1

MICROD

**National Library
of Canada**

**Bibliothèque nationale
du Canada**

Canadian Theses Service

Service des thèses canadiennes

Ottawa, Canada
K1A 0N4

NOTICE

The quality of this microform is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us an inferior photocopy.

Reproduction in full or in part of this microform is governed by the Canadian Copyright Act, R.S.C. 1970, c. C-30, and subsequent amendments.

AVIS

La qualité de cette microforme dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de qualité inférieure.

La reproduction, même partielle, de cette microforme est soumise à la Loi canadienne sur le droit d'auteur, SRC 1970, c. C-30, et ses amendements subséquents.

National Library
of Canada

Bibliothèque nationale
du Canada

Canadian Theses Service Service des thèses canadiennes

Ottawa, Canada
K1A 0N4

The author has granted an irrevocable non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of his/her thesis by any means and in any form or format, making this thesis available to interested persons.

The author retains ownership of the copyright in his/her thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without his/her permission.

L'auteur a accordé une licence irrévocable et non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de sa thèse de quelque manière et sous quelque forme que ce soit pour mettre des exemplaires de cette thèse à la disposition des personnes intéressées.

L'auteur conserve la propriété du droit d'auteur qui protège sa thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

ISBN 0-315-54364-7

Canada

The undersigned recommend to the Faculty of Graduate Studies and Research the acceptance of the thesis :

"Simultaneous Rapid Thermal Oxidation of Silicon and Polysilicon
for Analog CMOS Applications"

submitted by David Guy Richard LeBlanc in partial fulfillment of the requirements for the degree of Master of Engineering.

Prof. N. G. Tarr
Thesis Supervisor

Prof. A. R. Boothroyd
Acting Chairman,
Department of Electronics

October 1989

Abstract

This work investigates the use of simultaneous rapid thermal oxidation (RTO) to form both gate and inter-polysilicon capacitor dielectrics in an analog CMOS process. By forming the capacitor and gate dielectrics simultaneously, before the threshold adjust implant, the amount of redistribution of the threshold adjust impurity profile is minimized and as a result the short channel effects in the p-channel transistor are also minimized.

By growing the thin gate and inter-poly oxides at 1050°C in a rapid thermal processing system, breakdown fields of 8.5 MV/cm and 13.5 MV/cm were realized for the inter-poly and gate oxides respectively. These breakdown fields represent a significant improvement over standard furnace results. The average interface state density, as determined by charge pumping, showed a distinct correlation with both the temperature and duration of an in-situ post-oxidation inert ambient anneal. Protracted, high temperature ($T > 1100^{\circ}\text{C}$) post-oxidation anneals may be required to reduce the interface state densities of RTO oxides to their minimum value. The transistors fabricated with the simultaneous oxidation process showed reasonably good long channel performance for effective channel lengths down to 0.7 μm for the n-channel devices and 0.8 μm for the p-channel devices with no obvious degradation in performance caused by the formation of the inter-poly capacitors.

Acknowledgements

I would like to thank the following for having made it possible for me to perform this work: My thesis supervisor, Dr. Garry Tarr, for providing much advice and guidance especially relating to "full process" device design; Mr. Mike Rowlandson and Dr. Alex Kalnitsky for many enlightening discussions related to process technology and dielectric films; Dr. Joseph Ellul for the use of the rapid thermal processor; Mr. Robert Theriault for the original suggestion of the thesis topic and Northern Telecom Electronics Ltd. for access to its processing and computing facilities. I would also like to thank both Northern Telecom Electronics Ltd. and the Natural Sciences and Engineering Research Council of Canada (NSERC) for their financial support during the course of this work. Finally, yet most of all, I would like to thank my wife, Janice, for her patience and understanding during these trying times.

Table of Contents

List of Figures	vi
List of Tables	x
1. Introduction	1
2. Background.....	10
2.1 Justification for the Poly-2 Active Process	10
2.2 Rapid Thermal Processing of Silicon	16
2.3 RTP Equipment and the Peak ALP-6000.....	21
2.4 Characterization Methods for Thin Dielectrics	25
3. Experimental Procedure and Observations.....	36
3.1 System Calibration	36
3.2 Oxidation Kinetics and Uniformity.....	37
3.3 Simultaneous Oxidation of Silicon and Polysilicon	43
3.4 Experiments within a Full Process Context	56
3.4.1 Process Description	57
3.4.2 Dielectric and Interface Integrity.....	61
3.4.3 Transistor Characterization.....	74
4. Process Integration Concerns	89
4.1 Manufacturability	89
4.2 Improvements to the Process Flow Used in this Work.....	95
5. Conclusions and Recommendations for Future Work.....	98
References.....	101

List of Figures

Figure 1.1(a)	Alternate flows for analog CMOS processes	2
Figure 1.1(b)	Cross-sectional diagrams for the process flows of Figure 1.1(a).....	3
Figure 2.1	Simulated channel region impurity profile for the poly-1 active process.....	12
Figure 2.2	Simulated channel region impurity profile for the poly-2 active process.....	12
Figure 2.3	Simulated dependence of threshold voltage on channel length.....	14
Figure 2.4	Simulated dependence of punchthrough current on channel length	14
Figure 2.5	Simulated subthreshold transfer characteristics for 1 μm gate length transistors	15
Figure 2.6	Schematic diagram of the Peak ALP-6000 RTP System.....	24
Figure 2.7	Energy band diagram for Fowler-Nordheim tunneling with an n-channel device.....	30
Figure 3.1	Required emissivity setting as a function of set-point temperature	38
Figure 3.2	Temperature error resulting from fixed emissivity setting of 760	38
Figure 3.3	Isothermal dry oxidation curves for the Peak ALP-6000 RTP system.....	40
Figure 3.4	Arrhenius plot for the dry oxidation of Figure 3.3.....	40

Figure 3.5	Dependence of the oxide refractive index on the oxidant concentration	42
Figure 3.6	Dependence of oxide thickness uniformity on total gas flow.....	42
Figure 3.7	Isothermal oxidation curves used for fine tuning gate oxide thickness to 15 nm.....	44
Figure 3.8	Sample program used for the initial simultaneous RTO experiment.....	48
Figure 3.9	Polysilicon oxide thickness as a function of oxidation temperature	50
Figure 3.10	Gate oxide thickness as a function of oxidation temperature	50
Figure 3.11	Oxide thickness ratio as a function of oxidation temperature	52
Figure 3.12	Typical gate oxide I-V curves for the initial simultaneous RTO experiment.....	53
Figure 3.13	Fowler-Nordheim plot for the I-V curve of Figure 3.12.....	53
Figure 3.14	Typical inter-polysilicon oxide I-V curve for the initial simultaneous RTO experiment	55
Figure 3.15	Sample RTO program used for the full process experiment	60
Figure 3.16	C-V curves for an MOS capacitor from the full process experiment.....	62
Figure 3.17	C-V curves for an unconnected bond pad near device of Figure 3.16	62

Figure 3.18	High frequency C-V curve for an inter-polysilicon capacitor.....	65
Figure 3.19	Typical gate oxide I-V curve for the full process experiment.....	67
Figure 3.20	Fowler-Nordheim plot for the I-V curve of Figure 3.19.....	67
Figure 3.21	nMOS gate oxide breakdown field histogram.....	69
Figure 3.22	nMOS gate oxide breakdown field cumulative distribution.....	69
Figure 3.23	pMOS gate oxide breakdown field histogram.....	70
Figure 3.24	pMOS gate oxide breakdown field cumulative distribution.....	70
Figure 3.25	Inter-polysilicon oxide I-V curves for the full process experiment.....	73
Figure 3.26	Drain saturation characteristics of n- and p-channel MOS transistors ($L_{nom} = 0.9 \mu m$)	75
Figure 3.27	Substrate current characteristics of n- and p-channel MOS transistors ($L_{nom} = 0.9 \mu m$).....	76
Figure 3.28	Determination of ΔL , R_{sd} and μ_{eff}	79
Figure 3.29	Dependence of threshold voltage on effective channel length.....	81
Figure 3.30	Subthreshold transfer characteristics of long n- and p-channel transistors ($W/L = 12 \mu m / 12 \mu m$).....	83
Figure 3.31	Short n-channel subthreshold transfer characteristics ($L_{nom} = 0.9 \text{ \& } 1.2 \mu m$)	84
Figure 3.32	Short p-channel subthreshold transfer characteristics ($L_{nom} = 0.9 \text{ \& } 1.2 \mu m$)	85

Figure 3.33	Dependence of interface state density on anneal conditions for n- and p-channel transistors	87
Figure 4.1	Wafer to wafer variation in gate oxide thickness for the p-channel batch.....	90
Figure 4.2	Alternative strategies for the metal contact to the second polysilicon layer.....	94

List of Tables

Table 3.1	Estimated and refined times required to grow 15 nm of oxide on single crystal silicon	45
Table 3.2	Experimental matrix for the initial simultaneous RTO.....	47
Table 3.3	Experimental matrix for the full process experiment.....	59
Table 3.4	Threshold voltages and effective channel length parameters for various nominal channel lengths	78

1. Introduction

Many application specific integrated circuit (ASIC) designs are a mixture of analog and digital circuitry - often in the form of switched capacitor filters. This class of applications requires versatile, high performance capacitors. In MOS processes, these capacitors have traditionally been formed using polysilicon layers because of their high temperature compatibility and their ability to form linear, high per-unit-area capacitance structures with low parasitic resistance and high reliability. Two possible process flows for a generic CMOS process providing an analog option are shown in Figure 1.1(a) and cross-sectional "snapshots" at various points in each process are shown in Figure 1.1(b). The process on the left is referred to as "poly-1 active" since the gate electrode is the first polysilicon layer deposited while the second polysilicon layer is used for capacitors only.

If, in the interest of higher packing density, the per-unit-area capacitance of double polysilicon structures is increased by making the dielectric thinner, the reliability of the dielectric layer becomes a major concern. Current technologies form the inter-poly dielectric at temperatures near 900°C and result in average breakdown fields of approximately 4.5 MV/cm and critical fields for tunneling of about 2 MV/cm [1]. If the dielectric thickness were reduced to between 25 and 30 nm, the breakdown voltage would be between 11 and 13 V and the critical voltage for tunneling would be from 5 to 6 V. In switching applications, the effective bias across the capacitor can approach twice the power supply voltage during transients and reduced reliability in designs using 5 V power supplies voltages will occur.

To increase the reliability of the capacitor, it is necessary to increase the average electric field at which the undesirable conduction effects occur. For dielectrics on polysilicon, the irregular surface of the polysilicon has been shown to result in localized electric field enhancement at interface asperities for a given average electric field [1]. Since dielectric

Figure 1.1(a) Alternate flows for analog CMOS processes

Poly-1 Active Process Flow

Poly-2 Active Process Flow

Figure 1.1(b) Cross sectional diagrams for the process flows of Figure 1.1(a)

breakdown and high field tunneling are dominated by these localized effects, it is necessary to create a smoother dielectric to polysilicon interface to improve the effective breakdown field of the device.

The Deal-Grove model for the kinetics of silicon oxidation [2] proposes that two reaction regimes exist. The initial regime, limited by the reaction at the oxide-to-silicon interface, has a linear rate constant with an activation energy of 2 eV and is followed by a regime limited by oxidant diffusion through the oxide already grown. This second regime has a parabolic rate constant with an activation energy of 1.24 eV. Since the activation energy for the surface reaction is higher than the activation energy for the diffusion of the oxidant, higher temperatures result in more diffusion limited reactions. Carim and Sinclair have suggested that high temperature oxidations should result in smoother Si/SiO₂ interfaces due to both diffusion limited growth and the relaxation of stresses in the growing dielectric by viscous flow [3].

High temperature oxidations are apparently necessary to grow high quality thin dielectrics on silicon. This requirement is especially important for dielectrics grown on polysilicon due to the initial surface roughness. Surface limited reactions proceed quickly down grain boundaries creating large interface asperities, whereas a diffusion limited reaction presents a more uniform growth "front" and a smoother interface will result. A number of researchers [1,4,5] have shown that higher oxidation temperatures do indeed result in higher breakdown fields and lower leakage currents for oxides grown on polysilicon.

The initial surface roughness of the polycrystalline silicon films has a large effect on the final roughness of the grown interface. A smoother initial surface may be obtained by depositing the silicon film in the amorphous phase and then recrystallizing in a subsequent anneal cycle. If the dopant impurity is introduced during the deposition (in-situ doping) or implanted prior to recrystallization, the anneal will serve to activate the dopant. This

procedure of recrystallization after doping has also been shown to result in extremely high dopant incorporation within the growing polysilicon grains [6].

Since the deposition rate of amorphous silicon films (especially in-situ doped films) tends to be quite low, a stacked structure consisting of an undoped amorphous seed layer, an in-situ doped, polycrystalline layer to make up the bulk of the film thickness and an undoped amorphous capping layer, may be used to reduce deposition times. This structure, dubbed a Lo-Hi-Lo film, tends to have the same extremely smooth surface of an amorphous film while allowing substantially higher effective deposition rates [7].

The threshold voltage for an MOS transistor is a direct function of the substrate doping, the gate oxide thickness, the fixed charges in the gate oxide and the work function difference between the gate electrode and the substrate [8]. For process simplicity, the gate electrode is made of heavily doped n-type polysilicon. This simplification results in the p-channel work function difference being much less than the n-channel work function difference. The small work function difference, in modern devices with gate oxide thicknesses in the range 20-50 nm and well doping levels of approximately 10^{16} cm^{-3} , results in a p-channel threshold voltage magnitude which is too large and an n-channel threshold which is too small. A shallow implant (projected range much less than maximum depletion layer extent) is used to tailor the threshold voltages of the two transistors to their desired values by emulating the effect of a thin layer of oxide charge at the oxide to semiconductor interface. Since both threshold voltages must be made more positive, an acceptor such as boron, resulting in negative fixed charge, is implanted. In the interest of process simplicity, this implant is often common to both n- and p-channel transistors and any independent fine tuning of the two threshold voltages must be accomplished by adjusting the appropriate well doping level.

As the gate oxide becomes thinner, the oxide capacitance increases and therefore the implant dose required for a given threshold voltage shift becomes larger since, to a first order approximation, the voltage shift is proportional to the oxide thickness multiplied by the net implanted charge. For aggressive technologies with gate lengths less than 1 μm , the dose required for the threshold adjust implant is so large that usually a counter doped region is formed near the surface of the n-well. Chiang, Cham and Rung have shown that this counter doped region results in a buried channel for the p-channel transistor and that maintaining the depth of this region as shallow as possible is critical in reducing short channel effects [9].

Unfortunately, the activation energies for the diffusion of group III and V impurities in silicon, between 3 and 4 eV [10], are higher than either of the oxidation activation energies and, therefore, higher temperature oxidations imply increased dopant diffusion for a given oxide thickness. For a poly-1 active process, the threshold adjust impurity profile is exposed to this increased thermal budget and the depth of the counter doped region will be increased. The advantage of increased oxidation temperature on the inter-polysilicon oxide breakdown field is traded off against the increased redistribution of the threshold adjust impurity profile and the likelihood of short channel effects.

Figure 1.1 also shows a poly-2 active process flow, an alternative to the poly-1 active process. The poly-2 active process avoids exposing the threshold adjust impurity profile to the thermal budget of the capacitor dielectric formation by growing both the gate and capacitor dielectrics simultaneously. While the counter doped surface region of the p-channel device will result in some short channel effects, they can be minimized by making this region as shallow as possible.

There is another major advantage with the poly-2 active process related to circuit design. Not all designs require the use of capacitors and a substantial amount of processing is

avoided if they are not required. With the poly-1 active process, a significant difference in the electrical parameters of the transistors, such as threshold voltage, may exist between batches which include capacitor formation and batches which omit it. Since the capacitor dielectric formation is simultaneous with the gate dielectric formation in the poly-2 active process, there is no difference in the performance of transistors whether or not the capacitor formation is included.

The simultaneous oxidation, poly-2 active process is based upon the premise that appropriate oxide thicknesses may be grown on both the substrate and the polysilicon. The polysilicon oxide will almost certainly be required to be thicker than the gate oxide as the polysilicon to oxide interface could never be as smooth or "clean" as the single crystal silicon to oxide interface and it may have to withstand effective potential drops of up to twice the power supply voltage. In order for the oxidation rate of the polysilicon to be high enough, it is necessary to rely on dopant-enhanced oxidation [11]. That is, the polysilicon must be heavily doped with phosphorus at the time of oxidation. Since both the active channel region and the heavily doped polysilicon must be exposed during this high temperature step, the possibility exists that autodoping of phosphorus from the polysilicon into the active channel will occur causing a threshold voltage shift. Autodoping refers to the mass transport of a dopant species from one region of a wafer to another via the process gas.

Since the segregation coefficient for phosphorus at the silicon to oxide interface is greater than unity and the diffusion of phosphorus in oxide is quite low, the standard procedure for avoiding autodoping of phosphorus during oxidation is to ramp to the process temperature in an oxidizing ambient. This oxidizing temperature ramp is used in the hope of growing a thin capping layer of oxide before any significant dopant out-diffusion into the process gas may occur. The oxidation times being considered for this process are on the order of minutes due to the thin dielectrics and the high temperatures. The combination of the

requirement to ramp in an oxidizing ambient and these short times imply that the use of a standard furnace with its large thermal mass and volume would result in a process which could not be controlled in a manufacturing environment.

A recently developed alternative to the conventional furnace tube is the Rapid Thermal Processing (RTP) system. RTP systems use high fluxes of visible or near-visible light to heat single wafers extremely quickly to a process temperature. Because these systems are normally designed to process single wafers, the volume of process gas is small and may be changed quickly. RTP systems, therefore, offer a possible means of performing the simultaneous growth of gate and polysilicon oxides in a controllable way.

This work investigates the use of such an RTP system to simultaneously grow gate and capacitor oxides in the context of a poly-2 active analog CMOS process with submicron gate lengths. Target thicknesses are 15 to 17.5 nm for the gate oxide and 25 to 30 nm for the inter-polysilicon oxide. Process and device simulations in Chapter 2 provide justification for the arguments outlined above regarding the need for a poly-2 active process. This chapter also presents a critical review of recent trends in rapid thermal processing techniques and systems and the use of such systems to perform high temperature oxidation reactions. Chapter 2 finishes with a brief description of some of the more interesting analysis techniques used in this work.

The experimental work, presented in detail in Chapter 3, initially deals with the differential oxidation kinetics of single crystal and polycrystalline silicon for gate oxide thicknesses near 15 nm. After selection of the appropriate temperature and time to grow the desired thicknesses of oxide, experimental work focuses on the improvement of the quality of the dielectric and interface using an in-situ post-oxidation anneal intended to provide stress relief. Chapter 4 examines a number of issues related to the feasibility of the simultaneous rapid thermal oxidation process. Specifically, techniques used in the etching of the

polysilicon layers, process control difficulties and suggested improvements to the process used in this work are discussed. Finally, in Chapter 5, conclusions regarding the performance of the simultaneous rapid thermal oxidation process are drawn and recommendations for future work are proposed.

2. Background

2.1 Justification for the Poly-2 Active Process

In order to verify the hypothesis that, for thermally grown inter-poly oxides, the poly-1 active process results in enhanced short channel effects in p-channel transistors as compared to the poly-2 active process, numerical simulations of the processing and operation of p-channel transistors were performed. SUPREM III [12], an industry standard one dimensional process simulator, was used to determine the impurity concentration profiles in the channel and source/drain regions of transistors from the two process flows outlined in Chapter 1. The MINIMOS 4.0 [13] program, developed by the Technical University of Vienna, was then used to solve the semiconductor equations in two dimensions for the transistors at various applied biases. The impurity concentration profiles generated by SUPREM were used as input to MINIMOS so that the effect of process variations on electrical performance could be determined.

A poly-1 active process targeted at submicron effective channel lengths in which the gate oxide thickness is 17.5 nm, the inter-poly oxide thickness is 30 nm and the n-well doping concentration is $3 \cdot 10^{16} \text{ cm}^{-3}$ was considered. SUPREM was used to simulate the processing of the p-channel device including the threshold voltage adjust implant, the polysilicon oxidation process and all backend thermal steps (pre-implant source/drain re-oxidation and the combined source/drain anneal and BPSG reflow RTA). The implant was performed using $^{49}\text{BF}_2^+$ at an energy of 50 keV (effective energy for $^{11}\text{B}^+$ of 11.2 keV) and the dose was selected to achieve a threshold voltage of $-1.0 \pm 0.01 \text{ V}$ at the end of processing as determined by SUPREM. The polysilicon oxidation was carried out at 1050°C for 3.5 minutes followed by a nitrogen anneal for 5 minutes at 1050°C to simulate the minimum possible additional thermal budget due to the ramp and soak times in standard furnaces.

The simulated net impurity profiles in the channel region immediately after implant and at the end of processing are shown in Figure 2.1. This figure shows that the capacitor dielectric formation results in an excessive redistribution of the threshold adjust impurity profile. The junction depth of the counter doped region moves from 64.4 nm immediately after the implant to 109.8 nm after the reflow. The dose required to achieve the desired threshold of -1.0 V was $8.9 \cdot 10^{11} \text{ cm}^{-2}$. MINIMOS utilized the impurity profiles generated by SUPREM to perform a two dimensional simulation of a transistor with a 6 μm nominal gate length. The threshold voltage extracted from this simulation was -0.881 V.

The difference in estimated V_t values is due to the different algorithms used by the two simulation programs. SUPREM calculates the sheet conductance of the minority carriers in the substrate as a function of applied gate to substrate bias and extrapolates to zero conductance from the point of maximum slope. The intersection of the extrapolated line and the voltage axis defines the threshold voltage. MINIMOS, however, performs a two dimensional simulation of the channel and source/drain regions and iteratively determines the gate to substrate bias which results in a drain current of $(W/L) \cdot 10^{-7} \text{ A}$. While appearing somewhat arbitrary, this method has the added advantage of being able to predict the degree of short channel effects caused by a given impurity distribution. Both methods are likely to be somewhat in error compared to measured values but should provide accurate indications of trends.

The simulated net impurity profiles for the channel region of the poly-2 active process are shown in Figure 2.2. The only difference between the simulation of this process and the poly-1 active process is the exclusion of the distinct polysilicon oxidation step and modification of the dose of the threshold adjust implant. Since there was considerably less dopant redistribution with this process, the dose required to achieve the desired threshold voltage dropped by about 23%. A dose of $6.88 \cdot 10^{11} \text{ cm}^{-2}$ was found to yield the same

Figure 2.1 Simulated channel region impurity profile for the poly-1 active process

Figure 2.2 Simulated channel region impurity profile for the poly-2 active process

threshold voltage for the 6 μ m transistor as determined by MINIMOS. The smaller dose, in itself, helps to keep the depth of the counter doped region shallower. In the poly-2 active process simulation, we assumed that a suitable temperature may be found at which the gate and inter-poly dielectrics may be grown simultaneously. This temperature was arbitrarily set at 1050°C for the purposes of the simulation. The actual temperature used for the simultaneous oxidation is of little consequence to the impurity profiles as the threshold adjust implant is not performed until after the oxidation and the effect of the gate oxidation on the well profile is insignificant compared to that of the well drive and field oxidation cycles.

For the poly-2 active process, very little redistribution of the threshold adjust impurity profile is seen. The junction between the counter doped region and the well moves only from 60.7 nm after implant to 65.6 nm at the end of the process, representing a reduction of 89.2% compared to the poly-1 active process. A first order calculation, assuming complete depletion, finds the maximum depletion layer thickness for a substrate doping level of $3 \cdot 10^{16} \text{ cm}^{-2}$ to be 180 nm. The change in the depth of the counter doped region is therefore quite significant.

To study the short channel effects resulting from the threshold adjust impurity distribution, MINIMOS was used to simulate the effect of channel length on the device characteristics of both the poly-1 active and the poly-2 active processes discussed above. Specifically, the p-channel threshold voltages and the punchthrough currents (drain current per unit channel width at $V_G=0 \text{ V}$ and $V_D=-5.5 \text{ V}$) for both processes were extracted from simulations as the gate length was varied. Also, the subthreshold characteristics for both processes were simulated for a gate length of 1 μ m and drain biases of -0.1 and -5.0 V. Figures 2.3 and 2.4 show the results of the threshold voltage and punchthrough current simulations respectively while Figure 2.5 presents the results of the subthreshold simulations.

Figure 2.3 Simulated dependence of threshold voltage on channel length

Figure 2.4 Simulated dependence of punchthrough current on channel length

Figure 2.5 Simulated subthreshold transfer characteristics for $1\ \mu\text{m}$ gate length, $24\ \mu\text{m}$ gate width transistors

While Figure 2.3 shows that the poly-2 active process results in a slight lowering of the threshold voltage sensitivity to channel length, Figures 2.4 and 2.5 illustrate more concrete improvements. For channel lengths near $1\mu\text{m}$, the poly-2 active process represents a punchthrough current suppression improvement over the poly-1 active process approaching two orders of magnitude. Figure 2.5 shows that, with respect to subthreshold slope, both of these processes are exhibiting some short channel effects at $L_{\text{gate}}=1\mu\text{m}$ but the situation is worse for the poly-1 active process.

From these results, the benefit of reducing the post threshold adjust implant thermal budget is clearly seen to be a reduction in short channel effects and an extension of the minimum useable channel length. It should be noted that for these simulations, the substrate doping level ($3\cdot 10^{16}\text{ cm}^{-3}$) and the source/drain junction depth ($\approx 0.3\mu\text{m}$) are responsible for a large portion of the short channel effects. If the substrate doping were increased and/or the junction depth decreased, the relative importance of the threshold adjust impurity profile would be greater thereby making the advantage of the poly-2 active process more obvious.

2.2 Rapid Thermal Processing of Silicon

Rapid thermal processing (RTP) is one of a number of recently developed techniques which perform high temperature reactions within short time frames, generally employing a directed beam of energy (photons, electrons, RF) to directly heat a wafer or portion thereof to the process temperature. The techniques may be classified as operating within one of three regimes - adiabatic, heat diffusion and isothermal.

In the adiabatic regime, power levels are highest and process times are shortest. Typically, scanned pulsed beams (laser or e-beam) are used as the energy source. The resulting temperature profile in the wafer is extremely shallow and never reaches steady state. The time span for reactions in this regime is from nanoseconds to microseconds. The gas-

immersion laser doping (GILD) technique for ultra-shallow junction formation using a scanned pulsed XeCl excimer laser [14] is an example of a process technique operating in the adiabatic regime.

The heat diffusion regime refers to a situation where a nearly steady state but non-uniform temperature profile is established. CW scanned beams normally are operated in this regime where dwell times are typically much less than one second. The main application for processes operated in this regime is the zone melt recrystallization of silicon to form SOI structures.

The word, isothermal, implies the lack of a temperature gradient and characterizes the regime within which both RTP and standard furnaces are expected to operate though RTP systems tend to be used for the high-temperature/short-time range of this regime. For RTP, the entire wafer is at one temperature, ideally, and process times are on the order of tenths of a second to a few seconds. The two other regimes, adiabatic and heat diffusion, by their very nature, result in large thermal gradients and stresses which lead to the formation of defects and yield problems [15]. Conversely, isothermal processing inherently avoids this problem by attempting to minimize thermal gradients. For this reason, adiabatic and heat diffusion regime processing have been largely abandoned in recent years.

RTP systems achieve their speed of response by processing wafers individually to avoid the problems of rapid heating and cooling of the large thermal mass of a fully loaded wafer boat and furnace tube. In addition, the small volume of the chamber allows for quick changes of process gases. Multi-processing within a single chamber becomes a possibility by the "switching" on and off of reactions through rapid temperature change and reaction selection through rapid ambient change [16].

The original motivation for the development of RTP was the need for a method to activate implanted dopant species while minimizing diffusion so that electrical characterization of as-implanted impurity profiles could be performed. In this case, there are two competing, thermally activated reactions (diffusion and damage removal/dopant activation) with different activation energies.

Consider two competing reactions with Arrhenius dependence having activation energies E_a and E_b and pre-exponential constants R_{0a} and R_{0b} respectively :

$$R_a(T) = R_{0a} e^{\frac{-E_a}{kT}} \quad (2.1)$$

and

$$R_b(T) = R_{0b} e^{\frac{-E_b}{kT}} \quad (2.2)$$

The ratio of the reaction rates is given by

$$\frac{R_a(T)}{R_b(T)} = \frac{R_{0a}}{R_{0b}} e^{\frac{-(E_a-E_b)}{kT}} \quad (2.3)$$

and, if $E_a > E_b$,

$$\frac{\partial}{\partial T} \left[\frac{R_a(T)}{R_b(T)} \right] > 0 \quad (2.4)$$

Thus, the reaction with the higher activation energy is favoured by performing the reactions at a higher temperature. For example, the diffusion of group III and V impurities in silicon is thought, by some, to proceed via a vacancy mechanism and, as stated earlier, the activation energy for this reaction varies from 3 to 4 eV. Damage removal proceeds via Si self diffusion through vacancy generation and migration and the activation energy for this combined reaction has been found to be near 5 eV [17]. Therefore, by using higher temperatures and shorter times, it is possible to reduce the amount of diffusion occurring

while increasing the damage removal. Since the initial development of the technique for analyzing as-implanted profiles, the use of RTP has become de-rigueur in the formation of shallow implanted junctions in semiconductor processes.

Subsequent uses developed for RTP systems include the sintering of metal to silicon contacts [18] and the reflow of doped, deposited glasses for topography reduction in the backend of semiconductor processes [19], though both of these uses are still essentially inert ambient anneals. The formation of refractory metal silicides (specifically titanium disilicide) has also been performed in RTP systems to avoid the formation of metal oxides on the metal surface [20]. The oxidation rate of titanium is so fast in comparison with the silicidation reaction that all the titanium can be consumed before any appreciable amount of titanium disilicide is formed if there is any oxygen in the process ambient. By performing the reaction in an RTP system where tight control over the temperature and ambient gases may be achieved, the risk of this problem is minimized.

More recently, the use of RTP systems for growth and deposition processes in reactive ambients has been explored. Rapid thermal oxidation (RTO) [21-26] and nitridation [27,28], the so-called limited reaction processing (LRP) deposition of epitaxial and polycrystalline silicon [29, 30] and the in-situ deposition of tungsten gates following rapid thermal oxidation [31] are just a few examples.

It must be emphasized that the motivation for oxidation in an RTP system is not the same as the motivation for rapid thermal annealing. As stated in Chapter 1, the activation energy for the oxidation of silicon ranges from 1.2 to 2 eV depending on the reaction regime, while the activation energy for diffusion of group III and V impurities ranges from 3 to 4 eV. Thus, performing oxidations at increased temperatures results in an increase in the amount of dopant redistribution. The real motivation for RTO lies in the RTP system's ability to perform fast, repeatable ramps to and from process temperature in an oxidizing ambient -

providing a manufacturable access to short-time/high-temperature processes which result in higher quality dielectrics. Specifically for the poly-2 active process, the benefit of RTO is the ability to perform fast, controllable ramps to the process temperature in an oxidizing ambient.

RTO has been shown to live up to its promise of higher breakdown fields and improved interface quality. Fukuda et al. [32] found that thin oxides (3-10 nm) showed higher average breakdown fields and fewer infant failures when grown by RTO at 1050°C and 1150°C as opposed to 800°C, diluted oxygen in argon and 850°C wet furnace oxidations. They also found that the RTO oxides showed both lower mid-gap and minimum interface state densities (as determined by high frequency and quasi-static capacitance-voltage measurements) than the furnace oxides.

Maury et al. [33] and Alvi et al. [34] have both demonstrated that the use of RTO to form thin oxides on polycrystalline silicon results in substantially improved breakdown fields as compared to furnace grown films of comparable thickness. Maury et al. achieved a breakdown field of ~9 MV/cm by RTO of amorphous in-situ doped silicon films while Alvi et al. report breakdown fields in excess of 11 MV/cm for RTO oxides of polycrystalline silicon implanted with 10^{15} cm^{-2} of As, not annealed prior to oxidation. These results compare to the breakdown field of about 4.5 MV/cm for oxides grown at 900°C referred to in Chapter 1.

The work of Maury et al. also showed the importance of the initial roughness of the polysilicon surface prior to oxidation in determining the final conduction properties. They compared oxides grown on amorphous, in-situ phosphorus doped films to oxides grown on silicon films deposited in the polycrystalline phase and subsequently doped using a PBr_3 furnace cycle. The in-situ doped amorphous films showed breakdown fields of ~9 MV/cm compared with ~6 MV/cm for the furnace doped films. SEM photomicrographs

and surface profilometer traces both showed that the polycrystalline deposited films were substantially rougher than the amorphous deposited films which were almost featureless.

Alvi et al. did not consider in-situ doped films, rather they examined RTO of both amorphous and polycrystalline deposited films doped by ion implantation of phosphorus or arsenic at 100 keV. They found that the critical field for Fowler-Nordheim tunneling (the field where Fowler-Nordheim tunneling current dominates over the low-field leakage current) appeared to be well correlated to breakdown field and depended more upon the concentration of the dopant than the crystallinity of the starting film. Constant current stressing, however, showed that oxides grown on polycrystalline deposited films exhibited about four times the charge trapping of oxides grown on amorphous deposited films when electrons were injected into the oxide across the grown oxide to silicon interface.

2.3 RTP Equipment and the Peak ALP-6000

Other than single-wafer versus batch processing, the main difference between RTP systems and standard furnace tubes is that, for RTP systems, heating is generally achieved by direct irradiation of the wafer by visible or near-visible photons. Tungsten-halogen lamp banks and single arc-lamp/reflector assemblies are the most common optical sources. The tungsten-halogen emission spectrum is rather broad and favours free carrier absorption in the semiconductor (eg. $\lambda > 1.2 \mu\text{m}$). Arc-lamp emission spectra, however, are dependent on the gases in which the discharge occurs and it is possible to tune the emission of an arc lamp to the intrinsic (band-to-band) absorption region of the substrate material. If large variations in doping occur over a wafer or if different wafers will have different average doping densities, then the use of a tungsten-halogen based system may result in temperature variations or errors since the degree of free-carrier absorption is tied to the doping density [35].

Because of the high temperatures involved in typical rapid thermal processes, temperature control is a critical issue. Closed loop systems are essential in order to reproducibly attain the high ramp rates that are normally used and since the actual time for which the wafer is exposed to the process temperature is so short, temperature stabilization must be nearly immediate. In earlier systems, temperature measurement was achieved by the use of a thermocouple attached to a small piece of silicon in close proximity to the actual process wafer. Thermocouple samples have a large effective thermal mass due to the electrical connections required to the external world and since they are in a different position in the chamber, the process wafer may be at a different temperature. In addition, the heat loss through the thermocouple wires implies that the silicon immediately near the thermocouple will be at a lower temperature than the rest of the sample. As a result of these problems, thermocouple systems react too slowly and inaccurately for their use in fast ramp-rate closed-loop systems without the use of complicated adaptive control systems. A real-time, non-contact method of temperature measurement is required.

Optical pyrometry [36,37] is such a non-contact method which can measure the temperature of the process wafer itself. Planck's equation governing the spectral radiance of a blackbody is

$$W(\lambda, T) = \frac{C_1}{\lambda^5 \left(e^{\left(\frac{C_2}{\lambda T} \right)} - 1 \right)} \quad (2.5)$$

where $W(\lambda, T)$ is the power radiation per unit blackbody area, per unit wavelength interval at wavelength, λ , for a blackbody at absolute temperature, T . By examining the radiation emitted by the surface of an object, and comparing it to what is expected of an ideal blackbody source, an estimate of the object's temperature may be obtained. A wafer is, of course, not a blackbody but a brownbody with an emissivity less than unity. If the

emissivity of the wafer is known, however, the wafer temperature can be obtained readily from

$$\frac{1}{T} = \frac{1}{T_b} + \frac{\lambda}{C_2} \ln \epsilon(\lambda, T) \quad (2.6)$$

where T is the "true" absolute temperature, T_b is the absolute temperature assuming the surface is a blackbody, and $\epsilon(\lambda, T)$ is the emissivity of the surface at T and λ .

The emissivity of silicon is a strong function of both wavelength and temperature below 900°K and at wavelengths greater than 1 μ m [37]. Since most optical pyrometers examine radiation in the range of 1 to 5 μ m, this sensitivity can result in measurement inaccuracies at lower temperatures. For temperatures above 900°K, however, the emissivity saturates at a value of near 0.7 and varies little with wavelength thereby allowing relatively accurate control of high temperature processes. In RTP systems utilizing optical pyrometry, temperature calibration is achieved by adjusting the gain of the pyrometer. This gain setting is often called the emissivity setting though the relationship between the setting and the actual emissivity is tenuous at best.

The rapid thermal processing system used for this research work was the ALP-6000 manufactured by Peak Systems Ltd. of California. A schematic diagram of this system is shown in Figure 2.6. The ALP-6000 uses a proprietary arc-lamp design (gas mixture) tuned to the intrinsic band-band region of the silicon absorption spectrum to achieve dopant concentration independent wafer heating. The chamber wall is water-cooled, polished stainless steel to maintain the cleanliness of a cold-walled system while minimizing the thermal losses of the wafer due to radiation. The wafer is supported on four quartz pins which are sharpened to minimize thermal conductance and an optical grade quartz plate provides isolation between the process chamber and the arc-lamp.

Figure 2.6 Schematic diagram of the Peak ALP-6000 RTP System

The temperature control system is closed-loop computer-controlled with the temperature measurement performed via optical pyrometry. The pyrometer operates at a wavelength of $5\text{ }\mu\text{m}$, well outside of the emission range of the lamp, from 0.2 to $1.4\text{ }\mu\text{m}$, so that radiation from the lamp is not confused with radiation from the wafer. A CaF_2 window isolates the pyrometer from the process chamber itself.

Control of the ambient during processing is achieved through the use of microprocessor controlled mass-flow-controllers (MFC's). Due to the finite volume of the process chamber and its non-optimized gas flow characteristics, purges for 60 seconds at a flow of 5 SLM were used at the beginning of processing and subsequently whenever a process gas change was required. During actual thermal processes, however, gas flows were limited to 2 SLM to reduce the chances of excessive convective cooling at the wafer edges.

2.4 Characterization Methods for Thin Dielectrics

During the course of this work, numerous techniques were used to characterize the thin dielectric layers that were grown. For the initial experimentation, not dealing with patterned wafers, ellipsometry was used almost exclusively while the later work concentrated on electrical characterization. This section will discuss a few of the techniques used to characterize dielectric films in this work.

Ellipsometry is a technique which allows the measurement of the optical properties of material systems composed of various layers. If each layer of the structure may be considered to be homogeneous and uniform with respect to the parameters, index of refraction, n , extinction coefficient, k , and thickness, then fixed-wavelength ellipsometry provides a method of determining any single parameter or pair of parameters of the structure provided all others are known. The technique measures the change in phase difference, Δ , and the change in amplitude ratio, $\tan(\Psi)$, of a polarized light beam upon

reflection off the surface of the structure . The ellipsometry readings, Δ and Ψ , are related to the complex reflection coefficients, R^P and R^S , for light polarized parallel and perpendicular to the plane of incidence respectively by

$$\tan(\Psi) \cdot e^{i\Delta} = \frac{R^P}{R^S} \quad (2.7)$$

From basic optics theory, formulae may be developed which express R^P and R^S , and therefore Ψ and Δ , as a function of the unknown parameters in a given structure. An iterative numerical technique, such as Newton's method, may then be used to solve for the unknown parameters, from measured Ψ and Δ values [38].

The most common use for ellipsometry, by far, is the measurement of the thickness of oxide layers on silicon. For improved accuracy, it is desirable for the substrate material to be absorbing strongly at the measurement wavelength. Typically, the wavelength used for oxide measurements on silicon is 632.8 nm and since the substrate is so thick compared to the reciprocal absorption coefficient at the measurement wavelength, it may be considered infinite in extent for the purposes of the calculations. The measurement of oxides on polysilicon is more difficult since the structure being measured is substantially more complicated.

The polysilicon layer is typically less than 0.5 μm in thickness and is usually on top of field oxide less than 1 μm thick. The absorption coefficient in polysilicon at a wavelength of 632.8 nm is not high enough to prevent reflections from the poly-field oxide interface or the field oxide-substrate interface. A 4 layer structure would have to be assumed in the computations and any errors in the "known" parameters (such as the field oxide and polysilicon layer thicknesses and the refractive index and extinction coefficient of the polysilicon layer) would result in substantial uncertainty in the final computed results. For this reason, a wavelength of 405 nm is used to measure oxides on polysilicon.

At a wavelength of 405 nm, the absorption coefficient in polysilicon is high enough that reflections from the backside of the polysilicon do not interfere and the extent of the polysilicon layer may be considered infinite for the purposes of computation. Far fewer "known" parameters will be required for the effective two layer structure.

The effective substrate parameters (n_s and k_s) of the polysilicon are affected by grain size and orientation. Therefore, they must be re-measured for each wafer which experiences a different thermal cycle after polysilicon deposition or a different polysilicon deposition cycle itself [39]. The effective substrate coefficients are determined by measuring Δ and Ψ after etching off any grown oxide. Then, using the standard ellipsometry computations, assuming a structure of an infinitely thick substrate, the unknown refractive index and extinction coefficient may be evaluated. Since this procedure is destructive, the Δ and Ψ values for the complete structure must be measured before etching the oxide.

Ellipsometry handles intermediate film thicknesses (>20 nm) well and is the industry standard method for measuring oxide thickness. For very thin films, however, errors may result from the non-uniform refractive index of SiO_2 layers very near the interface or the modulation of the bulk refractive index by thin film stresses. There is disagreement, in this regime, as to whether the thin film's refractive index should be considered fixed or variable to gain accuracy in the thickness calculation. For these reasons, one must assume a possible error of at least ± 5 -10% on film thickness measurements below 20 nm obtained by ellipsometry. For the measurements which assumed a fixed oxide refractive index, values of $n=1.465$ at $\lambda=632.8$ nm and $n=1.470$ at $\lambda=405$ nm were assumed. For the single crystal silicon, substrate coefficients of $n_s=3.858$ and $k_s=0.018$ were used at $\lambda=632.8$ nm while $n_s=5.420$ and $k_s=0.328$ were used at 405 nm [40].

Though most initial work relied on ellipsometry to measure oxide thickness, second and third round experimentation favoured the use of capacitance as an indication of oxide

thickness. The bulk of the characterization performed in this work was electrical in nature. Both capacitor and transistor structures of various geometries were used to characterize the dielectrics, the interfaces and the overall process.

Capacitor structures, both MOS and inter-polysilicon, were used to measure capacitance-voltage (C-V) and current-voltage (I-V) characteristics. The analysis of MOS C-V curves to determine oxide thickness, effective uniform substrate doping density, flat-band voltage and oxide fixed charge was performed in the standard first-order manner as described by Nicollian and Brews [41].

Ideally, the capacitance of an inter-poly capacitor is invariant with bias, fixed at the capacitance of the oxide layer between the two plates. In theory, the degenerate doping of the polysilicon as well as the high concentration of dopant atoms on the grain boundaries pins the Fermi level firmly near the conduction- (or valence- for p^+ poly) band edge. In this case, no depletion regions may form in the polysilicon and therefore no lowering in the overall capacitance is observed as the bias is changed.

More realistically, the concentration of activated dopant within the grains in one or both of the polysilicon layers may not be high enough to ensure that no depletion occurs. If the concentration of activated dopant is not high enough, the capacitance of the structure will exhibit a bias dependence akin to that of an MOS capacitor though greatly attenuated due to the high doping levels. As insulating layers become thinner, to raise per-unit-area capacitance, this problem becomes more significant. First, since operating voltages are generally not scaled with the dielectric thickness, the fields that the structure must support become higher implying that depletion is even more likely. Second, as the capacitance of the dielectric increases, the effect, if any, of the extremely high, yet finite, depletion layer capacitance on the overall capacitance becomes greater.

The problem of depletion of polysilicon can be even more severe if arsenic is used as the dopant. Arsenic is an extremely slow diffuser in silicon and tends to segregate to the grain boundaries if implanted into polysilicon and annealed. Some researchers have even reported the distortion of polysilicon gate MOS capacitor C-V curves due to depletion of the polysilicon gate in the inversion regime [42,43,44].

The I-V characteristic for MOS diodes has a number of distinct regimes. Normally, we assume the insulating material to have a band structure analogous to a semiconductor with a very wide band-gap (approximately 9 eV for SiO₂). The large effective band-gap difference between the oxide and the silicon results in large barriers to electrons in the conduction band (3.2 eV) and holes in the valence band (4.7 eV) at both the polysilicon-to-oxide and the silicon-to-oxide interfaces. Due to the height of the potential barriers, the amount of thermionic emission is completely negligible at normal operating temperatures and for potential drops across the insulator less than the barrier height, the current through the capacitor structure is dominated by parasitic parallel leakage paths.

Once the potential drop across the gate insulator becomes comparable to the barrier height, the width of the barrier is effectively modulated. Figure 2.7 shows the band structure of an n-channel device biased so high into inversion that the effective barrier to electrons has been reduced to half its low field width. In such a situation, carriers may be able to tunnel from the conduction band of the gate electrode into the conduction band of the oxide where they are swept out by the high electric field. Since the probability of tunneling through a barrier is strongly related to the width of the barrier, once this width starts narrowing, the current through the structure rises dramatically. This conduction mechanism, known as Fowler-Nordheim tunneling, is typically the dominant conduction process for MIS diodes with SiO₂ as the insulator [45].

Figure 2.7 Energy band diagram for Fowler-Nordheim tunneling with an n-channel device

A simplified equation which describes Fowler-Nordheim tunneling is

$$J \propto E^2 \exp\left(\frac{-k_2}{E}\right) \quad (2.8)$$

where J is the current density through the structure, E is the electric field in the insulator and k_2 is a constant which depends on the barrier height, ϕ_B , and the effective mass of the carrier in the insulator, m_{eff} , and is given by

$$k_2 = \frac{8\pi (2m_{eff})^{1/2} (\phi_B)^{3/2}}{3qh} \quad (2.9)$$

By graphing the logarithm of measured $J \cdot E^{-2}$ values against E^{-1} , a straight line is obtained from which the barrier height may be determined if the effective mass is known (m_{eff} is usually taken to be $0.4 \cdot m_0$ in SiO_2). Often the extracted barrier height is less than the expected value near 3.2 eV. This effective barrier lowering is likely the result of electric field enhancement at asperities on the injecting interface. Thus, the values of barrier height extracted from this measurement may provide a means of measuring the roughness or quality of the grown interface.

Fowler-Nordheim conduction is usually accompanied by a polarization effect where some of the charges injected into the dielectric become trapped. This effect is readily observed as a change (reduction in the currents) in the I-V characteristic as subsequent sweeps of voltage are applied since the trapped charge distorts the electric field within the dielectric to the extent that current injection is suppressed. The polarization effect may also be seen as a change of current with time for a fixed bias voltage. The charge trapping ability of the structure should be minimized as it leads to instabilities in device parameters such as the threshold voltage.

The tunneling current through the device rises slightly less than exponentially with increasing bias until, eventually, the insulating layer suffers a catastrophic breakdown and

the current level is limited only by parasitic resistances. This is known as dielectric breakdown and, unlike avalanche breakdown in semiconductors, tends to be permanent for SiO_2 layers. As SiO_2 layers become thinner, the breakdown fields on single crystal silicon have been found to increase substantially and, as the thickness tends to zero, they approach the maximum Si-O bond strength of 30 MV/cm [43]. For dielectric thicknesses between 15 and 17.5 nm, breakdown fields on the order of 15-16 MV/cm are the expected upper limit.

As with tunneling characteristics, breakdown field measurements provide a means of measuring the quality of both the dielectric and the injecting interface. If the interface has many asperities, or the asperities which it has are substantial, localized electric field enhancement will reduce the overall measured breakdown field. Also, low field and infant failures may occur due to pin-holes or other defects in the film.

The characterization of the MOS transistors fabricated in this work followed standard analysis techniques. Measurements of threshold voltage, subthreshold slope and punchthrough current suppression as a function of channel length were performed. Measurement of effective channel length was carried out using the method of Laux [46] which correlates measured drain resistance with nominal channel length to determine an offset between nominal and electrical channel lengths.

The Laux technique relies on the accurate knowledge of the threshold voltages of each of the transistors being used to compensate for short channel effects. All threshold voltages referred to in this work were measured by the same method. A fixed, low reverse bias (0.1 V) is applied to the drain terminal and the I_d - V_g transfer characteristic is measured. This curve will have an inflection point at the gate bias where the transverse field begins to have a dominant effect on the mobility of carriers in the channel. The threshold voltage is obtained by extrapolation of a tangent line from this inflection point to the voltage axis. The slope of the tangent at the inflection point also gives the peak transconductance of the

transistor which can subsequently be used to determine mobility if the effective channel length is known.

To measure the average density of energy states in the forbidden gap at the gate oxide to silicon interface, the charge pumping technique was used. The charge pumping phenomenon, initially discovered by Brugler and Jespers in 1969 [47], was used by Groeseneken et al. to develop a reliable technique for the measurement of the the interface state densities of MOS devices [48]. The technique, in its simplest form, is able to measure the total integrated number of traps at the silicon to gate oxide interface over the area of the gate electrode less the region which is swept out by the source and drain depletion regions. In more complex forms the technique may also be used to obtain state density distributions in energy (over the forbidden gap) and in space (over the length of a transistor channel).

A small fixed reverse bias is applied to both the source and drain diffusions of an MOS transistor (or gated diode). By applying a periodic voltage to the gate electrode of sufficient amplitude, the condition at the silicon-to-oxide interface is alternated between strong accumulation and strong inversion. During the inversion portion of the cycle, minority carriers from the source/drain diffusions form the channel charge. Some of these channel carriers will be captured by the interface states. As the gate waveform drives the semiconductor surface back into accumulation, the mobile minority carriers in the channel drift back to the source/drain diffusions due to their reverse bias. The charges trapped by the interface states will recombine with majority carriers from the substrate giving rise to a net substrate current the polarity of which is determined by the majority carrier type.

The net charge which recombines with substrate majority carriers in one cycle, Q_{ss} , was given by Brugler and Jespers as simply

$$Q_{ss} = A_g \cdot q^2 \cdot \overline{D_{it}} \cdot \Delta \Psi_s \quad (2.10)$$

where A_g is the area under the active gate, $\overline{D_{it}}$ is the average (over energy and area) density of states in the forbidden gap of the semiconductor at the gate insulator to semiconductor interface and $\Delta\Psi_s$ is the total range of variation of the surface potential as a result of the gate voltage waveform.

Groesenek et al. developed a more in-depth model of the phenomenon (accounting for gate waveform shape, etc.) which expressed the charge recombined per-cycle for a triangular waveform as

$$Q_{ss} = A_g \cdot q^2 \cdot \overline{D_{it}} \cdot 2 \frac{kT}{q} \left[\ln(v_{th} n_i \sqrt{\sigma_n \sigma_p}) + \ln\left(\frac{|V_{fb} - V_t|}{|\Delta V_g|} \frac{1}{f} \sqrt{\alpha(1-\alpha)}\right) \right] \quad (2.11)$$

Here, v_{th} is the thermal velocity, n_i is the intrinsic carrier concentration in the semiconductor, σ_n and σ_p are the capture cross sections for electrons and holes, ΔV_g is the range of variation of the gate voltage, f is the measurement frequency and α is related to the geometry of the gate voltage waveform. The net current through the substrate due to the charge pumping effect is then simply the charge recombined per-cycle multiplied by the measurement frequency.

By measuring the dependence of the charge pumping current on frequency, the charge recombined per cycle, Q_{ss} , can be graphed versus the logarithm of the measurement frequency. Using this graph, both the average interface state density and the geometrical mean of the capture cross sections for electrons and holes may be determined from the slope and the frequency axis intercept, f_0 , respectively.

$$\sqrt{\sigma_n \sigma_p} = \frac{1}{v_{th} n_i} \cdot \frac{|\Delta V_g|}{|V_{fb} - V_t|} \cdot \frac{f_0}{\sqrt{\alpha(1-\alpha)}} \quad (2.12)$$

$$\overline{D_{it}} = \frac{\partial Q_{ss}}{\partial \log f} \frac{\log e}{2qkT} \quad (2.13)$$

Groeseneken et al. estimated a minimum measurement resolution for this technique of near $10^9 \text{ cm}^{-2} \text{ eV}^{-1}$ for a device with an active gate area of $100 \text{ } \mu\text{m}^2$ at a frequency of 100 kHz. This level of sensitivity represents an order of magnitude improvement over what is commonly achieved with capacitance-voltage techniques.

3. Experimental Procedure and Observations

Three major rounds of experimentation were carried out in this work. Initial work concentrated on characterizing the oxidation kinetics of the specific RTP system which was used. The second major experiment attempted to look more closely at the kinetics of oxidation reactions on both bulk silicon and polysilicon in the range of thicknesses of interest (15-17.5 nm for the gate oxide and 25-30 nm for the inter-poly dielectric). The third experiment investigated the performance of devices formed by this process as well as the implications of incorporation of this technique within a full-process context. The following sections will describe the experimental method utilized and the results obtained. First, however, it is appropriate to discuss the way in which the temperature control system was calibrated.

3.1 System Calibration

Since optical pyrometry may only measure relative temperature, it is necessary to perform a calibration against an absolute temperature measurement device such as a thermocouple. To obtain accurate temperature measurements, the thermocouple must be in intimate physical contact with a wafer which resembles the wafers to be used in the process. The backside preparation of the test wafer should be the same as process wafers so that the actual emissivities of the wafers will be similar. To achieve the intimate contact, a C-type (W-5%Re / W-26%Re) thermocouple was welded by e-beam heating into the center of a blank silicon wafer by Peak Systems Ltd. The leads from the thermocouple junction exited the wafer from the side which faces the lamp so that the pyrometer would not "see" anything but the silicon.

After positioning the thermocouple wafer inside the chamber of the RTP system, the leads were passed outside the chamber to an OMEGA Type C digital thermometer used as a readout device. A program was entered into the Peak RTP system control computer which

stepped the set-point temperature from 450°C to 1100°C in 50°C increments every 30 seconds in an ambient of pure N₂. At each temperature, the emissivity setting was adjusted until the temperature indicated by the thermocouple was within $\pm 1^\circ\text{C}$ (precision of the readout device) of the set-point temperature. The required emissivity setting was recorded for each set-point temperature and the procedure was repeated until consistent results were obtained. The emissivity values as a function of temperature for one such calibration exercise are shown in Figure 3.1.

It is impractical to change the emissivity setting with varying temperature during processing and since the most critical temperatures used in the experiments (oxidation and anneal temperatures) were above 900°C, it was proposed that a constant emissivity setting could result in acceptable temperature errors. To confirm this hypothesis, the calibration program was executed with a fixed emissivity setting of 760 while the thermocouple measured temperature was recorded as a function of set-point temperature. The temperature error as a function of set-point temperature for the fixed emissivity setting is presented in Figure 3.2. This figure shows that a fixed emissivity setting of 760 results in little detectable temperature error between 1000°C and 1050°C. This figure also shows that for temperatures between 900°C and 1100°C, errors of up to $\pm 5^\circ\text{C}$ should be expected.

3.2 Oxidation Kinetics and Uniformity

Once the system was calibrated as described in Section 3.1, it was necessary to determine the kinetics of oxidation of bulk silicon. Standard furnace oxidation kinetics data may not be valid for oxidation in an RTP system for a number of reasons. First, and foremost, a substantial portion of the oxidation for thin oxides at elevated temperatures occurs in the highly non-linear "initial oxidation" regime. Second, some researchers have detected a substantial enhancement in oxidation rates due to UV photon excitation of the oxidizing

Figure 3.1 Required emissivity setting as a function of set-point temperature

Figure 3.2 Temperature error resulting from fixed emissivity setting of 760

surface [22]. Since the Peak ALP-6000 uses an arc lamp optical source with substantial emission in the UV ($\lambda < 0.4 \mu\text{m}$), oxidation rates may be affected.

Oxidations were performed in pure dry O_2 for 20, 40, 80 and 140 seconds at 1050, 1100, 1150 and 1200°C. Fixed-index ellipsometry was used to measure oxide thickness at 9 points within 30 mm of the center of each wafer. The average oxide thickness is graphed versus oxidation time and temperature in Figure 3.3. Using linear regression to fit lines to the isothermal curves through 40, 80 and 140 seconds, a linear oxidation rate is determined for each temperature. These oxidation rates are presented as an Arrhenius relation ($\ln(\text{rate})$ vs $1/kT$) in Figure 3.4 and an approximate activation energy of 1.2 eV is obtained. The curvature in this graph is likely the effect of a lack of data points during the isothermal oxidation curve measurement. Moslehi has reported a linear oxidation regime activation energy of 1.24 eV [21] and Lassig obtained 1.3 eV for dry oxidations in an RTP system similar to the one used in this work [22]. The results presented in Figures 3.3 and 3.4 are reasonably close to the published results considering the lack of data points from which a more accurate measurement could be made.

To determine the effect of oxidant concentration in the gas flow on the oxidation kinetics and the effect of total gas flow rate on temperature (and therefore oxide growth) uniformity, an experiment was devised which varied the flow of Ar and O_2 each over the range 0 to 10 SLM. Flow rates were selected so that the concentration of O_2 in the ambient was either 25%, 50% or 100% and the total flows varied from 1 to 20 SLM. Since the effects of convective cooling are likely to be more important at higher temperatures, the oxidations were performed at 1200°C for 40 seconds. For this oxidation cycle and 100% O_2 , an oxide thickness near 200 Å was estimated from the previous results.

Oxide thickness was measured by ellipsometry with a fixed wavelength of 632.8 nm at 25 points per wafer. Two calculations of oxide thickness were made - one calculation

Figure 3.3 Isothermal dry oxidation curves for the Peak ALP-6000 RTP system

Figure 3.4 Arrhenius plot for the dry oxidation curves of Figure 3.3

assumed a fixed refractive index of 1.465 while the other solved for index as well as thickness. The oxide thickness uniformity was defined as the standard deviation of the fixed index calculated thickness expressed as a percentage of the mean of the same values.

Figure 3.5 shows a graph of the calculated oxide layer index of refraction against the oxidant concentration in the gas flow. The multiple points at each oxidant concentration are for the varying total flow rates. From this figure we note that the effect of the flow rate on the index of refraction appears to be negligible compared to the effect of the concentration. Regardless of whether the extracted refractive index is accurate, a change in it may imply either a stressed or non-stoichiometric film. In either case, the lower index for the lower oxidant concentrations is indicative of a film of lower quality and all subsequent oxidations, in the context of this work, were performed in 100% dry oxygen.

Figure 3.6 shows the dependence of the calculated uniformity upon total process gas flow for the three oxidant concentrations. From this figure, we see that, independent of the oxidant concentration, the uniformity followed essentially the same trend with flow. For flows higher than approximately 5 SLM, the uniformity became worse with increasing total flow at a rate of approximately 0.18%/SLM with an optimal uniformity of about 2% at a flow of 5 SLM. For lower flows, the uniformity also was degraded. The increasing lack of uniformity with increasing flow above 5 SLM was believed to be due to excessive convective cooling of the wafer while the lack of uniformity at flow rates below 5 SLM was believed to be caused by a faulty O₂ mass flow controller (MFC) which provided erratic control of flows less than 2.5 SLM. This MFC was subsequently replaced with one of a higher quality, providing accurate control from 0.05 to 4.95 SLM, though the uniformity experiment was not repeated.

The results indicate that the flow during any high temperature (> 900°C) step should be less than or equal to 5 SLM. This requirement is assumed to apply to high temperature steps in

Figure 3.5 Dependence of the oxide refractive index on the oxidant concentration

Figure 3.6 Dependence of oxide thickness uniformity on total gas flow

both oxidizing and non-oxidizing ambients. All future processes relating to this work used flows of either 0.5 or 2 SLM during the actual high temperature portion of the process and 5 SLM flows during ambient purge steps (immediately following a gas change) performed at temperatures near 700°C.

From the results of Figure 3.3, times were estimated to grow a 15 nm thick layer of oxide at temperatures ranging from 1050 to 1200 °C. Oxidations were performed at these temperatures for the predicted times and the same times $\pm 10\%$ in the expectation of bracketing the actual time required to grow 15 nm. The various conditions used are summarized in Table 3.1. The oxide thickness for each condition was measured by ellipsometry and the results are shown in Figure 3.7. A line was fitted to the results for each temperature by linear regression and its intersection with 15 nm was found (the result for the 49.5 second oxidation at 1125°C was abnormally low and was discarded). By this method, refined estimates for the time required to grow 15 nm at each temperature were obtained and are included in Table 3.1. These refined estimates were used in the subsequent round of experimentation, the simultaneous oxidation of polysilicon and silicon.

3.3 Simultaneous Oxidation of Silicon and Polysilicon

The second round of experimentation attempted to determine the relative dependence of oxidation rate on temperature for bulk silicon and polysilicon and the effect of the polysilicon doping technique (in-situ doped versus implanted) on the thickness and quality of the polysilicon oxide. Preliminary electrical characterization of both the gate and inter-poly oxide layers was attempted.

An experimental matrix of polysilicon doping technique, oxidation temperature (from 1075°C to 1200°C with time set to grow 15 nm of gate oxide) and in-situ post-oxidation

Figure 3.7 Isothermal oxidation curves used for fine tuning gate oxide thickness to 15 nm

Oxidation Temperature (°C)	Estimated Time (sec)			Interpolated Time (sec)
	$\Delta = -10\%$	$\Delta = 0\%$	$\Delta = +10\%$	
1050	153	170	187	169.62
1075	94.5	105	115.5	116.67
1100	72	80	88	85.64
1125	49.5	55	60.5	55.48
1150	36	40	44	40.21
1175	27	30	33	30.07
1200	18	20	22	21.48

Table 3.1 Estimated and refined times required to grow 15 nm of gate oxide

anneal cycle (30 seconds in pure Ar at 1050°C - if used) was developed. Fifty 4-6 Ω -cm, n-type, <100> silicon wafers were oxidized in pyrogenic steam to form a 700 nm field oxide. A layer of Lo-Hi-Lo LPCVD silicon (either in-situ doped or undoped) was deposited to a total thickness of 350 nm. If the polysilicon was deposited undoped, it was implanted with $^{31}\text{P}^+$ with a dose of 10^{16} cm^{-2} at an energy of 40 keV. All wafers were annealed, patterned and then etched so that polysilicon remained on only one half of each wafer. This remaining polysilicon layer was used as a mask to etch the field oxide in diluted hydrofluoric acid so that the final structure was a half wafer of bulk silicon and a half wafer of polysilicon on field oxide. This procedure resulted in a wafer where simultaneous oxidation of polysilicon and bulk silicon could be achieved and observed readily.

The wafers were oxidized according to the matrix shown in Table 3.2. All temperature ramps were performed at 50°C/second to or from the 700°C stabilization temperature with the process gas flowing at 2 SLM. If an anneal was performed, the chamber was purged with pure Ar for 30 seconds at 5 SLM at a temperature of 700°C before the flow was reduced to 2 SLM for the anneal step itself, 30 seconds at 1050°C in pure Ar. A typical program used to perform these oxidations (the temperature and gas flow profiles) is shown graphically in Figure 3.8. The fifty wafers were processed through the RTO system as two separate twenty-five wafer multiple recipe batches such that the first batch contained the wafers which had the in-situ doped first polysilicon layer and the second batch contained the wafers with implanted polysilicon.

Immediately following the oxidation, measurement of the ellipsometry parameters, Δ and Ψ , at a wavelength of 405 nm was performed for both the gate oxide and polysilicon oxide layers. Half of the wafers received a second deposition of in-situ doped Lo-Hi-Lo LPCVD silicon while the others were held for further ellipsometry studies. The second polysilicon

Oxidation Temperature (°C)	Oxidation Time (seconds)	Post-Oxidation Anneal (+/-)	Wafer #
1075	117	+	88DL24B-1,13
"	"	+	88DL24C-12
"	"	-	88DL24B-2,14
"	"	-	88DL24C-13
1100	85	+	88DL24B-3,15
"	"	+	88DL24C-2,14
"	"	-	88DL24B-4,16
"	"	-	88DL24C-3,15
1125	55	+	88DL24B-5,17
"	"	+	88DL24C-4,16
"	"	-	88DL24B-6,18
"	"	-	88DL24C-5,17
1150	40	+	88DL24B-7,19
"	"	+	88DL24C-6,18
"	"	-	88DL24B-8,20
"	"	-	88DL24C-7,19
1175	30	+	88DL24B-9,21
"	"	+	88DL24C-8,20
"	"	-	88DL24B-10,22
"	"	-	88DL24C-9,21
1200	21.5	+	88DL24B-11,23
"	"	+	88DL24C-10,22
"	"	-	88DL24B-12,24
"	"	-	88DL24C-11,24

Table 3.2 Experimental matrix for the initial simultaneous RTO

Figure 3.8 Sample program for the initial simultaneous RTO experiment

layer was patterned into C-V dots with a diameter of 1 mm (area $\approx 7.85 \cdot 10^{-3} \text{ cm}^2$). These wafers received a 45 minute sinter at 450°C in H_2 to reduce the effects of radiation damage during the polysilicon etch and finally received a layer of gold on the backside for improved electrical contact to the substrate.

The wafers held for further ellipsometry studies were used to determine the effective substrate constants of the polysilicon and the polysilicon oxide thickness in the manner described in Chapter 2. The wafers which received a second polysilicon deposition were used to perform initial electrical characterization of the rapid thermally grown oxides.

The first electrical parameters measured were the capacitance of the inter-poly dielectrics at zero bias and the capacitance of the gate oxides in accumulation, at a gate to substrate bias of +3 V. The oxide thickness was extracted from these capacitance measurements assuming a fixed relative permittivity of 3.8 for the oxide layer.

Figure 3.9 shows the dependence of the polysilicon oxide thickness on oxidation temperature for the two different doping techniques. Two thickness measurement techniques are shown - ellipsometry performed as discussed above and capacitance. The excellent agreement between the capacitance results for the two different batches and the obvious discrepancy between the ellipsometry and capacitance determined thicknesses resulted in the ellipsometry data being discarded. The ellipsometric technique requires two complex measurements of Δ and Ψ as well as an etch to remove all the oxide from the polysilicon. In addition, the ellipsometric technique assumes that both the oxide and polysilicon layers are smooth, homogeneous and uniform and that the effective thickness of the polysilicon is infinite. If a choice must be made between the two measurement techniques, the capacitance technique has fewer sources of error, is based upon fewer assumptions and gives the equivalent electrical thickness.

Figure 3.9 Polysilicon oxide thickness as a function of oxidation temperature (Oxidation times vary - see Table 3.2)

Figure 3.10 Gate oxide thickness as a function of oxidation temperature (Oxidation times vary - see Table 3.2)

Figure 3.10 shows the dependence of gate oxide thickness, as measured by accumulation capacitance, on the oxidation cycle. It is evident from this figure that the programs which were supposed to grow 15 nm of gate oxide did not achieve their goal. The gate oxide thicknesses for both of the simultaneous oxidation batches were lower than what was expected from the previous work. The batch using in-situ doped polysilicon was oxidized before the batch using implanted polysilicon and had consistently thicker gate oxides implying higher oxidation temperatures. The impact that the varying observed oxidation rates had on the manufacturability of the process is discussed in Section 4.1.

To obtain an estimate of the oxidation rate enhancement for the polysilicon relative to the bulk silicon, the ratio of the polysilicon oxide thickness to the gate oxide thickness was evaluated and is shown as a function of nominal oxidation temperature in Figure 3.11. This parameter follows a remarkably consistent trend with temperature given the variation in the oxide thicknesses seen in Figures 3.9 and 3.10. The thickness ratio for the in-situ doped polysilicon is slightly lower than that for the implanted polysilicon, confirming the suspected higher oxidation temperature. From the general trend of the thickness ratio with temperature, a temperature difference of about 20°C between the two batches is inferred.

The I-V characteristic for a gate oxide grown at 1075°C without an in-situ post-oxidation anneal is shown in Figure 3.12. This device was from the batch which used in-situ doped polysilicon for the bottom plate of the capacitor. For this curve, the device was biased into strong accumulation and the electrons injected from the substrate see a clean interface with few field enhancing asperities or trapping sites. The negligible difference between the first and second sweeps indicates that there is little polarization of the dielectric due to trapping of charge injected from the substrate. On the second sweep, the voltage was increased until catastrophic breakdown occurred. The breakdown field of 12.3 MV/cm, calculated from

Figure 3.11 Oxide thickness ratio as a function of oxidation temperature

Figure 3.12 Typical gate oxide I-V curves for the initial simultaneous RTO experiment

Figure 3.13 Fowler-Nordheim plot for the I-V curve of Figure 3.12

the capacitance determined oxide thickness of 12.5 nm, is somewhat lower than the predicted maximum field of 15-16 MV/cm.

The high-field portion of the I-V curve was transformed into the Fowler-Nordheim graph shown in Figure 3.13. From this curve, it is possible to extract the barrier height from the slope of the fitted line. If we assume an effective mass for electrons in the oxide of $m_{eff} = 0.4 \cdot m_0$, the barrier is found to be 3.14 eV which is reasonably close to the theoretical value of 3.2 eV.

The I-V curve for a poly-poly capacitor from the same wafer is shown in Figure 3.14. In this figure, substantial evidence of dielectric polarization is seen as the currents for the second pass are lower than those for the first pass. Due to the distortion of the electric fields in the oxide layer from the trapped charge and interface asperities, Fowler-Nordheim analysis would yield little useful information and was not performed. The breakdown field for this device, measured on the second sweep, is 7.1 MV/cm. This breakdown field represents a substantial improvement over the breakdown field of about 4-5 MV/cm for polysilicon oxides grown at 900°C. Due to the difficulties of probing directly into polysilicon (high resistance or intermittently open contacts), comprehensive electrical characterization of these batches was not performed. Of the devices examined, however, little difference existed between the two doping techniques.

If the trends seen in Figure 3.11 are extrapolated to an oxidation temperature of 1050°C, a thickness ratio between 1.65 and 1.7 is expected. Since this trend was found for gate oxide thicknesses near 13 nm, the resulting inter-poly oxide thickness would be 21-22 nm, lower than the desired range of 25-30 nm. By increasing the gate oxide thickness to 17.5 nm (still within the acceptable range of 15-17.5 nm), the polysilicon oxide thickness would be increased to 29-30 nm. For a gate oxide thickness of 17.5 nm, the longer oxidation time will be dominated less by the surface reaction and, as a result, the thickness ratio is likely to

Figure 3.14 Typical inter-poly oxide I-V curve for the initial simultaneous RTO experiment

be lower. Therefore, the polysilicon oxide thickness will probably be closer to the lower end of its acceptable range for the new gate oxide thickness. The breakdown field measurements indicate that an oxidation temperature of 1050°C should still be high enough to take advantage of stress relief due to viscous flow while allowing the desired oxide thicknesses to be achieved.

Thus, the appropriate oxidation cycle for future investigation of the simultaneous oxidation process was determined to be a 1050°C oxidation to grow about 17.5 nm of gate oxide. For such an oxidation cycle, the polysilicon oxide thickness is expected to be within the required range of 25-30 nm. The time required to perform this oxidation at 1050°C was determined using a fine tuning procedure similar to that of Section 3.2. From this procedure, the time required to grow the gate oxide at 1050°C was found to be 244 seconds.

3.4 Experiments within a Full Process Context

The final round of experimentation served a dual purpose. The primary goal was to determine the feasibility of incorporating the simultaneous rapid thermal oxidation procedure into a mixed digital/analog technology. The secondary goal was to determine the effect, if any, of an in-situ high-temperature post-oxidation anneal on the quality of both the grown dielectrics and the oxide to silicon interface for the gate dielectric. It is well known that a high temperature inert ambient anneal immediately following the oxidation reduces the oxide fixed charge density [49] and may reduce the interface state density, when combined with a subsequent low temperature anneal in H₂, possibly due to trace levels of H₂ or H₂O in the post-oxidation anneal ambient [50].

3.4.1 Process Description

To avoid the added complication of patterning to form separate n- and p-wells in a single substrate (eg. full CMOS), two batches (one nMOS and the other pMOS) were processed using substrates of the same type as the appropriate well. This decision allowed the formation of the wells by unmasked ion implants as well as electrical contact to the substrate through the backside of the wafer. This decision also allowed the use of an available mask set intended for the analysis of gate and inter-poly dielectrics as well as MOS transistors with nominal gate lengths down to 0.6 μm .

Starting material was either 10-20 $\Omega\text{-cm}$ p-type silicon (nMOS batch) or 4-6 $\Omega\text{-cm}$ n-type silicon (pMOS batch). After growing a 30 nm sub-nitride pad oxide, the well implant was performed ($2 \cdot 10^{12} \text{ cm}^{-2}$ of $^{11}\text{B}^+$ at 110 keV for the nMOS batch and $2 \cdot 10^{12} \text{ cm}^{-2}$ of $^{31}\text{P}^+$ at 180 keV for the pMOS batch). 120 nm of LPCVD nitride was deposited and patterned so that LOCOS-isolated device well regions could be formed. The 700 nm field oxide was grown at 950°C in pyrogenic steam; this oxidation also served to drive the well implants. After stripping the LOCOS nitride, and etching the pad oxide, a 70 nm thick wet oxide was grown at 950°C to avoid the Kooi effect [51]. 350 nm of in-situ phosphorus doped LPCVD Lo-Hi-Lo silicon was then deposited and annealed for 30 minutes at 850°C. This polysilicon layer was used to form the bottom plate of capacitors only and as such was etched off the device well regions using the 70 nm oxide as an etch stop. After removing the polysilicon from the wafer backside and etching back the wet pregate oxide in diluted hydrofluoric acid, the simultaneous rapid thermal oxidation step was performed.

The oxidation cycle for all wafers was 244 seconds at 1050°C in pure O_2 . Temperature ramps to/from the stabilization temperature of 700°C were performed at either $\pm 50^\circ\text{C}/\text{sec}$ or $\pm 150^\circ\text{C}/\text{sec}$ with some wafers receiving an in-situ post-oxidation anneal cycle in pure Ar while others did not. Table 3.3 illustrates the factors used in the design of the experiment

as well as which wafers from each of the two batches received which combination of factors. A typical program used in the RTP system for this experiment (for a wafer which had a 1050°C, 30 second in-situ post-oxidation anneal and used $\pm 50^\circ\text{C}/\text{sec}$ ramp rates) is shown in Figure 3.15.

Following the rapid thermal oxidation, the threshold voltage adjust implant was performed using $9 \cdot 10^{11} \text{ cm}^{-2}$ of $^{49}(\text{BF}_2)^+$ at an energy of 75 keV (corresponding to an effective $^{11}\text{B}^+$ energy of 16.8 keV). The gate poly deposition, like the first polysilicon deposition, resulted in a layer of 350 nm of in-situ doped Lo-Hi-Lo LPCVD silicon which also received the 30 minute, 850°C anneal. The second polysilicon etch used an $\text{SF}_6\text{-O}_2$ plasma chemistry with a 60% overetch after end-point detection as a compromise between the possibility of breaching the gate oxide over the source/drain regions and leaving second polysilicon stringers around the patterned first polysilicon. A discussion of the impact which the reversal of polysilicon roles has on the overall process, including etch considerations, is presented in Section 4.1.

A re-oxidation step was performed immediately prior to the source/drain implant to grow a screening oxide over the source/drain regions. The n-channel batch received an implant of $7 \cdot 10^{15} \text{ cm}^{-2}$ of $^{75}\text{As}^+$ at 100 keV followed by $3 \cdot 10^{13} \text{ cm}^{-2}$ of $^{31}\text{P}^+$ at 60 keV while the p-channel batch received a single implant of $3 \cdot 10^{15} \text{ cm}^{-2}$ of $^{49}(\text{BF}_2)^+$ at 80 keV (effective $^{11}\text{B}^+$ energy of 18 keV). A 700 nm thick layer of BPSG was deposited, contact etched and then reflowed to provide improved step coverage for metal over contact walls. The reflow cycle was performed in an RTA system at 1050°C for 20 seconds in N_2 and served to activate the source/drain implants at the same time.

$\text{Al}(1\% \text{ Si})$ was deposited to a thickness of 1 μm and patterned into interconnect and pads so that reliable probing could be accomplished. A sinter in H_2 at 450°C for 45 minutes was used to ameliorate the effects of any radiation damage induced during etching and to reduce

Ramp Rate (°C/sec)	Post-Ox. Anneal		nMOS Wafer #	pMOS Wafer#
	Temp. (°C)	Time (sec)		
±50	1025	1	19	11
"	"	30	13	9
"	"	60	4	2
"	"	120	2	20
±50	1050	1	8	10
"	"	30	22	22
"	"	60	9	18
"	"	120	11	19
±50	1075	1	14	14
"	"	30	17	7
"	"	60	25	17
"	"	120	7	3
±50	1100	1	23	4
"	"	30	6	21
"	"	60	24	12
"	"	120	16	16
±150	1050	1	20	5
"	"	60	12	15
"	1100	1	3	1
"	"	60	15	13
±50	-	-	18	24
±150	-	-	10	8
±50	-	-	21	-
±150	-	-	5	-

Table 3.3 Experimental matrix for the full process experiment

Lamp				Gases		
1	Lamp	Ignite	@ t = 60	1	O2 = 100%	@ t = 0
2	Ramp	to	700°C in 15	2	O2 = 10%	@ t = 40
3	Hold	@	700°C for 15	3	Ar = 100%	@ t = 358
4	Ramp	to	1050°C in 7	4	O2 = Off	@ t = 363
5	Hold	@	1050°C for 244	5	Ar = 10%	@ t = 403
6	Ramp	to	700°C in 7	6	Ar = Off	@ t = 547
7	Hold	@	700°C for 75			
8	Ramp	to	1050°C in 7			
9	Hold	@	1050°C for 30			
10	Ramp	to	700°C in 7			
11	Hold	@	700°C for 20			
12	Lamp	off				

1) All times in seconds
2) Ar and O2 flows are expressed as a percentage of 5 SLM.

Figure 3.15 Sample RTO program used for the full process experiment

the interface state density. Finally, gold was deposited on the wafer backside to provide good electrical contact to the substrate.

3.4.2 Dielectric and Interface Integrity

The initial measurements performed on this batch were to determine whether the basic quality of the grown interfaces and dielectrics was acceptable. That is, both gate and inter-poly dielectrics must have suitably high breakdown fields and suitably low leakage currents. Capacitance as a function of voltage should agree with MOS theory for the gate oxides and should be as close as possible to invariance for the inter-poly oxides. Also, a measure of the grown interface quality is the effective barrier height at the oxide to silicon interface to electrons in the conduction band of the substrate. This barrier height can be determined from examination of the Fowler-Nordheim tunneling regime of the I-V curves.

Figure 3.16 shows a typical set of high-frequency and quasi-static C-V curves for an MOS capacitor from the n-channel batch. The quasi-static C-V curve is badly distorted with a clearly visible slope in the accumulation and inversion regions resulting from a high leakage current. The leakage could be due to either a parallel leakage path or a leakage current directly through the oxide. Figure 3.17 shows the quasi-static C-V curve for an unconnected bond pad near the capacitor in question showing that the leakage current seen previously is most likely not due to the gate oxide itself. Rather, it is the result of a parallel path on the wafer surface. This leakage is the major problem with using high and low frequency C-V curves to determine interface state density, D_{it} . Any leakage current, whether through the oxide layer or not, will cause the extracted D_{it} values to be in error. These errors often mask any changes in interface quality caused by variations in the process. A more reliable method to determine the quality of the interface is the charge pumping technique described previously - results of which will be presented in the following section.

Figure 3.16 C-V Curves for an MOS capacitor from the full process experiment

Figure 3.17 C-V Curves for an unconnected bond pad near the device of Figure 3.16

The threshold adjust implant dose changes the threshold voltage by changing the flat-band voltage, V_{FB} , of the structure. If there has not been a threshold adjust implant, the measurement of V_{FB} leads to an estimate of the oxide fixed charge density, Q_f . The use of V_{FB} in determining Q_f is not possible for these batches since the precision of the threshold adjust implant dose is of the same order as the expected value of Q_f (on the order of 10^{10} cm^{-2}). That is, random variations in the threshold adjust dose may obscure the response of V_{FB} to process induced variations in Q_f . There was, however, no evidence for any unusually large variations in Q_f .

As a result of the above considerations, the high-frequency curves were used only to determine gate oxide thickness, effective substrate doping and to check for mobile ion contamination.

The gate oxide thickness was calculated from the accumulation capacitance of a number of devices from each wafer. For the n-channel batch, the average gate oxide thickness over the batch was $17.2 \pm 0.5 \text{ nm}$ ($1-\sigma$) (ignoring wafer#21 which had an anomalously thick oxide of $\approx 19.3 \text{ nm}$) while the average gate oxide thickness was $17.1 \pm 1.1 \text{ nm}$ over the p-channel batch. The rather large wafer to wafer variation in the gate oxide thickness is discussed in the context of the process manufacturability in Section 4.1. The equivalent uniform doping density in the substrate, as determined from the minimum high frequency capacitance value in inversion, was $2.1 - 2.2 \cdot 10^{16} \text{ cm}^{-3}$ in the n-well while for the p-well it was considerably higher at $5.7 - 5.8 \cdot 10^{16} \text{ cm}^{-3}$. The n-well doping density is lower than the originally desired $4 \cdot 10^{16} \text{ cm}^{-3}$ and the partial counter doping of the well by threshold adjust implant is somewhat to blame.

Under the influence of a positive gate to substrate field, positively charged mobile ions in the gate oxide (such as Na^+ and K^+) migrate towards the oxide/silicon interface causing a negative voltage shift of the C-V curve. If a subsequent negative field is applied across the

oxide, some of the mobile charge migrates back away from the interface resulting in a positive voltage shift of the curve. The density of mobile ions in the gate oxide can be estimated from the oxide capacitance and the shift in the flat-band voltage. The field enhanced migration of mobile ions occurs at room temperature, resulting in long-term instability in the characteristics of contaminated devices. Since the effect is greatly enhanced at elevated temperatures, a bias-temperature stress is used to detect the presence of mobile ion contamination.

Bias-temperature stresses were performed at 150°C for 30 minutes or longer at approximate fields of ± 2 MV/cm on field oxide isolated MOS capacitors. The electric field was applied before the ramp to the stress temperature and was not removed until the device had cooled to room temperature after the stress. No significant changes in any of the high-frequency capacitance analysis results were seen after the stresses implying that the mobile ion concentration in these oxides is quite low. This result was expected since the Peak RTP system is a cold-walled system and there are fewer "hot" sources of alkali metals near the wafer than in a conventional furnace tube. Bias-temperature stresses of control samples, known to be contaminated, showed significant changes in V_{FB} and V_t after stressing under the same conditions.

High frequency C-V characteristics of the poly-poly capacitors were also measured on a number of devices. A typical curve is shown in Figure 3.18. For this figure, the bias voltage is applied to the second poly plate with respect to the first poly plate. The curve is flat for negative bias but begins to decline as the bias becomes positive. As the top plate bias increases, negative charge accumulates on the bottom plate and positive charge accumulates on the top-plate. Since a reduction in high-frequency capacitance occurs as the top plate bias becomes more positive and both plates are doped n-type, the positive charge accumulated on the top plate may be a localized depletion charge near the polysilicon-to-

Figure 3.18 High frequency C-V curves for an inter-polysilicon capacitor

oxide interface. The reduction of capacitance is not witnessed for the opposite bias polarity indicating that only the second polysilicon plate is susceptible to this depletion effect.

Since the second polysilicon layer does not experience the high thermal budgets which the first polysilicon does, it is possible that its dopants may not be completely activated or may have become trapped in the grain boundaries. Such a lack of active dopant in the grains of the second polysilicon alone might explain the depletion of the second polysilicon but not the first. This problem could likely be minimized by optimization of the in-situ doped polysilicon deposition process and the post deposition polysilicon anneal.

Current-voltage (I-V) curves of both MOS structures and poly-poly capacitors were also measured. Most I-V curves, for both gate oxides and inter-poly oxides, were measured with the top plate of the capacitor biased positively, injecting electrons across the grown interface. The rationale for this convention was that the deposited interface quality, which would be emphasized by injecting electrons from the top plate, would be a strong function of the second poly deposition process and a weak function of the oxide growth process - the process being investigated. The positive gate bias would cause the n-channel devices to be swept into deep depletion if a source of minority carriers were not present. For this reason, a gated diode structure (with the source grounded) was used to measure the I-V characteristics of the gate oxides.

A typical gate oxide I-V curve from the n-channel batch is shown in Figure 3.19. The Fowler-Nordheim plot for this capacitor is shown in Figure 3.20 and indicates a barrier height for electrons injected from the substrate of 3.28 eV assuming an effective mass for electrons in the oxide of $m_{eff} = 0.4 \cdot m_0$. Both this barrier height and the 13.65 MV/cm breakdown field seen in Figure 3.19 were typical of the devices measured.

Figure 3.19 Typical gate oxide I-V curve for the full process experiment

Figure 3.20 Fowler-Nordheim plot for the I-V curve of Figure 3.19

The distribution of the gate oxide breakdown fields for the n-channel batch is shown in Figures 3.21 in the form of a histogram. Figure 3.22 shows the normally transformed cumulative probability distribution for this data excluding infant failures. For this axis transformation, normally distributed data appears as a straight line and, for a single mode distribution, the standard deviation of the data can be determined from the slope of the line. Two distinct normally distributed regimes are, therefore, detected.

The lower-field mode of the distribution is caused by small defects in the film such as pinholes, inclusions and particulates. Since the electrical effect of the defect is strongly dependent on its extent (size and/or location), the large standard deviation of the low-field mode of the distribution is explained. These low field failures accounted for approximately 30% of the devices tested on the n-channel wafers.

The larger mode of the distribution is a result of the intrinsic failure of the dielectric. If we define the intrinsic breakdown regime as being from 12 MV/cm to 15 MV/cm, the mean intrinsic breakdown field is found to be approximately 13.4 ± 0.5 MV/cm ($1-\sigma$). The dominant factor in the variance of the intrinsic regime breakdown field is the use of a fixed oxide thickness for each wafer in the calculation of breakdown fields. The actual variation of thickness over a wafer results in an effective variation in calculated breakdown field.

Figures 3.23 and 3.24 show the corresponding breakdown field distributions for the p-channel batch. The "intrinsic" breakdown regime does not appear to be as well behaved as for the n-channel batch (eg. it is trimodal). The number of low field failures, however, appears to be lower, accounting for only 5 to 10 % of the total number of devices tested on the p-channel wafers. The mean intrinsic breakdown field for this batch (including all three modes) was 13.3 ± 1.2 MV/cm ($1-\sigma$). The smaller number of low-field failures for the p-channel batch compared to the n-channel batch indicates that the low-field failures are not a result of the RTO process itself, but are more likely the result of particulate contamination.

Figure 3.21 nMOS gate oxide breakdown field histogram

Figure 3.22 nMOS gate oxide breakdown field cumulative distribution

Figure 3.23 pMOS gate oxide breakdown field histogram

Figure 3.24 pMOS gate oxide breakdown field cumulative distribution

The trimodal intrinsic regime for this batch may be due to a substantial fixed resistance (invariant with MOS device area) in series with the p-channel MOS capacitors. Higher tunneling current levels for larger devices would result in a higher potential drop across the fixed resistance and therefore would require a higher gate to substrate voltage to achieve the breakdown field across the dielectric. This resistance would need to be in the range of 100 to 2000 Ω and, therefore, could be a high metal to gate polysilicon contact resistance from the compensating effect of the p^+ implant or the resistance of the lightly doped substrate itself. If this explanation is valid, the actual intrinsic breakdown field for these oxides is more likely to be near the lowest of the three modes (≈ 12 MV/cm). A qualitative correlation between the breakdown field and the device area, supporting the theory of a fixed resistance, was observed by examining the I-V curves used to determine the breakdown field.

The inter-poly capacitors which were used to perform I-V measurements had a second polysilicon layer larger in extent than the first layer. If the inter-poly oxide and its interfaces with the polysilicon layers were ideal, breakdown events would occur exclusively at the edge of the structure where the geometry results in a localized electric field enhancement in the dielectric. By maximizing the amount of coverage of the edges of the first poly by the second poly, any weaknesses at this corner due to field enhancement are emphasized. Thus, the location of the failure is an indication of the quality of the dielectric in that breakdown events occurring somewhere within the area of the capacitor indicate a defect causing localized field enhancement. The large thermal stress of the high current density through a small breakdown region often results in a visible cracking of the polysilicon layers and/or the overlying BPSG. The location of the breakdown event can therefore be determined by applying a high bias to the structure after breakdown and examining it via microscopy.

I-V curves were measured for the inter-poly capacitors on both batches. Figure 3.25 shows a typical curve for the n-channel batch. In this figure, two consecutive sweeps are shown for both positive and negative second polysilicon bias. The curves for the negative second polysilicon bias exhibit a higher level of current conduction as well as a lower "critical" field for high-level conduction as compared to the positive bias situation. This confirms the expectation that the deposited interface is "messier" than the grown interface. The current plateaus for the first pass of both bias polarities are due to the extremely high admittance of the oxide when it is trapping charge causing oscillations in the measurement apparatus (an HP-4145A Semiconductor Parameter Analyzer) when range changes occur.

Breakdown fields (screened for infant failures) measured with a positive second polysilicon bias were 7.96 ± 0.45 MV/cm ($1-\sigma$) and 8.44 ± 0.34 MV/cm ($1-\sigma$) for the n-channel and p-channel batches respectively. For most devices with "good" breakdown fields (> 7.5 MV/cm), the breakdown event occurred where the second polysilicon layer crossed the edge of the first polysilicon layer while most low-field breakdown events occurred somewhere in the center of the device. No correlation between device area and breakdown field was observed for the inter-poly capacitors of either the n- or p-channel batches. This lack of correlation indicates that the series resistance suggested as an explanation for the trimodal p-channel gate oxide breakdown distribution is not due to the metal to polysilicon contact. If the resistance were due to the contact, a similar dependence as for the gate oxides would have been detected. Thus, if the trimodal gate oxide breakdown field distribution was caused by a series resistance, it must be due to the substrate.

A small number of capacitors on the n-channel batch had their breakdown field measured with the top plate biased negatively with respect to the bottom plate. The average breakdown field was slightly lower at 7.63 ± 0.40 MV/cm ($1-\sigma$) and almost all the breakdown events were located somewhere in the area of the capacitor. This result

Figure 3.25 Inter-polysilicon oxide I-V curves for the full process experiment

confirms that the grown interface between the first polysilicon layer and the oxide is of higher quality than the interface between the second polysilicon layer and the oxide.

No correlation was found between the inter-poly oxide breakdown fields and the post-oxidation anneal cycles.

3.4.3 Transistor Characterization

While the integrity of the gate and inter-poly dielectrics themselves is extremely important, equally important is the requirement that the technique of dielectric formation does not degrade overall device and circuit performance. For this reason, MOS transistors were fabricated and evaluated.

The mask used for the "full process" batches had an extensive set of transistors including variable length, constant width arrays and large area closed-geometry transistors. For the constant width array, nominal gate length varied from 0.6 μm to 3 μm while nominal width was fixed at 24 μm . The gate electrodes for all the devices within the array were common as were the sources and the substrate. Only the drain connections were used to select the transistor to be evaluated.

Figure 3.26 shows the drain saturation characteristics of representative n-channel and p-channel transistors with nominal channel lengths of 0.9 μm and nominal channel widths of 24 μm . The gate oxide thickness for both these transistors was near 17 nm. No punch-through is obvious from this measurement, though short channel effects are evident in the lack of saturation of the drain characteristic of the p-channel device. The curvature in the n-channel characteristic for gate biases of 2-3 volts is a result of hot electron generation in the pinched off region near the drain. Severe hot electron effects were expected since no attempt was made to form a lightly doped drain (LDD) structure.

Figure 3.26 Drain saturation characteristics of n- and p-channel MOS transistors ($L_{nom} = 0.9 \mu m$, $W = 24 \mu m$)

Figure 3.27 Substrate current characteristics of n- and p-channel MOS transistors ($L_{nom} = 0.9 \mu m$, $W = 24 \mu m$)

Figure 3.27 shows how the substrate current varied with gate voltage for the same n- and p-channel transistors. The peak ratio of substrate current to drain current is a measure of the hot-electron (or hot-hole for p-channel devices) immunity of the transistor. The p-channel device has a peak fraction of 0.02% for a 5 V drain bias while the corresponding fraction for the n-channel device is 3.39%. High peak substrate current fractions indicate excessive avalanche generation in the drain space-charge layer and result in hot electron injection into the gate oxide. It has been shown that the expected lifetime of the transistor decreases with the cube of the increasing substrate current ratio [52]. Therefore, under stress, the n-channel device would degrade extremely quickly. With the appropriate use of sidewall spacers (not precluded by this process flow) LDD n-channel transistors could be fabricated which would suffer much less from these effects [53].

The effective channel length offset from nominal channel length, ΔL , was determined in the manner described in Section 2.3.2 for a number of transistor arrays of both the n-channel and p-channel batches. The extracted slopes and intercepts of the regressed channel resistance versus nominal channel length lines as well as the threshold voltages used to set the gate overbiases are presented in Table 3.4. The graphs used to determine the parasitic source/drain resistance, the channel length offset and the effective mobility of carriers in the channel are shown in Figure 3.28. For this n-channel array, the parasitic resistance was found to be $R_{sd} \approx 57 \Omega$, the channel length offset was $\Delta L \approx 0.37 \mu\text{m}$ and the effective mobility of electrons in the channel was $\mu_{\text{eff}} = 549 \text{ cm}^2 \cdot (\text{V} \cdot \text{sec})^{-1}$. These values were typical of the n-channel devices measured. For the p-channel transistors, typical values were $R_{sd} \approx 79 \Omega$, $\Delta L \approx 0.45 \mu\text{m}$ and $\mu_{\text{eff}} \approx 168 \text{ cm}^2 \cdot (\text{V} \cdot \text{sec})^{-1}$.

Note that the effective channel lengths of the transistors whose characteristics are presented in Figures 3.26 and 3.27 are $0.53 \mu\text{m}$ and $0.45 \mu\text{m}$ for the n- and p-channel devices respectively. Some short channel and hot electron effects must be expected for transistors

L_{nom} (μm)	V_t (V)
0.6	0.440
0.9	0.806
1.2	0.844
1.5	0.845
1.8	0.838
2.1	0.835
2.4	0.831
3.0	0.830

$V_g - V_t$ (V)	"Slope" ($\Omega/\mu m$)	"Intercept" (Ω)
2	226.32	-26.797
2.5	185.93	-11.759
3	159.37	-1.4286
3.5	141.34	4.9692
4	128.48	9.4519
4.5	119.13	12.738
5	111.97	15.173

Table 3.4 Threshold voltages and effective channel length parameters for Laux's method

Figure 3.28 Determination of ΔL , R_{sd} and μ_{eff}

with effective channel lengths this short. A discussion of proposed improvements to the process used in this experiment is presented in Section 4.2

Sze [45] predicts bulk electron and hole mobilities of ≈ 1000 and $300\text{-}450 \text{ cm}^2\cdot(\text{V}\cdot\text{sec})^{-1}$ respectively for silicon doped to $3\cdot 10^{16} \text{ cm}^{-3}$. The measured mobility values of 549 and $168 \text{ cm}^2\cdot(\text{V}\cdot\text{sec})^{-1}$ both represent a factor of about two decrease from the predicted bulk values and as such are reasonable values for channels near the silicon surface. The large ΔL values, and correspondingly small L_{eff} values, reflect the lack of sidewall spacers during the formation of the source/drain diffusions as well as the non-negligible undercutting of the polysilicon etch.

Figure 3.29 shows the variation of threshold voltage magnitude with effective channel length for the transistor arrays used to measure ΔL . As can be seen, the p-channel device succumbs to short channel effects prior to the n-channel device, due to both the threshold adjust profile induced buried channel and the deeper source/drain junctions.

The subthreshold transfer characteristic can also provide information for the analysis of MOS transistors. The subthreshold slope, S , defined as

$$S = \left. \frac{\partial V_g}{\partial \log I_d} \right|_{V_d \text{ fixed}} \quad (3.1)$$

can be a useful measure of the quality of the oxide to silicon interface as well as the short channel immunity of the transistor. The subthreshold slope is simply the amount of subthreshold gate voltage change required to achieve an increase or decrease in the the drain current by a factor of ten and as such is a measure of how easily the transistor may be turned on or off. A simplified approximation to S for uniform doping, long-channel behaviour and small drain bias is given by Sze [45] as

$$S = \frac{kT}{q} \ln 10 \cdot \left(1 + \frac{C_{\text{depl}}}{C_{\text{ox}}} \right) \quad (3.2)$$

Figure 3.29 Dependence of threshold voltage on effective channel length

For our process with an n-well doping of $\approx 2 \cdot 10^{16} \text{ cm}^{-3}$, a p-well doping of $\approx 5 \cdot 10^{16} \text{ cm}^{-3}$ and a gate oxide thickness of about 17 nm, subthreshold slopes of approximately 74 mV/decade for the p-channel devices and 82 mV/decade for the n-channel devices are expected. Figure 3.30 shows the subthreshold transfer characteristics of long channel devices and the slopes are seen to be 87.9 and 78.8 mV/decade for the n-channel and p-channel devices respectively. The error between the measured and calculated values is most likely due to the non-uniform doping profile in the channel region caused by the threshold adjust implant. The doping profile can result in an error in the effective uniform well doping computed from C-V characteristics and therefore the calculated S value may be in error. In addition, a non-uniform profile near the surface of the channel has been shown to increase the subthreshold slope itself [43].

As gate lengths decrease, the channel can no longer be considered quasi-one dimensional. The drain bias affects the potential distribution over the entire length of the channel, reducing the energy barrier seen by the carriers. This effect, known as drain induced barrier lowering or, in extreme cases, punchthrough, can be readily observed on the subthreshold transfer characteristic of short channel transistors by varying the drain bias as a parameter. Figure 3.31 shows such a set of subthreshold curves with varying drain bias for two n-channel devices while Figure 3.32 shows the corresponding characteristics for two p-channel devices. The results shown in these figures are for nominal gate lengths of 0.9 and 1.2 μm with resulting effective channel lengths of 0.45 and 0.75 μm for the p-channel devices and 0.51 and 0.81 μm for the n-channel devices.

The n-channel devices are seen to be more immune to drain induced barrier lowering than the p-channel devices though some of this effect is attributable to the smaller L_{eff} for the p-channel transistors. A common definition of the limit of acceptable punchthrough is derived from DRAM storage time requirements as being a drain current greater than 10

Figure 3.30 Subthreshold characteristics of long n- and p-channel transistors ($W = L_{\text{nom}} = 12 \mu\text{m}$)

Figure 3.31 Short n-channel subthreshold transfer characteristics
(Drain bias varies from -0.1 to -5.1 V in -1 V steps)

Figure 3.32 Short p-channel subthreshold transfer characteristics (Drain bias varies from -0.1 to -5.1 V in -1 V steps)

pA/ μm of gate width for zero gate bias. For this definition, the nominally 0.9 μm gate length p-channel device suffers punchthrough for drain biases greater than about 4 V while the corresponding n-channel device does not. Both n-channel devices appear to be suffering from some parasitic conduction at 5 V drain bias and low gate bias - most likely hot electron generation in the drain depletion region. Both 1.2 μm nominal gate length devices show some, non-catastrophic, short channel effects. The 5 V drain bias resulted in an effective threshold voltage reduction of about 100 mV for the p-channel devices and 40 mV for the n-channel devices.

The DC characteristics of the transistors were found to be essentially invariant with respect to the various post-oxidation anneal cycles used as well as the ramp rate used for the oxidation cycle itself. The charge pumping measurement of average interface state density, however, exhibited a distinct correlation to the temperature and duration of the post-oxidation anneal, although no significant dependence on the ramp-rate was observed.

Figure 3.33 shows the extracted average interface state density, $\overline{D_{it}}$, as a function of the anneal conditions for both the n-channel and p-channel batches. A reference line is drawn to represent the interface state density for wafers which did not receive any in-situ post-oxidation anneal. The values presented in this figure were obtained using the charge pumping technique described in Section 2.3.2 with a fixed reverse bias on the source and drain junctions of 1 V and a symmetric triangular gate voltage waveform from -2.5 to 2.5 V varying in frequency from 15 to 80 kHz. The transistors investigated were large area closed geometry devices with an inside gate width of 1234 μm and a gate length of 24 μm translating to an area (neglecting the effect of the source/drain depletion regions) of $3.192 \cdot 10^4 \mu\text{m}^2$.

The average interface state density for both n- and p-channel devices drops for higher anneal temperatures and longer anneal times, although the largest reductions in the density

2

OF / DE

2

Figure 3.33 Dependence of interface state density on anneal conditions for n- and p-channel transistors

occur within the first 30 to 60 seconds of the anneal. No saturating minimum value of the interface state density was observed for the temperatures investigated. The actual values of the extracted interface state densities differ from n- to p-channel devices by an approximate factor of two with the n-channel state densities being higher. The reason for this discrepancy is unknown but may be the result of a buried channel for the pMOS device. The minimum values of the densities, $5.04 \cdot 10^9 \text{ (cm}^2 \cdot \text{eV)}^{-1}$ and $9.48 \cdot 10^9 \text{ (cm}^2 \cdot \text{eV)}^{-1}$ for p- and n-channel transistors respectively, are still slightly higher than typically obtained values for standard furnace oxidations ($3\text{-}4 \cdot 10^9 \text{ (cm}^2 \cdot \text{eV)}^{-1}$) and indicate the need for a more intense post-oxidation anneal.

These results imply that even though the interface state densities obtained with a 2 minute, 1100°C post-oxidation anneal are acceptable, higher temperature post-oxidation anneals (up to 1200°C for 30-60 seconds) may be required to reduce the interface state density to its minimum value. As "1/f" noise has traditionally been associated with the interface state density, this may be a necessary optimization for low noise circuit applications.

4. Process Integration Concerns

4.1 Manufacturability

The major obstacle standing in the way of a manufacturable implementation of the poly-2 active simultaneous RTO process, and of any other RTO process, is the lack of repeatability. Throughout this work, large variations in oxide thickness (wafer to wafer and batch to batch) were observed. From the initial work on simultaneous oxidation of silicon and polysilicon (Section 3.3), a difference of approximately 20°C between the actual process temperatures of two consecutive batches was inferred from oxide thickness measurements. Figure 4.1 shows the wafer to wafer variation of mean oxide thickness, as determined by accumulation capacitance, for the "full process" pMOS batch (Section 3.4). The shaded points are the thickness results for blank monitor wafers measured by ellipsometry. The peak variation of oxide thickness from the mean value over the batch is about 8% and is clearly unacceptable in a manufacturing environment. In addition, some wafers are seen to have unacceptably high cross-wafer non-uniformity.

The long term change in temperature calibration must be due to either drift in the performance of the optical pyrometer and the electronics which process its signals or a change in the optical transmission characteristics of the CaF₂ window through which the optical pyrometer is focussed upon the wafer. This type of variation should be easy to detect and possibly correct with the use of Statistical Process Control (SPC) techniques and periodic cleaning and recalibrations.

The large wafer to wafer variations within a batch are much more difficult to alleviate. Either the oxidation rate is higher on some wafers and lower on others, indicating a temperature control problem, or the oxidation time is not being well controlled, indicating a software problem with the control system.

Figure 4.1 Wafer to wafer variation in gate oxide thickness for the p-channel batch

Variations in the reflectivity of the wafer backside can result in variations of the emissivity of individual wafers and, therefore, the actual temperature at which they are processed. If the wafers within a batch are not all from the same silicon supplier and/or have received different processes affecting the backside surface, substantial variations in the oxidation rate would be expected. The results in Figure 4.1, however, are for wafers which all received the same pre-oxidation processes and were all from the same silicon supplier.

If the variations in temperature are not caused by variations in the wafer, they must be caused by the control system. On the RTP system used for this work, a customized Peak ALP-6000, the control program must continuously monitor the wafer temperature, gas flows and the chamber pressure as well as a number of inter-lock signals while concurrently regulating the lamp power supply and updating the status display panel. If, in a program run on the RTP system, a change in gas flow occurred within a few seconds of a temperature change, the response of the control system appeared to be adversely affected. Attempts were made to avoid this particular problem but there may have been other situations which resulted in oxidation times or temperatures being slightly in error. At the high temperatures used in this work, small errors in time or temperature can result in large errors in oxide thickness.

The lack of cross wafer temperature uniformity is another problem which could be alleviated through the use of SPC techniques and periodic maintenance. The cleanliness of the quartz isolation plate, the arc-lamp discharge tube and the chamber itself can all lead to large temperature non-uniformities. If the isolation plate or arc-lamp tube are not perfectly clean, a substantial amount of the radiation may be shadowed from areas of the wafer. If the chamber walls are not clean, their absorptivity may be high enough to cause excessive radiative cooling of the wafer in selected areas. Due to the number of reflections involved in the illumination system, the vertical position of the wafer in the process chamber can

have an effect on uniformity and periodic readjustment of the wafer height should therefore be performed.

Thus, the batch-to-batch and the cross wafer variation in oxide thickness can be controlled through the use of Statistical Process Control. The problem of the wafer-to-wafer variation within a batch is more difficult, however, as it appears to be related to the actual control system software. While the customized Peak ALP-6000 RTP system used for this work had to control too many parameters at once with limited computation resources, production equipment is substantially simpler and may be able to achieve more accurate control of time and temperature. The customized Peak ALP-6000 used for this work may not, therefore, be an appropriate vehicle for the implementation of an RTO process. Other, production oriented, RTP systems (including the newer model from Peak Systems Ltd.) would have to be investigated before any conclusion were made regarding the viability of the RTO process based solely upon uniformity considerations.

An SF_6/O_2 plasma etch chemistry was used for polysilicon etching. For this chemistry, a 30% overetch (etch time past endpoint detection) is used when etching the first polysilicon layer of a poly-1 active process as the only topography is due to the LOCOS defined device wells and the amount of undercutting during the etch is desired to be a minimum. In the same technology, a 120% overetch might typically be used when etching the second polysilicon layer to clear the poly-2 stringers around poly-1 patterns. There is a relatively thick oxide over the source/drain regions in this technology due to the source/drain screening oxide growth or deposition and the oxide loss during the 120% overetch can be tolerated.

A major concern with respect to the manufacturability of the poly-2 process is the etching of the second polysilicon layer to form both the gate electrode and the top plate of the capacitor. This etch process must meet three requirements. The etch must accurately limit

the amount of undercut of the polysilicon to minimize variation in the length of the gate and therefore channel. The selectivity of the etch for polysilicon over oxide must be high enough that it can stop on 17.5 nm of gate oxide. If this selectivity cannot be accomplished, the source/drain regions will be pitted or recessed by the polysilicon etch. The etch should remove second polysilicon stringers from around the first polysilicon patterns. As a compromise, a 60% overetch was used to attempt to minimize the effect of stringers while avoiding the breaching of the gate oxide over the source/drain regions. In a manufacturable process, this compromise should be eliminated.

The contact to second polysilicon was made off the capacitor stack for all devices tested in this work. In a process with this design rule, it is absolutely essential to remove any second polysilicon stringer as it will short all capacitors which share a common bottom plate. By changing the design rules so that contact to the top plate is made on top of the capacitor structure, as shown in Figure 4.2, any stringer becomes isolated and causes no immediate catastrophic effects. However, the presence of a conductor at a floating potential in intimate contact to the inter-poly capacitor may cause reliability problems and the stringer should be prevented if possible. The removal of any edge overlap of the second polysilicon over the first polysilicon layer also helps to increase the breakdown voltage of the structure since the field enhancement at the step is avoided.

There are two ways of preventing the formation of a poly-2 stringer around poly-1 patterns regardless of the design rule for the poly-2 contact. Either the topography at the edge of the poly-1 patterns must be reduced so that the relaxed overetch can clear the stringer or the selectivity of the polysilicon etch must be increased so that it does not breach the gate oxide during the required overetch. Only the former method will be considered here.

The topography at the edge of the first polysilicon layer can be improved by modifying the poly-1 etch technique to form a graded sidewall or by forming a sidewall spacer. The edge

Method i : contact poly-2 off the capacitor stack **Method i i: contact poly-2 on the capacitor stack** **Figure 4.2** Alternative strategies for the location of the poly-2 contact

profile of the polysilicon itself can be customized by alternating short isotropic and anisotropic etch steps. The isotropic steps result in almost equal vertical and horizontal etching while the anisotropic steps etch almost totally vertically. The resulting edge profile is staircase shaped and theoretically would cause fewer problems. The second technique, that of forming a sidewall spacer, is less complicated and has the added advantage that if a stringer does form unexpectedly (through variations in the etch performance, polysilicon thickness, etc.), it will be separated from the capacitor structure by the 100 - 200 nm width of the sidewall spacer. In this situation, the reliability of the capacitor might not be affected at all and it may not be necessary to totally remove the stringers.

The sidewall spacer must be formed just prior to the pregate oxide etchback before the RTO. In this sequence, the sidewall spacer etch must be able to stop on the wet pregate oxide to avoid damage to the underlying silicon. Also, the pregate oxide etchback must not remove the sidewall spacer around the poly-1 patterns. These two requirements rule out the use of an oxide sidewall spacer though some other material such as Si_3N_4 or even polysilicon could possibly be used.

Thus, the problem of second polysilicon etching and stringer removal can be alleviated by contacting the second polysilicon layer on the top of the capacitor stack and integrating a sidewall spacer into the bottom plate formation. The difficulties in patterning the second polysilicon contact on top of the capacitor stack are non-trivial as is the problem of selecting an appropriate sidewall spacer material. If these difficulties are overcome, however, the advantage of a thin, high reliability inter-poly oxide is substantial.

4.2 Improvements to the Process Flow Used in this Work

The electrical characterization of Section 3.4.3 demonstrated that improvements were required in the short channel performance of the p-channel devices and the hot electron

immunity of the n-channel devices. In addition, the long-channel threshold voltages of the n-channel and p-channel devices were not well matched. The measured values were 0.83 and -1.08 V n-channel and p-channel devices respectively. It is important to remember, however, that these experiments were designed to demonstrate the feasibility of implementing a simultaneous RTO poly-2 active CMOS process. The results were not expected to be optimal.

The experimental work showed that the equivalent uniform n-well doping concentration was near $2 \cdot 10^{16} \text{ cm}^{-3}$ as opposed to the desired value of $4 \cdot 10^{16} \text{ cm}^{-3}$. The lighter doping in the well results in larger depletion layers around the source and drain junctions which, if they extend under the gate, can result in enhanced short channel effects. By increasing the dose of the n-well implant, the well doping can be increased to its desired value and short channel effects should be somewhat suppressed. Unfortunately, the increase in the n-well doping level increases the p-channel threshold voltage and therefore implies an even larger dose for the threshold adjust implant. Thus, the benefits of increasing the n-well doping level are of a diminishing nature.

Energies of 75 keV for the threshold adjust implant and 80 keV for the source/drain implant were used with $^{49}(\text{BF}_2)^+$ in the experimental work. These energies resulted in deeper junctions than necessary, enhancing short channel effects. A more appropriate energy for both of these implants would be 50 keV (equivalent $^{11}\text{B}^+$ energy of 11.2 keV). The required increase in the dose of the threshold adjust implant because of the increase in the n-well doping level makes the reduction of the implant energy even more significant.

The increase in the threshold adjust implant dose to reduce the magnitude of the p-channel threshold will increase the magnitude of the n-channel threshold. Therefore, the p-well dose will have to be decreased to compensate this effect. Too much of a decrease in the p-well dose will result in short channel effects on the n-channel transistors. The n-channel

transistor is much less sensitive to the well doping than the p-channel, however, since it does not have a counter doped threshold adjust impurity profile.

A significant amount of the backend thermal budget was in the re-oxidation step used to grow a screening oxide for the source/drain implant. The same goal could be accomplished through the use of a thin layer of deposited oxide with much less redistribution of the threshold adjust implant profile therefore reducing the likelihood of short channel effects further.

The n-channel transistors showed significant hot electron generation in the drain space-charge region. To reduce the amount of hot carrier generation, a Lightly Doped Drain (LDD) structure can be formed to help reduce the peak field in this depletion region. To form the lightly doped drain, an n^- implant is performed prior to the formation of an oxide sidewall spacer around the patterned second polysilicon. After the sidewall spacer has been created, a heavy n^+ implant is used to reduce the source/drain parasitic resistance. These steps are all performed after the second polysilicon has been patterned and are completely compatible with the poly-2 active process.

5. Conclusions and Recommendations for Future Work

The simulations of Chapter 2 confirmed that the high temperature oxidations required to achieve adequate inter-polysilicon oxide strength and reliability represented too high a thermal budget to be used in a poly-1 active analog CMOS process due to the excessive redistribution of the threshold voltage adjust implant and the resulting short channel effects for the p-channel transistor. These simulations also showed that the poly-2 active process was a means of achieving the high breakdown fields and reliability without the excessive thermal budget.

By performing the simultaneous oxidation at 1050°C, suitable dielectric thicknesses for both the transistor and capacitor can be grown (17 and 26 nm respectively) due to the enhanced oxidation rate of the heavily phosphorus doped polysilicon. These thin dielectrics were found to exhibit high breakdown fields of approximately 8 to 8.5 MV/cm for the capacitor oxide and about 13.5 MV/cm for the gate oxide. While most of the credit for the high breakdown fields belongs to the high oxidation temperature, the smoothness of the surface of the in-situ doped Lo-Hi-Lo polysilicon film likely had a positive impact on the capacitor dielectric breakdown field.

The threshold voltages of the transistors were close to the expected values and no obvious signs of autodoping of phosphorus from the polysilicon into the channel region were detected. No degradation in the performance of the transistors could be traced to the simultaneous formation of the inter-polysilicon capacitors. Rather, all short comings in the performance of the p-channel transistors were attributable to less than optimal choices for the doses and energies of the source/drain and threshold adjust implants. In addition, the replacement of the source/drain re-oxidation with a CVD oxide layer deposition should help to improve the p-channel performance. The hot electron susceptibility of the n-channel transistor was due to the lack of a lightly doped drain structure.

None of the measured characteristics, except the interface state density as measured by charge pumping, exhibited any correlation to the in-situ post-oxidation anneal conditions. The different ramp rates investigated (± 50 or $\pm 150^\circ\text{C}/\text{sec}$) had no significant effect on any measured parameter. The average interface state density was found to decrease significantly with increasing post-oxidation anneal temperature and time and no saturation of this effect was detected in the range of temperatures and times explored (1025°C to 1100°C and 1 to 120 seconds). This result implies that higher temperatures and longer times may be required to reduce the interface state density to its minimum value. Since, traditionally, interface state density has been correlated to "1/f" noise, it may be necessary to use a protracted high-temperature post-oxidation anneal if the RTO gate oxide is to be used in a low noise application.

The simultaneous rapid thermal oxidation process could only be considered viable in a manufacturing environment if better temperature control could be achieved by the RTP system and an effective method of removing the second polysilicon stringers could be developed. The causes of the cross wafer uniformity problems and the batch to batch variation in temperatures should be detectable (and possibly controllable) using statistical process control techniques. The wafer to wafer variation within a batch, however, is likely a problem with the control system of the RTP system used for this work. Other RTP systems should be considered before considering this process unmanufacturable due to the temperature variations.

Contacting the second polysilicon layer on top of the capacitor stack is the most effective way of avoiding the catastrophic problem of second polysilicon stringers and, when combined with a sidewall spacer around the first polysilicon layer, should result in a high reliability structure without having to develop new etching techniques.

The non-linearity of the inter-polysilicon capacitors appeared to be caused by the formation of a depletion region in the top polysilicon plate. The deposition, doping and anneal processes for the second polysilicon layer should be varied in an attempt to optimize the linearity of the structure. Specifically, the investigation of completely in-situ doped films rather than Lo-Hi-Lo films as a second polysilicon layer should be performed.

Future work should also involve the investigation of the effects of post-oxidation anneals on interface state density at even higher temperatures and with more time resolution. This work may lead to the determination of an activation energy for this process, the value of which may help to indicate the mechanism involved in reducing the interface state density in high temperature anneals.

This work did not investigate the time dependent dielectric breakdown (TDDB) characteristics of these oxides to predict their long-term reliability. Investigation of the long-term reliability of these oxides should be a priority for any future studies in this area.

References

- [1] P. Heinmann, S. Murarka and T. Sheng, "Electrical conduction and breakdown in oxides of polycrystalline silicon and their correlation with interface texture", *Journal of Applied Physics*, Vol. 53, No. 9, pp.6240-6245, 1982.
- [2] B. Deal and A. Grove, "General Relationship for the Thermal Oxidation of Silicon", *Journal of Applied Physics*, Vol. 36, No. 12, pp.3770-3778, 1965.
- [3] A. Carim and R. Sinclair, "The Evolution of Si/SiO₂ Interface Roughness", *Journal of the Electrochemical Society*, Vol. 134, No. 3, pp.741-746, 1987.
- [4] R. Anderson and D. Kerr, "Evidence for surface asperity mechanism of conductivity in oxide grown on polycrystalline silicon", *Journal of Applied Physics*, Vol. 48, No. 11, pp.4834-4836, 1977.
- [5] H. Huff, R. Halvorson, T. Chin and D. Guterman, "Experimental Observations on Conduction Through Polysilicon Oxide", *Journal of the Electrochemical Society*, Vol. 127, No. 11, pp.2482-2488, 1980.
- [6] D.F. Waechter and N.G. Tarr, "Low Temperature Crystallization of In-Situ Phosphorus Doped LPCVD Amorphous Silicon", *Canadian Journal of Physics*, Vol. 65, pp.1030-1032, 1986.
- [7] H.Y. Tsoi, J.P. Ellul, S.P. Tay, J. Saeki and J.J. White, "Electron Conduction and Breakdown of Thin Polyoxides", *Extended Abstracts of the Electrochemical Society Fall Meeting*, p. 746, 1984.
- [8] S.M. Sze, The Physics of Semiconductor Devices, 2nd Ed., John Wiley and Sons, New York, 1981.
- [9] S. Chiang, K.M. Cham and R.D. Rung, "Optimization of Sub-micron p-Channel FET Structure", *IEDM Technical Digest*, pp.534-537, 1983.
- [10] J.C.C. Tsai, "Diffusion" in S.M. Sze, Ed., VLSI Technology, McGraw-Hill, New York, 1983.

- [11] H. Sunami, M. Koyanagi and N. Hashimoto, "Intermediate Oxide Formation in Double-Polysilicon Gate MOS Structure", *Journal of the Electrochemical Society*, Vol. 127, No. 11, pp.2499-2506, 1980.
- [12] C.P. Ho, S.E. Hansen and P.M. Fahey, "SUPREM III-A Program for Integrated Circuit Process Modelling and Simulation", Technical Report No. SEL84-001, Stanford Electronics Laboratories, Stanford University, Stanford, California, July 1984.
- [13] S. Selberherr, A. Schütz and H. Pötzl, "MINIMOS - A Two-Dimensional MOS Transistor Analyzer", *IEEE Transactions on Electron Devices*, Vol. ED-27, pp.1770-1780, 1980.
- [14] T. Sigmon, "UV Laser Processing of Semiconductor Devices", *MRS Symposium Proceedings on Photon, Beam and Plasma Stimulated Chemical Processes at Surfaces*, Vol. 75, pp.619-632, 1987.
- [15] T.O. Sedgwick, "Short Time Annealing", *Journal of the Electrochemical Society*, Vol. 130, No. 2, pp.484-493, 1983.
- [16] M.M. Moslehi, "Single Wafer Optical Processing of Semiconductors: Thin Insulator Growth for Integrated Electronic Device Applications", *Applied Physics "A" - Solids and Surfaces*, Vol. A-46, pp. 255-273, 1988.
- [17] T.E. Seidel, "Ion Implantation" in S.M. Sze, Ed., *VLSI Technology*, McGraw-Hill, New York, 1983.
- [18] C.S.Pai, E. Cabrerós, S.S. Lau, T.E. Seidel and I. Suni, "Rapid thermal annealing of Al-Si contacts", *Applied Physics Letters*, Vol. 46, No. 7, pp.652-654, 1985.
- [19] J.S. Mercier, "Rapid Flow of Doped Glasses for VLSIC Fabrication", *Solid State Technology*, pp.85-91, July 1987.
- [20] T. Yachi, "Formation of a $\text{TiSi}_2/\text{n}^+\text{Poly-Si}$ Layer by Rapid Lamp Heating and Its Application to MOS Devices", *IEEE Electron Device Letters*, Vol. EDL-5, No. 7, pp.217-220, 1984.

- [21] M.M. Moslehi, S.C. Shatas and K.C. Saraswat, "Thin SiO₂ insulators grown by rapid thermal oxidation of silicon", *Applied Physics Letters*, Vol. 47, No. 12, pp.1353-1355, 1985.
- [22] S.E. Lassig, T.J. Debolske and J.L. Crowley, "Kinetics of Rapid Thermal Oxidation of Silicon", *MRS Symposium Proceedings on Rapid Thermal Processing*, vol. 92, pp.103-108, 1987.
- [23] N. Chan Tung and Y. Caratini, "Rapid Thermal Oxidation of Silicon for Thin Gate Dielectric", *Electronics Letters*, vol. 22, no. 13, pp.694-696, 1986.
- [24] A.M. Hodge, C. Pickering, A.J. Pidduck and R.W. Hardeman, "Silicon Oxidation by Rapid Thermal Processing (RTP)", *MRS Symposium Proceedings on Rapid Thermal Processing*, Vol. 52, pp.313-319, 1986.
- [25] J.C. Gelpey, P.O. Stump and R.A. Capodilupo, "Oxide Growth using the Water-Wall Arc Lamp", *MRS Symposium Proceedings on Rapid Thermal Processing*, Vol. 52, pp.321-325, 1986.
- [26] J. Nulman, J.P. Krusius and A. Gat, "Rapid Thermal Processing of Thin Gate Dielectrics. Oxidation of Silicon", *IEEE Electron Device Letters*, Vol. EDL-6, No. 5, pp.205-207, 1985.
- [27] M.M. Moslehi, K.C. Saraswat and S.C. Shatas, "Rapid thermal nitridation of SiO₂ for nitroxide thin dielectrics", *Applied Physics Letters*, Vol. 47, No. 10, pp.1113-1115, 1985.
- [28] T. Hori, H. Iwasaki, Y. Naito and H. Esaki, "Electrical and Physical Characteristics of Thin Nitrided Oxides Prepared by Rapid Thermal Nitridation", *IEEE Transactions on Electron Devices*, Vol. ED-34, No. 11, pp.2238-2245, 1987.
- [29] C.M. Gronet, J.C. Sturm, K.E. Williams and J.F. Gibbons, "Limited Reaction Processing of Silicon: Oxidation and Epitaxy", *MRS Symposium Proceedings on Rapid Thermal Processing*, Vol. 52, pp.305-312, 1986.

- [30] J.C. Sturm, C.M. Gronet and J.F. Gibbons, "Limited Reaction Processing: *In-Situ* Metal-Oxide-Semiconductor Capacitors", *IEEE Electron Device Letters*, Vol. EDL-7, No. 5, pp.282-284, 1986.
- [31] M.M. Moslehi, M. Wong, K.C. Saraswat and S.C. Shatas, "In-Situ MOS Gate Engineering in a Novel Rapid Thermal/Plasma Multiprocessing Reactor", *Extended Abstracts of the IEEE VLSI Technology Symposium*, pp.21-22, 1987.
- [32] H. Fukuda, T. Iwabuchi and S. Ohno, "The Dielectric Reliability of Very Thin SiO₂ Films Grown by Rapid Thermal Processing", *Japanese Journal of Applied Physics*, Vol. 27, No. 11, pp.L2164-L2167, 1988.
- [33] A. Maury, S. Kim, A. Manocha, K. Oh, D. Kostelnick and S. Shive, "Application of Rapid Thermal Oxidation to the Development of High Dielectric Strength Polyoxides", *IEDM Technical Digest*, pp.676-679, 1986.
- [34] N. Alvi, S. Lee and D. Kwong, "Thin Polyoxide Films Grown by Rapid Thermal Processing", *IEEE Electron Device Letters*, Vol. EDL-8, No. 5, pp.197-199, 1987.
- [35] S. Onishi and K. Sakiyama, "The Influence of Free Carrier Absorption on the Rapid Thermal Sintering Process", *IEEE Electron Device Letters*, Vol. EDL-8, No. 4, pp.176-178, 1987.
- [36] J.C. Gelpey, P.O. Stump and J.W. Smith, "Process Control for a Rapid Optical Annealing System", *MRS Symposium Proceedings on Rapid Thermal Processing*, Vol. 52, pp.199-207, 1986.
- [37] R.E. Sheets, "Temperature Measurement and Control in a Rapid Thermal Processor", *MRS Symposium Proceedings on Rapid Thermal Processing*, Vol. 52, pp.191-197, 1986.
- [38] A. Kalnitsky, S.P. Tay and I.D. Caulder, "Theory and Implementation of a Multilayer Ellipsometric Model", *Journal of the Electrochemical Society*, Vol. 135, No. 5, pp.1271-1274, 1988.

- [39] Jaim Nulman, "Rapid Thermal Oxidation of Polysilicon and Silicon", *MRS Symposium Proceedings on Beam-Solid Interactions and Transient Processes*, Vol. 74, pp.641-646, 1987.
- [40] User Manual for the Rudolph Research AE-IV Ellipsometer, Rudolph Research Ltd.
- [41] E.H. Nicollian and J.R. Brews, MOS (Metal Oxide Semiconductor) Physics and Technology, John Wiley and Sons, New York, 1982.
- [42] C.Y. Lu, J.M. Sung, H.C. Kirsch, S.J. Hillenius, T.E. Smith and L. Manchanda, "Anomalous C-V Characteristics of Implanted Poly MOS Structures in n^+/p^+ Dual-Gate CMOS Technology", *IEEE Electron Device Letters*, vol. EDL-10, no. 5, pp. 192-194, 1989.
- [43] R.A. Chapman, C.C. Wei, D.A. Bell, S. Aur, G.A. Brown and R.A. Haken, "0.5 Micron CMOS for High Performance at 3.3 V", *IEDM Technical Digest*, 1988, pp.52-55.
- [44] C.Y. Wong, J.Y.-C. Sun, Y. Taur, C.S. Oh, R. Angelucci and B. Davari, "Doping of N^+ and P^+ Polysilicon in a Dual-Gate CMOS Process", *IEDM Technical Digest*, 1988, pp.238-241.
- [45] S.M. Sze, The Physics of Semiconductor Devices, 2nd Ed., John Wiley and Sons, New York, 1981.
- [46] Steven E. Laux, "Accuracy of an Effective Channel Length/External Resistance Extraction Algorithm for MOSFET's", *IEEE Transactions on Electron Devices*, Vol. ED-31, No. 9, p.1245, 1984.
- [47] J.S. Brugler and P.G.A. Jespers, "Charge Pumping in MOS Devices", *IEEE Transactions on Electron Devices*, Vol. ED-16, No. 3, pp.297-302, 1969.
- [48] G. Groeseneken et al., "A Reliable Approach to Charge-Pumping Measurements in MOS Transistors", *IEEE Transactions on Electron Devices*, Vol. ED-31, No. 1, pp.42-53, 1984.

- [49] B.E. Deal, M. Sklar, A.S. Grove and E.H. Snow, "Characteristics of the Surface State Charge of Thermally Oxidized Silicon," *Journal of the Electrochemical Society*, Vol. 114, p.266, 1967.
- [50] S.Wolf and R.N. Tauber, Silicon Processing for the VLSI Era. Vol. 1 – Process Technology, Lattice Press, Sunset Beach, California, 1986.
- [51] E. Kooi, J.G van Lierop and J.A. Appels, "Formation of Silicon Nitride at the Si-SiO₂ Interface During Local Oxidation of Silicon and During Heat Treatment of Oxidized Silicon in NH₃ Gas," *Journal of the Electrochemical Society*, Vol. 123, p.1729, 1976.
- [52] C. Hu, S.C. Tam, F.C. Hsu, P.K. Ko, T.Y. Chan and K.W. Terrill, "Hot-Electron-Induced MOSFET Degradation – Model, Monitor, and Improvement", *Transactions on Electron Devices*, Vol. ED-32, No. 2, pp.375-385, 1985.
- [53] M.L. Chen, C.W. Leung, W.T. Cochran, W. Jüngling, C. Dziuba and T. Yang, "Suppression of Hot-Carrier Effects in Submicrometer CMOS Technology", *IEEE Transactions on Electron Devices*, Vol. ED-35, No. 12, 1988.

END

1 3 0 9 9 0

FIN