

# The High-NA EUV Lithography: Architecture and Operation in ASML TwinScan EXE:5000 Systems

---

Author: Attilio Lo Magro, Eng.

June 2026

## Abstract

This educational manual explores the fundamental principles of High-NA EUV lithography and the architecture of next-generation ASML systems (EXE 5000 series). The analysis details the transition to 0.55 NA, the implementation of anamorphic optical systems, and the advanced physics of laser-produced plasma. The text further examines the intricate fabrication of reflective multilayer reticles and the ultra-precision engineering of the EXE 5000 platform, including examining the frictionless magnetic levitation (Maglev) stages and the laser interferometry systems capable of picometric measurement accuracy. By bridging the gap between complex VLSI engineering and technical accessibility, this work provides a structured reference for the 2 nm node era and the future of semiconductor manufacturing.

## Author's Note

This work stems from a commitment to apply the rigor of engineering analysis to High-NA EUV photolithography. Drawing from a background in electronic design and subsequently in telecommunications intellectual property protection, gained in prestigious companies such as GTE, Siemens, and Italtel, the author has curated this synthesis as a technical analyst. A distinctive feature of this study is the integration of advanced AI tools to navigate and structure complex theoretical domains, mirroring the methodological approach typical of a patent attorney.

---

## Legal Notice and License: Free Distribution – Not for Sale

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License (CC BY-NC-ND 4.0).

The author authorizes the free distribution and reading of this manual for personal, educational, and research purposes. Any commercial use or modification of the text without explicit authorization is prohibited.

## Disclaimer

This document is for educational and informative purposes only. All trademarks mentioned in the text, including ASML, Zeiss, Trumpf, Schott, Intel, TSMC, Samsung, STMicroelectronics, etc. belong to their respective owners and are used solely for descriptive and identification purposes. The technical concepts contained herein are derived from public domain scientific documentation, patents, white papers, and official academic publications.

How to Cite / Digital Object Identifier:

This monograph is permanently accessible via: **DOI: 10.5281/zenodo.20593018**

Prior Work Reference: This monograph represents the definitive, expanded development of the author's precursor technical study published in Italian: "*La Rivoluzione High-NA EUV*" (Prior DOI: [10.5281/zenodo.19914081](https://doi.org/10.5281/zenodo.19914081))

## PREFACE

You may wonder about the purpose of yet another publication on the ASML EXE:5000 series scanner, an instrument already frequently featured in international technical and scientific literature and heavily promoted on the Web — especially on YouTube — through highly visually impactful videos produced by the manufacturer itself. The answer lies in the author's desire to move beyond both commercial propaganda and the fragmented approach of single-topic monographs or patent descriptions. The goal is to provide a framework that is as "unified" as possible, guiding readers — who already possess a certain background in the technology discussed — to enable them to explore all the features of this incredibly complex machine, which are synergistically linked to overcome the numerous obstacles in the race toward extreme miniaturization in modern VLSI manufacturing. Achieving a comprehensive, specialized knowledge of this tool is nearly impossible given the vastness of the scientific fields involved; however, the author is convinced that those who have the patience to reach the end of this book will be rewarded with a thorough, 360-degree objective view of this scanner.

It must be stated as a premise that industrial standardization in this sector requires starting from circular silicon wafers with a diameter of 300 mm and a thickness of approximately 775  $\mu\text{m}$ . Integrated circuits (chips) are simultaneously fabricated across each of these wafers, with their dimensions varying according to their design complexity. Undoubtedly, the urgent demand for Artificial Intelligence (AI) and machine learning chips has pushed hardware integration to its maximum limits. Rare niche implementations are known to exist — such as the Cerebras WSE-3 — which occupy an entire silicon wafer with 4 trillion transistors, or the latest NVIDIA GPUs (Rubin, estimated at around 400 to 500 billion transistors), but these remain exceptions. If we focus instead on the domain of on-chip VLSIs (CPUs, GPUs, DRAMs) most widely utilized in mass-market commercial products — such as notebooks, desktop computers, gaming consoles, smartphones, SoCs, individual motherboards, and internet servers — the physical dimensions of the individual chips are likewise standardized. The bare chips produced by this scanner feature the following dimensions: a maximum area per single exposure field (Half-Field) of  $26 \times 16.5$  mm, which can be extended to  $26 \times 33$  mm (Full-Field). Within the standard half-field configuration, the circuit multilayer itself has a thickness of a few tens of micrometers, boasting a transistor density that is currently the highest in the world — exceeding 500 million transistors per square millimeter (equivalent to over 50 billion transistors per square centimeter) — achieved through a single, highly optimized exposure pass. Currently, the most of semiconductor manufacturers prefer this native half-field configuration for standard high-volume production, as it optimizes machine throughput and mask costs for mainstream processors and memories. Conversely, the full-field stitched layout — which explicitly requires two separate, synchronized scanning passes on the same silicon area utilizing two distinct masks — is exclusively reserved for massive "monster chips," such as high-performance artificial intelligence (AI) accelerators and data center graphics processing units (GPUs) that physically exceed the native geometric limits of the anamorphic optics.

## PLATES: INDUSTRIAL OVERVIEW



**Plate I:** The ASML TWINSKAN EXE High-NA EUV System in the Cleanroom  
Description: Full view of the operational platform during final hardware integration inside the fab environment. Technicians in protective cleanroom suits ("bunny suits") are visible in the foreground, establishing the real-world macroscopic scale and vertical proportions of the tool.



**Plate II:** Internal Structural Module of the Projection Optics Box (POB)  
Description: Detailed view of the scanner core assembly without external protection panels. The heavy structural frames house the nanometric Maglev stage tracks and the active thermal correction mechanisms engineered by Carl Zeiss SMT.



**Plate V:** Optical Path of the Illumination System and the Projection Optics Box (POB)  
Description: Internal three-dimensional schematic of the central scanner core, mapping the complete reflection sequence across the Carl Zeiss SMT mirror assembly. Note: The colored rays tracing the path from the illuminator system through the reflective reticle and into the anamorphic POB are digitally rendered for illustrative clarity; the actual extreme ultraviolet (EUV) radiation operates at a nanometric wavelength 13.5 nm propagating entirely within an ultra-high vacuum (UHV) environment and remaining physically invisible to the human eye.



**Plate III:** Sub-Fab Infrastructure: High-Power Laser Source and Beam Delivery Unit (BDU).  
Description: The massive underground support level situated directly beneath the cleanroom floor. This area contains the multi-stage TRUMPF pulsed CO<sub>2</sub> laser seed oscillators, industrial optical power amplifiers, and the rigid BDU pipeline that routes the infrared "photonic pump" toward the tin droplet generator core.



**Plate IV:** The Zeiss SMT high-NA EUV mirror, featured as Plate IV, utilizes specialized atomic layering for high-reflectivity molybdenum-silicon coatings within a cleanroom environment. Technical specifications include a 6-ton illuminator optics box with 25,000 components and a 12-ton projection optics box (POB) containing 40,000 components, totaling 18 tons of optical assembly. Learn more about Zeiss SMT's manufacturing technology at [ZEISS SMT](#).

**Legal Copyright Notice & Non-Infringement Statement**  
The photographic material compiled within this section consists of official press-kit illustrations and technical imagery publicly disclosed for educational and journalistic purposes by ASML Holding N.V., Intel Corporation, and the Interuniversity Microelectronics Centre (imec). All trademarks, logos, and original visual copyrights are the exclusive property of their respective corporate owners. This material is reproduced herein in good faith for non-commercial, academic analysis and technical commentary under the "Fair Use" provisions of international intellectual property frameworks. The author declares no corporate affiliation with the aforementioned entities and asserts no intent of copyright infringement.

## PREFACE

## PLATES: INDUSTRIAL OVERVIEW

## INDEX OF CONTENTS

Page

<i>TECHNICAL NOTE ON GEOMETRIC CONVENTIONS</i>	01
<b>Introduction:</b>	
<b>a) Physical limits of miniaturization.</b>	01
<b>b) Reticle fabrication and utilization.</b>	03
<b>c) Industrial Strategies and the Silicon Limit</b>	06
c1. From FinFET to GAAFET ( <i>Nanosheets</i> ) and CFET	06
c2. Beyond Silicon: Emerging Materials and Architectures	06
<b>d) VLSI Manufacturing Process: A Cyclic Evolution</b>	07
d1. The Lithographic Cycle and the Load-Lock System	08
d2. Scanner Entry: Precision Alignment and Topography Mapping	08
d3. Dual-Pass Exposure Process	09
d4. Wafer Unloading and Reticle Retention Management	09
d5. Development ( <i>Opening the "Windows"</i> )	09
d6. Pattern Transfer ( <i>The Choice of Intervention</i> )	09
d7. Cleaning and Reset (Stripping, Planarization, and Re-Coating)	10
d8. Cycle Completion, Die Dicing, and Packaging	10
d9. Macro-Economics of Advanced Sub-3nm Nodes	11
<b>1. The EXE:5000/5200 scanner: An overview</b>	12
1.1. System Overview and Architecture	12
1.1.1. The Remote Source Area	12
1.1.2. The Main Level	13
1.1.3. Auxiliary and Computational Systems	14
1.1.4. Conclusions on this chapter	15
<b>2. Vacuum Systems – Multi-stage Pumping Systems</b>	15
<b>3. Maglev Movement and Advanced Metrology</b>	16
3.1. Magnetic Levitation Motors ( <i>Maglev</i> )	16
3.1.1. Active Levitation and the Dual-Carriage System	17
3.1.2. Operating Cycle: Step-and-Scan	17
3.1.3. The serpentine scanning	18
3.2. Advanced Metrology	19
3.2.1. Operating Principles: Planar Optical Encoders	19
3.2.2. Insight into Wave-Quantum Optics and Metrology	21
3.2.3. Measurement Precision	22
3.2.4. Alignment Marks: The Diffraction "Fingers"	22
<b>4. The EUV Optical System</b>	23

4.1. Multilayer Mirrors ( <i>Bragg Gratings</i> )	23
4.1.1. Resonance and Constructive Interference	24
4.2. The Optical Path: From Plasma to Chip	25
4.3. The REMA Unit: Dynamic Exposure Framing	26
4.3.1. Structural Description	26
4.3.2. Functional and Kinematic Description	27
4.4. Reflective Reticule Projection	28
4.5. Mirror Manufacturing: Ion Beam Sputtering ( <i>IBS</i> )	29
4.6. Global Stability and Mirror Integrity	29
4.6.1. Zero-Expansion Substrates	29
4.6.2. Cooling, Stability, and Active Isolation	29
4.6.3. Adaptive Optics and Aberration Correction.	30
4.6.4. Predictive Feed-Forward Thermal Compensation and Illuminator Orchestration	31
<b>5. High-NA Geometry: Anamorphic Optics and Stitching</b>	<b>33</b>
5.1. The Problem: Angle of Incidence and Shadowing	33
5.2. Conceptual Formula for Shadow Length	33
5.3. The Solution: Anamorphic Optics	33
5.4. Reticule Design	34
5.5. Anamorphic Geometry, Field Width, and Kinematic Synchronization	34
5.6. Stitching: Nanometric "Seaming"	35
5.6.1. Strategies for Perfect Overlay	35
5.6.2. Wafer Layout	36
5.6.3. The Economic Paradox.	36
<b>6. Plasma Generation: The Tin Target</b>	<b>36</b>
6.1. MOPA Architecture ( <i>Master Oscillator Power Amplifier</i> )	37
6.1.1. The Master Oscillator	37
6.1.2. Staged Amplification: The Photonic "Pump"	37
6.1.3. Transport of the Laser Beam and Focusing	38
6.2. The Tin Droplet Generator	39
6.3. Spatial and Temporal Synchronization	39
6.4. The "Double-Hit" Strategy and Plasma Transformation	40
6.4.1. Timing of the Duty Cycle	40
6.4.2. Kinematic Energy Integration and Single-Die Transit Dynamics	41
<b>7. Beam Power: From Megawatts to Watts</b>	<b>42</b>
7.1. Decay of Useful Power	42
7.2. Energy Balance and the "Critical Point"	43
<b>8. Protection and Stability of the Laser System</b>	<b>43</b>
8.1. Optical Isolation and Self-Oscillation Protection.	44
8.2. Outgassing and Material Integrity under Ultra-High Vacuum	45
<b>9. The Intermediate Focus Shutter Mechanical System</b>	<b>45</b>
9.1. Structural Description	45
9.2. Functional and Kinematic Description	46
<b>10. In-Situ Maintenance and Clean-Optics Chemistry</b>	<b>47</b>

10.1. Structural Description	47
10.2. Functional and Chemical Description	47
10.3. Synopsis of Hydrogen Stabilization Effects	48
<b>11. Real-Time Distributed Computational Infrastructure</b>	49
11.1. Computational Requirements and Process Parallelism	49
11.2. Loop Speed Hierarchy and Priority Matrix	50
11.3. Local Edge Computing: In-Situ FPGAs	50
11.4. Sensor Distribution and External Optical Bus	51
11.5. Physical Architecture of the Supercomputer	51
11.6. Functional Architecture and Master Orchestration	51
<b>12. Automated Decommissioning and Controlled Shutdown System</b>	52
12.1. Structural Description	52
12.2. Functional and Chronological Description	53
<b>CONCLUSION</b>	55
INTRODUCTION TO THE APPENDICES	58
<b>APPENDIX A – From VLSI Design to the Reticle</b>	58
A.1 – Layout File	58
A.2 – OPC ( <i>Optical Proximity Correction</i> )	59
A.3 – M3D ( <i>Reticle 3D Effects</i> )	59
A.4 – SMO ( <i>Source Reticle Optimization</i> )	60
A.5 – Fracturing	61
A.6 – Typical File Sizes.	61
<b>APPENDIX B – Reticle Writing</b>	61
B.1 – Electron Resist	61
B.2 – Gaussian Beam Writer	62
B.3 – Multi-Beam Writer	62
B.4 – Writing Times	62
B.5 – Required Precision	63
B.6 – Development and Etching	63
B.7 – Cleaning and Typical Defects	63
B.8 – Reticle Inspection and Qualification.	64
<b>REFERENCES</b>	65
<b>APPENDIX C – Glossary of Acronyms</b>	66
<b>Figures (Slides 1 to 3) attached at the end of the article.</b>	

**Technical Note on Geometric Conventions:** To prevent semantic and geometrical ambiguity throughout this monograph, a formal clarification regarding the coordinate reference frame is required. In standard analytical geometry, the X-axis is conventionally represented as horizontal and the Y-axis as orthogonal to it within a two-dimensional Cartesian plane. Within the semiconductor industry and specifically across ASML lithographic architecture, all physical operations take place entirely on a horizontal plane, upon which both axes reside. Within this spatial reference frame, which is anchored to the physical kinematics of the stage translation system, the Y-axis is assumed by convention to denote the longitudinal scanning axis (the high-velocity direction of continuous stage travel). Conversely, the X-axis denotes the orthogonal, transverse stepping axis (the direction of lateral displacement between adjacent exposure fields).

## INTRODUCTION

### a) Physical limits of miniaturization

The miniaturization of transistors is not determined solely by materials engineering or device design: it is primarily constrained by the fundamental limits of optics. Since the circuit pattern is transferred onto silicon by projecting light via a mask onto a photoresist, the minimum achievable feature size depends directly on the physical limits of the optical system. Every lithographic system can only etch structures on the wafer that are larger than a certain minimum size, defined by Rayleigh's formula:

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

where:

- CD is the minimum printable Critical Dimension.
- $\lambda$  (*lambda*) is the wavelength of the light source.
- $NA = n \cdot \sin \alpha$  is the Numerical Aperture of the objective (*the final projector mirror*), where:  $n = 1$  is the refractive index of the medium (*vacuum*), and  $\alpha$  (*alpha*) is the half-angle of the maximum cone of light that the projection optics can collect. It represents the "light-gathering" capacity of the objective.
- $k_1$  is a process-related factor (*industrial limit*  $\approx 0.25$ ).

This relationship demonstrates that resolution can be improved in only two ways:

1. Reducing the wavelength  $\lambda$
2. Increasing numerical aperture  $NA$ .

### The Limit of DUV (*Deep Ultraviolet*) Lithography

For over twenty years, the industry pushed both variables to their limits, but DUV lithography stalled at  $\lambda = 193$  nm. At this point, physics prevents further significant progress. Even with a maximum  $NA$  of  $\approx 1.35$  (*immersion lithography*), Rayleigh's formula yields:

$$CD_{min} \approx 0.25 \cdot \frac{193}{1.35} \approx 36 \text{ nm}$$

This value represents a physical wall: DUV lithography cannot produce real structures smaller than  $\approx 36$  nm. Everything achieved below 20 nm using 193 nm light is the result of multi-patterning (*LELE, SADP, SAQP*), multiple overlays, and extreme OPC (*Optical Proximity Correction*), but it is not "true" optical resolution. Consequently, 193 nm technology cannot inherently reach the 5 nm scale.

### **EUV (Extreme Ultraviolet)**

EUV lithography utilizes a source with:

- $\lambda = 13.5$  nm
- $NA = 0.33$

Applying Rayleigh's formula:

$$CD_{min} \approx 0.25 \cdot \frac{13.5}{0.33} \approx 10 \text{ nm}$$

With advanced techniques (*OPC, SMO, M3D*), further reductions are achieved, albeit at the cost of significantly increased process complexity.

### **High-NA EUV: The Current Frontier**

This next generation of EUV scanners utilizes:

- $\lambda = 13.5$  nm
- $NA = 0.55$
- Anamorphic optics (*4x magnification in the X direction, 8x in the Y direction*).

Applying Rayleigh's formula:

$$CD_{min} \approx 0.25 \cdot \frac{13.5}{0.55} \approx 6 \text{ nm}$$

High-NA systems employ larger-diameter mirrors, capable of reflecting more EUV radiation generated by a more powerful source.

Modern miniaturization is driven by the physics of diffraction, not just by silicon properties.

High-NA represents the physical foundation for process nodes at 2 nm and beyond.

*Note: The term "2 nm node" is a commercial designation and does not correspond to a real physical dimension, such as the gate length (which is larger), at around 6 nm. The term "node" is used to indicate a specific generation of the chip manufacturing process. The following description only refers to the EUV technology.*

---

## b) Reticle fabrication and utilization

*(Details covered in Appendices A and B)*

*Note on Terminology: To prevent any structural ambiguity, this monograph strictly uses the term reticle to identify the electronic mask, while REMA (Reticle Masking) refers exclusively to the mobile mechanical blades used to frame the EUV light slot.*

The production of an integrated circuit involves two profoundly different lithographic phases:

1. Writing the reticle using an electron beam writer. A modern VLSI, such as a CPU or GPU, is a complex multi-layered circuit structure on silicon that requires the use of multiple reticles.
2. Exposing the photoresist on the wafer using an EUV scanner to reproduce the layout of the reticle and transfer it to the silicon.

Although both processes "draw" patterns, they operate on totally different physical, mechanical, and temporal principles. Understanding this distinction is essential to grasp why the reticle must be more precise than the scanner that utilizes it.

### b1 – FABRICATION of the Reticle

The physical substrate of an EUV reticle is based on a standard 6-inch square plate measuring exactly  $152 \times 152$  mm with a thickness of 6.35 mm. This structural plate provides a maximum active design area for the circuit layout measuring  $104 \times 132$  mm. Any pattern written within these boundaries represents the non-reduced macro-scale master copy of the integrated circuit.

A reticle for EUV lithography is not a transmissive plate, nor can it be reduced to an opaque pattern on a substrate. Instead, it is a complex multilayer mirror based on reflective Bragg grating technology. While standard EUV reticles use a uniform 4x magnification, High-NA EUV reticles utilize anamorphic optics: they maintain a 4x magnification in one direction X but increase to 8x in the other Y. This will be discussed extensively in the following sections. The components of the reticle are:

- **The Substrate:** The massive "body" of the mirror that provides its shape and curvature and supports the Bragg stratification.
- **The Mirror:** Composed of a Molybdenum/Silicon (Mo/Si) multilayer, responsible for reflecting the 13.5 nm radiation.
- **The Ruthenium (Ru) Capping Layer:** A very thin layer of Ruthenium (*approx. 2.5 - 5 nm thick*) is deposited over the multilayer. This layer is "transparent" to EUV radiation to allow reflection but acts as an inert chemical shield to protect the underlying silicon during subsequent preparation phases of the reticle.
- **The Absorber (TaBN):** An opaque layer of Tantalum Boron Nitride (TaBN) sits atop the Ruthenium. Its task is to absorb EUV radiation during the wafer exposure phase to define the "dark" areas of the circuit.

- **Polymeric Electron Resist:** This is spread over the absorber to define the pattern.

### b1.1 – Masking Process

- **Writing:** The electron beam of the e-beam writer does not act directly on the absorber; instead, it "draws" on the electron resist, modifying its solubility.
- **Development and Selective Etching:** After developing the resist (*which uncovers the absorber only at desired points*), the reticle undergoes a plasma etching process. This is a selective chemical removal mediated by reactive gases that transform the exposed absorber TaBN into gaseous compounds, physically removing it without damaging the multilayer, which is protected by the capping layer. The result is a plate etched with an extremely precise reflective pattern. Any minimal positional errors on the reticle are further reduced on the wafer thanks to the optical reduction (4x, 8x) of the EUV scanner, ensuring that the final resolution is limited only by the diffraction laws of EUV optics.

### b.2 – UTILIZATION: Pattern Transfer to Photoresist and Masking

In VLSI production, the EUV scanner does not "write" anything; it simply projects the pattern reflected by the reticle onto an EUV-sensitive photoresist previously deposited on the wafer.

**b2.1 – Projection:** The projection onto the photoresist does not happen like a static photographic "snapshot" but through a coordinated movement called scanning:

- **The Focal Plane:** Unlike traditional photography where a large depth of field is often desired, High-NA lithography operates with an extremely shallow DOF (*Depth of Focus*). This requires the wafer to be positioned within the focal plane with "picometric" tolerances. In ASML scanners, the reduction ratio (4x, 8x) is fixed, determined by the specific curvature and position of Zeiss mirrors. If the wafer shifts even slightly from this fixed focal plane, the image immediately becomes blurred, compromising the nanometric resolution.
- **Opposing Motion:** The movement of the reticle stage in the opposite direction to a wafer stage is an optical necessity. Since the projection system generates a real image rotated by 180° (*inverting both the X and Y axes*), the wafer must travel in the exact opposite direction of the reticle. For instance, as the reticle moves forward along the scanning Y-axis, the projected image moves backward. The wafer stage must therefore synchronize its motion to keep the projected image perfectly stationary relative to the wafer surface as the photoresist-coated wafer continuously passes through a stationary illumination slit (*described below*), ensuring that every point of the photoresist receives the correct radiation dose. This perfectly orchestrated "counter-dance" is vital to capture the pattern accurately. Without this precise countermovement, the pattern would "smear" and the energy balance would

be compromised, destroying the nanometric resolution. This mechanism also balances the inertial forces generated by the violent accelerations of the scanning stages (*driven by a magnetic levitation system described in Section 3.1.*), minimizing parasitic vibrations.

- **Scanning Velocity Ratio:** The reticle and wafer stages must move in perfect synchronization along the Y-axis, which is orthogonal to the slit of light (*described below*). Since the High-NA optics reduce the image by 8x in this scanning direction, the reticle stage must travel 8 times faster than the wafer stage to ensure the pattern is accurately transferred.

**b2.1.1 – The Slit of Light and the REMA Unit:** During the lithographic process, the unformed EUV illumination beam must be restricted to match the active field of the reticle. This is achieved by a REMA unit (*see Section 4.3.1.*), which shapes the light into a narrow, slightly curved slit. Through the asymmetric reduction of the anamorphic optics (4x along the X-axis and 8x along the Y-axis), (*See chapter 5.*), this beam is compressed down to a final chip scale measuring 26 mm in length along the cross-wise step X-axis and approximately 5 mm in width along the scanning Y-axis when projected onto the wafer surface. This sharp geometric framing isolates the active circuit pattern of the reticle, preventing the light from exposing surrounding non-active areas.

**b2.2 – Chemical Characteristics of the Photoresist:** The photoresist is a photosensitive polymer applied to the wafer via spin coating. Two main types of resists are used in EUV:

1. **CAR** (*Chemically Amplified Resists*): Based on organic molecules that, when struck by photons, release acids that catalyze the reaction.
2. **MOR** (*Metal-Oxide Resists*): A new frontier for EUV, based on metals (like tin or hafnium), which absorb EUV photons better than organic resists, allowing for thinner and sharper lines.

**b2.3 – Transformation under EUV Light:** Unlike visible light, a single EUV photon has very high energy (92 eV). When it hits the photoresist:

- **Photoelectron Generation:** The EUV photon does not "burn" the resist but knocks an electron out of the polymer molecules. This secondary electron travels a few nanometers and triggers a chemical chain reaction.
- **Positive Resist (*most common*):** Light breaks the polymer chains (*scission*). Struck areas become weak and soluble.
- **Negative Resist:** Light joins the chains (*cross-linking*). Struck areas become hard and insoluble.

*Conclusion: This complex interaction between EUV light and the photoresist concludes the exposure phase; however, the physical definition of the circuit occurs only when the wafer leaves the scanner to undergo the chemical and structural transformations detailed in the FAB processing cycle (see Section d4).*

---

## c) Industrial Strategies and the Silicon Limit

The semiconductor industry is currently in a phase of "diminishing returns," where increasing density requires multibillion-dollar investments in exchange for ever-decreasing performance benefits. Consequently, the future of the digital revolution will not rely exclusively on extreme geometric scaling, but on the capability to integrate these technologies with alternative architectures.

While High-NA EUV pushes transistor density toward the theoretical limits of silicon, technologies such as 3D Advanced Packaging and chiplets offer a complementary evolutionary path. Instead of printing massive monolithic chips that require a complex and costly field-stitching process (see *Section 5.6.*), the industry increasingly favors assembling smaller, heterogeneous chiplets. This approach allows for the combination of diverse computing blocks, utilizing the precious High-NA resolution only for the most critical processing components, thereby optimizing the manufacturing yield, costs, and overall system performance. In parallel, the rise of materials such as Silicon Carbide (*SiC*) for power electronics, alongside promising research into spintronics, suggests that the path forward will be a diversified technological ecosystem, where ASML's lithographic precision remains the fundamental pillar.

### c1. From FinFET to GAAFET (*Nanosheets*) and CFET

For decades, the FinFET (*Fin Field-Effect Transistor*) architecture dominated the market. However, below 3nm, the silicon "fin" becomes so thin that it can no longer effectively control electron flow, causing current leakage due to quantum tunneling and overheating. The industry is therefore migrating toward GAAFET (*Gate-All-Around*), also known as Nanosheets:

- **Structure:** The channel consists of stacked horizontal sheets.
- **Control:** The gate wraps around the channel on all four sides, ensuring total electrostatic control and power customization based on the width of the sheets.

**Future CFET:** The roadmap anticipates complementary FETs, where N and P transistors will be vertically stacked to halve the footprint.

### c2. Beyond Silicon: Emerging Materials and Architectures

The future of microelectronics will be heterogeneous:

- **Chiplets and Advanced Packaging (3D):** Instead of monolithic chips, interconnected "building blocks" are used. The main challenge remains the thermal dissipation of stacked chips, solvable only with complex cooling micro-channels.
- **Spintronics:** Leverages electron spin rather than charge, drastically reducing heat and energy consumption.

- **Silicon Carbide (SiC):** Vital for power electronics (*electric vehicles, renewable energy*). STMicroelectronics (*Catania hub*) is a leader in this sector due to the SiC ability to handle high voltages and temperatures superior to silicon.
- 

## d) VLSI Manufacturing Process: A Cyclic Evolution

The fabrication of a modern microchip is a cyclic and non-linear process that can last several months and include over 1,000 discrete operational steps. Far from being an overstatement, this immense complexity is mathematically dictated by the structural architecture of a modern Integrated Circuit (IC), which requires the vertical stacking of 60 to over 100 separate, atom-thin physical layers to build its interconnected routing network. Because the creation of *every single layer* requires the wafer to go through a cyclical sub-sequence of 10 to 15 highly specialized chemical and physical procedures — including surface cleaning, thin-film atomic deposition (ALD/CVD), photoresist coating, High-NA optical exposure, chemical development, plasma etching, resist stripping, and chemical-mechanical planarization (CMP) — the total cumulative step count inevitably scales into the thousands.

Each circuit layer is built upon the previous one, transforming a silicon wafer into a three-dimensional nanometric "skyscraper" through a continuous exchange between the high-NA optical scanner and the other specialized process chambers of the FAB (*Fabrication Plant*). To put this vertical architecture into perspective, the ASML High-NA scanner enables a structural transistor density exceeding 500 million transistors per square millimeter (over 50 billion transistors per square centimeter). Within the geometric boundaries of a single native half-field die (26 × 16.5 mm), this density allows for the integration of up to 214.5 billion raw transistors.

However, this raw transistor count represents merely the baseline raw material for logic design, rather than its final functional capacity. Within standard CMOS (*Complementary Metal-Oxide-Semiconductor*) technology, every basic electronic switch is fundamentally composed of two separate field-effect transistors (*a matched pair of n-channel and p-channel FETs*). Furthermore, the elementary digital units required for functional computation and data retention utilize multi-transistor topologies of varying complexity: a standard dynamic RAM cell (*1T-1C DRAM*) requires 1 transistor coupled with a capacitor, a high-speed static RAM cell (*6T SRAM*) utilizes 6 transistors, while complex edge-triggered registers, logical gates, and standard flip-flops dynamically consume between 10 and over 24 transistors each. Consequently, the raw density provided by the high-NA optics is rapidly fractionalized by the overhead of hardware logic, meaning that the raw transistor metric serves only as a starting point to evaluate the true architectural and computational potential of different microchip typologies.

This architectural complexity fundamentally explains why a modern microchip requires such a massive vertical multilayer stack, and what these internal layers actually contain. While the hundreds of billions of field-effect transistors are positioned exclusively on the very first, lowest layer directly on the raw silicon substrate (*the Front-End of Line, or FEOL*), they cannot function in isolation. The remaining 60 to over 100 superimposed layers (the Back-End of Line, or BEOL) constitute a colossal, three-dimensional nanometric routing highway made entirely of microscopic metallic wires—composed of copper, cobalt, or ruthenium—separated by chemical insulators. To prevent this astronomical number of electrical connections from intersecting and causing catastrophic short circuits, chip designers arrange them vertically like a matrix of multi-level highway overpasses. Signals travel upward, cross over underlying wires, and drop back down to the transistor level through microscopic vertical microscopic plugs called Vias. This vertical grid follows a strict hierarchical tree structure:

- **The Lowest Layers (closest to the silicon):** Contain the thinnest and shortest wires, measuring only a few nanometers in width, to interconnect adjacent transistors into localized functional cells.
- **The Intermediate Layers:** Utilize slightly thicker metal paths to route data between different logical blocks, such as arithmetic units and memory caches.
- **The Topmost Layers (at the summit of the skyscraper):** Feature the largest and thickest power lines to distribute stable electrical currents across the entire die and route macroscopic output signals toward the external boundary pads.

#### **d1. The Lithographic Cycle and the Load-Lock System**

To maintain the extreme vacuum integrity required for EUV radiation, a main vessel housing the optical scanner is never vented (*re-pressurized to atmospheric levels*) during operation. The transfer of wafers is managed as follows:

- **The Load-Lock Chamber:** This intermediate chamber acts as a "buffer" or airlock between the atmospheric pressure of the cleanroom and the ultra-high vacuum of the Vessel. A robotic arm places the wafer into a load-lock chamber, which is rapidly evacuated of air in a few seconds. Only then does the internal gate open to allow the wafer into the Main Vessel.
- **The FOUP (*Front Opening Unified Pod*):** Outside the machines, wafers travel inside sealed, automated containers called *FOUPs*, which move along overhead tracks to prevent any contact with air or dust between different processing stations.

#### **d2. Scanner Entry: Precision Alignment and Topography Mapping**

When the layout pattern requires advanced geometric stitching to print large chips (see *Section 5.6*), a single circuit layer requires the use of two distinct reticles. Before the exposure cycle begins, both the wafer and the two reticles enter the main vacuum vessel through the load-lock chamber (*introduced in Section d1*). Inside the vessel, the wafer is

precisely positioned onto its wafer carriage by a robotic loading system (*the wafer handler*). Concurrently, the first of the two reticles is loaded onto the single reticle stage, while the second reticle is temporarily stored within an internal vacuum storage system (*the reticle library*) located inside the vacuum vessel, ready for the automatic swap. Once all components are positioned in the ultra-high vacuum environment, the system performs two vital tasks:

- **Overlay Alignment:** Using diffractive alignment marks (*see Section 3.2*), the scanner synchronizes the new layout pattern with the existing underlying layers with sub-nanometric precision, coordinates the sequential exchange of the two reticles on the stage, and maintains perfect layer-to-layer registry.
- **Dynamic Focusing:** The laser metrology system maps the current 3D topography of the wafer, adjusting the focal plane in real-time to compensate for the cumulative thickness of the "nanometric skyscraper" already built.

### **d3. Dual-Pass Exposure Process**

The exposure for large chips (*see Section 5.6.2*) occurs in two distinct, highly synchronized sequential phases using the two separate reticles to form a single, functional circuit layer on the silicon surface. This dual-pass method enables the seamless integration of both halves of the layout across the entire wafer before proceeding to the subsequent manufacturing steps.

### **d4. Wafer Unloading and Reticle Retention Management**

Once the dual-pass exposure cycle is completed, the wafer handler robotically extracts the fully exposed wafer from the wafer carriage and transfers it back out of the main vessel through the load-lock chamber. Concurrently, the system manages the two reticles based on production requirements: if the scanner must process a new wafer of the same lot, both reticles remain securely stored within the internal vacuum storage and the reticle stage to repeat the exposure cycle immediately; otherwise, if a different layer or layout is scheduled, the reticles are retrieved and extracted from the vessel through the load-lock chamber.

### **d5. Development (*Opening the "Windows"*)**

Immediately after exposure, the wafer enters a developer unit:

- **Chemical Wash:** The wafer is treated with a developer solution that washes away the photoresist in the zones struck by EUV light (*in a positive resist process*).
- **The Masking Effect:** The remaining hardened photoresist acts as a protective mask, leaving the underlying material bare and "uncovered" only in the specific areas defined by the circuit design.

### **d6. Pattern Transfer (*The Choice of Intervention*)**

Through these open "windows," the FAB can perform different, often alternative, operations depending on the layer specific function:

- **Etching:** Plasma or chemical agents carve the uncovered material to create trenches or define transistor gates.
- **Ion Implantation:** Specific ions are injected through the windows into the silicon to modify its electrical properties (*doping*), creating the P-N junctions of the transistors.
- **Deposition and Metallization (*Contacts and Via-Holes*):** New atomic layers of conductors (*such as Tungsten or Copper*) or insulators are grown on the wafer using CVD (*Chemical Vapor Deposition*), PVD (*Physical Vapor Deposition*), or ALD (*Atomic Layer Deposition*). Specifically, ALD is critical for the nanometric precision required by modern architectures like GAAFET. This process fills the etched windows to create *Contacts (direct connections to the transistor)* and *via-holes (vertical connections between different metal layers)*, effectively building the "electrical plumbing" of the nanometric skyscraper.

#### **d7. Cleaning and Reset (*Stripping, Planarization, and Re-Coating*)**

- **Stripping:** Once the intervention is complete, the residual resist is no longer needed and is removed with a powerful solvent.
- **CMP (*Chemical Mechanical Polishing*):** Because the previous steps leave the surface uneven, CMP systems atomically level the wafer to a mirror-like finish. This is vital to ensure a perfectly flat foundation for the next layer.
- **Next Layer Preparation (*Spin Coating*):** This operation is conditional based on the manufacturing route: if the multi-layer fabrication is not yet finished and the wafer requires another exposure, a new film of photoresist is applied via spin coating, and the wafer is routed back to the scanner. Conversely, if the current layer completes the chip manufacturing process, this step is bypassed, and the wafer proceeds directly to the final packaging phase (*see Section d8*).

#### **d8. Cycle Completion, Die Dicing, and Packaging**

This final step represents the operational decision checkpoint for the manufacturing sequence: if more wafers from the same production lot require the same lithographic layer, the system routes a new wafer back to the beginning of the sequence to repeat the cycle. Conversely, if all the multi-layer fabrication steps on the silicon are fully completed, the wafer exits the cleanroom environment and is routed to the back-end assembly line. Here, the wafer undergoes a mechanical or laser dicing process, where it is precisely cut along the scribe lines to separate the individual silicon chips (*dies*). Each functional die is then placed onto a protective substrate or leadframe, wire bonded or flipped for electrical interconnections, and hermetically sealed within a plastic, ceramic, or metallic enclosure (*the package*). This encapsulation protects the nanometric structures from mechanical stress and chemical contamination, allowing thermal dissipation and providing the external pins necessary to solder the chip onto commercial printed circuit boards.

## **d9. Macro-Economics of Advanced Sub-3nm Nodes [Ref.1]**

**Costs and Throughput:** To transition from microscopic physics to industrial reality, the deployment of a High-NA scanner must be evaluated through the lens of macro-economics, manufacturing yield, and multi-million-dollar capital investments. When a fabless giant architecture a high-end CPU or GPU utilizing a sub-3nm process node at a premium foundry like TSMC (Taiwan Semiconductor Manufacturing Company), the financial and chronological metrics of a product lifecycle are structured across four deterministic phases:

- a) **Design Cycle and Software Infrastructure (Time and Cost):** The development of a next-generation sub-3nm architecture represents a massive research and engineering effort. The complete cycle — spanning architectural design, logic implementation, hardware verification, and the synthesis of Electronic Design Automation (EDA) software tools — lasts approximately 24 to 36 months. The total R&D cost for a single cutting-edge chip design scales between \$500 million and \$725 million. A significant portion of this capital is absorbed by complex physical verification and static timing closure to guarantee that no hardware bugs compromise the silicon layout before committing to manufacturing.
- b) **Reticle Set Fabrication (Time and Cost):** Once the design is finalized, the layout is converted into a physical template during the "tape-out" phase. A complete mask set for a sub-3nm advanced node requires between 40 and 70 individual 6-inch reflective masks to pattern every single FEOL and BEOL layer. Writing these nanometric patterns with ultra-precise electron-beam mask writers takes 3 to 4 weeks of continuous execution in a dedicated mask shop. The total cost for a single complete master mask set amounts to approximately \$15 million to \$20 million. If a critical logic bug is discovered after production begins, a catastrophic "re-spin" is required, forcing the manufacturer to pay millions more for new masks and delaying the product launch by 6 to 9 months.
- c) **Maximum Throughput and Yield Metrics (A 25-Wafer Lot Baseline):** In mass-volume industrial manufacturing, wafers are processed in standard automated pods containing a lot of 25 silicon substrates. Operating under a native half-field layout ( $26 \times 16.5$  mm), on 300 mm wafers, each substrate accommodates exactly 135 functional exposure fields, translating to a maximum geometric capacity of 3,375 raw dies per single production lot ( $135 \times 25$ ). At a standard foundry throughput of 150 wafers per hour, (*translating to exactly 20,250 raw dies per hour*), exposing a single 25-wafer lot requires a net scanning time of just 10 minutes ( $((25/150) \times 60)$ ).
- d) **Testing, Defect Density, and Scrap Costs:** Microchip manufacturing is never defect-free, particularly during the early phases of a process node where the initial

functional yield rate sits around 50% to 60% due to microscopic dust particles or crystal dislocations. Automated testing via post-silicon wafer probing requires 24 to 48 hours per lot to electrically validate every single die on the substrate. Within a 25-wafer lot containing 3,375 raw dies, a 60% yield means that 2,025 chips are fully functional, while 1,350 defective dies are scrapped during the dicing phase. Given that a single processed sub-3nm wafer costs the fabless company between \$20,000 and \$30,000 directly from TSMC, the total manufacturing cost of a single 25-wafer lot ranges from \$500,000 to \$750,000, meaning the scrap factor implicitly increases the manufacturing cost of every functional die.

- **The Final Commercial Selling Price** [Ref.2]: The final market price of a high-end processor is a direct consequence of this steep economic pyramid. While the net manufacturing and packaging cost for a high-end sub-3nm chip configuration begins at several hundred dollars under optimal conditions, the final commercial price easily escalates to \$3,000 to over \$5,000 for complex, large-scale server processors and data center AI accelerators. This substantial commercial markup is required by fabless companies to absorb the upfront \$700 million R&D design cost, pay for the \$15 million mask sets, compensate for the 40% scrap rate during manufacturing, and fund the next generation of semiconductor engineering.
- 

## 1. The EXE:5000/5200 scanner: An Overview [Ref.3]

The latest generation of scanners costs over \$500 million (2026). It represents the most complex engineering feat in existence, born from the technological synergy between ASML (*company*), Zeiss (*optics*), and Trumpf (*lasers*). While historically accelerated by strategic capital injections from leading semiconductor manufacturers, ASML operates today as a publicly traded corporation with institutional equity funds and global asset managers holding most of its shares. Its ultra-high-cost systems are sustained through continuous capital expenditures from industry giants like Intel, TSMC, and Samsung to ensure the continuation of Moore's Law. This machine is assembled by ASML technicians directly at the customers' sites; dozens of trucks and several large transport aircraft are used to ship the various parts. Physically, an EXE:5000/5200 scanner, only it, is a parallelepiped-shaped machine approximately 14 meters long, 4 meters wide, and 10 meters high.

### 1.1. System Overview and Architecture

Physically, the scanner is organized into a distributed architecture that separates high-energy generation and transport units — located in a Remote Source Area — from the high-precision lithographic core operating at a Main Level.

**1.1.1. The Remote Source Area:** A gigantic TRUMPF CO<sub>2</sub> laser is housed in a separate enclosure, typically on a lower floor (*the sub-fab*), to isolate its intense mechanical

vibrations and thermal emissions from the sensitive optics. This assembly serves as the primary energy source and comprises two interconnected subsystems:

- **Master Oscillator Power Amplifier (MOPA)** system that generates the initial low-power seed laser radiation at a  $10.6\ \mu\text{m}$  wavelength: corresponding to a frequency of approximately 28.3 THz within the Long-Wave Infrared spectrum (*LWIR*).
- **Staged Amplification Units** that operate as a high-power "photonic pump" to multiply the beam energy.

The resulting high-energy laser beam travels through:

- **a Beam Delivery Unit (BDU)** — a complex optical pipeline approximately 20 meters long — which guides the infrared light with nanometric precision toward a **Source Chamber (EUV Generation Chamber)** located at the Main Level interface. The BDU is physically isolated from the Source Chamber to protect its laser optics from the plasma heat and debris. An off-axis parabolic copper mirror is the final optical element in the BDU at the "clean" side of the interface with the Source Chamber; it focuses the laser beam through a narrow aperture in a wall of the Source Chamber (*the protective window*) onto tin droplets (*infrared beam focalization*).

**1.1.2. The Main Level:** This is the heart of the system, housed within a cleanroom, and in turn houses: the Source Chamber, and a massive Main Vessel including complex adaptive EUV optics.

- **The Source Chamber:** A dedicated module where the incoming infrared laser beam strikes tin droplets to create a plasma. The plasma emits Extreme Ultraviolet (*EUV*) isotropic radiation at a 13.5 nm wavelength. This wavelength corresponds to an energetic frequency of approximately 22.2 PetaHertz ( $22.2 \times 10^{15}$  Hz) operating far beyond the visible spectrum and immediately adjacent to the soft x-ray domain. This chamber is mechanically and physically coupled to the BDU at one side, and to the Main Vessel at the other side, but remains a distinct environment to contain the high-energy debris.
- **The Collector Mirror:** The Source Chamber contains this massive Collector Mirror: the first in the EUV optical system. Its primary function is to collect the isotropic EUV radiation emitted by the plasma and focus it with extreme precision onto a critical physical aperture of the Source Chamber wall communicating with the Main Vessel (*isotropic EUV focalization*). At this aperture is located the Intermediate Focus (*IF*) of the EUV beam, which represents the point of maximum convergence of the EUV beam and serves as the critical interface between the Source Chamber and the illumination optics inside the Main Vessel.
- **The Shutter:** Positioned into the Source Chamber immediately before the Intermediate Focus. It is at this exact location that a shutter system (*see Chapter 9*)

is positioned to regulate the immense flow of energy. The separation between the Source Chamber and the Vessel – both communicating at the Intermediate Focus only through the shutter – is essential to prevent tin contaminants and plasma heat from reaching the nanometric precision of the system optics inside the Vessel.

- **The Main Vessel:** A massive, high-vacuum stainless-steel container featuring rounded edges and a parallelepiped-shaped structure. It measures approximately 5 to 6 meters in height and 3 to 4 meters in width. Internally, its volume is architecturally divided into two distinct zones that constantly interface with each other: a central core and a surrounding peripheral space.
  - **The Central Core:** This interior volume of the vessel, permanently kept under ultra-high vacuum, hosts the active elements of the lithographic process. It contains: a dual-unit wafer stage system in the bottom position and a reticle stage mounted upside-down at the very top ceiling of the Central Core, along with the magnetic tracks (*stators*) of an internal Maglev system on which they travel. Centrally mounted between these stages, the Central Core houses a Projection Optics Box (*POB*): a massive, ultra-stable structural housing that rigidifies and thermally isolates a complex system of reflective EUV mirrors. Said POB features open optical apertures at its top and bottom, allowing the reflected EUV beams to enter from the reticle stage suspended above, perform their unobstructed zig-zag dance inside the optics block, and exit precisely onto the wafer surface below.
  - **The Surrounding Peripheral Space:** This boundary volume houses the fixed, heavy infrastructure of the scanner. It contains massive mirror supports, active cooling manifolds, and heavy structural mountings for the Maglev tracks. Crucially, the upper section of this periphery accommodates a vibration-isolated metroframe. Said metroframe acts as an independent, rigid reference framework manufactured from Zerodur, suspended via active magnetic isolators to remain dynamically decoupled from the Vessel walls. Positioned entirely within the upper peripheral boundary, the underside of this framework supports a two-dimensional matrix of stationary photovoltaic heads, each integrating an emission laser paired with an adjacent photovoltaic sensor. To maintain the ultra-high vacuum integrity of the central core, these heads interface with the exposure volume through dedicated, ultra-pure optical windows sealed with metal ConFlat gaskets, allowing the tracking beams to project into and read reflections from the central core without physical atmospheric leakage.

**1.1.3. Auxiliary and Computational Systems:** Vacuum pumps and cooling units are distributed around the Vessel and on the lower levels. These are flanked by electrical cabinets and high-performance computing clusters (*server racks*) located externally to the

main unit, which process the massive flow of metrological data and run real-time control algorithms necessary for automated operation.

**1.1.4. Conclusions on this chapter:** All this combined effort serves a single, precise purpose: to serially reproduce a single circuit layer from a reticle onto a thin, photoresist-coated silicon disc (*a wafer*). A machine of this magnitude and complexity is engineered to operate 24 hours a day, 7 days a week, continuously for years.

---

Having established this high-level architectural framework, the following chapters will provide a deep-dive analysis of each core subsystem. We will explore the extreme physics and engineering solutions — from vacuum integrity to nanometric stage dynamics — that allow this massive assembly to operate as a single, perfectly synchronized lithographic instrument.

---

## 2. Vacuum Systems – Multi-stage Pumping Systems

All phases of the production process take place in a high vacuum, which must not be interrupted to avoid any possible contamination of the chips by human operators.

Therefore, all stages are automated. The vacuum is generated by a hierarchy of pumps that operate without lubricants (*dry*) to avoid oil vapors (1 mbar = 100 Pa):

- **Dry Pumps (*Primary*):** Screw or lobe pumps that bring the pressure to an intermediate backing level of  $10^{-1}$  to  $10^{-3}$  mbar ( $10$  to  $10^{-1}$  Pa) and allow the subsequent turbomolecular pumps to operate. Turbomolecular pumps cannot exhaust directly into the atmosphere because they would encounter too much resistance (*at high speeds, air becomes as dense as molasses for their blades*). Instead, they discharge their residual gas into a dedicated backing line managed by the primary dry pumps, which collect the molecules, evacuate them toward the exhaust port, and finally expel them into the FAB facility exhaust system.
- **Turbomolecular Pumps:** Turbine blades rotating at 90,000 *rpm* (*supersonic tip speed*) suspended on magnetic bearings to achieve an ultra-high vacuum of  $10^{-6}$  to  $10^{-8}$  mbar ( $10^{-4}$  to  $10^{-6}$  Pa). They act by mechanically "kicking" gas molecules out of the clean space and into the primary backing line. At this stage, the mean free path of the molecules (*the average distance a molecule travels before colliding with another*) extends to several meters. This condition is essential because EUV photons would be instantaneously absorbed if they encountered even a few stray air molecules.
- **Cryogenic Pumps:** Specialized condensation surfaces placed in critical areas and cooled to temperatures near absolute zero (10–20 K). These pumps trap residual gas molecules by freezing them instantaneously upon contact, allowing the system

to achieve an extreme vacuum level of  $10^{-9}$  to  $10^{-11}$  mbar ( $10^{-7}$  to  $10^{-9}$  Pa). At this threshold, a residual gas molecule can travel for kilometers before colliding with another, ensuring that the weak EUV beam reaches the wafer without undergoing any atmospheric absorption.

---

### 3. Maglev Movement and Advanced Metrology

Inside the **Main Vessel**, nanometric positioning is achieved by decoupling the propulsion units from the measurement reference:

- **Propulsion (*Maglev Motors*):** The Maglev linear motors operate directly inside the ultra-high vacuum environment of the Main Vessel. Their magnetic tracks (*stators*) are integrated onto the heavy structural floor and walls of the Vessel base (*the peripheral space*). This positioning allows them to drive the wafer and reticle stages with frictionless levitation precisely where the EUV exposures occur, while remaining dynamically isolated from the metrology reference frame to prevent vibration transfer.
- **Reference (*The Metroframe*):** The spatial reference for all stage tracking is established by the vibration-isolated metroframe positioned within the upper peripheral boundary of the Vessel (*see Section 1.1.2.*). Because this structural framework levitates via active magnetic isolators and is dynamically and physically decoupled from both the Vessel walls and the Maglev motors, it remains completely unaffected by the intense accelerations of the stages, providing the absolute, immobile coordinate system required by the lasers and photovoltaic heads to measure real-time movements.
- **The Feedback Loop:** The laser eyes on the Metroframe track the stages' positions, while the control system adjusts the magnetic fields of the Maglev motors to correct the trajectory in real-time.

#### 3.1. Magnetic Levitation Motors (*Maglev*)

Unlike traditional mechanical guides, the Maglev system provides 6-DOF – Six Degrees of Freedom control – allowing the stage to float in a vacuum with no physical contact. This is achieved through Active Gravity Compensation, where magnetic repulsion counteracts the stage weight and compensates for inertial forces, maintaining a constant gap of a few microns even during extreme dynamic transitions. While linear motors handle long-range translations, a network of Lorentz-force magnetic actuators performs nanometric fine-tuning of height (*Z-focus*) and micro-angular displacements (*Roll, Pitch, Yaw*). This frictionless architecture eliminates parasitic vibrations and wear, enabling the violent accelerations required for high-throughput production without compromising positioning accuracy.

- **Zero Stick-Slip:** The absence of physical contact allows for movement without the "jerks" typical of mechanical motors. In a vacuum, static friction does not exist. Fluid displacements on the order of a picometer ( $10^{-12}$  m) can be achieved.
- **6 Degrees of Freedom:** The system (6-DOF) controls positions ( $X, Y, Z$ ) and the three rotations around those axes via magnetic actuators that keep the wafer parallel to the optics.
- **Performance:** Accelerations up to 20G (*higher than a jet*) with instantaneous stops at the exact position without mechanical vibrations. Linear motors are essential for high productivity (*throughput*).

### 3.1.1. Active Levitation and the Dual-Carriage System

The levitation of the stages (*maintained at a clearance of a few microns*) is an active process. To generate the repulsive magnetic fields and power the fine-tuning actuators, the stages require constant electrical power, which in turn generates heat.

- **Thermal Management:** Internal micro-channels circulate liquid coolant within the stage to prevent thermal expansion.
- **The "Tension-Free" Cable Handler:** To supply electrical power and conditioning fluids without introducing mechanical tension, each individual wafer unit utilizes a dual-carriage architecture. A secondary, synchronized cable-handler carriage tracks the movements of the main wafer stage within the central core, feeding the flexible power cables and fluid hoses in a loose, constant-loop configuration. This layout ensures that the main wafer stage remains completely isolated from any parasitic pulling forces or drag, allowing it to levitate and accelerate with absolute, frictionless freedom.

### 3.1.2. Operating Cycle: Step-and-Scan

- **The wafer stage:** the wafer is held by the wafer stage, positioned at the bottom of the central core. The stage upper perimeter is equipped with high-precision measurement diffraction gratings that serve as mobile targets for the laser metrology. By reflecting the beams back to the photovoltaic sensors mounted on the Metroframe, whose reference positions are known with absolute, sub-nanometric certainty, the system constantly calculates the stage exact location. This stage is driven by a 6-DOF Maglev system that ensures nanometric alignment and flatness during the high-speed scanning process.
- **The reticle stage:** the reticle is held by a specialized stage positioned at the top of the central core. To ensure nanometric precision, the sides of the reticle stage are equipped with high-precision measurement mirrors (*or diffraction gratings*) that serve as mobile targets for the horizontal beams of the laser metrology system. By reflecting the beams back to the photovoltaic sensors mounted on the metroframe, whose reference positions are known with absolute, sub-nanometric certainty, the

system constantly calculates the stage exact location. The entire assembly is driven by a high-acceleration Maglev motor to maintain perfect synchronization with the wafer stage.

### 3.1.3. The serpentine scanning

The performance of the Maglev motors is utilized in a continuous serpentine scanning trajectory: the half-field layout is executed in a single pass, whereas the full-field layout requires two separate sequential passes across the entire wafer surface. The serpentine movements are detailed with reference to the **figures 1 to 5**. **Figure 1** is a synoptic table reporting the real dimensions of the main elements involved in the exposure and the corresponding ones after optical reduction, as well as how they are scaled on the drawings. **Figure 2** shows the serpentine scan for the half-field configuration. **Figure 3** shows the transit of an exposure field (*the die*) under a fixed projection slit. **Figure 4** is a miniaturized schematic of a simplified two-row scanning path. **Figure 5** shows the full-field two-pass serpentine scan and the resulting stitched layout. In **figure 2**, the masking process utilizes a silicon wafer of 300 mm diameter (11.8 inches, commonly referred to as “12 inch”) with a thickness of 775  $\mu\text{m}$ . able to accommodate 135 half-field ( $26 \times 16.5$  mm) and 68 full-field ( $26 \times 33$  mm) functional exposure fields.

- **Half-field serpentine scanning.** With reference to **figure 2** the scanning process depicted herein is concluded at the central upper part of the complete exposure field. The remaining unexposed singular fields represent the next fields. The position of the illumination slit that projects the reticle layout onto the dies, is fixed with respect to both the wafer moving beneath and the reticle moving above (*not represented in the figure*). The fixed position of the slit is indicated by a blue line aligned with the left side of the slit, and arrows indicate the alternate scanning directions. The dimensions of a single exposure field on the wafer are ( $26 \times 16.5$ -mm).
- **Scan Phase:** This is the moment of exposure. The Maglev system moves the wafer stage and the reticle stage along the longitudinal Y-axis in opposite directions at synchronized speeds as the pattern is "brushed" through the illumination slit. The speed of the reticle stage is eight times the speed of the wafer stage, to precisely compensate for the 8 $\times$  optical reduction along the scan direction. This phase is better detailed in **Figure 3**, wherein an exposure field completes its run under the illumination slit.
- **Step and Reversal Phase:** Once a field row is completed along the longitudinal Y-axis (*the scanning axis*), the high-speed shutter closes (*see Chapter 9*). The Maglev system then performs a very rapid orthogonal displacement along the transverse X-axis (*the "step" axis*). This movement positions the next field row

under the optics, ready for a new scan in the opposite direction along the Y-axis. The dynamics of this process is simplified in **Figure 4**.

- **Dynamic Stability:** Despite the violent acceleration during the step, the system must dampen any residual vibration within milliseconds. The subsequent scan begins only when the laser metrology confirms that the stage is perfectly stable and aligned to the nanometer.
- **Full-field serpentine scanning and stitching.** With reference to **Figure 5**, the complete scan is subdivided into two synchronized passes executed with two different reticles. **Figure 5a** shows the first pass in which the Maglev locomotion system drives the wafer stage through a continuous serpentine trajectory to project the first halves of the exposure fields using the primary reticle. This pass is substantially identical to the one shown in **Figure 2**. Once the wafer exposure is completed, the stages are stopped, the reticle is taken out of the central core and swapped with the second reticle containing the remaining layout. The Maglev system then performs a precise wafer carriage translation along the longitudinal Y-axis to apply 16.5 mm spatial offset, accommodating the initial boundary of the second half of the exposure field directly beneath the projection optics, as shown in **Figure 5b**. The second pass repeats the exact same serpentine path to print the remaining halves, which allows for covering the entire wafer surface field by field, obtaining a full stitched layout (see *Section 5.6*).

### 3.2. Advanced Metrology

Laser interferometers are the "eyes" of the system. Rather than simply tracking absolute positions, they are physically anchored to the metroframe (see *Section 1.1.2*) to establish a dynamic and absolute coordinate system.

- **The Metrological Chain:** By using the vibration-isolated metroframe as a stable reference, the system constantly measures the wafer position relative to the reticle position. This "metrological chain" ensures that even the slightest systemic vibration or thermal expansion is detected and compensated for in real-time, maintaining nanometric overlay accuracy between the reticle and the wafer.
- **Real-time Synchronization:** This setup allows the scanner to orchestrate the movement of the Maglev stages so that the projected image and the wafer surface remain perfectly synchronized during the high-speed scanning process.

#### 3.2.1. Operating Principles: Planar Optical Encoders

To maintain sub-nanometric tracking across long-range horizontal movements, the system utilizes a parallel, face-to-face optical layout that avoids any geometrical misalignment.

- **Spatial Layout:** The laser source and the photovoltaic sensor are integrated into a stationary encoder head mounted to the underside of the metroframe, projecting its narrow beam vertically downward. The encoder diffraction grating is engineered as

a flat, continuous scale layout that follows the upper perimeter of the moving stage, framing the central wafer slot. This perimeter configuration provides two-dimensional coordinate tracking (*X and Y axes*) while leaving the wafer perfectly uncovered.

- **Beam Splitting via Diffractive Reflection:** When the stationary vertical laser beam strikes the horizontal moving scale, it does not reflect as a single mirror image. Instead, the periodic nanometric structure of the grating acts as a beam splitter, diffracting the incident light into multiple symmetrical, upward-returning paths, specifically the (+1) and (−1) diffractive orders, that form a regular divergent (*V-shaped*) returning wavefront.
- **Dynamic Phase-Shift Generation:** As the Maglev system drives the stage across either the X or the Y axis, the wide diffractive scale slides continuously beneath the overhead stationary laser. This horizontal relative movement introduces a differential Doppler shift between the (+1) and (−1) diffracted beams, altering their phase relationship proportionally to the nanometric displacement of the scale pitch. This effect can be explained by decomposing the vectors representing the two diverging diffracted waves into their components along the axis of movement of the stage and the axis orthogonal to it. In this way, it becomes evident that only the vector components aligned with the axis of movement point in opposite relative directions, thereby contributing to the differential Doppler effect.
- **Recombination and Photovoltaic Detection:** The two upward-returning diffracted beams re-enter the stationary encoder head on the metroframe, where they are optically recombined. Their spatial superposition creates a dynamic interference pattern of alternating light and dark fringes on the surface of high-speed photodiodes. These photodiodes convert the light intensity variations into sinusoidal electrical signals sent directly to an FPGA. The FPGA digitizes these signals and counts the number *N* of fringes traversed to determine the target position *P* in real time using the formula:

$$P = N \cdot \frac{p}{2}$$

where *p* represents the spatial pitch of the diffraction grating, and the factor of 2 accounts for the differential combination of the (+1) and (−1) diffractive orders. These systems operate with MHz-range feedback loops for real-time stage control and GHz-range sampling for high-resolution signal interpolation. This architecture ensures that the stage position remains absolute and deterministic even during the most violent accelerations, without the narrow laser beam ever losing the target.

### 3.2.2. Insight into Wave-Quantum Optics and Metrology

The sub-nanometric performance of the planar optical encoder relies on the dual wave-particle nature of light, merging classical wave optics with quantum detection.

- **The Quantum-Doppler Shift:** As the stage moves across either the X or the Y axis, the photons interacting with the moving nanometric grid undergo a differential Doppler shift. Photons scattered into the (+1) order collide with the advancing grating structures, slightly gaining momentum and shifting to a higher energy level  $E + \Delta E$ , while photons in the (-1) order lose momentum, dropping to a lower energy level  $E - \Delta E$ . This tiny energy split manifests as a dynamic frequency difference between the two returning wavefronts governed by the fundamental Einstein relationship:  $E = h \cdot f$ , where  $h$  is the Planck's constant and  $f$  is the photon frequency.
- **The Photovoltaic detection:** The final translation of light into data is governed by the photoelectric effect. When the recombined wave pattern — carrying the energy-differential information — strikes the high-speed photovoltaic sensors on the metroframe, individual photons transfer their discrete energy quanta to excite electrons. In this context, the photovoltaic effect acts as a solid-state subcategory of the broader photoelectric effect: rather than releasing electrons into the vacuum as in the historical definition, it specifically describes how photons provide enough energy for electrons to escape the valence band and jump into the conduction band across the internal P-N junctions of the semiconductor sensor. This quantum transition generates a measurable electrical current with a beat frequency directly proportional to  $2\Delta E$ , allowing the FPGA to execute high-frequency digital interpolation and effectively track the stage at atomic scales.
- **Feedback:** The system reacts in microseconds or fractions thereof to laser metrology readings, instantaneously correcting the current in the Maglev electromagnets.
- **Multi-Axis Network:** The scanner uses several simultaneous laser beams to calculate not only the position ( $X, Y, Z$ ) of the stage but also its minimal tilt (*rotations*).
- **Mirror Map:** The machine loads a map of the microscopic roughness of the measurement mirrors (*which are high-precision diffraction gratings*) into memory to cancel any systematic error during the EUV scan. This ensures that even sub-nanometric deviations in the measurement mirrors' topography do not affect the final overlay precision.
- **Laser Autofocus:** A stationary multi-spot laser assembly anchored to the lower lateral extensions of the Metroframe projects a measurement beam across the central core at a grazing angle, skimming the wafer surface. Rather than projecting downward, this horizontal-grazing beam strikes the topography of the wafer, and its

reflection is captured by a corresponding photovoltaic sensor on the opposite side of the framework. By detecting the lateral shift of each reflected beam, the system maps the exact 3D topography point-by-point, enabling real-time Z-focus and tilt corrections to maintain the perfect depth of focus.

- **TwinScan Technology:** The dual-unit wafer stage system (see Section 1.1.2.) maximizes throughput by operating two independent wafer exposure units in parallel. Each unit is an autonomous system consisting of a main wafer stage and its dedicated, synchronized power-slave carriage. While one unit executes the EUV exposure cycle, the second unit is mapped in 3D with nanometric precision in an adjacent measurement area. Once the exposure is finished, the Maglev system swaps the locations of both integrated units in less than a second along coordinated orbital paths, preventing their respective cable bundles from tangling. The scanner instantly applies the previously saved topographic map to begin the new exposure cycle without calibration delays.
- **Parallelism (*Dynamic Control*):** Parallelism relative to the last mirror of the projection system ensures the image is not distorted; if the wafer were tilted (*even by a few microradians*), the pattern would be printed with geometric distortion (*a square would become a trapezoid*). The 6-DOF actuators constantly correct tilt (*roll and pitch*) to ensure the wafer is perfectly orthogonal to the EUV beam.

**3.2.3. Measurement precision:** Each complete fringe detected by a photovoltaic sensor corresponds to a displacement equal to half the wavelength of the Helium-Neon (He-Ne) measurement laser  $\approx 316$  nm.

- **Phase Analysis:** To descend below the 316 nm limit, the electronics do not merely count the fringes; they measure the intensity of the intermediate sinusoidal brightness with extreme precision.
- **Wave Subdivision (*Interpolation*):** By converting the light intensity into a sinusoidal electrical signal, the system can electronically subdivide each individual wave into thousands of parts (*interpolation*). This process allows the detection of infinitesimal displacements, reaching a resolution of 10 to 20 picometers (0.01 to 0.02 nm), a dimension smaller than that of a hydrogen atom.
- **Summary:** Parallelism ensures the correct shape of the circuits (*no distortion*), while the focal length (*guided by the 3D map*) ensures sharpness (*focus*). The 6-DOF stage is the actuator that, reading the 3D map, tilts and lifts the wafer in real-time as it streaks beneath the beam.

**3.2.4. Alignment Marks: The Diffraction "Fingers"**

- **Diffraction Geometry:** Unlike simple visual indicators, Alignment Marks are microscopic diffraction gratings etched into the Scribe Lines. When the alignment

laser (*typically a multi-wavelength laser*) strikes these marks, the periodic structure of the grating reflects the light into specific diffraction orders.

- **The Phase-Shift Principle:** The scanner sensors analyze the phase and intensity of these diffracted beams. By comparing the signals from different wavelengths, the system can calculate the mark exact center with sub-nanometric precision, even if the mark is partially buried under layers of opaque material or distorted by the CMP (*Chemical Mechanical Polishing*) process.
- **Dynamic Calibration:** Upon every new wafer entry via the load-lock chamber, the machine performs a "Global Alignment":
- **Search:** The sensor quickly scans several primary marks to define the wafer overall orientation (*tilt and rotation*).
- **Fine Tuning:** It then measures a specific set of marks for each chip to compensate for local deformations (*linear expansion or "grid-warping"*) caused by previous thermal or chemical treatments.
- **From Marks to Maglev:** Once the sensors "lock" onto the marks, the global coordinate system is handed over to the Maglev laser interferometers. These interferometers, anchored to the stable Metroframe, guide the stage across the chip area, ensuring that the new pattern is superimposed on the previous one with a tolerance often lower than 1 nm (*Perfect Overlay*).

*Technical Note: To further enhance precision, the alignment sensor utilizes circularly polarized laser light. This choice is critical: it allows the system to remain insensitive to the orientation of the marks while effectively filtering out parasitic surface reflections. By analyzing the phase shift of the circularly polarized light reflected by the diffraction grating, the scanner can 'see' through the overlying material layers, accurately identifying the mark center even under challenging process conditions.*

---

## 4. The EUV Optical System

In EUV lithography, "light" cannot pass through glass lenses. The optical system, developed by Zeiss, is based exclusively on reflective mirrors that perform a selective reflection of the light generated by a tin plasma struck by a powerful CO<sub>2</sub> laser beam.

### 4.1. Multilayer Mirrors (*Bragg Gratings*)

These components are three-dimensional diffraction gratings consisting of a "sandwich" of 40-50 pairs of alternating layers of Molybdenum (Mo) and Silicon (Si). The operation of Bragg mirrors is based on the difference in electron density between the materials that compose the "sandwich" at the atomic level:

- **Silicon:** Having low electron density and a refractive index near unity, it acts as a "transparent" space or spacer, allowing light to penetrate into the subsequent layers.
- **Molybdenum:** Due to its higher density, it acts as a diffraction plane or reflector, providing the necessary optical contrast.
- **Protective Mirror Capping Layers:** A sub-nanometric protective coating of Ruthenium (Ru) is deposited as the final, chemically inert boundary layer over the multi-layer molybdenum-silicon (Mo/Si) Bragg structures of every mirror. This ultra-thin passivation layer acts as a permanent chemical shield to prevent surface oxidation and degradation caused by outgassing or active cleaning agents, ensuring long-term structural stability without diminishing the extreme ultraviolet (EUV) reflectivity of the underlying optics.

**4.1.1. Resonance and Constructive Interference:** Since the reflection from a single interface is extremely weak, the principle of the Distributed Bragg Reflector (*DBR*) is exploited, governed by Bragg's Law:  $n\lambda = 2d \sin \theta$ . Here,  $n$  is the refractive index and  $\theta$  is the angle of incidence relative to the layers. At the EUV wavelength ( $\lambda = 13.5 \text{ nm}$ ), the multilayer is engineered for near-normal incidence. The thickness of each pair  $d$  is precisely calibrated so that the path difference between reflections from successive layers leads to perfect constructive interference. However, as the Numerical Aperture (*NA*) increases in the EXE series, the range of incident angles becomes wider. This creates a "mismatch" with the Bragg condition across the mirror surface. To compensate, Zeiss uses "graded multilayers" where the thickness  $d$  varies across the mirror geometry to ensure that the resonance condition  $n\lambda = 2d \sin \theta$  is met at every point, despite their varying angles.

- **Spectral Bandwidth and Filtering:** The Mo/Si multilayer acts as a highly selective band-pass filter. Due to the low refractive index contrast between Molybdenum and Silicon at EUV frequencies, the reflective bandwidth is extremely narrow (approx. 2% of the central wavelength). This means the mirrors only reflect a specific "slice" of the emission spectrum generated by the tin plasma, centered at 13.5 nm. While this limits overall efficiency, it provides an essential benefit: it inherently filters out out-of-band radiation (*such as DUV or infrared*), preventing unwanted wavelengths from reaching the wafer and causing parasitic heating or blur.
- **Reflective Efficiency and Absorption:** The system reaches an overall efficiency of approximately 70% (*achieved by Zeiss*). Beyond 50-60 layers, the materials themselves (Mo/Si) absorb lighter than they can reflect, reaching a "saturation point" where adding more layers becomes counterproductive. The remaining 30% of the EUV radiation is absorbed by the materials and converted into heat, posing a massive thermal management challenge for the entire optical column.

- **Atomic Precision:** The surface tolerance of the individual layers must be absolute. To provide a scale, if the mirror were as large as Germany, the maximum imperfection would be only 1 mm. Any "atomic" roughness would cause light scattering (*flare*), drastically reducing image contrast and resolution.

*Technical Note on Graded Multilayers: In a theoretical mirror, or for small angles of incidence, the thickness of the layers is uniform across the entire surface. However, in High-NA mirrors — where the curvature is extreme and the angle of incidence  $\theta$  varies significantly from the center to the edges — Zeiss employs Graded Multilayer technology. In this case, it is the thickness of the entire "stack" (the period  $d$ ) that varies laterally across the mirror surface. This means that at any given point on the mirror, the layers are tuned to the specific local angle of incidence to satisfy the Bragg condition  $n\lambda = 2d \sin \theta$ . Without this lateral variation in thickness, the edges of the mirror would "fall out" of resonance, failing to reflect the EUV beam.*

#### **4.2. The Optical Path: From Plasma to Chip**

To correctly visualize the optical path, the system must be divided into two major functional sections: the Illumination System (*which shapes the light*) and the Projection System (*which prints the image*). The reticle is located exactly "in the middle," on the ceiling of the Central Core, acting as the watershed between these two sections.

After the Trumpf laser strikes the tin droplet and generates the EUV radiation, the light is "chaotic" and emitted in all directions. The optical train is arranged as follows:

- **Collector Mirror:** This is the first mirror of the EUV optical path, located inside the Source Chamber. It collects the isotropic light from the plasma and focuses it onto the narrow point called the Intermediate Focus (*IF*). The Collector Mirror is the largest one in the optics, featuring a massive, bowl-shaped body (*approx. 65 cm in diameter*). It is a Graded Multilayer Bragg Mirror, designed with variable-thickness coatings to ensure maximum EUV reflectivity across its entire curved surface. Its central aperture allows the infrared laser beam to pass through and strike the tin droplets falling in front of it. The substrate of multilayer is made of materials with ultra-low thermal expansion, featuring an internal network of water-cooling channels to dissipate the kilowatts of plasma heat.
- **Illuminator Mirrors:** After passing through the IF, the radiation enters the central core of the Main Vessel. Here, the light typically passes through 2 or 3 specialized mirrors (*the Field Facet Mirror and Pupil Facet Mirror*). These mirrors serve to collimate the beam and provide the correct shape and uniformity before it strikes the reticle. Consequently, there are approximately 3 to 4 reflections before the light reaches the reticle.
- **The Reticle (*The heart of the system*):** The radiation strikes the reticle at an angle of  $9^\circ$  relative to the optical axis. The reticle (*see Section b*) is a reflective multilayer

component containing the "design" of the chip layer. The light bouncing off it now carries the circuit information.

- **Projection Optics Box (POB):** Positioned centrally within the vacuum Central Core between the upper reticle stage and the lower wafer stage (*see Section 1.1.2.*), the POB is a massive, ultra-stable structural housing maintained at a constant temperature within a thousandth of a degree. The EUV light from the Illumination System is directed upward to strike the reticle floating above the POB. Once the beam has "read" the reticle, it reflects downward and enters the POB through an open input aperture, which is completely free of glass to prevent EUV absorption. Inside this thermally isolated shell, the beam is reduced in size and precisely projected onto the wafer through a series of ultra-high-precision aspherical mirrors produced by Zeiss. These mirrors are arranged in an unobscured zig-zag pattern, and the final element is a concave mirror that focuses the circuit image through an open output aperture directly onto the wafer below. In Standard-NA (0.33) machines, the POB typically contains 6 projection mirrors; in High-NA (0.55) machines, complexity increases, requiring 8 or more mirrors. Throughout the entire scanner, the beam undergoes a total of 10 to 12 reflections. The combined effect of these reflections is equivalent to a single "ideal objective" with a resulting focal length  $f$  capable of focusing the circuit pattern with nanometric resolution.
- **The Optical Efficiency Paradox:** This high number of mirrors introduces a monumental energy challenge: in a chain of 10 mirrors, the useful power reaching the wafer is only about 2% to 3% of the energy initially generated ( $0.7^{10} \approx 0.028$ ). To ensure enough light to print chips at industrial speeds, the initial source must be incredibly powerful, necessitating the use of the multi-kilowatt Trumpf laser systems.

### 4.3. The REMA Unit: Dynamic Exposure Framing.

#### 4.3.1. Structural Description

The REMA (*Reticle Masking*) unit is an adjustable optical framing apparatus positioned within the illumination system, located in close proximity directly beneath the upside-down reticle stage plane (*see Section 1.1.2.*). To prevent ambiguity, it must be clarified that the REMA unit is not the electronic mask (*the reticle*) containing the circuit design; rather, it is a system of mobile mechanical blades designed to frame the EUV light slot, ensuring the radiation isolates exclusively the active exposure field on the suspended reticle during the scan.

- **A Four-Blade Aperture Mechanism:** A matrix of four independent, razor-sharp motorized framing blades arranged in opposing pairs (*two governing the X-axis and two governing the Y-axis*). These blades do not block the global beam but intersect its peripheral edges to define a rectangular exposure field.

- **High-Bandwidth Linear Actuators:** Dedicated high-acceleration voice-coil motors attached directly to each individual blade are engineered for microscopic, high-frequency positioning adjustments.
- **The Slit Optic Output:** As previously noted in **Introduction b**, the geometric clearing between the four blades shapes the unformed EUV light into a narrow, slightly curved slit. To establish the standard operational layout, the width of this illumination slot along the longitudinal Y-axis is configured by the blades at 40 mm on the mask scale, while its length remains fixed at 104 mm to span the entire reticle. Through the asymmetric reduction of the projection optics (4x in X, 8x in Y), this massive slot is compressed onto the wafer plane to a final scanning strip measuring exactly 5 mm in width, and exactly 26 mm in length. Crucially, this scanning width remains perfectly constant during the active exposure passes of a given lot to guarantee absolute dose uniformity. To structurally maintain this spatial precision and endure the relentless energy absorption along the framing edges, the internal body of each blade integrates an independent network of water-cooling micro-channels operating with laminar flow, keeping the blade profiles geometrically stable and preventing thermal expansion. This standard 5 mm projected slit configuration remains architecturally identical across both native layouts and large-scale stitched fields (26 × 33 mm). Under the stitched operational mode, the optical system maintains this exact 5 mm aperture during the first global pass to expose the initial 16.5 mm of the die layout before the EUV radiation is blanked at the stitching boundary. Following the mechanical reticle swap and the stage translation, the second global pass utilizes the same 5 mm slit to selectively expose the remaining half, ensuring a seamless chemical and physical junction at the cross-over interface.

#### 4.3.2. Functional and Kinematic Description

The REMA unit dynamically restricts the exposure field boundaries in real-time coordination with the scanning stages to guarantee dose uniformity:

- **Synchronized Boundary Tracking:** As the Maglev system drives the reticle stage during a scanning pass, the REMA linear actuators drive the blades to continuously "shadow" and track the opaque borders (*black borders*) of the reticle. It must be clarified that these mechanical blades do not modulate the high-frequency 50 kHz laser pulses, nor do they shutter the entire slit during active exposure. Instead, their kinematic tracking acts as a moving framing aperture that opens progressively as the die enters the fixed slit and closes as it exits. Because the wafer stage translates at a perfectly constant velocity while the stationary illumination field sweeps over the photoresist layer, every point of the photoresist accumulates an identical number of energy pulses during its transit, ensuring absolute dose uniformity across the entire circuit area (*see Section 6.4.2.*) while preventing any stray light from bleeding into adjacent, previously exposed fields on the wafer. This

kinematic tracking ensures that the EUV slit exposes only the active circuit die, blocking any stray light from bleeding into adjacent, previously exposed fields on the wafer.

- **Pulse-Burst Stabilization:** To prevent partial-dose defects at the field edges, the blade movements are mathematically locked to the 50 kHz pulse rate of the TRUMPF drive laser. The control system calculates the blade velocities so that no blade edge is "mid-transition" across a circuit boundary during an active EUV burst. This strict synchronization prevents the clipping of individual pulses.
- **Banding and Stroboscopic Elimination:** By eliminating pulse clipping at the borders, the system cancels out stroboscopic illumination anomalies. This structural control prevents banding-unwanted periodic exposure patterns on the photoresist caused by local dose non-uniformity. This coordinated action achieves a razor-sharp boundary definition for the exposed field, eliminating any partial-dose transition defects. At the same time, the robust metallurgical design of the blades safely absorbs and dissipates the extreme thermal load concentrated along the edges of the narrow slit.

#### 4.4. Reflective Reticle Projection:

The reticle is also a multilayer mirror (*diffraction grating*). The pattern is formed by an absorber (*often Tantalum-based*) that "switches off" the reflection where no circuit should be present.

- **The Angle of Incidence:** The increased 0.55 NA requires larger mirrors, leading to tighter spatial constraints within the system. To prevent beam clipping and ensure a clear separation between the illumination and projection paths, the chief ray angle at the reticle was increased to 9°. This is achieved by precisely tilting the final mirror of the illumination system that projects the light onto the reticle.
- **The Shadowing Effect:** Caused by the thickness of the absorber. (*This aspect will be discussed in detail in Chapter 5*).
- **Real Image Formation:** For the circuit to be printed onto the photoresist, the projected image must necessarily be real and inverted. In a system of mirrors, this occurs when the object (*the reticle*) is located at a distance greater than the focal length  $f$  of the final mirror/objective. The chip (*wafer*) acts as the "screen" placed on the image plane.
- **Lens Equation and Magnification:** The reduction ratio  $M$  (4:1 or 8:1), defined as:

$$M = \frac{q}{p}$$

(where  $p$  is the optical distance from object to system and  $q$  is the distance from system to image), dictates that the reticle must be optically 4 or 8 times further from the optical center compared to the wafer.

- **Focus and Depth of Focus:** Given the extremely high Numerical Aperture (*NA*), the Depth of Focus (*DOF*) is very shallow, in the range of a few tens of nanometers. Any slight variation in the position of the reticle or wafer along the optical axis *Z* would result in a scale error or unacceptable blurring. The system must therefore compensate in real-time for the microscopic irregularities of the wafer during scanning to keep the image plane perfectly coincident with the photoresist layer.

#### 4.5. Mirror Manufacturing: Ion Beam Sputtering (*IBS*)

To satisfy the extreme physical constraints described above, mirror production cannot rely on standard deposition techniques. Zeiss utilizes Ion Beam Sputtering (*IBS*) in ultra-pure vacuum chambers:

- **Atomic Vaporization:** A high-energy ion beam strikes a target of pure material (*Molybdenum or Silicon*), "dislodging" individual atoms.
- **Layer-by-Layer Growth:** The vaporized atoms deposit onto the thick substrate (*made of Zerodur or ULE*), forming a uniform film. The growth is not timed but monitored by an interferometric laser control system that measures thickness in real-time.
- **Atomic Precision:** This process allows for deposition control with a tolerance smaller than the radius of a single atom, ensuring that the "pitch" of the Bragg grating is identical across the entire mirror surface.

#### 4.6. Global Stability and Mirror Integrity

In the EUV scanner, maintaining the shape and position of the mirrors is a multidisciplinary challenge. Despite advanced Bragg grating technology, each mirror absorbs approximately 30% of the incident light, converting it into heat. Without extreme control, this energy, combined with mechanical vibrations, would make nanometric precision impossible.

##### 4.6.1. Zero-Expansion Substrates

The massive "body" of the mirror, which determines its curvature and supports the Mo/Si layering, must be immune to temperature variations.

- **Materials:** To prevent deformations (*even of a few picometers*), special glass-ceramic substrates such as Zerodur (*produced by Schott*) or ULE (*Ultra-Low Expansion*) are used.
- **Properties:** These materials are designed to have a near-zero coefficient of thermal expansion. This ensures that the mirror supporting structure remains dimensionally stable while the cooling system and actuators manage the dynamic thermal load from the laser beam.

##### 4.6.2. Cooling, Stability, and Active Isolation

Managing mirror stability in a vacuum requires a synergy between thermal and mechanical control to cancel out vibrations and deformations. All active sub-systems are anchored directly to the internal framework of the POB (see *Section 4.2*):

- **Micro-Channel Cooling:** To dissipate absorbed heat without inducing mechanical jitter, cooling fluids circulate through microscopic channels manufactured within the mirror substrate. This process utilizes a perfectly laminar flow entirely free of turbulence.
- **Active Thermal Compensation:** Thousands of miniaturized heating and cooling elements located on the back of each mirror correct its shape in real-time. This localized thermal management compensates for micro-expansions induced by the high-power EUV energy.
- **Suspension and Adaptive Isolation (6-DOF):** The Zeiss mirror supports combine low-stiffness passive suspensions (*which bear the static weight load*) with an active Lorentz-force isolation system. The latter uses cooled magnetic actuators controlled by a dedicated internal laser metrology. This internal metrology continuously cross-references its coordinates with the absolute coordinate system of the metroframe (see *Section 1.1.2*). This architecture ensures complete decoupling from any structural vibrations of the main scanner frame, meaning each mirror remains dynamically locked to the absolute metrological reference of the machine while being physically isolated.

#### 4.6.3. Adaptive Optics and Aberration Correction

At the nanometric level, every single reflection inevitably introduces wavefront distortions. The system maintains absolute optical sharpness through two levels of structural and dynamic correction:

- **Complementary Design:** The optical train utilizes a chain of aspherical mirrors featuring opposing curvatures (*concave and convex*). Each mirror profile is mathematically optimized to cancel the wavefront errors introduced by the preceding reflections.
- **The Deformable Mirror Apparatus:** To eliminate dynamic aberrations (*such as astigmatism or coma*) and ensure perfect image flatness, the projection mirrors — specifically the final two elements in the optical chain housed inside the POB (see *Section 4.2*) — are engineered as adaptive structures. The capability to reshape the reflective boundary does not stem from substrate compliance but exploits the infinitesimal elastic deformability of ultra-rigid materials like Zerodur or ULE. Hundreds of specialized piezoelectric actuators are mounted in a dense matrix directly behind the mirror substrate, anchored to the internal rigid frame of the POB. These actuators are designed to exert localized mechanical micro-stresses —

applying both push and pull forces via a stable pre-loaded constraint system — inducing controlled surface deformations restricted to a few nanometers.

- **Phase Shift Management:** When a portion of the EUV beam undergoes a phase delay relative to the rest of the wavefront, the piezoelectric actuators induce targeted convex or concave micro-curvatures. This local correction alters the optical path difference, ensuring that all photons arrive at the wafer surface perfectly synchronized in phase, maintaining sub-nanometric node sharpness. Because the stress remains strictly within the material elastic limit, the substrate instantaneously snaps back to its nominal geometry when the current ceases, ensuring total process reversibility.
- **ILIAS Metrology Feedback Loops:** The operational kinematics of the adaptive mirrors combine predictive algorithms and real-time metrology to dynamically alter the mirror curvature within the linear elastic regime of the Zerodur. To validate and refine these predictive models, the scanner utilizes wavefront sensors (*such as Shack-Hartmann configurations*) and dedicated interferometric systems integrated directly onto the wafer stage: the ILIAS system (*Integrated Lens Interferometer At Scanner*). The ILIAS sensors periodically capture the projected image quality at the wafer plane, feeding high-frequency calibration data back to the processing units to continuously remodel the wavefront against thermal distortions or micro-vibrations, ensuring a flawless real image of the chip.

#### 4.6.4. Predictive Feed-Forward Thermal Compensation and Illuminator Orchestration

The extreme power density of the EUV radiation introduces severe, non-uniform thermal degradation across the optical path that must be dynamically counteracted. To guarantee nanometric wavefront integrity, the system implements an advanced predictive strategy that coordinates the static configuration of the illumination optics with the dynamic stabilization of the projection mirrors:

- **Illuminator Geometry and Static Configuration (The Asynchronous Loop):** Long before active exposure begins, the specific illumination mode — such as dipole, quadrupole, or annular configurations — is mathematically determined via Source-Mask Optimization (SMO) during the design phase. At the launch of a given production lot, the centralized computational infrastructure (*described in Chapter 11*) loads these predetermined coordinate maps. Localized field-programmable gate arrays (FPGAs) translate this data into exact electrostatic voltages, micro-positioning the microscopic MEMS actuators of the Pupil Facet Mirror. This architectural setup splits and organizes the raw EUV beam received from the Field Facet Mirror, establishing a permanent, structurally stable illumination envelope tailored to the diffraction marks of the specific reticle layout. This phase operates on

an asynchronous, non-real-time control loop, locking the facet configurations for the entire duration of the lot.

- **The Dual-Zone Thermal Problem (Reticle Absorption vs. Mirror Reflection):**  
Once the scanning sequence initiates at 1.2 m/s, the machine faces a highly localized thermal load that splits into two distinct thermodynamic environments across the optical system:
    - **The Reticle Thermal Loading:** The physical circuit patterns written on the mask are defined by a high-density metallic absorber layer. When hit by the EUV slot, these absorbing zones block the radiation by design, transforming nearly 100% of the intercepted photonic energy into localized heat. This induces microscopic thermal expansions directly on the reticle substrate.
    - **The Projection Optics Box (POB) Mirror Loading:** Conversely, the clear, unpatterned areas of the reticle (*the bare molybdenum-silicon multilayer*) reflect the maximum available EUV energy downward into the projection objective. Consequently, these high-energy reflected beams carry the architectural pattern into the POB, selectively striking subsequent mirror surfaces. Despite the ultra-high reflectivity of the Zeiss Bragg mirrors, the non-reflected residual energy (approximately 30% to 35%) is absorbed directly by the mirror faces, generating moving thermal "hot spots" that mirror the layout density of the mask.
  - **Predictive Feed-Forward Computation (The Dynamic Loop):** To mitigate these dual-zone distortions before they degrade the overlay precision (*see Section 5.6.1.*), the control software executes an advanced predictive feed-forward algorithm. By real-time mapping of the reticle layout, the computer calculates the exact thermal contrast between the absorbing zones on the mask and the corresponding high-energy reflections entering the POB. The supercomputer predicts precisely where the localized thermal load will induce micro-expansions on both the mask and the subsequent mirror surfaces on a millisecond scale. Before these nanometric physical distortions can manifest, the system applies compensating voltages to the embedded piezoelectric actuators behind the ultra-rigid mirror substrates (*such as Zerodur or ULE*), dynamically reshaping the optical profiles. This real-time loop is continuously cross-checked against calibration data from the in-situ ILIAS metrology sensors described in Section 4.6.3, ensuring absolute wavefront aberration correction during active scanning.
-

## 5. High-NA Geometry: Anamorphic Optics and Stitching

In High-NA machines (0.55 NA), such as the ASML EXE:5000, the increase in resolution has necessitated a radical change in projection geometry to manage the shadowing issue.

### 5.1. The Problem: Angle of Incidence and Shadowing

In EUV lithography, the reticle acts as a mirror. For optical reasons, light cannot strike it perpendicularly ( $90^\circ$  to the surface) because the reflected beam would travel back into the source. Therefore, an inclined angle of incidence is required:

- **0.33 NA Systems:** 6-degree angle of incidence.
- **0.55 High-NA Systems:** To accommodate the larger Numerical Aperture without the beams interfering with each other, the angle must increase to approximately 9 degrees.
- **Shadowing Effect:** Since the circuit patterns (*absorbers*) on the reticle are in relief, this  $9^\circ$  angle casts a significant shadow. If this shadow is too long, it masks adjacent circuit details, causing fatal printing errors and loss of resolution.

### 5.2. Conceptual Formula for Shadow Length (L)

$$L_{\text{shadow}} = h \cdot \tan(\theta),$$

where:

$h$  = height of the relief (absorber) on the reticle (*approx. 70 nm*).

$\theta$  = angle of incidence of the light.

#### Example and comparison:

With  $\theta = 6^\circ$  (*Standard EUV*),  $L_{\text{shadow}} \approx 7,3$  nm.

With  $\theta = 9^\circ$  (*High-NA*),  $L_{\text{shadow}} \approx 11,1$  nm.

On a 2 nm technology node or smaller, an 11 nm shadow is catastrophic as it physically exceeds the dimensions of the features being printed. Since the light arrives tilted in the ZY plane (*where Y is the scanning direction*), the shadow is cast exclusively along the Y-axis. This directional distortion is the fundamental reason why anamorphism — asymmetric magnification — is required exclusively in the Y direction.

### 5.3. The Solution: Anamorphic Optics

To mitigate the 11 nm shadow without drastically reducing the thickness of the reticle absorber — which would otherwise compromise absorption performance and image contrast — Zeiss developed an asymmetric optical system. Instead of a uniform reduction, the High-NA scanner applies two different magnification factors:

- **X-Axis:** Maintains a standard 4x reduction.
- **Y-Axis (*Scanning Direction*):** Increases the reduction to 8x. By shrinking the image twice as much along the Y-axis, the optical system effectively "compresses" the elongated shadow back into a nanometric scale that is compatible with the 2 nm node requirements. This dual-magnification approach is the core of anamorphic

lithography, allowing the system to achieve high resolution while managing the physical constraints of the 9-degree incidence angle.

#### 5.4. Reticle Design

Anamorphic optics require a deliberate deformation of the layout:

- **Stretching (OPC):** The computer designs the chip by stretching the entire pattern along the Y-scanning axis. On the reticle, squares appear as rectangles twice as long, and circles appear as ovals.
- **Anamorphic Scanning:** During projection, the Zeiss mirrors — characterized by a complex ellipsoidal and aspherical geometry — reduces the rectangle eightfold in the Y direction while maintaining a fourfold reduction in X. This asymmetric curvature allows the system to restore a perfect square on the silicon wafer. Effectively, the mirrors act with different focal lengths for each axis, precisely compressing the image only where the shadow distortion is present.

*Technical Note: The scanner optical reduction mitigates positional errors only, such as reducing a 4 nm reticle displacement to 1 nm on the wafer but cannot correct diffraction-related issues or Line Edge Roughness (LER). Because EUV lithography operates at the physical limit of resolution (6 nm Rayleigh limit), jagged reticle edges or imperfect absorbers produce anomalous diffraction that optics cannot correct.*

#### 5.5. Anamorphic Geometry, Field Width, and Kinematic Synchronization

To prevent any spatial blurring or stroboscopic distortion during exposure, the mechanical translation of the stages must be perfectly synchronized with the optical reduction factors of the anamorphic lens system 4x along the transverse X-axis and 8x along the longitudinal Y-axis. This optical asymmetry fundamentally dictates the physical layout dimensions and the velocity ratios of both the reticle and wafer carriages on the horizontal plane, building upon the baseline reticle and slit specifications previously established:

- **The Transverse Dimensions (X-axis):** Subject to the 4x optical reduction, the active reticle design length of 104 mm (*established in Introduction b*) projects a static field length of exactly 26 mm onto the wafer plane, defining the unchanging lateral boundary for both half-field and full-field configurations.
- **The Longitudinal Dimensions and Scan Width (Y-axis):** Along the longitudinal Y-axis, the maximum active mask pattern of 132 mm is reduced by the 8x optics to a final native half-field width of exactly 16.5 mm on the silicon wafer (*see Ref. b2.1.1*). To sweep this area, the REMA unit projects the illumination slit whose width — as detailed in Section 4.3.1 — is configured at 40 mm on the mask scale (*compressing to 5 mm on the wafer*).
- **Kinematic Symmetrical Timing:** To maintain a completely stationary, razor-sharp image projection on the moving photoresist, the greater relative velocity of the reticle stage must precisely compensate for the reduction ratio. During a standard

exposure pass, the wafer carriage translates at a constant scanning velocity  $V_{\text{wafer}}$  of 1.2 m/s (1200 mm/s) through its 5 mm slit. Simultaneously, the reticle carriage must travel along the longitudinal Y-axis at a velocity  $V_{\text{reticle}}$  of 9.6 m/s (9600 mm/s) — exactly 8 times faster than the wafer.

As a deterministic result of this velocity matching, the time required for the active mask layout to pass completely under the REMA slit matches to the microsecond the time required for the corresponding field area of the wafer to pass beneath the projected illumination strip. Using the baseline configuration of a 40 mm reticle slit (which projects a 5 mm strip on the wafer), the symmetrical transit times are calculated as follows:

$$t_{\text{half-field}} = \frac{16.5 \text{ mm}}{1200 \text{ mm/s}} = 0.01375 \text{ s} \quad (13.75 \text{ ms})$$

$$t_{\text{reticle}} = \frac{132 \text{ mm}}{9600 \text{ mm/s}} = 0.01375 \text{ s} \quad (13.75 \text{ ms})$$

The absolute synchronization ensures that the active mask window passes entirely under the radiation source in exactly 13.75 ms meaning the full field is swept seamlessly without any localized dose non-uniformity or focal distortion. In the case of full-field stitched printing  $26 \times 33 \text{ mm}$ , the scanner does not alter this kinematic velocity or compress the optical slit; it simply executes this identical 13.75 ms global scanning sequence across two separate passes utilizing two distinct masks to seamlessly seam the two contiguously patterned halves.

### 5.6. Stitching: Nanometric Seaming

The 8x reduction in the Y direction consumes the physical space of the reticle (*standard 6-inch reticle*) twice as fast as a 4x reduction. Consequently, this halves the maximum exposure area on the wafer:

- **0.33 NA Scanner:** 26 x 33 mm area (*full-field*).
- **0.55 NA (High-NA):** 26 x 16.5 mm area (*half-field*).

If a chip — such as a large AI GPU or a high-performance processor — exceeds the 16.5 mm limit, it must be printed in two distinct passes joined via a process known as stitching.

#### 5.6.1. Strategies for Perfect Overlap:

- **Picometer Metrology:** Maglev motors, guided by laser interferometry, ensure that the start of the second pass coincides with the boundary of the first with sub-nanometric precision.
- **Stitching/Overlay Zone:** The two half-images are slightly overlapped in a transition zone. EDA (*Electronic Design Automation*) algorithms design "comb-like" or

"interlocking" patterns so that circuit interconnections join seamlessly, even in the presence of infinitesimal displacements.

- **Dynamic Correction:** The scanner detects alignment marks from the first pass and applies real-time deformations to the second image. This compensates for any thermal expansion of the wafer caused by the intense energy of the EUV radiation during the first exposure.

**5.6.2. Wafer Layout:** The exposure for large chips occurs in two distinct and sequential phases, utilizing two separate reticles sequentially loaded onto the single reticle stage through the internal vacuum storage system fed by the load-lock chamber:

- **First Pass:** The first half of the circuit is printed across all fields of the wafer using the first reticle. At this stage, thousands of interconnection lines end abruptly at the boundary of the exposure field.
- **Second Pass:** Once the system automatically swaps the reticles on the stage, and after a complete stage displacement along the Y-axis — equivalent to half the scanning field (*approximately 16.5 mm*) — the second half is printed.
- **Seamless Completion:** Thanks to the extreme alignment guaranteed by the laser interferometers and the Maglev system, the interconnection lines from the second pass perfectly meet those of the first. This "stitching" along the X-axis boundary physically completes the circuit without any measurable discontinuity.

**5.6.3. The Economic Paradox:** While stitching allows to produce chips larger than the 16.5 mm half-field, it introduces a significant economic challenge. Printing a large chip in two steps doubles the exposure time for that specific layer, thereby reducing the machine overall throughput (*wafers per hour*). Manufacturers must therefore balance the need for massive "monster chips" against the significantly higher cost per transistor that stitching entails.

---

## 6. Plasma Generation: The Tin "Target"

The production of EUV light requires a source capable of generating high-energy plasma. ASML's technological choice fell on the interaction between a high-power CO<sub>2</sub> laser and molten tin droplets for two fundamental reasons:

- **Optimal Absorption and Conversion:** The wavelength of the CO<sub>2</sub> laser (10.6 μm) is ideal for ionizing tin and inducing the specific ionic states from Sn<sup>8+</sup> to Sn<sup>12+</sup> that emit a sharp spectral peak at 13.5 nm. Unlike other elements, tin concentrates most of its emission within this "in-band" frequency, maximizing the Conversion Efficiency (*CE*) from laser energy to EUV light.
- **Industrial Scalability:** The system is designed for high-volume manufacturing (*HVM*), allowing the source to maintain a power output of several hundred watts of

EUV radiation continuously (*high duty cycle*). This is made possible by the rapid and reliable delivery of tens of thousands of tin droplets per second.

### 6.1. MOPA Architecture (*Master Oscillator Power Amplifier*) [Ref.5]

The laser beam is not generated instantaneously at maximum power but follows a staged amplification path:

**6.1.1. The Master Oscillator:** The Master Oscillator is a low-power (*typically a few Watts*), sealed-cavity CO<sub>2</sub> laser. It serves as the high precision “seed” for the entire chain.

- **Molecular Source and Wavelength:** By exciting a gas mixture containing carbon dioxide CO<sub>2</sub>, it generates radiation at 10.6 μm, located in the Long Wave Infrared (*LWIR*) spectrum. While invisible to the human eye, this pure thermal radiation is perfectly tuned to be absorbed by the molten tin, acting as the ideal “match” to ignite the plasma.
- **Geometric Perfection:** The beam is generated with a diffraction limit near unity ( $M^2 \approx 1$ ). This represents a near-perfect Gaussian intensity profile and a spherical wavefront, which is essential for focusing the beam onto a tiny 30-micron droplet after a 20-meter journey through the Beam Delivery Unit (*BDU*).
- **Pulse Modulation (*The Electronic Shutter*):** Since a CO<sub>2</sub> gas discharge cannot be modulated at the extreme speeds required for EUV generation, the Master Oscillator generates a continuous beam that is “carved” into pulses using electro-optical modulators. By applying ultra-fast electrical signals to specialized crystals (*Pockels cells*), the system acts as a high-speed shutter, cutting the continuous emission into distinct pulse cycles. This serves as the fundamental clock of the scanner, ensuring that the laser energy is delivered in discrete packets perfectly synchronized with the droplet generator.

### 6.1.2. Staged Amplification: The Photonic “Pump”

The weak seed signal enters a chain of five massive CO<sub>2</sub> Power Amplifiers (*PA*), typically located in the facility sub-fab due to their immense size. This amplification chain — engineered by the German company **Trumpf** — stands as the most powerful industrial laser system ever built.

- **Amplification Mechanism:** The gaseous mixture (CO<sub>2</sub>, N<sub>2</sub>, He) is excited by 13.56 MHz Radio Frequency (*RF*) discharges. While CO<sub>2</sub> is the primary emitting species, Nitrogen N<sub>2</sub> acts as an energy buffer, transferring its vibrational energy to the CO<sub>2</sub> molecules via collisions to enhance excitation efficiency. Helium (He) provides critical molecular cooling, allowing the CO<sub>2</sub> molecules to return to their ground state rapidly, a vital requirement for maintaining a high repetition rate necessary for EUV production.
- **Volume of Gain and Thermodynamics:** The massive scale of these cylinders is dictated by two fundamental factors:

1. **Gain Volume:** To reach Megawatt peak power levels, the beam must interact with a massive volume of excited molecules. A larger diameter and length allow for maximum energy extraction from the gas.
2. **Thermal Stability:** To prevent the gas mixture from overheating — which would compromise beam quality and cause thermal lensing — the structure must be large enough to allow for rapid gas circulation and efficient heat dissipation through specialized heat exchangers.

### 6.1.3. Transport of the Laser Beam and focusing

- **Beam Delivery Unit (BDU):** The beam is guided through a 20-meter optomechanical arm. To prevent thermal turbulence and the "burning" of air (*ionization of O<sub>2</sub> and moisture*), the path is maintained in a controlled vacuum or a high-purity nitrogen environment. Nitrogen is used as a strategic alternative to vacuum because it is transparent to the 10.6  $\mu\text{m}$  wavelength and, when maintained in slight overpressure, acts as a "gas shield" that prevents external contaminants from entering the optical path.
- **Infrared Optics and Adaptive Focusing:** Standard optical glasses are opaque to 10.6  $\mu\text{m}$  radiation and would melt instantly. Therefore, the Beam Delivery Unit (BDU) utilizes a rigid pipeline incorporating a total of six specialized water-cooled copper mirrors or gold-coated silicon elements. These materials are chosen for their exceptional thermal conductivity and maximum reflectivity in the far-infrared, allowing them to manage the extreme thermal load without surface degradation. To compensate for mechanical vibrations and micro-thermal expansions, the first two copper mirrors within this BDU chain act as adaptive optics, equipped with high-speed piezoelectric actuators to dynamically stabilize beam pointing.
- **The final focusing: The Final Focusing:** Final focusing is performed by an actively water-cooled, off-axis parabolic copper mirror located just before the interface with the Source Chamber. This highly stressed mirror concentrates the multi-kilowatt beam into a microscopic focal point, passing through the aperture of the Source Chamber to strike the tin droplets with absolute spatial synchronization.
- **The Off-Axis Parabolic copper Mirror and Beam Steering:** To focus a laser beam of such immense power, a standard "centered" parabolic mirror cannot be used. In a hypothetical standard setup, the mirror would be located directly in the path of the incoming laser, obstructing the beam and causing thermal damage to the equipment. The final element of the BDU is the off-axis parabolic copper mirror, namely: a precisely cut "lateral section" of a larger, virtual "parent" parabola. This specific geometry allows the beam to be simultaneously tilted and focused. The laser beam thus enters the Source Chamber aperture at a precise angle,

converging toward the droplet impact point without any mechanical interference with the incoming path.

- **The Role of Preceding Mirrors:** This final focusing is only possible thanks to the preceding adaptive copper mirrors in the Beam Delivery Unit (*BDU*). These mirrors perform the “beam shaping”, adjusting the beam diameter and wavefront to match the off-axis segment perfectly. If the beam were not perfectly shaped and aimed by the earlier stages, it would miss the "sweet spot" of the parabolic section, leading to optical aberrations.
- **Functional Benefit:** This combined effort creates an unobstructed "clear path". The laser strikes the tilted parabolic surface and is reflected toward the droplet leaving the central space free for metrology sensors and allowing the generated EUV light to expand toward the collector without interference.

## 6.2. The Tin Droplet Generator

The "target" for the laser consists of a continuous, high-speed stream of ultra-pure molten tin droplets:

- **Ejection:** Tin is melted in a reservoir at approximately 250 °C and forced through a microscopic 30-micron nozzle using high-pressure inert gases.
- **Rayleigh-Plateau Instability:** A piezoelectric crystal vibrates the nozzle at a frequency of 50,000 Hz (*50 kHz*). This vibration triggers the Rayleigh-Plateau instability, which fragments the continuous liquid jet into exactly 50,000 identical droplets per second, traveling through the vacuum at a velocity of 80 m/s (*approx. 288 km/h*).
- **Equilibrium and Spacing:** The 50 kHz frequency is a strategic choice: it ensures the EUV light appears as a continuous flow to the photoresist while maintaining a 1.4 mm distance between successive droplets. This spacing is vital to prevent the massive shockwaves from one plasma explosion from disturbing the trajectory or integrity of the next incoming droplet.

## 6.3. Spatial and Temporal Synchronization

Intercepting a 30-micron droplet traveling at 80 m/s requires an ultra-fast, closed-loop control system:

- **Detection (*Look-Ahead*):** A dedicated "laser curtain" (*metrology system*) monitors the trajectory and velocity of each incoming droplet. This provides the real-time coordinates necessary for the firing sequence.
- **Timing (*FPGA Control*):** High-speed FPGA (*Field-Programmable Gate Array*) processors calculate the exact "time-of-flight." They coordinate the laser firing moment with microsecond precision to ensure the pulse meets the droplet at the exact focal point.

- **Fast Steering Mirrors (FSM):** If a droplet deviates even slightly from its predicted path due to pressure fluctuations, Fast Steering Mirrors of the BDU — driven by piezoelectric actuators — adjust the laser aim in real-time to track and hit the target.

#### 6.4. The "Double Hit" Strategy and Plasma Transformation

To maximize EUV yield, the laser doesn't strike the droplet once, but twice in rapid succession. This interaction triggers the final transformation:

- **The Pre-Pulse:** A low-energy pulse strikes the spherical 30-micron droplet. Instead of vaporizing it, the energy causes the tin to expand into a thin, saucer-shaped disk (*the "pancake"*). This increases the target surface area significantly.
- **The Main Pulse:** Microseconds later, the Megawatt-level main pulse strikes the expanded disk. Because the target is now wider and thinner, it absorbs the laser energy much more efficiently.
- **The Plasma Explosion:** This second impact triggers the near-instantaneous phase transition. The tin atoms are stripped of 8 to 12 electrons, creating a high-density plasma. As the plasma expands, it releases a massive, isotropic burst of EUV radiation at 13.5 nm.
- **Conversion Efficiency (CE):** This dual-pulse approach is what allows ASML to achieve the high "Conversion Efficiency" (CE) necessary for industrial production, turning raw laser power into a usable, high-intensity EUV beam.

##### 6.4.1. Timing of the Duty Cycle

To understand the extreme dynamics of the source, it is essential to break down the 20- $\mu$ s period (50 kHz) of a single droplet cycle:

- **Pulse Duration (Nanoseconds):** The actual laser emission is incredibly brief. The pre-pulse lasts approximately 50-100 ns, and the main pulse about 100-200 ns. Combined, the laser is "active" for less than 0.3  $\mu$ s per cycle.
- **Inter-pulse Delay (1.5 – 2  $\mu$ s):** After the Pre-pulse, the system waits about 2  $\mu$ s. This "pause" is vital: it allows the mechanical energy to reshape the tin droplet into a wide "pancake" disk, maximizing the surface area for the next impact.
- **The "Recovery" Window ( $\approx 18 \mu$ s):** After the main pulse vaporizes the tin into plasma, the laser remains silent for the remainder of the cycle. This window is necessary for:
  - **Debris Clearing:** Allowing the vacuum and hydrogen flow to sweep away the plasma remnants before the next droplet arrives.
  - **Amplifier Cooling:** Giving the CO<sub>2</sub> gas mixture in the Trumpf amplifiers time to return to its ground state.
  - **Droplet Travel:** Waiting for the next droplet to travel the 1.4 mm required to reach the focal point.

### 6.4.2. Kinematic Energy Integration and Single-Die Transit Dynamics

To establish a clear structural connection with the preceding analysis, Section 5.5. demonstrates that despite the asymmetrical scanning velocities between the reticle and the wafer stages, the integrated exposure window remains geometrically identical across both half-field and full-field layouts. Consequently, the actual radiation dose required to chemically transform the photoresist within that synchronized window is physically quantified and computed herein, based on the discrete pulse train of the 50 kHz plasma source.

To fully grasp the ultra-fast dynamics of the exposure process, the physical interaction between the moving substrate and the stationary illumination slit must be quantified. In High-NA configurations, the wafer stage translates along the longitudinal Y-axis at a peak scanning velocity of 1.2 m/s (1200 mm/s) beneath a fixed slit that measures exactly 5 mm in width at the wafer plane. Consequently, the temporal and energetic parameters governing a single exposure window evolve according to three deterministic metrics:

- **Single-Point Exposure Window:** Any discrete coordinate (or single atom) of the photoresist layer requires a transit time of exactly 4.16 ms:  $\left(t_{slit} = \frac{5 \text{ mm}}{1200 \text{ mm/s}}\right)$  to cross the illumination slit. Operating at a 50 kHz repetition rate, the laser delivering 50,000 pulses per second ensures that every single point of the circuit topography integrates exactly 208 discrete energy bursts ( $50,000 \times 0.00416$ ) during its brief exposure window.
- **Total Die Transit Time:** For an entire half-field circuit die — measuring 16.5 mm in length along the Y-axis to completely clear the 5 mm slit, the stage must sweep a total distance of 21.5 mm (16.5 mm + 5 mm). This total mechanical scan is executed in less than a heartbeat, lasting a mere 17.9 ms:

$$\left(t'_{half-field} = \frac{21.5 \text{ mm}}{1200 \text{ mm/s}}\right).$$

- **Surface Energy Target (Dose Absorption):** To establish a rigorous energy balance, the link between single-die absorption and total system power must account for mechanical overhead. Processing 150 wafers per hour with an average layout of 135 native half-field exposure fields per substrate requires the exposure of exactly 20,250 fields per hour, or 5.625 fields per second ( $20,500/3,600 \text{ s}$ ). Since each individual field physically absorbs 171.6 mJ (0.1716 Joules) of useful radiant energy, the net continuous power required on the wafer surface would mathematically be only 0.96525 Watt ( $\approx 1 \text{ Watt}$ ) ( $5,625 \times 0.1716 \text{ J}$ ). However, a scanner does not emit light continuously: out of the 3,600 seconds in an hour, the laser is active only during the pure scanning window, while the remaining time is lost to non-exposure mechanical overheads (wafer swap in the TwinScan system, stage deceleration, and stepping displacements along the X-axis). This operational

efficiency represents an exposure duty cycle of approximately 30% to 35%. Consequently, to deliver the required energy package within this compressed timeframe, the instantaneous power reaching the wafer during active scanning must be scaled up by a factor of three, establishing a baseline scanner requirement of 2.75 to 3.2 Watts for highly sensitive resists. In industrial high-volume manufacturing utilizing less sensitive, high-resolution formulations requiring greater exposure doses, this requirement scales dynamically up to 12 to 15 Watts of useful, in-band EUV radiation at the substrate level.

---

## 7. Beam Power: From Megawatts to Watts

Energy management represents the most critical engineering challenge of the entire system. The process is characterized by massive energy dispersion along the path separating the initial pulse generation from the final target on the silicon.

### 7.1. Useful Power Decay

The journey of EUV photons through the scanner is an "obstacle course" against material absorption:

- **At the Source (*Plasma*):** The tin plasma explosion emits approximately 250-500 Watts of "in-band" EUV radiation.
- **The Reflection Tax:** As analyzed in the mirrors chapter, every reflection involves a  $\approx 30\%$  loss. After an average optical train of 11 bounces (*between the illuminator and projector*), only 2-3% of the original light survives.
- **At the Wafer:** On average, only 5-15 Watts of useful power reach the final impact point (*typical for ASML EXE:5000 series scanners*). While this value may appear modest, the energy intensity is extremely high because this power is focused onto a nanometric area.

*Technical Note: EUV power fluctuates based on system generation and optical wear. While current sources generate 250 to 500 W nominally, the specific optical design and the progressive loss of reflectivity due to surface contamination (carbon or tin residues) determine the final wattage. Maintaining a stable output is essential to ensure a high industrial throughput (wafers per hour).*

*Technical Note on Plasma Physics: During the brief nanosecond window of the main pulse, the plasma reaches a peak instantaneous electron energy of 30-40 eV (also termed 30-40 eV electron temperature in technical language, approx. 400,000 °C). This energy level is the thermodynamic "sweet spot": it is high enough to ensure that the high-energy tail of the electron distribution constantly excites the  $\text{Sn}^{8+}$  to  $\text{Sn}^{12+}$  ions, yet low enough to prevent over-ionization, which would shift the emission away from the 13.5 nm target. The resulting 92 eV photons are the product of resonant electronic transitions (4d to 4f) within*

*these specific ionic states. It must be clarified that the 30-40 eV value represents the mean kinetic energy of the free plasma electron gas, acting as the impact trigger for excitation. Conversely, the higher 92 eV energy of the output photons is strictly dictated by the internal quantum mechanics of the tin ions: it represents the exact energy differential released when a bound electron drops back from the excited 4f orbital to the fundamental 4d orbital, naturally generating the target 13.5 nm extreme ultraviolet wavelength.*

## **7.2. Energy Balance and the "Critical Point"**

To achieve nanometric resolution, overall energy efficiency is sacrificed:

- **Electrical Consumption:** To power the Trumpf laser, vacuum pumping systems, and cryogenic cooling, a single EUV machine requires approximately 1 Megawatt (MW) of electricity - equivalent to the energy needs of a small residential neighborhood.
  - **Heat Management:** Over 95% of the input energy is not converted into useful light and is instantly transformed into heat. This massive thermal load must be dissipated to prevent optical deformations and structural damage, making the scanner one of the most complex thermal management systems in the world.
- 

## **8. Protection and Stability of the Laser System**

The Trumpf laser architecture and the Zeiss optical collection system require environmental and mechanical conditions of absolute stability. To maintain this framework, the drive laser beam and the tin plasma generation operate in a continuous cycle, even during exposure pauses — such as during stage stepping and wafer swaps when the shutter is closed. This uninterrupted operation satisfies two vital requirements:

- **Dynamic Thermal Balance:** Despite the extreme peak instantaneous temperature of the plasma (400,000 °C) the Source Chamber maintains a macro-structural steady state. The massive thermal energy injected by the 50 kHz laser pulses is continuously counterbalanced and removed by the active water-cooling networks integrated into the Collector mirror and the chamber walls, preventing any structural degradation. Even a brief shutdown would cause micrometric contractions of the mechanical supports and mirrors. Restoring the nanometric focus after such a thermal shift would require hours of recalibration.
- **Throughput Maximization:** The complex phase synchronization (*lock-in*) between the 50 kHz laser pulses and the tin droplet generator requires highly stable fluid dynamics and timing circuits, which take time to stabilize. Maintaining a continuous, uninterrupted pulse cycle ensures the machine is always in an operational steady-state, eliminating the steep productivity losses associated with system ramp-up. When the beam is not actively exposing a target — such as during wafer swaps or

stage steps — the energy is safely blocked and diverted by the high-speed shutter system (see *Chapter 9*) rather than shutting down the plasma source.

- **Dual-Shutter Architecture**

The scanner manages beam disruption through a coordinated two-tier architecture. High-frequency micro-pauses (*in the microsecond range*) are managed at the Master Oscillator (see *Section 6.1.1*) via the electro-optical shutter (*Pockels cell*), which modulates individual pulse profiles. Conversely, macro-pauses (*such as wafer swaps or major stage displacements*) are handled by the mechanical, water-cooled Intermediate Focus Shutter (see *Chapter 9*). This dual layout allows the Trumpf drive laser to maintain a continuous duty cycle, preventing thermal shocks in both the laser amplifiers and the plasma generation chamber.

### 8.1. Optical Isolation and Self-Oscillation Protection

In a high-power architecture, optical feedback represents a catastrophic threat to the system integrity.

- **The Risk of Self-Oscillation:** When the multi-kilowatt laser beam strikes the tin droplet, a portion of the infrared light is reflected directly back toward the beam delivery system. If this back-reflection were to re-enter the amplifier chain, it would travel in the opposite direction and undergo massive re-amplification. This parasitic feedback loop could destroy internal optical components within microseconds.
- **The Faraday Isolation Solution:** To neutralize this threat, advanced optical isolators based on the Faraday effect are integrated into the Beam Delivery Unit (BDU) (*as detailed in the vacuum system architecture*) outside the Source Chamber, positioned upstream of the off-axis mirror aperture. Operating as non-reciprocal optical "one-way valves," these isolators utilize magneto-optical rotation to shift the polarization state of the infrared returning light. This deployment configuration ensures that the parasitic back-reflection traveling back through the off-axis mirror hole is safely intercepted and diverted into water-cooled beam dumps before it can re-enter the main beam line and reach the Master Oscillator.
- **Geometric and Axial Filtering:** When the laser strikes the target, the light is scattered in all directions due to plasma expansion. Any radiation reflected off-axis strikes the water-cooled, absorbent walls of the Source Chamber, where it is safely dissipated as heat. Only the narrow fraction of infrared light that is back-reflected specularly along the exact optical axis can pass backward through the off-axis mirror aperture. Because this returning infrared light is perfectly aligned with the main beam line, it is guided directly into the Faraday isolator, ensuring that the dangerous axial feedback is 100% intercepted while the diffuse radiation is neutralized by the chamber geometry. Crucially, any trace amount of extreme ultraviolet (EUV) radiation back-scattered along this path poses zero threat to the

upstream optics: due to its high quantum energy, EUV radiation is instantaneously and fully absorbed by the outer atomic surface of any solid optical window, making the infrared-engineered Faraday isolator inherently immune to EUV-induced internal damage.

## 8.2. Outgassing and Material Integrity under Ultra-High Vacuum

To preserve the extreme vacuum threshold necessary to prevent EUV photon absorption (*as detailed in the vacuum system architecture*), strict structural material controls are enforced:

- **Outgassing Control:** To maintain an uninterrupted ultra-high vacuum  $10^{-9}$  to  $10^{-11}$  mbar ( $10^{-7}$  to  $10^{-9}$  Pa) within the central core, every internal component — including the encapsulated Maglev motor windings and structural brackets — undergoes an intensive bake-out process in specialized thermal ovens prior to assembly. This protocol forces out any trapped volatile molecules, preventing chemical outgassing that could otherwise contaminate the reflective coatings of the EUV mirrors or obscure the optical paths.
- 

## 9. The Intermediate Focus Shutter Mechanical System

The shutter assembly is a stationary-body apparatus mounted at the Intermediate Focus (*IF*) aperture, which defines the physical and pneumatic boundary interface between the Source Chamber and the Main Vessel Central Core (*see Section 1.1.2*). The physical architecture consists of the following interconnected structural sub-systems:

### 9.1. Structural Description

- **The Stationary Support Frame:** A heavy, water-cooled rigid bracket bolted directly to an interior dividing wall of the source chamber. This frame houses fixed permanent magnets (Neodymium-Iron-Boron) of a voice-coil actuator, defining a linear magnetic air-gap centered along the actuation axis.
- **The Resilient Suspension (*Flexures*):** A system of parallel, high-tensile metallic elastic strips, named flexures, that rigidly connect the stationary support frame to the moving assembly. This elastic suspension replaces all traditional mechanical bearings, joints, or pivots, forming a frictionless and pivotless hinge.
- **The Rigid Actuation Arm:** A lightweight, high-stiffness structural member suspended in space exclusively by the parallel flexure strips.
  - **The Proximal End (*Drive Interface*):** This end forms a monolithic body with a moving electrical coil of a voice-coil actuator, which hangs suspended inside the magnetic air-gap of the stationary frame without any physical contact.

- **The Distal End (*Optical Interface*):** This end extends directly toward the optical axis of the EUV beam and terminates with a mounting bracket for the shutter blade.
- **The Interception Blade:** A metallic blade (*or a synchronized pair of opposing blades*) secured to the distal end of the rigid arm. The blade is manufactured from refractory alloys combining high thermal resistance with extreme thermal conductivity (*such as specialized Molybdenum or Tungsten alloys*). The interior body of the blade housing integrates an internal network of liquid-cooling micro-channels designed for permanent laminar fluid flow.

## 9.2. Functional and Kinematic Description (*Apparatus Dynamics*)

The operational physics of the shutter system governs the high-speed temporal modulation of the EUV beam via non-contact electromagnetic forcing and elastic structural deformation:

- **Lorentz-Force Actuation:** When the scanner control unit commands a status change, a high-current electrical pulse is injected into the arm of the proximal coil. The interaction between this current and the permanent magnetic field of the stationary frame generates an instantaneous Lorentz force. This force drives the proximal coil linearly through the gap, accelerating the entire rigid arm assembly at rates of 10 to 15 G without any solid-to-solid friction.
- **Parallel-Flexure Kinematics:** As the voice coil pushes the arm, the parallel flexure strips undergo controlled micro-elastic bending. The parallel orientation of these elastic strips constrains the arm trajectory, converting the movement into a precise, horizontal planar translation.  
This geometric constraint keeps the distal end perfectly aligned, preventing any vertical dipping, twisting, or mechanical jitter under violent acceleration.
- **Gating and Interception:** Driven by this planar translation, the refractory blade at the distal end sweeps in front of the intermediate focus aperture across the optical axis. At the Intermediate Focus, where the EUV beam profile reaches its minimum microscopic spatial convergence, the blade completely cuts through or clears the focused photon path within a few milliseconds.
- **Thermal Energy Dissipation:** When the shutter is held in the closed position during macro-pauses (*such as stage stepping between dies or orbital wafer swaps in the TwinScan system*), the blade absorbs the full multi-kilowatt thermal load of the active plasma beam. The internal micro-channels continuously conduct this concentrated heat away from the optical axis to an external facility heat sink, protecting the downstream reticle and projection optics inside the POB (see *Section 4.2*) from thermal deformation while keeping the plasma source in an operational steady state.

*Technica Note on Hybrid Pulse Gating Mode (Electronic Synchronization): Because the mechanical shutter assembly has fundamental inertial limits that restrict its travel time to the millisecond range, the apparatus operates in a hybrid Pulse Gating mode to manage individual laser bursts. During the physical blade transit across the optical axis, the local control electronics command the TRUMPF laser to electro-optically "skip" or suppress specific plasma pulses via the Pockels cell (see Section 8). This advanced electronic gating switches off the active EUV emission within microseconds, working in flawless structural synchronization with the mechanical blade during wafer repositioning. This dual-tier modulation ensures that no partial-dose exposures or stroboscopic anomalies occur at the die boundaries, while the physical blade safely enters the closed position to absorb the continuous thermal load of the primary drive laser.*

---

## **10. In-Situ Maintenance and Clean-Optics Chemistry**

### **10.1. Structural Description**

Given the impossibility of mechanically cleaning the optics without permanently destroying the nanometric layers of the Mo/Si Bragg mirrors (see Section 4.2.) or breaking the ultra-high vacuum — an operation that would require days of thermal stabilization to restore (see Section 8.) — the scanner integrates a specialized chemical maintenance apparatus. This in-situ hardware is distributed across the Source Chamber and the Central Core of the Main Vessel, comprising the following foundational elements:

- **The Dynamic Gas Lock (DGL):** A pneumatic barrier assembly sustained by a continuous supersonic stream of ultra-pure hydrogen gas that emanates from cylindrical conduits machined directly within the structural separation wall between the Source Chamber and the Main Vessel. These integrated injection nozzles maintain a forced, high-pressure flow of ultra-pure molecular hydrogen gas (H<sub>2</sub>) rushing directly into the Intermediate Focus (IF) aperture, counter-current to the plasma expansion. This gaseous wall acts as a highly selective kinetic filter: while the high-energy extreme ultraviolet (EUV) photons slice straight through the light hydrogen molecules without absorption at an extreme frequency of 22.2 PHz (Petahertz - 10<sup>15</sup> Hz), the mass and velocity of the counter-flowing gas physically repel, slow down, and redirect incoming tin ions and ballistic debris back into the Source Chamber, away from the main optical path.

### **10.2. Functional and Chemical Description**

The in-situ maintenance system coordinates a continuous, non-contact molecular cleaning cycle during active lithographic operations through precise kinetic control and atomic reactions:

- **In-Situ Atomic Hydrogen Activators:** Rather than injecting external reactive agents, the system directly exploits the residual molecular hydrogen ( $H_2$ ) that inevitably migrates into the Main Vessel from the Intermediate Focus (IF) pneumatic barrier. To activate this residual gas near the sensitive mirror surfaces, the scanner utilizes a matrix of hot tungsten filaments or localized radiofrequency (RF) hydrogen plasma sources. These components act strictly as local thermal and electrical dissociators: by transferring energy to the passing  $H_2$  molecules, they break the molecular bonds to generate highly reactive, short-lived radical atomic hydrogen ( $H^*$ ). This in-situ generated atomic hydrogen chemically reacts with any deposited tin contaminants on the Zeiss optics, transforming solid tin into volatile stannane gas ( $SnH_4$ ) that is continuously evacuated by the ultra-high vacuum pumps.
- **Thermal Condensation Tin Catchers:** A series of structurally optimized, actively cooled extraction panels positioned strategically along the peripheral boundaries of the Source Chamber (see Section 8). When the stray tin vaporized during the 50 kHz plasma generation process escapes the primary kinetic barrier, it travels toward these freezing surfaces. Upon striking the extraction panels, the gaseous metal undergoes instantaneous physical condensation, solidifying out of the gaseous environment and onto the traps to prevent vacuum saturation and protect the upstream scanner elements.
- **Carbon Volatilization via Atomic Hydrogen:** Carbon contamination occurs when the intense EUV flux interacts with infinitesimal organic residual molecules, depositing an opaque carbon crust on the mirrors. To reverse this, the above mentioned highly reactive atomic hydrogen radicals  $H^*$  chemically attack the solid carbon deposits C, breaking them down into volatile methane gas  $CH_4$ . This gaseous hydrocarbon "evaporates" into the vacuum core and is immediately evacuated by the turbomolecular and cryogenic pumping systems.

### 10.3. Synopsis of Hydrogen Stabilization Effects

Beyond its primary role as a chemical detergent, the controlled introduction of hydrogen gas into the UHV core provides two secondary stabilization mechanisms for the scanner architecture:

- **Assisted Molecular Cooling:** While the system operates under a strict vacuum regime ( $10^{-9}$  to  $10^{-11}$  mbar), the low-density presence of hydrogen molecules facilitates micro-convective thermal transfer. This molecular gas movement assists the main micro-channel cooling network in dissipating localized heat gradients induced by the high-power radiation, helping to maintain the absolute structural steady-state of the POB.
- **Spectral and Dose Consistency:** By executing a continuous, real-time purge of carbon and tin contaminants from the mirror surfaces, the atomic hydrogen

chemistry prevents any degradation or drift in the reflectivity of the optics. This ensures that the reflective "pitch" of the Bragg layers remains unaltered over time, delivering a perfectly uniform and deterministic exposure dose to the photoresist on the wafer.

---

## 11. Real-Time Distributed Computational Infrastructure

### 11.1. Computational Requirements and Process Parallelism

To execute the lithographic exposure cycle, the system must process massive streams of multi-domain data simultaneously. The real-time operation of the scanner relies on the parallel execution of seven core computational tasks:

- **Sub-Nanometric Stage Metrology:** The continuous calculation of the wafer and reticle stage positions based on the GHz sampling of the planar diffractive photovoltaic sensors (*see Section 3.2.1*). This task involves resolving complex phase-shift trigonometry to output absolute 6-DOF coordinates.
- **Maglev Actuation Feedback:** The real-time computation of Lorentz-force vector profiles required to drive the wafer and reticle Maglev motors (*see Section 3.2.1*). The algorithms must dynamically compensate for micro-vibrations and structural accelerations up to 15G.
- **Real-Time Wafer Topography and Roughness Estimation:** The high-frequency processing of raw telemetry data streamed by the Laser Autofocus system (*see Section 3.2.1*). The computational matrix must instantaneously calculate local height variations and microscopic surface roughness across the wafer, translating these spatial micro-crests into real-time Z-focus and tilt adjustments for the Maglev stage.
- **Predictive Reticle Thermal Modeling (*Electronic Layout Analysis*):** The real-time mathematical analysis of the electronic reticle layout files during the scanning pass. Because the circuit areas with lower absorber density reflect higher EUV energy, the supercomputer executes predictive algorithms to model exactly where localized thermal hotspots will form on the reticle and downstream projection optics. This mathematical prediction allows the system to compute preventive deformation matrices for the adaptive mirrors (*see Section 4.6.3.*) before thermal expansion distortions can manifest.
- **Adaptive Wavefront Remodeling:** The continuous execution of feed-forward predictive thermal expansion algorithms and the processing of ILIAS wavefront sensor data (*see Section 4.6.3.*) to calculate the real-time voltage adjustments for hundreds of piezoelectric mirror actuators.

- **Plasma Pulse and Shutter Lock-In:** The multi-scale phase synchronization between the nanosecond-duration pulses of the 50 kHz TRUMPF drive laser, the electro-optical Pockels cell, the microsecond-paced tin droplet generator, and the millisecond-response high-speed mechanical Intermediate Focus shutter (see *Section 8 and Chapter 9*).
- **Reticle Masking (REMA) Border Shadowing:** The dynamic calculation of the trajectories for the four independent REMA framing blades to flawlessly track the opaque borders of the moving reticle without pulse clipping (see *Section 4.3.*).

## 11.2. Loop Speed Hierarchy and Priority Matrix

The distributed computing architecture organizes these processes into a strict hierarchical matrix based on execution bandwidth and deterministic priority:

- **Level 0 - Deterministic Nanosecond Range / GHz Sampling:** Raw signal digitization, photovoltaic fringe counting, and phase-shift capture. This level has absolute priority; any latency here causes immediate exposure aborts due to position uncertainty.
- **Level 1 - Microsecond Feedback Loop / MHz Control:** Active Maglev levitation adjustment, Lorentz-force current modulation, voice-coil shutter actuation, and laser pulse gating via the Pockels cell. This loop operates deterministically in the low-microsecond range.
- **Level 2 - Millisecond Real-Time Adjustments / kHz Loop:** Real-time adaptive mirror surface deforming via piezoelectric micro-stresses, REMA blade boundary tracking, and dynamic thermal balance monitoring.
- **Level 3 - Non-Deterministic System Management:** Wafer lot scheduling, automated file transfer of the reticle topography map, vacuum pump telemetry aggregation, and facility interface communications.

## 11.3. Local Edge Computing: In-Situ FPGAs

To achieve the zero-latency execution required by Level 0 and Level 1 loops, the scanner relies on Field-Programmable Gate Arrays (FPGAs) distributed as local edge-computing nodes physically inside the machine. Centralized software processing would introduce unacceptable propagation delays through standard operating system layers. FPGAs execute real-time processes via dedicated parallel hardware circuits carved into the silicon chip, achieving absolute determinism and microsecond-range processing times unaffected by software interrupts:

- **Metrology Edge FPGAs:** Mounted in close proximity to the metroframe within the surrounding peripheral space (see *Section 1.1.2*). These chips are hardwired to process raw sinusoidal signals from the high-speed photovoltaic sensors, executing real-time digital interpolation and fringe accumulation directly at the hardware layer.

- **Actuator Control FPGAs:** Embedded directly inside the structural enclosures of the Maglev stators and the shutter voice-coil assemblies. These devices receive the calculated coordinate corrections and instantly translate them into exact pulse-width modulation (*PWM*) current drivers for the magnetic coils.

#### 11.4. Sensor Distribution and External Optical Bus

Because the central core operates under ultra-high vacuum conditions  $10^{-9}$  to  $10^{-11}$  mbar (see Section 1.1.2), the high-frequency electronic data generated by thousands of distributed sensors cannot be routed via standard copper wires, which would cause severe electrical noise, outgassing, and thermal contamination.

- **Pneumatic and Electronic Isolation:** All local sensor data are immediately converted into optical signals inside the vacuum boundary.
- **The High-Bandwidth Fiber-Optic Bus:** Thousands of radiation-hardened fiber-optic lines bundle together to form an external optical data bus. This bus routes the digitized sensor data through hermetic vacuum feedthroughs, transferring gigabytes of real-time telemetry per second out of the Main Vessel and into an external computing facility located in the sub-fab area.

#### 11.5. Physical Architecture of the Supercomputer

Outside the scanner, located in a clean, vibration-isolated server enclosure in the sub-fab, sits the external computational facility. The physical architecture consists of the following hardware blocks:

- **The Mainframe Master Rack:** A central supercomputer rack housing high-core-count enterprise server nodes connected via an ultra-low-latency interconnect.
- **Dedicated Co-Processor Arrays:** Clusters of high-performance Graphics Processing Units (*GPUs*) or specialized application-specific integrated circuits (*ASICs*) optimized for massive matrix calculations.
- **The Hardware-in-the-Loop (*HIL*) Routing Matrix:** A high-speed digital switching network that terminates the incoming fiber-optic bus from the Main Vessel and cross-routes data packets to the respective processing cores.

#### 11.6. Functional Architecture and Master Orchestration

The external supercomputer architecture orchestrates the global machine state by segmenting processing tasks according to the priority matrix, bridging the distributed in-situ hardware with centralized computing:

- **The Level 2 Co-Processors Role:** Operating at the millisecond scale (*kHz loop*), the GPU/ASIC arrays process the **Level 2** data arriving directly from the fiber-optic bus. They execute the high-frequency matrix calculations for the Predictive Reticle Thermal Modeling and the Real-Time Wafer Topography and Roughness Estimation. By analyzing the reticle file and sensor feedback, these processors calculate the dynamic 3D deformation vectors for the adaptive projection mirrors

and the boundary coordinates for the REMA framing blades (see Sections 4.3. and 4.6.3).

- **The Master Orchestrator Role:** The Mainframe Master Rack acts as the supreme software intelligence of the scanner, managing Level 3 non-deterministic operations. It coordinates wafer lot scheduling, uploads and stores the complete 3D topographic maps generated during the TwinScan measurement phase, monitors vacuum pump telemetry, and calculates the macroscopic trajectory profiles for the next exposure cycle. This vacuum telemetry encompasses the high-frequency continuous data logging of absolute pressure thresholds, turbomolecular rotor rotation speeds, and structural motor temperature profiles across the Main Vessel boundaries, enabling the mainframe to execute automated system-safety interlocks and abort sequences in the event of a sudden vacuum degradation or hardware failure before the subsequent lot deployment.
  - **The Real-Time Semiconductor Relay:** Once the Level 2 co-processors compute the deformation and trajectory matrices, the Master Orchestrator synchronizes these matrices with the main scanning profile, streaming them as deterministic target vectors back down the fiber-optic bus. The local in-situ FPGAs (see Section 11.3.) receive these targets and maintain hardwired control over the Level 1 microsecond loops, executing the nanometric alignment of the Maglev stages and the pulse lock-in of the plasma source. If a local sensor reports a critical deviation, the local FPGA executes an immediate emergency pulse-skip via the Pockels cell, completely bypassing the external supercomputer to protect the optical train.
- 

## 12. Automated Decommissioning and Controlled Shutdown System

### 12.1. Structural Description

The safe, non-destructive termination of the scanner multi-megawatt operational state is governed by an integrated hardware apparatus designated as the Automated Decommissioning and Emergency Energy Mitigation System. This system is embedded within the distributed computational and fluidic infrastructure of the scanner, comprising the following structural components:

- **The Hardware Ramping Modules:** Dedicated solid-state power-delivery electronic circuits integrated into the TRUMPF laser amplifiers and the Maglev stator drive bays. These modules are hardwired to execute linear electrical current reductions (*decay ramps*) rather than instantaneous power cut-offs.
- **The Thermal Storage and Bypass Loop:** A secondary, high-volume hydraulic manifold paired with the main micro-channel cooling network of the collection and projection mirrors. This loop includes specialized pneumatic pressure reservoirs

capable of maintaining laminar coolant flow under residual heat expansion, completely independent of the main facility grid.

- **The Nitrogen Purge and Venting Block:** A gas delivery manifold consisting of ultra-fast piezo-valves and heating coils, connected to an ultra-pure nitrogen source. This block interfaces directly with the hermetic ConFlat flanges of the Main Vessel.
- **The Mechanical Docking Matrix:** A series of physical micro-locking pins and safety back-stops positioned at the boundaries of the common Maglev stator plane. These pins are designed to intercept and cradle the levitating wafer and reticle stages during loss-of-levitation sequences.

## 12.2. Functional and Chronological Description

The apparatus coordinates the thermal, pneumatic, and magnetic de-escalation of the scanner through three strictly timed, automated operational phases to eliminate shock-induced micro-contractions, outgassing contamination, or physical collisions:

- **Phase I: Optical and Energy Decoupling (*The Millisecond Ramping*):** Upon receiving a decommissioning command, the Master Orchestrator triggers the Level 1 control loops. The local Metrology FPGAs lock the system in an absolute **Pulse Gating mode**, commanding the Pockels cell to switch off individual laser bursts within microseconds. Concurrently, the voice-coil actuators drive the mechanical Shutter blade into the permanently closed position at the Intermediate Focus aperture, isolating the Central Core from the Source Chamber residual thermal emissions. The hardware ramping modules then execute a linear power decay across the main TRUMPF laser amplifiers, completely suppressing the 50 kHz tin plasma generation.
- **Phase II: Managed Thermal Dissipation (*The Cool-Down Decay*):** Following plasma suppression, the massive thermal energy stored within the collector mirror and the POB housing must be removed to prevent structural expansion or contraction drift. The **Thermal Storage and Bypass Loop** continues to pump chilled liquid through the mirror micro-channels at full operational bandwidth for a calculated duration. The Active Thermal Compensation heaters reduce their output along a mathematically optimized descending thermal gradient, ensuring that the structural materials (*such as Zerodur or cast-iron-alloys*) contract uniformly at a sub-nanometric rate, preserving the baseline mirror alignment.
- **Phase III: Pneumatic Venting and Magnetic Docking:** Once the internal temperature stabilizing loop reaches thermal equilibrium, the active magnetic dampers of the Metroframe and the main Maglev stages execute a controlled descent. The carriages are slowly steered into their designated Mechanical Docking Matrix, mechanically locking the stages to prevent secondary damage when the

levitation fields dissolve. Finally, the vacuum pumping system enters a staggered shutdown sequence: the cryogenic pumps isolate their chambers, and the turbomolecular pumps execute a monitored spin-down. To prevent the atmospheric back-flow of moisture or chemical contaminants, the Venting Block slowly injects ultra-pure, heated nitrogen gas into the Main Vessel, raising the pressure from the UHV threshold ( $10^{-9}$  to  $10^{-11}$  mbar), up to ambient atmospheric level in a linear, non-turbulent ramp.

---

## Conclusion:

High-NA EUV lithography does not merely represent an incremental improvement over previous technologies; it constitutes a total paradigm shift in semiconductor engineering. Every millimeter of these machines is the direct result of a challenge won against the extreme boundaries of physics, thermodynamics, and materials science. The integration of a 400,000 °C plasma power source, picometric metrology, deep vacuum, ultrasophisticated reflective mirrors, adaptive optics, Maglev actuators, and a massive distributed computational infrastructure, currently allows for the printing of circuits with atomic precision, enabling the mass production of 2 nm nodes and beyond. Yet, this technological adventure — featuring optics as the primary protagonist and its ability to print images onto a substrate — began long ago. It traces back to the 3rd century B.C., when Euclid founded geometrical optics by hypothesizing that light rays emanated from the eyes and propagated in straight lines. After centuries of subsequent debates regarding the nature of light propagation, the discovery of light diffraction by the Jesuit Father Francesco Maria Grimaldi in 1665 cast doubt on the pure ray theory. In 1678, Christiaan Huygens announced the wave propagation principle of light, establishing that every point on a wavefront can be considered a point source of secondary spherical waves. In contrast, Sir Isaac Newton published his treatise *Opticks* in 1704, advocating for the corpuscular nature of light. The wave theory was later rendered definitive by Thomas Young's double-slit experiment in 1801, which demonstrated that diffraction fringes were caused by the interference of sinusoidal light waves. James Clerk Maxwell, in his celebrated equations of 1864, unified electricity and magnetism, proving that light is an electromagnetic wave propagating at the velocity  $c$  of  $3 \times 10^8$  m/s. The famous Michelson-Morley experiment of 1887 subsequently proved that the speed of light was entirely unaffected by the motion of the reference frame, thereby demonstrating that it required no medium to propagate. This paved the way for Albert Einstein's Special Theory of Relativity in 1905, a time when universally accepted proof of the atomic constitution of matter did not yet exist. Concurrently, in December 1900, Max Planck proposed the quantization of light energy emitted by a blackbody, abandoning the classical dogma of continuity. The true quantum nature of light was proposed by Einstein, also in 1905, to explain the photoelectric effect: the quantum of light (later named the photon) was conceptualized as a particle endowed with momentum despite lacking a rest mass. This established the dual nature of light: displaying wave-like properties in certain contexts and particle-like characteristics in others. In 1911, Ernest Rutherford discovered the atomic nucleus and proposed a planetary atomic model, to which Niels Bohr applied Planck's quantum concepts to explain the orbital motion of electrons. This inaugurated the first formulation of quantum mechanics, providing a rigorous framework to explain both the generation and absorption of electromagnetic radiation, thereby giving a solid physical foundation to Christiaan Huygens' earlier observations.

All the knowledge accumulated over twenty-three centuries regarding light — and more broadly, the electromagnetic radiation of which visible light is a fraction — is more vital today than ever. The geometrical optics introduced by Euclid are still fundamentally employed in designing modern optical instruments such as spyglasses, telescopes, microscopes, camera lenses, and, most notably, in the fabrication of VLSI devices. However, as the integration density of electronic components escalates, geometrical optics alone is no longer sufficient; extreme miniaturization forces the manifestation of quantum effects that must be rigorously accounted for during design. Furthermore, quantum considerations are indispensable in generating the highly energetic radiation of short wavelengths capable of imprinting nanometric pattern widths onto photoresist layers. Thus, the ASML TwinScan EXE:5000 series, which pushes this collective historical knowledge to its absolute limits, is not just a tool for chip production, but a monument to the steadfast human willpower to transform what geometrical, quantum, and thermal barriers once deemed impossible into a repeatable industrial reality.

At this point in the narrative, it would be fitting to unplug the ASML scanner, as the author takes the liberty of embarking on a flight of fancy that steps outside the technology presented so far, leading us back to the dawn of human history. The drawings and graffiti discovered on the walls of ancient caves, though executed with crude techniques, already bore witness to a primitive intelligent thought seeking to reproduce the surrounding world. The need to know, dictated by the necessities of life but equally by curiosity, first drove ancient philosophers and subsequently naturalists and scientists to advance our collective knowledge. Over the centuries, ingenious experiments have been designed to verify theoretical hypotheses and achieve new discoveries. It has been a long journey which in recent years has undergone a growth so rapid as to appear unstoppable.

Yet, returning to our focus, the curiosity of our ancestors persists and amplifies today, as the existential questions remain unchanged: where do the stars in the sky and the world we observe come from, what is time, will we survive bodily death? Science has provided answers to some of these questions: consider, for instance, Darwin's theory on the origin of species, the nature of light, the atomic constitution of matter and of the atom itself, the origin of the universe, the evolution of stellar life, and so on. However, there are still no definitive answers — and perhaps there never will be — but if they did exist, we might not like them at all. This is because the intelligence of *Homo sapiens* has generated consciousness — that is, the mind ability to perceive itself and the surrounding world, far transcending mere survival instinct — and such consciousness makes it difficult to imagine that our ultimate destiny will be nothingness.

In the *Divine Comedy*, the supreme poet Dante Alighieri placed the ancient Greek philosopher Democritus in the Limbo of Inferno, "who ascribes the world to chance" (*che 'l mondo a caso pone*) due to his atomistic theory. This was contrasted with Dante's faith in

a Creator as the source of all wisdom, of whom he writes in Canto XXXIII of *Paradiso*: "...bound with love in one volume, what is scattered in pages through the universe..." (*...legato con amore in un volume ciò che per l'universo si squaderna...*). In other words, God represents the unique answer to our existential questions, and being a transcendent entity, He is not subject to human experimentation but is a matter of faith.

Leaving religious dogmas aside, the latest advancements in science, particularly in quantum mechanics, seem to vindicate Democritus, at least regarding the role of randomness in explaining natural phenomena on an atomic scale. Another anti-dogmatic aspect brought to light by quantum mechanics is Heisenberg's Uncertainty Principle, which decisively demolished the absolute determinism of classical mechanics (Laplace), placing an intrinsic limit on the boundaries of human knowledge. The quantum theory of matter and radiation, despite its probabilistic nature — so fiercely opposed by Einstein, who was nevertheless one of its foundational fathers — has become the cornerstone of modern physics. Current atomic theories derive from it and are verifiable using gargantuan machines such as the LHC at CERN in Geneva, which utilize proton beams accelerated to energies reaching 14 TeV in order to smash the atomic nucleus and reveal its elementary components.

Yet, quantum physics also pursues purposes far different from such high-energy endeavors. A prime example is found in photonics, and more specifically in spintronics, which exploits the electron's spin — its intrinsic magnetic moment — to design digital elements capable of switching between two voltage states rapidly and with minimal energy consumption. These spintronic devices have already been implemented in high-density memory production and actively coexist with traditional silicon architectures. Technology and computer science have now become an inseparable duo, driving both the mass manufacturing of advanced commercial products and the execution of the largest scientific experimental apparatuses to ever exist. Another intriguing phenomenon brought to light by the Schrödinger equation — the governing wave equation of quantum mechanics — is quantum entanglement. While entanglement could theoretically find applications in designing ultra-fast quantum computers based on qubits, the global industry remains far from a scalable, stable manufacturing reality for such systems. This operational gap ensures that the ASML scanner for High-NA EUV lithography will survive and dominate for a long time to come. In conclusion, the High-NA scanner harnesses highly energetic extreme ultraviolet photons produced by an artificial plasma — as hot as a fraction of the Sun's core — and concentrates that raw energy onto nanometric paths to carve billions of transistors, destined to configure the future silicon neurons of Artificial Intelligence.

## INTRODUCTION TO THE APPENDICES

The appendices that follow have a deliberately concise and summary character, conceived to serve as a streamlined technical reference. Given the vast and nearly endless scientific and industrial literature surrounding modern VLSI manufacturing, a fully exhaustive, narrative exposition of these secondary phases would easily expand into an entire encyclopedia. However, the most critical and foundational aspects involved in the mask data preparation and reticle engineering ecosystem are systematically listed and briefly commented upon, providing the reader with an essential, rigorous roadmap to bridge the gap between initial circuit design and physical high-NA lithographic exposure.

## APPENDIX A – From VLSI Design to the Reticle

### Introduction

The production of a reticle begins long before physical fabrication: it starts from the circuit design and undergoes a long chain of digital transformations. This phase is called Mask Data Preparation (*MDP*) and is essential for ensuring that the final pattern is printable on the wafer. A fundamental element in MDP is the absorber. While the absorber physically blocks the light, digitally it represents the main variable that necessitates M3D corrections due to its three-dimensional nature.

### A.1 – The layout file

The VLSI project originates as:

- schematic (*logic, netlist*)
- physical layout (*2D geometries on multiple layers*)

The layout is represented by the following files in standard formats:

- GDSII (Graphic Data System II): Historical, simple, and widely used. It is the de facto standard for the exchange of integrated circuit (*IC*) layout artwork.
- OASIS (Open Artwork System Interchange Standard): Modern and compressed, supporting more efficient structures. It was developed to address the data size limitations of GDSII.
- MEBES (Manufacturing Electron Beam Exposure System): Historical mask-writing standard used for decades. It is based on elementary geometric figures with low compression rates, typically generated at the end of the data prep chain.
- OASIS.mask (Advanced Mask Format): Modern and highly compact mask-writing standard, supporting complex hierarchies and advanced compression algorithms. It is currently the industry standard for optimizing data density on EUV and High-NA reticles.

These files contain:

- arbitrary polygons
- technological layers

- via-holes and contacts
- transistors and devices
- interconnections
- hierarchies (*cells, macros, IP blocks*)
- design rules (*DRC*)
- density and fill information

*Useful note: OASIS can reduce file size by up to 10x (and in some cases up to 20x) compared to GDSII thanks to embedded compression, optimization of geometric repetitions, and more compact coordinate representations.*

### **A.2 – OPC (*Optical Proximity Correction*)**

In the case of High-NA systems, OPC must manage asymmetric bias: since the anamorphic optical system reduces the image by 4x on the X-axis and 8x on the Y-axis (see *Chapter 5*), the correction software must calculate different deformations for the two axes so that the final result on the wafer is geometrically correct.

OPC corrects optical errors due to:

- diffraction
- flare (*internal scattering within the reticle*)
- shadowing (*especially in EUV*)
- anamorphism (*High-NA*)
- directional bias (*X/Y*)

Shapes are modified using:

- serifs (*to maintain sharp corners*)
- hammerheads (*widening of lines near terminations*)
- local bias (*controlled expansion/contraction*)
- SRAFs (*Sub Resolution Assist Features*) to improve resolution.

Modern OPC: Today, it is almost exclusively model-based, utilizing 3D optical simulations and iterative optimization. For EUV and High-NA, OPC is coupled with M3D and SMO.

### **A.3 – M3D (*Mask 3D Effects*)**

The EUV reticle is not a 2D object: it features relief structures, multilayers, absorbers, and topcoats. This three-dimensionality introduces effects such as:

- X/Y asymmetry: Due to the oblique incidence at 6°–9°.
- Shadowing: The absorber walls "cover" part of the pattern.
- Phase edge shift: The thickness of the absorber introduces unwanted phase delays at the line edges. These shifts alter the coherence of the wavefront (*a concept discussed in Chapter 11.2*) and must be partially compensated for both digitally on the reticle and dynamically through the scanner adaptive mirrors during exposure.
- Best focus shift: The optimal focus changes depending on the direction.

M3D correction modifies the shapes on the reticle to compensate for these effects, often in combination with OPC.

#### **A.4 – SMO (Source Mask Optimization)**

SMO is one of the most advanced techniques in modern MDP. It consists of the joint optimization of:

- the light source (*Source*)
- the mask (*Reticle*)

to achieve the best possible Process Window. The Process Window represents the combined tolerance range between light dose and focus within which the lithographic process guarantees the correct printing of circuits. The goal of MDP is to maximize this window to make production immune to inevitable environmental and mechanical micro-fluctuations.

#### **Why SMO is necessary:**

With advanced nodes (*EUV, High-NA*), OPC alone is no longer sufficient. Resolution depends on:

- illumination shape (*pupil shaping*)
- angle of incidence
- polarization
- pattern pitch and orientation
- reticle 3D effects

SMO allows finding the optimal combination between:

- illumination pupil (*e.g., dipole, quadrupole, annular, freeform*)
- reticle modifications (*OPC + M3D + SRAF*).

#### **How SMO works**

Typical workflow:

1. Define a set of critical patterns (*lines, vias, SRAM, dense logic*).
2. Simulate lithography with different pupil shapes.
3. Optimize the reticle for each candidate source.
4. Evaluate the Process Window (*focus, dose, CD, uniformity*).
5. Select the best combination.

It is a high-dimensional optimization problem, often solved using:

- genetic algorithms
- gradient-based optimization
- machine learning (*surrogate models*)

Typical SMO results:

- 20–40% increase in the Process Window
- reduction in CD (*Critical Dimension*) variability
- improved resolution for extreme pitches
- reduced sensitivity to flare and shadowing

- better matching between X-Y directions (*crucial for High-NA*).

*Important note: SMO is computationally massive: it can require hundreds of thousands of core hours and dedicated clusters.*

### **A.5 – Fracturing**

The GDS file contains complex polygons. However, the e-beam writer can only write:

- rectangles
- trapezoids

Every polygon is therefore "fractured" into thousands or millions of primitives.

Useful details:

Fracturing must minimize the number of figures (*to reduce write time*).

- It must avoid overly thin trapezoids (*dose-related issues*).
- It must respect dose and overlapping limits.
- Hierarchical fracturing is often used to reduce file size.

### **A.6 – Typical file sizes**

The final file can reach:

- 100–500 GB for a standard EUV reticle.
- >1 TB for High-NA reticles (*due to more aggressive OPC-M3D-SMO*).

This weight results from:

- the massive number of primitives after fracturing.
- SRAF and assist features.
- M3D corrections.
- highly dense patterns (*SRAM, interconnects*).

Managing files of this size requires not only massive computing clusters for SMO but also ultra-fast network infrastructures to transfer data from the design house to the reticle shop without bit errors.

## **APPENDIX B – Reticle Writing**

### **Introduction**

The e-beam writer is a machine that uses a focused electron beam to "draw" the pattern on the reticle with nanometric resolution. It is similar to a printer, but with sub-nanometer precision and a beam that must follow millions of geometric primitives. How it works:

### **B.1 – Electron Resist**

1. The reticle surface is coated with an electron resist.
2. An electron beam is generated and focused by electromagnetic lenses.
3. The beam is rapidly deflected to follow the primitives (*rectangles/trapezoids*).
4. Where the beam strikes, the resist changes its solubility.
5. After development, the resist pattern remains.
6. The pattern is transferred to the absorber via etching (*selective plasma removal*).

*Technical Note on Electron Resists: Electron resists are much more sensitive and fine-grained compared to optical resists. The most commonly used are **PMMA** and **ZEP**, chosen for:*

- high resolution
- low roughness
- predictable dose response (*see table below*):

Resist	Type	Resolution	Sensitivity	Writing Speed	Typical Use
PMMA	Positive	5–10 nm	Low	Slow	High-precision critical layer
ZEP	Positive	10–15 nm	High	Fast	Standard EUV mask patterning

### **B.2 – Gaussian Beam Writer (GBW)**

- Single beam
- Ultra-high resolution
- Very slow.

*Technical note: Gaussian profile allows for extreme resolution, but the dose is not uniform at the edges so it requires:*

- dose corrections
- very low speed
- absolute thermal stability.

*It is the preferred tool for the most complex EUV reticles (SRAM, dense logic, extreme pitches).*

### **B.3 – Multi-Beam Writer (MBW)**

- Thousands of parallel micro-beams
- Much higher speed
- Slightly lower precision
- Used for less critical reticles or to reduce turnaround time.

*Technical Note on Multi-Beam Kinematics: A modern MBW can have 50,000–100,000 micro-beams, each with modular dose control. While the resolution is slightly lower than a GBW:*

- writing time decreases drastically
- the dose is more uniform
- productivity increases.

*It is essential for modern EUV reticles, providing the throughput and precision required for the complex pattern geometries of High-NA nodes.*

### **B.4 – Writing Times**

- MDP requires days of computation.
- e-beam writing takes 10–30 hours.

*Note: The time depends on:*

- pattern density
- number of SRAFs
- OPC aggressiveness
- type of writer (*GBW vs. MBW*)
- required dose.

High-NA reticles will require even longer processing times.

### **B.5 – Required Precision**

- Positioning error < 1 nm.
- Extremely stable beam uniformity.
- Vibrations and temperature controlled to the millikelvin level.

*Technical Note on Sub-Atomic Tolerances: In High-NA systems, the extreme optical contrast targets mean that even sub-nanometric variations on the reticle layout directly influence the aerial image intensity profile. To ensure that software-driven optimization models — such as Optical Proximity Correction (OPC) — can precisely predict and compensate for diffraction, the mask writing phase requires exceptional electro-mechanical and thermal stability. Minimizing any localized electron-beam drifting is essential to prevent excessive line-edge roughness, thereby safeguarding the final chip performance and yield across sub-2nm nodes.*

### **B.6 – Development and Etching**

After writing:

1. Resist Development: The exposed resist is removed.
2. Absorber Etching: The pattern is transferred into the three-dimensional relief structures.
3. Residual Resist Removal.

*Technical Note on Refractory Etching Chemistry: The etching of the EUV absorber (TaBN) must:*

- *maintain vertical sidewalls.*
- *avoid micro-notching.*
- *not damage the underlying multilayer.*

*This is one of the most delicate phases of the entire process.*

### **B.7 – Cleaning and Typical Defects**

Advanced Cleaning:

- While post-fabrication cleaning eliminates residual particles, the prevention of carbon and tin defects during scanner use relies on atomic hydrogen chemistry (see *Chapter 10*). Particles must be removed without damaging the multilayer.

Typical Defects:

- Particles (*Additive Defects*): These are dust specks or metallic residues that settle on top of the absorber or multilayer. They act as small shields that block EUV light, creating unwanted "holes" in the circuit pattern on the wafer.
- Multilayer Pits (*Substrate Defects*): A "pit" is a microscopic cavity or depression on the surface of the Mo/Si mirror. Even if only a few nanometers deep, the pit alters the light reflection angle (*phase distortion*). This causes an uncorrectable deformation of the projected image, often rendering the reticle unusable.
- Absorber Defects: Etching errors that leave material residues where they shouldn't be (*bridges*) or remove too much material (*breaks/opens*).
- Stitching Error: "Seaming" errors between the electron beam writing fields. A misalignment of even 1 nm between two zones written at different times creates a discontinuity that breaks the circuit lines (*see Chapter 5*).
- Non-uniform CD (*Critical Dimension*): Unintentional variations in the width of the pattern lines. If the CD is not constant across the entire reticle, the transistors on the wafer will have different electrical performances (*some faster, others slower*), compromising the processor uniformity.
- Fracturing Errors: Digital errors occurring during the decomposition of complex polygons into rectangles and trapezoids (*see Appendix A*). If the software "fractures" a geometry incorrectly, the e-beam writer will produce an incorrect shape, creating a structural defect in the circuit.

## **B.8 – Reticle Inspection and Qualification**

Every reticle is inspected using:

- AIMS (*simulates the image on the wafer*).
- Ultra-high-resolution SEM (*Scanning Electron Microscope*).
- Mask Blank Inspection.
- Pattern Inspection.
- Pellicle inspection.

Why this phase is critical:

- An EUV reticle costs between \$300,000 and \$1,000,000.
- A single defect can compromise thousands of wafers.
- Tolerances are sub-nanometric.

*Final note: The EUV reticle is a unique and irreplaceable object: if damaged, complete repair is impossible. For this reason, the qualification process is extremely rigorous.*

---

## REFERENCES

**[Ref.1]:** For an authoritative analysis of sub-3nm chip design costs and mask set valuation metrics, see: Khan, S., Mann, A., and Peterson, D., *The High Costs of Advanced Silicon: An Analysis of R&D and Photomask Economics in Leading-Edge Nodes*, Center for Security and Emerging Technology (CSET), Georgetown University, Washington D.C., Report Series.

**[Ref.2]:** For comprehensive industry modeling on TSMC advanced wafer pricing (N3/N2), tool amortization, and high-end accelerator bill of materials (BOM), see: Patel, D., and Semianalysis Research Team, *Advanced Node Economics: Wafer Pricing, Yield Models, and the Silicon Cost Curve*, Semianalysis Report Series, Architecture and Foundry Strategy.

**[Ref.3]:** For the definitive system architecture, resolution thresholds (8 nm), and throughput modeling of the 0.55 NA platform, see: Van Schoot, J., Van de Kerkhof, R., and ASML Lithography R&D Group, *High-NA EUV Lithography: Pushing Advanced Density Boundaries via the TWINSCAN EXE Platform*, ASML Holding N.V. Whitepaper Series, Technology and Product Strategy, Veldhoven, The Netherlands.

**[Ref.4]:** For comprehensive wave-front modeling of the 0.55 NA anamorphic system, asymmetric demagnification mechanics (4x in X-axis and 8x in Y-axis), and the preservation of standard 6-inch reticle ecosystems, see: Kürz, P., Rapp, J., and Carl Zeiss SMT Engineering Group, *Anamorphic High-NA EUV Lithography Optics: Design Challenges and Mirror Resonances in the Projection Optics Box (POB)*, Carl Zeiss SMT GmbH, Optical Systems Division, Oberkochen, Germany.

**[Ref.5]:** For technical specifications on the high-power pulsed CO<sub>2</sub> laser architecture, staged MOPA optical amplification, and the high-frequency 50 kHz droplet plasma ignition, see: Kösters, M., and TRUMPF Laser Systems R&D Division, *Industrial High-Power Laser Pulses for Laser-Produced-Plasma (LPP) EUV Radiation Generation*, TRUMPF GmbH + Co. KG, Technical Presentation and Innovation Reports, Ditzingen, Germany.

## APPENDIX C – Glossary of Acronyms

- **6-DOF** – 6 Degrees of Freedom
- **AI** – Artificial Intelligence
- **AIMS** – Aerial Image Measurement System
- **ALD** – Atomic Layer Deposition
- **ASIC** – Application-Specific Integrated Circuit
- **ASML** – Advanced Semiconductor Materials Lithography
- **BDU** – Beam Delivery Unit
- **BEOL** – *Back-End of Line*
- **CAR** – Chemically Amplified Resist
- **CD** – Critical Dimension
- **CFET** – Complementary Field-Effect Transistor
- **CMP** – Chemical Mechanical Polishing
- **CPU** – Central Processing Unit
- **CVD** – Chemical Vapor Deposition
- **DBR** – Distributed Bragg Reflector
- **DGL** – Dynamic Gas Lock
- **DOF** – Depth of Focus
- **DRAM** – Dynamic Random-Access Memory
- **DRC** – Design Rules
- **DUV** – Deep Ultraviolet
- **e-beam writer** – Electron Beam Writer
- **EDA** – Electronic Design Automation
- **EUV** – Extreme Ultraviolet
- **FAB** – Fabrication Plant
- **FEOL** – *Front-End of Line*
- **FOUP** – Front Opening Unified Pod
- **FPGA** – Field-Programmable Gate Array
- **FSM** – Fast Steering Mirrors
- **GAAFET** – Gate-All-Around Field-Effect Transistor
- **GBW** – Gaussian Beam Writer
- **GDSII** – Graphic Data System II
- **GPU** – Graphics Processing Unit
- **High-NA** – High Numerical Aperture
- **HIL** – Hardware-in-the-Loop
- **HVM** – High-Volume Manufacturing
- **IBS** – Ion Beam Sputtering
- **IC** – Integrated Circuit
- **IF** – Intermediate Focus
- **ILIAS** – Integrated Lens Interferometer At Scanner
- **IP** – Intellectual Property

- **LELE** – Litho-Etch-Litho-Etch
- **LER** – Line Edge Roughness
- **LTEM** – Low Thermal Expansion Material
- **LWIR** – Long-Wave Infrared
- **M3D** – Mask 3D Effects
- **MBI** – Mask Blank Inspection
- **MBW** – Multi-Beam Writer
- **MDP** – Mask Data Preparation
- **MEBES** – Manufacturing Electron Beam Exposure System
- **MOPA** – Master Oscillator Power Amplifier
- **MOR** – Metal-Oxide Resist
- **MPI** – Mask Pattern Inspection
- **NA** – Numerical Aperture
- **OASIS** – Open Artwork System Interchange Standard
- **OPC** – Optical Proximity Correction
- **PHz** – PetaHertz
- **PMMA** – Polymethyl Methacrylate
- **POB** – Projection Optics Box
- **PVD** – Physical Vapor Deposition
- **PWM** – Pulse-Width Modulation
- **REMA** – Reticle Masking
- **RF** – Radio Frequency
- **Ru** – Ruthenium
- **SADP** – Self-Aligned Double Patterning
- **SAQP** – Self-Aligned Quadruple Patterning
- **SEM** – Scanning Electron Microscope
- **SMO** – Source Mask Optimization
- **Sn** – Tin
- **SnH<sub>4</sub>** – Stannane
- **SoC** – System on Chip
- **SRAFs** – Sub-Resolution Assist Features
- **SRAM** – Static Random-Access Memory
- **TaBN** – Tantalum Boron Nitride
- **THz** – TeraHertz
- **TMSC** – Taiwan Semiconductor Manufacturing Company
- **UHV** – Ultra-High Vacuum
- **ULE** – Ultra-Low Expansion glass
- **VLSI** – Very Large Scale Integration
- **WSE** – Wafer Scale Engine
- **ZEP** – Zeon Electron Resist Positive

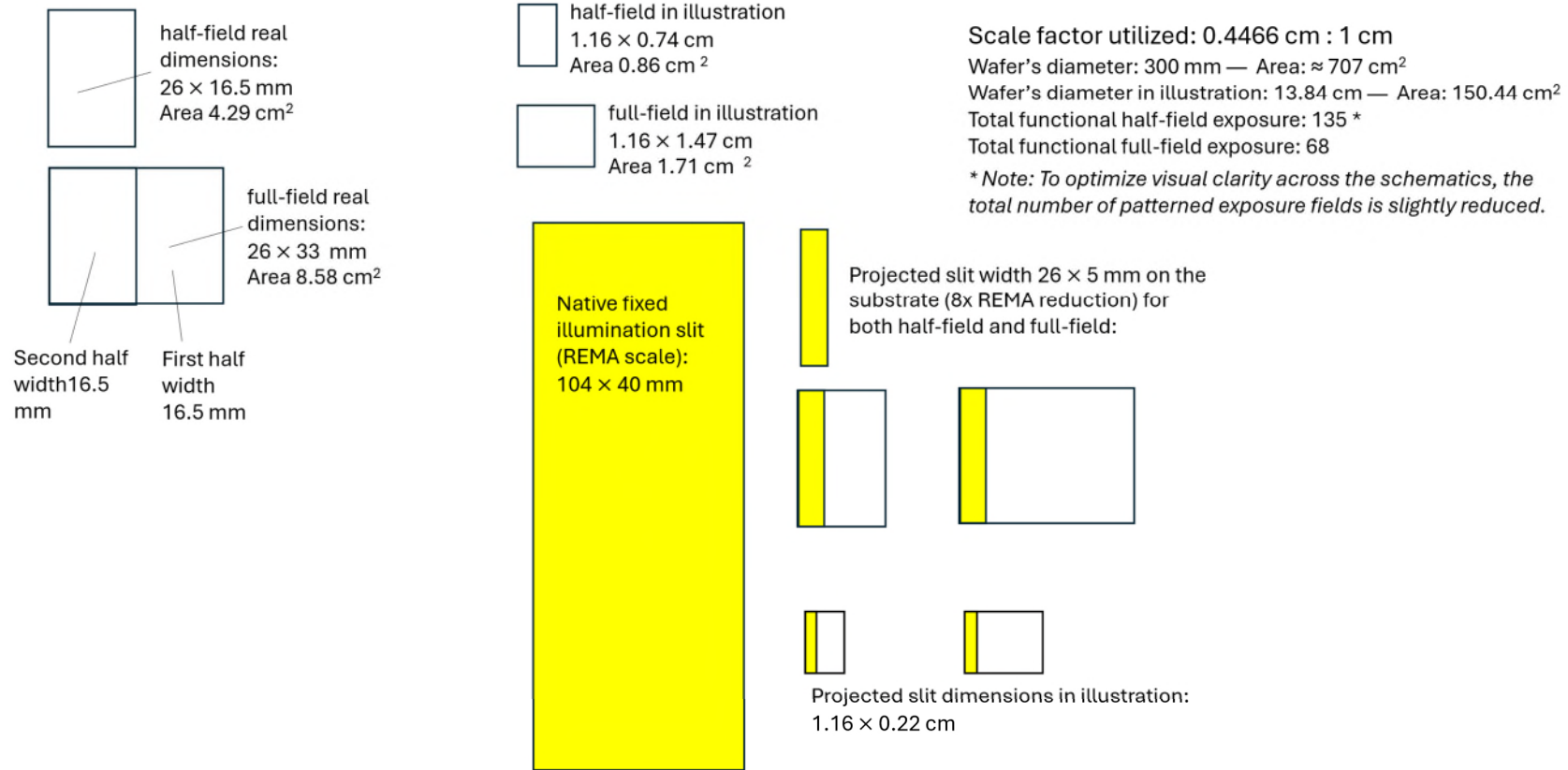


Figure 1 – System metric and scale parameters

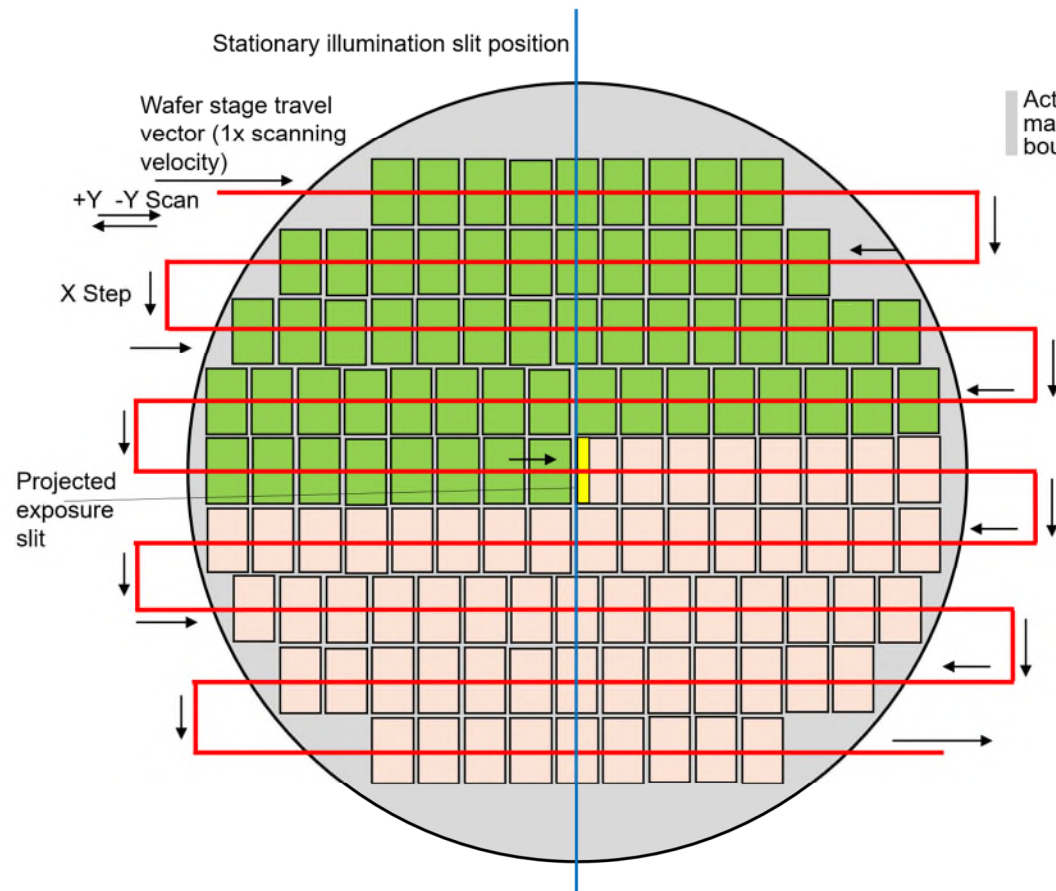


Figure 2 - EUV – half-field serpentine step-and-scan kinematics

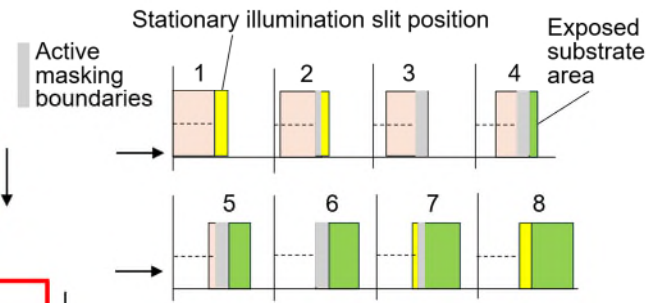


Figure 3 – Dynamic photoresist exposure integration

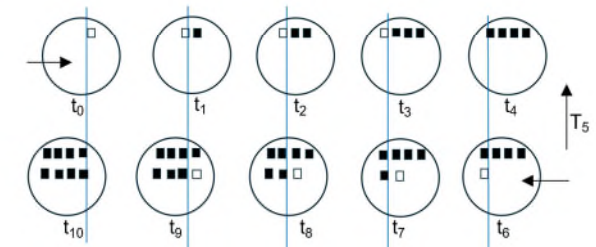


Figure 4 – Simplified two-row scan trajectory

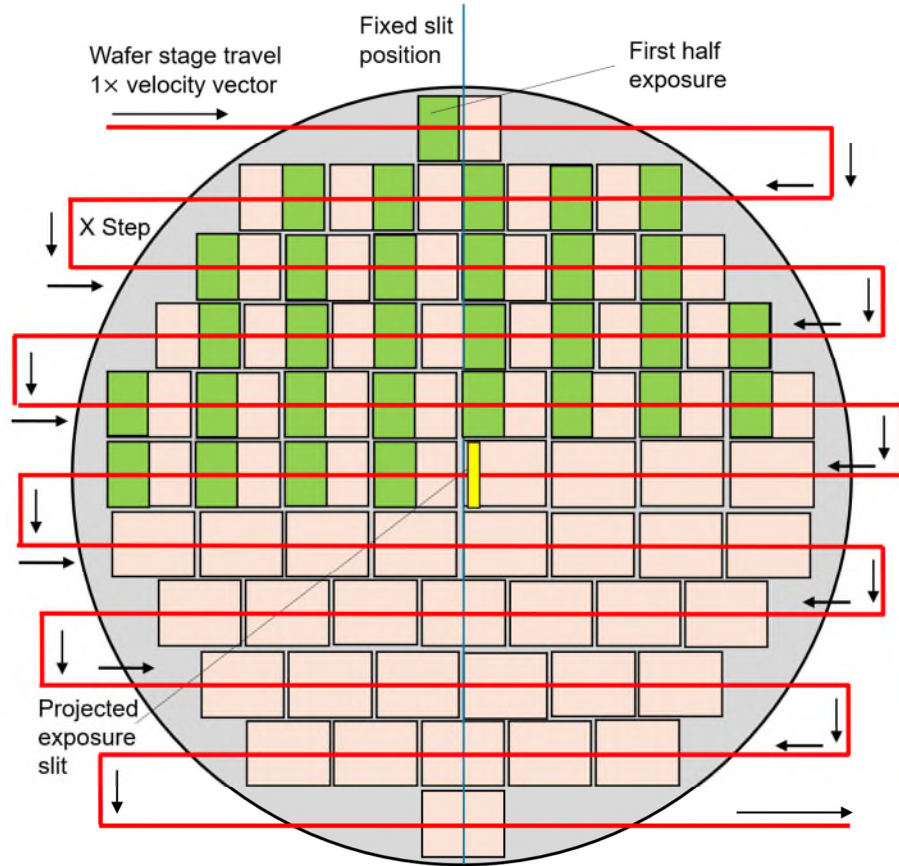


Fig. 5a: First-pass global exposure (Upper/Lateral fields)

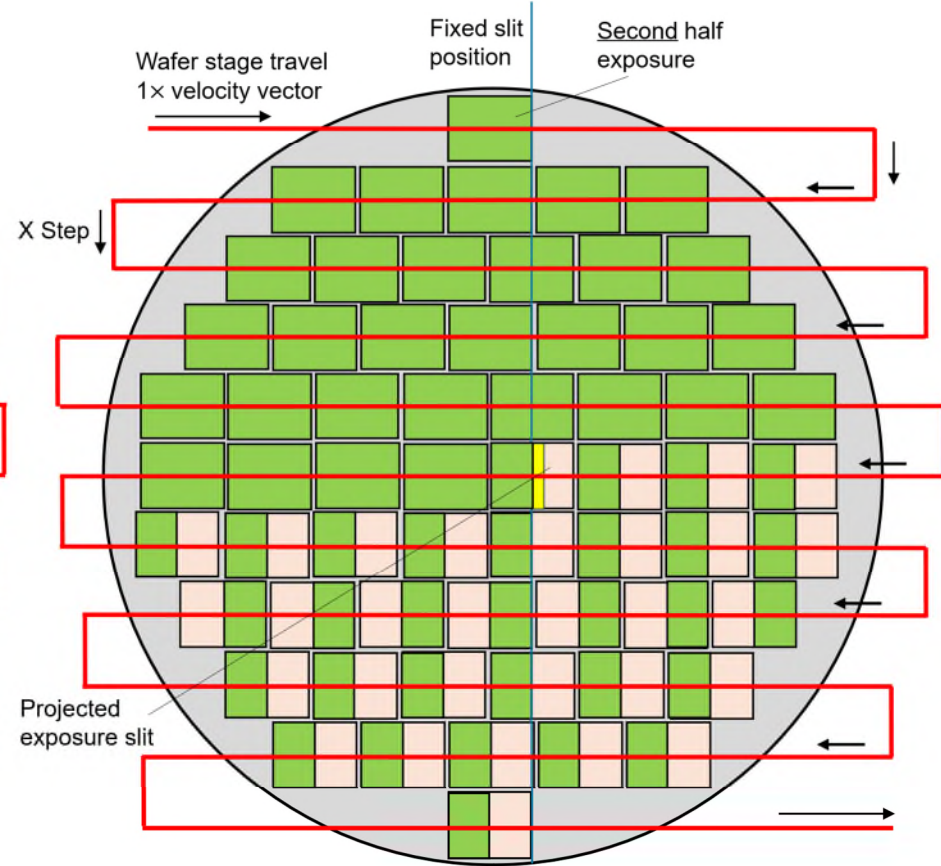


Fig. 5b: Second-pass global exposure with a 16.5 mm spatial offset (Lower/Lateral fields)

Figure 5 – High-NA EUV – Full-field exposure and dual-pass stitching