Advanced Gate Technologies for Deep-Submicron CMOSFETs

by

Hiu Yung Wong

B.Eng. (The Chinese University of Hong Kong) 1999

A dissertation submitted in partial satisfaction
of the requirements for the degree of

Doctor of Philosophy
in

Engineering - Electrical Engineering and Computer Sciences

in the

GRADUATE DIVISION

of the

UNIVERSITY OF CALIFORNIA, BERKELEY

Committee in charge:

Professor Tsu-Jae King, Chair
Professor Peter Y. Yu
Professor Vivek Subramanian

Fall 2006

The dissertation of Hiu Yung Wong is approved.

Chair Date

Date

Date

University of California, Berkeley

Fall 2006

Advanced Gate Technologies for Deep-Submicron CMOSFETs

Copyright © 2006

by

Hiu Yung Wong

Abstract

Advanced Gate Technologies for Deep-Submicron CMOSFETs

by

Hiu Yung Wong

Doctor of Philosophy in Engineering - Electrical Engineering and Computer Sciences

University of California, Berkeley

Professor Tsu-Jae King, Chair

The scaling of CMOSFETs into the deep-submicron regime necessitates the introduction of advanced gate technologies. In the near-term ($\geq 32nm$ technology nodes), poly-crystalline silicon (poly-Si) will continued to be used as the gate material, but requires the introduction of novel gate annealing techniques. This is because Rapid Thermal Annealed (RTA) poly-Si gate electrodes usually show either the gate depletion effect or boron penetration through the gate dielectric.

Pulsed Excimer Laser Annealing (ELA) is proposed as a near-term solution to the gate stack challenges. By using ELA, an amorphous Si gate electrode can be completely melted so that the implanted dopants diffuse rapidly ($D =\sim 10^{-4}cm^2/s$) to the interface and are activated upon Si crystallization. It is found that effective active dopant concentration $> 3 \times 10^{20}cm^{-3}$ can be attained for both nMOSFETs and pMOSFETs, so that the gate depletion effect is reduced to $\sim 0.1nm$ Equivalent Oxide Thickness (EOT). It is also found that, although the gate electrode is heated to very high temperature ($> 1200^oC$), the process time is sufficiently short ($\sim 100ns$) so that the underlying gate dielectric (SiO_2) is not damaged and there is no dopant

penetration. The ELA technique is also compatible with HfO_2 gate dielectric, and ELA PMOS capacitors have 5-7 orders of magnitude lower leakage current density than the RTA control samples.

After ELA, the effective Work Function (WF) of a p+ poly-Si gate is reduced by $\sim 470meV$. This is probably due to generation of vacancies and interstitials during the quenching process, resulting in Fermi-level pinning. However, the WF can be recovered by furnace annealing at moderate temperature.

In order to incorporate melt-ELA technique into a conventional CMOS process, we have proposed, studied and demonstrated two possible integration schemes (Selective ELA via Masking and ELA Self-Aligned STI process).

In the long term ($< 32nm$ technology nodes), novel transistor structures (e.g. thin-body MOSFETs) will require metallic gate materials with tunable WF to eliminate the gate depletion effect and to adjust transistor threshold voltage. Fully-silicided (FUSI) Nickel Silicide is one of the most promising candidates. In FUSI technology, nickel is deposited onto a poly-Si gate to form a fully silicided film, and the WF is tuned by incorporating impurities (e.g. B, P) by ion implantation before metal deposition.

We show that the WF of impurity incorporated FUSI NiSi gate is adjustable between $\sim 4.5eV$ and $\sim 4.9eV$. By using the pulsed excimer laser as an experimental tool, it is found that both dopant activation and initial dopant distribution are critical in the WF tuning. Gate oxide qualities (leakage current, TZDB, Q_{BD}) of boron doped FUSI gate are found to be better than those of p+ poly-Si reference in pMOSFETs, while those of phosphorus and undoped FUSI gate are comparable with those of n+ poly-Si reference in nMOSFETs. It is also found that doped

FUSI gates have better thermal stability to withstand backend process annealing.

Professor Tsu-Jae King
Dissertation Committee Chair

To my family

Contents

Contents **ii**

List of Figures **v**

List of Tables **xiii**

Acknowledgements **xiv**

1 Introduction **1**

1.1 CMOS Technology Scaling Challenges 1

1.1.1 Ideal Gate Stack 2

1.1.2 Short-term Challenges ($\geq 32nm$ Technology Nodes) 3

1.1.3 Long Term challenges (<32nm Technology Nodes) 6

1.2 Meeting the Challenges 7

1.2.1 Near-term 7

1.2.2 Long-term 7

1.3 Organization 8

1.4 References 9

2 Pulsed Excimer Laser Annealed (ELA) Poly-Si Gate **12**

2.1 Introduction 12

2.2 ELA Process and Simulation 14

2.3 Experiment 17

2.4 Properties of ELA Poly-Si Gate Films 18

2.4.1 Laser Adsorption and Melting of Si film 18

2.4.2 Dopant Distribution and Sheet Resistance 20
2.4.3 Effective Work Function of ELA Poly-Si Gate 22
2.5 Reduction of Gate Depletion 33
2.6 Gate Dielectric Qualities 37
2.6.1 SiO_2 gate dielectric 38
2.6.2 Channel Hole Mobility 39
2.6.3 HfO_2 gate dielectric 40
2.7 Summary 44
2.8 References 47
2.9 Appendix: Process Flow for Gate Last ELA MOSFETs 51

3 CMOS Process Integration of ELA Gate Technology 53

3.1 Introduction 53
3.2 Integration Issues of Melt-ELA in Conventional CMOS Processes . . 54
3.3 Integration with FLA 57
3.4 Selective ELA via Masking 59
3.5 Integration with Cypress SA-STI 65nm Process 64
3.5.1 Introduction 64
3.5.2 Process Flow 67
3.5.3 Gate Pre-Amorphization Implant (PAI) 68
3.5.4 Ablation Induced by Flourine 72
3.5.5 Device Results 75
3.6 Stitching Issue 80
3.7 Summary 81
3.8 References 83

4 Fully Nickel-Silicided (FUSI NiSi) Gate Technology 86

4.1 Introduction 86
4.2 Experiment 89
4.3 Adjustable WF Range of FUSI NiSi 90
4.3.1 Full Silicidation Issues 90
4.3.2 Effective Gate Work Function Extraction 92

4.3.3 Summary . . . 96
4.4 Dopant Pile Up . . . 97
4.4.1 Experiment . . . 98
4.4.2 Effect of Dopant Activation . . . 99
4.4.3 Effect of Initial Dopant Distribution . . . 103
4.4.4 Summary . . . 106
4.5 Gate Dielectric Reliability . . . 107
4.6 Thermal Stability . . . 112
4.7 Summary . . . 113
4.8 References . . . 115

5 Conclusion 120
5.1 Summary . . . 120
5.2 Recommendations for Future Work . . . 123
5.2.1 Integration of ELA with Dummy Gate SA-STI . . . 123
5.2.2 FUSI NiSi Gate . . . 127
5.3 References . . . 128

A MATLAB Simulation Codes for Gate Stack ELA Process 129
A.1 Overview . . . 129
A.2 main.m . . . 130
A.2.1 P_init.m . . . 134
A.2.2 Temperature.m . . . 135

List of Figures

1.1 Cross section of a bulk MOSFET. As the gate length, L_g, is reduced, the parasitic capacitors, such as drain-to-substrate junction capacitor, have stronger control over the body and increases off state leakage current. 3

1.2 The electron energy band diagram of an n-channel MOS (NMOS) capacitor with the channel in inversion condition (Left). Gate depletion is observed due to finite dopant concentration. Capacitance-vs.-Voltage curve of an nMOSFET (Right). Due to gate depletion effect (GDE), the effective gate capacitance is reduced. 5

2.1 ELA system used in this work. The source is either KrF (λ=248nm) or XeCl (λ=305nm) 15

2.2 Thermal simulation of the melting process for an amorphous-Si gate. A schematic cross section of the layered structure is shown at the right (55nm a-Si/2.3nm SiO_2/c-Si substrate). "T" or "B" indicates the temperature profile at the top or bottom of the a-Si layer, respectively. The laser pulse is turned on and off at 0ns and 38ns, respectively. . . 16

2.3 Melt duration of p+ and n+ a-Si films measured by in-situ IR reflectance. Thickness of a-Si is 50nm which is on top of 25$\AA$ of SiO_2 on c-Si substrate. 19

2.4 Measured reflected infrared (IR) intensity of p+ a-Si film on 25$\AA$ SiO_2/n-type c-Si substrate during ELA at different laser fluences. Since the reflectance of molten silicon increases substantially, the peaks show the moment Si gate melts. 300mJ/cm^2 is the melt threshold. Curves are shifted vertically for clear presentation. 19

2.5 SIMS profile of laser-annealed (640mJ/cm^2) poly-Si (55nm) film with 70nm LTO capping layer removed. 20

2.6 Resistivity of laser-annealed 55nm-thick poly-Si films capped with 43 nm LTO as a function of laser pulse energy. The 55nm film was implanted with $1 \times 10^{16} cm^{-2}$ boron at 2 keV. 21

2.7 Measured gate capacitance *vs.* gate voltage (C-V) characteristics for PMOS capacitors with p+ poly-Si gates annealed by ELA (640 mJ/cm^2, 10 shots) or by thermal annealing (10 hr 600^oC+ 5 s @ 850^oC). The gate was implanted with $1 \times 10^{16} cm^{-2}$ Boron at 2 keV. The C-V curve obtained with a quantum-mechanical C-V simulator (QMCV) is also shown. Parameters used for the QMCV curve are T_{ox}=24Å, V_{FB}=0.76V, N_{sub}=$5 \times 10^{17} cm^{-3}$ and N_{poly}=$2 \times 10^{20} cm^{-3}$. . 23

2.8 Work functions for different gate doping processes and annealing conditions. The gate was capped with 43 nm LTO during ELA and thermal annealing. 24

2.9 Fixed charge density for different gate doping processes and annealing conditions. The gate was capped with 43 nm LTO during ELA and thermal annealing. 24

2.10 Extracted V_{FB} values for PMOS and NMOS capacitors (2.4nm SiO_2 dielectric) after different annealing treatments. (The furnace anneals were 1 hour long). As indicated by the arrows, Φ_{Gate} for an ELA p+ poly-Si gate annealed at 400^oC in N_2 is low, and is increased by annealing at 400^oC in FGA. Φ_{Gate} returns to the low value upon further annealing at 400^oC in N_2. The ideal V_{FB} values for p+ and n+ doped poly-Si are also given. 27

2.11 Measured $I_D - V_G$ curves of $50\mu m/2\mu m$ PMOS transistors at $V_D = -20mV$. The inset magnifies the region at $I_D = 10^{-9}A$ at which the V_T is defined. The leftmost curve is that of the transistor after ELA. After that, from left to right, the transistor then went through 300^oC, 450^oC, 475^oC, 500^oC, 525^oC, 575^oC, and 600^oC furnace annealing, respectively. 29

2.12 Subthreshold slope of measured $I_D - V_G$ curves when $I_D = 10^{-9}A$ . . 30

2.13 WF recovery after ELA. "Long FA" doubled the annealing time of "Short FA". "FGA last" samples underwent 30mins 400^oC FGA annealing after FA. "FGA + FA Last" has this annealing sequence: 30 mins 400^oC FA + 30 mins 400^oC FGA + 30 mins 400^oC FA 30

2.14 Logarithmic value of WF recovery time constant *vs* inverse of temperature. The extracted at 550^oC E_A is $0.4eV \pm 0.1eV$. 32

2.15 Effective (electrical) oxide thickness of PMOS capacitor (at $V_G = V_{FB} - 2.16V$) as a function of laser pulse fluence and post-ELA annealing process. "SPC" = 4 hr at 600^oC. "RTA" = 5 s @ 850^oC. The samples were capped with 43 nm-thick LTO during the ELA. 34

2.16 PMOS C-V curves of laser annealed sample followed by different thermal treatments. QMCV is used to fit ELA+SPC+RTA. Fitting parameters: T_{ox}=2.35 nm, $N_{sub} = -1 \times 10^{15} cm^{-3}$, $N_{poly} = 4 \times 10^{20} cm^{-3}$. . 34

2.17 Measured gate capacitance *vs.* gate voltage (C-V) characteristic for a PMOS capacitor with $\sim 2.4nm$ SiO_2 gate dielectric and p+ poly-Si gate annealed by ELA followed by furnace annealing (1 hr @ 550^oC) and FGA. QMCV simulated curves are shown for comparison. The parameters used to generate the QMCV curves are $T_{ox} = 23.5\AA$, V_{FB}= 0.7V, $N_{sub} = 1 \times 10^{15} cm^{-3}$, and N_{poly}=$1 \times 10^{20} cm^{-3}, 3.5 \times 10^{20} cm^{-3}$, or infinity (metal gate) 35

2.18 Measured gate capacitance *vs.* gate voltage (C-V) characteristic for NMOS capacitors with $\sim 2.4nm$ SiO_2 gate dielectric and n+ poly-Si gate annealed by ELA followed by furnace annealing (1 hr @ 550oC) and FGA. QMCV simulated curves are shown for comparison. The parameters used to generate the QMCV curves are $T_{ox} = 24.3\AA$, V_{FB}=-0.85V, $N_{sub} = 1 \times 10^{15} cm^{-3}$, and N_{poly}=$1 \times 10^{20} cm^{-3}, 3.0 \times 10^{20} cm^{-3}$, or infinity (metal gate). 36

2.19 Gate leakage current density as a function of gate voltage. "ELA" indicates that the gate Si film was annealed with 10 laser shots at $640mJ/cm^2$. "RTA" indicates that the gate Si film was thermally annealed 10 hr at 600^oC followed by 5s at 850^oC. The a-Si films were capped with 70 nm LTO during ELA. 38

2.20 Time Zero to Dielectric Breakdown distributions for capacitors with threshold-adjust implant with ELA (10 shots, $640mJ/cm^2$) gates vs. RTA (10 hr at 600^oC + 5s at 850^oC) gates. The a-Si films were capped with 70 nm LTO during ELA. 39

2.21 Extracted hole mobility of PMOS transistor (L=$5\mu m$, W=$50\mu m$) annealed with ELA ($550mJ/cm^2$) and RTA ($800^oC 1min + 1000^oC$ spike) respectively. The universal hole mobility curve is also included as reference. 40

2.22 Measured gate capacitance *vs.* gate voltage (C-V) characteristics for PMOS capacitors with p+ poly-Si gates annealed by ELA (544 mJ/cm^2) followed by solid-phase crystallization (SPC) annealing at 600^oC for 4 hours. QMCV curve is also shown. Parameters used for the QMCV curve are T_{ox}=$23\AA$ and N_{poly}=$4 \times 10^{20} cm^{-3}$. 42

2.23 Measured leakage current density of high-κ PMOS capacitors (15nm HfO_2 with CET=3nm) for various gate-annealing processes. 42

2.24 Measured C-V curves of PMOS capacitors (boron implanted gate) with HfO_2 gate dielectric (initial EOT=2.3nm) after different thermal treatments. "600C" means 600^oC for 1min and "800C" means 800^oC for 10s. All samples received FGA at 400^oC for 20 minutes. . 43

2.25 Distribution of leakage current density at V_G=1V for boron-implanted p+ poly-Si-gated PMOS capacitors (HfO_2 gate dielectric) after different thermal treatments. Initial EOT = 2.3nm. "600C" means 600^oC for 1min and "800C" means 800^oC for 10s 44

2.26 Gate leakage current density vs. EOT for in-situ boron-doped poly-Si/HfO_2 gate stack after ELA. For reference, the leakage current density of SiO_2 is plotted, as well as the roadmap requirements for microprocessor unit (MPU), low operating power (LOP), and low standby power (LSTP) devices at the 45 nm and 65nm CMOS technology nodes [29] 46

3.1 Measured Capacitance-*vs*.-Voltage curves of PMOS capacitor at 100kHz after various post-ELA annealing conditions. All samples received 400°C Froming Gas Annealing (FGA). 55

3.2 Micrograph showing the morphology of patterned poly-Si gate on both active and isolation (LOCOS) areas after ELA ($480mJ/cm^2$). The fluence is not enough to fully melting the a-Si on the active area but is already too high such that the roughness of the gate pad is increased and debris are found. 57

3.3 Measured Capacitance-*v.s.*-voltage curves of ELA sample and ELA followed by FLA annealing. There is a 150mV shift after FLA. 59

3.4 Measured gate leakage densities of ELA + SPC and ELA + FLA samples. It is found that FLA does not increase the leakage current significantly. 60

3.5 Cross-sectional view of the proposed structure after a-Si masking-layer patterning. STI is shallow trench isolation and can represent any other isolation techniques. The stack is 70nm a-Si/43nm LTO/ 55nm gate a-Si/ Gate oxide/ STI or Si substrate. Laser is irradiated from the top. 61

3.6 Reflectance measurements of gate silicon layer ("Masked" = under mask silicon layer, "Not Masked"= exposed to the laser) after ELA at $430mJ/cm^2$. Measured values are relative to an unannealed sample 62

3.7 Measured $I_D - V_G$ curve of $W/L = 50/5$ pMOSFETs fabricated using the masked ELA process. 62

3.8 Cross-sectional SEM of sample illustrated in Figure 3.5 after high-fluence ($650mJ/cm^2$) ELA and Secco etch to delineate the a-Si boundary. The lateral melt distance underneath the edge of the a-Si mask is $\sim 100nm$. 63

3.9 Cross sectional view of conventional STI with associated problems: Gate-wrap-around and gate oxide thinning (lower left inset) and kink effect due to parasitic corner device (right inset) (After [10]). 65

3.10 Essential part of the EXTIGATE process. Top-Gate is usually metallic material overlapping the poly-Si gates and interconnecting transistors within/between active areas (AA) (After [10]). 66

3.11 Cross-sectional TEM picture of active and isolation areas after Ge PAI at 45keV and LTO deposition. 69

3.12 Simulation results using SRIM Monte Carlo simulation showing the Silicon (Si) and Oxygen (O) vacancy densities generated for different implantation conditions. All curves are for Ge implants with dose $= 6.8 \times 10^{14} cm^{-2}$, except NWI is that of N-well Implant and PCHI is that of P-channel implant. Si vacancy density corresponding to amorphization is also given (a-Threshold). 70

3.13 Amorphization depth of poly-Si gate after Ge implantation with dose $= 6.8 \times 10^{14} cm^{-2}$ at different energies. TEM is the measured TEM result. SRIM is the result of Monte Carlo simulation from Figure 3.12. R_p and Delta_R$_p$ are the projected range and straggle of Ge implantation into Si, respectively. 71

3.14 Gate dielectric breakdown voltage as a function of Ge PAI energies. . 72

3.15 Inversion mode gate leakage current as a function of Ge PAI energies. 72

3.16 Accumulation mode gate leakage current as a function of Ge PAI energies. 73

3.17 SEM showing the micro-scale active structure in SA-STI process after ELA at $717 mJ/cm^2$. The left a-Si gate was implanted with B while the right was implanted with BF_2. Ablation is observed with BF_2 implanted structure only. 74

3.18 SEM showing the nano-scale active structure in SA-STI process after ELA at $500 mJ/cm^2$. The structure are either implanted with BF_2 or P. No ablation is observed. The chosen laser fluence is larger than the full-melt threshold. 74

3.19 Cross-sectional TEM picture of active and isolation areas after Ge PAI at 45keV, LTO deposition and ELA at $309 mJ/cm^2$. 76

3.20 Measured $I_D - V_G$ curves ($V_D = 1.2V$) of ELA ($309 mJ/cm^2$) SA-STI nMOSFET ($L_G = 90nm$) with Ge PAI at 35keV and 45keV. Reference is the ordinary Cypress SA-STI wafer with Ge-PAI at 10keV (no ELA) and poly-gate RTA. 76

3.21 Measured $I_D - V_G$ curves ($V_D = -1.2V$) of ELA ($309 mJ/cm^2$) SA-STI pMOSFET ($L_G = 90nm$) with Ge PAI at 35keV and 45keV. Reference is the ordinary Cypress SA-STI wafer with Ge-PAI at 10keV (no ELA) and poly-gate RTA. 77

3.22 Measured $I_D - V_G$ curves of ELA ($309mJ/cm^2$) SA-STI NMOS ($L_G = 90nm$) with Ge PAI at 35keV and 45keV. Reference is the ordinary Cypress SA-STI wafer with Ge-PAI at 10keV (no ELA) but with poly-gate RTA. The ELA curves are shifted by -90meV to match the off current of the Reference. 78

3.23 Measured $I_D - V_G$ curves of ELA ($309mJ/cm^2$) SA-STI PMOS ($L_G = 90nm$) with Ge PAI at 35keV and 45keV. Reference is the ordinary Cypress SA-STI wafer with Ge-PAI at 10keV (no ELA) but with poly-gate RTA. The ELA curves are shifted by 55meV to match the off current of the Reference. 79

3.24 Current improvement of ELA ($309mJ/cm^2$) SA-STI NMOS and PMOS ($L_G = 90nm$) with Ge PAI at 35keV and 45keV over reference in saturation region ($|V_D|$=1.2V). I_{OFF} has been matched. 79

3.25 Schematic diagram showing how two dies can be annealed in each laser shot. Broken lines are the edges of the laser spots. Since scribe lines 3,5,7,9 are within the laser spots, the transistors on these scribe lines can still be used for automatic testing. However, all the vertical scribe lines cannot be used because they are at the edges of the laser spot. . 81

4.1 Schematic diagrams showing the process flow of fabricating FUSI NiSi capacitor structures. When LOCOS is used, step 4 is skipped. 90

4.2 Cross-sectional TEM micrographs (different magnifications) of NiSi gate formed with a 450°C 90 s RTA. Initial poly-Si thickness is 150nm and Ni thickness is 90nm capped with 30nm TiN. Most of the area is not full silicided. In some regions, less than 60nm of NiSi is formed. . 92

4.3 Measured $C - V$ characteristics of NiSi gate capacitors (T_{ox}=2.3nm) with different annealing conditions. The 50nm poly-Si was either undoped (upper) or doped with boron ($2.4 \times 10^{15}cm^{-2}$) (lower). 530°C for 60s is sufficient to fully silicide the poly-Si and therefore, with this annealing condition, undoped FUSI gate has the maximum accumulation capacitance and boron-doped FUSI has both the correct WF and maximum accumulation capacitance. SPC is 800^oC for 4 hours, which was used to activate the dopants in poly-Si before Ni deposition. . . . 93

4.4 Extracted gate work function relative to Si mid-gap, for NiSi gate with different implantation and RTA conditions. The initial poly-Si gate thickness is either 50nm or 100nm. For 100nm poly- Si, even with enough Ni (60nm), the gate is not fully silicided, so that the apparent tunable work function range is unreasonably large. (High=$7 \times 10^{20}cm^{-3}$, Low=$1 \times 10^{20}cm^{-3}$) 94

4.5 Effective gate WF extraction of FUSI NiSi using least squares fitting of V_{FB} against t_{ox} plots. P-type substrate was used. B - Boron, P - Phosphorus, high - $2.4 \times 10^{15} cm^{-3}$, low - $4 \times 10^{14} cm^{-3}$, Insitu - in-situ P doped poly-Si. 95

4.6 The WF and Q_f of FUSI NiSi gate under various impurity implantation conditions. B - Boron, P - Phosphorus, high - $2.4 \times 10^{15} cm^{-3}$, low - $4 \times 10^{14} cm^{-3}$, Insitu - in-situ P doped poly-Si. 95

4.7 Cross-sectional diagram showing the laser annealing process of a-Si gate before patterning. 50nm Low Temperature Oxide (LTO) (450°C) was used to cap the a-Si during annealing as anti-reflection coating. A layer of 150nm a-Si layer on top of LTO was deposited at 450°C and patterned to prevent the ablation of a-Si gate during ELA. 100

4.8 Sheet resistance ($R_\square$) of insitu boron doped silicon film (50nm) after ELA at different fluences. (left axis). $R_\square$ does not change after $600 mJ/cm^2$. This corresponds to the full melt fluence. Thus the melt depth at lower fluences can be calculated by using the corresponding $R_\square$ (right axis). 101

4.9 Measured reflected infrared (IR) density during ELA at different laser fluences. Since the reflectance of molten silicon increases substantially, the peaks show the moment Si gate is melted. $300 mJ/cm^2$ is the melt threshold. Curves are shifted vertically for clearer presentation. . . . 101

4.10 C-V curves of laser annealed in-situ boron doped sample after FUSI (tox=2.2nm). There is no gate depletion, which indicates that the gates are fully silicided. Higher laser fluence translates to deeper melt depth and higher percentage of dopant activation. 102

4.11 Extracted WF of laser annealed in-situ boron doped sample after FUSI versus the percentage of dopant activation. The percentage is calculated by using the melt depth found in Figure 4.8 103

4.12 Sheet resistance (R_S) of boron implanted (1keV, $4 \times 10^{15} cm^{-2}$) silicon film (55nm) after ELA. 104

4.13 Measured C-V characteristics of boron-implanted capacitors after ELA and FGA. 104

4.14 Measured C-V characteristics of boron implanted capacitors after ELA and FUSI. 105

4.15 Measured C-V characteristics of boron-implanted FUSI NiSi gates. The a-Si gate has gone through either no thermal treatment (amorphous), SPC or RTA before Ni deposition. 106

4.16 Measured C-V characteristics phosphorus implanted FUSI NiSi gates. The a-Si gate has gone through either no thermal treatment (amorphous), SPC or RTA before Ni deposition. 107

4.17 Leakage current density of boron doped (p+) FUSI NiSi gate *vs.* p+ poly-Si measured at $V_G = V_{FB} + 1.1V$. The area of the capacitors is $150\mu m^2$. "500C 10min" and "550C 10min" are the measured data of p+ FUSI NiSi after further annealing at 500^oC and 550^oC for 10 mins. 109

4.18 Leakage current density of undoped FUSI NiSi gate *vs.* n+ poly-Si measured at $V_G = V_{FB} - 1.03V$. The area of the capacitors is $150\mu m^2$. "500C 10min" is the measured data of FUSI NiSi after further annealing at 500^oC for 10 mins. 110

4.19 Time Zero to Dielectric Breakdown (TZDB) of boron doped (p+) FUSI NiSi gate *vs.* p+ poly-Si. The area of the capacitors is $50\mu m^2$. "500C 10min" and "550C 10min" are the measured data of p+ FUSI NiSi after further annealing at 500^oC and 550^oC for 10 mins. 110

4.20 Time Zero to Dielectric Breakdown (TZDB) of undoped and phosphorus doped (n+) FUSI NiSi gate *vs.* n+ poly-Si. The area of the capacitors is $50\mu m^2$. "500C 10min" is the measured data of FUSI NiSi after further annealing at 500^oC for 10 mins. 111

4.21 Charge to Breakdown (QBD) of boron doped (p+) FUSI NiSi gate *vs.* p+ poly-Si at $V_G = 4.1V$. The area of the capacitors is $25\mu m^2$. 111

4.22 Charge to Breakdown (QBD) of phosphorus doped (n+) and undoped FUSI NiSi gate *vs.* n+ poly-Si at $V_G = -3.85V$. The area of the capacitors is $25\mu m^2$. 112

5.1 Dummy gate ELA SA-STI process after dummy gate deposition. . . 125

5.2 Dummy gate ELA SA-STI process after STI planarization. 125

5.3 Dummy gate ELA SA-STI process after dummy gate removal. 125

5.4 Dummy gate ELA SA-STI process after gate deposition. 126

5.5 Dummy gate ELA SA-STI process during ELA. 126

A.1 Flowchart of the thermal simulation. 130

List of Tables

1.1 Bulk MOSFET parameters for High Performance (HP), Low Standby Power (LSTP) and Low Operating Power (LOP) [1]. * means that the parameters are for double-gate devices. 4

3.1 Spreadsheet showing that the p-to-n active seperation is larger than the single-sided lateral melting extent but smaller than the two-sided lateral melting extents. 64

3.2 Process Flow of ELA SA-STI. 67

4.1 Table summarizing the tunable WF range (in eV) of FUSI NiSi gate and the fabrication conditions. Second row of each WF entry shows the implant dose (cm^{-2}) and energy (keV). Thickness is in nm. Temperature is in ^oC. 97

Acknowledgements

At the end of the beginning, the thesis is completed. If one has not achieved what were contemplated at the beginning of the beginning, one should have at least realized that the completion of any process, significant or not, is the result of the tangible and intangible helps from numerous people and organizations. With this grateful heart, the main chapters of the life should be more successful and beautiful.

Thanks to those who have fought for the freedom and liberal environment in Berkeley, where I am always proud to be a part of.

Thanks to Prof. Tsu-Jae King, my thesis advisor, for her guidance, as well as the financial supports.

Thanks to Prof. Vivek Subramanian, Prof. Peter Y. Yu and Prof. Borivoje Nikolic for serving in my qualifying examination committee and for their advices on my research.

Thanks to Dr. Michael Ameen and Dr. Aditya Agarwal at Axcelis Technologies for arranging the high dose and ultra-low energy ion implantations. Without their help, both the ELA and FUSI experiments will be more difficult, if not impossible.

Thanks to Mr. Takuro Matsutoya for his friendship and advice. Especially, he has completed most of the gate-last transistor fabrication process.

Thanks to Dr. Qiang Lu. He has given me very valuable advices on my research, graduate life and job searching.

Thanks to Mr. Hideki Takeuchi. Without his help and comment in the Microlab, I would not have been able to learn and appreciate the right attitude of being an engineer. He has taught me many process, equipment and device characterization skills. He has also grown some of the HfO_2 films used in this study.

Thanks to Mr. Joseph Donnelly for installing the Ni target and maintaining the

sputtering machine, which is very critical to the FUSI NiSi experiment. Also to the other Microlab staffs who have usually work beyond their duties to help us complete our experiment in time.

Thanks to Dr. Eric J. Carleton, who was the superuser of Lextra, for maintaining, trouble-shooting the excimer laser equipment and giving me advices on laser processing.

Thanks to Mr. Kyoungsub Shin, my first project partner in Berkeley, for his friendship and advice. He also conducted the research in FUSI NiSi gate together with me at the early phase of the project.

Thanks to Dr. Wade Xiong in ADTF for providing the FUSI NiSi SIMS analysis.

Thanks to Dr. Patrick Smith of Kovio, Inc. for his training of using the laser system in Kovio.

Thanks to Dr. Joel W. Ager, Mr. Peng Zhang, and Mr. Wei Wang for the Raman scattering measurement and discussion in LNBL.

Thanks to Mr. Alvaro Padilla for helping deposit some of the HfO_2 films in this study.

While I was an intern in Cypress Semiconductor, I was so lucky to have Dr. Alain Blosse and Dr. Krishnaswamy Ramkumar as my mentors. They have given me freedom to conduct my research and help me complete the ELA SA-STI process.

I would like to thank the Sir Edward Youde Memorial Fund in Hong Kong for supporting my first three years of study and living in UC Berkeley, and also Semiconductor Research Corporation for providing the research fund for FUSI NiSi and Cypress Semiconductor for partially supporting the research in Excimer Laser Annealing.

I would like to say thank you to my parents for their unselfish love and never-

ending supports. Also to Miss Ai Ping Lee for her requesting of 10 additional pages of thesis draft every week!

The last but not the least, thanks to my friends! In the past five years, they have been the faithful readers of my essays and have listened to my complaints. Also to the friends and colleagues in Cory 373 (especially Donovan Lee, Mohan Dunga and Joanna Lai), for bringing me so much happiness and excitement, and letting me believe once again that Device Group is the best group in the world to join! Go Bears!

Chapter 1

Introduction

1.1 CMOS Technology Scaling Challenges

Scaling of Metal-Oxide-Semiconductor Field Effect Transistors (MOSFETs) has enabled the unparalleled advancement of the semiconductor industry. The main theme of scaling has been the reduction of the transistor size, through the introduction of more advanced lithography technologies [1]. As a result, better performing transistors and denser integrated circuits can be realized. In the course of scaling, two criteria must be met; firstly, the transistor on/off current ratio (I_{on}/I_{off}) has to be practically high for logic operations; secondly, the transistor leakage current (I_{leak}) has to be reasonably low due to standby power consideration [2]. In order to meet these two criteria, it is necessary to scale the gate stack so that the gate can retain dominant control over the conduction channel, while at the same time they are resistively decoupled. However, as the industry faces the 65nm technology node and beyond, burdens of gate stack scaling surface.

1.1.1 Ideal Gate Stack

The gate stack of a MOSFET consists of the gate electrode and gate dielectric. Together with the semiconductor channel, a capacitor is formed, with the dielectric as an insulator (Figure 1.1). The ideal property of the gate electrode is to be a perfect conductor with infinite charge carrier concentration, while that of the gate dielectric is to be a perfect insulator with infinite dielectric constant (κ). This ideal combination will result in maximum gate capacitance and, thus, absolute control of the channel charge by the gate voltage to turn the transistor ON and OFF.

Moreover, the gate electrode is expected to have appropriate work function (WF), on which the threshold voltage (V_{TH}) of the transistor depends. V_{TH} of a bulk MOSFET is given by [3],

$$V_{TH} = V_{FB} + 2\Phi_S + \frac{Q_B}{C_{gate}} \quad (1.1)$$

$$V_{FB} = \frac{1}{q}(\Phi_{Gate} - \Phi_S) - \frac{Q_f}{C_{gate}} \quad (1.2)$$

$$C_{gate} = \kappa\frac{\epsilon_o}{t_{ox}} \quad (1.3)$$

where Φ_{Gate} and Φ_S are the WF's of the gate and the semiconductor body, respectively, Q_f and Q_B are the areal oxide fixed charge and areal bulk depletion charge densities, respectively, q is the elementary charge, κ and t_{ox} are the dielectric constant and the thickness of the gate dielectric, respectively, and ϵ_o is the vacuum permittivity. Therefore, for an ideal gate stack, V_{TH} depends only on WF's of the gate and the semiconductor body.

1.1.2 Short-term Challenges ($\geq 32nm$ Technology Nodes)

In reality, ideal gate stack does not exist. Gate capacitance is finite and there are competitions between the gate capacitor and other parasitic capacitors (e.g. drain-to-body capacitor) over the control of the channel (Figure 1.1). When the gate length is reduced, off state leakage current increases and I_{on}/I_{off} reduces due to the stronger interference of the parasitic capacitors.

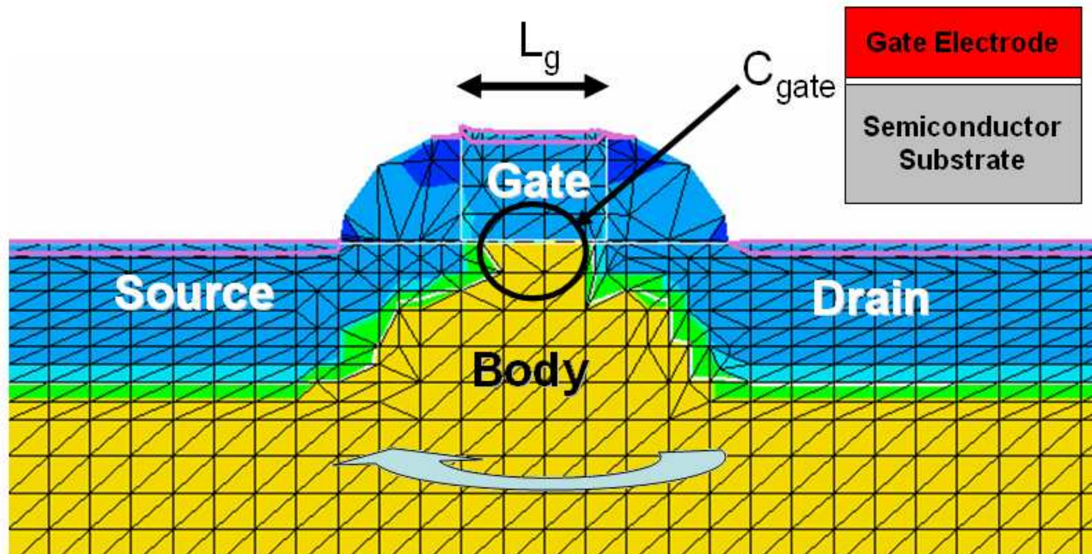

Figure 1.1. Cross section of a bulk MOSFET. As the gate length, L_g, is reduced, the parasitic capacitors, such as drain-to-substrate junction capacitor, have stronger control over the body and increases off state leakage current.

Traditionally, doped poly-crystalline silicon (poly-Si) and silicon dioxide (SiO_2, $\kappa = 3.9$) are used as the standard gate electrode and dielectric materials respectively. When the transistor is scaled, the thickness of SiO_2 has to be reduced in order to increase the gate capacitance. However, in the 65nm technology node and beyond, the physical thicknesses of the SiO_2 become very small (Table 1.1) and quantum tunneling of carriers from the electrode to the channel is so significant that the dielectric can no longer serve as an insulator. This is because while the capacitance is

inversely proportional to the dielectric thickness (t_{ox}), the tunneling current increases exponentially as t_{ox} is reduced [4].

	Year	2007	2010	2013
	Technology Node (nm)	65	45	32
HP	Physical Gate Length (nm)	25	18	13
	Equivalent Oxide Thickness $\AA$	11	6.5	6*
	Gate Leakage(A/cm^2)	8.00E+02	1.56E+03	8.46E+02*
LSTP	Physical Gate Length	45	28	20
	Equivalent Oxide Thickness $\AA$	19	14	12
	Gate Leakage (A/cm^2)	2.20E-02	3.60E-02	1.10E-01
LOP	Physical Gate Length	32	22	16
	Equivalent Oxide Thickness $\AA$	12	9	8*
	Gate Leakage(A/cm^2)	7.80E+01	1.10E+02	2.20E+02*

Table 1.1. Bulk MOSFET parameters for High Performance (HP), Low Standby Power (LSTP) and Low Operating Power (LOP) [1]. * means that the parameters are for double-gate devices.

Therefore, closer-to-ideal (larger κ) dielectric materials have been studied to replace SiO_2 so that the reduction of t_{ox} can be slowed down (Equation 1.3). Materials having much larger κ, such HfO_2 and ZrO_2, are promising [5]. However, their thermal stability is of concern due to oxygen vacancy formation or dielectric film crystallization [6; 7].

At the same time, poly-Si can only be doped to have finite carrier concentration. A depletion layer is formed in the region adjacent to the gate dielectric when the transistor is ON (Figure 1.2), which increases the effective oxide thickness. The gate depletion effect thus makes L_g scaling more difficult. In order to reduce gate depletion, higher thermal budget is required in gate dopant activation step. However, too high a thermal budget usually results in serious dopant penetration which has adverse impact on the transistor reliability [8; 9].

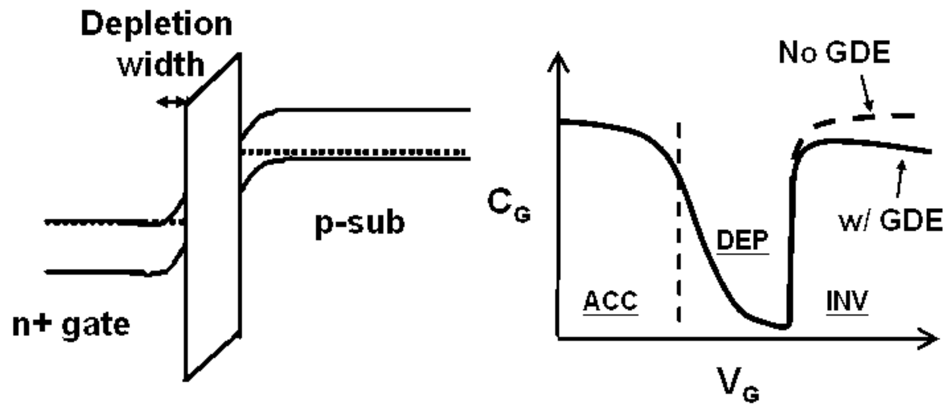

Figure 1.2. The electron energy band diagram of an n-channel MOS (NMOS) capacitor with the channel in inversion condition (Left). Gate depletion is observed due to finite dopant concentration. Capacitance-*vs.*-Voltage curve of an nMOSFET (Right). Due to gate depletion effect (GDE), the effective gate capacitance is reduced.

Despite the aforementioned drawbacks, poly-Si has many advantages over other gate materials. Its chemical and physical properties are well-known and well-studied. It is compatible with high temperature processing, which is required in a gate-first fabrication process. Its WF can be tuned easily for nMOSFET and pMOSFET via doping (n-type vs. p-type). Therefore, poly-Si is expected to remain as the preferred gate electrode material in the near future [1], and reducing the gate depletion effect in poly-Si is a major near-term transistor scaling challenge.

In view of this, the 2005 International Technology Roadmap for Semiconductors (ITRS '05) proposed the following short term challenges:

1. Control of boron penetration from doped poly-Si gate electrodes while minimizing depletion of dual-doped poly-Si electrodes

2. Introduction and process integration of high-κ gate stack materials and processes for high-performance, low operating and low standby power MOSFETs

3. Achievement of activated dopant concentration greater than solid solubility in dual-doped poly-Si gate electrodes

1.1.3 Long Term challenges (<32nm Technology Nodes)

In the long term, aggressive reduction of gate dielectric thickness with reduced or even eliminated poly-Si gate depletion effect will not be sufficient in meeting the scaling challenges [1]. Novel device structures, such as the Ultra-Thin-Body (UTB) MOSFET [10] or FinFET [11], may be required. Both UTB MOSFETs and FinFETs improve the scalability by reducing the parasitic capacitance or/and enhancing the gate capacitance.

Such novel devices require new device design criteria, among which the requirement of gate electrode with close-to-midgap tunable WF is one of most challenging ones [12]. This is because the channels will be lightly doped in order to reduce impurity scattering of carriers and dopant fluctuations [13]. Thus the adjustment of threshold voltage may require adjustability of the gate WF. Poly-Si as a gate electrode material has to be abandoned because its WF can only be adjusted in a binary fashion, so that alternative gate electrode materials have to be developed. Any candidate gate material should possess the following properties. Firstly, it should have very high carrier concentration to eliminate the gate depletion effect. Secondly, it should be integratable with the traditional CMOS processes. Thirdly, its WF should ideally be tunable for CMOSFET application. Finding metal gate electrodes with appropriate work function is thus stated in the ITRS'05 as one of the major long term scaling challenges.

1.2 Meeting the Challenges

1.2.1 Near-term

Poly-Si dopant activation, boron penetration issue, and high-κ dielectric degradation are the results of atomic rearrangement under high temperature thermal processing. Therefore, in order to meet the near-term challenges, a novel gate annealing technique has to be introduced. This technique should have the ability to speed up the desired atomic movements (dopant diffusion and activation in gate electrode) while preventing the undesirable ones (dopant penetration and high-κ crystallization). Melt excimer laser annealing (ELA) is a very promising candidate. An excimer laser can deliver highly concentrated (both spatially and temporally) laser pulses. It is expected to be able to melt the Si gate electrode without damaging the underlying gate dielectric due to its short duration of heating ($\sim 0.1\mu s$) [14; 15]. As a result, the dopants can diffuse quickly towards the gate electrode/ dielectric interface in liquid silicon (l-Si) [16] without penetrating the solid dielectric within the limited thermal process time. Upon crystallization, the dopants will be incorporated into substitutional lattice sites with activated concentration larger than solid solubility being possible [17].

1.2.2 Long-term

To meet the long-term challenge, metallic gate materials are needed because they provide the highest possible carrier concentration ($> 10^{22} cm^{-3}$). The gate depletion effect can thus be avoided and gate sheet resistance can also be reduced. There are various metal gate electrode materials proposed in the literature. Among them, Fully Silicided (FUSI) Nickel Silicide (NiSi) gate is the most promising one [18]. In a FUSI NiSi process, a traditional gate-first CMOS process is used. Ni is deposited after

source/drain dopant activation to fully silicide the poly-Si at low temperature ($\sim 450^oC - 550^oC$) for 1 to 10 mins. Therefore, FUSI NiSi poses minimum modification to a conventional CMOS process. During silicidation, dopants segregate to the gate electrode/ dielectric interface, and the effective work function can thus be adjusted by controlling the implanted dopant dose and types with the poly-Si precursor gate material. Moreover, NiSi has low resistivity ($\sim 15 - 20\mu\Omega \cdot cm$) [19] making FUSI NiSi a very suitable candidate for future high speed transistors.

1.3 Organization

This thesis is divided into five sections. After this introductory chapter, Chapter 2 and 3 discuss the use of pulsed Excimer Laser Annealing (ELA) to reduce the poly-Si gate depletion effect. Physical and electrical properties of ELA poly-Si gate electrodes are discussed in Chapter 2. In Chapter 3, the integration of ELA with conventional CMOS process is presented. This includes the study of the compatibility between ELA and flashing lamp annealing (FLA) used in source/drain activation, proposal and demonstration of selective ELA via masking scheme and fabrication of ELA transistor in a 65nm Self-Aligned Shallow-Trench Isolation (SA-STI) process. In Chapter 4, NiSi FUSI process is discussed with emphasis on the effective WF adjustment, dopant pile-up mechanisms, thermal stability and reliability of NiSi FUSI process. After the conclusion in Chapter 5, the Appendix documents the MATLAB codes used in ELA thermal simulation.

References

[1] *The International Technology Roadmap for Semiconductors*, Semiconductor Industry Association, 2005.

[2] A. Chandrakasan, W. J. Bowhill, and F. Fox, *Semiconductor material and device characterization.* IEEE Press, 2001.

[3] R. F. Pierret, *Semiconductor Device Fundamentals.* Addison-Wesley, 1996.

[4] D. K. Schroder, *Semiconductor material and device characterization.* NY, USA: John Wiley & Sons, 1998.

[5] G. D. Wilk, R. M. Wallace, and J. M. Anthony, "High-κ gate dielectrics: Current status and materials properties considerations," *Journal of Applied Physics*, vol. 89, pp. 5243–5275, 2001.

[6] H. Takeuchi, H. Y. Wong, D. Ha, and T.-J. King, "Impact of oxygen vacancies on high-k gate stack engineering," in *International Electron Devices Meeting. Technical Digest*, Dec. 2004, pp. 829–834.

[7] H. Kim, P. C. McIntyre, and K. C. Saraswat, "Effects of crystallization on the electrical properties of ultrathin HfO_2 dielectrics grown by atomic layer deposition," *Applied Physics Letters*, vol. 82, pp. 106–108, 2003.

[8] M. Cao, P. Vande Voorde, M. Cox, and W. Greene, "Boron diffusion and penetration in ultrathin oxide with poly-Si gate," *IEEE Electron Device Letters*, vol. 19, pp. 291–3, Aug. 1998.

[9] K. Krisch, M. Green, F. Baumann, D. Brasen, L. Feldman, and L. Manchanda, "Thickness dependence of boron penetration through O_2 and N_2O-grown gate oxides and its impact on threshold voltage variation," *IEEE Transactions on Electron Devices*, vol. 43, pp. 982–990, June 1996.

[10] B. Yu, Y.-J. Tung, S. Tang, E. Hui, T.-J. King, and C. Hu, "Ultra-Thin-Body Silicon-On-Insulator MOSFET's for Terabit-Scale Integration," in *International Semiconductor Device Research Symposium*, Dec. 1997, pp. 623–626.

[11] X. Huang, W.-C. Lee, C. Kuo, D. Hisamoto, L. Chang, J. Kedzierski, E. Anderson, H. Takeuchi, Y.-K. Choi, K. Asano, V. Subramanian, T.-J. King, J. Bokor, and C. Hu, "Sub 50-nm FinFET: PMOS," in *International Electron Devices Meeting. Technical Digest*, Dec. 2002, pp. 363–366.

[12] L. Chang, S. Tang, T.-J. King, J. Bokor, and C. Hu, "Gate length scaling and threshold voltage control of double-gate MOSFETs," in *International Electron Devices Meeting. Technical Digest*, Dec. 2000, pp. 37.2.1–37.2.4.

[13] L. Xia and X. Can, "Random dopant induced threshold voltage fluctuations in double gate MOSFETs," in *International Conference on Solid-State and Integrated Circuits Technology Proceedings*, 2005, pp. 138–41.

[14] T. Sameshima, M. Hara, and S. Usui, "Measuring the temperature of a quartz substrate during and after the pulsed laser-induced crystallization of a-Si:H," *Japanese Journal of Applied Physics*, vol. 28, pp. 2131–2133, 1989.

[15] J. S. Im, H. J. Kim, and M. O. Thompson, "Phase transformation mechanisms involved in excimer laser crystallization of amorphous silicon films," *Applied Physics Letters*, vol. 63, pp. 1969–1971, 1993.

[16] H. Kodera, "Diffusion coefficients of impurities in silicon melt," *Japanese Journal of Applied Physics*, vol. 2, pp. 212–219, 1963.

[17] Y. Takamura, S. H. Jain, P. B. Griffin, and J. D. Plummer, "Thermal stability of dopants in laser annealed silicon," *Journal of Applied Physics*, vol. 92, pp. 230–234, 2002.

[18] J. Kedzierski, D. Boyd, C. J. Cabral, P. Ronsheim, S. Zafar, P. Kozlowski, J. Ott, and M. Ieong, "Threshold voltage control in NiSi-gated MOSFETs through SIIS," *IEEE Transactions on Electron Devices*, vol. 52, pp. 39–46, Jan. 2005.

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

Chapter 2

Pulsed Excimer Laser Annealed (ELA) Poly-Si Gate

2.1 Introduction

Dual-doped p+/n+ polycrystalline-silicon (poly-Si) gate technology will continue to be used in CMOS technologies employing the conventional bulk-Si Metal-Oxide-Semiconductor Field-Effect Transistor (MOSFET) structure in the near term ($\geq 32nm$ technology nodes), and may even play an important role in non-planar devices towards the end of the CMOS technology roadmap [1]. However, there are many technological challenges to be overcome. According to the International Technology Roadmap for Semiconductor (ITRS'05), near-term challenges for CMOS technology advancement include: 1. Control of boron penetration from doped polysilicon (poly-Si) gate electrodes while minimizing depletion of dual-doped poly-Si electrodes, 2. Introduction and process integration of high-κ gate stack materials and processes for high-performance, low operating and low standby power MOSFETs, and 3. Achieve-

ment of activated dopant concentration greater than solid solubility in dual-doped poly-Si gate electrodes.

The first challenge by itself is a trade-off in terms of process optimization. This is because dopants are usually introduced into the poly-Si gate through ion implantation at reasonably low energy in order to minimize the damage to the gate dielectric. This results in the aggregation of inactive dopants in the top portion of the poly-Si. Thermal annealing is thus required to drive the dopants towards the gate electrode/dielectric interface and activate them in order to reduce the gate depletion effect. Although higher thermal budget gives higher dopant concentration and activation at the interface, boron penetration into the gate dielectric is also more serious. This is well known and studied for traditional rapid thermal annealing (RTA) process [2; 3; 4]. If, furthermore, high-κ gate stack materials such as hafnium dioxide (HfO_2) are to be integrated, the thermal budget has to be minimized. Otherwise, the dielectric quality and reliability will be severely degraded due to oxygen vacancy generation [5] or crystallization [6] of the high-κ material. Therefore, it is important to explore new annealing techniques that can maximize the ratio of the dopant diffusion coefficient in the gate electrode to that in the gate dielectric while being compatible with high-κ gate-dielectric materials.

Among the various annealing techniques, pulsed excimer laser annealing (ELA) is one of the most promising candidates to meet these challenges. Excimer laser is capable of delivering ultra-short (few 10's of nano-seconds (ns)) high fluence ($\sim 1Jcm^{-2}$) laser pulses. It has been used to form large-grained poly-Si for thin-film transistor applications [7; 8; 9]. Pulsed laser annealing has also been used to form ultra-shallow source/drain junctions in MOSFETs as follows [10; 11]: Firstly, the crystalline silicon (c-Si) is amorphized by ion implantation (of Si or Ge) to the desired junction depth. Afterwards, the dopants are implanted into the amorphous Si (a-Si). A laser pulse is then applied to momentarily melt the a-Si, but not the underlying/surrounding c-Si.

The dopants diffuse rapidly within the molten Si before it re-crystallizes, typically within $\sim 0.1\mu s$ [8], resulting in a super-saturated concentration of active dopants. It has been shown that $1.5 \times 10^{21} cm^{-3}$ electrically active boron can be attained through this process [12]. Therefore, ELA is a very attractive candidate to meet all the aforementioned near-term ITRS gate stack challenges if proper integration procedure is taken.

In this chapter, the ELA process and its thermal simulation results are first presented. Then poly-Si film properties after ELA, including the resistivity and effective work function, are discussed. Finally, the reduction of gate depletion effect and the preservation of gate dielectric qualities after ELA are shown.

2.2 ELA Process and Simulation

A typical gate stack cross-section is shown in Figure 2.2 (right). During ELA, a short (few 10's of ns) laser pulse is applied to the a-Si gate electrode. Excimer lasers are dimer lasers and typically have photon energies much larger than the bandgap energy of Si (248 nm KrF laser is $\sim 5eV$, 308nm XeCl laser is $\sim 4eV$), resulting in high absorption coefficients (KrF laser is $1.8 \times 10^8 m^{-1}$, XeCl laser is $1.5 \times 10^8 m^{-1}$) of the silicon [13] and, thus, most of the laser energy is absorbed within the first 20 nm of the a-Si; the absorbed energy is then transferred by thermal conduction to the rest of the film. For sufficiently high laser fluences, the film melts completely. Due to the extremely high dopant diffusivity in liquid Si (l-Si) ($\sim 10^{-4} cm^2 s^{-1}$ [14]), dopants redistribute evenly throughout the depth of the gate film and are readily incorporated into substitutional lattice sites upon solidification of the Si film, resulting in a very high level of dopant activation at the gate electrode/dielectric interface. To maximize the process window, a-Si is chosen as the gate material because it has a much lower

(by $\sim 200 \pm 50^{o}C$) melting point than c-S [15]. Therefore, the c-Si substrate can remain intact while the whole a-Si gate film is melted.

Figure 2.1 shows the equipment setup for generating spatially uniform ELA pulses from a Gaussian pulse excimer laser source. The pulses are directed onto the semiconductor wafer surface.

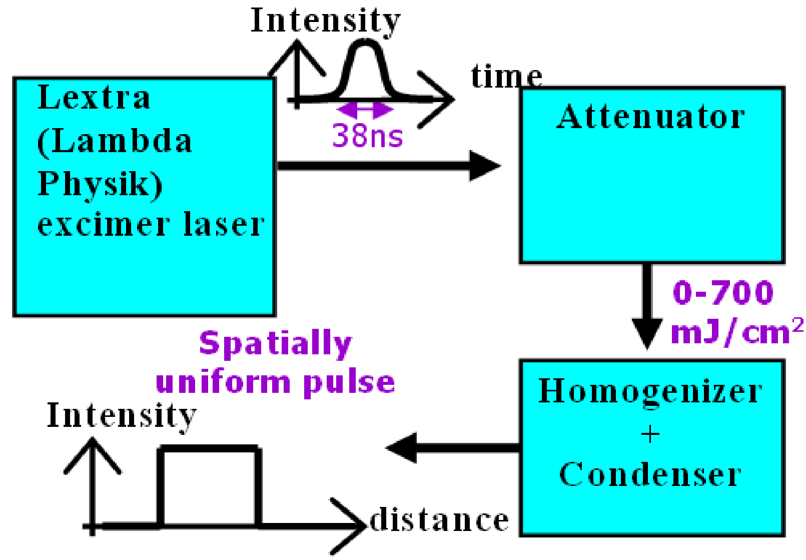

Figure 2.1. ELA system used in this work. The source is either KrF (λ=248nm) or XeCl (λ=305nm)

In order to understand the thermal process of ELA on the a-Si gate stack, the process was simulated by using finite difference method in MATLAB codes (Appendix A). For the sake of accuracy, the temperature dependences of laser reflectivity, thermal conductivity, and specific heat capacity were taken into account using the data given in [16]. Latent heat of a-Si during melting and re-crystallization were also modeled and found to have significant effect on the melting duration of the silicon film.

The simulation results (Figure 2.2) show that after the laser pulse is turned on at 0ns, the temperature at the surface of the Si gate rises quickly to the melting

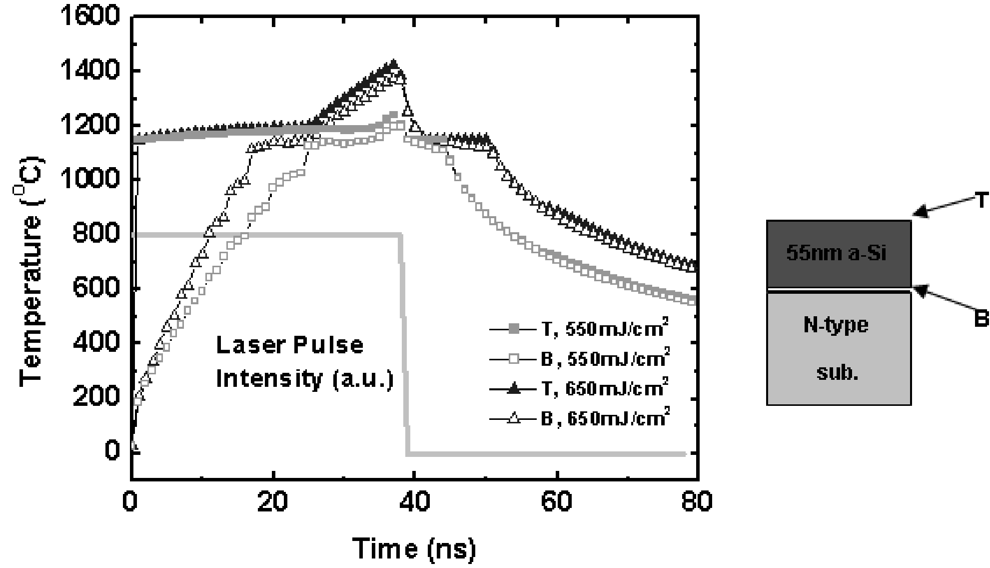

Figure 2.2. Thermal simulation of the melting process for an amorphous-Si gate. A schematic cross section of the layered structure is shown at the right (55nm a-Si/2.3nm SiO_2/c-Si substrate). "T" or "B" indicates the temperature profile at the top or bottom of the a-Si layer, respectively. The laser pulse is turned on and off at 0ns and 38ns, respectively.

point of a-Si (1150^oC) because of the high absorption coefficient of UV light in Si ($\sim 2 \times 10^6 cm^{-1}$) and the low thermal conductivity of a-Si. The bottom of the Si gate heats up more slowly and cools down sooner (after the laser pulse is turned off at 38 ns) than the top of the Si gate, because heat is conducted downward to the highly thermally conductive c-Si substrate. Thus, the melt duration at the bottom of the Si gate is shorter than at the top. A $550mJ/cm^2$ pulse is sufficient to fully melt the a-Si film. If the laser fluence is increased by 18% to $650mJ/cm^2$, the peak temperature at the bottom of the film reaches the melting point of c-Si (1410^0C). Therefore, there is about ±8% process window if $600mJ/cm^2$ is chosen, which should be good enough in semiconductor fabrication to avoid melting of the underlying Si substrate or under-melting of the a-Si film.

The simulations shows that ELA process of a-Si gate stack is a sub-μs process

with melting duration of $\sim 50ns$. This is confirmed by the results of in-situ Infra-Red reflectance measurement during the ELA of a similar structure (Figure 2.4).

2.3 Experiment

In order to study the electrical properties of ELA Si films and the gate depletion effect of ELA capacitors, CMOS capacitors were fabricated. Since for photons with large enough energies ($> 3.5eV$), the absorption coefficient of undoped Si and Si doped with different dopants are similar [17], detailed studies were carried out with boron doped PMOS capacitors because of the more stringent requirement for gate annealing due to the boron penetration phenomenon. 2.3 nm-thick gate oxide (by optical measurement) was thermally grown on 4" n-type wafers followed by the deposition of 55 nm of amorphous silicon (a-Si) by low-pressure chemical vapor deposition (LPCVD) at 550^oC. This gate layer was implanted with $1 \times 10^{16} cm^{-2}$ B at 2 keV. The wafers were then capped with either 70 nm or 43 nm of LPCVD SiO_2 deposited at 450^oC (Low Temperature Oxide, LTO) to prevent oxidation and contamination during laser processing. A pulsed KrF laser ($\sim 38ns$, $\lambda = 248nm$) was used to melt and crystallize the gate material and activate the implanted dopants. Since silicon dioxide has lower refractive index ($n = 1.45$) than silicon ($n = 3.5$), the 43 nm-thick LTO serves as an anti-reflection coating [18]. Laser fluences ranging from 380 mJ/cm^2 to 700 mJ/cm^2 were studied in this work. The LTO was then removed by HF dip, and the gate layer was patterned to form the gate electrodes.

To reduce the fixed charge caused by ion implantation, some other NMOS and PMOS capacitors were also fabricated on 4" p-type and n-type wafers, respectively. The process was the same as described above except that besides thermal SiO_2 ($\sim$2.4nm thick), RTCVD HfO_2 was also used as gate dielectric. The a-Si thickness was reduced to 45 nm and deposited by LPCVD at 530^oC. The a-Si was then

implanted with either boron ($4 \times 10^{15} cm^{-2}$ at 1 keV) or phosphorus ($4 \times 10^{15} cm^{-2}$ at 2 keV). Low doses of boron or phosphorus were then implanted to form "source/drain" regions (to provide the inversion-layer charge), and the devices were annealed in forming gas for 15 min at $450^{o}C$ or 20 min at $400^{o}C$.

2.4 Properties of ELA Poly-Si Gate Films

2.4.1 Laser Adsorption and Melting of Si film

The absorption coefficient of excimer laser is very large in Si, becoming almost constant when the photon energy is larger than 4eV ($\alpha = 1.5 - 2 \times 10^{8} m^{-1}$) [13]. Therefore, most of the laser energy is absorbed in the top several nm of the gate, which melts before the underlying layer rises in temperature (Figure 2.2). The higher the laser energy is, the deeper is the melt depth and the longer is the melt duration. When the surface melts, the reflectance increases dramatically due to the formation of l-Si. Therefore, the melt duration can be measured by monitoring the reflectance of infra-red (IR) light of the sample. Figure 2.3 shows the melting duration measured by in-situ IR reflectance on in-situ doped p+ and n+ a-Si during ELA by XeCl laser and Figure 2.4 shows the changes of IR reflectance during ELA processes. The samples are blanket (unpatterned) wafers with a structure consisting of either 50nm n+ a-Si/25$\AA$ SiO_2/p-type c-Si substrate or 50nm p+ a-Si/25$\AA$ SiO_2/n-type c-Si substrate. It can be seen that n+ and p+ a-Si have different melting thresholds. This may be due to the dopant effect, resulting in different IR reflectance on the a-Si. However, once the energy is large enough to melt the surface, the reflectance becomes very similar as the surface becomes l-Si within 1ns (Figure 2.2). Therefore, they have similar full melt threshold between $500 mJ/cm^2$ and $600 mJ/cm^2$. The full melt threshold can be

confirmed by the fact that there is a sharp increase in melt duration from $500mJ/cm^2$ to $600mJ/cm^2$ [8].

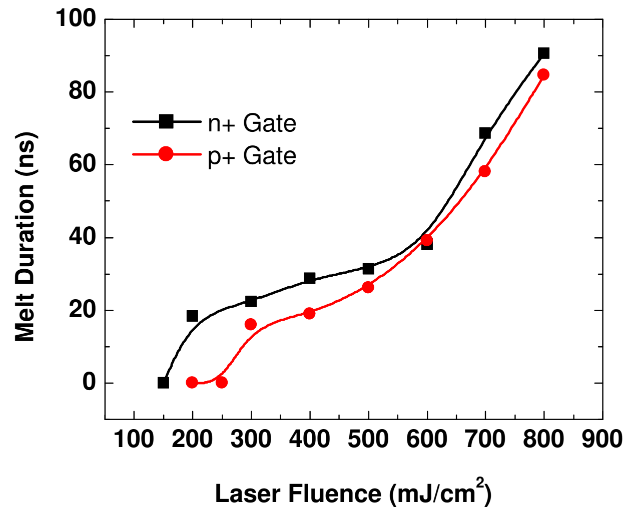

Figure 2.3. Melt duration of p+ and n+ a-Si films measured by in-situ IR reflectance. Thickness of a-Si is 50nm which is on top of $25\AA$ of SiO_2 on c-Si substrate.

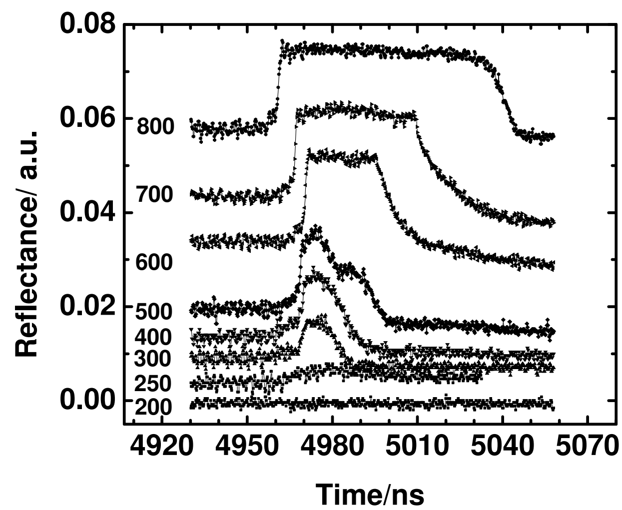

Figure 2.4. Measured reflected infrared (IR) intensity of p+ a-Si film on $25\AA$ SiO_2/n-type c-Si substrate during ELA at different laser fluences. Since the reflectance of molten silicon increases substantially, the peaks show the moment Si gate melts. $300mJ/cm^2$ is the melt threshold. Curves are shifted vertically for clear presentation.

2.4.2 Dopant Distribution and Sheet Resistance

Figure 2.5 shows the SIMS profile of a boron implanted gate film ($1 \times 10^{16} cm^{-2}$, 2 keV) after ELA. Although the melt duration is only about $50ns$, due to the high diffusivity (D_{Boron}) of boron in l-Si, the boron atoms distribute evenly throughout the film at a concentration of $\sim 2 \times 10^{21} cm^{-3}$. By taking $D_{Boron} = 2.4 \times 10^{-4} cm^2 s^{-1}$ [14], the $e^{-\frac{1}{2}}$ distance, $\sqrt{2D_{Boron}t}$ is $\sim 49nm$ for $t = 50ns$. Considering that the projected range of 2keV boron is $> 5nm$, the SIMS result agrees with the theoretical values. This shows that ELA is a very effective method for bringing impurities to the gate electrode/dielectric interface, which is necessary to reduce the gate depletion effect.

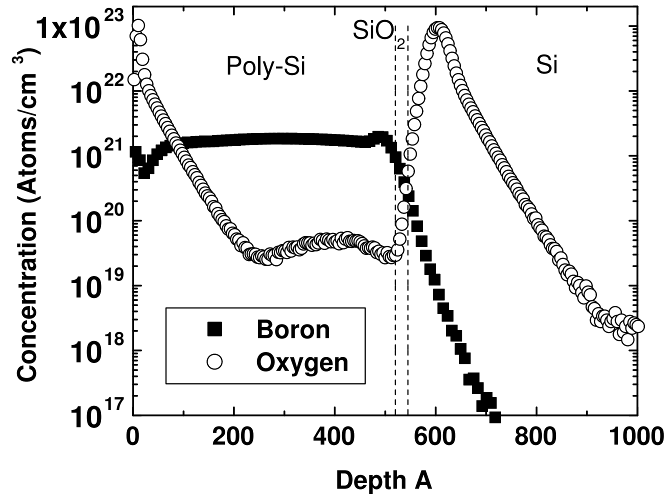

Figure 2.5. SIMS profile of laser-annealed ($640mJ/cm^2$) poly-Si (55nm) film with 70nm LTO capping layer removed.

Figure 2.6 shows the dependence of gate resistivity on laser fluence, for films capped with 43 nm LTO. When the laser fluence is below 550 mJ/cm^2, the resistivity decreases with increasing fluence, reaching $0.58m\Omega \cdot cm$ at $528mJ/cm^2$. When the fluence is greater than $550mJ/cm^2$, the resistivity increases by approximately one order of magnitude. This corresponds to the regime in which the film is completely

melted [8; 9] and supercooling causes spontaneous nucleation throughout the film, resulting in very small grains (less than a few hundred $\mathring{A}$) [19]. The higher density of grain boundaries results in a lower concentration of electrically active boron atoms. Indeed, according to [19], when the film is thin enough, overmelting can result in a-Si formation instead of poly-Si after solidification. Therefore, in order to reduce the gate depletion effect and achieve low resistivity, it is necessary to use a laser fluence just below the full-melt threshold. In this regime, also known as the "super lateral growth" (SLG) regime, the film is nearly completely melted with sparsely isolated small solid Si seeds remaining. The molten Si then crystallizes from these few isolated seeds, resulting in very large grains ($\sim 1\mu m$) [9].

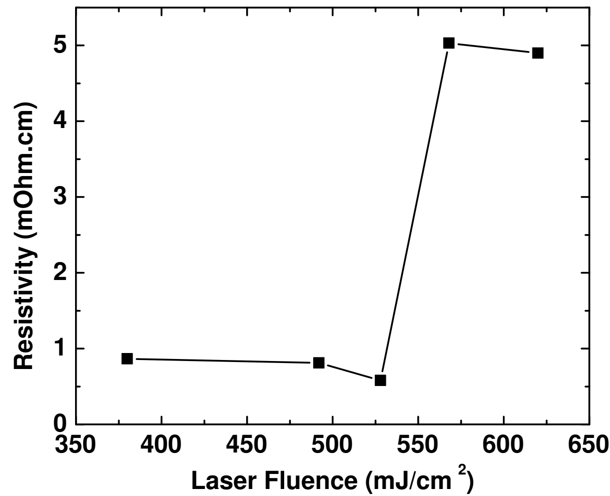

Figure 2.6. Resistivity of laser-annealed 55nm-thick poly-Si films capped with 43 nm LTO as a function of laser pulse energy. The 55nm film was implanted with $1 \times 10^{16} cm^{-2}$ boron at 2 keV.

2.4.3 Effective Work Function of ELA Poly-Si Gate

Work Function Shift and Recovery

Figure 2.7 shows the capacitance-*vs.*-voltage (C-V) characteristics of the fabricated PMOS capacitors which received a threshold-adjustment implant ($4.5 \times 10^{12} cm^{-2}$ As at 80 keV) before the growth of thermal oxide. These capacitors underwent the processes described in Section 2.3 except for the gate dopant activation step. The C-V curve of the laser-annealed gate is clearly shifted by approximately -0.4V as compared with that of the rapid-thermal-annealed (RTA) gate. Since the substrate is doped ($\sim 5 \times 10^{17} cm^{-3}$) and there is no significant change in the depletion capacitance (C_{min}), the shift is not due to boron penetration in the RTA device (shifting right). Indeed, a quantum mechanical C-V simulator [20] was used to match the C-V curve of the laser-annealed sample, and the flat-band voltage (V_{FB}) was found to be 0.76V, which is 0.3V lower than expected for a capacitor with a heavily doped p-type poly-Si gate and an n-type substrate with $5 \times 10^{17} cm^{-3}$ doping (expected $V_{FB} \simeq 1.06V$). This difference can be due to the contribution of oxide fixed charge (Q_f) and/or a change in gate work function (Φ_{Gate}) (Equation 1.2). However, for a 23$\AA$-thick gate oxide, an areal density of $3.8 \times 10^{12} cm^{-2}$ additional positive fixed charge would be necessary to completely account for the -0.4V shift in the C-V curve.

In order to determine the oxide Q_f and Φ_{Gate} for laser-annealed gate electrodes, some wafers were thermally oxidized and selectively partially etched back to achieve multiple gate-oxide thicknesses (195$\AA$, 126$\AA$, 87$\AA$, and 33$\AA$) on the same wafer. PMOS capacitors were then fabricated as described in Section 2.3, except that the gate layer (55nm) was implanted with B at 2 keV of different doses ($1 \times 10^{16} cm^{-2}$, $4 \times 10^{15} cm^{-2}$, and $1 \times 10^{15} cm^{-2}$) and a fixed laser fluence of 588 mJ/cm^2 was used, sufficient to completely melt the a-Si film.

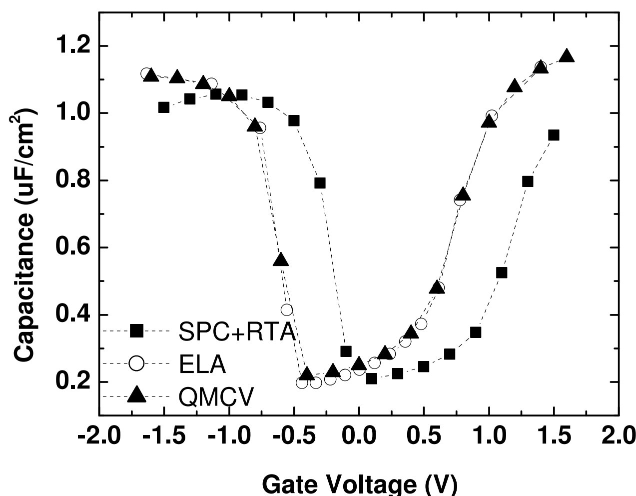

Figure 2.7. Measured gate capacitance *vs.* gate voltage (C-V) characteristics for PMOS capacitors with p+ poly-Si gates annealed by ELA (640 mJ/cm^2, 10 shots) or by thermal annealing (10 hr 600^oC+ 5 s @ 850^oC). The gate was implanted with $1 \times 10^{16} cm^{-2}$ Boron at 2 keV. The C-V curve obtained with a quantum-mechanical C-V simulator (QMCV) is also shown. Parameters used for the QMCV curve are T_{ox}=24$\mathring{A}$, V_{FB}=0.76V, N_{sub}=$5 \times 10^{17} cm^{-3}$ and N_{poly}=$2 \times 10^{20} cm^{-3}$.

High frequency (100kHz) capacitance-*vs*-voltage curves (CV) were taken for different oxide thicknesses and the V_{FB} was determined [21]. Equation 1.2 can be further expanded into the following equation:

$$V_{FB} \quad = \frac{1}{q}(\Phi_{Gate} - \Phi_{sub}) - \frac{Q_f}{\epsilon_{ox}} t_{ox} \qquad (2.1)$$

where V_{FB} is the flat-band voltage of the capacitor, q is the unit charge ($1.6 \times 10^{-19} C$), Φ_{sub} is the WF of silicon substrate in eV, Q_f is the fixed charge density in C/cm^{-2}, t_{ox} is the physical oxide thickness in cm and ϵ_{ox} is the SiO_2 dielectric permittivity ($3.45 \times 10^{-13} Fcm^{-1}$). Since the gate dielectrics were grown at the same time on the same wafer, Q_f can be assumed to be constant. By plotting V_{FB} against t_{ox}, after least square fitting, once can determine the Φ_{Gate} and Q_f from the y-intercept and slope of the fitting line respectively.

Figure 2.8 and 2.9 compare the extracted work function and oxide fixed charge

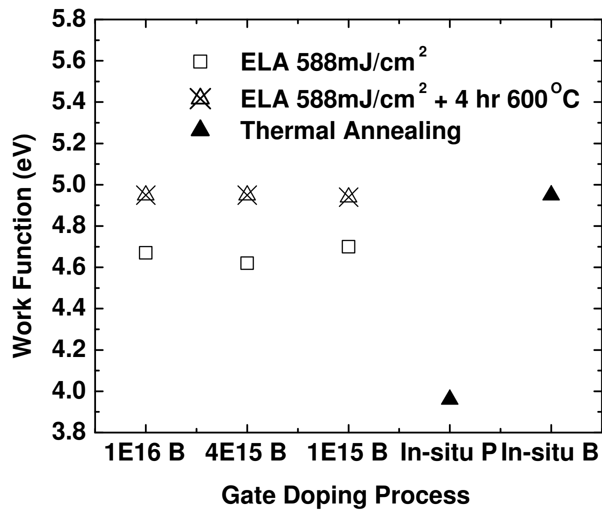

Figure 2.8. Work functions for different gate doping processes and annealing conditions. The gate was capped with 43 nm LTO during ELA and thermal annealing.

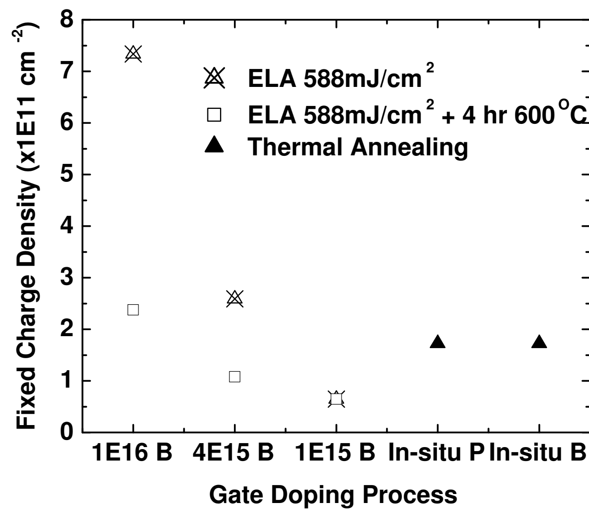

Figure 2.9. Fixed charge density for different gate doping processes and annealing conditions. The gate was capped with 43 nm LTO during ELA and thermal annealing.

density, respectively, for a laser-annealed implanted poly-Si gate, a thermally annealed (10 hours 600^oC + 5 s 820^oC) $in-situ$-doped p+ poly-Si gate, and a thermally annealed (950^oC for 30 mins) $in-situ$-doped n+ poly-Si gate. The work function

of the laser-annealed p+ gate is shifted by $-0.25eV$ to $-0.33eV$ from that of the $in-situ$-doped p+ gate. The data also shows that a higher gate implant dose results in a higher Q_f, but that laser annealing itself does not induce a significant amount of fixed charge. Indeed, with implantation dose of $1 \times 10^{15} cm^{-2}$, the Q_f after ELA is lower than that of $in-situ$ doped sample. The laser-annealed devices were subjected to an additional Solid Phase Crystallization (SPC) anneal at 600^oC for 4 hours in a nitrogen ambient. As shown in Figure 2.8, the gate work function returns to the normal value after SPC, regardless of the gate implant dose. SPC also reduces Q_f to be below $2.3 \times 10^{11} cm^{-2}$ in heavily implanted samples and is comparable to that in-situ-doped devices. This is not surprising because thermal annealing is known to be effective for reducing Q_f [22].

It should be noted that the thermally annealed sample in Figure 2.7 should have lower Q_f than the laser-annealed sample as it was annealed for 10 hr at 600^oC prior to the 5 s RTA at 850^oC. From Figure 2.9, Q_f of the RTA sample in Figure 2.7 can be estimated to be lower by $5 \times 10^{11} cm^{-2}$ than the ELA sample, which corresponds to $\sim 0.06V$ right shift in V_{FB}. Therefore, together with the shift of the effective WF, the differences of the CV curves in Figure 2.7 can be explained.

Origin of Work Function Shift

There are three possible causes of the change in effective Φ_{Gate} after ELA. Firstly, there may be bandgap narrowing (BGN) due to degenerate doping. However, from the fitting using QMCV simulator, the effective activated dopant concentration does not exceed $4 \times 10^{20} cm^{-3}$. According to [23], activate dopant concentration changes from $1 \times 10^{20} cm^{-3}$ (a typical RTA value) to $4 \times 10^{20} cm^{-3}$ (best ELA value achieved in the experiments) can only account for 25meV reduction of silicon bandgap, which is not sufficient to account for the change in effective Φ_{Gate} after ELA.

The second possibility is due to strain. On one hand, as shown in the SIMS analysis (Figure 2.5), the concentration of boron at the gate-oxide interface is $\sim 5\%$ (atomic). This can induce tensile stress in the gate film. However, it is also shown that the WF shift does not depend on a wide range of dopant doses (from $1 \times 10^{15} cm^{-2}$ to $1 \times 10^{16} cm^{-2}$ in Figure 2.9), therefore, high dopant concentration cannot be the cause. On the other hand, it is also possible that as the molten Si gate film quenches rapidly from the melt (a non-equilibrium process), the crystallized Si film may be highly strained, resulting in band-gap narrowing. And the strain can be relieved after thermal annealing. The strain, if there is any, is expected to be tensile because molten Si has lower density than crystalline silicon and upon rapid quenching, tensile stress is formed in the film due to the larger-than-equilibrium atomic distance.

The third possibility is that the atoms at the gate-dielectric interface do not have enough time during the solidification process to arrange themselves. As a result, there is a high degree of disorder at the interface giving rise to a high density of trap states (extrinsic energy states) within the band-gap of Si. This can also be referred as Disorder Induced Gap States (DIGS) [24]. These states are donor-like near the valence band and acceptor-like near the conduction band. Assuming they have comparable densities, the charge neutrality level (CNL) is also located within the bandgap of the silicon. Then the Fermi level of the poly-Si gate will be pinned to the CNL and the effective gate work function is reduced [25]. Thermal annealing can reduce the interface-trap density and, thus, recover the effective WF.

To further investigate the cause, NMOS and PMOS capacitors which received ELA and subsequently a furnace anneal in N_2 and/ or a $400^o C$ forming gas anneal were fabricated and characterized. The V_{FB} of the capacitors were then extracted using QMCV fitting [20]. Figure 2.10 shows that the V_{FB} shift for a p+ gate after ELA without any furnace annealing is about $-470meV$, but for a n+ gate it is negligible. This indicates that the change in effective Φ_{Gate} is not due to strain, because the

conduction-band edge should exhibit a larger offset than the valence-band edge due to strain in the gate electrode [26]. Thus, Fermi-level pinning is likely the reason why an ELA poly-Si gate electrode shows a shift in effective work function. Indeed, it is found that a forming gas anneal helps to restore V_{FB} towards the expected value (by $> 200meV$). The double-headed arrow in Figure 2.10 indicates that V_{FB} of the ELA+FGA p+ poly-Si sample is reduced by $\sim 50meV$ with annealing in N_2 ambient at 400^oC. If this sample subsequently undergoes another FGA, the V_{FB} increases by $\sim 50meV$ again. This further confirms that there are some interface states that can be passivated by hydrogen after ELA. Moreover, since V_{FB} for the NMOS devices does not change much with thermal annealing, it can be deduced that the charge neutrality level (CNL) is close to the conduction-band edge.

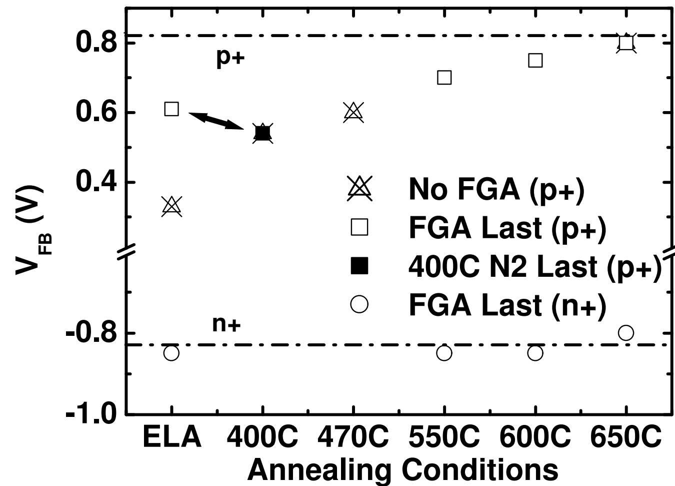

Figure 2.10. Extracted V_{FB} values for PMOS and NMOS capacitors ($2.4nm$ SiO_2 dielectric) after different annealing treatments. (The furnace anneals were 1 hour long). As indicated by the arrows, Φ_{Gate} for an ELA p+ poly-Si gate annealed at 400^oC in N_2 is low, and is increased by annealing at 400^oC in FGA. Φ_{Gate} returns to the low value upon further annealing at 400^oC in N_2. The ideal V_{FB} values for p+ and n+ doped poly-Si are also given.

Activation Energy of WF recovery

If the effective WF shift of ELA poly-Si is due to Fermi-level pinning caused by disordered states, there must be some activation energies associated with the WF recovery phenomenon. PMOS transistors were therefore fabricated and their threshold voltage (V_T) was extracted and compared after subsequent FA (with or without FGA) in order to extract the activation energies. Transistors allow more accurate extraction of WF shift than CV curves, especially when they are not treated with FGA.

A gate-last process was used to fabricate the transistors. After LOCOS and gate dielectric (SiO_2) formation, $Si_{0.4}Ge_{0.6}$ was deposited and patterned as a dummy gate. Source/Drain (S/D) implantation was carried out and annealed using RTA. The $Si_{0.4}Ge_{0.6}$ and SiO_2 were then removed by SC1 solution at 70^oC (5 H_2O:1 H_2O_2:1 NH_4OH) and HF, respectively. After that, $26\AA$ of thermal oxide was grown and in-situ B doped a-Si was deposited, annealed with ELA and patterned. The gate last process allows the S/D dopants to be activated without affecting the a-Si gate. The activated S/D regions have low enough resistance for current measurement and so that there is no need of metallization which would have otherwise limited the maximum FA temperature the transistors can receive. The detailed gate last process is outlined in Section 2.9.

Right after ELA, the I_D $vs.$ V_G curve of a $W/L = 50\mu m/2\mu m$ PMOS transistor was measured at low drain voltage (V_D=-20mV). The transistor then underwent subsequent annealing from 90^oC to 600^oC for different annealing time and the I_D $vs.$ V_G was measured between each annealing. At each temperature, transistor was first annealed for a short duration (15mins or 30mins). After measurement, longer annealing time was used at the same temperature. Higher temperature was only used when no significant V_T shift occurs at the same temperature for longer time. Figure 2.11 shows

some of those I_D $vs.$ V_G curves of $W/L = 50\mu m/2\mu m$ PMOS transistors and they are well-behaved. It is clear that the curves shift in parallel as the ELA transistors undergo different conditions of FA. The threshold voltage (V_T) is defined as where $I_D = 1nA$. Figure 2.12 shows the measured sub-threshold slopes of the curves given in Figure 2.11, taken near $V_g = V_T$. There is no big change of the sub-threshold slope with FA condition and thus the measurement of the shift of V_T is indicative of the shift of the gate WF.

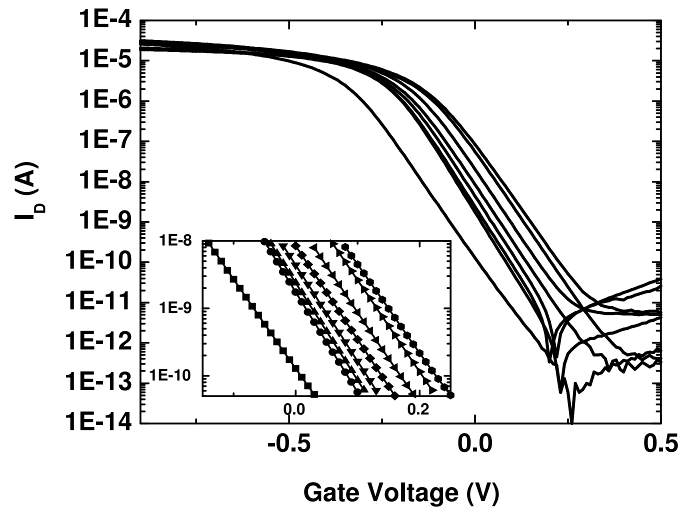

Figure 2.11. Measured $I_D - V_G$ curves of $50\mu m/2\mu m$ PMOS transistors at $V_D = -20mV$. The inset magnifies the region at $I_D = 10^{-9}A$ at which the V_T is defined. The leftmost curve is that of the transistor after ELA. After that, from left to right, the transistor then went through 300°C, 450°C, 475°C, 500°C, 525°C, 575°C, and 600°C furnace annealing, respectively.

The shift (recovery) of the V_T of the transistor after different thermal treatments are shown in Figure 2.13. The data for 300K (room temperature) correspond to ELA only, and serve as the reference. The differences between the "Short FA" and "Long FA" are small as compared with the change in WF with anneal temperature. This indicates that the recovery of the WF is very fast with a very small time constant.

To derive the activation energies, we assume the WF shift can be modeled by

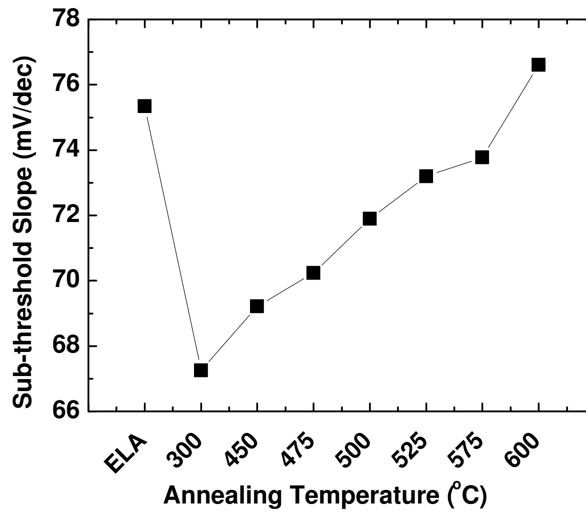

Figure 2.12. Subthreshold slope of measured $I_D - V_G$ curves when $I_D = 10^{-9}A$

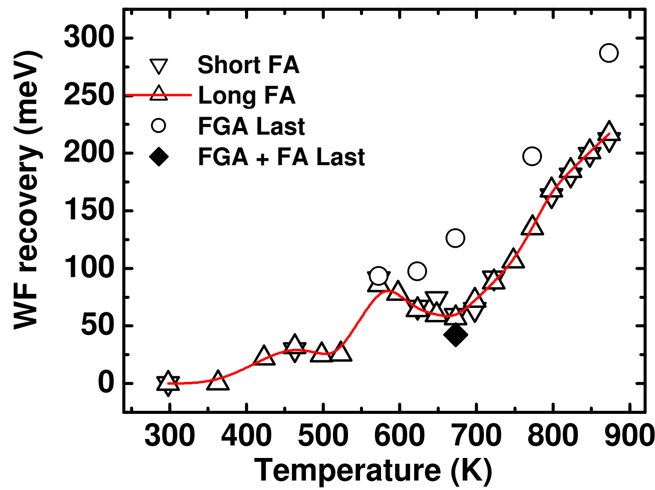

Figure 2.13. WF recovery after ELA. "Long FA" doubled the annealing time of "Short FA". "FGA last" samples underwent 30mins 400^oC FGA annealing after FA. "FGA + FA Last" has this annealing sequence: 30 mins 400^oC FA + 30 mins 400^oC FGA + 30 mins 400^oC FA

the formation of dipole moments (μ_d) with a density of N_{def0} at the gate electrode/ dielectric interface after the rapid quenching process. Therefore, the initial shift of

the WF right after ELA is,

$$\Delta WF = q\mu_d N_{def0} \tag{2.2}$$

For the dipoles to be annealed out, there is a certain transformation barrier to overcome, associated with a time constant $\tau = \tau(T)$ which is temperature dependent and has the following relationships,

$$\frac{dN_{def}}{dt} = -\frac{N_{def}}{\tau} \tag{2.3}$$

$$N_{def} = N_{def0}\ \exp(-\frac{t}{\tau}) \tag{2.4}$$

where $1/\tau$ is the rate of transformation and has the following property,

$$\tau \propto \exp(\frac{E_A}{kT}) \tag{2.5}$$

or,

$$\ln(\tau) = \frac{E_A}{k}\frac{1}{T} + Constant \tag{2.6}$$

When the ELA samples are annealed by FA, dipoles are annealed out and the WF is recovered (δWF). Using Equation 2.4 and 2.2, we have,

$$\delta WF = q\mu_d\,(N_{def} - N_{def0}) \tag{2.7}$$

$$= q\mu_d\ N_{def_0}\left(1 - \exp(-\frac{t}{\tau})\right) \tag{2.8}$$

δWF can be measured by the shift of V_T. Therefore, $\tau(T)$ can be found by annealing the sample at the same temperature for different durations and fitting with Equation 2.8. Then the activation energies can be found by using Equation 2.6.

Figure 2.13 summarizes the recovery of the gate WF after subsequent annealing at different conditions for different durations. It is clear that the WF has been recovered

to the saturated value in a very short time and there is no big difference between the long time and short time annealing. It is also surprising that the WF starts recovering at 150^oC. Moreover, there are at least four activation energies indicating four types of dipoles. One can be annealed before 190^oC (E_{A1}). One can be annealed at 300^oC (E_{A2}) and one is around 600^oC (E_{A3}). The last one actually gives opposite polarity dipoles (increases the WF) and is annealed in the temperature range between 400^oC and 500^oC (E_{A4}).

Since only the region from 525^oC to 575^oC shows differences in WF recovery for different annealing durations, it is only possible to extract the activation energy in that region. The shift is calculated by using to the minimum point due to E_{A2} as the reference because the dipoles associated with E_{A1} and E_{A2} can be assumed to be completely annealed out already. Figure 2.14 shows the Arrhenius plot for extracting E_{A3} and it is found to be $0.4 \pm 0.1eV$, which is very small as expected.

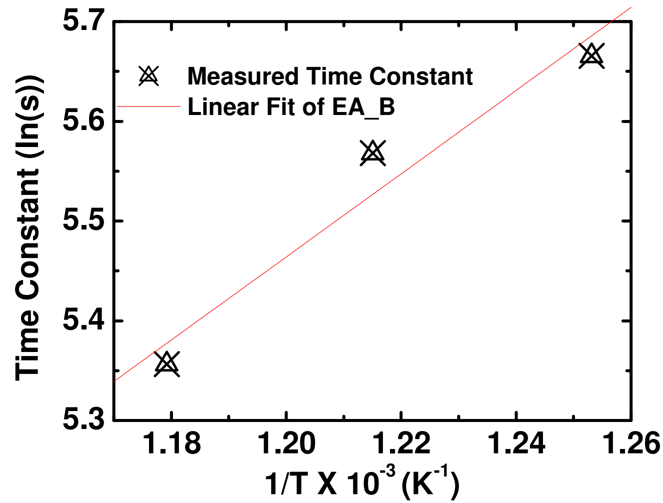

Figure 2.14. Logarithmic value of WF recovery time constant vs inverse of temperature. The extracted at 550°C E_A is $0.4eV \pm 0.1eV$.

2.5 Reduction of Gate Depletion

PMOS capacitors which have received $1 \times 10^{16} cm^{-2}$ B gate implant but not a V_T-adjust implant were used to investigate the reduction in gate depletion by ELA at various fluences. Figure 2.15 shows the effective (electrical) oxide thickness in inversion (at $V_G = V_{FB} - 2.16V$), $T_{ox,inv}$, *vs.* laser fluence. For low laser fluences, the gate film is not melted completely and hence there are few boron atoms at the gate oxide interface. As a result, there is significant gate depletion and $T_{ox,inv}$ is much greater than the physical oxide thickness. Complete melting of the gate film occurs at higher laser fluences, so that the concentration of boron at the gate oxide interface is high and the gate depletion effect is minimized. However, for very high fluences ($> 550 mJ/cm^2$), the average grain size in the poly-Si gate is small (for reasons described in Section 2.4.2) so that there is significant dopant segregation to grain boundaries, hence lower active boron concentration and increased gate depletion effect. The optimal laser fluence is about $\sim 550 mJ/cm^2$, sufficient to nearly completely melt the gate film. This is consistent with the resistivity data in Figure 2.6.

Additional thermal annealing, which is required to recovery the WF of ELA samples (Figure 2.8), was found to reduce $T_{ox,inv}$ further, as shown in Figure 2.15. This is attributed to boron diffusion to the gate-oxide interface for low laser fluences, and to secondary grain growth resulting in reduced dopant segregation for high laser fluences. The lowest effective oxide thickness achieved is $\sim 28\AA$. Considering that the optically measured oxide thickness is $23\AA$, this means that gate depletion and the inversion-layer charge centroid together contribute $\sim 5\AA$ to $T_{ox,inv}$. Figure 2.16 compares the C-V characteristics of PMOS capacitors which received a 540 mJ/cm^2 laser pulse. Based on the best fit for quantum C-V simulated curve, the electrically active dopant concentration in the gate is $\sim 4 \times 10^{20} cm^{-3}$, which is much higher than conventional RTA process [3].

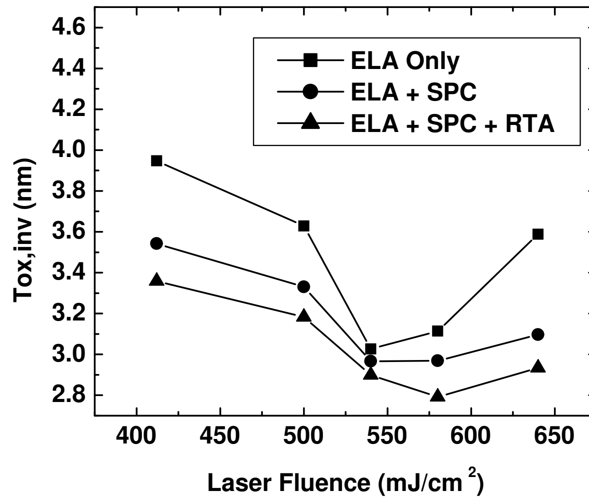

Figure 2.15. Effective (electrical) oxide thickness of PMOS capacitor (at $V_G = V_{FB} - 2.16V$) as a function of laser pulse fluence and post-ELA annealing process. "SPC" = 4 hr at 600^oC. "RTA" = 5 s @ 850^oC. The samples were capped with 43 nm-thick LTO during the ELA.

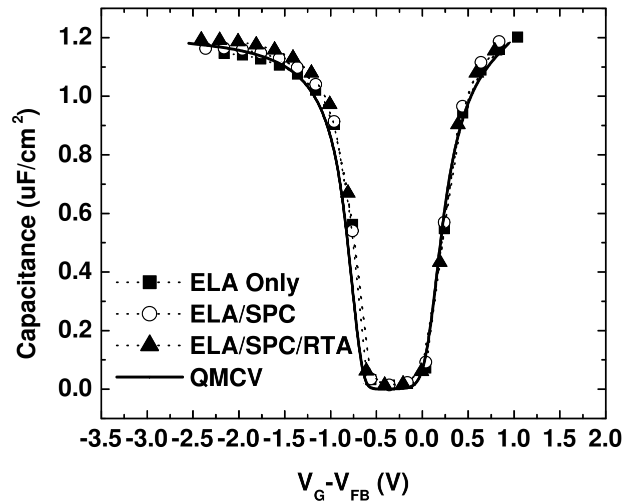

Figure 2.16. PMOS C-V curves of laser annealed sample followed by different thermal treatments. QMCV is used to fit ELA+SPC+RTA. Fitting parameters: T_{ox}=2.35 nm, $N_{sub} = -1 \times 10^{15} cm^{-3}$, $N_{poly} = 4 \times 10^{20} cm^{-3}$.

Since high implantation dose and energy increase the oxide fixed charge (Figure 2.9), low implantation dose at lower energy have been tried but with thinner a-Si (45nm) so as to maintain high enough dopant concentration for both NMOS and PMOS fabricated using the process described in Section 2.3, except that 46nm of LTO and 24$\AA$ SiO_2 was used and the PMOS was implanted with $4 \times 10^{15} cm^{-2}$ boron at 1 keV and NMOS with $4 \times 10^{15} cm^{-2}$ phosphorus at 2keV.

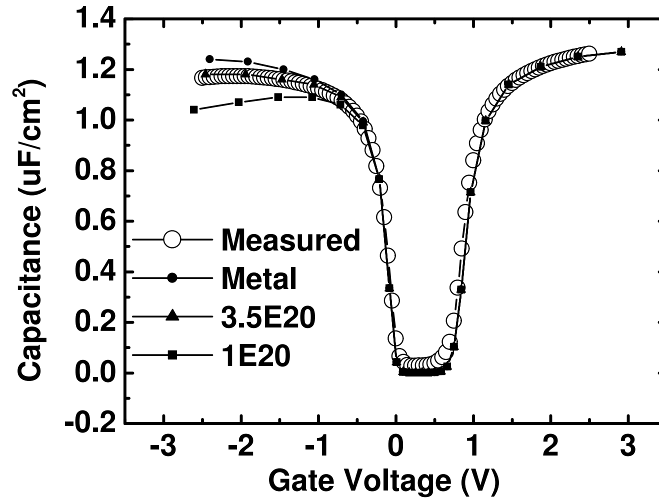

Figure 2.17. Measured gate capacitance *vs.* gate voltage (C-V) characteristic for a PMOS capacitor with $\sim 2.4nm$ SiO_2 gate dielectric and p+ poly-Si gate annealed by ELA followed by furnace annealing (1 hr @ 550^oC) and FGA. QMCV simulated curves are shown for comparison. The parameters used to generate the QMCV curves are $T_{ox} = 23.5\AA$, V_{FB}= 0.7V, $N_{sub} = 1 \times 10^{15} cm^{-3}$, and N_{poly}=$1 \times 10^{20} cm^{-3}, 3.5 \times 10^{20} cm^{-3}$, or infinity (metal gate)

Figures 2.17 and 2.18 show the measured capacitance-*vs.*-voltage (C-V) curves for PMOS and NMOS capacitors, respectively, with SiO_2 gate dielectric. The gate electrode was annealed by ELA followed by a furnace anneal (1 hour at 550^oC) and FGA. A quantum-mechanical C-V simulator (QMCV) was used to fit the measured data. The results indicate that the active dopant concentration in the p+ poly-Si gate near to the gate-dielectric interface is $3.5 \times 10^{20} cm^{-3}$, while it is $3 \times 10^{20} cm^{-3}$ in the n+ poly-Si gate. The gate sheet resistances of p+ and n+ gate are $114\Omega/\square$

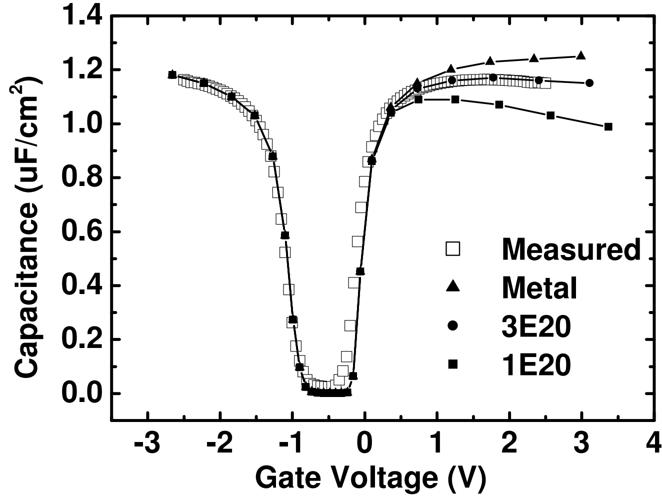

Figure 2.18. Measured gate capacitance *vs.* gate voltage (C-V) characteristic for NMOS capacitors with $\sim 2.4nm$ SiO_2 gate dielectric and n+ poly-Si gate annealed by ELA followed by furnace annealing (1 hr @ 550oC) and FGA. QMCV simulated curves are shown for comparison. The parameters used to generate the QMCV curves are $T_{ox} = 24.3\AA$, V_{FB}=-0.85V, $N_{sub} = 1 \times 10^{15} cm^{-3}$, and N_{poly}=$1 \times 10^{20} cm^{-3}, 3.0 \times 10^{20} cm^{-3}$, or infinity (metal gate).

and $158\Omega/\Box$, which translate to $\sim 0.5m\Omega \cdot cm$ and $\sim 0.7m\Omega \cdot cm$ respectively. These activation levels are much higher than those obtainable by rapid thermal annealing (RTA) without significant boron penetration ($\sim 5 \times 10^{19} cm^{-3}$) [3]. Figures 2.17 and 2.18 also show the QMCV-simulated curves for a poly-Si gate with $1 \times 10^{20} cm^{-3}$ active dopant concentration (representative of RTA poly-Si gate), as well as for a metal gate, as comparison. For the PMOS device at $V_G = V_{FB} - 2V$, the capacitance equivalent thickness (CET) achieved with ELA is $0.17nm$ thinner than that can be achieved with RTA, and only 0.1nm thicker than that can be achieved with a metal gate. For the NMOS device, at $V_G = V_{FB}$+2V, the CET achieved with ELA is 0.19nm thinner than that achieved with RTA, and only 0.1nm thicker than that achieved with a metal gate. The $\sim 0.2nm$ CET reduction provided by ELA is significant for sub-65nm CMOS technologies which will employ gate dielectrics with sub-1nm equivalent oxide thickness (EOT).

2.6 Gate Dielectric Qualities

Since high energy photons ($> 5eV$) and high temperatures ($\sim 1200^oC$) are involved in melt-ELA process, there is a concern that the gate dielectric qualities might be degraded. Although most of the photon are absorbed at the first 10nm of the a-Si film due to the high absorption coefficient and, indeed, for a film of 55nm and $\alpha = 2 \times 10^6 cm^{-1}$, only 0.001% of the photons can reach the dielectric, there are still $10^{16} photons/cm^2$ being able to reach the dielectric. In the worst case, these photons can create more than 10^{16} traps and damages per cm^2 if they are inelastically absorbed as in silicon. Fortunately, most of the dielectrics have bandgap larger than 5eV (SiO_2 has 9eV and HfO_2 has 5.7eV), so most of the photons pass through the dielectric without begin absorbed. Moreover, as the dielectric usually has much higher melting point than the crystalline silicon (e.g. SiO_2 has 1713^oC) and the thermal process is only in the order of 100ns, the dielectric is expected to be intact under the high temperature process. Indeed, according to [27], the time to breakdown for silicon oxynitride gate dielectric with a laser annealed poly-Si gate is better than that with a thermally annealed poly-Si gate.

In this section, experiments were conducted to confirm the aforementioned expectation that the qualities of SiO_2 should be preserved after ELA treatment. Experiments were also carried out with HfO_2 gate dielectric, which is used to indirectly confirm that the quality of dielectric is preserved due to the lack of movement of the dielectric atoms during the ultra-short process.

2.6.1 SiO_2 gate dielectric

Fixed Charge

It has been shown in Figure 2.9 that after ELA, the fixed charge of SiO_2 is similar to that of devices annealed with RTA. Therefore, ELA does not increase the fixed charge of SiO_2.

Gate Leakage Current

Figure 2.19 compares the leakage currents of the PMOS capacitors treated with ELA or RTA. The ELA capacitor exhibits higher leakage current in both the accumulation and depletion regimes (2-3 times). However, since the ELA device has thinner $T_{ox,inv}$, the leakage current is expected to be larger and this is not an indication of SiO_2 degradation.

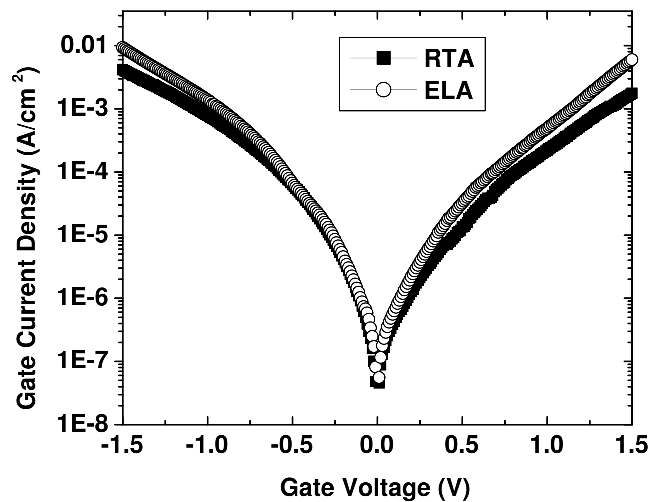

Figure 2.19. Gate leakage current density as a function of gate voltage. "ELA" indicates that the gate Si film was annealed with 10 laser shots at $640 mJ/cm^2$. "RTA" indicates that the gate Si film was thermally annealed 10 hr at 600^oC followed by 5s at 850^oC. The a-Si films were capped with 70 nm LTO during ELA.

Time Zero to Dielectric Breakdown

The gate-oxide quality of laser-annealed gate electrodes was compared with that of thermally annealed ones. Figure 2.20 shows the distributions of Time Zero to Dielectric Breakdown (TZDB), which indicates that the oxide qualities are comparable both cases.

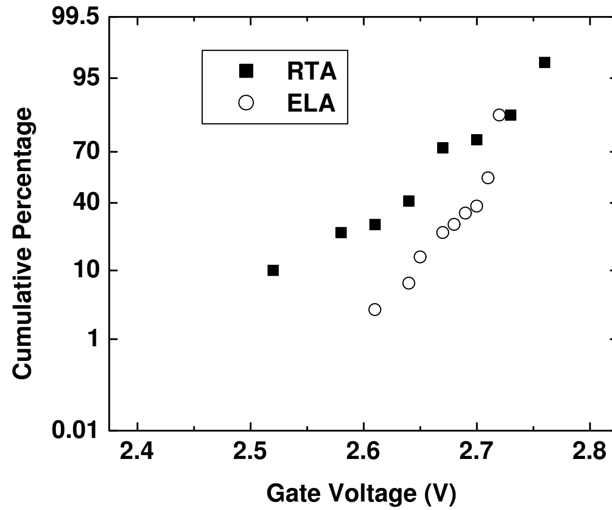

Figure 2.20. Time Zero to Dielectric Breakdown distributions for capacitors with threshold-adjust implant with ELA (10 shots, $640 mJ/cm^2$) gates vs. RTA (10 hr at 600^oC + 5s at 850^oC) gates. The a-Si films were capped with 70 nm LTO during ELA.

2.6.2 Channel Hole Mobility

PMOS transistor with L=$5\mu m$, W=$50\mu m$ was fabricated by using the gate last process described in Section 2.9. Either ELA or RTA was used to crystallize the a-Si gate and activate the dopants. Figure 2.21 shows the measured mobilities of the samples together with the universal mobility curve. It is found the hole mobility of ELA sample is similar to those of RTA sample as well as the universal curve at high transverse electric field. At low electric field, the mobility of RTA sample is degraded,

probably due to boron penetration. Therefore, from the hole mobility comparison, it can be further deduced that the ELA process doesn't degrade the SiO_2/Si channel interface quality and furthermore, it can avoid boron penetration.

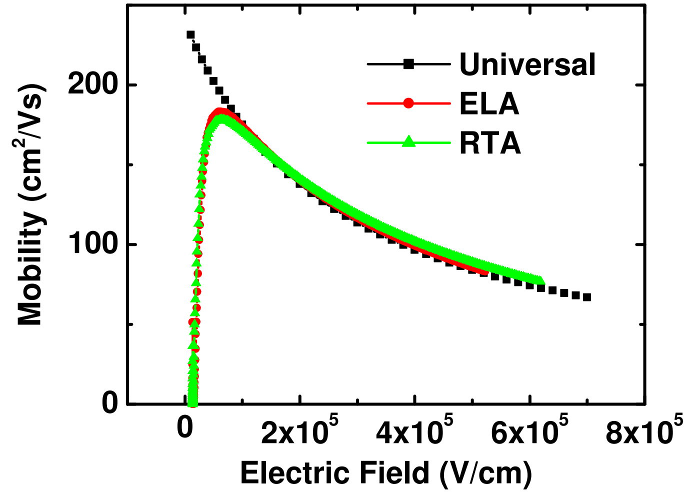

Figure 2.21. Extracted hole mobility of PMOS transistor (L=5μm, W=50μm) annealed with ELA (550mJ/cm^2) and RTA (800^oC1min + 1000^oC spike) respectively. The universal hole mobility curve is also included as reference.

2.6.3 HfO_2 gate dielectric

Since incorporating high-κ dielectric is considered as a possible solution to meet the near-term scaling challenges [1], the compatibility of ELA with high-κ gate dielectrics such as HfO_2 is worth studying. HfO_2 is known to be leaky after RTA due to either crystallization of the dielectric [6] or oxygen vacancy generation [5]. Such phenomena can be collectively understood as the atomic movement under the high temperature process of RTA. And as it has been proved that the ELA duration is short enough to prevent any boron penetration through SiO_2 (or atomic movement of boron in the solid SiO_2), it is, therefore, natural to expect that ELA might be

able to reduce or prevent atomic movement which leads to crystallization and oxygen vacancy generation in HfO_2.

PMOS capacitors using HfO_2 were prepared in a manner similar to the process described in Section 2.3. However, the a-Si was in-situ doped with boron. HfO_2 was deposited by Rapid Thermal Chemical Vapor Deposition (RTCVD) at 500^oC using hafnium t-butoxide as precursor. Excimer laser pulses with enough energy to fully melt the gate film were then applied to activate the dopants. For some samples, a solid-phase crystallization (SPC, 600^oC 4 hours) anneal was performed too.

Figure 2.22 shows the CV curve and QMCV fitting of an ELA+SPC PMOS sample. The simulation fits the measured data well and gate depletion is minimized with extracted effective poly-Si dopant activation concentration as large as $4 \times 10^{20} cm^{-2}$. Therefore, ELA can activate the dopants on HfO_2 capacitors as well as on SiO_2 capacitors. However, there is a strong Fermi-level pinning effect which results in V_{FB} of about $-0.2V$ and the V_{FB} cannot be restored to the ideal value after further SPC anneal. The pinning effect is likely due to the formation of oxygen vacancies in the HfO_2 even during the a-Si gate deposition [5] and is not the result of the rapid quenching process during ELA.

The gate leakage current densities of the PMOS capacitor was compared between the ELA+SPC sample and SPC+RTA samples. The results are shown in Figure 2.23. The ELA+SPC sample has 5 to 7 orders of magnitude lower leakage current density than the SPC+RTA samples. This indirectly proves that the short processing time of ELA substantially reduces atomic movement in the solid HfO_2 film and, thus, no degradation occurs. This is consistent with the previous result that boron penetration is suppressed in SiO_2.

RTA does not just increase the gate leakage current but also reduces the Equivalent Oxide Thickness (EOT) of the HfO_2. Figure 2.24 shows that if the ELA sample

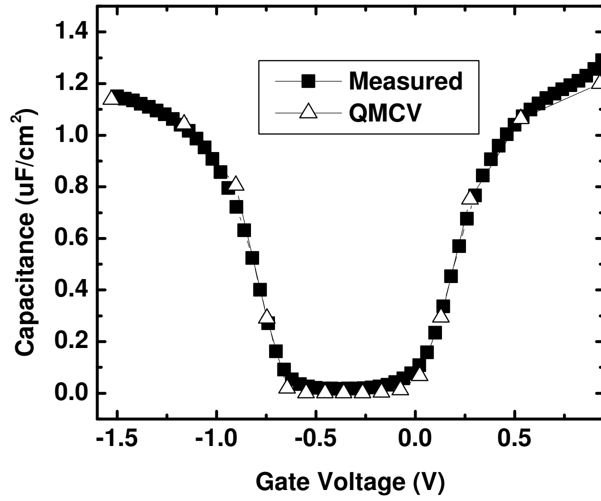

Figure 2.22. Measured gate capacitance *vs.* gate voltage (C-V) characteristics for PMOS capacitors with p+ poly-Si gates annealed by ELA (544 mJ/cm^2) followed by solid-phase crystallization (SPC) annealing at 600^oC for 4 hours. QMCV curve is also shown. Parameters used for the QMCV curve are T_{ox}=23$\mathring{A}$ and N_{poly}=$4 \times 10^{20} cm^{-3}$.

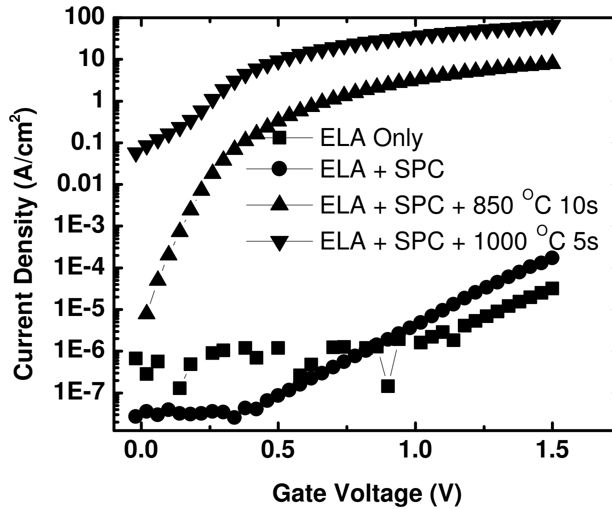

Figure 2.23. Measured leakage current density of high-κ PMOS capacitors (15nm HfO_2 with CET=3nm) for various gate-annealing processes.

is further annealed at 800^oC for 10 seconds (a typical gate implant thermal anneal performed prior to source/drain ion implantation in a CMOS process), the EOT in-

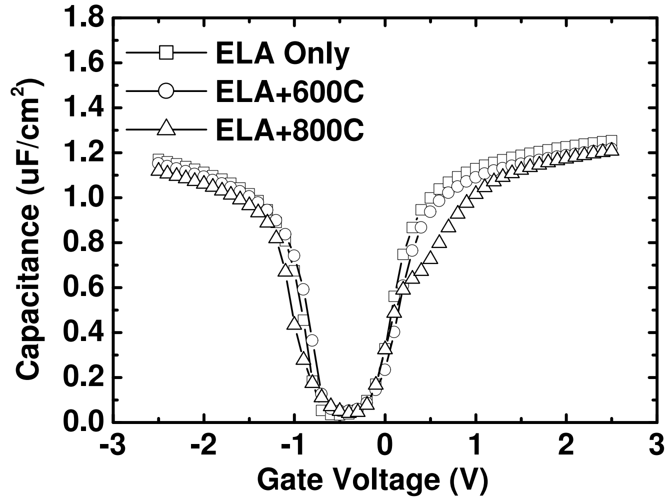

Figure 2.24. Measured C-V curves of PMOS capacitors (boron implanted gate) with HfO_2 gate dielectric (initial EOT=2.3nm) after different thermal treatments. "600C" means 600^oC for 1min and "800C" means 800^oC for 10s. All samples received FGA at 400^oC for 20 minutes.

creases, probably due to the formation of an interfacial layer at the Si channel interface [5]. Moreover, a high density of interface traps is formed close to the conduction band edge as shown in the CV curve, and cannot be eliminated by FGA (400^oC, 20min).

Figure 2.25 shows the distribution of leakage current density for HfO_2 PMOS capacitors after different thermal treatments. The leakage current increases with the annealing thermal budget, even though EOT increases at the same time (Figure 2.24). The increase in leakage current can be due to boron penetration (because the diffusivity of B in HfO_2 is 4 orders of magnitude higher than that in SiO_2 [28]) or oxygen vacancy formation. ELA is effective to avoid these problems because of its ultra-short process time. Therefore, the best integration of ELA with HfO_2 is to have SPC anneal at $\sim 600^oC$ after ELA to increase the dopant activation while minimizing the formation of an interfacial layer and gate leakage current.

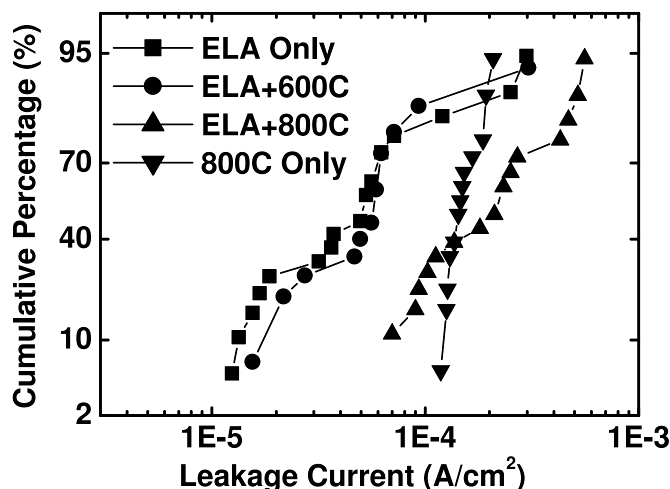

Figure 2.25. Distribution of leakage current density at V_G=1V for boron-implanted p+ poly-Si-gated PMOS capacitors (HfO_2 gate dielectric) after different thermal treatments. Initial EOT = 2.3nm. "600C" means 600^oC for 1min and "800C" means 800^oC for 10s

2.7 Summary

In this Chapter, we have shown that melt-ELA is a promising technique to meet the near-term gate stack scaling challenges. Resistivity of boron and phosphorus doped Si films as low as $\sim 0.5m\Omega\cdot cm$ and $\sim 0.7m\Omega\cdot cm$ can be achieved, respectively. It is found that, after ELA, there is a shift of effective gate work function (WF) of p+ poly-Si gate towards the mid-gap by $\sim 470meV$ and negligible shift of n+ poly-Si gate WF. It is therefore believed that the shift is due to Fermi-level pinning caused by the rapid quenching process rather than strain effects. The pinning was modeled by dipole formation and it is found that the activation energy to recover the pinning is very low ($\sim 0.4eV$) and the WF can be fully recovered by furnace annealing at moderate temperature ($\sim 600^oC$).

Melt-ELA maximizes the ratio of the dopant diffusion coefficient in the gate electrode to that in the gate dielectric by melting the gate film and keeping the di-

electric intact. With additional SPC anneal, effective dopant activation greater than $3\times10^{20}cm^{-2}$ in both PMOS and NMOS gates can be attained without any noticeable boron penetration, and the gate depletion effect is reduced to be $< 1\AA$.

ELA process does not degrade gate dielectrics (SiO_2 and HfO_2). The oxide fixed charge, leakage current density and Time-Zero-to-Dielectric-Breakdown data of ELA sample are similar to those of RTA sample. Moreover, the hole mobility is not degraded neither. This is probably due to the ultra-short annealing time ($\sim 100ns$) despite the high peak temperature of the dielectric. With such a short annealing duration, atomic movement in the solid dielectric is very limited. This is confirmed by the fact that the ELA HfO_2 capacitor shows 5-7 orders of magnitude lower leakage current density than the RTA one. However, ELA cannot solve the problem of Fermi-level pinning in a poly-Si/HfO_2 gate stack.

In summary, ELA is promising to meet the near-term gate-stack scaling challenges because 1) it can maximize the dopant activation in dual-gate CMOS process while preventing boron penetration and 2) it meets the thermal constraints of high-κ dielectrics. Figure 2.26 shows the leakage current density for PMOS capacitors with HfO_2 gate dielectric and in-situ-boron-doped gate electrodes, for different values of EOT (extracted using QMCV fitting) after ELA. The 65nm and 45nm CMOS technology requirements are also plotted for reference [29]. It can be seen that the roadmap specifications for gate leakage current density can be met by using ELA to activate dopants in poly-Si on HfO_2.

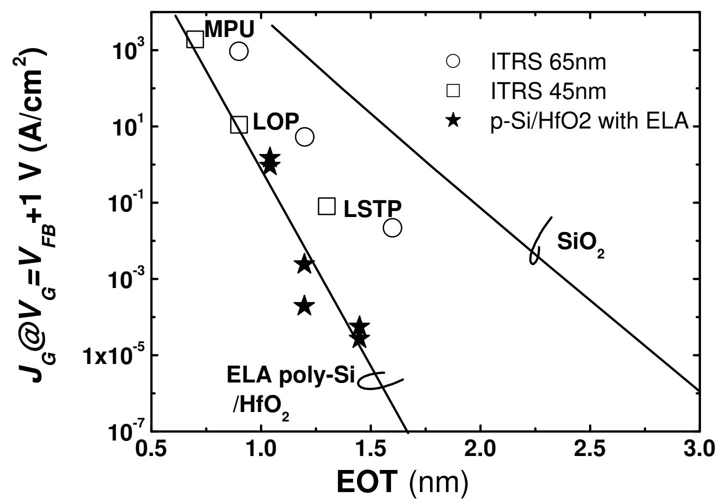

Figure 2.26. Gate leakage current density vs. EOT for in-situ boron-doped poly-Si/HfO_2 gate stack after ELA. For reference, the leakage current density of SiO_2 is plotted, as well as the roadmap requirements for microprocessor unit (MPU), low operating power (LOP), and low standby power (LSTP) devices at the 45 nm and 65nm CMOS technology nodes [29]

References

[1] *The International Technology Roadmap for Semiconductors*, Semiconductor Industry Association, 2005.

[2] H.-H. Tseng, D. O'Meara, P. Tobin, V. Wang, X. Guo, R. Hegde, I. Yang, P. Gilbert, R. Cotton, and L. Hebert, "Reduced gate leakage current and boron penetration of 0.18μm 1.5 V MOSFETs using integrated RTCVD oxynitride gate dielectric," in *International Electron Devices Meeting. Technical Digest*, Dec. 1998, pp. 793–6.

[3] M. Cao, P. Vande Voorde, M. Cox, and W. Greene, "Boron diffusion and penetration in ultrathin oxide with poly-Si gate," *IEEE Electron Device Letters*, vol. 19, pp. 291–3, Aug. 1998.

[4] K. Krisch, M. Green, F. Baumann, D. Brasen, L. Feldman, and L. Manchanda, "Thickness dependence of boron penetration through O_2 and N_2O-grown gate oxides and its impact on threshold voltage variation," *IEEE Transactions on Electron Devices*, vol. 43, pp. 982–990, June 1996.

[5] H. Takeuchi, H. Y. Wong, D. Ha, and T.-J. King, "Impact of oxygen vacancies on high-k gate stack engineering," in *International Electron Devices Meeting. Technical Digest*, Dec. 2004, pp. 829–834.

[6] H. Kim, P. C. McIntyre, and K. C. Saraswat, "Effects of crystallization on the electrical properties of ultrathin HfO_2 dielectrics grown by atomic layer deposition," *Applied Physics Letters*, vol. 82, pp. 106–108, 2003.

[7] A. Marmorstein, A. T. Voutsas, and R. Solanki, "A systematic study and optimization of parameters affecting grain size and surface roughness in excimer laser annealed polysilicon thin films," *Journal of Applied Physics*, vol. 82, pp. 4303–4309, 1997.

[8] J. S. Im, H. J. Kim, and M. O. Thompson, "Phase transformation mechanisms involved in excimer laser crystallization of amorphous silicon films," *Applied Physics Letters*, vol. 63, pp. 1969–1971, 1993.

[9] J. S. Im and H. J. Kim, "On the super lateral growth phenomenon observed in excimer laser-induced crystallization of thin Si films," *Applied Physics Letters*, vol. 64, pp. 2303–2305, 1994.

[10] B. Yu, Y. Wang, H. Wang, Q. Xiang, C. Riccobene, S. Talwar, and M.-R. Lin, "70 nm MOSFET with ultra-shallow, abrupt, and super-doped S/D extension implemented by laser thermal process (LTP)," in *International Electron Devices Meeting. Technical Digest*, Dec. 1999, pp. 509–512.

[11] H. Tsukamoto, "150 nm-gate p-channel MOSFET fabrication using laser annealing," *IEE Electronics Letters*, vol. 34, pp. 401–403, 1998.

[12] Y. Takamura, S. H. Jain, P. B. Griffin, and J. D. Plummer, "Thermal stability of dopants in laser annealed silicon," *Journal of Applied Physics*, vol. 92, pp. 230–234, 2002.

[13] D. E. Aspnes and A. A. Studna, "Dielectric functions and optical parameters of Si, Ge, GaP, GaAs, GaSb, InP, InAs, and InSb from 1.5 to 6.0 eV," *Phys. Rev. B*, vol. 27, pp. 985–1009, 1983.

[14] H. Kodera, "Diffusion coefficients of impurities in silicon melt," *Japanese Journal of Applied Physics*, vol. 2, pp. 212–219, 1963.

[15] M. O. Thompson, G. J. Galvin, J. W. Mayer, P. S. Peercy, J. M. Poate, D. C. Jacobson, A. G. Cullis, and N. G. Chew, "Melting temperature and explosive crystallization of amorphous silicon during pulsed laser irradiation," *Phys. Rev. Lett*, vol. 52, pp. 2360–2363, 1984.

[16] R. F. Wood and G. E. Giles, "Macroscopic theory of pulsed-laser annealing. I. Thermal transport and melting," *Phys. Rev. B*, vol. 23, pp. 2923–2942, 1981.

[17] R. F. Wood, C. W. White, and R. Young, Eds., *Semiconductors and Semimetals: Pulsed Laser Processing of Semiconductors.* Academic Press, 1984.

[18] C. Hill, "Laser annealing of device structures," in *Laser And Electron Beam Processing of Electronic Materials.* The Electrochemical Society, Inc., 1980, pp. 26–43.

[19] H. Kim and J. Im, "Multiple Pulse Irradiation Effects in Excimer Laser-Induced Crystallization of Amorphous Si Films," in *Mater. Res. Soc. Symp*, vol. 321, 1994, pp. 665–670.

[20] QMCV Simulator. [Online]. Available: http://www-device.eecs.berkeley.edu/qmcv

[21] D. K. Schroder, *Semiconductor material and device characterization.* NY, USA: John Wiley & Sons, 1998.

[22] S. Wolf, *Silicon Processing for the VLSI Era.* Lattice Press, 1995.

[23] J. W. Slotboom, "The pn-product in silicon," *Solid-State Electronics*, vol. 20, no. 4, pp. 279–83, 1977.

[24] H. Hasegawa and H. Ohno, "Unified disorder induced gap state model for insulator-semiconductor and metal-semiconductor interfaces," *Journal of Vacuum Science & Technology B*, vol. 4, pp. 1130–1138, 1986.

[25] Y.-C. Yeo, P. Ranade, Q. Lu, R. Lin, T.-J. King, and C. Hu, "Effects of high-κ dielectrics on the workfunctions of metal and silicon gates," in *Symposium on VLSI Technology. Digest of Technical Papers*, 2001, p. 49.

[26] M. M. Rieger and P. Vogl, "Electronic-band parameters in strained $Si_{1-x}Ge_x$ alloys on $Si_{1-y}Ge_y$ substrates," *Phys. Rev. B*, vol. 48, pp. 14 276–4287, 1993.

[27] Y. Chong, H.-J. Gossmann, M. Thompson, K. Pey, A. Wee, S. Talwar, and L. Chan, "Reduction of carrier depletion in p^+ polysilicon gates using laser thermal processing," *IEEE Electron Device Letters*, vol. 24, pp. 360–2, 2003.

[28] K. Suzuki, H. Tashiro, Y. Morisaki, and Y. Sugita, "Diffusion coefficient of B in HfO_2," *IEEE Transactions on Electron Devices*, vol. 50, pp. 1550–1552, 2003.

[29] *The International Technology Roadmap for Semiconductors, 2004 Update*, Semiconductor Industry Association, 2004.

2.9 Appendix: Process Flow for Gate Last ELA MOSFETs

Process Name	Process Specification	Equipment
Wafer Labeling	2 n-type 6" (N9, N10) and 1 p-type 6" wafers (P5)	
LOCOS formation	3339–3682A	
Dummy Gate Oxide		
Preclean	Piranha 10min and HF 1:25 90sec	Sink 6
Oxidation	950^oC DRY Oxidation 14min (15nm) N_2 annealing @ 1000^oC 20min	Tystar 1
Dummy Gate Formation		
SiGe Deposition	SiH_4=200, 300mT, 550^oC, 30sec. SiH_4=124, GeH_4=80, 16min. (150nm)	Tystar 19
Gate Patterning	coating / litho ($14.5mJ/cm^2$) / develop	ASML
Gate Etch	Recipe 5963, End 12sec, O.E. 12s	LAM5
S/D I/I	P5 : P, 50keV, 5E15 N9, N10: B, 20keV, 5E15	
Dummy gate removal		
Resist Removal		Matrix
Gate Removal	SC1 Solution at 70^oC (5 H_2O:1 H_2O_2:1 NH_4OH)	
S/D Activation and Damage Annealing		
Cleaning	Piranha 10min	Sink 8 and 6
SPC Anneal	550^oC 4 hours	Tystar 2
Gate Oxide Formation		
Preclean	Piranha 10min and HF 1:25 90sec	Sink 6
Oxidation	750^oC DRY Oxidation 14min.(2.6nm) POA 900^oC 20min.	Tystar 1
a-Si Gate Electrode Deposition		
N+ Deposition	530^oC 60min, SiH_4=200,PH_3=9, 600mT (50nm)	Tystar 19
P+ Deposition	450^oC 5.5min, Si_2H_6=100, B_2H_6=90, 300mT (50nm)	Tystar 19
ELA Protection Layer Formation		
LTO	2.5min 450^oC (50nm)	Tystar 11
n+ a-Si	480^oC Si2H6/PH3 100/9sccm 300mTr 20min. (50nm)	Tystar 19
Patterning	coating/litho/develop/etch/ashing/post cleanin	
ELA	laser annealing 308nm 0–1000mJ/cm2	Lextra
Si Removal	Recipe 5936	LAM5
LTO removal	BHF 5:1	Sink 8
FGA	H2/N2 400^oC	Tystar 18

Chapter 3

CMOS Process Integration of ELA Gate Technology

3.1 Introduction

Any novel gate annealing technique should be easily integratable into a traditional CMOS process, otherwise the benefits of using the new technique will be overshadowed. To integrate ELA gate annealing into a traditional CMOS process, the following integration issues (metrics) should be considered,

1. The extra dopant activated by ELA should be retained in the post-ELA steps. This is especially true for the source/drain (S/D) dopant activation step, during which high temperature is required to minimize the S/D sheet resistance.

2. The high dopant concentration at the gate electrode/dielectric interface should not cause extra dielectric degradation. Since traditional S/D activation involves high temperature Rapid Thermal Annealing (RTA), there is a concern of serious boron penetration when ELA is used before RTA.

3. There should not be too many modifications to the traditional CMOS process if melt-ELA is integrated, in order to limit the incremental process cost and retain high production yield.

4. The process should be scalable to deep-submicron feature sizes. Since ELA is proposed as a solution to the near-term gate stack challenges, its applicability down to 32nm technology node is discussed.

In this chapter, the compatibility of melt-ELA with rapid-thermal-annealing (RTA) and flash-lamp-annealing (FLA) is first discussed. RTA is currently widely used for S/D dopants activation and FLA is the promising candidate for the future S/D dopants activation in sub-65nm technology nodes [1]. After that, two integration approaches will be discussed to integrate melt-ELA into traditional CMOS processes, namely masked-ELA and self-aligned-shallow-trench-isolation (SA-STI) process.

3.2 Integration Issues of Melt-ELA in Conventional CMOS Processes

There are three major integration challenges when melt-ELA for gate annealing is introduced. The first one is the boron penetration issue during subsequent high-temperature processes. After ELA, a large amount of dopants are located at the gate electrode/dielectric interface (Figure 2.5), which does not normally occur in a conventional CMOS process. Therefore, during S/D dopant activation, which is usually done after gate activation, boron penetration can be more serious than for the traditional CMOS process.

Figure 3.1 shows the CV curves of an ELA sample after subsequent annealings. The curve consistently shifts to the right with higher annealing thermal budget. Below

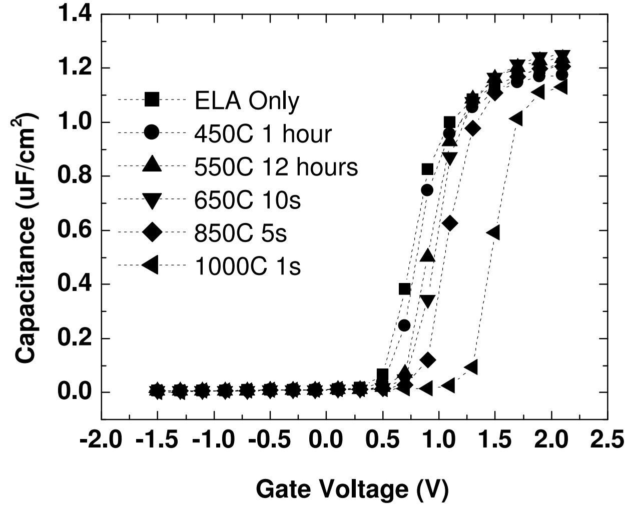

Figure 3.1. Measured Capacitance-*vs.*-Voltage curves of PMOS capacitor at 100kHz after various post-ELA annealing conditions. All samples received 400°C Froming Gas Annealing (FGA).

850^oC, the shift is due to the recovery of the effective Work Function (WF) of the ELA gate as discussed in Section 2.4.3. Above 850^oC, the shift is due to boron penetration. If we assume that the boron penetrating through SiO_2 into the Si substrate are all activated, the flat-band voltage (Equation 1.2) can be rewritten as [2]:

$$V_{FB} = \frac{1}{q}(\Phi_{Gate} - \Phi_S) - \frac{Q_f}{C_{gate}} + \frac{qN_B}{C_{gate}} \tag{3.1}$$

where N_B is the effective areal density of boron diffused into the channel. By using the V_{FB} difference between 850^oC and 1000^oC anneals, it can be calculated that there are about $4 \times 10^{12} cm^{-2}$ boron atoms that penetrated into the channel after 1000^oC annealing for 1 second, which is very serious. Therefore, ELA is not compatible with additional RTA or spike annealing.

Secondly, melt ELA should be done before the gate is patterned to reduce any edge effects (optical and thermal) and reduce the possibility of the silicon gate shorting

the S/D regions. This means that blanket laser annealing is required, during which the a-Si over both the isolation and active areas are annealed at the same time. However, as SiO_2 (commonly used material for device isolation) has much lower thermal conductivity ($\sim 1.4Wm^{-1}K^{-1}$) than c-Si ($\sim 148Wm^{-1}K^{-1}$), it is possible that at a laser fluence level which is still too low to fully melt the a-Si over the active area (to be patterned to be the gate electrode), the a-Si over the isolation (oxide) region is already over-melted or ablated. The resulting poly-Si over the isolation areas will have very rough topologies and, during gate patterning and other subsequent processes, debris and broken lines will form which probably will dramatically reduce the production yield.

Indeed, simulation shows that, for a structure similar to as that given in Figure 2.2, only $160mJ/cm^2$ is required to fully melt a-Si over 300nm isolation oxide, which is only 27% of the energy required to fully melt the a-Si over ~2nm gate oxide in the active region. Figure 3.2 shows the micrograph of a transistor with $5\mu m$ gate length after ELA and gate patterning. The fluence was $480mJ/cm^2$ and the deposited a-Si thickness was 60nm which requires about $680mJ/cm^2$ to be fully melted over the active region. However, the poly-Si over the isolation area is already rough and debris are found even with a laser fluence of $480mJ/cm^2$. Therefore, blanket ELA cannot be used because either the gate dopants are not fully activated at low fluence or broken lines will form over the isolation region at high fluence. Indeed, most of the $1\mu m$ gate-length transistor gate electrodes are broken at $680mJ/cm^2$ with blanket ELA. Therefore, a method has to be developed in order to selectively anneal the gate material in the active region.

The third integration challenge is that there may be intermixing of p+ and n+ dopants in the molten silicon during ELA as the dopants can diffuse laterally rapidly. Counter-doping can occur, resulting in gate depletion and even WF variations. This

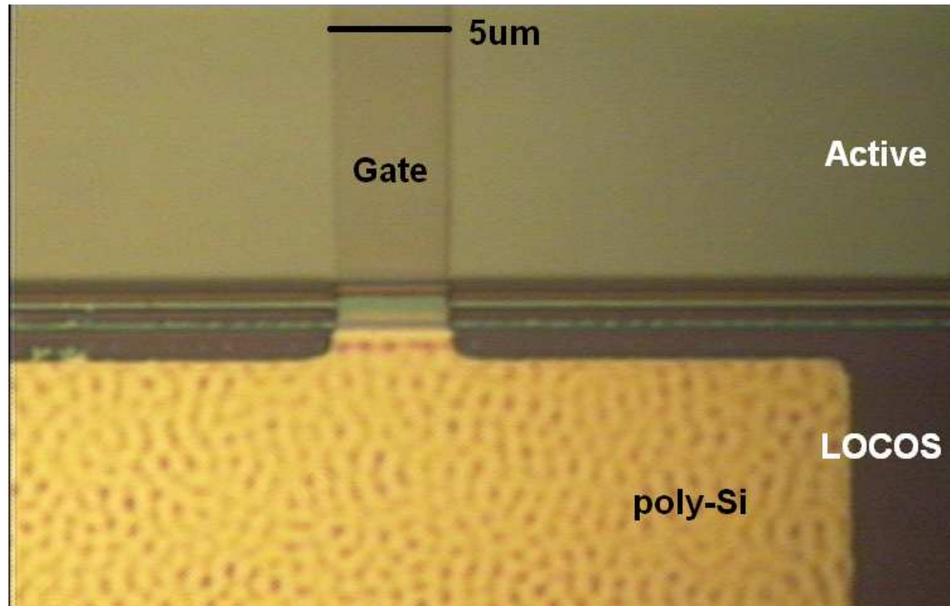

Figure 3.2. Micrograph showing the morphology of patterned poly-Si gate on both active and isolation (LOCOS) areas after ELA ($480mJ/cm^2$). The fluence is not enough to fully melting the a-Si on the active area but is already too high such that the roughness of the gate pad is increased and debris are found.

problem can be serious in scaled technologies as the pMOSFETs and nMOSFETs are getting closer to each other.

3.3 Integration with FLA

It is increasingly difficult for RTA and spike annealing to meet the requirements of junction abruptness and junction depth for ultra-shallow junction (USJ) formation in sub-90nm technology nodes due the inevitable dopant diffusion [3]. Flash lamp annealing (FLA) has been shown to be a promising S/D activation technique for 65nm and 45nm technologies [4; 5; 6; 7] and may be even adequate for 32nm technology. In a typical FLA setup, the wafer is pre-heated to about 700^oC to 800^oC by back-side irradiation with W-Halogen lamps. Xe flash lamp on the front side is used to bring the wafer surface up to a peak temperature of $\sim 1300^oC$ for less than 1 millisecond (ms), which then cools down rapidly using the substrate as a heat sink [5]. Therefore, FLA

can reduce the diffusion of dopants to be much less than for RTA and at the same time activate more dopants due to the elevated temperature [4]. However, usually gate dopants are implanted as shallow as possible to prevent dopants from penetrating through the gate dielectric and into the substrate and, since there is only negligible dopant diffusion under FLA, gate depletion effect will be very serious if FLA alone is used to anneal both the S/D and gate [8]. A separate pre-annealing step for gate dopant activation is thus required.

Because of the aforementioned properties of FLA, it is naturally expected to be compatible with melt-ELA. During gate dopant activation by ELA, gate dopants diffuse quickly towards the gate interface. And during S/D activation by FLA, S/D regions are activated without much dopant diffusion and thus boron penetration can be avoided. Moreover, since FLA's peak temperature is much higher than traditional RTA process ($< 1100^oC$), this can also prevent dopants deactivation in the gate.

Experiments were carried out to study the effect of FLA on the gate depletion effect and boron penetration. PMOS capacitors were fabricated and annealed by ELA as described in Section 2.3. Small dies were then put on a carrier wafer and the system was brought to an intermediate temperature of 760^oC (only for seconds) during FLA. This step is necessary because it has been shown that slip dislocations and cracks are generated in the wafer substrate due to thermal strain if this intermediate step is omitted [1]. Then the temperature was ramped up at a rate of $200^oC/s$ to be between $\sim 1260^oC$ and $\sim 1360^oC$ for less than 1 milli-second and then the wafer cooled by thermal conduction through the substrate. Figure 3.3 shows the measured CV curves after ELA and ELA+FLA (both received Forming Gas Annealing, FGA). The curves are essentially the same, except there is a shift of $\sim 150mV$ after FLA (Figure 3.3), which is much less that in the ELA+RTA case (Figure 3.1). And if the curves are shifted to overlap each other, ELA and ELA+FLA samples have almost the same

inversion capacitance. Therefore, benefit of ELA (reduced gate depletion effect) is not reduced by FLA.

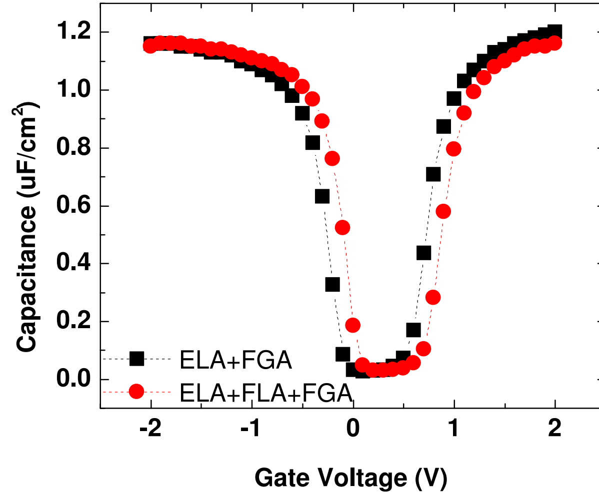

Figure 3.3. Measured Capacitance-*v.s.*-voltage curves of ELA sample and ELA followed by FLA annealing. There is a 150mV shift after FLA.

The V_{FB} shift of the ELA+FLA curve can be explained by recovery of WF or boron penetration. However, it is most likely due to the recovery of WF for two reasons. Firstly, the curve is shifted without any distortion in the inversion region. Secondly, the leakage current density and flat-band voltage after FLA is similar to that of ELA+SPC (550^oC annealing for 4 hours) sample (Figure 3.4). If there is any dopant penetration, the leakage current is expected to increase.

3.4 Selective ELA via Masking

In order to prevent the p-type dopants from intermixing with the n-type dopants during melt-ELA and also prevent the ablation of a-Si over the isolation areas mentioned in Section 3.2, a selective ELA via masking process is proposed. Figure 3.5 shows the idea of ELA via masking process. On top of the LTO, a masking layer

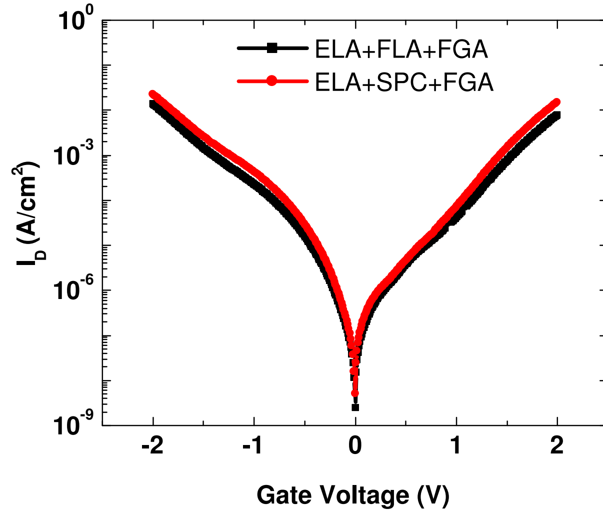

Figure 3.4. Measured gate leakage densities of ELA + SPC and ELA + FLA samples. It is found that FLA does not increase the leakage current significantly.

(Mask) is added and patterned using the inverse active mask, so that the laser photons over the isolation area are absorbed in the masking layer and, thus, prevent the melting of the a-Si underneath. The a-Si on active area is still allowed to melt and crystallize because it is not covered by any absorption mask. Since a-Si on the isolation does not melt, there is no way for the p-type and n-type dopants to cross-diffuse neither.

The fabrication process is exactly the same as the one described in Section 2.3. But after the LTO step, 70nm of a-Si in-situ doped with boron by LPCVD was deposited as the mask. The reason of using boron doped a-Si is because boron can enhance the deposition rate of the film and thus the film can be deposited at low enough temperature (450^oC) to prevent the crystallization of the a-Si gate. The masking a-Si is then patterned using chlorine-based dry etching and the LTO oxide can serve as an excellent etch stop. After ELA, the mask can be removed by either wet etch ($33\% H_2O/3\% NH_4F/64\% HNO_3$), which has very good selectivity over LTO, or by

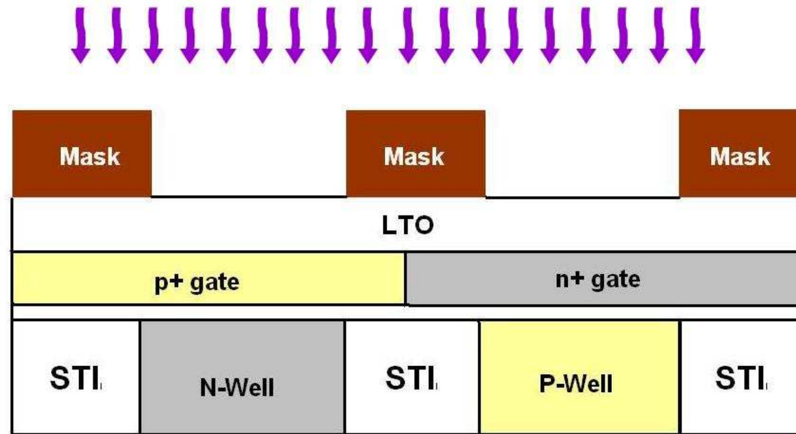

Figure 3.5. Cross-sectional view of the proposed structure after a-Si masking-layer patterning. STI is shallow trench isolation and can represent any other isolation techniques. The stack is 70nm a-Si/43nm LTO/ 55nm gate a-Si/ Gate oxide/ STI or Si substrate. Laser is irradiated from the top.

dry etch followed by HF dip to remove the LTO. Gate patterning and remaining conventional CMOS processing steps then follow to complete the fabrication.

In order to confirm the idea, a wafer with structure similar to that shown in Figure 3.5 without STI and gate doping was made and processed. After the mask-a-Si and LTO are moved, the reflectance of the structure at different wavelengths was taken with reference to an unannealed sample. Figure 3.6 shows that at the masked region, the reflectance is almost the same as that of the reference, while the unmasked region is very different. This confirms that the mask successfully prevented the underlying a-Si from melting while the a-Si in the unmasked area melted. Therefore, ablation over isolation (STI) area can be prevented.

pMOSFETs were fabricated using the masked ELA process with 300nm LOCOS isolation. A $580mJ/cm^2$ pulse was used to activate the gate. After the patterned a-Si masking layer is removed, a low fluence pulse ($160mJ/cm^2$) was used to active the a-Si on the isolation region so that electrical contact from the gate to the probe pad is possible. The poly-Si resistivity over the active and isolation (300nm LOCOS) regions is $< 0.6m\Omega \cdot cm$. SPC (600^oC 4 hours) and RTA (850^oC 5s) were used to

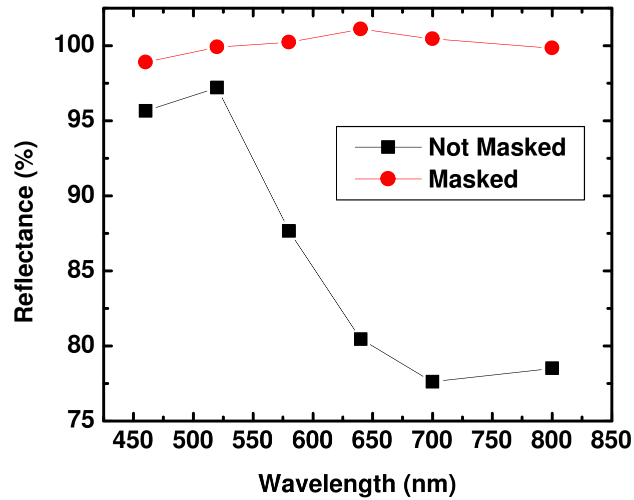

Figure 3.6. Reflectance measurements of gate silicon layer ("Masked" = under mask silicon layer, "Not Masked"= exposed to the laser) after ELA at $430mJ/cm^2$. Measured values are relative to an unannealed sample

activate the S/D dopants. There is no metalization. Figure 3.7 shows the measured $I_D - V_G$ curve of a W/L=50μm/5μm pMOSFET.

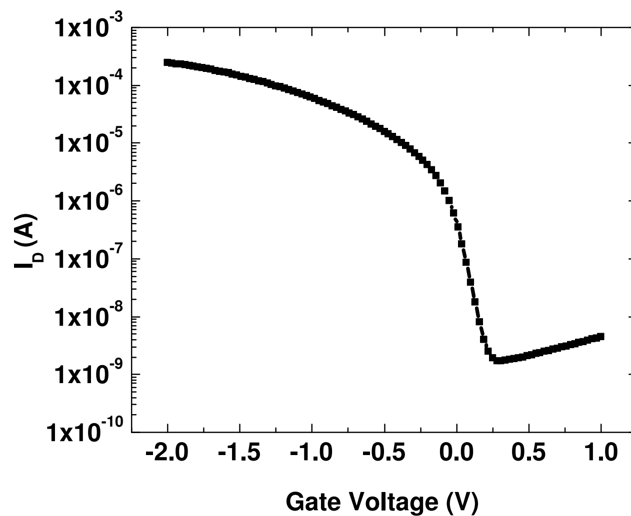

Figure 3.7. Measured $I_D - V_G$ curve of $W/L = 50/5$ pMOSFETs fabricated using the masked ELA process.

The above result shows that intermixing between p-type and n-type dopants can also be prevented via masking process because the a-Si on the isolation region does not melt. However, this is only true when the separation between the p-type and n-type active region is large enough. There is always lateral melting during ELA and if their seperation is too small, the a-Si on isolation region can still melt, although ablation can be prevented. In order to investigate the lateral melting extent, a cross-sectional SEM was taken after ELA but before the removal of the mask (Figure 3.8).

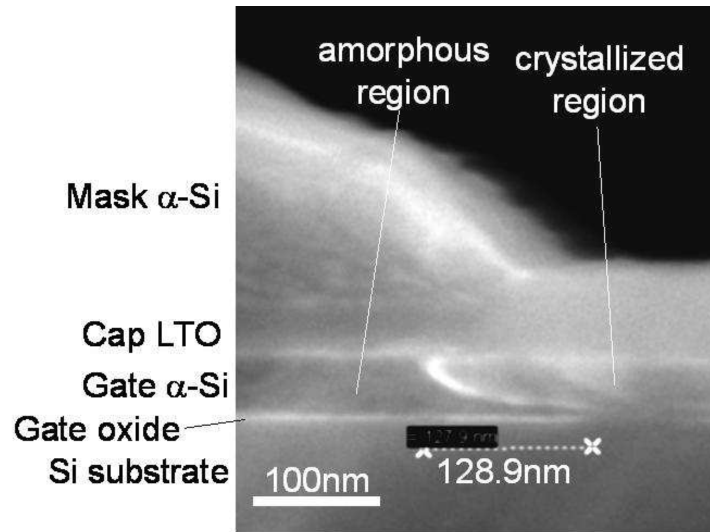

Figure 3.8. Cross-sectional SEM of sample illustrated in Figure 3.5 after high-fluence ($650mJ/cm^2$) ELA and Secco etch to delineate the a-Si boundary. The lateral melt distance underneath the edge of the a-Si mask is $\sim 100nm$.

From the figure, the lateral melting is found to be about two times the thickness of the gate electrode in one-sided melting process when full-melt fluence is used. It is assumed to be true for gate thicknesses with the same order of magnitude. Moreover, the lateral melting is minimum at the bottom of the a-Si. Theoretically, the lateral melting at the bottom of a-Si should be zero if the laser fluence is just enough to fully melt the a-Si on the active region. At the same time, according to ITRS 05, the gate thickness can be assumed to be two times the physical gate length [3] and usually, the n-type active to p-type active seperation is at least two times the half pitch in a technology node. Therefore, Table 3.1 can be constructed.

The Table shows that the p-to-n active seperation is larger than the single-sided lateral melting extent but smaller than the two-sided lateral melting extents. Therefore, the lateral melting is too much if both n and p regions are annealed (melt) at the same time (two-side melting), but intermixing can be prevented if there is only single-sided lateral melting. This can be implemented by repeating the mask ELA process twice with one mask exposing the p-region only and the other for n-region only, with the expense of additional mask cost. By using this scheme, the gate implantation has to be performed with the absorption mask so that there are no dopants in the a-Si underneath the mask. Otherwise, intermixing will still happen.

Year	2007	2010	2013
Technology Node (L_p/nm)	65	45	32
MPU Physical Gate Length (L_g/nm)	25	18	13
Expected Gate Thickness ($2L_g/nm$)	50	36	26
p-to-n Active Seperation ($2L_p/nm$)	130	90	64
Single-sided Lateral Melting at Full Melt Threshold ($4L_g/nm$)	100	72	52
Two-sided Lateral Melting at Full Melt Threshold ($8L_g/nm$)	200	144	104

Table 3.1. Spreadsheet showing that the p-to-n active seperation is larger than the single-sided lateral melting extent but smaller than the two-sided lateral melting extents.

3.5 Integration with Cypress SA-STI 65nm Process

3.5.1 Introduction

Shallow Trench Isolation (STI) has become the major CMOS isolation technique in sub-0.25μm technologies because it provides extreme planar surface topology for deep sub-micron lithography and abrupt transition between active and isolation areas

(elimination of the "bird's beak"), so that higher packing densities are possible [9; 10]. However, there are some drawbacks associated with STI. During the cleaning and etching steps after STI formation, there are often STI under-cuts next to the active region, resulting in dopant depletion, oxide thinning and the kink effect (in the $I_d - V_g$ characteristic) as poly-Si gate wraps around the active region in the under-cut area (Figure 3.9). Moreover, the Chemical-Mechanical Polishing (CMP) technique used in the STI process has to be very precisely controlled to avoid damaging the active area. During annealing, it is also possible to have n+ and p+ dopants cross-diffuse, resulting in counter-doping.

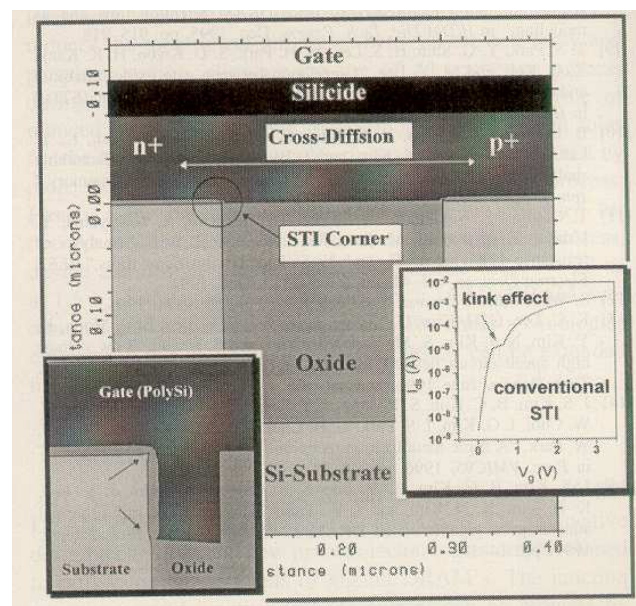

Figure 3.9. Cross sectional view of conventional STI with associated problems: Gate-wrap-around and gate oxide thinning (lower left inset) and kink effect due to parasitic corner device (right inset) (After [10]).

In view of these, U. Schwalke has proposed a novel technique called EXTIGATE (EXtended Trench Isolation GAte TEchnology) in which STI formation is after the gate dielectric and electrode deposition [11]. In this process, gate dielectric is first formed followed by poly-Si (gate electrode) deposition. The active areas are then patterned and the isolation regions are etched and filled with oxide. After planarization

using CMP, all the active areas are separated by STI and the poly-Si is self-aligned with the STI. Thus, dopant cross-diffusion and gate wrapping can be avoided (Figure 3.10).

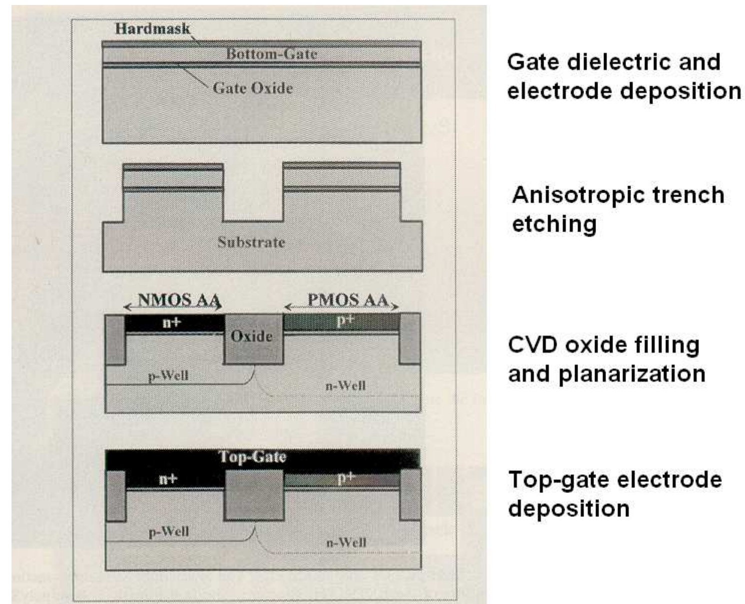

Figure 3.10. Essential part of the EXTIGATE process. Top-Gate is usually metallic material overlapping the poly-Si gates and interconnecting transistors within/between active areas (AA) (After [10]).

In Cypress Semiconductor, due to the self-aligned nature, the EXTIGATE process is also called Self-Aligned STI (SA-STI) [12]. In SA-STI, since the p+ and n+ poly-Si in the active areas are not interconnected and there is no poly-Si on top of the isolation regions, it provides an almost ideal structure for blanket laser annealing because ablation on isolation regions and p+/n+ cross-diffusion can be avoided. Therefore, an experiment was carried out to investigate the possibility of integrating ELA into Cypress 65nm SA-STI process.

3.5.2 Process Flow

The process flow of ELA SA-STI is outlined in Table 3.2. It is similar to the EXTIGATE process in Figure 3.10. Besides that the STI is formed after gate stack deposition, there are another three steps different from conventional STI processes.

Process Name	Process Description
O/N Gate Dielectric	LPCVD Nitride + Thermal Oxide + Annealing
Gate Electrode	76nm a-Si deposition by LPCVD
STI Formation	Gate stack and trench etching Liner annealing HDP-CVD oxide filling CMP
Amorphization Implant	Ge at different energies with dose = 6.8×10^{14}
Through Gate Implants	p-Well, n-Well, p-Channel and n-Channel implants
Poly Implant	B or BF_2 for PMOS and P for NMOS
LTO Deposition	40nm LTO deposited at 400^oC
Gate Annealing	Blanket ELA annealing/ RTA (control)
LTO removal	HF Dip
Gate Metal Formation	PVD W/WN_x
Nitride Capping	
Gate Patterning	Nitride etch and $W/WN_x/Poly$ etch
Selective Oxidation	Poly-Si side wall oxidation
S/D Extension	Halo and extension implants
Spacer Formation	Nitride Spacer
S/D Formation	n+/p+ S/D implantation Annealing
Backend Process	Up to Metal 1

Table 3.2. Process Flow of ELA SA-STI.

Firstly, there is a gate pre-amorphization implant (PAI) step by germanium (Ge). Since the a-Si gate is deposited before STI formation, the gate is crystallized to be poly-Si during the liner annealing process, which was conducted at high temperature (1000^oC) to cure the Si damage and round the corners in the trenches. Poly-Si has the same melting point as c-Si. PAI is thus used to amorphize the poly-Si, so

that lower laser fluence can be used to melt the gate. Secondly, channel and well implants are through the gate, so there are possible damages to the gate dielectric. This cannot be avoided in the SA-STI process because the liner anneal will cause too much dopant diffusion if the dopants were to be introduced before gate and STI formation. Fortunately, it has been shown not to be detrimental to the gate dielectric qualities [12; 10], probably due to their relatively high energies (since nuclear stopping power dominates only at low energies [9]) or relatively low implant doses. Thirdly, low temperature oxide (LTO) is deposited on top of STI and a-Si gate to prevent contamination and act as anti-reflection coating. It has to be removed after ELA and care has to be taken so that the STI is not over-etched.

3.5.3 Gate Pre-Amorphization Implant (PAI)

Gate pre-amorphization implant (PAI) by Ge is a critical step in ELA SA-STI process. The idea is to use Ge PAI to define the amorphization depth and thus the melt depth during ELA, as a-Si has lower melting point than poly-Si. Since amorphization depth defined by ion implantation is very uniform across the wafer, tolerance to laser fluence variation is enhanced. However, if the amorphization depth is not enough (low Ge implantation dose and energy), ELA will only bring dopants to a limited depth and there will be serious gate depletion effect. On the other hand, if too high an implantation dose or energy is used, gate oxide may be seriously degraded by the straggled Ge atoms. Therefore, it is necessary to study the amorphization process carefully.

In the standard Cypress SA-STI process, Ge PAI at 10keV is a standard step used to prevent channeling during gate implantation. It has been proved to be effective in amorphizing the poly-Si gate. Therefore, the same dose ($6.8 \times 10^{14} cm^{-2}$) was used with different implantation energies in this study.

Figure 3.11 shows a cross-sectional transmission electron microscopy (TEM) picture of active and isolation areas after Ge PAI at 45keV in SA-STI. 40nm LTO has already been deposited. It is clear that Ge-PAI is able to amorphize the poly-Si to a uniform depth. Moreover, as LTO was deposited at $400^{o}C$, the amorphized poly-Si was not recrystallized.

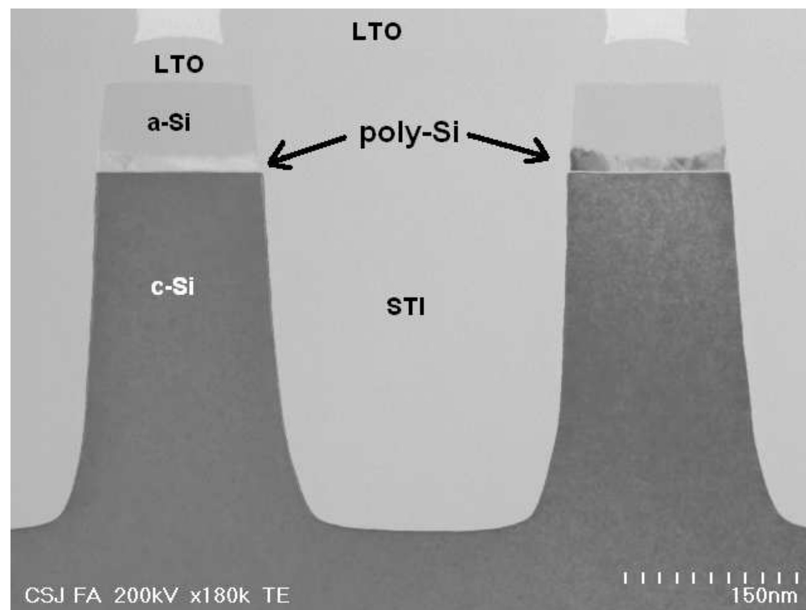

Figure 3.11. Cross-sectional TEM picture of active and isolation areas after Ge PAI at 45keV and LTO deposition.

To study the damage of gate dielectric by PAI, Monte Carlo simulations using SRIM 2003 [13] were performed. A simulation structure of $760\mathring{A}\ Si/20\mathring{A}\ SiO_2/1\mu m\ Si$ was defined. The simulation was completed with 10000 ions and full damage cascades model was turned on. Silicon and oxygen vacancies generated are recorded. The Si vacancy density is used to determine the amorphization depth and the oxygen vacancy is used for assessing the gate dielectric damage. N-well implant (NWI) and p-channel (PCHI) implant are also included in the simulation for comparison. The amorphization threshold (critical density) is defined as when the Si vacancy density is larger than 25% of the Si lattice density (i.e. $1.25 \times 10^{22} cm^{-3}$) [14].

Figure 3.12 shows the simulation results. As expected, the higher the implantation

energy, the deeper the amorphization depth. And the damage of gate dielectric increases as the implantation energy increases. It is of interest to note that when the Ge-PAI implantation energy is less or equal to 35keV, the dielectric damage is actually comparable to that of the through-gate NWI and PCHI.

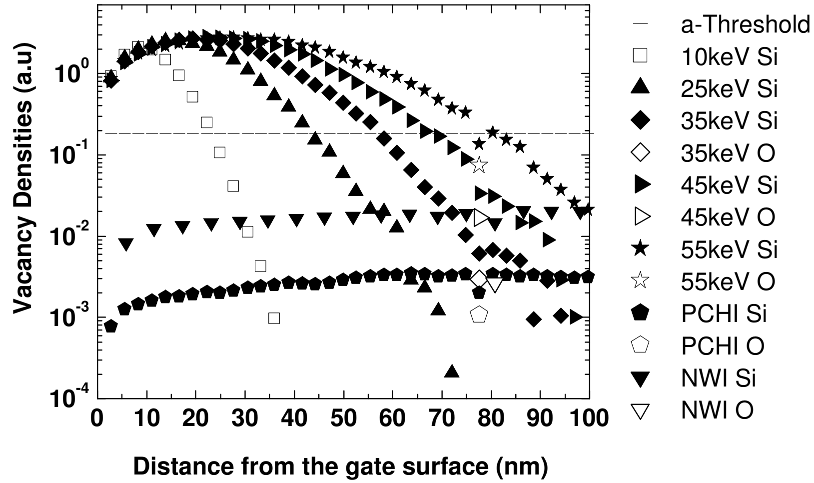

Figure 3.12. Simulation results using SRIM Monte Carlo simulation showing the Silicon (Si) and Oxygen (O) vacancy densities generated for different implantation conditions. All curves are for Ge implants with dose = $6.8 \times 10^{14} cm^{-2}$, except NWI is that of N-well Implant and PCHI is that of P-channel implant. Si vacancy density corresponding to amorphization is also given (a-Threshold).

Figure 3.13 plots the amorphization depths extracted from Figure 3.12. It is found that the amorphization depth can be fitted by the following equation:

$$Depth = R_P + 2.5\Delta R_p \tag{3.2}$$

where R_p and ΔR_p are the projected range and straggle of Ge implantation into Si respectively. Data from cross-sectional TEM measurement of 10keV and 45keV samples are also plotted and found to match pretty well with the simulation.

From Figure 3.12, the O vacancy density is about 5-6 times that produced by the

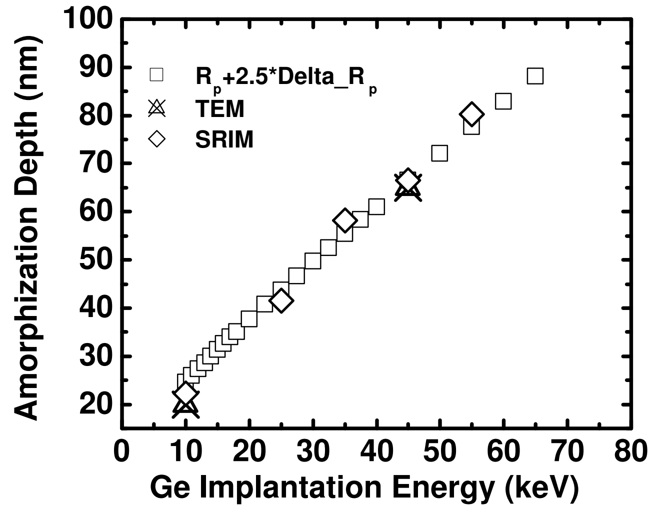

Figure 3.13. Amorphization depth of poly-Si gate after Ge implantation with dose $= 6.8 \times 10^{14} cm^{-2}$ at different energies. TEM is the measured TEM result. SRIM is the result of Monte Carlo simulation from Figure 3.12. R_p and Delta_R$_p$ are the projected range and straggle of Ge implantation into Si, respectively.

NWI when a 45keV Ge implant is used. This will probably cause additional damages to the gate dielectric. Device wafers with different Ge-PAI energies but without ELA were fabricated. The gate breakdown voltages and leakage currents were measured. In Figure 3.14, it is seen that the gate breakdown voltages of NMOS and PMOS are smaller at higher implantation energies. However, the change is less than 4% and 6% for NMOS and PMOS respectively and are well within the specifications.

Figure 3.15 and 3.16 show the gate leakage currents measured in inversion and accumulation modes, respectively. When the PAI energy increases from 10keV to 45keV, for inversion mode, the leakage current increases by 33% and 63% for NMOS and PMOS, respectively, and for accumulation mode, the leakage current increases by 90% and 170% in NMOS and PMOS respectively.

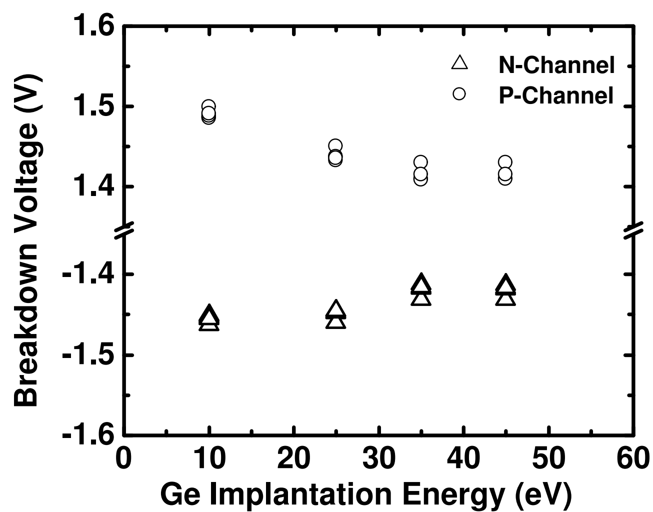

Figure 3.14. Gate dielectric breakdown voltage as a function of Ge PAI energies.

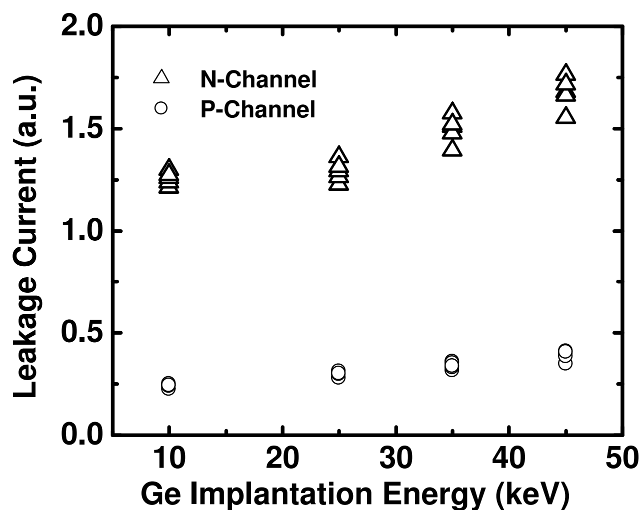

Figure 3.15. Inversion mode gate leakage current as a function of Ge PAI energies.

3.5.4 Ablation Induced by Flourine

A new ablation phenomenon was found when ELA was applied to BF_2 implanted PMOS gate. When low excimer laser fluence ($240mJ/cm^2$) was used to anneal blanket

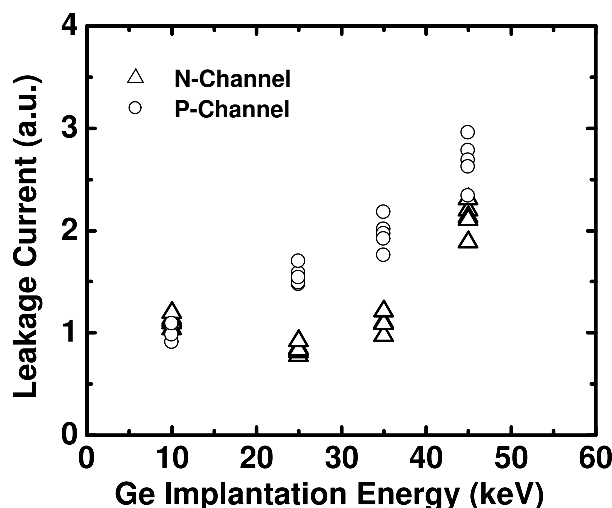

Figure 3.16. Accumulation mode gate leakage current as a function of Ge PAI energies.

(without STI) Ge-PAI gate (regardless of the PAI energy), ablation occurs and the film peeled off from the wafer. This does not happen to B and P doped wafers even at full-melt fluence. Figure 3.17 shows that ablation occurs only on BF_2 implanted region but not B implanted one, although they are close to each other and have the same area.

It is also found that it is easier for larger patterns to ablate. The ablation is due to the F atoms. During ELA, the highly concentrated F evolves explosively, similar to how H atoms act in a "smart cut" process [15]. The explosive F atoms exert an areal force on the pattern to ablate the Si film while the perimeter of the pattern, connected to the rest of the structure, tries to hold the film. The ratio of these two forces is thus proportional to $Area/Perimeter$. Therefore, large patterns, which have larger $Area/Perimeter$ ratio, ablate easier. Figure 3.18 shows that even at $500mJ/cm^2$, BF_2 implanted p-gate does not ablate when the pattern is in nanoscale. Therefore, BF_2 was used as PMOS gate implant in the rest of the experiment because F results

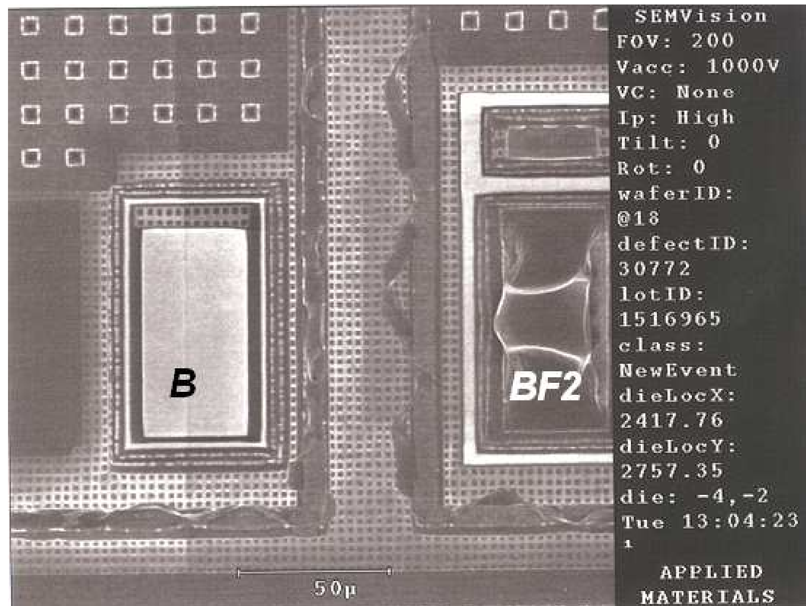

Figure 3.17. SEM showing the micro-scale active structure in SA-STI process after ELA at $717mJ/cm^2$. The left a-Si gate was implanted with B while the right was implanted with BF_2. Ablation is observed with BF_2 implanted structure only.

in less Negative Bias Temperature Instability (NBTI) and it gives a lower projected range.

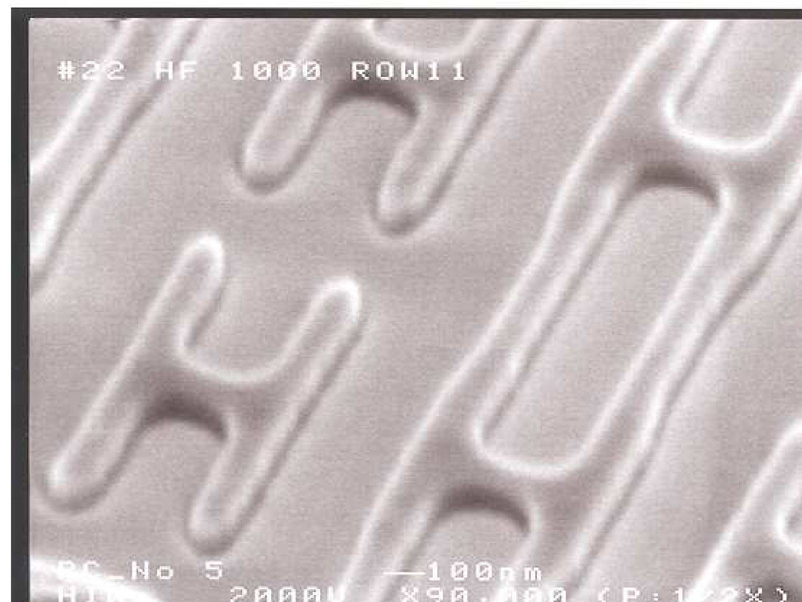

Figure 3.18. SEM showing the nano-scale active structure in SA-STI process after ELA at $500mJ/cm^2$. The structure are either implanted with BF_2 or P. No ablation is observed. The chosen laser fluence is larger than the full-melt threshold.

3.5.5 Device Results

Sub-100nm nMOSFETs and pMOSFETs were fabricated using ELA SA-STI process (Table 3.2). Ge PAI at 35keV or 45keV were performed and annealed at various laser fluences. Control sample (Reference) of Cypress SA-STI process was fabricated without ELA. However, the control sample also received 10keV Ge PAI as a default step (to prevent ion channeling during gate implantation) and has an additional gate annealing step by RTA. The gate implants are P and BF_2 for nMOSFETs and pMOSFETs, respectively. Although the mask set used is that of a 90nm SRAM process, the transistors are designed for the future 65nm technology. That means the gate dielectric, S/D dopings, gate, channel, well implants and interconnects are supposed to be well-designed for the minimum feature size transistor in 65nm SRAM technology.

Figure 3.19 shows the cross-sectional TEM picture of the active area of a SRAM cell after ELA at $309mJ/cm^2$. This sample received 45keV Ge PAI as the one shown in Figure 3.11. It is obvious that ELA has successfully re-crystallized the amorphized region. Therefore, $309mJ/cm^2$ is enough for re-crystallizing 45keV, and also 35keV, Ge PAI gates because the 35keV Ge PAI gate has shallower amorphization depth than the 45keV one.

Figures 3.20 and 3.21 show the measured $I_D - V_G$ curves of the minimum gate length transistors. In Section 3.5.3, it has been shown that when PAI is 35keV or 45keV, extra oxygen vacancies are generated in the gate dielectric and there is an increase of gate leakage current. Fortunately, this does not change the off state-leakage currents as shown by the curves.

However, it is found that the $I_D - V_G$ curves have shifted by about 90meV and -50meV in nMOSFETs and pMOSFETs, respectively. Since the shifts do not depend on the PAI energy, it should not be caused by Ge PAI or any degradation of the

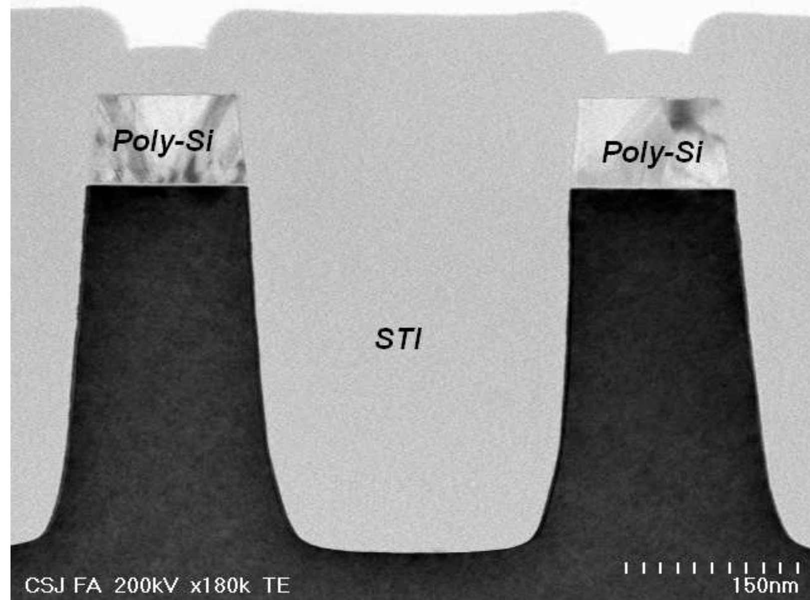

Figure 3.19. Cross-sectional TEM picture of active and isolation areas after Ge PAI at 45keV, LTO deposition and ELA at $309mJ/cm^2$.

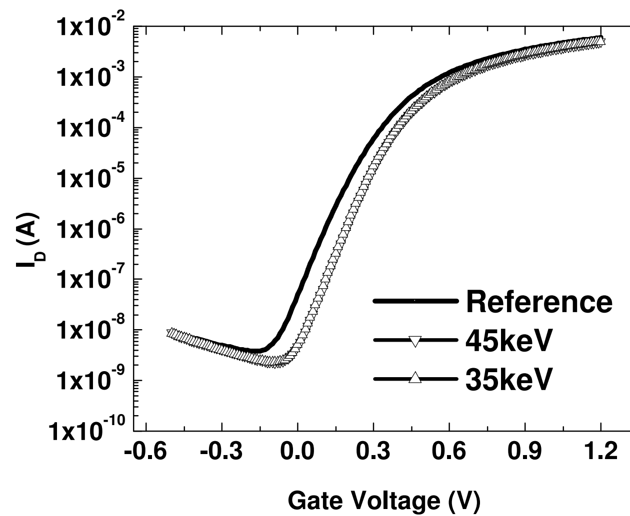

Figure 3.20. Measured $I_D - V_G$ curves ($V_D = 1.2V$) of ELA ($309mJ/cm^2$) SA-STI nMOSFET ($L_G = 90nm$) with Ge PAI at 35keV and 45keV. Reference is the ordinary Cypress SA-STI wafer with Ge-PAI at 10keV (no ELA) and poly-gate RTA.

dielectric. Moreover, the laser energy is large enough to melt the amorphized region but not the underlying poly-Si, the shift is not due to Fermi-level pinning after ELA. The last difference between the ELA and reference samples is that the reference has

a gate activation step using RTA. This is done after the channel and well implants. It is believe that this step further diffuses the dopants in the channel, resulting in lower effective channel dopant concentrations. Therefore, it has lower effective substrate work function Φ_S. According to Equations 1.1 and 1.2, the threshold voltage will be smaller. Since boron has larger diffusivity than phosphorus and arsenic, gate activation step RTA has also more profound effect on the nMOSFET (with boron doped substrate) than on the pMOSFET (with phosphorus or/and arsenic) doped substrate.

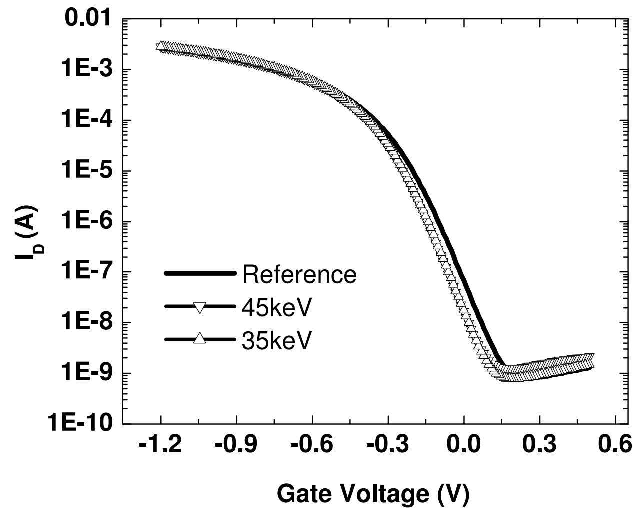

Figure 3.21. Measured $I_D - V_G$ curves ($V_D = -1.2V$) of ELA ($309mJ/cm^2$) SA-STI pMOSFET ($L_G = 90nm$) with Ge PAI at 35keV and 45keV. Reference is the ordinary Cypress SA-STI wafer with Ge-PAI at 10keV (no ELA) and poly-gate RTA.

In order to study the improvement of ON currents, the ELA sample curves are shifted to match the OFF current of the references at $V_G = 0V$. Linear plots of the curves of nMOSFETs and pMOSFETs are shown in Figures 3.22 and 3.23, respectively. For nMOSFETs, it is found that there is an improvement of I_D when $V_G < 0.83V$ and $V_G < 0.73V$ for 35keV and 45keV PAI, respectively. However, at

higher V_G, the ON current is degraded. Figure 3.24 shows that the degradation is 3.6% and 6.7% for 35keV and 45keV respectively at $V_D = 1.1V$.

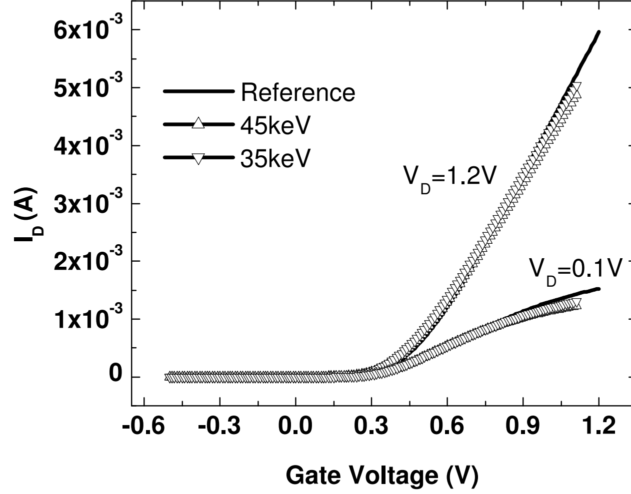

Figure 3.22. Measured $I_D - V_G$ curves of ELA ($309mJ/cm^2$) SA-STI NMOS ($L_G = 90nm$) with Ge PAI at 35keV and 45keV. Reference is the ordinary Cypress SA-STI wafer with Ge-PAI at 10keV (no ELA) but with poly-gate RTA. The ELA curves are shifted by -90meV to match the off current of the Reference.

For pMOSFETs, there are more than 8.8% and 11.1% current improvement for 35keV and 45keV in the saturation region, although the benefit reduces as V_G increases.

From the measured $I_D - V_G$ data, ELA probably has reduced the gate depletion effect. Since gate depletion effect is usually more serious in pMOSFETs, current improvement is thus more in pMOSFETs. However, as the V_G increases, the improvement is reduced and in nMOSFET case, there is even degradation of the ON current. This is probably due to PAI induced dielectric degradation. When the dielectric is degraded, surface scattering increases and, thus, mobility degrades at higher vertical electric field. Since 45keV generates more damages to the gate dielectric than 35keV, this can also explain why higher energy PAI gives less improvement (more degradation) of the ON current.

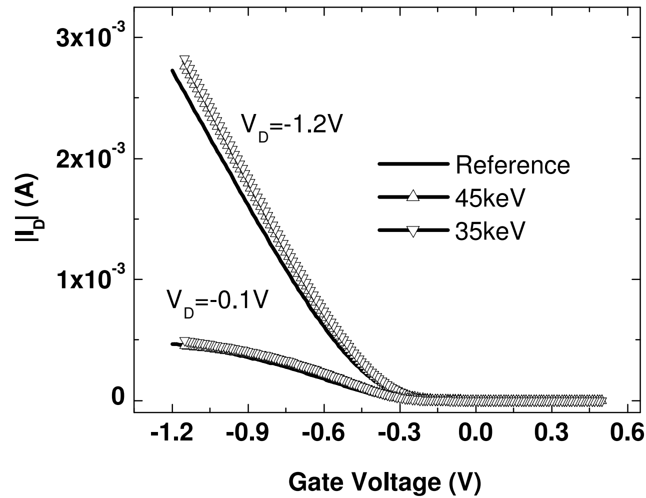

Figure 3.23. Measured $I_D - V_G$ curves of ELA ($309mJ/cm^2$) SA-STI PMOS ($L_G = 90nm$) with Ge PAI at 35keV and 45keV. Reference is the ordinary Cypress SA-STI wafer with Ge-PAI at 10keV (no ELA) but with poly-gate RTA. The ELA curves are shifted by 55meV to match the off current of the Reference.

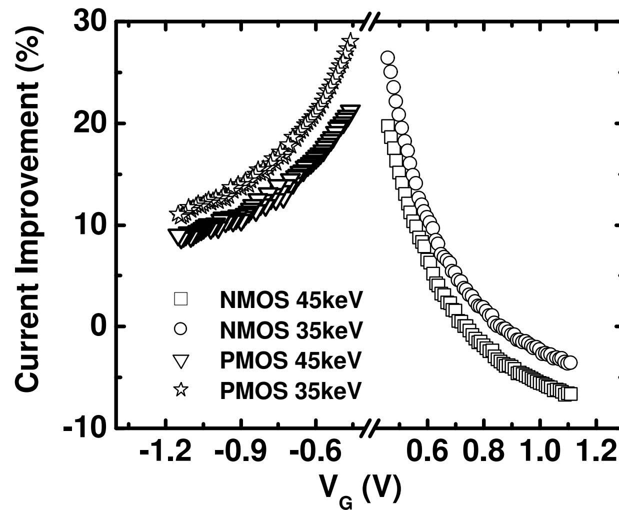

Figure 3.24. Current improvement of ELA ($309mJ/cm^2$) SA-STI NMOS and PMOS ($L_G = 90nm$) with Ge PAI at 35keV and 45keV over reference in saturation region ($|V_D|$=1.2V). I_{OFF} has been matched.

In order to prove this hypothesis, split CV curves and the mobilities have to be extracted on large area ($10\mu m \times 10\mu m$) transistors. However, it is found that the gate leakage current is too large for accurate split CV measurement (especially for nMOSFETs).

Another possible reason for the current degradation at high V_G is that the S/D contact resistance is higher in the ELA samples. At high V_G, the channel resistance is reduced and the S/D contact resistance dominates. Indeed, it is found that, for nMOSFETs, the measured S/D contact resistance is 19% higher in the ELA samples than in the RTA reference and 5.7% higher for pMOSFETs. The reason of contact resistance degradation may be due to Ge penetration into the S/D region during Ge PAI. This may change the optimal processing condition in S/D contact formation.

3.6 Stitching Issue

Although the laser spot is spatially uniform after passing the homogenizer (Figure 2.1), due to diffraction, the edges of the spot can have substantially lower energy. As a result, the amorphous-Si at the edges will be partially melted, trapping dopants in the poly-crystalline silicon (poly-Si) formed at the top of the electrode. Since poly-Si has a higher melting point than a-Si, even by stitching the laser beam spots, the partially crystallized area might not be fully melted in the second shot and there will be serious gate depletion effect at the edge of the spots. Therefore, to integrate ELA, at least one die has to be annealed in one shot. By doing so, the edges of the laser spots will be located at the scribe lines of the wafers, which is not used in the real products. Of course, in real production, the scribe line area between dies are used for fast automatic chip testing before wafer dicing. Therefore, excimer laser equipment which can deliver high energy and uniform large spot size to cover portion of the scribe line is required. Fortunately, nowadays, the technology is matured enough.

Guaranteed energy up to $15J$ per pulse can be delivered with pulse-to-pulse variation less than $\pm 2.5\%$. The associated beam homogenizer is able to provide a uniformity better than $\pm 3.5\%$ over a spot of $18cm^2$ with the energy density adjustable up to $650mJ/cm^2$ [16; 17]. According to ITRS'05, the maximum field area printed by exposure tools will be limited to $8.58cm^2$ until the end of the roadmap. Therefore, the maximum die area will not exceed $8.58cm^2$. This means that at least two dies can be annealed per laser shot and some parts of the scribe lines will be annealed under the same condition as the dies and can be used for electronic testing (Figure 3.25).

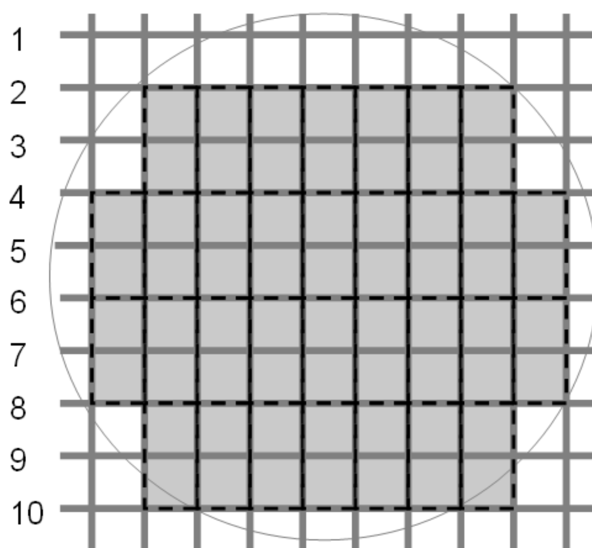

Figure 3.25. Schematic diagram showing how two dies can be annealed in each laser shot. Broken lines are the edges of the laser spots. Since scribe lines 3,5,7,9 are within the laser spots, the transistors on these scribe lines can still be used for automatic testing. However, all the vertical scribe lines cannot be used because they are at the edges of the laser spot.

3.7 Summary

In this Chapter, the integration issues of melt-ELA with traditional CMOS process have been discussed.

It is found that ELA is incompatible with RTA, which is commonly used for activating dopants in the transistor Source/Drain region. When RTA is used after ELA, there is significant dopant penetration through the gate oxide even after 1000^oC annealing for 1 second. Therefore, compatibility of ELA with FLA was studied due to the much reduced dopant diffusion in FLA. Experiment proved that ELA is compatible with FLA. The dopants activated by ELA are not deactivated and there is no significant boron penetration.

To solve the remaining integration issues, two integration experiments were performed. The first one is the ELA via masking process. In this process, absorption mask is used to prevent the a-Si on isolation region from melting and, thus, selectively annealing of active area is possible. It is proved that this method is effective in preventing the ablation of a-Si on isolation region and the intermixing of p+ and n+ dopants. It is scalable to 32nm technology node if extra masks and steps are used.

The second method is to integrate ELA with SA-STI process. Due to the nature of SA-STI process, a-Si ablation and dopant intermixing are not issues. A Ge PAI step is required to amorphize the gate electrode to reduce its melting point. However, the PAI step degrades the gate oxide qualities. Sub-100nm transistors were successfully fabricated with the ELA SA-STI process. The ON current of pMOSFETs is improved by 11.1%. However, that of nMOSFET is degraded by 3.6% at high vertical electric field due to increased surface scattering of the degraded oxide.

After the issues of integrating ELA with traditional CMOS processes have been identified, a new integration scheme is proposed, which is free from extra PAI steps, a-Si ablation and dopant intermixing, in the Future Work Section in Chapter 5.

References

[1] T. Ito, T. Iinuma, A. Murakoshi, H. Akutsu, K. Suguro, T. Arikado, K. Okumura, M. Yoshioka1, T. Owada1, Y. Imaoka, H. Murayama, and T. Kusuda, "10-15 nm ultrashallow junction formation by flash-lamp annealing," *Japanese Journal of Applied Physics*, vol. 41, pp. 2394–2398, 2002.

[2] K. Krisch, M. Green, F. Baumann, D. Brasen, L. Feldman, and L. Manchanda, "Thickness dependence of boron penetration through O_2 and N_2O-grown gate oxides and its impact on threshold voltage variation," *IEEE Transactions on Electron Devices*, vol. 43, pp. 982–990, June 1996.

[3] *The International Technology Roadmap for Semiconductors*, Semiconductor Industry Association, 2005.

[4] F. Ootsuka, H. Ozaki, T. Sasaki, K. Yamashita, H. Takada, N. Izumi, Y. Nakagawa, M. Hayashi, K. Kiyono, M. Yasuhira, and T. Arikado, "Ultra-low thermal budget CMOS process for 65nm-node low-operation-power applications," in *International Electron Devices Meeting. Technical Digest*, Dec. 2003, pp. 27.7.1–27.7.4.

[5] N. Acharya and P. Timans, "Fundamental issues in millisecond annealing," in *Advanced Short-Time Thermal Processing for Si-Based CMOS Devices II*, M. C. Ozturk, E. P. Gusev, L. J. Chen, D. L. Kwong, P. J. Timans, G. Miner, and F. Roozeboom, Eds. The Electrochemical Society, Inc., 2004, pp. 11–18.

[6] R. Lindsay, B. Pawlak, K. Henson, A. Satta, S. Severi, A. Lauwers, R. Surdeanu, S. Mccoy, J. Gelpey, X. Pages, and K. Maex, "Fundamental issues in millisecond annealing," in *Advanced Short-Time Thermal Processing for Si-Based CMOS Devices II*, M. C. Ozturk, E. P. Gusev, L. J. Chen, D. L. Kwong, P. J. Timans,

G. Miner, and F. Roozeboom, Eds. The Electrochemical Society, Inc., 2004, pp. 145–56.

[7] K. Adachi, K. Ohuchi, N. Aoki, H. Tsujii, T. Ito, H. Itokawa, K. Matsuo, K. Suguro, Y. Honguh, N. Tamaoki, K. Ishimaru, and H. Ishiuchi, "Issues and optimization of millisecond anneal process for 45 nm node and beyond," in *Symposium on VLSI Technology. Digest of Technical Papers*, 2005, pp. 142–43.

[8] T. Ito, K. Suguro, T. Itani, K. Nishinohara, K. Matsuo, and T. Saito, "Improvement of threshold voltage roll-off by ultra-shallow junction formed by flash lamp annealing." in *Symposium on VLSI Technology. Digest of Technical Papers*, 2003, pp. 53–4.

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

[10] U. Schwalke, M. Kerber, and K. K. H. J. Jacobs, "EXTIGATE: The ultimate process architecture for submicron CMOS technologies," *IEEE Transactions on Electron Devices*, vol. 44, pp. 2070–2077, Nov. 1997.

[11] U. Schwalke, German Patent DE 19 535 629C1, 1995.

[12] A. Blosse, K. Ramkumar, P. Gopalan, C. Hsu, S. Narayanan, G. Narasimhan, R. Gettle, R. Kapre, and S. Sharifzadeh, "A novel low cost 65nm CMOS process architecture with self aligned isolation and W cladded source/drain," in *International Electron Devices Meeting. Technical Digest*, Dec. 2004, pp. 669–672.

[13] The Stopping and Range of Ions in Matter (SRIM) Simulator. [Online]. Available: http://www.srim.org/

[14] *Taurus-Process*^{TM} *Reference Manual V-2003.12*, Synopsys, 2003.

[15] M. Bruel, B. Aspar, B. Charlet, C. Maleville, T. Poumeyrol, A. Soubie, A. Auberton-Herve, J. Lamure, T. Barge, F. Metral, and S. Trucchi, “”

[16] C. Prat, D. Zahorski, M. Stehle, J. Boulmer, D. Debarre, and G. Kerrien, “Excimer-laser activation of dopants in silicon : a high-energy excimer laser source for a single shot and uniform treatment over a whole die area.” in *9th International Conference on Advanced Thermal Processing of Semiconductors. RTP 2001*, 2001, pp. 146–7.

[17] C. Laviron, M. Semeria, D. Zahorski, M. Stehle, M. Hernandez, J. Boulmer, D. Debarre, and G. Kerrien, “Excimer-laser activation of dopants in silicon: a new concept for a uniform treatment over a whole die area,” in *Extended Abstracts of the Second International Workshop on Junction Technology, 2001*, 2001, pp. 91–94.

Chapter 4

Fully Nickel-Silicided (FUSI NiSi) Gate Technology

4.1 Introduction

As CMOS transistors are scaled to sub-32nm technology nodes, metallic gate electrode materials will be required [1]. This is because a metallic gate eliminates the gate depletion effect and thus reduces the equivalent oxide thickness (EOT) by 0.3-0.5nm [2]. Therefore, for the same EOT, a metal gate transistor has lower gate leakage current than a poly-silicon (poly-Si) gate transistor [3; 4]. Since metals have lower resistivity than poly-Si, the RC delay during transistor switching is also reduced. Moreover, unlike that of the poly-Si gate, the work functions (WF's) of metallic materials vary over a wide range and thus can be suitable for traditional bulk or partially-depleted silicon-on-insulator (PD-SOI) transistors (WF = 4.1-4.4eV for NMOS and 4.8-5.1eV for PMOS) [5] and novel transistor structures such as fully-depleted SOI (FD-SOI), FinFET and Tri-gate transistors (WF $\sim$ 4.7eV) [6]. It is

also expected that boron penetration can be avoided in metal gate technologies since high temperature annealing of boron doped Si-gate is obviated.

However, integration of metallic gate electrodes into CMOS process is not trivial. There are two major issues. Firstly, if a traditional CMOS flow is used, the gates are formed before the Source/Drain implantation and activation (gate-first process). There are concerns about contamination, dielectric integrity and metal etching processes. However, if gate-last process is used, where a dummy gate is removed and replaced after the major front-end processes, the removal of dummy gate can be difficult. Secondly, for optimized CMOS process, it is necessary to have symmetric threshold voltages (V_{TH}'s) and, thus, different metal WF's for nMOSFETs and pMOSFETs are required.

In the literature, various choices have been proposed. One of the approaches is using molybdenum (Mo) gate followed by nitrogen implantation and annealing to change the WF [7]. Another approach is using selective metal alloy. In this gate-last approach, Ti gate is used in nMOSFETs while Ni is selectively deposited to inter-diffuse with Ti in pMOSFETs [8]. Stacked metal layers as gates have also been demonstrated by changing the thickness of the bottom metal layer. The WF changes from that of the bottom metal to the top one as the bottom layer thickness decreases [9]. However, these technologies are not readily integratable with CMOS process because either metals have to be deposited before the high temperature source/drain annealing or gate-last process is required.

Fully silicided nickel silicide (FUSI NiSi) is another promising metallic gate technology, which differentiates itself from the others by the fact that, the metallic gate is formed (metal deposition and silicidation) after the high temperature front-end-of-line (FEOL) processes [10; 11; 4; 12]. In a FUSI process, a conventional CMOS FEOL is used, then the doped poly-Si gate is then fully silicided to form NiSi before the

back-end-of-line (BEOL) process. Therefore, it poses no major integration challenges as other metal gate technologies do, due to the much lower thermal budget in the BEOL than that in the FEOL processes.

In terms of the realization of dual WF's on the same wafer, it is also relatively easier in FUSI NiSi than other metal gate technologies. The WF can be adjusted by implanting appropriate impurities into the poly-Si gate before Ni deposition. During silicidation, dopants in the poly-Si gate segregate to the silicidation front and pile up at the electrode / dielectric interface, resulting in a shift of effective WF [12; 13; 14]. The direction and the amount of the shift depend on the type and the amount of the piled-up dopant, respectively [12]. Actually, as early as 1984, Kakumu et al. have already proposed the use of dopant segregation at the silicide/dielectric interface to control the effective WF of the silicide gate and this was demonstrated on a 64kb full CMOS RAM [15]. At that time, Mo_2Si was used. It was not until this century that FUSI gate has regained attention and, in 2001, B. Tavel et al. published the first mid-gap $CoSi_2$ metal gates by total gate silicidation [16]. In the same year, M. Qin et al. first studied the adjustment of WF through dopant implantation in FUSI NiSi gate [10]. They were able to adjust the FUSI NiSi WF from 4.6eV to 5eV. Then W.P. Maszara et al. were able to achieve a tunable range of FUSI NiSi between 4.5eV and 4.9eV [5]. They showed that the poly-gate depletion effect is reduced by 2.5$\AA$ and observed pile-up of B and As at the electrode / dielectric interface. Independently, J. Kedzierski et al. fabricated FUSI NiSi gate FinFET and FDSOI devices [17] and they extended the studies to include Al, As and Sb dopants in addition to P and B. They further confirmed that the WF change is due to dopant pile-up [12].

Owning to aforementioned characteristics of FUSI NiSi gate, the adjustable WF range, the dopant pile-up efficiency and mechanisms, and the thermal stability during the BEOL processes of the FUSI NiSi gate have become the important research topics. In this Chapter, the adjustable WF range of FUSI NiSi gate doped with boron and

phosphorus is first presented. Since boron doped FUSI NiSi is found to have much more profound effect on the WF tuning than phosphorus, it is then studied in more detail. The importance of boron activation and initial dopant profile on the WF tuning range is studied with the assistance of the excimer laser annealing tool. Finally, the reliability and thermal stability of FUSI NiSi gate is discussed.

4.2 Experiment

In order to study FUSI NiSi gate properties, CMOS capacitors were fabricated. 4" (100) p-type or n-type wafers were used as the starting wafers. 300nm-thick LOCOS (LOCal Oxidation of Silicon) was formed as isolation. After gate oxide (SiO_2) formation (1.8nm to 20nm), poly-Si or amorphous silicon (a-Si) (doped or in-situ doped) was deposited using Low Pressure Chemical Vapor Deposition (LPCVD). The undoped samples were then implanted with ultra low energy ions (2keV for phosphorus and 1keV for boron) at different doses (4×10^{14} to 2.4×10^{15} cm^{-3}). After gate patterning, various thermal treatments were performed (Excimer Laser Annealing (ELA), Solid Phase Crystallization (SPC) or Rapid Thermal Annealing (RTA)). Native oxide on Si gate was removed with 25:1 HF solution, followed by degas (150^oC for 5 minutes), and nickel (Ni) was sputtered and capped with titanium nitride (TiN). Silicidation was carried out with RTA (450°C to 560°C) for 60 seconds. TiN and unreacted Ni was removed by hot SPM solutions (H_2O_2:H_2SO_4=1:4). Forming gas annealing (FGA) was finally performed at 400°C for 20 minutes.

For some wafers, LOCOS was not used. Therefore, High-Temperature-Oxide (HTO)/ Nitride double spacer was used to prevent the gate from shorting to the substrate after FUSI. Figure 4.1 shows the schematic diagrams of the process flow.

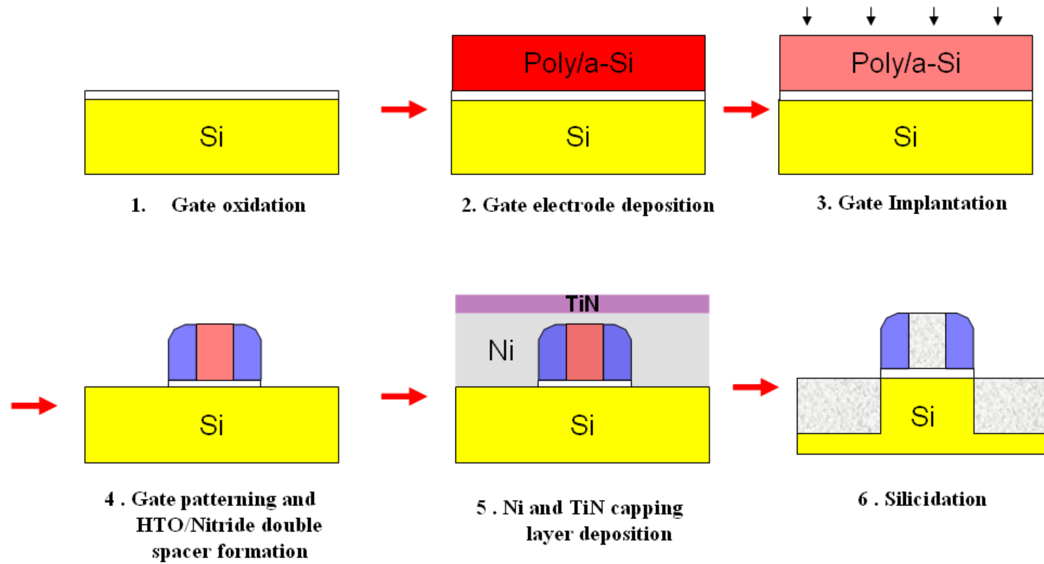

Figure 4.1. Schematic diagrams showing the process flow of fabricating FUSI NiSi capacitor structures. When LOCOS is used, step 4 is skipped.

4.3 Adjustable WF Range of FUSI NiSi

4.3.1 Full Silicidation Issues

There are many different phases in the Ni-Si system. Among them, nickel-rich phase (Ni_2Si), nickel mono-silicide (NiSi) and nickel disilicide ($NiSi_2$) are the most commonly used in the semiconductor industry. Ni_2Si is formed below 400^oC while NiSi is formed between 400^oC and 600^oC. NiSi has very low resistivity of about $14 - 20\mu\Omega \cdot cm$. But it is unstable at temperature above 650^oC because of the conversion to $NiSi_2$. According to Ottaviani [18], when Ni reacts with Si, Ni_2Si will form first until the whole Ni film is consumed. After that, Ni will diffuse and react with the remaining Si to form NiSi. The thicknesses of Ni required to consume per unit of Si is 1.10, 0.55 and 0.27 for Ni_2Si, NiSi and $NiSi_2$ respectively [19]. Therefore, to form FUSI NiSi, the annealing temperature has to be between 400^oC and 600^oC

and Ni to Si ratio should be larger than 0.55 but less than 1.10 so that the metal is NiSi at the region close to the dielectric.

However, although the formation temperature of NiSi is low, it is not trivial to fully silicide a poly-Si gate. Figure 4.2 shows that silicidation can be incomplete even when the thermal budget and the deposited Ni thickness are expected to be sufficient to fully silicide the 150nm poly-Si film. One possible reason is that there is a high oxygen content in a LPCVD Si film due to the leakage of the furnace. The O concentration is usually larger than $5 \times 10^{17} cm^{-3}$ [20], and it can be even as high as $10^{19} cm^{-3}$ in some films produced by the furnace used in our studies (Figure 2.5). As the silicidation front progresses, due to the much lower solubility of oxygen in NiSi than Si, oxygen atoms segregate into the poly-Si. Eventually the oxygen concentration is so high that silicon sub-oxide forms and the silicidation is retarded [21]. Since it is easier for the oxygen to diffuse through and form silicon sub-oxide at grain boundaries, the silicidation front, thus, is non-uniform and affected by the grain size.

Therefore, when the process thermal budget is limited, the poly-Si must be thin enough so that the film can be fully silicided. In order to determine the full silicidation conditions, NiSi gate capacitors with 2.3nm oxide were fabricated using the process described in Section 4.2. The measured C-V curves are shown in Figure 4.3. It is found that for undoped NiSi gate, the FUSI condition can be deduced by tracking the maximum accumulation capacitance. However, if it is doped before silicidation, FUSI condition can only be accurately identified by measuring the flat-band voltage (V_{FB}). This is because if the dopants are activated and the gate is not fully silicided, the capacitor can be regarded as a parallel combination of FUSI NiSi gate and poly-Si gate capacitors (Figure 4.2). As a result, high accumulation capacitance can still be obtained without full gate silicidation.

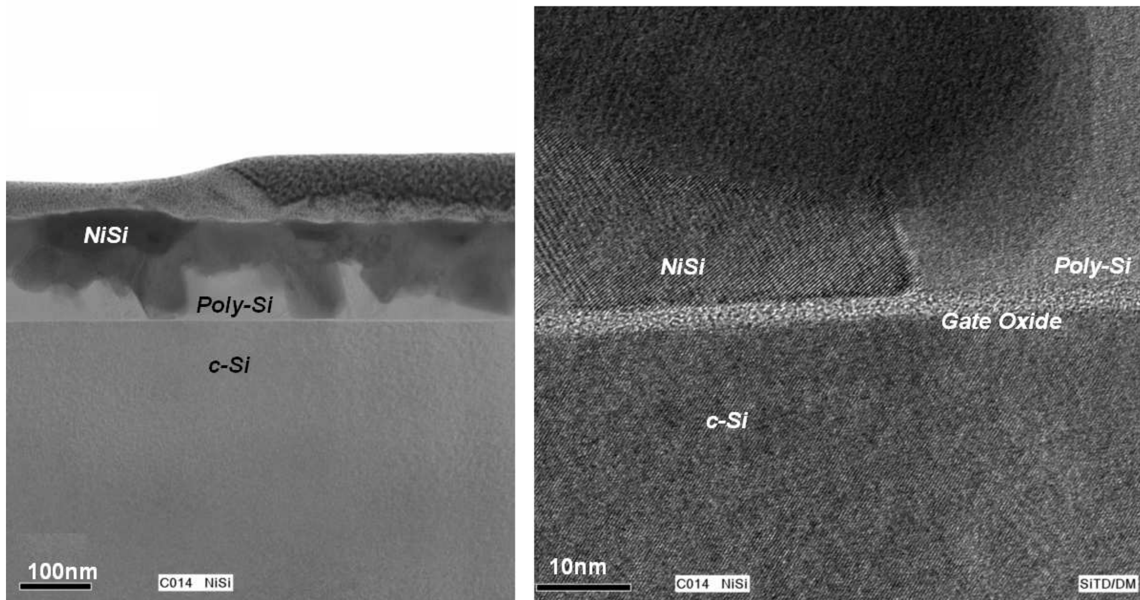

Figure 4.2. Cross-sectional TEM micrographs (different magnifications) of NiSi gate formed with a 450°C 90 s RTA. Initial poly-Si thickness is 150nm and Ni thickness is 90nm capped with 30nm TiN. Most of the area is not full silicided. In some regions, less than 60nm of NiSi is formed.

Indeed, if the doped but not fully-silicided capacitors are used for WF extraction, a misleadingly large tunable range of work functions is obtained (Figure 4.4).

From the above results, it was found that a 530°C 60 s RTA is required to fully silicide a 50nm poly-Si film. This silicidation condition was therefore used to fabricate FUSI NiSi gate capacitors with different oxide thicknesses for work function extraction.

4.3.2 Effective Gate Work Function Extraction

To determine the effective gate WF (Φ_{Gate}) of FUSI NiSi, $\sim 25nm$ thermal oxide (SiO_2) was formed and selectively time-etched back in 25:1 HF solution, so that a range of gate oxide thicknesses is realized on the same wafer [10], followed by the general process described in Section 4.2 to form capacitors. HTO/Nitride double spacer was also formed to prevent the gate from shorting with the Si substrate. The

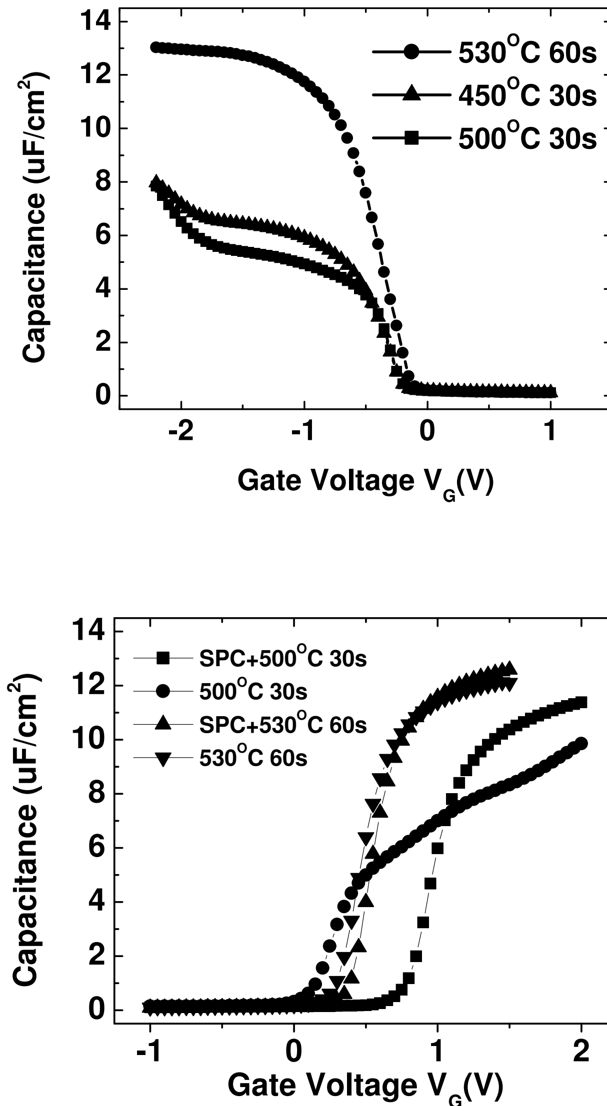

Figure 4.3. Measured $C - V$ characteristics of NiSi gate capacitors (T_{ox}=2.3nm) with different annealing conditions. The 50nm poly-Si was either undoped (upper) or doped with boron ($2.4 \times 10^{15} cm^{-2}$) (lower). 530°C for 60s is sufficient to fully silicide the poly-Si and therefore, with this annealing condition, undoped FUSI gate has the maximum accumulation capacitance and boron-doped FUSI has both the correct WF and maximum accumulation capacitance. SPC is 800^oC for 4 hours, which was used to activate the dopants in poly-Si before Ni deposition.

thermal budget of the spacer formation is equivalent to 800^oC furnace annealing for 4 hours.

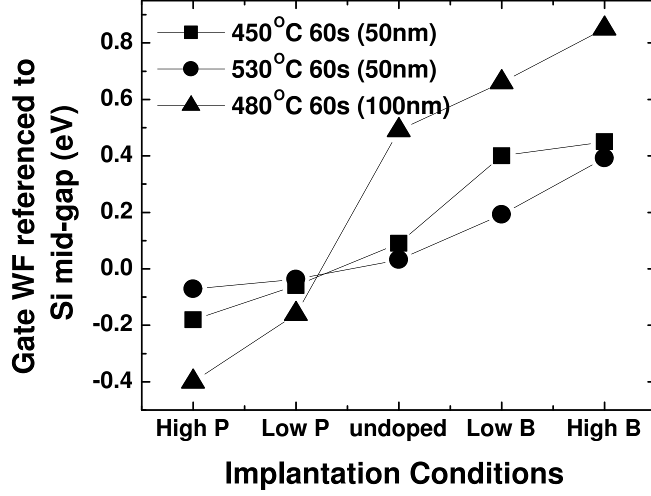

Figure 4.4. Extracted gate work function relative to Si mid-gap, for NiSi gate with different implantation and RTA conditions. The initial poly-Si gate thickness is either 50nm or 100nm. For 100nm poly- Si, even with enough Ni (60nm), the gate is not fully silicided, so that the apparent tunable work function range is unreasonably large. (High=$7 \times 10^{20} cm^{-3}$, Low=$1 \times 10^{20} cm^{-3}$)

High frequency (100kHz) capacitance-*vs*-voltage (CV) curves were taken for different oxide thicknesses and the V_{FB} was determined [22]. Then Equation 2.1, which is repeated here for convenience, was used to extract the effective gate WF,

$$V_{FB} = \frac{1}{q}(\Phi_{Gate} - \Phi_{sub}) - \frac{Q_f}{\epsilon_{ox}} t_{ox} \tag{4.1}$$

where q is the unit charge ($1.6 \times 10^{-19} C$), Φ_{Gate} and Φ_{sub} are the WF's of the silicon gate and substrate respectively (in eV), Q_f is the fixed charge density in C/cm^2, t_{ox} is the physical oxide thickness in cm and ϵ_{ox} is the SiO_2 dielectric permittivity ($3.45 \times 10^{-13} F cm^{-1}$).

By plotting V_{FB} against t_{ox}, after least squares fitting, one can determine the Φ_{Gate} and Q_f from the y-intercept and slope of the fitting line respectively. Figure 4.5 shows the results of the experiment with in-situ doped n+ p-Si gate as reference.

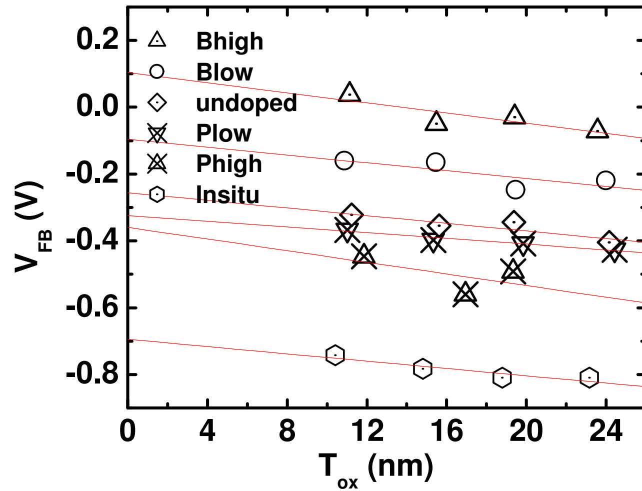

Figure 4.5. Effective gate WF extraction of FUSI NiSi using least squares fitting of V_{FB} against t_{ox} plots. P-type substrate was used. B - Boron, P - Phosphorus, high - $2.4 \times 10^{15} cm^{-3}$, low - $4 \times 10^{14} cm^{-3}$, Insitu - in-situ P doped poly-Si.

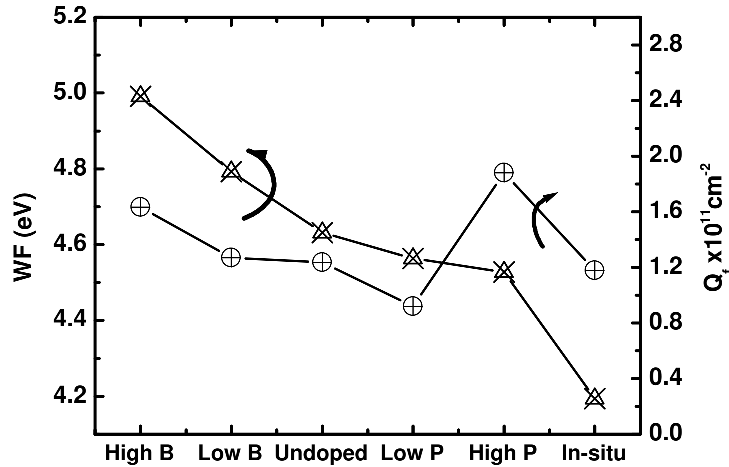

Figure 4.6. The WF and Q_f of FUSI NiSi gate under various impurity implantation conditions. B - Boron, P - Phosphorus, high - $2.4 \times 10^{15} cm^{-3}$, low - $4 \times 10^{14} cm^{-3}$, Insitu - in-situ P doped poly-Si.

The WF and Q_f of the samples are shown in Figure 4.6. It is found that the WF of the undoped FUSI NiSi gate is 4.63eV and very closed to the Si mid-gap (4.6eV). By

implanting B or P before the silicidation, the WF can be tuned towards the valence band and the conduction band edges by 0.36eV and 0.10eV respectively, which is enough for $L_G = 10nm$ FinFET application [23]. The amount of shift also depends on the implantation dose. The fixed charge density of the undoped FUSI NiSi gate is lower than the in-situ doped one, meaning that FUSI NiSi does not degrade the SiO_2 in terms of fixed charge density. The large fixed charge of high dose implanted B and P is possibly due the ion implantation damage. Therefore, it is preferred to use even lower implantation energy to reduce ion straggle.

4.3.3 Summary

Table 4.1 summarizes the tunable FUSI NiSi WF ranges from the literature and this work. Our work and many other works show that the WF of undoped NiSi is close to the mid-gap and, with B, the WF can be tuned to be close to the valence band edge (5eV). However, the tunability with P is limited and the lowest WF can be achieved is about 4.5eV.

If As is used, it is possible to turn the WF to be close to 4.4eV. However, it is found that there is adhesion problem with As-doped NiSi film after Forming Gas Annealing (FGA) (using the process described in Section 4.2 with $4 \times 10^{15} cm^{-2}$ As implanted at 2keV). While an almost ideal CV curve of As-doped NiSi can be obtained after FUSI process, the film detached and no measurement could be made after forming gas annealing at 400^oC even for just 5 mins. It is believed that the hydrogen gas in FGA reacts with the piled-up As at the gate electrode / dielectric interface to form high vapor pressure As-H compounds (for example, vapor pressure of Arsine (AsH_3) is already larger than 760mmHg at room temperature). And since the film was annealed at 400^oC, it can be peeled off easily. Adhesion problem of As doped FUSI NiSi was also found in another laboratory. Therefore, if As is to be used, the

reliability issue has to be addressed. For Sb, although it may help to extend the range further towards the conduction band edge, the variation is too large and might offset its usefulness in WF tuning [12].

NiSi WF						Fabrication Conditions		
	B	undoped	P	As	Sb	Temp	Si	Ni
This Work	4.99 2.4E15, 1	4.63	4.53 2.4E15, 2			530 1min	50	30
[10]	5 5E15, 40	4.9	4.6 5E15, 40			500 1min	115	65
[5]	4.9			4.5 4E15		NA	70	40
[11]	5.1 3E15, 5	4.87		4.58 2E15, 40		500 20sec	80	60
[12]	4.76 2E15	4.66	4.49 2E15	4.38 3E15	4.24 3E15	450	< 48	
[24]	4.99 5E15	4.67	4.47 5E15	4.39 5E15			190nm	

Table 4.1. Table summarizing the tunable WF range (in eV) of FUSI NiSi gate and the fabrication conditions. Second row of each WF entry shows the implant dose (cm^{-2}) and energy (keV). Thickness is in nm. Temperature is in ^{o}C.

4.4 Dopant Pile Up

Several pioneering studies have shown that there is a close relationship between the dopant pile-up and the effective WF shift in FUSI NiSi gate [5; 12]. This was further confirmed by J. H. Sim et al.'s experiment [25]. In their experiment, Ni/Si gate stack was capped by Titanium Nitride (TiN) layer before silicidation. It is found that boron tends to segregate into the TiN layer and results in less dopant pile-up at the electrode/gate dielectric interface, which then corresponds to smaller measured flat band voltage shift. Since boron penetration into the gate dielectric aggravates

the Negative-Biased-Temperature-Instability (NBTI), NiSi capped with TiN is thus expected to have better NBTI as shown in [26].

In most of the published experiments, rapid-thermal-annealing was performed after gate implantation. In our experiment, we also found that if FUSI NiSi is formed on unannealed implanted Si sample, the shift of WF is less compared to the annealed one (Figure 4.3). It has been shown that there is more As segregation if the poly-Si gate is annealed at higher temperature after implantation [27]. K. Hosaka et al. also showed that extra WF shift can be achieved if higher implantation energy is used for As or extra annealing for B doped poly-Si [14]. Therefore, it is believed that annealing is required to increase the pile-up of the dopants and thus increase the shift of the WF. However, it is not clear in most cases whether the extra WF shift after annealing is due to dopant activation or the diffusion of dopants towards the interface. Although D. Aime et al. showed that dopant distribution and location (substitutional or interstitial) are at the first order of minor importance [24], their poly-Si is 190nm, which is too thick for future CMOS technologies.

Therefore, in this study, we try to decouple the impacts of boron dopant location and activation on the WF adjustment.

4.4.1 Experiment

MOS capacitors were fabricated on 4" p-type ($\sim 10^{15} cm^{-3}$) or n-type ($\sim 5 \times 10^{14} cm^{-3}$) wafers. There were no channel or well implantations in order to keep the substrate uniformly doped so as to have an accurate estimation of the substrate WF. After 300nm-thick LOCOS formation, thermal silicon dioxide of 1.6nm or 2.2nm was grown followed by LPCVD of either 55nm undoped (at $530^{o}C$) or 50nm in-situ boron doped (at $450^{o}C$) amorphous Si (a-Si). The undoped a-Si then received high dose ($4 \times 10^{15} cm^{-2}$) low energy implantation of boron (1keV) or phosphorus (2keV),

all with implantation range, R_P, < 6nm. After gate patterning, different thermal treatments, including Solid-Phase Crystallization (SPC) at 580^oC for 10 hours, Rapid Thermal Annealing (RTA) at 1020^oC for 5 seconds, or melt pulsed Excimer Laser Annealing (ELA) (pulse width of ~30ns), were used to activate the dopants. Wafers were then dipped into 25:1 HF solution for 30 seconds and degassed at 150^oC for 5 mins before the sputtering of 35nm nickel. Silicidation was carried in N_2 ambient for 60 seconds at different temperatures and unreacted Ni was removed by fresh SPM solution (H_2SO_4+H_2O_2). Finally, wafers were forming gas annealed (FGA) at 400^oC for 20 minutes.

4.4.2 Effect of Dopant Activation

In order to study the role of dopant activation on the WF shift in FUSI NiSi gate, wafers with in-situ boron doped a-Si gate were used. P-type substrates were was used so that gate depletion, if it exists, can be observed under substrate accumulation mode. Since constant Si_2H_6 and B_2H_6 flow rates were maintained during the deposition, the concentration of boron is expected to be constant throughout the silicon gate. Pulsed excimer laser with different fluences were then applied to the gate. The set up of the excimer laser tool is shown in Figure 2.1. To prevent the ablation of a-Si on top of LOCOS area, selective annealing scheme (Section 3.4) was used and Figure 4.7 shows the cross sectional view of the structure during laser annealing. Due to the extreme high power of the laser pulses ($\sim 10^7$W/cm^2), a-Si is partially melted and recrystallized with most dopants in the melted region activated (substitutional). But as the process is completed within ~100ns [28], dopants in the non-melt regions remain inactivated.

Figure 4.8 shows the sheet resistance of the a-Si at different laser fluences. The sheet resistance becomes measurable at $300mJ/cm^2$. After $600mJ/cm^2$, the sheet

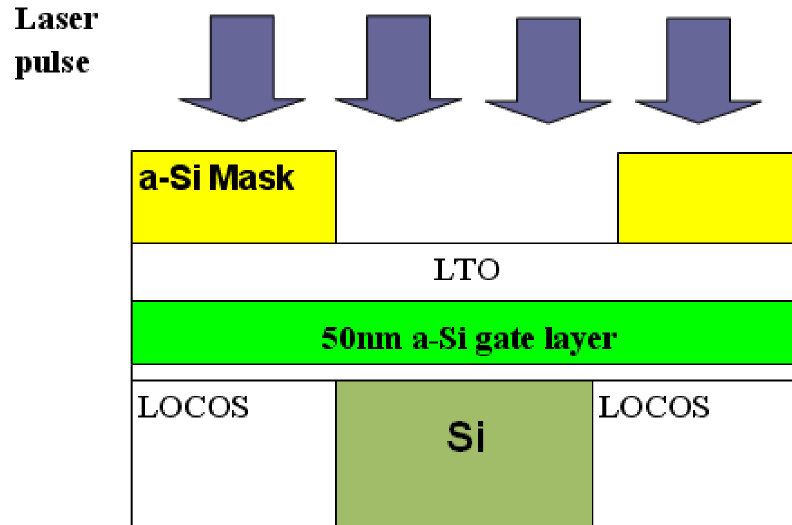

Figure 4.7. Cross-sectional diagram showing the laser annealing process of a-Si gate before patterning. 50nm Low Temperature Oxide (LTO) (450°C) was used to cap the a-Si during annealing as anti-reflection coating. A layer of 150nm a-Si layer on top of LTO was deposited at 450°C and patterned to prevent the ablation of a-Si gate during ELA.

resistance becomes constant. Therefore, it can be deduced that the melting threshold and full melt fluence are around $300mJ/cm^2$ and $600mJ/cm^2$ respectively. This is further confirmed by the infrared (IR) laser reflectance measurement on the gate surface during ELA processing. As shown in Figure 4.9, the reflectance increases at $300mJ/cm^2$ and the melt duration changes abruptly after $600mJ/cm^2$ [29].

The melt depth at different fluences thus can be estimated using the measured sheet resistance because the resistivity of the re-crystallized film can be assumed to be constant (Figure 4.8). After the removal of the masking a-Si and LTO layers, the capacitor gate electrodes were formed using the FUSI NiSi process outlined in Section 4.4.1.

Figure 4.10 shows the measured CV curves of the samples after Ni silicidation at 545^oC for 60 seconds. Quantum CV simulator [30] was then used to fit the curves and extract the flat band voltages (V_{FB}). It is found that the simulation and measured

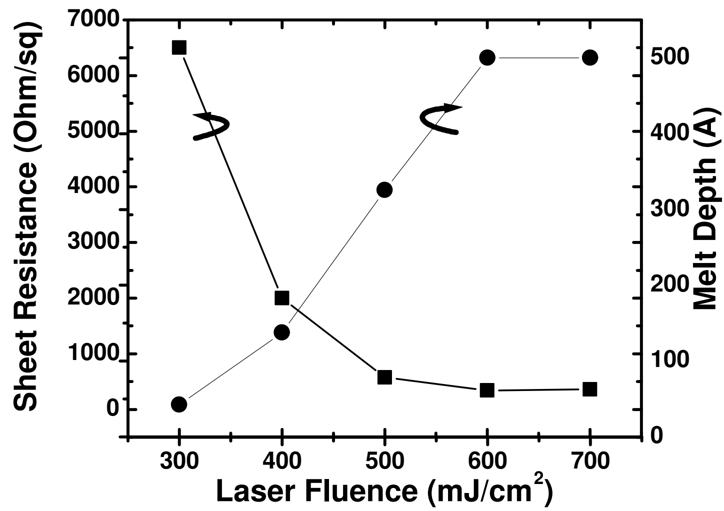

Figure 4.8. Sheet resistance ($R_\Box$) of insitu boron doped silicon film (50nm) after ELA at different fluences. (left axis). $R_\Box$ does not change after $600mJ/cm^2$. This corresponds to the full melt fluence. Thus the melt depth at lower fluences can be calculated by using the corresponding $R_\Box$ (right axis).

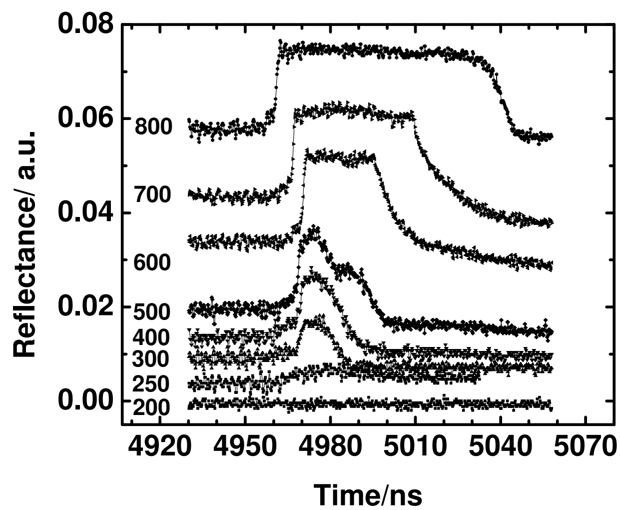

Figure 4.9. Measured reflected infrared (IR) density during ELA at different laser fluences. Since the reflectance of molten silicon increases substantially, the peaks show the moment Si gate is melted. $300mJ/cm^2$ is the melt threshold. Curves are shifted vertically for clearer presentation.

data fit every well and there is no observable gate depletion effect in the measured data even at the gate voltage of -2.5V. Therefore, it can be concluded that the gate is fully silicided. Since the fixed charge contribution to the V_{FB} can be neglected for very thin oxide and ELA does not increase the oxide fixed charge [31], the WF can then be calculated. Figure 4.10 shows that more WF shift is achieved at higher laser fluences. In this experiment, the total number and distribution (uniform) of dopants are identical in the laser-annealed in-situ boron doped gates. The grain sizes are also expected to be similar at different fluences before full melt threshold ($600mJ/cm^2$) [29]. The only difference is that they have different amount of activated dopants. Therefore, dopant activation is important to increase the WF in boron doped FUSI gate. Figure 4.11 plots the WF of the FUSI gate vs. the percentage of activated dopants.

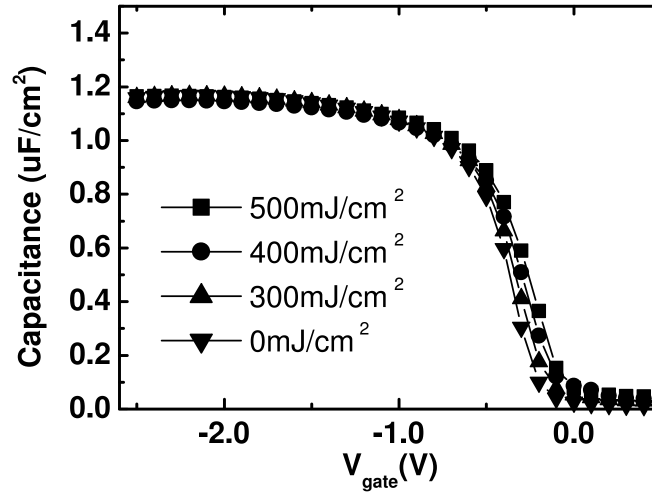

Figure 4.10. C-V curves of laser annealed in-situ boron doped sample after FUSI (tox=2.2nm). There is no gate depletion, which indicates that the gates are fully silicided. Higher laser fluence translates to deeper melt depth and higher percentage of dopant activation.

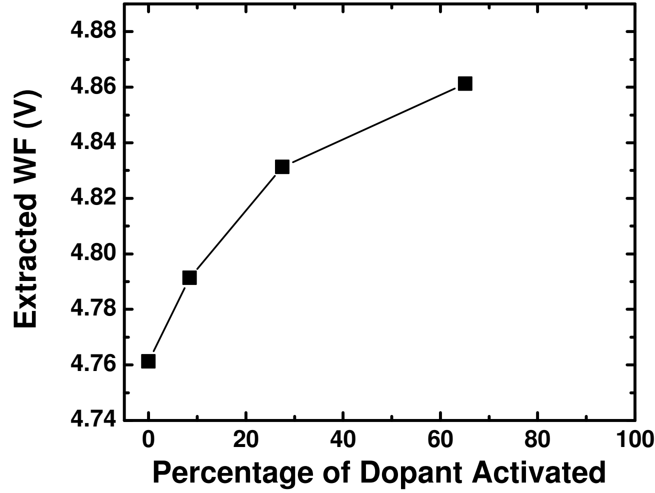

Figure 4.11. Extracted WF of laser annealed in-situ boron doped sample after FUSI versus the percentage of dopant activation. The percentage is calculated by using the melt depth found in Figure 4.8

4.4.3 Effect of Initial Dopant Distribution

To understand the dependence of NiSi FUSI gate WF shift on the initial dopant distribution, wafers with 1.6nm thermal oxide were used. Un-doped a-Si was deposited and implanted with boron (1keV, 4×10^{15}cm^{-2}). The wafers were annealed with pulsed excimer laser at different fluences. Figure 4.12 shows that the sheet resistance becomes constant after 350mJ/cm^2. However, unlike the in-situ-doped case, this does not correspond to full melt fluence. This is because, in Si, the implanted boron has a Gaussian distribution with the projected range, R_P, and straggle, ΔR_p, equal to 5.6nm and 4.5nm respectively. Once the melt depth is large enough to melt the layer of the a-Si containing most of the dopants, similar sheet resistances will be obtained regardless of the melt depth. To have 98% of the dopants activated, this corresponds to a melt depth $> R_P + 2\Delta R_p = 14.6nm$.

Figure 4.13 shows the CV curves of the capacitors after ELA and FGA but before any Ni deposition. As the laser fluence increases, the V$_{FB}$ and the accumulation

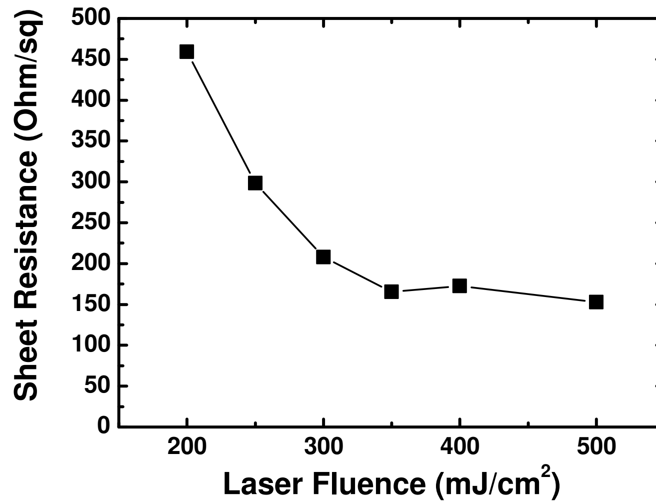

Figure 4.12. Sheet resistance (R_S) of boron implanted (1keV, 4×10^{15}cm^{-2}) silicon film (55nm) after ELA.

capacitance increases. This is because the deeper the gate is melted, the closer the dopants would diffuse to the electrode / dielectric interface and become activated.

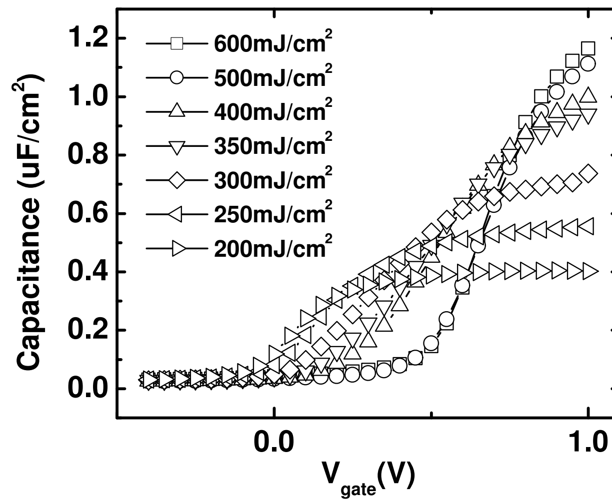

Figure 4.13. Measured C-V characteristics of boron-implanted capacitors after ELA and FGA.

At 400mJ/cm^2, the small accumulation mode capacitance and low flat band volt-

age indicates that it is still far from full melt condition. At 500mJ/cm^2, the gate is still not fully melted, because it has smaller accumulation mode capacitance than 600mJ/cm^2. At 600mJ/cm^2, the gate is fully melted and it fit very well with Quantum CV simulation. Therefore, for the samples laser annealed at fluences between 400mJ/cm^2 to 600mJ/cm^2, they have similar sheet resistance (Figure 4.12) and, thus, almost identical number of activated dopants. However, they have different melt depths and thus have different dopant distribution in the gate. The centroid of the activated dopants is closest to the electrode / dielectric interface after 600mJ/cm^2 annealing while that of 400mJ/cm^2 is farthest away.

Laser annealed samples that did not receive FGA were used for FUSI NiSi gate formation using the process outlined in Section 4.4.1. Figure 4.14 shows the measured CV curves of the FUSI NiSi capacitors. It shows that the sample annealed at 600mJ/cm^2 gives the most shift of the curve. Therefore, it can be concluded that as the activated dopants get closer to the electrode / dielectric interface, the WF shift increases.

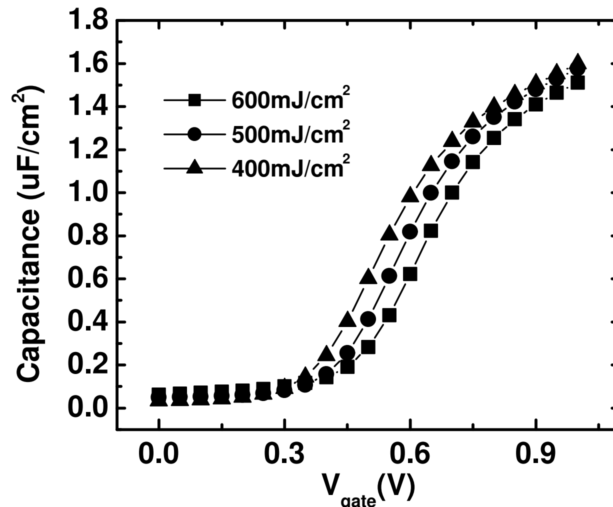

Figure 4.14. Measured C-V characteristics of boron implanted capacitors after ELA and FUSI.

4.4.4 Summary

From the previous results, it can be deduced that both dopant activation and initial dopant position affect the final tunable range of FUSI NiSi WF. The more the dopants are activated and the closer the initial dopant position is to the gate electrode/ dielectric interface, the greater the degree to which the WF can be adjusted. This is consistent with the results given in [27] and [14]. Therefore, it is important to choose the right implantation and annealing conditions before FUSI formation, so that maximum tunable WF range can be achieved.

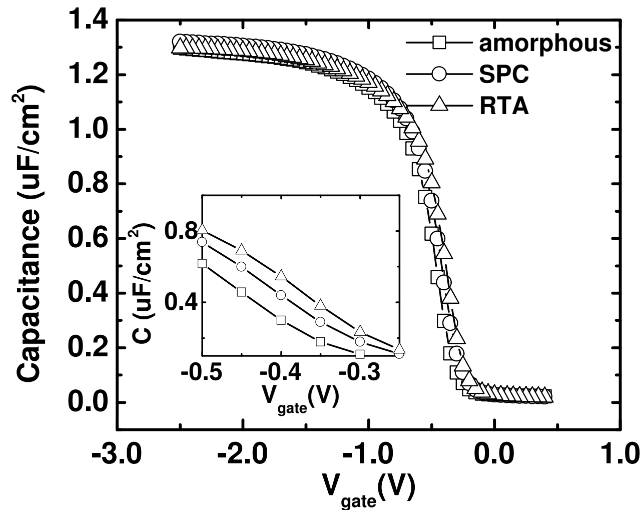

Figure 4.15. Measured C-V characteristics of boron-implanted FUSI NiSi gates. The a-Si gate has gone through either no thermal treatment (amorphous), SPC or RTA before Ni deposition.

Figure 4.15 and Figure 4.16 show the CV curves of boron and phosphorus doped NiSi FUSI gates. Undoped amorphous Si was deposited as the gate material and implanted with B or P. Different thermal annealing was performed as described in Section 4.4.1. For B doped NiSi, it is found that RTA sample has more WF shift (76meV) than SPC sample (47meV) with reference to the amorphous Si sample. While for P doped NiSi, the shifts with respect to amorphous sample is about -

49meV and -37meV for RTA and SPC respectively. The fact that the RTA samples have larger WF shift is again consistent with the previous results because RTA brings the activated dopants closer to the electrode / dielectric interface before silicidation.

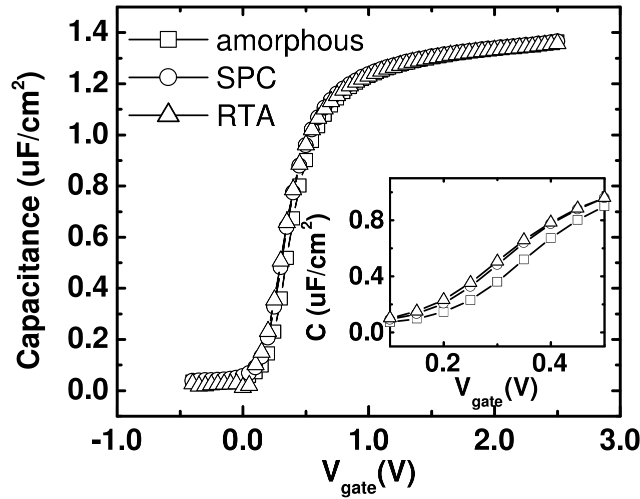

Figure 4.16. Measured C-V characteristics phosphorus implanted FUSI NiSi gates. The a-Si gate has gone through either no thermal treatment (amorphous), SPC or RTA before Ni deposition.

4.5 Gate Dielectric Reliability

As shown in Figure 4.2, the silicidation front is non-uniform in poly-Si due to the anisotropy of the granular structure. A more uniform silicidation front may be obtained if a-Si gate is used. However, as shown above, in order to obtain a large tunable WF range in FUSI NiSi gate, it is necessary to activate the dopants and to drive the dopants closer to the gate electrode/ dielectric interface. As a result, crystallization of poly-Si gate cannot be avoided. This implies that some poly-Si area could have been fully silicided earlier than the others. This happens within the same transistor gate when the gate area is larger than the grain size (wide transistor) or

across different transistor gates when the gate area is smaller than the grain size. Therefore, in order to have good uniformity within the same transistor and across the wafer, it is necessary to have large enough thermal budget in silicidation. Otherwise, incomplete FUSI will not just increase gate depletion effect, but also, more seriously, cause very large threshold voltage variations (Figure 4.3).

In FUSI NiSi gate, silicidation occurs throughout the gate electrode and the silicide finally is in contact with the gate oxide. A large enough thermal budget means that certain parts of the NiSi is already in contact with the gate oxide during annealing, thus will be "over-annealed". Therefore, there is concern about the qualities of the gate oxide at these "over-annealed" spots.

In order to study the gate dielectric qualities for FUSI NiSi gates, capacitors were fabricated with LOCOS isolation as described in Section 4.2. The a-Si gates were undoped and some of them were implanted with boron ($2.4 \times 10^{15} cm^{-2}$, 1keV) or phosphorus ($2.4 \times 10^{15} cm^{-2}$, 2keV). Since B-doped FUSI NiSi is likely to be used in pMOSFET due to its larger tunability from Si mid-gap to valence band edge, B doped FUSI NiSi was fabricated on n-type substrate ($5 \times 10^{14} cm^{-3}$). For undoped and P-doped FUSI NiSi, p-type ($1 \times 10^{15} cm^{-3}$) substrate was used. After implantation, all poly-Si were annealed at $800^{o}C$ for 4 hours. Then 30nm Ni was sputtered and capped with 40nm of TiN. The silicidation annealing was performed at $530^{o}C$ for 1 min followed by FGA.

Control samples were also fabricated following exactly the same process except that the metallization steps were omitted and the wafers were further spike annealed at $1000^{o}C$.

Firstly, the gate leakage current densities of the capacitors were measured and are shown in Figure 4.17 to 4.18. The gate bias is reference to the flat band voltage (V_{FB}) of the capacitors because FUSI NiSi and doped poly-Si gates have different

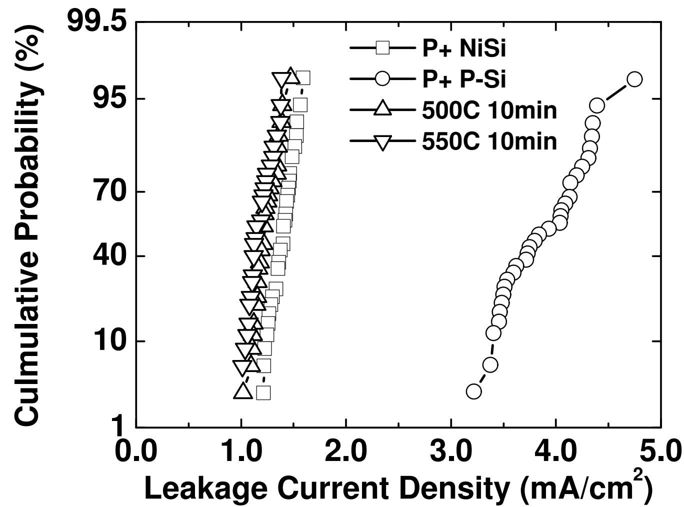

Figure 4.17. Leakage current density of boron doped (p+) FUSI NiSi gate *vs.* p+ poly-Si measured at $V_G = V_{FB} + 1.1V$. The area of the capacitors is $150\mu m^2$. "500C 10min" and "550C 10min" are the measured data of p+ FUSI NiSi after further annealing at 500^oC and 550^oC for 10 mins.

effective WF's. Since the gate oxide thickness is only 2.2nm, direct tunneling is the dominating gate leakage mechanism [22]. If the oxide is a perfect insulator, the leakage current will only depend on the carrier densities at the injection source (substrate) in deep accumulation mode. By offsetting the gate bias with reference to the V_{FB}'s, the carrier densities at the substrate will be the same and the difference between the oxide qualities can then be revealed from the leakage current density plots.

Figure 4.17 shows that the leakage current density of p+ poly-Si gate capacitor is three times larger than that of the B-doped FUSI NiSi gate. This is probably due to dielectric degradation caused by boron penetration in the poly-Si gate during the spike annealing, which is commonly implemented in traditional CMOS process. In Figure 4.18, it shows that undoped FUSI NiSi gate has about two times larger leakage current than n+ poly-Si but that of P-doped FUSI NiSi is similar to that of the reference.

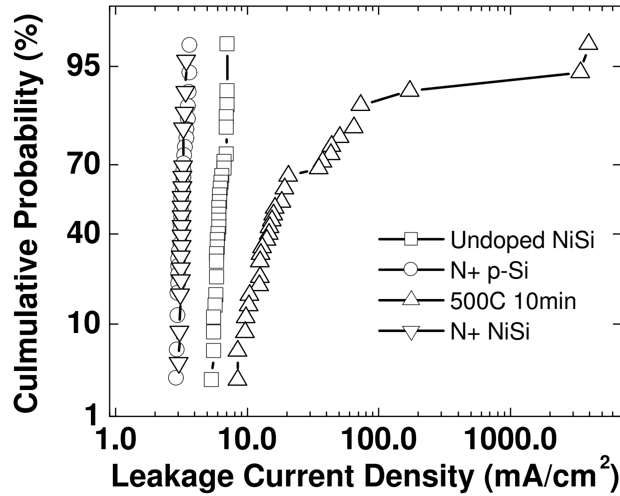

Figure 4.18. Leakage current density of undoped FUSI NiSi gate *vs.* n+ poly-Si measured at $V_G = V_{FB} - 1.03V$. The area of the capacitors is $150\mu m^2$. "500C 10min" is the measured data of FUSI NiSi after further annealing at 500^oC for 10 mins.

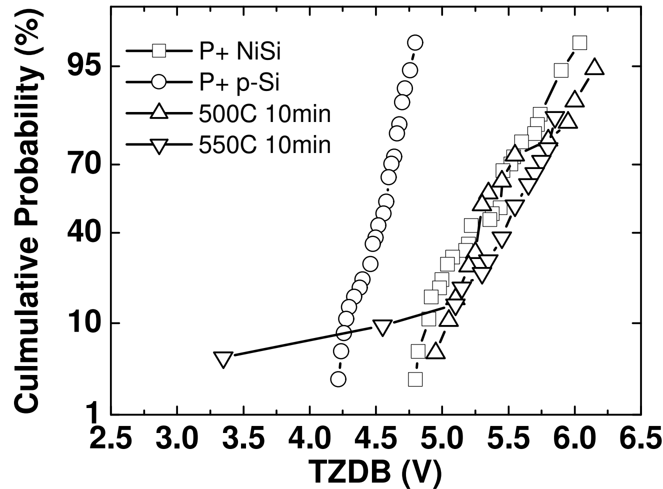

Figure 4.19. Time Zero to Dielectric Breakdown (TZDB) of boron doped (p+) FUSI NiSi gate *vs.* p+ poly-Si. The area of the capacitors is $50\mu m^2$. "500C 10min" and "550C 10min" are the measured data of p+ FUSI NiSi after further annealing at 500^oC and 550^oC for 10 mins.

Leakage current density represents the areal quality of the oxide. To further investigate if the weak spots are degraded or increase in FUSI NiSi gate, Time Zero

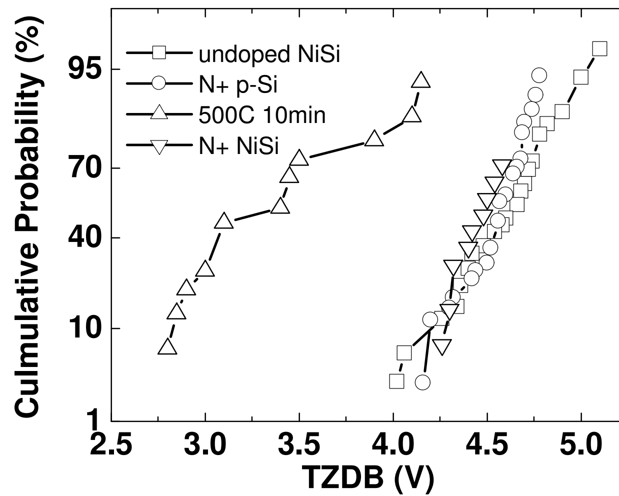

Figure 4.20. Time Zero to Dielectric Breakdown (TZDB) of undoped and phosphorus doped (n+) FUSI NiSi gate *vs.* n+ poly-Si. The area of the capacitors is $50 \mu m^2$. "500C 10min" is the measured data of FUSI NiSi after further annealing at $500^o C$ for 10 mins.

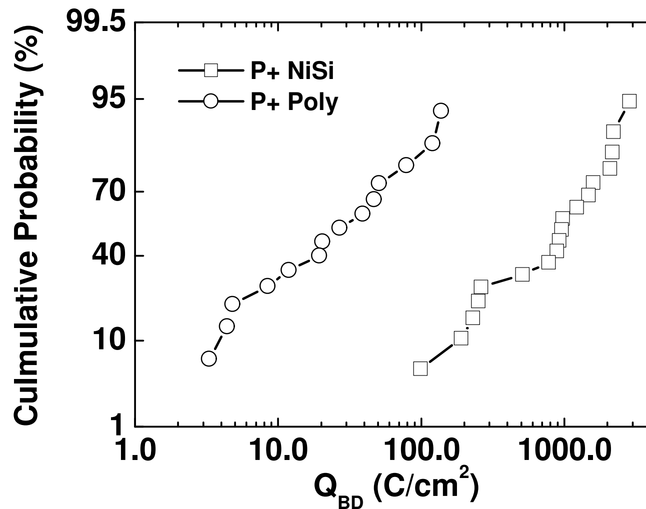

Figure 4.21. Charge to Breakdown (QBD) of boron doped (p+) FUSI NiSi gate *vs.* p+ poly-Si at $V_G = 4.1V$. The area of the capacitors is $25 \mu m^2$.

to Dielectric Breakdown (TZDB) and Charge to Breakdown (QBD) of the capacitors were also investigated. The results are shown in Figure 4.19 to 4.22. It can be seen

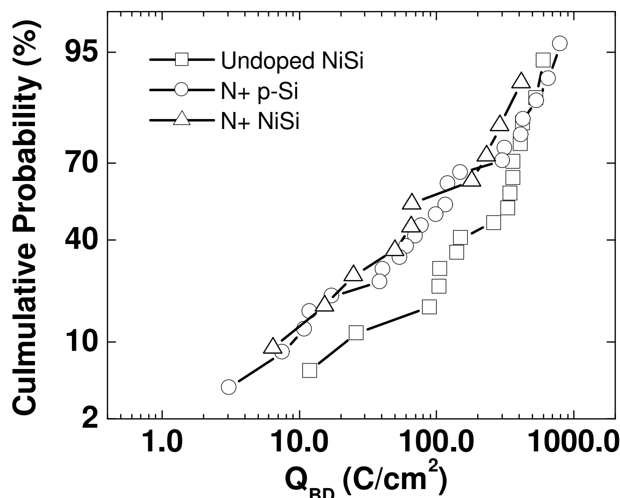

Figure 4.22. Charge to Breakdown (QBD) of phosphorus doped (n+) and undoped FUSI NiSi gate *vs.* n+ poly-Si at $V_G = -3.85V$. The area of the capacitors is $25\mu m^2$.

that result is consistent with the leakage current data. For boron doped FUSI NiSi, the quality is much better than the p+ poly-gate reference as boron penetration is reduced in FUSI NiSi gate. For undoped and phosphorus doped NiSi, it actually has comparable TZDB and QBD with the n+ poly sample. Therefore, the gate oxide qualities are not degraded in FUSI NiSi and is even better than p+ poly-Si reference as high temperature annealing of boron doped poly-Si gate is avoided.

4.6 Thermal Stability

FUSI NiSi gate technology avoids the high temperature annealing budget in the FEOL steps but still has to go through the thermal budget in the BEOL steps. As the NiSi is already in direct contact with the gate dielectric, there is also a concern about the degradation of the gate dielectric even the excess Ni has already been removed before the BEOL.

Therefore, the samples used in Section 4.5 were annealed for 10 minutes in N_2

ambient at 500^oC and then 550^oC respectively. Figure 4.17 shows that the gate leakage current of B doped FUSI NiSi does not change substantially after 550^oC annealing for 10 minutes. However, the yield starts degrading as shown in the TZBD plot (Figure 4.19). But for undoped FUSI NiSi gate, it degrades even at 500^oC 10 minutes annealing (Figure 4.18 and 4.20).

Since there is a significant dopant pile-up at the electrode-dielectric interface, it is believed that the boron has formed a barrier layer to prevent the Ni from diffusing into the SiO_2. This is absent in the undoped FUSI NiSi case. As a result, undoped NiSi is not stable upon further BEOL annealing.

4.7 Summary

In this chapter, we have studied the tunable WF range of FUSI NiSi gate. The effective WF of FUSI NiSi gate can be turned towards the valence band edge by implanting boron, or towards the conduction band edge by implanting phosphorus, into the poly-Si before Ni deposition. The WF range is tunable from 4.53eV to 4.99eV. This is comparable with reported values in the literature and is enough for 10nm FinFET application.

If a larger tunable WF range is required, As can be used instead of P. However, we found that As implanted FUSI NiSi can peel off easily during FGA annealing. It was also found that the shift of WF is limited if the implanted gate is not annealed. Therefore, it is necessary to find out the conditions which will lead to the maximum tunable WF range. With the help of excimer laser tool, it is found that dopant activation and initial dopant distribution in the gate is important to shift the WF. The higher the activated dopant concentration and the closer the dopant is to the electrode/ dielectric interface, the more the shift of the WF. As a result, the implantation

dose, energy and the annealing condition has to be optimized in order to maximize the WF shift.

Although metal is involved in the process and the NiSi is in direct contact with the gate dielectric in FUSI gate, the oxide qualities (leakage current, TZDB and QBD) are not degraded. The fixed charge is not increased as long as the implantation dose or the energy is low enough. The boron doped FUSI NiSi gate yields better gate dielectric qualities than the p+ poly-Si reference because high temperature annealing of boron doped poly-Si gate is avoided in FUSI process and thus, boron penetration is reduced. However, FUSI NiSi degrades in BEOL annealing even at 500^oC. The situation can be improved if the poly-Si has received boron implantation. This is probably because the pile-up dopants at the electrode / dielectric interface help to prevent the Ni from diffusing into the gate oxide. Therefore, FUSI NiSi should be used with dopant implanted if BEOL annealing is larger than 500^oC.

In conclusion, FUSI NiSi is very compatible with traditional CMOS process and it's WF can be turned over a fairly wide range. Therefore, it is a promising candidate for sub-32nm CMOS technologies.

References

[1] *The International Technology Roadmap for Semiconductors*, Semiconductor Industry Association, 2005.

[2] C. J. Cabral, J. Kedzierski, B. Linder, S. Zafar, V. Narayanan, S. Fang, A. Steegen, P. Kozlowski, R. Carruthers, and R. Jammy, "Dual workfunction fully silicided metal gates," in *Symposium on VLSI Technology. Digest of Technical Papers*, 2004, pp. 184–5.

[3] K. Takahashi, K. Manabe, T. Ikarashi, N. Ikarashi, T. Hase, T. Yoshihara, H. Watanabe, T. Tatsumi, and Y. Mochizuki, "Dual workfunction Ni-Silicide/HfSiON gate stacks by phase-controlled full-silicidation (PC-FUSI) technique for 45nm-node LSTP and LOP devices," in *International Electron Devices Meeting. Technical Digest*, Dec. 2004, pp. 91–94.

[4] Z. Krivokapic, W. Maszara, K. Achutan, P. King, J. Gray, M. Sidorow, E. Zhao, J. Zhang, J. Chan, A. Marathe, and M.-R. Lin, "Nickel silicide metal gate FDSOI devices with improved gate oxide leakage," in *International Electron Devices Meeting. Technical Digest*, Dec. 2002, pp. 271–274.

[5] W. Maszara, Z. Krivokapic, P. King, J.-S. Goo, and M.-R. Lin, "Transistors with dual work function metal gates by single full silicidation (FUSI) of polysilicon gates," in *International Electron Devices Meeting. Technical Digest*, Dec. 2002, pp. 367–70.

[6] L. Chang, Y. kyu Choi, D. Ha, P. Ranade, S. Xiong, J. Bokor, C. Hu, and T. King, "Extremely scaled silicon nano-CMOS devices," *Proceedings of the IEEE*, vol. 91, no. 11, pp. 1860–73, 2003.

[7] P. Ranade, H. Takeuchi, T.-J. King, and C. Hu, "Work function engineering of

molybdenum gate electrodes by nitrogen implantation," *Electrochem. Solid-State Lett.*, vol. 4, pp. G85–G87, 2001.

[8] I. Polishchuk, P. Ranade, T.-J. King, and C. Hu, "Dual work function metal gate CMOS transistors by Ni-Ti interdiffusion," *IEEE Electron Device Letters*, vol. 23, pp. 200–202, 2002.

[9] W. Gao, J. F. Conley, and Y. Ono, "Stacked metal layers as gates for MOSFET threshold voltage control," in *Mater. Res. Soc. Symp*, vol. 765, 2003, pp. D1.4.1–D1.4.6.

[10] M. Qin, V. Poon, and S. Ho, "Investigation of polycrystalline nickel silicide films as a gate material," *Journal of the Electrochemical Society*, vol. 148, no. 5, pp. G271–4, 2001.

[11] J. Sim, H. Wen, J. Lu, and D. Kwong, "Dual work function metal gates using full nickel silicidation of doped poly-si," *IEEE Electron Device Letters*, vol. 24, pp. 631–3, Oct. 2003.

[12] J. Kedzierski, D. Boyd, C. J. Cabral, P. Ronsheim, S. Zafar, P. Kozlowski, J. Ott, and M. Ieong, "Threshold voltage control in NiSi-gated MOSFETs through SIIS," *IEEE Transactions on Electron Devices*, vol. 52, pp. 39–46, Jan. 2005.

[13] M. Terai, K. Takahashi, K. Manabe, T. Hase, T. Ogura, M. Saitoh, T. Iwamoto, T. Tatsumi, and H. Watanabe, "Highly reliable HfSiON CMOSFET with phase controlled NiSi (NFET) and Ni_3/Si (PFET) FUSI gate electrode," in *Symposium on VLSI Technology. Digest of Technical Papers*, 2005, pp. 68–9.

[14] K. Hosaka, T. Kurahashi, K. Kawamura, T. Aoyama, Y. Mishima, K. Suzuki, and S. Sato, "A comprehensive study of fully-silicided gates to achieve wide-range work function differences (0.91 eV) for high-performance CMOS devices," in *Symposium on VLSI Technology. Digest of Technical Papers*, 2005, pp. 66–7.

[15] M. Kakumu and K. Hashimoto, “Work function controlled silicide technology for submicron CMOS,” in *Symposium on VLSI Technology. Digest of Technical Papers*, 1984, pp. 30–1.

[16] B. Tavel, T. Skotnicki, G. Pares, N. Carriere, M. Rivoire, F. Leverd, C. Julien, J. Torres, and R. Pantel, “Totally silicided ($CoSi_2$) polysilicon: a novel approach to very low-resistive gate $2\Omega/\square$ without metal CMP nor etching,” in *International Electron Devices Meeting. Technical Digest*, Dec. 2001, pp. 37.5.1–37.5.4.

[17] J. Fedzierski *et al.*, “Metal-gate FinFET and fully-depleted SOI devices using total gate silicidation,” in *International Electron Devices Meeting. Technical Digest*, Dec. 2002, pp. 247–50.

[18] G. Ottaviani, “Review of binary alloy formation by thin film interactions,” *Journal of Vacuum Science & Technology*, vol. 16, pp. 1112–1119, 1979.

[19] S. P. Murarka, Ed., *Silicides for VLSI Application.* Academic Press, 1986.

[20] P.Sallagoity, M.Juge, O.Diop, F.Oudin, G.Beaulieu, L.Seube, F.Gra, and P.Merenda, “Gate material properties induced $0.25\mu m$ SRAM marginality,” in *ESSDERC*, 2002, pp. 575–578.

[21] P. J. Grunthaner, F. J. Grunthaner, D. M. Scott, M.-A. Nicolet, and J. W. Mayer, “Oxygen impurity effects at metal/silicide interfaces: Formation of silicon oxide and suboxides in the ni/si system,” *Journal of Vacuum Science & Technology*, vol. 19, pp. 641–648, 1981.

[22] D. K. Schroder, *Semiconductor material and device characterization.* NY, USA: John Wiley & Sons, 1998.

[23] L. Chang, S. Tang, T.-J. King, J. Bokor, and C. Hu, “Gate length scaling and

threshold voltage control of double-gate MOSFETs," in *International Electron Devices Meeting. Technical Digest*, Dec. 2000, pp. 37.2.1–37.2.4.

[24] D. Aime, B. Froment, F. Cacho, V. Carron, S. Descombes, Y. Morand, N. Emonet, F. Wacquant, T. Farjot, S. Jullian, C. Laviron, M. Juhel, R. Pantel, R. Molins, D. Delille, A. Halimaoui, D. Bensahel, and A. Souifi, "Work function tuning through dopant scanning and related effects in Ni fully silicided gate for sub-45nm nodes CMOS," in *International Electron Devices Meeting. Technical Digest*, Dec. 2004, pp. 87–90.

[25] J. Sim, H. Wen, J. Lu, and D. Kwong, "Work function tuning of fully silicided NiSi metal gates using a TiN capping layer," *IEEE Electron Device Letters*, vol. 25, pp. 610–612, 2004.

[26] J. Liu, H. Wen, J. Lu, and D. Kwong, "Improving gate-oxide reliability by tin capping layer on $nisifusi$ metal gate," *IEEE Electron Device Letters*, vol. 26, pp. 458–460, 2005.

[27] J. Yuan and J. Woo, "Tunable work function in fully nickel-silicided polysilicon gates for metal gate MOSFET applications," *IEEE Electron Device Letters*, vol. 26, pp. 87–89, Feb. 2005.

[28] H. Y. Wong, H. Takeuchi, T.-J. King, M. Ameen, and A. Agarwal, "Elimination of poly-Si gate depletion for sub-65-nm CMOS technologies by excimer laser annealing," *IEEE Electron Device Letters*, vol. 26, pp. 234–6, 2005.

[29] J. S. Im, H. J. Kim, and M. O. Thompson, "Phase transformation mechanisms involved in excimer laser crystallization of amorphous silicon films," *Applied Physics Letters*, vol. 63, pp. 1969–1971, 1993.

[30] QMCV Simulator. [Online]. Available: http://www-device.eecs.berkeley.edu/qmcv

[31] H. Y. Wong, H. Takeuchi, T.-J. King, M. Ameen, and A. Agarwal, "Reduced poly-Si gate depletion effect by pulsed excimer," in *Advanced Short-Time Thermal Processing for Si-Based CMOS Devices II*. The Electrochemical Society, Inc., 2004, pp. 205–215.

Chapter 5

Conclusion

5.1 Summary

Transistor scaling has been the driving force for technology advancement in the semiconductor industry over the last few decades. In the course of scaling, the gate stack has to be scaled at a pace the same as, or exceeding that of the channel length in order to retain gate control within the transistors. However, at the 65nm technology node and beyond, gate stack scaling faces fundamental limits [1]. In this thesis, we presented two novel gate electrode technologies to meet the near term ($\geq$32nm technology nodes) and long term ($<$32nm technology nodes) challenges, namely excimer laser annealing (ELA) of poly-Si gate and fully silicided nickel silicide (FUSI NISI) gate, respectively. Promising results have been demonstrated.

In the near term, the gate depletion effect, gate leakage current density, and dopant penetration through the gate dielectric are expected to be very serious as the equivalent gate oxide thickness (EOT) is scaled down. Therefore, a novel gate annealing technique is required so that 1) the poly-Si gate depletion effect can be reduced or even eliminated, 2) boron penetration can be suppressed in dual-doped

poly-Si gate electrodes, and 3) high-κ dielectric can be introduced while its thermal budget constraint is met.

Melt excimer laser annealing (ELA) is proposed as a promising novel gate annealing technique. In this technique, amorphous Si (a-Si) is used as the precursor gate material and a pulsed excimer laser is used to fully melt the a-Si after gate dopant implantation. A finite-difference thermal simulator was written in MATLAB code to understand the laser annealing process on CMOS gate stacks. Since the a-Si has much lower melting point (by $\sim 250^oC$) than the crystalline silicon (c-Si), it is found that melting of the substrate c-Si can be avoided even there are process variations, and that the process window is about ±8% according to the simulation. Both simulation and in-situ infra-red reflectance measurement show that the melting time of the film is ~50ns, and the whole heating process is completed within $\sim 0.1\mu s$. Due to the high diffusivity of the dopants in the molten silicon, although the melting time is short, the dopants distribute evenly throughout the film, as shown by the SIMS data. And at the same time, there is no boron penetration because the underlying dielectric does not melt and the process time is short. The dopants are activated after ELA and resistivity of boron- and phosphorus-doped poly-Si of about $0.5m\Omega \cdot cm$ was achieved.

It is found that after ELA, the effective work function (WF) of the poly-Si gate is shifted towards the mid-gap. This is serious in p+ doped poly-Si but almost negligible in n+ doped gate. This is a result of the disorder induced gap states (DIGS) after the rapid quenching of the ELA process. It is found that there are multiple activation energies associated with the recovery of the disorder states, which are expected to be very low (one of them is found to be 0.36eV). Therefore, the WF can be recovered easily with furnace annealing at moderate temperature (~600°C). Even with additional furnace annealing, ELA is found to be able to reduce the gate depletion effect to be as low as 0.1nm.

The gate dielectric properties (SiO_2) after ELA do not degrade because the processing time is very short, despite the high processing temperature. Gate leakage current density, Time-Zero-to-Dielectric Breakdown (TZDB), and hole mobility of ELA devices are comparable to or even better than those of the RTA poly-Si references. The ultra-short processing time of ELA also makes itself very compatible with high-κ dielectrics such as HfO_2. It is found that, with HfO_2 as the gate dielectric, ELA samples have about 5-6 orders of magnitude lower gate leakage current density than the RTA references, and that the interfacial layer formation is reduced. It is also found that, in terms of Equivalent Oxide Thickness (EOT) and leakage current density, ELA poly-Si on HfO_2 can meet the ITRS 65nm and 45nm technology nodes requirements.

However, integration of ELA with a traditional CMOS process is not trivial. It is found that ELA is incompatible with traditional RTA, which results in severe boron penetration. Fortunately, ELA is compatible with flash-lamp annealing (FLA), which is expected to be widely used for S/D dopant activation in the future technologies. In order to prevent the intermixing of p+ and n+ dopants and the ablation of Si film over the isolation regions, a selective ELA via masking scheme was proposed and studied. It is found that although this scheme is projected to be applicable down to 32nm nodes, extra masks are required. Another attempt to integrate ELA with a Self-Aligned STI process was also studied and demonstrated to solve the intermixing and ablation issues. However, Ge PAI is required to amorphize the poly-Si gate film in the SA-STI process. This causes damage to the gate dielectric. It is found that by using this scheme, pMOSFETs have 8.8% to 11.1% improvement in drive current but there is degradation in the drive current of nMOSFETs (3.6% to 6.7%), probably due to the mobility degradation associated with PAI. However, the TEM pictures showed that integration of ELA with SA-STI process gives very good film

morphologies. Therefore, this is still very promising as long as the PAI step can be eliminated. Section 5.2 presents such a modified ELA SA-STI process.

In the long term, due to the advent of advanced transistor structures and in order to eliminate the gate depletion effect, metal gate with tunable WF is required. FUSI NiSi is a promising candidate and we showed that the gate depletion effect is eliminated and the tunable WF range is between 4.53eV and 4.99eV. It is found that it is necessary to anneal the implanted poly-Si gate before NiSi formation. Otherwise, the tunable WF range will be reduced. This can be due to the activation of the dopants or the change of the dopant profile after annealing. In order to make a distinction between these two effects, ELA was used as an experimental tool and we proved that, for boron, by activating the dopants and bring the dopants closer to the gate electrode/ dielectric interface, the WF tunable range of NiSi FUSI gate can be enhanced. We, therefore, demonstrated that by using RTA, more WF shift is achieved than using SPC in boron and phosphorus doped FUSI NiSi gates. The gate dielectric reliabilities of FUSI NiSi gate were also studied. It is found that good oxide reliability is retained. Since high temperature annealing of boron doped poly-Si gate is obviated in FUSI gate, the reliability of boron-doped FUSI NiSi is much better than that of the boron-doped poly-Si reference. Moreover, boron-doped FUSI NiSi samples do not degrade with up to 550°C backend of line annealing. This is due to the fact that the piled-up dopants prevent the Ni from diffusing into the gate dielectric.

5.2 Recommendations for Future Work

5.2.1 Integration of ELA with Dummy Gate SA-STI

As shown in Figure 3.19, SA-STI is an excellent process to integrate ELA for eliminating the gate depletion effect, and issues of ablation and dopant intermixing

can be solved. The only problem is that PAI is required in order to provide enough window for process variation (so that the gate can be completely melted while the substrate is intact). There are actually two solutions to this. The first is by using silicon-germanium as the gate material. Silicon germanium has lower melting point than Si. For example, $Si_{0.6}Ge_{0.4}$ has about 250°C lower melting point than c-Si [2], which will provide the same process window as the a-Si does. However, since SiGe has smaller electron affinity than Si, $Si_{0.6}Ge_{0.4}$ gate will results in about 0.2V larger threshold voltage than the Si gate in pMOSFETs [3]. This is not favorable for deep sub-micron CMOS, since low threshold voltages are required.

The second solution is to have a dummy gate during the high temperature STI formation. The dummy gate material can be either high Ge content Si_xGe_{1-x} ($x<$ 0.4), which can be easily removed by SC1 solution (5 H_2O:1 H_2O_2:1 NH_4OH) or poly-silicon which can be removed by silicon etchant (33% H_2O: 3% NH_4F: 64% HNO_3). After that, gate dielectric and electrode can be deposited and chemical-mechanically polished (CMP), so that it is still self-aligned with the STI.

The following shows the suggested process flow for dummy gate SA-STI process integrated with ELA. Since the gate material is fully melted, FLA is necessary for S/D activation.

1. Starting wafer up to gate dielectric formation
2. Dummy gate oxide formation (4nm)
3. Dummy Si or Si_xGe_{1-x} gate deposition (50nm) (Figure 5.1)
4. Self-Aligned STI formation (Trench etching and oxide filling)
5. STI planarization by CMP with dummy gate as stopping layer (Figure 5.2)
6. Dummy Si gate removal by Si etchant or SC1 solution

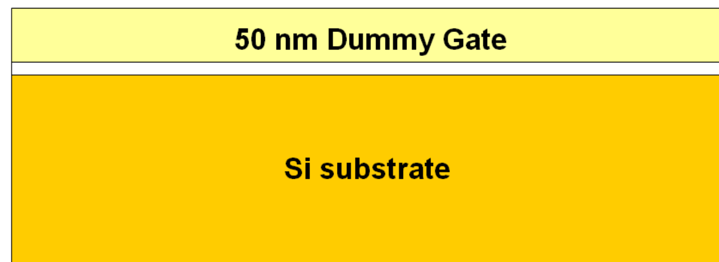

Figure 5.1. Dummy gate ELA SA-STI process after dummy gate deposition.

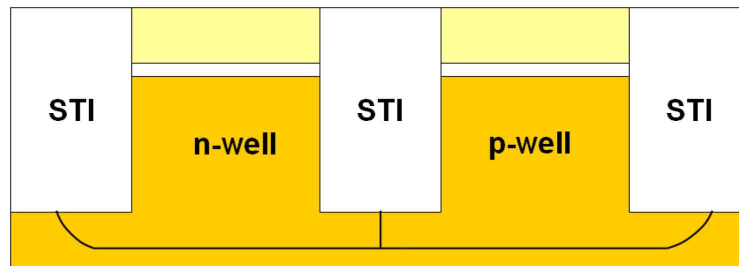

Figure 5.2. Dummy gate ELA SA-STI process after STI planarization.

7. Dummy gate oxide removal by HF solution (Figure 5.3)

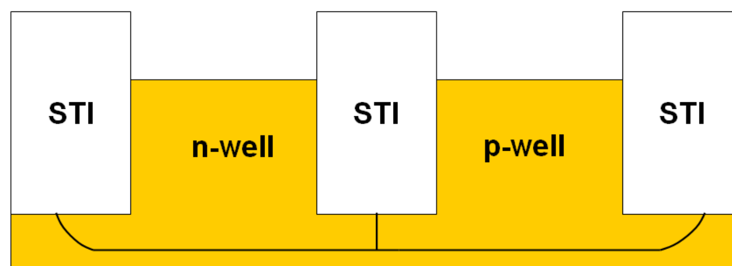

Figure 5.3. Dummy gate ELA SA-STI process after dummy gate removal.

8. Gate dielectric formation

9. a-Si gate deposition (Figure 5.4)

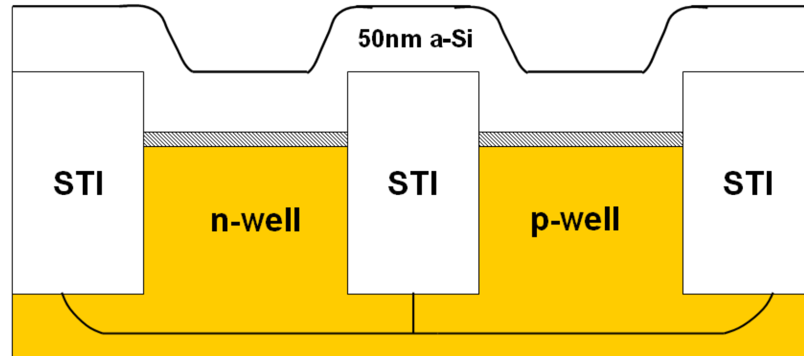

Figure 5.4. Dummy gate ELA SA-STI process after gate deposition.

10. CMP of a-Si with STI oxide as stopping layer (Figure 5.5)

11. Steps up to gate activation

12. Gate activation using blanket ELA (Figure 5.5)

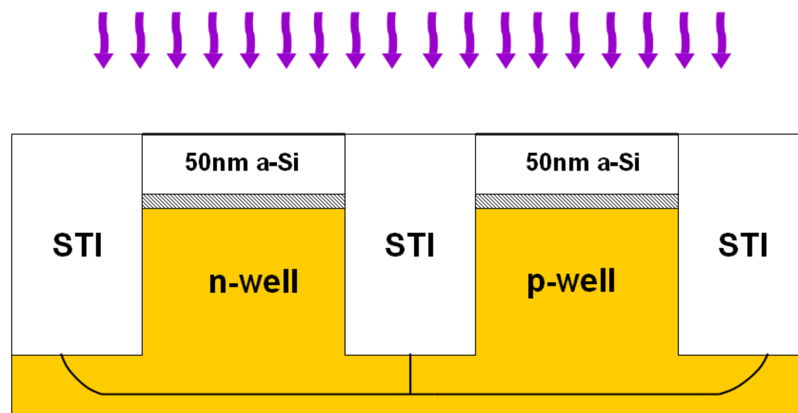

Figure 5.5. Dummy gate ELA SA-STI process during ELA.

13. Flash lamp annealing for S/D activation

14. BEOL steps

5.2.2 FUSI NiSi Gate

Although it is found that dopant activation and position affect the final tunable range of the WF in FUSI NiSi gate, the mechanism of dopant pile-up is not well understood. Moreover, dopant pile-up is known to be responsible for the WF shift, but the mechanism is also not clear. Therefore, it is interesting to study the piling-up process and WF shift by atomistic simulation. The results will be helpful in obtaining larger WF range and better WF uniformity.

It has been found that boron piled-up at the gate electrode/ dielectric interface helps stabilize the FUSI NiSi gate. Whether other dopant species (phosphorus, arsenic etc.) have the same effect is not known neither. Therefore, it is worthy to study the thermal stabilities of FUSI NiSi with different doses and types of dopants.

Arsenic doped FUSI NiSi can be peeled off easily upon forming gas annealing. Due to its larger work function (WF) tunable range than that of phosphorus, it is worthy to study the peeling-off mechanism and how it can be prevented.

References

[1] *The International Technology Roadmap for Semiconductors*, Semiconductor Industry Association, 2005.

[2] H. Stöhr and W. Klemm, "Über zweistoffsysteme mit germanium. i. germanium/aluminium, germanium/zinn und germanium/silicium," *Zeitschrift für anorganische und allgemeine Chemie*, vol. 241, pp. 305–323, 1939.

[3] T.-J. King, "Application of polycrystalline silicon-germanium thin films in metal-oxide-semiconductor technologies," Ph.D. dissertation, Stanford University, 1994.

Appendix A

MATLAB Simulation Codes for Gate Stack ELA Process

A.1 Overview

MATLAB codes were written to simulate the melt ELA process of CMOS gate stacks ($a - Si/SiO_2/c - Si$). The main routine is called *main.m*. This routine calls *P_init.m* to initialize the spatial and temporal distribution of the laser pulse. *Temperature.m* is then used to solve the thermal diffusion equations by calling other sub-routines. The sub-routines are *Cf.m*, *Rouf.m*, *Pf.m* and *Kf.m*, in which the temperature dependent specific heat capacities, material densities, heat energy generated due to photon absorption and temperature dependent thermal conductivities are calculated. Figure A.1 shows the flow chart of the program and Section A.2 lists the codes.

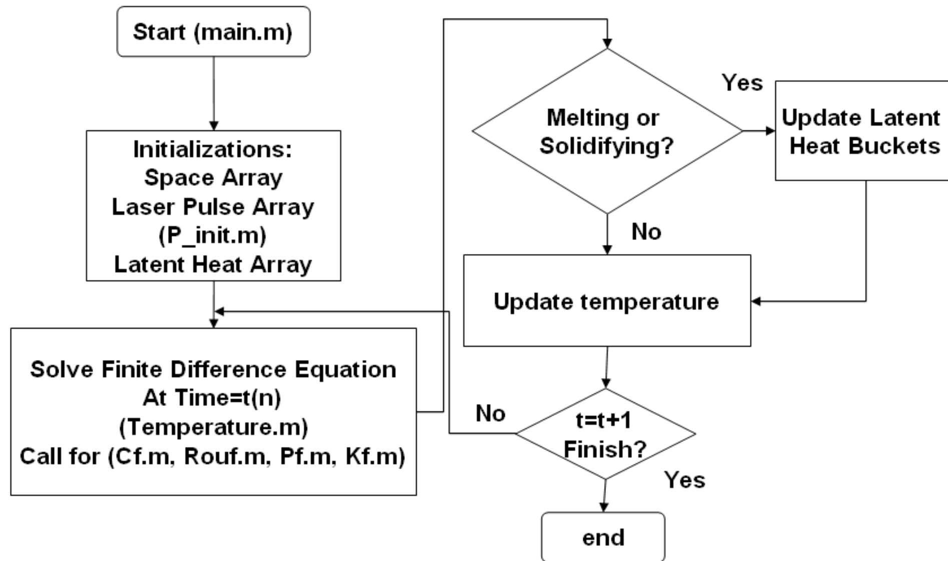

Figure A.1. Flowchart of the thermal simulation.

A.2 main.m

```
clear all;


%Universal Constants

global k q  eps0 h c;

k=1.38066*10^-23;       %Boltzmann constant in J/K

q=1.60218*10^-19;       %Electron Charge in C

eps0=8.854*10^-12;      %Vacuum Permittivity in F/m

h=6.626*10^-34;         %Planck Constant in J.s

c=2.998*10^8;           %light speed, m/s


%Intialization

global max_x max_t t P delta_t delta_x x_ind t_array material;

global P TEMP rou C K I0 xdistance;

global LH LHFa TEMP_mpa;


%input deck for time
```

```
%*************************
Total_t=80*10^-9;       %total simulation time in s
t_array=2000;           %size of time array
delta_t=5*10^-14;       %simulation time step
%*************************

%TIME STEP has to be small enough for accuracy
% when K increases, the simulation step has to be decreased!!

max_t=Total_t/delta_t;          %This has to be the multiple of 1000
iteration_t=max_t/t_array;      %number of iteration in t
t_ind=1:1:max_t;                %time index
t=(t_ind-1).*delta_t;           %time in s, index

%input deck for space
%***********
% define layers
layer1=55*10^-9;        %in meters
layer2=3*10^-9;         %in meters
layer3=4.2*10^-6;       %in meters
deltax1=5*10^-9;        %in meters
deltax2=3*10^-9;        %in meters
deltax3=30*10^-9;       %in meters
layerdummy=deltax1;
%*********
node0=1;                        % Node0 as dummy in Temperature calculation
node1=round(layer1/deltax1);
node2=round(layer2/deltax2);
node3=round(layer3/deltax3);
max_x=1+node1+node2+node3       %number of nodes
x_ind=1:1:max_x;                %depth index
x_ind=x_ind.';                  %convert into coloumn
delta_x=zeros(max_x,1);
```

```
delta_x(1:node0)=deltax1;
delta_x(node0+1:node0+node1+1)=deltax1;
delta_x(node0+node1+1:node0+node1+node2)=deltax2;
delta_x(node0+node1+node2+1:node0+node1+node2+node3)=deltax3;
xdistance(1)=0;

%%**** This is important. Can only be used in
%% Absorption part!!! Make Oxide transparent.

for node=1:node1
    xdistance(node+1)=deltax1*node;
end
for node=node1+1:node1+node2
    xdistance(node+1)=deltax1*node1+deltax2*(node-node1);
end
for node=node1+node2+1:node1+node2+node3
    xdistance(node+1)=deltax1*node1+deltax3*(node-node1-node2);
%% deltax2*node2 ignored => oxide is transparent
end

material=zeros(max_x,1);        %define materials (1=a-Si 2=SiO2 3=c-Si)
LHFa=zeros(max_x,1);            % array of latern heat
TEMP_mpa=zeros(max_x,1);        % array of melting point

%************************
%Data from Table 1, Pg 189 R.F Wood book and internet
off=1;
%node due to the dummy node
material(1:node1+off)=1;
material(node1+1+off:node1+node2+off)=2;
material(node1+node2+1+off:node1+node2+node3+off)=3;
%Latent heat of a-Si J/kg
LHFa(1:node1+off)=1320*1000*delta_x(1:node1+off)*2.33/1000*10^6;
```

```
% Latent heat of oxide
LHFa(node1+1+off:node1+node2+off)=0;
%Latent heat of c-Si J/kg
LHFa(node1+node2+1+off:node1+node2+node3+off)=1800*1000*...
           delta_x(node1+node2+1+off:node1+node2+node3+off)*2.33/1000*10^6;
TEMP_mpa(1:node1+off)=1150+273;
TEMP_mpa(node1+1+off:node1+node2+off)=1713+273;
TEMP_mpa(node1+node2+1+off:node1+node2+node3+off)=1410+273;

%% K, C, rou are temperature dependent, in seperate functions
%************************

%Step 1: Initialization of Heat Generation Matrix
TEMP=zeros(max_x,t_array);    %Temperature
LH=zeros(max_x,1);            %Store latent heat of different nodes
%Step 2: Initialization of Temperature matrix
% Call Temp solver
LH_start=zeros(max_x,1);
TEMP_start=zeros(max_x,1);
TEMP_start(:)=300;
Op=0;                         %starting time
    figure
    for i=1:iteration_t
    if(i==1)
        Op=input('Which time step to start? (ps),
n');
        %initialize the latent heat bucket and also temperature
        if(Op==0)                       %start from 0s
            TEMP_start(:)=300;
            LH_start(:)=0;
            P_init;           %Initialize the Power profile
        else
            IO=load('pulsefile');
```

```
            filename3=sprintf('%d3.dat',round(Op/delta_t));
            filename4=sprintf('%d4.dat',round(Op/delta_t));
           % plot(t_ind,I0);
            TEMP_start=load(filename3);
            LH_start=load(filename4);
        end
    end
    TEMP(:,1)=TEMP_start;
    LH=LH_start;
    [TEMP_start, LH_start]=Temperature(Op);
    Op=Op+t_array*delta_t
hold on
    plot(1:10,TEMP(1:10,10))
end
```

A.2.1 P_init.m

```
%This file initialize the pulse profile w.r.t. time
global max_t t delta_t I0;
% Material Constants
% All data from Jellison and Modine, 1982c, 308nm Excimer Laser
%Variables
pw=38*10^-9;             %pulse width in s
E_pulse=0.65*100^2;      %pulse energy in J/m^2
lambda=248*10^-9;        %wavelength in m
temp=300;
I=E_pulse/pw;            %Light intensity at surface
pulse_duration=min(round(pw/delta_t),max_t);
                         % choose the smaller one
I0=zeros(max_t,1);
```

```
for index_t=1:1:round(pw/delta_t);
   IO(index_t)=I;
end
save('pulsefile' ,'IO','-ASCII');
```

A.2.2 Temperature.m

```
function [TEMP_start,LH_start]=Temperature(Op)


global TEMP max_x delta_t delta_x LH LHFa TEMP_mpa t_array iteration_t material;
for n=1:t_array-1
    for m=1:max_x-2
%Temperature dependent physical parameters
 Km1=Kf(m+1,TEMP(m+1,n),TEMP(m+2,n));
        Km=Kf(m,TEMP(m,n),TEMP(m+1,n));
        Cm1=Cf(m+1,TEMP(m+1,n));
        roum1=rouf(m+1,TEMP(m+1,n));
% convert to integer type. Should already be integer
        Pm1=Pf(m+1,round(Op/delta_t)+n,TEMP(1,n));
        LHF=LHFa(m+1);
        TEMP_mp=TEMP_mpa(m+1);
        delta_Q=Pm1+delta_t*((Km1*TEMP(m+2,n)-Km1*TEMP(m+1,n)) ...
                /delta_x(m+1)-(Km*TEMP(m+1,n)-Km*TEMP(m,n))/delta_x(m));
        delta_Temp=delta_Q/Cm1/roum1/delta_x(m+1);
        Temp_old=TEMP(m+1,n);
        Temp_new=delta_Temp+Temp_old;
                if (((Temp_old-TEMP_mp)*(TEMP_mp-Temp_new))>=0)
                    if delta_Q>0            %melting
                        delta_Q=delta_Q-(TEMP_mp-Temp_old)*Cm1*roum1*delta_x(m+1);
                        delta_Qd=max(delta_Q-(LHF-LH(m+1)),0);
```

```
                    LH(m+1)=LH(m+1)+(delta_Q-delta_Qd);
                    TEMP(m+1,n+1)=TEMP_mp+delta_Qd/(Cm1*roum1*delta_x(m+1));
                else
                                                 %solidification
                    delta_Q=delta_Q-(TEMP_mp-Temp_old)*Cm1*roum1*delta_x(m+1);
                    delta_Qd=min(delta_Q+LH(m+1),0);
                    LH(m+1)=LH(m+1)+(delta_Q-delta_Qd);
                    TEMP(m+1,n+1)=TEMP_mp+delta_Qd/(Cm1*roum1*delta_x(m+1));
                    if (TEMP(m+1,n+1)<TEMP_mp)
% will affect the Thermal conductivity only
                        material(m+1)=3;
                    end
                end
            else
                TEMP(m+1,n+1)=Temp_new; %no phase transition
            end
    end
    TEMP(1,n+1)=TEMP(2,n+1);
    TEMP(max_x,n+1)=TEMP(max_x-1,n+1);      % Boundary conditions

end

filename1=sprintf('%d1.dat',round(Op/delta_t));
filename2=sprintf('%d2.dat',round(Op/delta_t));
filename3=sprintf('%d3.dat',round(Op/delta_t));
filename4=sprintf('%d4.dat',round(Op/delta_t));
TEMP_start=TEMP(:,t_array);
LH_start=LH;                                %the final temperature and LH bucket
save(filename3 ,'TEMP_start','-ASCII');
save(filename4 ,'LH_start','-ASCII');
```

Kf.m

```
function [K_return]=Kf(node,t,t1);
% Need 2 parts, otherwise left to right and right to left won't be symmetric
global max_x delta_x K material TEMP_mpa;
if (material(node)==1)
    if(t<TEMP_mpa(material(node)))
        K_return_1=0.02*100;             %P182    W/m/deg
    else
        K_return_1=0.6*100;
    end

elseif material(node)==2
    K_return_1=0.014*100;
elseif material(node)==3
   if(t<TEMP_mpa(material(node)))
        K_return_1=414.49778-1.38372*t+0.00201*t^2-1.43907E-6*t^3 ...
                   +5.01349E-10*t^4-6.76082E-14*t^5;
    else
        K_return_1=0.6*100;
    end
end

if (material(node+1)==1)
    if(t1<TEMP_mpa(material(node+1)))
        K_return_2=0.02*100;             %P182    W/m/deg
    else
        K_return_2=0.6*100;
    end

elseif material(node+1)==2
    K_return_2=0.014*100;
elseif material(node+1)==3
```

```
   if(t1<TEMP_mpa(material(node+1)))
        K_return_2=414.49778-1.38372*t+0.00201*t^2-1.43907E-6*t^3 ...
                   +5.01349E-10*t^4-6.76082E-14*t^5;
    else
        K_return_2=0.6*100;
    end
end
K_return=(K_return_1+K_return_2)/2;
```

Cf.m

```
function [C_return]=Cf(node,t);
global max_x delta_x C material;
if (material(node)==2)
       C_return=1*1000;
else
   C_return=1/6*t+721;        % by fitting J/Kg/K
end
```

Pf.m

```
function [P_return]=Pf(node,t,temp_1)
global k q  eps0 h c I0 x_ind delta_x delta_t xdistance material TEMP_mpa ;

if (temp_1<TEMP_mpa(material(1)))
        R=0.6;            % at 248nm, R=0.6

else
        R=0.6;            % l-Si also has R=0.6

end
alpha=2*10^6*100;         %alpha=alpha0*exp(T/T0)
I=(1-R)*I0(t)*exp(-alpha*(xdistance(node-1)+xdistance(node))/2);
```

```
%Light intensity profile
% (Node-1) for distance not to waste the first slide of power
% <=> having an insulation layer
%Assume rapid phonon emission (1ps)=thermal general profile
P_temp=alpha*I;           %Heat generated/unit lenght /unit time
P_return=P_temp*delta_x(node)*delta_t;
```

Rouf.m

```
function [rou_return]=rouf(node,t);
global max_x delta_x rou material;
if (material(node)==2)
        rou_return=2.27/1000*10^6;
else
        rou_return=2.33/1000*10^6;
end
```