4 Overview of FinFET Device Technology

4.1 INTRODUCTION

In Chapter 3 we discussed that in a multiple-gate or multigate metal-oxide-semiconductor (MOS) capacitor system, an inversion condition can be reached by a certain applied bias to the gates forming minority carrier concentration (e.g., electron) in the majority carrier (e.g., *p*-type) thin silicon body. Under this inversion condition, the thermally generated minority carriers form the inversion charge within the volume of the ultrathin silicon body. However, it is difficult to sustain this minority carrier charge in an undoped or a lightly doped majority carrier body from the thermal generation of carriers without a steady source of carrier supply. Therefore, a heavily doped minority carrier region (e.g., *n*+ region in a *p*-type body), called the *source*, is added to one end of the silicon body of the multigate MOS structure as a terminal for a steady supply of minority carriers at the inversion condition. And, another heavily doped region with the same doping-type as the source region, called the *drain*, is added as a terminal at the other end of the multigate MOS capacitor body to form a multiple-gate MOS field-effect transistor (FET) or MOSFET. These source and drain terminals contact the two opposite ends of the inverted silicon body so that a potential difference can be applied across the body and cause a current flow in the multiple-gate MOSFET structure. Such a multiple-gate MOSFET device with ultrathin vertical silicon body, called the "fin," on silicon pedestal, sidewall gate stack, and a source and a drain at the two ends of the gate length is called the "fin field-effect transistor" or "FinFET." Thus, a FinFET includes an ultrathin vertical silicon fin on a silicon substrate with a thin insulating layer such as SiO_2 grown on the sidewalls, a conducting metal layer, called the *gate electrode* deposited on the top of the gate oxide, and heavily doped source and drain regions formed from one end of the fin to the nearest gate edge and from the far edge of the gate to the far end of the fin, respectively. In reality, the gate can be placed on two, three, or four sides of the channel or wrapped around the channel as discussed in Chapter 1 [1,2].

Typically, a FinFET can be designed on a bulk-silicon substrate or silicon-on-insulator (SOI) substrate as described in Section 1.4 [1–3]. A bulk-FinFET is a four-terminal device with *gate*, *source*, *drain*, and *substrate or body*, whereas an SOI-FinFET is a three-terminal device with *gate*, *source*, and *drain* with floating *body*. Thus, bulk-FinFETs are more familiar to integrated circuit (IC) design engineers and the fabrication steps of the devices are compatible with those of the conventional planar complementary metal-oxide-semiconductor (CMOS) devices fabricated on bulk-silicon wafers [4]. The body terminal in bulk-FinFETs offer more flexibility of device operation in very large scale integrated (VLSI) circuits and systems. Bulk-FinFETs eliminate the problems associated with SOI-FinFETs such as

expensive wafer cost, high defect density, floating body effect, and poor heat dissipation [5]. The heat generated in the channel can be transferred to the substrate through the fin body that is connected to the substrate of the bulk-FinFETs. In addition, both the bulk and SOI FinFETs offer the same scalability while bulk-FinFETs have better heat dissipation characteristics [5].

A FinFET device is symmetrical and cannot be distinguished without the applied bias. In bulk-FinFET device architecture, shallow trench isolations (STI) are used to isolate various devices fabricated on the same substrate. In advanced VLSI circuits, *n*-channel FinFETs (*p*-type fin with n^+ source-drain) and *p*-channel FinFETs (*n*-type fin with p^+ source-drain) are fabricated together and is referred to as non-planar CMOS technology. In this chapter, an overview of the fabrication processes of FinFET devices in non-planar CMOS technology is presented to appreciate the operation of FinFET devices discussed in Chapters 5–10 of this book.

4.2 FinFET MANUFACTURING TECHNOLOGY

Figure 4.1 shows three-dimensional (3D) cross-sections of ideal FinFET structures with gate oxide on three sides (sidewalls and top) of the ultrathin-body fin on bulk-silicon substrate (Figure 4.1(a)) and SOI substrate (Figure 4.1(b)). As shown in Figure 4.1, the basic technology parameters for both the bulk and SOI FinFET devices include gate oxide thickness (T_{ox}), gate length (L_g), fin body thickness (t_{fin}), and fin height (H_{fin}). Additional technology parameters for bulk-FinFETs include field oxide or STI of thickness (T_{fox}) for isolation between neighboring devices. For FinFET manufacturing, the existing CMOS manufacturing technology is adopted [6–11]. However, some specific process steps or modules require additional restrictions and optimization, e.g., controlling the fin critical dimension (CD) and H_{fin}. In addition, the manufacturing flows for the bulk-FinFETs (Figure 4.1(a)) and SOI-FinFETs (Figure 4.1(b)) have subtle differences as shown in Figure 4.2. Furthermore, each

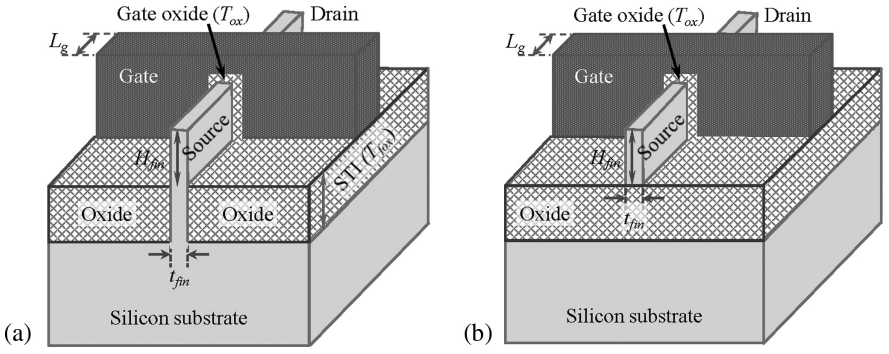

FIGURE 4.1 Ideal three-dimensional FinFET device structures with gate oxide on three sides (sidewalls and top) of the ultrathin fin body: (a) bulk-FinFET on silicon substrate; and (b) SOI-FinFET on silicon-on-insulator substrate. Here t_{fin}, H_{fin}, and L_g, are the fin thickness, fin height, and gate length of the devices, respectively; T_{ox} is the gate oxide thickness and T_{fox} is the field-oxide thickness of STI for device-to-device isolation.

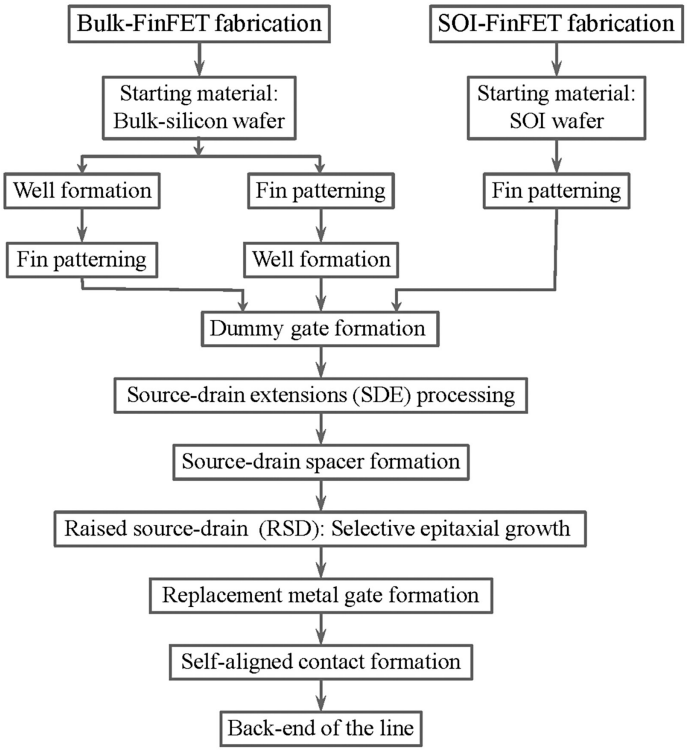

FIGURE 4.2 A typical flowchart for the fabrication of the bulk and SOI FinFET devices showing the major process modules of a representative VLSI technology.

IC manufacturer has its own proprietary fabrication processes different from that discussed in this chapter [6–11].

Figure 4.2 shows a typical flowchart of the major process modules for the manufacturing of bulk and SOI FinFETs at the nanometer node VLSI technology. Apart from the fin patterning and well formation, the major process flows for both the bulk and SOI FinFETs are identical. In the following section, we will present a brief overview of a typical FinFET fabrication technology. First of all, the bulk-FinFET fabrication process is overviewed. Then the major differences between the bulk and SOI FinFETs fabrication processes are highlighted.

4.3 BULK-FinFET FABRICATION

The fabrication process discussed in the following section is only to illustrate a representative FinFET manufacturing technology [7–12] and highlight the basic features of FinFET devices. In reality, a complementary FinFET or non-planar CMOS manufacturing technology is more complex than described in this section. Since the channel width of a FinFET is determined by the fin height H_{fin} (and also by fin thickness, t_{fin}) of the ultrathin-body fin, multiple-fin device architecture is used in FinFET layout to achieve the target drain current [13].

4.3.1 Starting Material

As shown in Figure 4.2, the starting material for the fabrication of FinFET ICs is ⟨100⟩ heavily doped p^+ silicon wafers. The front-side of the wafers is a lightly doped or undoped layer of epitaxial silicon with thickness determined by the target H_{fin} of FinFET device technology. The starting wafers are cleaned and a thin layer of pad or screen oxide is grown on the surface for the well implantation process as shown in the flowchart in Figure 4.2.

4.3.2 Well Formation

As outlined in Figure 4.2, there are two alternative methods to form wells in FinFET manufacturing for VLSI circuits: (1) at the beginning of the fabrication process; or (2) after the fin patterning.

4.3.2.1 *p*-Well Formation

In the well first FinFET fabrication process, a zero-level mask and subsequent etching process are used to define an alignment notch in the wafer with pad oxide. The wafers are then primed in hexamethyldisilazane (HMDS) to improve photoresist wetting and adhesion to oxidized silicon surface followed by photoresist coating and *top anti-reflective coating* (TARC). After the development of photoresist, *p*-well mask is used to expose the *p*-well region for ion implantation, hereafter referred to as the "implant," and cover the *n*-well region by photoresist pattern. Then the exposed *p*-well region of the wafers is implanted with p^+ (boron) anti-punchthrough (APT) dopants at the channel bottom region to suppress the leakage current [14] followed by *p*-type (boron) dopant implant to define the *p*-well region for *n*-channel FinFET (*n*FinFET) fabrication.

4.3.2.2 *n*-Well Formation

After the *p*-well formation, the photoresist is stripped and the wafers are cleaned. The wafers are then primed in HMDS followed by photoresist coating and TARC. Then the lithography and masking steps are performed so that the photoresist pattern covers the *p*-well region and exposes the *n*-well region. Next, the exposed region of the wafers is implanted with n^+ (phosphorus) APT at the channel bottom region to suppress the leakage current [14] followed by *n*-type dopant (phosphorus) implant to define the *n*-well region for *p*-channel FinFET (*p*FinFET) fabrication.

After the *n*-well ion implant processing step, the photoresist is stripped off and the wafers are cleaned followed by the removal of the initial pad oxide grown on the silicon front surface. After post-implant cleaning, the wafers are annealed using the rapid thermal annealing (RTA) process to activate the implanted dopants. Finally, the wafers are cleaned and a thin pad oxide is grown on the surface for silicon fin patterning in the next processing step.

4.3.3 Fin Patterning: Spacer Etch Technique

After the well formation, an array of ultrathin silicon fins is formed using the *spacer etch technique*, referred to as the *self-aligned double patterning* (SADP) [7,13,15–17].

It is to be noted that the dimensions of such ultrathin (≈7 nm) vertical structures are beyond the resolution of available optical (193 nm ArF Immersion) lithography. Although a high resolution extreme ultraviolet (EUV) lithography stepper shows great potential to print such fine lines, it is challenging to pattern such small fins in high volume production [17,18]. The major processing steps in patterning multiple-fins using SADP technique include: (1) carbon hard mask or mandrel patterning; (2) offset spacer formation; and (3) silicon fin formation as described below.

4.3.3.1 Mandrel Patterning

In this process step, a stack of layers is deposited over the previously grown pad oxide on the wafer: first of all, a thick layer of silicon nitride (Si_3N_4) using the *chemical vapor deposition* (CVD) process; then a thicker amorphous carbon sacrificial layer using the CVD process to support the fabrication of oxide spacers for fin patterning and is referred to as the "mandrel," which is removed after the spacer formation; a thin layer of dielectric anti-reflective coating that forms the *bottom anti-reflective coating* (BARC); and finally, photoresist coating and TARC.

After the deposition of the above layers on the wafers, the photoresist and masking steps are performed to expose multiple photoresist patterns on the BARC of the wafers. These photoresist strips are used as the mask to pattern the amorphous carbon mandrel layer using a highly anisotropic etch as shown in Figure 4.3(a). The Si_3N_4 layer deposited on the wafer is used as the etch stop layer (ESL) to ensure that the etch process stops on the front surface of the Si_3N_4 layer. Then the photoresist patterns and BARC are stripped in ionized oxygen plasma and the wafers with carbon mandrel patterns are cleaned to prepare the top surface for oxide spacer formation.

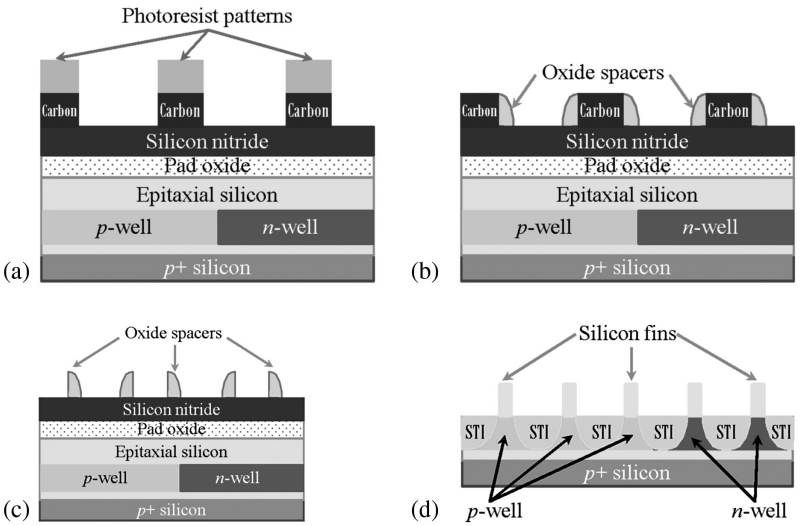

FIGURE 4.3 Two-dimensional cross-section of wafer for fin patterning: (a) mandrel patterning; (b) after selective oxide etch defining thin oxide spacers separated by mandrel strips; (c) after mandrel removal exposing oxide spacers; and (d) after TEOS etch to expose silicon fins along with the formation of STI between fins.

4.3.3.2 Oxide Spacer Formation

After mandrel patterning, a thin blanket layer of SiO_2 is deposited on the wafer using the CVD process. The deposited oxide layer surrounds the amorphous carbon mandrel at both ends of the carbon layer. In order to achieve straight lines at both ends of the oxide spacers, a photolithography process using a *cut mask* is used to trim the oxide spacers at both ends of the mandrel. Then a highly anisotropic etch process that is selective to oxide is used to form thin oxide spacers on the sidewalls of the amorphous carbon layer as shown in Figure 4.3(b).

After the oxide spacers formation, the mandrel strips between the oxide spacers are removed using a selective etch process as shown in Figure 4.3(c). The oxide spacers and the SiCN (silicon carbo-nitride) hard mask are unaffected by this etch process. After mandrel removal, the wafers are cleaned in a Megasonic bath [19]. Then using the oxide spacers as the hard mask, the underlying Si_3N_4 layer is removed by a highly anisotropic etching process to form the oxide spacer on the substrate.

4.3.3.3 Silicon Fin Formation

After the mandrel removal, the oxide spacers and nitride mask are used as patterns in an anisotropic etch that cuts down through the Si_3N_4, pad oxide, and epitaxial silicon into the well regions forming epitaxial silicon fins. The etching pressure, rate, and energy are extremely critical to create a sloped fin to achieve a smooth curved interface at the bottom of the fin. Furthermore, during the fin patterning process, the oxide spacers are etched and the shape of the Si_3N_4 hard mask strips underneath the oxide spacers are affected at the top. This etch process must terminate inside of the wells without an ESL to facilitate the process and is very challenging [17]. After cleaning the wafers, an ultrathin layer of thermal silicon dioxide (SiO_2) known as the "trench liner" is grown in the trenches to relieve the stress in the silicon around the upper and lower corners of the trenches [20]. Note that the Si_3N_4 is a diffusion barrier to oxygen preventing SiO_2 growth underneath. The etch process is followed by formation of exposed silicon fins and STI.

Silicon fin patterns and STI formation: The oxide trench liner is removed in an isotropic etch. After the linear oxide etch, the surface of the wafer is cleaned and a thick layer of tetraethyl orthosilicate (TEOS) oxide is deposited using the CVD process. Then the TEOS is densified by annealing the wafers to make it more resistant to wet etch. After TEOS densification, the wafers are polished using *chemical-mechanical planarization* (CMP) process using Si_3N_4 layer as the CMP stopper.

After the planarization of the surface, the residual Si_3N_4 layer is removed exposing the pad oxide grown on the wafer after the well formation (Section 4.3.2). The TEOS oxide is then etched back using a highly selective oxide etch to remove all of the oxide surrounding the fins and completely expose the fins as shown in Figure 4.3(d). This is a very challenging etch process to define the fin/well boundary [17,18]. The remaining oxide between the fins forms the STI (Figure 4.3(d)).

Fin removal: The SADP lithography process creates extra fins in bulk-FinFET wells that are removed to achieve the target drive current of FinFET devices at any technology node. In order to remove any odd fin, an ultrathin layer of thermal oxide is grown over the exposed silicon for ESL in the etching process. Next, a thin layer of SiCN hard mask is deposited over the wafer.

In the first step of the odd fin removal process, the wafers with deposited thermal oxide and SiCN are patterned with photoresist and developed. After photolithography and masking step, the photoresist pattern exposes target fin for removal. In the lithography process, a thin BARC layer is formed at the SiCN/photoresist interface. Then the exposed portion of the BARC layer is removed using an anisotropic etching process. In the next processing step, the exposed silicon fin is etched away using wet etch (typically, tetra-methyl-ammonium hydroxide or TMAH). Finally, the photoresist and BARC layers are etched away in an isotropic etch and a layer of oxide is grown on the exposed silicon base of the removed fin. After the fin removal process, the SiCN and oxide layers over the fins are etched away.

4.3.4 Alternative Well Formation Process

The alternative well formation approach after the fin patterning in a front-end FinFET fabrication technology enables the formation of uniform fin height H_{fin} after the STI etch-back processing step and mitigates the well/STI interface matching issue. The measured H_{fin} is then used for an accurate estimation of well implant energy to make the top of the wells coplanar with that of the STI. *This is the most common method for the formation of wells in a bulk-FinFET process flow* [15–17]. In this alternative well formation process flow, the SADP technique described in Section 4.3.3 is used for silicon fin patterning directly after the screen oxide growth on starting wafers (Section 4.3.1). The typical steps for the alternative well implant process including APT implants are similar to those described in Section 4.3.2 to form *p*-well and *n*-well regions for *n*FinFETs and *p*FinFETs, respectively. After the well formation, the wafers are cleaned and subjected to the RTA process to activate the implanted dopants in the wells. Next, the wafers are prepared for gate definition.

4.3.5 Gate Definition: Polysilicon Dummy Gate Formation

After odd fin removal and cleaning processes (Section 4.3.3.3) or alternative well formation process (Section 4.3.4), the fins are coated with an ultrathin oxide ESL and cleaned to define polysilicon FinFET gates. The major process steps include deposition, polish, and lithography processes to pattern amorphous silicon gate electrodes as described below.

In the first step of gate definition, a thick blanket layer of undoped amorphous silicon is deposited on the substrate with patterned fins using the CVD process. Then the amorphous silicon layer is polished back using CMP to form a smooth planar surface. After the planarization of the amorphous silicon surface, a thick layer of amorphous carbon hard mask is deposited using CVD. Then a thin layer of BARC is spun across the wafer. The wafers are then patterned with photoresist using gate mask to define the gate electrodes. Following the photoresist gate patterning, the carbon hard mask is etched using a highly anisotropic etching process to pattern the photoresist with hard mask underneath.

In the second step, the photoresist is etched away and the wafers are cleaned. Then, a highly anisotropic etch process is used to etch amorphous silicon and transfer the gate pattern from the hard mask into the amorphous silicon. After the amorphous

silicon etch step, the hard mask is removed resulting in the target number of continuous gate electrodes over the patterned fins on both the *p*-well and *n*-well. In order to form CMOS pair with *n*FinFET and *p*FinFET, the continuous gate electrodes are terminated to include a target set of parallel fins in the *p*-well and *n*-well, respectively by repeating the lithography processes.

In this step of forming CMOS pairs, a thick layer of amorphous carbon hard mask layer is deposited by the CVD process. Then a thin layer of BARC is spun across the wafer followed by resist deposition and the lithography process. Initially the lithography and photo mask processes are used to etch the hard mask using a highly anisotropic etch process. Then the photoresist is stripped and the wafers are cleaned. Next, a highly anisotropic etch recipe is used to transfer the pattern from the hard mask into the amorphous silicon. This etch terminates the gate electrodes to include the target set of fins in *n*-well and *p*-well to form a CMOS pair. *Note that the amorphous silicon gate electrodes have been subjected to the litho-etch-litho-etch* (LELE) process.

The minimum dimension of the amorphous silicon gate electrode defines the length L_g of a FinFET device at a technology node. And, the width W of a FinFET device is defined by $(t_{fin} + 2H_{fin})$. In reality, W of a FinFET can be increased by using gate electrode over multiple fins, thus, increasing the transistor drive current. Also, W can be increased by fabricating taller fins to increase H_{fin}. However, the ratio of H_{fin} and t_{fin} is a critical parameter affecting FinFET device performance.

4.3.6 Source-Drain Extensions Processing

After the formation of CMOS pairs, the photoresist, BARC, and the carbon hard mask layers are removed from the wafers exposing the fully formed polysilicon gate electrodes. Then wafers are cleaned and an ultrathin thermal oxide is grown on gate electrodes followed by the deposition of an ultrathin (≈1.5 nm) of CVD oxide to define the source-drain extension (SDE) *offset spacer* for source-drain extension processing [21].

4.3.6.1 *n*FinFET Source-Drain Extension Formation

After the oxide spacer formation, first a photoresist and then a TARC layer are deposited on the wafers. Next, the wafers are patterned with photoresist to cover the *n*-wells where the *p*FinFETs are fabricated. A dual Arsenic implant (typically, 2E15@1 Kev, ±10 degrees) is used to dope the fins in the *p*-well regions of the wafer to form SDE region of the *n*FinFET devices. This dual implant process is performed at +10 degrees and −10 degrees to ensure adequate coverage of the tall fins which are located alongside very tall photoresist coated structures. After the implant, the TARC and photoresist layers are removed and the wafers are cleaned to prepare for *p*FinFET SDE formation.

4.3.6.2 *p*FinFET Source-Drain Extension Formation

Again, photoresist and TARC layers are deposited on the wafers. Then the wafers are patterned with photoresist to cover the *p*-wells where the *n*FinFETs are fabricated. A dual boron SDE implant specification (typically, 2E15@1 Kev, ±10 degrees) is used

to dope the fins in the *n*-well regions of the wafer to form SDE region for the *p*FinFETs. Similar to the *n*FinFET SDE process, the dual implant step for *p*FinFET SDE is performed at +10 degrees and −10 degrees to ensure adequate coverage of the tall fins which are located alongside very tall photoresist coated structures. After *p*FinFET SDE implant, the photoresist is removed and the wafers are cleaned to prepare for dopant activation.

In order to activate the dopants for both the *n*FinFET and *p*FinFET SDE implants, the wafers are subjected to spike anneal followed by flash or laser anneal. During the annealing process, the SDE implants also diffuse under the polysilicon gate oxide in the fin-channel for a certain distance, thus, creating overlaps of the SDE implant in the ultrathin fin-channel under both ends of the polysilicon gate. This overlap is controlled by the thickness of the offset spacer which defines the effective channel length L_{eff} of the FinFET devices.

After SDE dopant activation, the wafers are prepared for raised source-drain (RSD) formation as outlined below.

4.3.7 Raised Source-Drain Processing

For RSD formation, a thick layer of Si_3N_4 is deposited on the wafers. Then a highly anisotropic etching process is used to form the Si_3N_4 spacers on the sidewalls of the polysilicon gates. However, even a highly anisotropic etching process leaves behind nitride residue on the sidewalls of the gate electrodes and fins. This is due to the two tall orthogonal vertical structures (gate electrodes and the fins) in the FinFET fabrication technology. The nitride spacer located on the fins is undesirable and must be removed for device reliability [18]. Finally, a highly anisotropic etch is used to remove the Si_3N_4 residue from the entire horizontal surfaces forming nitride sidewall spacers along the sides and ends of the gate electrodes and the fins. However, it is challenging to minimize the residual nitride spacer along the edge of the fins to maintain the target thickness of Si_3N_4 spacers along the gate electrode. After the Si_3N_4 spacer formation, RSD regions are formed for both types of devices.

4.3.7.1 SiGe *p*FinFET Raised Source-Drain Formation

After the Si_3N_4 spacer-etch process, a thin (≈30 nm) layer of SiCN hard mask is deposited over the surface of the wafers to cover all the underlying structures. Then the wafers are primed in HMDS followed by TARC and photoresist coat. After photoresist development and the masking process, the *n*FinFET region and the gate electrodes of the *p*FinFET region are covered in photoresist. After the lithography process, an anisotropic SiCN-specific etch is used to strip away the SiCN over the exposed *p*FinFET fins and the photoresist is stripped, and the wafers are cleaned. Then the *p*FinFET fins outside the Si_3N_4 spacer are etched away using a highly anisotropic etch process. The SiCN hard mask protects the *n*FinFET fins and the *p*FinFET gate electrodes.

After the *p*FinFET fin etch, the SiGe non-rectangular RSD regions are formed on the exposed silicon surface of the *p*FinFET source-drain fins by a selective epitaxial growth (SEG) of SiGe since every other part of the wafer is covered with nitride, or oxide, or SiCN hard mask. After SiGe growth, the remaining SiCN hard mask is

etched away using an etchant that is selective to SiCN and prepares the wafers for *n*FinFET RSD formation.

4.3.7.2 SiC *n*FinFET Raised Source-Drain Formation

After the *p*FinFET RSD formation, a thin layer of SiCN hard mask is deposited over the surface of the wafer to cover all of the underlying structures. Then the wafers are primed in HMDS followed by TARC and photoresist coat. After the lithography and masking process steps, the *p*FinFET region and the gate electrodes of the *n*FinFET region are covered in photoresist. The *n*FinFET RSD can be formed in two different ways: (1) grow epitaxial tip on the fins without *n*FinFET fin removal; or (2) by fin removal similar to *p*FinFET RSD formation process.

Epitaxial tip on fin-top: In this *n*FinFET RSD formation, an anisotropic SiCN-specific etchant is used to strip away the SiCN over the exposed silicon base of the *n*FinFET fins. Then the photoresist is stripped away and the wafers are cleaned. Next, the SiO_2 layer over the *n*FinFET fins is etched away to be followed by a deposition of silicon on the wafers using the SEG process. (Note that every other region of the wafers is masked by nitride, or oxide, or SiCN hard mask.) After RSD tip formation, the remaining SiCN hard mask is etched away. The epitaxial silicon caps on the top of the *n*FinFET fins are targeted to maximize the contact surface area of the *n*FinFET fins and provide a larger surface for the tungsten (W) contacts, thus, lowering the contact resistance and improving transistor speed.

SiC nFinFET RSD formation: For SiC RSD formation, the SiCN layer is deposited directly and is patterned with BARC and photoresist. Then the exposed SiCN is etched away in a highly anisotropic etch to expose the underlying fins of the *n*FinFET devices. Next, an isotropic SiCN-specific etch is used to strip away the SiCN over the exposed *n*FinFET transistor fins. The photoresist is then removed and the wafers are cleaned. After cleaning the wafer, the *n*FinFET fins outside the Si_3N_4 spacer are etched away in an anisotropic etch. The SiCN hard mask protects the *p*FinFET fins and the *n*FinFET gate electrodes. Then the SiC layer is formed on the exposed silicon surface of the *n*-well fins to form non-rectangular *n*FinFET RSD. After the SiC growth process, the remaining SiCN hard mask is etched away to prepare the wafers for silicidation.

4.3.7.3 Raised Source-Drain Silicidation

For RSD silicidation, the wafers are implanted with a silicon pre-amorphization implant. This step amorphizes the surface of the silicon and facilitates the formation of a more uniform, low resistance silicide in the next processing step [22]. After the silicon pre-amorphization implant, the wafers are prepared for RSD metallization and local contact formation.

Aluminum implant in SiGe RSD: First of all, the oxide layer on the gate electrodes as well as any native oxide on the surface of the RSD regions of the wafers are stripped away. Then the wafers are patterned with a photoresist layer to expose the *p*FinFET devices while covering the *n*FinFET devices. The patterned wafers are implanted with *p*-type dopants, aluminum (Al). The implant energy is chosen so that Al is located at the interface between the SiGe and the titanium (Ti) layer deposited in the next process step. This implant reduces the RSD contact resistance for the

*p*FinFET devices since the Al segregates to the top of the SiGe and being a *p*-type dopant, it lowers the Schottky barrier height for holes (typically, from about 0.4 eV to about 0.12 eV) and increases drive current by up to 19% [22, 23]. Next the photoresist layer is stripped away and the wafers are cleaned.

Titanium silicide formation: After the Al implant on the *p*FinFET RSD regions, the wafers are cleaned and a blanket thin layer of cold titanium (Ti) is deposited on the wafer using physical vapor deposition (PVD). Since Ti is a good getter, the oxygen and other contaminants are removed by the Ti layer [24]. Next, the wafers are rapidly heated to form titanium silicide ($TiSi_2$). After $TiSi_2$ formation, an additional thin layer of Ti is deposited while the wafer is at high temperature to form $TiSi_2$. This process produces in-situ $TiSi_2$ without voids and reduces production cycle time.

Unreacted titanium strip: After the annealing process, there are two types of Ti present on the transistor structure: unreacted Ti and reacted $TiSi_2$. The unreacted Ti located on the spacer sidewalls and on the top of the STI is etched away with highly selective etch. This leaves behind the reacted salicide on the top of the gate electrodes and over the epitaxial coated source-drain regions un-affected.

Oxide/nitride etch-stop layer deposition: After the silicidation process, the wafers are cleaned and a thin layer of SiO_2 followed by a thin layer of Si_3N_4 are deposited as an ESL for the next processing step of replacement metal gate and high-*k* gate dielectric.

4.3.8 Replacement Metal Gate Formation

In the previous sections, we have defined the gate geometry and processed SDE and RSD regions using polysilicon gate. This section outlines the process steps to replace the polysilicon gate with metal gate and high-*k* dielectric gate oxide. The major fabrication process steps include: (1) polysilicon gate removal; (2) high-*k* gate dielectric deposition; and (3) metal gate deposition and workfunction engineering.

4.3.8.1 Polysilicon Dummy Gate Removal

In the first step of replacement metal gate formation, a thick layer of phosphorus doped glass known as phospho-silicate glass (PSG) is deposited over the wafers with ESL using the plasma enhanced CVD (PECVD) process. This layer forms the first half of the pre-metal-dielectric (PMD). It is to be noted that the sharp corners and close proximity of the SiGe crystals "pinches-off" the PECVD disposition resulting in voids between the crystals. The void-free PECVD deposition is challenging, however, the small voids do not pose a significant problem on device performance. The PSG is then polished back using CMP to a lower thickness (≈120 nm).

After CMP, the top of the gate electrode is exposed and the amorphous silicon gate electrode is etched away using a very high selectivity silicon etch TMAH. This creates a cavity in the region of removed amorphous silicon. The entire cavity is lined with spacer oxide and the exposed fins are coated with an oxide ESL. *Note that this region of the fins does not have any SDE implants since they were covered by the amorphous silicon gate electrodes.*

After the removal of the amorphous silicon gate, the ESL inside the gate cavity is etched away exposing the inside walls of the cavity and the silicon fin.

4.3.8.2 High-*k* Gate Dielectric Deposition

The gate dielectric consists of an ultrathin SiO_2 interfacial layer and a high-*k* dielectric layer over the interfacial layer. The interfacial layer is thermally grown on silicon fins using a low temperature oxidation process [18]. This forms the bottom interface layer below the high-*k* dielectric and ensures smooth interface between the high-*k* material and silicon and prevents degradation of electron mobility.

After bottom interface SiO_2 gate layer growth, an ultrathin layer of hafnium oxide (HfO_2) high-*k* dielectric is deposited using atomic layer deposition (ALD) [18]. The high-*k* material is a blanket layer that covers the entire wafer; however, it is only required in the gate cavity over the fins.

4.3.8.3 Metal Gate Formation

pFinFET workfunction metal (TiN) deposition: After gate oxide deposition, the thin *p*FinFET workfunction metal gate is deposited using ALD. It consists of an ultrathin highly conformal layer of TiN (titanium nitride) that fills both the *p*FinFET and *n*FinFET cavities as well as coats the surface of the wafer. After TiN deposition, an ultrathin TaN ESL is deposited across the wafer using ALD. Next, a highly conformal thick layer of TiN is deposited using the ALD process to fill both the *p*FinFET and *n*FinFET cavities as well as coat the surface of the wafers.

A layer of photoresist is patterned on the wafer to cover only the *p*FinFET region and expose the TiN layer over the *n*FinFET region. After photoresist patterning to protect *p*FinFET gate metallization, the exposed TiN layer over the *n*FinFET gate fins is etched away using TaN as an ESL to this etch. After TiN removal from the *n*FinFET gate cavity, the photoresist is stripped and the wafers are cleaned to prepare for *n*FinFET workfunction metal gate.

nFinFET workfunction metal (TiAl) deposition: After the resist strip, a thin TiAl metal gate is deposited using advanced self-ionizing physical vapor deposition (SIPVD) technique. Then another highly conformal thin TiAl layer is deposited to cover the horizontal surfaces of both the *p*FinFET and *n*FinFET cavities and coat the surface of the wafer.

After TiAl deposition across the wafer, an anneal is performed to cause the Al in the TiAl to diffuse through the TaN barrier and creates the TiAlN (titanium-aluminum nitride) *n*FinFET workfunction metal on the top of the high-*k* dielectric in the *n*FinFET region. During the annealing process, Al diffuses rapidly into the TiN located on the *n*FinFET device region of the *gate cavity* forming TiAlN workfunction metal for *n*FinFETs devices. However, the thick TiN layer over the *p*FinFET region only blocks the diffusion of Al into the TiN *p*FinFET workfunction metal. Thus, the *p*FinFET workfunction metal does not become TiAlN.

Tungsten back fill: After anneal and metal gate workfunction engineering, a thick layer of highly conductive tungsten (W) is deposited using the CVD process to fill the gate cavities. Finally, the tungsten is polished back so that it is coplanar with the top of the gate electrodes.

4.3.9 Self-Aligned Contact Formation

After W deposition and planarization, self-aligned local contacts are formed for interconnection of the transistors. In the first step of the local contact formation

process, a recession is created in the *gate electrode cavity* by etch-back of W and metals in the surrounding gate cavity. Next, a thin layer of silicon oxynitride (SiON) is deposited on the wafers using the CVD process to fill the cavity of the recessed gate electrode with SiON. Then the SiON is polished back using CMP to make SiON coplanar with the PSG surface outside the gate electrode regions. Next, a thick layer of PSG is deposited on the wafers to complete the PMD deposition process and metallization for the formation of self-aligned local contacts for the fabricated devices.

4.3.9.1 Metallization

In the metallization process for local contact formation, a photoresist pattern is used to define the position of the contact trenches. Then the photolithography, masking, and anisotropic etch processes are used to cut down the contact holes to the gate electrode and the source-drain regions of FinFETs. Next the photoresist is removed and the wafers are cleaned to remove contaminants from the trenches as well as remaining polymeric residue and carbon contaminants. Then a thin Ti-liner followed by an ultrathin titanium nitride TiN barrier layer are deposited using the ion-metal plasma (IMP) physical vapor deposition (PVD) process [25]. Next, the wafers are annealed using the RTA process to react to the Ti/TiN layer and set the resistance of the Ti-liner.

Tungsten deposition and polish-back: After the deposition of the Ti/TiN barrier layer, a thin W seed is deposited in-situ on the wafers to line in the interior of the trenches and ensure conformal void-free bulk tungsten deposition in the trenches. Next, a thick layer of tungsten is deposited on the wafers by the CVD process. Then the wafers are polished using CMP so that the tungsten plugs are polished back to the top surface of the PSG, and the surface of the TEOS oxide is smoothed out to make the surface of the PSG coplanar with the top of the W-plugs.

Ta/TaN barrier metal patterning: After W deposition and CMP, a photoresist pattern is defined to determine the location of the contact trenches. Then lithography, masking, and etching operations are performed to cut down the contact holes to the gate electrode and the transistor source-drain regions. After opening the contact holes, the photoresist is stripped and the wafers are cleaned and appropriate processing steps are used to remove contaminants in the trenches and clean out any polymeric residue and carbon contaminants. Next, a thin TaN layer followed by a thicker Ta layer is deposited on the wafers using the IMP PVD process as the barrier metal to copper (Cu) trench contacts.

Copper fill: After the Ta/TaN barrier metal patterning, an ultrathin Cu seed is deposited in-situ on the wafers to line the interior of the trenches and ensure conformal void-free bulk-Cu deposition. Then a thick layer of bulk copper is deposited using electrochemical deposition process. Next, the Cu is subjected to low temperature annealing to anneal out the defects and reduce resistance. After the annealing step, the Cu is polished back using CMP which also removes the Ta that was on the upper surfaces of the TEOS. Then appropriate processing steps are used to smooth out the surface of the TEOS oxide and ensure that the surface of the TEOS is coplanar with the top of the copper lines. Finally, the wafers with Cu local contact metal are cleaned for back-end of the line (BEOL) fabrication processes.

4.4 SOI-FinFET PROCESS FLOW

As shown in the flowchart (Figure 4.2) for FinFET fabrication, the SOI-FinFET fabrication process eliminates the requirements for the formation of wells and STI to isolate neighboring FinFET devices. Thus, the major differences between the SOI-FinFET and bulk-FinFET fabrication are the starting material and fin patterning.

4.4.1 Starting Material

The starting wafers for SOI-FinFET fabrication consist of a layer of epitaxial silicon on a buried oxide (BOX) over a *p*-type base silicon substrate (Figure 4.1(b)). The thickness of the epitaxial silicon layer defines the height H_{fin} of the silicon body of a FinFET device. Similar to the bulk-FinFET process (Section 4.3.1), the wafers are cleaned and a thin layer of screen oxide is grown on the surface of the epitaxial silicon layer for fin patterning.

4.4.2 Fin Patterning: Spacer Etch Technique

As discussed in Section 4.3.3, the major processing steps in patterning multiple-fins using SADP technique include: (1) carbon hard mask or mandrel patterning; (2) offset spacer formation; and (3) fin formation as described below.

4.4.2.1 Mandrel Patterning

The mandrel patterning includes the deposition of a stack of layers over the screen oxide of the wafer: first of all, a thick layer of Si_3N_4 using CVD; then a thicker CVD amorphous carbon mandrel sacrificial layer to support the fabrication of oxide spacers for fin patterning; next a thin layer of dielectric BARC; and finally, photoresist coating and TARC on the top.

After the deposition of the stack of the above layers on the wafers, the photoresist and masking steps are used to create multiple photoresist patterns on the BARC of the wafers. These photoresist strips are used as the mask to pattern amorphous carbon mandrel layer using a highly anisotropic etch. The Si_3N_4 layer deposited on the wafer is used as the ESL to ensure that the etch process stops on the front surface of the Si_3N_4 layer. Then the photoresist patterns and BARC are stripped in ionized oxygen plasma and the wafers with carbon mandrel patterns are cleaned to prepare the top surface for oxide spacer formation.

4.4.2.2 Oxide Spacer Formation

After the mandrel patterning, a thin blanket layer of SiO_2 is deposited on the wafer using the CVD process. This oxide layer covers the wafers as well as surrounds the amorphous carbon mandrel at both ends of the carbon layer. In order to achieve straight lines at both ends of the oxide spacers, a photolithography process using a *cut mask* is used to trim the oxide spacers at both ends of the mandrel. Then a highly anisotropic etch process selective to oxide is used to form thin oxide spacers on the sidewalls of the amorphous carbon layer.

After the oxide spacers formation, the mandrel strips between the oxide spacers are removed using a selective etch process applicable to amorphous carbon. The

oxide spacers and the SiCN hard mask are unaffected by this etch process. After the mandrel removal, the wafers are cleaned in a Megasonic bath [19]. Then, using the oxide spacers as the hard mask, the underlying Si_3N_4 layer is removed by a highly anisotropic etching process to form the oxide spacer on the substrate. The wafers are then subjected to a Megasonic clean.

4.4.2.3 Silicon Fin Formation

After mandrel removal, the oxide spacers and Si_3N_4 hard mask patterns are used in a highly anisotropic etch to cut down through Si_3N_4, pad oxide, and epitaxial silicon to stop on the top of the BOX layer. The etching pressure, rate, and energy are extremely critical to create a sloped fin to achieve a smooth curved interface at the bottom of the fin. Furthermore, during the fin patterning process, the oxide spacers are partially etched and the shape of the Si_3N_4 hard mask strips underneath the oxide spacers are damaged at the top. Then the wafers are subjected to a Megasonic clean. In this etch process BOX is used as an ESL. Thus, the etch process to pattern silicon fins is significantly simpler than that of the bulk-FinFET fin formation (Section 4.3.3.3). As shown in Figure 4.2, there are no wells in an SOI flow and the BOX forms the STI. In addition, all of the fins are at a uniform height which greatly reduces the variability of device performance due to process variability [26,27]. Then, Si_3N_4 hard mask and pad oxide are removed exposing the epitaxial silicon fins standing on the top of the BOX on the base silicon substrate. Finally, any odd fin is removed following the process step described in Section 4.3.3.3.

After fin patterning, the remainder of the process flow is the same as that of the bulk-FinFET without the well and STI process modules.

4.4.3 Comparison of Bulk-Silicon FinFET and SOI-FinFET Fabrication Technology

From the discussion on bulk-FinFET and SOI-FinFET fabrication process flow, it is obvious that the major differences between bulk-FinFET and SOI-FinFET front-end CMOS fabrication processes are process complexities and hence cost. First of all, the bulk-FinFET fabrication requires formation of wells and STI to isolate each fin and therefore, each device on the substrate. The formation of STI requires deep etch to cut through the epitaxial silicon substrate to form trenches without the presence of an ESL. It is challenging to maintain the precise control of this etch process. Furthermore, it is challenging to maintain a uniform fin height as the trench filled TEOS oxide is etched back to the boundary between *n*-well and *p*-well to define H_{fin}, thus, causing variation in H_{fin} [18].

On the other hand, in the SOI-FinFET fabrication process each fin is patterned on the epitaxial silicon layer on the top of a BOX and they are already electrically isolated from each other. Therefore, wells and STI processes are eliminated. Again, H_{fin} is defined by the thickness of the starting epitaxial silicon layer on the BOX. Thus, a simple etch process with the top of the BOX as the ESL is used to pattern fins of uniform height. As a result, the variation in H_{fin} is less in SOI-FinFET fabrication process flow compared to the bulk-FinFET technology. Therefore, the SOI fabrication sequence reduces the process complexity and cost by eliminating a series

of challenging fabrication steps. However, SOI-FinFETs suffer from high defect density, floating body effect, and poor heat dissipation [5]. Furthermore, the raised source-drain formation by SEG is comparatively challenging in the SOI-FinFET fabrication flow.

Although the SOI-FinFET process flow presents a simpler front-end FinFET fabrication technology, the manufacturing cost is substantially higher than the bulk-FinFET technology due to the higher price of the SOI wafers compared to the bulk-silicon wafers. Therefore, cost-simplicity trade-off must be considered for volume production of FinFET fabrication technology.

4.5 SUMMARY

In this chapter an overview of the basic FinFET fabrication process flow is presented to introduce a representative FinFET device manufacturing technology. For the completeness of discussions, the bulk-FinFET as well as SOI-FinFET technology is highlighted. However, the bulk-FinFET process flow is described in some detail and only the differentiating features of the SOI-FinFET process flow are outlined. The major motivation of this chapter is to introduce the FinFET technology to appreciate the complexities of device fabrication and understand the device performance described in Chapters 5–10 of this book. Thus, some of the concepts may be different from the real silicon processing and presented often without explanation, or sketches, or justification, or references. Many companies and research laboratories build chips with final devices similar to the devices briefed in this chapter, however, with quite different process details. As, for example, the differences between the commercial process flows and that outlined here are the number and specifications of masks used. In commercial process with proprietary flows, the number of masks and process details varies from company to company. Some of the reasons for these differences between commercial process flows are equipment specific to the companies and application specific to the target technology node. Trade-offs in technology complexities and device performance may lead an individual company to process flow quite different than one we have overviewed. In spite of these differences between process flows from company to company, the final device structure and performances are comparable from all commercial manufacturers.

REFERENCES

1. J.-P. Colinge (ed.), *FinFETs and Other Multi-Gate Transistors*, Springer, New York, 2008.
2. J.P. Colinge, M.H. Gao, A. Romano-Rodriguez, H. Maes, and C. Clays, "Silicon-on-insulator 'Gate-all-around device.'" In: *IEEE Electron Devices Meeting Technical Digest*, pp. 595–598, 1990.
3. S.K. Saha, *Compact Models for Integrated Circuit Design: Conventional Transistors and Beyond*, CRC Press, Taylor & Francis Group, Boca Raton, MA, 2015.
4. J.D. Plummer, M.D. Deal, and P.B. Griffin, *Silicon VLSI Technology: Fundamentals, Practice and Modeling*, Prentice Hall, Upper Saddle River, NJ, 2000.
5. J.-H. Lee, T.-S. Park, E. Yoon, and J.J. Park, "Simulation study of a new body-tied FinFETs (Omega MSOFETs) using bulk Si wafers." In: *Proceedings of the Si Nanoelectronics Technical Digest*, pp. 102–110, 2003.

6. A. Gupta, M. Shrivastava, M.S. Baghini, *et al.*, "Part I: High-voltage MOS device design for improved static and RF performance," *IEEE Transactions on Electron Devices*, 62(10), pp. 3168–3175, 2015.
7. C. Auth, A. Aliyarukunju, M. Asoro, *et al.*, "A 10 nm high performance and low-power CMOS Technology Featuring 3rd generation FinFET transistors, self-aligned quad patterning, contact over active gate and cobalt local interconnects." In: *IEEE International Electron Devices Meeting Technical Digest*, pp. 673–676, 2017.
8. J. Markoff, *Intel Increases Transistor Speed by Building Upward*, New York Times, May 4, 2011. https://www.nytimes.com/2011/05/05/science/05chip.html.
9. C. Auth, C. Allen, A. Blattner, *et al.*, "A 22-nm-high performance and low-power CMOS technology featuring fully-depleted Tri-gate transistors, self-aligned contacts and high density MIM capacitors." In: *Symposium on VLS Technology*, pp. 131–132, 2012.
10. R. Merritt, *TSMC Taps ARM's V8 on Road to 16-nm FinFET*, October 16, 2012. www.eetimes.com/document.asp?doc_id=1262655.
11. D. McGrath, *Globalfoundries Looks to Leapfrog Fab Rival with New Process*, September 20, 2012. www.eetimes.com/document.asp?doc_id=1262552.
12. ThresholdSystems, Inc., Home page, https://secure.thresholdsystems.com/Home.aspx.
13. Y.-K. Choi, T.-J. King, and C. Hu, "A spacer patterning technology for nanoscale CMOS," *IEEE Transactions on Electron Devices*, 49(3), pp. 436–441, 2002.
14. K. Okano, T. Izumida, H. Kawasaki, *et al.*, "Process integration technology and device characteristics of CMOS FinFET on bulk silicon substrate with sub-10 nm fin width and 20 nm gate length." In: *IEEE Electron Devices Meeting Technical Digest*, pp. 721–724, 2005.
15. X. Huang, W.-C. Lee, C. Kuo, *et al.*, "Sub 50-nm FinFET: PMOS." In: *IEEE International Electron Devices Meeting Technical Digest*, pp. 67–70, 1999.
16. D. Hisamoto, W.-C. Lee, J. Kedzierski, *et al.*, "FinFET—A self-aligned double-gate MOSFET scalable to 20 nm," *IEEE Transactions on Electron Devices*, 47(12), pp. 2320–2326, 2000.
17. F.G. Pikus and A. Torres, "Advanced multi-patterning and hybrid lithography techniques." In: *Proceedings of the Asia and South Pacific Design Automation Conference*, pp. 611–616, 2016.
18. H.H. Radamson, Y. Zhang, X. He, *et al.,* "The challenges of advanced CMOS process from 2D to 3D," *Applied Sciences*, 7(10), p. 1047, 2017.
19. Z. Han, M. Keswani, and S. Raghavan, "Megasonic cleaning of blanket and patterned samples in carbonated ammonia solutions for enhanced particle removal and reduced feature damage," *IEEE Transactions on Semiconductor Manufacturing*, 26(3), pp. 400–405, 2013.
20. F. Nouri, O. Laparra, H. Sur, *et al.*, "Optimized shallow trench isolation for Sub-0.18-um ASIC technologies." In: *Proceedings of the SPIE Conference on Microelectronic Device Technology II*, vol. 3506, pp. 156–166, 1998.
21. S. Saha, "Design considerations for 25 nm MOSFET devices," *Solid-State Electronics*, 45(10), pp. 1851–1857, 2001.
22. H. Yu, M. Schaekers, S.A. Chew, *et al.*, "Titanium (Germano-)silicides featuring 10−9 Ω·cm2 contact resistivity and improved compatibility to advanced CMOS technology." In: *Proceedings of 18th International Workshop on Junction Technology*, pp. 1–5, 2018.
23. M. Sinha, R.T.P. Lee, K.-M. Tan, *et al.*, "Novel aluminum segregation at NiSi/*p*+-Si source/drain contact for drive current enhancement in *P*-channel FinFETs," *IEEE Electron Device Letters*, 30(1), pp. 85–87, 2009.
24. V.L. Stout and M.D. Gibbons, "Gettering of gas by titanium," *Journal of Applied Physics*, 26(12), pp. 1488–1492, 1955.

25. G.A. Dixit, W.Y. Hsu, A.J. Konecni, *et al*., "Ion Metal Plasma (IMP) deposited titanium liners for 0.25/0.18 μm multilevel interconnections." In: *IEEE Electron Devices Meeting Technical Digest*, pp. 357–360, 1996.
26. K.J. Kuhn, M.D. Giles, D. Becher, *et al*., "Process technology variation," *IEEE Transactions on Electron Devices*, 58(8), pp. 2197–2208, 2011.
27. S.K. Saha, "Modeling process variability in scaled CMOS technology," *IEEE Design & Test of Computers*, 27(2), pp. 8–16, 2010.