SPIN-ON GLASS
MATERIALS AND APPLICATIONS IN
ADVANCED IC TECHNOLOGIES

PROEFSCHRIFT

ter verkrijging van
de graad van doctor aan de Universiteit Twente,
op gezag van de rector magnificus,
prof. dr. F.A. van Vught,
volgens besluit van het College voor Promoties,
in het openbaar te verdedigen
op vrijdag 26 februari 1999 te 13:15 uur

door

NGUYỄN NHƯ TOÀN
geboren op 14 februari 1970
te Hanoi, Vietnam

Dit proefschrift is goedgekeurd door de promotoren

Prof. dr. P. H. Woerlee

Prof. dr. Nguyễn Đức Chiến

en de assistent promotor

Dr. J. Holleman

Nothing is too small to know,

and nothing too big to attempt.

~ Sir William Van Horne ~

To my parents and my family

Samenstelling van de promotiecommissie:

Voorzitter:	
Prof. dr. A. Brand	Universiteit Twente
Secretaris:	
Prof. dr. A. Brand	Universiteit Twente
Promotoren:	
Prof. dr. P. H. Woerlee	Universiteit Twente/ Philips Research
Prof. dr. Nguyễn Đức Chiến	Hanoi University of Technology/ ITIMS
Assistent promotor:	
Dr. J. Holleman	Universiteit Twente
Leden:	
Prof. dr. H. Wallinga	Universiteit Twente
Prof. dr. ir. P. Bergveld	Universiteit Twente
Prof. dr. C.A.J. Ammerlaan	Universiteit van Amsterdam
Prof. dr. A.F. Willoughby	University of Southampton

Title: Spin-on glass materials and applications in advanced IC technologies
Author: Nguyễn Như Toàn
ISBN: 90-365-12697

This research has been supported by the Netherlands University Foundation For International Co-operation (NUFFIC) and the International Training Institute for Materials Science (ITIMS) in Hanoi, Vietnam.

Contents

Preface

Smaller, faster and smarter electronic equipment is coming to the world every day. However, people are never satisfied and semiconductor technology has to face many difficult challenges as we enter the 21st century. According to Moore's Law, which states that device complexity doubles about every 18 months, among many challenges the feature size needs to be reduced to approximately 1.4 times for every generation. For future sub-0.1 μm MOSFET, new type of MOSFET structures or new materials are required. A scaled MOSFET (Metal Oxide Semiconductor Field Effect Transistor) requires a thinner gate, different gate materials and very shallow source/drain junctions. When the gate length is reduced to less than 0.1 μm, an extremely shallow source/drain junction of less than 40 nm is needed to suppress the short-channel effect. Furthermore, the shallow junction also needs to be highly doped for series resistance reduction. Very low-energy implantation could result in highly doped shallow junctions. However, these shallow junctions suffer from implantation damage. As a result, the junctions become much deeper after the annealing process due to transient enhanced diffusion. Diffusion from highly doped spin-on glass - SOG (or spin-on dopant - SOD) is one of the alternatives.

Shallow junction formation using SOD is the subject of this thesis. Very shallow junctions of less than 20 nm have been realized using spin-on dopant diffusion into Si. The diffusion of impurities from SOD into polysilicon on silicon structure was also investigated. Very high quality shallow junctions with low sheet resistance were obtained by this technique. The results were published and a poster of this subject was awarded as best poster for a scientific contribution in the 1st annual BENELUX workshop on "Semiconductor Advanced for Future Electronics" in Mierlo, the Netherlands (1998).

This diffusion technique has been applied successfully to many subjects, e.g. shallow polysilicon emitter formation in high frequency bipolar transistors and shallow junction formation in elevated source/drain MOSFET. For elevated source/drain MOSFET application, a new type of MOSFET called poly-spacer MOSFET has been invented. This poly-spacer MOSFET has all the features of future elevated source/drain MOSFET. In addition, the process is very straightforward and reproducible. A 0.25 μm poly-spacer PMOSFET has been realized successfully using SOD diffusion source to prove the new concept. A paper is in preparation and will be published soon. However, this research already has got appreciation in a recent technology briefing of "The Incredible Shrinking Transistor" published in Electronic Design journal (Nov. 2, 1998, p.18).

In the following paragraphs, a short review of each chapter in this thesis will be presented.

- *Chapter 1 - Introduction To Spin-On Glass Materials and Applications:* The sol-gel coating technology including SOG is presented in the first section of this chapter. Then structures and properties of three SOG material types, e.g. silicate SOG, siloxane SOG and dopant-organic compound, are discussed in the second section. In the last section, several important applications of SOG, e.g. dielectric planarisation and diffusion source and others, are presented.

- *Chapter 2 – SOG fabrication and characterization:* Fabrication of SOG and phosphorus doped SOD materials is discussed in the first section. Design of experiment method is introduced in realizing SOG materials. Effects of several factors in the sol-gel reaction of SOG fabrication are investigated. The fabrication of SOD (phosphorus doped) is then carried out using the same sol-gel technology as SOG. In the second section, characterization of two SOD materials is carried out using several advanced analyzing techniques. The analysis gives us information about elements and their concentration profiles in the SOD layer and the evolution of the SOD layers during processing can be understood.
- *Chapter 3 – Shallow PN junction formation using SOD:* The diffusion study of boron and phosphorus from SOD sources into Si and polysilicon on Si structure is presented in this chapter. Impurity concentration profiles in Si and poly-Si/Si are shown together with their diode characteristics. Simulation results are also shown here. Conventional furnace and rapid thermal processes are used in this study. The first section deals with impurity diffusion from SOD into silicon. The second section deals with impurity diffusion from SOD into polysilicon on silicon structure. Study in the second section is a new research approach for shallow junction formation. The polysilicon layer is used as a buffer layer for diffusion from SOD into Si. As a result, shallow junctions can be obtained and the sheet resistance of the junctions is reduced due to the highly doped polysilicon. This low resistance poly-buffered shallow junction is very important for poly-emitter bipolar transistors and for future elevated source/drain MOSFET. The results in this chapter are applied for these advanced devices presented in the next chapters.
- *Chapter 4 – Realization of high frequency bipolar transistor:* This chapter deals with optimization of the UT-bipolar device. The bipolar transistor process using very high energy implantation for collector formation is optimized using simulation. Three process variants are generated for different purposes. The first variant is proposed for minimum change of the standard UT-BiCMOS and better reproducibility. Rapid thermal annealing is implemented in the second variant for better device performance. The third variant is similar to the second one but the SOD material is used here for shallow poly-emitter formation. Optimization procedures and simulation results are presented in the first and the second sections. The third section deals with the fabrication process of three variants. Characterization results of the devices are also presented in this section.
- *Chapter 5 – Realization of submicron poly-spacer MOSFET:* Future sub-0.1 μm MOSFET requires elevated source/drain structure for better performance. In this chapter a new structure of MOSFET called poly-spacer MOSFET is proposed, designed and fabricated. This new structure is simple for processing but it also has all the requirement of a elevated source/drain MOSFET. The SOD is applied here for shallow source/drain junction formation. The first section in this chapter presents scaling laws of CMOS and several technology limitations. In the second section, the poly-spacer MOSFET is proposed and 0.25 μm PMOS is designed using simulation tools. The third section deals with fabrication of 0.25 μm PMOS. In this section, the photoresist ashing technique is studied for deep submicron gate length definition because our photo-lithography technology only allows for a minimum 1 μm feature. Processing details are presented next and device characterization results are shown at the end of this section. Conclusions and recommendations for process improvements are presented in the last section.

- *Chapter 6 – Conclusions and recommendations:* Final conclusions of our study in this thesis are presented in the first part of this chapter. Subsequently, several recommendations are proposed for future research and technology improvements.
- *Appendixes:* Details of SSUPREM-IV diffusion model and several process flows of bipolar transistors as well as 0.25 μm poly-spacer PMOS are presented here.

1

Introduction to Spin-On Glass Materials & Applications

1.1. Introduction

The technology of sol-gel thin film has been around for over 20 years and is now well accepted as a technology for forming thin films and coatings. The process is simple, a solution containing the desired oxide or non-oxide precursor is prepared and is applied to a substrate by spinning, dipping, draining or spraying. The process is able to apply a coating to the inside and the outside of complex shapes simultaneously. The films are typical a few hundreds of nanometers thick, uniform over large areas and adherent. The equipment is inexpensive, especially in comparison to any deposition technique that involves vacuum. Coatings can be applied to metals, plastics, and ceramics. Typically, the coatings are applied at room temperature, though most need to be calcined and condensed by heating. Both amorphous and crystalline coatings can be obtained. There are many useful applications of sol-gel technology such as coating for optical, electronic, sensor, abrasion, barrier, protective, and catalyst applications. In some cases the quality of the coatings obtained by the sol-gel process is limited in comparison to oxidation, evaporation or sputtering techniques. Furthermore, the shelf-life of sol-gel solutions is limited to typically 12 months.

In the semiconductor industry the sol-gel method is often used to deposit silicon dioxide. The method is known there as the Spin-On Glass method (SOG). Spin-On Glass materials have been widely used as a diffusion source or a planarizing dielectric for multilevel metalisation schemes in the fabrication of nowadays integrated circuits. SOGs are in general Si-O network polymers in organic solvents, and prepared through the hydrolysis-condensation reaction that implied the sol-gel technology. SOG materials can be divided into three groups: 1) Silicate based compounds, 2) Organosilicon compounds and 3) Dopant-organic compounds.

In the following section of this chapter the sol-gel coating technology will be described. The next section will focus on the three types of Spin-On Glass materials and their properties. In the last section we will discuss about some typical applications of SOG.

1.2. Sol-Gel coating technology

The sol-gel process is the name given to any process involving a solution or sol that undergoes a sol-gel transition. A solution is a truly single-phase liquid, while a sol is a stable suspension of colloidal particles. At the transition, the solution or sol becomes a rigid, porous mass by destabilisation, precipitation, or supersaturation. Most of the coatings (spinning, dipping, or draining) present the sol-gel transitions from true one-phase solutions to a rigid two-phase system of solid and solvent-filled pores. The sol-gel transition in this case is not reversible.

Films are normally formed from SOG sols using the spinning technique. Most of SOG sols need to be stored in a refrigerator and they must be at room temperature before applying it to the substrate. A substrate, usually a silicon wafer, must be cleaned before coating. The substrate is placed on a spinner and about one ml of SOG is dropped on the centre of the substrate. Then the substrate is immediately rotated with a speed of a few thousand cycles per minute. In most cases, a film thickness between 50 and 500 nm will result. Controlling the thickness is a matter of controlling the solution viscosity. Typical solution viscosity is 3-10 mPa-s. Typical surface tension is $30\text{-}50 \times 10^{-3}$ N/m. Knowing the viscosity and oxide concentration of the solution, a film thickness can be achieved by controlling the spinning rate.

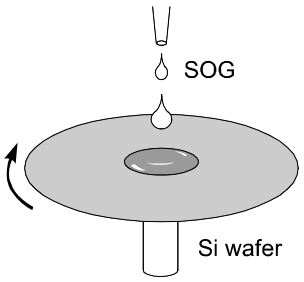


Figure 1-1: Thin film formation by spinning technique

When it comes to coatings, the SOG is said to go through a sol-gel transition, once the coating is on the substrate. What takes place on the molecular level is not well understood, but what is observed is that the solution film of perhaps 100 nm undergoes a sharp increase in viscosity. If the substrate has been handled properly, it is now covered uniformly by a tacky gel.

In the systems described above, the sol-gel transition is reached when one-phase liquid becomes a two-phase alcogel, solid plus liquid. The alcogel is an oxide polymer that condenses in the presence of solvent. Alcogel is used to differentiate gels prepared with alkoxides from those prepared from ion-exchanged solutions or colloidal sols. Those gels are called hydrogels. The transition in alcogels is irreversible and occurs with no change in volume. The time of the transition depends on the chemistry of the solution, but the chemical composition of the two phases at the transition is not unique. Once through the sol-gel transition, the solvent phase is removed to create xerogels by ordinary evaporation or aerogels by hypercritical evacuation. At this point, the dried gel is a micro-porous oxide.

After deposition, the film must be dried. During this step there are reductions by more than 50 % in weight and volume of the sol and gel. Yet the film remains adherent and continuous and maintains complete surface coverage. It has been shown repeatedly that all shrinkage is taken up in the dimension perpendicular to surface and not in the plane of the substrate. This holds as long as the thickness is about one micron or less. Drying is complete when there is no further weight loss. For the purpose of out-gassing, the drying temperature should be approximately 250^{o}C.

The next step is the heat treatment. To go from sticky gel to a hard gel usually takes about 30 minutes at temperatures less than 600^{o}C. Films can be dried quickly in air because of their small thickness. Water and solvent evaporate through interconnected pores which remain open at the surface. Then the hard gel may be heated to various degrees of collapse. The micro porosity in silica is not removed entirely until 1000^{o}C, but it may already behave as an oxidation barrier or passivation coating at 600^{o}C. This ability of the micro-porous film to behave in many ways like the bulk oxide is an attractive property of the sol-gel approach to coating.

1.3. SOG materials and their properties

SOG materials can be divided into three groups: silicate based compounds, organosilicon compounds and dopant-organic compound.

1.3.1. Silicate based compounds

The silicate SOG is formed from a condensation reaction of $Si(OH)_4$ by losing water. When the film is fully cured, the film should form a strong Si-O network and contain no -OH but it has fairly significant shrinkage. A rough description of the molecular structure [1] is presented in the following figure:

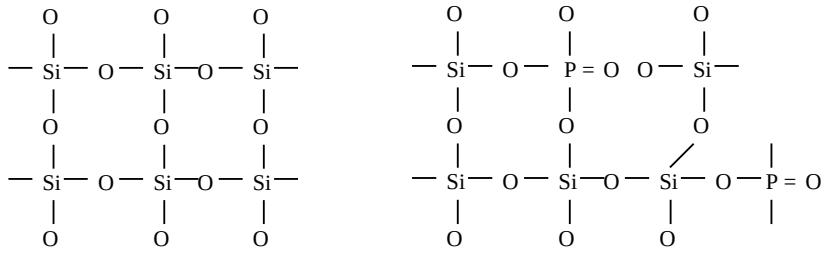


Figure 1-2: a) Silicate SOG and b) P-doped silicate SOG

The film shrinkage creates high tensile stress which may lead to cracking problem. A typical stress value of a fully cured film is about 500 MPa in tension. The viscosity of the silicate SOG solution is fairly low compared to other types. As a result, the obtained layer by each spin is very thin. The planarity is generally poor and it normally requires multiple spins to achieve good planarity. After condensation, the silicate SOG is thermally stable and is not sensitive to moisture and oxygen plasma. This provides the conditions for a non-etch back process [2, 3]. However, for planarisation application in IC technology the maximum temperature at which a SOG film can be cured is often limited to about 450^{o}C because of the presence of aluminium interconnects. After such low-temperature cures, the SOG film is far from being condensed, and may contain significant amounts of silanol, ≡Si-OH, and absorbed water. This effect could degrade the dielectric properties of this SOG layer. Of course, when there is no aluminium metalisation on the substrate, the SOG film can be condensed at high temperature, typical at 800^{o}C or 900^{o}C, to obtain a silanol- and water-free film.

Silicate SOG may be doped with phosphorus or other dopants as illustrated in figure 1-2b. Doping the glass, which modifies the Si-O network, will reduce the film stress to 200 MPa, typically. This helps against the cracking problem. A doped SOG is often called Spin-On Dopant (SOD) when it is used as a diffusion source. However, when the phosphorus concentration exceeds 4 % the film may become hygroscopic. We have observed water nodules on the wafer surface on a highly-P doped silicate SOG but it does not affect the diffusion properties.

It is more serious with boron doped SOD. A high concentration boron doped glass may have some problems such as boron skin and moisture effect. Since elemental boron is inert, boron diffusions are carried out by means of a surface reaction between B_2O_3 and Si that take place at diffusion temperatures:

$$2\ B_2O_3 + Si \Leftrightarrow 3\ SiO_2 + 4\ B$$

The elemental boron formed diffuses into the silicon forming a doped layer whose surface concentration of boron corresponds to the solid solubility limit at the

temperatures of the silicon substrate. If excessive amounts of B_2O_3 are present, SiB_x, species are formed. This layer, called boron skin, can be several hundred angstroms thick and is not easily removed by any chemical solution means. However, the silicon beneath it can be oxidized and the skin subsequently removed by undercutting in HF. Further more, B_2O_3 is very hygroscopic and reacts with water at room temperatures to form boric acids:

$$B_2O_3 + 3\ H_2O \Rightarrow 2\ H_3BO_3$$

$$B_2O_3 + H_2O \Rightarrow 2\ HBO_2$$

Since H_3BO_3 and HBO_2 are volatile compounds at the diffusion temperature, these give rise to variation in the amount of B_2O_3 reacted with silicon. Thus great care must be taken to keep the deposited wafer from coming in contact with moisture if reproducible diffusions are to be achieved. The problem can be solved by placing the coated wafers into a hot furnace immediately after baking to drive in the boron impurity.

1.3.2. Organosilicon compounds (Siloxane SOG)

The organosilicon compound SOG or Siloxane SOG contains $–CH_3$ or $–C_2H_5$ types of organic dopants.

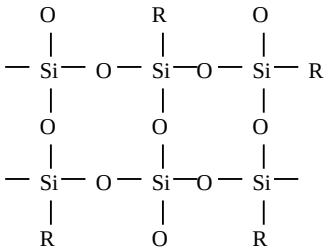


Figure 1-3: Molecular structure of Siloxane SOG

In this polymer structure, the organic group modifies the Si-O network which lowers the film stress to 200 MPa, typically. The viscosity and molecular weight are slightly higher than that of the silicate SOG. The film is thicker than that of the silicate SOG using the same spin speed and its planarity is better. It has been observed that some of the siloxane SOG films pick up atmospheric water and this may become a reliability issue or cause extensive processing constraints. Another major limitation of siloxane SOG is the decomposition of the organic group during the curing process particularly under oxygen plasma. The thermal decomposition may happen at temperature as low as 400^oC. It can be seen from the FTIR spectra before and after 400^oC baking that Si-OR and Si-R (R stand for organic groups) peaks disappear and Si-OH peaks show up. The -OH content not only reduces the device performance (due to high dielectric constant) but also creates reliability issues. Because of the instability caused by moisture and exposure to oxygen plasmas which will occur in subsequent processing steps, siloxane SOG has been more commonly used in the etch back process.

Increasing the organic content will further reduce the film stress and improve the crack resistance and planarity. However, it was found that the mechanical strength of this material is very low. This type of SOG suffer from potential thermal decomposition. Even though this type of SOG is expected to be similar to silicate SOG, it is not sensitive to moisture. In fact, this difference is not fully understood.

1.3.3. Dopant-organic compound [4]

The two types of SOG described above are SiO_2- based compounds. They suffer from moisture sensitivity and short-life because of inherent instability of the silicate glass matrix in solution, especially boron doped SOG. Therefore, a boron compound was sought that could be incorporate in low volatility polymer which is stable under attack by water and forms a stable solution in common organic solvents. This new spin-on glass is based on a boron-nitrogen backbone polymer dissolved in toluene. The polymer is synthesized from a class of boron-nitrogen compounds, known as borazole (or borazine), whose structure is represented by:

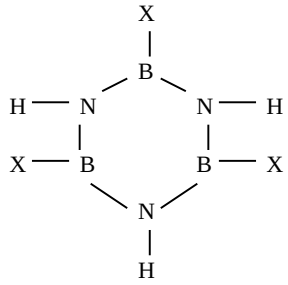


Figure 1-4: Molecular structure of borazole

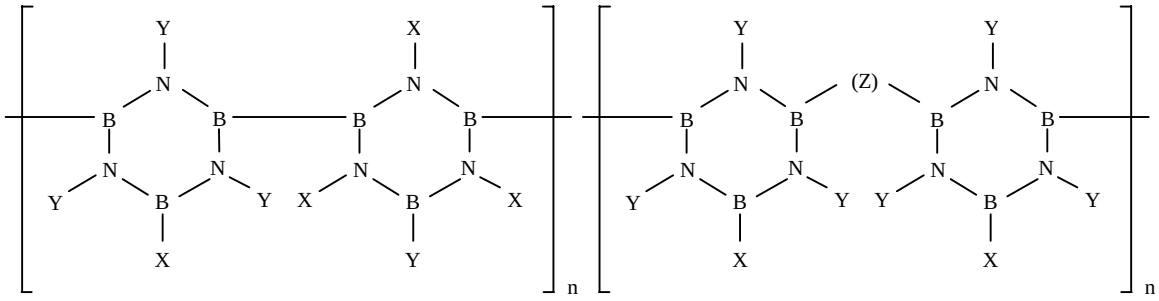
In preparing the polymeric dopant composition, a polymeric borazole is produced by first reacting BCl_3 and an amine at temperatures between -60°C and -5°C. in an inert solvent to form an intermediate triaminoborane. After removal of the amine hydrochloride salt, the intermediate is polymerized via thermal condensation at a temperature between 380°- 420°C. The polymerization process is controlled so as to yield a polymer that is soluble in toluene (and other solvents) in amounts of up to 50 % (wt/wt) and at the same time is not very volatile or sensitive toward moisture. Polymerization is monitored by string temperature. (Equivalent to melting point.) The polymerized material is "refined", that is, unreacted monomer, low molecular weight polymers and other volatile impurities are removed by purging the resin with dry nitrogen at polymerization temperatures and during the cool down. The polymer, in the form of a glassy resin (string temperature $140^oC \pm 30^oC$), is formed upon cooling and then dissolved in an organic solvent such as toluene or xylene at a 1 to 20 % (wt/wt) level. Cyclohexamine (5 %) or the like material is added to the solution of the polymer in toluene to stabilise the solution against precipitation of certain components from the solution during the spin-on application of the material on to the silicon wafers. A suitable commercially available surfactant is added to the solution to impart suitable properties for the formation of defect-free films of uniform thickness, by the spin-on technique. The surfactant is used at about 0.1 % (w/v) level. The resulting solution is then filtered through 0.2 micrometer filters.

In this method, the polymers derived from the borazines are of two general types:

a) polymers with borazine rings linked directly between a boron in one of the six-membered boron/nitrogen ring and a nitrogen atom in the adjacent six-membered boron/nitrogen ring (figure 1-5a) and

b) polymers, at least a substantial portion of whose backbone is borazine rings linked via other (different) groups A, i.e.: (figure 1-5b) wherein Z is selected from the group consisting of -N(R)-, -O-, -N(R)-R1 -N(r)-, and -R-, and wherein R and R1 denote organic radicals of 1 to 10 carbon atoms and X and Y are radicals selected from the group consisting of hydrogen, alkyl radicals of 1-8 carbon atoms, unsubstituted

aromatic radicals, heterocyclic aromatic, 1-6 carbon atom alkyl substituted heterocyclic aromatic, amino, hydroxy, alkoxy and halogen radicals.



a) *b)*

Figure 1-5: Polymeric boron nitrogen dopant structures

In employing the borazine polymers of this method, the traditional advantages of a spin-on dopant are maintained i.e., greater uniformity, less capital cost, higher throughput, more control, and less cleaning of diffusion tubes and boats.

The borazole polymer decomposes on heating to boron-nitrogen-hydrogen moieties (moiety is one of the portions into which something is divided) which in turn react with oxygen in the organic groups and occluded (absorbed) water from spin-on process or oxygen from the furnace ambient, to form B_2O_3. B_2O_3 then reacts with Si and boron diffuses into the silicon substrate.

This spin-on glass is supplied as a dark brown liquid of low viscosity and is often used for diffusion application. The shelf life of the solution can be more than one year at room temperature. Because the toluene is a hazardous material this kind of SOD must be handled carefully to prevent inhalation of SOD or contact with the skin.

1.4. Applications of SOG

1.4.1. Dielectric planarisation

The fabrication of reliable ULSI circuit structures require multiple layers of metalisation with uniform inter-metal dielectric thickness and subsequent patterning of metal layers with good line width control. The best way to achieve these processing goals is the interlevel planarisation of the substrate with dielectric materials prior to the metalisation step. Among various techniques for planarisation, techniques based on SOG films are advantageous because of the process simplicity, good adhesion characteristic of SOG, low level of stress and shrinkage in the SOG film, etc.

The limited film thickness of the available SOG materials, though not sufficient to allow their use as a stand-alone interlevel dielectric layer, is adequate for planarizing or smoothing a wide range of substrate topographies. Very little planarisation (defined as percent reduction in step height) is obtained over isolated line or lines separated by 3- or 4-microns-wide spaces. However, the 90-degree angle of the steps is reduced to about 45 degrees. This "smoothing" of the vertical-walled features is quite suitable for comformal deposition of subsequent layers with a high degree of step coverage. At smaller geometry, where the aspect ratio of the space between the lines become unity, a high degree of planarisation is produced by similar thin SOG films. At such geometry's, a mere smoothing would not be accepted. In either case, the SOG thickness above the lines is too small to provide interlevel insulation.

Thus, the SOG planarisation processes in use today employ SOG films primarily as a planarizing agent, with the bulk of the dielectric insulation functions being provided by CVD oxide layers. In some schemes, SOG films are used as a sacrificial planar layer. Four of the more common schemes for planarisation with SOG are [5]:

- CVD/SOG two-layer dielectric

- CVD/SOG/CVD sandwich dielectric

- Partial etch-back of SOG in a sandwich structure

- Total etch-back of SOG

However, there are still many processing problems. The most common types of problems encountered in the use of SOG processes include: particulate contamination, cracking of dielectric layer, poor adhesion (delamination), and poor via contacts. The three first problems have been eliminated by good processing equipment and handling or good SOG source. The last problem arises from the formation of insulating deposits at the via bottom by the reaction of aluminium with moisture outgased by SOG during metal deposition. The outgassing from SOG can be due to incomplete curing of the SOG, or it may be due to the inherent water content of SOG. In most cases, a dehydration step, carried out in-situ in the sputtering apparatus or externally, is effective in eliminating the problem. Above all, the SOG technique does not lead to global planarisation, especially when the number of metalisation layers increase. This is the reason why Chemical Mechanical Polishing (CMP) is used for planarisation in modern ULSI technology.

1.4.2. Diffusion source

SOG materials are also used as diffusion sources to make P-N junctions in semiconductor technology. In this case Spin-On Glass is called Spin-On Dopant (SOD). The diffusion using SOD which is based on the diffusion technique using liquid sources has been applied since the seventies but recently SOD has become more popular in IC processing.

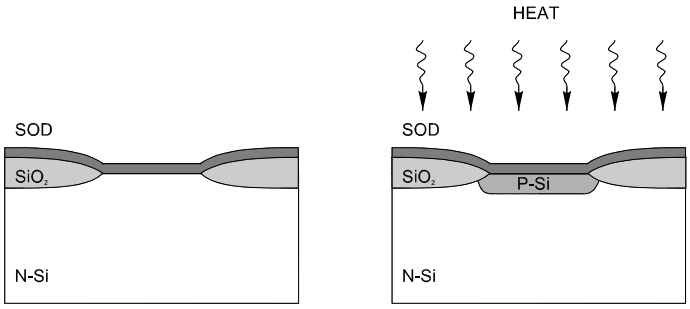


Figure 1-6: Diffusion from Spin-On Dopant

There are several techniques which use SOD as a diffusion source: one-step diffusion, two-step diffusion and proximity diffusion. The simplest technique is the *one-step* diffusion (figure 1-6). A wafer is coated with SOD by a spinning technique as described in the previous sections. After baking, there is a doped oxide layer on the top of the wafer. Then the coated wafer is placed into a furnace, a P-N junction is created since the impurities diffuse into the semiconductor material from the SOD layer under high temperature conditions. In the *two-step* diffusion technique, the first step is deposition then a constant-source diffusion is carried out similar to the one-step diffusion but at lower temperature for a short time. Then, the doped oxide layer is etched off. In the second step, called drive-in, the wafer is subjected to a high temperature for an

appropriate length of time to obtain the desired junction depth. A third technique called *proximity* diffusion, is a non contact diffusion that uses a dopant source (source wafer) made by spin-coating a thermomechanically stable disk with a SOD. In the furnace the source and the process Si wafers, stacked in proximity, are heated so the dopant can evaporate from the source wafer and be transported to the surface of the process Si wafer. In all techniques the furnace could be replaced by other heat sources such as Rapid Thermal Processing (RTP) or laser beam.

There are several advantages of the SOD diffusion technique:

- no implantation defects, thus no transient-enhanced diffusion and very shallow junctions could be obtained
- uniform and consistent doping with high yield
- elimination of use of toxic gases
- no storage of source wafers or use of costly ion implant equipment.
- application for both planar and 3-D structures

On the other hand, SOD technology has some drawbacks such as possible particulate defects in the films during processing or toxic nature of some SOD. Furthermore, for the diffusion techniques in which deposition of SOD directly on a silicon substrate, removal of the glass layer after the diffusion process can lead to serious device deterioration. For example, removal of thick dopant glass layers can lead to a decrease of the masking oxide thickness. Such oxide thinning is dangerous, especially at the junction edges, since it can result in high leakage currents as observed for shallow junctions doped from SOD. In addition, doping from SOD leaves a residual film on the silicon substrate and may deteriorate device operation. Moreover, SOD does not ensure junction uniformity in high aspect ratio trench structures and removal is especially difficult from such trenches. These disadvantages could be eliminated by the proximity diffusion technique.

In VLSI and ULSI (very and ultra large scale integration) circuits the small size of the individual devices imposes stringent requirements of source/drain dopant distribution and demands that the junctions be extremely shallow and heavily doped as illustrated in the following figure.

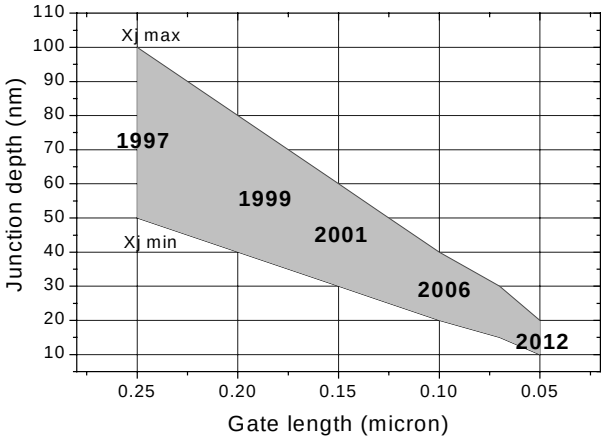


Figure 1-7: Junction depth scaling with MOSFET gate length

This figure is extracted from NTRS roadmap [6] showing the required junction depths scaled down from 0.25µm gate-length MOSFET generation in 1997 to an expected 0.05µm generation in 2012. Depending on the design, the junction depth has to be within the grey area which is optimum for MOSFET with a certain minimum gate

length. A high-performance MOSFET requires a deep source/drain junction for high-drive current. However, this junction also needs to be shallow enough to prevent short-channel effect and it has to be scaled together with the gate length of the MOSFET. Hence, an extremely shallow junction depth is required for future CMOS.

Recently, rapid thermal diffusion from SOD has emerged as a promising method for shallow junction formation. The Rapid Thermal Process (RTP) technique uses halogen lamps or arc-lamps to heat up the coated wafer at very high speed (100^o – 400^oC/sec.) so the temperature could be raised in very short time. As a result the thermal budget is reduced and the junction depth of the created P-N junction could be very shallow while the surface concentration is high. Shallow junctions less than 20 nm were first reported by Akira Usami [7]. They found that the RTD (Rapid Thermal Diffusion) of P and B was enhanced with the heating rate. This was ascribed to the stress field induced in the heating stage. The RTD from SOD has further been studied and applied to solar cells by L. Ventura and co-worker [8]. They showed that the efficiency of RTD from phosphorus SOG films deposited on Si wafers depends on the source composition and its thickness. Solar cells obtained by this technique have shown promising characteristics, which confirmed that the use of a remaining thin doped SOG layer is a suitable technique to make a good surface passivation without inducing minority-carrier diffusion length degradation in the base of a solar cell. A high-efficiency mono-silicon solar cell by RTD from SOD has been realised by D. Ruby (16.9 %) [9].

Recently, a new diffusion technique called proximity rapid thermal diffusion has been announced by W. Zagodzon [10]. Fabrication of sub-micron junctions by proximity rapid thermal diffusion of phosphorus, boron and arsenic has been studied [11]. Proximity RTD is presented as an encouraging process which results in very shallow junctions with high dopant concentration at the surface. This technique eliminates the disadvantages of the diffusion technique from top SOD layer deposited directly on silicon wafer. Moreover, diodes fabricated by proximity RTD have low defect density p-n junctions resulting in low leakage currents (≈10 nA/cm^2). However, proximity RTD has some limitations such as low efficiency of SOD source after first doping process or low efficiency diffusion of arsenic.

In this thesis we studied the shallow junction formation using SOD diffusion sources [12]. PN junctions with junction depths of less than 20 nm have been realised using RTD of SOD. However, very low series resistance shallow junctions are needed for future CMOS. Diffusion from silicide or polysilicon layers could solve this problem. We had studied shallow junction formation by diffusion from polysilicon doped by diffusion from SOD sources. Very shallow and high performance junctions were obtained using this new diffusion technique [13]. This technique has been applied successfully for shallow junction formation of poly-emitter in high-speed bipolar transistors and in elevated source/drain of submicron PMOSFET [14].

1.4.3. Other applications

Apart from the two above mentioned applications there are still many others. These include optical-, electronic-, abrasion-, barrier-, and protective coatings [15].

Multilayer metalisation of future ULSI requires high conductance metal, e.g. copper, and low k dielectric to increase the chips' speed. SOG is one of the candidate for porous low k dielectric and this approach seems to be very promising [16].

Some SOGs have been studied for nano-scale mask fabrication [17]. A 40 nm wide line could be obtained by writing with atomic force microscope tip. The etching rate difference between an SOG layer which is exposed to electron current and the normal

SOG could be as high as 20 times [18]. SOG has been used as positive resist together with focus ion beam technique to fabricate sub-100 nm feature lines [19].

SOD have been applied in micromachining technology such as a micro motor [20]. Fabrication of buried channel waveguides on silicon substrates using spin-on-glass is also an encouraging technique [21].

Thin films have been realised by sol-gel technology for numerous electronic and magnetic applications. Indium Tin Oxide (ITO) has been developed for electrically conducting transparent films or displays and has been widely used for solar cell fabrication. Optoelectronic layer like e.g. PZT (Lead-Zirconate-Titanate) have been prepared. Iron oxide layers have been used for magnetic films. Vanadium layers have been used for anti static films. Barium titanate have been prepared in thin layers for multilayer acitors. Some oxide films (Al_2O_3 or Ta_2O_3) have been used as sensitive membranes in ISFET (Ion Sensitive Field Effect Transistor). Ferroelectric layers like e.g. PZT have been also applied for memory devices.

When forming MOS (Metal Oxide Semiconductor) devices in GaAs, the growth of a native oxide (oxide of gallium and arsenic) at the semiconductor-oxide interface should be avoided. One method for preventing such growth is depositing a non-native oxide such as titania on the GaAs surface. The reason for choosing the sol-gel approach is that these coatings are applied at room temperature, meaning solid-state transport is negligible. Adhesion of the coating to the substrate is chemical rather than physical. The coating thickness is typically 50-200 nm and fairly uniform.

For optical coating, sol-gel technology has been applied for medical imaging [22]. Several optical materials, e.g. BaFCl:Eu and ZnS:Sm, have been realised and investigated. There are still many other applications of SOG or sol-gel technology in many aspects of industry and daily life.

References

[1] Chieng Chiang et al., "Understanding of Spin-On Glass properties from their molecular structure", VMIC Conference 1989 IEEE, p. 397

[2] S. Chen et al., VMIC 1988, p. 306

[3] D. Yen et al., VMIC 1988, p. 85

[4] Stephen W. Kirtley and George S. Wooster, “Polymeric boron nitrogen dopant”, US patent No. 4578283, March 1986

[5] K. Gupta et al.,"Spin-On Glass for dielectric planarization", Allied-Signal Inc.

[6] The National Technology Roadmap for Semiconductors, Semiconductor Industry Association 1997

[7] Akira Usami et al., "Shallow-junction formation on Silicon by rapid thermal diffusion of impurities from a spin-on source", IEEE Trans. Elect. Dev. , vol. 39 (1), Jan. 1992, p. 105

[8] L. Ventura et al., “Shallow-junction formation by rapid thermal diffusion into silicon from doped spin-on glass films”, Materials Research Society symposia proceedings, Vol. 342 , 1994, p.345

[9] D. Ruby et al., "High-efficiency silicon solar cells by RTP", Appl. Phys. Lett. 65 (16), Oct. 1994, p. 2087

[10] W. Zagozdzon-Wosik et al., "Silicon doping from phosphorus spin-on dopant source in proximity rapid thermal diffusion", J. Appl. Phys. 75 (1), Jan. 1994, p. 337

[11] W. Zagozdzon-Wosik et al., "Fabrication of submicron junctions - proximity rapid thermal diffusion of phosphorus, boron and arsenic", IEEE trans. Elec. Dev. 41 (12), Dec. 1994, p. 2281

[12] N. N. Toan, N. D. Chien, V. T. Son, N. T. Giang, J. Holleman, P.H. Woerlee, “Spin-On Glass (SOG) materials and applications”, 2nd International Workshop on Materials Science Proceeding, IWOMS’95, Hanoi 1995, p. 191

[13] N. N. Toan, N.D. Chien, J. Holleman, P.H. Woerlee, “Formation of shallow junctions by dopant diffusion from Spin-On Glass”, 1st annual workshop on Semiconductor Advanced for Future Electronics (SAFE) proceeding, The Netherlands, Nov. 1998.

[14] N.N. Toan et al., “Deep submicron Poly-spacer MOSFET for future ULSI”, to be published

[15] John L. Vossen, Werner Kern, "Thin film processes II", Academic Press, cop. 1991

[16] Nobuo Aoi, “Novel Porous Films Having Low Dielectric Constants Synthesized by Liquid Phase Silylation of Spin-On Glass Sol for Intermetal Dielectrics”, Japanese journal of applied physics, Vol. 36 (Issue 3), 1997, p. 1355

[17] M.Hoga et al., “practical method of phase-shifting mask fabrication”, SPIE Vol.2254 Photomask & X-ray mask technology, 1994, p. 14

[18] Park S. W. et al., “Nanometer scale lithography at high scanning speeds with the atomic force microscope using spin on glass”, Applied Physics Letters (16), 1995, p. 2415

[19] Kohei Suzuki, Motoji Yamashita, Nobuyuki Kawakami, Akimitsu Nakaue, “Focused Ion Beam Lithography Using Ladder Silicone Spin-on Glass as a Positive Resist”, Japanese journal of applied physics, Vol. 35 (Issue 12), 1996, p. 6517

[20] Conedera V. et al., “First steps towards the fabrication of electrostatic micromotors using SOG”, Sensors and actuators. A, Physical (1-3), 1995, p. 82

[21] A.S. Holmes et al., " Fabrication of buried channel waveguides on silicon substrates using spin-on-glass", Applied optics, 32(25), 1993, p. 4916

[22] T.K. Anh, L.Q. Minh, L.C. Anh, N.N. Toan, W. Strek, C. Barthou, “The role of sol-gel technology in the study of medical imaging materials”, Exploring New Tracks in Imaging ICPS'98 Proceeding, Antwerp, Belgium, 1998, p. 224

2

SOG Fabrication & Characterization

2.1. Introduction

Understanding SOG materials and the sol-gel fabrication technology is very important for their applications. SOG materials are commercially available from many manufactures. However, there is often lack of material information. Furthermore, development of the SOG technology is necessary for solar-cell and sensor fabrication in Vietnam. The technology of SOG fabrication could also be useful for other applications as presented at the end of chapter 1.

In this chapter, the technology of SOG fabrication is investigated using a Design Of Experiment (DOE) method. Effects and interactions of ingredients in the sol-gel reactions were studied. SOG and phosphorus doped Spin-On Dopant (SOD) materials were fabricated. The SOG and SOD fabrication and the DOE study will be described at the beginning of this chapter.

The second part of this chapter will present results of X-ray Photoelectron Spectroscopy (XPS), Auger Electron Spectroscopy (AES) and X-ray Fluorescence Spectroscopy (XFS) measurements of SOD layers. These measurements will give insight into the chemical bonding in the SOD layer during processing. These data are also useful to understand the processes which occur during the baking and diffusion process of SOD layers.

2.2. Spin-on glass fabrication

2.2.1. Alkoxide chemistry of making silicate-SOG

The SOG materials were prepared using the alkoxide method. The technology is straightforward. The starting materials are silicon alkoxide (tetraethoxysilane - TEOS or $Si(OC_2H_5)_4$), water (H_2O) , isopropanol ($(CH_5)_2CH(OH)$) and hydrochloric acid (HCl). The main reactions are :

Hydrolyzation: $\equiv Si\text{-}O\text{-}C_2H_5 + H_2O \Rightarrow \equiv Si\text{-}OH + C_2H_5OH$

Polymerisation: $\equiv Si\text{-}O\text{-}C_2H_5 + \equiv Si\text{-}OH \Rightarrow \equiv Si\text{-}O\text{-}Si\equiv + C_2H_5OH$

Because TEOS and water are immiscible, the reactions only begin when the solvent isopropanol is added. The reactions took place at 82^oC, the boiling temperature of isopropanol. The reaction rate can be controlled by adding an acid as catalyst; hydrochloric acid was used in this case.

One of the important parameters in this sol-gel reaction is the molar ratio (r) of H_2O:Si, i.e. the ratio between the amount of water and the amount of TEOS. Because water is produced as a by-product of the condensation reaction, an r value of 2 is sufficient for

complete hydrolysis and condensation to yield anhydrous silica as shown by the net reaction [1]:

$$n\ Si(OC_2H_5)_4 + 2n\ H_2O \rightarrow n\ SiO_2 + 4n\ C_2H_5OH$$

However, even with excess water (r >> 2), the reaction does not go to completion. Instead, a spectrum of intermediate species are generated. The intermediates, which remain soluble in the alcohol-water medium, are silanols, ethoxylanols, and polysiloxanes. Depending on this molar ratio, different kinds of product, e.g. bulk gel, film, fiber, and power, could be obtained. Beside the r value, the catalyst type and concentration, the solvent, temperature, and pressure also cause modifications in the structure and properties of the polysilicate products.

The final product (silicate SOG) will be a transparent sol of organic solvents (ethanol, acetone, isopropanol) containing Si-O network polymers. These above reactions will continue with time even at room temperature and may result in dense sol of larger network polymers. If this happens the SOG will not be able to be applied uniformly on a substrate because of its very high viscosity. In order to prevent these reactions the SOG should be stored at low temperature (3^oC is sufficient). Even though, the shelf life of silicate SOG is limited to six months.

The above technology can be used to make undoped spin-on glasses containing Si-O polymer network only. In order to make doped SOG or SOD the technology has to be modified by changing the HCl by another kind of acid containing the dopant we want. For example we used ortho-H_3PO_4 instead of HCl to prepare Phosphorus doped SOD.

The basic mechanism of this sol-gel reactions is well known. However, the recipes reported in the literature differ widely. Our goal was to make SOG which has similar characteristics as the commercial ones. The SOG layers have to be uniform and smooth after spinning it on a wafer. After baking the thickness should be more than 100 nm (with spinning speed of 3000 rpm for 20 seconds) and has SiO_2 - liked properties. This layer should also have less shrinkage after an annealing process at high temperature, i.e. $\geq 400^oC$. Furthermore, SOG materials should have good stability, e.g. lifetime longer than 6 months. After few experiments we could make SOG but the layers were not very uniform. We decided to use the design of experiment technique to optimise the SOG quality. In the next section we will present a brief overview of design of experiment technique.

2.2.2. Design of experiment

During process development a recipe is typically derived from experiments designed to identify an optimum set of process steps or reaction ingredients. Experiments are used to characterise the relationship between the independent (or controllable) variables (factor) and dependent (response) variables of the process. From this knowledge, the values of the controllable variables which give the best results are derived. The effective utilisation of experiments to optimise a process is based upon maximising the amount of information obtained, while minimising the cost of experimentation.

A structured approach, based on statistical design and analysis of experiments, provides investigators with a powerful tool for maximising the significance of experiments. The approach to be outlined is of use if the process or system being investigated exhibits the following behaviour: a) the response functions (i.e. the functions of the process system variables that respond to the controlling variables) are smooth over the range of the experimental conditions; b) the slope of such functions may under-go substantial change when the values of their independent variables are changed; c) The variables have interactions or they are related; d) the measured data is noisy due to experimental errors.

The statistical approach to the design of experiments is also useful in cases when no definite theory exists about the relationship between independent variables and their responses, and an empirical approach must be followed.

Table 2-1: Designs of experiment

	Design Type	**Purpose**	**Number of factors**
1	Screening	Identify important variables; crude predictions of effects	6 or more
2	Complete & fractional factorials	Linear effects & interactions: used for interpolation	3 to 8
3	Response surface	Linear effects, interactions, & curvature; used for interpolation	2 to 6

Traditional experimental design has often been conducted with "change one variable at a time, and keep all the others constant", as the single unifying principle. In fact this approach is very inefficient and potentially incomplete in terms of the information obtained. On the other hand, a formal experimental plan, consisting of a sequence of structured experiments (based on statistical principles) is far more likely to produce an optimum result, with a minimum expenditure of effort and resources. Depending on the objective of the experiment one could carry out as many of the steps as shown in table 2-1.

Details of these experimental designs can be found in references [2, 3]. In factorial experiment, the use of factorial arrangements for factors in an experiment makes each run yield information concerning every factor. The full factorial experiments, a set of runs in which all possible combinations of (k) factors at the levels specified by the experiment are conducted, e.g. for multiple (m) full factorial experiments m^k runs are required. Full-factorial experiments, as opposed to one-at-a-time experiments, not only give a complete exploration of the experiment space, but they also offer the additional advantages of hidden replication, and information about interaction effects among factors.

The factor effect is the difference in the two measured responses when the two different levels of a factor are applied in an experiment. Main effects for each individual factor are defined to be the difference between two average responses of the two levels of a factor. On the other hand, two or more factors are said to interact if the effect of one is different at different levels of other factors.

However, in the full factorial experiment, as quite common, the number of factors may turn out to be large (e.g. 6-30), and it is still not well known which factors have the most influence on results. Therefore an effective strategy at this point would be to eliminate (screen out) those factors which cause little or no impact, and to perform such screening with the minimum amount of effort. The fractional-factorial experiment is designed for this purpose. This experiment is still based on the principles which underlie the full-factorial experiment, but use a smaller number of runs for a given number of factors.

The objective of the response surface approach is to look for optimum points of the responses. In the case of two factors, response data can be plotted in a contour plot as in figure 2-1.

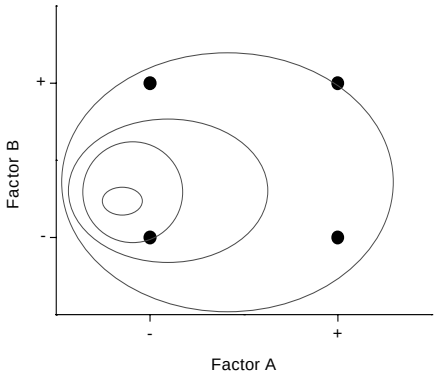


Figure 2-1: Example of a contour plot of a response vs. factor s A and B

If such response surfaces were constructed solely on the basis of 2-levels for each factor, only simple planar surfaces would result, and no information on local optimum points would be found. Therefore, at least three levels for each factor must be used for the response surface construction. In the next section, we will apply the DOE technique, starting with the screening method, for the SOG fabrication process.

2.2.3. DOE for SOG fabrication

The following experimental procedure was used: (all chemicals used have a VLSI grade purity)

- we kept the total volume of all ingredients constant at 37.5 ml
- partial volumes of TEOS, water and amount of HCl (mole) were chosen depending on a specific experiment
- the volume of isopropanol was calculated, by subtracting the volumes of TEOS and water from the a total volume
- the TEOS solution was then mixed with half of the total volume of isopropanol, in a two-way flask
- the HCl solution was made by diluting a 32 % HCl solution with water to obtain the desired amount of HCl.
- the acid solution was mixed with the other half of isopropanol
- the mixture of acid, water and isopropanol was carefully added, using a dropping funnel, to the mixture of isopropanol and TEOS while stirring
- the stirring was stopped, a reflux cooler was connected to the flask and a thermometer was lowered in the liquid
- the reaction mixture was then refluxed for 3 hours (including heating up) at 82°C (the boiling temperature of isopropanol)
- after refluxing, the reaction mixture was left standing until it was cooled down to room temperature
- the formed sol-gel mixture was diluted in acetone 1 : 1. It was ready for spinning or stored in a tightly closed container (to prevent further reaction with the moisture from atmosphere). The SOG container has to be stored at 3^o – 6^oC to prevent SOG from gelling.

For DOE, in our case there are four factors in the chemical reactions which are the amount of TEOS, water, HCl and the reaction time. The experiment time is long, we could carry out a maximum of only two experiments per day. We decided to chose the fraction factorial design for these 4 factors with a total of 16 runs (8 runs x 2 times).

In our experiments, the obtained SOGs were spin-coated on silicon wafers and were baked at 150^{o}C for 30 min. and then annealed at 400^{o}C for 30 minutes. The thickness,

refractive indexes of the layers after baking and annealing were measured by ellipsometer. The roughness was measured by topography profiler "Dektak". The results of the final thickness, the shrinkage, i.e. percentage value of the difference between the layer thickness before annealing (after baking) and after annealing divided by the thickness before annealing, and the average roughness were used as the responses for our design of experiment.

We used the statistical program MINITAB [4] to create and analyse the following half-fraction factorial experimental design in table 2-2. The trials were then randomised to perform in the order: 6, 1, 3, 8, 7, 4, 2, 5. This resulted in the scheme in table 2-3 (this scheme was performed 2 times).

Table 2-2: Fractional factor experimental design of SOG experiments

Trial No	TEOS	Water	Reaction time	HCl
1	-	-	-	-
2	+	-	-	+
3	-	+	-	+
4	+	+	-	-
5	-	-	+	+
6	+	-	+	-
7	-	+	+	-
8	+	+	+	+
Level (-)	10 ml	1 ml	1.5 hours	4.5 m mole
Level (+)	15 ml	4 ml	3.5 hours	7 m mole

Table 2-3: Factors inputs

Exp. No. (Trial No.)	TEOS (ml)	Water (ml)	Reaction time (hour)	HCl (m mole)
1 (6)	15	1	3.5	4.5
2 (1)	10	1	1.5	4.5
3 (3)	10	4	1.5	7.0
4 (8)	15	4	3.5	7.0
5 (7)	10	4	3.5	4.5
6 (4)	15	4	1.5	4.5
7 (2)	15	1	1.5	7.0
8 (5)	10	1	3.5	7.0

Table 2-4: Amounts of ingredients in moles and their percentages

Exp. No. (Trial No.)	TEOS		Water		Isopropanol		HCl		H_2O:Si (r) (molar ratio)
	m mole	(%)	m mole	(%)	m mole	(%)	m mole	(%)	-
1 (6)	67.65	16.6	55.33	13.6	280.79	68.8	4.5	1.1	0.8
2 (1)	45.10	10.0	55.33	12.3	346.09	76.7	4.5	1.0	1.2
3 (3)	45.10	8.0	221.32	39.0	293.85	51.8	7.0	1.2	4.9
4 (8)	67.65	12.6	221.32	41.2	241.61	44.9	7.0	1.3	3.3
5 (7)	45.10	8.0	221.32	39.2	293.85	52.0	4.5	0.8	4.9
6 (4)	67.65	12.6	221.32	41.4	241.61	45.2	4.5	0.8	3.3
7 (2)	67.65	16.5	55.33	13.5	280.79	68.4	7.0	1.7	0.8
8 (5)	45.10	9.9	55.33	12.2	346.09	76.3	7.0	1.5	1.2

2.2.4. Results and discussion

2.2.4.1. Spin-on glass preparation

The following table shows the results obtained from the sol-gel experiments as designed by the MINITAB program.

Table 2-5: Results of the sol-gel experiments

Exp. No. (Trial No.)	Thickness before annealing (nm)	Thickness after annealing (nm)	Shrinkage (nm)	Refractive index before annealing	Refractive index after annealing	Roughness (nm)
1 (6)	30	20	10	-	-	11.4
2 (1)	31	31	0	-	-	2.2
3 (3)	155	143	12	1.4289	1.4270	7.9
4 (8)	235	221	14	1.4704	1.4139	22.2
5 (7)	167	150	17	1.4273	1.4128	5.6
6 (4)	236	211	25	1.4366	1.4253	3.8
7 (2)	13	19	-6	-	-	77.4
8 (5)	23	19	4	-	-	92.6
1 (6)	61	70	-9	-	-	27.8
2 (1)	60	48	12	1.4190	1.4223	6.5
3 (3)	164	153	11	1.4323	1.4237	7.4
4 (8)	233	213	20	1.4465	1.4305	5.4
5 (7)	161	148	13	1.4331	1.4255	4.3
6 (4)	225	199	26	1.4325	1.4367	7.7
7 (2)	52	62	-10	-	-	54.9
8 (5)	30	42	-12	-	-	44.7

The thickness results in table 2-5 show that very thin SOG layers were formed after the reactions for all the experiments at low water level. This result indicates the hydrolysation-polymerisation reactions did not occur completely. At high water levels, i.e. with molar ratio r more than 3, thick SOG layers were obtained. These layers had refractive indexes in the range of 1.42 to 1.45. The refractive index decreased for every layer after the annealing process due to water and solvent evaporation and the layer became more porous.

In table 2-5 some refractive index values are not listed because they were unstable due to ellipsometry measurements on very thin SOG layers in the floating refractive index mode. Measurements with a fixed refractive index of 1.456 were carried out on these samples and the results (not presented here) indicated these layer were very thin, e.g. less than 10 nm.

The MINITAB program is also able to give some visual aids for the statistical analysis of the results, e.g. Pareto chart, main effect and effect interaction plots. Figure 2-2, 2-4, and 2-6 show Pareto charts of the effects caused by all the factors on the responses thickness, shrinkage, and the surface roughness. A Pareto chart is simply the bar chart listing all the factors and combinations of factors according to the effects, from the strongest effect to the weakest effect on a response. All terms with effects above the dotted line have a significant effect on the response. This line is calculated from the significant level alpha showing the maximum probability at which one could be wrong in the statement about the significance of a factor. In other words, the confidence level, $100(1-\alpha)$ % actually denotes the chance that the statement is correct. The value of alpha is chosen between 0 and 1. The smaller the value of alpha the more the confident the statement.

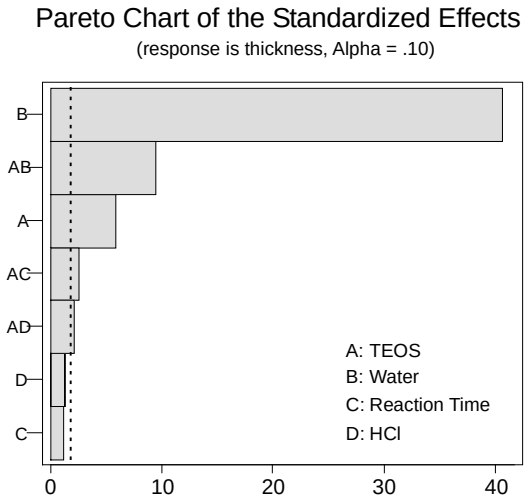


Figure 2-2: Pareto chart of the standardised effects on response thickness

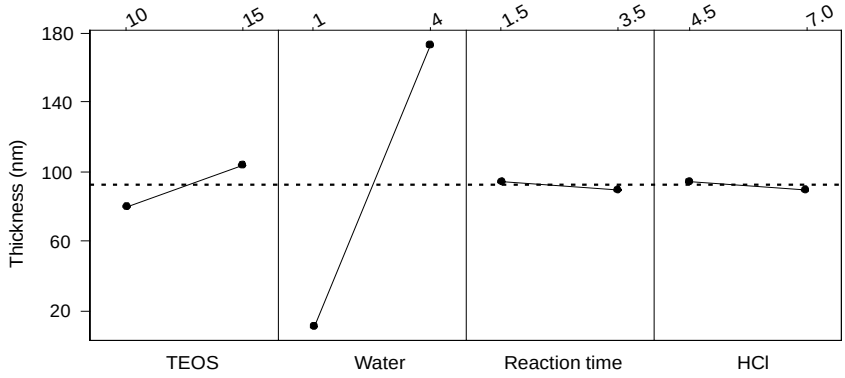


Figure 2-3: Main effects on thickness

The results for the layer thickness show that the amount of water (B) has the largest effect followed by the water/TEOS interaction (AB) and the amount of TEOS (A). This is most clearly seen in the Pareto chart in figure 2-2 with the significant level α of 0.10. Reaction time (C), the amount of acid (D) and the interactions AC, AD had small effects on the results. This result is understandable because water and TEOS are the two main ingredients in the hydrolysation and polymerisation reactions. The large influence of water on thickness is also illustrated in figure 2-3.

The small influence of the acid level on the thickness indicates that the acid only takes the role of catalyst to control the reaction speed. The effect of the reaction time, which does not have large influence on the thickness of the obtained SOG layer, suggests that the sol could be obtained after short time, i.e. less than 1.5 hour. However, the reaction time and the acid level also contribute to the quality of the SOG layer as seen in the following Pareto charts of the response shrinkage and the surface roughness.

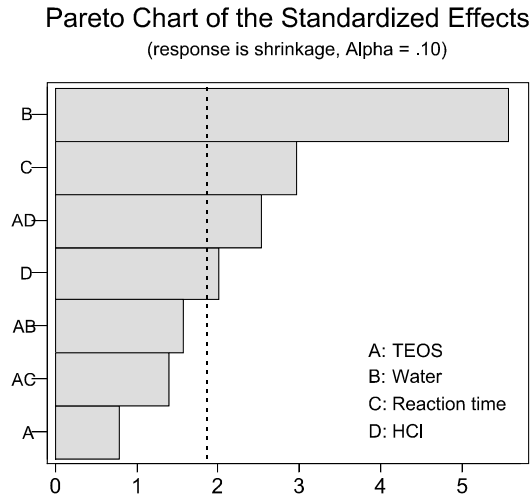


Figure 2-4: Pareto chart of the standardised effects on response shrinkage

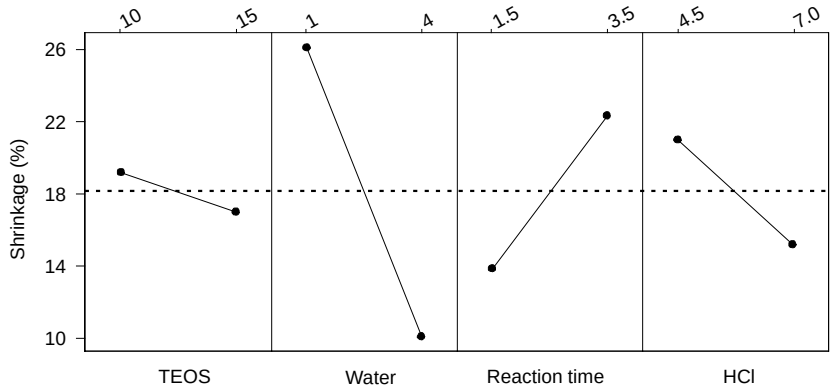


Figure 2-5: Main effects on shrinkage

For the shrinkage of SOG layers after annealing (figure 2-4), the effects are different. In this case, besides the amount of water, the reaction time, and the TEOS/HCl interaction also have important effects on the results. Figure 2-5 shows the shrinkage of SOG layers after annealing at 400^{o}C. During annealing the organic content in the SOG layer was evaporated and it resulted in porous layers with lower refractive indexes. According to figure 2-5, the more water the less shrinkage in SOG layers. At low water level, all experiments resulted in very thin SOG layers and the polymer structures were very weak. As a result, these layers changed more significantly than the layers made from high water level during annealing. The minus shrinkage of SOG layers in some cases at low water level was caused by the errors of ellipsometry of very thin layers in floating refractive index mode. Measurements under the fixed refractive index mode showed actually positive shrinkage of these layers.

In the case of the roughness (figure 2-6), the amount of water, the amount of acid and the water/time interaction have significant effects on the results.

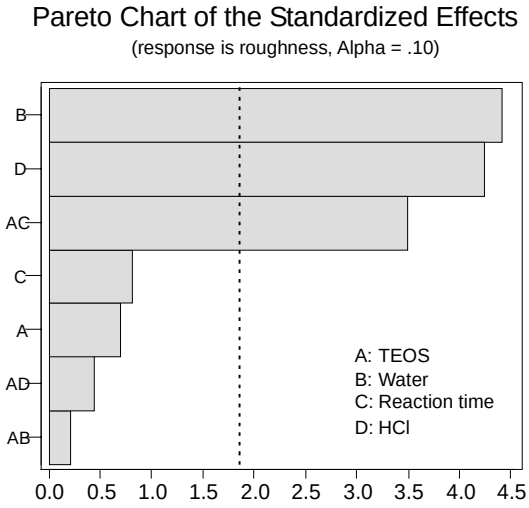


Figure 2-6: Pareto chart of the standardised effects on response surface roughness

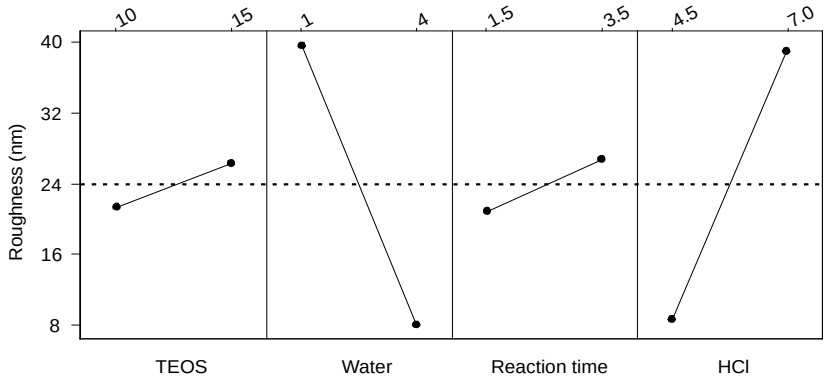


Figure 2-7: Main effects on roughness

As shown in table 2-5, the roughness of SOG layers made with low water level was higher than the thickness of the layers! This result indicated that there were particles on the wafers coated with these SOG. As explained for the shrinkage results, low-water-level experiments resulted in weakly bonding polymers containing particles. High-water-level experiments resulted in large polymer chains and less roughness as illustrated in figure 2-7. High acid concentration increased the reaction speed and resulted in more porous and rougher SOG layers.

Figure 2-8 shows the interaction plots of the factors on the obtained SOG layer. If two factors interact then the graphs showing the effects of two factors will have different slopes. Otherwise, if the graphs have equal slopes or no intersection, the interaction between two factors is small. As seen in figure 2-8 there are only two interactions on the response thickness; between the amount of TEOS and water, and between the reaction time and the acid level. These two interactions suggest that we could divide the four factors into two groups which do not affect each other. The group of TEOS and water is independent from the group of reaction time and acid. Therefore, in order to control the thickness one should change the parameters with a consideration of two interaction groups. For example, if one wants to add more TEOS into the sol-gel reactions the water volume must be reduced to obtain the similar thickness of a SOG layer while the acid and reaction time could be the same. The situation is the same for changing of the acid level and the reaction time.

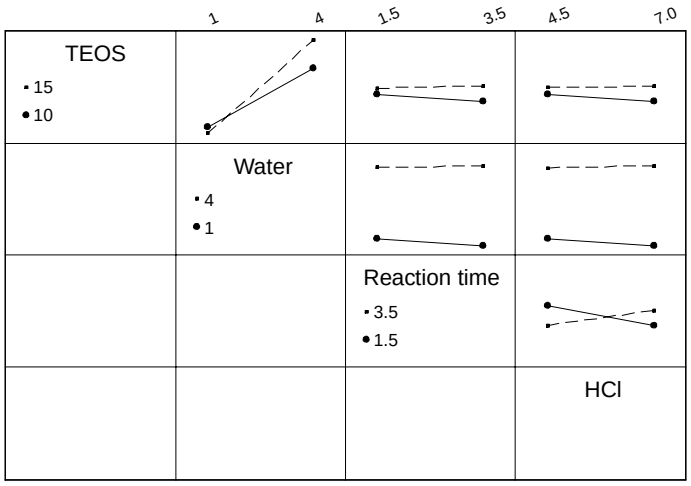


Figure 2-8: Effect interactions on thickness

In order to test the linearity of the effects we carried out 2 more experiments using the middle of the levels. The recipe was: 12.5 ml TEOS, 2.5 ml HCl 2.3 N, 22.5 ml isopropanol, 2.5 hours reaction time. The mole percentages of ingredients in this experiment were TEOS : Water : ISO : HCl = 11.4 : 28.0 : 59.5 : 1.2 and the H_2O/Si ratio (r) is 2.5. The same diluting, spinning and annealing procedures were used and compared with the commercial SOG (203AS from Allied Signal Inc.). The result in table 2-6 show that the responses are higher than the expected middle value as well as the maximum value from the main effect plots. This also implies that the responses were not as simple as linear lines as assumed by the models of factorial design and the results are only applied for the experiment points of the factors. For full exploration of the response ranges one should carry out the response surface design.

Table 2-6: Middle experiment results

Exp. Code	Before annealing (nm)	After annealing (nm)	Δ thickness (nm)	Δ thickness (%)	Roughness (nm)
030697	222	189	33	15	10.5
020797II	241	208	33	14	6.6
203AS	285	261	24	8.5	6.0

Our goal is to make a compatible SOG as the commercial one SOG-203AS (Allied Signal Inc.). Fortunately, from those experiments we already achieved some recipes which could create SOG which results in a relatively thick layer (>100 nm) and which is very smooth and uniform.

For the preparation of an undoped spin-on-glass the following recipes can be used:

- 10 ml TEOS, 23.5 ml Isopropanol, 4 ml HCl 1.75 N and 1.5 hours reaction time (r = 4.9). This should result in 145 nm SOG layer after annealing at 400^{o}C in N_2.
- 10 ml TEOS, 23.5 ml Isopropanol, 4 ml HCl 1.125 N and 3.5 hours reaction time (r = 4.9). This should result in 150 nm SOG layer after annealing at 400^{o}C in N_2.
- 15 ml TEOS 18.5 ml Isopropanol, 4 ml HCl 1.125 N and 1.5 hours reaction time (r = 3.3). This should result in 210 nm SOG layer after annealing at 400^{o}C in N_2.
- 12.5 ml TEOS, 22.5 ml Isopropanol, 2.5 ml HCl 2.3 N and 2.5 hours reaction time (r = 2.5). This should result in 230 nm SOG layer after annealing at 400^{o}C in N_2.

The sol-gels should be diluted in acetone or ethanol (both 1 : 1) and stored at low temperature, i.e. 3^o - 6^oC.

2.2.4.2. Spin-on dopant preparation

In order to make Spin-On Dopant (SOD) containing n-type impurities, e.g. phosphorus in this case, we used ortho-H_3PO_4 (85 %, 44.5 N) instead of HCl as a catalysing acid. Base on the recipes 3 (3) in table 2-4 the following experiments were conducted:

Isopropanol: 23.5 ml, TEOS: 10 ml, water: 4 ml, *H_3PO_4 : varied from 1N to 12N*, and reaction time of 1 hour. The reaction temperature was at the boiling temperature of isopropanol as in the SOG experiments, i.e. 82^oC.

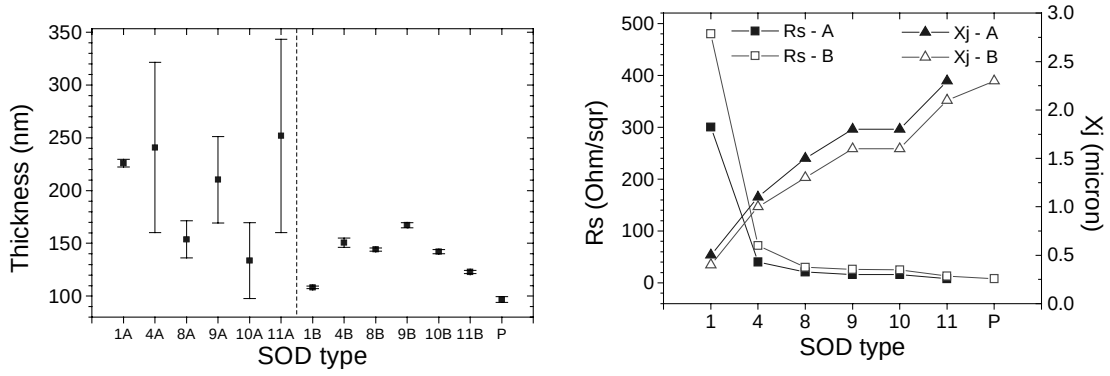
In these experiments, the amount of acid was varied in order to get SOD sols containing different phosphorus concentrations. The experiment with 12N H_3PO_4 resulted in a gelled SOD at the end of the reaction due to too high acid concentration. The reaction time was also reduced to 45 minutes in the cases of the acid concentration higher than 9N to prevent overreaction at low pH.

Each obtained SOD sol was divided in 2 parts after the sol-gel reaction finished. One half of SOD was mixed with acetone (50:50) for investigation of influence of acetone on the SOD properties. This part will be coded as nB, where n is the amount of acid concentration used in the sol-gel reaction, i.e. nB is from 1B to 11B in table 2-7. The other haft will be called nA for pure SOD.

To characterise the properties of SOD, the SOD sols were coated on different Si wafers (<100>, p-type, 10 Ωcm, 400 μm thick) by spinning. The coated wafers were baked at 150°C for 30 minutes and then annealed in N_2 at 1000°C for 1 hour. Thickness and refractive index of SOD layers were then measured by ellipsometry. After annealing, the SOD layers were removed with buffered HF and the resistivity of those wafers was measured by the 4-points probe method. Junction depths were measured by the ball-grooving technique. The next table shows an overview of the results obtained from the experiments with various amounts of acid. As a comparison, the commercial SOD-P8545 (Allied Signal Inc.) is also included in table 2-7 (P8545) and figure 2-9 (P).

Table 2-7: SOD experimental results

H_3PO_4 (N)	Thickness (nm)	Refractive index	Resistivity (Ω / sq)	Junction depth (μm)	Life time (stored at 3^oC)
1A	225 ± 03	1.480 ± 0.004	300	0.5	70 days
1B	108 ± 01	1.457 ± 0.000	480	0.4	> 6 months
4A	240 ± 80	1.963 ± 1.009	40	1.1	55 days
4B	150 ± 04	1.470 ± 0.002	72	1.0	> 6 months
8A	153 ± 17	1.602 ± 0.006	21	1.5	55 days
8B	144 ± 01	1.472 ± 0.001	30	1.3	> 6 months
9A	210 ± 40	2.193 ± 0.153	16	1.8	42 days
9B	167 ± 02	1.472 ± 0.001	26	1.6	> 6 months
10A	133 ±35	1.996 ± 0.540	16	1.8	35 days
10B	142 ± 02	1.471 ± 0.001	25	1.6	> 6 months
11A	251 ± 91	3.139 ± 2.380	8	2.3	30 days
11B	122 ± 01	1.476 ± 0.001	13	2.1	> 6 months
P8545	96 ± 02	1.451 ± 0.024	8	2.3	> 6 months



a) *b)*

Figure 2-9: a) Thickness of SOD layers and b) R_S and X_J of Si doped by SOD (1000 ^oC, 1h, N_2)

As seen in the above table and figures, the thickness of diluted SOD layers was in the range of 100 to 200 nm which depends on the acid concentration. The SOD was thickest for the case of H_3PO_4-9N which could be the optimum pH degree for the sol-gel reaction. The thickness of the pure SOD layers (A-series) was thicker than the SOD diluted with acetone (B-series).

The uniformity properties of SOD layers diluted with acetone were also much better than the pure SOD ones. The diluted SOD layers were very uniform with standard deviation (1σ) of less than 1 % for 25 point measurements over each wafer. The refractive index of these SOD layers was about 1.47 and the layers were very uniform, even better than the commercial one. On the other hand, the pure SOD layers had very scattered refractive index and thickness over the wafers. These layers looked also very hazy, especially for the ones with high acid concentration.

The hydrolyzation-polymerisation reactions of SOD was still happening at low temperature with a low reaction rate. However, the addition of acetone into the SOD sols significantly improved the properties of the layers. An addition of acetone or any solvent, e.g. ethanol, could decrease and even reverse this reaction [1]. As a result, the SOD could not form large chains of ≡Si-O- polymers and the viscosity of the diluted SOD was less than the pure ones. Therefore, layers formed by diluted SOD were thinner and had better uniformity when coated on the Si wafers. This effect of acetone or ethanol addition was also observed when we did experiments with SOG, the data for pure SOG layers was not collected.

The lifetime of SOD sols (table 2-7) also supports this explanation. The higher the acid concentration the shorter the lifetime of the SOD sol. The polymerisation reaction was still happening and gels were formed after a certain time even at 3^oC. A higher acid concentration resulted in a higher reaction rate and a shorter lifetime. Dilution of SOD sols with acetone reduced the reaction and improved the lifetime of all SOD sols.

Figure 2-9b illustrates the sheet resistance and junction depth of Si samples doped by the different kinds of SOD. These values are also listed in table 2-7. The results show that a lower sheet resistance and deeper junctions were obtained for samples doped by pure SOD than diluted SOD. It was due to higher phosphorus concentrations in the pure SOD layers than diluted ones. A higher H_3PO_4 concentration in the sol-gel reaction resulted in a higher phosphorus concentration in the SOD layer and a deeper junction in the Si samples. Compared with the commercial P8545, the SOD-11B showed similar doping properties.

In summary, SOD doped with different phosphorus concentrations have been successfully realised using sol-gel technology. The dilution of SOD by acetone increased the lifetime of the sols and gave better layer uniformity.

2.2.5. Conclusions

SOG materials were prepared by the sol-gel method using the statistical experimental design tools to optimise the process. Fractional factorial design of experiment for SOG fabrication was used. The results showed that the water volume had the largest effect in the sol-gel reactions on the thickness, shrinkage and the surface roughness of the obtained SOG layers. Although the response surface design has not been carried out yet but some of the recipes could offer good SOG with similar dielectric properties as a commercial SOG.

SOD sols containing different phosphorus concentration were also prepared. Very good uniform SOD layers were obtained by dilution of SOD sol with acetone. Because acetone addition can slow down the polymerisation reaction rate at low temperature it reduced the viscosity of the SOD sol and increased the SOD lifetime. The obtained SOD sols were also applied for doping of Si. Primary results were very encouraging but more diffusion study should be carried out to characterise these SOD properties.

2.3. Spin-on dopant characterisation

In this section we will present some results of SOD analysis using X-ray photoelectron Spectroscopy (XPS), Auger Electron Spectroscopy (AES) and other methods. Those analysis showed the chemical components of SOD materials and their concentrations.

2.3.1. X-ray Photoelectron Spectroscopy (XPS)

XPS can give information about the elements and bonds existing at the surface of a sample. XPS measurements were performed on our SOD samples using the system KRATOS - XSAM 800 at the university of Twente. The samples were Si wafers coated with SOD P8545 and B150 from Allied Signal Inc.. These SOD materials were used for our diffusion experiments in this thesis.

2.3.1.1. XPS results of P8545

XPS results of samples coated with P8545 showed that these SOD layers contain P,O, Si, C, N, F. The measured bonding energies were the same as the energies of signals coming from P_2O_5, SiO_2, $p(CF_2=CF_2)$ and NH_4NO_3 in the case of sample after baking at 150^{o}C for 30 minutes. In other words, P-O bonds, Si-O bonds and C-F bonds were found in our SOD P8545 samples.

After annealing at 850^{o}C for 30 minutes in N_2 this sample still contains the same elements but with different concentrations as illustrated in figure 2-10 and table2-8. It is clear that P8545 is a polymer of Si-O network containing P_2O_5. After diffusion the phosphorus evaporated, the atomic concentration was reduced from 11.2 at. % to 5.3 at. %.

Table2-8: Atomic percentages of elements in P8545 samples

P8545	P (%)	O (%)	Si (%)	C (%)	N (%)	F (%)
150^{o}C	11.2	65.5	13.2	7.3	1.2	1.5
850^{o}C	5.3	63.5	21.8	8.9	0.6	0

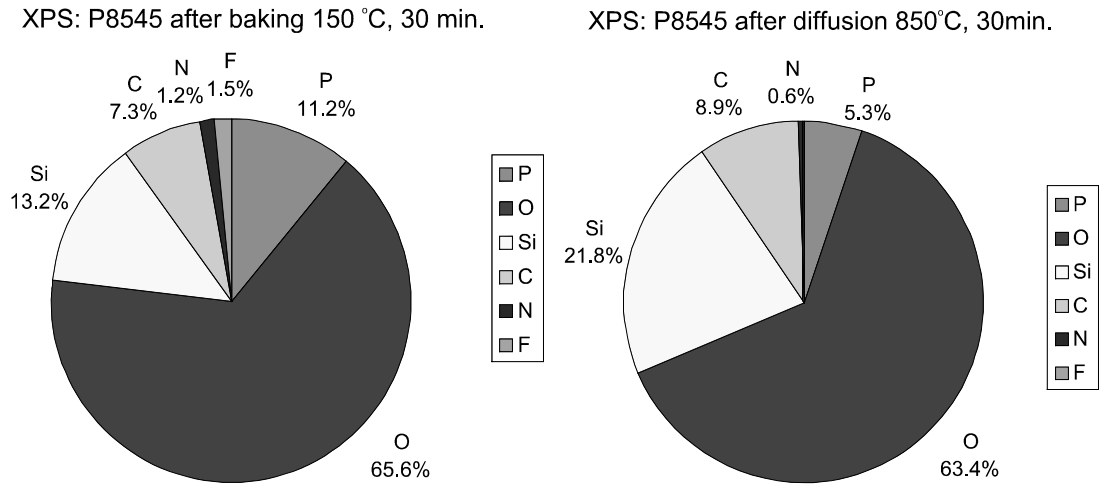


Figure 2-10: Concentration (at. %) of elements in the SOD-P8545 layer after a) baking at 150 °C for 30 min. and b) annealing at 850 °C for 30 min. in N2 ambient

2.3.1.2. XPS result of B150

Si wafers coated with B150 were also measured. XPS results show that these SOD samples contain B, N, O, Si, C and F. The sample which was baked only at 150°C for 30 minutes contained bonding like BN, p(CF_2=CF_2) and H_2O. However, after diffusion at 850°C for 30 minutes in N_2 this SOD converted into polymer containing B_2O_3, SiO_2 and little nitrogen left. Pie charts in figure 2-11 and table 2-9 show the percentage concentrations of the elements in the SOD after baking (150°C), after diffusion (850°C) and after etching in buffered HF for 3 minutes (BHF3').

Table 2-9: Atomic percentages of elements in B150 samples

B150	B (%)	N (%)	O (%)	Si (%)	C (%)	F (%)
150°C	12.3	12.5	13	0	60.8	1.4
850°C	4.3	0.3	62.7	24.4	8.3	0
BHF3'	2.79	0	61.82	31.27	3.94	0

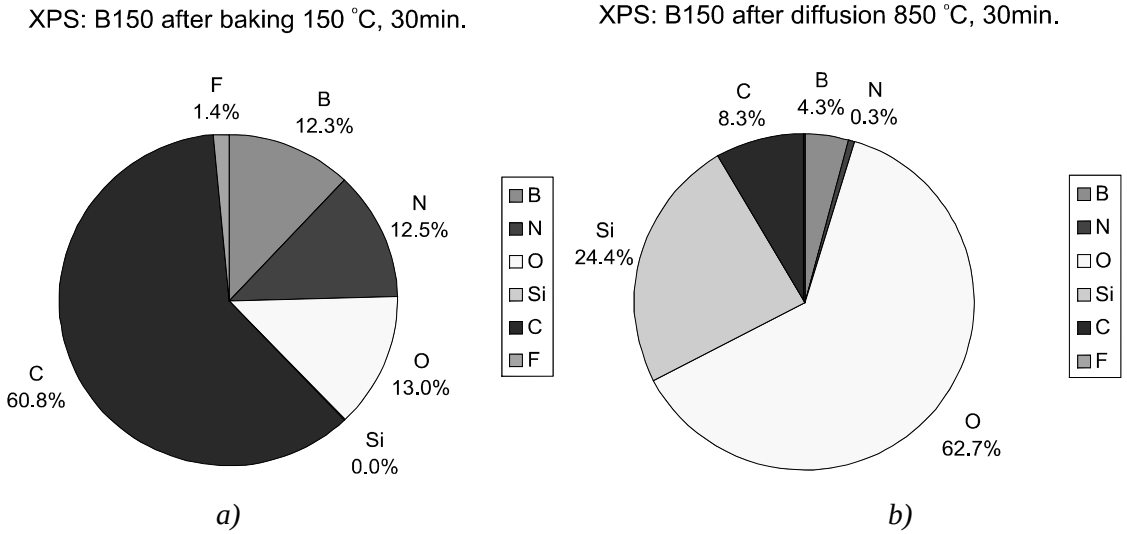


Figure 2-11: Concentration (at. %) of elements in the SOD-B150 layer after a) baking at 150 °C for 30 min. and b) annealing at 850 °C for 30 min. in N2 ambient

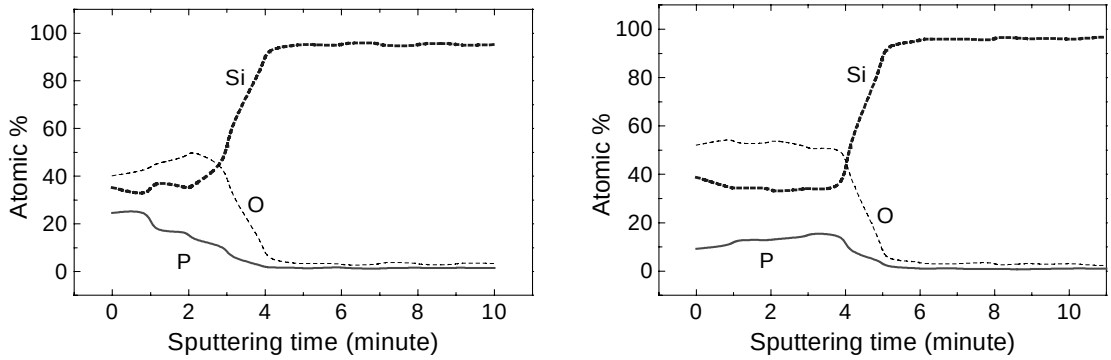
Compared with the XPS results of the P8545 layers, the chemical bonding was totally different in this case. In fact, this SOD B150 is synthesised from a class of boron-nitrogen compounds, known as borazole (or borazine) which is quite different from the alkoxide polymer like P8545 (see chapter 1). The above results suggest that during the diffusion process the B-N bonds were broken by the existing oxygen in the SOD layer and B_2O_3 was produced. Due to high temperature, boron diffused into the Si substrate and Si diffused out to the SOD layer. This explains why the Si content appeared in the SOD layer after diffusion at 850^{o}C.

After diffusion the SOD layer was removed in buffered HF for 3 minutes. However, there was still a layer of about 65 nm left on the Si wafer which could not be removed. This layer could cause difficulties during device processing and could deteriorate devices properties, e.g. high series resistance, as will be shown in the next chapter. Therefore, it is necessary to characterise this layer. The XPS result of this layer is in the last row of table 2-9 which was coded as " BHF3' ". Compared with the SOD layer after diffusion, this layer had lower boron, lower carbon concentration and higher Si concentration. This is due to Si diffusion out from Si substrate to form SiO_2 and some kinds of boron-rich silicon oxide which is very hard to be removed by buffered HF.

2.3.2. Auger Electron Spectroscopy (AES)

AES also can indicate the existence of different elements in a sample and their concentrations. In addition, profiles of element concentrations can be measured by using a sputtering gun. Our measurements were carried out using an AES system PERKIN - ELMER PHI 600 at the university of Twente. The samples were prepared as Si wafers coated with SOD, baked at 150^{o}C for 30 minutes and annealed at 850^{o}C for 30 minutes in nitrogen ambient.

2.3.2.1. AES results of P8545



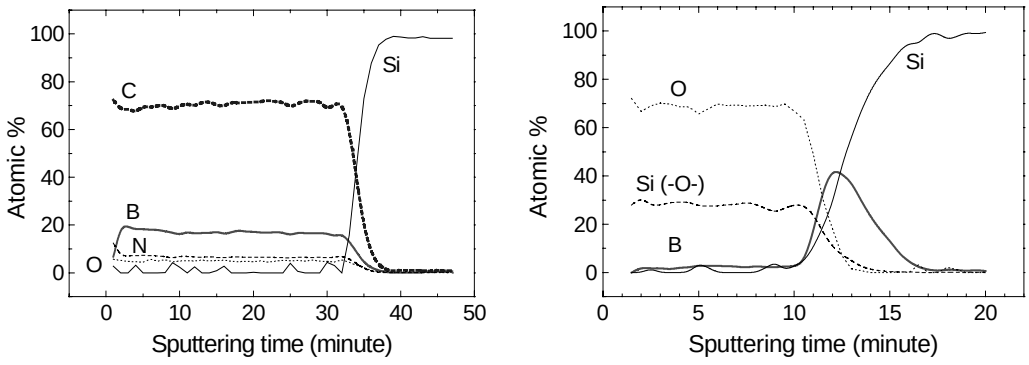
a) *b)*

Figure 2-12 : Concentration (at. %) profiles of elements in the SOD-P8545 layer on Si after a) baking at 150 ^{o}C for 30 min. and b) annealing at 850 ^{o}C for 30 min. in N2 ambient

Due to lower sensitivity of the AES technique, not all of the elements which were measured by XPS could be detected. As seen in figure 2-12, only P, O and Si profiles were measured in the P8545 samples. All the profiles were redistributed after annealing at 850^{o}C. The profiles were not very uniform throughout the P8545 layers after baking. This suggests that there were different polymer layers formed on the Si substrate after baking process. After the annealing process, the P, O and Si diffused and the profiles were flattened in the SOD layer. However, the phosphorus profile showed a depletion at

the surface due to out diffusion. The difference in the sputtering times in these profiles is due to the fact that SOD layer was densified after the diffusion process.

2.3.2.2. AES result of B150



a) *b)*

Figure 2-13: Concentration percentage profiles of elements in the SOD-B150 layer on Si after a) baking at 150 ^{o}C for 30 min. and b) annealing at 850 ^{o}C for 30 min. in N2 ambient

The above figures show AES atomic percentage profiles of elements in the B150 samples after baking and diffusion. Similar to the XPS results, B, C, N, O and Si were detected in B150 layer after baking. Differently from the baking sample of P8545, the profiles in this sample were very flat over the SOD layer. The C concentration in the B150 layer was very high; about 65 atomic %. There was almost no Si in B150 sample after baking. The Si element was only detected in the substrate region.

After diffusion, B, O and Si were detected in the SOD layer. The N and C could not be measured due to their low concentrations which were below the sensitivity of AES technique. During this measurement two peaks were detected in the Si spectra. One Si peak appeared in the SOD layer. This signal came from the Si atoms bonded with O or B which were found in the XPS measurement. The other peak was from the Si substrate. The O atomic percentage became very high in the SOD layer due to the low C concentration. Boron accumulated strongly at the SOD/Si interface and also diffused in to the Si substrate. The accumulation of boron was due to the lower diffusivity of boron in the Si than in the SOD layer. This effect will be explained more in the next chapter.

The AES results suggest that during diffusion there was decomposition of the B150 layer and most of the organic groups containing C and N were evaporated. Boron diffused into the Si substrate via a redox reaction of Si to form a SiO_2 like layer. This layer contained very high boron concentration, especially at the SOD/Si interface which was hardly removed by normal buffered HF etching.

2.3.3. Other measurements

Table 2-10:Concentration of the dopants in P8545 and B150 layers measured by different methods

No.	Sample description	Phosphorus (ions/cm^3)	Boron (ions/cm^3)	Element percentage (at. %)	Technique	Notes
342	P8545, 150^{o}C, 30 min. (air) 116 nm SOD/ 898 nm SiO2/Si	6.79x10^{21}		P=15.74, Si=25.12, O=49.14, C=7.30, N=1.20, F=1.50	XFS	Average P concentration of the P8545 layer, C % N % and F % are fixed & taken from XPS
344	P8545, 150^{o}C, 30 min. (air) 116 nm SOD/Si	7.15x10^{21}		P=16.24, Si=29.49, O=44.27, C=7.30, N=1.20, F=1.50	XFS	Average P concentration of the P8545 layer, C % N % and F % are fixed & taken from XPS
343	P8545, 800^{o}C, 30 min. (N2) 86 nm SOD/ 899 nm SiO2/ Si	9.47x10^{21}		P=23.43, Si=11.78, O=55.29, C=8.90, N=0.6, F=0	XFS	Average P concentration of the P8545 layer, C % N % and F % are fixed & taken from XPS
346 & 347	P8545, 150^{o}C, 30 min. (air) 157 nm SOD/Si	8.9x10^{21}		P=11.2, Si=13.2, O=65.5, C=7.30, N=1.20, F=1.50	XPS, weighing and ellipsometry	Surface P concentration of the P8545 layer
346 & 347	P8545, 800^{o}C, 30 min. (N2) 96 nm SOD/ Si	3.5x10^{21}		P=5.3, Si=21.8, O=63.5, C=8.9, N=0.6, F=0	XPS, weighing and ellipsometry	Surface P concentration of the P8545 layer
357 & 358	B150, 150^{o}C, 30 min. (air) 194 nm SOD/Si		7.4x10^{21}	B=12.3, Si=0, O=13, C=60.8, N=12.5, F=1.4	XPS, weighing and ellipsometry	Surface B concentration of the B150 layer
357 & 358	B150, 800^{o}C, 30 min. (N2) 81 nm SOD/ Si		2.4x10^{21}	B=4.3, Si=24.4, O=62.7, C=8.3, N=0.3, F=0	XPS, weighing and ellipsometry	Surface B concentration of the B150 layer

Other methods including X-ray Fluorescence Spectroscopy (XFS) and weighing were also carried out to characterize the SOD materials. From those measurements as well as from XPS and AES we calculated the concentration of the dopants in the SOD layers as shown in table 2-10. Within the errors of those measurements which may be within 10 % we can conclude that:

In the SOD-P8545 layer

- after baking at 150^{o}C for 30 minutes the average concentration of the phosphorus is about 7.6x10^{21} ions/cm^3.
- after diffusion at 850^{o}C for 30 minutes the average concentration of the phosphorus is about 6.5x10^{21} ions/cm^3.

In the SOD-B150 layer

- after baking at 150^{o}C for 30 minutes the surface concentration of the boron is about 7.4x10^{21} ions/cm^3.
- after diffusion at 850^{o}C for 30 minutes the surface concentration of the boron is about 2.4x10^{21} ions/cm^3.

These concentrations will be used as input parameters for simulation of the diffusion process.

2.4. Conclusions

Spin-On Glass (SOG) and phosphorus doped Spin-On Dopant (SOD) were prepared successfully using sol-gel technology and design of experiment method. Good materials and layers with similar properties as commercial SOG have been obtained. Technologies of boron doped SOD preparation can be investigated in future.

In the sol-gel reactions of silicate SOG, the water volume had the largest effect on the thickness, shrinkage and surface roughness of obtained SOG layers.

Dilution of the obtained SOG sols at the end of the sol-gel reactions with a solvent, e.g. acetone or ethanol, can increase the lifetime of the sols and improve significantly the properties of the coating layers.

Elements in the SOD layers and their concentrations profiles have been explored by advanced analysing techniques. The evolution of the SOD layer properties during baking and annealing steps were also investigated.

References

[1] C. Jeffrey Brinker and George W. Scherer, “Sol-gel science : the physics and chemistry of sol-gel processing”, Academic Press Inc., 1990.

[2] S. Wolf, “Silicon processing for the VLSI Era”, Volume 1 - Process Technology, chapter: “Structured approach to design of experiments for process optimisation”, p.618.

[3] Robert L. Mason, “Statistical design and analysis of experiments with applications to engineering and science”, John Wiley & Sons, Inc., 1989, ISBN 0 - 471 - 85364 - X

[4] MINITAB release 11 for Windows, Minitab Inc., 1996

3

Shallow P-N Junction Formation Using SOD

3.1. Introduction

The p-n junction is a crucial part in every modern semiconductor device, e.g. diode, bipolar junction transistors or metal oxide field effect transistors. Ultra shallow junction formation is a very critical process step for deep sub-micron devices in future ULSI. Ion implantation suffers from transient-enhanced diffusion and reduced dopant activation. Diffusion from highly doped glasses offers an alternative way to form ultra shallow junctions. In this chapter we will present a simple technology to form shallow p-n junctions in silicon using spin-on dopant (SOD) sources.

A study of boron and phosphorus diffusion from SOD sources in mono-crystalline silicon will be presented first. In section 3.2, we investigated the diffusion of boron and phosphorus in mono Si using conventional furnace and rapid thermal process. Results of measured concentration profiles and simulation will be presented. Electrical properties of p-n junctions were studied using diode structures. Current-voltage characteristics of diodes will be shown in relation to their doping profiles.

For ULSI technology, the sheet resistance of the shallow junctions formed by SOD diffusion into mono Si can be too high and the process window is small. In order to overcome this problem, we studied the junction formation of MOSFET elevated source/drain structures and emitters of bipolar devices by diffusion from polycrystalline and amorphous silicon layers doped by diffusion from spin-on-dopants. In this diffusion technique, influence of micro-structure of the as-deposited layers, polysilicon or amorphous silicon, to the uniformity of the junctions in the underneath mono-Si substrates was investigated. Critical thermal budgets required for good shallow P^+N and N^+P junctions were traced using furnace annealing and rapid thermal annealing of SOD at different temperatures and times. This diffusion technique will be discussed in the section 3.3 of this chapter.

3.2. Diffusion of B and P into mono-silicon from SOD

Spin-on dopant diffusion source is a kind of polymer containing dopant impurity and is dissolved in organic solvents. Details about the chemistry of SOD materials were presented in chapter 2. SOD sol is first applied to Si wafer using spinning technique. After deposition, SOD becomes an oxide like layer containing impurities, e.g. boron or phosphorus. Organic solvents are evaporated during subsequent baking and diffusion at high temperature. At the diffusion temperature, the impurities are driven into the semiconductor substrate, silicon. This diffusion technique is similar to conventional diffusion using chemical vapour deposition (CVD) doped oxide. However, in the case of SOD there must be some chemical reactions at the SOD/Si interface due to the organic nature of SOD materials. Depending on the kind of SOD, this chemical reaction

could be oxidation like. Therefore, the nature of diffusion of impurity from a SOD layer is a non-equilibrium process.

In this section, we will present our study of boron and phosphorus diffusion from SOD sources in Si. Diffusion using conventional furnace and rapid thermal process will be discussed separately for each dopant. The following subsection will present experimental details.

3.2.1. Experiments

Samples used in our experiment were 8 - 10 Ω.cm, <100> oriented, p-type (boron doped) and n-type (phosphorus doped) silicon wafers. The SOD sources were phosphorus doped silicate type P8545 and boron doped B150 (Allied Signal Inc.). The wafers were cleaned for 10 minutes in fuming HNO_3 (100 %), 10 minutes in boiling HNO_3 (65 %), and then etched in HF 1 % for 1 minute. Cleaned wafers were then coated with SOD by spin-on technique with 3000 rpm for 20 seconds. SOD solutions have been taken out of the refrigerator for warming up to room temperature before spinning. It is necessary to have SOD at room temperature for good coating reproducibility.

For testing the effect of prebaking temperature, the wafers were prebaked for 30 minutes at different temperatures: 90°C, 120°C, 180°C, 400°C and 600°C for the experiments of SOD-P8545. For SOD-B150 type it is not necessary to pre-bake because there is no water in this kind of SOD. The toluene solvent in SOD-B150 is evaporated almost completely during the spin coating process. Any baking of B150 layer will result in decomposition and evaporation of this layer.

The baked wafers were annealed in a conventional furnace in nitrogen ambient at different temperatures for different duration. SOD layers were removed totally in HF:NH_4F (1:6). The sheet resistance of diffused wafers was measured by four-point probe method. The ball grooving method was used for measuring junction depths. Electrically active doping concentration profiles of some samples were measured by Spreading Resistance Probing (SRP) technique.

For electrical characterisation of p-n junctions, diodes were fabricated using SOD diffusion sources. The process flow is illustrated in figure 3-1. Si wafers were first cleaned and oxidised to grow 100 nm SiO_2 for diffusion masking. Different active windows were then defined using photo-lithography. Oxide layers were etched away from those active areas by buffered HF. After cleaning wafers were coated with SOD containing different type of impurities, i.e. SOD-P8545 (phosphorus doped) was used on p-Si wafer to form N^+P junction and SOD-B150 (boron doped) on n-Si to form P^+N junction. Wafers coated with P8545 were baked at 120°C for 30 minutes for SOD densification and solvent evaporation. No baking step is required for wafers coated with B150. Thereafter the wafers were annealed in a conventional furnace or a rapid thermal system for different temperatures and times to diffuse impurities from the SOD source into the Si substrate. For conventional furnace diffusion the thermal budgets chosen were 900°C, 1000°C and 1100°C for 30 minutes. Rapid thermal diffusion experiments were carried out at temperatures of 950°C, 1000°C and 1050°C for 20 seconds with 50°C/second ramping rate. The PEAK system SP35X long arc, AC, gas-discharge lamp was used for the RTP process. All the diffusion processes were conducted in nitrogen. After drive-in diffusion, there was an additional annealing step at 650°C for 30 minutes in oxygen ambient to oxidise the SOD layers for easier removal by wet etching. This oxidation had negligible diffusion due to the low temperature.

After diffusion, contact holes were defined using photolithography and the SOD layer was removed locally in buffered HF. The photoresist layers were removed and wafers were coated with 1 μm thick aluminium film by sputtering. The wafers annealed using RTP were coated with 70 nm TiW layer and then 1 μm aluminium. The 70 nm TiW layer was used as barrier layer to prevent aluminium spiking on shallow junctions created by RTP. The metal layers were patterned by lithography and wet etching in an etchant composed of H_3PO_4, HNO_3 and CH_3COOH at 55°C for about 1 minute [1]. TiW was etched in H_2O_2 (31 %) for 30 minutes. The photoresist was then removed in fuming HNO_3 for 10 minutes. Subsequently, the wafers were annealed at 400°C for 10 minutes in wet nitrogen for metal sintering and surface state anneal. Finally, the wafers were coated with a thin layer of primer to protect wafers from water vapour in air during measurement. This primer coating can reduce the surface leakage current of diodes. I-V characteristics of diodes were measured by a Precision Semiconductor Parameter Analyser HP 4156A.

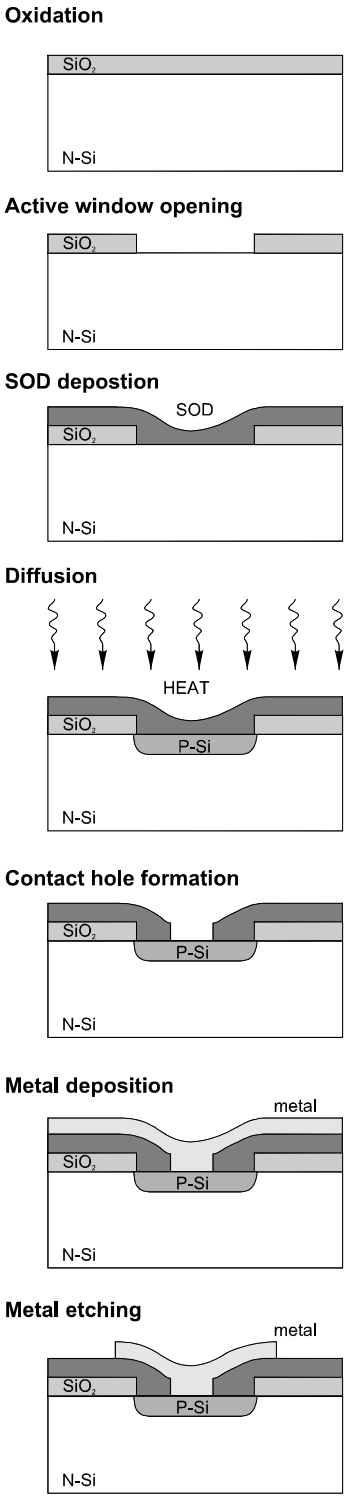


Figure 3-1: Diode process flow

In order to model the diffusion from SOD into silicon, we have used the SSUPREM-4 program embedded in the ATHENA software package. The diffusion model of SSUPREM-4 and parameters used for doping profile calibration are presented in appendix 1. There is no SOD diffusion model available in the SSUPREM-4 software. However, due to the similar diffusion behaviour of impurities from SOD and doped oxide into silicon, the diffusion of impurities from SOD has been simulated using the doped oxide model.

Only boron diffusion from SOD into Si was simulated because there were enough SRP profiles for these samples. The boron concentration in the doped oxide layer was chosen to be $7x10^{21}$ ions/cm^3, which is the concentration measured by XPS (see chapter 2). In order to fit the simulated profiles with the measured SRP profiles, parameters in the doped oxide diffusion model were varied simultaneously in an input simulation file. SRP profiles were extracted from simulated structures and compared with the measured SRP profiles.

This fitting procedure was carried out manually and it was actually very time consuming. Especially, when the number of tuning parameters is increased, the number of their combinations for the input file will be increased multiplicatively. In addition, parameters which were not important could be kept constant using default values. Therefore, we have chosen tuning parameters: the boron diffusivity in the SOD (oxide)

and the boron diffusivity in Si. For the diffusivity of boron in the SOD layer, the pre-factor (DIX.0) and the activation energy (DIX.E) were varied. The boron diffusivity in Si has two parts: intrinsic diffusivity of neutral vacancies and intrinsic diffusivity of single positively charged vacancies. The intrinsic diffusivity of neutral vacancy was not affecting significantly the simulated diffusion profiles and it was kept constant using the default value. The intrinsic diffusivity of singly charged vacancies was more important, especially for high boron concentration diffusion. This diffusivity also has the pre-factor (DIP.0) and its activation energy (DIP.E). As a result, 4 parameters, i.e. DIX.0, DIX.E of boron in oxide and DIP.0, DIP.E of boron in Si, were chosen for tuning. For more accurate fitting with more tuning parameters, an optimisation software is really needed to carry out the mass work of simulations.

3.2.2. Boron diffusion results

3.2.2.1. Furnace diffusion

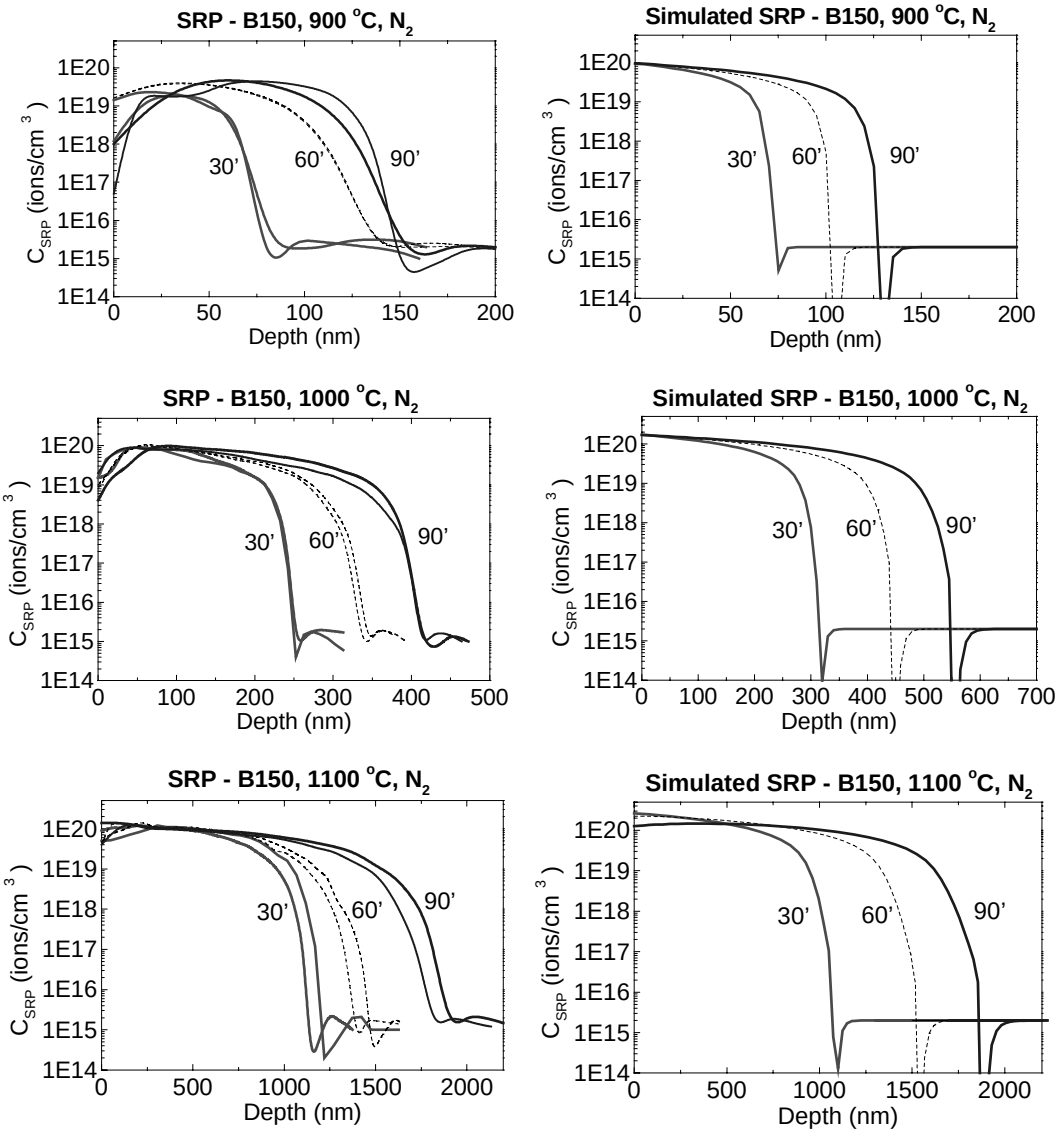


Figure 3-2: a) measured SRP and b) simulation SRP profiles of B150 samples diffused for 30', 60' and 90' at 900 °C, 1000°C and 1100 °C (from top to bottom)

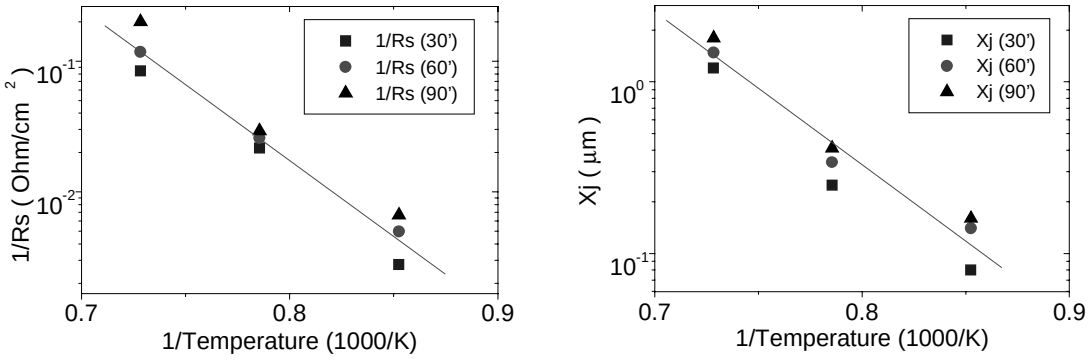
Figure 3-2 shows the measured SPR profiles and the simulation profiles of Si samples doped by B150 at 900°C, 1000°C and 1100°C for 30', 60' and 90' in N_2 ambient. Each

sample had two measured SRP profiles. The most shallow junction depth was 80 nm obtained for the sample diffused at 900°C for 30 minutes in N_2.

Table 3-1 shows the sheet resistance values (4pp), junction depths extracted from SRP profiles (SRP) of the samples and their simulated values (Sim.) at every experimental condition. Arrhenius plots of the sheet resistance values and junction depths are illustrated in figure 3-3.

Table 3-1: Sheet resistance and junction depth vs. diffusion temperature and time

	900°C				1000°C				1100°C			
t (minute)	Rs (Ω/cm²)		X_J(μm)		Rs (Ω/cm²)		X_J(μm)		Rs (Ω/cm²)		X_J(μm)	
	4pp	Sim.	SRP	Sim.	4pp	Sim.	SRP	Sim.	4pp	Sim.	SRP	Sim.
30	359.5	350.9	0.08	0.07	46.3	44.5	0.25	0.32	11.9	8.2	1.20	1.10
60	200.3	239.3	0.14	0.10	38.7	30.9	0.34	0.45	8.5	6.1	1.48	1.54
90	150.7	190.7	0.16	0.13	34.1	25.1	0.41	0.55	5.0	4.2	1.80	1.88



a) *b)*

Figure 3-3: Arrhenius plots of a) sheet resistance and b) junction depth of samples diffused by B150 for different annealing times.

The results of simulation are presented in figure 3-2b and in table 3-1. After changing many values of boron diffusivity in the SOD layer and in the Si substrate, the best fitted profiles were obtained for diffusion at 900°C and 1100°C. At 1000°C, the simulated profiles were always deeper than the measured profiles. In these simulation, the boron diffusivity in the SOD had the pre-factor DIX.0 of 1.00 cm^2/s and the activation energy DIX.E of 3.00 eV. Compared with the default diffusivity of boron in oxide, i.e. DIX.0 = 3.16x 10^{-4} cm^2/s and DIX.E = 3.53 eV, the boron diffusivity in SOD layer was much higher. For the boron diffusivity in Si, the intrinsic diffusivity of single positively charged vacancies had the pre-factor DIP.0 of 1.2 cm^2/s and an activation energy of 3.66 eV. Here, the boron diffusivity in Si was lower with higher activation energy than the default diffusivity, i.e. DIP.0 = 0.72 cm^2/s and DIP.E = 3.46 eV.

The different nature of the SOD layer compared to normal doped oxide could explain why the diffusivity of boron in SOD was much higher than in normal doped oxides and this value was lower in Si. It has been observed that during diffusion, the evaporation of B150 was significant and chemical reactions occured at the SOD/Si interface. This is especially at the beginning of the diffusion process. The ellipsometer measurement showed that thickness of B150 was about 400 nm with a refractive index of 1.52. After diffusion at 900°C for 30 minutes it remained about 100 nm thick layer with refractive

index of 1.48, i.e. only 25 % B150 left. The change in the refractive index also indicated that the properties of this layer has changed. After etching in BHF (about 5 minutes) there is still a layer of about 60 nm with a refractive index of 2.17 which could hardly be removed by BHF etching.

From the chemistry of B150 material and the XPS and AES measurements in chapter 2, the following process during deposition and diffusion of B150 layer on Si substrate is proposed:

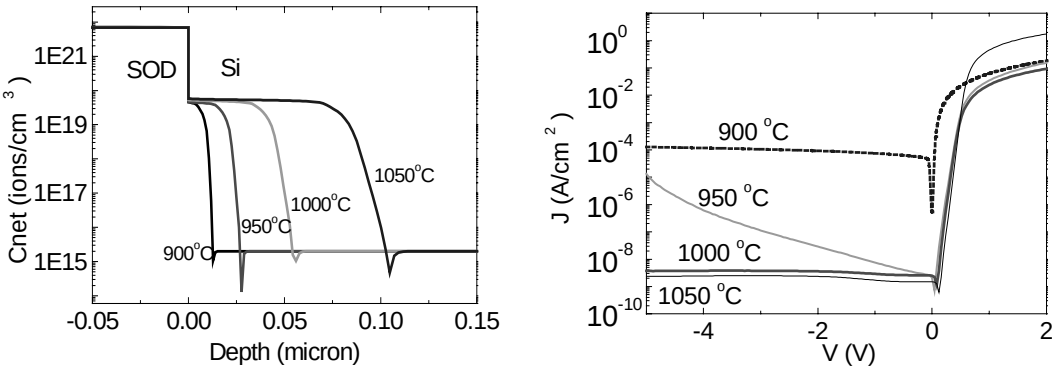
- during spin deposition at room temperature, most of the toluene in SOD solution was evaporated.
- after deposition by spinning the B150 layer became a borazine based polymer containing B, N, O, C and H.
- at diffusion temperatures, organic compounds in B150 layer were decomposed and the N and C evaporated. At the same time, boron diffused into the Si substrate and Si diffused out to the SOD layer. Because of the organic nature of the SOD layer, boron diffused in this SOD much faster than in a normal doped oxide layer. In addition, there must be a redox reaction of Si going on during this process evidenced by Si-O bonds found in XPS measurements. As a result, the grown oxide in turn acted as a barrier against the diffusion of boron and Si and it resulted in strong accumulation of boron ions at the SOD/Si interface as observed in AES profiles. Many other elements like C and H in the SOD layer, could also diffuse into the Si substrate and might change the diffusivity of boron in Si. The retrograde SRP profiles at the Si surface can be explained by the diffusion of H (and C) into Si which could inactivate part of the boron concentration.

B150 diffusion sources have been applied for diode realisation in conventional furnace. However, those diodes suffered from etching difficulty of the SOD layer after diffusion steps. Even when the wafers were oxidised at 650^{o}C in O_2 ambient for 30 minutes this layer could hardly be removed by buffered HF etching. As a result, the protecting field oxide surrounding the active areas was damaged because this layer was only about 100 nm thick. Consequently, those diodes had unacceptably high leakage current and are not discussed here. To solve this problem, a thicker field oxide or LOCOS structure is needed. Otherwise, there should be an additional oxidation step at higher temperature than 800^{o}C just after the diffusion in nitrogen. This oxidation step will convert the interface layer more into SiO_2 like and the BHF etching will be easier. However, it will result in much deeper junctions which is not of our interest for shallow junction formation. Rapid thermal diffusion of B150 layer can solve this problem and it also can form much shallower junctions. This diffusion technique will be presented in the next subsection.

3.2.2.2. Rapid thermal diffusion

Figure 3-4a shows the simulated net doping profiles of B150 rapid thermal diffusion (RTD) samples at different temperatures, 900^{o}C, 950^{o}C, 1000^{o}C and 1050^{o}C, for 20 seconds in N_2 ambient. Because we don't have any measured profiles of these samples, simulation of the diffusion process was carried out using diffusivity values of boron in the SOD layer found in the previous subsection, i.e. DIX.0 = 1.00 cm^2/s and DIX.E = 3.00 eV. All other parameters, including the boron diffusivity in Si, had the default values. The sheet resistance values were extracted and compared with the measured ones as listed in table 3-2. The diffusivity of boron in Si which was found in the previous section was also applied here but it resulted in very high sheet resistance values. Therefore, we used the default value of boron diffusivity in Si. Furthermore, this simulation was used only for estimation of the RTP profiles and their junction depths.

The sheet resistances were found about the same as the measured values. Therefore, the simulation profiles are expected to be similar to the experimental samples. According to these simulated profiles, shallow junctions of approximately 13 nm, 28 nm, 56 nm and 105 nm were obtained for samples diffused at RTP temperatures of 900°C, 950°C, 1000°C and 1050°C respectively.



a) *b)*

Figure 3-4: a) Simulated net doping profiles of B150 on Si RTD samples; b) IV characteristics of diodes fabricated using B150 RTD (diode area = 800x800 μm^2)

Measured electrical characteristics of diodes realised under these conditions are presented in figure 3-4b and in table 3-2. Differences between diode characteristics of these samples are clearly illustrated in figure 3-4b. Good diodes were obtained for samples 1000°C and 1050°C RTD with junction depths of approximately 50 nm and 100 nm. Shallower junctions resulted in worse electrical properties due to very high surface leakage current.

Details of diode parameters measured at room temperature are shown in table 3-2. For each sample diffused at a temperature shown in the first row, two diode parameters, leakage current density at reversed bias of 2.5 V (J_R) and ideality factor (n) [2], are listed for different diode areas.

Table 3-2: Parameters of diodes fabricated from B150 rapid thermal diffusion

T (°C)	900		950		1000		1050	
R_S (Ω/sqr)	>>		3900		500		250	
Simulated R_S (Ω/sqr)	5344.9		1703.1		648		290.2	
Simulated X_J (nm)	13		28		56		105	
Diode area (μm^2)	J_R (2.5V) (A/cm^2)	n	J_R (2.5V) (A/cm^2)	n	J_R (2.5V) (A/cm^2)	n	J_R (2.5V) (A/cm^2)	n
100x100	1.0E-04	3.53	4.5E-7	1.22	1.6E-8	1.01	1.1E-8	1.01
200x200	9.6E-05	6.00	2.6E-7	1.36	8.0E-9	1.01	6.1E-9	1.01
400x400	9.3E-5	3.43	9.6E-8	1.23	4.5E-9	1.02	3.3E-9	1.01
800x800	9.9E-5	4.32	5.5E-8	1.23	3.7E-9	1.02	2.4E-9	1.00

Both leakage current density and ideality factor are very important parameters presenting the quality of a diode. In general, when a diode is reversibly biased the leakage current often includes diffusion current and generation current. Diffusion

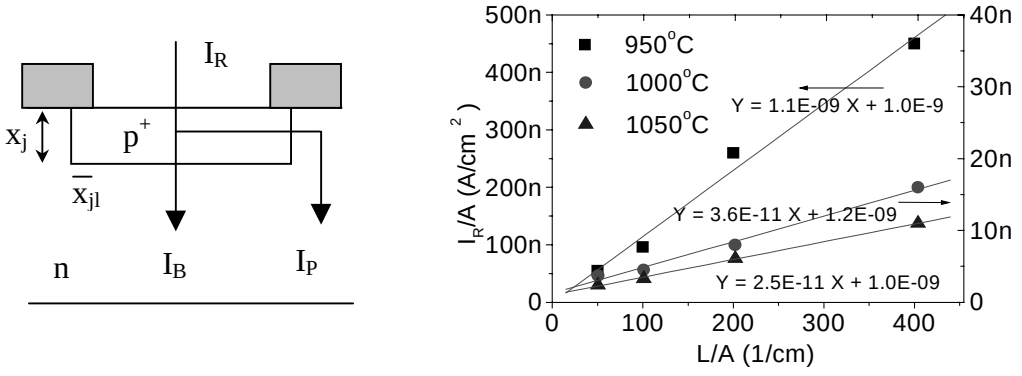
current density is very small and depends on the temperature and on the diode material, i.e. Si intrinsic properties. On the other hand, generation current shows the defect concentration presented in the diode volume, especially near the PN junction. The higher the defect concentration, the higher the leakage current. However, surface leakage current can also contribute significantly as in the cases of furnace diffusion samples in the first subsections. Defects also served as recombinations centres when the diode is forward biased and results in higher forward current in low bias region. Ideality factor n presents the deviation of forward current from the ideal diffusion current. The more this factor is close to 1 the more ideal the forward current and the less the defect density.

As seen in table 3-2, diodes fabricated at 1000°C and 1050°C show low leakage current densities of few nA/cm^2 and n values are nearly ideal, i.e. 1.01. The leakage current increased for diodes fabricated at lower temperatures and the ideality factor is far higher than 1. From this table we can see that the deeper the junction depth, the better the quality of its diode. The leakage current density is also decreased when the diode area is increased. This is because of the different contributions of the bulk and the peripheral between different areas. The current density shown in the table was calculated simply by taking the ratio of the total leakage current over the diode area. Actually, the total leakage current I_R can be presented with the bulk leakage current density I_B which contributed from the diode bulk area and peripheral leakage current density I_P as illustrated in figure 3-5a and in the following equation:

$$I_R = I_P \cdot L + I_B \cdot A \quad or \quad J_R = \frac{I_R}{A} = I_P \cdot \frac{L}{A} + I_B$$

Where L is the diode periphery and A is the diode area. If we plot the total leakage current density I_R/A as function of the ratio L/A we should obtain a linear relationship. The values of I_P and I_B can be derived by fitting to this line.

Indeed, figure 3-5b shows such a relationship, except for the diodes realised at 900°C for which the high leakage current could not be fitted because of other mechanisms.



a) *b)*

Figure 3-5: a) B150-RTP Diode leakage current density as function of L/A ratio; and b) leakage current model

From the fittings, the obtained bulk leakage current densities I_B are about 1 nA/cm^2 for all three samples. The peripheral leakage current densities I_P are decreased from 1.1 nA/cm to 36 pA/cm and 25 pA/cm when the diode junctions are going deeper at higher diffusion temperatures. This result can be explained by the fact that the lateral junction depth x_{jl} is often about 2/3 of the vertical junction depth x_j as illustrated in figure 3-5a. When the junction become too shallow then the shallower lateral junction will create a higher peripheral leakage current. At the same time, the SiO_2/Si interface properties will

also become important and it may increase the peripheral leakage current. The metal contact also becomes very important when the junction becomes too shallow. The metal/Si interface could increase the leakage current if the depletion layer was too close to this region in the case of very shallow junctions. On the other hand, a deeper junction resulted in lower leakage current and lower series resistance or higher forward current as shown in figure 3-5b.

In summary, high performance shallow junction have been obtained using RTD of B150 diffusion source into Si. All the junctions which were deeper than 20 nm had bulk leakage current densities of about 1 nA/cm^2 which is compatible with state-of-the-art diodes. The peripheral leakage current density increased as the junction is shallower due to higher leakage current at the SiO_2/Si interface. The simulated doping profiles should be checked by SIMS or SRP measurements.

3.2.3. Phosphorus diffusion results [3]

3.2.3.1. Furnace diffusion

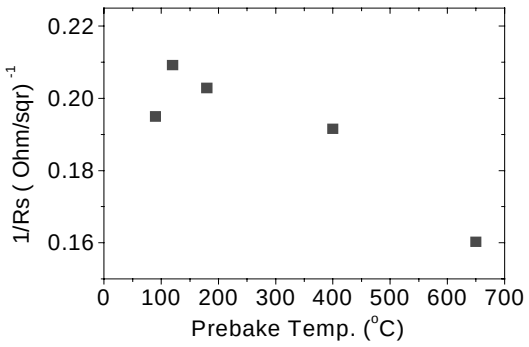


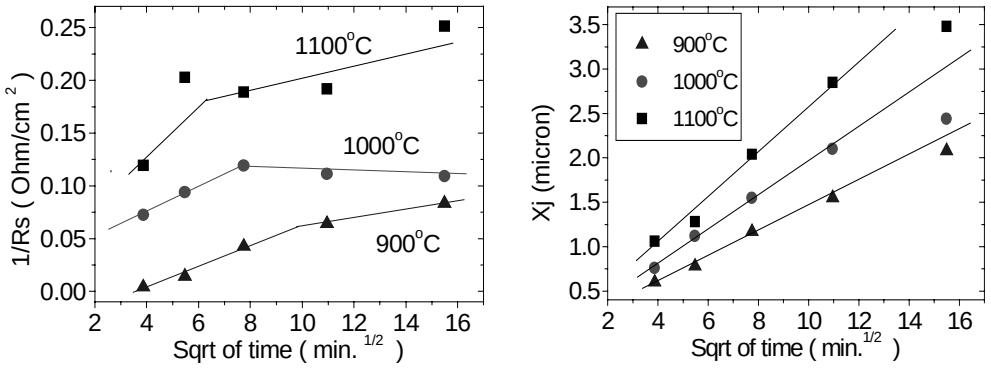
Figure 3-6: Dependence of sheet resistance after diffusion on pre-bake temperature

The influence of baking temperature on SOD diffusion source in presented in figure 3-6. The diffusion processes were carried out at temperature of 1100^{o}C for 30 minutes. Differently from the SOD B150, the SOD P8545 is a silicate based compound. During the low-temperature baking, polymer cross linking take place and a porous SOD structure was formed due to the solvent evaporation in the SOD layer. The experimental results showed that the most effective baking temperature was 120^{o}C, at which the sheet resistance was lowest. This is due to the changes in the SOD structure during the baking process at high temperatures. The higher baking temperature gives rise to the more dense SOD layer, and less efficient dopant release. However, if the baking temperature is less than 120^{o}C only a small amount of water and solvents are evaporated and the SOD structure is not an optimum porous source for dopant release.

Results of sheet resistance and junction depths of Si samples diffused from SOD P84545 at various temperatures and times are listed in table 3-3. In figure 3-7, the diffusion of phosphorus from SOD source is illustrated. The lines in these figures were linearly fitted with sheet resistances and junction depths vs. squared root of time. As seen in these figures, Fix's laws can not fit well to our results. The reason is that the phosphorus concentration in the SOD source was very high, i.e. about 8x10^{21} ions/cm^3, and the diffusion from SOD source is in non-equilibrium condition. From these figure we could recognise the indication of dopant depletion in long diffusion time experiments, where the lines are broken. The higher the diffusion temperature the sooner the SOD source is depleted. Figure 3-8 shows Arrhenius plots of phosphorus diffusion for 30 minutes at different temperatures.

Table 3-3: Sheet resistances and junction depths vs. temperature and time

	900°C		1000°C		1100°C	
t (minute)	Rs (Ω/sqr)	X_J(μm)	Rs (Ω/sqr)	X_J(μm)	Rs (Ω/sqr)	X_J(μm)
15	242.7	0.60	13.78	0.76	8.38	1.06
30	70.55	0.78	10.63	1.12	4.93	1.28
60	23.45	1.17	8.39	1.55	5.3	2.04
120	15.58	1.55	8.99	2.10	5.21	2.85
240	11.97	2.08	9.17	2.44	3.98	3.48



a) *b)*

Figure 3-7: Dependence on time$^{1/2}$ of a) reciprocal of sheet resistance and b)junction depth

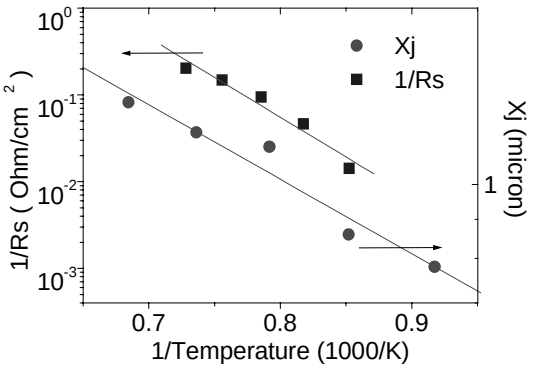


Figure 3-8: Arrhenius plots of sheet resistance and junction depth (diffusion time was 30')

In summary, the diffusion of phosphorus from P8545 source in a conventional furnace was investigated for the influence of diffusion time, temperature and also the baking temperature. The diffusion results did not fit with Fix's laws because of high concentration effect. The most effective baking temperature was found about 120°C for this P8545 diffusion source.

As in the case of B150 diffusion source, diodes fabricated from P8545 furnace diffusion also suffered from the etching problem. Better devices with shallower junctions were obtained by RTP which will be described in the following subsection.

3.2.3.2. Rapid thermal diffusion

Diffusion of phosphorus from P8545 source using RTP has been characterised using diode structures. RTD of P8545 resulted in sheet resistances of 790, 345, 92, and 78 Ω/sqr for samples diffused at temperature of 900°C, 950°C, 1000°C, and 1050°C for 20 seconds in nitrogen ambient. The junction depths of these samples were shallow due to low thermal budget, but no profile was measured.

Figure 3-9 shows the electrical characteristics of the diodes realised by different RTP temperatures. Although the characteristics here are better than those of the diodes fabricated using furnace diffusion of P8545, they are still not good. The leakage current density was more than 25 μA/cm^2 and the ideality factor was more than 2 for all diodes. Again the etching problem became severe and resulted in very high surface leakage currents. Guard-ring structure should be applied in these N^+P diodes to prevent surface leakage current.

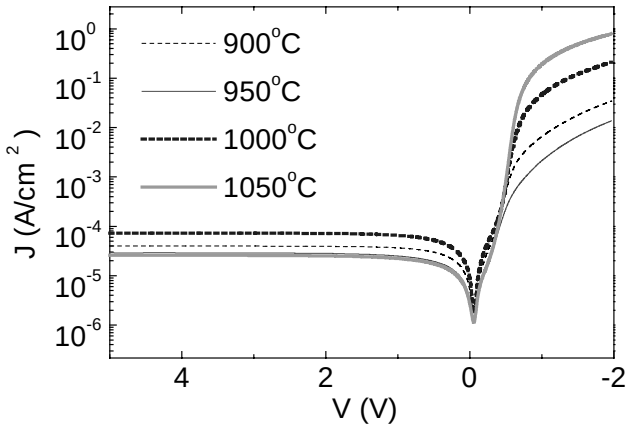


Figure 3-9: IV characteristics of diodes fabricated using P8545 RTD (diode area=800x800 μm^2)

3.2.4. Resume

Diffusion of boron and phosphorus into silicon from SOD sources have been investigated using conventional furnace and rapid thermal diffusion. Diffusion of boron was characterised using SRP and compared with simulation. This diffusion process from SOD was simulated as diffusion from a doped oxide source using equilibrium diffusion model. Diffusivity of boron in SOD was found to be much faster than in normal oxide and it was slower in Si. Diode realisation was carried out for electrical characterisation of these junctions. For the diodes fabricated by furnace annealing, the interface layer formed between the SOD layer and Si substrate during diffusion led to etching difficulty of the field oxide and resulted in bad diode characteristics. Fortunately, this problem became less profound in rapid thermal diffusion.

Very high performance shallow P^+N junctions, i.e. 50 nm deep, have been obtained by rapid thermal diffusion of B150 into Si. The diodes had very good characteristics with a low leakage current density of 1 nA/cm^2 and an ideality factor of 1.01. Junctions shallower than 20 nm were also achieved but diode quality suffered from very high surface leakage current. It was found that the peripheral leakage current increased when the junction depth decreased. It was explained by the contribution of SiO_2/Si interface leakage current. More SIMS measurements are needed to confirm the shallow P^+N junction depths.

Phosphorus diffusion from P8545 source into Si in conventional furnace and RTP was also carried out. Furnace diffusion junction depths and sheet resistances of samples at different diffusion temperatures and time were presented. Etching problem again resulted in defected diodes. This problem was less in RTP but still degraded the diodes' characteristics. The problem will be solved completely with poly-buffered diode structure which will be presented in the next section.

3.3. Diffusion of B & P into polysilicon/silicon [4]

As seen in the previous section, diffusion from SOD into Si can form ultra shallow junctions. However, the sheet resistance can be too high and the leakage current is increased when the junction become too shallow and the process window is small. Therefore, we studied the junction formation of elevated-contact diode structures by diffusion from polycrystalline and amorphous silicon layers doped by diffusion from spin-on dopants.

We deposit a polysilicon layer or amorphous silicon layer of about 300 nm on top of the Si substrate and then a SOD layer is deposited on top of this layer. At high temperature, dopant atoms will diffuse into this layer and then subsequently into the Si substrate to create a PN junction. This structure has advantages of low series resistance and the diode metal contact is brought far away from the semiconductor PN junction. As a result, the quality of diodes can be improved significantly. In addition, the buffered poly layer can be used as a sacrificial layer in modern silicidation processes for shallow junctions where silicon consumption is a big issue.

In this section we will present the technology and results on the study of diodes with as-deposited polysilicon and as-deposited amorphous silicon layers. Results of boron diffusion in conventional furnace and RTP will be presented first and then phosphorus.

3.3.1. Experiments

The experiments were carried out in almost the same way as the experiments in the previous section. The only difference was that there was a polysilicon or an amorphous silicon layer deposited on the mono-silicon before SOD deposition to serve as a buffer layer for dopant diffusion from the SOD source. This diffusion behaviour, is however, determined by the evolution of the polycrystalline-Si microstructure. This microstructure depends strongly on the initial deposited layer which is polycrystalline or amorphous and also depends on the doping impurities as well as the thermal budget. Small thermal budget is required for shallow junction formation. However, if this thermal budget is not enough the junction will be confined in the poly layer and it will result in a bad quality junction. Therefore, furnace diffusion and RTD have been applied to find minimum thermal budget for good quality shallow junctions.

In order to reduce the effect of the native oxide layer the wafers were dipped into 1 % HF solution for 1 minute before the polysilicon deposition process. 300 nm thick polysilicon or amorphous silicon was deposited on these wafers by LPCVD (Low Pressure Chemical Vapour Deposition) from SiH_4 at 625^{o}C and at a pressure of 0.1 mbar or 550^{o}C and 1 mbar respectively. Thickness of poly layer on each wafer was determined by weighing.

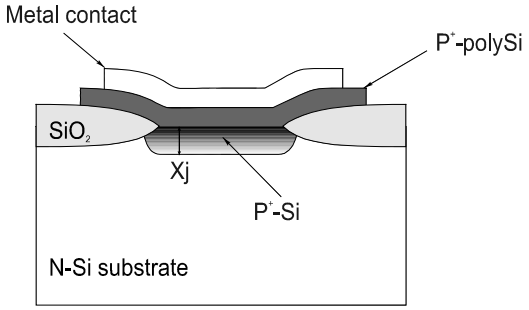


Figure 3-10: Polysilicon contacted diode structure

Extra steps were added to the diode process flow like poly-Si or α-Si deposition and patterning of this layer. The final structure of a poly-contacted diode is illustrated in figure 3-10.

The diffusion samples were profiled by Secondary Ion Mass Spectroscopy (SIMS), Spreading Resistance Probing measurement (SRP) and Differential Hall Effect technique (DHE). The I-V characteristics of the diode were also characterised by using a HP4146A semiconductor parameter analyser.

3.3.2. Boron diffusion results and discussion

3.3.2.1. Furnace diffusion

Diffusion of boron into the poly-Si or α-Si on Si structures was investigated using conventional furnace annealing at different temperatures from 800°C to 1000°C. All the samples were diffused for 30 minutes in nitrogen ambient. The etching of the SOD layer was not the problem any more because longer etching time could be used.

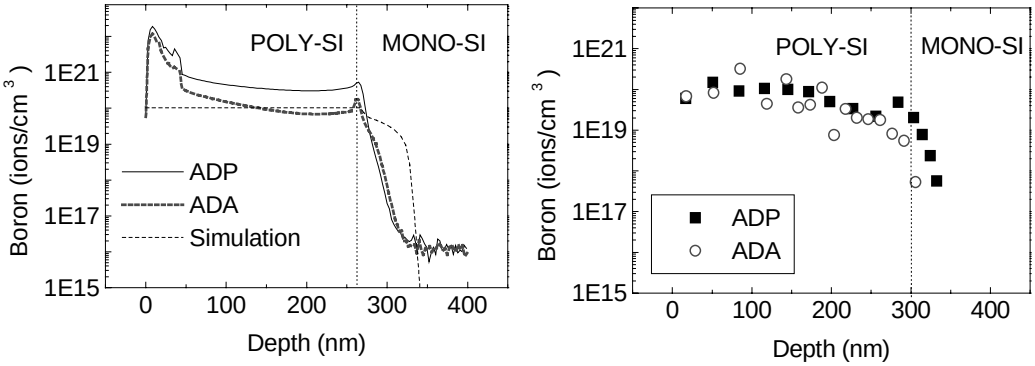


Figure 3-11: Profiles of sample diffused at 850°C 30 minutes in N_2: a) SIMS profiles of ADP (299 nm) and ADA (300 nm) on Si and a simulated profile; b) DHE profile of ADP (298 nm) and ADA(318 nm) on Si

The results showed that junctions in the Si substrate were very shallow after diffusion at 850°C for 30 minutes as illustrated for boron profiles in figure 3-11. Figure 3-11a shows SIMS profiles of both ADP/Si and ADA/Si samples diffused at 850°C and a simulated profile using default diffusivities. The thickness of the poly-layer was about 300 nm measured by weighing technique for all these SIMS samples. The poly/Si interface can be recognised by the small peak of boron concentration at about 300 nm deep and afterward the boron concentration drops quickly. This peak was caused by segregation effect of boron at the poly/Si interface. The poly thicknesses in these profiles were less than 300 nm because the SIMS measurement was calibrated with mono-crystalline Si. However, the thickness scale in the mono-silicon is more important for us and still correct. From that interface point we can define the junction depth in the silicon substrate. At 850°C the junction depths obtained were approximately 50 nm for both ADP and ADA samples.

The difference between the two concentrations may be due to the error during the measurement. It was difficult to remove this B150 layer at the beginning. There was accumulation of boron at the surface of the poly layer as illustrated in the SIMS profile. This accumulation was due to the thin residual SOD layer after etching. Simulation was carried out but there is no good model for diffusion in polysilicon available. The default model has been applied and the result showed a big difference between our measured profiles and the simulation.

Figure 3-11b shows DHE profiles of two samples ADP/Si (298 nm poly-Si) and ADA/Si (318 nm α-Si) diffused at 850°C . The DHE profile of ADP sample shows the active boron concentration in the poly-silicon of about $1x10^{20}$ ions/cm^3 and the junction depth is also approximately 50 nm similar to SIMS results. In the DHE profile of ADA sample, the boron concentration goes down already at the ADA/Si interface. The junction in this case was shallower than the ADP case.

Diodes were realised in the same conditions and were characterised. Table 3-4 shows parameters of poly-buffered diodes prepared at different temperatures and with as-deposited polysilicon (ADP) or as-deposited amorphous silicon (ADA) layer. In general, the diodes had very good characteristics with low leakage current and good ideality factor. All the ADP diodes had very good forward characteristic with ideality factor of 1.01. Although, the ideality factor was worse for the ADA diodes and it increased with diffusion temperature. The reason is not clear, may be it was due to contamination during processing of those diodes.

Table 3-4: Parameters of poly-buffered diodes fabricated from B150 furnace diffusion

	850°C				900°C				950°C				1000°C			
Poly gate	Poly		α		Poly		α		Poly		α		Poly		α	
Thickness (nm)	305		301		305		323		314		312		311		334	
RS (Ω/sq.) (poly/Si)	200		150		44		41		31		25		18		14	
Diode area (μm^2)	J_R (cm^{-2})	n	J_R (cm^{-2})	n	J_R (cm^{-2})	n	J_R (cm^{-2})	n	J_R (cm^{-2})	n	J_R (cm^{-2})	n	J_R (cm^{-2})	n	J_R (cm^{-2})	n
100x100	0.3 nA	1.01	0.9 nA	1.02	18.6 nA	1.01	3.3 nA	1.04	27.6 nA	1.01	3.5 nA	1.06	11.4 nA	1.01	1.2 nA	1.13
200x200	0.1 nA	1.01	0.3 nA	1.02	10.9nA	0.99	1.3 nA	1.05	11.1 nA	1.00	1.1 nA	1.07	5.3 nA	1.00	0.5 nA	1.16
400x400	87.5 pA	1.01	0.2 nA	1.02	4.5 nA	1.01	0.5 nA	1.05	4.8 nA	1.00	0.4 nA	1.10	2.4 nA	1.00	0.2 nA	1.18
800x800	51.6 pA	1.01	0.1 nA	1.02	1.5 nA	0.99	3.4 nA	1.04	0.9 nA	1.01	0.2 nA	1.06	0.6 nA	1.00	0.2 nA	1.20

Figure 3-12 shows the diode characteristics of all diodes. In this figure, the diode characteristics realised at different temperatures seem strange for both ADP and ADA cases. Especially in the forward region, at higher diffusion temperatures the forward current is lower even though the sheet resistance of the P^+ layer is lower. It means that at higher diffusion temperature the series resistance increased. A possible explanation is that the wafer backside was doped by boron out-diffusion from other wafers in the furnace. This may happen even when the wafers were positioned with SOD layers facing each other because SOD was evaporated strongly especially at high temperature. As a result, there were two P^+N junctions on both sides of the wafers. The higher temperature resulted in higher doping concentration and deeper P^+N junction on the back side. This parasitic junctions can increase the series resistance and resulted in lower forward current of the front side P^+N junction. This effect was also observed in all the cases of furnace diffusions from P8545 or B150 into both Si as well as poly/Si. However, it did not happen in RTP process because this is a single wafer process, see subsection 3.2.2.2.

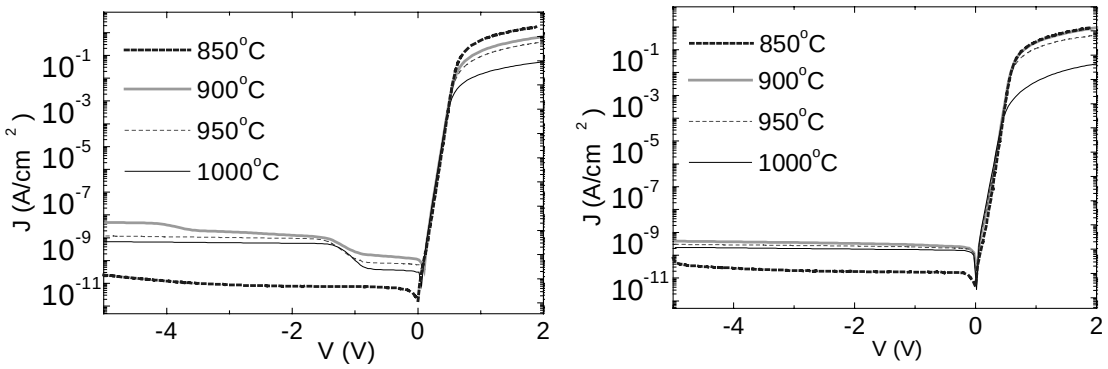


Figure 3-12: IV characteristics of poly-buffered diodes fabricated by furnace diffusion of B150 on a) ADP b) ADA; (diode area = 800x800 μm^2)

The leakage current of diodes decreased as the diffusion temperature increased, except for the case of 850°C. The diodes realied at 850°C were not in the same experimental series. They were fabricated afterward in order to see what was the minimum thermal budget to have good quality shallow junction. The backside diffusion effect also had influence on the leakage current of the diodes. An increase of reverse current at -0.7 V indicates the forward threshold of the backside junction, proving the above mentioned assumption.

The leakage current density of diodes are plotted vs. the L/A ratio for both ADP and ADA cases in figure 3-13. Linear fittings were still obtained but the bulk leakage current densities were negative. These fitting results could also be influenced by the backside diode characteristics.

Table 3-5: Linear fitting equations of diode leakage current density

	ADP	ADA
850°C	y = 7E-13x + 1E-11	y = 2E-12x - 3E-11
900°C	y = 5E-11x - 2E-10	y = 1E-11x - 5E-10
950°C	y = 8E-11x - 3E-09	y = 9E-12x - 3E-10
1000°C	y = 3E-11x - 8E-10	y = 3E-12x - 2E-11

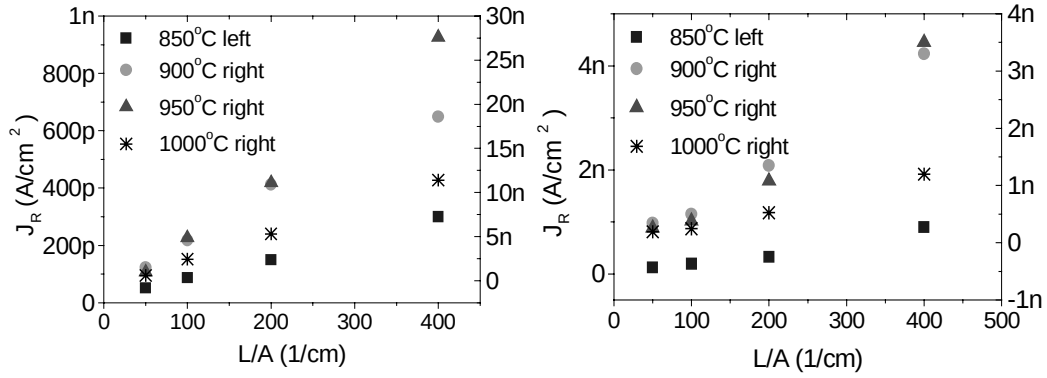


Figure 3-13: B150-FA Diode leakage current density as function of L/A ratio, a)ADP; b)ADA

As seen above, all the diodes realised at diffusion temperatures equal or above 850°C for 30 minutes had good electrical properties. At 850°C the junction in the mono-crystalline Si was already very shallow, i.e. approximately 50 nm, as illustrated in figure

3-11. This thermal budget was about the lower limit to form a good shallow junction in the Si substrate. A lower thermal budget or a thicker as-deposited layer would result in a worse diode because the junction in the Si becomes too shallow and the non-uniformity of the junction starts to degrade the diode characteristics. This effect is dependent on the micro-structure of the as-deposited layer, i.e. ADP or ADA. The difference in their diode characteristics is illustrated in figure 3-14.

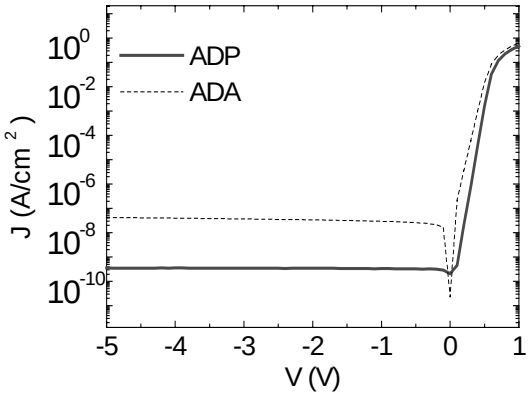


Figure 3-14: diode characteristics of ADP(298 nm) and ADA (318 nm) – B150, 850°C, 30'

In this figure the diodes were realised at 850°C for 30 minutes. Their boron profiles, measured by the DHE technique, are shown in figure 3-11b. It should be noticed that all the profiles in figure 3-11 were measured from the samples diffused at the same temperature of 850°C but the as-deposited layers of these samples had different thickness. The thickness of the as-deposited polysilicon layers was about the same in the DHE sample, the SIMS sample and the diode sample in table 3-4. Therefore, their profiles and the diode characteristics were similar.

On the other hand, the as-deposited amorphous layer in the DHE sample was 20 nm thicker than the ADA layer in the SIMS and diode sample in table 3-4. This thicker ADA layer resulted in a shallower junction in the mono-Si substrate as illustrated in figure 3-11b and worse diode characteristics as shown in figure 3-14. As seen in figure 3-14, the diode characteristic of ADA sample became worse with higher leakage current and non-ideal forward characteristics but the ADP diode still had good characteristics as the sample listed in table 3-4.

For the ADP diodes, the degradation effect occurred only when the diffusion temperature went down to 800°C. At this temperature the ADP diodes became worse with leakage current density of about tens of nA/cm^2 while the ADA diodes had totally bad characteristic with leakage current density of more than 20 μA/cm^2. From the diode characteristics we could conclude that in order to obtain a good shallow junction a minimum thermal budget of 850°C and 30 minutes diffusion is needed for the ADA diodes and 800°C, 30 minutes for ADP diodes.

The difference in minimum thermal budget can be explained by the influence from the micro-structure of as-deposited polysilicon layer and as-deposited amorphous silicon layer to the diffusion process from SOD as illustrated in figure 3-15.

Polycrystalline Si and amorphous silicon layer are constructed by many small mono-crystalline grains with different orientations. At grain boundary regions, the defect concentration is very high and impurities can diffuse very fast through boundaries at high temperature. At the same time, impurities can also diffuse to the grains and go through them with slower diffusivity. Figure 3-15 illustrates our technique using SOD diffusion source to dope the ADP on Si or ADA on Si structures. In this figure we can

see that boron atoms from SOD diffuse very fast through the poly-Si or α-Si layer and then subsequently into the mono-Si with lower diffusivity and got smeared out there. However, the difference in diffusivity of boron in the grains and grain boundaries will give rise to non-uniformity of the junction in the mono-Si, so the junction depths in the mono-silicon under the grain boundaries are often deeper than under the grains. This non-uniformity could cause higher leakage current of the junction in the mono-crystalline Si if this junction is too shallow. In the worst case, if the junction is confined within the poly layer the diode properties will be very bad due to defective structure of the poly layer. If the thermal budget is high enough the lateral diffusion of boron in the mono-Si will increase the junction uniformity and improve the electrical properties of the junction.

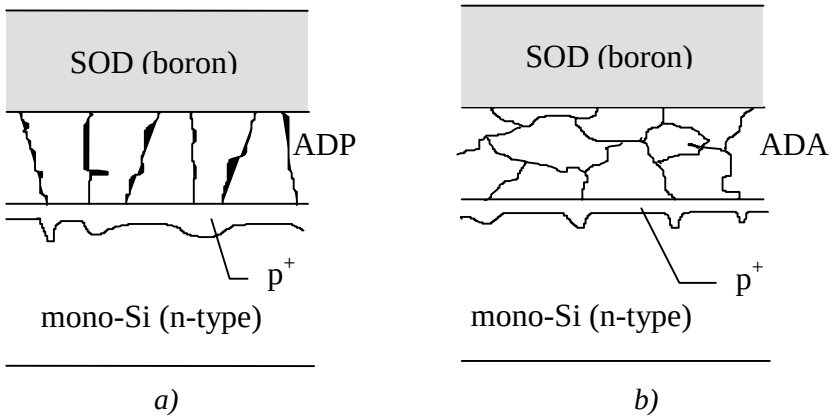


Figure 3-15: Models of boron diffusion from SOD through an ADP (a) or an ADA (b) layer to the mono-crystalline Si substrate

It is well known that polysilicon layers obtained from ADP often have columnar structure while it has random structure from ADA layer as illustrated in figure 3-15 [5]. According to the diffusion model, described above, the diffusion of boron in ADP should be faster than in the ADA layer due to the diffusion lines through boundaries are shorter in the ADP layer. Therefore, a larger thermal budget was required to form a good junction in ADA case than in ADP. This can explain the difference in measurement results of the ADP and ADA diodes.

In summary, boron diffusion from B150 into ADP and ADA on Si has been studied. Very high quality P^+N shallow junctions have been obtained for both ADP/Si and ADA/Si structures. At a thermal budget of 850^{o}C, 30 minutes in N_2 ambient the diffusion of boron from B150 resulted in both very good ADP/Si and ADA/Si P^+N junction with junction depths of about 50 nm in the mono-Si, a leakage current density of less than 0.1 nA/cm^2 and ideality factor of about 1.01 for more than 8 decades of forward current.

The quality of diodes depends strongly on the micro-structure of the as-deposited buffer layer especially at low thermal budget. At 850^{o}C, 30 minutes thermal budget the random structure of the as-deposited amorphous Si layer still resulted in good junction. However, when this amorphous layer was thicker or the diffusion temperature was lower then the diode characteristics were degraded because the junction was too close to the poly/mono Si interface. On the other hand, columnar structure of as-deposited polysilicon layer allowed lower thermal budget to form a shallower junction in the mono-Si substrate. However, as-deposited amorphous layer resulted in lower resistance and smoother surface of the formed polysilicon after diffusion.

The SOD etching difficulty has been eliminated using this poly-buffered diode structure and diode quality has been improved dramatically. However, evaporation of B150

during diffusion process in furnace caused wafer backside doping. RTP process or backside protection by field oxide should be employed to get rid of this problem. In the following subsection we will present results of rapid thermal diffusion study.

3.3.2.2. Rapid thermal diffusion

Rapid thermal processes were applied to study the diffusion of boron into the ADP and ADA on Si structures. Experiments were carried out at different temperatures from 950°C to 1250°C with a diffusion time of 20 seconds. Figure 3-16a shows SRP profiles of ADP/Si samples diffused at 1100°C and 1150°C and the junction depths in the mono-Si were about 90 nm and 110 nm respectively. SIMS profiles of ADP/Si and ADA/Si samples diffused at 1230°C are shown in figure 3-16b. In this figure we could see that diffusion of boron in ADP was faster than ADA and as a result, the junction in the mono-Si substrate was deeper in the ADP than in ADA sample. This is good evidence to support the model of impurity diffusion through a deposited Si layer into mono-Si substrate as discussed at the end of the previous subsection. Table 3-6 shows the parameters of diodes realised at temperature from 1000°C to 1150°C for the ADP case and from 950°C to 1050°C for ADA case.

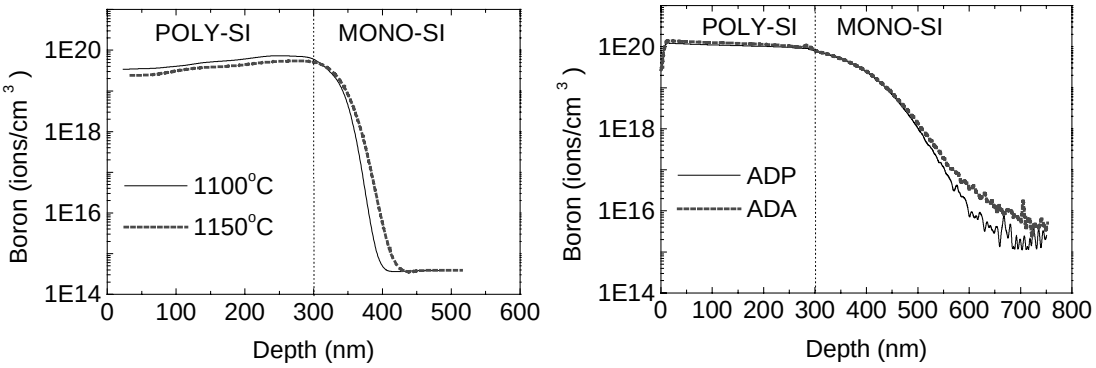


Figure 3-16: a) SRP profile of ADP/Si samples, B150 - RTD at 1050°C and 1100°C, 20s; b) SIMS profiles of ADP/Si and ADA/Si samples, B150 – RTD at 1230°C, 20s.

Table 3-6: Parameters of poly-buffered diodes fabricated from B150 rapid thermal diffusion

	1000°C				1050°C				1100°C		1150°C	
Poly gate	Poly		α		Poly		α		Poly		Poly	
Thickness (nm)	312		301		310		323		300		301	
RS (Ω/sqr) (poly/Si)	500		150		200		41		87		50	
Diode area (μm²)	J_R (cm^{-2})	n	J_R (cm^{-2})	n	J_R (cm^{-2})	n	J_R (cm^{-2})	n	J_R (cm^{-2})	n	J_R (cm^{-2})	n
100x100	10.6 μA	1.32	271 μA	1.84	29.3 nA	1.06	0.2 μA	1.33	8.4 nA	1.11	6.2 nA	1.09
200x200	9.3 μA	1.47	294 μA	3.61	15.0 nA	1.15	3.2 μA	1.36	3.1 nA	1.12	3.1 nA	1.12
400x400	4.2 μA	1.44	316 μA	3.32	10.7 nA	1.25	0.6 μA	1.40	1.5 nA	1.17	1.3 nA	1.10
800x800	3.0 μA	1.40	367 μA	2.91	3.1 nA	1.03	0.7 μA	1.37	1.2 nA	1.21	2.7 nA	1.17

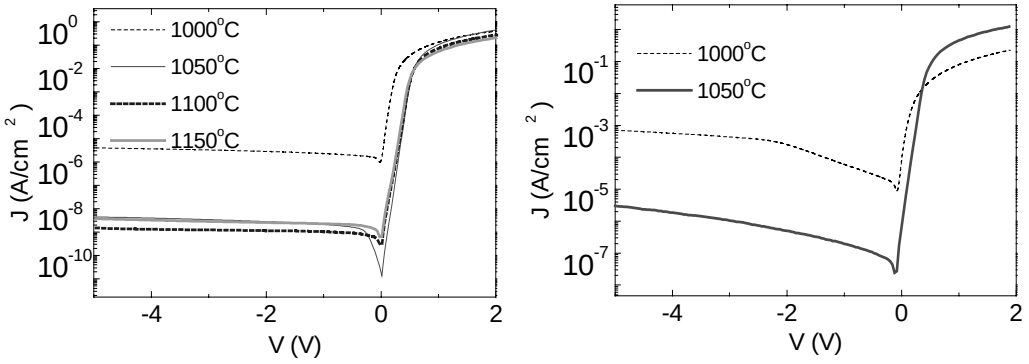


Figure 3-17: IV characteristics of poly-buffered diodes fabricated by RTD of B150 on a) ADP/Si and b) ADA/Si; (diode area = 800x800 μm^2)

Table 3-7: Linear fitting equations of diode leakage current density

	ADP
1000°C	y = 2E-08x + 3E-06
1050°C	y = 7E-11x + 1E-09
1100°C	y = 2E-11x - 4E-10
1150°C	y = 1.2E-11x + 1E-9

Linear fitting of leakage current density vs. the L/A ratio was obtained for ADP diodes and the equation are listed in table 3-7. Except for the case of 1100°C, we can see that the bulk leakage current density of good diodes was about 1 nA/cm^2 and the peripheral leakage current density was decreased from 70 pA/cm to 12 pA/cm when the diffusion temperature increased. In addition, there was no problem of wafer backside doping as was the case for furnace diffusion.

This leakage current result is similar to the results of diodes realised by RTD of B150 directly on Si. However, the results of poly-buffered diodes in the case of furnace diffusion of B150 is still much better than both RTD results. The reason may be caused by the rapid thermal process its self. The degradation of diodes quality due to RTP have been reported in literature [6]. In this work, junctions were formed by implanting dopants into silicides and then drive out by RTA (Rapid Thermal Anneal). Compared to furnace processed junctions, however, the RTA processed junctions showed a higher leakage current, coupled with a non-ideal diode behaviour. In addition, good junctions processed by furnace annealing showing ideal behaviour and low leakage current are converted to leakier non-ideal diodes after an additional RTA. On the other hand, leaky junctions processed by RTA can be "cured" by an additional furnace annealing. This RTP-defect induced leakage current may be due to stress enhancement on wafers under very high temperature ramping rate.

Compared between ADP and ADA diodes, good diodes were obtained at an RTP temperature of 1050°C or higher in the case of ADP but not for ADA diodes. The junction depth of ADP sample diffused at 1050°C should be less than 70 nm as compared with the SRP profiles in figure 3-16a. Compared between ADP and ADA diodes, the minimum thermal budget to obtain good diodes was lower for ADP than for ADA diodes, e.g. for ADP diode it was about 1050°C and 20 s diffusion, while it was higher for ADA. As seen in figure 3-17b, the leakage current densities of ADA diodes were still very high, in the order of μA. We only had samples of ADA diodes realised at

a temperature of 1230°C. At this very high thermal budget, the quality of ADA diodes were very good with leakage current densities of about 0.3 nA/cm^2.

In summary, high performance P^+N shallow junctions were obtained using RTD of B150 on both ADP/Si and ADA/Si. The minimum thermal budget was found higher for ADA diodes than for ADP diodes which was similar to the case of furnace diffusion experiments. However, the quality of RTD diodes were degraded due to the rapid thermal process which has been proven to create stress-induced defects in the shallow PN junctions.

3.3.3. Phosphorus diffusion results and discussion

3.3.3.1. Furnace diffusion

The diffusion study of phosphorus was carried out similar to the boron diffusion in the previous subsection. Table 3-8 shows the parameters of diodes realised at temperatures in the range 850°-950°C for both ADP and ADA cases.

Table 3-8: Parameters of poly-buffered diodes fabricated from P8545 furnace diffusion

	850°C				900°C				950°C			
Poly gate	Poly		α		Poly		α		Poly		α	
Thickness (nm)	314		307		315		290		314		290	
RS (Ω/sq.) (poly/Si)	440		270		90		30		12		11	
Diode area (μm^2)	J_R (cm^{-2})	n	J_R (cm^{-2})	n	J_R (cm^{-2})	n	J_R (cm^{-2})	n	J_R (cm^{-2})	n	J_R (cm^{-2})	n
100x100	0.8 μA	1.14	177 μA	2.04	1.2 nA	1.09	1.4 nA	1.11	3.7 nA	1.06	1.8 nA	1.04
200x200	1.5 μA	1.15	134 μA	1.59	0.6 nA	1.07	0.7 nA	1.08	0.8 nA	1.04	0.9 nA	1.04
400x400	5.4 μA	1.34	107 μA	1.41	0.4 nA	1.08	0.4 nA	1.09	0.5 nA	1.01	0.5 nA	1.10
800x800	7.9 μA	1.17	77 μA	1.30	0.3 nA	1.09	0.3 nA	1.12	0.5 nA	1.04	0.3 nA	1.09

Figure 3-18 shows the diode characteristics. The leakage current density of good diodes was very low, i.e. less than 0.5 nA/cm^2. The diodes characteristics were good at diffusion temperatures equal to or higher than 900°C. The minimum thermal budget was somewhere in between 850°C and 900°C for both ADP and ADA. At 850°C both ADP and ADA diodes showed bad characteristics but ADP diodes seemed better. This difference between ADP and ADA diodes was similar as in the case of boron in the previous subsection. However, the minimum thermal budget here was higher than for boron even though the diffusivity of phosphorus in mono-Si is higher than boron. Therefore, the diffusion of phosphorus within the poly layer must be different from boron.

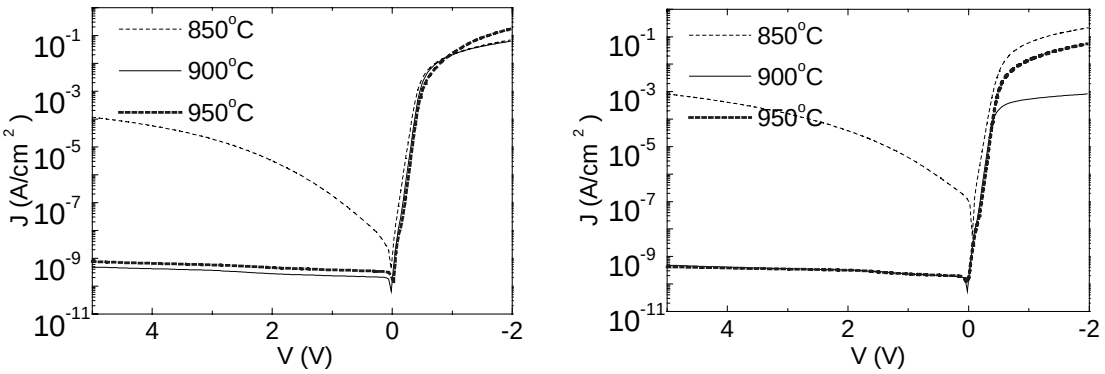


Figure 3-18: IV characteristics of poly-buffered diodes fabricated by furnace diffusion of P8545 on a) ADP/Si b) ADA/Si; (diode area = 800x800 μm^2)

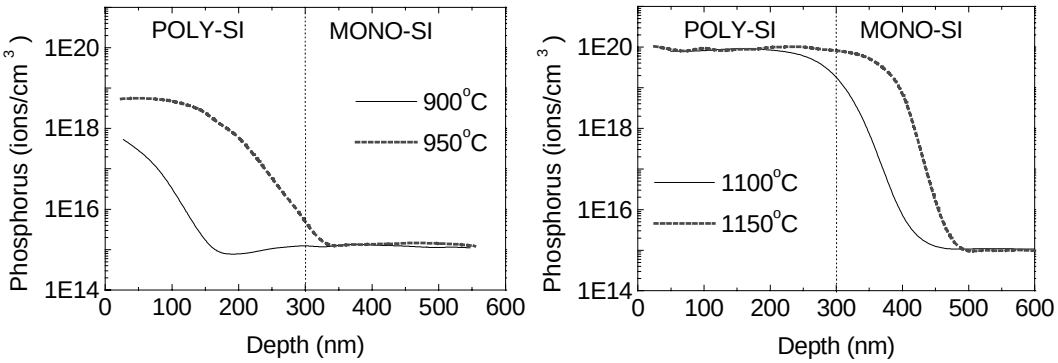
As the diffusion model of impurities through a poly layer into the Si substrate described in the previous subsection, the micro-structure of the as-deposited layer can determine the uniformity of the junction in the Si substrate. Furthermore, because the grains are growing during the diffusion process the micro-structure will be changed as well as the uniformity of the junction. This non-uniformity depends on the size of grains. Larger grains could lead to more lateral non-uniformity and hence leakier junction.

It is known that this secondary growth rate is determined by the type of doping impurity. Phosphorus or arsenic was found to substantially enhance grain growth while boron doping had little effect [7]. It is argued that phosphorus-enhanced grain growth occurs due to an increase in the grain boundary atomic mobility. A kinetic model is suggested which can be used to quantitatively predict the observed enhancement over wide ranges of temperature and doping. In this model, it is assumed that grain growth occurs through a diffusive process and/or a non diffusive process. Phosphorus doping enhances the diffusive process by increasing the number of charged vacancies and therefore the total number of vacancies. As a result, the grains are often bigger for phosphorus doping than for boron doping. In addition, phosphorus and arsenic segregate at the boundaries of the poly layer and diffuse slower in the poly layer than boron [8]. Hence, phosphorus diffusion in the poly layer could lead to more non-uniform junctions than for boron. This can explain the higher minimum thermal budget for P diffusion to obtain a good junction in the Si substrate than B.

In summary, high performance N^+P junctions have been obtained by furnace diffusion of phosphorus from SOD-P8545 source into the as-deposited polysilicon and as-deposited amorphous layer on Si. Similar to boron, the minimum thermal budget to form a good junction in the Si substrate was found higher for the ADA than for the ADP case. However, this minimum thermal budget was higher than in the case of boron diffusion. This was due to the lower diffusivity of phosphorus in the poly layer and due to the larger grains of phosphorus doped poly layer.

3.3.3.2. Rapid thermal diffusion

Similar to boron, RTD experiment of phosphorus from P8545 source were carried out for ADP and ADA samples. Junction depths in the silicon substrate were also very shallow as illustrated in figure 3-19. The junction depth was about 110 nm and 170 nm for 1100°C and 1150°C respectively.

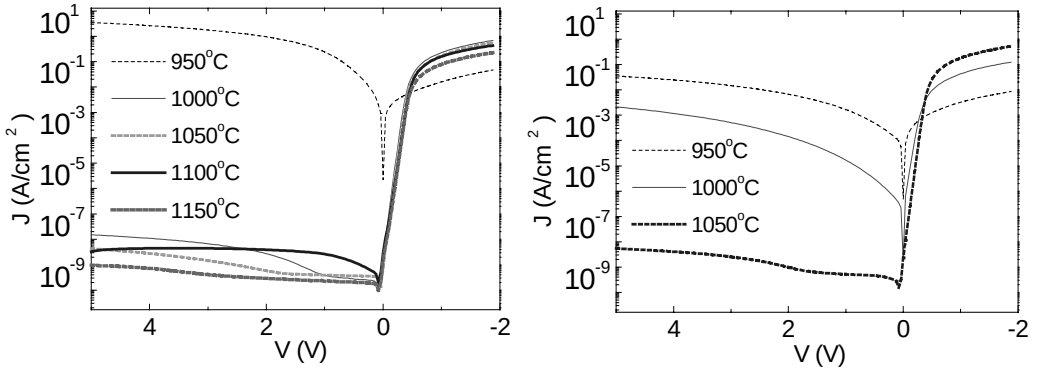


a) *b)*

Figure 3-19: SRP profile of ADP/Si samples, P8545 - RTD at a) 900°C and 950°C, 20s; b) 1100°C and 1150°C, 20s.

Table 3-9: Parameters of poly-buffered diodes fabricated from P8545 rapid thermal diffusion

	950°C				1000°C				1050°C				1100°C		1150°C	
Poly gate	Poly		α		Poly		α		Poly		α		Poly		Poly	
Thickness (nm)	297		307		293		311		309		315		300		297	
RS (Ω/sq.) (poly/Si)	110		165		70		115		40		81		24		16	
Diode area (μm^2)	J_R (cm^{-2})	n	J_R (cm^{-2})	n	J_R (cm^{-2})	n	J_R (cm^{-2})	n	J_R (cm^{-2})	n	J_R (cm^{-2})	n	J_R (cm^{-2})	n	J_R (cm^{-2})	n
100x100	1.39 A	6.94	29.4 mA	2.78	0.1 µA	1.11	68.7 µA	1.22	17.3 nA	1.05	8.4 nA	1.01	42.9 nA	1.08	4.3 nA	1.03
200x200	2.63 A	8.40	24.8 mA	4.15	40.7 nA	1.12	130 µA	1.26	3.6 nA	1.01	4.1 nA	1.01	14.4 nA	1.06	1.1 nA	1.03
400x400	2.62A	8.08	19.0 mA	5.43	19.8 nA	1.12	229 µA	1.32	2.4 nA	1.01	2.5 nA	1.01	15.6 nA	1.07	0.5 nA	1.01
800x800	1.34A	7.24	10.3 mA	5.75	4.6 nA	1.02	283 µA	1.42	0.9 nA	1.00	1.7 nA	1.00	4.2 nA	1.06	0.3 nA	1.02



a) *b)*

Figure 3-20: IV characteristics of poly-buffered N^+P diodes fabricated by RTD of P8545 on a) ADP/Si and b) ADA/Si; (diode area = 800x800 μm^2)

Good N^+P junctions were obtain for ADP diodes realised at temperatures of 1000^{o}C or higher for ADP diodes while this temperature was 1050^{o}C or higher for ADA diodes. Diode quality here was about the same as boron RTD and also worse than furnace diffusion diodes. Leakage current density fitting was carried out but good linear fitting could not obtained. The reason could be due to surface leakage current which often occurred in N^+P samples.

However, compared with RTD boron experiments the diffusion of phosphorus here seems faster in the poly layers. This effect was quite different from furnace diffusion where phosphorus diffused slower in the poly than boron. This difference may be caused by the secondary grain growing effect. The rapid thermal diffusion time was too short for growth of poly grains and the grain growth could be neglected. As a result, diffusion of phosphorus in polysilicon was not retarded by this growth effect and the diffusivity of phosphorus was higher than boron which is similar to the diffusion of P and B in mono-crystalline Si.

3.3.4. Resume

Diffusion of boron and phosphorus from SOD sources into as-deposited polysilicon on Si and into as-deposited amorphous silicon were investigated using both furnace annealing and rapid thermal annealing. Good quality shallow junctions were obtained for all cases. However, RTD resulted in worse junction quality than furnace diffusion due to rapid-thermal-induced stress-enhanced defects.

In both furnace and rapid thermal annealing, impurities diffused faster in as-deposited polysilicon than in as-deposited amorphous silicon due to difference in their micro-structure, i.e. columnar structure in ADP and random structure in ADA layer. However, as-deposited amorphous layers had advantages of lower resistance and smoother surface.

It was found that phosphorus diffusion in the poly layer was slower than boron during furnace annealing while it was faster during rapid thermal annealing. This effect was explained by the influence of secondary grain growth in the poly layer during diffusion.

3.4. Conclusions

Diffusion of boron and phosphorus into Si and as-deposited polysilicon (ADP) on Si and as-deposited amorphous (ADA) silicon on Si were investigated using both conventional furnace and rapid thermal processes. The main conclusions from this study are:

- High quality P^+N and N^+P shallow junctions, less than 50 nm, were obtained for both furnace diffusion as well as rapid thermal diffusion using SOD.
- Furnace diffusion resulted in better quality diodes than RTD because the RTD process created stress-induced defects in the PN junctions.
- Diode quality was improved significantly with the poly-buffered diode structure.
- In the case of impurity diffusion in the polysilicon/Si (or α-Si/Si) the micro-structure of the as-deposited layer and its evolution during diffusion process was very important for the properties of the junction in the mono-crystalline substrate. Random structure of ADA layer caused slower diffusion for both boron and phosphorus than in the ADP layer with columnar structure. Therefore, a higher thermal budget was required to form a good shallow junction using ADA than ADP.

- Phosphorus diffusion in the poly layer was slower than boron during furnace annealing while it was faster during rapid thermal annealing. This effect was due to the influence of secondary grain growth in the poly layer during diffusion.

This simple SOD diffusion technique has been proven to be very effective and promising for shallow junction formation in future ULSI technology.

References

[1] Sorab K. Ghandhi, "VLSI fabrication principles : silicon and gallium arsenide", 2nd ed. - New York [etc.] : Wiley, cop. 1994. 834 p. ISBN 0-471-58005-8

[2] S.M. Sze, Physics of semiconductor devices, 2nd ed. 1981, p. 279

[3] N. N. Toan, N. D. Chien, V. T. Son, N. T. Giang, J. Holleman, P.H. Woerlee, "Spin-On Glass (SOG) materials and applications", 2nd International Workshop on Materials Science Proceeding, IWOMS'95, Hanoi 1995, p. 191

[4] N. N. Toan, N.D. Chien, J. Holleman, P.H. Woerlee, "Formation of shallow junctions by dopant diffusion from Spin-On Glass", 1st annual workshop on Semiconductor Advanced for Future Electronics (SAFE) proceeding, The Netherlands, Nov. 1998.

[5] Ted Kamins, Ed., Polycrystalline silicon for integrated circuit applications, Kluwer Academic Publishers, 1988, p.53

[6] R. Liu, C.-Y. Lu, J.J. Sung, C.-S. Pai, N.-S. Tsai, "The effects of rapid thermal processing on ultra-shallow junctions for deep sub-micron MOSFETs", Solid state electronics Vol. 38 (Issue 8), 1995, p. 1473

[7] H.-J. Kim, "Kinetic modeling of grain growth in polycrystalline silicon films doped with phosphorus or boron", J. Electrochem. Soc.,135 (9), 1988, p.2312

[8] Theodore I. Kamins, "dopant segregation in polycrystalline silicon", J. Appl. Phys. 51 (11), Nov. 1980, p. 5755

4

Realization of High Frequency Bipolar Transistor

4.1. Introduction

Since the first point-contact bipolar transistor [1], invented in 1947, there has been much progress in the development of very high density integrated circuits (IC), often referred to as very large scale integration (VLSI). Although many new devices have been invented, the bipolar transistor is still one of the key devices in every IC. Bipolar devices offer the following advantages compared to CMOS: high current drive ability, high frequency response, good matching, analog applications, low noise and high transconductance.

To take benefit of the advantages of both CMOS and bipolar devices, several IC-processes have been developed that combine both types of devices on the same chip which are called BiCMOS processes. A cross-sectional view of an example of a BiCMOS device is given in figure 4-1.

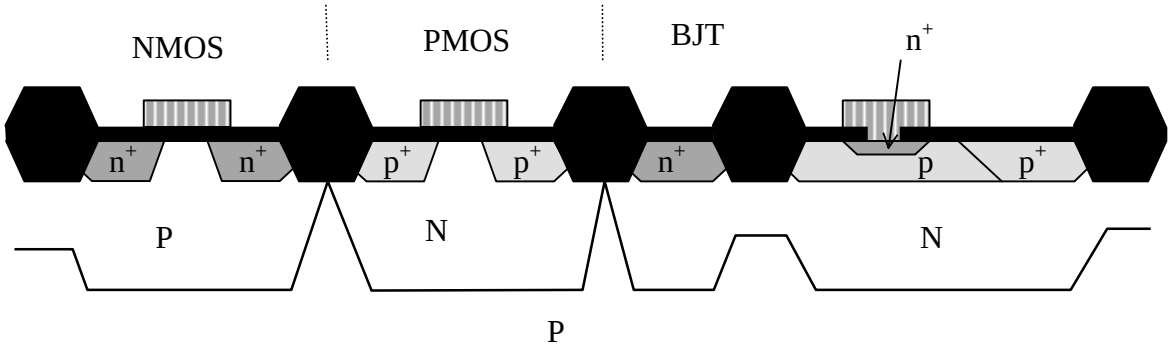


Figure 4-1: Cross-section of a BiCMOS device

The MESA research centre of the University of Twente (UT) has developed a BiCMOS process called UT-BiCMOS process. One of the characteristics of the UT-BiCMOS process is the application of high energy ion implantation for the formation of the CMOS wells, and the collector of the bipolar junction transistor (BJT) (figure 4-2a).

Conventionally, the collector of a bipolar transistor consists of a heavily doped buried layer with a lightly doped epitaxially grown layer on top of it (figure 4-2b), to obtain a low collector resistance as well as a minimum collector-base junction capacitance, a high Early voltage, and a sufficiently high collector-emitter breakdown voltage. However, this epitaxy technology is rather expensive. An alternative is using high energy ion implantation for the collector. The implementation of collector implantation is a routine in IC technology. Nevertheless, the implanted collector has some disadvantages and the process needs to be optimised. One of the challenges is the deep collector formation which requires very high implantation energy. Furthermore, the implantation often causes damages and high temperature annealing steps usually cause diffusion of impurities and change the implanted profiles. Therefore, the implantation profiles as well as the annealing processes have to be optimised.

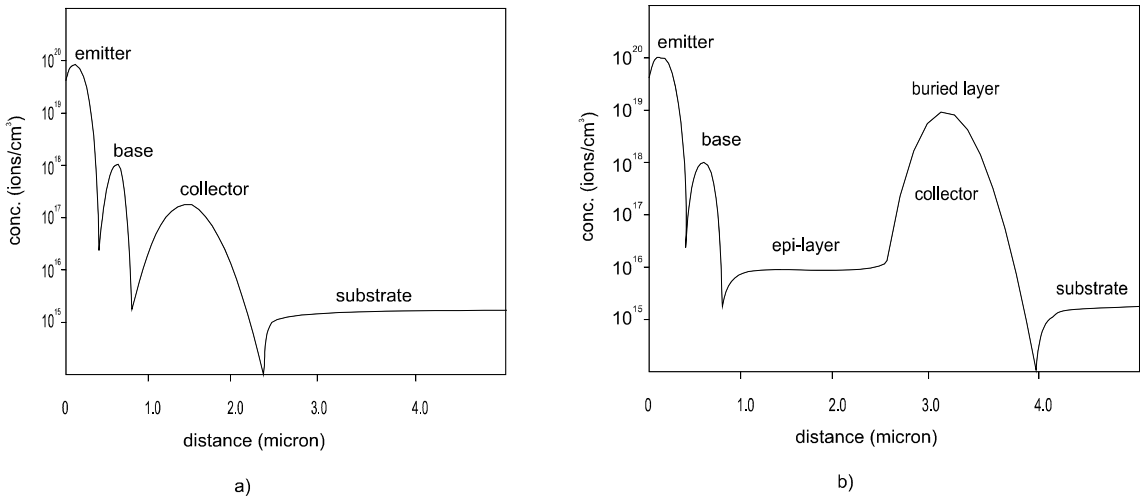


Figure 4-2: Typical net doping profiles of a BJT: a) implanted collector; b) conventional collector

❖ *The UT-BiCMOS process and its bottle-necks*

Table 4-1: Specification of the mask set needed for the UT-BiCMOS process.

	Mask name and number	CMOS	Bipolar	EPROM
1	AA (Active area)	x	x	x
2	DN (Deep N)	x		x
3	DP (Deep P)	x		x
4	PT (Preventing Trunk)			x
5	PS (PolySilicon)	x		x
6	BC (Base-Collector)		x	
7	EM (Emitter)		x	
8	SPS (Second PolySilicon)		x	x
9	CP (Collector Plug)		x	
10	SN (Shallow N)	x	x	x
11	SP (Shallow P)	x	x	x
12	CO (Contact Opening)	x	x	x
13	IN (Interconnect)	x	x	x
14	CB (Contact Bondpad)	x	x	x

The UT-BiCMOS process was developed by Wijburg [2]. This process was designed for double poly, double metal, LOCOS isolated, 5 V process, yielding NMOS, PMOS, and bipolar npn transistors, as well as a VIPMOS-EEPROM. The npn devices have a poly-silicon emitter with a minimum area of 2.5x2.5 μm^2, which is limited by 2.5 μm lithography. The minimum channel length of the MOS transistors is 1.5 μm due to lateral straggle and out-diffusion of the source/drain contacts, and the gate oxide thickness is 25 nm. The process has a so called modular structure, meaning that CMOS, bipolar and VIPMOS-EEPROM can be optimized rather independently. The different devices have only a few process steps in common, e.g. LOCOS isolation, contact holes formation and metallization. The reason for choosing this mutual independence at the cost of some extra masks, is the increased flexibility for modification of one or more parts of the process. In a university environment IC-processes usually are a matter of research, and therefore process flexibility is considered more important than process simplicity. Table 4-1 shows the complete set of 14 masks, and specifies which device is affected by a certain mask.

As seen in the table, the bipolar emitter, base, and collector implantation can be modified without affecting the CMOS transistors. Contrary of course to a modification of temperature processing steps which affect the entire wafer. The bipolar process flow is shortly illustrated in figure 4-3.

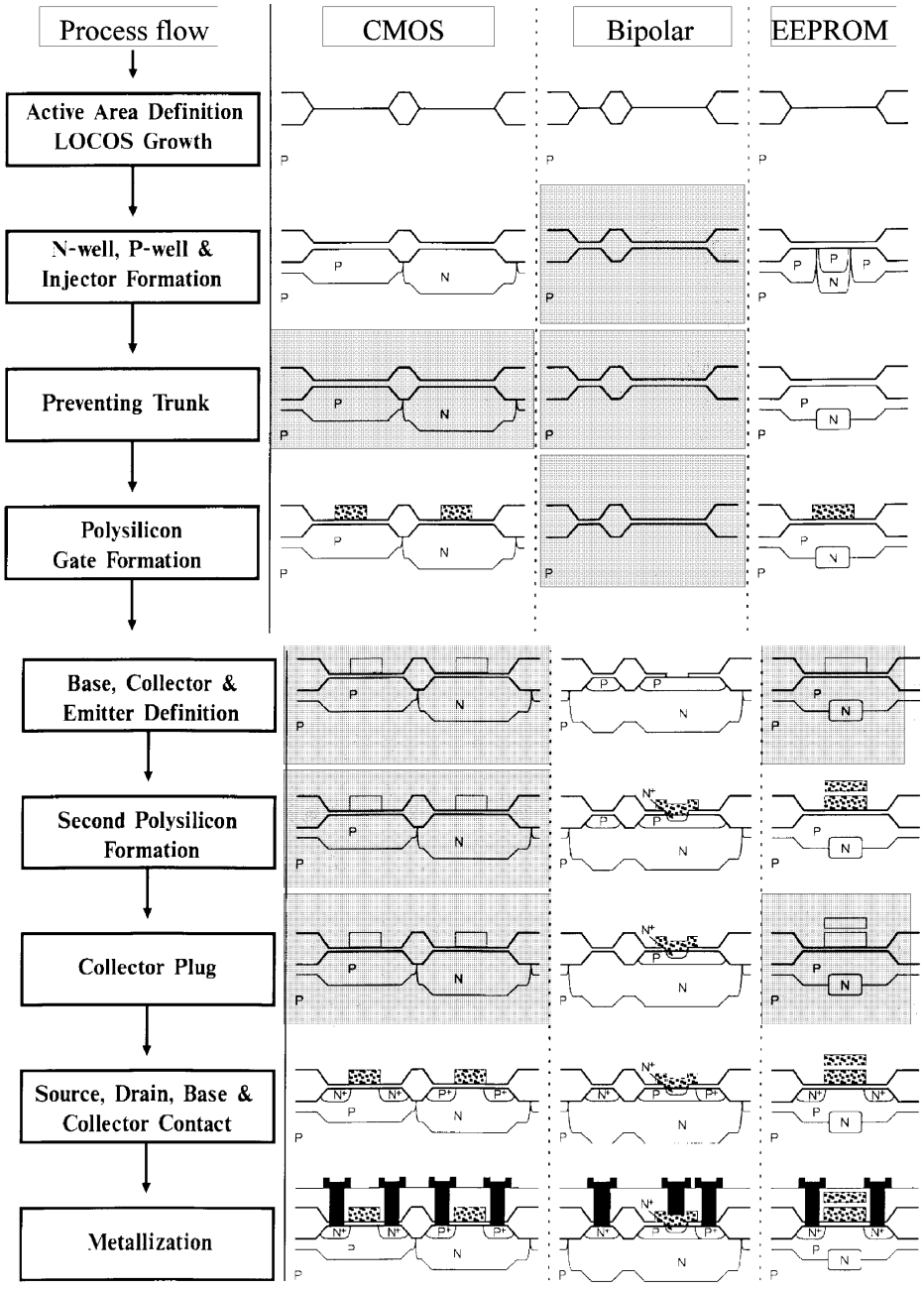


Figure 4-3: Standard UT-BiCMOS process flow

In the UT-BiCMOS process we have used very high energy implantation for collector formation. However, former development of the UT-BiCMOS process required a collector implantation at the highest possible acceleration voltage of the implanter (500 kV). Further more, triply charged phosphorus ions are needed to obtain good characteristics of those bipolar devices. Since these triply charged ions are not abundantly present in the ion source of our implanter, implantation may take quite some hours at maximum power to obtain a certain dose of phosphorus in the implanted

substrate. This process needs to be modified to improve the robustness of its bipolar part.

Besides, one important advantage of bipolar devices is their high frequency response compared to CMOS which has not been optimised yet in the UT-BiCMOS process. Hence, an improvement of the frequency behaviour of the bipolar transistor can be accomplished.

In this chapter we will present our results on the optimisation of the UT-BiCMOS process for the bipolar transistor fabrication. The process was optimised by means of simulation [3] and the bipolar transistors were fabricated at the MESA clean room, university of Twente. The SOD material was also used as a diffusion source for emitter doping in one of the variants. One of the most advantage of SOD technology is that there is no damage or contamination as compared to the implantation technology. As a result, better and shallower junctions can be obtained by SOD diffusion. The result of poly-buffered shallow junction formation using phosphorus diffusion from SOD source in chapter 3 has been applied here to form poly-emitter in one of the variants. In the next section, important characteristics of a bipolar transistor will be discussed and structure requirements to obtain a good transistor will be highlighted.

4.2. Transistor characteristics and requirements

A bipolar transistor consists of three regions of semiconducting material named respectively emitter, base and collector, which are separated by two p-n junctions, as illustrated in figure 4-4 for an npn device.

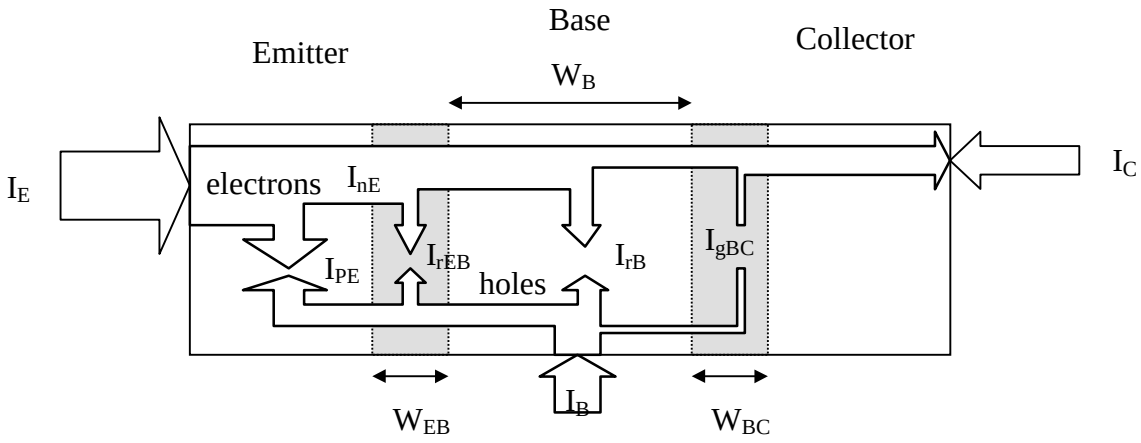


Figure 4-4: Current flow in a bipolar transistor

Bipolar transistor is often used as a current amplifier and the current gain β is defined as the ratio of the output collector current I_C and the input base current I_B. Figure 4-4 shows the components of all the currents flowing in a bipolar transistor.

In order to have a maximum current gain a transistor needs a heavily doped emitter and a very thin base. However, as the base thickness is decreased, the base concentration must be increased to prevent too high base resistance. Thus, a compromise has to be made between the current gain and the base resistance. Improvement of transistor current gain can be obtained using a polysilicon emitter structure.

4.2.1. Poly-emitter bipolar transistor

The implementation of polysilicon to form the emitter contact has had a major impact on the transistor current gain. In poly-emitter bipolar transistors a heavily doped

polysilicon layer is used as a buffer layer between the metal contact and the mono-crystalline Si emitter region. This poly layer can improve the current gain because of the lower base current than in the case of direct metal mono-silicon emitter contact. In the metal-contacted emitter, the effective recombination velocity (s_o) is very high, so that the base current rises with decreasing emitter depth. Lower s_o values are found experimentally for emitter diffused from polysilicon [4]. Consequently, the hole gradient in the emitter region is lower and the associated base current is reduced.

In addition, a very thin SiO_2 layer is often unintentionally formed at the poly/mono Si interface before polysilicon deposition. This thin oxide layer of 1-2 nm, depending on the technology, acts as a barrier layer for holes coming from the base into the emitter region. Furthermore, the segregation effect of donor impurities leads to an impurity pileup and band bending at the interface. Hence, it also provides a barrier for hole transport out of the single crystalline region, resulting in current gain improvement.

Using polysilicon, thin emitters can be formed with acceptable current gain. Thin emitters limit hole storage and allow thin base regions to be formed with adequate control and with higher base doping. From the technology point of view, the poly-emitter can effectively prevent the spiking effect of the metal, which often causes shortcut between emitter and collector regions. With all of these advantages, the polysilicon emitter formation becomes the preferred technology for bipolar transistor fabrication.

As mentioned in the first section of this chapter, optimisation of the collector implantation is necessary to obtain a device with good characteristics. Besides the cut-off frequency, the base sheet resistance and the current gain, the most important device characteristics are related to the collector-base junction, for instance breakdown voltage, Early voltage, and collector base capacitance. In the following subsections these parameters will be discussed and their relation to the collector-base junction will be made clear.

4.2.2. Breakdown voltage (BV_{CE0})

When the collector-base voltage is raised above a critical value, extremely large collector currents can be observed, even in the absence of the base current. This critical voltage is called the breakdown voltage, and determines the maximum allowable supply voltage since the transistor will not function properly above this collector-base voltage, and may even become severely damaged by high collector currents.

To obtain a breakdown voltage BV_{CE0} that is as high as possible, the concentration of the collector dopant must be as low as possible to obtain a thick collector depletion region. In that case the junction reverse bias results in a minimum electric field, and thus avalanche multiplication is minimised.

4.2.3. Early voltage (V_A)

Unfortunately, the current gain is dependent on the collector-base voltage. Due to reduction of the quasi-neutral base width by an increase of the collector-base reverse bias, the minority carrier gradient in the base is increased, resulting in an increase of the collector current. Thus in the active region, the collector current increases slightly with the collector-base voltage. This effect can be characterised by the Early voltage, by extrapolating the output characteristics to the negative voltage axis. The distance from the origin to the intersection with this axis is referred to as the Early voltage. A high Early voltage thus indicates that the collector current is only weakly affected by the collector-base voltage in the active region.

To obtain a maximum Early voltage, the collector concentration at the collector base junction must be as low as possible. The variation of the base depletion layer will be minimal in that case, resulting in a minimum variation of the minority carrier gradient.

4.2.4. Collector-base junction capacitance (C_{JBC})

The collector-base capacitance is very important in determining the AC performance of a bipolar transistor. When the base area is large, the collector-base capacitance can be very large. Moreover, it is seen at the input of the transistor as a capacitance which is about $(1+g_mR_L)$ times as large as the collector-base capacitance due to the Miller effect [5], with $g_m=\delta I_C/\delta V_{BE}$, and R_L is the load of the transistor. To obtain a high speed bipolar transistor, the collector-base capacitance has therefore to be minimised. Besides by reduction of the base area, this can be achieved by reducing the capacitance per area. A reduction of the capacitance per area also reduces the capacitance at the perimeter of a junction, which is becoming more important as device areas are scaled down. The capacitance per area can be reduced by minimising the collector concentration in the vicinity of the junction to obtain a maximum depletion layer width in the collector, and therefore a minimum collector-base capacitance.

4.2.5. High frequency performance

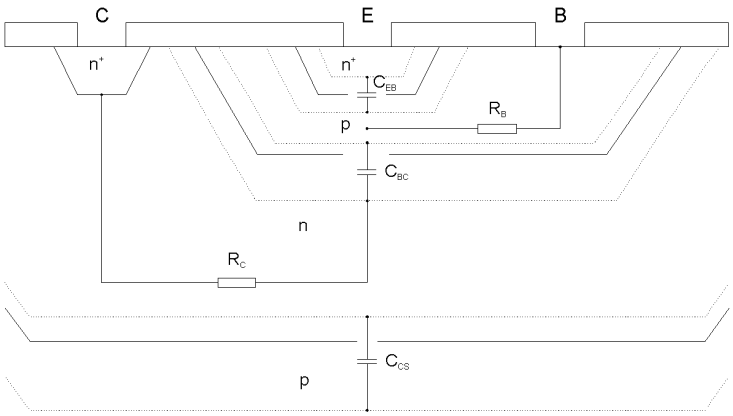


Figure 4-5: Cross-section view of a bipolar transistor including parasitic resistances and capacitances

To study high frequency behaviour of the bipolar transistor, a cross-sectional view of the transistor is given in figure 4-5, including its parasitic resistances and capacitances. For small-signal applications the model in figure 4-5 can be linearized to the small-signal hybrid-π model given in figure 4-6 [6].

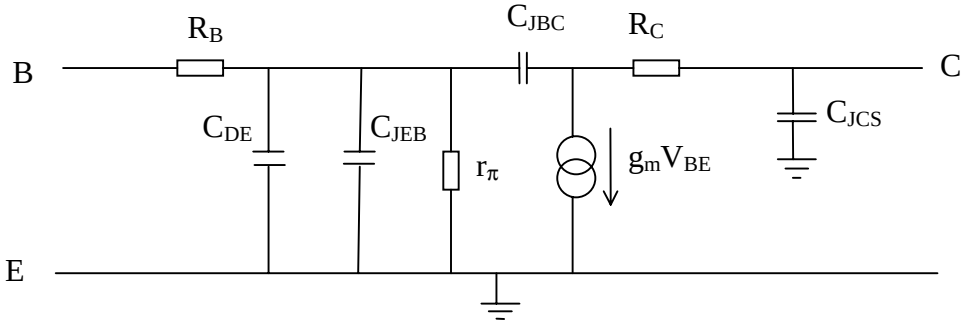


Figure 4-6: Small-signal hybrid-π model of a bipolar transistor

In this figure, the EB junction capacitances have been subdivided in a diffusion (C_{DE}) and the depletion capacitance (C_{jEB}). From this model the cut-off frequency of the transistor can be determined [7]. It is defined as the frequency at which the common emitter short-circuited load, small-signal current gain is unity. It can be written as a function of the device parameters and collector current:

$$f_T = \frac{1}{2\pi}\left[\frac{kT}{qI_C}\left(C_{jEB} + C_{jBC}\right) + \tau_F + R_C C_{jBC} + \frac{R_C C_{jCS}}{\beta_o}\right]^{-1}$$

where τ_F is the forward transit time, and β_o is the low-frequency current gain.

The cut-off frequency has been widely used as a figure of merit for the AC performance of the bipolar transistor, however it only reflects the specific situation of a short-circuited transistor. Therefore, the maximum oscillation frequency f_{MAX} for which the power gain is unity may be a more suited parameter for circuit applications:

$$f_{MAX} = \sqrt{\frac{f_T}{8\pi C_{jBC} R_B}}$$

From this equation it appears that besides a high cut-off frequency, a low collector-base junction capacitance and low base resistance are required to obtain a high speed transistor.

4.2.6. Summary

Table 4-2: Consequences of modifications of the base and collector implantations on device parameters

	High base dose	Low base implant energy	High collector dose	Low collector dose	Rapid thermal annealing
Cut-off frequency (f_T)	- (↓)	++ (↑)	- (↓)	+ (↑)	+ (↑)
Break-down Voltage (BV_{CE0})	+ (↑)		- (↓)	+ (↑)	+ (↑)
Current gain (β)	-- (↓)	+ (↑)			+ (↑)
Early voltage (V_A)	++ (↑)	- (↓)	- (↓)	+ (↑)	+ (↑)
Base-collector junction capacitance (C_{jBC})	- (↑)		- (↑)	+ (↓)	
Base sheet resistance (R_B)	++ (↓)	- (↑)			
Collector sheet resistance (R_C)			++ (↓)	-- (↓)	
Punch through	++ (↓)	- (↑)		+ (↓)	
Transient Enhance Diffusion	- (↑)	- (↑)	- (↑)	+ (↓)	+ (↓)
Implantation energy		≥ 40 keV		≤ 1350 keV	

This section states some implications which have to be taken into account for process optimisations. It may be clear that some different goals in the area of device characteristics are reached by actions which trade-off each other. Therefore none of the intended device characteristics can be fully satisfied, hence compromises have to be made. Table 4-2 illustrates that an improvement of a specific device parameter may lead to deterioration of one or more of the other device parameters. A plus means that a device parameter is improved by the action stated in the first row of the table, and a minus sign means that the device parameter is deteriorated. Arrow up or down means its value increases or decreases. Besides device parameters the limitations of the implanter of the MESA laboratory are shown.

The effects of the modifications of the base and collector processing on the bipolar device parameters are discussed quantitatively in the next section.

4.3. Optimisation of UT- bipolar device

This section presents the results of our simulations to optimise the bipolar device using the UT-BiCMOS process without altering the CMOS transistors. There are three variants with the same goal in order to find a solution for a better collector implantation process and higher cut-off frequency. Variant 1 was proposed to change only a minimum of the process flow from the UT-BiCMOS process in order to get a stable and reproducible process. Variant 2 proposed to use rapid thermal processing for a minimum thermal budget, resulting in devices with more attractive characteristics as shown in table 4-2, while the process order is kept the same. Variant 3 was proposed to be similar as variant 2 but use SOD as a diffusion source for emitter doping. Therefore, there is a small change in the process flow to keep the total thermal budget as low as possible. Figure 4-7 describes the process flows of the three variants using the UT-BiCMOS process.

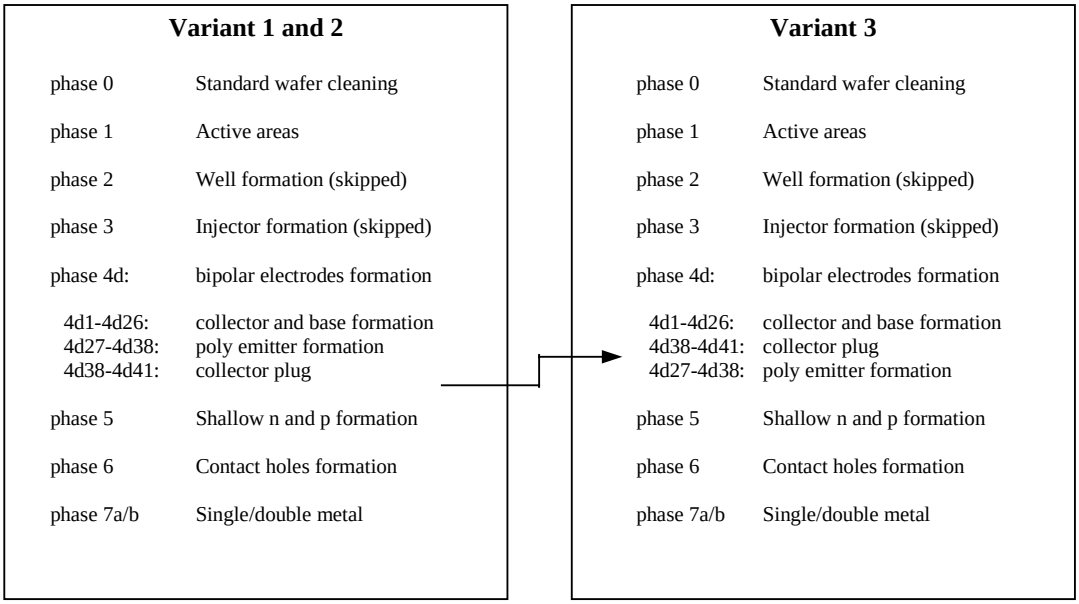


Figure 4-7: Process flows of three variants

4.3.1. Simulation calibration

The simulations were carried out using TSUPREM4 [8] for process simulation and MEDICI [9] for device simulation. To obtain simulation results which predict the device characteristics as good as possible, it is very important to select the correct

models and to adjust default parameters of the model to values which are obtained from experiments. Two simulators are needed to characterize the devices which are made in a certain process: a process simulator and a device simulator. In this work respectively TSUPREM4 and MEDICI are used.

4.3.1.1. Calibration of process simulation

Damage caused by implantation creates an excess of vacancies and interstitials in the substrate. The amount of damage depends on implant species, energy, dose, wafer temperature and orientation, dose rate and materials covering the substrate. During subsequent anneal steps, interstitials and vacancies recombine or diffuse into the substrate, while the excess interstitials must diffuse away to find a recombination site. Dopant atoms that were paired with those interstitials diffuse much faster, resulting in an unwanted broadening of the impurity profile. This effect is called transient enhanced diffusion (TED). The rapid thermal process is very effective for TED suppression.

In process simulation, the diffusion model needs to be calibrated for a specific process and equipment. Transient enhanced diffusions were calibrated for both conventional furnace and rapid thermal annealing.

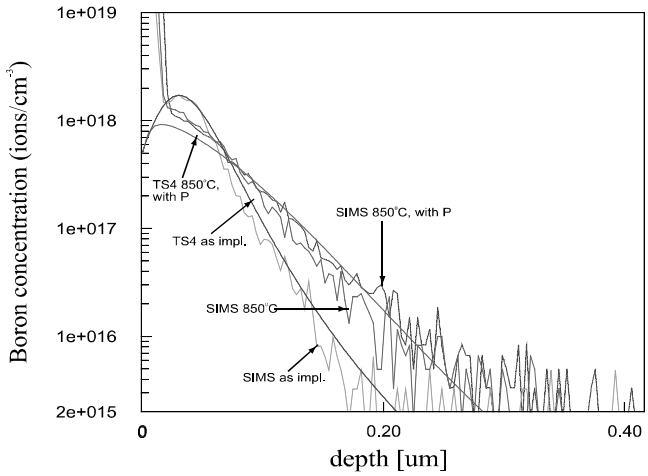


Figure 4-8: Boron SIMS and simulated profiles to calibrate process simulations with respect to transient enhanced diffusion.

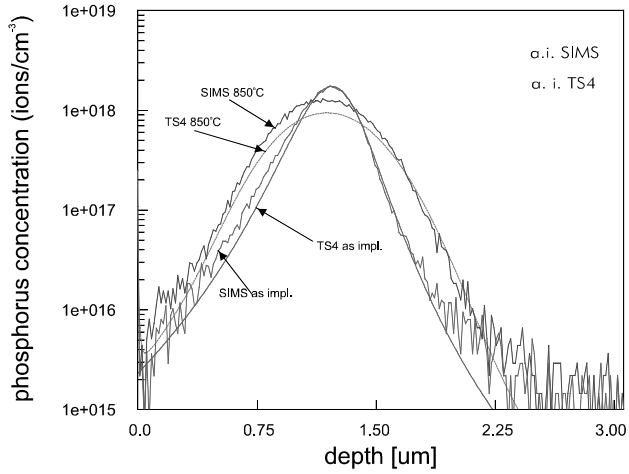


Figure 4-9: Phosphorus SIMS and simulated profiles to calibrate process simulations with respect to transient enhanced diffusion

The transient enhanced diffusion for the collector and base implantation was studied with SIMS profiles. Unexpectedly, it turned out from the experiment that substrate

damage caused by the phosphorus implantation does hardly result in any enhanced boron diffusion near the surface. This does not indicate that the amount of substrate damage is negligible, but that an excess concentration of vacancies may exist in the region of the substrate that contains boron, which suppresses the boron diffusion. The following figures show the profiles of boron and phosphorus for the TED calibration of the base and the collector implantation [3].

Calibration of the process simulator has been performed by selecting the plus one model for implantation damage and the PD.FULL model for the diffusion of point defects. The plus one model assumes that one interstitial is added to the interstitial concentration for each implanted ion. It turned out that diffused profiles of the simulations showed excellent agreement with the SIMS profiles as shown in figure 4-8 and figure 4-9, if the implantation moments were adjusted to the as implanted SIMS profiles and with the enhancement of the default phosphorus diffusivity by a factor 2.5. Again it must be emphasized that it is not intended to derive a complete diffusion model, the experiment is only meant to calibrate the simulation of the anneal step subsequent to the base and collector implantation. The moments used for simulation of the base and collector implantation can be found in reference [3].

4.3.1.2. Calibration of device simulation

What remains is optimal performance of the device simulator. Poly emitter effects have to be included. Additionally, care has to be taken to select the correct models for recombination, lifetime, (minority) mobility, and bandgap narrowing.

The mobility model used in MEDICI is the Philips unified mobility model (PHUMOB) [10,11]. This model takes different mobilities into account for majority and minority carriers, and is therefore very suitable for device simulations with bipolar devices. The parameters of the bandgap narrowing model have to be carefully adjusted to the mobility model, therefore these are chosen according to reference [12]. For recombination, the Shockley-Read-Hall and the Auger models are enabled, and the lifetime is changed according to values found by Slotboom [13]. Finally it must be mentioned that the impact ionization model is enabled with default coefficients. The impact ionization model may predict breakdown voltages which are slightly lower than measured values. This is caused by the fact that MEDICI uses a local impact ionization model while a non-local model would be more realistic. The local model assumes the carrier temperature to be increased at once from the lattice temperature T_L to the field temperature T_E for each carrier that enters the collector-base depletion region. In practice however, this happens gradually, and many carriers do not even reach T_E as they have traveled through the depletion region before they gained that much energy. Thus the generation of impact ionization is modeled somewhat too strong, especially for very narrow collector-base depletion layers. In this work however, the impact ionization model is assumed to calculate the breakdown voltage with sufficient accuracy.

4.3.2. Simulation results

4.3.2.1. Variant 1 – modified UT-BiCMOS process

The process flow of variant 1 consist of the same sequence of process steps as the original UT-BiCMOS process which only uses conventional furnace for every annealing step. The major differences between variant 1 and the UT-BiCMOS process are the base and collector implantation steps from which doses, energies and species are changed. To improve device manufacturing, the maximum accelerator voltage is adjusted at 450 kV for P^{3+}. Doubly charged phosphorus P^{2+} is still allowed at 500 kV because the implantation cycle takes much less time for P^{2+}, and implantation at 500 kV during a

short period does not result in equipment breakdown. For the base implantation BF_2 will be used instead of B, to obtain a shallow base, which results in higher cut-off frequency. A second advantage of a shallow base is that base and collector are effectively moved away from each other, resulting in a less critical junction with respect to transistor characteristics like breakdown voltage, Early voltage and collector-base junction capacitance.

The process flow of variant 1 is presented in appendix 2. The simulations are proposed to give a process window for the fabrication of the bipolar transistor. Cut-off frequency, current gain, breakdown voltage, Early voltage, collector-base junction capacitance, and base and collector sheet resistance are calculated for the range of base and collector implantation doses. The collector implantation energies used are 1000 keV and 1350 keV to investigate both P^{2+} and P^{3+} as collector dopant species. Simulation of the cut-off frequency and the breakdown voltage have been performed using a one and two-dimensional transistor profile for the most favourable transistors.

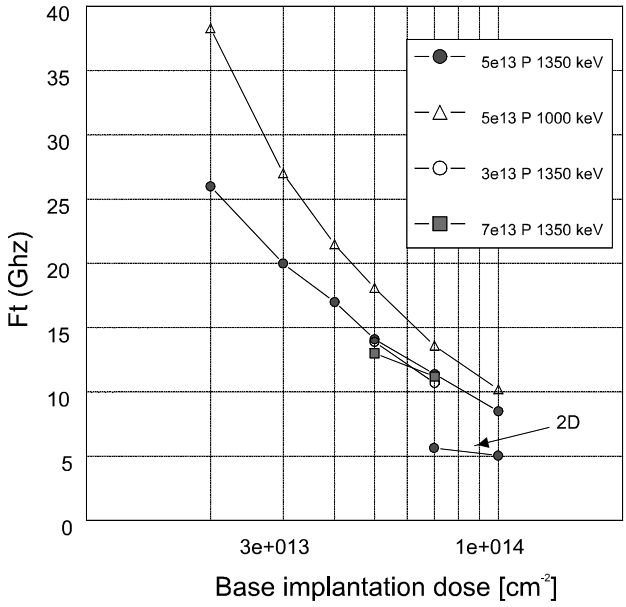


Figure 4-10: Simulated cut-off frequency versus base dose for several collector implantations (variant 1)

The results for the simulations of the cut-off frequency are presented in figure 4-10. One-dimensional simulations presented in this figure show that quite high cut-off frequencies can be obtained for transistors with a low base dose and a low collector implantation energy, which is not very surprising as it yields transistors with a minimal base thickness. However, two-dimensional simulations show that one-dimensional simulations are much too optimistic. A more realistic value of the cut-off frequency according to two-dimensional simulations is about 5 GHz for transistors with a $5x\ 10^{13}$ cm^{-2} P^+ 1350 keV implanted collector. This reduction of the cut-off frequency is caused by the increase of the transistor area, which results in a much higher collector-base capacitance and a higher collector and base resistance for the two-dimensional case. Therefore, a reduction of the transistor area is needed to increase the cut-off frequency to the maximum value that can be reached for the processing conditions shown in figure 4-10. From one-dimensional simulations it may be concluded that the cut-off frequency is hardly affected by a change of the collector dose at base doses of 5 and $7.\ 10^{13}$ cm^{-2}, however a large transistor area requires a minimum collector-base capacitance, and thus a low collector dose.

From the simulation results of Early voltage and base resistance, base implantation with a dose of less than 5x 10^{13} cm^{-2} resulted in transistor with low Early voltage, i.e. less than 15 V, and high base sheet resistance, i.e. higher than 8 kΩ/sqr. In literature [14] a maximum base sheet resistance of about 6 kΩ/sqr is recommended for high speed bipolar transistors. For transistors fabricated according to variant 1, this means that a minimum base implantation dose of 7x 10^{13} cm^{-2} is required, which also corresponds with a current gain of approximately 100.

As turned out from the simulation of the breakdown voltage, a transistor with a 1000 keV P^{2+} implanted collector may not be recommended as its breakdown voltage is too low, even for a 3.3 V process. Besides, its collector-base junction capacitance is very high, and its Early voltage is too low, especially for low base doses. From the transistors with a 1350 keV P^{3+} implanted collector, the devices with a collector dose of 3x 10^{13} cm^{-2} suffer from a collector sheet resistance which is unacceptably high. The collector dose of 7x 10^{13} cm^{-2} or 1x 10^{14} cm^{-2} is recommended.

Thus, the following implantations are recommended for variant 1:

collector: 5x 10^{13} cm^{-2} P^{3+} at 450 kV

base: 7x 10^{13} cm^{-2} BF_2 at 40 kV

The doping profile of this transistor is presented in the following figure with a base dose of 7x 10^{13} cm^{-2}. This transistor is supposed to work properly at a 3.3 V supply voltage.

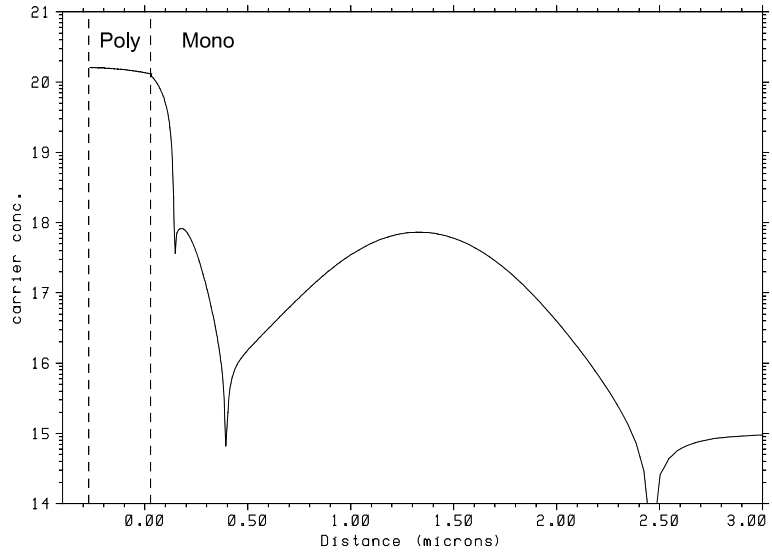


Figure 4-11: simulated net doping profile of variant 1

4.3.2.2. Variant 2 – RTP implementation

Variant 2 is based on a minimal diffusion of the base dopant, which is expected to result in a higher cut-off frequency due to a reduction of the quasi-neutral base width, and an improvement of breakdown voltage, Early voltage and collector-base junction capacitance due to an increased space between base and collector profile. To reduce the thermal anneal budget without affecting the CMOS devices, the bipolar process steps are carried out in a later stage of the process flow and used rapid thermal annealing instead of conventional furnace annealing in many processing steps. This results in minimal diffusion of the base and collector dopant. The process flow of variant 2 is presented in appendix 3. Simulation results of the cut-off frequency are presented in figure 4-12.

Figure 4-12 shows that cut-off frequencies in the range of 12-30 GHz can be achieved by variant 2 according to one-dimensional simulations. However, two-dimensional simulations result in cut-off frequencies of about 10 GHz for both P^{2+} and P^{3+} collector. The cut-off frequency as calculated by two-dimensional simulations turns out to be in the same range for transistors with a P^{2+} and a P^{3+} due to the high collector-base capacitance per area of the transistors with a P^{2+} collector. One-dimensional simulations show that a variation of the collector dose results in a variation of the cut-off frequency of roughly 2-3 GHz in this range of base dose. However, as a higher collector dose results in a higher collector-base capacitance per area, it is expected for larger transistor areas that cut-off frequencies for higher collector doses are in the same range or may be even lower as was found by two-dimensional simulations for the transistor with a 5x 10^{13} cm^{-2} P^{3+} implanted collector.

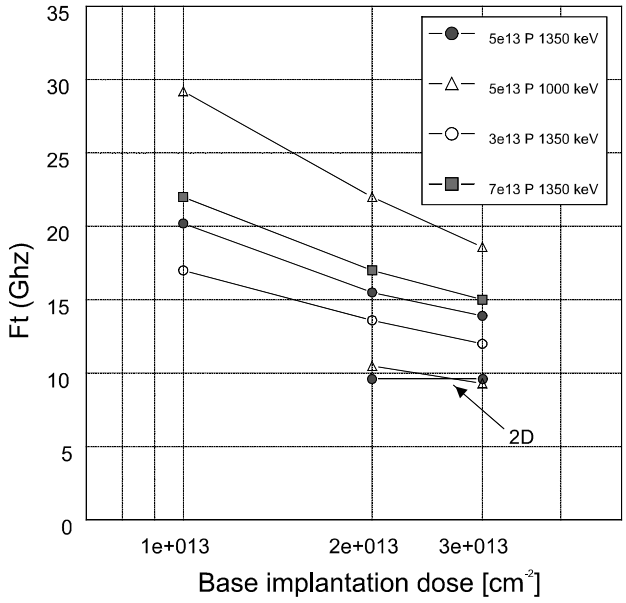
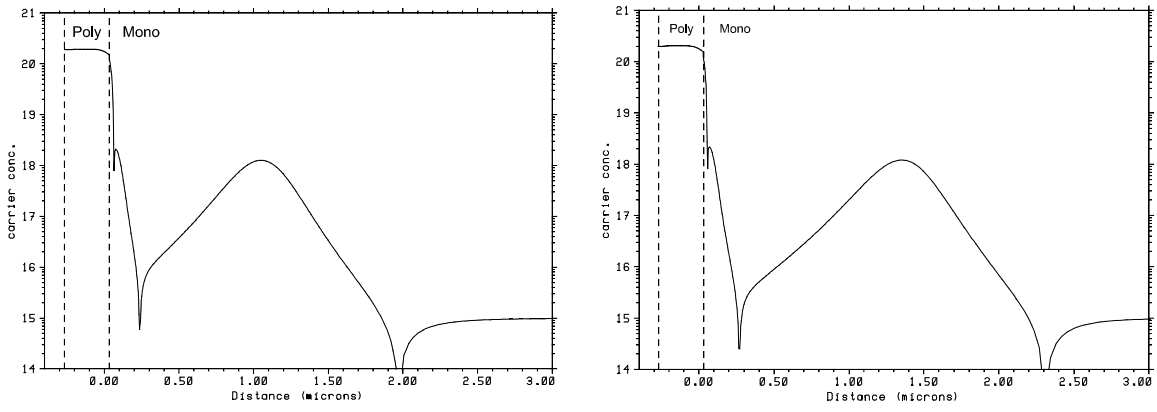


Figure 4-12: Simulated cut-off frequency versus base dose for several collector implantations (variant 2)



a) b)

Figure 4-13: simulated net doping profile of variant 2 with collector implantation of a) P^{2+} ions at 500 keV and b) P^{3+} ions at 450 keV

According to the simulations of the breakdown voltage, a 500 kV P^{2+} as well as 450 kV P^{3+} collector implantation may be used for 3.3 V and 5 V device respectively. The former is only recommended if P^{2+} ions have to be used for the collector implantation. The collector dose may be 5 or 7x 10^{13} cm^{-2} for a 450 kV P^{3+} implantation and

5x 10^{13} cm^{-2} for 500 kV P^{2+} implantation. The minimum required base dose to obtain an acceptable base resistance and sufficient current gain, is 2x 10^{13} cm^{-2}.

Thus:

collector: 5 or 7x 10^{13} cm^{-2} P^{3+} 450 kV

base: 2 or 3x 10^{13} cm^{-2} BF_2 40 kV

or for doubly charged phosphorus:

collector: 5x 10^{13} cm^{-2} P^{2+} 500 kV

base: 2 or 3x 10^{13} cm^{-2} BF_2 40 kV

The doping profile of this transistor is presented in figure 4-13 with a collector dose of 5x10^{13} cm^{-2}, and a base dose of 2x 10^{13} cm^{-2} for a 500 kV P^{2+} and a 450 kV P^{3+} implantation respectively.

4.3.2.3. Variant 3 – poly-emitter doping by SOD

Variant 3 used rapid thermal processing for most of the annealing steps as in variant 2. The base and collector implantation are the same as in variant 2. The only difference is that the emitter is doped by SOD source containing phosphorus instead of arsenic implantation. Therefore some process steps are changed in order to dope the emitter at the end of the process flow for a minimum thermal budget.

Figure 4-14 shows the simulated net doping profile of a bipolar transistor with emitter doped by SOD technique. Because the emitter is doped with phosphorus in this variant, the polysilicon emitter has lower doping concentration and a little deeper junction compared with the profiles of variant 2 transistors using arsenic emitter doping. As a result, the base is narrower and a higher base resistance is obtained. However, the base collector junction is not affected by the emitter doping and the high frequency performance of this transistor should be similar to the transistors of variant 2.

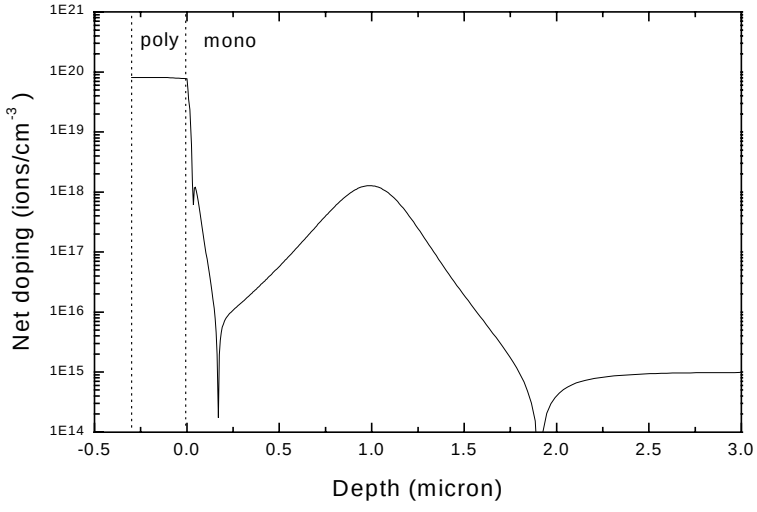


Figure 4-14: Simulated net doping profile of variant 3

Indeed, one-dimension simulation results also showed that slightly higher cut-off frequency, higher base resistance, higher current gain, and lower Early voltage were obtained for this transistor compared with variant 2 ones. This simulation result will be confirmed by measurement results of realised transistors in the device characterisation section.

4.3.3. Summary

Table 5-3 resumes the simulation results of devices made from the two first variants.

Table 5-3 Summary of simulated device characteristics

	variant 1	variant 2	
	5x 10^{13} cm^{-2} P^{3+} 450 kV	5x 10^{13} cm^{-2} P^{2+} 500 kV	5x 10^{13} cm^{-2} P^{3+} 450 kV
	7x 10^{13} cm^{-2} BF_2 40 kV	2x 10^{13} cm^{-2} BF_2 40 kV	
f_T (GHz)	5.6	10.6	9.6
BV_{CE0} (V)	5.4	5.3	7.7
β	120	135	130
V_{CC} (V)	3.3	3.3	5
V_{AF} (V)	32	32	45
C_{jBC} (nF/cm^2)	40	39	28
R_B (kΩ/sqr.)	4.5	5.0	5.0
R_c (Ω/sqr.)	300	360	370

From this table it can be concluded that devices for a 5 V process can only be fabricated according to variant 2, using P^{3+} as collector dopant. If it is preferred to use P^{2+} as collector dopant, a device for a 3.3 V process can be fabricated according to variant 2. It yields a device with a lower Early voltage and a higher collector-base junction capacitance and a comparable cut-off frequency. Finally, when changes to the present UT BiCMOS process must be limited to an absolute minimum, a device can be fabricated according to variant 1. Transistors of this type can only be used for 3.3 V applications. They have higher Early voltage and collector-base junction capacitance, their cut-off frequency is somewhat lower. If higher cut-off frequencies are required than could be obtained by variant 2, it is recommended to decrease the transistor area, for instance by the application of polysilicon as contact material for the base. By saving transistor area the cut-off frequency can be increased to a maximum value of 22 GHz and 15.5 GHz for transistors fabricated according to variant 2 with P^{2+} and P^{3+} as collector dopant respectively.

Rapid thermal diffusion from SOD source for shallow poly-emitter formation has the advantages of being defect-free and no TED. This technique is implemented in variant 3. Better device characteristics are expected for this variant over the others.

The next section will present process details of bipolar transistors realised according to the three variants.

4.4. Fabrication of the bipolar transistor

4.4.1. Experimental details

Bipolar transistors were realised according to the three simulated variants. The details of the process flows are described in the appendix 2, 3 and 4. The standard UT-BiCMOS process includes eight phases. However, in this run only steps relevant to the bipolar part were conducted. The starting material for our NPN bipolar transistors are p-type <100> epitaxial silicon wafers with resistivity of 5-10 Ωcm which corresponding to a boron doping concentration in the order of 10^{15} ions/cm^3. All three variants have the same phase zero and phase one for standard cleaning and for active areas definition. The phases from 2 to 4c of the UT-BiCMOS process are skipped because they are not for the

bipolar part. Phase 4d is continued to make collector, base and emitter of the transistors. Phase 5 is supposed to make heavily doped shallow layer around the base for good ohmic contact. The steps in these two phases are different from variants and will be described in detail in the next paragraph. The phase 6 and phase 7 are designed to make contact holes and metalisation. The last two phases are almost the same for all variants. In the following we will describe the details of the technology used in every variant and their differences.

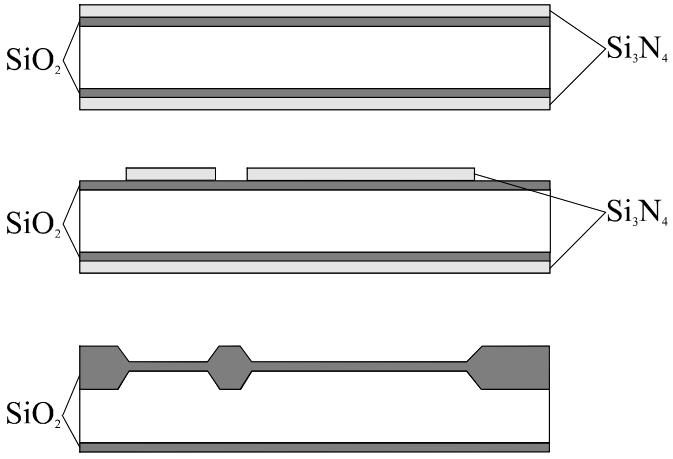


Figure 4-15: LOCOS formation of BJT

4.4.1.1. Variant 1

The active areas are defined by standard LOCOS (LoCal Oxidation of Silicon) in phase 1 (steps 1.1 to 1.12). These active areas are opened for transistors and other devices and characterisation structures (figure 4-15). The advantage of this LOCOS technique is a reduction in step heights between the active areas and the field oxide resulting in an improved reliability of the interconnects.

In phase 4d the collector, base and poly-Si emitter are formed (figure 4-16 and figure 4-17). First, very high energy implantation of phosphorus (1350 keV) were carried out to form the collector. The collector was implanted and annealed in two steps (half of total dose for each time) in order to reduce crystal defects near the Si wafer surface due to too high dose and high energy implantation. In this variant the collector is implanted with triply charged phosphorus ions P^{3+} at an energy of 450 kV. However, because the P^{3+} ions are not so abundant, the collector implantation took more than 10 hours to reach the dose of 2.5×10^{13} ions/cm^2. The implanter was broken few times during the process of variant 1. Therefore we decided to used doubly charged phosphorus ions P^{2+} with implantation energy of 500 kV for collector of the next variants. The base was implanted with BF_2^+ ions immediately after the last collector implantation. Photoresist was patterned by photolithography and used as a mask for all the implantation steps. After every implantation step the Si crystal is damaged and has to be annealed to remove these damages and to activate the implanted ions. In variant 1 we only used conventional furnace for annealing. The advanced rapid thermal annealing was used in variants 2 and 3.

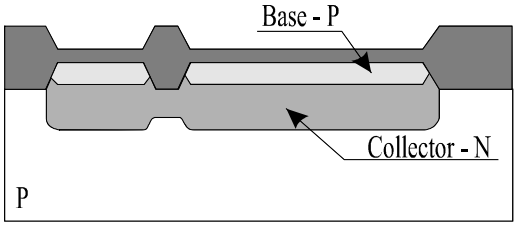


Figure 4-16: Base and Collector formation

The emitter area was then opened and the Si wafer was deposited with 300 nm of polysilicon by LPCVD (Low Pressure Chemical Vapour Deposition). During deposition the furnace was kept at 610^{o}C and at a pressure of 250 mTorr with SiH_4 flow. The thin native oxide is removed just before the polysilicon deposition step by diluted HF 1 %. However, there is always a thin oxide layer grown in between the poly-Si and the Si

substrate because the etched wafers are often exposed to the air for few minutes before they can be in the neutral ambient of the LPCVD furnace. Further more, some air may enter the furnace when opened. This might cause oxidation when the wafers enter the hot zone of the furnace and about 2 nm of oxide is grown. This native oxide combined with the polysilicon emitter can change the transistor operation because this poly-Si emitter can block the hole flow from the base to the emitter and increase the current gain. However, control of this native oxide thickness is difficult because we do not have a LPCVD furnace which can controllably remove or grow a certain oxide layer on Si wafers in the furnace just before any deposition. The polysilicon emitter was then implanted with a very high dose of As^+ ($6x10^{15}$ ions/cm^2) and low energy (100 kV) so all the As^+ ions can not penetrate through the 300 nm poly-Si emitter. During the annealing in the next steps these As^+ diffuse deeper through the poly-Si layer and into the Si substrate to form the emitter of the transistor.

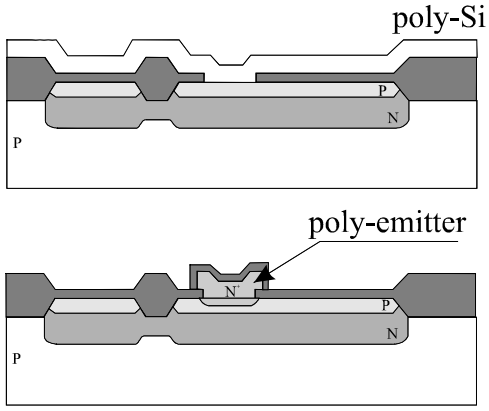


Figure 4-17: Poly-emitter formation

Definition and phosphorus implantation of collector plug was then conducted. This collector plug is designed to make an ohmic contact with aluminium. In the fifth phase the base was implanted with BF_2 to form a heavily doped shallow layer for ohmic contact. The next phase was then performed to open the contact holes. Before etching of the contact holes, the patterned wafer was planarised by BPSG (Boron Phosphorus Silicate Glass) deposition by APCVD (Atmosphere Pressure CVD). This planarisation step was added to improve the reliability of the subsequent interconnects. This BPSG layer also can getter metallic ions during high temperature annealing steps.

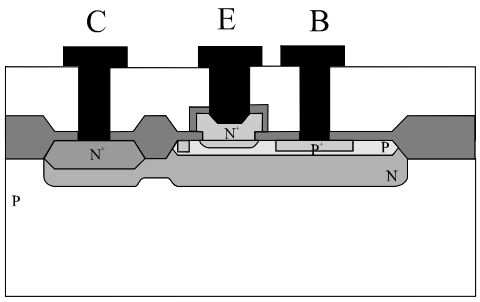


Figure 4-18: Final structure of a BJT

The surface state annealing was carried out at 500^{o}C in wet nitrogen ambient. This step can improve the quality of the oxide/Si interface as well as the poly-Si/Si interface. During annealing hydrogen atoms were created from water and diffused into the Si wafer and fill dangling bonds at the interfaces which cause charged states. The metalisation step was performed in phase seven. After removing the native oxide a layer of 70 nm Ti/W and 1 μm aluminium containing 1 % Si was sputtered. Patterning of the contact pads was carried out by lithography and dry etching. The final BJT structure was obtained as illustrated in figure 4-18. Wafers were then annealed at 400^{o}C in wet nitrogen ambient to sinter the Al layer for good ohmic contacts. Finally, electrical characterisations were carried out.

4.4.1.2. Variant 2

Variant 2 was designed for better performance transistor. The main difference of this variant compared with variant 1 is the large reduction of the thermal budget because Rapid Thermal Annealing (RTA) is used. Therefore, the diffusion of implanted ions is minimised leading to implanted doping profile being maintained almost unchanged.

Further advantage of RTA is that the Si wafer could be annealed at high temperatures of more than 1050^{o}C for very short time. Therefore, most of the implantation damages could be removed and implanted ions are more activated at high concentrations. As mentioned in variant 1 we had to change the collector implantation to doubly charged phosphorus ions P^{2+} because of the implanter breakdown problem. We decided to fabricate the transistor according to the solution for 3 V supply voltages of variant 2 in which the collector is implanted with P^{2+} at 500 kV and a dose of 5.10^{13} cm^{-2}. In this variant, step 6.5.a was inserted for flowing and gettering of the BPSG layer, for drive-in diffusion of poly-emitter and also for removal of all the implantation damages in the previous steps. The rest of the process was kept the same as in variant 1.

4.4.1.3. Variant 3

This variant is similar to variant 2 with only one different step namely the poly-emitter is doped by SOD source. This difference led to some changes in the process flow in order to keep the low thermal budget design as shown in figure 4-7. In this variant the doping step of the emitter was carried out at the end after all implantation steps for collector, base, collector plug and shallow p formation. There were only two annealing steps just after the collector implantations. During drive-in diffusion of SOD into the poly-emitter the wafer was also annealed and the damage caused by all the previous implantation steps was removed. This diffusion step was conducted with the rapid thermal process at 1050^{o}C for 20 seconds in nitrogen ambient. The diffusion source was P8545 (Aligned Signal Inc.) containing a very high concentration of phosphorus. The poly oxidation step was also inserted in order to oxidise the poly emitter for passivation and also for thickening the oxide on top of the base as well as for further annealing of the Si wafer. The contact hole formation phase and the metalisation phase were the same except that the step for 100 nm SiO_2 annealing and the BPSG flowing step were carried out at a temperature of 700^{o}C in order to minimise possible diffusions.

4.4.1.4. Transistor layout

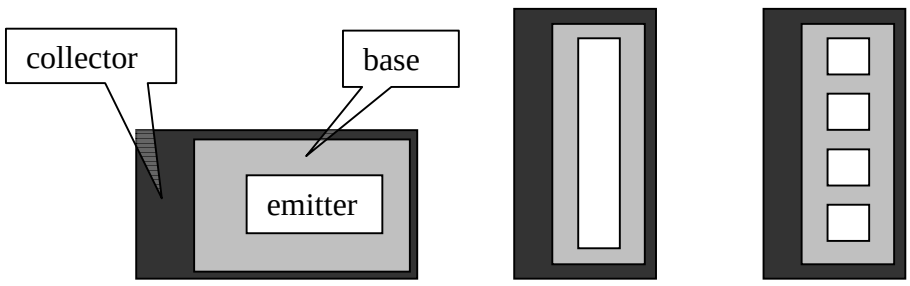


Figure 4-19: Top view of different transistors shapes: a) wide emitter b) narrow emitter c) large peripheral emitter

Transistor dimensions and shape are also important parameters for its performance. Different kinds of transistor were designed to characterise influence of active area size and shape. There are three kinds of shape which have different structures as shown in figure 4-19. The shape of transistor determines its electrodes' internal resistances. The narrow emitter transistor in figure 4-19b can have less collector resistance than the wide emitter transistor. Modern high frequency transistors also have very narrow emitter stripes of submicron for parasitic resistance and capacitance reduction. Our smallest transistor still has emitter area of 10x30 μm^2 because of the limitation of our photo-

lithography and misalignment tolerance of about 2 μm. Furthermore, our transistor has not self-aligned base. Improved design with self-aligned poly-base is recommended for a very small dimensions transistor. Each kind has three emitter sizes and they are named as A1, A2, etc. as listed in the following table.

Table 4: Sizes of emitter types

Wide emitter (μm^2)			Narrow emitter (μm^2)			Large peripheral emitter (μm^2)		
A1	A2	A3	B1	B2	B3	C1	C2	C3
10x30	20x60	40x120	30x110	60x220	120x440	4x10x30	4x20x60	4x40x120

Beside transistor structures, many structures were also designed for characterisation of electrodes' sheet resistance, capacitance and contact resistance. Those structures include Van der Pauw and Kenvin structures which are very basic for every semiconductor process.

The devices were then characterised by the precision semiconductor analyser HP4156. The electrical characteristics of the devices will be presented in the following section.

4.4.2. Device characterisations

4.4.2.1. Spreading resistance profiling

The doping profile is the most important factor which determines the device characteristics. Profiles could be measured by several techniques such as Secondary Ion Mass Spectroscopy (SIMS) for chemical profiling or Spreading Resistance Profiling (SRP) for active ion concentration profile. A Si sample is implanted with a certain dose of ions and then it has to be annealed for damage removal and activation of the implanted ions. In fact, the ion activation process is the process which uses the thermal energy to rearrange the implanted ions and to move them into lattice site positions. The degree of activation is dependent on the implantation dose and on the thermal budget. The higher the annealing temperature the higher the activated ion concentration. By measuring the spreading resistance of the bevelled samples the SRP measurement provides information about these free carrier concentrations. Figure 4-20 shows the SRP profiles of the transistors fabricated according to the three variants.

In this figure the first 0.3 μm is the poly-Si emitter, the rest is the mono-crystalline Si substrate. It is clear that there are only two peaks in all the three above profiles. The peak which is deep into the sample obviously is the collector region. The peak near the surface must be the emitter together with the base. However, the base is too shallow due to low implantation energy and the SRP resolution in these profiles is not good enough so that the junction between the emitter and the base could not be observed.

For variant 1 the junction depth of the collector is about 2.3 μm (the dashed line) while in variant 2 and 3 it is only about 2 μm (the solid lines). This is the result of the collector implantation by phosphorus ions with the energy of 1350 keV (P^{3+} 450 kV) for variant 1 and 1000 keV (P^{2+} 500 kV) for variant 2 and 3.

At the poly-emitter region the maximum of the net doping concentration is only about $1.5x10^{19}$ ions/cm^3 for variant 1, $4x10^{18}$ ions/cm^3 for variant 2 and $1x10^{19}$ ions/cm^3 for variant 3. The profile result of variant 2 seem not very correct compared with measurement result of poly-emitter sheet resistance which is 140 Ω/sqr and about the same as poly-emitter sheet resistance of 147 Ω/sqr for variant 3. Therefore, the net

doping concentration of poly-emitter should be higher for variant 2. The emitter is very deep in the Si substrate for variant 1 than two others. It is what we expected because the thermal budget is much higher in variant 1.

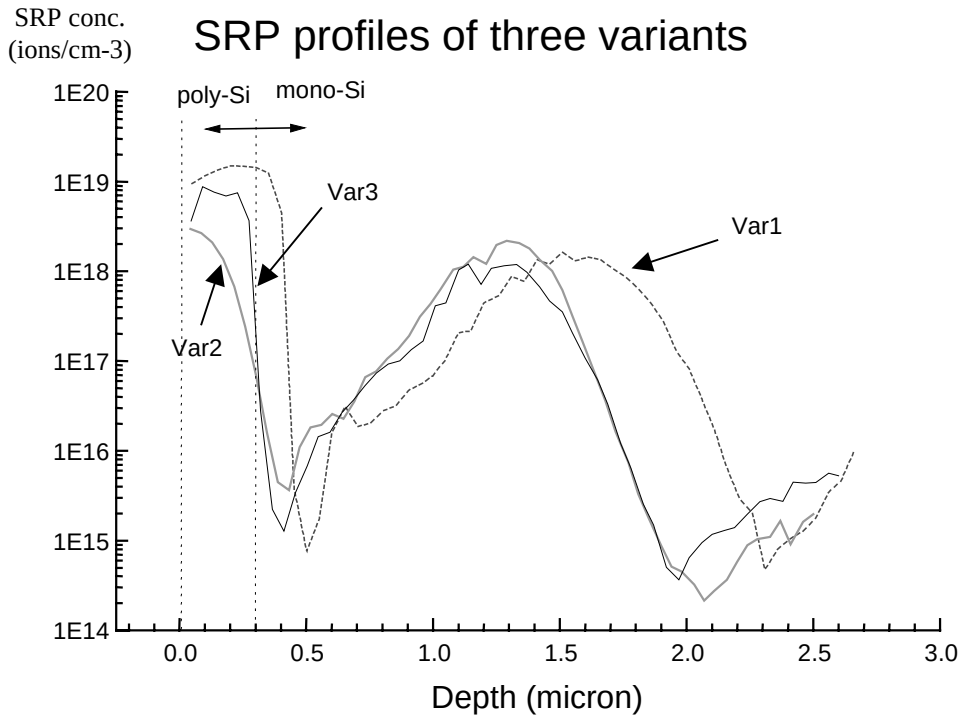


Figure 4-20: SRP profile of the bipolar transistor fabricated according to the three variant

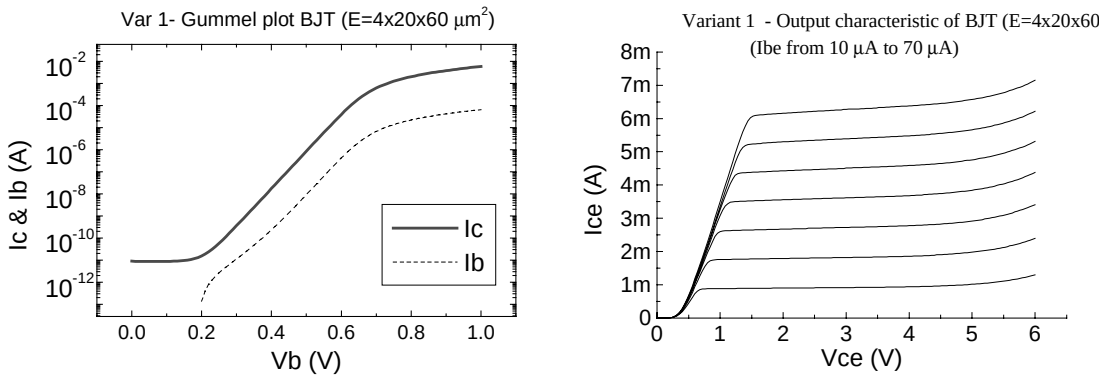
One of the important factors in the bipolar transistor is the capacitance between the base and the collector. One could see the base collector junction is widest for variant 3 followed by variant 2 and 1. It means that the spreading diffusion of ions from the implanted base and collector was least for variant 3. It is in agreement with the capacitance measurement which will be presented in the next section showing lowest base collector capacitance for variant 3. In fact, the thermal budget of variant 2 may be less than variant 3 but the spreading diffusion of the collector is larger for variant 2. It can be explained by the higher dose in the collector in variant 2 supposed that the SRP profiles in figure 4-19 are correct. In addition, in variant 2 the spreading diffusion from the base may be enhanced by the damage created from the emitter implantation while the emitter diffusion in variant 3 did not create any damage but annealed all the defects caused by previous implantations. There might be also the emitter-push effect [15] of phosphorus for variant 3 while it is not the case of arsenic for variant 2. This could explain why the base collector capacitance of the transistor fabricated from variant 3 is less than variant 2 and 1.

Electrical characteristics of fabricated transistors were characterised for all the transistor types as well as for all three variants. The following subsection will present measurement results of the transistors.

4.4.2.2. Electrical characteristics

4.2.4.2.1. Variant-1 BJT

Transconductance and output curves were measured for all the transistors. Good characteristics have been observed for some transistors. Figure 4-21 illustrates a Gummel plot and output curve of a transistor with large peripheral emitter (4x20x60 μm^2).



a) *b)*

Figure 4-21: a) Gummel and b) output characteristics of BJT (E=4x20x60 μm²) of variant 1

As seen in the above figure electrical properties of this transistor are quite good. The collector and base currents in the Gummel plot follows ideal exponential functions for more than 7 decades. The low leakage collector current is in order of pA. The current gain is about 90 which is reasonable. The output characteristic is also good with high breakdown voltage of more than 6 V and high Early voltage of about 48 V. The collector resistance which is the reverse value of slope of collector current in the linear region of the output characteristic is about 165 Ω for this device. This value is rather high and can be reduced by an additional phosphorus implantation for collector plug in step 4d.40.

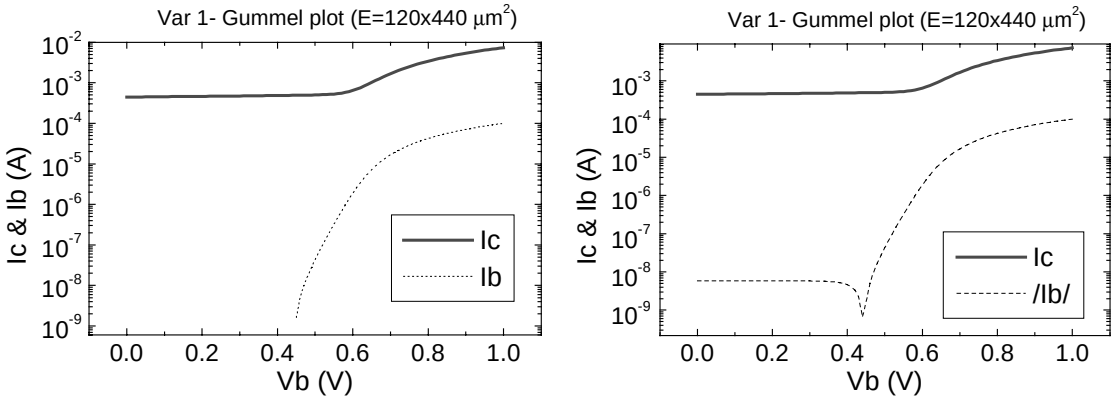
Table 4-5 Summary of the IV characteristics of BJT fabricated according to variant 1

	Bvceo (V)	β (Vb=0.5V)	Vcc (V)	Vaf (V)	Cjbc (5V) (nF/cm2)	Rse (Ω/sqr)	Rsb (Ω/sqr)	Rsc (Ω/sqr)
Measured	>6	90	5	47	31	101	1500	265
Simulated	5.4	120	3.3	32	40	-	4500	300

Different types of transistors were also characterised and compared. Following table lists parameters of those.

Table 4-6: measured parameters of transistor types of variant 1

Transistor type	A1	A2	A3	B1	B2	B3	C1	C2	C3
Collector leakage current at Vbe=0V **Ico**	20 μA	15 μA	10 nA	2 pA	0.8 mA	0.6 mA	4 pA	10 pA	15 μA
Negative base leakage current at low Vbe	Yes	Yes	Yes	No	Yes	Yes	No	No	Yes
Early voltage **Vaf (V)**	47	47	49	47	10	15	47	48	40
Collector resistance **Rc (Ω)**	364	222	214	110	156	162	164	165	130



a) *b)*

Figure 4-22: Gummel plot of a transistor having negative base currents at low base voltages a) collector and base current vs. base voltage; b) absolute base current vs. base voltage

All parameters of working transistors were measured and extracted. The sheet resistances and collector base junction capacitances were also measured and listed in table 4-5. The simulated values are also presented and compared with the measured parameters. It is noticed that the base resistance was measured with the Van der Pauw structure without the poly-emitter layer for all three variants. Hence, the measured base sheet resistance is lower than simulation. However, most of parameters are also different from the simulation. This result suggests some doubts about the simulation models.

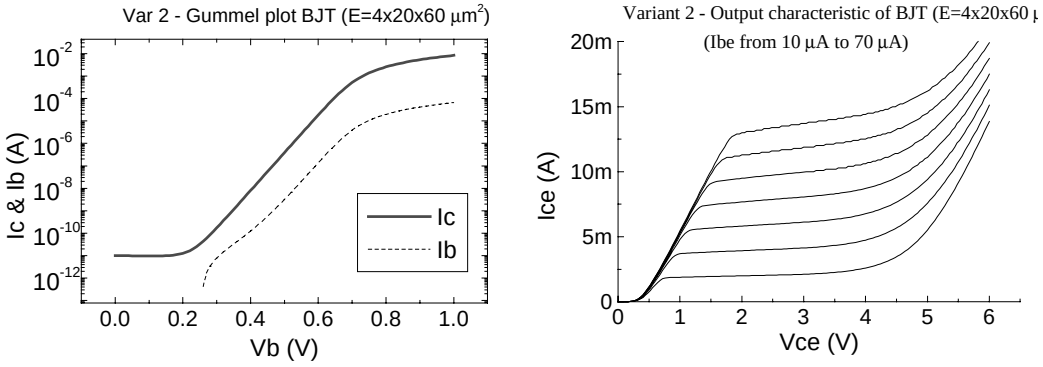
As seen in table 4-6 most transistors have negative base leakage current at low base voltages and very high collector leakage current in Gummel plots (figure 4-22). This means that there are short connections between emitter and collector. Those shorts are caused by diffusion pipes or diffusion spikes through the thin base width along dislocations which are the results of oxidation-induced stacking faults or other process-induced defects. The larger the emitter area the more chance the shortcuts appear. This phenomenon is also observed in table 4-6 where only small narrow emitter transistor or large peripheral emitter with small emitter areas do not suffer from this defect. The lower collector resistances of narrow emitter transistors compared to the wide emitter transistors are also illustrated in this table.

The uniformity over process wafers were characterised by comparing different dies of transistors on the wafers. Measurement results show about 33 % of devices which are mainly located in the centre of the silicon wafer are working well. The rest are not working properly due to defects. The reason why there are only 33 % of transistors which are working in this variant is still not understood. This yield is much higher for variant 2 and 3. Differently from other variants, the collector implantation in this variant was conducted with a very rare ion source, i.e. P^{3+}. As a result, it took very long implantation time, i.e. more than 10 hours per wafer. This slow and very high-energy implantation could contaminate and create a many defects in process wafers and they resulted in these bad devices.

However, the working transistors have good characteristics which are similar and even better than that predicted by simulation, e.g. the supply voltage could be 5 V compared with 3.3 V of simulation, the breakdown voltage Vceo and the Early voltage Vaf are higher, the base collector junction capacitance Cjbc is lower. The Gummel plots show ideal collector and the base currents.

If one compares the SRP and the simulated profiles of a transistor of variant 1 in the figure 4-11 and figure 4-19 it is easy to recognise that the base is much shallower in the fabricated transistor than in the simulation. This resulted in the improvements of the real transistors in comparison with the simulation. The reason why the base is shallower is that we used 10 nm oxide as a mask for collector and base in the simulation but in fact we still kept this thickness 25 nm during processing. We did this because we did not want to change much the UT-BiCMOS process and to etch the oxide from 25 nm down to 10 nm is not reliable. This shallow base is dangerous for devices which easily suffer from the emitter collector short cut. Base implantation with higher energy, e.g. BF_2 at 45 keV, would improve processing yield for this variant.

4.2.4.2.2. Variant-2 BJT



a) *b)*

Figure 4-23: The gummel plot (a) and the output I-V characteristic (b) of a transistor fabricated according to variant 2.

Table 4-7: Summary of the IV characteristics of BJT fabricated according to variant 2

	Bvceo (V)	β (Vbe=0.5V)	Vcc (V)	Vaf (V)	Cjbc (5 V) (nF/cm2)	Rse (Ω/sq)	Rsb (Ω/qr)	Rsc (Ω/qr)
Measured	>4	120	3.3	18	11	140	3700	314
Simulated	5.3	135	3.3	32	39	-	5000	360

Table 8: measured parameters of transistor types of variant 2

Transistor type	A1	A2	A3	B1	B2	B3	C1	C2	C3
Collector leakage current at Vbe=0V **Ico**	2 pA	5 pA	10 pA	5 pA	50 μA	70 pA	3 pA	10 pA	50 pA
Negative base leakage current at low Vbe	No	No	No	No	Yes	No	No	No	No
Early voltage **Vaf (V)**	44	20	19	17	17	17	18	17	18
Collector resistance **Rc (Ω)**	328	324	324	127	125	121	112	107	112

The same characterisations were carried out on the devices fabricated according to this variant. Good electrical characteristics are shown in figure 4-23 and measurement

parameters are also listed in table 4-7. Compared with variant 1, variant 2 has better characteristics. The current gain is higher, β=120 at Vbe=0.5 V, and the important base collector junction capacitance is much lower for this variant, Cjbc=11 nF/cm^2. Although the breakdown voltage is a little bit lower than in variant 1, the devices are still working well at a supply voltage of 3.3 V. However, the Early voltage is rather low, about 18 V, which is caused by a very shallow base width.

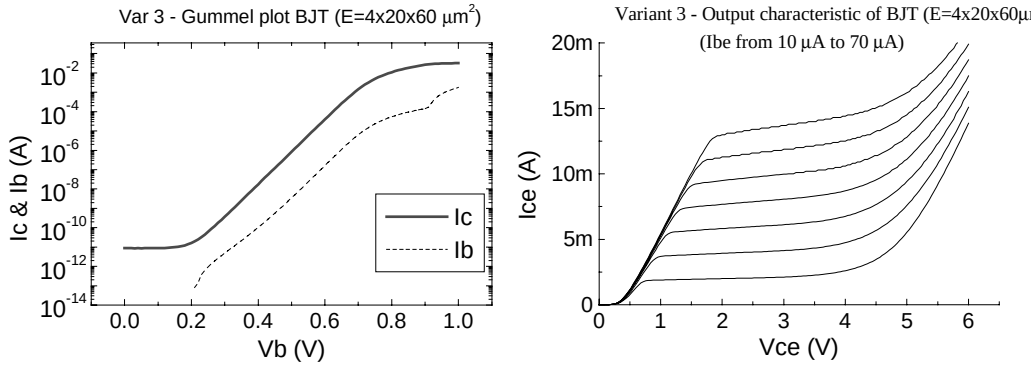
Similar to variant 1, the measured electrical parameters do not fit the simulated ones. Therefore, simulation should be calibrated again, especially for RTP profiles. The devices showed better DC and AC characteristics than simulation. The shallower base of the fabricated devices may be the reason for this improvement compared with the simulation.

The processing yield among different transistor types is much better for this variant. There is only one device having emitter collector shortcuts. Most of transistors are working well and collector leakage current is very low even for large emitter area transistor. This suggests the advantage of lower collector implantation energy which may result in less damage and more reliability.

However, there are still defective devices located randomly on variant 2 wafers. Furthermore, the base current has less range of ideality, only for 4 or 5 decades in the Gummel plots. This may be the result of the rapid thermal annealing which often cause non-uniform thermal distribution and stress-enhanced damage. Additional furnace annealing is recommended for damage removal.

4.2.4.2.3. Variant-3 BJT

Variant 3 has been characterised like two first variants and the results are presented in following figures and tables.



a) *b)*

Figure 4-24: The gummel plot (a) and the output I-V characteristic (b) of a typical transistor fabricated according to variant 3

Table 4-9: Summary of the IV characteristics of BJT fabricated according to variant 3

BLOCK	Bvceo (V)	β Vb=0.5V	Vcc (V)	Vaf (V)	Cjbc (3.3V) (nF/cm2)	Rse (Ω/sq)	Rsb (Ω/qr)	Rsc (Ω/qr)
Measured	4	240	3.3	17	7	147	3700	312

Table 4-10: measured parameters of transistor types of variant 3

Transistor type	A1	A2	A3	B1	B2	B3	C1	C2	C3
Collector leakage current at Vbe=0V **Ico**	8 pA	9 pA	11 pA	3 pA	8 pA	14 pA	4 pA	10 pA	14 pA
Negative base leakage current at low Vbe	No	No	No	No	No	No	No	No	No
Early voltage **Vaf (V)**	17	17	17	17	17	18	17	17	17
Collector resistance **Rc (Ω)**	348	325	317	130	125	119	117	106	107

Variant 3 is the best of all variants as the results shown in the above table with 100 % good devices. The electrical parameters of transistors in this variant are also much better than these other two, i.e. β is more than 200, collector and base current are more ideal, and Cbc is lower (7 nF/cm^2 compared with 39 and 11). The supply voltage is 3.3 V similar to variant 2. Besides, every device of different emitter structure is working very well and there is no sign of any defective device.

As in variant 2, this process flow also uses RTP to reduce thermal budget but it has additional annealing steps of SOD diffusion and poly-emitter oxidation. Therefore, devices of this variant have not only very attractive shallow emitter and base profiles but also better damage removal. This result proved that spin-on dopant can be applied very well for shallow poly emitter formation.

4.4.3. Summary

The transistors were fabricated according to the three simulated variants. Devices were electrically characterised and parameters were extracted. The results showed that in all the three variants the devices had better DC and AC characteristics than in simulation. The cut-off frequency could not be measured due to lack of equipment but the lower measured base collector capacitance implied a higher cut-off frequency than simulation predicted.

From the technology point of view, variant 1 should not be recommended because the collector implantation using P^{3+} is a very difficult process. Variant 2 and 3 are good, even though an extra annealing step should be added into variant 2 for better activation and damage removal.

Variant 3 with the use of spin-on dopant for emitter doping showed the best results of all variants. It also proved that rapid thermal annealing of spin-on dopant was a very effective technology for shallow junction formation.

4.5. Conclusions and recommendations

A bipolar transistor fabrication process has been developed and optimised for high energy implantation. Three variants of the process flows were proposed to improve the high energy collector implantation. With the help of simulation tools the processes were optimised for high cut-off frequency.

The bipolar transistors were also realised according to the three simulated variants. The devices were obtained and showed even better IV characteristics than simulation.

The first variant was proposed to keep the standard UT-BiCMOS process from changes and used P^{3+} ions at 450 kV for collector implantation. As a result, it was still very difficult to obtain P^{3+} ions and the breakdown problem was not solved completely. Thirty-three percent of the obtained devices are good. This might be due to implantation of the collector with the P^{3+} ion source.

The second variant was designed to use advanced rapid thermal annealing technique for better device performance. The collectors were implanted with P^{2+} at 500 kV and there was no problem. More than fifty percent of the devices are good but it could be better if an extra thermal annealing step were inserted after emitter implantation.

Variant 3 with SOD source for doping the emitter showed the best results. This variant was almost similar to variant 2 but all of the devices are very good showing the advantage of the diffusion process from SOD source. Variant 2 or 3 are recommended that for bipolar transistor process.

References

[1] J. Bardeen & W.H. Brattain, “The transistor, a semiconductor triode”, Phys. Rev., vol. 74, 1948, p. 230

[2] R.C. Wijburg, “High energy ion implantation for merged silicon technologies”, Ph.D. Thesis, University of Twente, the Netherlands, 1990

[3] K. Oldenziel, “High frequency bipolar transitor for the UT-BiCMOS process”, M.Sc. thesis, University of Twente, the Netherlands, 1996

[4] M. Takagi, “Improvement of shallow base transistors technology by using a doped polysilicon diffusion source”, J. Japan. Soc. Appl. Phys. Vol. 42, 1973, p. 101

[5] P.R. Gray and R.G. Meyer, “Analogue Integrated circuits”, John Wiley & Sons, New York, 1977

[6] P.R. Gray, “Analysis and design of analogue integrated circuits”, John Wiley & Sons, New York, 1977

[7] I.E. Getreu, “Modelling of the bipolar transistor”, Elsevier, Amsterdam, 1978

[8] TSUPREM4 manual v6.2, Technology Modelling Associates, USA, 1995

[9] MEDICI manuel v2.0, Technology Modelling Associates, USA, 1994

[10] D.B.M. Klaassen, "A unified mobility model for device simulation I, model equations and concentration dependence," Solid-state electronics, vol. 35, no. 7, 1992, p. 953

[11] D.B.M. Klaassen, "A unified mobility model for device simulation II, temperature dependence of carrier mobility and lifetime," Solid-state electronics, vol. 35, no. 7, 1992, p. 961

[12] D.B.M. Klaassen, J.W. Slotboom, and H.C. de Graaff, "Unified apparent bandgap narrowing in n- and p-type silicon," Solid-state electronics, vol. 35, no. 2, 1992, p. 125

[13] J.W. Slotboom, "Analysis of bipolar transistors," PhD-thesis, Technical University Eindhoven, The Netherlands, 1977

[14] T. H. Ning and D.D. Tang, "Bipolar trends," Proc. IEEE, vol. 74, no. 12, 1986

[15] K. Tsukamato, “Arsenic implanted emitter and its application to UHF power transistor”, Jpn. J. Appl. Phys., 17 (1), 1977, p.187

5

Realization of Sub-micron Poly-Spacer MOSFET

5.1. Introduction

The Metal-Oxide-Semiconductor-Field-Effect-Transistor (MOSFET) has been the major device for integrated circuits such as microprocessors and semiconductor memories over the past three decades. The principle of the surface field effect transistor was first proposed by Lilienfeld [1] and Heil [2] in the early thirties. The first MOSFET was realized with a thermally oxidized silicon structure by Kahng and Atalla [3] in 1960. Since then the MOSFET has been continuously developed and improved to become the most important device for ULSI (> 10^7 transistors on a chip). As compared with the bipolar transistor, the MOSFET is smaller and simpler to produce, and these advantages are needed for ULSIs.

The basic structure of a p-channel MOSFET is illustrated in figure 5-1. It has four electrodes and consists of a n-type semiconductor bulk into which two p^+ regions, called source and drain, are formed. Metal, heavily doped polysilicon or silicide material has been used as the gate material on top of the insulator layer. Presently, polysilicon is the standard material for gate formation. Important parameters of a MOSFET are gate length L_{eff}, gate width Z, gate oxide thickness d, source/drain junction depth X_j.

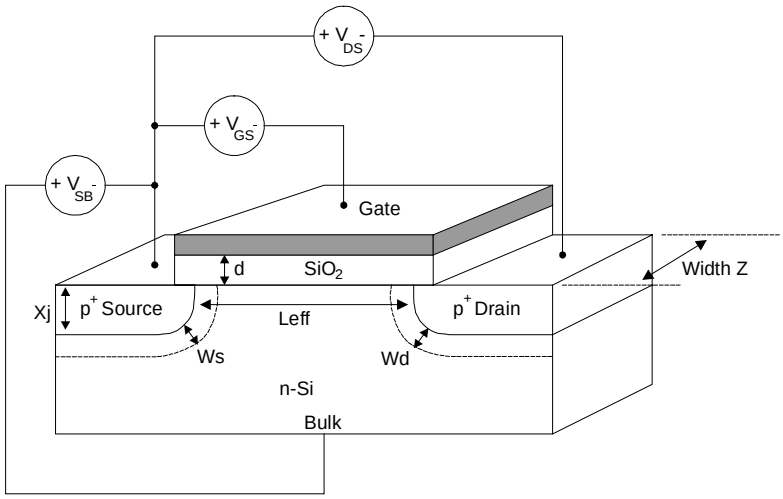


Figure 5-1: Schematic diagram of a MOSFET

The CMOS (Complimentary Metal Oxide Semiconductor) circuits, combining both NMOS (N-channel MOSFET) and PMOS (P-channel MOSFET), came to take over NMOS in the late seventies, as power dissipation and overheat issues could no longer be neglected.

With scaling, the performance of a semiconductor device can be increased by reducing all of its dimensions by a constant factor. At the current rate of miniaturization, CMOS technology is expected to reach a feature size limit of about 0.05 micron in the year

2012. At that point, quantum effects, such as tunneling of electrons will begin to play a role in semiconductor devices and will eventually dominate the operation of devices, with further decreases in feature size.

However, at present it is not well known what will happen. Economic factors may limit the scaling down and it may even stop before the 0.1 μm generation. Scaling down to less than 0.1 μm will lead to too many challenges not only with the technology but also with the circuit design, and testing. As a result, benefits from scaling will be reduced. Nevertheless, CMOS may last for a long time and research for new kind of MOS devices is attracting a lot of attention.

In this chapter, we will present methods to design an advanced sub-micron MOSFET. Many effects arise when the gate length is scaled to 0.25 micron region. The short-channel effect and the high series resistance of a sub-micron MOSFET are the most critical factors because of shallow source/drain junctions and low series resistance. Elevated source/drain MOSFET is an attractive structure for future sub 0.1 μm MOSFET. Our design is proposed with a MOSFET which has polysilicon spacers used as a buffer layer for shallow source/drain diffusion and for series resistance reduction. A optimization of the PMOS transistor process was simulated with the Athena and Atlas software packages [4]. The PMOS process was then realized in the MESA clean room and the devices have been characterized. The measurement results will be presented and some recommendations will be given for future research.

5.2. Scaling of MOSFET

5.2.1. The scaling laws [5]

Table 5-1: Three scaling laws

Parameter		Expression	Scaling law		
			Constant E	Generalized	Constant V
Dimension		W, L, d, x_j	$1/\lambda$ ($\lambda>1$)		
Voltage		V_{DD}, V_T	$1/\lambda$	$\kappa=\lambda$ $1/\kappa$ $\kappa=1$	1
Electric field		$E_\perp$, $E_\parallel$	1	λ/κ	λ
Doping density		N_B	λ	λ^2/κ	λ^2
Circuit	Capacitance	$C_G=A.\varepsilon_i/d$	$1/\lambda$		
	Current	I_D	$1/\lambda$	λ/κ^2	λ
	Gate delay time	$T_{pd}=C_G.V_{DD}/I_D$	$1/\lambda$	κ/λ^2	$1/\lambda^2$
	Power dissipation	$I_D.V_{DD}$	$1/\lambda^2$	λ/κ^3	λ
	Power density	P/A	1	λ/κ	λ
Interconnection	Line resistance	$R_L=\rho.l/A_l$	λ		
	Time constant	$R_L.C_L$	1		
	Current density	I_D/A_L	λ	λ^3/κ^2	λ^3

It is well known that ULSI has been achieved mainly by scaling of MOSFETs. Shrinking of the device size leads to many performance enhancements such as faster switching speed, lower power dissipation, and smaller circuit areas. In the mid seventies the first scaling concept was proposed by Dennard et al. from IBM [6]. This scaling scenario is called the *constant-field scaling.* The internal electric fields are kept constant by reducing the geometry and voltages by the single scaling factor λ. When the geometry is scaled the electrical field in the MOSFET structure is not reduced and two-dimensional effects at high electric field will degrade the device performances. Reducing the applied voltage together with increasing doping concentration by λ will reduce the depletion layer almost λ times. Therefore, undesirable short-channel effects and source-drain punch-through effects caused by a two-dimensional field effect will be reduced. This theory was confirmed by the successful fabrication of 1-μm MOSFETs exhibiting the expected, properly scaled characteristics with respect to those of the conventional 5-μm MOS technology.

In the mid eighties, the IBM researchers themselves modified their scaling theory to the *generalized scaling* law [7] which reduced the voltages by a factor of κ generally smaller than the geometry factor λ. By resetting the starting supply voltage to 2.5 V for the 1 μm channel FET, it was demonstrated that a 0.25 μm channel FET with 5 nm gate oxide (λ=4) operates well under a reduced (κ=2.5) bias of 1 V. The generalized design does not require major modifications of the state-of-the-art MOS technology since $\kappa < \lambda$, but as pointed out, it is rather an evolution toward thinner gate oxides and the proportionally reduced fabrication tolerances since the oxide thickness is scaled by the same factor λ.

However, voltages can not be reduced with the same factor as the device dimensions, mainly due to the following limiting factors:

a) The non-scalability of the sub-threshold swing S, which is a physical limit derived from the thermal voltage Φ_T (kT/q)

b) The temperature deviation of V_T (1 - 2 mV/^{o}C). This lead to a large threshold fluctuation if an extended operating temperature range (e.g., from –55^{o}C to +125^{o}C) must be guaranteed. This physical limit basically results from the temperature dependence of the Fermi potential Φ_F.

Both the above limits prevent V_T from being scaled proportionally with the device dimensions down to less than roughly 0.25 - 0.35 V, otherwise the OFF-state leakage current and so the power dissipation are undesirably increased.

c) The non-scalability of the transition region [8], where the inversion charge Q_I is nonlinearly increased with V_G. This results in a reduced current drive.

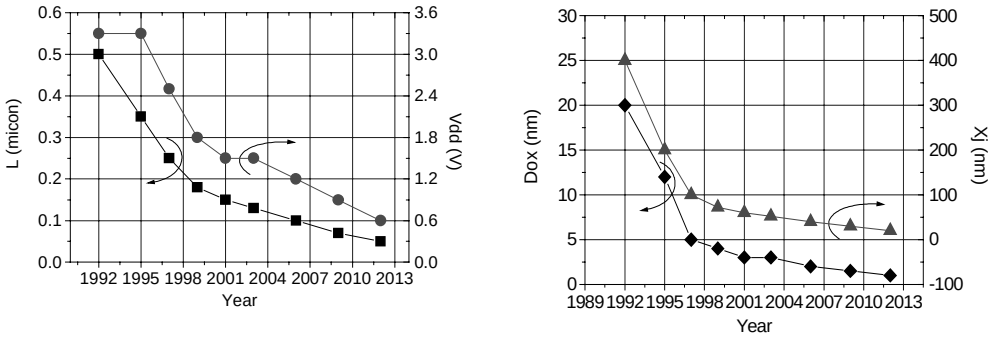
In addition to the above fundamental limits related with the physical parameters of Φ_t, Φ_F, and E_g, there are many practical limiting factors.

d) The circuit requirement on the functional capability with proper noise immunity prevents the power-supply voltage from being less than $\approx 4V_T$ in practice.

e) The practical requirement for the circuit-interface compatibility with the existing logic families (e.g., 5 V for TTL) tends to keep the circuit power-supply voltage constant as long as possible, unlike the constant-field scaling law. This is called *constant-voltage scaling.*

f) The non-scalability of the interconnection capacitances (fringe effects, lateral capacitance, etc.) lead to a reduced circuit performance with progress of scaling, unlike the constant-field scaling law.

All the above fundamental and practical limiting factors call for supply and threshold voltages to be reduced less than the conventional constant-field scaling would require. Especially due to the practical limits (e and f), the constant-voltage scaling scenario has been representative of the trends in circuit scaling down to the 0.35 μm generation.



a) *b)*

Figure 5-2: NTRS Road-map [9] for a) scaling of gate length and maximum supply voltage; and b) Scaling of maximum gate oxide thickness and maximum source/drain junction depth

With constant-voltage scaling, lateral and normal electric fields in a MOSFET have been considerably increased. Especially, since the power-supply voltage was maintained at 5 V over five device generation from 3 μm to 0.8 μm design rules, the insulator field E_I has already come to nearly its highest allowable value of ≈ 4 MV/cm. Accordingly, various effects resulting from large electric fields have started to limit MOSFET performance and reliability. Furthermore, when the minimum feature length goes down to the sub-0.1 μm range the clock frequencies approach 1 GHz, pressure to manage switching power, $CV^2_{dd}f$, is forcing power supply voltage, V_{dd} to decrease to even lower than 1 V. Therefore, constant-field scaling law has been applied again for the generations of lower than 0.35 μm gate length as illustrated in figure 5-2. In the next section we will present some of the issues related to the scaling.

5.2.2. Scaling limitations

The scaling trends described above lead to several limitations and issues of MOS technology. These limitations lead to trade-off between performance and reliability of a scaled device and it becomes the most important device design issue in actual ULSIs. The most important limitations are:

- small-geometry effects
- limited performance under high fields $E_\perp$ and $E_{//}$
- hot-carrier-induced device degradation
- gate-induced drain leakage current
- gate-dielectric reliability
- source/drain shallow junction and series resistance

The short-channel effect is the most important issue when a MOSFET is scaled down. Experimentally, the short-channel effect is observed to degrade the subthreshold characteristics, reduce the threshold value with decreasing L_{eff}.

Actually, there are only a few common criteria to distinguish a long-channel from short-channel device definitely. Empirically, the minimum available channel length L_{min} for which long-channel subthreshold behaviour can be observed, has been found to be the simple relation [10]:

$$L_{\min} \approx 0.9\left[dX_j (W_S + W_D)^2\right]^{1/3}$$

Where d, X_j, W_S and W_D are defined in figure 5-1. This equation is often used as a guideline for MOSFET miniaturization. It has been applied to a high performance 50 nm MOSFET, which was realized with a 2 nm gate oxide and very shallow junctions of 7 nm [11]. The scaling roadmap for the gate oxide thickness and the source/drain junction depth of present and future deep submicron MOSFETs are illustrated in figure 5-2b.

As junction depths are reduced, adverse effects arise. First, the parasitic series resistance of the source/drain region of the MOSFET increases and degrades the device performance. This series resistance of the source and drain is a significant part of the total device resistance for a short channel device. Secondly, reduced junction depth increases junction curvature, raising the electric field in the drain region. This increased field is undesirable because it creates energetic carriers that degrade the oxide and because it can limit the maximum voltage that can be used. Thirdly, shallow junctions are prone to damage during fabrication. This damage promotes leakage current, which is not desirable in CMOS design.

Elevated source/drain structures combined with silicide technology has been proposed for deep sub-micron MOSFET [12]. Our poly-spacer MOSFET also has a similar structure and it will be described in the next section.

5.3. Design of 0.25 μm P-MOSFET

5.3.1. Drain Engineering and Poly-spacer MOSFET concept

In scaling down devices, thinner gate oxides and higher-doped channels are needed to boost the punch-through voltage of the short channel devices. Both measures dramatically increase the lateral electrical field near the drain region where charge carriers, accelerated by this field are becoming hot, some can overcome the oxide barrier and are injected into the gate dielectric. Hot carriers trapped in the gate oxide will degrade the device performance by V_T shift and mobility reduction. The lightly doped drain (LDD) [13] was proposed to reduce this electrical field and to improve the device reliability.

In the LDD structure, the drain is formed by two implantations as shown in figure 5-3A. One of these is self-aligned to the gate electrode and the other is self-aligned to an oxide spacer at the edge of the gate. The spacer is formed by depositing an oxide layer of about 200 nm thick and then etch it back with blanket anisotropic reactive ion etching. All the oxide in the field will be cleared accept at the region next to sharp steps where the deposited oxide is the thickest. The first, lighter dose forms a lightly doped section of the drain and source at the edges near the channel. In the PMOS devices, this dose is normally 1 to $5x10^{13}$ cm^{-2} of BF_2. The second, heavier dose forms low-resistivity regions of the source and drain that merge into the lighter doped regions. This implantation is typically boron or BF_2 with a dose of 1 to $5x10^{15}$ cm^{-2}. These two implantations are separated and displaced in a self-aligned manner by spacers. As a

result, the electrical field can be reduced to an acceptable level near the gate edge because the impurity level is much lighter than that in a conventional drain. Since the heavier, second dose is further away from the gate edge than would be in a case of conventional drain structure, its depth can be made somewhat greater without adversely affecting the device operation. The increased junction depth also has advantages of lowering the sheet resistances and providing a better protection against junction spiking by the aluminium contacts.

However, when the gate length is scaled down to below 0.25 μm, the LDD structure become less useful because the spacer interface degradation from hot carrier injection is not severe. Furthermore, in this deep submicron region the channel resistance is very small and the series resistance contribution from the p^- region under the oxide spacers become more significant. As a result, the output characteristic of the LDD transistor is degraded when the gate length is further reduced to less than 0.25 μm.

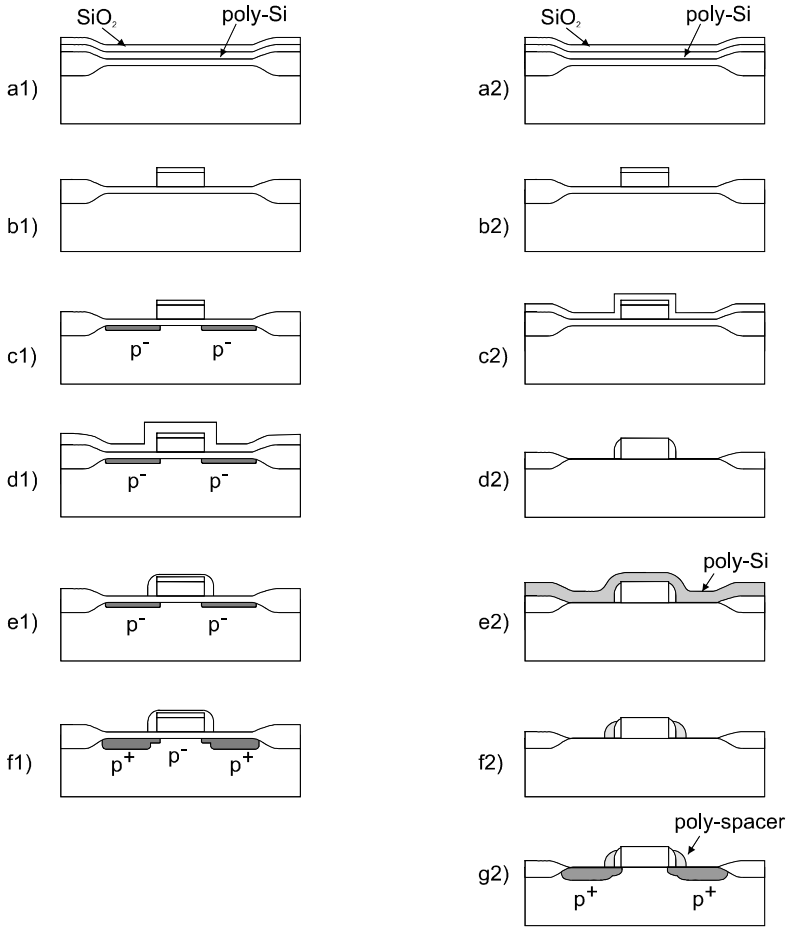


Figure 5-3: process flow of A) LDD-MOSFET and B) Poly-spacer MOSFET

There are solutions for sub-0.25 μm structures. One is to increase the LDD p^- doping density to the doses of $10^{14} – 10^{15}$ ions/cm^2. This structure can be called highly doped drain (HDD) structure. The HDD structure could reduce the series resistance but the V_{dd} also has to be reduced to lower the $E_{//}$ field. This approach is in consistence with the constant-field scaling law. There are alternative structures such as large-angle-tilt implanted drain [14] (LATID) and double-implant lightly doped drain (DI-LDD) [15]. However, all of these solution will become less effective for the sub-0.1 μm generations because the series resistance in these MOSFETs will be too high then.

Elevated source/drain with selectively grown silicon often combined with salicide technology is a very attractive solution for lower series resistance [16]. However, this technology is very complicated and selectively deposition is difficult and not reproducible. To solve the technology difficulties, we have proposed a simpler approach by using poly-spacers in the MOSFET structure which has elevated source/drain areas near the channel (see figure 5-3B and section 6.4.2 for more process details). The poly spacers can be simply created in the same way as the standard isolation spacers (figure 5-3B e2, f2). Implantation or diffusion techniques are then used for doping the source/drain and the gate simultaneously (figure 5-3B g2). The junction area under the polysilicon layer is shallower in comparison to the open silicon area where the metal contacts are placed because the poly layer acts as a buffer layer for both implantation and diffusion. Compared with LDD technology, this process has one doping step less. This approach is simple and effective for high-performance submicron MOSFET.

Details of our poly-spacer deep sub-micron MOSFET are illustrated in figure 5-4. A super steep retrograde well, indicated by n^+, a shallow extension junction depth and self-aligned oxide isolation with poly spacers were used to reach the high performance standards of a 0.25-µm PMOSFET.

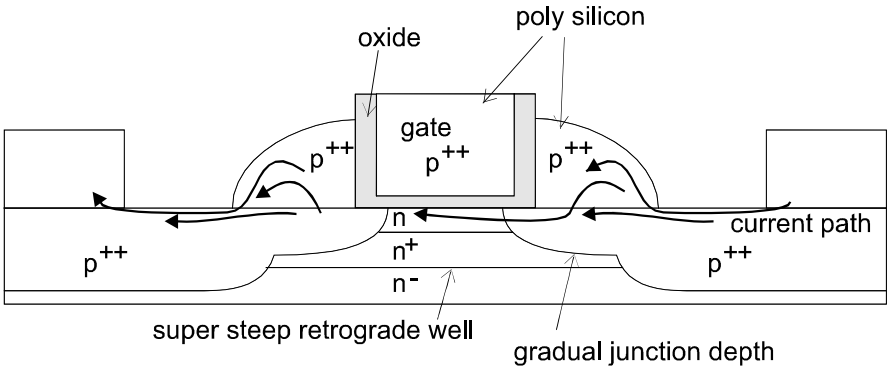


Figure 5-4 Poly-spacer quarter micron PMOSFET

Scaled gate oxide thickness of eight nanometers was used in this experiment to ensure processing reliability during gate oxide etching. Five nanometer gate oxide will give better MOSFET characteristics but the etching reliability has to be ensured. It is recommended for the next improvement of the device.

The channel profile has to be optimized for a required threshold voltage and for short-channel-effect suppression. In our design, the short channel effect was eliminated by the super-steep-retrograde channel combined with the shallow source/drain junctions. This channel has a peak of donor concentration at about the shallow source/drain junction depth in order to reduce the depletion widths and lower V_T. The surface channel concentration and the gate oxide thickness both determine the threshold voltage. The optimum channel doping profile must be found out with the help of simulation tools. The main goal of the simulation is to figure out the optimum process flow for a good performance quarter-micron PMOSFET. The next paragraph explains the set up of the simulations.

5.3.2. Simulation set up

5.3.2.1. The global approach

The simulation can be split up in two major phases. In the first phase, the 0.25-µm PMOST process must be simulated with all the processing steps. Especially, different thermal annealing steps are tried out to find good doping profiles for a quarter-micron

MOSFET. Different channel implantation energies and doses are implemented in the input files for the process simulation. The output results are different MOSFET structures which are used for device simulation inputs.

The second phase is to simulate the electrical behavior of the "produced" PMOSFETs and optimize electrical parameters. The correspondingly generated electrical characteristics are then used for evaluation. The optimized channel implantation is chosen as the one which gives the best electrical characteristics from the device simulation. Finally, three more variants of PMOS process flow were simulated. The above process flow (variant 1) uses conventional furnace annealing for diffusion of SOD. A new variant (variant 2) uses rapid thermal processing for SOD diffusion of boron. Two other variants , variant 3 and 4, use BF_2 implantation for boron doping using conventional furnace annealing and RTP respectively instead of SOD diffusion.

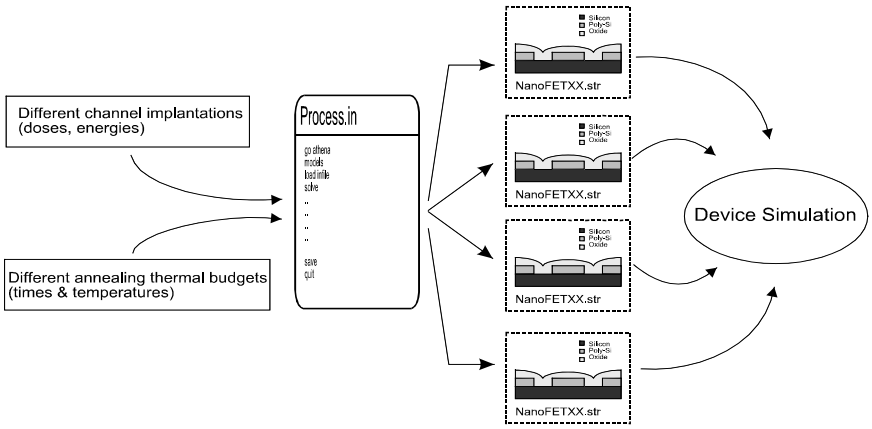


Figure 5-5: Process simulation

5.3.2.2. The simulation of the process

Process simulation was carried out using the ATHENA software package which is an advanced 2D process simulation tool set. The most important tool of ATHENA is SSuprem4. SSuprem4 is the state-of-the-art 1D and 2D semiconductor process simulator used in the design, analysis and optimization of silicon fabrication technologies and device structures. It incorporates a wide range of advanced physical models for diffusion, implantation, and oxidation modeling. All the steps of the PMOS process flow have been simulated from LOCOS formation to metalisation (see appendix 6). Details of the process flow will be discussed in the experimental section.

Adjustment of the diffusion models is very important in simulation. To approach the process as much as possible the implantation damage, i.e. interstitials and vacancies, are also taken into account using the Plus One Model. This model describes the diffusion of impurity atoms as function of point defects introduced by implantation. The effect on phosphorus of these Plus One model is shown in figure 5-6. A higher diffusion rate of the phosphorus atoms creates a more smoothed concentration of these implanted atoms and a higher concentration at the channel surface (x=0). Both simulation results, with and without the Plus One model, are plotted as different concentration lines. The third line shows the arsenic concentration from surface towards the bulk, which was almost the same in both cases.

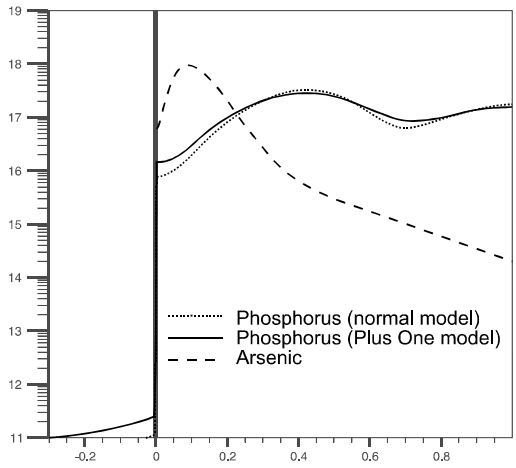


Figure 5-6: Point defect influence on phosphorus diffusion

The diffusion simulation model of implanted boron in the source/drain regions use default parameters. Especially, two dimensional diffusion effects under the spacer oxide is hard to be quantified. In addition, the full couple model of boron diffusion in mono Si could not be implemented in polysilicon diffusion. As a result, the transient-enhanced diffusion effect could not be taken into account during diffusion simulation of implanted source/drain and shallower junctions were obtained than expected in the open silicon region as illustrated in figure 5-14. However, the source/drain extension junction depths under the oxide spacer is more important for the device characteristic and there the default diffusivity can be used.

The diffusion model of boron from SOD source into the polysilicon/Si structure is not developed yet. Therefore, the simulation results should be used only to compare the difference between process variants. More calibration of boron diffusivity in polysilicon as well as mono-silicon should be carried out in order to obtain absolute simulation results.

5.3.2.3. The simulation of the device

The device simulation was conducted using the ATLAS software package containing the two dimensional semiconductor device modeling tool. To obtain accurate results for MOSFET simulations, it is necessary to take account for mobility effects associated with inversion layers. This is done by modeling mobility as having a dependence on the transverse field. The CVT [17] model was selected for our device simulation. Some other mobility models, e.g. Watt model [4] and Tasch model [4] were used for comparison and they gave similar results.

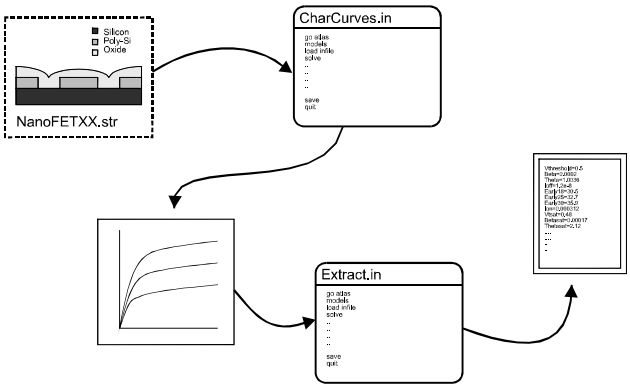


Figure 5-7: Device simulation

The device simulation has also been split up, see figure 5-7. The first part, which is the program *charcurves.in* simulates the transistor under normal operation conditions creating the characteristic output and transconductance curves. Another program *extract.in*, representing the second part, analyses the electrical parameters of these curves. It is convenient to separate this from the long solve command series in the file *charcurves.in*, if other parameters need to be examined. This file will extract the parameters of the transistor, e.g. the threshold voltage under saturation or triode regime, the transconductance, the Early voltage etc. Extracted data will be used to select the best performing transistor under the various simulation conditions.

5.3.3. Selection criterion

The wide range of selection criteria makes it necessary to define a straightforward selection procedure. Selection can be done by defining ideal areas in 2D-contour plots, drawn as function of the implanted dose and implantation energy, for each parameter. An area on this two dimensional plot, containing the best values within a certain range, can be determined for each parameter. Combining these plots leads to the choice for the best combination of parameters, and thus the best PMOSFET.

In our design, we are looking for a MOSFET with a maximum on-current and a minimum off-current. In practice circuits require a threshold voltage of approximately one fifth of the supply voltage V_{dd} which is 2.5 V for 0.25 μm MOSFET. It is desirable to strive for a threshold voltage of 0.5 V. The short channel effect can be checked with different gate lengths. In this work, a PMOSFET of 1.25 μm is used as a reference and it should have almost the same threshold voltage as one of a quarter micron. The difference between these two should not exceed 100 mV.

Although the main interest of the industry is to create a small gate length for digital logic, still it can be rewarding to examine analog characteristics of the transistors. Especially in the high frequency telecommunications market a high quality transistor is very interesting. The Early voltage describes the quality of the modeled current source. Limiting the voltage gain of one transistor stage, a high Early voltage is therefore a requirement in analog designs. This parameter correlates with the earlier mentioned short channel effect (SCE). The transconductance parameter β and mobility reduction factor θ will be determined for long and short channel PMOSFETs too.

5.3.4. Simulation Results

Simulations were conducted by varying the arsenic super-steep-retrograded doping concentration of the channel to find an optimum transistor within certain constraints. The implantation dose is restricted from 0.5x 10^{13}/cm^2 to 5x 10^{13}/cm^2, and the implant energy from 100 kV to 400 kV. Simulations are carried out with steps of 50 kV and for arsenic doping doses of 0.5, 1, 2, 2.5, 3, 4 and 5x 10^{13}/cm^2. To reduce simulation time at first 20 types are simulated in the range of 100 to 300 kV and from 0.5 to 2x 10^{13}/cm^2. The results of these transistors are used as guide for the higher dopant energy and higher dose transistor types. With this information initial guesses for further simulations are done with higher implant doses and acceleration voltages.

5.3.4.1. Threshold voltage of simulated devices

The required threshold voltage of a good transistor should be around 0.5 V, i.e. one fifth of the supply voltage V_{dd}. In addition, the voltage drop due to the short channel effect (SCE) should be less than 100 mV. To exaggerate the influence of the SCE, the MOSFET threshold voltage is also measured in saturation condition. This threshold

voltage is around 20-30 mV lower than the one measured at unsaturated operation for 1.25 μm gate length and around 40-50 mV lower for quarter micron gate length.

In figure 5-8 the threshold voltage of simulated quarter micron devices is shown as function of the process parameters, the implant energy and the arsenic dopant dose. During simulation, a longer furnace annealing time of the Spin-On-Dopant was necessary, from originally 14 to 17 minutes (the left and the right part in figure 5-8). As expected the threshold is higher when the arsenic dose is high combined with low implantation energy because the dopant concentration near the surface becomes higher. In contrast, a high energy and a low concentration create a low threshold voltage for the quarter micron device. In between these extremes, a region is present where the threshold approaches the desired value of 0.5 V.

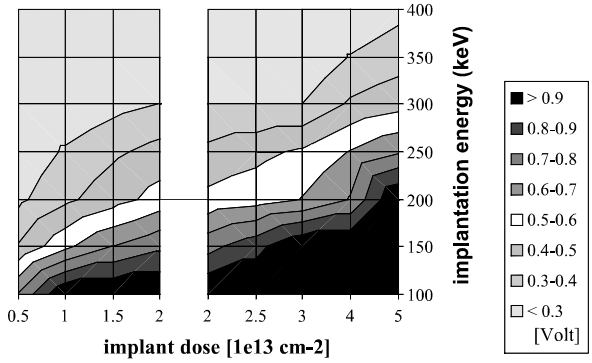


Figure 5-8: Threshold voltage at linear region (V_{dd}=-0.1 V) of 0.25μm PMOSFETs, using an anneal step of 14 minutes (left) or 17 minutes (right).

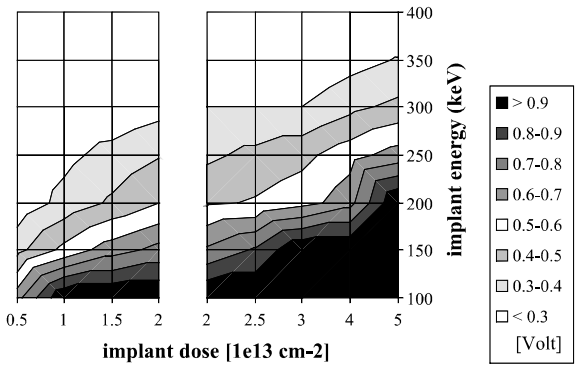


Figure 5-9: Threshold voltage at saturation (V_{dd}=-2.5 V) of 0.25μm PMOSFETs, using an annealing step of 14 minutes (left) or 17 minutes (right).

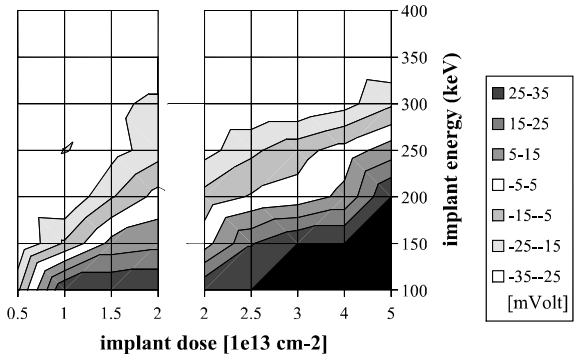


Figure 5-10: Threshold voltage drop at saturation of 0.25μm PMOSFETs, using an anneal step of 14 minutes (left) or 17 minutes (right)

In figure 5-8 the white area represents the area with a threshold voltage in between 0.5-0.6 V. At the upper border of this region, the desired threshold voltage is achieved. Small variations in the effective gate length may not disturb functioning of the transistors. Therefore, the threshold-voltage drop between short and long channel transistors must be reduced as much as possible. In figure 5-10 the threshold voltage drop of the simulated transistors at saturation is plotted as a function of two process parameters. While a high implant energy with a low concentration of dopant causes threshold voltage drop or short channel effects, a combination of high dose and low energy causes reverse SCE. This is the effect due to the offset of the channel-doping peak from the surface to the bottom while the source/drain doping profile is maintained constant.

These threshold voltages were determined with the method of extrapolating a linear curve from the I-V transconductance curve to the gate voltage-axis. As done with the previous figure also here one region close to zero voltage drop is marked white. Comparison of both figure 5-8 and figure 5-9 remarkably shows that both areas, with the ideal parameters, are almost the same.

5.3.4.2. On- and Off-current of simulated devices

On-current and Off-current are very important parameters of a transistor. With the goal to realize fast circuits, a very high on current must be achieved. Off-current should be as low as possible. However, when the gate length is scaled down for higher on-current, the off-current will increase simultaneously. Hence, a certain on-/off-current ratio is desired depending on a specific application. According to the NTRS roadmap, a high performance 0.25 μm PMOS should have a maximum on-current density of 280 μA/μm.

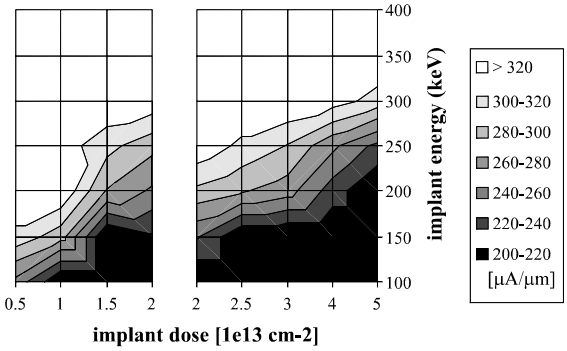


Figure 5-11: On-current of 0.25μm PMOSFETs, using an anneal step of 14 minutes (left) or 17 minutes (right).

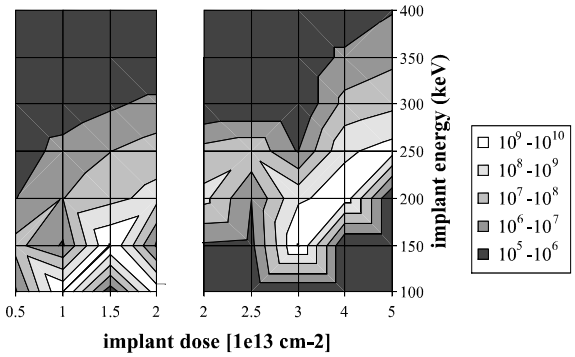


Figure 5-12: On/off current ratio of quarter micron transistors on logarithmic scale

In figure 5-11, the on-current is drawn as a function of the two implantation parameters. The difference between the transistors with 0.5 to 2x 10^{13} cm^{-2} (left) in comparison to the other types from 2 to 5x 10^{13} cm^{-2} (right) is remarkable. The latter ones have had a longer diffusion step of 17 minutes instead of 14 minutes for the low doped types. The influence of the diffusion time on the on-current is thus rather large, emphasizing the importance of this parameter. Also very clear is the influence of the threshold voltage for this parameter, the on-current. The dependency of the on-current and the threshold voltage on the implant process parameters show a great similarity. The ratio between the on and off current of the transistors is plotted in figure 5-11. The highest ratio, and thus the best behavior is found to be over 10^9 , white in this figure. With the goal to maintain a ratio of at least 3x 10^6 the white and the first two grey areas will fulfil this demand.

The data need further investigation as the reason for abrupt in- or decrease of the current ratio is unexplained. Possible cause is the concentration of dopants at the junction edges.

5.3.4.3. The Early voltage of the devices

A good analog transistor needs a high Early voltage to create a high voltage gain and a constant current source. Although small transistors are mainly made for high speed digital circuits, it can be interesting to view the influence of this parameter as function of the arsenic implant variables. In figure 5-13 this dependency is plotted. The highest Early voltage is found in two relatively small areas, but when this demand is less tight, a larger area containing very reasonable transistors, will satisfy.

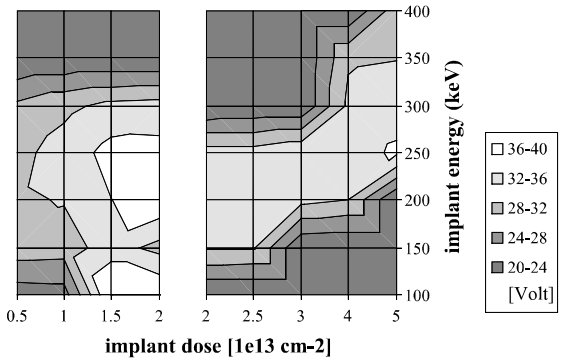


Figure 5-13: Early voltage of 0.25µm PMOSFETs, for both anneal steps of 14 minutes (left) or 17 minutes (right).

5.3.4.4. Comparison between variants

All previous simulations were done for variant 1 using conventional furnace annealing of SOD. During simulation, several annealing process and device simulations have been tried out. To ensure a working transistor the diffusion process is adjusted until the junction edge reaches the channel region. From simulation results a channel arsenic implantation with a dose of 2.8x 10^{13} ions/cm^2 at 250 keV is found to be optimum. With this implantation a PMOST has a threshold voltage of about 0.5 V at both transconductance and saturation conditions, reasonable Ion/Ioff ratio of 10^7, and Early voltage of 34 V.

However, rapid thermal annealing can be used instead of furnace annealing for SOD diffusion as in the variant 2. Simulation of diffusion was done at 1050°C for 40 seconds showing good transistor characteristics. Two other variants 3 and 4 using BF2 implantation and with furnace annealing and RTP respectively were also simulated. figure 5-14 shows net doping profiles of four variants from 1 to 4 corresponding to the top left, top right, bottom left and bottom right structures respectively.

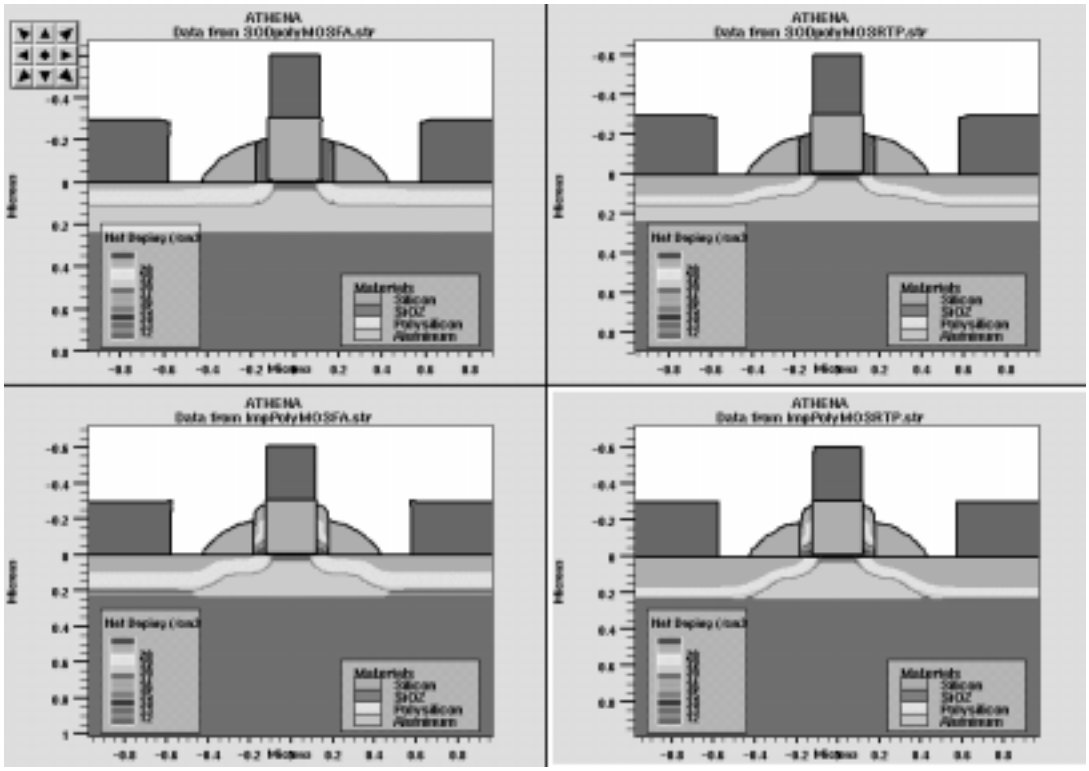
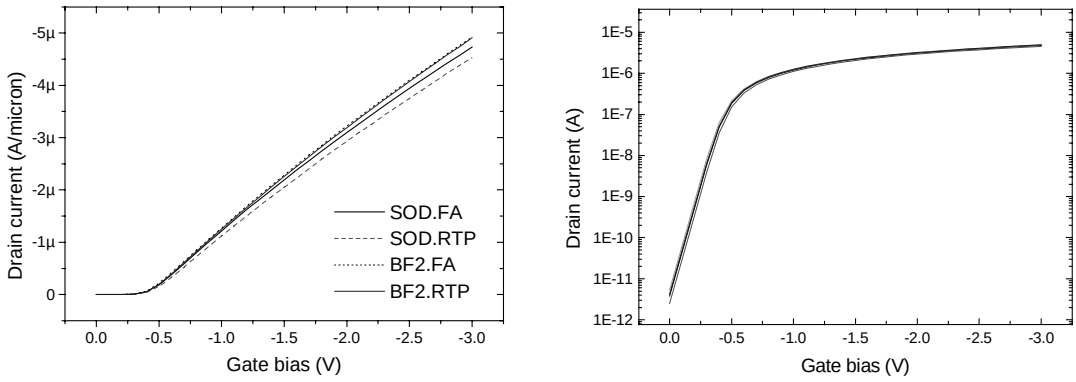


Figure 5-14: Net doping profiles of 0.25 μm PMOS variants: top-left: variant 1 (FA of SOD); top-right: variant 2 (RTP of SOD); bottom-left: variant 3 (FA of implanted BF_2); bottom-right: variant 4 (RTP of implanted BF_2)

The lateral doping profile underneath oxide spacers of a MOSFET is the crucial factor for operation of that MOSFET. As seen in figure 5-14, all the variants have similar doping profiles under the gate and the oxide spacers. However, the profiles under the poly spacers and the Si source/drain regions are quite different between variants. It is clear that variants with implantation have much deeper junction depth in the contact area. The reason is that boron ions are implanted and then diffuse deeper into Si substrate during annealing due to transient enhanced diffusion and high concentration effect. Furthermore, variants using RTP have deeper junction depth and higher surface doping concentration in the contact areas as well. It is due to higher annealing temperature of RTP (1050^{o}C), the solid solubility and the diffusion coefficient of boron is higher at this temperature than in the cases of furnace annealing at 900^{o}C. Deeper junctions would be obtained if the transient enhanced diffusion effect was implemented in the implantation variants. The deeper junction depth and higher surface concentration in the source/drain Si can give better and more reliable contacts with metal layers. Therefore, variants 2,3 and 4 would be more favorable from the technology point of view.

However, electrical characteristics of variant-2 transistor is comparable to variant-1 one and a even little worse as shown in figure 5-15 and figure 5-16 below. The reason is that the junction depth under the poly-spacer is a little bit shallower for RTP than in FA and the drain resistance is little higher for variant 2 than variant 1. Therefore, the output current is lower for variant 2. This simulation results are confirmed by device measurements in section 6.5. The junction depth is deeper for implanted cases and higher output current is shown as expected.



(a) *(b)*

Figure 5-15: a) Linearly scaled I_d-V_g characteristic and b) subthreshold characteristic of simulated transistors from different variants

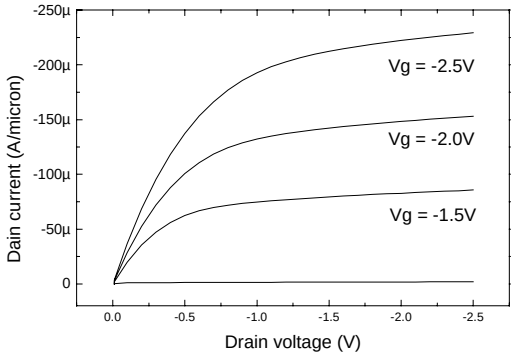


Figure 5-16: Output characteristic of simulated transistor fabricated according to variant 1

5.3.5. Summary

It has been shown that simulation is a very effective tool for optimization of a new technology. It is possible to find a compromise between various demands in order to create the best overall performing transistor. Implanting the mono silicon with $2.8 \cdot 10^{13}$ arsenic atoms at 250 kV leads to a reproducible device taking into account the deviations in the process. Normal furnace anneal of 17 minutes at 900°C is necessary to ensure a diffused junction edge under the gate. The reduced resistance in the source drain area is especially important for using these transistors in advanced circuits. Therefore, three other variants offering deeper source and drain junction depths were also simulated. Device characteristics can be better with different channel implantation because this was only optimized for variant 1 with furnace annealing.

5.4. Fabrication of sub-micron P-MOSFET

5.4.1. Sub-micron lines formation by photoresist ashing

5.4.1.1. Photoresist ashing technique

The decrease of feature length requires new lithography technologies. Industrial projection printing has evolved from Ultra-Violet (UV) light to deep UV light to meet the requirements. The far field diffraction is proportional to the wavelength of the light, as is the depth of focus, which demands the use of smaller wavelengths for photo-lithographic exposure. Modern wafers steppers use an excimer source laser for features

of 200-300 nm. For smaller dimension, e.g. sub 0.1 μm, electron beam or EUV lithography is needed.

Our lithography facility at the moment that this process was being realized can make only minimum feature sizes of 1 μm. Therefore, in order to make submicron gates a technique called resist ashing [18] was implemented for our prototype devices. The ashing technique is described in figure 5-17.

In this technique, patterned resist width is decreased by ashing of the resist in an oxygen plasma. First, the resist is exposed and developed as normally, see figure 16a, using conventional contact printing. The initial height of the resist (S1818) is about 1.8 μm while several patterns are made with an initial resist width of 1 to 2 μm. After that the resist is slowly burned away in an isotropic oxygen plasma, figure 16b. As the resist consists of polymer chains the ashing products are mainly H_2O and CO_2. Because the plasma chamber is at a temperature of more than 100^oC both gasses don't precipitate on process wafers. Both the width and the height of resist patterns will be reduced according to the ashing rate until a much smaller shape remains, see figure 16c. The remaining resist can now be used as mask for further processing. It is noticed that the reduced resist height should be large enough to withstand the chemicals used to pattern the underlying layer.

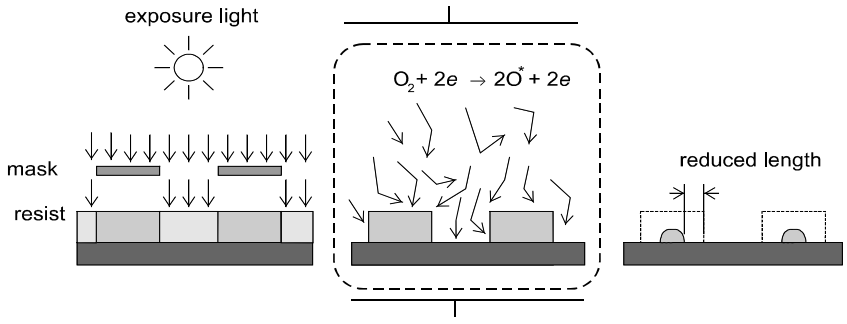


Figure 5-17: Reduction of the gate using resist ashing

5.4.1.2. Resist Ashing Results

The influence of process parameters on the ashing process was investigated. A typical proven recipe for resist ashing using O_2 plasma barrel etcher is listed as follows:

- Pump out till the pressure is less than 40mTorr.
- Purge with nitrogen till 5 Torr.
- Pump out till 40mTorr
- Stabilise pressure control at 2 Torr with nitrogen flow.
- Warm up till 85^oC using 750 Watt RF power.
- Removal of nitrogen until 40mTorr.
- Stabilise the pressure control at 600mTorr with oxygen.
- Resist ashing for a required time ***T*** with 150 Watt RF power under continuous flow of oxygen.
- Removal of residual gases till 40mTorr.
- Purge nitrogen into chamber combined with fast pump out during one minute

- Again purge nitrogen combined with slow pump out during five minutes.
- Fill chamber till 1 bar (air pressure) with nitrogen

(A more detailed description of the ashing recipe can be found in appendix 5)

Location of wafer in the oxygen plasma chamber is one of the important factors influencing the ashing rate. Due to the construction of the oxygen plasma etcher an influence of the position can occur.

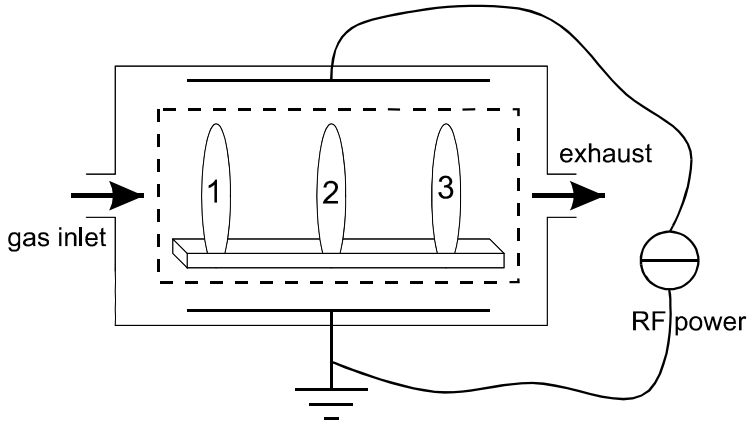
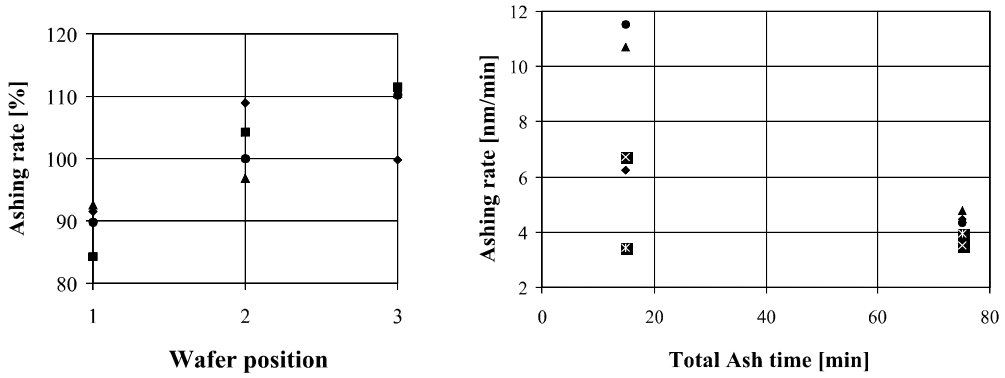


Figure 5-18: process wafers are located in oxygen plasma etcher for photoresist ashing



(a) *(b)*

Figure 5-19: The ashing rate as function of a) the position in the reactor chamber and b) the total process time.

Wafers in the front of the fresh oxygen stream are surrounded by a higher active oxygen concentration and less resist ashing products, while wafers in the exit of the etcher are situated in a lowered active oxygen concentration and an increased amount of ashing products. In figure 5-19a the ashing rate at three positions as in the following figure is plotted as percentage of the mean ashing rate. This experiment was conducted with different prebaking condition of photoresist and warming up the barrel etcher at the beginning stage which are presented in different kind of dots in the figure 5-19.

The dependence of etching rate on the total ashing time is shown in figure 5-19b. The difference in ashing rate can originate from several causes. At first, the large heat capacitance of the wafers reduces warming up, resulting in a lower temperature at the start of the ashing process. Secondly, the chemical compound of the resist could be altered slightly during the first ashing process, causing height alteration. Also the variation in resist height across a small surface or measurement errors can contribute to

this deviation. However, these influences became less when the ashing time became significantly longer. As a result, the scattering of the ashing rate became less as illustrated in figure 5-19b.

The following is the picture take from a submicron MOSFET gate after ashing for *T*=75 minutes using the ashing recipe shown in the previous section.

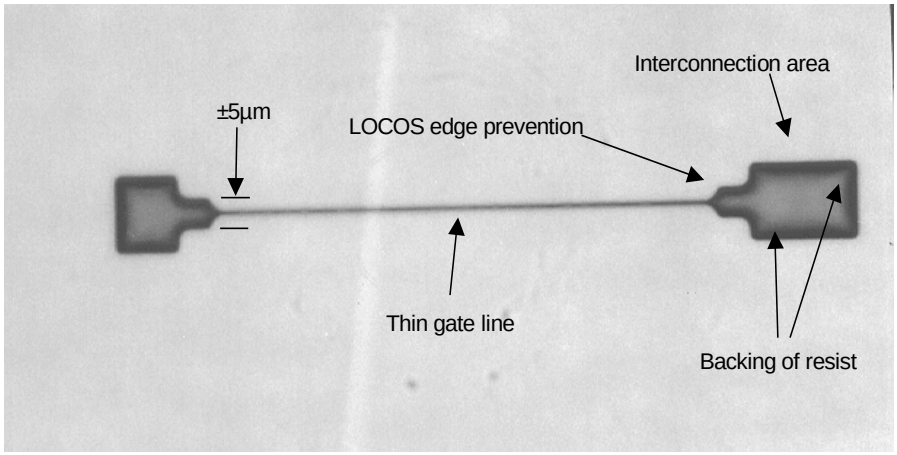
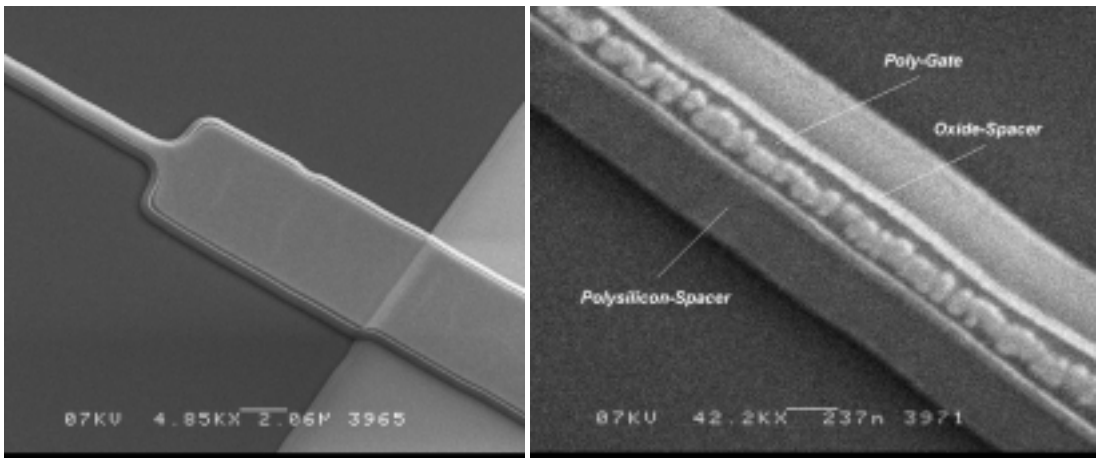


Figure 5-20: Submicron resist line of a MOSFET gate after ashing for 75 minutes



a) *b)*

Figure 5-21: SEM pictures of submicron poly-spacer MOSFET gates

In our process, the submicron PMOSFET is designed to use one mask for the smallest dimension, the gate, while other parts, like the spacers, are formed self-aligned. The shape of the poly silicon gate therefore needs to be highly rectangular. In figure 5-21 two Scanning Electron Microscope (SEM) pictures are shown. These SEM photos were taken during the processing of the device wafers. The gate poly silicon was etched and the resist on the gate pattern was removed. In the left photo, i.e. figure 5-21a, a submicron gate overlapping the LOCOS is illustrated. A SEM picture of a 0.2 µm gate is presented in figure 5-21b. There is a poly gate with a thin layer of silicon-oxy-nitride shown in the middle of the structure. The white band next to the poly-gate is the 50 nm oxynitride spacers. Then the two next gray bands are the poly-spacers formed after anisotropic etching of 300 nm poly. The rest is the source and drain silicon area.

Figure 5-22 shows a Focus Ion Beam (FIB) cross-section picture of a PMOS transistor after removed all of the top metal and oxide layer. The gate length of this transistor was

2 μm in the design before ashing. This gate length was reduced to approximately 1.2 μm after the photoresist ashing and etching processes. The poly-spacer widths were also less than 300 nm as compared with the SEM picture in figure 5-21b. This was caused by over-etching during removal of the silicon-oxy-nitride layer. The oxide-spacers could not be seen in this picture due to low contrast. The Si surface is rougher in this picture due to a decoration etching processes, e.g. for metal and oxide removal.

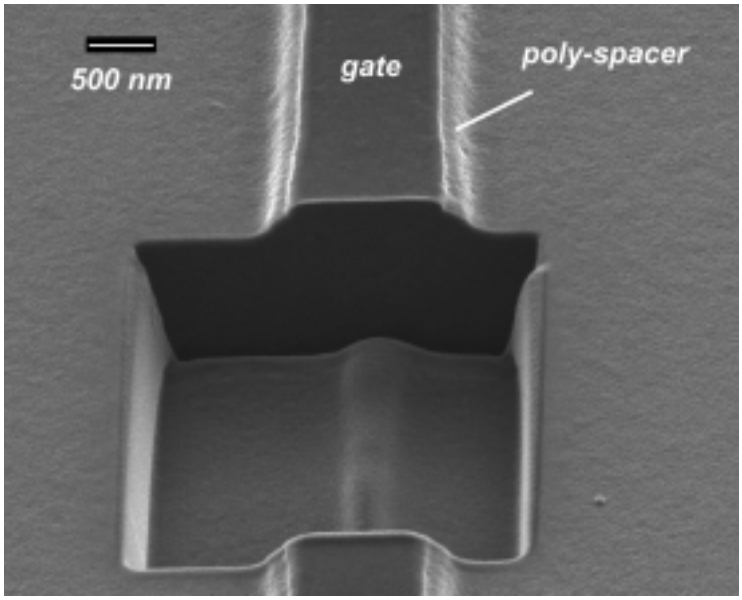


Figure 5-22: Focus Ion Beam (FIB) cross-section picture of a poly-spacer PMOS transistor

The pictures have shown that deep submicron gates were successfully fabricated by the simple resist ashing technique. However, one big disadvantage of this technique and with our equipment is the non-uniformity of the ashing rate over a wafer. The SEM pictures were taken on gates located at different positions on the process wafer. Although the starting gate lengths all were 1μm, after ashing they are different. The ashing rate is significantly higher near the edge of the wafer because the active oxygen concentration and temperature are higher near the edge than in the middle of the wafer during ashing. Therefore, much shorter gates are obtained at the edge of a wafer. This non-uniformity effect can be observed quite easily by normal optical microscope. Many gates were removed completely near the edge of the wafer. Results of measured effective channel lengths of devices which will be presented in the device characterization section also show this effect.

5.4.2. Poly-spacer PMOSFET process

In this section, a short description of the process flow is presented to explain the construction of the sub-micron poly-spacer PMOSFET.

5.4.2.1. Active Area

The realization of the process start with a blank n-type <100> oriented silicon wafer.

- A pad oxide of 25 nm is grown, followed by a 50 nm low-pressure chemical vapour deposition (LPCVD) of silicon nitride.
- The first mask active-area (AA) is defined using contact printing and standard definition procedure.
- Uncovered silicon nitride is removed using a CF_4/O_2 plasma.

- LOCOS with a thickness of 700 nm is grown using the standard UT-BiCMOS process step.
- The resist, the oxidised silicon nitride, the silicon nitride layer itself and the pad oxide are removed using several etching steps. The resulting structure is depicted in figure 5-23a.
- A sacrificial oxide is grown of 25 nm.

5.4.2.2. Well formation

- The super steep well is implanted using a normal UT-BiCMOS n-well implantation, an intermediate implantation at 300 keV, both using phosphorus, and an arsenic implant of $2.8 \cdot 10^{13}\,cm^{-2}$ at 250 keV.
- Lattice damage is removed in a N_2 environment at 800°C for 30 minutes.

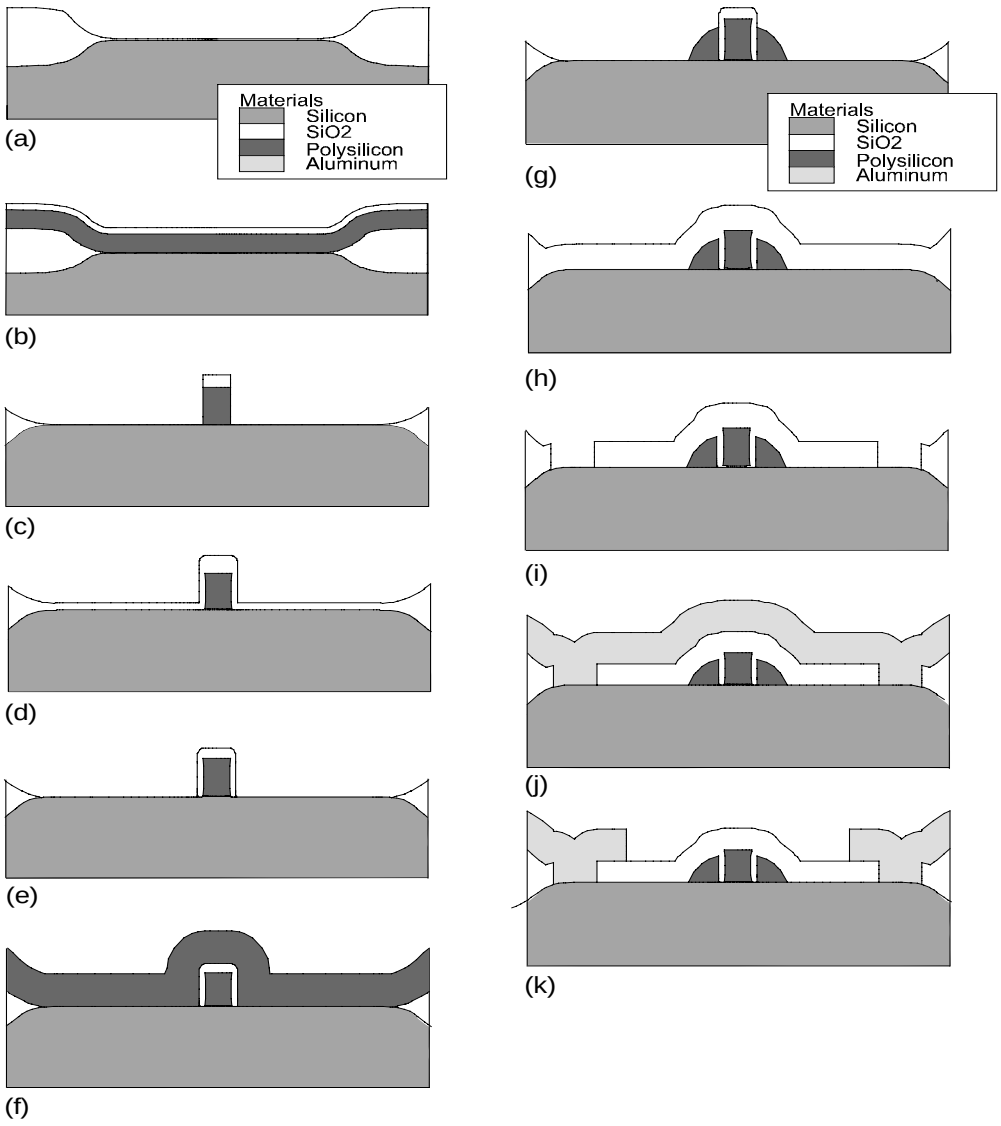


Figure 5-23 a.k: PMOSFET process flow

5.4.2.3. Polysilicon gate formation

- Sacrificial oxide is removed, directly followed by growing of a gate oxide of 8 nm thickness.
- 300 nm of undoped α-silicon deposited using LPCVD at 550^{o}C and then a layer of 50 nm LPCVD oxynitride is deposited on top of it, see figure 5-23b.
- The α-silicon gate pattern is defined using the PS-mask. Resist S1818 is used without baking to prevent deforming of the resist shape. Resist ashing is done to create submicron patterns.
- The silicon oxynitride and the α-silicon are etched anisotropically using a $Cl_2/SiCl_4$ plasma with end point detection (EPD) to stop over-etching of the gate oxide, see figure 5-23c.
- The α-silicon gate is oxidised in an O_2 ambient at 850°C for 30 minutes and 20 minutes in nitrogen. The α-Si gate is also converted into polysilicon during this oxidation step.
- A 50 nm silicon oxynitride layer is deposited, which will form the spacer between gate and drain/source, see figure 5-23d.
- The oxynitride is etched anisotropically using a CF_4/O_2 plasma and EPD as to control the process, see figure 5-23e.

5.4.2.4. Shallow junction source/drain formation

- 300 nm undoped α-Si is deposited, resulting in the structure depicted in figure 5-23f.
- Anisotropical etching of the α-Si using EPD creates poly source and drain spacers (figure 5-23g).
- The source-drain-disconnection (SD) mask is defined, wherefore again poly silicon is etched. This is done to disconnect the source and drain, see paragraph 6.4.3.2.
- Silicon oxynitride on top of the gate is removed using a CF_4/O_2 plasma. (more selectively etching to silicon-oxide-nitride is preferred)
- A spin-on-dopant (SOD) layer of 200 nm is deposited, see figure 5-23h. For implantation variants, this step is replaced by BF_2 implantation $5x10^{15}$ ions/cm^2 at 50 keV.
- The boron is diffused using normal furnace anneal (900°C, 17 min.) or rapid thermal anneal (1050°C, 40 sec.).
- The SOD is removed by BHF dipping and boiling HNO_3.

5.4.2.5. Contact holes formation

- 500 nm of APCVD oxide is deposited for passivation
- SiO_2 layer is densified at 650^{o}C in O_2 and subsequently at 500^{o}C in wet nitrogen to improve the gate oxide
- Contact holes are defined with resist using the CO-mask and oxide etching in BHF is carried out for contact holes formation. (This etching has to be done in the dark to prevent staining of the highly boron doped silicon)

- Standard cleaning step is carried out and a 1 % HF etch step of 1 minute to remove the interface oxide (also in dark), see figure 5-23i.
- A 70 nm thick titanium-tungsten alloy is deposited, followed by 1µm aluminium layer, see figure 5-23j.
- Photoresist patterns are defined using interconnect mask (IN)
- Uncovered Al is removed by wet etching in H_3PO_4 solution and then the TiW layer also is removed using H_2O_2 (40 %) solution. The resulting structure is drawn in figure 5-23k.
- Finally the wafers are annealed in a wet N_2 environment at 400°C for 10 minutes for alloying the metal/Si contact to decrease the contact resistance.

5.4.3. Mask set up

5.4.3.1. Mask requirements

The fabrication of submicron structures using conventional lithography has some disadvantages in comparison with a full submicron process. The alignment of different masks has a precision of around 1 to 2µm, therefore the significant parts of the PMOSFET structure must be made with one mask on submicron scale while the other parts can be realized on conventional scale using different masks. As the masks have the same critical dimensions as the normal UT-BiCMOS process, the same design rules are valid. The PS-mask, defining the gate, needs to be very smooth in comparison to the other masks. Disturbances along an initial gate resist pattern will result in an equivalent rough submicron line after ashing.

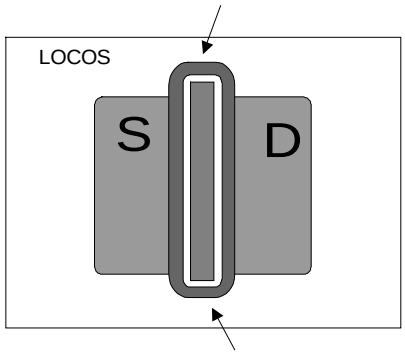


Figure 5-24: PMOSFET top view, arrows point at the disconnection for the next step

5.4.3.2. Source-drain disconnection

Source-drain disconnection is a special step for our poly-spacer PMOS process. As mentioned in paragraph 6.4.2.3, back etching of the poly silicon source and drain is followed by the same etching step to disconnect source and drain. This step is clearly necessary as seen in figure 5-24. The source and drain are indicated with S and D, while in the center a gray rectangular is drawn representing the gate. Around the gate the silicon oxynitride, depicted as white material, and the poly silicon, depicted as dark gray, are present. Due to this encircled pattern the source and drain are electrically connected. The SD-mask is designed to cover the active area while outside this region the wafer is uncovered. After defining this pattern source and drain will be disconnected by the second poly etch.

Table 5-2: Short description of the five designed masks

Colour	Name	Name (full)	Orientation	Comment
	AA	Active Area	Positive	To define transistor outline of source and drain (etch nitride for LOCOS).
	PS	Poly Silicon	Positive	Defines the gate, with widths of 1 up to 2 µm, which will be reduced with resist ashing.
	SD	Source-Drain protect	Positive	To prevent source and drain from etching during removal of source drain polysilicon short circuit.
	CO	Contact holes	*Negative*	Defines the areas which should be etched to create interconnections.
	IN	Interconnect	Positive	To connect transistors and test pads with aluminium.

5.4.3.3. A layout example

To explain the function of each mask a layout example of a medium sized transistor is drawn in figure 5-25. The active area, AA-mask, contains the real PMOSFET structure including the contact holes. The poly silicon gate, PS-mask, start at the interconnect electrode of the gate (bottom electrode) and continues outside the active area. Disconnection of the source drain poly can now be realized, using the SD-mask. The definition of the interconnection, CO-mask, follows logically as does the interconnect definition using the IN-mask.

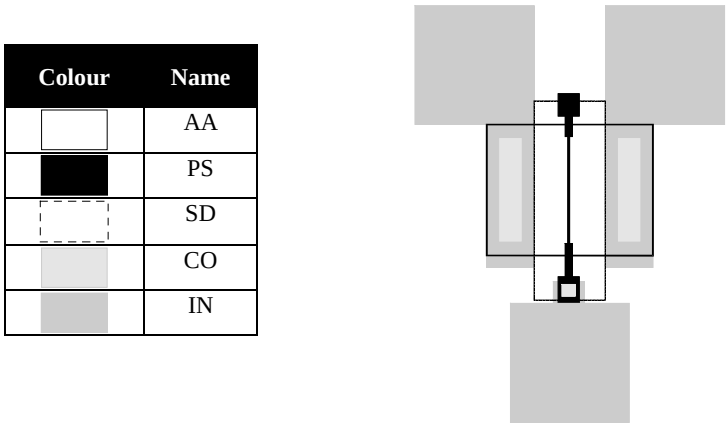


Figure 5-25: A medium sized PMOSFET layout

5.5. Results and discussion

Submicron PMOS transistors have been realized in the MESA clean room and characterized. Poly-spacer PMOS transistors of the four simulated variants were prepared. Two of the variants using SOD for boron doping were successful while the other two implantation variants were not because of boron implantation through the gate oxide. Therefore, in this section only the results from SOD variants are shown.

Determination of effective gate lengths of PMOS and their distribution will be discussed first. Electrical characteristics of the devices, e.g. capacitance-voltage curves of the oxide gate, transconductance and output characteristics will be presented thereafter.

5.5.1. Submicron effective gate lengths measurement

Electrical effective gate lengths of PMOSFETs were measured using the "Shift and Ratio" method [19]. Table 5-3 shows the gate length (L_{eff}) measured from transistors with different initial gate lengths before ashing (SOD-FA variant) located in different dies over the wafer.

Table 5-3: Effective gate lengths of SOD-FA PMOSTs W=110 μm measured by "Shift and Ratio" method

L_{eff} (μm)	Die Number											
L_{before} (μm)	1	2	3	4	5	6	7	8	9	10	11	13
1	0.31	0.32	0.34	0.29	0.45	0.34	0.22	0.56	0.20	0.00	0.00	0.00
1.2	0.37	0.39	0.50	0.39	0.51	0.41	0.30	0.49	0.27	0.00	0.00	0.16
1.4	0.56	0.55	0.69	0.49	0.71	0.59	0.44	0.59	0.55	0.23	0.19	0.36
1.6	0.79	0.83	0.90	0.74	0.94	0.89	0.61	0.98	0.74	0.41	0.39	0.57
1.8	1.02	1.03	1.12	0.94	1.20	1.12	0.87	1.26	1.01	0.61	0.55	0.76
2	1.26	1.24	1.22	1.13	1.35	1.26	1.03	1.30	1.14	0.82	0.72	0.98

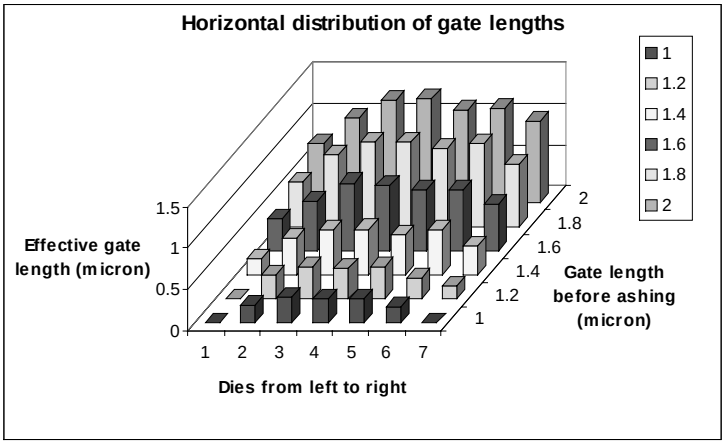


Figure 5-26: Horizontal distribution of gate lengths of PMOS transistors on across the SOD-FA variant wafer

This measurement method has the advantage of source/drain resistance and channel mobility independence. Furthermore, using this method a gate length of a submicron transistor can be extracted, using a reference transistor with very long gate length of tens of micron of which L_{eff} can be measured accurately by optical microscopy. The error of this method is only about 10 nm for a 0.25 μm gate length which is extracted from a 60 ± 1 μm gate length reference device. Therefore, it is not necessary to measure all the physical gate lengths as in the total source/drain resistance measurement method [20]. The results from the "Shift and Ratio" method were compared with the total resistance measurement method and they were in very good agreement.

As seen in table 5-3, effective gate lengths around 0.20 μm can be obtained with initial 1 μm gate length before photoresist ashing. The gate lengths are much shorter in the

dies located near the edges of the wafer due to ashing non-uniformity. Positions of the dies on the Si wafer and the gate lengths distribution across the wafer are illustrated in figure 5-26.

5.5.2. Electrical characterization

5.5.2.1. Capacitance-Voltage (C-V) characteristics

Transistors have been used for C-V characterisation, and several parameters can be extracted from its C-V curve. Figure 5-27 shows a capacitance model of a poly-spacer PMOS transistor. Beside the capacitor of thin gate oxide C_g , there are two parasitic capacitances of oxide spacers C_{sp} due to the P^+ poly-spacers.

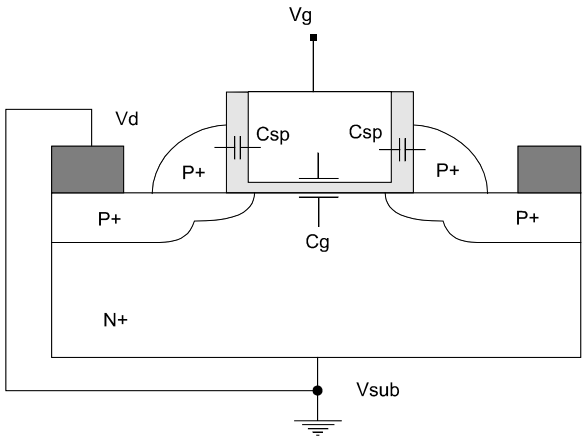


Figure 5-27: Capacitances in a poly-spacer PMOSFET

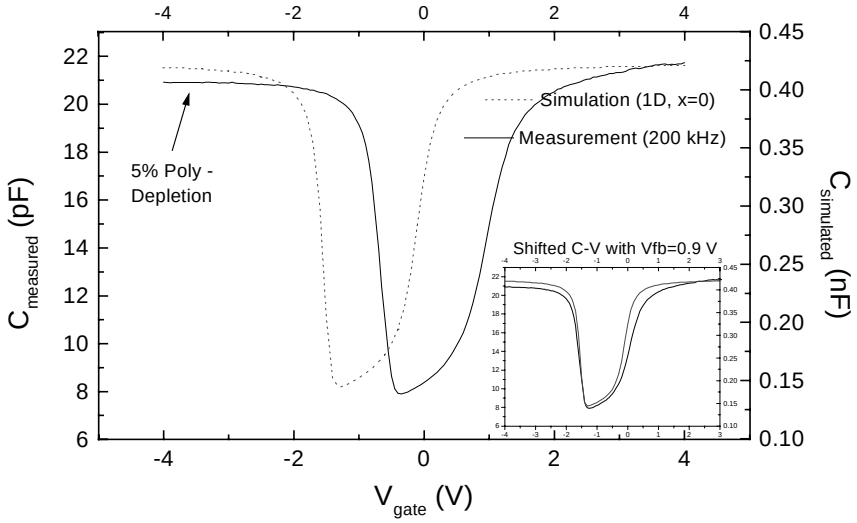


Figure 5-28: High Frequency C-V curve of a PMOST 40x140 μm^2

High-Frequency (HF) C-V measurements have been carried out on PMOSFETs with long gate lengths of 40 µm, 49 µm and 60 µm (widths of 140 µm). Figure 5-28 shows a typical HF (200 KHz) C-V curve of a PMOST measured with the connection as shown in figure 5-27 and compared with a 1-dimensional simulated C-V curve of the gate.

The 1-dimensional C-V curve was extracted from the simulated device (variant 1) at the middle point of the gate with an assumption that Φ_{MS}=0 (metal-semiconductor work function difference). The non-uniformity of the channel doping profile was also taken

into account for the C-V simulation. This curve looks similar to a low frequency C-V curve of a capacitor. It is because the source and drain can supply enough minority carriers in the inversion region of the substrate. The thickness of the gate oxide was extracted from the accumulation capacitance and it is about the same 8.8 nm for different gate sizes for both RTP and FA variants. This gate oxide thickness is higher than the 8 nm oxide as we expected because of a finite thickness of inversion and accumulation layers.

Depletion effect of the P^+ doped poly gate is also illustrated by a lower measured inversion capacitance value compared to the corresponding theoretical value. Our devices have a small poly gate depletion effect of less than 5 % capacitance which was calculated from the percentage of difference between the measured capacitance and the theoretical capacitance divided by that theoretical capacitance in the inversion region. The extracted flat-band voltage is about 0.9 V which is corresponds to the theoretical value for a P^+ poly gate. In figure 5-28, the small figure shows the same C-V curves in which the measured C-V curve was shifted by V_{FB}=0.9 V. The C-V curves look very good showing very high quality PMOS gate.

5.5.2.2. Threshold voltages and subthreshold characteristics

Threshold voltages of PMOS transistors were derived using method of G. Ghibaudo [21]. Using the following equation, the threshold voltage can be obtained by linear extrapolation of the Y(Vg) curves:

$$Y(V_g) = \frac{I_d}{\sqrt{g_m}} = \sqrt{\beta V_d}\,(V_g - V_T)$$

where, Id is the drain current, g_m is the transconductance in strong inversion ($g_m=\delta I_d/\delta V_g$); $\beta=WC_{ox}\mu_o/L$ is the transconductance parameter with W and L are the effective width and length of the gate, C_{ox} is the gate oxide capacitance, μ_o is the low field mobility; V_g and V_T are the gate- and threshold voltage respectively.

For the SOD furnace annealing variant, the threshold voltages are -0.71 V for MOSFETs with long channels (L_{eff} > 0.25 μm). V_T value is rolling off when the effective channel length is decreased to 0.20 μm as shown in figure 5-29. The same roll-off behaviour of threshold voltage is also observed with PMOS transistors of the RTP variant.

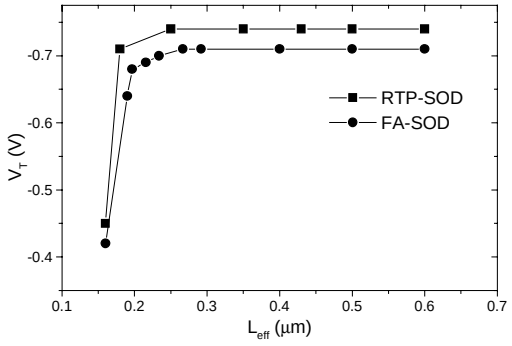


Figure 5-29: Threshold voltages of deep submicron PMOS transistors

The threshold voltage of transistors from the RTP variant is found to be lower than from the FA variant transistors, e.g. V_T is –0.74 V compared with –0.71 V. It is possible that the surface doping concentration of the channel is higher for the RTP variant transistor due to higher annealing temperature of RTP, i.e. 1050°C in RTP and 900°C in the FA

variant. The other possibility for the difference in V_T might be due to a difference in the work-function of the P^+ gates which were activated at different temperatures [22].

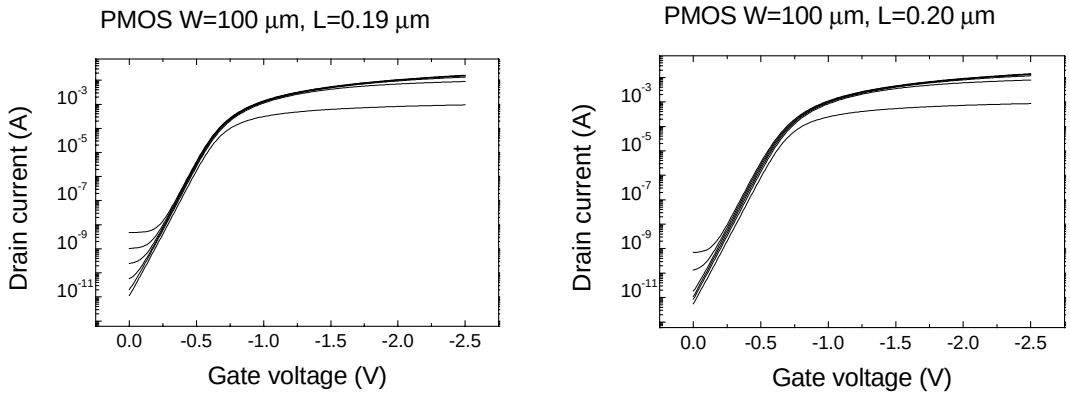
In table 5-4, effective gate lengths (L_{eff}) of transistors and their threshold voltages (V_T), subthreshold swings (S), transconductance parameters (β) and mobility reduction factors (Θ) are given.

Table 5-4: parameters of PMOS transistors (W=100 μm, die 9) from FA and RTP variants

	FA-PMOS					**RTP-PMOS**				
R_{SD}	20 Ω					14 Ω				
$L_{pre\text{-}ash}$ (μm)	L_{eff} (μm)	V_T (V)	S (mV/dec.)	β (mA/V)	Θ (V^{-1})	L_{eff} (μm)	V_T (V)	S (mV/dec.)	β (mA/V)	Θ (V^{-1})
1	0.19	-0.672	95	23.39	0.67	0.18	-0.699	97	22.98	0.50
						0.20	-0.707	96	20.09	0.47
1.2	0.29	-0.691	95	15.56	0.49	0.41	-0.736	95	10.39	0.33
1.4	0.54	-0.702	94	8.40	0.34	0.60	-0.743	95	6.99	0.29
1.6	0.84	-0.705	94	5.31	0.29	0.79	-0.745	95	5.18	0.25
1.8	1.02	-0.711	94	4.44	0.26	1.00	-0.748	95	4.09	0.24
2.0	1.20	-0.715	94	3.75	0.25	1.20	-0.746	95	3.38	0.22

In this table $L_{pre\text{-}ash}$ means gate length before the photoresist ashing. As seen in the table, the transconductance factors of the FA PMOS transistors are better than the RTP PMOS. However, the mobility reduction factors are less in the RTP PMOS because of their lower source/drain resistance R_{SD}.

Figure 5-30 shows good subthreshold characteristics of PMOS transistors with effective gate lengths of about 0.20 μm for the two variants. The physical gate lengths of those transistors are believed to be 0.25 μm as expected by simulation.



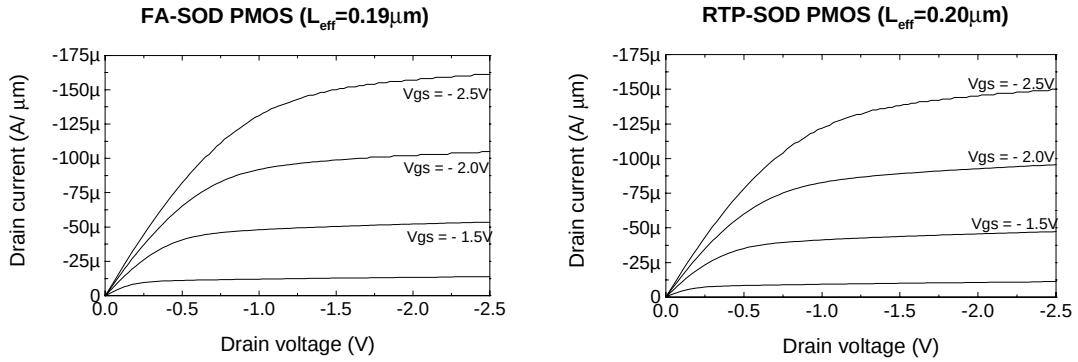
a) *b)*

Figure 5-30: Subthreshold characteristics: a) FA-SOD PMOS L_{eff}=0.19μm,W=100μm; b) RTP-SOD PMOS L_{eff}=0.20μm,W=100μm

5.5.2.3. Output characteristic

The output characteristics of 0.25 μm PMOST are presented in figure 5-31. The on-current of those transistors is about 160 μA/μm. This value is smaller than the simulated I_{on} of 230 μA/μm. This is because of the thinner simulated gate oxide thickness (8 nm). Furthermore, the simulation of boron might be not correct and the source/drain junction depths could be shallower for the realized devices. Mobility model of device simulation may also over estimated the drive current of the devices.

The higher threshold voltage in our device compared with simulation also reduces the current drive. If we do a simple extrapolation from our device to a device with 5.5 nm gate oxide thick and V_T=0.5 V a current of about 300 μA/μm is obtained, the maximum drive current that can be obtained from a state-of-art PMOS device.



a) *b)*

Figure 5-31: Output characteristics: a) FA-SOD PMOS L_{eff}=0.19μm,W=100μm; b) RTP-SOD PMOS L_{eff}=0.20μm,W=100μm

As predicted by simulation, the on-current of RTP-variant transistors is a little less than the FA one. The on-current values of different gate lengths of the two variants are compared in figure 5-32.

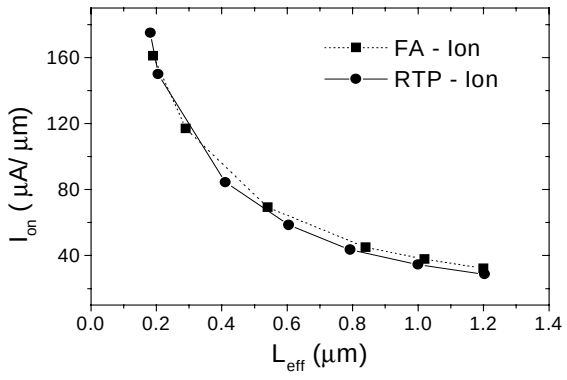


Figure 5-32: On-current of PMOS transistors from two variants

PMOS Off-current values were also measured. Table 5-5 shows the values of I_{on}, I_{off} and their ratio of PMOS with different gate lengths from two variants. FA-PMOS transistors have higher current drive and also higher leakage current. The lower values of off-current from RTP-PMOS may be due to less boron penetration from poly-gate and higher channel surface doping concentration. However, the off-currents in our devices also depends on the quality of the source/drain to substrate diodes which have

much large areas than the gate areas. This diode leakage current may dominate the off-current.

Table 5-5: Output currents of PMOS transistors of two variants (V_{supply}=2.5 V)

FA-PMOS				**RTP-PMOS**			
L_{eff} (μm)	I_{on} (μA/μm)	I_{off} (pA/μm)	I_{on}/I_{off}	L_{eff} (μm)	I_{on} (μA/μm)	I_{off} (pA/μm)	I_{on}/I_{off}
0.19	161	40	$3.9x10^6$	0.18	175	5	$3.6x10^7$
				0.20	150	6	$2.5x10^7$
0.29	117	5	$2.5x10^7$	0.41	84	4	$1.9x10^7$
0.54	69	12	$5.6x10^6$	0.60	58	6	$9.7x10^6$
0.84	45	40	$1.1x10^6$	0.79	43	5	$9.1x10^6$
1.02	38	3	$1.2x10^7$	1.00	35	5	$7.2x10^6$
1.20	32	30	$1.1x10^6$	1.20	28	5	$6.0x10^6$
Average	-	21	$8.2x10^6$	Average	-	5	$1.6x10^7$

5.5.3. Summary

A new poly-spacer PMOS device concept has been tested by simulation and realization. The photoresist ashing technique has been investigated using an O_2 plasma barrel etcher. Gate lengths of different sizes from 0.1 μm to 1 μm have been realised. All the FA-SOD and RTP-SOD PMOS transistors showed good electrical characteristics. The variant using furnace diffusion resulted in better PMOS transistors with higher threshold voltage, higher transconductance and higher output current. However, rapid thermal processing gave rise to transistors with lower source/drain resistance and lower leakage current. All the transistors may have thicker gate oxide and higher surface channel doping concentration than expected. Consequently, it resulted in lower threshold voltages and lower output current compared to simulated values.

5.6. Conclusions and recommendations

Poly-spacer 0.25μm PMOS transistors have been designed and realized successfully using SOD as a diffusion source. Channel implantation was optimized using simulation tools and four variants of process flow were produced. Better transistor characteristics may be obtained using rapid thermal annealing and BF_2 implantation but the channel optimization has to be further investigated. Decrease of the gate oxide thickness to lower than 6 nm can improve significantly the transistor characteristics. If so, the whole process flow has to be re-simulated and optimized again.

Deep submicron PMOS transistors have been realized and characterized. Quarter-micron poly-spacer PMOS has been obtained using SOD as a diffusion source for shallow-junction source/drain formation. Realized transistor characteristics are good but can be extensively improved still. There are some processing steps that must be modified for better devices and a more reliable process:

- The mask PS for poly gate definition can be redesigned for shorter starting gate lengths, e.g. from 0.8 μm or less up to 1.5 μm with difference of 0.1 μm between

adjacent gates. Therefore, the resist ashing time can be reduced and more uniform submicron resist lines can be obtained.

- The mask SD for source/drain poly-spacer disconnection also can be modified to protect other structures, e.g. capacitors and Kelvin structures, which are unwanted to be exposed to this etching process.
- Gate oxide can be grown thinner to 6 nm or less
- Deposition of 50 nm SiON layer on top of the poly gate may not be necessary because the resist layer after ashing is still thick enough, i.e. thicker than 1 μm, for the poly gate etching step. Furthermore, in the case of no SiON layer the removal of this layer is not required any more and this plasma etching step (see figure 21g) after the poly-spacer etching can be canceled. As a result, more reliable devices and a more reproducible process can be achieved. In the case this layer is still wanted for some reasons, the 50 nm SION layer can be replaced by a 50 nm SiO_2 layer or Si_3N_4 layer for easier removal of this layer using wet etching.
- Selectively deposited polysilicon deposition on source/drain is more desirable for elevated source/drain MOSFET than the poly-spacer structure. Therefore, study on selectively polysilicon deposition should be carried out.

References

[1] J.E. Lilienfeld, U.S. Patent No. 1745175, 1930

[2] O. Heil, Bristish Patent No. 439457, 1935

[3] D. Kahng and M.M. Atalla, “Silicon-silicon dioxide field induced surface devices”, IRE solid-state devices Res. Conf., Carnergie institute of technology, Pittsburg, Pa., 1960. D.Kahng, “A historical perspective on the development of MOS transistors and related devices”, IEEE trans. Electron Devices, ED-23, 1976, p. 655

[4] Athena and Atlas user’s manuals, Silvaco International 1997

[5] T. Hori, “Gate dielectrics and MOS ULSIs, principles, technologies, and applications”, Springer 1997

[6] R.H. Dennard, IEEE Journal SC-9, 1974, p. 256

[7] G. Baccarani, IEEE Trans. ED-31, 1984, p. 452

[8] C.G. Sodini, IEEE Trans. ED-31, 1984, p. 1386

[9] The National Technology Roadmap for Semiconductor Technology Needs, Semiconductor Industry Association 1997

[10] J.R. Brews, IEEE EDL-1, 1980, p. 2

[11] Ken-ichi Goto, “A High Performance 50 nm PMOSFET using Decaborane (B_{10} H_{14}) Ion Implantation and 2-step Activation Annealing Process”, IEEE IEDM (1997) Digest

[12] Hitoshi Wakabayashi, “A high performance 0.1 μm CMOS with elevated salicide using novel Si-SEG process”, IEEE IEDM (1997) Digest

[13] S. Ogura, et al., “Design and characteristics of the lightly doped drain-source (LDD) insulated gate field effect transistor”, IEEE Trans. Electron. Dev. 27(8), 1980, p. 1359

[14] T. Hori et al., “Deep-submicrometer large-angle-tilt implanted drain (LATID) technology”, IEEE Trans. Elec. Dev., 39 (10), 1992, p. 2312

[15] C. Codella and S. Ogura, “Halo doping effect in sub-micron DI-LDD device design”, IEDM Tech. Dig. , 1985, p. 230

[16] Hiroshi Iwai, "Recent advances and future trends of ULSI technologies", proceeding of 26th European solid state device research conference 1996, p. 45

[17] Lombardi et al., "A physically base mobility model for numerical simulation for non-planar devices", IEEE Trans. on CAD, 1988, p. 1164

[18] J.Chung et al., *Deep-Submicron MOS Device Fabrication Using a Photoresist-Ashing Technique*, IEEE Electron Device Letters, Volume 9, no. 4, April 1988, p. 15

[19] Yuan Taur, "A new "shift and ratio" method for MOSFET channel-length extraction", IEEE Elec. Dev. Lett. Vol. 13 No. 5, 1992, p. 267

[20] K. Terada, "A new method to determine effective MOSFET channel length", Jap. J. Appl. Phys. Vol. 18, No. 5, 1979, p. 953

[21] G. Ghibaudo, Electronics letters, Vol. 24, 1988, p. 543

[22] Takauki Aoyama et al., "Flat-band voltage shifts in P-MOS devices caused by carrier activation in P^+-polycrystalline silicon and boron penetration", IEDM 1997 proceeding, p. 627

6

Conclusions & Recommendations

Spin-on glass materials have been studied thoroughly for shallow junction formation in advanced semiconductor devices. Here are the main conclusions from this study.

- Spin-on glass materials were fabricated successfully using the sol-gel technology and the design of experiment method.
- Elements, concentration profiles and chemical bonding in spin-on dopant (SOD) layers have been characterised using advanced analysing techniques.
- Diffusion of boron and phosphorus from SOD into Si was investigated for shallow junction formation. Rapid thermal diffusion of SOD resulted in very shallow junctions in Si, i.e. less than 20 nm. However, high quality junctions were obtained only when the junction depth is equal to or greater than 50 nm. Diffusion using conventional furnace resulted in deeper junctions due to a large thermal budget.
- Diffusion of boron in Si from a SOD source was simulated using the doped oxide diffusion model. Compared with the doped oxide model, very high diffusivity of boron in the SOD layer was obtained due to the nature of SOD material.
- Diffusion of boron and phosphorus from SOD into polysilicon/Si and α-Si/Si was investigated for low series resistance shallow junction formation. Very high quality shallow junctions were obtained using both furnace and rapid thermal diffusion. Furnace diffusion resulted in better quality diodes than rapid thermal diffusion because the rapid thermal process created stress-induced defects in the PN junctions. Similar to the case of impurity diffusion directly into Si from SOD, 50 nm junction depth in the Si substrate was still the limit for a good quality poly-buffered junction. However, this kind of junction had the advantages of lower sheet resistance and a elevated contact structure which can be used for silicidation processes.
- In this diffusion technique, the random structure of as-deposited amorphous layer caused slower diffusion for both boron and phosphorus than in the as-deposited polysilicon layer with columnar structure. Therefore, a higher thermal budget was required to form a good shallow junction when using as-deposited amorphous silicon.
- High frequency bipolar junction transistor has been optimised and realized successfully using the UT-BiCMOS process. Transistors with a cut-off frequency of more than 10 GHz have been obtained. In one of the process variants, shallow junction poly-emitters were realized using SOD diffusion source. Compared with implanted poly-emitter, better quality transistors were obtained with SOD diffusion.
- A new concept of poly-spacers was proposed for future sub-0.1 μm MOSFET. This device has the elevated source/drain structure and low series resistance which are very important to obtain a high drive current. Furthermore, the technology is straightforward and reproducible.

- As a vehicle to test the concept, poly-spacer 0.25μm PMOS transistors were designed and realized successfully using SOD as a diffusion source. Although the device was not optimised yet, its electrical characteristics were very good and comparable with state-of-the-art devices. Above all, the new poly-spacer concept has been proven to be effective for future MOSFETs.

This simple SOD diffusion technique has proven to be very effective and promising for shallow junction formation in future ULSI technology. However, there is still much space for future research and device improvement. Here are the recommendations:

- More applications of SOD can be investigated, e.g. solar cells, sensors and actuators. Shallow junction is necessary to obtain high efficiency solar cells. On the other hand, a very deep junction is used for stop-etching in sensors and actuator technologies. SOD is a good material for these applications.
- Boron doped SOD and other thin film materials can be realized using sol-gel technology. There are so many applications in which sol-gel technology can be used. The sol-gel technology is simple and low cost. These advantages are very important for research & development in a developing country like Vietnam.
- Simulation of the impurity diffusion in Si and in polysilicon/Si from SOD can be investigated further. The non-equilibrium process of impurity diffusion from SOD into Si and polysilicon is really a challenge for research. Especially, diffusion of impurities in polysilicon is very dependent on the micro-structure of the poly layer and the type of the dopant itself, which makes it difficult to simulate. More research in the field of polysilicon deposition and micro-structure evolution at high temperatures has to be carried out first.
- Bipolar junction transistors can be redesigned using poly-contacted base structure and using SiGe base to increase the cut-off frequency.
- Sub-0.25 μm poly-spacer MOSFETs can be designed and realized using SOD diffusion or low-energy implantation. Masks for poly-gate definition can be redesigned for smaller dimensions, i.e. 1 micron or less. As a result, a deep submicron gate length, i.e. less than 0.15 μm, and better uniformity can be obtained using the photoresist ashing technique. Simulations have to be calibrated with impurities diffusion in polysilicon. Thin gate dielectric, i.e. Si-dioxide (< 6 nm) or other type of materials, should be studied. The poly-gate etching process has to be improved for better end-point detection on the very thin gate oxide layer.

APPENDIX 1

SSUPREM-4 Diffusion Model [1]

SSuprem4 is the state-of-the-art 1D and 2D semiconductor process simulator used in the design, analysis and optimisation of silicon fabrication technologies and device structures. It incorporates a wide range of advanced physical models for diffusion, implantation, and oxidation modelling.

The conventional furnace diffusion model of impurities and point defects in SSUPREM-4 is based on the diffusion theory as described in the beginning of this chapter. The form which is used to simulate impurity diffusion in SSUPREM-4 is:

$$\frac{\partial C_T}{\partial t} = \Delta\left[D_V C_A \frac{C_V}{C_V^*} \Delta \log\left(C_A \frac{C_V}{C_V^*} \frac{c}{n_i} \right) + D_I C_A \frac{C_I}{C_I^*} \Delta \log\left(C_A \frac{C_I}{C_I^*} \frac{c}{n_i} \right) \right] \qquad \textit{[Eq. 1]}$$

where:

C_I^* and C_V^* are the corresponding equilibrium values.

c can be n or p for specific impurities.

C_T is the total concentration.

C_A is the active concentration.

For intrinsic cases, $C_I = C_I^*$, $C_V = C_V^*$, and the intrinsic diffusivity D_I^{in} is obtained from the following equation:

$$D_i^{in} = D_I \frac{C_I}{C_I^*} + D_V \frac{C_V}{C_V^*} = D_I + D_V \qquad \textit{[Eq. 2]}$$

Above equation introduces the fraction interstitialcy factor F_I , which is impurity dependent, as

$$F_I = \frac{D_I}{D_I^{in}} = \frac{D_I}{D_I + D_V} \qquad \textit{[Eq. 3]}$$

In general, the diffusion coefficient D_I depends on impurity concentrations only when concentration is high. This dependence varies with the type of impurity. SSUPREM-4 models the concentration dependence of D_i^{in} using the vacancy-diffusion model expressed in the following equation:

$$D_i^{in} = D_i^x + D_i^- \frac{n}{n_i} + D_i^+ \frac{n_i}{n} + D_i^= \left(\frac{n}{n_i} \right)^2 + \ldots \qquad \textit{[Eq. 4]}$$

where

D_i^x , D_i^- , $D_i^=$ and D_i^+ are the intrinsic diffusivities of vacancy states neutral, single and double negative and positive, respectively.

n is the electron concentration that depends on all chemical concentrations of impurities and is approximately by:

---

[1] ATHENA user's manual, 2D Process Simulation Software, Silvaco International Inc., 1996, p.3-1

$$n = \frac{1}{2}\left[C_{net} + \sqrt{C_{net}^2 + 4n_i^2}\right] \qquad \textit{[Eq. 5]}$$

The activation energy relation describing the temperature dependence of the impurity of various vacancy states is of the form:

$$D_i = DIX.0 \cdot \exp\left(\frac{-DIX.E}{k_b T}\right) \qquad \textit{[Eq. 6]}$$

$$D_i^+ = DIP.0 \cdot \exp\left(\frac{-DIP.E}{k_b T}\right) \qquad \textit{[Eq. 7]}$$

$$D_i^- = DIM.0 \cdot \exp\left(\frac{-DIM.E}{k_b T}\right) \qquad \textit{[Eq. 8]}$$

$$D_i^= = DIMM.0 \cdot \exp\left(\frac{-DIMM.E}{k_b T}\right) \qquad \textit{[Eq. 9]}$$

Where ***DIX.0***, ***DIM.0***, ***DIMM.0***, ***DIP.0***, ***DIX.E***, ***DIP.E***, ***DIM.E***, and ***DIMM.E*** are accessible during simulation. Each impurity simulated in SSUPREM-4 uses only a subset of equation 4.

Boron Diffusion

Boron diffusion is assumed to take place primary via an interstitialcy mechanism with neutral and single positively charged defects. The diffusion equation is:

$$\frac{\partial C_T}{\partial t} = \Delta\left[D_V C_A \frac{C_V}{C_V^*}\Delta\log\left(C_A \frac{C_V}{C_V^*}\frac{p}{n_i}\right) + D_I C_A \frac{C_I}{C_I^*}\Delta\log\left(C_A \frac{C_I}{C_I^*}\frac{p}{n_i}\right)\right] \qquad \textit{[Eq. 10]}$$

Its intrinsic diffusivities are given by:

$$D_V = (1 - F_I)\left(D^X + D^+ \frac{p}{n_i}\right) \qquad \textit{[Eq. 11]}$$

$$D_I = F_I\left(D^X + D^+ \frac{p}{n_i}\right) \qquad \textit{[Eq. 12]}$$

For boron diffusion, the default value for $\boldsymbol{F_I}$ is 0.94.

Phosphorus Diffusion

The diffusion mechanism for phosphorus is not well understood. SSUPREM-4 model for phosphorus diffusion with explicit values for diffusivities for each point defect type and charge state. The diffusion equation is:

$$\frac{\partial C_T}{\partial t} = \Delta\left[D_V C_A \frac{C_V}{C_V^*} \Delta \log\left(C_A \frac{C_V}{C_V^*} \frac{n}{n_i} \right) + D_I C_A \frac{C_I}{C_I^*} \Delta \log\left(C_A \frac{C_I}{C_I^*} \frac{n}{n_i} \right) \right] \quad \textit{[Eq. 13]}$$

The intrinsic diffusivities for phosphorus are given by:

$$D_I = D_I^X + D_I^- \frac{n}{ni} + D_I^= \frac{n^2}{n_i^2} \quad \textit{[Eq. 14]}$$

$$D_V = D_V^X + D_V^- \frac{n}{ni} + D_V^= \frac{n^2}{n_i^2} \quad \textit{[Eq. 15]}$$

Impurity Segregation Model

In fact, the diffusion from SOD into Si is a non-equilibrium process but we used the equilibrium diffusion model to simulate our diffusion profiles. However, the effect at the SOD/Si interface can be characterised and corrected in the SSUPREM-4 using its impurity segregation model.

The interface segregation is modelled empirically by a first-order kinetic model.

$$F_S = h\left(\frac{C_1}{m} - C_2 \right) \quad \textit{[Eq. 16]}$$

Where:

$\boldsymbol{F_S}$ is the dopant flux across the interface from material 1 to material 2.

$\boldsymbol{C_1}$ and $\boldsymbol{C_2}$ are, respectively, the concentration of impurity in material 1 and material 2.

$\boldsymbol{h}$ is the transport coefficient specified by the parameters ***TRN.0*** and ***TRN.E*** and calculated as:

$$h = TRN.0 \cdot \exp\left(\frac{-TRN.E}{k_B T} \right) \quad \textit{[Eq. 17]}$$

The segregation coefficient m is modelled by an activated process:

$$m = SEG.0 \cdot \exp\left(\frac{-SEG.E}{k_B T} \right) \quad \textit{[Eq. 18]}$$

where ***SEG.0***, ***SEG.E*** as well as ***TRN.0*** and ***TRN.E*** values can be accessed for tuning diffusion profiles.

Default values of parameters are listed in tables A-1 and A-2 in the next page.

Table A-1: Default Impurity Diffusion Coefficients in SSUPREM-4

	Boron	Phosphorus	Boron	Phosphorus	Boron	Phosphorus
	Silicon		Polysilicon		Oxide	
DIX.0 (cm^2/s)	0.037	3.85	3.66	385.0	$3.16x10^{-4}$	$7.6x10^{-3}$
DIX.E (eV)	3.46	3.66	3.46	3.66	3.53	3.5
DIP.0 (cm^2/s)	0.72	0.0	72.0	0.0		
DIP.E (eV)	3.46	0.0	3.46	0.0		
DIM.0 (cm^2/s)	0.0	4.44	0.0	443.3		
DIM.E (eV)	0.0	4.0	0.0	4.05		
DIMM.0 (cm^2/s)	0.0	44.2	0.0	4420.0		
DIMM.E (eV)	0.0	4.37	0.0	4.37		
FI (unitless)	0.94	-	-	-		

Table A-2: Default Impurity Segregation Coefficients in SSUPREM-4

(Si/oxide and Poly/oxide)	Boron	Phosphorus
SEG.0 (unitless)	1126	30
SEG.E (eV)	0.91	0.0
TRN.0 (unitless)	27.9	1.5
TRN.E (eV)	2.48	1.99

APPENDIX 2

UT-BiCMOS process for BJT - Variant 1

(The process is similar to the UT-BiCMOS, except for the base and collector implantation doses)

Start with P-type Silicon wafer, <100>, 10 Ωcm

PHASE 0 Standard wafer cleaning

PHASE 1 Active areas

1.1 Skipped

1.2 Growth 25 nm pad oxide

1.3 Deposition of 50 nm LPCVD Si_3N_4

1.4 Definition of active area (AA-msk)

1.5 Plasma etching of Si_3N_4

1.6 Resist removal and cleaning

1.7 Growth of 700 nm LOCOS field oxide

1.8 Removal of oxidised Si_3N_4

1.9 Removal of Si_3N_4

1.10 Removal of pad oxide

1.11 Wafer cleaning

1.12 Growth of 25 nm sacrificial oxide

PHASE 2 Well formation (skipped)

PHASE 3 Injector formation for EEPROM (skipped)

PHASE 4d Polysilicon gate formation bipolar

4d.4 - 4d.18 are skipped,

modifications and extra process steps are printed bold below;

4d.1 Cleaning of furnace tube

4d.2 Removal of sacrificial oxide

4d.3 Growth 25 nm oxide

4d.19 definition base/collector area (BC-msk)

4d.20 (first) collector implantation

$2.5x10^{13}$ P^+ 1350keV

4d.21 resist removal and cleaning

4d.22 removal of lattice damage
$800^{o}C/N_2/30min.$
4d.23 definition base/collector area (BC-msk)
4d.24 (second) collector and base implantation
$2.5x10^{13}$ P^+ 1350keV
$7x10^{13}$ BF_2^+ 40keV (should be 45 keV)
4d.25 resist removal and cleaning
4d.26 removal of lattice damage
$800^{o}C/N_2/30min.$
4d.27 definition of emitter (EO-msk)
4d.28 opening emitter window in oxide
4d.29 resist removal and cleaning
4d.30 HF-etch for removal native oxide
4d.31 deposition of 300 nm LPCVD undoped polySi
4d.32 implantation of polySi
6.0e15 As^+ 100keV
4d.33 wafer cleaning
4d.34 removal of lattice damage
$800^{o}C/N_2/30min.$
4d.35 definition of poly pattern (EP-msk)
4d.36 plasma etching of the poly-Si
4d.37 resist removal and cleaning
4d.38 oxidation of the poly-Si
$550\text{-}950^{o}C/N_2/40min.$
$950^{o}C/O_2/30min.$
$950^{o}C/N_2/20min.$
$950\text{-}800^{o}C/N_2/20min.$
4d.39 definition of collector plug area (CP-msk)
4d.40 implantation of collector plug
1.0e14 P^+ 500keV
1.0e14 P^+ 200keV
5.0e15 P^+ 70 keV (should be added)
4d.41 resist removal and cleaning

PHASE 5 Shallow n and p formation
5.4 - 5.6 : shallow n steps are skipped

5.1	Definition of SP-area (SP-msk)
5.2	Implantation SP
5.3	Resist removal and cleaning
5.7	Removal of implantation damage
PHASE 6	Contact holes formation
6.1	deposition of 100 nm CVD oxide
6.2	anneal of 100 nm SiO_2
	$800^oC/N_2$/30min.
6.3	deposition of 500 nm BPSG oxide
6.4	wafer cleaning
6.5	flowing of BPSG and gettering
	$800\text{-}900^oC/N_2$/10min.
	$900^oC/N_2$/30min.
	$900\text{-}600^oC/N_2$/150min.
6.6	bare etching of wafer backside
6.7	resist removal and cleaning
6.8	surface states anneal
	$500^oC/N_2/H_2O$/1hr.
6.9	definition of contact windows (CO-msk)
6.10	plasma etching of contact windows
6.11	wet etching of contact windows
6.12	resist removal and cleaning
PHASE 7	Single metal
7.1	Removal of native oxide
7.2	Evaporation of 1μm aluminium
7.3	Definition of contact pads (AL-msk)
7.4	Wet etching of Aluminium
7.5	Removal of photoresist and cleaning
7.6	400 ^oC wet nitrogen annealing, 10 minutes
7.7	deposition of primer for measurement

APPENDIX 3

UT-BiCMOS process for BJT - Variant 2

(Process is the same as variant one but annealing steps used RTP)

PHASE 0 Standard wafer cleaning

PHASE 1 Active areas

The same as in variant 1 (appendix 2)

PHASE 2 Well formation (skipped)

PHASE 3 Injector formation for EEPROM (skipped)

PHASE 4d Polysilicon gate formation bipolar

4d.1 - 4d.3 are unchanged

4d.4 - 4d.18 are skipped,

4d.19 - 4d.41 are unchanged,

4d.21.a is added to have back-side silicon facing to the pyrometer of RTP (for correct RTP temperature measurement)

4d.31.a is added to prevent contamination from implantation

4d.1 Cleaning of furnace tube

4d.2 Removal of sacrificial oxide

4d.3 Growth 25 nm oxide

4d.19 definition base/collector area (BC-msk)

4d.20 (first) collector implantation

$2.5x10^{13}$ P^+ 1000 keV

4d.21 resist removal and cleaning

4d.21.a removal of the back-side pad oxide

4d.22 removal of lattice damage

RTP 1050 ^oC, 20 sec. , N_2

4d.23 definition base/collector area (BC-msk)

4d.24 (second) collector and base implantation

$2.5x10^{13}$ P^+ 1000keV

base implantation

$2x10^{13}$ BF_2^+ 40keV

4d.25 resist removal and cleaning

4d.26	removal of lattice damage
	RTP 1050 ^{o}C, 20 sec. , N_2
4d.27	definition of emitter (EO-msk)
4d.28	opening emitter window in oxide
4d.29	resist removal and cleaning
4d.30	HF-etch for removal native oxide
4d.31	deposition of 300 nm LPCVD undoped polySi
4d.31.a	**growth a thin poly oxide to prevent contamination from implantation**
	800 ^{o}C , 15 min. O_2
	800 ^{o}C , 5 min. N_2
4d.32	implantation of polySi
	6.0e15 As^+ 100keV
4d.33	wafer cleaning
4d.34	skipped (should have RTA)
4d.35	definition of poly pattern (EP-msk)
4d.36	plasma etching of the polySi
4d.37	resist removal and cleaning
4d.38	oxidation of the polySi - skipped (but should not)
4d.39	definition of collector plug area (CP-msk)
4d.40	implantation of collector plug
	1.0e14 P^+ 500keV
	1.0e14 P^+ 200keV
	5.0e15 P^+ 70 keV (should be added)
4d.41	resist removal and cleaning
PHASE 5	Shallow n and p formation
	5.4 - 5.6 the shallow n steps are skipped
5.1	Definition of SP-area (SP-msk)
5.2	Implantation SP
5.3	Resist removal and cleaning
5.7	**skipped**
PHASE 6	Contact holes formation
	the flowing and gettering of BPSG is performed by an RTA step
	6.1 - 6.4 are unchanged, 6.5 is replaced by 6.5.a,
	6.6 is skipped

6.7 - 6.12 are unchanged

6.1 deposition of 100 nm CVD oxide

6.2 anneal of 100 nm SiO_2

$700^oC/N_2/30min.$

6.3 deposition of 500 nm BPSG oxide

6.4 wafer cleaning

6.5.a RTP anneal for flowing, gettering of BPSG, drive-in diffusion of poly emitter and also removal of implantation damages in previous steps

RTP 1050 ^oC, 20 sec. N_2

6.6 skipped

6.7 skipped

6.8 surface states anneal

$500^oC/N_2/H_2O/1hr.$

6.9 definition of contact windows (CO-msk)

6.10 plasma etching of contact windows

6.11 wet etching of contact windows

6.12 resist removal and cleaning

PHASE 7 Single metal

The same as in variant 1 (appendix 2)

APPENDIX 4

UT-BiCMOS process for BJT - Variant 3

(Process flow is changed for using SOD diffusion source for emitter doping)

PHASE 0 Standard wafer cleaning

PHASE 1 Active areas

The same as in variant 1 (appendix 2).

PHASE 2 Well formation (skipped)

PHASE 3 Injector formation for EEPROM (skipped)

PHASE 4d Polysilicon gate formation bipolar

4d.4 - 4d.18 are skipped,

4d.26.a-4d.26.c: collector plug formation (4d.39-4d.41) are moved into front

4d.26.d-4d.26.f: shallow p steps (5.1-5.3) are moved into front

4d.32.a - 4d.32.c: The diffusion from SOD source will replace the implantation step 4.32 of variance 2.

4d.1 Cleaning of furnace tube

4d.2 Removal of sacrificial oxide

4d.3 Growth 25 nm oxide

4d.19 definition base/collector area (BC-msk)

4d.20 (first) collector implantation

$2.5x10^{13}$ P^+ 1000keV

4d.21 resist removal and cleaning

4d.21.a removal of the back-side pad oxide

4d.22 removal of lattice damage

RTP 1050 ^oC, 20 sec. , N_2

4d.23 definition base/collector area (BC-msk)

4d.24 (second) collector and base implantation

$2.5x10^{13}$ P^+ 1000 keV

base implantation

$2x10^{13}$ BF_2^+ 40keV

4d.25 resist removal and cleaning

4d.26 removal of lattice damage

RTP 1050 ^{o}C, 20 sec. , N_2

4d.26.a (4d.39) definition of collector plug area (CP-msk)

4d.26.b (4d.40) implantation of collector plug

1.0e14 P^+ 500keV

1.0e14 P^+ 200keV

5.0e15 P^+ 70 keV (should be added)

4d.26.c (4d.41) resist removal and cleaning

4d.26.d (5.1) Definition of SP-area (SP-msk)

4d.26.e (5.2) Implantation SP

4d.26.f (5.3) Resist removal and cleaning

4d.27 definition of emitter (EO-msk)

4d.28 opening emitter window in oxide

4d.29 resist removal and cleaning

4d.30 HF-etch for removal native oxide

4d.31 deposition of 300 nm LPCVD undoped polySi

4d.32.a deposition of SOD source (P8545)

4d.32.b RTP anneal for emitter diffusion

RTP 1050 ^{o}C, 20 sec., N_2

4d.32.c Removal of SOD layer

4d.33 skipped

4d.34 skipped

4d.35 definition of second poly pattern (EP-msk)

4d.36 plasma etching of second polySi

4d.37 resist removal and cleaning

4d.38 oxidation of the poly-Si (**should be shorter annealing time)**

550-950^{o}C/N_2/40min.

950^{o}C/O_2/30min.

950^{o}C/N_2/20min.

950-800^{o}C/N_2/20min.

PHASE 5 Shallow n and p formation (skipped)

PHASE 6 Contact holes formation

all are the same except the flowing and gettering of BPSG is at low temperature

6.1 deposition of 100 nm CVD oxide

6.2	**anneal of 100 nm SiO_2**
	700^{o}C/N_2/30min.
6.3	deposition of 500 nm BPSG oxide
6.4	wafer cleaning
6.5	**flowing of BPSG oxide**
	700 ^{o}C/N_2/30min.
6.6	bare etching of wafer backside (skipped)
6.7	resist removal and cleaning (skipped)
6.8	surface states anneal
	500^{o}C/N_2/H_2O/1hr.
6.9	definition of contact windows (CO-msk)
6.10	plasma etching of contact windows
6.11	wet etching of contact windows
6.12	resist removal and cleaning

PHASE 7	Single metal
	The same as in variant 1 (appendix 2)

APPENDIX 5

O_2 Plasma Ashing Program For The Barrel Etcher

Barrel etcher, recipe: 11 (total process time ≈ 1 hr)

Gas flow O_2 [sccm]	total pressure [mTorr]	process temperature [°C]	ashing rate [nm/min]	ashing time [min]
400	600	85	--	75

Step	gas flow [gas, valve]	Pressure [mTorr]	power [Watt]	vacuum pump	terminator
1				slow	P < 2 Torr
2				fast	P < 40 mTorr
3	N_2, purge				P > 5 Torr
4				fast	P < 40 mTorr
5	N_2, 1500 sccm	2000		fast	t = 1 min
6	N_2, 1500 sccm		750	fast	T = 85°C
7				fast	P < 40 mTorr
8	O_2, 400 sccm	600			t = 2 min
9	O_2, 400 sccm	600	150		t = 75 min
10				fast	P < 40 mTorr
11	N_2, purge			fast	t = 1 min
12	N_2, purge + 5000 sccm			fast	t = 5 min
13	N_2, purge + 5000 sccm				t = 10 min
14	N_2, purge				P < 2 Torr

APPENDIX 6

Process Flow of a Poly-Spacer PMOSFET Using UT-BiCMOS

Short overview

Phase 1	Active Areas
Phase 2	Well Formation
Phase 4e	Polysilicon Gate Submicron FET
Phase 5b	Shallow Source and Drain Formation
Phase 6	Contact Holes Formation
Phase 7	Contact Metal

Detailed description

Step no.	Action
1.1	Wafer cleaning Standard wafer cleaning
1.2	Growth of 25 nm pad oxide 950 °C, O_2, 30 minutes
1.3	Deposition of 50 nm LPCVD Si_3N_4 SiH_2Cl_2, NH_3, 800ºC
1.4	Definition of active area (AA-msk)
1.5	Plasma etching of Si_3N_4 CF_4/O_2
1.6	Resist removal and cleaning
1.7	Growing of 700 nm LOCOS field oxide 1050 °C, N_2/H_2O, 200 minutes
1.8	Removal of oxidized Si_3N_4 HF/NH_4F
1.9	Removal of Si_3N_4 H_3PO_4 (98%), 180°C, 20 minutes
1.10	Removal of pad oxide HF/NH_4F
1.11	Wafer cleaning Standard cleaning
1.12	Growing of 25 nm sacrificial oxide 950 ºC, 30 minutes

Step no.	Action			
2.1	Nwell implantation			
	Species:	P^{++}	$1.0 \cdot 10^{13}$	1000 kV
		P+	$1.0 \cdot 10^{13}$	300 kV
		As	$2.8 \cdot 10^{13}$	250 kV
2.2	Cleaning			
	Standard cleaning			
2.3	Removal of lattice damage			
	N2, 800°C, 30 minutes			

Step no.	Action
4.1	Cleaning of furnace tube
	$O_2/C_2H_2Cl_2$, 60 minutes
4.2	Removal of sacraficial oxide
	1% HF
4.3	Growing of 8nm gate oxide
	850ºC, 45 minutes
4.4	Deposition of 300nm LPCVD undoped α-silicon
	SiH_4, 550ºC, 2h30’
4.5	Deposition of 50nm LPCVD SiO_xN_y
	SiH_2Cl_2, NH_3, N_2O, 800ºC, 10 minutes
4.6	Definition of polysilicon pattern (PS msk)
4.7	Ashing of the resist layer
	O_2, 85ºC, 75 minutes
4.8	Plasma etching of SiO_xN_y and undoped polysilicon
	$Cl/ClSi_4$, End Point Detection
4.9	Resist removal and cleaning
	Barrel etcher recipe 2
	Standard cleaning
4.10	Oxidation of Polysilicon
	O_2, 900ºC, 45 minutes
4.11	Deposition of 50nm SiO_xN_y spacer
	SiH_2Cl_2, 800ºC, 10 minutes
4.12	Etch spacer oxide
	CF_4/O_2, End Point Detection
4.13	Cleaning of wafers
	Standard cleaning

Step no.	Action

5e.1 Deposition of 300nm LPCVD undoped α-silicon

SiH4, 550ºC, 2h30'

5e.2 Plasma etching of α-silicon

Cl_2, 40 minutes, End Point Detection

5e.3 Definition of source drain disconnect area (SD-msk)

5e.4 Plasma etching of undoped α-silicon

$Cl/ClSi_4$, End Point Detection

5e.5 Resist removal and cleaning

5e.6 Etching of 100nm SiO_xN_y

CF_4/O_2

5e.7 Resist deposition

5e.8 Etching backside wafer

CF_4/O_2

5e.9 Removal resist

5e.10 Deposition of 200nm Spin on Dopant

Or BF_2 implantation

BF_2 $5x10^{15}$,50 keV

5e.11 Rapid Thermal Annealing or Furnace Annealing

RTP: 1050ºC, 60 seconds FA: 900ºC, 17 minutes

5e.12 SOD removal

HF/NH_4F, 1 minutes

Boiling HNO_3 10 minutes

HF 1%, 2 minutes

Step no. Action

6e.1 Deposition of 500 nm APCVD SiO_2

6e.2 Definition of contact windows (CO-msk)

6e.3 Wet etching of contact windows

HF/NH_4F, 5 minutes

Step no. Action

7a.1 Removal of native oxide

1% HF, 1minute

7a.2 Depostion of 70 nm TiW and 1µm Al

7a.3 Definition of interconnect (IN-msk)

7a.4 Wet etching of aluminum

H_3PO_4 55ºC 1 minute

Wet etching of TiW

H_2O_2 (31%) 30 minutes

7a.5 Removal of resist

100% HNO_3, 10 minutes

7a.6 Contact anneal

wet N_2, 400ºC, 10 minutes

Summary

This thesis deals with the study of shallow PN junction formation by dopant diffusion from Spin-On Glass (SOG) for future deep sub-micron BiCMOS technology. With the advantages of no transient enhanced diffusion and no metal contamination, diffusion from highly doped SOG (also called spin-on dopant - SOD) is a good technology for shallow junction formation. In this thesis, diffusion of impurities from SOD into Si and polysilicon on silicon structure has been studied. This shallow junction formation technique using SOD has been applied in realisation of two important devices, i.e. high frequency bipolar transistor and deep sub-micron elevated source/drain MOSFET.

In chapter 2, realization and characterization of the SOG materials are described. Good undoped SOG and SOD (phosphorus doped) materials and layers with similar properties as commercial ones have been obtained. Using the design of experiment method, effects of several factors in the sol-gel reaction, e.g. water volume, TEOS volume, acid concentration and reaction time, were investigated. The results showed that the water volume had the largest effect on the thickness, the shrinkage and the surface roughness of obtained SOG layers. In addition, dilution of the obtained SOG sols at the end of the sol-gel reactions with a solvent, e.g. acetone or ethanol, can increase the lifetime of the sols and improve significantly the properties of the coating layers. Elements in SOD layers and their concentrations profiles have been explored by advanced analyse techniques, e.g. XPS and AES. The evolution of the SOD layer properties during baking and annealing steps was investigated.

In chapter 3, diffusion of boron and phosphorus from SOD into Si and polysilicon on Si structure was investigated using both conventional furnace and rapid thermal processing. The influence of the thermal budget to the electrical properties of the shallow junctions were investigated using diode structures. Rapid thermal diffusion of SOD resulted in very shallow junctions in Si, i.e. less than 20 nm. Diffusion using conventional furnace resulted in deeper junctions due to the large thermal budget.

The shallow junction formation using polysilicon/Si structures has the advantage of lower sheet resistance and elevated contacts in comparison with the conventional Si junction. Very high quality shallow junctions were obtained with this structure using both furnace and rapid thermal diffusion. Furnace diffusion resulted in better quality diodes than rapid thermal diffusion because the rapid thermal process created stress-induced defects in the PN junctions. In this diffusion technique, the diffusion of impurities from SOD into polysilicon/Si is very dependent on the micro-structure of the as-deposited layer, i.e. polysilicon or amorphous silicon. The random structure of as-deposited amorphous layers caused slower diffusion for both boron and phosphorus compared to the as-deposited polysilicon layer with columnar structure. Therefore, a higher thermal budget was required to form a good shallow junction using as-deposited amorphous silicon.

In both cases of dopant diffusion from SOD into Si and polysilicon/Si, a 50 nm junction depth in the Si substrate was the lower limit for a good quality shallow PN junction.

The SOD diffusion into polysilicon/Si technique was then applied to shallow junction formation for polysilicon-emitter in high frequency bipolar transistors and for elevated source/drain MOSFETs in chapter 4 and 5 respectively.

In chapter 4, the UT-BiCMOS bipolar transistor process was optimised for very high energy collector implantation and high frequency response. Three different process variants were generated and realized. The first variant was proposed to keep the standard UT-BiCMOS process from changes and used P^{3+} ions at 450 kV for collector implantation. However, it was still very difficult to obtain P^{3+} ions and the implanter breakdown problem was not solved completely. The second variant was designed to use advanced rapid thermal annealing technique for higher frequency response, i.e. more than 10 GHz. In this variant, the collectors were implanted with P^{2+} at 500 kV and there was no breakdown problem. Better devices were obtained for this variant with higher yield compared to variant 1. Variant 3 with a SOD source for doping the poly-emitter showed the best results. The process flow of this variant was almost the same as in variant 2 except for the emitter doping step. However, all of the devices had better electrical characteristics than variant 1 and 2 showing the advantage of the diffusion process from SOD sources.

In chapter 5, a new concept of poly-spacers was proposed for future sub-0.1 μm MOSFET. This device has the elevated source/drain structure and low series resistance which is very important to obtain a high drive current. Furthermore, the technology is straightforward and reproducible. As a vehicle to test the concept, poly-spacer 0.25 μm PMOS transistors were designed and realized using SOD diffusion source or BF_2 implantation. The photoresist ashing technique was investigated using a O_2 plasma for submicron gate length formation.

Several variants using both conventional furnace and rapid thermal processing were carried out. PMOS transistors with 8 nm gate oxide and gate lengths of 0.1 μm to 1 μm have been realised. All the PMOS transistors fabricated by SOD showed good electrical characteristics. The best working PMOS transistor had an effective gate length of 0.19 μm, a threshold voltages of approximately -0.67 V and -0.70 V for furnace diffusion and rapid thermal diffusion of SOD respectively, a subthreshold swing of 95 mV/decade and a current drive of about 160 μA/μm at a supply voltage of –2.5 V.

Although the device was not optimised yet, its electrical characteristics were very good and comparable with state-of-the-art devices. Above all, the new poly-spacer concept has been proven to be effective for future MOSFETs.

Tóm Tắt Nội Dung

Mục đích của bản luận án này nhằm nghiên cứu việc chế tạo lớp chuyển tiếp bán dẫn PN mỏng bằng phương pháp khuyếch tán từ vật liệu Spin-On Glass (SOG) dùng cho công nghệ BiCMOS trong tương lai với kích thước linh kiện siêu nhỏ. Với những ưu điểm như không có hiệu ứng khyếch tán nhanh gây bởi các khuyết tật do cấy ion và không có sự nhiễm bẩn kim loại, phương pháp khuyếch tán từ nguồn SOG có chứa nồng độ tạp cao (vật liệu này cũng được gọi là spin-on dopant - SOD) là một công nghệ rất tốt để chế tạo lớp chuyển tiếp mỏng. Trong bản luận án này, chúng tôi nghiên cứu sự khuyếch tán của các tạp chất từ lớp SOD vào đế silic và vào cấu trúc polysilicon trên đế silic. Chúng tôi cũng ứng dụng kỹ thuật chế tạo lớp chuyển tiếp mỏng sử dụng SOD này vào các công nghệ chế tạo của hai linh kiện bán dẫn quan trọng, đó là bipolar transistor tần số cao và MOSFET có các cực nguồn/ máng nâng cao với kích thước kênh nhỏ hơn rất nhiều so với 1 μm.

Trong chương 2 của bản luận án, chúng tôi miêu tả việc chế tạo và phân tích tính chất các vật liệu SOG. Chúng tôi đã chế tạo thành công các vật liệu cũng như các màng mỏng của SOG không chứa tạp và SOD (chứa tạp phốt pho) có tính chất tương đương với các thương phẩm. Sử dụng phương pháp thiết kế thí nghiệm, chúng tôi đã nghiên cứu sự ảnh hưởng của một số tham số quan trọng trong quá trình phản ứng sol-gel như thể tích nước, thể tích TEOS, nồng độ axit và thời gian phản ứng. Kết quả thí nghiệm cho thấy thể tích nước có ảnh hưởng lớn nhất đến chiều dầy, độ co và độ nháp bề mặt của lớp màng SOG nhận được. Hơn nữa, việc hoà tan các chất SOG nhận được sau quá trình phản ứng với các chất dung môi, ví dụ như acetone hoặc cồn, thì có thể làm tăng thời gian sống của chất SOG và nâng cao đáng kể các tính chất của lớp màng nhận được sau khi quay phủ. Chúng tôi đã dùng các phương pháp phân tích hiện đại như XPS và AES để đo các nguyên tố cũng như các profile nồng độ của chúng trong các lớp SOD. Nhờ đó mà sự chuyển đổi tính chất của các lớp SOD trong quá trình nung và quá trình khuyếch tán đã được khám phá.

Trong chương 3, chúng tôi nghiên cứu sự khuyếch tán của bo và phốt pho từ SOD vào silic và các cấu trúc polysilicon trên silic dùng lò khyếch tán thông thường và lò nhiệt nhanh. Chúng tôi đã nghiên cứu sự ảnh hưởng của lượng nhiệt nung đối với các tính chất của các lớp chuyển tiếp PN mỏng trên cơ sở dùng linh kiện diode. Dùng phương pháp khuyếch tán từ SOD trong lò nhiệt nhanh chúng tôi đã chế tạo được các lớp chuyển tiếp cực nông trong đế silic, mỏng hơn 20 nm. Phương pháp khuyếch tán bằng lò thông thường cho kết quả là các lớp chuyển tiếp quá sâu do lượng nhiệt nung lớn.

So sánh với lớp chuyển tiếp thông thường trong đế silic thì lớp chuyển tiếp nông chế tạo trên các cấu trúc polysilicon/Si có ưu điểm là điện trở màng thấp hơn và có tiếp xúc cách xa đế Si. Dùng cấu trúc này, chúng tôi đã nhận được các lớp chuyển tiếp cực mỏng với chất lượng cao chế tạo bằng cả phương pháp khuyếch tán thông thường lẫn khuyếch tán dùng lò nhiệt nhanh. Kết quả cho thấy là lò khuếch tán thông thường tạo được các lớp chuyển tiếp có chất lượng cao hơn so với lò nhiệt nhanh bởi vì lò nhiệt nhanh thường tạo ra các khuyết tật do ứng suất nội trong các lớp chuyển tiếp PN. Trong kỹ thuật khuyếch tán này, sự khuyếch tán của các tạp chất vào cấu trúc polysilicon/Si rất phụ thuộc vào cấu trúc vi mô của lớp phủ trên bề mặt silic là polysilicon hay là amorphous silic. Cấu trúc hỗn độn của lớp amorphous silic đã làm chậm lại sự khuyếch tán của cả bo lẫn phốt pho so với cấu trúc cột thẳng đứng trong lớp polysilicon. Do đó

cần phải có một lượng nhiệt lớn hơn để có thể tạo ra được một lớp chuyển tiếp mỏng có chất lượng tốt nếu muốn sử dụng lớp amorphous silic.

Trong cả hai trường hợp khuyếch tán từ SOD vào Si và vào cấu trúc polysilicon/Si thì 50 nm là giới hạn dưới của lớp chuyển tiếp PN vừa mỏng và vừa có chất lượng cao.

Tiếp theo chúng tôi đã ứng dụng kỹ thuật khuyếch tán từ SOD vào polysilicon/Si cho việc chế tạo lớp chuyển tiếp nông dùng cho polysilicon-emitter trong các bipolar transistor và cho MOSFET có cấu trúc nguồn/máng nâng cao lần lượt trong các chương 4 và 5.

Trong chương 4, chúng tôi đã tối ưu hoá công nghệ UT-BiCMOS để chế tạo bipolar transistor với cực collector tạo bằng phương pháp cấy ion năng lượng cao và để nâng cao tần số làm việc của transistor. Chúng tôi đã đưa ra và thực thi 3 giải pháp công nghệ. Giải pháp thứ nhất được đề nghị để giữ sao cho sự thay đổi so với công nghệ UT-BiCMOS chuẩn ở mức độ tối thiểu, đồng thời cực collector được chế tạo bằng phép cấy ion dùng ion P^{3+} với mức năng lượng cấy là 450 kV. Tuy nhiên do các ion P^{3+} rất khó nhận được cho nên hiện tượng sự cố của máy cấy ion vẫn bị xảy ra. Giải pháp thứ hai được thiết kế có sử dụng kỹ thuật khuyếch tán nhiệt nhanh để tạo ra các transistor có tần số làm việc cao, lớn hơn 10 GHz. Trong giải pháp này, do collector được chế tạo bằng phép cấy ion với ion P^{2+} ở mức năng lượng cấy là 500 kV cho nên không có hiện tượng sự cố máy cấy ion nữa. Chúng tôi đã nhận được các transistor có tính chất tốt hơn và với hiệu suất công nghệ lớn hơn từ giải pháp này so với giải pháp thứ nhất. Giải pháp thứ ba có sử dụng nguồn khuyếch tán SOD để pha tạp cực polysilicon-eimitter đã cho kết quả tốt nhất. Quá trình công nghệ của giải pháp này gần giống như giải pháp thứ hai trừ quá trình pha tạp cực emitter. Tuy nhiên, tất cả các transistor nhận được có các đặc trưng điện tốt hơn so với giải pháp 1 và 2. Kết quả này cho thấy tính ưu việt của phương pháp khuyếch tán dùng SOD.

Trong chương 5, chúng tôi đã phát minh ra một cấu trúc transistor mới gọi là poly-spacer transistor dùng cho MOSFET thế hệ có kênh ngắn hơn 0,1 μm trong tương lai. Linh kiện này có cấu trúc nguồn/máng nâng cao và có điện trở tiếp xúc nhỏ rất cần thiết để nhận được dòng làm việc lớn. Hơn nữa, công nghệ này rất là đơn giản và có tính lặp lại cao. Để kiểm chứng linh kiện mới này, chúng tôi đã thiết kế và chế tạo các poly-spacer transistor PMOS có kênh ngắn 0.25 μm tạo bằng nguồn SOD hoặc cấy ion BF_2. Chúng tôi đã sử dụng kỹ thuật ăn mòn dần lớp photoresist bằng plasma oxy để chế tạo các kênh ngắn hơn 1 μm.

Chúng tôi đã thực thi một số giải pháp sử dụng cả lò khuyếch tán thông thường và lò khuyếch tán nhanh. Transistor PMOS với lớp ô xit của cực cổng mỏng cỡ 8 nm và chiều dài kênh từ 0.1 μm cho đến 1 μm đã được chế tạo. Tất cả các transistor chế tạo bằng nguồn SOD cho đặc trưng điện tốt. Transistor tốt nhất có chiều dài hiệu dụng của kênh là 0.19 μm, điện áp ngưỡng xấp xỉ -0,67 V và -0,70 V lần lượt cho các transistor tạo bằng lò khuyếch tán thông thường và lò khuyếch tán nhiệt nhanh, độ dịch dưới ngưỡng là 95 mV/bậc và dòng làm việc là 160 μA/ μm ở điện áp nguồn là -2.5 V.

Mặc dù linh kiện này vẫn chưa được tối ưu hoá nhưng các đặc trưng điện của nó đã rất tốt và tương đương so với các linh kiện tiên tiến nhất. Vượt lên trên tất cả đó là thiết kế mới có sử dụng poly-spacer đã được chứng minh là rất có hiệu quả cho các MOSFET trong tương lai.

List of Publications

List of papers:

1. N. N. Toan, N. D. Chien, *"Two dimensional numerical simulation of MOS devices"*, 1st National Telecom and Communication Conference Proceeding, Hanoi, Vietnam, 1992
2. N. N. Toan, N. D. Chien, V. T. Son, N. T. Giang, J. Holleman, P.H. Woerlee, *"Spin-On Glass (SOG) materials and applications"*, 2nd International Workshop on Materials Science Proceeding, IWOMS'95, Hanoi, Vietnam, 1995, p. 191-194
3. N.D. Chien, N.N. Trung, N.N. Toan, L.M. Phuong, T.T. Thanh, *"Application of Spin-On Glass (SOG) materials and technology for the fabrication of semiconductor devices"*, Hanoi University of Technology Scientific Conference Proceeding, Hanoi, Vietnam, 1996, p. 86-91
4. T.K. Anh, L.Q. Minh, L.C. Anh, N.N. Toan, W. Strek, C. Barthou, *"The role of sol-gel technology in the study of medical imaging materials"*, Exploring New Tracks in Imaging ICPS'98 Proceeding, Antwerp, Belgium, 1998, p. 224-227
5. N.N. Toan, N.D. Chien, J. Holleman, P.H. Woerlee, *"Formation of shallow junctions by dopant diffusion from Spin-On Glass"*, 1st annual workshop on Semiconductor Advanced for Future Electronics (SAFE98) proceeding, Mierlo, the Netherlands, 1998, 6 pages (CD-ROM)
6. N.N. Toan, T.D. Quyen, N.D. Chien, T.K. Anh, C. Barthou, *"Study of spin-on dopant materials for micro-electronic technology prepared by sol-gel method"*, 7th National Telecom and Communication Conference Proceeding, Hanoi, Vietnam, 1998, p. 343-350

List of relevant presentations:

- N. N. Toan, N. D. Chien, J. Holleman, P.H. Woerlee *"Spin-On Glass (SOG) materials for advanced IC technology"*, annual MESA research institute meeting, Enschede, the Netherlands 1994 - Poster presentation.
- N. N. Toan, N. D. Chien, J. Holleman, P.H. Woerlee, *Shallow junction formation using Spin-On Glass"*, FOM days, Veldhoven, the Netherlands 1994 – Poster presentation.
- N. N. Toan, N. D. Chien, V. T. Son, N. T. Giang, J. Holleman, P.H. Woerlee, *"Spin-On Glass (SOG) materials and applications"*, 2nd International Workshop on Materials Science, IWOMS'95, Hanoi, Vietnam 1995 - Oral presentation.
- N. N. Toan, N.D. Chien, J. Holleman, P.H. Woerlee, *"Formation of shallow junctions by dopant diffusion from Spin-On Glass"*, 1st annual workshop on Semiconductor Advanced for Future Electronics (SAFE), Mierlo, the Netherlands 1998 - Poster presentation (Best Poster Award)

Acknowledgments

In the early nineties, a project agreement called ITIMS (International Training Institute for Materials Science) was signed between NUFFIC (the Netherlands University Foundation For International Co-operation) and the Vietnamese Ministry of Education and Training. As a result, some Vietnamese students were selected to be “sandwich” Ph.D. students. They were allowed to work on the project for one year in Holland, one year in Vietnam, one year in Holland, and one year in Vietnam. That is a “sandwich” of 4 years working. I was selected to work on the subject on shallow junction formation using spin-on glass in co-operation between ITIMS and the MESA institute, University of Twente. The combination between a student coming from a very poor country, where there was no clean-room laboratory, with a subject, which big companies like IBM and Intel are working on, was really a difficult challenge for me and my supervisors. With much effort, energy, and help from many sources, this thesis has been resulted.

First, I would like to thank my promoter, Prof. Dr. Pierre H. Woerlee, for his support and advises during my Ph.D. study. Without his valuable guidance this work could not have been done. I would like to acknowledge his fruitful discussions. I also very appreciate his patience in reading and criticizing the manuscript of this thesis.

I would like to thank my co-promoter, Prof. Dr. Nguyễn Đức Chiến, for his support and encouragement since the very beginning of my studies. He has continuously encouraged and supported me to study since I was an undergraduate student at the Hanoi University of Technology and also when I became a Ph.D. student of ITIMS. He has given me many useful advises during my research.

I would like to thank my assistant promoter, Dr. Jisk Holleman, for his support since I came to the MESA institute to work on this project. His kind support has helped me to overcome many difficulties to integrate into a new working environment. He has always encouraged and believed in my creativity and independence in research.

My sincere thanks are devoted to Prof. Dr. Thân Đức Hiền and Prof. Dr. Nguyễn Phú Thùy for all their help during my study.

I wish to thank Prof. Dr. Hans Wallinga for his kind support of my study in the group of Semiconductor Components, Electrical Engineering Department, University of Twente.

I would like to thank Dr. Võ Thạch Sơn, Dr. Trần Kim Lan and all of my teachers at the Faculty of Applied Physics, Hanoi University of Technology. I have learned from them the first lesson in micro-electronics.

I am very grateful to Dr. Trần Kim Anh, Dr. Lê Quốc Minh and their colleagues for the nice co-operation in sol-gel research.

I would like to thank Dr. F.F. Bekker, Ir. P. de Goeje and all the members of the International Adviser Board who have created the ITIMS project. Thanks to them, a nice ITIMS building with modern laboratories was built and many Vietnamese students like me have chance to learn very advanced technologies and physics.

I would like to give my sincere thanks to Tom Aarnink, Ronald Dekker, Eddy Ruiter, Arie Kooy, Gerrit Boom, Stan Krüger, Herman D. van Corbach and all other clean-room technicians for their great assistance. I thank Tom Aarnink for his kind support during my research and for his friendship. I also like to thank Marcel H.H. Weusthof and Ronald Dekker for the assistance with the RTP system.

Henk de Vries and Marcel H.H. Weusthof are also kindly acknowledged for their supporting in the MESA Test Centre.

I also would like to thank Mr. Tuấn, Mr. Hồng and Mr. Hà for their technical support during my research in Vietnam.

I would like to thank Bert Otter for the nice SEM and FIB pictures of my devices. I also would like to thank A. Lisfi for the XPS and AES measurements. Dr. Istvan Barsony and his colleagues in Hungary are kindly acknowledged for the SRP measurements. Thanks are also conveyed to Ms. Martine Gri and her co-workers at the CIME Institute (France) for SRP measurements of my samples. Philips Natlab research laboratory (the Netherlands) is acknowledged for the SIMS measurements.

I like to thank all the people who have contributed their work to my research during their M. Sc. final assignments. They are Nguyễn Thanh Giang, Trần Trí Thanh, Lê Minh Phương, Trần Đức Quyền, K. Oldenziel, P.C.S. Scholtens.

Herma van Kranenburg, Margie Remrev, Marie-Christine Predery, Sophie Kreulen, Cora Salm, Cor Bakker, Jan Hovius, Egbert Holl, Ton Mouthaan, Vincent Houtsma and all other colleagues and students in the group are kindly acknowledged for their support and for providing an enjoyable working environment in the HC group at the University of Twente.

I would like to give special thanks to Nguyễn Hoàng Việt, who has been a nice room mate and friend. He has shared every dinner with me in the MENSA. His sense of humor has given me a pleasant time during my stay in Holland.

P.T. Thuần, T.K. Tú, N.P. Loan, N.T. Hà, P.T. Hằng, Mr. Khiêm, V.N. Hùng, T.Q.Vinh, P.K. Hùng, N.V. Hoàng, C.V. Thắng, N.N. Trung, V.A. Minh, N.Đ. Chính and other ITIMS staff are kindly acknowledged for their help and for their friendships during my study in ITIMS.

Many thanks are due to ITIMS Ph.D. students: D. X. Thảo, Đ.V. Anh, L.A. Tuấn, D. T. Thành, T.M. Danh, L.D. Tùng, N.K. Mẫn, P.L. Kim, P. T. Huy, N. P. Dương, N.V. Hiếu, T.D. Hiển, P. D. Thắng and others.

I also like to thank all my friends in Vietnam as well as in Holland, who has shared their time to have fun with me. Life would be very boring if I don't have you. You are always in my mind forever.

The last words of thank are devoted to my beloved parents, my brothers and sisters. They have supported and encouraged me every moment since I was born. Nothing could show my great appreciation to them. I will try my best to deserve your love!

Enschede, Dec. 8th 1998

Nguyễn Như Toàn.

Lời Cảm Ơn

Vào đầu những năm 1990, dự án ITIMS (Trung tâm Quốc tế Đào tạo về Khoa học Vật liệu) đã được ký kết giữa Bộ Giáo dục & Đào tạo Việt nam với NUFFIC (Tổ chức hợp tác Quốc tế của các trường Đại học Hà lan). Nhờ có sự ký kết này mà một số sinh viên Việt nam được chọn làm nghiên cứu sinh dạng "sandwich". Những sinh viên này được phép nghiên cứu cho dự án trong thời gian 1 năm trong Việt nam, 1 năm ở Hà lan, 1 năm trong Việt nam, và 1 năm ở Hà lan. Đó là nghĩa tượng trưng của thời gian 4 năm làm việc xen kẽ giữa 2 nước. Tôi cũng được chọn làm nghiên cứu sinh "sandwich" cho đề tài "chế tạo lớp chuyển tiếp bán dẫn PN mỏng dùng vật liệu Spin-On Glass" trong dự án hợp tác giữa viện nghiên cứu ITIMS và viện nghiên cứu MESA, trường Đại học Twente. Sự kết hợp giữa một sinh viên từ một nước nghèo nàn lạc hậu, nơi chưa có một phòng thí nghiệm nào cấp quốc tế về nghiên cứu chế tạo linh kiện bán dẫn, với một đề tài mà các công ty hàng đầu thế giới như IBM và Intel cũng đang đầu tư nghiên cứu quả là một thử thách vô cùng khó khăn đối với tôi cũng như các giáo sư hướng dẫn. Với rất nhiều cố gắng, nỗ lực và sự giúp đỡ nhiệt tâm từ rất nhiều nguồn khác nhau bản luận án này đã được ra đời.

Trước tiên, tôi xin cám ơn giáo sư hướng dẫn của tôi, GS. TS. Pierre H. Woerlee, vì sự giúp đỡ cũng như sự hướng dẫn trong thời gian tôi làm luận án tiến sĩ. Không có sự hướng dẫn quí báu của giáo sư Woerlee có lẽ bản luận án này khó mà kết thúc thành công. Tôi cũng lấy làm biết ơn sự kiên trì của giáo sư trong việc đọc và góp ý sửa bản thảo của bản luận án này.

Tôi cũng xin chân thành cám ơn giáo sư đồng hướng dẫn, GS. TS. Nguyễn Đức Chiến, vì sự giúp đỡ và sự khuyến khích kể từ ngày đầu tiên tôi bắt tay vào nghiên cứu. Giáo sư Chiến đã không ngừng khuyến khích và giúp đỡ tôi từ lúc tôi làm luận án tốt nghiệp ở Trường Đại học Bách khoa Hà nội cũng như cho đến khi tôi đã là nghiên cứu sinh của ITIMS. Giáo sư Chiến đã cho tôi rất nhiều lời khuyên bổ ích trong quá trình nghiên cứu.

Tôi xin chân thành cám ơn thầy giáo hướng dẫn, TS. Jisk Holleman, vì sự giúp đỡ kể từ khi tôi làm việc tại viện nghiên cứu MESA. Sự giúp đỡ đầy tình người của tiến sĩ Holleman đã giúp đỡ tôi rất nhiều trong việc vượt qua những khó khăn ban đầu để hoà nhập vào môi trường làm việc mới. Tiến sĩ Holleman đã luôn luôn tin tưởng và khuyến khích trí sáng tạo và sự độc lập nghiên cứu của tôi.

Tôi cũng xin chân thành cảm ơn mọi sự giúp đỡ của GS. TS. Thân Đức Hiền và GS. TS. Nguyễn Phú Thùy.

Tôi xin cám ơn GS. TS. Hans Wallinga vì sự giúp đỡ trong quá trình tôi nghiên cứu tại nhóm Linh kiện Bán dẫn, khoa Điện Điện tử, trường Đại học Twente.

Tôi xin bày tỏ lòng cảm ơn đối với TS. Võ Thạch Sơn, TS. Trần Kim Lan và các thầy cô giáo tại khoa Vật lý Kỹ thuật, trường Đại học Bách khoa Hà nội. Tôi đã được học từ các thầy các cô bài học đầu tiên về Vi Điện tử.

Tôi cũng xin cám ơn TS. Trần Kim Anh, TS. Lê Quốc Minh và các đồng nghiệp về sự hợp tác tốt đẹp trong lĩnh vực nghiên cứu vật liệu sol-gel.

Tôi xin cám ơn TS. F.F. Bekker, P. de Goeje và các thành viên trong Ban Cố vấn Quốc tế (IAB), là những người sáng lập trung tâm ITIMS. Nhờ có họ mà toà nhà ITIMS với những phòng thí nghiệm hiện đại đã được xây lên và nhiều sinh viên giống như tôi có được cơ hội học tập nền khoa học công nghệ và vật lý tiên tiến của thế giới.

Tôi xin chân thành cám ơn Tom Aarnink, Ronald Dekker, Eddy Ruiter, Arie Kooy, Gerrit Boom, Stan Kruger, Herman D. van Corbach và các kỹ thuật viên phòng sạch khác vì sự giúp đỡ to lớn của họ. Tôi xin cám ơn Tom Aarnink vì sự giúp đỡ nhiệt tình cũng như tình bạn của anh đối với tôi. Tôi cũng xin cám ơn Marcel H.H. Weusthof và Ronald Dekker vì sự giúp đỡ đối với hệ RTP.

Tôi xin cám ơn Henk de Vries và Marcel H.H. Weusthof vì sự giúp đỡ của họ tại trung tâm đo lường của viện MESA.

Tôi cũng xin cảm ơn anh Tuấn, anh Hồng và anh Hà vì những giúp đỡ kỹ thuật trong quá trình tôi nghiên cứu tại Việt nam.

Tôi xin cảm ơn Bert Otter vì những tấm ảnh SEM và FIB rất đẹp chụp các linh kiện của tôi. Tôi cũng xin cám ơn A. Lisfi về các phép đo XPS và AES. Tôi xin gửi lời cảm ơn tới TS. Istvan Barsony và các đồng nghiệp ở Hungary vì các phép đo SRP. Tôi cũng xin cám ơn cô Martine Gri và các đồng nghiệp tại viện nghiên cứu CIME (Pháp) vì các phép đo SRP. Tôi cũng xin cám ơn phòng thí nghiệm Philips Natlap (Hà lan) về các phép đo SIMS.

Tôi xin cám ơn tất cả những sinh viên đã đóng góp công sức của họ vào bản luận án này khi họ làm luận án tốt nghiệp thạc sĩ. Họ là: Nguyễn Thanh Giang, Trần Trí Thanh, Lê Minh Phương, Trần Đức Quyền, K. Oldenziel, P.C.S. Scholtens.

Tôi cũng xin cám ơn Herma van Kranenburg, Margie Remrev, Marie-Christine Predery, Sophie Kreulen, Cora Salm, Cor Bakker, Jan Hovius, Egbert Holl, Ton Mouthaan, Vincent Houtsma và các đồng nghiệp cùng các sinh viên vì sự giúp đỡ của họ ở nhóm nghiên cứu Linh kiện Bán dẫn tại trường Đại học Twente.

Tôi xin dành sự cảm ơn đặc biệt gửi tới Nguyễn Hoàng Việt, người mà vừa là đồng nghiệp nhiệt tình và cũng là người bạn tốt. Việt đã chia sẻ những bữa cơm chiều với tôi hàng ngày ở MENSA. Sự vui tính của Việt đã góp phần làm cho cuộc sống của tôi ở Hà lan rất là dễ chịu.

Tôi xin chân thành cảm ơn: P.T. Thuần, T.K. Tú, N.P. Loan, N.T. Hà, P.T. Hằng, Chú Khiêm, V.N. Hùng, T.Q.Vinh, P.K. Hùng, N.V. Hoàng, C.V. Thắng, N.N. Trung, V.A. Minh, N.Đ. Chính và các thành viên khác của ITIMS vì mọi sự giúp đỡ cũng như tình bạn của họ dành cho tôi.

Tôi cũng xin gửi lời cám ơn đến các nghiên cứu sinh của ITIMS: D. X. Thảo, Đ.V. Anh, L.A. Tuấn, D. T. Thành, T.M. Danh, L.Đ. Tùng, N.K. Mẫn, P.L. Kim, P. T. Huy, N. P. Dương, N.V. Hiếu, T.D. Hiển, P. D. Thắng và những sinh viên khác.

Tôi xin chân thành cảm ơn tất cả các người bạn của tôi ở Việt nam cũng như ở Hà lan, những người đã chia sẻ thời gian của họ cùng vui với tôi. Cuộc đời này sẽ tẻ nhạt biết bao nếu như không có các bạn. Các bạn luôn hiện hữu mãi mãi trong trái tim tôi.

Những lời cảm ơn cuối cùng tôi xin dành cho cha mẹ và các anh chị yêu quí của tôi. Họ là những người luôn luôn giúp đỡ và khuyến khích tôi mọi nơi mọi lúc kể từ ngày tôi chào đời. Không có gì có thể diễn đạt được lòng yêu mến và biết ơn của tôi đối với họ. Tôi sẽ cố gắng hết sức để xứng đáng đối với tình thương yêu của mọi người.

Enschede, ngày 8 tháng 12 năm 1998

Nguyễn Như Toàn.

Biography

Nguyen Nhu Toan was born on 14 February 1970 in Hanoi, Vietnam. He received a B.Sc. in Applied Physics from Hanoi University of Technology in 1992 with honour "University Excellent Student". He worked on "Two dimensional numerical simulation of MOS devices" for his final assignment. In the same year, his results on simulation was published in the proceeding of Vietnam National Telecom and Communication Conference.

After graduation, he continued his work on simulation at the faculty of Applied Physics at Hanoi University of Technology for one year.

In 1994, he became a "sandwich" Ph.D. student of the International Training Institute for Materials Science (ITIMS) in Hanoi, Vietnam. The Ph.D. research was on "Spin-on glass materials and applications in advanced IC technologies". As a "sandwich" student he worked for two years at ITIMS (Vietnam) and two years at the Micro Electronic & Sensors and Actuators (MESA) research institute, University of Twente (UT), Enschede, the Netherlands. He was working under the supervision of Prof. Dr. P. H. Woerlee (UT and Philips), Prof. Dr. Nguyen Duc Chien (ITIMS) and Dr. J. Holleman (UT). The thesis resulted in papers in international conferences and a Best-Poster-Award.

The cover illustration shows the technology of spin-on glass in which there are two drops of spin-on glass falling on a rotating wafer. The spin-on glass material is used to dope the Si substrate for making electronic devices, which consist of positive and negative conducting parts. This rotating Si wafer is described by an ellipse, which looks similar to the symbol “Yin and Yang” of Taoism, a Chinese philosophy. This symbol also means that in the universe there are two principles, one negative, dark, and feminine (Yin), and one positive, bright, and masculine (Yang), whose interaction influences the destinies of creatures and things.