Semiconductor Technology from A to Z

Photolithography

Exposure and resist coating

Overview

In semiconductor manufacturing, structures are created on silicon wafers using lithographic processes. First, a radiation-sensitive film, usually a photoresist layer, is applied to the wafer, patterned, and then transferred into the underlying layer using etch processes.

Photolithography comprises the following process steps:

- applying an adhesion promoter and removing water from the wafer
- coating the wafers with resist
- stabilizing the resist layer
- exposure
- developing the resist
- curing the resist
- inspection

In some processes, such as ion implantation, the photoresist serves as a protective layer to exclude certain areas of the wafer from implantation. In this case, no transfer of the resist mask by an etch process takes place.

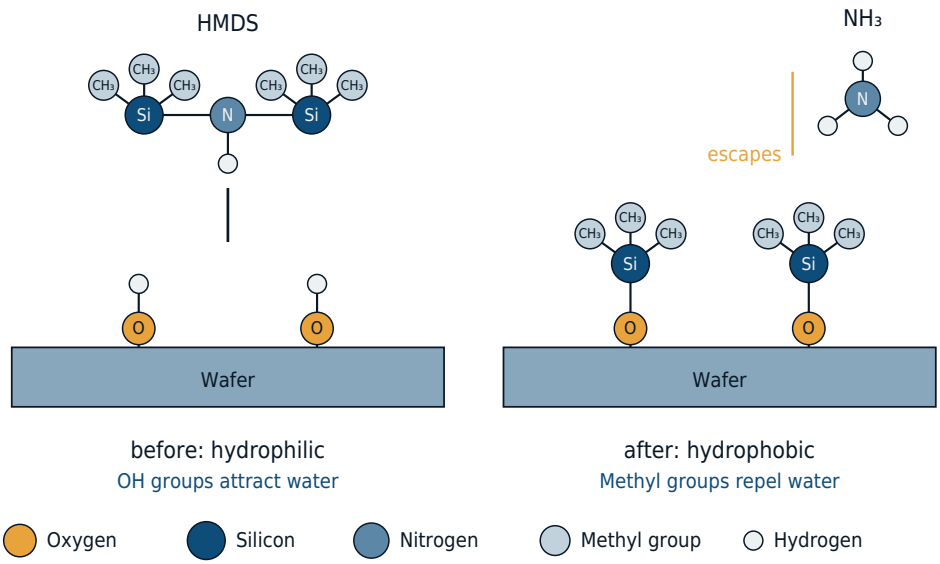
Applying an adhesion promoter

First, the wafers are cleaned and baked out (pre-bake) to remove adhering particles and any adsorbed water. The surface of the wafers is water-attracting (hydrophilic) and must be made hydrophobic – that is, water-repellent and thus resist-attracting – before the resist layer is applied. For this purpose, an adhesion promoter, usually hexamethyldisilazane (HMDS), is applied to the wafers. The wafers are exposed to the vapor of this liquid so that the wafer surfaces are wetted with it.

Due to the moisture in the ambient air, hydrogen H or hydroxide ions OH^- are always present at the wafer surface, even after baking. The HMDS splits into trimethylsilyl groups $Si(CH_3)_3$ and removes the hydrogen, forming ammonia NH_3.

Attachment of HMDS

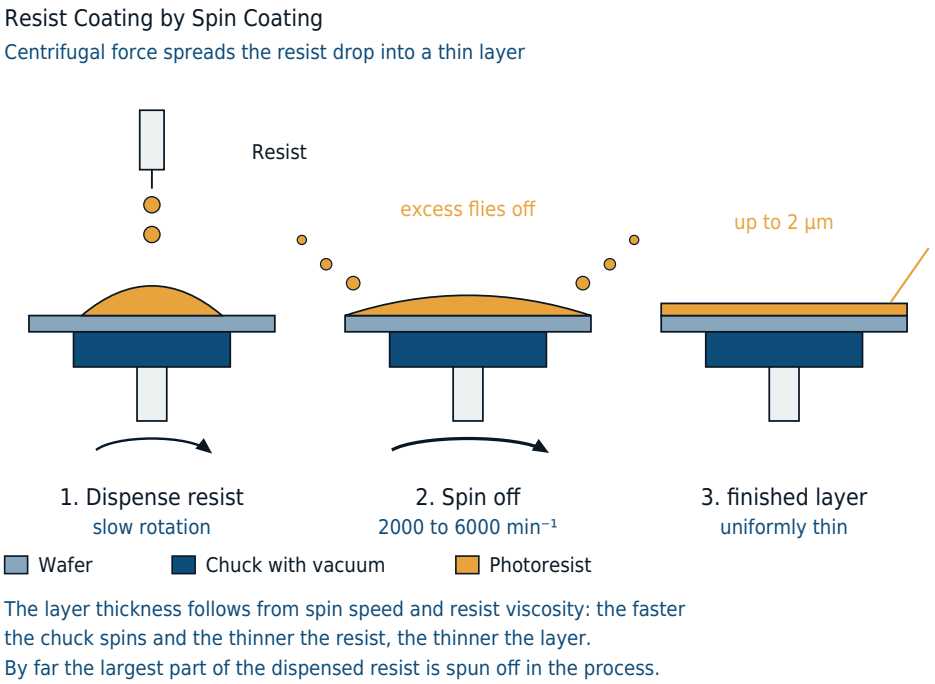
The water-attracting surface becomes water-repelling



The HMDS splits into two trimethylsilyl groups. Each bonds to an oxygen atom on the surface and takes its hydrogen along; together with the nitrogen, this forms ammonia.

Coating

The wafers are coated with resist by spin coating on a rotating platen with vacuum suction (chuck). At low rotational speed, resist is dispensed onto the center of the wafer and then, at 2000–6000 revolutions per minute, spread out into a homogeneous resist layer by centrifugal force. Depending on the subsequent process, its thickness is up to 2 µm. The thickness depends on the rotational speed and the viscosity of the resist.



So that the resist can flow evenly across the wafer, it contains water and solvents that soften it. To stabilize the resist layer, the wafer is subsequently baked out at about 100 °C (post-bake or soft-bake). Water and solvents are partially evaporated; a residual moisture content must be retained for exposure.

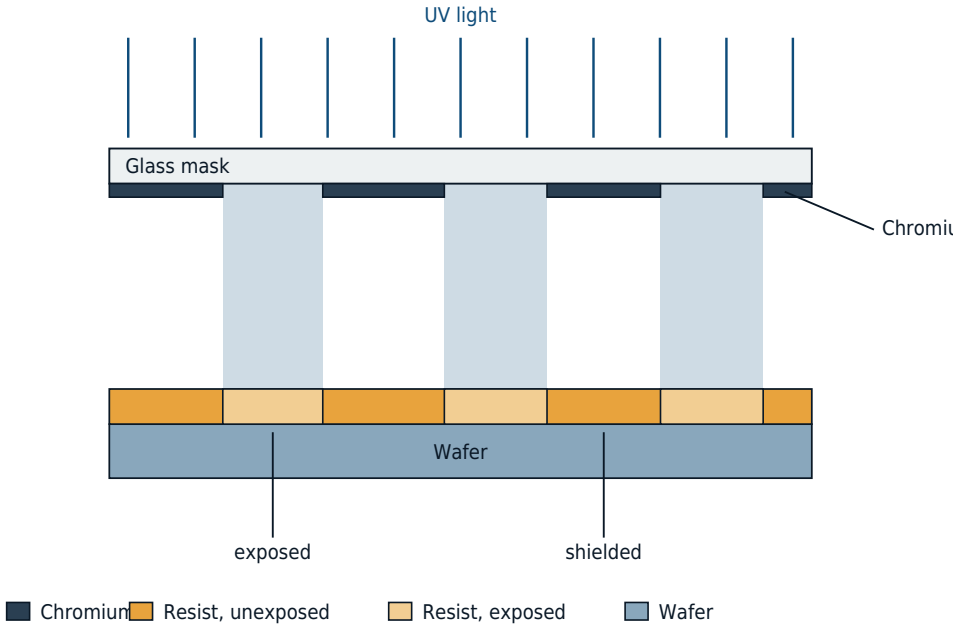
Nowadays, due to the small feature sizes, additional auxiliary layers are often used as well, which are applied to the wafer either before or after the photoresist. Among other things, these layers can serve as an anti-reflective coating or for additional planarization.

Exposure

An exposure tool contains a glass mask that is partially coated with chrome, whereby some areas of the resist-coated wafer are exposed while others are shadowed.

Exposure Through a Mask

Where chromium sits on the mask, the resist stays unexposed



The mask carries the pattern as a chromium layer on glass. During development, depending on the resist type, either the exposed or the unexposed part is removed – leaving behind the patterned resist layer.

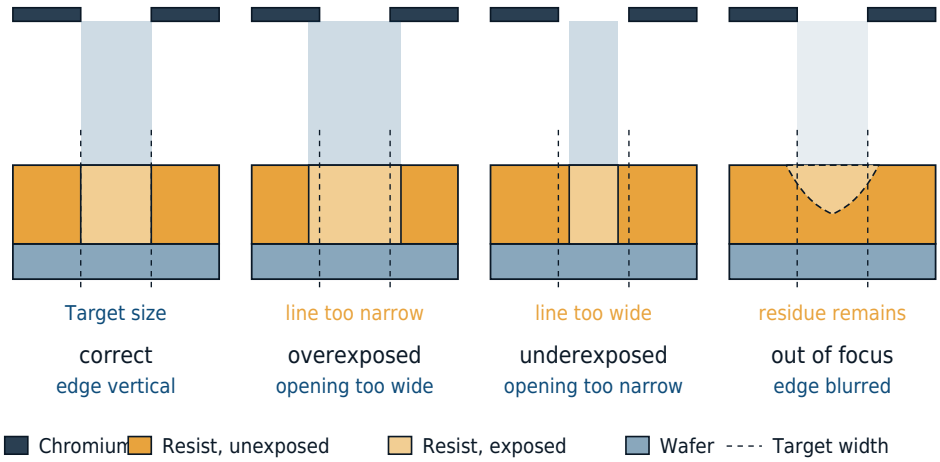
Depending on the type of resist, exposed portions become soluble or insoluble. Using a developer solution, the soluble areas are removed, leaving a patterned resist layer behind. In classic i-line positive resist, a nitrogen molecule N_2 is split off by the high-energy UV light. What remains is a keto-carbene, which, for energetic reasons, converts to ketene (CH_2=C=O). By absorbing moisture from the ambient air, the ketene forms a carboxylic acid. At 248 nm and below, this mechanism is replaced by chemically amplified resists, which release a photoacid (see photoresist).

The exposure time is very important so that the structures obtain the exact size. The longer the wafer is exposed to the radiation of the exposure tool, the larger the exposed areas become. A precise determination of the correct exposure duration to achieve the specified structure widths using one or more test wafers (precursors) is necessary, since the resist can behave differently depending on the ambient temperature.

With overexposure, resist lines – and thus the structures beneath them – become too small, while contact holes become too large. With too short an exposure time, the contact holes are not opened, interconnects are too wide, and may end up in contact with one another. In addition, poor focusing leads to unexposed areas, so that contact holes are not exposed during development and connections between interconnects remain, which can lead to short circuits.

Exposure Profiles

Exposure time and focus determine the width of the feature



The longer the exposure, the wider the exposed area: resist lines get narrower, contact holes larger. With too short an exposure or poor focus, resist remains at the bottom and the contact hole fails to open.

Depending on the subsequent process, the resist dimension – that is, the width of the resist lines or the size of the contact holes – has to be adjusted. In isotropic etch processes (etching occurs both vertically and horizontally), the mask is not transferred 1:1 into the layer to be patterned.

Exposure methods used

Depending on requirements, high-energy radiation such as UV light, x-rays, electron beams, and ion beams are used for exposure. The rule is: the shorter the wavelength of the radiation, the smaller the achievable feature widths. This relationship is described by the Rayleigh criterion:

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

Here, λ is the exposure wavelength, NA is the numerical aperture of the lens, and k_1 is a process factor that combines the illumination method, mask technology, and photoresist. The depth of focus decreases with the square of the aperture:

$$DOF=k_2\cdot\frac{\lambda}{NA^2}$$

Every improvement in resolution therefore comes at the cost of a smaller focus window – which is why planar wafer surfaces (see CMP) are a prerequisite for fine structures.

Overview of the wavelengths

Wavelength	Source	Use
436 nm (g-line) 365 nm (i-line)	Mercury vapor lamp	historical; the i-line is still used for non-critical layers, power semiconductors, and MEMS
248 nm	Krypton fluoride excimer laser	coarser layers, analog and power technology
193 nm	Argon fluoride excimer laser	workhorse of manufacturing, both dry and as immersion exposure
13.5 nm (EUV)	Tin plasma, laser-ignited	critical layers of all leading logic and memory processes

Originally, a fluorine laser at 157 nm was also planned as an intermediate step. It turned out to be impractical, since the calcium fluoride lenses required for it absorb moisture from the air; in addition, immersion exposure with water would not be possible, because water absorbs light at this wavelength.

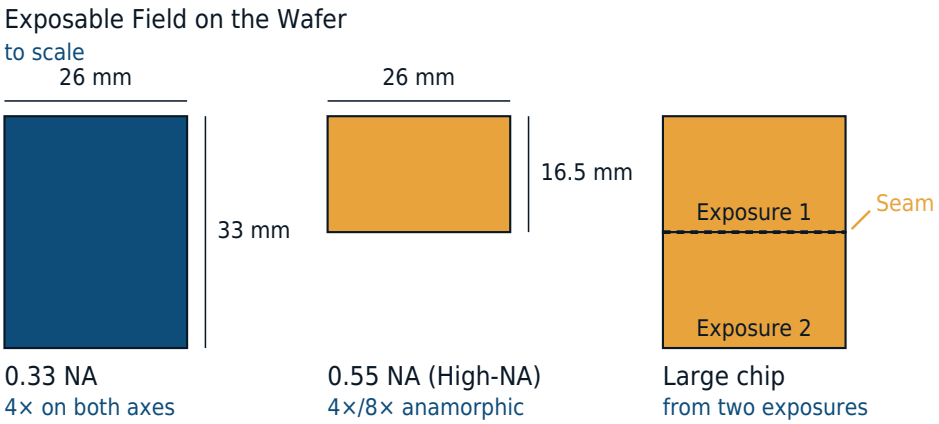
Immersion lithography

Instead of further shortening the wavelength, the numerical aperture can be increased. For this purpose, the gap between the last lens and the wafer is filled with highly pure water, whose refractive index at 193 nm is about 1.44. This allows apertures of up to 1.35, compared with about 0.93 in air. A single immersion exposure resolves structures down to about 38 nm half-pitch – this is the physical limit of this technique. Finer structures at 193 nm can only be produced by splitting a pattern across multiple exposures (see multipatterning).

EUV lithography

The jump to extreme ultraviolet at 13.5 nm has been in volume production since 2019. The radiation is generated by vaporizing tin droplets into a plasma using a high-power laser; since EUV is absorbed by air and by any glass, the entire tool operates under vacuum and images exclusively via mirrors. Today's generation of tools, with an aperture of 0.33, achieves about 13 nm half-pitch in a single exposure, thereby replacing two or more immersion steps on many layers.

Exposable field at 0.33 NA and 0.55 NA



Since 2024, the High-NA generation with an aperture of 0.55 has been added, resolving about 8 nm half-pitch. The price for this is an anamorphic optical system: the image is reduced by a factor of 4 in one direction and by a factor of 8 in the other, halving the exposable field to 26 × 16.5 mm. Large chips therefore have to be assembled from two adjoining exposures. In July 2026, High-NA-exposed layers were qualified in a production product for the first time; widespread adoption is expected for 2027 and 2028.

Other methods

X-ray and ion beam lithography have never reached production and remain confined to research. Electron beam writers, on the other hand, are indispensable – not for exposing the wafers, but for producing the photomasks.

Multipatterning

A single 193 nm immersion exposure resolves structures down to about 38 nm half-pitch. When devices dropped below this limit before EUV lithography became available, only one path remained: the pattern is split across several exposure and etch steps, each of which individually stays within the resolution limit. Together they produce a finer pattern than a single exposure could create.

Multiple exposure with two masks (LELE)

In the litho-etch-litho-etch process, the target pattern is decomposed into two sub-patterns, each with twice the feature spacing. The first is exposed and etched, followed by a second complete pass whose structures fall precisely into the gaps of the first.

The disadvantage lies in the alignment: the position of the second layer relative to the first directly affects the resulting feature width. Every alignment error

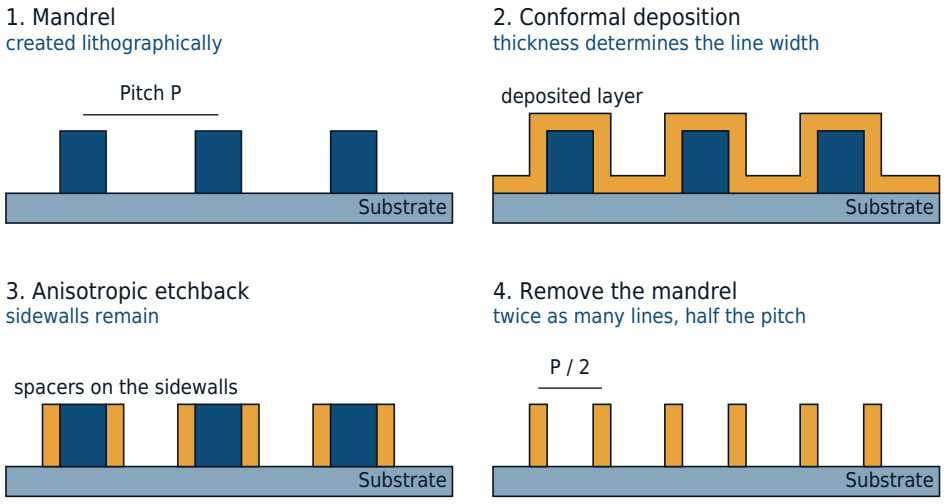
shows up as alternating too-wide and too-narrow spacings – a systematic error that makes the devices non-uniform. With three or four passes (LELELE and beyond), this problem grows, as does the number of process steps and thus the cost.

Self-aligned double patterning (SADP)

The self-aligned process avoids the alignment problem by depositing the fine structures rather than exposing them:

- An auxiliary pattern (mandrel) is created lithographically at normal resolution.
- A thin layer is deposited conformally on top of it, usually by atomic layer deposition. Its thickness determines the resulting feature widths – and it can be controlled far more precisely than an exposure step.
- An anisotropic etch step removes the layer from the horizontal surfaces, leaving lines standing at the sidewalls of the auxiliary pattern (spacers).
- The auxiliary pattern is removed selectively. What remains is twice as many lines as the auxiliary pattern had features.

Sequence of self-aligned double patterning



Since the lines are positioned by the mandrel itself, there is no alignment error between them. Repeating the process yields four structures per original one (SAQP). The price: closed loops always form around the mandrel, and the resulting pattern is inevitably regular. Both issues have to be corrected with additional masks that deliberately cut through unwanted sections (cut mask) or define line ends. The process is therefore suited to dense, regular patterns – interconnects, gate arrays, memory cell arrays – but not to arbitrary layouts.

Significance today

Multipatterning has not remained a stopgap solution. It continues to be used even with EUV, just in a different place: where EUV can complete the critical layers in a single exposure, the split is no longer needed; where structures fall below EUV's resolution, it becomes necessary again. At the same time, circuit design has changed permanently. Layouts today follow strict rules with uniform pitches and uniform preferred directions, because only such patterns can be decomposed at all – lithography dictates its constraints to circuit design.