# BACK SIDE CHARGE TRAPPING
# NANO-SCALE SILICON NON-VOLATILE MEMORIES

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

In Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Helena Gomes Silva

August 2005

BACK SIDE CHARGE TRAPPING

NANO-SCALE SILICON NON-VOLATILE MEMORIES

Helena Gomes Silva, Ph.D.

Cornell University 2005

A new alternative device structure for scalable silicon non-volatile memories was investigated. The difficulties in scaling current devices arise from the non-scalability of the gate stack formed by the tunneling oxide, floating gate and control oxide. The proposed device is based on storage of charge in silicon nitride traps in the back of a thin single crystal silicon channel. This is intrinsically different from conventional silicon non-volatile memory structures, in which charge is stored between the silicon channel and the gate.

The devices are fabricated on a modified silicon-on-insulator substrate that employs a stack of silicon oxide – silicon nitride – silicon oxide as the buried insulator. The charge trapping layer, silicon nitride, is separated from the silicon channel by a thin tunneling oxide and from a back gate by a thicker blocking oxide. The device is written and erased by applying an electric field between the back gate and source and drain that causes charge to tunnel between the silicon channel and the trapping layer. When there is no voltage applied, charge is retained in the silicon nitride, hence the non-volatility of the memory. Charges stored in the silicon nitride traps change the potential of the silicon channel resulting in a threshold voltage shift of the device that is sensed using the front gate. The decoupling of the read function (front) from the write and erase functions (back) gives this device a unique advantage in scalability and the ability to operate simultaneously as a high performance transistor and as a non-volatile memory.

Back side charge trapping non-volatile memory devices were demonstrated for the first time. The fabrication process is described and the electrical characteristics are presented. Fabricated devices exhibit memory operation down to 50 nm gate length and double gate operation down to 20 nm gate length. The memory characteristics of the devices, programming times, cycling endurance and retention time are comparable to those of conventional front side storage devices. The new device has the potential to be scaled to 10 nm gate length, a significant improvement from current devices, for higher density and lower power semiconductor non-volatile memory.

.

## BIOGRAPHICAL SKETCH

Helena Silva was born in Lisbon, Portugal. She received the *Licenciatura* degree in engineering physics in 1998 from the Technical University of Lisbon, Portugal. She joined the School of Applied and Engineering Physics of Cornell University in 1998 and received the M.S. and Ph.D. degrees in applied physics in 2002 and 2005 respectively. Her Ph.D. work focused on a new device structure for scalable silicon non-volatile memories. She is interested in semiconductor physics and devices, nano-fabrication techniques and device characterization techniques.

*To my parents, Maria Ester and Joao Luis*

# ACKNOWLEDGMENTS

This work would not have been possible without the help of many people.

I want to thank my advisor, Professor Sandip Tiwari, for all his guidance and help throughout my Ph.D. I also want to thank Professor Lester Eastman and Professor Joel Brock for serving in my exams committee and for their comments on this thesis.

This research was supported by the National Science Foundation through the Cornell Center for Materials Research. I gratefully acknowledge a fellowship from the Foundation for Science and Technology, Portugal, and the European Social Fund through the Third Community Support Framework.

I thank the members of our research group, especially those with whom I worked more closely, Uygar Avci, Soodoo Chae, Ali Gokirmak, Kevin Kim, Moon-Kyung Kim, Arvind Kumar, Hao Lin, Chris Liu, Jeremy Wahl and Lei Xue, for their help in innumerable occasions, fruitful discussions and friendship.

Special thanks to Arvind Kumar for reading my thesis and giving me valuable comments and suggestions. I also thank Chris Liu for reading parts of my thesis.

I thank all the staff members of the Cornell NanoScale Facility for their assistance with the fabrication process. Special thanks to Michael Guillorn, with Cornell NanoScale Facility, and Mick Thomas and John Grazul, with Cornell Center for Materials Research, for the cross-sectional scanning electron microscopy and transmission electron microscopy images presented in this thesis.

I thank my husband, Ali Gokirmak.

I also want to thank my friends, in Ithaca and elsewhere.

Lastly, I want to express my gratitude to my family for their unconditional help, support and encouragement.

TABLE OF CONTENTS

Biographical sketch……...iii

Acknowledgments……...v

Table of contents……vi

List of figures……ix

List of tables……xvi

**Chapter 1 -** Introduction……1

1.1 Semiconductor memory……..1

1.2 Flash memory scaling……3

1.3 Organization……5

**Chapter 2 -** Back-side trapping non-volatile memory devices……..6

2.1 Device structure and principle of operation……6

2.2 Writing and erasing mechanisms……9

2.2.1 Tunneling between the silicon channel and the nitride layer……9

2.2.2 Tunneling between the nitride layer and the control gate……10

2.2.3 Hole injection from the silicon channel into the nitride layer……..13

2.3 Threshold voltage modulation by back side trapped charge……14

2.4 Architecture for back side storage devices……18

2.5 Summary……...19

**Chapter 3 -** Fabrication of back-side trapping memory devices……..20

3.1 Substrate preparation……20

3.2 Transistors fabrication……...27

3.2.1 Optical lithography……..27

3.2.2 Electron-beam lithography……...28

3.2.3 Alignment marks……..29

3.2.4 Active area and device isolation……..30

Mesa isolation……...30

Shallow Trench Isolation (STI)……33

LOCOS isolation……...35

3.2.5 Gate stack deposition and patterning……...38

3.2.6 Thin gate oxide for device scaling……...41

3.2.7 Body and source/drain ion implantation……..43

3.2.8 Metallization……44

3.3 Summary……...45

**Chapter 4 -** Devices characterization……...46

4.1 Transistor characteristics……...46

4.1.1 Front gate transistors operation……46

4.1.2 Double gate operation……..53

4.2 Memory characteristics…….58

4.2.1 Charge injection and removal from the back ONO – memory window..58

4.2.2 Retention time……..64

4.2.3 Cycling endurance……66

4.2.4 Writing and erasing times……69

4.2.5 Small scale memory devices……71

4.3 Summary……...74

**Chapter 5 -** Electron mobility in charge trapping devices……...76

5.1 Effective mobility in MOSFETs……...76

5.2 Mobility at the front and back silicon interfaces……...78

5.3 Mobility at the front interface vs. back gate bias……..82

5.4 Mobility at the front interface in erased and written states……...86

5.5 Effect on the mobility of charge stored in close proximity to the channel…...88

5.6 Summary……...91

**Chapter 6 -** Individual trap characterization using Random Telegraph Signal……...92

6.1 Random Telegraph Signal (RTS)……..92

6.2 RTS measurements and analysis……...94

6.3 Results and discussion……99

6.4 Summary……104

**Chapter 7 –** Summary and future perspectives…..……105

7.1 Summary……105

7.2 Future perspectives……..107

Related publications……109

References……...110

LIST OF FIGURES

**Figure 1.1** Schematics of the two main classes of non-volatile memory: poly-silicon floating gate **(A)** and discrete storage nodes **(B)**. ……..…….3
**Figure 2.1** Schematics of the cross-section of a back side trapping device. The front part is a regular silicon-on-insulator (SOI) transistor. The storage function of the device is achieved with the back part, using a nitride trapping layer separated by the silicon channel through a thin tunneling oxide and by the back gate (control gate) by a thicker control oxide. ..……...6
**Figure 2.2** Band diagrams for a back side trapping device along the gate-to-substrate cross-section, starting from the front gate (most left) to the back gate (most right). ...7
**Figure 2.3** Schematics of the cross-section of a front side trapping device **(A)** and a back-side trapping device **(B)**. …..…….8
**Figure 2.4** Schematics of the write **(A)** and erase **(B)** mechanisms in the SONOS stack using direct tunneling. This applies to both the front side trapping device and the back side trapping device with the high field to write/erase applied between the silicon channel and the respective control gate (front or back). ..……11
**Figure 2.5** Schematics of the write **(A)** and erase **(B)** mechanisms in the SONOS stack using Fowler-Nordheim tunneling. The solid arrows indicate the tunneling across the tunnel oxide and the dashed arrows indicate tunneling across the control oxide between the nitride layer and the control gate. ……..……..12
**Figure 2.6** Band diagrams for a front side trapping device in the erased (solid lines, no charge in the nitride) and written (dashed lines, charge stored in the nitride) states. The threshold voltage shift corresponds to the difference between the dashed and solid lines on the silicon channel. ……....15
**Figure 2.7** Band diagrams for a back side trapping device in the erased (solid lines, no charge in the nitride) and written (dashed lines, charge stored in the nitride) states. The threshold voltage shift measured by the read (front) gate corresponds to the difference between the dashed and solid lines on the front silicon interface. ……..…….15
**Figure 3.1** Process sequence for the substrate preparation. …….21

**Figure 3.2** Typical host wafer after transfer of silicon single-crystal layer from donor wafer. The light areas are the transferred areas. ……...23
**Figure 3.3** AFM images of silicon transferred onto host wafer after exfoliation before (**A**) and after CMP (**B**). The vertical scale is 100 nm/div (A) and 10 nm/div (B). The RMS roughness after exfoliation is 9.3 nm and is reduced to 0.2 nm after CMP. …..24
**Figure 3.4** STEM images of a prepared substrate (single crystal silicon above a ONO stack) after exfoliation. The left image is a low magnification image showing the whole transferred silicon layer with the implantation damage close to the surface. The right image is a high magnification image of the ONO structure. ……...26
**Figure 3.5** STEM image of a cross section of a prepared substrate with a thin ONO stack. The back ONO stack is approximately 3, 4 and 7 nm respectively. ……..27
**Figure 3.6** SEM image of an exposed alignment mark, after being used for alignment with electron beam lithography. ……...30
**Figure 3.7** Mesa isolation process flow. ……..31
**Figure 3.8** AFM micrograph of a fabricated device using electron beam lithography to define both active and gate levels with Mesa isolation. Gate length is ~ 55 nm. ……33
**Figure 3.9** STI isolation process flow. ……34
**Figure 3.10** SEM image of a device active area isolated using STI. ……..35
**Figure 3.11** LOCOS isolation process flow. ……...36
**Figure 3.12** AFM image of a device active area isolated by LOCOS, prior to gate stack deposition. ……...37
**Figure 3.13** SEM image of a small gate length device after gate patterning using FOX 12 prior to the polysilicon etch. The polysilicon grains are visible in this image. …..38
**Figure 3.14** SEM image of a small gate length device after gate polysilicon etch. …39
**Figure 3.15** SEM cross-section of a 0.5 μm back-side trapping device. …….40
**Figure 3. 16.** SEM cross-section of a small back-side trapping device. The gate length is ~ 20 nm. ……40
**Figure 3.17** Thin gate oxide thickness measurement by spectroscopic ellipsometry. 41
**Figure 3.18** Electrical properties of thin gate oxide. **(a)** Capacitance versus gate voltage for a circular capacitor Al/$SiO_2$/p-Si of radius 106 μm. The oxide thickness as determined electrically from the maximum capacitance is 2.4 nm. **(b)** Current density

versus gate voltage for the same capacitor. ……..42
**Figure 3.19** AFM micrograph. Top view of a large device fabricated using optical lithography after vias opening and metal evaporation. ……44
**Figure 3.20** SEM cross-section of a small device. The gate length is ~ 20 nm. …….45
**Figure 4.1** Front channel transistor operation for a back-side trapping device. W = 3 μm, L = 0.75 μm. The silicon body is ~ 40 nm, the back ONO stack is ~ 7/20/80 nm and the front gate oxide is 7 nm. **(A)** Transfer characteristics for different back-gate bias. The solid line corresponds to $V_{BG}$ = 0 V. **(B)** Output characteristics. ……47
**Figure 4.2** Front channel transistor operation for a back-side trapping device. W = 150 nm, L = 200 nm. **(A)** Transfer characteristics. **(B)** Output characteristics. ……50
**Figure 4.3** Front channel transistor operation for a back-side trapping device. W = 200 nm, L = 50 nm. The back ONO stack is ~ 3/4/7 nm and the front gate oxide is 6 nm. **(A)** Transfer characteristics. **(B)** Output characteristics. ……51
**Figure 4.4** Front channel transistor operation for a back-side trapping device. W = 200 nm, L = 20 nm. The silicon body is ~ 15 nm, the back ONO stack is ~ 3/4/7 nm and the front gate oxide is 6 nm. **(A)** Transfer characteristics including the gate leakage current. **(B)** Output characteristics. ……..52
**Figure 4.5** Double-gate transistor operation for a back-side trapping device. W = 20 μm, L = 10 μm. **(A)** Front channel transfer characteristics for different back gate bias. **(B)** Back channel transfer characteristics for different front gate bias. ……...54
**Figure 4.6** Double-gate transistor C-V characteristics for a back-side trapping device. W = 20 μm, L = 10 μm. **(A)** Front gate to source/drain capacitance versus front gate voltage for different back gate bias. **(B)** Back gate to source/drain capacitance versus back gate voltage for different front gate bias. ……55
**Figure 4.7** Double-gate transistor operation for a back-side trapping device. W = 100 nm, L = 50 nm. **(A)** Front channel transfer characteristics for different back gate bias. **(B)** Back channel transfer characteristics for different front gate bias. ……...56
**Figure 4.8** Double-gate transistor operation for a back-side trapping device. W = 20 nm, L = 20 nm. **(A)** Front channel transfer characteristics for different back gate bias. **(B)** Back channel transfer characteristics for different front gate bias. ……...57
**Figure 4.9** Memory operation of a back-side trapping memory in erased (solid line)

and written (dashed line) states with $V_D$ = 1 V. The memory window at 1 nA is 0.68 V. The write and erase voltages were + 50 V and - 35 V 300 ms pulses applied to the back-gate while front-gate, source and drain were grounded. …….59

**Figure 4.10** Effect of back-gate bias in the erased (left set of curves) and written (right set) states of a back-side trapping memory. The same write and erase conditions were used as in Fig. 4.9. ……60

**Figure 4.11 (A)** Memory operation of a back-side trapping memory in erased (solid line) and written (dashed line). The write and erase voltages were + 45 V and - 45 V 300 ms pulses applied to the back-gate while front-gate, source and drain were grounded. **(B)** Effect of the same stored charge on the back channel transistor in the erased (solid line) and written (dashed line) states of the device for comparison between front side and back side storage. ……62

**Figure 4.12** Retention time characteristics for a back-side trapping memory. Threshold voltage as a function of the time elapsed after a 300 ms writing pulse of + 50 V (solid symbols) and after a 300 ms erasing pulse of – 35 V (open symbols) applied to the back-gate. W = 1.5 μm and L = 0.75 μm. ……...……64

**Figure 4.13** Endurance characteristics for a back-side trapping memory. Threshold voltage in the erased (open symbols) and written (solid symbols) states after up to $10^5$ write-erase cycles. W = 3.0 μm and L = 1.0 μm. The write and erase voltages were + 50 V and - 35 V applied to the back-gate for 300 ms. The front-gate, source and drain were grounded during both write and erase operations. ……...67

**Figure 4.14** Endurance characteristics for different write/erase conditions for write/erase optimization. W = 3 μm, L = 0.5 μm. **(A)** Over-writing causes threshold voltage of erased and written states to drift after $10^3$ cycles. **(B)** Over-erasing causes threshold voltage to drift after $10^3$ cycles. **(C)** Memory window approximately stable up to $10^5$ cycles. ……...68

**Figure 4.15** Programming and erasing time characteristics of a back-side trapping device. Threshold voltage shift, $\Delta V_T$, is plotted as a function of writing and erasing time for three different write/erase voltages, +/- 35 V, +/- 37.5 V and +/- 40 V. W = 3 μm, L = 0.5 μm. **(A)** Threshold voltage of the front transistor. **(B)** Threshold voltage of the back transistor. ……...70

**Figure 4.16** Memory operation of a back-side trapping memory device with gate length of 150 nm. The write and erase voltages were + 8 V and – 8 V 300 ms pulses applied to the back gate with front gate, source and drain grounded. ……..71
**Figure 4.17** Memory operation of a back-side trapping memory device with gate length of 100 nm. The write and erase voltages were + 8 V and – 8 V 300 ms pulses applied to the back gate with front gate, source and drain grounded. ……..72
**Figure 4.18 (A)** Memory operation of a 50 nm gate length back-side trapping memory in erased (solid line) and written (dashed line). The write and erase voltages were + 8 V and - 8 V 300 ms pulses applied to the back-gate while front-gate, source and drain were grounded. **(B)** Effect of the same stored charge on the back channel transistor in the erased (solid line) and written (dashed line) states. ……...73
**Figure 5.1** Front silicon channel characteristics, $I_D$-$V_{FG}$ and $C_{FG-SD}$-$V_{FG}$. ……..78
**Figure 5.2** Back silicon channel characteristics, $I_D$-$V_{BG}$ and $C_{BG-SD}$-$V_{BG}$, for the same device as in Figure 5.1. ……79
**Figure 5.3** Effective electron mobility for the front silicon channel derived from the $I_D$-$V_{FG}$ and $C_{FG-SD}$-$V_{FG}$ characteristics in Fig. 5.1. ……...81
**Figure 5.4** Effective electron mobility for the front silicon channel derived from the $I_D$-$V_{BG}$ and $C_{BG-SD}$-$V_{BG}$ characteristics in Fig. 5.2. ……..81
**Figure 5.5** Effective mobility for the front and back silicon channel in the same device (Figs 5.4 and 5.5) plotted as a function of inversion charge density. ……...82
**Figure 5.6** Front channel transfer characteristics $I_D$-$V_{FG}$ for different back-gate voltages. $V_{BG}$ varies from -7 V to +7 V in steps of 1 V. $V_D$ = 50 mV. L = W = 100 μm. ……..83
**Figure 5.7** Front channel $C_{FG-SD}$-$V_{FG}$ characteristics for different back-gate voltages for the same device as in Fig. 5.6. $V_{BG}$ varies from -7 V to +7 V in steps of 1 V. L = W = 100 μm. ……83
**Figure 5.8** Effective mobility for the front channel for different back-gate voltages derived from the $I_D$-$V_{FG}$ and $C_{FG-SD}$-$V_{FG}$ data in Fig. 5.6 and 5.7. $V_{BG}$ varies from -7 V to 4 V in steps of 1 V. ……..84
**Figure 5.9** Front channel peak mobility as a function of back-gate voltage for two different devices. The higher curve corresponds to the data in Fig. 5.8. …….85

**Figure 5.10** Front channel effective mobility in erased (no charge) and written (charge stored in the back ONO) states as a function of the front-gate voltage. W = 0.75 μm and L = 10 μm. ……86
**Figure 5.11** Front channel effective mobility in erased (no charge) and written (charge stored in the back ONO) states plotted as a function of inversion charge density. W = 0.75 μm and L = 10 μm. ……..87
**Figure 5.12** Transfer characteristics $I_D$-$V_G$ (**A**) and effective mobility as a function of the gate voltage (**B**) for a front-side trapping device in the initial state and after different writing times (accumulated). The write voltage was 12 V. The tunneling oxide, nitride and control oxide is ~ 30/70/300 Å. W = L = 2 μm. …….89
**Figure 5.13** Effective mobility in the initial state and after different writing times (Fig. 5.12) plotted as a function of inversion charge density. The peak mobility varies but the high field mobility is independent of trapped charge in the nitride. ……..90
**Figure 5.14** Variation of the peak mobility in Fig. 5.12 with the trapped charged density derived from the threshold voltage shift. …….90
**Figure 6.1** Schematics of a Random Telegraph Signal (RTS) event. When a carrier is captured by a trap located at the silicon – silicon oxide interface or within the silicon oxide, the current level switches from high to low and vice versa when the carrier is emitted back into the channel. ……..93
**Figure 6.2** Output (a) and transfer (b) characteristics for a 50 nm gate length back-side trapping device exhibiting RTS features. The front-gate voltage is 0, 0.2, 0.4 and 0.6 V in (a) and 0, -1, -2 and -3 V in (b). ……...95
**Figure 6.3 (a)** Random Telegraph Signal at the back silicon interface in a back-side storage memory. The physical gate length is 50 nm and the width is 200 nm. The back-gate voltage is increased from -3.2 V to -2.8 V in 0.1 V steps. **(b)** Histograms of the same RTS illustrating the occupation probability of the trap as the back-gate voltage is increased. ……..……...97
**Figure 6.4** RTS trace with the step function obtained from the Matlab code. The step function is used to calculate the statistics of the RTS trace. …….…..…..98
**Figure 6.5** Histogram of the duration of the steps for a particular RTS signal. The steps duration follows a Poisson distribution. ……..……98

**Figure 6.6** Average capture and average emission times ratio as a function of the back-gate bias for the RTS signal shown in Figure 5.3. The position of the trap responsible for this signal is determined from the slope of the fitted line. ……100
**Figure 6.7** RTS in the front interface (oxide only) of a back-side trapping memory device for three different front-gate voltages. ……101
**Figure 6.8** Fast and slow RTS events in a dual oxide device. The upper and lower traces are magnifications of the time windows indicated in the center trace. ……103
**Figure 6.9** Multi-level RTS signal at the back interface (ONO) of a back-side trapping device. $V_{BG}$ = - 1.45 V, $V_{FG}$ = -2 V, $V_D$ = 10 mV and $V_S$ = 0 V. The inset shows three levels of the signal. ……103

LIST OF TABLES

**Table 3.1** Smart-Cut parameters used for Silicon on ONO transfer. ……..25

**Table 6.1** RTS signals observed in different gate stack structures. Oxide only traps are faster than in oxide-nitride-oxide stack or oxide-oxide stack. ……...101

# Chapter 1

# Introduction

## 1.1 Semiconductor memory

Semiconductor memory can be divided into two main types, both based on CMOS technology, volatile and non-volatile memory. Volatile memory is fast but loses its contents when power is removed. Non-volatile memory is slower but retains the information without power supplied.

Volatile memories are SRAM and DRAM (Static and Dynamic Random Access Memory). SRAM is the fastest type of semiconductor memory, with write/read times in the range of 1-10 ns. An SRAM cell, which stores one bit of information, is usually made of 6 transistors. As a result, SRAM is the most expensive and lowest density memory and is only used for the highest performance applications such as memory cache. A DRAM cell is made of only one transistor and one capacitor and provides very dense memory. The write/read times for DRAM are in the order of 50 ns. Due to its relatively high speed, low cost and high density DRAM is used in a broad range of applications and is the largest fraction of the semiconductor memory market.

Most of the semiconductor non-volatile memory used today is referred to as *flash* memory. The name derives from the way in which cells are erased in an array (a large number of cells are erased at once). Other types of semiconductor non-volatile memory are ROM (read only memory), EPROM (electrically programmable ROM) and EEPROM (electrically erasable and programmable ROM). Flash memory developed from these and, as a result of its programming and erasing mechanisms,

combines the high density of EPROM (one transistor per cell) with the flexibility of electrical program and erase of EEPROM. Flash memory, and all semiconductor non-volatile memory, is slow, compared to SRAM or DRAM. The fastest write times are in the order of μs and the erase times are in the order of ms. The read time in flash memory is comparable to that of DRAM, sub-100 ns. The non-volatility and high density of flash memory give it a wide window of opportunities from code storage to mass data storage. It is present in virtually all portable electronic devices and is used for most of the memory cards. As a result, it makes already for more than half of the DRAM market and is currently the fastest growing memory segment. Each cell is made of a single transistor which has a floating node between the gate and the channel. Charge can be stored in this floating node and determines the state of the memory by changing the threshold voltage of the transistor. The major challenge for flash memory is scaling to smaller dimensions for denser lower power non-volatile memory. The current device structure (explained in more detail in the next section) makes scaling beyond 65 nm gate length very complicated and seemingly impossible beyond 32 nm gate length (expected to be reached by the end of this decade) [1]. Density can still be increased using the current device structure, making use of multi-level storage. The ability to precisely control the amount of injected charge allows more than two clearly distinct threshold voltage states [2]. Further scaling to higher density and lower power will require the use of new materials and probably the change to new device structures.

Current alternatives to transistor based charge storage memory are MRAM (magnetic RAM), FeRAM (ferroelectric RAM) and Phase Change Memory. Although with attractive properties such as power and speed, all at this point, appear to have important limitations regarding density or CMOS integration. Among the three, Phase Change Memory seems to be the most promising candidate since the phase change

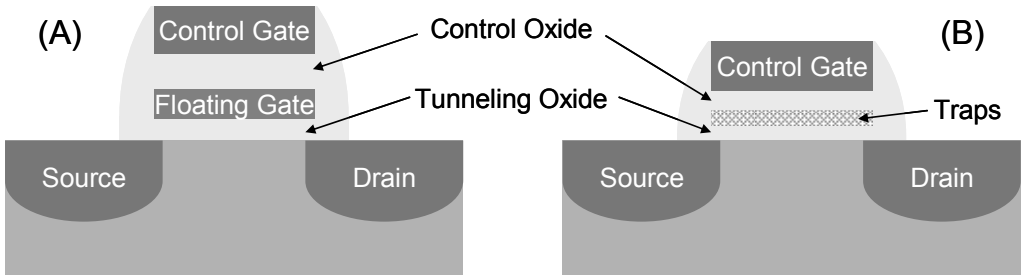



**Figure 1.1** Schematics of the two main classes of non-volatile memory: poly-silicon floating gate **(A)** and discrete storage nodes **(B)**.

material is compatible with backend CMOS processing and the cell can be scaled to very small areas. The main problem with Phase Change Memory is the requirement of high current to change the phase of the material. In any case, these technologies are still far from maturity and do not present an alternative to conventional flash memories in the near future [1].

## 1.2 *Flash* memory scaling

Figure 1.1 shows the schematic cross-section of the two main types of flash memory: polysilicon floating gate (A) and discrete storage nodes (B). The polysilicon floating gate device was proposed by Kahng and Sze in 1967 [3] and the discrete storage nodes device, making use of traps in silicon nitride, was proposed by Wegener *et al.* also in 1967 [4]. The injection and extraction of charge is done by tunneling across an insulator barrier, which results in the slow write and erase times. Charge can tunnel across this barrier (tunneling oxide) only when a sufficiently large voltage is applied. When there is no voltage applied the charge is retained in the floating gate, hence its non-volatility (charge is retained with no power supplied). Charge leakage into the control gate is prevented by a thicker insulator barrier (control oxide). The

charge stored in the floating gate causes a threshold voltage shift of the transistor and the state of the memory is read by sensing the current at a gate voltage between the two states threshold.

The floating gate device (Fig. 1.1 A) became the standard non-volatile memory until recently when discrete storage nodes devices, based on storage in traps in silicon nitride or in semiconductor or metal nanocrystals [5] started receiving increasingly more attention as more scalable devices. Isolated storage nodes can be placed closer to the channel (Fig. 1.1 B), resulting in a thinner gate stack with which smaller gate length devices can be implemented. This leads to higher density and lower voltage/power operation. The difficulties in making smaller silicon non-volatile memories with the current device structure, floating gate or discrete storage nodes, arise from non-scalability of the gate stack (tunneling oxide, storage medium, and control oxide) as gate length is reduced. The tunneling and control oxide cannot be thinned as required for smaller gate lengths and lower voltages operation without compromising the charge retention and the reliability of the memory.

This thesis investigates one alternative for flash memory scaling, based on back side trapping storage. The charge is stored in silicon nitride traps in the back of a thin silicon layer, between the silicon channel and a back gate. The use of thin fully depleted silicon-on insulator (SOI) together with thin buried insulator has attractive scaling properties since both interfaces of the channel are gated. As a result, the channel potential is better controlled and the device can be scaled to smaller gate lengths. Back side storage was proposed in 2001 by Kumar and Tiwari [6] as an alternative for flash memory scaling. The concept was demonstrated in 2004 by Avci *et al.* using a back polysilicon floating gate [7]. Back side trapping storage combines the advantages of discrete storage nodes, mentioned above, with this new more scalable geometry.

## 1.3 Organization

The organization of the thesis is as follows. In Chapter 2 the structure and the principle of operation of back side storage memory devices are explained, in comparison to conventional front side storage memory devices. Chapter 3 describes the fabrication of the devices using standard CMOS techniques. The unique part in the fabrication of the devices is the substrate preparation which involves placing a charge trapping layer underneath a thin silicon channel. The electrical characteristics of the fabricated back side trapping memories, transistor and memory operation, are shown in Chapter 4. Chapters 5 and 6 investigate two topics related to transport in the front and back silicon interfaces in back side trapping memories. In chapter 5 the electron mobility, in the front and back silicon interfaces, and with and without charge stored in the nitride, is studied. In Chapter 6 Random Telegraph Signal is used to determine the position of individual traps that affect the conduction in the front and back interface of the devices. A summary of the work and future perspectives for semiconductor non-volatile memory devices are given in Chapter 7.

# Chapter 2

## Back side trapping non-volatile memory devices

In this chapter the principle of back side storage for silicon non-volatile memories is explained. In these devices the charge is stored in a trapping layer formed by an ONO stack (silicon oxide - silicon nitride - silicon oxide) that is placed on the back of a thin single crystal silicon channel. The characteristics of back side storage are compared to those of conventional front side storage.

## 2.1 Device structure and principle of operation

Figure 2.1 shows a schematic cross section of a back side trapping device. It is a silicon-on-insulator (SOI) transistor in which a charge trapping layer (silicon nitride in this case) is placed within the buried insulator.

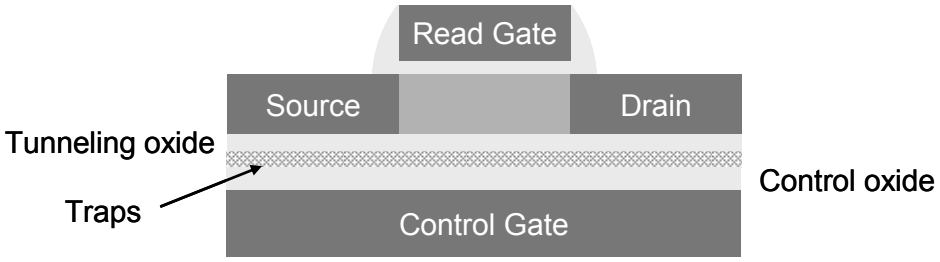



**Figure 2.1** Schematics of the cross-section of a back side trapping device. The front part is a regular silicon-on-insulator (SOI) transistor. The storage function of the device is achieved with the back part, using a nitride trapping layer separated by the silicon channel through a thin tunneling oxide and by the back gate (control gate) by a thicker control oxide.

The combined use of a thin silicon channel, thin front gate oxide and thin back ONO stack, with the possibility of storage in the back, gives this device unique scaling possibilities. Silicon nitride is used as a charge trapping layer because of its high density of traps, $10^{12} – 10^{13}$ $cm^{-2}$. Figure 2.2 shows the band diagrams for this structure, along the gate-to-substrate cross-section, in equilibrium, when there is no voltage applied between the back gate and the front gate and no charges stored in the nitride layer. When charge is stored in the nitride layer the potential in the silicon channel changes and causes a threshold voltage shift of the device.

The characteristics of back side storage are compared to those of front side storage which is the conventional geometry for silicon non-volatile memories. Figure 2.3 shows the schematic cross-sections of a front side trapping device (A) and a back side trapping device (B). As shown in the figure, in a front side storage device, the charge is stored between the gate and the silicon channel in a poly-silicon floating gate or a trapping medium. The charge is separated from the gate by a control (or blocking)

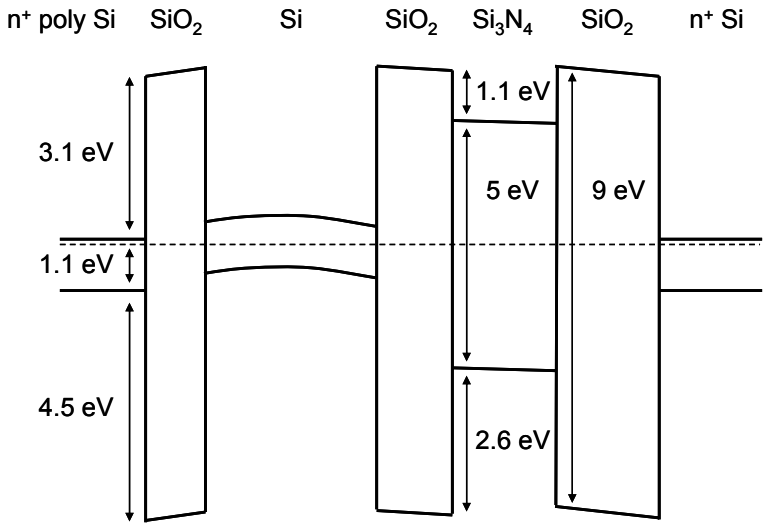



**Figure 2.2** Band diagrams for a back side trapping device along the gate-to-substrate cross-section, starting from the front gate (most left) to the back gate (most right).

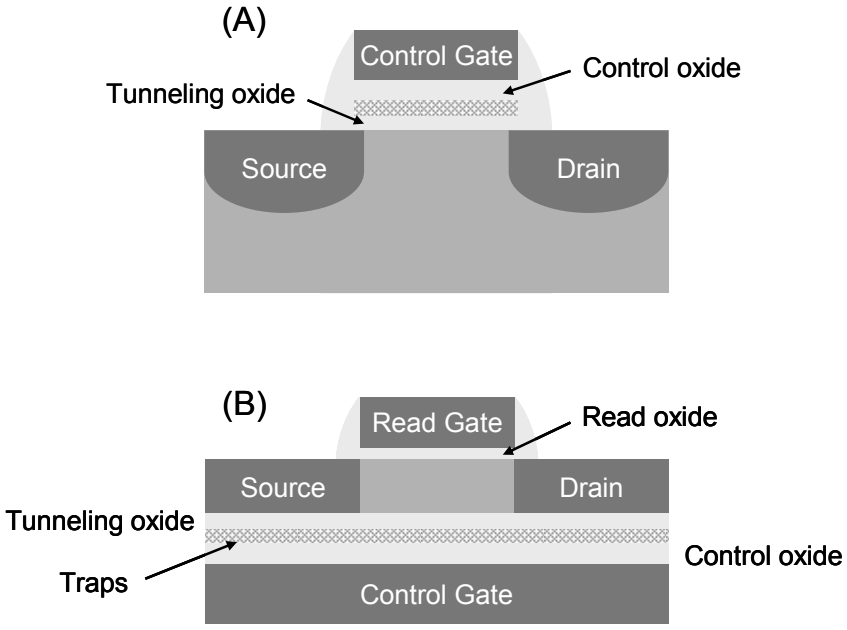



**Figure 2.3** Schematics of the cross-section of a front side trapping device **(A)** and a back-side trapping device **(B)**.

oxide and from the silicon channel by a thin tunneling oxide. The same gate is used to write and erase the device, using high voltage, and to read it, using low voltage. In the case of a back side storage there are two gates which allow the separation of the read and the write/erase functions: the front gate, separated from the channel by a thin oxide, is used to read the state of the device, at low voltage, and the back gate, separated from the channel by a thicker stack composed of the storage medium, the tunneling oxide and the control (blocking) oxide is used to write and erase the device, at higher voltage. Decoupling the charging mechanisms that lead to the memory function from the front gate transistor operation allows efficient scaling of the front gate without compromising the memory characteristics of the device that depend only on the back insulating films stack.

By applying a large voltage between the back-gate and the three front

terminals, charge can tunnel through the tunneling oxide between the silicon channel and the nitride layer. When there is no large voltage applied the charge cannot tunnel back into the silicon channel resulting in the non-volatility of the storage (charge is retained when power is removed).

## 2.2 Writing and erasing mechanisms

The write and erase mechanisms for back side storage are the same as those for front side storage with the control gate being the back gate in the case of back side storage. The processes involved in the writing and erasing of the devices are briefly described in this section. The physics of the different charge injection and removal processes are well known and can be found, specifically in relation to non-volatile memory devices in *Flash Memories* edited by Cappellotti *et al.* [8].

### 2.2.1 Tunneling between the silicon channel and the nitride layer

Tunneling across the injection oxide (or tunneling oxide), between the silicon channel and the nitride layer, is the dominant process during write and erase. Charge can be injected into the nitride traps using Fowler-Nordheim tunneling, direct tunneling or hot carrier injection (HCI), or a combination of these mechanisms, and it can be removed from the nitride traps by Fowler-Nordheim tunneling or direct tunneling. To write the device (inject electrons into the traps) a large positive voltage is applied to the control gate (back gate) while the front terminals (front-gate, source and drain are grounded). Electrons in the silicon inversion layer can tunnel into the unoccupied traps in the silicon nitride. To erase the device a large negative voltage is applied to the control gate while keeping the front terminals grounded and electrons tunnel back into the silicon channel.

Direct tunneling occurs for thin tunneling barriers, ~ 3 nm and below. Figure 2.4 shows the schematics of the write and erase mechanisms using direct tunneling. Fowler-Nordheim tunneling can occur through a thicker barrier on application of higher voltages that effectively reduce the barrier width by causing a triangular barrier for injection. Figure 2.5 shows the schematics of the write and erase mechanisms using Fowler-Nordheim tunneling. With hot carrier injection, electrons accelerated by a large drain voltage tunnel across the tunneling oxide near the drain end. Due to the higher energy of the electrons (hot carriers) this mechanism requires lower control gate voltage for injection. Direct tunneling is the preferred process for small scale devices due to lower voltages required, when compared to Fowler-Nordheim tunneling, and reduced tunneling oxide degradation, when compared to hot carrier injection.

### 2.2.2 Tunneling between the nitride layer and the control gate

During write and erase using Fowler-Nordheim tunneling, besides the intended tunneling between the silicon channel and the trapping layer, tunneling between the trapping layer and the control gate can also take place (see Fig. 2.5 where tunneling across the control oxide is indicated by dashed arrows). During the *write* process electrons can tunnel from the traps into the control gate and during the *erase* process from the control gate into the traps, in both cases countering the intended effect of increase or decrease of charge in the nitride. In poly-silicon floating gate devices this can be avoided by adjusting the ratios of the capacitances between the channel and the floating gate and between the floating gate and the control gate to ensure that the field in the control oxide is smaller than in the tunneling oxide thus preventing tunneling into the control gate. With silicon nitride or other discrete storage mechanism this

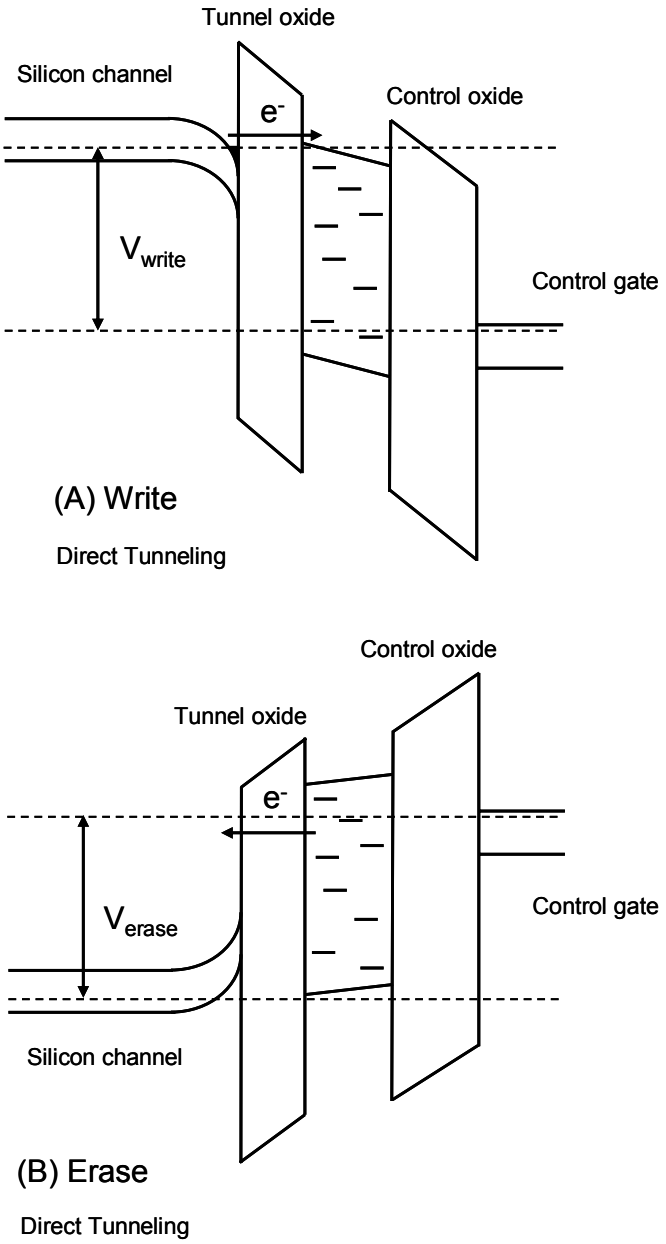



**Figure 2.4** Schematics of the write **(A)** and erase **(B)** mechanisms in the SONOS stack using direct tunneling. This applies to both the front side trapping device and the back side trapping device with the high field to write/erase applied between the silicon channel and the respective control gate (front or back).

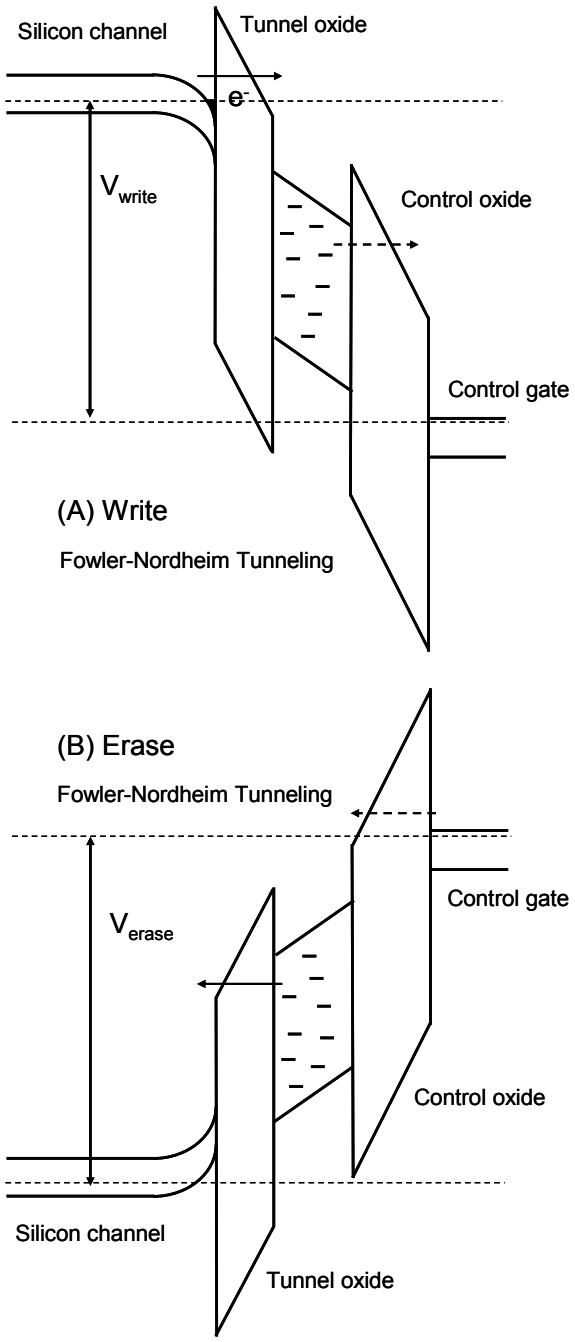



**Figure 2.5** Schematics of the write **(A)** and erase **(B)** mechanisms in the SONOS stack using Fowler-Nordheim tunneling. The solid arrows indicate the tunneling across the tunnel oxide and the dashed arrows indicate tunneling across the control oxide between the nitride layer and the control gate.

cannot be done since the charge is only stored under the gate and the area ratios cannot be adjusted. This is another reason, besides the lower voltage operation, why direct tunneling is the preferred process for injection/extraction with discrete storage centers. Fowler-Nordheim tunneling can still be used with discrete storage devices because initially, for both write and erase processes, tunneling across the tunneling oxide is more favorable than tunneling across the control oxide. During *write* there is only appreciable tunneling between the control gate and the nitride layer once there are enough available electrons in the traps to tunnel and, during *erase*, only when there are enough empty traps for electrons from the control gate to occupy. The field in the control oxide also favors tunneling into/from the control gate for long write/erase times due to trapped charge increase/decrease. During write, as more charge is trapped, the field in the tunneling oxide decreases and the field in the control oxide increases. During erase the reverse happens. As a result, for long write and erase times the tunneling between the control gate and the trapping layer dominates and a decrease of the stored charge (for write) or an increase of stored charge (for erase) takes place. This is observed in devices where Fowler-Nordheim tunneling is used due to a thick tunneling barrier (Chapter 4).

### 2.2.3 Hole injection from the silicon channel into the nitride layer

Other process that can take place in these devices is the injection of holes from the silicon channel into the nitride traps during erase. In Fig. 2.4 B it can be seen that if there are traps available for holes occupation in the silicon nitride band gap, holes can tunnel from the accumulation layer in the silicon-tunneling oxide interface into these traps. Unlike the previous effect injection of holes during ‘erase’ adds to the intended extraction of electrons from the traps in the sense that the threshold voltage of the device is further reduced, in this case by the injection of positive charge.

Nevertheless, since the retention time for electrons and holes can be different, the involvement of holes in the storage process is not desirable. This must be taken into account in the choice of programming voltages and times employed.

## 2.3 Threshold voltage modulation by back side trapped charge

Figure 2.6 and Figure 2.7 illustrate through band diagrams schematics the threshold voltage shift that is measured once charges have been injected into the nitride traps, for a front side trapping device and for a back side trapping device, respectively. The solid lines in these diagrams correspond to the erased state (no charges in the nitride traps) and the dotted lines correspond to the written state (with charge in the traps). In the front side trapping device, negative charges stored in the silicon nitride layer cause the silicon interface into accumulation (see Fig. 2.6). This accumulation layer can sustain a large charge density and the amount of charge and therefore threshold voltage shift that can be stored is only limited by the breakdown of the tunneling oxide. The threshold voltage difference between the erased and the written state in the case of front side trapping, $\Delta V_{T,fr}$ (the extra gate voltage that has to be applied in order to achieve the same inversion level at the silicon interface) is given by the simple capacitive coupling between the stored charge and the front gate:

$$\Delta V_{T,fr} = -\frac{\Delta Q_{tr}}{C_{tr-fg}} = -\frac{\Delta Q_{tr}\, t_{ins}}{\varepsilon_{ins}} \qquad (1)$$

where $\Delta Q_{tr}$ is the trapped charge density difference between the erased and written state, and $\varepsilon_{ins}$ and $t_{ins}$ are the equivalent dielectric permittivity and thickness of the insulator between the traps where the charge is stored and the front gate. If, for

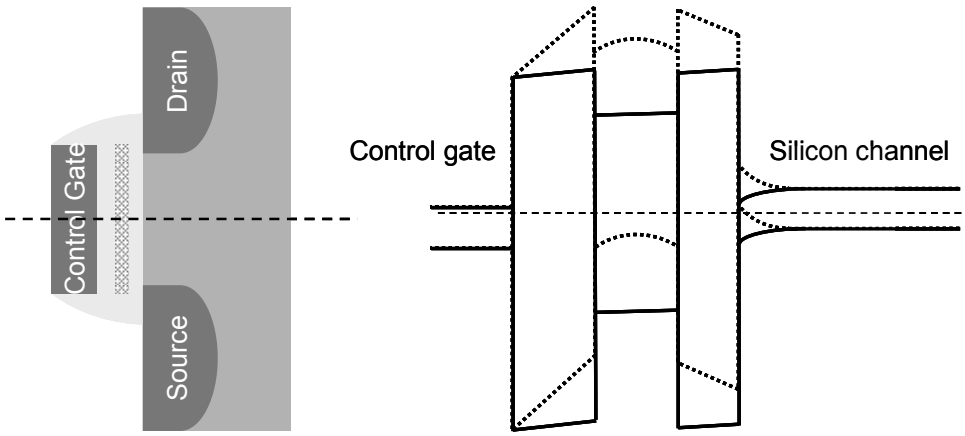



**Figure 2.6** Band diagrams for a front side trapping device in the erased (solid lines, no charge in the nitride) and written (dashed lines, charge stored in the nitride) states. The threshold voltage shift corresponds to the difference between the dashed and solid lines on the silicon channel.

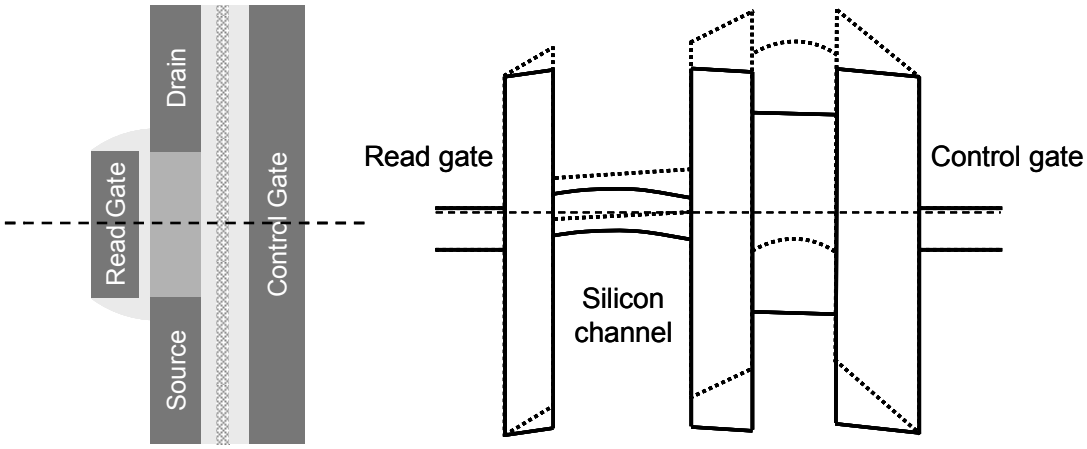



**Figure 2.7** Band diagrams for a back side trapping device in the erased (solid lines, no charge in the nitride) and written (dashed lines, charge stored in the nitride) states. The threshold voltage shift measured by the read (front) gate corresponds to the difference between the dashed and solid lines on the front silicon interface.

simplicity, we assume that all the charge is stored within the nitride, at a distance $t_{tr\text{-}co}$ from the nitride – control oxide interface, then the expression for the threshold voltage shift can be written as:

$$\Delta V_{T,fr} = -\Delta Q_{tr}\left(\frac{t_{tr-co}}{\varepsilon_{nit}} + \frac{t_{co}}{\varepsilon_{ox}}\right) \qquad (2)$$

where $t_{co}$ is the control oxide thickness. The threshold voltage shift for a front side trapping device (or front floating gate) is typically in the order of a few Volts.

In a back side trapping device (or floating gate) the threshold voltage shift on the front transistor due to charge stored in the back, $\Delta V_{T,bk}$, is not due to the direct effect of the charge stored in the nitride but to the change in potential of the back silicon interface (see Fig. 2.7). Once a charge layer forms at the back interface (either accumulation or inversion) the potential of the back interface is pinned and further increase in stored charge will not affect the potential of the front interface, and therefore the threshold voltage of the front transistor. The threshold voltage increase of the front transistor is therefore limited by a total variation of ~ 1 V, the silicon band gap. This is the same effect as a back gate bias in a silicon-on-insulator (SOI) transistor. It is possible to achieve an effective larger memory window not through a threshold voltage shift of the front transistor but through inversion of the back interface due to positive charge stored in the back. This causes the channel to conduct even at zero or negative front gate bias, in an effective low threshold voltage state.

A simple model based on capacitive coupling between the front (read) gate and the trap layer below the silicon channel, assuming a uniform (weighted by the dielectric permittivity of each layer) voltage drop between the front gate and the trapping layer, can be used to determine $\Delta V_{T,bk}$ as a function of the density of charge stored in the back. The potential of the front silicon channel can be approximated by:

$$V_{ch} = V_{fg} - \eta \left( V_{fg} - V_{tr} \right) \quad , \quad \eta = \frac{\frac{t_{ro}}{\varepsilon_{ox}}}{\frac{t_{ro}}{\varepsilon_{ox}} + \frac{t_{Si}}{\varepsilon_{Si}} + \frac{t_{to}}{\varepsilon_{ox}} + \frac{d_{tr-to}}{\varepsilon_{nit}}} \qquad (3)$$

where $V_{ch}$, $V_{fg}$ and $V_{tr}$ are the potentials at the front silicon channel, the front gate and the trapping layer, and $t_{ro}$, $t_{Si}$, $t_{to}$ are the thickness of the read oxide, the silicon body, and the tunneling oxide respectively. $d_{tr\text{-}to}$ is the distance between the trapped charge layer and the nitride – tunneling oxide interface, again assuming for simplicity that all the charge is within the nitride, concentrated at the same depth. The density of trapped charge can be related to the front gate and the back gate potentials by:

$$Q_{tr} = C_{fg-tr} \left( V_{tr} - V_{fg} \right) + C_{bg-tr} \left( V_{tr} - V_{bg} \right) \qquad (4)$$

From (3) and (4), and for the case when $V_{bg}$ = 0 V during the read operation, the threshold voltage shift of the front transistor due to charge stored in the back ONO, $\Delta V_{T,bk}$, can be written as:

$$\Delta V_{T,bk} = -\frac{\eta \, \Delta Q_{tr}}{C_{fg-tr} + (1-\eta) \, C_{bg-tr}} \qquad (5)$$

where $C_{fg\text{-}tr}$ is the capacitance between the front gate and the trapping layer and $C_{bg\text{-}tr}$ is the capacitance between the back gate and the trapping layer.

This is only an approximate expression for the bias range in which the back interface is not in accumulation or inversion and that neglects depletion charge effects in the silicon channel. Accurate potential profiles in the device can be numerically calculated using Poisson's equation. This was done for back floating gate memory devices by

Kumar *et al.* [9]. From measurements on back side trapping devices the amount of charge trapped in the back ONO can be determined accurately by measuring the threshold voltage shift on the back silicon interface. This is equivalent to a front side trapping device threshold voltage shift, $\Delta V_{T,fr}$, since the charge is stored between the silicon interface and the read gate, in that case, and the charge can be determined using (1).

## 2.4 Architecture for back side storage devices

If individual back gates are used, the same architecture schemes that are used with front side storage devices can be used with back gate storage devices. The control gates (back gates in the case of back side storage) are connected to the word lines and source and drain are connected to the bit lines. In the case of NOR architecture HCI or direct or Fowler-Nordheim tunneling can be used to write and direct or Fowler-Nordheim tunneling is used to block erase. If the whole substrate is used as a common back gate to all devices, which makes for a much simpler fabrication process, an addressable memory array can be implemented using Hot Carriers Injection (HCI) for write and direct or Fowler-Nordheim tunneling for block erase. Kumar *et al.* [9] and Avci *et al.* [7] have proposed two different ways in which this can be done.

## 2.5 Summary

The back-side trapping memory device has unique properties and promising applications in logic and memory integration since it is both a scalable memory and a scalable SOI transistor in the same structure, as a result of the decoupling of the read and the write/erase functions. The front transistor can be scaled to very short gate lengths (just like a regular SOI transistor) without compromising the memory characteristics of the device. During the write/erase, with high voltages, the field in the front oxide is minimized due to inversion/accumulation layers at the back silicon interface that shield the front transistor from the large fields in the back. Since the charge is stored between the silicon channel and the back-gate, and the device state is read using the front-gate, a much smaller read disturbance is to be expected in these devices when compared to front-side trapping memory cells. For the same ONO thickness for front side and back side trapping memories, the smaller read-disturb effect will, in principle, lead to better overall retention and reliability characteristics of the back-side trapping devices.

# Chapter 3

## Fabrication of back-side trapping memory devices

This chapter describes the fabrication of back-side charge trapping memory devices using a modified Smart-Cut substrate preparation process followed by standard CMOS processing with mixed (optical and electron-beam) lithography. The substrate is a complex silicon-on-insulator (SOI) substrate where instead of oxide alone a charge trapping multi-layer stack of oxide-nitride-oxide (ONO) is used as the buried insulator. Smart-Cut using ultra thin buried insulator is demonstrated. This capability of Smart-Cut is utilized to fabricate back side trapping memories with thin back gate dielectric stacks for low voltage and low power operation. Small scale devices, down to ~ 20 nm physical gate length, were fabricated.

## 3.1 Substrate preparation

Smart-Cut is a process to transfer a single crystal silicon layer onto a certain substrate, typically an oxidized silicon wafer. This technique is based on hydrogen ion implantation, wafer bonding and subsequent exfoliation through the implanted hydrogen region (the detailed process is explained below). Smart-Cut was introduced in 1995 by Bruel *et al.* [10] and in recent years has become a standard technique to fabricate SOI wafers due to its relatively simple and inexpensive process, good quality interfaces and scalability to larger substrates. Other techniques to produce SOI include epitaxial growth of silicon on insulator, re-crystallization of a deposited amorphous or polycrystalline silicon layer on insulator or bonding of two silicon wafers through an insulator and grinding or etching-back one of the wafers [11].

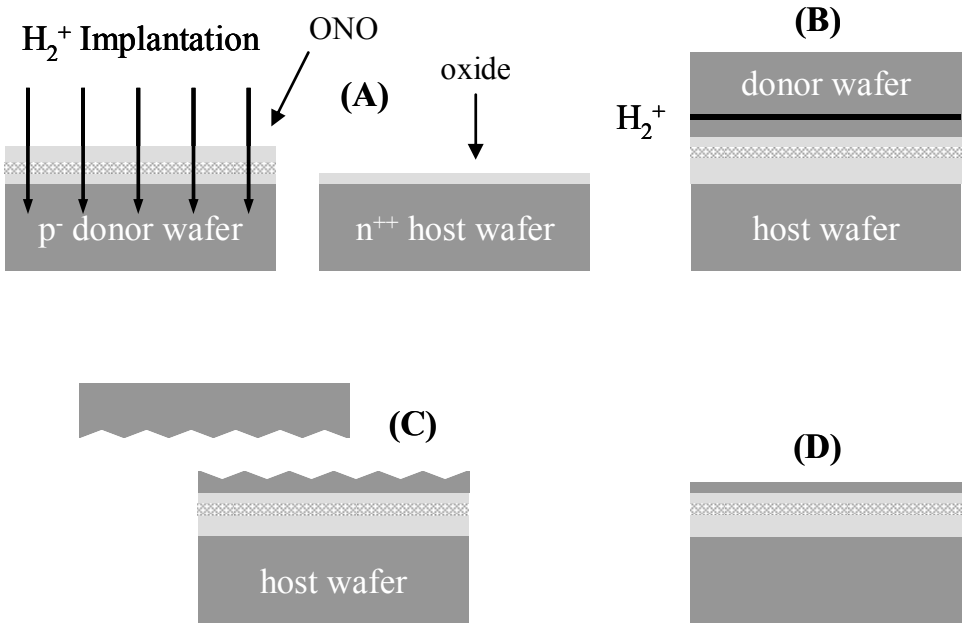



**Figure 3.1** Process sequence for the substrate preparation. **(A)** Starting wafers: (i) a donor $p^-$ wafer with a ONO stack is implanted with high dose of hydrogen (ii) a 'host' $n^{++}$ wafer with a thin thermal oxide. **(B)** The two wafers are bonded and annealed at low temperature, 250 $^0$C. **(C)** Exfoliation takes place through the hydrogen layer at 400 C, leaving a thin single-crystal silicon layer above a ONO stack onto the host wafer. The silicon surface after exfoliation is relatively rough with 6 - 9 nm RMS roughness as measured with AFM. **(D)** The single-crystal silicon layer can then be smoothed and thinned to the final desired thickness using CMP and sacrificial oxidation steps.

Figure 3.1 illustrates the Si on ONO substrate fabrication process sequence based on Smart-Cut. We start with two prime quality silicon wafers, a low doped p-type (or n-type) wafer, the *donor wafer*, and a heavily doped n-type wafer, the *host wafer* (Fig. 3.1 A). On the donor wafer a thin thermal silicon oxide is grown followed by a low pressure chemical vapor deposited (LPCVD) silicon nitride film and a low temperature (LTO) or high temperature (LPCVD) deposited silicon oxide. These films form the tunneling oxide, the charge trapping medium (the nitride layer and the

tunneling oxide – nitride interface) and part of the control oxide, respectively. Following the formation of the ONO stack, the 'donor wafer' is implanted with high dose of $H_2^+$. On a second silicon wafer, the 'host wafer', a thin thermal silicon oxide layer is thermally grown. The two wafers are directly bonded at low pressure at room temperature using a wafer bonder with a force of 1000 N (Fig. 3.1 B) and the bond is strengthened through a low temperature anneal, 250 C, in nitrogen ambient, for 12 hours. A very smooth surface, ~ 0.2 nm RMS roughness as measured with atomic force microscopy (AFM), obtainable either by chemical mechanical polishing (CMP) or by growth of a thin oxide (less than ~ 20 nm) on prime substrates, is a pre-requisite for formation of a strong bond between the ‘donor’ and the ‘host’ wafers. If the oxide layers on either the host wafer or the donor wafers are thicker than ~ 20 nm CMP is needed prior to bonding in order to achieve the necessary atomic smoothness. Thinner oxide films, both thermally grown and low temperature (LTO) or high temperature deposited (LPCVD), have RMS roughness less than 0.5 nm and can be bonded directly.

The control oxide in the device is the combination of the deposited oxide on the donor wafer and the grown oxide on the host wafer. It was found that the oxide layer in the host wafer can be thermally grown, low temperature LPCVD, or high temperature LPCVD, all resulting in good quality bond between the host and donor (that has LTO or LPCVD oxide, deposited immediately after the nitride deposition). PECVD oxide, deposited in the CNF GSI tool on the host wafer, does not produce a good bond, probably due to particle contamination.

After the low temperature anneal, the temperature is raised to 400 C, and the hydrogen micro cavities, located at the projected range of hydrogen ions in the donor wafer, cause the donor wafer to cleave leaving a rough (~10 nm RMS roughness) single crystal silicon layer bonded onto the host wafer (Fig. 3.1 C). Figure 3.2 shows a

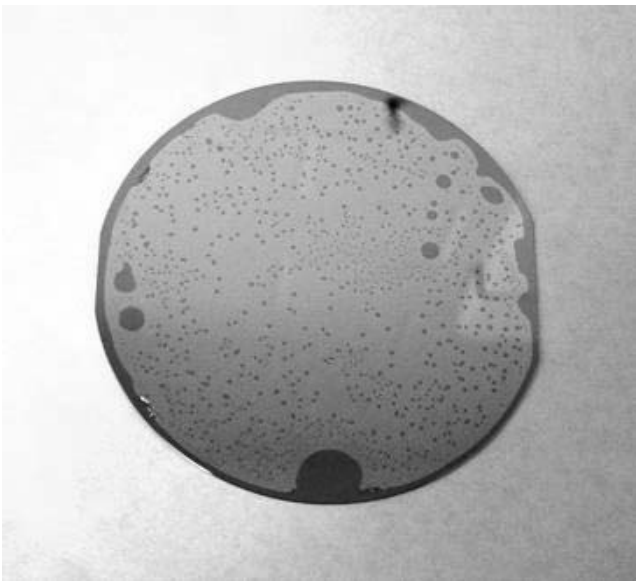
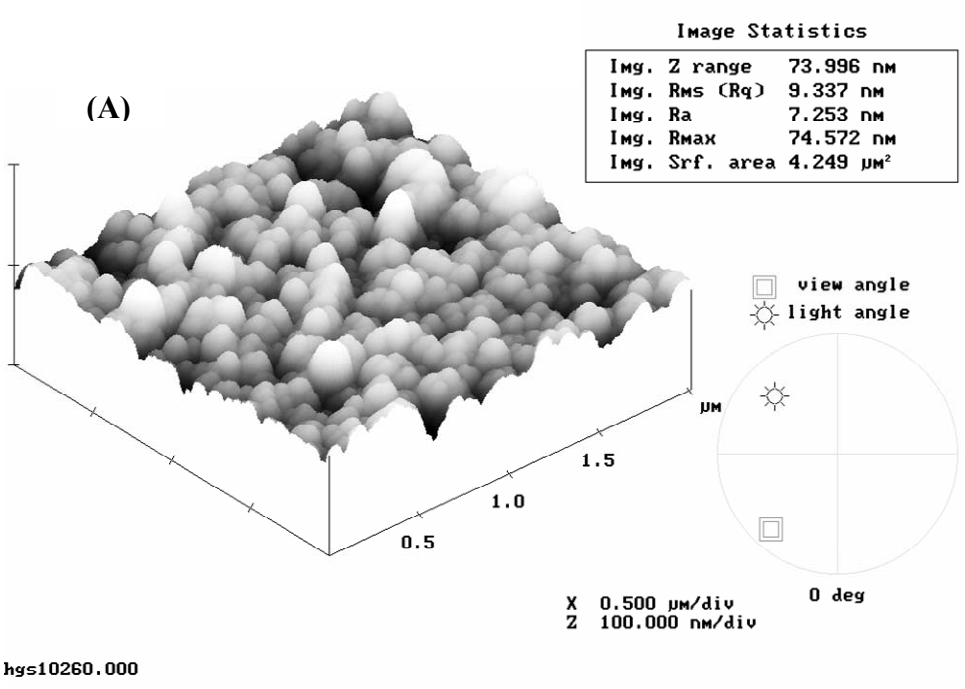


**Figure 3.2** Typical host wafer after transfer of silicon single-crystal layer from donor wafer. The light areas are the transferred areas. More than 90% of the area is transferred with small non transferred areas ('bubbles') within the transferred areas.

typical prepared substrate after exfoliation with ~ 90% area transfer. The dark areas are 'bubbles' where silicon did not get transferred. This is due to particle contamination that prevents the bond in these areas and it is not an intrinsic problem of Smart-Cut, as can be attested by the commercially available substrates. Details on the kinetics of the Smart-Cut splitting can be found in the work of Aspar et al. [12]. Figure 3.3 A shows an AFM image of a silicon surface after exfoliation. This rough surface is then smoothed by CMP and oxidation to achieve device quality silicon surface and the desired final thickness (Fig. 3.1 D and Fig. 3.3 B). Once the surface has been smoothed by CMP the silicon thickness can be reduced to the desired value by successive thin oxidations followed by hydrofluoric acid (HF) removal of these oxides. If these successive oxide films are thin enough, again less than ~ 20 nm, this procedure does not degrade the surface smoothness achieved by CMP.



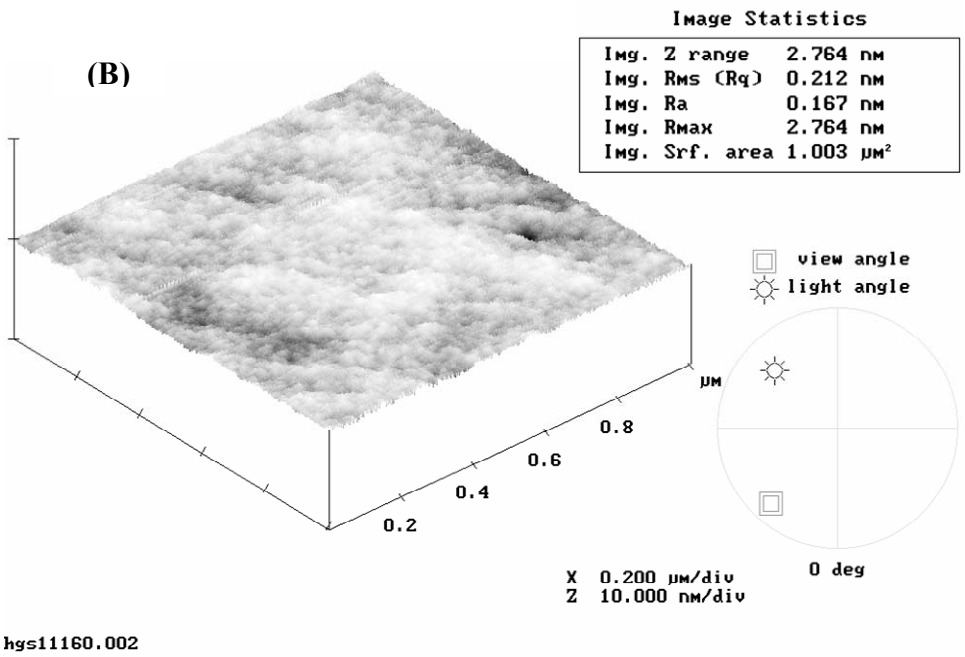




**Figure 3.3** AFM images of silicon transferred onto host wafer after exfoliation before (**A**) and after CMP (**B**). The vertical scale is 100 nm/div (A) and 10 nm/div (B). The RMS roughness after exfoliation is 9.3 nm and is reduced to 0.2 nm after CMP.

In our processes, final 40 nm and 15 nm (average values) layers of silicon were achieved after CMP from initial 600 nm and 120 nm thick silicon layers obtained after the hydrogen-induced exfoliation. Different conditions for the hydrogen implantation and the respective (Si + ONO) thickness transferred that were used are listed in Table 3.1. O/N/O back-trapping stacks of 7/20/80 nm, 7/15/40, and 3/4/7 nm were used in different runs as the devices were scaled down.

**Table 3.1** Smart-Cut parameters used for Silicon on ONO transfer.
All the implantations were done at $7^0$ tilt.

| Dose ($cm^{-2}$) | Species | Energy (KeV) | Transferred Si + ONO Thickness (nm) |
|---|---|---|---|
| $3 \times 10^{16}$ | $H_2^+$ | 140 | ~ 650 |
| $6 \times 10^{16}$ | $H^+$ | 70 | ~ 650 |
| $2 \times 10^{16}$ | $H^+$ | 16 | ~ 145 |

Figure 3.4 A shows a cross-sectional Transmission Electron Micrograph (TEM) of an example substrate after exfoliation. Here a hydrogen implantation of $3 \times 10^{16}$ $H_2^+$ $cm^{-2}$ dose at 140 KeV forms a hydrogen-rich layer approximately 600 nm deep in the silicon. The implantation damage extends ~ 150 nm from the exfoliated surface and this region must be polished or oxidized. The width of this damaged region is related to the width of the implanted ions distribution and is smaller for smaller energies, not setting a limit on how thin a silicon layer can be transferred. Energies as low as 16 KeV, corresponding to ~ 145 nm (Si + ONO) transferred were successfully used. In a higher magnification micrograph of the back ONO stack in the same substrate (Figure 3.4 B), the bonding interface is not discernable within the control oxide, attesting to the good quality of the oxide-oxide

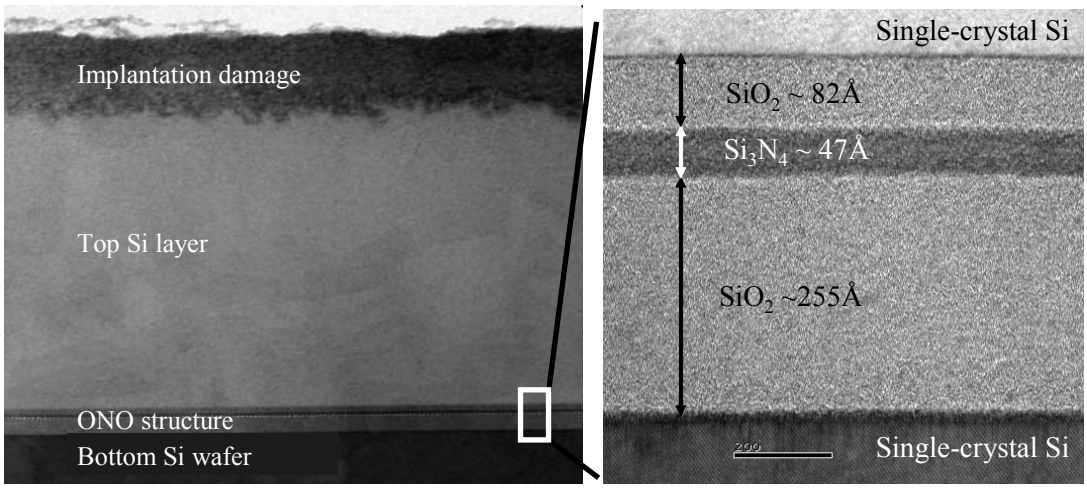



**Figure 3.4** TEM images of a prepared substrate (single crystal silicon above a ONO stack) after exfoliation. The left image is a low magnification image showing the whole transferred silicon layer with the implantation damage close to the surface. The right image is a high magnification image of the ONO structure.

bond and to the possibility of thinner back stacks using the same fabrication process. A TEM image of a substrate prepared with a thin back ONO stack, approximately 3/4/7 nm is shown in Figure 3.5. For simplicity of fabrication, in this work the substrate ($n^{++}$ silicon wafer) is used as a common back-gate to all the devices. With architecture schemes that require access to individual back-gates an additional lithography level is required to pattern the back-gate prior to the bonding and exfoliation steps. This capability of the Smart-Cut technique has been demonstrated by Aspar *et al.* to transfer patterned films [13] and by Avci *et al.* to implement back-gate MOSFETs [14].

Once the Si on ONO substrate is prepared, the rest of the fabrication of back – side trapping devices follows standard CMOS techniques with optical and electron-beam lithography.

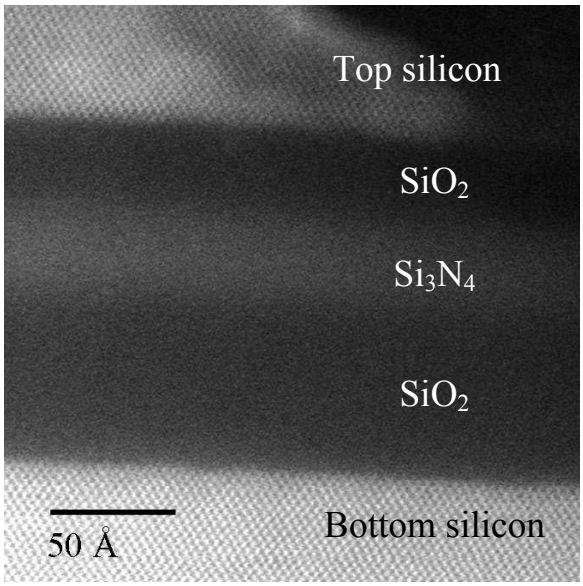



**Figure 3.5** TEM image of a cross section of a prepared substrate with a thin ONO stack. The back ONO stack is approximately 3, 4 and 7 nm respectively.

## 3.2 Transistors fabrication

The general aspects of the optical lithography and electron beam lithography processes used and the main steps of the transistors fabrication process are described in this section.

### 3.2.1 Optical lithography

The optical lithography was done with GCA AutoStep 200, an i-line (365 nm) lithography tool, following the standard processes used in CNF. The masks were made with the CNF GCA 3600F Pattern Generator. The resists used were OiR 620-7i for alignment marks, active area and gate levels and OiR 897-12i for the metal contacts level. In the large devices all levels were done using optical lithography. In the small

devices, the alignment marks, vias and metal contacts were patterned using optical lithography and the active area and gate were patterned using electron-beam lithography.

### 3.2.2 Electron-beam lithography

The electron beam lithography was done using Leica VB6-HR with a thermal field emission electron source operating at 100 kV. The earlier runs of small devices were done using negative tone resist NEB-31 in a 1:1 solution in MIBK (Methyl Isobutyl ketone) and the last run of small devices was done with FOX-12, an HSQ (hydrogen silsesquioxane) based negative resist for electron beam lithography. NEB 31 has etching properties similar to photo resists. The active area and gate definition is done using an oxide mask underneath the NEB-31 layer. The NEB-31 must be removed prior to the silicon or polysilicon etch in chlorine RIE. FOX-12 is a flowable oxide with etching properties similar to a PECVD deposited oxide and can be used as a mask to etch silicon or poly-silicon in chlorine RIE. Both resists were used with a thickness of approximately 90 nm. The larger features for both active and gate levels, the contact pads in small devices and the large area devices, were exposed with 10 nA and variable resolution limit (VRU) 4 (beam step size of 20 nm) and the small features, less than 100 nm, were exposed with 1 nA and VRU 1 (beam step size of 5 nm). The required doses vary between 20 and 250 $\mu C/cm^2$ for NEB-31 and between 350 and 2800 $\mu C/cm^2$ for FOX 12, for the largest and smallest feature size respectively. For the type of alignment required, gate line crossing the active level line, with a tolerance larger than 100 nm, die by die (~ 10 mm die) alignment is sufficient to achieve alignment better than ~ 30 nm. Standard routines for alignment in Leica VB6 were used. Both resists used for electron beam lithography, NEB-31 and

FOX-12, are very sensitive to variations in processing conditions and therefore it is important to keep the same parameters that are used for the dose tests, including the times between spinning and exposure and between end of exposure and development.

### 3.2.3 Alignment marks

Due to the thin silicon layers employed and different processes for active area isolation the alignment marks have to be defined separately on a different lithography level before the active area definition. The alignment marks were defined using optical lithography (for both optical and electron-beam lithography devices) with the AutoStep using photoresist 620-07i.

The alignment marks for electron-beam lithography must have high contrast under SEM. This can be done with evaporated metal features (if subsequent process steps are compatible) or deep etched trenches. In our process the alignment marks are etched ~ 1 μm deep into the silicon substrate.

Two sets of alignment marks are necessary for the two levels of electron-beam lithography, active and gate levels. During alignment, the electron beam crosses the four edges of the trench in order to determine its center. While locating the marks in the first level alignment the resist on the marks is exposed and the marks undergo the same process as the intended patterned areas making them unusable or harder to use in the second level alignment. This is illustrated in Figure 3.6, an SEM image of an exposed alignment mark.

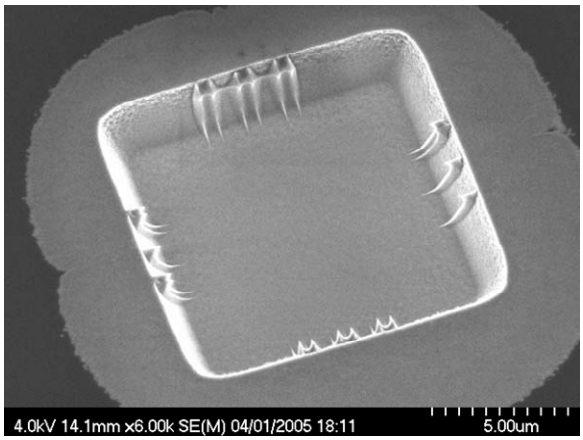



**Figure 3.6** SEM image of an exposed alignment mark, after being used for alignment with electron beam lithography.

### 3.2.4 Active area and device isolation

Different active area definition and device isolation techniques were used in the fabrication process to scale down and optimize the devices characteristics: (a) Mesa isolation (b) Shallow Trench Isolation (STI) and (c) LOCOS isolation [15].

#### Mesa isolation

The first runs of devices were fabricated using Mesa isolation. Figure 3.7 illustrates the Mesa isolation process flow for back-side trapping devices. Following the silicon thinning and the growth of a protective thin oxide layer an oxide layer is deposited and is used as a mask to etch the surrounding silicon to isolate the active area silicon islands. The oxide mask and the sacrificial oxide are removed and the gate oxide is thermally grown followed by the gate polysilicon deposition. While Mesa isolation is the simplest process for device isolation, it has two main drawbacks regarding the devices performance.

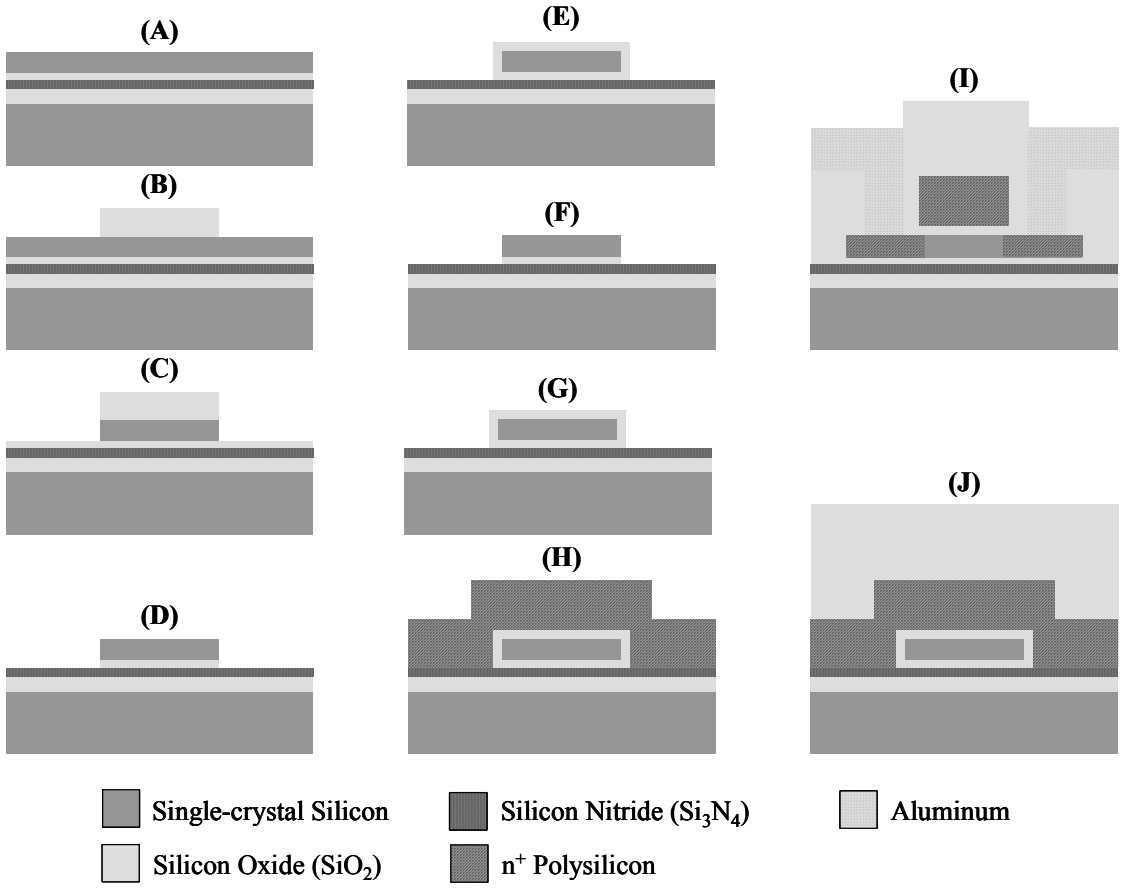



**Figure 3.7** Mesa isolation process flow. **(A)** Thin silicon on ONO substrate. **(B)** Grow and deposit silicon oxide and pattern the active area. **(C)** Dry etch silicon Mesa. **(D)** Wet etch oxide mask (bottom tunneling oxide is also removed). **(E)** Thermal oxidation for removal of side-wall damage from etch. **(F)** Wet etch of oxide. **(G)** Thermal oxidation (gate oxide). **(H)** Deposition, pattern and etch of polysilicon for gate. **(I)** Final cross-section view along Source-Drain (length direction) axis after field oxide deposition, vias opening and metal evaporation. **(J)** Final cross-section view along gate (width direction) axis.

The first one is specific to back-gated devices with thin buried insulators and is related to the isolation between the front terminals and the back-gate. With MESA isolation the front terminal pads are isolated from the substrate which is a common back-gate to all devices, only through the ONO stack, reduced by some necessary over-etch of the silicon (active area etch) and the polysilicon (gate etch). For thin ONO stacks, required for low-voltage operation, this isolation is not sufficient and the front terminals and the back-gate can short during the high-voltage write and erase operations, due to the reduced insulator thickness around the devices and higher probability of leakage paths across the pads large areas. A thin ONO stack will also result in high parasitic capacitance between the pads of the front terminals and the back-gate. The second drawback of mesa isolation is known as the mesa effect and is a kink observed in the transfer curves of the devices due to edge transistors with a different threshold voltage than the main channel due to different oxide thickness on the top and the edges of the mesa. There are also corners effects that alter the characteristics of the devices. These effects may not be significant if the channel area is much larger than the lateral areas but become important in small scale devices. Figure 3.8 shows an AFM image of a small scale device fabricated using the mesa isolation.

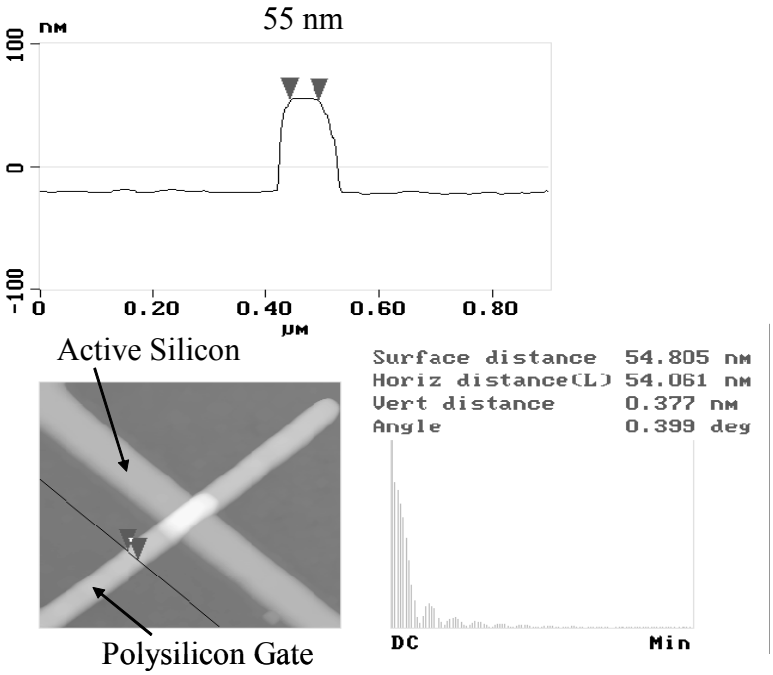



**Figure 3.8** AFM micrograph of a fabricated device using electron beam lithography to define both active and gate levels with mesa isolation. Gate length is ~ 55 nm.

**Shallow Trench Isolation (STI)**

STI is a standard isolation technique used for sud-100 nm silicon logic and memory devices. The STI isolation process for back-side trapping devices is illustrated in Figure 3.9. STI allows high density of devices, very good control of the width of the devices and makes for a more planar structure with lithography and device performance advantages. Figure 3.10 shows a cross-section SEM image of a thin single-crystal silicon layer above the back insulating films stack. The silicon substrate in this figure was etched for STI and the cross-section shows the active region of devices surrounded by insulator. STI isolation requires very precise control of the CMP step specially when using very thin silicon layers and buried insulator films.

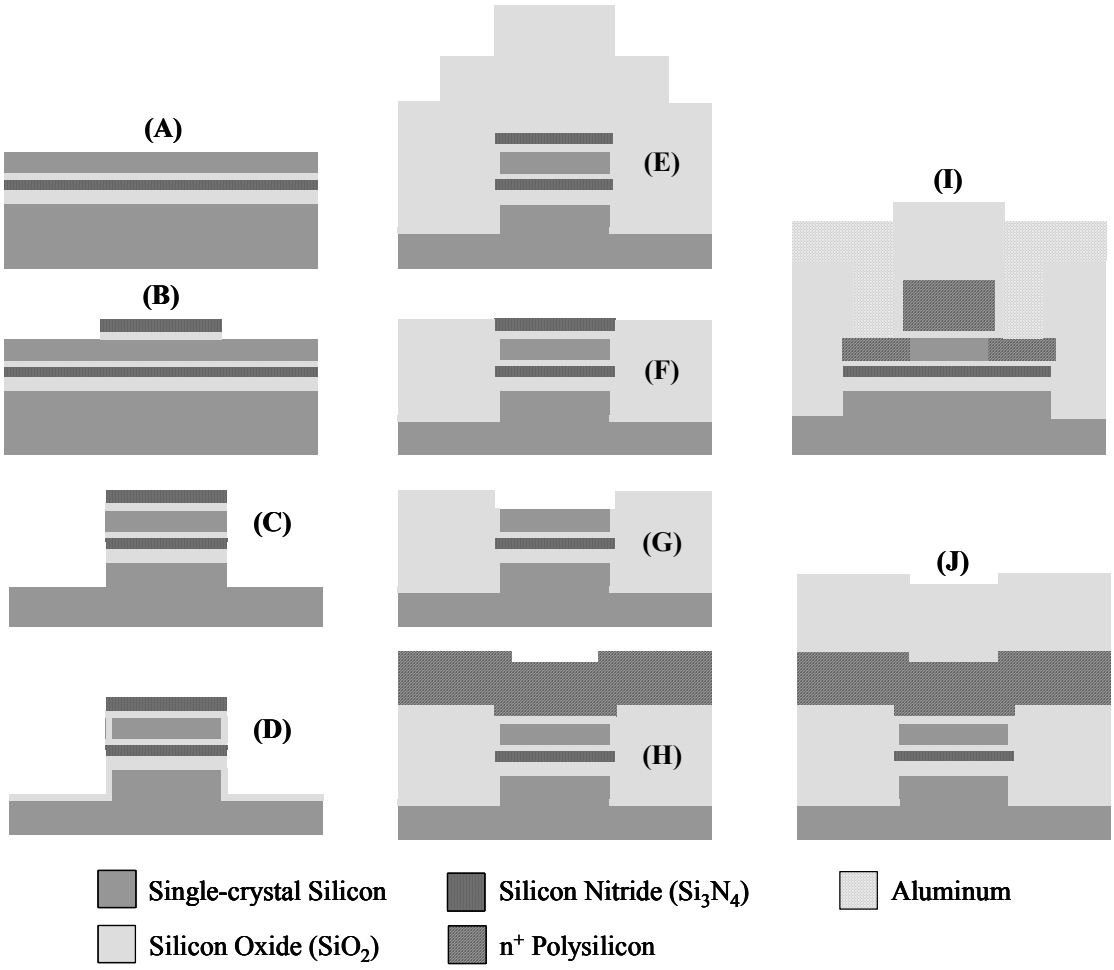



**Figure 3.9** STI isolation process flow. **(A)** Thin silicon on ONO substrate. **(B)** Grow thin cap silicon oxide and deposit silicon nitride. Pattern the active area, etch silicon nitride and silicon oxide. **(C)** Dry etch top silicon layer, ONO, and substrate trenches. **(D)** Thermal oxidation for side-wall damage passivation **(E)** Deposition of oxide to cover step of active area. **(F)** CMP (nitride acts as a stopping mask). **(G)** Wet etch of silicon nitride in hot phosphoric acid followed by wet etch of cap silicon oxide to expose silicon surface. **(H)** Thermal oxidation for gate oxide followed by deposition, pattern and etch of polysilicon for gate definition. **(I)** Final cross-section view along Source-Drain (length direction) axis after field oxide deposition, vias opening and metal evaporation. **(J)** Final cross-section view along gate (width direction) axis.

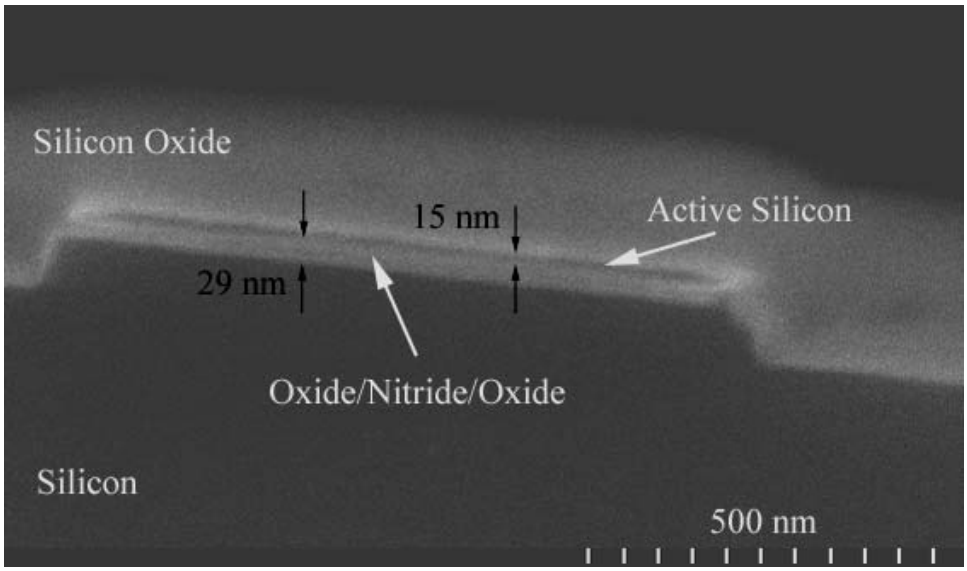



**Figure 3.10** SEM image of a device active area isolated using STI. The active area silicon is ~ 15 nm and the trenches are etched ~ 100 nm deep into the silicon substrate.

Polishing uniformity across the wafer is also a concern if the patterns are not very dense or uniformly distributed. CMP uniformity can be improved by using large area balancing structures but is not practical in electron-beam lithography due to extended exposure times. A mixed lithography step for the same level (e-beam and optical) can solve this problem with additional processing complexity.

**LOCOS isolation**

The last run of small devices were fabricated using LOCOS isolation due to the difficulties encountered with CMP control when using STI isolation.
Figure 3.11 illustrates the LOCOS isolation process flow for back-side trapping devices. The masked oxidation of the silicon around the active area causes a characteristic bump on the surface, *bird's beak*, which reduces the effective width of the device and results in different areas of the front and back channels.

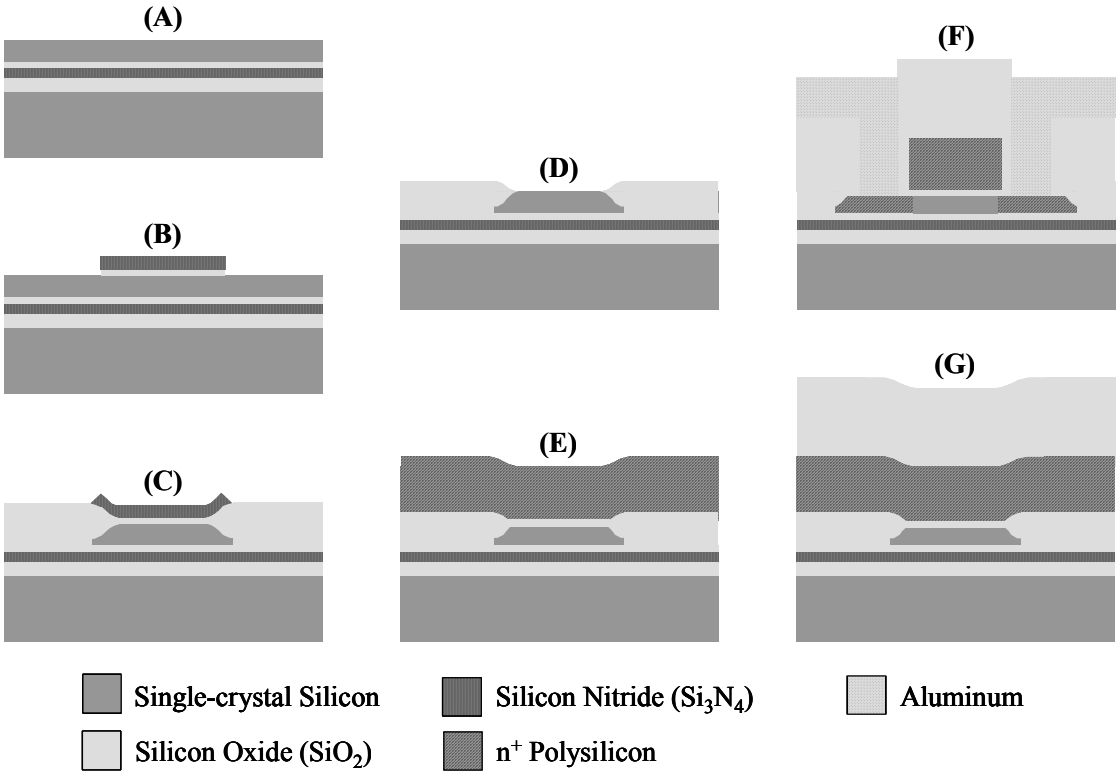



**Figure 3.11** LOCOS isolation process flow. **(A)** Thin silicon on substrate. **(B)** Grow thin cap silicon oxide (to reduce stress during subsequent oxidation) and deposit silicon nitride that will prevent the silicon oxidation in the patterned (active) area. Pattern the active area, etch silicon nitride and silicon oxide. **(C)** Thermal oxidation of silicon outside the active area resulting in a characteristic bump on the surface (*bird's beak*). **(D)** Wet etch of silicon nitride in hot phosphoric acid followed by wet etch of cap silicon oxide to expose silicon surface. **(E)** Thermal oxidation for gate oxide followed by deposition, pattern and etch of polysilicon for gate definition. **(F)** Final cross-section view along Source-Drain (length direction) axis after field oxide deposition, vias opening and metal evaporation. **(G)** Final cross-section view along gate (width direction) axis.

Another important aspect of LOCOS is the doping diffusion from the oxidized silicon areas into the edges of the active area. This increases the threshold voltage on the edges of the active area (or in the whole channel if this is narrow enough) effectively reducing the width or causing a MESA–like kink in the characteristics of the device. These effects have to be taken into account in the device design when using LOCOS isolation. For thin silicon layers on insulator, as employed in back-side trapping memories, the oxidation required to isolate the devices is much smaller than in conventional bulk devices and these LOCOS undesirable effects are relatively less important. Figure 3.12 shows an AFM image of a device active area isolated using LOCOS after the removal of nitride and cap oxide.

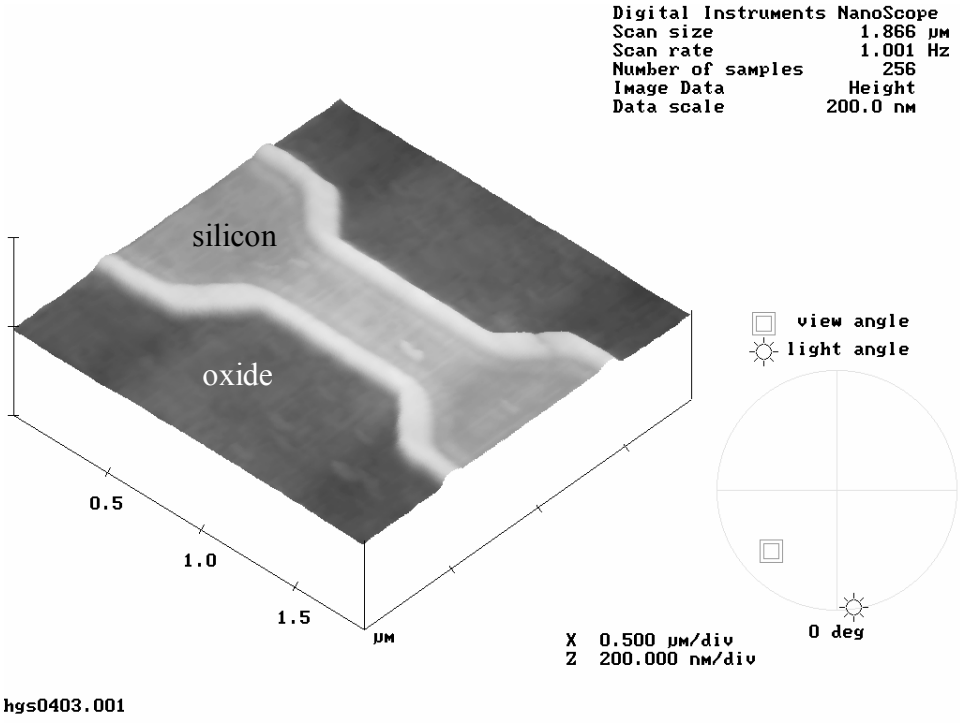



**Figure 3.12** AFM image of a device active area isolated by LOCOS, prior to gate stack deposition.

### 3.2.5 Gate stack deposition and patterning

After the active area definition the gate oxide is thermally grown followed immediately by the deposition of $n^+$ (for nFET) or $p^+$ (for pFET) polysilicon. The gate oxide (2-7 nm) is grown between 750 C and 900 C in different runs and in-situ doped polysilicon (60-80 nm) is deposited at 600 C. After polysilicon deposition an oxide layer is deposited as an etch mask for the polysilicon etch using $Cl_2$ and $BCl_3$ based RIE. As mentioned before, this is not required if FOX resist is used for the gate lithography. Although polysilicon etch in chlorine is very selective to oxide (~ 30:1) since the gate oxide is very thin overetch has to be minimized in order not to expose and etch the source and drain areas. This is particularly critical with the very thin silicon layers employed in these devices. Figure 3.13 shows an SEM image of a device after gate patterning prior to the polysilicon etch.

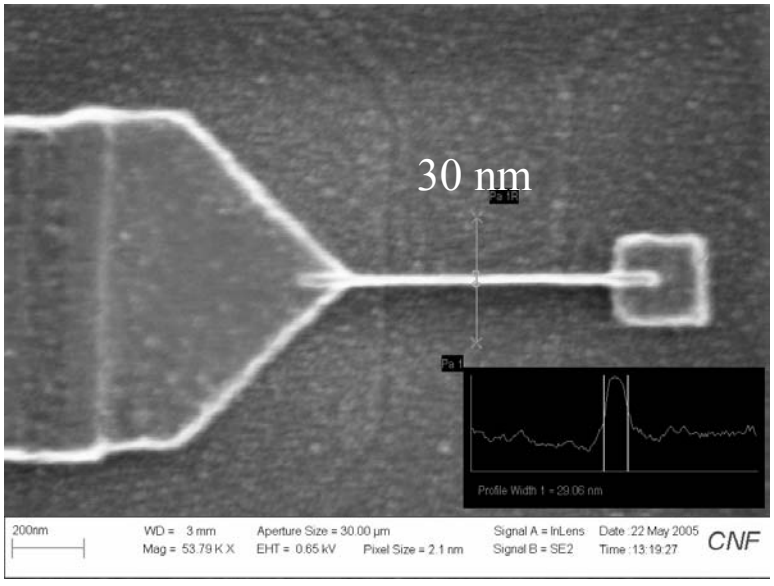



**Figure 3.13** SEM image of a small gate length device after gate patterning using FOX 12 prior to the polysilicon etch. The polysilicon grains are visible in this image.

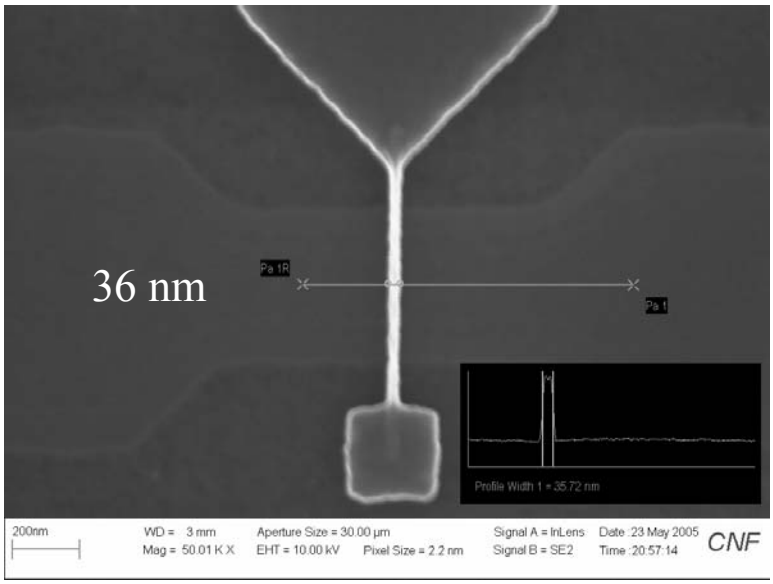



**Figure 3.14** SEM image of a small gate length device after gate polysilicon etch.

Figure 3.14 shows an SEM image of a device after the gate polysilicon etch. In both figures the devices were isolated using LOCOS. Figure 3.15 is an SEM image of a cross-section of a 0.5 μm gate length device with silicon channel ~ 40 nm thick. The cross-section of a very small device, gate length ~ 20 nm, where the ONO stack is visible underneath a thin silicon layer, ~ 15 nm, is shown in Figure 3.16.

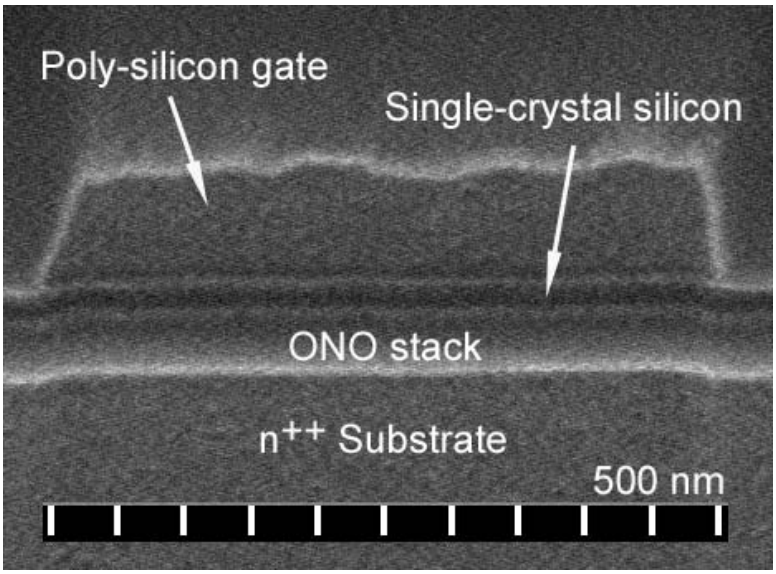



**Figure 3.15** SEM cross-section of a 0.5 μm back-side trapping device.

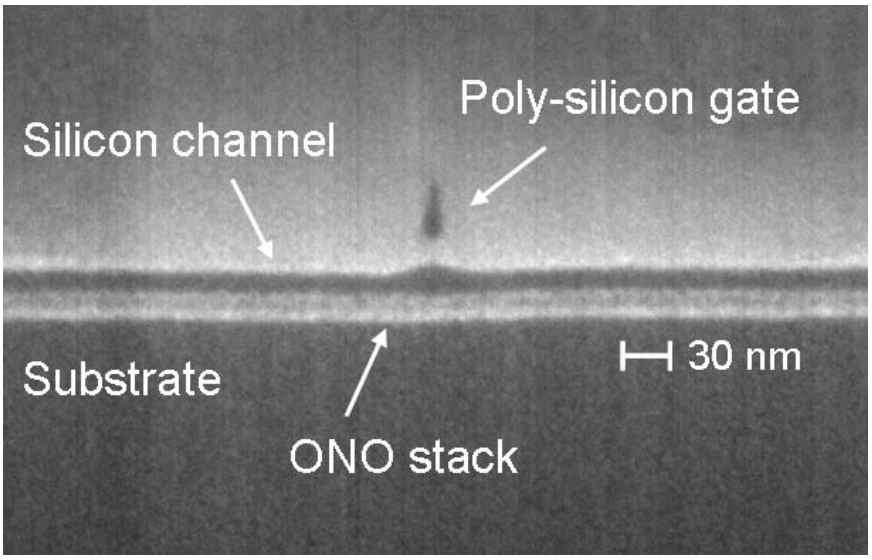



**Figure 3. 16.** SEM cross-section of a small back-side trapping device. The gate length is ~ 20 nm.

### 3.2.6 Thin gate oxide for device scaling

In the earlier runs, while developing the fabrication process, the front-gate oxide was kept at ~ 6 - 7 nm. This ensures minimal gate leakage current but results in reduced electrostatic control in the short gate length devices. In the last run of devices, for nFET and pFET memory devices, the gate oxide was reduced to ~ 2.4 nm to improve the performance of the small scale devices. This oxide was thermally grown at 750 C, for 10 min, in 20% partial oxygen ambience, followed by an anneal of 10 min at 750 C. The wafers were loaded and unloaded in the furnace at 650 C. Figure 3.17 shows the thickness results obtained with a spectroscopic ellipsometer for a p-type silicon monitor wafer.

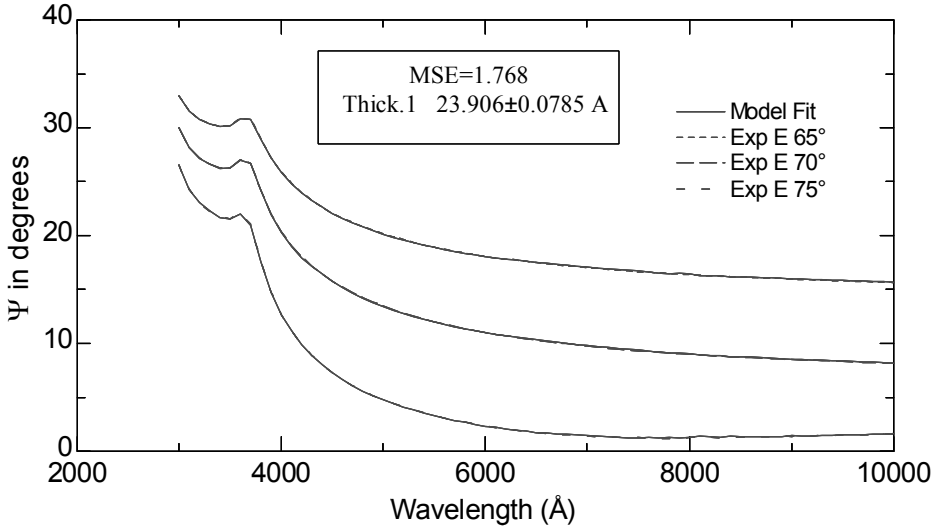



**Figure 3.17** Thin gate oxide thickness measurement by ellipsometry (CNF Woollam Spectroscopic Ellipsometer). The change in the polarization state of light is plotted as a function of wavelength for three angles of incidence. The spacing in wavelength was 10 nm. The oxide thickness is determined to be 23.9 ± 0.1 Å.

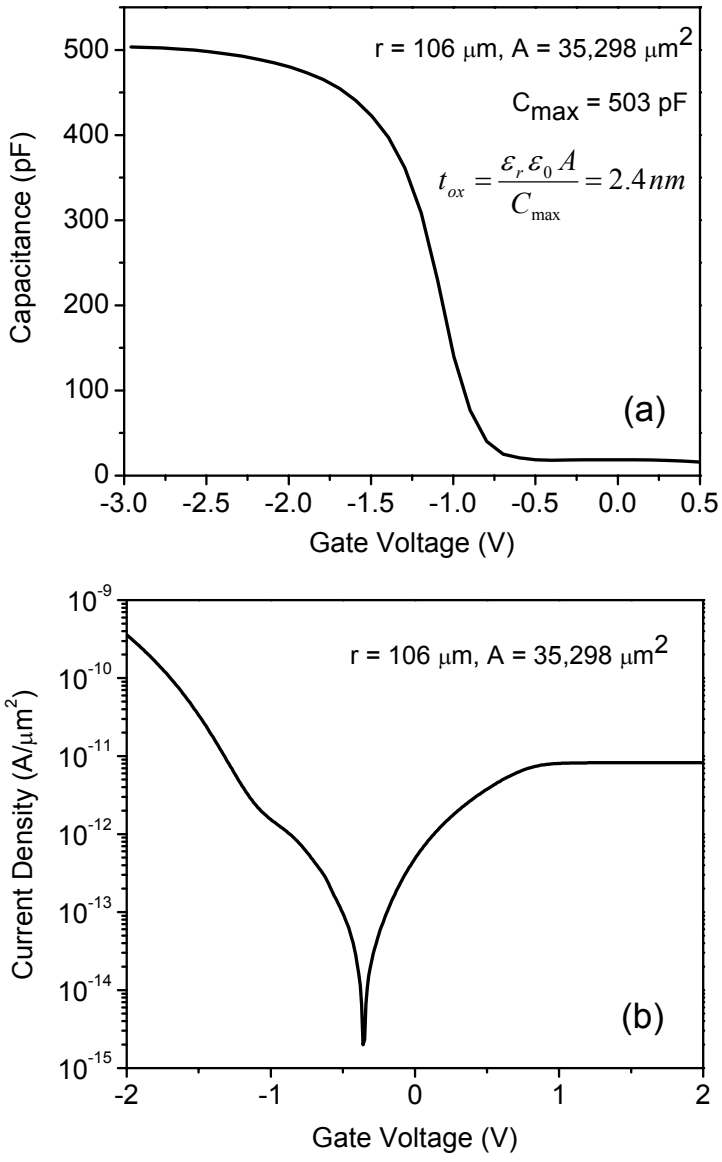



**Figure 3.18** Electrical properties of thin gate oxide. **(a)** Capacitance versus gate voltage for a circular capacitor Al/$SiO_2$/p-Si of radius 106 μm. The oxide thickness as determined electrically from the maximum capacitance is 2.4 nm. **(b)** Current density versus gate voltage for the same capacitor.

MOS capacitors were formed by evaporating aluminum through a shadow mask to characterize the electrical properties of the thin gate oxide. Measured current density and capacitance versus gate voltage characteristics for a circular capacitor of diameter 106 μm are plotted in Figure 3.18. Assuming a linear area scaling, which is probably an overestimation due to higher probability of leakage paths in large area capacitors, at $V_G = 1$ V, the gate current in a 1 $\mu m^2$ area device would be in the order of 5 pA. The oxide thickness derived electrically from the accumulation capacitance is in agreement with the thickness estimated using ellipsometry (2.4 nm in both cases).

### 3.2.7 Body and source/drain ion implantation

After gate patterning and etch, a thin oxide, ~ 5 – 7 nm, is grown to passivate the sidewalls of the gate that are damaged due to the plasma etch process. After this oxidation the source and drain regions are implanted with arsenic and boron, for nFET and pFET, respectively, with doses ranging from 1 x $10^{14}$ $cm^{-2}$ to 2 x $10^{14}$ $cm^{-2}$ and energies ranging from 2.5 KeV to 20 KeV. This results in source and drain doping concentration of approximately 5 x $10^{19}$ $cm^{-3}$. In order to restore crystallinity and electrically activate the dopants the devices are then annealed at 900 C for 15 min (short gate length devices) and 90 min (large devices). The doses and energies must be adjusted for different gate length devices and silicon thicknesses. A single implantation energy is sufficient for these devices since thin silicon layers, less than 50 nm, are used and the source and drain doping should be uniform across the silicon depth for front and back channel operation. The devices were fabricated with undoped channel (the starting substrate is estimated to have a doping concentration of ~ 8 x $10^{16}$ $cm^{-3}$).

### 3.2.8 Metallization

After the source and drain implantation anneal a thick (~ 200 nm) PECVD oxide is deposited. The via holes are patterned using optical lithography and etched using RIE and wet etch (buffered oxide etch). The metal contacts are done using a lift-off process with optical lithography and aluminum evaporation (~ 300 nm). At the end the devices are annealed in forming gas between 250 C and 300 C. Figure 3.19 shows an AFM image of a large device at the end of the fabrication process, after thick passivation oxide deposition, vias opening and aluminum evaporation for contacts.

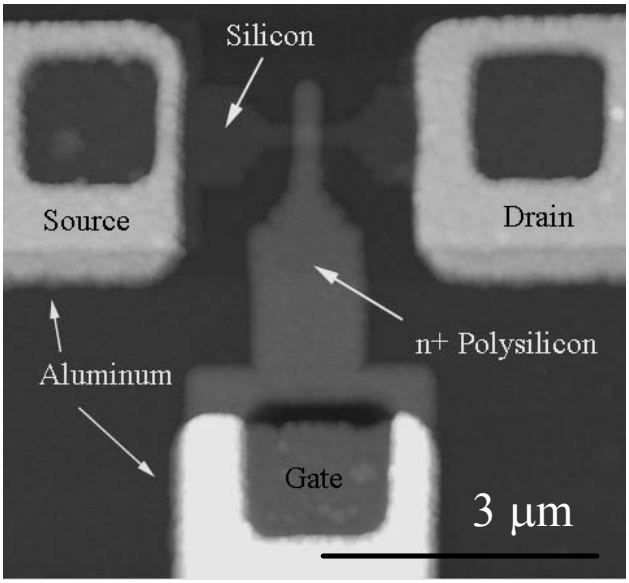



**Figure 3.19** AFM micrograph. Top view of a large device fabricated using optical lithography after vias opening and metal evaporation.

## 3.3 Summary

The fabrication process for back-side trapping memories was described in this chapter. The substrate, Si on ONO, was prepared using a Smart-Cut based technique. Large devices, down to 0.5 μm gate length, were fabricated using optical lithography and small scale devices, down to 20 nm gate length (Figure 3.20) were fabricated using electron-beam lithography.

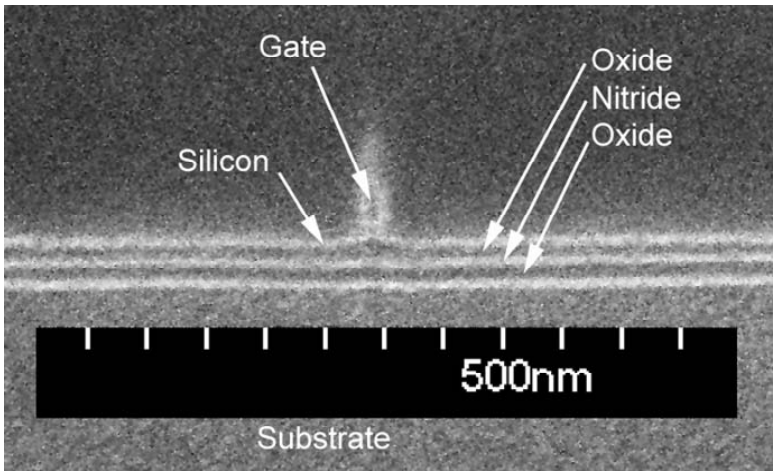



**Figure 3.20** SEM cross-section of a small device. The gate length is ~ 20 nm.

# Chapter 4

# Devices characterization

The electrical characteristics of the fabricated devices are discussed in this chapter. The first part describes the transistor behavior, front gate and double gate operation. The second part describes the memory characteristics of the devices making use of the ONO trapping medium between the silicon channel and the back gate.

## 4.1 Transistor characteristics

### 4.1.1 Front gate transistors operation

Figure 4.1 shows the transfer and output characteristics for the front channel transistor operation of a back-side trapping device with W = 3 μm and L = 0.75 μm in which the front gate oxide is 7 nm, the back ONO is ~ 7/20/80 nm and the silicon body is ~ 40 nm. In Fig. 4.1 A the transfer curve is shown for different back gate bias, between -2 V and + 2 V, indicating good coupling between the back gate and the front channel which can be used for threshold voltage tuning for power adaptive applications. The front channel transistors show the expected good sub-threshold characteristics. The sub-threshold slope for a fully depleted SOI MOSFET, neglecting the presence of interface states, is given by [11]:

$$S = \left( \frac{d(\log_{10} I_D)}{dV_G} \right)^{-1} = 2.3 \frac{kT}{q} \left( 1 + \frac{C_b}{C_{ox1}} \right) \qquad (1)$$

where $q$ is the electronic charge, $kT$ is the thermal energy, $C_b$ is the capacitance

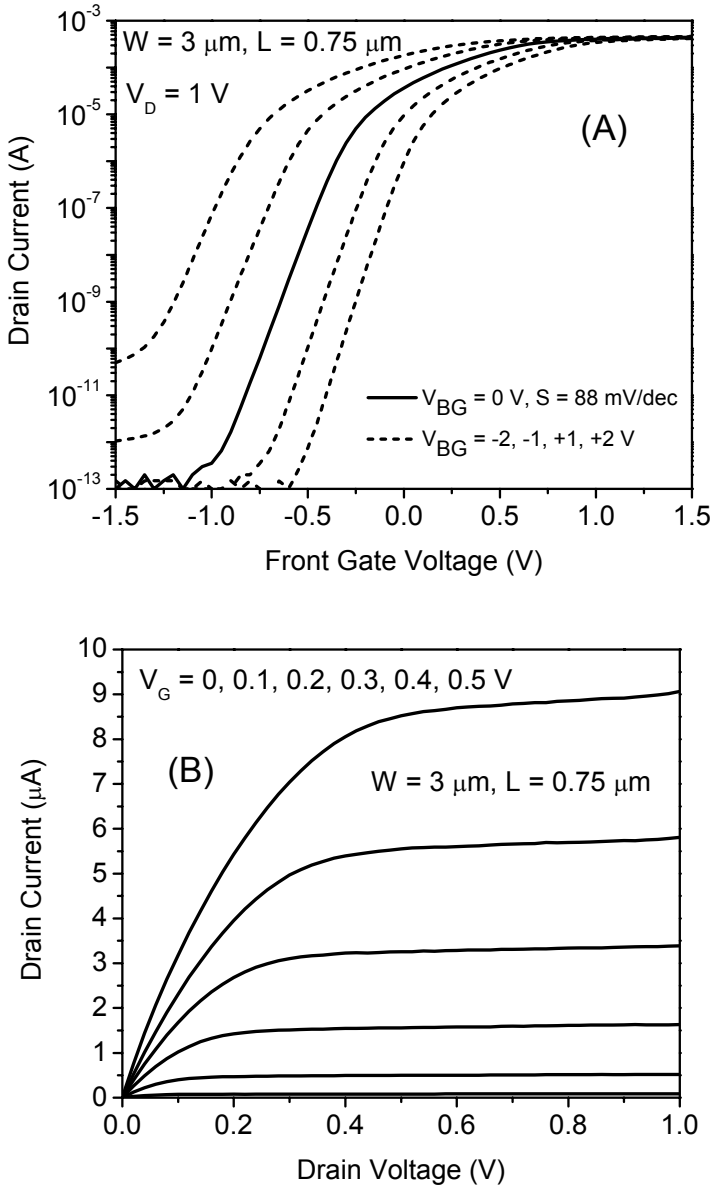



**Figure 4.1** Front channel transistor operation for a back-side trapping device. W = 3 μm, L = 0.75 μm. The silicon body is ~ 40 nm, the back ONO stack is ~ 7/20/80 nm and the front gate oxide is 7 nm. **(A)** Transfer characteristics for different back-gate bias. The solid line corresponds to $V_{BG}$ = 0 V. **(B)** Output characteristics.

between the inversion channel and the back gate and $C_{ox\,1}$ is the front gate oxide capacitance ($\varepsilon_{ox}/t_{ox\,1}$). $C_b$ is given by the silicon film capacitance for a fully depleted device with the back interface in accumulation (1) and by the series capacitance of the silicon film capacitance and the back gate oxide capacitance for a fully depleted device with the back interface depleted (2):

$$C_b = \frac{\varepsilon_{Si}}{t_{Si}} \qquad (2)$$

$$C_b = \frac{C_{Si}\,C_{ox\,2}}{C_{Si} + C_{ox\,2}} = \frac{\dfrac{\varepsilon_{Si}}{t_{Si}}\dfrac{\varepsilon_{ox}}{t_{ox\,2}}}{\dfrac{\varepsilon_{Si}}{t_{Si}} + \dfrac{\varepsilon_{ox}}{t_{ox\,2}}} \qquad (3)$$

where $t_{Si}$ is the silicon film thickness and $t_{ox\,2}$ is the back gate oxide thickness (in the case of back side trapping devices is the equivalent back ONO thickness). The sub-threshold slope in a fully depleted SOI transistor is higher (i.e. worse) when the back interface is accumulated (2) and lower (*i.e.* better) when the back interface is depleted (3). When the back interface is inverted the sub-threshold slope definition (1) is not meaningful.

Substituting the front gate oxide, the silicon thickness and the equivalent back ONO thickness by 7 nm, 40 nm and 100 nm respectively, the expected sub-threshold slope is 91 mV/decade if the back interface is in accumulation and 63 mV/decade if the back interface is depleted. In Fig. 4.1 A the sub-threshold slope is 90 mV/decade at $V_{BG}$ = 0 V, 84 mV/decade at $V_{BG}$ = -1 V and 78 mV/decade at $V_{BG}$ = -2 V. Due to undoped channel these devices have negative threshold voltage and at $V_{BG}$ = 0 V the back interface is in weak inversion and the sub-threshold slope definition is not valid. The fact that the sub-threshold slope is lower at $V_{BG}$ = - 2 V than at $V_{BG}$ = - 1 V

indicates that the back interface is still in weak inversion at $V_{BG}$ = -1 V and is either between inversion and depletion or in accumulation at $V_{BG}$ = - 2 V. The lowest sub-threshold slope of the device (with depleted back interface) is therefore less or equal to 78 mV/decade, in good agreement with the ideal value, 63 mV/decade. This attests to the good quality of the silicon surface. The output characteristics (Fig. 4.1 B) also show good electrical properties of the device, in particular good ohmic contacts.

Figures 4.2 to 4.4 show transfer curves (A) and output characteristics (B) for small devices, with gate length of 200 nm, 50 nm and 20 nm. In these devices the films thickness were reduced to allow lower memory operation voltages. The front gate oxide is ~ 6 nm, the back ONO is ~ 3/4/7 nm and the silicon body is ~ 15 nm. These curves show good electrostatic control, with on/off ratio larger than $10^6$ and off current below pA/μm, even at the shortest gate length. With the thicknesses employed in these small devices, the ideal sub-threshold slope (neglecting interface traps and short channel effects corrections) from (2) and (3) is 130 mV/decade and 81 mV/decade, with the back interface in accumulation and depletion, respectively. The measured sub-threshold slope values are 116 mV/decade for the 150 nm gate length device (Fig. 4.2), 82 mV/decade for the 50 nm gate length device (Fig. 4.3) and 207 mV/decade for the 20 nm gate length device (Fig. 4.4), all at $V_{BG}$ = 0 V. Since the front gate oxide and the back ONO thicknesses are very uniform across the wafer (grown and deposited very thin films), variations in sub-threshold slope from device to device are most likely due to silicon film thickness variations introduced by the CMP step.

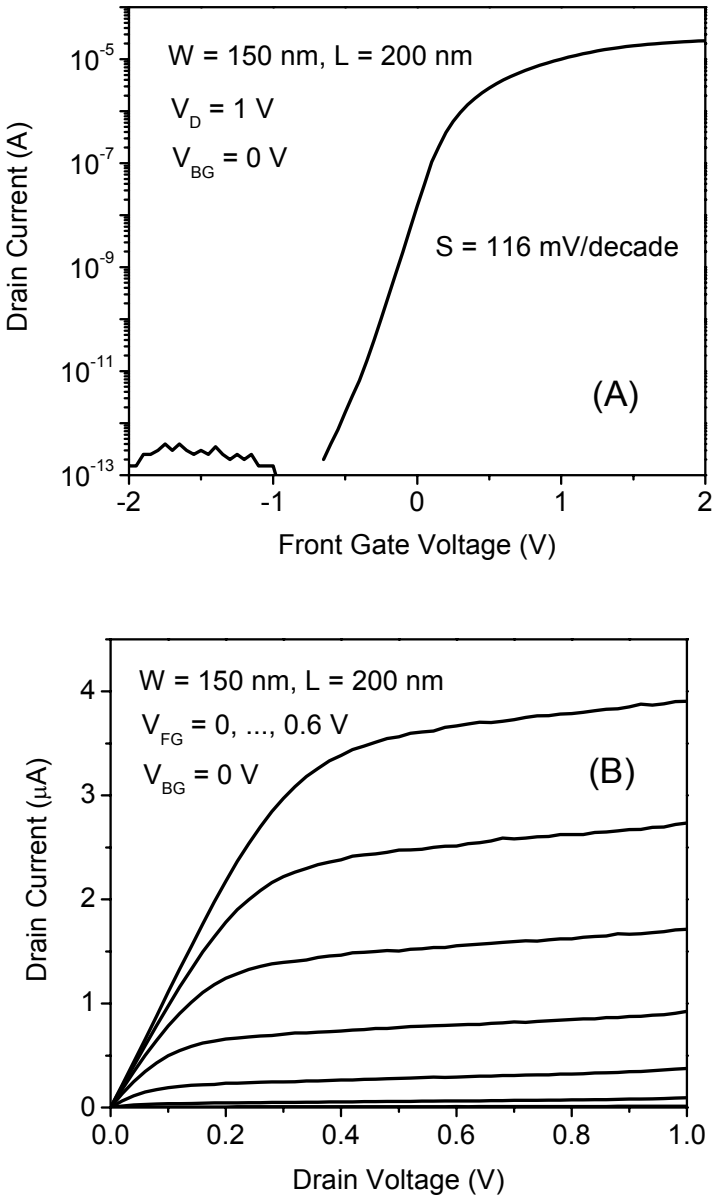



**Figure 4.2** Front channel transistor operation for a back-side trapping device. W = 150 nm, L = 200 nm. The silicon body is ~ 15 nm, the back ONO stack is ~ 3/4/7 nm and the front gate oxide is 6 nm. **(A)** Transfer characteristics. **(B)** Output characteristics.

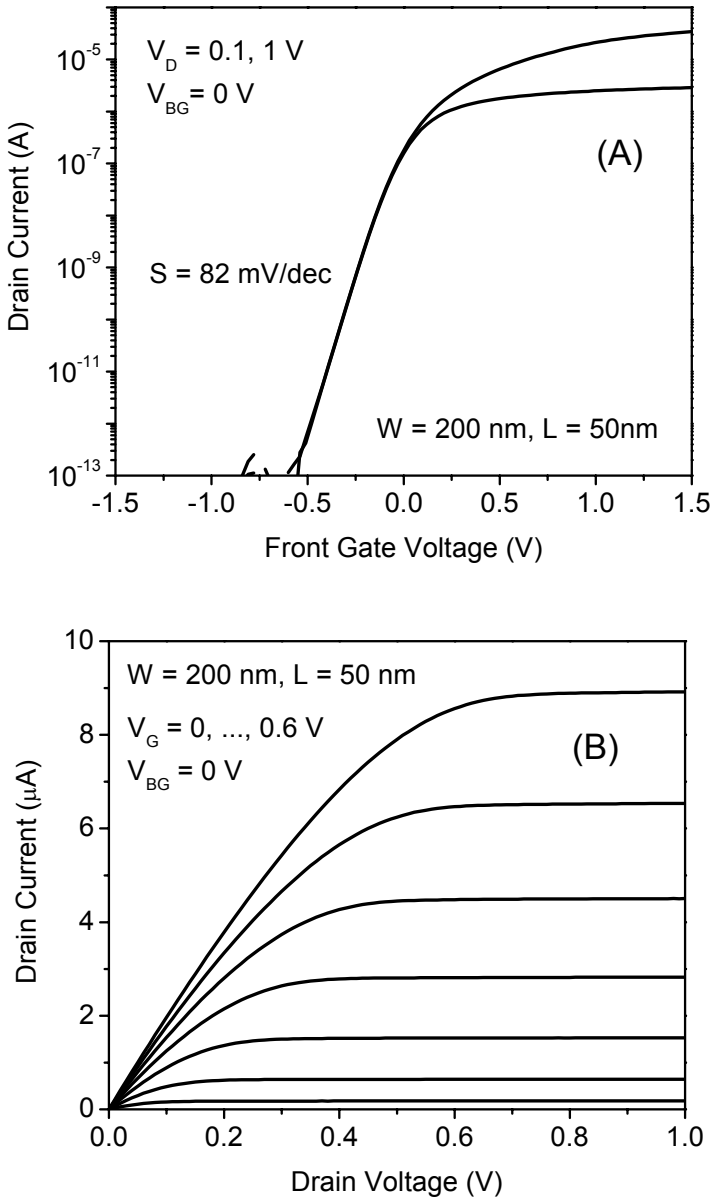



**Figure 4.3** Front channel transistor operation for a back-side trapping device. W = 200 nm, L = 50 nm. The back ONO stack is ~ 3/4/7 nm and the front gate oxide is 6 nm. **(A)** Transfer characteristics. **(B)** Output characteristics.

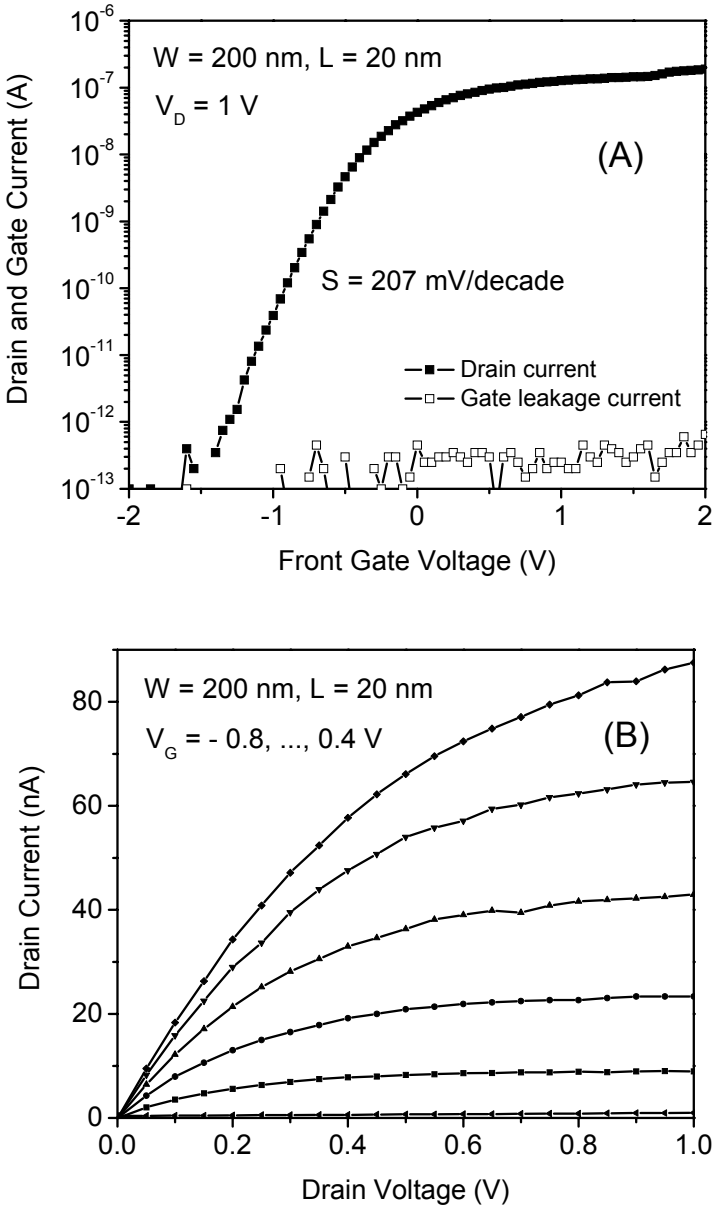



**Figure 4.4** Front channel transistor operation for a back-side trapping device. W = 200 nm, L = 20 nm. The silicon body is ~ 15 nm, the back ONO stack is ~ 3/4/7 nm and the front gate oxide is 6 nm. **(A)** Transfer characteristics including the gate leakage current. **(B)** Output characteristics.

### 4.1.2 Double gate operation

In addition to the above transistor behavior, there are interesting properties that can be observed because the thin silicon, doped back contact, and the thin insulators provide the structure with many of the characteristics expected from back-gate and double-gate geometries: inversion channels for transistor can be formed both on the front and the back silicon interfaces, and the threshold voltage in either of these modes can be modulated by the corresponding back-bias. Figures 4.5 and 4.6 show the I-V and C-V measurements for double-gate operation of these devices. These measurements were done on a device with active area of 20 μm x 10 μm. This geometry also allows us to accurately extract the silicon body thickness from capacitance measurements. Let us consider the front gate to source/drain capacitance characteristics, $C_{FG\text{-}SD}(V_{FG}, V_{BG})$ (Fig. 4.6 A). When both interfaces are off the measured capacitance is close to zero. When the front interface is on and the back interface is off, or when both interfaces are on, the capacitance is the front gate oxide capacitance. Finally, when the front interface is off but the back interface is on, the measured capacitance is the series capacitance of the front-gate oxide capacitance and the silicon film capacitance. From this the silicon film thickness can be calculated as follows:

$$\frac{1}{C} = \frac{1}{C_{Si}} + \frac{1}{C_{ox}} \quad \Rightarrow \quad t_{Si} = \varepsilon_0 \, \varepsilon_{Si} \, A \, ( \frac{1}{C} - \frac{1}{C_{ox}} ) \qquad (4)$$

where $C$ is the measured capacitance when the front interface is off and the back interface is on, $C_{ox}$ is the front gate oxide capacitance, measured when the front interface is on, $A$ is the active area of the device, and $C_{Si}$ is the silicon capacitance. In this case the silicon film thickness is determined to be 42 nm.

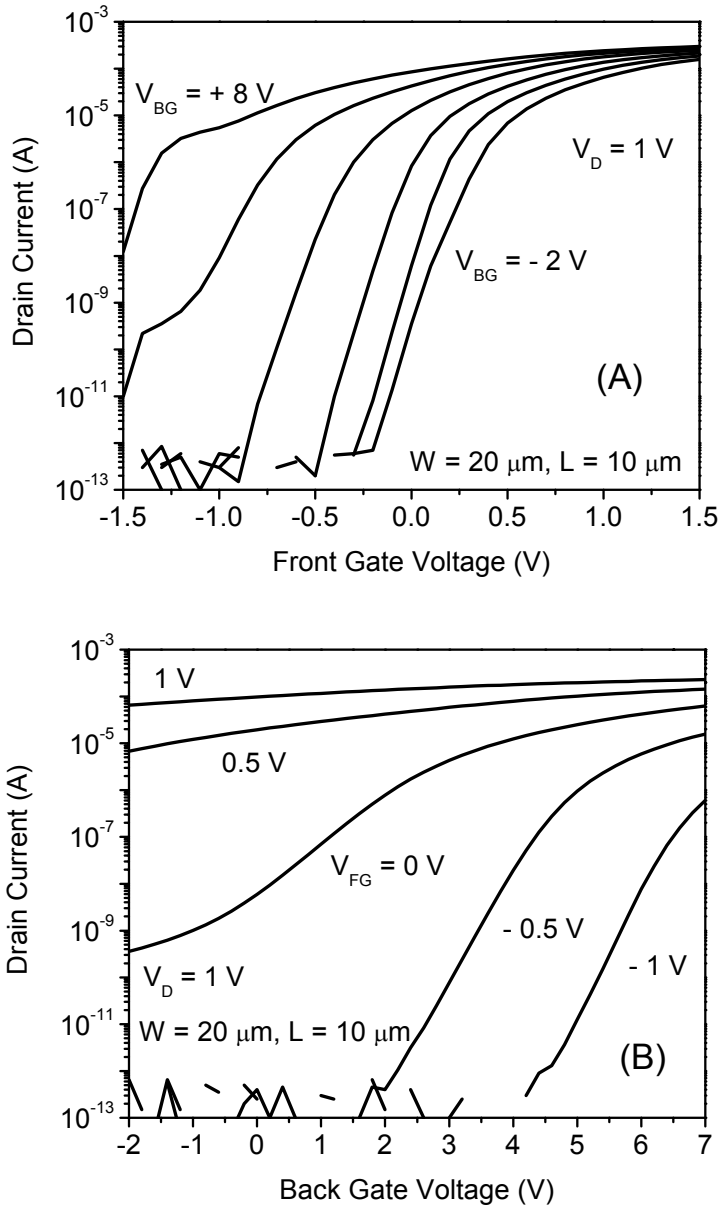



**Figure 4.5** Double-gate transistor operation for a back-side trapping device. W = 20 μm, L = 10 μm. The silicon body is ~ 40 nm, the back ONO stack is ~ 7/20/80 nm and the front gate oxide is 7 nm. **(A)** Front channel transfer characteristics for different back gate bias, from $V_{BG}$ = - 2 V to + 8 V in 2 V steps. **(B)** Back channel transfer characteristics for different front gate bias, from $V_{FG}$ = - 1 V to + 1 V in 0.5 V steps.

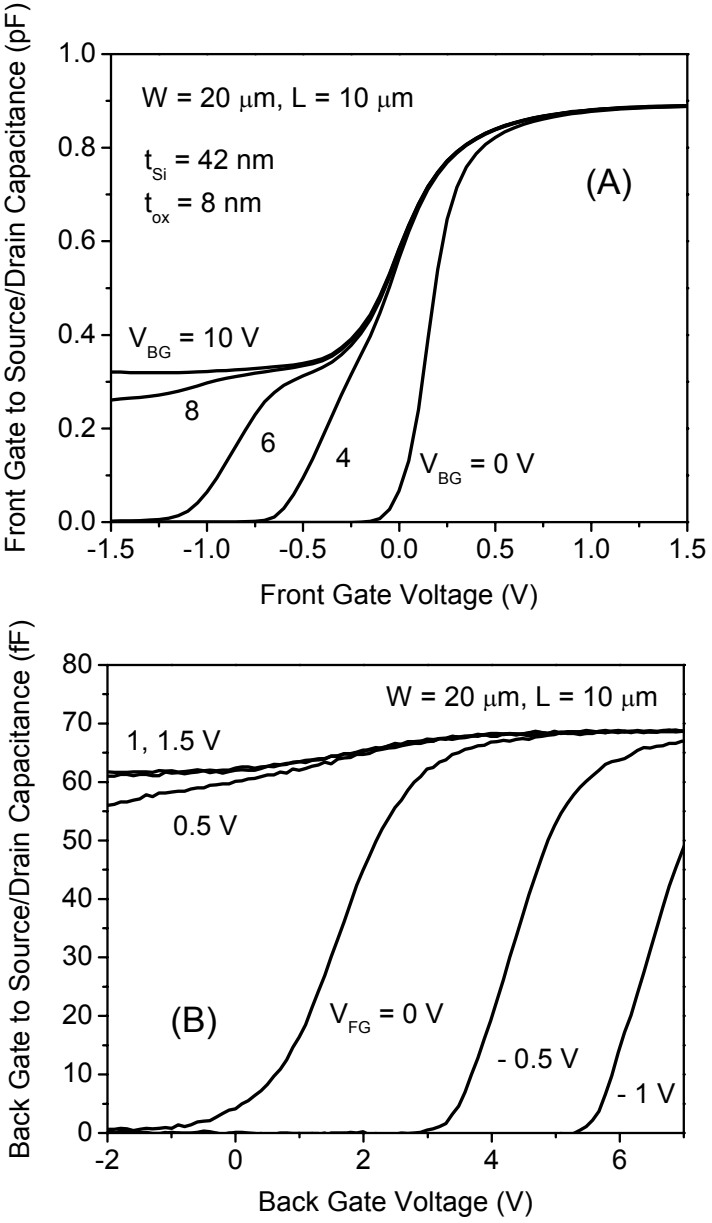



**Figure 4.6** Double-gate transistor C-V characteristics for a back-side trapping device. W = 20 μm, L = 10 μm. The silicon body is ~ 40 nm, the back ONO stack is ~ 7/20/80 nm and the front gate oxide is 7 nm. **(A)** Front gate to source/drain capacitance versus front gate voltage for different back gate bias, $V_{BG}$ = 0, + 4, + 6, + 8, + 10 V. **(B)** Back gate to source/drain capacitance versus back gate voltage for different front gate bias, $V_{FG}$ = - 1, - 0.5, 0, + 0.5, + 1 V.

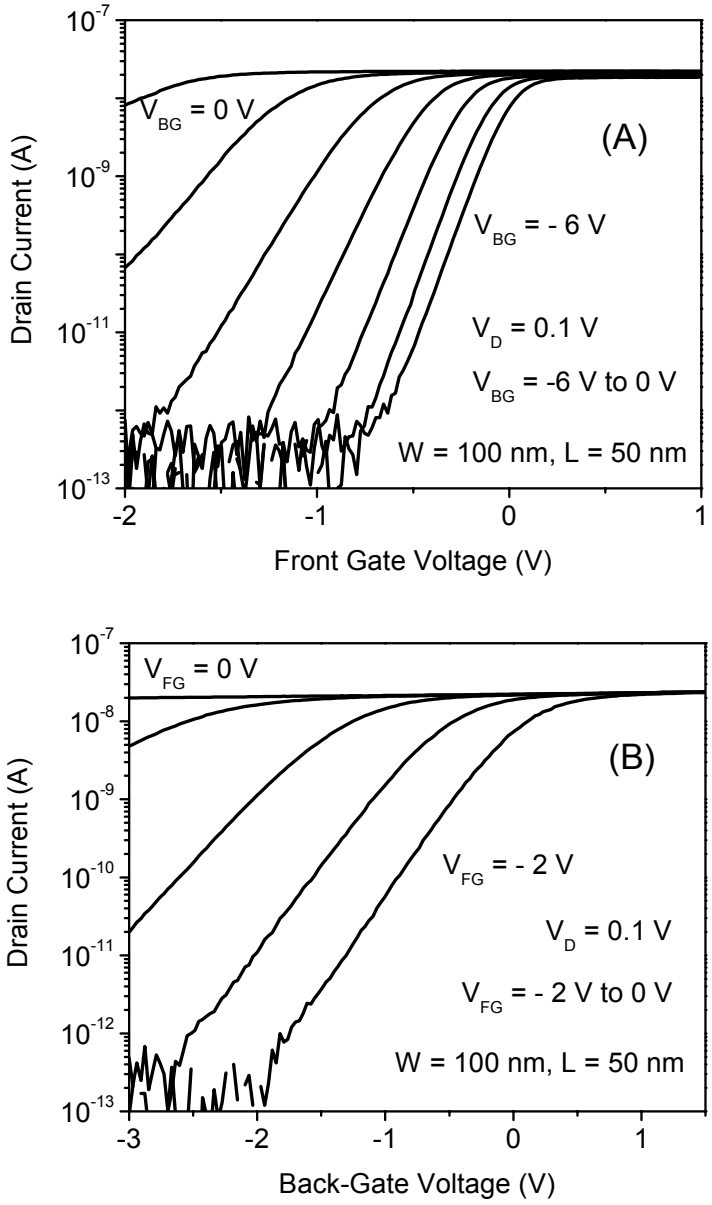



**Figure 4.7** Double-gate transistor operation for a back-side trapping device. W = 100 nm, L = 50 nm. The silicon body is ~ 15 nm, the back ONO stack is ~ 3/4/7 nm and the front gate oxide is 6 nm. **(A)** Front channel transfer characteristics for different back gate bias, from $V_{BG}$ = - 6 V to 0 V in 1 V steps. **(B)** Back channel transfer characteristics for different front gate bias, from $V_{FG}$ = - 2 V to 0 V in 0.5 V steps.

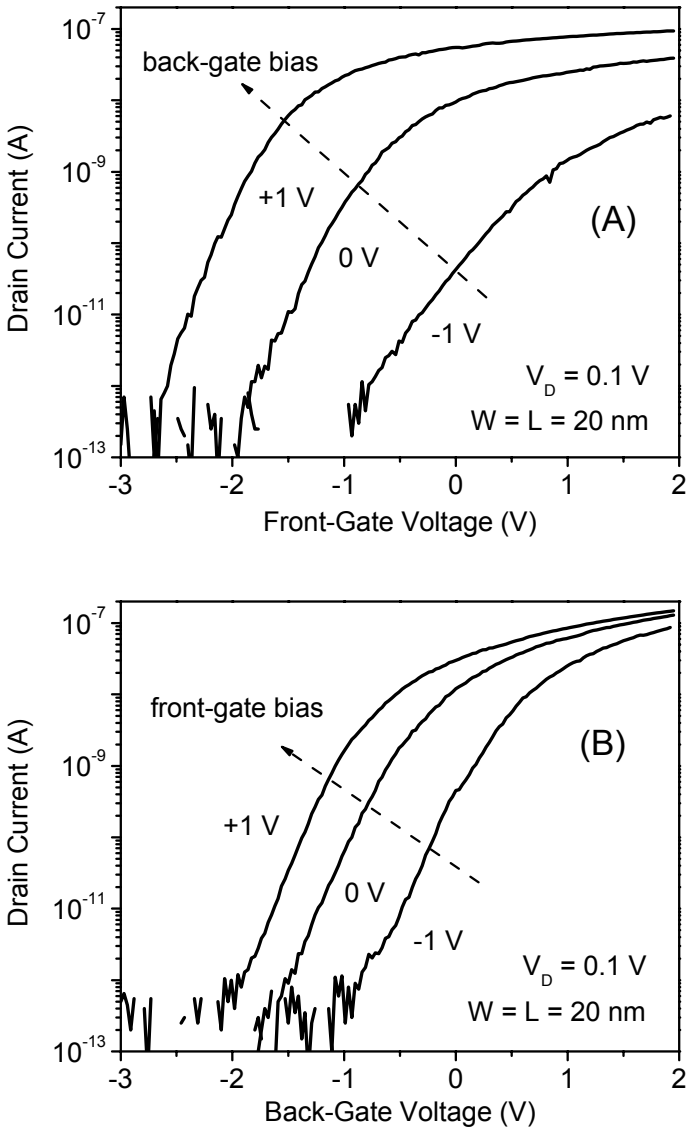



**Figure 4.8** Double-gate transistor operation for a back-side trapping device. W = 20 nm, L = 20 nm. The silicon body is ~ 15 nm, the back ONO stack is ~ 3/4/7 nm and the front gate oxide is 6 nm. **(A)** Front channel transfer characteristics for different back gate bias, $V_{BG}$ = - 1, 0, + 1 V. **(B)** Back channel transfer characteristics for different front gate bias, $V_{FG}$ = - 1, 0, +1 V.

The double gate characteristics for both front and back silicon interfaces for devices with gate length of 50 nm and 20 nm are plotted in Figures 4.7 and 4.8 respectively. The efficient threshold voltage modulation by the corresponding back gate is due to the thin dielectrics employed (front oxide and back ONO) and fully depleted silicon body.

## 4.2 Memory Characteristics

### 4.2.1 Charge injection and removal from the back ONO – memory window

Figure 4.9 shows the memory operation of a back-side trapping device. These are the transfer characteristics in the erased and written states (without charge and with charge stored in the ONO, respectively). The transfer characteristics show $I_{On}/I_{Off}$ ratio larger than $10^7$ and sub-threshold slope of 97 mV/decade in the erased state and 80 mV/decade in the written state. The negative charge stored in the back ONO causes an improvement in the sub-threshold slope of the front channel transistor. The reason is that in the erased state, with $V_{BG} = 0$ V, as mentioned before, the back interface is weakly inverted (the threshold voltage is negative due to undoped channel). This causes what looks like a worse sub-threshold slope for the front interface transistor because of higher off current which is due to back interface conduction. Once charge is injected into the back ONO it causes the back interface to go into depletion or accumulation. As a result the off current due to the back interfaces decreases and the sub-threshold slope of the front transistor improves. This effect was also observed by Kumar *et al.* in simulated results for undoped channel back floating gate memory devices [9]. The front gate oxide is 7 nm thick, the back ONO stack is 7/20/80 nm and the active silicon layer is approximately 40 nm. At 1 nA current drive a memory window ($\Delta V_T$) of 0.68 V is obtained. The thick back ONO stack used in the first set of

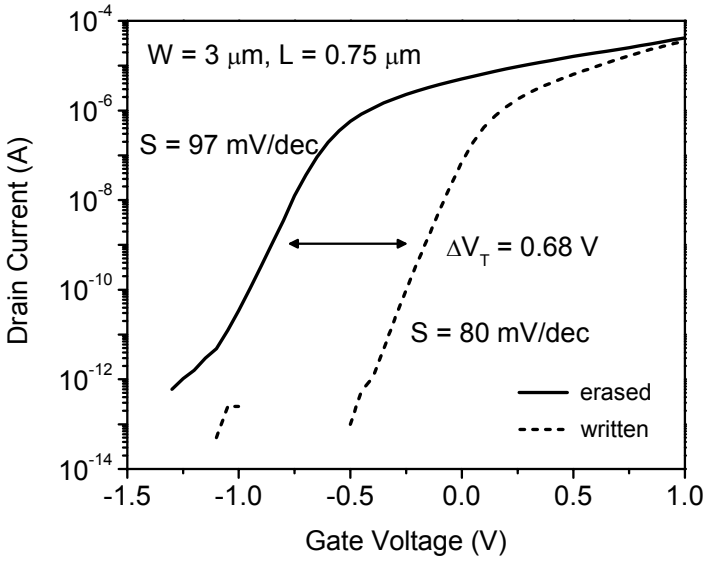



**Figure 4.9** Memory operation of a back-side trapping memory in erased (solid line) and written (dashed line) states with $V_D$ = 1 V. The memory window at 1 nA is 0.68 V. The write and erase voltages were + 50 V and - 35 V 300 ms pulses applied to the back-gate while front-gate, source and drain were grounded. Front gate oxide is 7 nm and the back ONO is ~ 7/20/80 nm. The silicon thickness is ~ 40 nm.

devices, ~ 7/20/80 nm, imposes the use of large programming voltages in order to achieve appreciable tunneling between the active silicon and the traps in the back ONO. Voltages in the order of 50 V, corresponding to an electric field of ~ 5 MV/cm, were used to write and erase these devices. The charge is transferred between the silicon channel and the traps in the back ONO using Fowler-Nordheim tunneling. The substrate (an effective back-gate) is biased at high voltages (positive to write and negative to erase), and front-gate, source and drain terminals are grounded. For the 0.68 V threshold-voltage shift in this structure, we estimate a trapped charge density of $2.7 \times 10^{12}$ $cm^{-2}$ in the back insulator stack assuming the whole charge is located

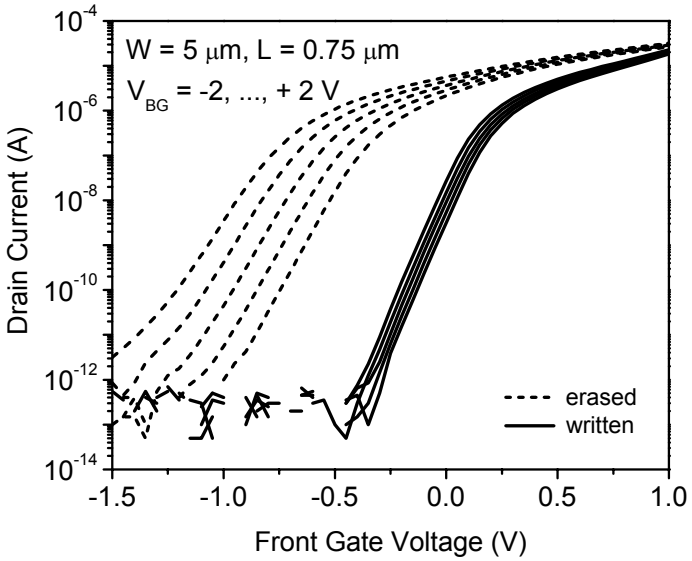



**Figure 4.10** Effect of back-gate bias in the erased (left set of curves) and written (right set) states of a back-side trapping memory. $V_{BG}$ = - 2, - 1, 0, + 1, + 2 V, from right to left in each set. $V_D$ = 1 V. The front gate oxide is 7 nm thick, the back oxide-nitride-oxide stack is 7, 20 and 80 nm respectively and the active silicon layer is ~ 40 nm. The same write and erase conditions were used as in Fig. 4.9.

half-way within the silicon nitride layer. In both write and erase bias schemes there is no voltage applied to any of the front transistor terminals. As a result there is no damage to the front (read) transistor during writing or erasing.

Figure 4.10 illustrates the effect of the back-gate bias on the characteristics of the front transistor in the erased state (left set of curves) and written state (right set of curves). For both the erased and the written states five transfer curves are obtained for five different back-gate voltages, -2, -1, 0, +1 and + 2 V. The threshold voltage modulation by the back gate is much larger in the erased state than in the written state. This is an evidence that, in the written state, charge is stored in the back, between the

silicon channel and the back-gate, causing an efficient shielding of the back-gate bias on the front device characteristics.

Figure 4.11 shows the transfer curves in the written and erased states for the front channel using the front gate (A) and for the back channel using the back gate (B) to compare back-side storage and front-side storage. Using the back gate to read the memory device is the same as operating a conventional front-side trapping device since the control gate (write/erase) is also the read gate in this case. As expected, the threshold voltage shift is much larger for the back channel, 3.4 V, than it is for the front channel, 0.55 V. As discussed in Chapter 2, this is because of the larger capacitive coupling between the front gate and the trapped charge as compared to the capacitive coupling between the back gate and the trapped charge that results in a smaller threshold voltage shift for the front gate for the same amount of trapped charge. Also, once an accumulation layer forms at the back channel, as a result of negative trapped charge density, the threshold voltage of the front transistor does not change further even if more charge is trapped in the back ONO. The accumulation layer pins the potential of the back interface preventing further coupling of the trapped charge with the front interface.

The charge density stored in the back ONO can be estimated from either the front or the back threshold voltage shift in Fig. 4.11. From the $\Delta V_T$ of the back transistor, $\Delta V_{T\text{-}BG}$, the charge density $\Delta Q_{TR}$ is estimated to be 8.6 x $10^{11}$ $cm^{-2}$ and from the $\Delta V_T$ of the front transistor, $\Delta V_{T\text{-}FG}$, the charge density is estimated to be 2.7 x $10^{12}$ $cm^{-2}$. The estimated value for $\Delta Q_{TR}$ using $\Delta V_{T\text{-}BG}$ is much more accurate since it is given by the capacitive coupling between the back gate and the trapped charge layer which are separated by insulator only (oxide and nitride):

$$\Delta Q_{TR} = \Delta V_{T-BG}\, C_{BG-TR} = \Delta V_{T-BG} \left( \frac{d_{co}}{\varepsilon_{ox}} + \frac{d_{nit-\overline{tr}}}{\varepsilon_{nit}} \right)^{-1} \qquad (5)$$

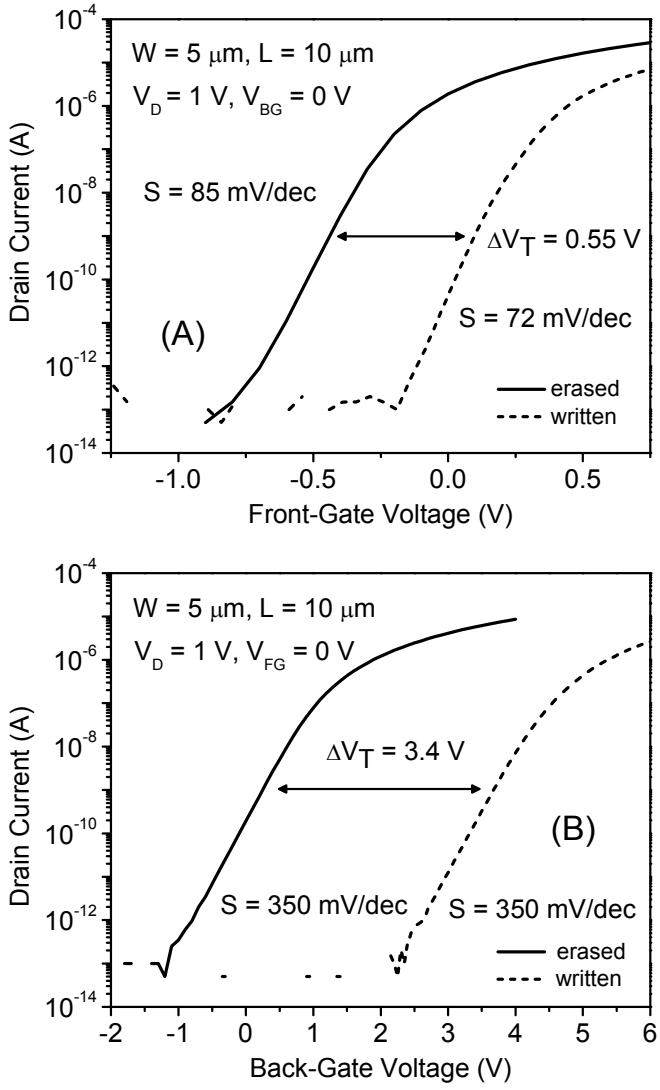



**Figure 4.11 (A)** Memory operation of a back-side trapping memory in erased (solid line) and written (dashed line). The write and erase voltages were + 45 V and - 45 V 300 ms pulses applied to the back-gate while front-gate, source and drain were grounded. Front gate oxide is 7 nm thick and back oxide-nitride-oxide stack is 7, 20 and 80 nm respectively. The silicon thickness is ~ 40 nm. **(B)** Effect of the same stored charge on the back channel transistor in the erased (solid line) and written (dashed line) states of the device for comparison between front side and back side storage. This is equivalent to a front side trapping memory.

where $d_{co}$ is the control oxide thickness and $d_{nit\text{-}tr}$ is the distance between the control oxide – nitride interface and the position of the stored charge within the nitride (assuming all the charge is concentrated at the same depth). $d_{nit\text{-}tr}$ was taken to be $d_{nit}/2$. $\varepsilon_{ox}$ and $\varepsilon_{nit}$ are the dielectric permittivity of silicon oxide and silicon nitride. The estimation of the trapped charge density from the front $\Delta V_{T\text{-}FG}$ is done using a simple model based on the capacitive coupling between the front gate and the trapped charge layer (Chapter 2). This model neglects charge effects that take place in the silicon channel, between the front gate and the trapped charge, such as inversion, depletion or accumulation charge. As a result, the charge density estimated from $\Delta V_{T\text{-}FG}$ is only a rough approximation and accurate values must be extracted from the threshold voltage shift on the back interface, $\Delta V_{T\text{-}BG}$.

The sub-threshold slope in the front channel operation, 72 mV/decade in the written state and 85 mV/decade in the erased state, is much better than the sub-threshold voltages of the back transistor, 350 mV/decade in both states (Fig. 4.11). The improved sub-threshold slope in the transfer curves compensates for the smaller memory window of the front transistor. In terms of the ability to distinguish between the written and erased states, by sensing the current at a voltage in between the two states threshold, a smaller memory window combined with a steeper sub-threshold slope is equivalent to a larger memory window with a poorer sub-threshold slope. The worse sub-threshold slope for the back transistor is due to the presence of the thick memory ONO stack between the silicon channel and the back gate compared to a thin gate oxide for the front transistor. The measured sub-threshold slopes for the front and back transistors are consistent with the ratios between the silicon body and the front and back gate dielectric thicknesses.

The differences between front side storage and back side storage observed in Fig. 4.11 clearly illustrate the possibility of further scaling using back side storage

devices. The silicon body potential can be efficiently controlled by the front gate, separated by a thin gate oxide, while charge is stored in a thicker stack in the back. As a result, the limits for gate scaling for back storage devices are the same as for fully depleted silicon MOSFETs.

### 4.2.2 Retention time

The industry standard for required retention time for non-volatile memories is a minimum of 10 years. Retention time is the time elapsed since the device was programmed to the time when the memory window narrows below a certain required

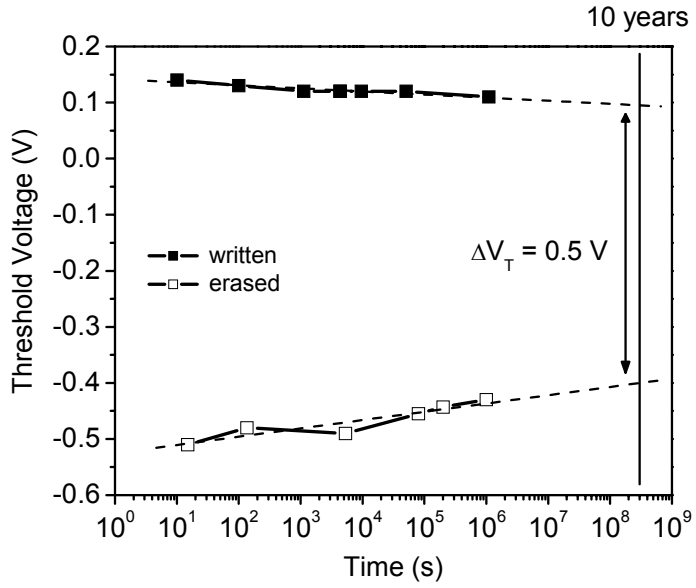



**Figure 4.12** Retention time characteristics for a back-side trapping memory. Threshold voltage as a function of the time elapsed after a 300 ms writing pulse of + 50 V (solid symbols) and after a 300 ms erasing pulse of – 35 V (open symbols) applied to the back-gate. The measurement was done at room temperature. W = 1.5 μm and L = 0.75 μm. Front gate oxide is 7 nm thick and back oxide-nitride-oxide stack is 7, 20 and 80 nm respectively. The silicon thickness is ~ 40 nm.

value. The retention time characteristics of a back side trapping memory device are shown in Figure 4.12.

The threshold voltage of the device shifts slowly over time due to charge leakage from the storage medium. The measurement was done by starting with two equivalent devices (neighbor devices so all the thickness parameters are identical), one in the written state and one in the erased state. The threshold voltage is measured for both devices in the initial state and at certain times afterwards. In between $V_T$ measurements there are no voltages applied to the devices. The measurement was done at room temperature and the last $V_T$ measurement was done 12 days after the devices had been written and erased. A memory window ($\Delta V_T$) larger than 0.5 V is maintained up to 12 days (approximately $10^6$ s).

The fact that the threshold voltage changes over time in both erased and written states points to charge leakage processes in both states. The $V_T$ decrease in the written state indicates negative charge leakage from the traps and the $V_T$ increase in the erased state indicates positive charge leakage from the traps. Positive charge can be injected into the traps during erase (over-erasing, discussed in 2.2.3). It is interesting to note that the change in $V_T$ in the erased state is faster than the change in $V_T$ in the written state. This indicates that the retention time for positive charge is worse than the retention time for negative charge. This may be due to more hopping sites for holes to leak from the traps or to a trap distribution for holes (spatial or energetic) that results in higher tunneling probability into the silicon layer, or control gate, compared to the trap distribution for electrons.

Lusky *et al.* [16] characterized the traps distribution within the silicon nitride band gap for electrons and holes using a GIDL (gate induced drain leakage) measurement technique. They found that the distribution of traps available for holes (bottom half of the band gap) spans a wider energy range than that for electrons (top

half of the band gap). Besides, the peak of the holes traps distribution is closer to the valence band edge than the peak of the electron traps distribution to the conduction band edge. These traps distributions would explain the worse retention time for holes compared to that for electrons.

As a result of the charge being stored in the back of the read transistor, minimal read-disturb effects are expected to lead to longer retention times in these devices when compared to front side trapping devices memories.

### 4.2.3 Cycling endurance

Cycling endurance is another important requirement for non-volatile memories. This refers to the ability of a device to withstand repeated write and erase cycles and still maintain the required memory window with the same write and erase conditions. In order to test the endurance of these memory devices, multiple write/erase cycles were applied, and the threshold voltage in both states measured after certain number of cycles. The erased and written characteristics continually reproduce approximately maintaining a good memory window larger than 0.6 V up to $10^5$ write/erase cycles (Figure 4.13). The written state of the device is very stable with minimal reduction or increase of threshold voltage after repeated cycling. The erased state is somewhat more erratic, with threshold voltage variations up to 0.2 V, but still maintains a reasonably wide memory window for the range of cycling tested.

In memory devices in which there is no self limiting mechanism for the amount of charge that can be stored, such as trapping memories in which a large density of traps is available, over-writing and over-erasing must be carefully avoided in order to achieve high cycling endurance. The over-writing or over-erasing of a back trapping device and its effect on the cycling endurance is illustrated in Figure 4.14. Here the same device is written and erased using different voltages and programming

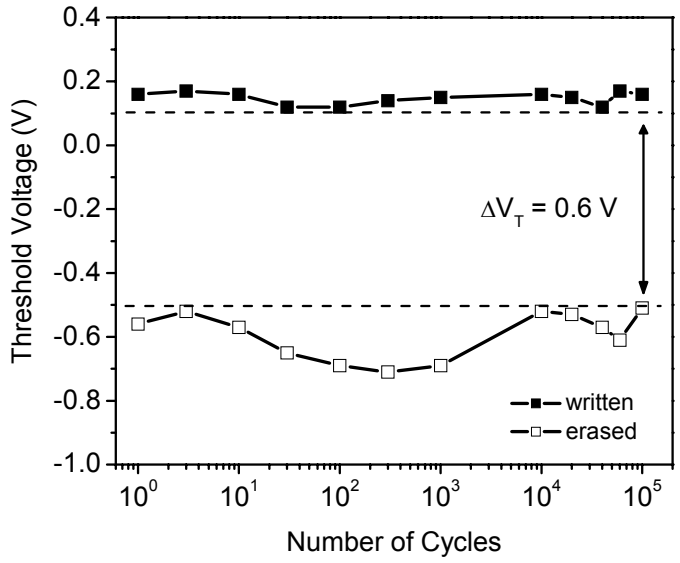



**Figure 4.13** Endurance characteristics for a back-side trapping memory. Threshold voltage in the erased (open symbols) and written (solid symbols) states after up to $10^5$ write-erase cycles. W = 3.0 μm and L = 1.0 μm. The write and erase voltages were + 50 V and - 35 V applied to the back-gate for 300 ms. The front-gate, source and drain were grounded during both write and erase operations. The back ONO stack is 7/20/80 nm and the active silicon layer is ~ 40 nm.

times. Figs. 4.14 A and B show write/erase conditions that result in over-writing (A) and over-erasing (B) of the device. Though not apparent up to 100 write/erase cycles, the accumulated over-writing/over-erasing effect becomes obvious after 1,000 cycles. In Fig. 4.14 C, by using programming conditions intermediate to those used in A and B, a stable memory window is achieved up to $10^5$ cycles. For different thickness and different compositions of materials employed, the programming voltages have to be selected after testing the device for a large number of write/erase cycles, typically more than $10^6$ cycles ($10^6$ cycles is a common endurance requirement).

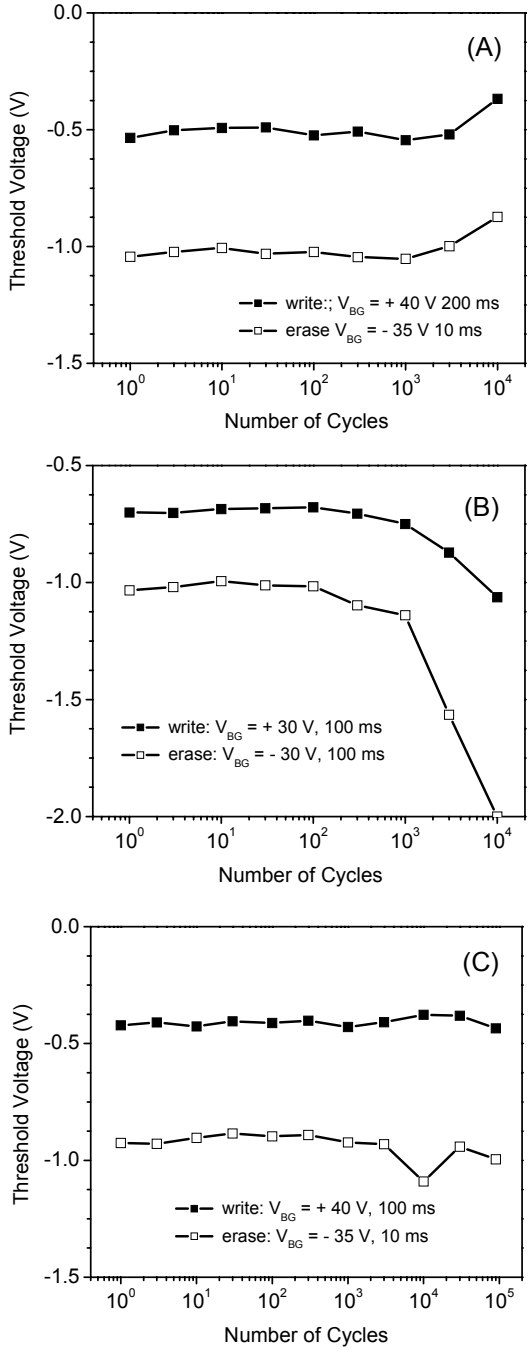



**Figure 4.14** Endurance characteristics for different write/erase conditions for write/erase optimization. W = 3 μm, L = 0.5 μm. **(A)** Over-writing causes threshold voltage of erased and written states to drift after $10^3$ cycles. **(B)** Over-erasing causes threshold voltage to drift after $10^3$ cycles. **(C)** Memory window approximately stable up to $10^5$ cycles.

### 4.2.4 Writing and erasing times

Other important parameters for non-volatile memories are the writing and erasing times, i.e., how fast the charge can be injected and removed, for a certain threshold voltage shift. Figure 4.15 A shows the threshold voltage for the front channel in a back trapping device, in written and erased states, as a function of writing and erasing time, starting from 1 μs and up to 300 s. The back ONO thickness is 7/20/80 nm and the write/erase voltages used are +/- 35 V, 37.5 V and 40 V. The threshold voltage for the back channel is also measured after the same programming times for accurate extraction of stored charge density (Fig. 4.15 B). For a front threshold voltage shift of 0.5 V, the writing time obtained is 30 ms, 250 ms, and 2.4 s for 40 V, 37.5 V and 35 V respectively. The erasing time for the same threshold voltage shift is 13 ms, 244 ms and 3.1 s for - 40 V, - 37.5 V and - 35 V respectively. The trapped charge saturates and starts decreasing at ~ 10 s of writing/erasing times. This corresponds to a maximum stored charge density, for the different voltages used, of 2.71 x $10^{12}$ $cm^{-2}$, 2.45 x $10^{12}$ $cm^{-2}$ and 1.89 x $10^{12}$ $cm^{-2}$, calculated from the threshold voltage shift on the back channel.

The charge saturation after long programming times is due to a change in the electric field direction once the trapped charge exceeds a certain value. This effect was discussed in section 2.2.2. The value of trapped charge at which saturation takes place increases with increasing programming voltage. The writing and erasing times for these devices are in the order of 10 - 30 ms for +/- 40 V (corresponding to a field of 4 MV/cm), for 0.5 V threshold voltage shift (corresponding to ~ 3 V if front side storage, see Fig. 4.11). These values are comparable to writing and erasing times for conventional non-volatile memory devices when programmed using Fowler-Nordheim tunneling. Since these times only depend on the thicknesses and voltages employed no difference is to be expected from the use of back side storage.

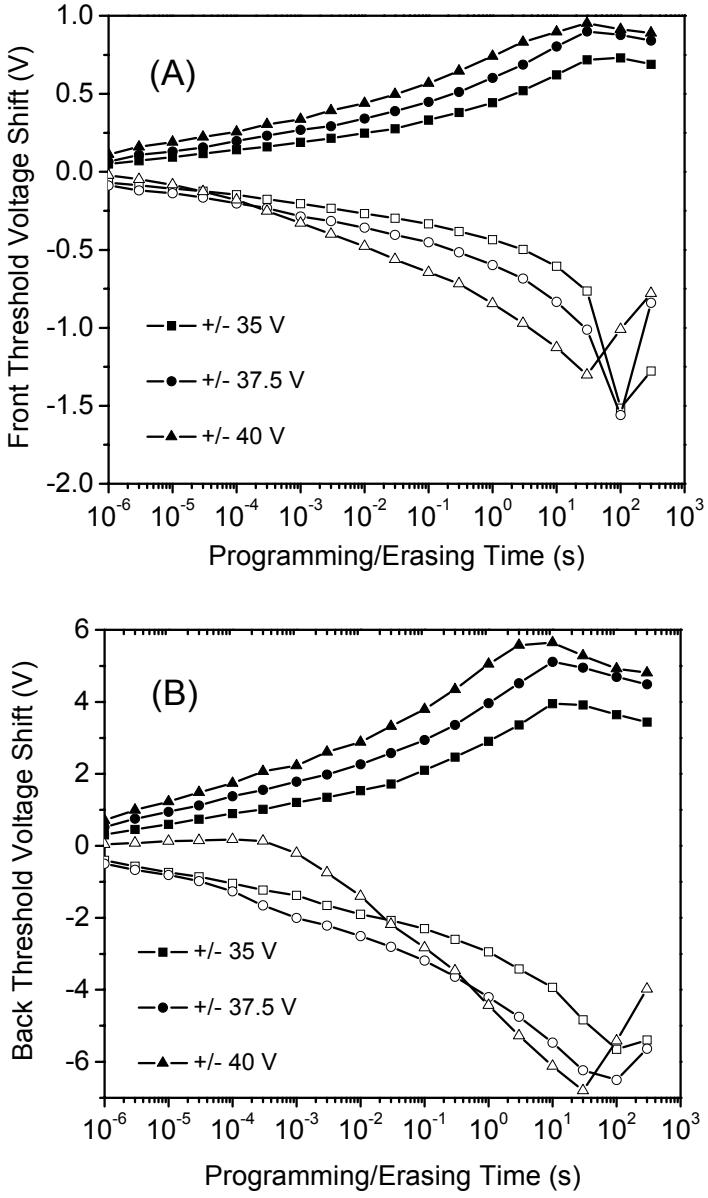



**Figure 4.15** Programming and erasing time characteristics of a back-side trapping device. Threshold voltage shift, $\Delta V_T$, is plotted as a function of writing and erasing time for three different write/erase voltages, +/- 35 V, +/- 37.5 V and +/- 40 V. W = 3 μm, L = 0.5 μm. **(A)** Threshold voltage of the front transistor. **(B)** Threshold voltage of the back transistor.

### 4.2.5 Small scale memory devices

Figures 4.16 to 4.18 show memory characteristics for shorter gate length back side trapping devices, 150 nm, 100 nm and 50 nm respectively. The front gate oxide is 6 nm thick, the back ONO stack was reduced to 3/4/7 nm, and the silicon body is ~ 15 nm. The use of thinner ONO allows the use of standard programming voltages, ~ 10 V, and the use of thinner silicon body makes for better coupling between the charge stored at the back and the front transistor channel (larger threshold voltage shift for the same amount of stored charge). For the 150 nm gate length device (Fig. 4.16) the sub-threshold slope is 127 mV/decade in the erased state and 94 mV/decade in the written state. The memory window is 0.75 V obtained with 300 ms write/erase pulses of +10/-10 V applied to the back-gate with the front terminals grounded.

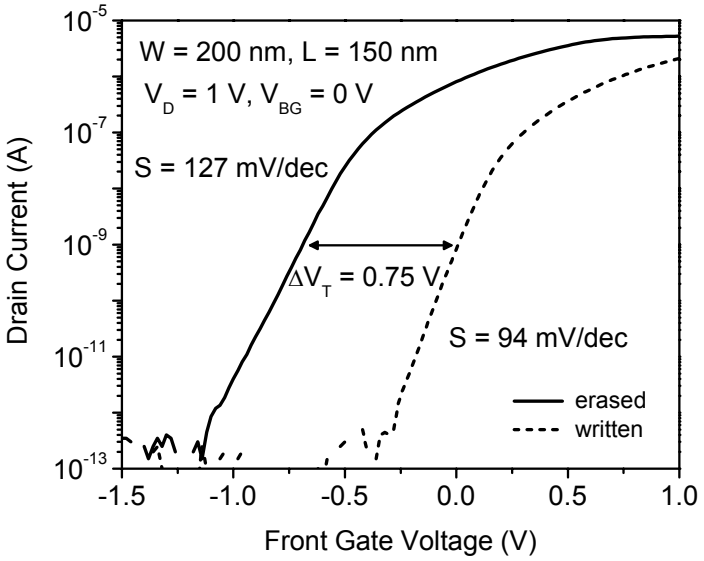



**Figure 4.16** Memory operation of a back-side trapping memory device with gate length of 150 nm. The write and erase voltages were + 8 V and – 8 V 300 ms pulses applied to the back gate with front gate, source and drain grounded. Front gate oxide is 6 nm and the back ONO stack is ~ 3/4/7 nm. The silicon thickness is ~ 15 nm.

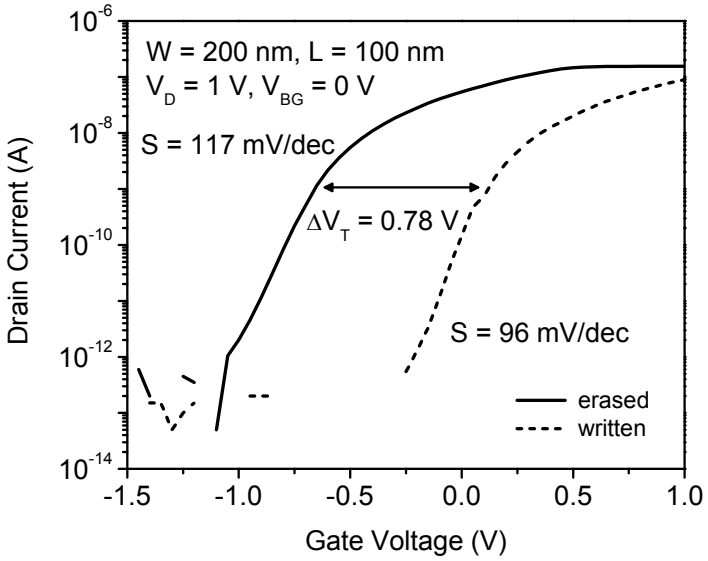



**Figure 4.17** Memory operation of a back-side trapping memory device with gate length of 100 nm. The write and erase voltages were + 8 V and – 8 V 300 ms pulses applied to the back gate with front gate, source and drain grounded. Front gate oxide is 6 nm and the back ONO stack is ~ 3/4/7 nm. The silicon thickness is ~ 15 nm.

For the 100 nm gate length device (Fig. 4.17) the sub-threshold slope is 117 mV/decade in the erased state and 96 mV/decade in the written state and the memory window is 0.78 V, with the same programming conditions. As mentioned earlier, the sub-threshold slope variations from device to device are most likely due to silicon film thickness variation. Fig. 4.18 shows the memory characteristics for a 50 nm gate length device. The transfer curves in written and erased state for the front channel are shown in Fig. 4.18 A. The effect of the same trapped charge on the back channel transfer characteristics is shown in Fig. 4.18 B. The device was written and erased with +/- 8 V applied on the back-gate with all front terminals grounded. The improved front channel sub-threshold slope in the written state, discussed in section 4.2.1, is also

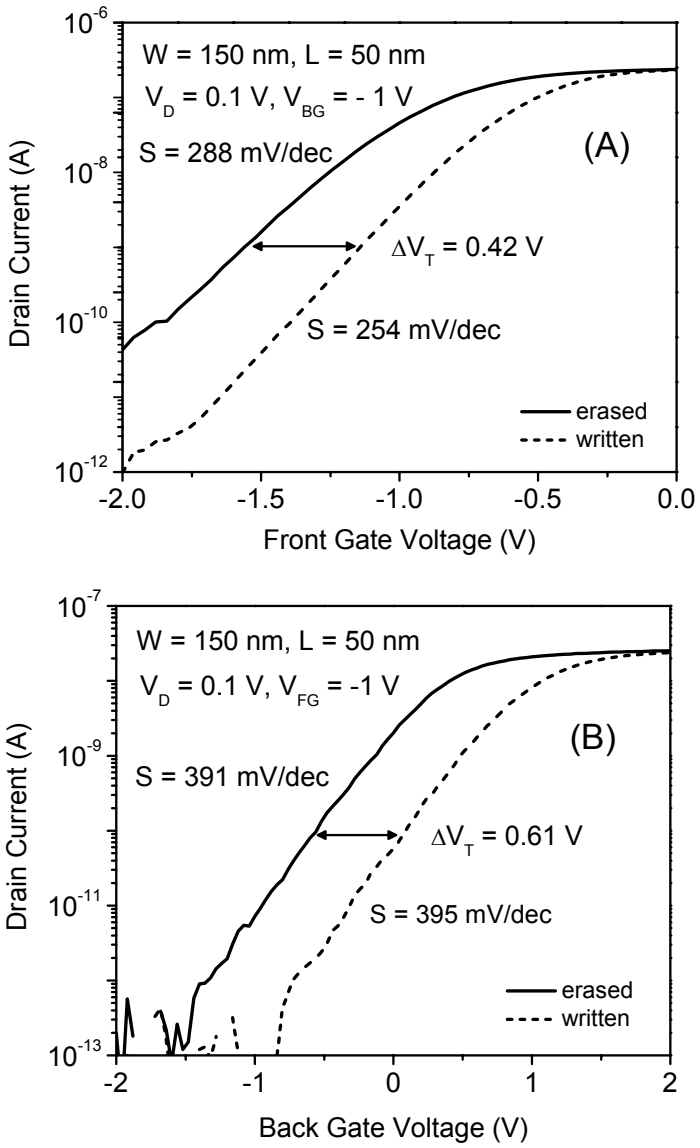



**Figure 4.18 (A)** Memory operation of a 50 nm gate length back-side trapping memory in erased (solid line) and written (dashed line). The write and erase voltages were + 8 V and - 8 V 300 ms pulses applied to the back-gate while front-gate, source and drain were grounded. Front gate oxide is 6 nm and the back ONO is 3/4/7 nm. The silicon body is ~ 15 nm. **(B)** Effect of the same stored charge on the back channel transistor in the erased (solid line) and written (dashed line) states.

observed in these characteristics. The sub-threshold slope in the erased state is 288 mV/dec compared to 254 mV/dec in the written state. As expected, for the back transistor there is no significant change in sub-threshold slope in the written state, 395 mV/dec, and the erased state, 391 mV/dec. The significantly better sub-threshold slope for the front (read) transistor compared to the back transistor, at 50 nm gate length, demonstrates once again the advantage of back-side storage versus front-side storage. The poorer sub-threshold slope observed in this device is probably due to thinner silicon body in this area of the wafer. The threshold voltage shift obtained for the front transistor is 0.42 V. The back ONO is ~ 3/4/7 nm and the active area of the device is 150 nm x 50 nm. From the threshold voltage shift on the back channel, $\Delta V_{T\text{-}BG} = 0.61$ V, the trapped charge density can be accurately determined using (5). Assuming all the charge is located in the middle of the nitride layer, this threshold voltage shift corresponds to approximately 120 electrons stored in the back ONO in the written state of this device.

## 4.3 Summary

The fabricated devices show good front and back interface transistor characteristics down to 20 nm gate length. The use of thin silicon body and thin front and back gate dielectrics allows efficient coupling of each interface with the corresponding back-gate. These double-gate characteristics can be used to improve channel potential control at very short gate lengths and for threshold voltage tuning of the opposite interface transistor for power adaptive applications. Reducing the front gate oxide will lead to improvements in the sub-threshold slope of the front transistor to near ideal values making for high-performance standard logic devices.

Back side trapping storage was demonstrated in devices down to 50 nm gate

length. Retention time, cycling endurance and writing and erasing times in these devices were characterized and found to comparable to those of conventional front side trapping memories. This was expected since they only depend on the thicknesses employed in the back. The possibility of further scaling for silicon non-volatile memories with back side trapping storage was demonstrated by the improved sub-threshold slope of the front transistors in written and erased states, compared to the sub-threshold slope of the back interface transistors. Improved sub-threshold slope allows scaling to smaller gate lengths due to low off current and low voltage operation. The thinning of the back trapping films thickness as well as the thinning of the silicon active channel can reduce the writing and erasing voltages, reduce programming time and increase the memory window but, as with front side storage devices, there will be retention and reliability issues associated with such a design.

# Chapter 5

## Electron mobility in charge trapping devices

Carrier mobility is a key performance parameter in semiconductor devices. In silicon non-volatile memories it partially determines the speed at which the devices can be read. From a more fundamental perspective, carrier mobility in a transistor channel also gives information on the properties of the interface between the inversion layer and the gate dielectric. Back-side trapping devices provide two interfaces for conduction, the front silicon interface and the back silicon interface, the latter being in close proximity to a region with large density of traps (the ONO system). Several interesting questions arise related to mobility in these structures. In this chapter some of these questions are addressed: (1) whether the mobility in the front and back silicon interfaces in back-side trapping devices is the same; (2) the effect of back bias on the mobility of the front silicon interface; (3) the effect of charge stored in the back ONO on the mobility of the front silicon interface; (4) the effect on mobility of charge stored in close proximity to the surface using front-side ONO trapping devices in which the tunneling oxide is 3 nm.

## 5.1 Effective mobility in MOSFETs

The effective mobility in a silicon MOSFET, $\mu_{eff}$, is significantly lower than the mobility in bulk silicon due to increased scattering mechanisms in the channel. $\mu_{eff}$ can be experimentally derived from the transfer characteristics $I_D(V_G)$ in the linear regime (small drain voltages), together with the capacitance characteristics, $C_{G\text{-}SD}$-$V_G$ [17].

In the linear region, the drain to source current $I_D(V_G)$ is given by:

$$I_D(V_G) = \frac{Q_{inv}(V_G)}{t} = \frac{Q_{inv}(V_G)\,v}{L} = \frac{Q_{inv}(V_G)\,\mu_{eff}(V_G)\,E}{L} = \frac{Q_{inv}(V_G)\,\mu_{eff}(V_G)V_D}{L^2} \qquad (1)$$

where $Q_{inv}$ is the inversion charge, $t$ is the transit time, $v$ is the average electron velocity, L is the channel length and $E$ is the longitudinal (drain induced) electric field. The inversion charge, approximated to be constant along the channel for small drain bias, is given by:

$$Q_{inv}(V_G) = \int_{V_T}^{V_G} C(V_G')\,dV_G' \qquad (2)$$

Once the $I_D(V_G)$ and $C(V_G)$ characteristics are known the effective mobility can be calculated:

$$\mu_{eff}(V_G) = \frac{I_D(V_G)L^2}{V_D \int_{V_T}^{V_G} C(V_G')\,dV_G'} \qquad (3)$$

The effective surface mobility decreases with increasing transversal field (increased $V_G$) as surface roughness scattering increases [17]. For the silicon-silicon oxide interface there is a 'universal mobility curve' (Takagi *et al*. [18]) as a function of effective transversal electric field that is independent of device parameters such as channel doping and silicon surface orientation. However for devices with channel doping concentration larger than 7 x $10^{16}$ $cm^{-3}$ the electron mobility deviates significantly from this curve. The devices used to study mobility in charge trapping memories have a doping concentration of ~ 8 x $10^{16}$ $cm^{-3}$ and therefore their mobility cannot be compared to the universal mobility curve.

The electron mobility values reported by Takagi *et al.* for devices with doping concentration of 7.2 x $10^{16}$ $cm^{-3}$ are ~ 500 $cm^2$/V.s in the low field region (near threshold) and below 200 $cm^2$/V.s in the high field region [18].

## 5.2 Mobility at the front and back silicon interfaces

A difference in mobility in the front and back silicon interfaces could be expected from different processes that both surfaces experience that could either degrade the mobility due to roughness and defects or enhance it due to induced stress from the bond via an oxide-nitride-oxide stack. Figures 5.1 and 5.2 show the $I_D(V_{FG})$ and $C(V_{FG})$ characteristics for the front and back silicon interfaces respectively, for the same back-side trapping device with charge stored in the back ONO . This device has a silicon body of ~ 40 nm, front-gate oxide is 7 nm and back ONO ~ 7/20/80 nm.

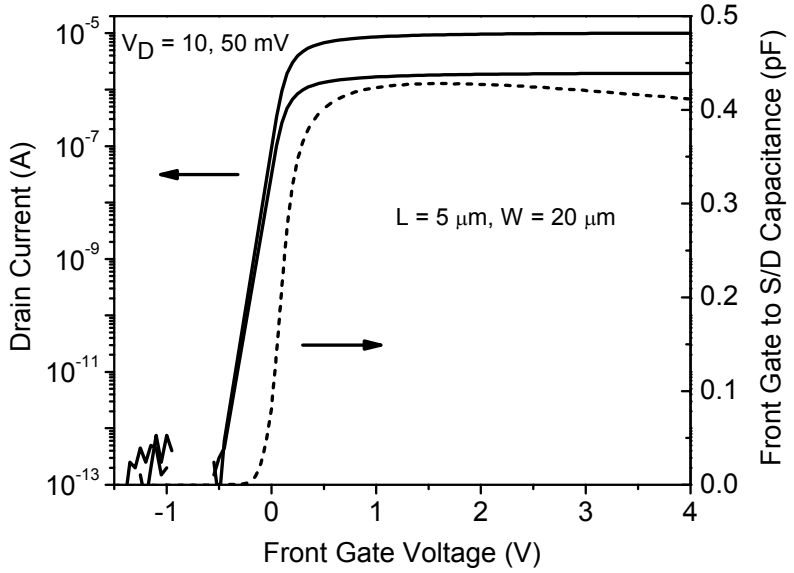



**Figure 5.1** Front silicon channel characteristics, $I_D$-$V_{FG}$ and $C_{FG\text{-}SD}$-$V_{FG}$. $V_{BG}$ = -1 V.

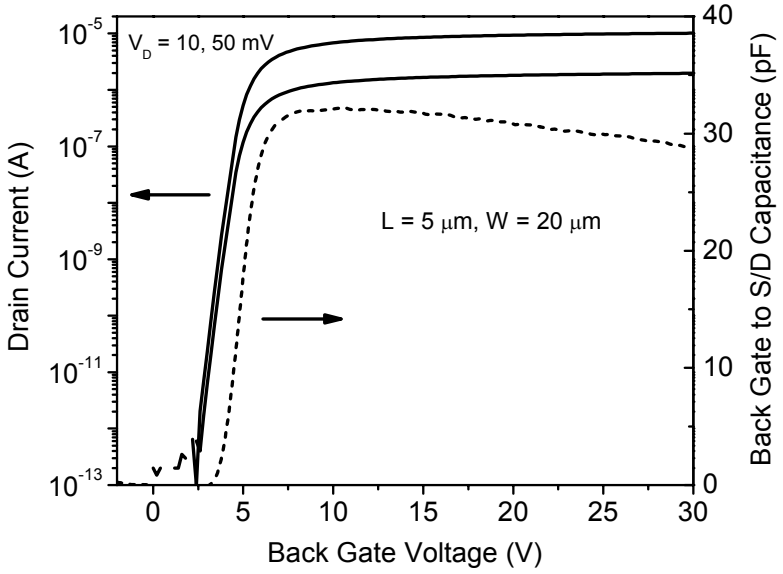



**Figure 5.2** Back silicon channel characteristics, $I_D$-$V_{BG}$ and $C_{BG\text{-}SD}$-$V_{BG}$, for the same device as in Figure 5.1. $V_{FG}$ = -1 V.

The channel width is 20 μm and channel length is 5 μm. The opposite gate in each case (back-gate for the front interface characteristics and vice-versa) is biased at -1 V so that the opposite interface is off and does not contribute to the drain current. If the back interface contributed to the drain current the mobility values for the front interface, calculated using the front inversion charge, would not be correct. The drain is biased at 10 mV and 50 mV. The capacitance is measured at 1 MHz with a 20 mV AC signal. The measured front-gate to source-drain capacitance is purely due to the channel capacitance (apart a very small contribution from the source and drain regions that overlap with the gate). The measured back-gate to source-drain capacitance however, includes the pads overlap capacitance since the whole substrate is used as a back-gate. This overlap capacitance, constant with $V_{BG,}$ has to be subtracted from the total capacitance in order to obtain the back-gate to source-drain capacitance to extract the correct back interface inversion charge.

From these characteristics the mobility for the front and back silicon interfaces is calculated. The effective electron mobility for the front interface and back interface is plotted in Figures 5.3 and 5.4 respectively. The peak mobility is approximately the same for both interfaces, ~ 350 $cm^2$/V.s. Since the thicknesses of the front and back-gate insulators are different the mobility for both interfaces must be compared at the same transverse electric field or inversion charge density. For a constant opposite gate bias (or constant charge stored in the back ONO) the surface electric field is proportional to the inversion charge density. In Figure 5.5 the mobility for both interfaces is plotted as a function of the inversion charge density ($Q_{inv}$/A), where A is the active area of the device. From this plot the high-field mobility is also found to be very similar for both front and back silicon interfaces, approximately 60 $cm^2$/V.s. It is important to note that this low value correspond to front and back-gate values of + 4 V and + 30 V, when the respective threshold voltages are – 0.2 V and + 3 V (see Figs. 5.1 and 5.2). This high field region is well beyond the normal operation region of the devices.

The similar values for front and back silicon interfaces in both low-field and high-field regions indicate that the two surfaces have very similar intrinsic properties. Since these measurements were done in the written state of the device (charge stored in the back ONO) this result also points to the fact that charge stored in nitride traps at a distance of 7 nm (the tunneling oxide) does not degrade the electron mobility at the neighboring silicon surface. The effect of charge stored in nitride traps in closer proximity to the silicon channel is discussed in section 5.4.

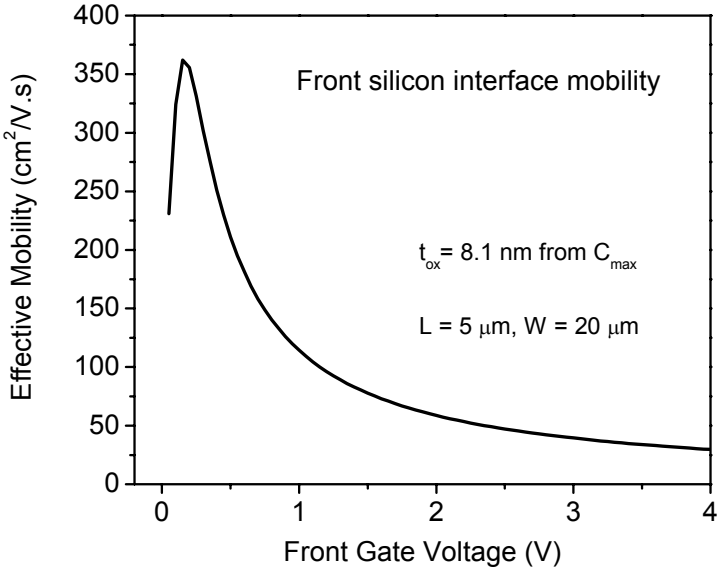



**Figure 5.3** Effective electron mobility for the front silicon channel derived from the $I_D$-$V_{FG}$ and $C_{FG-SD}$-$V_{FG}$ characteristics in Fig. 5.1.

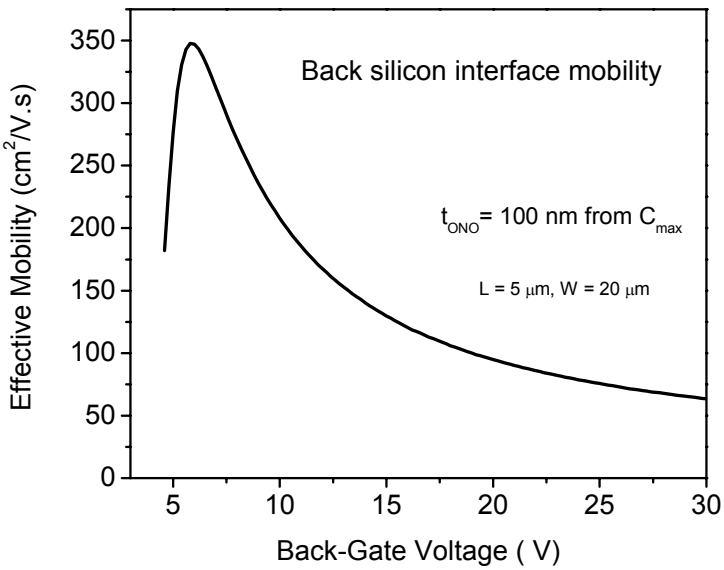



**Figure 5.4** Effective electron mobility for the front silicon channel derived from the $I_D$-$V_{BG}$ and $C_{BG-SD}$-$V_{BG}$ characteristics in Fig. 5.2.

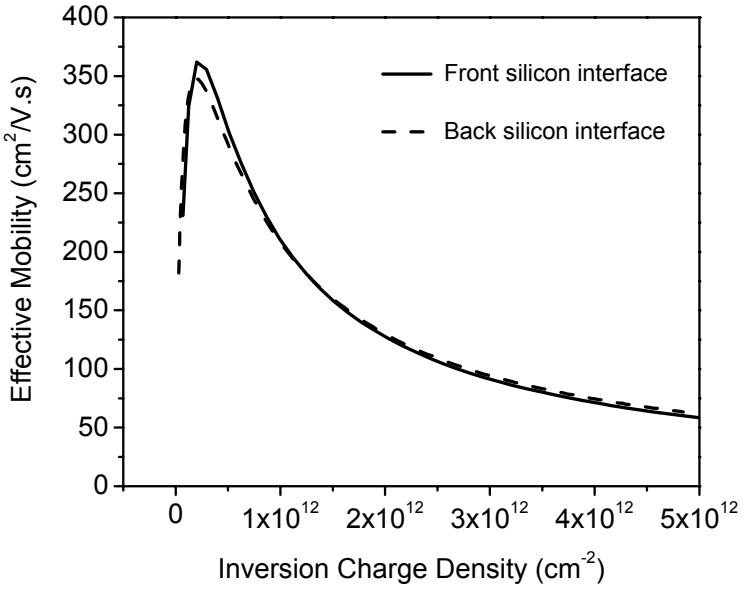



**Figure 5.5** Effective mobility for the front and back silicon channel for the same device (Figs 5.4 and 5.5) plotted as a function of the inversion charge density.

## 5.3 Mobility at the front interface vs. back gate bias

Figure 5.6 shows the transfer curves, $I_D$-$V_{FG}$, for the front silicon interface transistor, for different back-gate voltages, in the erased state of the device (no charge stored in the ONO). The silicon channel is ~ 40 nm thick, front-gate oxide is 7 nm and the back ONO is ~ 7/20/80 nm. The active area of the device is 100 μm x 100 μm. The front-gate is swept from - 2 to + 2 V and the back-gate is stepped between - 7 V and + 7 V in 1 V steps. These curves show the double-gate operation of these devices, in which there can be conduction at a single interface, front or back, or at both interfaces. When the back interface is off, there is efficient coupling between the back-gate and the front silicon interface.

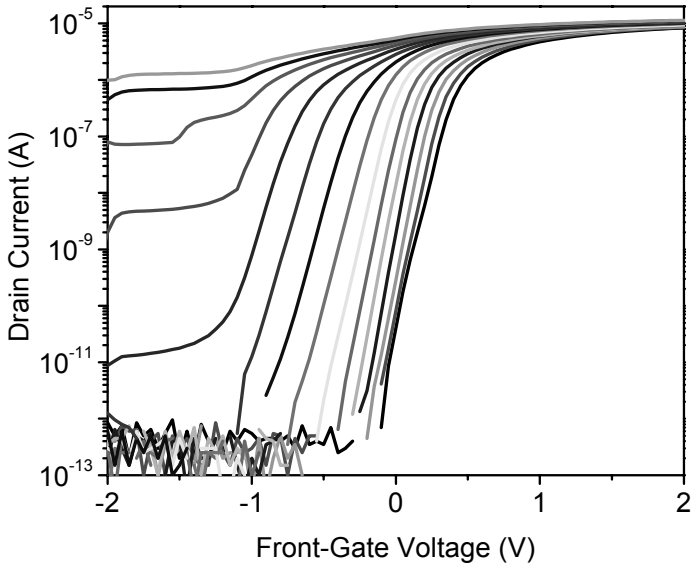



**Figure 5.6** Front channel transfer characteristics $I_D$-$V_{FG}$ for different back-gate voltages. $V_{BG}$ varies from -7 V (right most curve) to + 7 V (left most curve) in steps of 1 V. $V_D$ = 50 mV. L = W = 100 μm.

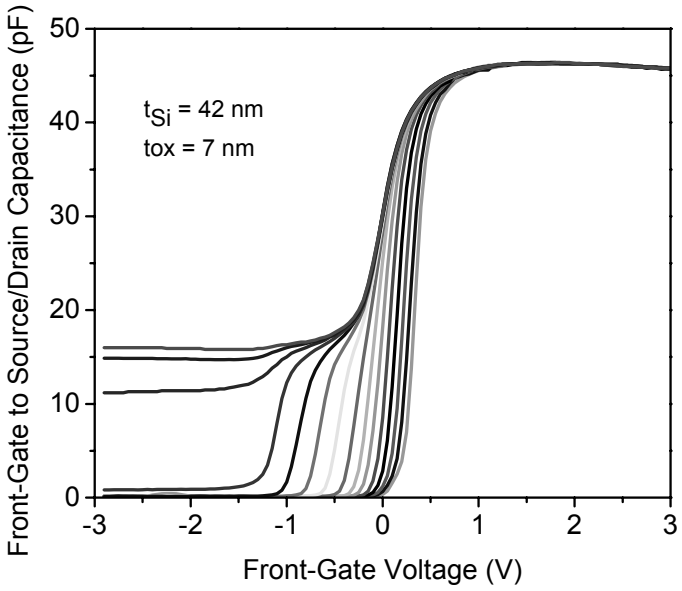



**Figure 5.7** Front channel $C_{FG\text{-}SD}$-$V_{FG}$ characteristics for different back-gate voltages for the same device as in Fig. 5.6. $V_{BG}$ varies from -7 V (right most curve) to + 7 V (left most curve) in steps of 1 V. L = W = 100 μm.

This is observed in the transfer curves as an increase of the threshold voltage and improved sub-threshold slope of the front transistor as the back-gate bias becomes more negative. These effects are observed in SOI devices [11], back floating gate devices [9] and double-gate devices [19]. Figure 5.7 shows the $C(V_{FG})$ curves for the same device, for the same back-gate bias in the transfer curves. Here again the double-gate characteristics are observed.

The mobility curves for the front interface for different back-gate bias are plotted in Figure 5.8. The mobility is calculated for back-gate bias from - 7 V up to + 4 V. Beyond + 4 V on the back-gate the conduction due to the back interface increases significantly (see Fig. 5.6) and the front inversion charge derived from the C-V measurement does not correspond to the total current measured $I_D(V_{FG})$.

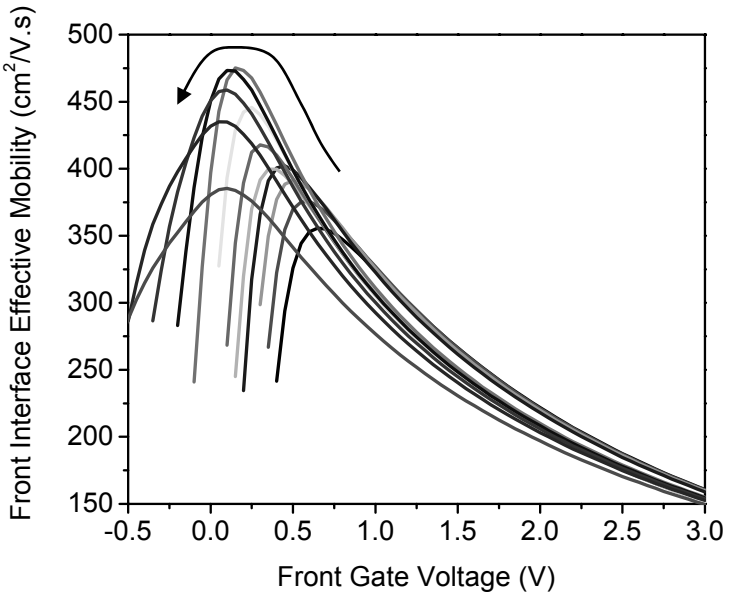



**Figure 5.8** Effective mobility for the front channel for different back-gate voltages derived from the $I_D$-$V_{FG}$ and $C_{FG-SD}$-$V_{FG}$ data in Fig. 5.6 and 5.7. $V_{BG}$ varies from -7 V (most right curve) to 4 V (most left curve) in steps of 1 V.

For the most negative back-gate bias the peak mobility is ~ 350 $cm^2/V.s$ and as the back-gate bias approaches zero this value increases to ~ 475 $cm^2/V.s$. The peak mobility versus the back-gate bias obtained for two different devices is plotted in Figure 5.9. The maximum peak mobility occurs at $V_{BG} = 0$ V for both devices, which probably corresponds to the lowest transverse electric field. For both positive and negative back gate bias the peak mobility decreases with increasing bias. In the high-field region however the mobility is between ~ 150 $cm^2/V.s$ and 160 $cm^2/V.s$ for all back-gate bias. The slight variation in the high field mobility is likely due to the different threshold voltage for different back-gate bias. The fact that the high field mobility is approximately independent of the back-gate bias indicates that, in a silicon channel ~ 40 nm thick, an accumulation layer at the back silicon interface does not change the electron mobility at the front silicon interface.

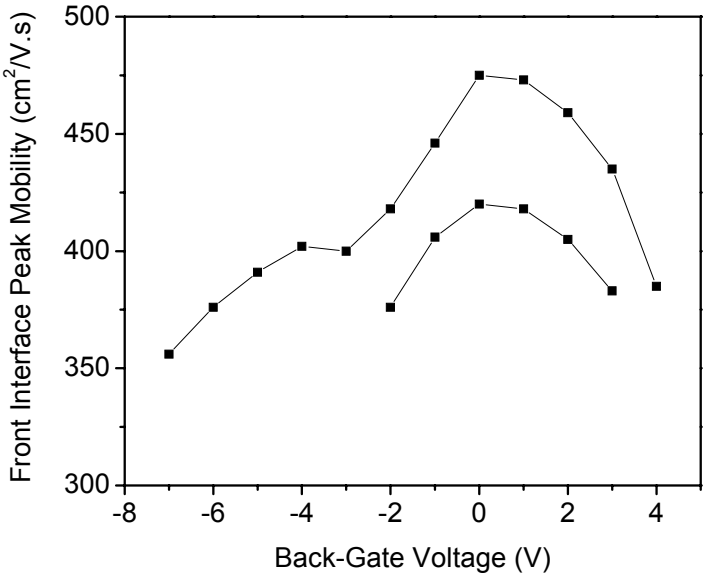



**Figure 5.9** Front channel peak mobility as a function of back-gate voltage for two different devices. The higher curve corresponds to the data plotted in Fig. 5.8.

## 5.4 Mobility at the front interface in erased and written states

The effect of charge stored in the back ONO on the mobility at the front interface is expected to be similar to the effect of back bias described in the previous section. To confirm this, mobility for the front silicon interface in the written and erased states of a back side trapping device was extracted. This device has W = 0.75 μm and L = 10 μm, the silicon thickness is ~ 40 nm and the back ONO is ~ 7/20/80 nm. Figure 5.10 shows the effective mobility for the front silicon interface when there is no charge and when there is charge stored in the back ONO.

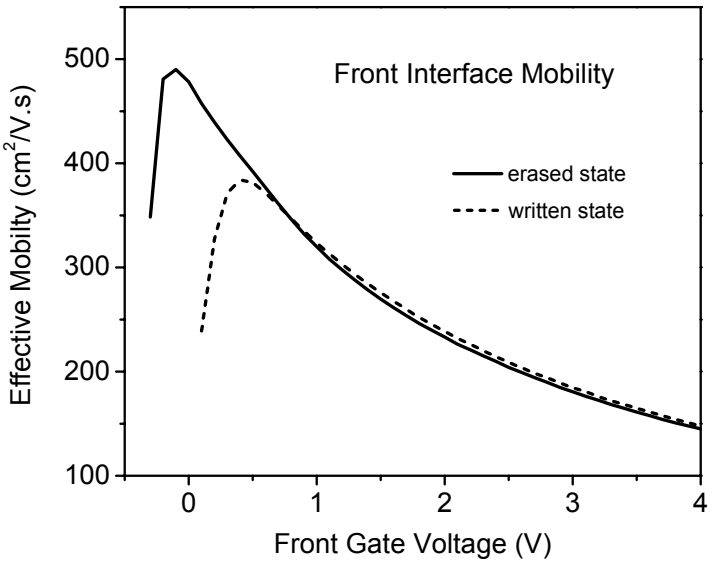



**Figure 5.10** Front channel effective mobility in erased (no charge) and written (charge stored in the back ONO) states as a function of front-gate voltage. W = 0.75 μm and L = 10 μm. The silicon thickness is ~ 40 nm and the tunneling oxide is 7 nm.

In Figure 5.11 the same mobility data is plotted as a function of the inversion charge density. The mobility is found to converge for the high field region, ~ 150 $cm^2/V.s$. The lower peak mobility in the written state, ~ 380 $cm^2/V.s$, compared to the erased state, ~ 480 $cm^2/V.s$, is again due to the higher effective transverse electric field at the silicon surface in the written state. As in section 5.2 it is observed that charge stored in the back ONO does not intrinsically degrade the mobility at the front silicon interface.

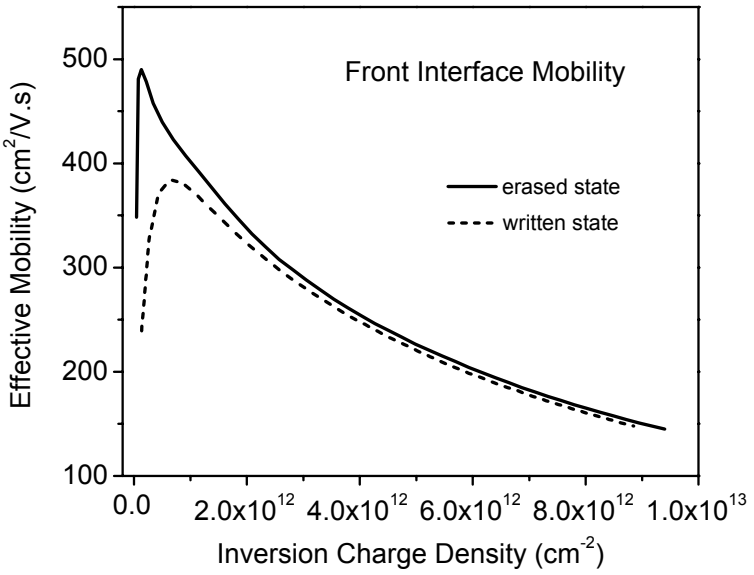



**Figure 5.11** Front channel effective mobility in erased (no charge) and written (charge stored in the back ONO) states as a function of inversion charge density. W = 0.75 μm and L = 10 μm.

## 5.5 Effect on the mobility of charge stored in close proximity to the channel

In section 5.2 it was concluded that charge stored in nitride traps at a distance of 7 nm does not degrade the mobility in the channel. In order to study the effect of charge stored in closer proximity to the channel, front charge trapping devices were used in which the tunneling oxide is ~ 3 nm. The mobility in the channel was extracted for different trapped charge densities. Figure 5.12 (A) shows the transfer curves with small drain bias in the initial state of the device and after different writing times, from 500 μs to 1 s. The device was written using Fowler-Nordheim tunneling with + 12 V applied on the gate. The $C(V_G)$ characteristics were measured at the same initial and written states and the resulting extracted mobility is shown in Figure 5.12 (B). In Figure 5.13 the mobility for the different states, initial and after different writing times, is plotted as a function of the inversion charge density. The mobility curves converge in the high-field region showing that the high-field mobility is independent of charge trapped in nitride at a distance of ~ 3 nm.

The trapped charge density can be calculated from the threshold voltage shift observed in the transfer curves in Fig. 5.12 A (see chapter 2). Figure 5.4 shows the peak mobility plotted versus the trapped charge density showing the initial significant drop corresponding to a trapped charge of ~ $4 \times 10^{10}$ $cm^{-2}$.

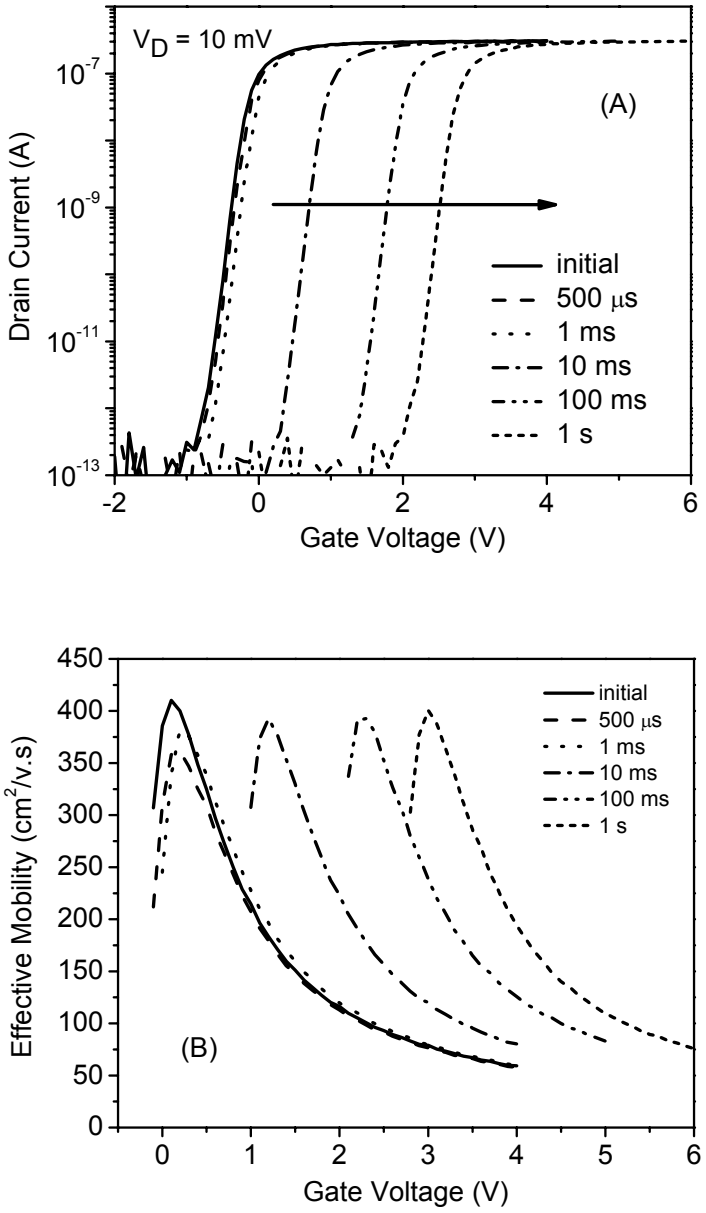



**Figure 5.12** Transfer characteristics $I_D$-$V_G$ (**A**) and effective mobility as a function of the gate voltage (**B**) for a front-side trapping device in the initial state and after different writing times (accumulated). The write voltage was 12 V. The tunneling oxide, nitride and control oxide is ~ 30/70/300 Å. W = L = 2 μm.

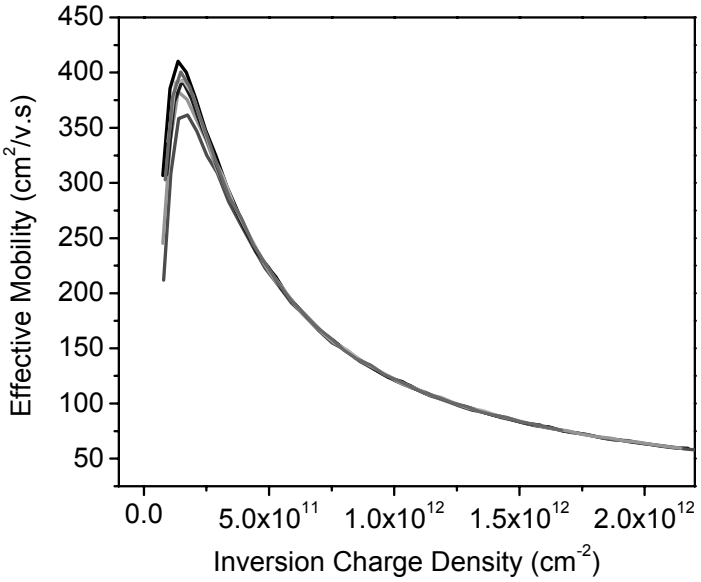



**Figure 5.13** Effective mobility in the initial state and after different writing times (Fig. 5.12) plotted as a function of inversion charge density. The peak mobility varies but the high field mobility is independent of trapped charge in the nitride.

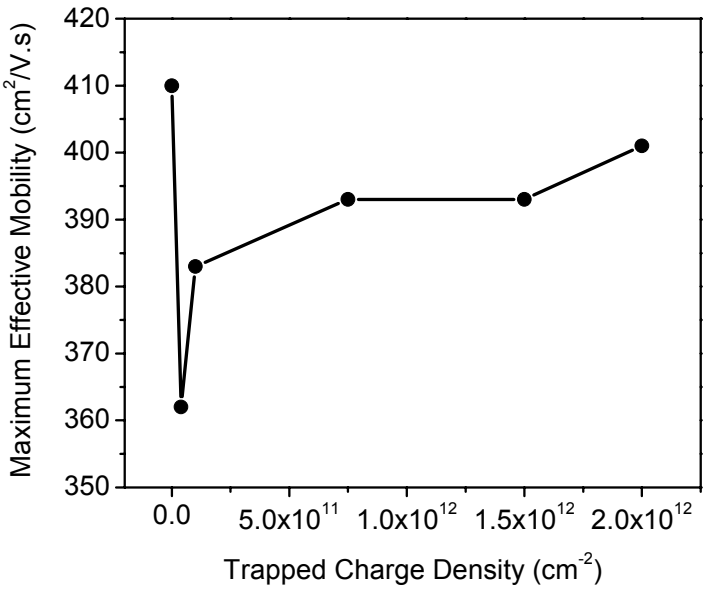



**Figure 5.14** Variation of the peak mobility in Fig. 5.12 with the trapped charged density derived from the threshold voltage shift.

It is important to note, regarding results presented in this chapter, that the peak mobility values, which correspond to bias conditions around the threshold voltage, are very sensitive to small threshold voltage shifts. A small threshold voltage shift between the I-V measurement and the C-V measurement due to a small loss/gain of charge by neighboring traps can cause a significant error in the calculated peak mobility. As a result, the peak mobility values are less accurate and reliable than the high field mobility values.

## 5.6 Summary

It was found that the electron mobility at the front and back silicon interfaces in back-side trapping devices is approximately the same which indicates that both interfaces have very similar properties. Charge stored in the back ONO or negative back-gate bias do not alter the front interface mobility and only increase the transversal electric field on the silicon surface resulting in a lower peak mobility. It was also determined that the silicon channel mobility is not degraded with charge stored in nitride traps as close as 3 nm from the channel. The mobility values obtained, ~ 450 $cm^2/V.s$ in the lowest transversal electric field (no charge stored in the ONO and zero back-gate bias) and ~ 150 $cm^2/V.s$ in high field (up to normal device operation conditions) are consistent with available data in the literature for the silicon-silicon dioxide interface with similar channel doping concentrations. The conclusion is that the mobility at both front and back silicon interfaces in charge trapping devices is not degraded nor improved when compared to conventional silicon MOSFETs.

# Chapter 6

# Individual trap characterization using Random Telegraph Signal

## 6.1 Random Telegraph Signal

Charge trapping and de-trapping events caused by individual traps near the silicon – silicon dioxide interface are observable in small devices as a discrete switching in the current between two or more levels under constant bias conditions. These events can give information about the traps involved in the process. In current technology the trap density in the silicon – silicon oxide interface is approximately $10^{10}$ $cm^{-2}$ and the charge density in the channel in weak inversion is approximately $10^{12}$ $cm^{-2}$; a small transistor, with an active area of 100 nm x 100 nm, will have roughly 1 – 10 traps at the silicon – silicon oxide interface and 100 electrons in the channel in weak inversion. The effect of an individual carrier being trapped and de-trapped by an individual trap has therefore a relatively large and noticeable effect in the characteristics of small devices.

This phenomenon has long been observed in sub-micrometer transistors and became known as Random Telegraph Signal (RTS) or random telegraph noise and has been a favored tool in the study of individual traps in the Si-$SiO_2$ system. Ralls' study on RTS in sub-micrometer transistors in 1984 [20] is perhaps one of the earliest. Later, Kirton and Uren [21] and Mueller and Schulz [22] have published review works with extensive information on RTS. RTS has also been observed and studied in a variety of devices other than transistors such as metallic nano-constrictions, memory devices, or photoluminescence devices [23, 24].

A schematic of the RTS phenomenon in a MOSFET is shown in Figure 6.1. The charge transfer events affect both the number and the mobility of the carriers in the channel resulting in a change in current that corresponds to carrier capture (low current level) and emission (high current level) by a single trap. From the bias dependence of the capture and emission times the location of the traps can be determined. The temperature dependence of these time constants gives information on the energy levels of the traps and their capture cross-section.

Important questions that arise are how deep into the gate dielectric can RTS reveal individual traps through its Coulombic interactions, and whether RTS can be used to study traps in devices where trapping is actively employed for achieving memory properties such as charge trapping non-volatile memories. In order to address these questions we have used RTS to characterize individual traps in back side charge trapping memories.

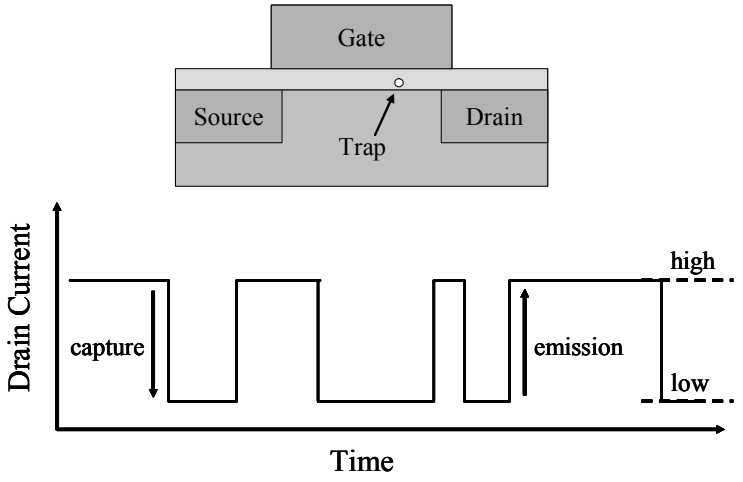



**Figure 6.1** Schematics of a Random Telegraph Signal (RTS) event. When a carrier is captured by a trap located at the silicon – silicon oxide interface or within the silicon oxide, the current level switches from high to low, and vice versa when the carrier is emitted back into the channel.

Due to the presence of an oxidized gated interface on front and back of a thin single-crystal silicon channel, back side trapping devices provide a tool for bias-dependent characterization at multiple interfaces and combinations of materials (see Fig. 2.1 for example).

## 6.2 RTS measurements and analysis

RTS was observed due to traps in both front (oxide alone) and back (ONO) gate stacks in back-side trapping memories as well as in other devices, such as dual oxide transistors. These are transistors in which the gate dielectric is formed by two oxides and their interface provides a high trap density that can be utilized for memory devices. Most of the results presented here are from back-side trapping devices in which the silicon body is approximately 15 nm thick, the front oxide is 6 nm, the back ONO stack is approximately 3, 4 and 7 nm, respectively. The active area of the devices is 100 nm x 50 nm and 200 nm x 100 nm (W x L). The devices were measured in weak inversion at room temperature with a drain current of ~ 1 nA.

RTS is most commonly observed when monitoring the drain to source current (or voltage) as a function of time when applying constant voltages to the different terminals as illustrated in Figure 6.1. The amplitude of the discrete switching can vary widely, depending on the proximity of the trap, the charged state of the trap when occupied and empty (it can be positive, negative or neutral in each state and it can also have more than one positive or negative charges) and the level of inversion of the channel. RTS amplitudes ranging from 0.01 % to 70 % of the drain current have been reported in the literature [22, 25]. In most cases, for a particular device, RTS is only observed for a relatively narrow range of voltages and when the current in the channel is moderate, in the range of 1 nA.

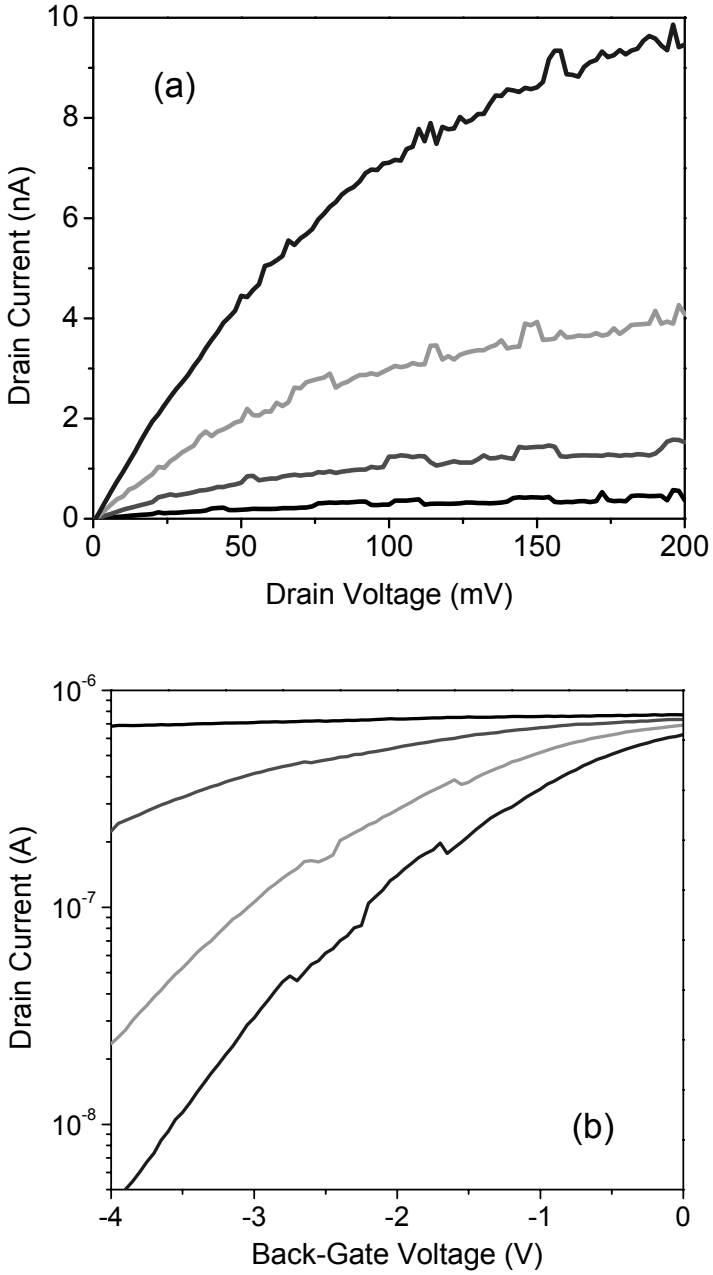



**Figure 6.2** Output **(a)** and transfer **(b)** characteristics for a back-side trapping device exhibiting RTS features. W = 100 nm, L = 50 nm. The front-gate voltage is 0, 0.2, 0.4 and 0.6 V in (a) and 0, -1, -2 and -3 V in (b).

The voltage range at which RTS can be observed, that depends on the energy level of the traps present and has to be searched carefully by monitoring the current or voltage as a function of time for different bias conditions. However, when the amplitude of an RTS event is large enough and when trapping occurs across a wide range of voltages (probably due to a larger number of traps in the system, or a wider distribution in energy) RTS can also be observed in the current-voltage characteristics of a device. Figure 6.2 illustrates one such case when RTS is observed in the transfer and output characteristics of a 50 nm physical gate length back-side trapping memory cell. Figure 6.3 shows the back-gate bias dependence of an RTS event due to a trap located in the back ONO stack and the occupational statistics of the individual trap as the gate bias is changed.

A parameter analyzer, with constant voltages applied to front-gate, back-gate, source, and drain, was used to record the RTS in the source-drain current as a function of time. A maximum of 10,000 points can be acquired with a minimum time interval of 80 μs. The sampling interval can be adjusted for a particular signal depending on its time scales in order to maximize the number of steps acquired while keeping the desired time resolution.

Due to the noise in the system, the two-level signals (interaction of a single electron with a single trap) have to be analyzed using a search algorithm. This algorithm was implemented in Matlab and takes as parameters the approximate low and high levels and a tolerance within which a data point is considered to be ‘up’ or ‘down’. The tolerance has to be adjusted depending on the noise level of the signal. The code generates a step function (Figure 6.4) that follows the original steps but doesn’t have noise and is used to verify the accuracy of the procedure by comparing it with the original trace. The step function is then used to compute the statistics relative

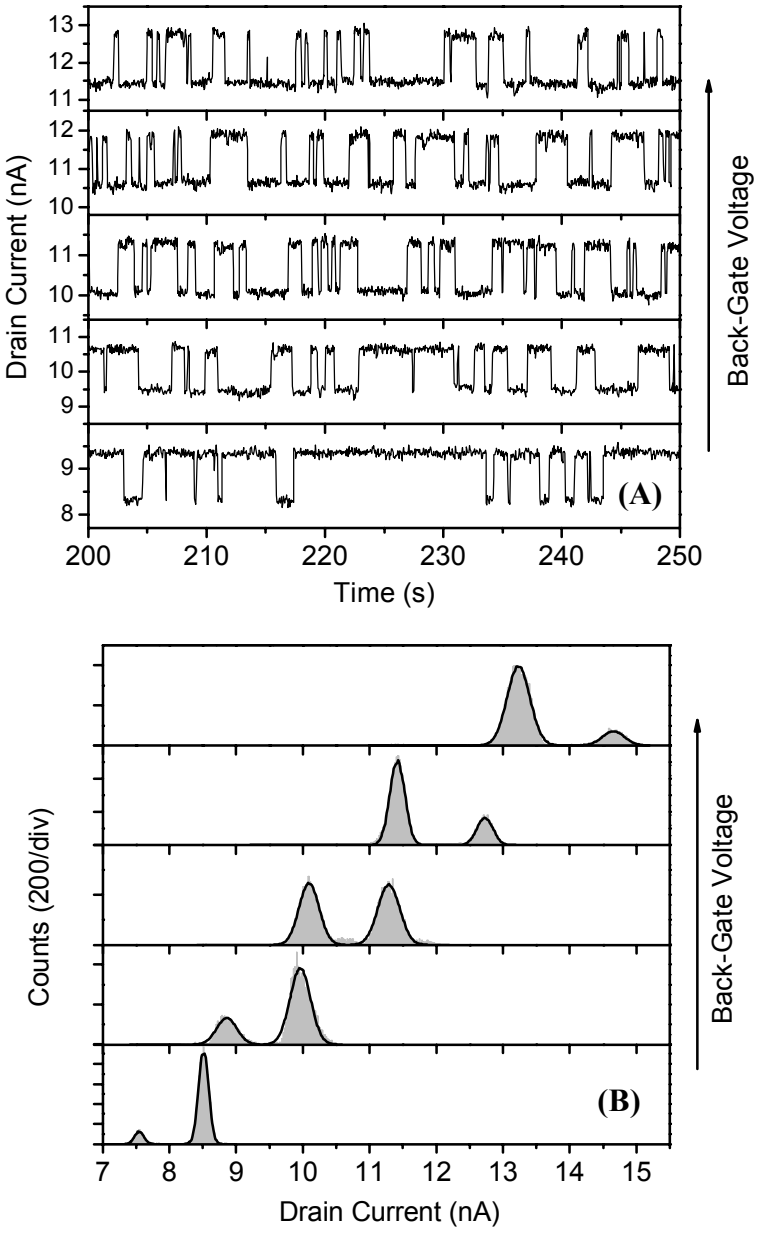



**Figure 6.3 (a)** Random Telegraph Signal at the back silicon interface in a back-side storage memory. The physical gate length is 50 nm and the width is 200 nm. The back-gate voltage is increased from -3.2 V to -2.8 V in 0.1 V steps. $V_{FG}$ = -2 V, $V_D$ = 10 mV and $V_S$ = 0 V.
**(b)** Histograms of the same RTS illustrating the occupation probability of the trap as the back-gate voltage is increased.

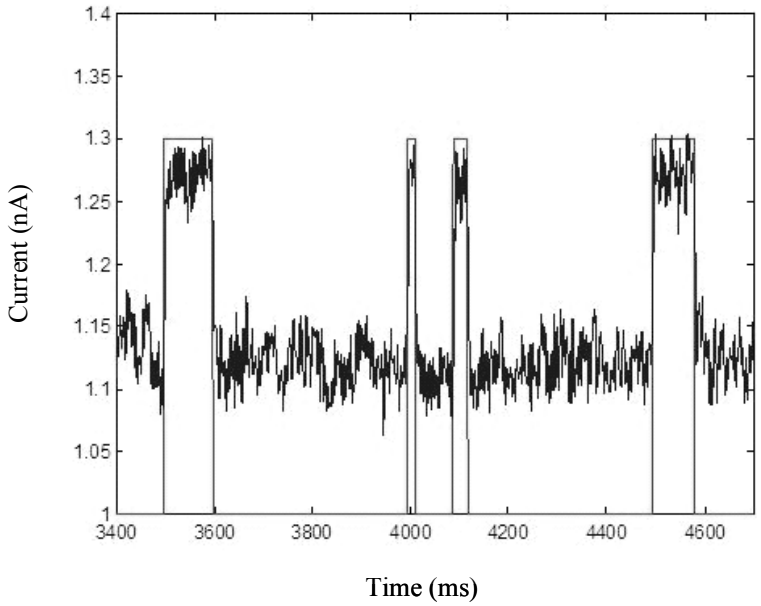



**Figure 6.4** RTS trace with the step function obtained from the Matlab code. The step function is used to calculate the statistics of the RTS trace.

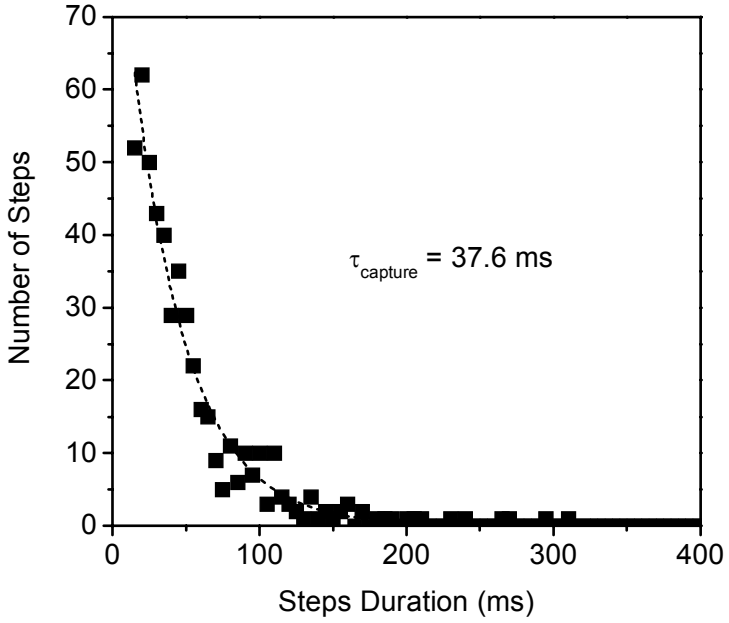



**Figure 6.5** Histogram of the duration of the steps for a particular RTS signal. The steps duration follows a Poisson distribution.

to the trace: average high and low values, average capture time (average time spent in the high-current state), average emission time (average time spent in the low-current state), and the total number of steps. In random switching events, the duration of the steps follows a Poisson distribution (Figure 6.5), and the average times are the averages of the distributions. The traces used have a minimum of approximately 200 steps per trace, i.e. 400 switching events, for reliable statistical information.

## 6.3 Results and Discussion

We can determine if a particular RTS signal corresponds to a trap in the front or back dielectric by its dependence on the front- or back-gate bias: the average capture (emission) time will decrease (increase) with the increase of the respective gate bias. The position of a trap is estimated by:

$$\frac{d(\ln \bar{\tau}_c / \bar{\tau}_e)}{dV_G} = -\frac{q}{kT}\frac{x_T}{t} \qquad (1)$$

where $\bar{\tau}_c$ and $\bar{\tau}_e$ are the average capture and emission times, $x_T$ is the position of the trap relative to the Si-$SiO_2$ interface and $t$ is the dielectric thickness. This equation is derived from the principle of detailed balance (same rate for capture and emission processes) together with the occupation probability of the trap as being given by Boltzmann distribution. It follows that the trap occupational probability is a function of the gate voltage through the difference between the energy level of the trap and the bottom of the conduction band in silicon. The underlying assumption is that the silicon surface potential changes slowly with the gate bias compared to the energy level of the trap in the oxide which is true following the onset of inversion [20-22].

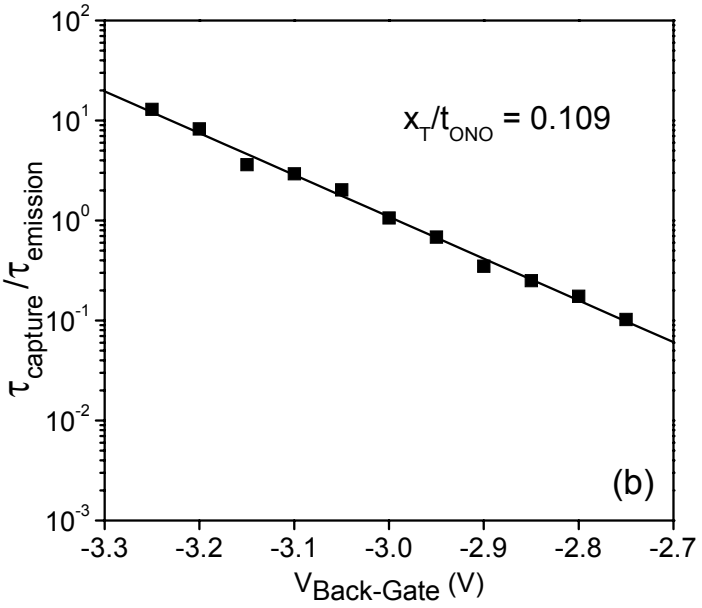



**Figure 6.6** Average capture and average emission times ratio as a function of the back-gate bias for the RTS signal shown in Figure 5.3. The position of the trap responsible for this signal is determined from the slope of the fitted line. $t_{ONO}$ is the equivalent thickness of the back ONO stack.

The location of the trap is determined based on the gate voltage dependence of the ratio of the capture and emission times. For the RTS signal in Figure 6.3 this dependence is plotted in Figure 6.6. From (1) this trap is found to be located 1.3 nm away from the Si-$SiO_2$ interface which corresponds to an oxide trap within the tunneling oxide. The RTS signals in this case have capture and emission time constants between 200 ms and 5 s. Figure 6.7 shows an example of an RTS observed in a device in which the gate dielectric is $SiO_2$ only, and its bias dependence. This trap is located 0.4 nm from the Si-$SiO_2$ interface and is therefore an interface trap.

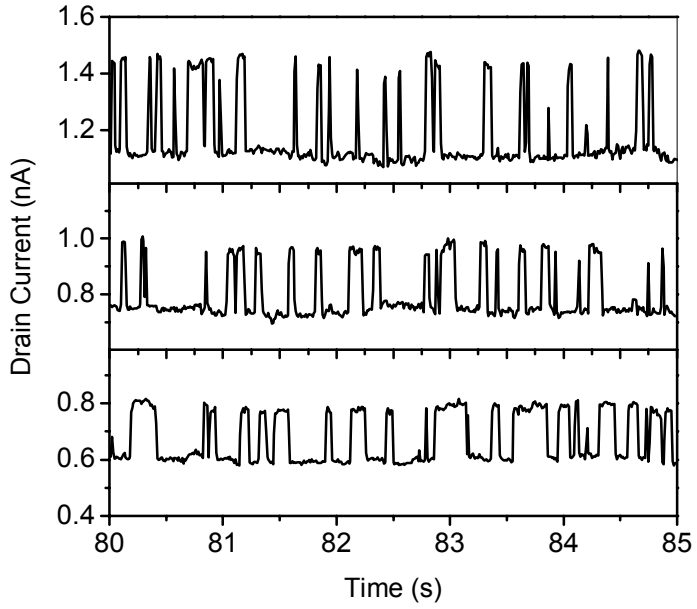



**Figure 6.7** RTS in the front interface (oxide only) of a back-side trapping memory device for three different front-gate voltages, - 0.94 V (top), - 0.98 V (middle) and -1.0 V (bottom).

Table 6.1 summarizes results from RTS signals observed at silicon interfaces with oxide-only and with ONO gate dielectrics. Traps observed in the front interface (oxide-only) show average capture and emission time constants between 8 and 340 ms, faster than those observed in the back interface (ONO) which show time constants in the order of 200 ms – 5 s.

**Table 6.1** RTS signals observed in oxide-only and ONO gate stack structures. The trap in the ONO stack is slower than those observed in oxide-only gate stacks.

| Gate dielectric | Time constants | Trap position ($x_t/t$) | Distance from Si |
|---|---|---|---|
| ONO (3 / 4 / 7 nm) | 200 ms – 4.8 s | 0.109 | 1.3 nm |
| Oxide (6 nm) | 8 ms – 30 ms | 0.067 | 0.4 nm |
| Oxide (6 nm) | 70 ms – 340 ms | 0.122 | 0.9 nm |
| Oxide (7 nm) | 14 ms – 190 ms | 0.242 | 1.9 nm |

Since both types of traps (in front oxide and back ONO) correspond to traps at the Si-$SiO_2$ interface and in the oxide, the differences in the observed time constants may be due to different surface properties of the front and back silicon interfaces. Prior RTS results in the Si-$SiO_2$ system have shown that traps can be shallow, located at the interface, or deeper, in the silicon dioxide within ~ 2 nm from the Si-$SiO_2$ interface. Our measurements on multiple interfaces also show shallow (located at the interface) and deeper traps. Shallow traps are faster, as expected and deeper traps can be both slow and fast traps. This is probably due to the energy level of the traps that can result in higher or lower transition probabilities for the same position in the dielectric. However, in any of these cases, the traps are within 2 nm of the interface.

Figures 6.8 and 6.9 show complex RTS signals observed in back-side trapping devices and in dual-oxide transistors. In Figure 6.8 two RTS signals, one fast and one slow, with approximately the same amplitude, are superimposed. The two events appear to be independent since the fast one appears in both the low and the high states of the slow one. The fast event turns on and off at random times during the measurement. Figure 6.9 illustrates a multi-level RTS that is due to either multiple traps or to a single trap with multiple occupational levels. These types of complex RTS are often observed but their interpretation is rather complicated due to the multitude of possible causes.

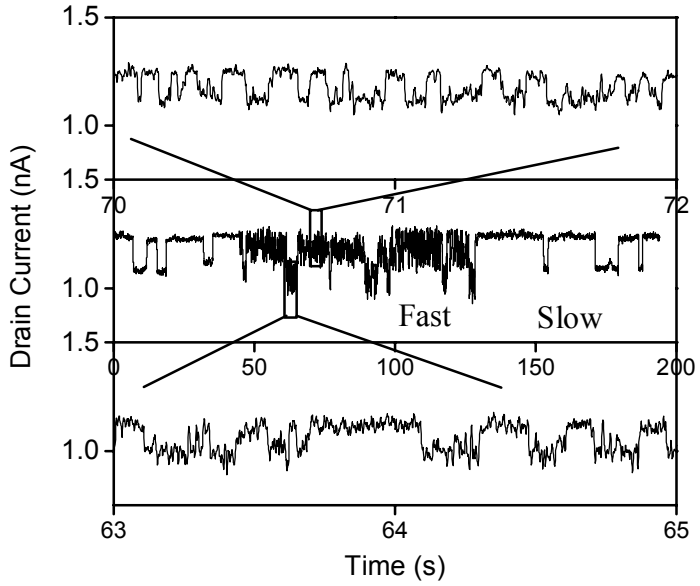



**Figure 6.8** Fast and slow RTS events in a dual oxide device. The upper and lower traces are zoomed in time windows indicated in the center trace.

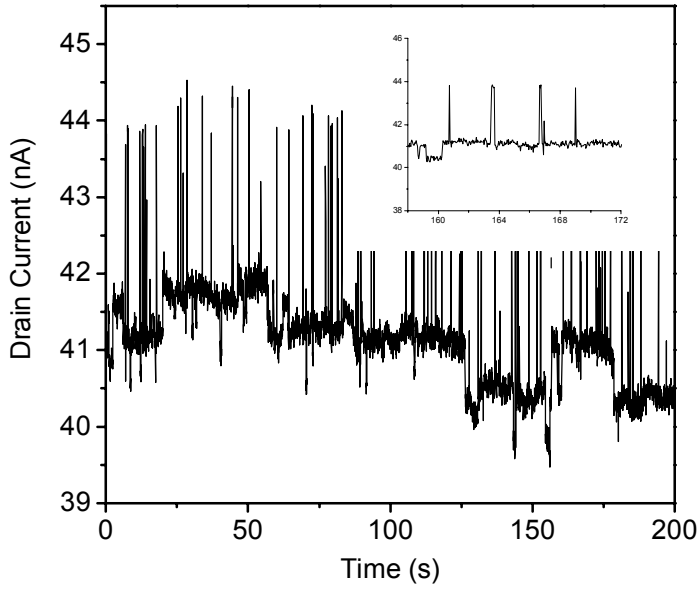



**Figure 6.9** Multi-level RTS signal at the back interface (ONO) of a back-side trapping device. $V_{BG}$ = - 1.45 V, $V_{FG}$ = -2 V, $V_D$ = 10 mV and $V_S$ = 0 V. The inset is a zoomed in time window showing three levels of the signal. W = 100 nm, L = 50 nm.

## 6.4 Summary

Random Telegraph Signals from both front and back silicon interfaces of back-side trapping memory devices were investigated. The traps observed both in the front (oxide only) and back (ONO) are located between 0.4 and 1.9 nm from the Si-$SiO_2$ interface, corresponding to interface and bulk traps in the oxide, in agreement with available literature on RTS and individual traps location. RTS due to deeper traps in the oxide or in the nitride layer, where the memory-related trapping takes place under more energetic injection conditions, were not observed.

In charge trapping devices with thinner tunneling oxides, it should be possible to observe RTS due to traps responsible for the memory properties, possibly with distinct characteristics from those in the Si-$SiO_2$. However, RTS arises from both capture and release processes, and observation of RTS signals would also indicate a very short storage time in such memory structures.

# Chapter 7

# Summary and future perspectives

## 7.1 Summary

Back side charge trapping devices for scalable silicon non-volatile memories were fabricated and characterized for the first time. The proposed device is based on storage of charge in silicon nitride traps in the back of a thin single crystal silicon channel. The new geometry allows the device to be operated as a scaled non-volatile memory device and as a scaled high performance transistor, suitable for logic. The device is written and erased by applying an electric field between the back gate and source and drain, high enough to cause charge to tunnel between the silicon channel and the trapping layer. When there is no voltage applied, charge cannot tunnel and is retained in the silicon nitride traps, hence the non-volatility of the memory. Charges stored in the silicon nitride traps change the potential of the silicon channel resulting in a threshold voltage shift of the device that is sensed using the front gate. As a result of the decoupling of read (front) and write/erase (back) functions, the front transistor can be scaled to very short gate lengths without compromising the memory characteristics of the device which depend only on the back gate stack. During the write and erase processes, with high voltages, the field in the front (read) oxide is minimized due to inversion/accumulation layers at the back silicon interface.

The fabrication process for these devices was described. The substrate, ‘Silicon on ONO’, was prepared using a Smart-Cut based technique. Optical lithography was used to define the non-critical dimensions patterns, alignment marks, vias and metal contacts and electron beam lithography was used to pattern the active area and the gate

of the devices. The transistors fabrication follows standard CMOS techniques. The fabrication process for back side trapping memory devices is fully compatible with CMOS processes. Like with current flash memory devices, this is a critical advantage compared to alternative emerging technologies that make use of non-standard CMOS materials or processes.

The fabricated devices show good front and back interface transistor characteristics down to 20 nm gate length. The use of thin silicon body and thin front and back gate dielectrics allows efficient coupling of each interface with the corresponding back-gate. The availability of two gates results in improved channel potential control at very short gate lengths and can be used for threshold voltage tuning for power adaptive applications.

Back side trapping storage was demonstrated in devices down to 50 nm gate length. Retention time, cycling endurance and writing and erasing times in these devices were characterized and found to be comparable to those of conventional front side trapping memories.

Extraction of mobility at both front and back interfaces in these devices indicates that both interfaces have very similar properties. It was also concluded that the mobility is not affected by charge stored in traps in close proximity to the channel. The values obtained for electron mobility for both interfaces show that the mobility in charge trapping devices is not degraded nor improved when compared to mobility in conventional silicon MOSFETs.

Random Telegraph Signals were used to determine the location of individual traps in back side trapping devices. Only traps located within ~ 2 nm from the front and back silicon interface were observed, corresponding to traps within the silicon oxide and at the silicon – silicon oxide interface.

## 7.2 Future perspectives

The possibility of further scaling for silicon non-volatile memories with back side trapping storage was demonstrated by the improved sub-threshold slope of the front transistors in written and erased states, compared to the sub-threshold slope of the back interface transistors. These results were recently confirmed by Ranica *et al* [26] with high performance transistor and memory operation with back side trapping devices at 50 nm gate length. In addition, the possibility of storing two bits per transistor in back side trapping memories was demonstrated in this work, effectively doubling the memory density.

Back side trapping devices can be further improved by introducing new materials for the charge trapping layer and for the injection and blocking barriers. The standard material for charge trapping devices has been silicon nitride due to its high density of traps (SONOS). However, as devices are scaled to smaller dimensions, a higher trap density material will be necessary to ensure device to device reproducibility. The trap density in silicon nitride varies between $10^{12}$ $cm^{-2}$ and $10^{13}$ $cm^{-2}$. Assuming a trap density of $10^{13}$ $cm^{-2}$ there will be only 40 electrons in a 20 nm x 20 nm area device and 10 electrons in a 10 nm x 10 nm device. As a result, alternative materials with higher density of traps will be necessary to scale charge trapping devices to the 10 nm gate length range. Regarding the injection and blocking barriers, materials other than silicon oxide can also result in better memory characteristics such as programming times and charge retention.

I will mention some examples of current research to identify new materials for improved non-volatile memory devices. One is the work of Tan *et al.* in which HfAlO (hafnium aluminum oxide) is used as the charge trapping layer with significant improvements over silicon nitride [27]. HfAlO combines the fast programming

provided by $HfO_2$ with the improved retention characteristics of $Al_2O_3$. Another example is the work of She *et al.* Here silicon nitride is used as the tunneling barrier with better retention and endurance compared to silicon oxide, while allowing the use of lower write and erase voltages [28]. Another interesting possibility of band engineering was proposed by Korotkov and Likharev [29]. They suggested the use of crested barriers, barriers in which the barrier height is field dependent. A lower barrier is created under high fields, leading to faster writing and erasing times, and a higher barrier is created under low fields, leading to longer charge retention time.

The vast amount of research that is currently oriented to the scaling of transistor based non-volatile memory devices will probably continue to result in higher density and lower power memory devices. Future devices will likely combine novel geometries, such as back side trapping or FinFet structures [30], improved materials and multi-level storage capabilities. The fabrication processes and architecture schemes for flash memory will tend to become more complex as devices are scaled to smaller dimensions, until a new technology matures and becomes a viable alternative.

RELATED PUBLICATIONS

H. Silva and S. Tiwari, “Individual trap characterization in charge trapping memories using random telegraph signal: depth limit in insulators”, *submitted* (2005)

H. Silva and S. Tiwari, “Back-side storage non-volatile memories: ultra-thin silicon single crystal silicon layers with complex thin film structure underneath”, Mater. Res. Soc. Symp. Proc. Fall 2004, 830, D1.4.1 (2005)

H. Silva, M. K. Kim, U. Avci, A. Kumar and S. Tiwari, “Nonvolatile silicon memory at the nanoscale”, MRS Bulletin, November 2004, 845-851 (2004)

H. Silva and S. Tiwari, “A nano-scale memory and transistor based on back-side trapping”, IEEE Transactions on Nanotechnology, 3, 2, 264-269 (2004)

H. Silva, M. K. Kim and S. Tiwari, “Scaled front-side and back-side SONOS memories”, 2003 IEEE Intl. SOI Conf. Proc., 105-106 (2003)

REFERENCES

[1] A. Fazio, “Future directions of non-volatile memory technologies”, Mater. Res. Soc. Symp. Proc. Fall 2004, 830, 3 (2005)

[2] J.-H. Park, S.-H. Hur, J.-H. Leex, J.-T. Park, J.-S. Sel, J.-W. Kim, S.-B. Song, J.-Y. Lee, J.-H. Lee, S.-J. Son, Y.-S. Kim, M.-C. Park, S.-J. Chai, J.-D. Choi, U.-I. Chung, J.-T. Moon, K.-T. Kim, K. Kim, B.-I. Ryu, “8 Gb MLC (multi-level cell) NAND flash memory using 63 nm process technology”, in Intl. Electron Devices Meeting Tech. Dig. 2004 p. 873 – 876 (2004)

[3] D. Kahng and S. M. Sze, “A floating gate and its application to memory devices”, Bell Syst. Tech. J., 46, 1288 (1967)

[4] H. A. R. Wegener, A. J. Lincoln, H. C. Pao, M. R. O'Connell, R. E. Oleksiak, “The variable threshold transistor, a new electrically alterable, non-destructive read-only storage device”, in Intl. Electron Devices Meeting Tech. Dig. (1967)

[5] S. Tiwari, F. Rana, H. Hanafi, A. Hartstein, E. F. Crabbe, K. Chan, “A silicon nanocrystals based memory”, Appl. Phys. Lett. 68, 10, 1377 (1996)

[6] A. Kumar and S. Tiwari, “Floating back gate electrically erasable programmable read-only memory (EEPROM)”, U.S. Patent 6248626 (2001)

[7] U. Avci, A. Kumar and S. Tiwari, “Back-floating gate non-volatile memory”, 2004 Intl. SOI Conf. Proc., 133–135 (2004)

[8] L. Selmi and C. Fiegna, Physical aspects of cell operation and reliability, in *Flash Memories*, edited by P. Cappellotti, C. Golla, P. Olivo, E. Zanoni, Kluwer Academic Publishers (1999)

[9] A. Kumar and S. Tiwari, “Scaling of Flash NVRAM to 10’s of nm by Decoupling of Storage From Read/Sense Using Back-Floating Gates”, IEEE Trans. on Nanotechnology, 1, 4, 247–254 (2002)

[10] M. Bruel, B. Aspar, B. Charlet, C. Maleville, T. Poumeyrol, A. Soubie, A. J. Auberton-Herve, J. M. Lamure, T. Barge, F. Metral, S. Trucchi, "Smart cut: a promising new SOI material technology", 1995 IEEE Intl. SOI Conf. Proc., 178-179 (1995)

[11] J.-P. Colinge, *Silicon-On-Insulator Technology: Materials to VLSI*, Kluwer Academic Publishers (1997)

[12] B. Aspar, C. Lagahe, H. Moriceau, A. Soubie, M. Bruel, A. J. Auberton-Herve, T. Barge, C. Maleville, "Kinetics of Splitting in The Smart-Cut Process", 1998 IEEE Intl. SOI Conf. Proc., 137-138 (1998)

[13] B. Aspar, M. Bruel, M. Zussy, A. M. Cartier, "Transfer of structured and patterned thin silicon films using the Smart-Cut(R) process", Electronics Letters 32, 21, 1985-1986 (1996)

[14] U. Avci and S. Tiwari, "Back-gated MOSFETs with controlled silicon thickness for adaptive threshold-voltage control", Electronics Letters 40, 1, 74–75 (2004)

[15] S. A. Campbell, "*The Science and Engineering of Microelectronic Fabrication*", Oxford University Press (1996)

[16] E. Lusky, Y. S.- Diamand, A. Shappir, I. Bloom, B. Eitan, "Traps spectroscopy of the Si3Ni4 layer using localized charge-trapping nonvolatile memory device", Appl. Phys. Lett., 85, 4, 669 (2004)

[17] Y. Taur and T. H. Ning, *Fundamentals of Modern VLSI Devices*, Cambridge University Press (1998)

[18] S. Takagi, A. Toriumi, M. Iwase, and H. Tango, "On the Universality of Inversion Layer Mobility in Si MOSFET's: Part I-Effects of Substrate Impurity Concentration", IEEE Trans. Electron Devices, 41, 12, 2357 (1994)

[19] U. Avci, A. Kumar, H. Liu and S. Tiwari, "Back-Gated SOI Technology: Power Adaptive Logic and Non-Volatile Memory Using Identical Processing", 34th European

Solid-State Device Research Conference Proc., 285–288 (2004)

[20] K. S. Ralls, W. J. Skocpol, L. D. Jackel, R. E. Howard, L. A. Fetter, R. W. Epworth, and D. M. Tennant, "Discrete resistance switching in submicrometer silicon inversion layers: individual interface traps and low frequency (1/*f*?) noise", Phys. Rev. Lett. 52, 3, 228-231 (1984)

[21] M. J. Kirton and M. J. Uren, "Noise in solid-state microstructures: a new perspective on individual defects, interface states, and low-frequency noise", Adv. Phys. 38, 367-468 (1989)

[22] H. H. Mueller and M. Schulz, Individual Interface Traps and Telegraph Noise *in* Characterization Methods of Submicron MOSFETS, Kluwer, Haddara (1996)

[23] O. Ce´spedes, G. Jan, M. Viret, M. Bari and J. M. D. Coey, "Random telegraph noise in a nickel nanoconstriction", Appl. Phys. Lett. 90, 10, 8433 (2003)

[24] M. Shima, Y. Sakuma, Y. Awano, and N. Yokoyama, "Random telegraph signals of tetrahedral-shaped recess field-effect transistor memory cell with a hole-trapping floating quantum dot gate", Appl. Phys. Lett. 77, 3, 442 (2000)

[25] H. M. Bu, Y. Shi, X. L. Yuan, J. Wu, S. L. Gu, Y. D. Zheng, H. Majima, H. Ishikuro, and T. Hiramoto, "Random telegraph signals and low-frequency noise in n metal–oxide–semiconductor field-effect transistors with ultranarrow channels", Appl. Phys. Lett. 76, 22, 3259 (2000)

[26] R. Ranica, A. Villaret, P. Mazoyer, S. Monfray, D.Chanemougame, P. Masson, C. Dray, P. Waltz, R. Bez, T. Skotnicki, "A new 40nm SONos structure based on backside trapping for nanoscale memories", Silicon Nanoelectronics Workshop (2004)

[27] Y. N. Tan, W. K. Chim, W. K. Choi, M. S. Joo, T. H. Ng and B. J. Cho, "High-K HfAlO Charge Trapping Layer in SONOS-type Nonvolatile Memory Device for High Speed Operation", Intl. Electron Devices Meeting Tech. Dig. 2004 (2004)

[28] M. She, H. Takeuchi and T.-J. King, "Silicon-Nitride as a Tunnel Dielectric for

Improved SONOS-Type Flash Memory", IEEE Elec. Dev. Lett. 24, 5, 309-311 (2003)

[29] A. Korotkov and K. Likharev, "Resonant Fowler-Nordheirn Tunneling through Layered Tunnel Barriers and its Possible Applications", Intl. Electron Devices Meeting Tech. Dig. 1999, 223-226 (1999)

[30] P. Xuan, M. She, B. Harteneck, A. Liddle, J. Bokor, T-J. King, "FinFet SONOS flash memory for embedded applications", Intl. Electron Devices Meeting Tech. Dig. 2003, 609-612 (2003)