

Semiconductor Technology from A to Z

Doping Techniques

Ion Implantation	3
<i>Ion implantation</i>	3
<i>Penetration Depth and Annealing</i>	4
<i>Channeling</i>	5
<i>Characteristics and Implanter Types</i>	5
<i>Plasma Doping (PLAD)</i>	6
<i>Halo and Well Implants in CMOS Integration</i>	7
<i>Amorphization and Defect Annealing</i>	8
<i>Dose Metrology</i>	9

Ion Implantation

Ion implantation

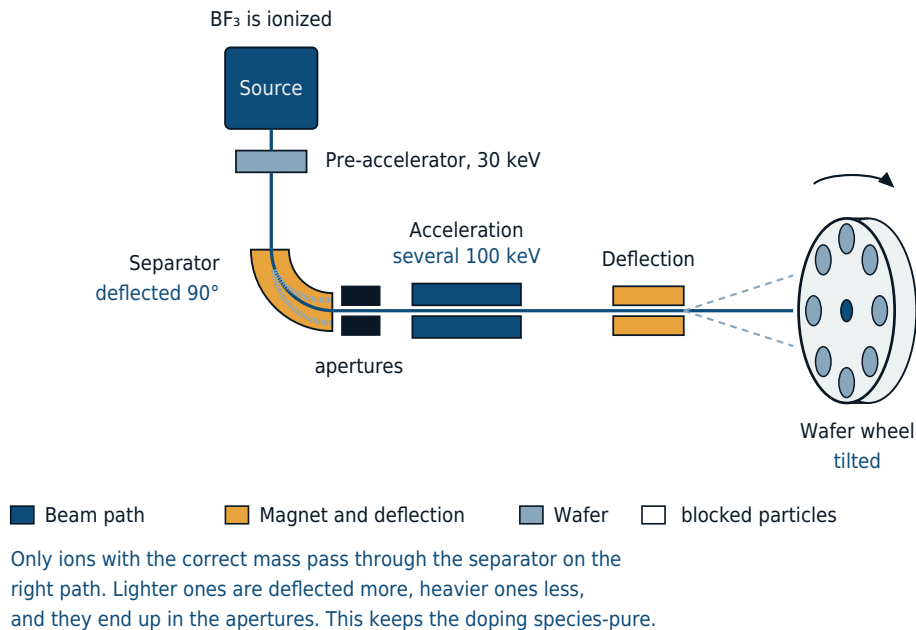
In ion implantation, charged dopants (ions) are accelerated in an electric field and directed onto the wafer. The penetration depth can be set very precisely by reducing or increasing the voltage used to accelerate the ions. Since the process takes place at room temperature, dopants introduced earlier cannot diffuse out. As with diffusion, regions that should not be doped are covered with a mask, though for implantation a photoresist mask is sufficient.

An implanter consists of the following components:

- Ion source: the dopant gas (e.g. boron trifluoride BF_3) is ionized (electrons emitted from a hot cathode collide with the gas particles; impact ionization continuously generates positive ions and free electrons)
- Pre-accelerator: the ions are drawn out of the ion source at roughly 30 kiloelectronvolts
- Mass separator: the charged particles are deflected by 90° in a magnetic field. Particles that are too light or too heavy are deflected more or less than the desired ions and are caught by screens behind the separator
- Acceleration lane: several hundred keV accelerate the particles to their final energy (200 keV accelerate boron ions to roughly 2,000,000 m/s)
- Lenses: lenses are distributed throughout the system to focus the ion beam
- Beam deflection: electric fields deflect the ions to irradiate the desired location
- Wafer station: the wafers are brought into the ion beam either individually or mounted on large rotating wheels

Illustration of an ion implanter

Ion implanter
The separator lets through only the desired ion species



Penetration Depth and Annealing

Penetration depth of ions in the wafer

Unlike diffusion, the particles do not penetrate the crystal through their own thermal motion but are shot into the crystal lattice at high velocity. Collisions with silicon atoms slow them down along the way. The impacts knock silicon atoms out of their lattice sites, while the dopant ions themselves mostly come to rest on interstitial sites. There, they are not electrically active, since they form no bonds with neighboring atoms that could give rise to free charge carriers. The displaced silicon atoms must be reincorporated into the crystal lattice, and the electrically inactive dopants must be activated.

Annealing the crystal lattice and activating the dopants

A temperature step at around 1000 °C moves the dopants onto lattice sites (beforehand, only about 5 % of the dopant atoms sit on lattice sites). Lattice damage from the collisions is already cured at around 500 °C. Because the dopant atoms keep moving through the substrate at these high temperatures, these steps are carried out for only a very short time.

This is exactly where the trade-off of the process lies: the temperature must be high enough to move the dopants onto lattice sites, and the time short enough that they do not migrate away while doing so. Both can only be achieved

together by continually shortening the heating duration. What began as an hours-long furnace process first became rapid thermal processing over seconds, then spike annealing, where the target temperature is only passed through rather than held, and finally flash-lamp or laser annealing in the millisecond range. Here, only the topmost layer of the wafer is heated while the substrate underneath stays cold.

For very shallow doping profiles, the ion mass is additionally increased: instead of implanting individual boron atoms, molecules containing several boron atoms are implanted. At the same acceleration voltage they carry less energy per boron atom and therefore penetrate less deeply – a way to create very shallow junctions without having to operate the implanter at impractically low voltages.

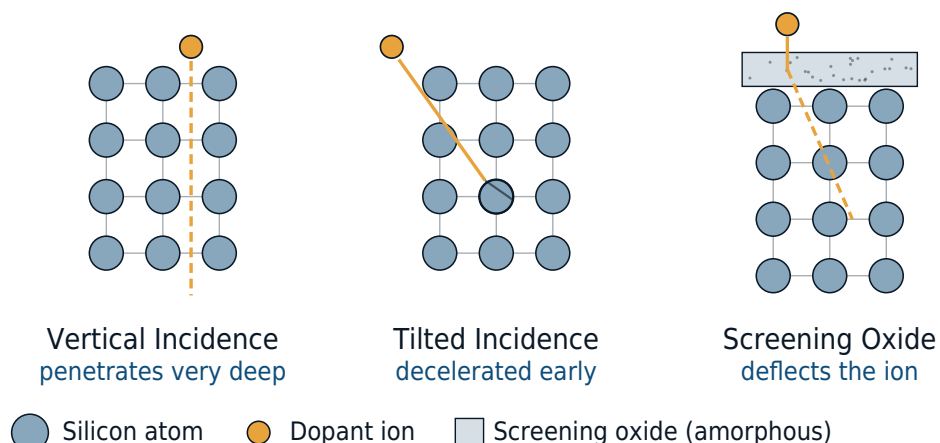
Channeling

The substrate used is a single crystal, meaning the silicon atoms are arranged periodically and form channels. Implanted dopant atoms that travel parallel to these channels are only weakly decelerated and penetrate very deeply into the substrate. There are two ways to prevent this:

- Wafer tilt: the wafers are tilted by roughly 7° relative to the beam direction. This means the ions are not shot in parallel to the lattice channels and are decelerated early through collisions.
- Screening oxide: a thin oxide is applied to the wafer surface, which deflects the ions and prevents them from arriving perpendicular to the substrate

Channeling and Countermeasures

How ions penetrate unintentionally deep – and how to prevent it



If an ion travels exactly parallel to a lattice axis, it finds almost no scattering centers and penetrates far deeper than intended. A tilt of about 7° or a thin screening oxide layer prevent this.

Characteristics and Implanter Types

Characteristics

- The reproducibility of ion implantation is very high
- The room-temperature process prevents other dopants from diffusing out
- A photoresist mask is sufficient; an oxide mask, as needed for diffusion, is not required
- Ion implanters are expensive, and the cost per processed wafer is high
- Dopants do not spread laterally underneath the mask (only minimally, due to collisions)
- Nearly every element can be implanted at the highest purity
- Similar to dopant deposits in the quartz tube during diffusion, ions can accumulate on walls or screens and later be dislodged and carried onto wafers during subsequent implants
- Vertical surfaces are difficult to dope with a directional ion beam. Tilting the wafer helps to a limited extent; for structures such as FinFET fins, doping is instead performed from a plasma whose ions arrive from all directions
- The implantation process takes place under high vacuum, which must be generated with several turbomolecular or cryopumps

Several types of implanters exist, with medium-current and high-current implanters used most often. Medium-current implanters are suited to small and medium ion doses ($1 \cdot 10^{11}$ – $1 \cdot 10^{15}$ ions/cm²), high-current implanters to doses of $1 \cdot 10^{15}$ – $1 \cdot 10^{17}$ ions/cm².

Due to its advantages over diffusion, ion implantation has largely become the standard.

Doping by Alloying

For completeness, it is worth mentioning that besides these two methods there is also doping by alloying. Since this method brings disadvantages such as cracking of the substrate, it is barely used in semiconductor technology today.

Plasma Doping (PLAD)

For structures that no longer lie flat on the wafer but rise into the third dimension like FinFET fins or gate-all-around nanosheets, the classic directional ion beam reaches its limits: sidewalls sit transverse to the beam and receive almost no dopant, even with a tilted wafer. Plasma doping (PLAD) solves this by abandoning the beam-line architecture of the classic implanter entirely.

The wafer sits on a pulsed high-voltage platen inside a plasma chamber, in which a process gas of the desired dopant (e.g. diborane B₂H₆ or phosphine PH₃)

is struck. Each high-voltage pulse builds up a plasma sheath around the wafer, across which ions are accelerated toward the surface – not as a narrow beam from one direction, but broadly from the entire plasma. Because the sheath follows the wafer's contour, sidewalls of fins and ridges are doped nearly conformally as well.

The price is giving up the mass separator: since no magnetic field sorts the ions by mass, every species ionized in the plasma reaches the wafer together – alongside the desired dopant ion also hydrogen, molecular fragments, and possible contaminants from the process gas. Dose and species purity are therefore controlled less precisely than in classic mass-separated implantation (see [Ion Implantation](#)). Because the entire wafer is exposed at once instead of being scanned point by point, throughput is considerably higher, and the typically very low pulse energies of a few kiloelectronvolts suit extremely shallow doping profiles well.

In production, PLAD is therefore used specifically where classic implantation fails on geometry alone – for example to dope fin sidewalls, or as a replacement for heavily tilted extension implants where shadowing by neighboring structures would otherwise become a problem (see also [Characteristics and Implanter Types](#)).

Halo and Well Implants in CMOS Integration

A finished transistor doesn't result from a single implant, but from the interplay of several implantation steps, each separated in timing and energy, each with its own depth and purpose.

Wells

Right at the start of front-end processing, well implants at MeV-range energies create large-area n- or p-doped regions deep in the substrate, in which transistors of the opposite type are later built – an n-well hosts p-channel transistors, a p-well n-channel transistors. These implants extend several hundred nanometers to a few micrometers deep and mostly determine the electrical behavior of the substrate as a whole.

Halo/Pocket Implants

Directly beneath the future source and drain edge sits an oppositely doped, narrow region, the halo or pocket implant. It locally raises the dopant concentration at the edge of the channel and thereby counteracts the short-channel effect: without it, the depletion zones of source and drain would expand

far enough at short channel lengths to touch (punch-through), and the transistor would lose control over the channel. Because this region must sit directly beneath the gate edge, it is implanted at a markedly steeper tilt angle than the other steps.

Four-Fold Rotation

Since transistors on a chip can be oriented in any of four directions relative to each other, a single tilted implant step would only dope symmetrically the gates whose channel direction happens to align with the tilt plane – all others would receive an asymmetric, orientation-dependent profile. The halo step is therefore split into four: the same dose is applied in four sub-steps at the same tilt angle, but rotated 90 ° around the wafer axis each time. In total, every gate receives a symmetric halo dose from all four sides regardless of its orientation on the chip.

LDD and Source/Drain Implants

The shallow, lightly doped extensions known as LDD (Lightly Doped Drain) sit right at the gate edge; they smooth the concentration profile between the channel and the deep, heavily doped source/drain implants, lowering the electric field at the drain edge – too abrupt a transition would otherwise generate hot carriers that damage the gate oxide. Only the final, high-dose source/drain implants provide the actual, low-resistance connection regions of the transistor.

Amorphization and Defect Annealing

At sufficiently high dose, especially with heavy ions, implantation disrupts the crystal structure so thoroughly that the affected layer turns amorphous – the silicon locally loses its regular lattice order entirely. Rather than avoiding this, modern processes exploit it deliberately.

Pre-Amorphization (PAI)

Before the actual dopant implant, a pre-amorphization step (Pre-Amorphization Implant, PAI) shoots electrically inactive ions such as germanium or silicon itself into the surface, deliberately creating a uniformly amorphous layer. The subsequent dopant then no longer hits a crystal lattice with open channels but disordered material – [channeling](#) is thereby ruled out from the start, regardless of tilt angle.

Solid-Phase Epitaxy and End-of-Range Defects

During the subsequent [anneal](#), the amorphous layer regrows single-crystalline

starting from the intact crystal structure underneath (solid-phase epitaxy). Right at the original boundary between amorphous and crystalline material, however, a band of defects often remains, the end-of-range (EOR) defects: dislocation loops of excess interstitial atoms that were shot past the actual amorphous zone during implantation and are not fully incorporated during annealing.

Transient Enhanced Diffusion (TED)

These excess interstitial atoms are not just electrically inactive crystal defects; they also accelerate the diffusion of nearby dopants – boron above all – by a large factor for the short duration of their own annealing. This effect is called transient enhanced diffusion (TED) and causes doping profiles to spread significantly further during annealing than equilibrium diffusion at the given temperature would predict. Because TED only acts as long as excess interstitial atoms are present, countermeasures target exactly this window: very short anneal times (see [Penetration Depth and Annealing](#)), which give the defects no time to migrate, or a carbon co-implant that specifically traps interstitial atoms before they can drive dopant diffusion.

Dose Metrology

How many ions actually made it into the wafer must be verifiable both during implantation and afterward – the two checks fundamentally measure different quantities.

Faraday Cup: Real-Time Dose Control

Inside the implanter itself, one or more Faraday cups handle dose control: they capture a known fraction of the ion beam and measure its current. Integrated over time, this yields the accumulated charge, and from that, relative to the exposed area and the charge per ion, the actually applied dose. The implanter automatically stops the process once the target dose is reached. Multiple cups at the edge and center of the wafer also allow the beam's uniformity across the wafer to be monitored. This measurement is fast and runs with every wafer, but says nothing about where in the wafer the ions actually ended up or how many are electrically active after annealing.

SIMS: The Depth Profile Afterward

Secondary ion mass spectrometry (SIMS) sputters the wafer surface layer by layer with a primary ion beam and mass-analyzes the secondary ions released in the process. This yields a highly precise concentration profile of the dopant

versus depth – the reference method for verifying range, straggle, and, by integrating the profile, the actually introduced total dose after the fact. SIMS is destructive and too slow to check every single wafer; it is used for process qualification and calibration, not in routine production.

For routine production, four-point-probe sheet resistance measurement is used instead: fast, non-destructive, and sensitive to the electrically active dose after annealing – though without depth resolution. Only the combination of routine Faraday-cup and sheet-resistance measurement with occasional SIMS calibration gives a complete picture of the doping.