



Literature review of Semiconductor photo-lithography development and its history in chip manufacturing

Utkarsh Pandey

Maharaja Surajmal Institute of
Technology
Fire Station Road, Janakpuri, New
Delhi, 110058
utkarshpandey190205@gmail.com

Mamta Rani

Maharaja Surajmal Institute of
Technology
Fire Station Road, Janakpuri, New
Delhi, 110058
mamatholia@msit.com

Rakhi Kamra

Maharaja Surajmal Institute of
Technology
Fire Station Road, Janakpuri, New
Delhi, 110058
rakhikamra@msit.com

How to Cite this Article:

Pandey, U., Rani, M. & Kamra, R. (2026). Literature review of Semiconductor photo-lithography development and its history in chip manufacturing. International Journal of Creative and Open Research in Engineering and Management, <i>02</i></i>(6).
<https://doi.org/10.55041/ijcope.v2i6.043>

License:

This article is published under the terms of the Creative Commons Attribution 4.0 International License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.

© The Author(s). Published by International Journal of Creative and Open Research in Engineering and Management.



<https://doi.org/10.55041/ijcope.v2i6.043>

Abstract—Advancements in technology with regard to electronics and especially the computing devices has been attributed to the miniaturization and evolution of chipset. Their making involves the process of semiconductor photolithography which in simpler terms means of etching or ‘tattooing’ the chip with necessary circuit patterns and transistor networks. Chronological developments range from the earliest achieved lithography of micrometer dimension scale of semiconductor evolved from physical transistors to the miniaturization of transistor networks on nanometer scales for a small chip in present times.

Keywords—lithography, photolithography, EUVL, DUVL, semiconductors, chips, photomask, photoresist

I. INTRODUCTION

Semiconductor photolithography sits at the heart of every modern integrated circuit. Without it, the extraordinary density of today’s chips — tens of billions of transistors packed into a fingernail-sized piece of silicon — would simply be impossible. This review traces the development of that technology from its earliest laboratory demonstrations in the late 1950s through to the extreme-ultraviolet tools now in volume production at leading fabs worldwide.

The driving force behind photolithography’s evolution has always been straightforward: print smaller, print faster, print more reliably. Each chip generation demands features that the previous generation’s tools could not resolve, pushing engineers to find new light sources, optical designs, resist chemistries, and mask architectures. The process works

by projecting a circuit pattern drawn on a transparent reticle through a precision lens onto a wafer coated with a light-sensitive polymer. Where light falls the polymer changes; where it does not, it is protected. The spatial contrast between exposed and unexposed regions is what defines a transistor. [1][3]

The later sections consider not just what has been achieved but where the technology is genuinely running into walls. Fundamental physics, material constraints, and economic realities are converging in ways that make continued scaling difficult. Emerging approaches — from directed self-assembly to high-NA EUV — are already being piloted, though none has yet fully displaced the incumbent. [1][5]



In concrete terms, the journey covered here spans roughly six orders of magnitude in feature size: from the 4 μm critical dimensions achieved by the earliest projection tools in the late 1960s to the sub-10 nm geometries now being patterned in commercial production — a compression achieved over roughly six decades of relentless incremental improvement. [1][13]

II. HISTORY

A. Initialization and working

The possibility of building transistors directly onto a silicon wafer — rather than assembling them as discrete components — emerged around the time of third-generation computing hardware in the early 1960s. That transition changed everything: a transistor on a wafer could be made far smaller, consumed less power, and could be interconnected with others without manual wiring. The question then became how to define such small structures with sufficient precision and repeatability, and the answer was photolithography. [3][4]

Between the late 1950s and mid-1960s, the planar process developed by Jean Hoerni at Fairchild Semiconductor established the template for all subsequent IC fabrication. The wafer was thermally oxidised to grow an insulating SiO_2 layer, then coated with a UV-sensitive polymer. A photomask carrying the circuit pattern was placed near the resist surface, and ultraviolet light transferred the pattern by selectively exposing the polymer. A chemical developer removed either the exposed or unexposed regions, leaving a stencil through which etching could cut into the underlying oxide. Metallisation and doping completed the structure. The elegance of the process lay in its repeatability: the same basic sequence, applied with different masks and materials, could build up arbitrarily complex circuits layer by layer. [3][5][6]

B. Framework in evolution

2nd generation of computing pushed for a further innovation of having multiple circuitries on a singular layer instead of transistors. During the 3rd generation, the transistors actually remained the root base for the all upcoming chip manufacturing and foundation stone semiconductor fabrication. The large scale integration and improvement jumps from previous photo chemical lithography proved to be a better iteration for chip fabrication. Now instead of the direct silica oxide, the Silica substrates became the alternatives which were made pre doped with elements and compounds like photo resist chemicals, doped mixture (GaAlS) with fine chemical layering and improvised photomasking.[11]

By the duration of 4th generation of computers, introduction of MOSFETs and their further integration to the VLSI of semiconductor making increased the number of layered transistors, microlevel graded transistor fabrication, shrinking down the size for having more transistor network set on the wafer area. [11][9]

III. COMPONENTS

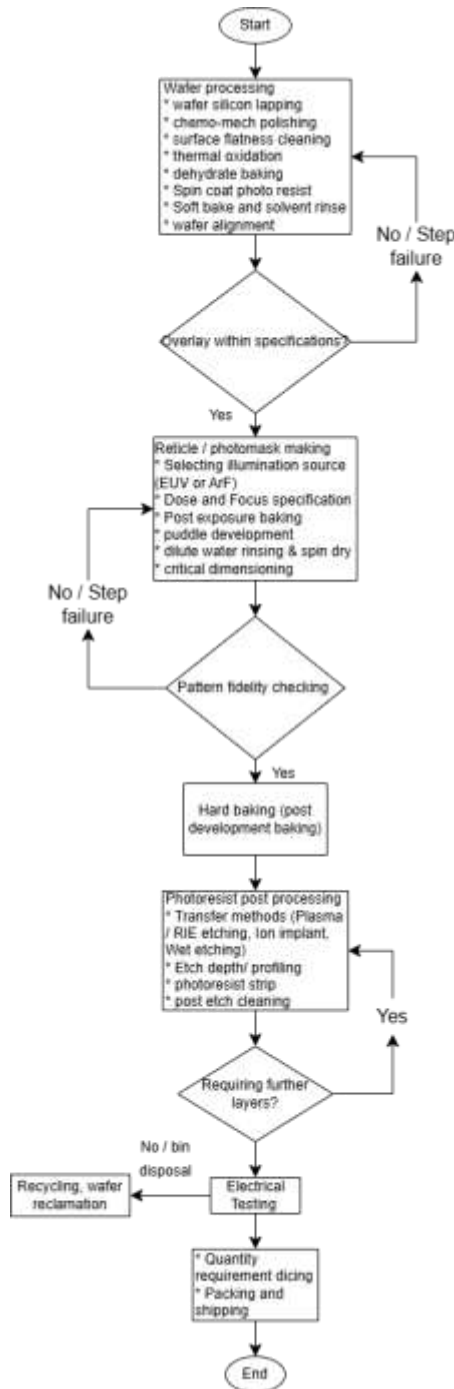
The earliest achieved pivoted breakthrough was during the mid-to-late 1950's when Jules Andrus and his colleague

Walter Bond developed the first contact proximity based photomasking method at the Bell Telephone Labs. Existing PCB manufacturing techniques were used to develop the prototype. That procedure was known as 'photo engraving' where a holed metal mask/stencil was placed over wafer top spraying it with Apiezon wax (an inorganic electronic paste/wax), carving the microscopic network and rinsing it with hydrochloric acid. [3][6]

A. Photomask

A photomask is a precision optical template — typically a polished quartz plate — that carries the circuit design for one layer of the chip. Its function is deceptively simple: when illuminated, it casts a shadow of the desired pattern onto the resist-coated wafer. In practice, the mask must be manufactured to accuracies of a fraction of a nanometre, inspected for defects smaller than the wavelength of the inspection light, and protected from airborne contamination throughout its working life. [8][6]

These stencils are undergone with a photo resist process where the wafer is coated with light sensitive materials, exposing to illumination and developing exposed material for desired pattern. Because the singular layer is not the whole system for the chip, different layers are etched over with different photo masks and each accumulating top layer is carefully carved out on the same photomask.[8]



Overall Semiconductor processing steps

Evolution wise, photo masks were made out of metal templates with desired holes and wax treatment. In next iteration, it involved two stage processes: firstly carving the circuit pattern on red colour sheet (called Rubylith, a 100x – 400x artwork of patterns), illuminated with green light for creating the photocopy at 10 times it's magnification.[6][5]

Further miniaturization of ICs became complicated manually. Hence a pattern generating system was designed in 1968, a 1600PG pattern generator using rotatable mechanic apertures. Modern photomasks are made using ultra-precise nanofabrication processes. A polished quartz or fused silica plate is coated with a thin chromium layer, then covered with electron-beam resist. Using electron-beam lithography,

intricate circuit patterns are written directly onto the resist at nanometer resolution. The exposed resist is developed, and the underlying chromium is etched away to form the opaque features, leaving transparent regions where light must pass. Advanced masks may include phase-shifting or multiple patterning layers for extreme ultraviolet (EUV) lithography. Finally, the mask undergoes rigorous inspection, defect repair with focused ion beams, cleaning, and pellicle mounting to ensure flawless semiconductor production.[8][9]

B. Silicon wafers

The base that is eventually laid for inclusion into microchip sets are the metalloids. Semiconductors help controlled and minimised usage while simultaneously performing charge carrier-based operations. Their evolution is closely tied to advancements in lithography and crystal growth tech.[9]

During the early developments, the initial wafer sizes in 1960s-70s were 1 inch or 25 mm for contact proximity exposure tools. By the late 1970s, their size was doubled to 2 inches or 50 mm. Subsequent extension led to 3 inch and 4 inches of wafers by 1976. The utmost minimum critical dimension achieved was 4 μm.[9]

For the next two decades following the expansion of size, it was an impressive boost for chip per hour production and reducing manufacturing costs. Their old treatment of pre doing similar to that of photoresist and photomasks were also made distinctive. In 1977, 5 inch or 125 mm wafers were introduced and by the end of that decade, 6 inch wafers had been brought in; although those 150 mm wafers necessitated larger photomasks, requiring additional investment in photo mask making production and scaling up of their treatment processing. By early 1990's, wafer sizes were stretched to 200 mm in area.[9]

Modern wafer sizes were increased further to another 100 mm of area. A proposed idea of increasing silicon wafers to 450 mm (or 18 inches) had been proposed in 2012 but was shelved off due to another massive funded investment requirement, as the world had been recovering from the aftermath effects of the 'Great Recession', holding off several projects and cancelling major decisions in various sectors of several countries.[9]

Diameter of wafers influences lithographic processes. Mirror projection systems used to handle upto 6 inches by exposing the wafer disc in a single scan. For larger wafers or simply improved resolution, the projection imaging with smaller image fields were made necessary to be covered for the discs, leading to development of wafer steppers. Increase in wafer diameters led to challenges in mask manufacturing, where the preset dimensions had to be reworked to fit the area, needing higher concave mirrors and photomasks for the refractive optical systems.[9]

Wafer steppers, like the Geological society of America (GSA)'s Mann 480 Direct Step-on Wafer or DSW brought in 1978 was designed to handle larger wafers by performing exposures in smaller, divided regions, minimizing issues like flatness and nonlinear process distortions. Similarly, Advance Semiconductor Materials Lithography aka the ASML Company made TWINSCAN machines, introduced in 2001 by utilizing two independent wafer stages for the



present 6 inch wafers, allowing alternative exposure and calibration, significantly boosting productivity.[9][1]

C. Photoresist

Photoresist is arguably the most chemically complex material in the lithographic stack. Early resists such as Kodak's KPR formulation from the 1950s were negative-tone: the polymer cross-linked where light struck it and remained after development, while the unexposed regions dissolved. Negative resists suffered from swelling during development which distorted fine features, and were eventually displaced for most applications by positive-tone chemistries in which the exposed regions dissolve. [6][5]

The transition to deep ultraviolet wavelengths in the 1990s demanded a fundamentally different resist architecture. Conventional photoactive compounds absorbed too weakly at 248 nm and 193 nm to provide adequate photospeed. The solution was the chemically amplified resist (CAR), in which a single absorbed photon generates an acid molecule that catalyses many deprotection reactions during the post-exposure bake — providing the sensitivity needed at DUV wavelengths without sacrificing resolution. CARs remain the dominant resist platform today, though their stochastic behaviour at EUV doses is a growing concern for sub-20 nm patterning. [14]

D. Exposure tools

The exposure tool — the scanner or stepper that projects the mask image onto the wafer — is the most capital-intensive piece of equipment in a semiconductor fab, and its capabilities largely define what critical dimensions are achievable. Early contact and proximity printers of the 1960s and early 1970s were relatively inexpensive but suffered from mask damage and diffraction at the proximity gap that limited resolution to roughly 2–3 μm . [5][7]

Projection lithography became prevalent with the introduction of Perkin-Elmer's unit-magnification wafer scanners in the 1970s, followed by wafer steppers from companies like GCA, Canon, Nikon, and ASML.[7][6]

These tools evolved significantly, moving from mercury arc lamps to deep ultraviolet (DUV) KrF and ArF excimer lasers, and eventually to extreme ultraviolet (EUV) lithography[13][7][1]

E. Light sources

The history of photolithography is, in large part, a history of light sources. Each shift to a shorter wavelength has unlocked a new resolution regime and required a new generation of optical materials, resist chemistry, and process engineering. Mercury arc lamps served the industry from the 1960s through the early 1990s, first at the 436 nm g-line and then at the 365 nm i-line, enabling features down to about 250 nm. When i-line reached its practical limit, the industry turned to excimer lasers: KrF at 248 nm in the mid-1990s, then ArF at 193 nm from the early 2000s onward. [6][7][8]

Extreme ultraviolet lithography at 13.5 nm represents a wavelength reduction by a factor of fourteen compared to ArF — a step that required reinventing virtually every component of the optical system, since no solid material transmits at that wavelength. EUV light is generated by a laser-pulsed tin plasma, collected by a large ellipsoidal

mirror, and steered through a vacuum-sealed path using multilayer Bragg mirrors that reflect only about 70% of the incident flux at each surface. The engineering challenge is extraordinary, but the resolution reward — sub-20 nm features in a single patterning step — justifies the investment at the leading edge. [6][7][8][16]

F. Imaging Optics

The optical systems within exposure tools, such as projection lenses, are vital for focusing and transferring the pattern. Early systems used lenses with varying numerical apertures (NA) and reduction ratios.[6][5]

Advancements included reduction lenses for photo-repeaters, with Nikon developing ultra-high-resolution optical systems in the 1960s. The development of these optics, including mirror-projection systems and later refractive lenses, was critical for achieving finer resolutions.[7][4]

IV. PROCESS

Semiconductor photolithography is best understood not as a single step but as a carefully ordered sequence of physical and chemical operations, each of which must succeed before the next can proceed. What follows traces that sequence from first principles, with the governing equations embedded where they arise naturally. [1][6][17]

A. Substrate preparation and oxidation

Everything begins with a bare silicon wafer that must be made atomically clean before any patterning can take place. The standard cleaning protocol — a two-stage RCA wash — strips organic residues in a hydrogen-peroxide and ammonium-hydroxide bath, then removes metallic ions in a dilute hydrochloric-acid solution. Once clean, the wafer enters a furnace at 900–1,200°C in oxygen or steam, and a thermal oxide (SiO_2) layer grows uniformly across the surface. The thickness of this oxide matters because it influences resist adhesion and the optical interference behaviour of the substrate during exposure. [1][6]

B. Photoresist coating and soft bake

With the oxidised wafer dehydrated to remove surface moisture, a liquid photoresist is dispensed at the centre and the wafer is spun at 1,500–4,000 RPM. Centrifugal force spreads the resist into a film whose thickness — typically 50–500 nm for advanced nodes — is governed by spin speed and resist viscosity. Thickness uniformity matters because even small variation translates directly into dose non-uniformity during exposure. The wafer is then soft-baked at 90–110°C for about sixty seconds to expel the casting solvent and stabilise the film. [6][5]

C. Alignment and exposure

The resist-coated wafer is loaded into the exposure tool, where alignment marks from earlier layers are read and the stage is corrected for translation, rotation, and magnification errors. The minimum half-pitch printable is governed by $HP_{\min} = k_1 \times \lambda / NA$, where λ is the exposure wavelength, NA is the numerical aperture of the projection lens, and k_1 is a process factor that cannot fall below 0.25 — setting a hard single-exposure resolution floor. For 193 nm ArF immersion



tools, water between the lens and wafer raises the effective NA to approximately 1.35, placing the single-exposure limit at roughly 36 nm half-pitch; going below this requires multiple patterning passes. [1][17]

D. Post-exposure bake and development

In chemically amplified resists (CARs), the latent image formed by exposure is a spatial map of photoacid concentration. A post-exposure bake at 110–130°C drives acid diffusion through the polymer matrix, catalysing deprotection reactions that alter resist solubility. The spread of acid during this step introduces blur that places a practical lower bound on feature sharpness. After the bake, the wafer is immersed in developer — typically 2.38% TMAH — which dissolves the exposed regions in a positive-tone process, leaving the circuit pattern as a relief structure in the resist film. [1][6][14]

E. Depth of focus and process window

The useful range of focus and dose conditions within which printed features remain within specification is governed largely by the depth of focus (DOF), expressed as $DOF = k_2 \times n\lambda / NA^2$, where n is the refractive index of the imaging medium and k_2 is a process constant between 0.4 and 1.0. This equation reveals a painful trade-off: raising NA to improve resolution shrinks DOF as the square of NA. At NA = 1.35 with 193 nm light, DOF is only about 55 nm — thinner than many resist films — forcing wafer flatness and focus control to sub-nanometre precision across the full scanner field. [1][17]

F. Pattern transfer: etch and implantation

Once the resist pattern has been verified by CD-SEM, it serves as a stencil for the underlying film. In plasma etching, reactive gases such as CF_4 , Cl_2 , or HBr are ionised into a plasma that transfers the resist pattern faithfully into the oxide, metal, or semiconductor beneath. Where ion implantation is needed instead, accelerated boron, arsenic, or phosphorus atoms are driven into the unmasked silicon at keV–MeV energies, altering its electrical properties in precisely defined regions. [1][6]

G. Image contrast and inspection

Running alongside every step is a concern for image contrast — the degree to which bright and dark regions of the aerial image are distinguishable to the resist. For a periodic grating, contrast C is defined as $C = (I_{max} - I_{min}) / (I_{max} + I_{min})$. A value of $C = 1$ represents a perfect image; $C = 0$ means the pattern is invisible. Lens aberrations, flare, mask quality, and resist blur all degrade contrast below unity, and optical proximity correction (OPC) together with source-mask optimisation (SMO) are applied computationally to partially recover it at the resolution limit. [1][2][6][17]

H. Strip, clean, and layer iteration

After etching or implantation, the photoresist is removed first by an oxygen plasma ash, then by a wet solvent rinse, and finally by an SC-1 clean that restores surface chemistry. The wafer then re-enters the deposition stage for the next layer, and the entire lithographic sequence repeats with a fresh mask. Modern chips require fifty to over a hundred such

iterations — the cumulative precision of every cycle determines whether a transistor switches correctly, whether a metal line conducts without leakage, and whether the finished die meets its specification. [1][9]

V. LITHOGRAPHY TECHNIQUES

Modern microfabrication relies on several lithography methods, each with distinct physics, resolution limits, and uses. Optical (photolithography) — using ultraviolet light and photomasks — remains the workhorse for high-volume chipmaking. Immersion lithography (an advanced form of optical lithography) uses a liquid medium to increase numerical aperture (NA) and push resolution further. Extreme Ultraviolet (EUV) lithography is the newest optical technique, using 13.5 nm wavelengths and mirrors for the finest critical dimensions.[12][13]

Electron-Beam (E-beam) lithography uses focused electrons for mask writing or very-high-resolution patterns at low volume. Nanoimprint Lithography (NIL) physically stamps patterns from a hard mold, bypassing optics. Directed Self-Assembly (DSA) uses block copolymers that automatically form regular patterns guided by lithographically pre-patterned templates.[12][13]

- Optical (Deep UltraViolet) Lithography

Conventional optical lithography projects a circuit mask image onto a photoresist-covered wafer using UV light and complex lenses. Common wavelengths include 248 nm (KrF laser) and 193 nm (ArF laser). By the 2020s, the industry almost exclusively uses 193 nm ArF (often chemically amplified resists) for advanced nodes.[13][1]

The resolution limit follows the Rayleigh criterion ($CD \approx k_1 \cdot \lambda / NA$), so increasing NA or reducing wavelength improves resolution. ASML's highest-NA 193 nm steppers reach $NA \approx 1.35$ (using very large lenses) and can thus print features <40 nm. In practice, immersion lithography is used: a high-purity water layer is placed between lens and wafer, boosting NA to 1.35.[13]

The large multi-element lens of a 193 nm immersion scanner is utilized. Water between the lens and wafer raises NA to 1.35, allowing sub-40 nm features at 193 nm. Optical lithography's strengths are its very high throughput (hundreds of wafers/hour) and maturity: it is used for most mask layers (logic, memory, analog) today.

Limitations are set by diffraction: below ~40 nm pitch (for 193 nm) it requires multiple patterning. Hence advanced nodes (7 nm/5 nm logic, <1 Ynm DRAM) typically use two- or quadruple-patterning with 193 nm light (splitting each mask into two or more exposures) to achieve smaller pitches. Multi-patterning greatly increases mask count and overlay complexity.[1]

Despite these workarounds, conventional 193 nm lithography remains cost-effective for all but the smallest features. It achieves high



yield and uniformity at low cost per wafer, but further shrinking beyond ~30 nm pitch requires fundamentally different approaches.[13]

- Extreme Ultra-Violet (EUV) Lithography

EUV lithography uses much shorter 13.5 nm wavelength light, projecting patterns with reflective mirrors (because EUV is absorbed by glass). ASML's NXE series (e.g. 3600D) use this technology in high-volume fabs. EUV beams are generated by laser-pulsed tin plasmas and directed through a vacuum chamber with multilayer Bragg mirrors. The standard EUV tool has $NA \approx 0.33$ (new high-NA EUV is 0.55).[1][8][12]

Inside an EUV scanner of ASML NXE3400 has multiple ultra-smooth mirrors reflecting the 13.5 nm beam (purple). EUV requires vacuum optics because no lens transmits EUV light. Because of its shorter wavelength, EUV can print much smaller features in one exposure – typically around 20–30 nm half-pitch today (≈ 7 nm logic node).[1][8]

ASML notes that even though high-NA EUV (0.55) uses “lower” NA than 193 nm immersion, the 13.5 nm wavelength yields smaller features. In practice, current EUV scanners ($NA=0.33$) print ≈ 40 nm pitch (20 nm half-pitch) with single patterning (TSMC 7nm logic, Intel 10nm, Samsung 8nm). Future high-NA EUV will push down to 2–3 nm nodes. Strengths of EUV: it reduces the need for multiple patterning, simplifying pattern stacks and improving alignment.[1][15]



EUV (condensed) process

It also enables higher aspect ratio features (thicker resists). However, EUV has limitations: source power is still low (limiting wafer throughput), mask blanks must be nearly defect-free, and resist stochastic effects are severe (requiring higher dose). EUV adoption is currently in high-volume use for only select layers (logic SRAM and M1 metal, some advanced DRAM bitcells) due to these challenges.[1][10][14]

- Electron-Beam (or E-Beam) Lithography

E-beam lithography uses a finely focused electron beam to “write” patterns directly into resist. Commercial tools exist for direct writing on wafers and for mask fabrication. E-beam resolution can easily reach single-digit nanometers (sub-5 nm), because electrons have very short wavelength.[5][12]



However, e-beam is serial – scanning beam across the surface – so throughput is extremely low (often only a few cm² per minute).[12]

As a result, e-beam is mainly used for mask writing, circuit prototyping, or specialty devices (photonics, MEMS) where volume is low. Multi-beam e-beam systems (array of many beams) are being developed to boost throughput for wafer patterning, but none have replaced optical litho in production yet. A recent review notes that combining e-beam with other techniques (EUV, DUV, NIL, DSA) can help achieve ultimate resolution.[12][13]

To put it simple, strength of beam is ultimate resolution with flexibility requiring no masks while it is slow throughput and costs high per wafer.[5][12]

- **Nanoimprint Lithography**

NIL is a mold-based technique: a hard template with nanostructures is physically pressed into a resist layer on the wafer, then cured to replicate the pattern. Because it is a contact process (rather than optical imaging), NIL is not limited by light wavelength. It can achieve extremely high resolution – sub-10 nm features have been demonstrated – and can pattern dense two-dimensional structures with ease.[12][13]

NIL also omits the photoresist development step (once imprinted and cured, there is nothing to dissolve), which simplifies processing and reduces chemical waste. notes that laboratory NIL has achieved <10 nm resolution, surpassing the diffraction limit of traditional photolithography. These advantages make NIL attractive for cost-sensitive, high-resolution patterning (e.g. bit-patterned media, MEMS, LEDs, or emerging chip technologies).[12]

However, NIL faces challenges: any defect or wear on the mold transfers to every wafer, so mask (stamp) fabrication must be near-perfect. Alignment of each imprint for multi-level alignment is also difficult. Throughput can be high in theory (a stamp can pattern an entire wafer at once), but handling (loading/unloading, mold release) adds cycle time. Despite these issues, NIL is being used in some manufacturing (e.g. hard drive media, and experimental semiconductor production) and in R&D.[12]

- **Directed Self-Assembly (DSA) Lithography**

DSA is a hybrid technique: first, a coarse “guide” pattern is made by conventional lithography (optical or e-beam). Then a thin film of block copolymer (two polymer blocks that phase-separate) is applied. The polymer self-assembles into a regular nanostructure (lines or holes) within the guide, which is transferred as

the final pattern. This can multiply the density of patterns by 2×, 4× or more beyond the guide pitch. DSA has produced sub-10 nm half-pitches in labs when combined with EUV or 193 nm guides.[12][13]

A recent review highlights that DSA can enhance resolution and defect tolerance: “when combined with common lithography techniques (EUV, DUV, EBL, NIL), DSA enhances pattern resolution, repairs defects, and improves critical dimension uniformity”. In practice, DSA is used to shrink and refine patterns (e.g. contact holes, via arrays) that would be at the limit of conventional litho. It also helps reduce random LER/ LWR by averaging polymer domains.[13]

The downsides are added process complexity (materials, etch integration) and the fact that DSA is not a stand-alone lithography – it always depends on a guiding pattern. Industry adoption is growing: companies like IBM, Intel, imec, and Samsung have piloted DSA for DRAM and logic layers (e.g. finFET gate fins, contact arrays) to extend 193 nm and EUV techniques.[11][13]

VI. LIMITATIONS

A. Physical limitations and multi-patterning

Every optical lithography system is ultimately constrained by the same wave physics: diffraction. When light passes through or reflects from a patterned mask, the features act as apertures that bend and spread wavefronts in ways that blur the projected image. The minimum half-pitch printable is set by the Rayleigh criterion, and the process factor k_1 can only be pushed so far below its physical minimum of 0.25 before the aerial image loses enough contrast that resist can no longer resolve the pattern. For 193 nm ArF with water immersion at NA = 1.35, this places a hard single-exposure floor at about 36 nm half-pitch, which is why the industry turned to multiple patterning to reach 7 nm and 5 nm nodes. [1][13]

Even Extreme-UV (13.5 nm) lithography is pushing its limits: practical single-pass EUV resolution is around 16–18 nm half-pitch ($k_1 \approx 0.3-0.4$). In other words, even though EUV’s shorter wavelength enables ~22 nm pitch features, stochastic effects and process window constraints create an extra “resolution gap” of ~5 nm beyond the optical limit. Pushing below this requires either very high-numerical-aperture tools (e.g. 0.55 NA EUV) or still more pattern splits, both of which have their own drawbacks.[9][1][10]

B. Stochastic effects and photoresist limitations

At EUV doses, the number of photons arriving at any given resist volume is small enough that shot noise becomes a dominant source of variation. Each EUV photon generates only a handful of photoelectrons, and their spatial distribution across the resist film is inherently random. Even with a mathematically perfect aerial image, this photon-counting statistics produces random variation in acid generation across the exposed area, translating into line-edge roughness,



stochastic bridging between adjacent features, and random breaks in lines that should be continuous. [10][1]

Such stochastic defects directly translate into yield loss (e.g. SRAM failures) and impose a need for much higher doses or repetition. However, higher dose lowers throughput. Moreover, resolving tiny features forces very thin resist films (to stay in focus), which amplifies line-edge roughness (LER) and collapse defects.[10]

In sum, resist chemistry and processing are heavily stressed: chemically-amplified resists (the current workhorse) struggle with shot noise and LER, and new materials (e.g. metal-oxide or inorganic resists) bring trade-offs in sensitivity, roughness and etch compatibility. The net effect is that lithography process windows narrow and stochastic defectivity becomes a hard limit on resolution.[10][14]

C. Overlay and alignment Challenges

Every lithographic layer must land accurately on top of the ones already printed. The overlay budget — the allowable misregistration between any two layers — scales roughly with the minimum feature size, and at advanced nodes it is measured in single-digit nanometres. For a 5 nm node chip with a 22 nm metal pitch, overlay error must be kept below 2–3 nm across a 300 mm wafer whose temperature may vary by a fraction of a degree during processing — a thermal expansion of several nanometres per metre. Achieving this requires active stage correction, wafer-level distortion models, and in some cases fully self-aligned process architectures. [1][9]

In EUV particularly, 3D mask topography (“shadowing” from the absorber pattern) and lens aberrations at field edges can warp printed features, making sub-2 nm overlay very hard. International roadmaps note that overlay budgets now consume ~40% of total lithography error allowance. To mitigate this, fabs increasingly use self-aligned process flows (self-aligned vias, gate-first approaches, etc.) that relieve overlay demands. But these add process complexity.[1]

Overall, alignment control is a major bottleneck; even small registration shifts can scrap entire wafers when billions of transistors are involved.[1][9]

D. Optical limitations

The optical train of an EUV scanner is unlike anything in conventional lithography. Tin droplets roughly 30 μm in diameter, generated at 50,000 per second, are struck by a high-power CO₂ laser that ionises them into a plasma hot enough to emit 13.5 nm light. A large ellipsoidal collector mirror captures the emitted radiation and focuses it into the illuminator. Every subsequent mirror in the projection optics reflects only about 68–72% of the incident EUV flux, so a system with six mirrors delivers roughly 10% of the collected power to the wafer — explaining why EUV source power remains the dominant throughput bottleneck. [1][8]

E. Mask Complexity and defectivity

Modern chips use dozens of mask layers (50–100+ per device), so mask quality is critical. EUV masks are especially demanding: they are reflective Mo/Si multilayers on quartz, and even a sub-nanometer defect buried in the stack can print as a major CD error. In fact, studies show that atomic-scale

height bumps (~0.3–0.6 nm tall, 50–100 nm wide) on the mask can distort pattern widths by ~10%.[1][9]

Yet current actinic (13.5 nm) inspection tools have difficulty finding these tiny defects. A single undetected defect on an EUV mask could ruin thousands of chips. Pellicles (protective films) help for DUV masks, but EUV pellicles must transmit 85–90% of EUV light and survive intense heating.[1]

Today's EUV pellicles (ultra-thin Si membranes) are fragile, expensive (\$10k each) and themselves cause pattern distortion and heat-loading. Even with pellicles, masks require frequent cleaning and eventually replacement, shortening lifetime. Altogether, mask-making and inspection are huge cost and risk factors. Roadmaps emphasize that, to keep up with more layers and smaller features, defect rates per layer (especially EUV exposures) would have to improve orders of magnitude each node – an extremely daunting prospect.[1][15]

F. Equipment and throughput constraints

Cutting-edge scanners are enormously expensive and maintenance-heavy. Current EUV scanners (e.g. ASML NXE:3600D) cost on the order of \$170–350 million each. Because ASML is essentially the sole source, lead times are long (often >12–18 months).[1][15]

Each subsystem in an EUV tool is pushed to its limits: high-power CO₂ lasers (for LPP sources) erode optical components in ~1200 hours, tin-droplet injectors clog weekly, vacuum pumps and seals wear, etc. [45†L192-L201]. Maintaining dozens of such complex tools is a major challenge – fabs increasingly rely on predictive ML maintenance.[1][7]

Throughput is another issue: raw EUV sources deliver only a few hundred watts, so exposing a full wafer takes significant time. Higher litho dose (to reduce stochastic defects) directly cuts wafers-per-hour, making EUV patterning slower (and thus more expensive) than indicated by spec sheets. By contrast, mature ArF immersion scanners have higher throughput per cycle but need two or more passes for critical layers, which also multiplies time and overlay steps.[1][10]

In summary, advanced lithography tools demand heavy infrastructure (ultra-clean vacuum, high-purity gases) and have operational costs that scale steeply with node.[1]

G. Materialistic and processing integration troubles

The lithography step is only one part of chip processing, and it must co-optimize with etch, deposition, CMP, etc. Pushing NA and resolution tightens focus depth: e.g., high-NA (0.55) EUV will have only a few hundred nanometers of focus, forcing very thin resist (<30 nm) and extremely flat wafers. Thinner resists amplify line collapse and roughness, so fabs are developing new hard-mask and organic underlayers, as well as dry-development processes.[1][10]

Each new material or process (e.g. exotic resists, high-index immersion fluids) adds its own R&D and compatibility headaches. Furthermore, as chipmakers try to extend 193 nm immersion (e.g. NA=1.35) with “hyper-NA” fluids or mirrors, they hit limits of chemistry and lens materials. In practice, every litho improvement (thinner resist, novel



topcoats, pattern transfer steps) tends to shrink process windows and require tighter controls.[1][15]

H. Economic, supply chain factors

The technical demands of advanced lithography translate directly into extraordinary costs. An EUV scanner from ASML carries a price tag in the range of \$170–350 million, and with only one company in the world capable of producing them, lead times can exceed eighteen months. The mask set for a new chip design at leading-edge nodes commonly runs to tens of millions of dollars, and a single defective mask layer can invalidate an entire production run. These economics mean that EUV lithography is effectively accessible only to a handful of the world's largest chipmakers, concentrating the leading edge in ways that carry significant geopolitical as well as commercial implications. [1][15]

The mask ecosystem is also constrained: there are only a few qualified suppliers of EUV mask blanks (e.g. Hoya, AGC) and inspection tools, so any bottleneck or export control (e.g. geopolitical restrictions) can disrupt node deployment. The photomask industry notes that modern chips often need 70–100 masks per layer; as nodes shrink, mask complexity grows and inspection/repair costs skyrocket.[1][15]

Analyst HJL Levinson points out that EUV reticle cycles (writing, inspecting) are so long that mask cycle time is becoming a rate-limiter for chip design, not just cost. Finally, the wafer economics show the impact: a 300 mm wafer at leading-edge (3 nm node) can cost on the order of \$20,000–\$22,000 (vs. ~\$3,000 at 28 nm), largely due to lithography.[1][16]

VII. THE FUTURE

Modern lithography continues to evolve, with ongoing research and development focused on overcoming current limitations and pushing the boundaries of miniaturization. Key areas of focus include high Numerical Aperture (NA) Extreme Ultraviolet (EUV) lithography, advanced illumination techniques, and the exploration of new material combinations.[1][5][8][15]

A. High-NA Lithography advance procurement

- The 0.55 NA EUV lithography is actively being developed by major semiconductor manufacturers to pattern features below 30 nm, particularly for DRAM storage-node patterns and critical logic integrated circuits. This technology aims to achieve a half-pitch (HP_{min}) of approximately 8 nm.[1][15]
- ASML began shipping the production version of its 0.55 NA EUV scanner, the EXE:5200B, in 2025, which promises higher throughput and tighter overlay capabilities than its predecessors.[1]
- The new Starlith® 5000, the optical core of ASML's EXE:5000 scanner, saw its first illuminators and Projection Optics Boxes (POBs) completed in 2023. A joint laboratory

between ASML and imec started operating this 0.55 NA EUV scanner in 2024.[1]

B. Higher NA Anamorphic optics

- The innovative optics architecture of the 0.55 NA EUV system, featuring anamorphic imaging, can be extended to achieve a "hyper NA" system with an NA of approximately 0.75. This is projected to target a half-pitch of around 6 nm with high productivity.[1][16]
- This anamorphic imaging system involves increasing de-magnification in the scanning direction while maintaining 4X in the orthogonal direction. This approach addresses mask-shadowing effects and allows for a 50% reduction in the size of the exposure field on the wafer, necessitating field stitching for larger chips.[1]

C. Illumination and mask optimization

- Advanced Illuminators: Significant efforts are being made to develop future illuminators that can substantially reduce the pupil filling ratio. This, combined with advanced resists and mitigation of the mask's 3D effect, is expected to enable a k_1 value of approximately 0.3, leading to an 8 nm half-pitch with 0.55 NA in production.[1] Modern illuminators offer a high degree of freedom in shaping the two-dimensional light distribution without light loss, thereby maintaining tool productivity.[1]
- Source Mask Optimization (SMO): Mask pattern enhancements, such as Optical Proximity Correction (OPC), and sub-resolution assisting features are being developed with the help of machine learning. Both off-axis illumination (OAI) and OPC are optimized together in a process called Source-Mask Optimization (SMO), which further reduces the k_1 factor.[1][17]

D. New Material Science combinations and Resist Technology

- For future lithographic optics using wavelengths shorter than 13.5 nm, new material combinations are required for the multilayer structures needed to build optical systems. These materials are not yet known, indicating a potential fundamental barrier for further wavelength reduction.[1][16]
- The development of high-resolution and high-sensitivity resists is crucial. Metal-oxide resists with smaller blur (σ) are being developed as a step towards achieving finer features, especially as pitch decreases.[1][14]

E. Low resolution, large field exposure tools application in AI processors.

As GPUs for AI applications become larger and are connected with DRAM chips via interposers, there is a growing need for exposure tools that can handle larger areas than the standard $26 \times 33 \text{ mm}^2$ exposure field of current scanners. Low-resolution but large-field exposure tools, such



as $52 \times 68 \text{ mm}^2$ i-line tools, are being considered for these applications.[1][15]

While significant progress has been made, the industry faces ongoing challenges in optical physics, material science, and system complexity, driving continuous innovation to meet the demands of advanced semiconductor manufacturing.[1][5]

REFERENCES

- [1] A. Yen, W. Kaiser and A. Suzuki (2026), "Development of Photolithography for Semiconductor Manufacturing – a Review," in *IEEE Electron Devices Reviews*, doi: 10.1109/EDR.2025.3650514
- [2] H. Aida, "Autobiography of Japanese Semiconductor Development (in Japanese)", NHK Press, 1991.
- [3] J. Andrus, "Fabrication of Semiconductor Devices," USP 3,122,817, filed on August 15, 1957 and granted on March 3 1964
- [4] Z. Koana and Z. Wakimoto, *Jpn. J. Appl. Phys.*, 3, pp. 113–120, 1964.
- [5] J. H. Bruning, "Optical lithography 40 years and holding," *Proc. SPIE*, 6520, pp. 652004-1– 652004-13, 2007
- [6] Moreau, Wayne M. (2012), *Semiconductor Lithography: Principles, Practices, and Materials*. Springer US.
- [7] L. Geppert, "Semiconductor lithography for the next millennium," in *IEEE Spectrum*, vol. 33, no. 4, pp. 33-38, April 1996, doi: 10.1109/6.486632
- [8] Jaizul-Aziz, Duta. "Principles of Photolithography in Semiconductor Manufacturing," n.d. 2020
- [9] Chandu, Hiranmaye Sarpana. "A survey of semiconductor wafer fabrication technologies: Advances and future trends." *Int. J. Res. Anal. Rev* 10, no. 04 (2023): 344-349.
- [10] Tu, Yuan. *Insights Into Semiconductor Technologies: Transistor, Interconnect, Packaging, Lithography and Memory*. N.p., Springer Nature Switzerland, 2026.
- [11] Kapoor, Rahul, and Ron Adner. "Technology interdependence and the evolution of semiconductor lithography." *Solid State Technology*, November 2007, 51+. Gale Academic OneFile.
- [12] Liu, Changhao. "Research on Advanced Photolithography Technology in Semiconductor Processes." In *2025 International Conference on Electronics, Electrical and Grid Technology (ICEEGT 2025)*, pp. 168-175. Atlantis Press, 2026
- [13] Chae, Heejoon, Hyunje Park, and Dae Joon Kang. 2026. "The Evolution of Lithography: From Resolution Scaling to Manufacturing Constraints" *Micromachines* 17, no. 2: 261. <https://doi.org/10.3390/mi17020261>
- [14] Cyriac, Jincemon, Jisty Ann Jiji, Joice K. Joseph, Dan Mathew Sabu, Ashtami R. Nair, Tina Sebastian, V. N. Sheeja, and Tom Cherian. "A Pervasive Study on the Fabrication Techniques of Nanotechnology." In *A Journey from Nanotechnology to Picotechnology: Atomic and Molecular Scale Engineering*, pp. 45-90. Cham: Springer Nature Switzerland, 2026
- [15] Jain, Manish, and Nilesh Jain. "Recent Developments in Semiconductor Wafer Fabrication: Materials, Processes, and Innovations." *Journal of Global Research in Electronics and Communication* 2, no. 2 (2025).
- [16] Chris A. Mack, "The Future of Semiconductor Lithography? Look to Flash," *J. Micro/Nanolith. MEMS MOEMS* 12(3) 030101 (8 August 2013) <https://doi.org/10.1117/1.JMM.12.3.0301>
- [17] *Handbook of VLSI Microlithography: Principles, Technology, and Applications*. Ukraine, Noyes Publications, 2001
- [18] C. A. Mack, *Fundamental Principles of Optical Lithography: The Science of Microfabrication*, John Wiley & Sons, Chichester, UK, 2007.
- [19] H. J. Levinson, *Principles of Lithography*, 3rd ed., SPIE Press, Bellingham, WA, 2011.
- [20] B. J. Lin, "The ending of optical lithography and the prospects of its successors," *Microelectronic Engineering*, vol. 83, no. 4–9, pp. 604–613, 2006.