Chapter 1

Chemistry in the “Front End of the Line” (FEOL): Deposits, Gate Stacks, Epitaxy and Contacts

1.1. Introduction

The manufacturing of a transistor involves several hundred production and control stages and is a process that takes several weeks, generally between four and twelve depending on the complexity of the electronic component. Production is considered to be either “front end” or “back end”.

A “front end” plant will produce a component consisting of between hundreds of millions and several billion transistors, all interconnected via different stages. Today the most advanced technologies with 20 nm nodes use between eleven and thirteen metal levels to interconnect several billion transistors by means of a specific map that is determined by the feature of the component. The component is tested and the functional chips are identified and selected in order to calculate what is known as the yield, that is the number of chips made to specifications versus the total number made. A “back end” plant will receive the “front end” components and finalize the packaging through a second pathway and specific processes. Tests then

Chapter written by François MARTIN, Jean-Michel HARTMANN, Véronique CARRON and Yannick LE TIEC.
guarantee the reliability of the component, generally over a period of 10–20 years.

In the “front end” plant, the manufacturing of the component is itself divided into two main sectors: first, the “front end of the line” (FEOL), which could be considered as the technological heart of the operation. It involves the extreme optimization of each individual transistor: dimensioning, type (N or P), arrangement, characteristic (high performance or low consumption, for example) and application (logic, analog, etc.). The second area of manufacturing is known as the “back-end of the line” (BEOL) and will be explained in Chapter 2 of this book, simplified, its aim is to optimize the interconnection of the individual transistors made by the FEOL sector.

This chapter focuses on the FEOL sector: it can be divided into several subsectors of activity; through processes, we will successively list zones for deposition, mechanical–chemical polishing, dry and wet etching, surface preparation/cleaning, doping or lithography. For each technological node, there are groups of experts aiming to define the specifications required for these different processes, in order to obtain the desired features for various applications. This is done by the international group, International Technology Roadmap for Semiconductors (ITRS) [ITR 12].

A modern transistor must address the various technological aspects that are shown in Figure 1.1. The challenges concern the starting material “A” (generally silicon, eventually modified to a version called silicon on insulator (SOI) corresponding to an ingenious Si/SiO₂/Si stack where the highest active layer of silicon is isolated from the substrate by an embedded layer of oxide), the insulation between the transistors (B), the determination of the dopant profiles (target addressed by points C, E and G), the gate stack (D and F), the contact zones of the gate as well as the “Source” and “Drain” zones of the transistor (H and I), etc. [ITR 11].

This chapter describes three important elements. In Figure 1.1, these are described mainly by points F, H and I, and somewhat by D and E as well. First, the chemistry involved in the gate stack is described as this is directly related to a fine tuning of the transistor: the optimization of the gate insulator and the gate itself. The choice of the chemical precursors is of key importance. Next, crystalline heterostructures (based on SiGe in particular) are explained and the involvement of the precursors, temperature, pressure
and surface preparation is detailed. Finally, the chemistry involved to address efficient contacts is also precisely described.

![Figure 1.1. The challenges of the “front end of the line” as defined by the ITRS [ITR 11]](image)

1.2. Arrangement of the gate

1.2.1. Generalities

After the discovery of the planar structure in the 1960s, for a long time the gate stack of the metal oxide semiconductor (MOS) transistors was made of a silicon (SiO$_2$) insulator covered by a gate for the channel control (itself made of N and P doped polycrystalline silicon). Looking at the chemical processes used in the production of this stack, first a thermal oxidation of the silicon takes place in the heat treatment furnace containing several tens of wafers (reaction [1.1]). Then the silicon gate is deposited by chemical deposition in the vapor phase low pressure chemical vapor deposition (LPCVD) from silane in a furnace under vacuum (reaction [1.2]).

\[
\text{Si} + (O_2 \text{ or } 2 \text{H}_2\text{O}) \rightarrow \text{SiO}_2 \text{ (+2 H}_2\text{)} \]  

[1.1]
\[
\text{SiH}_4 \rightarrow \text{Si} + 2 \text{H}_2 \]  

[1.2]

For dimensions of about 0.18 μm and smaller, additional constraints arose: the continuation of transistor size scaling fixed the thickness of the insulation gate (Moore’s Law) and while the current tunnel through the oxide increased drastically, the boron doping diffused from the P-type metal-
oxide semiconductor (PMOS) side through the gate insulator to the transistor’s channel. This chapter will explain how scientists first improved the properties of the silicon oxide, before introducing into the 45 to 28 nm nodes an innovation that would replace a large part of SiO₂ by a high-permittivity insulator, known as “High K”: by depositing a greater thickness of a material with a higher permittivity, the leakage current is indeed reduced without adversely affecting the capacitive coupling. To guarantee a greater physicochemical compatibility with the High K material as well as an optimal capacitive coupling, the second major innovation was to deposit a metal gate, rather than a gate made from doped polycrystalline silicon. The latter had the disadvantage of having a slightly depleted interface doping, adding a parasitic capacity, which adversely affected the transistor’s performance. In this section, the chemical mechanisms that come into play inside the fine layers of the gate stacks will be defined as well as the precursors used for depositing the films in the manufacturing environment. The motivation behind the choice of the High K material and the interaction with its environment will only be described briefly (for more information, the reader may refer to review papers about this topic) [WIL 01, LOC 06, CHO 11].

1.2.2. Silicon nitriding processes

1.2.2.1. Thermal nitridation

In the history of the technological breakthroughs in this field, the first requirement was to increase the resistance of the silicon oxide to electrical tension (its reliability) and against the diffusion of the boron through the insulator. Before the 90 nm nodes, boron diffusion was controlled by an optimization of the implantation conditions as well as by the activation annealing of the polycrystalline silicon gate and junctions. Processes used to increase the reliability of the Si/SiO₂ interface were investigated by introducing into the 2–3 nm oxide layers a very specific quantity of nitrogen. Experiments were first conducted by annealing silicon oxide in the presence of NH₃, thus nitriding the oxide surface by optimizing the temperature of the process; however, the presence of reducing species resulting from the thermal decomposition of NH₃ at high temperatures caused the apparition of electrical traps in the silica. This disturbance was resolved by carrying out complex reoxidizing steps on the nitrided oxide, thus increasing the difficulties for controlling the process.
Next, scientists introduced nitriding molecules that contained no hydrogen. The first attempts were carried out with N$_2$O; however, it appeared that its complex decomposition involves the nitridation/oxidation of the interface, giving a Si–N–O bonding state, which is difficult to control uniformly over a batch of wafers [HEG 97]. By identifying that the main product of N$_2$O decomposition was, in fact, nitrous oxide (NO), it seemed easier to introduce it directly into the annealing furnace. As the diffusivity of the NO molecule is dependent on the insulator’s growth conditions (and its stress) [SAG 96], the annealing process was optimized in order to obtain the necessary nitrogen concentration in the internal interface, depending on the silicon’s thickness. In addition, the bonds formed at the interface were of the Si–N type [HEG 97] and so by lightly reoxidizing it, the nitrogen was then in an environment of the same type as that formed by treatment with N$_2$O, as shown by X-ray photoelectron spectroscopy (XPS) [HAL 98]. With this sort of process, approximately 0.1% nitrogen may be obtained in an optimal chemical environment [WIL 01], that is when conditions lead to the improvement of the interface’s characteristics without degrading any other properties of the transistor, such as mobility and leakage current.

1.2.2.2. Plasma nitridation

From the 90 nm technological nodes up to the advent of High K materials, a more efficient nitridation process was developed, which granted an increase in permittivity by nitriding the silica to the core. It improved the balance in the compromise between the electric thickness of the insulator (Equivalent Oxide Thickness (EOT)) and its leakage current, since permittivity increases with nitrogen incorporation, while providing a more efficient barrier to the diffusion of the boron. As the nitridation process in the presence of NO only allowed a reduced amount of nitrogen, approximately $1 \times 10^{13}$ atm/cm$^2$ at the Si/SiO$_2$ interface, a new insulator-nitriding technology was required. It needed to be able to introduce a given percentage of nitrogen (between six to 15% depending on the devices) while preserving the silicon interface from nitridation for the reasons explained previously (degradation of transistor properties). The new process uses a low-energy nitrogen plasma (in the order of 1 eV), allowing nitrogen implantation in the surface and the volume of the 1.5 to 2.5 nm-thick oxide layers. The experimental procedure for this process is extensively documented [CUB 04]: the nitriding plasma first incorporates the nitrogen into the silicon oxide in the form of three Si–N bonds and unstable Si–N=O complexes. The latter are then removed by lightly oxidizing annealing,
which reorganizes the Si/SiO₂ interface. This must be achieved without exposure to air and so is carried out in a single-wafer treatment system that contains the nitriding plasma reactor as well as the annealing reactor. Otherwise, non-reproducible exposure to the atmosphere gives way to a reaction of non-stabilized nitrogen atoms to form N₂O or NO [CUB 04]. This technology did not evolve until 40 nm nodes for the low-power technologies. To scale further electrical oxide thickness while keeping a low gate leakage current, High K/metal stacks were introduced from the 45 nm nodes for microprocessor technologies (high performance) and from the 32 nm nodes for the “low-power” technologies.

1.2.3. The introduction of the High K/metal gate stacks

1.2.3.1. The choice of High K materials in a conventional process

The selection of the best High K material for use as an insulation gate was the subject of intensive sampling by the end of the 1990s. The principal criteria of success being, besides a sufficiently high permittivity (at least four times that of SiO₂), a thermodynamic stability in relation to the silicon substrate [LOC 06], a sufficiently large forbidden band and the positioning of the bands in relation to the silicon (known as the band “overlap” of the insulator), at least 1.5 eV, respectively, above or below silicon conduction band energy (Ec) and silicon valence band energy (Ev) to limit the leakage current [ROB 00]. The material presenting the best compromise in relation to all these properties is hafnium oxide, HfO₂, mainly due to its thermal stability with underlaying silicon that sets it apart from ZrO₂ (Zr is in the same column as Hf in Mendeleyev’s periodic table but is less stable on silicon/thin silicon dioxide couple). However, following an extensive study of the properties of the material, scientists decided in favor of hafnium silicate HfₓSi₁₋ₓO₄, a compromise between SiO₂ and HfO₂. Indeed, the aim was first and foremost to maintain the compatibility of the High K material with a polysilicon gate, which is not performed by HfO₂, due to the interaction between reducing species (resulting from the decomposition of the silane) with Hf–O bonds. Silicon in hafnium silicate also increases the crystallization temperature of the insulator compared with that of HfO₂; since crystallization does not interact with the leakage current [KIM 04a] in very thin films, researchers quickly discovered that it is generally responsible for the creation of defects that affect the mobility of the channel carriers [ZHA 06, KIR 06a]. Adding even small quantities of silicon is
favorable to suppress crystallization [KIR 06a]. The same also applies, more
generally, to the decrease in thickness of the High K, stopping the nucleation
of the crystallites [ZHA 06]. Thermal stability of the hafnium silicate is
improved by plasma nitridation. Later this also increases the permittivity of
hafnium silicate as well as its chemical compatibility with the metal gate.
However, unlike plasma nitridation of silicon dioxide described earlier, very
unstable Hf–N bonds are also produced and must be eliminated during the
annealing process. This is known as “post-nitridation anneal (PNA), and is
carried out to leave only nitrogen–silicon bonds [INU 05]. The control of the
nitridation plasma parameters is critical, particularly the pressure; indeed the
nitrogen atoms replace oxygen atoms which diffuse into vacuum at a low
working pressure. On the contrary, at higher pressures, it diffuses till Si/SiO2
interface and contributes to an increase in the thickness of the interfacial
silicon oxide layer, thus diminishing the global electric capacitance of the
insulator [INU 05, KAM 08a]. The optimization of a nitrided hafnium
silicate increases the final gain on the leakage current by about two orders of
magnitude in relation to the equivalent SiON electrical thickness, in the
1.2 to 1.3 nm equivalent oxide thickness (EOT) range (corresponding to a
28 nm “low-power” node), in spite of a reduction of the HfSiO band-gap on
the valence band side due to the addition of nitrogen [BAR 06]. A metal gate
compatible with the insulator must also be introduced, providing the same
work function as N+ and P+ doped polycrystalline silicon (respectively ~4.1
and 5.15 eV).

1.2.3.2. Constraints and possibility of High K/metal integration

1.2.3.2.1. “Gate first” assembly

The assembly that consists of maintaining polysilicon on silicate HfSiON
was indeed only attempted in transistors whose principal objective was
limiting the leakage current (application of the dynamic random access
memory transistors (DRAM), rather than improving overall performance. In
complementary metal oxide semiconductor (CMOS) logic, scientists first
researched the insertion of the metal gate on the High K insulator in the
standard way where the gate is made before the activation of the junction
dopants (a rapid thermal treatment carried out above 1,050°C). It seemed
that controlling the potential of the gate at the edge of the silicon band could
not be done by simply using metals with corresponding theoretical
potentials. This impossibility is due to the interaction of the heated metal
with the High K insulator, namely its chemical affinity for the oxygen,
which induces a complex phenomenon: the creation of charged vacancies in
the High K [UED 06] and/or the modification of the interfacial SiO₂ layer [BER 10]. These phenomena lead to a value of the final work function close to the silicon mid-gap [AKI 07], which makes the fabrication of the metal for the PMOS transistor in the range of fine electrical thicknesses (about 1 nm) impossible. This problem was partially solved for the N-type metal-oxide semiconductor (NMOS) transistor by introducing a dipole at the IL/HfSiON interface [KIR 08]. The most frequently used elements are lanthanum for the NMOS transistor [KIR 08] and aluminum for the PMOS transistor [JUN 05] on the condition that the latter operates with an FDSOI-type substrate (see Figure 1.2, [WEB 10]): the channel is made from undoped silicon that releases the specifications to 4.8 eV for PMOS and 4.4 eV for NMOS transistors. On bulk silicon, where the channel is highly doped, it is impossible, even with a dipole, to pilot the required work function near 5.1 eV; the channel’s potential is then modified using a SiGe alloy [KRI 11] (see section 1.3). Moreover nitrogen incorporated into the insulator (voluntarily and mainly by diffusion from the grid after high-temperature annealing) plays an essential role in the mobility control and the reliability of the transistor [GAR 08]. The integration of a High K/metal stack in a conventional “gate first” scenario is then actually feasible but quite complex for nodes beyond 28 nm. Then major manufacturers of integrated circuits chose to modify the transistor assembly completely in order to avoid these high-temperature interactions between the metal gate and the insulator.

1.2.3.2.2. “Gate last” assembly

A “gate last” assembly uses a “dummy” polysilicon gate during the high-temperature sequencing of the process before being replaced by a gate made from High K/metal. It was used by Intel® (see Figure 1.3) for 32 nm “high-performance” transistors [PAC 09]. In these stacks, the gate potential of the NMOS transistor is set by a TiAl alloy deposited on thin TiN [VEL 11]. The lack of high-temperature interaction between the metal and the insulator, HfO₂, means it is easier to obtain lower electrical thicknesses [RAG 05] and to adjust the PMOS gate potential with titanium nitride [PAC 09]. Looking at the subsequent technological nodes (20 nm followed by 14 nm), it is obvious that manufacturing a “planar” gate stack insulator (which has existed for nearly half a century) equivalent to 0.6 nm silica is difficult. The solution used nowadays involves a three-dimensional deployment of the transistor to increase the capacitive coupling surface [WU 10]. Others may prefer solutions where the SiO₂ interface layer (IL) is reduced to the very minimum due to the “scavenging” phenomenon [KIM 04b], based on a “suction” of the
High K’s oxygen through a metal gate with a high affinity for oxygen: the High K material, which is thermodynamically more stable than silicon, compensates for the loss of oxygen by reducing the underlying SiO₂ IL and by releasing silicon into the substrate. We also mention the attempts aiming to modify the crystalline structure of HfO₂ in order to increase the value of its permittivity, either by stabilizing in a cubic or quadratic phase, rather than monoclinic. The apparition of this phase may be obtained by adding another compound, such as a small quantity of silicon [MIG 08], or by forming an HfZrO alloy [TRI 08]. Let us note that phase stabilization is a complex, stress-dependent phenomenon, and therefore depends on the gate structure and the thermal budget of the device [TRI 06].

1.2.3.3. Chemistry of High K deposition processes

After having described the evolution of the gate integration sequences as well as the role played by solid chemistry in silicon manufacturing, this section will describe the chemical precursors used for deposition of the layers mentioned previously.

1.2.3.3.1. Deposition by CVD

Metal organic chemical vapor deposition (MOCVD) allowed the deposition of hafnium silicate from Si₂H₆ together with a hafnium alkoxy, the Tetrakis(tert-butoxy)Hafnium (Hf(OtBu)₄) (TBH) diluted in octane. This method involves the sectioning of the alcohol ligands by the silicon hydride, leading to the formation of an oxide with a 60% Hf/Hf+Si ratio, without any addition of oxygen in a “batch” furnace at low temperature (260°C), and so
without any risk of regrowth of the SiO₂ interface sublayer [INU 05]. However, it is critical to control the mechanisms of deposition over a large surface in the furnace, which is why the single wafer systems have become more efficient for High K oxide deposition during the last decade. The first criterion for hafnium precursor selection is the capability to perform deposition at a controlled rate, which will allow for a sufficient uniformity in the thickness on the wafer. It must have a high enough vapor pressure, a property that is generally obtained from the liquid phase, either heated or not, to deliver a sufficient quantity of the precursor into the reactor. The second criterion is the thermal stability of the precursor, which guarantees the absence of a particle producing solid residue [GIR 07]. Based on the latter, hafnium’s most frequently used precursor is an ethylamine, tetrakis diethylamino hafnium (see Table 1.1) only used for HfO₂ deposition [VAN 04a] or with amino precursors of silicon such as tris(dimethylamino)silane (TDMAS) or bis(dimethylamino)disilane (BDMAS) [GRO 12, KAT 08] in the presence of oxygen in a range of deposition temperatures greater than 500°C to avoid carbonated contamination [VAN 04a]. Octane is always used as a solvent as it distributes small quantities of the chemical precursor into the reactor. More sophisticated processes also allow the injection of less-volatile precursors and so improve the CVD deposition recipe by injecting pulses of reactant into the oxygen with an optimal frequency [COS 04]. It is still a CVD process but the deposition is controlled to the monolayer (known as “AVD™”) and is close to the way of deposition known as ALD (atomic layer deposition), which will be detailed in the next section.

<table>
<thead>
<tr>
<th>Metal</th>
<th>Amines (Main ligand)</th>
<th>Cyclopentadienyl</th>
<th>Alkoxy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hf</td>
<td>TDMAH</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>mixte</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>ethyl</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Si</td>
<td>n-DMAS (n=2–4)</td>
<td>TDEAH</td>
<td>HTB</td>
</tr>
<tr>
<td>Zr</td>
<td>TDMAS</td>
<td>ZrCp₂Cl₂</td>
<td></td>
</tr>
</tbody>
</table>

Table 1.1. The main organometallic precursors described in this chapter for gate stack deposition; the others are chlorides: MCl₄ with M=Ti, Zr, Hf; the latter two are solids, TiCl₄ is liquid
Table 1.1. (Continued) The main organometallic precursors described in this chapter for gate stack deposition; the others are chlorides: $MCl_4$ with $M$=Ti, Zr, Hf; the latter two are solids, $TiCl_4$ is liquid

<table>
<thead>
<tr>
<th>Ti</th>
<th>TDMAT</th>
<th>TEMAT</th>
<th>TDEAT</th>
<th>TiCp$_2$Cl$_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ta</td>
<td>PDMAT</td>
<td>TAIMATA</td>
<td>TBTDET</td>
<td>Mg(EtCp)$_2$</td>
</tr>
<tr>
<td>Mg</td>
<td></td>
<td></td>
<td></td>
<td>La(i-PrCp)$_3$</td>
</tr>
</tbody>
</table>

1.2.3.3.2. ALD: atomic layer deposition

For gate application in microelectronic technology, a specific method was introduced to deposit High K insulator: ALD in which deposited gate oxide films typically have a thickness of 1.5–2.5 nm, plus or minus one atomic layer. Unlike CVD where the reactants are simultaneously injected into the reactor, the metallic precursors and the oxidizing reactants are separately injected by pulses in ALD, with an inert gas purge between the two (see Figure 1.4). Dedicated reactor designs have been developed to minimize the time taken by the pulse and purge while avoiding any possibility of a mixture of reactants and oxidizing reagents, leading to a CVD parasitic growth. Here, the deposition kinetic is controlled by precursor and reactant chemisorption mechanisms on the surface sites. In the case of an oxide deposition, hydroxyl groups are used whose surface density decreases by recombination when the temperature is increased. Then, unlike CVD deposition, the ALD deposition rate will decrease with any increase in reaction temperature. The total quantity of precursor (or reactant) injected over the course of each cycle must be optimized in order to cover all the reactive sites in as short a time as possible (typically between a few tens or hundredths of a millisecond).

Figure 1.4. Schematic of the ALD deposition process: a) pulse and chemisorption of the metal precursor, b) purge of the inert gases, c) pulse and chemisorption of the reactant and d) reactant purge
Figure 1.5. An amine molecule with a cyclopentadienyl ligand used for zirconium oxide deposition (R being an organic radical), from [GIR07]

The optimal temperature range for chemisorption and desorption reactions is between 200 and 350°C as it allows the use of suitable metallic precursors: they must be stable inside this range and optimized in terms of solid residue, such as cyclopentadienyl amine derivatives, called MyALD (with M = Hf or Zr, see Figure 1.5), or ozone, O3, as an oxidative reactant [GIR07]. These precursor pairs have been developed specifically to manufacture metal-insulator-metal (MIM)-type capacitive structures that are not, however, inside the scope of this chapter: applying ozone to the gate insulator deposition indeed helps to start the reaction on a deoxidized surface [KIR 06b], however, it leads finally to a thicker SiO2 interface and so water is the preferred reactant even if it does require the presence of an initial silicon dioxide layer to start the reaction. There are many studies covering the advantages of different reactors and precursors, but it is acknowledged that if using water as the oxidative reactant it is simpler to use hafnium tetrachloride, HfCl4, as the hafnium precursor. Others such as tetrakis dimethylamino hafnium (TDMAH) and tetrakis ethylmethylamino hafnium (TEMAH) react with H2O, but have a much lower thermal stability (leading to an increase in dry residue [GIR 07]), meaning the reactions must be carried out below 300°C. Then carbonated residues will form in the layers, leading to a decreased film density [TRI05].

Hafnium chloride, which is a by-product of zirconium purification and is relatively of low cost, is a solid precursor (unlike amines) and has a vapor pressure of approximately 1 torr at 160°C [PAS 63], compatible with the doses injected during each ALD cycle. Dedicated sources working in this temperature range have been developed to deposit hafnium oxide at a higher
reaction chamber temperature, in order to avoid any possible recondensation of the chloride. Their design takes into account the extremely high reactivity of HfCl₄ in humid conditions, guarantees a sufficient amount of precursor entering the reactor for a continued delivery of reactant, especially in the case of deposition on larger areas than the chip itself. Indeed, if the gates are present as cavities (see Figure 1.3), the total amount of precursor added must be increased depending on the actual developed surface of the chip, and must also take into account the time required for the diffusion of the precursor/reactant pair inside the cavities during the pulses. Models of this process have been investigated in detail [STO 06]. Due to its present use in industry, specific details of the HfCl₄/H₂O chemistry will be examined here, as well as the behavior of the residues of chemical precursors inside the films during the following process steps. Indeed, chlorine was identified early as a contaminant and then mixed chlorinated cyclopentadienyl (C₄H₄FCl₂) molecules were developed to reduce its effects. Results, however, were disappointing as the gain achieved was countered by an increase in carbonated contaminants [NIN 06].

1.2.3.3.3. HfO₂ ALD deposition with HfCl₄

The effect of the deposition parameters. The deposit mechanism based on HfCl₄/H₂O on silica was established in an experimental reactor [AAR 99] and globally corresponds to the chemisorption on surface hydroxyls, given in reaction [1.3]:

\[
\text{HfCl}_4 + n (-\text{OH})_s \rightarrow n (-\text{O} \cdot \text{HfCl}_{4-n}) + n \text{HCl} \quad \text{with } n = 1 \text{ or } 2 \quad [1.3]
\]

Next is the hydrolysis of the Hf–Cl bonds by water, producing once more HCl as well as an Hf–OH-type surface, which may then host another cycle. The actual mechanisms, which are more complicated, are modeled in terms of growth per cycle (GPC) [PUU 04]. The chemisorption mechanisms on the substrate during the first cycle have led to much research into the potentially crucial role of the heterointerface in the electrical properties of the part. Researchers also found that the first deposition cycle leaves a greater quantity of chemisorbed chlorine [BLI 03, TAL 08] than a permanent deposition regime (approximately ten times more, see Figure 1.6). This is even more pronounced on a chemical (rather than thermal) oxide, due to readsoption of HCl by the silica’s hydroxyl groups. The chemisorption of chlorine increases with the decrease in reaction temperature and may affect the charge density of the oxide interface [SRE 06]; by increasing both the
time taken by the water pulses and the deposition temperature, the surface and volume concentration of chlorine may be reduced [BLI 03, DEL 06] (see Figure 1.7). These steps are necessary as it is then harder to eliminate chlorine afterwards by annealing, especially at the SiO$_2$/HfO$_2$ interface [SRE 06, FER 02] where a silicate-like compound is identified [REN 02] before the appearance, after a few cycles, of the permanent deposition regime, corresponding to a speed of approximately 0.055 nm of HfO$_2$ deposited per cycle at 300°C. The chlorine and hydroxyl residues have a direct effect on the properties of the layers: they have an unstable structure, as shown by the strong capacity of the amorphous HfO$_2$ to exchange oxygen during the annealing process at low-temperature as shown by O$^{18}$ annealing experiments [GAN 04]. They will also increase the crystallization temperature [VAN 04b] as well as the regrowth of the SiO$_2$ oxide IL during the following annealing processes at 650°C [FER 06] (see also Figure 1.8). The undesirable regrowth at high temperature of the interfacial layer is reduced when annealing is carried out under vacuum rather than at atmospheric pressure [BAR 04]. Recent studies have developed a more specific, two-stage, desorption of the hydroxyls during the vacuum annealing of HfO$_2$: first the surface groups at 350°C, then those bonded to the core of the layer at 500°C [RAG 09]. This technique controls the thickness of the SiO$_2$/HfO$_2$ bilayer with an EOT of less than 1 nm.

![Figure 1.6. TXRF measurement of the total amount of chlorine adsorbed versus temperature and the number of ALD cycles (HfCl$_4$/H$_2$O) on a chemical silica surface, from [BLI 03]](image)

Precursors for alloy and functional layer deposition. In section 1.2.3.2, the possibility of modifying the structural properties of the High K layer (the crystallinity), or fixing the work function of the gate stack with an La- (Mg) or Al-based dipole was mentioned. A group of precursors that allow the deposition of these alloys and functional layers exists (see Table 1.1): for
example HfZrO₄, first obtained from organometallic precursors [TRI 08] is also deposited with chlorinated Hf and Zr precursors and water reactant, while the hydrolysis of SiCl₄ permits the addition of Si in HfO₂ using ALD [FED 07]. Aluminum oxide is deposited using trimethylaluminum due to its high vapor pressure, which allows the liquid source to be maintained at ambient temperature. As for the possibility of depositing a fine layer of lanthanum by ALD, a first approach based on diketonate/ozone [NIE 01] was discarded (due to the solid source with a low vapor pressure and the formation of carbonates in the deposit) in favor of cyclopentadienyl compounds, typically La(i-PrC₅p)₃. Similarly, thin layers of MgO have been deposited using Mg(EtCp)₂ [KAM 08b].

![Figure 1.7](image1.png)

**Figure 1.7.** Chlorine’s profile in HfO₂ and at the interface with silicon measured by secondary ion mass spectroscopy (SIMS) versus the deposition temperature and water pulse time, from [BLI 03]: the effect of increasing the water pulse time at 350°C is shown by the black arrow.

![Figure 1.8](image2.png)

**Figure 1.8.** View by transmission electron microscopy of the HfO₂ bilayer and the interfacial layer: on the left hand side, before annealing, there is a slightly abrupt interface corresponding to a silicate; on the right hand side, after annealing in the presence of N₂ at 800°C, the crystalline HfO₂ and the interfacial SiO₂ layer are perfectly distinct.
1.2.3.4. Deposition of the metallic gate and associated precursors

The most innovative gate last stacks (see Figure 1.3) require physical vapor deposition (PVD)-type deposition, especially in cases of non-conformal layers (e.g. no deposition on the walls of a cavity) or non-stoechiometric deposition. These kinds of deposits are outside the scope of this chapter that will mainly explain the chemistry of the processes. In this framework, CVD or ALD deposits of conventional materials for gate stacks will be described, namely titanium and tantalum compounds or nitrides. The most frequently used precursors for metal compounds are typically the same structures of amines and chlorides molecules as used for oxide deposition, but with an nitridizing reactant molecule (see Table 1.1). The nitride deposition temperature could be less critical, unlike the insulator deposit, where the oxidizing compound may directly influence the regrowth of the interfacial SiO₂ layer. However, the diffusion of the nitridizing species and its interaction with the insulator may damage certain properties [GAR 08]; so depositions at moderate (<500°C) temperatures will be favored. In the next section, the molecules that may be used in industrial depositions of metallic gates will be described.

1.2.3.4.1. Chlorinated precursors

Similarly to the chlorinated solids used for hafnium and zirconium oxide deposition, liquid titanium chloride is used to deposit TiN by the CVD and ALD processes. The latter method is preferred for reasons already described, that is a deposition temperature less than 500°C. Unlike with chloride hydrolysis, their nitridation with ammonia is less energetically favorable [RIT 99], so the deposition temperature must be above 400°C and the pulse time is increased until saturation of the Ti–Cl bond nitridation reaction occurs to facilitate the desorption of the less-volatile compounds (which may saturate the ulterior chemisorption sites of TiCl₄); the two half-reactions, instead of [1.4] and [1.5], involved are:

Chloride chemisorption (only one site):

\[ \text{TiCl}_4 (g) + \text{NH} (s) \rightarrow \text{N-TiCl}_3 (s) + \text{HCl} (g) \]  \hspace{1cm} [1.4]

Nitridation of the Ti–Cl bonds (simplified):

\[ \text{N-TiCl}_3 + 3 \text{NH}_3 \rightarrow \text{N-Ti-N}_3 + 3 \text{HCl} \]  \hspace{1cm} [1.5]
We finally end up with a low deposition speed (~0.02 nm/cycle) and relatively long cycles [RIT 99], which limit the application of TiCl₃ to thin films (batch treatment of wafers are in this case more favorable for the throughput). Thanks to their higher deposition rate, the organometallic precursors were more rapidly integrated into manufacturing for deposition of metal compounds and were introduced first into interconnection levels such as barrier materials or electrodes for capacitors.

1.2.3.4.2. Organometallic precursors

MOCVD and CVD metallic nitride deposits use amino compounds from Table 1.1, which, in the case of tetravalent titanium, have the same structure as those used for hafnium, namely the tetrakis-dimethylamino titanium (TDMAT), tetrakis-ethylmethylamino titanium (TEMAT) and tetrakis(diethylamino) titanium (TDEAT) (tetrakis amino compounds). Pentavalent tantalum also generates a family of amino compounds: the metal is linked either to five nitrogen atoms in the pentakis dimethyaminotantalum molecule (PDMAT) (Ta[N(CH₃)₃]₃) [MUS 09] or to four nitrogen atoms including one double bond in terbutylimido-tris(diethylamino) tantalum (TBTDET) (NEt₂)₃Ta=NBu [PAR 02]; a methyld variation Ta(N-t-C₃H₁₁)[N(CH₃)₃] is still known by its more commercial name, TAIMATA [PAR 03] (see Figures 1.9 and 1.10).

![Figure 1.9. TBTDET molecule, tantalum’s precursor](image)

![Figure 1.10. TAIMATA molecule](image)

However, CVD- and ALD-made deposits (with these organometallic precursors) contain carbonated residues and have a lower density than would be expected for the material. This gives way to oxygen diffusion, for example when using TDMAT and NH₃ [ELA 03, FIL 05]. Their specific resistance is high, particularly in the case of tantalum nitride layers. In fact,
the carbonated contamination is decreased by increasing the length of the titanium ligands, that is by using TDEAT rather than TDMAT [KIM 03], or by a section of tantalum precursor like TAIMATA ligands by hydrogenated silicon compounds, but leading to the formation of TaSiN [NAR 04]. To reduce the carbonated contamination of the layers, a plasma treatment can be applied after the deposit. Finally, the films are now formed directly by plasma-enhanced atomic layer deposition (PEALD), similarly to ALD, but with a reactive cycle including hydrogenated and/or nitriding reducing plasma [MUS 09]. Using this method, the mechanisms leading to the formation of a layer of TiCₓNᵧ are described in [CAU 08] and in the following reactions, [1.6] and [1.7]:

Adsorption of TDMAT on amino sites:

\[
2 \text{NH} + \text{Ti[N(CH₃)₂]₄} \rightarrow \text{N₂Ti[N(CH₃)₂]₂} + 2 \text{HN(CH₃)}₂ \quad [1.6]
\]

Ligand division by hydrogenated plasma:

\[
\text{N₂Ti[N(CH₃)₂]₂} + \text{H₂ plasma} \rightarrow \text{TiCN} + \text{CH₃NH₂} \quad [1.7]
\]

The titanium, initially bonded to four nitrogen atoms, in its final state can be bonded to a carbon atom via a transportation mechanism where a Ti–N bond is substituted by a carbon from a methyl ligand [CAU 08]. The plasma power conditions allow the control of the concentrations of Ti₂C and TiN. These mechanisms give similar results with tantalum precursors such as TBTDET [PAR 02], leading to a mixture of Ta–C and Ta–N, again with the plasma controlling the proportions. Equally, pure metal deposits are possible but only in the presence of hydrogenated plasma from chlorinated metals and at very low deposition speeds [KIM 02]. For an application to gate stacks, it is crucial to remember that the gate insulator is a sensitive material, which may be structurally modified in the presence of active hydrogen plasma [GAR 05]. Effects on HfO₂ have been proven after PEALD of a TaCN gate, combined with the apparition of interfacial states (depending on the type of plasma used) [PAR 08]. The manufacturing of structures that include PEALD deposits in contact with the gate insulator is probably rather difficult to achieve, except perhaps in the top part of the stack, when the insulator is already protected from the active species generated during the metal deposition.
1.2.4. Conclusion

This section non-exhaustively describes the manufacturing challenges of an advanced CMOS gate stack, and identifies the role played by chemistry in the growth and transformations of the gate insulator and metal electrode. Compared with processes used since the invention of planar technology, a rise in material innovation has taken place on gate stacks in these last few years. The control of the film properties of the nanomaterial thickness’ critical functions was done with much intense research and development: first identifying the molecules and processes that were most favorable and then optimizing the deposition equipment for the treatment or modification of layers. The development of stable chemical precursors is crucial, at the lowest possible costs and with a high level of purity. They illustrate how, in the last decade, chemistry has taken an important role in the critical stages of chip manufacturing, which are key elements in our communication tools (telephony, computers). This continuous innovation in process chemistry allows controlling the manufacturing of billions of transistors on a single chip: the thickness has a precision of approximately 1 Å (a tenth of an nm) on the thinnest layers such as the gate insulator of these transistors. Compared with processes used in the previous decades, critical deposits are being carried out at lower temperatures by ALD, a deposition method that may be extended to other applications not mentioned in this chapter (for instance MIM capacitive insulation or insulation on emerging semiconductor materials). The reduction of the thermal budgets shows how the energy consumption of the manufacturing process is being better managed for integrated circuits, an approach that will certainly take more and more importance in the next coming years.

1.3. Chemistry of crystalline materials

1.3.1. Generalities

In [BAU 11], Chapter 4, entitled “Epitaxy of Strained Si/Si$_{1-x}$Ge$_x$ Heterostructures”, describes in great detail the reduced pressure – chemical vapor deposition (RP-CVD) of Si/SiGe(C) heterostructures for nano- and optoelectronics [BAU 11]. That chapter consists of the following parts: section 4.1 describes: (1) the evolution of Si and SiGe industrial
epitaxy by CVD, (2) one of the two epitaxy tools used by CEA-LETI for research purposes and (3) a few general notions of epitaxy, which may be useful later on. The work performed on Si/Si$_{1-x}$Ge$_x$ and Si/Si$_{1-y}$Cy strained heterostructure epitaxy for use as MOSFET conduction channels is detailed in sections 4.2 and 4.3. The optimization of the selective epitaxial growth of Si and SiGe:B raised Sources and Drains (RSD) on SOI substrates is described in sections 4.4 and 4.5. In section 4.6, the structural properties of SiGe virtual substrates used as templates for the fabrication of tensily strained SOI (or sSOI) substrates are described alongside the electrical gains they lead to (namely because of the dual channels of these structures). The electrical and optical properties of devices based on thick layers of pure Ge grown on blanket and patterned wafers are described in section 4.7. In section 4.8, the architectural possibilities offered by the lateral selective etching of SiGe versus Si are detailed. Finally, the most important findings are summarized in section 4.9 and several perspectives are presented.

The aim of this section here is to focus on two points that have so far scarcely been mentioned in [BAU 11]: (1) the preliminary Si (or SiGe) surface preparation used prior to epitaxy of monocrystalline layers and (2) the various advantages and drawbacks of gaseous (or liquid) precursors used to deposit high-quality Si/Si$_{1-x}$Ge$_x$ heterostructures.

1.3.2. A few basic ideas about epitaxy

Rather than proceeding directly to the results, this section will give some basic definitions and facts about epitaxy. To carry out an epitaxy, a monocrystalline layer is deposited that will adopt the same crystalline structure as the monocrystalline substrate on which it is deposited (which is the case of the semiconductors from the III–V or IV–IV columns, which crystallize in the diamond or sphalerite structure such as GaAs, Si and Ge), that is a structure dictated by the underlying substrate, sometimes with rotations in the crystal’s direction (such as for metals on insulating substrates, etc.). This is known as homoepitaxy when the substrate and the layer are of the same type (e.g. Si on Si) or heteroepitaxy when they are not (e.g. SiGe on Si).

Intrinsic or in situ doped Si/SiGeC stacks (known as heterostructures) are most of the time grown on Si substrates with a (100) surface orientation.
Those substrates can be bulk-type or not (presence of a buried oxide layer), blanket or patterned, etc. It may be useful at this stage to mention the main characteristics of the elementary compounds of a SiGeC alloy: Si, Ge and C crystallize in the diamond phase, that is two overlapping sphalerite structures shifted one from the other by a quarter of the large diagonal (see Figure 1.11). The lattice parameter of C (i.e. the length of one of the sides of the cube in Figure 1.11) is much smaller than that of Si, which in itself is much smaller than that of Ge (a_C = 3.567 Å ⇔ a_Si = 5.43105 Å ⇔ a_Ge = 5.65785 Å). The lattice parameter increase is accompanied by a sharp decrease in the energy bandgap: E_g (C) = 5.48 eV ⇔ E_g (Si) = 1.11 eV ⇔ E_g (Ge) = 0.66 eV. Mixing Si, Ge and C in a SiGeC alloy will induce strong band-gap modifications; built-in stress will also play a role. Incorporating C by substitution into a Si or Ge matrix is otherwise quite difficult. The miscibility of C in Si is only approximately 10^{-3}% at thermodynamic equilibrium (i.e. at 1,400°C), with an unfortunate tendency toward SiC precipitates formation at higher concentrations. Si and Ge, however, are miscible in all proportions. During heteroepitaxy, there will be discrepancies between the lattice parameters of the deposited layer and the substrate, which will lead to an accumulation of elastic energy in the layer. For thin layers, the substrate’s atomic columns will be extended into the pseudomorphic epitaxial layer. A layer is described as being in tension (or compression) when its lattice parameter a_L is lower (or higher) than that of the substrate a_S. For a pseudomorphic layer, the in-plane lattice parameter of the layer is equal to that of the substrate (i.e. a_L^// = a_S). For a layer in tension, we then have a_L^− < a_L < a_L^// = a_S; for a layer in compression, a_L^+ > a_L > a_L^// = a_S. The perpendicular lattice parameter is given by a_L^⊥ = a_L + 2c_{12}/c_{11}(a_L - a_S), with 2c_{12}/c_{11}(Si) = 0.77, 2c_{12}/c_{11}(Ge) = 0.74 and 2c_{12}/c_{11}(C) = 0.61. However, when the thickness of the layer increases, it can become energetically favorable to introduce misfit dislocations to minimize the elastic energy that has accumulated and accommodate the lattice parameter mismatch between the layer and the substrate. The layer is then said to be plastically relaxed. The thickness at which the transition from pseudomorphic to plastically relaxed layer occurs is called critical thickness for plastic relaxation. Schematic diagrams in Figure 1.12 show the different configurations of the {epitaxial SiGe layer on Si substrate} system.
Through the 1980s, researchers from several institutes proved that the lower the epitaxial temperature, the higher the critical thickness for plastic relaxation of SiGe on Si(001) was (kinetic barrier to the formation and propagation of dislocations in metastable layers). Values 10 times larger than those predicted by equilibrium theories were obtained [VAN 75, HOU 91]. Solid source molecular beam epitaxy (SS-MBE) was used in all these studies for the growth, between 400 and 750°C, of variable thickness Si\textsubscript{1-x}Ge\textsubscript{x} layers (with a concentration of Ge between 15 and 70%). Their degree of pseudomorphicity with the underlying Si(001) substrate was determined by X-ray diffraction (XRD) and/or by Rutherford back scattering [BEA 84, HOU 91]. Researchers from the American Bell laboratories were then able to propose the well-known “People and Bean” relationship giving the critical thickness for plastic relaxation \( h_c \) at 550°C as a function of the lattice parameter mismatch \( f \) between SiGe and Si [PEO 85, PEO 86]:

\[
\frac{h_c}{d} = \frac{1}{f} - \frac{1}{f_c}
\]

where \( d \) is the critical thickness, \( f_c \) is the critical mismatch, and \( f \) is the actual mismatch.
Recent studies would suggest that $h_c$ is higher than the values predicted by People and Bean [BUC 06, TOM 10].

\[ h_c \equiv \left( \frac{1.9 \times 10^{-2}}{f^2} \right) \ln \left( \frac{h_c}{4A} \right), \]  

[PEO 86] with \( f = \frac{a_{\text{SiGe}} - a_{\text{Si}}}{a_{\text{Si}}} \approx 0.042x \)  \[1.8\]

We have thus revisited the critical thickness for plastic relaxation of SiGe on Si(100). For this, 200 mm Si(001) substrates (of a higher crystalline quality than those used in the studies referenced in [HOU 91], that is with a lower pre-existing dislocation density) were used as templates for the RP-CVD of variable composition and thickness SiGe layers.

The growth temperature was reduced from 700 down to 550°C as the Ge concentration increased from 12 to 52% in order to reduce surface roughness. Because of the differences between SS-MBE and RP-CVD in terms of Si and Ge precursors (solid ingots of Si and Ge, which are sublimated by MBE $\Rightarrow$ ultrapure gaseous precursors for CVD), from the way very low partial pressures of impurities such as oxygen or water are obtained (ultrahigh vacuum in MBE $\Leftrightarrow$ large fluxes of ultrapure hydrogen in CVD), etc., we would expect differences in findings [BEA 84, HOU 91, HAR 11a].

Figure 1.13 shows the summary of the results [HAR 11a]. The dotted line shows those of Bean [BEA 84], the colored squares correspond to SiGe layers seen as fully compressively strained in XRD and the squares with a
cross in their center represent partially relaxed SiGe layers. $h_c$ is in fact two to three times higher than predicted by People and Bean. However, for high Ge concentrations, certain layers seen as pseudomorphic in XRD are actually characterized by the presence of a limited amount of misfit dislocations in their midst. The surface signature of their propagation (“plough” lines along the $<110>$ crystallographic directions) was detected using atomic force microscopy (hatched zone shown in Figure 1.13).

1.3.3. **Surface preparation prior to epitaxy**

The substrate surface preparation that precedes any kind of epitaxy is of paramount importance. The goal is indeed to elongate the atomic columns of the substrate into the layer. The presence of amorphous silicon oxide layers (even very thin layers), polymers or etching residues (on masked substrates) on the surface has indeed a crippling effect on the quality of the epitaxy. Depending on the starting configuration, two strategies may be used to obtain Si surfaces free of oxide.

The starting substrate may be bulk Si or, to a lesser degree, an SOI-type substrate (i.e. a thin layer of Si/thermal buried layer of SiO$_2$/Si substrate stack, usually fabricated using the Smart Cut™ technique [BRU 95]). It is then possible to carry out high-temperature bakes, during which dozens of liters of ultrapure H$_2$ are injected into the reaction chamber (typically only a few ppb of impurities will be present because of dedicated gas purifiers), in order to remove the 0.8 to 1 nm layer of native or chemical SiO$_2$, which is initially present on the surface and smooth it [BEN 94]. The term chemical oxide is meant to design the oxide that is created by oxidizing wet cleanings such as RCA or dynamic diluted clean (DDC) [TAR 99, ROU 02], which remove surface particles, organic and metallic contaminants. The simplified chemical reaction for the removal of the oxide layer is:

$$\text{SiO}_2(s) + 2\text{H}_2(g) \rightarrow \text{Si}(s) + 2\text{H}_2\text{O}(g)$$

[1.9]

An image taken by scanning tunneling microscopy (STM) of a Si(001) surface after DDC cleaning, followed by *in situ* H$_2$ bake at 1,100°C for 1 min, is shown in Figure 1.14. The surface is smooth at the atomic level (average roughness: 0.05 nm only). There are monoatomic steps, alternately smooth ($S_A$ type) or rough ($S_B$ type) [MO 91, VOI 97], limiting (1 × 2) or (2 × 1) reconstructed terraces slightly larger than 100 nm. The extended
holes observed on the terraces are likely due to partial desorption of the silicon atoms during the H$_2$ bake.

![Image](image.jpg)

**Figure 1.14.** STM image (500 nm $\times$ 500 nm) of a Si surface (001) after DDC cleaning followed by H$_2$ bake at 1,100°C and 20 Torr for 1 min. The insert shows a (2 × 1) surface reconstruction.

The requirement for high thermal budgets (typically higher than 1,050°C for 1 min; 1,100°C for 2 min as the CEA-LETI standard) prohibits the use of this surface preparation on patterned wafers (i.e. with active Si zones surrounded by dielectrics) or ion-implanted substrates. The active zones could facet due to the strain generated by masking dielectrics, especially if they are in a Shallow Trench Isolation (STI) configuration. An exodiffusion of preimplanted atoms and a consequent autodoping of the epitaxial layer may also take place [CHA 85, JER 99]. In the case of SOI substrates, even blanket substrates, this type of bake is prohibited for thin Si films ($<20$ nm). If these criteria are not followed, moat recess and islanding of the films will take place [ISH 02]. It is also wise to optimize the power delivered to the lamps used to heat up the SOI substrates so that the temperature at the surface is uniform; otherwise slip lines will form.

The surface preparation described here (where a native or chemical oxide present at the surface is removed by *in situ* H$_2$ bake in suitable conditions) is far from being universal given the high thermal budgets needed. In order to minimize the latter, the following sequence is then used:

1) The native SiO$_2$ is removed by an “HF-last” wet cleaning [ABB 04]. During the next to last step of such a cleaning, the wafer is dipped in hydrofluoric acid diluted in deionized and deoxygenated water
(the dilution is typically between 0.2 and 1%), to etch the surface SiO$_2$. The dissolution reaction is as follows:

$$\text{SiO}_2 + 4\text{HF} \rightarrow 2\text{H}^+ + \text{SiF}_6^{2-} + 2\text{H}_2\text{O} \quad [\text{BUH 97}]$$

[1.10]

The etch rate of thermal SiO$_2$ (the most resistant of all) is of the order of 1.2–1.4 nm/min for 0.2% HF. During the last step, the wafer is rinsed in deionized and deoxygenated water to remove all traces of HF. The wafer is then dried by isopropyl alcohol (IPA) vapors (at 190°C), before being charged as quickly as possible in the load-lock chambers of the epitaxy tool, in an inert atmosphere (purified N$_2$). After such a wet cleaning, approximately 85% of the Si dangling bonds are occupied by hydrogen atoms. The remaining 15% are mainly occupied by fluorine atoms as well as oxygen and carbon contaminants [MEY 90, TRU 90].

2) An H$_2$ in situ bake then takes place at temperatures close to 900°C to remove all O, F or C surface contaminants (if compatible with the technology it will be used in). A surface perfectly passivated by hydrogen atoms is then obtained, which is ideal for epitaxy.

An “HF-last” surface is only stable for a few dozens of minutes up to a couple of hours (before the regrowth of a native oxide a few Å thick can be detected by spectroscopic ellipsometry). Minimizing the time between an “HF-last” wet cleaning and the loading of wafers inside chambers purged with high-purity inert gases (such as N$_2$) is thus mandatory. A perfectly hydrogen-passivated Si surface obtained by a high-temperature H$_2$ bake is by contrast stable for several days [HER 01].

We will now focus on the impact of the H$_2$ bake temperature on the properties of the epitaxial layers and the interfacial contamination, for Si, SiGe and pure Ge surfaces of (100) and (110) crystallographic orientations.

The secondary ion mass spectrometry (SIMS) depth profiles of the C, O and F atoms present in 50 nm-thick Si layers grown at 650°C on Si (001) substrates after an “HF-Last” wet cleaning followed by a 2 min, 20 Torr H$_2$ bake, either at 750 or 775°C, are plotted in Figure 1.15 [ABB 04].

A temperature of 750°C is too low to get rid of F and O atomic contamination. Peaks are indeed present at the interface between the substrate and the epitaxial layer. No such peaks were observed at 775°C, however (the same as findings in [KOR 08]). These results are in full
agreement with those of Brabant, who showed that 750°C H₂ bakes lasting as long as 10–20 min did not completely rid the Si(001) surface of O contamination [BRA 03].

**Figure 1.15.** SIMS depth profiles of the O, C and F atomic concentrations in epitaxial Si layers grown at 650°C on blanket Si(001) substrates, after a “HF-Last” wet cleaning and H₂ bake at 750°C (filled symbols) or 775°C (open symbols) for 2 min (P = 20 Torr)

**Figure 1.16.** STM images (500 nm × 500 nm) of Si(001) surfaces after a “HF-Last” wet cleaning (a), followed by H₂ bakes at 750°C (b) or 775°C (c), 20 Torr for 2 min

500 nm × 500 nm STM images of the Si (001) surface after (1) “HF-last” wet cleaning followed by in situ H₂ bakes for 2 min at 750°C (2) or 775°C (3) are plotted in Figure 1.16. After an “HF-last” wet cleaning, the surface is
structured. After H$_2$ bake at 750°C, short terraces delimited by bi-atomic steps appear. After bake at 775°C, the surface is atomically smooth (average roughness: 0.09 nm), with, as in Figure 1.14, the appearance of monoatomic steps that are alternatively smooth (S$_A$ type) and rough (S$_B$ type) and that delimit several dozens of nm wide (2 × 1) or (1 × 2) reconstructed terraces [VOI 97].

It is interesting to note that the H$_2$ bake threshold temperature (between 750 and 775°C for Si (001)), above which the C, O and F surface contamination is eliminated, does not seem to be influenced by the crystallographic orientation of the Si surface. A low interfacial O peak was evidenced by Destefanis et al. [DES 08] after the H$_2$ bake at 775°C, for 2 min (after “HF-last” cleaning) of Si (110) substrates. This peak disappeared entirely at 800°C.

Similarly, changing, for (001) surfaces, from Si to SiGe or even pure Ge has little impact on the threshold temperature. O and F contamination peaks were evidenced by Abbadie et al. [ABB 04] after H$_2$ bakes at 775°C for 2 min (post “HF-last” cleaning) of SiGe 20 and 33% virtual substrates (VS) prior to their encapsulation with thin, tensily-strained Si layers [ABB 04]. These contamination peaks disappeared at 800°C. H$_2$ bakes at 750°C for 2 min enables us to get rid of interfacial contamination peaks in pure Ge (low temperature re-epitaxy of Ge on thick, cyclically annealed Ge layers, themselves on Si (001); [HAR 04a]); there is, however, a low O peak after H$_2$ bakes at 700°C.

The temptation here would be, given this information, to systematically adopt a high H$_2$ bake temperature (typically between 800°C and 900°C), for relatively long times (2 min), in order to be safe. This is not always possible.

Let us first deal with SiGe VS grown on Si (001), that is thick, linearly graded SiGe layers (used to confine misfit dislocations) capped with nearly fully relaxed constant Ge composition SiGe layers (several microns total thickness). The surface of such SiGe VS is usually mechanically and chemically polished prior to any re-epitaxy (of, for example, thin, tensily strained Si layers), in order to get rid of the <110> surface cross-hatch inherent to such stacks. A H$_2$ bake with a too high thermal budget leads to a resurgence of the surface roughness [HAR 08a]. When the H$_2$ bake temperature increases, the same type of surface roughening has clearly been shown for thick epitaxial layers of Ge grown directly on Si (001) and then
polished [CLA 06]. The optimal compromises are, for virtual SiGe substrates, shorter H₂ bakes (15 s) at 850°C, and for thick layers of Ge on Si (001), H₂ bakes at 750°C for 2 min. For these two types of stacks, the “HF-Last” (where HF is hydrofluoric acid) wet cleaning used prior to H₂ bake and epitaxy must also be optimized. Indeed, (1) the etch rate in most of the solutions used in microelectronics to get rid of particular or metallic contamination increases exponentially with the Ge concentration [ABB 06] and (2) the HF concentration must be increased as the Ge concentration increases in order to obtain hydrophobic SiGe surfaces (i.e. without oxides) [ABB 08].

On ultrathin SOI substrates with mesa isolation, moat recess at the edges of the active Si zones and even some islanding of the Si film may occur if the thermal budget of the in-situ H₂ bake is too high. The H₂ bake is typically carried out at 650°C, 20 Torr for 2 min prior to the selective epitaxial growth of Si-raised sources and drains on ultrathin SOI (i.e. with a starting Si layer thickness of 3–4 nm) [JAH 05]. Too high a H₂ bake temperature may also lead, for MOSFET transistors, to a detrimental regrowth of the interfacial oxide layer present between the Si channel and the high permittivity gate dielectrics (such as HfO₂) or a detrimental diffusion of ions that had previously been implanted to dope the extensions.

What about the structural and electronic properties of the epitaxial layers on partially contaminated Si surfaces? What of the structural and electronic properties of epitaxial layers on partially contaminated Si surfaces? The impact of a H₂ bake temperature which is too low to completely remove O, C or F interfacial contamination (i.e. between 650°C and 750°C) is limited on (001) surfaces. The surface stays smooth after epitaxy [DES 08], Si/SiGe/Si stacks are of high crystalline quality in X-ray diffraction, the electrical performances of fully depleted transistors built on top of SOI or SiGe/Si dual channels [LE 11] with Si [AND 07] or SiGe:B [BAU 10] raised sources and drains are excellent, etc. However, H₂ bakes at too low a temperature (i.e. T < 775°C for 2 min at 20 Torr) have a catastrophic impact on the crystalline quality of epitaxial Si layers grown on Si(110) substrates [DES 08].

Whatever the surface epitaxy is carried out on, specific care must be taken during lithography and etch steps to remove all polymers, High K dielectrics (oxides from actual transistor gates), residues, etc. Failure to do so delays growth, prevents epitaxy from taking place if the H₂ bake
temperature is too low, etc. [HAR 11b]. It may even lead to a significant deterioration of the layer crystalline quality [LIU 12].

As a conclusion for this section, let us note that it is possible to carry out the removal of the native or chemical SiO₂ not only *ex situ*, in automated wet benches (wafers processed either individually or in batches) [KOR 08], but also *in situ*. HF vapors in a dedicated chamber connected to the epitaxy cluster tool reactor can be used to get rid of surface oxide [KUI 92]. Dry plasma etching based on NH₃ and NF₃ gases in a “Siconi” chamber can also be used to that end [YAN 10].

### 1.3.4. Low-temperature Si and SiGe growth: the comparison of three precursors (silane, disilane and dichlorosilane)

Dichlorosilane (SiH₂Cl₂) is the silicon precursor of choice for the selective epitaxial growth of Si and SiGe in the Si windows of patterned Si substrates, the masking layers being dielectrics such as SiO₂ or Si₃N₄. This Si gaseous precursor is used for the growth of recessed or raised Si or SiGe:B sources and drains [HAR 04b, HAR 08b, BAU 10, HAR 11a], SiGe/Si epitaxial bilayers that constitute the core of localized silicon-on-insulator-type devices [MON 10] or the high-mobility channels of advanced p-type MOSFET transistors [HUT 10], etc. Dichlorosilane does, however, have relatively low growth rates at low temperatures [HAR 07], as we will see later on. Hydrogenated Si precursors such as silane (SiH₄) [HAR 02], disilane (Si₂H₆) [HAR 12] or trisilane (Si₃H₈) [BAU 07, GOU 09, TAK 10, VIN 10] yield much higher Si and SiGe growth rates at low temperatures. These precursors are not selective versus dielectrics, however. Monocrystalline Si and SiGe layers then grow in the Si windows and polycrystalline layers on the masked zones. This difficulty may be overcome with advanced cyclic deposition/etch (CDE) processes [LOU 12, HE 12]. During a CDE cycle, a few nanometer thick monocrystalline layer is then non-selectively deposited over the wafer surface. This layer is monocrystalline in the Si windows and polycrystalline on the masking dielectrics. The latter is removed selectively during the etch step that follows (with Cl₂ [BAU 12a] or a mixture of HCl + GeH₄ [BAU 12b]). By repeating the cycle several times, a perfect selectivity may be obtained for several tens of nanomolar thick layers. The use of hydrogenated precursors may be advantageous if a low thermal budget (and thus chemically abrupt interfaces) or high concentrations of substitutional carbon atoms are aimed for.
Trisilane, which yields very high growth rates at low temperatures, has, however, several flaws: (1) it is extremely expensive to produce with a quality high enough for nanoelectronics and (2) being liquid, it requires the use of a dedicated bubbler (in which high partial pressures of H₂ are used to obtain vapor phase Si₃H₈) prior to injection into the epitaxy reactor. It is also not always possible to take full advantage of its decomposition at very low temperatures; indeed, Si epitaxial layers grown with Si₃H₈ at temperatures below 500°C are of lesser crystalline and electronic quality [VIN 10]. Using Si₃H₈ fluxes (and thus growth rates) that are too high leads above 500°C to defective Si layers [GOU 09, SHI 12]. Disilane, on the contrary, is gaseous and yields higher growth rates than those obtained with silane (the Si-Si bond energy, 226 kJ mol⁻¹, is lower than the Si-H bond energy (318 kJ mol⁻¹); forming SiH₃ reaction by-products will therefore be easier with disilane than with silane). Disilane therefore seems like a good choice for the low temperature growth of Si and SiGe layers. In section 1.3.2, SiH₄, Si₂H₆ and SiH₂Cl₂ will be compared (in a 300 mm industrial RP-CVD reactor) for use in Si and SiGe growth. A more accurate description of their strengths and weaknesses will be given [HAR 12].

The growth rate of Si at 20 Torr is shown in Figure 1.17 (vs. the inverse of the absolute growth temperature). The flow of disilane is half that of silane or dichlorosilane, due to the fact that there are two Si atoms in a disilane molecule (i.e. F(Si₂H₆)/F(H₂) = 0.006 ⇔ F(SiH₄ or SiH₂Cl₂)/F(H₂) = 0.012). Similar growth rates, which have little dependence on the temperature, are obtained for the same mass flows of Si at higher temperatures (T > 950°C for SiH₂Cl₂ ⇔ T > 850°C for SiH₄ and Si₂H₆).

For SiH₄ and SiH₂Cl₂, as expected at low temperature, there is an exponential growth rate increase with temperature, the limiting factor being here the desorption of H and Cl atoms from the surface. Their activation energies are close to those predicted by the theory (Eₐ = 2.13 and 2.52 eV here ⇔ Eₐ = 2.17 and 2.30 eV in another 200 mm RP-CVD reactor, with 1 eV = 23.053 kcal·mol⁻¹). The situation with Si₂H₆ is more complicated, with a growth rate “plateau” between 575°C and 675°C (as with Si₃H₈ [GOU 09]). The growth rate increase with temperature is not the same below and above the plateau (Eₐ = 1.39 eV (T > 675°C) ⇔ Eₐ = 2.30 eV (T < 575°C)). For T < 575°C, growth rates are approximately 10 times higher with Si₂H₆ than with SiH₄, which are in turn approximately 10 times greater than the growth rates extrapolated for SiH₂Cl₂ (the practical Si
The growth rate limit of the order of 0.5 nm·min⁻¹ is reached at ~450°C with Si₂H₆ ⇌ ~550°C with SiH₄ ⇌ ~650°C with SiH₂Cl₂).

The three growth rate regimes observed for Si₂H₆ below 850°C may be explained this way: for T ≤ 575°C, the disilane molecules are adsorbed on the Si (001) surfaces according to Si₂H₆(g) → 2SiH₃(a). The adsorbed silyl, SiH₃(a), decomposes into silylene, SiH₂(a) plus an adsorbed H atom H(a) following SiH₃(a) → SiH₂(a) + H(a). SiH₂ is mobile on the underlying Si dimer rows and reacts to form Si monohydride following 2 SiH₂(a) → 2 SiH(a) + H₂(g). The residual hydrogen atoms are desorbed following 2 SiH(a) → 2 Si(a) + H₂(g); the two Si adatoms then contribute to growth [BRA 94]. For T ≥ 575°C (i.e. inside the “plateau”), the reaction Si₂H₆(g) + 2 SiH(a) → 2 SiH₄(g) + 2 Si(s) suggested in [KUL 90] becomes predominant over Si₂H₆(g) → 2 SiH₃(a). As the growth temperature increases, a gradual transition toward a behavior close to that of SiH₄ takes place. The growth rate of Si using Si₂H₆ is indeed close to that of SiH₄ for T ≥ 675°C (with the same Si mass flows).

The main features of the 20 Torr growth kinetics of SiGe using a gaseous mixture of dichlorosilane and germane (GeH₄) will be described in the
following pages. The SiGe growth rate and Ge concentration functions of the F(GeH₄)/F(SiH₂Cl₂) mass flow ratio (MFR) are provided in Figure 1.18, for growth temperatures between 550°C and 750°C. As shown previously, in another 200 mm epitaxial reactor [HAR 07], \( x \), the concentration of Ge increases parabolically with the GeH₄ flux. Such behavior is accurately described by a relationship of the type \( \frac{x^2}{1-x} = n \times \frac{F(GeH_4)}{F(SiH_2Cl_2)} \) [SUH 00], with \( n = 0.98 \) (700°C), 1.34 (650°C), 2.01 (600°C) and 2.52 (550°C).

The exponential increase of \( n \) as \( T \) (the absolute growth temperature) decreases may be modeled by \( n = 5.2 \times 10^{-3} \exp(0.45 \text{ eV}/k_B T) \). The Suh and Lee model, which gives a physical description of the system and includes the \( \frac{x^2}{1-x} = n \times \frac{F(GeH_4)}{F(SiH_2Cl_2)} \) formula, translates the fact that each surface Cl atom migrates toward a neighboring Ge atom, before being desorbed. The SiGe growth rate otherwise increases linearly with the mass flow of GeH₄. This is due to the catalyzed desorption of surface H and Cl atoms (freeing dangling bonds for the adsorption of Si and Ge atoms) due to the presence of Ge atoms (Ge-H and Ge-Cl bonds, respectively, 1.6 and 2.2 eV, are weaker than Si-H and Si-Cl bonds, 2.0 and 3.9 eV). However, the growth rate decreases and the Ge concentration increases when the growth temperature drops, making the 550°C epitaxy of SiGe layers rather complicated, at least with a SiH₂Cl₂ + GeH₄ mixture (prohibitively low growth rates for Ge contents below 35%).

The complexity of the situation increases when hydrochloric acid (HCl) is added to the gaseous mixture, which may be done, for example, to enable selective epitaxial growth versus Si₃N₄ [HAR 07]. At 650°C and 20 Torr, a growth rate drop of approximately 50% occurs with F(HCl)/F(H₂) = 0.0015 (Figure 1.19(a)). The SiGe growth rate increase with the germane flow, while being smaller, is also linear. Adding HCl leads to an increase in the concentration of Ge (Figure 1.19(b)).

In the gaseous phase, the added Cl atoms bond preferentially with Si, as the Si chlorides and chlorosilanes are more stable than the corresponding molecules formed with Ge (hence the Ge content increase). The increase of \( x \) with the F(GeH₄)/F(SiH₂Cl₂) MFR is also parabolic, with \( n = 2.06 \). If adding B₂H₆ does lead to a slight increase in the growth rate (in this case, no more than 25%, with a \( 2 \times 10^{20} \text{ cm}^{-3} \) boron atomic concentration), it is still difficult to selectively grow SiGe:B recessed or raised sources and drains.
below $650^\circ C$ (at least for Ge concentrations below 40% [HAR 08b, HAR 11b]).

![Figure 1.18](image)

**Figure 1.18.** a) SiGe growth rate and b) Ge concentration at 20 Torr versus the $F(\text{GeH}_4)/F(\text{SiH}_2\text{Cl}_2)$ mass flow ratio for different temperatures between $550^\circ C$ and $750^\circ C$. The $F(\text{SiH}_2\text{Cl}_2)/F(\text{H}_2)$ MFR, equal to 0.003 for $550^\circ C \leq T \leq 700^\circ C$, was multiplied by 4 at $750^\circ C$ (i.e. 0.012) to obtain low Ge percentage.

Higher SiGe growth rates (and lower Ge concentrations) than those obtained with dichlorosilane may be achieved at $550^\circ C$ with silane and especially disilane. The SiGe growth rate and the Ge concentration associated with SiGe layers grown at $550^\circ C$ and 20 Torr with GeH$_4$ and either SiH$_4$, Si$_2$H$_6$ or SiH$_2$Cl$_2$ as the Si precursor are given in Figure 1.20 as a function of the $F(\text{GeH}_4)/F(\text{Si precursor})$ MFR.
Figure 1.19. a) SiGe growth rate and b) Ge concentration at 650°C and 20 Torr (vs. the $F(\text{GeH}_4)/F(\text{SiH}_2\text{Cl}_2)$ Mass Flow Ratio) with or without HCl ($F(\text{HCl})/F(\text{H}_2) = 0.0015$). The $F(\text{SiH}_2\text{Cl}_2)/F(\text{H}_2)$ MFR is equal to 0.003.

Figure 1.20. a) SiGe growth rate and b) Ge concentration at 550°C and 20 Torr functions of the $F(\text{GeH}_4)/F(\text{Si precursor})$ MFR. The same Si flows were used ($(F(\text{SiH}_2\text{Cl}_2)/F(\text{H}_2) = 0.003 \iff F(\text{Si}_2\text{H}_6)/F(\text{H}_2) = 0.0015)$.
As expected with SiH₄, x increases linearly with the GeH₄ mass flux. Such a behavior is accurately described by \( x/(1 - x) = m \times F(\text{GeH}_4)/F(\text{SiH}_4) \) relationships [ROB 91], with \( m = 2.61 \) at 550°C, 20 Torr (here) \( \Leftrightarrow m = 2.7 \) at 650°C, 20 Torr in another 200 mm epitaxy reactor [HAR 02]. It would seem that, for SiH₄ + GeH₄ gaseous mixtures, the growth temperature has only a small impact on the Ge concentration (at least between 550°C and 650°C). However, for reasons we still do not understand, the increase of x with the \( F(\text{GeH}_4)/F(\text{Si precursor}) \) MFR is parabolic for a Si₂H₆ + GeH₄ chemistry \( (x^2/(1 - x) = 0.47 \times F(\text{GeH}_4)/2F(\text{Si}_2\text{H}_6) \) at 550°C, 20 Torr). For given mass flows, lower Ge concentrations are obtained with SiH₄ and especially Si₂H₆ than with SiH₂Cl₂. The SiGe growth rates increase linearly with the GeH₄ mass flow for SiH₂Cl₂ and Si₂H₆, and exponentially with SiH₄.

They are, for given Ge concentrations, much higher at 550°C and 20 Torr with SiH₄ and especially Si₂H₆ than with SiH₂Cl₂, as shown in Figure 1.21. Such results clearly illustrate the interest of disilane for the low temperature, non-selective epitaxial growth of high crystalline quality SiGe layers in RP-CVD reactors [HAR 12].

![Figure 1.21. SiGe growth rate (vs. the Ge concentration) for the three Si gaseous precursors investigated at 550°C and 20 Torr. The same Si flows are used ((F(SiH₂Cl₂)/F(H₂)) and F(SiH₄)/F(H₂) = 0.003 \( \Leftrightarrow F(\text{Si}_2\text{H}_6)/F(H_2) = 0.0015 \))](image)

**1.3.5. Integration and conclusion**

The following paragraph will describe a successful use of low temperature H₂ bake and selective epitaxial growth in pMOSFET transistors
with state-of-the-art electrical performances [LE 11]. To modulate the threshold voltage, \( V_T \), and to have a high mobility channel for the holes, a \{Si\(_{0.8}\)Ge\(_{0.2}\) 3 nm/Si 2 nm\} bilayer was first of all grown at 650°C and 20 Torr with SiH\(_2\)Cl\(_2\) + GeH\(_4\) (using the operating points from Figure 1.18) on a 300 nm ultrathin SOI substrate (the starting thickness of Si was around 3 nm). A “HF-Last” wet cleaning followed by an \textit{in situ} H\(_2\) bake at 650°C and 20 Torr for 2 min (in order to avoid any islanding of the very thin Si starting layer) was used to prepare the surface prior to growth. A second epitaxy was then carried out later on in the process flow (after gate stack deposition and patterning) in order to selectively grow highly \textit{in situ} B doped Si\(_{0.7}\)Ge\(_{0.3}\) raised sources and drains (\([B] = 2 \times 10^{20} \text{ cm}^{-3}\)). This drastically reduces the access resistance of the transistors and to a certain extent, introduces a uniaxial compression in the SiGe channel, increasing the hole mobility. As with the channel, a temperature of 650°C was used for both the \textit{in situ} H\(_2\) bake and the SiGe:B epitaxy (with SiH\(_2\)Cl\(_2\) + GeH\(_4\) + HCl + B\(_2\)H\(_6\) chemistry; see Figure 1.19). Cross-sectional transmission electron microscopy (TEM) images of a pMOSFET transistor (gate length: 30 nm) and of the channel beneath its gate are shown in Figure 1.22. The \{Si\(_{0.8}\)Ge\(_{0.2}\)/Si\} bilayer has a perfect crystallinity (no extended defects; abrupt interfaces). The deposition process used for the Si\(_{0.7}\)Ge\(_{0.3}\):B raised sources and drains is otherwise perfectly selective versus the Si\(_3\)N\(_4\) spacers. \{111\} facets are, however, present at the boundaries between the source and drain regions and the gate.

![TEM image of pMOSFET transistor](image.png)

**Figure 1.22.** Cross-sectional high-resolution TEM images of a pMOSFET transistor (of 30 nm gate length) with a SiGe/Si bilayer as channel for the holes and silicided, 18 nm thick Si\(_{0.7}\)Ge\(_{0.3}\):B raised sources and drains.
1.4. Contact areas between the gate and the “source” and “drain”

1.4.1. Generalities

1.4.1.1. A history of silicides in microelectronics

The first metallic silicides were initially introduced to microelectronics at the end of the 1970s to improve the contact of MOS devices on the polysilicon gate [IWA 02]. The polycide/polysilicon gate was directly patterned using photolithography and etching of a deposited polycide/polysilicon stack layer according to a quite simple integration scheme. MoSi$_2$ was the first polycide to be used and was replaced by WSi$_2$ from the 1980s onwards, due to its lower resistivity (about 70 $\mu\Omega\cdot$cm vs. approximately 100 $\mu\Omega\cdot$cm for MoSi$_2$). However, the device miniaturization was accompanied by an increase in series resistance in the source and drain region, so a metallic silicide layer with low resistivity at both the gate and the source and drain zones was introduced. Thus, the self-aligned silicidation process (SALICIDE process) was introduced during the 1990s for sub-1 µm technological nodes. This ingenious process allows the simultaneous manufacturing of silicide for the gate, as well as the source and drain, by reaction between the metal deposited and the silicon without a photolithographic stage [LAU 82]. The silicide is formed all over the source and drain surfaces, increasing the contact area on the silicon compared with an approach without silicide (see Figure 1.23): contact resistance is thus considerably improved. Furthermore, since the silicide has metallic conduction properties, the doped junction layer’s contribution ($R_{\text{HDD}}$) to the resistance is then reduced as, in this case, the metallic layer is brought as close as possible to the channel.

Titanium silicide (TiSi$_2$), was the first silicide to be integrated by the self-aligned process as it presents low resistivity in the C54 crystalline phase (10–15 $\mu\Omega\cdot$cm). The formation of TiSi$_2$-C54 is done via a first more resistive crystalline phase, called TiSi$_2$-C49. The polymorphic transformation of the C49 phase into C54 is controlled by the nucleation of the grain boundaries and so becomes difficult when the dimensions of the silicide zones are reduced [GAM 98]. For this reason, from 0.25 µm nodes onwards, TiSi$_2$ was replaced by cobalt silicide (CoSi$_2$) that has a similar resistivity but with fewer nucleation problems. However, limitations in the use of this type of silicide also appeared for devices smaller than 0.1 µm: the large amounts of silicon are required to form CoSi$_2$ (1 nm of Co consumes approximately
3.6 nm of Si to form ~3.5 nm of CoSi₂ [MAE 95]) as well as its high formation temperature (above 800°C) make it incompatible with the presence of ultra-shallow junctions.

Figure 1.23. Schematic diagrams of half a MOS transistor showing the different contributors to the series resistance in the source and drain region a) a transistor with direct metallic contact on silicon; b) a transistor with silicided zones. $R_{\text{acc}}$: resistance of the channel’s layer; $R_{\text{ext}}$: accumulation resistance; $R_{\text{HDD}}$: resistance of the highly doped silicon extensions’ layer; $R_{\text{c1}}$: resistance of the metal contact on silicon; $R_{\text{c2}}$: resistance of the silicide on silicon; as a first approach, the contact resistance of metal on silicide may be neglected

1.4.1.2. Nickel silicide

Nickel silicide (NiSi) had been known to science for a long time [TU 75] and has become the silicide of reference for the 65 nm node for CMOS on bulk silicon substrates [FRO 04], such as CMOS on SOI substrates [STR 07]. Compared with its predecessors TiSi₂ and CoSi₂, NiSi has many advantages: its formation does not consume much silicon (1 nm of Ni consumes ~1.8 nm of Si to form ~2.2 nm of NiSi [MAE 95]) and it is formed at relatively low temperatures (~300–400°C). Unlike TiSi₂ and CoSi₂, its growth is not controlled by nucleation but by the diffusion of the metal, which leads to smoother layers. Its resistance is similar to that of CoSi₂ (~15–20 µΩ-cm) and is compatible with the formation of contacts on active zones made up of SiGe that, as explained later on in this chapter, is crucial. Since its integration into 65 nm node, numerous adaptations of the manufacturing protocol have taken place in order to adapt to the ever changing specifications of the semiconductor industry. NiSi is, to this day, still the silicide of choice for the most advanced technologies [ITR 11].

This section will describe the chemical treatments used during the nickel silicide (NiSi) formation process, the SALICIDE process. The chemical mechanisms will be succinctly described, particularly the adaptations of the SALICIDE process that were developed in order to overcome the constraints
of NiSi integration. The impact of the recent technological orientation and the miniaturization of technology will also be discussed.

1.4.2. Introduction to the conventional NiSi process for sub-90 nm nodes

1.4.2.1. General description of the self-aligned silicidation process

Applied to nickel silicide, the SALICIDE process has five successive steps (see Figure 1.24): the first step is the surface preparation of the semiconducting zones before the metallic layer deposition. Then, the deposit of a thin metal layer (less than 20 nm thick) is followed by that of a capping layer of titanium nitride (TiN). A first low-temperature annealing leads to the formation of a nickel-rich silicide. The TiN is then wet-etched and the excess nickel (especially the non-transformed nickel present on the insulating zones) removed via a second wet-etching process, selective toward the silicide. Finally, a second thermal treatment is performed to transform all the Ni-rich silicide into the monosilicide NiSi.

This chapter is dedicated to steps that use chemical treatments: step 1 (surface preparation) and 4 (wet etching). Steps 2 (deposition), 3 (first annealing) and 5 (second annealing) will only be described when they influence or are influenced by the chemical treatment steps.

Figure 1.24. Schematic diagrams of successive self-aligned silicidation process steps (SALICIDE) applied to a MOS transistor
1.4.2.2. The Ni/Si system and the formation of silicide by solid state reaction

1.4.2.2.1. Growth mechanisms

The annealing steps 3 and 5 lead to solid state reactions that constitute the core of the silicidation process. This part will first give some general information about Ni/Si systems and silicide formation. For more details in this area, the reader should refer [DHE 86], to which the following paragraphs refer.

The phase diagram of the Ni/Si system at thermal equilibrium is relatively complex as it is made up of 11 phases of which only six are stable at standard temperature [NAS 87]. These phases are likely to grow simultaneously or competitively when the system undergoes a thermal treatment but only the less resistive NiSi phase is desired at the end of the SALICIDE process. The nickel silicide growth during thermal treatments occurs typically in three steps:

– silicide nucleation.

– lateral growth of the nuclei into a uniform layer.

– increase in the layer’s thickness by reactive diffusion (reaction at the interface and diffusion through the formed layer) following Deal and Grove’s model [DEA 65] developed for silicon oxides.

Thankfully, thin film growth only leads to a limited number of phases. Indeed, some phases that are present on the binary diagram are sometimes unable to form because a too high energy or a too low diffusion coefficient is required for their nucleation [DHE 86]. Furthermore, the thin film growth of the silicide is generally sequential rather than simultaneous [GOE 82, DHE 86, GAS 93] due to the competition between reaction kinetics at the interface and the diffusion of metal in the growing phase. From the reactive diffusion equations, it can be shown that for an M/Si system leading to two phases I and II, phase I must reach a minimum critical thickness before phase II starts growing. This minimum thickness is of the same order as the initial thickness of the deposited metal. Consequently, phase II will only appear when the metal has been entirely consumed by the reaction [GAS 93]. There are never more than two phases present simultaneously.
Taking into account the specifics of thin film growth, the usual sequence of nickel silicides growth follows reaction [1.11] [OLO 76]:

$$\text{Ni} \rightarrow_{250^\circ C} \text{Ni}_2\text{Si} \rightarrow_{350^\circ C} \text{NiSi} \rightarrow_{800^\circ C} \text{NiSi}_2 \quad [1.11]$$

The transformation curve of the Ni/Si system (which represents the resistance of the system’s layer vs. the annealing temperature) shows the apparition of the three various phases (see Figure 1.25). The NiSi phase with the lowest resistivity corresponds to intermediate temperatures. For the smallest initial nickel thickness, the progressive increase in resistance for temperatures above 500°C corresponds to the agglomeration of the monosilicide NiSi. This phenomenon is the main cause of NiSi degradation and must be avoided [KIT 03, MUR 03]. The agglomeration of the silicide is influenced by many parameters such as the metal’s thickness, the doping of the silicon, the impurities content in the silicon and/or in the layer of nickel, whether it is controlled or not.

**Figure 1.25.** a) Evolution of the sheet resistance of nickel silicide versus the silicidation temperature (60 s in argon) for three initial thicknesses of deposited nickel: 9 nm, 15 nm, 21 nm; b) Scanning electron microscopy (SEM) of a 15 nm thick Ni sample surface annealed at 700°C. RTP = rapid thermal process

While the growth of Ni$_2$Si and NiSi is controlled by the metal diffusion in the silicide, the growth of NiSi$_2$ is controlled by a nucleation step. In general, the nucleation takes place above 800°C (see Figure 1.25), however a localized or extended decrease in nucleation activation energy (due to the presence of grain boundaries, dislocations, amorphous zones, etc.) may favor the apparition of NiSi$_2$ at lower temperatures (<500°C), thus changing the
usual process sequence [DHE 88]. Under certain conditions, other phases with nucleation controlled growth have been transiently observed during the thin film formation of nickel silicides: \( \text{Ni}_3\text{Si}_{1.2} \) [KIT 07], \( \text{Ni}_3\text{Si}_2 \) [LAV 03] as well as the \( \theta \)-\( \text{Ni}_2\text{Si} \) phase (non-stoichiometric, actually \( \text{Ni}_2\text{Si}_{1+x} \)), sometimes mistaken for the \( \text{Ni}_3\text{Si}_2 \) phase [MAN 09].

1.4.2.2.2. Adapting to the constraints of integration

Management of the nickel diffusion by two-step annealing. Referring only to the transformation curves obtained on a Si(100) substrate and with no device (see Figure 1.25), the formation of the NiSi monosilicide could be limited to a surface preparation, a nickel deposition and a thermal treatment in the 400–500°C range. However, these annealing conditions applied on a masked substrate favor a strong inhomogeneity in the silicide’s thickness both on the gate and active zones on the same chip. This phenomenon is linked to the nickel diffusion that is the main diffusing species during the \( \text{Ni}_2\text{Si} \) and NiSi growth [IWA 02]. On a masked substrate, the nickel present on the insulating zones such as Shallow Trench Isolation (STI) or nitride spacers diffuse toward the silicon, leading to an excessive silicidation at the pattern edges (see Figure 1.26) [IWA 02, FRO 04]. As line width decreases, the contribution of the edges becomes more and more dominant until a thicker silicide forms as the edges meet. Consequently, industrial SALICIDE process with two annealing steps has been used from the 65 nm node onwards [FRO 04]. The sequence of \( \text{Ni}_2\text{Si} \) and NiSi phases makes such a choice even more relevant: the first annealing step is carried out at low temperature (below 350°C) and leads to the formation of a nickel-rich phase, generally \( \text{Ni}_2\text{Si} \). At low temperature, the lateral diffusion of nickel atoms is much slower as diffusion phenomena are thermally activated. The second annealing step, which aims to transform \( \text{Ni}_2\text{Si} \) into NiSi, is carried out in the 350–500°C range after the selective removal of the metal. Afterwards, there is no more metal on the insulators to diffuse toward the silicide. By allowing a better control of the nickel diffusion, the SALICIDE process with two annealing steps reduces the junction leakages in the MOS transistors.

NiSi stabilization by platinum incorporation. The main issue with nickel monosilicide is its low morphological (i.e. agglomeration) and thermodynamic (transforms into \( \text{NiSi}_2 \)) stability (Figure 1.25). Indeed, during the manufacturing process of chips, the silicide formed must be able to withstand the highly cumulated thermal budgets, especially during contacts fabrication and interconnection levels. Many studies have been
carried out on the effect of incorporating a third element to the NiSi structure (W, Ti, Ta, Co, Pd, Pt, etc.) on its thermal and morphological stability [DET 06, DED 07]. Platinum was chosen by most of the suppliers [STR 07]. Adding platinum into NiSi was initially studied to delay the nucleation of the disilicide, NiSi$_2$ [MAN 99]. Because NiSi and PtSi phases have the same crystallographic structure, they can form a solid solution Ni$_{1-x}$Pt$_x$Si where some nickel atoms are substituted by platinum atoms from the crystal lattice. On the contrary, Pt is insoluble in NiSi$_2$, and so delays the phase formation, which only appears on a large scale above 900°C. More recent studies show that the use of platinum improves the morphological stability of monosilicide and delays its agglomeration at higher temperatures [DET 04]. As such, platinum shows a double advantage and the majority of the industry now uses between 5 and 10 atomic percent platinum in the SALICIDE process.

![50 nm](image)

**Figure 1.26.** Scanning electron microscopy (SEM) cross-section image that illustrates the edge effects induced by the lateral diffusion of Ni: the silicide formed at the edge is thicker due to a greater nickel injection from the insulating zones. This figure is extracted from [FRO 04]

1.4.2.3. **Steps of the SALICIDE process: techniques and processes**

1.4.2.3.1. Surface preparation

The surface preparation before deposition is carried out aiming to remove the native silicon oxide and to entirely clean the silicon surface before the metallic layer deposition. The presence of oxygen at the Si/Ni interface may change the phase sequence and even stop the silicide growth if the amount is too large [CHA 04]. Moreover, oxygen at low content may also lead to a rougher silicide surface.

The situation is similar to surface preparation before epitaxy (see Chapter 3) except for the fact that thermal cleaning in the presence of H$_2$
fluxes is never carried out prior to metal deposition due to the presence of ultra-shallow junctions. Three types of surface preparations may be used:

– *Wet chemical cleaning*. Wet chemical cleaning is a complex sequence of successive chemical treatments aiming to prepare the silicon surface for silicidation while preserving the patterns (such as spacers, the gate and insulators). The last step is a chemical deoxidization using a HF solution diluted in deionized and deoxygenated water (the dilution is in the range of 0.2–1%), which will be described in Chapter 3 of this book. The overall dissolution reaction may be summed up in reaction [1.12]:

\[
\text{SiO}_2 + 4\text{HF} \rightarrow 2\text{H}^+ + \text{SiF}_6^{2-} + 2\text{H}_2\text{O} \quad [1.12]
\]

Chemical etching is preceded by rinsing and cleaning steps that will not be described here. They are carried out aiming to remove any possible traces of organic or particular contamination from the silicon surface. The last active step is the HF-based treatment, leading to the “HF-Last” name of this type of process. The wafer is then rinsed with deionized and deoxygenated water and dried by anhydrous vapor fluxes (such as isopropyl alcohol (IPA)). After this treatment, the silicon atoms at the surface are exposed and the dangling bonds are passivated by hydrogen atoms\(^1\). Such a surface is stable for dozens of minutes that leaves enough time to load the wafers into the deposition tool where they are protected from oxidation by an inert atmosphere (N\(_2\) or Ar generally). Any exposure to air before loading must be minimized.

– *Physical cleaning with argon plasma*. Often performed in addition to chemical cleaning, plasma cleaning is based on mechanical abrasion (by Ar\(^+\) ions bombardment) of the silicon’s surface [FRO 04, IMB 07, BON 07, YAN 10]. The aim is to remove the final oxygen and fluorine residues on the silicon surface that may still remain after the HF cleaning, as well as any possible traces of carbon. The plasma chamber is connected to the deposition equipment that allows the sequence of cleaning and metallic deposition steps. Direct plasma cleaning may, however, lead to some problems: due to the bombardment, defects may appear on the Si surface. In addition, injection of charges in the gate oxide may alter the electric behavior of the components. Plasma cleaning may also lead, under certain plasma generation conditions, to the creation of a layer presenting a high defects density near

---

\(^1\) For more details about the mechanisms of SiO\(_2\) etching with hydrofluoric acid, the reader may refer to Chapter 3 of this book, which is dedicated to the chemistry of surface preparations.
the silicon surface [IMB 07, BON 07]. This layer contains implanted argon atoms as well as contaminants such as oxygen, fluoride or carbon. It will induce kinetic modifications to the nickel diffusion and during the first silicidation steps will favor the apparition of Ni₃Si₂ rather than Ni₂Si. On a blanket substrate, this change in the sequence has little influence in fine on the texture and resistivity of the monosilicide [IMB 07]. However, it negatively impacts the yield of the electrical devices [BON 07].

– **Surface preparation by reactive plasma.** Surface preparation by reactive plasma (or remote plasma) was recently introduced to overcome the drawbacks of argon direct plasma cleaning. As with argon plasma, the cleaning chamber is connected to the deposition equipment and so allows *in situ* substrates surface cleaning without any exposure to air before the metal deposition. The cleaning is no longer physical but chemical as the active species are separately generated in the reactor and directed toward the substrate. The most common cleaning based on reactive plasma used for the silicidation is called Siconi™ that is based on the reaction between a fluoride component (NH₄F) and silicon oxide [YAN 10]. The “SICONI” process consists of three main steps that are shown in reactions [1.13], [1.14] and [1.15]:

– the generation of active species in the plasma chamber

\[
\text{NF}_3 + \text{NH}_3 \rightarrow \text{NH}_4\text{F} + \text{NH}_4\text{F.HF} \quad [1.13]
\]

– the chemical etching of the silicon oxide at approximately 30°C

\[
\text{NH}_4\text{F} + \text{SiO}_2 \rightarrow (\text{NH}_4)_2\text{SiF}_6 + \text{H}_2\text{O} \quad [1.14]
\]

– sublimation of the solid species (NH₄)₂SiF₆ at about 100°C

\[
(\text{NH}_4)_2\text{SiF}_6 \xrightarrow{100^\circ\mathrm{C}} \text{SiF}_4 + \text{NH}_3 \quad [1.15]
\]

Reactive plasma cleaning has improved the electrical performances and the yields of the devices to which it is applied [BON 07, YAN 10]. Several explanations have been proposed: for one, SICONI does not damage the silicon’s surface, unlike argon plasma. The lack of defect on the surface may limit the random diffusion of the nickel atoms, which reduces the junction leakage. Moreover, it has been proven that the “SICONI” process leaves a non-negligible quantity of fluorine atoms at the surface [IMB 07]. This contamination could act as a barrier to the nickel’s diffusion, further reducing junction leakage.
1.4.2.3.2. Deposition of the metallic layer and the silicidation annealing

Metallic layers are generally deposited under vacuum at room temperature by physical vapor deposition techniques such as magnetron sputtering. The metallic target is pulverized by an Ar+ ions beam and atoms of metal that are ripped from the surface are then deposited on the silicon substrate. The metallic thickness required for sub-65 nm technology is less than or equal to a dozen nanometers. The nickel (or Ni(Pt) alloy with a few percentage of platinum atoms) sputtering is usually followed by the deposition (in a different chamber of the same equipment) of a titanium nitride capping layer (TiN). The TiN is supposed to protect the nickel layer from oxidization during the exposure to air that precedes the charging in the annealing equipment.

The silicidation annealing is done by rapid thermal annealing (RTA) technique under inert conditions (i.e. in the presence of Ar or N₂). It lasts less than 120 s, which is more than enough to fully transform the metal layer into a silicide layer. The temperature is chosen depending on the phase needed: between 250°C and 350°C for a nickel-rich phase and between 350°C and 550°C for the monosilicide NiSi, according to the transformation curves of the Ni/Si system (see Figure 1.25).

1.4.2.3.3. Selective wet etching of the metallic layer

After the first silicide is formed, the TiN cap as well as the non-transformed metal that is present on the insulating zones (SiO₂ insulators and Si₃N₄ spacers) must be removed. The aim of this step is to chemically etch the TiN and metal with very high selectivity with regard to the silicide and the other constituting elements of the device (spacers, insulators). The process used must meet these specifications. The TiN layer is generally etched by a mix of aqueous ammonium hydroxide NH₄OH and hydrogen peroxide H₂O₂, commonly known in microelectronics as SC1 (standard clean 1) [KER 70]. It reacts poorly with nickel and mainly contributes to passivate the Ni surface via the formation of a hydroxide layer. Several processes may then be used to remove the Ni or Ni(Pt).

_Sulfuric acid/hydrogen peroxide mixture (SPM)._ The chemical solution that is most commonly used to etch nickel selectively versus the silicide is a

2 The chemistry described in this chapter is detailed in Chapter 3.
mixture of concentrated sulfuric acid (H₂SO₄ 96%) and hydrogen peroxide (H₂O₂) diluted at 30% with deionized water, commonly named SPM (sulfuric peroxide mixture). The mixing of the H₂SO₄ and H₂O₂/H₂O is highly exothermic and leads to the formation of Caro’s acid H₂SO₅ by oxidation of the sulfuric acid following reaction [1.16]:

\[
\text{H}_2\text{SO}_4 + \text{H}_2\text{O}_2 \leftrightarrow \text{H}_2\text{SO}_5 + \text{H}_2\text{O}
\]  

Caro’s acid is a highly oxidizing compound: the standard potential of the H₂SO₅/H₂SO₄ couple is 1.81 V (relative to the standard hydrogen electrode (SHE) against 1.776 V for H₂O₂/H₂O [POU 63]). For the selective etching step, the SPM is used in varying proportions of H₂SO₄ 96%/H₂O₂ 30%, ranging from 8:1 to 2:1 with the mixture sometimes being diluted in deionized water. Without dilution, the temperature can reach 130°C in the minutes following mixing. The decomposition of the active species (especially H₂O₂) is fast, meaning the chemical etching of nickel must be carried out straightaway after mixing in the reactor. The selectivity of the etching is based on the formation at the silicides surface of a passivating layer of SiO₂ under the oxidizing action of H₂O₂ and/or H₂SO₅ (depending on the dilution) [RAN 74, SHE 98]. On the insulating zones, the nickel is rapidly oxidized and then dissolved into the solution (see reaction [1.17]), while a competition between the oxidization/dissolution of the nickel and oxidization of silicon happens on the surface of the silicide. The oxidation of silicon, which is thermodynamically more favorable, (see reaction [1.18]) is instantaneous and the oxide formed protects the underlying silicide from etching. Figure 1.30 illustrates the NiSi passivation surface mechanism. The selectivity, which may be defined as the ratio of the nickel etching rate to that of silicide, is quasi-infinite [FRO 03].

\[
\text{Ni} + \text{H}_2\text{O}_2 + 2\text{H}^+ \leftrightarrow \text{Ni}^{2+} + 2\text{H}_2\text{O}
\]  

\[
2\text{Si} + \text{H}_2\text{O}_2 + 2\text{H}_2\text{O} \leftrightarrow 2\text{SiO}_2 + 6\text{H}^+
\]

Hydrochloric acid (HCl)-based mixtures: aqua regia and HPM. The incorporation of platinum stabilizes the nickel silicide but also creates additional constraints for the use of SPM for the selective etching step. Platinum, as a noble metal is thus difficult to oxidize and so is not etched when using this solution. Platinum atoms that are contained in the nickel
layer, sputtered from a Ni(Pt) alloy target, are then etched by the “lift-off” phenomenon during the selective SPM removal: as they are in a small enough amount, they are removed during the oxidation/dissolution reaction of the surrounding nickel atoms. However, platinum residues sometimes remain at the surface of the insulating zones after the SPM etching. These residues are thus potential sources of short circuits between the gate and the active zones of the transistor. Their chemical etching requires the presence of platinum complexing agents in the selective etching solution that may lower the apparent oxidation potential of the metal. Using chloride-based chemistries is a possibility since platinum at oxidation states +2 and +4 forms soluble complexes with chlorides, respectively, tetra- and hexachloroplatinates (PtCl$_4^{2-}$ and PtCl$_6^{2-}$).

Experimental selective etchings based on aqua regia (a well-known platinum etcher based on a mixture of nitric acid (HNO$_3$) and hydrochloric acid (HCl)) were carried out successfully [GOH 07]. However, the treatment of the effluent streams is somewhat delicate and aqua regia is seldom used in microelectronics to this day. The mixture of hydrochloric acid and hydrogen peroxide (also called SC2 or hydrochloric peroxide mixture (HPM)) may equally be a possible alternative. In this case, H$_2$O$_2$ oxidizes the platinum atoms that are then dissolved in the form of hexachloroplatinic acid (H$_2$PtCl$_6$). This chemistry is not, however, without its own problems: under certain conditions of silicon manufacturing, it leads to an important oxidation of the silicon at the surface of the silicide [IMB 08]. This phenomenon appears mainly at very low silicidation temperatures, when the silicide formed is nickel-rich. The silicon oxide’s thickness can measure up to several dozens of nanometers. It highlights a decomposition of the underlying silicide: the nickel and the silicon of the silicide are both oxidized, respectively, in Ni$^{2+}$, which is then dissolved, and SiO$_x$. The remaining platinum in the SiO$_x$ layer could catalyze the decomposition of the silicide into SiO$_x$ by galvanic corrosion.

As such, SPM remains despite everything the favored solution nowadays to perform the selective etching step for the latest technological nodes. Actual studies are leading toward the use of high temperature SPM mixtures (>150°C) that are useful for removing the final platinum residue in cases in which Ni(Pt) layers have a platinum content of 10% [IMB 09]. The mechanisms of platinum dissolution are still not well understood.
1.4.2.4. Abnormal migration of nickel under the gate: how chemistry rescued the latest dimensions

The integration of silicides in tridimensional devices such as MOS transistors induced a whole series of modifications in the formation conditions of silicide. As described previously, the phenomena of the masked substrates (such as the lateral diffusion of nickel from the insulating zones) have lead the industry to adapt the SALICIDE process. At the very smallest dimensions, a new adverse phenomenon of transistor yield appeared (see Figure 1.27): the abnormal nickel migration under the gate and/or the spacers of the transistor [STR 07, SEG 05]. This phenomenon is called “encroachment” and its frequency of apparition increases considerably with use of smaller dimensions and high structures density. It is very difficult to detect as it may only involve one transistor in 100,000 and its effect is mainly seen in yield measurements (percent of functional devices). The nickel encroachment is due to numerous parameters: the substrates orientation (i.e. (110) or (100)), the dopants type (boron, arsenic, phosphorus, etc.), the presence of species such as oxygen, fluorine or hydrogen in the silicon or the silicide, the thickness of the deposited metal layer, the platinum content in the metal layer and the annealing conditions of the silicidation process (temperature, time). Several “encroachment” morphologies have been observed: in some cases, the phase formed was analyzed and found to be NiSi$_2$ that has certain privileged orientations, notably [110] [KUD 08]. Other times, the spherical defects corresponding to the NiSi phase were observed near contact pads [IMB 10]. Without going into too much detail, it appears that the suppression of this diffusive phenomenon requires the preferential formation of Ni$_2$Si during the first anneal. It is crucial to avoid the formation of other nickel-rich species. Therefore, argon plasma surface preparation is to be avoided as it favors, as previously explained, the apparition of compounds such as Ni$_3$Si alongside Ni$_2$Si. The replacement of argon plasma by SICONI cleaning (after initial cleaning with a hydrofluoric acid) favors the formation of Ni$_2$Si exclusively [IMB 07] and contributes to reduce the Ni encroachment [BON 07]. Better yields are then obtained [BON 07, YAN 10]. The encroachment may be reduced by increasing the platinum content, using even thinner Ni(Pt) layers and by reducing the temperature of the first annealing [IMB 10].
1.4.3. Implications for the SALICIDE process of the recent technology evolutions

1.4.3.1. Introduction of germanium into MOS devices

From the mid 2000s and onwards, an increasing interest was taken in germanium-based electronic devices [THO 04, ZHA 05, LI 06, ZHU 05, DE 07]. The manufacturing of a MOSFET device with a Ge channel has become very worthwhile in terms of electrical performance due to the higher mobility of the electrons and holes in Ge compared with Si [LI 06, ZHU 05]. As well as pure germanium, an Si$_{1-x}$Ge$_x$ alloy also has very interesting channel, source and drain properties, especially when made by epitaxial
growth on Si(001) [ZHA 05, LI 06]. One of the most common applications for SiGe is the filling of the etched cavities by heteroepitaxy in the silicon’s active zones [THO 04]. This technique consists of generating lateral and uniaxial strain on the silicon channel in order to increase the mobility of the holes in a pMOS$^3$ device. The main obstacle to the integration of germanium is the chemical instability of its GeO$_2$ oxide. Indeed, it is soluble in water and requires the implementation of adapted surface cleanings that are specific to germanium and germanium-rich SiGe alloys (Ge $\geq$ 30%) [BRU 08a]. Aqueous and oxidizing chemistries (especially those with a high H$_2$O$_2$ content) are forbidden as they lead to a high degradation of the Ge layer via the dissolution of the GeO$_2$ formed at the surface following reaction [1.19]:

$$\text{GeO}_2 + 2\text{H}_2\text{O} \rightarrow \text{H}_2\text{GeO}_3$$

[1.19]

As with silicon, Ge and SiGe devices require the formation of low resistance contacts in the source and drain regions. Many metals have been evaluated by Gaudet for the formation of the contacts on germanium (called germanide or germanosilicide) [GAU 06]. Nickel is, once again, considered to be the most useful due to the resistance and thermal stability of the germanide obtained, subject to a few adaptations to be done in the self-aligned process. The cleaning of germanium before the metallic deposition is relatively simple due to the fact that the native germanium oxide is soluble in water. Diluted HF solutions [CAR 06] or, more simply, water-based cleanings [BRU 08b] are enough to remove the native oxide. Unlike the silicon process, the residual presence of oxygen does not decrease the formation rate of nickel germanide [NEM 08]. The transformation curve of the Ni/Ge system is shifted toward lower temperatures with respect to that of Ni/Si (see Figure 1.28). Below 250°C, Ni$_3$Ge$_3$ and NiGe phases coexist and grow simultaneously [BRU 08b, NEM 06]. Above 250°C, NiGe is the only phase present [BRU 08b]. The agglomeration of the germanide layer takes place at temperatures 100°C cooler than for the silicide and leads to a rapid increase in the resistance of the layer. The ideal temperature range to form a robust nickel germanide layer is approximately 50–100°C less than that of NiSi. In the case of germanosilicide, the situation is intermediate [ZHA 04, CAR 06]. The higher the germanium content in SiGe, the more the transformation curve looks like that of pure Ge. However, for the same amount of germanium, the strain and the relaxation level of the SiGe layer

---

3 SiGe heteroepitaxy is covered in section 1.3 of this chapter.
greatly affect the morphological stability (or the agglomeration) of the germanosilicide: relaxation increases the morphological stability [ZHA 04].

As previously explained, the selectivity of the nickel chemical etching performed by SPM mixtures toward the silicide is due to the formation of SiO₂ at the silicide’s surface that protects it from etching. In the presence of germanium, the highly oxidizing power of hydrogen peroxide (H₂O₂) leads to the formation of GeO₂; as this compound is very unstable unlike SiO₂, the NiGe is rapidly etched (see Figure 1.30). A germanium content in Si₁₋ₓGeₓ above which SPM is incompatible may be defined. Figure 1.29 illustrates this phenomenon is the case of diluted SPM [CAR 06]. For a Ge content of 20% in Si₁₋ₓGeₓ (x = 0.2), SPM treatment applied to germanosilicide only leads to a small increase in the layers resistance, mainly at low annealing temperatures, corresponding to the formation of the Ni₂SiGe phase. So the germanosilicide is mainly left intact. For a Ge content of 30%, the increase in resistance after SPM is high, but still acceptable. At 50% Ge, the high resistance leads to a complete etching of not only the germanosilicide (when formed at low temperature (<400°C)), but also of the underlying SiGe. In the temperature range corresponding to mono-NiSiGe formation (T>350°C), the increase in resistivity is less but still leads to a strong deterioration of the phase. Beyond 500°C, the redistribution of the germanium between

**Figure 1.28.** Evolution of the nickel silicide NiSi and of nickel germanide NiGe sheet resistance versus the annealing temperature (60 s in Ar). Ni initial thickness is 9 nm
the substrate and the germanosilicide leads to a reduced content of Ge in the NiSiGe [JAR 02]. It is thus less sensitive to SPM treatment, meaning the variation in the resistivity of the layer is smaller. However, high annealing temperatures are not compatible with the fabrication of a MOS device. As a result, germanosilicide must be formed at less than 450°C and this germanosiliciure must resist the selective etching step.

Several approaches for selective etching in the presence of germanium may be found in the literature. A simple strategy is to use non-selective chemical reactions and to base the process on precise control of the etching time. “Selectivity” in these cases is based on the difference in the layer thickness of the obtained NiGe and the deposited Ni layers: NiGe is 2.4 times thicker than Ni so it is possible to completely etch the nickel layer while leaving enough NiGe. Aqueous and acid mixtures (HNO₃/H₂O) [LI 06] or alkaline (NH₄OH/H₂O₂/H₂O) [ZHU 05] may be used. However, the process control is delicate as for the same process time a slight variation of temperature or mixture composition may strongly modify the final thickness of the remaining NiGe.

Another approach is to use acid aqueous chemical reactions that are poorly oxidizing, such as a mixture of HCl/H₂O with [BRU 08b] or without few percents of HF [NEM 08]. In this case, nickel etching is favored by the presence of Cl⁻ ions that are complexing agents of Ni²⁺. The oxidation
potential of Ni is thus lowered and Ni is oxidized by the $\text{H}_3\text{O}^+$ ions, whereas in the case of Ge (or rather, of its oxide) the dissolution is very slow.

Finally, a third approach is to use anhydrous and non-oxidizing mixtures. Concentrated sulfuric acid at 96% is a good example [CAR 06] as it slowly etches the nickel while it leads to the formation of the insoluble GeO$_2$ oxide on the germanide’s surface. The germanide is then protected from etching during the treatment. In this case, the main NiGe consumption comes from the rinsing step performed in deionized water after the sulfuric acid treatment. During this step, the surface GeO$_2$ is dissolved in water following reaction [1.19]. It is, however, mandatory to carry out the rinsing step in order to remove the sulfide residues left by the decomposition of the sulfuric acid. It leads to a consumption of less than 10% of the initial NiGe layer. Figure 1.30 summarizes schematically the different surface evolutions of NiSi and NiGe under SPM or 96% $\text{H}_2\text{SO}_4$ treatment followed by rinsing with deionized water.

![Figure 1.30. Schematic representation of the surface evolution of the NiSi layer a) and the NiGe layer b) and c) during the nickel selective etching process. Etching performed using diluted SPM a) and b) and with 96% $\text{H}_2\text{SO}_4$ c) followed by a rinsing step with deionized water. This figure is extracted from [CAR 06]](image-url)
1.4.3.2. Planar and 3D co-integration pMOS/nMOS

As previously explained, with a few adaptations to the self-aligned NiSi process, an equivalent NiGe process would be possible and the literature shows numerous examples of such a MOS device on Ge with NiGe metallization [WEB 05, BRU 08]. The use of strained SiGe layers has become widespread, including at industrial level [AND 05, LE 11]. However, germanium is not interesting for nMOS transistors, meaning the requirement to co-integrate on the same substrate a Ge-based pMOS and a Si-based nMOS has emerged. Examples can be found of planar and three-dimensional (3D) co-integration [LE 08] (see Figure 1.31). The aim of 3D integration is to increase the density of the devices on a chip. Contact metallurgy for pMOS and nMOS transistors are carried out separately. As nMOS is buried, it must be able to resist the thermal budget of the pMOS manufacturing. With this approach, very good results in terms of electrical performance have been obtained for superior pMOS on a GeOI (germanium on insulator) substrate as well as SOI substrate [BAT 09]. In the case of the planar approach, a simultaneous silicidation of the pMOS and nMOS may be performed. The self-aligned metallization process must meet the specifications of silicon as well as germanium or SiGe alloy. However, it is possible to find an acceptable compromise on the formation temperature of the silicide and on the selective etching mixture. Here, as well, examples may be found of successful planar nMOS on SOI and pMOS on GeOI co-integration with nickel metallization that show the incredible flexibility of the self-aligned process [LER 10].

Figure 1.31. Schematic representation of two pMOS Ge/nMOS Si co-integration strategies; co-integration is a) planar or b) 3D [LER 08]
1.4.4. Conclusion

To meet the ever increasing specifications such as miniaturization and performance of devices, the Ni SALICIDE process, which was introduced initially for the 65 nm node, has since been routinely adapted. For every technological node and each new integration approach, the steps of the process had to be redesigned. The previous sections have shown a few examples of the evolution of the various chemical treatments steps (for the surface preparation and the selective etching of the metal) used in the process. These steps are very sensitive to the integration of new materials (platinum to improve the morphological stability of NiSi, germanium to increase the carrier mobility in the channel). For future generations, one of the main challenges is the ever decreasing size of the contact surface between the metal and the junction [ITR 11], the effect of which is increased in the case of devices on SOI where the contact may be lateral. In these conditions, the reduction of the silicide/doped silicon contact resistance is a dominant parameter and several minimization approaches are available: engineering of the silicide/silicon interface via the incorporation of new species, double silicidation (one for the pMOS and the other for the nMOS with a low Schottky barrier height for, respectively, the holes and the electrons). Here, as well, the processes will need to adapt to the integration of new materials. Finally, another challenge for future generations is the control of the junction consumption by the silicidation process. As the junction is no bigger than a dozen nanometers, the thickness of the metal must be less than 5 nm and its uniformity controlled, a difficult feat to achieve with sputtering deposition techniques. These techniques may have to give way to chemical deposition techniques that allow the selective deposition of the metal, such as MOCVD or electrochemical deposition. To integrate such techniques into general production would certainly mean the end of the SALICIDE process as we know it. This prospect is still distant, however, and the SALICIDE process still has a few years ahead of it before being dropped from use.

1.5. General conclusion

This chapter has described the first stages in the manufacturing of a transistor. The “front end of the line” steps are essential as they define the final transistor performance and could be described as the precise dimensioning of the heart of the transistor in order to give it what will be its
main characteristic: either high performance (with a strong passing current and an acceptable leakage current) or low consumption (with an acceptable passing current and a very low leakage current). The gate stack is a key element of the FEOL, as are the growth of the materials by epitaxy and the manufacturing of efficient contacts, two other crucial sequential modules that have been described in this chapter.

From there, the latest transistors designed with 20 nm or 14 nm gate length show already some significant improvements over the previous generation. New “3D” shape transistors (called FinFETs) have been recently developed and presented [YAM 11, AUT 12, JAN 12] when standard planar CMOS technology also improved in parallel through a fully depleted SOI technology by keeping the historical planar integration scheme and extending the high performance and the low power consumption targets through innovative pathways [LIU 11, CHE 12, GRE 12, LE 12]. Finally, going down the road of the technology roadmap, new transistors are also on their way [KUH 12], “nanowires” transistors being currently a more widely investigated option in research [ERN 08, SAI 10, BAR 12, RIE 12].

1.6. List of Abbreviations

MOS: Metal Oxide Semiconductor
MOSFET: Metal Oxide Semiconductor Field Effect Transistor
FinFET: Fin (shape) Field Effect Transistor
CMOS: Complementary Metal Oxide Semiconductor (N-type+P-type)
NMOS: N-type Metal Oxide Semiconductor
PMOS: P-type Metal Oxide Semiconductor
EOT: Equivalent Oxide Thickness
CVD: Chemical Vapour Deposition
PVD: Physical Vapour Deposition
MOCVD: Metal Organic Chemical Vapour Deposition
ALD: Atomic Layer Deposition
PEALD: Plasma Enhanced Atomic Layer Deposition
ALCVD: Atomic Layer Chemical Vapor Deposition
RPCVD: Reduced Pressure Chemical Vapor Deposition
BDMAS: Bis_DiMethylAmino Silane
TDMAS: Tris(DiMethylAmino) Silane
TDMAT: Tetrakis(DiMethylAmino) Titanium
TEMAT: Tetrakis(EthylMethylAmino) Titanium
TDEAT: Tetrakis(DiEthylAmino) Titanium
GPC: Growth Per Cycle
MFR: Mass Flow Ratio
RSD: Raised Source Drain
SSMBE: Solid Source Molecular Beam Epitaxy
XRD: X-Ray Diffraction
DDC: Diluted Dynamic Clean
STM: Scanning Tunneling Microscopy
IPA: IsoPropyl Alcohol
HF: HydroFluoric acid
SIMS: Secondary Ion Mass Spectrometry
VS: Virtual Substrate
CDE: Cyclic Deposition Etch
TEM: Transmission Electron Microscopy
STI: Shallow Trench Isolation
RTA: Rapid Thermal Annealing

1.7. Bibliography


[LE 11] LE ROYER C., VILLALON A., CASSE M., et al., “First demonstration of ultrathin body c-SiGe channel FDSOI pMOSFETs combined with SiGe(:B) RSD: drastic improvement of electrostatics (Vth, p tuning, DIBL) and transport (μ0, Isat) properties down to 23 nm gate length”, Proceedings of 2011 IEDM Conference, pp. 16.5.1–16.5.4, 2011.


