Semiconductor Technology from A to Z

Wet Chemistry

Etch processes

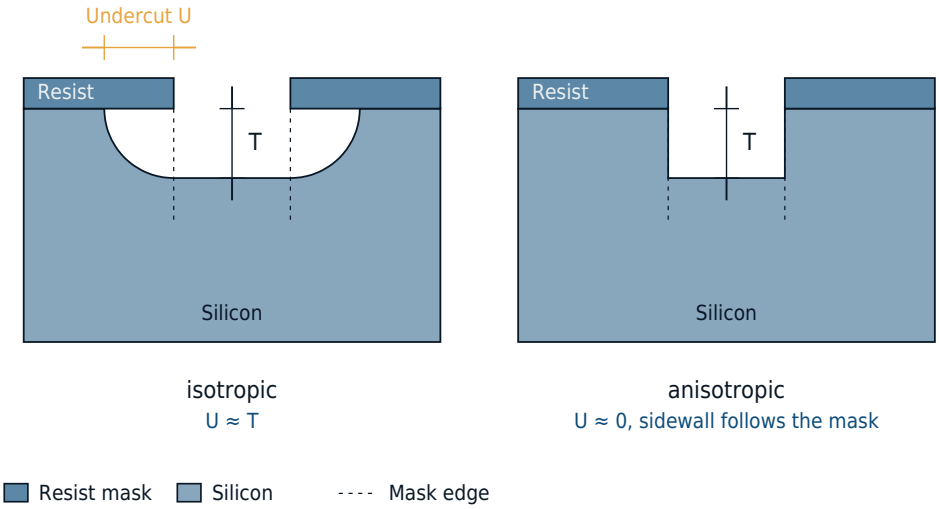
Overview

In semiconductor manufacturing, a wide variety of layers have to be etched. This may be either to remove a layer over the full wafer area, or to transfer a patterned photoresist film into an underlying layer. Etch technology can be divided into wet chemical etching and dry etching. In addition, a distinction is made between isotropic and anisotropic processes, as well as between the chemical and physical character of the etch process.

In an isotropic etch process, etching occurs in all directions. This means layers are removed not only in thickness but also in their lateral extent. In anisotropic etching, the layer is removed in only one direction. Depending on the process, either an isotropic or an anisotropic etch process may be desired.

Isotropic and Anisotropic Etching

With ideal isotropy, the undercut is about as large as the etch depth



Wet chemical etching mostly attacks without direction: the solution reaches the area under the mask just as it reaches the trench bottom – the structure becomes wider than the mask opening as a result. Anisotropic processes like RIE etch directionally instead, transferring the mask geometry almost unchanged.

An important parameter of etch processes is the selectivity. This indicates the removal ratio between two layers. If the selectivity is 2:1, one layer is removed by the etch process twice as fast as the other.

The wet chemistry described below, however, is not used solely for etching layers, but is also employed in other processes:

- Wet etching: for the full-area removal of doped and undoped oxide layers
- Wafer cleaning
- Resist removal
- Backside processing: removal of layers that formed on the wafer backside during furnace processes
- Polymer removal: removal of byproducts that form during plasma etching and deposit on the wafers
- Edge bevel removal: targeted removal of layers at the very edge of the wafer, where they would later flake off and generate particles

Because of its generally isotropic etch profile, wet etching is hardly ever used for patterning. An exception is micromechanics. Owing to the lattice structure of single-crystal silicon, wet chemical etch solutions can produce well-defined structures, such as those with sidewall angles of 90° or 54.74°.

However, the loss of its patterning role does not mean that wet chemistry has become less important – quite the opposite. The smaller the structures have become, the more process steps have been added, and the more sensitive the device has become to contamination. A modern process flow contains more wet chemical steps than ever before; it is just that these are now predominantly cleaning, removal, and pre-treatment steps rather than patterning steps. At the same time, the equipment technology has shifted from batch processing to single-wafer processing.

Wet etching

Principle

The principle of wet etching is the conversion of the solid material of the layer into liquid compounds with the help of a chemical solution. The selectivity is very high, since the chemicals used can be precisely tailored to the layers present; for most solutions, it exceeds 100:1.

The process proceeds in three sub-steps, which must occur one after another:

1. The reactive species travel from the solution to the surface of the layer to be etched.
2. The chemical reaction takes place at the surface, forming a soluble compound.
3. The reaction product detaches from the surface and is transported away into the solution.

Whichever of these steps is the slowest determines the behavior of the entire

process. If the actual reaction is the bottleneck, the process is called reaction-limited: the etch rate then depends strongly on temperature and increases roughly exponentially with it, but the process etches uniformly across the entire wafer. If, on the other hand, the transport to and from the surface is the bottleneck, the process is diffusion-limited. The temperature dependence is then weaker, but any non-uniformity in the flow directly affects the removal rate – the edges of the wafer are etched faster than the center.

For practical purposes, this means: reaction-limited processes are controlled through very precise temperature control, while diffusion-limited processes are controlled through movement of the solution, i.e. through circulation, rotation of the wafer, or ultrasound. In narrow trenches and contact holes, every process becomes diffusion-limited, because the solution there is hardly exchanged at all. This is one of the reasons why wet etching is not suitable for fine structures.

Requirements

The following requirements have to be met by the chemical solutions:

- the masking layer must not be attacked
- the selectivity must be high, both with respect to the masking and to the underlying layer
- the etch process must be able to be stopped by dilution with water
- no gaseous reaction products may form, since bubbles can locally shadow certain areas
- constant etch rate over a long period
- the reaction products must dissolve directly, so that no particles contaminate the etch solution
- the solution must fully wet the surface; otherwise water-repellent areas and narrow trenches will not be reached
- no introduction of alkali or heavy metal ions, which would enter the oxide as mobile charges and render devices unusable
- good environmental compatibility and easy disposal

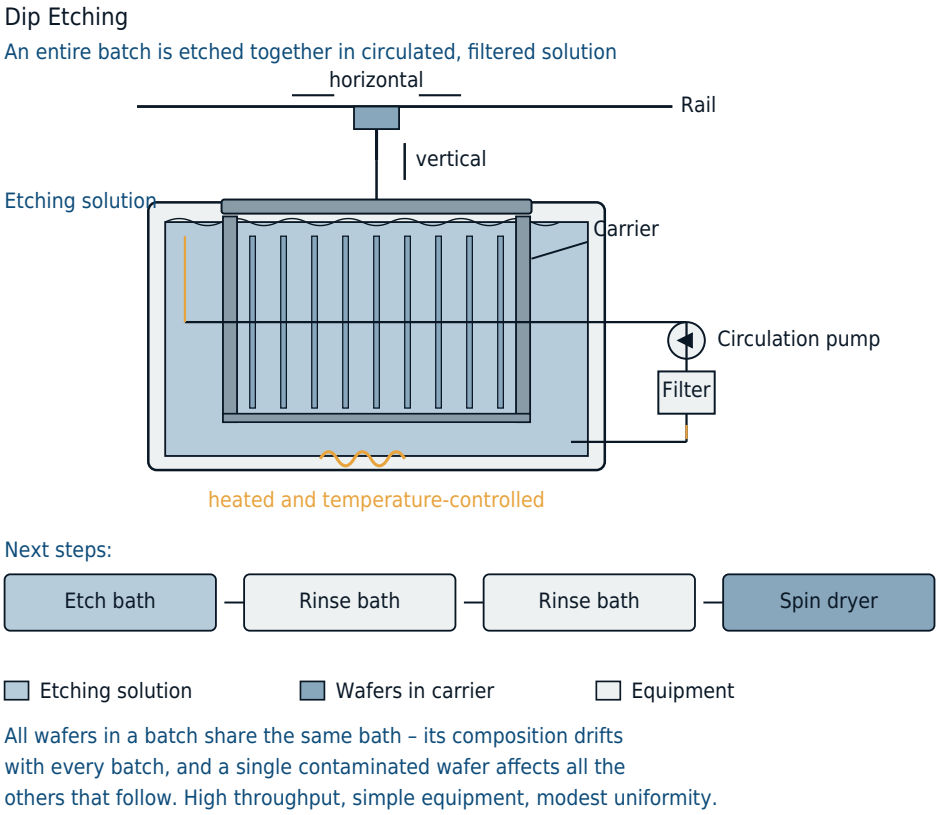
The last requirement is the most difficult one in practice. The most effective etch solutions are also the most hazardous, and some of the chemicals used cannot be replaced by more benign alternatives. The effort required for exhaust air, wastewater, and occupational safety therefore accounts for a considerable share of the operating costs of a wet chemical facility.

Dip etching

In the dip etch process, an entire batch of wafers is processed at once in a bath containing the etch solution. Filters and circulation pumps keep particles away

from the wafers. Since the concentration of the etch solution decreases as more wafers are processed, it often has to be renewed.

The etch rate, i.e. the material removal per unit time, has to be precisely known in order to achieve reproducible etch results. Precise temperature control of the etch solution is required, since the etch rate of most chemical solutions increases with temperature.



The lifting mechanism can transport the batch both horizontally and vertically. After the wafers have been etched, the etching process is stopped by rinsing in several dip baths; the wafers are then dried in a spin dryer.

The advantages of dip etching are the high wafer throughput and the relatively simple construction of the etch equipment. However, the uniformity of layer removal is low.

A fundamental problem of the process is the bath itself: every wafer in a batch sees the same solution, which changes over time. Its composition shifts with every batch processed, dissolved material accumulates, and a single contaminated wafer can affect the entire batch as well as all subsequent ones. The concentration is therefore continuously monitored and replenished, and the service life of a bath is limited.

For critical steps, dip etching has therefore largely been replaced by single-

wafer processing. It has, however, persisted wherever long etch times at high temperatures are required and the uniformity requirement is moderate – the classic example being the removal of silicon nitride in hot phosphoric acid, which, depending on layer thickness, takes an hour or longer and would be uneconomical to perform wafer by wafer.

Spray etching

Spray etching is comparable to spray development in photolithography. Owing to the rotation of the wafer on the chuck, combined with a continuous supply of fresh etch solution, the uniformity is very good. Bubbles cannot form here because of the fast rotation; however, the disadvantage here too is that each wafer has to be processed individually.

As an alternative to sequential processing, spray etching of several batches at once is also possible in a drum, in which the wafers rapidly rotate around a central spray fixture. Immediately afterward, the wafers are dried under rotation in a hot nitrogen atmosphere.

What started out as a disadvantage is today the norm: single-wafer processing has established itself for nearly all critical wet chemical steps. The reason lies in control. Every wafer sees only fresh chemistry, the process time can be set to the second, and any error always affects only a single wafer rather than an entire batch.

A modern single-wafer tool has several pivoting nozzles that apply different media to the rotating wafer one after another – etch solution, rinse water, drying agent – without the wafer ever leaving the process chamber. Consumption per wafer is orders of magnitude lower than that of a bath, which considerably reduces the costs for chemicals and disposal.

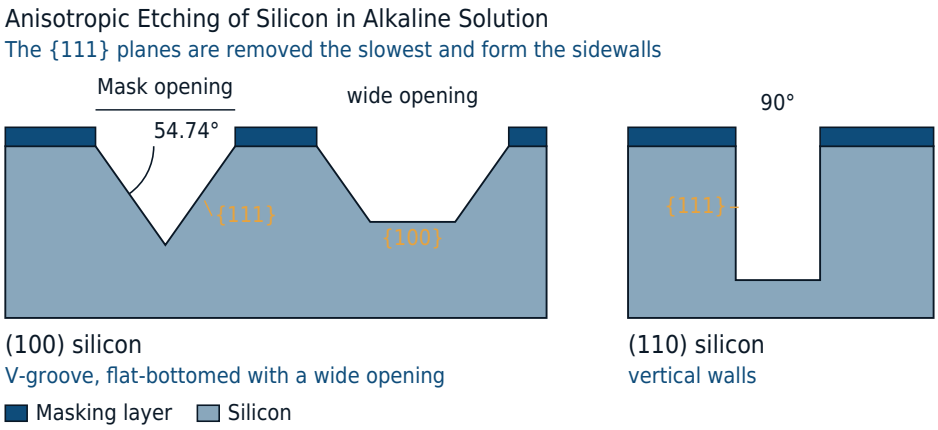
This also makes possible something that is fundamentally impossible in a bath: spatially confined treatment of a wafer. A nozzle at the edge can be used to strip only the outermost bevel, while a supply from below can treat only the backside, with the front side protected by a stream of nitrogen.

Anisotropic etching of silicon

Although the particles in an etch solution can etch the layer on the wafer in every direction, there are also processes that allow an almost anisotropic etch profile to be obtained during wet etching. These make use of the different etch rates on the various crystal planes.

The cause lies in the number of bonds with which a silicon atom is held in the respective plane of the crystal. In the {111} plane, the density of bonds is highest, and it is correspondingly removed most slowly. The {100} and {110} planes are etched considerably faster – the ratio in potassium hydroxide reaches more than 100:1, depending on concentration and temperature. As a result, the etch proceeds in depth until it reaches {111} planes; these remain standing and form the sidewalls of the structure. The shape of the result is therefore not determined by the etch time, but by the crystal orientation of the wafer.

Etch profiles depending on crystal orientation



On (100)-oriented wafers, the {111} planes intersect the surface at 54.74°, producing V-shaped trenches and pits with a pyramidal cross-section. If the mask opening is wide or the etch time short, a flat bottom remains and the structure is trapezoidal. On (110)-oriented wafers, the {111} planes stand perpendicular to the surface, so that trenches with vertical walls and a very high depth-to-width ratio can be etched.

Etch solutions

Etching is carried out with potassium, sodium, or lithium hydroxide (KOH, NaOH, LiOH), with tetramethylammonium hydroxide (TMAH), or with an EDP solution (a mixture of water, pyrazine, catechol, and ethylenediamine). In every case, the reaction is driven by the OH^- group (hydroxide group) of these substances:

$$Si + 2OH^- + 2H_2O \rightarrow SiO_2(OH)_2^{2-} + 2H_2$$

The choice of solution is less a question of etch rate than one of compatibility with the rest of the manufacturing process. Potassium hydroxide etches with the strongest anisotropy but introduces potassium ions: alkali ions are mobile in the gate oxide, shift the threshold voltage of transistors, and render a device

unusable. In a production line where integrated circuits are also fabricated, potassium hydroxide is therefore ruled out. TMAH contains no metal ions and is compatible with CMOS manufacturing; its anisotropy is lower, but it barely attacks silicon dioxide. EDP is largely avoided today because of its toxicity. TMAH, as a processing solution, is by no means harmless either, being highly toxic through skin contact – in a much more dilute form, namely as a 2.38 % solution, the same substance is the standard developer used in photolithography.

Limiting the etch depth

Since the etch stops on its own only at {111} planes, the depth has to be defined in some other way if a membrane of a defined thickness is to be produced. Two approaches are common: a layer very heavily doped with boron, which is barely attacked by the lye any longer, or electrochemical control, in which a p-n junction is placed under voltage and the etch comes to a halt at it.

Applications

Anisotropic wet etching plays no role in patterning integrated circuits. Outside of that field, however, it is a standard process:

- Micromechanics: membranes for pressure sensors, proof masses for accelerometers, nozzles for inkjet print heads
- Solar cells: a brief treatment in dilute lye produces a dense array of small pyramids on (100) silicon. This texture causes incident light to strike the surface multiple times and considerably lowers the reflectance.
- Optics and packaging: V-grooves for holding and aligning optical fibers, alignment structures for assembly

Etch solutions for isotropic etching

Different etch solutions are available for the respective materials. In the semiconductor industry, oxide layers are etched with hydrofluoric acid HF:

$SiO_2 + 6HF \rightarrow H_2SiF_6 + 2H_2O$

The solution is buffered with NH_4F to keep the HF concentration constant. In a mixture of a 40 % NH_4F solution and 49 % hydrofluoric acid (ratio 10:1), the etch rate on thermal oxide is 50 nm/min. TEOS oxides are etched considerably faster, at about 150 nm/min, and PECVD oxides at about 350 nm/min. The selectivity relative to crystalline silicon, silicon nitride, and polysilicon is much greater than 100:1.

The fact that the etch rate depends so strongly on the origin of the oxide is not a side effect but a useful tool: conversely, the etch rate in buffered hydrofluoric

acid can be used to assess the density – and thus the quality – of a deposited oxide layer.

A brief treatment in highly diluted hydrofluoric acid, the HF dip, removes the native oxide that forms on any silicon surface in air within a short time. It therefore precedes many processes that require a clean contact to the silicon, such as epitaxy, gate oxidation, or metal deposition. Afterward, the surface is saturated with hydrogen and thus water-repellent – water visibly beads off it. Since new oxide forms again immediately, the subsequent step must follow promptly.

Silicon nitride is etched with hot phosphoric acid H_3PO_4 at about 160 °C. Here, however, the selectivity to SiO_2 is very low, at 10:1. For polysilicon, the selectivity to nitride is essentially determined by the concentration of the phosphoric acid. To increase the selectivity relative to oxide, dissolved silicon is added to the solution so that it becomes saturated with respect to silicon dioxide and thus barely attacks it any longer.

Crystalline and polycrystalline silicon are first oxidized with nitric acid (HNO_3), after which the silicon dioxide is etched with hydrofluoric acid:

1. $3Si + 4HNO_3 \rightarrow 3SiO_2 + 4NO + 2H_2O$
2. $SiO_2 + 6HF \rightarrow H_2SiF_6 + 2H_2O$

Aluminum is etched at about 50 °C with a mixture of phosphoric, nitric, and acetic acid. The nitric acid oxidizes the aluminum, the phosphoric acid dissolves the resulting oxide, and the acetic acid lowers the surface tension so that the solution fully wets the structure and the hydrogen generated does not remain adhered as bubbles. Titanium is etched with a solution of ammonia water NH_4OH, hydrogen peroxide H_2O_2, and water (ratio 1:3:5). Since this solution also attacks silicon once the peroxide is consumed, its bath lifetime is short. The same mixture, in a more dilute form, serves as a cleaning solution for wafer cleaning.

Handling hydrofluoric acid

Among process chemicals, hydrofluoric acid occupies a special position and deserves an explicit warning. It is only a weak acid and initially causes little pain on the skin, but penetrates deep into the tissue. There, the fluoride binds the body's calcium, which can lead to chemical burns reaching down to the bone and to disturbances of the heart rhythm. Symptoms often appear only hours after contact, by which time the damage has already occurred. Handling it therefore requires appropriate protective equipment, a readily available calcium gluconate gel, and proper training – in manufacturing, hydrofluoric acid is handled exclusively in closed systems.

Assessment

In general, wet chemical etching is well suited for removing entire layers on a wafer. The selectivity between the layer to be etched and the one beneath it is usually very good, so there is little risk of removing the wrong layer. In addition, the removal rate per unit time is very high, and in dip etching many wafers can be etched simultaneously. For small structures, however, wet etching cannot be used, since its isotropic etch profile is unsuitable for this purpose. In this case, layers are removed anisotropically using dry etch processes.

For very fine structures, a second reason comes into play that has nothing to do with the etch profile anymore: at the end, the liquid has to be removed from the wafer again, and it is precisely during this step that the exposed structures can be destroyed.

Drying the wafers

At the end of every wet chemical step comes rinsing with ultrapure water, followed by drying. What sounds like a minor step is, for fine structures, the most critical part of the entire process.

Why drying is a problem

As long as liquid remains between two neighboring fins or lines, its surface tension pulls them toward each other. The force grows the more closely spaced and the taller the structures are. Once two lines touch, they adhere to each other through bonds between their surfaces and do not spring back apart. This structural collapse – known as *stiction*, from *static friction* – is the most common cause of failure for slender lines at the end of a wet process. Affected above all are free-standing elements in micromechanics and tall, narrow resist lines in photolithography.

Drying by simple spin-off or evaporation makes the problem worse, because the receding liquid front then runs directly through the structure. All of the following methods therefore aim either to defuse the transition from liquid to dry, or to avoid it altogether.

Methods

Spin drying
: The wafer rotates rapidly, the liquid is flung outward, and a stream of nitrogen finishes the drying. Simple and fast, but unsuitable for fine structures and prone to leaving behind dry spots.

Isopropanol drying
: The rinse water is displaced by isopropanol, whose surface tension is only

about one-third that of water. The capillary forces are correspondingly lower.

Marangoni drying
: The wafer is slowly withdrawn from the water while isopropanol vapor is deliberately supplied at the water surface. The resulting difference in surface tension along the interface pulls the water film away from the wafer, so that it emerges dry from the bath and no liquid has to evaporate on it.

Supercritical carbon dioxide drying
: The liquid is replaced by carbon dioxide, which is then brought into the supercritical state through pressure and temperature. In this state, there is no longer an interface between liquid and gas, and thus no surface tension either; the carbon dioxide can be released without leaving any residue. This is the most elaborate but also the gentlest method – it is used wherever all the others reach their limit.

The same reasoning underlies the development of dry alternatives to individual wet steps: wherever a layer can be removed with a gas instead of a liquid, the problem is avoided from the outset.

Wafer cleaning

Cleanroom

Semiconductor manufacturing takes place in cleanrooms in order to protect the highly complex circuits of the semiconductor devices from contamination that could affect their functionality. Cleanrooms are classified according to the size of the particles and their number per cubic foot (= cbf; 1 foot = 30.48 cm) or, according to DIN/ISO, per 1 m^3:

Permitted number of particles per 1 m^3

Class ↓	≥ 0.1 µm	≥ 0.2 µm	≥ 0.3 µm	≥ 0.5 µm	≥ 1.0 µm	≥ 5.0 µm
ISO 1	10	2				
ISO 2	100	24	10	4		
ISO 3	1,000	237	102	35	8	
ISO 4	10,000	2,370	1,020	352	83	
ISO 5	100,000	23,700	10,200	3,520	832	29
ISO 6	1,000,000	237,000	102,000	35,200	8,320	293
ISO 7	10,000,000	2,370,000	1,020,000	352,000	83,200	2,930

For example, in a class 3 cleanroom there may be a maximum of 1,000 particles with a diameter ≥0.1 µm, 237 particles ≥0.2 µm, 102 particles ≥0.3 µm, 35

particles ≥0.5 µm, and 8 particles ≥1 µm. In an operating room, the cleanroom class is 2 or 3; in a volume of 1 m^3 of city air, there are 400 million particles of size 5 µm.

The air in cleanrooms for microelectronics is cleaned through ultra-fine filters and then introduced through the ceiling. Air is extracted through holes in the floor, so that this laminar flow carries particles downward and away through the floor. To prevent contamination from entering the cleanroom from outside, a slight overpressure is generally maintained there. To protect the wafers as effectively as possible against particles, they are transported from one production tool to the next in transport boxes. In modern cleanrooms, these are so-called FOUPs (Front Opening Unified Pods), which dock directly onto the loading stations of the tools and are opened in such a way that the wafers pass from the FOUP into the tool without being exposed to the cleanroom air.

Since the wafers inside the FOUPs are completely shielded from the ambient cleanroom air, a different cleanroom class can prevail inside the transport boxes than in the cleanroom itself. This makes it possible to operate the large production areas at a relatively high cleanroom class of ISO 5, while a cleanroom class of ISO 1 or ISO 2 prevails inside the FOUPs. This allows enormous cost savings, since only a small volume needs to meet the highest cleanliness class.

Illustration of a class 1 cleanroom

(Click to enlarge, 700 KB · source: unknown)

In the illustration, the production area is only the region between the pink-colored ventilation system beneath the roof and the yellow-colored floor area. Below that is the basement housing the supply equipment (pumps, chemical tanks, etc.). Some cleanrooms are surrounded by a so-called grey room in which the tools themselves are located, so that, separated by a wall, only the control panels and the load ports for bringing wafers into the tools are accessible from within the cleanroom.

Personnel wear special cleanroom suits that do not shed particles. Depending on requirements, the suit covers the entire body. The head is covered either with a simple hairnet or a hood and a face mask, or with a full mask. In addition, there are special low-particle shoes, undergarments, and gloves. At the entrance to the cleanroom there are often air showers that blow particles off the suit once more before entry.

Types of contamination

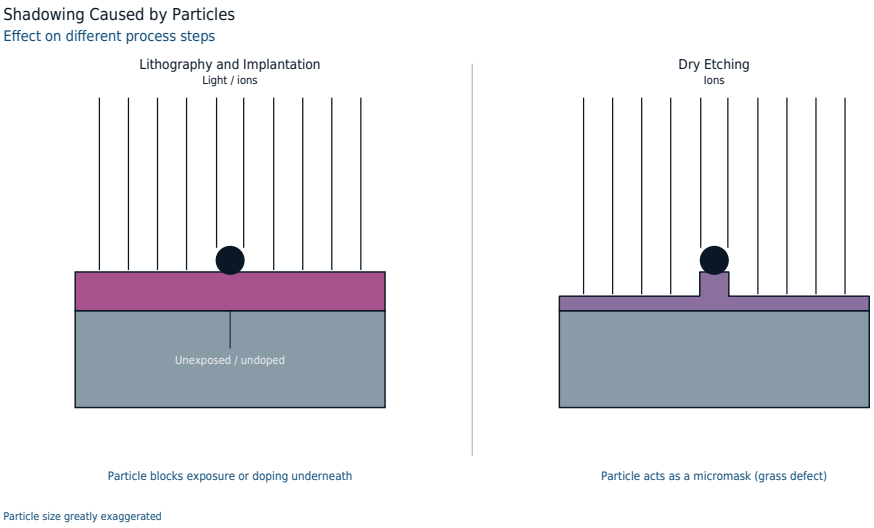
Despite the cleanroom, there are various types of contamination, which are mainly caused by personnel in the production line, the ambient air, chemicals (gases, solutions), and the equipment itself:

- Microscopic contamination: e.g. particles from the ambient air or from gases
- Molecular contamination: e.g. hydrocarbons from oil in the pump systems
- Ionic contamination: e.g. hand sweat
- Atomic contamination: e.g. heavy metals from solutions, abrasion from solid parts

Microscopic contamination

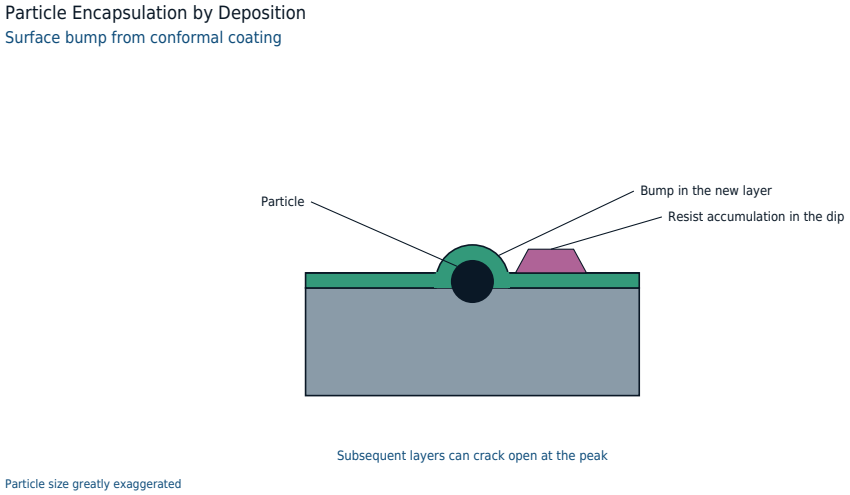
Microscopic contamination consists of particles that attach to the wafer surface. Sources of this contamination include the ambient air, personnel clothing, abrasion from moving parts in process equipment, inadequately filtered liquids such as etch and cleaning solutions or developer, or etch residues remaining after dry etch processes.

Microscopic contamination causes shadowing effects when particles on the wafer surface prevent exposure of the photoresist; in contact exposure, poor resolution results when relatively large particles are located between the mask and the wafer. In addition, they can shield accelerated ions from the wafer surface during implantation processes or dry etching.



Particles can also become enclosed by layers, creating unevenness. Subsequent layers can crack open at these locations, or resist can accumulate there, resulting in areas that are not fully exposed due to the excessive resist

thickness.



Molecular contamination

Molecular contamination results from resist and solvent residues on the wafers or from oil mist deposits originating from vacuum pumps. This contamination can be present on the wafer surface as well as diffuse into the layers. Contamination adhering to the surface partly hinders the adhesion of subsequently deposited layers, considerably so in the case of metallizations (thin interconnects). Diffusion into oxides degrades the electrical robustness of these layers.

Alkaline and metallic contamination

The main source of this contamination is humans, who constantly release salts through the skin as well as through breath. In addition, insufficiently deionized water introduces (alkali) ions such as sodium or potassium onto the wafers. Heavy metals present in etch solutions can also lead to contamination. Radiation-based processes in equipment (implanters, dry etching) can sputter material off the chamber walls, which then deposits onto the wafers.

Ionic contamination affects, for example, the electrical properties of MOS transistors, since their charge alters the threshold voltage – that is, the voltage above which the transistor becomes conductive. Heavy metals such as iron or copper provide free electrons, causing the power consumption of diodes to increase. Metals can also form recombination centers for free charge carriers, so that insufficient free charge carriers remain available in the circuit for proper operation.

Cleaning techniques

The wafers are rinsed with ultrapure water after every wet chemical process step, but wafers are also freed from contamination after other processes. There are various cleaning techniques for removing the different types of contamination. The consumption of ultrapure water in a semiconductor fab is extremely high, amounting to several million liters per year. In ultrapure water, almost no contamination remains; 1–2 ppm (parts per million = number of contaminant particles per million water molecules) is permitted (example, as of 2002). The water is usually purified directly on site in treatment plants.

One method of wafer cleaning is the ultrasonic bath, in which the wafers are placed into a solution of water together with ultrasonic cleaning agents and wetting agents. Ultrasonic excitation loosens particles from the surface, while metals and molecular contaminants are partially bound by the cleaning agent. Particles that do not adhere strongly can also be blown off with nitrogen.

Solvents such as acetone or ethanol are suitable for removing organic contamination such as grease, oil, or skin flakes. However, these can leave carbon residues behind.

Ionic contamination (ions of potassium, sodium, etc.) is removed by rinsing with deionized water. Cleaning with rotating brushes together with a cleaning liquid is also possible. However, on patterned wafers, particles accumulate at edges, and the brushes can damage the wafer surface. Contact holes or other recesses can be rinsed out using high-pressure cleaning at about 50 bar. This method, however, does not remove ionic or metallic contamination.

Cleaning with water and various cleaning agents is often not sufficient on its own. Contaminants are frequently removed using aggressive etch solutions, or the surface is deliberately removed to a minimal extent. Mixtures of hydrogen peroxide with sulfuric acid or ammonia, or Caro's acid/piranha solution (sulfuric acid with hydrogen peroxide), dissolve organic contamination through oxidation at about 90 °C.

A mixture of hydrochloric acid and hydrogen peroxide converts alkali metals into readily soluble chlorides (salts), while heavy metals form complexes and go into solution. Native oxide can be removed with hydrofluoric acid, whereas a deliberately deposited oxide layer using hydrogen peroxide can instead be used to protect the surface.

In addition to rinsing the wafers after every wet chemical process step, wafers undergoing a full cleaning pass through a series of cleaning steps performed one after another. The order of the cleaning processes is important here, since some can partially interfere with each other's cleaning effect. A cleaning sequence might, for example, proceed as follows:

- blowing off particles with nitrogen
- cleaning in an ultrasonic bath
- removing organic contamination with Caro's acid (H_2SO_4–H_2O_2)
- removing finer organic contamination with an NH_4–H_2O_2 solution
- removing metallic contamination with hydrochloric acid and hydrogen peroxide
- drying the wafers in a spin dryer under a hot nitrogen atmosphere

After every cleaning step, the wafers are rinsed with ultrapure water, and, where necessary, native oxide is removed with hydrofluoric acid. Depending on the current wafer surface, the cleaning sequences differ, since the solutions can attack certain layers. As structures continue to shrink, cleaning becomes increasingly difficult as well. Not only are small openings harder to reach, but surface tension and capillary effects can also cause structures to topple over, destroying the circuit.