Dielectric engineering

Characterization, development and process damage minimization of various silicon oxides

Jan Germain Gabriel Ackaert

ISBN : 90-365-2069-x

DIELECTRIC ENGINEERING

CHARACTERIZATION, DEVELOPMENT AND PROCESS DAMAGE MINIMIZATION OF VARIOUS SILICON OXIDES

DISSERTATION

to obtain
the doctor's degree at the University of Twente,
on the authority of the rector magnificus,
prof.dr. F.A. van Vught,
on the account of the decision of the graduation committee,
to be publicly defended
on Wednesday 8th of September 2004 at 15.00

by

Jan Germain Gabriel Ackaert

born on 31st of December 1958

in Kortrijk, Belgium

Dit proefschrift is goedgekeurd door de promotor
prof.dr.ir. F.G Kuper

'…. doe nou maar wat je zegt,
dat is al moeilijk genoeg …..'
naar Bram Faas†

Aan Jonas, Katrijn, Karel en Hilde voor de
vele uren dat papa weer zat te 'tikken'
en mijn ouders voor hun nooit aflatende steun

De promotiecommissie:

Voorzitter:	
Prof.dr. W.H. Zijm	Universiteit Twente, EWI
Secretaris:	
Prof.dr. W.H. Zijm	Universiteit Twente, EWI
Promotor:	
Prof.dr.ir. F.G. Kuper	Universiteit Twente / Philips Semiconductors
Referent:	
Dr. A. Scarpa	Philips Semiconductors Nijmegen
Leden:	
Prof.dr. J. Schmitz	Universiteit Twente, EWI
Prof.dr.ir. A.J. Mouthaan	Universiteit Twente, EWI
Prof.dr.ir. G. Groeseneken	KU Leuven

Levensloop

Jan Ackaert werd op 31 december 1958 geboren te Kortrijk in de Belgische provincie West-Vlaanderen. Hij behaalde in 1976 zijn Humaniora diploma aan het Onze-Lieve-Vrouwe College te Oostende. Hierna studeerde hij Chemie aan de Industriële Hogeschool te Oostende waar hij afstudeerde in de richting Biochemie op de gaschromatografische bepaling van aminozuren in serum. Het ingenieurs diploma werd in 1980 ‘met lof’ behaald. Hij vervulde zijn militaire dienst waarna hij in 1982 in dienst trad bij Centraal Laboratorium UCB te Oostende. In 1983 trad hij in dienst bij Bayer te Brugge als verantwoordelijke voor het analytisch laboratorium en milieu. Sinds 1988 werkt hij bij de halfgeleider fabrikant AMIS te Oudenaarde. Na een aantal jaren als proces support ingenieur in de dunne films en implant afdeling actief te zijn geweest werd hij in 1994 aangesteld als proces module ontwikkeling ingenieur. Hier werkte hij aan de ontwikkeling van diverse proces modules: metallisatie, intermetaal diëlectrica, planarisatie, pre-epi proces voor bipolaire en hoogspannings technologiëen, diëlectrica als basis voor selectieve epi depositie, diëlectrica voor metaal-metaal capaciteiten, tunnel en interpoly oxides in geheugen cellen. Als rode draad door de ontwikkeling van de proces modules was er de impact van de effecten van het plasma tijdens het fabriceren van de halfgeleider. Over deze materie werden dan ook vele publicaties gemaakt onder andere in Solid State Electronics, Journal Microelectronics Reliability en IEEE Transactions on Electron Devices en op de conferenties P2ID, ESSDERC, ESPID, IPFA, ESREF en ICICDT. Hij werd uitgenodigd als conference committee member voor de ICICDT conferentie bij de werkgroepen plasma damage and advanced memory devices.

Biography

Jan Ackaert was born in Kortrijk, Belgium in 1958. He received the M.S. degree in Chemistry ‘with honor’ from Ostend Industrial High School in 1980 with a specialization in biochemistry. After one year with the central laboratory of UCB in Ostend, he joined Bayer as the manager or the analytical laboratory and environment. In 1988 he joined AMIS in Oudenaarde as process support engineer. In 1992 he became process module development engineer covering various process modules as metal interconnects, inter metal dielectrics, buried layers for bipolar and high voltage technologies, dielectrics for selective epi depostion, the dielectrics for metal-metal capacitors, tunnel and interpoly dielectrics for memory cells. As a constant through the development of the process modules, there was interest for the effects of the plasma processes on the devices during the manufacturing of the semiconductor. In this field many publications are made among others in Solid State Electronics, Journal Microelectronics Reliability and the IEEE Transactions on Electron Devices, and on the conferences P2ID, ESSDERC, ESPID, IPFA, ESREF and ICICDT. He is invited as conference committee member of the ICICDT conference with the workgroups plasma damage and advanced memory devices.

Dankwoord

Vele mensen hebben de afgelopen jaren bijgedragen aan de totstandkoming van de artikelen die in dit proefschrift bijeengebracht zijn. Als eerste dank ik hierbij mijn collega's van de Proces Module Development Group van AMIS:. Ze hebben mij geassisteerd en gestimuleerd mijn eerste wetenschappelijke publicaties te maken, met name Sylvie Boonen, Paul Colson, Peter Coppens, Malik Fatkhoudinov en Hocine Ziad (in alfabetische volgorde). Heel erg belangrijk en gewaardeerd hierbij zijn het management van Eddy De Backer en de directie van Marnix Tack. In deze groep heb ik geleerd dat werken best met veel plezier te combineren valt waar de genoemde groepsleden veel aan bijgedragen hebben.

In Oudenaarde is het werk klimaat in technology research en development groep en de nauwe samenwerking met de process support engineers in de waferfab en de designers, device fysici en product engineers een uitstekende voedingsbodem: Een ruime keuze aan interessante en relevante onderwerpen, voldoende middelen om deze gedegen uit te zoeken, goede discussie partners en altijd een plezierige en positieve samenwerking, ook als het business belang grote druk op de groep legde. Onmisbare bijdragen, zowel in materiele als immateriële zin, zijn echter ook geleverd door de vele TRD collega's. Hierbij wil ik in het bijzonder Klara en Antony danken voor het zeer nauwgezette nalezen van dit werk.

Belangrijke steun kwam tevens vanuit IMEC. Hierbij niet te vergeten (en hierbij besef ik dat deze lijst steeds onvolledig zal zijn) zijn de bijdragen van Guido Groeseneken, Ludo van de Bempt, Geert van den Bosch, Martin Creusen, Wade Zawalsky, Jan Van Houdt, Robin Degraeve, Luc Haspeslagh, Paul Hendrickx en nog zo veel meer hoogstaande wetenschappers.

Een zeer bijzonder en welgemeend woord van dank wil ik richten aan Zhichun Wang. Tijdens haar stage periode op ons bedrijf raakte de studie van plasmaschade in een echte stroomversnelling. Talrijk waren onze uitvoerige, kritische en vruchtbare discussies. In vele van de publicaties, opgenomen in dit werk, heeft zij persoonlijk een zeer belangrijke bijdrage geleverd. Hierbij wil ik ook Cora Salm niet vergeten die de eerste contacten legde tussen AMIS en UT.

De grote inspiratoren van dit proefschrift mogen ook apart geëerd worden: Fred Kuper, mijn promotor en Jurriaan Schmitz. Zonder hun stimulans en vertrouwen in mij was het waarschijnlijk nooit zover gekomen. Fred en Jurriaan, hiervoor hartelijk dank evenals voor het grote geduld dat je ten toon gespreid hebt bij de voltooiing van dit proefschrift.

Verder natuurlijk dank aan Jonas, Katrijn, Karel en Hilde, voor hun grote geduld tijdens al die uren dat 'Papa weer een verhaal aan het schrijven was' en waarbij ze de PC niet zelf konden gebruiken.

Tenslotte wil ik nog mijn ouders vermelden die het studeren en het tot stand komen van dit werk altijd door dik en dun gesteund en aangemoedigd hebben, 'Pa en Ma', bedankt!

Abstract

Over the various chapters, this thesis describes the characterization and development of a number of applications of silicon dioxides. An oxynitride is developed allowing a much higher SiGe epitaxial deposition rate in a bipolar process. Also a tunneloxide for non volatile memory application is developed and characterized. Once the oxide has been formed and defined it is exposed to the sometimes harsh environments during the following processing steps. Possible process damage due to this exposure is characterized, evaluation methods are developed, process steps are engineered for damage reduction and protection methods are developed.

Chapter 2 describes the use of a PECVD oxynitride layer instead of the nitride layer to allow a much higher deposition rate of the selective epitaxial in the heterojunction bipolar transistor interpoly dielectric layer module of a SiGe BiCMOS technology. The composition of the oxynitride layer was chosen as a function of selectivity, etch rate ratio and stress. This layer allows a significant improvement of growth rate of the selective epitaxial deposition by a factor 4 to 5. Good electrical performance and yield have been proven.

A principal problem of floating gate based non-volatile memories such as flash memories or EEPROMs is anomalous charge loss which leads to threshold voltage (V_t) shifts on a time scale of months or years at room temperature. In chapter 3 we compare different types of tunnel oxides and determine the impact on the moving bit issue with different measurement techniques. It is demonstrated that a fast Q_{BD} test of the channel tunnel oxide is generating exactly the same qualitative ranking as the moving bit test of virgin devices and the Q_{BD} test of the erase junction tunnel oxide is a measure for the moving bit behaviour of cycled devices. The superior behaviour of the wet oxides in terms of moving bits remains over the full life cycle of the memory cell.

These delicate dielectrics are easily damaged during the further processing of the device. Plasma processing is a major source of process induced damage. To evaluate plasma process induced damage induced by the different process modules on gate oxides and metal-insulator-metal capacitors, test structures with large antennas connected to the gate electrode have been developed. Chapter 4 is describing the test structures required for evaluation of plasma damage on MOS and metal-insulator-metal capacitors and HIMOS^{TM} flash memory cells devices. In the majority of cases however, protection structures need to be foreseen already in the design phase of the circuit. This chapter is providing designer guidelines, possible scenarios and required limitations for protecting semiconductor devices against plasma damage. Gate leakage measurements proved to be the most simple, fast and sensitive method.

In chapter 5 electron shading effects are discussed as a function of reactor- and transistor-type. Also a new method is proposed to study plasma process-induced latent damage on gate oxide, by using single-layer antenna and multi-layer antenna structures. With this method one can study the plasma process induced damage without suffering from artefacts induced by traditional constant current stress method. It is demonstrated that also the performance of non volatile memory cells are affected by plasma damage. For these cases the suitable protective devices and connection schemes have been developed. The correlation found between low levels of gate leakage and both HC degradation and oxide breakdown is described. This proves that plasma damage affects not only yield but also reliability and device lifetime.

Chapter 6 demonstrates that not only thin gate oxide but also dielectricums thicker than 20nm as used in metal-insulator-metal capacitors are affected by plasma damage. Also for these cases the suitable protective devices and connection schemes have been developed. Evidence is demonstrated that even after the deposition of a thick insulating inter metal dielectric over semiconductor devices, there is still an eminent risk for charging induced damage of gate oxides.

Chapter 7 describes how after a step-by-step investigation of a conventional 5 metal layer CMOS process, a number of plasma damage inducing process steps have been identified and eliminated. During the investigation also a severe case of damage to metal capacitors was discovered. It was found that ESD like discharges are causing severe defects on the metal-insulator–metal capacitors during the post processing of the devices.

Samenvatting

Dit proefschrift behandelt de ontwikkeling en de karakterisatie van een aantal toepassingen van siliciumdioxide in de halfgeleiderindustrie. Een oxynitride werd ontwikkeld dat toelaat om SiGe met een veel hogere snelheid selectief te deponeren in een hetrojunctie bipolaire technologie. Ook werd een tunneloxide in niet-vluchtig geheugen toepassing ontwikkeld en gekarakteriseerd. Nadat deze delicate diëlektrica zijn gevormd worden ze dikwijls tijdens de verdere processtappen blootgesteld aan een zeer agressieve omgeving. De mogelijke proces-schade tijdens deze verdere processtappen werd gekarakteriseerd en geëvalueerd. Waar noodzakelijk werden de processtappen aangepast om de processchade te minimaliseren. Beschermingsmethoden werden ontwikkeld en toegepast.

Hoofdstuk 2 beschrijft de ontwikkeling van een PECVD oxynitride laag ter vervanging van een siliciumnitride laag. De nieuwe laag laat toe de SiGe bij een veel hogere snelheid selectief te deponeren bij de fabricatie van een heterojunctie bipolaire transistor. De samenstelling van de oxynitride laag werd gekozen op basis van de SiGe depositieselectiviteit, de etssnelheid en de inwendige spanning van de laag. De depositie snelheid van de SiGe kon worden verhoogd met een factor 4 tot 5. Goed elektrisch gedrag en hoge opbrengst werden gedemonstreerd.

Een belangrijk probleem bij ‘floating gate’ gebaseerde niet-vluchtige geheugens zoals flash geheugens of EEPROMs is het abnormaal ladingsverlies dat leidt tot verlies van gegevens. In hoofdstuk 3 worden verscheidene types tunneloxides vergeleken en wordt hun invloed op het abnormaal ladingsverlies bestudeerd met verscheidene meettechnieken. Er wordt aangetoond dat een snelle Q_{BD} test op het kanaal-tunnel oxide dezelfde kwaliteit-volgorde geeft als het abnormale ladingsverlies test voor nieuwe geheugencellen. De Q_{BD} test op het junctie-tunneloxide geeft meer informatie over het gedrag van veel gebruikte geheugencellen. Tunnel oxide gegroeid met een nat oxidatieproces toont voor de beide gevallen de beste resultaten.

Deze delicate diëlektrica worden tijdens de verdere processtappen zeer gemakkelijk beschadigd. Plasma processen zijn de belangrijkste oorzaak van dit type beschadiging. Hoofdstuk 4 beschrijft de ontwikkeling van de teststructuren om de invloed te bepalen op transistoren, metaal-metaal capaciteiten en geheugencellen. Eveneens wordt de meest geschikte evaluatietechniek bepaald. Om plasma processchade te voorkomen dienen in de

meeste gevallen speciale beschermingsstructuren te worden gebruikt. Dit hoofdstuk biedt richtlijnen aan hoe en waar deze beschermingsstructuren te gebruiken.

In hoofdstuk 5 wordt dieper ingegaan op het mechanisme van de plasmaschade op gate- en tunneloxides. ‘Electron shading’ effecten worden besproken als functie van het transistor- en reactortype. Een nieuwe methode wordt voorgesteld om latente plasmaschade te bepalen aan de hand van meerlagige test structuren. Er wordt ook aangetoond in welke omstandigheden geheugencellen kunnen worden aangetast door plasmaschade. Geschikte beschermings-structuren worden voorgesteld om plasmaschade te voorkomen. Tevens wordt het verband beschreven dat werd gevonden tussen ‘gate’ lek metingen en ‘hot carrier’ degradatie. Hiermee wordt aangetoond dat plasmaschade niet enkel de opbrengst beïnvloedt maar tevens de betrouwbaarheid van de IC’s.

Hoofdstuk 6 toont aan dat niet enkel dunne gate oxides maar evenzeer oxides van meer dan 20nm zoals die worden gebruikt in metaal-metaal capaciteiten kunnen worden aangetast door plasmaschade. Ook hier worden geschikte beschermings-structuren voorgesteld om plasmaschade te voorkomen.
Tevens wordt aangetoond dat ook ‘dikke’ lagen diëlektrica onder de geschikte omstandigheden voldoende elektrische lading kunnen transporten om plasmaschade te induceren.

Hoofdstuk 7 beschrijft hoe na een stap-voor-stap evaluatie van een volledig CMOS proces, een aantal plasmaschade veroorzakende processtappen werden geïdentificeerd. Voor elk van deze gevallen werd het proces aangepast om de plasmaschade in voldoende mate te minimaliseren. Tevens werden grote defecten vastgesteld, veroorzaakt door een ‘ESD’-achtige ontlading van metaal –metaal capaciteiten tijdens de verdere processtappen. Ook voor deze schade werd het mechanisme bepaald en werd de oorzaak verwijderd.

Chapter 1

Introduction

1.1 Introduction

Silicon dioxide is one of the most commonly encountered substances both in daily life and in electronics manufacturing. Beach sand is mostly silicon dioxide. The whole of planar electronics processing and the modern IC industry has been made possible by the unique properties of silicon dioxide: the only native oxide of a common semiconductor which is stable in water and at elevated temperatures, an excellent electrical insulator, a good dielectric, a mask to common diffusing species and capable of forming a nearly perfect electrical interface with its substrate. Deposited silicon dioxide, mostly manufactured by chemical vapour deposition, is almost as old as thermal growth on the substrate, and has been employed in various ways in IC fabrication due to its familiarity, versatility and reliability.
Silicon dioxide has the wonderful ability to have its chemical, physical and electrical properties changed and moulded over a wide range according to the way it has been processed.
Therefore silicon dioxide has a premium role in the semiconductor industry for a very wide range of applications such as dielectric, insulator, topographical component and even as substrate for subsequent process steps. Over the various chapters, this thesis describes the characterization and development of a number of applications of silicon dioxides.
Once the oxide has been formed and defined it is exposed to the sometimes harsh environments during the following processing steps. It is a challenge to keep the delicate oxides intact all through the processing of the IC. Possible process damage due to this exposure is characterized, evaluation methods are developed, process steps are analysed for damage and protection methods are developed.

In this thesis, various aspects of dielectric engineering are considered. Chapter 2 describes the development of a PECVD oxynitride layer as a replacement for the nitride layer to allow a much higher growth rate of selective epitaxy in the heterojunction bipolar transistor. In chapter 3 we compare various types of tunnel oxides and determine the impact on anomalous charge loss with various measurement techniques. The delicate dielectrics are easily damaged during the further processing of the device. Plasma processing is a major source of process induced damage. Chapter 4 is describing the test and protection structures required for evaluation and prevention of plasma damage on MOS and metal-insulator-metal capacitors and HIMOS^{TM} flash memory cells devices. In chapter 5, the in depth investigation of the effects of plasma damage on gate oxide and tunnel oxides are described. Chapter 6 describes the effects of plasma damage on thick (>20nm) dielectrics. And finally, chapter 7 describes how after a step-by-step investigation of a conventional 5 metal layer CMOS process, a number of plasma damage inducing process steps have been identified and eliminated. For each of these subjects, some background will be given in the following paragraphs.

1.2 Bipolar Transistors

The bipolar transistor, one of the most important semiconductor devices, was invented by a research team at Bell Laboratories in 1947. It has had an unprecedented impact on the electronics industry in general and on solid-state research in particular. Prior to 1947 semiconductors were only used as thermistors, photodiodes and rectifiers. In 1948 John Bardeen and Walter Brattain announced the development of the point contact transistor [1]. In the following year William Shockley's classical paper on junction diodes and transistors was published [2].
Since then the transistor theory has been extended to include high-frequency, high-power and switching behaviour. Many breakthroughs have been made in transistor technology, particularly in the alloy-junction [3] and grown-junction techniques [4] and in zone-refining [5], diffusion [6], epitaxial [7], planar [8], beam-lead [9], ion implantation [10], lithography, and dry etching [11] technologies. These breakthroughs have helped to increase the power and frequency capabilities as well as the reliability of transistors. In addition, application of semiconductor physics, transistor theory and transistor technology has broadened our knowledge and improved other semiconductor devices as well.
Bipolar transistors are now key elements, for example, in high-speed computers, in vehicles and satellites, and in all modern communication and power systems. Many books have been written on bipolar transistor physics, design and application. Among them are standard texts by Philips [12] and Gartner [13] and a series of books by the Semiconductor Electronics Education Committee [14], Pritchard [15], Ghandhi [16], and Muller and Kamins [17].

1.2.1 Modern Silicon Bipolar Transistors

A representative for integrated silicon bipolar transistors is shown in Figure 1.1. Buried n+ subcollectors are formed by implantation of arsenic or antimony into a p substrate, followed by epitaxial growth of highly doped silicon to form the collector layer. This is followed by isolation steps based on localized oxidation of the surface.

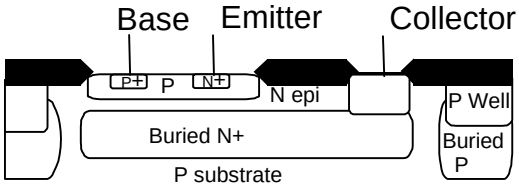


Figure 1.1:A representative for integrated silicon bipolar transistors.

The emitter is formed with the implantation of arsenic or phosphorus followed by base formation with boron ion implantation. Collector contact “plugs” may then be implanted. The structure is completed with the deposition of several layers of interconnect metal, separated by interlevel dielectric.

In recent years, advances have been made in the fabrication technology of silicon bipolar transistors, which have led to considerable reductions in vertical and lateral device dimensions, and corresponding increases in performance. This section discusses several of the common themes of present technology [18].

1.2.2 Self-Aligned Emitter and Base Contact

Significant improvements over conventional technology are possible by employing emitters and extrinsic bases diffused from polysilicon layers, which remain on the wafer forming the respective contacts. The polysilicon layers withstand high process temperatures and can be oxidized, so they lend themselves to a variety of sophisticated processing techniques. In particular, the separation between the emitter contact and the base contact can be self-aligned, using a dielectric sidewall spacer whose width can be accurately controlled in the range of several hundred nanometers. The width of the emitter is equal to the width of the opening in the base polysilicon (defined photolithographically) minus the sidewall dimensions. A representative device cross-section is shown in Figure 1.2. By repeated application of sidewall spacers, emitter widths down to 0.35 µm have been made, starting from 1 µm lithographic patterns. The emitter polysilicon can extend over a considerably wider region than the area of contact, reducing its series resistance. The base-collector junction area is kept at a minimum, typically three times the emitter area, even for the very narrow emitters. Dramatic savings in device area are made over non-self-aligned devices, which require wide separations between base and emitter to allow for photolithographic alignment tolerances [18].

1.2.3 Polysilicon Emitter

The use of polysilicon to form the emitter contact has had a major impact on transistor current gain, and therefore on vertical scaling. The value of the current gain appropriate to polysilicon depends on detailed process conditions, particularly on the amount of SiO_2 at the interface between single crystal and polysilicon. The amount of SiO_2 is typically governed by an HF dip prior to deposition of polysilicon, the subsequent thermal treatment and exposure of the resultant “hydrogen-passivated” surface, and on heat treatments after polysilicon deposition (which tend to break up the oxide). It is necessary to produce devices with relatively well-controlled values of this thickness; if the oxide is too thick, the series resistance of the emitter contact becomes excessive; if it is too thin, the current gain suffers [18].
With polysilicon, thin emitters can be produced with acceptable current gain. Thin emitters limit hole storage, and allow thin base regions to be formed with adequate control and with higher base doping. An advantage of polysilicon from the processing standpoint stems from the fact that implants of the emitter dopant (As) are made into the polysilicon material, followed by a drive-in anneal. Therefore, implant damage is kept away from the single-crystal material and by appropriate control of the anneal, extremely shallow emitters can be reproduced.
The formation of the base regions is also a critical fabrication step. The conventional method of forming the base layer is by boron implantation prior to emitter deposition. However, boron atoms have a pronounced tendency to channel, giving rise to implant distributions with deep tails. Thus, very low implant energies must be used, with the result that base charge becomes very sensitive to emitter depth [18].

1.2.4 Sidewall Contact Process

The sidewall base contact structure transistor makes innovative use of polysilicon contacts to achieve considerable improvement in device performance and to increase the possible circuit uses of the device. A sidewall base contact transistor is shown in Figure 1.2. Thick oxide layers separate the base polysilicon from the collector, so that extrinsic base collector capacitance is dramatically reduced. With appropriate doping profiles, a nearly identical geometry can be obtained for current flow downward from emitter to collector and upward from collector to emitter. High-quality upward transistors can be used in circuit applications to simplify layout [18].

1.2.5 Epitaxial Base

To overcome the problems of base layer formation by ion implantation and allow thinner base regions with controlled doping profiles, epitaxial growth of the base with in situ doping has been explored. The epitaxial process can be configured so that deposition takes place only on exposed single-crystalline regions of the silicon substrate. Alternatively, deposition can be made all across the wafer surface, although in regions covered by silicon dioxide, the silicon deposits are polycrystalline. The control achieved during epitaxy permits thinner bases and higher cutoff frequency. It is critical, however, to control the thermal budget of processing subsequent to the base growth, so that the boron dopant does not diffuse excessively. This is a particular concern during the emitter implantation step [18].

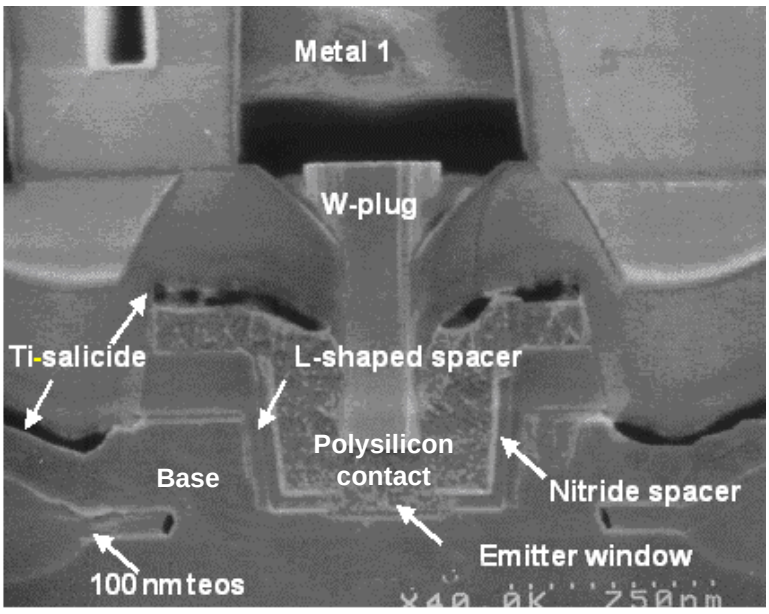


Figure 1.2: The sidewall base contact structure transistor makes innovative use of polysilicon contacts to achieve considerable improvement in device performance and to increase the possible circuit uses of the device. Thick oxide layers separate the base polysilicon from the collector, so that extrinsic base collector capacitance is dramatically reduced. Epitaxial growth of the base with in situ doping is used to overcome the problems of base layer formation by ion implantation and allow thinner base regions with controlled doping profiles.

Although polysilicon emitters and advanced processing have allowed shallower emitters and thinner base layers to be produced, there are limits to the vertical scaling achievable.

- As the base is thinned, base resistance is increased unless doping is also increased. This limit can be counteracted to a considerable extent by making narrow emitters, so that the distance between base contact and centre of the emitter stripe is kept at a minimum.
- It is necessary to avoid base punch-through at reasonable operating voltages. This requires increasing values of base doping as the structure is scaled.
- For higher base doping charge storage in the emitter is increased, cutoff frequency is decreasing and current gain drops.
- As the collector doping is increased to limit collector transit time and allow increasing current density, the breakdown voltage associated with collector avalanching decreases.

1.2.6 Silicon-Based Heterojunction bipolar transistors

The constraints on vertical scaling of bipolar transistors may be overcome if the semiconductor composition can be changed appropriately within the device. This has a critical effect on current gain. In conventional silicon transistors, considerable improvement in transistor performance can be obtained by intentionally changing the semiconductor composition. The ability to tailor this layer provides a powerful new degree of freedom in the design of bipolar devices.
The implementation of heterojunction approaches in bipolar devices was delayed for decades because of the technological problem of providing interfaces between dissimilar materials that were free from imperfections, either impurities or structural defects. Even at the present time, high-performance heterojunction bipolar transistors are limited to relatively few materials systems. The most explored systems are those involving semiconductors that have identical lattice constants (GaAlAs/GaAs and InGaAs/InAlAs/InP) and those that use thin layers that adapt to the lattice constant of the substrate (SiGe/Si) [18].

To obtain the benefits of heterojunctions combined with standard silicon bipolar technology, considerable efforts have been made to identify a suitable semiconductor to be used as base within silicon devices.

The pseudomorphic growth of a SiGe epitaxial layer results in compressive strain, low defect levels and improved electrical transport properties. By incorporating SiGe in a Si bipolar transistor the frequency limit of the devices is increased to the range normally associated with GaAs.

Experimental results have confirmed a variety of advantages of SiGe heterojunction bipolar transistors over transistors based on silicon alone, including:

- SiGe heterojunction bipolar transistors exhibit a nearly threefold improvement in cutoff frequency to over 130 GHz compared to silicon. Grading the Ge content of the base region also helps increase f_T by providing a built-in field.

- Typical characteristics of SiGe and GaAs devices are quite comparable, GaAs having the superior f_T; but SiGe enjoying lower noise parameters comparable to those of GaAs transistors.

- Early voltage increases with the use of germanium. The product of current gain and Early voltage is an important figure-of-merit for linear (analog) applications. Values of this figure-of-merit are increased by 20-100 times with the addition of germanium.

- By incorporating SiGe in a Si bipolar transistor, a higher base punch-through voltage is achieved.

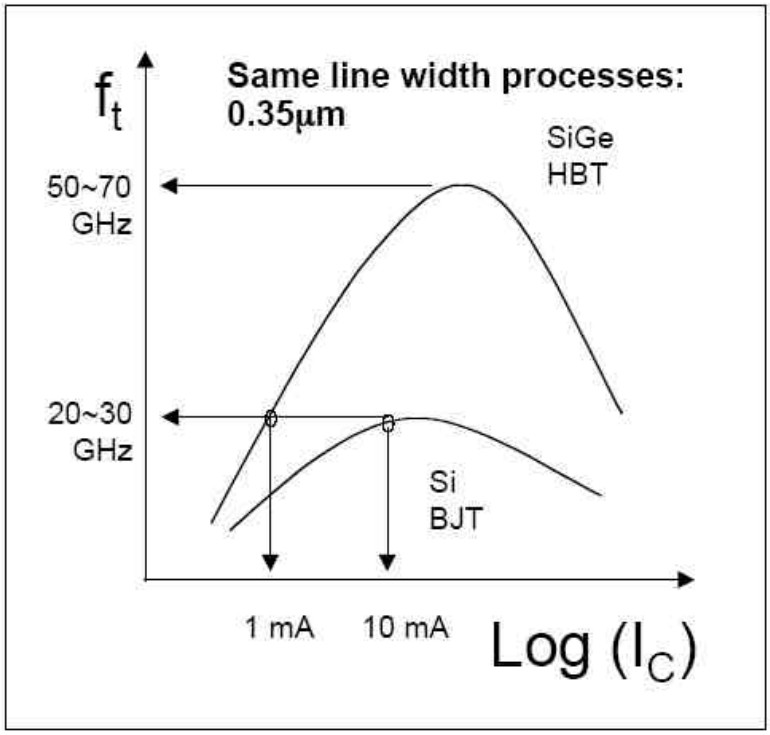


Figure 1.3: By maximizing the built-in field and reducing other delay components a significantly higher cutoff frequency f_T, up to 115 GHz, has been demonstrated. With respect to standard silicon bipolar devices, with identical processing, SiGe allows reduced current for equal performance.

The high-speed performance of SiGe heterojunction bipolar transistors is excellent. Figure 1.3 illustrates the advantage obtainable in f_T with respect to standard silicon bipolar devices with identical processing. By maximizing the built-in field and reducing other delay components, significantly higher f_T, up to 115 GHz, has been demonstrated [18].

1.2.7 SiGe Growth and HBT Fabrication

Techniques for the growth of SiGe include molecular beam epitaxy, with either solid or gaseous sources; limited reaction processing, in which epitaxy is carried out from the vapour phase on a controlled basis using heating by high intensity lamps; and ultrahigh vacuum chemical vapour deposition, among others. A key feature in the ultrahigh vacuum chemical vapour deposition technique is the virtual elimination of oxygen, which allows growth to take place at lower temperatures than in ordinary silicon growth. This technique can provide uniform growth over multiple wafers.

Several possibilities exist for the growth of the SiGe for the base. Conditions can be chosen for:

- Non-selective growth, in which SiGe is deposited everywhere (although it will be polycrystalline when covering oxide or nitride layers). After the base deposition, the SiGe is covered with polysilicon to form the emitter.

- Selective growth of the SiGe so that it will deposit only in windows opened to silicon dioxide or silicon nitride films.

Etching is an important step in the fabrication of SiGe devices and circuits. Chemical etching of SiGe is in general similar to the etching of silicon, but presents quite some challenges. Device oxides can be prepared on SiGe. SiGe alloys with Ge content below 50% can be dry oxidized for MOS devices. Ge segregation occurs at the interface causing an unacceptable level of interface states.
For these reasons a choice was made for selective growth. With this technique, epitaxial silicon is deposited only in areas with exposed silicon. The remaining area of the wafer is covered with a suitable dielectric.
So far, interpoly dielectric layer, isolating the polybase from the subsequent poly-emitter, consisted of silicon nitride. This layer withstands the wet etch during further processing, but shows a limited selectivity during the epitaxial deposition. To maintain sufficient selectivity, the deposition rate of the epitaxial layer has to be kept very low. This is done introducing a high flow of HCl in the gas chemistry. In addition, the nitride layer has often to be used in combination with an extra underlying oxide layer [3]. The factors mentioned above have a severe negative impact on manufacturability and processing cost. Chapter 2 describes the use of a PECVD oxynitride layer instead of the nitride layer to allow a much higher deposition rate of the epitaxial layer. It describes the development, properties and implementation of the oxynitride into the HBT interpoly dielectric layer module of a SiGe BiCMOS technology [18].

1.3 Characterization of Tunnel oxides for flash memories

Chapter 3 of this thesis is dedicated to the development of a reliable tunnel dielectric layer which can be applied in solid state non-volatile memories as the Electrical Erasable Programmable Read Only Memories (EEPROMs). The improvement of the electrical properties of the tunnel oxide and the development of a fast method for evaluation was the main challenge of the research described in chapter 3 of this thesis. First an introduction will be given in the basic principles of non-volatile memories, which is followed by a brief description of the mechanisms that can be used to transport charge carriers through insulators.

1.3.1 Non-volatile memories

Solid-state non-volatile memories can store information for a very long time without the need of a power supply. This is in contrast with volatile memories like DRAM and SRAM, which do need a non-interrupted power supply.
Non-volatile data storage can be performed optically or magnetically on disks or tapes. Although a lot of data can be stored, the power consumption is high during read, write and idle state and the speed is relatively low. Electrical data storage however has the advantage of operating without mechanical parts, leading to much shorter access times and therefore a higher performance. Erasable Programmable Read Only Memories (EPROM) device can be programmed by hot carrier injection into the floating gate and can be erased by ultraviolet light exposure. The programming and erasure can be carried out several times by the user [21]. However for erasing these memories have to be taken out of the system, which is a disadvantage. One Time Programmable (OTP) EPROMs are EPROMs without the UV transparent window in the package. As a result, OTP's cannot be erased and they can therefore be classified as PROM. By giving up reprogrammability, OTP can use cheaper and smaller packages [21]. The most advanced memory is the EEPROM or E^2PROM. This type of memory can be programmed and erased electrically without removing the memory chip from the system.

Nowadays a very popular memory is the so called Flash EEPROM, in which a large number of cells is erased at the same time (block or page). This is in contrast with EEPROMs where every byte can be erased separately. The most important advantage of Flash FEEROM over EEPROM is that the cell size is considerably smaller and is even approaching the cell size of DRAM and EPROM. It is mainly because of this reduced cell size and therefore lower cost and the fact that flash EEPROMS can contain huge amounts of data, that flash EEPROMs are now the dominant type of non-volatile memory [22, 23, 24].

Most non-volatile memory cell concepts are based on a standard MOS transistor. For instance the threshold voltage can be altered electrically between a low and a high value to represent logic 1's and 0's. A first method is varying the threshold voltage by changing the amount of charge that is stored in the gate dielectric, so a floating gate is not required. An example of this type of memory is known as a Metal Nitride Oxide Silicon (MNOS) transistor [40]. Charge is stored in the traps at the interface of the oxide and the nitride layer. In 1967, Kangh and Sze [19] introduced the use of a MOS transistor with a so called floating gate, which is a polycrystalline silicon layer that is completely surrounded by an insulating dielectric. This floating gate is capacitively coupled to other parts of the device. A second gate is placed on top of the floating gate, in order to control the floating gate potential through capacitive coupling.

1.3.2 Charge carrier transport through insulators

The floating gate is electrically completely isolated from the other terminals by a surrounding dielectric layer. Under normal operating conditions, no charge can flow to or from this gate, leading to a typical storage time of more than 10 years.
Charge stored on the floating gate causes a shift in the threshold voltage of the MOS transistor. The threshold voltage of the MOS transistor can be manipulated by injecting negative charge on the floating gate. This operation is known as programming. For n-channel MOS transistors it results in a higher threshold voltage, which is defined as a logic '0'. By removing the negative charge from the floating gate, which is known as erasure, the device returns to the lower threshold voltage state, defined as a logic '1'.

To change the amount of charge that is stored on the floating gate of an (E)EPROM, charge carriers have to be transported through the dielectric layer that surrounds the floating gate. Carriers can be transported through the insulator by two mechanisms. One mechanism is tunnelling. With the second mechanism, electrons can gain sufficient energy to surmount the barrier. The energy to surmount the barrier can be obtained by an electron through interaction with radiation, e.g. photons [29], Fowler-Nordheim tunnelling [20], channel hot electron injection [21- 23] and substrate hot electron injection [24 - 28].

- UV Stimulated injection
With the energy, gained through interaction with ultraviolet (UV) radiation, the electron can cross the oxide barrier [29]. If electrons are accelerated in an electric field they gain kinetic energy. This energy can be lost by collisions with lattice atoms. However, if the mean free path, i.e. the average distance between subsequent collisions is long enough [24], they can have more energy than the thermal energy corresponding to the lattice temperature. These highly energetic electrons are then called "hot" [41] and some are able to surmount the barrier. Two hot electron injection techniques are described in literature: Charnel Hot Electron injection [22] and Substrate Hot Electron injection [25].

- Channel hot electron injection
Programming by channel hot electron injection is done by making use of the pinch-off region, where a high electric field is available [42] [43] so that the electrons are accelerated from the channel into the pinch-off region. A disadvantage of this method is that the electric field in the oxide under these conditions is not directed towards the floating gate, since both the floating gate and drain potentials are providing the pinch-off condition. Only a small fraction of the hot electrons are accelerated into the desired direction, which leads to a low programming efficiency. The first EEPROM device using channel hot electron injection was presented in 1973 [22].
A third gate can be used to change the direction of the electric field and to improve the efficiency significantly. This method is known as source side hot electron injection [44, 45, 46]. A disadvantage is that the additional gate enlarges the cell and makes it more complex. By using hot electron injection the gate oxide is locally stressed in a small area near the drain junction or the source side, which is also a disadvantage compared to the more homogeneous injection with lower current density, achieved by substrate hot electron injection.

- Substrate hot electron injection

Substrate hot electron injection does not stress oxides locally, since electron injection takes place over the whole gate area. The hot electrons are generated by a vertical electric field under the gate, which accelerates the electrons in the direction of the insulator. Therefore a much more efficient programming is possible. The electrons can be provided for example by a forward biased p-n junction [25]. This method has a low efficiency since only a limited fraction of injected electrons reaches the insulator. The efficiency has been improved by the development of the Vertical Injection Punch-through based MOS (VIPMOS) EEPROM, in which a buried injector underneath the floating gate acts as a very efficient source of electrons [26].

- Fowler-Nordheim tunnelling

A very important mechanism for charge carrier transport through insulating layers is Fowler-Nordheim tunnelling [20]. Electrons have a finite probability to tunnel through a potential barrier. The tunnelling probability is determined by the height and width of the potential barrier. By applying an electric field the width can be decreased, leading to electron tunnelling from the conduction band of a semiconductor through the oxide barrier into the conduction band of an insulator.

In addition to Fowler-Nordheim tunnelling a different mechanism becomes significant for dielectric layers with a thickness below 5 nm, which is known as direct tunnelling. In that case electrons can tunnel through the barrier without using the conduction band of the insulator. However, for non-volatile memories this process is not useful since such barriers do not isolate the floating gate sufficiently.

Fowler-Nordheim tunnelling is used to erase the floating gate transistor in the High Injection (efficiency) MOS (HIMOS^{TM}) cell. It takes place from the floating gate to the drain of the MOS. The oxide layer in this overlap region can be processed as a special thin oxide, called tunnel oxide. Tunnelling in this oxide occurs when a sufficiently large potential difference is applied between the floating gate and the drain or source. When a potential difference is applied between the floating gate and the substrate, tunnelling will occur in the gate oxide. In this case the control gate will be used to bias the floating gate. The first EEPROM using tunnelling for write and erase was introduced by Intel which is the well known Flotox EEPROM [47].

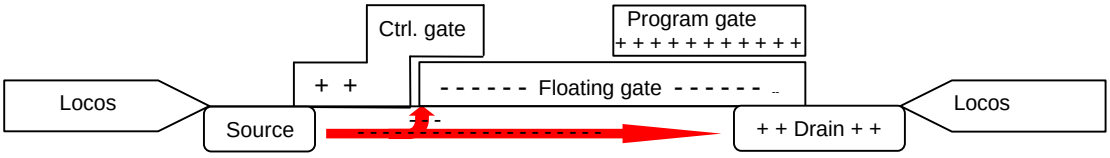
Fowler-Nordheim tunnelling deteriorates the insulating properties of the tunnel oxide which may result in a malfunctioning of the memory device, i.e. stored charge may leak away (poor data retention) or the number of program/erase cycles is very low (poor endurance).

By applying these methods, currents are forced to flow through the dielectric layer, in this case gate oxide at high electric fields. This is rather contradictory since dielectric layers are primarily known for their good isolating properties. Therefore a lot of energy is put in the development of the dielectric layers in EEPROM devices, since the electrical behaviour of the devices is dominated by the insulating properties of these layers.

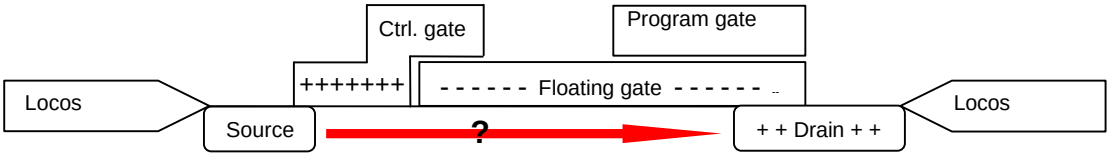
A lot of effort already has been put into improving the electrical properties of the interpoly dielectric layers in the last few decades: different types of oxide layer [30], stacks of oxide and/or nitride layers [31, 32] and different post deposition treatments [30, 33] have been investigated in order to improve the reliability of the memory cells. The improvement of the electrical properties of the tunnel oxide and the development of a fast method for evaluation was the main challenge of the research described in chapter 3 of this thesis.

1.3.3 The HIMOS^{TM} structure

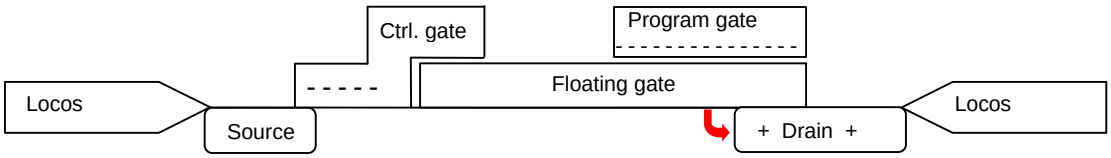
The HIMOS^{TM} structure is depicted under three different operating conditions: writing (A), reading (B) and erasing (C). The arrows indicate the direction of movement of the electrons.



1.4.A: During writing the MOS between source and drain is operated in a hot carrier regime by a limited voltage on the control gate. A high voltage is applied to the program gate to collect electrons in the floating gate.



1.4.B: During reading the MOS is partially switched on by the control gate. Depending on the charges collected in the floating gate, electrons are flowing to the drain.



1.4.C: During erasure a strong negative voltage is applied to the program gate. Due to this field the electrons collected in the floating gate are tunnelled to the drain.

The voltages necessary for operating the memory cell are listed in table 1.1.

	Write	Erase	Read
Source	0V	-	0V
Drain	3.3V	2.7V -> 5.2V	1V
Prog. gate	9V	-8.5V	0V
Contr. Gate	0.9V	-3.6V	1.7V

Table 1.1 : Terminal voltages for operating the memory cell of figure A, B and C.

Besides the concepts in which only single HIMOS^{TM} structures are used, this structure is also applicable in large memory arrays. Arranging the cells in an array gives the possibility to share terminals of cells and to connect many cells by the same wires. In this manner, the total chip area can be reduced. The source/injector terminals can be shared by the cells next to each other. The erase gates are connected to each other; therefore a complete row of cells is erased at the same time. The effective size of every cell is 135 μm^2.

During writing and especially during erasing, a current is forced through the tunnel oxide in between the substrate and the floating gate. This degrades the insulating properties of the tunnel oxide. When the tunnel oxide is degraded too far, charges leak away from the floating gate to the substrate. This anomalous charge loss leads to threshold voltage shifts on a time scale of months or years at room temperature. We refer to the bits showing such low-temperature long-term V_t shifts as anomalous SILC or moving bits. The number of moving bits is greatly affected by the nature and properties of the tunnel oxide that is used in the non-volatile memory technology. Chapter 3 compares different types of tunnel oxides and determines the impact on the moving bit issue. Together with the impact of the different types of oxide, different measurement methods are compared and evaluated.

1.4 Plasma damage

Besides the formation of a dielectric meeting all the required specifications of performance and reliability, it is a major challenge to keep this dielectric intact throughout the full manufacturing process of the integrated circuit. Especially for gate oxides, a major issue is the use of plasma processing during the manufacturing process. The plasma process ambient is a very harsh environment which influences the yield and reliability of finished semiconductor devices.

In the 50 years since the invention of transistor silicon integrated circuit technology has made astonishing advances. A key factor that makes these advances possible is the ability to have precise control on material properties and physical dimensions. The introduction of plasma processing in pattern transfer and in thin film deposition is critical for enabling further downscaling in the semiconductor technology. In state of the art silicon integrated circuit manufacturing process, plasma is used in more than 20 different critical steps.
Plasma is sometimes called the fourth state of matter (other than gas, liquid and solid). It is a mixture of ions (positive and negative), electrons and neutrals in a quasi-neutral gaseous steady state, sustained by an energy source that balances the loss of charged particles. It is a very harsh environment for the delicate integrated circuits. Highly energetic particles such as ions, electrons and photons bombard the surface of the wafer continuously. These bombardments can cause all kinds of damage to the silicon devices that make up the integrated circuit.

Very early on in the introduction of plasma processing in silicon IC manufacturing, people became aware of the potential danger due to plasma damage to the devices. Much effort had been devoted to understand the damage mechanisms and to find ways to avoid or fix them. There is a huge body of literature on the subject. Most of the damage studies focused on physical damage to crystalline silicon by energetic ions and electrical damage to SiO_2 and its interface by energetic photons.
More recently a different kind of plasma damage has become important. This is charging damage. The first reported observation of plasma charging damage was in 1984 [63]. A high electric field apparently developed across the gate-oxide of a metal-oxide-semiconductor-field-effect transistor during plasma processing. This high electric field leads to gate-oxide breakdown or wear out (lifetime shortening). It was quite unclear at first how such an electric field could develop.

In more recent years, plasma process induced damage has become a major industry-wide problem requiring an interdisciplinary approach. In modern VLSI manufacturing extremely tight control on every production step is required. Plasma processing steps are no exception. Plasma processing tools have been developed at great expense to achieve excellent uniformity across the wafer. Gone are old tools such as the barrel etcher with poor plasma uniformity. Yet, in modern IC manufacturing plasma charging damage is more prevalent: the reason for this increase in incidence of damage is due mostly to the continued scaling down of gate-oxide thickness. As with many other problems in the integrated circuit industry, the impact of plasma process induced damage on real products has been aggravated by the relentless scaling of device and interconnect dimensions. Currently, interconnecting 10^7-10^8 transistors constitutes a major technological challenge requiring the use of multi-level metallization schemes. To achieve this goal, high density plasma enhanced deposition and etching techniques have to be used. However, the use of high density plasma reactors and the accompanying higher density of charging species (10^{12} ions/cm^3) impose a major yield and reliability threat. This negative effect is further aggravated by the use of ever thinner gate oxides. The combination of high-density plasmas and thinner gate oxides therefore constitutes one of the major threats for the operational lifetime of integrated circuits [48] [49]. A thinner gate-oxide requires a lower voltage to support Fowler-Nordheim tunnelling: it is more vulnerable to plasma non-uniformity.
On the other hand, if one were to calculate the plasma potential variation across the wafer based on the measured ion current or etch rate in most modern plasma processing tools, one would have found that the voltage variations are often lower than what is needed to cause damage. Clearly the actual floating potential variation across the wafer must be higher than what is suggested from these measurements. It is, of course, quite possible that the measurement method is inadequate and that the real variation in plasma potential is larger. Still, apart from plasma non-uniformity, other mechanisms generating plasma damage should be considered. These other possibilities will be explored in chapter 4, 5 and 6 of this work.

1.4.1 Plasma damage mechanisms

The three main plasma process induced damage models are based on surface charging, the non-uniformity of the plasma and on the electron shading effect [50].

- Surface charging

Some processes are capable of generating a considerable quantity of charges on a surface. These charges can be generated simply by friction. Under the right conditions, the charging of a surface can also be the effect of a plasma. Figure 1.5 shows the mapping of the charges collected on a wafer surface by a high speed water spray on the centre of a silicon oxide wafer surface. When the silicon substrate is grounded, the full electric field is distributed over the thin oxide. When the electric field exceeds the field that the oxide can support, a destructive Fowler-Nordheim current starts to flow through the oxide.
Similarly, voltages built up by friction can easily reach the range of several thousands of volts and can be very destructive. These phenomena are grouped under the name electrostatic discharge.

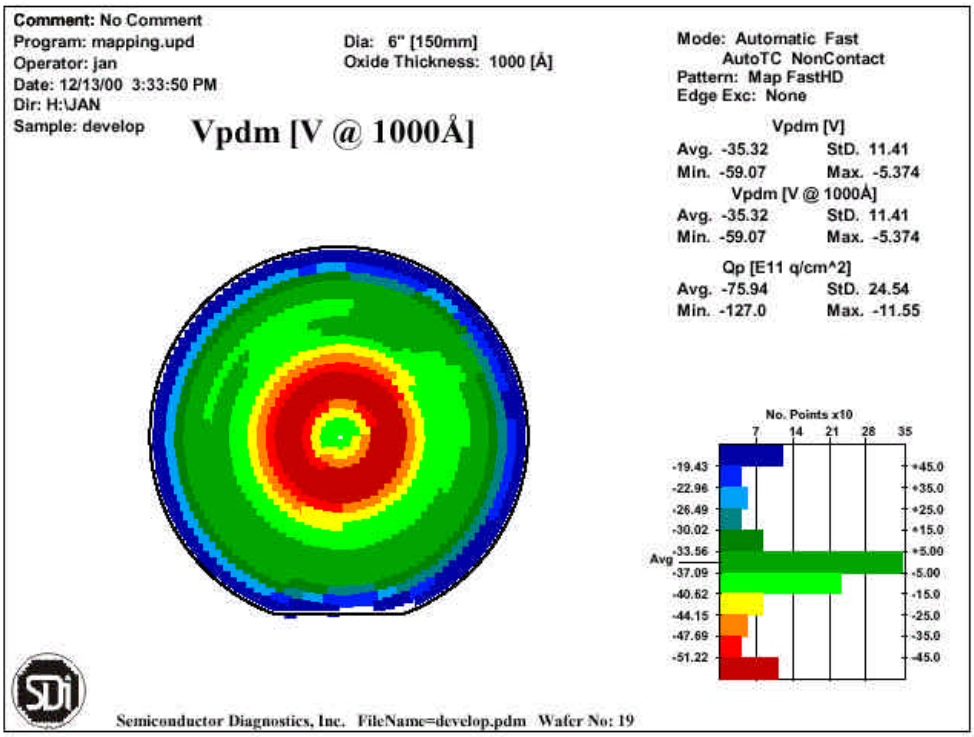


Figure 1.5: For mapping of the charges collected on a wafer surface, a plasma damage monitoring tool manufactured by Semiconductor Diagnostics Inc. was used. This technique measures the build-up of the deposited plasma charges during plasma processing on top of an unpatterned SiO_2 film. In this example a high speed water spray generated –59 V charging in the centre ring of the wafer.

- Plasma non-uniformity

Non-uniform plasma discharge causes a local imbalance in ion and electron currents on the wafer surface. The edges of a wafer surface can be charged opposite compared to the charging in the centre surface of the wafer. The bulk of the wafer is conductive and can support the current required for balancing the non-uniform charging of the wafer surface.
When etching a conductive pattern connected to a gate oxide, the charging of the conductor continues until the maximum field is reached that can be supported by the thin gate oxide connected to the conductive layer. When the electric field is high enough, destructive Fowler-Nordheim current starts to flow through the oxide as depicted in Figure 1.6.

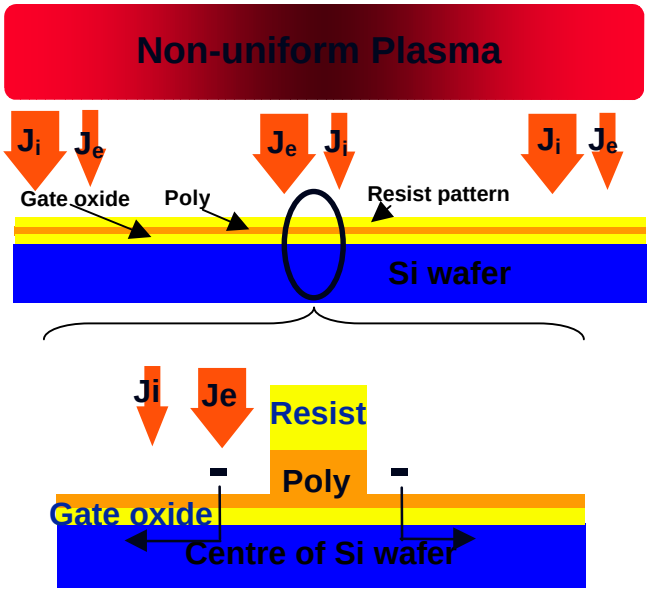


Figure 1.6 Non-uniform plasma discharge causes a local imbalance in ion and electron currents. The top picture shows a plasma inducing more negative charges by a negative electron current J_e in the centre of a wafer. The edges of the wafer are objected to a positive ion current J_i. The bottom picture shows how due the dominating negative electron current, the wafer centre is charged negatively and an electric field builds op over the gate oxide. Plasma damage occurs once the field built up over the gate oxide exceeds the voltage that can be supported by the gate oxide. At that moment a FN current (depicted by the arrows) flows through gate oxide and induces damage.

- Electron shading

Even in uniform plasmas charging can occur. Most important is the electron shading effect in which negative charges on the photoresist prevent electrons from penetrating into the trenches being etched. As depicted in figure 1.7 this causes positive charging of the gate electrode in case of high aspect ratio patterns even for perfectly uniform plasmas [50],[51]. This phenomenon occurs for any etch/ash with insulator over conductor as is the case during poly or metal patterning.

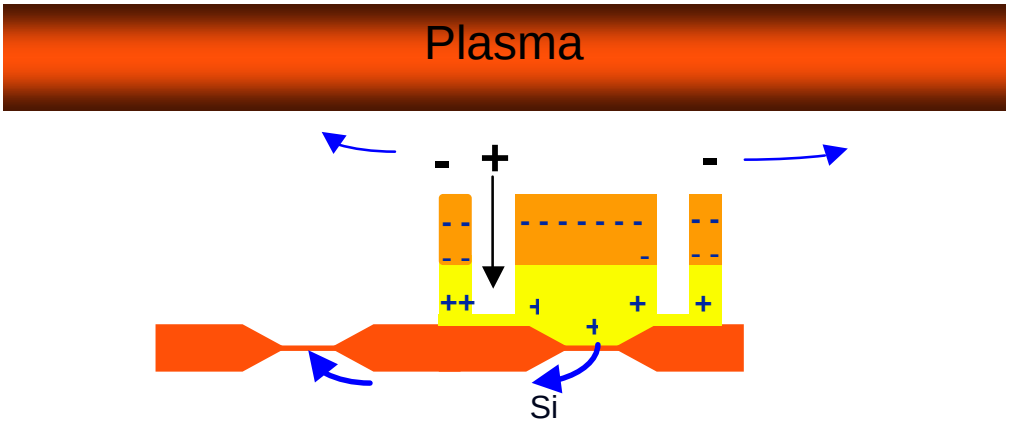


Figure 1.7 When an antenna has a dense resist pattern, the resist charges negatively and “shades” electrons from entering the antenna. This is causing positive charging of the conductor: electrons are deflected by the negative charge built up in the resist as depicted by the top arrows. Mainly positive ions can still reach the conductor through the trenches in the resist. The positive charged conductor induces an electric field over the gate oxide: a FN tunnelling current starts to flow through the gate oxide (as depicted by the bottom arrows) and plasma damage is occurring.

1.4.2 Metal-insulator-metal capacitors

Apart from the increasing use of high density plasmas in combination with very thin gate oxides, CMOS devices are being extended to RF applications, for the fast expanding field of telecommunications. Typical for the BiCMOS-RF technologies is the use of inductors in combination with capacitors. High performance mixed signal and RF circuits require integrated capacitors with low voltage coefficients, good capacitor matching, precision control of capacitor values and low parasitic capacitance along with high reliability and low defect density. Conventional high-density capacitors such as poly-substrate, see for example [52], poly-poly, see for example [53] and metal-poly [54] [55] structures suffer from high voltage coefficients due to voltage induced depletion effects as well as higher parasitic capacitance due to their proximity to the substrate. Metal-insulator-metal (MIM) capacitors

have low parasitic capacitance, especially when fabricated on metal 2 or higher, as well as high quality factor for RF circuit applications [56]. High-density metal-insulator-metal capacitors with capacitance densities of 1.0 to 2.0 fF/µm² using PECVD nitride dielectric have been integrated into the backend metallization layers of a CMOS or BiCMOS process flow. Capacitor breakdown voltage, linearity and reliability meet mixed-signal circuit requirements. Less than 0.5 defects/cm² has been achieved for 2.0 fF/µm² capacitance density. A quality factor larger than 10 at 1 and 2GHz was measured indicating its usefulness in RF applications. As depicted in figure 1.8 these capacitors are made as metal-insulator-metal capacitors. The application of the thin dielectric together with the metal processing in a high density plasma environment makes these metal-insulator-metal capacitors very vulnerable to plasma damage [56]. In this thesis, special attention is paid to the mechanisms on how plasma damage affects these devices and how they can be protected.

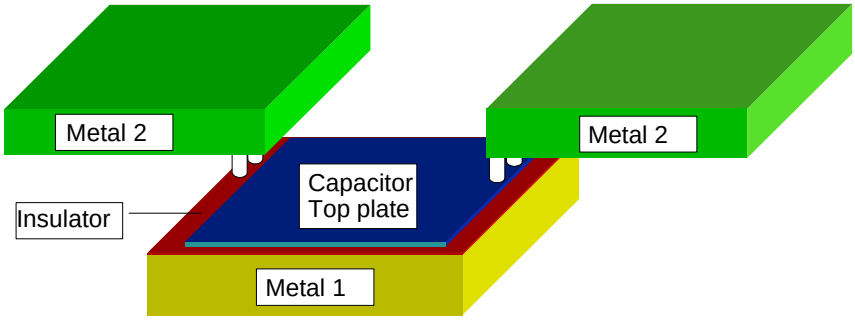


Figure 1.8. BiCMOS RF technologies use inductors in combination with capacitors. These capacitors are made as metal-insulator-metal capacitors.

1.4.3 Evaluation and prevention of plasma damage

To evaluate the plasma process induced damage caused by the different process modules on gate oxides and metal-insulator-metal capacitors, test structures with large antennas connected to the gate electrode have been developed. Typically, large plate and comb antennas are used for the poly and metal patterning processes whereas antennas with a large number of contact and via holes are used to assess the plasma process induced damage introduced during contact and via processing. Besides that, electron shading effects are studied by varying the spacing of the comb antennas. However, also more dedicated test structures have been designed to achieve a better understanding of the more fundamental plasma process induced damage aspects, such as timing and pattern density related aspects [57] [58]. Chapter 4 of this work describes the test structures required for evaluation of plasma damage on MOS and metal-insulator-metal capacitors. As a special case of CMOS, test structures are developed to investigate the impact of plasma damage on HIMOS^{TM} flash memory cells.
To assess the impact of topography, evaluation methods are required that allow qualification of processes in an efficient but reliable way using the antenna structures. Chapter 4 of this work compares evaluation techniques such as voltage breakdown, charge pumping, gate

leakage and threshold voltage measurements. The measurement methods are compared in terms of sensitivity, speed and ease of use.

Non-contact surface charge measurements provide a fast and indirect way to detect plasma process induced damage. This type of measurement is typically used to monitor the plasma (non-) uniformity by measuring the surface charge on non-patterned oxide wafers. After the plasma process, the potential on the oxide surface can be measured by means of a plasma damage monitoring tool (Semiconductor Diagnostics Inc.). The absolute potential (Vpdm) and the potential difference (ΔVpdm) reflect the potential charging contribution of final stages of the deposition process. This technique, however, does not detect electron shading effects [59]. In chapter 5 the application of this measurement in order to detect possible sources of plasma damage is described.

In the majority of cases, however, protection structures need to be foreseen already in the design phase of the circuit [60]. Chapter 5 provides design guidelines, possible scenarios and required limitations for protecting semiconductor devices against plasma damage. In other cases, plasma damage can be prevented by the development of the proper process. Chapter 6 describes how process optimization can lead to plasma damage reduction.

1.4.4 Impact of plasma damage on reliability

Apart from the impact on the gate oxides, plasma damage can have impact on other reliability parameters. In chapter 7 of this work it is the intention to correlate two reliability parameters: hot-carrier degradation behaviour and breakdown of the gate oxide to the accumulated damage during plasma processing.
Hot-carrier stress typically occurs under high silicon and relatively low oxide field conditions. In the technologies that are considered here, the dominant degradation mechanism is interface trap generation in NMOS and electron trapping in PMOS [61]. This is clearly different from the high-field oxide degradation case where the dominant degradation mechanism finally leading to oxide breakdown is bulk trap generation [62]. The higher density of bulk oxide electron traps in devices that suffered from plasma charging damage might influence the trapping behaviour under hot-carrier stress conditions. For this reason it is worthwhile to investigate if and to what extent hot-carrier degradation rate is influenced by the accumulated damage during plasma processing.
If a clear correlation could be found between gate leakage and both HC degradation and oxide breakdown, it would prove that plasma damage affects not only yield but also reliability and lifetime of the devices. On top it would suggest that gate leakage current can be used to monitor device reliability under plasma stress, thereby saving a lot of measurement time and prevent possible reliability problems in the field.

1.5 References

[1] J. Bardeen and W. H Brattain, “The Transistor, A Semiconductor Triode”, Phys. Rev., 74, 230, 1948.

[2] W. Shockley, “The Theory of p-n Junctions in Semiconductors and p-n Junction Transistors”, Bell Syst. Tech. J., 28, 435, 1949.

[3] R.N. Hall and W. C. Dunap, “p-n Junctions Prepared by Impurity Diffusion”, Phys. Rev., 80, 467, 1950.

[4] G.K. Teal, M. Sparks, and E. Buehlor, “Growth of Germanium Single Crystals Containing p-n Junctions”, Phys. Rev., 81, 637, 1951.

[5] W.H. Pfann, “Principles of Zone- Refining”, Trans. AIME, 194, 747, 1952.

[6] M. Tanenbaum and D. E. Thomas, “Diffused Emitter and Base silicon Transistor”, Bell Syst. Tech. J., 35, 1, 1956.

[7] H.C. Theuerer, J. L Kleimack, H. H. Loar, and H. Christenson, “Epitaxial Diffused Transistors”, Proc. IRE, 48, 1642, 1960.

[8] J.A. Hoerni, “Planar Silicon Transistor and Diodes”, IRE Electron Devices Meet., Washington D.C., 1960.

[9] M.P. Lepselter, “Beam-Lead Technology”, Bell Syst. Tech. J., 45, 233, 1966.

[10] W. Shockley, U.S. Patent 2,787,564 (1954), For a view on ion implantation, see J. F. Gibbons, “Ion Implantation in Semiconductors-Part 1, Range Distribution Theory and Experiments”, Proc. IEEE, 56, 295, 1968.

[11] E.F. Labuda and J.T. C1emens, “Integrated Circuit Technology”, in R. E. Kirk and D. F. Othmer, Eds., Encyclopedia of Chemical Technology, Wiley, New York, 1980.

[12] A.B. Phillips, Transistor Engineering, McGraw-Hill, New York, 1962.

[13] W.W. Gartner, Transistors, Principle, Design and Application, D. Van Nostrand, Princeton, N.J., 1960.

[14] SEEC (Semiconductor Electronics Education Committee), 4 vols.
(1) R.B. Adler, A.C. Smith, and R.L. Longini, Introduction to Semiconductor Physics, SEEC Vol. 1, Wiley, New York, 1966.
(2) P.E. Gray, D. DeWitt, A.R. Boothroyd and E.P. Gibbons, Physical Electronics and Circuit Models of Transistors, SEEC Vol. 2, Wiley, New York, 1966.
(3) C.L. Searle, A.R. Boothroyd, E.J. Angelo, P.E. Gray, and D.G. Pederson, Elementary Circuit Properties of transistors, SEEC, Vol. 3, Wiley, New York, 1966.
(4) R.D. Thornton, D. DeWitt, E.R. Chenette, and P.E. Gray, Characteristics and Limitations of Transistors, SEEC, Vol. 4, Wiley, New York, 1966.

[15] R. L. Pritchard, Electrical Characteristics of Transistors, McGraw-Hill, New York, 1967.

[16] S.K. Ghandhi, Semiconductor Power Devices, Wiley, New York, 1977.

[17] R.S. Muller and T.I. Kamins, Device Electronics for integrated Circuits, Wiley, New York, 1977.

[18] S.M. Sze, Modern Semiconductor Device Physics, Wiley, New York 1998.

[19] D. Kangh and S.M. Sze, "A floating gate and its application to memory devices", in Bell Syst. Tech. J., Vol. 46, pp. 1238, 1967.

[20] M. Lenzlinger, E. H. Snow, "Fowler-Nordheim tunnelling into thermally grown SiO_2", Journal of Applied Physics, Vol. 40, No. 1, pp. 278-283, 1969.

[21] C.N. Berglund, R.J. Powell, "Photoinjection into $Si0_2$:Eletctron Scattering in the Image Force Potential Well", Journal of Applied Physics, Vol. 42, No. 2, pp. 573-579, 1971.

[22] H.C. Card and A.G. Worall, "Reversible floating-gate memory", Journal of Applied Physics, Vol. 44, No. 5, pp. 2326-2330, 1973.

[23] G. Perlegos, S. Pathak, A. Renninger, W. Johnson, M. Holler, I. Skupnak, M. Reitsma, C. Kuhn, "A 64K EPROM using scaled MOS Technology", ISSCC Technical Digest, pp. 142-143, 1980.

[24] T.H. Ning, C.M. Osburn, H.N. Yu, "Emission Probability of Hot Electrons from Silicon into Silicon Dioxide", Journal of Applied Physics Vol. 48, No. 1, pp. 286-293, 1977.

[25] B. Eitan, J.L. McCreary, D. Amrany, J. Shappir, "Substrate Hot-electron EPROM", IEEE Transaction on Electron Devices Vol. 31, No. 7, pp. 934-942, 1984.

[26] R.C. M. Wijburg, G.J. Hemink, J. Middelhoek, H. Wallinga and T.J. Mouthaan, "VIPMOS a novel buried injector structure for EPROM applications", IEEE Transactions on Electron Devices, Vol. 38, No. 1, pp. 111-120, 1991.

[27] G.J. Hemink, R.C.M. Wijburg, R. Cuppens, J. Middelhoek, "High Efficiency Hot electron Injection for EEPROM Applications Using a Buried-Injector", Extended abstracts 21st Conference on Solid State devices and Materials (SSDM), pp. 133-136, 1989.

[28] T.H. Ning, "Hot-Electron Emission from Silicon into Silicon Dioxide", Solid State Electronics, Vol. 21, pp. 273-282, 1978.

[29] G. van Steenwijk, K. Hoen, H. Wallinga, "A Nonvolatile Analog Programmable Voltage source Using the VIPMOS EEPROM Structure", IEEE Journal of Solid State Circuits, Vol. 28, No. 7, pp. 784-788, 1993.

[30] J. Lee, I.C. Chen, C. Hu, "Comparison Between CVD and Thermal Oxide Dielectric Integrity", IEEE Electron Device Letters, Vol. 7, pp.506-509, 1986.

[31] R. Moazzami, C. Hu, "A high-Quality Stacked Thermal/LPCVD gate Oxide Technology for ULSI", IEEE Electron Device Letters, Vol. 14, No. 2, pp. 72-73, 1993.

[32] T. Watanabe, A. Menjoh, M. Ishikawa, I. Kumagai, “Stacked $SiO_2/Si_3N_4/SiO_2$ dielectric layer for reliable memory capacitor”, in Proceedings international Electron Device Meeting (IEDM), pp. 173-177, 1984.

[33] J.H. Klootwijk, H. van Kranenburg, M.H.H. Weusthof, P.H. Woerlee, H. Wallinga, “RTP annealings for high quality LPCVD inter-polysilicon dielectric layers”, Microelectronics & Reliability Journal, No. 38, pp. 277-280, 1998.

[34] D.W. Greve, “Programming Mechanism of Polysilicon Resistor Fuses”, IEEE Transaction an Electron Devices, Vol. 29, No. 4, pp. 719-724, 1982.

[35] P. Thai, S.C. Chang, M.C. Yang, “A 35 ns 128K Fusible Bipolar PROM”, ISSCC Technical Digest, pp.44-45, 1986.

[36] K. Alexander, J. Hicks, T. Soukop, “Moisture Resistive, U.V. Transmissive Passivation for Plastic Encapsulated EPROM device”, Proceedings International Reliability Physics Symposium (IRPS), pp. 218-222, 1984.

[37] G. Grisenza, R. Annunziata, E. Camerlenghi, P. Cappelletti, “Non-volatileile memories Issues, Challenges and Trends for the 2000's scenario”, in Proceedings European Solid-State Device Research Conference (ESSDERC'96), pp.121-130, 1996.

[38] T.S. Jung, Y.J. Choi, K.D. Suh, B.H. Suh, J.K. Kim, Y.H Lim, Y.N. Koh, J.W. Park, K.J. Lee, J.H. Park, K.T. Park, J.R. Kim, J.H. Yi, H.K. Lim, “A 117mm2 3.3-V Only 128 Mb Multilevel Flash Memory for Mass Storage applications”, IEEE Journal of Solid State Circuits, Vol. 31 No. 11, pp. 1575-1584, 1996.

[39] Y. Iwata, M. Momodomi, T. Tanaka, H. Oodira, Y Itoh, R. Nakayama, R. Kirisawa, S. Aritome, T. Endoh, R. Shirota, K. Ohuchi, F. Masuoka, “A High-Density NAND EEPROM with Block-Page Programming for Microcomputer Applications”, IEEE Journal of Solid-State Circuits, Vol. 25, No. 2, pp. 447-424, 1990.

[40] J.J. Chang, “Theory of MNOS Memory Transistor”, IEEE Transactions of Electron Devices, Vol. 24, No. 5, pp. 511-518, 1977.

[41] R.S. Muller and T.J. Kamins, “Device Electronics for Integrated Circuits”, second edition, John Wiley and Sons, ISBN 0-471-84309-1, 1986.

[42] G. Yaron, W.A. Lukaszek arid D. Frohnan-Bentchkowsky, “E2FAMOS-An Electrical Erasable Reprogrammable Charge Storage Device”, IEEE transactions on Electron Devices, Vol. 26, No. 11 pp. 1754-l759, 1979.

[43] K. Yoshkawa, S. Mari, Y Kanako, Y. Oshima, N. Arai and E Sakagaini, “0.6 µm EPROM cell design based on a new scaling scenario”, Proceedings international Electron Device Meeting (IEDM), pp. 587-590, 1989.

[44] A.T. Wu, T.Y. Chan, P.K. Ko and C. Hu, “A Novel High Speed 5-volt programming EPROM structure with source side injection”, Proceedings International Electron Device Meeting (IEDM), pp.584-587, 1986.

[45] K. Naruke, S. Yamada, B. Obi, S. Taguchi and M. Wada, “A new flash erase EEPROM cell with a sidewall select-gate on its source side”, Proceedings International Electron Device Meeting (IEDM), pp. 603-606, 1989.

[46] J. van Houdt, L. Haspeslagh, D. Wellekens L. Deferm, G. Groeseneken and H.E. Maes, “HIMOS - A High Efficiency Flash E2PROM Cell for Embedded Memory Applications”, IEEE Transactions on Electron Devices, Vol. 40, No.12, pp. 2255-2263, 1993.

[47] W.S. Johnson, G. Perlegos, A. Renninger, G. Kahn, T.R. Ranganath, “A 16 Kb Electrically Erasable Non-vo1ati1e Memory”, Digest Technical Papers IEEE international Solid-State Circuits Conference (ISSCC), pp. 152-153, 1980.

[48] Proceedings International Symposia on Plasma Process-Induced Damage (P2ID) 1-3, 1996-1998.

[49] R.W. Thomas, “An analysis of the quality and reliability supplement to the SIA roadmap”, Microelectronics Reliability 38, 1998, pp. 861-868.

[50] K. Hashimoto, “New phenomena of charge damage in plasma etching: heavy damage only through dense-line antenna,” Jpn. J. Appl. Phys., part 1, vol. 32, pp. 6109-6113, 1993.

[51] K. Hashimoto, “Charge damage caused by electron shading effect,” Jpn. J. Appl. Phys., part 1, vol. 33, pp. 6013-6018, 1994.

[52] J.L. McCreary, “Matching properties, and voltage and temperature dependence of MOS capacitors”, IEEE J. Solid State Circuits, IEEF- J. Solid State Circuits, vol. SC-16, pp. 608-616, Dec., 1981.

[53] Sing-Pin Tay and J.P. Ellui, “Thin gate and analog capacitor dielectrics for submicron device fabrication”, J. Electronic Materials, vol. 21, no. 1, pp. 45-55, 1992.

[54] C. Kaya, H. Tigelaar, J. Paterson, M. de Wit, J. Fattaruso, S. Kiriakai, K.S. Tan and F. Tsay, “ Polycide /metal capacitors for high precision A/D converters”, Proceedings of IEDM, pp. 782-785, 1988.

[55] A. Yin, J. White, A. Karroy and C. Hu, “Integration of polycide/metal capacitors in advanced device fabrication”, 5th ICSICT Proceedings, pp. 131-134, 1998.

[56] K. Stem, J. Koeis, G. Hueekel, E. Eld, T. Bartush, R. Groves, N. Greco, D. Harame and T. Tewksbury, “High reliability metal insulator metal capacitors for silicon germanium analog, applications”, Proceedings of BCTM,pp.191-194, 1997.

[57] S. Fang and S. Murakawa, “A new model for thin oxide degradation from wafer charging in plasma etching,” in IEDM Tech. Dig., pp. 61-64, 1992.

[58] T. Yamada, K. Eriguchi, Y. Kosaka, and K. Hatada, “Impacts of antenna layout enhanced charging damage on MOSFET reliability and performance,” in IEDM Tech. Dig., pp. 727-730, 1996.

[59] G. Horner, I.J. Gupta, “RaP2ID detection of charging damage with non-contact electrical analysis,” in Proceedings International Symposium on Plasma Process-Induced Damage (P2ID) 3, pp. 144-147, 1998.

[60] S. Krishnan, A. Amerasekera, S. Rangan and S. Aur, “Antenna device reliability for ULSI processing,” in IEDM Tech. Dig., pp. 601-604, 1998.

[61] P. Heremans et al., “The mechanisms of hot carrier degradation”, in “Hot carrier design considerations for MOS devices and circuits”, Ed. C.T. Wang, Van Nostrand-Reinhold, 1992.

[62] R. Degraeve et al., “Degradation and breakdown in thin oxide layers: mechanisms, models and reliability prediction”, Microelectronics Reliability 39, pp. 1445, 1999.

[63] T. Watanabe, Y. Yoshida, “Dielectric breakdown of gate insulator due to reactive ion etching”, Solid State Technol., pp. 263, 1984.

Chapter 2

Oxynitride development for selective epitaxy in HBTs

The results in this chapter have partially been published in:

J. Ackaert, P. Chevalier, J.-L. Lohéac, H. Ziad, E. De Backer, M.Tack, "Use of Oxynitride Dielectric to Maximise the Growth Rate of Selective Epitaxial Base Layer in a Self-Aligned Double-Polysilicon SiGe Bipolar Transistors", 32nd European Solid-State Device Research Conference (ESSDERC), pp. 267-270, 2002.

Part of this work was funded by IWT and MEDEA T204 project, IMEC provided technical support.

SiGe BiCMOS technology offers high-performance together with high-integration capabilities for low-noise high-speed and low-power wireless and optical communications [1]. The epitaxial growth of the SiGe base layer can be done in a non-selective or a selective way. The selective method has the advantage of producing an almost ideal self-aligned structure with reduced collector-base parasitics and then improved f_{MAX} over f_T ratio of the heterojunction bipolar transistor device [2]. In the previous state-of-the-art the so-called interpoly dielectric layer, isolating the polybase from the subsequent poly-emitter, consisted of silicon nitride. This layer withstands the wet etch during further processing, but selective epitaxial deposition of the SiGe can only be achieved when a very low deposition rate is used. If the deposition rate is increased, then the SiGe statrts to deposit as a polycrystalline layer on the nitride layer. To maintain sufficient selectivity, the deposition rate of the epitaxial layer must be kept very low. This is done by introducing a high flow of HCl in the gas chemistry. In addition, the nitride layer often has to be used in combination with an extra underlying oxide layer [3]. The factors mentioned above have a severe negative impact on manufacturability and processing cost. This chapter describes the use of a PECVD oxynitride layer instead of the nitride layer to allow a much higher deposition rate of the epitaxial layer. This chapter describes the development, properties and implementation of the oxynitride into the heterojunction bipolar transistor interpoly dielectric layer module of a SiGe BiCMOS technology.

In this chapter, "nitride" and "silicon nitride" are referring to Si_3N_4, "oxynitride" is referring to SiO_xN_y and "poly" is referring to polycrystalline silicon.

2.1 Process description

The process starts with the implantation of n+ and p+ buried layers in a p-substrate followed by the epitaxial growth of a 1 µm n-type bipolar collector layer. After definition of active areas with poly-buffered LOCOS isolation, CMOS retrograde wells and bipolar collector sinker are implanted. This is followed by the gate oxide growth and the polygate deposition and patterning. A standard spacer module allows the implantation of the CMOS LDD's. Then the n+ source/drain implantation of the NMOS is performed. Compared to the Si-BiCMOS technology the p+ source/drain implant of the PMOS is moved beyond the fabrication of the bipolar transistor. This is done to prevent severe p+ boron penetration through the gate oxide during the H_2 bake performed before the bipolar base epitaxy growth. This led to positive V_t's. Next the bipolar fabrication starts by opening the oxide/nitride stack deposited to protect the CMOS.
A 100 nm oxide (TEOS) sacrificial layer for the epitaxy of the SiGe base is deposited and densified. The 200 nm polybase layer deposited on top is first lowly boron doped to form 1 kΩ/square high ohmic poly-resistor and then masked to receive the implantation needed for the bipolar base lead [4].
An oxynitride layer is deposited by PECVD and the oxy-nitride/polybase stack is patterned and etched, stopping on the TEOS layer to open the internal base region. Phosphorous implantation of a Selective Implanted Collector (SIC) follows. Nitride sidewalls are formed to protect the polybase during the selective epitaxy. Then the TEOS layer is laterally wet etched under the polybase. After an HF-dip the selective growth of the base is performed in an ASM Epsilon 2000 RP-CVD epi reactor. This process sequence is illustrated in Figure 2.1.

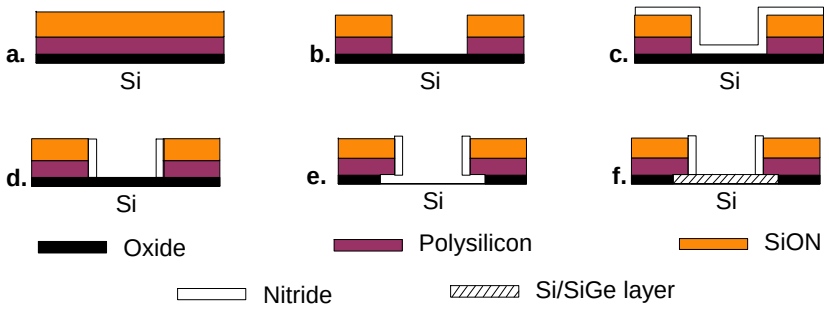


Figure 2.1: Process steps used in the manufacturing of the bipolar transistor. An oxynitride layer is deposited by PECVD (a) and the oxynitride/polybase stack is patterned and etched, stopping on the TEOS layer to open the internal base region (b). Nitride sidewalls are formed (c-d), then the TEOS layer is laterally wet etched under the polybase (e). After an HF-dip the selective SiGe deposition of the base is performed (f).

A H_2 bake at 850°C is applied prior to the deposition. The process gases used are H_2 as carrier gas, $SiCl_2H_2$ as silicon source, GeH_4 as Ge source, B_2H_6 as doping gas and HCl to obtain selectivity. The SiGe layer is grown with a linearly graded profile with a targeted maximum Ge fraction of 15 %, followed by a Si capping layer of 30 nm. The top 55 nm of the epitaxial layer is boron doped with a uniform concentration of 8.10^{18} cm^{-3}. The first 40 nm of SiGe are undoped in order to allow boron diffusion from the base and arsenic push from the emitter, without forming potential barriers in the conduction band at the collector-base junction. SiGe and Si layers are respectively grown at 750°C and 810°C, both at 20 Torr. Figure 2.2 gives SIMS profiles of the drift base bipolar right after the SiGe growth.

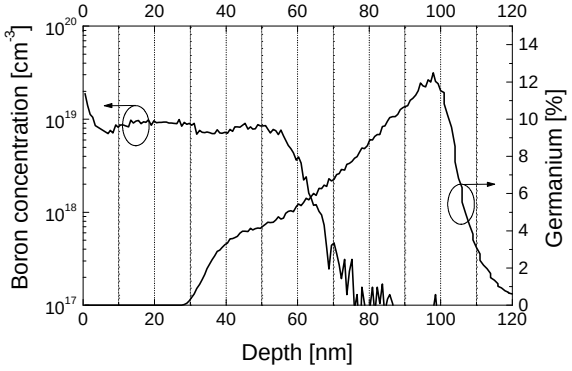


Figure 2.2: SIMS profiles of the "as grown" SiGe epitaxial base layer.

In the previous state-of-the-art the so-called interpoly dielectric layer (isolating the polybase from the subsequent polyemitter) consisted of silicon nitride. Indeed, this layer has to withstand the wet etch of the TEOS layer that opens the intrinsic base area. But this nitride mask limits the growth rate of the epitaxial process since a high HCl ratio has to be introduced in the gas chemistry to allow a selective growth. The innovation brought to the technology is the use of a PECVD oxynitride layer instead of the nitride one [5]. The epitaxy growth rate is then multiplied by a factor of 10 for Si and by about 5 for SiGe (depending of the Ge content). Once the SiGe epitaxy is performed, a composite oxide/polysilicon inside spacer module is applied to reduce the emitter window to 0.35 µm. After the oxide etch, a 50 nm polysilicon layer is deposited and implanted with As at 5.10^{15} cm^{-2} to form the emitter. A second 250 nm thick As in-situ doped polysilicon layer is deposited to reduce the emitter resistance.
The polyemitter, interpoly dielectric and polybase layers are then patterned to define the bipolar device. The oxide/nitride stack previously deposited to protect the CMOS is etched and the PMOS source/drain implant is performed. A protection oxide/nitride stack is again deposited followed by an RTP anneal. A Ti-salicide module is used to reduce the gate resistance, source/drain resistance and external base resistance. Figure 2.3 shows a SEM cross-section of the finalised emitter/base region of the heterojunction bipolar transistor.

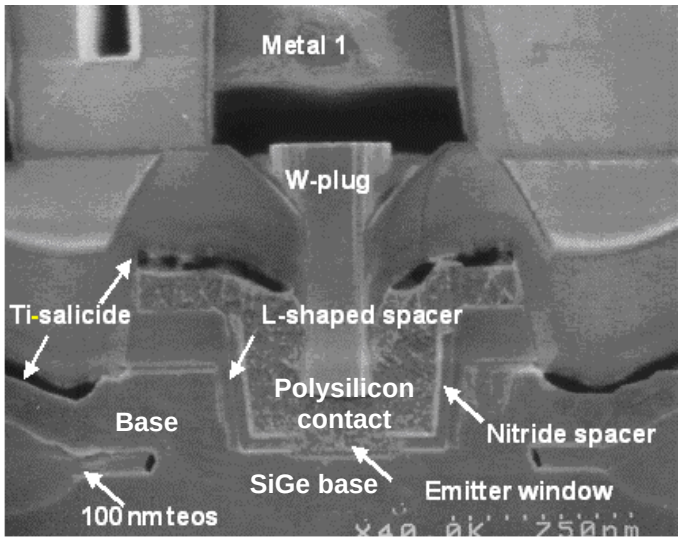


Figure 2.3: SEM cross-section of the emitter/base region of the heterojunction bipolar transistor.

2.2 Goals for the oxynitride optimization

The interpoly dielectric film used in the self-aligned bipolar module requires the following characteristics:

- The layer should show a good selective behaviour towards SiGe deposition in the emitter window.
- The film needs to have a low stress and good adhesion to avoid delamination.
- The etch rate during the opening of the intrinsic base should be as low as possible.
- A good thickness uniformity over the wafer is required to ensure a low spread of electrical characteristics.

LPCVD nitride has a low etch rate in HF solutions but induces high stress and shows limited selectivity towards epitaxial SiGe deposition: a very low deposition rate of the SiGe is required. PECVD oxide introduces low stress and shows high selectivity towards SiGe deposition: a high SiGe deposition rate can be used. On the other hand, this layer etches too fast during the opening of the intrinsic base. Therefore properties of PECVD oxynitride have been explored. To investigate the impact of the N:O ratio of the layer, several ratios of NH_3 and N_2O flows were used. To investigate the impact to the RF power during the deposition, the layers have been deposited both with 350 W RF power and with 400 W RF power. The experimental conditions are represented in Table 2.1.

Recipe	SiON35-95	SiON50-80	SiON65-65
NH_3 (sccm)	35	50	65
N_2O (sccm)	95	80	65
SiH_4 (sccm)	75	75	75
RF power (W)	350/400	350/400	350/400

Table 2.1: Experimental conditions of the PECVD oxynitride depositions. Several ratios of NH3 and N2O flows were used. Each recipe has been applied with 350 W RF power and with 400 W RF power.

2.3 Selectivity of epitaxial deposition

With LPCVD nitride as an interpoly dielectric layer, the selectivity is the limiting factor for the deposition rate of the SiGe epitaxy: only a maximum of 9nm/min could be achieved. With the oxynitride as interpoly dielectric layer, the deposition rate could be increased by a factor 4 to 5. This is done by reducing the HCl flow during the epitaxial deposition. The Ge concentration profile needs to have a ‘staircase’ shape. For this reason, the deposition consists of a number of small steps, each with a decreasing Ge concentration. For a process of this nature a very tight process control is required. For high deposition rates, the individual process steps become too short to ensure the required process control: the control of the epitaxial deposition itself now becomes the limiting factor instead of the selectivity. For this reason, the maximum deposition rate is limited to 33 nm/min.

The sharp increase of selectivity can be explained by the hydrogen terminated surface of the PECVD interpoly layer. The hydrogen would reduce the Si (respectively SiGe) nucleation on the oxynitride, in the same way the hydrogen desorption is believed to act for the precursors in the Si-H-Cl system, i.e. as a reaction rate limiting process [6].

2.4 Stress

High mechanical stress in the interpoly dielectric layer can lead to delayering at the level of the underlaying poly (see figure 2.4). Therefore, it was necessary to optimise the oxynitride recipe towards stress. Both the stress as deposited and after the densification need to be considered. No delayering has been observed with stress levels in between -0.2 and +0.2 GPa.

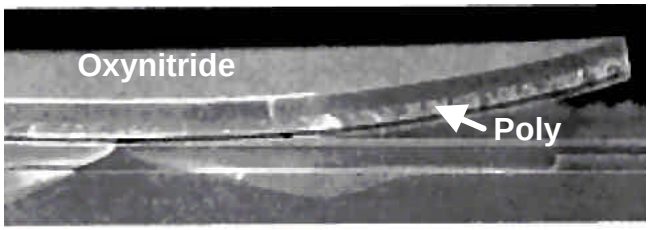


Figure 2.4: SEM cross-section of the
delayering of a high stress interpoly layer.

Figure 2.5 shows the stress evolution as a function of the composition of the layer. The curves are plotted for the 350 W and 400 W layers after deposition and after anneal. Also the curves showing the increase ("delta") of the stress due to the anneal are plotted, both for the 350 W process as for the 400 W process.

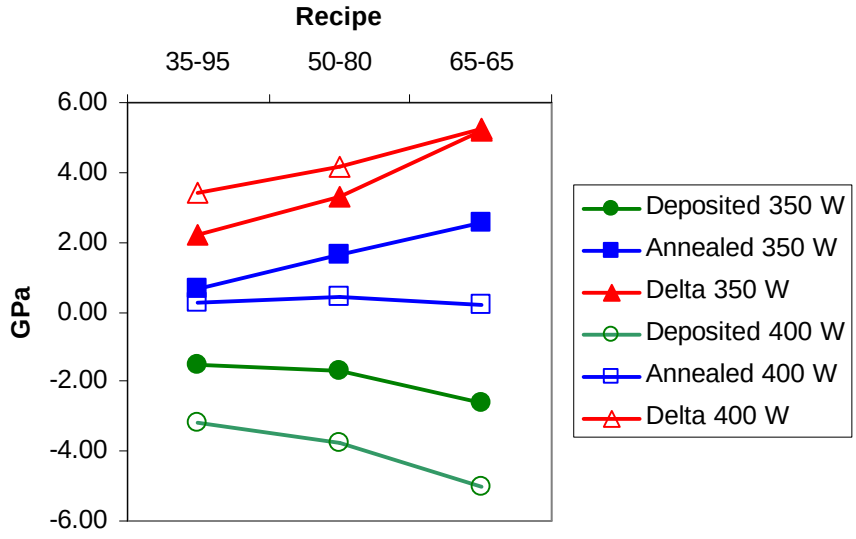


Figure 2.5: Stress evolution as a function of the composition of the layer as deposited and after densification. The change of the stress due to the densification is not constant: a lower NH_3/N_2O ratio results in a smaller change of the stress during densification. An increased RF power results in a more tensile layer as deposited and after densification.

From the graph, we can conclude that

- The densification changes the stress from tensile to compressive.
- A lower NH_3/N_2O ratio results in a less compressive stress as deposited.
- A lower NH_3/N_2O ratio results in a less tensile stress after densification.
- The change of the stress due to the densification is not constant: a lower NH_3/N_2O ratio results in a smaller change of the stress during densification.
- An increased RF power results in a more tensile layer as deposited and after densification.

Based on the experiments described above, one can conclude that the two best recipes have NH_3/N_2O ratio of 35/95 with and without an increased RF power. Since the thickness-uniformity is better for the recipe SiON 35-95, this one was selected. When used on full integration wafers no cracks were noticed. The 35-95 oxynitride will be used for further development of the bipolar module.

2.5 Etch rate

The etch rate of the oxynitride layer should be as low as possible in comparison with the TEOS layer in the emitter window. The following paragraphs are describing the reduction of the etch rate by a densification of the layer and by the selection of the proper etchant.

2.5.1 Densification

A reduction of the etch rate of the deposition can be achieved by a densification of the layer by means of a temperature treatment. Two densification processes have been compared: a densification for 30' at 850°C in N_2 and a densification for 30' at 800°C in O_2 environment. Figure 2.6 shows the impact of the type of densification on the etch rate as a function of the composition of the oxynitride layer.

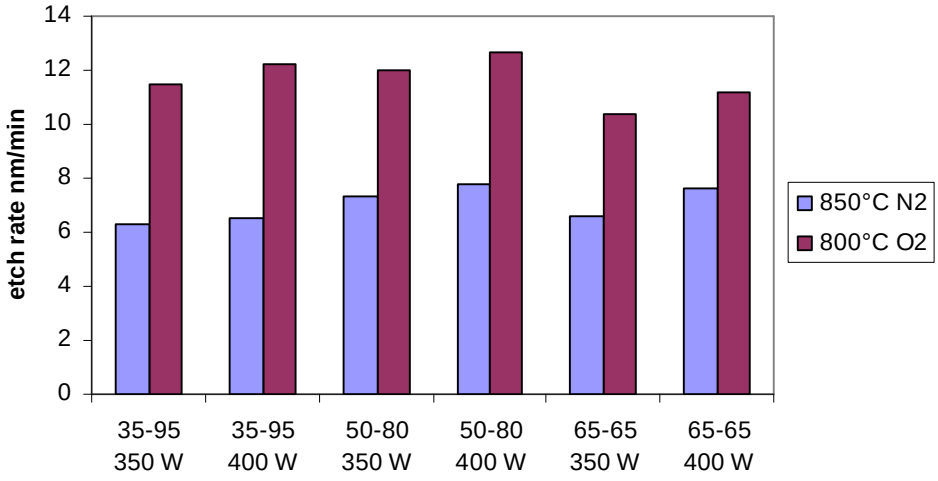


Figure 2.6: The impact of the type of densification on the etch rate of an oxynitride film for all investigated layers. The densification for 30 minutes at 850°C in N_2 results in an etch rate that is about 30% lower compared to a densification for 30' at 800°C in O_2.

For all investigated layers, the densification for 30' on 850°C in N_2 results in an etch rate that is about 30% lower compared to a densification for 30' on 800°C in O_2. Clearly the 850°C densification in N_2 is preferable. From this experiment one cannot conclude whether the impact on the etch rate is due to the temperature or the ambient.

2.5.2 Etchant vs. etch rate

So far an acceptable loss of oxynitride could only be achieved when a high stress oxynitride layer is used. In spite of many experiments of which only a few have been described in the previous paragraphs, no layer could be found, combining a low stress and a low etch rate in 1% HF (1% HF in water). Since adjusting the layer properties proved to be insufficient as long as the 1% HF is being used, other etchants have been explored.
Two types of etchant are introduced in the experiment and compared with 1% HF. The two etchants are 14% HF (14% HF in water) and BHF (6.6% HF and 35.7% NH_4F in water). The etchants have been evaluated on the selectivity for oxynitride vs. TEOS and on etch behaviour in the emitter window.

These etchants have been tried out of five types of depositions.

- Low stress oxynitride 35-95 (SiON 35-95)
- High stress oxynitride 80-50 (SiON 80-50)
- PECVD nitride (ON 70-0)
- LPCVD TEOS (TEOS)
- Thermal oxide (THOX)

For each layer, the etch rate has been measured in each etchant. The results are shown in Table 2.2 and Figure 2.7.

Oxide etch rate (nm/min)	**SiON 35-95**	**SiON 80-50**	**ON 70-0**	**TEOS**	**THOX**
1%HF	13.3	8.4	0.8	25.1	6.9
BHF	58.4	24.5	3.8	305.7	133.8
14%HF	76.6	35.0	3.5	113.8	33.0

Table 2.2: Etch rate in nm/min for several oxide and oxynitride types and etchants.

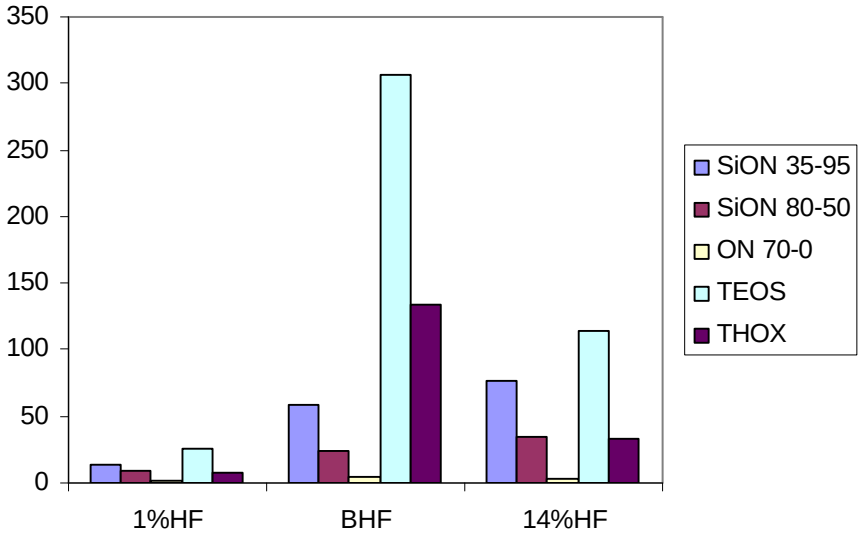


Figure 2.7: Etch rate in nm/min for several oxide and oxynitride types and etchants.

From these data, we cannot only extract the etch rate but also the selectivity: the etch rates normalised towards the etch rate of the TEOS are given in table 2.3 and figure 2.8.

Oxide etch rate ratio	SiON 35-95	SiON 80-50	ON 70-0	TEOS	THOX
1%HF	0.532	0.336	0.032	1.000	0.276
BHF	0.191	0.080	0.012	1.000	0.438
14%HF	0.673	0.308	0.030	1.000	0.290

Table 2.3: Normalised etch rate as a function of oxide type and etchant.

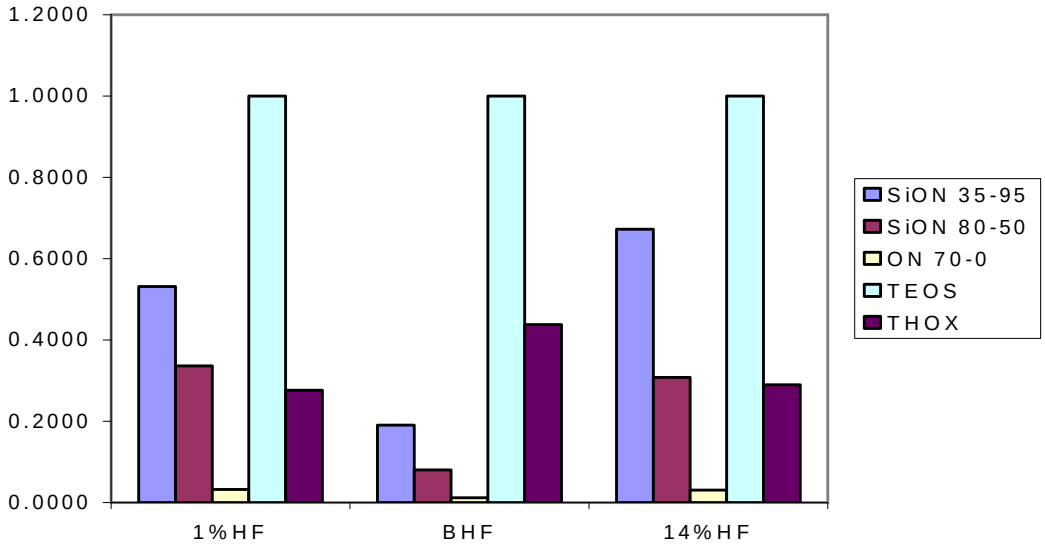


Figure 2.8: Normalised etch rate as a function of oxide type and etchant. BHF has the best selectivity towards low stress oxynitride: 0.2:1. 1% HF has a selectivity of 0.5:1 and with 14% HF selectivity degrades to 0.7:1.

When comparing 14% HF with 1% HF, selectivity has degraded. The 14% HF cannot be considered as an improvement.

On the other hand, from the data above, we can conclude that in BHF, the selectivity towards low stress oxynitride is improved to 0.2:1. When etched in 1% HF, this selectivity is 0.5:1. Using the BHF is stead of the 1% HF reduced the loss of oxynitride by 60%.

This results with BHF look very promising but the following aspects still need to be evaluated:

- Since the etching in the BHF is quite short, reproducibility and uniformity in the emitter window might be an issue.
- Particle performance of the BHF etch: the available BHF tank is also used for resist patterned wafers. This may raise some particle concerns.

2.5.3 Etch behaviour, uniformity and reproducibility

To evaluate the etch behaviour, uniformity and reproducibility, patterned wafers have been etched during 30, 45 and 60 seconds. The 60 seconds etch has been repeated 3 times. To evaluate the undercut in the emitter window, a cross-section of each wafer has been made after epitaxy. To evaluate uniformity over the wafer, cross-sections has been made on the top, centre and bottom of one wafer. To make a clear layer identification possible, an extra nitride layer has been deposited. On each cross-section, the undercut has been measured. Figure 2.9 shows one of the cross-sections.

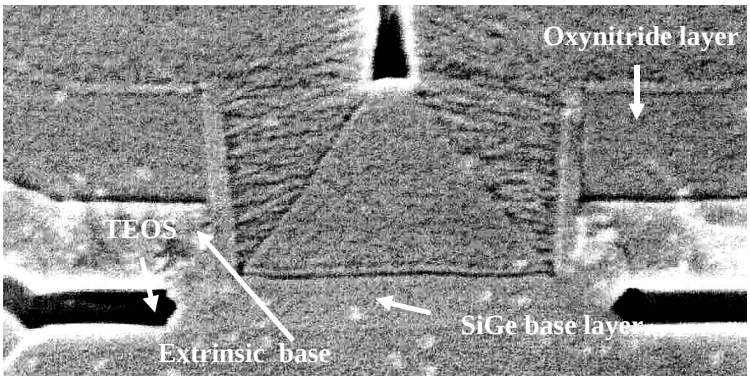


Figure 2.9: Cross-section of the emitter window.

The results of the measurements are summarised in the graph below. The undercut for 30, 45 and 60 seconds has been plotted and fitted by a curve. In order to estimate the repeatability, the 60 seconds etching has been repeated on 3 wafers. Each wafer has been evaluated in the centre. One wafer is evaluated in the centre, top and bottom. The results of the repeatability test and the uniformity test have been added to figure 2.10.

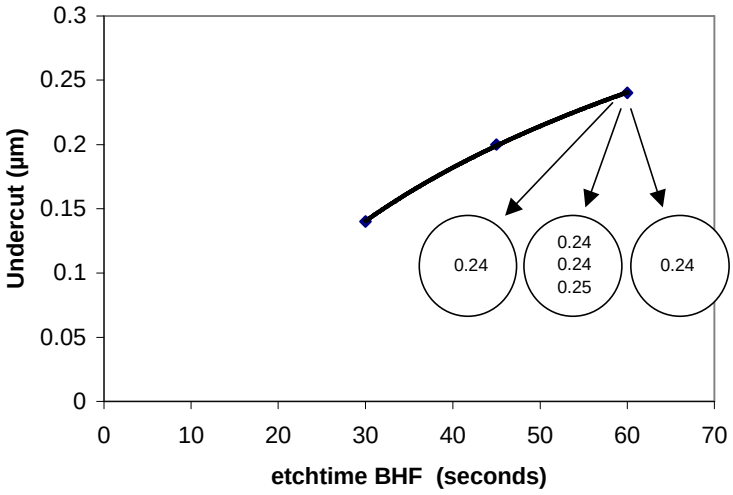


Figure 2.10: Emitter window undercut in function of etch time. The undercut is very controllable and proportional to the etch time. Three repeated wafers show exactly the same result: reproducibility is very good. The bottom of the wafer only shows 10 nm more undercut than the centre or the top; uniformity is excellent.

From the graph one can conclude that:

- The undercut is very controllable by adjusting the etch time
- Reproducibility is very good: all 3 repeated wafers show exactly the same result.
- Uniformity is excellent: the bottom of the wafer only shows 10 nm more undercut than the centre or the top.
- The curve is flattening for longer etch times: the deeper the undercut already is, the slower the etching becomes. The function fits very well a logarithmic function: a saturation is not expected.

An undercut between 0.2 and 0.25 µm is expected to ensure good electrical results. The 60" process is selected for implementation in the process flow.

2.5.4 Cleaning and Particle performance

To insure that all native oxide is removed from the emitter window prior to the epitaxy, an extra 10" 1% HF clean is introduced.

To evaluate the particle performance the following test has been performed:

- 4 wafers have been deposited with 200 nm LPCVD TEOS, 4 wafers are left blank.
- Particles are measured after deposition.
- All wafers have been etched in BHF until bare Si.
- Particles have been measured.
- Wafers are cleaned with an SC1 clean; some wafers are left uncleaned as a reference.
- All wafers get the extra 10” 1% HF clean.
- Particles have been measured.

Table 2.4 shows the results of the test. For completeness it should be added that immediately after the BHF etching, there was a problem with the IPA vapour dryer: a load of particles were generated. These make this test a very worst case test. Extra particles were accidentally added by some scratches at the bottom of the wafers during the TEOS deposition.

	TEOS dep	**TEOS dep**	**TEOS dep**	**TEOS dep**	**ref**	**ref**	**ref**	**ref**
Particles after TEOS dep + densification	6	57*	2	9	0	1	1	4
	BHF	**BHF**	**BHF**	**BHF**	**BHF**	**BHF**	**BHF**	**BHF**
Particles after BHF	391	422*	546	471	5017	4662	3112	2900
	ref	**SC1**	**ref**	**SC1**	**ref**	**SC1**	**ref**	**SC1**
	10”HF	**10”HF**	**10”HF**	**10”HF**	**10”HF**	**10”HF**	**10”HF**	**10”HF**
Particles after SC1 and 10”HF	237	68*	358	231*	4397	11	2740	9

(*) scratch

Table 2.4: Particle performance etching and cleaning. BHF etching, especially when done on bare Si, is resulting on huge numbers of particles. Additional cleaning is required. 10 seconds HF in removing particles very poorly. An extra SC1 clean is removing practically all particles.

From the table, we can conclude that BHF etching, especially when done on bare Si, results in huge numbers of particles. Additional cleaning is required. 10 seconds HF removes particles very poorly. An extra SC1 clean is removes practically all particles.

2.6 Thickness Uniformity

Electrical properties of the SiGe heterojunction bipolar transistor are very sensitive to the deposition temperature of the SiGe [3]. The thickness of the interpoly dielectric influences the emissivity and as such the temperature during the SiGe deposition. As a consequence, thickness nonuniformity has to be limited. A range of ± 3% was found to be acceptable.

2.7 Process window definition

Mechanical stress is found to be minimal for low NH_3/N_2O ratio. The ratio of the etch rate of the oxynitride towards the TEOS should be as low as 0.2 to ensure sufficient remaining

interpoly dielectric. For these reasons the composition SiON35-95 is selected. With this layer, the selectivity towards the epitaxial deposition is no longer the limiting factor for the SiGe growth rate. The epideposition rate is now determined only by process control limitations.
In order to define the full process window, oxynitride layers of different compositions have been evaluated on stress and etch rate ratio as a function of composition as determined by RBS. Figure 2.11 shows the evolution of both the etch rate ratio between the densified oxynitride and the TEOS in 1:9 BHF, and the mechanical stress in the layer, as a function of the nitrogen content in oxynitride.

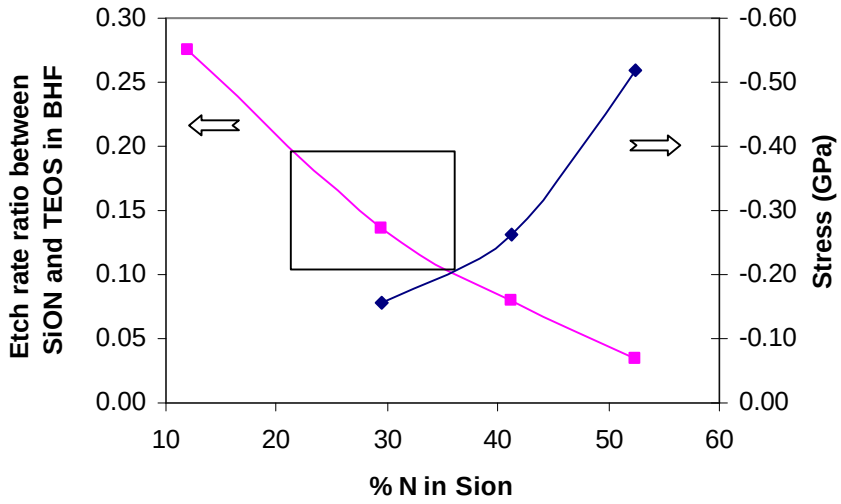


Figure 2.11: Etch rate ratio between oxynitride and TEOS and stress versus the nitrogen content (RBS) in the oxynitride layer. The striped area represents the process window. The maximum etch rate ratio defined the lower N content as 21%. The upper N content of 37% is defined by the stress limit of -0.2 GPa.

The process window limits are defined on one hand by the etch rate ratio oxynitride vs. TEOS that has to be lower than 0.2 and on the other hand by the absolute value of the stress that has to be lower than 0.2 GPa. Nitrogen content in the layer is therefore between 21% and 37%. The oxynitride layer that is currently used in the process contains about 30% nitrogen.

2.8 Performance and yield of the bipolar transistor

The Gummel plots of a single 0.5*1.7 µm² SiGe heterojunction bipolar transistor, and of a yield array of 10000 of these devices connected in parallel, are both shown in figure 2.12. This illustrates the good electrical yield obtained on our 0.35 µm BiCMOS technology.
The average current gain of the heterojunction bipolar transistor is 110. Good process uniformity is demonstrated by standard deviation of the current gain of 8.5%. BVceo is 4.1 V and Early voltage is –125V. f_T and f_{MAX} are respectively about 40 GHz and 85 GHz at V_{ce} of 1.5V. The f_{MAX}/f_T ratio larger than 2 demonstrates the benefit of the fully self-aligned architecture.

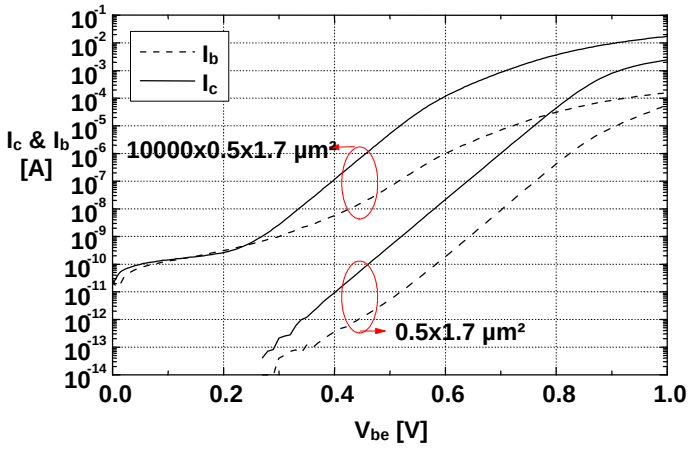


Figure 2.12: Gummel plots of a single 0.5x1.7 µm² SiGe heterojunction bipolar transistor and of a yield array of 10000 devices connected in parallel. Collector and base current of the yield array increased by a factor of 10000 compared to the single transistor. This proves good yield and functionality.

2.9 Conclusion

The manufacturability of the selective SiGe processing is significantly improved by replacing LPCVD nitride as interpoly dielectric by a single layer PECVD oxynitride layer. The composition of the oxynitride layer was chosen as a function of selectivity, etch rate ratio and stress. This layer allows a significant improvement of growth rate of the selective epitaxial deposition by a factor 4 to 5. The oxynitride interpoly dielectric is implemented in an industrial 0.35 µm SiGe BiCMOS technology. Good electrical performance and yield have been proven.

2.10 References

[1] M. Soyuer, “The Impact of the SiGe BiCMOS Technology on Microwave Circuits and Systems”, 30th European Solid-State Device Research Conference (ESSDERC), pp. 34-41, 2000.

[2] A. Gruhle, “Prospects for 200 GHz on Silicon with SiGe Heterojunction Bipolar Transistors”, Proceedings of the 2001 Bipolar/BiCMOS Circuits and Technology Meeting (BCTM), pp. 19-25, 2001.

[3] W. Klein and U. Klepser, “75 GHz Bipolar Technology for the 21st Century”, European Solid-State Device Research Conference (ESSDERC), pp. 88-91, 1999.

[4] P. Chevalier, F. De Pestel, H. Ziad, M. Fatkhoutdinov, J. Ackaert, P. Coppens, J. Craninckx, S. Günçer, A. Pontioglu, P. Tasci, F. Yayil, B. S. Ergun, A. I. Kurhan, E. Vestiel, E. De Backer, J.-L. Lohéac and M. Tack, “An Industrial 0.35 µm SiGe BiCMOS Technology

for 5 GHz WLAN Featuring an Improved Selective Epitaxial Growth Process", IEEE International Symposium on Electron Devices for Microwave and Optoelectronic Applications (EDMO), 2001.

[5] Patent: Process for selective epitaxial growth and bipolar transistor made by using such process, Patent number: EP1280189, Publication date: 2003-01-29, Inventors: J. Ackaert, J. Vastmans, F. Depestel, P. Chevalier.

[6] S. A. Campbell, "The Science and Engineering of Microelectronic Fabrication", Oxford University Press, pp. 349, 1996.

Chapter 3

Characterization of tunnel oxides for non-volatile memory applications

The results in this chapter have partially been published in:

J. Ackaert, A. Lowe, S. Boonen, T. Yao, J. Rayhem, B. Desoete, J. Prasad, M. Thomason, J. Van Houdt, R. Degraeve, L. Haspeslagh and P. Hendrickx, "Characterization of Tunnel Oxides for non-volatile memory applications", International Semiconductor Device Research Symposium (ISDRS), 2003.

3.1 Introduction

A principal problem of floating-gate-based non-volatile memories such as flash memories or EEPROMs is anomalous charge loss which leads to threshold voltage (V_t) shifts on a time scale of months or years at room temperature. The number of these moving bits is greatly affected by the nature and properties of the tunnel oxide that is used in the non-volatile memory technology. In this chapter we compare different types of tunnel oxides and determine the impact on the moving bit issue with different measurement techniques.
The appearance of moving bits is demonstrated by a low field test. This test is a close match to real life conditions and as such gives a good indication of the moving bit performance of the device. Nevertheless this test is very time consuming and it is therefore not feasible for use as monitor during manufacturing. For the first time it is demonstrated that a fast Q_{BD} test of the channel tunnel oxide is generating exactly the same qualitative ranking as the moving bit test of virgin devices; and the Q_{BD} test of the erase junction tunnel oxide is a measure for the moving bit behaviour of cycled devices. These observations are valid over a wide range of oxide qualities. The differences in the initial trap density are reflected in Q_{BD} data. As such the Q_{BD} test, when performed under the given circumstances, is a valid tool for non-volatile memory reliability assessment both for virgin and for cycled devices.
Silicon dioxide formed at low temperature in a water vapor environment proves to be far superior in terms of moving bit performance compared to the oxides grown at high temperatures in an oxygen environment. An oxidation in an oxygen environment process inevitably generates weak and/or dangling bonds near the Si/SiO_2 interface and a high trap density in the oxide [1]. It is confirmed that the hydrogen incorporated by the wet oxidation tends to passivate the dangling bonds as well as the oxide traps.
The superior behaviour of the wet oxides in terms of moving bits remains over the full life cycle of the memory cell.

3.2 The experiment

All experiments have been performed on 1-Mb 0.35µm flash memory circuits based on the HIMOS^{TM} cell concept [3]–[6]. The cells are programmed by source-side injection (SSI) and erased by Fowler–Nordheim tunneling. Figure 3.1 shows a schematic overview of the cell and the normal operating conditions.

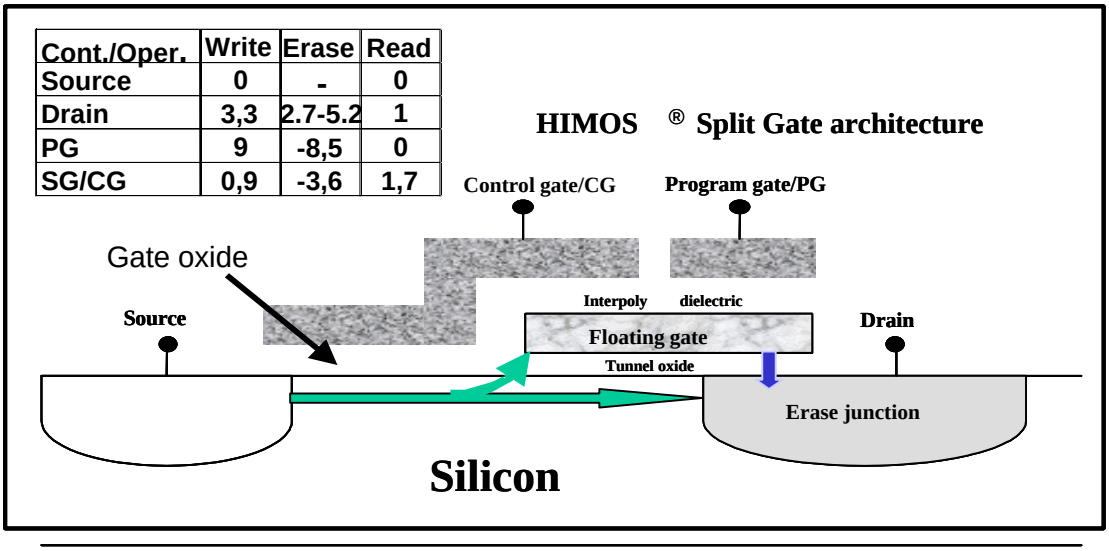


Figure 3.1**:** A schematic drawing of the HIMOS^{TM} cell. The green arrow represents the writing of the cell by hot electron injection into the floating gate. The blue arrow represents the erasing of the cell by tunneling these electrons back into the drain. The table gives the normal read, write and erase operating conditions in Volt.

A tunnel oxide thickness in the range of 8 to 10 nm is used. Different oxidation process conditions have been compared:

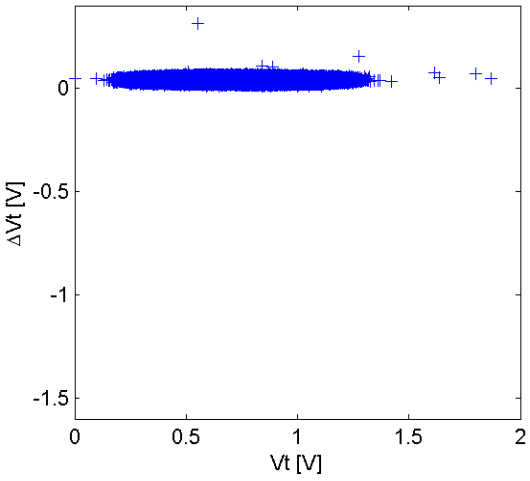
- Dry oxidation at 900°C: the oxide is grown in a pure oxygen atmosphere.
- 5% O_2 diluted oxidation at 900°C or 960°C: the oxides are grown in a 5% oxygen atmosphere diluted with N_2.
- Wet oxidation at 750°C. the oxide is grown in a water vapor atmosphere

3.3 The measurements

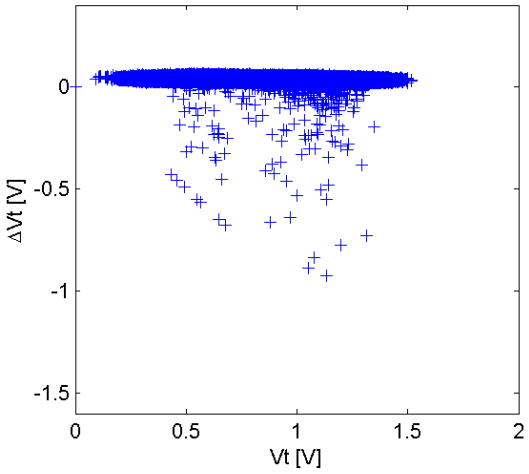
Since it is impossible to check data retention for ten years or more, accelerated testing and models are needed [6] which can be used for long-term extrapolations of charge loss and for product failure-rate estimations. In this chapter different measurement techniques are used that can provide the input for these models. First moving bits are discussed (3.3.1), next the leakage current of the dielectric analyzed (3.3.2) and finally the charge-to–breakdown (Q_{BD}) is investigated (3.3.3).

3.3.1 Moving bits measurements

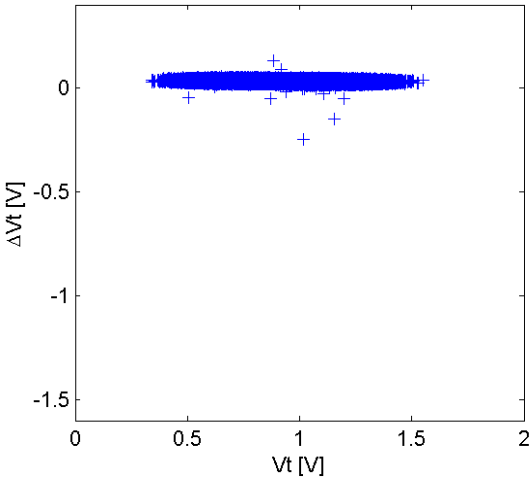
The most favorable and direct approach to assess the real life moving bit behavior is a moving bit measurement under accelerated conditions. This method is as close as possible to real life conditions. The main drawback of this method is that it is very time consuming.
The moving bit investigation presented in this chapter has been performed on the programmed level (V_t of about 1 V). In order to increase the field, the voltage on the drain or bit line is kept at +3.4V instead of the normal +1 V. After certain storage times, the complete memory array has been read out and the V_t of each individual cell has been determined. The V_t is defined as the program gate voltage at which the cell current is 10μA. Figure 3.2 shows V_t shift as a function of the initial V_t for various oxide types and oxide thickness. Each figure contains 250000 measurements. In the figures, a negative V_t shift corresponds to the loss of electrons from the floating gate.
The dielectric grown at 900°C in 5% O_2 ambient with a thickness of 10.28 nm shows no sign of V_t shift after stress (See Figure 3.2(A)). However below 10 nm, several moving bits start to emerge as shown in Figure 3.2(B). In order to reduce the moving bit issue, a wet 750°C oxidation was introduced. This wet low temperature oxidation gives superior results even for a minimal oxide thickness of 8.89 nm as shown in Figure 2(C). Figure 3.2(D) shows that reducing the thickness to less than 9.0 nm with the 5% O_2 dry oxidation at 900°C is causing a increase of the number moving bits. Still referring to this process, increasing the temperature to 960°C, as shown in Figure 3.2(E), generates even more moving bits. Using a 100% O_2 process also generates a high number of moving bits as shown in Figure 3.2(F).
The diagonals showing up on the graphs with a large number of moving bits are cells where the V_t shift is equal to the to initial V_t. This is a measurement artifact: the system was set up to measure positive V_t's only, When the V_t reduction was equal of higher than the initial V_t, in other words, when the V_t became negative, then a zero value was returned.



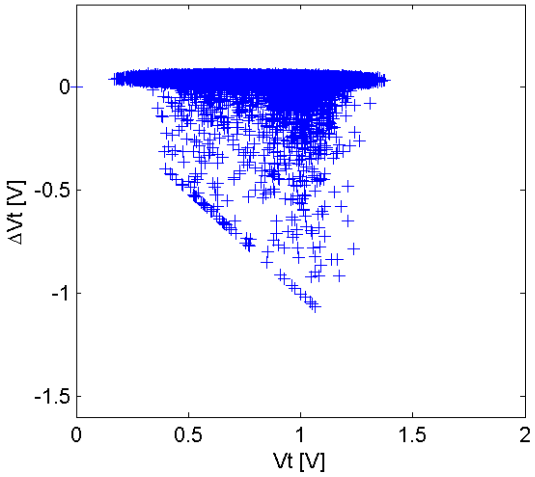
(A) 5% O_2 diluted oxide at 900°C 10.28 nm



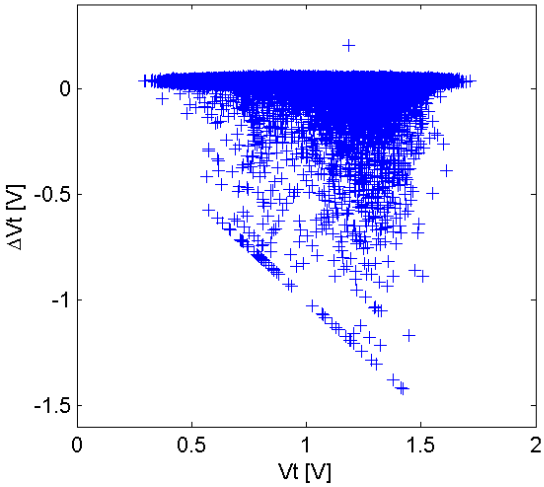
(B) 5% O_2 diluted oxide at 900°C 9.22 nm



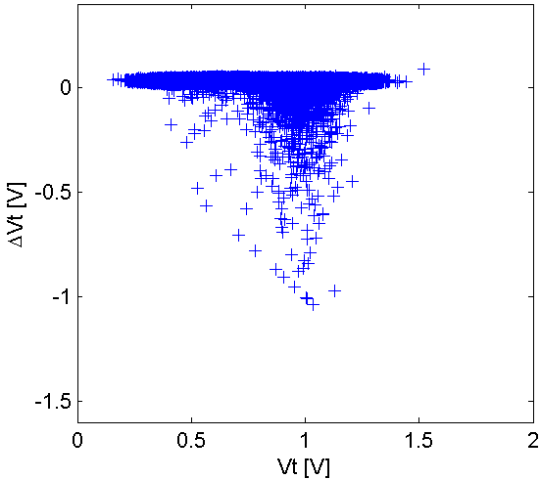
(C) Wet oxide at 750°C 8.89 nm



(D) 5% O_2 diluted oxide at 900°C 8.96 nm



(E) 5% O_2 diluted oxide at 960°C 9.30 nm



(F) 100% O_2 dry oxide at 900°C 9.34 nm

Figure 3.2**:** V_t shift, represented on the Y axis, as a function of the initial V_t for various oxide types and oxide thickness.

During writing and especially during erasing of the cells, currents are forced through the tunnel oxide. Even though these currents are minimal they are affecting the oxide quality anyhow. Therefore the quality of the cells not only has to be assessed in the initial virgin state but in fact over the full life cycle. To assess the behaviour over the lifetime of the memory cell, the number of moving bits has been measured after writing and erasing the cell for 10000 times. In the following graphs, the distributions of the moving bits are compared for the 4 oxide types before and after 10.000 write-erase cycles.

Figure 3.3 is comparing the number of moving bits of virgin cells using dry oxide at 900°C, diluted oxides at 900°C and 960°C and wet oxide at 750°C as tunnel oxide. Samples have been selected with as much as possible the same oxide thickness. This graph is using the same data as graph 3(A), 3(C), 3(D) and 2(F).

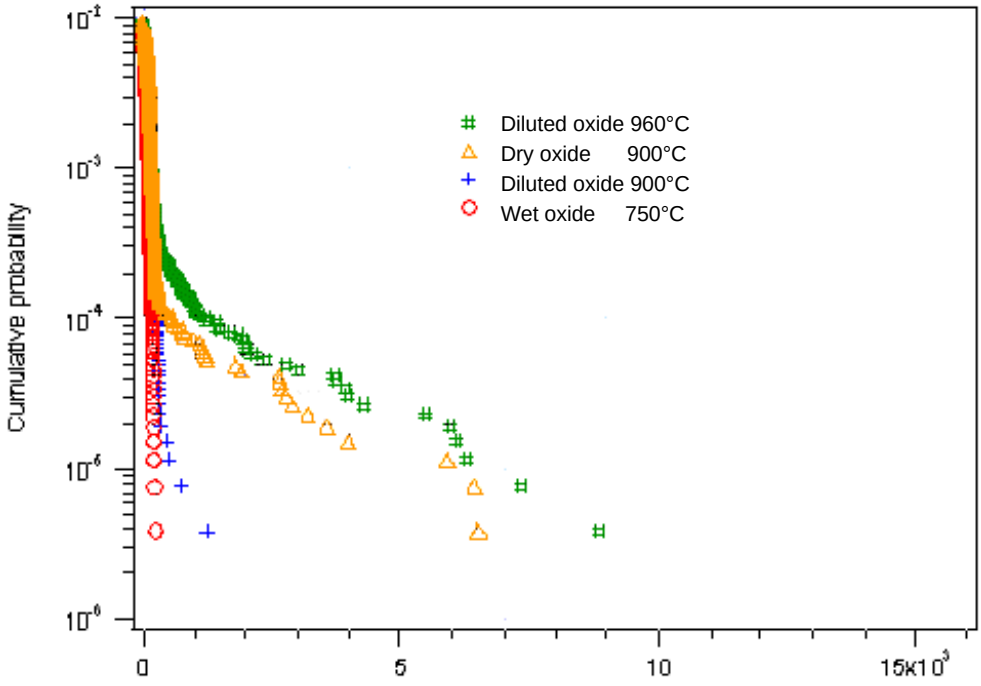


Figure 3.3**:** The number of moving bits on virgin samples for various oxide types. Similar to figure 3.2 the wet low temperature oxidation gives superior results: in virgin cells no moving bits are observed. The diluted oxidation at 900°C is causing an increased number of moving bits. Still referring to this process, increasing the temperature to 960°C generates even more moving bits. Using a 100% O_2 process results in an intermediate quality in between the diluted oxide at 900°C and the diluted oxide at 960°C.

Figure 3.4 compares the number of moving bits of virgin cells. For dry oxide at 900°C, diluted oxides at 900°C and wet oxide at 750°C. Unfortunately the 960°C could not be evaluated after cycling any more: as already mentioned, the moving bit measurements are very time consuming. Because the available time was limited to only 3 samples, the worst sample (diluted 960°C) had to be omitted.

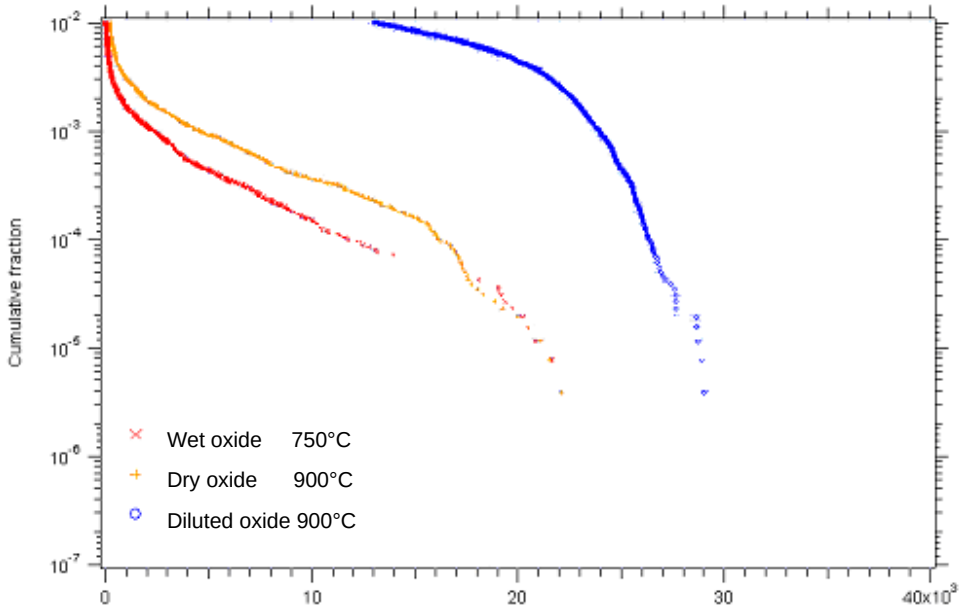


Figure 3.4: The number of moving bits on cycled samples for various oxide types. The wet oxide also starts to show moving bits. While on virgin cell, the diluted 900°C oxide performed better towards the number of moving bits, now the 100% O_2 dry oxide is better than the 5% O_2 diluted oxide. Due to the cycling, the 5% O_2 diluted oxide degraded much faster than the 100% O_2 dry oxide.

After write–erase cycling, the wet oxide also starts to show moving bits due to the traps induced by the charge transport through the tunnel oxide. Still, similar to figure 3.3, when compared to the other oxides, the wet low temperature oxidation gives the best results. For the two other oxides, something remarkable happened. While on virgin cell, the diluted 900°C oxide performed better towards the number of moving bits, now the 100% O_2 dry oxide is better than the 5% O_2 diluted oxide. Due to the cycling, the 5% O_2 diluted oxide degraded much faster than the 100% O_2 dry oxide.

3.3.2 Leakage measurements

In order to find a faster evaluation method, it was investigated whether oxide leakage measurements showed any relation with the results of the moving bits test.
Leakage measurements have been performed on a 650000 μm^2 flat Si-poly capacitor for the different types of oxide. Figure 3.5 shows the measured voltage required for a leakage current of 140 pA as a function of oxide thickness. For all types of oxide, the leakage current is determined by the thickness of the oxide. From these measurements, no conclusions can be made about the moving bit issue.

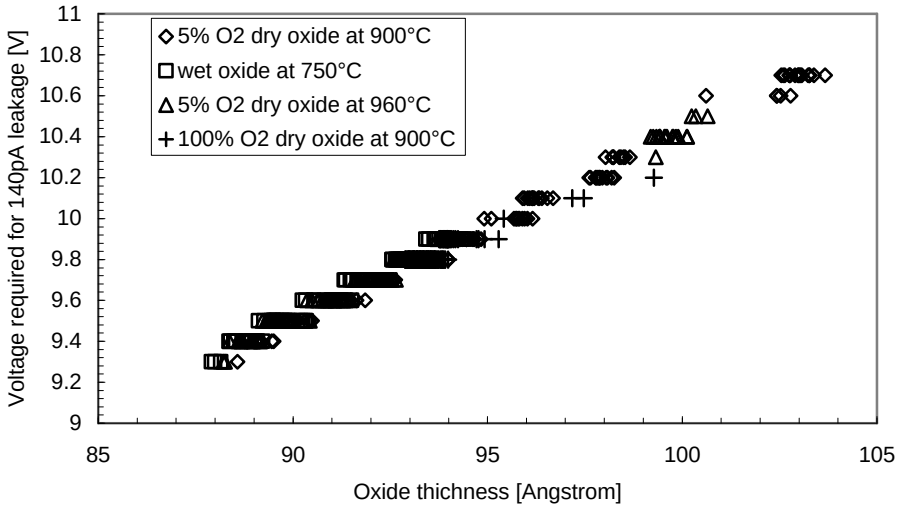


Figure 3.5: The measured voltage required for a leakage current of 140pA as a function of oxide thickness.

3.3.3 Q_{BD} measurements

Q_{BD} values were measured on square floating gate poly (FG)-to-PWELL capacitors. The configuration of these capacitors is representative of the channel of the transistor. 100 mA/cm^2 was applied as stress current density. Depending on the oxide thickness, intrinsic breakdown occurred between –11 and –13 V. For the evaluation of the different recipes, the same samples have been selected as used for the moving bit evaluation. Figure 3.6 shows the Weibull distribution for the different recipes.

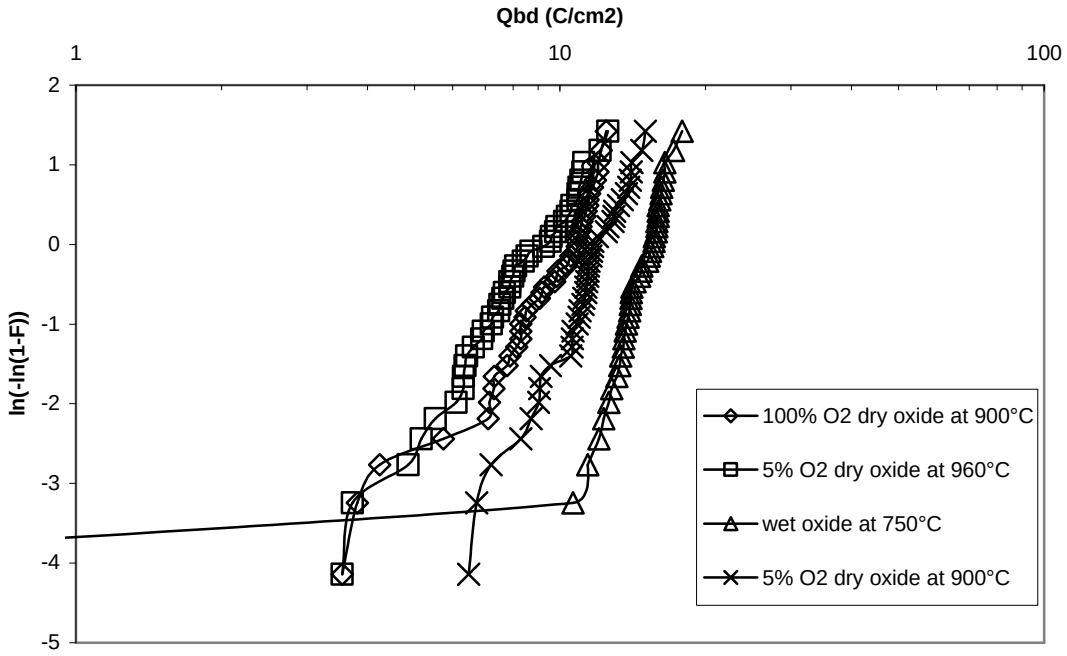


Figure 3.6: Q_{BD} of the different types of channel tunnel oxide. The wet oxide is the best in terms of intrinsic Q_{BD}. The high-temperature diluted oxide has most moving bits and was also worst in terms of Q_{BD}. The intermediate cases the 100% O_2 dry oxide has more moving bits than device with the 5% O_2 dry oxide at 900°C and this is also reflected in terms of Q_{BD}.

The 4 types of oxide that have been compared show clearly a different distribution. The ranking made for the oxide quality in terms of breakdown (based on intrinsic Q_{BD} values) perfectly reflects the performance of the different oxides in terms of initial moving bit. The wet oxide already has no moving bits yet before cycling despite its thinner oxide compared to the other cases. This wet oxide was also best in terms of Q_{BD}. The high-temperature diluted oxide has most moving bits and was also worst in terms of Q_{BD}. So the worst and the best cases correspond in terms of Q_{BD} and moving bits. Also the intermediate cases, the correlation is matching: The 100% O_2 dry oxide has more moving bits than device with the 5% O_2 dry oxide at 900°C and this is also reflected in terms of Q_{BD}.

This observation shows that a good performance in terms of oxide breakdown correlate very well with a good performance in terms of moving bits for the virgin cells. This is quite surprising when considering very different nature of the stress conditions in both measurements. Breakdown involves stress-induced oxide trap densities of ~10^{20} cm^{-3} which is several orders of magnitude higher than the oxide trap densities seen by moving bits. Even though the nature of silicon-induced leakage current and oxide breakdown is very different and even though the difference of the stress applied in both measurements is large, we observe a good correlation. We can conclude that for the measurement in the given conditions, differences in the initial trap density in virgin cells (e.g. because of a different oxide recipe) are reflected in the Q_{BD} data.

At the erase junction, the side of the floating gate is implanted together with the drain by the erase junction implant. To asses the impact of this implant on the tunnel oxide behaviour, gate poly (FG)-to-erase junction capacitor test structure has been investigated. Q_{BD} values were measured on square floating gate poly (FG)-to-erase junction capacitor. 100 mA/cm^2 was applied as stress current density. The tunnel oxide between FG-to-erase junction breaks down between –20 and –23 V. For the evaluation of the different recipes, the same samples have been selected as used for the moving bit evaluation. Figure 3.7 shows the Weibull distribution for the different recipes.

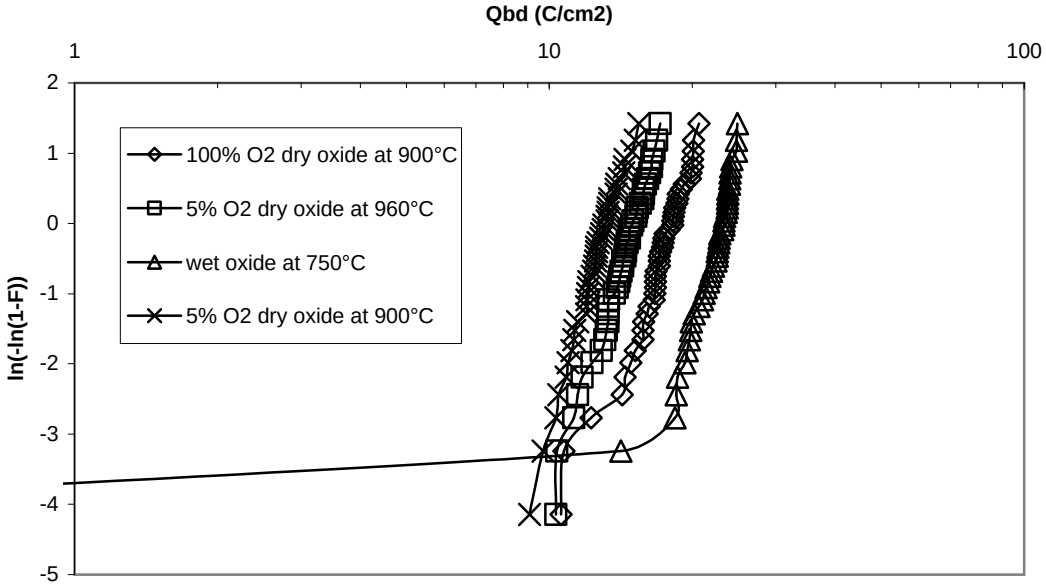


Figure 3.7: Q_{BD} of the different types of erase junction tunnel oxide. The wet oxide performs the best in terms of Q_{BD}. When comparing the 100% O2 dry oxide at 900°C with the diluted dry oxide at 900°C, now the 100% O2 dry oxide performs better.

The 4 types of oxide that have been compared show clearly a different distribution. Again the wet oxide performs the best in terms of Q_{BD}. When comparing the 100% O2 dry oxide at 900°C with the diluted dry oxide at 900°C, again something remarkable can be observed: where for the (FG)-to-PWELL capacitor, the diluted oxide at 900°C performed better, now the 100% O2 dry oxide performs better. This is exactly the same behaviour as what was observed when comparing the moving bits test results of virgin and cycled cells. Where Q_{BD} measurements of the channel tunnel oxide correlate with the moving bit results of the virgin devices, the Q_{BD} data of the erase junction tunnel oxide correlates with the moving bit results of the cycled devices. For measurement, the diluted dry oxide at 960°C now performs better than the diluted dry oxide at 900°C. Unfortunately, there is no moving bit data to compare with this data.

3.4 Discussion

In previous sections it is demonstrated that the moving bit behaviour of virgin cells correlates with the Q_{BD} performance of the channel tunnel oxide. The moving bit behaviour of cycled cells correlates with the Q_{BD} performance of the erase junction tunnel oxide. The anomalous charge loss in virgin cells is determined by the quality of the channel tunnel oxide: when comparing the breakdown voltages of both oxides this is indeed the weakest oxide. Occurrence of anomalous charge loss is as such determined by the properties of this channel tunnel oxide. This explains the correlation between both measurements.
However, during cycling, the situation is affected by the different mechanisms used for writing and erasing. The writing is performed by hot carrier electron charge transfer over the tunnel oxide. This charge transfer does not affect the tunnel oxide quality in the channel of the cell. However, for erasing the cell, a high field is applied between the floating gate and the erase junction. This causes electrons to tunnel through the erase junction tunnel oxide. Since the oxide thickness is too large to allow direct tunneling, the tunnel mechanism is Fowler-Nordheim tunneling. This tunneling mechanism degrades the erase junction tunnel oxide. Since during erasing, current is only passed through the erase junction tunnel oxide, the channel tunnel oxide is not affected. Through cycling of the cell, the erase junction tunnel oxide becomes weaker than the channel tunnel oxide. As such the erase junction tunnel oxide determines the anomalous charge loss. This explains why, after cycling, moving bit behaviour is determined by the nature and properties of the erase junction tunnel oxide rather than by the properties of the channel tunnel oxide.
Appealing is the correspondence between moving bit and Q_{BD} measurements. Still one has to take in account the big difference between the stress conditions in both measurements and the number of oxide traps involved before and after stress. Breakdown involves stress-induced oxide trap densities of $\sim 10^{20}$ cm^{-3} which is several orders of magnitude higher than the oxide trap densities seen by moving bits, even after cycling. A reduction in the initial trap density (e.g. because of a different oxide recipe) measured as a reduced number of moving bit, does not have to lead by definition to an improved Q_{BD} performance. Nevertheless, all our experimental data collected so far demonstrates a correspondence between moving bit and Q_{BD}.

3.5 Conclusion

The appearance of moving bits is demonstrated by a low field test. This test is a close match to real life conditions and as such gives a good indication of the moving bit performance of the device. Nevertheless this test is very time consuming and it is therefore not feasible for use as monitor during manufacturing. For the first time it is demonstrated that a fast Q_{BD} test of the channel tunnel oxide is generating exactly the same qualitative ranking as the moving bit test of virgin devices, and the Q_{BD} test of the erase junction tunnel oxide are a measure for the moving bit behaviour of cycled devices. These observations are valid over a wide range of oxide qualities. The differences in the initial trap density are reflected in Q_{BD} data. As such the Q_{BD} test, when performed under the given circumstances, is a valid tool for non-volatile memory reliability assessment both for virgin and for cycled devices. Still, since the nature of the mechanisms of trap-assisted tunneling (responsible for moving bits) and oxide breakdown is very different, the moving bit measurements should not be replaced completely by the Q_{BD} measurements.
Low temperature wet oxide proves to be far superior in terms of moving bit performance compared to the dry oxides grown at high temperatures. A dry oxidation process inevitably

generates weak and/or dangling bonds near the Si/SiO_2 interface and a high trap density in the oxide [1]. It is confirmed that the hydrogen incorporated by the wet oxidation tends to passivate the dangling bonds as well as the oxide traps.
The superior behaviour of the wet oxides in terms of moving bits remains over the full life cycle of the memory cell.

3.6 References

[1] P. Hahn and M. Henzler, "The Si-SiO2 interface: correlation of atomic structure and electrical properties" J. Vac. Sci. Technol., vol. A.2, No.2, pp. 574, 1984.

[2] Tempel et al., "Abnormal Charge loss of Flash cells at medium temperatures", NVSM Workshop, 2000.

[3] J.Van Houdt et al.," Method for programming a semiconductor memory device," U.S. Patent 5 583 810, 1996.

[4] J.Van Houdt et al., " Transistor structure for erasable and programmable semiconductor memory devices," U.S. Patent 5 583 811, 1996.

[5] J.Van Houdt et al., "Transistor structure for use in erasable and programmable semiconductor memory devices", European patent no. 0501941, Oct.23, 1996.

[6] D. Wellekens et al., "Embedded HIMOS^{TM} flash memory 0.35-µm and 0.25-µm CMOS technologies", IEEE Trans. Electron Devices, vol. 47, pp. 2153–2160, 2000.

[7] R. Degraeve et al., "Analytical model for failure rate prediction due to anomalous charge loss of Flash memories", IEDM Tech. Dig., pp. 699-702, 2001.

Chapter 4

Test structures, protection methods and evaluation methods for process induced damage

The results in this chapter have partially been published in:

J. Ackaert, Z. Wang, E. De Backer, P. Colson, P. Coppens, “Non Contact Surface Potential Measurements for Charging Reduction During Manufacturing of Metal-Insulator-Metal Capacitors”, Microelectronics Reliability, vol. 41, pp. 1403-1407, 2001.

J. Ackaert, E. De Backer, P. Coppens, M. Creusen, “Plasma damage antenna test structure matrix description, application for optimization high density plasma oxide deposition, metal etch, Ar-preclean and passivation processing in sub-half micron CMOS processing”, 1st European Symposium on Plasma Induced Damage (ESPID), 1999.

J. Ackaert, Z. Wang, E. De Backer, P. Coppens, “Charging Damage in Floating Metal-Insulator-Metal Capacitors”, 6th International Symposium on Plasma Process-Induced Damage, pp. 120-123, 2001.

J. Ackaert, Z. Wang, E. De Backer, P. Coppens, “Plasma Damage in floating Metal-Insulator-Metal Capacitors”, 8th International Symposium on the Physical & Failure Analysis of Integrated Circuits (IPFA), pp. 224-227, 2001.

Z. Wang, J. Ackaert, C. Salm, F.G. Kuper, “Plasma Process-Induced Latent Damage on Gate Oxide Demonstrated by Single-layer and Multi-layer Antenna Structures”, 8th International Symposium on the Physical & Failure Analysis of Integrated Circuits (IPFA), pp. 220-223, 2001.

J. Ackaert, Z. Wang, E. Backer, P. Colson, P. Coppens, “Non Contact Surface Potential Measurements for Charging Reduction During Manufacturing of Metal-Insulator-Metal Capacitors”, 12th European Symposium on Reliability of Electron Devices, Failure Physics and Analysis (ESREF), pp. 1403-1408, 2001.

Z. Wang, J. Ackaert ,C. Salm, F.G. Kuper, “Charging Induced Damage on Complex-antenna Test Structures”, Proceedings Semiconductor Advances for Future Electronics (SAFE), pp. 220-223, 2001.

J. Ackaert, K. Bessemans, E. De Backer, “Plasma Charging Damage Induced by a Power Ramp Down Step in the end of Plasma Enhanced Chemical Vapor Deposition (PECVD) Process”, 14th European Symposium on Reliability of Electron Devices, Failure Physics and Analysis (ESREF), 2003.

4.1 Introduction

Besides the formation of a dielectric meeting all the required specifications of performance and reliability, it is a major challenge to keep this dielectric intact throughout the full manufacturing process of the integrated circuit. Especially for gate oxides, a major issue is the use of plasma processing during the manufacturing process. The plasma process ambient is a very harsh environment which influences the yield and reliability of finished semiconductor devices.

Since already a number of years, plasma process induced damage is considered to be a major industry-wide problem requiring an interdisciplinary approach. As with many other problems in the integrated circuit industry, the impact of plasma process induced damage on real products has been aggravated by the relentless scaling of device and interconnect dimensions.

In the last decade, integrated circuits interconnecting 10^7-10^8 transistors became a major technological challenge requiring the use of up to 8 levels or even more of metallization. To achieve this goal, high density plasma enhanced deposition and etching techniques have to be used. However, the use of high density plasma reactors and the accompanying higher density of charging species (10^{12} ions/cm^3) impose a major yield and reliability threat. This is the case especially for those technology nodes using gate oxides in a thickness between 10 nm and 3 nm. In general, above 10nm the gate oxide can support the field induced by a plasma in modern process equipment. Between 3 and 10 nm the electrical field generated by a plasma is strong enough to induce Fowler-Nordheim tunnelling current through the gate oxide. This tunnelling current affects the oxide quality. Below 3 nm, direct tunnelling mechanism starts to replace the Fowler-Nordheim tunnelling mechanism to transfer charges through the gate oxide. Direct tunnelling has far less impact on the oxide quality than Fowler-Nordheim tunnelling. The combination of high-density plasmas and 3 to 10 nm gate oxide gate oxides, therefore, constitutes one of the major threats for the operational lifetime of the involved integrated circuits [1] [2].

The three main plasma process induced damage models are based on surface charging, the non-uniformity of the plasma and on the electron shading effect as described by Hashimoto. Most important is the electron shading effect which causes positive charging during the patterning of high aspect ratio structures. This occurs even in a perfectly uniform plasma [3] [4].

Apart from the increasing use of high density plasmas in combination with very thin gate oxides, CMOS devices will be extended to RF applications, for the fast expanding field of telecommunications. Typical for the BiCMOS RF technologies is the use of inductors in combination with capacitors. These capacitors are made as metal-insulator-metal capacitors. The application of the very thin dielectric together with the metal processing in a high plasma density plasma environment makes these metal-insulator-metal capacitors very vulnerable to plasma damage [5].

To evaluate plasma process induced damage induced by the different process modules on gate oxides and metal-insulator-metal capacitors, test structures with large antennas connected to the gate electrode have been developed. Typically, large plate and comb antennas are used for the poly and metal patterning processes whereas antennas with a large number of contact and via holes are used to assess the plasma process induced damage induced during contact and via processing. Besides that, electron shading effects are studied by varying the spacing of the

comb antennas. However, also more dedicated test structures have been designed to achieve a better understanding of the more fundamental plasma process induced damage aspects, such as timing and pattern density related aspects [6] [7]. This chapter describes the test structures required for evaluation of plasma damage on MOS devices and metal-insulator-metal capacitors and HIMOS^{TM} flash memory cells.

Non-contact surface charge measurements provide a fast and indirect way to detect plasma process induced damage. This type of measurement is typically used to monitor the plasma (non-) uniformity by measuring the surface charge on non-patterned oxide wafers. This technique, however, is not detecting electron shading effects [8]. This chapter describes the application of this measurement in order to detect possible sources of plasma damage.

To assess the impact of the topography, evaluation methods are required that allow to qualify a process in an efficient but reliable way on the antenna structures. This chapter compares evaluation techniques, such as voltage breakdown, charge pumping, gate leakage and threshold voltage measurements. The measurement methods are compared in terms of sensitivity, speed and ease of use.

In some cases, plasma damage can be prevented by the development of the proper process. In the majority of cases however, protection structures need to be foreseen already in the design phase of the circuit [9]. This chapter provides designer guidelines, possible scenarios and required limitations for protecting semiconductor devices against plasma damage.

4.2 Test structures

By now it is well established that plasma process induced damage is a result of high field Fowler-Nordheim tunnelling through the gate oxide. The plasma charge, inducing this high field, is collected by the conducting area connected to the gate electrode. This conducting area, which is called the antenna, includes all poly lines, contact holes, via holes, and metal lines connected to the gate electrode. The antenna ratio is typically defined as the total conducting area on field oxide divided by the active gate area for CMOS devices or the capacitor area for metal-insulator-metal capacitor devices. The charge inducing process can be detected by connecting the small active gate areas to a large variety of antennas for the different process levels.

The two types of antennas which are commonly used are area intensive and perimeter intensive antennas. The area intensive antennas are sensitive for plasma process induced damage induced by processes in which the plasma charge is collected by the complete area, such as ashing and deposition processes. The perimeter intensive "comb" antennas are used for the patterning processes in which only the edge of the antennas can collect the plasma charge.

To distinguish the Electron Shading effect from the plasma (non)-uniformity effect, a non-contact surface charge measurement technique needs to be performed in parallel with the evaluation of the antenna test structures.

To detect antenna effects during contact and via etching processes, antennas with different numbers of holes must be applied. New test structures are, however, necessary to study more fundamental plasma process induced damage aspects such as timing and pattern density related aspects.

4.2.1 Concept of an antenna test structure

An antenna test structure generally consists of a charging sensitive device, a small capacitor or transistor, with some form of charge collecting "antenna" connected to the floating gate. The current collected from the plasma charges the floating gate and stresses the gate oxide of the charging sensitive element. Electrical characterization at end-of-line assesses the impact of the charging stress on the relevant electrical parameters. As a general rule, the impact increasing with the size of the antenna is indicative for plasma charging damage. The size of an antenna is quantified by its antenna ratio which is the ratio of the charge collecting area of the antenna (on isolation) to the gate oxide area of the charging sensitive element. To assess antenna ratio dependent charging, at least 3 different antenna sizes should be foreseen (small – medium – large) of each type of antenna and at each level. The gate oxide area of the charging sensitive device is often kept fixed for simplicity and easy comparison. The area is quite small, of the order of 1 μm^2, to limit the total chip area consumption of the antenna test structure set.

Both PMOS and NMOS charging sensitive devices have to be implemented as considerable differences in plasma charging damage have been observed between PMOS devices (in nwell) and NMOS devices (in p-type substrate).

A capacitor as the charging sensitive element has the advantages of simpler processing (which is particularly advantageous in dedicated short-loop experiments) and minimal probe pad

count, but limits the amount of electrical characterization that can be done. On the other hand, a small transistor as the charging sensitive element might be more complex to process, requires extra probe pads for source and drain connections but offers the highest flexibility for electrical characterization.

4.2.2 Standard set of MOS structures

The purpose of this standard set is to assess the plasma charging damage contributions of the individual process steps and/or modules on the one hand, and of the cumulative plasma damage experienced during the complete process flow on the other hand.

The test structures can be fabricated in any CMOS flow. In general, to limit the size of the test structures, antennas are connected to a transistor with a close to minimal gate area. For a 0.35 µm technology, this can be 0.35 µm x 2 µm. The antennas consist of conductors like poly and metal of various designs. Also contact and via structures are included. All test structures can be made available in both NMOS and PMOS.

Antennas of varying size and layout are connected to the floating gates to collect plasma charging current during process steps/modules sensitive to this particular antenna layout. The following general classes of antennas can be distinguished: the area intensive plate antenna, perimeter intensive comb antennas, charging damage from contact/via hole processing and cumulative antennas are sensitive to all process steps of a complete flow.

The area intensive plate antenna

The area intensive plate antenna, which has a large antenna area and minimized perimeter (edge length), primarily assesses charging damage during photoresist ashing, dielectric deposition and implantation processes. For these types of processes the collected charge will be proportional to the area of the antenna. In the case of ashing the antenna surface can collect plasma charge when the resist has (almost) been stripped. On the other hand, for dielectric deposition processes the charging will mainly take place at the beginning of the deposition process as an insulating layer of increasing thickness forms. Table 4.1 gives an overview of the plate antennas that have been fabricated. The test structures have been made for NMOS as well as PMOS transistors. Plate antennas have been made out of salicided and not salicided gate poly and in each metal layer. Figure 4.1 shows an example of some poly plate antennas.

	Poly		Metal level				
	Salicid.	Unsalic.	1	2	3	4	5
Antenna ratio.	0	0		0	0	0	
	1000	1000	1000	1000	1000	1000	
	10000	10000	10000	10000	10000	10000	
	100000	100000	100000	100000	100000	100000	100000

Table 4.1: Overview of the layers and antenna rations of plate antennas. Reference structures without antenna are marked by Antenna ratio 0.

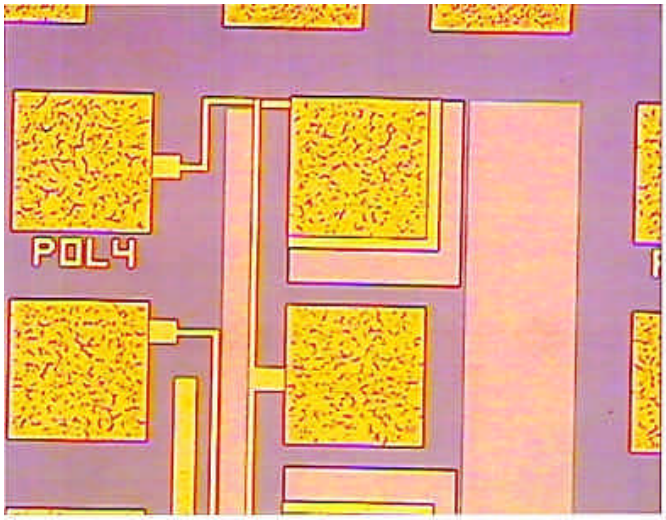


Figure 4.1: Example of poly plate antennas.

Perimeter intensive comb antennas

Perimeter intensive comb antennas, which have a long edge length and minimized area, are primarily used to assess charging in conductor patterning processes (poly and metal etch). Charging only occurs within a limited time frame of the total etch as schematically represented in figure 4.2. There is no charging as long as the etched layer is continuous, i.e. during main etch. After clearance of the layer, in the overetch regime, the antenna becomes isolated (floating) and charging is possible. During the patterning process the conductor lines remain covered with photoresist hence only the sidewalls of the patterned features can collect charge. Therefore, by using a comb antenna the charge collection during the patterning process can be maximized. In the presence of microloading effects, the narrow spacings in the comb clear later than the open area surrounding the antenna and a so-called latent (or giant) antenna exists temporarily with a charge collection area much larger than the actual sidewall area [10]. With respect to plasma charging, this time frame in the etch process is referred to as the latent antenna regime. In the overetch regime the charging gradually decreases as the collecting area formed by the “feet” of the sidewalls gradually decreases to the final etched profile.

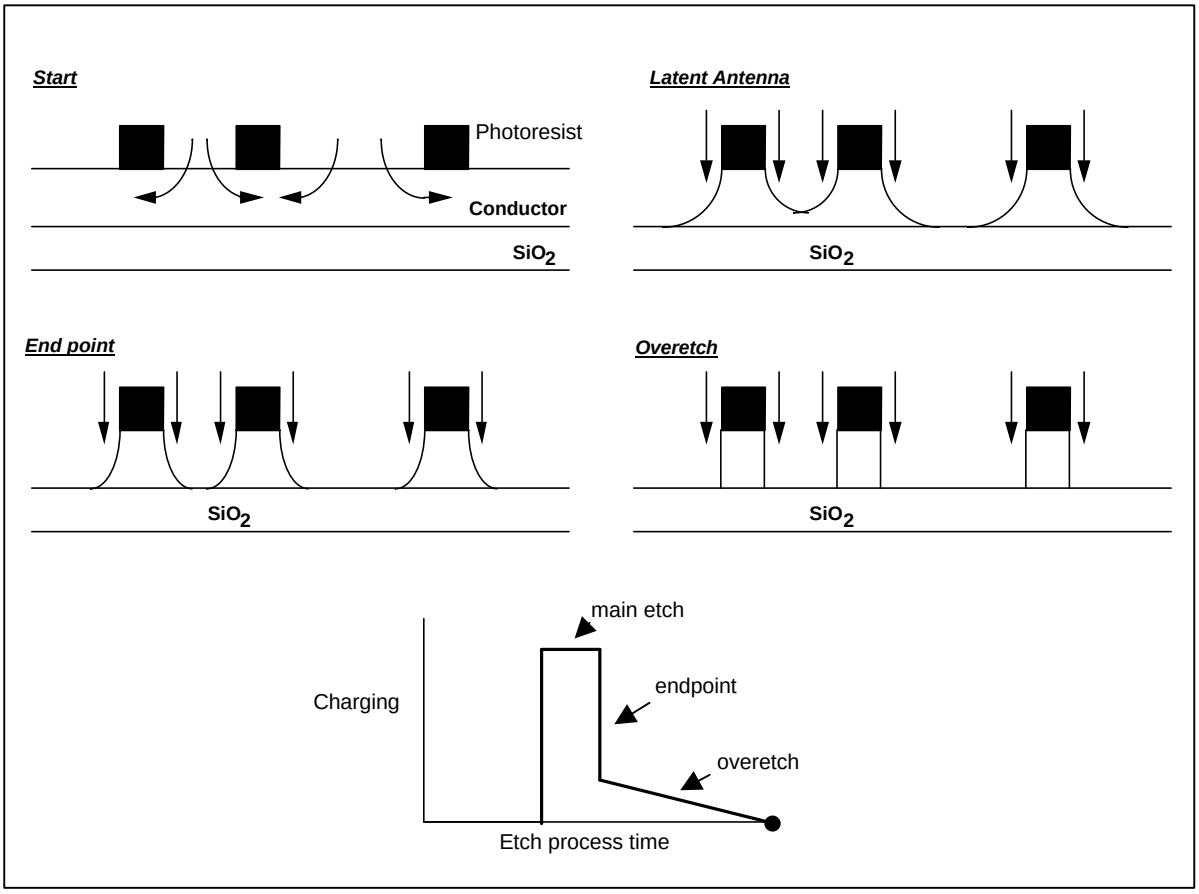


Figure 4.2: Charging only occurs within a limited time frame of the total etch process.

Minimizing the area of a comb antenna is often not practically feasible, so in general a comb antenna will be sensitive to area damage as well. A more useful approach is then to adjust the width of the comb fingers to achieve equal antenna area ratios for corresponding plate and comb antenna sizes. In this way the plasma damage contributions of non-relevant area-dependent process steps can be more easily distinguished from the contribution of the patterning step of interest. Patterning causes additional damage over the background level of area dependent steps as quantified by the plate antennas.

The spacing between the conductors in the comb should be varied in order to catch all relevant effects with respect to so-called “aspect ratio dependent charging”. Plasma charging increases with increasing aspect ratio (narrower spacing) through the electron shading [11] phenomenon acting in the latent antenna regime. On the other hand, the related “extended” electron shading [12] phenomenon acting in the overetch regime tends to increase with decreasing aspect ratio (wider spacing) [13].

Table 4.2 gives an overview of the comb antennas that have been fabricated. Comb antennas have been made out of salicided and not salicided gate poly and in each metal layer. The test structures have been made for NMOS as well as PMOS transistors. Figure 4.3 shows an example of some metal comb antennas.

Gatetype	Poly		Metal level				
	Salicid.	Unsalic.	1	2	3	4	5
Antenna ratio							
Large spacing 5 µm	1000	1000	1000	1000	1000	1000	
	10000	10000	10000	10000	10000	10000	10000
	100000	100000	100000	100000	100000	100000	100000
Minimal spacing 0.6 µm	1000	1000	1000	1000		1000	
	10000	10000	10000	10000	10000	10000	10000
	100000	100000	100000	100000	100000	100000	100000

Table 4.2: Overview of the layers and antenna ratios and line spacing of comb antennas.

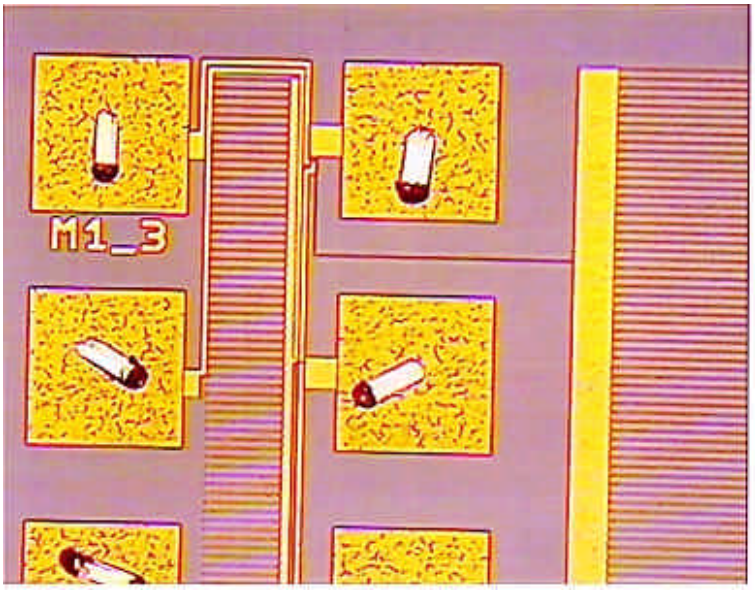


Figure 4.3: Example of large spacing metal comb antennas.

.

Charging damage from contact/via hole processing

Charging damage from contact/via hole processing (i.e. dielectric etch, contact/via fill related steps) can be assessed by contact/via antennas. The antenna consists of a bottom conductor plate on which the holes are landing, and on top a conductor plate by which the holes are covered. The size of the antenna is varied by varying the number of vias/contact holes. The influence of the top and bottom plates should be minimized or at least be kept fixed. Relevant parameters are furthermore the size of the holes, their aspect ratio, the density of holes. Table 4.3 gives an overview of the contact and via antennas that have been fabricated. Via antennas have been made in each via layer. Figure 4.4 shows an example of some via antennas.

Number of contacts, via's	Contacts	Via level			
		1	2	3	4
On small plate	1	1	1	1	1
	100	100	100	100	100
	1000	1000	1000	1000	1000
On large plate	1	1	1	1	1
	100	100	100	100	100
	1000	1000	1000	1000	1000
	10000	10000	10000	10000	10000

Table 4.3: Overview of the layers and numbers of contacts or vias on the contact and via antennas

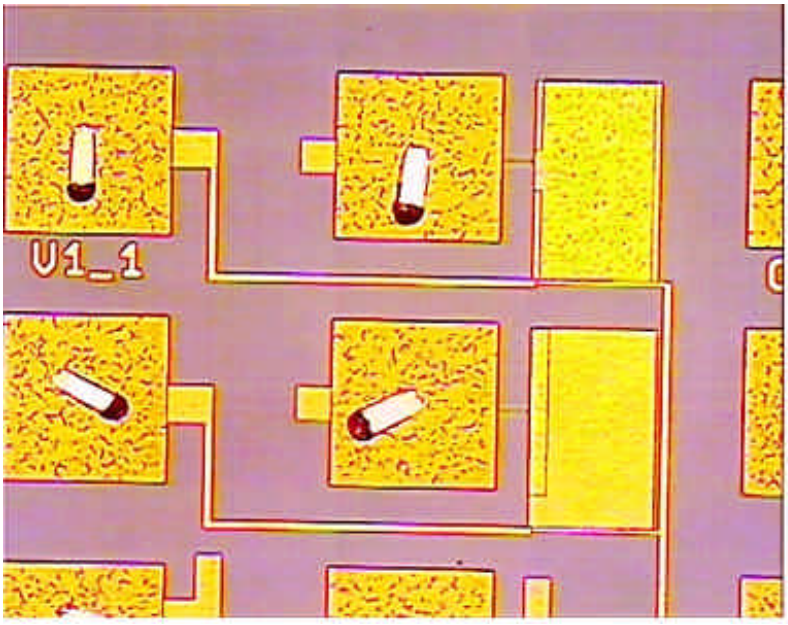


Figure 4.4: Example of vias on large plate antennas.

Cumulative antennas, sensitive to all process steps of a complete flow

Cumulative antennas are sensitive to all process steps of a complete flow. Typically, this type of antenna is used as a plasma damage monitor for the full process. The global charge collecting antenna consists of elementary antennas at all levels with the maximum antenna ratio allowed by the design rules. Under these circumstances, the charging sensitive element must remain essentially damage free at end of line. Table 4.4 gives an overview of the cumulative contact-via antennas and poly and metal comb antennas that have been fabricated.

Cumulative comb structure	poly	Metal1	Metal2	Metal3	Metal4	Metal5
Antenna ratios Set1	1000	1000	1000	10000	1000	10000
Antenna ratios Set2	300	300	300	300	300	300
Cumulative Contact & Via structures	Contacts	Via1	Via2	Via3	Via4	
Numbers of contacts and vias Set1	1000	1000	1000	1000	1000	
Numbers of contacts and vias Set2	83	83	83	83	83	

Table 4.4: Overview of the layers, antenna ratios and numbers of contacts or vias on cumulative contact-via antennas and poly and metal comb antennas.

4.2.3 Specific MOS structures

Fuses

The timing of the electron shading effects is extracted with help of the Transient (T)-fuse scheme [14]. This T-fuse consists of several closely spaced conductor (metal or poly) lines which are separated only after clearance of the open regions of the conductor layer in these narrow spacings, i.e. at the end of the latent antenna regime. The fuse is placed either between the antenna and the gate or alternatively between the antenna and ground. The test structure is able to separate the contributions of main etch and overetch. With the fuse in between the gate and the antenna, the metal antenna is disconnected from the gate of the transistor at the end of the etching step. The gate of the transistor only experiences the impact of the plasma damage during the beginning and main part of the metal etching. This structure has been fabricated

only in the metal 1 level. It is fabricated with an antenna ratio of 1000 and 10000. Figure 4.5 shows an example of a metal antenna with a fuse.

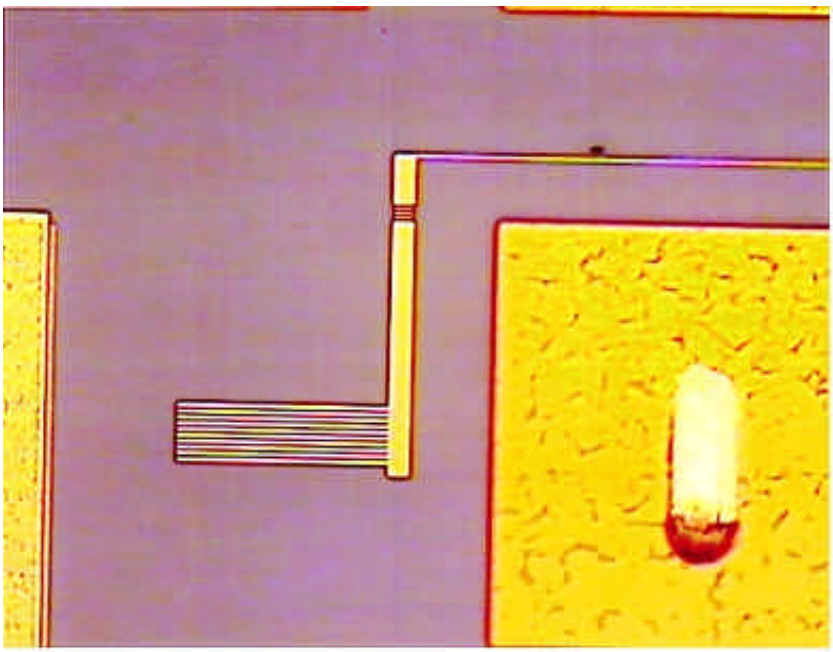


Figure 4.5: Antenna with transient (T)-fuse scheme.

Proximity antenna structures

The impact of nearby conductors is investigated by metal and poly antennas enclosed by large extra combs which can be either grounded or left floating.

4.2.4 Specific metal-insulator-metal capacitor structures

Figures 4.6a and 4.6b show a schematic and a SEM cross-section of the metal-insulator-metal capacitor built on metal 2. Standard metal 2 deposition of a Ti/TiN AlCu TiN stack also acted as the bottom plate of the capacitor. A thin capacitor dielectric film (PECVD nitride) was deposited followed by the deposition of a PVD TiN/AlCu/TiN top plate. The top plate was patterned first and etched with a selective etch chemistry with etch stop on the capacitor dielectric. The bottom electrode (metal 2) was subsequently patterned and the stack of residual dielectric and metal 2 was etched in sequence followed by metal photo resist strip. The process flow continued with the standard CMOS back-end flow sequence of inter-metal dielectric deposition, via 2 formation followed by deposition and patterning of metal 3 and finished with passivation. The metal-insulator-metal capacitor was implemented using the above scheme in a CMOS back-end flow with less than 0.5 µm line and 0.5 µm space design rules. The bottom plate is always extending the top plate by 2 µm.

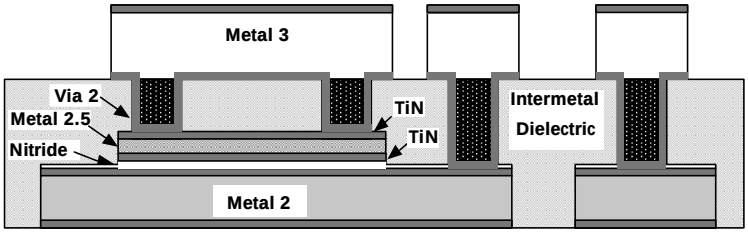


Figure 4.6a: Schematic diagram of Metal-Insulator-Metal capacitor integrated on metal2 of a CMOS backend.

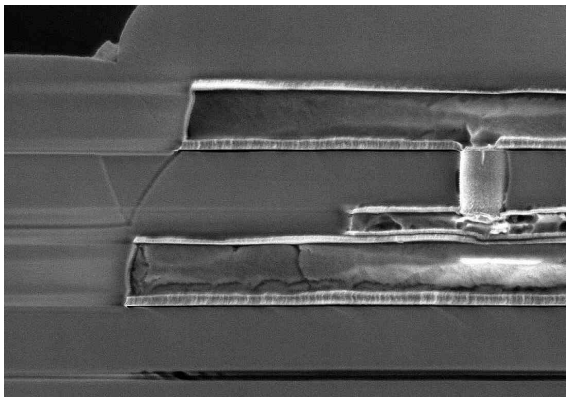


Figure 4.6b: Cross-section SEM picture across a typical metal-insulator-metal capacitor. Note that the shallow via 2 provides connectivity between top plate of capacitor to metal 3. Connection of bottom plate (metal 2) to metal 3 is provided by a deep via2 not depicted.

More than 400 systematically arranged test structures to simulate the metal-insulator-metal capacitor interconnection layout design, have been designed and processed in order to investigate and characterize the effect of charging during the processing on the yield and reliability of the metal-insulator-metal capacitor.

Metal-insulator-metal capacitors have been designed with an area of 5 by 5 µm, 20 by 20 µm and 20 times 5 by 5 µm in parallel.

Antennas connected to top and bottom plate were designed in a range from 20µm² up to 120.000 µm² or consist just of a small connection (1.36 µm²) to the next metal level.

To investigate the antenna perimeter effect, antennas have been designed both as combs and plates. Figure 4.7 shows a layout example of a metal-insulator-metal capacitor test structure.

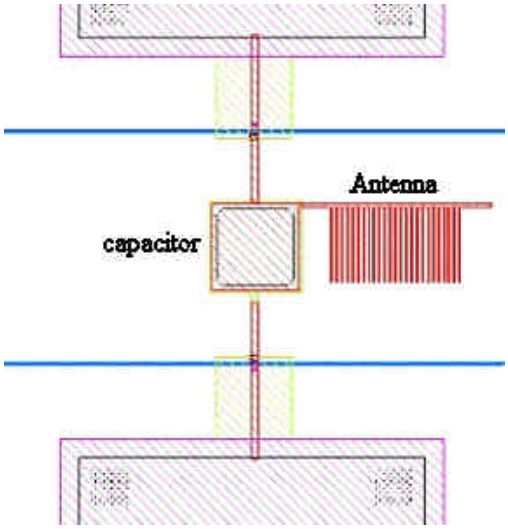


Figure 4.7: Example of layout of metal-insulator-metal capacitor with metal comb antenna connected to the top plate of the capacitor.

The metal-insulator-metal capacitor structures were designed in such a way that they also can be processed without the manufacturing of the diodes. Like this a lot more information was achieved about the effect of protective diodes, without having to design more test structures.

4.2.5 Specific flash memory structures

Flash memory cells have been discussed in detail in the previous chapter. Since the cells consist of a complex combination of tunnel and gate oxides, combined with a dense frame of metal inter connections, the probability that these cells may be affected by plasma damage is eminent. In order to investigate if the plasma damage affects the flash memory cells the appropriate test structures have been designed, manufactured and measured. The test structures include structures to generate plasma damage as well as possible protective structures to prevent plasma damage. This chapter is describing these test structures.

Plasma damage can be defined as any deviation of the normal operational behavior of the cell due to plasma processing steps. The occurrence of plasma damage therefore is correlated to the size of conductors connected to the cell (antenna) and there is a correlation with the size or number of cells in parallel. Comparing this cell with a normal MOS in terms of plasma damage, the most likely point for plasma damage to occur is the gate oxide between the control gate and the silicon.

Other failure mechanisms are less likely:

- Although there is the thin tunnel oxide between the floating gate and the silicon, the floating gate is never connected so it does as such not have an antenna.

- Damage of the interpoly oxide between control gate/ program gate and floating gate is less likely since the oxide is quite thick (15-25 nm) and is expected to be less vulnerable.

Test structures have been designed with comb antenna structures with variations in the size of 100, 1000 and 10000 μm^2 connected to the control gate. In addition, to investigate the impact of plate antennas, extra metal plates have been added to the testers that already had a metal comb antenna with an area 1000 and 10000 μm^2. To investigate the impact of the gate oxide area, cells have been put in parallel: in combination with the different antenna sizes, a matrix of test structures with 1, 16, 128 and 1028 parallel cells has been fabricated.

4.3 Plasma process induced damage protection schemes

To protect both CMOS and metal-insulator-metal capacitor devices against plasma process induced damage, typically protection diodes in parallel with the gate electrode are used to bypass the charge collected by the antenna. The polarity of the damaging tunnel current through the gate oxide is, however, not fixed and depends on the plasma reactor, the timing of the plasma process, and the place on the wafer. This hampers the choice of the correct protection method.

4.3.1 Diode protection

After the concerning process step/module, the gate electrode is connected to one or two diodes to avoid additional damage from subsequent steps/modules. For example, in pwell an n+ diode is connected. The idea is that the plasma current during back end of line processing will flow through the diode hence the maximum gate potential is limited to a safe value. In the case of positive charging the diode will be reversed biased, in this case its current carrying capability (hence maximum gate potential) depends on the diode area and on the level of illumination from the plasma and the temperature of the silicon substrate. To assure effective protection in both charging polarities, a second p+/nwell diode can be added in anti-parallel. Diodes can only be effective from the metal 1 level on.

The use of protection diodes has important consequences for characterization of the plasma damage as it limits the type of biases that can be applied at the gate. This is a problem in particular in the double-diode protection case, in which counterbiasing must be applied at the nwell, requiring an extra probe pad in the test module. “Bipolar” types of measurement such as charge pumping and CV measurements are simply not possible in diode protected structures as one of the two diodes would become forward biased.

MOS devices are protected against plasma damage by the use of diodes put in parallel with the gate oxide. For MOS transistors a single diode can be used since they are operated always with the same gate polarity. For the flash memory cell this is more complex: during writing and reading there is positive voltage on the control gate while during erasing there is a negative voltage on the control gate. The application of a back to back diode is a possible way

to maintain functionality and to protect against plasma damage. The technology in which the flash memory cells are manufactured, provides a number of layers to construct a back to back diode capable of supporting the field applied to the cells during writing, reading and erasing of the cells. Test structures have been manufactured to investigate to efficiency of the back the back diode. A 1000 or a 10000 µm² comb antenna has been connected to the control gate. The diode was placed in parallel over the control gate. As such the efficiency of the diodes can be measured as a function of the antenna size. The described test structures have been designed both with a single memory cell and with 1024 cells in parallel. In this way, the impact of the gate oxide area can be assessed.

4.3.2 Capacitor protection

After the concerning process step/module, the gate electrode is connected to a relatively large field-overlapping capacitor. The idea is that the antenna ratio for the remaining back-end-of-line processes will be much closer to one and the extra back-end-of-line damage will be minimal. Also this type of protection is only effective from metal 1 on (although one can think of a kind of reference test structure in which the capacitor is connected already at poly level, i.e. from the very first damaging step on).

For electrical evaluation purposes it is mandatory that the capacitor is located inside a well type opposite to the well type of the device to which it is connected. This is to avoid disentangle the tunnelling current, which is drawn to the capacitor when the gate oxide is sufficiently thin, from the regular current of the charging sensitive device.

As mentioned in the previous paragraph, the control gate of flash memory cell is being operated both with a positive and a negative polarity. Besides the use of the back to back diode, protection against plasma damage with a capacitor is a valid alternative. The requirements are that the capacitor oxide is thick enough to withstand the voltages used over the control gate in a reliable way. On the other hand, the oxide should not be too thick: it should still be capable of transferring the charges accumulated on the control gate during the patterning and the depositions steps.

As with the diode, a 1000 or a 10000 µm² comb antenna has been connected to the control gate. A 400 µm² or a 10000 µm² capacitor was placed in parallel over the control gate. As such the efficiency of the capacitors can be measured as a function of the antenna size. The described test structures have been designed both with a single memory cell and with 1024 cells in parallel. In this way, the impact of the gate oxide area can be assessed.

4.3.3 Metal bridges

When a conductor to be connected with a charge sensitive device becomes larger, the risk for inducing plasma damage increases. A simple but efficient way to reduce the size of the conductor without affecting the functionally, is to introduce a small metal bridge on the next metal level. As such, the large conductor is interrupted close to the charge sensitive device. On the next metal layer the connection is restored with a short metal line. In this way, during the processing of the conductor, only a small part is directly connected to the charge sensitive device. At the moment when the connection is restored on the next metal level, the initial conductor is already isolated and cannot pick up charges any longer.

This protection is investigated in combination with MOS devices metal-insulator-metal capacitor and flash memory cells.

4.4 Evaluation techniques

4.4.1 Surface charges

Non-contact surface charge measurements provide a fast and indirect way to detect plasma process induced damage. For our measurement, a plasma damage monitoring tool manufactured by Semiconductor Diagnostics Inc was used. This technique measures the build-up of the deposited plasma charges during plasma processing on top of an unpatterned SiO_2 film. With up to 6000 measured points per wafer, a map of detailed information about the charge distribution over the entire wafer can be made. Often this technique is compared with the use of antenna structures. This technique, however, is not detecting any effects induced by patterns. As such, the plasma damage monitoring measurements are rather complementary to antenna structures rather than a replacement of the antenna structures. A major advantage of this measurement technique is the very short learning cycle, and the possibility to do in-line monitoring on critical equipment. [8]. As an illustration, figures 4.8 compare the charge distribution mapping of a wafer processed with a uniform plasma (Figure 4.8a) with a wafer processed with a non-uniform plasma (figure 4.8b). Application of this measurement to detect and solve plasma damage issues is discussed in detail later in the work.

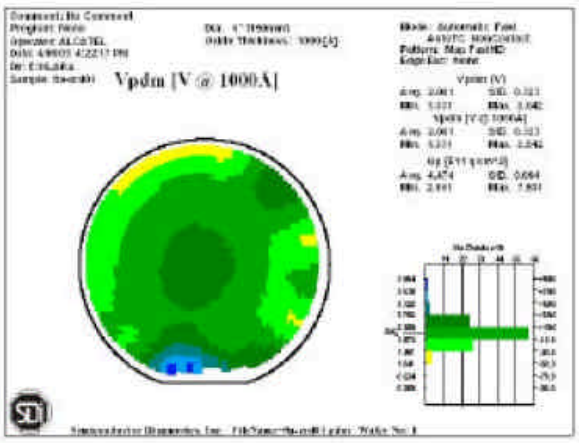


Figure 4.8a: plasma damage monitoring mapping of an oxide layer deposited with a uniform plasma. The surface voltage range is 2.3 V.

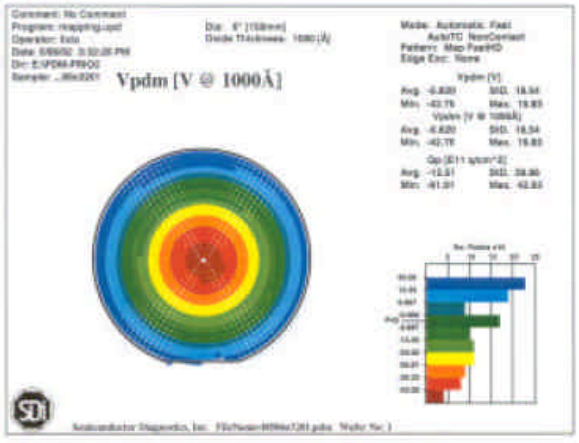


Figure 4.8b: plasma damage monitoring mapping of an oxide layer deposited with a non-uniform plasma. The surface voltage range is 62.6 V.

4.4.2 MOS Antenna structures

In this chapter, plasma damage was already defined as any parametrical change to a device induced by plasma processing. An evaluation has been made of the various parameters of the transistor connected to the antenna. In order to find the optimal measurement technique for the evaluation of plasma damage with antenna structures samples are prepared with deliberately induced plasma damage. As a test case metal 1 antenna structures with various sizes have been etched both in a high density plasma etcher and a medium density plasma etcher. This paragraph is describing the relation between the antenna/gate on threshold voltage (V_t), Voltage breakdown, charge pumping and leakage current measured through the gate oxide in between the floating gate and the substrate (gateleakage). Figure 4.9 is comparing the results of charge pumping, gateleakage and voltage breakdown measurements as a function of high density or medium density plasma etching and the antenna ratio of the test structure.

Each of the investigated measurement techniques has its own typical sensitivities advantages and disadvantages. Gate leakage measurement is fast, simple and shows a good sensitivity. Voltage breakdown is rather slow, simple but more sensitive than gate leakage measurements. Charge pumping takes even more time than voltage breakdown, it is complex but very sensitive. Gate leakage, voltage breakdown and charge pumping all show a good correlation with the antenna ratio and as such with the induced plasma damage.

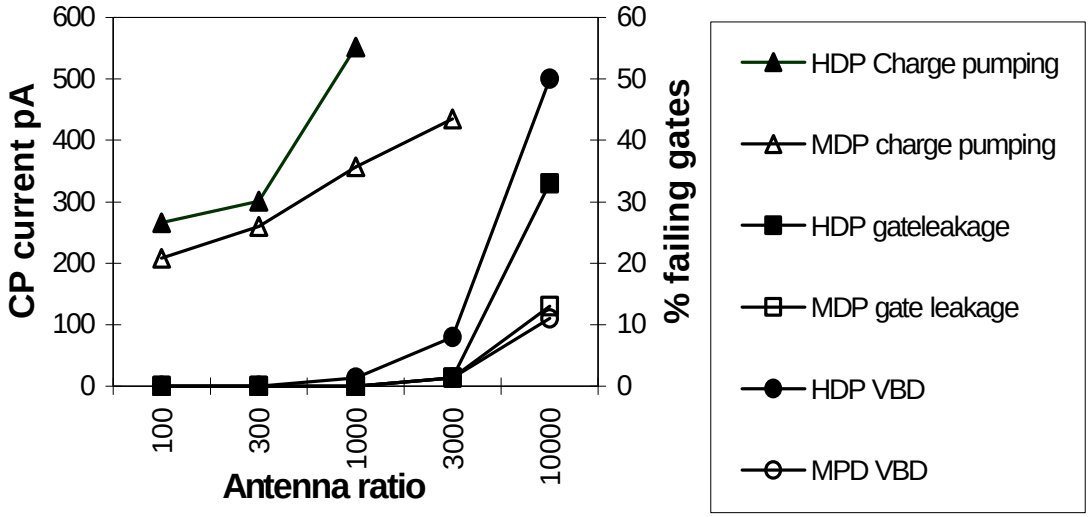


Figure. 4.9: Charge Pumping current, Voltage breakdown and gate leakage as a function of Antenna Ratio for the High Density Plasma and Medium Density Plasma metal etching.

A V_t measurement is fast, but is not sensitive. Figures 4.10a and 4.10b show the correlation between V_t measurements and gate leakage measurements. The V_t measurements show a poor correlation with gate leakage. For the NMOS devices, good unaffected gates show a gate leakage current of less than 0.1 nA. Damaged devices show a gate leakage current between 1

µA and 10 nA Very little measurements can be found in between 0.1 nA and 10 nA. This makes the evaluation of number of damaged devices rather straight forward. When evaluating V_t's, the damaged devices only show a minimal shift toward lower V_t. As such, V_t's cannot be used for monitoring plasma damage.

For PMOS devices there a distinction between affected and not affected devices around 10 A. The limit is less clear than in the case of NMOS. Also here, when evaluating V_t's, there is hardly any shift distinguishable between good or affected gates.

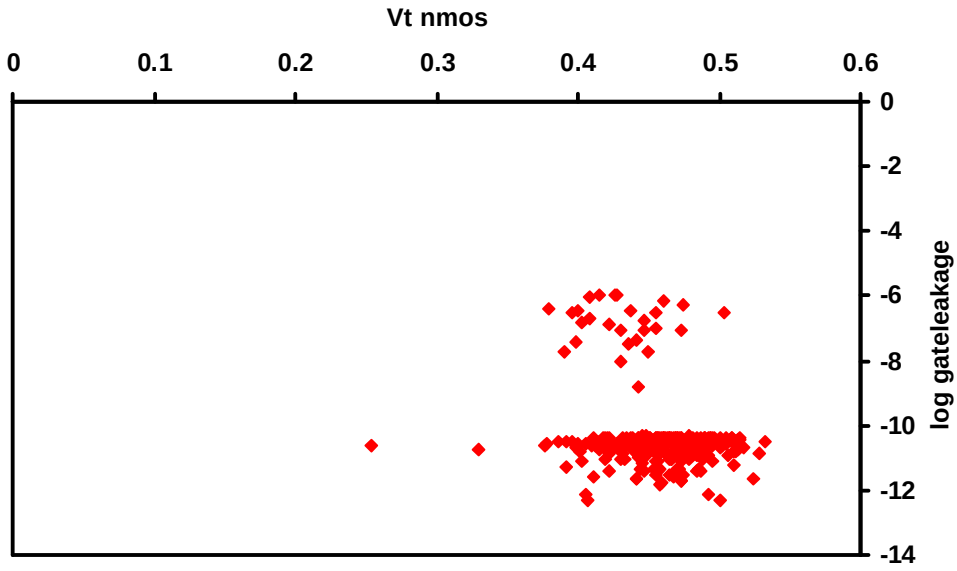


Figure. 4.10a: Correlation between gate leakage measurements and the V_t measurements for the NMOS devices.

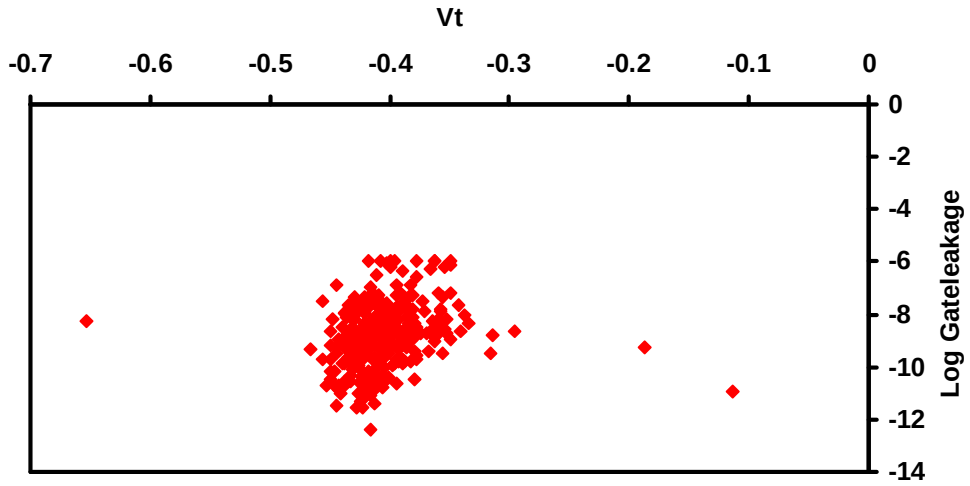


Figure. 4.10b: Correlation between gate leakage measurements and the V_t measurements for PMOS devices.

We can conclude that gate leakage measurements are the fastest and simplest measurements. Although it is not the most sensitive measurement, it is showing good correlation with the

more sensitive measurements like charge pumping and voltage breakdown. V_t measurements show no practical use for evaluating plasma damage on antenna structures.

4.4.3 Measurements on flash memory cells

Initial problems were observed with HIMOS™ cells when the erase V_t is being reduced from –3V to close to 0 V. In this case, the memory cell is losing functionality since there is no difference between written and erased state. Since this failure mechanism was occurring more on wafers processed with 4 layer metal compared to wafers processed with 2 layer metal, plasma damage was considered as a major cause for this problem. Figures 4.11a and 4.11b compare erase V_t distribution for 4 types of cells with 2 and 4 level metal processing.

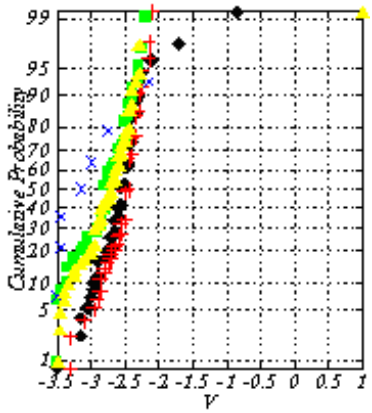


Figure 4.11a: Erase V_t distributions with 2 layers of metal processing.

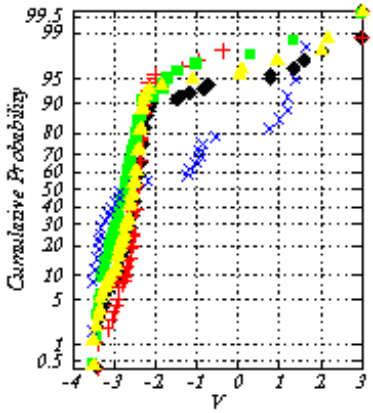


Figure 4.11b: Erase V_t distributions with 4 layers of metal processing.

Figure 4.12 shows the correlation between gate leakage measurements and the delta- V_t (Erase vs. Write) measurements. Good unaffected gates show a gate leakage current of less than 0.1 nA. Damaged devices show a gate leakage current between 1 µA and 10 nA. Very little measurements can be found in between 0.1 and 10 nA. Also here, this makes the evaluation of number of damaged devices rather straightforward. When evaluating the delta-V_t's, only a fraction of the damaged devices only show a significant reduction delta V_t. Also with the flash memory cells, the gate leakage measurements show to be a good tool to asses the plasma damage.

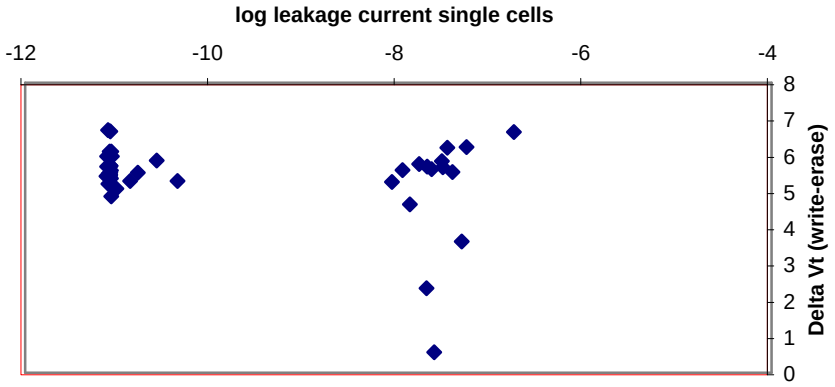


Figure. 4.12: correlation between gate leakage measurements and the delta V_t (Erase vs. Write) measurements for the flash memory devices.

4.5 Conclusion

This chapter is describing the test structures required for evaluation of plasma damage on MOS and metal-insulator-metal capacitor devices. As a special case of CMOS, test structures are developed to investigate the impact of plasma damage on HIMOS^{TM} flash memory cells.

Surface charges can easily be measured and mapped by the plasma damage monitoring tool. A major advantage of this measurement technique is the very short learning cycle, and the possibility to do in-line monitoring on critical equipment. It is not giving any information about topography relates sources of plasma damage.

To evaluate topography relates sources of plasma damage, various parametric measurements on transistors have been compared. V_t measurements are not sensitive enough. Charge pumping and voltage breakdown both are sensitive but rather complex or time consuming. Gate leakage measurements proved to be the most simple, fast and sensitive method. It can easily be applied on MOS transistors and on HIMOS^{TM} memory cells. On metal-insulator-metal capacitors the gate leakage measurement is replaced by the capacitor leakage measurement.

4.6 References

[1] Proceedings International Symposia on Plasma Process-Induced Damage (P2ID) 1-3, 1996-1998.

[2] H.C. Lin et al., “Characterization of plasma charging damage in ultra thin gate oxide”, IRPS, pp. 312, 1998.

[3] K. Hashimoto, “New phenomena of charge damage in plasma etching: heavy damage only through dense-line antenna,” Jpn. J. Appl. Phys., part 1, vol. 32, pp.6109-6113, 1993.

[4] K. Hashimoto, “Charge damage caused by electron shading effect,” Jpn. J. Appl. Phys., part 1, vol. 33, pp.6013-6018, 1994.

[5] Ken Stem, Joe Koeis, Gary Hueekel, Ernie Eld, Tom Bartush,Rob Groves, Nancy Greco, Dave Harame, and Ted Tewksbury,”High reliability metal insulator metal capacitors for silicon germanium analog applications,” Proceedings of BCTM,pp.191-194, 1997.

[6] S. Fang, S. Murakawa, and J.P. Mc Vittie, “A new model for thin oxide degradation from wafer charging in plasma etching,” in IEDM Tech. Dig., pp.61-64, 1992.

[7] T. Yamada, K. Eriguchi, Y. Kosaka, and K. Hatada, “Impacts of antenna layout enhanced charging damage on MOSFET reliability and performance,” in IEDM Tech. Dig., pp. 727-730, 1996.

[8] G. Horner, I.J. Gupta, “RaP2ID detection of charging damage with non-contact electrical analysis,” in Proceedings International Symposium on Plasma Process-Induced Damage (P2ID) 3, pp. 144-147, 1998.

[9] S. Krishnan, A. Amerasekera, S. Rangan and S. Aur, “Antenna device reliability for ULSI processing,” in IEDM Tech. Dig., pp. 601-604, 1998.

[10] Eds. K.P. Cheung et al., “On the mechanism of plasma enhanced. dielectric deposition charging damage” Proceedings of the 1st Intern. Symposium on Plasma Process Induced Damage, 1996.

[11] J.F. Verwij and J.H. Klootwijk, “Dielectric breakdown I: A review of oxide breakdown”, Chapter of American Vacuum society, Microelectronics Journal, 27, pp.611, 1996.

[12] A.B. Joshi et al., “Gate Oxide Thickness Dependence of RIE-Induced Damage on n-Channel MOSFET Reliability”, Proceedings of IEEE International Reliability Physics Symposium, Dallas, TX, 1996.

[13] M. Creusen, G. Groeseneken, J. Ackaert, E. De Backer, “Impact of reactor- and transistor-type on electron shading effects”, Proceedings of the 4th Intern. Symposium on Plasma Process Induced Damage, 1999.

[14] S. Krishnan, K. Brennan and G. Xing , “A transient fuse scheme for plasma etch damage detection”, Proceedings of the 3th Intern. Symposium on Plasma Process Induced Damage June 4-5, 1998.

Chapter 5

Impact of plasma damage on gate and tunnel oxides

The results in this chapter have partially been published in:

M. Creusen, J. Ackaert, E. De Backer and G. Groeseneken, “Impact of reactor- and transistor-type on electron shading effects”, Proceedings of the 4th International Symposium on Plasma Process-Induced Damage (P2ID), pp. 8-11, 1999.

Z. Wang, J. Ackaert, C. Salm, F.G. Kuper, M. Tack, E. De Backer, P. Coppens, L. De Schepper, B. Vlachakis, “Charging damage during Metal-Insulator-Metal capacitor manufacturing: test structures, model and prevention during design-phase”, IEEE Transactions on Electron Devices, pp. 1017-1024, June 2004.

Z. Wang, J. Ackaert, C. Salm & F.G. Kuper, “Plasma Process-Induced Latent Damage on Gate Oxide Demonstrated by Single-layer and Multi-layer Antenna Structures”, 8th International Symposium on the Physical & Failure Analysis of Integrated Circuits, pp. 220-223, 2001.

J. Ackaert, A. Lowe, S. Boonen, T. Yao, J. Van Houdt and L. Haspeslagh, “Plasma Damage in HIMOS^{TM} Non-Volatile Memory (NVM) cells”, ICICDT 2004.

J. Ackaert, Z. Wang, E. De Backer & C. Salm, “Correlation between Hot Carrier Stress, Oxide Breakdown and Gate Leakage Current for Monitoring Plasma Processing Induced Damage on Gate Oxide”, Proceedings International Symposium on Plasma Process-Induced Damage, pp. 45, 2002.

Z. Wang, J. Ackaert, C. Salm , E. De Backer, G. van den Bosch and W. Zawalsky, “Correlation between Gate Leakage Current Measurement and Reliability Measurement in Monitoring Plasma Process-Induced Damage on Gate Oxide”, Proceedings International Symposium on Physics and Failure Analysis, pp. 242, 2002.

For the last 2 papers in this list, the authors are indebted to Geert van den Bosch and Wade Zawalsky of IMEC. Part of the work was funded by the Flemish Institute for Science and Technology (IWT) in project PINDA (Plasma Induced Damage), AUT/990241, in collaboration with IMEC, Belgium.

5.1 Introduction

A more fundamental understanding of the effects of plasma processing and the loss of MOS reliability requires a comprehensive study of plasma process induced damage. First of all, it is necessary to asses the level of charging that both a gate oxide and other dielectrics for metal-insulator-metal capacitor applications can withstand.
For thin dielectrics such as gate oxides and tunnel oxides, the charge and discharge level can be defined by using Fowler-Nordheim tunnelling, so that a safe and critical level of charge or discharge can be defined. Besides that, the important electron shading effect has been studied. The classical and extended electron shading-effect in both NMOS and PMOS transistors has been compared for two different metal etchers. It has been investigated what the impact is of the aspect ratio of the used antenna pattern and also the reactor- and transistor-type on the type of electron shading effect that is dominating during an etching process. In addition to the effects on MOS devices also the effects on non volatile memory applications are investigated and discussed.
A low level of Fowler-Nordheim tunnelling can affect the gate oxide without being detectable by gate leakage measurement. This level of damage is called latent damage. It is investigated if latent damage effect can anyhow be detected by leakage measurements by comparing the impact of multi layer antennas with the impact of single layer antennas.
Plasma damage affects not only yield but also reliability and device lifetime. The correlation is described between gate leakage degradation and HC degradation.

The following paragraphs describe the impact of plasma damage on gate and tunnel oxides. Special attention is paid to

- The interaction between plasma density and the effect of electron shading and extended electron shading.
- The occurrence and detection of latent damage in gate oxides.
- The impact of plasma damage on non-volatile memory cells.
- The correlation between hot carrier stress, oxide breakdown and gate leakage current for monitoring plasma processing induced damage on gate oxide.

5.2 Impact of reactor- and transistor-type on electron shading effects

In this paragraph electron shading effects are discussed as a function of reactor- and transistor-type. It was found that the classical electron shading-effect, which occurs during the latent antenna regime, dominates for the current high-density plasma-reactors, independent of the transistor-type used. However, in less dense plasma-reactors, the so-called extended electron shading-effect, which occurs during the overetch regime, can overrule the classical electron shading-effect depending on the transistor type. In this paragraph, we study the classical and extended electron shading effect in both NMOS and PMOS transistors and this for two different metal etchers. We will show that both electron shading effects take place during a metal etch process. However, not only the aspect ratio of the antenna pattern used but also the reactor- and transistor-type determine which electron shading effect dominates during an etching process.

In a first lot, nMOSFETs (L=0.5 µm W=0.8 µm) were processed up to metal 2 using a 0.5 µm CMOS technology. The lot included a split on metal 1 etching between a High Density Plasma (HDP) metal etcher and a medium density plasma MERIE-type metal etcher. In a second lot, nMOSFETs and pMOSFETs were fabricated using a 4-level metal 0.35 µm CMOS technology. In this lot, both nMOSFETs and pMOSFETs were etched in the medium density plasma MERIE-type metal etcher. All metal etch-processes were based on a Cl_2/BCl_3 chemistry. The 7 nm gate oxides were grown in an O_2 ambient at 925°C. The inter metal dielectric in the first lot was a PECVD/SOG/etchback/PECVD stack. In the second lot the inter metal dielectric-stack was composed of a high density plasma oxide combined with a PETEOS-layer followed by CMP. Before measurement the wafers were annealed in a forming gas ambient for 60 minutes at 450°C.

To study the electron shading effects, metal plate- and comb-antennas have been used as described in the previous chapter. The timing of the electron shading effects was extracted, for the metal comb-antenna, with help of the Transient (T)-fuse scheme as previously described.

The yield loss of the NMOS-devices as a function of antenna area ratio for both the high density plasma-etcher and the medium density plasma etcher is plotted in figures 5.1. The overall plasma damage level is higher for the high density plasma-tool compared to the medium density plasma-tool. For the high density plasma-etcher, the 0.8 µm comb antenna devices show a higher yield loss than the 1.6 µm antennas as a result of the classical electron shading effect.

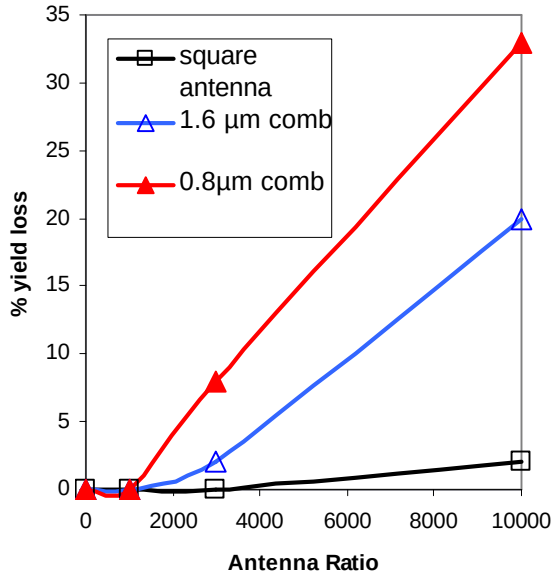


Figure 5.1a: The classical electron shading effect is observed for the high density plasma etcher (NMOS devices).

However, for the medium density plasma-MERIE reactor the 1.6 µm spaced antenna devices show higher yield loss, which is probably due to the extended electron shading effect. Finally, only minimum plasma damage is observed for the square plate antenna-devices for both metal etchers.

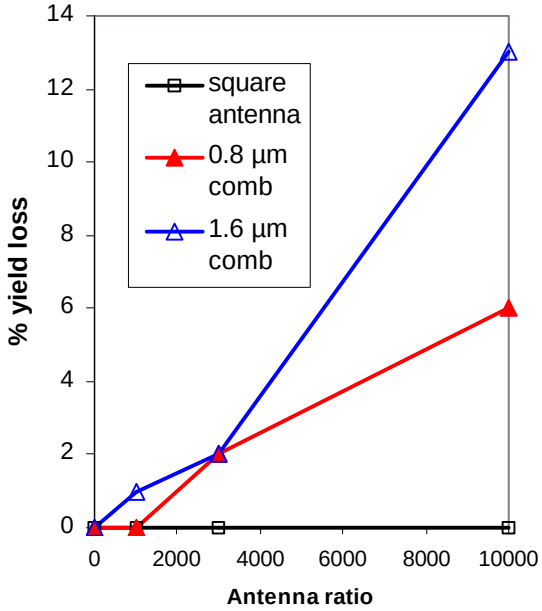


Figure 5.1b: The extended electron shading effect is observed for the medium density plasma etcher (NMOS devices).

These results can be explained as follows. In case of the classical electron shading effect, the transistor gate is charged positively during the latent antenna regime of the etching process due to different ion- and electron-velocity distributions at the wafer surface. Only the highly anisotropic ions are able to penetrate the high aspect ratio trenches. Plasma density and classical electron shading effect are thus directly correlated. On the other hand, the ion flux impinging on the wafer is significantly increased for the current high density sources. Consequently, for the high density plasma etcher we observe the classical electron shading effect. In case of the less dense MERIE-plasma, the classical electron shading effect is diminished and overruled by the opposite extended electron shading effect which occurs during the overetch regime of the etching process.

The yield loss as a function of the antenna perimeter for both NMOS and PMOS devices, etched with the medium density plasma-MERIE etcher, is plotted in figure 5.2. The overall plasma damage-level is lower for the NMOS devices than for the PMOS devices. As in the previous lot, the extended electron shading effect results in more damage for the larger spaced (5 µm) NMOS comb devices. However, for the PMOS devices the classical electron shading effect is observed. These electron shading effects are also observed at other metal levels and in other lots.

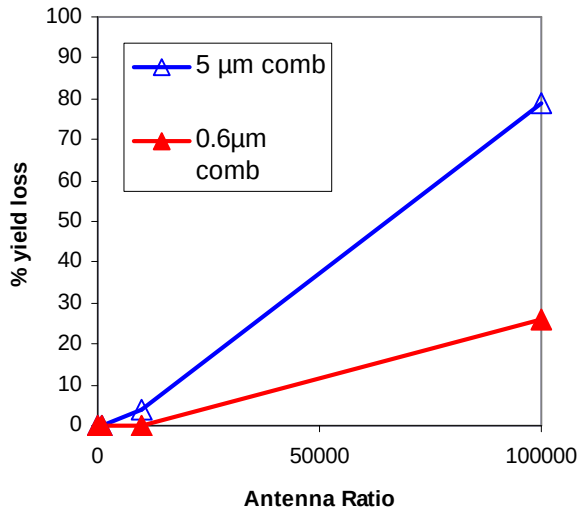


Figure 5.2a: The extended electron shading effect is observed for the NMOS devices etched with the medium density MERIE-type plasma etcher.

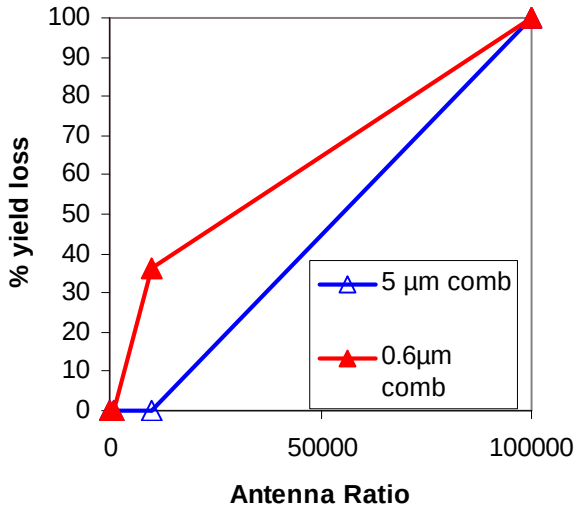


Figure 5.2b: The classical electron shading effect is observed for the PMOS devices etched with the medium density MERIE-type plasma etcher.

The timing of the NMOS charging events was studied with help of the T-fuse scheme. The T-fuse results (see table 5.1) show that the charging of the NMOS devices occurs during the overetch regime of the metal etch process and during the inter metal dielectric deposition process. This correlates with the results of Hasegawa et al. [1] who showed that the breakdown failure rate induced by the extended electron shading effect increased with over-etch time.

Device:	Yield loss(%)
T-fuse between 10 mm antenna and gate	0
T-fuse between 10 mm antenna and substrate	13

Table 5.1: Timing of the NMOS charging events for the MERIE-type reactor (not the same lot as in Figure 5.2a.

The gate oxide stress current and voltage is usually determined by the intersection of the plasma IV-curve and the gate oxide Fowler- Nordheim curve. For the current thin oxides, the plasma is thereby modelled as an imperfect current source [2]. However, the NMOS and PMOS IV-characteristics are completely different for the two electron shading effects. For example, in case of the classical electron shading effect, the positively charged gate leads to a maximum oxide stress for the accumulated PMOS-devices. However, for the NMOS devices the positive charging leads to depletion effects in the underlying p-type substrate and consequently to less oxide stress. In addition, the resulting higher gate voltage for the NMOS devices leads to a less repulsive field for the electrons and thus to more neutralization of the excess ion current. Both effects diminish the classical electron shading effect during the

metal-etching process. This explains why we observe the classical electron shading effect only for the PMOS devices in case of the medium density plasma-tool.

Summarizing, the experiment with different reactors and transistors lead to the following conclusions. In case of the high density plasma-reactors the classical electron shading effect dominates the extended electron shading effect due to the higher plasma density, for both NMOS and PMOS devices. In case of the medium density plasma-reactors the situation is not so unambiguous. For the NMOS devices, the extended electron shading effect results in maximum gate-oxide stress whereas the classical electron shading effect results in less gate-oxide stress due to depletion effects in the p-type substrate. On the other hand, for the PMOS devices the classical electron shading effect overrules the extended electron shading effect as a result of depletion effects in the n-well. In this chapter we have shown that both classical and extended electron shading phenomena occur during a metal-patterning process. Which type of electron shading effect is dominating during the etching process depends not only on the aspect ratio of the antenna structures but also on the used reactor- and transistor-type.

5.3 Plasma process-induced latent damage on gate oxide

In this section, a simple experimental method is proposed to demonstrate directly the pure latent damage without any impact of additional defects generated by constant current stress revealing technique. Single-layer antenna test structures are used to evaluate plasma process induced damage of each stand-alone plasma process step. And the cumulative plasma process induced damage of a few plasma processes is evaluated by multi-layer antenna test structures. The test structure is considered to suffer plasma process induced damage when there is an antenna present during a certain plasma process. Therefore, the number of layers determines how many plasma process steps are used to introduce damage to the structures. The used single-layer test structure and multi-layer test structure are described in section 2, and the experimental results are presented and discussed in section 3. The experimental data clearly demonstrate the existence of latent damage, since the multi-layer structures that were exposed, but did not fail from antecedent plasma process induced damage, are more susceptible to subsequent plasma process compared with fresh single-layer structures that are free from antecedent plasma process induced damage.

Test structures have been used as described in the previous Chapter. In this study, some wafers with single-layer test structure and multi-layer test structure have been exposed to a 0.35 μm CMOS backend-of-line process. After that, the charging sensitive antenna structures [3] of these wafers are evaluated by leakage measurements as described in chapter IV. In Figure 5.3, the schematic of the multi-layer test structure used in our study is depicted. The antenna ratio of all poly and metal comb antennas is 10000. The contact or via antenna has 1000 contacts or 1000 vias on a small plate.

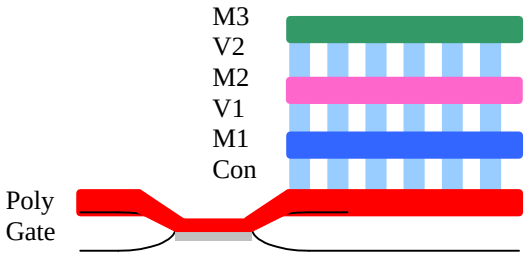


Figure 5.3**:** Cross-section of a combined poly, contact, metal1, via1, metal2, via2, and metal3 (Poly+Con+M1+V1+M2+V2+M3) multi-layer antenna structure. This structure is supposed to suffer from plasma process induced damage during the whole process.

In order to compare quantitatively the plasma damage to different structures we use the failure fraction, defined as the number of failed devices divided by the total number of investigated devices. In the following, the equation is derived, which can extract the induced failure fraction by one antenna level out of the failure fractions of two consecutive multi-layer antennas.

$Y_{Poly+con+M1}$ is defined as the yield of Poly+Con+M1 structures, $Y_{Poly+con}$ as the yield of Poly+Con structures, and $Y_{M1,ML}$ as the yield of M1 plasma process with multi-layer antenna structures.

Since

$$Y_{poly+con+M1} = Y_{poly+con} \cdot Y_{M1,ML} \quad (5.1)$$

and

$$F_{M1,ML} = 1 - Y_{M1,ML} \quad (5.2)$$

therefore

$$F_{M1,ML} = 1 - ((1-F_{poly+con+M1})/(1-F_{poly+con})) \quad (5.3)$$

where $F_{poly+con+M1}$ is the failure fraction of Poly+Con+M1 structures, $F_{poly+con}$ is the failure fraction of Poly+Con structures, and $F_{M1,ML}$ is the M1 plasma process induced failure to multi-layer antenna structures. Because $F_{poly+con+M1}$ and $F_{poly+con}$ can be obtained from the experiment, $F_{M1,ML}$ can be calculated. With equation (3), one can extract the failure fraction induced by one plasma process to multi-layer antenna structures out of the failure fractions of two corresponding multi-layer antenna structures.
The failure fraction comparison of different antenna structures is presented in Table 1. $F_{'one\ process',SL}$ is one plasma process induced failure fraction to single-layer antenna structures. $F_{'multi-\ processes',ML}$ is a serial plasma processes induced failure fraction to Multi-layer antenna structures. In the table, $F_{'one\ process',SL}$ and $F_{'multiprocesses',\ ML}$ are experimental data. $F_{'one\ process',ML}$ is the calculated induced failure fraction of one plasma process to multi-layer antenna structures, which is extracted from the failure fractions of two corresponding multi-layer antenna structures with equation (3).

In Table 5.2, $F_{poly+con+M1}$ = 40% and $F_{poly+con}$ = 32%, with equation (3), we can get $F_{M1,ML}$ = 12%. Therefore, in our case, 12% of the devices that are not damaged during poly and contact plasma processes are damaged during M1 plasma process. However, the same M1 plasma process causes only 0.4% failures on the single-layer antenna structures. The reason for this phenomenon is plasma process-induced latent damage. After the poly and contact plasma process, 32% of the total devices failed and the other 68% of the total devices are survived. Those 68% devices do not show failure, but they are much weaker because of the latent damage generated by plasma processes. Therefore, they are more susceptible to subsequent M1 plasma process. On the other hand, single-layer antenna structures have only one antenna. The devices are fresh and have not suffered plasma damage from antecedent plasma processes. They are more robust. This is the reason why the failure fraction of the M1 plasma process to multi-layer antenna structure is much higher than to single-layer antenna structure. With equation (3), the failure fraction of other plasma processes to multi-layer antenna structure are also calculated and listed in Table 5.1. The comparison of the calculated values to the measured failure fraction of the same plasma process to the single-layer antenna structure reveal the existence of the plasma induced latent damage: Due to the existence of latent damage in multi-layer structures, the values of $F_{'one\ process',ML}$ are higher than that of $F_{'one\ process',SL}$.

F 'one process ',SL		F 'multi-process ',ML		F 'one process ',ML (calculated)	
Fpoly,SL	17%	Fpoly,ML	17%	Fpoly,ML	
Fcon,SL	10%	Fpoly+con,ML	32%	Fcon,ML	18%
FM1,SL	0.4%	Fpoly+con+M1,ML	40%	FM1,ML	12%
FV1,SL	1%	Fpoly+con+M1+V1,ML	44%	FV1,ML	7%
FM2,SL	6%	Fpoly+con+M1+V1+M2,ML	66%	FM2,ML	39%
FV2,SL	0.4%	Fpoly+con+M1+V1+M2+V2,ML	65%	FV2,ML	-3%
FM3,SL	4%	Fpoly+con+M1+V1+M2+V2+V3,ML	86%	FM3,ML	60%

Table 5.2**:** Failure fraction comparison by layer. Since multi-layer calculated values are exceeding measured single-layer values the occurrence of latent damage is demonstrated.

In order to make the comparison more clear, the failure fraction of the same plasma process to the single-layer structure and multi-layer structure are depicted in Figure 5.4. From the figure, it can be observed clearly that the same plasma process always causes more damage on multi-layer antenna structures. The reason is that the multi-layer structures are suffering from latent damage from antecedent plasma processing.

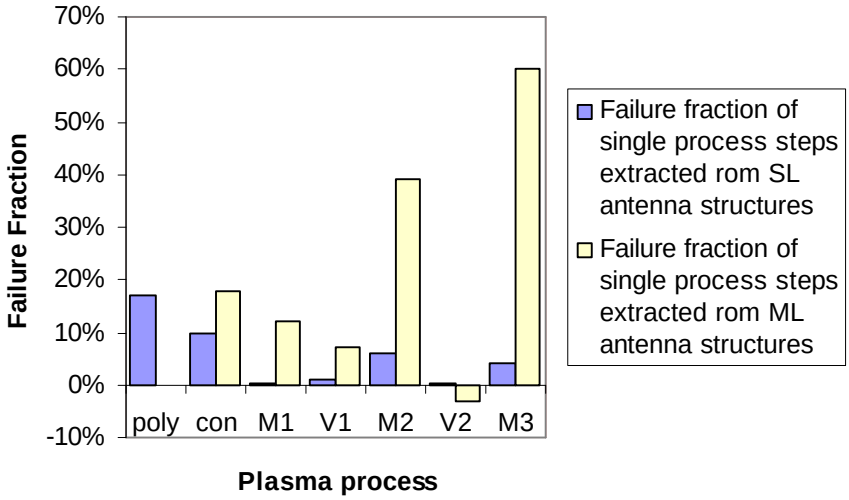


Figure 5.4: Comparison of one-plasma-process induced failure fraction between single-layer antenna structures and multi-layer antenna structures. The small negative value of V2 is probably caused by the noise of the measurement.

These results show that the failure fraction of the same plasma process to the multi-layer antenna structure is higher than the failure fraction of the single-layer antenna structure, which demonstrates the existence of the plasma-induced latent damage. Hence, a plasma process may generate not only active damage but also latent damage, which may show itself and cause reliability problems in the subsequent processes and in the field.

5.4 Plasma Damage in HIMOS^{TM} Non-Volatile Memory cells

In this section, plasma induced damage on floating gate based non-volatile memory cells is studied. Since the cells consist of a complex combination of tunnel and gate oxides, combined with a dense frame of metal interconnect, the probability that these cells may be affected by plasma damage is evident. In order to investigate if the plasma damage affects the flash memory cells, the appropriate test structures have been designed, manufactured and measured. The test structures include structures to generate plasma damage as well as possible protective structures to prevent plasma damage.

All experiments have been performed on 0.35 µm flash memory circuits based on the HIMOS^{TM} cell concept. The HIMOS^{TM} technology allows adding a full flash technology by adding only 2 masks to any technology capable of providing the voltage required for the cell operation. The HIMOS^{TM} NVM technology which is investigated is implemented in the AMIS I3T80 high voltage technology. Test frames have been manufactured containing 1, 16, 128 and 1024 cells. A tunnel oxide thickness in the range of 8 to 10 nm is used. The interpoly dielectric is 20 to 25nm.

Figure 5.5 shows the effect of the size of the antenna on the yield of the HIMOS^{TM} cells. The yield is calculated as the % of cells with gate leakage less than 1 nA. The yield on the single cells is heavily affected by the large antennas and less by smaller antennas. For the multiple cells, this effect is decreasing while increasing number of cells in parallel.

A single pad connected to the control gate on one metal level (M4 pad) is not enough yet to damage a fraction of the gate oxide of single cell test structures significantly. Pads on all metal levels (M234 pads) already damage single cells and the 16 cell structure. With more metal connected, the damage gets more severe. This is a clear evidence that plasma damage is affecting the HIMOS^{TM} NVM cells.

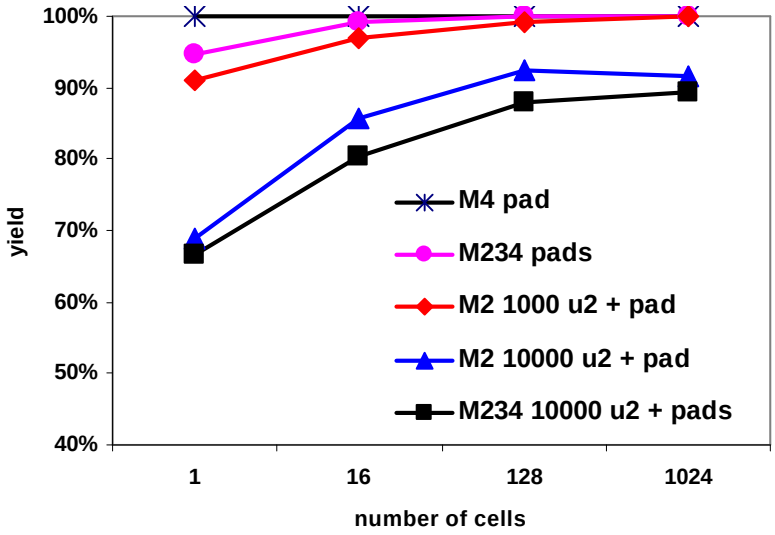


Figure 5.5: plot of the effect of the size of the antenna and the number of cells on the yield of the HIMOS^{TM} cells.

Figure 5.6 shows the impact of the protection against plasma damage by small 400 µm² capacitors and large 10000 µm² capacitors on single cells and 1K HIMOS^{TM} cells. Plasma damage has been generated with an antenna of 1000 µm² and 10000 µm². Again, the yield is calculated as the % of cells with gate leakage less than 1 nA. The 1K cell frame remains unaffected by the plasma damage generated by the 1000 µm² antenna. The single cell with the 10000 µm² antenna yields the lowest for both capacitors. The 400 µm² capacitor is too small for the 10000 µm² antenna: yields are low for both the single and the 1K cells. For the 10000 µm² capacitor, yields are much higher for the 1K cells with the 10000 µm² antenna. The single cell with 1000 µm² antenna decreased in yield. Probably here the yield is affected by the plasma damage collected by the patterning of the large capacitor.

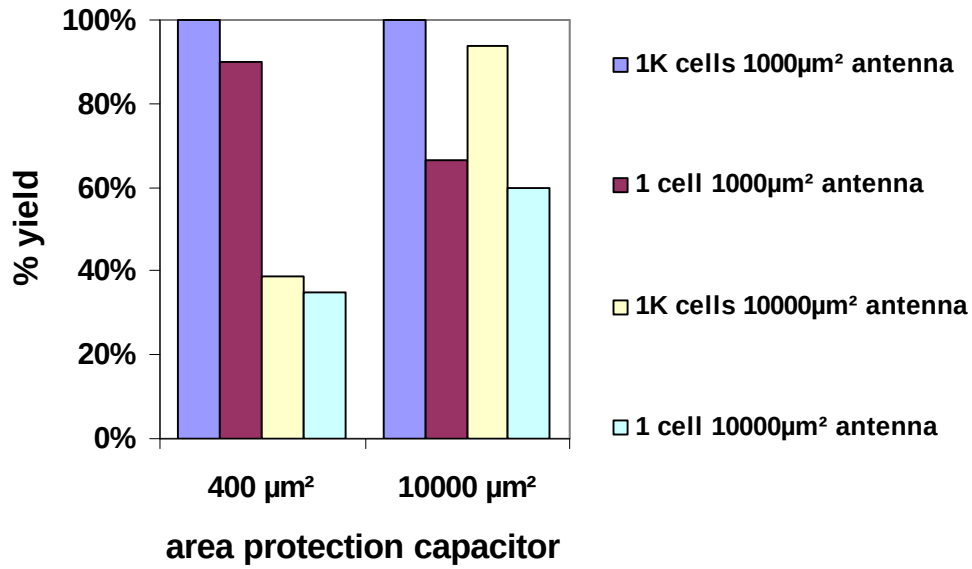


Figure 5.6: Yield of single and 1K cells with 1000 µm² and 10000 µm² antennas on the control gate protected by 400 µm² and 10000 µm² capacitors.

As an alternative to capacitor protection, back to back diodes were evaluated. With the double back to back diode parallel over the control gate, the yield is 100%. Even for the single cells with a large antenna, not a single gate is leaking. The double back to back diodes prove to be very effective to protect HIMOS^{TM} cells against plasma damage.

Besides the gate leakage, also the impact of plasma damage on the behaviour of the HIMOS^{TM} cell functionality has been investigated. Figure 5.7 shows the delta V_t (V_t write – V_t erase) plotted against the leakage of the control gate for the single cells. Unaffected cells show a gate leakage of less then 0.1 nA. The write-erase delta V_t has a very narrow distribution around 6V. Affected cells all show a gate leakage between 10 nA and 1 µA. The write-erase delta V_t of affected cells has a very wide distribution. For some cells this delta is close to 0: these cells show no longer a V_t difference between the written or erased state. As such these cells are no longer functional. The cells that still have a good delta V_t but have a high leakage current are a concern for reliability.

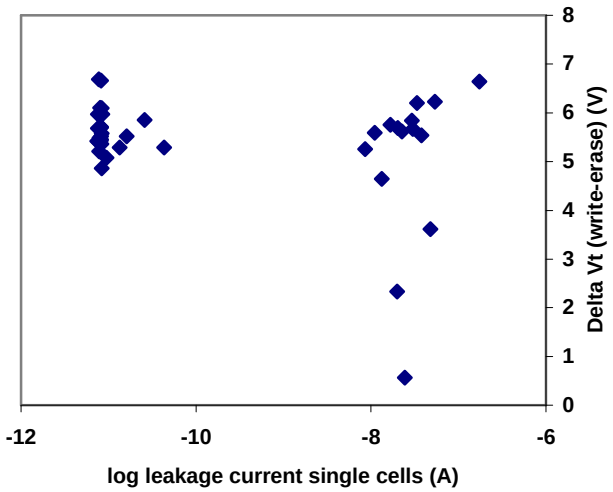


Figure 5.7: The delta V_t (Write-Erase) plotted against the leakage of the control gate for the single cells.

When implementing the diode on the control gate in the layout of the memory cell, it was observed that the diode was not providing the expected protection. Only by adding an extra diode to the program gate, the plasma damage could be completed. Figure 5.8 is comparing the delta V_t distributions of the unprotected cell, the cell with the diode protection on the control gate and the cell with a diode both on the control and program gate.

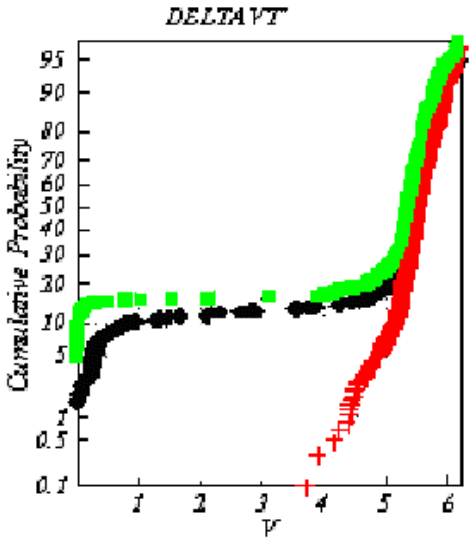


Figure 5.8: Delta V_t distributions of the unprotected cell (o), the cell with the diode protection on the control gate (*) and the cell with a diode both on the control and program gate (+).

The reason for the failures is similar to plasma damage reported earlier on metal-insulator-metal capacitors [4]. With only one gate connected to the substrate and the other one left floating but with a significant area of metal exposed to a plasma, a substantial potential can be

built over the dielectric capable of affecting oxides thicker than 20nm. With both gates protected with a diode, this risk no longer exists.

In conclusion, the yield on the single cell test structures is heavily affected by the large antennas and less by smaller antennas. For the multiple cell test frames, this effect decreases with increasing number of cells. This is a clear signature that plasma damage is affecting HIMOS^{TM} cells. Charges collected by one single pad connected to the control gate on each metal level is already enough to damage a fraction of the gateoxide of single cell test structures significantly. With more metal connected to the CG, more charges are collected from the plasma and the damage gets more severe. The capacitors used as a protection against plasma damage are not effective to protect HIMOS^{TM} cells. With a double back to back diode parallel over the control gate, the gate oxide yield of the HIMOS^{TM} cells is 100%. Even for the single cells with a large antenna, not a single gate is leaking. The double back to back diodes prove to be very effective in order to protect HIMOS^{TM} cells against plasma damage. For cells affected by plasma damage a number of gates are leaky but are still having a good functionality. As with MOS transistors, the gate leakage measurements prove to be more sensitive than V_t measurements. A fraction of the leaky cells clearly show an affected functionality: there is hardly a difference left between the written and the erased state. Even though the interpoly oxide is 20 to 25nm, there is still a risk that it is affected by plasma damage. It is required to protect both control gate and program gate with a diode in order to prevent plasma damage completely independent from interconnect schemes.

5.5 Correlation between hot carrier stress, oxide breakdown and gate leakage current.

In this chapter it is the intention to correlate two reliability parameters: hot-carrier degradation behaviour and voltage break down of the gate oxide due to the accumulated damage during plasma processing.
Hot-carrier stress typically occurs under high silicon and relatively low oxide field conditions. In the technologies that are considered here, the dominant degradation mechanism is interface trap generation in NMOS and electron trapping in PMOS [5]. This is clearly different from the high-field oxide degradation case where the dominant degradation mechanism finally leading to oxide breakdown is bulk trap generation [6]. The higher density of bulk oxide electron traps in devices that suffered from plasma charging damage might influence the trapping behaviour under hot-carrier stress conditions. For this reason it is worthwhile to investigate if and to what extent hot-carrier degradation rate will be influenced by the accumulated damage during plasma processing.

In this chapter, we do not only compare the HC stress but also oxide breakdown results with the fast gate leakage current measurement.

If a clear correlation could be found between gate leakage and both HC degradation and oxide breakdown, it would prove that plasma damage is affects not only yield but also reliability and lifetime of the devices. On top it would suggest that gate leakage current can be used to monitor device reliability under plasma stress, thereby saving a lot of measurement time.

Test structures with a metal antenna ratio of 1K, 10K and 100K have been used, as described in chapter 3. Evaluation of the plasma damage has been performed according to the procedures described there.

Special precautions in the measurement setup have to be taken to reach high enough accuracy with the low-level leakage measurements. As shown in figure 5.9, the devices with $I_{g,}$leak higher than 0.1 nA diverge from the intrinsic slope, indicating extrinsic damage induced by plasma charging.

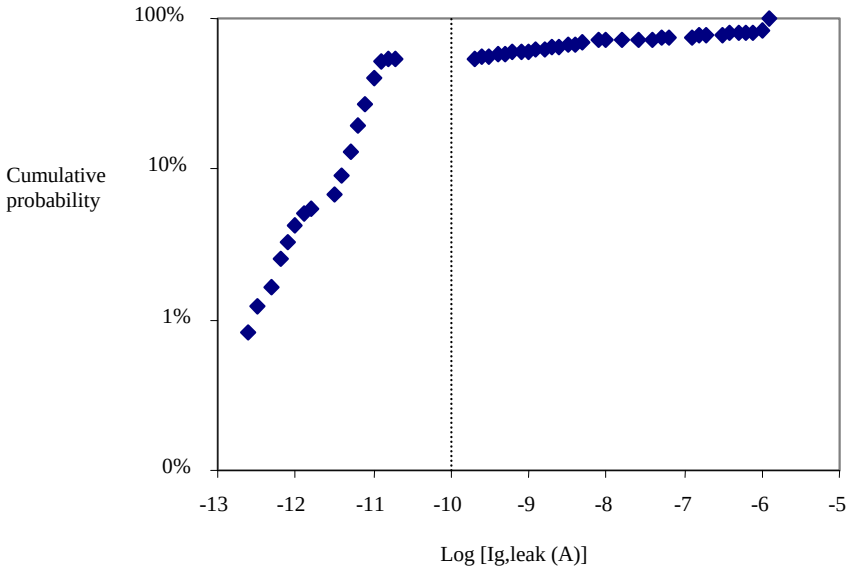


Figure 5.9: Distribution of gate leakage current. The devices with $I_{g,}$leak higher than 0.1 nA, diverge from the intrinsic slope, indicating extrinsic damage induced by plasma charging.

Hot carrier (HC) stress was performed on pMOSFETs. The pMOSFET was stressed at V_d=-6.5V, V_g=-1.9V. This V_g was chosen to maximize bulk current I_b. The HC lifetime, defined as the time the MOSFET takes for a 10 mV V_t shift, is compared with the initial I_g,leak at V_g = 3.6V measured on the same devices.

The oxide breakdown measurement was performed with a ramping voltage stress on more than one thousand devices. The gate oxide was stressed on 3.5V, corresponding with 4.7MV/cm for 1 second (step 1), 4.0V or 5.3MV/cm for 1 second (step 2), 4.5V or 6.0MV/cm for 1 second (step 3) and so on until 7.0V or 9.3 MV/cm for 1 second (step 8). In between each step gate leakage is measured. The results are plotted as a function of the initial $I_{g\text{-}leak}$ with V_g = 3.6V. This test method was developed to reveal latent as well as actual damage, for a wide range of gate oxide thicknesses in a very fast way.

The HC lifetime of pMOSFETs are shown in figure 5.10 for devices with metal 1 antennas. Figure 5.11 shows the HC lifetime of devices with metal 2 antennas. For failing devices with $I_{g\text{-}leak}$ above 0.1 nA, the HC lifetime is very much degraded. The devices with $I_{g\text{-}leak}$ in the range of 1 to 10 pA have longer lifetimes, while the others with $I_{g\text{-}leak}$ in the range of 10 pA to

0.1 nA value have clearly shorter lifetimes. A clear correlation is also observed with the size of the antenna connected to the gate of the transistor: increasing the antenna size by a decade, the $I_{g\text{-}leak}$ increases and the HC lifetime decreases by more than a decade. No impact of the spacing between the fingers is observed: in this case, electron shading is not the major cause of the degradation.

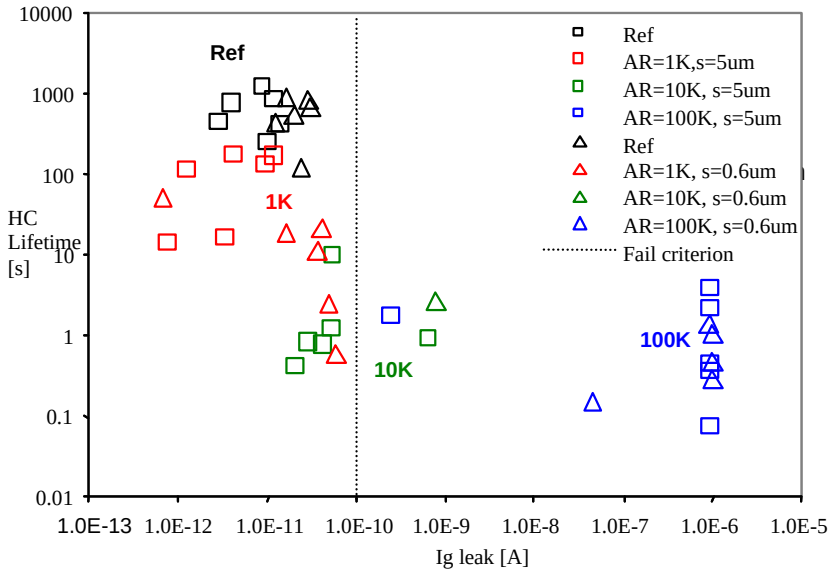


Figure 5.10: HC lifetime as function of gate leakage current measured at 3.6V on the same devices with Metal 1 finger antenna. S is the spacing between two fingers. For failing devices with $I_{g\text{-}leak}$ above 0.1 nA, the HC lifetime is degraded very much. The devices with $I_{g\text{-}leak}$ in the range of 1 to 10 pA have longer lifetimes, while the others with $I_{g\text{-}leak\ k}$ in the range of 10 pA to 0.1 nA value have clearly shorter lifetimes.

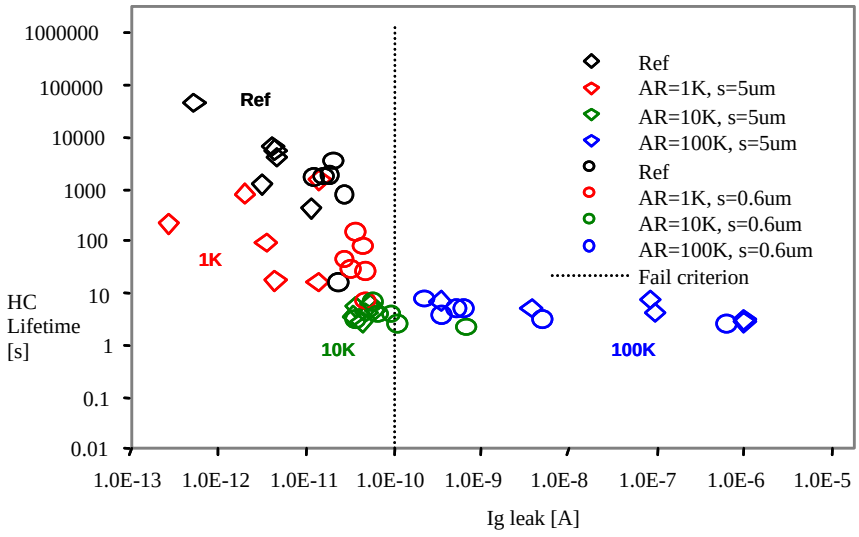


Figure 5.11. HC lifetime as function of gate leakage current measured at 3.6V on the same devices with Metal 2 finger antenna. S is the spacing between two fingers.

The strong correlation with the antenna ratio is considered as a signature for plasma damage being the cause of the degradation [7],[8]. In this case the plasma damage can be caused by the etching of the metal pattern or/and the oxide deposition by a high density plasma process. Plasma damage during metal etching is typically driven by electron shading and as such influenced by the spacing between the fingers of the antenna [9]-[12]. Since no impact is observed of the spacing between the fingers, it is more likely that the HC degradation is caused by the HDP oxide deposition. The mechanism and the required improvement on the HDP oxide deposition have been published before [13].

As is well known, the HC mechanism of pMOSFET under Ib, max stress condition is the channel shortening effect due to electron trapping near the drain region. This leads to increased drain current and reduced V_t. [14].

An almost 1:1 ratio between HC lifetime degradation and the antenna size is observed.

The results of breakdown measurement performed with a ramped voltage stress are compared with the gate leakage current on the same devices in figure 5.12 for devices with metal 1 antennas and figure 5.13 for devices with metal 2 antennas.

All the reference structures without extra antenna connected to the gate have an initial leakage current in the range from 10pA to 0.1 nA and break down after stress at 8MV/cm. The antenna devices that are as good as reference devices and with low initial $I_{g\text{-}leak}$ value also have breakdown values after stress at 8MV/cm, as shown in the figures. For the antenna devices with higher initial $I_{g\text{-}leak}$ value, there is a clear linear trend between $I_{g\text{-}leak}$ and the breakdown field.

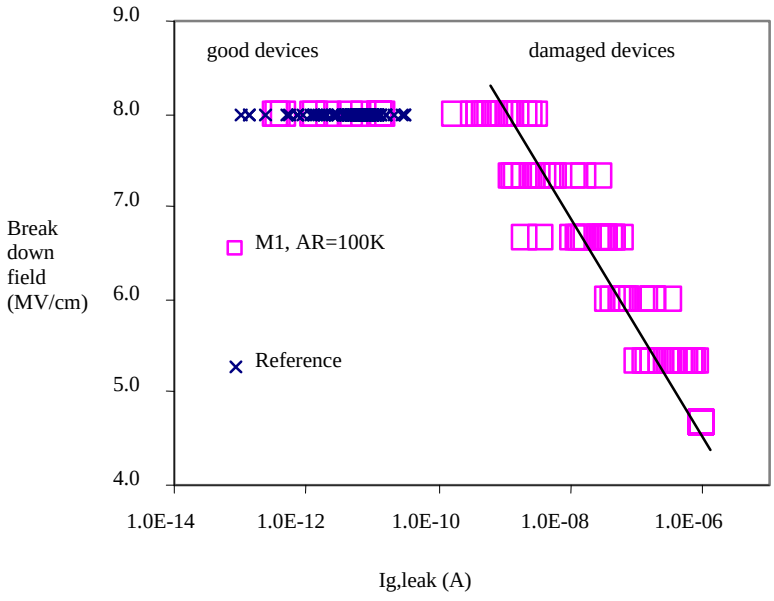


Figure 5.12: Breakdown field as function of initial gate leakage current measured on nMOSFETs with M1 finger antenna. Antenna devices with higher initial $I_{g\text{-}leak}$ value, there is a clear linear trend between $I_{g\text{-}leak}$ and the breakdown field.

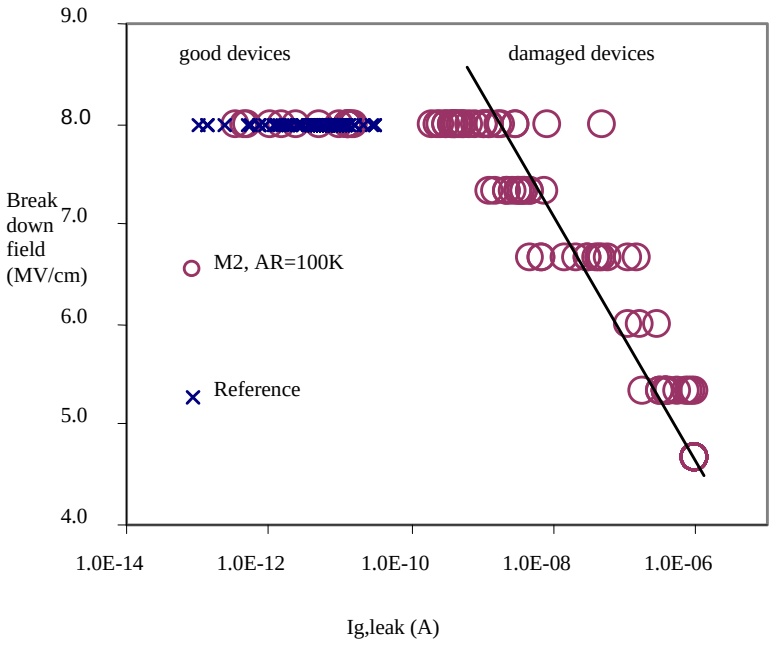


Figure 5.13: Breakdown field as function of initial gate leakage current measured on the same nMOSFETs with M2 finger antenna.

Summarizing, extensive voltage break down and channel hot-carrier stress characterisation has been done on PMOS antenna transistors in a 0.35 µm technology. A clear correlation is found between low levels of gate leakage and both HC degradation and oxide breakdown. This proves that plasma damage is affects not only yield but also reliability and device lifetime. We demonstrate that the value of the gate leakage current is not only a failure indicator in the region about 1 nA but also a good indicator of the reliability of the devices in the region between 1 pA and 1 nA. Thus, from the value of gate leakage current, one can estimate the impact of plasma damage on reliability of the devices, saving precious measurement time.

5.6 Conclusions

In this chapter electron shading effects are discussed as a function of reactor- and transistor-type. It was found that the classical electron shading effect, which occurs during the latent antenna regime, dominates for the current high-density plasma-reactors, independent of the used transistor-type. However, for less dense plasma-reactors the so-called extended electron shading effect, which occurs during the overetch regime, can overrule the classical electron shading effect depending on the transistor-type.

A new method is proposed to study plasma process-induced latent damage on gate oxide, by using single-layer antenna and multi-layer antenna structures. With this method one can study the plasma process induced damage without suffering from artefacts induced by traditional constant current stress method. The presented results show that the failure fraction of the same

plasma process to the multi-layer antenna structure is higher than that to the single-layer antenna structure, which demonstrates the existence of the plasma-induced latent damage. Hence, plasma process may generate not only active damage but also latent damage, which may show itself and cause reliability problem in the subsequent processes and application.

It is demonstrated that also the performance of non volatile memory cells is affected by plasma damage. The failure fraction of the NVM test structures is proportional to the antenna ratio. Large antennas increase the failure fraction while increasing the gate areas is decreasing the failure fraction. Both phenomena are a clear signature that plasma damage is affecting HIMOS^{TM} cells. A single pad connected to the control gate on each metal level is already enough to damage a fraction of the gateoxide of single cell test structures significantly. With more metal connected to the control gate, the damage gets more severe.

The capacitors used as a protection against plasma damage are not effective to protect HIMOS^{TM} cells.

With a double back to back diode parallel over the control gate, the gateoxide yield of the HIMOS^{TM} cells is 100%. Even for the single cells with a large antenna, not a single gate is leaking. The double back to back diodes prove to be very effective to protect HIMOS^{TM} cells against plasma damage.

For cells affected by plasma damage a number of gates are leaky but still have a good functionality. As with MOS transistors, the gate leakage measurements prove to be more sensitive than V_t measurements. A fraction of the leaky cell clearly show an affected functionality: there is hardly a difference left between the write and the ease status.

It is required to protect both control gate and program gate with a diode in order to prevent plasma damage completely independent from interconnect schemes.

Also the correlation is demonstrated between HC degradation, oxide breakdown and low levels of gate leakage and both. This proves that plasma damage is affects not only yield but also reliability and device lifetime

5.7 References

[1] A.Hasegawa, F. Shimpuku, M. Aoyama, K. Hashimoto and M.Nakamura, “Direction of topography dependent damage current during plasma etching”, in Proc. 1998 Int. Symp. Plasma Process Induced Damage (P2ID), pp. 168-171, 1998.

[2] H.Shin, K.Noguchi, C.Hu, “Modelling oxide thickness dependence of charging damage by plasma processing”, IEEE Electron Device Lett., vol. 14, no. 11, pp.509-511, 1993.

[3] J. Ackaert, E. De Backer, P. Coppens, and M. Creusen, “Plasma Damage Antenna Test Structure Matrix Description, Application for Optimization HDP Oxide Deposition, Metal Etch, Ar-Preclean and Passivation Processing in Sub-half Micro CMOS Processing”, 1st European Symposium on Plasma Process Induced Damage, pp. 70, 1999.

[4] J. Ackaert “Charging Damage in Floating Metal-Insulator-Metal Capacitors”, Proceedings 6th International Symposium on Plasma Process-Induced Damage, pp. 120-123, May 14, 2001.

[5] P. Heremans et al., “The mechanisms of hot carrier degradation”, in “Hot carrier design considerations for MOS devices and circuits”, Ed. C.T. Wang, Van Nostrand-Reinhold, 1992.

[6] R. Degraeve et al., “Degradation and breakdown in thin oxide layers: mechanisms, models and reliability prediction”, Microelectronics Reliability 39, pp. 1445, 1999.

[7] F. Shone, K. Wu, J. Shaw, E. Hokelek, S. Mittal and A. Haranahalli, “Gate oxide charging and its elimination for metal antenna capacitor and transistor in VLSI CMOS double layer metal technology”, Symposium on VLSI Technology, pp. 73, 1989.

[8] H. Shin, C. C. King, T. Horiuchi, and C. Hu, “Thin Oxide Charging Current During Plasma Etching of Aluminum”, IEEE Electron Device Letters, vol. 12, no. 8, pp. 404-406, 1991.

[9] K. Hashimoto, “New phenomena of charge damage in plasma etching: Heavy damage only trough dense-line antenna”, Jpn. J. Appl. Phys., vol 32, pp. 6109-6113, 1993.

[10] K. Hashimoto, “Charge damage caused by electron shading effect”, Jpn. J. Appl. Phys., vol 33, pp. 6013-6018, 1994.

[11] A. Hasegawa, F. Shimpuku, M. Aoyama, K. Hashimoto and M. Nakamura, “Direction of topography dependent damage current during plasma etching”, Proceedings of the 3th Intern. Symposium On Plasma Process Induced Damage, pp.168, 1998.

[12] M. Creusen, J. Ackaert, E. De Backer and G. Groeseneken, “Impact of reactor and transistor-type on electron shading effects” Proceedings of the 4th Intern. Symposium On Plasma Process Induced Damage, pp. 8-11, 1999.

[13] J. Ackaert, E. De Backer, P. Coppens, M. Creusen “Prevention of plasma induced damage on thin gate oxide of HDP oxide deposition, metal etch, Ar preclean processing in BEOL sub-half micron CMOS processing”, Proceedings of the 5th Intern. Symposium On Plasma Process Induced Damage, pp. 120-123, 2000.

[14] Pei Yao “Hot-Carrier Degradation Study of High Density Plasma (HDP) Oxide Deposition Process in Deep-submicron NMOSFETS”, SPIE, Vol. 3884, pp. 116-123, 1999.

Chapter 6

Impact of plasma and charging on thick (>20nm) dielectrics

The results in this chapter have partially been published in:

J. Ackaert, K. Bessemans, E. De Backer, “Charging induced damage by photoconduction through thick inter metal dielectrics”, Microelectronics Reliability, vol. 43, pp.1525-1529, 2003.

J. Ackaert , Z. Wang, E. De Backer, P. Coppens, “Charging Damage in Floating Metal-Insulator-Metal Capacitors”, 6th International Symposium on Plasma Process-Induced Damage, pp. 120-123, May 14, 2001.

Z. Wang, J. Ackaert, C. Salm, F.G. Kuper, M. Tack, E. De Backer, P. Coppens, L. De Schepper, B. Vlachakis, “Charging damage during Metal-Insulator-Metal capacitor manufacturing: test structures, model and prevention during design-phase”, IEEE Transactions on Electron Devices, pp. 1017-1024, June 2004.

J. Ackaert, Z. Wang, E. De Backer, P. Coppens, “Plasma Damage in floating Metal-Insulator-Metal Capacitors”, 8th International Symposium on the Physical & Failure Analysis of Integrated Circuits (IPFA), pp. 224-227, July 9-13, 2001.

Z. Wang, J. Ackaert, C. Salm, F.G. Kuper, “Charging Induced Damage on Complex-antenna Test Structures”, Proceedings Semiconductor Advances for Future Electronics (SAFE), Veldhoven the Netherlands, pp. 220-223, 2001.

6.1 Introduction

For thicker dielectrics, above 20nm, completely different failure mechanisms are observed, investigated and discussed. For metal-insulator-metal capacitor application, during the deposition of the dielectric, the temperature is limited because of the exposed metal. Under these conditions, oxide quality in terms of break down and reliability is marginal. Failure mechanisms different from those observed on thin thermal oxide can be expected and are investigated. Methods for protecting the metal-insulator-metal capacitor already implemented during the design phase are explored.
Very thick oxide depositions, above 500nm, are not as highly insulating as might be expected. Under certain conditions, also these oxides can transfer enough charges to affect thin gate oxides as will be shown in the last section of this chapter.

The following paragraphs describe the impact of plasma and charging on thick (>20nm) dielectrics.

Special attention is paid to

- The impact of plasma and charging on metal-insulator-metal-capacitor test structures, model and prevention during design-phase.
- Charging induced damage by photoconduction through thick inter-metal dielectrics.

6.2 Charging damage during Metal-Insulator-Metal capacitor manufacturing

In this section, charging induced damage to metal-insulator-metal capacitors, is discussed. The damage is caused by the build up of a voltage potential difference between the two plates of the capacitor. Currently a comprehensive physical model has been developed along with a practical measurement. A logarithmic model is derived, explaining the relation between the damage by this voltage potential and the ratio of the area of the exposed antennas connected to the plates of the metal-insulator-metal capacitor. This function makes it possible to predict damage in metal-insulator-metal capacitor devices with long interconnects.

To predict and automatically flag the potential charge-damage during the phase of IC layout design, new charge antenna design rules are proposed to apply while layout checking with the consideration of the interactions of the interconnection lines which become antennas connected to both plates of the metal-insulator-metal capacitor. Furthermore, layout solutions are introduced.

The dielectric in this metal-insulator-metal capacitor is much thicker than MOS gate oxides and the ratios of the antenna area vs. the capacitor area are typically low and even near unity. One therefore does not expect these capacitors to suffer from charging induced damage easily compared to thin gate oxides. However, previous studies found thicker oxides to be very sensitive to charging damage. It was argued that thicker oxides require higher voltage but less current to breakdown [1]. It was stated before that metal-insulator-metal capacitor s would be damaged only when the bottom plate is grounded and the top plate is connected to a large conductor (antenna) that is exposed to a plasma [2]. In this section it is shown that floating metal-insulator-metal capacitor s, without any substrate contact are very sensitive to charging induced damage: When both plates of the metal-insulator-metal capacitor are connected to a

large antenna with enough difference in area, severe charging induced damage occurs as soon as these antennas are exposed to charging. The impact of the size and the shape of the capacitor and the antennas is investigated and discussed.

Figures 6.1a and 6.1b show a schematic and a SEM cross-section of the MIM capacitor built on metal 2. Standard metal 2 deposition of a Ti/TiN AlCu TiN stack also acted as the bottom plate of the capacitor. A thin capacitor dielectric film (PECVD nitride) was deposited followed by the deposition of a PVD TiN/AlCu/TiN top plate. The top plate was patterned first and etched with a selective etch chemistry with etch stop on the capacitor dielectric. The bottom electrode (metal 2) was subsequently patterned and the stack of residual dielectric and metal 2 was etched in sequence followed by metal photo resist strip. The process flow continued with the standard CMOS back-end flow sequence of inter-metal dielectric deposition, via 2 formation followed by deposition and patterning of metal 3 and finished with passivation. The MIM capacitor was implemented using the above scheme in a CMOS back-end flow with less than 0.5 µm line and 0.5 µm space design rules. The bottom plate is always extending the top plate by 2 µm.

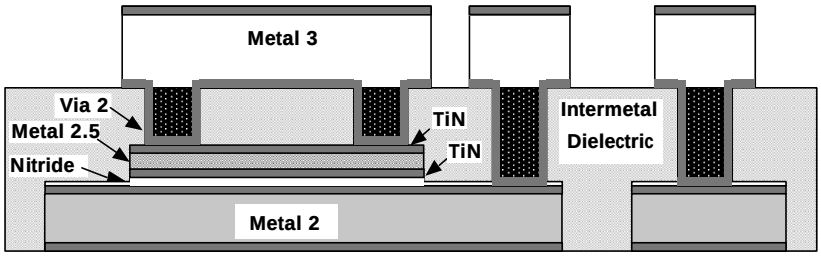


Figure 6.1a: Schematic diagram of metal-insulator-metal capacitor integrated on metal2 of a backend CMOS. The metal-insulator-metal capacitor has metal 2 of bottom plate, nitride as dielectric and metal 2.5 as top plate.

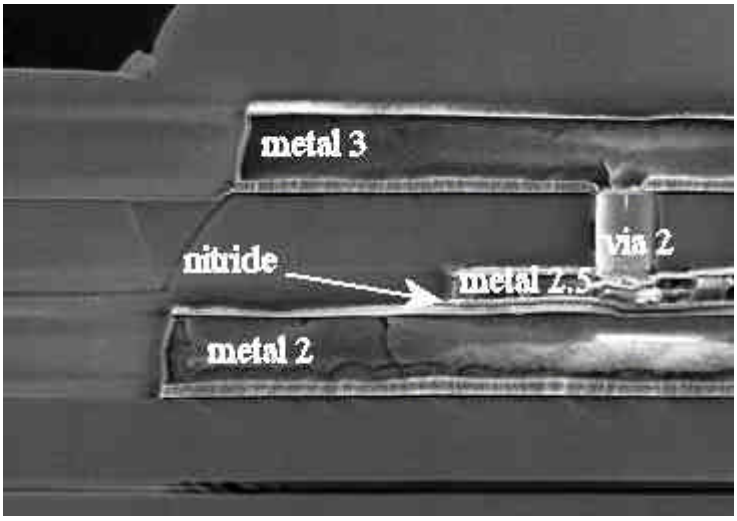


Figure 6.1b: Cross-section SEM picture across a typical MIM capacitor. Note that the shallow via 2 provides connectivity between top plate of capacitor to metal 3. Connection of bottom plate (metal 2) to metal 3 is provided by a deep via2 is not depicted.

Even more than the MOS gates, damaged metal-insulator-metal capacitors show a very high leakage current. Also here the fast leakage current can be applied. The failure criterion is leakage current.

The failure mechanism can be dominated by current or by voltage. In the case that current drives the failure mechanism, bigger capacitors show less damage since current density is lower. In Figure 6.2 one can see the impact of the antenna area connected to the top plate and of the area of the metal-insulator-metal capacitor on the failure fraction. The bottom plate is connected to a 100 µm² antenna on the same level as the top plate antenna. For the same antenna area, the bigger capacitors show a significantly higher failure fraction. This is the opposite from what one would expect when current is the driving force behind the failure mechanism. This means that the failure mechanism is not dominated by current but in fact by voltage: this breakdown is determined by the probability of weak spots and this probability is area dependent.

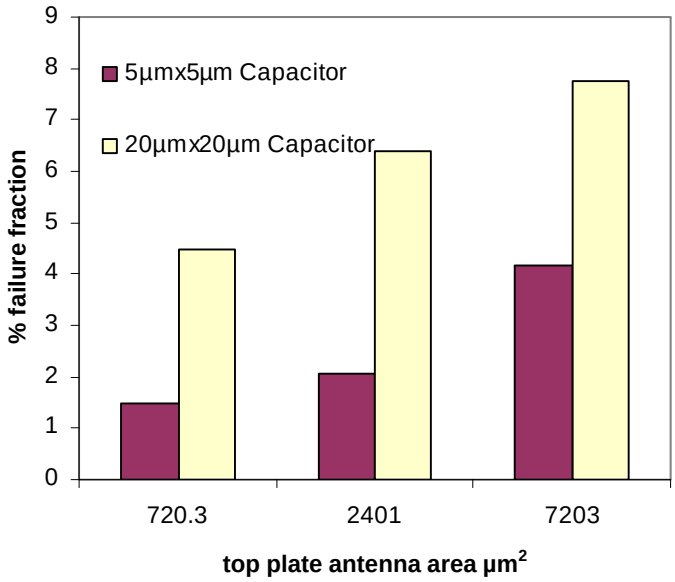


Figure 6.2: The effect of the capacitor area size on the failure fraction: for one antenna size, there is more damage on a large metal-insulator-metal capacitor compared to a small metal-insulator-metal capacitor.

Figure 6.3 shows the effect of the area ratio of the top to the bottom antenna on the failure fraction, with both antennas on the same metal level and the capacitor isolated from the substrate. When both of the antennas on the top and the bottom plate have the same size, the failure fraction falls back to almost zero. The failure fraction rises with an increasing difference in antenna size. This confirms that the charging generates a voltage potential on the capacitor plate according to the size of the antenna exposed to the charging. When both of the antennas have the same size, there is no potential difference and as a consequence, there is no charging induced damage. When the ratio between the area of the antenna connected to one plate to the area of the antenna connected to the other plate is large, then there is a large voltage potential difference causing a high failure fraction.

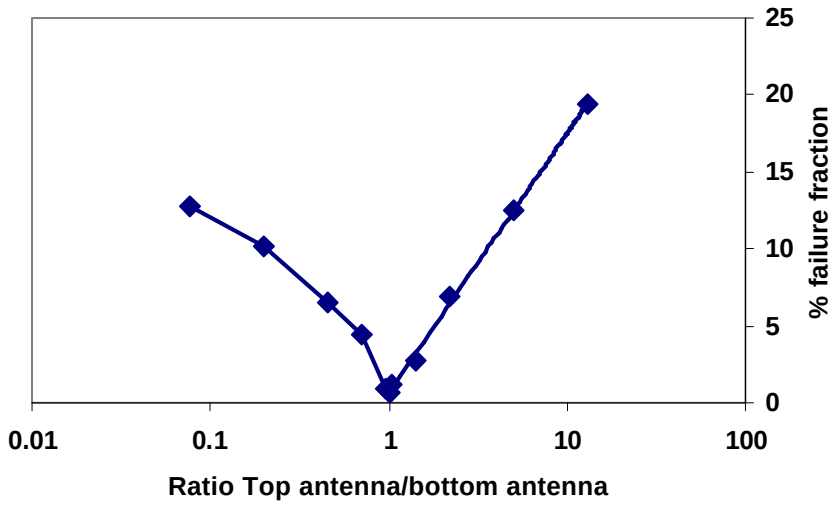


Figure 6.3: the effect of the area of the antenna size of both top and bottom antenna on the failure fraction.

Plasma induced damage on MOS gate oxides is many times more severe with a comb shaped antenna compared to a plate shaped antenna with the same area. This is caused by electron shading [3]. Figure 6.4 shows that for charging induced damage on metal-insulator-metal capacitor this is not the case: a comb shaped antenna causes just a little more charging induced damage than a plate antenna with the same area. Unlike plasma induced damage on MOS gate oxides, there is very little electron shading effect. Typically, electron shading is very severe on metal etching and inter metal dielectric oxide deposition [4]. Because we hardly see any difference between the impact of comb and plate, we can exclude electron shading as being the cause of the damage.

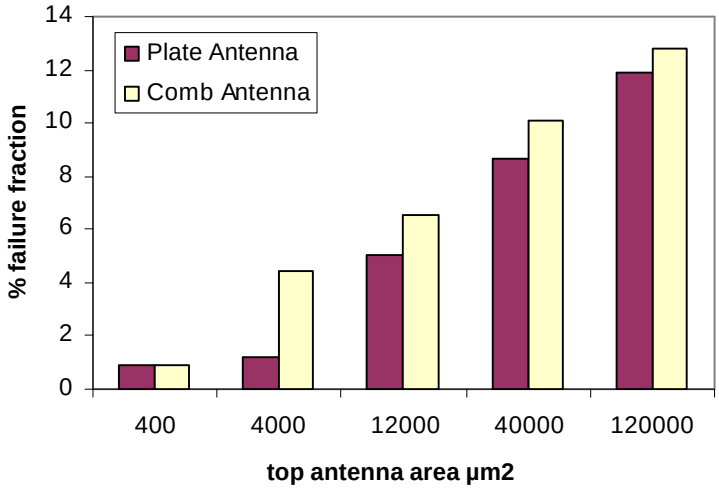


Figure 6.4: The effect of the antenna shape on the metal-insulator-metal capacitor failure fraction for a range of antenna areas.

Figure 6.5 depicts the leakage current failure fraction of the antenna test structures as function of the antenna ratio. It is evident that the increase of the failure fraction with the antenna ratio is nonlinear.

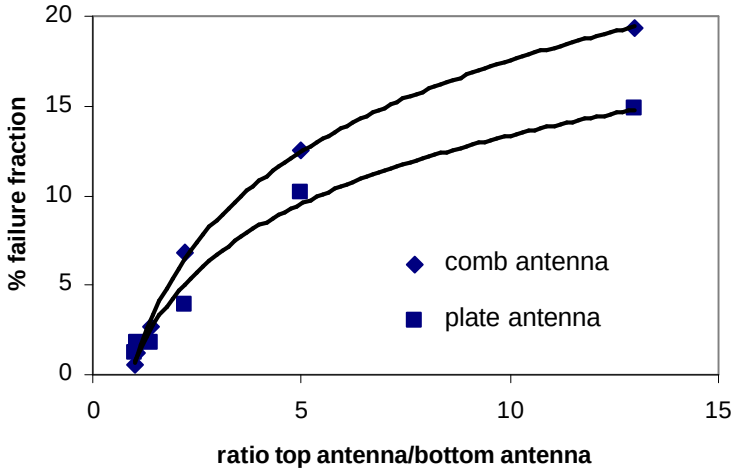


Figure 6.5: Antenna ratio dependence of failure fraction for comb and plate antennas.

From Figure 6.2 we concluded before that the breakdown is determined by the probability of defects in the dielectric. For these cases yield follows the Poisson distribution:

$$Y = e^{-A*D} \qquad (6.1)$$

where Y is the yield, A is the capacitor area and D is the defect density of the dielectric [5].

The relation between the number of defective gates and the antenna ratio is logarithmic [6]. We define:

$$A*D = n * \ln(AR) \qquad (6.2)$$

where in the case of metal-insulator-metal capacitor the antenna ratio is the area ratio of the larger antenna vs. the smaller antenna and where n is a constant factor. The yield Y can be expressed as (1 – F) where F is the failure fraction. This leads to the following equations:

$$Y = (1\text{-}F) = e^{-AD} = e^{-n*\ln(AR)} = AR^{-n} \qquad (6.3)$$

Figure 6.6 shows that this model fits the measured data very well. In this experiment, n equals 0.0812 for comb shaped antennas and 0.0607 for plate shaped antennas. The power n is defined by the antenna shape, the quality of the dielectric, the area of the capacitor and the processing itself.

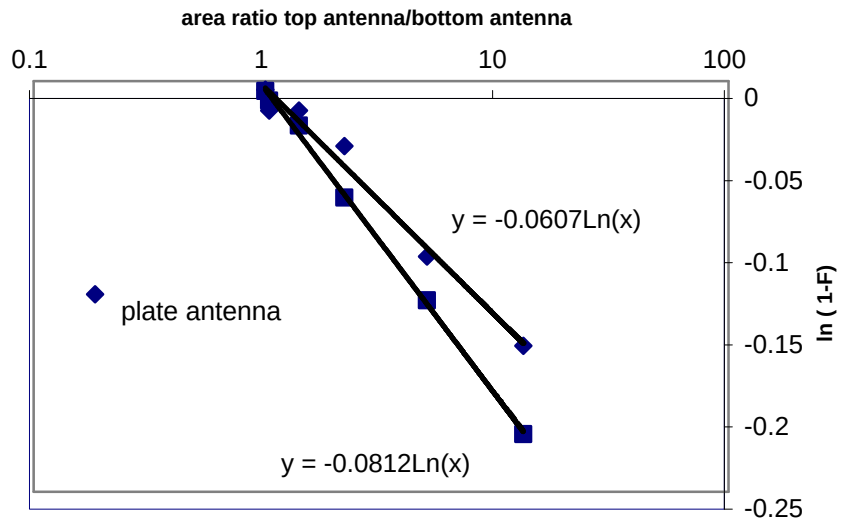


Figure 6.6: The simulated and measured failure fraction as a function of the antenna area ratio.

The model illustrates that any deviation of the antenna area ratio from unity results in yield loss.

For a test structure where one of the antennas is limited to a very small area required to make the connection to a higher metal level, the failure fraction is found to be almost zero even when the other antenna is very large. Figure 6.7 is comparing the effect of a very small antenna (1.36 µm²) on one plate vs. a large antenna (10000 µm²) on that plate for a range of antenna sizes on the other plate of the capacitor.

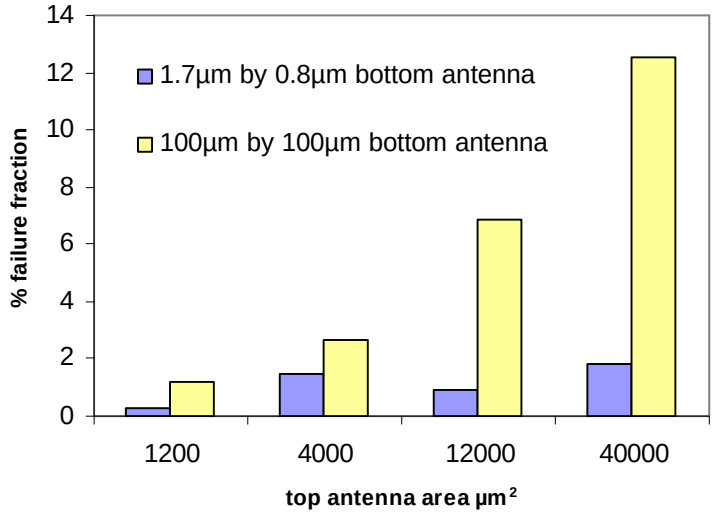


Figure 6.7: The effect on charging induced damage of a very small antenna connected one plate of the MIM capacitor: when the area of one of the antennas is very small, charging induced damage is almost completely prevented.

For the small bottom antenna, the failure fraction remains low and is independent of the top antenna size on the other plate. This indicates that there is a threshold value for the antenna area in order to damage the metal-insulator-metal capacitors. Even though the ratio between both antenna areas is very high: when one of the antennas is very small, the metal-insulator-metal capacitors are not damaged with one of the antennas below the threshold area.

Large antennas have a higher current-carrying capability. With a small antenna, only a small charge can be exchanged with the environment. As long as the leakage of the metal-insulator-metal capacitor dielectric can support this charge transfer, the capacitor is not damaged.

This phenomena is especially interesting because it opens the possibility for the introduction of metal bridges as a way of protecting the metal-insulator-metal capacitor: When the ratio of the two conductors connected to the plates of the metal-insulator-metal capacitor becomes too large, one or both of the antennas can be disconnected from the capacitor at the level of the antenna and reconnected again at a higher metal level by means of very small metal area. As such any charging induced damage during the processing is avoided. Figure 6.8 is showing a possible layout for the protection of metal-insulator-metal capacitor by the use of a metal bridge.

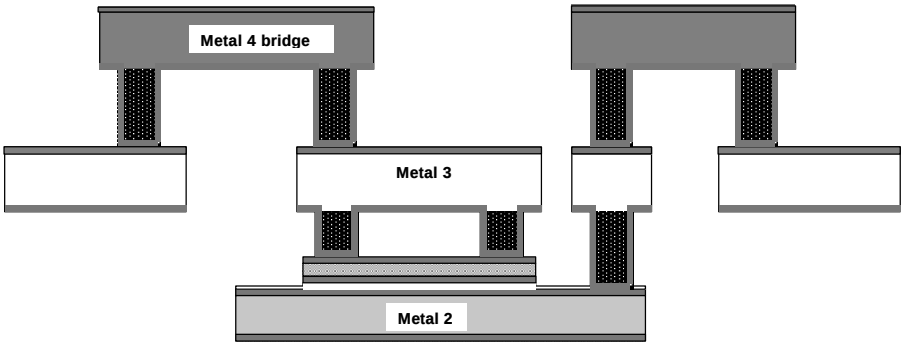


Figure 6.8: possible layout for the protection of metal-insulator-metal capacitors by the use of a metal bridge.

MOS transistors are often protected against plasma induced damage by the use of diodes [7]. These diodes are placed in parallel with the MOS gate-substrate capacitor of the MOS transistor. In this way, a leakage path is provided for any charge collected at the MOS gate – substrate capacitor during the processing. In the case of the metal-insulator-metal capacitor, a substrate diode cannot be placed in parallel over the capacitor. Still both plates can be connected to the substrate. In this way, both plates can be discharged over the substrate. As such, charging induced damage on the metal-insulator-metal capacitor is avoided. Figure 6.9 compares the damage to metal-insulator-metal capacitors as a function of the antenna ratio, both for unprotected and diode protected devices as a function of the top antenna area. For these test structures, the bottom antenna area is 10000 µm² None of the capacitors protected by diodes are damaged by charging induced damage, no matter how large the antennas are. The protective diodes prove to be an excellent way to prevent charging induced damage on metal-insulator-metal capacitors. Figure 6.10 is showing a possible layout of the protection of metal-insulator-metal capacitors by the use of diodes. Care must be taken that both capacitor plates are connected to the diodes by the same metal level. This prevents that during the processing, one plate is connected to the substrate and the other one is floating. In that case the risk on charging induced damage is increased because an extra current path is provided.

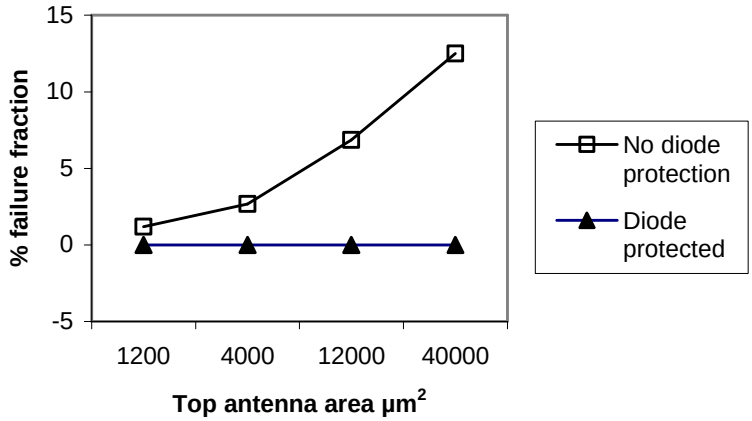


Figure 6.9: The effect on charging induced damage of diodes connected to both plates of the metal-insulator-metal capacitor: when both plates on the metal-insulator-metal capacitor are connected to a protective diode, charging induced damage is completely prevented.

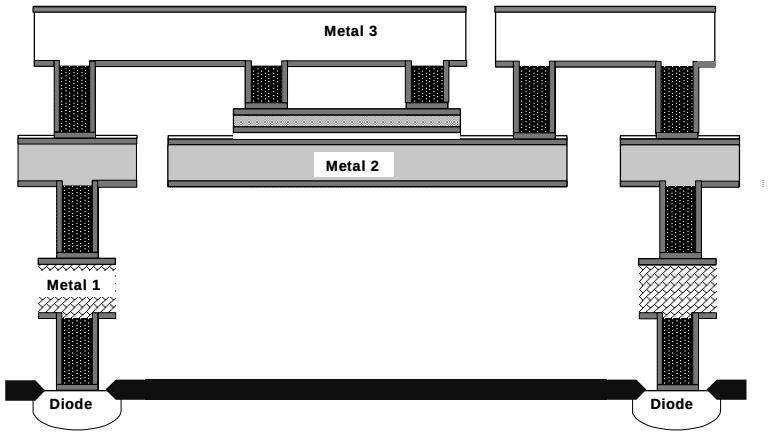


Figure 6.10: possible layout for the protection of metal-insulator-metal capacitors by the use of diodes.

The results of the experiments can be summarized in the following simple design rules:

- It is recommended to keep the ratio between the areas of the antennas connected to both plates of capacitor for each individual metal level close to unity except for the case that the area of one of the antennas is only a few μm^2.

- Contacting only one of the capacitor plates to the substrate by a contact, a diode or even by a MOS gate has to be avoided.

Use of protective diodes or metal bridges as proposed in Figures 5.16 and 5.18 can solve problems that might rise when these rules cannot be followed.

Charging induced damage to metal-insulator-metal capacitor s is related to the size and shape of the antennas connected to both capacitor plates. The failure mechanism is not dominated by current but rather by voltage. This makes large capacitors more vulnerable than small capacitors for the same antenna area. The damage is caused by the build up of a voltage potential difference on the two plates of the capacitor. It has been proven that there is a direct relation between the failure fraction and the area ratio of antennas connected to the plates of the metal-insulator-metal capacitor. A simple log function describes very well the relation between the failure fraction of the metal-insulator-metal capacitor and the ratio of the two antenna areas exposed to charging. For a very small antenna, the failure fraction remains low and is independent of the antenna size connected to the other plate. The use of metal bridges

or protective diodes are effective ways to protect metal-insulator-metal capacitors against charging induced damage. Antenna design rules are proposed to apply on layout checking with the consideration of substrate contacts and the interactions of the interconnection lines which become antennas connected to both plates of the metal-insulator-metal capacitor. Applying the design rules during the phase of IC design prevents the potential charge induced damage during manufacturing. Furthermore, layout solutions are introduced.

6.3 Charging Induced damage by photoconduction through thick inter metal dielectrics

This section brings evidence, that even after the deposition of a thick insulating inter metal dielectric over semiconductor devices, there is still an eminent risk for charging induced damage of gate oxides. The charging induced damage is induced through photo-conductance of the inter metal dielectric. This charging induced damage requires the simultaneous presence of elevated temperatures, UV light, a large perimeter conductor and a highly non-uniform plasma as used for particle reduction. The combination of these factors is used quite commonly in commercial tools for semiconductor processing. This charging induced damage of gate oxides leads to loss of reliability or in the worst cases to direct yield loss. The location and magnitude of the charge was determined with non contact surface potential measurement and confirmed by gate leakage measurements on antenna structures. In this way further yield loss could be prevented. A model for the relation between the surface charging potential and the reduced gate oxide integrity is presented.

We present and analyze a new very strong charging phenomenon due to plasma power ramp down in the PECVD chamber occurring when a thick oxide layer already covers and insulates the metal lines. The evaluation is provided with non-contact surface potential measurement on an in-line plasma damage monitor, and gate oxide integrity measurements on antenna test structures.

The used PECVD tool is a medium density plasma reactor which is used to deposit a SiO_2 layer on 0.5 µm metal lines. The medium density plasma is generated by a capacitor coupled plasma source from a SiH_4 and N_2O chemistry. The source frequency is 13.56 MHz. We have studied two PECVD deposition processes: one is a deposition of a 0.5 µm thick oxide layer, the other studied processes consist of the same deposition of a 0.5 µm thick oxide layer but now an extra plasma power ramp down step was added. This extra plasma power ramp down step is added in order to improve the particle performance of the process.

By comparing the charging and the degradation between both PECVD processes, the impact of the plasma power ramp down step is measured. The impact of further processing is measured by comparing the degradation of the structures with and without the diode protection.

Figures 6.11 and 6.12 compare the charging of a deposited oxide layer with and without the plasma power ramp down step.

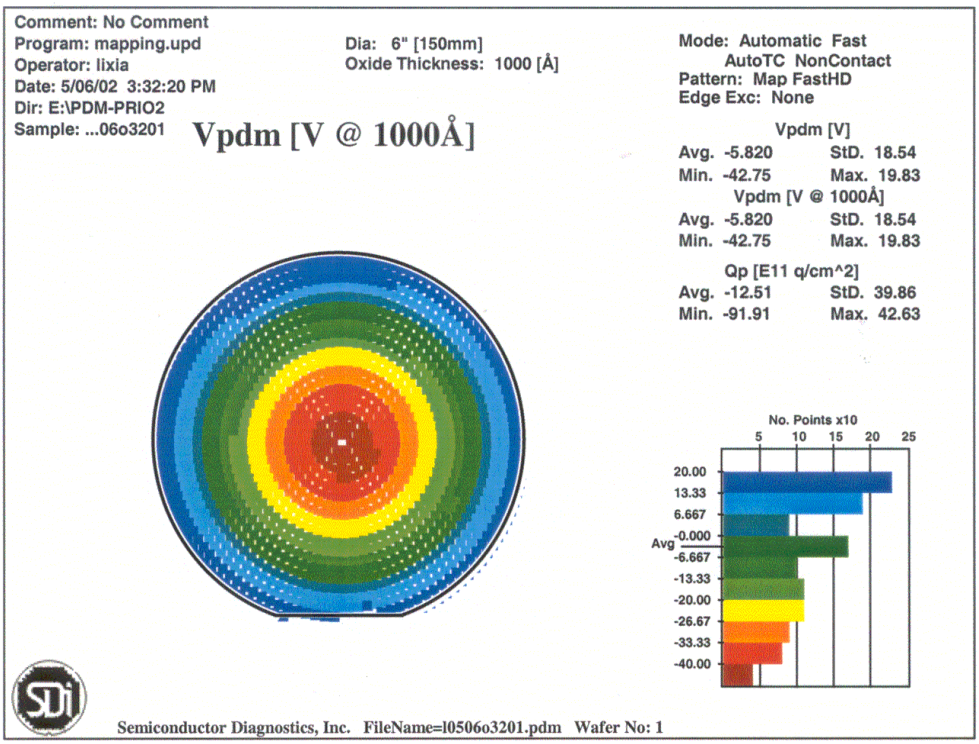


Figure 6.11: plasma damage monitoring mapping of a 500 nm deposited oxide layer followed by the power ramp down step showing 62.13 V difference between the centre and the edge of the wafer.

The extreme levels of charging are measured in the centre and on the edges of the wafer. In the centre of the wafer a minimum voltage of –42.75 V was measured, while on the edges a maximum voltage of 19.83 V was measured.

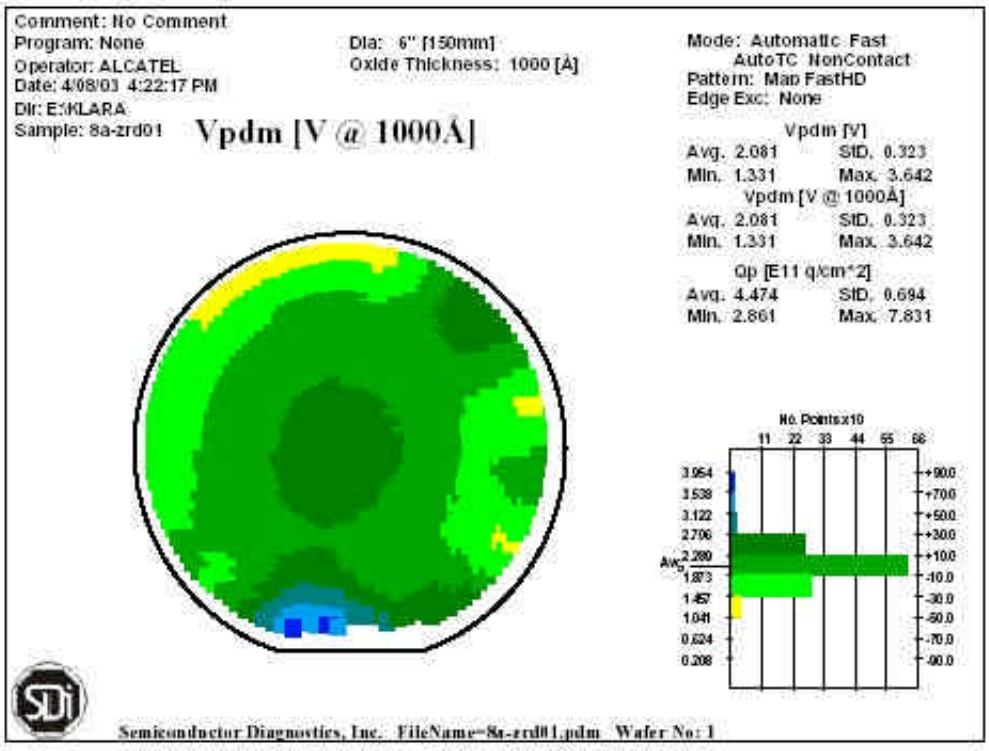


Figure 6.12: plasma damage monitoring mapping of a 500 nm deposited oxide layer not followed by the power ramp down step shows 2.3 V difference over the wafer.

Figure 6.13 is showing the failure fraction of the transistor gate connected to an antenna with antenna ratio of 100 000 as a function of the shape of the antenna, the diode protection and the plasma power ramp down.

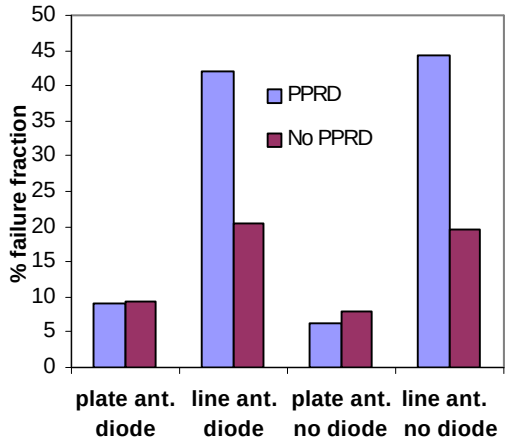


Figure 6.13: Failure fraction of the transistor gate connected to an antenna ratio 100 000 antenna as a function of the shape of the antenna, the diode protection and the plasma power ramp down step.

The transistor gates connected to the line shape antennas are showing a sharp increase of the failure fraction when processed with the plasma power ramp down step. This effect does not occur on the plate-shaped antennas. The increase of plasma damage correlates with the perimeter of the antenna.

There is no significant difference between the structures with and without a diode protection at the next metal level: the plasma damage occurs during the oxide deposition covering the antenna structures.

To correlate the location of the highly charged regions with the location of the failing sites, two wafer mappings of the failing antenna structures have been made. Figure 6.14 shows the mapping of a wafer processed without plasma power ramp down. Figure 6.15 shows the mapping of a wafer processed with plasma power ramp down. The numbers in the squares represent the log of the leakage current.

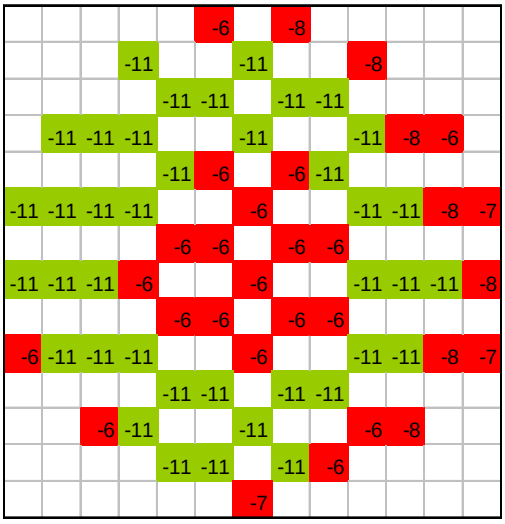


Figure 6.14: Typical yield mapping of a wafer processed with the plasma power ramp down process.

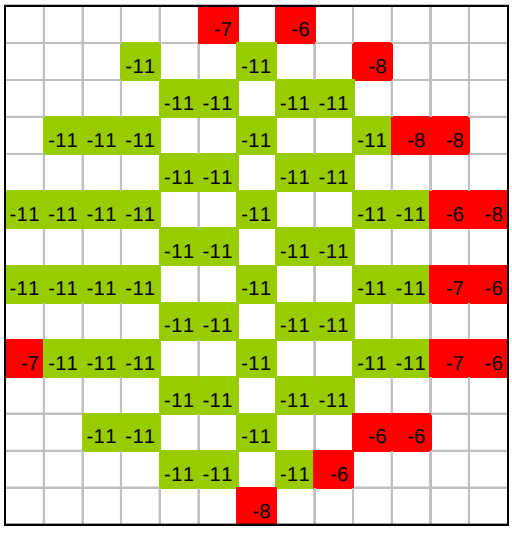


Figure 6.15: Typical yield mapping of a wafer processed without the plasma power ramp down process.

The yield in this experiment has mainly been affected in the centre and on the edges of the wafer. The yield loss at the edges of the wafers is the same for both experiments. This edge effect was proven to be caused by the metal patterning prior to the inter metal dielectric deposition. The affected centre region however coincides very well with the charged area as measured with the plasma damage monitoring. This mapping confirms the causal relation between the charges measured by the plasma damage monitoring and the gate defects as measured on the antenna structures.

The failure mechanism could have two causes: the first is a charge induced into the antenna, mirroring the charge on top of the oxide surface, the second is a leakage current through the inter metal dielectric.

A charge on an insulator induces a mirror charge in any conductor on the other side of the insulator. When this charge exceeds the Q_{BD} of the gate oxide of the transistor gate connected to the conductor, this gate oxide will be damaged and starts to leak. The Q_{BD} of the gate oxide, especially for small areas, is typically larger than 10 C/cm^2 [8]. This was confirmed by measurements. With a given gate area of 0.4 μm^2, the Q_{BD} per gate is $4*10^{-8}$ C. The charge in the metal lines connected to the gate, required to mirror the charge Q on the oxide surface on top of these metal lines, can be calculated.

With $C_{fF} = \frac{e_{ox}.e_o}{T_{ox}}.A$ and $Q = C*V$ we find $Q = 350\,pC$

The charge induced by mirroring the surface charge is less than 1% of the available Q_{BD}. Although Q_{BD} is reported to reduce to some extent at elevated temperature, it is not reduced by 2 orders of magnitude [9]. This is also confirmed by the fact that the phenomenon is mainly perimeter-dependent: it does not occur on metal plates with the same area. Metal plates and lines with the same area require the same charge to mirror the oxide surface charge. This hypothesis has to be abandoned.

Charging induced damage by conduction through inter metal dielectric has been reported in literature before and contributed to the elevated temperature [9] and the presence of highly energetic UV photons resulting in photoconduction of the dielectric [10]. In both cases the suggested models cannot explain the proposed topographical effects and the effects could only be observed for very thin inter metal dielectric as it is present only at the beginning of the deposition.

J.P. Carrère et al. explained the perimeter effect by introducing the nonconformality of the deposition in the model. This nonconformality causes the inter metal dielectric oxide to remain thin at the bottom of the gap in between the metal lines. These sites remain more sensitive to photoconduction. But again, this could only be demonstrated with thin (100nm) depositions [11].

Now, for the first time it is demonstrated that the risk of damaging gate oxides by charging induced damage remains present even after a thick oxide layer (600nm) has been deposited. The damage can easily occur in commercially available tools, using the vendors “best known method” for particle reduction.

The occurrence of the phenomenon can be explained by the simultaneous presence of four factors:

- Elevated temperatures: Plasma enhanced dielectric deposition, irrespective of the tool and plasma source being employed, is invariably done in the 350ºC to 400ºC range.
- The presence of a large perimeter conductor connected to the gate of a transistor.
- The presence of UV light, inevitable in a PECVD oxide deposition process.
- The presence of a highly non-uniform plasma: In order to improve the particle performance of the process, the equipment vendor suggests to reduce plasma power at the end of the deposition process. This reduced plasma power however, is no longer capable of supporting a plasma over the full wafer surface: only in the centre of the wafer there is a small zone of plasma remaining. This zone is charging highly negative as confirmed by the plasma damage monitoring measurements.

The quite common combination of these factors, is leading to charging induced damage of gate oxides. Damaged gate oxides are causing a reduced reliability and, in the worst case, are leading to direct yield losses.

In summary, evidence is brought forward that even after the deposition of a thick insulating inter metal dielectric over semiconductor devices, there is still an eminent risk for charging induced damage of gate oxides. The charging induced damage is induced through photo conductance of the intermetal dielectric. This charging induced damage requires the quite common simultaneous presence of elevated temperatures, UV light, a large perimeter conductor and a highly non-uniform plasma as used for particle reduction. The location of the charges is measured with plasma damage monitoring and confirmed by gate leakage measurements on antenna structures.

6.4 Conclusion

Charging induced damage to metal-insulator-metal capacitor s is related to the size and shape of the antennas connected to both capacitor plates. The failure mechanism is not dominated by current but voltage is the main driver. This makes large capacitors more vulnerable than small capacitors for the same antenna area. The damage is caused by the build up of a voltage potential difference on the two plates of the capacitor. It has been proven that there is a direct relation between the failure fraction and the area ratio of antennas connected the plates of the metal-insulator-metal capacitor. A simple log function describes very well the relation between the failure fraction of the metal-insulator-metal capacitor and the ratio of the two antenna areas exposed to charging. For a very small antenna, the failure fraction remains low and is independent of the antenna size on the other plate. The use of metal bridges or protective diodes is effective to protect metal-insulator-metal capacitors against charging induced damage. Antenna design rules are proposed to apply on layout checking with the consideration of substrate contacts and the interactions of the interconnection lines which become antennas connected to both plates of the metal-insulator-metal capacitor. Applying the design rules during the phase of IC layout design prevents the potential charge induced damage during the phase of IC manufacturing. Furthermore, layout solutions are introduced.

Evidence is found that even after the deposition of a thick insulating inter metal dielectric over semiconductor devices, there is still an eminent risk for charging induced damage of gate oxides. The charging induced damage is induced through photo conductance of the inter metal dielectric. This charging induced damage requires the quiet common simultaneous presence of elevated temperatures, UV light, a large perimeter conductor and highly non-uniform plasma as used for particle reduction. The location of the charges is measured with plasma damage monitoring and confirmed by gate leakage measurements on antenna structures.

6.5 References

[1] S. Fang and J. P. McVittie, "Oxide damage from charging: breakdown mechanism and oxide quality", IEEE Trans. Electron Dev. 41(6), 1034, 1994.

[2] E. B. Harris, "Charging damage in metal-oxide-metal capacitors", Proceedings of 3th International Symposium on Plasma Process-Induced Damage, June 4-5, 1998.

[3] K. Hashimoto, "New phenomena of charge damage in charging etching: Heavy damage only through dense-line antenna", Japanese journal of Applied Physics, Vol 33, Part 1, No. 10, pp. 6013-6018, 1994.

[4] J. Ackaert et al., "Antenna test structures matrix description, application for optimization HDP oxide deposition, metal etch, Ar preclean and passivation processing in sub-half micron CMOS processing", Proceedings of the 5th Intern. Symp. P2ID, May 22-24, 2000.

[5] J. Cunningham, "The use and evaluation of yield models in integrated circuit manufacturing", IEEE Transactions on Semiconductor Manufacturing, Vol. 3 no.2, pp. 60-71, May, 1990.

[6] Z. Wang, "Relation between Plasma Process-Induced Oxide", Proceedings of the 6th Intern. Symp. P2ID, May 14-15, 2001.

[7] H. Shin et al., "Impact of plasma charging damage and diode protection on scaled thin oxide", IEEE IEDM, pp. 467-470, 1993.

[8] Degraeve R et al., "New Insights in the Relation Between Electron Trap Generation and the Statistical Properties of Oxide Breakdown", IEEE Transactions on Electron Devices, 45(4), pp. 904-11, 1998.

[9] Kin P. Cheung, "On the mechanism of plasma enhanced dielectric deposition charging damage", Proc. 5th P2ID, pp. 161-163, 2000.

[10] Mayur Joshi et al., "Direct Experimental Determination and Modeling of VUV Induced Bulk Conduction in Dielectrics during Plasma Processing", Proc. 5th P2ID, pp. 157-160, 2000.

[11] J.-P. Carrère et al., "Topographical Dependence of Charging and New Phenomenon During Inductively Coupled Plasma CVD Process", Proc. 5th P2ID, pp. 164-167, 2000.

Chapter 7

Identification, evaluation and optimization of critical process steps for process induced damage

The results in this chapter have partially been published in:

J. Ackaert, E. De Backer, P. Coppens, M. Creusen, “Prevention of plasma induced damage on thin gate oxide of high density plasma oxide deposition, metal etch, Ar preclean processing in BEOL sub-half micron CMOS processing”, Proceedings of the 5th International Symposium on Plasma Process-Induced Damage, pp.77-80, 2000.

J. Ackaert, Z. Wang, E. De Backer, P. Coppens, “Charging Damage in Floating Metal-Insulator-Metal Capacitors”, 6th International Symposium on Plasma Process-Induced Damage, pp. 120-123, 2001.

J. Ackaert, Z. Wang,E. De Backer, P. Colson, P. Coppens, “Non Contact Surface Potential Measurements for Charging Reduction During Manufacturing of Metal-Insulator-Metal Capacitors”, 12th European Symposium on Reliability of Electron Devices, Failure Physics and Analysis. (ESREF), pp. 1403-1408, 2001.

7.1 Introduction

The previous chapter described plasma charging damage on specific devices like the flash memory cells and metal-insulator-metal capacitors. This chapter describes how after a synergetic step-by-step investigation of all major plasma process steps in a conventional CMOS process, a number of process steps were pinpointed that were generating plasma damage in the MOS devices. It is described how these process steps have been characterized and optimized.

The areas with the highest risk of inducing plasma damage are patterning of conductors as poly and metal, via etching and preclean, high density oxide deposition and passivation processing.

During dry etching of a conductor, the plasma damage is induced by edge-collected plasma charge and occurs during the last part of the etch step. The collected charge is proportional to the perimeter of the conductive layer. In addition to contact and via etch, sputter cleans prior to plug filling in the contact and via modules can also create plasma damage. The collected charge is proportional to the number of contact- and via holes. Dry resist strip, ion implantation and plasma-assisted deposition create damage proportional to the total area of the underlying conductive material. A first assessment has shown that steps critical to plasma process induced damage are mainly situated in the backend. This is also the place in which, in a typical design, also the largest charge collecting antennae can be found.

Metal and poly patterning etch are a major concern. Advanced etch tools provide very uniform high density plasma in which plasma damage is mainly caused by electron shading. Many groups already investigated the relation between charge collecting antenna area that can be allowed for a given pattern density and aspect ratio of the etched features. This phenomenon is reported in numerous papers [1-8].

In addition to contact and via etch process, sputter cleans prior to plug filling in the contact and via modules can also create plasma damage. The sputter clean process is optimized in terms of plasma density and energy and process time for minimal plasma damage.

Low pressure, high density plasmas are more and more implemented in deep submicron technology patterning processing. A second application of high density plasma is found in the pre-metal dielectric and inter metal dielectric modules where a good gap fill can be achieved by combining high density plasma deposition with simultaneous bias sputtering. To avoid direct contact of the high density plasma with the underlying conductor, a liner can be deposited with a less damaging deposition technique. It is investigated to what extent the damage is reduced by the liner and what are the optimal process parameters of the high density plasma inter metal dielectric deposition process.

In some cases, the use of protective devices disturbs the functionality of the device. Under these conditions, the process inducing the plasma damage has to be optimized or even replaced by a plasma damage free process. This chapter describes a case where the reactive ion etch process for opening the pad-area of the chip was the cause of severe plasma damage. In this particular case a diode was not allowed to protect the pad from plasma damage for reasons of functionality of the design. It was investigated if the process for opening the pad could be replaced by remote plasma, a plasma damage free process.

During the investigation a severe case of damage to metal capacitors was discovered. Metal-insulator-metal capacitor processing includes dielectric deposition, plasma cleaning, oxide etching, metal depositions and patterning and resist stripping. Each of these process steps may have an impact on the amount of plasma process induced damage on the metal-insulator-metal capacitor dielectric. Each of these processing steps had to be investigated, characterized and optimized with the aim of minimizing the plasma process induced damage. As described in the previous chapter interconnect design limitations and protection methods have to be defined. It was discovered that not only the processing of the metal-insulator-metal capacitor implies risks of charging damage. Also the post-processing of these devices can damage the device. Not only plasma induced damage has to be considered but rather the more general charge induced damage. Post processing had to be optimized to eliminate charging induced damage risks for the metal-insulator-metal capacitor during the post processing.

In any case, the solutions reducing the plasma damage still have to perform as well as the original process in terms of yield, defectivity, cost, electrical parameters or process capability.

7.2 Optimization of high density plasma oxide deposition, Ar preclean and passivation processing in sub-half micron CMOS processing

Process optimization results are presented on high density plasma inter metal dielectric deposition, Ar-preclean, metal etch and passivation etch. For high density plasma oxide deposition, the relation between plasma damage and the deposition to sputter ratio is investigated. Ar-preclean induces plasma damage depending on the applied ion energy and plasma density. The optimal combination of these parameters is investigated. In cases where reactive ion etching for pad opening induces severe plasma damage, a plasma damage free alternative is investigated.

7.2.1 Introduction

Degradation of gate oxide MOS devices due to plasma processing has become technologically important and has driven intense research since the device dimensions approach the sub-half micron regime [9][10] and gate oxide thickness continues to decrease [11]. This section describes the application of the set of test structures described in chapter3 in order to identify and optimize major process steps responsible for plasma induced damage in a sub-half micron CMOS process with 5 layers of metal.

7.2.2 High density plasma inter metal dielectric oxide deposition

High density plasma oxide deposition is combining PECVD oxide deposition and the removal of oxides by means of Ar sputter etching. The ratio between the deposition and sputter rate can be chosen as a process parameter. The choice of deposition to sputter ratio (D/S) is determined by the aspect ratio of the space between two metal lines that needs to be filled with the oxide. A low aspect ratio allows a fast deposition process with high deposition to sputter ratio while a high aspect ratio requires a slow process with a low deposition to sputter ratio. During the process, various deposition to sputter ratios can be used. During the deposition, the aspect ratio is reducing: a higher deposition to sputter ratio is allowed towards the end of the process. To protect the metal from excessive bombardment by the Ar ions, the process can be stated with a "liner" oxide deposition without the Ar-sputter etch component.

Four processes have been compared.

- "Liner + D/S 4 + D/S 5.6": a 100 nm liner oxide is followed by 400 nm deposition with a deposition to sputter ratio of 4. On top there is 400 nm deposition with deposition to sputter ratio of 5.6.
- "No liner D/S 4 + D/S 5.6": a 400 nm deposition with a deposition to sputter ratio of 4. On top there is 500 nm deposition with deposition to sputter ratio of 5.6.
- "Liner + D/S 5.8 + D/S 3.2": a 100 nm liner oxide is followed by 400 nm deposition with a deposition to sputter ratio of 5.8. On top there is 400 nm deposition with deposition to sputter ratio of 3.2.
- "Liner + D/S 3.2": a 100 nm liner oxide is followed by 800 nm deposition with a deposition to sputter ratio of 3.2.

The impact of these processes has been investigated in terms of plasma damage. Figure 7.1 shows the percentage of failing gates for the various deposition processes processed on metal 1 and on metal 2 antennas.

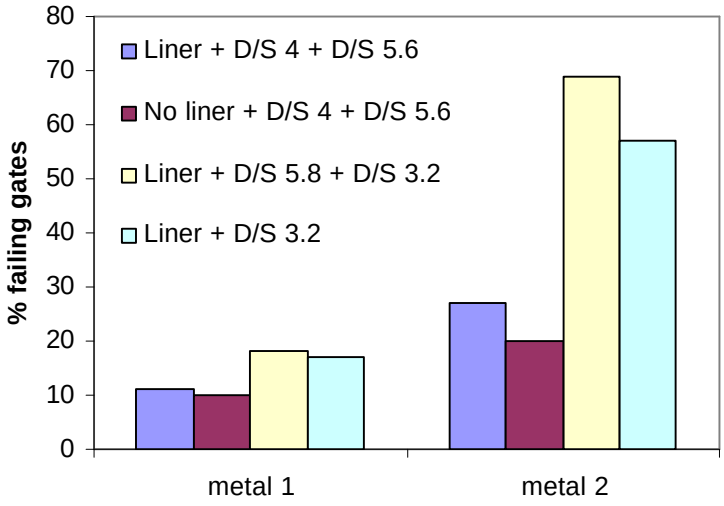


Figure 7.1: Plasma damage induced by high density plasma oxide depositions for metal 1 and 2. The processes with only deposition to sputter ratios of 4 and higher induce minimal plasma induced damage compared to the process combinations including deposition to sputter ratio of 3.2.

This deposition to sputter ratio was found to be a major parameter determining the plasma damage induced by high density plasma oxide deposition. The processes with only high deposition to sputter ratios are inducing minimal plasma induced damage compared to the process combinations including lower deposition to sputter ratio. Even when low deposition to sputter ratio is used after high deposition to sputter ratio, plasma induced damage is high. Introducing a PECVD SiO_2 liner has no significant impact on the plasma induced damage. This means that simulations reported in [12] cannot be confirmed.

The reason for this is discovered by examining the SEM cross-sections of the partially deposited layers. High density plasma deposition with low deposition to sputter ratio keep the top corners of the metal structures open to the impact of the plasma during a large part of the processing (Figure 7.2 left). A process with a higher deposition to sputter ratio keeps the metal lines covered and isolated from direct contact with the plasma. (Figure 7.2 right). In this way fewer charges are picked up from the plasma and fewer gates are being damaged.

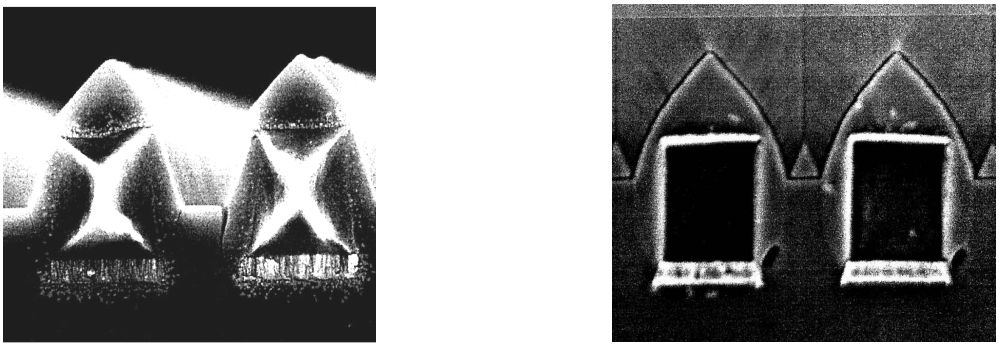


Figure 7.2: SEM Cross section of partial high density plasma deposition at low deposition to sputter ratio (left) and high deposition to sputter ratio (right). When the metal lines remain exposed to plasma during the process, plasma damage is increasing.

In conclusion, an increased deposition to sputter ratio results in reduced plasma damage. When the metal lines remain exposed to plasma during the process which is the case when a deposition to sputter ratio of 3.2 is used, plasma damage increases. Introducing a PECVD SiO_2 liner has no significant impact on the plasma induced damage. At the upper corners of the metal lines, deposited oxides are removed by the Ar sputter etch component of the process. When using the deposition to sputter ratio of 3.2, the deposited liner is removed as well together with the HDP deposited oxide and the metal lines become re-exposed to the plasma anyhow. In order to keep plasma damage induced by a high density plasma oxide deposition minimal, the deposition to sputter ratio has to be chosen as high as possible. As an extra, the high deposition to sputter ratio has a higher final deposition rate. This results in a higher throughput of the equipment and a reduced operation cost.

7.2.3 Ar-preclean processing

Prior to the filling of contact- and via holes, all contaminants remaining from previous process steps as well as the native oxide on the underlying metal or Si, needs to be removed. This is done by means of an Ar plasma. In addition to contact and via etch, the Ar sputter clean can also create plasma damage. The sputter clean process needs to be optimized in terms of plasma density, plasma energy and process time for minimal plasma damage and minimal ohmic resistance. A process combining minimal plasma damage with minimal via and contact resistance has been developed. In the following paragraph, the exploration of the full sputter etch window in terms of plasma power setting in relation to resistance and plasma damage is reported. The hardware that was used for the experiment is the Preclean II installed on the Endura mainframe of Applied Materials.

Preclean II is used to remove native oxide and post etch fluorocarbon residue from contacts and vias on device wafers prior to metallization. Preclean II features a high oxide removal rate at independent wafer DC bias potentials, with minimal electrical device damage resulting from processing. The Preclean II chamber consists of an upper inductor or coil powered by a 400 kHz RF supply and a cathode powered by a 13.56 MHz RF supply (Figure 7.3).

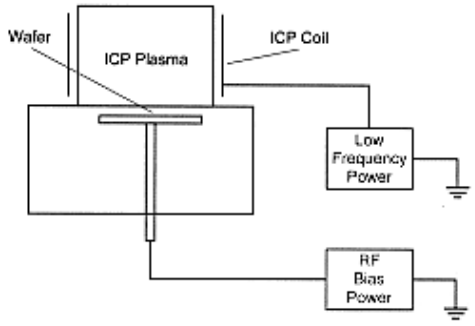


Figure 7.3: The Preclean II chamber with upper inductor or coil powered by a 400 kHz RF and a cathode powered by a 13.56 MHz RF.

The coil is a copper strip wrapped around a quartz cylinder. The electrical and magnetic fields induced inside the cylinder cause the electrical breakdown of Ar to create a plasma. Plasma density increases inside the cylinder as coil power is increased. A higher plasma density inside the coil creates more Ar ions in the plasma, resulting in a higher oxide etch rate for a given DC bias on the wafer.

To etch a wafer, a negative bias potential is applied to attract ions from the plasma. The positively charged Ar ions are attracted to the wafer surface and subsequently knock off material on impact. In the Preclean II chamber, bias is induced by applying RF energy at 13.56 MHz to the wafer pedestal. The effective DC bias developed on the wafer is proportional to the RF energy applied to the wafer pedestal and inversely proportional to the RF power applied to the upper coil. The oxide etch rate depends on the wafer DC bias and also on the power applied to the upper coil. The advantage of the Preclean II chamber thus lies in the fact that etch rate can be increased without increasing the DC bias by independently

increasing the power to the upper coil and creating a denser plasma. Both for the RF power in the upper coil as for the RF energy applied to the wafer pedestal, the maximum sustainable power is 400 W.

The optimum process recipe for a given application is highly dependent on the geometry of the contact or via to be cleaned and is also function of the previous steps used in the process flow. On the wafer, DC bias controls the angularity of the incoming ion flux. High DC bias creates a more collimated beam, giving a higher ratio of bottom etch- rate to etch rate at the top of the contact. Higher DC bias values are required at higher aspect ratios to obtain low via resistance and good uniformity of via resistance.

A matrix of plasma power and plasma density was processed. Time was adjusted to keep oxide removal constant. Plasma damage was measured as gate leakage on gates connected to 10000 via's. The results can be fitted with a linear model. Figure 7.4 shows the contour lines of the model indicating the % of failing gates. The percentage of failing gates is decreasing mainly by increasing plasma power and less by increasing plasma density.

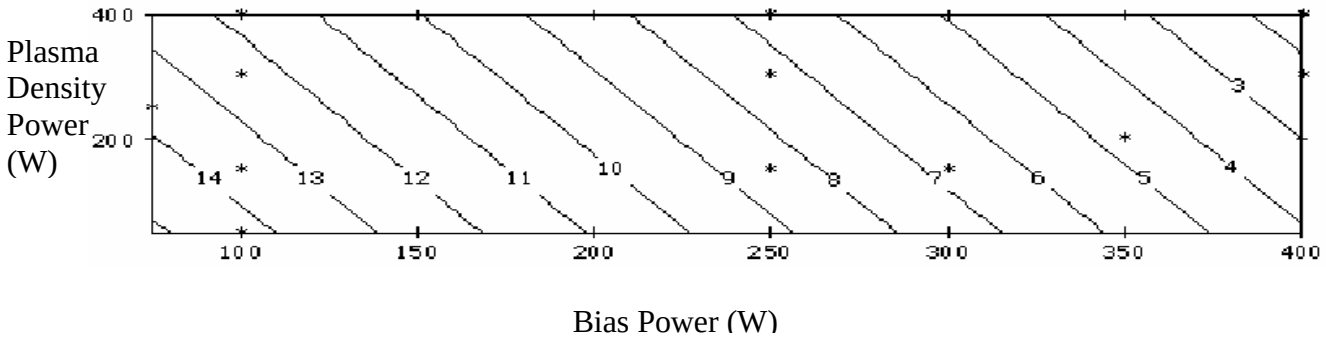


Figure 7.4: Contour lines of Plasma damage expressed as % leaking gates, as a function of bias power and plasma density power. The percentage of failing gates is decreasing mainly by increasing bias power and less by increasing plasma density power. The experimental settings are marked by a "∗".

For the same process window, the resistance of the vias was modelled as a function of the two plasma power settings (Figure 7.5). A low damage zone is found diagonal over the investigated process window. When plasma density power and bias power are similar, then the via resistance is minimal.

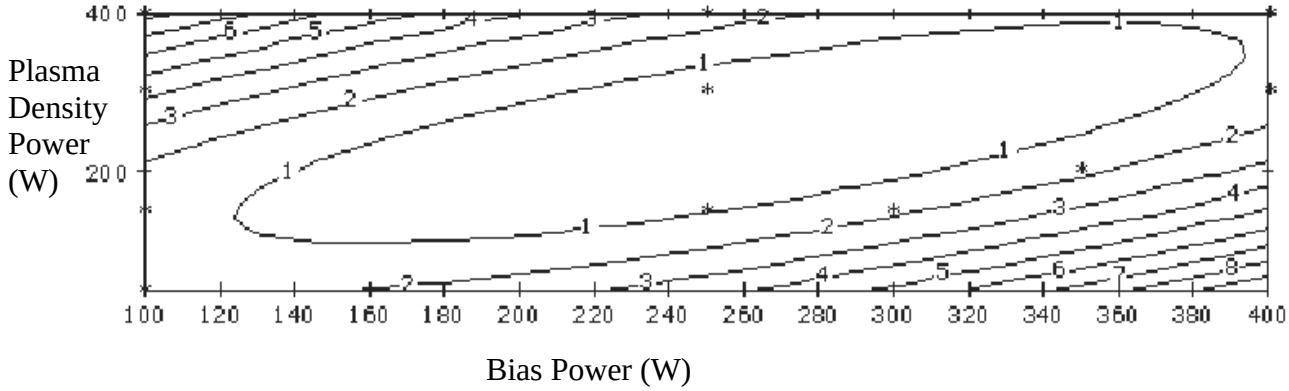


Figure 7.5: Contour lines of via resistance (Ω) as a function of plasma power and plasma density. In the region where plasma density power and bias power are similar then the via resistance is minimal. The experimental settings are marked by a "∗".

When combining the results of both plasma damage and via resistance optimization, the maximal 400 W bias power with the maximal 400 W plasma density is the optimal combination for minimal plasma damage combined with a low via resistivity. This combination also has the highest sputter etch rate and the lowest process time. The negative impact of the process time indicates that under the given process conditions, the time that the gate oxide is stressed by the charging dominates the effects of the plasma density or plasma energy.

Still to avoid having to run at the maximal available power, it was evaluated which setting can be reduced without to much impact on the plasma damage and via resistance. When plasma density power is reduced from 400 W to 300 W, there is no impact on the via resistance and the plasma damage is increasing by 1%. When bias power would be reduced from 400 W to 300 W, then the via resistance improves a little but the plasma damage is increasing by 5%. The bias power should remain at the maximum value. In addition there is neither indication nor experience that the model should be discontinuous over 400 W. Considering the maximal available power, the lesser impact of the plasma density power reduction and the large impact of the bias power; the optimal working point was defined as 300 W plasma density power and 400 W bias power.

When combining the conclusions from both plasma damage and via resistance optimization, and considering the maximal available power, the combination 400 W bias power with 300 W plasma density power was found to be the optimum combination. The process with the reduced plasma damage again has the shorter process time. Again the low plasma damage process is more cost effective. Under the given process conditions, the time that the gate oxide is stressed by the charging dominates the negative effects of the plasma density or plasma energy.

7.2.4 Pad opening etching: reactive ion etch vs. downstream

In some cases the use of protective devices disturbs the functionality of the device. Under this condition, the process inducing the plasma damage has to be optimized or even replaced by a plasma damage free process. This paragraph describes a case where the reactive ion etch process for opening the pad-area (passivation etching) of the chip was the cause of severe plasma damage. In this particular case the use of a diode was not allowed to protect the pad from plasma damage for reasons of functionality of the device. The process for opening the pad had to be replaced by remote plasma downstream etch process.

With a reactive ion etch process, the plasma is very close to the surface of the wafer. Charged ions and electrons are reaching the surface of the wafer continuously. This process can be made highly directional or anisotropic. As a consequence a current source is provided from the plasma to the wafer. This can result in plasma damage.

With a remote plasma downstream etcher, the plasma is isolated from the wafer by a grounded grid. This grid catches any charged ion and electrons. Only neutral radicals are able to pass the grid and reach the surface of the wafer. This process is not directional or isotropic. Since only neutral atoms and molecules can reach the wafer surface, there is no chance for plasma damage to occur.

For this experiment, a test structure was used without the usual diode protection. The test structures have a comb antenna on the metal 1 level. The antenna ratio of this antenna varies from 100 to 100 000. These antennas are connected to a pad on metal 2 level. After patterning the pad, silicon nitride is deposited as a protective layer. Next the pad area is opened by etching. In one case, the etching is performed by reactive ion etch; in the other case a downstream etch process is used. As a reference, the results are compared to test structures that did not get the final silicon nitride deposition nor the etching to open the pads.

Figure 7.6 shows the % of failing gates for the various antenna ratios processed with the two etch processes and the reference without the etching. The reference and the downstream etcher show the plasma damage proportional to the antenna ratio. This plasma damage is caused by the metal etching and/or inter metal dielectric deposition. The reactive ion etching generates a severe plasma damage independent of the metal 1 antenna ratio.

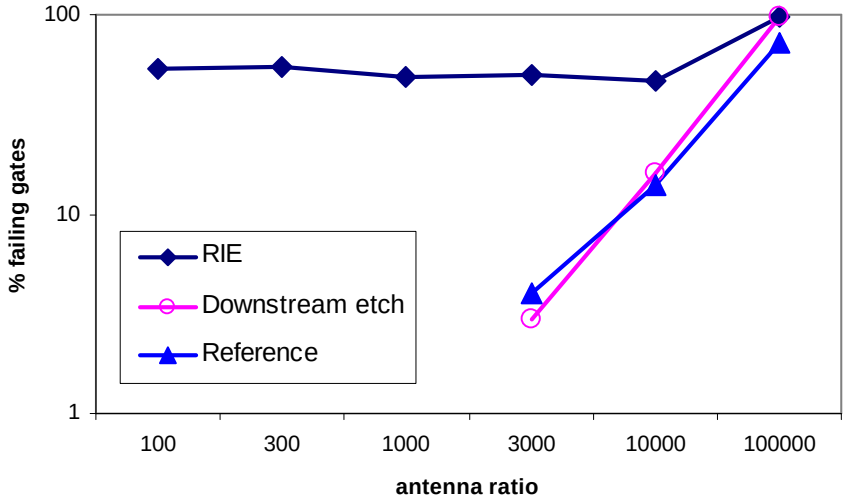


Figure 7.6: The percentage of failing gates for the various antenna ratios processed with the 2 etch processes and the reference without the etching. Reactive ion etch process used for opening the pads induces a severe plasma damage. By replacing the reactive ion etch process by a downstream etch process, all plasma damage is eliminated.

Concluding, reactive ion etch process used for opening the pads induces a severe plasma damage. By replacing the reactive ion etch process by a downstream etch process, all plasma damage is eliminated.

7.3 Charging induced damage by water rinsing

In this paragraph, charging induced damage to metal-insulator-metal-capacitor is reported. The damage is caused by the build up of charges on an oxide surface during a water rinsing step. The excessive charging over a large capacitor area results in a discharge over the inter metal dielectric layer towards a grounded structure. This charging induced damage leads to direct severe yield loss. The charging has been detected, measured and reduced with the help of a non contact surface potential measurement. In this way further yield losses have been prevented. A model for the relation between the surface charging potential and the voltage difference between the capacitor and the grounded structure is presented.

In this section, charging induced damage to metal-insulator-metal-capacitor, often used in RF circuits [13], is reported. The metal-insulator-metal capacitor is formed with the metal 1 layer as bottom plate, 20 to 60 nm PECVD SiO_2 as insulator and an additional, thin metal layer (TiN-AlCu-TiN) as top plate (see Figure 7.7). The area of the structure is typically 100 to 1000 µm² and the capacitance is around 1fF/µm².After fabrication of these capacitors in the 0.35 µm IC circuit, further processing is required to cover the metal-insulator-metal capacitor with the inter metal dielectric and process the further interconnects.

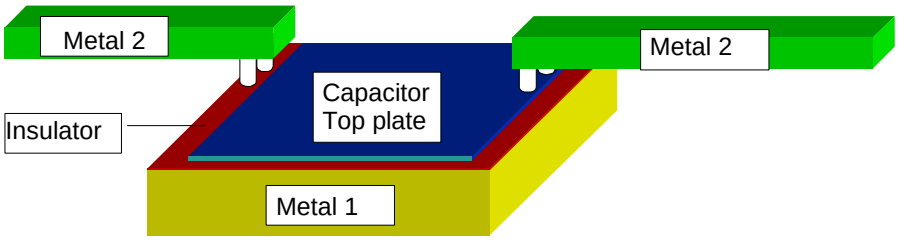


Figure 7.7: Schematic presentation of the construction of the metal-insulator-metal-capacitor.

During this further processing of the wafers, severe defects in the inter metal dielectric were discovered on the edges of these metal-insulator-metal capacitors (see Figure 7.8). This work describes the detection of the source of these defects, the modelling of the failure mechanism and the prevention. The evaluation is provided with in-line plasma damage monitor plots, inline wafer inspection and end-of-line electrical evaluation and yield data.

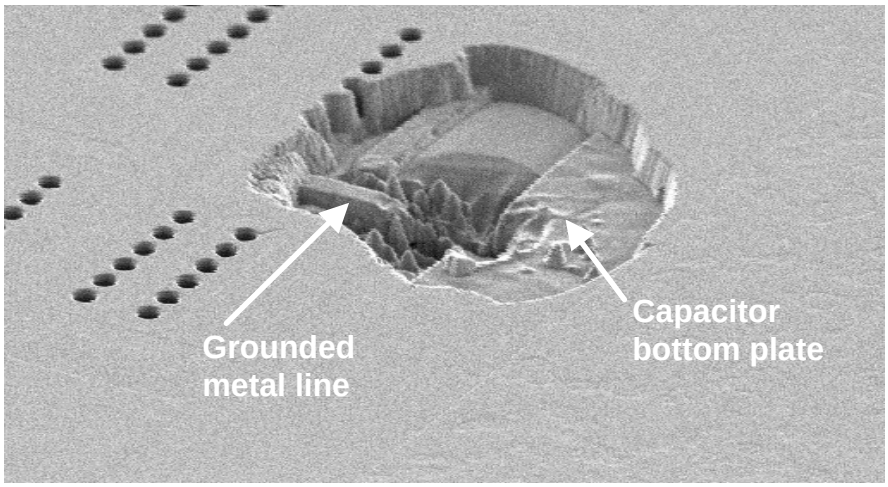


Figure 7.8. SEM picture of severe defects on the edges of metal-insulator-metal capacitor structures. Material from the intermetal dielectric has been ejected by a discharge between the capacitor bottom plate and the grounded metal line.

The defects were discovered right after the definition of the via mask on top of the inter metal dielectric covering the metal-insulator-metal capacitors. This process mainly consists of coating the wafer with resist, exposing the resist with the mask, developing the resist and rinsing the wafer with water after development.

Partitioning showed that the water rinsing step at the end of the process is responsible for generating the defects in the inter metal dielectric. The defect itself is caused by a discharge between the metal-insulator-metal capacitor and a metal line connected to the substrate in the very near proximity of the metal-insulator-metal capacitor. Figure 7.9 shows the damaged region before the deposition of the inter metal dielectric. In order to investigate the relation

between the water rinsing and the discharge like defect generation, we have investigated how much the water rinsing can charge the surface of the wafer. Therefore rinsing processes with different spin speeds were compared.

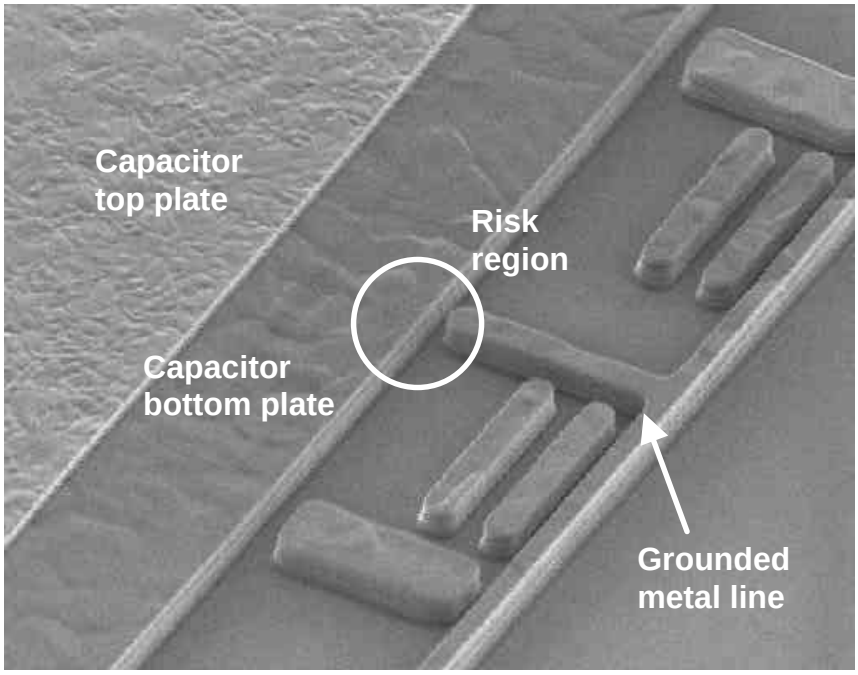


Figure 7.9: SEM picture of the damaged region before the inter metal dielectric deposition. On the left side there is the metal capacitor. On the right side there is the grounded metal line.

Very much like oxide breakdown caused by electrostatic discharge, an electric field over the inter metal dielectric in between the metal-insulator-metal capacitor and the grounded metal line had sufficient magnitude to cause breakdown, allowing charge collected in the capacitor to pass. The discharge is confined to a very small region. Because the oxide is a poor thermal conductor, the discharge causes high energy densities and a rapid rise of temperature [14].

To quantify the charging of the wafer surface caused by the water rinsing step, Plasma Damage Monitor measurement technique has been applied as described in Chapter 3.

In line defect detection has been performed on automatic optical defect detection system. For this application a KLA2550 was used. After etching the vias, the wafer is coated with a TiN layer to protect the inter metal dielectric oxide during further W plug processing. This layer provides an excellent optical contrast, making any topographical defects easily detectable.

Yield of the product has been tested. The results of the yield test have been split up into the different failure mechanisms of the chip. Wafermaps are provided for each wafer and failure mechanism.

The defects were initially discovered after the water rinsing step of the mask development step. Since at that point the wafer is rotated at high speed, the most likely mechanism to generate charges on the wafer surface is by friction. To investigate the impact of the friction, an experiment was set up that compares the impact of the water rinsing process with a high and a low spin speed.

Plasma damage monitor results

Figure 7.10 and Figure 7.11 compare the charging of an oxide layer after spin rinsing at high and at low speed. As can be expected, the low spin speed rinse process generates considerably less charging compared to the high spin speed rinse process.

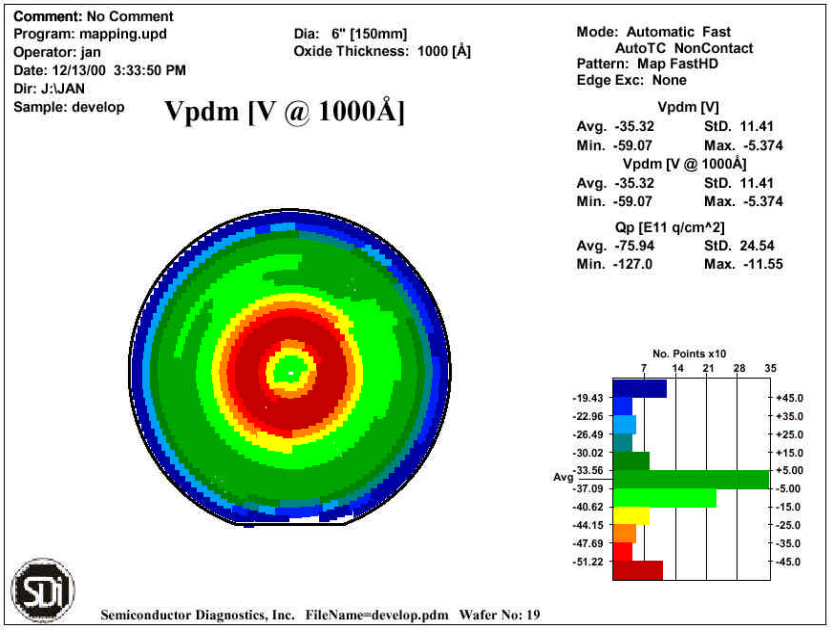


Figure 7.10: Plasma damage monitor mapping of a 100 nm oxide layer after a high spin speed rinse process, showing –59 V charging in the centre ring.

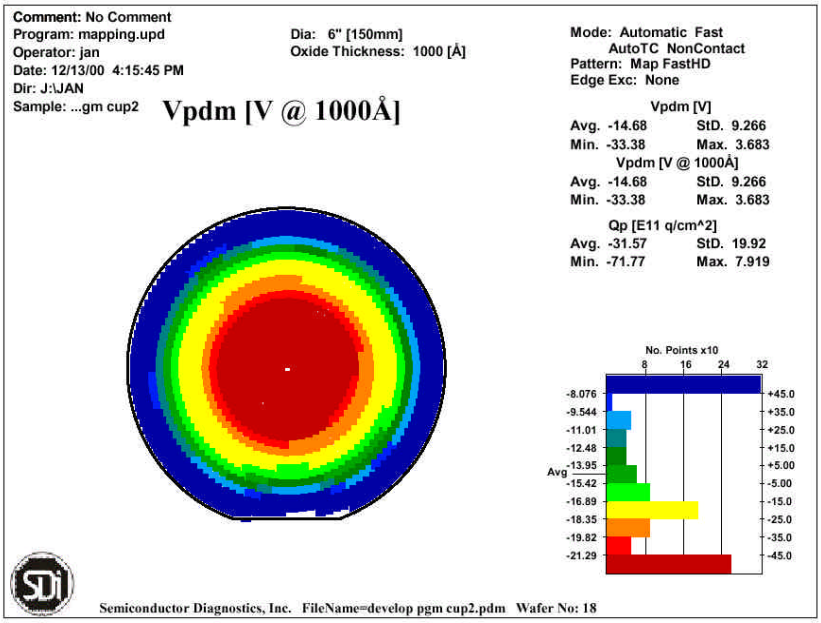


Figure 7.11: Plasma damage monitor mapping of a 100 nm oxide layer after a low spin speed rinse process showing –33 V charging near the centre of the wafer.

The highest levels of charging are measured near the centre of the wafer. During the rinsing, the water jet is sprayed close to the centre of the wafer. K. Yatsuzuka *et al.* describe that ultra pure water can generate a high negative voltage by friction on a thin insulator surface. They found that the voltage reaches an extreme at the contact point of the waterjet and the wafer [15]. This explains why the most negative voltage is measured near the centre of the wafer.

Optical Defect review

Figure 7.12 is showing the mapping of the optical defect review after the via processing of a wafer processed with the high spin speed rinse process. The inspection was done with a KLA5220.

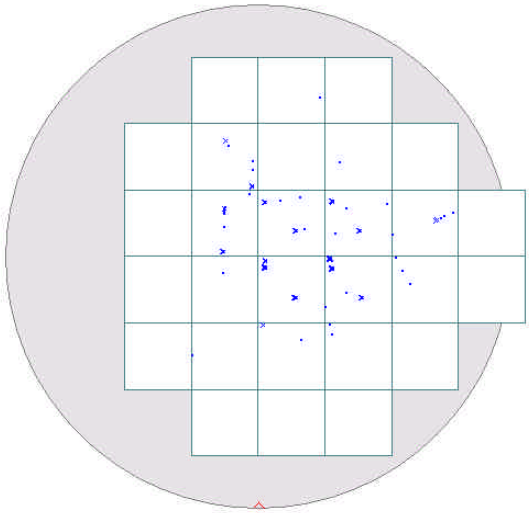


Figure 7.12: Mapping of a of the optical defect review after the via processing of a wafer processed with the high spin speed rinse process: the centre zone shows a concentration of defects.

The optical defect inspection shows defects visible after the via processing. The defects show up exclusively in the region with the high charging.

Further SEM cross-sections show the nature of the defects as depicted in Figure 7.8.

This mapping proves the causal relation between the charges measured by the plasma damage monitor and the ESD like defects on the edges of the metal-insulator-metal capacitors.

Yield Mapping

The impact of the charging during the via processing is also evaluated by electrical yield. The experimental vehicle used for this study is a BiCMOS chip with metal-insulator-metal capacitors. Figure 7.13 shows typical mapping of the dysfunctional dies. Dies are failing because of high electrical leakage. This failure mechanism is strongly related to large defects.

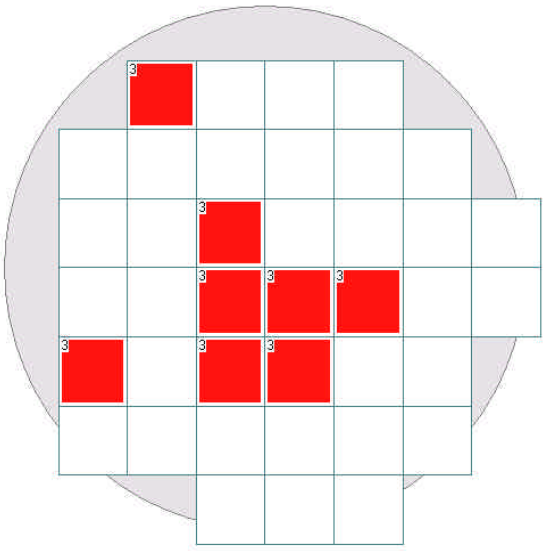


Figure 7.13: Typical yield mapping of a wafer processed with the high spin speed rinse process: the yield in the centre zone is highly affected.

Yield of this experiment is mainly affected in the centre of the wafer. This area coincides very well with charged area and the optical defect mapping. This mapping confirms the causal relation between the charges measured by the plasma damage monitor, the ESD like defects on the edges of the metal-insulator-metal capacitors and the direct yield impact of the failure mechanism.

The voltage between the capacitor and the grounded structure is induced by the charges on the surface of the oxide. In order to calculate this voltage, parasitic capacitances have to be taken in account. Figure 7.14 shows the schematic representation of the relevant capacitances. On the topside, one has the capacitance C1 between the oxide surface and the capacitor top plate. On the bottom side, one has the capacitance C2 between the metal-insulator-metal capacitor bottom plate and the Si substrate. The main capacitance is the metal-insulator-metal capacitor C3.

Considering that the dielectric constant of all capacitor-dielectrics is very similar for all capacitors involved, the voltage over C2 is directly in proportion with the ratio of the thickness of C2 versus the thickness of C1.

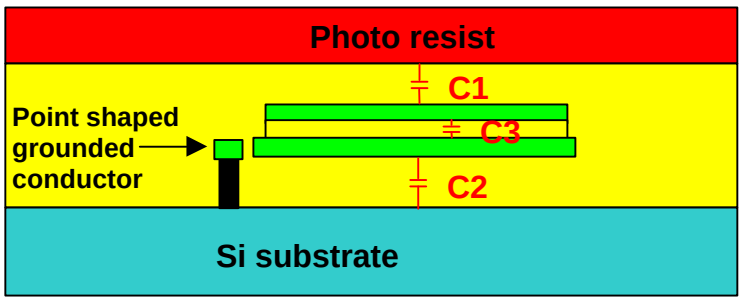


Figure 7.14: Schematic presentation of the parasitic capacitance metal-insulator-metal capacitor.

The thickness if C1 is typically 0.6 µm while the thickness of C2 is typically 1.45 µm. With 60 V measured on the surface of the oxide, about 40 V therefore remains between the capacitor bottom plate and the grounded conductor.

This 40 V is too low to cause the observed ESD-like defects. One has to realize that ESD type defects are not only caused by the voltage potential but in fact by electric field. In the given circumstances, there is a point shaped conductor in front of a large conductor. Under these conditions, electric field can be increased easily by an order of magnitude [16].

To conclude, ESD like discharges are causing severe defects in the inter metal dielectric layer between a metal-insulator-metal capacitor and a close, grounded conductor. This discharge is caused during a high speed water spin rinse. The location of the charges has been measured with plasma damage monitor and correlated with optical defect detection and functional yield loss. The charge density on the oxide surface is maximal in the region where the water jet hits the wafer surface. This effect is enhanced by the spin speed of the wafer.

The combination of the voltage induced on the capacitor by the charges on the oxide surface, the energy stored in the large capacitor and the point shaped conductor results in ESD discharge. The defect mechanism has been avoided by reducing the spin speed during the rinse.

7.4 Conclusion

This chapter describes how after a step-by-step investigation of a conventional 5 metal layer CMOS process, a number of plasma damage inducing process steps have been identified. The investigation has been using the single and multilayer structures described in Chapter 3. For each of the identified cases the plasma damage could removed completely or be reduced to an acceptable low level. This has been done by optimizing the process or where possible replacing the process by a plasma damage free alternative. Each of the process optimizations has been developed, evaluated and implemented taking into account concerns towards yield, defectivity, cost, electrical parameters and process capability. In a number of cases, the low plasma damage version even proved to be also the cost effective option.

High density plasma oxide deposition, Ar preclean sputter etch and reactive ion etch for pad opening were detected as sources of plasma damage. For each of these processes an adequate solution could be implemented. During the investigation also a severe case of damage to metal capacitors was discovered. It was found that ESD like discharges are causing severe defects on the metal-insulator–metal capacitors during the post processing of the devices.

Low pressure, high density plasmas are more and more implemented where a good gap fill is required. This is achieved by combining high density plasma deposition with simultaneous bias sputtering. The high density of the plasma makes the application critical in terms of plasma damage. To avoid direct contact of the high density plasma with the underlying conductor, a liner oxide can be deposited with a less damaging deposition technique. It was found out that the liner deposition does not have any impact on the amount of plasma damage induced.

At the upper corners of the metal lines, deposited oxides are removed by the Ar sputter etch component of the process anyhow. When using the low deposition to sputter ratio, the deposited liner is removed as well together with the HDP deposited oxide and the metal lines become re-exposed to the plasma

When using a high deposition to sputter ratio, the upper corners of the metal lines are covered with oxide already in the early phase of the deposition. This keeps the time that the metal lines are exposed directly to the plasma limited. These results in a reduction of the level of plasma damage.

In order to keep plasma damage induced by a high density plasma oxide deposition minimal, the deposition to sputter ratio has to be chosen as high as possible. As an extra, the high deposition to sputter ratio has a higher final deposition rate. This results in a higher throughput of the equipment and a reduced operation cost.

Ar sputter cleans prior to plug filling in the contact and via modules was found to be one of the sources of plasma damage. The sputter clean process is optimized in terms of plasma density and bias power and process time for minimal plasma damage. The process with the maximal power of plasma density and bias power and also the shortest process time, is the optimum in terms of plasma damage. Under the given process conditions, the time that the gate oxide is stressed by the charging dominates the negative effects of the plasma density or plasma energy.

When considering both plasma damage and via resistance optimization, and the maximal available power, the combination 400 W bias power with 300 W plasma density power was found to be the optimal combination. The process with the reduced plasma damage again has the shorter process time and is more cost effective.

In a number of IC products, the use of protective devices disturbs the functionality of the device. Under these conditions, the process inducing the plasma damage has to be optimized or even replaced by a plasma damage free process. A reactive ion etch process used for opening the pads after the passivation deposition was pinpointed to induce severe plasma damage. Since no protective diodes could be used in the design of the IC, plasma damage was successfully prevented by replacing the reactive ion etch process by a remote plasma downstream etch process.

During the investigation a severe case of damage to metal capacitors was discovered. Metal-insulator-metal capacitor processing includes dielectric deposition, plasma cleaning, oxide etching, metal depositions and patterning and resist stripping. Each of these process steps have an impact on the amount of plasma process induced damage on the metal-insulator-metal capacitor dielectric For this reason protective devices have been implemented in the design. Nevertheless still ESD like discharges have been observed to cause severe defects in the inter metal dielectric layer. The discharges were located between a metal-insulator-metal capacitor and a close by, grounded conductor. Unexpectedly this discharge was caused during a high speed water spin rinse. Friction of the high purity water with the wafer surface rotating at high speed was the source of the charging. The location of the charges has been measured with a plasma damage monitor and correlated with optical defect detection and functional yield loss. The charge density on the oxide surface is highest in the region where the water jet hits the wafer surface. This effect is enhanced by the spin speed of the wafer.

The combination of the voltage induced on the capacitor by charges on the oxide surface, the energy stored in the large capacitor and the point shaped conductor results in ESD discharge. The defect mechanism has been avoided by reducing the spin speed during the rinse.

In two cases, the level of plasma damage could be reduced by introducing a process with a shorter process time. This demonstrates that:

- the time that the gate oxide is stressed dominates the impact of changes plasma
- introducing low plasma damage processing can even be cost effective.

After all sources of plasma damage and excessive charging have been eliminated, the CMOS process flow can be considered as plasma damage free. No yield loss due to plasma damage could be demonstrated anymore.

7.5 References

[1] Proceedings International Symposia on Plasma Process Induced Damage 1-8, 1996-2003.

[2] R.W.Thomas, “An analysis of the quality and reliability supplement to the SIA roadmap”, Microelectronics Reliability 38, pp. 861-868, 1998.

[5] S.Fang, S.Murakawa, and J.P. McVittie, “A new model for thin oxide degradation from wafer charging in plasma etching”, IEDM Tech. Dig., pp.61-64, 1992.

[6] T.Yamada, K. Eriguchi, Y.Kosaka, and K.Hatada, “Impacts of antenna layout enhanced charging damage on MOSFET reliability and performance”, IEDM Tech. Dig., pp. 727-730, 1996.

[7] R.Viswanathan, “PD-free multi-level metal backend for a sub-half micron CMOS process”, Proceedings VMIC, pp.461, 1997.

[8] S. Krishnan, A. Amerasekera, S. Rangan, S. Aur, “Antenna device reliability for ULSI processing” IEDM Tech. Dig., pp. 601-604, 1998.

[9] Eds. K.P. Cheung et al., Proceedings of the 1st Intern. Symp. On Plasma Process Induced Damage, Northern California, May 13-14, 1996.

[10] J.F. Verwij and J.H. Klootwijk, Chapter of American Vacuum Society, Microelectronics Journal, 27, pp. 611, 1996.

[11] A.B. Joshi et al., Proceedings of IEEE International Reliability Physics Symposium, Dallas, TX, 1996.

[12] G.S. Hwang and K.P.Giapis, Proceedings of the 3th Intern. Symposium On Plasma Process Induced Damage, June 4-5, 1998.

[13] S. Decoutere, “A 0.35 µm SiGe BiCMOS Process Featuring a 80 GHz Fmax HBT and Integrated High-Q RF Passive Components”, IEEE BCTM 3.6., 2000.

[14] D.G. Pierce, “Electrostatic Discharge Failure Mechanisms”, ESD Tutorial Notes, ESD Association, 1996.

[15] K. Yatsuzuka, Y. Higashiyama, K. Asano, “Electrification of polymer surface caused by sliding ultrapure water”, Industry Applications Society Annual Meeting, Conference Record of 1994 IEEE, Vol 2. p1400-1405, 1994.

[16] N. Jonassen, “Fundamentals of Electrostatics”, ESD Tutorial Notes, ESD Association, 1996.