Chapter 4

Fabrication Technology

YUQING JIAO, JEROEN BOLK, HUUB AMBROSIUS, AMER BASSAL, KLEMENS JANIAK

Figure 4.1 gives a schematic impression of the waveguides structure of InP-based PICs. The technology used for the fabrication of PICs determines which designs are manufacturable and permitted (i.e. comply with the design rules). It also determines the process windows (the range of process parameters that will yield the required results). Therefore, it is important for designers to have a basic knowledge of the fabrication technology, to understand the impact of process variations on the performance of the device, and to bear the fabrication tolerances in mind when designing. This chapter starts with an overview of the generic InP technology used to fabricate InP-based PICs. It is a very versatile technology thanks to the efficient integration of high-performance lasers, optical amplifiers, modulators and detectors, offered by the InP-based material system.

4.1 The generic process flow

This section provides a step-by-step overview of the major fabrication steps of the generic integration process, developed at TU Eindhoven[1], as shown in Figures 4.4-4.11. A detailed description of the different process steps is given in the follow-up sections. It should be noted that the illustrated process flow is only one way to achieve the targeted structures, many variations are possible to get the same result. In a qualified generic process each of the process steps has been extensively tested so that the chance on failures is small. Changing some steps in the processing will often lead to problems in the performance and will usually require extensive re-testing. Designing in a qualified generic process will, therefore, lead to substantially shorter fabrication times than designing in a process under development. Even if it is not yet fully stable, designing in

generic process

Yuqing Jiao, Jeroen Bolk and Huub Ambrosius are with the Eindhoven University of Technology. Amer Bassal and Klemens Janiak are with the Fraunhofer Heinrich Hertz Institut in Berlin and contributed Section 4.2. Support from Tjibbe de Vries and Akanksha Kapoor is gratefully acknowledged.

Support from Tjibbe de Vries and Akanksha Kapoor is gratefully acknowledged.

[1] The process is licensed to SMART Photonics, who uses it for its MPW-runs with a few modifications.

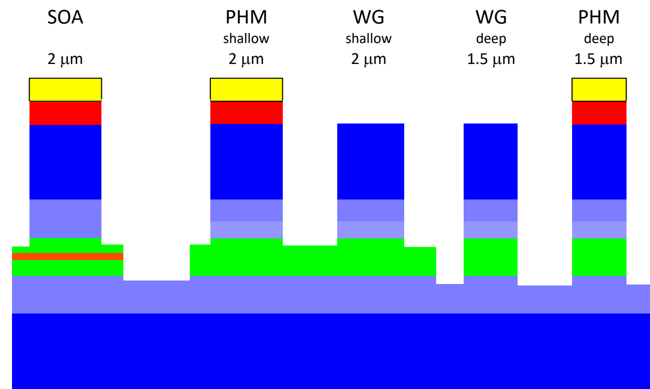


Figure 4.1: Schematic waveguide structure of generic InP-based PICs

such a process will be much faster than designing in a newly developed process, which may take several years for complex processes.

4.1.1 Epitaxial (re)growth

The process flow starts from an InP substrate (Figure 4.2a), which can be n-doped or semi-insulating (SI). For details see Section 4.6. In this and the following figures, the process for an n-doped InP substrate is shown. On semi-insulating (SI) substrates the process is slightly different, mainly in the metallization steps, because in this process both the p- and the n-contacts are at the upper side of the chip.

MOVPE
MOCVD
separate-confinement heterostructure
Q1.25

Typically, the surface of the substrate is epi ready, it can be loaded into a Metal-Organic Vapor Phase Epitaxy reactor (MOVPE or MOCVD: Metal-Organic Chemical Vapor Deposition) to grow the required layer stack. For the generic process the grown layer stack contains multi-quantum wells (MQWs) as the active medium, separate-confinement heterostructures[2] (SCHs) as well as the p- and n-doped cladding layers (Figure 4.2b, for details see Section 4.5). In the figure, p-InP and n-InP denote p- and n-doped InP materials, respectively, and n.i.d. denotes non-intentionally doped materials. Q1.25 denotes the InGaAsP quarternary material, lattice matched to InP, whose bandgap corresponds to a wavelength of 1.25 μm. These naming conventions will be used throughout the rest of the chapter.

active-passive integration
regrowth
butt-joint interface

To achieve butt-joint active-passive integration, a regrowth process is applied (Figure 4.2c-g). The key idea is to keep the MQWs in the areas for lasers and amplifiers intact, while removing the MQWs in the rest of the areas on the wafer. Then a passive layer stack is regrown for the phase modulators and the waveguides. The regrown stack has the same thickness as the original active stack, forming a so-called butt-joint interface.

PECVD
RIE

The process starts with depositing a dielectric hard mask layer (Silicon Nitride SiN_x and/or Silicon Oxide SiO_x) on the grown wafer using Plasma Enhanced Chemical Vapor Deposition (PECVD, for details see Section 4.10.2). A lithography step is applied to transfer the mask patterns of the active mesas to a photoresist layer, and subsequently to the SiN_x layer by means of a reactive ion etching step (RIE, for details see Section 4.11.2), as shown in Figure 4.2d. The pattern is transferred to the active stack

[2] A separate confinement heterostructure is a waveguide structure in which the optical waveguiding is provided by a quaternary layer in which the active QW layers are embedded.

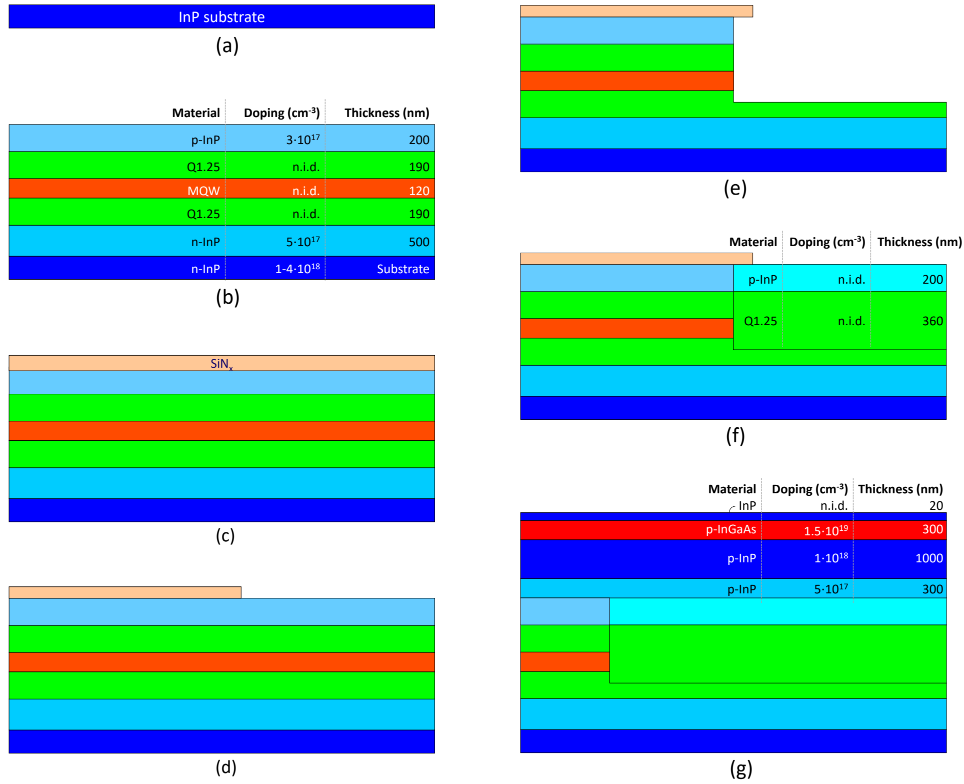


Figure 4.2: Epitaxial regrowth

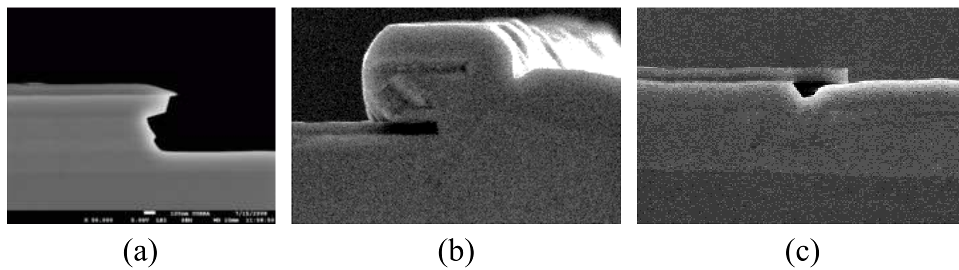


Figure 4.3: (a) SEM-photograph of the undercut profile. (b) Undercut too small: regrowth over mask. (c) Undercut too large: hole formation under mask.

wet etching *growth rate enhancement*

by means of wet etching (Figure 4.2e, for details see Section 4.11.1). An undercut is deliberately created underneath the edge of the SiN_x hard mask, for compensation of the growth rate enhancement (GRE) effect which occurs near the mask edges during the regrowth [107] to achieve a flat butt-joint interface. Figure 4.3a shows a SEM-photograph of the undercut profile. After etching the active stack the wafer is loaded back into the MOVPE reactor for the regrowth (Figure 4.2f, for details see Section 4.7.2). Under proper circumstances the growth will only happen at exposed semiconductor surfaces, and not on the SiN_x hard mask thanks to its non-crystalline nature. Note that when the SiN_x hard mask is too large (e.g., in the order of hundreds of microns), undesired growth of crystalline particles on the mask can occur, which can cause problems in subsequent process steps. The width of the active regions is, therefore, restricted to a few tens of micrometers.

The size of the undercut is important to get a smooth butt-joint as shown in Figure 4.2f. If the undercut is too small, regrowth will occur over the mask (so-called rabbit ears), as shown in Figure 4.3b. If the undercut is too large, a hole will form under the mask, as shown in Figure 4.3c. In both cases the butt-joint between the active and the passive waveguide will show a discontinuity which will cause coupling loss and reflection.

The third (and final) MOVPE growth-step is to realize the p-doped InP cladding layer and an InGaAs contacting layer as well as a thin InP protection layer, as shown in Figure 4.2g. After the (re)growth process, the wafer will be subject to multiple lithography and etching steps to form a few different waveguide structures. For reasons of clarity, the butt-joint interface in Figure 4.2g is shifted towards the left-hand side, to make room for drawing multiple waveguides in the passive area.

4.1.2 Waveguide etching

ICP-RIE *isolation section*

After the epitaxial growth the first process module is the etching of the waveguides. A SiN_x hard mask layer is deposited by PECVD on the wafer, followed by spin-coating of a photoresist layer (Figure 4.4h, for details see Section 4.8.6). For patterning of waveguides, a positive photoresist with high resolution is used. In a following lithography step, all active and passive waveguide patterns are transferred to the photoresist layer, and subsequently to the SiN_x layer by RIE (Figure 4.4i). The remaining photoresist is removed by an O_2-plasma. To create different etch depths for deep and shallow waveguides, additional lithography is performed to cover the shallow waveguides and only etch the deep waveguides. This is achieved by using a thick photoresist, as shown in Figure 4.4j. Using the photoresist and the exposed SiN_x together as the etching mask, an inductively coupled plasma RIE (ICP-RIE, for details see Section 4.11.2) is performed to define the depth difference between the deep and shallow waveguides (Figure 4.4k). After the etching, the photoresist layer is removed (Figure 4.4l). The second ICP etching defines the depth difference between the bottom of the shallow waveguide and the isolation section. It is achieved by ICP etching into the p-InP cladding using the original SiN_x mask (Figure 4.4m).

BHF

The definition of the electrical isolation sections is done with a photoresist mask. A new layer of AZ4533 photoresist is spin-coated and patterned to protect the wafer except at the isolation waveguide structures (Figure 4.4n). Since the isolation sections require high resistivity, the p-doped cladding must be removed. Therefore a wet etch step using buffered hydro fluoric acid (BHF) is performed to remove the SiN_x mask on top of the isolation waveguides (Figure 4.4o). This wet etch is highly selective and does not attack the photoresist and semiconductor materials significantly. The photoresist layer is then removed (Figure 4.4p). The third ICP etching step will realize the differ-

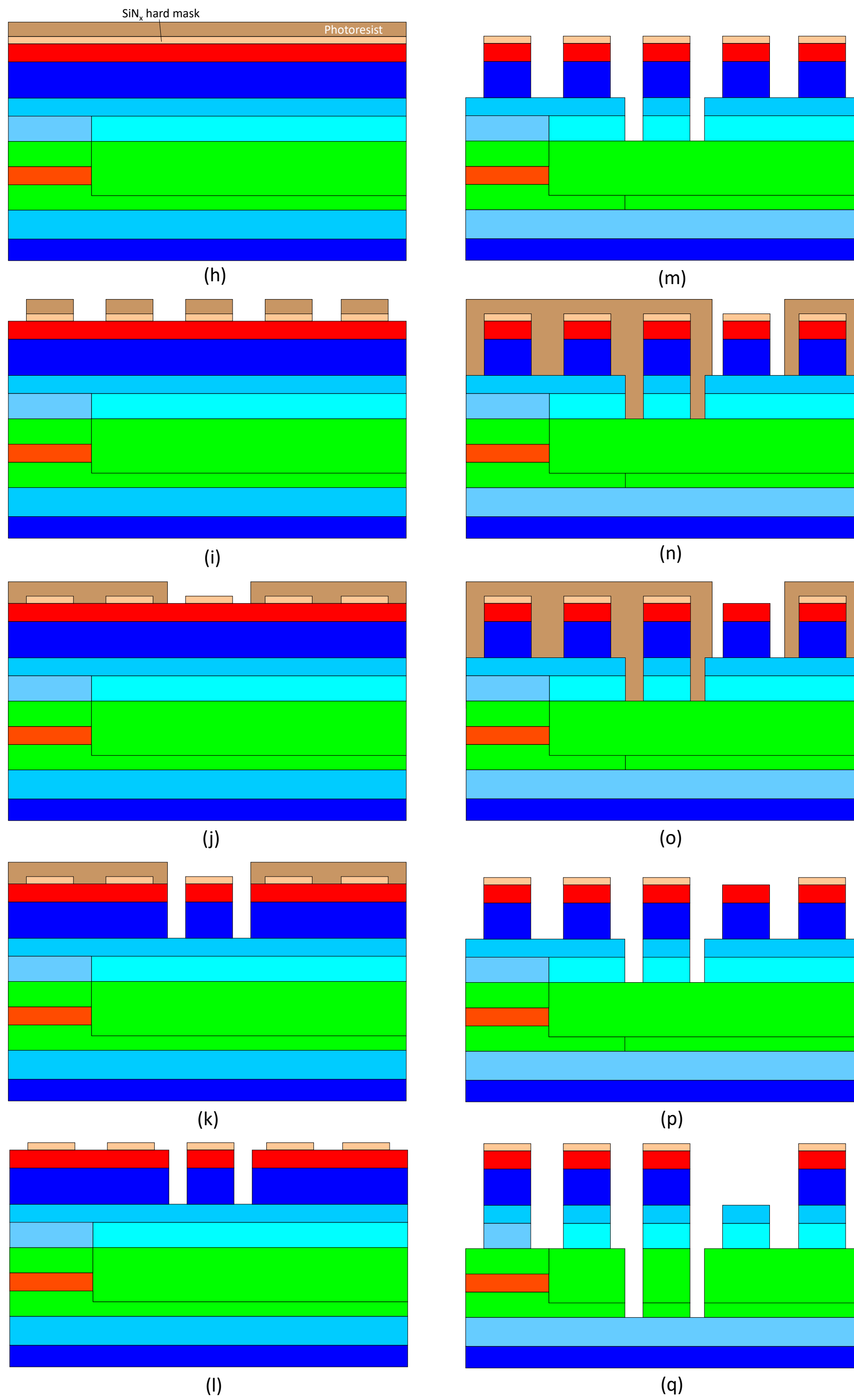


Figure 4.4: First three waveguide RIE steps.

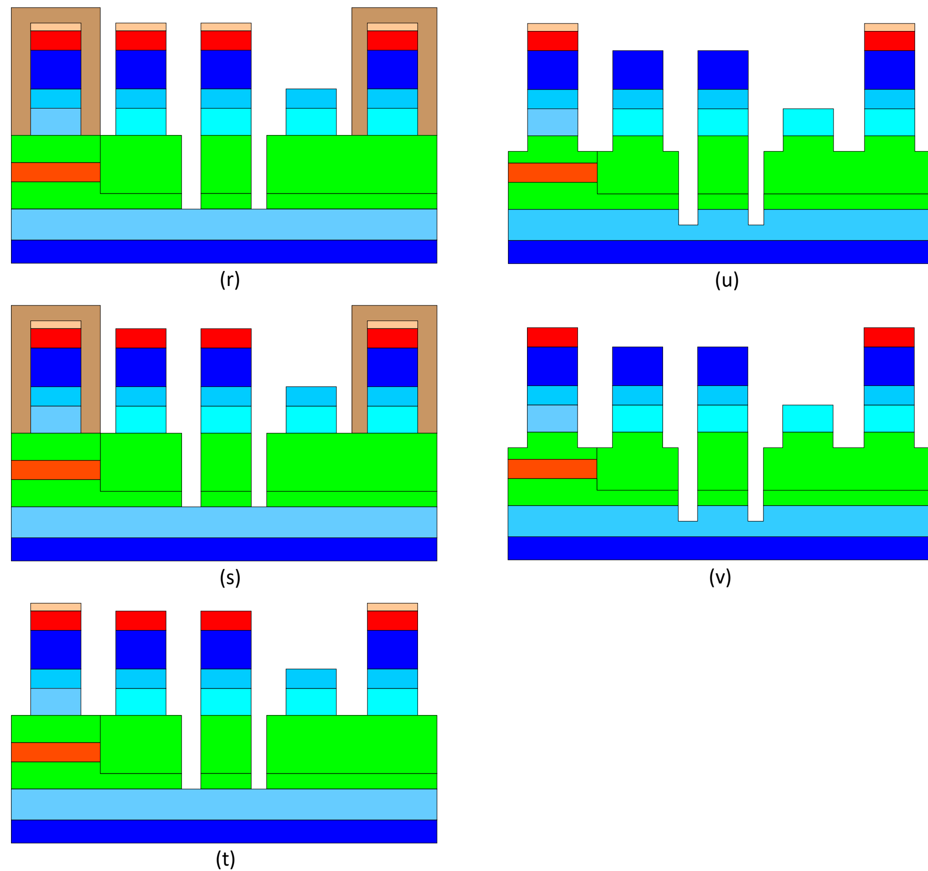


Figure 4.5: Fourth waveguide RIE step.

ence between the isolation waveguides and the top of the shallow waveguides (Figure 4.4q). Since there is no protection over the isolation sections, their p-claddings will be removed during the etching step. Only a lightly p-doped InP layer of approximately 300 nm thick will eventually remain and act as the waveguide ridge.

The fourth ICP etch step, which is the final one, removes the heavily p-doped InGaAs layers on top of the passive waveguide. This is essential for low loss of the waveguides. A new photoresist layer is used to protect the SiN_x hard masks in the active devices; the amplifiers, detectors and phase modulators which require the InGaAs layer for contact metallization (Figure 4.5r). A BHF wet etch step removes the SiNx material in unprotected areas (Figure 4.5s), followed by a resist stripping step to remove the photoresist (Figure 4.5t). The fourth ICP etch step is then carried out to etch away the exposed InGaAs materials, and at the same time complete the deep and shallow waveguide definition (Figure 4.5u). At this stage, all waveguide structures are realized. Therefore, the remaining SiN_x can be removed completely (Figure 4.5v). This step finalizes the multi-stage etching process.

The lithographic process is very sensitive to disruptions in the surface topology. Figure 4.6 illustrates waveguide defects caused by particles (left) and by discontinuities (rabbit-ears) at the junction between the active and the passive region. Around a particle the resist will be thicker and and during development it will not disappear in a

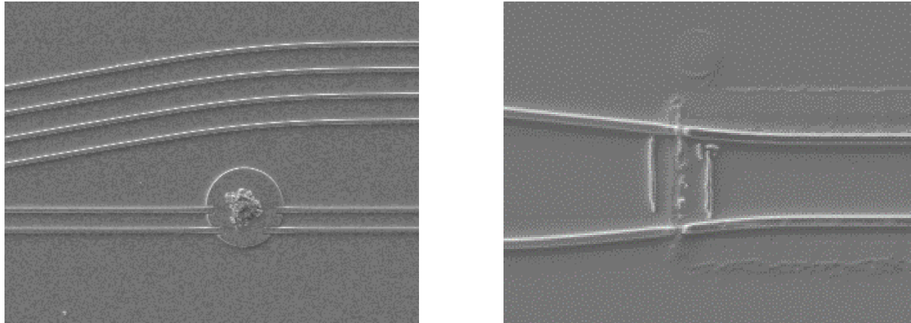


Figure 4.6: Waveguide defects caused by thickness variations in the photoresist around a particle (left) or at an imperfect junction between the active and the passive region (right)

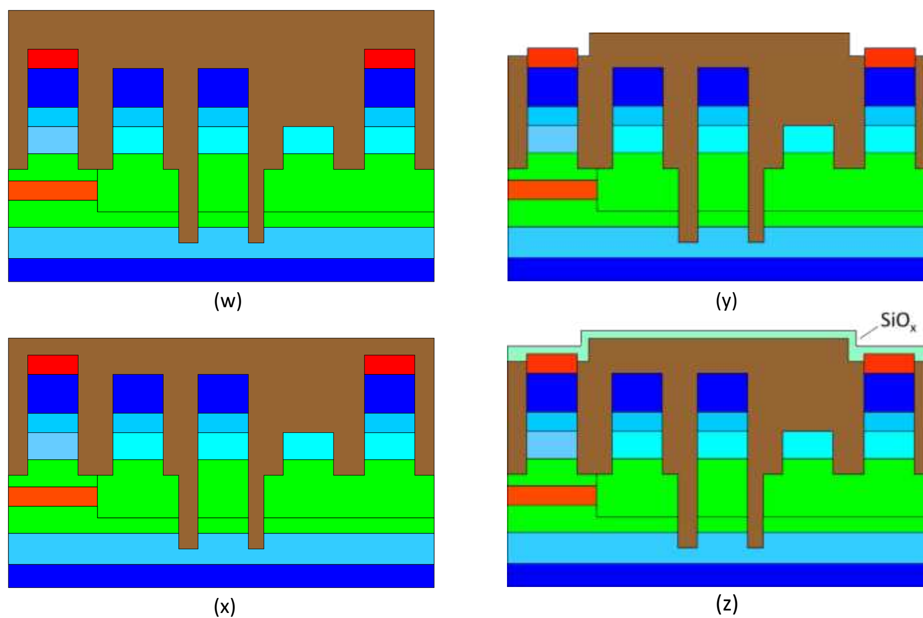


Figure 4.7: Contact opening.

region around the particle, so that the waveguide is not defined there. At rabbit ears it will be thinner and it may open also on the waveguide ridge and cause an interruption. Thorough cleaning of the wafer, as described in Section 4.4, and a well controlled surface topolgy are, therefore, of key importance for a good process yield.

4.1.3 Contact opening

metal contacts
interconnections
polyimide (PI)

Before the deposition of metal contacts and interconnections the wafer surface need to be planarized. To do so, multiple layers of polyimide (PI) are spin-coated on the wafer (Figure 4.7w). PI has good planarization properties and is very stable after curing. To access the p-InGaAs contact layer, the planarized surface needs to be etched back. This is done using a RIE process with a mixture of O_2 and CHF_3 to etch back until approximately 100 nm above the p-InGaAs layer (Figure 4.7x). Next a lithography with

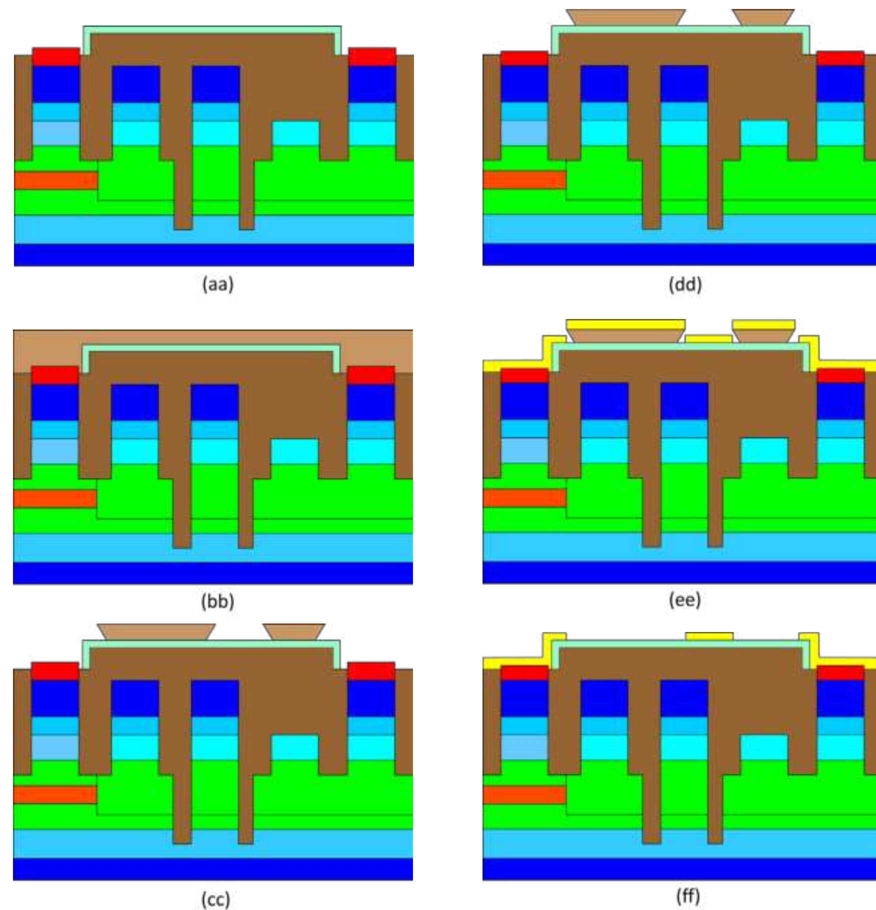


Figure 4.8: Contact and interconnect metallization.

positive photoresist is performed to open the area on top of the InGaAs contacts. A RIE etch step is used to remove the PI material on top of the InGaAs (Figure 4.7y). To improve the adhesion between the PI and the metals, a thin layer (50 nm) of silicon oxide (SiO_x) is deposited with PECVD (Figure 4.7z).

4.1.4 Contact and interconnect metallization

metallization The next step is the contact and interconnect metallization. The SiO_x on top of the InGaAs is removed by a lithography step using the same mask as for opening the contacts and a RIE etch step afterwards (Figure 4.8aa). For the metallization process, a negative photoresist is used to expose the areas where metal electrodes or interconnections are desired (Figure 4.8bb). By carefully controlling the exposure dose, a negative slope in the resist pattern can be achieved, as shown in Figure 4.8cc. This is essential for *lift-off process* the following lift-off process (for details see Section 4.10.4). Before the metal deposition, a cleaning step using a very diluted solution of H_2SO_4 : H_2O_2 is applied (Figure 4.8dd). This step will etch away a thin layer of the top InGaAs material, which may have defects accumulated during previous processing, and yield a fresh InGaAs surface. A *electron-beam evaporation* metal stack of Ti/Pt/Au is deposited onto the wafer using electron-beam evaporation. Thanks to the negative slope in the resist profile, the metal at the resist edge is discon-

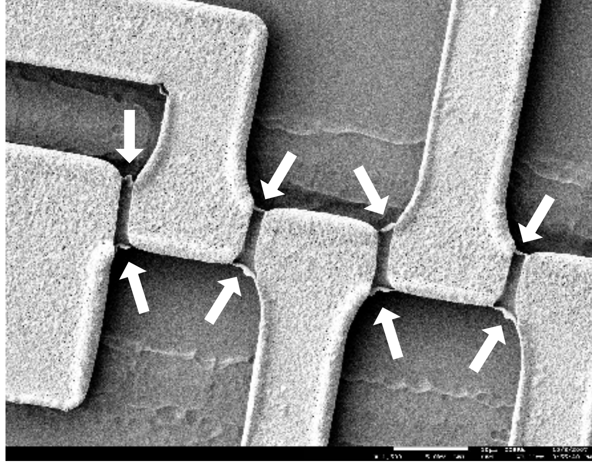


Figure 4.9: Short circuits between closely spaced metal contacts.

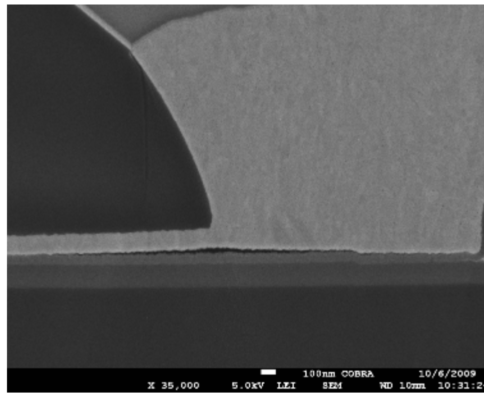


Figure 4.10: SEM-photo of an electroplated metal contact, showing a large tension at the edge of the plating window (dark layer is photoresist): the metal is teared of the underlying SiO_2 layer.

nected (Figure 4.8ee). The metal layer on top of the resist will be removed by soaking the wafer in an organic solvent such as acetone, resulting in the desired metal patterns for contacts and interconnections (Figure 4.8ff).

The SiO_x exposed to air is removed with a RIE step and a metal stack (Ti/Pt/Au) is deposited on the backside of the substrate for the n-contact, as shown in Figure 4.11gg. For optimal contact resistance between the metal and the semiconductor, a rapid thermal annealing (RTA) process step is applied at high temperature (in the order of 400 °C) during a very short time (10s of seconds).

rapid thermal annealing (RTA)

Figure 4.9 shows a number of (too) closely spaced metal contacts on an active waveguide section. It is seen that at the sidewall of the ridge there is metal left between the contacts, which will cause a short circuit. Reliable processing requires a combination of good process control and proper design rules for the minimum distances between contacts.

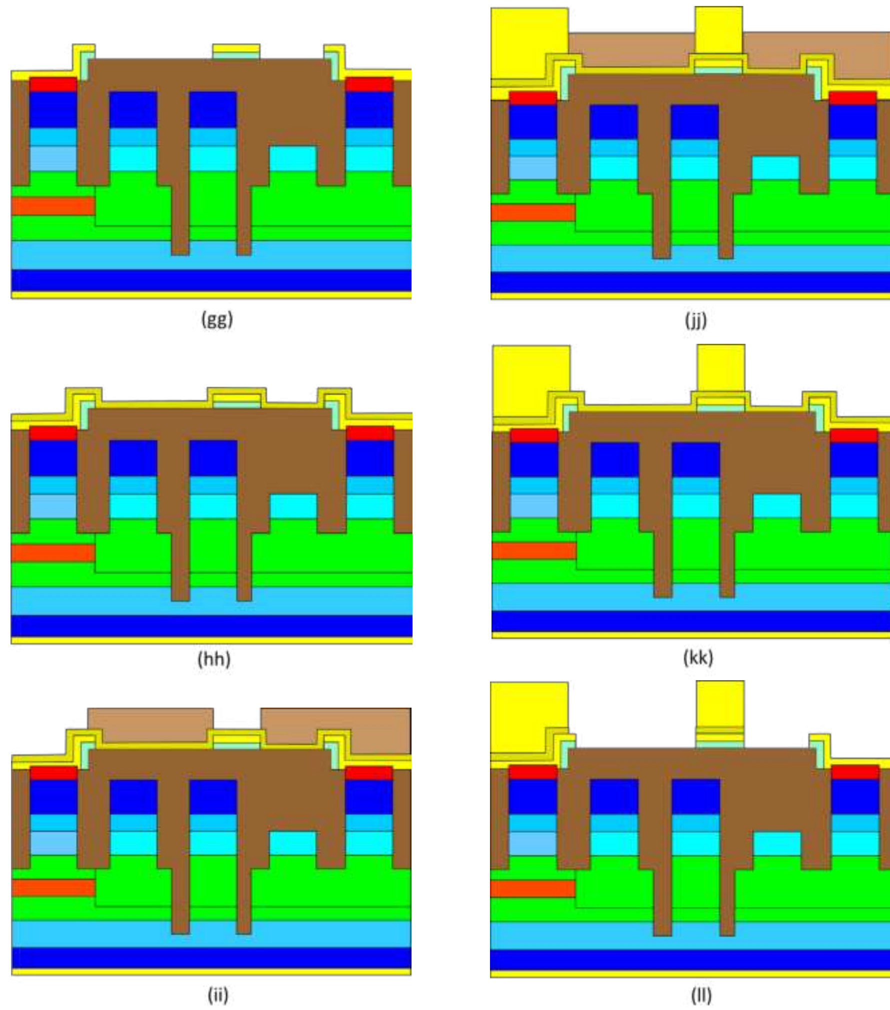


Figure 4.11: Electroplating.

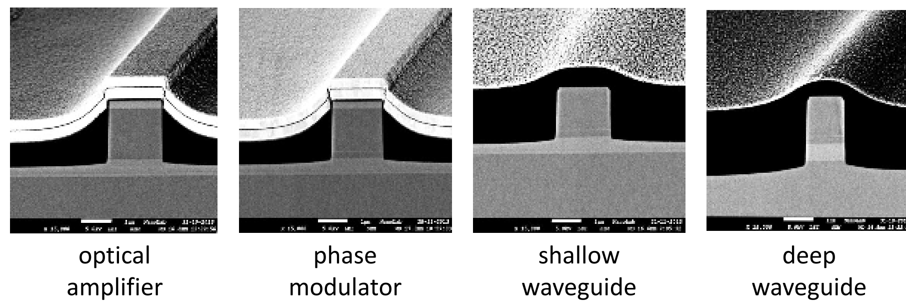


Figure 4.12: Cross sections of five Basic Building Blocks (SEM photographs).

4.1.5 Electroplating

For the optimal operation of amplifiers and lasers, which relies on a uniform current distribution, the metal layers need to be thick. This is achieved by electroplating. The applied approach for electroplating is shown in Figures 4.11hh-ll. It begins with sputter deposition of a thin seed layer of Ti/Au (50 nm) over the wafer surface (Figure 4.11hh). Sputtering is used for this step for its better step coverage as compared to evaporation. Next a lithography step is applied to cover the entire wafer surface except for the areas of amplifiers and lasers (Figure 4.11ii). In the meanwhile, the backside of the substrate is also protected with a photoresist layer (not shown). The wafer is placed in the gold-plating bath, and connected to the cathode of a DC power supply. As a result, Au will accumulate in the open areas on the wafer surface. Typically a plated Au-film thickness of 2-3 μm is realized. The parameters of the plating process should be chosen properly to avoid large stress in the plated layer. Figure 4.10 shows a plated gold layer that has detached from the underlying SiO_2-layer due to large internal strain.

electroplating

sputter deposition

Finally, the photoresist used for the plating is removed (Figure 4.11kk). The remaining seed layer is etched away wet-chemically. At this point, the wafer fabrication is completed, as shown in Figure 4.11ll). Backend processing can be applied for die singulation, polishing and coating (for details see Section 4.12).

backend processing

4.2 The generic process flow of HHI

Figure xx shows a cross-section of the waveguide structure used by HHI. The process of transforming the bare InP wafer into fully functional PICs with more than 40 Building Blocks requires a very complex and sensitive process with many variables which requires regular monitoring and adjustments. The main underlying reason for this complexity is the combination of the active BB types in a common monolithic substrate. The HHI-process uses three different quantum well (QW) stacks for the three main types of active BBs: lasers, modulators and photodiodes (PDs). Therefore, the process involves at least three QW epitaxy steps along with various etching, deposition and metallization steps, with each step having its own influence and variables. Thus, controlling and keeping track of those variables is a crucial part of the fabrication process.

4.2.1 Gain stack epitaxy

Similar to the generic process at SMART photonics, the process at HHI starts with an epitaxy step, on a semi-insulating (SI) InP wafer . On the bare wafer, the base epitaxy is applied using an epitaxy reactor creating the quantum well stack for the lasers and optical amplifiers. The base epitaxy also includes the n-doped layer for the n-contact to be formed later in the process and a p-doped cap layer on top (Figure 4.14a). Once the whole wafer is covered with the gain stack, a lithography step is then performed to define areas for the stack to stay and etch away the rest (Figure 4.14b). At this stage, the active regions for the SOAs, DFBs and EAMs are defined and protected with SiN_x.

epitaxy

semi-insulating (SI)

4.2.2 Modulator stack regrowth and grating formation

To integrate Electro-Absorption Modulators and tunable gratings into the platform, another epitaxy step is applied to create the modulator quantum wells. In this epitaxy

Electro-Absorption Modulator (EAM)

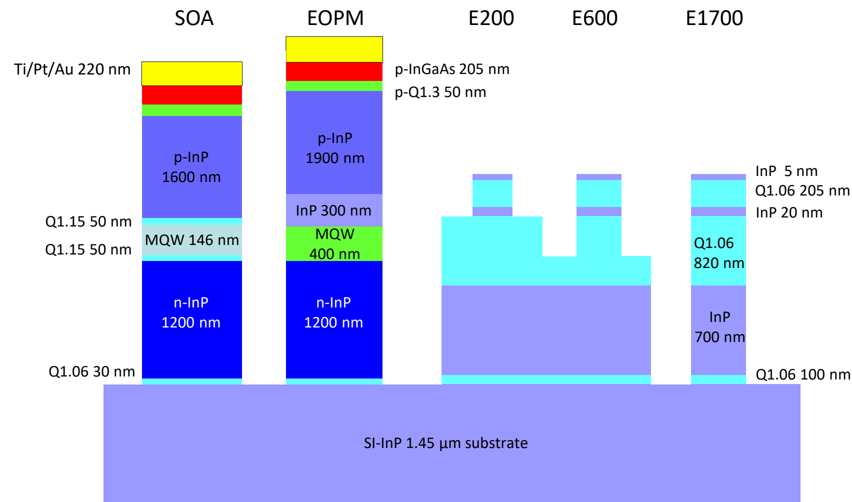


Figure 4.13: Cross-section of the HHI waveguide structure.

tunable grating
butt joint

regrowth process, the modulator stack is grown, which includes an n-doped layer and a non-intentionally doped (n.i.d) InP (or simply i-InP) cap layer on top (Figure 4.14c). At this point, the gain-modulator butt joints are created which have an angled interface with respect to the waveguide to reduce back reflections. The modulator stack is engineered such that it contains different materials suitable for the subsequent selective etching steps. This allows the simultaneous chemical etching of the gain and modulator stacks to remove the p-cap layers above the gain MQW in preparation for the common p-contact (Figure 4.14e).

DFB laser
DBR laser

Two e-beam lithography steps are then applied to define the gratings on the gain and modulator QWs including the InP cap layer of the modulator stack, thereby creating the DFB lasers and the tunable Bragg gratings that are also used in DBR lasers. These two steps are made using special e-beam sensitive resists. The first e-beam lithography step is selectively done on the modulator stack using a protective resist applied elsewhere, in which the modulator cap layer is etched to create the DBRs. Then the second e-beam lithography step is performed to define the gratings in the gain stack necessary for the DFBs. Then, a final etching step is done to etch both DBR and DFB gratings such that they are deep enough to go beyond the cap layer and through a small thickness of the QWs (Figure 4.14d). This completes the creation of the vertical gratings for the DFB and DBR BBs.

However, to create the p-contacts, there still needs to be a final epitaxy regrowth of a thick p-doped cap layer on top, which is done after removing the residual resists and protective layers (Figure 4.14e). The p-contacts are then defined with a lithography step where the p-InP is etched to create isolated islands of p-InP for the p-contacts of all devices, both in the gain and modulator stacks. The isolation areas between the contacts are then filled with an epitaxial layer of i-InP to insulate the contacts. This forms the HHI BB ISO-section, which contains the modulator QWs (that are intrinsic) but with no p-contacts.

At this point, the gain and modulator stacks are finished, along with their butt joints, p-contact cap layers, and gratings. And in preparation for the next steps involving the development of the passive BBs and PDs, the wafer is cleared from all the unused active stacks to make room for the passive and PD stack epitaxial overgrowth process. The

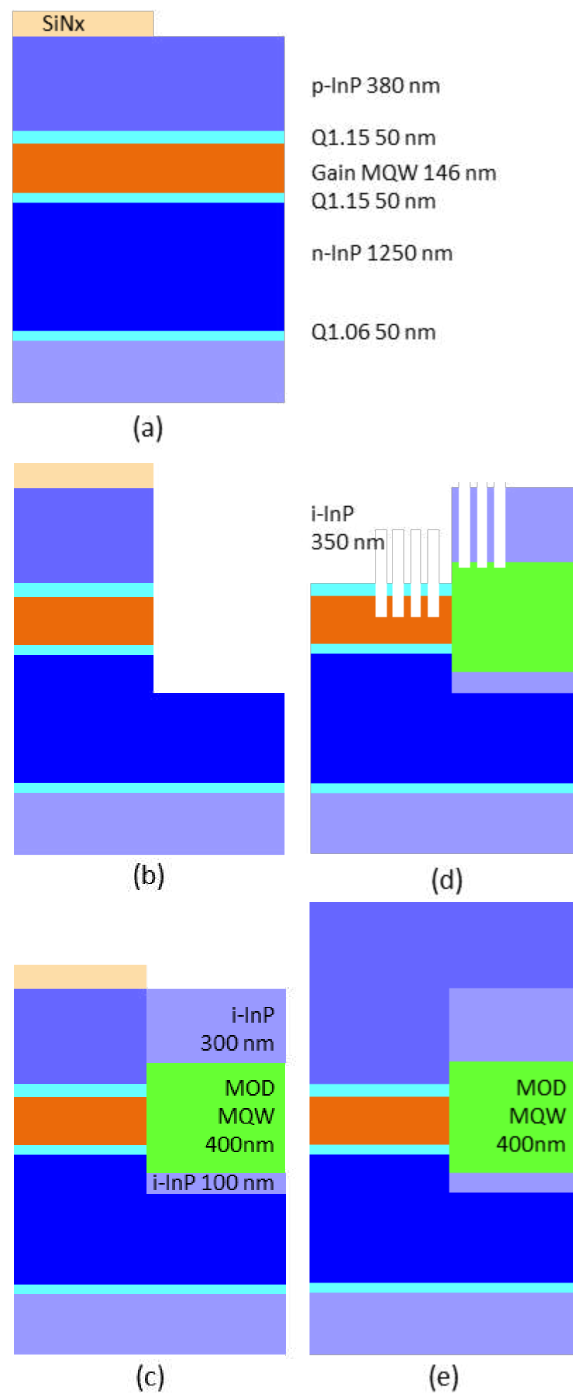


Figure 4.14: Gain stack base epitaxy growth and SiN_x hard mask lithography (a), gain stack etching (b) and modulator stack regrowth (c), gain p-layer etching and grating formation (d) and common p-contact regrowth (e).

active BB areas are protected with SiN_x, applied through lithography, and the rest of the wafer is etched all the way down to the InP substrate (Figure 4.15h).

4.2.3 Passive and active WG formation

The active BBs and the newly formed PDs are then protected with resist for the deposition of SiN_x on the WG areas, also covering the metal contacts on the gain and modulator stacks. Then, with a lithography step, the whole wafer is etched, to define the shallow WGs (E200) with SiN_x on top as a mask for subsequent etching steps. This WG etching step (~200 nm) also includes the gain and modulator active WGs and the passive structure below the PD mesas.

E200 waveguide

A lithography step is applied with an etch window around shallow WGs to etch a deep trench thereby creating the deeply etched gain and passive WGs (E1700). The etching step on the shallow WGs brings the etch depth to ~1700 nm, while WGs to be kept shallow are protected with photoresist. The modulator WGs are etched even further with a separate lithography step. Another lithography step similar to the E1700 etch lithography is applied with etch windows and with a similar etching step to create the E600 WGs by further etching the shallow E200 WGs to a depth of ~600 nm. In the same lithography and etching steps, the ~400 nm wet etch exposes the n-InP underneath the MQWs in the gain and modulator BBs for the creation of n-contacts in subsequent metallization steps.

E1700 waveguide

E600 waveguide

All the passive BBs are now complete except for the polarization BBs and the Spot-Size Converters (SSCs) which require a special etching recipe to form the beveled side-walls necessary for polarization operations and mode broadening. For the SSC with bevels on both sides, the process starts with E200 WGs and an SSC lithography with etch windows for further etching and then a second etching step to create the beveled round edges with a total depth of ~1700 nm. Polarization BBs undergo this step as well, however, only on one side of the WG to form the asymmetric WGs required for polarization rotation. The SSC WG becomes much wider at its bottom because of the two beveled side-walls, thus pushing the mode down towards the substrate. Without any vertical boundaries, the mode will disperse in the substrate, therefore two deep trenches are etched a few micrometers away from both sides of the WG. This creates a wide WG to couple the “deep mode” to a single mode fiber with the correct mode field diameter (MFD).

Spot-Size Converter (SSC)

polarization rotation

mode field diameter (MFD)

4.2.4 Passive and PD stack regrowth, contacts and PD formation

The third main stack in the process is now epitaxially regrown on the previously created “white space” on top of the InP substrate, while the active areas are covered in SiN_x (Figure 4.15i). The stack includes the passive layers, made up of Fe-doped InP, and the pin Photo-Diode (PD) layers on top which consist of the highly absorbing material InGaAs. This creates the passive-active butt joints that have an angled interface with respect to the waveguides defined by the SiN_x hard mask. If any p-diffusion is to be performed, it happens at this stage through an opening in the SiN_x resist.

Photo-Diode

All the epitaxy process steps are now completed, and the following fabrication steps are etching, metallization and deposition to form the functional BBs, as shown in Figure 4.16i. Before etching, a thin metallization layer consisting of Ti, Pt and Au is applied through liftoff, which forms the p-contacts and acts as a hard mask for subsequent etching steps. The first BB to be defined is the PD, starting with the etching of the passive/PD stack to form the p-mesa of the PD, while other active areas are protected

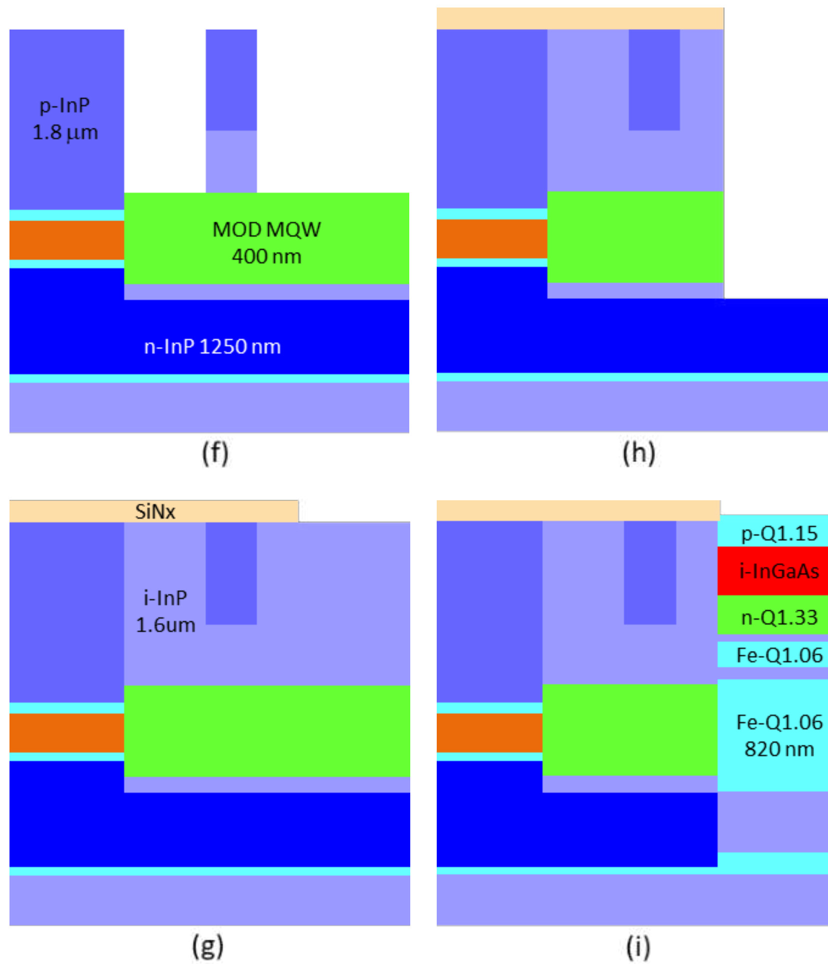


Figure 4.15: Etching isolation areas (f) and i-InP isolation regrowth before SiNx hard mask (g), etching of passive / PD areas (h) and the butt-joint regrowth of passive and PD stacks (i).

with resist. A lithography with an etching step is applied to define the n-mesa below the p-mesa of the PD and above the passive stack.

4.2.5 Contact formation and metallization

heater

The final stage of the fabrication is to create the heaters, gold contacts and interconnects. Gold is not directly deposited on InP since the latter is slightly conductive and could absorb the gold with annealing and aging. Therefore, metal is deposited after depositing a SiN_x isolation layer which covers all active BBs, except PDs and EAMs, and areas to be used for interconnects and heaters. Since the adhesion of gold on SiN_x and InP is poor, a thin layer of titanium and platinum are deposited with a lithography step before depositing the thick layer of gold. The Ti-layer is used for improving adhesion and the Pt-layer to prevent diffusion of gold into the InP or InGaAs layer. This Ti/Pt layer forms the n-contact of EAMs and also the heating element of thermally tuned BBs. These heaters are then covered with another SiN_x isolation layer. The Pt

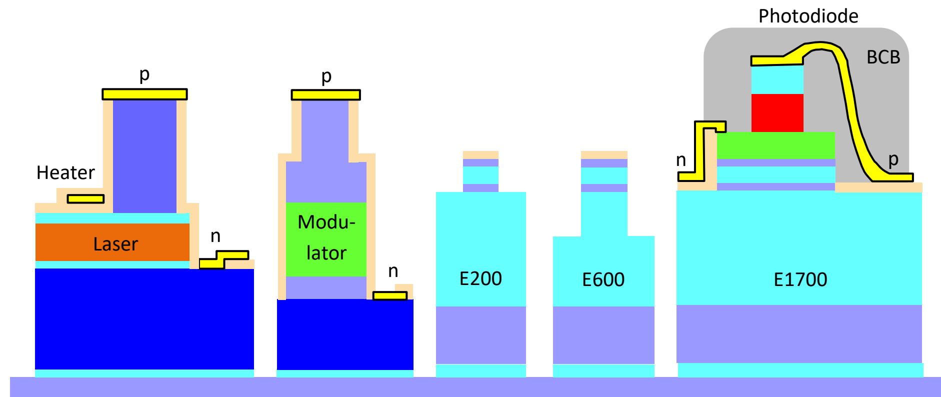


Figure 4.16: Functional Building Blocks.

sandwiched between SiN_x layers can also be used for creating capacitors and metal crossings.

Then, openings are etched trough SiN_x for the metals to contact the heaters, the n- and p-contacts of active BBs previously created and buried in SiN_x, whereas PDs however don't receive any SiN_x on top. NiCr is then deposited for the resistive elements like 50 Ω terminations.

NiCr

The fabrication process is completed with the last metallization step. It is done by first depositing a thin layer of Ti, Pt and Au on a photoresist after a lithography process, then a thick layer of gold is electroplated on the exposed thin gold. This step also creates the air bridges through a special combination of metallization steps. PDs are by now also contacted since they did not receive any SiNx in the beginning. Finally, BCB is applied to all PDs to reduce their dark noise and to all air bridges as protection.

4.3 Introduction to clean room environment

A clean environment is essential for the fabrication of PICs, just like for electronic ICs. Particles or chemical substances that stick to the surface of the wafer can be detrimental to the device performance. Particle contamination occuring during a process step, may not only affect that process step, but also the following process steps, especially for those processes that are sensitive to the surface quality, such as photoresist spin-coating and photolithography. Other sources of contamination, such as organic molecules will affect the surface properties. To minimize contamination, the PIC fabrication must be performed in a cleanroom.

cleanroom

A cleanroom is an environment that has a much lower particle count than a normal environment. The first cleanrooms were built in hospitals to control the airborne bacteria contamination. Soon after the semiconductor industry developed, engineers realized the importance of contamination control, and adopted cleanroom technology in transistor and IC fabrication. Before fresh air enters the cleanroom, it will pass through several stages of high-efficiency particulate absorbing (HEPA) filters to remove dust and other contamination. Also, the temperature and humidity in the cleanroom are strictly regulated as they will influence the process quality as well.

HEPA filter

Cleanrooms are classified using either the US Federal Standard or the ISO Standard,

Figure 4.17: View of the TU/e cleanroom with a variety of processing equipment. On the background the (yellow) lithography room.

Table 4.1: ISO 14644-1 Cleanroom Standards.

	Maximum particles/m^3						FED STD 209E
Class	≥ 0.1µm	≥ 0.2µm	≥ 0.3µm	≥ 0.5µm	≥ 1µm	≥ 05µm	equivalent
ISO 1	10	2	1				
ISO 2	100	23	10	4			
ISO 3	1,000	237	102	35	8		Class 1
ISO 4	10,000	2,370	1,020	352	83	3	Class 10
ISO 5	100,000	23,700	10,200	3,520	832	29	Class 100
ISO 6	$1 \cdot 10^6$	237,000	102,000	35,200	8,320	293	Class 1,000
ISO 7	$1 \cdot 10^7$	$2.37 \cdot 10^6$	$1.02 \cdot 10^6$	352,000	83,200	2,930	Class 10,000
ISO 8	$1 \cdot 10^8$	$2.37 \cdot 10^7$	$1.02 \cdot 10^7$	$3.52 \cdot 10^6$	832,000	29,300	Class 100,000
ISO 9	$1 \cdot 10^9$	$2.37 \cdot 10^8$	$1.02 \cdot 10^8$	$3.52 \cdot 10^7$	$8.32 \cdot 10^6$	293,000	Room air

see table 4.1. The two are interchangeable. The Federal Standard (FED STD) uses a combination of metric and English units. A class 10 (FED STD) cleanroom is defined as having less than 10 particles with diameter larger than 0.5 µm per cubic foot. This is equivalent to the ISO 4 standard. More details on the standards can be found in [108]. A modern fab for fabricating electronic IC chips needs a class 1 environment at critical regions in the cleanroom to achieve an acceptable yield. In comparison, there are more than 500,000 particles with diameter larger than 0.5 µm per cubic foot inside a reasonably clean house. Typically, for chip assembly and some less-critical fabrication processes, class 10,000 is a popular choice since it is relatively easy to obtain.

For more critical processes, at least class 100 is required. Good filtering techniques and protective clothing for personnel are needed to obtain class 100. For high volume production, class 10 or even class 1 is needed in order to achieve a reasonable yield. Rooms, where photoresist is used, are illuminated with yellow light so that resist can be handled freely without being exposed. These areas often require chemical filtering to prevent contamination of chemically amplified resists.

4.4 Substrate and wafer cleaning

Many process steps are extremely sensitive to surface contamination. Therefore, wafer cleaning prior to critical process steps is very important. In this section we give a brief overview where wafer cleaning is required.

Surface cleaning **Surface cleaning before processing.** During epitaxial growth, besides the functional layers (e.g., p and n contacts and waveguide core), protective layer(s) will be grown at the end. For instance for the active-passive layer stack shown in Section 4.1, usually an additional InP cap layer, in the order of 50 nm, is grown on top of the p-doped InGaAs contact layer. This is to protect the functional stack from oxidation in the air or contamination due to particles. Before the processing starts, the cap layer will be removed using selective wet etching (HCl in case of an InP cap layer), which removes not only the cap layer itself, but also any thin oxidized materials and particles on the surface.

Organic residuals **Cleaning of process residuals.** Organic residuals can occur on the wafer surface after certain processes. Examples include the polymer residuals after CH_4/H_2 based plasma etching for InP and CHF_3-based plasma etching for SiN_x/SiO_x dielectrics. Those organic residuals will hinder the processing that follows by acting as micromask, source of roughness and they can even block the process completely. An effective way to remove the residuals is an O_2 plasma-ashing process, which is highly isotropic and effective to remove residuals from sidewalls and corners.

native oxides **Cleaning of native oxides on surface.** It is common in III-V processing that the semiconductor surface is oxidized. Even at room temperature and in air, a monolayer of III-oxide (such as In_2O_3 on InP surface) can be formed by reacting with O_2 in the air. Oxidation will also occur when the wafer is exposed to O_2-containing plasma processes. To restore a good semiconductor surface, a short H_3PO_4-acid treatment effectively removes the oxides without damaging the semiconductor crystal.

DI-water **DI-water cleaning step.** Cleaning steps using chemicals are usually finished with a cleaning step in a DI-water bath to remove residues of the chemicals. DI-water stands for de-ionized water: ultrapure water which leaves no residues on the wafer after drying. A DI-water bath is connected with continuous flow of DI water, the resistivity of which is constantly monitored. The wafer, after chemical treatment, should be immediately rinsed in the DI water bath. The wafer can only be taken out when the resistivity reading of the water increases to its original level (i.e., the stable resistivity of running DI water, typically $8\,M\Omega \cdot cm$ or above). After this cleaning step the wafer will be dried by a Nitrogen flow or by a spin-dry step.

4.5 Introduction to vacuum

4.5.1 Concept of vacuum

vacuum Most of the process tools rely on vacuum techniques. A high-vacuum environment is required for epitaxy and dry etching processes. Even for a simple process such as spin-coating, vacuum is essential to clamp the wafer tightly on the chuck. The word

Table 4.2: Values of molecular density n, molecular incidence rate ϕ, mean free path λ , and time τ to form a monolayer, as a function of pressure P, for air at 25°C.

P (Torr)	n (molec/cm^3	ϕ (molec/cm$^2\cdot$sec)	λ (cm)	τ (sec)
760	$2.46\cdot10^{19}$	$2.88\cdot10^{23}$	$6.7\cdot10^{-6}$	$2.9\cdot10^{-9}$
1	$3.25\cdot10^{16}$	$3.78\cdot10^{20}$	$5.1\cdot10^{-3}$	$2.2\cdot10^{-6}$
10^{-3}	$3.25\cdot10^{13}$	$3.78\cdot10^{17}$	5.1	$2.2\cdot10^{-3}$
10^{-6}	$3.25\cdot10^{10}$	$3.78\cdot10^{14}$	$5.1\cdot10^{3}$	2.2
10^{-9}	$3.25\cdot10^{7}$	$3.78\cdot10^{11}$	$5.1\cdot10^{6}$	$2.2\cdot10^{3}$
10^{-12}	$3.25\cdot10^{4}$	$3.78\cdot10^{8}$	$5.1\cdot10^{9}$	$2.2\cdot10^{6}$
10^{-15}	$3.25\cdot10^{1}$	$3.78\cdot10^{5}$	$5.1\cdot10^{12}$	$2.2\cdot10^{9}$

Table 4.3: Values of ϕ, λ and τ for various gases at 25°C and 10^{-3} Torr.

Gas	ϕ (molec/cm$^2\cdot$sec)	λ (cm)	τ (sec)
H_2	$14.4\cdot10^{17}$	9.3	$1.0\cdot10^{-3}$
He	$10.4\cdot10^{17}$	14.7	$2.3\cdot10^{-3}$
N_2	$3.85\cdot10^{17}$	5.0	$2.1\cdot10^{-3}$
O_2	$3.60\cdot10^{17}$	5.4	$2.4\cdot10^{-3}$
A	$3.22\cdot10^{17}$	5.3	$2.6\cdot10^{-3}$
Air	$3.78\cdot10^{17}$	5.1	$2.2\cdot10^{-3}$
H_2O	$4.80\cdot10^{17}$	3.4	$1.1\cdot10^{-3}$
CO_2	$3.07\cdot10^{17}$	3.3	$1.7\cdot10^{-3}$

vacuum originates from the Latin word "vacuus", which means empty. However practically applied vacuum is far from being empty without any matter. At the lowest pressures which can be obtained by modern pumping methods there are still hundreds of molecules per cubic cm of evacuated space. These residual molecules will stick to exposed surfaces and will influence the surface properties. The degree of vacuum can be categorized into low, medium, high and ultrahigh vacuum. Each category of vacuum has specific kinds of pumps and measuring instruments. In order to describe the various vacuum ranges, it is useful to first define the following concepts:

- Molecular density n: the average number of molecules per unit volume. *molecular density*
- Mean free path λ: the average distance that a molecule travels in a gas between two successive collisions with other molecules of that gas. *mean free path*
- Time constant τ to form a monolayer: the time required for a freshly cleaved surface to be covered by a layer of the gas of one molecule thickness. This time is given by the ratio between the number of molecules required to form a compact monolayer (in the order of 10^{15} molecules/cm^2) and the molecular incidence rate at which molecules strike a surface.

Table 4.2 lists the values of these parameters for different pressure levels of air. Table 4.3 provides an overview of the behaviour for several commonly seen gases. The unit of pressure often used in semiconductor processing is Torr. It is related to other units of pressure according to 1 atm = 1.013 bar = 760 mm Hg = 760 Torr = 101,325 Pa *pressure units*

Vacuum ranges and their physical characteristics are:

- Low and medium vacuum: the number of the molecules in the gas phase is large compared to that covering the surfaces (chamber walls). The pumping is di- *low and medium vacuum*

rected towards rarefying the existing gas phase. Pressure extends from atmospheric to 10^{-2} Torr. Typically used in tools for fixing or handling wafers.

high vacuum

- High vacuum: the gas molecules are located principally on surfaces. The mean free path is in the order of or greater than the dimensions of the chamber. The pumping is evacuating or capturing the gas molecules leaving the surfaces and individually reaching the pump. Particles can travel in the enclosure without colliding with other particles. Pressure ranges from 10^{-3} to 10^{-7} Torr. Typically used for plasma deposition and etching, evaporation and sputtering tools.

ultrahigh vacuum (UHV)

- Ultrahigh vacuum (UHV): is equal or larger than the usual time for laboratory measurements, in the order of hours or days, thus clean surfaces can be prepared and studied before the adsorbed gas layer is formed. Pressure extends from 10^{-9} to 10^{-16} Torr. Used for high-precision epitaxy such as in MBE and ALD tools.

Note that the gas composition changes while the total pressure is decreasing (when pumping air). In low or medium vacuum the gas resembles the atmospheric composition while in high vacuum it contains mainly water vapour (70-90%) coming from the surfaces (chamber walls). As pumping is continued and proper baking/heating is applied, the CO ratio increases and finally in UHV H_2 becomes dominant coming mostly from the bulk materials (permeation).

4.5.2 Boiling point and vapour pressure

During processing, for instance etching and deposition, it is important to understand the behaviour of the gases and by-products in the chamber and their impact on the quality of the fabricated structures. An important measure for this behaviour is the boiling point. The boiling point of a liquid is defined as the temperature at which the vapor pressure of the liquid is equal to that of the surrounding pressure. The vapor pressure of a given substance at a given temperature is thus the pressure at which the substance becomes gaseous.

boiling point
vapor pressure

partial pressure

In gas mixtures the partial pressure of one gas is the pressure of that gas when considered alone in the same enclosure (volume). Hence the total pressure of the gas mixture is equal to the sum of the partial pressures of all constituents. The first example is trimethylindium, $In(CH_3)_3$ or TMI, used as a precursor in MOVPE for growing InP and other quaternary and ternary compounds. It is crucial to have a precise vapour pressure model of TMI so that an accurate deposition efficiency can be achieved at a particular growth condition. Knowing the exact value of the vapor pressure of each of the precursors is even more crucial to the accurate composition in the solid compound [109]. A second example is the dry etching of InP. A Cl_2-based gas mixture is widely used for dry etching of InP and related compound materials. However, one of the by-products, $InCl_3$, is relatively non-volatile. If it cannot be removed efficiently from the sample surface, indium-containing droplets or islands will occur and eventually block the etch and cause a rough surface [110]. Its boiling point at 1 atmosphere is as high as 600 °C [111]. Even at a vacuum pressure of a few milliTorr, typically found in a plasma etch tool, its boiling point is still about 160 °C [112]. Therefore the Cl_2 based dry etching processes typically operates at 160-200 °C.

4.5.3 Vacuum pumps

As mentioned earlier, various pumping techniques are used to achieve different degrees of vacuum. For high vacuum a fore pump is required before the high-vacuum

Table 4.4: Comparison of several pumps for semiconductor processing tools.

Type	Principle	Degree of vacuum	Fore pumping required?	Suitable as fore pump?
Rotary	Positive displacement	Medium 10^3 to 10^{-2} mBar	No	Yes, but not oil-free
Scroll	Positive displacement	Medium 10^3 to 10^{-2} mBar	No	Yes
Turbo molecular	Momentum transfer	High to UHV 10^{-3} to 10^{-11} mBar	Yes	-
Cryo	Entrapment	High to UHV 10^{-3} to 10^{-11} mBar	Yes	-

pump because high-vacuum pumps cannot start at atmospheric pressure. In Table 4.4, several common types of pumps are briefly explained and compared.

4.6 Substrates and epitaxial growth

4.6.1 Substrates

wafers

Optical and optoelectronic devices are built on high-quality mono-crystalline wafers. The material must withstand very high (> 10 GWatt/cm^2) optical power density and > 10 kA/cm^2 current density. Any crystal grain boundaries in a poly-crystalline material or a defected wafer will cause significant optical absorption and nonradiative carrier recombination, both of which are detrimental for many devices. It should be emphasized that the quality of the functional layers grown on top of the substrate is in best case as good as the quality of the substrate. The substrate also acts as a mechanical base for everything created on it.

substrate crystal

The process flow for the substrate production is shown in Figure 4.18. First a large single crystal (e.g., InP) is grown from a melt (see Section 4.6.2). It is called boule or ingot and it has a cylindrical shape and a narrow tip due to the growth process. The end parts are cut and it is grinded to achieve a round shape. It is then sliced into wafers, followed by an edge rounding process. The sharp edges resulting from the slicing are grinded into rounded corners, so that the substrates will not easily crack from the sharp edges during handling. Each substrate will be laser-marked with a unique serial number, followed by a polishing step to achieve an atomically smooth surface. Single-side or double-side polished substrates are both common. Finally the substrates are characterized and inspected before they are packed in boxes. The first step of creating the ingot is the most critical step as it determines the quality as well as the diameter of the substrate (e.g., 3 or 4 inch for InP and 200-300 mm for Si).

4.6.2 Substrate growth techniques

In this section the technique used to grow InP is briefly described. Note that all III-V compounds and Si have Zinc Blende lattice structure (see details in Chapter 3). The growth methods of III-V (InP, GaAs, etc.) materials can be categorized into Bridgman and Czochralski growth. Bridgman methods are widely used for fabricating monocrystalline InP ingots.

Bridgman methods

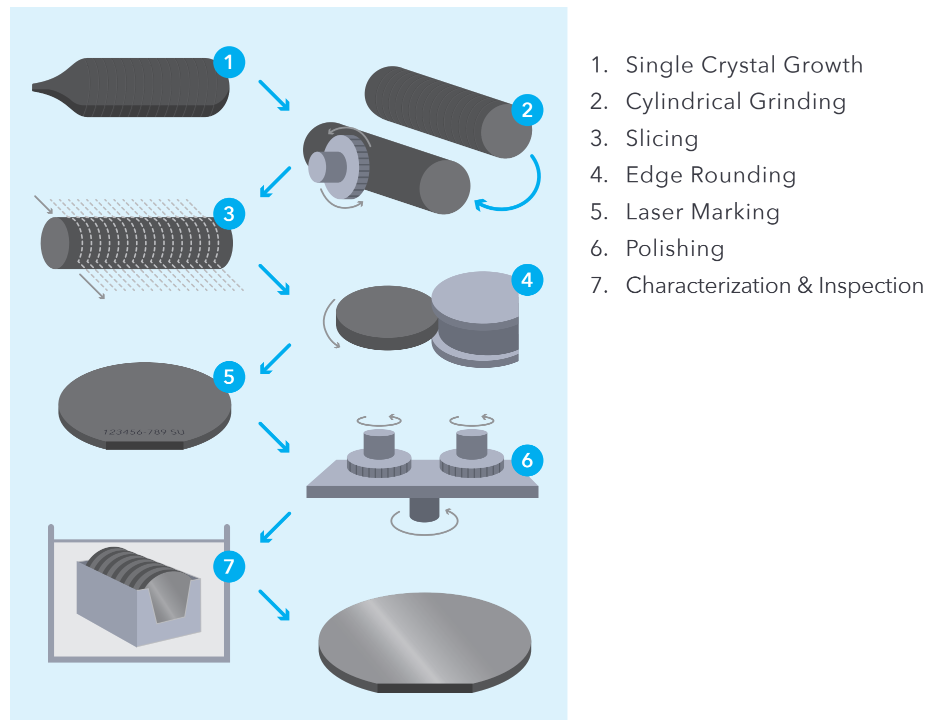


Figure 4.18: Process flow of substrate production.

The schematic of the Vertical Bridgman (VB) method is shown in Figure 4.19. The growth starts with a polycrystalline InP ingot made with relatively easy and low-cost methods. It is melted in a crucible as indicated in the figure. The furnace has a temperature profile as shown in Figure 4.19(B), with the highest temperature in the center of the furnace and a significantly lower temperature at both ends. The crucible has a narrow end at the bottom. The furnace is controlled such that in the center the temperature is beyond the melting point of InP and at the lower end below the crystallization point. When the crucible is moved down crystallization will start in the narrow part at the bottom and proceed to the higher part of the crucible if it is moved sufficiently slow. Special measures are taken to prevent evaporation of phosphorus from the melt. More sophisticated methods use multi-section furnaces with an accurately controllable temperature gradient, so that the crystallization can be controlled by moving the temperature gradient instead of the crucible (Vertical Gradient Freezing VGF). More information about Bridgman methods can be found in Dutta [113]

Czochralski methods

Another widely used growth method is the Czochralski (CZ) method. In CZ methods the crystal is directly pulled from the melt in vertical direction, as shown in Figure 4.20, thus it is not in contact with the crucible. However for InP growth, P overpressure is required. To avoid decomposition of the melt, liquid encapsulation using B_2O_3 is used, leading to the liquid encapsulated Czochralski (LEC) method [114]. A further improved version is called vapor-pressure controlled Czochralski (VCZ) where an inner chamber is built to isolate the growing crystal from the water-cooled outer wall, thus significantly improving the temperature uniformity inside the growing crystal [115].

Currently the growth of InP ingots mainly uses LEC, VCZ or VB/VGF methods[113]. Their characteristics are compared in the Table 4.5 below. Since VB/VGF methods have more precise control over the temperature across the melt and crystal, typically they

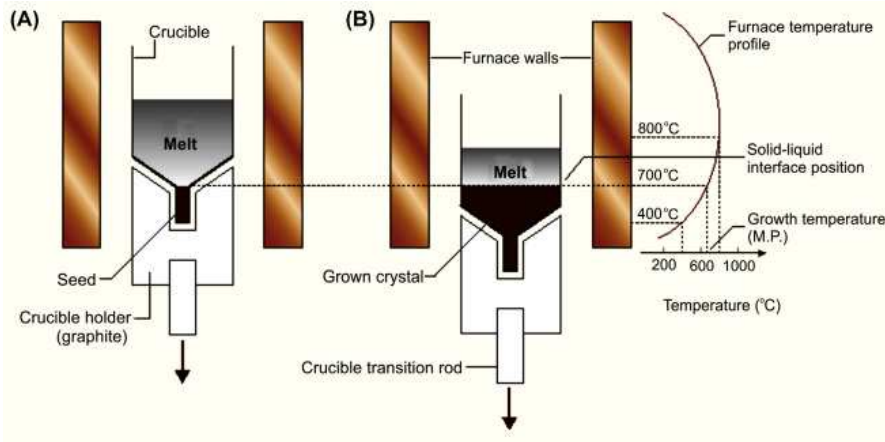


Figure 4.19: Schematic diagram of a vertical Bridgman crystal growth process in a single-zone furnace: (A) in the hot zone at the beginning of the experiment, and (B) while moving through the cold zone with partially grown crystal.

Table 4.5: Qualitative comparison of LEC, VCZ and VB methods for InP ingot growth.

	LEC	VCZ	VB
Defect density	High	Low	Very low
Residual strain	High	Medium	Low
Material loss	Large	Large	Small
Throughput	High	High	Low

produce superior crystal quality in terms of defect density and strain. On the other hand, LEC/VCZ methods are low cost and high throughput, suited for volume manufacturing of relatively large-sized wafers.

4.6.3 Characteristics of substrates

It is important to choose the right substrate for different kinds of applications. The main characteristics of a substrate are summarized below:

- Size: indicated as wafer diameter. Typical sizes are 3 or 4 inch for InP, 6 inch for GaAs and 300 mm for Si. *wafer diameter*
- Surface orientation: indicating the crystallographic plane corresponding to the wafer surface. (100) and (111) are common for InP. The surface orientation is very important for the directional properties of the epitaxial growth and for the directional etch rate of selective wet-etching processes. The surface orientation is usually slightly offset from the ideal crystal plane (< 1°). This is to avoid forming of large terraces of several mono-layers thick due to surface diffusion during the growth. The small offset angle avoids the formation of such large terraces. *surface orientation*
- Doping: can be n-type, p-type, undoped and/or semi-insulating. The concentration is specified in doping atoms per cm^3 (or cm^{-3}). The choice of dopant depends on the substrate material. Zinc is used as p-dopant for both InP and GaAs, and S or Sn is used as n-dopant for InP. Semi-insulating InP substrates are formed by incorporating Fe doping. *doping*

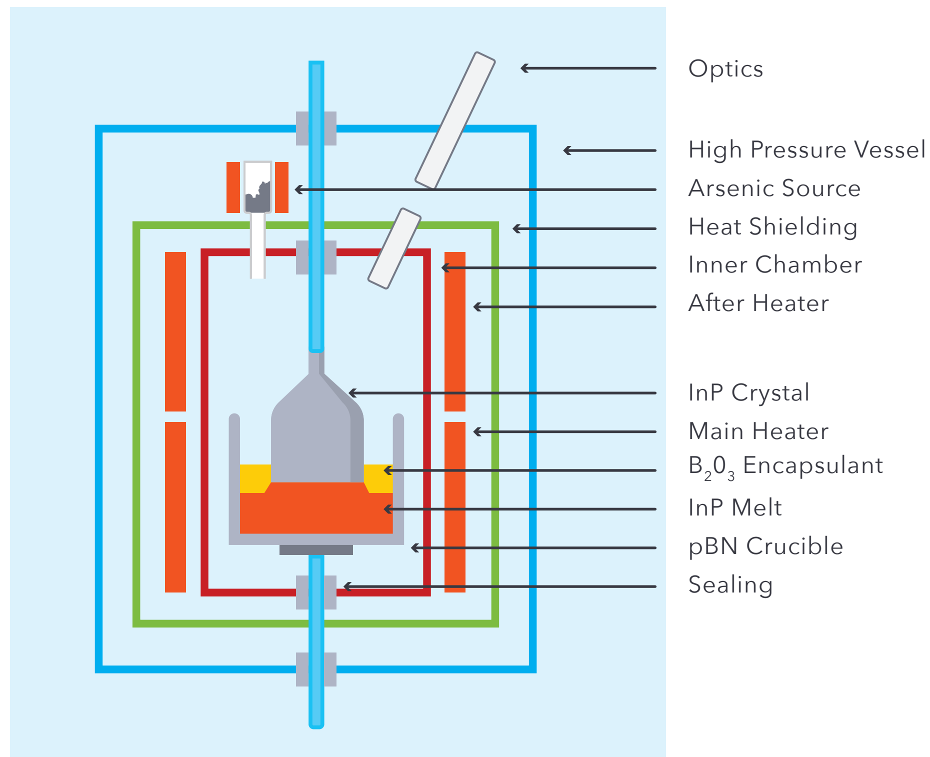


Figure 4.20: Schematic diagram of a Czochralski crystal growth reactor.

surface finish

- Surface finish: the substrates can have one side or both sides polished. However even for double-side polished wafers, two sides have different qualities and only one side can be used for epitaxy and processing.

total thickness variation (TTV)

- The flatness of substrates is important as it will impact the quality of the following processes such as epitaxy and lithography. It is usually specified as the Total Thickness Variation (TTV), the difference between the highest and the lowest point of the wafer, referenced to the back side.

4.7 Epitaxial growth

4.7.1 Vegard's law

epitaxy *lattice-matched* *bandgap engineering*

Epitaxial growth is usually the first step of any process flow. Epitaxy is defined as growth of a crystalline layer with the same crystallographic structure as the substrate, in other words, growth of lattice-matched crystal on a crystalline substrate. III-V semiconductors enable a powerful variety of layers with different electronic and optical properties by varying the composition of the crystals, while keeping the lattice constant unchanged. It is often called bandgap engineering, because with the composition we can vary the bandgap wavelength from 0.9 µm for InP to 1.65 µm for InGaAs lattice-matched to the InP substrate. In this section, the theory of lattice matching and bandgap engineering is introduced.

lattice constant

The material bandgap and the lattice constant are indicated in Figure 4.21 for a number of III-V semiconductors with varying compositions. The dots indicate the bandgap

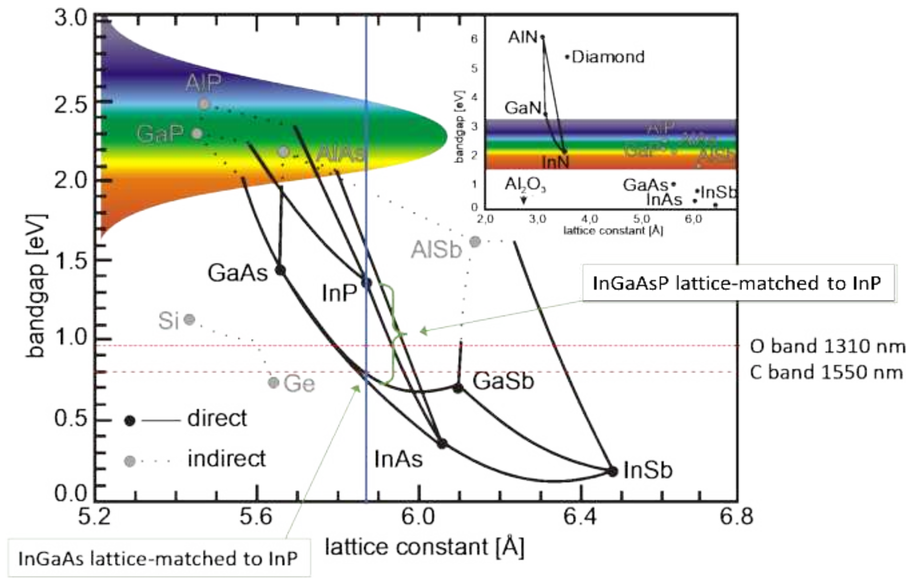


Figure 4.21: Material bandgap vs. lattice constant.

and lattice constant of the most important binary compounds, the black lines between them indicate how the lattice constant and the bandgap changes if we mix them in varying concentrations. For InP we can go either to GaP by replacing a fraction x of the In atoms by Ga atoms, or to InAs by replacing a fraction y of the P atoms by As atoms. The black lines depict how the bandgaps and lattice constants of $In_{1-x}Ga_xAs$ and $InAs_yP_{1-y}$ change as a function of x and y.

binary compound

A line of particular interest is the curve between GaAs and InAs, which describes the ternary compound $In_xGa_{1-x}As$. Its bandgap is dependent on the compositional ratio between In and Ga. For $x = 0.53$, the material $In_{0.53}Ga_{0.47}As$ has the same lattice constant (so called lattice matched) as InP. The vertical line between InP and $In_{0.53}Ga_{0.47}As$ is most interesting as it forms the quaternary compound $In_xGa_{1-x}As_yP_{1-y}$. Along this line we can achieve a wide range of bandgaps and at the same time lattice matching to InP. Lattice matched InGaAsP covers the two most important wavelength bands in photonics and optical communication: the O band and the C band, as indicated with the red lines in the plot.

ternary compound

The material composition for ternary and quaternary compounds can be derived from Vegard's law. The actual lattice constant of a ternary or quaternary compound is a weighted average of the lattice constants of the constituent binary crystals. For ternary $In_xGa_{1-x}As$, the lattice constant is

Vegard's law
quaternary compound

$$a_{\mathrm{InGaAs}} = x a_{\mathrm{InAs}} + (1-x) a_{\mathrm{GaAs}} \tag{4.1}$$

or

$$x = (a_{\mathrm{InGaAs}} - a_{\mathrm{GaAs}})/(a_{\mathrm{InAs}} - a_{\mathrm{GaAs}}) \tag{4.2}$$

For lattice matched InGaAs $a_{\mathrm{InGaAs}} = a_{\mathrm{InP}} = 5.8687$ Å, and with $a_{\mathrm{GaAs}} = 5.6533$ Å, $a_{\mathrm{InAs}} = 6.0583$ Å, and $a_{\mathrm{GaP}} = 5.4505$ Å, we find $x = 0.53$. Similarly for $In_xGa_{1-x}As_yP_{1-y}$, its lattice constant is

$$a_{\mathrm{InGaAsP}} = xy a_{\mathrm{InAs}} + x(1-y) a_{\mathrm{InP}} + (1-x) y a_{\mathrm{GaAs}} + (1-x)(1-y) a_{\mathrm{GaP}} \tag{4.3}$$

Problem: Calculate the material composition for InGaAsP quarternary materials Q1.05 and Q1.3. Note that 1.05 and 1.3 corresponds to the bandgap equivalent wavelength of those compounds in units of μm.

Solution: For Q1.05 we find $x = 0.89$ and $y = 0.23$, and for Q1.3 $x = 0.72$ and $y = 0.59$.

Problem 4.1: Vegards Law.

For lattice-matched $In_xGa_{1-x}As_yP_{1-y}$ the lattice constant should equal that of InP (5.8687 Å) for all combinations of x and y. This results in a relation between x and y:

$$x = \frac{(0.4174 - 0.2021y)}{(0.4174 - 0.0123y)} \tag{4.4}$$

For lattice matched $In_xGa_{1-x}As_yP_{1-y}$, the bandgap depends on y according to

$$E_g = (1.344 - 0.738y + 0.138y^2)\ \text{eV} \tag{4.5}$$

and the corresponding wavelength follows from

$$\lambda_g = 1.24/E_g\ (\mu m) \tag{4.6}$$

Q1.25 *Q1.55* If we want to grow transparent waveguide material with a bandgap of 1.25 μm (Q1.25), we can solve y from Equation 4.5 and x from Equation 4.4. For this material we find $x = 0.76$ and $y = 0.53$. For Q1.55 material we find in the same way $x = 0.59$ and $y = 0.88$. With PhotoLuminescence (PL) and X-Ray Diffraction (XRD) we find $x = 0.76$ and $y = 0.53$ for Q1.25, and $x = 0.59$ and $y = 0.88$ for Q1.55 material [3].

4.7.2 MOVPE

MOVPE *MOCVD* *PVD* *trimethyl gallium* *trimethyl indium* *arsine* *phosphine*

This section focuses on metalorganic vapor-phase epitaxy (MOVPE), also called metalorganic CVD, MOCVD) It is the most widely applied growth technique for InP PICs. Vapor-phase epitaxy is usually classified by the transport mechanism of the gaseous species: physical-vapor deposition (PVD), or chemical-vapor deposition (CVD). CVD is often further classified according to the chemistry of the source gases, such as metalorganic CVD (MOCVD), chloride VPE (ClVPE), and hydride VPE (HVPE). In MOVPE, III-species are room temperature stable metalorganic compounds, e.g., trimethyl gallium, $(CH_3)_3Ga$ and trimethyl indium, $(CH_3)_3In$, while V-species are normally hydrides such as arsine (AsH_3) or phosphine (PH_3). The major advantages of MOVPE over other techniques are the suitability for large-scale production and its versatility [116].

A basic MOVPE reactor is illustrated in Figure 4.22. The group-V gas precursors are directly injected into the reactor. For group-III precursors, the metalorganic-IIIs are liquid at room temperature. Therefore to generate their vapor, a flow of H_2 is passed through a bubbler containing the liquid. The H_2 gas flow out of the bubbler will carry

[3] The values may vary slightly based on the reference used. In practice, the bandgap wavelength and the lattice mismatch are verified with PhotoLuminescence (PL) and X-Ray Diffraction (XRD), respectively.

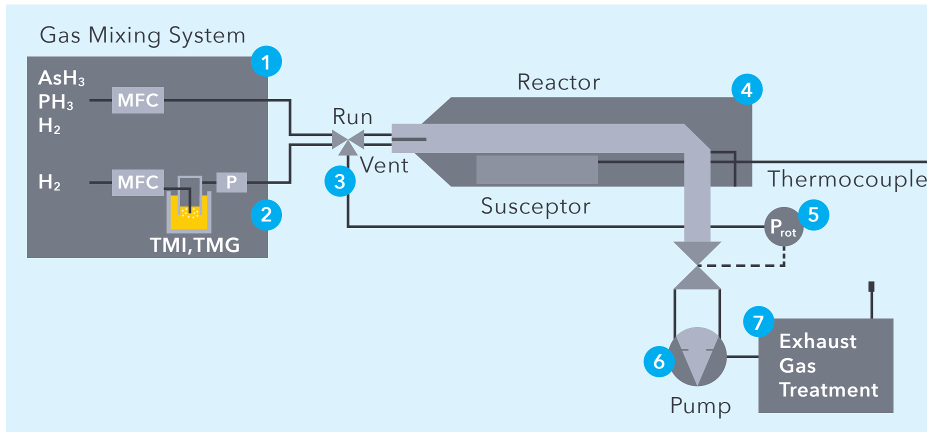


Figure 4.22: Schematic diagram of an MOVPE reactor.

saturated vapor of the desired metalorganic-III compound. All gas flows are controlled precisely by mass flow controllers (MFCs), before they are mixed and injected into the reactor chamber. The dissolution of the precursors in the hot zone above the wafer leads to deposition of the growth materials and results in the growth of a new epitaxial layer.

Depending on the growth temperature, the MOVPE process may fall in one of the following three regimes [117]. (see Figure 4.23).

- A: Desorption or gas phase limited growth. High temperature causes the elements to desorb from the surface, leading to a decrease in growth rate.
- B: Mass transport limited (diffusion limited) growth. The growth rate is nearly temperature independent over a wide range of temperatures. Thermodynamic equilibrium is achieved at the interface. Growth in this regime is highly controllable and reproducible.
- C: Surface kinetic limited growth. At relatively low temperature, the reaction kinetics dominate the growth rate, which decreases as the temperature decreases.

In the MOVPE chamber, the wafers sit on a susceptor, which controls the temperature of the wafer. The susceptor is usually made of graphite, which conducts heat but does not react with the precursors. It should be noted that the actual temperature of a substrate depends on its doping, even at a fixed reactor temperature. For instance, the temperature of an n-type InP substrate can be more than 10°C different from an SI InP substrate.

4.7.3 MBE

MBE (molecular beam epitaxy)

So far, InP generic integration processes are fully based on MOVPE. However, for some applications Molecular Beam Epitaxy (MBE) is used. The concept of an MBE reactor is illustrated in Figure 4.25. Here beams with different species are provided by thermal heating in effusion cells. The MBE process takes place in ultra-high vacuum (UHV, typical 10^{-8}–10^{-12} Torr). The slow but highly controllable deposition rate enables growth of nano-scale layers such as quantum wells and quantum dots with high precision,

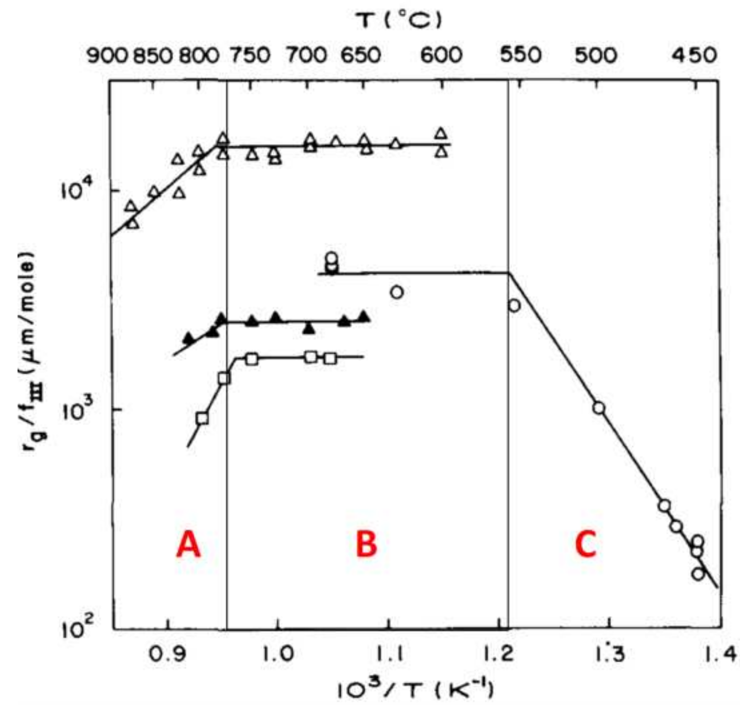


Figure 4.23: MOVPE growth regimes.

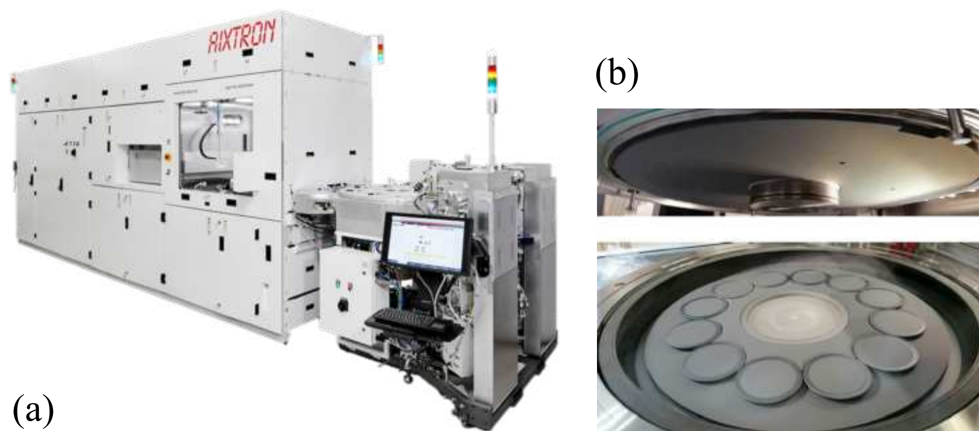


Figure 4.24: (a) Modern MOVPE reactor and (b) closer view of the reactor chamber with rotating multi-wafer susceptor and top lid with shower head.

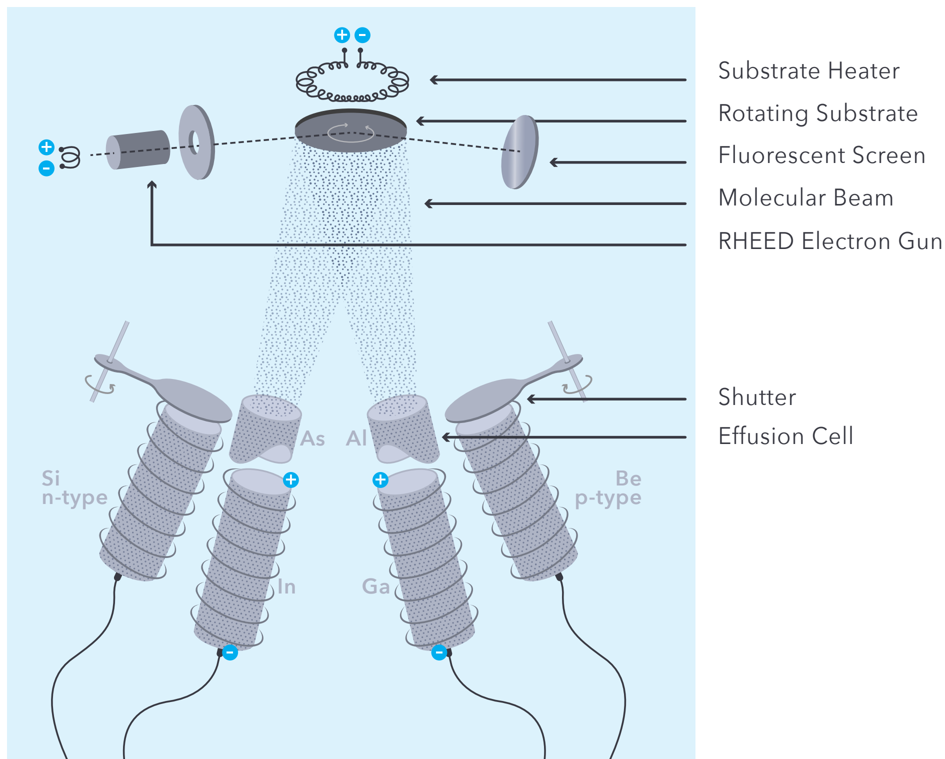


Figure 4.25: Illustration of an MBE reactor.

and the relatively low deposition temperature enables the creation of very large doping gradients, as used, for example, in Avalanche Photo Diodes. The absence of carrier gasses and the use of a UHV environment contributes to the highest achievable material purity. Therefore, it is mainly used in research facilities and sometimes in production for some critical layers.

doping gradient

In the reactor, each element is contained and heated in a so-called effusion cell, in ultra-pure form. The gaseous element will effuse and eventually condense on the wafer surface. In case a compound is needed, multiple effusion cells operate at the same time. Multiple elements will condense on the surface and react with each other to form the compound. Doping of the material follows the same principle. The wafer is typically rotated during the growth process to achieve high uniformity. The growth is monitored by Reflection High-Energy Electron Diffraction (RHEED). The control of layer thickness can be as precise as a single atom layer.

effusion cell

RHEED

4.7.4 Characterization of epitaxy quality

There are many tools and techniques to characterize different aspects of the epitaxy. The most important quality aspects and their characterization methods are summarized in Table 4.6. To characterize the morphology, different types of microscopes can be used, for instance optical microscope, or scanning or transmission electron microscopes.

Conventional optical microscopes are limited to the diffraction limit of light (~200 nm). However by using interference and acquiring light path length difference in the sample,

SEM *TEM* one could achieve resolution and contrast similar to that of a SEM (Scanning Electron Microscopy. TEM (Transmission Electron Microscopy) is used when atomic level resolution is needed. Benefitting from the extremely small de Broglie wavelength of electrons, SEM and TEM work in a similar way as an optical microscope, but with much better resolution.

SIMS *Hall Effect* The doping profile of a semiconductor sample can be measured by several methods. The Capacitance-Voltage (C-V) method measures the capacitance of a diode with varying applied voltage. The capacitance is dependent on the depletion width, thus providing insight into the material doping profile. However the measurement cannot be performed across the intrinsic region (e.g., from p region through to n region), as a diode junction is required to remain existing during the measurement. SIMS (Secondary Ion Mass Spectrometry) provides an alternative way to extract doping profile and it applies to any structure. It sputters the sample surface with a focused ion beam and collects and analyses the elements that come off the sample surface. Therefore it can measure the concentration of all elements, not only the dopants. A third method not so regularly used in daily practice, is the Hall Effect [118] that can be used to measure majority carrier concentration and mobility in semiconductor.

photoluminescence *cathodoluminescence* When a material is grown (e.g., a bulk InP or an InGaAsP quantum well (QW) layer), it is convenient to characterize its bandgap to confirm the quality of the growth. Photoluminescence and cathodoluminescence are common methods. The main difference is the use of higher-energy photons or electrons to excite carriers in the material, which, after relaxation, will emit new photons corresponding to the direct bandgap. Photoluminescence focuses on a specific transition path since the photon excitation is narrow band (due to the nature of the laser), while cathodoluminescence can probe on multiple possible transition paths by directly injecting free electrons into the material.

XRD Furthermore, X-ray diffraction (XRD) is widely used to examine the crystallography of the grown material. By examining the behaviour (e.g., intensity and angle distributions) of electrons diffracted by the crystal lattice, one can reconstruct the assembly of atoms in the crystal. Therefore it is very important tool to analyze crystal lattice, strain and doping.

4.8 Optical Lithography

4.8.1 Introduction

lithography *optical lithography* *photolithography* Lithography, as used in the manufacturing of ICs and PICs, is the process of transferring geometric shapes on a mask to the surface of a semiconductor wafer. These shapes make up the parts of the circuit, such as waveguides, electrodes, contact windows, metal interconnections, and so on. Although most lithography techniques used today were developed in the past 40 years, the process was actually invented in 1796 by Alois Senefelder. In this first process the pattern or image was transferred from a stone plate (lithos). Nowadays, optical lithography has become the most critical (and the most expensive) step in the semiconductor process. In this chapter, the terminologies of optical lithography and photolithography are used interchangeably. After circuit design and simulation, the first step in fabricating an IC or PIC is to generate the pattern of geometric shapes. A composite drawing of the circuit is broken into mask layers for subsequent processing steps: e.g. waveguides on one level, electrodes on another, and so on. These are called mask layers. Interactive graphic displays and digitizers convert the geometrical layout to digital data, which is used to fabricate a set of photomasks. The final IC or PIC is made by sequentially transferring the features from each mask,

Table 4.6: Common methods for the characterization of epitaxy quality.

Quality aspect	Method	Feature
Morphology	Optical microscopy (Nomarski interference contrast microscopy)	~1000× magnification, high contrast
	Scanning Electron Microscopy (SEM)	~10,000× magnification, high resolution
	Transmission Electron Microscopy (TEM)	Åresolution, crystallographic features
Doping	C-V measurement	Carrier concentration and mobility of majority carrier
	Secondary ion mass spectroscopy (SIMS)	Concentration of all elements
	Hall effect	Concentration and mobility of majority carrier
Bandgap	Photoluminescence	Direct transitions dominated
	Cathodoluminescence	Excite multiple transition paths
Crystallography	X-Ray diffraction (XRD)	Thickness, composition, doping and strain of the crystal

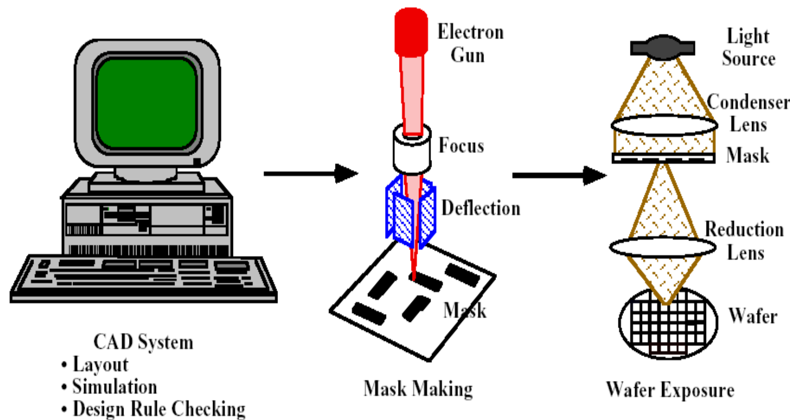


Figure 4.26: Lithography pattern transfer process.

level by level, to the surface of the wafer. This is illustrated in Figure 4.26. Between each successive image transfer one or more processing steps will take place, such as wet or dry etching, passivation and/or metallization.

4.8.2 Contact and proximity lithography

Contact lithography and proximity lithography both belong to the category of 1:1 exposure systems, which means that the patterns obtained on the semiconductor wafer have the same size as the original ones on the mask. The principles of contact and proximity lithography are illustrated in Figure 4.27(a). In both methods a highly uniform flux of light passes through the optical mask and shines on a thin layer of pho-

contact lithography

proximity lithography

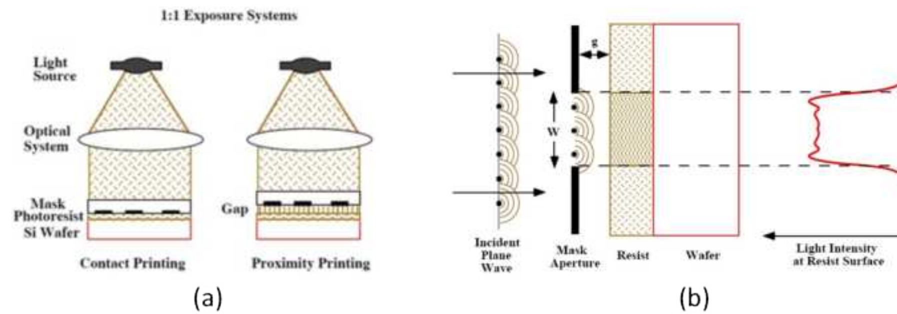


Figure 4.27: (a) Illustration of contact and proximity lithography. (b) Illustration of near-field diffraction the 1:1 exposure system.

toresist coated on the wafer. The major difference in the two methods is the distance between the mask and the wafer.

Proximity printing consists of placing the mask in close proximity but not in direct contact with the wafer. The diffraction that occurs at pattern edges causes the light to diverge. This divergence depends on the spacing between the mask and the sample. Typical spacing is a few to tens of microns. Contact printing is a special case of proximity printing. The mask is placed in proximity to the sample for alignment; then it is clamped directly against the sample for exposure. During exposure, the mask is in direct contact with the surface of the photoresist layer. The resolution of such systems can be estimated with Fresnel diffraction theory, since the wafer is placed at the near field of the mask aperture, as illustrated in Figure 4.27(b). Due to the presence of diffraction, the image on the wafer surface is not perfect, showing non-abrupt transitions and oscillation of the intensity at the edge of the image. In this approach, the minimum resolution (minimum feature size of the aperture on the mask) is related to the wavelength of the light and the spacing between mask and wafer [119]:

resolution

$$W_{\min} \approx \sqrt{\lambda g'}, \text{ when } \lambda < g' = (g + t/2) < W^2/\lambda \tag{4.7}$$

where W is the aperture width, λ the wavelength of the light, g' the effective spacing which is the spacing g between mask and wafer plus half of the photoresist thickness t. It can be seen that the resolution can be improved by reducing the wavelength (see Section 4.8.4), the spacing g as well as the photoresist thickness t. For instance, in a proximity lithography tool with $g = 10\,\mu$m, a photoresist thickness of 400 nm and a wavelength of 365 nm, the resolution is approximately 2 μm. The resolution can be improved to about 660 nm by reducing the gap to 1 μm. Note that a perfect contact lithography (g = 0) is nearly impossible to achieve due to non-flatness of the wafer surface (e.g. TTV of the wafer, non-uniformity of photoresists, etc).

Advantage/Disadvantage Practically, the resolution of a 1:1 exposure system is limited to about 500 nm, when using common i-line UV light (365 nm wavelength). Higher resolution can be achieved using deep ultraviolet (DUV) light, however it is not commonly seen in contact lithography. In contact printing, the direct contact between mask and sample can damage the mask either directly or from the frequent cleaning required to remove photoresist residues attached to the mask. This makes the technique more suitable for low-end tasks with low resolution patterns (> 1 μm) and lab R&D. Due to the relatively simple optics, light source and mechanics, the 1:1 exposure tool is very cheap as compared to projection lithography.

Problem: Assume we need to perform a photolithography using MA6 mask aligner (with i-line light). The machine allows for vacuum contact lithography (g = 2 µm). The thickness of AZ4533 photoresist is about 2 µm.

1) What is the approximate resolution that can be achieved?

2) If the wafer has a small particle on it, with diameter of 1 µm, what is then the resolution?

3) If the wafer has a big defect (standing out of the surface) with height of 20 µm, what is then the resolution?

Solution:

1) $W = \sqrt{0.365 \cdot (2+1)} = 1.05\,\mu\text{m}$.

2) The gap between mask and resist can accommodate the small particle, therefore in principle the impact to the resolution is negligible.

3) The particle is larger than the mask/resist gap. $W = \sqrt{0.365 \cdot (20+1)} = 2.77\,\mu\text{m}$.

Problem 4.2: Contact Lithography.

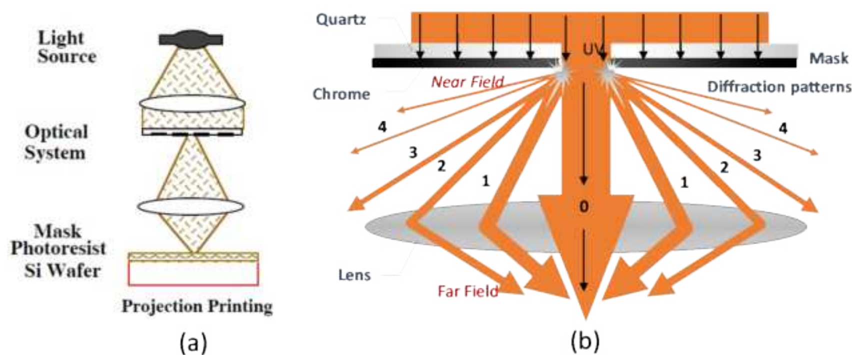


Figure 4.28: (a) Illustration of projection lithography. (b) Illustration of the far field imaging in projection lithography.

4.8.3 Projection lithography

Projection lithography was developed to circumvent the main disadvantages of contact lithography. The mask is moved away from the substrate to solve the physical contact and defectivity issues. A projection lens is positioned between the mask (called reticle) and substrate to image the diffracted light pattern back onto the substrate with a typical de-magnification factor of 4 or 5. This results in increased tolerances on the mask manufacturing side. An illustration of the projection system is shown in Figure 4.28(a).

projection lithography
reticle

Since the wafer is located at the far field of the aperture on the mask, the behaviour of the light, before being refocused by the lens, can be described by Fraunhofer diffraction. As shown in Figure 4.28(b), the diffraction occurs at the narrow aperture of the mask, causing a complicated diffraction pattern. Smaller features will result in higher diffraction orders, which in free space represent larger diffraction angles. It requires a large lens (large numerical aperture (NA)) in order to capture as much of the higher

numerical aperture (NA)

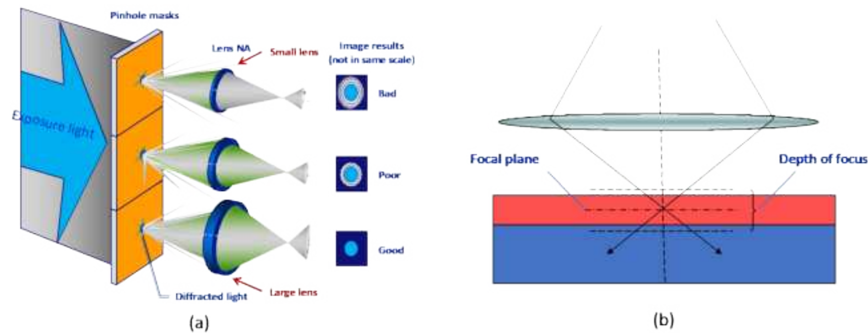


Figure 4.29: (a) Illustration of the relation between image quality and lens NA. (b) Illustration of the relation between image quality and DoF.

orders as possible, so that the reconstructed image is as close as possible to the original pattern on the mask. As can be seen from Figure 4.28(a), a large NA is required to achieve good image quality.

Critical Dimension (CD)
Rayleigh criterion

Critical dimension. The Critical Dimension (CD) is the minimum feature size that can be reliably imaged with a projection lithography system. It can be derived from the Rayleigh criterion [120] .

$$CD = k_1 \frac{\lambda}{NA} \tag{4.8}$$

As shown in the equation, the minimum CD that can be resolved, is linearly dependent on the wavelength of the light source (λ) and inversely proportional with the numerical aperture (NA) of the projection lens. The constant k_1 represents the specific lithography process conditions. This constant includes resist and developer sensitivity, mask engineering, illumination conditions, and advanced process integration-based enhancements such as double patterning.

From this equation it can be derived that to shrink the minimum CD, it is advantageous to use a shorter wavelength source. This trend has been seen in the industrial lithography systems [121, 122] where traditional filtered 365 nm Mercury light sources were replaced by 248 nm excimer lasers and later 193 nm lasers [123, 124, 125]. Currently extreme ultra-violet (EUV) sources with 13.5 nm wavelength are being deployed for large scale ultra-small CD production systems [123]. For more details see Section 4.8.4.

Next to the source, an improved projection lens can shrink the minimum CD as well [121]. Better and more complex lens systems were developed over the years, increasing the NA from approximately 0.4 to around 0.9, which is approaching the physical limit of what a lens with practical size and excellent quality can achieve. To improve the NA even further the air between the projection lens and the substrate was replaced by water (refractive index of 1.44) allowing the systems to reach an NA as high as 1.35 with so-called immersion systems [121]. For EUV systems, traditional optical lenses do not work anymore, and ultra-flat mirror systems were introduced instead. This has caused an effective decrease in NA to approximately 0.33. While improvements to the NA are being developed, it is still clear from the above equation that due to the much smaller wavelength EUV still yields a net minimum CD shrink. Since in practice the lens and wavelength of a specific system are hardware defined, people have also worked on improved process technology to lower the k_1-factor. Some examples of this

are high contrast resists, off-axis illumination, phase-shift masking, optical proximity correction and double patterning. This postponed the need to replace existing expensive lithography infrastructure with next generation hardware at a particular node.

Depth of focus. The large NA lens brings challenges to the lithography as well, which is illustrated in Figure 4.28(b). The depth of focus (DoF) describes the range over which the image is clearly resolved. It can be described by the equation below:

depth of focus (DoF)

$$DoF = \pm k_2 \frac{\lambda}{NA^2} \tag{4.9}$$

where k_2 is an experimental factor representing the tool and process conditions. During exposure, the light should be focused on the center of the photoresist layer, and in order to achieve a sharp image its DoF should cover the entire thickness of the photoresist. A larger NA lens, as well as a shorter wavelength, will lead to a shorter DoF, meaning the photoresist thickness must be reduced accordingly to maintain a good image quality. In modern DUV exposure tools, the DoF can be extremely narrow. For instance for the ASML PAS5500/1100B scanner at Nanolab@TU/e, the DoF is only about $\pm 150\ nm$. The reduced photoresist thickness may become a limiting factor in the subsequent processing. For instance it will limit the maximum etch depth when used as etching mask. A narrow DoF also puts a tighter demand on the substrate quality, in particular the TTV of the substrate (see Section 4.4.3). Any variation on the surface may cause the local image to be out of focus, leading to a poor image quality. For instance for application in modern DUV scanner tools, a TTV < 1 μm is required for the substrate.

Stepper vs. scanner. The conventional projection lithography tool is called a stepper. For this tool, a mask is commonly referred to as a reticle that contains the image of only part of the wafer. This pattern is imaged onto the wafer, after which the wafer is moved and the exposure is repeated to fill the wafer with replicas, or even different images from the same or other reticles. This “step and repeat” procedure continues till the entire wafer is exposed. The pattern on the reticle is typically 4x larger than the pattern on the wafer although other lens magnifications have been used in industry as well.

stepper
reticle
step and repeat

A more advanced exposure tool is called a scanner. Instead of exposing the entire field at once, the scanners use a narrow exposure area called “exposure slit”. The narrow side of the field is only a few mm wide. In the other direction the dimension is the full widt of the exposure field. The full field is now exposed by scanning the exposure slit across the exposure area. The main advantage of the scanner system is that the requirements for the optical lens quality are reduced, because in one direction the lens is now used only for imaging in the center of the image plane, and the aberrations which increase farther from the center do no longer degrade the image quality. This permits a smaller optical system, and a better optimized lens to suppress optical aberrations. For scanning the slit, these systems add a significant degree of complexity. On a scanner, each image is exposed while both the reticle and the wafer are moving (in opposite directions) through the narrow slit of light that passes through the projection lens. This dynamic method of exposure requires accurate control of the position and speed of both the reticle and the wafer relative to each other. The principles of the stepper and the scanner are compared and illustrated in Figure 4.30.

scanner
exposure slit

In practice, the scanner requires more stringent process conditions but provides the highest possible quality. For modern PIC manufacturing, it is mainly used in the most critical lithography steps of the process. The majority of the lithography steps are less

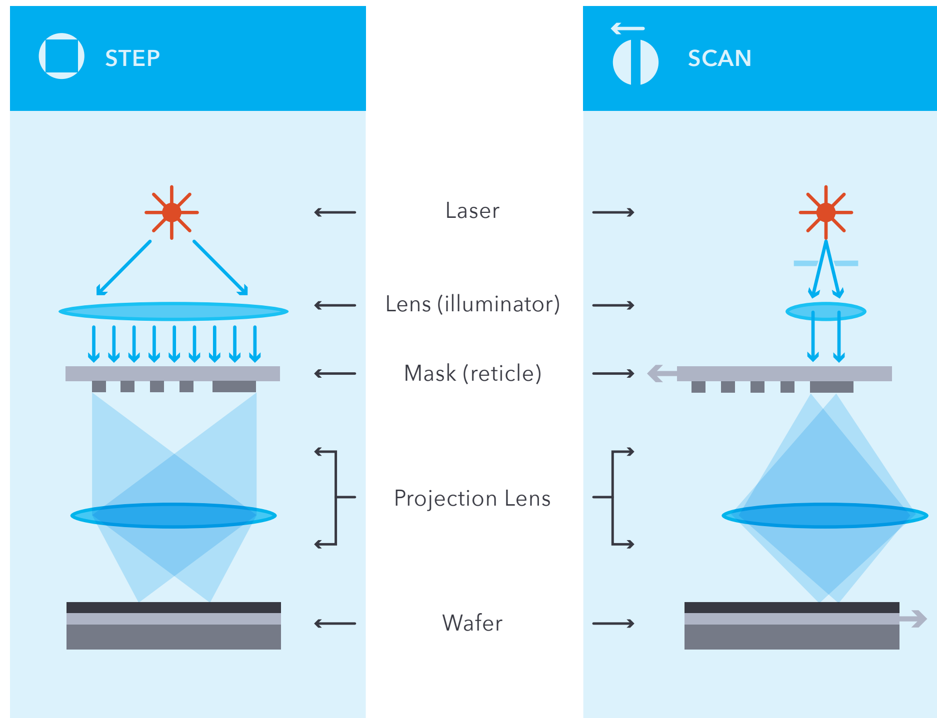


Figure 4.30: Comparison of the working principles of a stepper (left) and a scanner (right).

critical and are exposed by the stepper. Contact or proximity lithography is barely used in PIC manufacturing due to the low resolution and high defectivity.

4.8.4 Evolution of wavelength

As discussed in the previous sections, the wavelength used in the lithography is a main driving force in reducing the device size and keeping Moore's Law forward [126]. Figure 4.31 illustrates the evolution of the lithographic resolution along with the development of the technology nodes. Table 4.7 summarizes the wavelengths used in the history of semiconductor manufacturing and their basic light source technology. In the 1980s and 1990s, Mercury lamps were used to generate ultraviolet (UV) light; first g- and h-line, later also i-line (365 nm). The achievable resolution was typically larger than 0.5 µm. After that the technology evolved to a deep UV (DUV) wavelength of 248 nm and later 193 nm, which is still the most commonly used for advanced electronic chip manufacturing. Later, immersion imaging systems were introduced to enhance the NA of the optical system. DUV light is generated by excimer lasers with Fluoride-based gasses operating with emission of KrF or ArF. In recent years, the high demand for sub-10 nm CMOS pushed the light source technology to the 13 nm extreme UV (EUV) wavelength, produced by laser-generated plasma (LPP) sources. This wavelength is currently the key enabler for sub-10 nm CMOS technology nodes and will continue to play an exclusive role (with higher NA systems) for future 2 nm or even 1 nm nodes.

i-line
deep UV (DUV)
excimer laser
KrF
ArF
extreme UV (EUV)

As discussed earlier, the NA plays an important role together with the wavelength. Figure 4.31 provides a comprehensive summary of the development of the optical systems

Table 4.7: Summary of wavelengths and light source technologies used in photolithography.

UV wavelength (μm)	Wavelength name	EV emission source
436	g-line	Mercury arc lamp
405	h-line	Mercury arc lamp
365	i-line	Mercury arc lamp
248	Deep UV (DUV)	Mercury arc lamp or Krypton Fluoride (KrF) excimer laser
193	Deep UV (DUV)	Argon Fluoride (ArF) excimer laser
157	Vacuum UV (VUV)	Fluorine (F_2) excimer laser
13	Extreme UV (EUV)	Tin (Sn) laser-produced plasma source

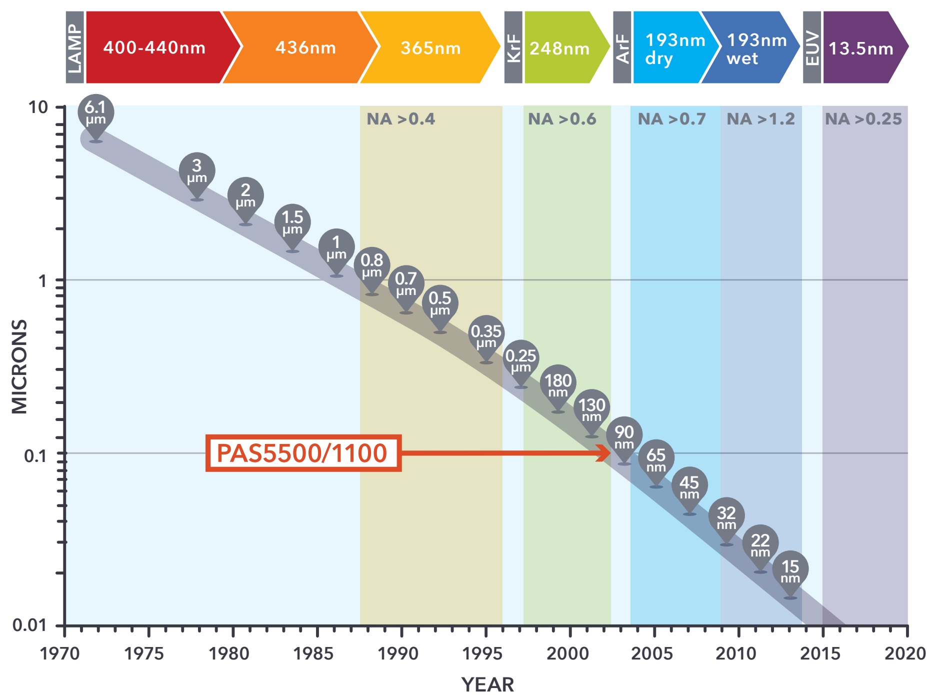


Figure 4.31: Evolution of wavelengths and numerical apertures used in lithography tools. The PAS5500/1100 scanner used in Nanolab@TU/e is indicated with an arrow.

used in lithography tools.

4.8.5 Masks

The mask is an essential part which carries the design features that, during the lithography, are transferred to the semiconductor wafer. A typical mask is based on a glass or quartz plate which is transparent to the UV spectrum. In high-end applications, quartz is preferred since it offers better UV transmission and a superior low thermal expansion coefficient. The glass or quartz plate is coated with a thin layer of chromium metal, which carries the design patterns. When the patterns of the chromium represent the

Figure 4.32: Side by side photograph of a 6-inch scanner 4X reticle (left) and a 4-inch 1X contact lithography mask (right) both displayed with chrome pattern side up.

actual design and other areas of the mask are transparent, the mask is called clear-field mask. In the opposite case when the transparent areas in the mask represent the actual design and the other areas are covered by the metal, the mask is called dark-field mask. The properties of clear or dark field masks are discussed in Section 4.8.6. In projection lithography, the term "reticle" is widely used instead of "mask". The patterns on the reticle are typically 4x – 5x larger than the actual design, due to the nature of the projection lithography. Figure 4.32 shows a 1:1 mask and a 4:1 reticle side by side for comparison.

clear-field mask

dark-field mask

reticle

Besides the conventional binary masks where only the intensity is controlled (i.e., no exposure for metal pattern, full exposure for transparent area), one can also use phase-shifting masks [127], which add an extra degree of freedom to control the phase of the light. With this technique higher resolution can be achieved. For PIC manufacturing, it can be especially beneficial to improve the imaging of DBR/DFB gratings in lasers [128].

4.8.6 Photoresists

photoresist

In the lithographic process, a photosensitive polymer film (the photoresist) is applied to the wafer, dried, and then exposed to UV light or other radiation through a mask with the proper geometrical patterns. After exposure, the wafer is soaked in a solution that develops the images in the photosensitive material. Depending on the type of polymer used, either exposed or non-exposed areas of the film are removed in the developing process. The procedure is described in detail in Section 4.8.7.

Photoresist contains several components. The most important ones are the resin and the sensitizers. The resin is a mixture of polymers used as binder. It gives the resist the desired chemical and mechanical properties. The sensitizers are photosensitive components which will trigger the reactions during light exposure. There can also be additives in the resist, which are chemicals that control specific aspects of the materials. All of them are dissolved in solvent, which gives resist its fluid characteristics. In liquid form, the resist can easily be spin-coated on any substrate.

chemically amplified photoresist

Modern semiconductor processes are all using chemically amplified photoresist. Instead of traditional photoresists where several photons are needed to trigger a photochemical reaction, the chemically amplified resists have significantly higher sensitivity. When a photochemical reaction occurs in the chemically amplified resist, the photochemically generated acid acts as a catalyst, which induces a cascade of chemi-

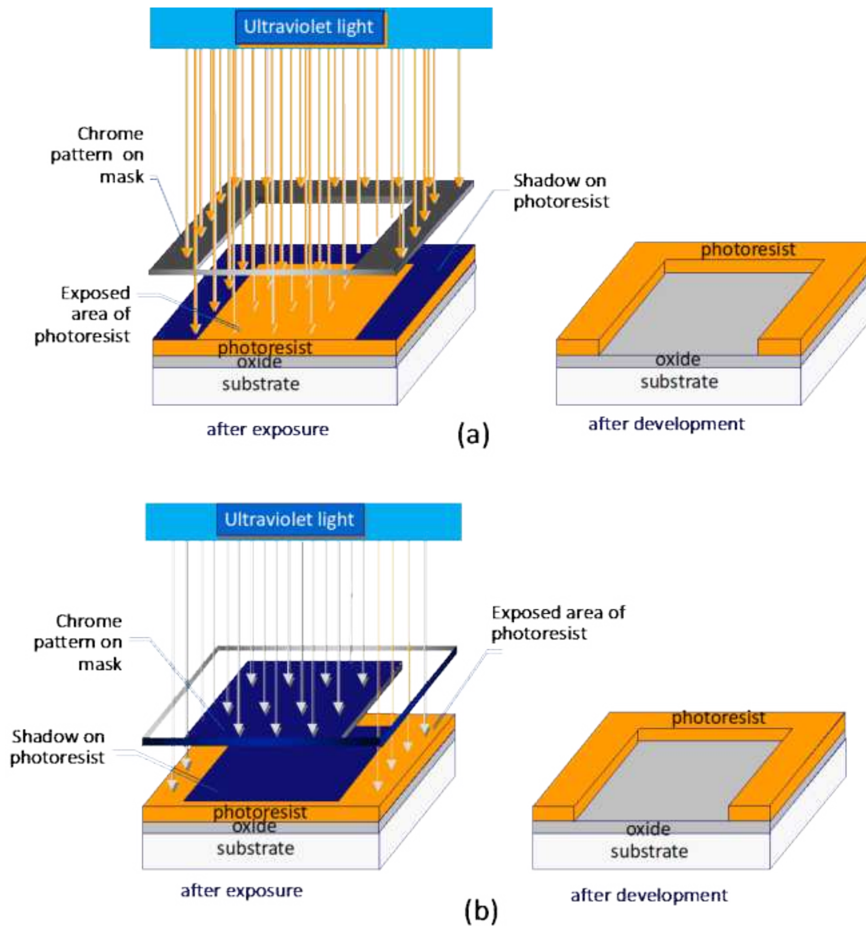


Figure 4.33: (a) Lithography with negative resist. (b) Lithography with positive resist. From [119]

cal transformations in the resist film, providing a gain mechanism [129]. Such resist is dominant since the introduction of DUV lithography.

However, the chemically amplified resists are more sensitive to contaminations, even from the air. This is mainly because once coated, the chemically amplified resist is susceptible to interaction with volatile compounds that can dissolve into the layers and influence the intended chemical reaction of the lithography process [130]. Therefore any significant waiting time or delay needs to be taken into account, even for just a few hours of transporting the wafers [128].

Photoresists can be categorized into two types: positive resist and negative resist. As illustrated in Figure 4.33, the positive resist will result in a pattern on the wafer which is identical to the pattern on the mask. In the positive resist, the exposed part becomes softened and soluble in the developer solution. Contrary, the negative resist will result in an inverted pattern. The exposed part becomes hardened and insoluble in the developer solution.

It can also be observed from Figure 4.33 that, by using a proper combination of photoresist (positive or negative) and mask (clear or dark field), one can achieve the same pattern on the wafer. This is depicted in Figure 4.34.

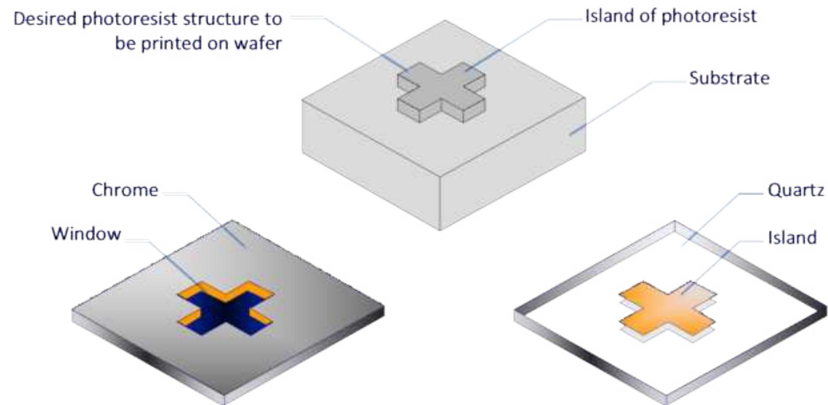


Figure 4.34: Relationship between polarity of the mask and the photoresist, to achieve the same pattern on the wafer.

4.8.7 Lithography procedure

A complete lithography procedure consists of eight major steps, which are illustrated in Figure 4.23 below.

- (a) Vapor prime promotes good photoresist-to-wafer adhesion by coating the wafer with a monolayer of Hexamethyldisilazane (HDMS). Note that this step depends on the material. For example, for InP it is needed, but for Si it is not needed.

- (b) Spin coating is widely used to develop a uniform layer of photoresist on wafer. Photoresist is dispensed on the wafer, which is held onto a vacuum chuck. The spinning of the wafer results in a uniform resist layer, the thickness of which depends on the viscosity of the polymer and the spin speed.

- (c) Soft bake has several effects. It improves photoresist-to-wafer adhesion, promotes resist uniformity on the wafer and most importantly removes solvent in the resist.

- (d) Alignment and exposure are carried out in a lithography tool. The UV light transfers patterns in the mask to the resist-coated wafer. After exposure, photosensitive components in the resist are activated and change the properties of the resist locally. In many cases, the patterns on the mask need to be aligned to the existing patterns on the wafer prior to the exposure. It is called overlay exposure (see Section 4.6.8).

- (e) Post-exposure bake is required for some photoresists. It is performed immediately after the exposure. It helps to complete the chemical reaction in the resist, release mechanical stress and smoothen the exposed/non-exposed interface.

- (f) Development is performed to dissolve the unwanted part of the photoresist. Each photoresist requires a specific type of developer for optimal results.

- (g) Hard bake is performed on some photoresists. It helps dehydrate the wafer surface and also further improves the resist-to-wafer adhesion.

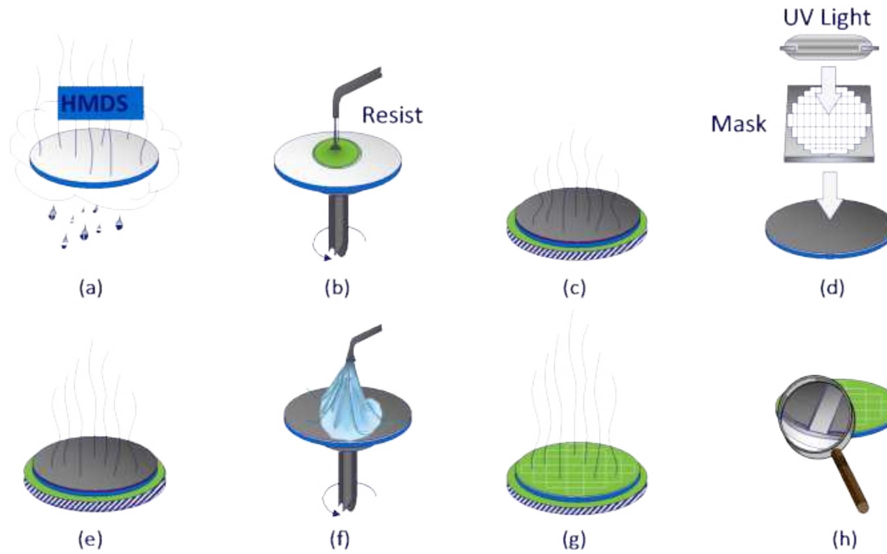


Figure 4.35: Lithography procedure: (a) vapor prime; (b) spin coat; (c) soft bake; (d) alignment and exposure; (e) post-exposure bake; (f) develop; (g) hard bake; (h) inspect.

- (h) Inspection is essential to assess the quality of the lithography. In an optical or electron microscope, one can measure the line width of the patterns, inspect any particles and defects and assess the uniformity across the entire wafer. If the quality turns out unacceptable, it is still possible (in most cases) to remove the photoresist completely and redo the lithography procedure from beginning.

The presented lithography process is a basic example. It is possible to add various features to this flow to enhance the lithography results. For instance, a multi-layer coating is commonly applied for ArF lithography to reduce reflections during exposure. The layers may contain anti-reflection coatings on top and at the bottom of the photoresist layer, which improve resolution and pattern fidelity [128].

4.8.8 Mask overlay

overlay

The term overlay is used in semiconductor manufacturing to describe the accuracy with which one fabrication layer can be put on top of a previous layer [120]. When trying to minimize the overlay errors, it is necessary to distinguish between the different contributions to these errors. Overlay performance is typically quantified as the distribution and size of relative position errors of specific locations within two lithographic layers on a wafer. For the overlay performance, random stage positioning errors can be the cause of relative shifts between the layers. Apart from calibrating stages and maintaining their performance, these are mostly hardware limited and cannot be corrected for. Secondly, since both scanners and steppers repeatedly move the wafer to a specific location to perform a single exposure, systematic errors can occur. These can be divided in two categories as illustrated in Figure 4.36. There are overlay errors between different exposure fields (inter-field), which can be caused by wafer deformation, rotation, translation errors, marker processing or readout. There are also overlay errors within exposure fields (intra-field), which are typically caused by magnification errors, skew or trapezoid deformation of the image, lens distortions or reticle rotation.

overlay errors

inter-field overlay errors

intra-field overlay errors

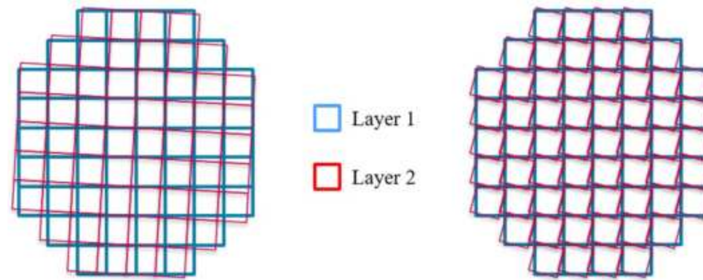


Figure 4.36: Systematic overlay errors between blue layer 1 and red layer 2: inter-field (left). From [128]

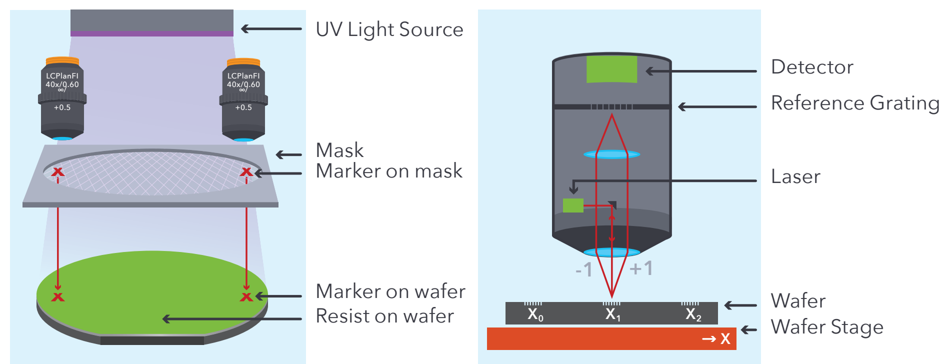


Figure 4.37: (a) Illustration of overlay with images. (b) Illustration of overlay with interferometry.

There are two ways of performing mask alignment in photolithography: image-based alignment and interferometry/scatterometry based alignement. The former one captures the images of the alignment markers on the wafer and on the mask, and mechanically moves and rotates the wafer and/or the mask so that the markers match to each other. This is illustrated in Figure 4.37(a). The alignment accuracy is limited by the resolution of the optical microscope system to about 200-300 nm. This method is mainly used in contact and proximity lithography. The same principle is also used in some mask-less lithography tools such as Electron-Beam Lithography (EBL, see Section 4.9) and direct laser writers (not discussed in this book). The interferometry and scatterometry methods not only utilize the intensity, but also the phase of the light. A typical example is the alignment using phase markers in the ASML scanner tool. As illustrated in Figure 4.37(b), when coherent light exposes a prefabricated phase marker (based on gratings) on the wafer, multiple orders of scattered light will be emitted by the grating. The wafer position information is contained in the phase of the scattered light. By capturing and interfering the different orders of scattered light, the phase information will be converted to an intensity pattern (i.e. a fringe pattern) which can be easily measured and analysed. The alignment accuracy using this method is extremely high. Sub-nanometer alignment accuracy is achievable [131].

Table 4.8: Lithography capability comparison, showing resolution, overlay, throughput and defectivity, assuming that the EBL-written pattern is 1% of the full wafer area, for determining throughput in 300 mm wafer equivalents.

Lithography (type)	Resolution	Overlay (Single Machine)	Throughput (300 mm)	Defectivity
Contact	500 nm	250 nm	160 WPH	*
Projection	13 nm	1.4 nm	125 WPH	***
EBL	6 nm	5 nm	0.005 WPH	***

*** Very Good, ** Good, * Moderate

4.8.9 Lithography for InP PICs: solutions and challenges

Table 4.8 summarizes the strengths and weaknesses of the different lithography technologies. EBL is described in Section 4.9. Most of the rated properties in this table are equally important for generic photonic integration. In essence, photonic integration requires exceptionally high dimensional control and low roughness. Both are highly dependent on the lithographic resolution. Some photonic components also require a challenging degree of overlay control. Lastly, generic InP integration aims to exploit the scalability of the technology platform to leverage the PIC prices. Therefore, throughput, indicated as wafers per hour (WPH), and defectivity performance are also required to deliver large volumes of high-yield wafers. Projection lithography comes out as an excellent candidate for InP photonic integration.

wafers per hour (WPH)

In recent years, Silicon photonics platforms have been able to exploit the existing high-resolution manufacturing infrastructure of the traditional electronics industry. This advantage has allowed the manufacturing of devices with ArF 193 nm immersion lithography used for technology nodes as low as 28 nm [132, 133]. It is, however, important to create precision photonic components also on InP substrates that enable monolithic integration with lasers and high efficiency optoelectronic devices. So far, generic InP platforms have been restricted to the use of e-beam, stepper and contact lithography. For volume production of PICs, stepper lithography is needed, but minimum feature sizes have been limited to 250 nm [1]. This limitation comes from the fact that the shortest wavelength illumination and highest performance imaging have only been available for the latest large silicon wafer diameters. Whilst 4-inch InP wafers are used for high volume photonic devices and 6-inch is starting to become available, the InP industry has mostly been operating on substrates of 3-inch for high functionality PICs. In 2011 an ASML PAS5500/1100B scanner tool was installed at the NanoLab@TU/e cleanroom in Eindhoven and modified to be the first scanner worldwide to allow for fast 100 nm resolution exposures with 15 nm overlay on InP substrates as small as 3-inch. In 2013, a PAS2500/40 stepper was also installed with up to 800 nm resolution and 250 nm overlay. This machine replaced most traditional contact lithography steps and enabled an upgrade in terms of overlay, dimension control and defectivity. A photograph of both tools is shown in Figure 4.38.

Challenges. Due to physical limits that determine the interaction with light, minimum feature sizes of photonic components [134] do not scale down as far as state-of-the-art electronics [135]. However, uniformity tolerances on CD are exceptionally strict for photonic components despite their size, because in waveguides the CD directly influences the effective index [132, 136]. Opposite to the requirements on dimensional control stands the cost of exploitation of the fabrication technology. The choice for a particular manufacturing infrastructure is, therefore, always based on a

Figure 4.38: ASML tools at Nanolab@TU/e: PAS5500/1100B scanner (left) and PAS2500/40 stepper (right).

compromise where the cost needs to be met with appropriate volume. The ArF scanner model installed in NanoLab@TU/e, enables reproducible imaging down to 100 nm CD with very good overlay specifications. Less critical steps can be exposed with the installed i-line Stepper that can exploit existing resist process modules to deal with the patterned wafer topography. The transition of moving from i-line contact lithography to ArF scanner lithography has taken the silicon electronics industry close to 20 years to complete [122]. The application of such lithography in InP photonics inherently poses unique and different challenges that require extensive research as well. To create waveguide geometries in InP, etch depths of over 1 μm are often required, while ArF resists suitable for 100 nm CD patterning only have a thickness of around 200 nm. One challenge is in fully covering the wafer topography with a uniform, high-quality resist layer and anti-reflection coatings. The aggressive CD shrink that happened in electronics, also led to scaling in layer thickness of the integrated circuit materials. Improvements in terms of wafer topography and flatness were required to deal with reduced depth of focus (DoF) implied by high-resolution lithography systems. Planarization can be applied to reduce wafer topology before critical lithography. Similarly, flatness of InP substrates needs significant improvement to enable high-resolution lithography. Many PICs require edge coupling of light to and from the chip, while maintaining phase, polarization, wavelength and intensity. Because PIC dicing is often performed by cleaving along the crystallographic plane, this requires accurate matching of the chip coordinate system to the orientation of the substrate. Another challenge is to transfer a thin ArF lithography pattern into the semiconductor material. This requires highly selective hard mask patterning techniques, through the development and optimization of several uniform etch steps. Several process modules were developed at TU/e to achieve this for a specific set of materials and applications [128].

Some InP photonic components will push the limits of what the TU/e scanner can resolve. In that respect, it is required to investigate methods for imaging improvement to stretch tolerances and limits of this system. Double patterning is a method that is exploited in electronics to achieve this and was investigated. Secondly, lithographic simulation and application of off-axis illumination by using diffractive optical elements was investigated. Lithographic simulations are also used to apply optical proximity correction (OPC) to designs [137], to enable structures that are otherwise impossible to image with traditional binary masking.

off-axis illumination
optical proximity correction (OPC)

4.9 Electron-beam lithography

The resolution of optical lithography depends mainly on the light wavelength and the optics used. To achieve resolutions in the order of tens of nm, the machines become very expensive.

Electron-beam lithography (EBL) is the most common lithographical technique for obtaining feature sizes down to a few nm. It offers, besides a higher resolution, an easy pattern modification and absence of mask defects. The high resolution of E-beam pattern generators (EBPG) results from the ability to focus the e-beam to spot sizes of 10 nm. Another advantage of EBL is a larger practical depth of focus in the order of 25 μm. It means that, contrary to optical lithography, sample flatness is not a limiting factor in EBL.

electron-beam lithography (EBL)

A schematic illustration of an EBL system is shown in Figure 4.39(a). An electron gun generates the electron beam. Modern high-resolution systems use field electron emission sources [138] for their lower energy spread and enhanced brightness. The generated electron beam will experience several stages of lenses in order to shape the beam to the desired quality. Magnetic lenses are mostly used, the magnetic field of which is generated by the coils as shown in the figure. At the final lens stage, the beam can be deflected within a small field (the write field) with dimensions in the order of 100 μm to 1 mm. To expose a larger area, precision stage movement is required. This leads to a relative positioning error between write fields, the stitching error. Typical stitching errors can be maintained < 50 nm. It is an extremely small error, but it can still cause considerable effects in photonic devices. For instance, a single-mode InP or Si membrane waveguide is about 400 nm wide. A stitching error of 50 nm will lead to significant scattering loss and also back reflection. And in InP-based PICs an error of 50 nm corresponds to 10% of a wavelength which will have a large effect in interferometric devices.

write field

stitching error

The beam spot size is not the final achievable resolution. Electron scattering in the resist layer and backscattering from the substrate will limit the practical resolution to 20-30 nm. Figure 4.27 (b) illustrates the electron scattering in a medium. An incident electron (purple) produces secondary electrons (blue). Sometimes, the incident electron may itself be backscattered as shown here and leave the surface of the resist (amber). This effect, which causes the actual exposed pattern to be wider than the scanned pattern, is called the proximity effect. As a result the resist outside the scanned pattern receives a non-zero dose, leading to a deterioration and deformation of the pattern. However, the electron scattering behavior in the medium (e.g., resist, InP, Si) can be simulated and thus computationally corrected (so called proximity effect correction algorithm) [139]. Another limitation of this technique is the lower speed at which samples can be patterned using this method of direct writing. Hence this technique is extensively and commercially used to fabricate high resolution optical masks and reticles that are used in optical lithography, but much less for exposure in volume production. There it is used in conjunction with optical lithography where the E-beam technique is reserved only for the critical patterns which require the high resolution and/or the high depth of focus, for instance for distributed Bragg gratings [140].

electron scattering resolution

proximity effect

proximity effect correction

direct writing

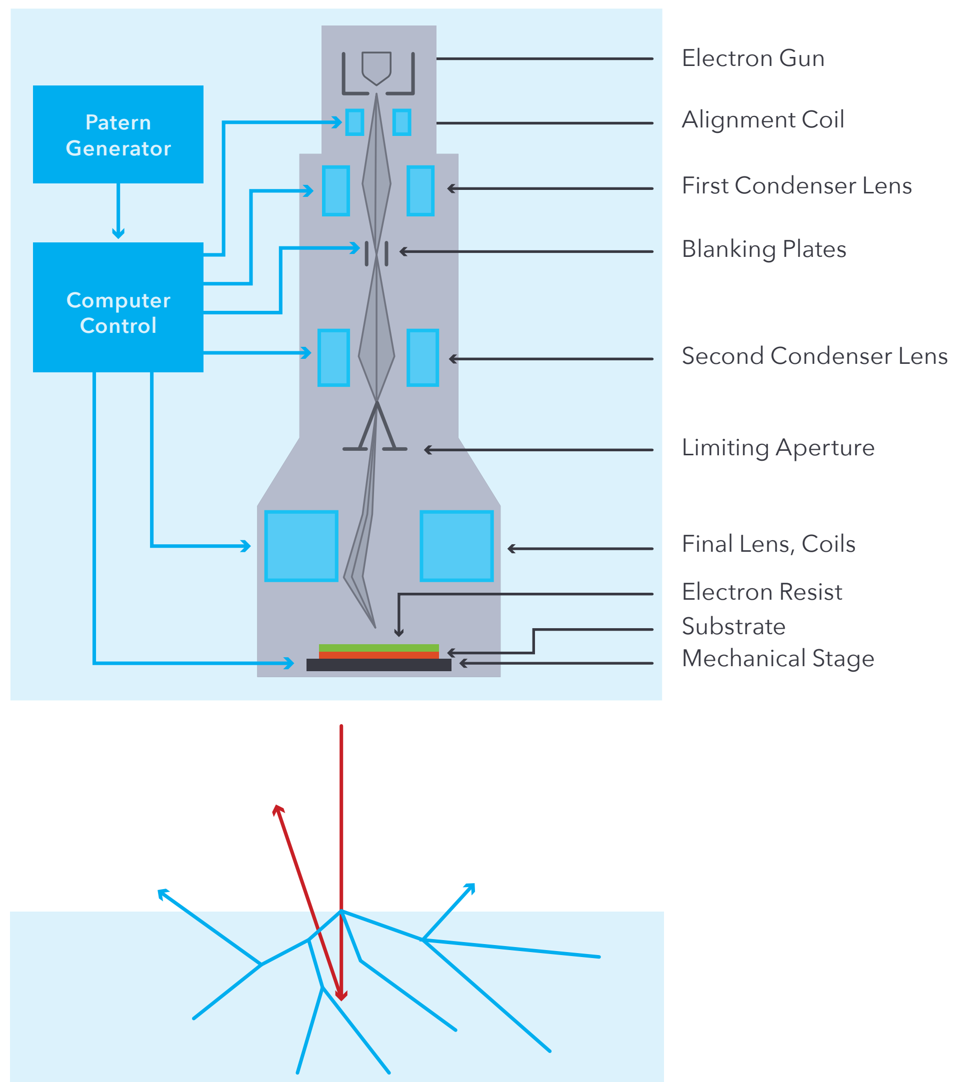


Figure 4.39: (a) Schematic of an EBL system. (b) Illustration of electron scattering in a medium.

4.10 Deposition techniques

4.10.1 Plasma Processes

In Plasma-based deposition processes the materials that are to be deposited, e.g. SiN or SiO, are formed in a plasma. A plasma is a thin gas (at low pressure) in which a small part of the gas atoms or molecules are ionized. As the plasma contains both the positive ionized atoms and free electrons that are generated during the ionisation, the plasma is neutral and there is no electric field inside the plasma because an electric field is neutralized by displacements of the free carriers (ions and electrons). An electric field will only exist at the edges of the plasma, where carriers can escape from the plasma, usually at the walls of the vacuum chamber. Here a space charge region will be formed which counteracts the escape process. Dependent on the gas pressure this space charge region is a few millimeters up to a few centimeters wide.

plasma

space charge region

When a free electron and an ion collide, they recombine under emission of light, that's why a plasma glows. The color is dependent on the gas species, nitrogen plasma is blue, argon plasma purple and oxygen plasma yellow. To remain stable the plasma needs a constant ionization process to compensate for the carriers which are lost by recombination. This can be achieved by bringing the plasma in contact with a negatively charged electrode. The electrode will attract and accelerate the positive ions, which will collide with the electrode and generate secondary electrons, which will accelerate towards the plasma and compensate for the electrons which are lost by recombinations. Ionizing collisions of accelerated particles will also contribute to maintaining the plasma. The other electrode, which can be the chamber wall or a second electrode, will get positively charged relative to the plasma, which will get a negative bias by the loss of positive ions at the other electrode. The potential which the plasma will take is a complex balance between the carrier extraction and generation mechanisms in the space charge regions around the electrodes, and strongly dependent on the gas pressure and the electrode surface area and condition. It can make a big difference whether the chamber wall is clean or covered with a thin layer of deposition or etch residues. Therefore, keeping the chamber conditions constant is extremely important for getting reproducible process specifications.

In order to avoid that the plasma generation process is slowed down or even stopped by surface charge arising at the surface of an electrode, an RF plasma is often applied, in which surface charge built up in the positive part of a cycle, is eliminated in the negative part.

RF plasma

Plasma processes are applied with advantage for getting increased chemical reaction rates at relatively low substrate temperatures, e.g. in Chemical Vapour Deposition processes. They are also applied in Reactive Ion Etchers, where the ions that are accelerated in the space charge region at the edges of the plasma, are used for etching material away from a wafer, which is exposed to the plasma. And in sputter deposition the energy of the accelerated gas ions is used for breaking atoms out of a target electrode, and let them condense on a substrate which is mounted opposite to the target electrode, with the plasma in between.

substrate temperature

4.10.2 Plasma Enhanced Chemical Vapour Deposition

SiO_x and SiN_x are commonly used as hard mask and/or for passivation. Plasma enhanced chemical vapour deposition (PECVD) is a common and widely used technique for depositing silicon dioxide (SiO_x) or silicon nitride (SiN_x) layers. Chemical vapor deposition (CVD) refers to the process of chemically reacting volatile gasses to produce

plasma enhanced chemical vapour deposition (PECVD)

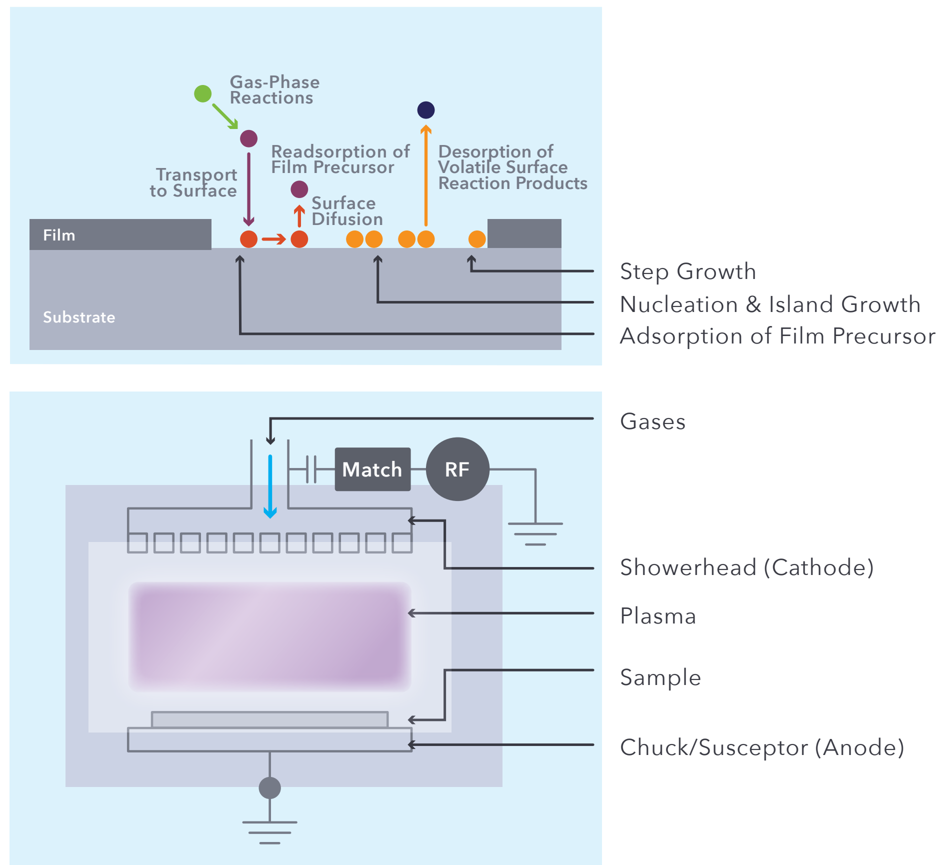


Figure 4.40: (a) Illustration of a CVD process. (b) Schematic of a PECVD chamber.

chemical vapor deposition (CVD)

a non-volatile solid that deposits atomically on a substrate. The process is schematically depicted in Figure 4.40. The gasses react in the proximity of the hot substrate surface, the resulting precursors diffuse and adsorb onto the surface, where they nucleate. Volatile by-products are desorbed, but a small fraction is built into the grown film, which can degrade the film properties. In plasma enhanced CVD, the gasses are activated by RF or microwave energy, and the process can take place at a lower substrate temperature. The plasma contains not only neutral molecules or atoms, but also many other components, such as ions (when electrons are taken away from a molecule or atom), radicals (when an atom is taken away from a molecule), free electrons and photons. This makes the reaction much more efficient, so that high-quality films can be formed at relatively low substrate temperatures. A schematic of a PECVD chamber is shown in Figure 4.40(b). Gasses are injected into the chamber through a showerhead which produces a uniform distribution of the gases. Power from an external RF source is coupled in the chamber and turns the gasses into a plasma. The substrate is located close to the plasma to “collect” the precursors.

PECVD can deposit films at relatively low substrate temperatures (e.g. 300°C). Although the electron temperature in the plasma may be near 10^5 °C, the sample temperature remains close to the set electrode temperature. This technique is mainly used to deposit SiN_x and SiO_x films. SiN_x is deposited by reacting silane and ammonia or nitrogen, while SiO_x is deposited using silane and nitrous oxide. The reaction formulas

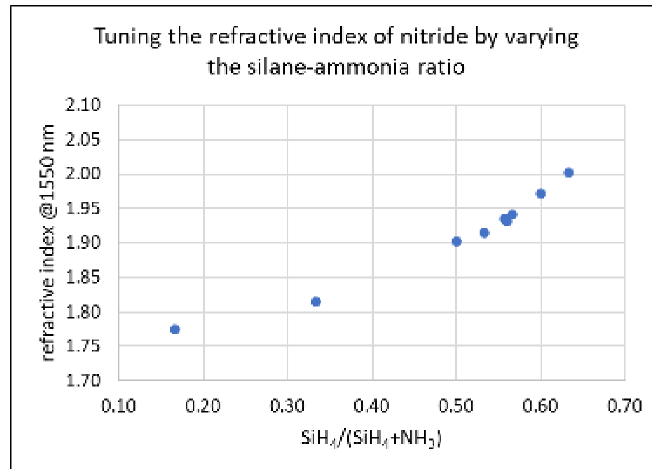
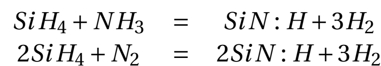


Figure 4.41: Refractive index tuning by gas ratio between SiH_4 and other gasses.

for forming nitride are



$$\begin{aligned} SiH_4 + NH_3 &= SiN:H + 3H_2 \\ 2SiH_4 + N_2 &= 2SiN:H + 3H_2 \end{aligned}$$

and for forming oxide

$$SiH_4 + 4N_2O = SiO_2 + 4N_2 + 2H_2O$$

In PECVD of dielectrics the flow ratio of silane/ammonia (SiN_x) or silane/nitrous oxide ($SiOx_)$) influences the chemical composition of the corresponding dielectric layer. For instance increasing the ratio would result in more incorporation of amorphous Si in the dielectric leading to an increase of the refractive index to even higher values than that of stoichiometric Si_3N_4 or SiO_2, which are 2.04 and 1.46, respectively, at a wavelength of 633 nm. The case for SiN_x is shown in Figure 4.41. Also, using silane, ammonia and nitrous oxide in the same process leads to deposition of SiO_xN_y with various oxygen and nitrogen compositions, which are strongly dependent on the used ammonia and nitrous oxide flows.

sticking probability

A very important and practical concept in PECVD is the sticking probability s, which describes the probability that molecules are trapped on surfaces and adsorb chemically. When depositing materials on topologies, especially narrow slots, the effect caused by the sticking probability is most significant. As can be seen from Figure 4.42(a) and (b), gasses with high sticking probability tend to accumulate at the top entrance of a trench, rather than diffusing deeper into the trench. This results in a thicker deposition at the top opening than at the bottom. For narrow trenches, the top opening may already be closed while the bottom is still not completely filled, as shown in Figure 4.42(c). It is most often observed in PECVD SiO_x films. There are other technologies that can solve this problem. One can use gasses with lower sticking probability, for instance Tetra Ethyl Ortho Silicate (TEOS) instead of SiH_4. Another method is to use atomic layer deposition (ALD) which can grow the materials atomic layer by atomic layer, resulting in extremely conformal deposition.

TEOS
atomic layer deposition (ALD)
conformal deposition

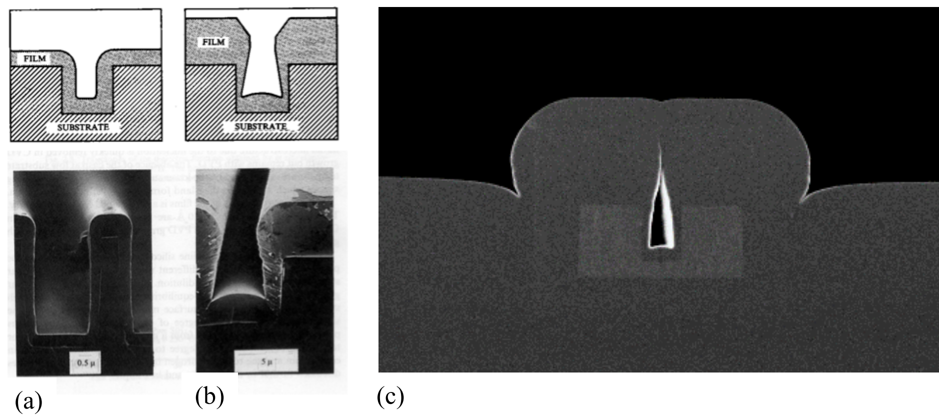


Figure 4.42: Deposition profiles with (a) low and (b) high sticking probabilities. (c) Example of PECVD SiO_x film with a void in the slot due to high sticking probability.

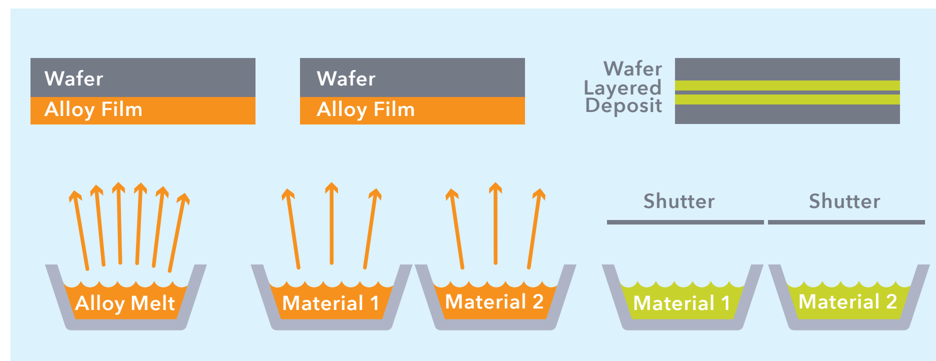


Figure 4.43: Different methods to evaporate alloys and layered materials.

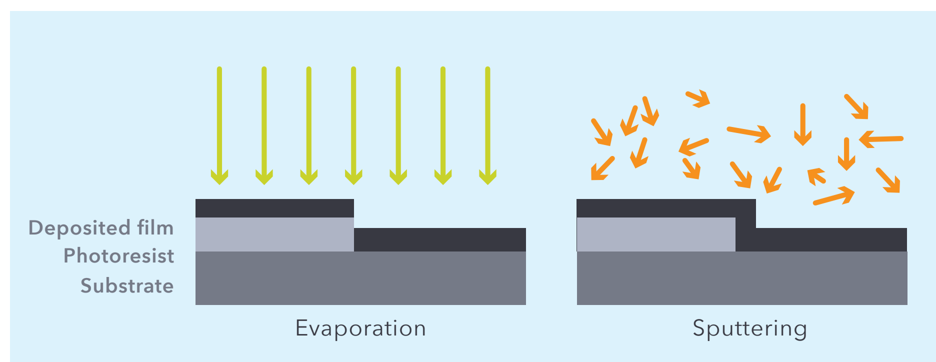


Figure 4.44: Illustration of step coverage in evaporation and sputtering processes.

4.10.3 Evaporation

Evaporation is a simple yet effective method to deposit high quality metals as well as oxides and nitrides. The basic principle can be seen from Figure 4.43. The source material, an ultrapure solid, is evaporated in vacuum. The vacuum allows the evaporated particles to travel to and condense on the target substrate. The evaporation process is done by either heating up the entire source material (thermal evaporation) or heating the surface of the material by an electron beam (E-gun evaporation). Alloys or layered materials can also be deposited using evaporation, as shown in Figure 4.43. Main advantage of evaporation is that it gives a high purity due to the low pressure and very pure starting materials. A disadvantage is the poor step coverage due to the high directionality of the evaporated particles, as illustrated in Figure 4.44. However, the high directionality can be beneficial in some particular processes, for instance in the metal lift-off process (see Section 4.10.4 for details).

evaporation

thermal evaporation

E-gun evaporation

4.10.4 Metal lift-off

Metal patterns on Si are typically achieved by deposition and selective etching. A metal is applied on a Si-wafer, then spinning resist over the metal, patterning the resist and etching away all exposed metal. This process cannot be applied for InP or GaAs wafers, however, because the metal stacks usually contain gold or platinum, which is very difficult to etch. Therefore metal patterning in InP is done with a so-called lift-off process, through a resist pattern with an adequate profile (top wider than bottom, creating an undercut) which allows for easy removal of the metal on top of the resist. A comparison of the two methods is shown in Figure 4.45.

lift-off process

undercut

The choice of the deposition method is important. Evaporation is used here because the high directionality of the evaporated metal causes the metal in the contact openings to be disconnected from the metal on the resist. The dimensions of the resist profile are critical. A negative resist, such as MaN-400 resist, can be used. In such a negative resist, the profile of the resist can be fine tuned by exposure time. As shown in Figure 4.46, the undercut profile of the resist becomes more and more significant with increased exposure time. Too short exposure time will lead to insufficient undercut and a higher risk of lift-off failure. Too long exposure will result in the closing of the resist gap, therefore reducing the actual width of the metal pattern. Soft bake of the resist can also control the resist profile to a certain extent. Figure 4.46 also shows that a higher soft bake temperature will result in a harder resist and, consequently, a less steep sidewall slope. The resist profile determines the actual achievable dimensions of the metal. An empirical rule for the thickness of the metal is that it should not exceed half of the thickness of the resist. Otherwise the metal may not break properly at the resist edge. Furthermore, the thickness of the resist also determines the smallest possible width of the metal. It is generally more difficult to properly resolve and open a narrow gap for thicker resist. In other words, the resist thickness determines the upper boundary of the metal thickness as well as the lower boundary of the metal width.

evaporation

undercut profile

soft bake

4.10.5 Sputtering

Sputtering is based on ion bombardment of a source material. Typically the ion is created from the plasma of an inert gas, such as Ar^+ from Ar gas. And the source material can be both metal or dielectric material, which is mounted on a plate which acts as one of the electrodes of a parallel plate reactor, as shown in Figure 4.47. It is called the target. When the plates are provided with DC or RF bias, then the Ar^+ ions will be

sputtering

target

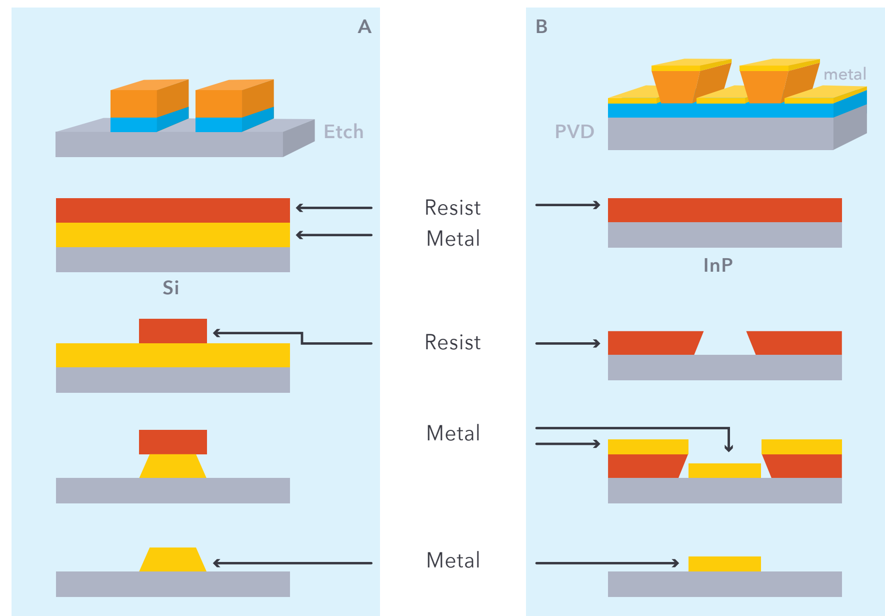


Figure 4.45: Simplified process flow of metal etching for Si and metal lift-off for InP.

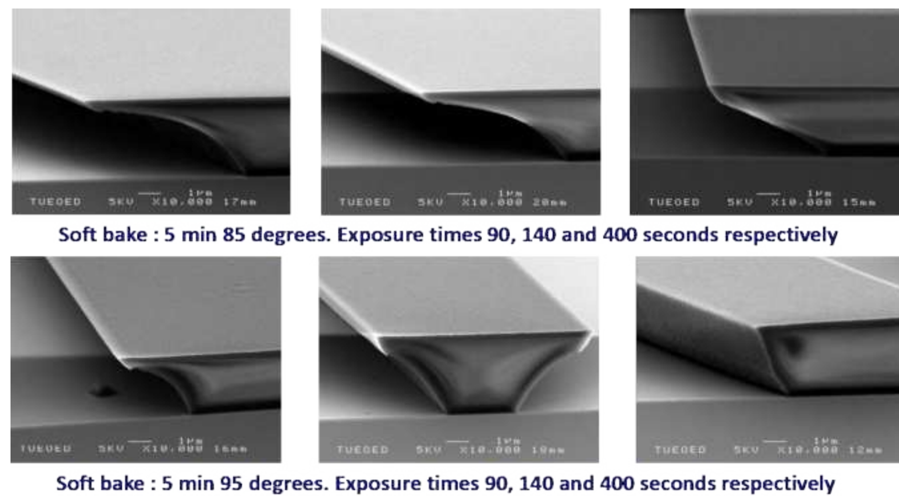


Figure 4.46: Sidewall profile engineering of the MaN-400 negative resist profile as a function of exposure time and soft bake temperature.

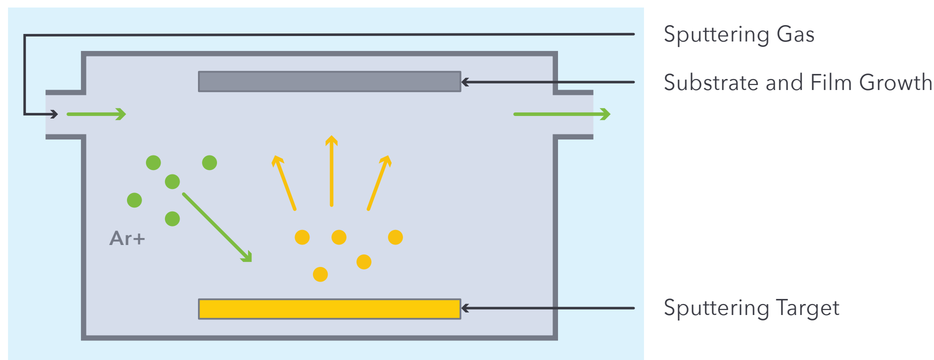


Figure 4.47: Schematic of sputtering deposition.

accelerated by the electric field between the plates. When they gain sufficient kinetic energy and hit the surface of the source material, source atoms will be ejected and re-deposited on the wafer which is mounted on the plate opposite to the target. Due to the nature of the physical bombardment, the atoms will be ejected in a wide range of directions. Therefore, sputtering is suitable to cover height steps, as shown in Figure 4.44, but not suitable for lift-off processes (Figure 4.45). The purity of sputtered material is not as good as that of evaporated material, because gas pressures in plasma processes are much higher than for evaporatio. Therefore, gas ions will be buried into the growing film and become impurities.

step coverage

There is one application where sputtering is the only solution, namely in electro-plating. As shown in Figure 4.11, a thin TiAu conductive seed layer is required to cover the entire surface of the wafer, despite of the height steps present. This conductive layer is essential for the Au electroplating process. Obviously, sputtering is the only choice for this layer.

electro-plating

Materials that can be deposited with sputtering are not limited to metals. Metal oxides, for instance, can also be obtained. This can be done from a metal target by injecting O_2 gas into the chamber for oxidizing the metal atoms. But usually the oxides are directly sputtered from an oxide target. Nitride and other films can be sputtered in the same way.

4.11 Etching techniques

4.11.1 Wet etching

Wet etching is the use of liquid chemical solutions to remove material. III-V semiconductors can be etched to remove damaged material, to create certain structures, to assist polishing, as a cleaning process or as a diagnostic technique. Wet etching proceeds through chemical reactions occurring at the surface of the material. The etchant species must reach the surface (diffusion), the appropriate reaction must take place, and the resulting etch products must be removed from the surface (diffusion). If the diffusion steps are slower than the reaction itself, the process is defined as diffusion-limited or mass-transport-limited etching. In the opposite case the whole is defined as reaction-rate-limited, surface-limited, or kinetic-limited etching.

Wet etching

diffusion-limited

reaction-rate-limited

Diffusion-limited etchants tend to be more isotropic with respect to crystal orientation. They tend to give a more polishing effect as surface protrusions are more exposed

isotropic

to incoming reactant species and hence are etched more rapidly than a smoother surface. Stirring the solution can greatly affect mass transport, etch rates are highly sensitive to the form and degree of agitation.

anisotropic selective etching Reaction-limited etches tend to preserve surface morphology, although this tendency can be completely dominated by a second tendency: to be highly anisotropic in selectively etching crystalline structures through mask patterns. These etchants are less sensitive to agitation/stirring.

temperature sensitivity Etching is also sensitive to temperature. A 10°C increase in temperature may double the etch rate. Freshly mixed etchants may be hot due to an exothermic character of the dissolving process. Thus it is important to monitor the temperature of the mixed solution. Also etching in a small amount of solution can result in an increase of temperature during etching. Some etchants may change in composition during storage, such as H_2O_2. Some etchants may produce gas (H_2) bubbles, which can adhere to the surface and cause non-uniform etching. Other etchants may produce toxic by-products (etching of InP with HCl will produce phosphine, PH_3) which must be ventilated properly.

non-uniform etching
toxic

oxidation Almost all InP etchants operate by first oxidising the surface and then dissolving the oxide, thereby removing some of the In and P atoms. Hydrogen peroxide (H_2O_2) is the most oxidising agent used in etching III-V semiconductors. Typically a mixture of H_2SO_4/ H_2O_2/H_2O is an excellent etchant for InGaAs and InGaAsP. H_2O_2 or H_2SO_4 alone do not etch InGaAs(P). The oxidation by the H_2O_2 creates indium oxide (InOx) at the surface. The InOx is then effectively removed by H_2SO_4, and fresh semiconductor surface is exposed, allowing oxidation to take place again. The process continues iteratively. Similarly, etchants for InP typically involve H_3PO_4 as the oxidising agent (H_2O_2 cannot be used for InP due to a too low chemical potential) and HCl as the etchant.

selective etching **Selective etching.** Selective etching is a very powerful tool in semiconductor processing. It is useful in surface cleaning, in butt-joint regrowth and in creating MEMS structures. It occurs for etching solutions (single etchant or a mixture) which have sufficient chemical potential to one type of material, but not sufficient for another type. The example of H_2O_2 to InP mentioned above is a good example. For InP processing, it is well known that the H_2SO_4/ H_2O_2/H_2O is highly effective for InGaAs(P) but not for InP, while H_3PO_4/HCl is highly effective for InP but not for InGaAs(P).

anisotropic etching **Anisotropic etching.** The crystalline nature of III-V semiconductor lattices leads to anisotropic etching in almost all cases where masking materials are used for patterning. Figure 4.48(a) illustrates this anisotropic etching behavior. Depending on the orientation of the mask edge, one can obtain different etched sidewall profiles. This is a direct result of the lack of symmetry in the zinc-blende lattice and the dependence of the etch rate on the crystal orientation. The etching characteristics of InP tend to be dominated by the {111}A planes with the slowest etch rate. Contrary, {111}B planes have the highest etch rate. Therefore the etch tends to stop at {111}A planes.

etched sidewall profile
zinc-blende lattice
crystal orientation

[111]A plane
[111]B plane

The principle is explained in Figure 4.48(b). The {111} planes consist either of all In-atoms or all P-atoms and are referred to as {111}A or {111}B planes, respectively. The A and B planes have very different chemical properties. For the {111}A plane, In atoms are exposed. Each In atom is bonded to 3 P atoms, leaving no free electrons left. Therefore it is chemically harder to etch. For the {111}B plane, P atoms are exposed. After bonding with 3 In atoms, there are still two free electrons,which make the material chemically more reactive. For rate-limited reactions the etch rate will be ordered {111}B > {100} > {111}A. This description is simplistic and in practice, removal of

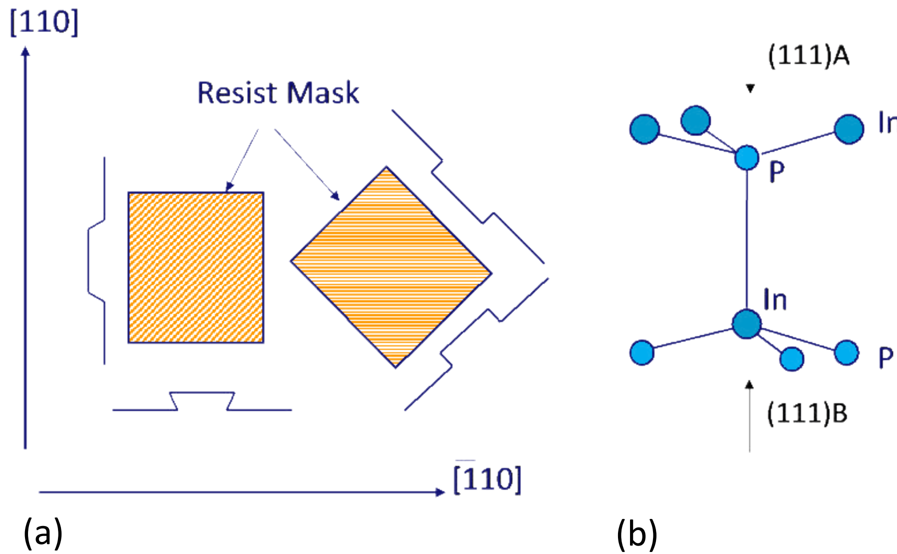


Figure 4.48: (a) Anisotropic etching behavior with a mask. (b) Illustration of two different (111) directions.

a surface-atom not only exposes the under-laying atoms but also allows lateral etching. This can lead to micro-faceting and roughening of the surface. *micro-faceting*

Anisotropic etching can be used to create a particularly desired profile. One can create either a V-groove or a dovetail profile in InP using bromine/methanol, simply by controlling the orientation of the crystal plane with respect to the mask opening [141]. This is shown in Figure 4.49(a). The principle can be further extended to complex InP/InGaAsP layer stacks, for example to create an efficient polarization converter with a triangular-shaped waveguide [142], as shown in Figure 4.49(b). In practice, it is also important to know that the etch profiles can be affected by other factors like the masking material itself and its adhesion to the surface. If the adhesion is week the etchant may creep under the mask. Post-baking the photoresist, after development, may affect the etch profile significantly. *V-groove profile* *dovetail profile* *bromine/methanol* *Post-baking*

4.11.2 Reactive ion etching

As discussed in Section 4.10.1, plasma can be generated from reaction gasses by applying external RF energy. Plasma contains original molecules as well as ions and radicals, along with many other components. The principles of the reactive ion etching (RIE) process are illustrated in Figure 4.50. Ions and radicals can be used to etch materials with different mechanisms. Radicals are single atoms or ions that normally only occur in larger molecules, where they share their valence electrons with other atoms. In radical form they possess loosely bound electrons, which makes them much more chemically reactive than their original molecule. Therefore they can be used for enhanced chemical etching. Ions can be accelerated by an external electric field and be used to bombard the material surface and contribute to physical etching (similar to sputtering). The physical bombardment also attacks masking materials, causing the erosion of the mask. In practice, both processes co-exist and contribute to the etching. The etch rate of RIE can be significantly increased as compared to cases where only one etching mechanism is employed [143, 144]. *reactive ion etching (RIE)* *radicals*

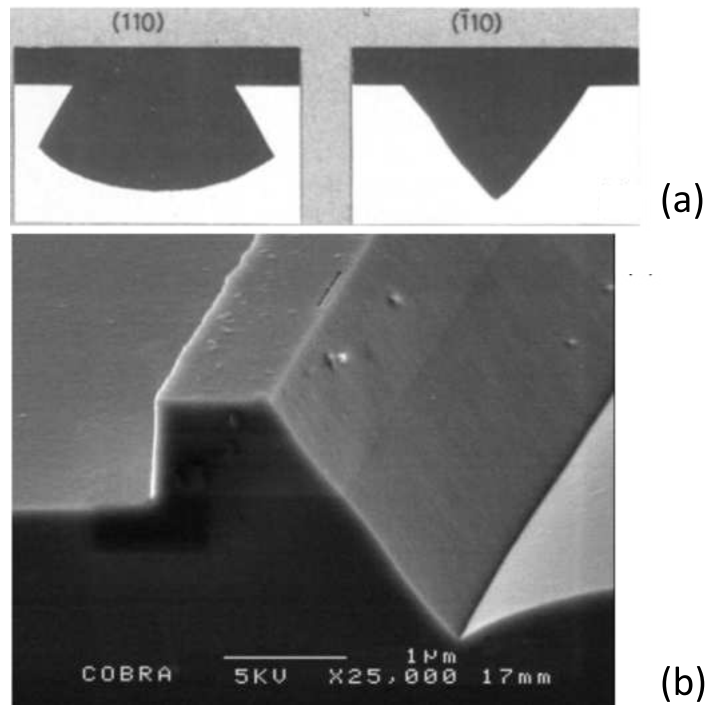


Figure 4.49: (a) Etching profiles of lnP(001) etched in bromine/methanol. from [141] (b) Triangular-shaped polarization converter.

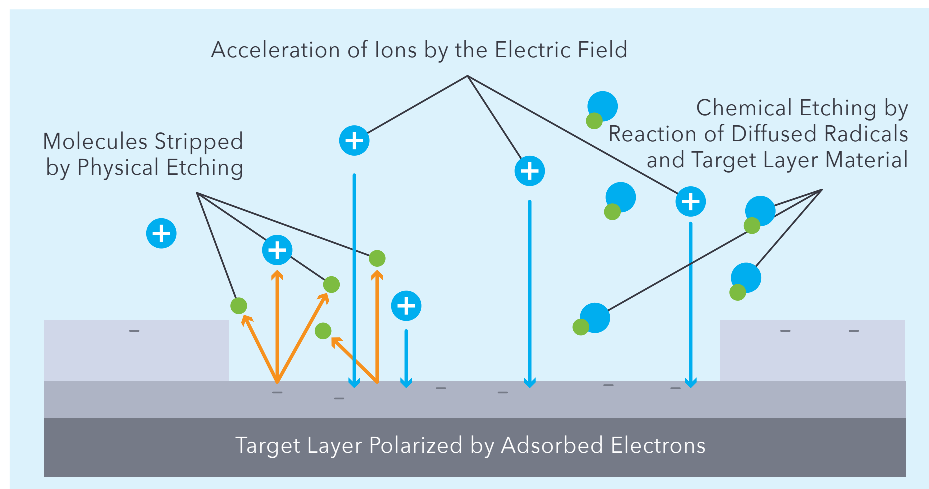


Figure 4.50: Principles of reactive ion etching.

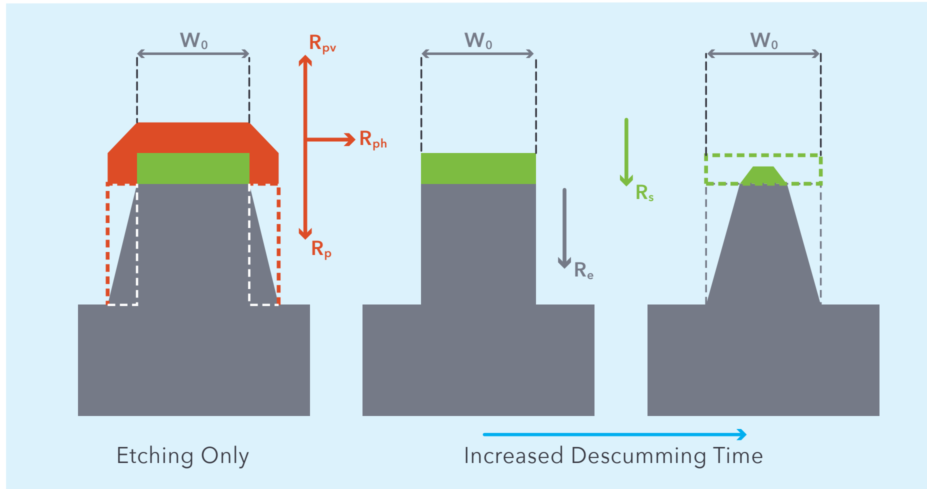


Figure 4.51: Illustration of the role of polymer passivation to the etched profile.

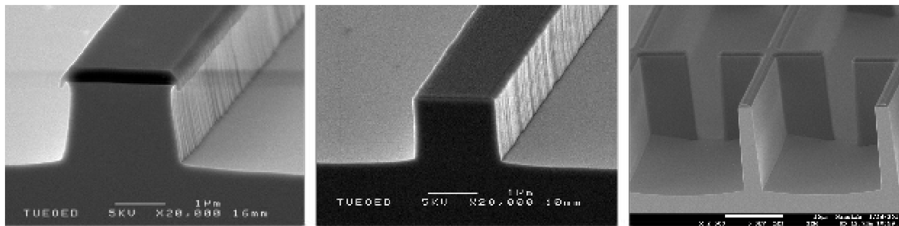
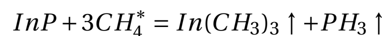


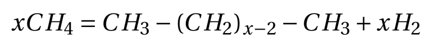
Figure 4.52: SEM pictures of 3 passivation scenarios in InP etching, from left to right: too much polymer passivation, balanced etching and passivation, and erosion of hard mask.

In many RIE processes, passivation of the etched surface is critical to achieving close-to-vertical profiles. Here the example of InP etching using a CH_4/H_2 gas mixture and O_2 polymer descum is discussed. CH_4/H_2 gas mixture is an effective choice to etch InP and its ternary and quaternary compounds. The reaction formula can be written as *polymer descum*



$$InP + 3CH_4^* = In(CH_3)_3 \uparrow + PH_3 \uparrow$$

The reaction generates two volatile products which can be removed easily by vacuum pumps. The "*" means a mixture of CH_3, CH_2, CH and H radicals and ions, all contributing to the reaction. During the same process, polymer will be generated as well. It can be expressed as



$$xCH_4 = CH_3 - (CH_2)_{x-2} - CH_3 + xH_2$$

This polymer will deposit not only on the etched surface (vertical sidewall and horizontal surface), but also on the mask. Thus, on one hand it may slow down or even block the etching process, on the other hand it can act as an extra protection of the mask against erosion. This is called passivation (note that it is a different concept than the surface passivation of a semiconductor surface to reduce dangling bonds and the corresponding surface recombination). A sequence of etch steps and O_2-descum steps, by ignition of an O_2 plasma for a very short time (seconds), can be used to balance the *passivation*

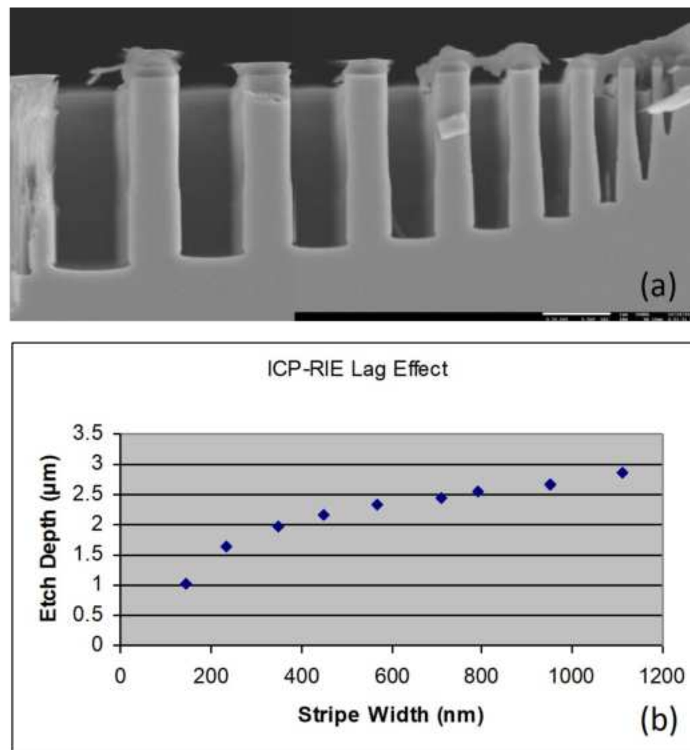


Figure 4.53: Lag effect in InP RIE process. (a) SEM image (b) Measured data.

polymer deposition, as illustrated in Figure 4.51. When there is no or little control of polymer deposition by O_2 descum, excess polymer may accumulate on the mask and create overhangs. This increases the width of the mask during etching, which leads to a non-vertical sidewall. When to much polymer is removed, the mask lacks extra protection, leading to mask erosion and reduction of the width (and thickness) of the mask. This will result in a non-vertical sidewall. Only a balanced polymer passivation yields a near-vertical sidewall. This principle is applied in the generic InP PIC process as used by TU Eindhoven. By optimising the strength and duration of the O_2 descum, a near-vertical sidewall of 89° is achieved, as shown in Figure 4.52.

lag effect Another important effect in the RIE process is the lag effect. As can be seen from Figure4.53(a), the etch rate in narrower mask openings is significantly lower than in wide openings. This is because the etching relies on diffusion of the etchants into the slot. For narrower slots, the etchants diffuse slower into the slot. Another scenario where lag effect can occur is due to the size of the mask. Etch rate in the nearby open area increases as the mask stripe width increases. This is due to the larger concentration of etchants caused by excess etch precursors diffused from the mask area to etched area. A quantitative measurement result of the effect is given in Figure4.53(b).

Last but not least, a good RIE process relies on a well prepared and clean surface. If any organic (e.g. residues of photoresist due to bad development) or inorganic (e.g. residues of SiNx hard mask due to incomplete etch) traces remain on the surface to be etched, they will act as local masks. As a consequence, one will observe spikes due to micro-masking after the etching process, as shown in Figure 4.54.

inductively coupled plasma RIE (ICP-RIE) Regarding the applied RIE processes, currently two are widely used. One is a basic RIE process, as shown in Figure 4.55(a). Another is inductively coupled plasma RIE (ICP-

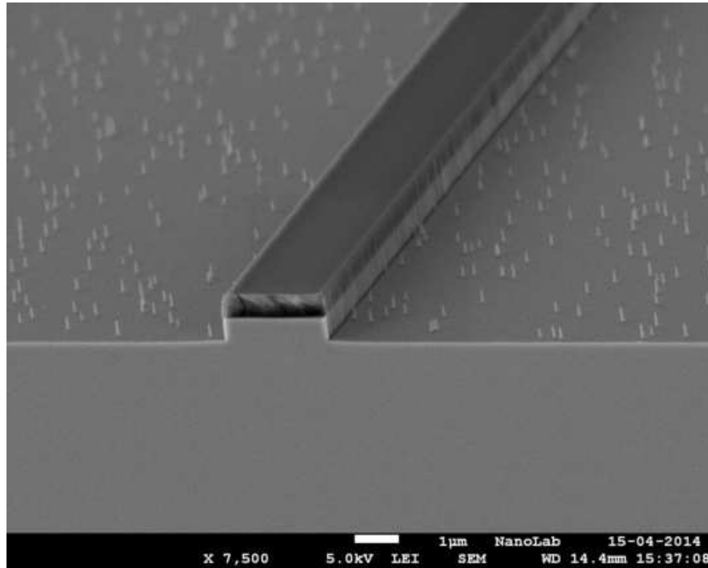


Figure 4.54: SEM image showing spikes caused by micro-masking around the etched waveguide.

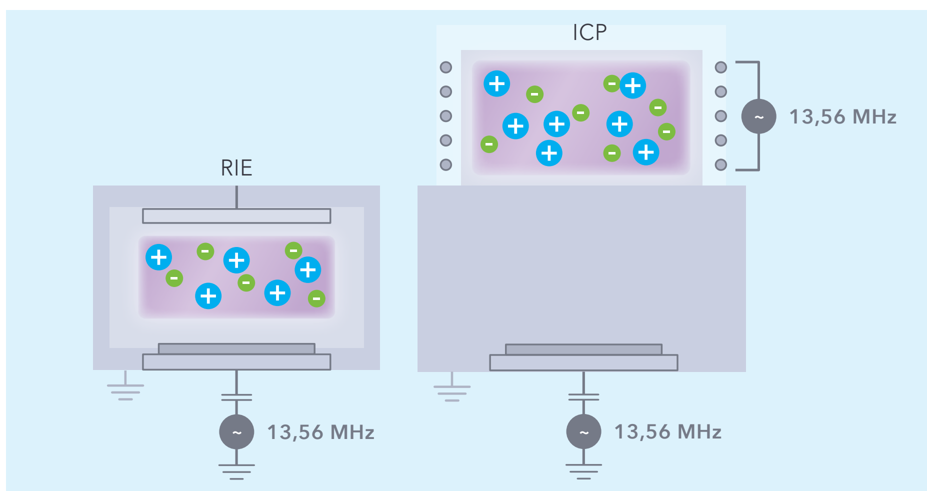


Figure 4.55: Configuration of (a) basic RIE machine and (b) ICP-RIE machine.

Table 4.9: Comparison of key parameters in RIE and ICP-RIE generated plasmas.

Parameter	RIE	ICP-RIE
Pressure p (mTorr)	10-1000	0.5-50
Power P (W)	50-2000	100-5000
Frequency f (MHz)	0.05-13.56	0-2450
Volume v (L)	1-10	2-50
Plasma density n (cm^3)	10^9-10^{11}	10^{10}-10^{12}
Electron temperatre T_e (V)	1-5	2-7
Ion acceleration energy ε_i (V)	200-1000	20-500
Fractional ionization X_{iz}	10^{-6}-10^{-3}	10^{-4}-10^{-1}

Problem: A Cl_2 and H_2 gas combination is also widely used to dry etch InP.

1) Write down all the products in the RIE plasma.

2) Explain which product(s) have sputtering effect, and why.

3) Explain which product(s) can enhance the reaction, and why.

4) Write down the reaction formulas.

Solution:

1) Molecules of Cl_2 and H_2; Ions Cl^+, H^+; Radicals Cl^*, H^*; electrons; visible and UV photons.

2) Ions contribute to the sputtering effect, because they will gain kinetic energy before reaching the wafer surface due to the electric field applied on them.

3) Radicals enhance the reaction because they contain highly reactive unpaired bonds.

4) $3\ Cl_2 + 3\ H_2 + 2\ InP = 2\ InCl_3 + 2\ PH_3$

Problem 4.3: RIE etching.

RIE), as shown in Figure 4.55(b). The principle of the basic RIE machine is similar to that of a sputter tool. Plasma is generated between two parallel plates by RF energy. Reactant gasses are injected into the chamber through a showerhead. The wafer is placed on the bottom electrode to get a good interaction with the plasma. For ICP-RIE, the major difference is the location where the plasma is generated. Instead of generating the plasma very close to the wafer in basic RIE, the plasma in ICP-RIE is generated remote by a magnetic coil. The generated plasma will diffuse to the wafer, which is located relatively far away. The main motivation is to balance the etch rate and plasma damage. It is obvious that a higher plasma density will lead to an increased etch rate. However the physical bombardment effect is also significantly enhanced which will cause roughness on the wafer surface and increased mask erosion. Further it may cause defects in the semiconductor material. For ICP-RIE, the kinetic energy of the ions reaching the wafer is significantly less, despite the fact that the plasma density is much higher. The key parameters of RIE and ICP-RIE generated plasmas are compared in Table 4.9. Based on these parameters, ICP-RIE is most widely used for deep etching and for etching structures with very high aspect ratio (narrow width and large depth).

ICP-RIE

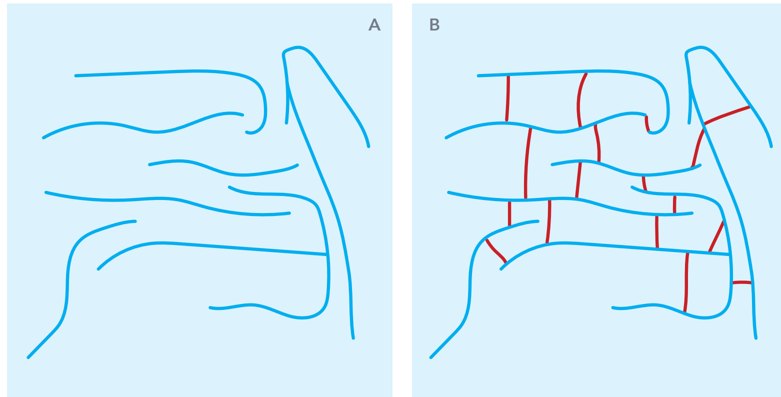


Figure 4.56: (a) Original polymer molecules (b) Cross-linking of the polymer

Table 4.10: Comparison of cross-linking polymers to photoresists.

	Photosensitive?	T_g(°C)	Cross-link temperature (°C)	Dissolve in
HPR504	Yes	~100	/	Acetone
MaN220	Yes	~100	/	Acetone
BCB 3022 series	Can be	> 350	150-300	/
Polyimide HD4100	Can be	> 400	200-350	/

4.12 Planarization

Planarization is an essential step in the generic InP process, as it creates a mechanical support and electrical isolation for the metal electrodes and interconnects. In the current generic InP process, planarization is achieved with cross-linking polymers such as polyimide (PI) or benzocyclobutene (BCB).

polyimide (PI)
benzocyclobutene (BCB)

A cross-linking polymer is a special type of polymer. With a proper heat treatment, original polymer molecules will bond to neighbouring molecules, forming one big molecule (so called cross linking). This is illustrated in Figure 4.56 below. The benefit of such cross-linking is obvious. The bigger molecule will be much more resilient to chemicals such as solvents and some acids. It will also be mechanically and thermally more stable than its original form. Some properties of PI and BCB are summarized in Table 4.10, and compared to commonly used photoresists based on small polymer molecules.

cross-linking
PI
BCB

Because of their excellent stability, both PI and BCB are used as planarization materials. Both can be easily spin-coated on a wafer. Due to the high topologies (> 2μm) in the generic InP wafer after the waveguide etching steps (see Figure 4.5(t)), one layer of PI or BCB (typically around 2 μm thick) may not be sufficient. Multiple layer planarization can be applied to improve the planarity. Figure 4.57 shows the improvement on the planarization as more layers of PI are applied.

4.13 Rapid Thermal Processing

Rapid thermal processing refers to a process that brings the wafer to a high temperature (several hundreds degrees Celsius) in a very short time. For the generic InP pro-

metal-to-semiconductor contact

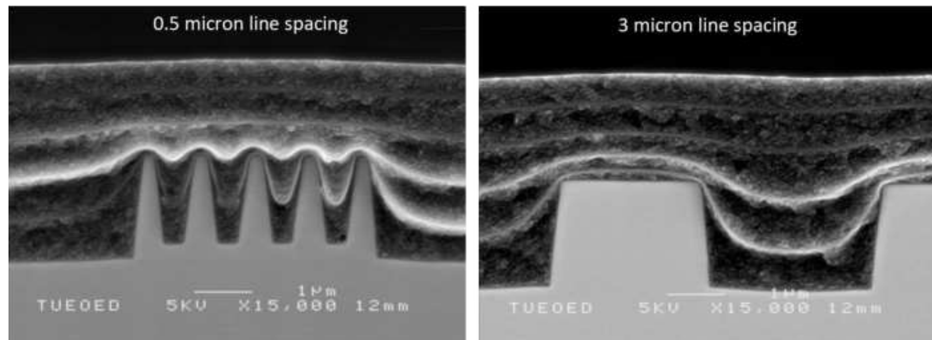


Figure 4.57: Planarization using multiple layers of 1 μm thick PI. A thin SiN layer is added intentionally between two layers of PI to enhance the contrast and make the interface between the PI layers visible in the SEM.

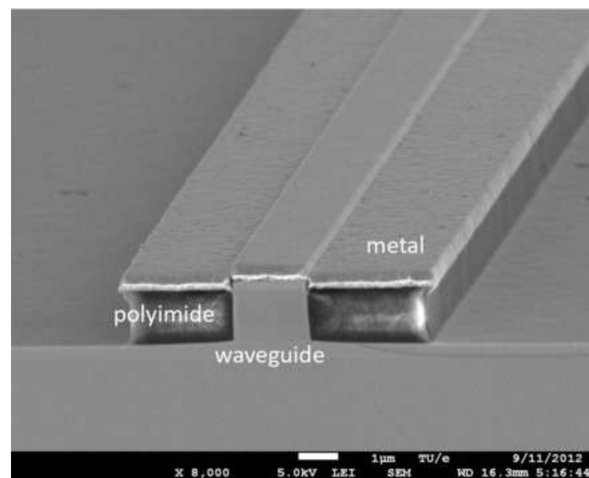


Figure 4.58: A fabricated SOA structure with a good planarization (6 layers of PI) for metal contacting.

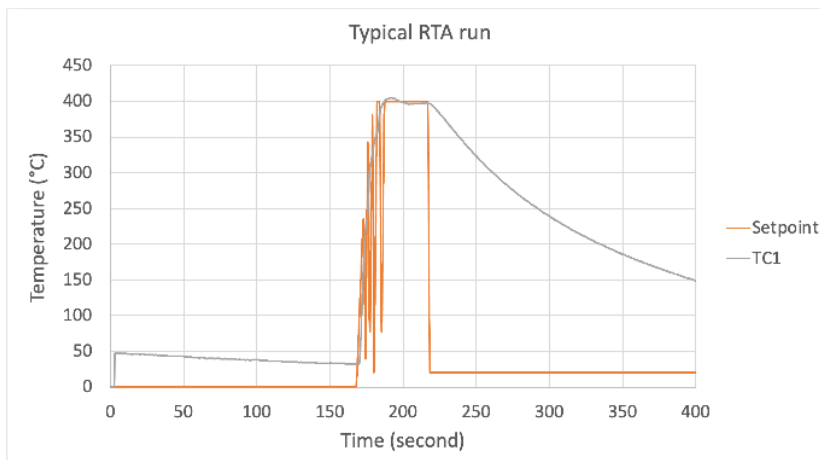


Figure 4.59: Typical temperature curve for an RTA process.

cess, it is used to anneal the metal-to-semiconductor contact to achieve low contact resistance. This particular process is called rapid thermal annealing (RTA). A typical temperature curve of the RTA process is shown in Figure 4.59. Because of the short rise time it is very difficult to avoid overshoot by optimizing the PID controller. We tackled the problem by heating with pulses. We incorporated a number of steps, alternating between a pulse of the lamps at a certain setpoint value, and lamps off (setpoint room temperature).

rapid thermal annealing (RTA)

The RTA process is most effective for moderately or low doped semiconductors, and less so for heavily doped semiconductors. For instance, for heavily doped p-InGaAs with doping concentration up to $2\cdot10^{19}\text{cm}^{-3}$, RTA does not reduce the contact resistance between p-InGaAs and Ti/Pt/Au metal stack much [145]. However the improvement in contact resistance is significant for moderately doped n-InP with doping concentration of $2\cdot10^{18}\text{cm}^{-3}$. The contact resistance is as high as $5\cdot10^{-5}\Omega\text{cm}^2$ without any RTA after metal evaporation and lift-off process. The value drops drastically to $2\cdot10^{-6}\Omega\text{cm}^2$ and $4\cdot10^{-7}\Omega\text{cm}^2$ when annealed at 350 °C for 15s and 400 °C for 15s, respectively [145]. Low contact resistance is key to the performance of active devices, as it is a significant source of heat generation in the diode as well as a limiting factor of the RC bandwidth at high speed.

contact resistance

4.14 Backend Processing

4.14.1 Cleaving and dicing

The fact that atoms in a crystal are arranged in planes is important for the mechanical properties of the material. For example, crystals can often be cleaved along some atomic planes, resulting in atomically flat surfaces. Semiconductors with zinc blende lattices present the lowest strength along certain directions, which are called cleavage planes. Si cleaves most easily along {111} planes while InP and GaAs cleave better along {011}.

cleavage plane

From a fully processed wafer into individual dies and chips, the wafer will experience a processed called dicing. There are many dicing techniques for different materials. For

die
dicing

InP, a wafer scriber is commonly used. The principle is simple: a small groove is first scratched at the edge of the wafer using a diamond tip. Under a small external force, the wafer will begin to cleave from the scratch and propagate along a favoured cleavage plane. The step is repeated until all chips are diced out.

4.14.2 Facet coating

high reflection coating (HRC)
antireflection coating (ARC)

The as-cleaved chips usually have an optical high-quality facet thanks to the cleavage along a crystal plane. Therefore no additional facet polishing is required. However some applications require the facet to have high reflectivity (e.g. $> 99\%$ for a Fabry-Perot laser mirror) or highly suppressed reflection (e.g. $< 0.1\%$ for an optical amplifier output). These require high reflection coating (HRC) and antireflection coating (ARC), respectively. The coatings also protect the waveguide facets against contamination. The simplest ARC is a quarter wave coating layer. However, it only works in a relatively narrow wavelength and incidence angle range. The performance can be significantly improved using multilayer coatings [146]. For a HRC, typically a stack of multiple quarter wave films is applied, and the reflectance mainly depends on the index contrast between the layers [147]. Commonly used materials for multilayer HRC and ARC coatings include Al_2O_3 and Si for high index and SiO_2 and MgF_2 for low index films. An E-gun evaporator is commonly used to deposit the layers. See for more information on optical coatings [148]