DEUTERIUM IN THE GATE DIELECTRIC OF CMOS DEVICES

Albert Jan Hof

Samenstelling promotiecommissie

Voorzitter:	
prof. dr. W.H.M. Zijm	
Secretaris:	
prof. dr. W.H.M. Zijm	Universiteit Twente
Promotor:	
prof. dr. J. Schmitz	Universiteit Twente
Assistent promotor:	
dr. A.Y. Kovalgin	Universiteit Twente
Deskundige:	
dr. P.H. Woerlee	Philips Research Laboratories
Leden:	
prof. dr. ir. F.G. Kuper	Universiteit Twente Philips Semiconductors
prof. dr. H. Wallinga	Universiteit Twente
prof. dr. J.H.W. de Wit	Technische Universiteit Delft TNO

Title: Deuterium in the Gate Dielectric of CMOS Devices
Author: Albert Jan Hof
ISBN: 90-365-2093-2

This research has been supported by Philips Semiconductors.

DEUTERIUM IN THE GATE DIELECTRIC OF CMOS DEVICES

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 29 oktober 2004 om 13.15 uur

door

Albert Jan Hof

geboren op 11 augustus 1976

te Wolvega

Dit proefschrift is goedgekeurd door:

de promotor prof. dr. J Schmitz
de assistent promotor dr. A.Y. Kovalgin.

Contents

Summary

Most of the electronic integrated circuits used today are Complementary MOS (CMOS) circuits, which consist mainly of Metal-Oxide-Semiconductor Field-Effect Transistors (MOSFETs). In the last forty years there has been a tremendous reduction of the MOSFET dimensions. This reduction will continue, enabling even faster and more complex integrated circuits. But, there are a number of hurdles on the road. One of these hurdles is the thickness reduction of an essential electrically isolating layer inside the MOSFET, the so-called gate dielectric. This gate dielectric is becoming so thin, it starts to leak electrical current under operating conditions. This increases the power consumption and can lead to a non-functional transistor. Reliability is also of concern, because the gate dielectric deteriorates under device operation, leading to even larger leakage currents.

The degradation of the gate dielectric is due to the creation of defects in its bulk and at its interface with the substrate underneath. Hydrogen is believed to play a crucial role in this. Some reports in literature suggest that replacing hydrogen with its isotope deuterium during device fabrication can improve the resistance to degradation. This claim is investigated in this thesis. The gate dielectric used throughout this thesis is silicon oxide.

The most direct method to incorporate deuterium is during the growth of the gate dielectric, using a deuterated ambient. To this means, silicon oxidation data, *i.e.* oxide thickness as a function of oxidation time, has been obtained for silicon oxidation in ultra-diluted (partial pressure 0.7 kPa) H_2O and D_2O ambient at different temperatures. The oxidation data shows that the silicon oxidation rate in a D_2O ambient is 18 % lower than in a H_2O ambient. This is independent of oxidation time and oxide thickness. A difference of 15 to 17 meV in activation energy suffices to give a change of 18 % in oxidation rate in the investigated temperature range of 750 to 950 °C. An

attempt to model the obtained oxidation data with existing physical oxidation models is not satisfactory. The Deal-Grove, Reisman and Wolters model are equally good in describing the data. Since the Reisman and Wolters model do not specify how the model parameters depend on oxidation temperature and pressure, the Deal-Grove model is preferred. However, the parameter values needed to fit the Deal-Grove model to the obtained data are not in accordance with values reported in literature and the relative error still remains in the range of 20 %. In this thesis, a new silicon oxidation model has been developed, taking into account electron-stimulated dissociation of the oxidising species and the out-diffusion of the dissociated species. It is a full physical model with a limited number of parameters. The model is able to fit the silicon oxidation data within 30 % accuracy. It is expected that this can be improved if the electron-dissociated stimulation is not only taken to occur at the SiO_2 surface and the Si-SiO_2 interface, but also in the bulk of the silicon oxide.

The obtained knowledge on the oxidation kinetics of silicon in a deuterated ambient has been used to fabricate CMOS devices with deuterium in the gate dielectric. Deuterium has also been introduced during other processing steps. Measurements regarding the bulk quality of the gate dielectric of these devices show no improvement of the resistance to degradation. Deuterated gate dielectrics degrade as fast as hydrogenated gate dielectrics under Fowler-Nordheim stressing conditions. This is independent of the processing stage in which deuterium was introduced. The results are well acceptable in view of our current understanding of SiO_2 degradation, but do contradict with several experimental observations reported in literature. On the other hand, measurements regarding the interface quality, do show an improvement of the resistance to degradation. However, this only occurs when the deuterium is introduced during the post-metal anneal at the end of the process. Earlier incorporation of deuterium does not show improved resistance to hot carrier degradation. Most likely, because the early introduced deuterium is replaced with hydrogen during high temperature processing steps that follow the step introducing deuterium. Hydrogen is abundantly available in CMOS manufacturing, so this is a plausible idea.

In conclusion, it can be stated that for deuterium to be beneficial for CMOS devices, it has to be incorporated at the end of the CMOS manufacturing process, to avoid replacement with hydrogen. The only benefit will be a higher resistance to hot carrier degradation. This will allow for higher internal electric fields in the CMOS devices.

Hot carrier degradation has been a primary reason for reduction of the supply voltage. The incorporation of deuterium can relax this demand for a decreased supply voltage.

Samenvatting

De meeste huidige geïntegreerde schakelingen zijn Complementaire MOS (CMOS) schakelingen, welke hoofdzakelijk bestaan uit veldeffect transistoren, de zogenaamde MOSFETs. Gedurende de laatste veertig jaar zijn de afmetingen van deze MOSFETs enorm gereduceerd. Deze reductie gaat voorlopig nog door en maakt nog snellere en meer complexe schakelingen mogelijk. Maar er zijn een aantal obstakels. Eén van deze obstakels is de reductie van de dikte van een essentieel elektrisch isolerend laagje in de MOSFET, het zogenaamde *gate*-diëlektricum. Dit *gate*-diëlektricum wordt zo dun, dat er elektrische stroom doorheen begint te lekken onder gebruikerscondities. Dit verhoogt de energieconsumptie en kan tot niet-functionele transistoren leiden. Een ander heikel punt is de betrouwbaarheid. Tijdens gebruik 'slijt' het *gate*-diëlektricum als het ware, wat leidt tot nog grotere lekstromen.

De degradatie van het *gate*-diëlektricum wordt veroorzaakt door het ontstaan van defecten zowel in het *gate*-diëlektricum als op het grensvlak van het *gate*-diëlektricum met het onderliggende substraat. Er wordt verondersteld dat waterstof hierin een cruciale rol speelt. In een aantal wetenschappelijke artikelen wordt gesteld dat het vervangen van waterstof met diens isotoop deuterium tijdens de fabricage van de MOSFETs de weerstand tegen degradatie kan verhogen. Deze claim is nader onderzocht in dit proefschrift. Het *gate*-diëlektricum dat gebruikt wordt in dit proefschrift is siliciumoxide.

De meest directe manier om deuterium in te brengen, is tijdens de groei van het *gate*-diëlektricum, in een gedeutereerde omgeving. Hiertoe is data over de oxidatie van silicium, d.w.z. oxidedikte tegen oxidatietijd, vergaard voor siliciumoxidatie in een ultra-verdunde (partieel druk 0.7 kPa) H_2O en D_2O omgeving bij verschillende oxidatietemperaturen. De oxidatiedata laat zien dat de snelheid van siliciumoxidatie in een D_2O omgeving 18 % lager is dan in een H_2O

omgeving. Dit is onafhankelijk van de oxidatietijd en oxidedikte. Een verschil van 15 tot 17 meV in activeringsenergie is voldoende om het verschil van 18 % te verklaren in het onderzochte temperatuurbereik van 750 tot 950 °C. De modellering van de verkregen oxidatiedata met bestaande modellen van siliciumoxidatie is niet toereikend. De modellen van Deal en Grove, Reisman en Wolters zijn alle even goed in staat om de data te beschrijven. Het model van Deal en Grove verdient de voorkeur, aangezien de modellen van Reisman en Wolters niet specificeren hoe de modelparameters afhangen van de oxidatietemperatuur en -druk. Echter, de parameterwaarden die nodig zijn voor het model van Deal en Grove om de verkregen data te beschrijven zijn niet in overeenstemming met de waarden uit de vakliteratuur. Bovendien blijft de relatieve fout rond de 20 %. In dit proefschrift wordt een nieuw model voor siliciumoxidatie voorgesteld, waarin elektron-gestimuleerde dissociatie van het oxiderende gas en de uitdiffusie van de gedissocieerde gassen worden meegenomen. Het is een volledig fysisch model met een beperkt aantal parameters. Dit model is in staat om de verkregen oxidatiedata te beschrijven binnen een nauwkeurigheid van 30 %. Naar verwachting kan dit verbeterd worden door de elektron-gestimuleerde dissociatie niet alleen plaats te laten vinden aan het SiO_2 oppervlak en aan het Si-SiO_2 grensvlak, maar ook binnen in het siliciumoxide.

Het inzicht dat verkregen is in de oxidatiekinetiek van silicium in een gedeutereerde omgeving is gebruikt om CMOS componenten te maken met deuterium in het *gate*-diëlektricum. Deuterium is ook geïntroduceerd tijdens andere processtappen. Metingen naar de kwaliteit van het binnenste van het *gate*-diëlektricum van deze componenten laten geen verbetering zien van de weerstand tegen degradatie. *Gate*-diëlektrica met deuterium degraderen even snel als *gate*-diëlektrica met waterstof. Dit is onafhankelijk van de processtap waarin deuterium is geïntroduceerd. Dit resultaat is acceptabel gezien de huidige kennis van SiO_2 degradatie, maar strookt niet met verschillende experimentele observaties in de vakliteratuur. Aan de andere kant, laten metingen naar de kwaliteit van het grensvlak van de *gate*-diëlektrica wel een verbetering van de weerstand tegen degradatie zien. Dit gebeurt echter alleen wanneer deuterium wordt geïntroduceerd tijdens de laatste processtap, de *post-metal anneal*. Eerdere introductie van het deuterium laat geen verbetering zien. Waarschijnlijk wordt vroegtijdig geïntroduceerde deuterium tijdens latere processtappen op hoge temperatuur vervangen door waterstof. Aangezien waterstof rijkelijk beschikbaar is tijdens het productiepro-

ces, is dit een plausibel idee.

Concluderend kan er gesteld worden dat deuterium nuttig is voor CMOS schakelingen, mits het aan het eind van het productieproces geïntroduceerd wordt, om vervanging door waterstof te voorkomen. Het enige voordeel is een hogere weerstand tegen zogenaamde *hot carrier* degradatie aan het grensvlak. Dit laat hogere interne elektrische velden toe in de CMOS componenten. *Hot carrier* degradatie is de belangrijkste reden voor de reductie van de voedingsspanning. De introductie van deuterium in het *gate*-diëlektricum kan de eis voor een verlaagde voedingsspanning verzachten.

Chapter 1

Introduction

The development of electronic integrated circuits has had and still has a large impact on society, enabling a highly automated world. Most of the electronic integrated circuits used today are Complementary MOS (CMOS) circuits, which consist mainly of Metal-Oxide-Semiconductor Field-Effect Transistors (MOSFETs). In the last forty years the semiconductor industry has been able to reduce the dimensions of the MOSFET tremendously. According to the International Technology Roadmap for Semiconductors [1] this reduction continues until at least 2018, enabling even faster and more complex integrated circuits. But, there are a number of hurdles on the road. One of these hurdles is the reduction of the gate dielectric thickness, the subject of this thesis.

Figure 1.1 shows the basic drawing of a MOSFET. The MOSFET consists of a semiconductor (usually silicon) substrate, a source, a drain and a gate. For a n-channel MOSFET, the substrate is doped such that the majority free electric charge carriers are holes, and the source and drain are doped such that the majority free electric charge carriers are electrons. For a p-channel MOSFET the doping is reversed. The gate is electrically isolated from the substrate by the gate dielectric. When no voltage is applied to the gate, there is no electrical connection between the source and drain. When a voltage is applied to the gate (positive for a n-channel MOSFET and negative for a p-channel MOSFET), a channel of minority charge carriers will form directly underneath the gate dielectric. This channel electrically connects the source and drain. With a voltage bias applied between source and drain, an electrical current will flow. The magnitude of

the electrical current is regulated with the gate-to-substrate voltage.

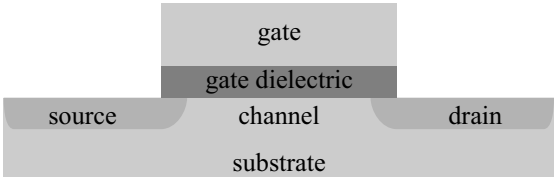


Figure 1.1: Basic layout of a Metal-Oxide-Semiconductor Field-Effect Transistor (MOSFET).

In state of the art MOSFETs, the thickness of the gate dielectric is only 1 to 2 nm. The dielectric is only a few atomic layers thick. Actually, the layer has become so thin, it starts to leak electrical current and the power consumption of the electronic circuit increases. When the leakage current becomes so large that it becomes comparable to the drain current, the transistor loses its functionality. In reality, this situation is not reached, because the increase in power consumption puts severe constraints on the maximum allowable gate leakage current. Especially for mobile applications (*e.g.* laptop and mobile telephone) which rely on battery power, this poses a problem.

Reliability of the gate dielectric is also of concern. Under device operation conditions, the gate dielectric deteriorates, which finally results in the local breakdown of the dielectric. The dielectric has become a conductor. This strongly increases the power consumption and depending on the function of the MOSFET in the overall integrated circuit, the circuit will not function properly anymore.

A special CMOS device extra sensitive to gate dielectric degradation is the floating gate device. It is used as a non-volatile memory device and can be found in many applications. Figure 1.2 shows the basic drawing of a floating gate device. Shown is a MOSFET with a stacked double gate. The top gate is the control gate and the bottom gate is the floating gate. The floating gate is completely electrically isolated by surrounding dielectric material. The dielectric between the floating gate and the substrate is called the tunnel dielectric. When proper voltages are applied to the control gate, source, drain and substrate, electric charge can be forced through the tunnel dielectric from the substrate to the floating gate. The reverse is also possible. The resulting charge on the floating gate represents the

stored information and can be read by sensing the drain current for a certain control gate voltage. Degradation of the tunnel dielectric implies that charge can leak away from the floating gate. This is loss of the stored information. The floating gate device does not function properly anymore.

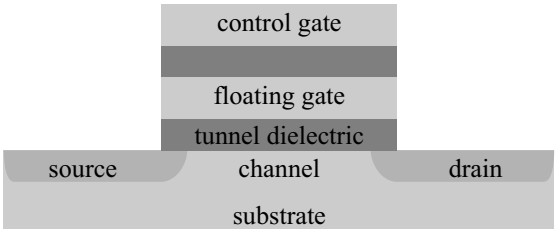


Figure 1.2: Basic layout of a floating gate type non-volatile memory device.

Degradation of the gate (tunnel) dielectric occurs due to the creation of defects in the dielectric and at the interface between the dielectric and the substrate. These defects are created during device operation. They are local chemical alterations, like broken chemical bonds, to the dielectric and it's interface with the substrate. There is evidence that hydrogen in CMOS devices plays a crucial role in both the creation of and passivation of defects. Hydrogen is present during many steps in the CMOS manufacturing process and is therefore abundantly available in the resulting CMOS devices.

Scanning tunnelling microscope experiments by Avouris *et al.* [2] showed that deuterium is harder to desorb from a silicon surface than hydrogen. This raised the idea to replace hydrogen during some stage in the CMOS manufacturing process with deuterium to improve the passivation of defects and therefore the device stability. Reports (on hot carrier degradation [3], charge-to-breakdown [4] and plasma-process induced damage [5]) indicate that the reliability of the gate dielectric in MOSFETs can be improved by introducing deuterium in the device. Improved light stability of amorphous silicon for solar cell applications [6] and improved reliability of electroluminescence [7] from MOS tunnelling diodes due to deuterium incorporation have also been reported.

Deuterium[1] is a hydrogen isotope. In addition to the one proton

[1]The chemical notation for deuterium is ^{2}H or D. The last is adopted throughout this thesis.

and one electron a hydrogen atom consists of, a deuterium atom also has one neutron. Since the mass of a proton and a neutron are almost equal and much larger than the mass of an electron, a deuterium atom is twice as heavy as a hydrogen atom. This is why deuterium is sometimes called heavy hydrogen. The chemical properties of an element are controlled by the number of electrons of the element, therefore hydrogen and deuterium are chemically almost equal, but physically the factor 2 difference in mass can play a significant role. For example, the vibrational frequency of a chemical bond is proportional with the square root of the mass of the atoms involved. Consequently, a bond between a silicon atom and a hydrogen atom will vibrate at 1.4 times the frequency of a bond between a silicon atom and a deuterium atom.

The objective of this thesis is to investigate the impact of deuterium incorporation in the gate dielectric of CMOS devices on the electrical stability of these devices. The gate dielectric used throughout this thesis is silicon oxide. Deuterium can be introduced into the CMOS manufacturing process during a number of processing steps. The most direct method to incorporate deuterium into the gate dielectric is to use a deuterated method to grow the gate dielectric, *i.e.* to use heavy water (D_2O) or a mix of deuterium and oxygen as oxidising ambient. Little is known about the growth of silicon oxide in a deuterated ambient. To be able to grow in a deuterated ambient a silicon oxide layer with the desired thickness, a detailed knowledge of the oxidation kinetics is necessary. Therefore, chapter 2 provides a set of experimental thickness data on silicon oxidation in both a H_2O and a D_2O ambient for a variety of oxidation conditions. An attempt is made to model this data with existing silicon oxidation models. This modelling is not satisfactory and therefore a new silicon oxidation model is developed in chapter 3.

The growth of the gate dielectric is an early step in the process of CMOS manufacturing and many steps follow. Some of these steps introduce hydrogen. It is possible that this hydrogen replaces the deuterium introduced earlier. Chapter 4 describes a general CMOS manufacturing process and identifies the hydrogen-containing steps. The hydrogen in these steps can be replaced with deuterium. This leads to later or extra incorporation of deuterium in the CMOS manufacturing. The two extremes are formed by an early incorporation of deuterium as stated above during the gate dielectric growth and a late incorporation of deuterium during the so-called post-metal anneal.

The existing literature on the impact of deuterium incorporation

on the CMOS device stability is inconsistent. The work presented in this thesis aims at resolving this inconsistency. After the description of a general CMOS manufacturing process, the specific details of the manufacturing of the devices used in this work are presented at the end of chapter 4. The electrical measurements performed at these devices are presented in chapters 5 and 6. The measurements presented in chapter 5 investigate the impact of deuterium incorporation in the CMOS manufacturing process on the bulk quality of the gate dielectric. The measurements presented in chapter 6 investigate the impact of deuterium incorporation in the CMOS manufacturing process on the quality of the interface between the substrate and the gate dielectric.

The final conclusions of the work presented in this thesis are given in chapter 7.

Chapter 2

Wet oxidation of silicon

2.1 Introduction

An important candidate processing step to introduce deuterium in the CMOS gate dielectric is the actual growth of the gate dielectric. This gate dielectric usually is silicon oxide, or nitrided silicon oxide. Present MOS devices demand the growth of high quality silicon oxide layers of only a few nanometres. High quality refers to low leakage current, not exceeding the Fowler-Nordheim tunnelling current, negligible fixed and mobile oxide charge and a low defect density ($< 10^{10}\,\mathrm{cm}^{-2}$) at the interface with the underlying silicon. These layers have to be grown in a well controlled manner, ensuring good thickness control and uniformity. The growth is carried out in either a dry or wet ambient, where the electrical quality of wet oxides has been shown to exceed that of dry oxides [8–10]. Recent studies [4, 11] indicate that replacement of hydrogen with deuterium in wet oxides, that is the use of heavy water instead of water, can further enhance the electrical quality of thin oxides. Wet oxides are also preferred in view of the lower thermal budget [12]. This is because wet oxides grow faster than dry oxides. Therefore, to assure reasonable oxidation times and good process control, thin wet oxides have to be grown at either low temperature or reduced water vapour pressure.

Little experimental oxide thickness versus oxidation time data are available on the growth of thin wet oxides. Even less data are available on the growth of thin deuterated wet oxides. In this chapter more insight and understanding into the growth of thin (1 to 60 nm) wet

and deuterated wet oxides is given. For the first time, a wide range of growth rate data of H_2O and D_2O oxides in an ultra-diluted ambient at different temperatures will be presented. In an attempt to gain better understanding of the oxidation process, the experimental data will be compared with various existing oxidation models.

In section 2.2 an overview of existing silicon oxidation models will be given. Three models will be treated in more detail. Section 2.3 gives the details of the experimental setup used to obtain growth data of thin wet and deuterated silicon oxides, followed by the results and discussion of the experiments in section 2.4 and the conclusions in section 2.5.

2.2 Existing silicon oxidation models

Silicon oxidation models are important for CMOS manufacturing. Silicon oxide is widely used during the processing and the most important step involving the growth of silicon oxide is the gate dielectric growth. The thickness of this gate dielectric is only a few nanometers and it has to be precisely controlled. For this reason a proper model of the silicon oxide thickness as a function of processing conditions is important.

There exists a broad range of models describing oxidation of silicon. In general, these models can be divided in three different groups:

1. Physical models, *i.e.* models derived from laws of physics, *e.g.* the Deal-Grove model [13]. These models are preferred when trying to understand the physics behind oxidation. The mathematical complexity of some physical models, *e.g.* the Beck model [14], forms a drawback for its use, especially in process simulators.

2. Mathematically altered physical models, *i.e.* physical models which are extended or altered in a purely mathematical manner without physical background, *e.g.* the Massoud [15] extension to the Deal-Grove model. The mathematical extension is added to overcome fitting problems of the physical model in mostly the thin oxide thickness range.

3. Mathematical models, *i.e.* mathematical models lacking any physical background. These models are intended to fit the available oxidation data in a quick and easy mathematical way.

In this work, the interest is in the physical understanding of thin oxide growth of silicon. Therefore, only physical models will be considered.

A good physical model will be able to describe the oxidation with a limited number of parameters, which are preferably independent of oxidation parameters like temperature and pressure and have a physical meaning, like activation energy or diffusivity. Furthermore, like any model, it should be able to describe measured data within a certain accuracy. Finally, to be of practical use, the complexity of the model should be limited. Because of this last requirement, only the Deal-Grove, the Reisman and the Wolters model will be considered. All three models have a physical basis and result in a simple mathematical relationship between oxidation time and resulting oxide thickness. Other models, like for instance the Beck model [14], result in more complex mathematical relationships.

Deal-Grove model

In 1965, Deal and Grove postulated a linear-parabolic growth model for silicon oxidation [13]. As a first assumption for their model, they assume that 'the oxidation proceeds by the inward movement of a species of oxidant rather than by the outward movement of silicon'. As a consequence the transported species must go through three stages:

1. The species is transported from the bulk of the oxidising gas to the outer silicon oxide surface.

2. The species is transported across the silicon oxide film towards the silicon.

3. The species reacts with silicon to form silicon oxide.

A graphical presentation of this transport is given in figure 2.1. This picture shows the concentration of the oxidising species in the oxidising gas (C_{g}), at the outer surface of the silicon oxide (C_{s}) and at the interface between the silicon oxide and the silicon (C_{i}). Furthermore it shows the flux of oxidising species from the oxidising gas to the outer surface of the silicon oxide (F_{g}), the diffusion flux of oxidising species across the silicon oxide (F_{o}) and the flux of oxidising species corresponding to the oxidation reaction (F_{r}). The silicon oxide thickness is represented by x_{o}.

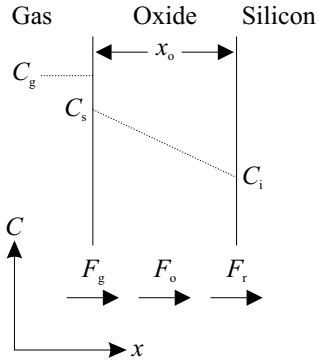


Figure 2.1: Concentration profile and transport of oxidising species during silicon oxidation according to the Deal-Grove model. This picture shows the concentration of the oxidising species in the oxidising gas (C_g), at the outer surface of the silicon oxide (C_s) and at the interface between the silicon oxide and the silicon (C_i). Furthermore it shows the flux of oxidising species from the oxidising gas to the outer surface of the silicon oxide (F_g), the diffusion flux of oxidising species across the silicon oxide (F_o) and the flux of oxidising species corresponding to the oxidation reaction (F_r). The silicon oxide thickness is represented by x_o.

The three fluxes are described by:

$$F_g = h(C_g - C_s) \tag{2.1}$$

$$F_o = D_{eff}\frac{\partial C}{\partial x} = D_{eff}\left(\frac{C_s - C_i}{x_o}\right) \tag{2.2}$$

$$F_r = kC_i \tag{2.3}$$

$$F_g = F_o = F_r = F \tag{2.4}$$

In these equations, h, D_{eff} and k are a gas-phase transport coefficient, the effective diffusion coefficient for the oxidising species in silicon oxide and the reaction rate constant for the formation of silicon oxide from the oxidising species and silicon. The second assumption of Deal and Grove is that of steady-state conditions, that is there is no accumulation or depletion of oxidising species or equivalently the flux of oxidising species is independent of position. This is implemented in equation 2.2 in the rightmost term and by the equality of the three fluxes described in equation 2.4.

Solving equations 2.1 to 2.4 results in:

$$F = \frac{kC_{\mathrm{g}}}{1 + \frac{k}{h} + \frac{kx_{\mathrm{o}}}{D_{\mathrm{eff}}}} \tag{2.5}$$

With N the number of oxidant molecules incorporated in a unit volume of the silicon oxide layer, the growth rate can be expressed as:

$$\frac{dx_{\mathrm{o}}}{dt} = \frac{F}{N} = \frac{kC_{\mathrm{g}}}{N\left(1 + \frac{k}{h} + \frac{kx_{\mathrm{o}}}{D_{\mathrm{eff}}}\right)} \tag{2.6}$$

Deal and Grove integrate this equation starting from an initial silicon oxide thickness x_{i} stating that this initial silicon oxide layer can be regarded as the thickness of the layer grown before the above assumptions become valid. The result is:

$$\frac{x_{\mathrm{o}}^2}{B} + \frac{x_{\mathrm{o}}}{B/A} = (t + \tau) \tag{2.7}$$

with

$$B/A = \frac{C_{\mathrm{g}}}{N\left(\frac{1}{h} + \frac{1}{k}\right)} \tag{2.8}$$

$$B = \frac{2D_{\mathrm{eff}}C_{\mathrm{g}}}{N} \tag{2.9}$$

and

$$\tau = \frac{x_{\mathrm{i}}^2}{B} + \frac{x_{\mathrm{i}}}{B/A} \tag{2.10}$$

The parameters B/A and B are the linear and parabolic rate constants. The linear rate constant is determined by the interface reaction constant, while the parabolic rate constant depends on both the interface reaction constant and the diffusion constant of the oxidation species in silicon oxide. The constant τ gives the time needed to grow an initially present oxide thickness x_{i}. Since temperature dependence of the diffusion and reaction rate constants can be described with Arrhenius equations, the temperature dependence of the linear and parabolic rate constants can also be described with Arrhenius equations. The pressure dependence of the rate constants is linear as this is determined by the concentration of the oxidising species in the oxidising gas.

The Deal-Grove model is a simple and complete model. Moreover, it is able to describe a large portion of the existing silicon oxidation data. Therefore, it is widely applied. However, it is known to be unsuccessful in explaining the growth of thin (0 to 20 nm) oxides in dry oxygen [16], namely oxidation occurs faster than the model predicts.

Reisman model

The Deal-Grove model cannot describe the growth of thin (0 to 20 nm) oxides in dry oxygen. This is why Reisman *et al.* [17] propose a different silicon oxidation model. They show that a wide variety of dry oxidation data can be modelled with a simple power law. They used data from several sources, for both <100> and <111> silicon within a oxidation temperature range of 700 to 1000 °C and a oxidation pressure range of 1×10^{-5} to 20 atm. Later Nicollian and Reisman [18] provided a physical basis for the observed power law.

On the basis of the values of the fitting constants found in [17], Reisman and Nicollian argue that the silicon oxide forming reaction is the rate-limiting step at all times. This is in contrast with the Deal-Grove model which also considers the diffusion of the oxidising species to be a rate-limiting step. The key point in the Reisman model is the assumption that the surface reaction is retarded by the viscous flow of the silicon oxide to accommodate the volume expansion that occurs when silicon oxidises. They state that 'oxidation cannot proceed unless free volume is provided to accommodate the newly forming oxide which occupies approximately 2.2 times the volume of the silicon from which it is formed'.

The above considerations lead to the following description of the growth rate:

$$\frac{dx_{\mathrm{o}}}{dt} = k_{\mathrm{v}}(t,p,T)C_{\mathrm{os}}^{n} \tag{2.11}$$

In this equation, C_{os} is the concentration of the oxidising species at the reaction surface and n is the partial order of the surface reaction. Due to the assumption that diffusion of the oxidising species is not a rate-limiting step, C_{os} is determined solely by the partial pressure of the oxidising species in the oxidising gas.

The reaction rate constant k_{v} depends on oxidation time, pressure and temperature, since it is controlled by the silicon oxide viscosity. In

general, the reaction rate constant can be described by an Arrhenius equation:

$$k_{\mathrm{v}}(t,p,T) = k_{\mathrm{A}} e^{-\frac{[E_{\mathrm{v}}(t,p,T)+E_{\mathrm{r}}]}{kT}} \tag{2.12}$$

The activation energy consists of two parts. The first part is the oxidation time, pressure and temperature dependent energy E_{v} required to produce the free volume by viscous flow to accommodate the volume expansion. The second part, E_{r} is 'the remaining energy required for the reaction. The latter includes the activation energy required to break Si–Si bonds, for example.'

At the same time, the energy E_{v} also represents the activation energy of the average viscosity of the silicon oxide:

$$\eta(t,p,T) = \eta_{\mathrm{o}} e^{-\frac{E_{\mathrm{v}}(t,p,T)}{kT}} \tag{2.13}$$

The crux of the physical basis of the Reisman model lies in a proper time dependent description of the average viscosity of the silicon oxide. To arrive at the desired power law for the silicon oxide growth, and in line with other people describing the time dependence of glass viscosity [19], Nicollian and Reisman use a power law to describe the time dependence of the average viscosity:

$$\eta(t,p,T) = G\,(t-\tau)^{\phi(p,T)} \tag{2.14}$$

In this equation, G is a constant, τ represents an incubation time for initial oxidation and ϕ determines the shape of the η-t curve.

Combining equations 2.11 to 2.14 results in:

$$\frac{dx_{\mathrm{o}}}{dt} = \frac{k_{\mathrm{A}}\eta_{\mathrm{o}}C_{\mathrm{os}}^{n}}{G\,(t+\tau)^{\phi(p,T)}} e^{-\frac{E_{\mathrm{r}}}{kT}} \tag{2.15}$$

Integrating this equation starting at an initial oxide thickness leads to the desired power law:

$$x_{\mathrm{o}} = a\,(t+\tau)^{b} \tag{2.16}$$

with

$$b = 1 - \phi(p,T) \tag{2.17}$$

$$a = \frac{k_{\mathrm{A}}\eta_{\mathrm{o}}C_{\mathrm{os}}^{n}}{Gb} e^{-\frac{E_{\mathrm{r}}}{kT}} \tag{2.18}$$

and

$$\tau = \sqrt[b]{\frac{x_\mathrm{i}}{a}} \tag{2.19}$$

The oxidation temperature and pressure dependence of the model parameters a and b is more complicated than in the case of the Deal-Grove model. A specific temperature and pressure dependence for $\phi(p, T)$ is not available and probably quite complicated. The only simple temperature and pressure dependence that can be derived is that for the product ab. The pressure dependence of this product is determined by C_os. Therefore, the product is similar to the nth power of the pressure. The temperature dependence is described by the Arrhenius term with an activation energy of E_r.

The Reisman model is a rather sophisticated model and not complete. The temperature and pressure dependence of the direct model parameters is as stated in the former paragraph not known. It is capable of describing a wide variety of oxidation data [17].

Wolters model

The last physical model considered, is the model by Wolters and Zegers-van Duijnhoven [20]. Wolters and Zegers-van Duijnhoven have three reasons to propose a different silicon oxidation model: The Deal-Grove model cannot describe the growth of thin oxides in dry oxygen, since silicon oxidation occurs even at room temperature the silicon oxide forming reaction cannot be a rate-limiting step, and the work of Jorgensen [21] shows that the application of an external electric field influences the oxidation rate of silicon and can even completely stop oxidation. The last reason is why Wolters and Zegers-van Duijnhoven consider the oxidation of silicon to occur due to ionic transport. A mechanism which is also used to describe the oxidation of metals.

The derivation of the Wolters model is done for the case of silicon oxidation in an oxygen ambient, but it is also applicable to the case of wet oxidation. The model assumes an electron flow from the Si-SiO_2 interface into the growing silicon oxide, reacting with oxygen diffusing into the oxide from the ambient to form oxygen ions: $O_2 + 2\,z\,e^- \longrightarrow 2\,O^{z-}$. This reaction is assumed to be in a local equilibrium throughout the oxide. Taking this into account as well as the requirement that there is no net electrical current, the model arrives

at:

$$J_{O^{z-}} = -\frac{\left(\sigma_e^{-1} + \sigma_{ion}^{-1}\right)^{-1}}{2z^2q^2}\frac{d\mu_{O_2}}{dx} \tag{2.20}$$

In this equation, $J_{O^{z-}}$ represents the flow of oxygen ions, σ_e and σ_{ion} the conductivity of electrons and oxygen ions in the growing silicon oxide and μ_{O_2} the oxygen thermodynamical potential. When assuming that the conductivities σ_e and σ_{ion} are independent of the position in the silicon oxide, equation 2.20 will result in the classical parabolic ionic-transport model developed by Wagner [22]. However, it is known that a parabolic model is not sufficient to describe silicon oxidation. The Wolters model assumes the oxygen ions to migrate through the silicon oxide layer by jumping from one potential well to another, resulting in a field-dependent ion conductivity:

$$\sigma_{ion} = \sigma_o e^{\left(-\frac{zqaE}{kT}\right)} \tag{2.21}$$

In this equation a represents half the hopping distance and σ_o the ion conductivity at zero electric field.

According to equation 2.21, the ion conductivity decreases exponentially with increasing electric field. A local higher electric field will decrease the ionic transport. According to the Wolters model, to sustain the necessary ion flow $J_{O^{z-}}$, the local electric field has to increase. This is a self-supporting effect with as a result that most of the thermodynamic potential difference stands over a thin layer of silicon oxide at the Si-SiO_2 interface with low ion conductivity. The ion conductivity in the remaining silicon oxide layer is high. Taking the series connection of these two conductivity regions into account in equation 2.20 the Wolters model arrives after a few approximations at a power-parabolic law:

$$x_o^2 + Ax_o^{2-\alpha} = Bt \tag{2.22}$$

with

$$\alpha = \frac{a\left(\sigma_e^{-1} + \sigma_{ion}^{-1}\right)^{-1}}{kT}\frac{\Delta\mu_{O_2}}{x_Q} \tag{2.23}$$

$$A = \frac{2}{2-\alpha}e^{\alpha}x_Q^{\alpha}\frac{\sigma_e}{\sigma_o + \sigma_e} \tag{2.24}$$

and

$$B = \frac{\left(\sigma_e^{-1} + \sigma_{ion}^{-1}\right)^{-1}\Delta\mu_{O_2}}{Nz^2q^2} \tag{2.25}$$

With N the number of oxidant molecules incorporated in a unit volume of the silicon oxide layer and x_Q, the thickness of the low ion conductivity layer. The model reduces to a linear-parabolic law in case $\alpha = 1$ and to a power law in case $A \gg x^\alpha$. The first has the same thickness-time relation as the Deal-Grove model and the latter as the Reisman model. As with the Reisman model, there is no clear relation between the model parameters α, A and B and the oxidation temperature and pressure.

2.3 Experimental setup

A wide range of oxidation experiments in ultra-diluted wet ambient were carried out to investigate the silicon oxide growth rate. The experiments were conducted using 100 mm p-type wafers with a <100> orientation and a boron doping concentration of $6 \times 10^{14}\,\mathrm{cm}^{-3}$. Prior to oxidation the wafers were cleaned and dipped in a 1 % HF solution until the wafer surface was hydrophobic. Directly following the final rinse, the wafers were loaded into the oxidation furnace.

The oxidation was performed in a horizontal furnace. Depending on the valve settings, dry nitrogen flowed either directly or via a small tank into the furnace, see figure 2.2. The nitrogen flow rate was set at 4 l/min. The small tank contained either H_2O or D_2O allowing the nitrogen to take up (heavy) water vapour. The temperature of the tank, which is the minimum temperature in the system, determined the water vapour pressure of the outcoming wet nitrogen gas. For H_2O the tank temperature was set at 17.5 °C and for D_2O at 20 °C. These temperatures correspond to a saturated vapour pressure of 0.020 atm for both species [23].

The actual vapour pressure of the outcoming nitrogen gas was estimated by measuring the mass loss of the tank during the oxidation. The measured mass loss is plotted versus oxidation time in figure 2.3. The slope of the linear regression lines is 1.21 ± 0.01 mmol/min for H_2O and 1.16 ± 0.01 mmol/min for D_2O. There is a difference of 4 %. Taking into account the nitrogen flow of 4 l/min, the precursor partial pressure is calculated to be 7×10^{-3} atm.

The wafers were loaded at a temperature of 700 °C under dry nitrogen flow. Nonetheless, during loading, clean room air is transported into the furnace along with the wafers. To minimise the effect of this clean room air, the furnace temperature was maintained for five minutes at maximum nitrogen flow. Then, the temperature was

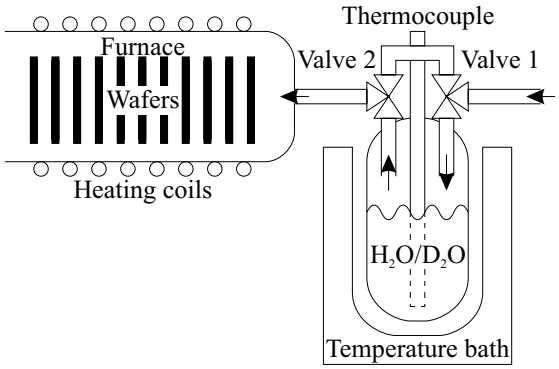


Figure 2.2: The oxidation system. During standby, loading, temperature ramp up and down and unloading, the valve settings are such that dry nitrogen flows directly into the furnace. During oxidation the valves are switched to lead dry nitrogen through the small tank taking up H_2O or D_2O vapour before entering the furnace. The temperature of the small tank and therefore the saturated vapour pressure of the (heavy) water is controlled with a temperature bath and thermocouple. The temperature of the furnace is controlled with heating coils.

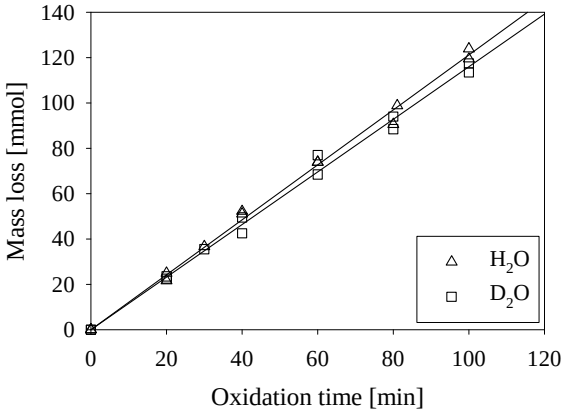


Figure 2.3: Amount of oxidising species introduced into the furnace as a function of oxidation time.

ramped up to the oxidation temperature with 10 °C/min. Oxidation time and temperature were varied from 0 to 960 min and from 750 to 950 °C. After the oxidation temperature was reached, the nitrogen was passed through the tank allowing it to take up (heavy) water

vapour. After oxidation, the temperature was ramped down under dry nitrogen flow to 700 °C with 2.5 °C/min, before the wafers were unloaded.

The resulting oxide thickness was measured at 80 points across the wafer and averaged. The edge exclusion was set to 5 mm. The measurements were carried out with a PLASMOS ellipsometer using a He:Ne laser ($\lambda = 632.8\,\mathrm{nm}$) at an incident angle of 70°. The refractive index was fixed at 1.456.

The measured oxidation data were compared with existing oxidation models by fitting the models with a least squares fit minimising the relative error between the data and the model.

2.4 Results and discussion

Oxide thickness was measured as a function of oxidation time and temperature. The measurements across the wafers showed that the thickness non-uniformity of the grown oxide layers was less than 5 %. The measured oxide thickness is plotted versus time in figure 2.4. The values are listed in table 2.1. Figure 2.4 indicates that, for the same oxidation time and temperature, the oxide grown in H_2O ambient is considerable thicker than the oxide grown in D_2O ambient.

The distance between the curves on the log-log scale appears to be constant which implies a constant ratio between the oxide thickness grown in H_2O and D_2O ambient independent of time and temperature. This observation is confirmed in figure 2.5. For the investigated oxidation pressure the ratio is:

$$\frac{x_{\mathrm{H_2O}}(t,T)}{x_{\mathrm{D_2O}}(t,T)} = 1.18 \pm 0.07 \tag{2.26}$$

The difference of 18 % in oxide growth could be due to either the measured difference in water partial pressure, or the different oxidation ambient. However, it follows from [16] that typically the pressure dependence of oxide growth is described by a power law:

$$x \sim p^n \tag{2.27}$$

with $0.5 \leq n \leq 1.0$. This implies that the 4 % difference in H_2O/D_2O partial pressure will at most result in 4 % difference in oxidation rate. The difference in oxidation pressure cannot completely explain the measured 18 % difference in oxidation rate. Therefore, the difference

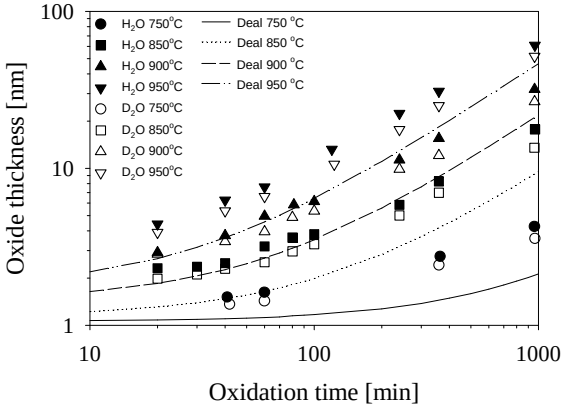


Figure 2.4: Silicon oxide thickness (measured by ellipsometer) as a function of oxidation time for oxidation in a H_2O and D_2O ambient (partial pressure is 7×10^{-3} atm) in the temperature range of 750 to 950 °C. The markers indicate the measured data, while the lines indicate the prediction of the Deal-Grove model for a H_2O ambient using the parameter values extracted by Deal [24]. These values are listed in table 2.2.

Table 2.1: Measured silicon oxide thickness (in nm) for oxidation in a H_2O and D_2O ambient (partial pressure is 7×10^{-3} atm) for an oxidation time of 0 to 960 min and an oxidation temperature of 750 to 950 °C.

Time	750 °C		850 °C		900 °C		950 °C	
(min)	H_2O	D_2O	H_2O	D_2O	H_2O	D_2O	H_2O	D_2O
0	1.06	1.06	1.14	1.14	1.43	1.43	1.71	1.71
20	-	-	2.31	1.98	2.90	2.84	4.41	3.90
30	-	-	2.36	2.11	-	-	-	-
40	1.51	1.36	2.49	2.29	3.73	3.42	6.28	5.35
60	1.62	1.43	3.18	2.52	4.96	3.94	7.61	6.62
80	-	-	3.62	2.96	5.85	4.88	-	-
100	-	-	3.81	3.29	6.15	5.34	-	-
120	-	-	-	-	-	-	13.24	10.63
240	-	-	5.86	5.01	11.31	9.89	22.44	17.71
360	2.74	2.42	8.27	6.99	15.52	12.11	30.96	25.14
960	4.26	3.57	17.73	13.55	31.87	26.65	60.99	51.77

in oxidation rate must largely be attributed to an isotope effect. For example, a difference of 15 to 17 meV in activation energy suffices to give a change of 18 % in oxidation rate in the temperature range of 750 to 950 °C.

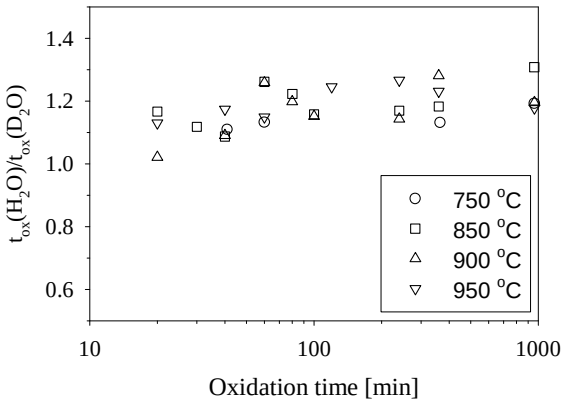


Figure 2.5: Oxide thickness for H_2O oxidation divided by oxide thickness for D_2O oxidation as a function of oxidation time in the temperature range of 750 to 950 °C.

Table 2.1 also lists the oxide thickness for zero oxidation time. This is the oxide grown in clean room ambient at room temperature, during loading, temperature ramp-up, temperature ramp-down and unloading, without actual oxidation in H_2O or D_2O ambient. This oxide has to be attributed mainly to clean room ambient coming into the furnace during loading and is taken to be the initial oxide thickness x_i for actual oxidation as intended in this experiment.

The measured oxidation data were first compared with the Deal-Grove model. Deal [24] measured the model parameters B and B/A for pyrogenic H_2O oxidation at a partial pressure of 640 torr. These values are commonly used [16]. There are no known values for oxidation in a D_2O ambient. Table 2.2 lists the parameter values measured by Deal for a partial pressure of 0.84 atm (640 torr) and the extrapolated values for 7×10^{-3} atm, using the linear pressure dependence of the parameters. The oxidation curves predicted by using these values for H_2O oxidation are depicted in figure 2.4 by the lines. The initial oxide thickness x_i is set equal to the measured oxide thickness for zero oxidation time as listed in table 2.1. Comparison of the Deal-Grove model with the H_2O oxidation data shows that the Deal-Grove model consequently underestimates the oxide thickness in the investigated range by 15 to 40 %.

Table 2.2 also lists fitted parameters for both H_2O and D_2O oxidation. These parameters are obtained by minimising the relative square deviation between model and data. The fitted values for the

Table 2.2: *Deal-Grove* model parameter values for a partial pressure of 0.84 atm (640 torr) as measured by Deal [24] and extrapolated to 7×10^{-3} atm. Also given are the fitted parameter values and fitted initial oxide thickness, x_i, for a relative least square error fit of the measured oxidation data.

	B/A		B		x_i		
	C (μm/hr)	E_a (eV)	C (μm^2/hr)	E_a (eV)	T (°C)	(nm) H_2O	D_2O
0.84 atm	9.70×10^7	2.05	386	0.78	750	1.26	1.16
7×10^{-3} atm	8.07×10^5	2.05	3.21	0.78	850	1.40	1.29
fit for H_2O	2.28×10^6	2.05	0.624	0.82	900	1.50	1.50
fit for D_2O	1.73×10^6	2.05	0.463	0.82	950	1.77	1.75

activation energies are equal or close to the values obtained by Deal, but the pre-exponential factors are deviating strongly. The classical Deal-Grove model clearly fails. Figure 2.6 compares the Deal-Grove model using the fitted parameters to the measured data from table 2.1. The deviation between the model and the data is large for both short and long oxidation times. The relative error between fitted model and measured data still is in the range of 20 %. It can be concluded that, though simple and applicable to a wide range of data, the Deal-Grove model does not fit the measured data for a wet ultra-diluted ambient in a satisfactory manner.

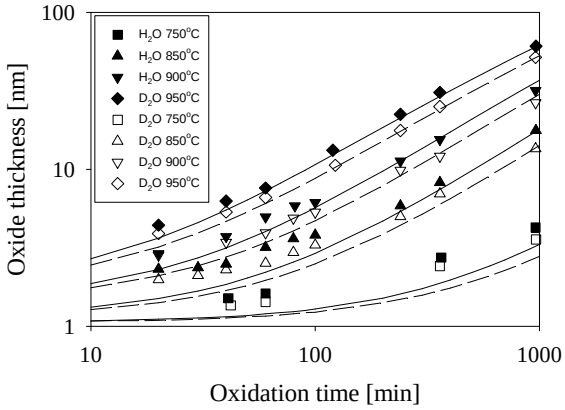


Figure 2.6: Comparison of the measured data from table 2.1 and the Deal-Grove model using the fitted parameter values listed in table 2.2. The solid lines are the fitted curves for H_2O and the dashed lines are the fitted curves for D_2O.

Table 2.3: Reisman a parameter values and fitted initial oxide thickness, x_i, which best fits the measured oxidation data of table 2.1. The best fitting activation energy and pre-exponential factor for the product ab are 0.40 eV and 8.87 nm/min^b for H_2O and 0.42 eV and 8.96 nm/min^b for D_2O respectively.

T (°C)	a (nm/min^b)		x_i (nm)	
	H_2O	D_2O	H_2O	D_2O
750	0.232	0.176	1.08	1.07
850	0.243	0.195	1.17	1.17
900	0.248	0.201	1.42	1.46
950	0.251	0.207	1.76	1.77

The second model that is compared with the oxidation data is the Reisman model. For this model no known values are available for silicon oxidation in either a H_2O or D_2O ambient. Secondly, the model only provides an explicit temperature dependence for the product of the parameters a and b. This implies that for every temperature, a has to be fitted separately and b follows from a and the Arrhenius pre-exponential factor and activation energy of the product ab.

The fitted parameter values for the Reisman model are listed in table 2.3. The relative error between the Reisman model using these parameter values and the measured data is plotted in figure 2.7. This figure also shows the relative error for the Deal-Grove model. The error is on average less for the Reisman model, but the maximum error still is 15 %. Furthermore, the fact that there is no explicit temperature dependence for a forms a drawback for using this model. Therefore, although the fit for the Reisman model is slightly better, there is no benefit in using this model instead of the Deal-Grove model.

The last model that is compared with the measured oxidation data is the Wolters model. This model does not supply an obvious temperature dependence for its parameters. Therefore, the three model parameters α, A and B have to be fitted separately for every temperature. This accumulates to the matrix of parameters listed in table 2.4. As with the Reisman model there are no known values in literature to be compared to. The fact that none of the parameters have an explicit temperature dependence forms a severe drawback for using this model. Furthermore, as can be seen in figure 2.7 the relative error between the fitted Wolters model and the measured data is almost identical to the relative error obtained for the Reisman

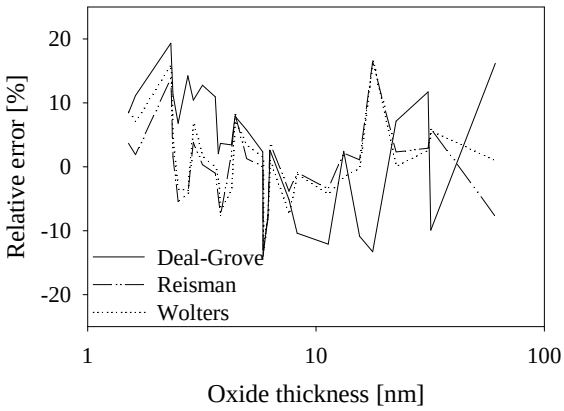


Figure 2.7: Relative error between the measured data from table 2.1 and the three oxidation models considered in this chapter as a function of oxide thickness. The parameters values used in the models are listed in tables 2.2, 2.3 and 2.4.

Table 2.4: Wolters model parameter values which best fits the measured oxidation data of table 2.1.

T (°C)	α (-)		a (nm^{α})		b (nm^2/min)		x_i (nm)	
	H_2O	D_2O	H_2O	D_2O	H_2O	D_2O	H_2O	D_2O
750	5.31×10^{-4}	6.56×10^{-4}	2.15	2.03	0.0579	0.0384	1.12	1.10
850	0.561	0.660	7.29	5.78	0.601	0.324	1.20	1.20
900	0.758	0.730	30.2	25.8	3.06	2.16	1.46	1.48
950	0.870	0.806	63.9	85.1	10.5	12.0	1.73	1.74

model. Taking into account the lack of a proper description of the temperature dependence of the model, the Deal-Grove model is still preferred, although the parameters values needed for the best possible fit are deviating strongly from the values obtained by Deal [24].

2.5 Conclusions

Oxidation data, *i.e.* oxide thickness as a function of oxidation time, has been obtained for silicon oxidation in ultra-diluted (partial pressure 0.7 kPa) H_2O and D_2O ambient at different temperatures. The oxidation data shows that the silicon oxidation rate in a D_2O ambient

is 18 % lower than in a H_2O ambient. This is independent of oxidation time and thickness. A difference of 15 to 17 meV in activation energy suffices to give a change of 18 % in oxidation rate in the investigated temperature range of 750 to 950 °C.

An attempt to model the obtained oxidation data with existing physical oxidation models is not satisfactory. The Deal-Grove, Reisman and Wolters model are equally good in describing the data. Since the Reisman and Wolters model do not specify how the model parameters depend on oxidation temperature and pressure, the Deal-Grove model is preferred. However, the parameter values needed to fit the Deal-Grove model to the obtained data are not in accordance with values reported in literature and the relative error still remains in the range of 20 %.

Chapter 3

New oxidation model

3.1 Introduction

In the former chapter, oxidation data, *i.e.* oxide thickness as a function of oxidation time, has been obtained for silicon oxidation in ultra-diluted H_2O and D_2O ambient at different temperatures. An attempt to model the obtained oxidation data with existing physical oxidation models is not satisfactory. This raises the need for a new oxidation model. As already stated in section 2.2 a proper silicon oxidation model is necessary to be able to predict the obtained silicon oxide thickness for a certain set of processing conditions. Especially for the gate dielectric a good control of the oxide thickness is essential. Furthermore, a sound physical model can give more understanding in the silicon oxidation process. This better understanding might be linked to the quality of a silicon oxide layer for a certain set of processing conditions.

In this chapter a new silicon oxidation model is proposed to explain the data obtained in the former chapter. Before postulating the new silicon oxidation model, section 3.2 takes a closer look at the initial oxidation regime of silicon. Based on this, the new oxidation model incorporates a tunnelling regime. The concept of electron tunnelling will first be introduced in section 3.3 and will be followed by the postulation of the new silicon oxidation model in section 3.4. In section 3.5, the new oxidation model will be verified on the oxidation data obtained in the former chapter. Finally, conclusions will be drawn in section 3.6.

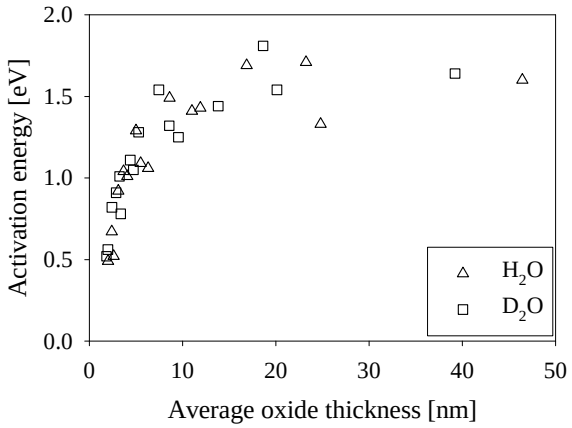


Figure 3.1: Activation energy extracted from the measured oxidation data in table 2.1 as a function of average oxide thickness for H_2O and D_2O ambient. The activation energy is extracted from every pair of adjacent measurement points for the same oxidation time, but different oxidation temperatures.

3.2 Initial oxidation regime

The three models discussed in chapter 2 are not completely satisfactory in explaining the measured data in a solid physical way. While a solid physical use of the Deal-Grove model results in a poor fit, the other two models fit just slightly better and do not relate one or more parameters to temperature, resulting in a separate set of parameters for every temperature.

To obtain further insight in the oxidation process, the activation energy for the oxide thickness is calculated for different thicknesses. The result is shown in figure 3.1. Two observations can be made from this graph. Firstly, the activation energy for H_2O and D_2O grown oxides is similar. Secondly, the activation energy is changing going from ultra-thin oxides to thin oxides. This indicates a change in oxidation mechanism. However, a reservation has to be made.

The initial oxide thickness x_i influences the measured oxide thickness as a function of time. If the initial oxide had not been present before the actual oxidation process had started, the final oxide would have been thinner. For thick oxides, the influence is small, but for thin oxides, the influence can be substantial. This implies that the activation energy for ultra-thin oxides will be higher than presented

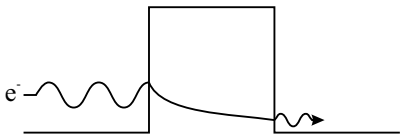


Figure 3.2: Electron tunnelling through an energy barrier.

in figure 3.1.

The trend of a lower activation energy for ultra-thin oxides can be expected, considering the fact that a thin native oxide layer forms on a bare silicon wafer even at room temperature. The growth of this native oxide practically stops when a certain layer thickness is reached. The examined models do not predict this fast initial oxidation at room temperature. A new model extension is necessary to describe the oxide growth in the regime below 10 nm.

The initial oxidation rate is high, but slows down abruptly after a few nanometres. The initial oxidation regime is also relatively weakly temperature dependent because a native oxide will already form at room temperature. This native oxide grows rapidly, but is limited to 1 to 2 nm. After this thickness, the oxidation rate at room temperature is so low that the oxidation practically stops. A viable candidate to explain this initial oxidation behaviour of silicon is a tunnelling-stimulated process. Electron tunnelling is only weakly dependent on temperature and the electron tunnelling probability depends exponentially on the oxide thickness. This can explain the transition from fast to slow oxidation.

3.3 Electron tunnelling

Figure 3.2 shows an electron encountering an energy barrier. In classical mechanics the electron with an energy ϕ_b below the energy barrier is totally reflected by the barrier. However, in quantum mechanics the electron can be described by a probability wave. This probability wave will also be reflected by the barrier, but a small evanescent probability field will exist inside the barrier. When the barrier is thin enough, the evanescent probability field will be non-zero at the other side and the electron has a certain possibility of turning up at the other side of the barrier. This is called tunnelling. The tunnelling probability of an electron in case of a square barrier can be approxi-

mated by [25]:

$$P = e^{\frac{-2x_{\mathrm{o}}\sqrt{2m_{\mathrm{e}}\phi_{\mathrm{b}}}}{\hbar}} \tag{3.1}$$

with x_{o} the thickness of the barrier, m_{e} the effective electron mass inside the barrier, ϕ_{b} the barrier height and *hbar* the reduced Planck constant.

In case of a silicon substrate covered by a thin layer of silicon oxide, the energy of the tunnelling electron can be the bottom of the conduction band and the barrier is formed by the large forbidden band gap of the silicon oxide. The energy level where the electron will tunnel to can be an energy level of an adsorbed molecule at the silicon oxide surface. At practical oxidation temperatures, the concentration of electrons in the silicon, will equal the intrinsic electron concentration $C_{\mathrm{e,i}}$. Due to the tunnel probability of the electrons through the barrier, adsorbed molecules will see a virtual concentration of electrons at the ambient side of the barrier equal to:

$$C_{\mathrm{e,s}} = C_{\mathrm{e,i}}P = C_{\mathrm{e,i}}e^{\frac{-2x_{\mathrm{o}}\sqrt{2m_{\mathrm{e}}\phi_{\mathrm{b}}}}{\hbar}} \tag{3.2}$$

This is the concentration of electrons available for reaction at the surface of the silicon oxide.

3.4 New oxidation mechanism

In an attempt to overcome the problems of the models used in the former chapter, a new oxidation model is proposed. A viable candidate for describing the rapid initial oxidation is electron-stimulated dissociation. Electron tunnelling heavily depends on the silicon oxide thickness, while it hardly depends upon temperature. This electron tunnelling will be embedded in a modified Deal-Grove model.

To accommodate for the electron-stimulated dissociation, it is postulated that the reaction between the oxidising species and the silicon proceeds in two steps. Step I is the electron-stimulated dissociation of the oxidising species, and step II is the formation of silicon oxide from the reaction product $\mathrm{O_iH_j^{b-}}$ and silicon. In chemical formula:

Reaction I $\quad a\,\mathrm{H_2O} + c\,\mathrm{e^-} \xrightarrow{k_{\mathrm{I}}} d\,\mathrm{O}_i\mathrm{H}_j^{b-} + f\,\mathrm{H_2} + g\,\mathrm{e^-}$

Reaction II $\quad d\,\mathrm{O}_i\mathrm{H}_j^{b-} + m\,\mathrm{Si} \xrightarrow{k_{\mathrm{II}}} g\,\mathrm{SiO_2} + n\,\mathrm{H_2} + b\,\mathrm{e^-}$

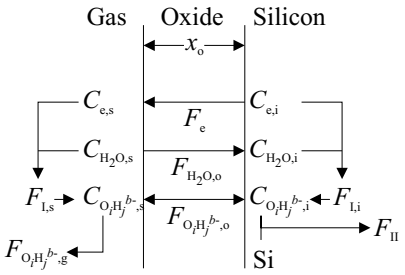


Figure 3.3: Transport of the different species according to the new oxidation mechanism.

The product of reaction I ($O_iH_j^{b-}$) is a species consisting of oxygen and possible hydrogen. The key-point in this model is the fact that the overall reaction is split in two and stimulated by electrons.

Reaction II takes place at the Si-SiO_2 interface. Reaction I can take place both at the Si-SiO_2 interface and at the silicon oxide surface. At the interface electrons are freely available in the silicon, while at the surface, the electrons have to be supplied by tunnelling. An overall picture of the reactions and fluxes of species is given in figure 3.3. In this figure $F_{I,s}$ and $F_{I,i}$ are the fluxes associated with reaction I at respectively the silicon oxide surface and the Si-SiO_2 interface, F_{II} is the flux associated with reaction II at the interface, $F_{O_iH_j^{b-},o}$, $F_{O_iH_j^{b-},g}$, $F_{H_2O,o}$ are fluxes due to the diffusion of the reaction product in the silicon oxide and in the oxidising gas and of the oxidising species in the silicon oxide. The concentrations that play a role are the electron concentrations at the silicon oxide surface $C_{e,s}$ and at the Si-SiO_2 interface $C_{e,i}$, the oxidising species concentrations at the silicon oxide surface $C_{H_2O,s}$ and at the Si-SiO_2 interface $C_{H_2O,i}$ and the reaction product concentrations at the silicon oxide surface $C_{O_iH_j^{b-},s}$ and the Si-SiO_2 interface $C_{O_iH_j^{b-},i}$. The flux of reaction product due to diffusion from the silicon oxide surface into the oxidising gas is included, because due to the production of the reaction product at both the interface and the surface, it is possible that the interface concentration may exceed the surface concentration, resulting in out-diffusion of the reaction product. This definitely occurs for thick oxides when tunnelling is negligible.

The different fluxes are described with the following formulas:

$$F_{I,s} = k_I C_{e,s} C_{H_2O,s} \tag{3.3}$$

$$F_{\mathrm{I,i}} = k_{\mathrm{I}} C_{\mathrm{e,i}} C_{\mathrm{H_2O,i}} \tag{3.4}$$

$$F_{\mathrm{II}} = k_{\mathrm{II}} C_{\mathrm{O_iH_j^{b-},i}} \tag{3.5}$$

$$F_{\mathrm{O_iH_j^{b-},o}} = D_{\mathrm{O_iH_j^{b-},o}} \frac{C_{\mathrm{O_iH_j^{b-},s}} - C_{\mathrm{O_iH_j^{b-},i}}}{x_{\mathrm{o}}} \tag{3.6}$$

$$F_{\mathrm{O_iH_j^{b-},g}} = D_{\mathrm{O_iH_j^{b-},g}} C_{\mathrm{O_iH_j^{b-},s}} \tag{3.7}$$

$$F_{\mathrm{H_2O,o}} = D_{\mathrm{H_2O,o}} \frac{C_{\mathrm{H_2O,s}} - C_{\mathrm{H_2O,i}}}{x_{\mathrm{o}}} \tag{3.8}$$

In these formulas, k_{I} and k_{II} are the reaction rate constants for reaction I and II, $D_{\mathrm{O_iH_j^{b-},o}}$ and $D_{\mathrm{O_iH_j^{b-},g}}$ are the effective diffusion constants for the reaction product in the silicon oxide and in the oxidation gas and $D_{\mathrm{H_2O,o}}$ is the effective diffusion constant for water in silicon oxide. The reaction steps are considered to be of first order and steady-state conditions are imposed on the diffusion fluxes. The assumption of steady-state conditions, i.e. the total flux is constant is further translated in the following three formulas:

$$F_{\mathrm{I,s}} = F_{\mathrm{O_iH_j^{b-},o}} + F_{\mathrm{O_iH_j^{b-},g}} \tag{3.9}$$

$$F_{\mathrm{I,i}} = F_{\mathrm{H_2O,o}} \tag{3.10}$$

$$F_{\mathrm{II}} = F_{\mathrm{I,i}} + F_{\mathrm{O_iH_j^{b-},o}} \tag{3.11}$$

Combining the above equations for the fluxes with the equation for the oxidation rate:

$$\frac{dx_o}{dt} = \frac{F_{II}}{N} \tag{3.12}$$

results in:

$$\frac{dx_{\mathrm{o}}}{dt} = \frac{C_{\mathrm{H_2O,s}}}{N} \frac{D_{\mathrm{O_iH_j^{b-},g}} D_{\mathrm{H_2O,o}} + k_{\mathrm{I}} D_{\mathrm{O_iH_j^{b-},o}} C_{\mathrm{e,s}} + \frac{D_{\mathrm{H_2O,o}} D_{\mathrm{O_iH_j^{b-},o}}}{x_{\mathrm{o}}} \left(1 + \frac{C_{\mathrm{e,s}}}{C_{\mathrm{e,i}}}\right)}{\left(1 + \frac{D_{\mathrm{H_2O,o}}}{x_{\mathrm{o}} k_{\mathrm{I}} C_{\mathrm{e,i}}}\right) \left(x_{\mathrm{o}} D_{\mathrm{O_iH_j^{b-},g}} + D_{\mathrm{O_iH_j^{b-},o}} \left(1 + \frac{D_{\mathrm{O_iH_j^{b-},g}}}{k_{\mathrm{II}}}\right)\right)} \tag{3.13}$$

where N represents the number of oxidant molecules incorporated in a unit volume of the silicon oxide layer. In the case that $D_{\mathrm{O_iH_j^{b-},g}} \to 0$,

that is when out-diffusion of the reaction product does not occur, the above formula can be simplified to:

$$\frac{dx_{\mathrm{o}}}{dt} = \frac{C_{\mathrm{H_2O,s}}}{N}\left(k_{\mathrm{I}}C_{\mathrm{e,s}} + \frac{k_{\mathrm{I}}D_{\mathrm{H_2O,o}}C_{\mathrm{e,i}}}{D_{\mathrm{H_2O,o}} + x_{\mathrm{o}}k_{\mathrm{I}}C_{\mathrm{e,i}}}\right) \tag{3.14}$$

This formula describes a parallel oxidation process with a Deal-Grove component and an electron tunnelling stimulated component. For thick oxides, the tunnelling component will become negligibly small and the new model is exactly the Deal-Grove model, except that in the new model, the reaction between the oxidation species and silicon occurs in two successive steps using electron-stimulated dissociation.

3.5 Verification of the new model

The new oxidation model incorporates a number of parameters; k_{I}, k_{II}, $D_{\mathrm{O_iH_j^{b-},o}}$, $D_{\mathrm{O_iH_j^{b-},g}}$ and $D_{\mathrm{H_2O,o}}$. All five parameters will be described by an Arrhenius equation. Since the new model converges to the Deal-Grove model for thick oxides, the value of $D_{\mathrm{H_2O,o}}$ and k_{I} can be fitted to match the Deal-Grove parameters. The other three parameters are used as fitting parameters.

Figure 3.4 shows the result of fitting the measured oxidation data with the new model. It is clear from this figure that the new model does not provide a perfect fit. The curves for 950 °C can be nicely fit, but for lower oxidation temperatures, the model underestimates the silicon oxide thickness for the middle range of oxidation times. Still, the fit is reasonable and much better than the Deal-Grove model, taking into account the physically correct values. This is due to the addition of the two terms of backstream- and out-diffusion of the reaction product $\mathrm{O_iH_j^{b-}}$. The model supplies a clear temperature and pressure dependence and is therefore also to be preferred over the Reisman and Wolters models.

A remarkable thing to note is that for the used parameter values, the tunnelling component does not play a role. This implies that there is room for improvement of the model. Tunnelling should play a role at least at low temperatures, to be able to provide the growth of for example native oxide at room temperature. The model incorporates the electron-stimulated dissociation only at the surface and the interface. An improvement can be expected when the electron-stimulated dissociation process takes place also inside the oxide. This will increase the importance of the electron tunnelling component and

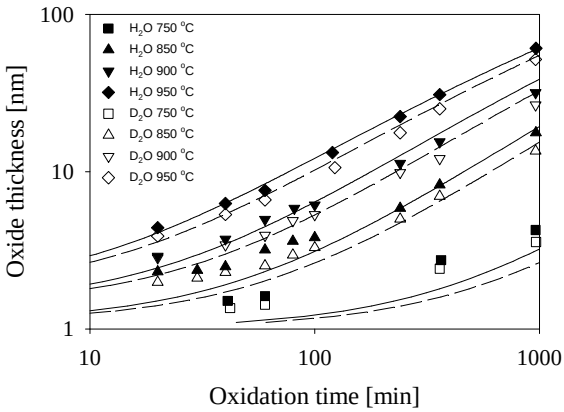


Figure 3.4: Comparison of the measured data from chapter 2 with the oxidation model proposed here.

might be able to remove the underestimation of the model for the middle range oxidation times as well as the prediction of native oxide growth.

3.6 Conclusions

In this chapter, a new silicon oxidation model has been proposed, taking into account electron-stimulated dissociation of the oxidising species and the out-diffusion of the dissociated species. The model is able to fit the silicon oxidation data from chapter 2, but there is room for improvement.

Chapter 4

Device fabrication

4.1 Introduction

Hydrogen is abundantly available during the process of CMOS manufacturing. It is introduced during various processing steps where hydrogen-containing species are used at high ($\geq 400\,^{\circ}\mathrm{C}$) temperature. Substitution of hydrogen with deuterium can in principle be done at any of these hydrogen-containing steps. However, care has to be taken that the deuterium introduced is not replaced by hydrogen during a subsequent processing step. This can occur during a high temperature step, when the already available hydrogen in the structure can diffuse, or during a hydrogen-containing step. These processing steps are the topic of this chapter.

In section 4.2 the general flow of processing steps during the manufacturing of a CMOS electronic circuit is described. In section 4.3 the processing steps which might be important for the final hydrogen and/or deuterium concentration in the device are highlighted, followed by a description of hydrogen diffusion through the several layers in a CMOS circuit in section 4.4. After this general description, the process flows of the specific devices used for the experiments described in chapters 5 and 6 are explored. For these experiments, two sets of devices have been used, fabricated at two different places. In section 4.5 the process flow of the devices fabricated at the University of Twente is treated followed by the process flow of the devices fabricated by Philips Research Leuven in section 4.6.

4.2 General CMOS manufacturing

This section briefly describes a modern general CMOS process flow, that is the sequence of processing steps needed to manufacture a CMOS circuit. Detailed descriptions can be found in a number of textbooks, for example [16].

A general process flow can be divided in a few parts. Cross-sectional views of the different processing stages are depicted in figure 4.1.

The first part is the active area formation. The active area is the area on the silicon substrate where the actual devices are placed. The various devices are laterally electrically isolated. In modern CMOS manufacturing shallow trench isolation (STI) is used for this aim. STI is fabricated by first growing a silicon oxide on the silicon substrate, followed by the deposition of a silicon nitride layer. The silicon oxide layer relieves stress induced by the silicon nitride layer, which is used as a polishing stop layer. A trench is etched through the silicon nitride and silicon oxide into the silicon. The sidewall and bottom of the silicon trench are thermally oxidised. This is to provide a proper Si-SiO_2 interface. The trench is further filled with deposited silicon oxide. The excess deposited silicon oxide is polished away until the silicon nitride layer is reached. Finally, the remaining silicon nitride is etched away.

The second part is the well formation. The N-well and P-well are formed with implantation. To avoid channelling a sacrificial silicon oxide layer is grown first, followed by implantation of the wells. Finally, an anneal restores the crystalline structure and activates the impurities.

The third part is the gate formation. First the gate dielectric is grown, followed by deposition of poly-crystalline silicon. Optionally, this poly-crystalline silicon can be pre-doped and annealed for a higher final gate doping density, resulting in a lower resistivity and less gate depletion. After locally etching the poly-crystalline silicon to define the gates, a thin silicon oxide layer is grown on the top and sidewalls of the poly-crystalline silicon gates.

After the gate formation, the source and drain regions are formed. First the so-called extensions are implanted and annealed. These extensions are lightly doped tips of the source and drain, which are close to the gate. The relatively low doping density reduces the electric fields induced during device operation at the junction, thus reducing hot carrier degradation effects. A silicon nitride layer is deposited

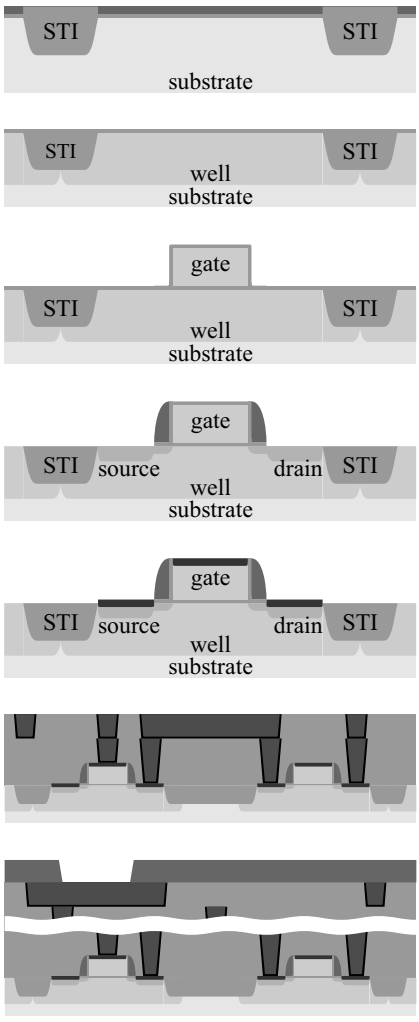


Figure 4.1: Intermediate stages in a general modern CMOS process flow.

and etched with an anisotropic process. The result is a silicon nitride spacer on the poly-crystalline silicon sidewalls. This spacer protects the lightly doped extensions. What follows is a high dose implantation to form both the heavily doped regions of the source and drain and to dope the poly-crystalline silicon gate. The final step is an implantation anneal.

The resistivity of even highly doped silicon is relatively high. The same applies to the contact resistance of a direct contact between silicon and metal. To reduce the resistivity and the contact resistance, modern CMOS processes use a metal silicide. Optionally, silicidation can be done selectively, to exclude certain areas like poly-crystalline silicon resistors and protection devices for electrical discharge, where the high resistivity is an advantage. For this option, a silicon nitride layer is deposited and etched selectively. This is followed by the removal of the silicon oxide covering the source, drain and gate regions. Titanium is deposited and annealed to form titanium silicide with the underlying silicon in the source, drain and gate regions. At the top, titanium reacts with the nitrogen in the annealing ambient to form titanium nitride. This titanium nitride is removed and a second anneal decreases the resistivity of the titanium silicide.

The devices are now completed and what follows is the metal interconnection. Here, the single damascene process will be described. An isolation material, like silicon dioxide, is deposited followed by the deposition of silicon nitride, which is used as a polishing stop layer. Contact holes are etched through the silicon nitride and isolation material. First a barrier and adhesion layer of titanium nitride is deposited and then the contact holes are completely filled with tungsten. The excess tungsten is polished away, followed by the removal of the silicon nitride layer. A second layer of isolation material is deposited, again followed by silicon nitride and the gaps where the metal lines will be are etched. Aluminum or copper (with a barrier layer) are deposited, and the excess metal is polished away, followed by the removal of the silicon nitride layer. This sequence of processing steps is repeated for every level of metal interconnect. Due to the amount of devices available in one circuit, modern processes need up to nine metal interconnect levels to be able to connect all devices properly [1].

Finally, a post-metal anneal is performed to improve the conductivity of the metal lines, to reduce the contact resistance between the metal and the silicide and to reduce the number of interface states at the Si-SiO_2 interface of the transistor channel. This is followed by the deposition of a silicon nitride layer, serving as a scratch protection layer. It also serves as a barrier layer for unwanted species from the outside ambient. After etching contact windows in the silicon nitride layer, the bonding pads can be contacted from the outside world and the circuit is complete.

The processing sequence described above is briefly repeated in

table 4.1. This table shows the intermediate processing stages and lists the processing steps important for the final hydrogen and/or deuterium content in the devices.

Table 4.1: Sequence of processing steps and intermediate stages in a general modern CMOS process flow. Only processing steps that might be important for the final hydrogen and/or deuterium content in the devices are given. The ambient and temperature given are used for the production of the devices at Philips Research Leuven as described in section 4.6 and can differ for other specific processes.

Processing stage	Processing step	ambient	temp in °C
Active area formation:	Silicon oxide	O_2	900
	Silicon nitride	$\mathbf{SiH_4/NH_3}$	750
	Trench oxidation	O_2	1050
	Trench filling	$\mathbf{Si(OC_2H_5)_4}$	700
Well formation:	Sacrificial oxide	O_2	1050
	Well Anneal	N_2	850
Gate formation:	Gate oxide growth	O_2	850
	Gate deposition	$\mathbf{SiH_4}$	600
	Gate pre-doping anneal	N_2	950
	Gate re-oxidation	O_2	750
Source/drain formation:	Extension anneal	N_2	1000
	Spacer deposition	$\mathbf{SiH_4/NH_3}$	750
	S/D anneal	N_2	1100
Silicide formation:	Silicide protection	$\mathbf{SiH_4/NH_3}$	750
	Silicidation Anneal I	N_2	900
	Silicidation Anneal II	Ar	850
Interconnect formation:	Isolation deposition	$\mathbf{Si(OC_2H_5)_4}$	350
	silicon nitride	$\mathbf{SiH_4/NH_3}$	350
	Isolation deposition	$\mathbf{Si(OC_2H_5)_4}$	350
	silicon nitride	$\mathbf{SiH_4/NH_3}$	350
Final processing:	Sintering anneal	$\mathbf{H_2/N_2}$	450
	Scratch protection layer	$\mathbf{SiH_4/NH_3}$	350

4.3 Hydrogen-containing processing steps

As listed in table 4.1, there are a number of high-temperature steps involved in a modern CMOS process flow and a number of them intro-

duce a substantial amount of hydrogen. These hydrogen-containing steps are, the deposition of silicon oxide layers with either silane (SiH_4) and oxygen or tetraethylorthosilicate ($Si(OC_2H_5)_4$ abbreviated to TEOS), the deposition of silicon nitride with silane (SiH_4) and ammonia (NH_3), the deposition of poly-crystalline silicon with silane (SiH_4) and the post-metal anneal in a mixture of hydrogen and nitrogen.

The first occurrence of silicon oxide deposition is the trench filling for the shallow trench isolation. This step occurs in the beginning of the process flow and does as such not directly influence the hydrogen and deuterium content in the still to grow gate dielectric. The resulting silicon oxide layer however, will contain a substantial hydrogen concentration and can serve as a reservoir for future diffusion. The other occurrences of silicon oxide deposition are the repeated deposition of isolation layers during the interconnect formation. These can directly influence the hydrogen and deuterium content in the gate dielectric.

The single occurrence of poly-crystalline silicon deposition is directly after the growth of the gate dielectric. The hydrogen and deuterium concentration in the gate dielectric are therefore greatly influenced by the hydrogen released during the silicon forming reaction from silane. Per silicon atom deposited, 2 hydrogen molecules are formed.

The deposition of silicon nitride occurs several times during the process. The first time, during the active area formation, is in the beginning of the process, where it cannot directly introduce hydrogen into the gate dielectric. Furthermore, it is removed later on, and does not serve as a possible diffusion reservoir. This last remark also applies to the silicon nitride layers deposited during the interconnect formation, but they can directly introduce hydrogen into the gate dielectric. The most critical use of silicon nitride is the spacer. This spacer is formed soon after the gate dielectric is grown and is situated in the gate dielectrics near vicinity. It can both directly introduce hydrogen into the gate dielectric and serve as a hydrogen source during subsequent high temperature steps.

The post-metal anneal occurs at the end of the process and intentionally introduces large amounts of hydrogen in the devices. The post-metal anneal is performed in hydrogen gas, typically at 400 to 450 °C for 30 min. To avoid explosion danger, the hydrogen is diluted with nitrogen to 5 to 10 %. This process step is intended to improve the conductivity of the metal lines, reduce the contact resis-

tance between the metal and the silicide and to reduce the number of interface states at the interface between the silicon substrate and the overlying dielectrics. It is not only important that the interface states underneath the gate dielectric are reduced, but also interface states underneath the shallow trench isolation have to be reduced to avoid parasitic transistors from turning on.

4.4 Hydrogen diffusion

The hydrogen introduced during the process has, depending on the stage of processing, to diffuse through one or more layers of material before it reaches the gate dielectric, where it can replace earlier introduced deuterium. As an example, the last processing step where hydrogen (or deuterium) is introduced, the hydrogen diffusion during the post-metal anneal will be discussed.

On top of the gate dielectric there is a complicated stack of materials: poly-crystalline silicon, silicide, silicon nitride, isolation material and metal stack and the hydrogen from the post-metal anneal has to diffuse through all these layers to reach the gate dielectric, as is depicted in figure 4.2.

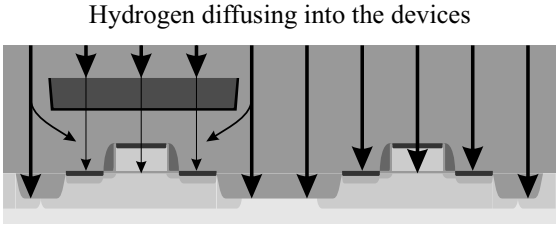


Figure 4.2: Diffusion of hydrogen introduced during the post-metal anneal through the stack of material layers.

The isolation material, deposited silicon oxide, is highly permeable for hydrogen (and deuterium) during processing conditions. Hydrogen quickly diffuses through this layer. The metal stack used for the interconnect, is not so permeable to hydrogen. For instance, Tuinhout *et al.* [26] noticed an unusually large mismatch between matched pair transistors, when one of the transistors was completely covered by a metal layer, composed of a stack of Ti-TiN-AlCu-TiN. The mismatch is due to incomplete passivation of the interface states in the metal covered transistors. This indicates that the metal stack is obstruct-

ing the diffusion of hydrogen towards the interface between the gate dielectric and the substrate. This is confirmed by secondary ion mass spectroscopy (SIMS) by Yamaha *et al.* [27]. They show that Ti, TiN and AlCuSi films block hydrogen diffusion. Also silicon nitride is reported to be an efficient barrier for hydrogen diffusion [28].

The fact that several layers act as diffusion barriers to hydrogen, indicates that hydrogen cannot directly diffuse towards the gate dielectric. The position of the metal layers can be random compared to the position of the transistors. This means that, as indicated above, one transistor gate dielectric can receive more hydrogen than another and the impact of the interconnect is uncertain. The silicon nitride is positioned on the sidewalls of the gate as spacer material and has a fixed position compared to the gate dielectric.

4.5 Devices fabricated at the university

A first set of devices was produced at the University of Twente. The process used to fabricate these devices greatly differs from the general process described in section 4.2. This is due to practical reasons and the desire for a quick process. As a trade-off only MOS capacitors can be fabricated with the used process.

The process is as follows. A 300 nm silicon oxide layer is thermally grown on the silicon substrate, using H_2O at 1050 °C. This forms the field oxide isolating the individual devices. Active area is defined by selectively etching the field oxide, using a buffered HF solution. Next, the gate dielectric is grown, directly followed by the deposition of 300 nm poly-crystalline silicon at 610 °C. Before implantation, a 25 nm silicon oxide layer is grown on top of the poly-crystalline silicon. The poly-crystalline silicon is doped using implantation of arsenic with an implantation energy of 100 keV and an implantation dose of $3 \times 10^{15}\,\mathrm{cm}^{-2}$. After implantation 10 nm silicon oxide is removed using a 1 % HF solution. The implantation damage is annealed in N_2 at 900 °C for 30 min. The gates are defined by selectively etching the poly-crystalline silicon using TMAH (tetramethylammonium-hydroxide) at 60 °C. For good electrical contact, aluminum contacts are defined on top of the gates and at the back of the silicon substrate using sputtering and wet chemical etching. Finally, the devices receive a post-metal anneal for 30 min at 450 °C. This results in MOS capacitors as drawn in figure 4.3.

The deuterium is introduced during both the growth of the gate

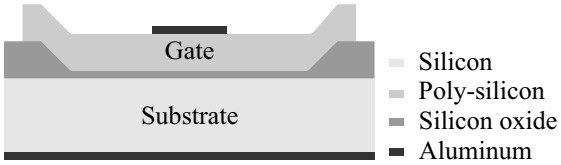


Figure 4.3: Cross sectional drawing of the MOS capacitors fabricated at the university.

dielectric and the post-metal anneal. The gate dielectric is grown following the same recipe as used for the study of the oxidation kinetics described in chapter 2. Either H_2O or D_2O diluted in N_2 was used with a partial pressure of 0.7 kPa at an oxidation temperature of 950 °C. The oxide thickness was varied by changing the oxidation time. The resulting thickness as measured by elllipsometer is given in table 4.2. The post-metal anneal was performed in a mixture of H_2O and N_2, when the gate dielectric was grown using H_2O and in a mixture of D_2O and N_2 when the gate dielectric was grown in D_2O.

Table 4.2: Oxide thickness (in nm) measured by ellipsometer for the devices fabricated at the University of Twente. The left column gives the thickness for devices with both the gate dielectric grown and a post-metal anneal performed in H_2O, and the right column gives the thickness for devices with both the gate dielectric grown and a post-metal anneal performed in D_2O.

time (min)		thickness (nm)	
H_2O	D_2O	H_2O	D_2O
31	41	5.24	5.34
43	56	6.50	6.72
57	74	7.95	8.10

Prior to the gate dielectric growth, the field oxide was grown, using H_2O. The resulting silicon oxide layer is rich with hydrogen. This hydrogen can diffuse and replace the introduced deuterium in subsequent high temperature steps. Similarly, after the gate dielectric growth, poly-crystalline silicon is deposited using SiH_4. During this deposition, an abundance of hydrogen is formed, which can also replace the previously introduced deuterium. The following post-implantation anneal can aid further in this. Only the deuterium introduced during the post-metal anneal cannot be replaced, since this is the last processing step.

4.6 Devices fabricated by Philips

A second set of devices was fabricated by Philips Research Leuven (batch PLI8509, MINOXG mask set). The process used to produce these devices follows the general process description of section 4.2. The only exception is that the process is finished after the first metal layer. The set of devices contains MOS transistors, matched pair transistors, MOS capacitors with and without a drain edge as well as other test devices. The gate oxide is 7 nm for all devices.

Three processing steps are varied for the different devices: the gate dielectric growth, the post-oxidation anneal and the post-metal anneal. The gate dielectric is either grown in O_2 at 900 °C or in a mixture of H_2 and O_2 or D_2 and O_2 at 850 °C. The post-oxidation anneal is either omitted or performed in either a N_2 or D_2 ambient at 900 °C for 30 min. Finally, the post-metal anneal was performed in either a mixture of H_2 and N_2 or D_2 and N_2 at 450 °C for 30 min. The combination of processing variants as used in this thesis are listed in table 4.3. The remarks made in section 4.3 regarding the introduction of deuterium and the replacement with hydrogen during subsequent processing steps fully apply to this set of devices.

Table 4.3: Combination of processing variants for the devices fabricated at Philips Research Leuven used in this thesis. The ambient during the gate oxide growth, the post-oxidation anneal and the post-metal anneal are varied.

Processing variant		1	2	3	4	5	6	7	8	9	10	11	12
Gate oxidation	O_2	+	+	+	+								
	H_2/O_2					+	+	+	+				
	D_2/O_2									+	+	+	+
post-oxidation	No anneal	+	+			+	+			+	+		
anneal	N_2			+				+				+	
	D_2				+				+				+
post-metal	H_2/N_2	+				+				+			
anneal	D_2/N_2		+	+	+		+	+	+		+	+	+

Chapter 5

Bulk quality

5.1 Introduction

The bulk quality of a dielectric is determined by its ability to insulate. Silicon oxide is in this respect, when properly manufactured, a nearly perfect electrical insulator. Nearly perfect, because electrical conduction is possible due to a quantum mechanical phenomenon called tunnelling as already explained in section 3.3. This is an intrinsic property of dielectrics. After electrical stress, the bulk quality of a silicon oxide film will degrade and the film starts to conduct more current than the intrinsic tunnelling current. This is the topic of this chapter. More precisely, the impact of deuterium, incorporated in the gate dielectric of CMOS devices during manufacturing, on the bulk quality of this gate dielectric. It is triggered by conflicting reports in literature.

The concept of tunnelling in the absence of an electric field is already introduced in section 3.3. Section 5.2 extends this concept to the case of an applied electric field. Section 5.3 deals with the increase in leakage current due to electrical stress and lists the three theories postulated in literature to explain this degradation of the dielectric bulk quality. Some work is already done by others on the influence of deuterium on the dielectric bulk quality. This work is summarised and examined in section 5.4. After this section, the experimental setup and results of the measurements conducted on the subject of this chapter are described in sections 5.5 and 5.6. In section 5.7, these results are discussed and compared to literature. Finally, conclusions are drawn in section 5.8.

5.2 Field-assisted tunnelling

Section 3.3 introduced the concept of tunnelling. The description remained constrained to the tunnelling probability of electrons in the absence of an electric field. In gate dielectrics under operating conditions, there will be an electric field across the dielectric, the oxide electric field. This electric field assists the electrons with tunnelling. Two situations exists as depicted in figure 5.1.

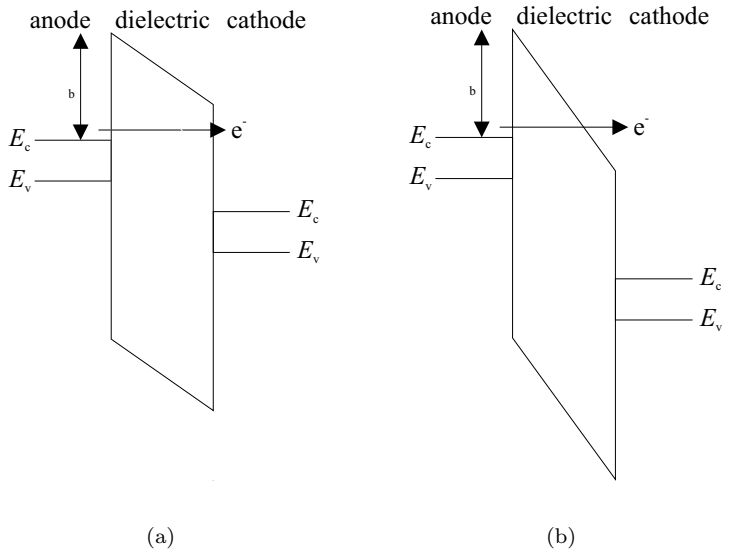


Figure 5.1: Two regimes for field-assisted tunnelling: (a) Direct tunnelling, the potential across the dielectric is smaller than the energy barrier height; and (b) Fowler-Nordheim tunnelling, the potential across the dielectric exceeds the energy barrier height.

In the first situation, depicted in figure 5.1(a), the applied oxide potential $q_e V_{ox}$ is smaller than the barrier height ϕ_b and electrons tunnel directly from the conduction band in the cathode to the conduction band in the anode. This is termed *direct tunnelling*. In the second situation, depicted in figure 5.1(b), the applied oxide potential exceeds the barrier height ϕ_b and electrons tunnel from the conduction band in the cathode first to the conduction band of the dielectric, before entering the conduction band of the anode. This is

termed *Fowler-Nordheim (FN) tunnelling.*

Fowler and Nordheim derived an analytical solution for the FN-tunnelling current [29]:

$$J_{\mathrm{FN}} = AE_{\mathrm{ox}}^2 e^{-\frac{B}{E_{\mathrm{ox}}}} \tag{5.1}$$

with

$$A = \frac{q_{\mathrm{e}}^3}{16\pi^2\hbar\phi_{\mathrm{b}}}\frac{m_{\mathrm{e}}}{m_{\mathrm{ox}}} \tag{5.2}$$

and

$$B = \frac{4\sqrt{2m_{\mathrm{ox}}}\phi_{\mathrm{b}}^{\frac{3}{2}}}{3q_{\mathrm{e}}\hbar} \tag{5.3}$$

The direct tunnelling current cannot be easily described in a closed analytical form as the FN-tunnelling current. Several approximate solutions exists of which one is [30]:

$$J_{\mathrm{DT}} = A'E_{\mathrm{ox}}^2 e^{-\frac{B'}{E_{\mathrm{ox}}}} \tag{5.4}$$

with

$$A' = \frac{A}{\left(1-\sqrt{1+\frac{q_{\mathrm{e}}V_{\mathrm{ox}}}{\phi_{\mathrm{b}}}}\right)} \tag{5.5}$$

and

$$B' = B\left(1-\left(1-\frac{q_{\mathrm{e}}V_{\mathrm{ox}}}{\phi_{\mathrm{b}}}\right)^{3/2}\right) \tag{5.6}$$

In these formulas, E_{ox} denotes the electric field across the dielectric, V_{ox} the electric potential across the dielectric, q_{e} the electron charge, m_{e} the rest electron mass, m_{ox} the effective electron mass in the dielectric, ϕ_{b} the energy barrier height and $\hbar$ the reduced Planck constant. The FN tunnelling current solely depends on the oxide electric field and is, when plotted as a function of the oxide electric field, independent of the dielectric thickness. The direct tunnelling current depends both on the oxide electric field and the oxide electric potential and is therefore highly dependent on the dielectric thickness when plotted as a function of the oxide electric field.

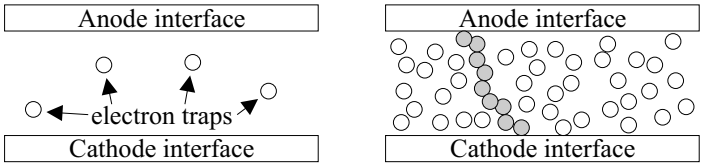
For a silicon oxide gate dielectric, general used values for the effective mass m_{ox} and the energy barrier height ϕ_b are $0.34m_e$ and 3.1 eV respectively. This indicates that for a silicon oxide gate dielectric the transition of the direct tunnelling regime to the FN-tunnelling regime occurs at an oxide electric potential of 3.1 V. During typical operation of a MOSFET, the oxide electric field is 6 MV/cm. This indicates that for devices with a silicon oxide gate dielectric thinner than 5 nm direct tunnelling becomes important. Actually, in state of the art devices, the gate dielectric thickness is well below 5 nm [1] and the direct tunnelling current is obstructing further reduction of the gate dielectric thickness. For floating gate devices, high electric fields, exceeding 10 MV/cm are used during program and erase. For the current tunnel dielectric thickness of around 8 nm this implies program and erase are done in the Fowler-Nordheim tunnelling regime.

5.3 The impact of FN-tunnelling stress

During device operation, the insulating properties of silicon oxide degrade due to electrical stressing of the film. In 1982, Maserjian and Zamani [31] noticed for the first time an increase in tunnelling current through a silicon oxide layer at low electric fields after applying a high electric field for some time. This excess current is known as stress-induced leakage current (SILC). SILC causes malfunctioning of floating gate devices, since a leakage current causes loss of data on the floating gate. When continuing the electrical stress, there will be a sudden jump in the oxide current and the insulating properties of the dielectric are completely lost. Dielectric breakdown has occurred.

Most researchers working on bulk silicon oxide degradation have adopted the idea that the increase in tunnelling current and the eventual breakdown is due to the creation of neutral electron traps within the silicon oxide. In the beginning of the electrical stress, just a few neutral traps will be created. These traps can function as stepping stones for the electrons tunnelling from cathode to anode. This greatly enhances the tunnel probability for electrons at low electric field and therefore the tunnelling current. When stressing continues, trap generation continues and after a critical trap density is reached, a conducting path will have formed between cathode and anode. Dielectric breakdown has occurred. Both situations are depicted in figure 5.2.

The formation of a breakdown path is a statistical process. Not



(a) (b)

Figure 5.2: Formation of neutral traps during electrical stress. (a) In the beginning, the traps acts as stepping stones for the tunnelling electrons, (b) but after a critical trap density is reached, a conducting path will have formed between cathode and anode.

all dielectrics will fail after the same stress time. Time- and charge-to-breakdown for dielectrics are described with a Weibull distribution [32]. This is a distribution widely used to describe weakest link statistics. The Weibull cumulative distribution function is given by:

$$F(t) = 1 - e^{-(t/\eta)^{\beta}} \tag{5.7}$$

In this equation η is the modal value of the distribution, *i.e.* η is the value of t where the the cumulative distribution reaches 63 %. This implies that in 63 % of the devices breakdown has occurred. The second parameter is the shape factor β better known as the Weibull slope. The modal value of the distribution for dielectric breakdown depends on the area of the dielectric layer. A large area has a higher probability to contain a weak spot than a small area. It can be shown that for two sets of further identical devices, one set with a dielectric area of A_1 and the other with a dielectric area of A_2 [33]:

$$\frac{\eta_1}{\eta_1} = \left(\frac{A_2}{A_1}\right)^{1/\beta} \tag{5.8}$$

The Weibull slope β does not depend on the dielectric area.

Although there is consensus on the fact that neutral trap generation is the cause of dielectric degradation, the mechanism of neutral trap generation in itself is still subject of debate. Three different models are postulated in literature.

McPherson and Baglee [34, 35] postulated the so-called *thermochemical model*. This model states that an applied electric field across the oxide gives rise to an ionic displacement. At an electric field

of 10 MV/cm the ionic displacement is 0.04 Å. This is 2 % of the normal bond distance of 1.7 Å for a Si–O bond. This is considered large, since SiO_2 will fracture at 7 % bond distortion and melt at 8 %. The strained bond is expected to interact with thermal phonons and occasionally gains enough thermal energy to cause bond breakage. A broken bond is a possible trap centre. This model predicts that the time-to-breakdown t_{bd} for a dielectric under electrical stress is described by:

$$\ln t_{\mathrm{bd}} \sim \frac{\Delta H}{kT} - \gamma E_{\mathrm{ox}} \tag{5.9}$$

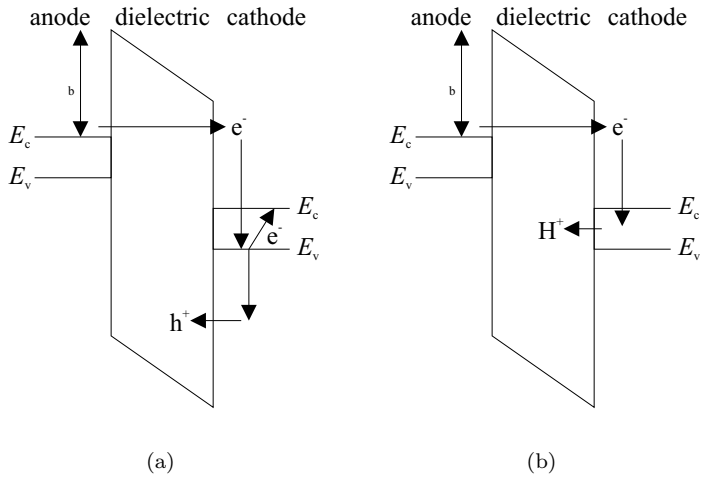
In this equation, ΔH is the activation enthalpy for bond breakage, k is the Boltzmann constant, T is the absolute temperature and γ is a field acceleration parameter.

Chen *et al.* [36, 37] noticed a similarity between generated hole tunnelling currents through thin silicon oxide layers and breakdown of the silicon oxide. Therefore they postulate that the hole current generates the neutral traps in the oxide. In the current version [38] of this so-called *Anode Hole Injection (AHI) model*, the measured hole current is generated by energetic electrons tunnelling through the silicon oxide. These electrons lose their energy as they reach the anode, creating energetic holes in the silicon valence band. A fraction of these holes tunnel back into the silicon oxide, forming the measured hole current. This mechanism is illustrated in figure 5.3(a). This model predicts that the total generated number of neutral electron traps is proportional to the total generated hole fluence which is proportional to the total electron tunnelling fluence. Therefore Chen *et al.* state that the charge-to-breakdown Q_{bd} is a more proper measure for dielectric breakdown than the time-to-breakdown t_{bd}. The model predicts that:

$$\ln Q_{\mathrm{bd}} \sim \frac{1}{E} \tag{5.10}$$

This is quite different from the prediction made by the thermochemical model. Extrapolation of t_{bd} or Q_{bd} measured at stressing conditions to operating conditions results in a large difference in predicted lifetime for devices for the two different models.

DiMaria *et al.* [39, 40] noticed that the degradation of silicon oxide subjected to a hydrogen plasma is similar to the degradation of silicon oxide subjected to electrical stress. Therefore, they postulated the *Hydrogen Release (HR) model*. This model is similar to the AHI



(a) (b)

Figure 5.3: Mechanism of neutral traps generation: Electrons tunnel from the cathode to the anode. In the anode they loose their excess energy, thus creating (a) hole-electron pairs according to the AHI-model or (b) releasing hydrogen from the interface according to the HR-model. The positive holes/hydrogen ions move into the dielectric due to the field and create neutral electron traps.

model, with the adjustment that not only holes are generated by the energy-loss of the tunnelling electrons at the anode side, but also hydrogen ions are released from passivated dangling bonds. According to the HR model, these hydrogen ions generate the neutral traps and not the holes.

From the three models mentioned above, only the HR model can readily explain a possible deuterium isotope effect on the bulk quality of silicon oxide. As will be shown in chapter 6 the creation of interface states is harder when dangling bonds are passivated with deuterium, than with hydrogen. This is true at least in the case of hot carrier degradation, where the breaking of the Si–H (or Si–D) bond is a multi-step event. The thermochemical model and the AHI model cannot readily explain a possible deuterium isotope effect on the bulk quality of silicon oxide, because the Si–H bond does not play a role in either of these models. In case of these two models an explanation has to be found in the trap itself and not in the mechanism of trap creation.

5.4 Work by others

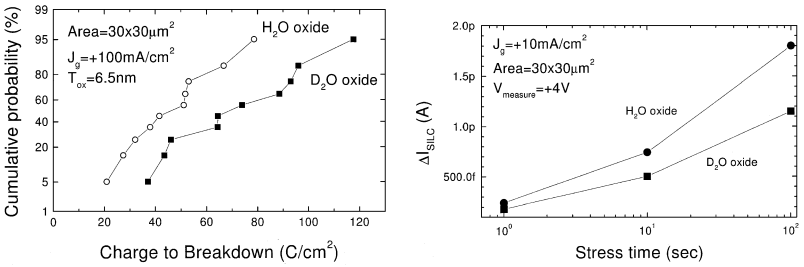
Six groups report about the influence of deuterium incorporation on the bulk quality of thin silicon oxide films. Some of these groups do report an isotope effect on the bulk quality, while others see no isotope effect. Several methods are applied to incorporate the deuterium during the device fabrication. Next to this, oxide thickness and electrical stressing conditions vary from group to group. A brief overview of the conditions and results of the groups is given in table 5.1. The apparent contradicting results are the prime motivation for the work presented in the remainder of this chapter. Therefore the results of these published findings are discussed in more detail in this section.

Table 5.1: Overview of available studies on the influence of deuterium incorporation in semiconductor processing on the bulk quality of thin silicon oxide films.

Group	Process step	T [°C]	x_o [nm]	Isotope effect
Hwang	Oxidation	835	6.5	Yes
Mitani	Oxidation	850	7.7	Yes
Liu	Pre- and	1000	1.6–2.5	Yes/No
	post-oxidation anneal	900		
Lin	Anneal at end of process	400–425	7.0	No
			2.7	Yes
Wu	Forming gas anneal	450	2.2–3.7	No
Esseni	Forming gas anneal	425–450	5.0	No
This work	Oxidation	850–950	6.0–9.0	No
	Post-oxidation anneal	850	7.0	No
	Post-metal anneal	450	7.0	No

The first group to report on the effect of deuterium incorporation on the bulk quality of thin silicon oxide films is the group of Hwang [4, 41, 42]. This group grows the gate oxide of MOS capacitors in either H_2O or D_2O steam ambient at 835 °C. The resulting oxide thickness is 6.5 nm. Apparently, the process is finished immediately after polycrystalline silicon deposition and gate definition. Measured charge-to-breakdown statistics for a stressing current of 100 mA/cm^2, reveals an isotope effect. The substitution of hydrogen with deuterium improves the Q_{bd} with 50 %, as can be seen in figure 5.4(a). The deuterated dielectrics are also reported to show less SILC after comparable stress as shown figure 5.4(b). The SILC was measured at a sense voltage of 4 V after applying a stress current of 10 mA/cm^2. Finally, Hwang and coworkers extend the research to other gate insulator materials

and shows a deuterium isotope effect on the bulk quality of oxynitride [43, 44], ZrO_2 [45], Ta_2O_5 [46] and HfO_2 [47].



(a) (b)

Figure 5.4: (a) Charge-to-breakdown and (b) stress-induced leakage current improvement as measured by Kim *et al.* [4].

Mitani *et al.* [11,48,49] use pyrogenic oxidation at 850 °C in either a H_2/O_2 or D_2/O_2 ambient to create a 7.7 nm gate oxide. Their poly-silicon gates are deposited using either SiH_4 or SiD_4. The final MOS capacitors and MOSFET's were annealed in either H_2/N_2 or D_2/N_2 at 450 °C. The process for the MOS capacitors ended with this anneal which was performed directly after poly-crystalline silicon gate deposition, implantation and patterning. It was found that SILC is reduced for both stressing polarities and sensing polarities in case of deuterated pyrogenic oxidation in combination with deuterated poly-silicon deposition. If conventional poly-silicon deposition was used, the SILC was not reduced for gate injection conditions. The deuterium anneal at the end of the process did not influence the SILC.

The last group that incorporated deuterium in an early stage of processing at high temperature, is the group of Liu [50–52]. Liu and coworkers anneal the silicon in H_2 or D_2 at 1000 °C for 2 min, followed by the rapid thermal growth of the dry gate oxide. This gate oxide is annealed in H_2, D_2 or N_2 for 10 min at 900 °C. Processing was finished by deposition and etching of aluminum gates. The gate oxide thickness ranged from 1.6 to 2.5 nm. The NMOS devices were stressed at different stress current levels, while the gate voltage was monitored. The results showed that above a stress voltage of 4 V, no isotope effect occurred, while below this stress voltage an distinct isotope effect was

seen. A similar experiment on PMOS devices reveals no isotope effect at all [52].

Wu *et al.* [53] incorporate the deuterium at the end of the process. MOSFETs with a 2.2 nm and 3.7 nm gate oxide were fabricated. The devices with a gate oxide thickness of 3.7 nm received a forming gas anneal, while the devices of 2.2 nm gate oxide did not. The process ended with an anneal in 100 % H_2 or D_2 at 450 °C at 1 atm or 5 atm. The results show no isotope effect on Q_{bd} or SILC for both PMOS and NMOS devices.

The results of Wu *et al.* are confirmed by the work of Esseni *et al.* [54, 55]. In this work MOSFETs are annealed at 425 °C for 2 h in 10 % H_2 or at 450 °C for 5 h in 10 % D_2. The fabrication of these devices included silicon oxide spacers and was ended after the first metal layer. The devices were subjected to hot carrier stress and the gate current was measured using a floating gate technique. The SILC measured in this way shows no isotope effect.

Finally, Lin *et al.* [56] manufacture MOS capacitors. At the end of the process, *i.e.* after poly-crystalline silicon gate deposition, doping and patterning, they incorporate deuterium with an anneal at 400 to 425 °C for 30 min in pure H_2 or D_2. They measure the SILC for devices with a gate oxide of 7.0 nm and 2.7 nm. Their results indicate no isotope effect for the 7 nm gate oxide, but a distinct isotope effect for the 2.7 nm oxide.

Concluding, it can be stated that there is no consensus on the influence of deuterium incorporation on the bulk oxide quality. Reviewing the results published in literature, no completely consistent theory can be formulated, but the idea arises that an early incorporation of deuterium at a high temperature (≥ 850 °C) is beneficial for the bulk quality. The work of Liu and coworkers only partly agrees to this conclusion. Incorporation at a late stage in processing, at a relatively low temperature (≤ 450 °C) does not improve the bulk quality of thin silicon oxide layers. In this case the work of Lin *et al.* contradicts this conclusion partly.

5.5 Experimental setup

The stress-induced leakage current of thin silicon oxide films with or without deuterium incorporated in the fabrication process was investigated on the Philips set of wafers. The devices in this set have a gate oxide grown in either an O_2, H_2/O_2 or an D_2/O_2 ambient.

A post-oxidation anneal is either omitted, or performed in a N_2 or D_2 ambient and finally the post-metal anneal is performed in either a H_2/N_2 or D_2/N_2 ambient. The full process flow is described in section 4.6.

The stress-induced leakage current was investigated by stressing MOS capacitors with a drain edge with a constant stress current of $1\,\mathrm{mA/cm^2}$. The stress polarity corresponds to electrons tunnelling from the silicon substrate to the poly-silicon gate. Periodically, the stress was interrupted to measure the current-voltage characteristics of the capacitor. The area of the measured devices is $900\times900\,\mu\mathrm{m}^2$.

Before and after the cycle of IV-measurements and stress, quasi-static capacitance-voltage characteristics were measured. These CV-characteristics were used to calculate the oxide electric field. By definition, the measured differential capacitance equals:

$$C = \frac{dQ}{dV} \tag{5.11}$$

At the same time, it follows from the boundary conditions at a conductor surface that:

$$E_{\mathrm{ox}} = \frac{Q}{\varepsilon} \tag{5.12}$$

where ε is the permittivity of the dielectric. Combining equation 5.11 and 5.12 and integrating from flatband voltage, that is zero oxide electric field, to the desired gate voltage gives the corresponding oxide electric field:

$$E_{\mathrm{ox}}(V_{\mathrm{g}}) = \frac{1}{\varepsilon}\int_{V_{\mathrm{fb}}}^{V_{\mathrm{g}}} C dV \tag{5.13}$$

Two components are critical when using this method to estimate the oxide electric field. Firstly, to start the integration at the correct boundary, the flatband voltage has to be known accurately. This is achieved by the so-called Mott-Schottky method [57]. Secondly, the gate voltage area of interest where the oxide electric field has to be known, lies outside the measurement range of the CV-measurement. This implies that the capacitance needed for the integration in this area is unknown and has to be estimated. This is done by assuming that the capacitance outside the measurement range equals the measured capacitance at the boundary of the CV-measurement, that is this capacitance is assumed to be the oxide capacitance. In practice

this leads to an underestimate of the capacitance and therefore the oxide electric field.

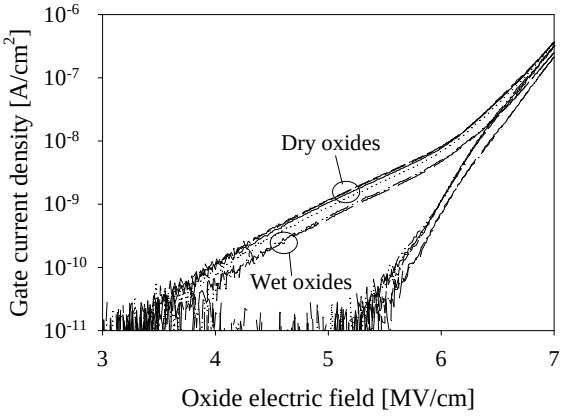
The measurements were performed using an automated measurement setup consisting of a Cascade probe station, an HP E5250A switching matrix and an Agilent 4156C semiconductor analyser. The equipment was operated using a dedicated Metric measurement program.

The breakdown statistics of thin silicon oxide films with or without deuterium incorporated in the fabrication process were investigated on both the university and Philips set of wafers. The university set has a gate oxide thermally grown in either H_2O or D_2O ambient at 950 °C and at a partial pressure of 0.7 kPa. The post-metal anneal is performed in a H_2O ambient for the H_2O grown oxides and in a D_2O ambient for the D_2O grown oxides. The full process flow is described in section 4.5. The Philips set has a gate oxide grown in either an O_2, H_2/O_2 or an D_2/O_2 ambient. A post-oxidation anneal is either omitted, or performed in a N_2 or D_2 ambient and finally the post-metal anneal is performed in either a H_2/N_2 or D_2/N_2 ambient. The full process flow is described in section 4.6.

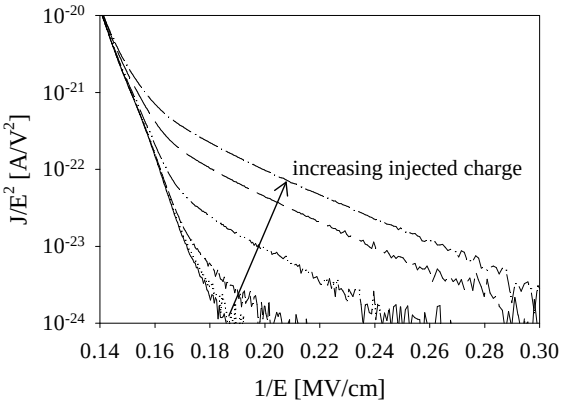
Breakdown statistics were investigated by measuring the charge-to-breakdown under constant current stressing with a stress current of $-100\,\mathrm{mA/cm^2}$. This corresponds to electrons tunnelling from the poly-silicon gate to the silicon substrate. For the university set 121 devices were stressed per wafer, while for the Philips set 20 devices were stressed per wafer. The used devices are MOS capacitors with an area of $10\times10\,\mu\mathrm{m}^2$ for the university set and $50\times200\,\mu\mathrm{m}^2$ for the Philips set.

5.6 Results

Figure 5.5(a) shows the gate current density as a function of oxide electric field for devices both before and after a stress of $1\,\mathrm{C/cm^2}$. The shown curves are the average for 5 devices per wafer. Figure 5.5(b) shows current-voltage measurements in a Fowler-Nordheim representation for one of the devices. The linear behaviour in the plot is typical for all the devices measured and shows that the fresh current indeed is a Fowler-Nordheim current. Next to this, it shows the trend with increasing injected charge. As can be seen in figure 5.5(a), the current density of the fresh devices with a wet oxide, i.e. the oxides grown in H_2/O_2 or D_2/O_2 ambient, is slightly lower than the current



(a)



(b)

Figure 5.5: (a) Gate current density versus oxide electric field before and after stress. (b) Fowler-Nordheim plot for one of the devices shown in (a).

density of the fresh devices with a dry oxide, *i.e.* the oxides grown in O_2. This slight difference was not further studied. The figure does not show a difference between the current densities of the H_2/O_2 and D_2/O_2 grown oxides after stress. At the same time, all four curves

for the dry oxide after stress are almost overlapping, indicating that the post-oxidation and post-metal anneal do not significantly influence the SILC. The last two observations are valid for a broad range of injected charge, as can be seen in figure 5.6. Regardless of the processing step introduced, deuterium does not affect the SILC characteristics.

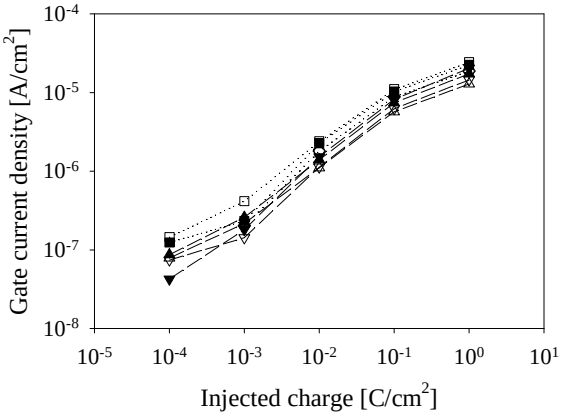


Figure 5.6: SILC versus injected charge for $900\times900\,\mu m^2$ MOS capacitors in the Philips set of wafers. The stress current was maintained at $1\,mA/cm^2$. The gate oxide is grown in either O_2 at 900 °C or in either H_2/O_2 or D_2/O_2 ambient at 850 °C. The post-oxidation anneal was either omitted, or performed in either N_2 or D_2 ambient at 900 °C for 30 min. The post-metal anneal was performed in either H_2/N_2 or D_2/N_2 ambient. The dotted lines indicate the dry oxides and the dashed lines the wet oxides. The symbols are the same as used in figure 5.7.

Part of the observations regarding SILC and the moment of deuterium introduction in the process is confirmed by the measured Q_{bd} characteristics shown in figure 5.7. The charge-to-breakdown does not depend on the post-oxidation anneal or the post-metal anneal. Even more, the post-oxidation anneal does not have an effect at all[1]. But in contrast to the stress-induced leakage current, the Q_{bd} does depend on the oxidation ambient. As can be seen in figure 5.7(a), oxidation in a O_2 ambient results in a lower Q_{bd}. There is no observable difference between oxidation in a H_2/O_2 or D_2/O_2 ambient.

[1] In general this is a remarkable result since a post-oxidation anneal is performed to improve the bulk quality.

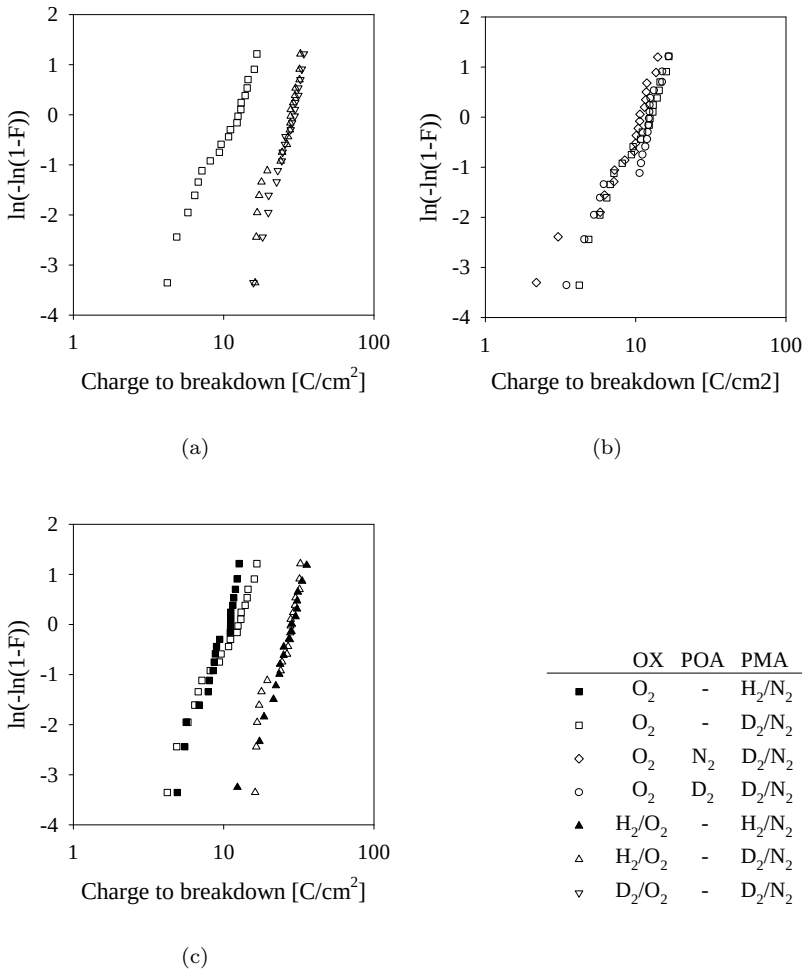


Figure 5.7: Charge-to-breakdown characteristics for $50\times200\,\mu m^2$ MOS capacitors in the Philips set of wafers. The stress current was maintained at $-100\,mA/cm^2$. The gate oxide is grown in either O_2 at 900 °C or in either H_2/O_2 or D_2/O_2 ambient at 850 °C. The post-oxidation anneal was either omitted, or performed in either N_2 or D_2 ambient at 900 °C for 30 min. The post-metal anneal was performed in either H_2/N_2 or D_2/N_2 ambient; (a) depicts the influence of oxidation ambient on charge to breakdown; (b) depicts the influence of post-oxidation anneal on charge to breakdown; (c) depicts the influence of post-metal anneal on charge to breakdown.

For a complete overview of the obtained results, the Weibull slope and modal Q_{bd} value were extracted from the measured data. These values are calculated using a maximum likelihood estimation. The extracted values are depicted in figure 5.8. Only 20 devices were tested per wafer, so one can expect a large uncertainty in the extracted values [58]. As can be seen in the figure, this is certainly true for the Weibull slope. The spread in the modal Q_{bd} value per oxidation is restricted to 15 %, but the trends observed in figure 5.7 are also clearly visible in figure 5.8, *i.e.* :

1. a distinct difference between the dry oxide and the wet oxides, but no difference between the wet oxides;
2. no difference for different post-oxidation annealing ambients, even more, the anneal does not have an effect on bulk properties at all;
3. no difference between the forming gas anneals.

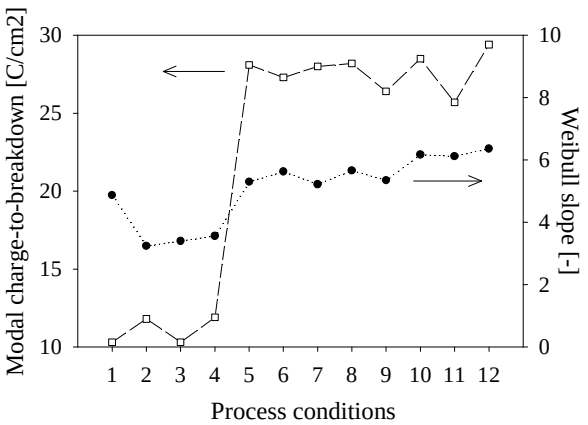
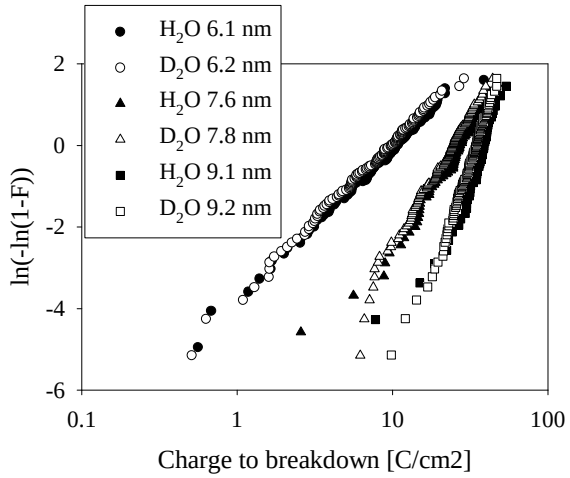
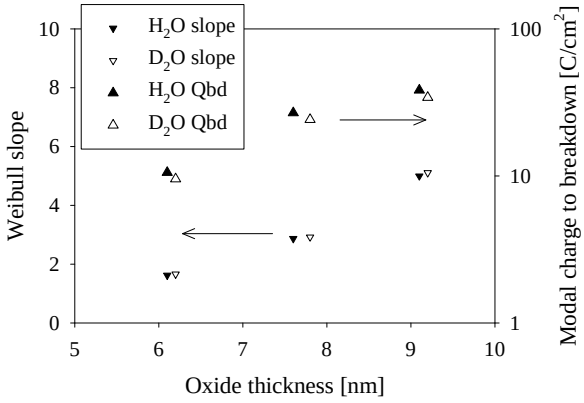


Figure 5.8: Weibull slope (β) and modal Q_{bd} (α) for $50\times200\,\mu\mathrm{m}^2$ MOS capacitors in the industrial set of wafers. The stress current was maintained at $-100\,\mathrm{mA/cm^2}$ The gate oxide is grown in either O_2 at 900 °C or in either H_2/O_2 or D_2/O_2 ambient at 850 °C. The post-oxidation anneal was either omitted, or performed in either N_2 or D_2 ambient at 900 °C for 30 min. The post-metal anneal was performed in either H_2/N_2 or D_2/N_2 ambient. The numbering of the processing conditions is the same as used in table 4.3.



(a)



(b)

Figure 5.9: Charge to breakdown characteristics for $10\times10\,\mu m^2$ MOS capacitors with varying oxide thickness in the university set of wafers. The stress current was $-100\,mA/cm^2$. The gate oxide is grown in either ultra-diluted H_2O or D_2O ambient at $950\,^\circ C$; (a) depicts the cumulative distribution and (b) depicts the Weibull parameters extracted from the data in (a) as a function of oxide thickness.

Finally the charge-to-breakdown characteristics were measured on the MOS capacitors on the university set of wafers. The results, shown in figure 5.9(a), indicate that the Q_{bd} characteristics of the H_2O and D_2O devices are similar.

A small precaution has to be taken though, since the oxide thickness for the two different precursors do not exactly match, as indicated in the legend of figure 5.9(a). To avoid this problem of thickness mismatch, the Weibull slope and modal (or 63 %) Q_{bd} value are plotted as a function of oxide thickness in figure 5.9(b). This graph confirms the observation of figure 5.9(a), *i.e.* there is no distinguishable difference between the H_2O and D_2O devices. The graphs for the Weibull slope overlap, while the difference in the modal Q_{bd} value is small, in favour of the H_2O devices.

5.7 Discussion

After examining the work of others in section 5.4, the idea arose that an early incorporation of deuterium at a high temperature ($\geq 850\,^{\circ}\mathrm{C}$) would be beneficial for the silicon oxide bulk quality. The results presented in the former section, prove this idea wrong. The results clearly indicate that the incorporation of deuterium as done in this work does not improve the silicon oxide bulk quality. The Q_{bd} and SILC for the set of devices with a gate oxide grown in H_2/O_2 or D_2/O_2 are similar. The same applies to the set of devices with a gate oxide grown in H_2O or D_2O. The sets originate from two different laboratories and while the oxidation temperature for the first set was $850\,^{\circ}\mathrm{C}$, the oxidation temperature for the second set was $950\,^{\circ}\mathrm{C}$. This strongly suggests that the obtained results do not depend on the specific oxidation temperature or other processing conditions used.

This is in contradiction with the results obtained by both Mitani *et al.* [11, 48, 49] and Hwang and coworkers [4, 41, 42]. Hwang and coworkers need 10 min oxidation time at an oxidation temperature of $835\,^{\circ}\mathrm{C}$ to grow 6.5 nm silicon oxide in a D_2O ambient. For comparison, the oxidation data obtained in section 2.4 indicate that for the ultra-diluted D_2O ambient used to grow silicon oxide in this work, ≈ 300 min oxidation time at an oxidation temperature of $850\,^{\circ}\mathrm{C}$ is necessary to achieve the same oxide thickness. The D_2O partial pressure must be substantially higher for the case of Hwang and coworkers. Next to this, they observe an almost 100 % difference in oxidation rate for a H_2O and D_2O ambient. This factor two difference in oxidation

rate will have more influence on the silicon oxide bulk quality than the factor 1.18 observed in section 2.4.

Mitani *et al.* use pyrogenic oxidation in a H_2/O_2 or D_2/O_2 ambient at 850 °C. They need ≈ 6 min to grow 6.5 nm silicon oxide in a D_2/O_2 ambient [11]. This is in the same order of magnitude as Hwang and coworkers. Surprisingly enough though, Mitani *et al.* notice only a 10 % difference in oxidation rate for H_2/O_2 and D_2/O_2 oxidation ambient. Both groups oxidise at a substantially higher rate than done in this work. This suggests that due to the low oxidation rate used in this work the silicon oxide bulk quality is already so high that incorporation of deuterium does not help anymore. This also explains why a post-oxidation anneal does not influence the silicon oxide bulk quality anymore.

The fact that the results of section 5.6 show that the introduction of deuterium during the post-metal anneal does not influence the silicon oxide bulk quality is in agreement with the work of Lin [56], Wu *et al.* [53] and Esseni [54,55]. The hot carrier degradation measurements presented in chapter 6 reveal that the deuterium incorporated during the post-metal anneal is able to suppress the hot carrier degradation. This indicates that the incorporated deuterium is at least present at the Si-SiO_2 interface. Combining these two results, one might jump to the conclusion that the hydrogen release model must be incorrect. However, the excess electron energy transferred to the Si–H or Si–D bond is different for the two degradation measurements. In case of hot carrier degradation, the electrons do not have enough energy to break the Si–H or Si–D bond at once, bond breakage occurs through multi-excitation. In case of the Fowler-Nordheim tunnelling stress used in this chapter, the electrons have enough energy to break the Si–H or Si–D at once. The benefit of using deuterium in this case vanishes. Therefore, the results obtained in this work are not able to exclude one of the possible neutral trap creation mechanisms mentioned in section 5.3.

Concluding, the work presented in this chapter indicates that the incorporation of deuterium during CMOS manufacturing does not improve the silicon oxide bulk quality of an already high quality silicon oxide layer. It will therefore not enable the reduction of the thickness of tunnel dielectrics in floating gate devices.

5.8 Conclusions

In this chapter, the impact of deuterium on the bulk quality of silicon oxide was investigated. Stress-induced leakage current and charge-to-breakdown measurements on two sets of devices fabricated in two different laboratories obtained the same result. Regardless the processing step deuterium is introduced or the used silicon oxide thickness or oxidation temperature, deuterium does not improve the bulk quality of silicon oxide as a gate dielectric.

Chapter 6

Interface quality

6.1 Introduction

The interface between the silicon substrate and the gate dielectric in a CMOS device is not perfect. The defects at this interface are in electrical contact with the channel and have a large impact on the performance of the device. Electrical stress, either intentional or during device operation, creates more defects at the interface, degrading the electrical properties of the CMOS device. This degradation is the topic of this chapter. More precisely, the impact of deuterium, incorporated in the gate dielectric of CMOS devices during manufacturing, on the interface quality of this gate dielectric.

First, the origin of the defects is described in section 6.2. This is followed by the impact of electrical stress on the interface quality in section 6.3. Work done by others on the impact of deuterium on the interface quality of silicon oxide is summarised in section 6.4. After this section, the experimental setup and results of the measurements conducted on the subject of this chapter are described in sections 6.5 and 6.6. In section 6.7, these results are discussed and compared to the work summarised in section 6.4 and conclusions are drawn.

6.2 Silicon oxide charges

No matter how well controlled, there will always be charge in a silicon oxide layer. Deal [59] recognises four important type of charges in silicon oxide, as depicted in figure 6.1. These charges are mobile oxide charge, fixed oxide charge, oxide trapped charge and interface

trapped charge. The mobile oxide charge is due to ionic impurities like Na^+ and K^+ in the silicon oxide layer. In the early days of CMOS manufacturing these ions were of high concern, but in modern manufacturing, they are well controlled and generally do not pose a problem. The fixed oxide charge is due to structural imperfections in the silicon oxide layer originating from the oxide growth and depends on the growing conditions. This charge can also be well controlled during manufacturing.

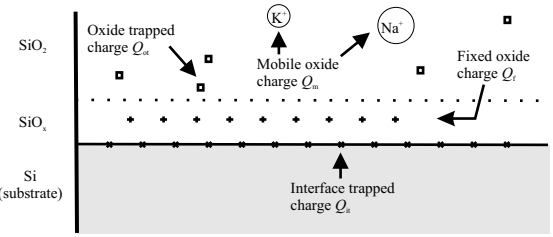


Figure 6.1: Schematic representation of the electrical charges in a silicon oxide layer.

The last two types of charges are more of interest. The oxide trapped charge is formed by holes and/or electrons which are trapped at defect sites in the silicon oxide layer. During device operation, some holes and/or electrons can surmount the energy barrier and enter the silicon oxide layer where they can create defect sites, or get trapped at existing defect sites. Under proper bias conditions, they can also be released from the defect site again.

Interface trapped charge is formed by electrons and/or holes, that are trapped at interface states. Interface states are energy states at the Si-SiO_2 interface, which lie in the forbidden energy gap between the top of the Si valence band and the bottom of the Si conduction band. The amount of electrons and holes trapped at interface states depends on the biasing conditions and changes during device operation. As the charge is situated at the interface, it communicates with the transistor channel.

The interface states arise from unsaturated bonds. The lattice constant of the silicon oxide gate dielectric does not match the lattice constant of the mono-crystalline silicon substrate. Even more, the silicon oxide is amorphous with only short range order. As a result not all the necessary bonds can be formed at the interface between

the silicon and the silicon oxide. This is depicted in figure 6.2.

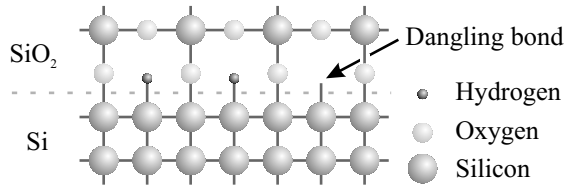


Figure 6.2: Simple representation of the lattice mismatch at the Si-SiO_2 interface, resulting in dangling bonds. Part of the dangling bonds are passivated during the CMOS manufacturing process with hydrogen.

Figure 6.2 shows that part of the unsaturated bonds, termed dangling bonds, can be bonded to hydrogen. This is common practice in CMOS manufacturing. As one of the last steps, a post-metal anneal is performed. This post-metal anneal is performed in a hydrogen containing ambient. During the anneal, hydrogen atoms from the annealing ambient diffuse towards the Si-SiO_2 interface, where they dissociate and bond to the dangling bonds, thus reducing the interface trapped charge.

As stated above, oxide and interface trapped charge can change during device operation. The oxide defect sites and the interface states can trap and release charge. The trapped charge comes from the transistor channel and the released charge returns to the transistor channel. This directly influences the current flowing through the transistor channel and therefore the device operation.

The macroscopic trapped charge is controlled by the device biasing conditions, but the trapping and release of charge is in principle a random process. It can be described by an average capture and emission time. This causes fluctuations in the trapped charge. This is thought to be the origin for the observed $1/f$ noise in the drain current of MOS devices.

6.3 Impact of hot carrier stress

During device operation, the interface quality of silicon oxide degrades due to electrical stressing of the film. In the transistor channel, charge carriers are accelerated by the electric field, gaining kinetic energy. A small number of charge carriers is able to gain a substantial amount of

kinetic energy before colliding with the silicon lattice. These charge carriers are termed *hot*.

The hot carriers give rise to a number of phenomena depending on their kinetic energy and the electric field distribution in the channel. For instance, the hot carrier can cross over the energy barrier at the Si-SiO_2 interface when the kinetic energy is large enough. After crossing the energy barrier, the hot carrier can be trapped in the silicon oxide and thus contribute to the oxide trapped charge, or release the excess kinetic energy by breaking Si–H bonds at the interface creating dangling bonds, thus interface states. This mechanism is termed *channel hot carrier (CHC)* degradation and mainly occurs when the gate voltage and drain voltage are similar.

A second mechanism occurs when the hot carrier releases the excess kinetic energy to the silicon by creating a hole-electron pair. Under influence of the electric field perpendicular to the transistor channel, either the created hole or electron might be able to cross the Si-SiO_2 energy barrier and create the same damage as described above. This mechanism is termed *drain avalanche hot carrier (DAHC)* injection and is favoured when the maximum substrate current is reached, that is when the gate voltage is roughly half the drain voltage.

The created oxide trapped charge and interface states result in a change of the transistor parameters. Depending on the polarity of the extra oxide charge, the threshold voltage will increase or decrease. Due to extra coulomb scattering, the transconductance will decrease.

6.4 Work by others

A number of groups [3, 60–65] have shown that the incorporation of deuterium during some stage of processing, improves the resistance to hot carrier degradation of gate dielectrics. In most cases, this concerns a gate dielectric of silicon oxide, but also silicon oxynitride [61] and hafnium oxide [65] have been shown to benefit. Most groups introduce the deuterium at the end of the process during the forming gas or post-metal anneal and report a substantial (10 to 90 times [66]) improvement in lifetime of MOS transistors subjected to hot carrier stress. This is confirmed on technologies ranging from 0.7 to 0.1 µm [3, 67] and from single metal processing [68] to 6 metal layers [64].

A few groups show that earlier incorporation of the deuterium also results in an isotope effect. Mitani *et al.* [11] introduce deuterium

during the growth of the gate dielectric. They show that MOSFETs with a gate dielectric grown in a D_2/O_2 ambient, when subjected to hot carrier stress, have a higher lifetime than MOSFETs with a gate dielectric grown in a H_2/O_2 ambient. Clark *et al.* [62] use a silicon nitride layer deposited using ND_3 and SiD_4 on top of MOS devices as both a reservoir for deuterium diffusion towards the gate dielectric and as a barrier for hydrogen (introduced in subsequent processing steps) diffusion towards the gate dielectric.

In all this work, the hot carrier degradation measurements were done for n-channel MOSFETs at maximum substrate current stress conditions. Li *et al.* [67] show that the benefit of deuterium is not restricted to this stress condition and also applies to the condition where the gate voltage equals the drain voltage. Monzio Compagnoni *et al.* [63] show that the benefit of deuterium also applies to p-channel MOSFETs when the degradation is due to the creation of interface states and not fixed oxide charge.

In all this work, also the virgin transistor characteristics were not altered by the incorporation of deuterium. The isotope effect only applies to the degradation. In contrast with this, Ohguro *et al.* [69] show a difference in $1/f$ noise in non-stressed MOSFETs. The measured $1/f$ noise level in MOSFETs with hydrogen terminated dangling bonds is higher than in MOSFETs with deuterium terminated dangling bonds. They used devices with a gate length of 0.11 µm with a silicon oxynitride gate dielectric of 1.9 to 2.1 nm thick and the deuterium is incorporated at the end of the process during the post-metal anneal.

According to Van de Walle and Jackson [70], the observed isotope effect on the interface quality can be explained by the difference in bending mode frequency of the Si–H and Si–D bond. Due to the mass difference, the bending mode frequency of the Si–D bond is lower than of the Si–H bond, with as a result that the bending mode frequency of the Si–D bond lies in the phonon spectrum of the silicon lattice. This gives rise to a strong frequency coupling, so a Si–D bond can effectively lose the excess energy to the silicon lattice. The same does not apply for the Si–H bond. This explains why it is much harder to desorb deuterium from a silicon surface (or interface) than hydrogen. During hot carrier stressing, the Si–H (or Si–D) bond adsorbs excess energy from hot carriers during collisions. The Si–H bond has no efficient coupling to the silicon lattice and cannot lose this excess energy quickly. When the next hot carrier collides and transfers energy to the bond, this energy adds to the already gained energy.

This continues until the excess energy is large enough for the bond to break. The Si–D bond on the other hand quickly loses the excess energy to the silicon lattice due to the vibrational frequency coupling. Although the energy needed to break a Si–D bond is almost equal to the energy needed to break a Si–H bond, this frequency coupling makes it much harder to supply sufficient energy to break the bond. The result is that a MOSFET with Si-SiO_2 interface dangling bonds passivated with deuterium is more resistant to hot carrier stress than one with hydrogen passivated dangling bonds.

This qualitative picture is supported by both simulations and measurements. Biswas *et al.* [71] use tight-binding molecular dynamics calculations to show a fast decay for the Si–D bending mode excitation in 1–2 oscillations (0.2 ps due to coupling with the silicon phonon spectrum, while the bending mode excitation for the Si–H lasts at least 8 ps. These calculations are done for a <111> silicon surface. Watanabe [72] uses surface infrared spectroscopy on a unreconstructed <111> silicon surface to show coupling of the Si–D bending mode with the silicon phonon spectrum. Shih *et al.* [73] do the same using Raman spectra of amorphous hydrogenated and deuterated silicon. Chen *et al.* [74] use fourier-transform infrared spectroscopy and show that the Si–D bending mode not only couples to the Si–Si transverse-optical bending mode, but also to the Si–O transverse-optical rocking mode. This indicates that there are two coupling paths. Finally, Chen *et al.* [75] use hot carrier degradation measurements to show that the isotope effect depends on the number of hot electrons and not on the electron energy. This supports the concept of multiple vibrational excitation in the above described theory.

6.5 Experimental setup

The interface quality of thin silicon oxide layers with hydrogen or deuterium incorporated during the process, has been tested on single MOS transistors and matched pair MOS transistors on the Philips set of wafers. The devices in this set have a gate oxide grown in either an O_2, H_2/O_2 or a D_2/O_2 ambient. A post-oxidation anneal is either omitted, or performed in a N_2 or D_2 ambient and finally the post-metal anneal is performed in either a H_2/N_2 or D_2/N_2 ambient. A detailed description of the device fabrication can be found in section 4.6. For the work described in this chapter only devices where

the post-oxidation anneal was omitted, were tested.

The degradation of transistor parameters, like threshold voltage and maximum transconductance, due to hot carrier stressing was measured on single MOS transistors with a gate length of 0.25 µm and a gate width of 10 µm. During the hot carrier stressing, the source and substrate contacts were grounded, the drain voltage was set at 3.5, 3.75 or 4.0 V and the gate voltage was set at 1.7, 1.9 or 1.9 V. This corresponds to peak substrate current conditions. The stress was periodically interrupted, to measure the linear input characteristics. During these measurements, the source and substrate were grounded, the drain voltage was set at 0.1 V and the gate voltage was stepped from 0 to 2.5 V with steps of 0.01 V.

In accordance with the definition, the transconductance g_m is extracted from the linear input characteristics using:

$$g_\mathrm{m} = \frac{\Delta I_\mathrm{d}}{\Delta V_\mathrm{g}} \tag{6.1}$$

To reduce the noise in the resulting g_m-V_g graph, ΔI_d was not taken for every two adjacent measurement points, i.e. for a ΔV_g of 0.01 V, but instead for a ΔV_g of 0.05 V.

The threshold voltage V_t was extracted from the linear input characteristics by calculating the intercept of the tangent at the point of maximum transconductance $g_\mathrm{m,max}$. This is just one of many ways of extracting the threshold voltage [76], but as the shift in threshold voltage is the interesting parameter, using the same extraction method throughout is more important than obtaining the absolute value.

As a second measure of the interface quality, the degradation of the noise behaviour was measured at matched pair MOS transistors with a gate length of 0.5 µm and a gate width of 2 µm. Both transistors have their own drain contacts. The gate, source and well contacts are common. During the hot carrier stressing, the source and substrate contacts were grounded, the drain voltage was set at 4.5 V and the gate voltage was set at 2.1 V. This corresponds to peak substrate current conditions. The stress was periodically interrupted to measure the linear input characteristics and the noise behaviour. The linear input characteristics were measured in the same way as on the single transistors and g_m and V_t were extracted using the same procedures.

The noise measurements were done in a differential setup [77] to cancel out the common mode noise arising from the measurement setup. The LF noise measurements were performed in the saturation

regime under a fixed drain current of 10 μA. To allow for a reliable comparison between the noise measured before and after stress the gate voltage was compensated for the increase in threshold voltage due to the hot-carrier stress.

6.6 Results

As mentioned in section 6.4, in literature it is reported that the incorporation of deuterium during processing does not influence the virgin transistor parameters, and therefore the virgin input characteristics. Figure 6.3 shows the linear input characteristics of five devices with a gate oxide grown in O_2. These devices did not receive a post-oxidation anneal, while the post-metal anneal was performed in D_2/N_2 an ambient. This picture is typical for all processing variants tested. Figure 6.4(a) shows there are small differences between the average input characteristics of the measured wafers, but there is no direct relation with the difference in processing. In fact, this wafer-to-wafer non-uniformity is to be expected and does not exceed the within-wafer uniformity. This is shown in figure 6.4(b). This figure shows the standard deviation in the measured drain current for devices with the same processing variant and compares this with the standard deviation of the average measured drain current per processing variant. This last standard deviation is in the same range as

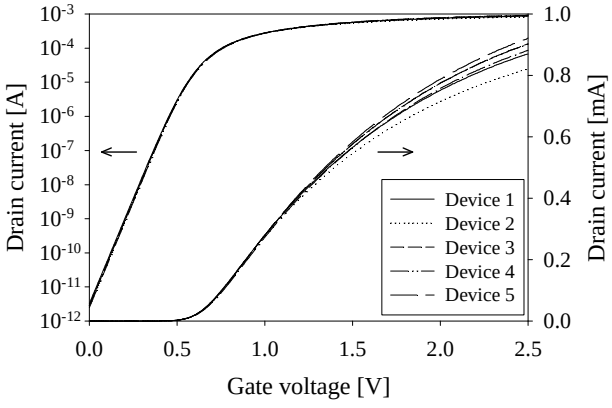
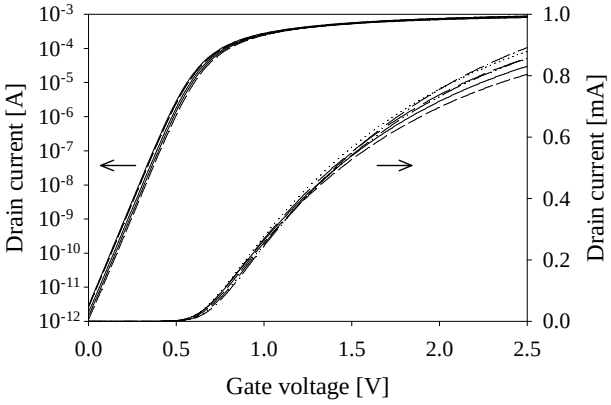
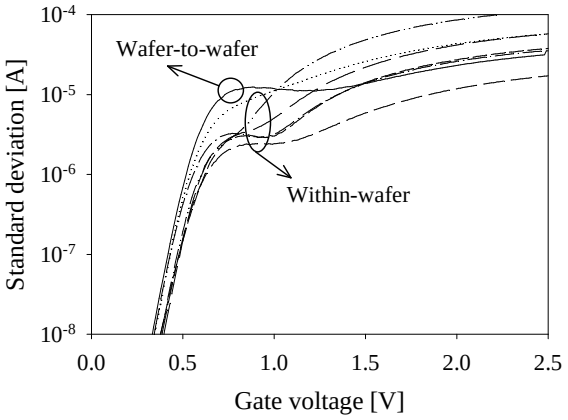


Figure 6.3: Within-wafer non-uniformity of the linear input characteristics for 5 devices with a gate oxide grown in O_2, no post-oxidation anneal and a post-metal anneal in H_2/N_2.



(a)



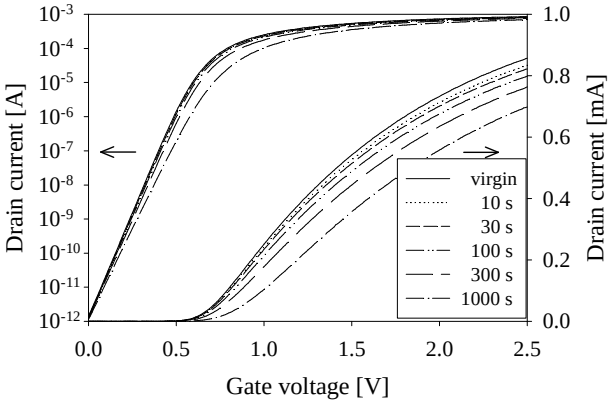
(b)

Figure 6.4: Wafer-to-wafer non-uniformity of the linear input characteristics. (a) Average linear input characteristics for the six measured processing variants in which the oxidation ambient (O_2, H_2/O_2 or D_2/O_2) and the post-metal annealing ambient (H_2/N_2 or D_2/N_2) were varied. (b) Comparison of standard deviation of the average linear input characteristics of (a) with the within-wafer standard deviation of the six processing variants.

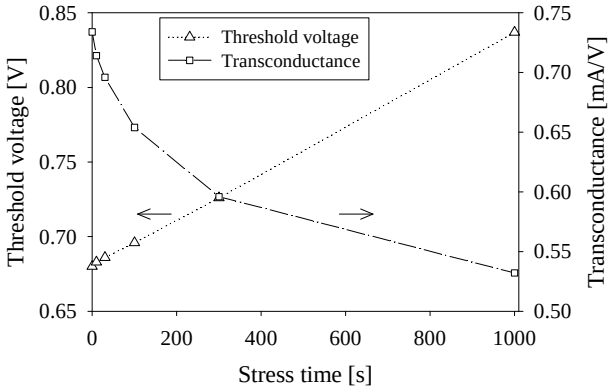
the others, implying that the wafer-to-wafer non-uniformity is merely a consequence of the general processing non-uniformity and cannot be traced to specific differences in processing.

Figure 6.5 shows the typical trend in the average linear input characteristic and the corresponding threshold voltage and transconductance with increasing hot carrier stress time, in this case for 5 devices with a gate oxide grown in H_2/O_2, no post-oxidation anneal and a post-metal anneal in H_2/N_2. This trend is observed for all processing variants. Figure 6.6 compares the rate of hot carrier degradation for the different processing variants. It is clear from this picture that the transistor parameters degrade less due to hot carrier stressing when hydrogen is replaced with deuterium during the post-metal anneal. Earlier incorporation of the deuterium does not have an effect, neither is there a difference between dry and wet oxidation. These observations do not only apply to the specific stress conditions used for the measurements of figure 6.6, but are general. This becomes clear in figure 6.7, the hot carrier lifetime of the devices is depicted as a function of substrate current. The lifetime is defined as a 5 % decrease in transconductance or 50 mV threshold voltage shift and is interpolated or extrapolated from the parameter versus time graphs.

As a second measure of the interface quality, the low-frequency noise characteristics were measured before and after hot carrier degradation. Figure 6.8(a) shows the low frequency noise spectra for a device post-metal annealed in H_2/N_2 and a device post-metal annealed in D_2/N_2, before and after hot carrier stress. These are just two devices and the spread in noise characteristics between devices on the same wafer can be substantial. To take this into account, several transistors were characterised for each wafer. The distribution of the noise level at 100 Hz for both processing variants is depicted in figure 6.8(b). This graph shows that the initial noise level for the deuterium annealed devices is slightly larger than for the devices annealed in hydrogen. This is in contrast with the results obtained by Oghuro *et al.* [69]. After hot carrier stress, the noise level for the hydrogen annealed devices is larger. In fact, while the distribution for the hydrogen annealed devices has shifted already substantially after 1000 s hot carrier stress, the distribution for the deuterium annealed devices does not shift at all. Even at 4000 s, the distribution of the deuterium annealed devices has not shifted as much as the distribution of the hydrogen annealed devices after 1000 s.

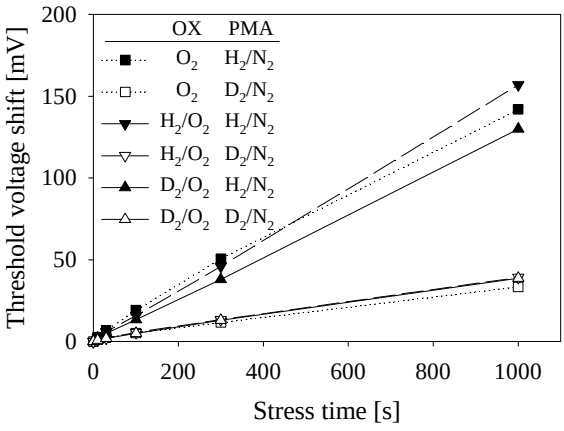


(a)

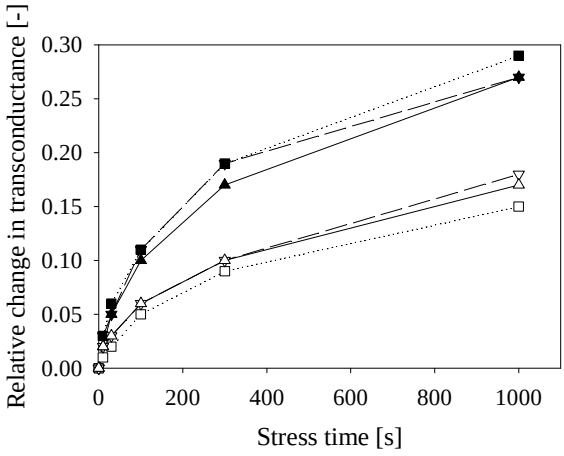


(b)

Figure 6.5: (a) Average linear input characteristics after different hot carrier stress times for 5 devices with a gate oxide grown in H_2/O_2, no post-oxidation anneal and a post-metal anneal in H_2/N_2 and (b) extracted threshold voltage and transconductance as a function of hot carrier stress time. The devices were stressed with a drain voltage of 4 V and a gate voltage of 1.9 V, the source and substrate were grounded. This corresponds to maximum substrate current conditions.

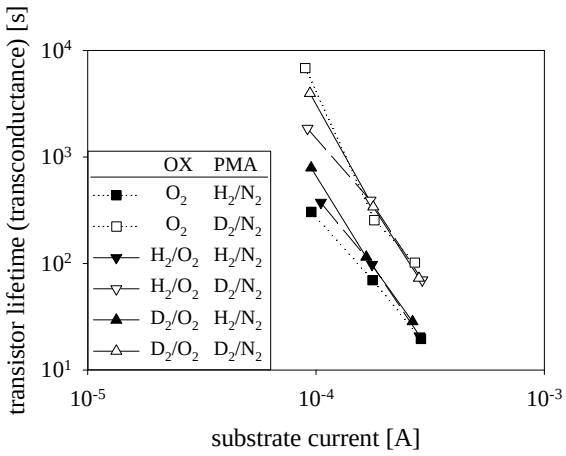


(a)

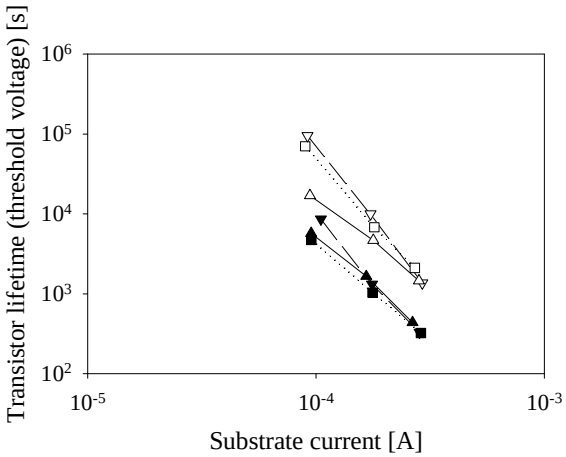


(b)

Figure 6.6: (a) and (b) Threshold voltage shift and relative change in transconductance as a function of hot carrier stress time, for the different processing variants. The devices were stressed with a drain voltage of 4 V and a gate voltage of 1.9 V, the source and substrate were grounded. This corresponds to maximum substrate current conditions.

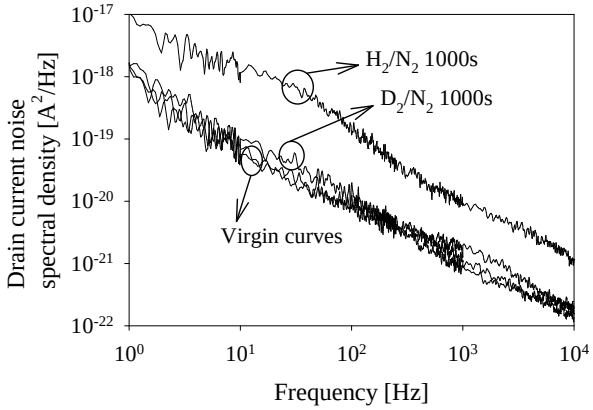


(a)

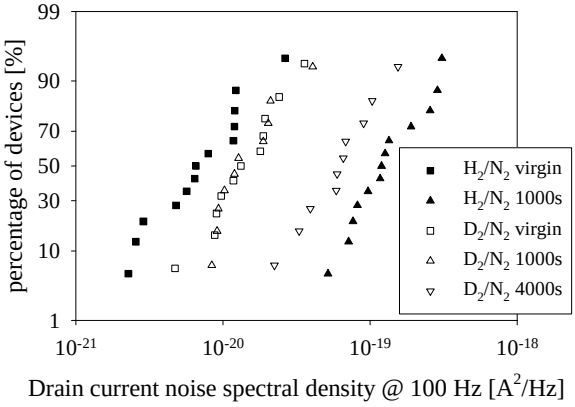


(b)

Figure 6.7: Transistor lifetime as a function of hot carrier stress substrate current, for the different processing variants. The devices were stressed with a drain voltage of 4 V and a gate voltage of 1.9 V, the source and substrate were grounded. This corresponds to maximum substrate current conditions. (a) The lifetime criterium was set at 5 % degradation in transconductance (b) The lifetime criterium was set at 50 mV threshold voltage shift.



(a)



(b)

Figure 6.8: Drain current noise spectral density for devices post-metal annealed in H_2/N_2 or D_2/N_2, before and after hot carrier stress. The devices were stressed with a drain voltage of 4.5 V and a gate voltage of 2.1 V, the source and substrate were grounded. This corresponds to maximum substrate current conditions. (a) depicts the noise spectrum of two devices and (b) depicts the noise distribution.

6.7 Discussion and conclusions

The results obtained in this chapter unambiguously show that the incorporation of deuterium into the gate dielectric has virtually no effect on transistor operation, but can improve the resistance to hot carrier degradation. This only applies to the case where deuterium is incorporated during the post-metal anneal. Introduction of deuterium during the gate dielectric growth gives no improvement. The isotope effect appears equally in wet and dry grown silicon oxide.

The fact that only deuterium introduced during the post-metal anneal is beneficial indicates that the deuterium introduced earlier in the process is substituted with hydrogen during subsequent processing. This most likely already happens during the deposition of the poly-crystalline silicon as gate material. During this step SiH_4 is used to form silicon. This means that with every deposited silicon atom, 4 hydrogen atoms become available and since the gate is on top of the gate dielectric these hydrogen atoms are close to the gate dielectric.

The fact that there is no noticeable difference between the dry and wet oxides, indicates that in both cases the silicon oxidation results in an equal quality interface. In both cases the minimum number of interface states after the post-metal anneal are achieved.

In this chapter, the impact of deuterium on the interface quality of silicon oxide was investigated. The transistor parameters threshold voltage and transconductance and the $1/f$ noise were measured before and after hot carrier stressing. Devices with deuterium incorporated during the gate dielectric growth showed no difference with devices without deuterium. Devices with deuterium incorporated during the post-metal anneal did show improved resistance against hot carrier degradation.

Chapter 7

Conclusions

Based on promising results in literature, in this thesis the impact of deuterium on the gate dielectric of CMOS devices has been investigated. To this aim, deuterium has been introduced at different stages in the CMOS manufacturing and the resulting devices have been electrically tested. Statistical data is obtained on two sets of devices fabricated in two different laboratories. The results have been tested with theory and data available in literature.

One of the processing stages to introduce deuterium is the gate dielectric growth. To be able to grow a reproducible gate dielectric using deuterium, the oxidation kinetics of silicon has been investigated in chapter 2, both for the case of a H_2O oxidising ambient and a D_2O oxidising ambient. The obtained oxidation data show that silicon oxidises faster in a H_2O ambient compared to a D_2O ambient. To obtain the same silicon oxide thickness, a longer oxidation time has to be used in case of D_2O oxidation. The observed difference in oxidation rate of 18 % can be explained by a difference in activation energy of 15 to 17 meV.

An attempt was made to model the data with existing physical silicon oxidation models. The results were not satisfactory. This is not entirely surprising because thin oxides were grown and the existing models are known not to describe initial oxide growth properly. To improve the description of the oxidation kinetics, a new silicon oxidation model has been proposed in chapter 3. This model considers the oxidation of silicon to be an electron-stimulated process. In the initial oxidation regime, the electrons are supplied at the silicon oxide surface due to electron tunnelling. For thicker silicon oxides, the

electrons are supplied at the Si-SiO_2 interface. The model is able to explain the measured oxidation data, but there is room for improvement. The major improvement is to be expected when the supply of electrons is not restricted to the surface and the interface, but occurs throughout the silicon oxide layer.

The obtained knowledge on the oxidation kinetics of silicon in a deuterated ambient has been used to fabricate CMOS devices with deuterium in the gate dielectric. Deuterium has also been introduced during other processing steps. The bulk quality of the gate dielectric of the resulting devices is the subject of chapter 5. The results showed no improvement of the bulk quality. Deuterated gate dielectrics degrade as fast as hydrogenated gate dielectrics under Fowler-Nordheim stressing conditions. This is independent of the processing stage deuterium was introduced. The results are well acceptable in view of our current understanding of SiO_2 degradation, but do contradict with several experimental observations reported in literature.

In chapter 6 the interface quality of deuterated gate dielectrics was investigated. It is shown that the incorporation of deuterium in the gate dielectric can strengthen the CMOS devices against hot carrier degradation. However, this only occurs when the deuterium is introduced during the post-metal anneal at the end of the process. Earlier incorporation of deuterium does not show improved resistance to hot carrier degradation. Most likely, because the early introduced deuterium is replaced with hydrogen during high temperature processing steps that follow the step introducing deuterium. Hydrogen is abundantly available in CMOS manufacturing, so this is a plausible idea.

Concluding it can be stated that for deuterium to be beneficial for CMOS devices, it has to be incorporated at the end of the CMOS manufacturing process, to avoid replacement with hydrogen. The only benefit will be a higher resistance to hot carrier degradation. This will allow for higher internal electric fields in the CMOS devices. Hot carrier degradation has been a primary reason for reduction of the supply voltage. The incorporation of deuterium can relax this demand for decreased supply voltage.

The incorporation of deuterium will not increase the resistance to gate dielectric bulk degradation. Therefore it cannot be used to reduce the tunnel dielectric thickness of floating gate devices.

Appendix

In this appendix the derivation of equation 3.13 is given.

First, combination of equations 3.4, 3.8 and 3.10 gives:

$$C_{\mathrm{O_iH_j^{b-}},\mathrm{s}} = \frac{x_\mathrm{o} k_\mathrm{I} C_{\mathrm{e,s}} C_{\mathrm{H_2O,s}} + D_{\mathrm{O_iH_j^{b-}},\mathrm{o}} C_{\mathrm{O_iH_j^{b-}},\mathrm{i}}}{D_{\mathrm{O_iH_j^{b-}},\mathrm{o}} + x_\mathrm{o} D_{\mathrm{O_iH_j^{b-}},\mathrm{g}}} \tag{1}$$

Second, combination of equations 3.3, 3.6, 3.9 and 3.7 gives:

$$\begin{aligned} C_{\mathrm{O_iH_j^{b-}},\mathrm{s}} \quad = \quad & \frac{D_{\mathrm{O_iH_j^{b-}},\mathrm{o}} + x_\mathrm{o} k_\mathrm{II}}{D_{\mathrm{O_iH_j^{b-}},\mathrm{o}}} C_{\mathrm{O_iH_j^{b-}},\mathrm{i}} - \\ & \frac{x_\mathrm{o} k_\mathrm{I} D_{\mathrm{H_2O,o}} C_{\mathrm{e,i}} C_{\mathrm{H_2O,s}}}{D_{\mathrm{O_iH_j^{b-}},\mathrm{o}} \left(D_{\mathrm{H_2O,o}} + x_\mathrm{o} k_\mathrm{I} C_{\mathrm{e,i}} \right)} \end{aligned} \tag{2}$$

Third, combination of equations 1 and 2 gives:

$$C_{\mathrm{O_iH_j^{b-}},\mathrm{i}} = \frac{\frac{x_\mathrm{o} k_\mathrm{I} C_{\mathrm{e,s}} C_{\mathrm{H_2O,s}}}{D_{\mathrm{O_iH_j^{b-}},\mathrm{o}} + x_\mathrm{o} D_{\mathrm{O_iH_j^{b-}},\mathrm{g}}} + \frac{x_\mathrm{o} k_\mathrm{I} D_{\mathrm{H_2O,o}} C_{\mathrm{e,i}} C_{\mathrm{H_2O,s}}}{D_{\mathrm{O_iH_j^{b-}},\mathrm{o}} \left(D_{\mathrm{H_2O,o}} + x_\mathrm{o} k_\mathrm{I} C_{\mathrm{e,i}} \right)}}{\frac{D_{\mathrm{O_iH_j^{b-}},\mathrm{o}} + x_\mathrm{o} k_\mathrm{II}}{D_{\mathrm{O_iH_j^{b-}},\mathrm{o}}} - \frac{D_{\mathrm{O_iH_j^{b-}},\mathrm{o}}}{D_{\mathrm{O_iH_j^{b-}},\mathrm{o}} + x_\mathrm{o} D_{\mathrm{O_iH_j^{b-}},\mathrm{g}}}} \tag{3}$$

Rewriting equation 3 and combining with equations 3.5 and 3.12 leads to:

$$\frac{dx_\mathrm{o}}{dt} = \frac{C_{\mathrm{H_2O,s}}}{N} \frac{D_{\mathrm{O_iH_j^{b-}},\mathrm{g}} D_{\mathrm{H_2O,o}} + k_\mathrm{I} D_{\mathrm{O_iH_j^{b-}},\mathrm{o}} C_{\mathrm{e,s}} + \frac{D_{\mathrm{H_2O,o}} D_{\mathrm{O_iH_j^{b-}},\mathrm{o}}}{x_\mathrm{o}} \left(1 + \frac{C_{\mathrm{e,s}}}{C_{\mathrm{e,i}}} \right)}{\left(1 + \frac{D_{\mathrm{H_2O,o}}}{x_\mathrm{o} k_\mathrm{I} C_{\mathrm{e,i}}} \right) \left(x_\mathrm{o} D_{\mathrm{O_iH_j^{b-}},\mathrm{g}} + D_{\mathrm{O_iH_j^{b-}},\mathrm{o}} \left(1 + \frac{D_{\mathrm{O_iH_j^{b-}},\mathrm{g}}}{k_\mathrm{II}} \right) \right)} \tag{4}$$

which is equation 3.13.

Bibliography

[1] "International Technology Roadmap for Semiconductors homepage." [Online]. Available: www.itrs.org

[2] P. Avouris, R. E. Walkup, A. R. Rossi, T. C. Shen, G. C. Abeln, J. R. Tucker, and J. W. Lyding, "STM-induced H atom desorption from Si(100): isotope effects and site selectivity," *Chemical Physics Letters*, vol. 257, no. 1-2, pp. 148–154, July 1996.

[3] J. W. Lyding, K. Hess, and I. C. Kizilyalli, "Reduction of hot electron degradation in metal oxide semiconductor transistors by deuterium processing," *Applied Physics Letters*, vol. 68, no. 18, pp. 2526–2528, Apr. 1996.

[4] H. Kim and H. Hwang, "Electrical and Reliability Characteristics of Ultrathin Gate Oxide Prepared by Oxidation in D_2O," *Japanese Journal of Applied Physics*, vol. 38, no. 2A, pp. L99–L101, Feb. 1999.

[5] S. Krishnan, S. Rangan, S. Hattangady, G. Xing, K. Brennan, M. Rodder, and S. Ashok, "Assessment of charge-induced damage to ultra-thin gate MOSFETs," in *IEDM technical digest*, Washington DC, USA, Dec. 1997, pp. 445–448.

[6] S. Sugiyama, J. Yang, and S. Guha, "Improved stability against light exposure in amorphous deuterated silicon alloy solar cell," *Applied Physics Letters*, vol. 70, no. 3, pp. 378–380, Jan. 1997.

[7] C. W. Liu, C.-H. Lin, M. H. Lee, S. T. Chang, Y.-H. Liu, M.-J. Chen, and C.-F. Lin, "Enhanced reliability of electroluminescence from metaloxidesilicon tunneling diodes by deuterium incorporation," *Applied Physics Letters*, vol. 78, no. 10, pp. 1397–1399, Mar. 2001.

[8] I. W. Wu, M. Koyanagi, S. Holland, T. Y. Huang, J. C. Mikkelsen, Jr., R. H. Bruce, and A. Chiang, "Breakdown Yield and Lifetime of thin Gate Oxides in CMOS Processing," *Journal of the Electrochemical Society*, vol. 136, no. 6, pp. 1638–1645, June 1989.

[9] W.-J. Cho, Y.-C. Kim, E.-S. Kim, and H.-S. Kim, "Effects of Oxidation Ambient and Low Temperature Post Oxidation Anneal on the Silicon/Oxide Interface Structure and the Electrical Properties of the Thin Gate Oxide," *Japanese Journal of Applied Physics*, vol. 38, no. 1A, pp. 12–16, Jan. 1999.

[10] M. H. Cho, J. S. Shin, Y. S. Roh, I. W. Lyo, K. Jeong, C. N. Whang, J. S. Lee, J. Y. Yoo, N. I. Lee, K. Fujihara, and D. W. Moon, "Characteristics of ultrathin SiO_2 films using dry rapid thermal oxidation and Pt catalyzed wet oxidation," *Journal of Vacuum Science and Technology A*, vol. 21, no. 4, pp. 1004–1008, July/Aug. 2003.

[11] Y. Mitani, H. Satake, H. Ito, and A. Toriumi, "Highly Reliable Gate Oxide under Fowler-Nordheim Electron Injection by Deuterium Pyrogenic Oxidation and Deuterated Poly-Si Deposition," in *IEDM technical digest*, San Francisco, California, USA, Dec. 2000, pp. 343–346.

[12] P. A. Stolk, A. C. M. C. van Brandenburg, and A. H. Montree, "Oxidation enhanced diffusion during the growth of ultrathin oxides," *Material Science in Semiconductor Processing*, vol. 2, no. 1, pp. 29–33, Apr. 1999.

[13] B. E. Deal and A. S. Grove, "General Relationship for the Thermal Oxidation of Silicon," *Journal of Applied Physics*, vol. 36, no. 12, pp. 3770–3778, Dec. 1965.

[14] R. B. Beck and B. Majkusiak, "The model of growth kinetics of very thin thermal silicon oxide layer," *Electron Technology*, vol. 21, no. 1-2, pp. 65–79, 1988.

[15] H. Z. Massoud and J. D. Plummer, "Thermal Oxidation of Silicon in Dry Oxygen Growth-Rate Enhancement in the Thin Regime I: Experimental Results," *Journal of the Electrochemical Society*, vol. 132, no. 11, pp. 2685–2693, Nov. 1985.

[16] J. D. Plummer, M. D. Deal, and P. B. Griffin, *Silicon VLSI Technology; Fundamentals, Practice and Modeling.* New Jersey, USA: Prentice Hall, Inc, 2000.

[17] A. Reisman, E. H. Nicollian, C. K. Williams, and C. J. Merz, "The Modelling of Silicon Oxidation from 1×10^{-5} to 20 Atmospheres," *Journal of Electronic Materials*, vol. 16, no. 1, pp. 45–55, Jan. 1987.

[18] E. H. Nicollian and A. Reisman, "A New Model for the Thermal Oxidation Kinetics of Silicon," *Journal of Electronic Materials*, vol. 17, no. 4, pp. 263–272, July 1988.

[19] T. Y. Tan and U. G osele, "Growth kinetics of oxidation-induced stacking faults in silicon: A new concept," *Applied Physics Letters*, vol. 39, no. 1, pp. 86–88, July 1981.

[20] D. R. Wolters and A. T. A. Zegers-van Duijnhoven, "Silicon Oxidation and Fixed Oxide Charge," *Journal of the Electrochemical Society*, vol. 139, no. 1, pp. 241–249, Jan. 1992.

[21] P. J. Jorgensen, "Electrolysis of SiO_2 on Silicon," *Journal of Chemical Physics*, vol. 49, no. 4, pp. 1594–1598, Aug. 1968.

[22] C. Wagner, "Beitrag zur Theorie des Anlaufsvorgangs," *Zeitschrift für physikalische Chemie B*, vol. 21, no. 1-2, pp. 25–41, Jan. 1933.

[23] D. R. Lide, Ed., *CRC Handbook of Chemistry and Physics*, 73rd ed. Boca Raton, Florida, USA: CRC Press, 1992, special student edition.

[24] B. E. Deal, "Thermal Oxidation Kinetics of Silicon in Pyrogenic H_2O and 5% HCl/H_2O Mixtures," *Solid-State Science and Technology*, vol. 125, no. 4, pp. 576–579, Apr. 1978.

[25] S. M. Sze, *Modern semiconductor device physics.* New York, USA: John Wiley and Sons, Inc., 1998.

[26] H. Tuinhout, M. Pelgrom, R. Penning de Vries, and M. Vertregt, "Effects of Metal Coverage on MOSFET Matching," in *IEDM technical digest*, San Francisco, California, USA, Dec. 1996, pp. 735–738.

[27] T. Yamaha and F. Masuoka, "Influence of Retarding Hydrogen Diffusion in Boron Phosphosilicate Glass on Annealing Damage of Metal-Oxide Semiconductor Transistors," *Journal of the Electrochemical Society*, vol. 146, no. 8, pp. 3065–3069, Aug. 1999.

[28] B. E. Deal, E. L. MacKenna, and P. L. Castro, "Characteristics of Fast Surface States Associated with SiO_2-Si and Si_3N_4-SiO_2-Si Structures," *Journal of the Electrochemical Society*, vol. 116, no. 7, pp. 997–1005, July 1969.

[29] R. H. Fowler and L. Nordheim, "Electron Emission in Intense Electric Fields," in *Proceedings of the Royal Society London*, vol. A119, May 1928, pp. 173–181.

[30] K. F. Schuegraf, C. C. King, and C. Hu, "Ultra-thin Silicon Dioxide Leakage Current and Scaling Limit," in *Symposium on VLSI technical digest*, Seattle, Washington, USA, June 1992, pp. 18–19.

[31] J. Maserjian and N. Zamani, "Behaviour of the Si/SiO_2 interface observed by Fowler-Nordheim tunneling," *Journal of Applied Physics*, vol. 53, no. 1, pp. 559–567, Jan. 1982.

[32] W. Weibull, "A statistical theory of the strength of materials," in *Proceedings of the Royal Swedish Institute for Engineering Research*, vol. 151, 1939, pp. 1–45.

[33] R. Degraeve, B. Kaczer, and G. Groeseneken, "Degradation and breakdown in thin oxide layers: mechanisms, models and reliability prediction," *Microelectronics Reliability*, vol. 39, no. 10, pp. 1445–1460, Oct. 1999.

[34] J. W. McPherson and D. A. Baglee, "Acceleration factors for thin gate oxide stresing," in *Proceedings of the IRPS*, vol. 23, Orlando, Florida, USA, Mar. 1985, pp. 1–5.

[35] J. W. McPherson and H. C. Mogul, "Underlying physics of the thermochemical E model in describing low-field time-dependent dielectric breakdown in SiO_2 thin films," *Journal of Applied Physics*, vol. 84, no. 3, pp. 1513–1523, Aug. 1998.

[36] I. C. Chen, S. Holland, and C. Hu, "A quantative physical model for time-dependent breakdown in SiO_2," in *Proceedings of the IRPS*, vol. 23, Orlando, Florida, USA, Mar. 1985, pp. 24–31.

[37] I. C. Chen, S. Holland, K. K. Y. C. Chang, and C. Hu, "Substrate hole current and oxide breakdown," *Applied Physics Letters*, vol. 49, no. 11, pp. 669–671, Sept. 1986.

[38] K. F. Schuegraf and C. Hu, "Hole Injection SiO_2 Breakdown Model for Very Low Voltage Lifetime Extrapolation," *IEEE Transactions on Electron Devices*, vol. 41, no. 5, pp. 761–767, May 1994.

[39] D. J. DiMaria and J. W. Stasiak, "Trap creation in silicon dioxide produced by hot electrons," *Journal of Applied Physics*, vol. 65, no. 6, pp. 2342–2356, Mar. 1989.

[40] D. J. DiMaria and E. Cartier, "Mechanism for stress-induced leakage currents in thin silicon dioxide films," *Journal of Applied Physics*, vol. 78, no. 6, pp. 3883–3894, Sept. 1995.

[41] H. Kim and H. Hwang, "High-quality ultrathin gate oxide prepared by oxidation in D_2O," *Applied Physics Letters*, vol. 74, no. 5, pp. 709–710, Feb. 1999.

[42] H. Kwon and H. Hwang, "Electrical Characteristics of Ultra-thin Gate Oxide Prepared by Oxidation in D_2O," in *Proceedings of the ESSDERC*, vol. 29, Leuven, Belgium, Sept. 1999, pp. 416–419.

[43] H. Kwon and H. Hwang, "Electrical characteristics of ultrathin gate oxide prepared by postoxidation annealing in ND_3," *Applied Physics Letters*, vol. 76, no. 6, pp. 772–773, Feb. 2000.

[44] H. Kwon, I. Yeo, and H. Hwang, "Electrical Characteristics of Ultra-thin Oxynitride Gate Dielectric Prepared by Reoxidation of Thermal Nitride in D_2O," *Japanese Journal of Applied Physics*, vol. 39, no. 4A, pp. L273–L274, Apr. 2000.

[45] S. Jeon and H. Hwang, "Electrical characteristics of ultrathin ZrO_2 prepared by wet oxidaton of an ultrathin Zr-metal layer," *Journal of Vacuum Science and Technology B*, vol. 20, no. 1, pp. 400–403, Jan./Feb. 2002.

[46] H. Jung, K. Im, D. Yang, and H. Hwang, "Electrical and Reliability Characteristics of an Ultrathin TaO_xN_y Gate Dielectric Prepared by ND_3 Annealing of Ta_2O_5," *IEEE Electron Device Letters*, vol. 21, no. 12, pp. 563–565, Dec. 2000.

[47] H. Sim and H. Hwang, "Effect of deuterium postmetal annealing on the reliability characteristics of an atomic-layer-deposited HfO_2/SiO_2 stack gate dielectrics," *Applied Physics Letters*, vol. 81, no. 21, pp. 4038–4039, Nov. 2002.

[48] Y. Mitani, H. Satake, H. Ito, and A. Toriumi, "A Study of the Effect of Deuterium on Stress-Induced Leakage Current," *Japanese Journal of Applied Physics*, vol. 39, no. 6B, pp. L564–L566, June 2000.

[49] Y. Mitani, H. Satake, H. Ito, and A. Toriumi, "Suppression of Stress-Induced Leakage Current After Fowler-Nordheim Stressing by Deuterium Pyrogenic Oxidation and Deuterated Poly-Si Deposition," *IEEE Transactions on Electron Devices*, vol. 49, no. 7, pp. 1192–1197, July 2002.

[50] C.-H. Lin, M. H. Lee, and C. W. Liu, "Correlation between Si-H/D bond desorption and injected electron energy in metal-oxide-silicon tunneling diodes," *Applied Physics Letters*, vol. 78, no. 5, pp. 637–639, Jan. 2001.

[51] M. H. Lee, C.-H. Lin, and C. W. Liu, "Novel Methods to Incorporate Deuterium in the MOS Structures," *IEEE Electron Device Letters*, vol. 22, no. 11, pp. 519–521, Nov. 2001.

[52] C.-H. Lin, F. Yuan, B.-C. Hsu, and C. W. Liu, "Isotope effect of hydrogen release in metal/oxide/n-silicon tunneling diodes," *Solid-State Electronics*, vol. 47, no. 6, pp. 1123–1126, June 2003.

[53] J. Wu, E. Rosenbaum, B. MacDonald, E. Li, J. Tau, B. Tracy, and P. Fang, "Anode Hole Injection versus Hydrogen Release: The Mechanism for Gate Oxide Breakdown," in *Proceedings of the IRPS*, vol. 38, San Jose, California, USA, Apr. 2000, pp. 27–32.

[54] D. Esseni, J. D. Bude, and L. Selmi, "Deuterium Effect on Interface States and SILC Generation in CHE Stress Conditions: A Comparative Study," in *IEDM technical digest*, San Francisco, California, USA, Dec. 2000, pp. 339–342.

[55] D. Esseni, J. D. Bude, and L. Selmi, "On Interface and Oxide Degradation in VLSI MOSFETs—Part I: Deuterium Effect in CHE Stress Regime," *IEEE Transactions on Electron Devices*, vol. 49, no. 2, pp. 247–253, Feb. 2002.

[56] B. C. Lin, Y. Cheng, A. Chin, T. Wang, and C. Tsai, "Deuterium Effect on Stress-Induced Leakage Current," *Japanese Journal of Applied Physics*, vol. 38, no. 4B, pp. 2337–2340, Apr. 1999.

[57] S. R. Morrison, *Electrochemistry at Semiconductor and Oxidized Metal Electrodes*. New York, USA: Plenum Press, 1980.

[58] J. Jacquelin, "Inference of Sampling on Weibull Parameter Estimation," *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 3, no. 6, pp. 809–816, Dec. 1996.

[59] B. E. Deal, "Standardized terminology for oxide charges associated with thermally oxidized silicon," *IEEE Transactions on Electron Devices*, vol. 27, no. 3, pp. 606–608, Mar. 1980.

[60] R. A. B. Devine, J. L. Autran, W. L. Warren, K. L. Vanheusdan, and J. C. Rostaing, "Interfacial hardness enhancement in deuterium annealed 0.25 μm channel metal oxide semiconductor transistors," *Applied Physics Letters*, vol. 70, no. 22, pp. 2999–3001, June 1997.

[61] S. Aur, T. Grider, V. McNeil, T. Holloway, and R. Eklund, "Remote plasma nitridation, deuterium anneal and pocket implant effects on NMOS hot carrier reliability," *Microelectronics Reliability*, vol. 39, no. 5, pp. 673–679, May 1999.

[62] W. F. Clark, T. G. Ference, S. W. Mittl, J. S. Burnham, and E. D. Adams, "Improved Hot-Electron Reliability in High-Performance, Multilevel-Metal CMOS Using Deuterated Barrier-Nitride Processing," *IEEE Electron Device Letters*, vol. 20, no. 10, pp. 501–503, Oct. 1999.

[63] C. Monzio Compagnoni, A. Pirovano, and A. Lacaita, "Degradation Dynamics for Deep Scaled p-MOSFET's during Hot-Carrier Stress," in *Proceedings of the ESSDERC*, vol. 32, Firenze, Italy, Sept. 2002, pp. 559–562.

[64] S. Watanabe and Y. Tamura, "Reliability Improvement in Deep-Submicron nMOSFET's by Deuterium," *FUIJTSU Science and Technology Journal*, vol. 39, no. 1, pp. 84–93, June 2003.

[65] R. Choi, K. Onishi, C. S. Kang, H.-J. Cho, Y. Kim, S. Krishnan, M. Akbar, and J. Lee, "Effects of deuterium anneal on MOSFETs with HfO_2 gate dielectrics," *IEEE Electron Device Letters*, vol. 24, no. 3, pp. 144–146, Mar. 2003.

[66] J. Lee, K. CHeng, Z. Chen, K. Hess, J. W. Lyding, Y.-K. Kim, H.-S. Lee, Y.-W. Kim, and K.-P. Suh, "Application of High Pressure Deuterium Annealing for Improving the Hot Carrier Reliability of CMOS Transistors," *IEEE Electron Device Letters*, vol. 21, no. 5, pp. 221–223, May 2000.

[67] E. Li, E. Rosenbaum, J. Tau, G. C.-F. Yeap, M.-R. Lin, and P. Fang, "Hot Carrier Effects in nMOSFETs in 0.1 μm CMOS Technology," in *Proceedings of the IRPS*, vol. 37, San Diego, California, USA, Mar. 1999, pp. 253–258.

[68] I. C. Kizilyalli, J. W. Lyding, and K. Hess, "Deuterium Post-Metal Annealing of MOSFET's for Improved Hot Carrier Reliability," *IEEE Electron Device Letters*, vol. 18, no. 3, pp. 81–83, Mar. 1997.

[69] T. Ohguro, Y. Okayama, K. Matsuzawa, K. Matsunaga, N. Aoki, K. Kojima, H. S. Momose, and K. Ishimaru, "The impact of oxynitride process, deuterium annealing and STI stress to 1/f noise of 0.11 *mu*m CMOS," in *Symposium on VLSI technical digest*, Kyoto, Japan, June 2003, pp. 37–38.

[70] C. G. van de Walle and W. B. Jackson, "Comment on "Reduction of hot electron degradation in metal oxide semiconductor transistors by deuterium processing" [Appl. Phys. Lett. 68, 2526 (1996)]," *Applied Physics Letters*, vol. 69, no. 16, p. 2441, Oct. 1996.

[71] R. Biswas, Y. P. Li, and B. C. Pan, "Enhanced stability of deuterium in silicon," *Applied Physics Letters*, vol. 72, no. 26, pp. 3500–3502, June 1998.

[72] S. Watanabe, "Isotopic shift and broadening of Si–D bending vibration on Si(111)," *Applied Surface Science*, vol. 162-163, pp. 146–151, Aug. 2000.

[73] A. Shih, S.-C. Lee, and C. ta Chia, "Evidence for coupling of Si–Si lattice vibration and Si–D wagging vibration in deuterated amorphous silicon," *Applied Physics Letters*, vol. 74, no. 22, pp. 3347–3349, May 1999.

[74] Z. Chen, J. Guo, and P. Ong, "Evidence for energy coupling from the Si–D vibration mode to the Si–Si and Si–O vibration

modes at the SiO_2/Si interface," *Applied Physics Letters*, vol. 83, no. 11, pp. 2151–2153, Sept. 2003.

[75] Z. Chen, P. Ong, A. K. Mylin, V. Singh, and S. Chetlur, "Direct evidence of multiple vibrational excitation for the Si—H/D bond breaking in metal-oxide-semiconductor transistors," *Applied Physics Letters*, vol. 81, no. 17, pp. 3278–3280, Oct. 2002.

[76] A. Ortiz-Conde, F. J. G. Sánchez, J. J. Liou, A. Cerdeira, M. Estrada, and Y. Yue, "A review of recent MOSFET threshold voltage extraction methods," *Microelectronics Reliability*, vol. 42, no. 4-5, pp. 583–596, Apr./May 2002.

[77] A. P. van der Wel, E. A. M. Klumperink, S. L. J. Gierkink, R. F. Wassenaar, and H. Wallinga, "MOSFET $1/f$ Noise Measurement Under Switched Bias Conditions," *IEEE Electron Device Letters*, vol. 21, no. 1, pp. 43–46, Jan. 2000.

Dankwoord

Zoals de titel al aangeeft, wil ik op deze plek graag een aantal mensen bedanken, mensen die me hebben geholpen en gesteund tijdens mijn onderzoek en het uiteindelijke schrijven van dit proefschrift.

Ik wil graag beginnen bij mijn collega's op vloer 3. Ik heb ruim vier jaar met plezier mijn werk verricht, mede door de goede werksfeer. De Twentse kwartiertjes in de koffiekamer en de wandelingen tijdens de lunch waren een prettige afwisseling op het werk.

Een aantal mensen van vloer 3 wil ik graag in het bijzonder bedanken. Mijn begeleiders, Pierre en Herma en later Alexey en Jurriaan. Zonder jullie was dit boekje er zeker niet geweest. Tom en Ton, bedankt voor de hulp in de clean room en Marcel en Henk in het meetlab. Marie-Christine, Margie, Annemiek en Gerdien, bedankt voor de secretariële ondersteuning en Cor en Frederik voor de computerondersteuning. Viet, Nader, Zhichun en Jay, het was gezellig in 3124.

Niet alle elektronische componenten, die ik doorgemeten heb tijdens mijn onderzoek, heb ik zelf gefabriceerd. Een deel komt bij Philips Research Leuven vandaan. De fabricage van deze componenten is opgestart en begeleid door Rob van Schaijk en de Q_{bd}-metingen hieraan zijn uitgevoerd door Wilko Baks. Jongens, bedankt.

Verder heb ik een aantal studenten mogen begeleiden bij het uitvoeren van een meetopdracht, waarbij het resultaat bruikbaar was voor mijn eigen onderzoek. Edward, Jan Laurens, Mark, Anna-Jo, Wouter, Eric en Erik, bedankt. Cora, bedankt voor het aanleveren van deze 'meetslaafjes'.

Tenslotte heb ik niet alleen veel steun mogen ontvangen op de werkvloer, maar ook daarnaast. Mijn ouders hebben me altijd en onvoorwaardelijk gesteund in alles wat ik doe. Pap en mam, bedankt. Hetzelfde geldt voor mijn vrouw Nanda. Waar zou ik zijn zonder jou? En als allerlaatste, Lotte, zo jong en toch al zo onmisbaar.

List of publications

Journal papers

- **A.J. Hof**, A.Y. Kovalgin, R. van Schaijk, W.M. Baks, J. Schmitz, *The Impact of Deuterated CMOS Processing on Gate Oxide Reliability*, IEEE Transactions on Electron Devices, submitted for publication.
- **A.J. Hof**, A.Y. Kovalgin, P.H. Woerlee, J. Schmitz, *On the Oxidation Kinetics of Silicon in Ultra-Diluted H_2O and D_2O Ambient*, Journal of the Electrochemical Society, submitted for publication.
- V.H. Nguyen, **A.J. Hof**, H. van Kranenburg, P.H. Woerlee, *Copper Chemical Mechanical Polishing Using Slurry-free Technique*, Micro-Electronic Engineering, Volume 55, pages 305–312, 2001.

Conference contributions

- **A.J. Hof**, A.Y. Kovalgin, J. Schmitz, R. van Schaijk, W.M. Baks, *Gate Oxide Reliability and Deuterated CMOS Processing*, Integrated Reliability Workshop (IRW), Lake Tahoe, USA, 2004, oral presentation.
- **A.J. Hof**, A.Y. Kovalgin, J. Schmitz, *On Oxidation of Silicon in Ultra-Diluted H_2O and D_2O Ambient*, Workshop on Dielectrics in Microelectronics (WoDiM), Cork, Ireland, 2004.
- J. Schmitz, M.H.H. Weusthof, **A.J. Hof**, *Leakage Current Correction in Quasi-Static C-V Measurements*, International Conference on Microelectronic Test Structures (ICMTS), Japan, 2004.
- **A.J. Hof**, A.Y. Kovalgin, P.H. Woerlee, J. Schmitz, *On Oxidation Kinetics and Electrical Quality of Gate Oxide Grown in H_2O*

or D_2O Ambient, Proceedings of the Semiconductor Advance for Future Electronics (SAFE), Velthoven, The Netherlands, 2003.

- **A.J. Hof**, A.Y. Kovalgin, P.H. Woerlee, *Comparison of H_2O and D_2O Oxidation Kinetics of <100> Silicon*, Proceedings of the Semiconductor Advance for Future Electronics (SAFE), Velthoven, The Netherlands, 2002.

- **A.J. Hof**, J. Holleman, P.H. Woerlee, *Gate Current for p^+-poly PMOS Devices under Gate Injection Conditions*, Proceedings of the Semiconductor Advance for Future Electronics (SAFE), Velthoven, The Netherlands, 2001.

- **A.J. Hof**, H. van Kranenburg, P.H. Woerlee, *Calculation of Interface State Tunnelling*, Proceedings of the Semiconductor Advance for Future Electronics (SAFE), Velthoven, The Netherlands, 2000.

- V.H. Nguyen, **A.J. Hof**, H. van Kranenburg, P.H. Woerlee, *Copper Chemical Mechanical Polishing Using Fixed-abrasives Polishing Pad*, Proceedings of the Semiconductor Advance for Future Electronics (SAFE), Mierlo, The Netherlands, 1999.

- V.E. Houtsma, **A.J. Hof**, J. Holleman, C. Salm, F.P. Widdershoven, P.H. Woerlee, *DC-SILC in p^+ Poly MOS Capacitors with Poly-Si and Poly-$Si_{0.7}Ge_{0.3}$ Gate Material*, Semiconductor Interface Specialist Conference (SISC), San Diego, USA, 1998.

- V.E. Houtsma, **A.J. Hof**, J. Holleman, P.H. Woerlee, *The Influence of Fluorine on the Stress-Induced Leakage Current Characteristics in n^+ Poly-Si MOS Capacitors*, Proceedings of the Semiconductor Advance for Future Electronics (SAFE), Mierlo, The Netherlands, 1998.

Patent

- Nguyen, Hoang, V.; **Hof, Albert, J.**; Van Kranenburg, Herma; Woerlee, Pierre, H.; *Method and Apparatus for Chemical-Mechanical Polishing (CMP) Using Upstream and Downstream Fluid Dispensing Means*; WO 02/051589 A1, 4 juli 2002.

About the author

Albert Jan Hof, better known as André Hof, was born on 11 August 1976 in Wolvega, the Netherlands. During his years at secondary school, his interest was drawn to science and technology. In 1994, he started to study Electrical Engineering at the *University of Twente*, the Netherlands. Here, among a number of interesting directions, he focussed on microsystems and -electronics. During his internship at the *Institute of Maritime Technology* located near Capetown (South Africa) he designed the electronics for a small robotic arm. His master thesis was on the topic of chemical mechanical polishing (CMP) for microelectronic applications. The research for this, was carried out in the group of Hans Wallinga, *Semiconductor Components*, under the supervision of Pierre Woerlee. André graduated with honours in 1999. Directly after graduation (and a holiday) he started working towards his PhD degree in the same group, starting under the supervision of Pierre Woerlee. After Pierre left the group, supervision was taken over by Jurriaan Schmitz. The underlying thesis is the result of this PhD work.