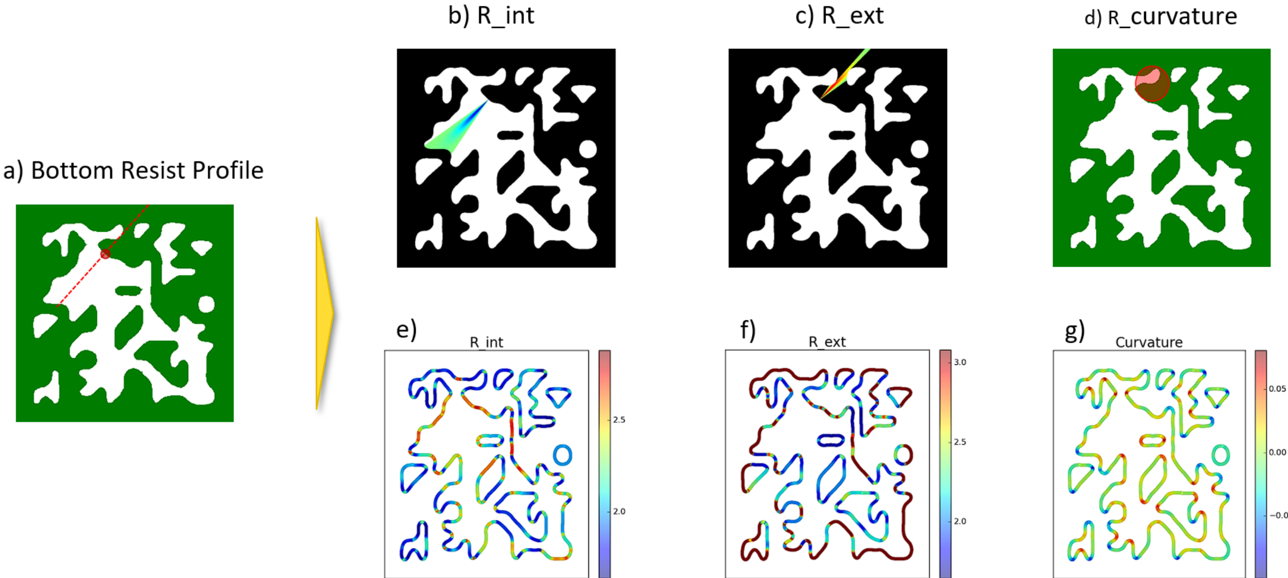


Fig. 4: *Lithography and etch features extracted from the resist bottom contours(a). The geometrical features (b,c,d) are extracted perpendicularly to the resist bottom contours. The contour plots (e,f,g) are generated by moving the cutline along the bottom contours while evaluating the different features.*

The last features are defined to approximate the 3D resist profiles of the patterns (see Fig. 5a). They represent important properties for the etch process. The R3D_density feature estimates, at a point along the resist bottom contour, the local resist density by convolving the resist 3D profiles with a Gaussian of chosen width as shown in Fig. 5b. The R3D_swa assesses the resist sidewall angle (see Fig. 5c) by computing the minimum distance between two resist contours simulated at the bottom and close to the top of the resist, as shown in Fig. 5d. It is a good estimator of resist top-loss or resist pinching.

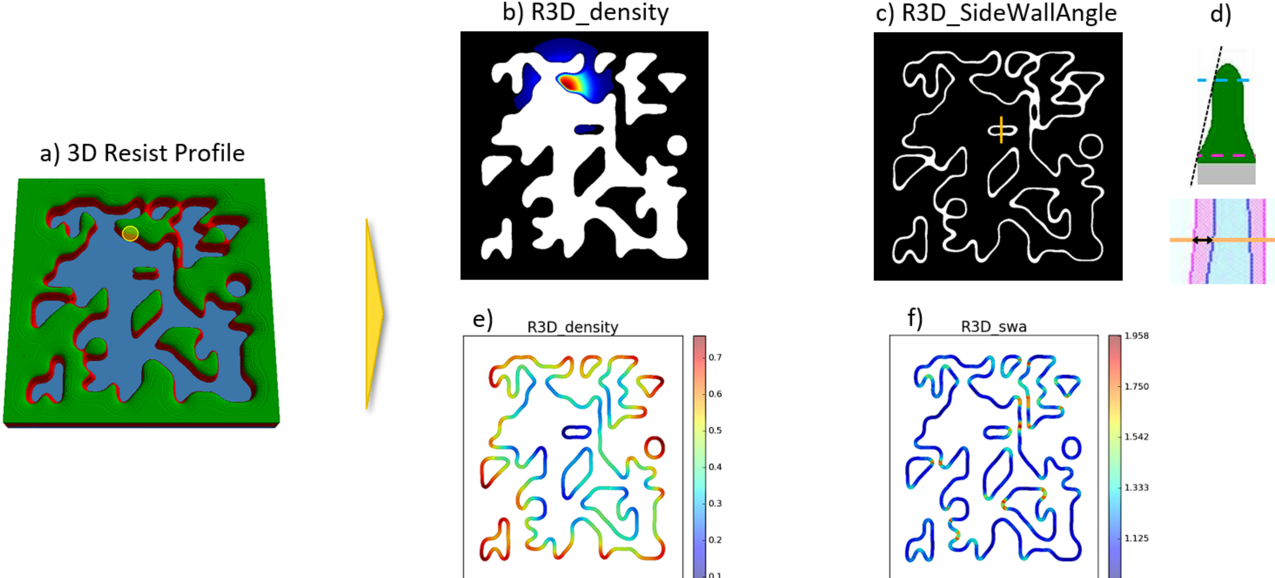


Fig. 5: *Etch features approximating the resist 3D profiles (a). The resist density (b) and resist sidewall angle (c) are evaluated locally for each point along the bottom profiles to generate the contour plots (e,f). The resist sidewall angle is estimated by measuring the distance between two resist contours simulated at different heights in the resist(d).*

3.2 Pattern Feature Space

Considering a given pattern and its response at each point along the reference contours to each of the optical and geometrical features, one can populate a multi-dimensional space representing the pattern signature with respect to these features. But such a signature is difficult to manipulate and too abstract to understand. One solution to simplify the data is to reduce the dimension of the space with principal component analysis (PCA) or singular value decomposition (SVD). The drawback is to lose the connection with the physical properties of the features. Instead, we propose to project the n-dimensional feature space into a set of 2D-subspaces very convenient to visualize and handle. Fig. 6a shows an example of a 2D-subspace where the values of a feature are plotted against another feature values while moving along the reference contours. This generates a cloud of points, characteristic of the pattern in this 2D-feature space. Fig. 6b shows a part of the 2D-feature space matrix combining all the 2D-spaces. This is a helpful way to see all the various facets of the pattern signature in the feature space. It is important to verify that there is no strong correlation between the various features.

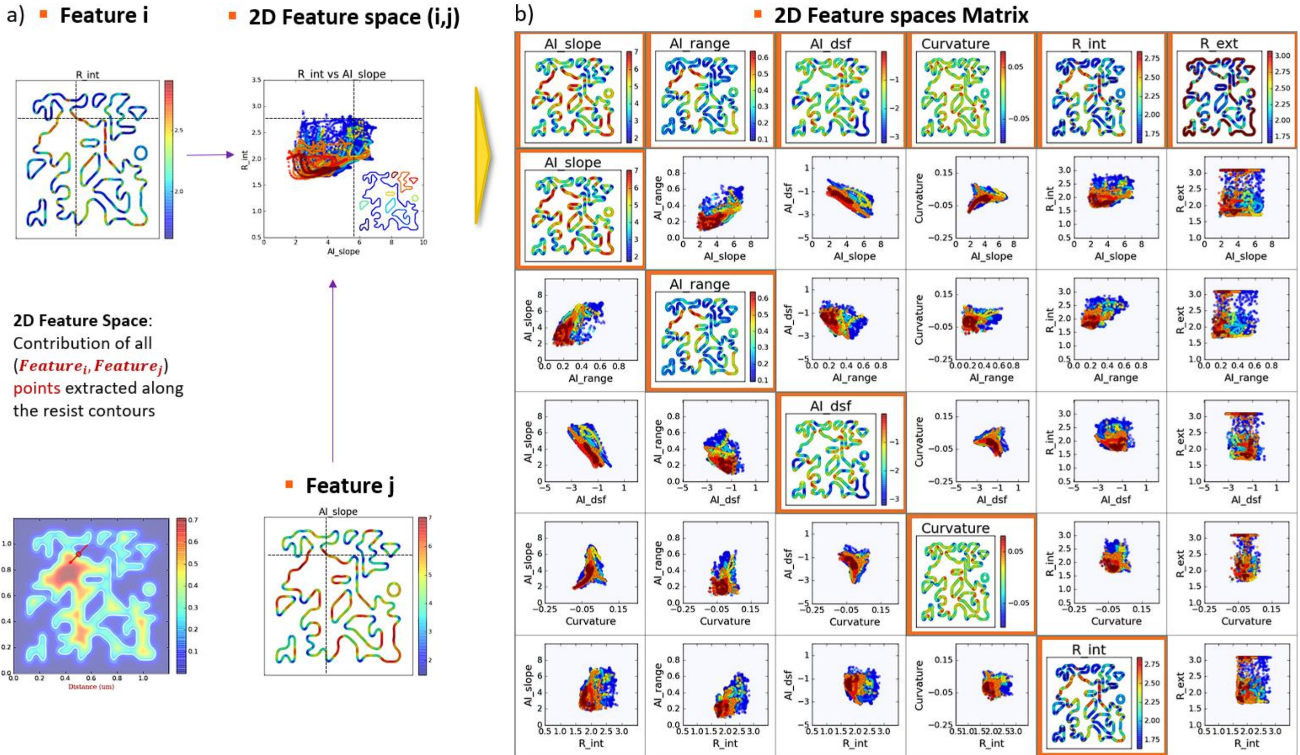


Fig. 6: *Pattern signature projection in a 2D-feature space. (a) Plotting feature$_i$ against feature$_j$ while moving along the reference contours to generate a cloud of points. The colors identify the contours they are coming from. (b) Extract of the 2D-feature space matrix*

At this point, one can wonder how to use the pattern signature in the feature space to assess the quality of a pattern in terms of design representativity or diversity. Fig. 7 compares the spread of three patterns with increasing design complexity in a few 2D-feature spaces. As expected, the simple CD Test pattern (see Fig. 7a), has a much narrower distribution of points in all 2D-feature spaces compared to the more complex patterns (see Fig. 7b,c). However, one cannot clearly identify which of the two last patterns has the better feature space coverage even when trying to use data reduction techniques like Single Value Decomposition with two components. This observation stresses the need for a more qualitative approach.

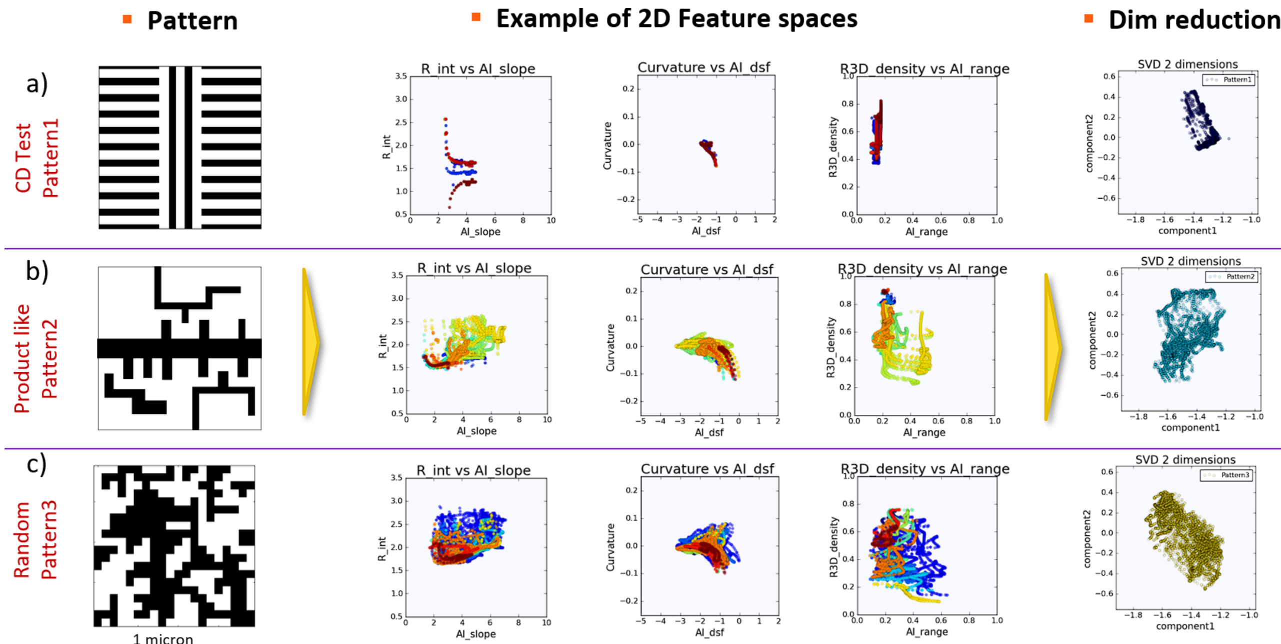


Fig. 7: *Example of pattern signatures in 2D-feature spaces. The design complexity of the patterns is increasing from (a),(b) to (c), as well as, the spread of the pattern signatures in the feature spaces.*

3.3 Pattern Feature Score

In this section, we propose a metric to measure qualitatively the signature of a pattern in a 2D-Feature space. This metric, associated with a pair of features (i,j), is named the feature score and is noted $fScore_{i,j}$. It is a combination of two positive sub-scores that measure two essential properties of the pattern signature: the coverage of the space and the redundancy of the points in the space, as illustrated in Fig. 8a. The computation of the scores requires two types of normalization to be independent of the pair of features and the geometry of the pattern.

- First, each feature is normalized with a min-max scaling to bring all point values into the range [0,1] interval. The min and max values are defined a priori, estimating the possible variations of each feature. This ensures that the scores coming from various pairs of features are homogeneous.
- Then, the 2D-feature space is discretized as a 1000x1000 matrix and filled with the points of the pattern signature.
- In addition, each score must be normalized with respect to the number of points populating the feature space. Indeed, the choice of the simulation grid modulates the number of points sampled around the reference contours (see section §2.1). Furthermore, the perimeter of the reference contours, and the number of points associated, are also strongly dependent on the pattern geometry.

Based on this data transformation, the *coverage score* (see Fig. 8b) measures the filling coverage of the points generated by the pattern in the 2D-feature space. This score is high if the points are well distributed in the space, ideally maximizing the distance of each point to its closest neighbors. The calculation used a distance map based on the 2D-feature space matrix to speed up the computation. The assumption is that different parts of the feature space correspond to different design geometries, and therefore a good coverage of the space by the points means a good design representativity.

On the other hand, *the redundancy score* (see Fig. 8c) is low if the points overlap in the feature space. The score is computed by convolving the points with a gaussian to generate a density map in the feature space and integrating the map to get an average density. The score applies a penalty if points accumulate in the same region of the feature space, with the assumption that some design configurations would be overrepresented, and some parts of the design would not be optimally used.

Finally, the two scores are combined to form the feature score, expressing the representativity of the test pattern with respect to the pair of features. The score is set to negative for the purpose of optimization by minimizing its value (see section §4).

$$\boldsymbol{fScore_{i,j} = f(Score_{coverage}, Score_{redundancy})}$$

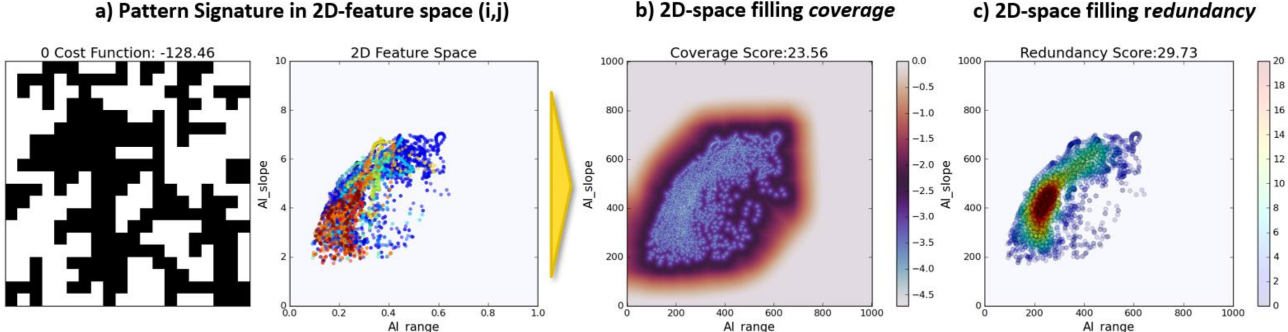


Fig. 8: *Pattern score in 2D-feature space. (a) Example of a pattern and its signature in a 2D-feature space. (b) The space-filling coverage measures the distribution of the points in the space. (c) The space-filling redundancy is based on the point density distribution in the space The red color highlights regions of high point density, i.e., with high redundancy.*

3.4 Pattern Feature Score: Matrix and Vector

We have established that by projecting the signature of a pattern in a given 2D-feature subspace, we can derive the score of the pattern in this space (see Fig. 9a). If we repeat this operation for all the combinations of pairs of features (i,j), we can build a symmetric matrix containing all the scalars $fScore_{i,j}$ (see Fig. 9b). This matrix, called $fScore\ Matrix$, characterizes completely and qualitatively a pattern.

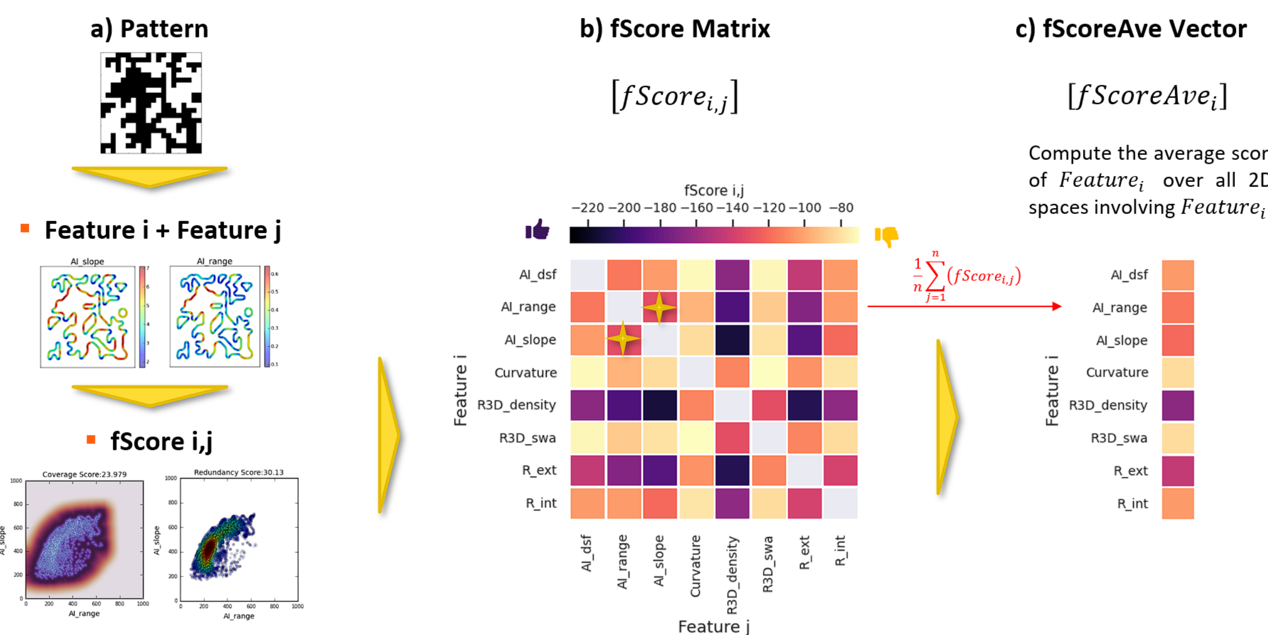


Fig. 9: *Feature Score Matrix and Vector. (a) Given a pattern and two features (i,j), the signature of the feature values in the 2D-feature space is used to compute the pattern feature score* $\boldsymbol{fScore_{i,j}}$. *(b) Computing the* $\boldsymbol{fScore_{i,j}}$ *of all feature combinations to derive the symmetric* $\boldsymbol{fScore\ Matrix}$. *The smallest values (black) correspond to better scores, the stars refer to the example on the left. (c)Derivation of the* $\boldsymbol{fScore\ Vector}$ *by averaging the columns of the* $\boldsymbol{fScore\ Matrix}$.

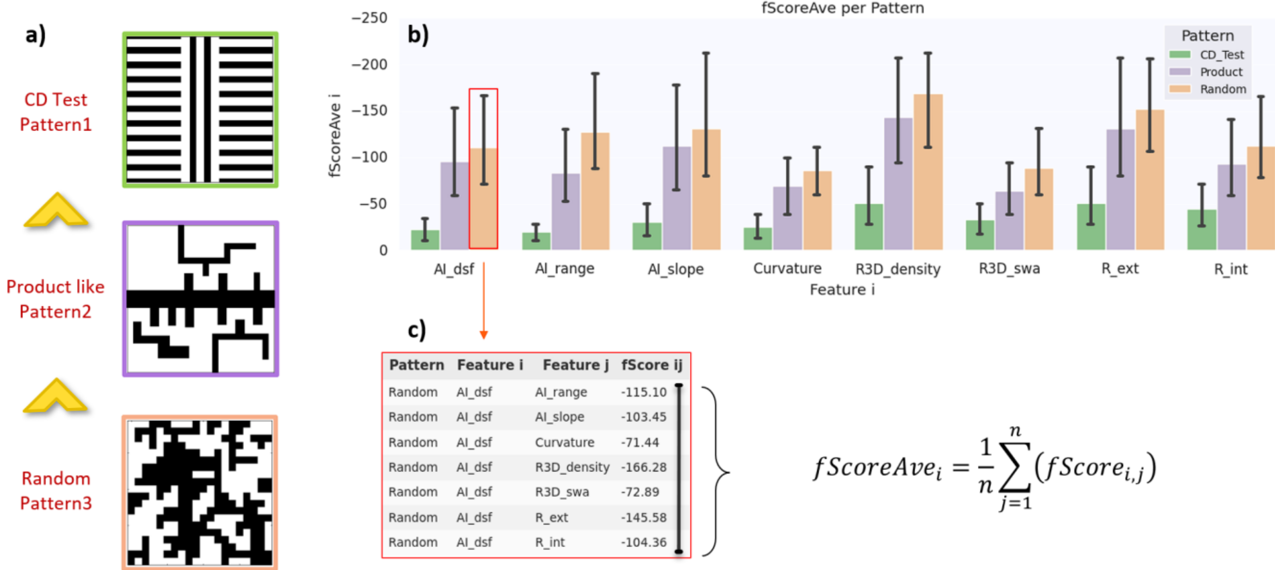


Fig. 10: *Example of pattern ranking based on the* $\boldsymbol{fScore\ Vector}$. *(a) Patterns with increasing design complexity. (b) Plotting the average feature scores* $\boldsymbol{fScoreAve_i}$ *of the 3 patterns. For each feature$_i$ on the x-axis, the error bar indicates the range of the* $\boldsymbol{fScore_{i,j}}$ *evaluated on all the 2D-spaces involving feature$_i$ (c).*

But we can go a step further to have a more synthetic description of a pattern by averaging all the $fScore_{i,j}$ of the same row to derive a vector whose size is the number of features, called the $fScoreAve\ Vector$ (see Fig. 9c). With this vector, it is easy to compare the responses of various patterns concerning a feature. Coming back to the example of Fig. 7, it is now possible to rank the 3 patterns relative to each feature, as shown in Fig. 10 and prove that the 'random pattern3' has a better representativity, whatever the features.

4. PATTERN SHAPE OPTIMIZATION

Thanks to the $fScore$ metric, we have an objective way to evaluate the usefulness of a pattern with respect to a pair of features (i,j). If we consider this metric as a cost function, it is now possible to optimize the shape of a pattern to get a better $fScore_{i,j}$.

4.1 Differential evolution

Starting with a pattern parametrized as a binary matrix with specific geometrical constraints (see section §2.2), the goal is to optimize its shape with respect to the $fScore_{i,j}$ cost function. The parameters are the 'pixels' of the binary matrix that can be flipped in accordance with the design rule constraints. Due to the number of degrees of freedom of the problem and the computation time to evaluate the feature scores, it is not possible to use the gradient of the cost function. Besides, the optimization space is highly non-convex and discontinuous but still discrete. To address this combinatorial optimization problem, we selected a differential evolution algorithm [14, 15], which is an iterative stochastic global optimization technique that is not gradient-based and can converge quickly even with very large solution spaces. The advantage of this algorithm is that it can run multiple searches in parallel by co-optimizing multiple vectors at the same time improving at the end the capability to escape local minima.

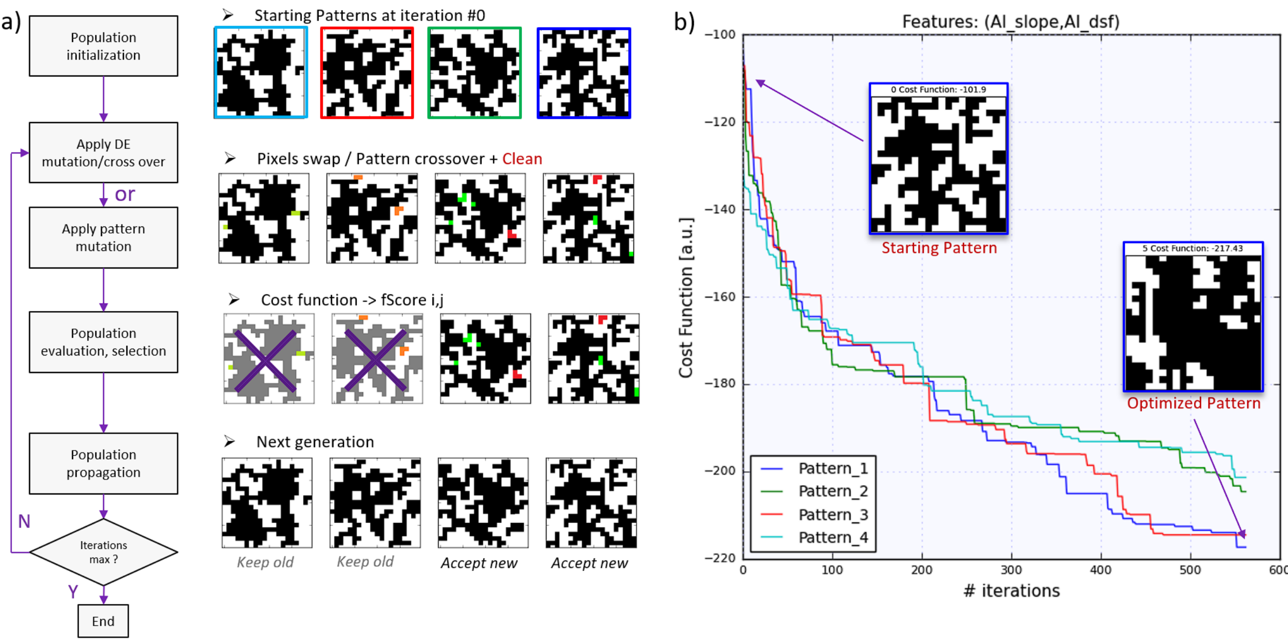


Fig. 11: *(a) Flow chart of the modified differential evolution algorithm showing the evolution of the binary patterns during one iteration. The pixels highlighted in red/orange are removed, whereas the green ones are added during the mutation/crossover step. (b) Convergence of each pattern over many iterations. The plateaus of the curves indicate repeated subsequent rejections of new pattern candidates.*

The flow of the algorithm is based on the differential evolution algorithm with a modification to allow an additional pattern mutation. The principle, like the genetic algorithm, is to start from a population of vectors (patterns in our case) and

generate new trial vectors to increase the diversity of the perturbed populations and then select only the new vectors that yield an improvement of the cost function. The flow illustrated in Fig. 11 can be summarized as follow:

- Initialization: the initial population is generated randomly in accordance with the design rules and a global ‘pixel’ density constraint. A population of at least 10 patterns is required to achieve a good convergence.
- Loop over all members of the population in parallel to modify and evaluate each candidate (target vector) one by one. The perturbation of the candidate pattern is randomly decided to be done either directly with differential evolution or with a single pattern mutation.
    - Differential evolution: First, a target vector is selected, and a muted vector is generated by adding the weighted difference vector between two population members to a third one, all chosen randomly. Then a crossover is performed between the muted vector and the target vector to get a trial vector. The crossover operation randomly blends two binary matrices by swapping single pixels or blocks of pixels between them.
    - Pattern mutation: Generate the new trial vector by randomly flipping one or multiple pixels of the pattern.
    - Selection/Propagation: if the trial vector improves the cost function of the candidate, then the modification is propagated to the next generation; otherwise, it is rejected, and the old candidate is kept.
- Go to the next generation until the maximum number of iterations is reached.

Note that each time a trial pattern is generated and before it is evaluated, it is cleaned to enforce the geometrical constraints.

4.2 Analysis of the optimization

Since the pattern optimization is driven by only two features, it is interesting, given a reference pattern, to compare the final optimized patterns generated by different pair of features. To this end, starting from the same reference pattern, we run a few optimizations with the cost function based on a few pairs of features described in Fig. 12. Then we did a comparison of the $fScore\ Matrices$ of the optimized patterns with respect to the reference one, to assess how the various single $fScore_{i,j}$ are modified after the optimization. Fig. 12 shows the relative evolution of the $fScore\ Matrices$. As expected, there is a clear improvement of the $fScore_{i,j}$ involving the features (i,j) used for the optimization, but for other $Feature_k,\ with\ k \neq i, j$, the score can worsen. This observation confirms that the choice of features is critical for generating a pattern that gets the best response or sensitivity for a given application.

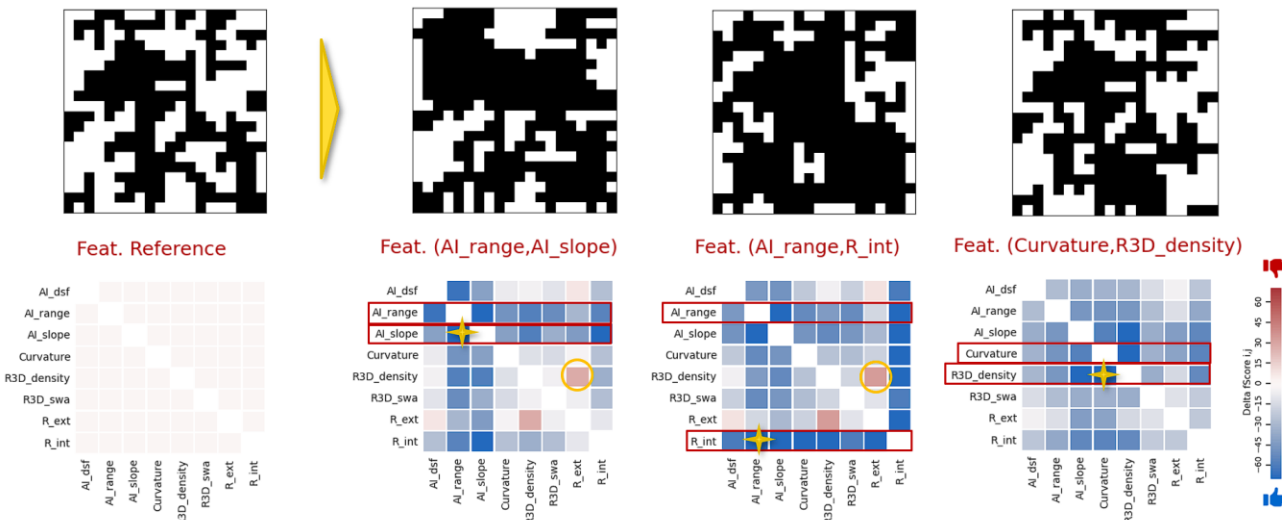


Fig. 12: *Evolution of the* $\boldsymbol{fScore\ Matrix}$ *of a reference pattern after optimization for different pairs of features. We subtracted the* $\boldsymbol{fScore\ Matrix}$ *of the reference pattern to get the relative* $\boldsymbol{fScore\ Matrices}$*, with the blue color indicating an improvement of the single* $\boldsymbol{fScore_{i,j}}$*. The yellow stars indicate the pair of features used for the optimization. The yellow circles highlight fScores that deteriorate after optimization.*

4.3 Application

In this section, we present an application of the optimization method to generate a test pattern intended for lithography hotspots or focus monitoring. The best candidate features we selected are the AI_dsf feature (see Fig. 3d) which is sensitive to a change of slope of the aerial image with a defocus change, and the R3D_swa feature (see Fig. 5c) sensitive to resist profile changes and as such well suited to catch resist top loss, pinch or bridging.

The result of the pattern optimization is presented in Fig. 13 which shows the simulation of the resist 3D profile through focus. We can observe a concentration of patterning challenges like resist bridging at negative defocus and resist pinching with severe top loss at positive defocus. With the help of advanced SEM contour metrology [16], provided enough measurements are taken, it is possible to evaluate the deviation of measured resist profiles of a monitoring pattern with respect to ideal reference golden contours. Fig. 13 shows that the generated compact pattern could react with high sensitivity to lithography focus variations to catch slight drift from the scanner or measure intra-field or intra-wafer focus variations. Lastly, one could take advantage of the strong asymmetry between the resist profiles across focus to set up a SEM image-based focus monitoring.

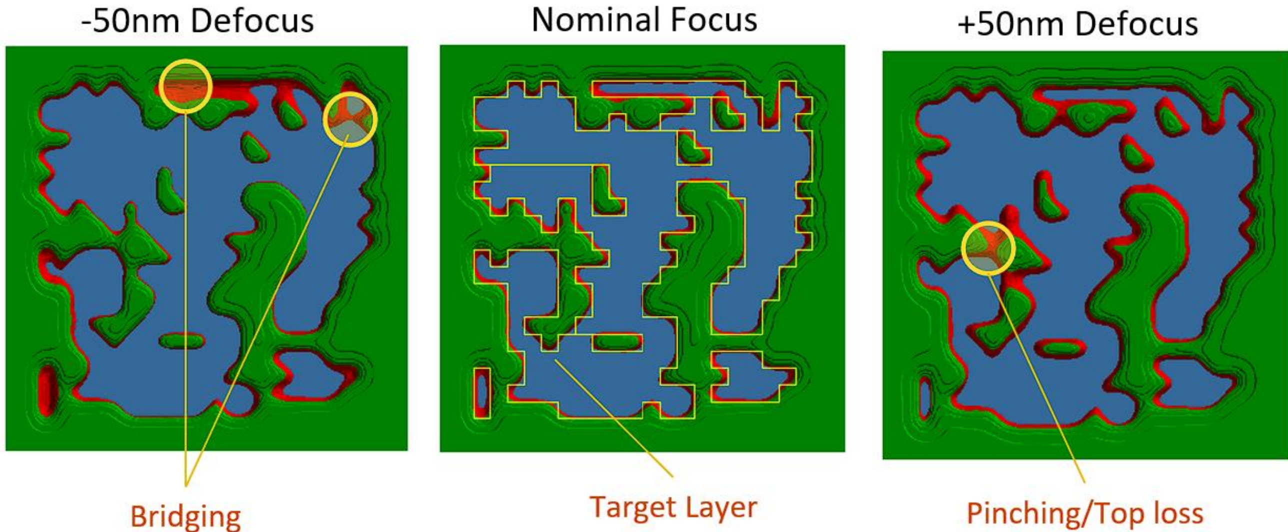


Fig. 13: *Example of a test pattern optimized for lithography process control. Rigorous simulation of the 3D resist profiles across focus showing patterning failures at defocus like bridging, pinching or resist top loss.*

5. CONCLUSIONS

We presented a flow to generate specific test patterns by optimizing their signature in a feature space. For that, we defined a limited number of features, specifically adapted to represent the lithography and etch processes based on aerial images and resist profiles. We showed how the features, extracted along reference simulated resist contours, are combined to populate as a set of 2D-feature spaces. We derived feature scores measuring the coverage and redundancy of a pattern signature in any 2D-feature space. The feature score is used as a cost function to drive the shape optimization of patterns with the help of a differential evolution algorithm. This flow is a first step to generate compact test patterns with high design diversity customized to specific applications that should help to anticipate, represent or monitor well process challenges.

The next step involves enhancing the definition of features as well as improving the convergence of the optimizations. Finally, this work could reach more applications if it also supports curvilinear patterns.

6. REFERENCES

[1] A. Hamouda et al., "Enhanced OPC recipe coverage and early hotspot detection through automated layout generation and analysis," in *Proceedings Volume 10147, Optical Microlithography XXX; 101470R (2017) https://doi.org/10.1117/12.2260769*.

[2] S. Shim and Y. Shin, "Topology-oriented pattern extraction and classification for," in *Journal of Micro/Nanolithography, MEMS, and MOEMS, Vol. 14, Issue 1, 013503 (January 2015). https://doi.org/10.1117/1.JMM.14.1.013503*.

[3] W. Wei Zhang et al., "Modeling Sampling Strategy Optimization by Machine Learning Based Analysis," in *978-1-6654-2079-2/21/$31.00 ©2021 IEEE*.

[4] E. Tejnil, "Application of optical similarity in OPC model calibration," in *Proc.SPIE 10147, Optical Microlithography XXX, 101471Q (24 March 2017); doi:10.1117/12.2258054*.

[5] G. Viehoever, B. Ward and H.-J. Stock, "Pattern selection in high-dimensional parameter spaces," in *Proceedings Volume 8326, Optical Microlithography XXV; 832618 (2012) https://doi.org/10.1117/12.916352*.

[6] Y. Feng Yaobin et al., "Layout pattern analysis and coverage evaluation in computational lithography," in *Optics Express Vol. 31, Issue 5, pp. 8897-8913 (2023) •https://doi.org/10.1364/OE.485206*.

[7] J. Zhu et al., "Automatic generation of representative and diversified pattern samples from a full chip layout," in *Proceedings Volume 12495, DTCO and Computational Patterning II; 124951B (2023) https://doi.org/10.1117/12.2662024*.

[8] C. Gangmin Cho et al., "Integrated Test Pattern Extraction and Generation for Accurate Lithography Modeling," in *EEE TRANSACTIONS ON SEMICONDUCTOR MANUFACTURING, VOL. 35, NO. 3, AUGUST 2022*.

[9] P. Kareem et al., "Layout Pattern Synthesis for Lithography Optimizations," in *IEEE Transactions on Semiconductor Manufacturing, vol. 33, no. 2, pp. 283-290, May 2020, doi: 10.1109/TSM.2020.2982989*.

[10] H. Yang et al., "DeePattern: Layout Pattern Generation with Transforming Convolutional Auto-Encoder," in *DAC '19: Proceedings of the 56th Annual Design Automation Conference 2019June 2019Article No.: 148Pages 1–6 https://doi.org/10.1145/3316781.3317795*.

[11] K. Patterson et al., "Improving model-based OPC performance for the 65-nm node through calibration set optimization," in *Proceedings Volume 5756, Design and Process Integration for Microelectronic Manufacturing III; (2005) https://doi.org/10.1117/12.601166*.

[12] T. Roessler, B. Frankowsky and O. Toublan, "Improvement of empirical OPC model robustness using full-chip aerial image analysis," in *Proceedings Volume 5256, 23rd Annual BACUS Symposium on Photomask Technology; (2003) https://doi.org/10.1117/12.518055*.

[13] F. Weisbuch, A. Lutich and J. Schatz, "Introducing etch kernels for efficient pattern sampling and etch bias prediction," in *Journal of Micro/Nanolithography, MEMS, and MOEMS, Vol. 17, Issue 1, 013505 (March 2018). https://doi.org/10.1117/1.JMM.17.1.013505*.

[14] K.Price and R. Storn, "Differential Evolution – A Simple and Efficient Heuristic for global Optimization over Continuous Spaces,," in *J. Glob. Optim. 11(4), 341–359 (1997)*.

[15] S. Campbell et al., "“Review of numerical optimization techniques for meta-device design," in *Opt. Mater. Express 9(4), 1842–1863 (2019)*.

[16] F. Dettoni et al., "The image lab sandbox, pulling image computing in wafer fab metrology environment," in *EMLC 2023, to be published.*.

A Deep Learning Facilitated Approach for SEM Image Denoising Towards Improved Contour Extraction for 1D and 2D Structures

Bappaditya Dey[a], Stewart Wu[b], Victor Blanco[a], Samir Bhamidipati[c], Amr Essam[d], Kevin Ahi[b], Sandip Halder[a], and Germain Fenger[b]

[a]imec, Kapeldreef 75, 3001 Leuven, Belgium
[b]Siemens Digital Industries Software, Inc. (United States)
[c]Siemens Digital Industries Software, Inc. (India)
[d]Siemens Digital Industries Software, Inc. (Egypt)

ABSTRACT

Flow from circuit design to manufacturing a mask is a complex process. A mask is used to print a particular design feature. Extraction of accurate CD contour geometries from ADI (after developed inspection) SEM images plays a pivotal role for a qualitative lithographic process as well as to verify device characterization in aggressive pitches. In our previous work, it has been shown how deep learning based de-noising is helping to improve the contour detection accuracy. We analysed and validated our result with noisy/denoised image pair for categorically different geometrical patterns, such as, L/S (line-space), T2T (tip-to-tip), pillars with different scan types etc., by using a programmable tool (Calibre®SEMSuite™) for contour extraction and further analysis metrics. The comparative analysis demonstrated that denoised images have significantly higher confidence analysis metrics, reduced number of missing patterns as well as false bridges against raw noisy images while keeping the same parameter settings for both data inputs. We have demonstrated that our proposed method is capable to extract contours on the body of the noisy SEM images with accuracy in close proximity with design data. By combining these advanced algorithms as options in Calibre®SEMSuite™, users would be able to process large amount of information for data cleaning, classification & further model calibration intelligently.

Keywords: C D-SEM, metrology, denoise, image quality enhancement, machine learning, deep learning, supervised, unsupervised, Power Spectral Density (PSD), defect inspection, contour detection, Dynamic random-access memory (DRAM), bit-line-periphery (BLP), storage node landing pad (SNLP)

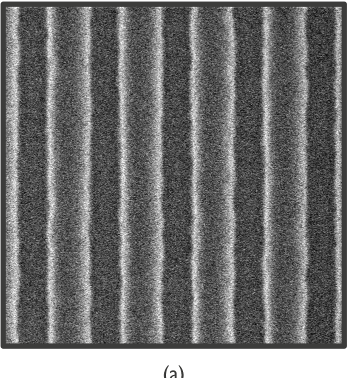
1. INTRODUCTION

With the continuous development of advanced manufacturing techniques in the semiconductor sector, the size of the physical features of integrated circuits (IC) continues to shrink year after year. This leads to an urgent demand for the advancement of metrology and inspection techniques to ensure the quality as well as the yield of end products. Typically, in the semiconductor manufacturing lines, wafer quality inspection tasks such as overlay, edge placement errors, and critical dimension (CD) are performed by using optical microscopes or scanning electron microscopes (SEM). When the scale of pattern size is in the nanometers range, electron beam-based inspection techniques are usually preferred, as they can achieve very high resolution and have been shown to be very effective in detecting defects of fabricated patterns. However, SEM images inherently contain a substantial level of background noise as shown in Fig. 1. According to Ref. 1, an analytical noise model for generalized SEM image has been modeled as five cascaded signal conversion stages as: (1) primary emission, (2) secondary emission, (3) scintillation process, (4) photocathode, and finally (5) signal multiplication in the photomultiplier tube. The authors also assumed that signal conversion at each stage to be Poisson distributed. The analytical expression of the total SNR can be expressed as:

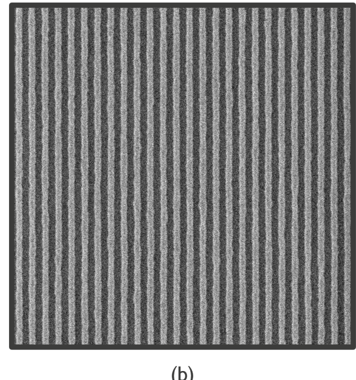
$$SNR_5 = \left(\frac{\sigma_1^2}{\mu_1^2} + \frac{\sigma_2^2}{\mu_1 \mu_2^2} + \frac{\sigma_3^2}{\mu_1 \mu_2 \mu_3^2} + \frac{\sigma_4^2}{\mu_1 \mu_2 \mu_3 \mu_4^2} + \frac{\sigma_5^2}{\mu_1 \mu_2 \mu_3 \mu_4 \mu_5^2} \right)^{-0.5} \quad (1)$$

where, σ_k is the standard deviation of each Poisson distributed stage k, and μ_k is the corresponding expected value. Their experimental analysis demonstrates that (a) SNR_5 generally decreases at each subsequent stage k, (b) the influence of each stage on SNR_k decreases accordingly. Therefore, the first two stages (primary emission and secondary emission) enforce maximum contribution on SNR_{total}.

38th European Mask and Lithography Conference (EMLC 2023), edited by
Uwe F.W. Behringer, Jo Finders, Proc. of SPIE Vol. 12802, 1280207
© 2023 SPIE · 0277-786X · doi: 10.1117/12.2680884



(a)



(b)

Figure 1: Noisy CD-SEM images of line structures. (a) Pitch 90. (b) Pitch 32

Noisy images of measured sites often cause erroneous defect detection and inaccurate metrology, which can lead to serious problems in the later steps. To reduce the noise and achieve higher contrast, frame averaging of the same inspected structures can be applied. Nevertheless, a high number of averaged frames can significantly affect the development time constraints and cost of production. Moreover, high averaging can cause shadowing artifacts and the increase of CD shrinkage,[2] as well as wafer damage. Therefore, in order to attain good metrology quality and accurate defect detection, while preserving a reasonable yield, image quality enhancement (IQE) techniques are usually employed.

ML and deep learning (DL) architectures and techniques have been applied in various tasks in the production line including overlay metrology,[3–5] wafer leveling and alignment,[6] defect detection and classification,[7,8] SEM images denoising,[9] and mask optimization,[10–13] and they have shown great improvement compared to conventional algorithms both in terms of accuracy and speed. In this research, we have demonstrated the application of our deep learning denoiser assisted framework towards enabling SEM contour extraction possible for all the edges in raw noisy DRAM SEM images with bit-line-periphery (BLP) and storage node landing pad (SNLP) patterns. We have demonstrated that our proposed method is capable to extract contours on the body of the noisy SEM images (across different geometrical patterns) with accuracy in close proximity with design data. By combining these advanced algorithms as options in Calibre®SEMSuite™, users would be able to process large amount of information for data cleaning, classification and further model calibration intelligently.

The rest of this paper is organized as follows. Section 2 briefly discusses some related research works on image denoising and contour extraction. Section 3 presents in detail our proposed methods including the theoretical background, data preparation strategy, training steps involved, obtained results, and finally discussions. Section 4 introduces significant applications which can be aided by the proposed deep learning denoiser. In Section 5, we extensively discussed proposed deep learning denoiser assisted framework for robust SEM contour extraction across different geometrical patterns. Finally, section 6 concludes the paper.

2. RELATED WORK

In this section, we briefly walk through and discuss some recent related research methodologies and techniques for denoising/ enhancing the quality of SEM images. For the application of conventional ML techniques in image denoising, Block Matching and 3D filtering (BM3D) proposed by Dabov et al. is one of the most popular ones.[14] It is an internal statistics method based on the assumption that natural images regularly consist of similar blocks. BM3D reduce the noise of an image by clustering repeated patterns together to form 3D cylinder-like shapes and sequentially filtering them. The drawback of this method is the high computational cost during prediction. Recently, there has been a variety of DL denoisers that have been able to outperform the traditional ML-based ones. Jain et al. were the first authors who proposed a denoiser based Convolutional Neural Networks (CNNs),[15] where the denoising was considered as a supervised regression task, and the DL networks try to minimize the loss between predicted and clean ground truth pixels of the images. This is a simple but effective idea and is still applied by many State-of-The-Art (SOTA) denoising techniques. Regarding the methods that

do not need clean data, Jaakko Lehtinen et al. introduced noise2noise scheme,[16] which is based on the premise of conditional independence of noise on the signal in an image. In this scheme, a pair of noisy images with identical structures but sampled independently were used to train the CNNs. The regressor network tries to predict the noisy pixels of one image with the corresponding input pixels from the other image. Although the mapping from noisy pixels to noisy pixels was performed, the model will eventually converge due to the aforementioned premise. Building upon the idea of noise2nosie, Hairong Lei et al. proposed a predesigned additive noise algorithm to augment and reduce the noise of input SEM images, which has proved to be very effective when the training dataset is small.[9] This algorithm adds various Gaussian noise level to the input images to increase the dataset as well as enhance the noise coverage of the input images, then it will train the denoiser by applying noise2noise model in either denoise-to-next or denoise-to-best scheme. In Ref. 2, a framework based unsupervised DL for image quality enhancement called uMLIQE was introduced. The idea of this framework is similar to the noise2noise scheme, which is based on the assumption that the statistical expectation of the noisy observation is equal to the real signal. As a result, the model trained to predict noisy corresponding data pair will finally converge and is able to predict the clean signal. However, the results produced by this method reported some artifacts and the loss of contrast in the denoised images.

In Ref. 17, the authors have presented a deep learning-based method to realize accurate and stable CD measurement from SEM images with low SNR. They constructed the dataset used in deep learning training using a CD-SEM image model designed from mathematical and statistical line edge roughness (LER). This approach has limitations due to induced artifacts. N. Chaudhary et al. Ref. 18 also proposed supervised Deep Convolutional Neural Network (DCNN) architectures such as LineNet1 and LineNet2, to denoise and predict edges from low-dose scanning electron microscope (SEM) images concurrently. They have simulated a dataset of 100,800 SEM images constructed with the help of the Thorsos method and the ARTIMAGEN library. In Ref. 19, N. Chaudhary et al. also proposed supervised DCNN architectures as LineNet1 and LineNet2, to denoise and predict edges from low-dose SEM images concurrently. However, they used simulated single-line and multiple-line SEM images with Poisson noise and other artifacts using the ARTIMAGEN library developed by the National Institute of Standards and Technology (NIST). Their proposed approach facilitated edge estimation in multiple-line images and reduced the memory usage for edge estimation in single-line images with compromised accuracy. In Ref. 20, the authors proposed a deep learning denoising of SEM images towards accurate LER measurements. Towards denoiser model training and deployment, they used synthesized SEM images. Finally, experimental results were compared with SOTA methods. In Ref. 21, the major contributions can be demonstrated as proposing a GAN-CNN-based framework towards blind denoising of images. The framework involves two stages. In the first stage, GAN is utilized to build paired training datasets. In the second stage, CNN architecture is employed for denoising. The utilization of GAN architecture primarily helped to model the unknown noise distribution. The authors proposed in Ref. 22, a deep CNN architecture for image denoising. The proposed architecture, integrated with batch normalization and residual learning, is capable to handle the blind Gaussian denoising with unknown noise levels. In addition, the proposed approach can also be applied for single image super-resolution with multiple upscaling factors as well as JPEG image deblocking with different quality factors. N. Chaudhary et al. demonstrated the effectiveness of deep learning towards SEM image denoising and contour geometry estimation.[23] Conclusion can be derived from their experimentation as, simulated SEM training dataset of limited capabilities combining with appropriate parameter variation can be proven useful towards applications like real SEM image denoising, roughness estimation, and contour geometry estimation. However, authors did not validate the effectiveness of their proposed approach on real SEM images with (a) low dose, (b) thin resist and (c) unfavorable topographical and/or contrast differences.

In Ref. 24, the authors proposed an analytical linescan model (ALM), inspired by a physical model for linescan generation, to restrict the possible shape of a linescan from inherently noisy SEM images. Their proposed approach enabled reliable detection of edges in noisy SEM images without using any conventional image filtering technique as well as without altering the true roughness of the features by means of inverting a calibrated model. The usage of ALM generally depends on (a) materials property and geometric explanation of the feature (such as CD, Pitch, SWA, etc.). Using this information ALM estimates the edge positions/linescan. However, reverse modeling is also possible as, given the measured linescan, the model could estimate/predict the feature geometry as discussed above. In Ref. 25, the authors addressed off-line metrology to study the acceptable noise level by decreasing the dose in simulated top-down SEM-like images. They also quantified the limit of required edge-based images for estimating LER-related parameters. However, these traditional techniques have certain limitations as in Ref. 24, the usage of ALM for edge detection depends on certain pre-defined parameters as well as feature edge positions are approximated towards best-fitted measured linescan. In Ref. 25, authors used synthetic images as well as randomly generated line edges. They also processed those images by adding Poisson-distributed noise to every

pixel of the noise-free images, after choosing an average electron density.

However, after careful assessment of all the above research works and datasets used for denoising, the quantification of noise for SEM images does not fit truly Gaussian or Poisson distributions Ref. 26, as noise bias varies with metrology settings and sample properties.

3. UNSUPERVISED LEARNING BASED DENOISER

To transcend the limitations mentioned above in the supervised and/or semi-supervised method, we applied another denoising model based on unsupervised deep learning. This model does not require any clean ground truth data, and the training, as well as validation, can be done with only one single noisy image. We have used only real wafer data (noisy ADI SEM images of different CD/Pitch, of certain resist profiles and different thicknesses) for the training, validation, and test, for categorically different geometrical patterns such as L/S (line-space), C/H (contact-hole), pillars with different scan types, SRAM structures, Logic structures and DRAM (SNLP+BLP) 2D-structures.

3.0.1 Denoising Model

The network structure of our proposed denoiser is identical to the one presented in Ref. 27. The proposed denoiser of SEM images is based on the work introduced by Krull et al.,[28] where the Noise2Void scheme was applied. Considering the construction of a noisy image x:

$$x = n + s \tag{2}$$

Where s is the real signal, and n is the noise of the image. The goal of any denoising algorithm is to remove the noise (n) from the pixel value of the original image, without affecting the signal part (s). In the context of Noise2Void, as there is no need for clean ground truth data, denoising can be done based on two assumptions: 1. the pixels in a certain area of the image are statistically dependent, and 2. the noise value n_i of pixel value x_i are conditionally independent given the real signal s_i, and the noise has zero mean. These two assumptions can be mathematically formalized as follow:

- Image x is generated from the joint distribution function:

$$p(s, n) = p(s)p(n|s) \tag{3}$$

- Assumption 1 can be expressed as:

$$p(s_i|s_j) \neq p(s_i) \tag{4}$$

For two pixel i, j in a certain radius of each other.

- Assumption 2 can be expressed as:

$$p(n|s) = \prod_i p(n_i|s_i) \tag{5}$$

$$E[x_i] = s_i \rightarrow gt[noiselessimage] iff E[n_i] = 0 \tag{6}$$

Having the above assumptions, to map the image from x to s, we used a convolutional neural network with U-Net architecture. The proposed denoising framework based on Ref. 28 is illustrated in Fig. 2 in comparison with the conventional one. In this framework, the center pixel of the receptive field is replaced by another random neighboring pixel (blind-spot mechanism). The blind-spot mechanism allows the network to predict the value of the center pixel based on the surroundings (the statistical dependence of the signal part) without learning its identity. The predicted result after training is the

clean signal part, as the network cannot predict the noise of the center pixel based on its surrounding due to assumption 2 (the statistical independence of noise).

In the training phase, the goal is to minimize the loss function shown in Equation (7) by fine-tuning the parameter weights w, where $\tilde{x}^j_{RF(i)}$ is the receptive field of the model as depicted in Fig. 3.

$$arg_w \; min \sum_j \sum_i (g(\tilde{x}^j_{RF(i)}; w), x^j_i) \tag{7}$$

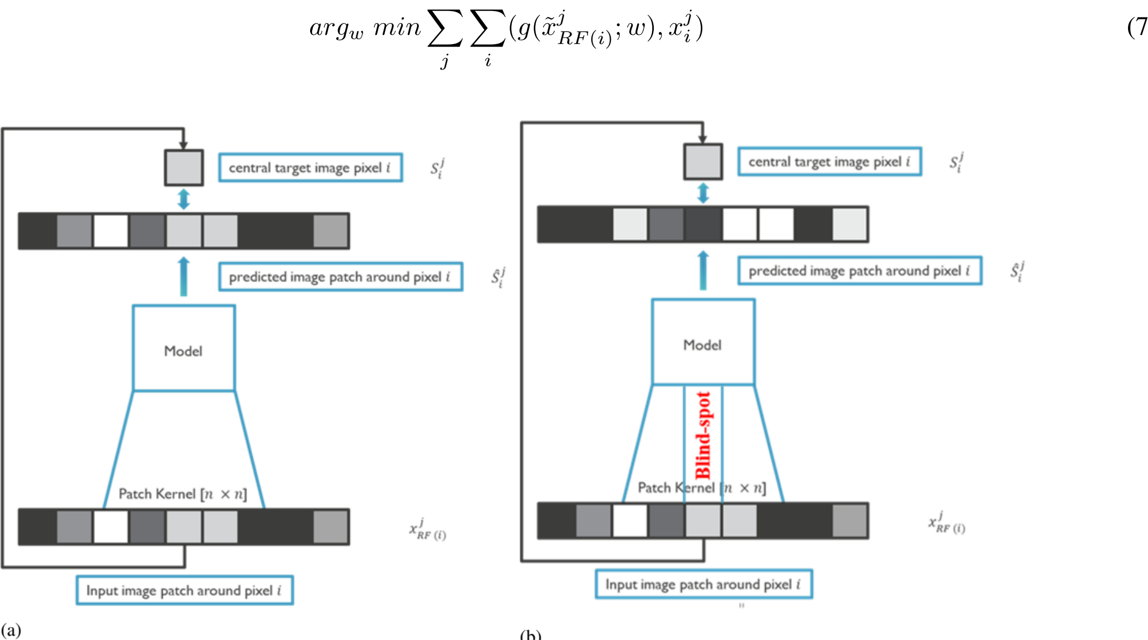


Figure 2: Schematic of conventional network and proposed network. (a) Conventional network. (b) Proposed network.[27]

3.0.2 Data Preparation and Training

For the training step, a number of patches of size [64x64] from a single noisy SEM image (in *TIFF* format) were randomly extracted as the dataset. We have reiterated the training for other configurations of the images as well (patch size of 32x32, 96x96, etc. as well as different image sizes such as 256, 512, 1024, and 2048). For the DRAM dataset, we have used one ADI dataset and 2 AEI (After Etch Inspection) datasets for our experimentation. This dataset is then divided into a training set and validation set, with a ratio of 60:40. Based on the convergence performance of the model, we set the number of training iterations between 100 to 150 with the number of training steps per epoch in the range of 200-400. We set the training batches size as 128, the first layer's convolution kernels size as 3, the number of feature channels in the first layer as 32, the initial learning rate as 0.0001, depth of the network as 2. The denoising model was written in Python using

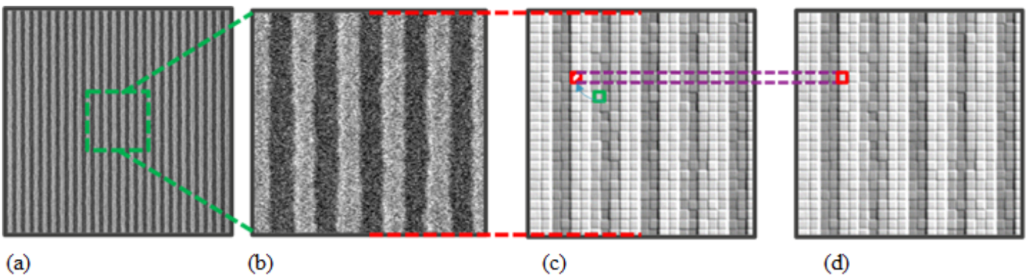
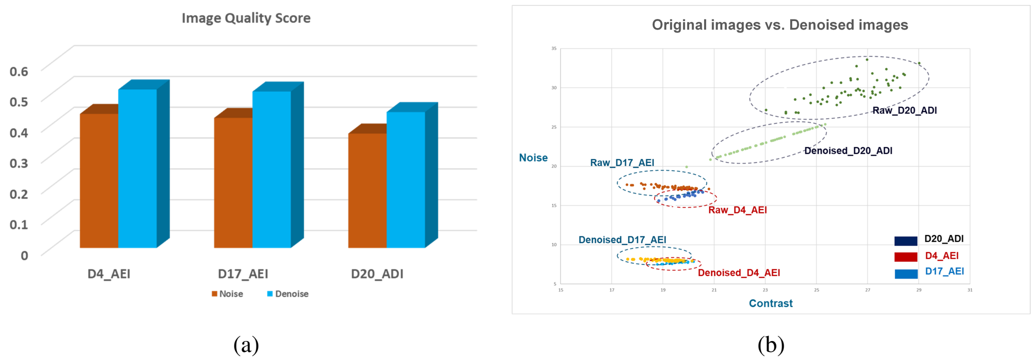


Figure 3: Blind-spot masking framework. (a) Original noisy SEM image. (b) Extracted image patch from (a). (c) Modified image by replacing the target pixel (labeled red) by a random neighboring pixel (labeled green), which created a blind-spot. (d) Original input patch corresponding to (c) without any modification.[27]



(a) (b)

Figure 4: Validation of Noisy/Denoised image pair for DRAM [SNLP+BLP] structure by leveraging 'Image Quality Assessment' module in Calibre®SEMSuite^{TM} across both ADI/AEI group of images.

Table 1: SUMMARY OF ANALYSIS METRIC FROM CALIBRE®SEMSUITE^{TM} (**LINE/SPACE** PATTERNS)[32]

Images (Noisy/Denoised)	Gauge ID	# of Bridges	
		TRUE	FALSE
Noisy Raw Image	1	3	**17**
Denoised_normal_withoutCP	2	3	**0**

Keras framework[29] and Tensorflow platform.[30] The model was packaged in an Anaconda environment (version 4.6.8). To accelerate the training, Batch Normalization[31] was added before each activation function. The training of our model is done by using Lambda TensorBook with NVIDIA RTX 2080 MAX-Q GPU.

3.0.3 Results and Discussion

Fig.4 shows qualitative and quantitative assessment for both ADI and AEI dataset of DRAM [SNLP+BLP] structures, after applying proposed unsupervised denoising algorithm. By leveraging 'Image Quality Assessment' module in Calibre® SEMSuite^{TM}, both raw noisy images and denoised images are validated. The image score (value:0∼1) demonstrates that proposed deep learning approach would gain the overall image quality ∼20% improvement (including brightness, contrast, noise level etc.) as shown in Fig.4 (a). From the scatter plot Fig.4 (b), the 'Noise' index shows the improvement against corresponding raw noisy images across 3 groups (1 ADI and 2 AEI DOEs).

Fig.5 shows Calibre®SEMSuite^{TM} panel with contour extraction result metrics for DRAM structures (an example scenario). With proposed deep learning denoiser assisted framework integrated into this programmable tool, the 'Image Quality Assessment' module helps to automatically detect the quality of a batch of images by leveraging a set of metrics that are indicative of visual perception of the images. Some of possible applications with Calibre®SEMSuite^{TM} are:

- to compare datasets quantitatively to get an upfront idea about potential success of running the dataset through other tools.
- to filter out bad quality images upfront from being considered for model calibration, alignment with layout, defect detection, contour extraction, etc.
- to add some meaning to confidence of alignment and contour extraction results.

From Fig.5, it can be seen that all the polygons within the denoised image can be extracted successfully in these 2 pattern types (SNLP+BLP) by applying same image preprocessing settings (as for noisy) in Calibre®SEMSuite^{TM}. The "high confidence Overall %" is indicating to the extracted contour result (labelled in green) after calculating with the high score within all the edges. Also, within the same image processing step setting, the denoised image could achieve no leakage extraction result in comparison with noisy image.

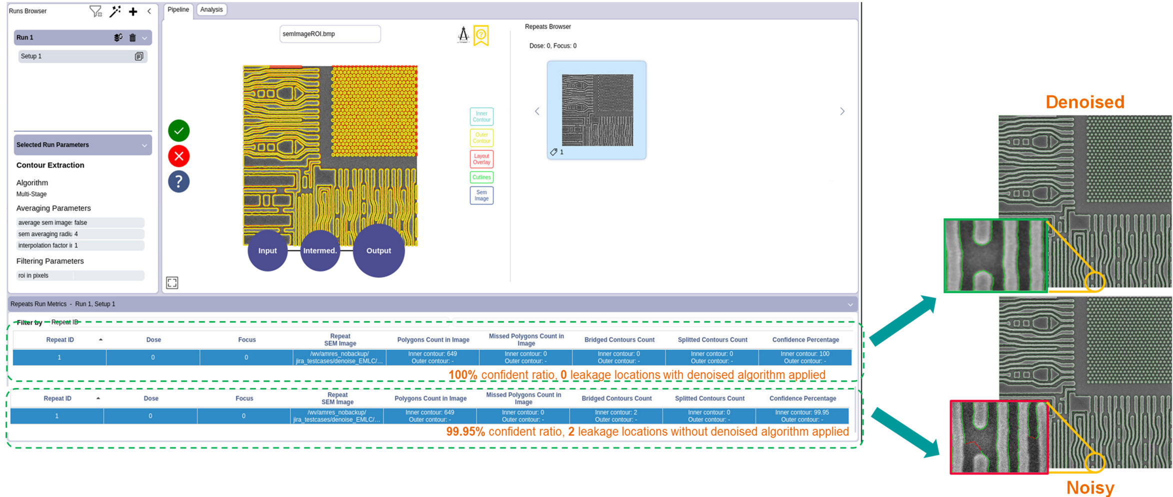


Figure 5: Calibre®SEMSuite^{TM}: Contour extraction result metrics panel.

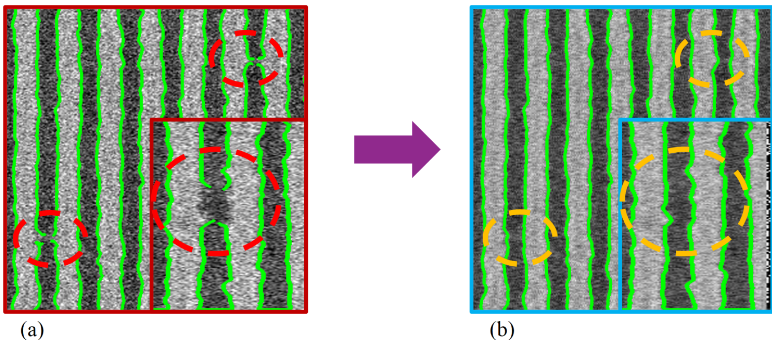


Figure 6: False positive removal of micro bridge defects in a noisy image when applying the denoising model.[32]

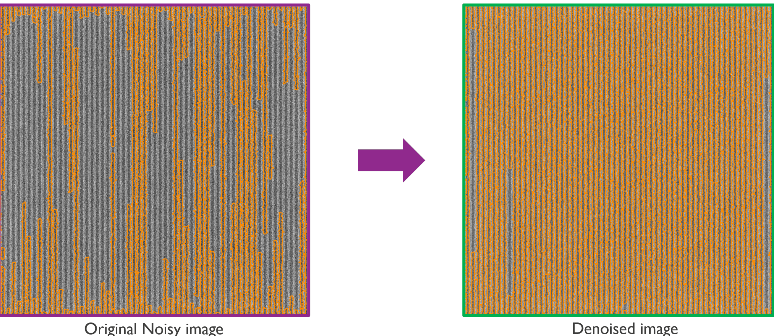


Figure 7: Gaussian Mixture Model-based contour extraction results in noisy (left) vs denoised (right) for Thin resist low-ft Line-Space image.[27]

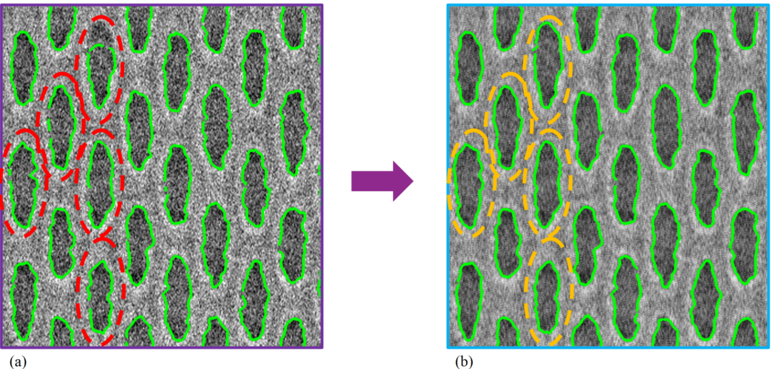


Figure 8: Hole/contact structures' contour extraction. (a) Original noisy image. (b) Denoised image.[32]

Table 2: SUMMARY OF ANALYSIS METRIC FROM CALIBRE®SEMSUITE™ (**CONTACT/HOLE** PATTERNS)[32]

Images (Noisy/Denoised)	Gauge ID	High Confidence Overall %
Noisy Raw Image	1	96.61
Denoised_normal_withoutCP	2	**98.85**

4. APPLICATIONS

4.1 Defects Detection and Classification

As the size of printed features becomes very small, the number of available pixels for defect detection and classification approaches single-digit numbers. Thus, noise presence in SEM images will considerably degrade the accuracy of the metrology analysis. As a result, denoising is critical to ensure robust performance. In this work, we showed several examples of how noise filtering helps with the task of defect detection. Table 1 and Fig. 6 demonstrated a comparison of the performance of the detecting micro-bridge errors task between noisy and denoised SEM images. The analysis was done by using Calibre® SEMSuite™.[32] We can see that by denoising the image before analyzing it, the number of false positive detection is significantly reduced.

4.2 Contour extraction

Noise filtering also plays an important role as an aided tool for extracting the CD or contour in SEM images. To show an example, we tested an unsupervised Gaussian Mixture Modelling (GMM) contour extraction method for both noisy and denoised SEM images (Fig. 7). We can see that the algorithm failed to extract full contour for a lot of parts in the image due to the lack of granularity for key pixels in noisy images. Meanwhile, the same model demonstrates improved contour extraction performance with the denoised image, without the need for fine-tuning. In Fig. 8 and Table 2, we analyzed the contour extraction performance on SEM images of contact/hole patterns by the Calibre® SEMSuite™ programmable tool.

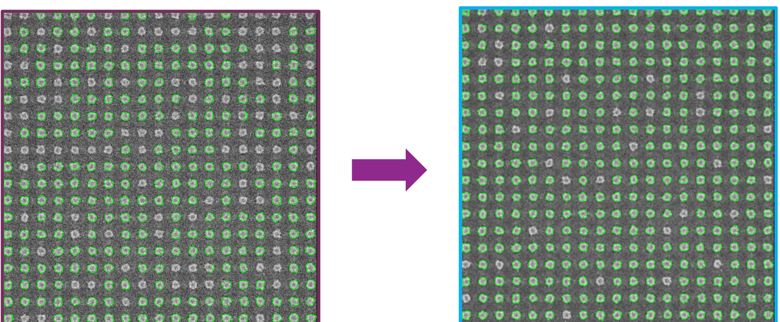


Figure 9: Contour extraction result analysis on Calibre®SEMSuite™ for orthogonal pillar array (design CD=20nm, pitch= 40nm) with a specific scanning direction [Scan_1]. (a) Noisy image (b) Denoised image.[32]

Table 3: Summary of analysis metric from Calibre®SEMSuite™ for **Pillar** patterns with 3 different scanning directions.[32]

Scanning direction	Images (Noisy/Denoised)	Gauge ID	Layout patterns	# patterns w/ properly extracted contours	# patterns w/ partially extracted contours	# missed patterns
Scan_1	Raw	1	361	**282**	**24**	**55**
	Denoised_uniform_withCP	2		319	2	40
	Denoised_normal_withoutCP	3		**327**	2	32
	Denoised_normal_fitted	4		316	2	43
	Denoised_normal_additive	5		326	6	**29**
Scan_2	Raw	1	361	**279**	2	**80**
	Denoised_uniform_withCP	2		348	0	13
	Denoised_normal_withoutCP	3		342	2	17
	Denoised_normal_fitted	4		**353**	1	7
	Denoised_normal_additive	5		352	1	8
Scan_3	Raw	1	361	**274**	0	**87**
	Denoised_uniform_withCP	2		335	1	25
	Denoised_normal_withoutCP	3		**350**	2	9
	Denoised_normal_fitted	4		344	11	**6**
	Denoised_normal_additive	5		332	9	20

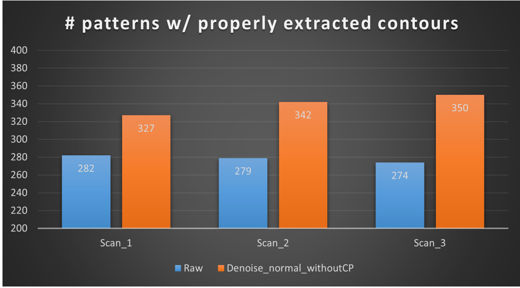


Figure 10: Number of patterns with fully extracted contours in noisy versus denoised Pillar pattern (only shown for gauge ID3) images with 3 different scan types. Contour extraction result analysis on Calibre®SEMSuite™.[32]

In the denoised image, the number of fully extracted contours increases substantially compared to its noisy counterpart, where there are many partially extracted contours. For C/H scenario, "High Confidence Overall %" is indicative of the ratio of the number of green contour points on extracted contour to total number of points on extracted contour, across all patterns in the image. In Table 2, we have shown the impact when the unsupervised algorithm is applied. The "High Confidence Overall %" is relatively higher in the denoised image than the raw noisy one even when all the polygons are extracted for both. Fig. 9 shows the contour extraction results comparison for Scan_1 pillar array pattern, for noisy/denoised image pair. Table 3 demonstrates that our proposed unsupervised denoising method significantly increases the number of patterns

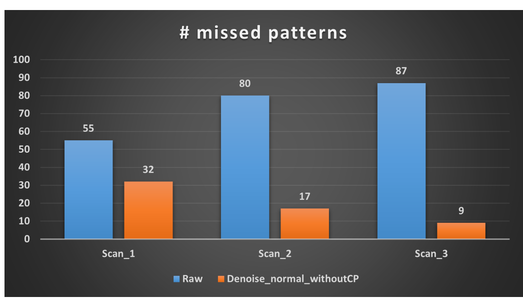


Figure 11: Number of missed patterns (through contour extraction) in noisy versus denoised Pillar pattern (only shown for gauge ID3) images with 3 different scan types. Contour extraction result analysis on Calibre®SEMSuite™.[32]

with fully extracted contours, as well as decreases the number of missed patterns for all categorically different geometrical structures, specifically presented here for pillar patterns with 3 different scanning directions as well as for all four different experimental blind-spot pixel manipulation mechanisms. During training of the model, we created the blind-spot pixels by manipulating neighborhood pixels of an image patch in 4 different experimental ways as (1) uniform_with_CP $\rightarrow$ samples neighborhood pixels according to a continuous uniform distribution with the center pixel of the patch kernel, (2) normal_without_CP $\rightarrow$ samples neighborhood pixels according to a normal gaussian distribution without the center pixel of the patch kernel, (3) normal_additive $\rightarrow$ sampling a random number from a Gaussian distribution with mean $(\mu) = 0$ and standard deviation $(\sigma) = radius$ of patch kernel and adding this random number to the original pixel value and (4) normal_fitted $\rightarrow$ sampling a random number from a gaussian distribution with mean $(\mu) = mean(neighborhood_pixels)$ and standard deviation $(\sigma) = standarddeviation(patch_kernel)$. This is to keep the network away from learning the identity $f(inp_{pix} \rightarrow out_{pix})$ as well as to observe the denoising effect on the contour extraction performance. Fig. 10 and Fig. 11 are the visual representations (only shown for gauge ID3) for the above two significant metrics, against target layout pillar patterns number [#361].

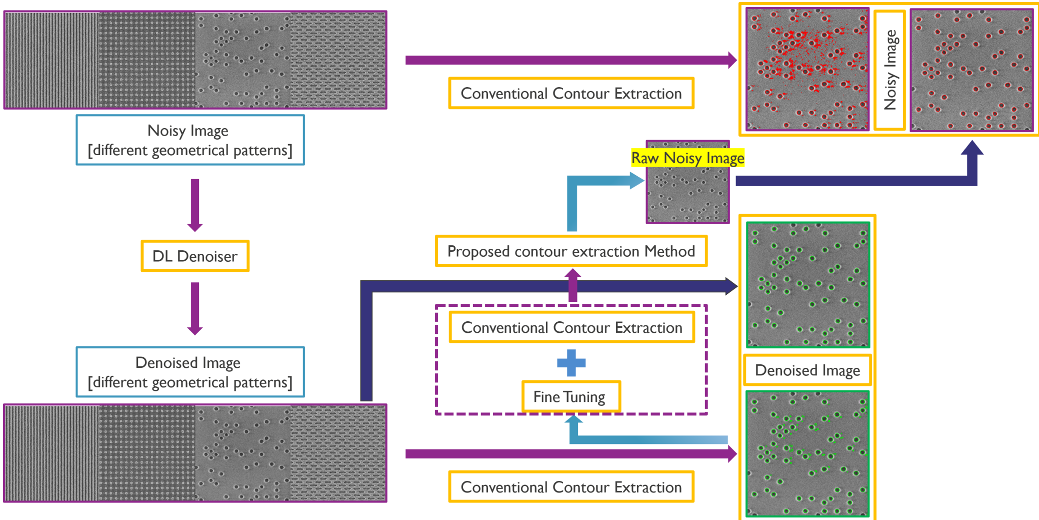


Figure 12: Proposed DL Denoiser assisted Framework[33].[34]

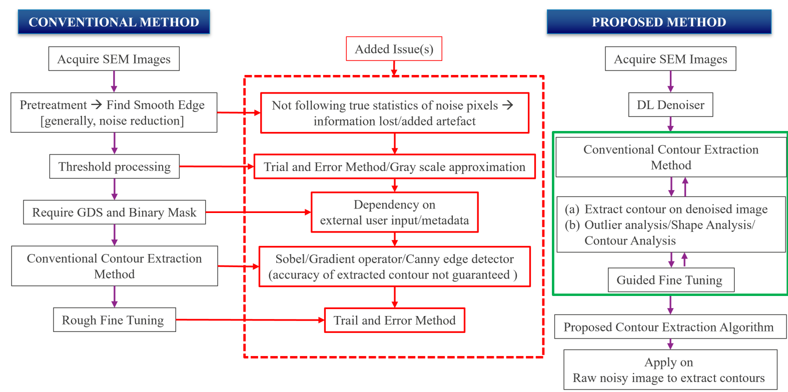


Figure 13: Flowchart of Contour extraction steps: Conventional vs. Proposed contour extraction method[33].[34]

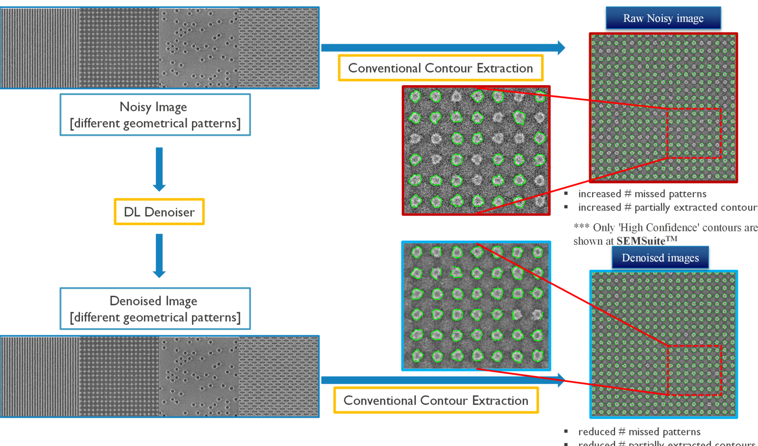


Figure 14: Analysis with Calibre®SEMSuite™ [Pillar pattern]: contour extraction performance on raw noisy image vs denoised image[33],[34]

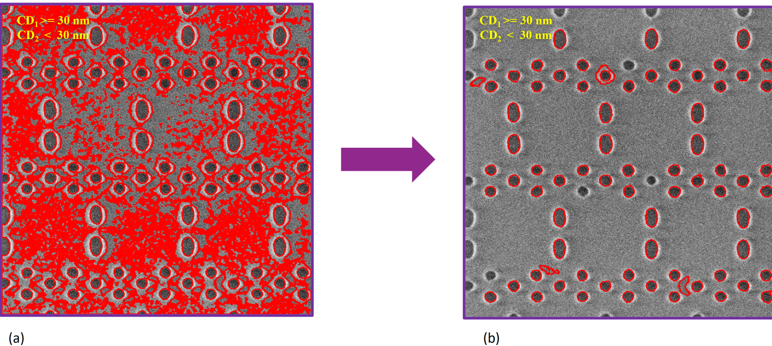


Figure 15: SRAM structure [minimum C2C 56nm] (ADI): (a) Contour extraction performed using conventional image processing technique on noisy raw CD-SEM image, (b) Contour extraction performed using the proposed method on the same noisy raw CD-SEM image[33],[34]

5. PROPOSED DL DENOISER ASSISTED FRAMEWORK FOR ROBUST SEM CONTOUR EXTRACTION

We have extended our work towards enabling SEM contour extraction possible for all the edges in the raw noisy SEM image itself. We have proposed a deep learning denoiser-assisted framework for the extraction and analysis of SEM contours with main contributions as: (a) a novel and effective edge extraction technique, (b) with minimum/no requirement of external user input or metadata (like GDSII / OASIS data, CSV-like meta-data set, etc.) to extract and analyze information from noisy SEM images, (c) an improved contour extraction algorithm capable to extract contours on the body of noisy raw image itself with a posteriori knowledge derived from its denoised twins. The proposed contour extraction algorithm, aided by deep learning denoiser, specifically eliminates the presence of any outlier patterns by analyzing the shape properties such as #sides, perimeter, area, etc., then differentiating from the target patterns. We have analyzed, compared, and validated our contour extraction results for each noisy/denoised image pair for categorically different geometrical patterns such as L/S (line-space), C/H (contact-hole), pillars with different scan types, SRAM structures, Logic structures and DRAM (SNLP+BLP) 2D-structures, respectively. We have demonstrated that our proposed method is capable to extract contours on the body of the noisy SEM images with accuracy in close proximity to design data. The proposed framework is shown in Fig. 12.

Fig. 13 demonstrates a comparative analysis of the step-by-step approach of contour extraction algorithm between the conventional and proposed method. Roughness and low thickness of EUV resists, pose significant challenges, especially After Development Inspection (ADI). SEM images are usually noisy and low contrast as (a) low k1 factor limited lithogra-

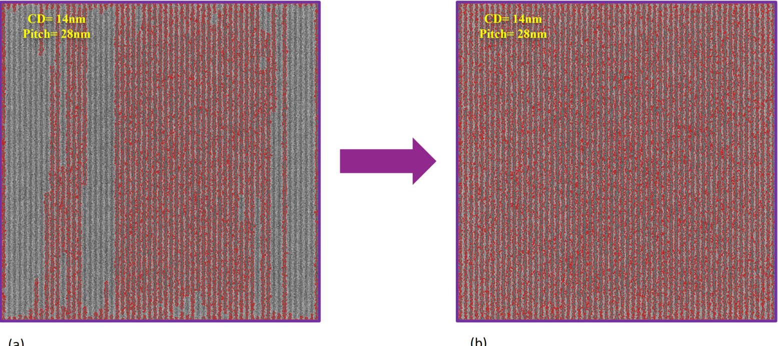


Figure 16: L/S pattern [P28_8nm] (ADI) thin resist: (a) Contour extraction performed using conventional image processing technique on noisy raw CD-SEM image, (b) Contour extraction performed using the proposed method on the same noisy raw CD-SEM image[33].[34]

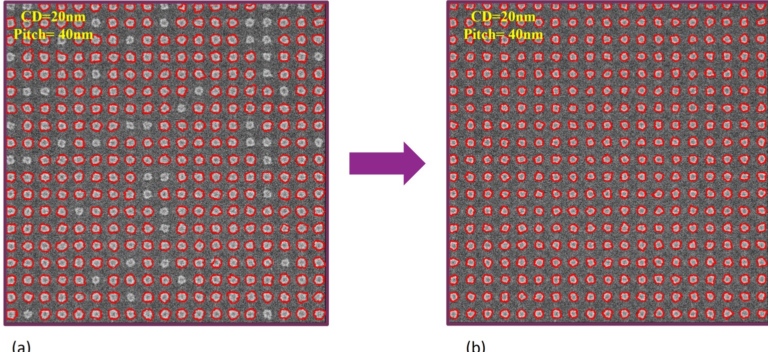


Figure 17: Orthogonal pillar array (design CD=20nm, pitch= 40nm) with Scan_1 type: (a) Contour extraction performed using conventional image processing technique on noisy raw CD-SEM image, (b) Contour extraction performed using the proposed method on the same noisy raw CD-SEM image[33].[34]

phy process, (b) a smaller number of frames are used to avoid resist shrinkage. Conventional contour extraction techniques, such as Sobel operator, gradient operator, Canny edge detector, etc., do not survive the requirements of the semiconductor industry at these advanced nodes. Though applying these traditional methods have limited advantages in terms of simple scripting and speedy computation only, in contrast, the major disadvantages are: (a) the existence of stochastic noise significantly impacts contour extraction itself, (b) the accuracy of an extracted contour cannot be guaranteed in terms of the width of a single pixel and therefore may appear either as (1) edge of an isolated one or, (2) a small continuous section. Fig. 14 demonstrates these consequences associated with raw noisy images as well as with conventional contour extraction algorithms. In this evaluation of the extracted contours, only the "high confidence" contours, labeled in green, are shown without any "low confidence" contour(s). Inside of Calibre®SEMSuite™, the "high/low confidence" contour is categorized simultaneously. In addition, there are multiple indexes in the metric for further analysis when the extracted contour is provided. "High Confidence Overall %" is indicative of the ratio of the number of green contour points on extracted contour to the total number of points on extracted contour, across all patterns in the images. The main drawbacks of conventional contour extraction techniques can be summarized as:

- Image pre-processing to find the smooth edges → noise reduction method in general and not true noise removal. Therefore, this approach does not follow true statistics of noise pixels and leads toward information loss and/or added digital artefacts.
- Threshold processing → to generate binary mask a trial-and-error method towards gray scale approximation. Susceptible towards actual morphology of the feature (geometric precision may be altered).

Table 4: Summary of analysis metric from our proposed framework in comparison with Calibre®SEMSuite™ (Pillar patterns)[33],[34].

Scanning Direction	Images (Noisy/ Denoised)	Gauge ID	Method	Extracted contours/ Layout Patterns	Error Percentage (%)	Method	Extracted contours/ Layout Patterns	Error Percentage (%)
Scan_1	Raw Noisy	1	Conventional	310/361	14.13	Proposed Method	361/361	0
	Denoised_ uniform_ withCP	2		349/361	3.32		361/361	0
	Denoised_ normal_ withoutCP	3		319/361	11.63		361/361	0
	Denoised_ normal_ fitted	4		327/361	9.42		361/361	0
	Denoised_ normal_ additive	5		335/361	7.2		361/361	0
Scan_ 2	Raw Noisy	1		283/361	21.6		361/361	0
	Denoised_ uniform_ withCP	2		348/361	3.6		361/361	0
	Denoised_ normal_ withoutCP	3		344/361	4.7		361/361	0
	Denoised_ normal_ fitted	4		354/361	1.94		361/361	0
	Denoised_ normal_ additive	5		353/361	2.22		361/361	0
Scan_3	Raw Noisy	1		275/361	23.82		361/361	0
	Denoised_ uniform_ withCP	2		335/361	7.2		361/361	0
	Denoised_ normal_ withoutCP	3		349/361	3.32		361/361	0
	Denoised_ normal_ fitted	4		345/361	4.43		361/361	0
	Denoised_ normal_ additive	5		336/361	6.9		361/361	0

- Design GDS + Binary mask requirement → Dependency on external user input (like CSV metadata).
- Conventional contour extraction algorithm → accuracy of extracted contours not guaranteed due to previous steps, mostly. However, rough fine-tuning of parameter settings (with manual intervention and expertise with trial-and-error method) may improve contour extraction results to a certain extent.

The comparative analysis demonstrates that de-noised images have higher confidence contour metric than their noisy twins while keeping the same parameter settings for both data inputs. When the ML algorithm is applied, the contour extraction results would have higher confidence numbers compared with the ones only applied to the conventional Gaussian or Median blur de-noise method. Most importantly, this work demonstrates with applied deep learning-based denoising algorithm, the stochastic noise has been removed which in turn optimizes the varying non-uniform background intensity for better contour detection without altering the PSD.

Therefore, in the first phase, only applying the proposed deep learning-based denoising algorithm with the conventional contour extraction algorithm significantly reduced "# missed patterns/contours" as well as increased "# full/partially extracted contours" in denoised images.

In the second phase, we analyzed the arc-lengths, perimeters, and areas of all the extracted contours/patterns on the denoised images. This analysis (numerical parameters for arc-length, perimeter, and area) will help to explicitly eliminate the presence of any outlier pattern(s) as unintended patterns, edge of an isolated one etc., which do not fit a target threshold. This data helps a guided fine-tuning of the conventional contour extraction algorithm hyperparameters. Now, with a posteriori knowledge derived from these denoised images, the resulting contour extraction algorithm can be applied precisely to the body of the noisy raw image itself and therefore capable to extract critical contours.

Table 5: Summary of analysis metric from our proposed framework in comparison with conventional contour extraction method (L/S pattern [P28 8nm] (ADI) thin resist)[33].[34]

Images (Noisy/Denoised)	Method	Extracted contours/Layout Patterns	Error Percentage (%)
Raw Noisy	Conventional	37318/58	64241.38
Denoised_uniform_withCP	Conventional	6248/58	10672.41
Raw Noisy	Proposed Method	236/58	306.9
Denoised_uniform_withCP	Proposed Method	113/58	94.83

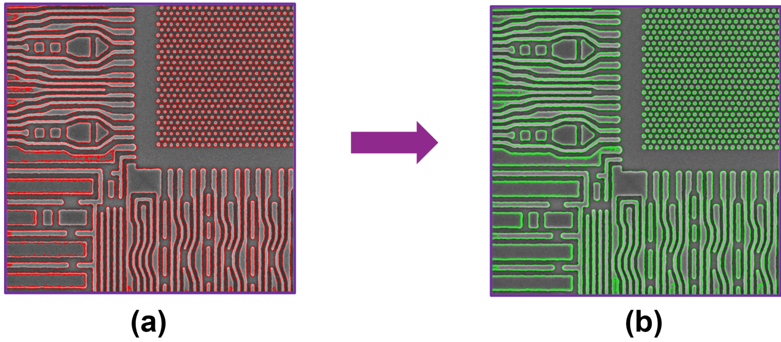


Figure 18: DRAM [SNLP+BLP](design Pillar CD=20nm, pitch= 36nm, AEI): (a) Contour extraction performed using conventional image processing technique on noisy raw CD-SEM image, (b) Contour extraction performed using the proposed method on the same noisy raw CD-SEM image.

5.1 Experimental Results

Fig. 15-19, (b) demonstrates a comparison of extraction of contours on the body of the noisy images itself using our proposed framework, against using conventional image processing technique as shown in Fig. 15-19 (a). With the conventional method, for noisy images we are missing the granularity for major pixels and thus CD extraction is failing. For the same, proposed method makes better CD extraction possible. Our proposed approach is capable to extract contours on the body of the raw noisy image itself, where the contour extraction algorithm is only fine-tuned by the contour extraction parameters [based on shape analysis/contour analysis] of the denoised images. Varying parameters, for the conventional method, has a significant impact on the number of properly/partially extracted contours as well as missed patterns. However, our proposed contour extraction method shows robustness. It has also been observed for certain geometrical patterns, if contour extraction can be performed on noisy images without any image preprocessing feature, the contour extraction performance gain can be achieved using only its denoised twins approximately more than 100%. This is a significant metric to evaluate the performance of the proposed framework as to how accurately we can extract contours without using any additional image preprocessing steps. As shown in Table. 4, for pillar patterns we were able to extract all contours without any "# missed patterns" on both raw noisy and denoised images. As shown in Table. 5, for L/S pattern [P28, ADI SEM with 8nm thin resist], our deep learning-based contour detection framework performs better on both noisy/denoised image pair without requiring the Mask-GDS against classical contour extraction method. However, the evaluation has been carried out by considering the fact that leakage is predominant in the case of 8 nm thin resist.

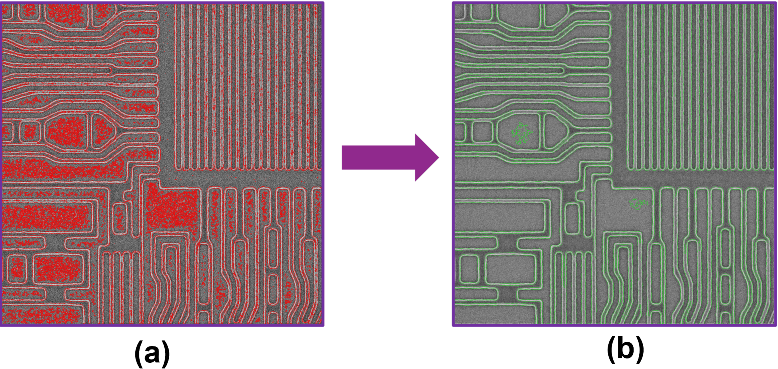


Figure 19: DRAM [BLP] (ADI): (a) Contour extraction performed using conventional image processing technique on noisy raw CD-SEM image, (b) Contour extraction performed using the proposed method on the same noisy raw CD-SEM image.

6. CONCLUSION

In this work, we have shown how applied deep learning denoiser demonstrates better image quality score against raw noisy images (for both ADI and AEI dataset across different geometrical patterns). Proposed deep learning denoiser assisted contour extraction framework, being integrated into Calibre®SEMSuite™,can significantly improve the performance of the post-processing software after obtaining the images by SEM tools in different tasks, including defects detection and contour extraction. Moreover, the programmable tool assists to filter bad quality images upfront from being considered for model calibration and alignment with layout. Contrast can be further enhanced by image processing module in Calibre®SEMSuite™ in order to reach the high precision extraction results. By applying our proposed framework in the production line, the yield and quality of IC devices will increase greatly, since the errors are more well-controlled. For future work, our goal is to improve the robustness and accuracy of the denoising framework by integrating techniques that can learn better the statistical distribution of noise such as noise histogram analysis or GMM.

7. ACKNOWLEDGMENTS

The authors used the imec Leuven research computing facility to conduct part of the research. We thank Danlio De Simone, Mihir Gupta, Victor Blanco, and Sayantan Das for their assistance in data preparation. We thank Micea Dusa and other colleagues who provided insight and expertise that helped us to improve the manuscript.

REFERENCES

[1] Timischl, F., Date, M., and Nemoto, S., "A statistical model of signal–noise in scanning electron microscopy," *Scanning* **34**(3), 137–144 (2012).

[2] Yu, L., Zhou, W., Pu, L., and Fang, W., "Sem image quality enhancement: an unsupervised deep learning approach," in [*Metrology, Inspection, and Process Control for Microlithography XXXIV*], **11325**, 1132527, International Society for Optics and Photonics (2020).

[3] Lee, H.-G., Schmitt-Weaver, E., Kim, M.-S., Han, S.-J., Kim, M.-S., Kwon, W.-T., Park, S.-K., Ryan, K., Theeuwes, T., Sun, K.-T., et al., "Virtual overlay metrology for fault detection supported with integrated metrology and machine learning," in [*Metrology, Inspection, and Process Control for Microlithography XXIX*], **9424**, 94241T, International Society for Optics and Photonics (2015).

[4] Schmitt-Weaver, E., Kubis, M., Henke, W., Slotboom, D., Hoogenboom, T., Mulkens, J., Coogans, M., ten Berge, P., Verkleij, D., and van de Mast, F., "Overlay improvements using a real time machine learning algorithm," in [*Metrology, Inspection, and Process Control for Microlithography XXVIII*], **9050**, 496–502, SPIE (2014).

[5] Kuo, H.-F. and Faricha, A., "Artificial neural network for diffraction based overlay measurement," *IEEE Access* **4**, 7479–7486 (2016).

[6] Schmitt-Weaver, E. and Bhattacharyya, K., “Pairing wafer leveling metrology from a lithographic apparatus with deep learning to enable cost effective dense wafer alignment metrology,” in [*Optical Microlithography XXXII*], **10961**, 35–40, SPIE (2019).

[7] Nakazawa, T. and Kulkarni, D. V., “Anomaly detection and segmentation for wafer defect patterns using deep convolutional encoder–decoder neural network architectures in semiconductor manufacturing,” *IEEE Transactions on Semiconductor Manufacturing* **32**(2), 250–256 (2019).

[8] Cheon, S., Lee, H., Kim, C. O., and Lee, S. H., “Convolutional neural network for wafer surface defect classification and the detection of unknown defect class,” *IEEE Transactions on Semiconductor Manufacturing* **32**(2), 163–170 (2019).

[9] Lei, H., Teh, C., Yu, L., Fu, G., Pu, L., and Fang, W., “Denoising sample-limited sem images without clean data,” in [*Metrology, Inspection, and Process Control for Semiconductor Manufacturing XXXV*], **11611**, 116111A, International Society for Optics and Photonics (2021).

[10] Yu, D., Liu, Y., and Hawkinson, C., “The application of a new stochastic search algorithm “adam” in inverse lithography technology (ilt) in critical recording head fabrication process,” in [*Optical Microlithography XXXIV*], **11613**, 116130N, International Society for Optics and Photonics (2021).

[11] Kwon, Y. and Shin, Y., “Optical proximity correction using bidirectional recurrent neural network with attention mechanism,” *IEEE Transactions on Semiconductor Manufacturing* **34**(2), 168–176 (2021).

[12] Geng, H., Zhong, W., Yang, H., Ma, Y., Mitra, J., and Yu, B., “Sraf insertion via supervised dictionary learning,” *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems* **39**(10), 2849–2859 (2019).

[13] Alawieh, M. B., Lin, Y., Zhang, Z., Li, M., Huang, Q., and Pan, D. Z., “Gan-sraf: Subresolution assist feature generation using generative adversarial networks,” *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems* **40**(2), 373–385 (2020).

[14] Dabov, K., Foi, A., Katkovnik, V., and Egiazarian, K., “Image denoising by sparse 3-d transform-domain collaborative filtering,” *IEEE Transactions on image processing* **16**(8), 2080–2095 (2007).

[15] Jain, V. and Seung, S., “Natural image denoising with convolutional networks,” *Advances in neural information processing systems* **21** (2008).

[16] Lehtinen, J., Munkberg, J., Hasselgren, J., Laine, S., Karras, T., Aittala, M., and Aila, T., “Noise2noise: Learning image restoration without clean data,” *arXiv preprint arXiv:1803.04189* (2018).

[17] Midoh, Y. and Nakamae, K., “Image quality enhancement of a CD-SEM image using conditional generative adversarial networks,” in [*Metrology, Inspection, and Process Control for Microlithography XXXIII*], *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series* **10959**, 109590B (Mar. 2019).

[18] Chaudhary, N., Savari, S. A., and Yeddulapalli, S. S., “Line roughness estimation and Poisson denoising in scanning electron microscope images using deep learning,” *Journal of Micro/Nanolithography, MEMS, and MOEMS* **18**(2), 1 – 16 (2019).

[19] Chaudhary, N. and Savari, S. A., “Simultaneous denoising and edge estimation from sem images using deep convolutional neural networks,” in [*2019 30th Annual SEMI Advanced Semiconductor Manufacturing Conference (ASMC)*], 1–6 (2019).

[20] Giannatou, E., Papavieros, G., Constantoudis, V., Papageorgiou, H., and Gogolides, E., “Deep learning denoising of sem images towards noise-reduced ler measurements,” *Microelectronic Engineering* **216**, 111051 (2019).

[21] Chen, J., Chen, J., Chao, H., and Yang, M., “Image blind denoising with generative adversarial network based noise modeling,” *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition* , 3155–3164 (2018).

[22] Zhang, K., Zuo, W., Chen, Y., Meng, D., and Zhang, L., “Beyond a gaussian denoiser: Residual learning of deep CNN for image denoising,” *CoRR* **abs/1608.03981** (2016).

[23] Chaudhary, N., Savari, S. A., Brackmann, V., and Friedrich, M., “Sem image denoising and contour image estimation using deep learning,” in [*2020 31st Annual SEMI Advanced Semiconductor Manufacturing Conference (ASMC)*], 1–6, IEEE (2020).

[24] Mack, C. A. and Bunday, B. D., “Analytical linescan model for sem metrology,” in [*Metrology, Inspection, and Process Control for Microlithography XXIX*], **9424**, 117–139, SPIE (2015).

[25] Verduin, T., Kruit, P., and Hagen, C. W., “Determination of line edge roughness in low dose top-down scanning electron microscopy images,” in [*Metrology, Inspection, and Process Control for Microlithography XXVIII*], Cain, J. P. and Sanchez, M. I., eds., **9050**, 148 – 164, International Society for Optics and Photonics, SPIE (2014).

[26] Marturi, N., Dembélé, S., and Piat, N., “Scanning electron microscope image signal-to-noise ratio monitoring for micro-nanomanipulation,” *Scanning* **36**(4), 419–429 (2014).

[27] Dey, B., Halder, S., Khalil, K., Lorusso, G., Severi, J., Leray, P., and Bayoumi, M. A., “Sem image denoising with unsupervised machine learning for better defect inspection and metrology,” in [*Metrology, Inspection, and Process Control for Semiconductor Manufacturing XXXV*], **11611**, 1161115, International Society for Optics and Photonics (2021).

[28] Krull, A., Buchholz, T.-O., and Jug, F., “Noise2void-learning denoising from single noisy images,” in [*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*], 2129–2137 (2019).

[29] Chollet, F. et al., “Keras: Deep learning library for theano and tensorflow,” *URL: https://keras. io/k* **7**(8), T1 (2015).

[30] Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., Isard, M., et al., “{TensorFlow}: A system for {Large-Scale} machine learning,” in [*12th USENIX symposium on operating systems design and implementation (OSDI 16)*], 265–283 (2016).

[31] Ioffe, S. and Szegedy, C., “Batch normalization: Accelerating deep network training by reducing internal covariate shift,” in [*International conference on machine learning*], 448–456, PMLR (2015).

[32] Dey, B., Wu, S., Das, S., Khalil, K., Halder, S., Leray, P., Bhamidipati, S., Ahi, K., Pereira, M., Fenger, G., et al., “Unsupervised machine learning based sem image denoising for robust contour detection,” in [*International Conference on Extreme Ultraviolet Lithography 2021*], **11854**, 88–102, SPIE (2021).

[33] Dey, B., Halder, S., Das, A., DAS, S., WU, S., and Fenger, G., “Deep-learning denoiser-assisted framework for robust SEM contour extraction and analysis for advanced semiconductor node,” in [*International Conference on Extreme Ultraviolet Lithography 2022*], Itani, T., Naulleau, P. P., Gargini, P. A., and Ronse, K. G., eds., **PC12292**, PC122920Z, International Society for Optics and Photonics, SPIE (2022).

[34] Dey, B., “Deep Learning based SEM image Denoising Approaches for Improving Metrology and Inspection of Thin Resists,” (3 2023).

Understanding the impact of rinse on SEM image distortion and hole patterning variability

Elvire Soltani[a,b], Bertrand Le-Gratiet[a], Sébastien Bérard-Bergery[a], Jonathan Pradelles[b], Thibaut Bourguignon[a,b], Aurélie Le Pennec[a], Nathalie Charras[a], Raluca Tiron[b]

[a] STMicroelectronics, 850 rue Jean Monnet, 38926 Crolles Cedex, France
[b]Univ. Grenoble Alpes, CEA, Leti, F-38000 Grenoble, France

ABSTRACT

Background: Stochastic effects are manifested by local variabilities in critical dimension (CD), shape, size and placement of patterns. On previous works, a commercial software has been used to extract contours from CD-SEM images, from which various metrics are computed to quantify these variabilities. Last year, an image processing system has been implemented to monitor these indicators in a high-volume manufacturing (HVM) environment. This system automatically extracts these contour-based metrics from the in-line images.
Aim: The study has been divided into two parts. The first objective was to show the ability to detect a mismatch between two lithography clusters, not revealed by traditional CD in-line trends. Since investigations identified that this mismatch was caused by differences in the rinse recipes, the second objective was to gain a deeper understanding of the impact of rinse on both metrology and patterning.
Approach: In addition to the traditional CD measurements, the following metrics are also monitored: Local CD Uniformity (LCDU), Pattern CD Uniformity (PCDU) and Centroid Shift Uniformity (CSU).
Results: Baselines for the 28 nm contact process have been established, providing new indicators to compare and monitor processes and tools in the fab. PCDU computed on post-lithography images proved to be an effective detector of clusters mismatching, not identified by the CD criterion. Upon investigation, engineers discovered differences in the rinse recipes between the two lithography tools. To analyze the influence of rinse, an R&D experiment was conducted, repeating this very same step 2, 4 and 6 times in a row. As rinse steps number increased, two observations have been made: SEM images are more distorted, and the shape variability indicator (PCDU) increases. The most likely explanation is that increased friction with the deionized water during the rinse results in more stored electrical charges at the wafer surface, which affects the resist (increased "roughness") and disturbs the SEM image acquisition.
Conclusions: By implementing a remote metrology system to perform extensive analysis of in-line CD-SEM images with a contour extraction software, a cluster mismatching (not from the CD point of view) was identified in a HVM environment. This enhanced in-line monitoring enabled us to find the parameter to be adjusted to match them, thereby avoiding any adverse effects. Experiments confirmed that rinse negatively affects both SEM image fidelity (increased distortion) and process uniformity (decreased shape uniformity).
Keywords: stochastics, contour extraction, in-line metrology, resist development process, local variability, image distortion, immersion lithography, computational metrology

1. INTRODUCTION

In EUV lithography, stochastic effects are leading to catastrophic failures, while in DUV (our topic of interest) they manifest themselves as local variability of CDs, shapes and placement [1, 2]. To ensure failure rates below 1 ppm for markets like automotive, these stochastic events must be characterized and predicted. In particular, CD local variability visible within a single CD-SEM image accounts for 90% of total CD variability [3]. An averaging of the data per image as typically done for CD measurement would necessarily hide this local

38th European Mask and Lithography Conference (EMLC 2023), edited by
Uwe F.W. Behringer, Jo Finders, Proc. of SPIE Vol. 12802, 1280208
© 2023 SPIE · 0277-786X · doi: 10.1117/12.2675586

component, and the monitored variability would be under-estimated. This observation motivated the use of a contour extraction software, to maximize the amount of information that can be extracted from an image [4]. Particularly, stochastic effects can be quantified inside SEM images using specific contour-based metrics. Conserving all the intra-pattern CD values and making statistical calculations between various Regions Of Interest (ROIs) help us to characterize local variability of a given process [5]. The benefit of using these metrics for process monitoring is to identify at early stage (i.e. during lithography) some potential issues, like open circuits (open contacts), that may not become apparent until later in the process [6]. These metrics have been monitored in a HVM environment thanks to a newly implemented image processing system [7].

The first part of this paper will present the in-line monitoring of contacts of a 28nm process at STMicroelectronics, which revealed the negative impact that rinse may have on shape's uniformity. Due to the friction with the deionized water, rinse is known to create a storage of negative electrical charges on the surface of the wafer [8]. Does rinse really deteriorates the patterning process itself, or only the SEM image fidelity? This question will be investigated in the second part.

2. DETECTION OF CLUSTERS MISMATCHING

2.1 Contour-based metrics definitions and remote metrology system

Contours are extracted from CD-SEM images with the software SIMPL, from the french company Aselta Nanographics. The SEM contour metrology flow is presented on figure 1. The image is aligned on a layout (in blue) and mathematical model is calibrated to extract the contours (in green). Once the extraction is done, gauges are placed within the hole: it allows to analyze the CD in all directions. The Local CD Uniformity (LCDU) is calculated as the variance of the mean CD values of the holes present in one image. A low LCDU value indicates that all the holes inside the image are uniform in size. Similarly, the Pattern CD Uniformity (PCDU) is calculated as the variance of the CD within each hole, quantifying the shape's uniformity. The lower the PCDU value, the rounder is the hole. Lastly, horizontal and vertical centroid displacements ΔXc and ΔYc are obtained by subtracting the X and Y centroid's coordinates of the holes in the layout, to the ones of the aligned extracted contours. By computing the norm of the vector constructed with these two displacements, the centroid shift metric is obtained. To compensate for the impact of SEM image distortion, centroid shift values are corrected using a six parameters model [9], which is detailed in section 3.5. Then, Centroid Shift Uniformity (CSU) is calculated as the variance of the centroid shift values inside one image, enabling the placement's variability quantification. The three metrics are schematically represented figure 1.

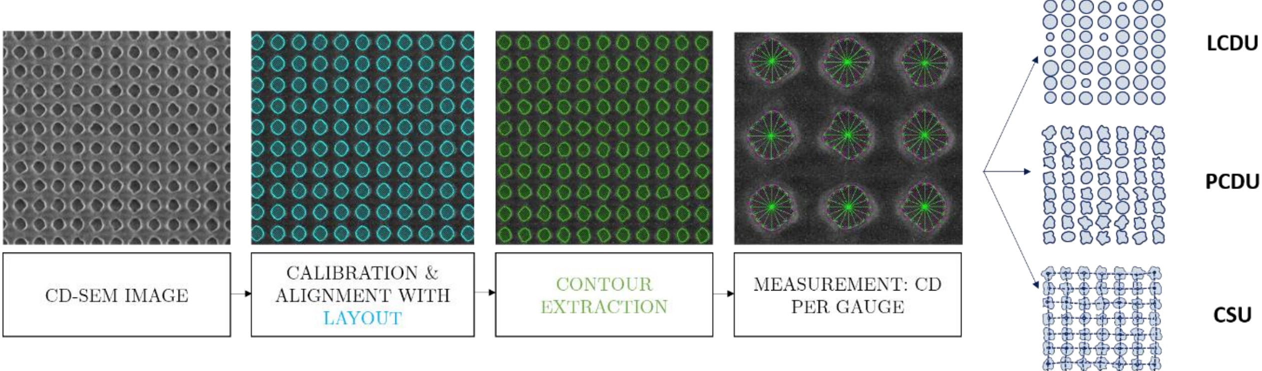


Figure 1. SEM contour metrology flow, with CD-SEM image as input. Measurements can be (among others) CD by gauges placed within the extracted contour. From that, LCDU PCDU and CSU are computed to quantify local variabilities.

Le Gratiet *et al.* presented in a previous publication the image processing system implemented in Crolles fab [7]. This system automatically transfers the in-line CD-SEM images to a Virtual Machine where the previously

mentioned contour extraction software is installed. It extracts contours and calculates the desired measurements on production wafers, to compare processes, tools or recipes. In this work, PCDU, LCDU and CSU are automatically calculated on each CD-SEM image taken during layer process control.

2.2 In-line trends

This work is focused on the contact layer of one 28nm FD-SOI product. The figure 2 (top) shows the CD in-line trend (in percentage, normalized to the mean value) of 600 lots measured in production. The CDs are extracted from post-lithography images (After Develop Inspection, ADI) inside the metrology targets, acquired with Hitachi CG5000 SEMs. The colors indicate from which lithography cluster the data come from, referenced as cluster 1 and cluster 2. The lines schematically represent the average trend for each cluster. One point refers to one CD value of a single lot, averaging several wafers on which multiple images were acquired. The exact in-line metrology sampling cannot be detailed here. The first conclusion from this figure is that, based on the CD criteria, cluster 1 and 2 are matched below 0.1%.

However, if we look at the PCDU in-line trend (bottom), the conclusion is quite different. The data come from the very same ADI images. With this indicator, a clear mismatch can be seen between the clusters, with a difference of 6% between the PCDU values. With a higher PCDU value, it means that the cluster 2 tends to print more distorted contacts compared to cluster 1. Upon investigation, engineers discovered differences in the rinse recipes between both tools. In contrast to traditional CD measurement, PCDU is a very sensitive detector, which demonstrates the benefit of this contour-based metrology remote system. This mismatch had no impact on yield and engineers adapted the rinse recipes, to match the two clusters together (indicated by the star). LCDU baselines (not shown here) did not reveal this rinse based mismatch.

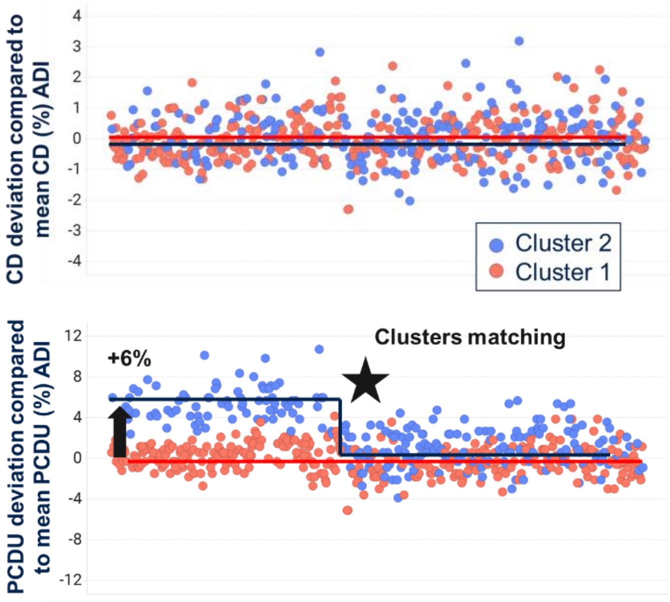


Figure 2. CD and PCDU (shape variability) trends on ADI images. One point represents one CD (top) or one PCDU (bottom) lot, while the color of the point indicates the cluster used. A difference of shape variability on contacts coming from the cluster 2 has been noticed (+ 6% compared to the mean PCDU of cluster 1), while CD remained constant.

To determine if this shape's degradation is only occurring in lithography, the post-etch in-line trend has also been studied (After Etch Inspection, AEI). The figure 3 represents the CD and PCDU trends of the same lots after etch. We notice that contacts etched after a lithography step done with the cluster 2 are also less circular.

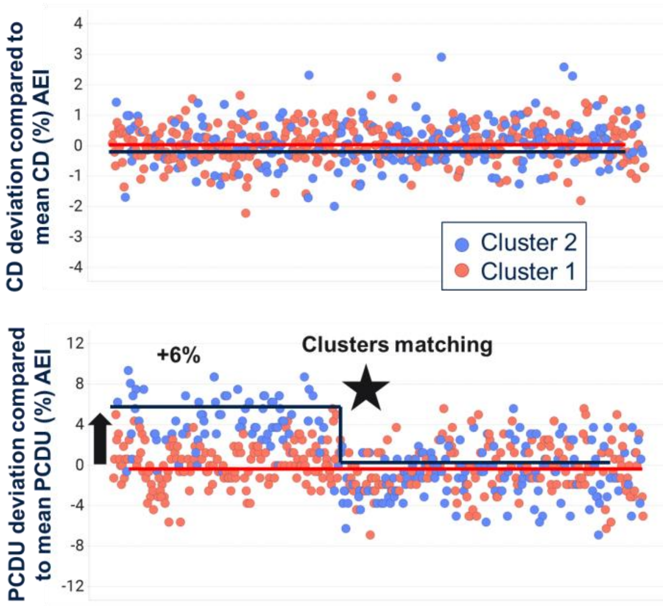


Figure 3. CD and PCDU (shape variability) baselines on AEI images. A difference of shape variability on contacts etched, previously coming from the cluster 2 has been noticed, although it is less straightforward.

This leads us to the second objective: the understanding of the rinse impact on shape variability.

3. IS RINSE AFFECTING PATTERNING?

3.1 Surface charging caused by rinse spinning process

Rinse processing tools operate on a simple principle: the water is dispensed by a nozzle, positioned directly above the center of the spinning wafer. Due to the nature of the spin process, interaction between the tool and the wafer may result in various surface-charging issues. Other wet chemical processes used in integrated circuit manufacturing, like post-etch residue removal or pre-deposition surface treatment, can bring these different types of wafer surface charging issues. The importance of surface charging cannot be understated since excessive current flowing through the surface leads to shifts in device performance and can also lower device yield [10].

The radial pattern of charge caused by rinse - greater negative charge in the center of the wafers and lesser at the edge - were originally shown by Murali *et al*. in 1990 [11]. In their work, Mui *et al*. proposed a physical model to fundamentally explain surface charging induced by a single-wafer wet spin tool [10]. Their model predicted well the charging potential trends with respect to radial position, wet process time and wafer spin speed. Firstly, they displayed through a surface potential wafer map the center-negative and edge-positive potential profile expected (with good repeatability). Secondly, they showed that the surface potential increases in magnitude with time, without a significant change in the overall center-to-edge potential profile. Although, they concluded that the charging process is not linear over time but instead is following an exponential time dependence. Finally, they observed that multiplying the number of dry-to-wet and wet-to-dry transitions did not result in a significant change in the charging potential profile. In another study, Halladay *et al*. showed that the faster the spin on a single wafer tool the more charge is accumulated [12].

There is no consensus in the literature on the mechanism leading to this type of charging. The mechanism can be either chemical, or tribo-electrical. A detailed explanation of the chemical reaction between the water and a SiO_2 surface is detailed in the handbook of cleaning for semiconductor manufacturing [13]. In this case, a deprotonation of the silanol group (found in silicon dioxide) in the water leaves behind a negatively charged

siloxane group. This deprotonation of the terminal silanol groups is caused by the mere action of immersing a wafer in water: silica surfaces acquire a negative charge upon immersion in H_2O without any spinning process. Since insulators do not have a conductive path, charge builds up on the surface and cannot be dissipated easily. In parallel, it has long been known that friction between two materials creates static charge (tribo-electricity).

Very recently, Chen *et al*. reported the influence of the wafer surface charging effect on the precision of measuring ADI features during photo development process using CD-SEM [14]. They made the same observation than the previous authors: that deionized water used in via photo development process would induce negative charge accumulation on the wafer surface. These charges are distributed between the resist and the hard mask. In between, an asymmetric electric field is created on the wafer surface. Because of this field, when the e-beam of the SEM arrives at the wafer surface, the incidence direction will be changed: it will affect the auto-focus. It results in astigmatism, which is the root cause of SEM image blur. In their work, a kind of functional rinse containing surfactant was employed to eliminate the surface charge.

Finally, these charges can be dissipated brutally in contact with a conductor material, it is the electrostatic discharge (ESD). Yang *et al*. listed the defects caused by the ESD, such as inter metal dielectric crack, via hole corrosion and active area damage, all of them induced by the lithography process related charging effects [8].

3.2 Experimental set-up

In this study, we analyze the influence of rinse on CD-SEM image distortion and on patterning of vias of a 28nm process. Four wafers have been developed, repeating consecutively the very same rinse process 1 (process of reference, POR), 2, 4 and 6 times. The rinse recipe has not been modified; it has simply been repeated several times in its entirety. According to the literature, we expect an increase in wafer surface charging. For each wafer, 65 CD-SEM images have been acquired. To avoid astigmatism previously mentioned by Chen *et al.*, correction functions are proposed on the Hitachi QCG5000 SEMs that we are using. It also helps to reduce the other effects of surface potential on wafers, for example the deflection of beams which may causes fluctuations of magnifications values, and hence errors in the CD measurements. Precisely, an option called the Surface Potential Measurement (SPM) is available. The SPM function measures the surface potential with potentiometers, to correct the decelerating voltage during the SEM image acquisition. Although the image blur could be avoided, the wafer surface charge still exists.

Our objective is to determine if the shape's degradation of holes induced by rinse that have been noticed during the in-line monitoring is only a metrology artefact. To do so, the correlation between PCDU and rinse number need to be first determined. The second part aims to confirm that rinse induces charge storage on the wafers we have developed. Our hypothesis is that by creating an accumulation of charges, rinse produces SEM image distortion. To determine if the degradation in shape uniformity is due only to this, we correct it and reanalyze the PCDU based on rinse. The objectives are summarized in the figure 4.

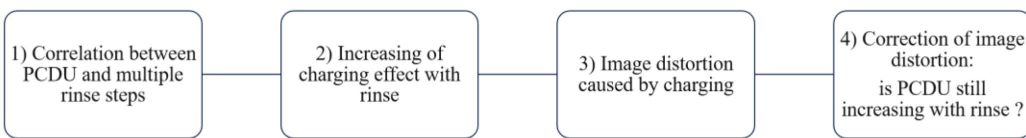
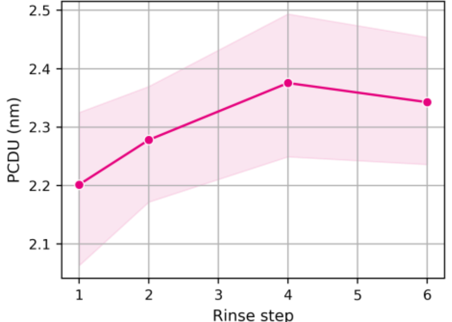


Figure 4. Steps to determine the impact of rinse on shape variability on 4 wafers with 1, 2, 4 and 6 rinse steps.

3.3 Correlations between the amount of rinse steps and shape variability indicator

The figure 5 shows the correlation between rinse steps number (on the x-axis) and the averaged PCDU value on each wafer (on the y-axis). The PCDU is calculated for each hole using the formula on the right of the figure. The PCDU values are then averaged per wafer (population: about 6 500 holes per wafer). The band around the mean value corresponds to the percentiles 10 and 90, showing the dispersion of the metric. This high dispersion

is the natural variability of this indicator. This figure shows that by increasing rinse steps number, holes "roughness" increased, likely reaching a saturation after 4 rinse steps.



$$\text{PCDU (nm)} = \sqrt{\frac{\sum_{i=1}^{N}(CD_{gauge\ n°i} - \mu_{contact})^2}{N-1}}$$

for one contour measured by N gauges, with $\mu_{contact}$ the average value of the N CDs intra-pattern.

Figure 5. PCDU (shape variability) with rinse. One point = one value averaged per wafer.

3.4 Increased charge accumulation with rinse

As mentioned previously, the potential on wafer is measured by the SEM with the SPM function. The SEM returns the parameters a, b and c, allowing a modeling of the potential as a function of the radius with a polynomial function. The figure 6 (left) represents the surface potential (in Volt) at the center (radius=0) for each wafer as a function of rinse step. We see that negative charges accumulation on the wafers increased from -15 to -25 V with rinse and saturated after 4 steps. Figure 6 (right) confirms a negative linear relation between the potential at the center of the wafer and the averaged PCDU.

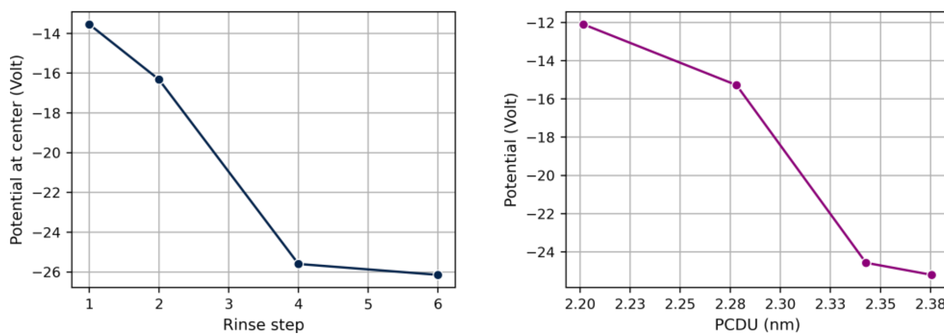


Figure 6. Potential with rinse (left) and SPM as a function of PCDU (right): correlation between both.

3.4.1 Aside

A few years ago, the lithography engineers at Crolles carried out an experiment to study the evolution of the charging of a gate layer as a function of rinse. The surface charges have been measured with a Plasma Damage Monitoring (PDM) tool. The principle of this measurement is to place near the wafer surface a pick-up electrode to form a capacitor with the wafer. The voltage across the capacitor is then equal to the contact potential difference between the electrode and the wafer. The result is showed on the figure 7: potential mappings from a first wafer rinsed one time, and another rinsed three times. The mappings confirm that at the center of wafer -where more water is dispensed- the charging is higher. The radial signature of potential does not change, but the surface charges doubled while tripling the rinse process.

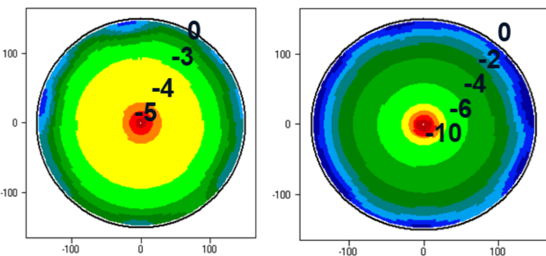


Figure 7. Radial signature of surface voltage obtained with Plasma Damage Monitoring tools. Left mapping corresponds to 1 rinse, right mappings corresponds to 3 rinses: surface potential doubled.

3.5 Image distortion caused by charging potential

As mentioned previously, CD-SEM images are subject to distortion due to the charges at wafer surface. Centroid shift values ΔXc ($Xc\ Wafer - Xc\ Reference$) and ΔYc ($Yc\ Wafer - Yc\ Reference$) are modeled with the following equation, similarly to the overlay's modeling:

$$\begin{cases} \Delta Xc = TransX + x.MagX - y.RotX \\ \Delta Yc\ = TransY + y.MagY + x.RotY \end{cases} \quad (1)$$

With x and y the coordinates of the holes, and Rot, Trans, and Mag the terms of rotation, translation and magnification, schematized on the figure 8.

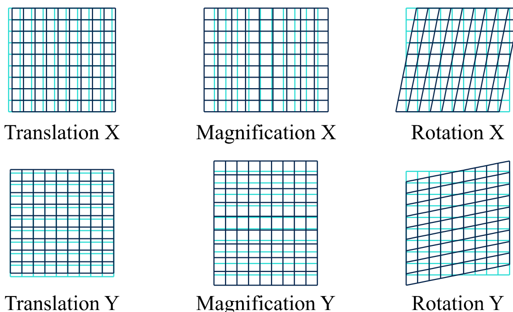


Figure 8. Interpretation of the corrective terms in equation 1. Adapted from [15].

The systematic phenomenon of image distortion can be seen figure 9 on a CD-SEM image of vias. The contours extracted (in green) are shifted from the layout (in blue) at the top left and bottom right corners of the image. The figure 9.b represents these ΔXc and ΔYc raw values. The arrows are colored by the value of the centroid shift vector. The equation (1) is used to model the data and the result is shown on the figure 9.c. Residuals are obtained by subtracting modeled data to the raw data, plotted figure 9.d. Residuals are less than 3 nm and are random: the 6-parameter fit is adequate.

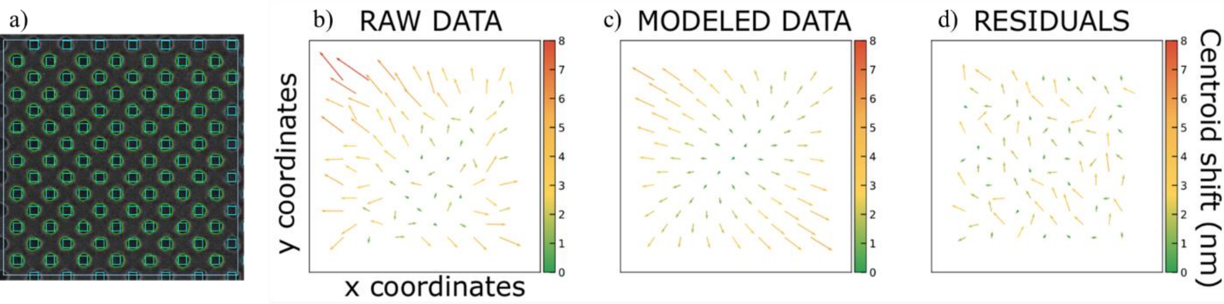


Figure 9. Correcting the impact of SEM image distortion on the centroid shift values: a) raw image, b) raw centroid shifts in this image, c) modeled centroid shifts and d) residuals centroid shifts.

The figure 10 shows the modeled centroid shifts of one image taken at the center of the wafers POR and 6-rinses. For comparison purposes, the scale of the modeled centroid shift data is the same for both wafers. We observe that the modeled signature of the wafer POR is less "pronounced" than the signature of the 6-rinses wafer, with a maximum centroid shift of 3.5nm, compared to 6nm. This gives an initial indication that rinse increases image distortion.

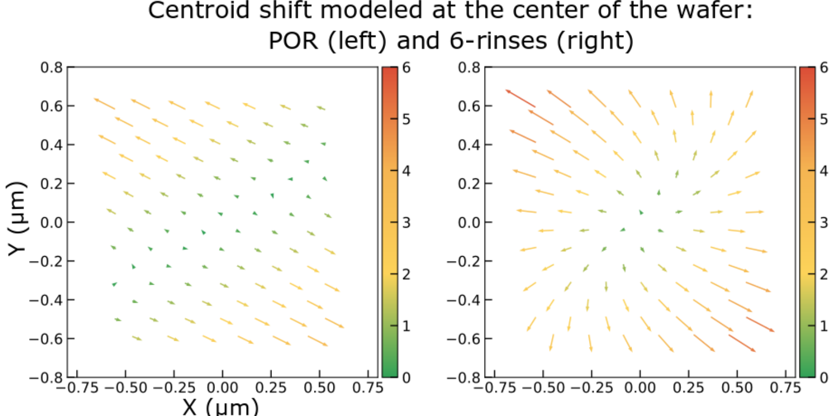


Figure 10. Centroid shift modeled for one image (center of the wafer) for both POR and 6-rinses: higher SEM image distortion is observed on the 6-rinses wafer.

To verify this hypothesis, we have modeled centroid shifts of the 65 images. For each image, we get Trans, Rot and Mag parameters. The figure 11 (left) shows the evolution of the Mag_X parameter (quantifying the horizontal scaling of the image) as a function of the radius at which the image was acquired. The pink and the blue curves correspond respectively to the 1 and 6 rinse steps wafers. For both wafers, the scaling parameters are the highest at the center of the wafer. The first conclusion to be drawn is that image distortion decreases with radius. Moreover, as we suspected, a clear shift between the two curves is seen: rinse increases image distortion. This behavior seems to be generalized for the two other wafers (figure in the Appendix).

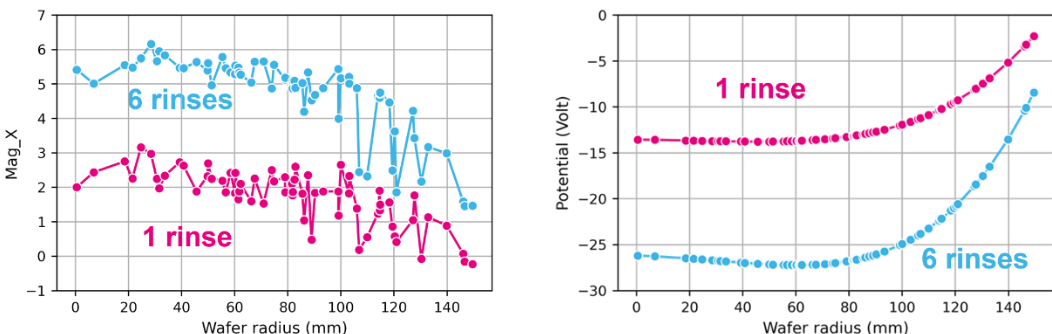


Figure 11. Scaling parameter (left) and Wafer potential (right) as a function of radius of the image. Pink points represent the POR (1 rinse) and blue ones the 6 rinses process.

Likewise, surface voltage is also plotted (on the right) as a function of the wafer radius. A very similar radial signature is observed, as well as a higher quantity of negative charges for 6 rinses than for 1. It confirms the hypothesis that rinse creates charging which is most likely the cause of image scaling. The multiplication of rinse steps also seems to increase the potential difference between the center and edge of wafer (around -10 V for the POR wafer, compared with almost -20 V for the 6-rinses wafer). At this point, the question remains: is PCDU's degradation with rinse a real patterning effect, or is it merely a consequence of these greater image distortions?

3.6 Correction of image distortion phenomenon

As rinse number increased, more pronounced image distortions were observed. To remove its impact, the following methodology is used. From the previous centroid shift modeling, the 6 parameters defining the translation, rotation and magnification are kept. These parameters are injected in the equation (1), this time to calculate each new coordinates of the contours. A resulting correction is shown figure 12, where we see contours before and after the image distortion correction. Now PCDU can again be computed, to investigate its evolution with rinse, this time without this metrology artefact.

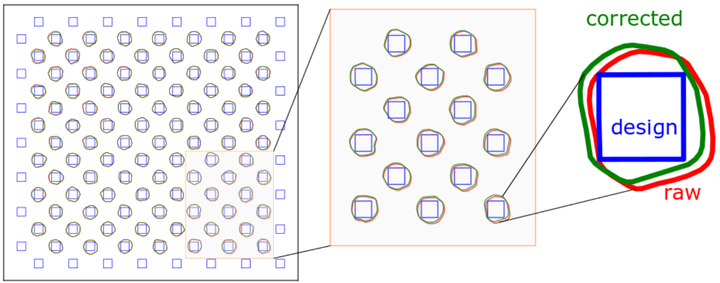


Figure 12. Correction of the whole contour, thanks to the modeling of centroid shift values.

Both the corrected (green) and raw (pink) PCDU values as a function of rinse steps are represented figure 13. A schematic red dashed line represents the PCDU values we would obtain after correction if image distortion were the only cause of PCDU's degradation. From this figure, the first conclusion to be drawn is that PCDU measured on the corrected contours is still increasing with rinse. Moreover, the offset between the raw and corrected PCDUs remains constant: this means that the impact of image distortion on shape degradation is constant, even though we previously found a higher distortion for the last two wafers.

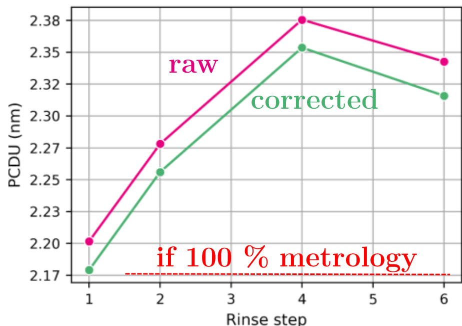


Figure 13. PCDU as a function of rinse steps: raw data in pink, corrected (no image distortion) data in green: impact of rinse on PCDU's degradation is still visible.

We conclude that increasing rinse steps number produces holes "roughness". Resist swelling is a cause we reject, because of the hydrophobic property of the 193nm lithography resist. The hypotheses of a mechanical effect (impact of water destroying part of the resist?) or a reaction between the negative charges and a component of the resist cannot be ruled out.

These vias from this study were not etched, but from the previously shown AEI in-line baselines (figure 3), it appeared that this effect was also observed on post-etch holes, confirming again that this phenomenon is not only metrology-related.

3.7 CD's behavior with rinse

The way CD is evolving with rinse is not very clear to us yet. Our first intuition was that rinse should not affect CD. Nevertheless, the figure 14 shows an increase of 3% (compared to the CD target) between the wafer POR and the second wafer. This increasing is too high to be due only to intra-lot variability.

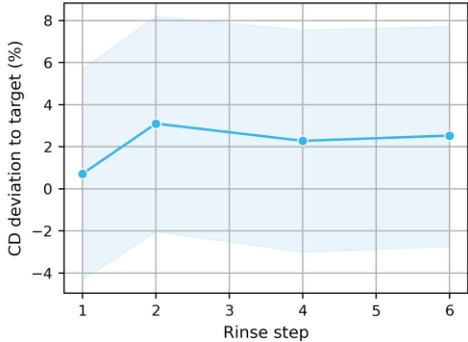


Figure 14. CD as a function of rinse step number. One point = one value averaged per wafer.

4. CONCLUSION

The monitoring of one metric quantifying post-lithography local shape variability of hole (PCDU) have revealed a clusters mismatching in a HVM environment. Post-etch in-line data of the same lots showed a similar trend. Hence, remote contour-based metrology enabled us to detect an issue and to start the investigation to understand its root cause. Engineers have adapted the rinse recipe to match the two clusters together. A subsidiary experiment confirmed that rinse has an impact on metrology since it exacerbates image distortions due to the charges accumulation. Regardless of the rinse recipe, we observed that image distortion decreases with the wafer radius. After isolating the patterning effect (by correcting image distortions), we confirmed that rinse creates shape variability of holes inside the resist.

Appendix:

The previously commented figure 11 is completed with the data coming from the two other wafers. The charging increases with the rinse (figure 15 right), reaching a saturation at 4 rinse steps. Likewise, the image distortion increases with the rinse (figure 15 left), always with the same radial signature.

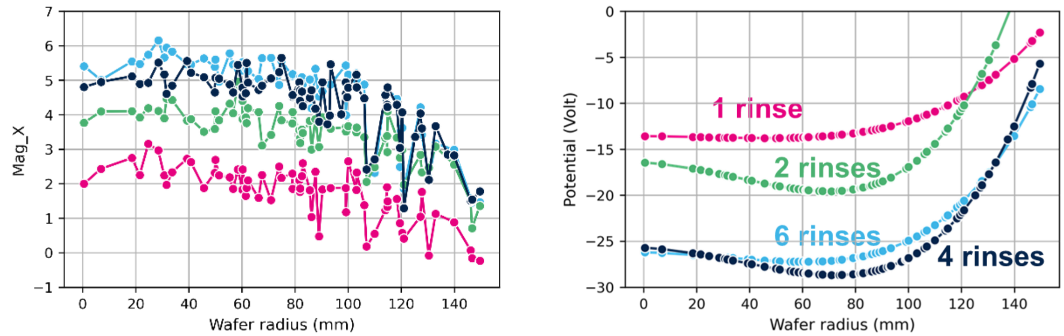


Figure 15. Scaling parameter (left) and Wafer potential (right) as a function of radius of the image.

REFERENCES

[1] P. De Bisschop, E. Hendrickx, "Stochastic printing failures in EUV lithography," Proc. SPIE 10957, Extreme Ultraviolet (EUV) Lithography X, 109570E (26 March2019)

[2] Mack, C. A. "Stochastic limitations to EUV lithography". 2018 International Symposium on VLSI Technology, Systems and Application, VLSI-TSA 2018

[3] B. Le-Gratiet, et al. "Investigating process variability at ppm level using advanced massive eBeam CD metrology and contour analysis", Proc. SPIE 10959, Metrology, Inspection, and Process Control for Microlithography XXXIII, 109591A (26 March 2019)

[4] Le-Gratiet, B., Bouyssou, R., Ducoté, J., Dezauzier, C., Ostrovsky, A., Beylier, C., Gardin, C., Petroni, P., Millèquant, M., Chagoya-Garzon, A., & Schiavone, P. (2019). « «Contour based metrology: getting more from a SEM image ».

[5] Soltani, E., Le-Gratiet, B., Bérard-Bergery, S., Pradelles, J., Bange, R., Schuch, N. G., Figueiro, T., & Tiron, R. (2022). "Highlighting stochastics manifestations in 193nm immersion lithography with contour-based metrology metrics". Proc. SPIE 1205308(May),

[6] Soltani, E., Le-Gratiet, B., Bérard-Bergery, S., Pradelles, J., Desmoulins, S., Sicurani, E., & Tiron, R. (EMLC 2022). "Predicting DUV open contact risk with scarce sampling using new contour-based metrics."

[7] Le-Gratiet, B., Bouyssou, R., Ducoté, J., Dettoni, F., Bourguignon, T., Morin, V., Bange, R., Schuch, N. G., Nicoulaud, J., Renault, G., Robert, F., & Figueiro, T. (2022). "A proof of concept of remote contour-based SEM metrology integration in HVM environment". June 2022

[8] Yang, X., Ye, Y. Z., Zou, Y., & Zhu, X. (2015). "Lithography develop process electrostatic discharge effect mechanism study". Metrology, Inspection, and Process Control for Microlithography XXIX, 9424(March 2015), 94242D

[9] Dillen, H., Kiers, T., Halder, S., Wallow, T. I., & van Roey, F. (2017). "CD-SEM distortion quantification for EPE metrology and contour analysis". Metrology, Inspection, and Process Control for Microlithography XXXI, 10145 (March 2017)

[10] D. S. L. Mui, E. H. Lenz, C. Cyterski, K. Venkataraman and M. Kawaguchi, "Wafer Surface Charging Model for Single-Wafer Wet-Spin Processes," in IEEE Transactions on Semiconductor Manufacturing, vol. 24, no. 4, pp. 552-558, Nov. 2011.

[11] Murali, V., Wu, A. T., Chatterjee, A., and Fraser, D. B., VLSI Symp. Tech. Dig., p. 103 (1990).

[12] Halladay, J., Teeter, B., Newcomb, R., Usry, W., Yoo, J., Lam, K., Lansford, J., and Brennan, B., *SEMATECH Surface Preparation and Cleaning Conference (SPCC),* p. 02-10 (March 24, 2008).

[13] Reinhardt, Karen A., Reidy, Richard F., Marsella, John A. "Handbook of Cleaning for Semiconductor Manufacturing: Fundamentalhs and Applications" (2010).

[14] Chen, H., Hu, S., Liu, X., Shi, H., Shen, M., & Yang, Y. (2022). Study of Negative Charge Accumulation Mechanism and Removal Method on the Wafer Surface in Via Photo Development Process. *2022 China Semiconductor Technology International Conference, CSTIC 2022*, 15–17.

[15] Bourguignon T., "Exploration of a non-destructive on device overlay metrology on FD-SOI technologies", PhD to be published in 2024.

Training Dataset Optimization for Deep Learning applied to Optical Proximity Correction on non-regular hole masks

Mathis URARD [a b], Clément PAQUET [a], Charlotte BEYLIER [a], Jean-Noël PENA [a], Alice CAPLIER [b], Mauro DALLA MURA [b c], Romain BANGE [a], Roberto GUIZZETTI [a]

[a] STMicroelectronics, 850 rue Jean Monnet, 38926 Crolles Cedex, France
[b] Univ. Grenoble Alpes, Gipsa Lab, 11 rue Maths, 38400 Saint Martin d'Hères, France
[c] Institut Universitaire de France (IUF), Paris, France

Abstract

With the machine learning breakthroughs in the past few years, the number of studies applying this principle to lithography steps is increasing constantly. In this article, the focus does not concern the learning models for OPC masks improvement, but the optimization of the data used for such learning. This part is essential for a good learning process, but has rarely been studied, despite its impact on the output results quality being as important as an improvement of the learning model. Several optimization methods are discussed, each with a specific objective: either reducing learning time, increasing the obtained results quality, or both. To evaluate these different results, classical optical proximity correction simulation tools are used, allowing for a complete evaluation in line with production standards.

Keywords: Mask data preparation, Deep Learning, Dataset optimization, OPC, Generative Adversarial Network

38th European Mask and Lithography Conference (EMLC 2023), edited by
Uwe F.W. Behringer, Jo Finders, Proc. of SPIE Vol. 12802, 1280209
© 2023 SPIE · 0277-786X · doi: 10.1117/12.2675612

1. Introduction

Optical Proximity Correction (OPC) is a technique used in photolithography to improve the accuracy and resolution of printed features on a semiconductor wafer [1]. It involves making small adjustments to the mask patterns used in the lithography process to compensate for the non-linearities and distortions that occur during the printing process. These adjustments are based on computer simulations of the lithography process and are used to optimize the mask patterns to print the targeted solution on the wafer. OPC is an essential technique for producing high-performance semiconductor devices with ever-shrinking feature sizes. The conventional OPC methods include different techniques:

- *Rule-based* using predefined rules to correct design pattern. This technique is advantageous for less advanced technologies, but not so accurate for advanced technologies due to its inability to take non-linear effects into account.
- *Model-based* applying physical models and simulations for more advanced corrections. This technique can handle complex lithography effects, offering higher precision for cutting-edge technologies but requiring more time and resources to build and use models.
- *Hybrid* combining rule-based and model-based methods to obtain the benefits of both techniques but can be more complex to implement and optimize than purely rule-based methods.
- *Inverse Lithography Transform* (known as ILT) used to determine the design patterns which, when applied in lithography, will produce the desired result. This technique is capable of solving complex lithographic problems with optimal results for advanced technologies but requires significant computational resources and can be slower due to the complexity of inverse techniques.

The choice of OPC method depends on the design complexity, lithography process parameters, and desired level of accuracy, resolution and computational resource.

In this paper, Deep Neural Networks (DNN) are used to generate masks mimicking the conventional model based OPC technique. The goal is to obtain a final mask faster and with the same level of quality as the one generated by traditional OPC models. As the DNN learns directly from optimized masks, the generated results quality can at best reach the quality of the data it learns from.

A part of the scientific literature concerning DNN applied to OPC (for example [2-4]), focuses on improving learning methods, especially DNN models architecture exploration and optimization, rather than optimizing the data used for learning. This optimization is mandatory as designs used are non-regular (i.e., composed of irregular shapes patterns), which increases the diversity of pattern types compared to regular designs used in the literature, and consequently, the complexity of DNN training. It is then essential to optimize data for consistent learning.

This study aims at improving learning results by mainly acting on the data used for training. In the AI general case, randomly distributed training datasets can generate biases and reduce DNN models performance since a DNN will learn on all data equally. In general cases, this is not a problem because DNN models results are sufficient in most applications which is not true in our case. However, because of the huge pattern diversity of optimized masks, the resulting training database usually does not show a homogenous pattern distribution, as shown in this paper. Therefore, one strategy relies on carefully selecting patterns from the training database to achieve better training results by learning from a balanced dataset. The study aims at finding a suitable method for composing a robust dataset to train a DNN that will generate OPC (and so optimized masks) for hole layers. This approach could provide runtime saving by removing deemed unnecessary images from DNN training and could improve quality by supporting the learning for specific shapes and patterns to achieve the desired outcome.

The remainder of the paper is organized as follows: section 2 introduces the application of DNN in OPC, followed by the presentation of different methods to optimize the training dataset in section 3. The results obtained for each method are summarized and compared in section 4. Concluding remarks are drawn in section 5.

2. Machine Learning based method to generate OPC

The chosen machine learning approach is convolutional neural network learning (CNN). This DNN approach enables image generation, here used in supervised learning. In this type of learning, the algorithm receives labeled input and output examples and learns to associate them by adjusting its internal parameters. The goal of a supervised learning model is to minimize a loss function that measures the difference between the generated output and the correct output. This training loss value can also be a general metric of the quality of a network. Two sets of data are required to differentiate the input and expected output, distributed in a dataset. These data are assembled by corresponding input/output to form a training dataset to feed to the DNN. As this process is iterative, once the DNN has learned from the entire dataset, this step is repeated to allow the network to adjust its internal parameters.

This study focuses on the impact of the dataset used for learning of non-regular designs, on the final quality of the DNN output, therefore any other parameter, such as network architecture, is identical across every test that will be presented.

To achieve the image-to-image translation task of generating OPC mask from an input target, a Generative Adversarial Network (GAN) [5] is used, composed of two separate neural networks as shown in Figure 1. The Generator is the main network of this architecture, as it is the one mapping the input target to the post-OPC mask. The Discriminator judges if the output of the Generator seems true or fake compared to a database of pre-optimized OPC mask, this means that the Generator must always get better to fool the Discriminator. Among the various existing GAN types, the one called Pix2Pix [6] is used as a basis and adapted to our problem. Although expected results differ from classical image generation, as it requires precision in the output instead of a good-looking result, only minor changes were necessary from state-of-the-art model architecture to generate good results. These changes affect losses to focus the learning on the crucial points: shapes themselves and not the empty space around them.

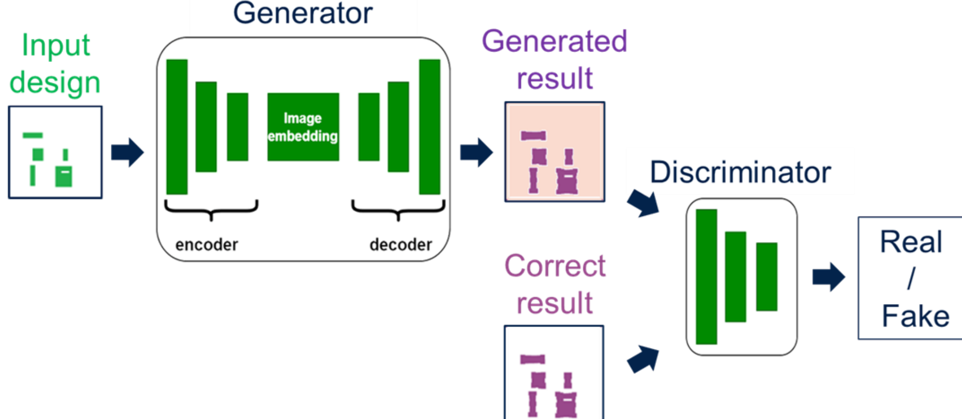


Figure 1: Proposed DNN architecture based on pix2pix for OPC generation

In our case, the training loss metric is not sufficient since changing the dataset will also change the images on which the loss is computed as explained in the following paragraphs.

3. Strategies to build a dataset for training a DNN

Studying a dataset means looking at the composition of its population. In the present case, these are pairs of images representing a possible input and its associated output. These images are made up of several shapes that represent design configurations. Details of these different shapes are given in section 4.1. Using in-house classification tools developed in ST, we can determine the composition and number of shapes present in each image. All this data is available directly in the database used for optimization methods.

As usual in AI general case, a dataset is split into training and validation sub-datasets with respectively 80% and 20% of the total data randomly selected. The training sub-dataset is used to properly train the DNN, to map input targets with optimized masks. On the other hand, the validation sub-dataset is only used to evaluate the DNN at each epoch on data not used for training, and therefore to provide a more objective measure of its quality.

3.1. Detailed optimization methods

The first dataset creation method covers the entire database to create what we call the basic dataset. This is the most basic approach, with no specific selection. The goal of this basic dataset's training is to assess the ability of our DNN model to learn from data generated by conventional OPC tools. After this first step it will be possible to correct some learning inconstancies by acting directly on the Basic dataset.

The second method, called reduction, used to obtain a training dataset, is to reduce the quantity of images by sorting them according to their composition. The goal of this approach is to remove all non-essential images to optimize learning time, while balancing the impact of different image types. Remember that the principle of neural network learning is to adjust the network's internal parameters according to its training images. It is therefore necessary to have information redundancy to enable efficient learning without omitting any case. The principle used is simple: a selection of the training images is made based on the information about their composition stored in the database. Selected images are then added to the reduced dataset.

The next method, called augmentation, consists in increasing the proportion of images on which the network delivers incorrect learning results. In general AI, dataset augmentation refers to the duplication of images by adding a rotation and/or a flip. This augmentation method has been tested, without giving good training results: the dataset used in OPC might be oriented. This means that changing the orientation of the image frames using traditional augmentation methods is not an option. In our approach, an initial training is made on the basic dataset followed by the evaluation of the obtained results using the methods presented in section 3.3. New images centered on the detected defects are created and added to the training sub-dataset to produce an augmented dataset. This method is iterative since any training result can be used as an augmentation baseline

to improve future training results. In this study, only one iteration will be presented to evaluate this method. This method is summarized in Figure 2.

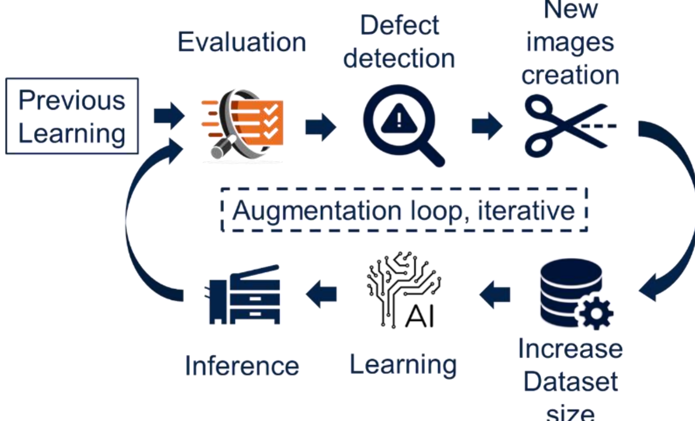


Figure 2: Illustration of Augmentation process

The final method to obtain a training dataset consists in applying a reduction on an augmented dataset. The key feature of this method is to reduce the number of images considered useless for learning without removing images from the augmentation. Reducing an augmented dataset produces a harmonized dataset. The goal of this approach is to only take the augmentation and reduction method's good contributions, i.e. filling learning deficits and reducing training time.

3.2. Pros &cons of each method

The method leading to the basic dataset aims at determining the data distribution impact as training data. This is the first step before trying any optimization methods. As this method trains on the entire database, studying its results can reveal representativeness problem of the different patterns and shape types of the database. To prevent training on the least recurring patterns being overwritten by the most recurring patterns in the dataset, it is important to balance their occurrence. This principle is a basic one in DNN and cannot be respected in the case of the basic dataset due to the high disparity in distribution of certain zones of the design chosen to create the database in Figure 3. In this design, there are two main zones: one where shapes are sparsely placed without frequent pattern repetition, called digital design, and a much denser and repetitive part called memory design. We know that classical OPC mathematical rules differ between these two zones, which causes a representativeness problem due to a difference in terms of shapes composition. In the design used for learning, memory design represents less than 3% of the total design and this is about the same proportion as in the database.

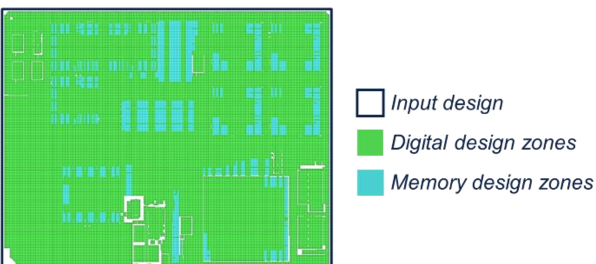


Figure 3: Example of an optimized design used as learning input. Zones containing shapes are colored, white zones do not contain any shape.

As we know, each image is composed of several types of shape, it's therefore impossible to create a perfectly balanced dataset with the reduction method. In addition, as the shape patterns present in digital design are extremely diversified, it's essential to keep a relevant quantity of this zone to have correct learning. The goal of this method is to improve results by acting on dataset distribution while reducing the time needed for learning.
Regarding the augmentation method, the whole point of this method is to correct learning issues by adding more elements where the network makes learning mistakes. In theory, this method seems very promising, although it only aims at increasing the size of the database, which directly impacts training time. However, this method requires learning results evaluated on a different design than the one used for training, but having the same mathematical rules from OPC point of view. It is therefore necessary to have at least three similar but not identical designs, to:

- perform the first training on the first design dataset
- evaluate this training on the second design
- create images from the defects highlighted by the evaluation
- add those images to the first design dataset, which becomes augmented
- train the second network on the augmented dataset
- evaluate the second training on the 3rd design.

Obviously, this not only increases the second network's training time, but also requires a considerable amount of time to complete each step. However, the goal is to improve the neural network's results and push the network beyond the limits reached with the basic dataset.
The last method, harmonization, is based on both previous methods. It therefore requires just as much preparation and different designs as the augmentation method, but also a selection of the data from the augmented dataset. The time saved in the learning part by reducing the dataset is to be considered as part of the overall method. The goal is to reduce the time needed to improve the results obtained with the basic dataset.

3.3. Evaluation methods

Two testing levels, through its learning loss and with the classical OPC simulation tools, have been set up to evaluate the training efficiency. First, we look at the learning losses, which measure the difference between DNN model's predictions and expected results. This evaluation level takes place directly during a training session.
The loss used in a learning is a Fraction of Variance Unexplained (FVU):

$$FVU = \frac{MSE(f)}{var[Y]}; with \begin{matrix} MSE\ (f)\ the\ Mean\ Square\ Error\ of\ the\ regression\ function \\ Y\ the\ measured\ variable \end{matrix} \quad (1)$$

This loss can be summarized as follows:

$$FVU=1- R^2,\ \ with\ R^2\ the\ correlation\ coefficient \quad (2)$$

It gives a first estimation of the results' quality. Since all the tests presented in this study were launched with exact same learning parameters, it is possible to compare their FVU, which gives a quick comparison of the learning quality between the different tests.
The last step in the evaluation process is to use classical OPC simulation tools. This method allows us to obtain an exact output quality estimation, with a minor drawback: it takes time to run. This evaluation uses the trained DNN model to generate OPC on a 42mm² design with about 1 billion shapes. This design is completely new to the network and follows the same design rules as the ones used for training. This method is used to simulate the mask generated by the DNN and evaluate it with the same criteria used for classical OPC solutions, to confront different learning results.

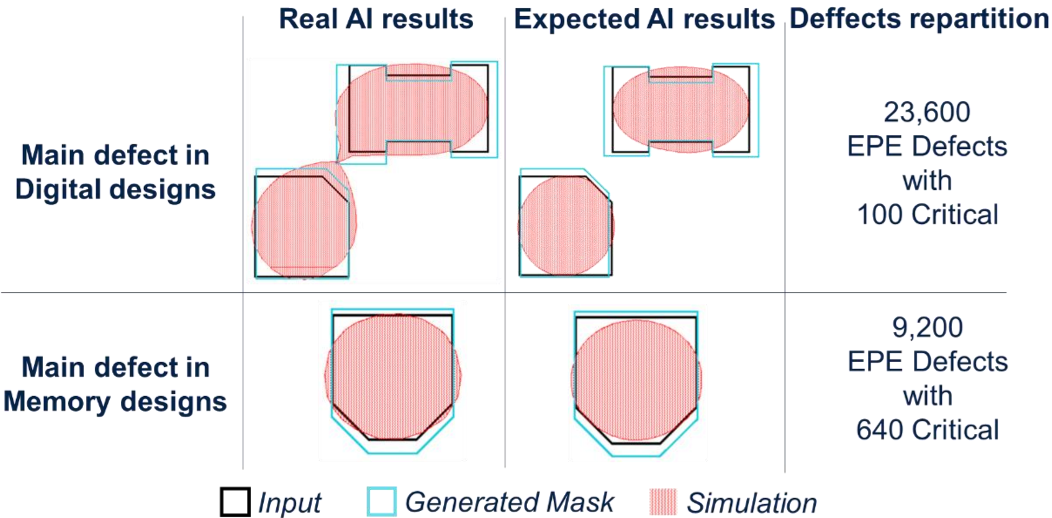


Figure 4: Example of the most recurrent critical defects in digital and memory design from a basic training. The digital defect shown is called a bridge and is referenced on 88% of the defects detected in digital design. The defect detected in memory design is a simulation-size defect and is found in 75% of the defects in its design section.

Figure 4 illustrates an example of a simulated defect caught by classical OPC simulation tool. Thanks to ST's expertise, defects are divided into 2 categories: EPE for Edge Placement Errors and Critical for patterning defects. In general, the objective is to eliminate critical defects while reducing the number of EPE defects as much as possible. Indeed, very small EPE defects might be acceptable, since the impact on simulation is smaller than process variability.
The use of classical OPC simulation tools also means that the DNN is evaluated with the same quality standard as the one used in production.

4. Experimental analysis

As preliminary results, details of the datasets obtained by the different methods and their composition will be described in section 4.1, followed by the different learning results comparison in 4.2 and discussions about these results in 4.3.
A complete iteration over the training dataset is called an epoch and a complete training usually takes many epochs to converge. In our case, we consider that the convergence is satisfying after around 40 epochs. The machine used for this study is composed of a Nvidia A100 GPU with 80GB VRAM and takes around 4 days to complete the required 40 epochs.

4.1. Experimental Set-Up

The first case to study is the basic dataset composition, which contains the entire database before any modification. We know that a design is made up of many different types of shapes and assume that each one represents a distinct mathematical rule. It is therefore important to look at the quantity and the occurrence of each type of shape to know the exact composition of the dataset. In-house detection algorithms are used to identify and classify the shapes present in the images and store information in the SQL database.

Figure 5 shows the composition of the basic dataset, with the occurrence of each shape type. It's also interesting to know how these types of shapes are distributed across images. This indicates whether the shape is uniformly

distributed within the optimized mask or if it is exclusive to specific areas. Inversely, if a few clusters of this shape are identified in only few images, this could indicate that a specific mathematical rule applies on this pattern. As an example, let's look at the distribution of *Rectangle_2* and *Merged_rectangles* shapes in Figure 5. The diagram a) indicates that *Merged_rectangles* are occurring in less images than the *Rectangles_2*; The diagram b) focuses on shapes distribution, which is comparable for these ones (0.93% / 1.21%). A first hypothesis would be that *Rectangle_2* are found in more zones of the mask, leading to a higher variability than *Merged_rectangles*. The term variability refers to a variation in the proportions in height or width of the respective shape compared to its general type of shape.

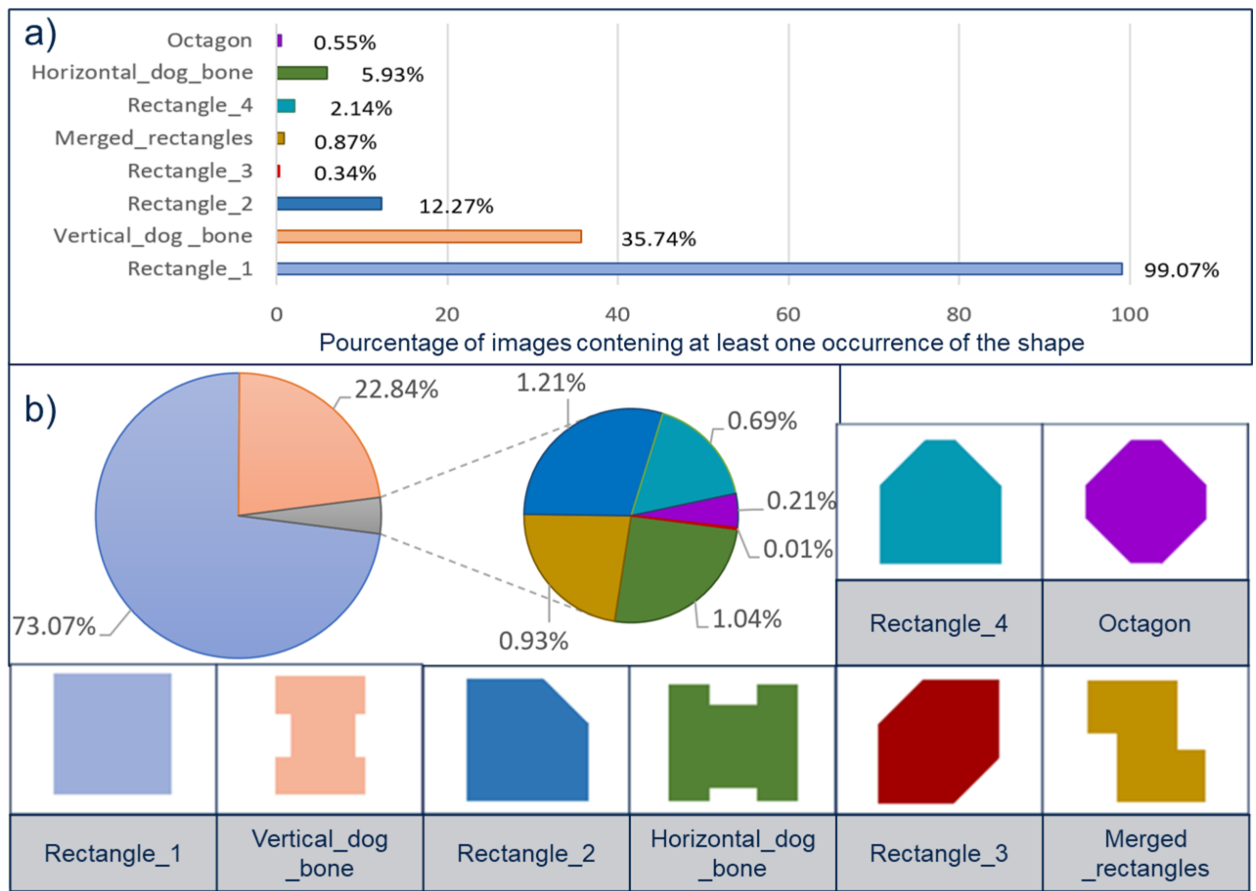


Figure 5: Composition and shapes details present in the basic dataset. A) percentage of images containing at least one occurrence of the shape. B) percentage of shape in the whole database.

Figure 5 shows the disparity of the shapes distribution. The difference between Figure 5.a) and Figure 5.b) is the way sampling is performed: in a), sampling is based on images in the dataset, in b) sampling is based on the total amount of shapes of the dataset (basic dataset size: 85900 images). By looking at the quantity of each type of shape, we can expect that a training would not give optimal results. This is due to a learning difference by the DNN for each type of shape, directly linked to their apparition rate in the given dataset. It's therefore important to have enough samples for the DNN to learn effectively.

About the reduction method, the principle is simple: according to their composition, images can be selected using an SQL database. The first attempt was to reduce the number of images containing only rectangles. The most conclusive dataset composition for this type of reduction is referenced in Figure 6. a. A second reduction,

involving *Vertical_dog_bone* in addition to *Rectangle_1*, was tested. The goal here is to reduce the quantity of the two over-represented shapes in the dataset. Several proportions were tested and Figure 6.b references the one providing the best results. The last type of reduction tested proportionally reduces the quantity of images containing certain types of shapes according to the results obtained with a previous training made on the basic dataset, referenced in Figure 6.c.

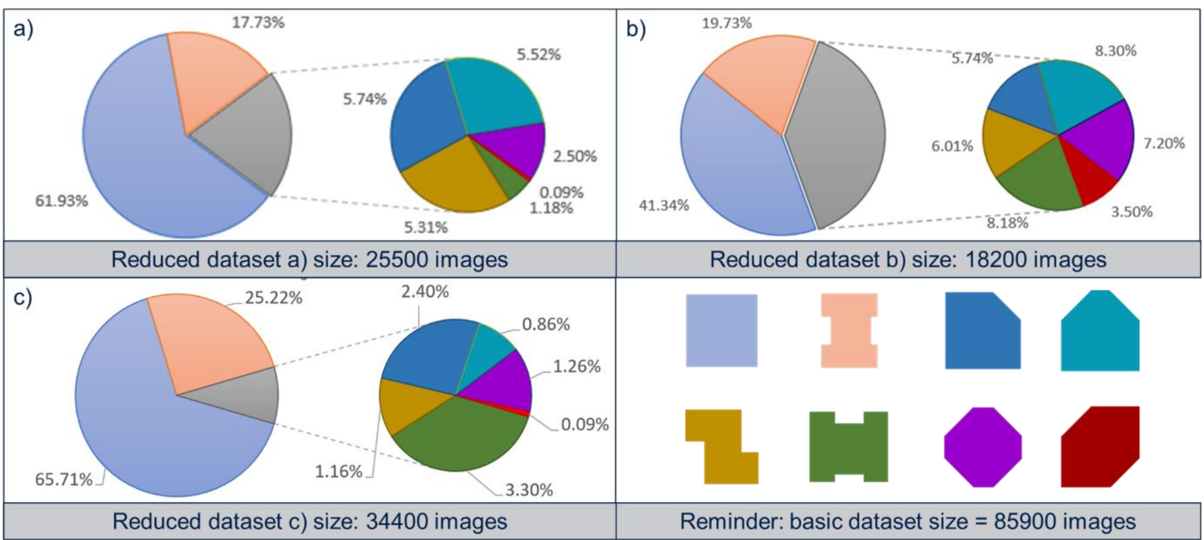


Figure 6: Different reduced dataset distribution comparison. Reducing the size of the reference basic dataset will reduce the time needed to learn an epoch.

Regarding dataset b), it should enable high learning performance, thanks to its fairer distribution of under-represented shapes with a decent dataset size.

For the other two dataset optimization methods, details of the obtained datasets are shown in Figure 7.
Diagram a) shows the distribution of shapes obtained with the dataset augmentation method. Many under-represented shape types have been augmented, notably *Merged_rectangles* (yellow) and *Rectangle_4* (cyan). The proportion of *Rectangle_2* (dark blue) and *Octagon* (purple) have also been correctly increased. On the other hand, the increases made on *Horizontal_dog_bone* (green) or *Rectangle_3* (red) were not significant. For shapes shown in red, their low presence in the chosen design for augmentation can be a root cause.
For the green shape type low augmentation, there are several hypotheses:
- As *Horizontal_dog_bone* is a type of shape equivalent to *Vertical_dog_bone* but rotated by 90°, the *Vertical_dog_bone's* proportion in the training dataset should help to generate this type of shape.
- If the first learning used as a basis for augmentation correctly generated this type of shape, the augmentation will create only few images from the augmentation.
- The *Horizontal_dog_bone* shape type has little variability in its proportion and is therefore easy to learn.
For the first hypothesis about shape rotation, it is known in global DNN approaches that convolution networks specific mathematical rules do not lead to natural invariance to rotation, so this hypothesis is not validated. Referring to Figure 7.a), we understand that in the basic dataset, this shape is represented in around 6% of images. In addition, we know that this type of shape is mainly present in the Digital Design section, which invalidates the hypothesis of low variability.

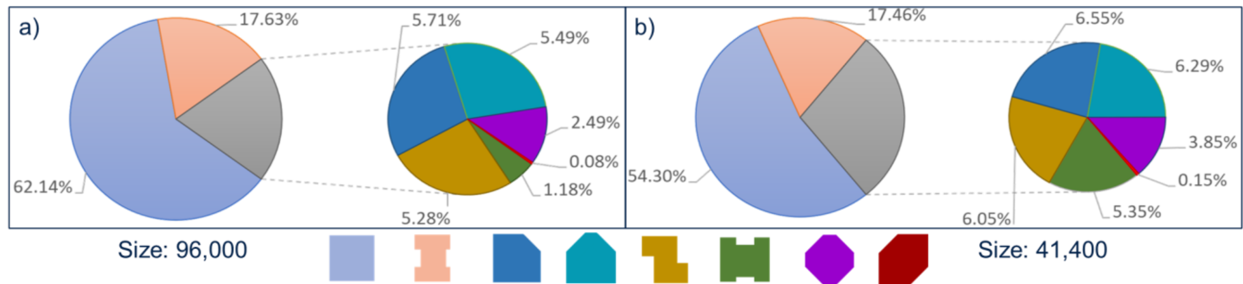


Figure 7: Details of dataset size and composition generated by two different methods: augmentation in a) ; harmonization in b)

Figure 7.b) shows the distribution obtained using the harmonization method. By starting from the augmented dataset and applying a reduction on the two most represented shape types, the size of the resulting harmonized dataset is divided by two. The idea of this harmonization is to obtain a more balanced dataset compared to the augmented dataset. In the result, the proportion of *Horizontal_dog_bone* (green) has been increased in the harmonized dataset, and it almost doubles the proportion of red shapes, which is still insignificant.

4.2. Results and discussion

After having examined and compared the different datasets obtained with each method, let's now look at the training results obtained. To enable effective comparison, all the results shown in this section are obtained by applying the trained DNNs on the same test design. As a reminder, the machine used for this study is composed of a Nvidia A100 GPU with 80GB VRAM and takes around 4 days to complete the required 40 epochs on the basic dataset.

First, it is necessary to choose the best reduction dataset. To do this, the three tested reductions are evaluated using classical OPC analysis tools, and the detected defects are divided into two categories. All results are summarized in the table in Figure 8.

Reduced dataset	Dataset size	Proportion of under-represented shapes	Time required to reach 40 epochs	EPE defects	Critical defects	Total defects
Reduction a)	25500	20%	2 days	387000	10400	397400
Reduction b)	18200	39%	1.7 days	Over 1M	687000	Over 1.6M
Reduction c)	34400	9%	2.5 days	131800	1260	133060
Basic	85900	4%	4 days	19800	860	20660

Figure 8: Error distribution for all tested reductions

In terms of result quality of the different reduction trials, reduction c) gives the best results, but is the weakest reduction. The other two reductions give poor results, with over 10,000 critical defects for a) and 687,000 for b). As a reminder, this method's goal is to reduce the learning time without altering the results' quality.

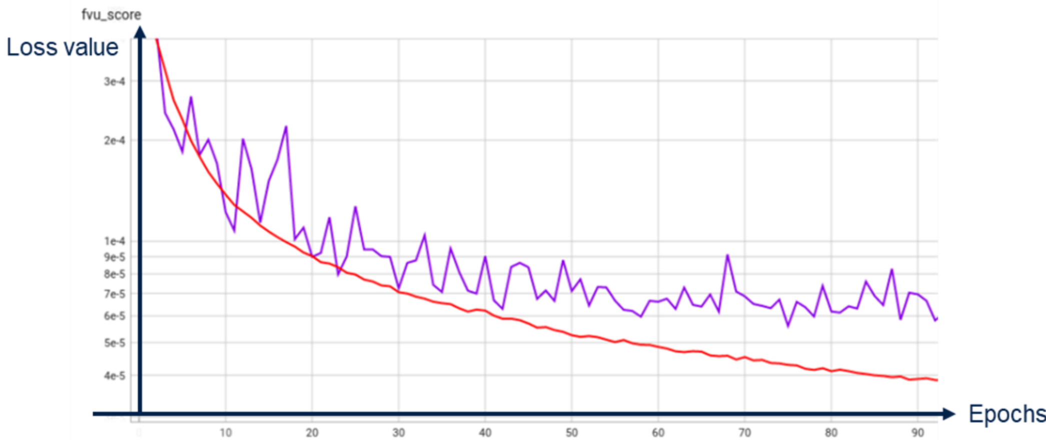
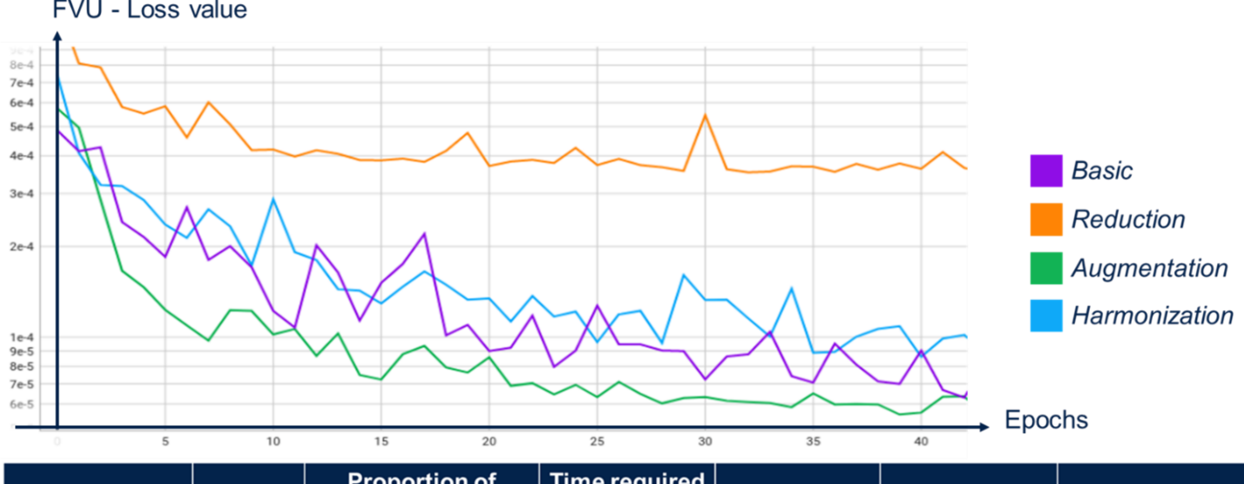


Figure 9: Example of learning loss evolution through learning epochs, the training sub-dataset score appears in red and validation sub-dataset in purple.

Figure 9 shows the FVU (1) loss evolution for a training performed on the basic dataset. Only validation sub-dataset loss can reveal real learning progress. In this case, we consider that learning is not really evolving after 40 epochs and not at all after 55 epochs

Now let us compare the results provided by all dataset optimization methods together. Concerning the Basic dataset, shown in purple in Figure 10, the resulting learning loss on the validation sub-dataset is encouraging. Note that the validation loss has a lot of variation compared to the training loss, as shown in Figure 9. Such a difference can be explained by a poor distribution of shape types between the training & validation sub-datasets. As this image repartition is made from a random selection, it is probable to have distribution disparities.



Tested dataset	Dataset size	Proportion of under-represented shapes	Time required to reach 40 epochs	EPE defects	Critical defects	Total defects
Basic	85900	4%	4 days	32800	740	33540
Reduction	34400	9%	2.5 days	131800	1260	133060
Augmentation	96000	20%	5 days	750	330	1080
Harmonization	41400	28%	3.5 days	19800	860	20660

Figure 10: Comparison of the results of each optimization method. The upper part shows the evolution of the loss functions and the table references the results obtained.

When looking only at the learning curves, it's clear that the reduction method doesn't give good results compared to the basic dataset curve. This observation is confirmed by the number of defects: almost 4x more EPEs defects and 2x more critical defects.

The Harmonization method produces comparable results to the basic dataset. Unfortunately, the learning curves for the basic dataset and the harmonized dataset are too close to make any conclusions based on the curves themselves. However, the defect distribution is interesting here and leads for a real comparison: the result contains fewer EPE defects, but more critical defects. The network seems to have learned to better generalize the solution, at the cost of certain special cases that appear as critical defects. However, the network was able to learn and make fewer EPE defects with a shorter training time, which is enough to keep this method and continue to improve it.

In Figure 10 results, the augmentation method yields the best quality. Indeed, despite the small difference in the training curves, the classical OPC simulation tool shows a real quality gain from the network. As far as EPE defects are concerned, the quantity has been divided by 43, while the number of critical defects has been divided by 2.2.

To better understand the impact of dataset composition on result quality, a study of the number of EPE and critical defects was also performed for each type of shape, resumed in Figure 11.

Dataset									Total
Basic	5k	3k	4k	2k	11k	3k	5k	0.5k	33.5k
Reduction	17k	13k	16k	12k	28k	25k	17k	5k	133k
Augmentation	66	36	252	72	377	126	148	3	1080
Harmonization	5k	3k	3k	1k	5.4k	1k	2k	0.2k	20.6k

Figure 11: Distribution of simulation defects number by shape type for each dataset optimization

Looking at the defect distribution presented in Figure 11, several effects are interesting to analyze for Augmentation and Harmonization.

The first 2 columns of Figure 11 show that the proportion of errors on over-represented shapes dropped from 8k/33.5k (around 24%) on the basic dataset to 102/1080 (around 9%) on the augmented dataset. This means that adding images containing mainly learning errors on under-represented shapes also improved results on over-represented shapes.

Concerning the defect distribution on harmonized dataset, the opposite effect occurred with 40% (8k/20.6k) of errors coming from over-represented shapes. This is due to the selection made to reduce training time, which may not yet be optimal. However, all errors on under-represented shapes were reduced, especially for *Rectangle_2* (dark blue) where the number of defects was divided by 2.

5. Conclusion

To end on the exploration of different optimized training dataset building strategies, if an optimization method based on the results quality would to be chosen, then augmentation would be the right choice. If no distinction is made between simulation defect types, then the harmonization solution can be chosen, as its total number of defects is much lower than those given by basic dataset training for an optimized training time. Finally, the reduction method alone does not provide conclusive results, so it should not be used outside Harmonization. However, the results shown in this paper were achieved after just one iteration of Augmentation, so we can expect further improvements with more Augmentations' iterations. Therefore, it is entirely possible to greatly improve the generation quality of an DNN by playing on its training dataset alone. The image classification methods we've implemented are key to mastering the training dataset. Without them, the quality generation evaluation is blind and cannot be used correctly.

However, the shape sorting approach is not optimal, as the direct neighborhood of a shape has a direct impact on the generation of its OPC. A solution based on the neighborhood rather than the shapes themselves would give better results. Another way to achieve better results would be to do more augmentation loops. It is therefore important to optimize training data in addition to optimizing the architecture used for the DNN.

6. References

[1] Y. Hou and Q. Wu, "Optical Proximity Correction, Methodology and Limitations," *2021 China Semiconductor Technology International Conference (CSTIC)*, Shanghai, China, 2021, pp. 1-5, doi: 10.1109/CSTIC52283.2021.9461507.

[2] Chen, G., Chen, W., Ma, Y., Yang, H., & Yu, B. DAMO: Deep agile mask optimization for full chip scale. In Proceedings of the 39th International Conference on Computer-Aided Design (pp. 1-9). (2020, November).

[3] Jiang, B., Liu, L., Ma, Y., Yu, B., & Young, E. F. Neural-ILT 2.0: Migrating ILT to Domain-Specific and Multitask-Enabled Neural Network. IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, 41(8), 2671-2684. (2021).

[4] Ye, W., Alawieh, M. B., Lin, Y., & Pan, D. Z. LithoGAN: End-to-end lithography modeling with generative adversarial networks. In Proceedings of the 56th Annual Design Automation Conference 2019 (pp. 1-6). (2019, June).

[5] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. Advances in neural information processing systems, 27, 2014.

[6] Isola, P., Zhu, J. Y., Zhou, T., & Efros, A. A. Image-to-image translation with conditional adversarial networks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1125-1134). (2017).

[7] Yang, H., Li, S., Ma, Y., Yu, B., & Young, E. F. GAN-OPC: Mask optimization with lithography-guided generative adversarial nets. In Proceedings of the 55th Annual Design Automation Conference (pp. 1-6). (2018, June).

[8] Alawieh, M. B., Lin, Y., Zhang, Z., Li, M., Huang, Q., & Pan, D. Z. “GAN-SRAF: Sub-resolution assist feature generation using conditional generative adversarial networks”. In Proceedings of the 56th Annual Design Automation Conference 2019 (pp. 1-6). (2019, June)

[9] Apoorva Oak, Soobin Hwang, Ruoxia Chen, Shinill Kang, and Ryan Ryoung-han Kim. "Machine learning based recursive partitioning for simplifying OPC model building complexity", Proc. SPIE 11614, Design-Process-Technology Co-optimization XV, 116140O (22 February 2021) ; doi: 10.1117/12.2584704

[10] Weilun Ciou, Tony Hu, Yi-Yien Tsai, Chung-Te Hsuan, Elvis Yang, Ta-Hung Yang, Kuang-Chao Chen, "Machine learning OPC with generative adversarial networks," Proc. SPIE 12052, DTCO and Computational Patterning, 120520Z (26 May 2022); doi: 10.1117/12.2606715

[11] Hesham Abdelghany, Kevin Hooker, "Effective data sampling techniques for machine learning OPC in full chip production," Proc. SPIE 11613, Optical Microlithography XXXIV, 1161307 (16 March 2021); doi: 10.1117/12.2586176

[12] Shi, X., Yan, Y., Li, C., Xia, M., Pan, B., Gao, Y., & Yuan, W. Can optical proximity correction solution be learned? The learning limit and a general learning framework. Journal of Micro/Nanopatterning, Materials, and Metrology, 21(4), 043203-043203. (2022)

Invited Paper

LITHOSCALE® Features Accomplish Dual Exposure and High–Resolution Patterning

Ksenija Varga[a*], Thomas Uhrmann[a], Roman Holly[a], Tobias Zenger[a]
[a]EV Group, St. Florian am Inn, Austria, *K.Varga@EVGroup.com
Chris Milasincic[b], Mel Zussman[b], Ron Legario[b], Michael Knaus[b]
[b]HD MicroSystems L.L.C., 250 Cheesquake Road Parlin, NJ 08859-1241, USA

ABSTRACT

The concept of dual layer pattering as an alternative to POR in interconnect formation has been evaluated on maskless exposure technology. The novel dielectric materials negative tone PBO were used for the concept prove. The process optimization was conducted by varying the exposure and spin coating paraments. The lowest feature size accomplished was even <4 µm proving the two–step structures with FT of the first layer 50% of the dual layer FT. This new concept represents contribution to the improved CoO setup in the advanced packaging processes.

Keywords: LITHOSCALE®, dual–layer exposure, PBO positive tone dielectrics, via patterning, RDL patterning, advanced packaging, heterogenous integration.

1. INTRODUCTION

The fast–growing advanced packaging and heterogenous integration industries are facing several challenges. In terms of lithographic patterning, in the development of next generation advanced packaging technologies, application of steppers is facing a severe limitation. Accurate reconstitution of the wafers is a key parameter in integrating dies from various wafer manufacturers in a multi–die solutions. The current lithography steppers and others mask–based systems struggle to cope with inaccuracies from die placement and die shift variations caused by the over–molding.

The given reticle size and optics dimensions of static exposure systems limit the exposure area. This is particularly challenging in large die interposer fabrications, where stich–lines and or mismatches overlap regions of reticle exposure field can affect the electrical properties within the RDL. The ability to generate a stitch–less pattern for interposers exceeding current reticle size is important for advanced devices needed for complex layouts, such as advanced graphic processing, AI and high–performance computing (HPC) [1].

The solution to overcome the difficulties caused by the steppers in advanced packages is usage of maskless exposure lithography.

Maskless exposure technology refers to a concept where lithographic patterning is achieved without the use of masks or reticles. Instead, the patterning is performed directly on the substrate using a digital light source or other means of localized exposure. This eliminates the need for intermediate alignment steps typically required when using masks or reticles. MLE Technology was evaluated for a dual damascene process with the goal of a 50 % reduction in lithographic steps. Positive tone thick resist was exposed with two different doses using the multi–level exposure feature of MLE. Well–defined RDL structures and vias were simultaneously generated at different height levels with lateral dimensions of <5 µm. These results clearly demonstrate good process feasibility towards a new, cost–efficient implementation of the dual–damascene process. The proposed process flow is a general approach for different UV–patternable resists and materials needed to tackle many lithographic challenges in advanced packaging and heterogenous integration industry [2]. In terms of the materials, the challenge was to develop new high temperature dielectrics that are capable of curing at temperatures much lower than previously required (e.g. 350 °C – 380 °C). Low temperature cure dielectrics such as HD8900 Series are cured around 200 °C and thus are compatible with epoxy molding materials used in FO–WLP packages. These new low temperature cure dielectrics were also developed for markets like MRAM, RF, INFO, CMOS, MEMS, and backside RDL applications where

38th European Mask and Lithography Conference (EMLC 2023), edited by
Uwe F.W. Behringer, Jo Finders, Proc. of SPIE Vol. 12802, 128020A
© 2023 SPIE · 0277-786X · doi: 10.1117/12.2675899

the base substrate, other materials, or the device packages itself are temperature sensitive and require a low cure dielectric [3].
The motivation for the present investigation was to develop an alternative concept for the interconnect formation to industry–established dual damascene process. The goal was to pattern the novel low–temperature cure dielectric material by dual–exposure on maskless exposure technology (EVG's LITHOSCALE®). By one coating step, dual–layer exposure and one development step (without intermediate alignment), the overall process steps could be reduced contributing to the cost–efficient back–end processes.

2. EXPERIMENTAL

The present paper focuses on dual–exposure patterning of dielectric materials designed for FO WLP, the low–temperature cure positive tone PBO (HD8961, supplier HD MicroSystems).
A set of preliminary experiments in terms of the wavelength, exposure dose, bake, cure and development were performed before the most suitable process conditions were set to achieve the lower resolutions and the low film thickness, summarized in Tab. 1.

Table 1. Process Recipe for Coating, Exposure and Development

Process	Parameters
Spin Coating	EVG®120 Advanced Resist Processing System 1500 rpm / 300 rpm/s / 60s
Soft Bake	110 °C / 180s (FT 12.5µm) for HD8961
Exposure	405nm: Dual layer dose 65/200 $mJcm^{-2}$
Development	4 standard puddles (30s) with intermediate cleaning Aqueous Developer AZ 726 MIF FT 10.5 µm
Cure	EVG®105 System N_2 250 °C 2h (1h 125 °C; 1h ramp, 2h 250 °C) FT 9.0 µm

Since the target application of this dual–exposure process is the advanced packaging, the layout was chosen with compact lines (RDL) and openings (VIA) to simulate the FO WLP design as shown in Fig. 1.

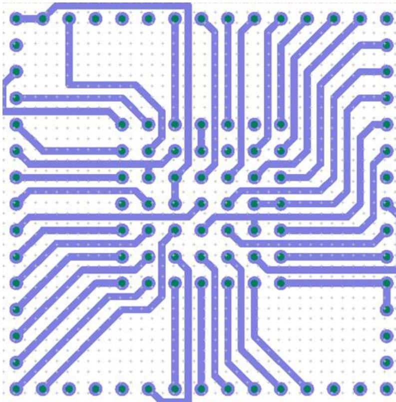


Figure 1: Patterning design illustrating the compact lines and openings

The maskless exposure lithography offers unique capability of efficient process optimization. As illustrated in Fig. 2 on one single wafer variety of the exposure parameters tested without frequent mask or reticles changes in stepper/scanner exposure technologies.

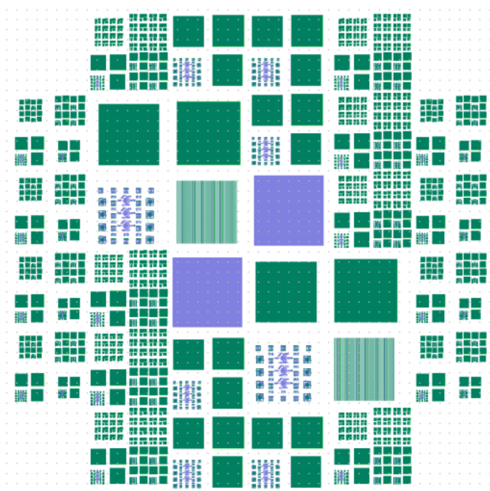


Fig. 2.: LITHOSCALE® test layout containing various test fields scaled up/down to different sizes written on one wafer

3. RESULTS

3.1 Layer Thickness Uniformity

The film thickness of the dielectrics was measured after development and post exposure cure processes. In sum, 41 measuring points were taken for the statistical calculations including histograms of the FT distribution. For the layer thickness uniformity, the spectral reflectance method is used utilizing a wavelength range of 200nm up to 1700nm. The spot size up to 1.5 mm and the measurement accuracy can reach values down to 0.2 nm. Spectral reflectance can be utilized for a broad range of planar film thicknesses and materials including dielectric films.
The results of the FT measurements are after spin–coating and after post exposure bake are summarized in Tab. 2.

Table 2: Dielectric film thickness and nonuniformity data after spin–coating and post exposure bake

HD8961	**Spin–coating**	**Post Exposure Cure**
FT Min [µm]	12.160	7.1759
FT Max [µm]	12.420	7.9710
FT Mean [µm]	12.299	7.7436
FT Std. Dev.	0.05811	0.17771
Nonuniformity [%]	±1.1	±5.1

Analyzing the histograms of film thickness distribution (film thickness distribution vs. measurement point counts) in Fig. 3, it can be stated for HD8961 that the novel dielectric materials perform with normal Gauss distribution on the spin–coating process, i.e., the novel dielectric materials is optimized for the advanced packaging application and fulfill the quality requirements for the dielectrics in FO WLP processes.

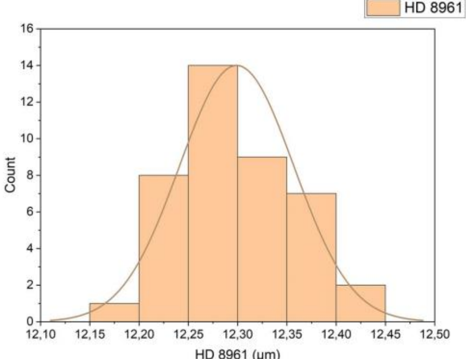


Figure 3. Film thickness distribution HD8961 after spin–coating on EVG®120

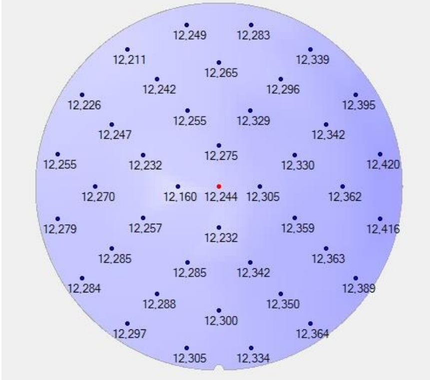


Figure 4. HD8961 coating thickness after spin coating

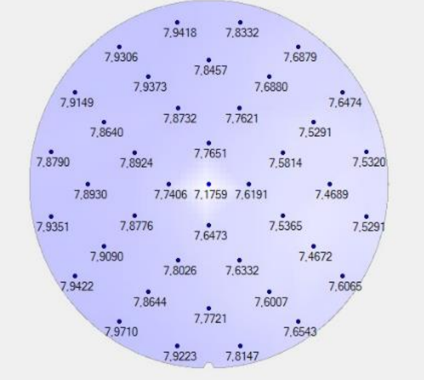


Figure 5. HD8961 coating thickness after development and post exposure cure

An automated puddle development process has been set up. The PBO material is partially removed (even unexposed areas) during the development leading to a reduced layer thickness of 6.6 µm and an increased nonuniformity of 5.1%, illustrated in Fig. 4. The development process needs to be further optimized to improve the uniformity in the central area of the wafer. Due to this issue structures in the central area have been excluded from further evaluation. The cure further reduces the layer thickness down to 7.7 µm but has no impact on the uniformity, illustrated in Fig. 5.

3.2 Microscopic Investigation of the Dual–Layer Structures

The optical microscope inspection has been performed after the final process step (cure). The resolution for VIA type structures has been evaluated. The smallest resolved diameter of inner VIA was 3.8 µm, as shown in Fig. 6. The SEM inspection has been performed after the final process step (cure). The shape of the two step structures has been inspected using tilted top view as well as cross–section images.

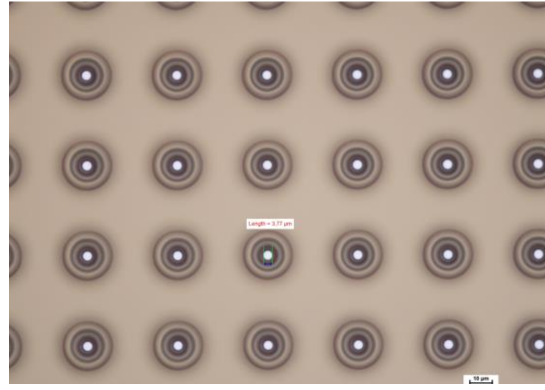


Figure 6. HD8961 via and via pattern with best resolution: inner via = 3.8 µm

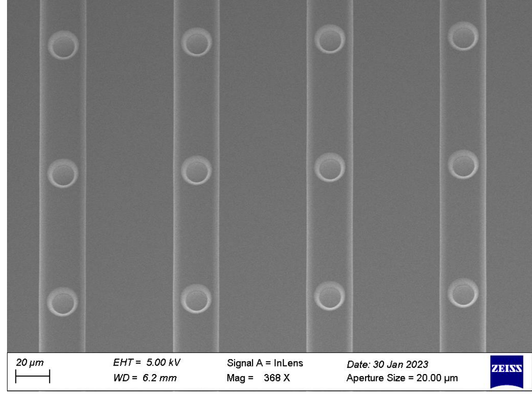


Figure 7: HD8961 SEM tilted top view: via within pattern line

Figure 8: SEM Image: cross section of line within line pattern

The optical microscopy images with high magnification for HD8961 (Fig. 6) show the same patterns as digital file design entered in tool's layout design illustrated in Fig. 2.

Fig. 7 demonstrates the regular profiles and via openings in the RDL lines. Looking at the profiles (Fig. 8) the dielectric show straight side walls in the first layers, nevertheless, due to the physical nature of the positive tone PBO dielectric material (high shrinkage), the side walls of the second layer became concave (rounded inward) after the development process. Besides the shrinkage of the polymer during the curing as a material characteristic, the second parameter influencing the shrinkage is the development process, i.e., development time.

In the further investigation of dual–layer exposure concept of the dielectric materials for FO WLP, an influence of the development parameters is going to be study in details.

3.3 Stylus Profiler Method

Film thickness measurement via stylus profiler utilizes a contact mode where a stylus moves over the surface of a substrate and measures the substrate topography / surface profile. Therefore, in planar films a step e.g. scratch is required. The standard stylus diameter is 2µm. The measurement resolution in z-direction (height) is in the range of 0.1nm – 16nm depending on the overall z–height of the measured topography. The step height repeatability is <0.16nm 1sigma. The stylus profiler method also works for any sort of topography up to 1mm z–height including films with high roughness and absorbing materials like color/black resists.

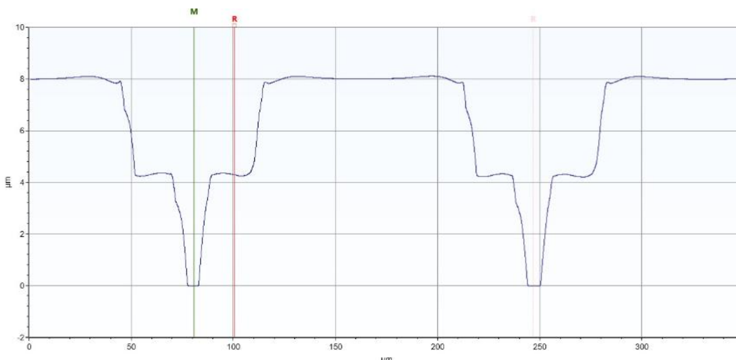


Figure 9: Stylus profile measurement: line in line 2 step structure

Stylus profiler measurement has been performed on 6 identical structures across each wafer (one wafer for each material). The height of the middle step (layer thickness) has been measured for each position. For the dielectric material HD8961, the measured middle step height (first layer thickness) is 3.98 µm with a uniformity value 8.54%. The results clearly demonstrate that the dual–layer exposure concept, besides low feature size and Gaussian–type CD distribution, provides 50% thickness of the first layer in the dual–exposed structures.

4. CONCLUSIONS

The dual–layer exposure concept supports the industry efforts to develop the cost–efficient POR solutions in the interconnect formation in FO WLP.
The capability to pattern RDL traces and vias in dual–layer exposure, achieving more efficient and streamlined lithographic processing was demonstrated in the present paper.
The ability to achieve sub–4 µm resolution in the via structures is significant as it allows for the fabrication of smaller and more precise interconnects.
This can lead to improved performance, increased packaging density, and enhanced functionality of semiconductor devices.
From general perspective, MLE technology helps to overcome the limitation in lithographic patterning associated with the stepper technology.
Besides in advanced packaging and heterogenous integration, MLE is versatile technology finding applications in photonics, MEMS, next–generation advanced packaging, and emerging, for instance medical markets.
To prove the suitability of this concept in the interconnect formation in advanced packaging, the full integration process PVD, ECD and CMP will be conducted on the dual–exposed structures on LITHOSCALE®.

REFERENCES

[1] B. Matuskova et al.: Maskless Lithography Addresses Shift Toward 3D Integration, White Paper, EV Group, September 2020.
[2] B. Dielacher at all: Digital Lithography for Advanced Packaging and Heterogenous Integration, 2022 IEEE 24th EPTC, December 03rd – December 06th 2022, Singapore.
[3] T. Enomoto et al: Advanced Dielectric Materials (Polyimides and Polybenzoxazoles) for Fan–Out Wafer–Level Packaging (FO–WLP), Advances in Embedded and Fan–Out Wafer–Level Packaging Technologies, Wiley IEEE Press, ISBN 9781119314134.

New wave front phase sensor used for 3D shape measurements of patterned silicon wafers

Kiril Ivanov Kurtev[a], Juan M. Trujillo-Sevilla[a], Miguel Jiméneza[a], Rubén Abrantea[a], Guillermo Castro Luis[a], Jan O. Gaudestad[b]

[a]Wooptix SL Av. Trinidad nº61, 7º, 38204 La Laguna, Tenerife Canary Islands, Spain; [b]Wooptix SL, San Francisco, California, USA

ABSTRACT

On product overlay (OPO) is one of the most critical parameters for the continued scaling according to Moore's law. Without good overlay between the mask and the silicon wafer inside the lithography tool, yield will suffer[1]. As the OPO budget shrinks, non-lithography process induced stress causing in-plane distortions (IPD) becomes a more dominant contributor to the shrinking overlay budget[2]. To estimate the process induced in-plane wafer distortion after cucking the wafer onto the scanner board, a high-resolution measurement of the freeform wafer shape of the unclamped wafer, with the gravity effect removed, is needed. A high-resolution wafer shape map using a feed-forward prediction algorithm, as has been published by ASML, can account for both intra and inter die wafer distortions, minimizing the need for alignment marks on the die and wafer in addition to that it can be performed at any lithography layer[3]. Up until now, the semiconductor industry has been using Coherent Gradient Sensing (CGS) interferometry or Fizeau interferometry to generate the wave front phase from the reflecting wafer surface to measure the free form wafer shape. However, these techniques have only been available for 300mm wafers[3,4,5].

In this paper we introduce Wave Front Phase Imaging (WFPI), a new technique that can measure the free form wafer shape of a patterned silicon wafer using only the intensity of the reflected light. In the WFPI system, the wafer is held vertically to avoid the effects of gravity during measurements. The wave front phase is then measured by acquiring only the 2-dimensional intensity distribution of the reflected non-coherent light at two or more distances along the optical path using a standard, low noise, CMOS sensor. This method allows for very high data acquisition speed, equal to the camera's shutter time, and a high number of data points with the same number of pixels as available in the digital imaging sensor. In the measurements presented in this paper, we acquired 7.3 million data points on a full 200mm patterned silicon wafer with a lateral resolution of 65µm. The same system presented can also acquire data on a 300mm silicon wafer in which case 16.3 million data points with the same 65µm spatial resolution were collected.

Keywords: Patterned wafer geometry, nanotopography, metrology, semiconductor manufacturing, wave front phase imaging, wafer shape, wafer stress

1. INTRODUCTION

One approach to predict the process-induced in-plane displacement (IPD) is to use free-form wafer shape measurements without the effects of gravity pulling on the silicon wafer. In this case, an accurate measurement of the free-form shape of the unclamped wafer is used as the input to a model to estimate the in-plane wafer distortion after clamping[3].

1.1 Fizeau Interferometry

The current state of the art wafer shape system used for inline high-volume semiconductor metrology, both blank and patterned wafers, is using a dual-side Fizeau interferometry setup that acquires data on the silicon wafer from both sides. Light from a tunable laser diode with $\lambda \approx 632$nm is guided through an optical fiber to each of the two interferometer channels. In each channel, a polarization beam splitter directs the beam through a quarter-wave plate ($\lambda/4$) aligned at 45° compared to the polarization direction of the polarizing beam splitter. The light beam then passes through the collimator lens where the light beam is collimated. The collimated beam then propagates to the reference flat, which acts as the reference surface of the interferometer. At the reference flat, part of the light is reflected, and another part is transmitted. The transmitted collimated beam then hits the silicon wafer where it is reflected to the reference flat where it is then transmitted and mixing with the reference beam to form interference pattern. The combined light beam then passes through the collimator lens, the quarter-wave plate ($\lambda/4$), and then through the beam splitter to be imaged at the digital image sensor where the interference patterns are detected. The wavelength of the laser is varied to create a linear phase shift in the

38th European Mask and Lithography Conference (EMLC 2023), edited by
Uwe F.W. Behringer, Jo Finders, Proc. of SPIE Vol. 12802, 128020B
© 2023 SPIE · 0277-786X · doi: 10.1117/12.2675624

interferograms. Several camera frames with phase-shifted fringe patterns are acquired and processed in the computer, where phase algorithms are employed to convert each camera pixel from the fringe-phase to surface height[6].

Fizeau interferometry wafer metrology systems reconstruct the wafer surface by calculating the wrapped surface phase maps from a sequence of measured interferometry image frames while modulating the wavelength (frequency) using a tunable laser and then unwrapping these phase maps. The wave front phase is proportional to the surface slope [6,7,8]. The large wafer shape of up to >10μm can lead to slopes exceeding 500nm/mm, which is a slope of 0.028°, while the measurement requirements for flatness are in the nanometer range, which represents a huge dynamic range for the interferometric measurements. Interferometry tools require special tool design to minimize the systematic measurements errors for high slope condition. Lastly, to keep measurements errors to a minimum, the reference flats are placed very close together, generally at the order of around 100mm apart[6].

Phase signals of a patterned wafer surface obtained using a Fizeau interferometry wafer metrology tool, contains sharp phase transitions induced by the surface pattern structures causing large errors and failures when using a general two-dimensional phase unwrapping method. These sharp phase transitions will lead to severe artifacts on a reconstructed surface map of a patterned silicon wafer. To overcome this, a method has been developed where the phase of the front surface is being subtracted from the phase of the back surface to flatten large phase shape components[9]. The backside of 300mm silicon wafers is polished, however 200mm silicon wafers are not generally polished[10]. The subtracted phase map is first smoothed before being unwrapped and used to represent the shape of the front, patterned, side of the silicon wafer[9]. Such a system acquires wafer shape maps with spatial resolution of 150μm-500μm[11].

1.2 Wave Front Phase Imaging (WFPI)

The newly invented phase measuring technique Wave Front Phase Imaging (WFPI) is based on registering the intensity distribution of the same field of view at two different optical planes of the reflected light from the silicon wafer. This can be done by either having a single camera on a linear translation stage and moving the camera a short distance to acquire both intensity images, or by splitting the light and guiding the rays onto two different cameras located at different distances away from the silicon wafer being imaged. The wave front phase is defined as the surface perpendicular to the direction of propagation of the light rays. The image sensor assumes geometrical propagation of light, and in this regime the light can be considered as a collection of light rays which bends according to Snell's law and reflects on a silicon wafer surface keeping its angle with respect to the surface normal[12].

Since WFPI is only utilizing intensity images, no interference is needed nor wanted. In fact, interference should be avoided to prevent intensity changes in the light rays reaching the images sensor that are not due to the surface topography. For this reason, WFPI is using a non-coherent Light Emitting Diode (LED), in this case with a wavelength of $\lambda = 590$nm, that generates a red-orange light beam with an undisturbed phase ($\varphi = 0$). The light beam with $\varphi = 0$ will then pass through the main lens, which is larger than the silicon wafer to be measured, and then onto the silicon wafer. The light will be reflected off the silicon wafer surface and carry a phase change ($\varphi \neq 0$) that is attributable to the surface topography. The reflected light beam with $\varphi \neq 0$ will the pass through the main lens and then onto the image sensor, or two image sensors if a dual camera sensor is used, to acquire the two intensity images at two different distances away from the silicon wafer[13].

A special algorithm is used to calculate the phase change based on these two intensity images. The surface phase map is proportional to the relative surface height map[13]. A WFPI system acquires wafer shape maps with spatial resolution of 65μm with an optical limitation of 3.2μm if higher pixel image sensor arrays are added[14]. To test for max slope the system can handle, a very flat, high quality reference wafer is rotated until vignetting, a reduction of the intensity on the periphery of the wafer, starts to show in the intensity images. The max slope has been found to be more than ±0.25° which equals at least a 4,363nm/mm slope (4.36μm/mm).

1.3 Patterned Wafer Geometry Challenges

The Transport Intensity Equation (TIE), $\frac{\partial I}{\partial z} = -\frac{\lambda}{2\pi}\left(\vec{\nabla} I \cdot \vec{\nabla}\emptyset + I \cdot \nabla^2 \emptyset\right)$, where I is intensity, z the propagation distance, λ is wavelength and ϕ is the phase[15], is a computational approach to reconstruct the phase of a complex wave in an optical system, describing the relationship between the intensity and phase distribution[16]. When imaging the reflected light from a blank silicon wafer, one can assume the intensity distribution is constant and hence the term $\vec{\nabla} I \cdot \vec{\nabla}\emptyset = 0$ the term solved by Shack-Hartman[17]. However, for a patterned wafer with highly reflected copper metal lines separated by a highly light absorbing space filled with dielectric material, generally glass-like optical features, one must assume that $\vec{\nabla} I \cdot \vec{\nabla}\emptyset \neq 0$ and the full TIE needs to be solved. Due to the large intensity difference in the reflected light from a patterned silicon wafer,

interferometry techniques generally use a phase difference map calculated based on the smoothed front and back phase differences to improve the accuracy of phase unwrapping[9]. The second term of the TIE, $I \cdot \nabla^2 \emptyset$, is the part that is attempted to be solved in the classic TIE inversion methods[16]. The novel approach of WFPI consists of solving this equation entirely, using a proprietary numerical method.

1.4 WFPI System Setup

Two images, I_1 and I_2, are captured around a conjugated plane of the reflective sample, in this case a silicon wafer, with symmetric positions (in front of and behind the reflective sample) with a total distance between the images equal to z (see Fig. 1). A collimated light beam, with an assumed flat phase map, is generated from an incoherent light source using a light-emitting diode (LED). In principle, any type of light can be used, but the use of incoherent light has the advantage of not producing speckle, which worsens the performance of the WFPI. The light beam passes through lens L_3 and is then reflected on the beam splitter changing direction 90°, and then passing through lens L_2 and then lastly lens L_1 to generate the collimated light beam with the same size, or larger, than the sample being measured, to hit the silicon wafer. The collimated light is then reflected from the reflective sample, which modifies the shape of the collimated beam proportional to the optical behavior of the reflective sample surface, thus generating a distorted outgoing phase map proportional to the wafer surface. The imaging plane P of the reflected light is then being translated into a conjugated plane, P′, after the distorted phase map passes through lens L_1 and lens L_2 and then through the beam splitter without changing direction, to then being recorded by one or more digital image sensors which acquires the two required intensity images, I_1 and I_2. These intensity images are acquired around the conjugated plane P′ in equidistant planes along the z-axis. Choosing focal lengths of lens L_1 and lens L_2 appropriately allows for adjusting the sample size to the image sensor size[14].

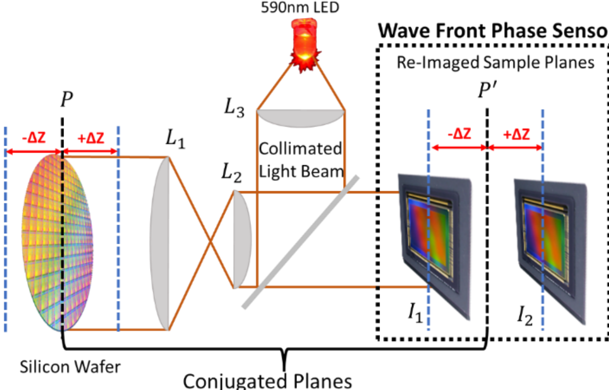


Figure 1. Basic optical setup of the wavefront phase imaging (WFPI) technique for reflective samples.

2. THE WFPI LAB SYSTEM

The newly developed WFPI system is using a 310 mm diameter clear aperture lens to allow for the entire 300mm silicon wafer to be measured in a single image shot. The larger clear aperture diameter of the lens allows for additional room for wafer placement within several mm while still generating the same quality of data (See Fig. 2a). The wafer is held vertically in a specially made wafer-holder-ring exposing the wafer from both back and front side, where the wafer rests on two metal studs while being balanced vertically using the top-knob. The three points keeping the wafer vertically is covering less than a millimeter of the edge of the wafer, meaning only the wafer edge exclusion zone will miss some data points at these locations while the silicon wafer where chip production takes place is fully covered in the wafer geometry data (See Fig. 2b).

A digital camera, using thermoelectric cooling to improve noise performance, is attached to a linear translation stage that acquires two images at two different locations along the optical path using a 1 second shutter speed and approximately 12 seconds to move the camera making it a total of 14 seconds to acquire the full data set. The intensity data is then processed by a regular PC, which takes around the same amount of time to process the data sets with the current number of data points acquired, to generate the wave front phase and the surface slope. To remove some of the effects of air turbulence between the lens and the wafer, the camera averaging time was set to 1 second, while under normal operations in a more controlled environment, the shutter speed would have been set to 0.1 second or less, which would be the time a dual camera system acquires all the data.

While the camera has a 62 mega pixel imaging sensor (9576×6388), only pixels within the circular wafer projection are utilized for data generation. In addition, extra pixels outside of the wafer projection allow for freedom with wafer placement location and to compensate for large bow and other shape features that can guide the light rays outside of the wafer projection, hence slightly more pixels are captured for the wave front phase calculation. The reflection coming from the 200mm wafer equaled just over 7.3 million pixels in the imaging sensor with a pixel size of 65µm encompassing the full wafer geometry data set. Since the wave front phase is generated from the intensity of a collimated light beam, the distance between the main 310mm lens and the silicon wafer can be anywhere within a range of ±1.6mm without degrading the resolution. The system can also handle vibrations of up to a half pixel size, around 32µm; an amplitude of vibrations that is easily reached with a simple breadboard, passive vibration isolation setup.

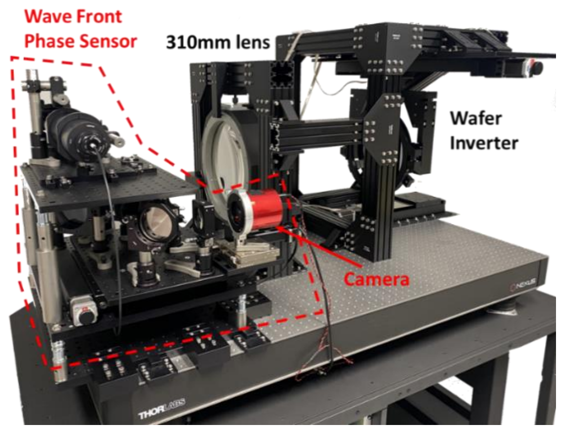


Figure 2a. Picture of the current lab system with a 310mm lens and the wafer on the sample plate.

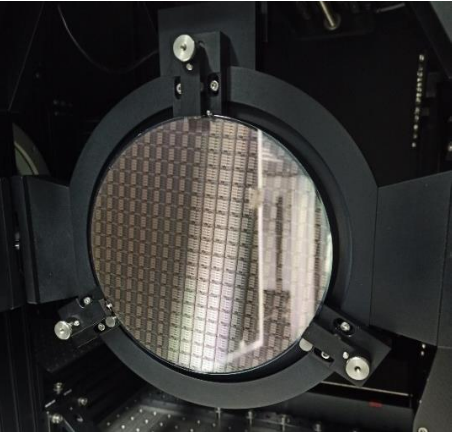


Figure 2b. Shows the wafer holder for the 200mm patterned wafer resting on 3 steel balls and the location of each steel ball.

The pixel size of 65 µm is the same whether measured on a 200mm or 300mm silicon wafer. The resolution could be improved by increasing the pixels in the imaging sensor up to the optical limit of the lens, which limit is 3.2µm; however, at such small pixel size a camera with more than 1.7 billion pixels would be required. There are image sensors commercially available at 151 million pixels, which would generate just over 80 million data points on a 300mm silicon wafer at no extra data acquisition time.

Since the system can measure surfaces with a slope of up to 0.25°, data is collected on the silicon wafer all the way out to the edge of the wafer to within a single pixel (in this case 65 µm) away from where the edge roll-off exceeds 0.25°. The Semi standards allow for a wide variation in wafer edge profiles manufactured by silicon suppliers [18,19] making it hard to set the exact exclusion zone for where the 0.25° angle limit hits, however the default edge roll off measurement point set by Semi standards is 1mm away from the wafer edge[19].

3. WFPI APPLIED ON A 200MM PATTERNED SILICON WAFER

A 200mm patterned silicon wafer built using a BCD (Bipolar – CMOS – DMOS) process, an advanced family of silicon processes, was used for this experiment. BCD is a key technology for power ICs (Integrated Circuits) that combines the strengths of three different process technologies onto a single chip: Bipolar for precise analog functions, CMOS (Complementary Metal Oxide Semiconductor) for digital design and DMOS (Double Diffused Metal Oxide Semiconductor) for power and high voltage elements. This specific BCD process was using an 0.35µm technology with 3 metal layers. The top metal layer (M3) has 2.2µm wide copper metal lines with 1.8µm spacing between the metal lines. The middle copper metal layer (M2) had 0.7µm wide metal lines with 0.8µm spacing and the bottom copper metal layer connecting directly to the transistors (M1) had 0.5µm wide metal lines with 0.6µm spacing between the metal lines. There was a passivation layer covering the top metal layer (M3). The patterned silicon wafer, built by a US based foundry, was provided by the fabless semiconductor company ElevATE, a leading supplier of low power, high density components for the design of Automated Test Equipment (ATE) used by the semiconductor industry[12].

The challenge of having a patterned wafer with highly reflective copper metal lines exposed at the top is the very large differences between high intensity reflected light and low intensity or dark areas in between the reflective copper lines.

The full frame intensity image of the 200mm silicon wafer is shown in Fig. 3a with a zoom in intensity image shown in Fig. 3b. Fig. 3a shows the large freedom of where to place the silicon wafer without having any effect on the shape data, making it ideal for high-speed wafer fab automation applications.

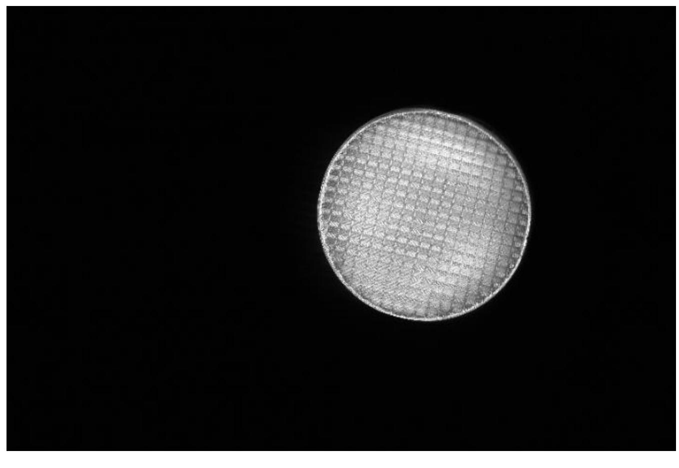


Figure 3a. Full frame intensity image of the 200mm silicon wafer.

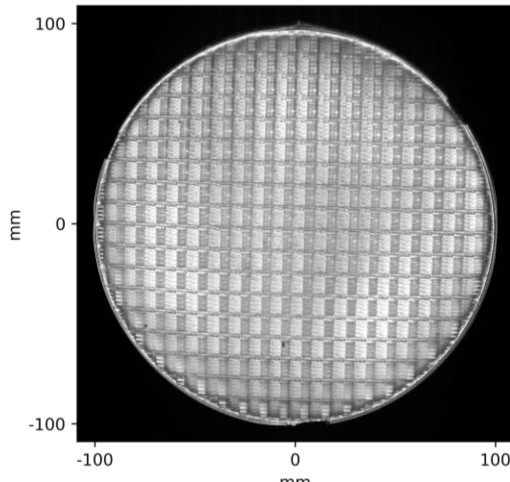


Figure 3b. Zoom in intensity image of the 200mm silicon wafer.

The wafer was manually placed into the vertical wafer holder and then the wafer shape data was acquired from the front side of the silicon wafer (Fig. 3b). It was not possible to acquire data on the backside of the wafer due to its very poor reflectivity. A silicon wafer on the back side tends to be very non-reflective after the furnace step in the fab. The acquired free form wafer shape (Fig. 4a) of the frontside of the patterned wafer was then generated from the intensity images acquired of the silicon wafer (Fig. 3a and Fig. 3b). Due to lack of high-quality data along the edges, 1.3mm of the edge was removed, making the diameter of the shape data 197.4mm (radius of 98.7mm). The notch is located at (0, -100) in all data sets presented (Fig. 4a, Fig. 4b).

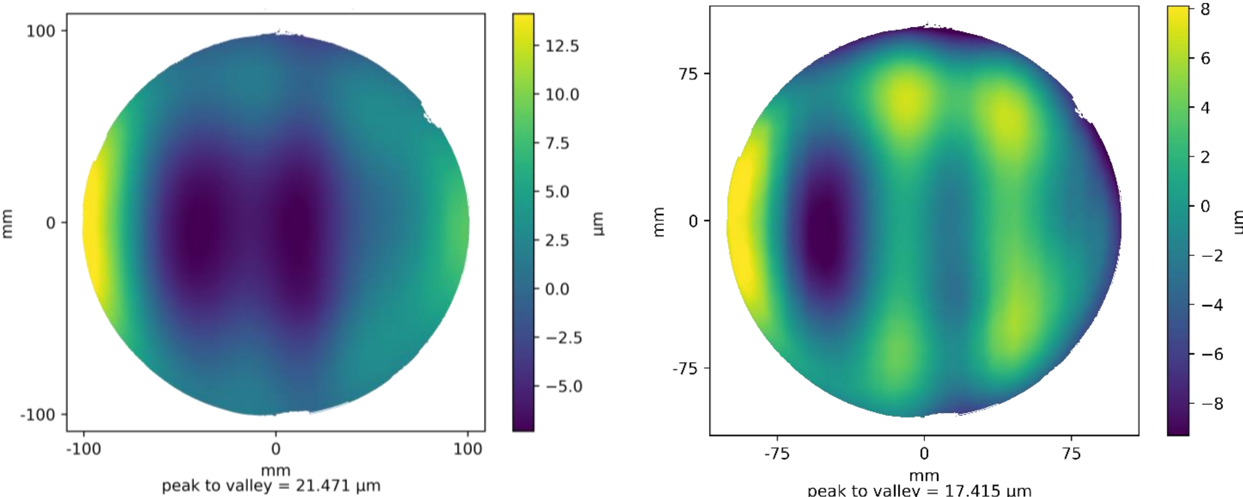


Figure 4a. Wafer shape measurement of the front side of the 200mm silicon wafer.

Figure 4b. Wafer shape after removing 2nd order features

The free form wafer shape data was found to have PV = 21.471µm when removing all data points below 1% and above 99% (Fig. 4a). Uniform stress in a silicon wafer can easily be corrected for in the scanner by removing the 2nd order shape from the global shape data (Fig. 4b). PV after removing 2nd order, was 17.415µm (using data withing 1% - 99%),

which is not much improvement from the global shape indicating that most of the shape in this wafer is not coming from uniform stress but rather from higher order wafer shapes.

ASML has implemented a Higher Order Wafer Alignment model up to the third order (HOWA3) into their scanner to correct for process induced wafer shape, however it cannot correct for higher order shape distortions (above 3rd order). One can always add many more alignment marks on the die; however, this takes space in the scribe lines. Also, no alignment marks can be placed inside the die hence intra-die stress cannot be corrected for. With a very high map of the free wafer shape, the wafer stress can be predicted, and a correction can be made using a feed forward algorithm in the lithographic scanner without adding alignment marks[3].

An attempt was also made to image the backside, however due to the low-reflective backside, no data was able to be acquired. As the backside of the wafer had a blueish color, it is expected that using a shorter wavelength for the shape measurement, such as a blue LED, would make it possible to also acquire shape measurement of the backside, however this is yet to be tested.

The data was then further analyzed to calculate certain important flatness values as stated by the Semi standards MF1390[20] and MF1530[21]. Bow and warp may be caused by unequal stresses on the two exposed surfaces (front side and backside) of the wafer, however, it cannot be determined from measurements on a single exposed surface. However, since a backside image was not possible to acquire on the silicon wafer due to poor backside reflectivity, the median surface was not possible to generate and therefore the frontside was used as a representative of the median surface. Bow is calculated in two ways; bow best fit (Bow_{BF}), which is the algebraic difference between the corrected median surface at the center of the wafer to a reference plane, and bow 3-points (Bow_{3P}), which is the algebraic difference between the corrected median surface at the center of the wafer to a 3-point reference plane out to 6mm away from the wafer edge. Warp, or warp best fit ($Warp = Warp_{BF}$), is a measurement of the differentiation between the max and min of the entire median surface of a wafer and a reference plane, not just the center point[20]. It is important to note that PV values are calculated by removing data below 1% and above 99%.

Table 1. Bow and warp[20] are listed in the table below using the mean calculations.

Parameter	Values
Bow_{BF}	-6.72 µm
Bow_{3P}	-12.57 µm
Warp	23.55 µm

The next step is to reveal nanotopography. Nanotopography is the non-planar deviation of the whole front wafer surface within a spatial wavelength range of approximately 0.2 to 20 mm and within a fixed quality area (FQA) specified area, which in our case is the entire wafer less the 3.9mm edge exclusion zone. Typical examples of nanotopography include dips, bumps or waves on the wafer surface that vary in PV height from a few nanometers to a several hundred nanometers[22]. The first step in removing the long spatial wavelengths on the wafer is by using a high-pass filter, a double Gaussian filter according to Semi standards, effectively creating a global reference surface with PV values from the entire surface[23].

The wafer shape data, showing the longer lateral wavelength features (Fig. 3b), was then filtered using a double Gaussian with an 800µm lateral cutoff wavelength to reveal the nanotopography for the entire 200mm silicon wafer (Fig. 5). A 99.7% data sensitivity threshold was added and the nanotopography to remove outliers.

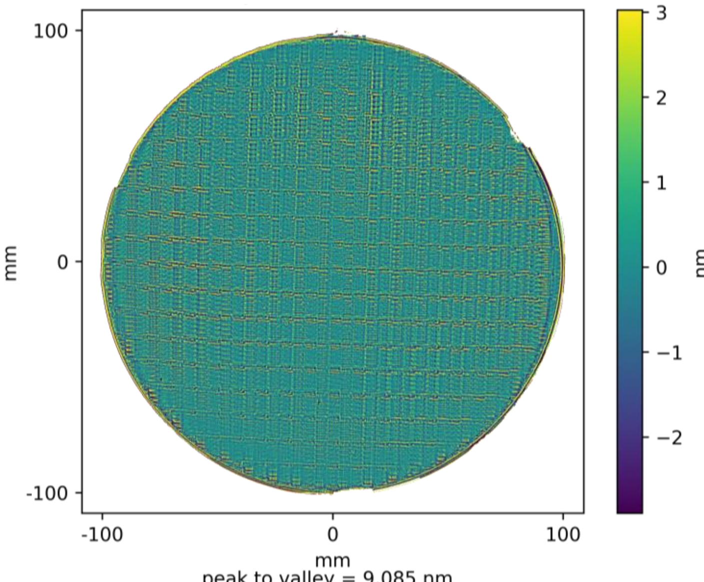


Figure 5. Nanotopography of the full wafer using 800µm lateral cutoff wavelength using a double Gaussian filter.

The PV of the Naotopography across the entire wafer was found to be 9.085nm. The nanotopography reveals the top surface features caused by the copper metal lines. Since this wafer was a finished wafer with very large metal lines exposed at the top, the nanotopography data indicates this wafer is good enough for complying with the overlay budget required for such large metal lines.

4. CONCLUSION - DISCUSSION

Understanding the fundamental relationship between localized wafer shape changes and EUV lithography overlay error is crucial in solving defocus problems, particularly in advanced nodes where feature sizes are smaller. In the present work, we show a high-speed wafer geometry system that can capture all necessary topography information of the entire 200mm patterned silicon wafer.

One requirement to measure stress is a free shape wafer measurement before and after the wafer process step, hence stress measurements is a measurement of the change in the stress incurred during a specific process or processes. As it has been difficult to get access to a wafers pre-process and then again post-process, no process induced stress measurements have been performed using WFPI, however the key to understanding stress measurements is to first measure the free shape wafer geometry and then use mathematical equations such as a finite element (FE) modeling technique[5], to generate the IPD fingerprints to predict the overlay error.

WFPI has several favorable features, such as data acquisition speed, high slope sensitivity, as much as 10× higher slope than what has been published when using Fizeau interferometry[6], and a very high number of data points. Adding that WFPI can handle vibrations very well, in the 10s of microns, WFPI is a perfect candidate for high volume manufacturing. Current work is ongoing to integrate the WFPI sensor into a fully automatic fab system in a dual camera setup with either a single or a dual lens imaging system to satisfy the needs of semiconductor fabs for a very high wafer throughput.

REFERENCES

[1] "INTERNATIONAL ROADMAP FOR DEVICES AND SYSTEMS™", 2022 EDITION on METROLOGY

[2] Leon van Dijk, Jeffrey Mileham, Ilja Malakhovsky, David Laidler, Harold Dekkers, Sven Van Elshocht, Doug Anberg, David M. Owen, Richard van Haren, "Wafer-shape based in-plane distortion predictions using superfast 4G metrology," Proc. SPIE 10145, Metrology, Inspection, and Process Control for Microlithography XXXI, 101452L (28 March 2017)

[3] Richard van Haren, Ronald Otten, Subodh Singh, Amandev Singh, Leon van Dijk, David Owen, Doug Anberg, Jeffrey Mileham, Yajun Gu, Jan Hermans, "Intra-field stress impact on global wafer deformation," Proc. SPIE 10959, Metrology, Inspection, and Process Control for Microlithography XXXIII, 109591I (26 March 2019)
[4] T. Brunner, V. Menon, C. Wong, N. Felix, M. Pike, O. Gluschenkov, M. Belyansky, P. Vukkadala, S. Veeraraghavan, S. Klein, C. H. Hoo, J. Sinha, "Characterization and mitigation of overlay error on silicon wafers with nonuniform stress," Proc. SPIE 9052, Optical Microlithography XXVII, 90520U (31 March 2014)
[5] Timothy A. Brunner, Vinayan C. Menon, Cheuk Wun Wong, Oleg Gluschenkov, Michael P. Belyansky, Nelson M. Felix, Christopher P. Ausschnitt, Pradeep Vukkadala, Sathish Veeraraghavan, Jaydeep K. Sinha, "Characterization of wafer geometry and overlay error on silicon wafers with nonuniform stress", J. Micro/Nanolith. MEMS MOEMS 12(4), 043002 (Oct–Dec 2013)
[6] Klaus Freischlad, Shouhong Tang, Jim Grenfell, "Interferometry for wafer dimensional metrology," Proc. SPIE 6672, Advanced Characterization Techniques for Optics, Semiconductors, and Nanotechnologies III, 667202 (10 September 2007); doi: 10.1117/12.732546
[7] US6847458B2 patent "Method and apparatus for measuring the shape and thickness variation of polished opaque plates" 2003
[8] US9074873B1 patent "Measurement of thickness variation and shape of wafers" 2013
[9] US9632038B2 patent "Hybrid phase unwrapping systems and methods for patterned wafer measurement", 2015
[10] SEMI M1 "SPECIFICATION FOR POLISHED SINGLE CRYSTAL SILICON WAFERS"
[11] US20160071260A1 patent "Enhanced Patterned Wafer Geometry Measurements Based Design Improvements for Optimal Integrated Chip Fabrication Performance" 2014
[12] Juan M. Trujillo-Sevilla, Óscar Casanova-González, Alex Roqué-Velasco, Miguel Jesús Sicilia, Javier González Pardo, José Manuel Ramos-Rodríguez, Jan O. Gaudestad, "New wave front phase sensor used for 3D shape measurements of patterned silicon wafers," Proc. SPIE 12292, International Conference on Extreme Ultraviolet Lithography 2022, 122920P (1 December 2022); doi: 10.1117/12.2642287
[13] Juan M. Trujillo-Sevilla, Álvaro Pérez-García, Óscar Casanova-González, Miriam Velasco-Ocaña, Sabato Ceruso, Ricardo Oliva-García, Óscar Gómez-Cárdenes, Javier Martín-Hernández, Alex Roqué-Velasco, José Manuel Rodríguez-Ramos, Jan O. Gaudestad, "New high repeatability wafer geometry measurement technique for full 200mm and 300mm blank wafers," Proc. SPIE 12008, Photonic Instrumentation Engineering IX, 120080A (5 March 2022)
[14] Juan M. Trujillo-Sevilla, Óscar Casanova-González, Alex Roqué-Velasco, Javier González Pardo, Jose Manuel Ramos-Rodríguez, Jan O. Gaudestad, "New wafer shape measurement technique for 300mm blank vertically held silicon wafer," Proc. SPIE 12428, Photonic Instrumentation Engineering X,124280U (8 March 2023); doi: 10.1117/12.2651981
[15] Juan M. Trujillo-Sevilla, Alex Roqué-Velasco, Miguel Jesús Sicilia, Óscar Casanova-González, Jose Manuel Ramos-Rodríguez, Jan O. Gaudestad, "Wave front phase imaging for silicon wafer metrology", Proc. SPIE Optics + Photonics 1221605 Novel Optical Systems, Methods, and Applications XXV (August 22, 2022).
[16] Transport Intensity Equation (TIE) explained: https://en.wikipedia.org/wiki/Transport-of-intensity_equation
[17] Schack-Hartman wavefront sensor from ThorLabs: https://www.thorlabs.com/newgrouppage9.cfm?objectgroup_id=5287
[18] SEMI M73 "TEST METHOD FOR EXTRACTING RELEVANT CHARACTERISTICS FROM MEASURED WAFER EDGE PROFILES"
[19] SEMI M77 "TEST METHOD FOR DETERMINING WAFER NEAR-EDGE GEOMETRY USING ROLL-OFF AMOUNT, ROA"
[20] SEMI MF1390-0218 TEST METHOD FOR MEASURING BOW AND WARP ON SILICON WAFERS BY AUTOMATED NONCONTACT SCANNING
[21] SEMI MF1530-0707 (Reapproved 1018)"TEST METHOD FOR MEASURING FLATNESS, THICKNESS, AND TOTAL THICKNESS VARIATION ON SILICON WAFERS BY AUTOMATED NONCONTACT SCANNING"
[22] SEMI M43-0418 GUIDE FOR REPORTING WAFER NANOTOPOGRAPHY
[23] SEMI M78-0618 GUIDE FOR DETERMINING NANOTOPOGRAPHY OF UNPATTERNED SILICON WAFERS FOR THE 130 nm to 22 nm GENERATIONS IN HIGH VOLUME MANUFACTURING

High-Performance Computing Architecture (HW/SW) for Mask CDSEM Design Based Metrology

Tatsuro Okawa[1], Yusuke Kakinuma[1], Yoshiaki Ogiso[1], Naoyuki Tanaka[1], Kazuo Mukawa[1], Soichi Shida[1], Shinichi Kojima[2], Toshimichi Iwai[1]
[1]Advantest Corporation Saitama R&D Center, 1-5 Shin-tone, Kazo-shi, Saitama, Japan
[2]Advantest America Inc., 3061 Zanker Rd., San Jose, CA 95134, USA

ABSTRACT

We have developed novel Design Based Metrology (DBM) technology which enables metrology engineers to utilize not only 2-dimensional metrology with high precision but also high number CPU cores to reduce calculation time. It is crucial to maximize efficiency of parallel processing with high number of CPU cores and reduce overhead. We designed new DBM software based on the concepts of our novel DBM technology and build the DBM PC Cluster system consisting of the software and the latest computer system which meets the concept. The DBM PC Cluster system processing performance shows greater than several thousand images per hour capacity. In this paper, we will report the evaluation results and scalability for future mask metrology.

Keyword: mask CD-SEM, Design based Metrology, DBM, Edge placement error, EPE, Contour extraction, 2D CD

1. INTRODUCTION.

Importance of mask metrology for mask fabrication and wafer lithography is increasing while minimum dimension of features on wafer and mask are decreasing. The state-of-the-art mask, EUV and 193i optical, have many lithography assist features and are starting to include curvilinear features, To control and certify patterning accuracy of assist features and curvilinear features, it is necessary to measure not only multiple 1-dimentional(1D)-CD but also multiple 2D-CD. Figure 1 shows that limitation of 1D-CD metrology for complex patterns. Increasing complexity of features makes it more difficult to define CD as indexes to indicate the fidelity of mask to design and to qualify mask fabrication. Therefore 2D-CD by using design-based metrology (DBM), which extracts contour based on design data precisely and evaluate edge placement error (EPE) with using design data, is one of most essential solution[1,2,3].

Progress of mask and lithography technology enhancements, or SRAF, ILT, OPC, multiple patterning, phase shift mask, lithography simulation[4] and Multi-beam mask writer, have been requiring improvement of both speed and precision of mask DBM metrology. On the viewpoint of speed, high precise mask needs significant number of locations to be measured. Smaller dimensions of pattern and higher density of patterns on mask increase data volume to be processed. These trends require higher computer load and longer calculation time, especially 2D-CD measurement. Our motivation of this study is development of novel DBM technology to enhance productivity of mask CD metrology in current and feature node.

2. CONCEPT OF HIGH SPEED DBM

Figure 2 shows that data flow of Design based metrology (DBM). Mask CD-SEM acquires SEM images with recipe data and extracts contours based on EDA design data to output 1D-CD, Contour, 2D-CD and EPE by using DBM. DBM system loads single recipe file with large volume and high number of SEM images to process images parallelly. Parallel processing with high efficiency is essential solution for achieving fast image processing.

38th European Mask and Lithography Conference (EMLC 2023), edited by
Uwe F.W. Behringer, Jo Finders, Proc. of SPIE Vol. 12802, 128020C
© 2023 SPIE · 0277-786X · doi: 10.1117/12.2675545

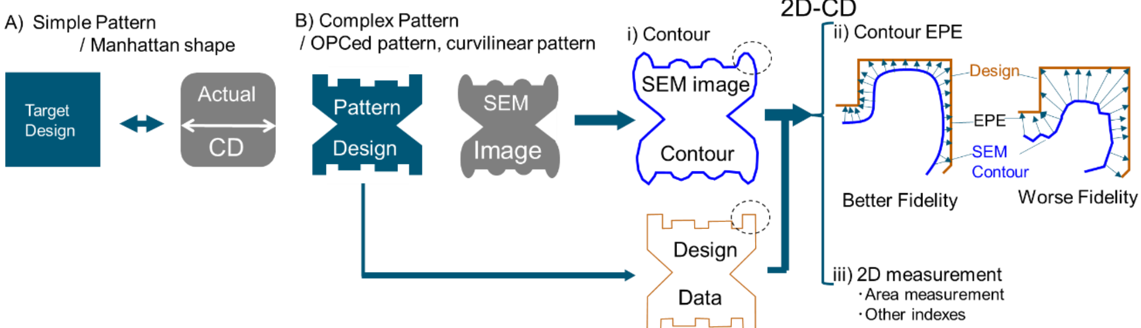


Figure 1. Contour and 2D measurement of DBM

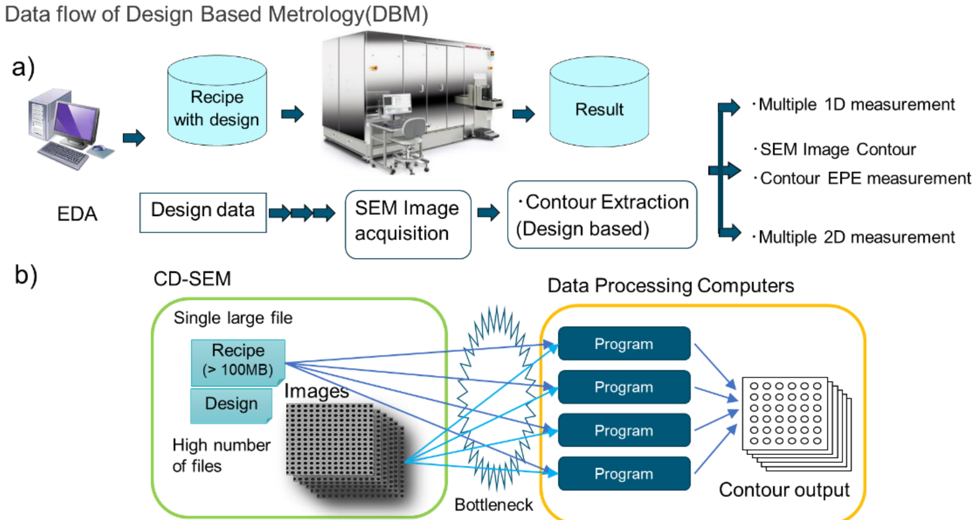


Figure 2. Data flow of DBM: a) Data input and output, b) Bottleneck

There are many methodologies to improve parallel processing efficiency in high power computing. Major strategies of fast parallel processing are i) higher number of fast CPUs, ii) fast file access, iii) fast communication between PC/CPUs (LAN). Rapid progress of computing technology gives us the greatest opportunity to improve capability of i), ii) and iii). However, it is mandatory to build the algorithm to process images with minimum overhead, which might prevent to utilize performance of high number of CPUs. One of the largest overhead contributors is data file access. Both parallel file access to a single recipe file by several programs and file access to high number of image files might be bottlenecks to cause overhead, as shown in Figure 2b).

We employed the three concepts to minimize overhead in Advantest's novel DBM technology or software algorithm, as described in Figure 3. We developed new software, which we call as Intelligent Task Splitter (ITS), based on these concepts. 1) Pre-processing of ITS splits a single task into small tasks prior to image processing without additional overhead. 2) Small task size makes it easier to adjust workload of each CPUs. 3) By predicting work schedule of programs, ITS can provide a new task to the program with high efficiency.

It is necessary to optimize task size and its number because excessive number of tasks might cause overhead.

ITS employed new algorithm to optimize task size and number to minimize overhead.

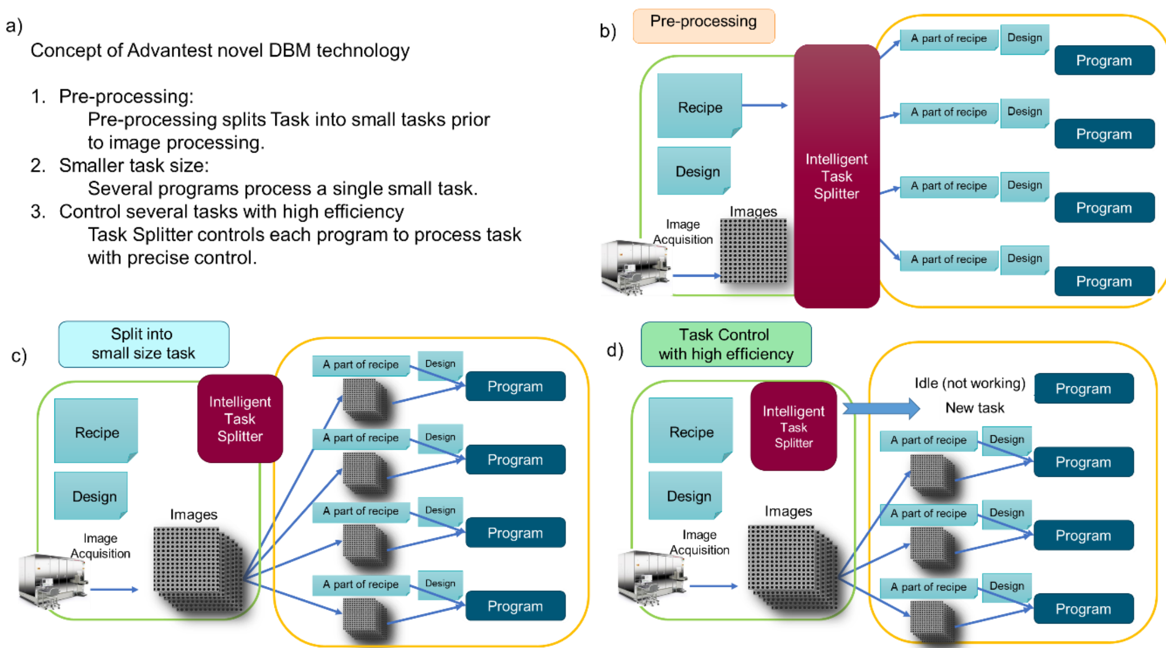
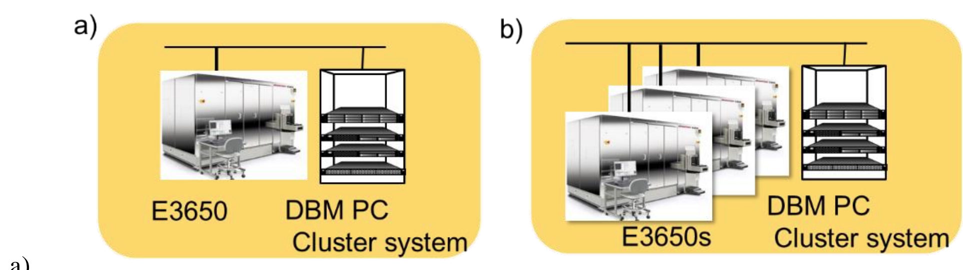


Figure 3. Concepts of Advantest's novel DBM technology

a) Concepts, b) Pre-processing, c) Small task, d) Intelligent control



a)

Figure 4. Example of configuration of CDSEM and DBM Cluster system:

a) Online, one CDSEM by one, b) Central server, N CDSEMs by one

Advantest's novel DBM technology needs the optimal computer hardware for our concepts and strategies. We must consider not only high-performance computing, such as high number of CPU cores, but also economical point of view, such as smaller electric power footprint, lower maintenance cost and smaller computer footprint. Therefore, we employed the latest CPU architecture with high number of CPU cores and computer virtualization technology which make it easy to minimize these footprints.

DBM PC Cluster system is consisted of combination of software of the novel DBM technology and state-of-the-art computer cluster. Figure 4 shows different kinds of configurations of the system to fit the customer's application or use-case. DBM PC Cluster system can serve a single CDSEM (a) or can work with multiple CDSEMs as a central server of image processing(b).

3. DESIGN OF EXPERIMENT

To prove the concept of Advantest's novel DBM technology is effective to improve processing speed, we built the system described in Table 1. Number of CPU cores was changed to evaluate the effect of increasing CPU numbers or CPU scalability. (see 3.2)

In this experiment, we selected Cluster PCs in commercial PC market and virtualization software running with windows OS. Our word, "Cluster PC" means a system containing high number of CPUs cores, PC chases, storage and internal communication devices in single computer rack. We employed the virtualization technology of hardware/computer operating system on on-premises environment.

#	Item	Description
1	Software	Software: Software of Advantest's novel DBM technology, newly developed, running as an optional software of PMsite on Advantest Mask CDSEM E3650 PC operation system: Windows server and virtualization software
2	Hardware	・Number of CPU cores/system: 128, 256, 368 ・Data storage: Enterprise SSD ・Connection: High speed internal LAN(within system and to CDSEM)

Table 1. DBM PC Cluster system

3.1 Evaluation performance improvement with combination of software and hardware

We perform the evaluation with three conditions (a, b, c) of combinations of software and hardware as Table 2. Condition a) is reference. Condition b) is expected to clarify how software design improves the speed to qualify software design. Condition c) must show the importance of combination of software and hardware to prove the concept of the novel DBM technology.

#	Software	Hardware	Purpose
a	Conventional software	Conventional DBM PC (96 CPU cores)	Reference
b	New software based on concept	Conventional DBM PC	To qualify software
c	New software based on concept	DBM PC Cluster PC (384 CPU cores)	To prove the concept

Table 2. Evaluation condition of combination of software and hardware

3.2 CPU Scalability evaluation

To discuss about parallel processing efficiency and future improvement, it is important to evaluate the CPU scalability. CPU scalability means the ability that increasing number of CPU can enhance processing speed in proportion to number of CPUs. Evaluation of CPU scalability can estimate the efficiency of parallel processing and might be able to predict the future performance with higher number of CPUs. We evaluated the CPU scalability with changing number of CPU cores on DBM Cluster System, a) 128, b) 256, c) 384 CPU cores.

3.3 Evaluation of processing speed

We evaluated processing speed by estimating processing time to process 5000 images with contour extractions and calculating number of images processed in an hour as speed, images/hour or IPH. The conditions of images used in this evaluation is shown in Figure 5 and Table 3.

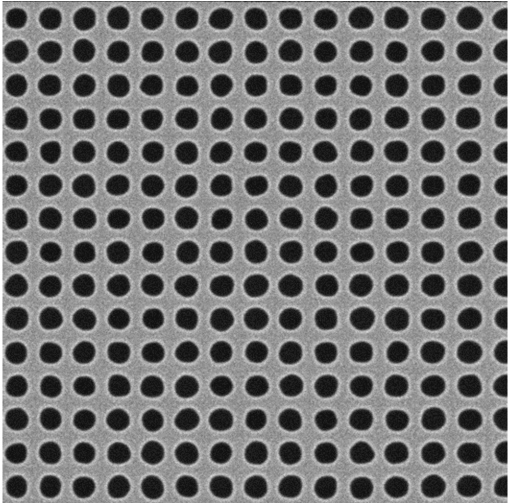


Figure 5. SEM images: One of the 5,000 SEM images in evaluation

	Items	Condition
1	Processing	processing 5000 Images with contour extraction.
2	Image	1.44um FOV, 1024pixel format.
	Feature in image	169 Contact hole/image (Advantest standard mask)

Table 3. Image condition

4. RESULTS AND CONCLUSION

4.1 Evaluation results of performance improvement with combination of software and hardware

Figure 6 indicates the processing speed to process images with conditions of Table 2. Processing speed of running conventional software on the conventional computer was 500 image per hour (IPH). New software running on the conventional computer improved speed to 1200 IPH or 240% faster than reference or that of conventional software/hardware. This result shows that software design is qualified for speed improvement on conventional computers.

Processing speed of new software running on the DBM Cluster PC achieved up to 7200 IPH or 1440% faster than reference. This result illuminates that combination of the new software and DBM cluster PC can improve performance by 10 times faster and the new software can utilize CPU capability fully.

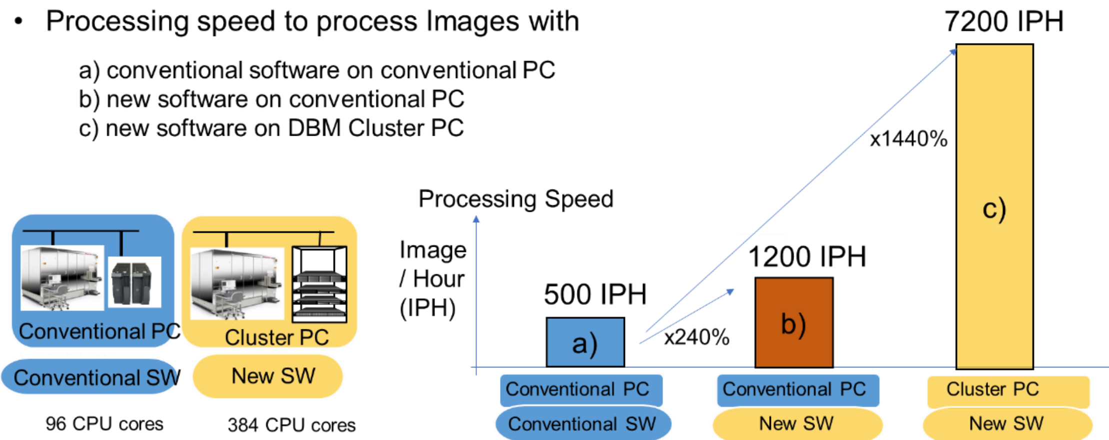


Figure 6. Speed evaluation results by combination software and hardware

4.2 CPU Scalability evaluation

Figure 7 shows evaluation results of CPU scalability of DBM Cluster system. By increasing number of CPU cores, 128, 256 and 384, processing speed is improved from 2,400 (image/hour or IPH), to 4900 and 7200 IPH. The new software on Cluster PC has shown to utilize CPU capability with good efficiency.

Good CPU scalability proved the new software to be able to support extensions over 7200 IPH and meets the future customer's demand. We concluded it might be possible to extend performance up to double or triple of current evaluation results, 7200 IPH by extending the current strategy and concepts described in section 2.

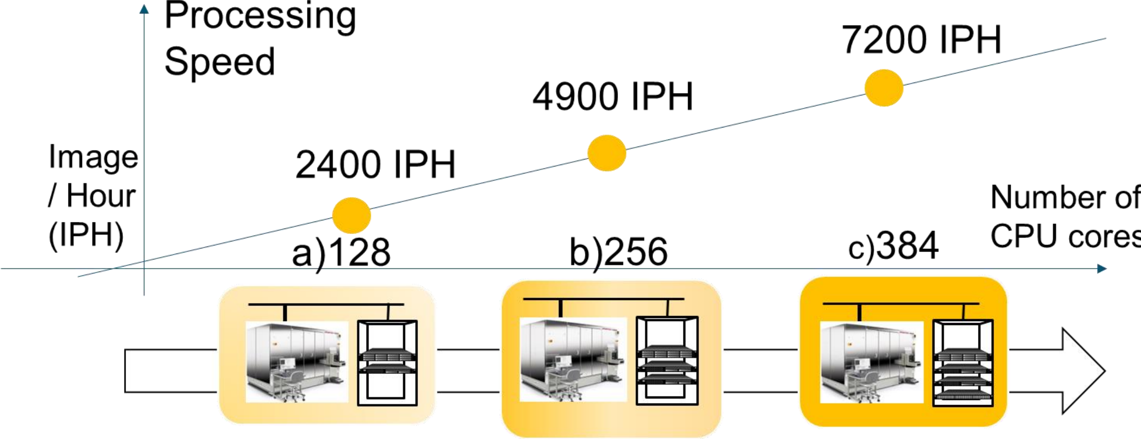


Figure 7. CPU scalability evaluation results by combination software and hardware

4.3 Outcome of the DBM technology

Figure 8 shows the CDSEM images and image contour outputted by DBM Cluster system. We examined the output of faster processing or contour extraction results. We confirmed that fast processing has no side effect on image processing results or contour extraction accuracy. We concluded that the DBM Cluster system or combination of the novel DBM technology, latest Cluster PC and Advantest's Mask CDSEM, enables us to

acquire accurate image and output precise SEM image contour.

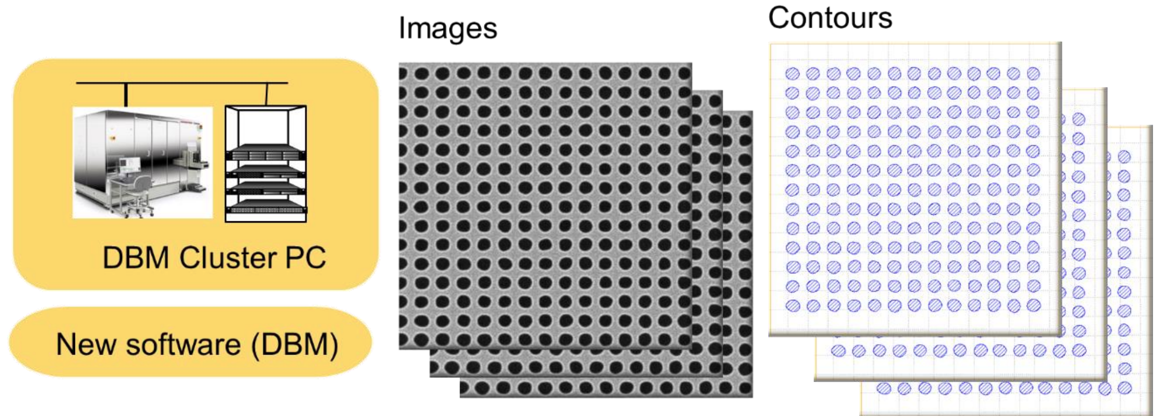


Figure8. Contour outputted from DBM cluster system

5. CONCLUSION

1. Advantest developed DBM technology (software algorithm) and DBM Cluster system.
2. Evaluation results show newly designed software and latest Cluster PC can process over 7000 Images per hours. Concepts of Advantest's novel DBM technology were proved.
3. The CPU scalability of our software indicated that this algorithm will be able to support shrinking features and increasing number of measurement patterns to meet future customer's demand.

Acknowledgement

We would like to express our deepest gratitude to Samsung mask metrology team for discussion and advice.

Thanks to our colleagues to work with this study and supported us.

REFERENCE

[1] Won Joo Park, et. al, "The study on quantifiable analysis for complex OPCed patterns based on Mask CD SEM contour information" Proc. of SPIE Vol. 9984 998407-12 (2016);
[2] Yuta Chihara, et. al, "New CDSEM Technology and its Performance for Multiple Patterning Process" Proc. of SPIE Vol. 8441 844109-8 (2012)
[3] Yuki Yamada, et. al, "Novel CD-SEM technology for a shadow phenomenon in SEM image" Conference presentation in IMS Chips workshop "Beams and More" 2019 (2019)
[4] Shingo Yoshikawa, et. al, "The capability of lithography simulation based on MVM-SEM" Proc. of SPIE Vol. 9635 96351X-1 (2015)

Development of an e-beam/i-line stepper intra-level mix and match approach with the photoresist mr-EBL 6000.5 for PIC related structures such as waveguides, ring resonators and coupling structures

M. Reinhardt*[a], C. Helke[a,b], S. Schermer[a], S. Hartmann[b], A. Voigt[c], D. Reuter[a,b]
[a]Fraunhofer Institute for Electronic Nanosystems ENAS, Technologie-Campus 3, 09126 Chemnitz, Germany; [b]Center for Microtechnologies (ZfM), Technical University of Chemnitz, Reichenhainer Str. 70, 09126 Chemnitz, Germany; [c]micro resist technology (mrt) GmbH, Koepenicker Str. 325, 12555 Berlin, Germany

ABSTRACT

In this paper the development of an intra-level mix and match (ILM&M) process, an expression for the exposure of one resist layer with at least two different exposure technologies, for the negative tone resist mr-EBL 6000.5 (micro resist technology, Germany) is demonstrated. Process development is conducted on a layout with photonic integrated circuit (PIC) related waveguides (WG), ring resonators and coupling structures on 150 mm silicon wafers with a 1000 nm SiO_2 layer and a 450 nm low pressure (LP) Si_3N_4 layer on top. In order to match the intended structure dimensions perfectly, the ideal exposure dose has to be determined with an i-line wafer stepper and in parallel with an e-beam lithography (EBL) system. In addition, different post exposure bake (PEB) processes and their influence on resulting structures, which are investigated by means of CD-SEM and profilometer measurements are investigated. It is shown, that regarding pattern fidelity, coupling structures exposed by EBL match the layout design better than those exposed by the i-line stepper. For the purpose of further optimizing the matching of generated coupling structures to the targeted design, different proximity effect correction (PEC) parameter sets are applied. CD-SEM measurements reveal the PEC parameter set which is most suitable for generating the targeted coupling structures. By combining the measurement results of structures exposed with different exposure doses and selecting the best PEC parameter set regarding structure dimensions and pattern fidelity, a processing recipe for an e-beam/i-line stepper ILM&M with the negative tone resist mr-EBL 6000.5 is successfully established.

Keywords: intra-level mix and match, e-beam lithography, mr-EBL, i-line stepper lithography, PIC, waveguide

1. INTRODUCTION

To precisely fabricate structures at the nanoscale, EBL is a suitable patterning technique [1-4]. EBL can also be applied for PIC related structures such as gratings [5-6] and WGs [6]. However, due to its direct writing method, EBL is time-consuming and therefore expensive. In contrast, e.g. i-line stepper lithography offers a higher throughput but is limited in resolution. In the mid-1990s a new lithography process combining UV and electron beam exposure on just one resist layer has been presented in several papers [7-9]. This kind of approach is called intra-level mix and match lithography (ILM&M) and has the advantage of time saving because only structures below the UV lithography's resolution limit are written by EBL whereas structures above the resolution limit of UV lithography are exposed by UV radiation [2, 9-11]. In addition, only one resist layer has to be processed, which also reduces the lithographic processing time. In order to apply e-beam/UV stepper ILM&M lithography successfully, the resist has to be sensitive to UV and electron beam radiation. In recent years, the epoxy-based negative-tone resist series “mr-EBL 6000” (from micro resist technology, Germany) showed great performance in structuring resist layers with e-beam systems with a sub-100 nm lateral resolution dependent of the layer thickness [3-4].

38th European Mask and Lithography Conference (EMLC 2023), edited by
Uwe F.W. Behringer, Jo Finders, Proc. of SPIE Vol. 12802, 128020D
© 2023 SPIE · 0277-786X · doi: 10.1117/12.2675558

2. EXPERIMENTAL SECTION

2.1 Layout design

The development of the ILM&M process is conducted on a layout design with PIC related WGs, ring resonators and coupling structures. Figure 1a shows an optical microscope image of these structures and CD-SEM images (Figure 1b, 1c and 1d) which are taken at corresponding sites. The CD-SEM image of WG, ring resonator and the gap in between shown in Figure 1b is rotated by 90 degrees clockwise compared to the optical microscope image because in CD-SEM, edges are depicted better with scanning direction perpendicular to the corresponding edge. A 900 nm wide straight WG is broadened to 21.4 µm at both ends, where it is merged to a coupling structure. The coupling structure consists of a 17 x 33 holes array with square holes of 800 nm edge length separated by 400 nm wide grating patterns. In the middle of the WG, a 900 nm wide ring resonator is located beside the WG with a gap of 1000 nm between ring resonator and WG. In addition, a second design with a gap of 500 nm between ring resonator and WG is realized in this work but not shown in Figure 1.

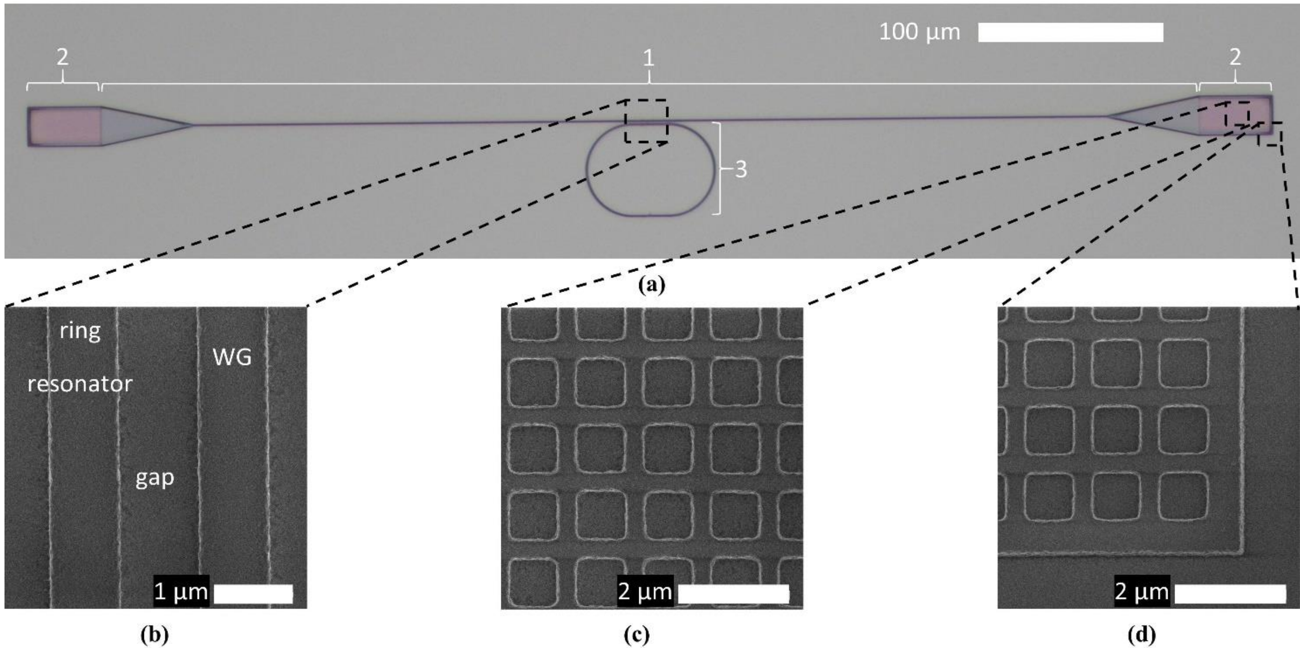


Figure 1. (a): Optical microscope image of (1) implemented WG with (2) coupling structures at both ends and (3) ring resonator structure beside the WG in the middle of the WG structure. CD-SEM top view images of (b) ring resonator, WG and gap in between, (c) center of coupling structure and (d) edge of coupling structure. The CD-SEM image with WG and ring resonator is rotated 90 degrees clockwise compared to the optical microscope image.

2.2 Lithography process

For this study 150 mm silicon wafers with a 1000 nm SiO_2 and a 450 nm LP-Si_3N_4 layer on top are used. In order to achieve a good overlay accuracy, 200 nm high alignment marks are etched into the Si_3N_4 previous to the applied ILM&M process. Thus, both following exposures can align in a two-step alignment first to global marks for rough alignment and subsequently chip marks for fine alignment. After the application of a dehydration bake for 30 min in a convection oven at 200 °C as substrate pre-treatment, wafers are spin coated at 3000 rpm for 120 s with 3.5 ml of the negative tone resist mr-EBL 6000.5 in a semi-automatic spin-coater tool SM-200 (Sawatec AG). Subsequently, the resist layer is soft baked for 180 s at 120 °C on a semi-automatic hot-plate HP-200 (Sawatec AG) to evaporate remaining solvents from the resist layer resulting in a homogenously thickness of about 500 nm. Afterwards the exposures with both e-beam and i-line stepper were applied. The e-beam exposure is performed with the VISTEC SB254 (50 kV acceleration voltage, 12 A/cm² current density, shaped beam, CD > 20 nm, standardized PEC). The i-line wafer stepper exposure is done by a NIKON NSR 2205i11D (CD > 350 nm). In a first step the exposure dose which leads to a 1 to 1 print is determined. This means, that the exposed structure should exactly match the layout.

After exposure, a PEB is applied on the semi-automatic hot plate, which accomplishes cross-linking of the resist which is initialized during the exposure. Two different PEBs are investigated: 5 min at 120 °C, as recommended in the datasheet of the resist and 1 min at 90 °C. After the PEB the puddle development was applied at room temperature by the use of developer mr-Dev 600 for 60 s at a rotation speed of 75 rpm in a semi-automatic development tool LRD-250 (Sawatec AG).

2.3 CD-SEM and profilometer measurements

The CD-SEM CS4800 (Hitachi) allows to conduct fully automated measurements on wafer level which are used for the characterization of the manufactured structures. In order to determine the width of the ring resonator, the WG and the gap in between in one image, the site where the ring resonator is located beside the WG is selected as measurement position as shown in Figure 1a. For the measurement of the square sized hole diameter and the width of the grating patterns of the coupling structures, the edge of the coupling structure is selected as measurement position (Figure 1d). In addition, a beam shift of 12 µm in negative x and 12 µm in positive y direction is conducted in order to measure these structures in the center of the coupling structure (Figure 1c).

For determining the resist thickness, profilometer measurements are performed on the ring resonator, the WG as well as on the coupling structures by means of a Dektak 8 Stylus Profiler (Veeco Instruments Inc.). With a tip radius of 2.5 µm of the Dektak tool it is not possible to reach the bottom of a square sized hole of the coupling structures. However, profiles are measured across a whole coupling structure to cover the bottom outside the coupling structure and the grating patterns of a coupling structure. Measured resist thickness is determined to be the height difference of the profile outside the coupling structure and on a grating pattern of a coupling structure.

3. RESULTS AND DISCUSSION

3.1 i-line stepper lithography

In first experiments a PEB of 5 min at 120 °C is applied according to the recommended values from the datasheet of the mr-EBL 6000 series. The applied exposure doses ranging from 220 mJ/cm² to 580 mJ/cm² in 10 mJ/cm² steps with the i-line stepper. The resulting structures are characterized by means of CD-SEM measurements. Figure 2a shows the ring resonator and the WG widths as well as the width of the gap in between in dependence of the applied exposure dose for a range of 290 mJ/cm² to 350 mJ/cm². A clear trend can be derived that a higher exposure dose leads to a broadening of the ring resonator and the WG. Corresponding to that, the gap between ring resonator and WG is reduced with increasing exposure dose.

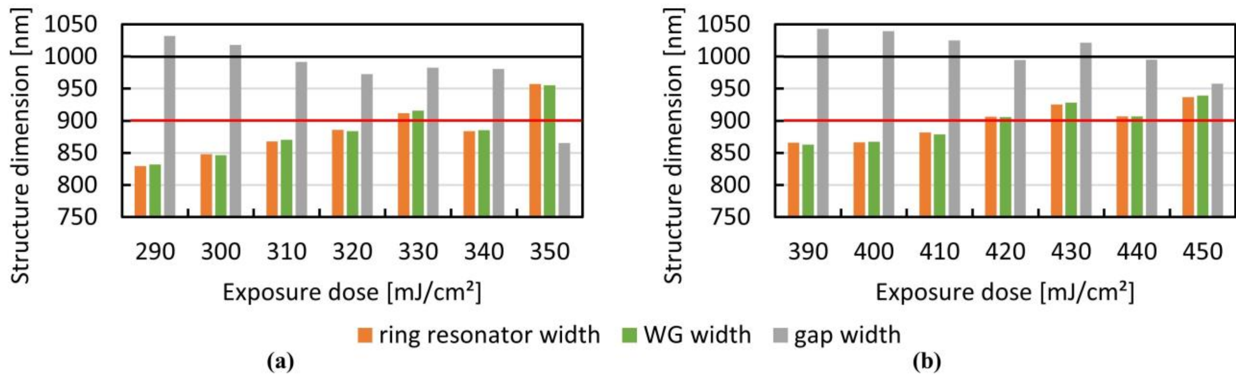


Figure 2. i-line stepper exposure dose dependencies of WG, ring resonator and gap dimensions (targeted gap: 1000 nm) for a PEB of (a) 5 min at 120 °C and (b) 1 min at 90 °C. Black and red lines in diagrams refer to the target dimensions of gap width and WG or ring resonator width, respectively.

The required exposure dose to realize ring resonator and WG widths of 900 nm with a gap of 1000 nm in between is about 330 mJ/cm². However, considering the corresponding CD-SEM image (Figure 3a), the bottom of the gap between WG and ring resonator has additional features, indicating crosslinked resist residues inside the gap. For the design of 500 nm gap between WG and ring resonator, crosslinked resist residues inside the gap are found at an exposure dose of 230 mJ/cm² (Figure 3c) and at even lower exposure doses whereas ring resonator and WG widths at this exposure dose

are only about 800 nm. These results led to the decision to apply a slightly gentle PEB to reduce crosslinking degree. Based on previous experiments, a PEB of 1 min at 90 °C is applied, which results in a lower crosslinking degree of the resist initialized during exposure. This process allowed to fabricate the following structures: 1000 nm as well as the 500 nm gap between WG and ring resonator without any crosslinked resist residues inside the gap (Figure 3b and 3d), ring resonator widths of 900 nm and WG widths of 900 nm. For the 1000 nm and the 500 nm gap design a required exposure dose of 420 mJ/cm² (Figure 2b and 3b) and 400 mJ/cm² (Figure 3d) is found, respectively. According to these results, further experiments were conducted only with a PEB of 1 min at 90 °C. By means of profilometer measurements, the resist thickness of the WG, the ring resonator and the coupling structures at 400 mJ/cm² and at 420 mJ/cm² is determined and it is in the range of 417 nm to 455 nm and 412 nm to 454 nm, respectively.

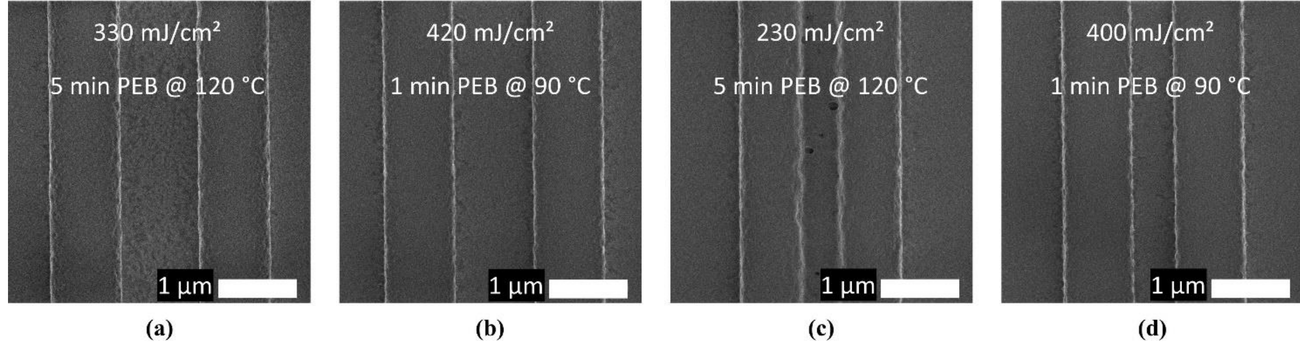


Figure 3. CD-SEM images of ring resonator (left) and WG (right) and a gap of (a) and (b) 1000 nm and of (c) and (d) 500 nm in between.

The with different exposure doses exposed coupling structures were characterized with the CD-SEM. At an exposure dose of 360 mJ/cm², the square holes and the grating patterns match the layout's design dimensions for the coupling structure. In addition, the dimensions measured at the edge of the coupling structure and in the center are the same within the uncertainties of the measurement systems. So basically, it is possible to manufacture coupling structures with the targeted dimensions by means of i-line stepper lithography. However, as shown in Figure 4a and 4b, the shape of the holes clearly differs from the squared shape from the layout design. The corners are strongly rounded which could cause a lower coupling efficiency. This deviation from the layout design is attributed to UV light reflection at the LP-Si_3N_4 surface during exposure. This effect can be corrected by using software solutions, which apply an optical proximity correction (OPC) to the layout. This is a similar solution like using a PEC in EBL. But instead of varying the exposure dose, compensation structures are added to the reticle. Since this method is extremely cost intensive and time consuming, it is not commonly used at universities and R&D institutes.

In addition to the deviation from the 1 to 1 print, the bottom of the holes shows similar artefacts like the bottom between ring resonator and WG in Figure 3a, which indicates that the holes are not completely free of resist residues. For these reasons, EBL is applied to realize coupling structures, which match the layout's design more precisely than the ones exposed by the i-line stepper.

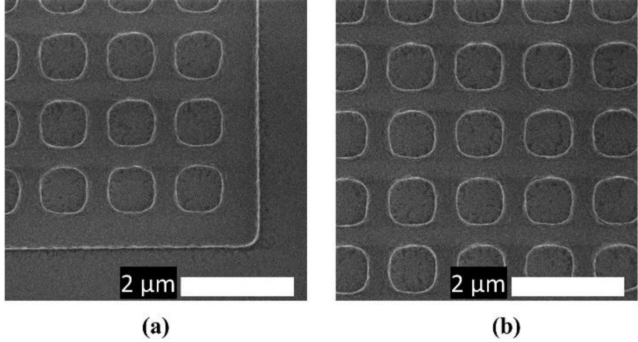


Figure 4. CD-SEM images of a coupling structure exposed by the i-line stepper at an exposure dose of 360 mJ/cm² conducted at (a) the edge of the coupling structure and (b) in the center of the coupling structure.

3.2 E-Beam lithography

In accordance to i-line stepper lithography different exposure doses were applied ranging from 1.5 µC/cm² to 4.5 µC/cm² in order to manufacture coupling structures, which match the layout's design concerning targeted shape of the holes and dimensions. A selection of SEM images taken at different applied exposure doses is shown in Figure 5. It is clearly visible, that with increasing exposure dose the quality of the square holes decreases and the crosslinking of the resist inside the holes increases.

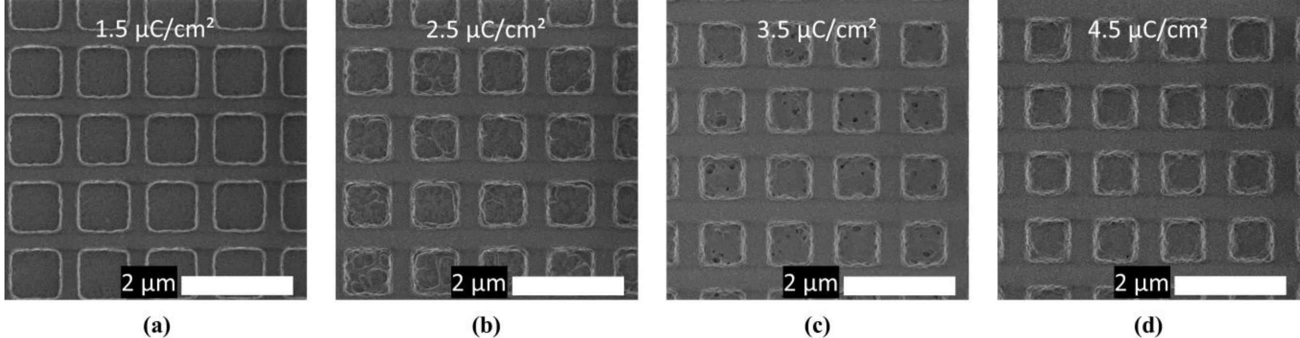


Figure 5. CD-SEM images depicting the center of the coupling structures taken at an exposure dose of (a) 1.5 µC/cm², (b) 2.5 µC/cm², (c) 3.5 µC/cm² and (d) 4.5 µC/cm².

The generated structures were investigated by means of CD-SEM measurements and the square hole diameter and grating pattern widths were determined at the edge as well as in the center of the coupling structure. Figure 6 shows these values in dependence of the applied exposure dose ranging from 1.5 µC/cm² to 3 µC/cm² for a predefined Vistec PEC's standard (Figure 6a) and for an optimized PEC parameter set (Figure 6b).

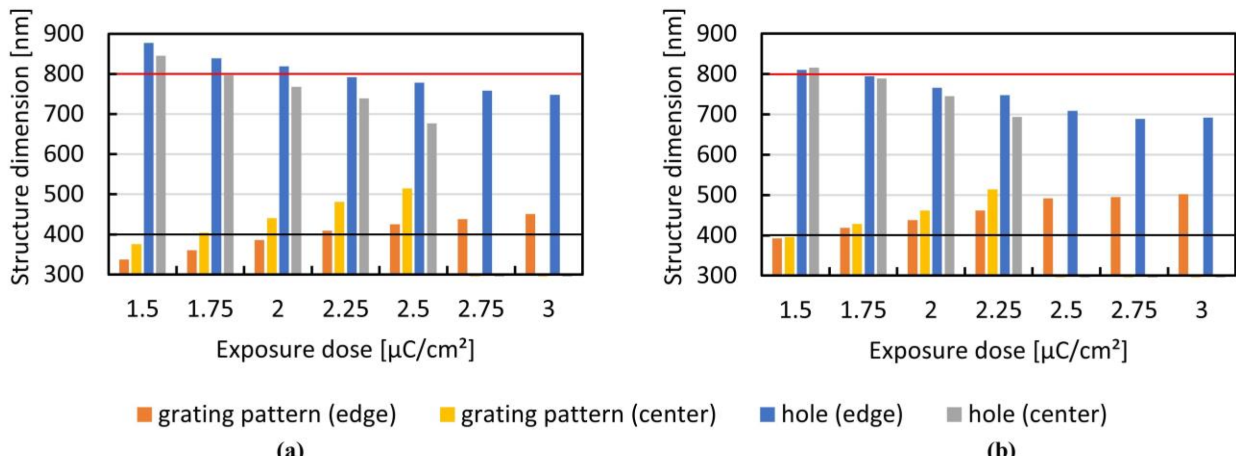


Figure 6. Hole diameter and grating pattern width measured at the edge and in the center of the coupling structure in dependence of the applied exposure dose manufactured with (a) a predefined Vistec PEC's standard applied and (b) an optimized PEC parameter set which delivers structures which match the layout's design more precisely. Black and red lines in diagrams refer to the target dimensions of grating pattern width and hole diameter, respectively.

In order to have a narrow size distribution of square hole diameters across the coupling structures it is important to adjust the exposure dose as a function of position. This is done by optimizing the PEC parameter set, which will assign a calculated exposure dose value to a structure in relation to the position within the layout, as shown in Figure 7a and 7b. First a Monte Carlo Simulation of electron scattering in a resist-substrate model is performed and the radial energy distribution is fitted with two Gaussian functions. The obtained parameters are used to calculate the electron energy distribution within the layout yielding the relative exposure dose factors shown as different colors in the layout in Figure 7.

According to a larger number of electrons initiating the crosslinking of the resist, with increasing exposure dose increasing grating pattern widths and decreasing square hole diameters were measured. For the lowest applied exposure dose and the predefined Vistec PEC's standard, grating pattern width at the edge of the coupling structure differs about 40 nm compared to the width in the center of the coupling structure (Figure 6a). This difference increases with increasing exposure dose. In accordance to that, the difference of the square hole diameter measured at the edge and in the center of the coupling structure is about 32 nm at an exposure dose of 1.5 µC/cm² and increases with increasing exposure dose. Indeed, for structures exposed with the optimized PEC parameter set, it is possible to decrease this difference between edge and center of the coupling structure from 40 nm down to about 6 nm at an exposure dose of 1.5 µC/cm² (Figure 6b). This difference also increases with increasing exposure dose, however, structure dimensions match the layout design dimensions nearly perfectly at an exposure dose between 1.5 µC/cm² and 1.75 µC/cm².

CD-SEM images of structures manufactured with the optimized PEC parameter set at an exposure dose of 1.5 µC/cm² are shown in Figures 7e and 7f. Coupling structures exposed by EBL show better pattern fidelity concerning hole squareness as compared to stepper lithography (compare Figure 4 & Figure 7). In addition, there are no resist residues visible at the bottom of the square holes at the edge of the coupling structure indicating, that the square holes are completely free of crosslinked resist. Even if there is a light structure visible at the square hole bottom in the center of the coupling structure, it is much less pronounced compared to the structures exposed by the i-line stepper shown in Figure 4b, revealing less remaining crosslinked resist inside the square holes.

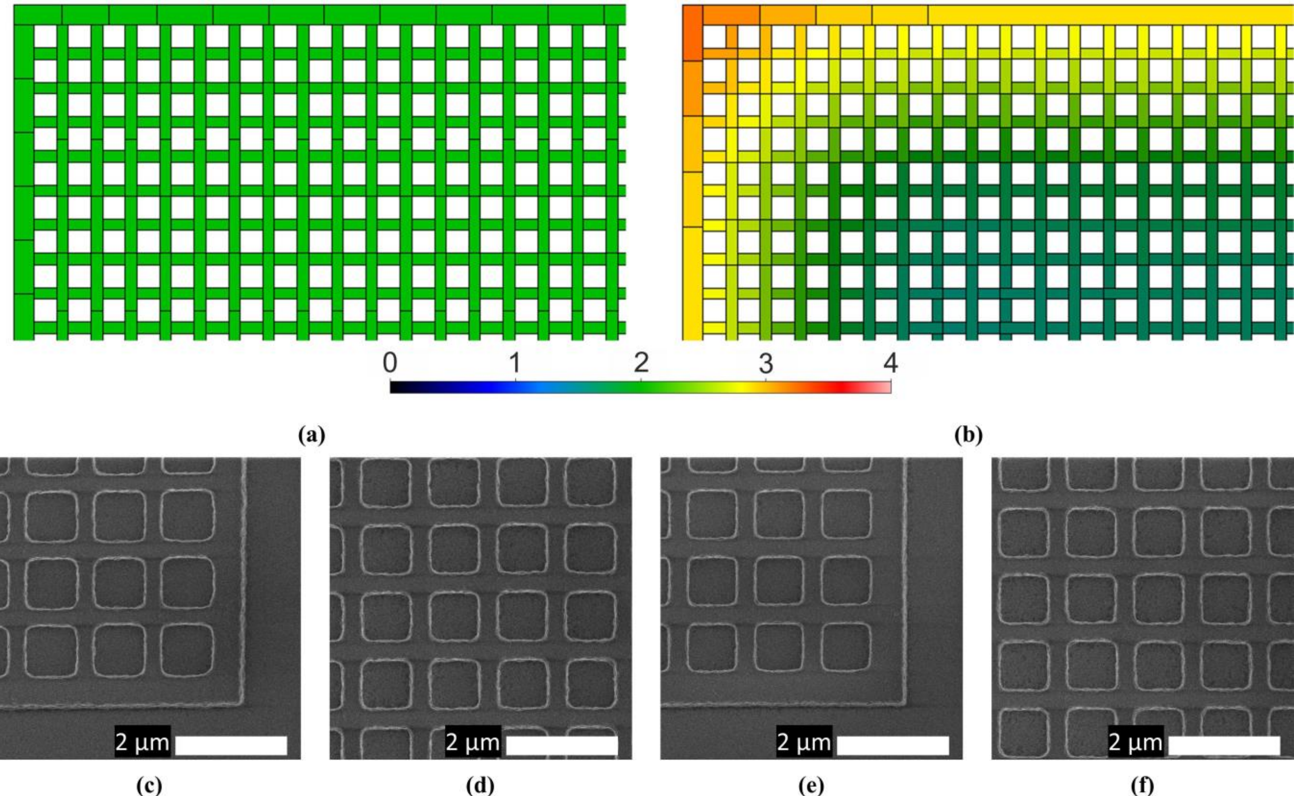


Figure 7. Comparison between a (a) constant e-beam exposure dose and (b) applied exposure doses with optimized PEC parameter set (color scale indicates relative exposure doses). (b) shows the by PEC calculated relative exposure dose values indicated by color as a function of position. Colored structures being exposed. Bottom line depicts CD-SEM images of the coupling structures exposed with the above exposure dose correction. Left CD-SEM images (c) and (d) show a structure width mismatch in x and y-direction for a constant exposure dose compared to the PEC corrected exposed structures at the right with a hole diameter variation less than 10 nm ((e) and (f)).

The resist thickness of the manufactured coupling structures was determined by profilometer measurements. Figure 8 shows the measured resist thickness dependent on the applied exposure dose for the two applied PEC parameter sets discussed above. Remaining resist thickness increases with increasing applied exposure dose.

For the final device the generated PIC resist patterns have to be transferred by a dry etching process into the LP-Si_3N_4 substrate, having a targeted thickness of about 450 nm. By assuming an etching selectivity of 1 to 1 for resist to substrate, the resist thickness has to be 450 nm. However, resist thickness at 1.5 µC/cm² is only about 330 nm. For that reason further experiments regarding determination and fine-tuning of etching rate of mr-EBL6000.5, as well as use a thicker mr-EBL 6000 resist layer will be conducted in order to meet the targeted 450 nm of LP-Si_3N_4 pattern depth.

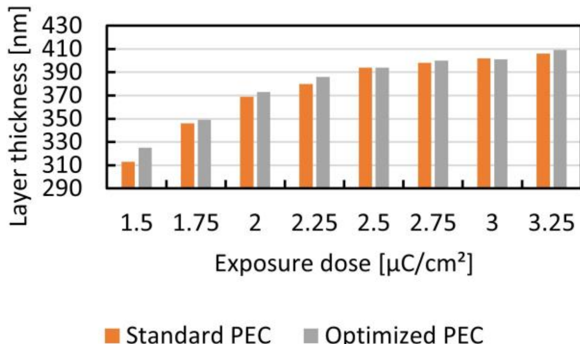


Figure 8. Remaining resist thickness dependent on the applied exposure dose for different PEC parameter sets tested.

4. CONCLUSION

In this paper the manufacture of PIC related WGs, ring resonators and coupling structures having dimensions ranging from 400 nm widths to about 900 nm of a given layout design by means of i-line stepper and EBL on separate wafers with the negative tone resist mr-EBL 6000.5 is demonstrated. The generated resist structures were investigated concerning shape and dimensions with CD-SEM measurements. With the application of a PEB with 1 min at 90 °C for the with an i-line stepper exposed PIC layer, the manufacture of the 500 nm gap between the WG and the ring resonator patterns with an exposure dose of 400 mJ/cm² is successfully demonstrated. For the 1000 nm gap between WG and ring resonator pattern the required exposure dose is 420 mJ/cm². The large coupling structures basically can be manufactured by means of i-line stepper, but they differ from the layout design as the square sized hole corners are strongly rounded. For coupling structures generated by EBL a better structure fidelity concerning hole squareness was observed. Two different PEC parameter sets were applied in order to improve the structure uniformity at the edge and in the center of the coupling structure. Indeed, a PEC parameter set which delivers uniform structures at an e-beam exposure dose of 1.5 µC/cm² was demonstrated. Resist thickness is determined and is about 330 nm. In further experiments, it will be investigated if this resist thickness is sufficient to etch 450 nm LP-Si_3N_4. Combining results from i-line stepper and EBL, provides the ability to establish an ILM&M lithography process for the given layout design.

ACKNOWLEDGEMENT

This work is co-financed by: tax funds on the basis of the budget approved by the Saxon State Parliament. The authors acknowledge Dr. Chris Stöckel, Julia Wecker, Toni Grossmann, Dr. Alexey Shaporin, Dr. Jörg Martin and Franz Tank for many good discussions and the preparation of the layout design.

REFERENCES

[1] A. A. Tseng et al. "Electron beam lithography in nanoscale fabrication: recent development", IEEE Transactions on Electronics Packaging Manufacturing 26(2), p.141-149, (2003).

[2] C. Helke et al. „Intra-level mix and match lithography with electron beam lithography and i-line stepper combined with resolution enhancement for structures below the CD-limit", Micro and Nano Engineering, (2023).

[3] C. Martin et al. “Electron beam lithography at 10keV using an epoxy based high resolution negative resist”, Microelectronic Engineering 84, 5-8, p.1096-1099, (2007).
[4] A. J. Taal “mr-EBL: ultra-high sensitivity negative-tone electron beam resist for highly selective silicon etching and large-scale direct patterning of permanent structures”, Nanotechnology 32, (2021).
[5] W. Guo et al. “InP photonic integrated circuit with on-chip tunable laser source for 2D optical beam steering”, 2013 Optical Fiber Communication Conference and Exposition and the National Fiber Optic Engineers Conference (OFC/NFOEC), p.1-3, (2013).
[6] Suraj et al. "Polarization independent grating in a GaN-on-sapphire photonic integrated circuit", Opt. Express 31, 23350-23361, (2023).
[7] P. Zandbergen and H. J. Dijkstra “Evaluation of a Positive Tone Chemically Amplified Deep UV Resist for E-beam Applications”, Microelectronic Engineering 23, 1-4, p.299-302, (1994).
[8] R. Jonckheere et al. “Electron beam / DUV intra-level mix-and-match lithography for random logic 0.25μm CMOS”, Microelectronic Engineering 27, 1-4, p.231-234, (1995).
[9] F. Benistant et al. “A 0.10 μm NMOSFET, made by hybrid lithography (e-beam/DUV), with Indium pocket and specific gate reoxidation process”, Microelectronic Engineering 30, 1-4, p.459-462, (1996).
[10] S. Magoshi et al. "Improved Electron-Beam / DUV Intra-Level Mix-and-Match As A Production Viable Lithography With 100-nm Resolution”, Digest of Papers. Microprocesses and Nanotechnology'98. 198 International Microprocesses and Nanotechnology Conference (Cat. No.98EX135), Kyoungju, Korea (South), p.42-43, (1998).
[11] M. Narihiro et al. “Intra-Level Mix-and-Match Lithography Process for Fabricating Sub-100-nm Complementary Metal-Oxide-Semiconductor Devices using the JBX-9300FS Point-Electron-Beam System”, Japanese Journal of Applied Physics 39, p.6843-6848, (2000).

Fabrication of gate electrodes for scalable quantum computing using CMOS industry compatible e-beam lithography and numerical simulation of the resulting quantum device

Varvara Brackmann[1], Malte Neul[2], Michael Friedrich[1], Wolfram Langheinrich[3], Maik Simon[1], Pascal Muster[3], Sebastian Pregl[3], Arne Demmler[1], Norbert Hanisch[1], Maximilian Lederer[1], Katrin Zimmermann[1], Jan Klos[2], Felix Reichmann[4], Yuji Yamamoto[4], Marcus Wislicenus[1], Claus Dahl[3], Lars R. Schreiber[2], Hendrik Bluhm[2], and Benjamin Lilienthal-Uhlig[1]

[1]Fraunhofer IPMS, CNT, An der Bartlake 5, 01099 Dresden, Germany
[2]JARA-Institut for Quantum Information, RWTH Aachen University, 52056 Aachen, Germany
[3]Infineon Technologies Dresden GmbH & Co. KG, Königsbrücker Str. 180, 01099 Dresden, Germany
[4]IHP Leibniz-Institut für innovative Mikroelektronik, Im Technologiepark 25, 15236 Frankfurt (Oder), Germany

ABSTRACT

Universal quantum computers promise the possibility of solving certain computational problems significantly faster than classically possible. For relevant problems, millions of qubits are needed, which is only feasible with industrial production methods. This study presents an electron beam patterning process of gate electrodes for Si/SiGe electron spin qubits, which is compatible with modern CMOS semiconductor manufacturing. Using a pCAR e-beam resist, a process window is determined in which structure sizes of 50 nm line and 30 nm space can be reproducibly fabricated with reasonable throughput. Based on electrostatic simulations, we implemented a feedback loop to investigate the functionality of the gate electrode geometry under fabrication-induced variations.

Keywords: Electron beam lithography, quantum computing, Si/SiGe, quantum dots, CMOS, e-beam

Contents

1 Introduction **2**

2 Materials and methods **3**

2.1 Gate electrode geometry for qubits . . . 3

2.2 Fabrication process of gate electrodes . . . 3

2.3 Electron beam lithography . . . 4

3 Results and Discussion **6**

3.1 Exploration of the e-beam patterning parameter space . . . 6

3.2 Exploration of the optimal gate electrode geometry . . . 9

4 Simulations **12**

Further author information: (Send correspondence to V.B.)
V.B.: E-mail: varvara.brackmann@ipms.fraunhofer.de, Telephone: +49 351 26073015
M.N.: E-mail: malte.neul@rwth-aachen.de, Telephone: +49 241 8027004

38th European Mask and Lithography Conference (EMLC 2023), edited by
Uwe F.W. Behringer, Jo Finders, Proc. of SPIE Vol. 12802, 128020E
© 2023 SPIE · 0277-786X · doi: 10.1117/12.2675943

5 Conclusion 15

1. INTRODUCTION

Error-tolerant universal quantum computers promise to process some categories of problems exponentially faster and to simulate quantum systems with large Hilbert spaces. The base unit of a quantum computer is the quantum bit (qubit). There are several approaches to define the qubits, particularly by means of superconducting circuits, color centers, trapped ions, topologically protected systems and semiconductor quantum dots (QDs).[1,2] This paper deals with QDs in silicon, which are qubits defined by the single electron spin confined to the QD in a magnetic field.[3] Solving computational problems, which are currently intractable for classical super-computer, often require hundreds of error-corrected logical qubits represented by millions of physical spin-qubits each confined to a QD,[4,5] but only a six spin-qubit chip has been demonstrated in silicon as of yet.[6] Hence, scaling up the number and interconnectivity of qubits is the essential challenge. There are several device architectures of silicon quantum computers suggested,[7–11] but they all have a very high number of gate electrodes to be fabricated in common. Most current efforts for scaling up quantum chips are limited by the low-yield and low-throughput fabrication processes in academic cleanrooms. Therefore, the scaling up with good reproducibility and fabrication yield is only possible in a highly automated industrial semiconductor fabrication environment. Moreover, classical semiconductor circuits produced in this environment combined with the quantum chip signal enable the readout and reinforcement of quantum technology.[10,12,13] Therefore, scalable quantum computing architectures which are compatible with industrial CMOS (complementary metal-oxide-semiconductor) silicon foundries are necessary.[14]

Spin qubits in silicon are excellent candidates for large scale quantum computing for several reasons. QDs are extremely dense compared to other qubit platforms and thus offer the potential for millions of qubits on a chip.[10] Moreover, silicon quantum-chips are compatibile with mature semiconductor fabrication technologies. Isotopically purified nuclear-spinless ^{28}Si is an excellent host material for spin qubits, for its weak hyperfine interaction and low spin-orbit coupling that reduce coupling to a noisy environment.[3,15] Consequently, silicon-based electron-spin qubits show single- and two-qubit gate,[16–21] as well as readout[22–24] fidelities reaching the prerequisite for topological quantum error correction.[20]

The most common approaches for Si-based qubits[25] are Si/SiGe heterostructures, planar metal oxide semiconductor (MOS) devices, fully depleted silicon-on-insulator (FDSOI) devices, and donors in purified silicon.[26] In Si/SiGe heterostructure spin qubits, electrons are vertically confined in a tensile-strained and isotopically purified ^{28}Si layer, which is epitaxially grown between two layers of SiGe. Lateral QD confinement is done by planar gate electrodes. Compared to MOS and FDSOI, Si/SiGe heterostructures ensure improved control as the upper SiGe layer separates the QDs from the defect-rich semiconductor/oxide interface.[11]

Semiconductor spin qubits have many similarities to scaled transistors and hence are advantageous for the integration into the semiconductor manufacturing process technology. There are approaches using all-optical[27,28] as well as e-beam "mix and match" lithography[29,30] integration approaches. E-beam lithography provides high resolution and flexibility of the exposed pattern, because no mask is required, unlike for optical lithography. In this paper we propose an e-beam lithography approach for Si/SiGe qubits which is compatible with a 200 mm industrial CMOS semiconductor manufacturing line.

In most studies on Si/SiGe heterostructures for qubits, the gate electrodes are fabricated using academic e-beam lift-off nanostructuring. This fabrication technique suffers from low yield, poor uniformity, high particle load and hence is not applicable on an industrial scale.[31–38] In this study we apply subtractive electrode fabrication, using plasma etching, which is common for industrial CMOS fabs.[39] Our work is process-oriented and dedicated to the CMOS compatible gate electrode fabrication process and electrode geometry for Si/SiGe qubits, which has not been described in detail before.

The geometry and dimensions of gate electrodes needed for semiconductor QDs vary depending on the applied semiconductor host material and working principle of the electron confinement. To make the quantisation of the electron number in a structure measurable, the spatial confinement of an electron must lead to a level splitting larger than the thermal energy and strong enough to prevent several minima to form due to disorder. Both criteria favor a small effective electron mass. For planar silicon devices having strong confinement in the

crystallographic (100) direction (growth-direction of the Si/SiGe heterostructure), the relevant effective mass is 0.19 m_0, where m_0 is the free-electron mass.[3] The critical feature parameters differ between Si-MOS and Si/SiGe qubits: the closer proximity of the inversion-layer to the gate electrodes in Si-MOS results in typically sharper potentials and stronger disorder. Tunnel-coupling between QDs, which is at the heart of 2-qubit manipulation, thus requires very narrow gates. In many devices this is not achieved and consequently devices lack control of the tunnel-barrier.[40] In Si/SiGe heterostructures, on the other hand, the 30 nm to 50 nm SiGe spacer-layer leads to a widening of the QD potentials.[3] Since this effect limits the sharpness, a reduction of the gate widths below approx. the thickness of the spacer and oxide yield no further advantage. Hence, in this work we optimise the lithography fabrication process targeting a gate electrode width of 30-50 nm, and a pitch which is feasible for e-beam nanostructuring, etching and consequent top gate material deposition. In general the gap between the gate electrodes shall be as small as possible to have good electrostatic control, to screen oxide defects and to achieve higher orbital splitting of the QDs.

In summary, this work is focused on the setup of a CMOS industry compatible e-beam lithography fabrication approach of the gate electrodes for Si/SiGe heterostructure spin qubits. The gate electrode width should range from 30 nm to 50 nm. The distance between the electrodes should be as small as possible from the processing point of view (e-beam lithography process, and consequent etching, gate oxide and self-aligned TiN gap filling). Sections 2.1 and 2.2 show the desired device and electrode geometry, as well as process flow. Section 2.3 describes the CMOS compatible e-beam lithography process applied for the patterning of the gate electrodes. In the results section, Section 3, we first describe the feasibility and overall process window of the e-beam nanostructuring process targeting the 30-50 nm line width and smallest possible pitch. Secondly the obtained knowledge about the lithography process window is applied for further process optimisation to fabricate a specific gate electrode geometry, as described in 2.1. Finally, electrostatic control of QD occupation, tuning tunnel-barriers and robustness against fabricational imperfections are simulated by a finite-element COMSOL model of the device.

2. MATERIALS AND METHODS

2.1 Gate electrode geometry for qubits

The gate electrode arrangement is shown in Fig. 1 a). In addition to the first, structured gate layer, a second metal layer is self-aligned to the first. The inner gates are fanned out to form pads that are later contacted by a higher metal layer (not shown in this picture) using vias. Four channels are formed in the direction of the implanted areas, which are controlled in the final device by additional accumulation gates in a higher layer. The main region of interest for this study is shown in Fig. 1 b) and contains the elements with critical feature size as well as larger gates that have an influence on the e-beam exposure via proximity effects. The active qubit region is shown in Fig. 1 c). It consists of two sets of five (or three) fine gate electrodes separated by a separation gate. Each adjacent set of three fine gates can form a quantum dot. Here, each QD can either be operated in transport mode as a charge-sensitive single-electron transistor (SET) or successively emptied down to the last electron, thus functioning as a qubit. The device is then operated by forming a SET on one side of the separation gate and a qubit on the other side, as exemplified by the red/green dot. Both quantum dots are capacitively coupled, which enables the electron occupation of the qubit-dot to be read out by the SET.

Furthermore, the five fine gates can be used to form two coupled qubit-dots on one side. This allows all the basic functionalities of a qubit (initialisation, readout, single/two qubit gates). The formation of SETs and qubit-dots at different positions, as highlighted by the dashed circles, allow to draw conclusions about the reproducibility of the gate electrodes and the homogeneity of the underlying heterostructure.

2.2 Fabrication process of gate electrodes

Fig. 2 represents the scheme of the fabrication flow applied in this study. Operational blocks, which are not part of this study (e.g. implantation/annealing of ohmic contacts, etching of a mesa, ...) are omitted in this Figure. The incorporated Si/SiGe heterostructure is isolated by an CVD deposited silicon dioxide (SiO_2) layer. Afterwards, a global titanium nitrate (TiN) layer and a sacrificial hardmask are deposited. This is covered with a chemically amplified resist with positive tonality (pCAR) and patterned via electron beam lithography. Using a pCAR the written line ends up as a space in resist after development. Afterwards, the resist mask is transferred

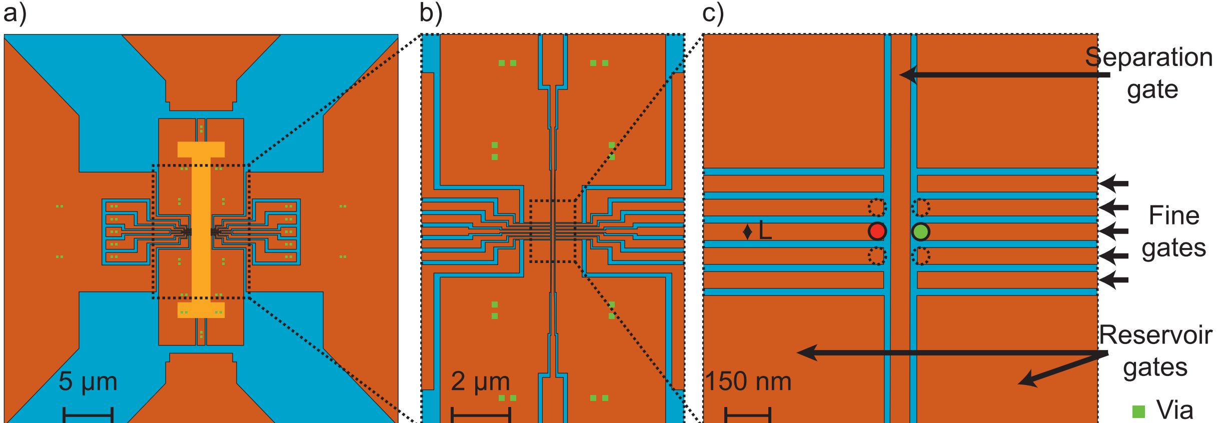


Figure 1. Studied gate layout with different zoom-ins: a) Gate fan-out up to pads designed to be contacted by higher metal layers later in the fabrication process (orange) with vias indicated as green squares. The second, self-aligned gate layer is highlighted in the middle of the image for improved visibility. b) Zoom-in with focus on the large area metal pads in close proximity to the active qubit region. The second gate layer is omitted to reveal the structure of the gates. c) Further zoom-in on the active qubit region, showing the two sets of five fine gates, separated by an additional gate and the four reservoir gates. As an example, a possible position of a SET (red) and that of a qubit dot (green) is shown. These can also be formed as desired at the dashed positions.

into the hardmask, which is subsequently used to pattern the TiN by reactive ion etching (RIE). This etching step introduces an etching bias ΔS, widening the spaces, which is used to adjust the final gate geometry. The deposited SiO_2 layer in the opened section acts as etch stop. The potentially damaged oxide is successfully removed in a further RIE step, followed by the conformal deposition of a second SiO_2 layer. A second TiN layer finally forms the self-aligned second gate layer.

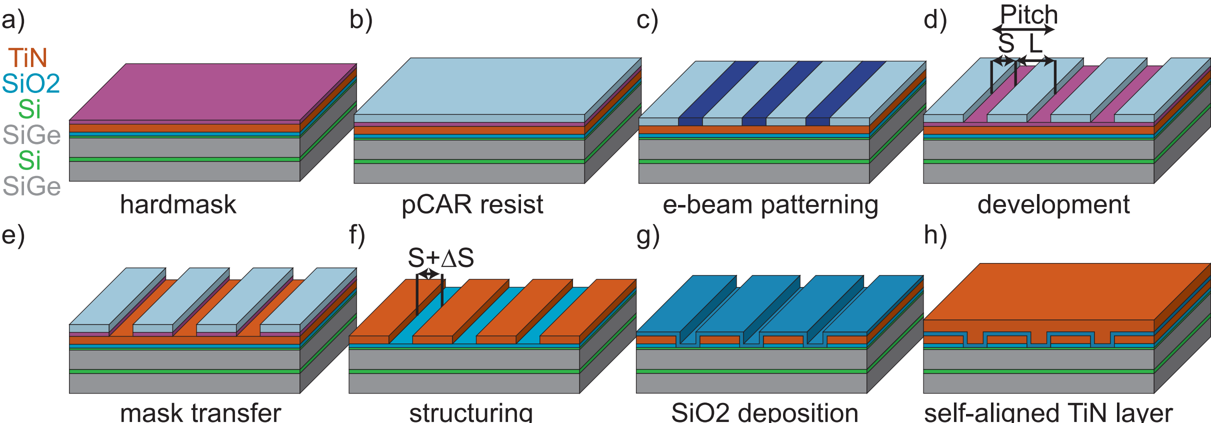


Figure 2. Process flow for device fabrication: a) The heterostrucutre is isolated by depositing SiO_2. On top of this, a global TiN layer G1 is deposited and capped by an sacrificial hardmask. b) The e-beam resist is spin coated on the wafer and c) subsequently patterned. d) After development, the written structure is transferred into the hard mask by reactive ion etching (RIE), f) which is afterwards used to pattern the TiN layer beneath, stopping at the SiO_2 layer. g) The potentially damaged oxide is successfully removed in a further RIE step, followed by the conformal deposition of a second SiO2 layer. h) A second TiN layer G2 finally forms the self-aligned second gate layer.

2.3 Electron beam lithography

For scaling up of the number of qubits a reliable, defect free, high-resolution nanopatterning process is indispensable. Furthermore, the process should be CMOS compatible, which implies among others strict metal

Table 1. Experimental details of the industry compatible nanopatterning process applied in present study, in comparison to the lab scale processing of the gate electrodes, taken from.[41]

	Lab scale e-beam nanostructuring	CMOS compatible e-beam nanostructuring
Patterning process	Additiv (lift-off)	Subtractiv*
e-beam Tool	Gaussian type Raith, 5200	VSB type Vistec SB3050DW
Beam shape	Gaussian	VSB
Resist type	CSAR	CAR, positive
Developer	AR 600-55 (MIBK Basis)	TMAH
Dose $\mu C/cm^2$	800 (extremely high to get good LER)	90
e-gun acceleration voltage kV	100	50
Beam current	150 pA / 3 nA	20 µA
Beam size	<5 nm	4 μm^2
Wafer size	6" Wafer	12" for the investigation of the proces window and stability, 8" for device fabrication

*CMOS compatible

contamination limits, no particle contamination and high throughput. Table 1 summarises the main characteristics of the CMOS industry compatible e-beam lithography system applied in this work. In comparison, an example is given of an e-beam device used in academic research, which was used to fabricate a structure for shuttling single electrons in conveyor-mode.[41]

An important aspect of the lithography process in our work is the application of CMOS compatible chemicals. CMOS compatibility defines weather the process or material is suitable for the standard CMOS semiconductor manufacturing. This definition implies many factors, for example low annealing temperatures to prevent dopant redistribution in the transistors or using specific materials to avoid any defect formation. All materials circulating in the semiconductor CMOS fabs should meet the specification of the maximum allowed element concentrations. Such resist, as well as developer and solvent which meet the CMOS requirements, are also used in this study.

CMOS compatibility also means minimal particle contamination on the wafer surface. The main issue with lab scale lift-off processes is the high particle load and low reproducibility, which are avoided in our work by application of a subtractive etch defined process. Additionally, the e-beam tool (Vistec SB3050DW) is constructed with fully automated wafer handling and alignment systems. This minimises the human contact with the wafers to be exposed, and reduces the time consuming manual handling. Fig. 3 shows fidelity and reproducibility of the 100 nm line/space (L/S) pattern exposed with e-beam over a 300 mm wafer. The measurement was done on structured resist with a CD-SEM tool from Applied Materials (Verity 4i).

The high throughput of our nanopatterning process (compared to other e-beam techniques) is further supported by the variable shaped beam (VSB) principle. The VSB tool has fixed shaped apertures defining the electron beam so that the actual pixel exposed onto a wafer is not a single Gaussian beam of tiny size, but a large beam which fills a particular shape of the exposure pattern (max. shot size 1.6 µm). As a result, the whole wafer with 232 chips containing support structures and 11 quantum devices per chip with an overall exposure area (Fig. 4 a)) of 1358.36 mm^2 requires around 13 hours of e-beam exposure time with the used system. The exposure time for one quantum device with an area of 0.036 mm^2 (see Fig. 4 c)) is just 2 seconds, as calculated from the exposure time for all quantum devices (2 h 23 min) divided by the 232x11 quantum devices per wafer.

Exposure time is also saved due to the application of CAR. CAR contains a photoacid generator (PAG). PAG exposed with electron beam produces protons $H+$. Each generated proton acts as a catalyst and initiates the avalanche of the consequent reactions of deblocking the resin from the large organic radicals, making it soluble in the aqueous developer. Hence CAR requires significantly lower exposure dose and time than nonCAR due to the impact of PAG.[42]

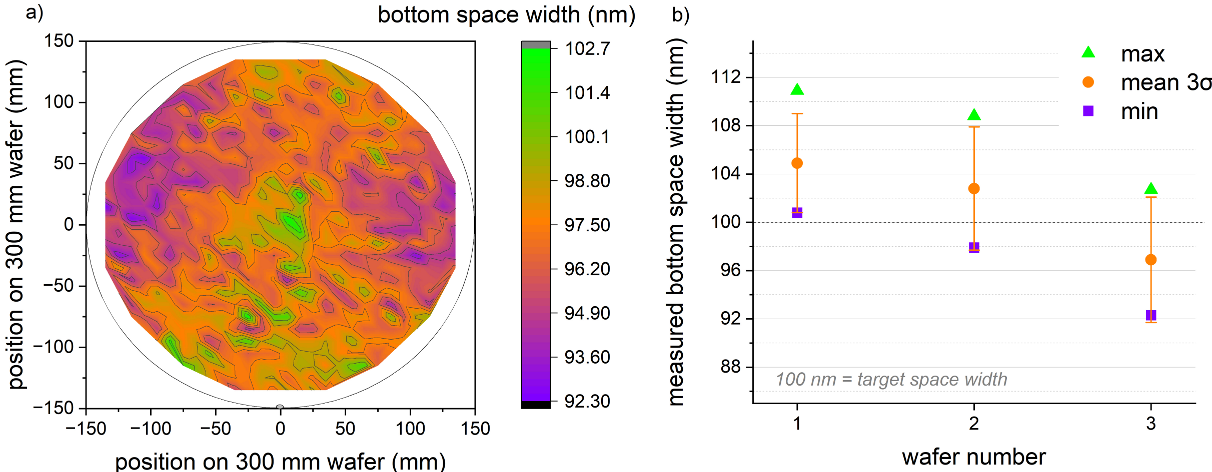


Figure 3. Parameters of the process stability on 300 mm wafers - pattern fidelity, uniformity and wafer to wafer reproducibility. Distribution of a 100 nm trench width with pitch 1/1 a) measured with CD-SEM over the 300 mm wafer b) for three different wafers.

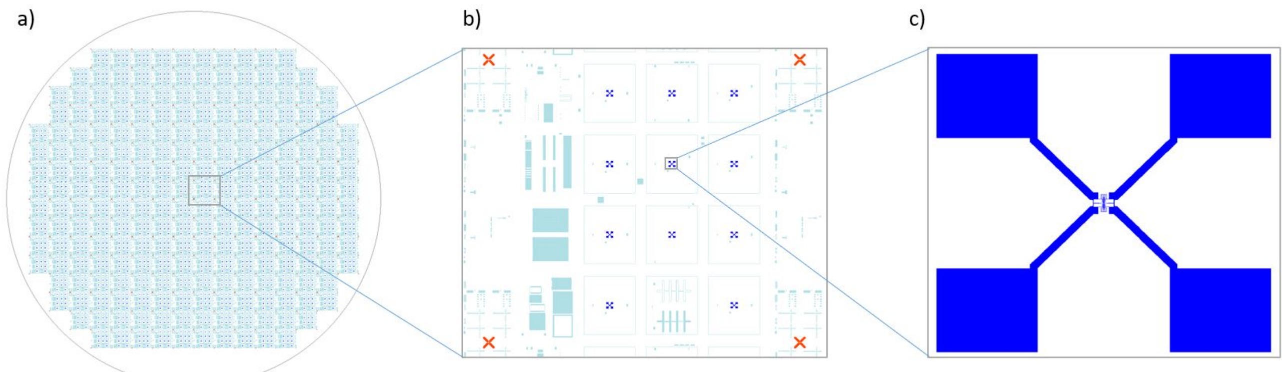


Figure 4. a) Wafermap - arrangement of the 232 chips on a 200 mm wafer, b) layout exposed by e-beam lithography on one chip 1 mm x 1 mm, containing support and device structures, c) layout of the device with four contact pads and gate electrodes in the middle. For further zoom in see Fig. 1.

Besides sufficient throughput and CMOS compatibility, the fabrication of a high numer of gates electrodes for qubit control requires high spatial resolution. This is realized by the specific properties of CAR resists, together with the low resist thickness (ca. 70 nm).

It is also worth mentioning, although not in the scope of this paper, that an optimised resist processing including coating, adhesion, development condition as well as pre and post baking are crucial for the lithographic step.

Last but not least, Proximity Effect Correction (PEC), which is described in detail in reference,[43] plays an essential role in the spatial resolution and pattern quality.

3. RESULTS AND DISCUSSION

3.1 Exploration of the e-beam patterning parameter space

In this part we describe the investigation of the e-beam lithography process window to find a stable and reproducible working window matching the desired gate size around 30-50 nm and smallest possible space in between. For precise illumination of the e-beam lithography process window under optimised conditions, we exposed a

special layout matrix, containing variation of line (bar) and space (trench) - (L/S) dimensions: space varies from 20 nm to 50 nm, and L/S ratio varies from 0.5 to 20, as shown in Fig. 5. The space dimensions after the resist development step were systematically measured by CD-SEM.

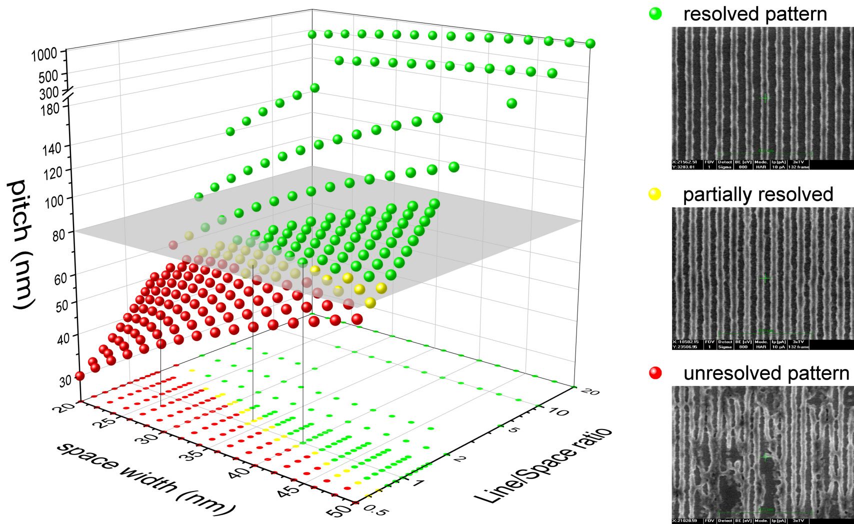


Figure 5. Lithography process window obtained after evaluation of the systematic CD-SEM measurement of the space width (trench width) in the resist mask on the gate stack for various L/S dimensions. The colours represent the quality of the pattern. Grey surface represents the transition region around pitch of 80 nm from unresolved to resolved L/S combination. Three CD-SEM images show examples of L/S pattern in resist.

The results of the CD-SEM measurements are shown in Fig. 5. The pattern quality, given by the color of the data points, is shown in dependence of space width, line/space ratio and pitch. Green represents a stable fidelity for this space - pitch combination even with ±5% exposure dose variation. Yellow indicates that the pattern fidelity is given at least under best dose condition. A deviation of the exposure dose might lead to bridging between resist structures (under exposure) or pattern collapse (over exposure). Red shows that there is no possibility to resolve this space – pitch variation in the chosen resist.
From the matrix in Fig. 5 one can see that at L/S ratio of 1 the space down to a 40 nm can be achieved, resulting in 80 nm pitch. The smallest S in our investigation is 20 nm. But to resolve this space a L/S ratio of at least 3 is required. That means that to ensure proper resolution, the distance between the 20 nm trenches should be at least 60 nm, corresponding to a pitch of 80 nm. When we bring the spaces closer to each other the L/S ratio is getting smaller and pitch decreases. For 30 nm spaces the smallest possible L/S ratio is around 1.7. So the minimal possible bar between the 30 nm trenches is around 50 nm. This is close to the desired gate electrode width. In general one can conclude that the L/S ratios and trench width combinations that result in a pitch of 80 nm or more achieve resolved patterns. (see Fig. 5 the 80 nm pitch is marked with the grey surface cutting the plot).

For proof of pattern stability, Fig. 6 a) shows the fidelity of the resist pattern within a small exposure dose variation. If the pattern is stable when varying the dose in the range of ±5%, a certain amount of deviation in pattern dimensions will not result in a damage of the resist structure. This is visible in Fig. 6, where all points are acquired by CD-SEM measurement on well resolved patterns. The deviation from the target value at nominal dose of 90 $\mu C/cm^2$ is not higher than 2.2 nm for all structure types. It can be concluded that the pattern is stable in case of some process instabilities (for example dose or resist thickness fluctuations). Good pattern fidelities for various pitches also show that the proximity effect correction is well adjusted and correctly compensates electron scattering effects. However, it should be taken into account that at 5 % dose variation the space width may be affected by 4-6 nm.

Fig. 6 b) jumps ahead and shows the same dependence as 6 a), but for the various gate electrode geometries. In this case the size of the resist pattern decreases at higher dose because the line width was measured with

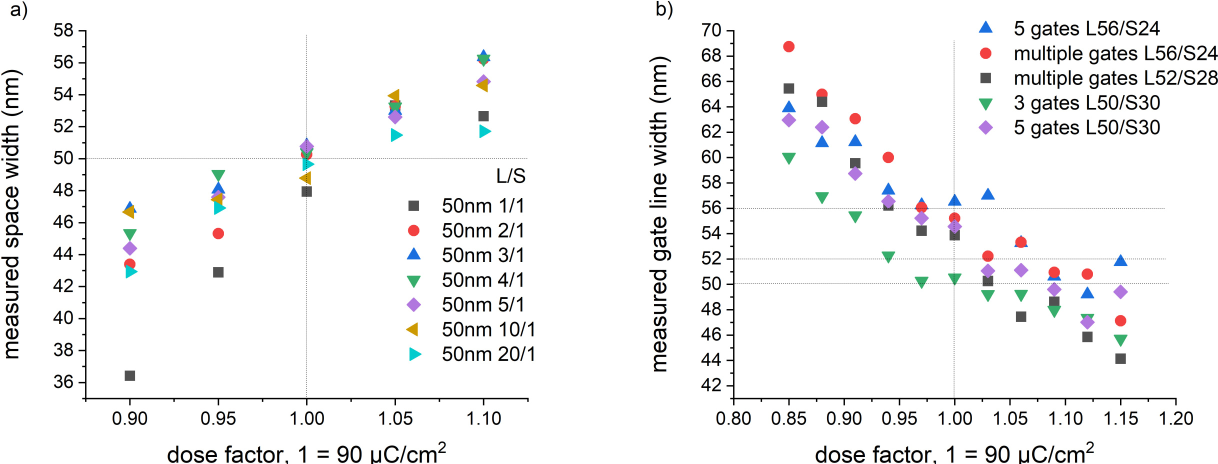


Figure 6. Dependence of the resist pattern width measured with CD-SEM on the applied e-beam exposure dose. a) 50 nm target space width (trench width) with various pitches L/S from 1/1 to 20/1. b) Gate line width (bar width) with various electrode number and L/S dimensions.

CD-SEM unlike plot a), where the space width was measured. So, the width of resist trench increases a) and the width of resist bar decreases b) at higher dose. Nonetheless, also in b) we observe that a variation of the exposure dose in the range of 5 % causes variation of the gate width from 2 to 6 nm, depending on the geometry. Fig. 6 b) reveals that the width of three finger gate electrodes reacts differently than five- and multiple finger. This is the effect of electron scattering from the neighborhood, which is different for these patterns: a three finger structures are subjected to more scattering effects from adjacent exposed areas and hence result in narrower measured resist bars.

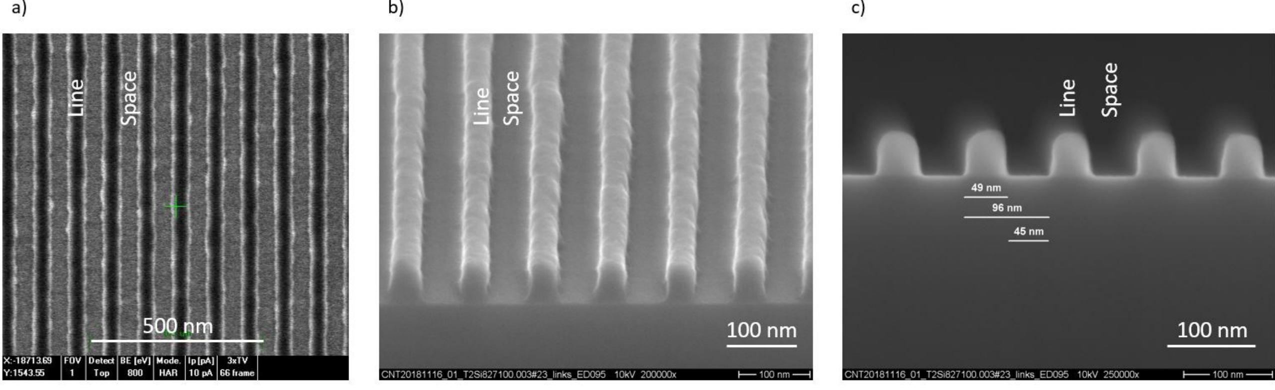


Figure 7. 50 nm L/S pattern in resist with pitch 1/1, a) top view, field of view FOV 1 µm mesured with CD-SEM, b),c) tilted and cross section view correspondingly measured with SEM S5000 from Hitachi.

To verify the reliability of the CD-SEM measurement, the cross section of the 50 nm L/S=1 resist pattern was acquired by SEM. Fig. 7 presents the top view CD-SEM image and SEM images of the 50 nm resist lines cross sections. The cross section SEM measurement (Fig. 7 c)) is in a good agreement with the space values measured by CD-SEM (Fig. 6 a)). From the cross section one can observe that the resist thickness is lower than 70 nm. However, the cross-sectional SEM method is not reliable enough for measuring resist thickness because the resist material is shrunk by the electron beam during SEM measurement. The ellipsometer measurements of the resist thickness over the wafer showed 70 nm thickness, and after resist development the thickness reduced to 67 nm. The dark erosion of the applied resist is around 3 nm. The measured with CD-SEM line edge roughness LER of the resist lines on their bottom revealed 3 nm.

Summarising this part, with application of given resist and optimised processing conditions the target gate width of 50 nm can be reliably achieved at the pitch of at least 80 nm and above. Hence, the minimal possible space between the 50 nm gate electrodes is 30 nm. It should be considered that after etching the gate width slightly reduces due to the etching of the side walls. So the distance between the gates (space) is getting correspondingly wider after etching (see also Fig. 2 ΔS). This aspect will be discussed in the following section in detail.

3.2 Exploration of the optimal gate electrode geometry

In this part the lithography process is further optimised for specific gate electrode geometry with around 50 nm gate electrode width and pitch of 80 nm.

Following parameters of the device geometry were considered during the study: exposure area around the gate electrodes, gate electrodes width and space in between, gap distance from the gate electrode to the separation gate and 3D gate electrode shape after etching.

E-beam exposure area around the gate electrodes It was observed that the quality of the gate electrode structures in resist is vastly affected by the area exposed close to them. Fig 8 shows the comparison of the two layout variants: a) complete area around the electrodes is exposed, b) just to the area which is needed to electrically isolate the electrodes is exposed, so the remaining resist after development can be seen in the SEM images. Exposed area without resist will be free of TiN layer after etching. So, in case of a) the electron backscattering has larger impact as compared to b) due to the larger exposed surface. This is directly reflected on the process window as can be seen in Fig. 8. As described in the previous section green framed images are well resolved and CD is on target, even at slight dose variations. Yellow framed images show the resolved, but unstable at slight dose variation, or partially collapsed pattern. Finally red frame means unresolved region. Consequently, variant b) is selected for further experiments because it provides a wider process window. This does not limit the qunatum device performance as the exposed area would have otherwise been covered by the self-aligned second metal layer. Moreover, by adding unique contacts to the metal plates, a greater flexibility for device tuning can be achieved this way.

Gate electrode width and space in between Fig. 8 presents the influence of the gate electrode width and space in between, which were varied keeping the pitch of 80 nm constant. It is observed, that at constant pitch the width of line and space defines the pattern quality: The thinner the line (gate electrode width) the worse is the pattern, i.e. it collapses. At a line width of L=38 nm (Fig. 8 b)) one can observe a pattern collapse, and at a line width of L=60 nm the pattern looks stable and smooth. The transition point is around dimensions of L=50 nm and S=30 nm. This result is in agreement with the results obtained from the test pattern described in the previous section. Fig. 9 presents the fidelity and wafer to wafer repeatability of the three-gate a), and five-gate b) electrode structure with L/S 50nm/30nm after resist development. This result shows that the gate width deviation after the lithography processing can vary in the range of ± 3 nm.

Distance from the gate electrodes to the separation gate Besides L and S, the gap distance from gate electrode to the separation gate and the separation gate width are essential for the device performance. On the other hand, these values are not restricted by the lithography process and can be chosen more flexibly. One should only take into account that, when electrodes are too close to the separation gate (< 14 nm), they may start to merge. Therefore, the distance was chosen based on simulations to ensure a robust formation of quantum dots (QDs) (see Sec. 4 for the information about the optimal gap distance, as well as about the separation gate width).

Geometry of contact lines and pads Besides gate electrode geometry, we also adapted the shape of contact lines and pads. The common contact geometry applied for the Gaussian type e-beam in the lab scale is not perfectly applicable to the VSB-type e-beam tool. As indicated in the literature (e.g.[41]), the contacts of the lab scale geometry are tilted to various angles. For the VSB tool it is preferable to apply a 45° degree angles or avoid tilted structures at all. This may improve the pattern roughness and reduce the exposure time. The reason is that the VSB-type tool applies apertures with defined geometry, which is either rectangular or with 45° angle. Hence patterns with different than 45° and 90° angles or some curvilinear patterns need to be approximated with rectangular or 45°-triangular shapes, and can cause additional source of line edge roughness and longer exposure

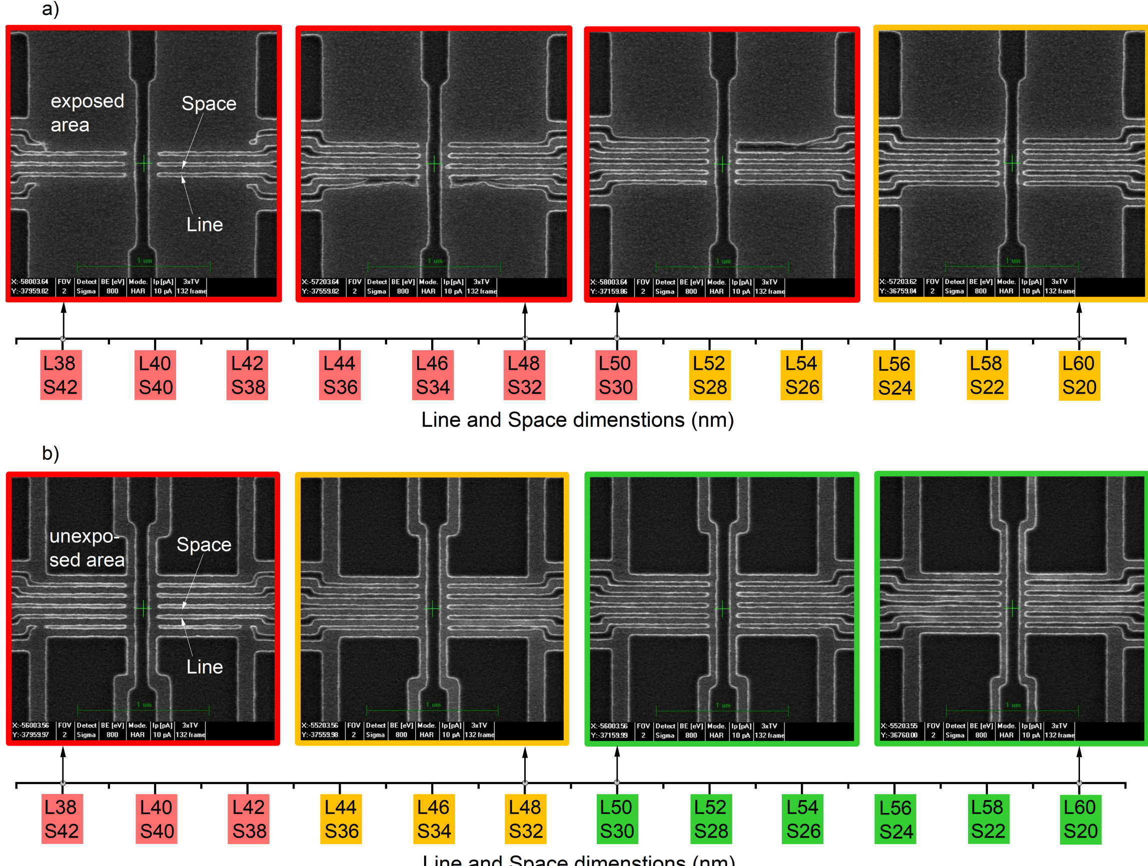


Figure 8. Lithography process window for two different gate electrode geometries and different L/S variations at constant pitch of 80 nm: with a) large and b) reduced exposure area around the electrodes. Green colour represents the optimal L/S ratios providing stable pattern in resist mask, yellow - pattern resolved only at some exposure dose, red - unresolved pattern, pattern collapse.

time. Therefore we applied mostly rectangular shapes for the complete device (see Fig. 1 a) and b)). For the quantum performance this modification is irrelevant, and the fanout can be chosen freely.

Gate electrode 3D shape after etching When speaking about electrode geometry it should be considered that after etching (transfer of the resist mask into the gate electrode material TiN) the lines are getting narrower because they are also etched from the side walls. Hence the spaces are getting wider. In present study, we observed an etch bias ΔS of about 12 nm. That means, that 50 nm line and 30 nm space in the resist transform after etching into a 38 nm line and 42 nm space. After SiO_2 deposition the space width reduces to 32 nm. The widening of the space by etching comes as an accommodation for the subsequent deposition of the self-aligned TiN layer (Fig. 2 (h)): The greater the distance between the electrodes, the better the gap can be filled with TiN. Ideally, this distance should be at least 35 nm to ensure reliable filling (Fig. 2 (f)). In addition, the slightly positive sidewall angle of the etched gate electrodes is used to allow homogeneous TiN deposition without voids that would occur with negative sidewalls.

Fig. 10 a) and d) represents the final optimised gate electrode arrangement after etching into the resist mask. The cross-sectional images b) and e) show the gate width of around 40 nm (ca. 38 nm). The images illustrate also that the gate width fluctuates, which is explained by the line width roughness (LWR) observed in f). To avoid or reduce the LWR it is reasonable to reduce the etch bias or reduce the exposure dose to get wider lines which do not wobble. Nonetheless, the final stack measured by TEM c) shows a good agreement with the desired structure. Fig. 10 c) corresponds to the schematic image in Fig. 2 h). The width of the fine gates in G_1 is

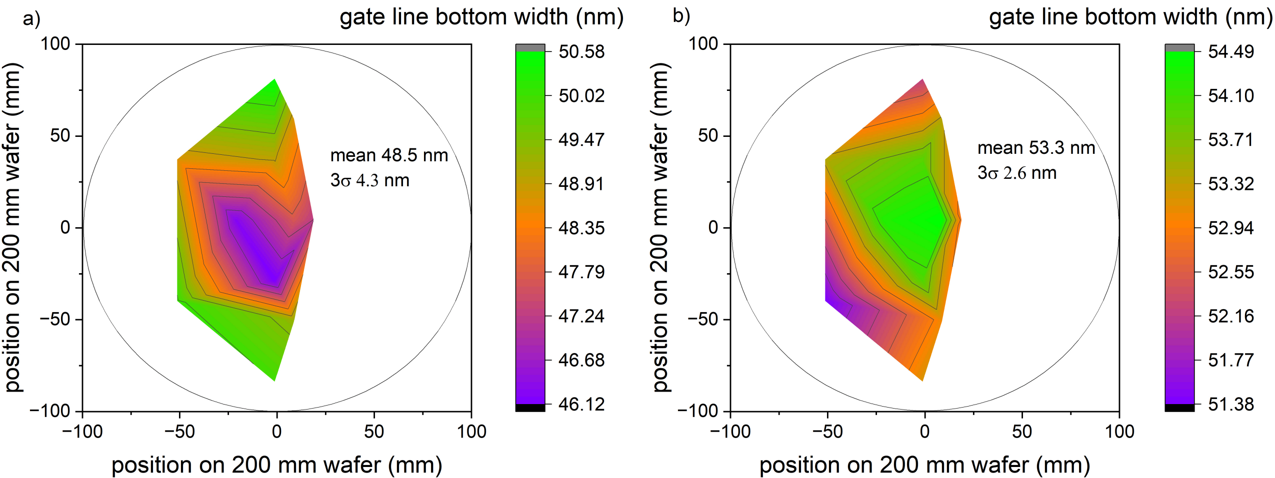


Figure 9. Gate pattern uniformity after development of the e-beam exposed resist, measured with CD-SEM on the a) 3 gate electrode structure, and b) on the 5 gate electrode structure.

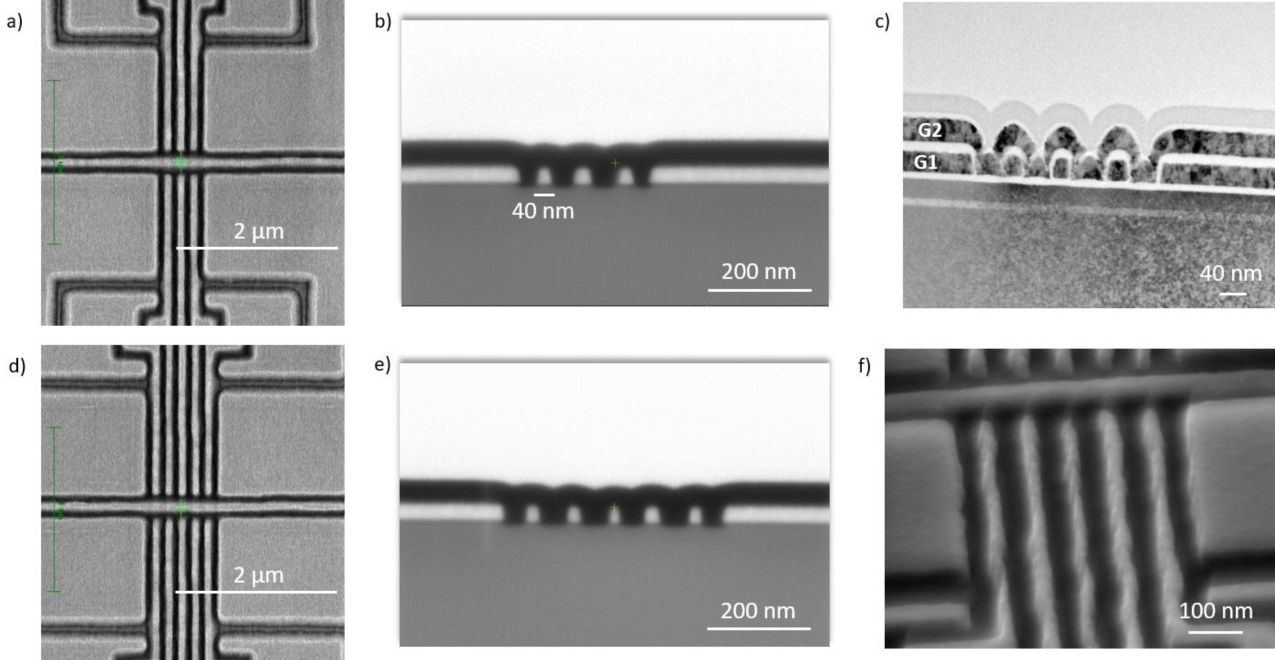


Figure 10. Images of the final gate electrode devices with optimised geometry (L=50 nm S=30 nm post litho, and L=38 nm post etch). All images were acquired after etching step. Etching was performed at Infineon FAB in Dresden. a), d) CD-SEM images of the final G1 structure after etching with three and five gates correspondingly. b), e) FIB cross section SEM images of the G1 electrodes after etching. c) TEM image of the complete gate structure according to the Fig. 2 h). f) tilted view of the etched five gate electrode arrangement.

slightly below 40 nm, and G_2 shows a good gap filling without voids due to the positive side wall angles of the G_1.

To summarise, the ideal gate electrode geometry after the lithography step, under the condition of reduced surrounding exposure area, is L=50 nm and above, S=30 nm and below, at constant pitch of 80 nm (see Fig. 8 b). Larger pitches between the 50 nm electrodes are also possible. The space of 30 nm is widened by the etching process, resulting in $w_{\mathrm{G1}} = 38\,\mathrm{nm}$ gate electrode width in TiN layer for the structured layer. After deposition of

5 nm SiO_2, the remaining space between the G1 electrodes is $w_{\mathrm{G2}} = 32\,\mathrm{nm}$, which is then subsequently filled with TiN (see Fig. 2 g) and h)). In the sum L, S and SiO_2 layer result in a 80 nm pitch. These geometrical parameters, along with the positive taper angle and the line width variations are studied in terms of corresponding qubit performance by COMSOL simulation, as described in the next section.

4. SIMULATIONS

For the gate layout in Fig. 1, we estimate suitable geometrical parameters such as gate pitch, widths, and spacing. These considerations are based on the general behavior of QDs in Si/SiGe, as described in Sec. 1. Starting from the estimated device parameters in the previous section, we establish geometric boundaries within which the device could be reliably fabricated. In this section, we adjust the simulation using the gate electrode geometry after the etching step described in Sec. 3.2 and verify the functionality of the gate pattern in terms of QD confinement, capatitive cross-coupling and tunnel-barriers to reservoirs: The gate pattern is designed to form a QD on one side, which can be tunnel-coupled to two electron reservoirs. On the other side, a single electron transistor (SET) shall be formed, which is capacitively coupled to the QD and thus operates as a proximate charge detector of the QD. The goal is to maximize the sensitivity of the SET to the QD and to control the QD filling down to the zero to few electron regime, which requires narrow tunnel barriers and a sufficiently strong confinement. As the gate pattern is designed to be symmetric with respect to the separation gate, the location of QD and SET can be swapped. By swapping the SET and QD functionality, the symmetry of the device in terms of local potential disorder or non-ideal electrode shapes can be tested.

Our finite-element simulation-model (COMSOL Multiphysics) of the device electrostatics is based on the e-beam lithography results, target etch bias and gate oxide (GOX) thickness. It includes the structured gate layer G_1 (as shown in Fig. 1 c)), as well as a second, self-aligned gate layer G_2 (indicated in Fig. 1 a)). The schematic cross-section of the device is shown in Fig. 11 a). A structured gate width of $w_{\mathrm{G1}} = 38\,\mathrm{nm}$ and a $w_{\mathrm{GOX}} = 5\,\mathrm{nm}$ thick conformally deposited oxide layer for electrical isolation between G_1 and G_2 are taken into account. We assume the patterning pitch of $P = 80\,\mathrm{nm}$ for the structured layer. This results in a gate width of the self-align gates, interleaving the G_1 finger gates, of $w_{\mathrm{G2}} = P - w_{\mathrm{G1}} - 2 \cdot w_{\mathrm{GOX}} = 32\,\mathrm{nm}$. The Si/SiGe heterostructure is modelled with a homogeneous dielectric constant $\epsilon_{\mathrm{Si}} = 13$ and the Si quantum well is implemented as charge density plane at a depth of $z_{\mathrm{2DEG}} = 35\,\mathrm{nm}$ below the semiconductor/oxide interface at $z = 0$.

We use Poisson equation to numerically solve self-consistently for the electron potential $V(x, y)$ and electron density $\rho(x, y)$ in the 2DEG using the semi-classical Thomas-Fermi approximation.[44] In order to account for the filling difference between the sensor QD of the SET, which is usually operated in a multi electron regime, and the qubit QD, operated in the single-electron regime, the electron density for the QD region (here right-hand side of the separation gate) is set to zero only in the QD region.

We use the following strategy to iteratively determine the voltages applied to a given gate-electrodes pattern within the simulation: Before tuning the QDs, we increase the voltages on the four reservoir plates and the two outermost fine gates of each side until a sufficiently high electron density is achieved. Then, we tune the gate voltages applied to the SET gates (highlighted in green in Fig. 11 b)) until the tunnel barriers of the SET to the reservoirs reaches a targeted potential shape. For the QD side (highlighted in red in Fig. 11 b)), we coarsely pre-tune the shape of the desired QD. Then, we iterate the calculation of first and second orbital by Schrödinger equation and fine-tuning of voltages until only the lowest orbital state is found below the Fermi-energy. All voltages refer to the middle of the band-gap of silicon[45] and all simulation parameters are summarized in Table 2.

After tuning the voltages by the simulation, the calculated electrostatic potential $V(x, y)$ and electron density $\rho(x, y)$ is obtained as shown in Fig. 11 b). The color plots are overlaid with the G_1 gate layout while G_2 is omitted but globally placed on top of the displayed device section. We are able to form two QDs in the potential, separated by the gate electrode. From the preliminary simulations, we find a separation gate width w_{Gsep} which is far away from fabricational boundaries. Therefore, the width of this gates is picked based on the following considerations: The width of the separation gate G_{sep} controls the capacitive coupling of the QD and the sensing QD of the SET. It also separates opposing electron reservoirs. We chose a well-balanced width of 60 nm, which allows to separate the electron reservoirs at $V_{\mathrm{sep}} = 0\,\mathrm{V}$ and thus simplifies the operation of the device. On the

Table 2. Simulation parameters used in the presented COMSOL model.

Description	Variable	Si/SiGe
Electrode width	w_{G1}	38 nm
Self-aligned width	w_{G2}	32 nm
Thickness oxide	w_{GOX}	5 nm
Pitch	P	80 nm
Separation gate width	w_{Gsep}	60 nm
Distance tip	d_{tip}	70 nm
Effective electron mass	m^*	0.19 m_e^*
Valley splitting	E_{VS}	70 μeV
Fermi energy	E_{F}	585 meV
Permittivity of heterostructure	ϵ_r	13
Permittivity of oxide	$\epsilon_{\mathrm{oxide}}$	4
Min. element Size (ES)	d_{min}	5 nm
Max. ES	d_{max}	60 nm
Min. 2DEG ES	$d_{\mathrm{min,2DEG}}$	3 nm
Max. 2DEG ES	$d_{\mathrm{max,2DEG}}$	10 nm
Depth of 2DEG	$\mathrm{z}_{\mathrm{2DEG}}$	35 nm

one hand, a larger gate width would reduce the capacitive coupling between QD and SET, and thus lowers the sensitivity of the charge detector. On the other hand, a smaller separation gate width would require more negative voltages to isolate the reservoirs from each other, resulting a higher potential difference to the surrounding gates and thus an increased chance of current leakage across gates.

Each dot is controlled by three of the five finger gates indicated in green (red) on the left (right) side of the separating gate. The distance between the tip of the fine gates and the separation gate was another input parameter not limited by fabrication. We found that $d_{\mathrm{tip}} \approx 70\,\mathrm{nm}$ is suitable, which is a bit larger than the typical diameter of a QD. By reducing d_{tip}, on the one hand, the accumulation control of the G_2 gate is more screened, which must be compensated by a more positive plunger gate voltage. This might conflict with upper bound of the plunger gate voltage around the accumulation threshold, since charge accumulation underneath the long plunger gate leading to an unintentionally deformed QD must be avoided. On the other hand, making d_{tip} too large would require strong negative voltages on the barrier gates to form the tunnel barriers as well as resulting in broader barriers. The left QD is visible in the electron density, while the right QD shows no accumulated electrons. This is consistent with the previously discussed goal of using the left QD as SET, sensing the charge occupation of the right QD, which operates in single electron mode.

We calculate the probability densities $|\Psi(x, y)|^2$ of the first and second orbital state (OS), by solving the Schrödinger equation in the potential of the right QD (shown in top (first OS) and bottom (second OS) panel of Fig. 11 c)). We observe the occurrence of a second maximum along the y-axis for the second orbital state. This orientation of the second OS is expected due to the elongated shape of the dot. We calculate the energy difference of the first and second eigenstate (denoted as orbital splitting) $\Delta E_{\mathrm{OS}} = E_{\mathrm{OS},2} - E_{\mathrm{OS},1} = 1.0\,\mathrm{meV}$. Thus, the simulated QD-confinement is within the typical range of 1-1.5 meV,[46] which indicates that it is robust with respect to disorder-induced separation into several dots.

2-dimensional line cuts of the potential and charge carrier density are given in Fig. 11 d), corresponding to the SET (left/ green) and single electron dot (right/ red) following the path indicated by the dashed lines in Fig. 11 b). This path was calculated using *Dijkstra's algorithm*, minimizing a cost function combining the semiclassical Wentzel-Kramers-Brillouin approximation and a path-of-lowest-potential (see[47] for more information). We can form a set of two barriers for each QD, given by the regions above E_{F} (indicated by the dashed line), allowing localization of electrons inside the QDs. The position of the reservoirs close to the QD, as indicated by the

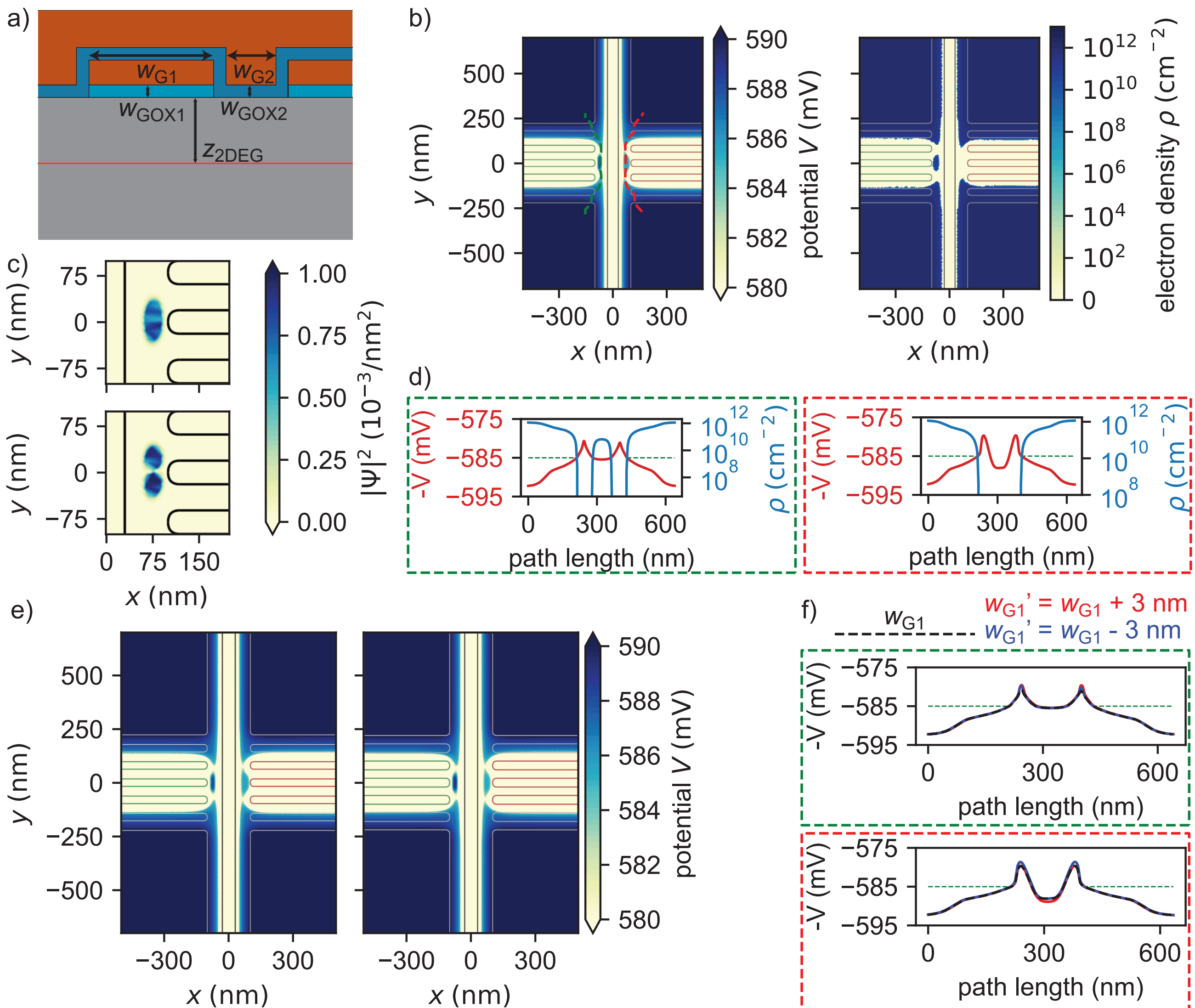


Figure 11. Simulation results: a) Schematic cross-section of the gate stack implemented into our simulation. b) Electrostatic potential $V(x, y)$ formed in the quantum well (left) and electron density $\rho(x, y)$ in the quantum well (right), overlaid with the G_1 gate layout (G_2 is not shown). c) Probability density $|\Psi(x, y)|^2$ of the first (top) and second (bottom) orbital state calculated by solving the Schrödinger equation on the simulated quantum well potential. d) Line cuts through the left (green) / right (red) QD along the path indicated by the colored dashed lines in panel b, respectively. e) Electrostatic potential formed in the quantum well for variations of the structured gate width of -3 nm (left) / +3 nm (right), after repeated gate voltage tuning. f) line cuts corresponding to the paths shown in b) for the original layout (dashed lines) as well as for the two studied variations.

rapidly increasing ρ on the other side of the barrier with regards to the QD, indicates a good capacitive coupling. Furthermore, we achieve electron densities up to $1 \times 10^{12}\,\mathrm{cm}^{-2}$, indicating sufficient 2DEG accumulate far beyond the metal-insulator-transition.

In a next step, we consider the line width variation in the size of $3\sigma_w = 3\,\mathrm{nm}$ of the patterning process, as shown in Fig. 9 b), and study its impact on the device. In this way, we ensure that the functionality of the device as a spin qubit is not compromised by unavoidable variations in the fabrication process. Therefore, we assume the positional accuracy of the e-beam to be sufficiently high and therefore keep the patterning pitch fixed to $P = 80\,\mathrm{nm}$. An increase/decrease in the width of the gate electrodes in the structured layer $w'_{G1} = w_{G1} \pm 3\,\mathrm{nm}$ therefore results in a decrease/increase in the width of the self-aligned interleaved gates $w'_{G2} = w_{G2} \mp 3\,\mathrm{nm}$. We

repeat the simulation for the two extreme values to estimate the worst case. After applying these changes to the simulated gate geometry, the described voltage tuning process is repeated. The potential $V(x, y)$ obtained for the smaller (larger) structured gate width w'_{G1} is shown on the left (right) in Fig. 11 d). We were able to adjust the voltages for both directions of variation to recover the qubit/SET QD formation. This required voltage changes in the order of $\mathcal{O}(100\,\text{mV})$ on the sets of three gates forming the QDs. This is consistent with voltage fluctuations observed in experimental studies of industrially fabricated quantum structures.[48] We obtain slightly increased orbital splittings of 1.1 meV and 1.2 meV and single-electron occupation, enabling well isolated operation of the qubit QD in the first OS.

In Fig. 11 f), we compare the potentials for the original gate electrode geometry and the two variants, obtained along the line cuts for the SET QD (green) and the qubit QD (red) as shown in Fig. 11 b). The original potential is plotted in black and dashed, while the potentials for smaller (larger) structured gate width are added in blue (red). Except for minor variations in the shape of the barriers, the applied voltage changes allowed the original potential for both variations to be reproduced.

Using simulations, we have successfully demonstrated shaping of a single-electron QD with sufficient confinement and a proximate charge-sensor tunnel-coupled to well-accumulated reservoirs within the fabricational boundaries and tolerated geometrical variations. These results affirm the viability and operability of the targeted device, and thus the fabrication process, according to our intended specifications.

5. CONCLUSION

In this work we setup a CMOS industry compatible fabrication process of gate electrodes for Si/SiGe heterostructure spin qubits based on e-beam lithography. Our work is dedicated to the e-beam lithography process and the interplay of pattern fidelity with gate electrode geometry. We developed a process providing high yield, uniformity, reproducibility and throughput of the resulting nanostructured electrodes. The following parameters of the qubit device geometry were considered during the study: exposure area around the gate electrodes, gate electrodes width and space in between, gap distance from the gate electrode to the separation gate and 3D shape of the gate electrode after etching. We observed that with given resist, reduced surrounding exposure area, and optimised lithography processing conditions a gate width of 50 nm can be reliably achieved for a pitch of 80 nm or more. We characterized the stability of the process and obtained a line width variation of $3\sigma_w = 3\,\text{nm}$. Process details of the fabrication of the second, self-aligned gate layer resulted in line-to-gate transfer bias of 12 nm, when etching the resist pattern into the TiN gate metal, as well as positive flanks. Taking into account the 5 nm SiO_2 isolation layer, we obtain a final gate electrode width of $w_{G1} = 38\,\text{nm}$ for the structured layer and $w_{G2} = 32\,\text{nm}$ for the self-aligned interleaved gates on a pitch of $P = 80\,\text{nm}$. We verified the operability of the device using a COMSOL Multiphysics model based on the obtained geometric parameters and successfully formed a qubit quantum dot (QD) and a charge-sensitive single-electron transistor (SET QD). The qubit QD could be brought to single-electron occupancy, showing a sufficiently large orbital splitting of 1 meV, and the tunnel barriers indicate favorable transport properties. With the simulations, we were able to show that the fabricated devices can be operated as intended, even when process variations were taken into account. This proves the applicability of our method for large-scale, high-throughput patterning of quantum devices.

Acknowledgements

The presented work has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 951852, the FET Flagship on Quantum Technologies (FETFLAG-05-2020) and the Federal Ministry of Education and Research BMBF under the Project QUASAR with Grant Number 13N15652.

REFERENCES

[1] Daniel Loss and David P. DiVincenzo. Quantum computation with quantum dots. *Phys. Rev. A*, 57:120–126, Jan 1998.

[2] Gabriel Popkin. Quest for qubits. *Science*, 354(6316):1090–1093, 2016.

[3] Floris A. Zwanenburg, Andrew S. Dzurak, Andrea Morello, Michelle Y. Simmons, Lloyd C. L. Hollenberg, Gerhard Klimeck, Sven Rogge, Susan N. Coppersmith, and Mark A. Eriksson. Silicon quantum electronics. *Rev. Mod. Phys.*, 85:961–1019, Jul 2013.

[4] Austin G. Fowler, Matteo Mariantoni, John M. Martinis, and Andrew N. Cleland. Surface codes: Towards practical large-scale quantum computation. *Phys. Rev. A*, 86:032324, Sep 2012.

[5] Craig Gidney and Martin Ekerå. How to factor 2048 bit RSA integers in 8 hours using 20 million noisy qubits. *Quantum*, 5:433, April 2021.

[6] Stephan G. J. Philips, Mateusz T. Mądzik, Sergey V. Amitonov, Sander L. de Snoo, Maximilian Russ, Nima Kalhor, Christian Volk, William I. L. Lawrie, Delphine Brousse, Larysa Tryputen, Brian Paquelet Wuetz, Amir Sammak, Menno Veldhorst, Giordano Scappucci, and Lieven M. K. Vandersypen. Universal control of a six-qubit quantum processor in silicon. *Nature*, 609(7929):919–924, Sep 2022.

[7] L. C. L. Hollenberg, A. D. Greentree, A. G. Fowler, and C. J. Wellard. Two-dimensional architectures for donor-based quantum computing. *Physical Review B*, 74(4):045311, July 2006.

[8] M. Veldhorst, H. G. J. Eenink, C. H. Yang, and A. S. Dzurak. Silicon CMOS architecture for a spin-based quantum computer. *Nature Communications*, 8(1):1766, December 2017.

[9] Ruoyu Li, Luca Petit, David P. Franke, Juan Pablo Dehollain, Jonas Helsen, Mark Steudtner, Nicole K. Thomas, Zachary R. Yoscovits, Kanwal J. Singh, Stephanie Wehner, Lieven M. K. Vandersypen, James S. Clarke, and Menno Veldhorst. A crossbar network for silicon quantum dot qubits. *Science Advances*, 4(7):eaar3960, July 2018.

[10] Jelmer M. Boter, Juan P. Dehollain, Jeroen P.G. van Dijk, Yuanxing Xu, Toivo Hensgens, Richard Versluis, Henricus W.L. Naus, James S. Clarke, Menno Veldhorst, Fabio Sebastiano, and Lieven M.K. Vandersypen. Spiderweb array: A sparse spin-qubit array. *Phys. Rev. Appl.*, 18:024053, Aug 2022.

[11] Veit Langrock, Jan A. Krzywda, Niels Focke, Inga Seidler, Lars R. Schreiber, and Łukasz Cywiński. Blueprint of a scalable spin qubit shuttle device for coherent mid-range qubit transfer in disordered si/sige/sio$_2$. *PRX Quantum*, 4:020305, Apr 2023.

[12] Arne Hollmann, Daniel Jirovec, Maciej Kucharski, Dietmar Kissinger, Gunter Fischer, and Lars R. Schreiber. 30 GHz-voltage controlled oscillator operating at 4 K. *Review of Scientific Instruments*, 89(11), 11 2018. 114701.

[13] L. M. K. Vandersypen, H. Bluhm, J. S. Clarke, A. S. Dzurak, R. Ishihara, A. Morello, D. J. Reilly, L. R. Schreiber, and M. Veldhorst. Interfacing spin qubits in quantum dots and donors—hot, dense, and coherent. *npj Quantum Information*, 3, 09 2017.

[14] Lars R Schreiber and Hendrik Bluhm. Toward a silicon-based quantum computer. *Science*, 359(6374):393–394, 2018.

[15] Tom Struck, Arne Hollmann, Floyd Schauer, Olexiy Fedorets, Andreas Schmidbauer, Kentarou Sawano, Helge Riemann, Nikolay V. Abrosimov, Łukasz Cywiński, Dominique Bougeard, and Lars R. Schreiber. Low-frequency spin qubit energy splitting noise in highly purified 28 Si/SiGe. *npj Quantum Information*, 6(1):1–7, May 2020.

[16] Jun Yoneda, Kenta Takeda, Tomohiro Otsuka, Takashi Nakajima, Matthieu R. Delbecq, Giles Allison, Takumu Honda, Tetsuo Kodera, Shunri Oda, Yusuke Hoshi, Noritaka Usami, Kohei M. Itoh, and Seigo Tarucha. A quantum-dot spin qubit with coherence limited by charge noise and fidelity higher than 99.9%. *Nature Nanotechnology*, 13(2):102–106, February 2018.

[17] D. M. Zajac, A. J. Sigillito, M. Russ, F. Borjans, J. M. Taylor, G. Burkard, and J. R. Petta. Resonantly driven cnot gate for electron spins. *Science*, 359(6374):439–442, 2018.

[18] T. F. Watson, S. G. J. Philips, E. Kawakami, D. R. Ward, P. Scarlino, M. Veldhorst, D. E. Savage, M. G. Lagally, Mark Friesen, S. N. Coppersmith, M. A. Eriksson, and L. M. K. Vandersypen. A programmable two-qubit quantum processor in silicon. *Nature*, 555(7698):633–637, 2018.

[19] X. Xue, T. F. Watson, J. Helsen, D. R. Ward, D. E. Savage, M. G. Lagally, S. N. Coppersmith, M. A. Eriksson, S. Wehner, and L. M. K. Vandersypen. Benchmarking gate fidelities in a Si/SiGe two-qubit device. *Phys. Rev. X*, 9:021011, Apr 2019.

[20] Xiao Xue, Maximilian Russ, Nodar Samkharadze, Brennan Undseth, Amir Sammak, Giordano Scappucci, and Lieven M K Vandersypen. Quantum logic with spin qubits crossing the surface code threshold. *Nature*, 601(7893):343–347, January 2022.

[21] Akito Noiri, Kenta Takeda, Takashi Nakajima, Takashi Kobayashi, Amir Sammak, Giordano Scappucci, and Seigo Tarucha. A shuttling-based two-qubit logic gate for linking distant silicon quantum processors. *Nature Communications*, 13(1):5740, 2022.

[22] Elliot J. Connors, JJ Nelson, and John M. Nichol. Rapid high-fidelity spin-state readout in Si/Si-Ge quantum dots via rf reflectometry. *Phys. Rev. Applied*, 13:024019, Feb 2020.

[23] Akito Noiri, Kenta Takeda, Jun Yoneda, Takashi Nakajima, Tetsuo Kodera, and Seigo Tarucha. Radio-frequency-detected fast charge sensing in undoped silicon quantum dots. *Nano Lett.*, 20(2):947–952, 2020.

[24] Tom Struck, Javed Lindner, Arne Hollmann, Floyd Schauer, Andreas Schmidbauer, Dominique Bougeard, and Lars R. Schreiber. Robust and fast post-processing of single-shot spin qubit detection events with a neural network. *Sci Rep*, 11(1):16203, Aug 2021.

[25] Lars R Schreiber and Hendrik Bluhm. Silicon comes back. *Nature nanotechnology*, 9(12):966–968, 2014.

[26] Hendrik Bluhm and Lars R Schreiber. Semiconductor spin qubits—a scalable platform for quantum computing? In *2019 IEEE International Symposium on Circuits and Systems (ISCAS)*, pages 1–5. IEEE, 2019.

[27] R. Pillarisetty, N. Thomas, H.C. George, K. Singh, J. Roberts, L. Lampert, P. Amin, T.F. Watson, G. Zheng, J. Torres, M. Metz, R. Kotlyar, P. Keys, J.M. Boter, J.P. Dehollain, G. Droulers, G. Eenink, R. Li, L. Massa, D. Sabbagh, N. Samkharadze, C. Volk, B. P. Wuetz, A.-M. Zwerver, M. Veldhorst, G. Scappucci, L.M.K. Vandersypen, and J.S. Clarke. Qubit Device Integration Using Advanced Semiconductor Manufacturing Process Technology. *Technical Digest - International Electron Devices Meeting, IEDM*, 2018-Decem:6.3.1–6.3.4, 2019.

[28] A. M.J. Zwerver, T. Krähenmann, T. F. Watson, L. Lampert, H. C. George, R. Pillarisetty, S. A. Bojarski, P. Amin, S. V. Amitonov, J. M. Boter, R. Caudillo, D. Corras-Serrano, J. P. Dehollain, G. Droulers, E. M. Henry, R. Kotlyar, M. Lodari, F. Lüthi, D. J. Michalak, B. K. Mueller, S. Neyens, J. Roberts, N. Samkharadze, G. Zheng, O. K. Zietz, G. Scappucci, M. Veldhorst, L. M.K. Vandersypen, and J. S. Clarke. Qubits made by advanced semiconductor manufacturing. *Nat Electron 5*, pages 184–190, 2022.

[29] B. Govoreanu, S. Kubicek, J. Jussot, B.T. Chan, N.I. Dumoulin-Stuyck, F.A. Mohiyaddin, R. Li, G. Simion, Ts. Ivanov, D. Mocuta, J. Lee, and I.P. Radu. Moving spins from lab to fab: A silicon-based platform for quantum computing device technologies. *2019 Silicon Nanoelectronics Workshop (SNW)*, pages 1–2, 2019.

[30] R. Li, N. I. Dumoulin Stuyck, S. Kubicek, J. Jussot, B. T. Chan, F. A. Mohiyaddin, A. Elsayed, M. Shehata, G. Simion, C. Godfrin, Y. Canvel, Ts. Ivanov, L. Goux, B. Govoreanu, and I. P. Radu. A flexible 300 mm integrated Si MOS platform for electron- And hole-spin qubits exploration. *Technical Digest - International Electron Devices Meeting, IEDM*, 2020-Decem:38.3.1–38.3.4, 2020.

[31] Fabio Ansaloni, Christian Volk, Anasua Chatterjee, and Ferdinand Kuemmeth. Characterization of top-gated si/sige devices for spin qubit applications. In *2019 Silicon Nanoelectronics Workshop (SNW)*, pages 1–2. IEEE, 2019.

[32] X Mi, JV Cady, DM Zajac, J Stehlik, LF Edge, and Jason R Petta. Circuit quantum electrodynamics architecture for gate-defined quantum dots in silicon. *Applied Physics Letters*, 110(4):043502, 2017.

[33] Tom Struck, Arne Hollmann, Floyd Schauer, Olexiy Fedorets, Andreas Schmidbauer, Kentarou Sawano, Helge Riemann, Nikolay V Abrosimov, Łukasz Cywiński, Dominique Bougeard, et al. Low-frequency spin qubit energy splitting noise in highly purified 28si/sige. *npj Quantum Information*, 6(1):1–7, 2020.

[34] Kenta Takeda, Jun Kamioka, Tomohiro Otsuka, Jun Yoneda, Takashi Nakajima, Matthieu R Delbecq, Shinichi Amaha, Giles Allison, Tetsuo Kodera, Shunri Oda, et al. A fault-tolerant addressable spin qubit in a natural silicon quantum dot. *Science advances*, 2(8):e1600694, 2016.

[35] Kenta Takeda, Akito Noiri, Takashi Nakajima, Jun Yoneda, Takashi Kobayashi, and Seigo Tarucha. Quantum tomography of an entangled three-qubit state in silicon. *Nature Nanotechnology*, 16(9):965–969, 2021.

[36] Daniel R Ward, Dohun Kim, Donald E Savage, Max G Lagally, Ryan H Foote, Mark Friesen, Susan N Coppersmith, and Mark A Eriksson. State-conditional coherent charge qubit oscillations in a si/sige quadruple quantum dot. *npj Quantum Information*, 2(1):1–6, 2016.

[37] DM Zajac, TM Hazard, X Mi, K Wang, and Jason R Petta. A reconfigurable gate architecture for si/sige quantum dots. *Applied Physics Letters*, 106(22):223507, 2015.

[38] Ruichen Zhao, Tuomo Tanttu, Kuan Yen Tan, Bas Hensen, Kok Wai Chan, JCC Hwang, RCC Leon, Chi Herng Yang, Will Gilbert, FE Hudson, et al. Single-spin qubits in isotopically enriched silicon at low magnetic field. *Nature communications*, 10(1):1–9, 2019.

[39] Aaron J Weinstein, Matthew D Reed, Aaron M Jones, Reed W Andrews, David Barnes, Jacob Z Blumoff, Larken E Euliss, Kevin Eng, Bryan Fong, Sieu D Ha, et al. Universal logic with encoded spin qubits in silicon. *arXiv preprint arXiv:2202.03605*, 2022.

[40] C. H. Huang, W.and Yang, K. W. Chan, T. Tanttu, B. Hensen, R. C. C. Leon, M. A. Fogarty, J. C. C. Hwang, F. E. Hudson, K. M. Itoh, A. Morello, A. Laucht, and A. S. Dzurak. Fidelity benchmarks for two-qubit gates in silicon. *Nature*, 569:532–536, 2019.

[41] Inga Seidler, Tom Struck, Ran Xue, Niels Focke, Stefan Trellenkamp, Hendrik Bluhm, and Lars R. Schreiber. Conveyor-mode single-electron shuttling in Si/SiGe for a scalable quantum computing architecture. *npj Quantum Information*, 8(1):100, August 2022.

[42] Hiroshi Ito. Chemically amplified resists: past, present, and future. *Proceedings of the SPIE*, 3678:2–12, 1999.

[43] Varvara Brackmann, Michael Friedrich, Clyde Browning, Norbert Hanisch, and Benjamin Uhlig. Influence of the dose assignment and fracturing type on patterns exposed by a variable shaped e-beam writer: simulation vs experiment. In Uwe F.W. Behringer and Jo Finders, editors, *35th European Mask and Lithography Conference (EMLC 2019)*, volume 11177, page 1117713. International Society for Optics and Photonics, SPIE, 2019.

[44] A. Siddiki and Rolf Gerhardts. Thomas-Fermi-Poisson theory of screening for laterally confined and unconfined two-dimensional electron systems in strong magnetic fields. *Physical Review B - Condensed Matter and Materials Physics*, 68(12):1–12, 2003.

[45] W. Bludau, A. Onton, and W. Heinke. Temperature dependence of the band gap of silicon. *Journal of Applied Physics*, 45(4):1846–1848, 1974.

[46] Arne Hollmann, Tom Struck, Veit Langrock, Andreas Schmidbauer, Floyd Schauer, Tim Leonhardt, Kentarou Sawano, Helge Riemann, Nikolay V. Abrosimov, Dominique Bougeard, and Lars R. Schreiber. Large, Tunable Valley Splitting and Single-Spin Relaxation Mechanisms in a Si/Six Ge1-x Quantum Dot. *Physical Review Applied*, 13(3):1, 2020.

[47] Inga Seidler, Malte Neul, Eugen Kammerloher, Matthias Künne, Andreas Schmidbauer, Laura Diebel, Arne Ludwig, Julian Ritzmann, Andreas D. Wieck, Dominique Bougeard, Hendrik Bluhm, and Lars R. Schreiber. Tailoring potentials by simulation-aided design of gate layouts for spin qubit applications. *arXiv preprint arXiv:2303.13358*, pages 1–10, 2023.

[48] P. L. Bavdaz, H. G. J. Eenink, J. van Staveren, M. Lodari, C. G. Almudever, J. S. Clarke, F. Sebastiano, M. Veldhorst, and G. Scappucci. A quantum dot crossbar with sublinear scaling of interconnects at cryogenic temperature. *arXiv preprint arXiv:2202.04482*, pages 1–6, 2022.

Opportunities of Polarization-Resolved EUV Scatterometry on Photomasks

Victor Soltwisch[a], Till Biskup[a], Michael Kolbe[a], and Frank Scholze[a]

[a]Physikalisch-Technische Bundesanstalt (PTB), Abbestraße 2–12, 10587 Berlin, Germany

ABSTRACT

The development of EUV scatterometry, as a potentially interesting option for the characterization of photomasks, has made significant progress in the last decade. The at-wavelength performance is directly assessed, and the measurements are potentially fast, a relevant precondition for the development of production-worthy instruments. However, an accurate prediction of the imaging properties requires a high effort in simulation and the precise determination of fundamental parameters such as the optical constants to correctly model the interaction of the EUV radiation with the nano-structured EUV mask. Another possibility to further increase the sensitivity and minimize the uncertainties in the reconstruction could be a polarization-resolved analysis of the scattered EUV radiation. The high degree of linear polarization of the beamline is ideal for investigating this question. However, it remains to be proven whether the additional measurement effort leads to a reduction of the uncertainties in the geometric reconstruction of the nanostructures. Besides theoretical modeling addressing this question, we also present benchmark measurements performed by PTB at the soft X-ray beamline of BESSY II.

Keywords: EUV scatterometry, polarization-resolved analysis, nano-metrology, photomasks

1. INTRODUCTION

Most current semiconductor devices consist of geometric features in the lower nanometer range that need to be accurate on a sub-nanometer scale.[1] Furthermore, producing state-of-the-art integrated electronic circuits is a multi-step process where any misplacement of the edge position can lead to a complete failure of the features.[2] Hence, monitoring the manufacturing employing nano-metrology methods accounts already for more than 50% of the overall processing time.[3] Ideally, the metrology method of choice will be non-invasive and non-destructive, fast, and applicable to in-line monitoring. Traditionally, optical critical dimension (OCD) metrology has been used and allows for fast and non-destructive inspections of structured surfaces in high volume manufacturing.[4] However, due to using wavelengths in the visible range, the sensitivity limits of OCD are in the nanometer range. Alternative methods already employed include atomic force and electron microscopy that allow for direct imaging, but are limited in data acquisition speed, as they require scanning the sample area. Due to their wavelength below 0.1 nm, hard X-rays with photon energies above about 1 keV are suitable for sub-nanometer resolution.[5] Grazing-incidence small-angle X-ray scattering (GISAXS) has therefore been proposed as an alternative method for characterizing nanostructured surfaces. However, the elongated footprint leads to experimental[6] and data analysis challenges.[7] A recent comparison between Extreme Ultraviolet (EUV) scatterometry and GISAXS has shown the potential of using EUV radiation to characterize nanostructured surfaces.[8] Compared to GISAXS, EUV scattering allows for investigating samples using steeper angles, thus reducing the footprint of the beam on the sample while keeping the surface sensitivity of the scattering technique. Additionally, EUV scatterometry provides better optical contrast for materials relevant in the semiconductor industry, including oxide layers. Examples of applying EUV scatterometry include at-wavelength investigations of MoSi multilayer mirrors[9] and analyzing the carbon contamination of photomasks.[10,11] Current developments of laboratory EUV scatterometers illustrate the potential of these methods and can already compete with synchrotron-based instruments.[12]

Characterizing nanostructures by means of EUV scatterometry relies on optimization methods that are based on a forward model. Hence, reliably reconstructing parameters and their uncertainties requires these models to be as faithful to the experimental conditions and as versatile as possible.[13,14] This can only be achieved by

Further author information: victor.soltwisch@ptb.de

38th European Mask and Lithography Conference (EMLC 2023), edited by
Uwe F.W. Behringer, Jo Finders, Proc. of SPIE Vol. 12802, 128020F
© 2023 SPIE · 0277-786X · doi: 10.1117/12.2675921

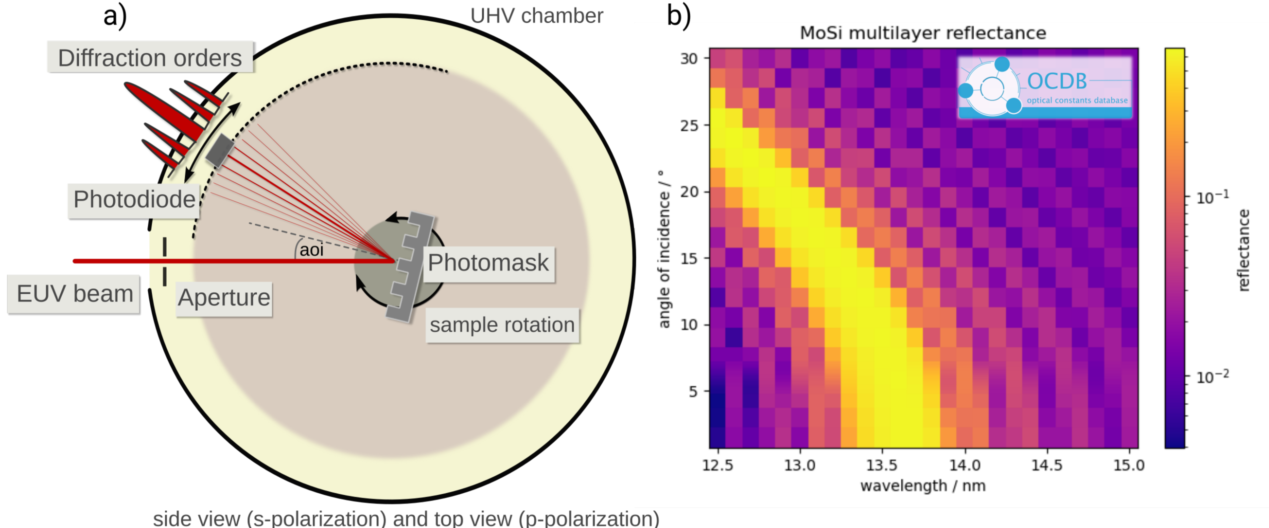


Figure 1. a) Schematic representation of EUV scatterometry. Measurements with different polarizations can be realized by changing the sample position from horizontal (s-polarization) to vertical (p-polarization). This procedure is much easier than rotating the synchrotron. b) This setup is also used for determining the optical constants of layered systems in the EUV spectral range (https://www.ocdb.ptb.de). The figure shows the measured MoSi multilayer reflectance of the photomask around 13.5 nm.

precisely determining fundamental parameters such as the optical constants to correctly model the interaction of the EUV radiation with the nano-structured EUV mask. PTB, the German metrological institute, is developing a public database for optical constants (n&k) of different materials and compositions with several European partners. This optical constant database (OCDB, https://www.ocdb.ptb.de) already contains various n&k measured values from the EUV and soft X-ray spectral range and is continuously being expanded. Another possibility to further increase the sensitivity and minimize the uncertainties in the reconstruction could be a polarization-resolved analysis of the scattered EUV radiation. With their high degree of linear polarization the EUV beamlines at the PTB laboratory at the electron storage ring BESSY II are ideally suited to investigate this question.

2. EXPERIMENT AND SIMULATION

Polarization-resolved scatterometry measurements were performed in the radiometry laboratory of the PTB located at the electron storage ring BESSY II, and here at the PTB soft X-ray beamline.[15–17] This beamline is dedicated to EUV radiation at 13.5 nm spanning the photon energy range from 50 eV to 1800 eV. The beamline is well-optimized regarding its beam profile, with minimal divergence, strong collimiation, and highly suppressed higher diffraction orders using metallic foils. The radiation can be monochromatized with a spectral resolution better than 0.25 eV. The experimental end-station of the PTB soft X-ray beamline consists mainly of a vacuum tank situated in a clean-room environment. The vacuum tank houses a reflectometer-scatterometer setup enabling scatterometry experiments in linear s- and p-polarization by rotating the sample (see Fig. 1). The mechanics of the setup are lubricant-free to reduce the environmental contamination known to affect optical surfaces. For data acquisition, GaAsP photodiodes are used to detect and quantify the diffraction efficiencies. The infrared encoders used at the goniometer provide an independent reference of the sample position.

The sample investigated is an EUV photomask that has been well-characterized in previous studies[18,19] with respect to the optical constants and the multilayer stack. The absorber structure on the MoSi layer consists of a newly developed TaTeN[20] material with a pitch of 180 nm. The EUV scatterometry was performed in the classical dispersion geometry with the grating lines vertical to the incident beam with a wavelength of 13.5 nm. The angle of incidence (AOI) was then varied from 3° to 21° (near normal) and the resulting diffraction intensities

were measured with the photodiode. Identical measurements were then repeated in p-polarization after rotating the sample by 90° and measuring the diffraction orders horizontally.

The simulations in this study were performed with the software package JCMsuite. The Maxwell solver,[21,22] based on finite elements, allows a very high numerical precision and the free modeling of any shape. A side length of 5 nm and a polynomial degree of 4 for the approximation of the near-field distributions was sufficient to push the numerical uncertainty one order of magnitude below the experimental one. With periodic boundary conditions, the far-field intensity (diffraction intensities) can be rigorously determined via a FFT in a post-process. The incident radiation was described as a 100% linearly polarized plane wave.

3. RESULTS

The measured EUV scatterometry data obtained from the TaTeN photomask sample[20] are shown in Fig. 2. Scatterometry at 13.5 nm uses the resonance of the multilayer which has been optimized for maximum reflection around an angle of incidence of 6° and 13.5 nm. This also means that with increasing incidence angle the diffraction intensities decrease very fast. This gives a high sensitivity to the Bragg mirror and at the same time to the absorber structure. The intensity of the zeroth order drops by up to two orders of magnitude between an angle of incidence (AOI) of 3° (top) and 21° (bottom). Therefore, we chose a logarithmic representation to show the very weak higher diffraction orders. A noticeable feature of the measurements with p-polarized EUV radiation is their increased background at detection angles above 40°. This is probably due to the sensitivity of the GaAsP photodiodes to infrared light from the position encoders of the goniometer. However, this can be readily solved for future measurements by deactivating the encoders, which is easily possible for simple 2θ movements as in EUV scatterometry due to the high mechanical precision of the goniometer. The intensities of the individual diffraction orders were extracted for s- and p-polarized radiation (blue and red dots) and the asymmetry (Intensity(S) - Intensity(P)) plotted versus the diffraction order in the right figure panel (green dots). For comparison, the theoretically expected intensities were calculated based on the multilayer and absorber model reconstructed in Ref. 18 (pink dots). Qualitatively, measured and simulated asymmetry agree already very well. However, the experimental uncertainty (green dashed line) reveals that measuring the partly very small differences in the diffraction intensities is experimentally challenging. Also remarkable is the AOI of 15°, where theory predicts a clearly larger asymmetry than the value observed experimentally. This observation can have several reasons. On the one hand, the basis of the finite element method model used is based on a reconstruction from an older measurement[18] where only s-polarized radiation with an AOI of 3–12° was used. On the other hand, an additional variation of the sample position compared to the older measurements together with a slight inhomogeneity of the structured surface or aging effects of the EUV multilayer could lead to significant shifts of the intensities.

Another more likely possible explanation emerges when examining the theoretical modeling of the asymmetry in more detail. For this purpose, the expected asymmetry was continuously calculated over the entire range of incidence angles, and the results are shown in Fig. 3. In order to avoid false conclusions due to theoretically extremely large asymmetries in the high diffraction orders with very small diffraction intensities which are hardly measurable, all simulated intensities $< 10^{-5}$ were not considered in the representation. For a clearer representation, the relative asymmetries of the diffraction intensities versus the AOI are shown as well (Fig. 4, bottom left). Obviously, the AOI around 15° is very sensitive to the asymmetry. Smallest variations in the AOI can cause very strong changes in the asymmetry. This also indicates that the experimental uncertainty of the AOI (better than $\pm$ 0.01) should be considered as an important input parameter in future polarization-resolved EUV scatterometry reconstructions. In addition to the uncertainty of the AOI, the divergence and degree of linear polarization of the incident radiation may also need to be considered in future simulations.

The potential of polarization-resolved scatterometry for dimensional metrology of EUV photomasks also becomes visible when varying the linewidth of the absorber in simple forward calculations. This is shown in Fig. 4 (right panel). The polarization-dependent shift in the Kiessig fringes can lead to high asymmetries in the diffraction intensities. A linewidth variation of 80–90 nm at an AOI of 14° was simulated here. To take the experimental limitations into account, the intensities for the asymmetry calculation were limited once to $> 10^{-3}$ and once to $> 10^{-5}$. At relative intensities above 10^{-3}, which nowadays can be achieved by good laboratory EUV reflectometers/scatterometers with HHG sources, the diffraction orders that can be evaluated are already clearly

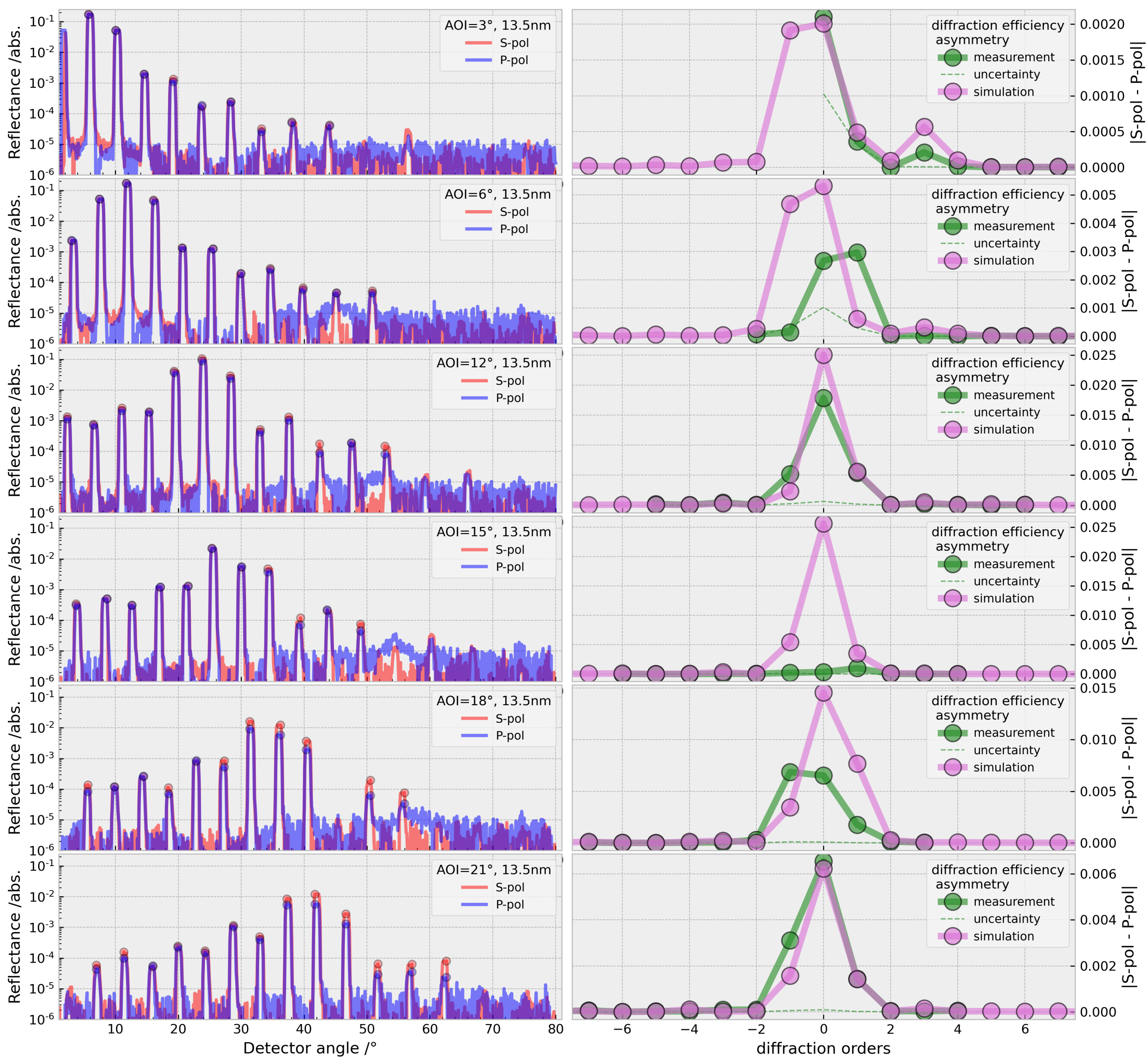


Figure 2. Left: Measured diffraction efficiencies for linear s- (red) and p-polarization (blue) at a wavelength of 13.5 nm are plotted on a logarithmic scale versus the detection angle for different angles of incidence (AOI) from 3° to 21°. The AOI is defined with respect to the surface normal. Right: Comparing the measured asymmetry (pink) in the diffraction intensities with the expected simulated asymmetry (green) on a linear scale. The experimental uncertainty is visualized as dashed green line.

limited in their number. But theoretically, there should be changes in the asymmetry of a few percent when varying the line width. In the range above 10^{-5}, which is a challenge even for very good soft X-ray beamlines, the evaluation of the asymmetries of the higher diffraction orders probably means a significant increase in sensitivity for scatterometry.

For more detailed conclusions, the different experimental and possible numerical uncertainties of the physical models have to be investigated in more detail. Also, the very precise measurement of the MoSi multilayer independently of the nm patterned surface and the characterization of the optical constants of the different materials will be crucial for further sensitivity improvements in EUV scatterometry.

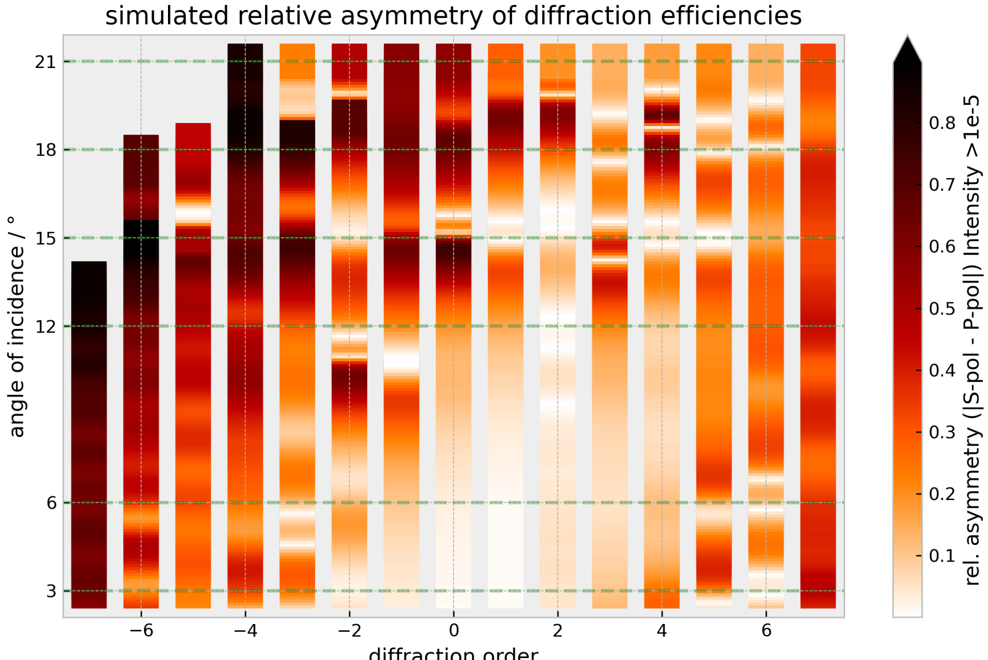


Figure 3. The theoretically expected relative asymmetry of diffraction efficiencies measured with linearly polarized EUV scatterometry. Intensities $< 10^{-5}$ were not considered in the asymmetry calculation.

4. CONCLUSIONS

Polarization-resolved EUV scatterometry was performed on a newly developed TaTeN EUV photomask at the PTB soft X-ray beamline situated at the electron storage ring BESSY II. The measured asymmetries of the diffraction efficiencies under linear s- and p-polarized EUV radiation at 13.5 nm show the possible potential for dimensional metrology and a further increase in sensitivity. The simulations carried out on the basis of the finite element method and rigorous Maxwell solvers already show a good agreement with the experimental results. At the same time, however, they also reveal remaining challenges in the experimental field. The very small intensity variations in the expected asymmetries of the higher diffraction orders are a tough problem for the precision and flux of laboratory EUV scatterometers or reflectometers. Further statistical analyses including all possible experimental uncertainties and optimizations of the measurements are necessary to quantitatively analyze the potential of the method.

ACKNOWLEDGMENTS

The authors gratefully acknowledge imec and in particular Vicky Philipsen for providing the sample that has been investigated. This project has received funding from Horizon Europe under grant agreement 101096772–14ACMOS, as well as from the EMPIR programme 20IND04 ATMOC. These Joint Undertakings receive support from the European Union's Horizon Europe research and innovation program alongside the Netherlands, France, Belgium, Germany, Czech Republic, Austria, Hungary, and Israel.

REFERENCES

[1] IEEE, "International roadmap for devices and systems (IRDS™): More Moore," tech. rep., IEEE (2021).

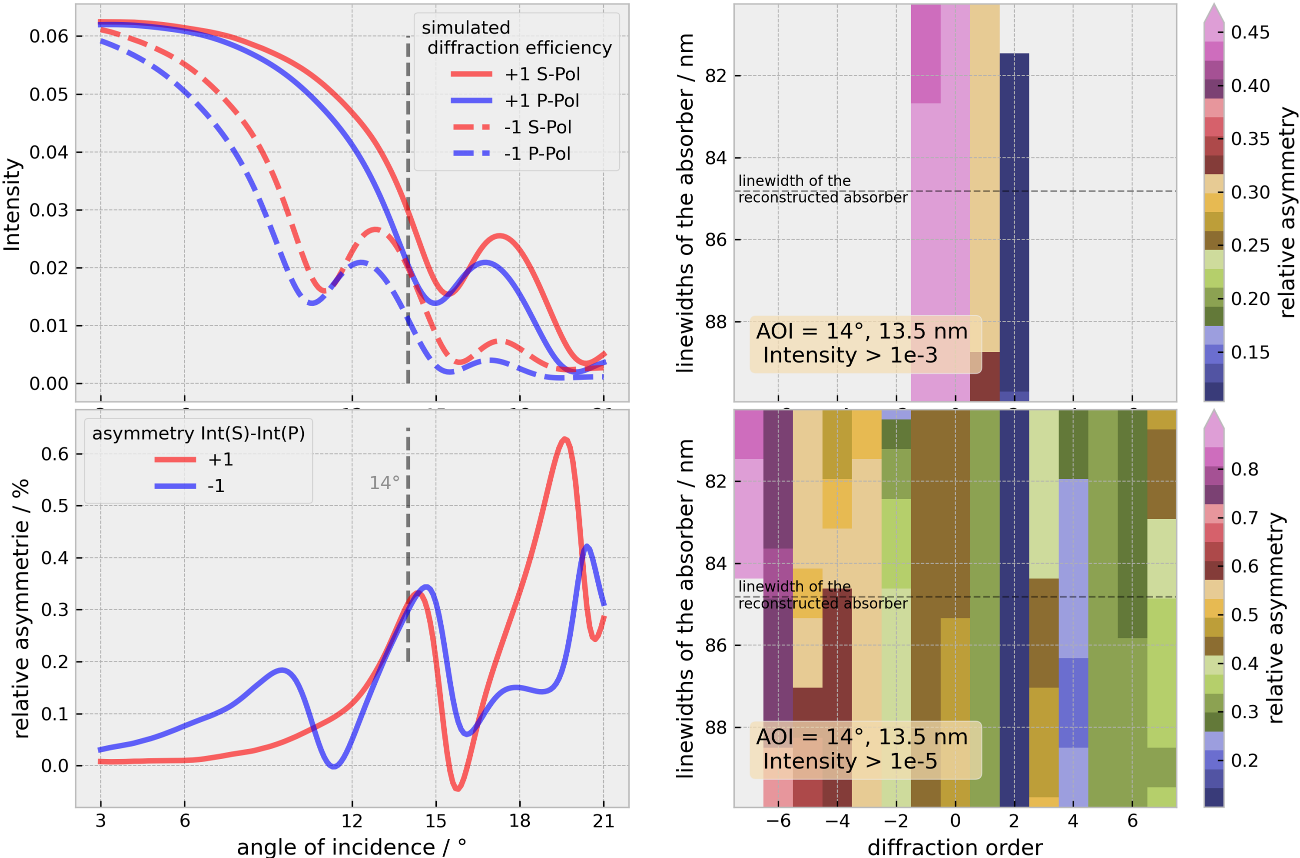


Figure 4. Left: Theoretical intensity modulation of the positive and negative first diffraction orders as a function of the AOI. The resonance structure of the MoSi multilayer is clearly visible. The shift in the Kiessig fringes can lead to high relative asymmetries. Right: Simulated variation of the linewidths (CD) of the TaTeN absorber structure in relation to the expected asymmetry at an AOI of 14°. Intensities below $< 10^{-3}$ and $< 10^{-5}$ were not considered to visualize experimental limitations. The reconstructed linewidth of the TaTeN absorber is marked with a dashed line.

[2] Bhattacharyya, K., "Tough road ahead for device overlay and edge placement error," in [*Metrology, Inspection, and Process Control for Microlithography XXXIII*], Ukraintsev, V. A. and Adan, O., eds., **10959**, 1095902, International Society for Optics and Photonics, SPIE (2019).

[3] Orji, N. G., Badaroglu, M., Barnes, B. M., Beitia, C., Bunday, B. D., Celano, U., Kline, R. J., Neisser, M., Obeng, Y., and Vladar, A. E., "Metrology for the next generation of semiconductor devices," *Nat. Electron.* **1**, 532–547 (2018).

[4] IEEE, "International roadmap for devices and systems (IRDS™): Metrology," tech. rep., IEEE (2021).

[5] Attwood, D., [*Soft X-Rays and Extreme Ultraviolet Radiation: Principles and Applications*], Cambridge University Press (1999).

[6] Roth, S. V., Autenrieth, T., Grübel, G., Riekel, C., Burghammer, M., Hengstler, R., Schulz, L., and Müller-Buschbaum, P., "In situ observation of nanoparticle ordering at the air-water-substrate boundary in colloidal solutions using x-ray nanobeams," *Appl. Phys. Lett.* **91**(9), 091915 (2007).

[7] Pflüger, M., Soltwisch, V., Probst, J., Scholze, F., and Krumrey, M., "Grazing-incidence small-angle X-ray scattering (GISAXS) on small periodic targets using large beams," *IUCrJ* **4**, 431–438 (2017).

[8] Herrero, A. F., Pflüger, M., Puls, J., Scholze, F., and Soltwisch, V., "Uncertainties in the reconstruction of nanostructures in euv scatterometry and grazing incidence small-angle x-ray scattering," *Opt. Express* **29**, 35580–35591 (2021).

[9] Haase, A., Soltwisch, V., Braun, S., Laubis, C., and Scholze, F., "Interface morphology of mo/si multilayer systems with varying mo layer thickness studied by euv diffuse scattering," *Opt. Express* **25**, 15441–15455 (2017).

[10] Fan, Y.-J., Yankulin, L., Thomas, P., Mbanaso, C., Antohe, A., Garg, R., Wang, Y., Murray, T., Wüest, A., Goodwin, F., Huh, S., Cordes, A., Naulleau, P., Goldberg, K., Mochi, I., Gullikson, E., and Denbeaux, G., "Carbon contamination topography analysis of EUV masks," in [*Extreme Ultraviolet (EUV) Lithography*], Fontaine, B. M. L., ed., **7636**, 76360G, International Society for Optics and Photonics, SPIE (2010).

[11] Levinson, H. J. and Brunner, T. A., "Current challenges and opportunities for EUV lithography," in [*International Conference on Extreme Ultraviolet Lithography 2018*], Ronse, K. G., Hendrickx, E., Naulleau, P. P., Gargini, P. A., and Itani, T., eds., **10809**, 1080903, International Society for Optics and Photonics, SPIE (2018).

[12] Lohr, L. M., Ciesielski, R., Glabisch, S., Schröder, S., Brose, S., and Soltwisch, V., "Nanoscale grating characterization using euv scatterometry and soft x-ray scattering with plasma and synchrotron radiation," *Appl. Opt.* **62**, 117–132 (Jan 2023).

[13] Soltwisch, V., Fernández Herrero, A., Pflüger, M., Haase, A., Probst, J., Laubis, C., Krumrey, M., and Scholze, F., "Reconstructing detailed line profiles of lamellar gratings from GISAXS patterns with a Maxwell solver," *J. Appl. Cryst.* **50**, 1524–1532 (2017).

[14] Herrero, A. F., Pflüger, M., Probst, J., Scholze, F., and Soltwisch, V., "Applicability of the debye-waller damping factor for the determination of the line-edge roughness of lamellar gratings," *Opt. Express* **27**, 32490–32507 (2019).

[15] Scholze, F., Beckhoff, B., Brandt, G., Fliegauf, R., Gottwald, A., Klein, R., Meyer, B., Schwarz, U. D., Thornagel, R., Tuemmler, J., Vogel, K., Weser, J., and Ulm, G., "High-accuracy EUV metrology of PTB using synchrotron radiation," in [*Metrology, Inspection, and Process Control for Microlithography XV*], Sullivan, N. T., ed., **4344**, 402–413, International Society for Optics and Photonics, SPIE (2001).

[16] Laubis, C., Kampe, A., Buchholz, C., Fischer, A., Puls, J., Stadelhoff, C., and Scholze, F., "Characterization of the polarization properties of PTB's EUV reflectometry system," in [*Extreme Ultraviolet (EUV) Lithography*], Fontaine, B. M. L., ed., **7636**, 76362R, International Society for Optics and Photonics, SPIE (2010).

[17] Richter, M. and Ulm, G., "Metrology with synchrotron radiation – a brief introduction," *PTB-Mitt.* **124**(3), 3–6 (2014).

[18] Soltwisch, V., Glabisch, S., Andrle, A., Philipsen, V., Saadeh, Q., Schröder, S., Lohr, L., Ciesielski, R., and Brose, S., "High-precision optical constant characterization of materials in the EUV spectral range: from large research facilities to laboratory-based instruments," in [*37th European Mask and Lithography Conference*], Behringer, U., ed., **12472**, SPIE (2023).

[19] Ciesielski, Richard and Saadeh, Qais and Philipsen, Vicky and Opsomer, Karl and Soulie, Jean-Philippe and Wu, Meiyi and Naujok, Philipp and van de Kruijs, Robbert W. E. and Detavernier, Christophe and Kolbe, Michael and Scholze, Frank and Soltwisch, Victor, "Determination of optical constants of thin films in the EUV," *APPLIED OPTICS* **61**(8), 2060–2078 (2022).

[20] Wu, M., Thakare, D., de Marneffe, J.-F., Jaenen, P., Souriau, L., Opsomer, K., Soulié, J.-P., Erdmann, A., Mesilhy, H. M. S., Naujok, P., Foltin, M., Soltwisch, V., Saadeh, Q., and Philipsen, V., "Study of novel EUVL mask absorber candidates," *J. Micro/Nanopattern. Mater. Metrol.* **20**, 021002 (2021).

[21] Pomplun, J., Burger, S., Zschiedrich, L., and Schmidt, F., "Adaptive finite element method for simulation of optical nano structures," *Phys. Status Solidi B-Basic Solid State Phys.* **244**, 3419–3434 (2007).

[22] Burger, S., Zschiedrich, L., Pomplun, J., Blome, M., and Schmidt, F., "Advanced finite-element methods for design and analysis of nanooptical structures: applications," in [*Emerging Liquid Crystal Technologies VIII*], Chien, L.-C., Broer, D. J., Chigrinov, V., and Yoon, T.-H., eds., **8642**, 864205, International Society for Optics and Photonics, SPIE (2013).

EUV Mask Defect Inspection for the 3nm Technology Node

Yannick Hermans*[a], Tilmann Heil[b], Renzo Capelli[c], Bartholomaeus Szafranek[b], Daniel Rhinow[b], Gerson Mette[b], Patrick Salg[b], Christian Felix Hermanns[b], Bappaditya Dey[a], Luc Halipre[a], Darko Trivkovic[a], Paulina Rincon Delgadillo[a], Thomas Marschner[a], Sandip Halder[a]
[a]imec, Kapeldreef 75, Heverlee 3001, Belgium; [b]Carl Zeiss SMT GmbH, Industriestraße 1, 64380 Roßdorf, HE, Germany; [c]Carl Zeiss SMT GmbH, Rudolf-Eber-Straße 2, 73447 Oberkochen, BW, Germany

ABSTRACT

The paradigm switch to a reflective mask design for EUV lithography has proven to be challenging. Within the Horizon2020 PIn3S program Zeiss and imec are collaborating to address some of these challenges. In this work, an EUV mask with a collection of programmed defects representative for the 3nm technology node was reviewed. Defect printability at wafer level was analyzed after exposure on the ASML NXE:3400B by SEM. Furthermore, the mask was analyzed on the Zeiss AIMS® EUV platform and by SEM. For P36 (1x) 1:1 L/S programmed extrusions we have demonstrated that AIMS® EUV can be used to predict ADI local defect widths as well as (µ)bridge printability. Moreover, from P36 to P32 the mask spec regarding allowed opaque L/S extrusion widths needs to be tighter considering an earlier onset of ADI (µ)bridge printability and a stronger than expected ADI defect width increase through pitch.

Keywords: 3nm technology node, EUV lithography, Defectivity, Mask review, CDSEM, Wafer print, Zeiss AIMS® EUV

1. INTRODUCTION

Reflective mask development for EUV lithography has proven to be challenging: increasing feature density, smaller feature size, a wide range of absorber materials and the need for multilayer mirror-blanks are just a few of the challenges that come with EUV masks. With these challenges mask defect and variability management also become more stringent. A series of research works have for instance shown how mask variations can contribute to stochastic defect printing as well as wafer edge placement errors in defocus.[1–4]

A way to systematically investigate mask defect detectability, repairability and wafer printability for a particular feature is through a series of intentionally placed programmed defects with increasing size. The defect printability and defect width for after development inspection (ADI) and for after development inspection (AEI) can be verified using critical dimension secondary electron microscopy (CDSEM). For instance, Das et al. found for a set of programmed opaque P32 (1x scale) line/space (L/S) mask defects that defect size increases for AEI CDSEM compared to ADI CDSEM.[5] Moreover, below a programmed defect size of about 100 nm^2, AEI CDSEM showed that the size of the defect print was subject to variation and did not even print all the time.

Besides performing CDSEM measurements on wafer prints, programmed defects can be assessed using actinic mask disposition tools which do not require the use of a lithography scanner.[6] Actinic mask disposition allows to establish an aerial image of specific mask regions by using an illuminator equivalent to that of a lithography scanner and projection optics that allows the magnification of the aerial image onto a CCD camera. The Zeiss AIMS® tools are widely used in industry for actinic defect disposition after mask patterning to identify killer defects.[7] After those defects are repaired the Zeiss AIMS® tool can be used to determine the quality of the repair. In recent work, the Zeiss AIMS® EUV was successfully used to qualify the effective repair of programmed 2.25 nm (1x) extrusions for a P44 L/S 1:1 pattern.[8] Despite the effective repair, it was not clear whether that extrusion would have led to a defect at wafer level and whether repair was needed after all.

* yannick.hermans@imec.be

38th European Mask and Lithography Conference (EMLC 2023), edited by
Uwe F.W. Behringer, Jo Finders, Proc. of SPIE Vol. 12802, 128020G
© 2023 SPIE · 0277-786X · doi: 10.1117/12.2678392

In this work, we have addressed for P36 (1x scale) 1:1 L/S to what extent opaque programmed (internal) extrusions can be detected and from which size they start to print on wafer. With AIMS® EUV, we have established that extrusions down to a width of 1.25 nm (1x) could be detected. Furthermore, the EUV programmed defect mask was exposed with the ASML NXE:3400B (0.33 NA) using an SOC/SOG/CAR resist stack to determine the defect printability and local CD of the programmed extrusions at wafer level with ADI CDSEM. We have shown that (μ)bridges do not print for extrusion widths of 1.75 nm and below and that at 1.75 nm the local CD deviates on average 11.5% from target. Moreover, a good correlation could be found between the ADI local CD and AIMS® local CD, allowing to predict from AIMS® measurements alone whether a particular opaque P36 L/S mask defect would print μbridges on wafer and what local ADI CD could be expected. Additionally, ADI CDSEM measurements for P34 1:1 L/S and P32 1:1 L/S showed that the onset of (μ)bridge printing takes place at smaller extrusion widths and that the relative local CD deviation at the programmed defect width increases stronger than what could be expected from the pitch decrease. Thus, the mask spec regarding allowed opaque defect widths needs to be tighter for lower L/S pitches.

2. METHODOLOGY

For this study a programmed defect mask has been designed and manufactured in the framework of the EU ECSEL PIn3S project, therefore the mask is called "PIn3S mask" in the following. This mask contains among others P32 to P36 1:1 vertical (vL/S) and horizontal (hL/S line/space defects as well as P45 orthogonal contact hole (oCH) defects. Mask SEM review was done on a Zeiss MeRiT LE repair tool to image the defects at mask level. The aerial image intensity at a subset of programmed defect locations was probed using AIMS® EUV measurements. The experimental setup as well as the capabilities of AIMS® EUV have been described elsewhere.[9] A horizontal dipole, vertical dipole and a quasar (Figure 1a) were used as sources during the AIMS® EUV measurements of the vL/S, hL/S and oCH programmed defects, respectively. The local CD at the programmed defects was evaluated from the AIMS® EUV aerial image maps by determining an aerial image intensity threshold based on the target CDs, being 18 nm, 17 nm and 16 nm for the P36, P34 and P32 L/S patterns.

The mask was also used for wafer prints using the ASML EUV NXE:3400B scanner. A horizontal dipole, vertical dipole and an annular (Figure 1b) flexray source were used for the EUV lithography wafer prints. The target CDs were 18 nm, 17 nm and 16 nm for the P36, P34 and P32 L/S patterns, and 32 nm for the oCH pattern. EUV compatible SOC/SOG/CAR resist thin film stacks were used during the wafer prints. ADI CDSEM was carried out on the printed wafers using the Hitachi CG5000. To study the ADI defectivity and local CD at the L/S programmed extrusions (Figure 1c) there were 984 CDSEM measurements (Figure 1d) for which the space CD at the programmed defect (MP1) as well as the neighboring space CDs (MP2-MP19) were measured (Figure 1e).

3. RESULTS

3.1 PIn3S programmed defect mask review

As a first part of this work a set of programmed P36 vL/S, P36 hL/S and P45 oCH programmed defects were studied using mask SEM, AIMS® EUV and ADI CDSEM. Many of the studied programmed defects had relatively large sizes as can be seen from the 1x design dimensions of the examples in Figure 2. Consequently, these defects could be readily distinguished in mask SEM images and AIMS® EUV aerial image maps. Moreover, ADI CDSEM images showed that these large size programmed defects systematically printed on wafer. The main aim of studying these systematic defects is to verify the success of a mask repair process, which is part of the mask repair work package of the PIn3S project and will be reported at a later stage.

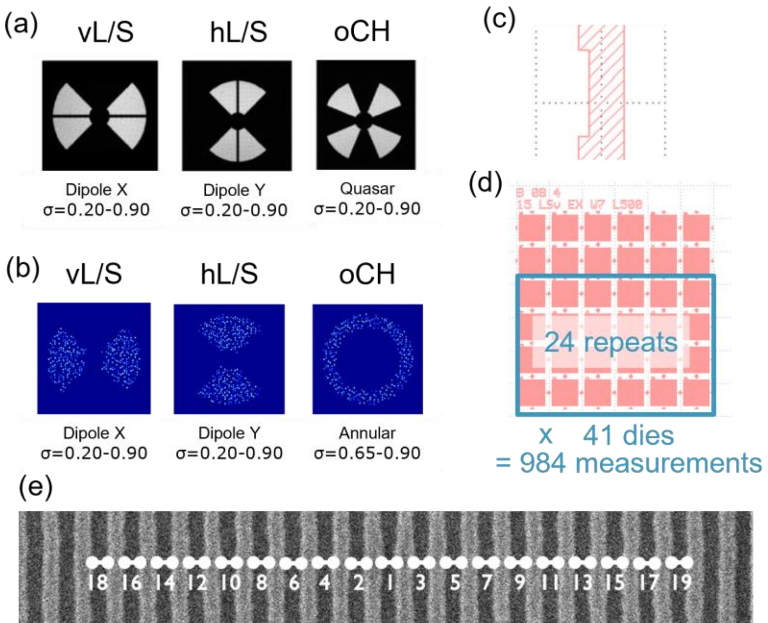


Figure 1: Sources used for the a) AIMS® EUV measurements and b) wafer prints on the ASML NXE:3400B for vertical line/spaces (vL/S), horizontal line/spaces (hL/S) and octahedral contact holes (oCH). (c) Example of a programmed extrusion on the PIn3S programmed defectivity mask. (d) ADI CDSEM measurement series for one programmed defect. (e) All local space ADI CDSEM measurement points (MP) with MP1 being the location of the programmed defect.

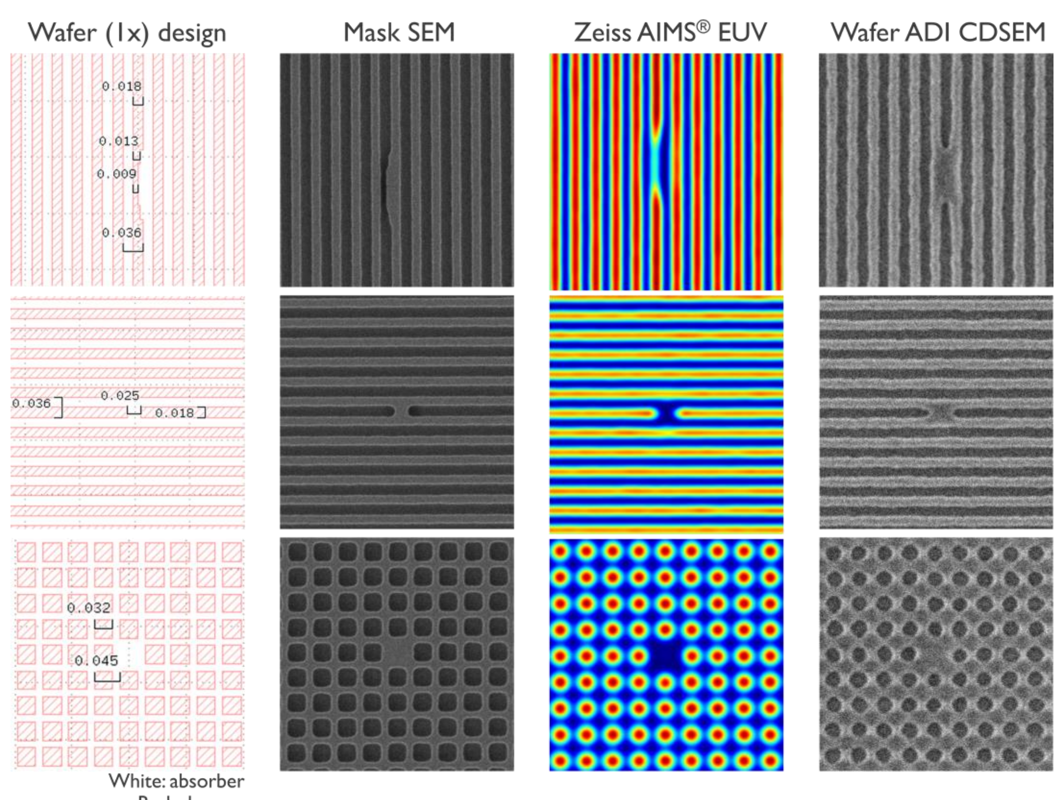


Figure 2: Examples of programmed defects for P36 vertical L/S, P36 horizontal L/S and P45 octahedral contact holes which print systematically on wafer. The programmed defects were analyzed using mask SEM, AIMS® EUV, and wafer print ADI CDSEM.

3.2 P36 L/S programmed (internal) extrusion detection and printability

As the programmed defect dimensions decrease, the ADI CD deviation as well as the likelihood of visible defectivity after ADI decreases.[5] To systematically study the printability of small L/S P36 programmed defects, a set of programmed extrusions (EX) and internal extrusions (EXI) with varying width and length (Figure 3) were measured using AIMS® EUV and ADI CDSEM. From the AIMS® EUV aerial image maps (Figure 4) all the programmed extrusions under study down to a 1x width of 1.25 nm gave rise to a considerable decrease in local aerial image intensity, allowing for their detection.

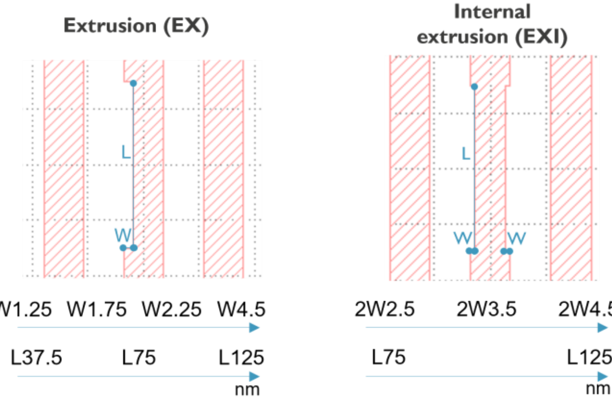


Figure 3: Series of programmed P36 L/S extrusions (EX) and internal extrusions (EXI). The total width (W (extrusions), 2W (internal extrusions)) and length (L) of the probed defects are illustrated.

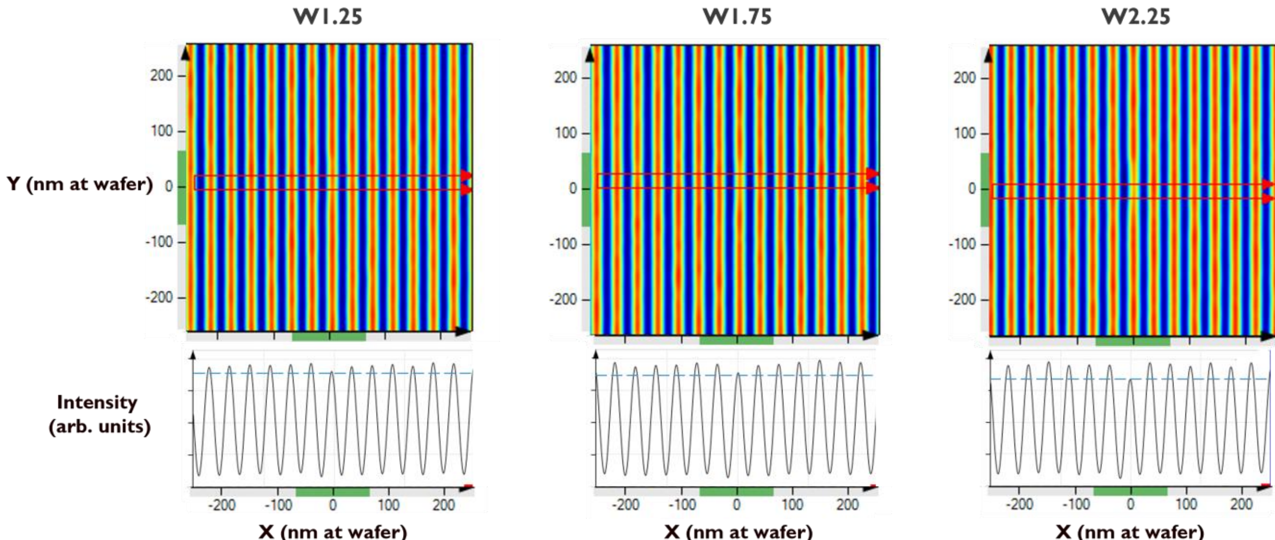


Figure 4: AIMS® EUV measurements for programmed P36 L/S extrusions with 75 nm length and various width. The upper graphs show the areal image intensity maps around the programmed extrusions. The lower graphs are plots of the aerial image intensity along the x-axis defined by the red arrows. The programmed extrusions situate at about (x: 0 nm, y: 0 nm).

The programmed mask extrusions were also further studied using wafer print ADI CDSEM in which the space CD at the local defect (measurement point 1 (MP1)) was compared to the space CD of the neighboring measurement points (MP2 to MP 19) (Figure 5). From W1.25 to W4.5 the boxplots, visualizing the CD distributions per MP, show that only the CD of MP1 was clearly affected by the programmed defect situated at MP1. Moreover, at a programmed defect width of 4.5 nm a tail of high CD outliers could be noticed for MP1. These outliers originate from faulty CD measurements caused by

bridging. Indeed, the MP1 CD scatter plot distributions in Figure 6 show that almost all the high CD outliers for W4.5 are colored red, which is used here to indicate visible (µ)bridges. For the programmed extrusions with a width of 1.75 nm and below, no visible µbridges (of 984 CDSEM images taken) could be detected. Moreover, for W2.25 only one µbridge could be detected for L75 and L125.

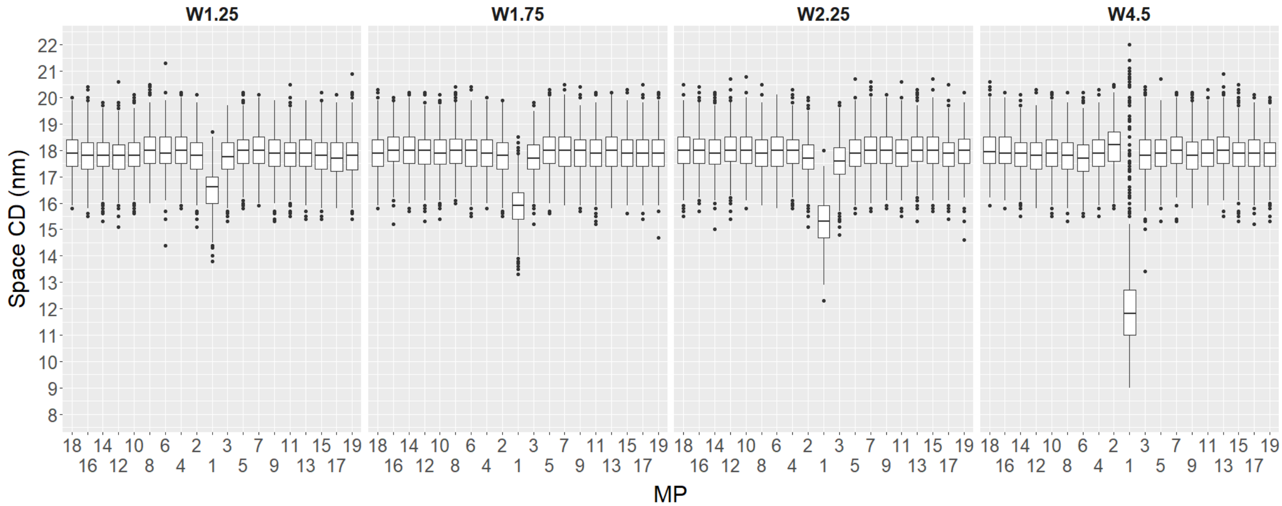


Figure 5: Wafer ADI CDSEM space CD boxplots for programmed P36 L/S extrusions with 75 nm length and various width. MP1 corresponds to the location of the programmed defect while the other MPs represent the neighboring spaces.

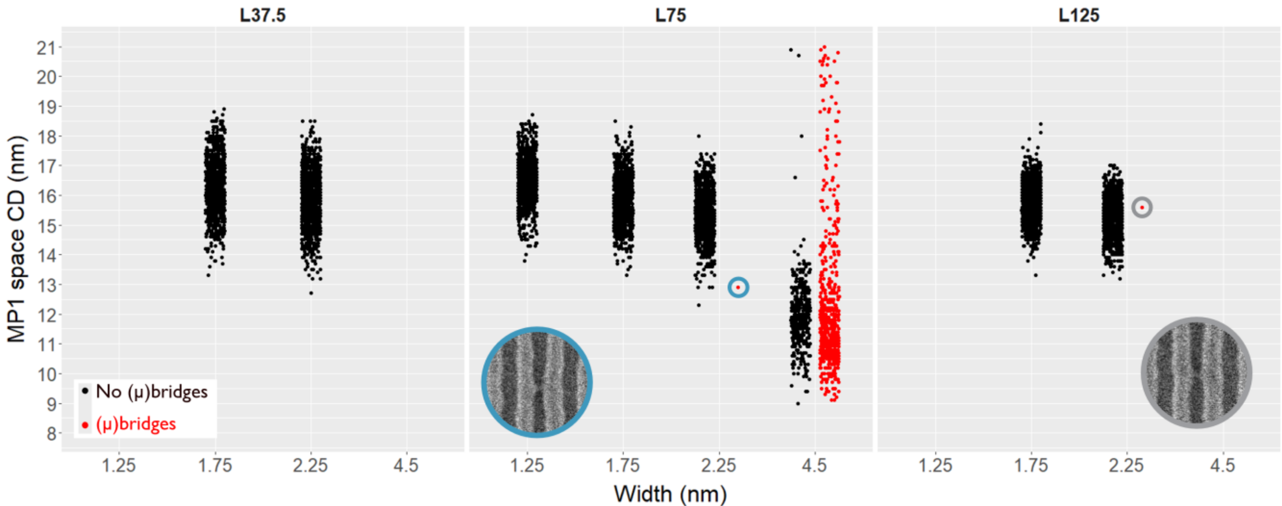


Figure 6: Wafer ADI CDSEM MP1 space CD scatter plots for programmed P36 L/S extrusions (EX) with various length and various width. Every point represents one measurement. The distributions are split based on whether a (µ)bridge could be seen at the MP1 programmed defect (red) or not (black). For some measurement points the ADI CDSEM image is shown.

For the programmed internal extrusions, the onset of µbridges lies at a total width of 2.5 nm, as a µbridge could be observed for internal extrusion lengths of 75 nm and 125 nm (Figure 7). Although a single opaque defect of 1.25 nm would not give rise to a (µ)bridge, two neighboring µbridges would. The internal extrusion CD scatterplots also show that the likelihood of a (µ)bridge rises when the length of the programmed defect increases. As such, increasing the programmed defect length could potentially be a way to test with fewer CDSEM measurements whether a programmed defect prints systematically in a particular lithography process.

For the P36 EX and EXI L/S programmed defects, Figure 8 shows a correlation of the AIMS® EUV space CD and the average ADI CDSEM CD with a slope of 1 and an offset of about 0.5 nm. This correlation allows for a reliable

prediction of the average ADI CDSEM CD and number of ADI (µ)bridges based on AIMS® EUV measurements, on the condition that a calibration curve as in Figure 8 is available. Roughly, three zones can be established based on ADI CD deviation and (µ)bridge printing. There's a green zone where the average MP1 ADI CD deviation amounts to about 2 nm and where no (µ)bridges could be identified. Then, in the orange zone the average CD deviation is smaller than 3.5 nm and 1 to 10 µbridges could be seen. Last there is a red zone with CD deviations larger than 3.5 nm and more than 10 (µ)bridges. Importantly, despite AIMS® EUV being able to detect the defects in the green zone, those defects likely do not require a repair, raising another advantage of knowing how a series of programmed defects propagate after lithography wafer prints.

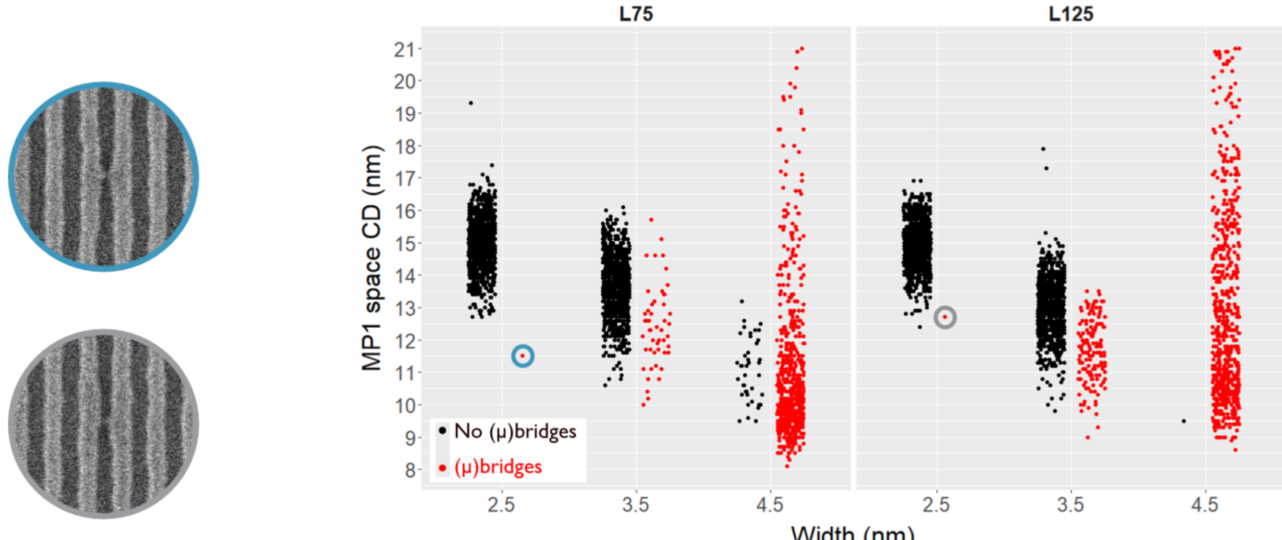


Figure 7: Wafer ADI CDSEM MP1 space CD scatter plots for programmed P36 L/S internal extrusions (EXI) with various length and various width. Every point represents one measurement. The distributions are split based on whether a (µ)bridge could be seen at the MP1 programmed defect location (red) or not (black). For some measurement points the ADI CDSEM image is shown.

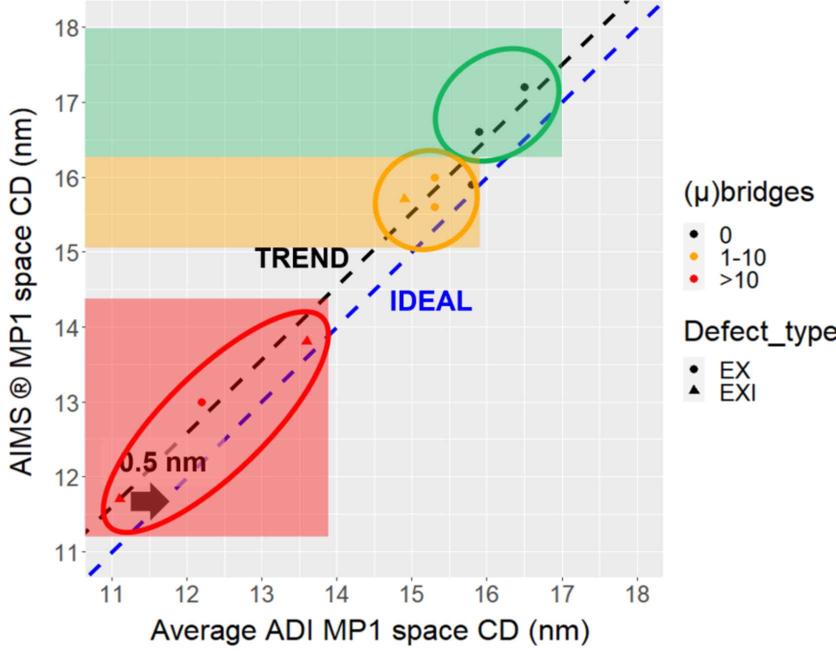
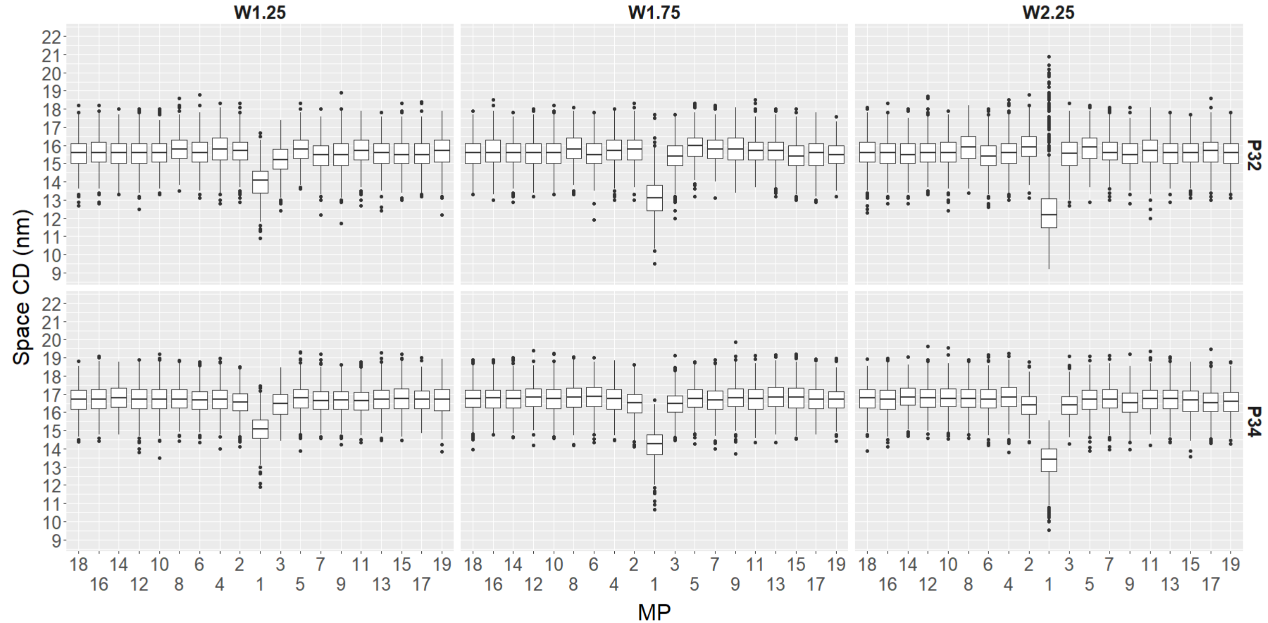


Figure 8: AIMS® EUV MP1 space CD versus average ADI CDSEM MP1 space CD. Green zone depicts the absence of (µ)bridges, yellow zone corresponds to 1-10 µbridges and the red zone represents the programmed extrusions with more than 10 µbridges.

3.3 P34, P32 L/S programmed (internal) extrusion printability

The study into the limits of (internal) extrusion L/S printability was also extended to P34 and P32. The boxplot ADI space CD boxplots for P34 and P32 (Figure 9) also show that the effect of the programmed extrusions on MPs other than MP1 is limited. Furthermore, at P32 the high CD outlier tail can now already be observed at W2.25. The ADI space CD distributions through pitch (Figure 10) show that the onset of (µ)bridges happens at earlier programmed defect widths as the pitch becomes smaller. Indeed, while for P34 and P36 (µ)bridges do not print at programmed extrusion widths of 1.25 nm and 1.75 nm (Figure 11), (µ)bridges already appear from 1.25 nm onwards for P32. Furthermore, the average ADI MP1 space CD for the same mask programmed defect width increases with decreasing pitch and more rapidly than what can be expected from the decrease in target CD through pitch. This observation can be more clearly deduced from the relative average ADI CD deviation (ΔCD/CD) versus relative programmed defect width plot in Figure 12. Figure 12 clearly summarizes that the relative CD deviation increases (e.g. for W1.25 from 8% to 11%) as the L/S pitch decreases from P36 to P32. This shows, together with the earlier onset of (µ)bridge printing, that mask specs regarding allowed mask defect sizes should tighten when going from P36 to P32 L/S.



.

Figure 9: Wafer ADI CDSEM space CD boxplots for programmed P34 and P32 L/S extrusions with 75 nm length and various width. MP1 corresponds to the location of the programmed defect while the other MPs represent the neighboring spaces.

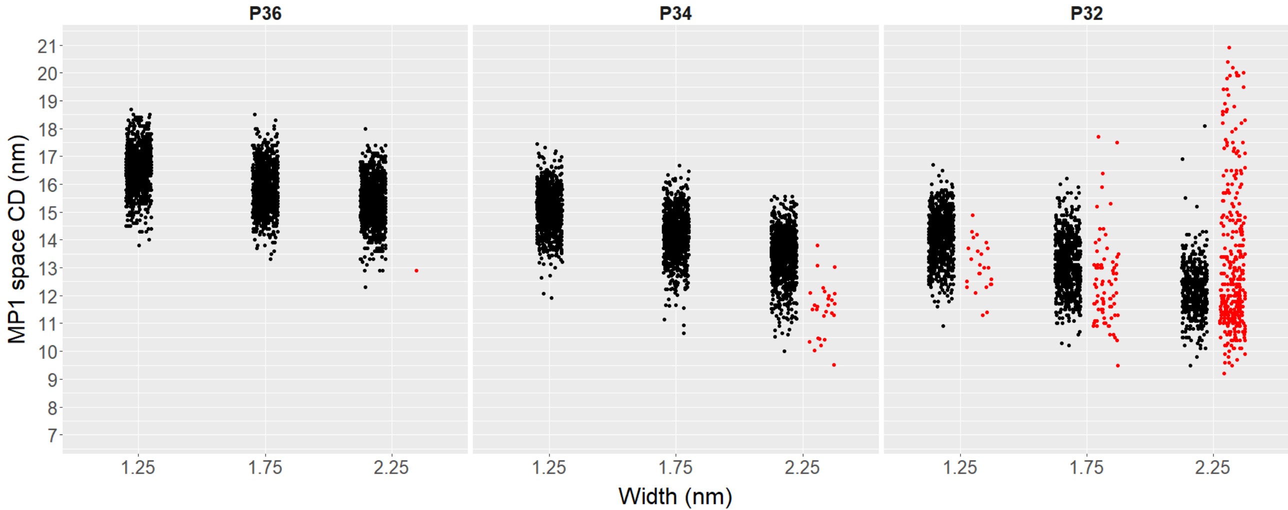


Figure 10: Wafer ADI CDSEM MP1 space CD scatter plots for programmed P36, P34 and P32 L/S extrusions (EX) with various length and various width. Every point represents one measurement. The distributions are split based on whether a (µ)bridge could be seen at the MP1 programmed defect location (red) or not (black).

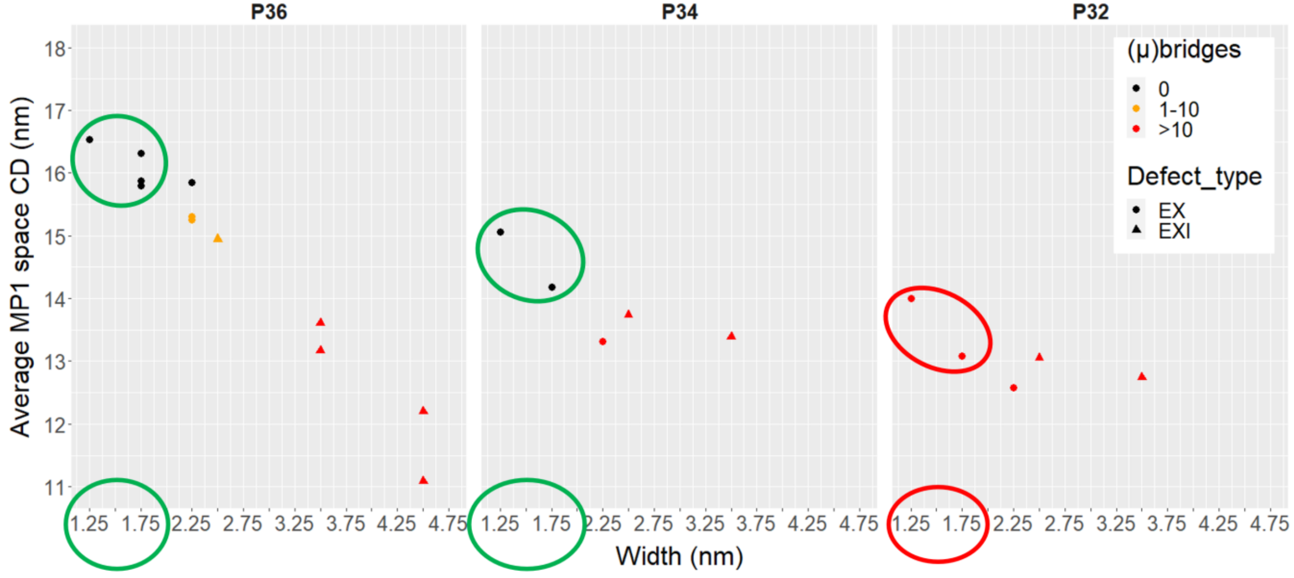


Figure 11: Average wafer ADI CDSEM MP1 space CD scatter plots for programmed P36, P34, P32 L/S extrusions (EXI) with various length and various width. Every point represents one measurement. The distributions are split based on whether a (µ)bridge could be seen at the MP1 programmed defect location (red) or not (black).

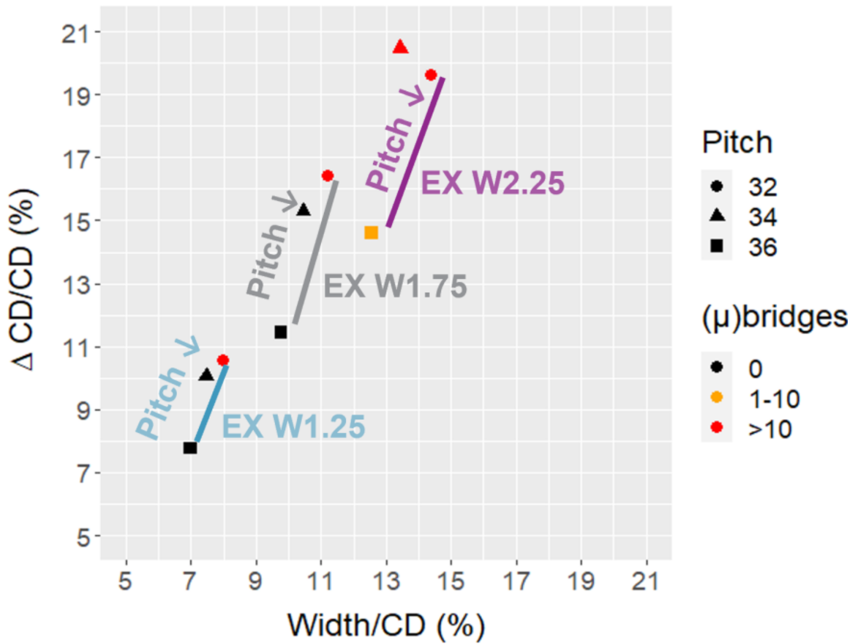


Figure 12: average ADI CDSEM MP1 ΔCD/CD (with ΔCD being MP1 CD – target CD and CD being target CD) versus relative programmed defect width/CD for extrusions with 75 nm length and various width.

4. CONCLUSIONS AND OUTLOOK

We have shown for P36 (1x) 1:1 L/S programmed extrusions that ADI local defect widths as well as (μ)bridge printability can be predicted using AIMS® EUV measurements if the correlation between AIMS® EUV and ADI wafer CDSEM has been established. Furthermore, ADI CDSEM inspections on P36 to P32 1:1 L/S programmed extrusions showed an earlier onset of ADI (μ)bridge printability and a stronger than expected ADI defect width increase through pitch. Consequently, the mask spec regarding allowed opaque L/S extrusion widths needs to tighten towards lower pitches.

Next, the PIn3S programmed defect mask will undergo a series of repair experiments. Then, AIMS® EUV and wafer print CDSEM are to be used to verify the quality of the repairs. Furthermore, how the programmed extrusion defect printability changes after etch should be understood better.

This project has received funding from the ECSEL Joint Undertaking (JU) under grant agreement No 826422. The JU receives support from the European Union's Horizon 2020 research and innovation program and Netherlands, Belgium, Germany, France, Romania, Israel.

REFERENCES

[1] Rik Jonckheere., “Overcoming EUV mask blank defects: what we can, and what we should,” Proc.SPIE 10454, 104540M (2017).
[2] Rik Jonckheere and Lawrence S. Melvin III., “Stochastic printing behavior of non-local mask deficiencies in EUV lithography,” Proc.SPIE 11517, 1151710 (2020).
[3] R. Jonckheere, D. Van den Heuvel, N. Takagi, H. Watanabe, and E. Gallagher., “Correlation of actinic blank inspection and experimental phase defect printability on NXE3x00 EUV scanner,” Proc.SPIE 9422, 942216 (2015).

[4] Rik Jonckheere, Lawrence S. Melvin III, and Renzo Capelli., “Stochastic printing behavior of ML-defects on EUV mask,” Proc.SPIE 11147, 111470P (2019).
[5] Poulomi Das, Alain Moussa, Christophe Beral, Mihir Gupta, Mohamed Saib, Sandip Halder, Anne Laure Charley, and Philippe Leray., “Printability and propagation of stochastic defects through a study of defects programmed on EUV mask,” Proc.SPIE 11854, 118540Z (2021).
[6] Larissa Juschkin and Daniel C. Wack., “Source performance metrics for EUV mask inspection,” J. MicroNanopatterning Mater. Metrol. 21(2), 021204 (2022).
[7] Renzo Capelli, Nathan Wilcox, Martin Dietzel, Dirk Hellweg, Scott Chegwidden, and Joseph Rodriguez., “AIMSTM EUV first insertion into the back end of the line of a mask shop: a crucial step enabling EUV production,” Proc.SPIE 10810, 108100S, Photomask Technology 2018 (2019).
[8] Tilmann Heil, Michael Waldow, Renzo Capelli, Horst Schneider, Laura Ahmels, Fan Tu, Johannes Schöneberg, and Hubertus Marbach., “Pushing the limits of EUV mask repair: addressing sub-10 nm defects with the next generation e-beam-based mask repair tool,” J. MicroNanopatterning Mater. Metrol. 20(3), 031013 (2021).
[9] Capelli, R., Kersteen, G., Krannich, S., Koch, M., Fischer, L., Roesch, M. and Gwosch, K., “AIMS EUV evolution towards high NA: challenge definition and solutions implementation,” Opt. EUV Nanolithography XXXV 12051, 60–69, SPIE (2022).

Divide et impera: A short talk about the importance of feature and model engineering

Jacob König-Otto*[a], Stefan Meusemann[a]

[a]Advanced Mask Technology Center Dresden GmbH & Co.KG, Rähnitzer Allee 9, 01109 Dresden, Germany

ABSTRACT

The Advanced Mask Technology Center (AMTC) is a joint venture of GlobalFoundries and Toppan Photomasks. AMTC delivers supreme quality photo masks for an enormous variety of different products. To compete with the constantly increasing quality requirements, data pre-processing becomes more and more a key component to reach required yield performances. To reach the quality requirements for the critical dimension (CD) of our masks we predict correction maps in order to improve CD uniformity and a process bias to improve CD mean to nominal (MTN). The process bias is defined as the difference of written structure size to the final structure dimensions on the completed mask. In this talk we present our path to our first machine learning based process bias model with an overall improvement of CD mean to nominal capability of around 30%. We take a closer look on model performance depending on the general model and feature design.

Keywords: Data Science, Machine Learning, Model Design, Critical Dimension, CD Targeting

1. INTRODUCTION

The data flow from incoming customer mask data to a final data set, which is handed over to production, involves multiple fully automated and mask specific correction steps. Depending on the correction target and the process of record for a single mask, the techniques to yield a specific correction artefact can vary from constants and rather simple analytical models up to fully grown machine learning models. With a growing customer base and increasing versatility of products, the latter start to take a growing role.

To reach the quality requirements for the critical dimension (CD) of our masks we predict correction maps in order to improve CD uniformity and a process bias to improve CD mean to nominal. The process bias is defined as the difference of written structure size to the final structure dimensions on the completed mask.

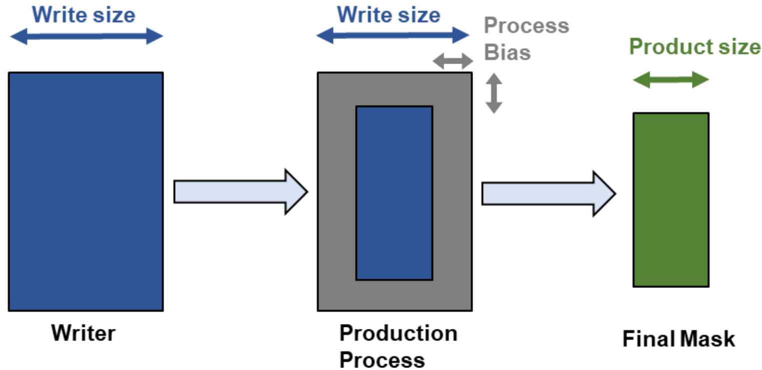


Figure 1. Effect of process bias to final mask dimensions

Whereas we have years of experience with machine learning approaches to create correction maps the process bias (PB) used to be predicted with analytical models in the past. In this talk we present our path to our first machine learning based process bias model.

*jacob.koenig-otto@amtc-dresden.com

38th European Mask and Lithography Conference (EMLC 2023), edited by
Uwe F.W. Behringer, Jo Finders, Proc. of SPIE Vol. 12802, 128020H
© 2023 SPIE · 0277-786X · doi: 10.1117/12.2673254

2. MODEL TARGET AND MODEL DESIGN

2.1 Model target

The main requirement for machine learning techniques is the presence of a huge and self-contained data set. Those are two hurdles for product specific machine learning models in a mask shop. The product volume is comparably low and there are non-neglectable differences between individual masks that have influence on final mask properties. Therefore, next to zero configuration deep learning techniques are out of reach and careful engineering of features and model target is required.

There exists a plethora of variables that can influence the PB. Whereas the unknown obviously can't be included into our model, even known influences can be inaccessible to be employed in a predict, because their specific value is not known at time of data preparation. Though, the core challenge for successful application of machine learning is the identification of accessible covariates and the untangling of the predictable and unpredictable terms of the PB.

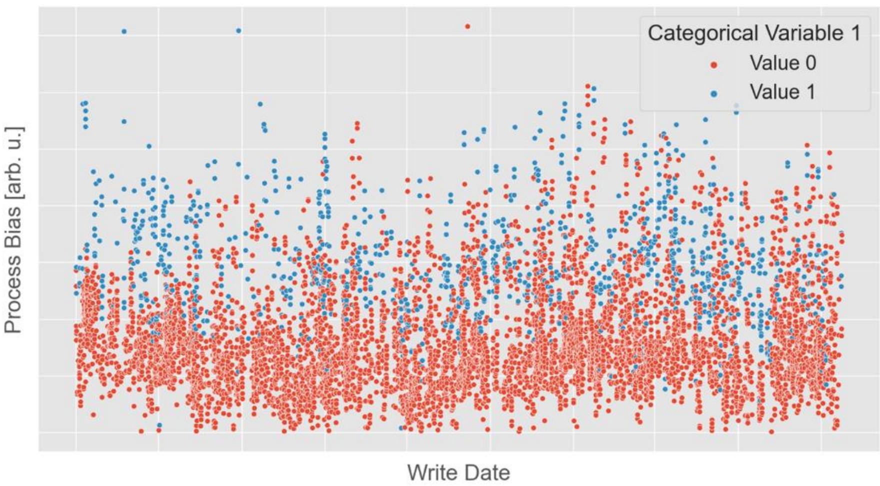


Figure 2. Process bias of different masks over write time.

Figure 2 shows the variation of the PB with time. One can deduce two observations from this plot. Firstly, at a given time point there exists a certain spread of PB. Secondly, the centre of the distribution of the PB is varying with time. There is a notable offset in the PB between the two subgroups of masks with respect to a categorical variable, which explains the spread of PB partially. Since this variable appears to have a direct influence on the PB and the value is known before production, it is an obvious candidate as a covariate for a ML model. The variation of the PB with time can to some extend be explained in hindsight by tool drifts, raw materials et cetera. But this knowledge can not be employed to create a covariate for a ML model to predict the PB in advance. On the other hand, excluding these influences in a model to predict PB, would lead to a probably oversimplified solution with reduced performance.

2.2 Solution idea

Let there be n variables that explain the PB via the function f i.e.:

$$PB = f(x_1, x_2, \dots, x_n)$$

If we assume that there are two subsets of variables x_{ij} and that f can be separated into two independent functions g and h, we can write:

$$PB = f(x_1, \dots, x_n) = g(x_{i1}, \dots, x_{ik}) + h(x_{j1}, \dots, x_{jl})$$

We now reduce complexity for our machine learning approach by using a rolling mean to model the hard to predict function g, depending partly on unknown variables, as a function of time. This is depicted as grey line in the upper panel of figure 3. If we subtract the rolling mean from the PB, we yield a new target for our ML model. The residual shown in the lower

panel of figure 3 is assumed to be depending on an accessible subset of covariates x_j and to be explained by the less complicated function h that can be modelled via machine learning approaches.

Figure 3. Split PB prediction in a classical part (rolling mean) and a ML part (residual).

The decision whether to include a covariate into the machine learning part or to model it outside with an analytical approach has the most significant effect of overall model performance. In a first approach the categorical variable 1 was excluded from the ML part of the solution. The ML model was only trained on data where this variable has value 0. To make predicts for mask where the variable has value 1, the difference of the means of the two distributions was added as a constant offset (see figure 4). Although this approach reduced complexity for the ML part, it had poor performance since categorical variable 1 is not unentangled from other covariates used as ML model input.

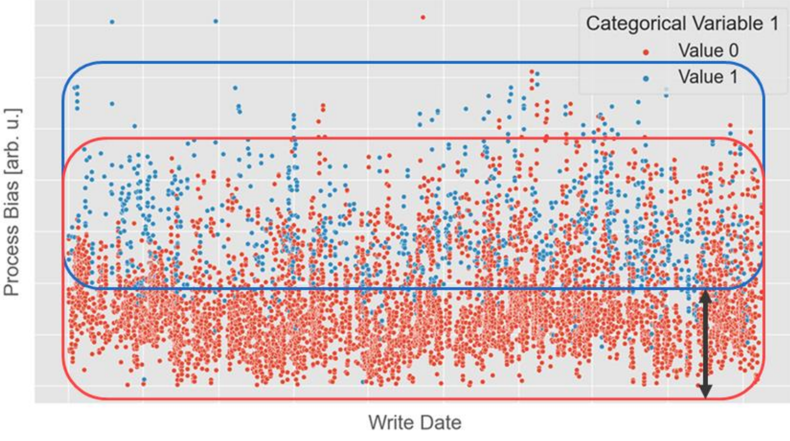


Figure 4. First Approach: Reduce complexity by focusing on one group and model the difference of the two groups via a constant offset (black arrow).

3. MODEL EVALUATION

The final ML model part of our solution was trained on the categorical variable 1 and three additional numerical covariates, which are specific for a given mask. To evaluate a trained model, it is not sufficient to look solely at the ML model part. Instead, the final verdict on the performance of our solution must be drawn by an integrated comparison of the new solution and our old process of record (POR). A first conclusion can be drawn by comparing measured PB with the predicts of the POR and our new model.

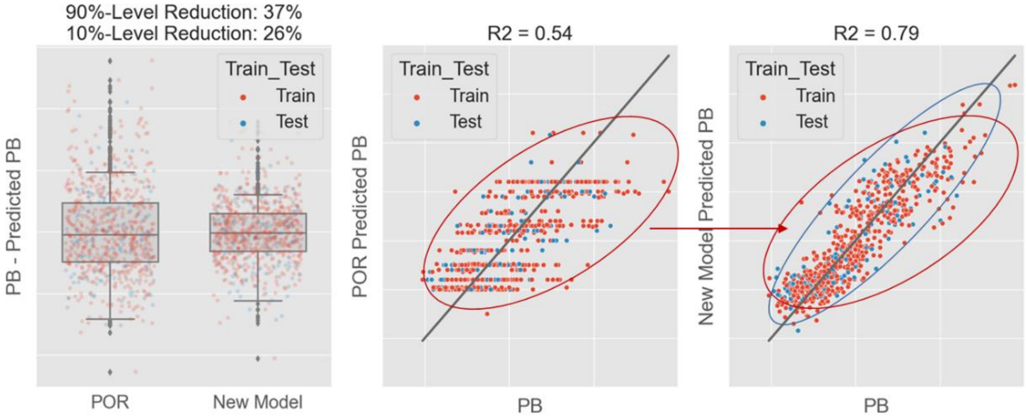


Figure 5. Comparison of measured PB and predicted PB for POR and new model. Ellipses are shown as guide for the eyes to compare the distributions.

The box plots in the left subplot indicate an improved performance of the new model. Comparing the distributions in the middle subplot for POR and in the right subplot for the new solution one can see that the new model explains a bigger part of the variance, which is reflected in an improved R-square value of 0.79 compared to 0.54.

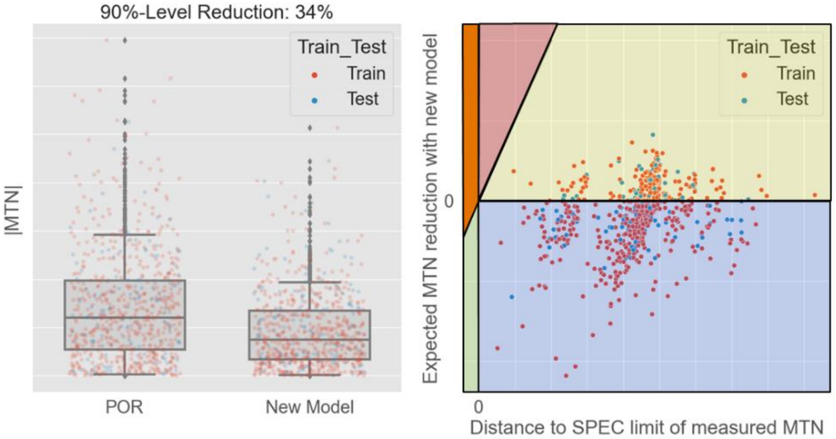


Figure 6. Predicted MTN performance for new model and old POR (left). Detailed insight of model performance (right). Lower half (green and blue area) indicate improvement with new model, upper half (yellow and reddish area) indicate worsening with new model.

As the MTN performance is directly correlated to the quality of the PB prediction, it is no surprise that we see an improvement in MTN capability for our new model compared to POR (see figure 6). The 90% level is expected to drop by 34%. For a subgroup of masks with a special focus the improvement is even higher with a drop of 51%. Although this is a promising result, it is not sufficient as it is lacking insight how the individual mask is improved or worsened. The right

plot in figure 6 shows the difference in MTN on the y-axis and the distance to the respective MTN specification limit on the x-axis. Masks in the blue area would have been improved with the new model. Masks in the yellow area would have seen a worse performance. You can see that some masks are in this yellow area, but the worsening is not high enough that masks go into the red area over unity. In that case, they would not have reached their specification limit. It is to no surprise that not every predict of the new solution will be superior to the old POR, but the expected worsening is still acceptable.

4. REAL PRODUCTION PERFORMANCE

The presented model was deployed to production, which allows us to present the real-world performance and compare it to model evaluation.

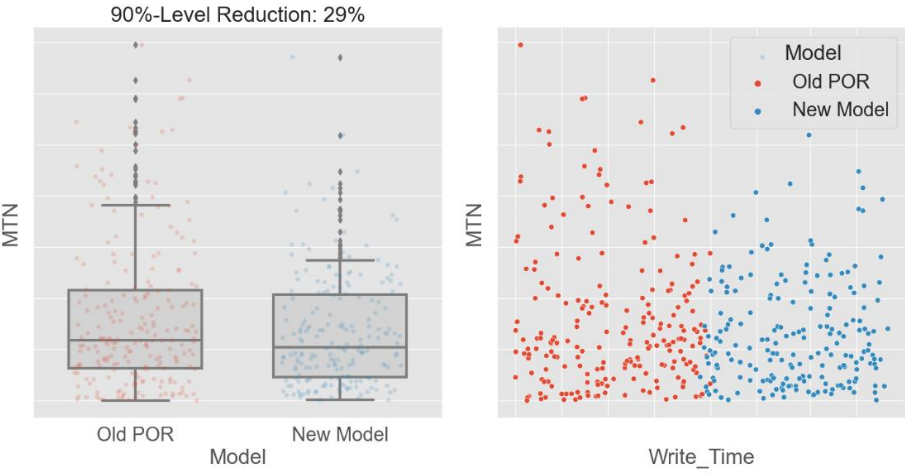


Figure 7. Production performance before and after rollout of new PB model.

The box plot in left panel of figure 7 shows the narrower distribution of the absolute of MTN. This is reflected in a reduction of 29% in the 90%-level. This is in good agreement with the expected reduction of 34% from model evaluation. The right panel shows the distribution of MTN over mask production color-coded by the used tool for PB predict. One can see that the change in performance is abrupt and homogeneous over time. That assures that the increased performance is due to the rollout and not to a general trend.

The model evaluation predicted an even higher reduction of 51% for a subgroup of masks with special focus in the project. Figure 8 shows the performance comparison for this group of masks. There is less data for that group, but the measured reduction of 55% is in perfect agreement with the prediction.

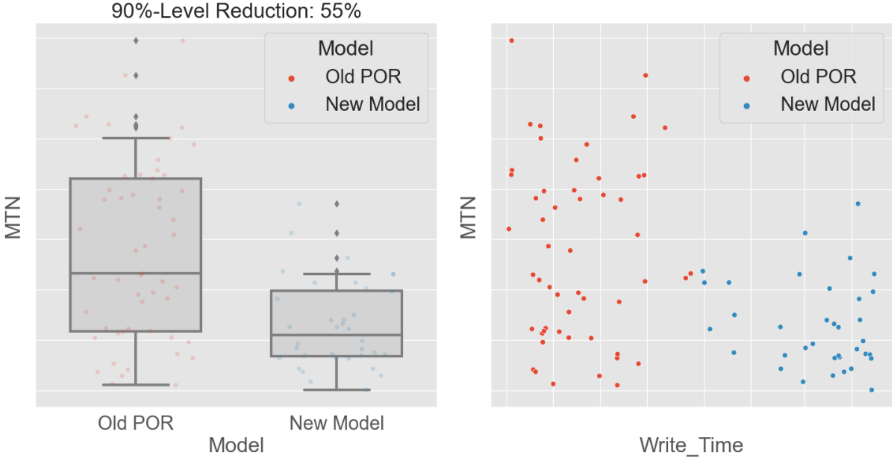


Figure 8. Production performance before and after rollout of new PB model for special focus subgroup of masks.

5. SUMMARY

We presented a new solution to predict PB to improve MTN at AMTC. We reached an improvement in MTN capability in production of 29% in general and 55% for a special focus group of masks. Our solution tackles the challenges of limited data and influences on PB that are inaccessible for ML models, by dividing our solution in a classical statistical part and a ML part. The ML focusses on influences caused by global mask characteristics and the classical part covers route and material related influences. This separation of concerns has the most significant effect on overall performance on the solution.

6. ACKNOWLEDGEMENTS

AMTC is a joint venture of Globalfoundries and Toppan Photomasks, and gratefully acknowledges the financial support by the Free State of Saxony in the framework of the technology grants based upon European Regional Development Funds and funds of the Free State of Saxony under contract number 100392764 (SMART MASK).

Efficient mask characterization through automated contour and corner rounding extraction

Rainer Zimmermann[a], Joost Bekaert[b], Mariya Braylovska[a], Balakumar Baskaran[b], Vicky Philipsen[b], Martin Bohn[a], Michael Bachmann[a], Ulrich Klostermann[a], Eric Hendrickx[b], Wolfgang Demmerle[a]
[a]Synopsys GmbH, Karl-Hammerschmidt-Strasse 34, Aschheim/Dornach D-85609, Germany;
[b]IMEC, Kapeldreef 75, Leuven BE-3001, Belgium

ABSTRACT

Wafer CD Uniformity (CDU) and pattern fidelity are useful properties to monitor when considering yield improvement and scaling to smaller dimensions. Besides control of process fluctuations (e.g., focus, dose), wafer stack film thickness uniformity, and image quality (e.g., contrast), quality of mask manufacturing and OPC models are essential to optimize these properties. Therefore, a proper characterization of the written mask dimensions is becoming more and more important, especially as the mask pattern complexity increases through Inverse Lithography Technology (ILT) as well as the need for curvilinear or all-angle designs.
Applying contour-based mask quality assessment instead of traditional gauge-based characterization of mask dimensions allows to intrinsically capture mask imperfections like corner rounding (CR) of the absorber for complex shapes which would be hard to measure and categorize by using only a few gauges.

Thus, in our study, we examine ways to use contour-based mask characterization methods, including CD and area uniformity, linearity, and CR determination to evaluate mask quality. We present a method and flow to automatically extract contours and determine values for mask CR from top-down mask SEM images. Contour-based metrology and data evaluation is then used to quantitatively address the above-mentioned mask properties of interest. Finally, as an initial approach to investigate impact of mask quality on wafer printing, we apply a generic EUV model to ideal and imperfect masks and compare the simulated contour results. Using realistic mask patterns for lithography modeling and simulation is considered essential to achieve the required accuracy for advanced nodes and technologies.

Keywords: contour, mask, corner rounding, SEM image, lithography, modeling

1. INTRODUCTION

Advances in pattern fidelity and process control, as reflected by Edge Placement Error (EPE) and CDU, are key enablers for yield improvements and scaling to smaller dimensions. Thus, mask dimensions characterization and control are increasingly important[1-4]. To reduce non-uniformities on wafer, it is crucial to ensure required quality and uniformity already on mask patterns by appropriate control of the mask manufacturing process. In addition, precision of pattern fidelity in wafer printing is strongly dependent on quality of model-based mask operations like OPC or ILT. This is even more true as we are now dealing with enhanced mask complexity with the need to handle all-angle or curvilinear designs (Fig. 1). Thus, in addition to high-quality mask manufacturing itself, lithography models taking realistic mask patterns into account are essential for correct prediction of wafer printing (as required for OPC or ILT). Consequently, model calibration needs to be based on realistic mask patterns revealing effects like CR (reflecting the physical limitations within the mask writing process) instead of ideal (target) shapes, especially when using contour-based model calibration.

In this paper we present flows and methodologies based on contour extraction (CE) using Synopsys S-Metro[5], applied to high quality SEM images from EUV masks, which allow enhanced characterization of the mask quality and CR of complex shapes, rather than relying on pure gauge-based assessment.

The structure of the paper is as follows: In the next section, we shortly describe the experiments and applied evaluation procedure. In the main part of the paper, we explain the methods and results from different contour-based mask quality assessments including following aspects:

- Mask CDU extraction on contact arrays and comparison to CDU measured from mask SEM images. Corresponding lithography simulation could then be used to separate the origin of wafer CD non-uniformities as

38th European Mask and Lithography Conference (EMLC 2023), edited by
Uwe F.W. Behringer, Jo Finders, Proc. of SPIE Vol. 12802, 128020I
© 2023 SPIE · 0277-786X · doi: 10.1117/12.2675486

coming from either mask variations or stochastic effects on wafer, which is of significant importance in EUV lithography.

- Determine area uniformity within contact arrays.
- Measure CD linearity from contours at line-space (L/S) patterns and compare with experimental results.
- Extract CR from the mask SEM images and discuss ways to use the results for (contour-based) calibration or simply application of a lithography model.
- Use CE across die locations on a random pattern to evaluate mask stability or variability.
- Initial steps to a methodology to assess impact of mask imperfections on wafer results using a generic lithography model.

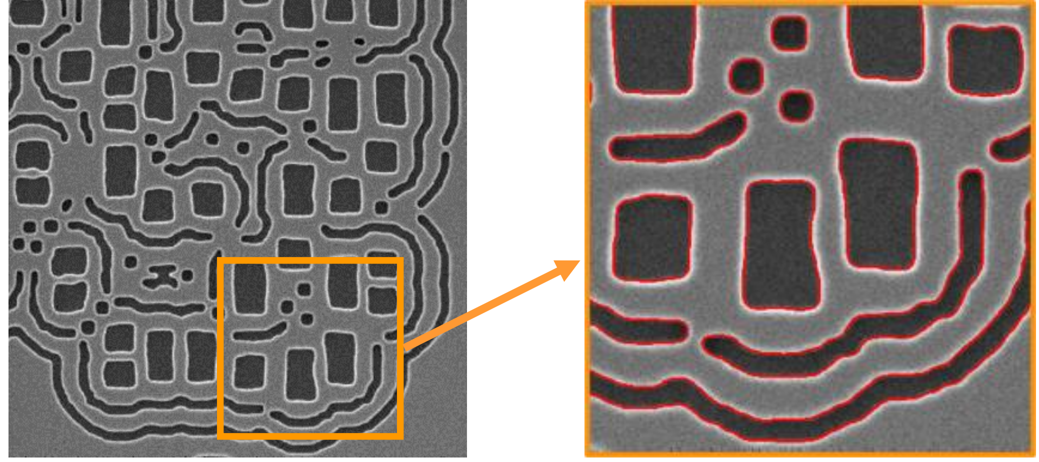


Figure 1: SEM image from an ILT-generated curvilinear mask with contact-like main shapes and Sub-Resolution Assist Features (SRAF). The detailed view shows part of the image with extracted mask contours in red.

2. EXPERIMENT AND EVALUATION PROCEDURE

2.1 Experiment

As mentioned before, the procedures presented in this paper to evaluate mask quality are based on contours extracted from mask SEM images. The images were measured on test patterns (TP) with a variation of pattern type (line-space, contact hole), dimensions (CD, space, pitch), SRAF placement, orientation, and tonality. The smallest applied CD and space is 10nm (40nm on mask). A mask containing the TP was written with a Multi-Beam Mask Writer (MBMW). SEM images and mask CDs from a selection of the full TP portfolio were measured based on an appropriate metrology plan using the Advantest E3650 mask CD-SEM tool. We restricted the SEM image settings to 1024 pixels in X and Y direction, and the Field of View (FoV) in most cases to 1.44µm (mask dimension), corresponding to 100k magnification. In a few cases, 2.88µm was used (50k magnification).

2.2 Evaluation procedure

Contour Extraction (CE) was done with Synopsys S-Metro[5], allowing to use an appropriate CE mode for best quality contour results. Based on the extracted contours, we used S-Metro also for further data generation as needed. This step mainly included CD measurements, area measurements, and CR extraction steps, required to evaluate the following mask pattern properties:

- CDU
- Area uniformity
- Linearity
- Corner rounding
- Local and die variability

- Lithography simulation

The applied evaluation flow based on S-Metro is shown in Fig. 2. S-Metro functionalities take SEM images and related image data files as main input, which are imported in the first step of the flow, together with the mask layout. An initial CE is done from the SEM images, without the need to refer to the related mask layer. The initially extracted contour is then aligned with the mask layer, and this alignment then allows for improved CE using a layout-guided mode.

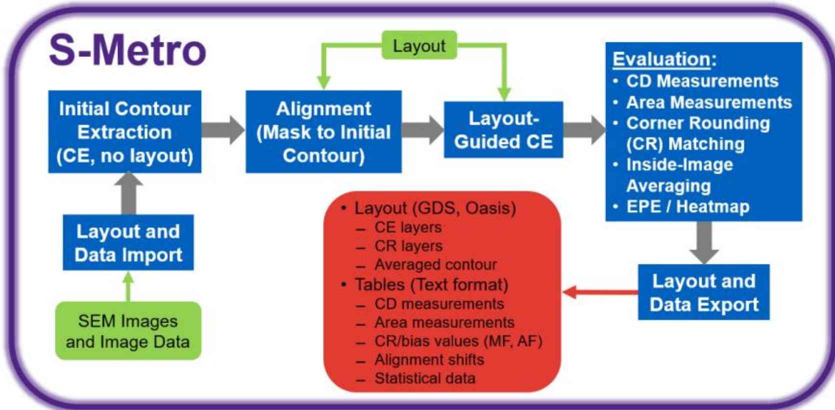


Figure 2: CE and evaluation flow applied in this work using Synopsys S-Metro. Note the Alignment step between first and second CE (see Fig. 3).

The extracted contours, together with the aligned mask layer can then be used for further qualitative and quantitative analysis, which is needed to create the layers and data required for assessment of the above-mentioned mask pattern properties. Supported evaluations are for example CD and area measurements, CR matching, contour averaging inside the FoV of an image, and representation of results as heatmaps or EPE statistic plots. Finally, all data and created layers (including the extracted contours) can be saved and exported in tables and GDS or Oasis layers, ready for further evaluation steps. Fig. 3 shows a part of the ILT mask SEM image given in Fig. 1, together with the mask (green) and extracted contour (red) layers before and after alignment.

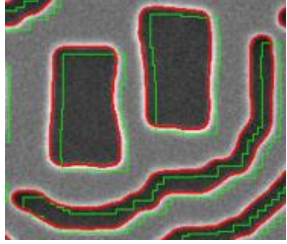
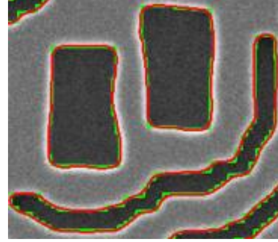
 

Figure 3: CE in red before (left) and after (right) alignment between SEM image and target features (green) in Synopsys S-Metro. Alignment allows for enhanced CE and is required for subsequent postprocessing steps.

3. RESULTS

We now present the evaluation results from a series of CE-based applications, as announced in the previous section. For each topic, we select some representative examples to point out the key aspects of the application. Note that all given dimensions are referring to 1x wafer dimensions, if not explicitly stated otherwise.

3.1 CD uniformity on contact arrays

We start with a comparison of CDU extracted on contact arrays, using the experimentally measured SEM CDs on one side, and the CD measurements based on extracted contours on the other. We consider this a valid method to verify whether CE is sufficiently accurate to be used for lithography model calibration.

Using S-Metro, we extracted contours from SEM images of contact arrays containing nominally identical contact hole shapes, and looked at variants regarding target CD, pitch, and designed CR. Within the CE tool, we then measured contact CDs on the extracted contours in horizontal and vertical direction and compared the results with the corresponding SEM-based measurements. These extracted and experimental CDs are then used for CDU evaluation. For simplicity, evaluation is restricted to 5x5 contact holes around the center of the FoV.

Fig. 4 shows an example of measurement locations from an SEM image of a contact array and the corresponding measurement gauges as used for CD determination based on the extracted contours. To mimic the SEM measurements, the final CD measurement value is the average of measurements around the center gauge covering a box with height of about 50% of the contact edge length.

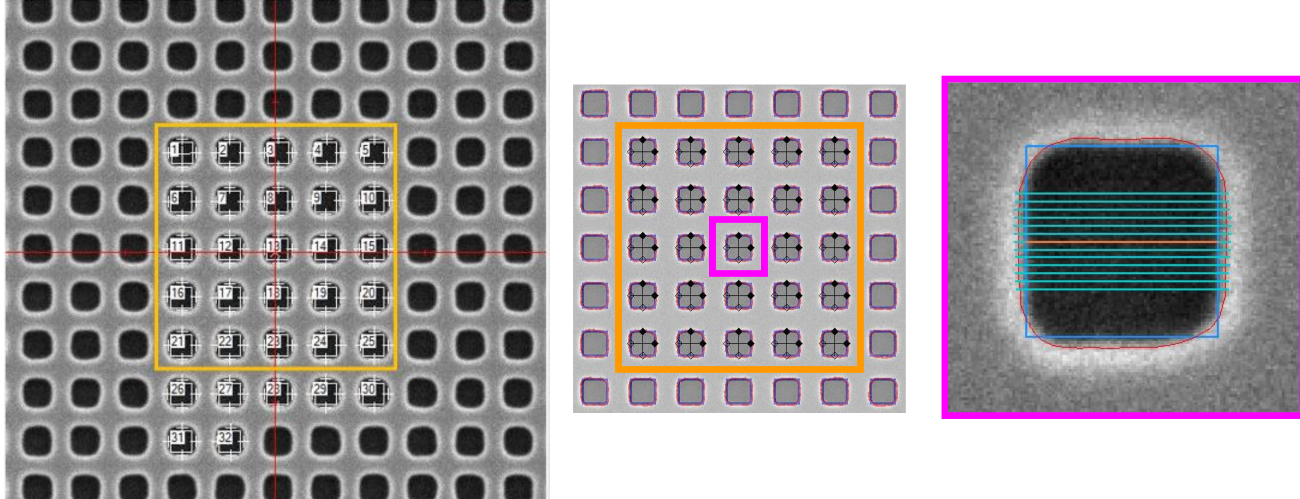


Figure 4: Measurement locations for CDU evaluation from a SEM image of a contact array. Note horizontal and vertical measurement gauges on 5x5 selected contacts. Left: Measurement locations in the SEM. Center: Contacts selected for the contour-based CD measurement. Right: Detailed view of the S-Metro CD measurement locations (→ averaged CD) together with the layout (blue) and extracted contour (red).

We now discuss two representative results. The plot in Fig. 5 shows the difference between the CD from the extracted contours and the CD from the SEM for 3 different 1:1 pitches (28/36/48nm), measured at the contacts shown in Fig. 4. For the three pitches, the variation of the differences is not larger than 0.3nm, meaning that CDU from CE and SEM agree well. This is also supported by the fact that the RMS values of the individual measurements match well as shown in the two boxes within the plot.

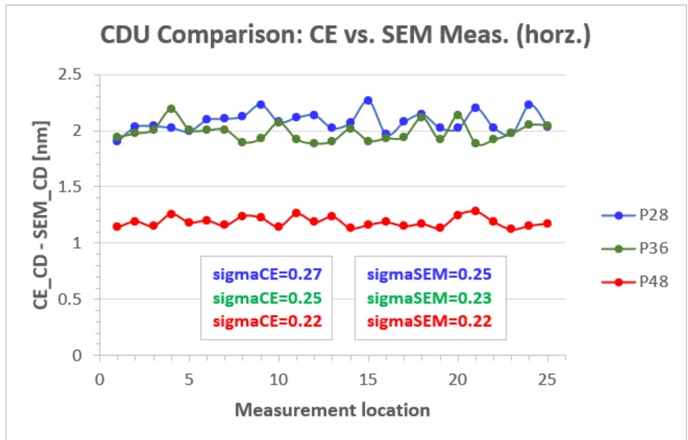


Figure 5: CDU comparison using measured and CE-based CDs for contact arrays of different 1:1 pitches. Boxes show excellent agreement between sigma values of the discrete measurements.

However, we also realize that there is an offset between CDs taken from extracted contour and the one measured with the SEM tool, in the range of about 1-2nm. This is most likely caused by the fact that CE and SEM do not use the same CD assessment methodology. First tests have shown that the offset can be minimized by tuning CE parameters or by aligning the CD assessment methodologies.

As second example of CDU evaluation we look at contact arrays of pitch 56nm, mask CD 20nm, with different nominal corner rounding applied on the mask. The bottom of Fig. 6 shows the representative contact pattern without CR (blue), and with CR (red) of 6nm (left) and 10nm (right) applied, the latter leading to a circular contact shape. We only consider the contour-extracted data (no SEM measurements) and plot the deviation between measured CD and target CD for the three CR cases, separated for horizontal (left) and diagonal (right) measurement. Note that "targetCD" refers to the actual target including the applied CR, meaning that for the diagonal measurement our reference is the red line in the contacts shown at the bottom, and not the square (blue) contact.

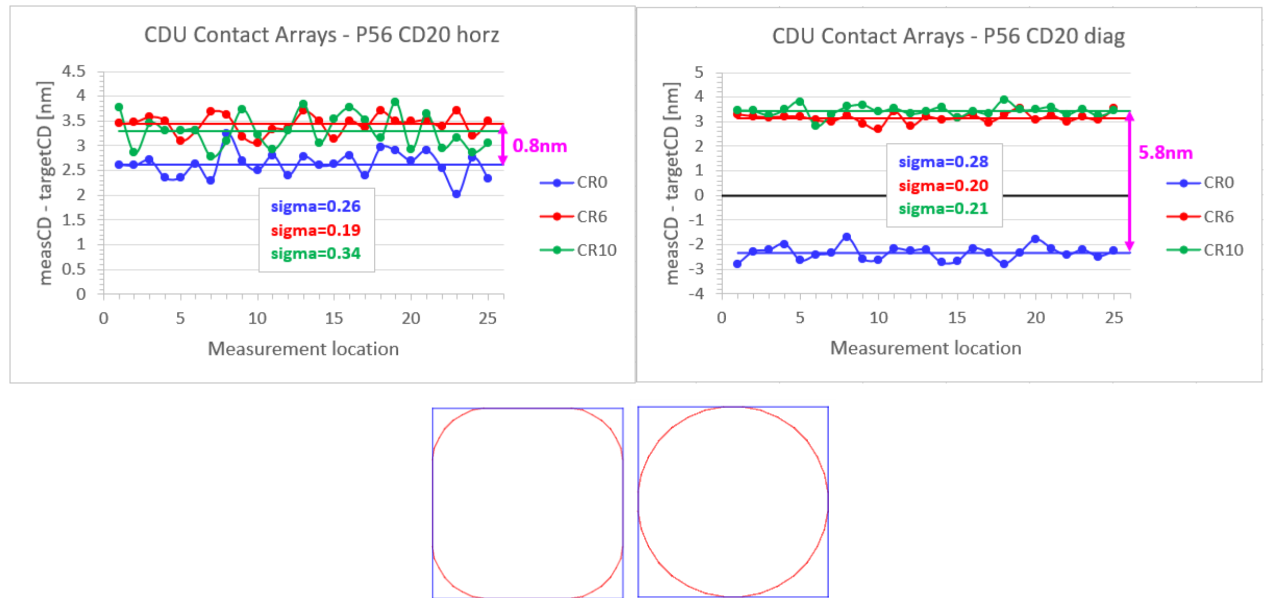


Figure 6: CE-based CDU plots for Pitch 56nm, CD 20nm contact patterns with CR 0/6/10nm applied, as shown at the bottom. Left/right plot correspond to horizontal/diagonal measurement, respectively. No measurement averaging across locations was used. Colored horizontal lines represent mean CD values. Sigma values are given in the boxes.

For the horizontal case we find a small offset (≈0.8nm) between CR0 and CR6 or CR10 cases, while for the diagonal case the offset between CR0 and CR6 or CR10 cases is much larger. Of course, this is a direct consequence of the CR as occurring in the SEM image: If no CR is used in the mask (CR0 case), the physical limit of the mask making process results in the observed difference in diagonal direction. As soon as a CR is applied to the mask, and it is beyond the physical limit of the process, no dependence of the CE-based CD on the CR value is observed, as seen in the right plot where the red and green graphs match quite well.

Overall, our examples show that CE is an appropriate means to evaluate CDU-related properties, because extracted contours correlate well with the CD behavior as measured from SEM images.

3.2 Area uniformity on contact arrays

We now use extracted contours to investigate area uniformity (AU) of contacts within a contact array, similar to what we did for CDU in the last section. Interest in AU is increasing in the frame of AR/VR device manufacturing, as it turned out that pattern area is even more important for the optical device performance (e.g. in meta lenses) than pattern fidelity[6]. We apply CE within contact arrays for different contact types (e.g. variation of CD, pitch, tone, …). Then we determine the areas of the individual contacts of the arrays using appropriate gauges (see Fig. 4, only 1 gauge per contact required).

Results are shown in Fig. 7. In the left graph the extracted areas from the 25 contacts within the array are plotted for different 1:1 pitch pattern, ranging from pitch 28nm to 48nm. The extracted areas are flat across the measurement

positions for each pitch. In addition, the plot on the right side compares the area as measured from CE to the area calculated by just multiplying the extracted CDs in x and y direction of the individual contacts. As expected, the area measured from CE is smaller than the calculated area, clearly representing the effect of CR. The deviation is slightly increasing with increasing pitch.

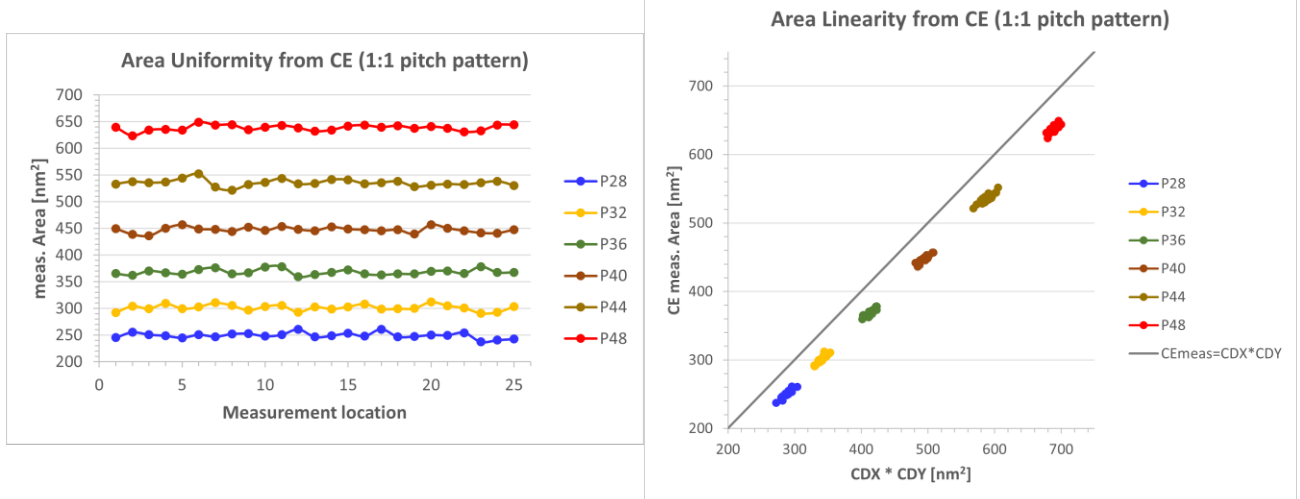


Figure 7: Area Uniformity (left) and Linearity (right) from CE for different 1:1 pitches.

Now we examine AU dependence on mask CR, similar to what we did in previous section for CDU (see Fig. 6). Snapshots from the contact arrays of the pitch 56nm, CD 20nm patterns with different CR applied are shown in Fig. 8, where we overlay the SEM images with extracted contours (red) and the corresponding masks (yellow).

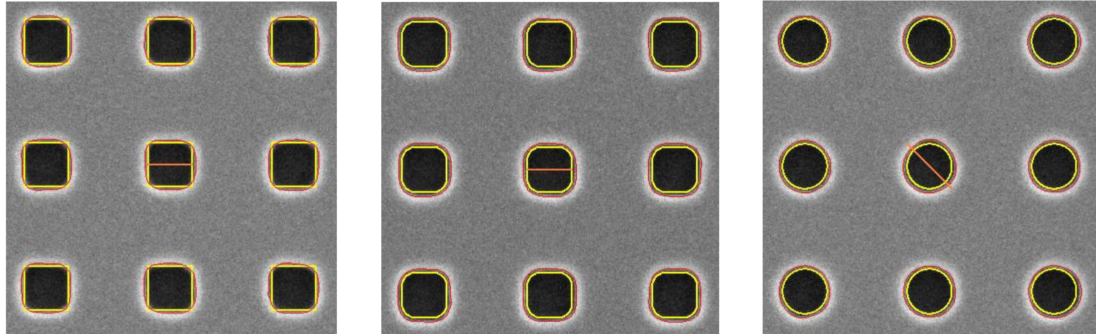


Figure 8: SEM images, CE (red), and target layers (yellow) for pitch 56nm, CD 20nm patterns with target CR 0/6/10nm applied (from left to right).

The result is shown in Fig. 9, where we plot the difference between the areas as measured from the extracted contours and the areas as calculated for the different corner-rounded targets. Similar to Fig. 6 for CD, "targetArea" refers to the actual area calculated from the ideal mask pattern (i.e. square for CR0, corner-rounded square for CR6, and circle for CR10). In all cases we find some variation in area uniformity, demonstrating the capabilities of this assessment methodology.

The second message from Fig. 9 is, similar to the case of CD measurements (Fig. 6), that it looks like the deviation from the corresponding (corner-rounded) target is no more affected once the applied mask CR is beyond the physical limit of the mask making process. AU for CR0 (blue) is clearly separated from the overlapping AU curves for CR6 (red) and CR10 (green). This implies that there is no CR contribution coming from the mask making process which is not already taken into account by the CR applied to the mask pattern. In other words: Applying sufficiently large CR to the mask pattern allows user a precise control of the area fidelity, as no additional uncontrolled mask CR effects from mask making process are to be expected.

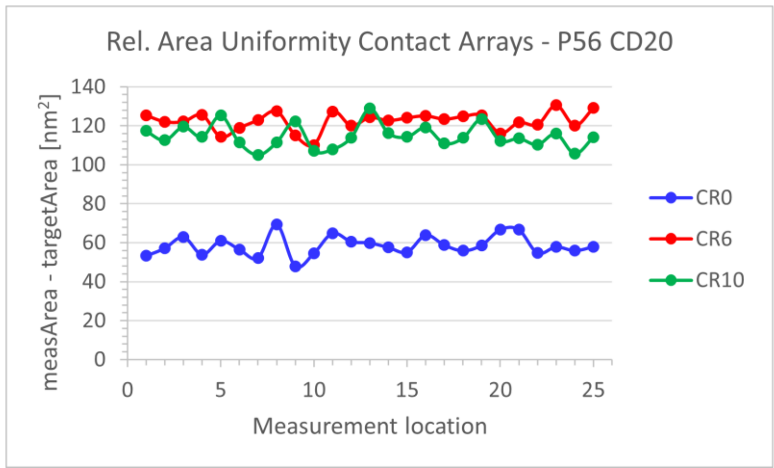


Figure 9: AU plots for pattern from Fig. 8 across 25 measurement locations.

3.3 Linearity and HV bias at line-space patterns

Now we switch from contact holes to line-space (L/S) patterns, and compare CDs measured from CE with SEM-based CD measurements. We study linearity and horizontal-vertical (HV) bias with the intention to clarify whether extracted contours are appropriate to be used for application with modeling and simulation tools.

We take CD measurements from SEM images on horizontal and vertical L/S patterns for constant pitch 80nm with variation of CD, and for lines with designed CD of 16nm and 32nm through pitch. Running CE on these patterns, contour-based CDs are determined as well by averaging 21 individual measurements along a range of 100nm (Fig. 10). Based on SEM- and CE-CDs we evaluate and compare linearity and HV bias between both measurement approaches.

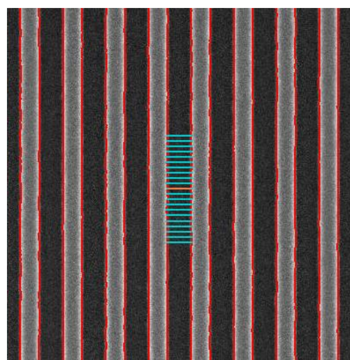


Figure 10: SEM image from L/S pattern with CE (red) and measurement locations (blue) arranged along 100nm. Max. sigma found from all measurements is 0.44nm.

Linearity results are shown in Fig. 11, where we plot the measured CD coming either from SEM or from CE against the Design CD for 80nm pitch pattern. We can only see the data for the vertical case (red and orange) because they perfectly overlap those for the horizontal case (blue and light-blue), indicating a negligible HV bias. In addition, we find that there is a small offset between the CDs from SEM and those from CE, confirming what we already observed for the CDU measurements (section 3.1). Measured CDs are slightly smaller than mask (Design) CDs, but the behavior for SEM-based and CE-based linearity is identical.

In Fig. 12 we plot CDs measured from SEM and CE for 16nm and 32nm horizontal and vertical design lines through pitch. In both cases, we find that no HV bias is seen (neither in SEM data nor in CE data), but there is a slight offset of about 1.2nm between CDs measured from SEM and those extracted from the extracted contours. This agrees with what we already saw in the linearity plot (Fig. 11), and supports the findings already observed in the CDU evaluation (Fig. 5).

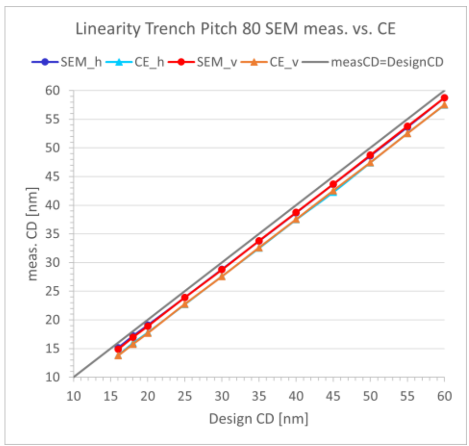


Figure 11: Linearity plot for pitch 80nm L/S patterns from 16nm to 60nm (1x), horizontal and vertical orientation. Both SEM plots and both CE-based plots are well aligned, so the horizontal data (blue) are hardly visible.

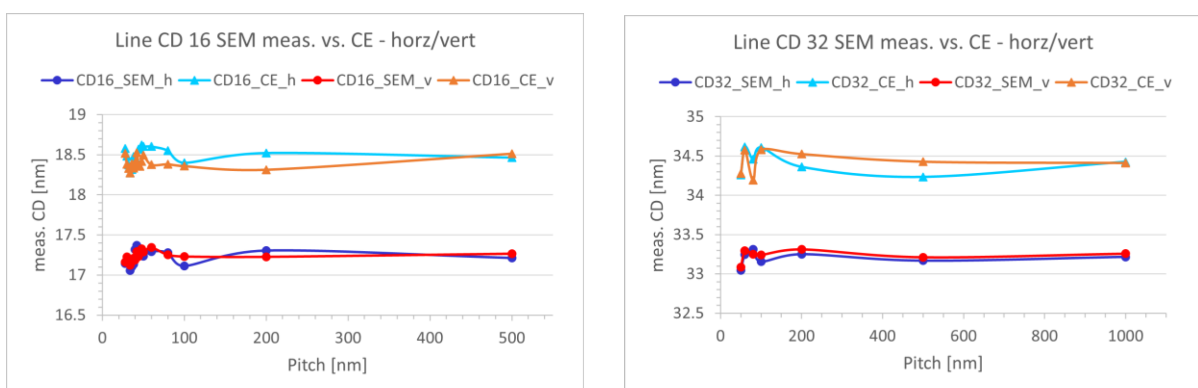


Figure 12: SEM-based and CE-based CDs for 16nm (left) and 32nm (right) horizontal and vertical design lines along pitch.

3.4 Mask corner rounding study

We now have a closer look at mask CR by using Synopsys S-Metro to find mask CR best matching the SEM. The use case we have in mind is to apply detected mask CR in lithography simulation and modeling.

To achieve this, we run CE on a variety of contact patterns, including SRAFs, as shown for example in Fig. 13. Based on the extracted contours, we determine and evaluate CR and export the layer with applied CR to GDS or Oasis layout. We then use heatmap and EPE distribution plots for quality assessment, and compare the results to intra-image averaging, another feature provided by the CE tool.

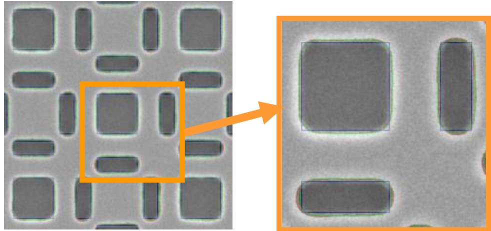


Figure 13: Example showing CR matching result from a staggered contact array with SRAFs. Target layer in blue, target layer with applied CR in green, and CE layer in red are overlaid to the SEM image.

Results are presented in Fig. 14 for a series of 1:1 pitch pattern in the left plot, and the nine combinations of CD and Space using 22.5/50/175nm for positive contacts (blue, light-blue) and negative, i.e. inverted contacts (red, orange). We

evaluate both, contacts and inverted contacts, to compare CR values for pattern of nominal identical sizes, but with outer corners in one case and inner corners in the other. Both plots reveal that the extracted CR and accompanying biases are quite stable across the pattern types and sizes. Inner CR (red curve on the right) was found to be about 1nm smaller than outer CR (blue) in given example, while biases seem to be similar for both.

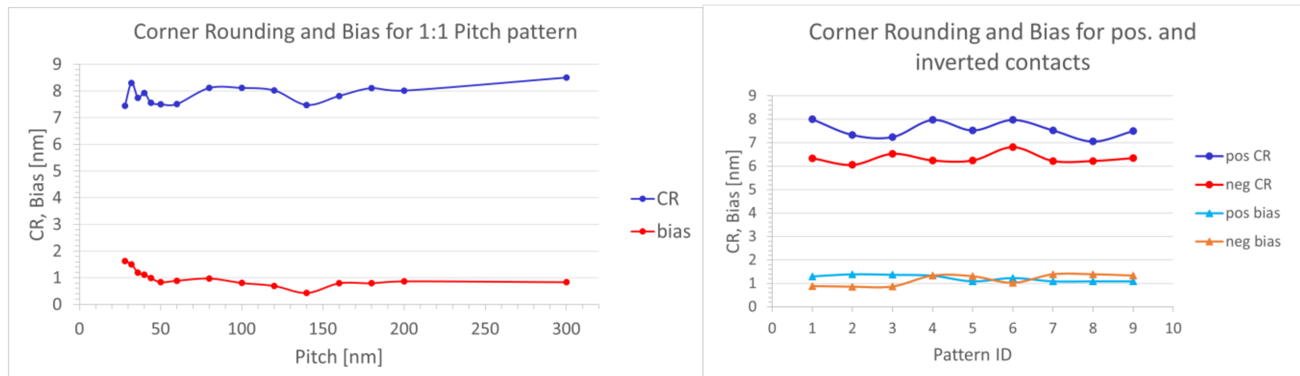


Figure 14: Left: CR/bias found from 1:1 pitch pattern contours. Right: CR/bias found from contours on contacts and inverted contacts, for the 9 CD and Space combinations based on 22.5/50/175nm.

Now looking at CR extraction within a contact array, we detect a non-uniformity of the individual contacts (Fig. 15). In the top part of the figure, we present CE and CR layer overlaid to the SEM image and original target layer, where the part on the left side refers to a case where the extracted CR layer shows better agreement to the original CE layer than the case presented on the right side. This becomes more obvious when looking at the EPE heatmap plots in the bottom part of the figure, where we clearly see a larger deviation (highlighted by red heatmap color) in the upper left corner of the contact for the right case as compared to the left case. However, the EPE RMS and sigma values determined across all patterns of the array are still quite small.

As another means to evaluate pattern uniformity within a contact array, we take a closer look at the averaged contour extracted from the FoV of the array. In Fig. 16 we present the images corresponding to those in Fig. 15, but now using the averaged CE layer instead of the CR layer for comparison. We find that the averaged contour seems to be smoother regarding EPE deviation from exact extracted contour, since the EPE heatmap does not show red areas. However, comparing the EPE distribution plots in Figs. 15 and 16 it becomes clear that the difference is actually quite small.

As follow-up we intend to evaluate the sensitivity of lithography modeling to the observed pattern non-uniformity within a contact array, for example to see whether the observed deviation found from the CR is more critical than using averaged or even exact CE for model calibration.

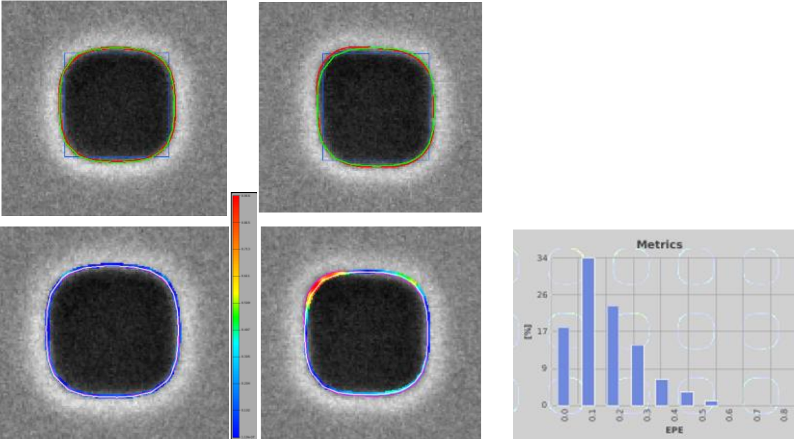


Figure 15: Top: SEM images, target (blue), CE (red) and CR matched (green) from a Pitch40 CD20 pattern. Left good match, right worse match between CE and CR layers. Bottom: Corresponding heatmap images. Heatmap legend scales with increasing EPE from blue to red. Corresponding EPE distribution shown on the right with EPE RMS=0.188, sigma=0.121 (determined across all patterns).

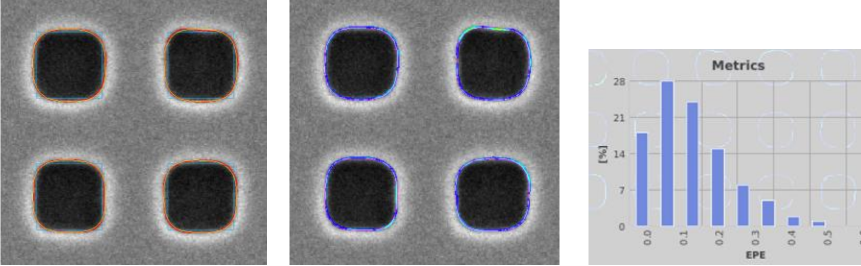


Figure 16: Left: Pitch40 CD20 pattern region with averaged contour (yellow) and extracted contour (red). Right: EPE heatmap and distribution images for matching of averaged contour (white) to extracted contour (magenta). EPE RMS=0.152, sigma=0.099 (across all patterns).

3.5 Variability study

In this section we present a method to use CE and related CD measurements to investigate the stability of the mask making process. We placed a random test pattern repeatedly across a full die and extracted the contours based on the SEM images. This is illustrated in Fig. 17, where the left picture shows a section of the SEM image from the pattern of interest, together with the extracted contour, and the center sketch displays the measurement regions on the die (upper left UL, CENTER, lower right LR). To assess local variability in addition to within-die variability, we repeated measurements four times respectively at UL, CENTER, and LR regions. To avoid charging effects, the repeated measurements were done at locations offset by 1 or 2 pitches between each other. Finally, to quantitatively evaluate the differences, we placed some random gauges across the patterns where CE-based CD measurements were done (right image in Fig. 17).

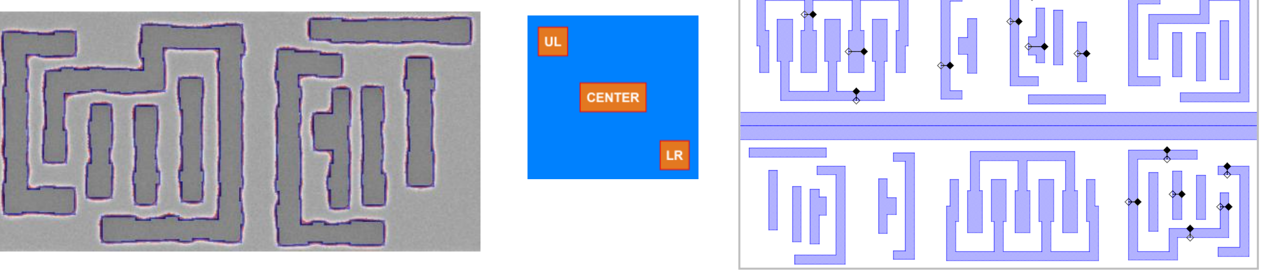


Figure 17: Left: SEM image, mask (blue), and CE (red) from a random pattern. Center: Pattern placements on the die. Each die location involves four repeated measurements. Right: Test pattern design with gauge locations for CE-based CD measurements.

Fig. 18 shows the overlay of four contours as extracted locally at CENTER region from the corresponding SEM images. Obviously, there is no remarkable difference seen between the four locations. Similarly, we find in Fig. 19 that the contours extracted at equivalent locations from UL, CENTER, and LR part of the die match very well.

Finally, we compare in Fig. 20 the CE-based CD measurements taken at the locations shown in the right part of Fig. 17. Obviously, and corresponding to the results from extracted contours, these CD measurements also match quite well across die (left plot) and locally (right plot). Note that the lines between the individual measurements are just eye guides.

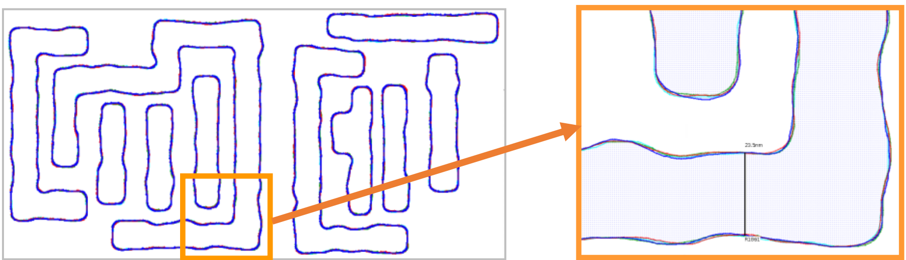


Figure 18: Overlays of locally extracted contours. Different colors correspond to the four slightly different locations at die CENTER (see Fig. 17).

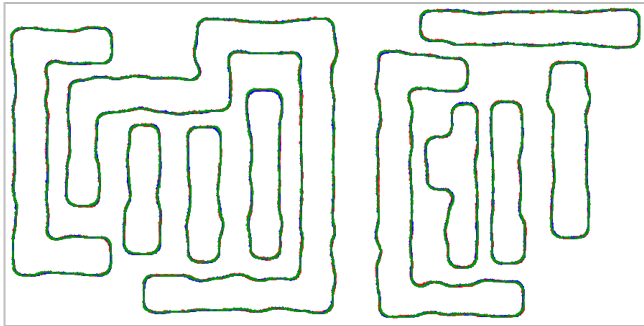
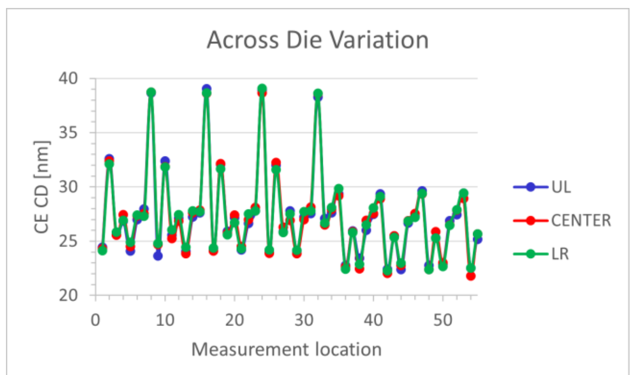


Figure 19: Results of contour overlay within die (UL vs. CENTER vs. LR, represented by different colors), taken at equivalent local positions.



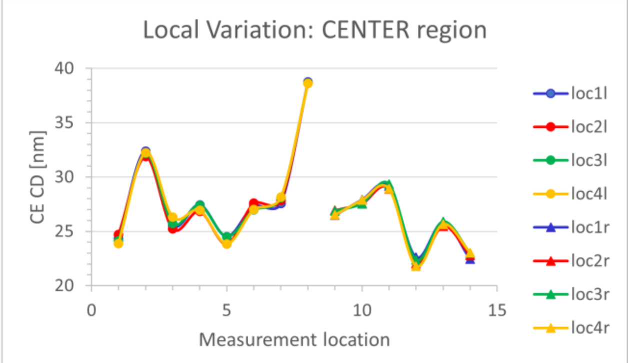


Figure 20: CD measurements taken from extracted contours. Different colors correspond to different locations. Left: Variation within die. Right: Local variation, using center die region as example. In both cases, no relevant differences could be found.

In summary, we show in this section that CE and CE-based CD measurements are appropriate means for assessment of required data to evaluate mask stability locally and within die.

3.6 Initial assessment to lithography simulation

In the last section of this paper, we are starting to look into methods to use extracted contour and evaluation layers to investigate impact of mask imperfections (e.g., CR) on wafer printing. The intention is to assess a methodology of using CE results in lithography model calibration, both considering compact models as applied in OPC or ILT flows, but also rigorous models as for example calibrated with Synopsys S-Litho™ [7] with enhanced options to examine physical effects not covered by speed-optimized OPC models.

For the lithography simulations we prepared a generic Synopsys Proteus™ [8] EUV model. If needed, we adapted the threshold as to provide acceptable resolution for the patterns of interest within a given experiment. In order to evaluate impact of mask differences on wafer results, we use S-Metro to run CE on mask images, and then export the extracted contours as well as contours averaged across image FoV as layout (GDS, Oasis) layers. In addition, layers with extracted mask CR are exported as well. Then we use these different layers as mask layers in the lithography simulation applying the EUV model to create the lithography wafer contours, which can now also be exported as layout layers and overlaid for comparison. This experiment was run on a variation of patterns.

Fig. 21 depicts a snapshot where the lithography simulation contours based on ideal OPC mask (no CR or CD bias) and CE mask are overlaid. The differences are obvious.

In Fig. 22 we overlay the litho simulation contours and corresponding mask pattern of contact and inverted contact patterns for different target CDs. We compare the results for ideal (square shaped) mask pattern, CE mask pattern, and average CE mask pattern. We observe that the litho simulation contours for averaged CE layer and exact CE layer are very similar, and (as expected), the difference to the result from ideal mask decreases with increasing target CD.

Figure 21: Overlay of lithography simulation contours from ideal OPC mask (red) and CE mask (green). The thick lines representing the corresponding lithography contours show some differences.

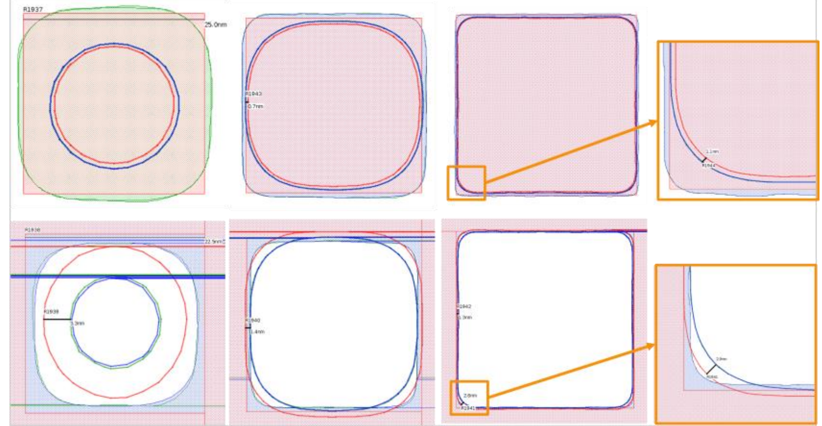


Figure 22: Overlay of Litho simulation contours created from ideal mask (red), CE mask (green), and CE averaged mask (blue) for qualitative analysis on pos. (top) and inverted (bottom) contact patterns. Thick lines represent the corresponding litho contours. Positive pattern CDs 25/50/175nm, negative (inverted) pattern CDs 22.5/50/175nm.

Finally, in Fig. 23 we add the corner-rounded mask layer to the comparison of the results for compact holes of size 50nm and 175nm. In addition to the good agreement between CE mask and averaged CE mask, we now also find that the lithography simulations between averaged CE and CR masks match nicely.

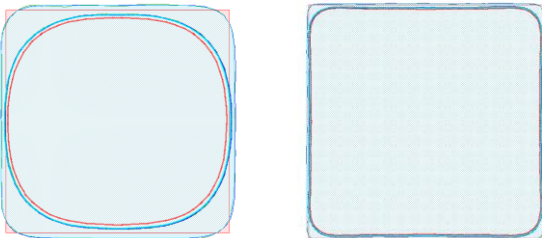


Figure 23: Overlay of all masks and their simulations (ideal red, CE green, averaged CE blue, CR light blue) for CD 50nm (left) and 175nm (right). Litho contours from averaged CE and CR layer are nearly identical in this case.

In short, we showed CE and related evaluation being appropriate to assess impact of mask differences or imperfections on wafer results. In a follow-up work we'd like to investigate sensitivity of model calibration to these differences.

4. CONCLUSION AND OUTLOOK

4.1 Conclusion

The Synopsys S-Metro tool was used to extract and evaluate contours from high-quality mask SEM images, taken with the Advantest E3650 mask CD SEM tool on a variety of patterns with the intention to assess pattern uniformity and CR. We proposed ways to use the results to investigate mask quality determining CD and area uniformity, linearity, CR, and impact on the wafer result by lithography simulation using a generic EUV model.

Our results imply that the available precision of CE and subsequent evaluation of related data is an appropriate and mandatory means to investigate mask quality and imperfections and their impact on wafer printing. For example, the CD uniformity and linearity measured from extracted contours nicely match the ones as measured with the SEM tool. Area uniformity assessment and averaging from contact arrays was demonstrated and proposed as another means to examine mask quality. Given CR examples imply that CR is representative for the mask manufacturing process, rather than for the patterns themselves. Preliminary investigation of impact of mask imperfections on wafer printing results implies that a sensitivity study of mask quality on wafer lithography simulation and corresponding modeling are of specific interest, especially since extracted contours can be exported to GDS/Oasis for subsequent processing or direct usage in model calibration/simulation tools (esp. important for complex patterns).

All work is supported by automated CE and CR determination, metrology, analysis, and visualization within acceptable time using distributed processing.

4.2 Outlook

Regarding simulation and modeling, instead of only using wafer simulations like in current work, direct evaluation of the impact of mask imperfections on calibration of a lithography model is of strong interest. For example, we may use the ideal (as-designed) mask as input to run the model calibration, and then, for comparison, use the CE contour layers from SEM images. As an intermediate approach we may simply apply extracted mask CR to the input mask for modeling. We know that by using more realistic mask patterns the calibrated lithography model will reveal better predictability than a model based on perfect masks, and CE provides another useful means to validate this fact.

Finally, we would like to use mask CE to investigate another interesting aspect of mask manufacturing and quality becoming more and more important for advanced nodes, namely the impact of mask proximity correction on lithography model simulation and calibration.

ACKNOWLEDGEMENTS

The authors would like to thank the Advantest team for strong support during SEM image creation and extraction of SEM CD values, and for discussions regarding CD measurements from SEM images and extracted contours.

REFERENCES

[1] Vetter, A., Yan, C., Kirner, R., Scharf, T., Noell, W., Voelkel, R. and Rockstuhl, C., “Computational rule-based approach for corner correction of non-Manhattan geometries in mask aligner photolithography”, Optics Express Vol. 27, No. 22 (2019)

[2] Choi, Y., Fujimura, A. and Shendre, A., “Curvilinear masks: an overview”, Proc. SPIE 11855, Photomask Technology (2021)

[3] Van Look, L., Gillijns, W. and Gallagher, E., “Impact of mask corner rounding on pitch 40 nm contact hole variability”, SPIE Proceedings Volume 11854, International Conference on Extreme Ultraviolet Lithography (2021)

[4] Chou, S.-Y., Shin, J.-J., Shu, K.-C., You, J.-W., Shiu, L.-H., Chang, B.-C., Gau, T.-S. and Lin, B., “Study of mask corner rounding effects on lithographic patterning for 90-nm technology node and beyond”, SPIE Proceedings Volume 5446, Photomask and Next-Generation Lithography Mask Technology XI (2004)

[5] Synopsys S-Metro data sheet: https://www.synopsys.com/silicon/mask-synthesis/s-metro.html

[6] Chalony, M., Melvin, L., Zimmermann, R., Küchler, B., Viasnoff, E., Scarmozzino, R., Herrmann, D., Saad, Y., Stopford, P., Dam, T., Klostermann, U., Demmerle, W., Blais, A. and Stoffer, R., “Computational patterning and process variation impact on photonics devices”, SPIE Proceedings Volume 12499, Advanced Etch Technology and Process Integration for Nanopatterning XII (2023)

[7] Synopsys S-Litho data sheet: https://www.synopsys.com/silicon/mask-synthesis/sentaurus-lithography.html

[8] Synopsys Proteus data sheet: https://www.synopsys.com/silicon/mask-synthesis/proteus.html

Image sharpness score CD-SEM CDSEM monitoring

Nicolas Kubler, Bertrand Le Gratiet, Florent Dettoni
STMicroelectronics, 850 rue Jean Monnet, 38926 Crolles Cedex, France

ABSTRACT

Background: In the semiconductor industry, CD-SEM (Critical Dimension Scanning Electron Microscope) are key metrology tools used in the control of integrated circuit manufacturing processes [1]. In recent years, several approaches have been evaluated to better monitor these measurement tools from the images generated. One of the promising methodologies is the image sharpness score calculated from the grey levels present on an image.

The industrialization of this metric is a major and essential challenge in the CD-SEM image quality monitoring improvement, particularly in the detection of blurred images that can cause erroneous CD (Critical Dimension) measurements.

However, the implementation of such an indicator can be complex due to the number of different tool generations and the different beam conditions that can co-exist within the same factory. The diversity of technologies, patterns type, layers and materials add an additional constraint to its industrialization.

Aim: This study presents the work done from sharpness score extracted from CD-SEM images to industrialize this score as a metrology indicator.

It is based on analyses performed in a production environment comprising a diversity of 4 CD-SEM generations (18 tools) and various products. The objective is to define grouping rules allowing the industrialization of this metric with the aim of detecting equipment drift while offering a better piloting of the metrology equipment capabilities. The interest in this automation will also be to improve out-of-control events treatment efficiency, which are today dependent on human appreciation of image quality.

Approach: To systematically save the images and calculate the image sharpness score in real time, a database system [2,3] was used, coupled with a Virtual Machine (VM) on which software from Aselta Nanographics was installed. The IQ (Image Quality) module of this software was used in this study capable of providing a score on any CD-SEM image without the need to perform any edge detection or pre-calibration.

This Image Lab computing infrastructure has enabled the extraction of image sharpness score from over 35 million CD-SEM images collected in addition to the CD measurements already provided by the measurement equipment [3].

Results: The industrialization of the sharpness score required the evaluation and study of the different parameters that could influence this metric. An overview of these results is presented showing a dependence of the image sharpness score on the different pixel sizes that can be used in CDSEM recipes. The use of different materials can also lead to additional sensitivity and variability which may be different depending on the type of CD-SEM used.

Further results will be given to better understand the behavior of the sharpness score and to highlight situations that could not be detected in time currently.

Conclusion: The analysis of the image sharpness scores revealed the complexity of industrializing this metric and the need to create smart groupings by context such as pattern type, pattern density, material/process type,

38th European Mask and Lithography Conference (EMLC 2023), edited by
Uwe F.W. Behringer, Jo Finders, Proc. of SPIE Vol. 12802, 128020J
© 2023 SPIE · 0277-786X · doi: 10.1117/12.2675924

pixel size and measurement conditions. Adjusting the right sensitivity to the detection of the sharpness score also appeared to be one of the other key criteria in its industrialization. The evaluation of the contour-based sharpness score could provide an answer in the ease of grouping certain contexts.

Keywords: CD-SEM, industrialization, image blurr, sharpness score, measurement indicator, parameters, pixels size, materials, SEM type, sensitivity, variability, contour based

Contacts :

[a] e-mail: nicolas.kubler@st.com
[b] e-mail: bertrand.le-gratier@st.com
[c] e-mail: florent.dettoni@st.com

INTRODUCTION

With increasing demand for electronic components, many microelectronics companies have decided to invest and massively increase their production capacity to meet the growing demand for semiconductors. Nanotechnologies require rigorous control to meet the challenges facing the microelectronics industry in the race to mass industrialization and the miniaturization of electronic components.

In the semiconductor industry, the scanning electron microscope is one of the key metrology equipment used in the control of critical dimensions required for the proper functioning of integrated circuits.

To ensure the quality demanded by customer specifications, measurements need to be increasingly accurate, robust, and reliable. Increased production capacity requires effective monitoring of dimensional control, particularly regarding CD-SEM equipment reliability. Consequently, the equipment must return the same measurement information so that it can be used indifferently. Companies are faced with having to manage and coexist with different equipment generations that do not have the same technical characteristics or performance levels. Image quality drift of one of these equipment's can lead to erroneous measurements, which can have a significant impact on high production volumes.

The millions of measurement images generated each year by CD-SEM equipment are a mine of information which, when analyzed, can give various indications of their operating condition in addition to the measured CD value. For a few years, software has been available to quantify the blur level present in an image as a numerical value. The sandbox, presented at SPIE2022 [2], allows us to collect sharpness score data from all the images obtained from CD-SEM measurements.

Based on the data collected, the objective of this study was to establish a metrological indicator to ensure better detection of equipment drift and more out-of-control events treatment efficiency.

Industrializing sharpness score is a challenge to improve CD-SEM image quality monitoring. Setting control limits requires evaluation and study of the different parameters (pixel size, materials, patterns type, measurement conditions) that can have an influence on this metric. Three of the most influential parameters are presented in this study. Examples of CD-SEM excursion detection are shown to illustrate the benefits of such image quality monitoring.

PIXEL SIZE EFFECT

Each CD measurement recipe has its own settings and a measurement magnification that is adapted according to the pattern size to be measured and its design. So, there are many recipes for CD-SEM generating images with different pixel sizes.

The trend (**figure1**) illustrates the evolution of the sharpness score according to the different pixel sizes used in different recipes. The image pixel size has an influence on the image sharpness score:

(1) Image sharpness score increases with pixel size.
(2) If the pixel increases, a higher mismatching is observed between the different equipment generations.

Two distinct areas are clearly identified which are functions of the pixel size used in the CD-SEM recipes:

- **Area 1 [1nm – 4nm]**: the most advanced equipment generation (SEM_TYPE_3 and SEM_TYPE_4) have the highest values, which coincides with the fact that they have better resolution and therefore produce better image quality. The sharpness score is stable in this area.
- **Area 2 [4nm – 9nm]**: the oldest equipment generations (SEM_TYPE_1 and SEM_TYPE_2) are those which show the best scores. The sharpness score value is sensitive to the variation of the pixel size.

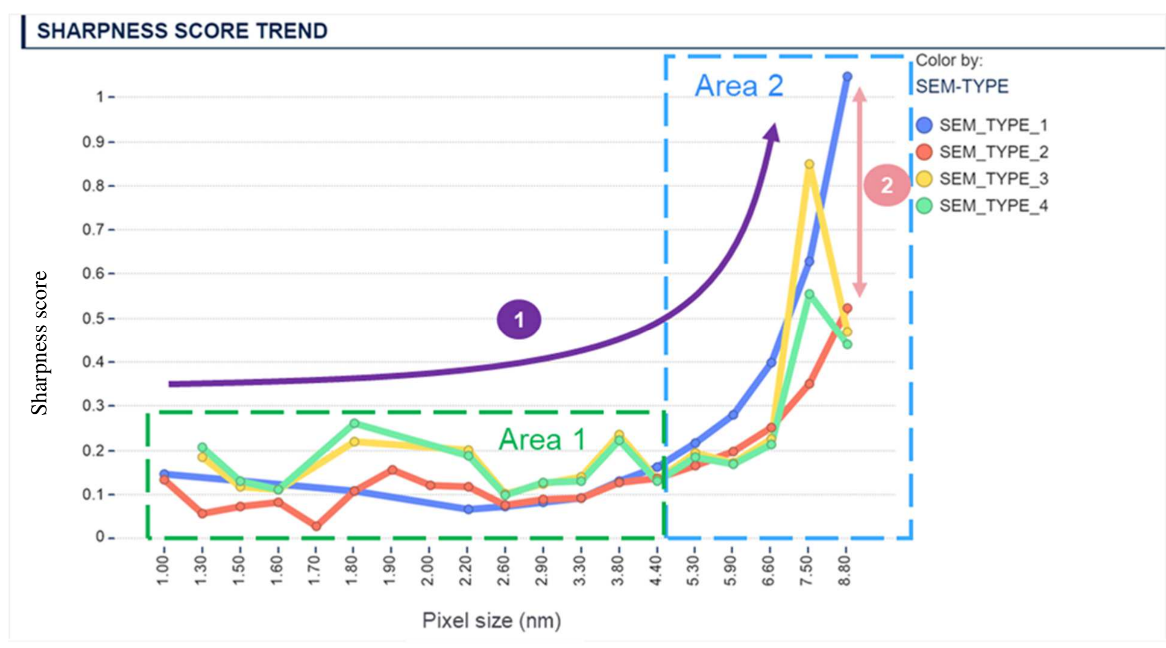


***Figure 1*:** Comparison of CD-SEM sharpness score depending on pixel size used in CD-SEM recipes. (Each color line represents a different type of CD-SEM)

(1) Images sharpness score increasing with pixel size (2) At high pixel size appears an increase between the different types of equipment (3) Two distinct areas are clearly identified which are functions of the pixel size used in the CD-SEM recipes. The most advanced generation equipment (SEM_TYPE_4) does not always appear with the highest metric.

These results show that the image sharpness score is dependent on the pixel size used in CD-SEM recipes. To industrialize this indicator, it is necessary to create groups by pixel size.

MATERIAL EFFECT

Many materials are used during the process steps allowing the manufacture of an integrated circuit, ranging from conductor to semiconductor to insulator.

A comparison of five processes was studied, each with different material properties (**figure 2**). These images, carried out after the photolithography and etch steps, contain resists, oxide (Ox), polysilicon (Poly-Si) and silicon (Si).

To avoid any effect related to pixel size, the study focused on recipes and images with the same pixel size (7.5nm). The patterns type measured (isolated trench-type patterns), the CD target (~1µm) and the measurement conditions (800eV) used were equivalent across the images whose scores were compared.

The image sharpness score (figure 2) shows a dependency on the material used and the type of CD-SEM. Indeed, the presence of certain materials such as polysilicon generates high sensitivity

and significant variability, particularly in the most advanced SEM_TYPE_3 and SEM_TYPE_4 equipment.

At the opposite for older generation equipment (SEM_TYPE_1 and SEM_TYPE_2), the sharpness score is not sensitive to changes in materials.

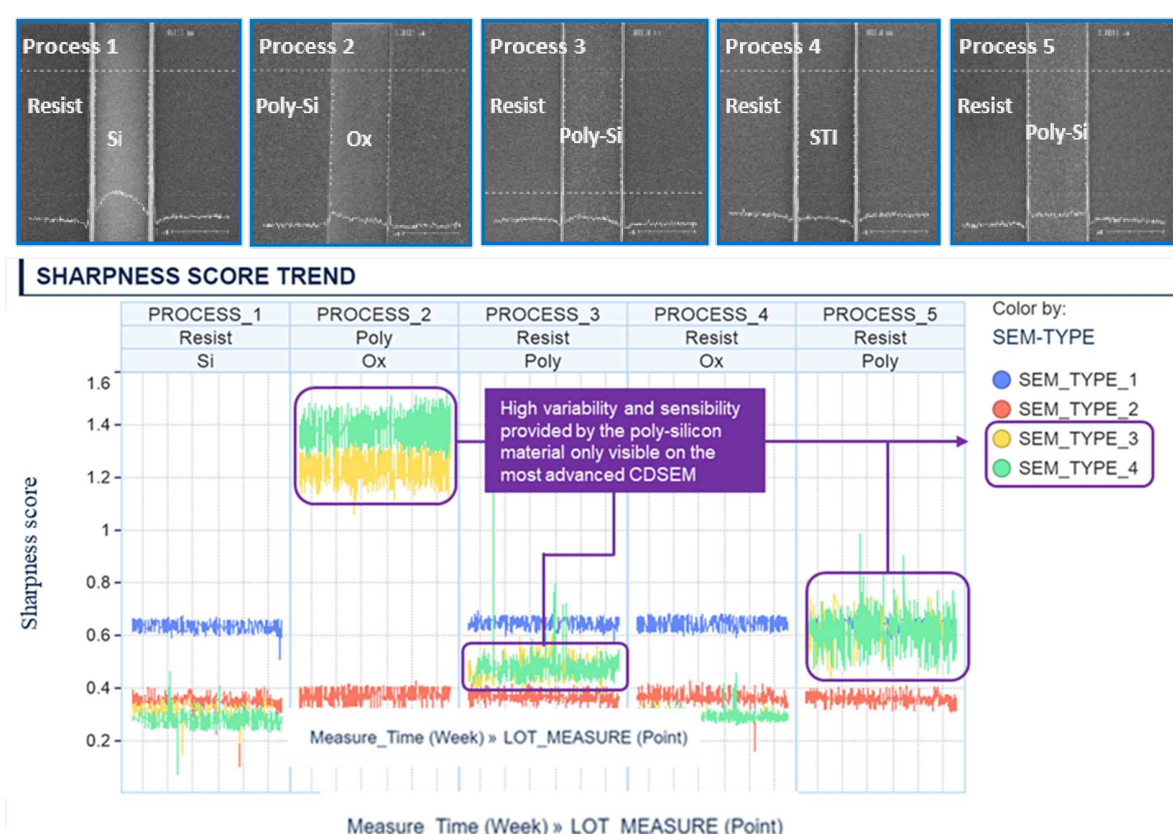


Figure 2: *Sharpness score extracted per process type with the different type of CD-SEM used*

Trend up is an overview of image with a different process used for sharpness score measurements. The trend down is the sharpness score extracted per process type with, in color, the different type of CD-SEM used.

The use of different materials can also lead to additional sensitivity and variability which may be different depending on the type of CD-SEM used. This reinforces the interest in creating relevant groups by material and adapting control limits according to the equipment generation used.

PATTERNS EFFECT

Several structure types are monitored during the manufacturing process of an integrated circuit, which may include different densities (isolated, dense) and geometric shapes. To assess their impact on image sharpness score, a comparison of structures such as trenches, holes, and product

(**figure 3**) was carried out using images collected after the photolithography and etch process steps.

To eliminate as much variability as possible, these images were compared using the same measurement conditions, the same pixel size (1.32nm), the same number of pixels per image (1024 pixels) and the same generation of equipment.

The results illustrated in **figure 3** show that the most isolated structures (product type) give the lowest image sharpness scores. On the other hand, trench-type patterns have the highest values, ahead of hole type patterns.

This first observation can be explained by the fact that the dense patterns (trench type) have a greater number and length of edges within the same image.

Moreover, the difference in score observed between the three trench type structures reveals that the sharpness score is dependent on the process step at which they were checked. The sample's nature and the structure's slope influence the yield of the secondary electrons collected on the detector. This will have a direct impact on the grey level present on the CD-SEM image and therefore on the sharpness score value. This second observation reinforces the previous remarks which mentioned that its sensitivity and its variability depended on the observed material.

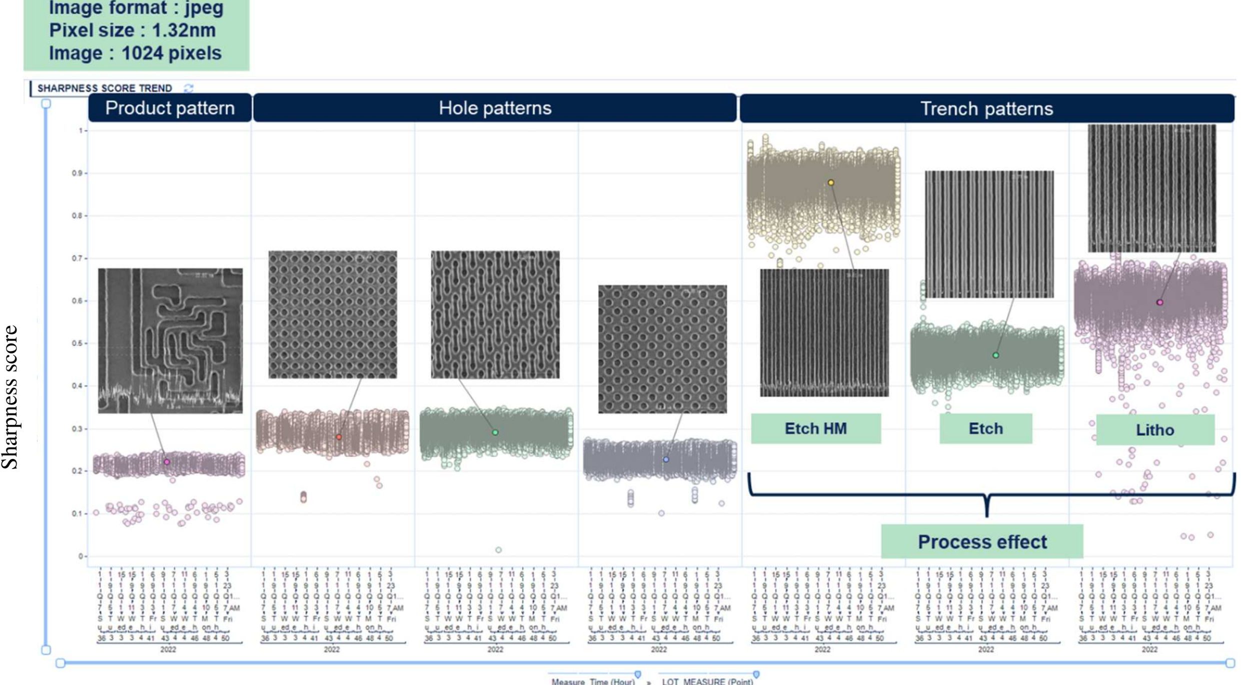


Figure 3: *Sharpness score values according to the patterns type measured*

Each point on the graph corresponds to a sharpness score value for an image acquired on a wafer at a specific time and date for an equipment.

These differences make it difficult to group the structures together and confirm the need to dissociate them if we wish to ensure optimal monitoring of the equipment.

USE CASE: EXCURSION CD-SEM

Previous results have shown that the sharpness score was sensitive to many parameters. They revealed the complexity of industrializing the sharpness score and the need to create smart groupings by context such as pattern type, pattern density, material/process type, pixel size and measurement conditions.

Contour based sharpness score can help to group certain parameters. This indicator has the advantage, compared to sharpness score, of providing a normalized score that can be used to group structures that do not have the same number of edges in the image. In addition, the effects associated with the random centering of the structure in the CD-SEM field of view are removed. By reducing this variability, CD-SEM equipment excursions can be better identified.

With a normalization by equipment type, contour-based sharpness score can identify equipment drifts. Examples of atypical behavior are illustrated in **figure 4** from the defined tolerance limits. A link with maintenance actions has been correlated to these weak signals:

1) Periodical repeated image quality excursions are observed on SEM_TYPE_1 generation equipment. A beam adjustment temporarily resolved the problem. The main cause has not yet been identified.

2) A deterioration in image quality and high variability was observed over a given period. The equipment then returned to normal following maintenance. A change of diaphragm position and column alignment, allowing the electron beam to be centered and uniform, restored optimum image quality to the equipment.

3) A drift in the image sharpness scores has been observed on the equipment. It is now disqualified for the most critical measurement recipes. A change to the electron gun tip is planned to improve image quality.

4) A point outside the high limits has been detected. This corresponds to a one-off drift in the peak current. The current value read during the measurement is higher than the expected value. This difference means that more electrons are sent to the sample, which explains the score observed and the difference in image quality. As a result, the image quality is better than normal.

5) One of the most advanced generations of equipment SEM_TYPE_4 randomly displays noisy images related to electromagnetic interference. The root cause has now been identified and is on the way to being fixed by installing a field compensator to eliminate external disturbances.

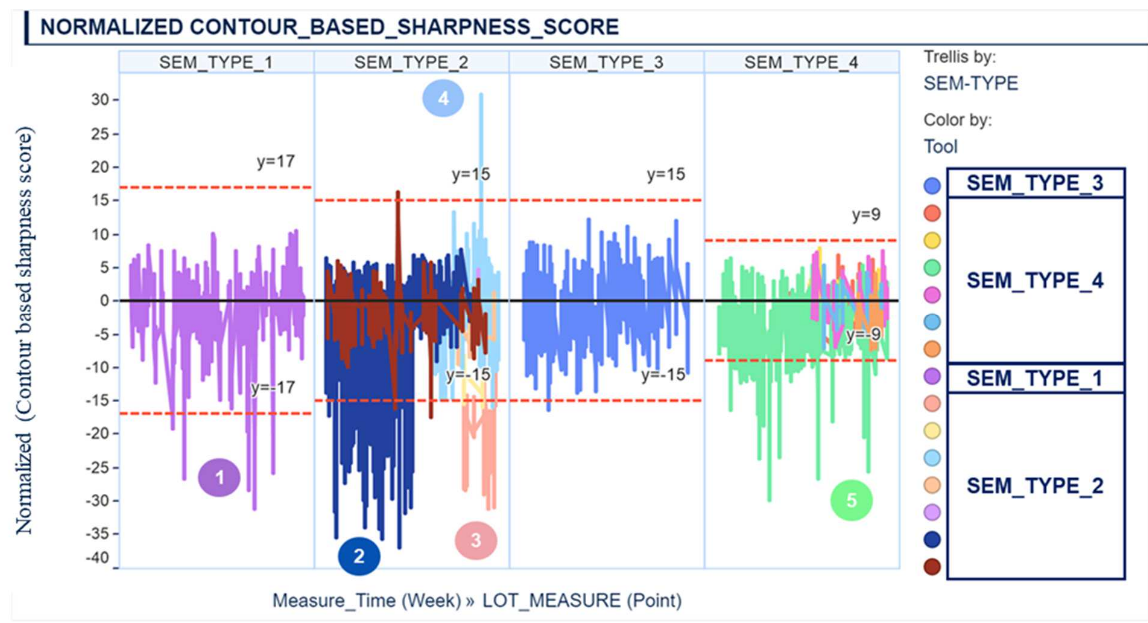


Figure 4: *Contour based sharpness score normalized highlighting atypical equipment behaviors*

CONCLUSION

Contour based sharpness score is an interesting metric to improve and uniformize the monitoring of a heterogeneous CD-SEM fleet (including recent machines and older ones).

This work has put in evidence the effectiveness of this metric to detect or understand CD-SEM machine image acquisition drifts so that appropriate actions can be taken to secure measurement quality. Contour based sharpness score is a robust solution, but it requires for its industrialization a normalization and grouping of measurement conditions (pattern type, material type, measurement condition and pixel size) so that appropriate control limits can be used. This indicator is now used to ensure greater reactivity in detecting excursions, help improve matching and will lead to greater efficiency and expertise in resolving maintenance problems. It is a valuable aid in diagnosing complex equipment problems and provides employees with guidelines for their work.

REFERENCES

[1] ***SEM image quality assessment for mask quality control***
Nivea G. Schuch, Alexandre Moly, Charles Valade, Nassim Halli, Mohamed Abaidi, Jordan Belissard, Frederic Robert, and Thiago Figueiro
Proc. SPIE 11855, Photomask Technology 2021, 1185505 (27 September 2021)
https://doi.org/10.1117/12.2601081

[2] ***A proof of concept of remote contour-based SEM metrology integration in HVM environment***
Bertrand Le Gratiet, Régis Bouyssou, Julien Ducoté, Florent Dettoni, Thibaut Bourguignon, Vincent Morin, Romain Bange, Nivea Schuch, Julien Nicoulaud, Guillaume Renault, Frederic Robert, Thiago Figueiro
Proceedings Volume PC12053, Metrology, Inspection, and Process Control XXXVI; PC120530 (2022) https://doi.org/10.1117/12.2615199

[3] The image lab sandbox, pulling image computing in wafer fab metrology environment
Florent Dettoni, Bertrand Le Gratiet, Vincent Morin, Nicolas Kubler, Aurélie Le Pennec, Romain Bange, Julien Ducoté, Elvire Soltani, Thibaut Bourguignon, Thomas Pelletier, Nour Osta, Mathilde Dmitrieff, Nicolas Bigarne. EMLC 2023

Actinic EUV mask qualification for next generation lithography

Matthias Roesch*, Renzo Capelli, Lukas Fischer, Klaus Gwosch, Grizelda Kersteen, Carolin Mueller, Robert Nicholls, Andreas Verch, and Alexander Winkler
Carl Zeiss SMT GmbH, Rudolf-Eber-Straße 2, D-73447 Oberkochen, Germany

ABSTRACT

During the last decade, the introduction of EUV lithography in high-volume chip manufacturing has been accompanied by the development of technological prerequisites for a future support of the node scaling roadmap. As core element, the next generation EUV scanner with an increased NA of 0.55 will be implemented into wafer fabs within the upcoming few years. In addition to its enhanced resolution, the High-NA exposure tool improves image contrast, and consequently reduces local CDU and defect printing on wafer. To take full advantage of this next leap in lithography, the whole infrastructure including EUV photomask technologies and metrology must keep pace with the scanner progress.

In this context, actinic EUV mask measurement represents a unique and variously usable way for the qualification of the mask printing performance under scanner-equivalent conditions. The aerial image metrology is targeted to match the corresponding scanner aerial image by means of the emulation of imaging-relevant scanner properties including wavelength, mask-side NA, through-slit chief ray angle, illumination schemes, and aberration level. To qualify High-NA masks of the anamorphic scanner, a methodology was developed that allows the simultaneous measurement of both NA=0.33 and NA=0.55 reticles based on one isomorphic optical projection optics design.

Here, we describe the challenges and corresponding solutions combined with two intrinsically diverse emulation types, NA=0.33 isomorphic and NA=0.55 anamorphic, in one single metrology. Special attention is paid to the emulation of the elliptical scanner NA at reticle, the contrast impact due to vector-effects in High-NA scanner imaging, wafer defocus of an anamorphic system for focus-dose process window determination, the pupil obscuration of the High-NA scanner projection optics, and the scanner facetted illumination by means of physical free-form blades, and by a completely digital solution.

Keywords: Mask metrology, aerial image, defect review, EUV optics, scanner emulation, High-NA

1. INTRODUCTION

The huge progress of the EUV scanner technology of the last few years boosted confidence into the new tools and led to a ramp-up in semiconductor chip industry. [1] In parallel to the NA=0.33 roadmap for scanner evolution enabling improved imaging, overlay, productivity, and defectivity, the mask infrastructure had to create a suitable ecosystem to support this technological leap. With the development and initial production of the next-generation anamorphic NA=0.55 EUV lithography systems, [2] new challenges appear for photomask technology and metrology.

*matthias.roesch@zeiss.com; phone +49 7364 20-2504; fax +49 7364 20-8000; zeiss.com

38th European Mask and Lithography Conference (EMLC 2023), edited by
Uwe F.W. Behringer, Jo Finders, Proc. of SPIE Vol. 12802, 128020K
© 2023 SPIE · 0277-786X · doi: 10.1117/12.2681647

One of the fundamental prerequisites for robust NA=0.55 high-volume manufacturing is the supply of a defect-free mask for EUV exposure. In this context, aerial image metrology of photomasks is a powerful and credible way to qualify the reticle printing performance under scanner-matching conditions. In a closed loop together with a corresponding mask-repair process, by means of aerial image mask review different types of printing defects can be identified and finally get fixed. [3] Since mask-shops will have to provide defect-free reticles for both isomorphic NA=0.33 and anamorphic NA=0.55 EUV lithography in the future, it will be a cost-efficient solution to qualify all mask types based on one single metrology. Consequently, the question arises how the aerial image measurement technique using an isomorphic optical column can be refined and extended to emulate anamorphic High-NA imaging as well.

In this paper, we will first show the metrology challenges combined with the anamorphic lithography tool of higher NA. Subsequently, a detailed overview of hardware- and software-based solutions is presented that allow a precise mask review of High-NA photomasks using aerial image measurements in combination with an isomorphic projection optics.

2. HIGH-NA CHALLENGES FOR PHOTOMASK METROLOGY

The basic principle of an aerial image measurement technique that delivers scanner-matching results, i.e., that sees the mask in the same way as the wafer does in the scanner, is the usage of an optical column being as close as possible to the one in a corresponding lithography system (Figure 1). Within an EUV scanner, the reticle structure is illuminated by a dedicated illumination system under a chief-ray angle (CRA), and the reflected and diffracted light is subsequently imaged into the wafer with a certain size reduction of < 1 x. The optical resolution of the imaging process is mainly determined by the radiation wavelength, the size of the mask side NA, and the illumination profile used. Aerial image measurements of photomasks are combined with a similar optical architecture that consists of an illumination system, the reticle, and a magnifying projection optics (> 500 x). Provided that scanner matching is ensured, the illumination system can be a top hat design combined with corresponding setting blades. A similar blade can emulate the pupil shape of the scanner POB. The mask image is finally detected by an EUV-sensitive camera system.

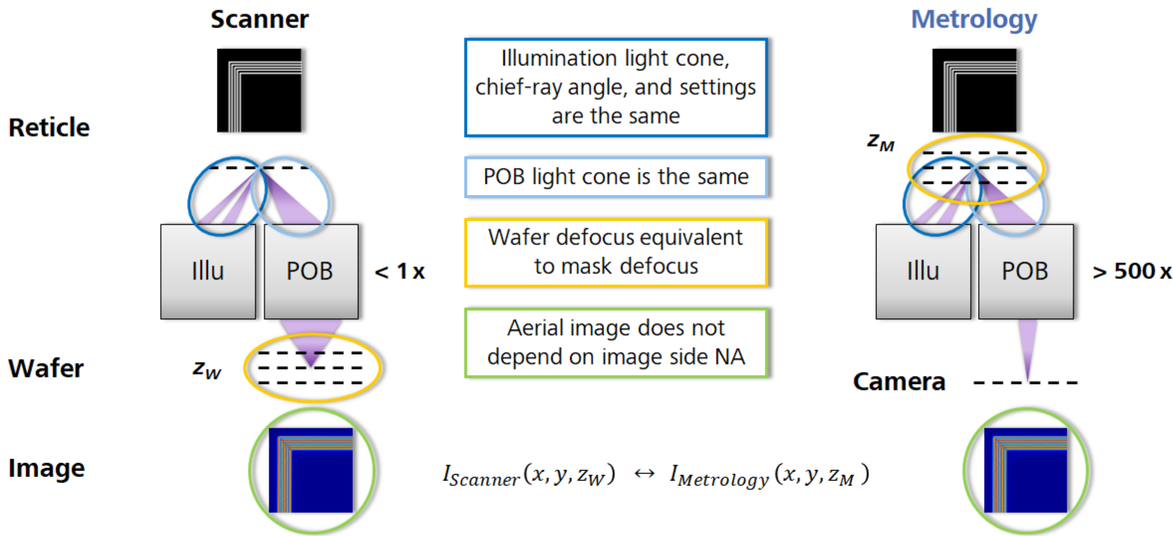


Figure 1. Optical architecture of an isomorphic EUV scanner (left) and of a corresponding photomask metrology (right). The main difference of both compositions is the large magnification of the metrology optics as compared to the typical reduction of the scanner optical column, and the placement of an EUV metrology camera at the position of the wafer in the scanner.

Likeliness of the aerial images of a scanner and the corresponding metrology is obtained by several properties of the metrology optical architecture. First, the illumination wavelength must be identical (actinic metrology), and its light cone, the CRA, and the illumination profile must meet the scanner architecture. Furthermore, the reticle-side light cone of the metrology POB must fit the NA and shape of its scanner counterpart, too. One essential functionality of photomask metrology is the emulation of scanner wafer defocus for process window characterization. For an isomorphic review projection optics, one can fully mimic wafer defocus of the scanner by mask stage defocus movement, i.e., mask defocus is equivalent to wafer defocus. Finally, for scanner matching the aerial image profile must not depend on the image side NA, which is clearly different for scanner reduction imaging and metrology microscopic magnification.
In contrast to the isomorphic NA=0.33 EUV scanner, the High-NA EUV scanner with its anamorphic optical architecture leads to additional challenges for the design of a scanner-matching metrology. The increased scanner NA is not only combined with increased light cones at reticle but will also clearly enhance the impact of vector effects on imaging. Due to the anamorphic magnification of the scanner, which also leads to a different scaling with respect to isomorphic metrology aerial images of the same mask, wafer defocus cannot be emulated by reticle defocus alone. In contrast to the NA=0.33 scanner, the POB exit pupil of the High-NA exposure tool shows a central obscuration that impacts imaging. Finally, source-mask optimized (SMO) illumination schemes, that cannot be simply emulated by the previously used standard exposure settings, become more and more important for High-NA EUV lithography. [4]

3. METROLOGY ARCHITECTURE FOR HIGH-NA PHOTOMASKS

3.1 Increased Numerical Aperture

The increase of the Numerical Aperture of the EUV lithography optics from NA=0.33 to NA=0.55 for the next generation scanner has consequences for the metrology architecture both on reticle and on image side. Provided the optical column will support the increased light cones, illumination- and POB-blades that exhibit the reticle-side NA values of the anamorphic scanner must replace the circular blades combined with isomorphic NA=0.33 emulation (Figure 2).

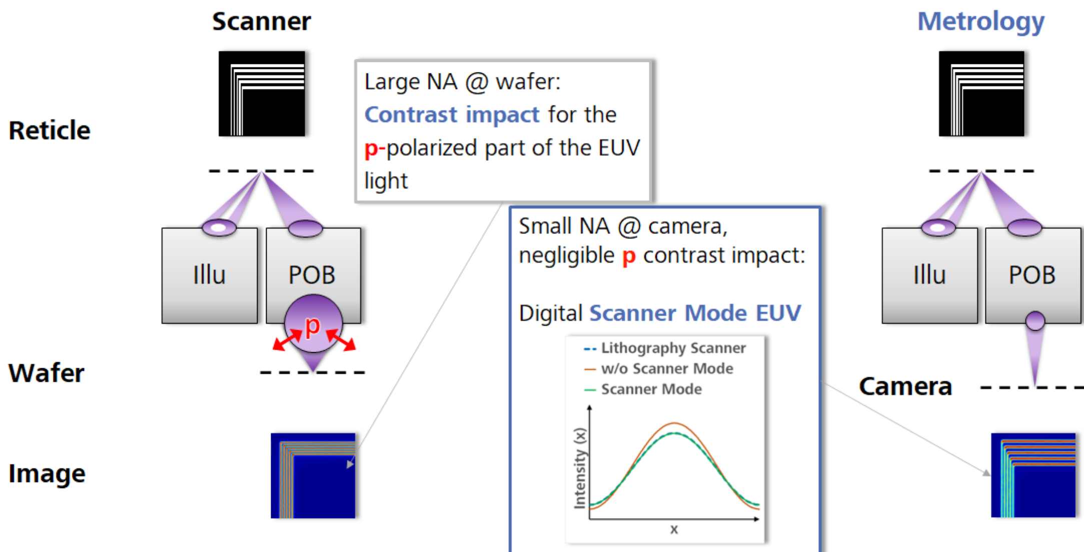


Figure 2. Optical architecture of an anamorphic High-NA EUV scanner (left) and of a corresponding isomorphic photomask metrology (right). Imaging combined with the NA of the anamorphic scanner requires suitable illumination- and NA-blades of corresponding size, and a digital correction of the metrology results to correct for the contrast impact for the p-polarized part of the EUV light in a NA=0.55 lithography scanner.

On wafer side, the expansion of the High-NA scanner light cone is combined with a polarization driven contrast impact on aerial images [5]. In detail, the p-polarized part of the EUV light shows decreasing interference with increasing opening angle of the illumination profile since the polarization vectors of both interfering rays then become less and less parallel. Due to its large magnification and consequently small camera-side NA, the metrology results are essentially not affected by polarization. To account for this difference in polarization impact between scanner and metrology, a suitable digital image correction (Scanner Mode EUV) must be applied for the metrology data.

3.2 Anamorphic Magnification

The most prominent property of the High-NA EUV scanner optical column is its anamorphic image reduction. Consequently, compared to photomasks for the isomorphic NA=0.33 scanner, the reticle structures for NA=0.55 are stretched in one direction with respect to the wafer target pattern. If masks for anamorphic scanner exposure are qualified using an isomorphic measurement technique, the arial images thus must be suitably processed to show the same scaling as the scanner aerial images (Figure 3).

The anamorphic imaging principle of the High-NA lithography system has extensive consequences for the emulation of scanner wafer defocus, e.g., for process window characterization with respect to dose and focus, if the isomorphic measurement technique is applied (Figure 3). The challenge can be understood in terms of isomorphic depth-of-focus that is proportional to λ/NA^2, where λ is the wavelength of the EUV light. Given the elliptical shape of the NA at mask level, the depth-of-focus of features extending along the direction of smaller reduction will be in defocus earlier than features extending along the opposite direction.

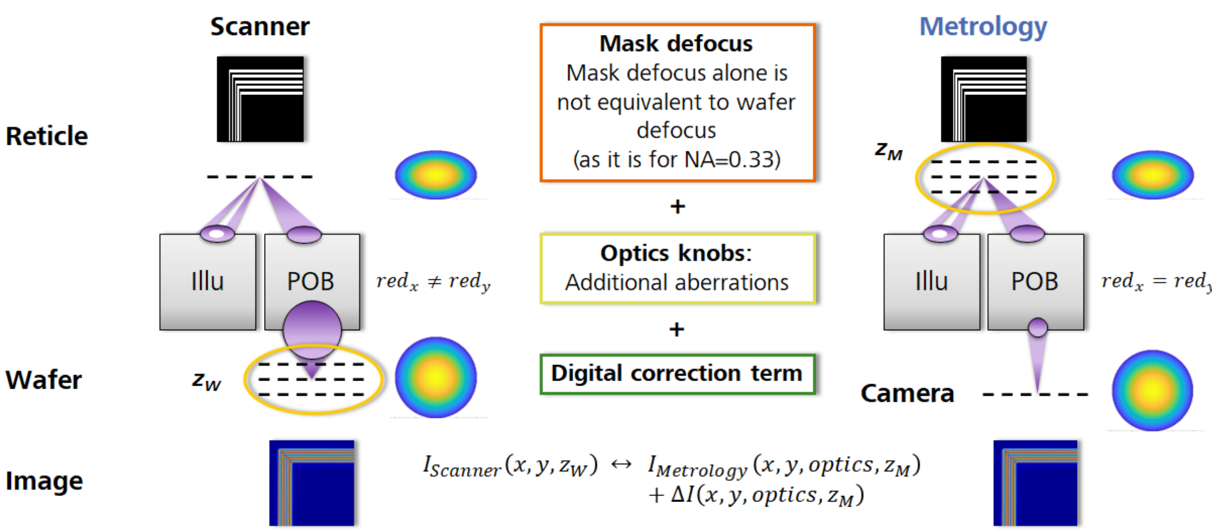


Figure 3. Wafer defocus z_W of an anamorphic scanner can be emulated within an isomorphic metrology by means of a combination of mask defocus z_M, turning optics knobs, and a proper digital correction term.

If the wafer in an anamorphic scanner is moved along the focus direction (defocus z_W), the corresponding wafer-side wavefront change is of Zernike Z4 type. On reticle side, due to the anamorphic optics the wavefront shape is compressed for the direction of larger reduction, and thus of elliptical Z4-like type. If one now tries to emulate wafer defocus by mask defocus alone (z_M), as it is way-of-working for emulation of isomorphic scanner wafer defocus, the resulting wavefront change at reticle side first is of circular Z4 shape. Since an elliptical POB blade is inserted at this position in the metrology, however, the circular Z4 wavefront is cropped correspondingly, resulting in a wavefront pattern that is clearly different from the target given by the scanner. An extensive approximation to the reticle-side wavefront change in

the scanner is obtained by means of turning dedicated optics knobs of the metrology projection optics. In this way, additional aberrations are introduced that nearly align the mask defocus wavefront change to the scanner target. The remaining gap can be covered by an additional so-called digital correction based on a sophisticated object reconstruction algorithm. [6]

3.3 Central Obscuration

The exit pupil of the High-NA scanner POB shows a central obscuration that, dependent on illumination setting and lithographic pattern, can block part of the transmitted and diffracted EUV light originating from the photomask. Therefore, for certain features and pitches within a mask structure a contrast loss with respect to an optics without central obscuration might occur, despite potential imaging enhancement techniques that try to minimize the effect. [7]
Image contrast in turn can have significant impact on defectivity printing and detection, therefore matching of the mask review metrology to the High-NA scanner requires the emulation of the scanner POB central obscuration. The scanner-POB pupil shape can be mimicked by means of a suitable mechanical NA-blade within the metrology optical path. Consequently, one can expand the POB NA-blade with an additional central obscuration that in shape is as close as possible to the central obscuration of the scanner (Figure 4).

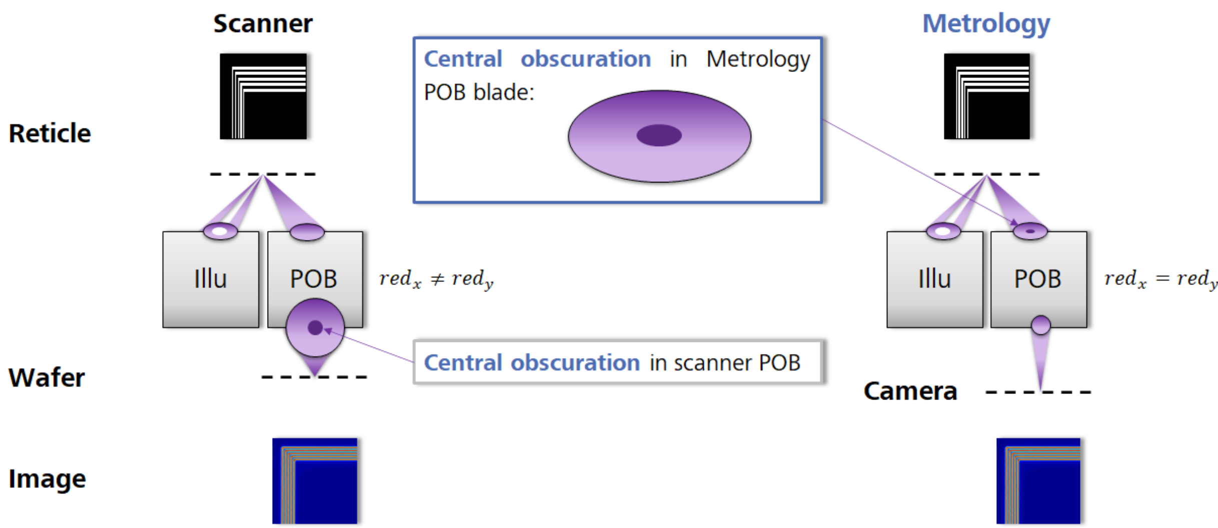


Figure 4. To match the aerial images of the photomask metrology to the aerial images of the High-NA scanner with its central obscuration in the POB exit pupil, photomask metrology is equipped with a POB NA-blade that exhibits a central obscuration that is as close as possible in shape to the one in the scanner POB.

3.4 Advanced Illumination

Modern EUV scanner systems are equipped with a free-form illumination functionality via the degrees of freedom in a facetted mirror system. [4] These modules form the pupil of the scanner illuminator and allow to simultaneously optimize the source shape (illumination setting) and the mask layout (SMO) to increase image contrast and therefore wafer throughput. Here we present essentially two different ways for photomask metrology to emulate a complex flexible spot illumination based on mechanical illumination blades (Figure 5).
The most obvious way to emulate scanner spot settings is to code a top-hat translation of the facetted scanner illumination into the mechanical metrology sigma aperture. However, although quite similar in shape, a top-hat translation might often not be the best illumination profile with respect to scanner matching. A more sophisticated solution is based on the optimization of imaging properties of the sigma aperture. Here, the illumination design contains

information over the whole optical path to compensate for any difference within the light cones between metrology and the corresponding scanner system. [8] The technique has, however, the drawback of a missing flexibility, since for every SMO setting in use within a production environment a dedicated hardware blade is required.

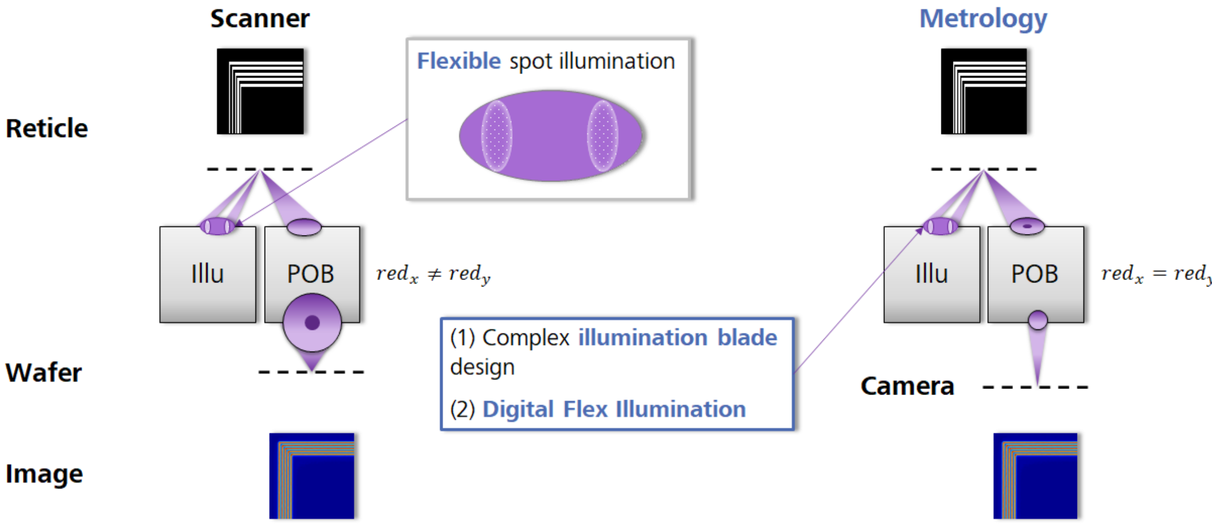


Figure 5. EUV lithography scanners are equipped with a flexible facetted spot illumination system that allow to provide nearly every required radiation profile. For optimized scanner matching, photomask metrology might either make use of (1) imaging-optimized mechanical sigma blades or (2) apply a fully digital solution that consists of a combination of metrology steps and sophisticated algorithms.

To meet the requirements of both scanner matching and a full and fast flexibility to scanner SMO illumination profiles, the photomask metrology technique can be complemented by a suitable process that combines aerial image measurement with powerful imaging algorithms (Digital Flex Illumination). [8] The basic idea of the procedure is to first capture aerial image information under various exposure conditions like defocus and illumination shift, and then to derive a very essential and useful property of the mask structure, the so-called object spectrum, from an optimization loop. The object spectrum of the mask in general is dependent on the angle of incidence, so a proper mask modelling is required to obtain a preferably compact form of the diffraction pattern information. Once the sigma-dependent object spectrum of the mask is known, synthetical aerial images with profound scanner matching quality can be obtained via a forward imaging calculation using the scanner spot illumination profile. [8]

4. CONCLUSIONS

With the upcoming introduction of the next-generation anamorphic High-NA lithography scanner, several challenges arise for actinic aerial image photomask review. To obtain optimum scanner matching, the vector-effect driven contrast impact on NA=0.55 imaging must be emulated within metrology via digital correction. Since mask defocus of the metrology is not equivalent to anamorphic scanner wafer defocus, for focus-dose process window characterization the metrology wavefront has sophistically to be controlled via additional scanner knob operation and digital fine tuning. Finally, there are two ways to emulate the scanner SMO free-form illumination. One solution is based on elaborately imaging-optimized physical free-form blades. More flexibility is combined with a completely digital solution that allows the generation of synthetic aerial images by a combination of standard image acquisition and powerful algorithms.

The presented metrology solution for High-NA reticle qualification offers a photomask review flow with supreme scanner matching for both NA=0.33 and NA=0.55. It represents an essential technology for the whole infrastructure of current and next generation EUV masks.

REFERENCES

[1] Eric Verhoeven, Ron Schuurhuis, Marcel Mastenbroek, Peter Jonkers, Frank Bornebroek, Arthur Minnaert, Harrie van Dijck, Parham Yaghoobi, Geert Fisser, Payam Tayebati, Klaus Hummler, and Roderik van Es, "0.33 NA systems for high-volume manufacturing", Proc. SPIE 11609, 1160908 (2021).

[2] Jan van Schoot, "High-NA EUVL exposure tool: program progress and mask interaction", Proc. SPIE PC12325, 123250A (2022).

[3] Renzo Capelli, Nathan Wilcox, Martin Dietzel, Dirk Hellweg, Scott Chegwidden, and Joseph Rodriguez, "AIMS™ EUV first insertion into the back end of the line of a mask shop: a crucial step enabling EUV production", Proc. SPIE 10810, 108100S (2019).

[4] Rongkuo Zhao, Fan Zhou, Jialei Tang, Jeff Lu, Yunbo Liu, Dezheng Sun, Ming-Chun Tien, Stephen Hsu, Rachit Gupta, Youping Zhang, and Joerg Zimmermann, "Computational lithography solutions to support EUV high-NA patterning", Proc. SPIE 12495, 124950R (2023).

[5] Donis G. Flagello, Bernd Geh, Steven G. Hansen, and Michael Totzeck, "Polarization effects associated with hyper-numerical-aperture (>1) lithography", J. Micro/Nanolith. MEMS MOEMS 4 (3), 031104 (2005).

[6] Renzo Capelli, Grizelda Kersteen, Sven Krannich, Markus Koch, Lukas Fischer, Matthias Roesch, and Klaus Gwosch, "AIMS EUV evolution towards high NA: challenge definition and solutions implementation", Proc. SPIE 12051, 120510A (2022).

[7] Eelco van Setten, Gerardo Bottiglieri, Laurens de Winter, John McNamara, Paul Rusu, Jan Lubkoll, Gijsbert Rispens, Jan van Schoot, Jens Timo Neumann, Matthias Roesch, and Bernhard Kneer, "Edge placement error control and Mask3D effects in High-NA anamorphic EUV lithography", Proc. SPIE 10450, 104500W (2017).

[8] Renzo Capelli, Klaus Gwosch, Grizelda Kersteen, Matthias Roesch, Carolin Mueller, Alexander Winkler, and Andreas Verch, "The power of algorithmic employed in a metrology system: AIMS EUV Digital Flex Illu", Proc. SPIE 12292, 122920N (2022).

SEMI-CenterNet: A Machine Learning Facilitated Approach for Semiconductor Defect Inspection

Vic De Ridder[a, b, *], Bappaditya Dey[b, *], Enrique Dehaerne[b,c, *], Sandip Halder[b], Stefan De Gendt[b, d], and Bartel Van Waeyenberge[e]

[a]Faculty of Engineering, Ghent University, Belgium
[b]imec, Kapeldreef 75, 3001 Leuven, Belgium
[c]Faculty of Science, KU Leuven, 3001 Leuven, Belgium
[d]Dept. of Chemistry, KU Leuven, 3001 Leuven, Belgium
[e]Dept. of Solid State Sciences, Ghent University, 9000 Ghent, Belgium
[*]These authors contributed equally to the work.

ABSTRACT

Continual shrinking of pattern dimensions in the semiconductor domain is making it increasingly difficult to inspect defects due to factors such as the presence of stochastic noise and the dynamic behavior of defect patterns and types. Conventional rule-based methods and non-parametric supervised machine learning algorithms like k-nearest neighbors (kNN) mostly fail at the requirements of semiconductor defect inspection at these advanced nodes. Deep Learning (DL)-based methods have gained popularity in the semiconductor defect inspection domain because they have been proven robust towards these challenging scenarios. In this research work, we have presented an automated DL-based approach for efficient localization and classification of defects in SEM images. We have proposed SEMI-CenterNet (SEMI-CN), a customized CN architecture trained on Scanning Electron Microscope (SEM) images of semiconductor wafer defects. The use of the proposed CN approach allows improved computational efficiency compared to previously studied DL models. SEMI-CN gets trained to output the center, class, size, and offset of a defect instance. This is different from the approach of most object detection models that use anchors for bounding box prediction. Previous methods predict redundant bounding boxes, most of which are discarded in postprocessing. CN mitigates this by only predicting boxes for likely defect center points. We train SEMI-CN on two datasets and benchmark two ResNet backbones for the framework. Initially, ResNet models pretrained on the COCO dataset undergo training using two datasets separately. Primarily, SEMI-CN shows significant improvement in inference time against previous research works. Finally, transfer learning (using weights of custom SEM dataset) is applied from ADI dataset to AEI dataset and vice-versa, which reduces the required training time for both backbones to reach the best mAP against conventional training method (using COCO dataset pretrained weights).

Keywords: semiconductor defect inspection, metrology, lithography, stochastic defects, supervised learning, deep learning, defect classification, defect localization, CenterNet

1. INTRODUCTION

The continuous shrinking of wafer pattern dimensions is causing increasing difficulties in wafer inspection due to factors such as noise, contrast changes, and lower resolution. SEM has been shown as a useful tool for semiconductor defect inspection due to its high spatial resolution and relatively fast throughput making it suitable for in-line inspection of small defects. Hence SEM imaging is used extensively between different (Litho/Etch) process steps to inspect the patterned wafer. For SEM-based wafer defect inspection, DL-based techniques have been demonstrated as an advantageous technique to deal with the challenging conditions for defect detection caused by shrinking pattern dimensions.[1] With continuous developments in Industry 4.0, AI and the Internet

Further author information: (Send correspondence to both authors)
Vic De Ridder: E-mail: vic.deridder.ext@imec.be
Bappaditya Dey: E-mail: bappaditya.dey@imec.be

38th European Mask and Lithography Conference (EMLC 2023), edited by
Uwe F.W. Behringer, Jo Finders, Proc. of SPIE Vol. 12802, 128020L
© 2023 SPIE · 0277-786X · doi: 10.1117/12.2675570

of Things, a significant increase in chip demand is inevitable. This may lead to even larger-scale fabs with an exponential increase in the amount of wafer data to be inspected per unit time. To deal with these challenges associated with advanced node defect inspection, fast and lightweight DL models can be beneficial.

While several works have investigated the use of DL methods for semiconductor defect detection,[2–4] little work has pursued using DL techniques with lower inference time and smaller models. Ref. 5 proposed a lightweight model for automatic classification of semiconductor defects, based on multiple neural networks in a decision tree. While this model achieved good results, it could not localize defects or find multiple defect instances inside one image. This work was an initial attempt at developing lightweight DL-based models with subsequently faster inference for semiconductor defect classification and localization. The CenterNet object detection framework[6] is investigated and adapted to the semiconductor defect detection task. CenterNet has been demonstrated as a lightweight and fast model family on the object detection task[6,7] because it predicts bounding boxes only for probable object centers, which is computationally more efficient compared to anchor-based approaches which predict numerous and redundant bounding boxes in background regions. Additionally, in the current literature, little investigation has been carried out in applying transfer learning between different semiconductor datasets (ADI-to-AEI and vice-versa) for defect detection application. Since semiconductor wafer SEM images from different datasets share certain features, a model trained from scratch or fine-tuned on one semiconductor SEM dataset can be proven advantageous by sharing weight parameters for extracting and learning subtle local and global features of other SEM semiconductor datasets with numerous defect patterns, compared to models with randomly initialized weights or models trained on other object detection tasks like COCO[8] or PASCAL-VOC.[9] Hence it can be expected that using transfer learning from one semiconductor SEM dataset could lead to a lower number of training epochs on other SEM semiconductor dataset(s) required for satisfactory performance to be achieved because appropriate initial weight parameters help backbones/models to avoid (early) convergence to a suboptimal local minimum of the loss function. In industry, numerous models have already been trained on custom datasets and the transfer learning strategy can emerge as a possible solution for improving Automatic Defect Classification and Detection frameworks.

The main contributions of this research work are i) SEMI-CenterNet, a model designed for fast inference time is proposed for the ADCD task, ii) two ResNet backbone variants integrated into SEMI-CenterNet framework are trained and evaluated for precision and inference time, and compared against previous DL-ADCD models, iii) it is shown that transfer learning between two ADCD datasets can cause shorter training time and comparable performance.

2. CENTERNET OBJECT DETECTION FRAMEWORK

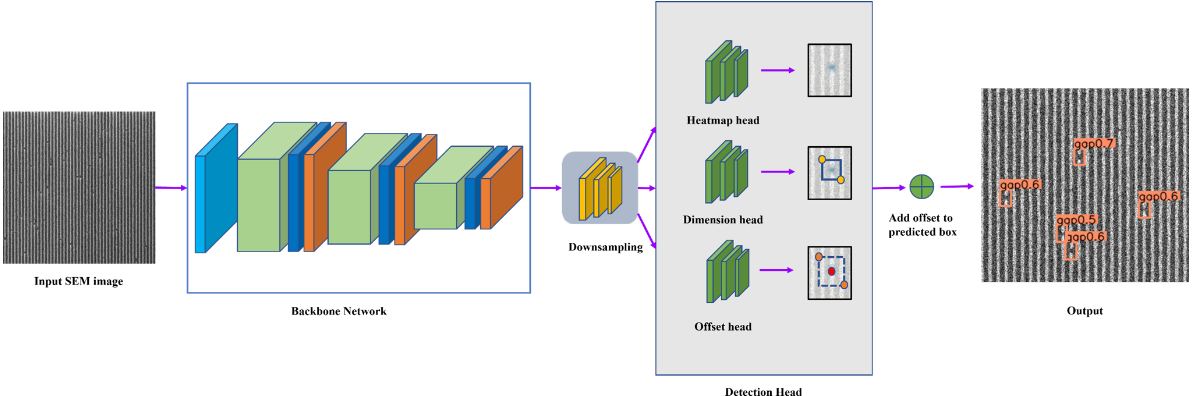


Figure 1. High-level overview of the SEMI-CenterNet based defect detection method adapted from Albahil S. et al.[10]

In this section, the CenterNet model is introduced and compared to other object detection architectures. First, DL-based object detection frameworks which were previously studied in the paradigm of semiconductor ADCD are discussed. Afterwards, the conceptual differences between CenterNet and other one-stage/two-stage object

detection models are explained and how these differences can lead to inference speed advantages is discussed. Finally, an in-depth look at the CenterNet object detection framework and its underlying mechanisms is provided.

Previous works first investigated simple CNN-based defect detection frameworks. In Ref. 2, a simple CNN framework was proposed which generates bounding boxes by classifying sliding window patches as either background or a specific defect type. This work demonstrated better accuracy of CNN-based models compared to other ML-based approaches in semiconductor defect detection. However, the sliding window approach caused a limitation since it creates predetermined bounding box dimensions. This is critical in use cases where large varieties of defects with different sizes exist.

To allow for flexible bounding box sizes and achieve higher accuracy on complex datasets, more complex CNN-based object detection frameworks have been investigated in recent works.[11,12] These complex models all use a feature extractor backbone.[13] This is a pretrained CNN that extracts relevant high-, mid-, and low-level features from the image. Afterwards, another sequence of CNN-related operations is applied to the extracted feature map to produce the networks' final outputs, often bounding boxes and corresponding class predictions. Due to backbone pretraining, the extracted feature map is more useful for making predictions compared to the original image, resulting in fewer object detection training iterations and often a better final performance compared to models that do not contain a pretrained backbone and are trained for detection from scratch.

The most common, well-studied DL-based object detection frameworks are the Region-based Convolutional Neural Networks (R-CNN)[14] and You Only Look Once (YOLO)[15] model families. Both paradigms function through bounding box proposals. A model outputs a large number of bounding boxes and corresponding class confidence scores which are then post-processed using an algorithm such as Non-maximum Suppression (NMS) to discard low confidence predictions and fuse predictions pertaining to the same object. These post-processed predictions are then shown to the user.

Models in the R-CNN family first predict Regions Of Interest (ROIs) through either a CNN or classical approach such as selective search. A bounding box with corresponding confidence scores will then be predicted for each ROI. The YOLO family bypasses the time-intensive ROI generation stage. An image is divided into n same-sized grids (often called anchors) and the model is tasked with predicting k object proposals for each anchor. By using constant ROIs (the grids), YOLO achieves faster inference compared to R-CNN.

The aforementioned object detection model families have been successfully applied to the semiconductor defect detection task as demonstrated in Ref. 3 and Ref. 16. Moreover, architectures are being developed specifically for the task of semiconductor defect detection. In Ref. 17, a model which makes use of a number of regular convolutions in an encoding stage to produce a feature map of the image from which defect segmentations and classifications are predicted using transposed and regular convolutions respectively. However, as of now, no research work has been done to optimize the inference speed of the models.

A downside of the approaches taken by YOLO and R-CNN is, in cases where objects constitute small regions of the input image, a lot of computations will be involved in predicting bounding boxes for background regions, which is not useful for object detection or localization and thus leads to inefficient use of computational resources. Several works attempt to avoid this issue, among which CenterNet, whose architecture (adapted in the form of SEMI-CN) is shown in figure 1. CenterNet first predicts probable object centers, and only afterwards constructs bounding boxes for the most likely centers. In this way, computation is first put into detection and center localization and afterward into prediction of object bounds.

To save resources, CenterNet predicts a heatmap of lower resolution than the original image. A stride R is chosen such that for an $n \times n$ pixels image, the model predicts an $\frac{n}{R} \times \frac{n}{R}$ heatmap. To produce the target $\frac{n}{R} \times \frac{n}{R}$ ground truth heatmap, the ground truth center x,y from the $n \times n$ map is scattered on the n/R heatmap around $\frac{x}{R}, \frac{y}{R}$ with heat intensity decaying with increasing distance from $\frac{x}{R}, \frac{y}{R}$ in a gaussian manner with object size dependent standard deviation.

In model use, bounding boxes eventually need to be displayed for the original image dimensions. The network's heatmap output only consists of discrete points so a discretization error is introduced: the actual ground truth center coordinates in the original image is c, the corresponding scaled-down version will be $\frac{c}{R}$ which are not necessarily discrete points. The network makes an offset prediction for each predicted center to resolve this. If a model has predicted center x coordinate $x1$ on the heatmap with corresponding offset $\hat{o}$, the upscaled center x

coordinate in the original image becomes $(x1 + \hat{o}) * R$. Then the bounding boxes are predicted for each object center using a head separate from the heatmap or offset prediction head, as shown in figure 1.

3. DATASETS

The proposed framework is evaluated on two datasets, each containing line-space pattern SEM images taken either After Develop Inspection (ADI dataset) or After Etch Inspection (AEI dataset). Both datasets were already introduced in previous works (Ref. 18 and Ref. 16, respectively). Each dataset was divided into training, validation and test splits as shown in tables 1 and 2. Each image is stored in the TIFF greyscale format and each defect has a corresponding bounding box and class annotation generated by a human expert. No synthetic defects or images are used to ensure proper approximation of real fab data and conditions. All defects encountered are stochastic in nature, no intentionally placed defects are present in the datasets. The AEI dataset consists of 480×480 pixel images that contain the following defect types: line collapse, single bridge, thin bridge, multi bridge non-horizontal and multi bridge horizontal. Examples of these defects are shown in figure 2. The ADI dataset consists of 1024×1024 pixel images that contain the following defect types: gap, probable gap (p-gap), microbridge, bridge, line-collapse. Examples of these defects are shown in figure 3.

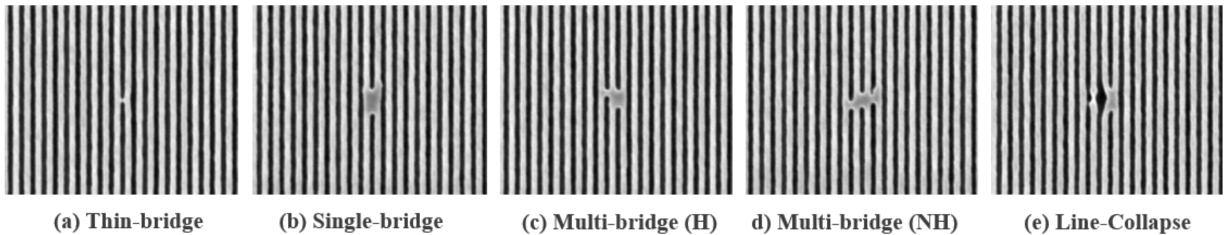


Figure 2. Example defect types in the AEI dataset

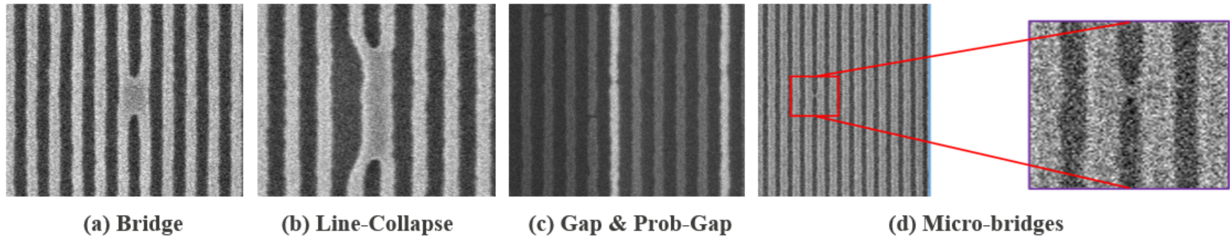


Figure 3. Example defect types in the ADI dataset

Table 1. Splits and defect distribution of AEI dataset.

Class Name	Train (920 images)	Val (120 images)	Test (120 images)
Linecollapse	202	34	40
Single bridge	240	31	29
Thin bridge	241	29	29
Multi bridge (non horizontal)	160	19	21
Multi bridge (horizontal)	80	10	10
Total instances	923	123	129

Table 2. Splits and defect distribution of ADI dataset.

Class Name	Train (1053 images)	Val (117 images)	Test (154 images)
gap	1046	156	174
p_gap	315	49	54
microbridge	380	47	78
bridge	238	19	17
line_collapse	550	66	76
Total instances	2529	337	399

4. METHODOLOGY

4.1 Benchmarking

This work will investigate the precision and inference time of the proposed SEMI-CenterNet framework for defect detection on semiconductor wafer images (SEM-based). Our implementation is based on a publicly-available GitHub repository.[19] A performance comparison is made between the use of two different feature extractors/backbones in the proposed framework, ResNet 50 and ResNet 101.[20] Both backbones were trained, initialized with COCO-pretrained weights, on ADI and AEI datasets separately for the semiconductor defect detection task. The inference speed is also recorded for both backbones in terms of required total inference time and pure compute time per image. Per-class Average Precision (AP) and mean Average Precision (mAP) across all defect types are used as performance metrics. The (m)AP results depend on the Intersection over Union (IoU) and confidence thresholds used. The confidence score ranges between 0.0 to 1.0 and is given by the model to express its confidence in the corresponding defect class prediction. By using a confidence threshold, only defect predictions with a confidence score above the threshold are considered when calculating (m)AP. IoU expresses how much the ground truth and predicted bounding box overlap. IoU threshold determines how much this overlap has to be for a prediction to be considered correct.

For each dataset, models are trained for 1000 epochs and its performance on the validation dataset is recorded every 50 epochs at confidence threshold 0.33. The per defect class AP and mAP achieved for IoU's between 0.5 and 0.95 at an interval of 0.05 as well as mAP at an IoU threshold of 0.5 are recorded. The epoch with the best mAP 0.5:0.95 is reported along with the other performance metrics at that epoch.

4.2 Model transfer

Another experiment has been conducted (by means of transfer learning), where instead of COCO-pretrained weight initialization, weight parameters corresponding to the highest mAP @IoU0.5:0.95 as described in 4.1 are initialized for finetuning on each dataset. This means the best weight parameters for ADI dataset have been finetuned on AEI dataset and vice versa. This is to validate the possibility of optimizing the model training time (requiring fewer epochs) and other computational resources by applying transfer learning strategy (which means instead of using random weights or weights from completely different data distribution [like COCO], we will initialize weight parameters trained on SEM image features). Hence, the performance is recorded every 20 epochs to monitor this initial behavior. The epoch where the highest mAP IoU0.5:0.95 is achieved on the validation dataset is recorded. Models are finetuned for 500 epochs and the precision and required epochs are compared to the experiments from section 4.1. The best performance achieved by the finetuned model (initialized with custom SEM dataset weight) is compared to that of the models from section 4.1 (trained with COCO-pretrained weights).

5. RESULTS AND DISCUSSION

5.1 Benchmarking

5.1.1 ADI dataset

The model was evaluated every 50 training epochs and we report the performance of the models for the epoch where the highest validation mAP IoU0.5:0.95 was recorded. Figure 4 shows detection prediction examples on the ADI dataset. Tables 3 and 4 show the performance on the validation and test datasets respectively. ResNet

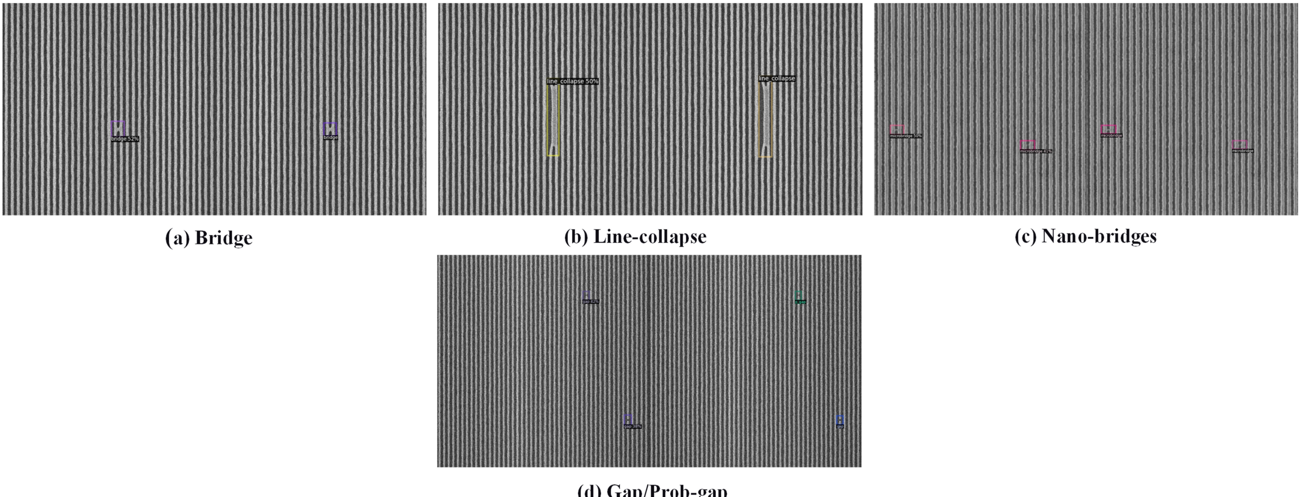


Figure 4. Detection results for each defect type on ADI data. For each pair, (left) model prediction against (right) ground truth.

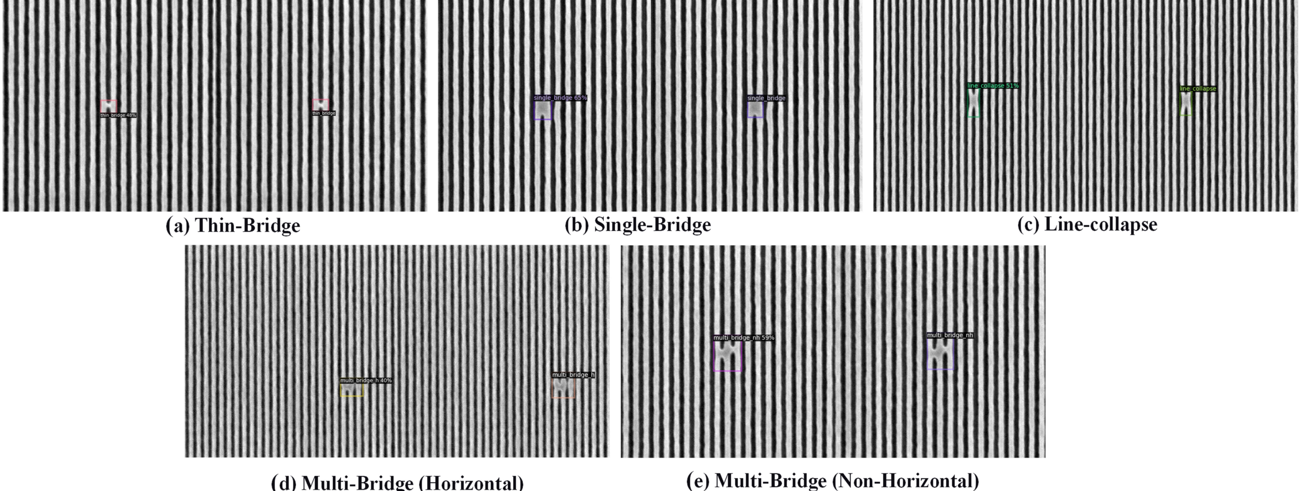


Figure 5. Detection results for each defect type on AEI data. For each pair, (left) model prediction against (right) ground truth.

50 significantly outperformed ResNet 101 on mAP on both validation and test dataset. However, there is a significant drop (per class AP) for bridge defect for ResNet 50 backbone, from **35.69** in the validation phase to **8.78** in the test phase. This drop is a clear outlier compared to other defect types, which requires further investigation.

Table 3. Per class and overall average precision on ADI validation data of backbones trained with COCO-pretrained weights with confidence threshold 0.33. Epoch with best mAP IoU0.5:0.95 on validation was selected. Best values in **bold**.

Backbone	**Epoch**	**Per class AP IoU0.5:0.95**					**mAP**	**mAP**
		microbridge	gap	bridge	p_gap	line_collapse	IoU0.5:0.95	IoU0.5
ResNet 50	750	**15.50**	**31.50**	**35.69**	**3.39**	72.61	**31.74**	**50.97**
ResNet 101	450	5.42	15.23	32.57	0.94	**74.28**	25.69	42.24

Table 4. Per class and overall average precision on ADI test data of backbones trained with COCO-pretrained weights with confidence threshold 0.33. Epoch with best mAP IoU0.5:0.95 on validation was selected. Best values in **bold**.

Backbone	**Epoch**	**Per class AP IoU0.5:0.95**					**mAP**	**mAP**
		microbridge	gap	bridge	p_gap	line_collapse	IoU0.5:0.95	IoU0.5
ResNet 50	750	**14.96**	**30.50**	8.78	**3.26**	70.63	**25.63**	**42.77**
ResNet 101	450	8.77	14.64	**24.06**	1.06	**73.29**	24.36	38.34

5.1.2 AEI dataset

Similar to the previous section (5.1.1), the model was evaluated every 50 epochs and the best performance on the validation dataset is reported. Figure 5 shows detection prediction examples on the AEI dataset. Tables 5 and 6 show the performance on the validation and test dataset respectively. ResNet 101 significantly outperforms ResNet 50. While the performance difference between the two backbones remains significant on both test and validation, the per class AP by ResNet 101 on the Multi Bridge Horizontal (MBH) defect type drops from **34.85** on validation to **4.36** on test. Similar to section 5.1.1, the reasoning behind such a large drop can be investigated in future work. Additionally, ResNet 50 fails to detect any Multi Bridge Non Horizontal (MBNH) defects and achieves extremely small AP values of **4.32** and **6.95** on thin bridge defect type. Stochastic failure analysis on certain defect types may be done in future work.

Table 5. Per class and overall average precision on AEI validation data of backbones trained with COCO-pretrained weights with confidence threshold 0.33. Epoch with best mAP IoU0.5:0.95 on validation was selected. Best values in **bold**.

Backbone	**Epoch**	**Per class AP IoU0.5:0.95**					**mAP**	**mAP**
		thin_bridge	single_bridge	MBNH	MBH	line_collapse	@0.5:0.95	@0.5
ResNet50	200	4.32	15.65	0.00	26.48	**65.84**	22.46	39.46
ResNet101	450	**26.99**	**38.78**	**28.59**	**34.85**	40.406	**33.92**	**53.18**

Table 6. Per class and overall average precision on AEI test data of backbones trained with COCO-pretrained weights with confidence threshold 0.33. Epoch with best mAP IoU0.5:0.95 on validation was selected. Best values in **bold**.

Backbone	**Epoch**	**Per class AP IoU0.5:0.95**					**mAP**	**mAP**
		thin_bridge	single_bridge	MBNH	MBH	line_collapse	@0.5:0.95	@0.5
ResNet50	200	6.95	32.60	0.00	**20.17**	**57.75**	23.49	39.14
ResNet101	450	**24.67**	**42.77**	**32.22**	4.36	45.23	**29.85**	**54.25**

5.2 Model Transfer

Tables 7 and 8 show the comparison in validation precision for models trained in sections 5.1.1 and 5.1.2 against models that have been fine-tuned with weight parameter initialization, trained on SEM image features. On the ADI dataset, finetuning the model from the AEI dataset pretrained weight shows a significant decrease in the number of required epochs before the best precision on validation data is reached. Retraining the ResNet 50-based model required only 80 epochs instead of 750 epochs before it reached its highest precision for the transfer learning strategy. For ResNet 101, 100 epochs instead of 450 epochs of training before convergence was observed as well as an improvement in detection mAP. On the AEI dataset, both significantly outperform against backbones trained with COCO-pretrained weight parameter. ResNet 50 and ResNet 101 (finetuned with ADI dataset-pretrained weight parameter) improve AEI test mAP @IoU0.5:0.95 by **37.42%** and **19.50%**, against models trained with COCO-pretrained weight parameter. In general observation, all models trained/fine-tuned with SEM dataset pretrained weights converge faster than models initialized with random/COCO-pretrained weights, especially on ADI dataset training. In general, AEI dataset is less complex compared to the ADI dataset. Most images have only one defect instance per image and attribution of noise pixels and contrast change scenario is less significant. Hence our experimental observations indicate that weights pretrained on semiconductor image datasets can be advantageous for overall training from scratch and/or fine-tuning as well as when mainly using a model pretrained on complex semiconductor data and finetuning it for less complex data.

Table 7. mAP scores and epoch needed on ADI dataset for backbones trained with COCO-pretrained weights against backbones trained with AEI dataset-pretrained weights. Best metric achieved per backbone in **bold**.

Backbone	Normal initialization			Finetuning from AEI			
	Epoch	Val mAP IoU0.5:0.95	Test mAP IoU0.5:0.95	Epoch	Val mAP IoU0.5:0.95	Test mAP IoU0.5:0.95	Test mAP IoU0.5
ResNet 50	750	**31.74**	**25.63**	**80**	22.92	23.23	38.64
ResNet 101	450	25.69	24.63	**100**	**28.43**	**28.28**	47.89

Table 8. mAP scores and epoch needed on AEI dataset for backbones trained with COCO-pretrained weights against backbones trained with AEI dataset-pretrained weights. Best metric achieved per backbone in **bold**.

Backbone	Normal initialization			Finetuning from ADI			
	Epoch	Val mAP IoU0.5:0.95	Test mAP IoU0.5:0.95	Epoch	Val mAP IoU0.5:0.95	Test mAP IoU0.5:0.95	Test mAP IoU0.5
ResNet50	200	22.46	23.49	**160**	**34.68**	**32.28**	61.04
ResNet101	450	**33.92**	29.85	**440**	31.23	**35.67**	61.47

5.3 Comparison to previous work

Table 9 shows the inference speed and mAPs achieved on the ADI dataset by the proposed SEMI-CenterNet and previously studied models from Ref. 1. mAPs are all reported on the test data split. While the mAP achieved by SEMI-CenterNet-based models is worse compared to previous work, it does achieve the lowest inference time yet. Compared to the second fastest model (YOLOv7), SEMI-CenterNet with ResNet 50 backbone achieves **55.94%** improvement in inference time per image. This demonstrates models based on the CenterNet architecture are capable of faster inference and show potential, but further investigation on lower precision is required before it can be proposed as a valid alternative to models from previous works.[11,16,18]

Table 9. Inference time and mAP performance comparison between different models on ADI dataset. Best values in **bold**.

Model	mAP IoU0.5	Inference time (ms / image)
Faster R-CNN	0.825	56.8
Faster R-CNN (+NWD)	0.827	60.2
DINO (ResNet-50)	**0.865**	108.7
DINO (SWIN-Tiny)	0.769	119.2
RetinaNet (ResNet-152)[18]	0.788	78.2
YOLOv7[11]	0.843	20.2
Proposed SEMI-CenterNet (ResNet-50)	0.472	**8.9** 8.7 (pure compute time)
Proposed SEMI-CenterNet (ResNet-101)	0.479	13.4 12.8 (pure compute time)

6. CONCLUSION

In this work, the SEMI-CenterNet framework was proposed to reduce the computation time for DL-based semiconductor Automated Defect Classification and Detection. Due to its anchorless architecture, the model makes fewer predictions on background image sections, leading to faster inference. ResNet 50 and 101 backbones were investigated on both an ADI and AEI dataset. Despite its smaller size, ResNet 50 performed best on the ADI dataset, while ResNet 101 showed better performance on AEI dataset. Moreover, it was shown that using the model weights obtained by training on one semiconductor wafer dataset to retrain/fine-tune and retraining them on other wafer datasets (ADI/AEI) causes models to converge to similar performance faster since the model had learned various semiconductor defect pattern features and distributions. Finally, we compared our proposed framework against previous research works on the same ADI dataset, both on precision and inference time. Future work can be extended toward analyzing the deteriorated performance of certain classes in certain situations to improve detection precision metrics.

REFERENCES

[1] Dehaerne, E., Dey, B., and Halder, S., "A comparative study of deep-learning object detectors for semiconductor defect detection," in [*2022 29th IEEE International Conference on Electronics, Circuits and Systems (ICECS)*], 1–2, IEEE (2022).

[2] Cheon, S., Lee, H., Kim, C. O., and Lee, S. H., "Convolutional neural network for wafer surface defect classification and the detection of unknown defect class," *IEEE Transactions on Semiconductor Manufacturing* **32**(2), 163–170 (2019).

[3] Dey, B., Dehaerne, E., and Halder, S., "Towards improving challenging stochastic defect detection in sem images based on improved yolov5," in [*Photomask Technology 2022*], **12293**, 28–37, SPIE (2022).

[4] Yang, Y.-F. and Sun, M., "Semiconductor defect detection by hybrid classical-quantum deep learning," in [*2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*], IEEE (jun 2022).

[5] Li, Z., Wang, Z., and Shi, W., "Automatic wafer defect classification based on decision tree of deep neural network," in [*2022 33rd Annual SEMI Advanced Semiconductor Manufacturing Conference (ASMC)*], 1–6, IEEE (2022).

[6] Zhou, X., Wang, D., and Krähenbühl, P., "Objects as points," *arXiv preprint arXiv:1904.07850* (2019).

[7] Duan, K., Bai, S., Xie, L., Qi, H., Huang, Q., and Tian, Q., "Centernet: Keypoint triplets for object detection," in [*Proceedings of the IEEE/CVF international conference on computer vision*], 6569–6578 (2019).

[8] Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., and Zitnick, C. L., "Microsoft coco: Common objects in context," in [*Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V 13*], 740–755, Springer (2014).

[9] Hoiem, D., Divvala, S. K., and Hays, J. H., "Pascal voc 2008 challenge," *World Literature Today* **24**(1) (2009).

[10] Albahli, S. and Nazir, T., "Ai-centernet cxr: An artificial intelligence (ai) enabled system for localization and classification of chest x-ray disease," *Frontiers in Medicine* **9**, 955765 (2022).

[11] Dehaerne, E., Dey, B., Halder, S., and De Gendt, S., "Optimizing yolov7 for semiconductor defect detection," in [*Metrology, Inspection, and Process Control XXXVII*], **12496**, 635–642, SPIE (2023).

[12] Kim, J., Nam, Y., Kang, M.-C., Kim, K., Hong, J., Lee, S., and Kim, D.-N., "Adversarial defect detection in semiconductor manufacturing process," *IEEE Transactions on Semiconductor Manufacturing* **34**(3), 365–371 (2021).

[13] Elharrouss, O., Akbari, Y., Almaadeed, N., and Al-Maadeed, S., "Backbones-review: Feature extraction networks for deep learning and deep reinforcement learning approaches," *arXiv preprint arXiv:2206.08016* (2022).

[14] Ren, S., He, K., Girshick, R., and Sun, J., "Faster r-cnn: Towards real-time object detection with region proposal networks," *Advances in neural information processing systems* **28** (2015).

[15] Redmon, J., Divvala, S., Girshick, R., and Farhadi, A., "You only look once: Unified, real-time object detection," in [*Proceedings of the IEEE conference on computer vision and pattern recognition*], 779–788 (2016).

[16] Dey, B., Dehaerne, E., Halder, S., Leray, P., and Bayoumi, M. A., "Deep learning based defect classification and detection in SEM images: a mask R-CNN approach," in [*Metrology, Inspection, and Process Control XXXVI*], **PC12053**, PC120530K, SPIE (2022).

[17] Nag, S., Makwana, D., Mittal, S., Mohan, C. K., et al., "Wafersegclassnet-a light-weight network for classification and segmentation of semiconductor wafer defects," *Computers in Industry* **142**, 103720 (2022).

[18] Dey, B., Goswami, D., Halder, S., Khalil, K., Leray, P., and Bayoumi, M. A., "Deep learning-based defect classification and detection in sem images," in [*Metrology, Inspection, and Process Control XXXVI*], PC120530Y, SPIE (2022).

[19] Wang, F., "Centernet-better." https://github.com/FateScript/CenterNet-better (2020).

[20] He, K., Zhang, X., Ren, S., and Sun, J., "Deep residual learning for image recognition," in [*Proceedings of the IEEE conference on computer vision and pattern recognition*], 770–778 (2016).

3D SEM metrology of microstructures for high volume manufacturing

Zeinab Abdallah[a], Aurélien Fay[a], and Stéphane Bonnet[a]

[a]Univ. Grenoble Alpes, CEA, Leti, F-38000, Grenoble, France

ABSTRACT

Scanning Electron Microscopy (SEM) uses surface-emitted Secondary Electrons (SE) to produce grayscale level images. From these SEM images, an observer can mentally represent the topography of the imaged object. Numerical techniques can additionally provide a quantitative assessment of the topography, which is one of the major challenge in the today microelectronics industry for advanced nodes and more-than-moore devices. In this paper, we propose and validate a non-destructive, fast 3D metrology strategy for microscopic object. Various microlens-like 3D structures are patterned on 300 nm wafers by grayscale i-line lithography. The lenses are imaged by a SEM tool with a four-quadrant SE detector, and also measured by Atomic Force Microscopy (AFM) which serves as 3D reference metrology. Both, AFM and SEM data are co-registered and background is removed. An analytical SEM model is used both for SEM detector calibration and parametric reconstruction. First, we observe that this model poorly describes the entire SEM images of the device under test, partly due to electron screening and reemission effects. Secondly, by weighting the image depending on the quadrant orientation, we show that this model can still be used for both calibration and parametric reconstruction. The parametric reconstructions of microlenses with different footprints and heights match well the reference geometries.

Keywords: Scanning electron microscopy, 3D metrology, microlens, grayscale lithography, image sensor, four-quadrant detector system.

1. INTRODUCTION

CMOS image sensors are valuable due to their cost-effectiveness and compatibility with integrated circuit manufacturing.[1] This makes them ideal to benefit from the developments in silicon process driven by Moore's law, resulting in improved quantum efficiency and the ability to address wider application fields. These sensors are involved in multiple applications, including automotive sensors, spatial technologies and medical instrumentation, besides the conventional video and photography. The primary factor in increasing the sensor performance at every generation step is the reduction of the pixel size.[2] Fig. 1 shows that inside each pixel of the image sensor, there is a silicon photodiode that captures photons. These diodes are connected to transistors that generate, collect, transfer, and measure these charges. It is important to take into consideration that the region of the photodiode is sensitive to light, but not all the pixel surface, which means not all the pixel surface can collect photons. A layer of microlenses is used to focus the incident light onto these photodiodes[2–4]

To increase the efficiency of the CMOS sensor, every microlens should have a high coverage of the pixel, be light-transparent, and must have a target shape. Consequently, the shape of the microlens should be optimized and controlled. The main focus of our research is to investigate the shape of these microlenses using SEM technology.

The most common techniques to get access to the third dimension are by Scanning Probe Microscopy (SPM). Two widely employed methods within SPM are the Atomic Force Microscope (AFM) and Scanning Tunneling Microscope (STM). Despite the high resolution and non-destructiveness of AFM, it is relatively slow when acquiring three-dimensional data. On the other hand, STM offers an atomic-scale resolution with low throughput, it is generally considered as a more destructive technique due to its requirements of the deposition of surface layers.[6] Another method that can be used to obtain three-dimensional information is Focused Ion Beam Scanning

Further author information: (Send correspondence to Zeinab Abdallah)
Zeinab Abdallah: E-mail: zeinab.abdallah@cea.fr, Telephone: +33 (0)4 38 78 27 03

38th European Mask and Lithography Conference (EMLC 2023), edited by
Uwe F.W. Behringer, Jo Finders, Proc. of SPIE Vol. 12802, 128020M
© 2023 SPIE · 0277-786X · doi: 10.1117/12.2675620

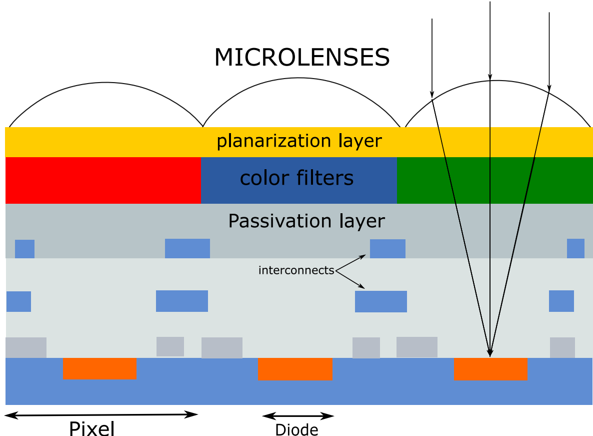


Figure 1. Sectional side view of CMOS image sensor, drawing inspired from [5].

Electron Microscopy (FIB-SEM). FIB-SEM combines the capabilities of both ion milling and SEM imaging to sequentially remove material layers and capture images at each step, allowing for the reconstruction of three-dimensional structures. It is important to note that FIB-SEM is a destructive technique as it involves the removal of material during the imaging process [7]. SEM images demonstrate a realistic shading effect that enables the observer to mentally recover the topography of the sample. Shortly after the introduction of SEM, researchers were taking advantages of this contrast and developing different techniques to reconstruct the topography of the sample using SEM images.[8–12] Expanding on these progressions, the main objective of this paper is to use shaded SEM images to achieve fast, in-line and non-destructive 3D reconstruction of microlenses.

The paper is structured into several sections. The first section outlines the experimental setup, including details on microlens fabrication and imaging. In the methodology section, the reconstruction problem formulation is presented, along with the SEM model formulation and image calibration process. The results and discussion section evaluates the calibrated SEM image and compares the reconstructed height maps with experimental data obtained from AFM. Finally, the conclusion and perspectives section summarizes the key results and outlines future research directions. The appendices contain additional technical details and the derivation of the forward SEM model to support the main content of the paper.

2. EXPERIMENTS

Experiments are performed in this study to evaluate and validate the proposed SEM-based 3D reconstruction strategy. In this section, we describe the experimental procedure. First, we discuss the fabrication process of microlenses. Next we proceed into the imaging process which involves using both Atomic Force Microscopy (AFM) and Scanning Electron Microscopy (SEM).

2.1 Microlens Fabrication

Arrays of microlenses of different heights, radii and pitches were fabricated using grayscale lithography, which is an alternative i-line lithography method that enables the patterning of 3D microstructures in a photoresist material. The process involves the use of sub-resolution dots on the optical mask and a low-contrast resin.[13–15] Fig. 2.(a) shows a unitary cell of the optical mask used. This cell consists of chromium dots that have characteristic pitches below 225 nm. It is worth noting that pitch is smaller than the wavelength of the incoming light ($\lambda_{light} = 365$ nm). In our experimental setup, grayscale lithography was performed on four different 300 mm wafers, with each wafer being exposed to a specific dose of light at a constant and optimal focus. This enables the creation of a variety of microlens structures with different geometric parameters. Fig. 2.(b) shows an AFM-characterized microlens obtained using grayscale lithography.

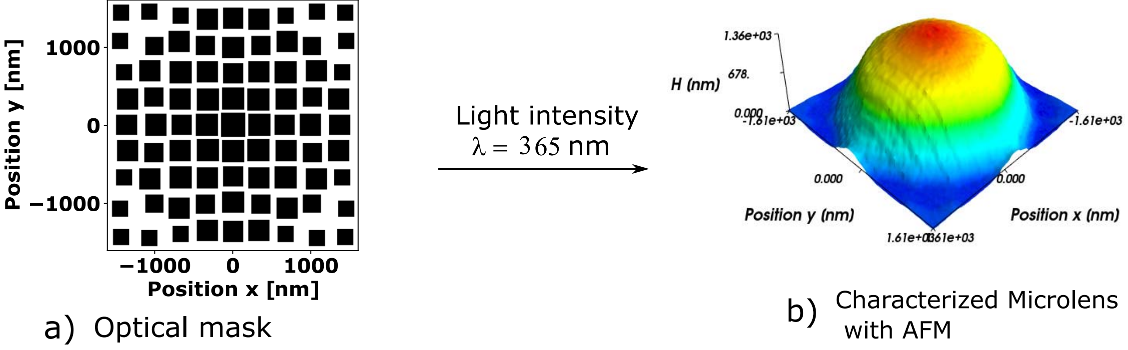


Figure 2. a) Unitary cell of the optical mask used in grayscale lithography process, b) AFM-characterization of microlens obtained from grayscale lithography, taken from [13].

2.2 Microlens Imaging

After the microlens fabrication, SEM imaging is performed. In our study, the SEM tool is equipped with four secondary electron (SE) quadrants configuration, where each quadrant gives as an output one SEM image. This means that for each 3D object, we obtain four 2D SEM images corresponding to the four SE quadrants labeled as A, B, C and D, as shown in Fig. 3.(a). In this study, we selected different microlenses with varying radius and spacing in order to investigate the influence of those parameters on the methodology performance. Fig. 3.(b) shows the output of quadrant A with a central line profile, within this profile, we observe a peak in the SE intensity. This peak indicates a higher concentration of SE intensity in a specific direction. More specifically, the peak is observed in the direction of the active detector.

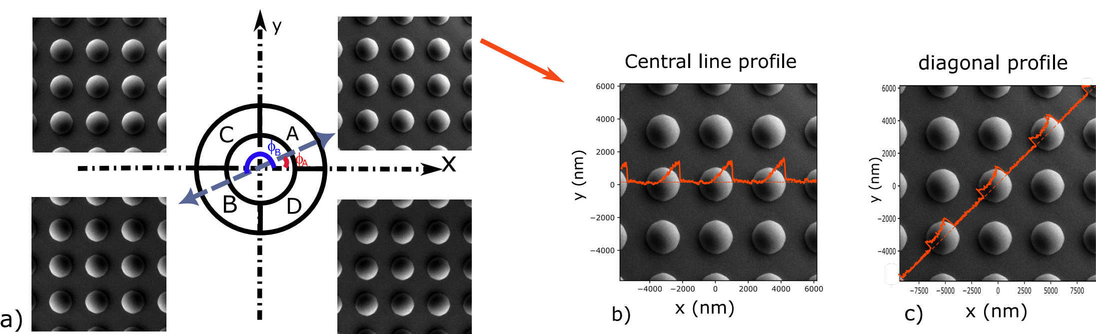
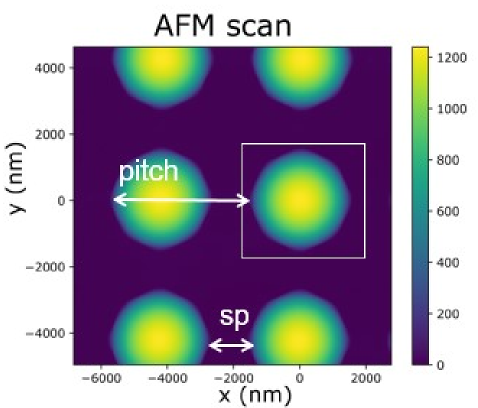


Figure 3. a) Four SEM images of microlens obtained from the SEM tool, b) output of quadrant A with a central profile cut, c) output of quadrant A with a diagonal profile cut

In parallel to SEM imaging, topographical-data of the microlenses are obtained using AFM and serve as reference data, providing the ground truth profile of the microlens surface.

After having collected both SEM and AFM data, a flow of preprocessing steps has been performed. The SEM images were registered with respect to their corresponding AFM scans to guarantee a precise spatial alignment between the two imaging modes. Additionally, unitary cells were extracted from both the registered SEM and AFM arrays. Fig. 4 shows an AFM array of microlenses with a central line profile of an extracted-registered unitary cell.



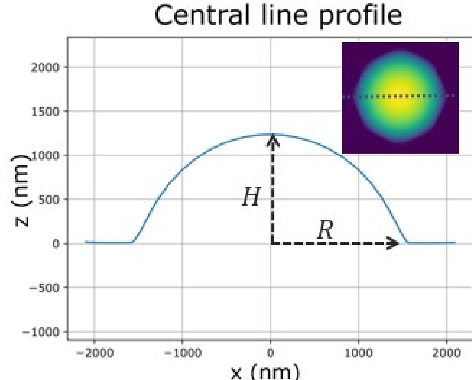


Figure 4. (left) Microlens array characterized using AFM. (right) central line profile from a registered AFM cell.

3. METHODOLOGY

In the following, the 3D reconstruction algorithm will be explained in detail. The work is divided into two main parts, the forward SEM model and the inverse SEM model as shown in Fig. 5. The forward model involves going from the 3D surface of the microlens into the SEM images taking into account the physics of the imaging process (blue arrow in Fig. 5). On the other hand, the inverse model is focused on performing the 3-D reconstruction, which means going from the images to the 3D object itself, by solving a minimization mathematical problem (red arrow in Fig. 5).

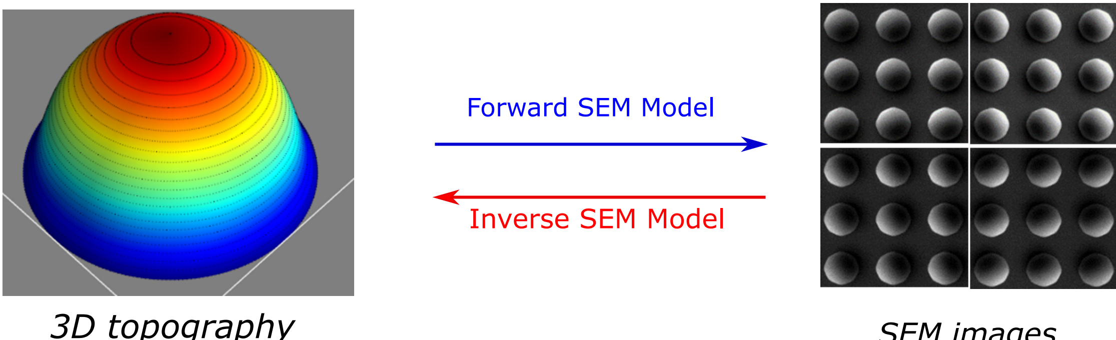


Figure 5. Schematic illustrating the two main parts of the 3D reconstruction algorithm: left) 3D microlens, right) SEM images with a four-quadrant SEM tool.

3.1 Forward SEM model

In this section, we will first provide an overview of the existing forward SEM models. We will then introduce the forward model designed to work in multiple quadrant arrangements, that will be adapted to our experimental setup

3.1.1 History and background

Several forward SEM models have been developed for simulating SEM images, including physically-based models like Monte Carlo simulations,[16,17] with JMONSEL software being the most well-known in the microelectronics field[17] due to its main advantage which is the ability to realistically simulate SE emission.[18,19]

Alternatively, there are analytically-based models such as the analytical slope-based model, where the intensity I at a point (x, y) is given as a function of the slope of the surface at the same referred point (x, y) on the surface. More precisely, I is given as a function of the incident angle θ_N (angle between the electron beam and the normal to the surface). This relationship is established through the reflectance map R that provides

information about the surface's reflectance properties at different incident angles.[8,12,20,21] The concept of a reflectance map was first introduced by Horn [20]. The SE intensity can be written as:

$$I = I_0 R(\theta_N), \tag{1}$$

where $R(\theta_N)$ is the normalized reflectance map (i.e: $R(0°) = 1.0$) and I_0 represents the angular current density at $\theta_N = 0°$. It is expressed as a function of the primary electron current I_{PE} as: $I_0 = I_{PE}\delta_0$ with δ_0 the SE yield at $\theta_N = 0°$. Defining the SE intensity as I_{SE}, the SE yield at a position (x, y) is defined by the amount of SE detected from this position by:

$$\delta = \frac{I_{SE}}{I_{PE}} \tag{2}$$

Over the years, several approximations of the reflectance map for SE emission have been proposed and compared to real SE emission behavior.[12,20,22] However, this model is typically used with a standard detector configuration placed directly above the imaged surface. Our research involves the use of multiple quadrant setups in SEM, which indicates the forward model associated to each quadrant is a fraction of the overall system described in Eq. 1. Hence our forward model relies on additional variables beyond just the angle of incidence such as the quadrant position. Note that the slope-based model holds only when the characteristic size of the electron-matter interaction volume is small compared to the curvature radius at any point of the imaged surface.[23,24]

3.1.2 Forward SEM model in four-quadrant arrangement

In multiple-quadrant configuration, each shaded image has its own intensity value, which is determined by both the characteristic angles of the surface and the detector's position. A system of four detectors placed above the specimen as shown in Fig. 6 produces four detector signals: I_A, I_B, I_C, I_D. These signals are linked into a set of four images shown in Fig. 3.(a), where the summation of these signals is equal to the overall intensity described in Eq. 1. The theoretical model of the multi detector approach relies on Lambert's angular distribution of SEs.[25]The SE intensity flowing along a straight trajectory toward quadrant A as in Fig. 6 can be written as:

$$I_A = I_0 R(\theta_N) \cdot \frac{1}{\pi} \int_{\Omega_A} \cos(\gamma) d\Omega \tag{3}$$

where Ω_A represents the solid angle of quadrant A, γ is the take-off angle of the secondary electron. From the geometry described in Fig. 6, we have: $\cos(\gamma) = \mathbf{n}^T \mathbf{v}$, where $\mathbf{n}$ is the surface normal vector and $\mathbf{v}$ is the initial velocity vector of the SE. Expressing the vectors in spherical coordinates and integrating this formula over our quadrant A surface with performing additional mathematical operations described in the Appendix A, Eq. (3) takes the form:

$$I_A = I_0 R(\theta_N) \cdot \frac{1}{\pi} [d_A \cos(\phi_A - \phi_N) \sin(\theta_N) + c_A \cos(\theta_N)], \tag{4}$$

where ϕ_N, and θ_N are respectively the surface azimuth and surface inclination angles. Parameters c_A and d_A are coefficients depending on quadrant's geometry and position, and are defined as:

$$c_A = \Omega_A \cos(\theta_A) \cos\left(\frac{\Delta\theta_A}{2}\right) \tag{5}$$

$$d_A = [\Delta\theta_A - \cos(2\theta_A) \sin(\Delta\theta_A)] \sin\left(\frac{\Delta\phi_A}{2}\right) \tag{6}$$

From the various available reflectance maps, we have selected the inverse cosine reflectance function [26], that is given by:

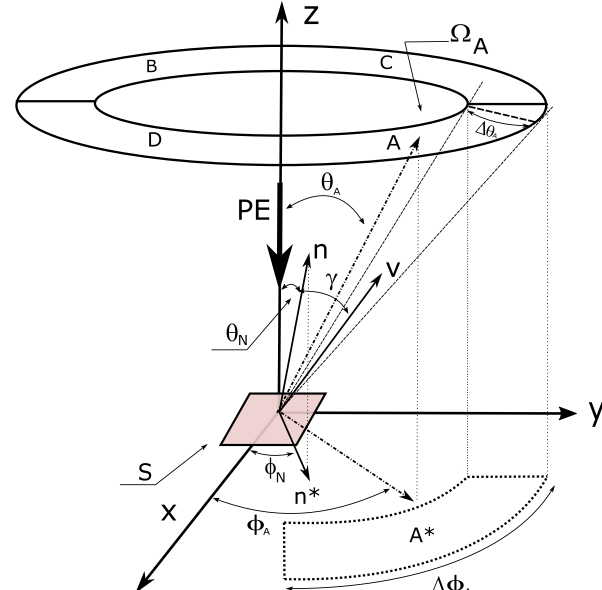


Figure 6. Characteristic angles in a simplified detector system: A, B, C, D = quadrant detector, PE = primary electrons, n= surface normal vector, ϕ_N, θ_N- surface azimuth and inclination angles.

$$R(\theta_N) = \frac{1}{\cos(\theta_N)} \tag{7}$$

Substituting $R(\theta_N)$ from Eq. (7)by its value in Eq. (4), it simplifies into:

$$I_A = \frac{I_0}{\pi} \left[d_A \cos(\phi_A - \phi_N) \tan(\theta_N) + c_A \right] \tag{8}$$

The above formulation of Eq. (8) was originally derived by W. Slówko *et al.* [27]. In their study, the detectors were assumed to be a part of a hemisphere. Despite this difference in the setup, we arrived to the same formula as described in Appendix A. This formula has been extended and used by many other research studies that involve reconstruction using multiple detector arrangements.[27–30]

A last ingredient is to include the detector gain s_A to match the experimental SEM image. The simulated SEM image is finally given by

$$I_A^{\text{SIMU}}(\theta_N, \phi_N) = s_A I_A = \tilde{d}_A \tan(\theta_N) \cos(\phi_A - \phi_N) + \tilde{c}_A, \tag{9}$$

with $\tilde{d}_A = s_A I_0 d_A / \pi$ and $\tilde{c}_A = s I_0 c_A / \pi$.

Eq. (8) and Eq. (9) serve as our comprehensive forward model, effectively used for both calibration and reconstruction purposes. Before proceeding, it is important to know how to compute the characteristic angles of the surface θ_N and ϕ_N as well as the azimuth angles of the quadrants $(\phi_A, \phi_B, \phi_C, \phi_D)$.

Characteristic angles computation (θ_N, ϕ_N) To accurately compute (ϕ_N, θ_N), which are dependent on the surface orientation, it is essential to compute the surface normal at each position (x, y). The unit normal of the surface can be written in the form:

$$\mathbf{n} = \begin{pmatrix} \cos\phi_N \sin\theta_N \\ \sin\phi_N \sin\theta_N \\ \cos\theta_N \end{pmatrix} = \cos\theta_N \begin{pmatrix} \cos\phi_N \tan\theta_N \\ \sin\phi_N \tan\theta_N \\ 1 \end{pmatrix} = \frac{1}{\sqrt{1 + p^2 + q^2}} \begin{pmatrix} -p \\ -q \\ 1 \end{pmatrix} \tag{10}$$

where $p = \frac{\partial z}{\partial x}, q = \frac{\partial z}{\partial y}$ denotes the gradient of the surface with respect to spatial coordinates. This formula relates the spatial derivatives of the surface height to the normal spherical coordinates. Following Eq. (10) and the geometry of the figure, θ_N and ϕ_N are given by:

$$\theta_N = \arccos\left(\frac{1}{\sqrt{1+p^2+q^2}}\right) \tag{11}$$

$$\phi_N = \arctan\left(\frac{q}{p}\right) \tag{12}$$

Azimuthal quadrant shift computation $(\phi_A, \phi_B, \phi_C, \phi_D)$ Having explained how to determine ϕ_N and θ_N, the next step is to compute the azimuthal shift of the quadrants, which represents the angular displacement of each quadrant with respect to the x-axis. ϕ_A and ϕ_B are illustrated in Fig. 3 (a). In the computation of the four angles, we took the advantage of the rotation invariance of the microlenses. This property guarantees that any circular profile extracted from the microlens and centered at $(0, 0)$ will exhibit a constant height and slope along this circular profile, constant slope implies constant polar angle, and this means:

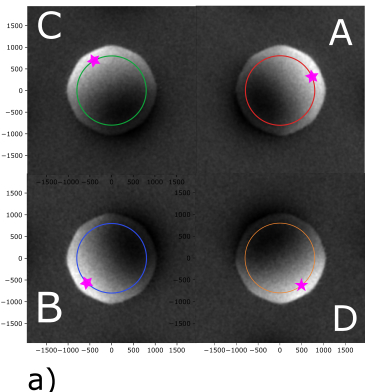
$$\begin{aligned} &\theta_N = \text{const} \\ &\quad \Rightarrow \tan(\theta_N) = \text{const} \\ &\quad \xrightarrow{\text{by Eq. (8)}} I_A = \frac{I_0}{\pi}\left[d_A \text{const} \cos(\phi_A - \phi_N) + c_A\right] \end{aligned} \tag{13}$$

Therefore, the intensity I_A reaches its maximum value if and only if $\cos(\phi_A - \phi_N) = 1.0$, which implies that $\phi_N = \phi_A$. Hence, to determine the four angles of interest, it is sufficient to identify the azimuthal angle at which one obtains maximum yield along an extracted circular path from the SEM image. Consequently, we performed circular profile extraction on smoothed and registered unitary cells shown in Fig. 7 (a). The variation of SE yield along the circular path as a function of the azimuthal angle is displayed in Fig. 7 (b). The maximum yield points are indicated by stars, and they correspond to the quadrant positions. Table (1) presents the azimuth angles for each quadrant, obtained from our methodology (column 1) and the reference angle values in (column 2) derived from the machine. The table demonstrates a close agreement between the values obtained and the reference values, furthermore, it is worth noting that opposite quadrants are approximately 180° apart . For our analysis, we will utilize the azimuth angle values obtained from our methodology.

Table 1. $(\phi_A, \phi_B, \phi_C, \phi_D)$ values obtained from our methodology (Column 1) and reference angles obtained from the SEM machine (Column 2).

Quad ID	ϕ_{quad} (°) - from SEM microlens images	ϕ_{quad} (°) - reference calibrated values
A	28	27.6
B	209	207
C	116	114
D	299	298.49

To complete the inputs of the forward model, we need to compute two parameters d_A and c_A. These parameters are essentially defining the gain and the offset characteristics of quadrant A, as will be explained in the next section.



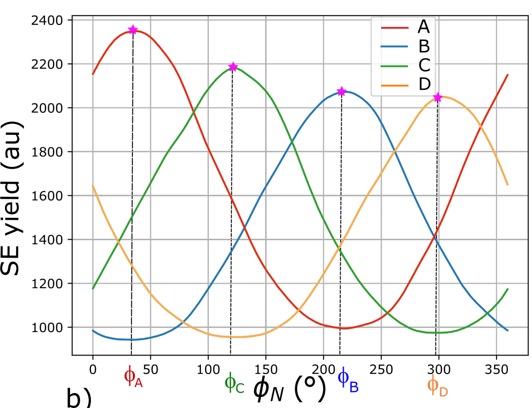


Figure 7. a) Experimental SEM images from the four quadrants indicating the circular paths representing the trajectory along which the SE yield is measured. b) SE yield as a function of the azimuthal angle along these circles.

3.1.3 Forward SEM model calibration

Calibrating the forward model involves determining the calibration parameters $\tilde{c}_A$ and $\tilde{d}_A$. In Eq. (9), it can be observed that $\tilde{c}_A$ corresponds to the intensity of the flat region i.e at $\theta_N = 0°$:

$$I_A^{\text{SIMU}}(0, \phi_N) = \tilde{c}_A \tag{14}$$

We choose the top of the microlens as the flat region to avoid any potential physical phenomena or variation in the background. It is important to note that for every individual image we determine an offset value $\tilde{c}_{A,..}$.

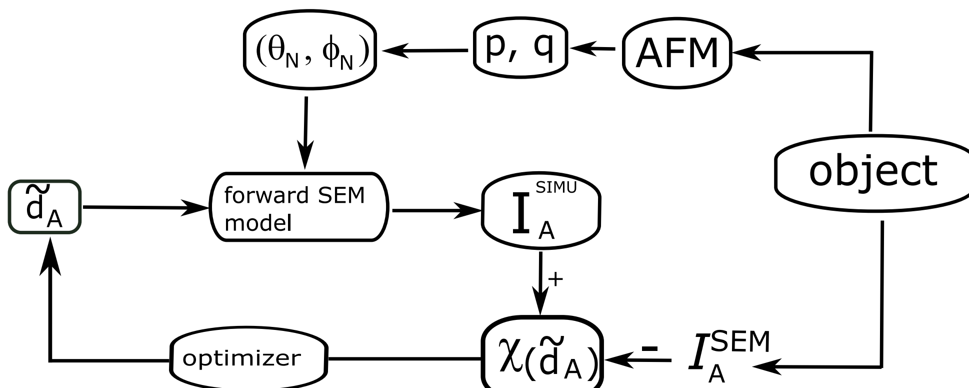


Figure 8. Schematic presentation of the calibration process for determining scaling factor $\tilde{d}_A$ in the forward model for quadrant A.

To determine the scaling factors $(\tilde{d}_A, \tilde{d}_B, \tilde{d}_C, \tilde{d}_D)$, an inverse problem should be solved. First, a trained lens is selected and $\tilde{c}_A$ is computed using Eq. (14). θ_N and ϕ_N are then calculated using AFM scan data of the lens. However, computing θ_N and ϕ_N requires the determination of the partial derivatives p and q. These derivatives are obtained by performing a 2D convolution of the AFM height map with the derivatives of the Gaussian function $G_\sigma(x, y)$. The expressions for the derivatives are given by:

$$p = z^{\text{AFM}} * \frac{\partial}{\partial x} G_\sigma(x,y) \qquad\qquad q = z^{\text{AFM}} * \frac{\partial}{\partial y} G_\sigma(x,y),$$

Next, by plugging the obtained values of θ_N, ϕ_N, $\tilde{c}_A$, and the pre-computed ϕ_A into Eq. (9), an intensity I_A^{SIMU} is obtained as described in Fig. 8 The objective is to find the optimal value of $\tilde{d}_A$ that minimizes the discrepancy between the simulated intensity I_A^{SIMU} and the experimental intensity I_A^{SEM}. This is achieved by minimizing the objective function:

$$\chi(\tilde{d}_A) = \frac{1}{2} \| \mathbf{W}_A \odot (\mathbf{I}_A^{\text{SIMU}}(\tilde{d}_A) - \mathbf{I}_A^{\text{SEM}}) \|_F^2, \tag{15}$$

where the weight $\mathbf{W}_A$ is used to assign importance to the SEM/A intensity measurement (will be explained in the next section). We proceed to determine the scaling factors $\tilde{d}_B$, $\tilde{d}_C$, and $\tilde{d}_D$ correspond to quadrants B, C, and D using the same trained lens. With the completion of the input parameters for the forward SEM model, we can now solve the inverse problem of reconstructing the microlens geometry from the SEM images.

3.2 Inverse SEM Model

In this section, we shift our focus towards the inverse SEM model. SEM-based topography reconstruction can be broadly categorized into two groups: multi-view and shape from shading, the first one uses techniques similar to stereo photography where the sample is imaged from two distinct angles allowing the topography reconstruction by parallax.[31–33] While these methods have generated successful results in reconstructing topography, they typically require the sample stage to be tilted at a known angle. The proposed method in this article goes in the second class. The shape from shading method uses the shading image to derive the slope of the surface and then perform reconstruction. In this section, we will provide an overview on reconstruction using shape from shading that had been used in such setup. We will then proceed to introduce our adapted inverse SEM model.

3.2.1 Shape from Shading

Within the state of the art of the shape from shading problem, the reconstruction process is divided into two parts: the first part involves using the shading images to compute the slope of the surface, while the second part is reconstructing the topography by integrating the derived slope information. By pairing detectors A-B and C-D and defining x' as a unit vector along A-B and y' along C-D, we can form two sets of orthogonal diagonally opposed detectors.[28] Using these detector pairs, the two gradient images are constructed as:

$$\frac{\partial z}{\partial x'} = \frac{\tilde{c}_B I_A - \tilde{c_A} I_B}{\tilde{d}_B I_A + \tilde{d}_A I_B}, \qquad\qquad \frac{\partial z}{\partial y'} = \frac{\tilde{c}_D I_C - \tilde{c}_C I_D}{\tilde{d}_D I_C + \tilde{d}_C I_D} \tag{16}$$

Where I_A, I_B, I_C and I_D are the forward model derivation described in Sec. 3.1.2. Integrating these variables enables to determine the surface topography profiles in x' and y' directions. Researchers have extensively employed this approach to achieve 3D recosntruction using shaded images in multiple detectors arrangement.[27–30, 34]

3.2.2 Proposed Inverse SEM model: Parametric Reconstruction

Microlens parametrization In contrast to the method described, our methodology involves the use of parametric reconstruction techniques. Given the nature of our target objects, which are microlenses, we take advantage of being able to analytically parameterize the geometry with a set of few parameters denoted as $\mathbf{p}$ such as:

$$z(\mathbf{p}; x, y) = H \left[1.0 - \left(\frac{x^2}{R^2} + \frac{y^2}{R^2} \right) \right]^{\frac{1}{k}}, |x| \le R, |y| \le R, \tag{17}$$

Eq. (17) represents mathematical model of a microlens. $\mathbf{p}$ represents the parameter vector that characterizes the microlens and in this case it is: $\mathbf{p} = \{R, H, k\}$, where H denotes the maximum height of the microlens, R represents its radius, and the exponent k controls the steepness of the surface profile (for $k = 2$ we obtain a perfect hemisphere). Fig. 9 visually demonstrates the change in profile associated with different values of k and highlighting R and H. Observing the AFM scans, the edges of microlenses do not perform a right angle with the footprint of the lens which highlight the non- ideal nature of the lens shape. Hence, the steepness exponent k in Eq. (17) covers this characteristic by allowing for more realistic representation of the lens geometry and better alignment with the AFM.

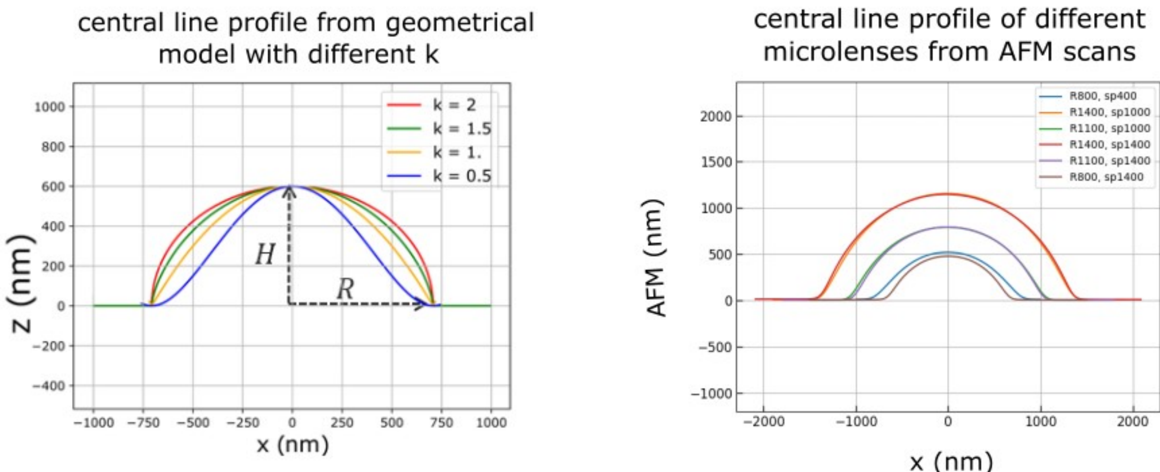


Figure 9. left) Central line profile of microlens geometry as defined in Eq. 17, right) central line profile from AFM scans of actual microlenses.

Reconstruction Flow Fig. 10 presents the inverse problem schematic, the microlens geometry is parameterized using $\mathbf{p}$. The next step after parameterizing the geometry, is to find the simulated intensity with respect to $\mathbf{p}$ by applying the forward SEM model explained in Sec. 3.1.2. The inputs of this model (Eq. 8) are the characteristic angles of the surface (ϕ_N, θ_N), the azimuth shift of the active detector (q) and the the parameters $\tilde{d}_q, \tilde{c}_q$ corresponds to detector q as shown in Fig. 10. The azimuth shifts of the detectors, (ϕ_A, ϕ_B, ϕ_C, ϕ_D) have been computed and validated in the previous section. Eq. (11), (12), provides the formulae for computing the characteristic angles of the surface. Hence, it is sufficient to compute the partial derivatives of the parameterized surface. With the analytical representation of the surface geometry, we take the advantage of being able to derive the analytical expressions of the partial derivatives (p, q) as follows:

$$p = \frac{\partial z}{\partial x}(\mathbf{p}; x, y) = (-2x)\,\frac{H}{kR^2}\left[1 - \frac{x^2}{R^2} - \frac{y^2}{R^2}\right]^{\frac{1-k}{k}} \tag{18}$$

$$q = \frac{\partial z}{\partial y}(\mathbf{p}; x, y) = (-2y)\,\frac{H}{kR^2}\left[1 - \frac{x^2}{R^2} - \frac{y^2}{R^2}\right]^{\frac{1-k}{k}} \tag{19}$$

Substituting Eqs (18), (19) in Eqs. (11), (12), we get θ_N and ϕ_N. By plugging θ_N, ϕ_N and ϕ_A into Eq. (9), we obtain the simulated SEM intensity by quadrant A. The aim is to find the optimal parameter values that minimizes the error between the simulated and experimental intensities, i.e minimizing the objective function:

$$\psi(\mathbf{p}) = \sum_{q\in\{A,B,C,D\}} \|\mathbf{W}_q \odot (\mathbf{I}_q^{\mathrm{SIMU}}(\mathbf{p}) - \mathbf{I}_q^{\mathrm{SEM}})\|_F^2 \tag{20}$$

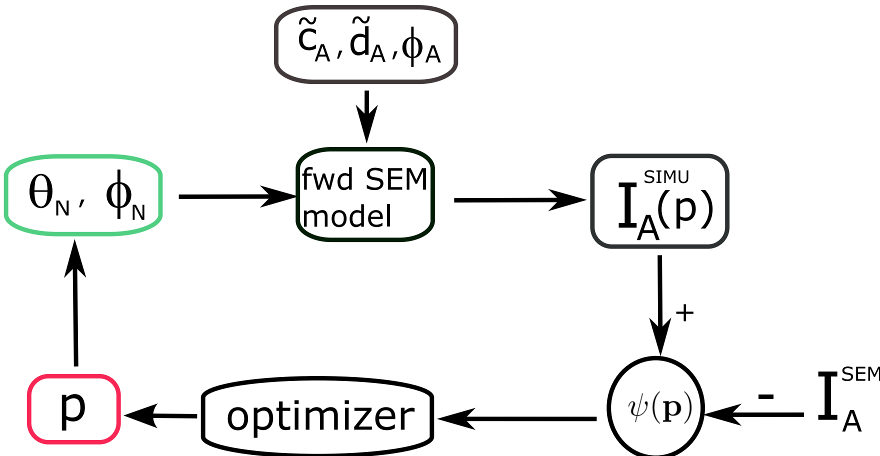


Figure 10. Schematic illustration of the inverse problem for obtaining the estimated parameters $\hat{\mathbf{p}}$

Where the weights $\mathbf{W}_q$ are used to assign importance to SEM intensity by the q-th detector. Hence, the estimated parameters $\hat{\mathbf{p}}$ are given by:

$$\hat{\mathbf{p}} = \arg_{\mathbf{p}} \min \psi(\mathbf{p}) \tag{21}$$

Ground truth parameters Once the estimated parameters are found, we compare them to the true parameters. The true parameters denoted as $\mathbf{p}^{\star}$ are found using AFM scans. The computation of true parameters is again obtained by solving an inverse problem as illustrated in Fig. 11. The goal is to find the optimal parameter values that result in the best fit between the geometrical model of the microlens as defined in Eq. (17) and AFM, which means minimizing the objective function:

$$\eta(\mathbf{p}) = \frac{1}{2} \|\mathbf{z}(\mathbf{p}; x, y) - \mathbf{z}^{AFM}\|_F^2 \tag{22}$$

Hence,

$$\mathbf{p}^{\star} = \arg_{\mathbf{p}} \min \eta(\mathbf{p}) \tag{23}$$

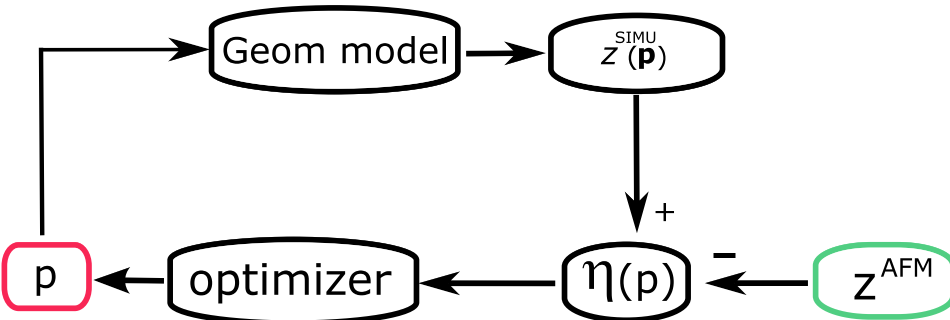


Figure 11. Schematic illustration of the inverse problem for obtaining true parameters $\mathbf{p}^{\star}$

4. RESULTS/ DISCUSSIONS

In this section, we will first present the results of finding the ground truth parameters using AFM. We will also show the calibrated model that has been generated based on the proposed methodology. Furthermore, we will provide results on lens reconstruction using SEM images.

4.1 AFM Reconstruction, True parameters

We were able to determine the $\mathbf{p}^{\star}$ of the different microlenses using the geometrical model defined in Eq. (17). To check the validity of the fit, we performed central line profile from both the geometrical model and AFM height maps shown in Fig.12, for a small microlens ($R < 800$ nm). and Fig. 13 for a large microlens ($R > 1100$ nm). The shape of the microlens shrinks as the dose increases providing a valuable insight into the fabrication process and its impact on lens properties

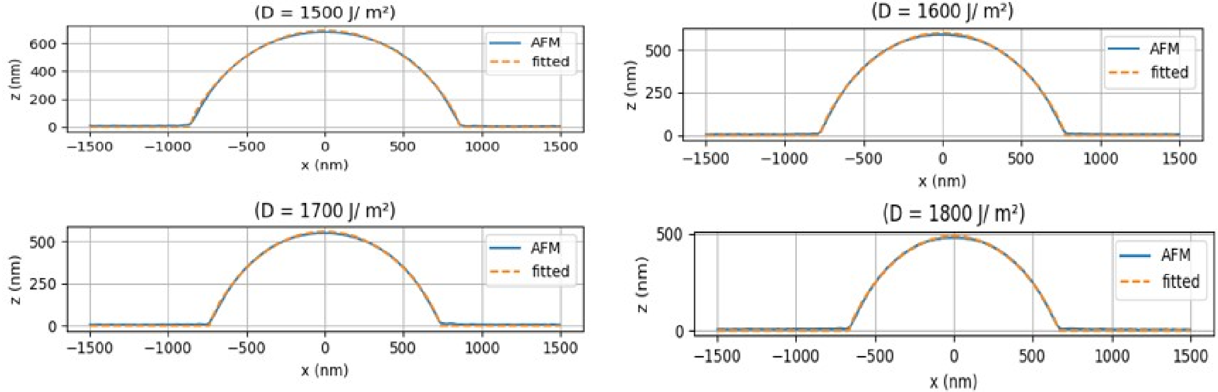


Figure 12. Fitting Geometrical model (Eq. 17) with AFM data for small lens ($R < 800$ nm)

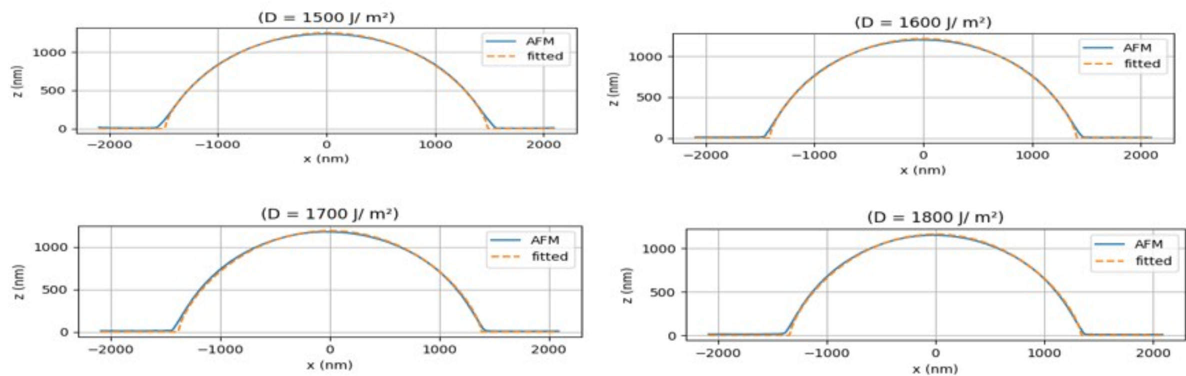


Figure 13. Fitting Geometrical model (Eq. 17)with AFM data for big lens ($R > 1100$ nm)

It is observed that the geometrical model demonstrates a strong fit with AFM data which validates our choice in Eq. (17).

4.2 Forward SEM model calibration

Model validity Theoretically speaking, the model presented the system environment under ideal SEM imaging scenario. To validate the model's applicability in real- world experiments, we performed line profile cut on both the experimental and simulated SEM intensities in the direction of the active quadrant (i.e where $\phi_N = \phi_A$), as shown in Fig 14. The simulated intensity demonstrates symmetry with respect to the origin. However,

the experimental intensity does not exhibit the same behaviour. This can be attributed to various artifacts introduced by the SEM imaging, primarily the shadowing and reemission effects. These effects prevent the direct calculations of geometrical gradients from Eq. (16)

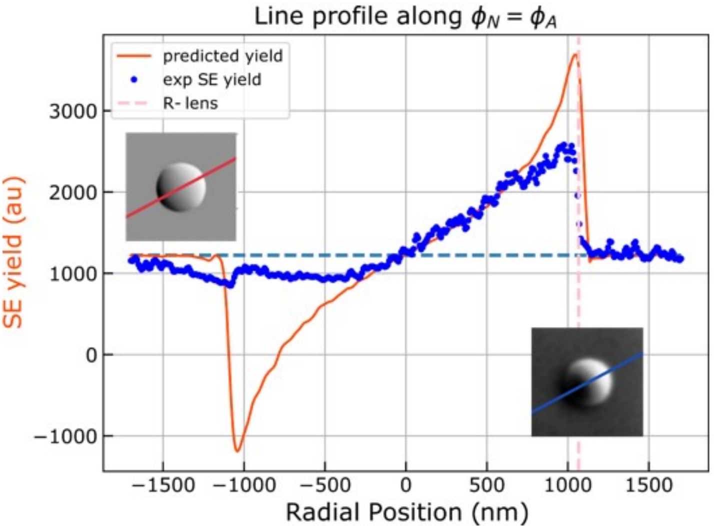


Figure 14. Line profile cut in the direction of the engaged quadrant (in this case A) for simulated and experimental intensities, $R-$ corresponds to the radius of the lens.

Weighting Filter Observing Fig. 14, there is a noticeable agreement between the simulated and experimental SEM intensities in the illuminated region of the image, specifically, the region facing the active detector. To further reinforce this agreement, we have introduced a weighting binary filter defined as:

$$\mathbf{W}_A(\phi_N) = \begin{cases} \mathbf{1} & \text{if } |\phi_N - \phi_A| \leq \Delta\phi \quad \text{and} \quad H > H_{min} \\ \mathbf{0} & \text{otherwise} \end{cases} \tag{24}$$

By applying this filter, we can emphasize the importance of the region facing the detector, while de-emphasizing the remaining regions. H_{min} is specific height threshold to help prevent any undesired artifacts or distortions that may arise from the edges of the microlenses, more details about the weighting filter are given in Appendix B. Fig 15 displays the experimental SEM image on the left and the resulting image after applying the weighting filter to the right. With the obtained weighting filter, we can integrate it into the defined objective functions in Eq. (15) and Eq. (20).

Model Calibration With the proposed method, we successfully calibrated the four quadrants, accounting for their varying gains and offsets (Fig. 7). To test the calibration validity, we normalized the SEM images according to Eq. (9) with respect to the calibration parameter as:

$$\tilde{I}_A^{SEM} = \frac{I_A^{SEM} - \tilde{c_A}}{\tilde{d_A}} \tag{25}$$

On the normalized SEM images ($\tilde{I}_A^{SEM}$), we performed line profile in the direction of the active detector before calibration 16, left) and after calibration (Fig. 16 right). Through the calibration process, we have effectively adjusted the gain and offset, showing that the amplitudes of the normalized intensities from all four quadrants are now equal as it should be bacause of the rotational invariance.

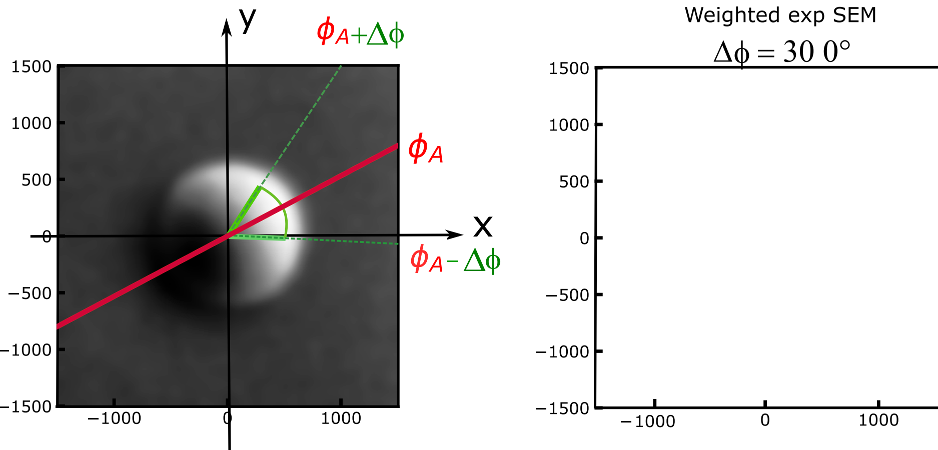


Figure 15. left: experimental SEM image with lines indicating the weighted region, right: experimental SEM image after applying the weighting filter

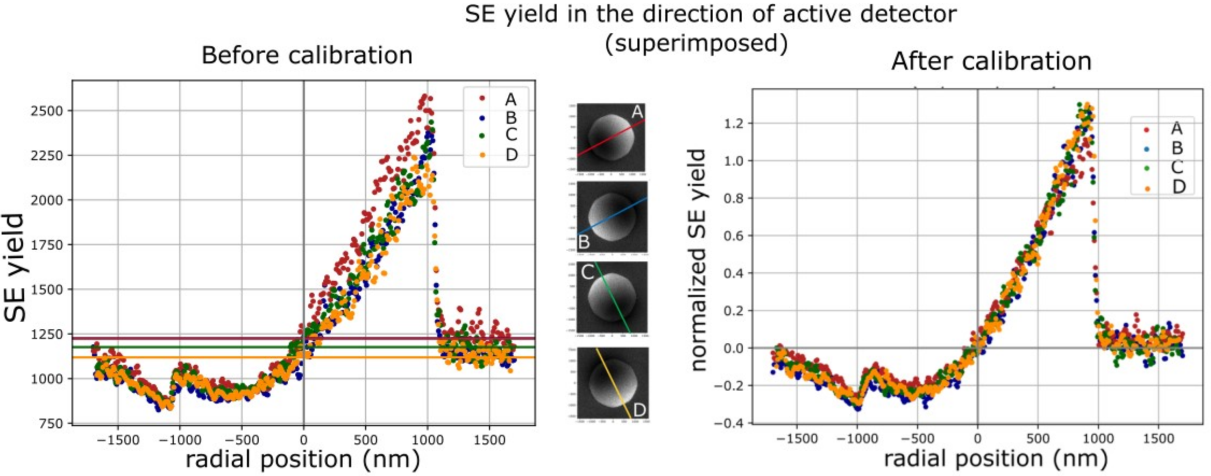


Figure 16. Line profile in the direction of active detectors for the 4 SE images of: left) before calibration, right) after calibration

4.3 Reconstruction Using SEM images

Using the proposed methodology, we were able to accurately reconstruct the microlens profile from SEM images. We achieved this by fixing the radius parameter to the ground truth value obtained from AFM scans, and then optimizing over the remaining two parameters (H and k). Figs. 17 and 18 shows central line profile from reconstructed height map and AFM superimposed for small lenses ($R < 800$ nm) and for big lenses ($R > 1100$ nm) respectively. The reconstruction process achieves a remarkable accuracy of approximately 2% (estimated height $\hat{H}$ vs. the true height $H^{\star}$), ensuring a precise characterization of the microlenses' geometry.

Reconstruction using SEM images provided an accurate, fast, and non-destructive method for characterizing microlenses. The developed model showed promising results and can be applied to a wide range of microlenses with varying dimensions and pitches.

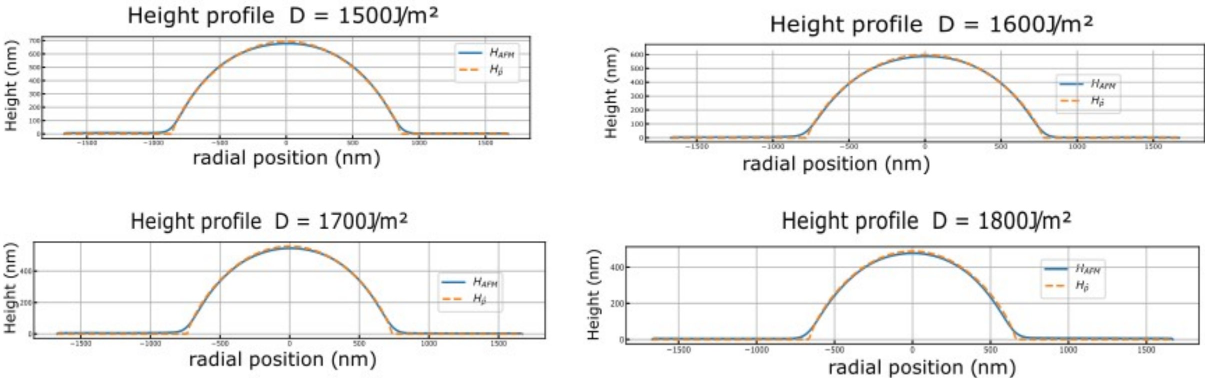


Figure 17. Central line profile from the reconstructed heights and AFM heights for lens (R <800 nm)

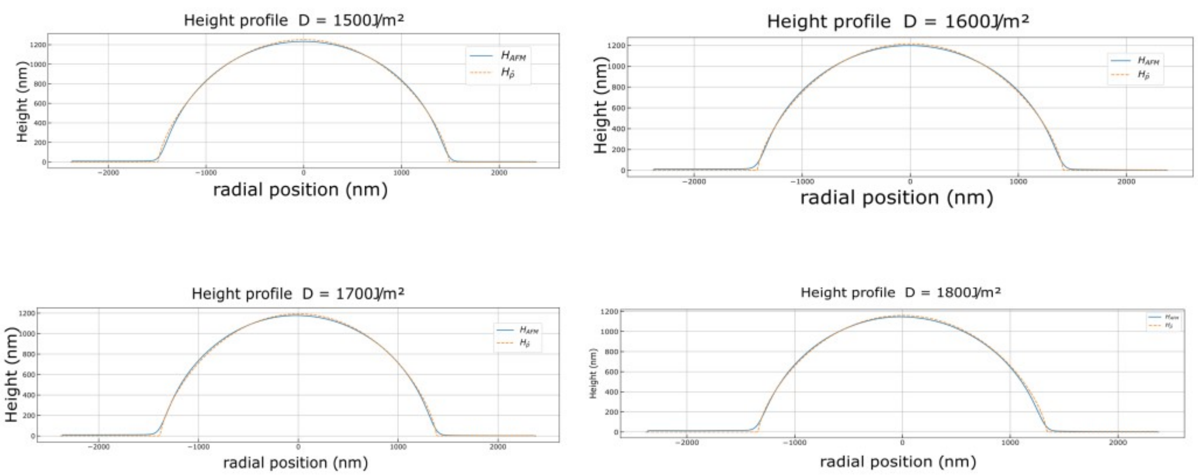


Figure 18. Central line profile from the reconstructed heights and AFM heights for lens (R >1100 nm)

5. CONCLUSIONS AND PERSPECTIVES

In this study, we proposed a methodology for the reconstruction of microlenses using multiple SEM images. First, we fabricated different arrays of microlenses with different radii, heights and pitches using grayscale lithography. Subsequently, we performed four-quadrant SEM imaging to capture the surface of the fabricated microlenses. In parallel, we obtained reference height maps of the imaged microlenses' surface using AFM. We adapted a physically-based model capable of simulating SEM images based on the microlens' geometry. Using the analytical geometry parametrization approach, we were able to analytically parametrize the microlens surface, hence performing parametric reconstruction. By integrating the geometrical and physical models, we successfully calibrated the SEM imaging model and accurately estimated the microlens parameters through the reconstruction process for microlenses of different sizes.

There are several areas where the proposed methodology can be further improved and expanded. First, the model accuracy can be enhanced by considering additional factors such as electron shadowing and re-emission effects. Another potential path of improvement is the estimation of the microlens radius from SEM image contour extraction. Currently, the ground truth radius from AFM data is used as a fixed parameter in the reconstruction process. Furthermore, the proposed methodology can be expanded to accommodate other geometries beyond microlenses. By adapting the model and reconstruction approach to different geometrical structures, we can extend its applicability to a wider range of micro and nano-scale features.

ACKNOWLEDGEMENT

I would like to extend my gratitude to Alexis Royer and Remi Coquand for their invaluable contributions in conducting the experiments and image acquisition, which were integral to the success of this project.

APPENDIX A. FORWARD MODEL INTEGRATION

As mentioned in Sec. 3.1.2, our forward model (Eq. 8) was initially derived by W. Słówko et al [27]. In their architecture, the detectors were assumed to be part of a hemisphere Fig. 19 (b). We obtained the same formulation, as shown in Fig. 19 (a), despite the variation in the experimental setup. In this appendix we will carry out the integration.

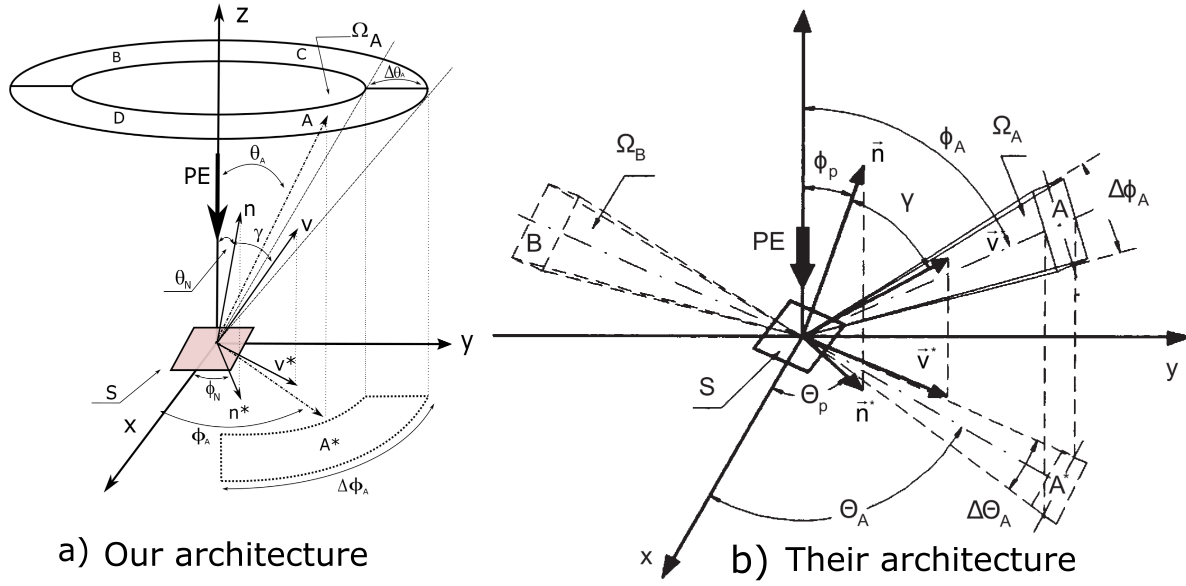


Figure 19. a) The architecture of the of the system setup that had been used in this paper. b) the system where they had obtained Eq. (8), where the detectors are part of a hemisphere

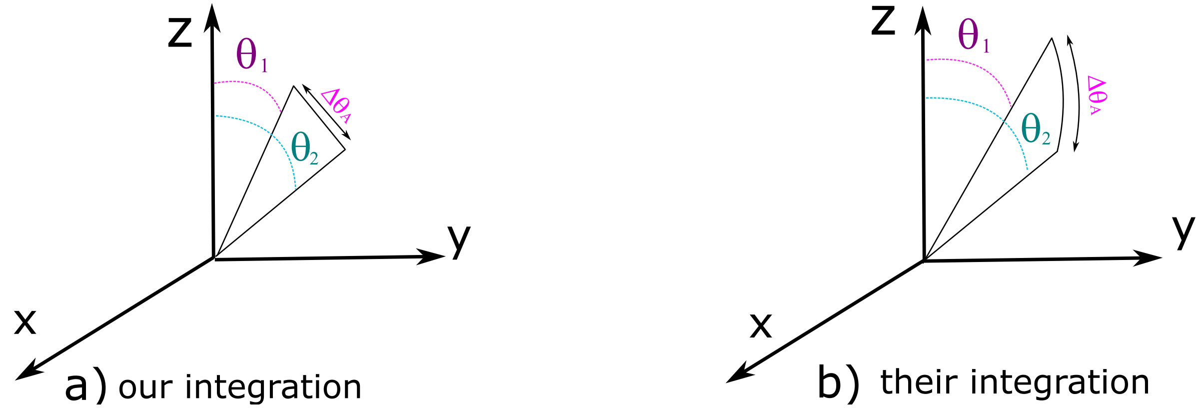


Figure 20. a) Surface's integration o setup that had been used in this paper. b)) Surface's integration where Eq. 8, where the detectors are part of a hemisphere

A.1 Derivation of the forward model

The SE yield generated by the primary electron beam depends on the beam incident angle (θ_N), δ denotes the SE emission coefficient:

$$\delta = \delta_0 R(\theta_N) \tag{26}$$

δ_0 is the secondary yield at $\theta_N = 0$, δ_0 depends on the energy of the primary electron beam. The topographic contrast creation depends mainly on the angular distribution of SE emission.

Following Eq. (3), performing the integral term:

$$\int_{\Omega_A} \cos(\gamma) d\Omega = \int_{\Omega_A} \mathbf{n}^T \cdot \mathbf{v} d\Omega \tag{27}$$

Parametrizing in spherical coordinates:

$$\mathbf{n} = \begin{pmatrix} \sin\theta_N \cos\phi_N \\ \sin\theta_N \sin\phi_N \\ \cos\theta_N \end{pmatrix}, \qquad \mathbf{v} = \begin{pmatrix} \sin\theta \cos\phi \\ \sin\theta \sin\phi \\ \cos\theta \end{pmatrix} \tag{28}$$

And:

$$d\Omega = \sin\theta d\phi d\theta \tag{29}$$

$$\cos(\gamma) = \mathbf{n}^T \cdot \mathbf{v} = (\sin(\theta_N)\cos(\phi_N))(\sin(\theta)\cos(\phi)) + (\sin(\theta_N)\sin(\phi_N))(\sin(\theta)\sin(\phi)) + \cos(\theta_N)(\cos(\theta))$$
$$\cos(\gamma)\sin(\theta) = \sin(\theta_N)\cos(\phi_N)\sin^2(\theta)\cos(\phi) + \sin(\theta_N)\sin(\phi_N)\sin^2(\theta)\sin(\phi) + \frac{1}{2}\cos(\theta_N)\sin(2\theta)$$

The quantity to compute:

$$\begin{aligned}
\int_{\Omega_A} \cos(\gamma) d\Omega &= \int_{\theta_1}^{\theta_2} d\theta \int_{\phi_1}^{\phi_2} d\phi \cos(\gamma)\sin(\theta) \\
&= A\left[\cos\phi_N(\Delta\sin\phi) - \sin\phi_N(\Delta\cos\phi)\right]\sin(\theta_N) + \frac{B}{2}(\Delta\phi)\cos\theta_N \\
&= 2A\sin\frac{\Delta\phi_A}{2}\left[\cos\phi_N\cos\phi_A + \sin\phi_N\sin\phi_A\right]\sin\theta_N + \left[\frac{1}{2}\sin 1\theta_A \sin\Delta\theta(\Delta\phi)\right]\cos\theta_N \\
&= 2A\sin\Delta\phi_A\cos(\phi_N - \phi_A)\sin\theta_N + \left[\frac{1}{2}\sin\theta_A\sin\Delta\theta(\Delta\phi)\right]\cos\theta_N \\
&= 2A\sin\Delta\phi_A\cos(\phi_N - \phi_A)\sin\theta_N + \left[\Omega_A\cos\theta_A\cos\frac{\Delta\theta_A}{2}\right] cos\theta_N \\
&= d_A\cos(\phi_N - \phi_A)\sin\theta_N + c_A\cos\theta_N
\end{aligned}$$

where:

$$\Omega_A = \int_{\phi_1}^{\phi_2} d\phi \int_{\theta_1}^{\theta_2} d\theta \sin\theta = -(\Delta\phi)(\Delta\cos\theta) = 2(\Delta\phi)\sin\theta_A\sin\frac{\Delta\theta_A}{2}$$

Pre- computation

$$\begin{aligned}
A &= \int_{\theta_1}^{\theta_2} d\theta \sin^2\theta = \frac{1}{2}\left(\Delta\theta - \frac{1}{2}\Delta\sin 2\theta\right) = \frac{1}{2}\left(\Delta\theta\cos 2\theta_A\sin\Delta\theta\right) \\
B &= \int_{\theta_1}^{\theta_2} d\theta \sin 2\theta = -\frac{1}{2}\Delta\cos 2\theta = \sin 2\theta_A\sin\Delta\theta
\end{aligned}$$

$$\Delta \sin\xi = \sin a - \sin b = +2\cos\left(\frac{a+b}{2}\right)\sin\left(\frac{a-b}{2}\right) = +2\cos\xi_A \sin\frac{\Delta\xi_A}{2}$$
$$\Delta \cos\xi = \cos a - \cos b = -2\sin\left(\frac{a+b}{2}\right)\sin\left(\frac{a-b}{2}\right) = -2\sin\xi_A \sin\frac{\Delta\xi_A}{2}$$

APPENDIX B. WEIGHTING FILTER

In this appendix, we will address a very important factor about the accuracy of our metrology: the shadowing effect. The shadowing effect refers to the phenomenon where the presence of neighboring microlenses casts shadows on the surface of the microlens of interest, resulting in changes in the observed SEM signal intensity. Superimposing Fig. 14 with θ_N as shown in Fig. 21 left), we observe that at $\theta_N = 45°$, there are two different SE values one high value ~ 2400 and another low value ~ 1100. This goes back to the shadowing effect. In Fig. 21 (right) we see a schematic demonstration of the shadowing effect. For the same incident angle (angle between the normal and the electron beam) some SEs are getting detected while others are shadowed.

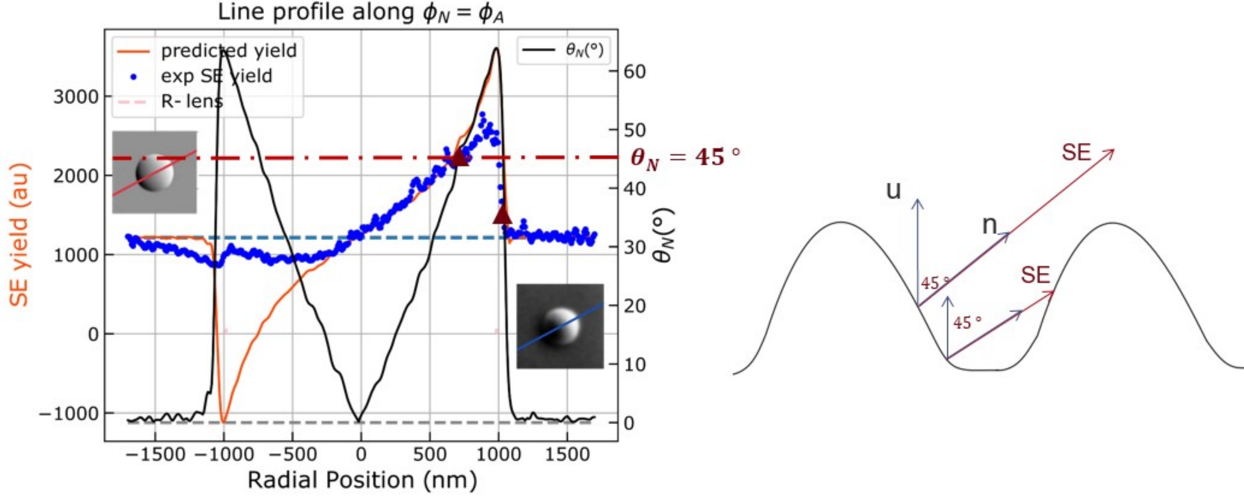


Figure 21. Line profile cut in the direction of the engaged quadrant (in this case A) for: predicted intensity, experimental intensity and the incident angle (θ_N); R- lens -radius.

In our proposed methodology, we address this effect by incorporating a minimum height threshold, denoted as H_{min}, which is used to filter the height of the lens. This minimum height threshold is integrated within the weighting filter introduced earlier in Eq. (24). By applying this threshold, we ensure that only valid height values are considered, effectively mitigating the artifacts and distortions that may arise at the edges of the microlenses.

REFERENCES

[1] Bigas, M., Cabruja, E., Forest, J., and Salvi, J., "Review of CMOS image sensors," *Microelectronics Journal* **37**, 433–451 (may 2006).

[2] Audran, S., *Etude des propriétés et des mécanismes de mise en forme des résines photolithographiques pour une application capteurs d'images CMOS avancés*, PhD thesis, Université Louis Paster (Strasbourg) (2007).

[3] Audran, S., Faure, B., Mortini, B., Aumont, C., Tiron, R., Zinck, C., Sanchez, Y., Fellous, C., Regolini, J., Reynard, J. P., Schlatter, G., and Hadziioannou, G., "Study of dynamical formation and shape of microlenses formed by the reflow method," in [*SPIE Proceedings*], Lin, Q., ed., SPIE (mar 2006).

[4] Audran, S., Vaillant, J., Farys, V., Hirigoyen, F., Huss, E., Mortini, B., Cowache, C., Berthier, L., Mortini, E., Fantuz, J., Arnaud, O., Depoyan, L., Sundermann, F., Baron, C., and Reynard, J.-P., "Grayscale lithography process study applied to zero-gap microlenses for sub-2μ m CMOS image sensors," in [*SPIE Proceedings*], Allen, R. D., ed., SPIE (mar 2010).

[5] Lakcher, A., Bidault, L., Ducoté, J., Mortini, E., Ostrovsky, A., Le-Gratiet, B., Berthier, L., Jamin-Mornet, C., and Besacier, M., "SEM contour based metrology for microlens process studies in CMOS image sensor technologies," in [*Optical Microlithography XXXI*], Kye, J. and Owa, S., eds., SPIE (mar 2018).

[6] Guo, D., Nakagawa, Y., Ariga, H., Suzuki, S., Kinoshita, K., Miyamoto, T., Takakusagi, S., Asakura, K., Otani, S., and Oyama, S., "STM studies on the reconstruction of the ni_2p surface," *Surface Science* **604**, 1347–1352 (aug 2010).

[7] Sharma, P., Chiu, T. S., Biring, S., Chang, T.-F., Chu, C.-H., and Hsieh, Y.-F., "FIB-SEM investigation and auto-metrology of polymer-microlens/CFA arrays of CMOS image sensor," in [*Proceedings of the 21th International Symposium on the Physical and Failure Analysis of Integrated Circuits (IPFA)*], IEEE (jun 2014).

[8] Horn, B. K. P., "Shape from shading; a method for obtaining the shape of a smooth opaque object from one view," (1970).

[9] Lebiedzik, J., "An automatic topographical surface reconstruction in the SEM," *Scanning* **2**(4), 230–237 (1979).

[10] Carlsen, I. C., "Reconstruction of true surface-topographies in scanning electron microscopes using backscattered electrons," *Scanning* **7**(4), 169–177 (1985).

[11] Horn, B. K. and Brooks, M. J., "The variational approach to shape from shading," *Computer Vision, Graphics, and Image Processing* **33**, 174–208 (feb 1986).

[12] Ikeuchi, K. and Horn, B. K., "Numerical shape from shading and occluding boundaries," *Artificial Intelligence* **17**, 141–184 (aug 1981).

[13] Chevalier, P., Quemere, P., Berard-Bergery, S., Henry, J.-B., Beylier, C., and Vaillant, J., "Rigorous model-based mask data preparation algorithm applied to grayscale lithography for the patterning at the micrometer scale," *Journal of Microelectromechanical Systems* **30**, 442–455 (jun 2021).

[14] Chevalier, P., Quéméré, P., Beylier, C., Bérard-Bergery, S., Allouti, N., Paris, M., Farys, V., and Vaillant, J., "Beyond contrast curve approach: a grayscale model applied to sub-5μm patterns," in [*Novel Patterning Technologies for Semiconductors, MEMS/NEMS, and MOEMS 2019*], Panning, E. M. and Sanchez, M. I., eds., SPIE (mar 2019).

[15] Allouti, N., Pierre, C., Bérard-Bergery, S., Rousset, V., Mortini, B. P., Quéméré, P., Tomaso, F., and Coquand, R., "Grayscale lithography process study for sub 5μm microlens patterns," in [*Novel Patterning Technologies for Semiconductors, MEMS/NEMS, and MOEMS 2019*], Panning, E. M. and Sanchez, M. I., eds., SPIE (mar 2019).

[16] van Kessel, L. and Hagen, C., "Nebula: Monte carlo simulator of electron–matter interaction," *SoftwareX* **12**, 100605 (jul 2020).

[17] Villarrubia, J. S., Ritchie, N. W. M., and Lowney, J. R., "Monte carlo modeling of secondary electron imaging in three dimensions," in [*SPIE Proceedings*], Archie, C. N., ed., SPIE (mar 2007).

[18] Lowney, J. R., "Monte carlo simulation of scanning electron microscope signals for lithographic metrology," *Scanning* **18**, 301–306 (dec 2006).

[19] Li, H. M. and Ding, Z. J., "Monte carlo simulation of secondary electron and backscattered electron images in scanning electron microscopy for specimen with complex geometric structure," *Scanning* **27**, 254–267 (dec 2006).

[20] Horn, B. K. P. and Sjoberg, R. W., "Calculating the reflectance map," *Applied Optics* **18**, 1770 (jun 1979).

[21] Prados, E. and Faugeras, O., "Shape from shading," in [*Handbook of Mathematical Models in Computer Vision*], 375–388, Springer-Verlag (2006).

[22] Horn, B. K., "Understanding image intensities," *Artificial Intelligence* **8**, 201–231 (apr 1977).

[23] Valade, C., Hazart, J., Berard-Bergery, S., Sungauer, E., Besacier, M., and Gourgon, C., "Tilted beam sem, 3d metrology for industry," 32 (03 2019).

[24] Labbé, S., Faouzi, T., Belissard, J., and Hazart, J., "Limits of model-based cd-sem metrology," 6 (09 2018).

[25] Reimer, L., Böngeler, R., and Desai, V., "Shape from shading using multiple detector signals in scanning electron microscopy," *Scanning Microscopy* **1**(3), 10 (1987).

[26] Schou, J., "Secondary-electron emission from solids by electron and proton-bombardment," *Scanning Microscopy* (1988).

[27] Czepkowshi, T. and Słówko, W., "Some limitations of surface profile reconstruction in scanning electron microscopy," *Scanning* **18**, 433–446 (2006).

[28] Paluszyński, J. and Slówko, W., "Measurements of the surface microroughness with the scanning electron microscope," *Journal of Microscopy* **233**, 10–17 (jan 2009).

[29] Slówko, W., "Specific features of the miniature ionisation BSE multi-detector unit for 3d imaging in environmental conditions," *Micron* **126**, 102752 (nov 2019).

[30] Neggers, J., Héripré, E., Bonnet, M., Boivin, D., Tanguy, A., Hallais, S., Gaslain, F., Rouesne, E., and Roux, S., "Principal image decomposition for multi-detector backscatter electron topography reconstruction," *Ultramicroscopy* **227**, 113200 (aug 2021).

[31] Beil, W. and Carlsen, I. C., "Surface reconstruction from stereoscopy and "shape from shading" in SEM images," *Machine Vision and Applications* **4**, 271–285 (sep 1991).

[32] Ellison, T. P. and Taylor, C. J., "Calculating the surface topography of integrated circuit wafers from SEM images," *Image and Vision Computing* **9**, 3–9 (feb 1991).

[33] Carli, L., Genta, G., Cantatore, A., Barbato, G., Chiffre, L. D., and Levi, R., "Uncertainty evaluation for three-dimensional scanning electron microscope reconstructions based on the stereo-pair technique," *Measurement Science and Technology* **22**, 035103 (feb 2011).

[34] Schwarzband, I. and Weinberg, Y., "Three dimensional mapping using scanning electron microscope."

Direct patterning of functional materials using nanoimprint lithography

Michael J. Haslinger[a], Sonja Kopp[a], Viktorija Jonaityte[a], Amiya Moharana[a], Helene Außerhuber[a], Michael M. Mühlberger[*]
[a]PROFACTOR GmbH, Im Stadtgut A2, 4407 Steyr-Gleink, Austria

ABSTRACT

Nanoimprint lithography is a large area, high resolution and cost-effective replication technology for micro- and nanostructures. An interesting possibility in the nanoimprint process is the use of materials that remain a functional part of the final component or device. Such an additive approach offers interesting opportunities in terms of novel applications but also cost reduction and could also contribute to sustainability aspects. This paper is aiming at providing a short non-comprehensive overview on the direct patterning of various functional materials by using NIL for optics and life science applications.

Keywords: Nanoimprint Lithography, UV-NIL, functional materials, optics, life sciences

1. INTRODUCTION

Nanoimprint lithography (NIL) is a micro- and nanostructure replication technology, which offers high resolution on large areas and great process and application flexibility [1–3]. In a typical nanoimprint process in a first step a stamp is fabricated. This is made by preparing a negative copy of a master structure (e.g. ref[4,5]), which itself was made by electron beam lithography[6], two-photon polymerization 3D printing or other micro and nanofabrication technologies. The stamp is often made either by a casting process (e.g. with PDMS (poly-dimethyl-siloxane)) or in a UV-NIL process (e.g. with OrmoStamp® see e.g. ref[4]). This approach reduces cost of ownership since the actual master structures is not used for a nanoimprint process and therefore is less in danger to get a defect. Additionally, with this approach, in general a higher flexibility is achieved in the master fabrication as well as in stamp fabrication and nanoimprinting. This aspect is especially relevant in the context of functional materials. We focus on UV-based nanoimprinting, where UV-curable materials, which are typically liquid at room-temperature are coated on a substrate and patterned using a UV-transparent stamp. To achieve this the stamp is brought into contact with the coated substrate. The liquid material fills the micro- and nanocavities in the stamp due to a capillary filling process. When this is done, the material is cured by UV-radiation (through the UV-transparent stamp). Finally, the stamp can be removed and reused for the next imprinting step. The UV-curing material can then can remain as a functional component in the device to be manufactured. Such a material is a functional material in the sense of this contribution. We call such a process also an additive NIL process. Of course, the nanoimprint material can also be used as an etching mask or in a lift-off process[7], in a subtractive NIL process.

2. OPPORTUNITIES, SUSTAINABILITY, CHALLENGES

The approach briefly described above offers several opportunities as far as applications and process design are concerned. Some examples will be given below. Additionally, it also has aspects in terms of sustainability. Considering sustainability, firstly, the nanoimprint process itself can be performed in a sustainable way. This is relevant for energy consumption and material consumption. Tools for optical lithography like DUV and EVU lithography often consume significantly more energy than nanoimprint tools [8],[9]. Considering a growing market for semiconductor applications and the need to shift to fully renewable energy supply, reducing the energy consumption is important. The use of functional materials adds to this aspect, since a reduction of the number of process steps is possible. Furthermore, nanoimprinting can be performed in an additive way using functional materials, meaning that the patterned material does not serve as an etch mask, but remains an integral component of the final device. If such a process is possible for a certain application, it can reduce the number

*michael.muehlberger@profactor.at; phone +43 7252 885 253; fax +43 7252 885 101; www.profactor.at

38th European Mask and Lithography Conference (EMLC 2023), edited by
Uwe F.W. Behringer, Jo Finders, Proc. of SPIE Vol. 12802, 128020N
© 2023 SPIE · 0277-786X · doi: 10.1117/12.2674052

of processing steps and related energy and material consumption. Also cost and time savings could potentially be achieved with such processes. Additionally, inkjet printing can be used as a material deposition technology for NIL[10, 11] (see also figure 2), which produces less waste than typical spin coating processes and even provides additional opportunities. Secondly, the versatility of the nanoimprint process can contribute to many applications with sustainability aspects. In this context, often the structures are bio-inspired[3]. Friction reduction is an important topic. Nanoimprint patterning was used to replicate the reduce the friction on ceramic surfaces (see figure 2) [12, 6]. For airplanes sharkskin-inspired structures are investigated [13] and for ships surfaces are investigated that reduce the adhesion of marine plants and animals, which, when attached to a ship hull, increase the fuel consumption of the ship significantly [14]. In photovoltaic applications anti-reflective moth-eye structures [15] and self-cleaning structures inspired by the surface of rose petals[16], both made by nanoimprinting, can be of interest. This can also be relevant for lenses [17]. Other examples where nanoimprinting plays a role can be found in the life sciences and lighting field. In the life sciences the use of in vitro tests to replace in vivo test with lab animals often is desired, both to reduce costs and animal suffering. Here, the capability of nanoimprinting to directly pattern biocompatible materials with complex structures can be of use (e.g. [18, 19]). The possibility to pattern micro-optical elements for indoor and outdoor lighting using nanoimprinting can help to reduce light pollution, which in turn benefits nocturnal animals and humans alike.

Not only but also and especially for the use of functional materials in additive nanoimprint processes the material properties are critical. The materials must fulfil the requirements for the nanoimprint process and has to fulfil the requirements of the final application, in terms of pattern fidelity, mechanical and optical properties, temperature and humidity requirements to biocompatibility, autofluorescence aspects, to name a few. In the following, a few examples will be given.

3. APPLICATION EXAMPLES

3.1 Optics

The refractive index plays an essential role in the functionality and efficiency of devices in the field of optics. It determines the refraction, reflection, or in case of waveguides, the quality of light guiding. NIL offers the possibility of a fast-direct patterning of materials with different refractive indices. UV-curable nanoimprint materials come in a wide range of refractive indices starting around 1,3 (e.g. 1,34 for MD700[20] and going up to 1.9 (e.g. IOC-133[21]). In addition, some of these materials are now also available as inkjet versions (e.g. IPO-912). Table 1 below gives a few examples.

Table 1. Examples for nanoimprint materials with different refractive indices

Material	Refractive Index (@ 589nm)	Curing mechanism
IOC-133 (Inkron)	1.8-1.9x [21]	UV curing
OrmoClear (micro resist GmbH)	1,56 [22]	UV curing
MD700 (solvay solexis)	1,34 [20]	UV curing
IOC-560C (Inkron)	1,1x [21]	thermal

In our first example a multilayer multimaterial optical device has been fabricated. Figure 1 shows the result of preparing two nanoimprinted layers with different refractive indices on top of each other. The layer sequence is sketched in the inset. First a low refractive index material was patterned with a hexagonal pillar array, followed by a high refractive index material of the same pattern. The layer stack was protected with an unpatterned layer of the low refractive index material. The substrate that was used was glass. Although this special layer sequence only serves demonstrational purposes it shows the potential of preparing several nanoimprinted layers on top of each other. For nanoimprinting of optical components alignment in the range of 100nm has been shown[23–25]. For the sample in Figure 1 the high index material was just droplet dispensed, with the hexagon-shape developing due to the underlying micropattern, which consisted of a hexagonal array of pillars. Figure 2 shows an example, where the high refractive index material was inkjet printed to obtain the pattern of the butterfly in this case.

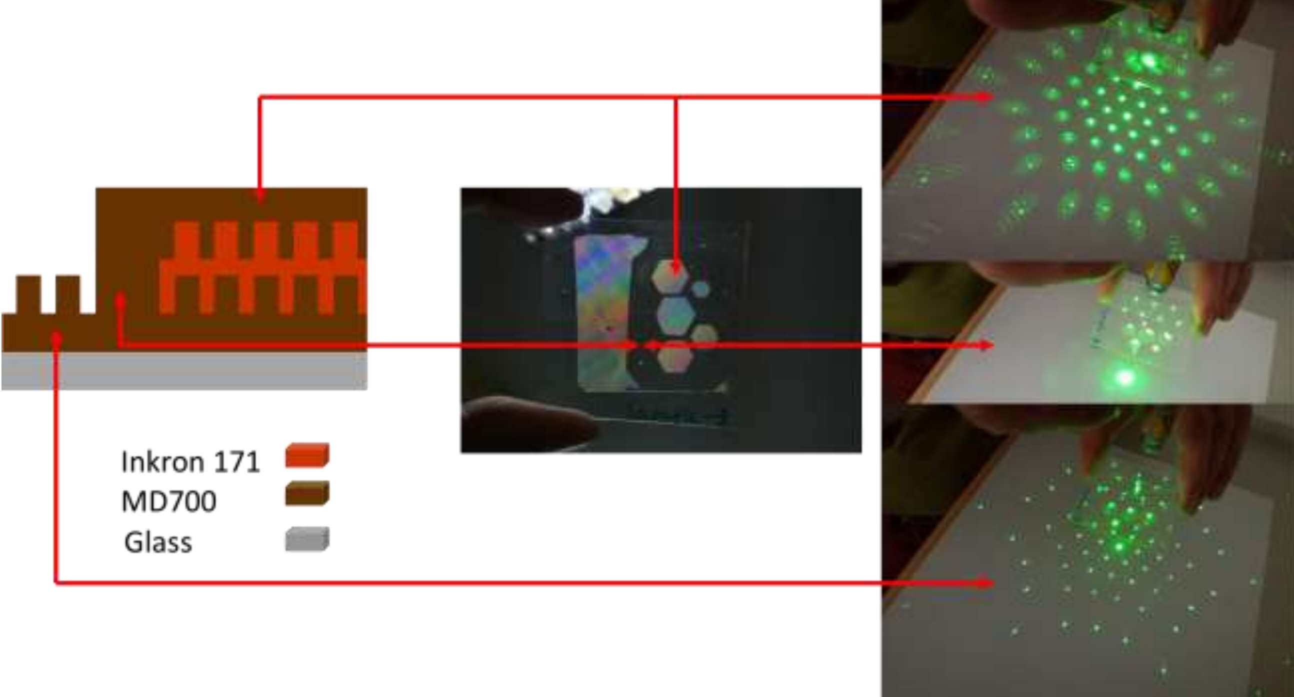


Figure 1 example for a multilayer multimaterial nanoimprinted pattern. On a glass substrate a low refractive index layer was nanoimprinted (MD700), followed by a high refractive index layer (Inkron 171). The layer stack was covered by a low refractive index protection layer, which was unpatterned. In the left the schematic layer stack is show, in the center a photograph of the sample, to the right the diffraction patterns resulting in the different regions of the sample. From top to bottom: all three layers, only unpatterned low refractive index layer, only the first patterned layer.

Figure 2: sample according to the layer stack in figure 1, but with inkjet deposited high refractive index material.

Another example in the field of optics is the fabrication of biomimetic anti-reflection moth-eye structure coatings as shown in Figure 3. Biomimetic anti-reflection moth-eye[26,27] structures are known to provide a gradual change in refractive index coming from the refractive index of air changing to the refractive index of the moth-eye (imprint) material. This leads to a significantly reduced reflection caused by the moth-eye structures. In our example, the nanoimprinting was performed on a curved lens-like surface. The ability to pattern the surfaces of non-flat objects is an additional interesting feature of nanoimprinting[28]. Figure 3a shows a photograph of a watchglass patterned on both sides (lower part of the substrate) and an unpatterned area (top part of the substrate). The detailed process is described in Haslinger et al.[17] .Figure 3b shows an atomic force microscope image of the moth-eye pattern. Figure 3c shows transmission measurements through a sample

patterned on both sides with two different materials. As can be seen, different materials show different performances, but both result in a significant increase in transmission. The reference transmission without coating and nanostructures is around 91.8%. The material with the better performance has a refractive index close to the one of the substrates. Depending on the imprint material, transmission values of up to 98.2% for a both side-imprinted lens could be achieved.

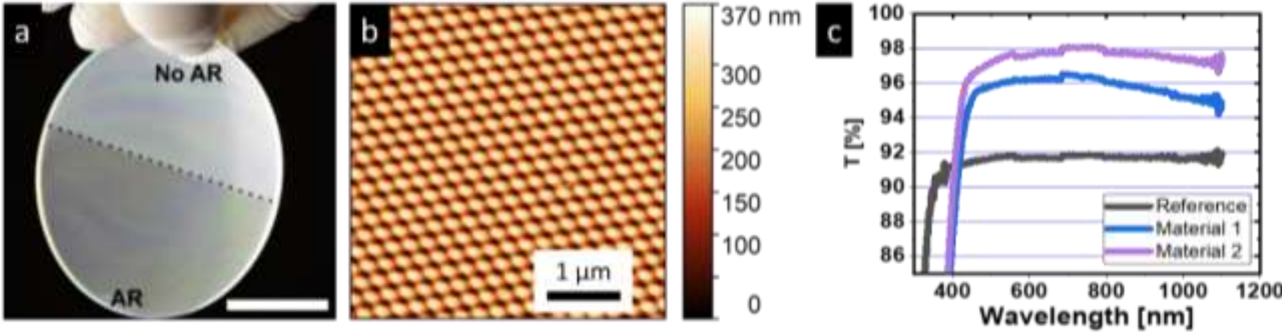


Figure 3 Photograph, AFM image and transmission measurement of an imprinted lens. a) The photograph shows a 70 mm diameter lens with one half-side imprinted with AR moth-eyes on front and back surface. Scale bar is 1.5cm. b) AFM image of imprinted moth-eyes. c) Transmission measurements of two samples with AR moth-eye structures imprinted with two different coatings. [17]

As demonstrated by the moth-eye structure, the apparent refractive index can not only be a function of the material, but also of the nanostructure. Another example found in nature are the structural colors found in many species, e.g., in butterflies (e.g. ref[3] and references therein). Figure 4 below gives another example where nanoimprinting into a functional material was used to realize a biomimetic structure inspired by the Morpho butterfly, which has metallic blue wings, resulting from a complex arrangement of nanostructure, providing a Bragg type of a reflector, which reflects the blue color into all directions[29,30]. The details on the imprinting process and master fabrication can be found in references [5,31]. This is an example where the requirements from the nanoimprint process side were especially high, since the removal of the stamp on the nanoimprint process is especially difficult for these undercut structures. A material was formulated to exhibit the right mechanical properties to allow this. Figure 4 shows from left to right first the master structure, which has a T-shaped cross section. The center image in figure 4 shows the stamp, which was fabricated from the master structure, without destroying it, and finally on the right the imprint is shown, exhibiting the same polarity as the master structure. It can be noted that also the slanted angle of the T cross bar has been faithfully replicated in the nanoimprinting process.

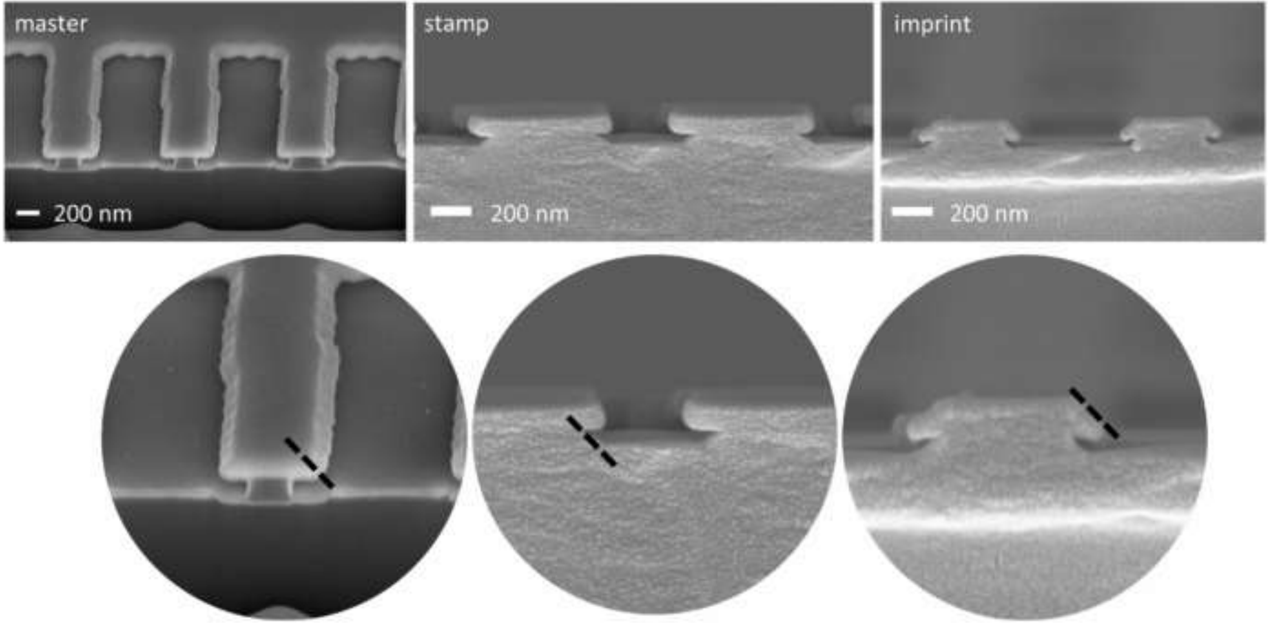


Figure 4 SEM cross-sectional images of the sample sequence from the master structure[31] (left) over the stamp in NILCure® JR5 (center) and the imprinted structures in NILcure® JR5 (right). The second row shows close-ups from the images in the upper row [5]

A final example relating to optics is a more traditional structure relating to microlens arrays. Figure 5a shows a master structure fabricated using two photon polymerization 3D printing. This master structure was fabricated at the Medical University of Vienna using a two photon polymerisation 3D printer (upnano, Vienna). The size of the patterned area is approx. 4x6 mm². From this master a PDMS (Sylgard 184) copy was made and used in a step and repeat nanoimprint process in a self-built step and repeat nanoimprint tool (figure 5b) resulting in a sample as shown in figure 5c. It consists of 96 imprints and can either be a final product or can be used in a further step to fabricate a stamp for full area printing of the 96 optical elements in a single nanoimprint step, e.g. in a roll-to-plate nanoimprint process. For such optical elements there are often demanding requirements if they are to be used e.g. in automotive applications, which go much beyond just only the refractive index. The imprints in this case were fabricated using OrmoClear® (micro resist technology).

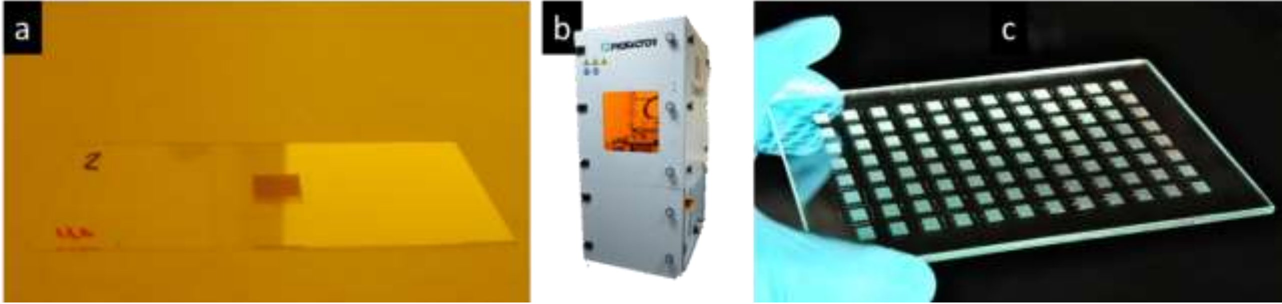


Figure 5 a) Photograph of the master structure on a 75x25mm² glass slide. b) self-built step and repeat nanoimprint tool for dedicated to microstructure replication. c) final imprinted structure of 12x8 imprints.

3.2 Life Sciences

For life science applications biocompatibility of the material is often one of the main requirements. OrmoComp ® is one of the materials that is often used. The example given in figure 6 uses OrmoComp® as imprint material. Here, we were aiming at providing a substrate for the growth of cardiomyocyte fibers. Details on the background, the imprinting and the first cell growth results can be found elsewhere[32]. Here it should be noted that several criteria had to be met by the

imprinting material: i) biocompatibility, i.e. the cells have to survive, if placed on such a surface, ii) good imprinting properties (also for roll-to-plate nanoimprint processes) and iii) excellent optical properties for observation of the cells in optical inverted microscopes. Nanoimprinting provided a fast and cost-efficient way to fabricate the cell growth substrates on various substrates in different configurations[32].

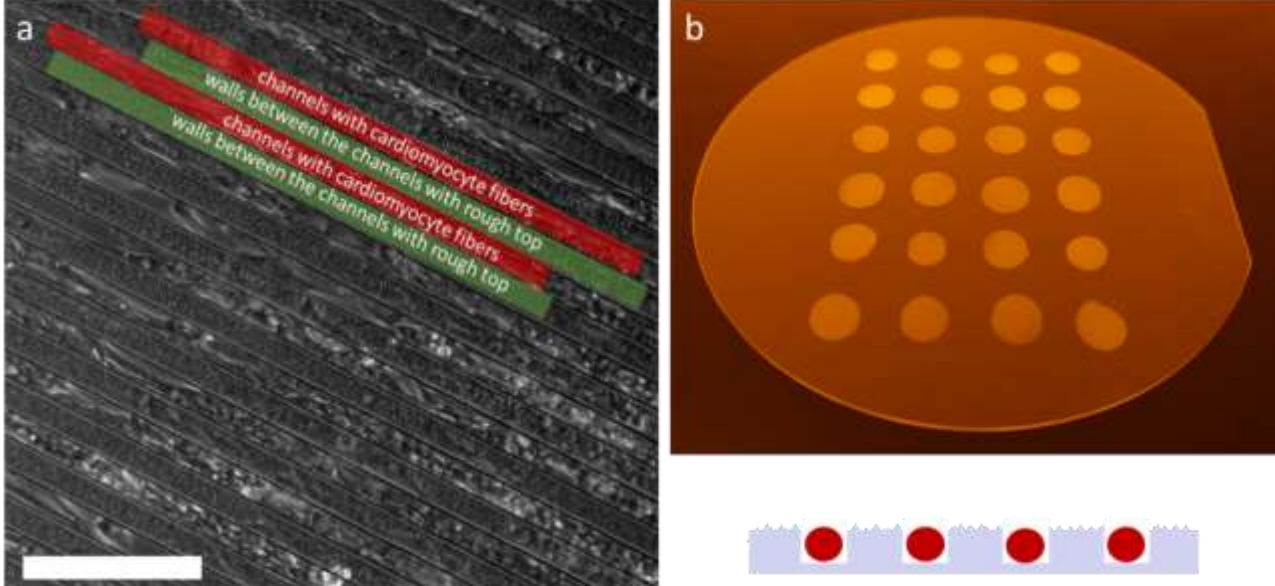


Figure 6 Cardiomyocytes forming fibers inside nanoimprinted channels: The cells tend to grow inside the channels between the walls. a: optical micrograph with some of the wall and channels areas highlighted. b: 24 imprints made on a 150 mm glass wafer to act as an intermediate large area master for printing plate fabrication for roll-to-plate nanoimprinting. The inset to the lower right sketches the arrangement of the cardiomyocyte fibers located insid the trenches. [32]

OrmoComp® was also used to nanoimprint the nanoneedles shown in figure 7. Here[33] it was the aim to mechanically improve the stability and electrically functionalize the nanoneedles by coating them after imprinting with first a layer of Au followed by SiN to provide electrical insulation. Figure 7a shows the imprinted needle. The master in this case was a nanoneedle fabricated by FEBID as detailed in reference[33]. Figure 7b shows the final functionalized needle. The top of the needle was opened to expose the underlying Au layer. In figure 7c the schematic cross section of the functionalized needle is shown. For this application the mechanical stability of the cured material was essential in addition to the imprint properties and the biocompatibility (there are potentially uncovered areas of the imprint material in contact with the cells).

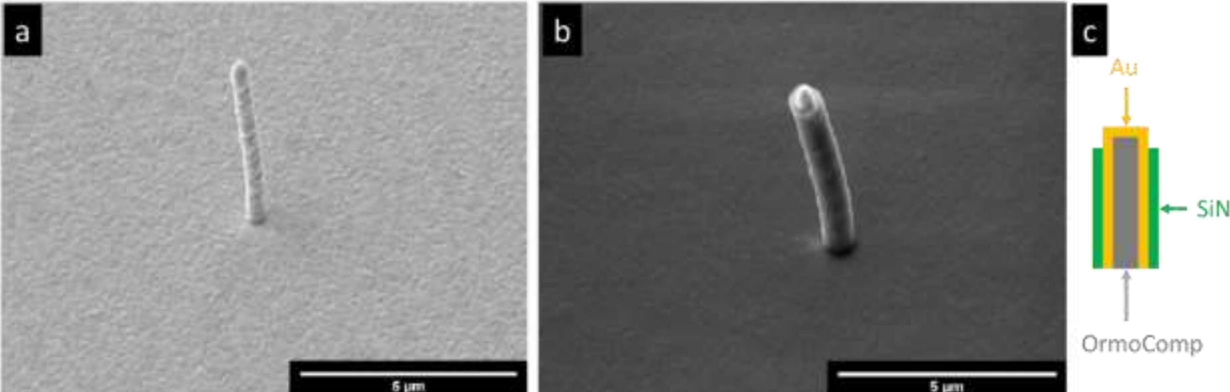


Figure 7 a: SEM image of a nanoimprinted nanoneedle. The imprint material is OrmoComp®. b: imprinted nanoneedle, coated with Au and SiN, after tip opening. c: schematic cross section of the functionalized needle in b.

Microneedles can be interesting for long term electrodes[34] (ecg eeg), trans dermal drug delivery or vaccination purposes[35]. Using nanoimprinting, it is also possible to replicate such structures. Figure 8 shows such examples. On the left in figure 8a a small array of microneedles is shown. The master was 3D printed and used for stamp fabrication. The imprint material was OrmoComp® in this case. The high replication fidelity of the nanoimprint process makes the fabrication of sharp microneedles feasible, which is beneficial in the application.[35,36]

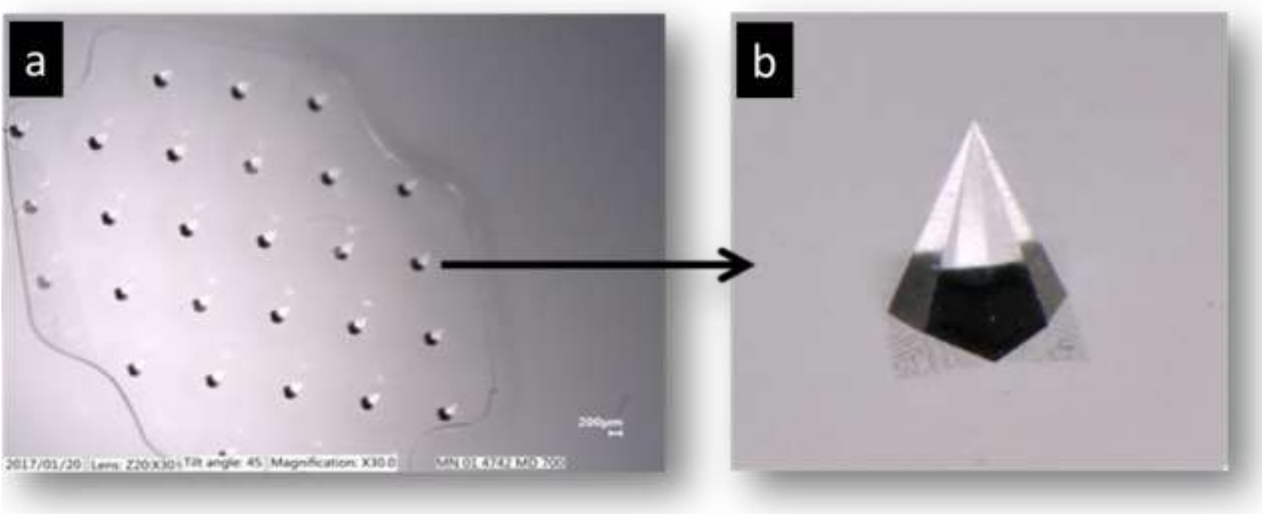


Figure 8 a: nanoimprinted microneedle array. The material is OrmoComp®. b: close-up of one of the microneedles of the array shown in a.

3.3 Other examples from literatures

Other applications where nanoimprinting was used to structure functional materials include organic photovoltaics[37], a dual damascene process[38], riblet structures for wind turbines[13], anti-fouling structures for ships[14], microfluidics[39], waveguides[40], and many more.

4. CONCLUSIONS AND OUTLOOK

The fact that nanoimprinting can directly pattern a wide range of functional materials enables a broad range of applications and makes nanoimprinting a very versatile micro- and nanofabrication technology. We showed examples from the fields of optics and life sciences. Using the right materials, nanoimprinting has the potential to be a sustainable, cost-efficient micro- and nanopatterning technology for a large number of applications. We therefore hope that many novel materials will become available in the future to support additive nanoimprint processes.

ACKNOWLEDGEMENTS

The authors like to acknowledge funding from various projects: MINALEM (moth-eye structures, eurostars funding), ANIIPF and multilink (high and low index multimaterial imprints, bmvit funding), M3Dres (microoptics example, FFG funding), rollerNIL (Morpho butterfly structures, FFG funding), NEAT (cardiomyocytes and nanoneedles, FFG funding), Microneedle (microneedles, FFG funding). We acknowledge the Medical University Vienna, Department of Biomedical Engineering for fabricating the 2PP master structure for the micro-optics step&repeat NIL shown here.

REFERENCES

[1] Schift, H., “Nanoimprint lithography: An old story in modern times? A review,” Journal of Vacuum Science & Technology B: Microelectronics and Nanometer Structures **26**(2), 458 (2008).

[2] Schift, H., “Nanoimprint lithography: 2D or not 2D? A review,” Appl. Phys. A **121**(2), 415–435 (2015).

[3] Muehlberger, M., “Nanoimprinting of Biomimetic Nanostructures,” 1, Nanomanufacturing **2**(1), 17–40 (2022).

[4] Mühlberger, M., Bergmair, I., Klukowska, A., Kolander, A., Leichtfried, H., Platzgummer, E., Loeschner, H., Ebm, Ch., Grützner, G. and Schöftner, R., “UV-NIL with working stamps made from Ormostamp,” Microelectronic Engineering **86**(4–6), 691–693 (2009).

[5] Muehlberger, M., Ruttloff, S., Nees, D., Moharana, A., Belegratis, M. R., Taus, P., Kopp, S., Wanzenboeck, H. D., Prinz, A. and Fechtig, D., “Nanoimprint Replication of Biomimetic, Multilevel Undercut Nanostructures,” Nanomaterials **11**(4), 1051 (2021).

[6] Kirchner, R., Guzenko, V. A., Rohn, M., Sonntag, E., Mühlberger, M., Bergmair, I. and Schift, H., “Bio-inspired 3D funnel structures made by grayscale electron-beam patterning and selective topography equilibration,” Microelectronic Engineering **141**, 107–111 (2015).

[7] Haslinger, M., Mitteramskogler, T., Shoshi, A., Schotter, J., Schrittwieser, S., Mühlberger, M. and Brückl, H., “UV-nil based fabrication of plasmon-magnetic nanoparticles for biomolecular sensing,” Plasmonics: Design, Materials, Fabrication, Characterization, and Applications XVI **10722**, 107222O, International Society for Optics and Photonics (2018).

[8] “NIL and the Environment – Canon Nanotechnologies.”, <http://cnt.canon.com/352-2/> (16 February 2022).

[9] Nakasugi, T., “The current status of nano-imprint lithography and its future outlook toward carbon neutrality by 2050,” Photomask Japan 2022: XXVIII Symposium on Photomask and Next-Generation Lithography Mask Technology **12325**, 22–31, SPIE (2022).

[10] Longsine, W., Traub, M. C. and Truskett, V. N., “Inkjet for Nanoimprint Lithography,” [Handbook of Industrial Inkjet Printing], John Wiley & Sons, Ltd, 851–868 (2017).

[11] Sreenivasan, S. V., “Nanoimprint lithography steppers for volume fabrication of leading-edge semiconductor integrated circuits,” Microsystems & Nanoengineering **3**(1) (2017).

[12] Mühlberger, M., Rohn, M., Danzberger, J., Sonntag, E., Rank, A., Schumm, L., Kirchner, R., Forsich, C., Gorb, S., Einwögerer, B., Trappl, E., Heim, D., Schift, H. and Bergmair, I., “UV-NIL fabricated bio-inspired inlays for injection molding to influence the friction behavior of ceramic surfaces,” Microelectronic Engineering **141**, 140–144 (2015).

[13] Leitl, P. A., Stenzel, V., Flanschger, A., Kordy, H., Feichtinger, C., Kowalik, Y., Schreck, S. and Stübing, D., “Riblet Surfaces for Improvement of Efficiency of Wind Turbines,” [AIAA Scitech 2020 Forum], American Institute of Aeronautics and Astronautics (2020).

[14] Atthi, N., Dielen, M., Sripumkhai, W., Pattamang, P., Meananeatra, R., Saengdee, P., Thongsook, O., Ranron, N., Pankong, K., Uahchinkul, W., Supadech, J., Klunngien, N., Jeamsaksiri, W., Veldhuizen, P. and ter Meulen, J. M., “Fabrication of High Aspect Ratio Micro-Structures with Superhydrophobic and Oleophobic Properties by Using Large-Area Roll-to-Plate Nanoimprint Lithography,” 2, Nanomaterials **11**(2), 339 (2021).

[15] Deniz, H., Khudiyev, T., Buyukserin, F. and Bayindir, M., “Room temperature large-area nanoimprinting for broadband biomimetic antireflection surfaces,” Appl. Phys. Lett. **99**(18), 183107 (2011).

[16] Fritz, B., Hünig, R., Guttmann, M., Schneider, M., Reza, K. M. S., Salomon, O., Jackson, P., Powalla, M., Lemmer, U. and Gomard, G., “Upscaling the fabrication routine of bioreplicated rose petal light harvesting layers for photovoltaic modules,” Solar Energy **201**, 666–673 (2020).

[17] Haslinger, M., Moharana, A. R. and Mühlberger, M., “Antireflective moth-eye structures on curved surfaces fabricated by nanoimprint lithography,” 35th European Mask and Lithography Conference (EMLC 2019), U. F. Behringer and J. Finders, Eds., 31, SPIE, Dresden, Germany (2019).

[18] Elsayed, M. and Merkel, O. M., “Nanoimprinting of topographical and 3D cell culture scaffolds,” Nanomedicine **9**(2), 349–366 (2014).

[19] Nowduri, B., Schulte, S., Decker, D., Schäfer, K.-H. and Saumer, M., “Biomimetic Nanostructures Fabricated by Nanoimprint Lithography for Improved Cell-Coupling,” Advanced Functional Materials **30**(45), 2004227 (2020).

[20] “MD700.pdf.”, MD700 product data sheet, <https://www.acota.co.uk/wp-content/uploads/2018/11/MD700.pdf> (21 February 2023).

[21] “Optical coatings.”, Inkron, <https://inkron.com/optical-coatings/> (21 February 2023).

[22] “Resists & Photopolymers – Microresist.”, <https://www.microresist.de/en/products/> (21 February 2023).

[23] Mühlberger, M., Bergmair, I., Schwinger, W., Gmainer, M., Schöftner, R., Glinsner, T., Hasenfuß, Ch., Hingerl, K., Vogler, M., Schmidt, H. and Kley, E. B., “A Moiré method for high accuracy alignment in nanoimprint lithography,” Microelectronic Engineering **84**(5–8), 925–927 (2007).
[24] Mühlberger, M., Schwinger, W., Gmainer, M., Schoftner, R., Glinsner, T., Hasenfu\s s, C., Hingerl, K., Schmidt, H. and Kley, E., “High Precision Alignment in Multi-layer Nanoimprint Lithography,” Physics of Semiconductors; Part B **893**, 1495–1496 (2007).
[25] Kimura, A., Takabayashi, Y., Iwanaga, T., Hiura, M., Sakai, K., Morohoshi, H., Asano, T., Hayashi, T. and Komaki, T., “Nanoimprint system alignment and overlay improvement for high volume semiconductor manufacturing,” Novel Patterning Technologies for Semiconductors, MEMS/NEMS and MOEMS 2020, E. M. Panning and M. I. Sanchez, Eds., 9, SPIE, San Jose, United States (2020).
[26] Chen, Q., Hubbard, G., Shields, P. A., Liu, C., Allsopp, D. W. E., Wang, W. N. and Abbott, S., “Broadband moth-eye antireflection coatings fabricated by low-cost nanoimprinting,” Applied Physics Letters **94**(26), 263118 (2009).
[27] Sun, J., Wang, X., Wu, J., Jiang, C., Shen, J., Cooper, M. A., Zheng, X., Liu, Y., Yang, Z. and Wu, D., “Biomimetic Moth-eye Nanofabrication: Enhanced Antireflection with Superior Self-cleaning Characteristic,” 1, Scientific Reports **8**(1), 5438 (2018).
[28] Köpplmayr, T., Häusler, L., Bergmair, I. and Mühlberger, M., “Nanoimprint Lithography on curved surfaces prepared by fused deposition modelling,” Surface Topography: Metrology and Properties **3**(2), 024003 (2015).
[29] Saito, A., “Material design and structural color inspired by biomimetic approach,” Science and Technology of Advanced Materials **12**(6), 064709 (2011).
[30] Tabata, H., Kumazawa, K., Funakawa, M., Takimoto, J. and Akimoto, M., “Microstructures and Optical Properties of Scales of Butterfly Wings,” OPT REV **3**(2), 139–145 (1996).
[31] Taus, P., Prinz, A., Wanzenboeck, H. D., Schuller, P., Tsenov, A., Schinnerl, M., Shawrav, M. M., Haslinger, M. and Muehlberger, M., “Mastering of NIL Stamps with Undercut T-Shaped Features from Single Layer to Multilayer Stamps,” Nanomaterials **11**(4), 956 (2021).
[32] Mühlberger, M., Kopp, S., Deyett, A., Pribyl, M., Haslinger, M. J., Siegel, A., Taus, P., Guille, E., Torres Caballero, A., Baltov, B., Netzer, M. A., Prado-López, S., Yde, L., Stensborg, J., Mendjan, S., Hering, S. and Wanzenboeck, H. D., “Nanoimprinted Hierarchical Micro-/Nanostructured Substrates for the Growth of Cardiomyocytes,” Nanomanufacturing submitted.
[33] Haslinger, M. J., Maier, O. S., Pribyl, M., Taus, P., Kopp, S., Wanzenboeck, H. D., Hingerl, K., Muehlberger, M. M. and Guillén, E., “Increasing the Stability of Isolated and Dense High-Aspect-Ratio Nanopillars Fabricated Using UV-Nanoimprint Lithography,” Nanomaterials **13**(9), 1556 (2023).
[34] Wang, R., Jiang, X., Wang, W. and Li, Z., “A microneedle electrode array on flexible substrate for long-term EEG monitoring,” Sensors and Actuators B: Chemical **244**, 750–758 (2017).
[35] O’Mahony, C., Bocchino, A., Haslinger, M. J., Brandstätter, S., Außerhuber, H., Schossleitner, K., Clover, A. J. P. and Fechtig, D., “Piezoelectric inkjet coating of injection moulded, reservoir-tipped microneedle arrays for transdermal delivery,” Journal of Micromechanics and Microengineering **29**(8), 085004 (2019).
[36] Schossleitner, K., O’Mahony, C., Brandstätter, S., Haslinger, M. J., Demuth, S., Fechtig, D. and Petzelbauer, P., “Differences in biocompatibility of microneedles from cyclic olefin polymers with human endothelial and epithelial skin cells,” Journal of Biomedical Materials Research Part A **107**(3), 505–512 (2019).
[37] Kim, M.-S., Kim, J.-S., Cho, J. C., Shtein, M., Kim, J., Guo, L. J. and Kim, J., “Flexible conjugated polymer photovoltaic cells with controlled heterojunctions fabricated using nanoimprint lithography,” Applied Physics Letters **90**(12), 123113 (2007).
[38] Schmid, G. M., Stewart, M. D., Wetzel, J., Palmieri, F., Hao, J., Nishimura, Y., Jen, K., Kim, E. K., Resnick, D. J., Liddle, J. A. and Willson, C. G., “Implementation of an imprint damascene process for interconnect fabrication,” Journal of Vacuum Science & Technology B **24**(3), 1283–1291 (2006).
[39] Lohse, M., Thesen, M. W., Haase, A., Smolka, M., Iceta, N. B., Ayerdi Izquierdo, A., Ramos, I., Salado, C. and Schleunitz, A., “Novel Concept of Micro Patterned Micro Titer Plates Fabricated via UV-NIL for Automated Neuronal Cell Assay Read-Out,” Nanomaterials **11**(4), 902 (2021).
[40] Prajzler, V., Chlupaty, V., Kulha, P., Neruda, M., Kopp, S. and Mühlberger, M., “Optical Polymer Waveguides Fabricated by Roll-to-Plate Nanoimprinting Technique,” Nanomaterials **11**(3), 724 (2021).

The image lab sandbox, pulling image computing in wafer fab metrology environment.

F. Dettoni, B. Le-Gratiet

[1]STMicroelectronics, 850 rue Jean Monnet, 38920 Crolles, France

ABSTRACT

Background:

In a fast-growing wafer fab environment with multiple technologies in development and production, enabling capabilities to investigate new metrology or process control solutions in production condition is key. One type of data is of interest to us: the images. To better exploit images, and for now CDSEM (Critical Dimension Scanning Election Microscope) images, an image processing sandbox environment has been put in place to enable a real time image process in a "shadow" mode with respect to production. This sandbox is used to validate solutions and propose them for full integration.

Aim:

Images are produced in large amounts in a wafer fab. They contain a lot of relevant information that can be extracted thanks to diverse types of software, either internally developed or from external suppliers. Simple demonstration of capabilities is often not enough to evidence the return on investment of a solution. The image lab sandbox is an infrastructure able to compute real time the images generated during production and send the results in a database environment which is connected to our data analytic environment. By doing so engineers can generate trends which are emulating control charts.

Approach:

The sandbox construction has been presented at SPIE 2022 [1] and more details and results are being proposed for publication in JM3 journal. Figure1 describe the main components are image backup infrastructure, ETL (Extract Transform Load) scripts to send image meta data to a database, scripts (trigger and daemon) that are used to launch "on demand" the usage of chosen software based on the image context, a data base system gathering fab measurement meta data and measurement results from software solution being tested, and a data analytic platform which is used to join data from the sand box to the global fab data system and provide data visualization solution.

Results:

The sandbox is today mainly used to support the deployment of contour-based metrology solutions but is aimed to interface with any other software solution. In 2022, the sandbox has processed more than 30 million images for diverse kinds of purposes. Currently three different software are running all dedicated to CDSEM images. CDSEM image quality monitoring, various contour-based metrology recipes, and more advanced testing like in device overlay or hotspot monitoring shown in this paper.

38th European Mask and Lithography Conference (EMLC 2023), edited by
Uwe F.W. Behringer, Jo Finders, Proc. of SPIE Vol. 12802, 128020O
© 2023 SPIE · 0277-786X
doi: 10.1117/12.2675906

Conclusion:

The image Sandbox is now running as a laboratory to evaluate solutions and validate them for production integration. Applications are either for manufacturing control, tool qualification, R&D. Use cases are today multiple and some of them will be presented in this paper.

Keywords:

CDSEM, image blur, sharpness, contour metrology, Hotspot monitoring, scanner matching, MetroSpection

Contacts :

[a] e-mail: florent.dettoni@st.com
[b] e-mail: bertrand.le-gratier@st.com

INTRODUCTION

This paper talks about a fab story related to data access, extraction, and valorization for engineers. As discussed in the problematic part of this paper, a wafer fab like Crolles300mm is a 20-year-old fab running several tens of technologies, essentially More than Moore, based on technology nodes ranging from 120nm down to 18nm. The diversity of metrology equipment, suppliers along with the various generations of equipment that were, and are, entering the fab makes data accessibility a challenge for engineers.

It all started with the desire to provide a universal interface that would give access to some data that are not handled by the fab automation, SPC (Statistical Process Control), APC (Advanced Process Control) systems. Amongst those data, this paper will focus on CD-SEM images which are "unstructured" but rich data sources if engineers are able to use image processing solutions like contour extraction. From this use case, metrology engineers from the wafer fab manage to build a sandbox able to run in real time remote image processing (mean outside the CDSEM) using contour extraction software. This sandbox is a unique environment (that could also be called Image Lab) helping engineers to evaluate image processing solutions in a HVM (High Volume Manufacturing) environment. The motivation being that, if one wants to put in evidence the benefit of deploying such solution, it cannot come from a simple demo on few wafers but from a monitoring of a running process, whether it is in production already or in R&D phase.

This paper will go in 3 parts, the first one related to the problematic that has driven the development of such SandBox, then a description of the Sandbox itself. It will be followed by a few use cases illustrating the role of the SandBox to build proofs of concepts. Finally, some perspectives will be given to go beyond CDSEM image.

1. PROBLEMATIC

Year after year the semiconductor industry is getting more complex. This evolution is led by an increase of the technology's offer through More Moore and More Than Moore approaches, by an increase of the factories size, the complexification of technologies mix in fab, and products yield enhancement requirement. One of the main levers to deal with this challenge

is using more metrology types, techniques, and steps through or in parallel of the products fabrication. This raises the problematic of managing more metrology data, but also and especially handling an important diversity of metrology data types coming from different tool suppliers. In the semiconductor industry the standard method to collect data from metrology tools and aggregate them in host systems, as Advanced Process Control or Statistical Process Control, is based on SECS GEM protocols. Nevertheless, a large amount of data is still captive from metrology machines, off-line measurements are missing and big files (images, spectra, etc...) are not addressed.

Without metrology data, technicians and engineers are blind. Metrology data access is instrumental for technicians and engineers' efficiency and productivity. Furthermore, for several years and with massive adoption of data sciences there is a clear trend leading to the creation of new metrology capabilities outside the metrology tools that provide raw data. Intrinsic metrology tool capabilities are still the main ones in the fab but global fab metrology capability is augmented by data sciences through third party software or in-house developed algorithms. This clearly drives the need for a universal metrology data infrastructure.

Metrology data use is a daily task. Universal metrology data infrastructure addressing all metrology data (in-line, off-line, images, spectra, context …) providing a unique self-service for data shopping, wrangling, and viewing and allowing both basic and complex analysis is a powerful tool for R&D and production. Such capability can be enabled by deployment of basic data infrastructure constituted by 3 parts: automation, storage and Graphical User Interface as shown in Figure 1.

Automation or ETL for Extract Transform Load aims to connect to the tool, download dataset then formatting it before uploading the formatted dataset to the storage part of the architecture. ETL needs dedicated attention because it directly interacts with the production tools and thus any potential impact should be carefully considered when developing the solution. ETL is the most process related part since it will require a deep knowledge of metrology data to efficiently transform the data. As the first part of the architecture, automation should be used as semantic layer to uniform data format between all metrology types, techniques, and supplier to reduce development effort of the next architecture parts. On the other hands because of the data diversity this part will require laborious development and like a Rosetta stone is instrumental to translate all metrology tools data language into a universal one.

The second part of the architecture, the storage, is key since it will link metrology data to the end users. It will be the data source for the end users and can be file system storage or database. Performance and cost are the two drivers of the final choice, but Graphical User Interface (GUI) should also be considered since some can only use database as data source. In most of the case, mixing database and file system storage is the solution using the indexation power of the database to enable appropriate criteria research in GUI while minimizing the database size by using file system to store main part of the data volumetry as for example images or spectra. In this configuration other aspects should also be taken into consideration such as network performance, file system type, volumetry requirement, etc... Database scheme(s) table(s), keys and indexes configuration will drive the performances of the system, enable analysis capabilities through the GUI and finalize semantic layer application. One of the main requirements of this architecture is to provide data access to metrology production data in real time to enable analysis capabilities as production control tool. In this context Cloud solutions are not considered.

Finally, the last part, the Graphical User Interface or GUI will be the interactive part between the end users and the metrology data. GUI enables analysis capabilities such as cartography display, trends extraction, correlation, etc... This part needs to be as agile as possible to fit the fab needs seamlessly and rapidly.

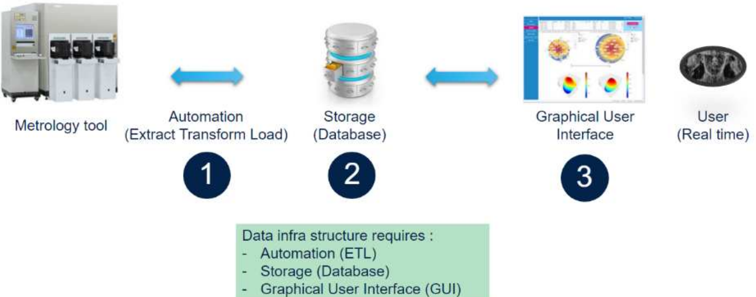


Figure 1: Basic metrology data infrastructure

This basic and universal metrology data infrastructure provides R&D and production teams with a powerful tool to help them in their daily tasks. On besides of this it can be used as an enabler for most complex metrology data architecture as discussed in the next part.

2. A FAB IMAGE LAB

As presented in the previous part of this paper, having such basic metrology data infrastructure is instrumental for the production line but can also be a powerful enabler for most complex metrology data infrastructure. Being able to handle in real time CD-SEM data, and images, opens new opportunities. As illustrated in Figure 2, a trigger part can be used to feed a Virtual Machine hosting images analysis capability. This trigger part can be configured to "listen" users defined context as a technology, a product, a step used for several mask sets or other contexts of interest and then apply to the context related images a dedicated image processing. Image analysis capability is hosted on a Virtual Machine and could be third party software, open-source capabilities like CNN (Convolutional Neural Network) or homemade solutions. Results from the images analysis VM (Virtual Machine) need to be stored and then merged with production context through GUI to be provided to the end users. Basic metrology data infrastructure evolves in a Fab Image Lab like shown in Figure 2.

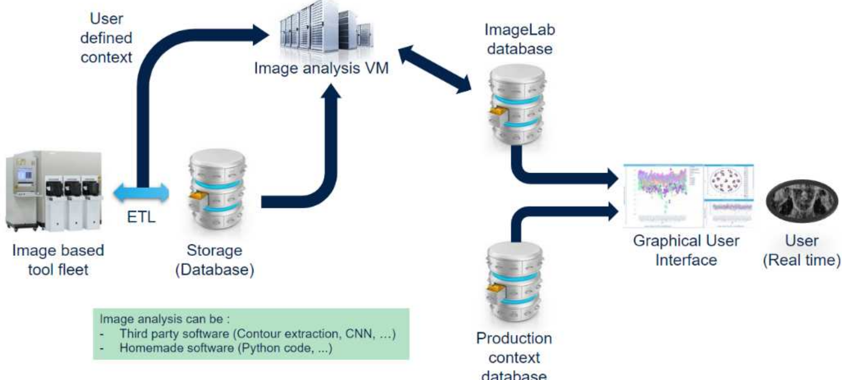


Figure 2: Image lab infrastructure

As illustrated in Figure 3, this environment allows to go beyond single lot analysis by emulating data collection on production line in parallel of the real production environment. Fab Image Lab infrastructure can also be used in

asynchronous mode through retro-processing capability. In this context this system allows reactivity for crisis needs, but also evaluation possibility with pre-production in shadow mode before production release as demonstrated in the next part.

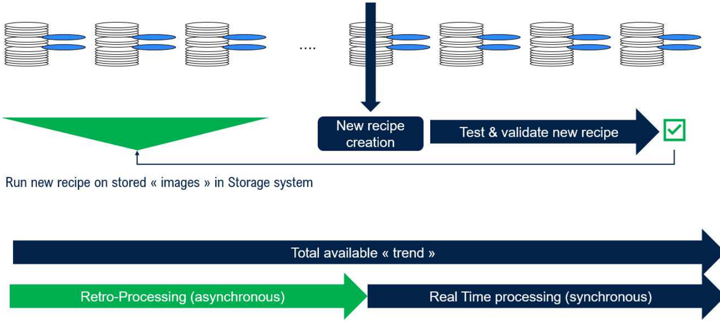


Figure 3: Fab Image Lab replay modes

3. PROOF OF CONCEPT

As said previously, the motivation related to a SandBox is to be able to run (in this case) image processing in real time on a remote environment. The goal is to evidence the added value by a proof of concept from which the return on investment can be clearly shown to wafer fab management. With the SandBox engineers can activate a shadow mode and highlight the added value by capturing use cases rather than waiting for a use case to show up. Many publications have been made from our side illustrating the interest in contour-based metrology. With the SandBox it is now possible to show situations that are relevant for production process or equipment engineering as well as R&D engineering. During the EMLC2023 conference two use cases [2,3] were presented as independent papers and few more are presented in this paper.

1. Use Case one: CDSEM monitoring through image blur control [2]
2. Use Case two: Rinse effect on local variability on 193 immersion process [3]
3. Use Case three: 193immersion Hotspot monitoring [1]
4. Use Case four: Edge Placement difference metrology.
5. Use Case five: Combined Local variability and Overlay metrology on double patterning process.

The first use case relates to the Crolles300mm CDSEM fleet monitoring. Most of the work is described in [2]. To resume this work let us say that our wafer fab is a 20-year-old fab. Therefore, the CDSEM fleet is composed of several generations of CDSEM equipment (old and new systems). With the SandBox, it is possible to provide the same level of analysis whatever the CDSEM type or model. The idea is to show that it is possible to use image blur metrology to build up trends with specs to detect CDSEM image quality drift prior out of control CD situations. The difficulty lies in the sensibility of the image blur computing with respect to SEM type, image acquisition conditions, patterns being measured and material present on the wafer. Solving the equation is not simple, however since the blur value is an analogical value it is possible to compute normalizations so that for a given context specs can be computed and used to detect excursion. In Figure 4.

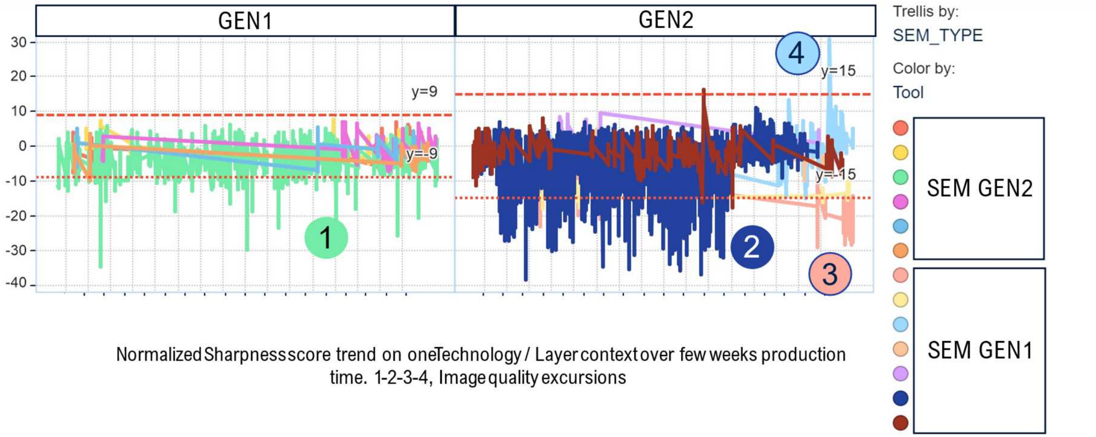


Figure 4: Normalized Image sharpness score applied to CDSEM pictures taken from 13 different CDSEM running on a backend Hard Mask process. 4 image quality excursions are highlighted.

On, trends of normalized CDSEM image sharpness scores computed through the SandBox are being shown for two different models of CDSEM running the same measurements. Excursion 1 is related to regular image quality losses, and it has been evidenced that it was related to electromagnetic interferences that are being fixed. Excursion 2 related to a long-standing situation on a machine that has finally been recovered by a diaphragm change and a column realignment. Excursion 3 is a drift on one machine, the gun tip will be changed, and the tool is “disqualified” for measurement until it is fixed and recovered. The last excursion 4 is a one off and related to a tip current excursion. The SandBox shows here the opportunities that equipment engineers have to trend the image quality and observe the behavior of the tools so that they can take appropriate actions and build recovery plans. They can also instantly observe the results of their actions, gaining time in recovery and limit process tools downtime due to metrology and not process related CD excursion.

The second use case relates to a lithography tool mismatch that has been evidenced by putting in listening mode the local pattern CD variability. On Figure 5, it is shown that although the 2 lithography clusters are matched in term of CD, the local Pattern CD uniformity is not. As explained in [3] it has been found to be related to a rinse process mismatch. By fixing the process brought the tool back to matched state with the other.

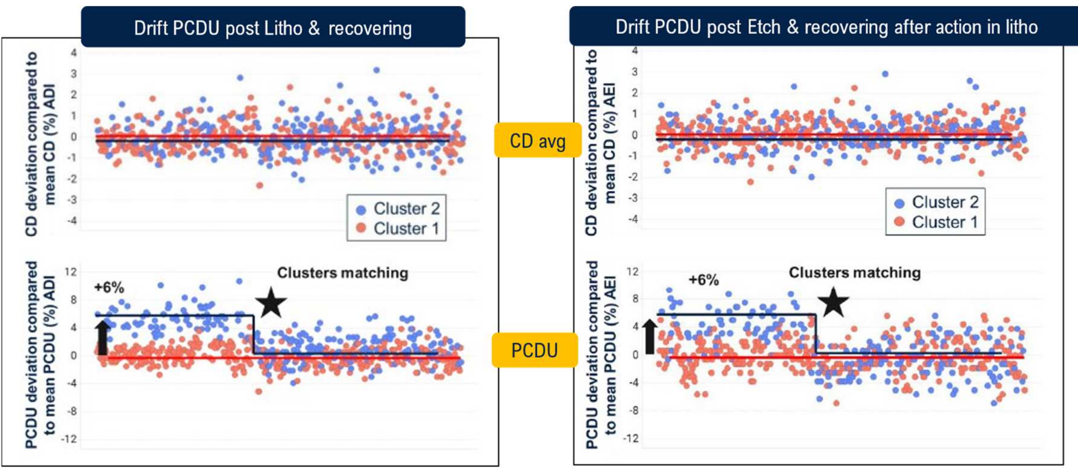


Figure 5: monitoring of CD average along with Pattern CDu (Critical Dimension Uniformity) trends showing cluster mismatch on local variability (on 28nm contact hole process) while average CD was stable. Process fix is shown by having the Pattern CDu matched after the corrective action.

The third and fourth use case relates to in-line CD Hotspot monitoring. The purpose is to show that in production environment Hotspots are showing significant variability while process is still yielding. From this trend a golden reference can be built (from a reference machine / mask) so that any new mask or machine to be qualified can be directly compared to this envelop giving complementary information regarding matching to strengthen the qualification process. Contour metrology can be used in a simple manner, i.e., extract contour and place metrology boxes at some specific locations Figure 6a or, more relevantly, an Edge Placement difference statistic can be computed over the whole contour below Figure 6b.

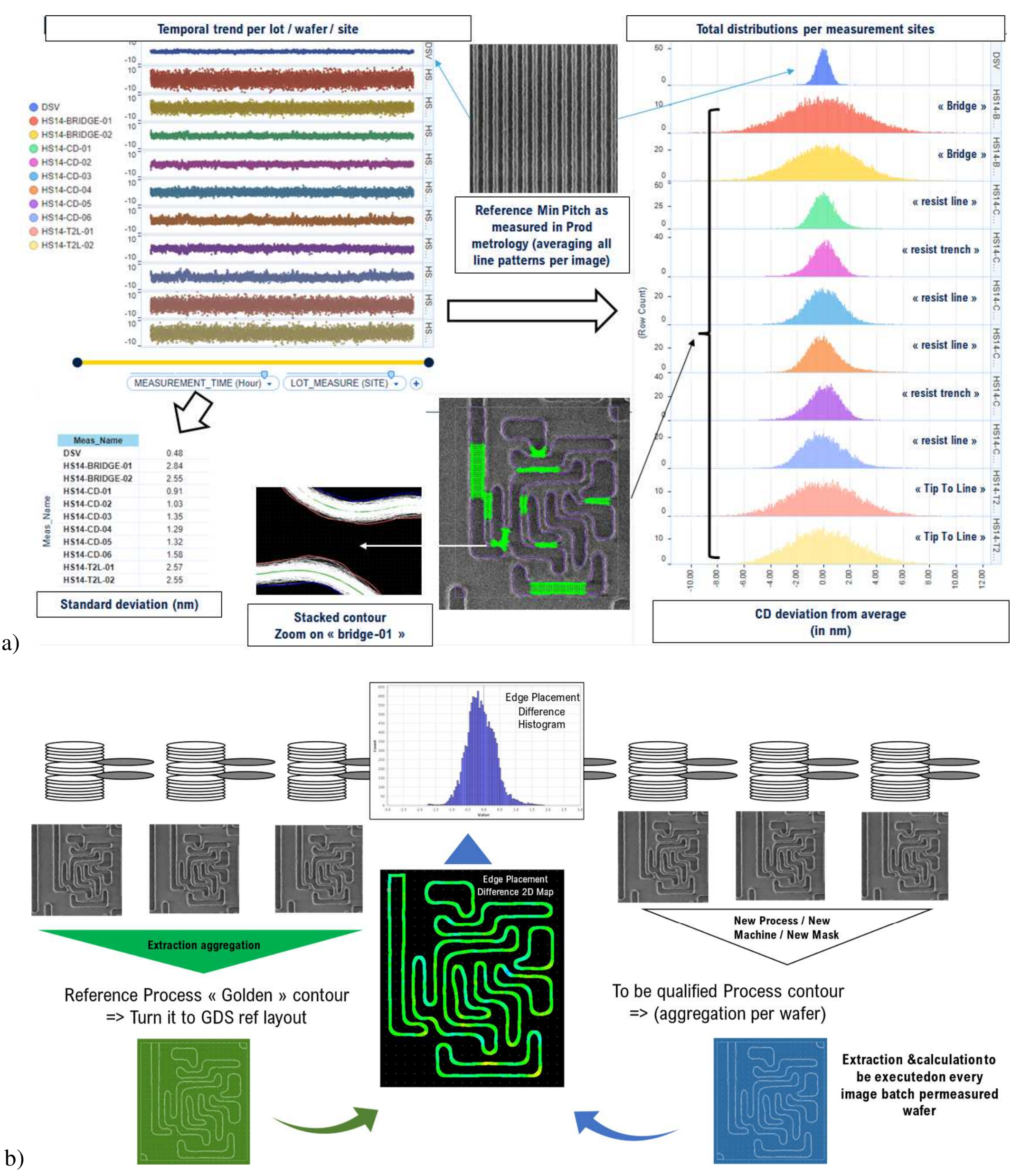


Figure 6: (a) in-line monitoring of various chosen locations within the Hotspots clip – trend (b) extension of contour metrology to full contour usage and comparison of average contour from Golden machine to average contour of process to be qualified – 2D maps and matching histograms.

Finally, the last and fifth use case presented in this paper relates to the capability to couple several kinds of metrology on the same double patterning process image if design layout of the pattern is known. On this figure, the CDSEM takes a picture of a dual layout layer Litho1 Etch1 & Litho2 Etch2 and thanks to the contour metrology in the Sandbox 4 measurements can be performed at the same time (1) the averaged CD of the pooled layers (2) the local CD variability of the pooled layer, goes high if CD mismatch between Litho Etch 1 and Litho Etch 2 (3-4) Overlay measurement by centroid shift between the Litho1 and Litho2.
For process owners, it can be shown that CD can be unstable and due to a mismatch between the two layers and that overlay can show local distortions especially in Y direction where a small non-zero overlay can be observed. On the top graph green mean overlay tends to 0, red to positive values and blue to negative values. Since the structure is not the one on which overlay is being regulated, this local on device overlay measurement highlights a trend to a systematic negative overlay shift. This process is in R&D phase and such data collection, easy and fast to deploy in the SandBox provides a first level of information without requiring a whole configuration in the fab automation system. Further on, overlay models can be tuned to reduce the non-zero overlay and the CD APC control can also benefit from CD mismatch information.

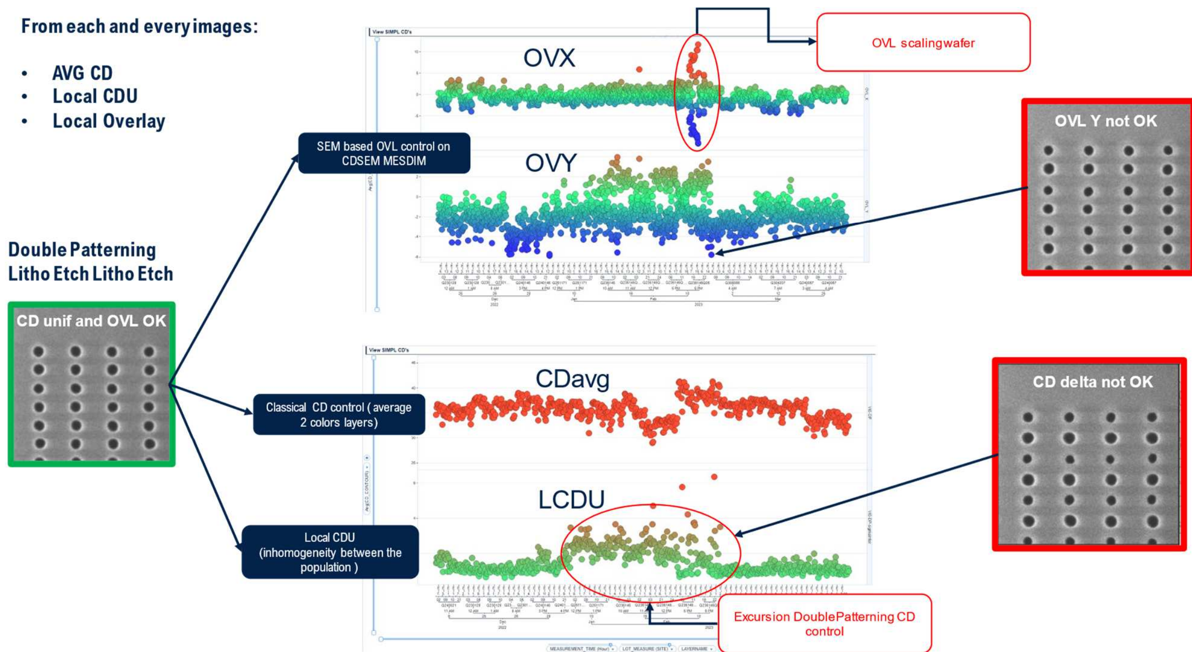


Figure 7: Double pattern process control using contour solution to compare two processes visible in the single image. The contour algorithm is aware of which pattern belongs to which design layer.

4. BEYOND IMAGES

Through all the examples presented in the previous part it has been demonstrated the high potential value for such metrology data infrastructure. Nevertheless, it is important to note that part 2 and 3 focus on a valuable but relatively small part of the metrology data. CD-SEM images interest cannot be argued because they are used to control lithography and etch processes, that take a key place in the product creation flow, but they are not the only metrology images available in Front-End manufactures. Image Based Overlay images, Diffraction Based Overlay images, Atomic Force Microscopy images, optical profilometer images but also alignment images coming from all metrology tools form another important metrology data stock. Furthermore, previous techniques refer only to local metrology techniques. Full wafer metrology types as nano-topography, photoluminescence, shape, etc., also produce small images quantity but information's fat data. Considering all images produced by metrology tools fleet gives us a glimpse of how a Fab Image Lab could become a powerful tool. Of course, enabling such capability requires an important integration cost to raise system versatility. Finally, the most important data source of a metrology tool fleet is still missing. Indeed, images are the easiest metrology data

source to exploit but not the bigger one in quantity. The main data source in a metrology tool fleet is spectra. Of course, spectra use outside the metrology tools is not straight forward but cannot be neglected (Figure 8).

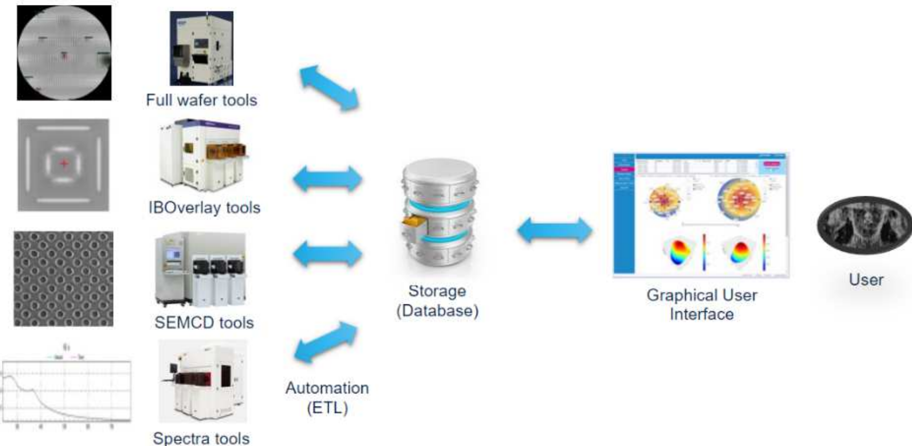


Figure 8. Universal metrology data infrastructure

Having the capability to analyze metrology data outside of the tools opens the possibility to use metrology data for inspection purposes. On other hand similar data infrastructure for defectivity tool fleet could also be interesting to use defectivity data for metrology purposes. In this context of global data emancipation borders between metrology and defectivity are vanishing opening the path for the MetroSpection concept (Figure 9). Several MetroSpection examples have already been published [4-8].

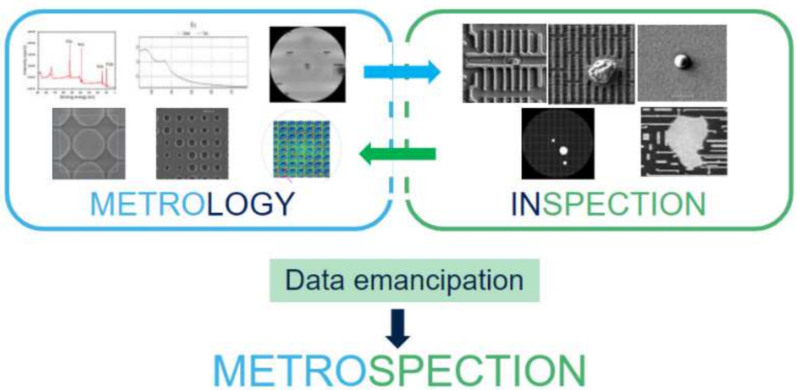


Figure 9: Introduction of MetroSpection concept.

5. CONCLUSIONS

This paper is part of a series of papers published by our team about the engineering put in place to highlight the interest of using contour-based metrology. It has been possible to perform this work thanks to an engineering platform that has been built to support engineers and technicians in the fab to access more data than fab automation standardly proposed. First image processing has been used as a driver, and several use cases have been shown. However, the main purpose of this paper is to show that it is possible and useful to provide versatile and flexible data extraction and analytic environment to engineers so that they can build proofs of concepts with regards to the deployment of new metrology solutions that are not necessarily embedded inside the metrology tools. We intend to extend this paradigm beyond image process and forecast to build an infrastructure customized for all measurement machines we have in the fab and a large panel of complex data. eBeam images, full wafers images from all kinds (topography, photoluminescence, acoustic microscopy, visual inspection), optical images, spectra are rich sources of information that represent opportunities to better, faster react in the fab. This is also for us the beginning of the era of what we call MetroSpection.

REFERENCES

[1] ***A proof of concept of remote contour-based SEM metrology integration in HVM environment*** Bertrand Le Gratiet, Régis Bouyssou, Julien Ducoté, Florent Dettoni, Thibaut Bourguignon, Vincent Morin, Romain Bange, Nivea Schuch, Julien Nicoulaud, Guillaume Renault, Frederic Robert, Thiago Figueiro. Proceedings Volume PC12053, Metrology, Inspection, and Process Control XXXVI; PC120530 (2022) https://doi.org/10.1117/12.2615199

[2] ***CD-SEM image sharpness score monitoring,*** Nicolas Kubler, Bertrand Le Gratiet, Florent Dettoni EMLC 2023

[3] ***Understanding the impact of rinse on SEM image distortion and contact patterning***, Elvire Soltani, Bertrand Le-Gratiet, Sebastien Bérard-Bergery, Jonathan Pradelles, Thibaut Bourguignon, Aurélie Le Pennec, Raluca Tiron EMLC 2023

[4] ***Deployment of convolutional neural network solutions for image computing in semiconductor manufacturing environment***, Bertrand Le Gratiet, Delphine Le Cunff, Laurent Bidault, Thomas Alcaire, Sébastien Desmaison, Régis Bouyssou. Journal of Micro/Nanopatterning, Materials, and Metrology, Vol. 21, Issue 4, 041603, (October 2022) https://doi.org/10.1117/1.JMM.21.4.041603

[5] ***Spectroscopic ellipsometry Imaging for Process Deviation Detection via Deep Learning Approach***, T. Alcaire, D. Le Cunff, V. Gredy and J.H Tortai. ASMC 2020

[6] ***Detection and correlation of yield loss induced by color resist deposition deviation with a deep learning approach applied to optical acquisitions,*** T. Alcaire, D. Le Cunff, JH Tortai, S. Soulan, V. Gredy, M. Templier, M. Kessar, R. Bianchini and A. Berthoud. SPIE 2021

[7] ***Spectroscopic ellipsometry Imaging for Process Deviation Detection via Deep Learning Approach,*** T. Alcaire, D. Le Cunff, V. Brouzet, R. Duru, C. Euvrard, S. Soulan and JH Tortai. ASMC 2022

[8] ***Sensitivity analysis of optical scatterometry technique for high aspect ratio trench measurements,*** J. Grasland, D. Le Cunff and JH Tortai. SPIE 2023

Improving the on-product overlay performance after optimization of the etch-induced contributions

Oktay Yildirim*[a], Richard van Haren[a], Orion Mouraille[a], Ilirjan Aliaj[a],
Jan Hermans[b], Christiane Jehoul[b]
[a]ASML Netherlands B.V., De Run 6501, 5504 DR Veldhoven, The Netherlands; [b] IMEC, Kapeldreef 75, B-3001, Leuven, Belgium

ABSTRACT

To achieve the best on product overlay performance, process induced contributors should be understood, quantified, and addressed. Global wafer overlay penalties have been observed since the etch direction is not always perpendicular to the wafer surface especially at the wafer edge due to the etch chamber geometry and plasma parameter settings. Focus ring concept has been used to compensate etch tilt at the edge however it may be challenging to have zero tilt even with a new focus ring. Etch tool setting optimization can be a solution for accurate overlay control at the wafer edge. Going one level deeper and considering the more local etch-induced overlay penalties, we previously showed that for the inner parts of the wafer there is significant field pattern density dependent etch contribution up to 1-nm at each exposure field although the etch tool is not exposure field aware.

In this experimental work, etch tool parameters such as power and pressure have been varied and two different reticles with different pattern densities are used to reveal the nature of global and die-level overlay penalties. Pressure and power settings are found to be critical for the wafer edge. The Spin-On-Glass (SOG) anti-reflection coating is shown to be the main contributor for the intra-field etch-induced overlay.

We conclude with this work that etch-induced overlay is successfully mitigated at wafer edge and eliminated at intra-field thanks to the understanding on how etch process generates overlay penalties at global and at the die level.

Keywords: Overlay, On-product, Etch, Intra-field, Wafer edge, Exposure field, Etch bias, Trench width

1. INTRODUCTION

The on-product overlay (OPO) as measured on a production wafer has contributors both from scanner and processing. There are solutions available from scanner side to minimize overlay contribution at various length scales and reach overlay performance levels of ~1.6-nm and below [1, 2]. Typically, the etch process is set up and tuned for optimal defectivity and critical dimension (CD) performance of a single layer. The impact of etch step on overlay was secondary. With tighter overlay requirements etch step contribution to overlay gained more attention and solutions for mitigation have been implemented. One major aspect is the high etch contribution to overlay at the wafer edge [3]. During etching a boundary is formed between the wafer and the plasma which is known as “sheath” [4]. The flatness of the sheath enables the incoming trajectory of ions to be straight (perpendicular to wafer plane). Due to the sharp transition of the electric field and plasma at the wafer edge the sheath bends around the wafer edge and causes non-normal ion trajectory [4,5] eventually causing high overlay at the wafer edge [3]. To mitigate the electrical field discontinuity, etch tools are equipped with a focus ring (FR) placed around the wafer, which improves the overlay at wafer edge [3, 6]. The focus ring has a limited lifetime and its performance drifts due to wear (Figure 1). It is not straightforward to achieve zero tilt at the wafer edge since the flatness of the sheath depends on the geometry of the chamber and the focus ring design. An alternative is to achieve flatter sheath and low overlay by impacting the plasma conditions via the applied etch tool settings.

*oktay.yildirim@asml.com; phone +31644835281; www.asml.com

38th European Mask and Lithography Conference (EMLC 2023), edited by
Uwe F.W. Behringer, Jo Finders, Proc. of SPIE Vol. 12802, 128020P
© 2023 SPIE · 0277-786X · doi: 10.1117/12.2675473

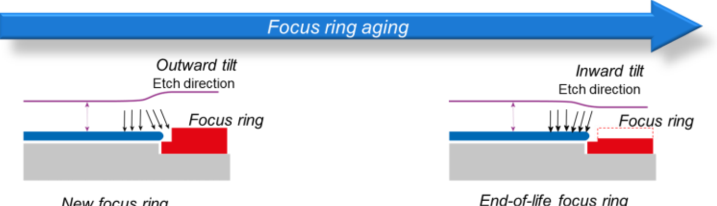


Figure 1. At the wafer edge the electric field is not perpendicular to the wafer surface, due to sharp discontinuity of the electric field. This impacts the trajectory of etch front and eventually causes etch-induced overlay. Focus ring is used as a solution to correct for the tilt. Depending on design of hardware and state of focus ring, tilt can be inward or outward. On top of the geometry, the aging and wear of the focus ring further impact the actual tilt. Etch settings may be tuned to impact the plasma in a way to improve the sheath flatness regardless of the tilt direction.

In the inner part of the wafer there is a significant etch contribution up-to 1-nm at each exposure field although the etch tool is not exposure field aware. To form an efficient elimination strategy at etch tool level a deep understanding on how etch process contributes to the generation of overlay penalties at die level is required. There is a relation with pattern density distribution across exposure field [7].

Despite the continuous improvements on scanner side, due to tighter requirements, handling high spatial frequency overlay penalties at small length scales (sub-mm level) becomes more challenging [8]. This length scale corresponds to die level which repeats at each exposure field where we have reported an extensive study on both etch-induced [7,9] and stress related intra-field overlay [10]. Interestingly, the etch-induced intra-field overlay fingerprints were similar for wafers at different stress types and levels including no stress [7,9,10]. The common part was the Spin-On-Glass (SOG) anti-reflecting coating layer. This hinted that SOG step is the main contributor for the etch-induced overlay at the intra-field.

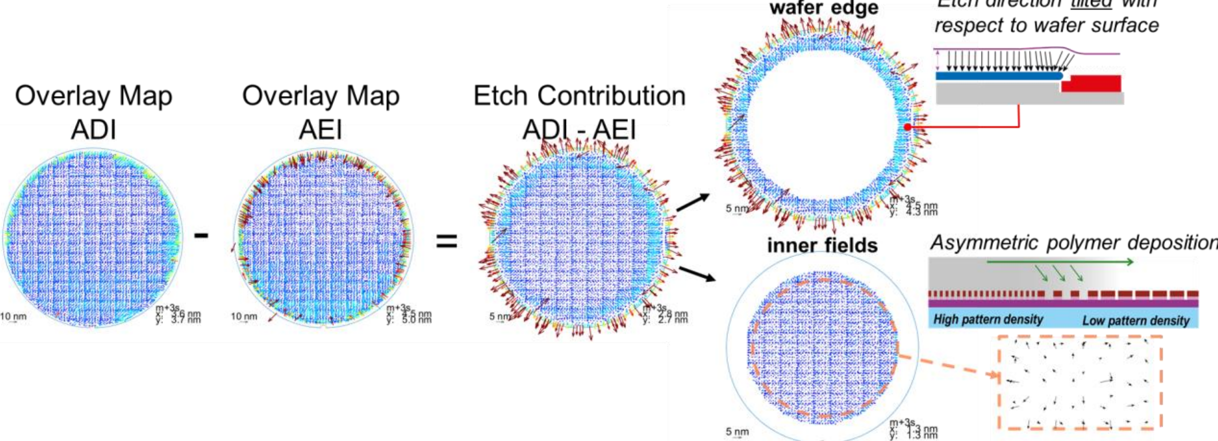


Figure 2. ADI overlay map shows a certain level of overlay for a given use case. After the etch step, overlay measurements are performed on the same wafer at the same locations. The AEI overlay map clearly shows higher overlay at the wafer edge due to etch. To quantify the etch contribution, ADI-AEI difference is calculated where two regions can easily be identified: wafer edge and inner part. At the wafer edge, the high etch contribution to overlay is traditionally compensated by hardware solutions (focus ring) to some extent. However, guaranteeing zero tilt via focus ring is not always straightforward. In that case additional optimization via etch settings can bring further gain. Moreover, the focus ring impact does not extend to inner regions. For the inner part where the etch contribution is less compared to wafer edge but still significant, a different mechanism is dominating, and a solution direction is to optimize the etch settings.

The etch process transfers the resist pattern, but also introduces additional variability known as lithography to etch (LE) bias [11, 12]. SOG etch step was also found to be the main contributor for LE bias [12] which motivated us to check the width of trench features for optimized and Process of Record (POR) etch cases before and after etch.

Overlay after development (ADI) is usually different from the overlay measured after etch (AEI) level. Etch contribution can be quantified simply from ADI-AEI overlay difference, both at wafer [3] and intra-field level [7] as shown in Figure 2. This approach enables isolating the pure etch contribution regardless of the overlay contributions of mask and lithography at ADI level (mask writing errors, lens and reticle heating, aberration impact).

In the current work, we present the results of an experimental study showing the impact of the etch tool settings on etch-induced overlay at the wafer edge. Going one level deeper towards the inner part of the wafer, we show the impact of pattern density distribution and the etch tool settings on etch-induced overlay effects at the intra-field. Based on our findings we propose a hypothesis which explains all our observations regarding the etch-induced overlay at the intra-field. In section 2 the experimental details are shown followed by the results in section 3. In section 4 the results are discussed.

2. EXPERIMENTAL DETAILS

2.1 Stack information and etch-induced overlay as function of the etch tool settings

The stack and overlay measurement methodology to isolate etch contribution is shown in Figure 3. As explained in our previous publication [9], we use the Super Nova 2 process stack which is representative for an N7 LE^3 Metal 1 BEOL process patterning process at IMEC. In the original Litho-Etch (LE) flow, the M1A, M1B, and M1C patterns are first etched into a 20-nm thick SiO_2 memorizing layer. Later on, the three sub-layers are transferred into a 25-nm thick TiN layer. In the current modified flow, the bottom layer (M1A) is etched into the TiN, and it acts as a reference layer to study the etch impact on the overlay targets as defined in the M1C layer. Before exposing the M1C mask, a 150-nm thick SiO_2 layer is deposited followed by a Spin-On-Carbon (SOC), a Spin-On-Glass (SOG), and a photo resist layer. The bottom overlay metrology targets in the 25-nm thin TiN layer have high contrast and negligible side wall target asymmetry. At ADI level, the top layer is defined in photoresist. At AEI the reference bottom layer remains same while the top layer is defined in SiO_2 layer. To study the intra-field pattern density impact, M1B top layer exposure is performed similar to M1C. Diffraction based overlay targets (µDBO) were selected to characterize the etch-induced overlay effects and measured on the YieldStar S-375 metrology tool. To have dense overlay information, 51 positions per field are measured. All etch results were obtained on a TEL™ CCP etch tool at IMEC. Scanning electron microscopy (SEM) measurements are done at IMEC using Hitachi CD SEM CG-5000, with 500V and 1024x1024 pixels, to monitor the width of the trench features at ADI and AEI.

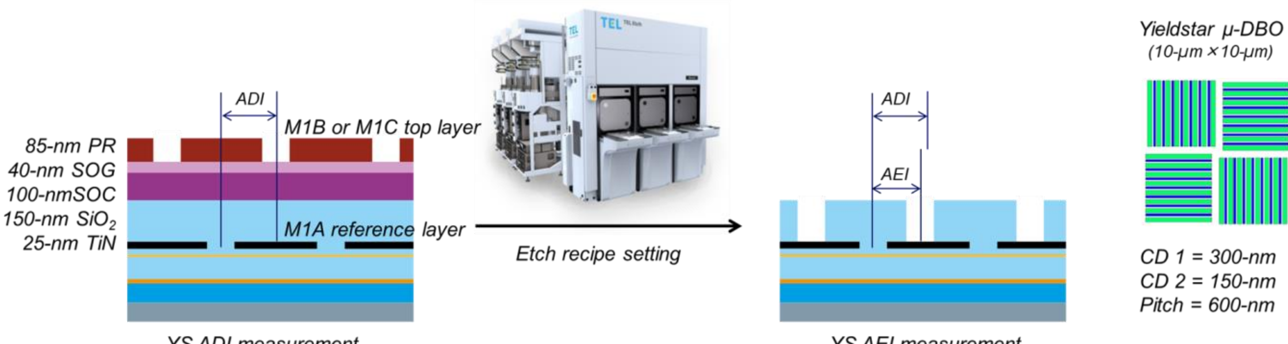


Figure 3. Yieldstar (YS) overlay measurements are performed after photo resist development (ADI) and after etch (AEI) for many different etch recipes on the YS (interlaced) overlay targets. Etch contribution is quantified by (ADI-AEI) on (the same) interlaced POR YS target on YS-375. All the etch contributors which are present at ADI level are also present at AEI, with the addition of etch contribution. By taking ADI-AEI difference, contributors other than etch drop out. In this way pure etch step contribution is quantified.

In table 1 we show the two wafers where we use process-of-record (POR) etch conditions and optimized etch conditions to eliminate etch contribution to intra-field overlay. We show only the relevant conditions which belongs to the critical SOG etch step. There are more etch settings and more layers which went through etch [7,9].

Table 1. Normalized etch tool settings for the critical layer SOG for optimized etch and POR etch

Wafer	Pressure	Chemistry	Temperature
POR etch	1	CF_4/CHF_3	1
Optimized etch	0.3	CF_4	4

3. RESULTS

3.1 Etch tool settings impact on global etch-induced overlay fingerprint during SiO_2 etch

Figure 4 shows the impact of increasing the pressure at the SiO_2 etch step on etch-induced overlay (ADI-AEI) for the wafer (top) and for the intra-field (bottom). Wafer plots clearly show that higher pressure conditions are favorable regarding overlay mitigation at wafer edge. We observe ~7-nm overlay gain at the wafer edge with increased pressure compared to POR case. From this, we conclude that the etch direction is more perpendicular to the wafer surface at high pressure conditions. Contrary to the wafer edge, intra-field etch contribution fingerprint increases with pressure, which will be explained later on in relation with the proposed intra-field mechanism.

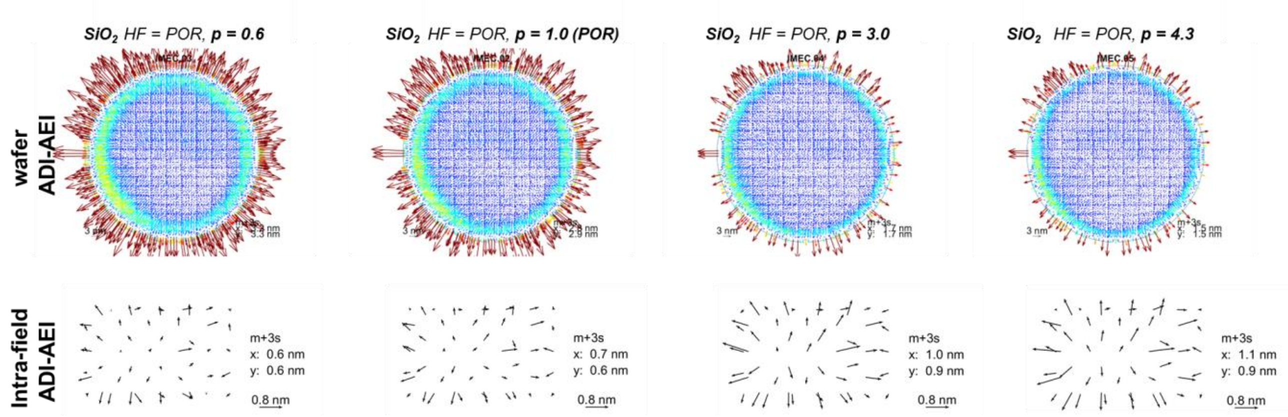


Figure 4. Pressure at SiO_2 etch is changed from 0.6·POR to 4.3·POR. ADI-AEI plots show the etch-induced overlay significantly drops at the wafer edge with increased pressure. Smaller overlay is possibly due to a flatter sheath due to a denser plasma. At the intra-field, which is shown below wafer plots, the pressure increase causes an increase at the etch-induced overlay, which will be explained further in the paper at related sections.

In Figure 5 we show the ADI-AEI (etch contribution) results of cases where HF (high frequency) power at SiO_2 etch step is increased while other etch settings (including pressure) are kept at POR values. Increasing HF has clear improvement at the etch-induced overlay at the wafer edge (~7-nm). An advantage of using HF setting is that it causes a smaller negative impact on the intra-field performance compared to pressure increase. Here, again the plasma – wafer interface seems to be flatter when HF is increased. Both HF and pressure are effective settings at SiO_2 etch step to improve the edge tilt. Changing etch settings while keeping the same wafer stack and chamber (geometry), seems to impact plasma properties as well as sheath flatness. Pressure and HF increase create more polymerizing conditions. This may be another mechanism at work at the wafer edge, however the ion trajectory (tilt) dictated by sheath flatness seems to be dominating.

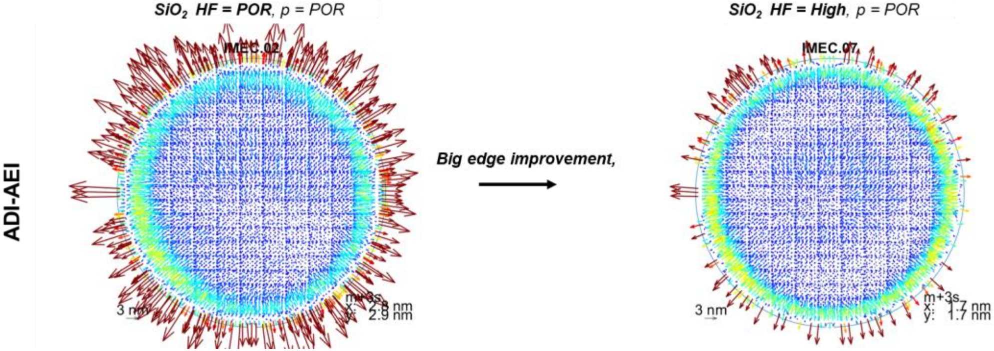


Figure 5. HF power at SiO_2 etch is increased. ADI-AEI plots show the etch-induced overlay significantly drops at the wafer edge with increased HF power. Smaller overlay is possibly due to a flatter sheath due to a denser plasma. Unlike increasing pressure, increasing HF only slightly increases intra-field fingerprint [9].

3.2 Etch tool settings impact on global etch-induced overlay fingerprint during SOG etch

From earlier work where SiN and SiO_2 containing stacks are etched [7, 9, 10], SOG etch step is identified as critical step for intra-field etch-induced overlay contribution. It is the common layer for all cases and eventually resulting in similar etch fingerprints where same reticles (pattern density) are used. SOG etch step conditions optimization with lean chemistry and lower pressure is expected to improve intra-field etch contribution fingerprint. We first wanted to see the impact of such SOG etch setting modification on overlay at the edge. We saw that etch settings favoring intra-field optimization has no negative impact on the wafer edge and produced even a slight improvement (Figure 6). This gave us confidence to further focus on the SOG etch step optimization without compromising the wafer edge. The main etch-induced intra-field overlay contributors and mitigation via SOG etch step optimization will be discussed in the following sections.

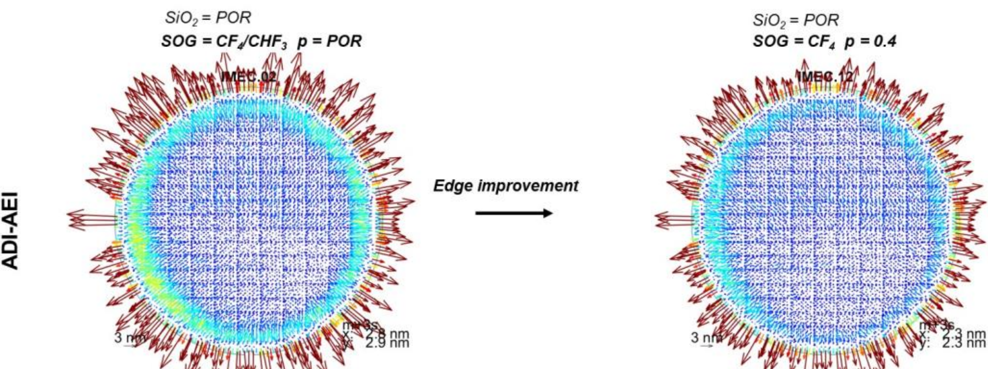


Figure 6. SOG step is critical for intra-field etch-induced overlay. Using favorable etch conditions (lean chemistry and lower pressure) for intra-field also results in improvement at the wafer edge as seen in ADI-AEI plots.

3.3 Pattern density distribution impact on intra-field etch-induced overlay fingerprint

In our earlier work we have shown that the etch-induced overlay fingerprint at the intra-field has a relation with the pattern density variation within the field [7]. In this work we take an extra step and compare two cases where pattern density is changed by using different reticles in the top layer exposures. To be able to make a fair comparison, all the stack and etch steps are kept the same (POR) where M1A acts as a reference layer in both cases. Using the same etch process means having the same chemistry, the same plasma conditions and the same etch duration for both cases. The main difference is the pattern density gradient between regions within the exposure field and the rate at which etch by products are generated depending on the pattern density/surface area. M1C reticle has a higher variation along the field

while M1B has a more uniform distribution in terms of pattern density. For both cases, pattern density distribution explains the etch-induced overlay fingerprints obtained by dense sampling (51 points per field).

Smaller pattern density variation within the field in case of M1B results in smaller arrows (=smaller etch-induced overlay) when compared to larger arrows at M1C (Figure 7). The yellow-range high pattern density region seems to have high impact (M1C). Its counterpart at M1B has smaller pattern density variation, eventually resulting in smaller impact. The arrows at the edge in the upper part of M1C are similar to the M1B case, which may suggest that the high density region in the center has a limited impact at that length scale (~5-mm). The impact becomes apparent when the row of arrows halfway between upper edge and yellow blocks is considered: in case of M1C that row has large arrows, while M1B case corresponding row has much smaller arrows. This comparison gives insight for the proposed mechanism and the extent of the pattern density variation impact (up to ~2.5-mm).

It becomes apparent that to mitigate the etch impact on overlay, smaller pattern density (variation) should be aimed at, which may not be possible for all layers. Additional precautions are needed to achieve low overlay. That is to solve the problem at the root cause, in other words optimizing the etch step itself to generate conditions where the etch is straight with no additional contribution to overlay. In the following sections etch optimization with the high pattern density variation (M1C) case will be explained.

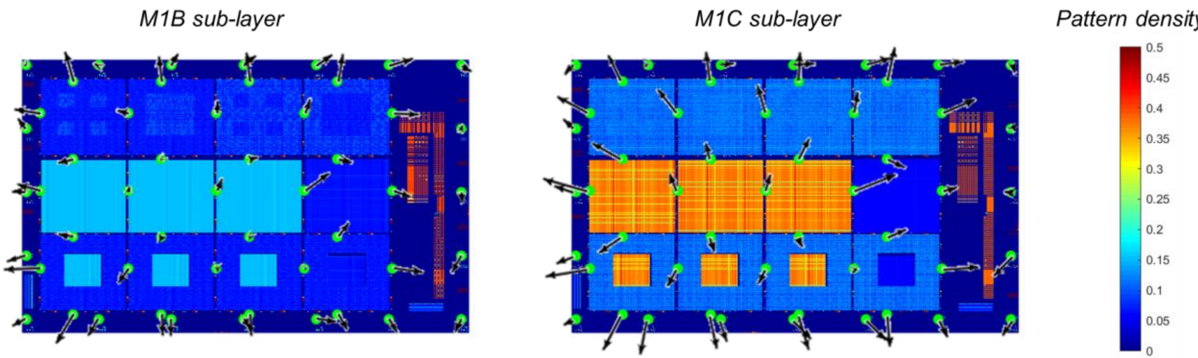


Figure 7. The intra-field etch-induced overlay maps for two different pattern density distributions. M1C and M1B cases use same POR stack and same POR etch conditions, only the reticle (pattern density) is changed. Smaller pattern density gradient (M1B) results in smaller etch contribution. This is reflected in the smaller (ADI-AEI) vector sizes within the (3×4) pattern density blocks at M1B compared to M1C. The intra-field (ADI-AEI) signatures show that the etch-induced overlay penalties have a long-range character.

3.4 Etch tool settings impact on intra-field etch-induced overlay fingerprint during SOG etch

SOG etch step is identified as critical layer to create the etch-induced intra-field overlay signature at several cases in combination with pattern density effect. Here we want to tune SOG etch step in a way that even with high pattern density we can create conditions where etch impact is minimal. The strategy is to tune the conditions in such a way that etch is clean, straight, polymerization and material re-deposition are minimal. One way is to change the chemistry from CF_4/CHF_3 to CF_4, to mitigate polymerization by increased F/C ratio. Lower pressure also helps mitigating polymerization. Increasing the temperature further reduces the sticking coefficient enabling a better exhaust of the etch by-products. Figure 8 shows that changing the chemistry and decreasing the pressure already brings significant reduction of the etch fingerprint (middle) compared to POR case (left). Starting from these new power and pressure conditions (middle) which determine the ion energy and trajectory, further improvement can be achieved by increasing the temperature, as seen in Figure 8 on the right. This particular result gives insight about the underlying mechanism, showing that the impact is related to polymerization and deposition rather than the ion trajectory change for the inner regions.

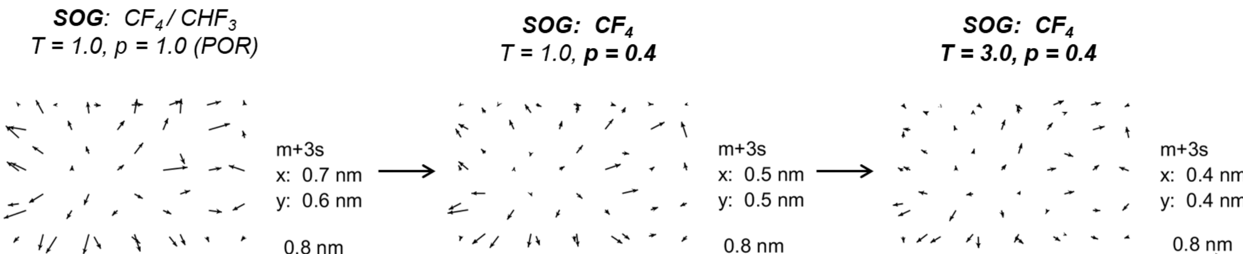


Figure 8. The POR intra-field etch-induced overlay map is shown on the left. Clean conditions created to lower polymerization by chemistry change to higher F/C ratio and lower pressure significantly reduces etch contribution fingerprint (middle). Further improvement is shown by reducing polymer sticking achieved by temperature increase (right). By driving only the SOG conditions in to a clean etch, etch contribution is significantly reduced.

3.5 Mitigation of the intra-field etch-induced overlay fingerprint

To fully eliminate the etch contribution to overlay at the intra-field we apply the knowledge gained in the previous section and tune the settings further to enable even cleaner conditions. When we increase the temperature and reduce the pressure further, we can eliminate the etch fingerprint (Figure 9, right) and show significant gain compared to POR (Figure 9, left). The etch fingerprint in the POR case has a certain shape and magnitude, formed in relation with pattern density. When the SOG etch is optimized the etch fingerprint is gone and randomized. The fingerprint is no longer related to pattern density distribution although the same M1C reticle was used in the exposure. In this way we show successful elimination of etch-induced intra-field overlay penalties and a significant gain of 0.4-nm.

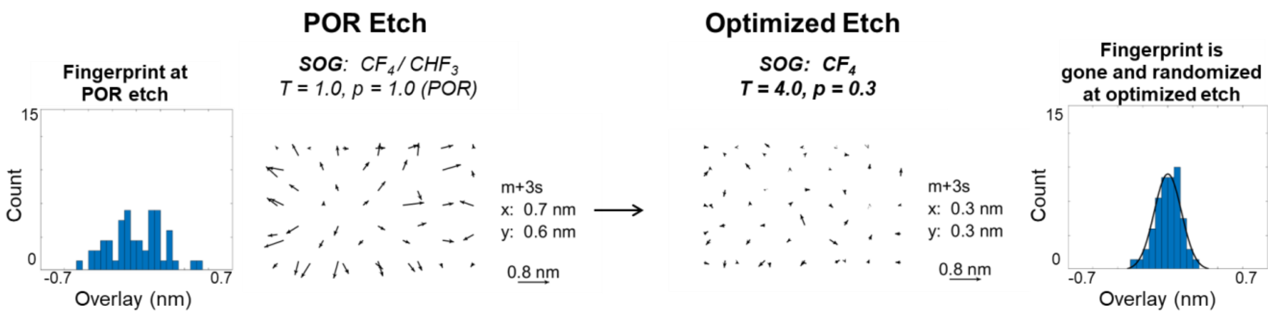


Figure 9. The etch-induced overlay fingerprint for the POR SOG etch recipe is shown on the left-hand-side. When we clean up the SOG etch reducing hydrogen content by only using CF_4, reducing pressure setting and further increasing the temperature (less polymer sticking) the fingerprint disappears. At POR case there is a specific etch fingerprint and its distribution which is randomized and became gaussian-like when etch contribution is eliminated.

3.6 Proposed mechanism for the intra-field etch-induced overlay

Combining our learnings about pattern density impact and (SOG) etch parameters impact, we propose a mechanism (shown in Figure 10) to explain how the etch step results in overlay penalties. This mechanism explains our observations, moreover, gives us ability to tune the etch process in order to eliminate the etch-induced overlay (Figure 8, 9).

Etch by-products are generated at a higher rate at high pattern density regions within the field. There is diffusion of materials above etched features from high to low pattern density regions due to concentration gradient. These drifting species redeposit at the side walls of features being etched in an asymmetric manner which in the end results in an overlay shift (non-zero ADI-AEI difference). As discussed earlier, the impact can be mitigated either by reducing concentration differences (lower pattern density variation) or via mitigation polymerization with proper etch settings.

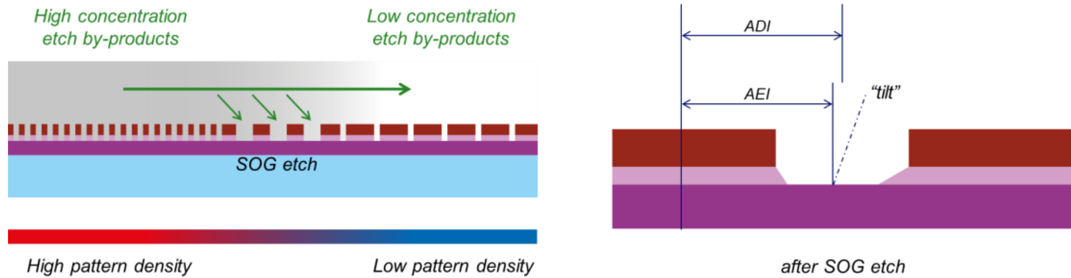


Figure 10. High concentration of etch by-products are formed above the high pattern density areas. They diffuse away from these areas towards the low pattern density areas. The non-volatile etch products re-deposit asymmetrically on the overlay metrology grating side walls. This results in a "tilted" etch (asymmetric tapering) and consequently an (ADI-AEI) difference.

3.7 Trench width behavior after improving the etch-induced intra-field overlay

A natural question may come up as what happens to the trench width after optimizing the etch tool recipe for overlay. Etch conditions are known to have impact on LE bias [11, 12]. To monitor the impact of the etch step on the width of trench features, SEM measurements are done at ADI and AEI. The trench width behavior due to etch step is schematically shown in Figure 11. A reduction of the trench width is expected due to the (asymmetric) polymer deposition in case of POR etch. This may be a function of trench width. On the other hand, no change in trench width is expected for optimized etch, since the etch conditions are tuned to eliminate polymer deposition. We validated this on a trench width of 47-nm after resist development (ADI). The observations show indeed the expected trench behavior: trench width remains 47-nm for the optimized etch while it reduces to 36-nm for the POR etch. It is clear that if we want to maintain the 36-nm trench width after etch for the optimized etch, we need to apply a bias at mask level.

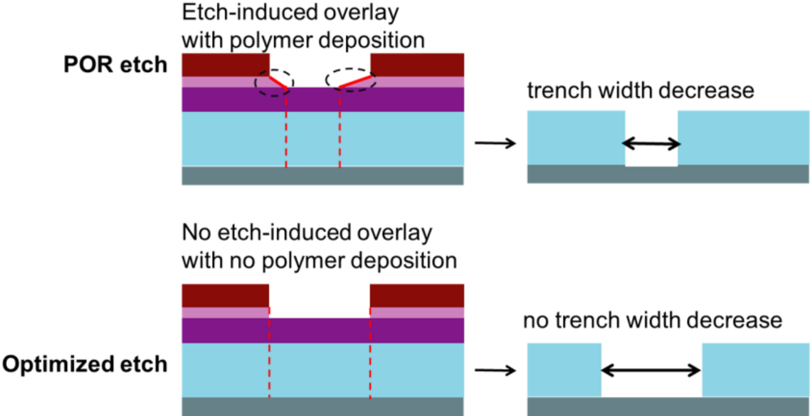


Figure 11. Trench width behavior is as expected upon etch as measured on device-like features. POR etch process results in smaller trench width due to polymer deposition. Optimized etch process maintains the trench width as expected.

4. DISCUSSION

The etch step has significant contribution to overlay both at wafer edge and at the intra-field. The impact at the wafer edge is much larger compared to inner regions. This already hints towards the fact that these two regions need separate attention and approach both for understanding the mechanism and for mitigating the etch-induced overlay. At the wafer edge the dominating mechanism seems to be the trajectory of incoming ions to the wafer being under an angle due to the sharp discontinuity of the electric field at the wafer edge. The impact seems to be long range and extends towards the

center for a few millimeters. The presence, age, and design of a focus ring strongly influences the end-result. It is commonly thought that the boundary between plasma and wafer (sheath) is made flatter to impact the incoming ion trajectory straight by making use of the focus ring. On the other hand, we cannot rule out other potential mechanisms which may have an impact, such as polymerization/deposition of materials coming from the focus ring direction, however this may be possibly smaller compared to sheath flatness aspect. Regardless, the use of focus ring does not guarantee full elimination of etch-induced edge overlay. Whether the mechanism is sheath flatness or something else, additional precautions are needed to mitigate high etch contribution close to wafer edge and that is tuning the etch settings as well on top of optimizing hardware and geometry. Settings such as pressure or HF power at the SiO_2 etch yields very large gain (~7-nm) at the wafer edge (ADI-AEI). This is a result of the improved and denser plasma in combination with the flatter sheath at the wafer edge (Figure 12). Although a large overlay gain can be observed at the wafer edge, it negatively impacts the etch-induced intra-field overlay performance as shown in Figure 4. However, if these intra-field overlay penalties can be absorbed, the pressure and HF power settings can be used to improve the overlay at the wafer edge.

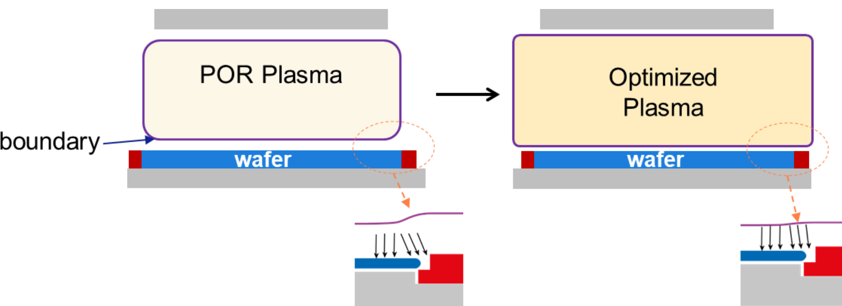


Figure 12. Optimizing etch conditions result in reduced etch-induced overlay at the wafer edge (ADI-AEI gets smaller). The potential mechanism may be the denser and extended plasma resulting in a flat plasma-wafer boundary, in the end enabling straighter trajectory for the incoming ions around wafer edge.

The second aspect is the etch contribution at the intra-field, which is much less in magnitude compared to etch-induced overlay at the wafer edge. However, with tighter requirements, it is very important to address the underlying mechanism and mitigation strategy for etch-induced overlay at the intra-field as well. We have shown that the intra-field etch contribution to overlay can be fully eliminated by optimizing the etch tool recipe settings. However, the overlay penalties at the wafer edge remain, see Figure 6.

It looks like the etch-induced overlay contribution cannot be optimized for the intra-field and wafer edge simultaneously by only changing the chemistry and etch tool recipe settings. This is a direct consequence of the different mechanisms that play a role. While the chemical discontinuities drive the overlay performance within each exposure field, the electrical discontinuities dominate at the wafer edge.

This knowledge provides us a solution path to solve the etch-induced overlay across the full wafer. The first step would be to eliminate the intra-field etch-induced overlay penalties by optimizing the chemistry and etch tool recipe settings. By doing so, the chemical discontinuities can be mitigated. Subsequently, the overlay penalties at the wafer edge due to electrical discontinuities can be mitigated by the application of a tunable focus ring [3].

5. CONCLUSIONS

In this work, we have studied the etch-induced overlay penalties that are present both at the wafer edge and at the inner regions (exposure field level). The etch-induced overlay contribution can be quantified by subtracting the overlay that is measured after etch (AEI) from the overlay that is measured after photo resist development (ADI) on the same overlay metrology targets. We observe at wafer edge ADI-AEI is heavily dependent on etch parameters such as pressure and HF power at SiO_2 etch step. With higher Pressure or higher HF power ~7-nm overlay gain can be achieved at the wafer edge.

Etch conditions have a big impact on the etch contribution to intra-field overlay as well (0.3 to1.7-nm). Unlike the wafer edge, inner regions have a much smaller but still significant etch-induced overlay contribution. The SOG etch step has been identified as the main contributor. The etch fingerprint heavily depends on the pattern density distribution which we validated by using two different masks. The reticle with high pattern density variation resulted in significantly higher etch-induced overlay contribution although the stack and etch conditions were the same. Avoiding pattern density variations at the masks is not always possible in practice. The main mechanism that is responsible for the etch-induced overlay fingerprint at the intra-field is pattern density dependent asymmetric polymer deposition. Depending on the use case, etch-induced overlay at the wafer edge or at the intra-field can effectively be mitigated by optimizing the etch chemistry and etch tool settings. However, it is hard to optimize the etch-induced overlay penalties for both the intra-field and wafer edge simultaneously. We conclude that the intra-field etch-induced overlay penalties should be mitigated by tuning the etch tool chemistry and recipe settings first. The remaining overlay penalties at the wafer edge should be taken away with a tunable focus ring.

Acknowledgements: We thank Yannick Feurprier and Kaushik Kumar from TEL™ for the etch optimization work and discussions.

REFERENCES

[1] Roelof de Graaf, Stefan Weichselbaum, Richard Droste, Matthew McLaren, Bert Koek, Wim de Boeij, "NXT:1980Di immersion scanner for 7nm and 5nm production nodes," Proc. SPIE Vol. 9780, 978011 (2016).

[2] Roderik van Es, Mark van de Kerkhof, Arthur Minnaert, Geert Fisser, Jos de Klerk, Joost Smits, Roel Moors, Eric Verhoeven, Leon Levasier, Rudy Peeters, Marco Pieters, Hans Meiling, "EUV for HVM: towards an industrialized scanner for HVM NXE3400B performance update," Proc. SPIE Vol 10583, 105830H (2018).

[3] Richard van Haren, Victor Calado, Leon van Dijk, Jan Hermans, Kaushik Kumar, Fumiko Yamashita, "Wafer edge overlay control solution for N7 and beyond," Proc. SPIE 10589, 105890D, (2018).

[4] Doosik Kim, Demetre J. Economou, "Simulation of a two-dimensional sheath over a flat insulator–conductor interface on a radio-frequency biased electrode in a high-density plasma," J. Appl. Phys., Vol. 95, No. 7, (2004)

[5] Mingmei Wang, Mark J. Kushner, "High energy electron fluxes in dc-augmented capacitively coupled plasmas I. Fundamental characteristics," J. Appl. Phys. 107, 023308, (2010).

[6] Natalia Y. Babaeva, Mark J. Kushner, "Penetration of plasma into the wafer-focus ring gap in capacitively coupled plasmas," J. Appl. Phys. 101, 113307, (2007).

[7] Richard van Haren, Oktay Yildirim, Orion Mouraille, Leon van Dijk, Kaushik Kumar, Yannick Feurprier, Jan Hermans, "Intra-field etch-induced overlay penalties, "Proc. SPIE 11329, 1132910 (2020).

[8] D. M. Slotboom, P. Hinnen, J. Mulkens, "On-product overlay solutions for DUV and EUV mix-scanner usage in an EPE-driven patterning world," Proc. of SPIE Vol. 12051 120510L-1, (2022).

[9] Richard van Haren, Oktay Yildirim, Orion Mouraille, Leon van Dijk, Kaushik Kumar, Yannick Feurprier, Christiane Jehoul, Jan Hermans, "Mitigation of the etch-induced intra-field overlay contribution," Proc. SPIE, 120560D (2022).

[10] Richard van Haren, Orion Mouraille, Oktay Yildirim, Leon van Dijk, Kaushik Kumar, Yannick Feurprier, Jan Hermans, "On product overlay characterization after stressed layer etch," Proc. SPIE, 116150N (2021).

[11] Prem Panneerchelvam, Ankur Agarwal, Chad M. Huard, Alessandro Vaglio Pret, Antonio Mani, Roel Gronheid, Marc Demand, Kaushik Kumar, Sara Paolillo, Frederic Lazzarino, "Evolution of lithography-to-etch bias in multi-patterning processes," J. Vac. Sci. Technol. B 40(6), (2022).

[12] Prem Panneerchelvam, Chad M. Huard, Ankur Agarwal Alessandro Vaglio Pret, Antonio Mani, Roel Gronheid Marc Demand, Kaushik Kumar, "Trilayer hardmask lithography and etch for BEOL manufacturing," Proc. SPIE 1205320, (2022).

The Use of Cross-Validation for Overlay Model Selection

Sungwoo Jung*[a], Dr. Clemens Utzny[a]
[a]Qoniac GmbH, Dr.-Külz-Ring 15, 01067 Dresden, Germany

ABSTRACT

Overlay requirements are becoming more demanding as the semiconductor device design dimensions continue to reduce in size. Thus, it is increasingly crucial to control overlay under tightened specifications to maximize product yield. To meet these requirements, selecting the best overlay model is important to best manage the lithography process with confidence. A well-characterized overlay model is essential in the wafer production process as it becomes the gauge to optimize product yield, reduce rework process costs and shorten cycle time. In this paper, we will introduce the powerful machine-learning method of Cross-Validation (CV) which helps to improve the prediction capability of an overlay model. This method provides a numerical value that can indicate how well the overlay model predicts the misalignment of a new wafer. Our test result shows that the 5th-order model exhibits an overfitting problem, while the 4th-order model shows good performance. We also discuss how we apply the CV to the correction per exposure (CPE) models that are commonly used.

Keywords: Overlay, CPE, Model, Cross-Validation

1. INTRODUCTION

As semiconductor devices increase in complexity, the control of overlay becomes increasingly more demanding under tightened specifications [1]. Lithographic compensation in Electron-beam based masked writing differs from the options for photon-based wafer exposure [2]. To achieve this, many overlay models have been introduced in recent years, such as high-order process correction (HOPC), intra-field high-order process correction (iHOPC), and correction per exposure (CPE) [3]. These models serve as input into the exposure settings of a scanner, correcting systematic overlay errors, and thus helping manufacturers achieve the required overlay control. However, a high-order overlay model sometimes can cause overfitting, leading to unexpected wafer rework. Such rework is due to the poor prediction capability of models with overfitting. While using a high-order overlay model is advantageous in reducing the residual mean + 3sigma (M3S), it does not ensure a good capability of predicting new wafers' overlay. That is why many engineers would value an objective measure for an overlay model that best fits their process. The key objective is to figure out which overlay model can deliver a good balance between prediction capability and the smallest overlay residual.

In this paper, we introduce the machine-learning technic Cross-Validation (CV) that can help engineers characterize which overlay model has the best prediction capability for new wafers and which overlay models have the potential to cause an overfitting issue. Since the CV method can be applied to wafer-based overlay models like HOPC and iHOPC, and field-based overly models like CPE, this approach would help engineers select a suitable overlay model for their process.

2. CROSS-VALIDATION ALGORITHM

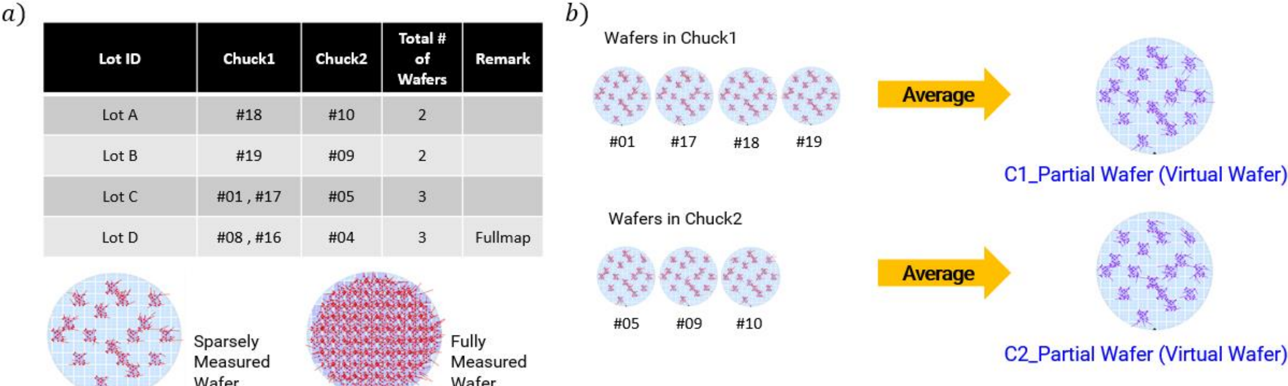
There are several types of CV such as Leave-One-Out Cross-Validation (LOOCV), K-Fold Cross-Validation, and Hold-Out Cross-Validation. In this paper, we utilized the LOOCV, a special case of K-Fold CV [4], and a common method applied to overlay measurement data. The CV gives a Root Mean Square Error (RMSE) for each overlay model as an output. This RMSE value can be used as an indicator for selecting which overlay model has the best prediction capability and which model shows overfitting problems. [5]

*Sungwoo.Jung@Qoniac.com; phone 49 177 239-1781

38th European Mask and Lithography Conference (EMLC 2023), edited by
Uwe F.W. Behringer, Jo Finders, Proc. of SPIE Vol. 12802, 128020Q
© 2023 SPIE · 0277-786X · doi: 10.1117/12.2675560

2.1 Data Set Introduction for CV

Figure 2-1 (a) shows the overlay measurement data of ten wafers. Seven of the ten wafers are sparse measurement wafers and the other three wafers are fully measured. Since most manufacturers typically use sparse overlay maps due to the measurement time and their productivity, we also decided to focus on the seven sparsely measured overlay wafers. Unlike the actual production process flow, where a lot influences the subsequent ones by giving and taking feedback from Advanced Process Control (APC) correction, the simplest CV process is more like a discontinuous flow where lots are treated as temporarily unordered. To reflect the common model set ups for a process flow in a fab with the CV process as best as possible, some virtual wafers need to be defined by calculating the average of wafers based on their context. In this case we aim to determine the optimal chuck-specific models.



Lot ID	Chuck1	Chuck2	Total # of Wafers	Remark
Lot A	#18	#10	2	
Lot B	#19	#09	2	
Lot C	#01 , #17	#05	3	
Lot D	#08 , #16	#04	3	Fullmap

Figure 2-1. (a) The information of the wafers used for the CV. Seven wafers are based on a sparse measurement map; the other three are fully measured. (b) Definition of two virtual wafers that can represent each chuck. Each virtual wafer is the average of all single wafers in chuck one and chuck two, respectively.

Figure 2-1 (b) shows how the virtual wafers can be defined. Chuck 1 Partial Wafer can be defined by calculating the average of all single wafers belonging to Chuck 1, point by point. This virtual wafer is used to represent all the wafers in the Chuck 1 since it is the average of them. Likewise, the Chuck 2 Partial Wafer can also be defined in the same way. Those two virtual wafers are then used as a data set for calculating the CV process.

2.2 Essential Algorithm of the CV

Assume that the total number of overlay data in one of the virtual wafers is N. This can be divided into two distinct data sets. The first data set is called the Training Data Set, which will be used for calculating the CV for each wafer model, and the other is called the Validation Data Set, which will be used for computing scores for both X and Y directions. Depending on the user's choice, the fraction of overlay data included in both data sets could be changed. In general, however, the typical case would have more overlay data in the training data set than in the validation data set.

Figure 2-2 (a) shows how the CV process works. The iteration count is determined depending on the number of overlay data in the training data set, and the rest of the overlay data is assigned as the validation data set. Let us assume that there is the K number of overlay data in the training data set. In each iteration, one test overlay data is excluded from the training data set, and the other overlay data, consisting of the number K-1, is used for fitting the overlay model, as shown in Figure 2-2 (b). After fitting the overlay model, the Square Error (SE) of the test overlay data can be calculated based on its predicted overlay data ($(\hat{y_i})$) and the actual measured overlay data of the sample (y_i). By iterating this process K times, a K number of SE values can be obtained. The final output of the CV for the overlay model, which is Root Mean Square Error (RMSE), can be found by calculating the average of all SE values and taking the square root, as shown in Figure 2-2 (c).

The cross-validated RMSE is an indicator of the prediction capability of the given overlay model. An overlay model with good prediction capability has a small RMSE value, while a large RMSE means the given overlay model is not good at predicting a new wafer's overlay. This constitutes a marked difference from the model residual. By comparing

the RMSE value of each overlay model, we can know which overlay model shows an overfitting issue that can cause trouble such as a rework. The overfitting problem will be discussed later.

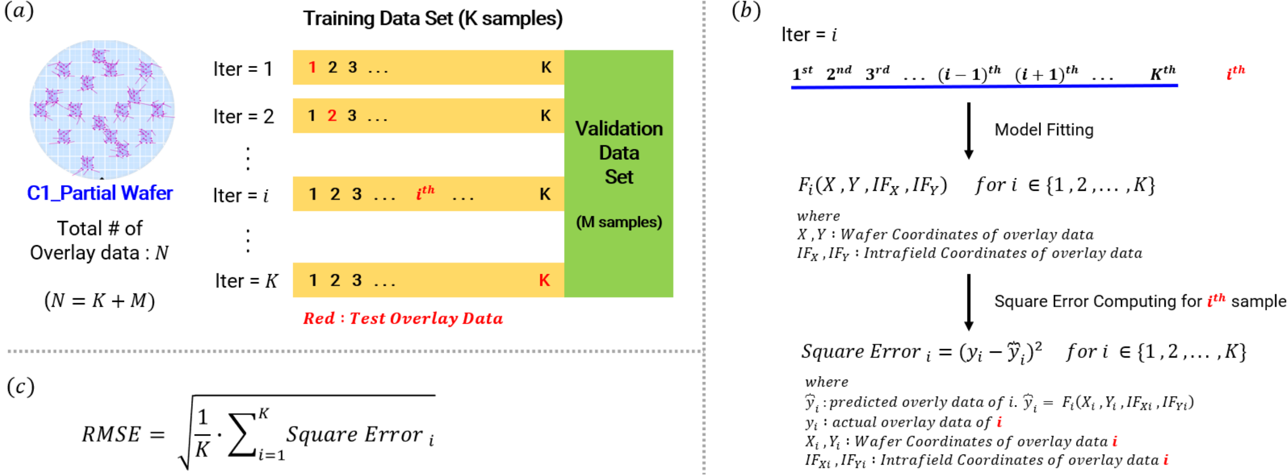


Figure 2-2. (a) The overlay data with N observations and its division. The training data set includes the K number of overlay data, and the validation data set has the M number of overlay data. Each iteration with one left overlay data has shown. (b) Process of the iteration. The training data set does model fitting except for one test data, and then the Square Error of the test sample is calculated. (c) Compute Square Error's $Root\ Mean\ Square\ Error$ ($RMSE$), which is an output of the CV for each overlay model.

3. THE RESULT ANALYSIS

We conducted the CV with nine wafer-based overlay models to check which overlay model has the best prediction capability and which model shows an overfitting problem. Figure 3-1 shows the CV result of the Chuck 1 and Chuck 2 Partial Wafers. The Y-axis indicates the average RMSE values of the X and Y directions, and X-axis shows the nine wafer-based overlay models used in CV. F1, shown in X-axis, is the intra-field linear model, which includes the magnification and the rotation, and the F3 model is the iHOPC15.

3.1 The CV Result Based on the Training Data Set

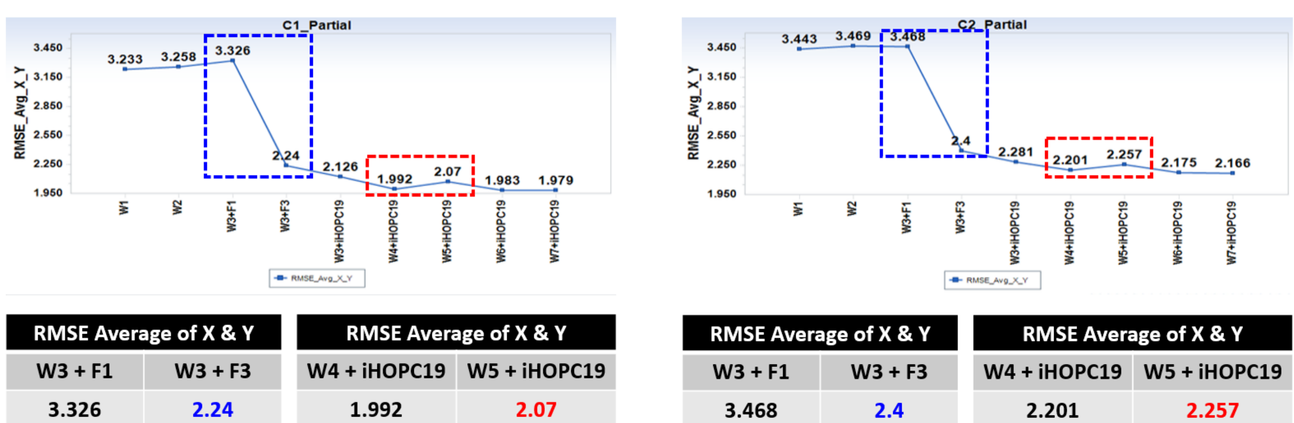
The most evident result in the graph is that the RMSE value of the W3+F3 model is dramatically decreased in comparison to the simpler model W3+F1. This means that the F3 model has many contributions to predicting the overlay of a new wafer, so manufacturers can benefit from using this model to control their overlay data based on the given overlay measurement map.

Meanwhile, the other two models to mention are W4+iHOPC19 and W5+iHOPC19. The RMSE value of the W5+iHOPC19 model has been slightly increased compared to the previous model, which means there is an overfitting issue in the W5+iHOPC19 model. In this case, using the W5+iHOPC19 model could worsen this Key Performance Index (KPI) due to this overfitting issue.

According to the result shown in Figure 3-1, the W6+iHOPC19 and W7+iHOPC19 are the best overlay models in this data set. However, the best overlay model should be determined based on many CV results calculated by each training data set. This can be achieved by selecting different overlay samples as a validation data set among all overlay measurement data for each CV process.

Figure 3-2 shows how the CV processes with distinct training data sets can be conducted. In the beginning, some overlay samples were selected as a validation data set, and the remaining overlay data were used for calculating RMSE, an output of the CV. It means that depending on which overlay samples are chosen as a validation data set, the training data set used for the CV could be different. We tested the CV 75 times with distinct training data sets based on this idea.

Figure 3-2 shows the CV results with distinct training data sets. In some cases, there was no overfitting issue. In most cases, however, a significant overfitting issue was found in W5+iHOPC19, W6+iHOPC19, and W7+iHOPC19 models. Interestingly, among the 75 CV results, none of the W3+iHOPC19 and W4+iHOPC19 models show any overfitting problem. Thus, the conclusion that the W3+iHOPC19 and W4+iHOPC19 would be the best overlay model based on this overlay measurement map can be made.



RMSE Average of X & Y	
W3 + F1	W3 + F3
3.326	2.24

RMSE Average of X & Y	
W4 + iHOPC19	W5 + iHOPC19
1.992	2.07

RMSE Average of X & Y	
W3 + F1	W3 + F3
3.468	2.4

RMSE Average of X & Y	
W4 + iHOPC19	W5 + iHOPC19
2.201	2.257

Figure 3-1. The CV result of two virtual wafers. The prediction capability is dramatically improved from the model W3+F3, while the W5+iHOPC19 model shows an overfitting issue in both virtual wafers.

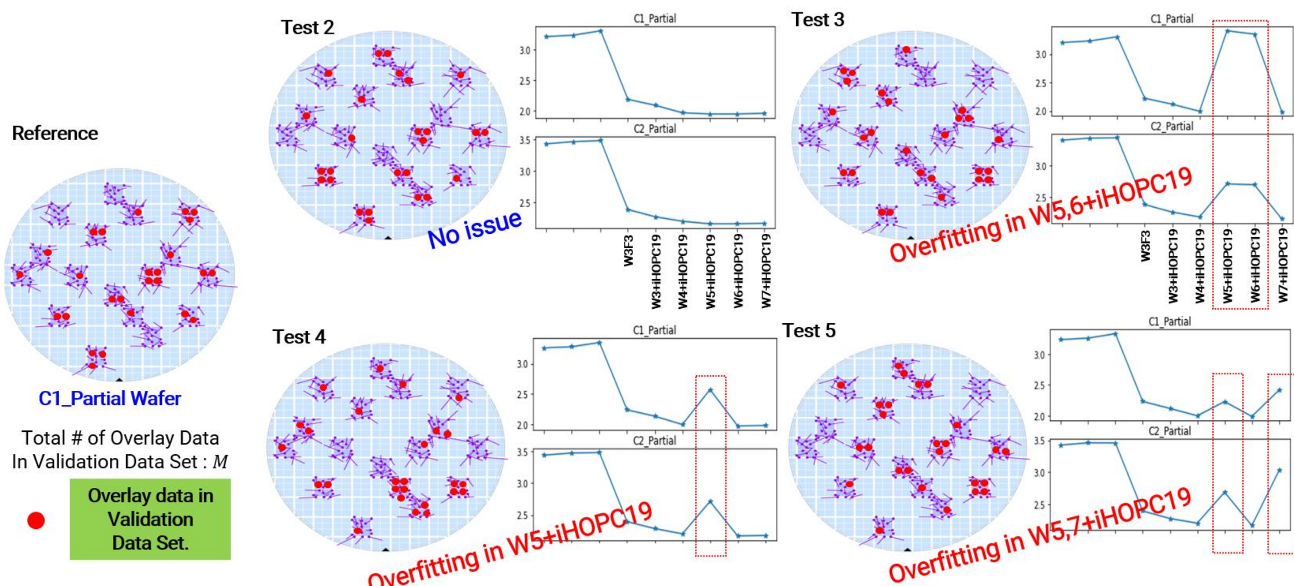


Figure 3-2. The CV results with each different training data set. There is an overfitting issue in W5, W6, and W7+iHOPC19 models, while W3 and W4+iHOPC19 have not shown any overfitting problem.

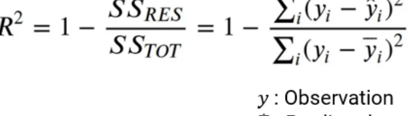
The CV result could be changed depending on which overlay measurement map we use. By using the CV, engineers have an opportunity to understand which overlay model would have the best prediction capability based on the overlay measurement map they use. That is the crucial concept of using the CV.

3.2 R Square Score Calculation Based on the Validation Data Set

Since the validation data set has not yet been used, it is time to apply the data set to calculate the R square scores of each overlay model for X and Y directions. In general regression, the R square score is always positive as the total Sum of Squares (SS_{TOT}) is always greater than the residual Sum of Squares (SS_{RES}). In our case, however, since the overlay samples in the validation data set were not used for fitting the overlay model, SS_{RES} could be greater than SS_{TOT}, which means that the R square score could be negative, as shown in Figure 3-3.

Figure 3-3 shows the R square score values for each overlay model. In the table, the R square score becomes stable from the model W3+F3, where the RMSE value starts to be stable.

Virtual Wafer	Direction	Overlay Model	RMSE	R^2 Score
C1_Partial	X	W1	3.807	-0.0729
C1_Partial	X	W3+F1	3.914	-0.2116
C1_Partial	X	W3+F3	2.699	0.3793
C1_Partial	X	W3+iHOPC19	2.492	0.4379
C1_Partial	X	W5+iHOPC19	2.373	0.5549
C1_Partial	X	W7+iHOPC19	2.258	0.5549
C1_Partial	Y	W1	2.659	-0.0198
C1_Partial	Y	W3+F1	2.739	-0.1098
C1_Partial	Y	W3+F3	1.782	0.7115
C1_Partial	Y	W3+iHOPC19	1.759	0.6777
C1_Partial	Y	W5+iHOPC19	1.768	0.6350
C1_Partial	Y	W7+iHOPC19	1.700	0.6347



$$R^2 = 1 - \frac{SS_{RES}}{SS_{TOT}} = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \bar{y}_i)^2}$$

y : Observation
$\hat{y}$: Predicted
$\bar{y}$: Mean

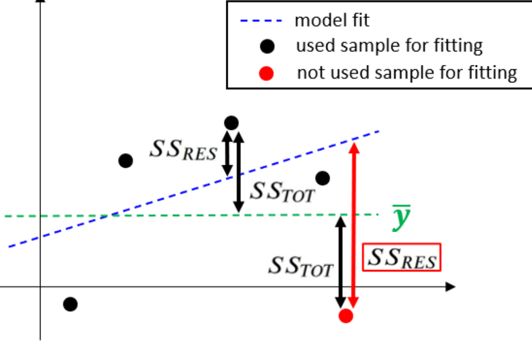


Figure 3-3. The result of R square score of overlay models. R square could be negative if the validation samples were not used for fitting.

4. CROSS-VALIDATION FOR CPE MODEL

So far, we have discussed how the CV works for wafer-based overlay models. Next, we will expand the idea of CV to the CPE overlay models and see how it works. Unlike the wafer-based overlay model, which considers all overlay measurement data in a wafer as a data set, the CPE model typically considers each field as a data set. Therefore, which data set needs to be used for the CV in CPE models should be defined first.

4.1 The Data Set Definition

There are several ways of defining the data set used for the CV for the CPE model. As we did earlier in this paper, defining a virtual field that can represent all fields in a wafer residual overlay map will be introduced. Figure 4-1 shows a residual overlay map of the Chuck 1 Partial Wafer using a specific wafer-fine overlay model. Since there are 17 fields on the residual overlay map, a virtual field that can represent those 17 fields is obtained by calculating the average of 17 fields point by point. This data set will be used for the CV for the CPE overlay model. Unlike the CV for a wafer-based model where the total number of overlay data is more than a hundred in general, the number of overlay samples in the data set for the CPE overlay model is comparatively small as they use a field-based data set. Thus, a validation data set was not considered for the CPE overlay model.

4.2 The CV Result Analysis for CPE Model

Figure 4-2 shows the result of the CV for the CPE models. The result can be divided into two sections, one is based on the residual overlay data of a low-order wafer-based overlay model, and the other is based on a relatively high-order

wafer-based overlay model's residual overlay data. It indicates that the CPE15 model would have better prediction capability than relatively low-order CPE models like CPE2 or CPE6 if the data set is based on a low-order wafer-based model like W1, as shown in Figure 4-2 (a). In contrast, the low-order CPE models like CPE2 or CPE6 would be better if a wafer-based overlay model order is relatively high, as shown in Figure 4-2 (b). Interestingly, the RMSE values of the CPE2 or CPE6 models considered the best model in part (b) are lower than that of the CPE15 model in part (a).

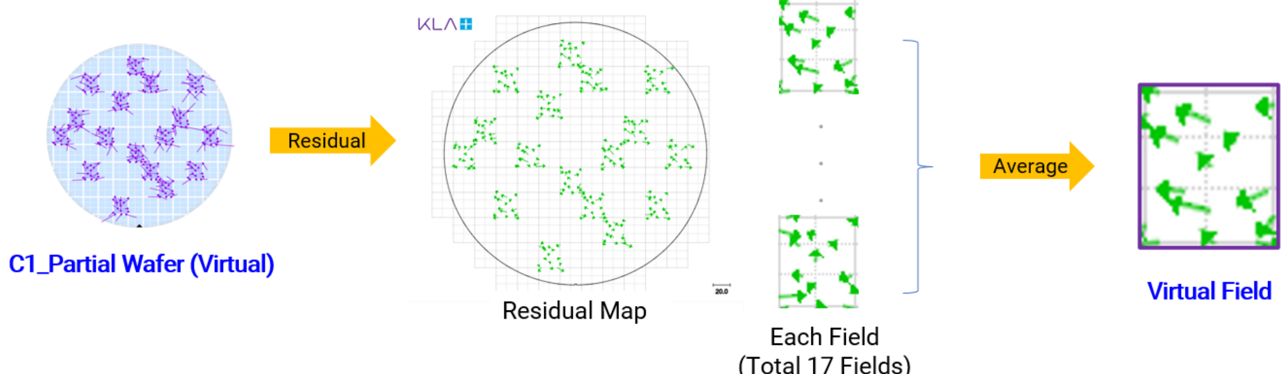
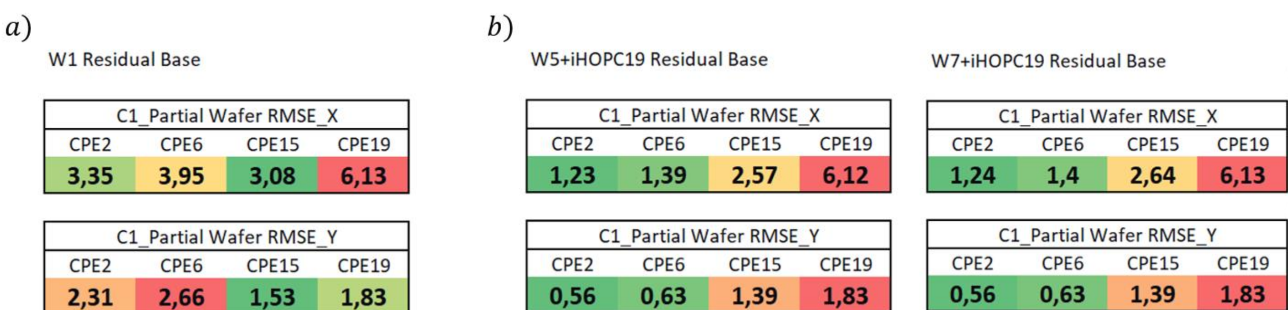


Figure 4-1. Defining the data set for the CV for CPE overlay model. The data set can be obtained by calculating the average of all fields on a residual overlay map point by point.



a)

W1 Residual Base

C1_Partial Wafer RMSE_X			
CPE2	CPE6	CPE15	CPE19
3,35	3,95	3,08	6,13

C1_Partial Wafer RMSE_Y			
CPE2	CPE6	CPE15	CPE19
2,31	2,66	1,53	1,83

b)

W5+iHOPC19 Residual Base

C1_Partial Wafer RMSE_X			
CPE2	CPE6	CPE15	CPE19
1,23	1,39	2,57	6,12

C1_Partial Wafer RMSE_Y			
CPE2	CPE6	CPE15	CPE19
0,56	0,63	1,39	1,83

W7+iHOPC19 Residual Base

C1_Partial Wafer RMSE_X			
CPE2	CPE6	CPE15	CPE19
1,24	1,4	2,64	6,13

C1_Partial Wafer RMSE_Y			
CPE2	CPE6	CPE15	CPE19
0,56	0,63	1,39	1,83

Figure 4-2. (a) The CV result of CPE models based on the W1 wafer-based overlay model. The CPE15 model shows better RMSE than other CPE models. (b) High-order wafer-based models CV result of CPE models. Unlike case (a), the low-order CPE models like CPE2 or CPE6 have better RMSE than CPE15 or CPE19.

4.3 Consideration of Fallback and Interpolation for a Data Set

The overlay map in this paper has enough overlay samples to apply a high-order CPE overlay model in each field. In case some sparse overlay map does not have enough overlay samples, there are two ways of applying the CV. The first method is to apply fallback, where the less-order CPE model would be applied if there is not enough overlay data in a field. The advantage of using fallback is that it can be applied to any field no matter how many overlay samples each field has. Hence, the CV process could be applied to a sparse overlay measurement map. Another method is using interpolation or extrapolation, which can predict overlay data of unmeasured points to generate virtual overlay data in a field that does not have enough overlay measurements. Virtual overlay samples can be added to a sparse overlay map to look like a fully measured map, where any CPE overlay models can be applied.

The use case of these two methods is worthy of consideration.

5. FURTHER DISCUSSION

In this paper, we have studied CV validation and how it can be applied to wafer-based and CPE models. The basic idea of using the CV is to determine which overlay model has good prediction capability for a new wafer's overlay. However, unlike an actual production process flow where each lot is heavily related to one another by giving and taking feedback within APC, the CV process is more like a discontinuous process flow where each lot and wafer has nothing to do with each other. In other words, the CV can be applied between wafer and wafer and returns each wafer's RMSE as their output, so there is no connection or interaction between them. That is why we defined some virtual wafers by calculating the average of actual wafers to reflect their connection as best as possible. One of the subsequent research topics of the CV could be how we consider an inter-lot feedback system like an actual production process flow.

REFERENCES

[1] Dana Klein, Eran Amit, Guy Cohen, Nuriel Amir, Michael Har-Zvi, Chin-Chou Kevin Huang, Ramkumar Karur-Shanmugam, Bill Pierson, Cindy Kato, Hiroyuki Kurita, "Quality metric for accurate overlay control in <20nm nodes," Proc. SPIE 8681, Metrology, Inspection, and Process Control for Microlithography XXVII, 86811J (18 April 2013); https://doi.org/10.1117/12.2011454

[2] Clemens Utzny, "Splendidly blended: a machine learning set up for CDU control ," Proc. SPIE 10446, 33rd European Mask and Lithography Conference, 104460N (28 September 2017); https://doi.org/10.1117/12.2279430

[3] Nyan Aung, Woong Jae Chung, Pavan Samudrala, Haiyong Gao, Wenle Gao, Darius Brown, Guanchen He, Bono Park, Michael Hsieh, Xueli Hao, Yen-Jen Chen, Yue Zhou, DeNeil Park, Karsten Gutjahr, Ian Krumanocker, Kevin Jock, Juan Manuel Gomez, "Overlay control for 7nm technology node and beyond," Proc. SPIE 10587, Optical Microlithography XXXI, 105870A (20 March 2018); https://doi.org/10.1117/12.2295828

[4] Berrar, Daniel. "Cross-Validation." (2019): 542-545.

[5] Viana, F.A.C., Haftka, R.T. & Steffen, V. Multiple surrogates: how cross-validation errors can help us to obtain the best predictor. Struct Multidisc Optim 39, 439–457 (2009). https://doi.org/10.1007/s00158-008-0338-0

YOLOv8 for Defect Inspection of Hexagonal Directed Self-Assembly Patterns: A Data-Centric Approach

Enrique Dehaerne[a,b], Bappaditya Dey[b], Hossein Esfandiar[c,b], Lander Verstraete[b], Hyo Seon Suh[b], Sandip Halder[b], and Stefan De Gendt[a,b]

[a]Faculty of Science, KU Leuven, 3001 Leuven, Belgium
[b]Interuniversity Microelectronics Centre (imec), 3001 Leuven, Belgium
[c]Dept. of Chemistry, Materials and Chemical Engineering, Giulio Natta; Politecnico di Milano; I-20133 Milano; Italy

ABSTRACT

Shrinking pattern dimensions leads to an increased variety of defect types in semiconductor devices. This has spurred innovation in patterning approaches such as Directed Self-Assembly (DSA) for which no traditional, automatic defect inspection software exists. Machine Learning-based SEM image analysis has become an increasingly popular research topic for defect inspection with supervised ML models often showing the best performance. However, little research has been done on obtaining a dataset with high-quality labels for these supervised models. In this work, we propose a method for obtaining coherent and complete labels for a dataset of hexagonal contact hole DSA patterns while requiring minimal quality control effort from a DSA expert. We show that YOLOv8, a state-of-the-art neural network, achieves defect detection precisions of more than 0.9 mAP on our final dataset which best reflects DSA expert defect labeling expectations. We discuss the strengths and limitations of our proposed labeling approach and suggest directions for future work in data-centric ML-based defect inspection.

Keywords: Semiconductor defect inspection, machine learning, deep learning, object detection, YOLOv8, supervised learning, data-centric learning, directed self-assembly

1. INTRODUCTION

Shrinking semiconductor pattern dimensions has led to increased manufacturing complexity. This complexity has led to a greater number of stochastic defects to inspect. Scanning Electron Microscopy (SEM) is a workhorse fab tool since it can image at high resolution relatively fast. However, SEM images are inherently noisy. Capturing too many frames to reduce this noise can cost time and can be destructive for materials to be inspected such as thin resists. Machine Learning (ML)-based defect detection methods for SEM images are becoming a popular research topic due to their ability to deal with noise better than conventional image processing algorithms.[1] Additionally, ML-based methods are more easily adaptable to pattern changes (e.g. CD/Pitch variation) and imaging conditions (contrast, field-of-view, etc.). Supervised learning, where an ML model learns by comparing its predictions against (pseudo-)ground truth labels (generally defined by a human expert), usually leads to the best performance. However, the performance of supervised learning depends on the quality of the labels provided by a human labeler.

In this study, we apply the latest object detection model of the popular YOLO[2] (You Only Look Once) family, YOLOv8,[3] to detect challenging defects in Hexagonal Directed Self-Assembly (DSA) patterns. Hexagonal DSA patterns make for an interesting case study as they can contain unique defects that traditional defect inspection software cannot detect or classify. We investigate the effect of variance between human labelers. We consider three labeled datasets, each containing the same SEM images labeled by different human labelers. We propose a data-centric approach to combine labels from multiple labelers and align them with DSA expert expectations. We show that training YOLOv8 models on this final, combined dataset results in detection performance that best matches DSA expert expectations.

Send correspondence to Enrique Dehaerne, e-mail: enrique.dehaerne@imec.be

38th European Mask and Lithography Conference (EMLC 2023), edited by
Uwe F.W. Behringer, Jo Finders, Proc. of SPIE Vol. 12802, 128020R
© 2023 SPIE · 0277-786X · doi: 10.1117/12.2675573

2. RELATED WORK

2.1 Data-centric ML

Data-centric ML refers to improving an ML system's performance by enhancing the dataset used to train a given model.[4] This stands in contrast to the model-centric ML approach which, for a given dataset that is assumed to be fixed, makes changes to the ML model to improve the performance of the system. Broadly, methods for data collection/curation,[5,6] preparation/augmentation,[7,8] and labeling[9,10] can all be considered as data-centric improvements for ML systems.

2.2 Semiconductor defect inspection

Traditionally, semiconductor defect inspection uses hand-crafted image-processing algorithms.[11–13] While these algorithms typically perform well for specific inspection scenarios, they struggle to adapt to variations in patterns and imaging conditions. Defect inspection for DSA patterning has been studied using more traditional techniques.[14–16] To the best of our knowledge, no ML-based approaches for defect inspection of DSA patterns have been published. ML-based SEM defect detection methods, especially Convolutional Neural Network (CNN) models, are becoming popular due to their improved robustness to these variations. Two main paradigms for training defect inspection models exist, supervised[1,17–19] and unsupervised[20–22] learning. Supervised learning, where an ML model learns by comparing its predictions to labels provided by a human expert, usually leads to the best performance for defect detection and classification. Note, the performance of supervised learning depends on the quality of the labels. The majority of data-centric works in the semiconductor defect inspection domain have focused on data augmentation techniques.[23,24] In this study, we propose improving label quality as another method for improving the performance of CNN-based semiconductor defect detection systems.

3. HEXAGONAL DIRECTED SELF-ASSEMBLY SEM IMAGE DATASET

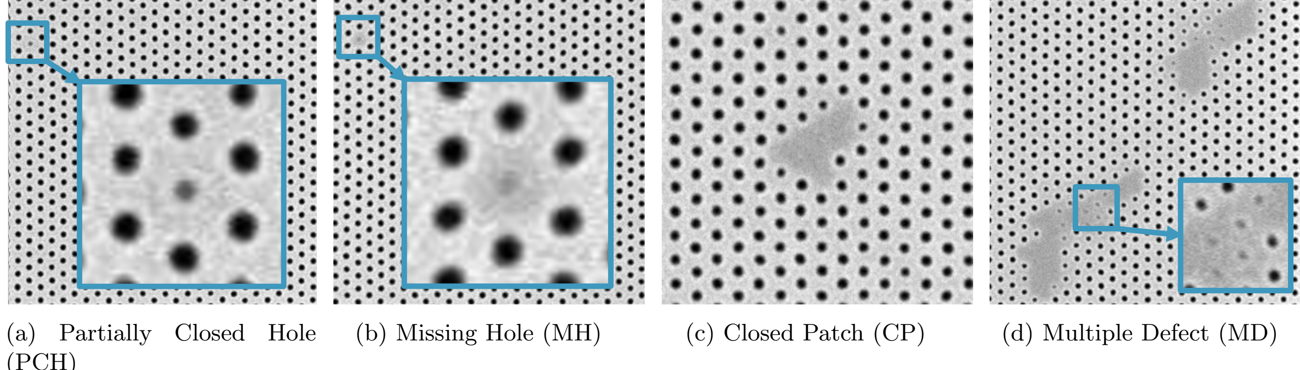
For this research work, we collected a novel dataset of hexagonal DSA SEM images with defects. DSA is a patterning technique that uses the inherent period of block-copolymer (BCP) materials to define line/space or hexagonal hole patterns from the bottom up.[25,26] To generate well-aligned patterns, a guiding pattern for the BCP materials is needed. While this guiding pattern is typically created by conventional lithography techniques, the final pattern characteristics are defined by the self-assembled BCP materials. This allows DSA to overcome stochastic issues commonly associated with EUV lithography. However, DSA is known to have specific defects that are otherwise not observed in conventional lithography patterns. These DSA specific defects are associated with the dynamic processes underlaying the self-assembly process. Typical DSA specific defects are dislocations in line/space patterns, which represent a kinetically trapped metastable structure.[16,27] Similarly, HEXagonal Contact Hole (HEXCH) arrays may suffer from misaligned domains and even non-vertically oriented structures. When etching the latter into the underlying hard mask, closed holes may be observed.

Our novel dataset consists of SEM images of HEXCH DSA pattern defects after etch into a SiN layer. In these images, we localize (in the form of bounding boxes) and classify three different classes of defects. These defect classes are Partially Closed Holes (PCH), Missing Holes (MH), and Closed Patches (CP). PCH and MH defects each relate to one hole while a CP defect relates to a contiguous area of closed holes. As DSA HEXCH patterns are not yet routinely considered in the semiconductor industry, defects in these patterns are not covered in traditional inspection methodology. The DSA dataset images were collected by a DSA patterning expert. The expert inspected each image and partitioned the collection into four sub-collections based on the defects shown in each image. The names given to these partitions correspond to the three defect classes listed above (PCH, MH, and CP) as well as a Multiple Defect (MD) sub-collection. These MD images contained more than one defect of one or more different defect classes. Figure 1 shows examples of defects of these four partitions. The full dataset organized into these four sub-collections was given to three different labelers with instructions to draw and classify bounding boxes that best capture the extent of each defect in each image. In the rest of this paper we refer to the labelers as *labeler A*, *labeler B*, and *labeler C*. Note that this initial partitioning of the dataset is not necessary for the labeling methodology proposed in Section 4.

The dataset was randomly split into train, validation, and test subsets with a ratio of 70:15:15, respectively. The image-level statistics of the dataset are provided in Table 1. Each SEM image has a resolution of 1000×1000 pixels.

Table 1: SEM image counts for each split and partition of the HEXCH DSA dataset.

Partition	Train	Validation	Test	All
Partially Closed Hole (PCH)	73	13	18	104
Missing Hole (MH)	29	8	5	42
Closed Patch (CP)	72	13	19	104
Multiple Defect (MD)	63	17	9	89
Total	237	51	51	339



(a) Partially Closed Hole (PCH) (b) Missing Hole (MH) (c) Closed Patch (CP) (d) Multiple Defect (MD)

Figure 1: Defect examples for each raw image partition of the full hexagonal DSA dataset. The images are cropped (and zoomed in where necessary).

4. METHODOLOGY

4.1 Semi-automated labeler consensus

To mitigate individual labeler errors and biases, we propose a semi-automated method for combining labels from multiple labelers. A naive method for combining multiple labeled datasets would be to have a DSA expert manually compare labels for each image and remove or modify each label individually. This naive method is inefficient for two main reasons. First, a majority of labels will most likely have consensus in the sense that many labels between multiple labelers overlap to a high degree and are classified as the same class. Second, misunderstanding labeling instructions may lead to common patterns of erroneous labels. Oftentimes, these common patterns can be easily identifiable and automatically corrected without requiring a DSA expert's manual intervention. To mitigate these inefficiencies, respectively, our proposed method includes: (i) using voting for reclassification and Weighted-Box-Fusion (WBF)[28] for bounding box extent to combine overlapping labels and (ii) programmatically correcting common labeling error patterns as identified by a DSA expert. In this way, the proposed method only requires DSA expert intervention two times: (a) breaking ties between labelers during reclassification voting and (b) identifying common erroneous label patterns.

Figure 2 shows a flowchart of the proposed labeling procedure. First, the raw SEM image dataset is labeled by

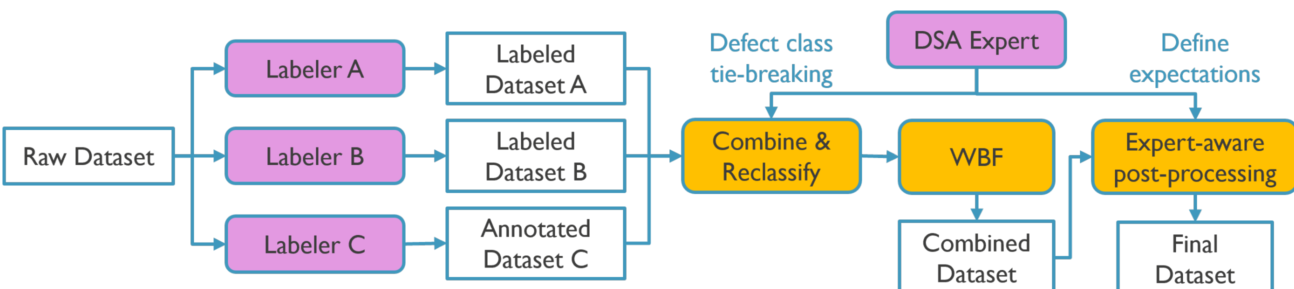
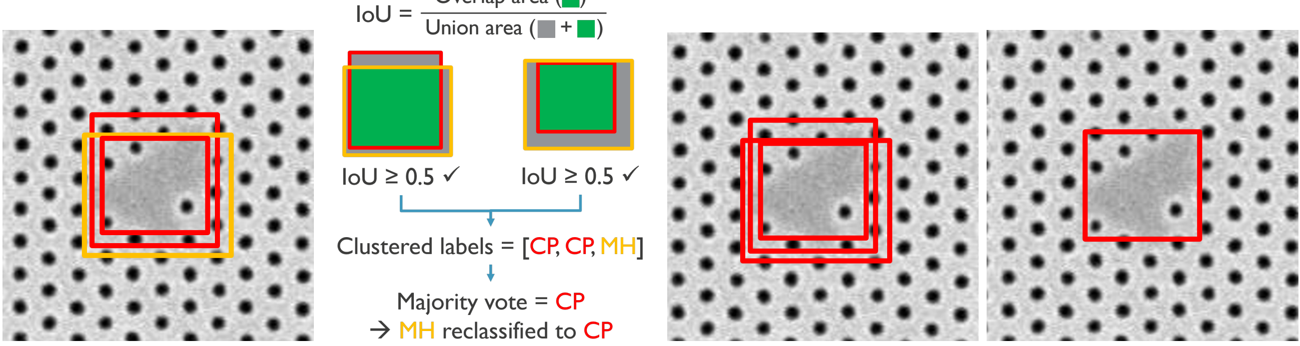


Figure 2: Flowchart of the proposed labeling and consensus methodology.



(a) Example labels for a defect. (b) IoU calculation for overlapping labels. (c) IoU calculation for overlapping labels. (d) Example labels after WBF.

Figure 3: An example scenario with three overlapping labels, two CP labels and one MH label, (3a). Using the orange label as a randomly selected starting label, the labels are clustered together based on having IoU values greater or equal to 0.5 (3b). The most common class in this label cluster is CP so the MH label is changed to a CP label (3b and 3c). The cluster of labels is then fused into a single label using the WBF algorithm (3d).

multiple labelers (three labelers in the case of this study) to obtain multiple labeled datasets. Note that this part does not require any consensus between the labelers which means each labeler can work independently and in parallel, optimizing labeling speed. The labels of all the labelers are then combined on all corresponding images. Standard labeling software was used by the labelers. LabelImg[29] was used by labelers B and C. Coco-annotator[30] was used by labeler A.

Second, overlapping clusters of labels are reclassified based on voting between these labels. A cluster of labels is formed by selecting a label and adding all other labels in the image that have an Intersection-over-Union (IoU) value greater or equal to 0.5 to a cluster along with the selected label. This is similar to how the WBF algorithm clusters labels to be fused together. Within a cluster, a vote is performed using each label's class. All labels in the cluster are then reclassified to the class with the most votes. Reclassification vote ties are resolved manually by a DSA expert.

The (possibly reclassified) overlapping clusters of labels are fused into one label via WBF. The WBF algorithm expects as input a list of bounding boxes, each with corresponding labels and confidence scores. Compared to other bounding-box filtering methods such as Non-Maximum Suppression (NMS), WBF does not rely on bounding-boxes having varying confidence scores. This means that we can assign the same confidence score, 1, to all labels since we do not know the confidence of a label. Additionally, WBF has the desirable property that it forms a new bounding box using properties from all boxes in a cluster rather than simply throwing away boxes such as with NMS. After WBF is applied, a *combined dataset* is obtained which can more easily be inspected by a DSA expert than the raw combined labels of all labelers. Figure 3 shows an example of how labels are analyzed for overlap, reclassified via voting, and fused using WBF. The combined dataset is obtained after WBF has been applied to all images of the dataset results in the *combined* dataset.

While it is possible to let the DSA expert thoroughly inspect and correct each label instance of the *combined* dataset, we find that a majority of the labeling errors follow a pattern that the DSA expert can identify by only inspecting a few labeled image samples from the combined dataset. We also find that these common label error patterns can be corrected using automatic post-processing, mitigating the need for the DSA expert to correct these errors manually. The exact required post-processing steps will be different for different labelers and different datasets. The post-processing steps used in our study are discussed in Section 5.1. If very few errors are found, it will most likely be faster to correct these individually. As with all steps in the proposed methodology, using post-processing for error correction scales better with dataset size and erroneous label counts compared to manual label correction.

Following the methodology described above, three datasets labeled by individual labelers (*labelers A*, *B*, and *C*) and two labeled datasets from combining and processing these three datasets (*combined* and *final*,

respectively) were obtained. The final dataset is considered to be the labeled dataset most similar to the DSA expert's ideal dataset. The defect detection performance results on the final dataset are therefore considered to be the most important.

4.2 YOLOv8 defect detection model

The original YOLO model[2] was developed to achieve a good trade-off between detection performance and inference speed. This is a very desirable characteristic for industrial image-based defect inspection applications where throughput is important. YOLOv8[3] is a recent YOLO-style, single-stage object detection model that has been shown to outperform previous YOLO models in terms of mean Average Precision (mAP) on the COCO dataset.[31] YOLOv8 comes in five different size variants: *nano*, *small*, *medium*, large, and *extra-large*. In our experiments, we train, validate, and test each variant on the various labeled datasets obtained from the labeling methodology proposed in the previous subsection. We report the average test results over all of the variants to obtain our final AP results. No consistent differences between model size variants were observed for the experiments conducted.

All experiments were run on a Tesla V100 GPU with 32GB of VRAM unless mentioned otherwise. Images were resized to the closest multiple of 32, 992×992, since YOLOv8 models have a maximum stride of 32. A batch size of 32 was used for the *nano*, *small*, and *medium* YOLOv8 model variants. A batch size of 16 was used for the *large* and *extra-large* variants. All models were trained for a maximum of 200 epochs. The checkpoint that achieves the best fitness value ($fitness = 0.1 \times mAP@0.5 + 0.9 \times mAP@0.5 : 0.95$), as calculated at the end of each epoch, was used for evaluation. The training for all experiments was automatically stopped if the fitness value did not improve after 50 epochs. All other hyperparameters were kept the same as the default settings of the YOLOv8 GitHub repository.[3] Most training runs were completed within 2 hours.

5. RESULTS & DISCUSSION

5.1 Differences in labeled datasets

Figure 4 shows examples of prominent labeling behavior differences between the three labelers. The labeling differences in each row and how these differences were handled are described below:

1. The first row shows labeler A missing or not considering certain holes to be partially closed when labeler B does. Combining labels from all the labelers ensures that as few defects as possible are missed in the final labeled dataset. The first row also shows labeler C wrongly classifying all defects in the MD partition as a separate MD defect class. This is corrected through class voting. For example, in this particular instance, the label from labeler C would be reclassified as a CP since it is the majority vote when considering the overlapping labels from the three labelers.

2. The exact extent of closed patches bounding boxes is different between the three labels in the second row. This is solved by using WBF to fuse the labels into a single label of the same class that has an extent somewhere between all the overlapping labels.

3. The third row shows labelers B and C did not consider labeling partially closed and missing holes instances inside of a closed patch instance separately, while labeler A labeled all of these instances separately. While the image features captured by labeler A's labels may very well reflect true MH or PCH defect instances, the DSA expert preferred B and C's labeling. Therefore, for the final dataset, all combined image labels were processed to remove MH and PCH labels within CP labels.

4. The last row again shows some of the labeling differences from the previous row but also shows irregularly shaped closed patches that were broken up into multiple bounding boxes by each labeler differently. While these multiple boxes more precisely encapsulate CP features without capturing non-defective background features, the DSA expert preferred them to be labeled as one CP instead of multiple, separate CPs. For the final dataset, combined image labels were processed to combine overlapping CPs (in this case using an IoU threshold of more than zero) into the smallest box that completely covers all overlapping CPs labels. For this post-processing step, we use an IoU threshold of 0.0001 since most of these adjacent CP boxes barely overlap.

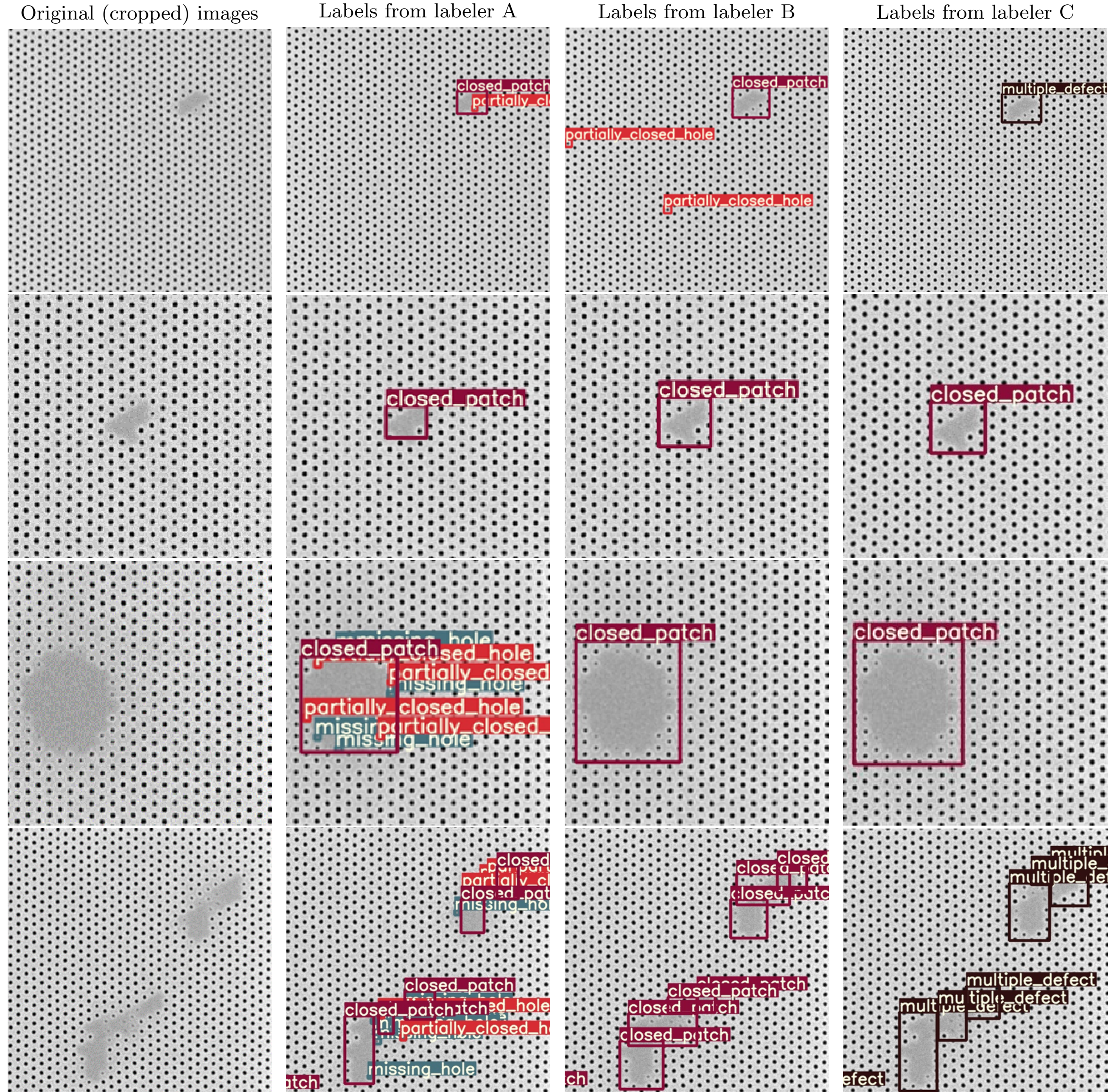
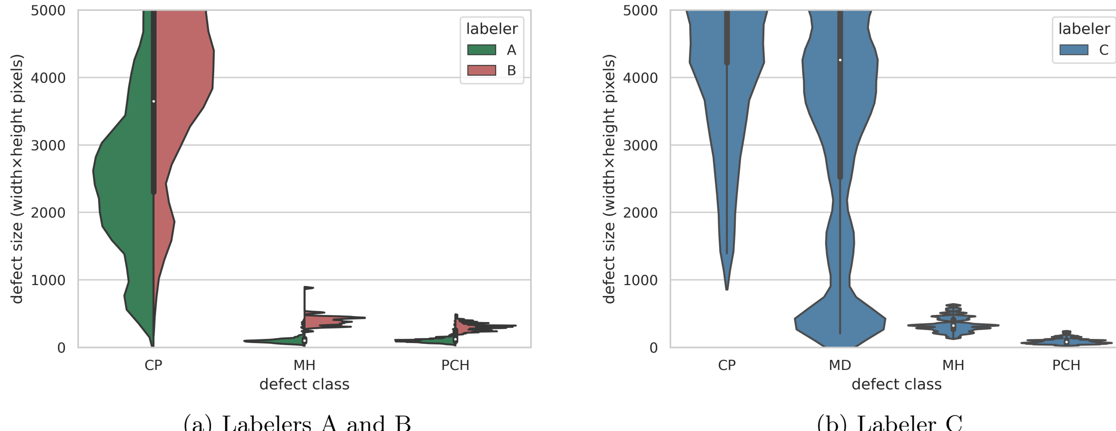


Figure 4: Examples of labeling differences between labelers A, B, and C.



(a) Labelers A and B (b) Labeler C

Figure 5: Violin plots for the defect size of labels of the three individually labeled datasets. Each violin is scaled to have the same width. The plots are cut off at a defect size of 5000 to improve readability.

After combining and labeling class voting, MD labels that did not overlap with labels from labelers A or B still remained. As can be seen in Figure 5, the sizes of CP, MH, and PCH labels for labeler C were particularly well separable. Therefore, the remaining MD labels were reclassified to be CPs if their size was larger than the largest MH or PCH label from labelers C. Otherwise, the remaining MDs were reclassified to be PCHs. This was done because the sizes of the remaining MD labels were much more similar to the average size of labeler C's PCH labels. It can be assumed that PCH labels are more easily to be accidentally missed by labelers than MH labels which further motivated the choice to reclassify the remaining MDs as PCHs instead of MHs.

Figure 6 shows corresponding combined and final dataset labels for the examples shown in Figure 4. It can be seen that the combined dataset contains far more labels than the final dataset. This is confirmed in Figure 7a which shows label count statistics for all images in each labeled dataset. Figure 7b shows the counts for all images except for MD images. The differences in label counts between datasets in Figure 7a are greater than those in Figure 7b which shows that most differences in labeling happen in MD images. We investigate the effects of training and evaluating with/without MD images, which have the highest likelihood to have erroneous labels due to the disagreement between labelers, in the next subsection.

5.2 YOLOv8 defect detection performance

First, we investigate the performance of YOLOv8 models trained and tested only on the individual labeler datasets. Table 2 shows the results of all combinations of train and test datasets from the three labelers. These results show that models trained and evaluated on data from the same labeler (A or B, not C due to the bad classification of MD labels) perform relatively well. Labeler C is an exception for this due to the model's classification confusion due to the erroneous MD labels. This behavior can be seen in the last row of Figure 8. The results from models trained on data from one labeler and tested on another labeler are significantly worse (again, when ignoring effects from MD labels). This shows that differences in labeling have a significant impact on model performance. The biggest differences are in the cross-labeler results involving labeler A since their label count was far greater than labelers B and C (see Figure 7a) which led to many false positive predictions compared to labeler B or C's test labels. This behavior can also be seen in Figure 8.

Second, we investigate the performance of models trained and/or tested on the combined and final datasets. Table 3 shows the test results for all of these experiments. Similar to Table 2, we find that the best results come from models trained and evaluated on the same labels. The models trained on the final training data achieved the best result on the target test data, the test final dataset, of 0.919 mAP. This is the third-best overall mAP score achieved. The best overall mAP of 0.944 is achieved by the models trained on the final dataset without MD images and tested on the same data distribution. This suggests that our labeling combination and processing method is more coherent and complete than relying on individual labelers. The MD images are the

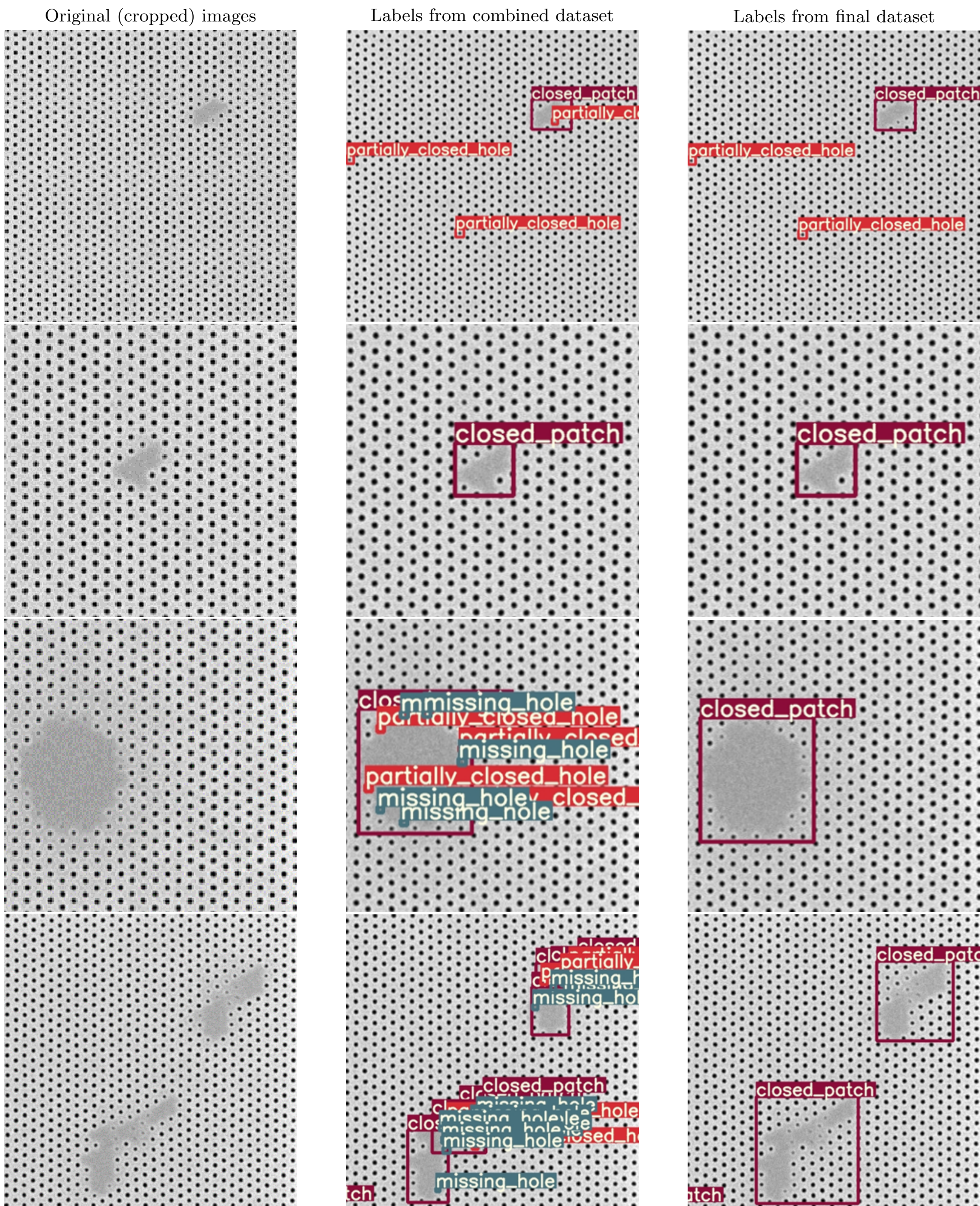
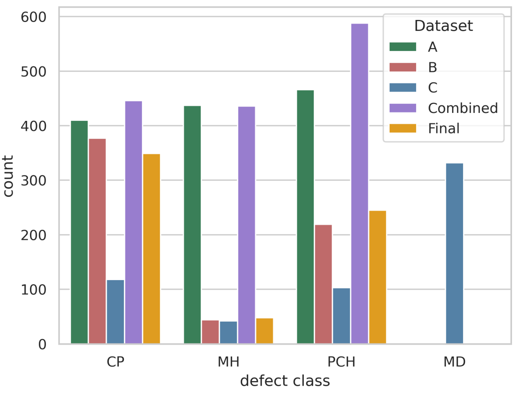
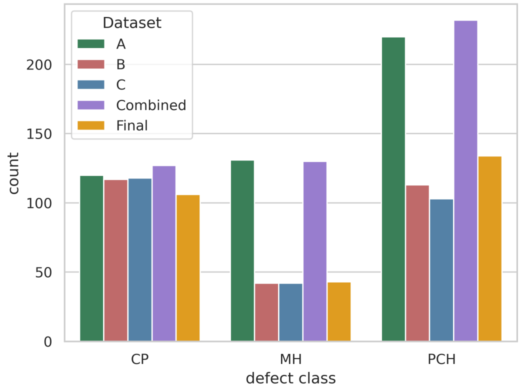


Figure 6: Examples of labeling differences between the combined and final datasets.



(a) All images



(b) All images except for MD images

Figure 7: Label counts for each labeled dataset.

Table 2: Per-class AP and mAP results for models trained and tested on datasets labeled by individual labelers. These results are the average of all YOLOv8 size variants. The first three rows of results use models that were trained and validated using the labels of one labeler and tested on the labels of the same labeler. The following rows use labels from different labelers for training and testing.

Labeler		**AP (@0.5 IoU)**				
train	**test**	**CP**	**MH**	**PCH**	**MD**	**All (mAP)**
A	A	0.883	0.873	0.769	–	0.842
B	B	0.986	0.968	0.765	–	0.906
C	C	0.421	0.716	0.213	0.576	0.482
A	B	0.632	0.072	0.020	–	0.260
	C	0.396	0.052	0.231	0.0	0.226
B	A	0.504	0.096	0.020	–	0.207
	C	0.975	0.796	0.186	0.0	0.652
C	A	0.418	0.079	0.297	–	0.265
	B	0.975	0.804	0.195	–	0.658

most challenging which explains why the test results were better when the models are tested without them. Interestingly, however, models trained on the final dataset without MD images outperformed models trained with MD images for the CP and MH defect classes when tested using MD images (see red values in Figure 3). This suggests that the final labels for the MD images are still a bit noisy. On the other hand, the improved performance of the PCH class when trained on MD images suggests that the optimal training dataset for the models should have a balance between label quality and quantity.

We find that the performance of PCH detection, the most challenging defect class overall, improves most when trained on final dataset labels compared to training on individual labeler data. We also find that the performance of PCH detection increases significantly when no MD images are considered for testing. This could be indicative of false positive PCH labels in the final dataset (especially in MD images) due to the somewhat ambiguous characteristic of PCH defects. An example of such an erroneous label could be the PCH label just below the center of the input image from Figure 8 where only labeler B, potentially erroneously, labeled this defect. Figure 9 shows that models trained on the combined or final datasets did not predict this defect at all.

The models trained on labeler A's dataset perform decently well on the combined test dataset and poorly on the final dataset. The opposite is true for models trained on labeler B and C's datasets which perform better on the final dataset than the combined dataset. This is because many of the labels in the combined dataset come exclusively from labeler A.

Applying the same post-processing steps used to create the final dataset to the predictions of the model

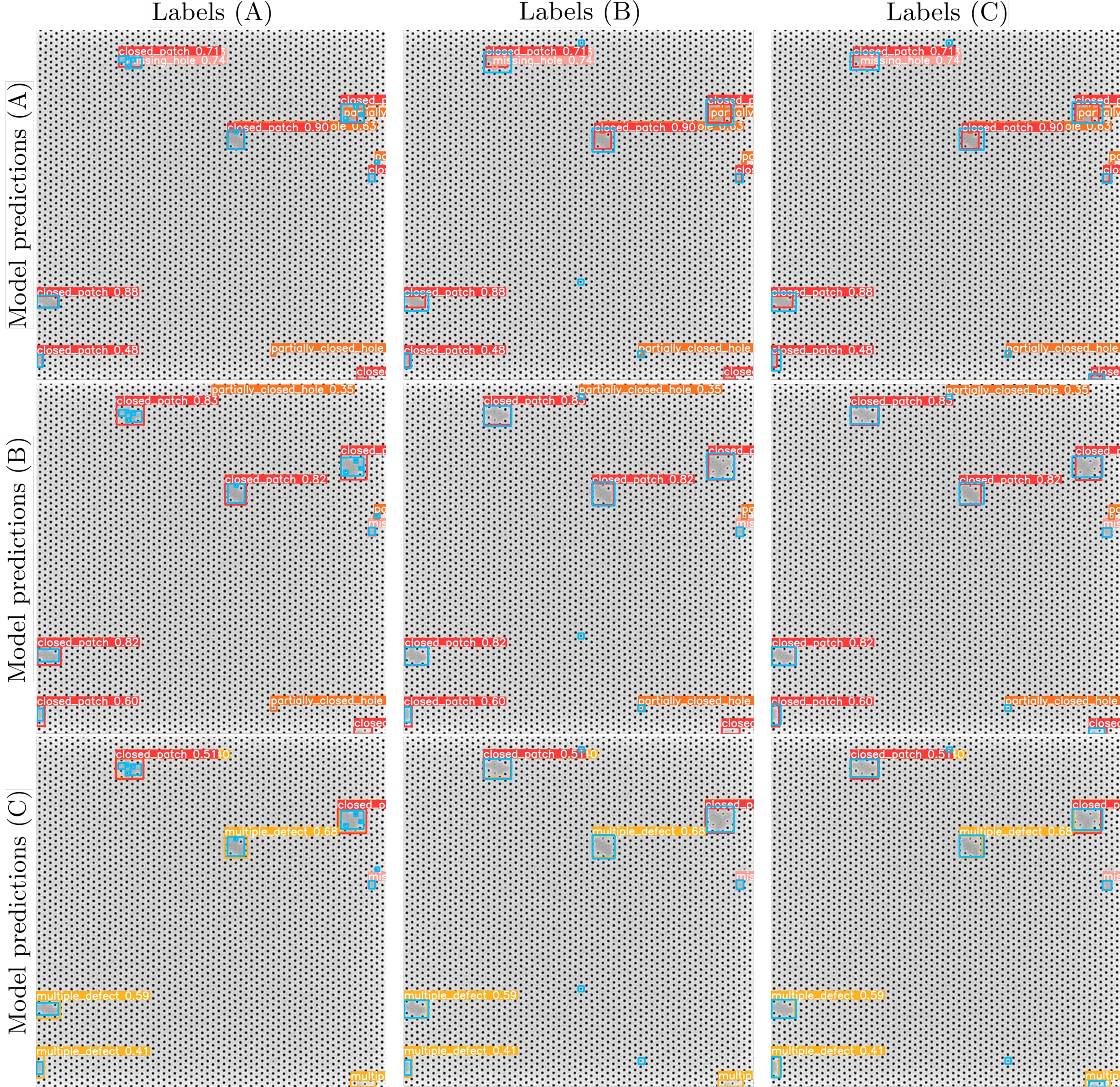


Figure 8: Prediction examples on a test image for YOLOv8m (medium) models trained on the labels from labelers A, B, and C (top row, middle row, and bottom row, respectively). Displayed predictions include confidence scores. Each prediction example is overlayed with labels of the test image from labelers A, B, and C (left column, middle column, and right column, respectively). The labels are shown in blue without class information to improve visibility. Best viewed digitally and zoomed-in for the best visibility of the prediction and label details.

Table 3: Per-class AP and mAP results for models trained and/or tested on the combined and/or final datasets. These results are the average of all YOLOv8 size variants. Post-processing refers to applying the "expert-aware post-processing" rules from Figure 2 to model predictions before comparison with test labels. Blue and red values are the best test values for combined and final data, respectively.

Dataset		**AP (@0.5 IoU)**			
train	**test**	**CP**	**MH**	**PCH**	**All (mAP)**
A	Combined	0.856	0.872	0.655	0.794
	Final	0.810	0.072	0.268	0.383
B	Combined	0.892	0.190	0.235	0.438
	Final	0.976	0.968	0.383	0.775
C	Combined	0.911	0.125	0.290	0.441
	Final	0.931	0.792	0.460	0.728
Combined	Combined	0.945	0.904	0.795	0.881
	Final	0.944	0.222	0.550	0.572
	Final (+ post-processing)	0.974	0.810	0.578	0.787
	Final (no MDs + post-processing)	0.986	0.911	0.719	0.872
Final	Final	0.977	0.984	0.797	0.919
	Final (no MDs)	0.952	0.980	0.844	0.925
Final (no MDs)	Final	0.992	0.995	0.695	0.892
	Final (no MDs)	0.995	0.984	0.853	0.944

trained on the combined dataset significantly improves the performance when tested on the final test dataset. However, this improved performance is still noticeably lower than the results obtained by the models trained on the final dataset itself. Post-prediction-processing in this manner could still be useful in a dynamic environment where new post-processing rules are regularly introduced and re-training of the model is not wanted.

6. LIMITATIONS & FUTURE WORK

In our proposed labeling combination method, we took a liberal approach to combining labels in the sense that we assume all labels are valid labels for defects. This minimizes false negatives. However, this means that any false positive labels from individual labelers will also be included in the combined dataset. This can be an issue for defects that can sometimes be similar to the background pattern such as PCH as discussed in the previous section. Future work could experiment with more conservative label combination approaches or automatic erroneous label detection and removal methods.

Another limiting assumption we make is that each image is labeled by multiple labelers. This is inefficient in the sense that a lot of duplicate work is performed, especially for labeling "obvious" defects. This could be improved in future work given access to larger datasets by (i) having different partitions of the dataset labeled only once by different labelers (see Figure 10a), (ii) training separate models on each partition (see also Figure 10a), (iii) using these models to predict pseudo-labels on the other image partitions it was not trained on (see Figure 10b), and (iv) applying a label combination algorithm on all the manual- and pseudo-labels for each image in the dataset. This allows larger datasets to be labeled using similar amounts of manual labeling effort while benefitting from the coherency and completeness of our proposed label combination method. Future work should take care to identify major labeling errors (such as the MD misclassifications by labeler C which led to poor performance even when tested on their own dataset as shown in Table 2) and ensure that each labeler labels enough images to train a model that can reliably predict pseudo-labels similar to the labeling behavior the model was trained on.

7. CONCLUSION

ML has the potential to reliably detect defects for a wide variety of patterns. This includes exploratory patterns such as HEXCH DSA patterns. However, the performance of supervised ML-based defect detection models is dependent on the quality of the labels of its training data. In this work, we proposed a method of combining

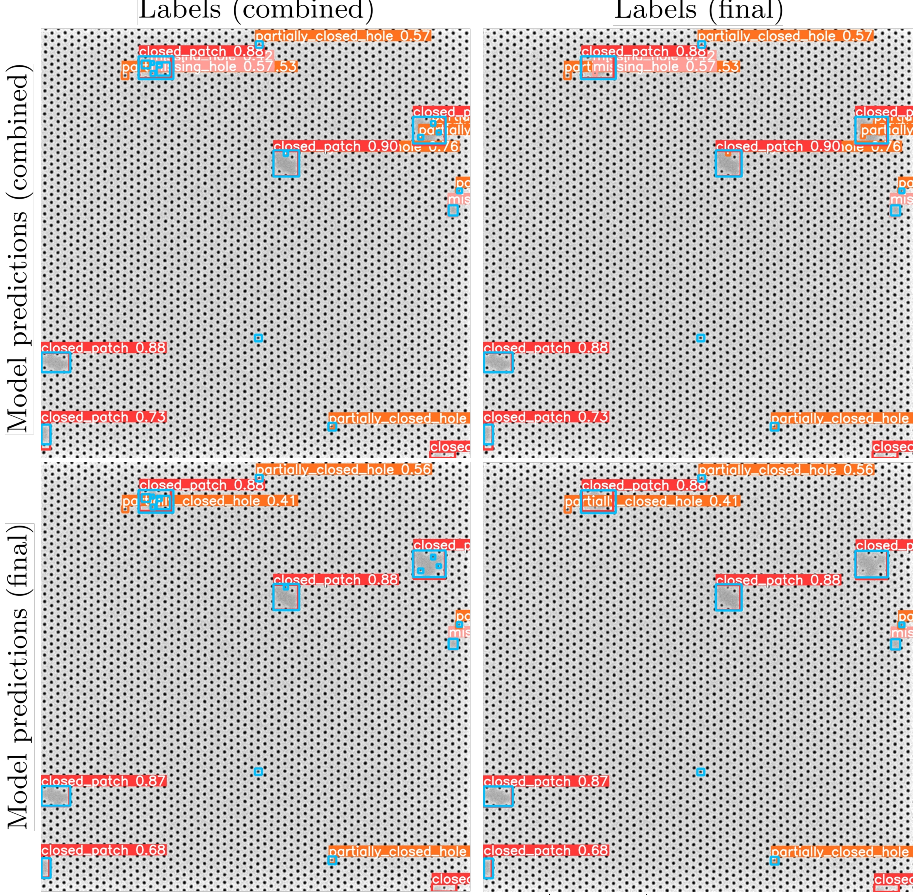
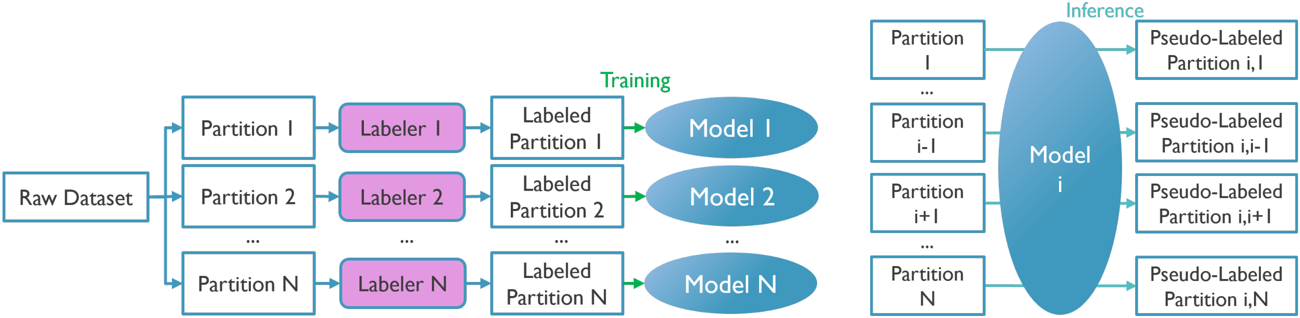


Figure 9: Prediction examples on a test image for YOLOv8m (medium) models trained on the labels from the combined and final datasets (top row and bottom row, respectively). Displayed predictions include confidence scores. Each prediction example is overlayed with labels of the test image from the combined and final datasets (left column and right column, respectively). The labels are shown in blue without class information to improve visibility. Best viewed digitally and zoomed-in for the best visibility of the prediction and label details.



(a) Per partition labeling and model training. Note each model is the same architecture but is trained on different partitions.

(b) Prediction of pseudo-labels by a model trained on labeled partition i.

Figure 10: Suggested method for semi-automated labeling of large datasets in future work.

labels from multiple human labelers to create a more coherent and complete final dataset. This final dataset best reflected the labels expected by a DSA expert without requiring exhaustive manual inspection and label correction from the DSA expert. Experimental results showed that state-of-the-art YOLOv8 models can achieve high detection precisions of more than 0.9 mAP when trained and tested on the final dataset. Finally, we suggested future work continue in the direction of data-centric ML by exploring automatic defect label removal and pseudo-label prediction.

REFERENCES

[1] Dey, B., Goswami, D., Halder, S., Khalil, K., Leray, P., and Bayoumi, M. A., "Deep learning-based defect classification and detection in SEM images," in [*Metrology, Inspection, and Process Control XXXVI*], Robinson, J. C. and Sendelbach, M. J., eds., SPIE (June 2022).

[2] Redmon, J., Divvala, S., Girshick, R., and Farhadi, A., "You only look once: Unified, real-time object detection," (2015).

[3] Jocher, G., Chaurasia, A., and Qiu, J., "YOLOv8 by Ultralytics," (1 2023).

[4] Zha, D., Bhat, Z. P., Lai, K.-H., Yang, F., Jiang, Z., Zhong, S., and Hu, X., "Data-centric artificial intelligence: A survey," (2023).

[5] Oquab, M., Darcet, T., Moutakanni, T., Vo, H., Szafraniec, M., Khalidov, V., Fernandez, P., Haziza, D., Massa, F., El-Nouby, A., Assran, M., Ballas, N., Galuba, W., Howes, R., Huang, P.-Y., Li, S.-W., Misra, I., Rabbat, M., Sharma, V., Synnaeve, G., Xu, H., Jegou, H., Mairal, J., Labatut, P., Joulin, A., and Bojanowski, P., "Dinov2: Learning robust visual features without supervision," (2023).

[6] Bogatu, A., Fernandes, A. A. A., Paton, N. W., and Konstantinou, N., "Dataset discovery in data lakes," in [*2020 IEEE 36th International Conference on Data Engineering (ICDE)*], 709–720 (2020).

[7] Khurana, U., Samulowitz, H., and Turaga, D., "Feature engineering for predictive modeling using reinforcement learning," in [*Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence and Thirtieth Innovative Applications of Artificial Intelligence Conference and Eighth AAAI Symposium on Educational Advances in Artificial Intelligence*], *AAAI'18/IAAI'18/EAAI'18*, AAAI Press (2018).

[8] Cubuk, E. D., Zoph, B., Mané, D., Vasudevan, V., and Le, Q. V., "Autoaugment: Learning augmentation strategies from data," in [*2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*], 113–123 (2019).

[9] Ouyang, L., Wu, J., Jiang, X., Almeida, D., Wainwright, C., Mishkin, P., Zhang, C., Agarwal, S., Slama, K., Ray, A., Schulman, J., Hilton, J., Kelton, F., Miller, L., Simens, M., Askell, A., Welinder, P., Christiano, P. F., Leike, J., and Lowe, R., "Training language models to follow instructions with human feedback," in [*Advances in Neural Information Processing Systems*], Koyejo, S., Mohamed, S., Agarwal, A., Belgrave, D., Cho, K., and Oh, A., eds., **35**, 27730–27744, Curran Associates, Inc. (2022).

[10] Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A. C., Lo, W.-Y., Dollár, P., and Girshick, R., "Segment anything," (2023).

[11] Breaux, L. and Singh, B., "Automatic defect classification system for patterned semiconductor wafers," in [*Proceedings of International Symposium on Semiconductor Manufacturing*], 68–73 (1995).

[12] Ritchison, J. W., Ben-Porath, A., and Malocsay, E., "SEM-based ADC evaluation and integration in an advanced process fab," in [*Metrology, Inspection, and Process Control for Microlithography XIV*], Sullivan, N. T., ed., **3998**, 258 – 268, International Society for Optics and Photonics, SPIE (2000).

[13] McGarvey, S. and Kanezawa, M., "Automated defect review of the wafer bevel with a defect review scanning electron microscope," in [*Metrology, Inspection, and Process Control for Microlithography XXIII*], Allgair, J. A. and Raymond, C. J., eds., **7272**, 72723N, International Society for Optics and Photonics, SPIE (2009).

[14] Dixit, D., Green, A., Hosler, E. R., Kamineni, V., Preil, M., Keller, N., Race, J., Chun, J. S., O'Sullivan, M., Khare, P., Montgomery, W., and Diebold, A. C., "Optical critical dimension metrology for directed self-assembly assisted contact hole shrink," *Journal of Micro/Nanolithography, MEMS, and MOEMS* **15**, 014004–014004 (March 2016). https://www.spiedigitallibrary.org/journals/journal-of-micro-nanolithography-mems-and-moems/volume-15/issue-1/014004/Optical-critical-dimension-metrology-for-directed-self-assembly-assisted-contact/10.1117/1.JMM.15.1.014004.full ; https://nanolithography.spiedigitallibrary.org/article.

aspx?doi=10.1117/1.JMM.15.1.014004 ; http://ui.adsabs.harvard.edu/abs/2016JMM&M..15a4004D/abstract.

[15] Ito, C., Durant, S., Lange, S., Harukawa, R., Miyagi, T., Nagaswami, V., Delgadillo, P. R., Gronheid, R., and Nealey, P., "Inspection of directed self-assembly defects," in [*Alternative Lithographic Technologies VI*], Resnick, D. J. and Bencher, C., eds., **9049**, 90492D, International Society for Optics and Photonics, SPIE (2014).

[16] Pathangi, H., Chan, B. T., Bayana, H., Vandenbroeck, N., den Heuvel, D. V., Look, L. V., Rincon-Delgadillo, P. A., Cao, Y., Kim, J., Lin, G., Parnell, D., Nafus, K., Harukawa, R., Chikashi, I., Polli, M., D'Urzo, L., Gronheid, R., and Nealey, P. F., "Defect mitigation and root cause studies in 14 nm half-pitch chemo-epitaxy directed self-assembly LiNe flow," *Journal of Micro/Nanolithography, MEMS, and MOEMS* **14**(3), 031204 (2015).

[17] Cheon, S., Lee, H., Kim, C. O., and Lee, S. H., "Convolutional neural network for wafer surface defect classification and the detection of unknown defect class," *IEEE Transactions on Semiconductor Manufacturing* **32**(2), 163–170 (2019).

[18] Wang, Z., Yu, L., and Pu, L., "Defect simulation in SEM images using generative adversarial networks," in [*Metrology, Inspection, and Process Control for Semiconductor Manufacturing XXXV*], Adan, O. and Robinson, J. C., eds., **11611**, 116110P, International Society for Optics and Photonics, SPIE (2021).

[19] Dey, B., Dehaerne, E., and Halder, S., "Towards improving challenging stochastic defect detection in SEM images based on improved YOLOv5," in [*Photomask Technology 2022*], Kasprowicz, B. S., ed., **12293**, 1229305, International Society for Optics and Photonics, SPIE (2022).

[20] Chang, C.-Y., Chang, J.-W., and der Jeng, M., "An unsupervised self-organizing neural network for automatic semiconductor wafer defect inspection," in [*Proceedings of the 2005 IEEE International Conference on Robotics and Automation*], 3000–3005 (2005).

[21] Ofir, N., Yacobi, R., Granoviter, O., Levant, B., and Shtalrid, O. IEEE (oct 2022).

[22] Dey, B., Cerbu, D., Khalil, K., Halder, S., Leray, P., Das, S., Sherazi, Y., Bayoumi, M. A., and Kim, R. H., "Unsupervised machine learning based CD-SEM image segregator for OPC and process window estimation," in [*Design-Process-Technology Co-optimization for Manufacturability XIV*], Yuan, C.-M., ed., **11328**, 113281G, International Society for Optics and Photonics, SPIE (2020).

[23] Dehaerne, E., Dey, B., Halder, S., and Gendt, S. D., "Optimizing YOLOv7 for semiconductor defect detection," in [*Metrology, Inspection, and Process Control XXXVII*], Robinson, J. C. and Sendelbach, M. J., eds., **12496**, 124962D, International Society for Optics and Photonics, SPIE (2023).

[24] Kondo, N., Harada, M., and Takagi, Y., "Efficient training for automatic defect classification by image augmentation," in [*2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*], 226–233 (2018).

[25] Delgadillo, P. A. R., Thode, C. J., Nealey, P. F., Gronheid, R., Wu, H., Cao, Y., Neisser, M., Somervell, M. H., and Nafus, K., "Implementation of a chemo-epitaxy flow for directed self-assembly on 300-mm wafer processing equipment," *Journal of Micro/Nanolithography, MEMS, and MOEMS* **11**(3), 031302 (2012).

[26] Singh, A., Nam, J., Lee, J., Chan, B. T., Wu, H., Yin, J., Cao, Y., and Gronheid, R., "Manufacturability of dense hole arrays with directed self-assembly using the CHIPS flow," in [*Alternative Lithographic Technologies VIII*], Bencher, C., ed., **9777**, 97770P, International Society for Optics and Photonics, SPIE (2016).

[27] Doise, J., Koh, J. H., Kim, J. Y., Zhu, Q., Kinoshita, N., Suh, H. S., Delgadillo, P. R., Vandenberghe, G., Willson, C. G., and Ellison, C. J., "Strategies for increasing the rate of defect annihilation in the directed self-assembly of high- block copolymers," *ACS Applied Materials & Interfaces* **11**(51), 48419–48427 (2019). PMID: 31752485.

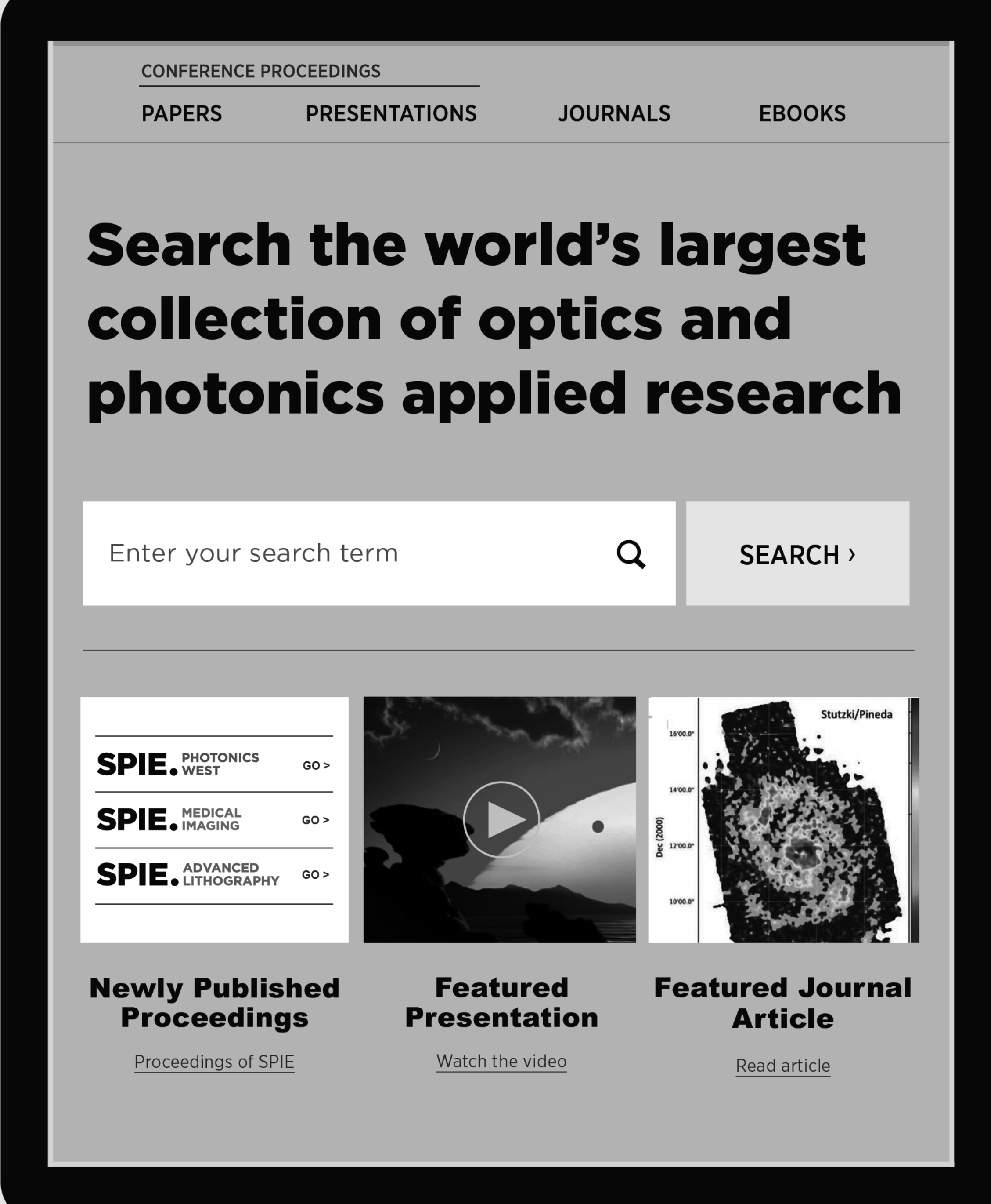
[28] Solovyev, R., Wang, W., and Gabruseva, T., "Weighted boxes fusion: Ensembling boxes from different object detection models," *Image and Vision Computing* , 1–6 (2021).

[29] Lin, T., "labelImg." https://github.com/heartexlabs/labelImg (2021).

[30] Brooks, J., "COCO Annotator." https://github.com/jsbroks/coco-annotator/ (2019).

[31] Lin, T.-Y., Maire, M., Belongie, S., Bourdev, L., Girshick, R., Hays, J., Perona, P., Ramanan, D., Zitnick, C. L., and Dollár, P., "Microsoft coco: Common objects in context," (2014).

SPIE. DIGITAL LIBRARY



Featuring 490,000 papers from SPIE Conferences, Journals and eBooks.

www.spiedl.org